ANALYSIS OF ACYLGLUCOSE AND ACYLINOSITOL BIOSYNTHETIC MECHANISMS
IN SOLANACEAE SPECIES

By

Bryan James Leong

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Biology - Doctor of Philosophy

2019

ABSTRACT

ANALYSIS OF ACYLGLUCOSE AND ACYLINOSITOL BIOSYNTHETIC MECHANISMS
IN SOLANACEAE SPECIES

By

Bryan James Leong

Plant specialized metabolites are compounds produced in specific lineages whose synthesis most

likely arose as adaptations to specific ecological conditions. One group of these compounds are

acylsugars, which are produced in the glandular trichomes of many species in the Solanaceae

family. Research indicates that differences in acylsugar core and acyl chains alter the effects of

these compounds on insect deterrence or mortality. We can use natural chemical diversity to

study what effects different acylsugars have on insects using *in vitro* techniques, but *in planta*

methods are preferred. Identification of biosynthetic genes allows us to genetic engineer these

pathways into plants to more precisely study the impact of these compounds *in planta*. This

study is comprised of two projects characterizing biosynthetic pathways present in *Solanum*

pennellii and *Solanum quitoense*. We are interested in how acylglucoses are made in *S. pennellii*

– the first project focused on identification of an acylsucrose fructofuranosidase in *S. pennellii*, a

wild relative of cultivated tomato. This *S. pennellii* glycosyl hydrolase converts acylsucroses to

acylglucoses, the predominant acylsugar in several accessions of this species. *In vitro* and *in*

planta data show this enzyme accepts *S. pennellii* P-type acylsucroses as substrates, but not F-

type acylsucroses from cultivated tomato. Whether the plant produces P- or F-type acylsucroses

is determined by different enzymatic activities of acyltransferases that alter acylation pattern in

cultivated tomato and *S. pennellii* – representing a three gene epistatic interaction between those

genes and the acylsucrose fructofuranosidase. Acylinositols represent a novel type of acylsugar

produced in the *Solanum* genus, however, the biosynthetic pathway has not been studied. My

second project involved characterization of the biosynthesis of these *myo*-inositol containing acylsugars in *Solanum quitoense*. VIGS analysis suggested this BAHD acyltransferase acetylates triacylinositols and i*n vitro* analysis confirmed that hypothesis. Further VIGS analysis and i*n vitro* assays identified an inositol acyltransferase that possessed characteristics matching other acylsugar biosynthetic enzymes, but this BAHD acyltransferase acylated *myo*-inositol at an aberrant acylation position. Alternative hypotheses were offered to reconcile the *in vitro* and *in vivo* results. Together, these two projects represent first steps towards understanding how the array of acylsugars in Solanaceae species are produced.

ACKNOWLEDGEMENTS

Finally, I want to thank my parents and my brother. I don't really know what to say – not because I have so little to thank you for, but because there is far too much. I would not have made it this far without your guidance, support, and love. I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

ACN - Acetonitrile

aiC5 – Anteiso C5

ANOVA – One-way analysis of variance

ASAT – Acylsugar acyltransferase

ASFF1 – Acylsucrose fructofuranosidase 1

ASH – Acylsugar acylhydrolase

BAHD – acyltransferase class; each letter represents one of the first four enzymes identified in this class

