PLANT SCIENCES

Tomato root specialized metabolites evolved through gene duplication and regulatory divergence within a biosynthetic gene cluster

Rachel E. Kerwin[1]*, Jaynee E. Hart[1], Paul D. Fiesel[1]†, Yann-Ru Lou[1,2], Pengxiang Fan[1,3], A. Daniel Jones[1,4], Robert L. Last[1,5]*

Tremendous plant metabolic diversity arises from phylogenetically restricted specialized metabolic pathways. Specialized metabolites are synthesized in dedicated cells or tissues, with pathway genes sometimes colocalizing in biosynthetic gene clusters (BGCs). However, the mechanisms by which spatial expression patterns arise and the role of BGCs in pathway evolution remain underappreciated. In this study, we investigated the mechanisms driving acylsugar evolution in the Solanaceae. Previously thought to be restricted to glandular trichomes, acylsugars were recently found in cultivated tomato roots. We demonstrated that acylsugars in cultivated tomato roots and trichomes have different sugar cores, identified root-enriched paralogs of trichome acylsugar pathway genes, and characterized a key paralog required for root acylsugar biosynthesis, *SlASAT1-LIKE* (*SlASAT1-L*), which is nested within a previously reported trichome acylsugar BGC. Last, we provided evidence that *ASAT1-L* arose through duplication of its paralog, *ASAT1*, and was trichome-expressed before acquiring root-specific expression in the *Solanum* genus. Our results illuminate the genomic context and molecular mechanisms underpinning metabolic diversity in plants.

INTRODUCTION

Plants synthesize and store a vast array of metabolites. Essential and ubiquitous metabolites, including sugars, amino acids, and lipids, are synthesized through highly conserved general, or primary, metabolic pathways. In contrast, most of the chemical diversity observed among plants is produced through taxonomically restricted specialized metabolic pathways that use general metabolic precursors (*1*). Specialized metabolites are commonly synthesized and stored in dedicated cells, tissues, or organs, and serve critical roles in herbivore defense (*2*–*5*), pollinator attraction (*3*, *6*, *7*), and abiotic stress mitigation (*8*–*10*). Specialized metabolite variation influences herbivory, disease, and fitness in the natural environment (*11*–*13*), reflecting their adaptive roles.

Specialized metabolic pathways often emerge through gene duplication and divergence, co-opting enzymes and precursors from other aspects of plant metabolism, including general metabolism (*14*–*17*). Typically, enzymes involved in specialized metabolism exhibit broader substrate specificities and narrower expression patterns than those of general metabolism (*15*, *16*, *18*). However, the molecular mechanisms by which specialized metabolic pathway genes acquire their distinct expression patterns remain poorly understood, representing an important gap in our knowledge.

The genes of some specialized metabolic pathways colocalize on chromosomes into biosynthetic gene clusters (BGCs) (*19*, *20*), which can facilitate pathway coregulation and localized metabolite production (*21*). Established BGCs promote coinheritance of clustered genes due to reduced recombination, thereby maintaining metabolic pathway integrity (*22*, *23*). Within BGCs, exchange of coding or regulatory sequences, for example, through recombination, can lead to gene duplication or altered gene expression patterns (*24*, *25*), yet the precise roles of BGCs in the evolution of specialized metabolic pathways, particularly in the context of regulatory divergence, are not fully elucidated.

Acylsugar biosynthesis is an ideal system to investigate the molecular mechanisms underlying metabolic pathway evolution. These specialized metabolites are synthesized in the tip cells of type I/type IV glandular trichomes of Solanaceae species (*26*–*29*) and protect aerial plant surfaces against herbivory, disease, and desiccation (*27*, *30*–*34*). Composed of a sugar core decorated with straight or branched acyl chains, acylsugars in cultivated tomato (*Solanum lycopersicum*) trichomes are assembled by four clade III BAHD (BEAT AHCT HCBT1 DAT) acyltransferase enzymes, ACYLSUGAR ACYLTRANSFERASE 1 (SlASAT1; *Solyc12g006330*), SlASAT2 (*Solyc04g012120*), SlASAT3 (*Solyc11g067670*), and SlASAT4 (*Solyc01g105580*), which sequentially esterify acyl chains to a central sucrose core (*35*–*37*). Acylsugar biosynthesis also involves enzymes co-opted from general metabolism, including ISOPROPYLMALATE SYNTHASE 3 (IPMS3; *Solyc08g014230*) (*38*), ACYLSUGAR ACYL-COA SYNTHETASE 1 (AACS1; *Solyc07g043630*) (*39*), and ACYLSUGAR ENOYL-COA HYDRATASE 1 (AECH1; *Solyc07g043680*) (*39*). Notably, *SlASAT1*, *SlAECH1*, and *SlAACS1* are colocalized to a trichome acylsugar BGC spanning syntenic regions of chromosomes 7 and 12 (*39*).

In addition to their presence in glandular trichomes, a recent study unexpectedly revealed acylsugars in the roots of cultivated tomato (*40*), suggesting the existence of a previously uncharacterized root-specific biosynthetic pathway. Here, we examined the structural diversity, biosynthesis, and evolutionary origin of these cultivated tomato root acylsugars.

Our results are consistent with the hypothesis that a distinct acylsugar biosynthetic pathway operates in tomato roots. Through detailed metabolite characterization, we found that cultivated tomato roots accumulate acyldissacharides that are distinct from the tri- and tetraacylsucroses synthesized in trichomes. Nuclear magnetic resonance (NMR) spectroscopic analysis demonstrated that the most

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA. [2]Department of Plant Biology, University of California, Davis, Davis, CA 95616, USA. [3]Department of Horticulture, Zhejiang University, Hangzhou, China. [4]Department of Chemistry, Michigan State University, East Lansing, MI 48824, USA. [5]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA.
*Corresponding author. Email: kerwinra@msu.edu (R.E.K.); lastr@msu.edu (R.L.L.)
†Present address: Minnesota Department of Health, Saint Paul, MN 55164, USA.

abundant cultivated tomato root acylsugar has a glucosylinositol disaccharide core. Analyzing coexpression across cultivated tomato tissues, we identified a set of root-expressed paralogs of characterized trichome acylsugar pathway genes. We functionally characterized *SlASAT1-LIKE* (*SlASAT1-L*; *Solyc07g043670*), the root-expressed paralog of *SlASAT1*, the first core trichome acylsugar biosynthesis pathway gene, yielding results consistent with a role early in root acylsugar biosynthesis. First, genetic ablation of *SlASAT1-L* led to loss of detectable root acylglycosylinositols, with no impact on trichome acylsucroses, providing strong evidence for separate biosynthetic pathways. Second, consistent with the hypothesis that *SlASAT1-L* catalyzes the first step of acylglycosylinositol biosynthesis, the enzyme acylated *myo*-inositol, but not sucrose, in vitro.

Notably, root-expressed *SlASAT1-L* and paralogs of two other trichome acylsucrose biosynthetic genes are nested within the previously identified trichome acylsugar BGC (*39*). We analyzed synteny and coexpression within the BGC, revealing that *ASAT1-L* arose via *ASAT1* duplication and was originally trichome-expressed before evolving root-specific expression in a subset of *Solanum* species. Further, we found a correlation between the spatial expression pattern of *ASAT1-L* and the site of inositol-based acylsugar accumulation among extant *Solanum* species. Our findings demonstrate that gene duplication, regulatory divergence, and enzyme evolution within a BGC can modify the cell- or tissue-specific localization of a biosynthetic pathway, shedding light on the molecular mechanisms contributing to metabolic diversity in plants.

RESULTS

Gene coexpression network analysis reveals candidates for root acylsugar biosynthesis

To interrogate whether the trichome acylsugar pathway is active in roots, we evaluated transcript abundance across cultivated tomato tissues using publicly available data (table S1). We found that characterized trichome acylsugar biosynthesis genes, including *SlASAT1-SlASAT4*, *SlIPMS3*, *SlAACS1*, and *SlAECH1*, were highly enriched in trichomes compared to shaved stems devoid of trichomes (table S2), consistent with previous reports (*35*–*39*). In contrast, these genes showed negligible expression in root tips, root hairs, and whole roots (table S2), suggesting that they are unlikely to be involved in root acylsugar biosynthesis.

To identify candidates for root acylsugar biosynthesis, we performed weighted gene coexpression network analysis (WGCNA) using 274 publicly available transcriptomes of various cultivated tomato tissues and developmental stages, including trichomes, root tips, root hairs, and whole roots (table S1). The resulting network was composed of 27 modules, each with an average of 941 genes (range, 204 to 1584) with similar expression patterns (Fig. 1, Table 1, and tables S2 and S3). Within the WGCNA network, we identified a 529-member module ("grey60") predominantly containing trichome-enriched genes, including characterized trichome acylsugar biosynthesis genes (Fig. 1, Table 1 and tables S2 and S3). This "trichome" module also included genes involved in trichome development and trichome-localized terpene biosynthesis, including *Hair2* (*Solyc10g078990*), *SlCYCB2* (*Solyc10g083140*), *Expression of Terpenoids 1* (*EOT1*, *Solyc02g062400*), and *Terpene Synthase 5* (*TPS5*, *Solyc01g105890*) (*41*–*45*). In addition, we identified three modules whose 1325 ("red"), 631 ("midnight blue"), and 483 ("light yellow") gene members were enriched principally in root hairs, root tips, and whole roots, respectively (Table 1 and tables S2 and S3). The red "root hair" module contained putative orthologs of characterized root hair regulators, *ROOT HAIRLESS 1* (*RHL1*, *Solyc12g010170*) and *ROOT HAIR DEFECTIVE 6* (*RHD6*, *Solyc12g088380*), further supporting its classification (*46*–*48*). Because acylsugar accumulation in cultivated tomato root hairs was previously documented (*40*), we searched the root hair module for putative root acylsugar pathway candidates.

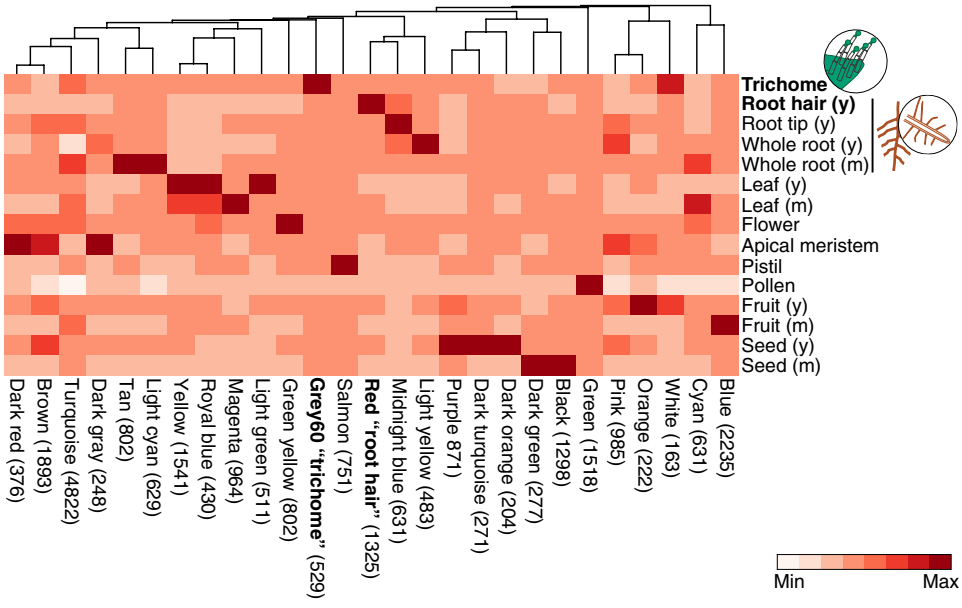

Fig. 1. Coexpression analysis across cultivated tomato tissues reveals root acylsugar pathway candidates. Heatmap shows average transcript abundance for genes in the 27 coexpression modules across 15 cultivated tomato tissues, with values centered and scaled within each module. The color gradient provides a visual marker to rank the transcript abundance from max (red) to min (white). Dendrogram represents coexpression similarity across modules based on correlations between their eigengene (first principal component) values. The number of genes in each module is indicated in parenthesis.

Table 1. WGCNA reveals groups of genes enriched in cultivated tomato tissues. Table summarizes results of Fisher's exact test identifying WGCNA coexpression modules with tissue-biased expression patterns across cultivated tomato tissues. Columns show gene counts in each of four discrete categories, (i) significantly enriched and in module (SI), (ii) not significantly enriched and in module (NI), (iii) significantly enriched and outside of module (SO), and (iv) not significantly enriched and outside module (NO), as well as odds ratio and *P* values from Fisher's exact test. The coexpression module with the highest odds ratio for each pairwise comparison is denoted with a "†." Results for the grey60 (trichome) and red (root hair) modules are in bold. RH_y, root hair (young); RT_y, root tip (young); RW_y, whole root (young); TR, trichome; SS, shaved stem. See table S1 for more information on transcriptome samples.

	TR versus SS						RH_y vs. TR						RT_y vs. TR						RW_y vs. TR					
Module	**SI**	**NI**	**SO**	**NO**	**Odds**	***P* value**	**SI**	**NI**	**SO**	**NO**	**Odds**	***P* value**	**SI**	**NI**	**SO**	**NO**	**Odds**	***P* value**	**SI**	**NI**	**SO**	**NO**	**Odds**	***P* value**
Black	48	1250	1095	23,971	0.84	8.9E−01	75	1223	3607	21,459	0.36	1.0E+00	67	1231	2920	22,146	0.41	1.0E+00	95	1203	3432	21,634	0.50	1.0E+00
Blue	39	2196	1104	23,025	0.37	1.0E+00	152	2083	3530	20,599	0.43	1.0E+00	48	2187	2939	21,190	0.16	1.0E+00	146	2089	3381	20,748	0.43	1.0E+00
Brown	16	1877	1127	23,344	0.18	1.0E+00	80	1813	3602	20,869	0.26	1.0E+00	243	1650	2744	21,727	1.17	1.9E−02	186	1707	3341	21,130	0.69	1.0E+00
Cyan	14	617	1129	24,604	0.49	1.0E+00	200	431	3482	22,251	2.97	7.0E−31	196	435	2791	22,942	3.70	1.2E−41	231	400	3296	22,437	3.93	2.1E−50
Dark green	14	263	1129	24,958	1.18	3.2E−01	51	226	3631	22,456	1.40	2.3E−02	44	233	2943	23,144	1.49	1.3E−02	55	222	3472	22,615	1.61	1.6E−03
Dark gray	4	244	1139	24,977	0.36	9.9E−01	15	233	3667	22,449	0.39	1.0E+00	20	228	2967	23,149	0.68	9.6E−01	49	199	3478	22,638	1.60	3.1E−03
Dark orange	4	200	1139	25,021	0.44	9.8E−01	30	174	3652	22,508	1.06	4.1E−01	25	179	2962	23,198	1.09	3.7E−01	33	171	3494	22,666	1.25	1.4E−01
Dark red	10	366	1133	24,855	0.60	9.7E−01	34	342	3648	22,340	0.61	1.0E+00	68	308	2919	23,069	1.74	6.6E−05	39	337	3488	22,500	0.75	9.7E−01
Dark turquoise	6	265	1137	24,956	0.50	9.8E−01	32	239	3650	22,443	0.82	8.7E−01	31	240	2956	23,137	1.01	5.1E−01	37	234	3490	22,603	1.02	4.7E−01
Green	60	1458	1083	23,763	0.90	7.9E−01	224	1294	3458	21,388	1.07	1.9E−01	120	1398	2867	21,979	0.66	1.0E+00	125	1393	3402	21,444	0.57	1.0E+00
Green yellow	39	763	1104	24,458	1.13	2.5E−01	85	717	3597	21,965	0.72	1.0E+00	75	727	2912	22,650	0.80	9.7E−01	87	715	3440	22,122	0.78	9.9E−01
Grey60 (trichome)	**383**	**146**	**760**	**25,075**	**86.35**	**0.0E+00** †	0	529	3682	22,153	0.00	1.0E+00	0	529	2987	22,848	0.00	1.0E+00	0	529	3527	22,308	0.00	1.0E+00
Light cyan	0	629	1143	24,592	0.00	1.0E+00	173	456	3509	22,226	2.40	1.8E−19	171	458	2816	22,919	3.04	1.2E−28	179	450	3348	22,387	2.66	6.3E−24
Light green	16	495	1127	24,726	0.71	9.3E−01	32	479	3650	22,203	0.41	1.0E+00	25	486	2962	22,891	0.40	1.0E+00	28	483	3499	22,354	0.37	1.0E+00
Light yellow	44	439	1099	24,782	2.26	3.2E−06	262	221	3420	22,461	7.78	1.5E−98	212	271	2775	23,106	6.51	5.3E−75	414	69	3113	22,768	43.87	5.0E−291 †
Magenta	30	934	1113	24,287	0.70	9.8E−01	110	854	3572	21,828	0.79	9.9E−01	73	891	2914	22,486	0.63	1.0E+00	100	864	3427	21,973	0.74	1.0E+00
Midnight blue	34	597	1109	24,624	1.26	1.1E−01	352	279	3330	22,403	8.49	3.4E−138	455	176	2532	23,201	23.68	2.7E−292 †	377	254	3150	22,583	10.64	5.0E−169
Orange	9	213	1134	25,008	0.93	6.3E−01	26	196	3656	22,486	0.82	8.6E−01	25	197	2962	23,180	0.99	5.4E−01	24	198	3503	22,639	0.78	8.9E−01
Pink	1	984	1142	24,237	0.02	1.0E+00	56	929	3626	21,753	0.36	1.0E+00	38	947	2949	22,430	0.31	1.0E+00	197	788	3330	22,049	1.66	2.5E−09
Purple	37	834	1106	24,387	0.98	5.8E−01	76	795	3606	21,887	0.58	1.0E+00	73	798	2914	22,579	0.71	1.0E+00	79	792	3448	22,045	0.64	1.0E+00
Red (root hair)	73	1252	1070	23,969	1.31	2.1E−02	**1059**	**266**	**2623**	**22,416**	**34.01**	**0.0E+00** †	556	769	2431	22,608	6.72	9.6E−192	507	818	3020	22,019	4.52	1.1E−121
Royal blue	16	414	1127	24,807	0.85	7.7E−01	30	400	3652	22,282	0.46	1.0E+00	28	402	2959	22,975	0.54	1.0E+00	29	401	3498	22,436	0.46	1.0E+00
Salmon	64	687	1079	24,534	2.12	2.3E−07	105	646	3577	22,036	1.00	5.1E−01	89	662	2898	22,715	1.05	3.4E−01	113	638	3414	22,199	1.15	9.7E−02
Tan	33	769	1110	24,452	0.95	6.5E−01	161	641	3521	22,041	1.57	8.8E−07	105	697	2882	22,680	1.19	6.4E−02	153	649	3374	22,188	1.55	2.8E−06
Turquoise	9	4813	1134	20,408	0.03	1.0E+00	92	4730	3590	17,952	0.10	1.0E+00	52	4770	2935	18,607	0.07	1.0E+00	48	4774	3479	18,063	0.05	1.0E+00
White	42	121	1101	25,100	7.91	3.1E−21	9	154	3673	22,528	0.36	1.0E+00	7	156	2980	23,221	0.35	1.0E+00	12	151	3515	22,686	0.51	9.9E−01
Yellow	45	1496	1098	23,725	0.65	1.0E+00	80	1461	3602	21,221	0.32	1.0E+00	81	1460	2906	21,917	0.42	1.0E+00	59	1482	3468	21,355	0.25	1.0E+00

Within the root hair module, we found a subset of genes belonging to the clade III BAHD acyltransferase family, including nine paralogs of trichome-expressed *SlASAT1-SlASAT4* (Fig. 2, fig. S1, and table S2). Notably, paralogs of other trichome acylsugar biosynthesis genes, including *SlAACS1*, *SlAECH1*, and *SlIPMS3*, were also root hair module members (figs. S2 to S4 and table S2). We speculated that these root-expressed paralogs may be involved in acylsugar biosynthesis, prompting a detailed investigation of these metabolites in cultivated tomato roots.

Cultivated tomato roots accumulate a class of acylsugars not detected in trichomes

To characterize acylsugar variation across tissues, we screened metabolite extracts from (i) homogenized roots and (ii) leaf, hypocotyl, and stem surfaces (i.e., trichomes) of 10-day-old cultivated tomato *cv.* M82 seedlings using liquid chromatography–mass spectrometry (LC-MS) combined with collision-induced dissociation (CID). Annotating metabolites based on chromatographic retention times and CID fragmentation patterns, we observed distinct acylsugar profiles in roots and trichomes (Fig. 3A, fig. S5, and tables S4 and S5). We identified 50 putative root-specific acylsugars, comprising 18 distinct molecular masses plus apparent isomers (table S5). Annotations for 22 acylsugars were further supported by data-dependent LC-MS/MS (table S5).

To describe the reported structures, we adopted the conventional acylsugar nomenclature (*26*): one or two letters designating the sugar core (e.g., "S" for sucrose) followed by the number of acyl chain attachments, and then a colon followed by the total number of acyl chain carbons, with individual acyl chain lengths identified in parentheses. For example, the most abundant acylsugar in cultivated tomato trichomes, S4:17 (2,5,5,5), is composed of a sucrose sugar core bearing four acyl chains with a total of 17 carbons (Fig. 3A).

All of the acylsugars detected in cultivated tomato root extracts are composed of a hexose-hexose disaccharide sugar core substituted with multiple acyl chains. While reminiscent of acylsucroses found in trichomes (*35*–*39*), these root acylsugars differ in acyl chain quantities, lengths, and positions. For instance, we readily detected acyldisaccharides with two to five acyl chain substitutions in roots, while detectable trichome acylsucroses only have three or four acyl chains (Fig. 3A and table S5). Further, root acylsugars are decorated with short acyl chains that are two (C2), five (C5), six (C6), or seven (C7) carbons in length, whereas trichome acylsugars are substituted with short (C2, C4, and C5) and medium (C10 to C12) acyl chains (*37*–*39*). To infer the distribution of acyl chains across hexose sugar core rings, we performed positive mode CID, which produces fragment ions arising from glycosidic bond cleavage (fig. S5) (*49*). This analysis revealed that root acylsugars bear two to three acyl chain substitutions on one hexose ring and one or two on the other. In contrast, trichome acylsucroses carry two to three acyl chains on the pyranose ring and just a single acyl chain on the furanose ring (*37*). Further, C2 and C5 acyl chains co-occur on the pyranose ring of trichome acylsucroses but reside on separate hexose rings of root acyldisaccharides. These details underscore the structural differences between trichome and root acylsugars in cultivated tomato (Fig. 3A and table S5).

On the basis of a previous report of acylsugars in cultivated tomato (*40*), we expected all detectable root acyldisaccharides to have a sucrose core. However, our LC-MS–based methods could not provide definitive structural resolution necessary to verify this claim because the individual hexose groups have the same masses. Therefore, to elucidate exact structures of cultivated tomato root acyldisaccharides, we attempted to purify the most abundant acylsugar in root extracts [labeled GI5:26 (2,5,6,6,7) in Fig. 3], yielding two isomers (fig. S6) with indistinguishable LC-MS/MS spectra in both positive- and negative-ion modes. With this sample, we used an assortment of NMR spectroscopy techniques for detailed structural analysis (data S1). Contrary to expectations, NMR spectra of our semipurified root acyldisaccharide sample revealed a sugar core composed of 4-*O*-β-glucopyranosyl *myo*-inositol (glucosylinositol) rather than sucrose (Fig. 3B and data S1). On the basis of changes in chemical shifts at the sugar carbons where acyl groups are attached, we inferred that

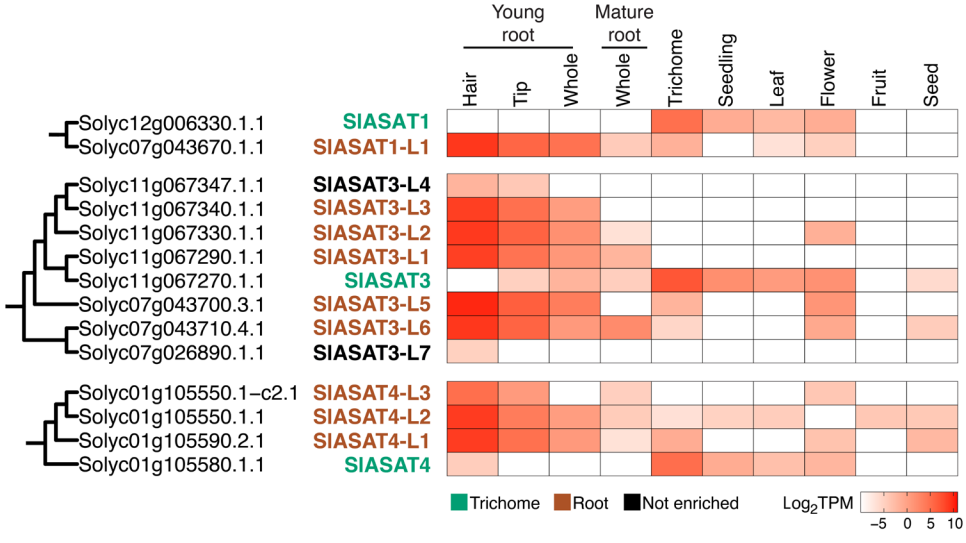

Fig. 2. Close paralogs of trichome ACYLSUGAR ACYLTRANSFERASE (ASAT) enzymes are expressed preferentially in cultivated tomato roots. Phylogenetic trees were subset from a maximum likelihood phylogeny of 98 predicted BAHDs (PF002458) in the cultivated tomato *cv.* M82 reference genome. The BAHD phylogeny was inferred from amino acid sequences using IQ-TREE v2.1.3 with 100,000 ultrafast bootstrap replicates (fig. S1). Heatmaps show absolute transcript abundance (log_2TPM) across 10 of 84 cultivated tomato tissues included in the analysis. The color gradient provides a visual marker to rank the transcript abundance from high (red) to low (white). TPM, transcripts per million.

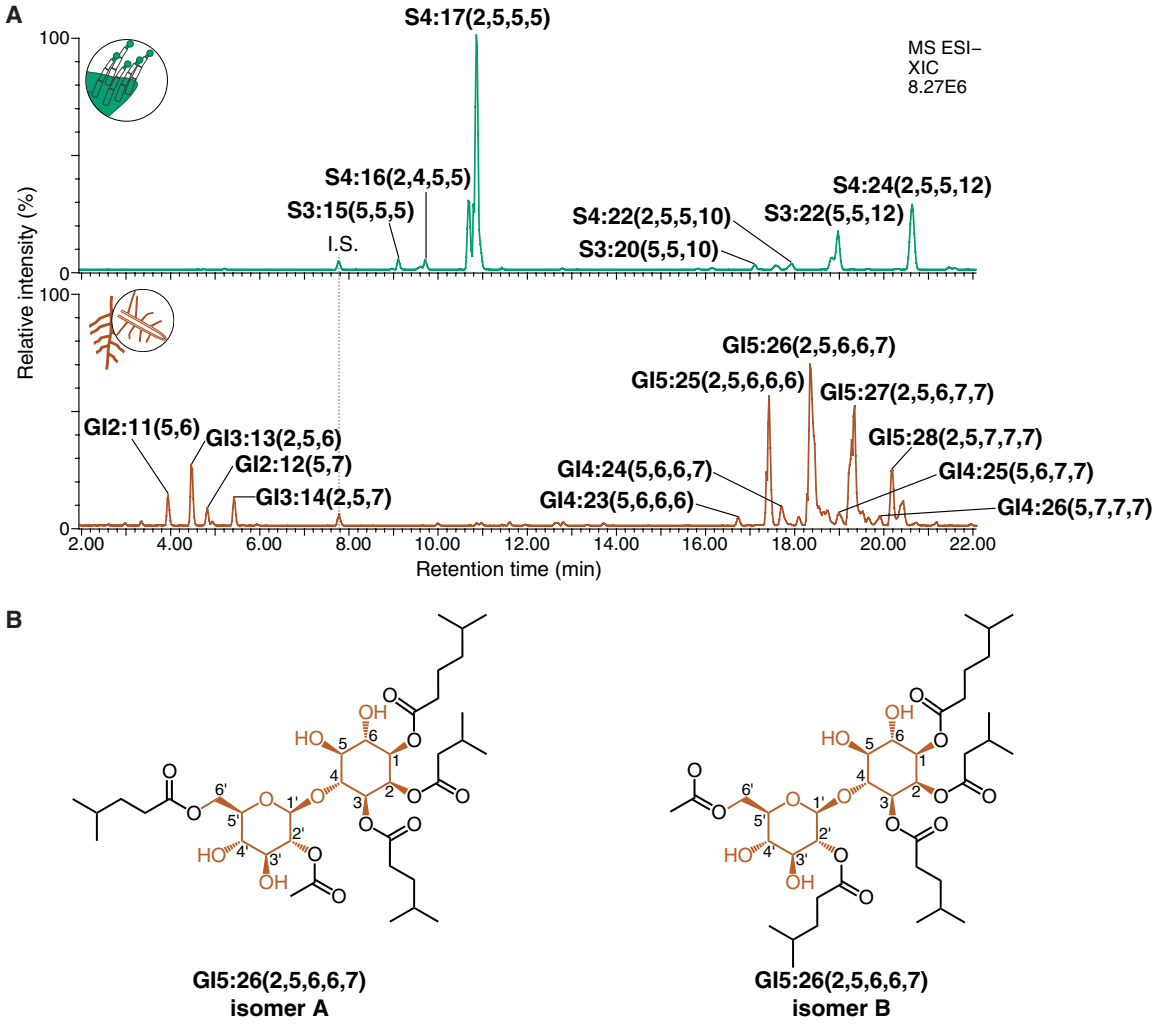

Fig. 3. Cultivated tomato roots accumulate a distinct class of acylsugars. (**A**) Metabolite profiling of cultivated tomato trichomes (top trace) and roots (bottom trace) reveal nonoverlapping acylsugars in each tissue. Extracted ion chromatogram (XIC) peaks specify the internal standard (I.S.) telmisartan (*m/z* 513.23) and formate [M + FA-H]$^-$ adducts of acylsugars labeled in figure (*m/z* values listed in tables S4 and S5). Acylsugars analyzed using LC-MS-CID in ESI$^-$ mode. Chromatograms scaled to the intensity of the major peak across both traces [i.e., S4:17 (2,5,5,5)]. (**B**) NMR structural elucidation of GI5:26 (2,5,6,6,7) revealed two isomers, each with a glycosylinositol sugar core. Acylsugars annotated using the following nomenclature: one- to two-letter sugar core abbreviation (i.e., "GI" for glycosylinositol); acyl chain count, colon, and total acyl chain carbons; and then individual acyl chain carbon lengths in parentheses.

the *myo*-inositol ring of GI5:26 (2,5,6,6,7) isomer A is substituted at positions 1, 2, and 3 with short, iso-branched iC7, iC5, and iC6 acyl chains, and positions 2′ and 6′ of the glucose ring are esterified by iC6 and a C2 acetyl group, respectively (Fig. 3B). Support for the hypothesis that GI5:26 (2,5,6,6,7) isomer B (fig. S6) differs by swapped placement of the C2 and iC6 acyl chains at positions 2′ and 6′ comes from the indistinguishable MS/MS spectra of the two isomers.

Given these data, we propose that cultivated tomato root acylsugars are composed of glycosylinositol cores, distinguishing them from the acylsucroses produced in trichomes, which is reflected in our metabolite annotations (Fig. 3, fig. S5, and table S5). Further, we hypothesize that acylsucroses and acylglycosylinositols in cultivated tomato are synthesized through independent biosynthetic pathways operating in trichomes and roots, respectively. To test this idea, we used CRISPR-Cas9 mutagenesis.

CRISPR-Cas9 gene editing validates a role for *SlASAT1-L* in root acylsugar biosynthesis

SlASAT1, which encodes the first step of trichome acylsugar biosynthesis, has one close paralog, *SlASAT1-L* (*Solyc07g043670*), which is highly and preferentially expressed in roots (Fig. 2). Further, *SlASAT1-L* is localized to an acylsugar BGC that harbors characterized trichome acylsugar genes, including *SlASAT1* (*39*). On the basis of these criteria, we hypothesized that *SlASAT1-L* plays a crucial role in root acylsugar biosynthesis. To compare the in vivo functions of *SlASAT1* and *SlASAT1-L*, we used CRISPR-Cas9–mediated mutagenesis in cultivated tomato *cv.* M82 (*50*). We recovered one *SlASAT1* mutant, *slasat1-1*, carrying a 123–base pair (bp) deletion, and two independent *SlASAT1-L* mutants, *slasat1-l-1* and *slasat1-l-2*, containing a 139- and 1-bp deletion, respectively (fig. S7).