BIL – Backcrossed introgression line

Bp – Basepair

BPI – Base peak intensity

Cas9 – CRISPR associated protein 9

cDNA – Complimentary deoxyribonucleic acid

CDS – Coding sequence

CoA – Coenzyme A

CRISPR – Clustered regularly interspaced short palindromic repeats

Da – Dalton

dATP – deoxyadenosine triphosphate

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic acid

DTT – Dithiothreitol

dTTP – deoxythymidine triphosphate

DW – Dry weight

E-64 – *trans*-epoxysuccinyl-L-leucylamido(4-guanidino)butane

EDTA – ethylenediaminetetraacetic acid

ESI – Electrospray ionization

F-type – Furanose type

GFP – Green fluorescent protein

GH – Glycosyl hydrolase

GOI – Gene of interest

HPLC – High performance liquid chromatography

IAT – Inositol acyltransferase

iC5 – Iso C5

IL – Introgression line

IPA - Isopropanol

IPMS – Isopropylmalate synthase

IPTG - Isopropyl-β-D-thiogalactoside

IS – Internal standard

JTT – Jones-Taylor-Thornton

Kb – Kilobase

LB – Lysogeny broth

LC-MS – Liquid chromatography-mass spectrometry

Leu – Leucine

LIC – Ligation-independent cloning

Mb – Megabase

MSA – Multiple sequence alignment

Mya – Million years ago

m/z – mass-to-charge ratio

Ni-NTA – Nickel-nitrilotriacetic acid

NMR – Nuclear magnetic resonance

OD600 – Optical density at 600 nm

ORF – Open reading frame

P-type – Pyranose type

PCR – Polymerase chain reaction

PDS – Phytoene desaturase

PMSF - Phenylmethansesulfonylfluoride

PPFD – photosynthetic photo flux density

PVPP – Polyvinylpolypyrrolidone

qRT-PCR – Quantitative reverse transcriptase polymerase chain reaction

QTL – Quantitative trait loci

qToF – Quadrupole time-of-flight

RNA – Ribonucleic acid

RNAseq – RNA sequencing

SCPL – Serine carboxypeptidase-like

sgRNA – Small guide RNA

TAIAT – Tri-acylinositol acetyltransferase

UDP – Uridine diphosphate

UPLC – Ultra high performance liquid chromatography

VIGS – Virus induced gene silencing

WT – Wild-type

Chapter 1. Literature review

Body

Introduction to specialized metabolites and specialized structures

Plant specialized metabolites are a group of compounds that are restricted in their phylogenetic distribution, in sharp contrast to general metabolites – those present in most, if not all plants. These "specialized" metabolites are produced by plants to mediate the environmental interactions, ranging from coping with abiotic stresses, to fending off herbivores, or attracting pollinators or beneficial microbes. Estimates of the number of plant specialized metabolites have been made in the hundreds of thousands or more (Pichersky and Lewinsohn, 2011). The sheer size of the plant kingdom, combined with the mostly unexplored metabolic landscape, leads to the hypothesis that there is a tremendous amount of metabolic diversity yet to be discovered. The canonical classes of specialized metabolites include alkaloids, terpenoids, and phenylpropanoids, but plenty of smaller classes of specialized metabolites have been characterized. Glucosinolates are a prominent example of these smaller classes that are primarily restricted to the Brassicales order (Halkier and Gershenzon, 2006). These sulfur-based compounds are deterrents or toxic to an array of plant enemies that include: birds, insects, nematodes, bacteria, fungus, and mammals, but can also attract specialists (Halkier and Gershenzon, 2006). Another example is acylsugars, a specialized metabolite class produced across the Solanaceae family. Unfortunately, many of these specialized metabolites can have negative effects on the plants themselves due to the nature of their biological activities.

Plants mitigate these effects through compartmentalization of these specialized metabolites. Enzyme compartmentalization from substrates is one mechanism demonstrated by glucosinolates. The activating enzymes – myrosinases – catalyze the hydrolysis of glucose from the glucosinolate leading to a highly reactive aglycone (Halkier and Gershenzon, 2006).

Myrosinases are localized in the idioblasts – cells that are separate, but adjacent to those containing the glucosinolates in *Arabidopsis thaliana* (Halkier and Gershenzon, 2006). Another case of compartmentalization is in monoterpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. Different parts of the biosynthetic pathway are localized to the parenchyma, the epidermis, and laticifers/idioblasts, spanning multiple cellular compartments in each cell type (Courdavault et al., 2014). Additionally, some of the products – such as vindoline – are stored in the vacuole, presumably to sequester these compounds (Courdavault et al., 2014).