We profiled metabolites in homozygous null mutants using the same LC-MS–based approach applied to wild-type tomato tissues. The

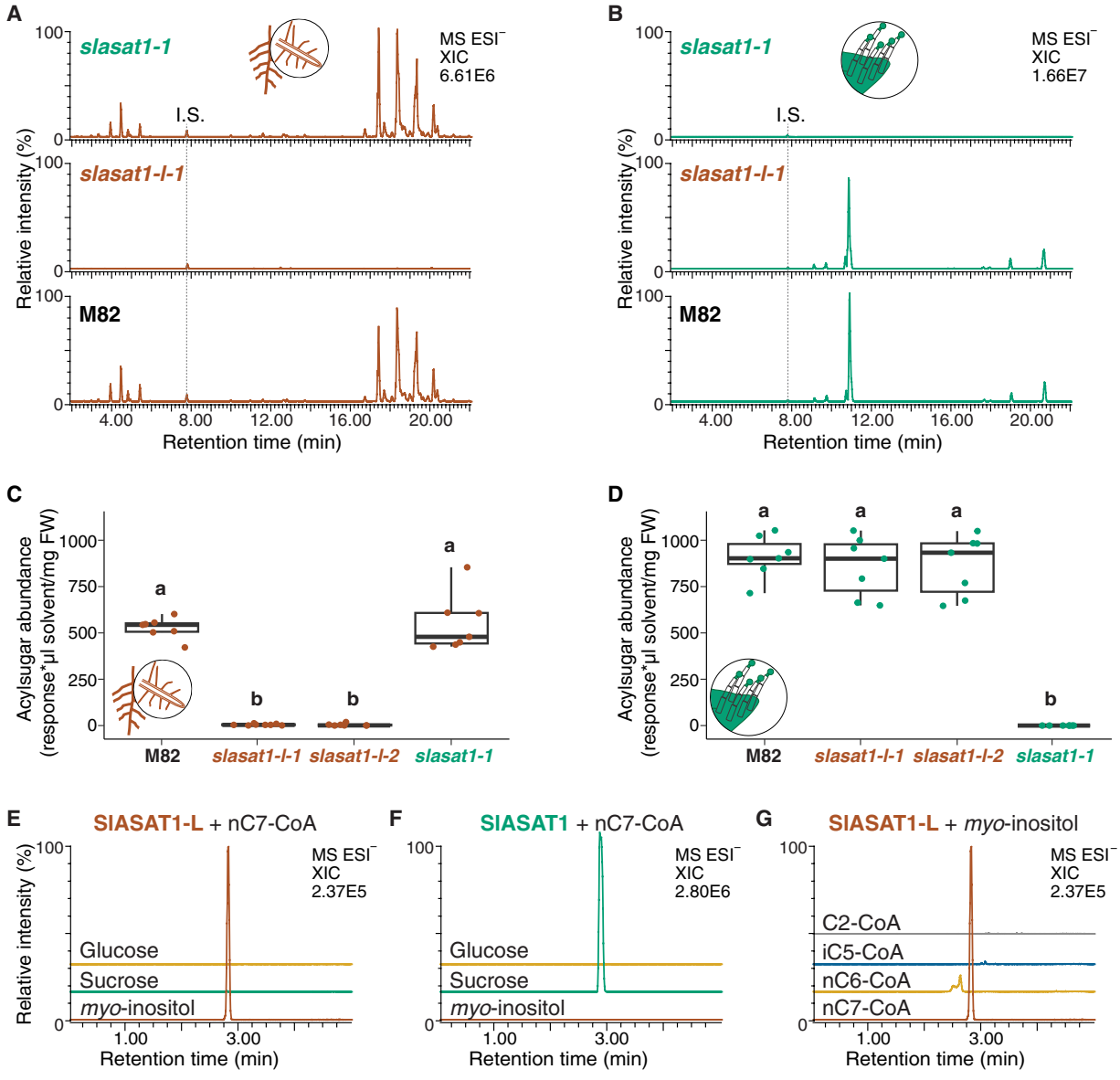

Fig. 4. SlASAT1-L is required for acylsugar biosynthesis in cultivated tomato M82 roots. (**A** to **D**) Genetic evidence that root-expressed *SlASAT1-L* is required for acylsugar biosynthesis in cultivated tomato [(A) and (C)] roots, but not [(B) and (D)] trichomes, and trichome-expressed *SlASAT1* is required for acylsugar biosynthesis in cultivated tomato [(B) and (D)] trichomes, but not [(A) and (C)] roots. (A) and (B) XIC peaks specify the internal standard telmisartan (*m/z* 513.23) and formate $[M + FA\text{-}H]^-$ adducts of root and trichome acylsugars labeled in Fig. 3 (*m/z* values listed in tables S4 and S5). (C) and (D) Boxplots illustrate total acylsugar abundances in (C) roots and (D) trichomes across *slasat1-l-1*, *slasat1-l-2*, *slasat1-1*, and wild-type M82. Acylsugar abundance in each sample was quantified by integrating peaks illustrated in (A) and (B) and normalizing to the internal standard area, solvent volume (in microliters), and tissue fresh weight (in milligrams). Letters above boxplots reflect Tukey's post hoc test results (table S6). (**E** to **G**) Analysis of SlASAT1-L and SlASAT1 substrate preferences in vitro. To test acyl acceptor preferences, (E) SlASAT1-L and (F) SlASAT1 were assayed with nC7-CoA and glucose, sucrose, or *myo*-inositol. (G) To test acyl donor preferences, SlASAT1-L was also assayed with *myo*-inositol and acetyl-, iC5-, nC6-, or nC7-CoA. XIC peaks specify the formate $[M + FA\text{-}H]^-$ adducts of G1:7, S1:7, I1:7, I1:6, I1:5, and I1:2 (*m/z* values listed in table S4). Acylsugars were analyzed using LC-MS in ESI^- mode. Stacked chromatograms in (A) and (B) and (E) to (G) are scaled to the intensity of the major peak in each panel.

results were notable: CRISPR-Cas9 inactivation of *SlASAT1-L* resulted in complete loss of detectable root acylsugars, while the trichome acylsugar levels remained unaffected (Fig. 4 and tables S6 and S7). In contrast, genetically ablating *SlASAT1* led to the expected total loss of detectable trichome acylsugars but had no effect on root acylsugars (Fig. 4 and tables S6 and S7). Notably, while trichome acylsugars were detected as minor constituents in a subset of wild-type M82 root extracts, they were completely absent in all *slasat1-1* tissues (table S6). This result strongly suggests that trichome exudate contamination before or during tissue collection was the source of sporadic and minor trichome acylsugars in root extracts. The absence of root acylsugars in the *slasat1-l* mutants is consistent with the hypothesis that SlASAT1-L enzyme catalyzes an early step in root acylsugar biosynthesis. These results also provide compelling evidence that independent biosynthetic pathways synthesize acylsugars in cultivated tomato trichomes and roots.

In vitro characterization of root-expressed SlASAT1-L protein

On the basis of sequence similarity between SlASAT1-L and SlASAT1 and genetic evidence that *SlASAT1-L* is required for acylsugar accumulation in cultivated tomato roots, we hypothesized that SlASAT1-L catalyzes the first acylation in root acylsugar biosynthesis. To test this idea, we performed in vitro enzyme assays using recombinant SlASAT1-L purified from *Escherichia coli*. Initially, we intended to use the glycosylinositol core extracted from root acylsugars as a substrate, but it proved to be highly labile in aqueous solution. Therefore, we examined the substrate preferences of SlASAT1-L using sucrose as a representative disaccharide along with *myo*-inositol and glucose, the individual components of the root sugar core (Fig. 4). For comparison, we performed these assays concurrently using trichome-expressed SlASAT1 and SqASAT1H, the trichome-expressed *Solanum quitoense* ortholog of SlASAT1-L (Fig. 4 and fig. S8) (*51*). We found that SlASAT1-L produced monoacylsugars in vitro using *myo*-inositol as an acyl acceptor, but not using glucose (Fig. 4E). In addition, SlASAT1-L could acylate sucrose in vitro, but only by increasing the acceptor concentration 10-fold, to 10 mM (fig. S8A). Further, SlASAT1-L substrate specificity reflects that of its ortholog SqASAT1H, which preferentially acylates *myo*-inositol in vitro (fig. S8B), but differs from SlASAT1, which only acylates sucrose under these assay conditions (Fig. 4F). This similarity between the two ASAT1-L orthologs is reinforced by assays in which we gave both SqASAT1H and SlASAT1-L nC10-CoA and *myo*-inositol, the native substrates of SqASAT1H (fig. S8C). The products of both enzymes comigrated in reversed-phase LC-MS, consistent with the hypothesis that, given the same substrates, these enzymes produce indistinguishable products.

Root acylsugars contain C2, C5, C6, and C7 acyl chains (Fig. 4 and table S5). To determine which acyl chain additions could be catalyzed by SlASAT1-L, we conducted in vitro assays using acetyl-, iC5-, nC6-, and nC7-CoA acyl donors, and *myo*-inositol as acyl acceptor. Because of the absence of commercially available iC6- and iC7-CoA, we initially used nC6- and nC7-CoA as substitutes. SlASAT1-L esterified nC6 and nC7 acyl chains to *myo*-inositol while showing a clear preference for nC7-CoA as substrate (Fig. 4G). We observed no acylation activity using either acetyl- or iC5-CoA as acyl donors, suggesting that enzymes other than SlASAT1-L are responsible for transferring these moieties to root acylsugars in vivo. Upon synthesizing iso-branched iC6- and iC7-CoAs (*52*), we found that SlASAT1-L prefers iC7-CoA as an acyl donor over nC7-CoA, but has no apparent activity with iC6-CoA (fig. S8D). Together, the generation of monoacylsugars by SlASAT1-L in vitro and the phenotypes of *slasat1-l-1* and *slasat1-l-2* support the hypothesis that SlASAT1-L catalyzes the first acylation in root acylsugar biosynthesis.

The tomato acylsugar BGC contains both root- and trichome-expressed enzymes

Previous work from our group identified a Solanaceae-specific acylsugar BGC harboring BAHD, ECH, and ACS enzyme-encoding genes. This cluster is spread across two intraspecific syntenic regions, located on chromosomes 7 and 12 of cultivated tomato (*39*). Four genes within the cultivated tomato acylsugar BGC are trichome-expressed, including *SlASAT1*, *SlAECH1*, and *SlAACS1* involved in trichome acylsucrose biosynthesis (*35*, *39*). Intriguingly, we found that *SlASAT1-L*, the root-expressed *SlASAT1* paralog (Fig. 2) required for root acylsugar biosynthesis in cultivated tomato (Fig. 4), is also localized to the acylsugar BGC on chromosome 7. Upon closer inspection, we found that 7 of the 12 genes positioned within the cultivated tomato acylsugar BGC (*39*) are preferentially expressed in roots (table S2).

These observations prompted us to investigate the evolutionary relationship between formation of the acylsugar BGC in the Solanaceae and root acylsugar biosynthesis in cultivated tomato. We identified chromosomal regions that are syntenic with the acylsugar BGC across 14 eudicot species with sequenced genomes, including 11 Solanaceae species and 3 outgroup species, *Ipomoea trifida*, *Coffea canephora*, and *Vitis vinifera* (Fig. 5). We also assessed orthology across the 14 sequenced eudicots plus 2 other *Solanum* species, *S. nigrum* and *S. quitoense*, whose genomes have not been sequenced (table S8). Combining synteny and orthology information into our phylogenetic analysis extended our previous observations about the history of the acylsugar BGC (*39*) by revealing several key events associated with its formation, as described below (Fig. 5).

BAHD, *ECH*, and *ACS* genes localized to the acylsugar BGC in cultivated tomato separated into five orthologous groups (orthogroups; table S5). *ECH* and *ACS* homologs formed *Acylsugar ECH* (*AECH*; yellow arrows in Fig. 5) and *AACS* (blue arrows in Fig. 5) orthogroups, respectively. In contrast, *BAHD* homologs separated into three orthogroups, *ASAT1* (green arrows in Fig. 5), *ASAT1-L* (brown arrows in Fig. 5), and *ASAT3* (black arrows in Fig. 5). Consistent with the hypothesis that the acylsugar BGC emerged in the Solanaceae, we found *AECH*, *AACS*, *ASAT1*, *ASAT1-L*, and *ASAT3* orthologs in syntenic regions of other Solanaceae species, but not the outgroup species (Fig. 5 and table S8).

Our analysis suggests that formation of the acylsugar BGC began when an *AECH* homolog was recruited to the syntenic region in the Solanaceae ancestor, following divergence from the Convolvulaceae (Fig. 5, event 1). Thereafter, whole-genome triplication (WGT) within the Solanaceae (*53*) duplicated the syntenic region, resulting in an *AECH* copy on both chromosomes, designated A and B (Fig. 5, event 2). This hypothesis is supported by the fact that *Nicotiana* and *Petunia* carry an *AECH* homolog on chromosomes A and B (Fig. 5). Next, we propose that *ASAT1* arose within the nascent BGC on chromosome B, shortly after the WGT event (Fig. 5, event 3). Consistent with this hypothesis, we identified *ASAT1* orthologs in all 11 Solanaceae species, and always on chromosome B (Fig. 5). Thereafter, we infer that *AACS* entered the BGC on chromosome A after *Petunia* split from the rest of the Solanaceae (Fig. 5, event 4), given that *AACS* homologs are absent from the syntenic region in *Petunia*. Next, we propose that *ASAT3* was inserted into the acylsugar BGC on chromosome A in the common ancestor of *Lycium* and *Solanum* (Fig. 5, event 5), followed by subsequent duplications and deletions. Then, the cluster incurred substantial changes in the *Datura-Solanum* lineage. First, the *AECH* homolog on chromosome B was permanently lost (Fig. 5, event 6). Second, and most notably, chromosome A sustained *AECH* and *AACS* duplications and the appearance of *ASAT1-L* (Fig. 5, event 6). These changes effectively duplicated the full complement of acylsugar biosynthetic enzymes in the BGC, which may have potentiated formation of an independent pathway in roots. If true, root acylsugars would be restricted to species carrying root-expressed paralogs of trichome acylsugar pathway genes within the cluster, including *S. lycopersicum* and close relatives.

To test this hypothesis, we performed metabolite screening of root extracts from wild tomato (*Solanum pennellii acc.* LA0716), black

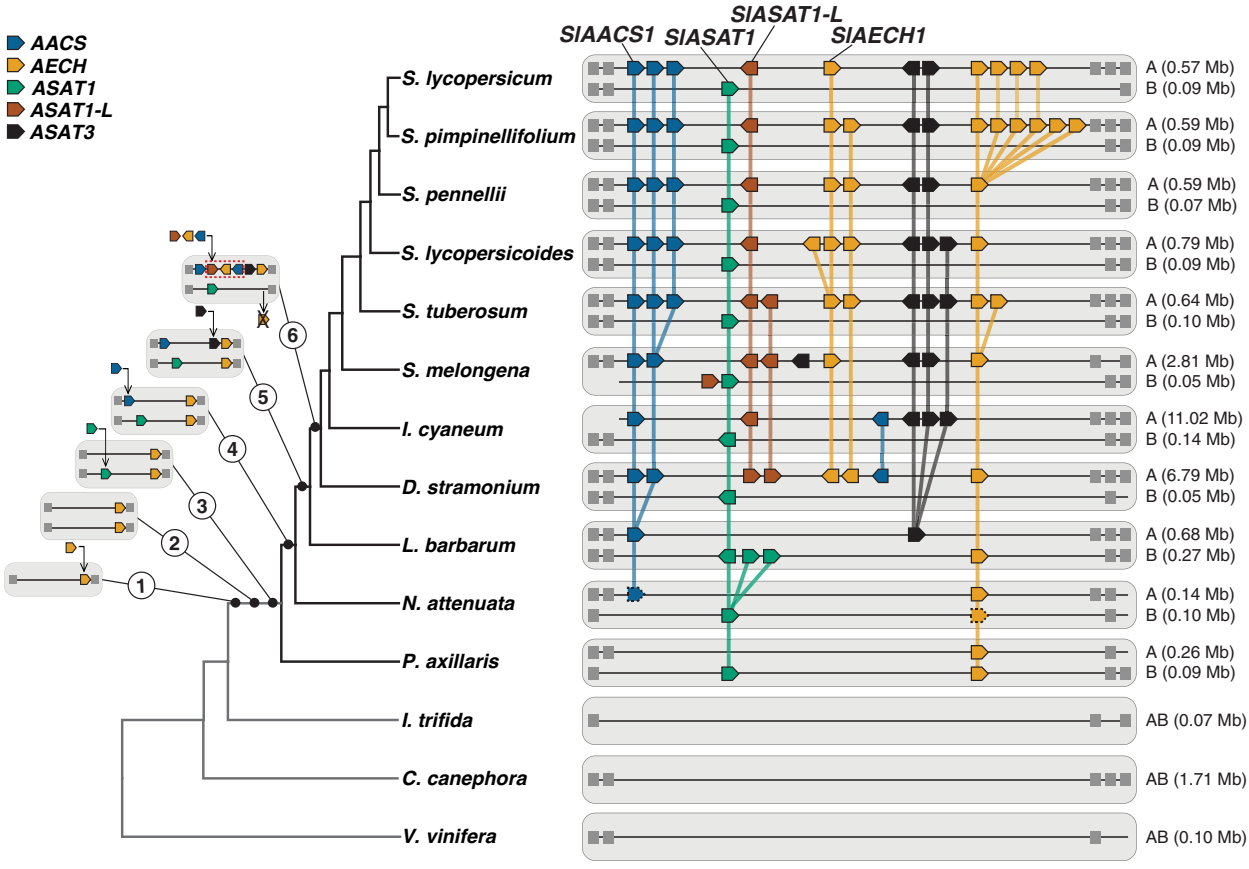

Fig. 5. *ASAT1-L* emergence within the acylsugar BGC is associated with root acylsugar pathway evolution. Schematic view of collinearity within a syntenic region shared across 16 eudicot species harboring the Solanaceae-specific acylsugar BGC. Members of five orthologous groups (orthogroups) identified by OrthoFinder (table S8) located within the acylsugar BGC are represented by colored arrows: blue, *AACS1* (OG0000391); green, *ASAT1* (OG0011176); brown, *ASAT1-L* (OG0015608); yellow, *AECH1* (OG0000616); and black, *ASAT3-L* (OG0000384). Anchor genes used to identify the syntenic region across species are represented by gray boxes. Orthologous syntenic chromosomal regions are denoted AB, A, or B, with the sizes listed in parentheses. Genes denoted with dashed outline were unannotated in the reference genome and identified using TBLASTN (see Materials and Methods for details). Numbered circles represent six key events leading to formation of the acylsugar BGC present in extant Solanaceae species. Characterized cultivated tomato acylsugar biosynthesis genes are labeled.

nightshade (*S. nigrum*), and brinjal eggplant (*Solanum melongena acc.* 67/3). Applying our established LC-MS–based methods, we detected acylsugars in root extracts from *S. pennellii* (fig. S9 and table S9), but not *S. nigrum* or *S. melongena* (fig. S10). Thereafter, we examined gene expression within the cluster using publicly available RNA sequencing (RNAseq) data from all three species (*39*, *54*, *55*). *S. pennellii* has one *ASAT1-L* ortholog within the cluster, *SpASAT1-L* (*Sopen07g023230*), which is highly expressed in roots, consistent with the observed acylsugar phenotype (Fig. 6 and tables S8 and S10). In contrast, *S. melongena* has three divergently expressed *ASAT1-L* orthologs within the cluster, *SmASAT1-L1* (*Smel4.1_07g013870*), *SmASAT1-L2* (*Smel4.1_07g013850*), and *SmASAT1-L3* (*SMEL4.1_12g018370*) (Fig. 6 and tables S8 and S11). *SmASAT1-L1* is highly and preferentially expressed in trichomes (Fig. 6 and table S11). *SmASAT1-L2* is enriched in roots, although its overall expression is very low (Fig. 6 and table S11). *SmASAT1-L3* expression is undetectable in both trichomes and roots (Fig. 6 and table S11). Although *S. nigrum* root RNAseq data were unavailable, we identified one *ASAT1-L* ortholog, *SnASAT1-L* (*Snig_c64578_g2*), which was highly enriched in trichomes versus shaved stems (Fig. 6 and tables S8 and S12). These findings show correlation between strong *ASAT1-L* root expression and root acylsugar biosynthesis. Furthermore, the phylogenetic distribution of *ASAT1-L* expression variation (Fig. 6) suggests that this gene was likely trichome-expressed when it first arose and subsequently evolved root-specific expression in a subset of *Solanum* species, ultimately leading to root acylsugar biosynthesis in cultivated tomato and its close relatives, including *S. pennellii*.

DISCUSSION

Trichome acylsugar biosynthesis has proven to be an exemplary system for studying metabolic evolution across hundreds of thousands to tens of millions of years. Several attributes make these specialized metabolites well-suited to evolutionary studies. First, acylsugars are structurally diverse, despite being synthesized from simple components, typically with a sugar core based on sucrose, glucose, or inositol decorated with C2-C12 acyl esters. Second, analysis of trichome-enriched metabolites, transcriptomes, and proteomes is facilitated by their presence on the surface of aerial tissues. Last, the metabolic enzymes can be heterologously expressed and tested in vitro to analyze their activity singly or in combination.

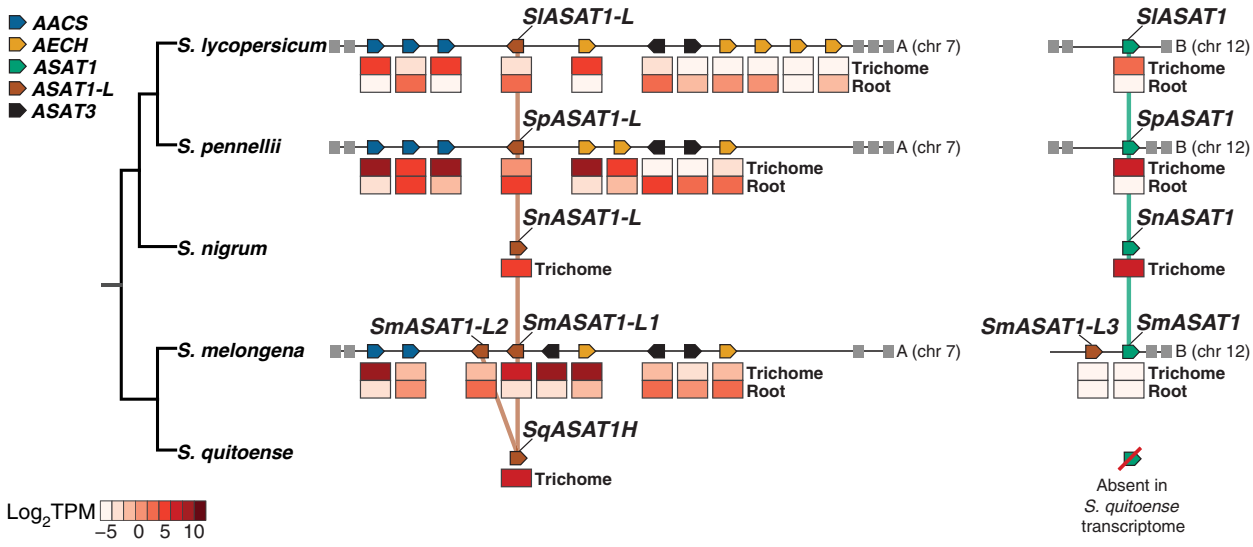

Fig. 6. Gene expression evolution within the acylsugar BGC is associated with root acylsugar pathway evolution. Detailed view of gene expression variation within the acylsugar BGC across *Solanum* species. *ASAT1* and *ASAT1-L* orthologs (table S8) in each species are labeled. Orthologs in *S. nigrum* and *S. quitoense* are shown without genomic context because genome sequences are not available. Characterized acylsugar biosynthesis genes in cultivated tomato are labeled. Colored arrows represent orthogroups within the acylsugar BGC, consistent with Fig. 5: blue, *AACS1* (OG0000391); green, *ASAT1* (OG0011176); brown, *ASAT1-L* (OG0015608); yellow, *AECH1* (OG0000616); and black, *ASAT3-L* (OG0000384). Anchor genes used to identify the syntenic region across species are represented by gray boxes. Heatmaps show absolute transcript abundance (log_2TPM) in trichomes and young whole roots. Because root RNAseq data are not available for *S. nigrum* and *S. quitoense*, only trichome expression is shown for these species.

This study was inspired by two observations: our identification of root-expressed paralogs of previously characterized trichome acylsugar biosynthetic enzyme genes (Figs. 2 and 6 and figs. S1 to S4) and a recent publication from the Aharoni group (*40*) demonstrating exudation of acylglucoses and acylsucroses from roots of 28-day-old cultivated tomato plants. Our results demonstrated that 10-day-old tomato roots accumulate a suite of 50 detectable root-specific acylsugars with a glycosylinositol core (Fig. 3 and table S5). Discrepancies between our observations and those of Korenblum and colleagues (*40*) may reflect differences in plant developmental stages, growth conditions, extraction methods, or our use of NMR and acylsugar-specific LC-MS analysis techniques.

Gene coexpression network analysis in cultivated tomato identifies root-expressed paralogs of trichome acylsugar pathway components

Analysis of 274 public domain cultivated tomato gene expression datasets (table S1) revealed the existence of root hair–expressed paralogs of previously characterized trichome acylsugar pathway genes (Fig. 2, figs. S1 to S4, and tables S2 and S8). For example, *SlASAT1*, which encodes the first core trichome pathway enzyme (*35*), has one close paralog, *SlASAT1-L*, which is highly and specifically expressed in roots (Figs. 2 and 6 and tables S2 and S8). In addition, the third and fourth trichome pathway enzymes, SlASAT3 and SlASAT4 (*36*, *37*), each have multiple root-expressed paralogs (Fig. 2, fig. S1, tables S2 and S8). A similar trend was observed for paralogs of upstream acylsugar pathway genes involved in acyl-CoA substrate production, including *SlIPMS3* (fig. S4 and tables S2 and S8) (*38*), *SlAACS1* (fig. S2 and tables S2 and S8), and *SlAECH1* (fig. S3 and tables S2 and S8) (*39*). This pattern was also seen for acylsugar modifying hydrolase (ASH) enzymes (*56*). Trichome-expressed *SlASH3* (*Solyc09g075710*) and *SlASH4* (*Solyc04g005230*) both fall into the trichome module, while their close paralog, *Solyc09g075680*, is root hair–enriched and clusters with the root hair module (tables S2 and S8). While this is consistent with the observation that gene duplicates often show divergent expression patterns (*57*–*59*), the notable extent to which the trichome pathway enzymes are "replicated" in roots led us to investigate root acylsugars in detail.

Evidence for distinct acylsugar biosynthetic pathways in tomato roots and trichomes

Using LC-MS and NMR techniques, we obtained evidence for the presence of acylsugars with glycosylinositol cores in roots of 10-day-old cultivated tomato seedlings (Fig. 3 and table S5). Our NMR-based approaches annotated one tomato root acylsugar as containing a glucosylinositol disaccharide core substituted with five acyl chains including C2, iC5, iC6, and iC7 (Fig. 3B and table S5). LC-MS studies revealed several dozen similar acyldisaccharides composed of hexose-hexose sugar cores that we infer are glycosylinositol based on parsimony (Fig. 3A and table S5). Most acylsugars detected in tomato roots are acylated on both hexose rings, and the acyl chain combinations are similar, suggesting that smaller acylsugars represent biosynthetic pathway intermediates or degradation products. Genetic evidence that accumulation of all root-specific acylsugars in cultivated tomato depends on *SlASAT1-L* (Fig. 4 and table S6) suggests the existence of a common acylglycosylinositol biosynthesis pathway anchored by root-expressed *SlASAT1-L* in cultivated tomato (Figs. 2 and 6). Further, functional differences between SlASAT1 and SlASAT1-L (Fig. 4, fig. S8, and table S6) support the hypothesis that acylsucroses and acylglycosylinositols are synthesized through separate pathways.

Using the same LC-MS approach, we identified acyldisaccharides decorated with similar short C2, C5, C7, C8, and C9 acyl chains in young 10-day-old roots of the wild tomato, *S. pennellii* LA0716 (fig. S9 and table S9). In addition, wild tomato root acylsugars lack

the medium-length C11-C12 acyl chains present in trichomes (*39*, *60*, *61*). Accumulation of different acylsugars in trichomes and roots, coupled with the root-specific expression of *ASAT1-L* in both wild and cultivated tomato (Fig. 6), reinforces the idea that independent biosynthetic pathways operate above and below ground.

CRISPR-Cas9 mutagenesis of the root-expressed *SlASAT1-L* and trichome pathway *SlASAT1* (Fig. 4) also helped us address observations of root acylsucroses from the literature (*40*). We found that root extracts contained metabolites that resemble trichome acylsucroses based on molecular masses and fragmentation patterns. Coelution with trichome metabolite extracts provided strong evidence that these root "trichome-like" acylsucroses are identical to trichome acylsugars. However, their abundances in root extracts were substantially lower than seen in trichome extracts, and these trace quantities were not found in roots of *slasat1-l* null mutants (Fig. 4 and table S6). This result, combined with the fact that SlASAT1-SlASAT4 transcripts are not highly expressed in roots (Fig. 2 and table S2), led us to hypothesize that our nutrient agar plate–grown plant root samples were contaminated with trichome acylsucroses before or during sample collection. A less likely possibility is that trichome-produced acylsucroses are transported to roots.

Hypotheses for cultivated tomato root acylsugar biosynthesis

On the basis of our acylsugar annotations (Fig. 3 and table S5) and in vitro SlASAT1-L substrate preferences (Fig. 4E and fig. S8D), we propose that acylglycosylinositol biosynthesis in cultivated tomato roots begins with acylation of *myo*-inositol with an iC7 chain by SlASAT1-L. As all detectable cultivated tomato root acylsugars are composed of a disaccharide core decorated with two to five acyl chains, we hypothesize that the resulting monoacylinositols undergo subsequent acylations and a glycosylation, although the precise order is currently unknown. At some point, the multiply-acylated inositol is likely to become a poor substrate for glycosylation, although the acylsugar accumulation pattern observed in *S. quitoense* trichomes suggests that triacylinositols may be glycosylated in vivo. Detectable acylsugars in *S. quitoense* trichomes include acylinositols with acyl chains at positions 1, 2, and 3 and cognate acylinositol-disaccharides with identical acyl chains at positions 1, 2, and 3 plus a sugar moiety at position 4 (*62*). Tetraacylinositols acylated at positions 1, 2, 3, and 4 are also detected in *S. quitoense* trichomes, suggesting that the acylsugar biosynthetic pathway in this species branches at triacylinositols (*62*).

Coexpression analysis identified several candidate root acylsugar biosynthesis genes in the red root hair module (Fig. 1, Table 1, and tables S2 and S3). If acylinositol biosynthesis is analogous to characterized acylsucrose pathways, where each ASAT catalyzes a single acylation (*35*–*37*), we expect a total of five acyltransferases, including SlASAT1-L, to comprise the core acylglycosylinositol pathway in cultivated tomato roots. High-priority ASAT candidates include root-expressed paralogs of *SlASAT3* and *SlASAT4* (Figs. 2 and 6, fig. S1, and table S2). Plant-specialized metabolites are often glycosylated through the action of uridine diphosphate–dependent glycosyltransferase (UGT) enzymes (*63*). Top acylinositol glycosylation candidates include 17 root-expressed UGTs in the root hair module (table S2). Testing these candidates using a combination of in vitro biochemistry and CRISPR-Cas9 mutagenesis could provide a blueprint for acylinositol pathway elucidation efforts in other *Solanum* species.

Potential biological functions of root acylsugars

The biological role served by root acylsugars remains an open question. Given that trichome acylsugars defend aerial tissues against pests, it is tempting to posit that root acylsugars protect subterranean tissues. Acylsugars are exuded by cultivated tomato roots challenged with soil microbes (*40*), demonstrating that their transport, and possibly their biosynthesis, is triggered by biotic stimuli. However, it is unknown how acylsugars affect rhizosphere microbiota. It is possible that root acylsugars are exuded to select for association with some microbes while deterring others. Potato cyst nematodes (*Globodera* spp.) are specialized root pests able to colonize Solanaceae species, including tomato. Host plant root exudates are sufficient to trigger nematode eggs to hatch, initiating infestation (*64*). Future experiments examining how root acylsugars structure soil microbiota and influence beneficial and harmful animal pests will shed light on the biological functions of these metabolites. The work performed in this study, including the availability of the *slasat1-l* knockout line, which does not accumulate detectable root acylsugars, should facilitate these downstream studies.

Evolution of inositol-based acylsugars: A tale of nested BGCs

Our observation that the cultivated tomato acylglycosylinositol pathway gene, *SlASAT1-L*, is a root-expressed component of the previously described trichome acylsugar BGC containing three trichome-expressed genes, *SlASAT1*, *SlAACS1*, and *SlAECH1* (*39*), prompted us to reexamine the evolutionary history of this cluster (Figs. 5 and 6). We found evidence suggesting the BGC's inception dates to an "early" period in Solanaceae evolution and changed over time, leading to nested trichome and root-expressed BGCs in tomato.

As illustrated in Fig. 5, acylsugar BGC formation likely began approximately 30 to 50 million years ago (Ma) in the Solanaceae ancestor (*53*, *65*), driven initially by the integration of an *AECH* homolog into a syntenic chromosomal region conserved among Eudicots. The subsequent Solanaceae-specific polyploidization event (*53*, *65*) led to its representation on both chromosomes (Fig. 5). Intriguingly, the genesis of *AECH* homologs within the BGC can be attributed to duplication of an ancient ECH homolog, *Solyc12g011160* (table S8), localized to an adjacent region, roughly 3 Mb away from the chromosome 12 acylsugar BGC in cultivated tomato (table S2). Thereafter, we infer that *ASAT1* and *AACS* homologs were recruited to this region in the lineages leading to *Petunia axillaris* and *Nicotiana attenuata* (Fig. 5), establishing the trichome acylsugar BGC approximately 24 to 28 Ma (*65*). This early acylsugar BGC formation is consistent with broad distribution of trichome sucrose-based acylsugars across the Solanaceae, including *Salpiglossis sinuata, Petunia* spp., *Nicotiana* spp., *Hyoscyamus niger*, *Datura wrightii*, *Iochroma cyaneum*, *Physalis origanifolia*, *Jaltomata sinuosa*, and multiple *Solanum* spp., including wild and cultivated tomatoes (*31*, *39*, *55*, *66*–*75*).

Our synteny analysis also revealed evidence of a notable expansion of the established trichome acylsugar BGC approximately 19 to 21 Ma, marked by duplication and nested integration of a gene array that included *ASAT1-L*, *AECH*, and *AACS* homologs (Fig. 5) in a common ancestor of extant *Datura* and *Solanum* species (*65*, *76*, *77*). While the nested *AACS* duplicate appears to have been subsequently lost in the *Solanum* genus (Fig. 5), the nested *AECH* and *ASAT1-L* were both recruited into acylsugar metabolism.

Previous genetic and biochemical evidence indicates that cultivated tomato *SlAECH1* (Fig. 5) and its orthologs in wild tomato, *SpAECH1* and *S. quitoense*, *SqAECH1*, produce medium-length

(C10-C12) acyl-CoA substrates for acylsugar biosynthesis in trichomes (*39*). Further, the distribution of medium C10-C12 chain containing acylsucroses across the Solanaceae correlates with the presence of *AECH* duplicates nested between *ASAT1-L* and *ASAT3* homologs in the acylsugar BGC. For example, while only short chain (<C10)–containing acylsugars have been documented in *Petunia* spp. and *Nicotiana* spp., acylsugars decorated with both short and medium acyl chains have been identified in *I. cyaneum* (*55*) and many *Solanum* species (*26*, *38*, *39*, *55*). These observations suggest that duplication of *AECH* within the gene cluster enabled incorporation of medium-length acyl chains into acylsugar biosynthesis.

In addition to cultivated tomato roots, inositol-based acylsugars accumulate in trichomes of several dozen species in different clades of the large and diverse *Solanum* genus, including *S. melongena* (brinjal eggplant), *S. quitoense*, and *S. nigrum* (*51*, *54*, *62*, *78*, *79*), and their spatial distribution across species is correlated with *ASAT1-L* expression (Fig. 6). For example, cultivated tomato *SlASAT1-L* is expressed in roots that accumulate acylglycosylinositols (Figs. 3 and 6 and tables S2 and S5). In contrast, *S. melongena SmASAT1-L1* and *S. nigrum SnASAT1-L* are trichome-expressed (Fig. 6 and tables S11 and S12), and these species accumulate detectable inositol-based acylsugars in trichomes, but not roots (fig. S9). Previous work indicates that *S. quitoense* trichomes also accumulate acylinositols and express *SqASAT1H* (*51*, *55*, *62*), an *SlASAT1-L* ortholog (Fig. 6 and tables S8 and S13). Further, inositol-based acyldisaccharides in *S. quitoense* and *S. melongena* trichomes and GI5:26 (2,5,6,6,7) in cultivated tomato roots (Fig. 3B) are all acylated on the inositol ring at positions 1, 2, and 3 (*54*, *62*, *79*). These results support the hypothesis that root and trichome inositol-based acylsugars are synthesized through similar biosynthetic routes anchored by *ASAT1-L*. If true, we would expect the acyldisaccharides detected in *S. pennellii* roots (fig. S9) to be inositol-based and synthesized through a pathway anchored by root-expressed *SpASAT1-L* (Fig. 6 and table S9). Future experiments examining acylsugar biosynthesis in *S. pennellii* roots will test these hypotheses and deepen our understanding of acylsugar pathway evolution.