A prime example of tissue compartmentalization is in specialized structures called secretory glandular trichomes. These are epidermal protuberances present on the surface of many angiosperms. There are numerous examples of glandular trichomes playing host to biosynthetic pathways for protective phytochemicals. Glandular trichomes in the *Mentha* genus are the production and storage site for monoterpene oils containing natural pesticides such as pulegone, menthone, and carvone (Ahkami et al., 2015). The Nobel prize related anti-malarial drug, artemisinin, is produced in glandular trichomes of *Artemisia annua*. Methylketones are insecticidal compounds produced in the glandular trichomes of *Solanum habrochaites* ssp *glabratum* (Fridman et al., 2005). In cultivated tomato, there are multiple trichome types. One example of a glandular trichome in cultivated tomato is type I/IV glandular trichomes that contain long or short stalks, respectively. These two trichome types have a single gland cell at the tip of the stalk. These glandular cells serve as biochemical factories producing a variety of phytochemicals – one such group is acylsugars.

Acylsugar structural diversity in the Solanaceae

Acylsugars are a group of compounds – characterized in a fraction of species in the Solanaceae – mainly produced in the glandular trichomes of species in the Solanaceae, though there are documented examples in other plant families such as the Martyniaceae, Rosaceae, and Geraniaceae (Liu et al., 2019). Chemical screening revealed the presence of these acylsugars throughout the Solanaceae in several genera, including *Salpiglossis* (Moghe et al., 2017)*, Petunia* (Liu et al., 2017), *Nicotiana* (Matsuzaki et al., 1989; Matsuzaki et al., 1991; Shinozaki et al., 1991; Matsuzaki et al., 1992), *Datura* (King and Calhoun, 1988) , *Physalis* (Maldonado et al., 2006; Zhang et al., 2016), and the large *Solanum* genus (King et al., 1986; Burke et al., 1987; Herrera-Salgado et al., 2005; Ghosh et al., 2014).

Acylsugars are composed of two components: a sugar core, commonly sucrose, and acyl chains derived from amino acids or fatty acids. Acyl chain diversity is one component that can result in an expansion of acylsugar structural diversity. Acyl groups in acylsugars have been documented to be as short as two carbons in acetyl groups (Ghosh et al., 2014), and as long as twenty carbons in *Solanum lanceolatum* (Herrera-Salgado et al., 2005) (Figure 1.1). Acyl chains vary in branching pattern – either iso- or anteiso-branched acyl chains – apparently derived from branched chain amino acid metabolism. Acylsugars in the *Petunia* genus are also acylated by a malonyl moiety (Liu et al., 2017) (Figure 1.1). Finally, sugar cores play a role in the diversity of acylsugars produced in different Solanaceae spp (Figure 1.1).

Much of the analytical chemistry characterization of acylsugars has been on sucrose-based compounds (King et al., 1986; Maldonado et al., 2006; Ghosh et al., 2014; Liu et al., 2017), however there are instances of different sugar cores such as glucose (Burke et al., 1987; King and Calhoun, 1988) and *myo*-inositol (Herrera-Salgado et al., 2005). Sucrose is the

predominant sugar core and was identified in genera across the family, including *Solanum* (Ghosh et al., 2014), *Petunia* (Liu et al., 2017), and in *Salpiglossis sinuata* (Moghe et al., 2017). Acylsugars containing hexose cores were characterized in a small number of species dispersed across the family, such as acylated inositols in *Solanum quitoense* (Hurney, 2018) and acylglucoses in *Solanum nigrum* (Moghe et al., 2017, Unpublished). Glucose-based acylsugars were demonstrated to be sporadically distributed in species of the *Solanum* (Shapiro et al., 1994; Leong et al., 2019), *Nicotiana* (Matsuzaki et al., 1991), and *Petunia* (Chortyk et al., 1997) genera.