We infer that *ASAT1-L* arose through duplication of its sucrose-acylating *ASAT1* ancestor and then diverged functionally to acquire *myo*-inositol acylating activity. On the basis of biochemical evidence demonstrating that both SlASAT1-L and its *S. quitoense* ortholog, SqASAT1H (*51*), preferentially acylate *myo*-inositol over sucrose in vitro (Fig. 4 and fig. S8), we hypothesize that these functional changes must have occurred by 14 Ma, when the *Solanum* genus began radiating (*65*, *76*). We hypothesize that other enzymes involved in inositol-based acylsugar biosynthesis arose through duplication of acylsucrose pathway components and acquired inositol or acylinositol-disaccharide acylating activity during this time, as well. This neofunctionalization was likely enabled by the well-known process of relaxation of selective pressure following gene duplication (*18*, *80*, *81*).

Existing data suggest that inositol-based acylsugar biosynthesis first arose in trichomes and was redirected to roots in a subset of *Solanum* species, including cultivated and wild tomato, due to *ASAT1-L* expression divergence (Figs. 3 and 6 and fig. S9). On the basis of the observed *ASAT1-L* spatial expression distribution (Figs. 2 and 6 and tables S10 to S12), we infer that root-specific *ASAT1-L* expression arose 2 to 14 Ma, after the *Solanum* genus began radiating and before cultivated and wild tomato split (*65*, *76*, *77*). Gene expression divergence is common among closely related species, due to changes in the *cis* (e.g., *cis*-regulatory elements) and *trans* (e.g., transcription factors) elements governing transcription (*82*–*84*). One intriguing possible mechanism is that spatial expression divergence of one or a few transcription factors regulating the expression of inositol-based acylsugar biosynthesis enzymes led to redeployment of this pathway in the roots.

Inositol-based acylsugar biosynthesis, with its recent origin and restricted distribution within the Solanaceae, presents a rich system for detailed explorations into how metabolic pathways form and diverge. Investigating the causal mutations and mechanisms underlying the observed changes in substrate preferences and spatial expression patterns of *ASAT1-L* and other pathway components could transform our understanding of biochemical evolution. We also anticipate that further study of inositol-based acylsugar biosynthesis will impart wisdom concerning the fates of gene duplicates (*85*), the structure and function of BGCs (*22*, *86*), and enzyme neofunctionalization.

MATERIALS AND METHODS

Plant lines and growth conditions

Cultivated tomato *S. lycopersicum* cv. M82 seeds were obtained from the C.M. Rick Tomato Genetic Resource Center (https://tgrc.ucdavis.edu/; RRID:SCR_014954). Unless otherwise noted, plants were grown under sterile conditions. Tomato seeds were surface sterilized in 40% bleach for 20 min with gentle rocking, rinsed five times with sterile water, and then sown in a straight line on Falcon 150 mm–by–25 mm cell culture dishes (Corning Inc., Corning, NY, USA) containing half-strength Murashige and Skoog basal salts (Caisson Labs, Smithfield, UT, USA), supplemented with Gamborg's B-5 vitamins (Caisson Labs, Smithfield, UT, USA), 3% sucrose (Thermo Fisher Scientific Inc., Waltham, MA, USA), and 0.7% Phytoblend agar (Caisson Labs, Smithfield, UT, USA) adjusted to pH 5.8. The plates were wrapped in Micropore surgical tape (3M Company Inc., Maplewood, MN, USA), then oriented vertically in steel alloy 4.25″ by 4.5″ by 13.5″ plate racks (Spectrum Diversified Designs LLC, Solon, OH, USA), and stratified in the dark at 25°C for 5 to 7 days to germinate seeds. Thereafter, vertically oriented plates were transferred to a growth chamber and grown under a 12-hour diurnal cycle, 25°C/18°C, 175 μE light intensity.

Reference genome selection and transcriptome alignment

We used the *S. lycopersicum* M82 genome (*87*) as our cultivated tomato reference. Gene annotations for M82 were derived from ITAG4.0 (*88*) and the tomato pan-genome (*89*), which are missing *SlASAT1* (*Solyc12g006300*). To correct this, we performed a BLASTn search to determine the coordinates for *SlASAT1* in the M82 genome, and then updated the M82 annotation file accordingly. We generated mRNA, coding sequence (CDS), and protein fasta files from the M82 genome using R (*90*), University of California Santa Cruz Genome Browser utilities (kentUtils; https://github.com/ucscGenomeBrowser/kent), and BEDtools (*91*). The mRNA fasta file was the simplest to construct. First, we converted our modified M82 annotation file to a BED12 file using gff2ToGenePred, genePredToBed, and bedSort from kentUtils. The mRNA transcript sequences specified in resulting BED12 file were extracted from the M82 genome fasta file with the BEDtools getfasta command with -split -s -name parameters specified. To construct the CDS and protein fasta files, we first used R to select only CDS-encoding exon features from our modified M82 annotation, and then used kentUtils and BEDtools as before to generate the CDS fasta file. To construct the protein fasta file, we

translated the CDS sequences using the faTrans function from kentUtils, with the -stop parameter specified.

We performed transcriptome analysis in cultivated tomato using publicly available data accessed from two sources: the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) repository (*92*) and the Genome Sequence Archive (GSA) maintained by the Beijing Institute of Genomics (BIG) Data Center (*93*). We selected 274 transcriptomes, representing 84 tissue-developmental stage groups, including four root tissues: young root hairs, young root tips, young whole roots, and mature whole roots (Table RNAseq). All transcriptomes selected in this study were collected from the *S. lycopersicum* M82 cultivar except root hair samples, which were from the Heinz 1706 cultivar. To download transcriptomes housed in the INSDC repository, we used the fasterq-dump command available from NCBI's SRA Toolkit (https://github.com/ncbi/sra-tools). We downloaded raw fastq sequence files from the SRA and GSA repositories using the fasterq-dump command from the NCBI SRA Toolkit and the Wget utility, respectively.

To prepare sequences for mapping, we trimmed adapters and low-quality bases and then filtered paired- and single-end reads shorter than 65 and 36 bp, respectively, using Trimmomatic (*94*). These filtering steps removed 0.04 to 24.4% of sequences, yielding a high-quality set of 49- to 147-bp reads for alignment (Table RNAseq). We aligned the processed reads to the M82 reference genome using STAR in two-pass mode to optimize splice junction discovery and mapping sensitivity (*95*, *96*). On average, 85.5% (range: 46.5 to 96.2%) of the reads in each sample mapped to a single, unique genomic location (Table RNAseq). We filtered the resulting transcriptome alignments, according to Genome Analysis Toolkit (GATK) best practices (*97*, *98*). Briefly, we removed optical and polymerase chain reaction (PCR) duplicates with the MarkDuplicates command from Picard tools (http://broadinstitute.github.io/picard), parsed reads into exon segments, and removed intron-spanning bases using the SplitNCigarReads command from GATK (*98*) and then nonunique alignments (i.e., reads with a mapping quality score below Q60) using the SAMtools view command (*99*). We generated raw read counts from the filtered alignments using the htseq-count command from HTSeq (*100*).

Differential gene expression and weighted gene coexpression analyses

Raw read counts generated by HTSeq-count were used to perform differential gene expression analysis in edgeR (*101*). To restrict comparisons to expressed genes, we required at least one read count per million in two or more samples for further analysis. This filtering step removed 8256 (23.8%) of the 34,620 genes in our modified M82 annotation, yielding a set of 26,364 expressed transcripts for differential gene expression analysis. Next, we normalized read counts across samples using the default trimmed mean of *M* values (TMM) method in the calcNormFactors function and then tested for gene expression differences across our 84 tissue-developmental stage groups using generalized linear modeling (GLM) with a quasi-likelihood (QL) approach, which permits any combination of sample comparisons to be made. To set up our comparisons, we generated an experimental design matrix specifying the 84 groups with the model.matrix function and then used the glmQLFit function to fit our model to a QL-GLM framework. To identify genes with a $\log_2$ fold change (FC) > 2 between comparisons, we used the glmTreat function in edgeR. The glmTreat function performs threshold hypothesis testing, which is a rigorous statistical approach that evaluates variance and magnitude to detect expression differences greater than the specified value (e.g., $\log_2$ FC > 2), and then applies false discovery rate (FDR) *P* value corrections on the resulting gene list. We categorized genes as significantly differentially expressed between two groups if $\log_2$ FC > 2 and FDR-corrected *P* value ≤ 0.05.

To identify clusters of genes that exhibit similar expression patterns across the tissue-age groups in our transcriptome dataset, we generated a coexpression network using the WGCNA package in R (*102*, *103*). To obtain the absolute expression values needed as input for WGCNA, we used the calculateTPM function from the R package, scater, to calculate transcripts per million (TPM) for all expressed genes in our dataset (*104*). We selected TPM as a measure of absolute transcript abundance rather than fragments per thousand base pairs per million reads mapped (FPKM) because TPM is more comparable across samples and experiments when used in combination with transcript normalization techniques, such as the TMM method used by the calcNormFactors function in edgeR (*105*). We took the one-step module construction approach using the blockwiseModules function in WGCNA, specifying the following parameters: power = 15, TOMType = "signed," networkType = "signed," detectCutHeight = 0.995, maxBlockSize = 30,000, minModuleSize = 50, reassignThreshold = 1e-6, mergeCutHeight = 0.15, pamStage = TRUE, pamRespectsDendro = FALSE, and deepSplit = 2. We specified signed as the network type, which only clusters positively coexpressed genes. In contrast, in an unsigned network, both positively and negatively coexpressed genes may be grouped together. The resulting coexpression network was composed of 27 modules of genes with positively correlated expression profiles across our 274 transcriptomes.

Phylogenetic analysis

To identify candidates for root acylsugar biosynthesis, we searched for homologs of characterized trichome acylsugar genes, *ACYLSUGAR ACYLTRANSFERASE 1* (*SlASAT1*; *Solyc12g006330*), *SlASAT2* (*Solyc04g012020*), *SlASAT3* (*Solyc11g067270*), *SlASAT4* (*Solyc01g105580*), *ISOPROPYLMALATE SYNTHASE 3* (*SlIPMS3*; *Solyc08g014230*), *ACYLSUGAR ACYL-COA SYNTHETASE 1* (*SlAACS1*; *Solyc07g043630*), and *ACYLSUGAR ENOYL-COA HYDRATASE 1* (*SlAECH1*; *Solyc07g043680*) (*35*–*39*), in the cultivated tomato M82 reference genome based on the presence of characteristic Protein family (Pfam) domains (*106*). Pfam domains represent evolutionarily conserved functional units and are identified through sequence similarity. We queried our reference trichome acylsugar sequences against the Pfam domain library using PfamScan (https://ebi.ac.uk/Tools/pfa/pfamscan/), which returned four Pfam domains, Transferase (PF02458), HMGL-Like (PF00682), AMP-binding (PF00501), and ECH_2 (PF16113). To identify related sequences in cultivated tomato, we screened the representative hidden Markov model (HMM) profile for each Pfam domain against our modified M82 proteome (see the "Reference genome selection and transcriptome alignment" section for details) using the hmmsearch tool from HMMER (hmmer.org).

We reconstructed phylogenetic relationships among the resulting PF02458 ($N = 98$), PF00682 ($N = 7$), PF00501 ($N = 60$), and PF16113 ($N = 32$) M82 proteins using MAFFT v7.471 (*107*) and IQ-TREE v2.1.3 (*108*). First, we used MAFFT in E-INS-i mode to construct multiple sequence alignments (MSAs) for each group of proteins through iterative refinement. The E-INS-i algorithm performs local alignment using a generalized affine gap cost suitable for distantly related sequences characterized by a conserved domain embedded within relatively unconserved regions, which are difficult to align (https://mafft.cbrc.

jp/alignment/software/algorithms/algorithms.html). Next, we implemented the ModelFinder tool in IQ-TREE v2.1.3 to select the optimal sequence evolution model for estimating phylogenetic relationships based on the MSAs (*109*). Last, we used IQ-TREE v2.1.3 to infer maximum likelihood trees for each Pfam based on the selected model, with branch support obtained from 100,000 ultrafast bootstrap approximations (*110*). To visualize phylogenies with absolute transcript abundance (log_2 TPM) heatmaps, we used the ggtree R package (*111*).

Metabolite extraction

Roots and shoots of 10-day-old seedlings grown under sterile conditions were separated at the base of the hypocotyl using a straight edge razor blade and collected for metabolite extractions. Roots were collected into preweighed 2-ml screw-cap tubes (Dot Scientific Inc., Burton, MI, USA) containing three 3.9-mm stainless steel grinding balls (Spex SamplePrep, Metuchen, NJ, USA), weighed quickly, and then flash frozen in liquid nitrogen. Frozen root tissue was homogenized mechanically using a 1600 MiniG (Spex SamplePrep, Metuchen, NJ, USA). Briefly, tubes were placed in aluminum cryo blocks (Spex SamplePrep, Metuchen, NJ, USA) submerged in liquid nitrogen, transferred to the MiniG, ground for 30 s at 1600 rpm, and then quickly transferred back to liquid nitrogen. Samples were inspected individually, and this grinding cycle was repeated at least three times until root tissue was completely powdered. Thereafter, 500 µl of extraction solvent [90% methanol (v/v) containing 0.1% formic acid and 1 µM telmisartan (internal standard)] was added and samples were incubated overnight at −20°C. The next day, root samples were centrifuged at 17,000 rpm for 20 min at 4°C, and the supernatant was transferred to Whatman Mini-UniPrep LC vials (0.5 ml) equipped with 0.22-µm polytetrafluoroethylene (PTFE) syringeless filters (Cytiva Life Sciences, Marlborough, MA, USA). In parallel, shoots (cotyledons and hypocotyls) were collected for aerial surface (i.e., trichome) metabolite extraction into preweighed 2-ml screw-cap tubes (Dot Scientific Inc., Burton, MI, USA), weighed, and immediately flash frozen. Thereafter, 1000 µl of extraction solvent was added to each sample and vortexed briefly, and 500 µl of supernatant was transferred to Mini-UniPrep LC vials.

LC-MS method

Metabolite profiling was performed using LC-MS on an Acquity ultra-performance LC (UPLC) I-class LC system coupled to a Xevo G2-XS quadrupole time-of-flight (QToF) mass spectrometer (Waters Corporation, Milford, MA) equipped with an electrospray ionization (ESI) probe operating in positive (ESI^+) and negative (ESI^-) modes. The ionization source temperature was set to 100°C, desolvation temperature was 350°C, capillary voltage was 2 kV, sampling cone voltage was 40 V, cone gas flow was 20 liters/hour, and desolvation gas flow was 600 liters/hour. We injected 10 µl of each metabolite extract onto a 100 mm–by–2.1 mm inside diameter (i.d.), 1.7-µm BEH C18 UPLC column (Waters Corporation, Milford, MA) maintained at 40°C. Metabolites were separated using a 30-min binary solvent gradient, composed of aqueous 10 mM ammonium formate with 0.1% formic acid (solvent A) and 100% acetonitrile (solvent B), flowing at 0.3 ml/min. The gradient program was as follows: 5 to 25% B at 0 to 1 min, 25 to 100% B at 1 to 26 min, 100% B at 26 to 28 min (wash), and 5% B at 28.01 to 30 min (initial conditions). To generate fragment ions useful for acylsugar annotation, we performed CID using three quasi-simultaneous (0.2-s acquisition times) collision potentials in negative (0, 10, and 30 V) and positive (0, 10, and 25 V) ionization modes. Spectra were acquired over mass/charge ratio (*m*/*z*) 50 to 1500 and stored in separate acquisition functions. To increase mass accuracy, lock mass correction using leucine enkephalin as the reference was applied during data collection.

Acylsugar annotation and data-dependent tandem MS

We annotated acylsugars based on retention times, molecular and pseudomolecular ion masses, and fragmentation patterns obtained from LC-MS-CID (*49*, *78*). The basic acylsugar structure consists of a central sugar core substituted with acyl chain esters, which will fragment into its component parts under appropriate CID conditions. To identify putative acylsugars, we first searched the high-energy data function for coeluting ions corresponding to (i) free acyl chain fragments and (ii) sugar core fragments after losses of all acyl chains. Acyl chain fragments, as carboxylate anions (e.g., *m/z* 101.06 for C5), are readily apparent in negative mode CID spectra at our highest collision potential, 30 V. Sugar core identity has a major impact on fragmentation characteristics in positive and negative mode CID. For example, ions corresponding to deprotonated sugar cores remaining after neutral losses of all acyl chains are apparent in negative mode CID spectra at our lowest (0 V) or highest (30 V) collision potentials for acylglucoses (glucose core *m/z* 161.05, 143.03) or acyldisaccharides (disaccharide core *m/z* 341.11, 323.09), respectively. In contrast, acylinositols containing a *myo*-inositol sugar core are resistant to fragmentation in negative-ion mode, while their dehydrated sugar core fragment ions (*m/z* 127.04) are apparent in positive mode CID at our highest collision potential level, 25 V. Therefore, we search for acyl chains in negative mode and search for sugar cores in both negative and positive modes.

Next, we search for intact pseudomolecular ions of acylsugars calculated on the basis of observed sugar core and acyl chain fragment ions. In MS spectra from positive and negative mode CID at 0 V, adducts of ammonium, $[M + NH_4]^+$, and formate, $[M + FA\text{-}H]^-$, are the most abundant acylsugar ions. The presence of calculated pseudomolecular ion masses coeluting with sugar core and acyl chain fragment ions is considered strong evidence for an acylsugar. Fragment ions resulting from successive neutral mass losses of the ester groups as ketenes (e.g., 84 Da for C5 ketenes) are apparent in negative mode CID at 10 to 30 V. Acylsugars are often decorated with acetate (C2) esters whose fragment ions are below our *m/z* 50 mass spectra acquisition limit. Instead, acetyl groups are inferred from neutral loss of a C2 ketene (42 Da). These ions can be used to confirm acyl chain and sugar core annotations for individual acylsugar compounds.

CID in positive mode provides information about the number and sizes of acyl chains attached to each hexose ring. The dominant fragment ions observed in positive mode CID at 10 V arise from glycosidic bond cleavage, leaving the esterified acyl chains intact. For acylsugars with a disaccharide (hexose-hexose) core, including those observed in cultivated tomato trichomes and roots, positive mode CID yields two major fragment ions, one for each acylated hexose ring. By comparing fragment and precursor ions, the acyl chain identity on each ring can be inferred.

To support our annotations for highly abundant acylsugar compounds detected in wild-type cultivated tomato root extracts, we performed data-dependent LC-MS/MS using the same instruments, ion source parameters, column, and gradient method described above for LC-MS metabolite profiling analyses. For all acylsugar peaks

analyzed, formate $[M + FA\text{-}H]^-$ adducts generated using negative mode CID at 0 V were selected as precursor ions and fragment ions were generated using a collision potential of 25 V. For each acylsugar precursor ion, full-fragment ion spectra (*m/z* 50 to 1500) were acquired over the relevant retention time window (using 0.2 s per scan acquisition times) and stored in a separate acquisition function.

The following nomenclature is used to describe acylsugar structures: one to two letters defining the sugar core (e.g., GI for glycosylinositol) followed by the total number of acylations; then, a colon followed by the combined number of carbon atoms across all acyl chains; last, the numbers of carbon atoms in individual acyl chains are listed in parentheses. For example, an acylglycosylinositol with one C2, one C5, two C6, and one C7 acyl chains is written as GI5:26 (2,5,6,6,7).

Acylsugar quantification and statistical analysis

To quantify relative acylsugar abundance across root and trichome extracts from wild-type M82, *slasat1-1*, *slasat1-l-1*, and *slasat1-l-2*, we performed LC-MS with an LC-20ADvp ternary pump (Shimadzu Corporation, Kyoto, Japan) coupled to a Xevo G2-XS mass spectrometer (Waters Corporation, Milford, MA). Metabolites (10-µl injections) were separated on an Ascentis Express 100 mm–by–2.1 mm i.d., 2.7-µm HPLC C18 column (Supelco Inc., Bellefonte, PA, USA), maintained at 40°C, using the same 30-min binary solvent gradient used for root metabolite characterization, flowing at 0.4 ml/min. Ion source parameters were as described above for LC-MS metabolite profiling analyses. Acylsugar compounds were quantified using extracted ion chromatograms of formate $[M + FA\text{-}H]^-$ adducts in negative mode CID at 0 V.

The TargetLynx tool in MassLynx V4.2 (Waters Corporation, Milford, MA) was used to integrate extracted ion chromatograms across specified retention time windows and then normalize acylsugar peak areas to internal standard peak area (telmisartan $[M\text{-}H]^-$; *m/z* 513.23). We adjusted the resulting acylsugar response values for each sample to solvent volume (in microliters) and fresh weight (in milligrams). We constructed analysis of variance models to test for differences in acylsugar abundance across the four genotypes in each tissue using the aov function in the R stats package (*90*). Then, we performed Tukey's pairwise comparisons of acylsugar abundance between genotypes using the glht function from the multcomp package in R (*112*).

NMR analysis

To obtain purified GI5:26 for NMR analysis, approximately 100 root metabolite extracts were pooled and subjected to semi-preparative LC (semi-prep LC). Solvent was removed in vacuo by centrifugal evaporation, and dried residue was dissolved in 2 ml of acetonitrile:water:formic acid (80:20:0.001). Semi-prep LC was performed using a Waters 2795 HPLC (Waters Corporation, Milford, MA, USA) equipped with an Acclaim 120 4.6 mm by 150 mm, 5-µm HPLC C18 column (Thermo Fisher Scientific, Waltham, MA, USA), maintained at 40°C. Repeated injections of 100-µl extract aliquots were separated using a 44-min binary solvent gradient, composed of water with 0.1% formic acid (solvent A) and 100% acetonitrile (solvent B), at a flow rate of 1.5 ml/min. To purify GI5:26, we used the following LC gradient: 5 to 60% B at 0 to 2 min, 60 to 80% B at 2 to 40 min, 80 to 100% B at 40 to 42 min, and 5% B at 42.01 to 44 min. Fractions were collected automatically at 0.25-min intervals using a 2211 Superrac fraction collector (LKB Bromma, Stockholm, Sweden), and identical fractions were pooled between runs.

Pooled semi-prep fractions were analyzed for GI5:26 presence and purity using two methods: LC-MS and MS with direct infusion. LC-MS analysis of semi-prep fractions was performed using an LC-20ADvp ternary pump (Shimadzu, Kyoto, Japan) coupled to a Waters Xevo G2-XS QToF mass spectrometer (Waters Corporation) equipped with an ESI probe operating in negative mode, with the following parameters: 3-kV capillary voltage, 40-V sampling cone voltage, 100°C source temperature, 350°C desolvation temperature, 20-liter/hour cone gas flow, and 600-liter/hour desolvation gas flow. We injected 10 µl of each fraction onto a Ascentis Express 100 mm–by–2.1 mm i.d., 2.7-µm HPLC C18 column (Supelco Inc., Bellefonte, PA, USA) maintained at 40°C and then separated acylsugars using a 14-min binary solvent gradient, composed of aqueous 10 mM ammonium formate with 0.1% formic acid (solvent A) and 100% acetonitrile (solvent B), flowing at 0.3 ml/min. The gradient program was as follows: 5 to 60% B at 0 to 2 min, 60 to 100% B at 2 to 10 min, 100% B at 10 to 12 min, and 5% B at 12.01 to 14 min. We performed negative mode CID with three collision potentials (0, 25, and 60 V) and acquired spectra over *m/z* 50 to 1200 into separate functions with a scan time of 0.1 s. For direct infusion analysis, we injected 10 µl of each fraction directly into the mass spectrometer and then performed negative mode ionization without CID (0 V) and acquired spectra over *m/z* 50 to 1500 with a scan time of 0.1 s. To increase mass accuracy, lock mass correction was applied during data collection using leucine enkephalin as the reference.

To prepare purified GI5:26 for NMR, solvent was removed in vacuo by rotary evaporation, and dried residue was dissolved in either 300 or 600 µl of chloroform-d_1 (99.96 atom % D; MilliporeSigma, Burlington, MA, USA) and placed in a solvent-matched 5-mm Shigemi tube (Shigemi Co. Ltd., Tokyo, Japan) or a 5-mm Kontes tube (MilliporeSigma, Burlington, MA, USA), respectively. The compound was analyzed with a Bruker Avance NEO 600-MHz spectrometer equipped with a 5-mm nitrogen cryogenic HCN Prodigy probe (Michigan State University Max T. Rogers NMR Core) and with a Bruker Avance NEO 800-MHz spectrometer equipped with a 5-mm helium cryogenic HCN probe (University of Michigan Biomolecular NMR Core). See data S1 for details. Collected spectra were referenced to chloroform-d_1 ($^{\delta}H = 7.26$ and $^{\delta}C = 77.20$ parts per million). NMR spectra were processed and analyzed with TopSpin v4.1.1 (Bruker, Billerica, MA, USA).

The GI5:26 structures were determined with a series of one-dimensional (1D) and 2D NMR experiments (data S1), an approach used successfully in previous studies (*54*, *70*, *79*). We used total correlation spectroscopy (TOCSY) to identify which proton signals correspond to each ring in the sugar core and then determined correlations between individual ^1H signals with correlation spectroscopy (^1H-^1H COSY), heteronuclear single quantum coherence (HSQC), heteronuclear multiple bond correlation (HMBC), and heteronuclear two-bond correlation (H2BC) experiments. Spin-spin splitting patterns determined ring identities to be *myo*-inositol and glucose, and HMBC revealed that the sugars are glycosidically linked at position 4 of *myo*-inositol and position 1′ of glucose. We identified the acylated ring positions based on upfield shifted proton signals and connected specific acyl chains to those positions with HMBCs between the carbonyl acyl chain carbon and sugar ring proton.

Protein expression and purification

The pET28b vector (MilliporeSigma, Burlington, MA, USA), which has a 6× His tag followed by a T7 tag, was modified to contain a tobacco etch virus (TEV) protease recognition site to enable tag cleavage. Inverse PCR was used to add the sequence CAACGACCGAAAACCTGTATTTTCAGGGC, which encodes the TEV recognition sequence (ENLYFQG) and a spacer (PTT), between the pET28b T7 tag and the Eco RI restriction site in the multiple cloning site. The PCR product was Dpn I–digested, phosphorylated by T4 polynucleotide kinase, and then recircularized using T4 DNA ligase all following standard NEB protocols. The modified and recircularized plasmid was transformed into *E. coli* and confirmed by Sanger sequencing. The M82 *SlASAT1-L* cDNA sequence was cloned into the modified pET28b using Gibson Assembly (*113*), following the NEB protocol.

SlASAT1-L in TEV-28b was expressed in BL21(DE3) *E. coli* cells (MilliporeSigma, Burlington, MA, USA). For protein expression, overnight cultures of TEV-28b SlASAT1-L were prepared in Luria Broth (LB) media containing 1% glucose (w/v). The next day, 1 liter of LB media was inoculated with 15 ml of the overnight culture and grown at 37°C until it reached an OD600 (optical density at 600 nm) of 0.5. Thereafter, the culture was equilibrated at 18°C for 30 min, induced with a final concentration of 0.5 mM isopropyl-β-D-thiogalactopyranoside, and incubated at 18°C overnight. The cultures were pelleted the next day and resuspended in 15-ml purification buffer [50 mM tris (pH 7.2) and 150 mM NaCl containing 10% glycerol] before freezing at −80°C.

For protein purification, the frozen *E. coli* pellet was thawed and lysed by sonication. The disrupted cells were centrifuged at 20,000*g* for 45 min to obtain the soluble fraction. The soluble fraction was passed through a 0.45-μm filter and applied to an AKTA start fast protein LC instrument equipped with a Cytiva 5-ml HisTrap column (Cytiva, Marlborough, MA, USA). The column was washed with 10 volumes of purification buffer containing 40 mM imidazole, followed by elution for 11 volumes with purification buffer containing 250 mM imidazole. Eluted fractions with the highest A280 were pooled. Two milligrams of His-tagged TEV protease was added to the pooled fractions, which was dialyzed overnight into the purification buffer without imidazole using SnakeSkin dialysis tubing (Thermo Fisher Scientific Inc., Waltham, MA, USA) with a 10,000 molecular weight cutoff.

The next day, cleaved and dialyzed protein was separated from TEV protease and the cleaved tag by applying the dialysate to the Histrap column. The flowthrough contained the cleaved protein, which was collected and concentrated to 0.5 ml. An equal volume of 80% glycerol was added to the concentrated protein for enzyme assays and storage at −20°C.

Enzyme assays and detection of enzymatic products

To assess enzyme activity in vitro, purified recombinant enzyme (10% v/v of the reaction volume) was combined with 100 mM sodium phosphate buffer (pH 7.4), 0.1 mM acyl-CoA, and 1 mM sugar and incubated at 30°C for 1 hour (*114*). The reaction was stopped with the addition of two volumes of 1:1:0.001 acetonitrile:isopropanol:formic acid and incubated at −20°C for 20 min to precipitate the protein. The precipitated protein was pelleted by centrifugation, and the supernatant was removed for LC-MS. For substrates, iso-branched acyl-CoAs were synthesized precisely as described in (*52*), while *myo*-inositol, nC6-CoA, and nC7-CoA were purchased from Sigma-Millipore.

Enzyme assays were evaluated by LC-MS using an LC-20ADvp ternary pump (Shimadzu, Kyoto, Japan) coupled to a Waters Xevo G2-XS QToF mass spectrometer (Waters Corporation) equipped with an ESI probe operating in negative mode, with the following parameters: 2.15-kV capillary voltage, 60-V sampling cone voltage, 100°C source temperature, 350°C desolvation temperature, 50-liter/hour cone gas flow, and 600-liter/hour desolvation gas flow. We injected 10-μl samples onto a 5 cm–by–2.1 mm reversed-phase Ascentis Express C18 HPLC column (Supelco Inc., Bellefonte, PA, USA) maintained at 40°C and then separated enzyme assay products using a 5-min binary solvent gradient composed of aqueous 10 mM ammonium formate with 0.1% formic acid (solvent A) and 100% acetonitrile (solvent B), flowing at 0.4 ml/min. The gradient program was as follows: 2% B at 0 to 0.5 min, 2 to 35% B at 0.5 to 2 min, 35 to 40% B at 2 to 2.1 min, 40 to 99% B at 2.1 to 3 min, 99% B at 3 to 4 min, and 2% B from 4.1 to 5 min. Co-retention analysis was done as above but with a 16-min gradient: 2 to 45% B from 0 to 11 min, 45 to 98% B from 11 to 12 min, holding at 98% B from 12 to 14 min, and holding at 2% B from 14 to 16 min to reequilibrate the column. We performed negative mode ionization using a CID ramp of 20 to 80 V collision potential (0.2-s acquisition times) and acquired spectra over *m/z* 50 to 1500. To increase mass accuracy, lock mass correction using leucine enkephalin as the reference was applied during data collection. Monoacylsugars were detected in negative-ion mode as formate $[M + FA\text{-}H]^-$ adducts.

CRISPR-Cas9 mutagenesis and plant transformation

We performed CRISPR-Cas9 mutagenesis of *SlASAT1* (*Solyc12g006330*) and *SlASAT1-L* (*Solyc07g043670*) in cultivated tomato following previously described methods (*39*, *50*, *115*). Putative Cas9 targets were identified for each gene through the Find CRISPR Sites tool in Geneious (Dotmatics LLC, Boston, MA, USA), and then we selected two guide RNAs (gRNAs) that target a 100- to 300-bp exonic region. We synthesized ~350-bp gene blocks (gBlocks; Integrated DNA Technologies, Coralville, Iowa) carrying 23- to 24-bp gRNAs, designed to be used in the Golden Gate cloning system. For each CRISPR construct, two gBlocks and four plasmids, pICH47742-35Spro::Cas9 (Addgene no. 49771), pICH41780 (Addgene no. 48019), pAGM4723 (Addgene no. 48015), and pICSL11024 (Addgene no. 51144), were mixed for DNA assembly using the Golden Gate assembly kit (New England Biolabs, Ipswich, MA, USA). The final CRISPR constructs were transformed into *Agrobacterium tumefaciens* strain AGL-0, which was used to transform cultivated tomato accession M82 using tissue culture, as previously reported (*39*, *115*, *116*). To identify mutated transgenic T0 plants, we performed PCR with primers flanking the gRNAs to amplify the target region and then performed Sanger sequencing on PCR products after clean-up. T0 plants with confirmed mutations were self-pollinated, and then T1 progeny were genotyped to identify homozygous mutants. Mutated T0 and T1 plants were grown in growth chambers or in a greenhouse. The growth chamber conditions are described above. The greenhouse conditions consisted of a 16-hour diurnal cycle, 25°C/18°C, with supplemental sodium iodide lighting. Plants were watered as needed using 0.5× Hoagland's solution.

Orthology and synteny analysis

To investigate ASAT1-L evolutionary history, we performed genome-wide orthology inference across 16 eudicot species, including 13 Solanaceae species and three outgroup species using OrthoFinder v2.4.1 (*117*, *118*). Using proteomes as input, OrthoFinder separates

genes into orthogroups or sets of genes descended from a single gene in the common ancestor of the species under consideration, which includes all orthologs and paralogs. We included 14 sequenced genomes—*S. lycopersicum* M82_MAS2.0 (*87*), *S. pimpinellifolium* (*119*) *S. pennellii* (*120*), *S. lycopersicoides* (*121*), *S. tuberosum* v6.1 (*122*), *S. melongena* v4.1 (*123*), *I. cyaneum* (*124*), *Datura stramonium* (*125*), *Lycium barbarum* (*126*), *N. attenuata* (*127*), *P. axillaris* (*128*), *I. trifida* (*129*), *C. canephora* (*130*), and *V. vinifera* (*131*)—and two transcriptomes—*S. nigrum* (*55*) and *S. quitoense* (*55*).