Variation in acyl chains and acylation positions on the sugar suggest tremendous potential to generate acylsugar diversity. For example, much of the existing acylsugar diversity is unexplored and uncharacterized. Liquid chromatography-mass spectrometry (LC-MS) analysis of *Petunia axillaris* leaf metabolites reveal numerous chromatographic peaks that were predicted as acylsugars (Liu et al., 2017). LC-MS characterized metabolites extend far beyond the smaller number of acylsugars that were structurally resolved through NMR characterization. This point is reinforced by work on *Salpiglossis sinuata*, where NMR structures were established for 16 acylsucroses (Hurney, 2018), out of 400 chromatographically-separable acylsucroses annotated by LC/MS of leaf-surface extracts (Hurney, 2018). It is clear at this point that the assortment of acylsugars is relatively unexplored. One of the major ways to understand how acylsugar structural complexity is generated in different plant species is to study the biochemical and genetic basis for acylsugar biosynthesis.
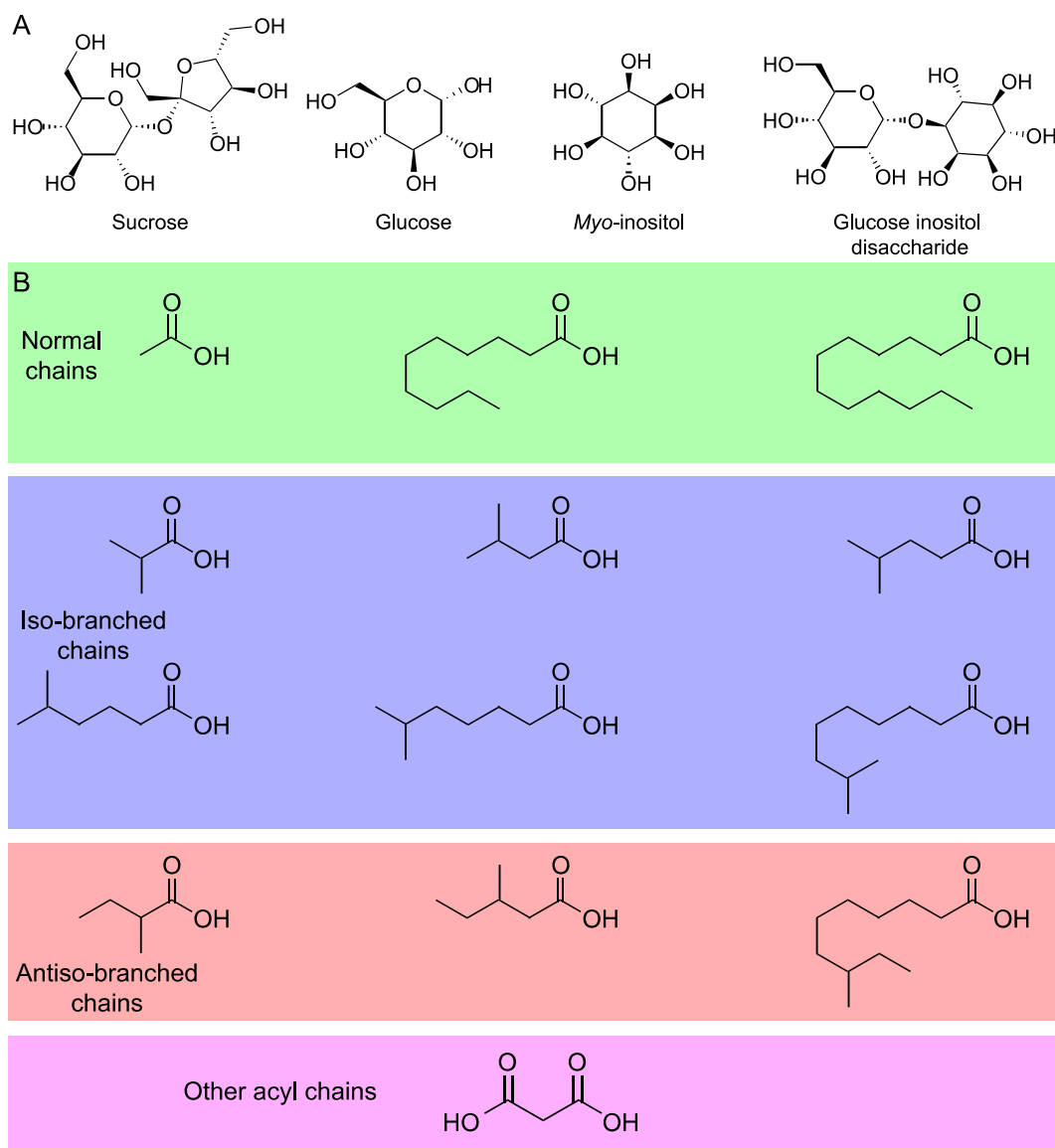
Figure 1.1. Acylsugar diversity in the Solanaceae

(A) The component sugars used in acylsugar biosynthesis in the Solanaceae with sucrose being the most prominent. There are examples of glucose, *myo*-inositol, and glucose-inositol-derived acylsugars. (B) The complement of acyl chains incorporated into acylsugars in different Solanaceae species. The majority of acyl chains fit into three categories: Normal chain, iso-branched chains, and anteiso-branched chains. There are exceptions in the case of malonate chains in *Petunia axillaris* and other *Petunia* spp.

Acylsugar diversity beyond the Solanaceae

The Solanaceae family is where most acylsugar characterized has occurred, although there are documented examples of acylsugars in other plant families. One such example is in the Martyniaceae family containing several metabolites pertaining to glucose-centric compounds. Both *Ibicella lutea* and *Proboscidea louisiana* have glucose molecules containing hydroxylated fatty acids of varying lengths ranging from 18 to 22 carbons in length conjugated to glucose through an ether linkage to a hydroxyl group near the middle of the acyl chain (Asai et al., 2010) (Figure 1.2). Several of the acylsugars contain acetyl groups at the R6 position of glucose as well (Asai et al., 2010) (Figure 1.2). This pattern of acylation with hydroxylated fatty acid is distinct from acylsugars in the Solanaceae, in which structures of isolated compounds reveal acyl chains conjugated through ester linkages to the various sugar hydroxyls.
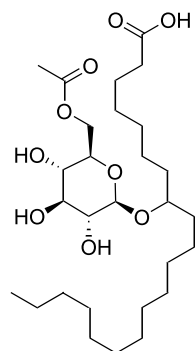
P. louisiana
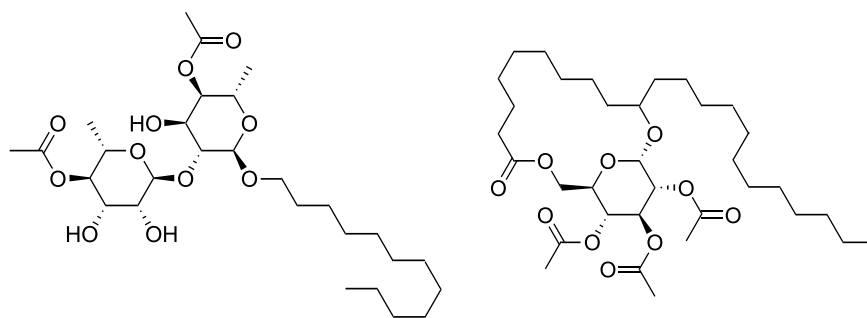
E. pelargoniflorum   *C. glomeratum*

Figure 1.2. Examples of acylsugars from non-Solanaceae plant families

Multiple examples of acylsugars from different plant families including the Martyniaceae (*Proboscidea louisiana*), Rosaceae (*Cerasus yedoensis*), Geraniaceae (*Erodium pelargoniflorum*), and Caryophyllaceae (*Cerastium glomeratum*).

Another example of acylsugars is in the Geraniaceae family in *Geranium carolinianum* and *Erodium pelargoniflorum*. *G. carolinianum* produces an interesting mixture of disaccharides (rhamnose-1,2-glucose disaccharide) with multiple substitutions of acyl chains at different positions of both sugars (Asai et al., 2011) (Figure 1.2). Further, the R1 position of glucose is conjugated to an octanol moiety. Other positions contain acyl chains similar to those encountered in Solanaceae acylsugars, such as isobutyrate or 2-methylbutyrate at the R4 position of rhamnose and isobutyrate at the R6 position of glucose (Asai et al., 2011) . *E. pelargoniflorum* possesses acylsugars as well, with a different disaccharide (rhamnose-1,2-fucose disaccharide) (Sakai et al., 2013). The variation from those acylsugars in *G. carolinianum* is present in the acyl chains, specifically, with a dodecanol moiety at the R1 position of fucose. Additionally, the fucose molecule is acetylated at the R5 position, while the rhamnose is substituted at the R4 and R5 positions by different acyl moieties (Sakai et al., 2013).