We also performed synteny reconstruction with the 14 eudicot species with sequenced genomes using the Multiple Collinearity Scan (MCScanX) toolkit (*132*). To generate required input files, we performed all-by-all BLASTp, retaining only the top five hits for each query sequence, as recommended. With this input, MCScanX identified putative homologous chromosomal regions, defined as colinear blocks of at least three genes within and between genomes. To aid synteny analysis, we selected *I. trifida*, *C. canephora*, and *V. vinifera* as outgroup species. *C. canephora* and *V. vinifera* were chosen because they experienced the paleohexaploidy event shared by all eudicots but did not undergo subsequent whole-genome duplication events. *I. trifida* is more closely related to the Solanaceae than *C. canephora* and *V. vinifera* but underwent an additional Convolvulaceae-specific WGT. Using all three species as outgroups enabled us to confidently identify synteny and reconstruct the evolutionary events leading to formation of the acylsugar BGC.

To ascertain whether apparent absences of syntenic orthologs in each species are simply due to poor genome annotations, we performed TBLASTN searches. Using protein sequences of *S. lycopersicum* acylsugar BGC genes as queries, we searched against genomic sequences of the 14 sequenced eudicot species in our analysis, including cultivated tomato. If one of the six-frame translated sequences from a target genome had significant similarity to *S. lycopersicum* query sequences and was located within the syntenic region but unannotated, we defined it as a putative syntenic ortholog. This analysis revealed two unannotated genes colocalized to the acylsugar BGC in *N. attenuata*, an *AACS* and an *AECH* homolog, denoted with dashed outlines in Fig. 5.

Supplementary Materials

This PDF file includes:
Figs. S1 to S10
Data S1
Legends for tables S1 to S13

Other Supplementary Material for this manuscript includes the following:
Tables S1 to S13

REFERENCES AND NOTES

1. E. Pichersky, D. R. Gang, Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective. *Trends Plant Sci.* **5**, 439–445 (2000).
2. B. Li, C. Förster, C. A. M. Robert, T. Züst, L. Hu, R. A. R. Machado, J.-D. Berset, V. Handrick, T. Knauer, G. Hensel, W. Chen, J. Kumlehn, P. Yang, B. Keller, J. Gershenzon, G. Jander, T. G. Köllner, M. Erb, Convergent evolution of a metabolic switch between aphid and caterpillar resistance in cereals. *Sci. Adv.* **4**, eaat6797 (2018).
3. E. Pichersky, J. Gershenzon, The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* **5**, 237–243 (2002).
4. R. N. Bennett, R. M. Wallsgrove, Secondary metabolites in plant defence mechanisms. *New Phytol.* **127**, 617–633 (1994).
5. B. G. Hansen, R. E. Kerwin, J. A. Ober, V. M. Lambrix, T. Mitchell-Olds, J. Gershenzon, B. A. Halkier, D. J. Kliebenstein, A novel 2-oxoacid-dependent dioxygenase involved in the formation of the goiterogenic 2-hydroxybut-3-enyl glucosinolate and generalist insect resistance in *Arabidopsis*. *Plant Physiol.* **148**, 2096–2108 (2008).
6. K. J. Byers, H. D. Bradshaw Jr., J. A. Riffell, Three floral volatiles contribute to differential pollinator attraction in monkeyflowers (Mimulus). *J. Exp. Biol.* **217**, 614–623 (2014).
7. E. Pichersky, J. P. Noel, N. Dudareva, Biosynthesis of plant volatiles: Nature's diversity and ingenuity. *Science* **311**, 808–811 (2006).
8. T. Tohge, L. Perez de Souza, A. R. Fernie, On the natural diversity of phenylacylated-flavonoid and their in planta function under conditions of stress. *Phytochem. Rev.* **17**, 279–290 (2018).
9. L. G. Landry, C. Chapple, R. L. Last, Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. *Plant Physiol.* **109**, 1159–1166 (1995).
10. F. Loreto, V. Velikova, Isoprene produced by leaves protects the photosynthetic apparatus against ozone damage, quenches ozone products, and reduces lipid peroxidation of cellular membranes. *Plant Physiol.* **127**, 1781–1787 (2001).
11. J. R. Smeda, H. A. Smith, M. A. Mutschler, The amount and chemistry of acylsugars affects sweetpotato whitefly (*Bemisia tabaci*) oviposition and development, and tomato yellow leaf curl virus incidence, in field grown tomato plants. *PLOS ONE* **18**, e0275112 (2023).
12. R. E. Kerwin, J. Feusier, J. Corwin, M. Rubin, C. Lin, A. Muok, B. Larson, B. Li, B. Joseph, M. Francisco, D. Copeland, C. Weinig, D. J. J. Kliebenstein, Natural genetic variation in *Arabidopsis thaliana* defense metabolism genes modulates field fitness. *eLife* **4**, e05604 (2015).
13. M. G. Bidart-Bouzat, D. J. Kliebenstein, Differential levels of insect herbivory in the field associated with genotypic variation in glucosinolates in Arabidopsis thaliana. *J. Chem. Ecol.* **34**, 1026–1037 (2008).
14. Y.-R. Lou, E. Pichersky, R. L. Last, Deep roots and many branches: Origins of plant-specialized metabolic enzymes in general metabolism. *Curr. Opin. Plant Biol.* **66**, 102192 (2022).
15. B. M. Moore, P. Wang, P. Fan, B. Leong, C. A. Schenck, J. P. Lloyd, M. Lehti-Shiu, R. Last, E. Pichersky, S.-H. Shiu, Robust predictions of specialized metabolism genes through machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 304873 (2019).
16. T. Shoji, T. Hashimoto, Recruitment of a duplicated primary metabolism gene into the nicotine biosynthesis regulon in tobacco. *Plant J.* **67**, 949–959 (2011).
17. L. Chae, T. Kim, R. Nilo-Poyanco, S. Y. Rhee, Genomic signatures of specialized metabolism in plants. *Science* **344**, 510–513 (2014).
18. G. D. Moghe, R. L. Last, Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* **169**, 1512–1523 (2015).
19. H.-W. Nützmann, A. Osbourn, Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.* **26**, 91–99 (2014).
20. S. Boycheva, L. Daviet, J.-L. Wolfender, T. B. Fitzpatrick, The rise of operon-like gene clusters in plants. *Trends Plant Sci.* **19**, 447–459 (2014).
21. N. Yu, H.-W. Nützmann, J. T. MacDonald, B. Moore, B. Field, S. Berriri, M. Trick, S. J. Rosser, S. V. Kumar, P. S. Freemont, A. Osbourn, Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res.* **44**, 2255–2265 (2016).
22. G. Polturak, Z. Liu, A. Osbourn, New and emerging concepts in the evolution and function of plant biosynthetic gene clusters. *Curr. Opin. Green Sustain. Chem.* **33**, 100568 (2022).
23. S. J. Smit, B. Lichman, Plant biosynthetic gene clusters in the context of metabolic evolution. *Nat. Prod. Rep.* **39**, 1465–1482 (2022).
24. X. Qi, S. Bakht, M. Leggett, C. Maxwell, R. Melton, A. Osbourn, A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8233–8238 (2004).
25. J. Trowsdale, The gentle art of gene arrangement: The meaning of gene clusters. *Genome Biol.* **3**, (2002).
26. A. Schilmiller, F. Shi, J. Kim, A. L. Charbonneau, D. Holmes, A. Daniel Jones, R. L. Last, Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J.* **62**, 391–403 (2010).
27. S. P. Slocombe, I. Schauvinhold, R. P. McQuinn, K. Besser, N. A. Welsby, A. Harper, N. Aziz, Y. Li, T. R. Larson, J. Giovannoni, R. A. Dixon, P. Broun, Transcriptomic and reverse genetic analyses of branched-chain fatty acid and acyl sugar production in *Solanum pennellii* and *Nicotiana benthamiana*. *Plant Physiol.* **148**, 1830–1846 (2008).
28. S. Mandal, W. Ji, T. D. McKnight, Candidate gene networks for acylsugar metabolism and plant defense in wild tomato *Solanum pennellii*. *Plant Cell* **32**, 81–99 (2020).
29. W. Ji, S. Mandal, Y. H. Rezenom, T. D. McKnight, Specialized metabolism by trichome-enriched Rubisco and fatty acid synthase components. *Plant Physiol.* **191**, 1199–1213 (2023).
30. H. Feng, L. Acosta-Gamboa, L. H. Kruse, J. D. Tracy, S. H. Chung, A. R. Nava Fereira, S. Shakir, H. Xu, G. Sunter, M. A. Gore, C. L. Casteel, G. D. Moghe, G. Jander, Acylsugars protect *Nicotiana benthamiana* against insect herbivory and desiccation. *Plant Mol. Biol.* **109**, 505–522 (2021).
31. N. M. Van Dam, J. D. Hare, Biological activity of *Datura wrightii* glandular trichome exudate against *Manduca sexta* larvae. *J. Chem. Ecol.* **24**, 1529–1549 (1998).
32. O. T. Chortyk, J. G. Pomonis, A. W. Johnson, Syntheses and characterizations of insecticidal sucrose esters. *J. Agric. Food Chem.* **44**, 1551–1557 (1996).
33. B. M. Leckie, D. M. De Jong, M. A. Mutschler, Quantitative trait loci increasing acylsugars in tomato breeding lines and their impacts on silverleaf whiteflies. *Mol. Breed.* **30**, 1621–1634 (2012).

34. V. T. Luu, A. Weinhold, C. Ullah, S. Dressel, M. Schoettner, K. Gase, E. Gaquerel, S. Xu, I. T. Baldwin, O-Acyl sugars protect a wild tobacco from both native fungal pathogens and a specialist herbivore. *Plant Physiol.* **174**, 370–386 (2017).
35. P. Fan, A. M. Miller, A. L. Schilmiller, X. Liu, I. Ofner, A. D. Jones, D. Zamir, R. L. Last, *In vitro* reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E239–E248 (2016).
36. A. L. Schilmiller, G. D. Moghe, P. Fan, B. Ghosh, J. Ning, A. D. Jones, R. L. Last, Functionally divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. *Plant Cell* **27**, 1002–1017 (2015).
37. A. L. Schilmiller, A. L. Charbonneau, R. L. Last, Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. *Proc. Natl. Acad. Sci.* **109**, 16377–16382 (2012).
38. J. Ning, G. D. Moghe, B. Leong, J. Kim, I. Ofner, Z. Wang, C. Adams, A. D. Jones, D. Zamir, R. L. Last, A feedback-insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. *Plant Physiol.* **169**, 1821–1835 (2015).
39. P. Fan, P. Wang, Y.-R. Lou, B. J. Leong, B. M. Moore, C. A. Schenck, R. Combs, P. Cao, F. Brandizzi, S.-H. Shiu, R. L. Last, Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. *eLife* **9**, e56717 (2020).
40. E. Korenblum, Y. Dong, J. Szymanski, S. Panda, A. Jozwiak, H. Massalha, S. Meir, I. Rogachev, A. Aharoni, Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 3874–3883 (2020).
41. V. Falara, T. A. Akhtar, T. T. Nguyen, E. A. Spyropoulou, P. M. Bleeker, I. Schauvinhold, Y. Matsuba, M. E. Bonini, A. L. Schilmiller, R. L. Last, R. C. Schuurink, E. Pichersky, The tomato terpene synthase gene family. *Plant Physiol.* **157**, 770–789 (2011).
42. S. Gao, Y. Gao, C. Xiong, G. Yu, J. Chang, Q. Yang, C. Yang, Z. Ye, The tomato B-type cyclin gene, *SlCycB2*, plays key roles in reproductive organ development, trichome initiation, terpenoids biosynthesis and *Prodenia litura* defense. *Plant Sci.* **262**, 103–114 (2017).
43. E. A. Spyropoulou, M. A. Haring, R. C. Schuurink, Expression of Terpenoids 1, a glandular trichome-specific transcription factor from tomato that activates the terpene synthase 5 promoter. *Plant Mol. Biol.* **84**, 345–357 (2014).
44. J.-I. Chun, S.-M. Kim, H. Kim, J.-Y. Cho, H.-W. Kwon, J.-I. Kim, J.-K. Seo, C. Jung, J.-H. Kang, *SlHair2* regulates the initiation and elongation of type I trichomes on tomato leaves and stems. *Plant Cell Physiol.* **62**, 1446–1459 (2021).
45. C. C. N. van Schie, M. A. Haring, R. C. Schuurink, Tomato linalool synthase is induced in trichomes by jasmonic acid. *Plant Mol. Biol.* **64**, 251–263 (2007).
46. J. D. Masucci, J. W. Schiefelbein, The rhd6 mutation of *Arabidopsis thaliana* alters root-hair initiation through an auxin- and ethylene-associated process. *Plant Physiol.* **106**, 1335–1346 (1994).
47. K. Schneider, J. Mathur, K. Boudonck, B. Wells, L. Dolan, K. Roberts, The *ROOT HAIRLESS 1* gene encodes a nuclear protein required for root hair initiation in *Arabidopsis*. *Genes Dev.* **12**, 2013–2021 (1998).
48. B. Menand, K. Yi, S. Jouannic, L. Hoffmann, E. Ryan, P. Linstead, D. G. Schaefer, L. Dolan, An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* **316**, 1477–1480 (2007).
49. B. Ghosh, A. Jones, Profiling, characterization, and analysis of natural and synthetic acylsugars (sugar esters). *Anal. Methods* **9**, 892–905 (2017).
50. C. Brooks, V. Nekrasov, Z. B. Lippman, J. Van Eck, Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol.* **166**, 1292–1297 (2014).
51. B. J. Leong, S. Hurney, P. Fiesel, T. M. Anthony, G. Moghe, A. D. Jones, R. L. Last, Identification of BAHD acyltransferases associated with acylinositol biosynthesis in *Solanum quitoense* (naranjilla). *Plant Direct* **6**, e415 (2022).
52. A. Kawaguchi, T. Yoshimura, S. Okuda, A new method for the preparation of acyl-CoA thioesters. *J. Biochem.* **89**, 337–339 (1981).
53. The Tomato Genome Consortium, The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641 (2012).
54. P. D. Fiesel, R. E. Kerwin, A. D. Jones, R. L. Last, Trading acyls and swapping sugars: Metabolic innovations in *Solanum* trichomes. bioRxiv 542877 [Preprint] (2023). https://doi.org/10.1101/2023.06.05.542877.
55. G. D. Moghe, B. J. Leong, S. Hurney, A. D. Jones, R. L. Last, Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife* **6**, e28468 (2017).
56. A. L. Schilmiller, K. Gilgallon, B. Ghosh, A. D. Jones, R. L. Last, Acylsugar acylhydrolases: Carboxylesterase-catalyzed hydrolysis of acylsugars in tomato trichomes. *Plant Physiol.* **170**, 1331–1344 (2016).
57. Y. Wang, X. Wang, A. H. Paterson, Genome and gene duplications and gene expression divergence: A view from plants. *Ann. N. Y. Acad. Sci.* **1256**, 1–14 (2012).
58. E. W. Ganko, B. C. Meyers, T. J. Vision, Divergence in expression between duplicated genes in Arabidopsis. *Mol. Biol. Evol.* **24**, 2298–2309 (2007).
59. D. J. Kliebenstein, A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PLOS ONE* **3**, e1838 (2008).
60. P. Fan, A. M. Miller, X. Liu, A. D. Jones, R. L. Last, Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. *Nat. Commun.* **8**, 2080 (2017).
61. D. B. Lybrand, T. M. Anthony, A. D. Jones, R. L. Last, An Integrated analytical approach reveals trichome acylsugar metabolite diversity in the wild tomato *Solanum pennellii*. *Metabolites* **10**, 401 (2020).
62. S. M. Hurney, "Strategies for profiling and discovery of acylsugar specialized metabolites," thesis, Michigan State University, East Lansing, MI (2018).
63. T. Louveau, A. Osbourn, The sweet side of plant-specialized metabolism. *Cold Spring Harb. Perspect. Biol.* **11**, a034744 (2019).
64. J. M. Nicol, S. J. Turner, D. L. Coyne, L. den Nijs, S. Hockland, Z. T. Maafi, Current nematode threats to world agriculture, in *Genomics and Molecular Genetics of Plant-Nematode Interactions*, J. Jones, G. Gheysen, C. Fenoll, Eds. (Springer, 2011), pp. 21–43.
65. T. Särkinen, L. Bohs, R. G. Olmstead, S. Knapp, A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evol. Biol.* **13**, 241 (2013).
66. A. W. Johnson, R. F. Severson, Leaf surface chemistry of tobacco budworm resistant tobacco. *J. Agric. Entomol.* **10**, 23–32 (1984).
67. X. Liu, M. Enright, C. S. Barry, A. D. Jones, Profiling, isolation and structure elucidation of specialized acylsucrose metabolites accumulating in trichomes of *Petunia* species. *Metabolomics* **13**, 85 (2017).
68. O. T. Chortyk, S. J. Kays, Q. Teng, Characterization of insecticidal sugar esters of *Petunia*. *J. Agric. Food Chem.* **45**, 270–275 (1997).
69. C. Dutra, M. V. Cesio, P. Moyna, H. Heinzen, Acyl sucroses from *Salpichroa origanifolia*. *Nat. Prod. Commun.* **3**, 539–542 (2008).
70. B. Ghosh, T. C. Westbrook, A. D. Jones, Comparative structural profiling of trichome specialized metabolites in tomato (*Solanum lycopersicum*) and *S. habrochaites*: Acylsugar profiles revealed by UHPLC/MS and NMR. *Metabolomics* **10**, 496–507 (2014).
71. S. L. Blauth, G. A. Churchill, M. A. Mutschler, Identification of quantitative trait loci associated with acylsugar accumulation using intraspecific populations of the wild tomato, Lycopersicon pennellii. *Theor. Appl. Genet.* **96**, 458–467 (1998).
72. D. M. Jackson, O. T. Chortyk, M. G. Stephenson, A. W. Johnson, C. D. Harlow, A. M. Simmons, V. A. Sisson, Potential of *Nicotiana* species to produce sugar esters. *Tob. Sci.* **42**, 1–9 (1998).
73. T. Matsuzaki, Y. Shinozaki, S. Suhara, M. Ninomiya, H. Shigematsu, A. Koiwai, Isolation of glycolipids from the surface lipids of *Nicotiana bigelovii* and their distribution in *Nicotiana* species. *Agric. Biol. Chem.* **53**, 3079–3082 (1989).
74. T. Matsuzaki, Y. Shinozaki, S. Suhara, T. Tobita, H. Shigematsu, A. Koiwai, Leaf surface glycolipids from *Nicotiana acuminata* and *Nicotiana pauciflora*. *Agric. Biol. Chem.* **55**, 1417–1419 (1991).
75. Y. Shinozaki, T. Matsuzaki, S. Suhara, T. Tobita, H. Shigematsu, A. Koiwai, New types of glycolipids from the surface lipids of *Nicotiana umbratica*. *Agric. Biol. Chem.* **55**, 751–756 (1991).
76. E. Gagnon, R. Hilgenhof, A. Orejuela, A. McDonnell, G. Sablok, X. Aubriot, L. Giacomin, Y. Gouvêa, T. Bragionis, J. R. Stehmann, L. Bohs, S. Dodsworth, C. Martine, P. Poczai, S. Knapp, T. Särkinen, Phylogenomic discordance suggests polytomies along the backbone of the large genus *Solanum*. *Am. J. Bot.* **109**, 580–601 (2022).
77. T. Särkinen, P. Poczai, G. E. Barboza, G. M. van der Weerden, M. Baden, S. Knapp, A revision of the Old World Black Nightshades (Morelloid clade of *Solanum* L., Solanaceae). *PhytoKeys* **106**, 1–223 (2018).
78. Y.-R. Lou, T. M. Anthony, P. D. Fiesel, R. E. Arking, E. M. Christensen, A. D. Jones, R. L. Last, It happened again: Convergent evolution of acylglucose specialized metabolism in black nightshade and wild tomato. *Sci. Adv.* **7**, eabj8726 (2021).
79. B. J. Leong, S. M. Hurney, P. D. Fiesel, G. D. Moghe, A. D. Jones, R. L. Last, Specialized metabolism in a nonmodel nightshade: Trichome acylinositol biosynthesis. *Plant Physiol.* **183**, 915–924 (2020).
80. J. A. Birchler, H. Yang, The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* **34**, 2466–2474 (2022).
81. M. Benderoth, S. Textor, A. J. Windsor, T. Mitchell-Olds, J. Gershenzon, J. Kroymann, Positive selection driving diversification in plant secondary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9118–9123 (2006).
82. P. J. Wittkopp, Evolution of gene expression, in *The Princeton Guide to Evolution*, J. B. Losos, D. A. Baum, D. J. Futuyma, H. E. Hoekstra, R. E. Lenski, A. J. Moore, C. L. Peichel, D. Schluter, M. C. Whitlock, Eds. (Princeton Univ. Press, 2013), pp. 413–419.
83. D. Tautz, Evolution of transcriptional regulation. *Curr. Opin. Genet. Dev.* **10**, 575–579 (2000).
84. G. A. Wray, M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer, M. V. Rockman, L. A. Romano, The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419 (2003).
85. N. Panchy, M. Lehti-Shiu, S. H. Shiu, Evolution of gene duplication in plants. *Plant Physiol.* **171**, 2294–2316 (2016).

86. H.-W. Nützmann, A. Huang, A. Osbourn, Plant metabolic clusters—From genetics to genomics. *New Phytol.* **211**, 771–789 (2016).
87. M. Alonge, X. Wang, M. Benoit, S. Soyk, L. Pereira, L. Zhang, H. Suresh, S. Ramakrishnan, F. Maumus, D. Ciren, Y. Levy, T. H. Harel, G. Shalev-Schlosser, Z. Amsellem, H. Razifard, A. L. Caicedo, D. M. Tieman, H. Klee, M. Kirsche, S. Aganezov, T. R. Ranallo-Benavidez, Z. H. Lemmon, J. Kim, G. Robitaille, M. Kramer, S. Goodwin, W. R. M. Combie, S. Hutton, J. Van Eck, J. Gillis, Y. Eshed, F. J. Sedlazeck, E. van der Knaap, M. C. Schatz, Z. B. Lippman, Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**, 145–161.e23 (2020).
88. P. S. Hosmani, M. Flores-Gonzalez, H. van de Geest, F. Maumus, L. V. Bakker, E. Schijlen, J. van Haarst, J. Cordewener, G. Sanchez-Perez, S. Peters, Z. Fei, J. J. Giovannoni, L. A. Mueller, S. Saha, An improved *de novo* assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv 767764 [Preprint] (2019). https://doi.org/10.1101/767764.
89. L. Gao, I. Gonda, H. Sun, Q. Ma, K. Bao, D. M. Tieman, E. A. Burzynski-Chang, T. L. Fish, K. A. Stromberg, G. L. Sacks, T. W. Thannhauser, M. R. Foolad, M. J. Diez, J. Blanca, J. Canizares, Y. Xu, E. van der Knaap, S. Huang, H. J. Klee, J. J. Giovannoni, Z. Fei, The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* **51**, 1044–1051 (2019).
90. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
91. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
92. R. Leinonen, H. Sugawara, M. Shumway, The Sequence Read Archive. *Nucleic Acids Res.* **39**, D19–D21 (2011).
93. Y. Wang, F. Song, J. Zhu, S. Zhang, Y. Yang, T. Chen, B. Tang, L. Dong, N. Ding, Q. Zhang, Z. Bai, X. Dong, H. Chen, M. Sun, S. Zhai, Y. Sun, L. Yu, L. Lan, J. Xiao, X. Fang, H. Lei, Z. Zhang, W. Zhao, GSA: Genome Sequence Archive. *Genomics Proteomics Bioinformatics* **15**, 14–18 (2017).
94. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
95. A. Dobin, T. R. Gingeras, Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics* **51**, 11.14.1–11.14.19 (2015).
96. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
97. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, M. A. DePristo, From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.1-11.10.33 (2013).
98. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
99. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
100. S. Anders, P. T. Pyl, W. Huber, HTSeq–A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
101. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
102. P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
103. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
104. D. J. McCarthy, K. R. Campbell, A. T. L. Lun, Q. F. Wills, Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).
105. A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szczesniak, D. J. Gaffney, L. L. Elo, X. Zhang, A. Mortazavi, A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13 (2016).
106. J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, A. Bateman, Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
107. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
108. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
109. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
110. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
111. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Y. Lam, GGTREE: Anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
112. T. Hothorn, F. Bretz, P. Westfall, Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363 (2014).
113. D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison III, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
114. P. Fan, G. D. Moghe, R. L. Last, Comparative biochemistry and *in vitro* pathway reconstruction as powerful partners in studies of metabolic diversity. *Methods Enzymol.* **576**, 1–17 (2016).
115. B. J. Leong, D. B. Lybrand, Y. R. Lou, P. Fan, A. L. Schilmiller, R. L. Last, Evolution of metabolic novelty: A trichome-expressed invertase creates specialized metabolic diversity in wild tomato. *Sci. Adv.* **5**, eaaw3754 (2019).
116. S. McCormick, Transformation of tomato with *Agrobacterium tumefaciens*, in *Plant Tissue Culture Manual*: *Supplement* 7, K. Lindsey, Ed. (Springer, 1991), pp. 311–319.
117. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
118. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
119. X. Wang, L. Gao, C. Jiao, S. Stravoravdis, P. S. Hosmani, S. Saha, J. Zhang, S. Mainiero, S. R. Strickler, C. Catala, G. B. Martin, L. A. Mueller, J. Vrebalov, J. J. Giovannoni, S. Wu, Z. Fei, Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat. Commun.* **11**, 5817 (2020).
120. A. Bolger, F. Scossa, M. E. Bolger, C. Lanz, F. Maumus, T. Tohge, H. Quesneville, S. Alseekh, I. Sørensen, G. Lichtenstein, E. A. Fich, M. Conte, H. Keller, K. Schneeberger, R. Schwacke, I. Ofner, J. Vrebalov, Y. Xu, S. Osorio, S. A. Aflitos, E. Schijlen, J. M. Jiménez-Goméz, M. Ryngajllo, S. Kimura, R. Kumar, D. Koenig, L. R. Headland, J. N. Maloof, N. Sinha, R. C. H. J. Van Ham, R. K. Lankhorst, L. Mao, A. Vogel, B. Arsova, R. Panstruga, Z. Fei, J. K. C. Rose, D. Zamir, F. Carrari, J. J. Giovannoni, D. Weigel, B. Usadel, A. R. Fernie, The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* **46**, 1034–1038 (2014).
121. A. F. Powell, A. Feder, J. Li, M. H.-W. Schmidt, L. Courtney, S. Alseekh, E. M. Jobson, A. Vogel, Y. Xu, D. Lyon, K. Dumschott, M. McHale, R. Sulpice, K. Bao, R. Lal, A. Duhan, A. Hallab, A. K. Denton, M. E. Bolger, A. R. Fernie, S. R. Hind, L. A. Mueller, G. B. Martin, Z. Fei, C. Martin, J. J. Giovannoni, S. R. Strickler, B. Usadel, A *Solanum lycopersicoides* reference genome facilitates insights into tomato specialized metabolism and immunity. *Plant J.* **110**, 1791–1810 (2022).
122. G. M. Pham, J. P. Hamilton, J. C. Wood, J. T. Burke, H. Zhao, B. Vaillancourt, S. Ou, J. Jiang, C. R. Buell, Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience* **9**, giaa100 (2020).
123. L. Barchi, M. T. Rabanus-Wallace, J. Prohens, L. Toppino, S. Padmarasu, E. Portis, G. L. Rotino, N. Stein, S. Lanteri, G. Giuliano, Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *Plant J.* **107**, 579–596 (2021).
124. A. F. Powell, J. Zhang, D. Hauser, J. A. Vilela, A. Hu, D. J. Gates, L. A. Mueller, F.-W. Li, S. R. Strickler, S. D. Smith, Genome sequence for the blue-flowered Andean shrub *Iochroma cyaneum* reveals extensive discordance across the berry clade of Solanaceae. *Plant Genome* **15**, e20223 (2022).
125. F. Zhang, F. Qiu, J. Zeng, Z. Xu, Y. Tang, T. Zhao, Y. Gou, F. Su, S. Wang, X. Sun, Z. Xue, W. Wang, C. Yang, L. Zeng, X. Lan, M. Chen, J. Zhou, Z. Liao, Revealing evolution of tropane alkaloid biosynthesis by analyzing two genomes in the Solanaceae family. *Nat. Commun.* **14**, 1446 (2023).
126. Y.-L. Cao, Y. Li, Y.-F. Fan, Z. Li, K. Yoshida, J.-Y. Wang, X.-K. Ma, N. Wang, N. Mitsuda, T. Kotake, T. Ishimizu, K.-C. Tsai, S.-C. Niu, D. Zhang, W.-H. Sun, Q. Luo, J.-H. Zhao, Y. Yin, B. Zhang, J.-Y. Wang, K. Qin, W. An, J. He, G.-L. Dai, Y.-J. Wang, Z.-G. Shi, E.-N. Jiao, P.-J. Wu, X. Liu, B. Liu, X.-Y. Liao, Y.-T. Jiang, X. Yu, Y. Hao, X.-Y. Xu, S.-Q. Zou, M.-H. Li, Y.-Y. Hsiao, Y.-F. Lin, C.-K. Liang, Y.-Y. Chen, W.-L. Wu, H.-C. Lu, S.-R. Lan, Z.-W. Wang, X. Zhao, W.-Y. Zhong, C.-M. Yeh, W.-C. Tsai, Y. Van de Peer, Z.-J. Liu, Wolfberry genomes and the evolution of *Lycium* (Solanaceae). *Commun. Biol.* **4**, 1–13 (2021).
127. A. Navarro-Quezada, K. Gase, R. K. Singh, S. P. Pandey, I. T. Baldwin, *Nicotiana attenuata* genome reveals genes in the molecular machinery behind remarkable adaptive phenotypic plasticity, in *The Tobacco Plant Genome* (Springer Nature, 2020), pp. 211–229.
128. A. Bombarely, M. Moser, A. Amrad, M. Bao, L. Bapaume, C. S. Barry, M. Bliek, M. R. Boersma, L. Borghi, R. Bruggmann, M. Bucher, N. D'Agostino, K. Davies, U. Druege, N. Dudareva, M. Egea-Cortines, M. Delledonne, N. Fernandez-Pozo, P. Franken, L. Grandont, J. S. Heslop-Harrison, J. Hintzsche, M. Johns, R. Koes, X. Lv, E. Lyons, D. Malla, E. Martinoia, N. S. Mattson, P. Morel, L. A. Mueller, J. Muhlemann, E. Nouri, V. Passeri, M. Pezzotti, Q. Qi, D. Reinhardt, M. Rich, K. R. Richert-Pöggeler, T. P. Robbins, M. C. Schatz,

M. E. Schranz, R. C. Schuurink, T. Schwarzacher, K. Spelt, H. Tang, S. L. Urbanus, M. Vandenbussche, K. Vijverberg, G. H. Villarino, R. M. Warner, J. Weiss, Z. Yue, J. Zethof, F. Quattrocchio, T. L. Sims, C. Kuhlemeier, Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* **2**, 16074 (2016).

129. S. Wu, K. H. Lau, Q. Cao, J. P. Hamilton, H. Sun, C. Zhou, L. Eserman, D. C. Gemenet, B. A. Olukolu, H. Wang, E. Crisovan, G. T. Godden, C. Jiao, X. Wang, M. Kitavi, N. Manrique-Carpintero, B. Vaillancourt, K. Wiegert-Rininger, X. Yang, K. Bao, J. Schaff, J. Kreuze, W. Gruneberg, A. Khan, M. Ghislain, D. Ma, J. Jiang, R. O. M. Mwanga, J. Leebens-Mack, L. J. M. Coin, G. C. Yencho, C. R. Buell, Z. Fei, Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nat. Commun.* **9**, 4580 (2018).
130. F. Denoeud, L. Carretero-Paulet, A. Dereeper, G. Droc, R. Guyot, M. Pietrella, C. Zheng, A. Alberti, F. Anthony, G. Aprea, J.-M. Aury, P. Bento, M. Bernard, S. Bocs, C. Campa, A. Cenci, M.-C. Combes, D. Crouzillat, C. D. Silva, L. Daddiego, F. De Bellis, S. Dussert, O. Garsmeur, T. Gayraud, V. Guignon, K. Jahn, V. Jamilloux, T. Joët, K. Labadie, T. Lan, J. Leclercq, M. Lepelley, T. Leroy, L.-T. Li, P. Librado, L. Lopez, A. Muñoz, B. Noel, A. Pallavicini, G. Perrotta, V. Poncet, D. Pot, M. R. Priyono, M. Rouard, J. Rozas, C. Tranchant-Dubreuil, R. Van Buren, Q. Zhang, A. C. Andrade, X. Argout, B. Bertrand, A. de Kochko, G. Graziosi, R. J. Henry, R. M. Jayarama, C. Nagai, S. Rounsley, D. Sankoff, G. Giuliano, V. A. Albert, P. Wincker, P. Lashermes, The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **345**, 1181–1184 (2014).
131. O. Jaillon, J.-M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Hugueney, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyère, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M. E. Pè, G. Valle, M. Morgante, M. Caboche, A.-F. Adam-Blondon, J. Weissenbach, F. Quétier, P. Wincker, The French–Italian Public Consortium for Grapevine Genome Characterization, The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
132. Y. Wang, H. Tang, J. D. Debarry, X. Tan, J. Li, X. Wang, T. H. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger, A. H. Paterson, *MCScanX*: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).

Acknowledgments: We thank members of the Michigan State University RTSF Mass Spectrometry and Metabolomics Core Facilities for LC-MS analysis support, especially A. Schilmiller, as well as D. Holmes and L. Xie of the Michigan State University Max T. Rogers NMR Core. We also thank D. Kliebenstein, S.-H. Shiu, and current and former members of the Last laboratory for useful feedback during manuscript preparation. **Funding:** This work was supported by the following: National Science Foundation grants 1546617 (R.L.L. and A.D.J.) and 2218206 (R.L.L. and R.E.K.), National Institutes of Health National Institute of General Medical Sciences grant NIGMS-T32-GM110523 (P.D.F.), and Natural Science Foundation of Zhejiang province, China, grant LZ22C150005 (P.F.). **Author contributions:** Conceptualization: R.E.K. and R.L.L. Methodology: R.E.K., J.E.H., and P.D.F. Investigation: R.E.K., J.E.H., P.D.F., and P.F. Formal analysis: R.E.K., P.D.F., A.D.J., and Y.-R.L. Resources: R.E.K., Y.-R.L., P.F., and R.L.L. Visualization: R.E.K. and P.D.F. Supervision: R.E.K. and R.L.L. Writing—original draft: R.E.K., J.E.H., and P.D.F. Writing—review and editing: R.E.K., J.E.H., P.D.F., Y.-R.L., P.F., A.D.J., and R.L.L. Funding acquisition: R.L.L., A.D.J., and R.E.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Supplementary tables have been deposited to Dryad: https://doi.org/10.5061/dryad.8cz8w9gzh. Requests for biological materials should be submitted to R.L.L. at lastr@msu.edu or R.E.K. at kerwinra@msu.edu.

Submitted 11 December 2023
Accepted 20 March 2024
Published 24 April 2024
10.1126/sciadv.adn3991