Acylsugar-like metabolites have been detected in the Caryophyllaceae family in *Cerastium glomeratum* and *Silene gallica*. *C. glomeratum* possesses glucose-centric compounds esterified to a docosanoyl moiety at the R6 position of glucose (Asai et al., 2012) (Figure 1.2). The same acyl group is conjugated through an ether linkage between a C9-C11 hydroxyl group to the R1 anomeric carbon of glucose forming a cyclic molecule (Asai et al., 2012) (Figure 1.2). The sugar is often substituted at the R2-R4 positions as well with acetyl groups, and some O-methylation (Asai et al., 2012). *S. gallica* possesses a glucose molecule esterified to octadecanoic acid at the R2 position of glucose, with the ether linkage at the R1 anomeric carbon, completing the ring. The ether linkage is through a hydroxyl group at the 12-13th carbon of octadecanoic acid. Additionally, glucose is further acylated at different positions with acetyl, malonyl, or a malonyl methyl ester groups.

One last published example is in *Brassica rapa* roots where acylsucroses with multiple iC5 chains conjugated accumulate (Wu et al., 2013). The acylation pattern is different from those in the Solanaceae, but they appear similar to acylsucroses produced in the nightshade family.

Interestingly, the presence of acylsugar-like compounds in distinct plant families suggesting two possible general mechanisms. The first is a single ancient origin of acylsugar biosynthesis, but this seems unlikely based on the sparse representation of acylsugars across the plant kingdom. What seems more probable is independent origins of these acylsugar-like compounds through relatively simple reactions.

Acylsugar biosynthesis in the *Solanum*

The enzymes that catalyze acylsugar biosynthesis fall into the BAHD family of acyltransferases. This family of enzymes is named after the first four enzymes characterized in this family – <u>B</u>EAT, <u>A</u>HCT, <u>H</u>CBT, and <u>D</u>AT – that catalyze the acylation of a range of structurally diverse plant metabolites. This enzyme family catalyzes a reaction using dual substrates, an acyl acceptor and acyl donor. Some of these acceptors lead to products including volatile floral compounds in *Clarkia breweri* (Dudareva et al., 1998), monoterpenoid indole alkaloids in *Catharanthus roseus* (St-Pierre et al., 1998), hydroxycinnamoyl compounds in phenylpropanoid metabolism (Fujiwara et al., 1998), and epicuticular waxes (Negruk et al., 1996). The acyl donor is an activated acyl-CoA molecule that supplies the acyl group to be conjugated to the acyl acceptor, releasing free coenzyme A. The characterized BAHD acyltransferases that catalyze acylsugar biosynthesis across the family are located in clade III of this enzyme family (Moghe et al., 2017). This clade traditionally uses alcohol substrates, while many of the enzymes use acetyl-CoA as the major acyl donor (D'Auria, 2006).

Most genetics and biochemistry work related to acylsugar biosynthesis has been performed in *Solanum lycopersicum* and its wild relatives. *S. lycopersicum* produces a mixture of acylsucroses acylated either three or four times with acyl chains ranging from C2 to C12 in length (Ghosh et al., 2014). Acyl chains are present on both the five- (furanose) and six-membered (pyranose) rings, making these acylsugars, 'F-type' acylsucroses. Genetic resources resulting from single or multiple *Solanum pennellii* chromosomal introgressions into a cultivated tomato background are called introgression lines (ILs) or backcrossed introgression lines (BILs) respectively. Differences in acylsugar biosynthesis in *S. lycopersicum* and *S. pennellii* result in detectable phenotypes in several of the introgression lines that were used to identify acylsugar biosynthetic genes.

Sl-ASAT1 catalyzes the first step in acylsugar biosynthesis in cultivated tomato (Fan et al., 2016)(Figure 1.3). This enzyme uses iso-C5-CoA (iC5) and sucrose to catalyze the formation of mono-acyl sucrose, demonstrated using *in vitro* and *in vivo* approaches (Fan et al., 2016) (Figure 1.3). This mono-acyl sucrose product is acylated at the R4 position of the pyranose ring of sucrose as demonstrated by NMR. Mono-acyl sucrose is converted to di-acyl sucrose by Sl-ASAT2 using anteiso-C5-CoA (aiC5) or straight-chain C12 (nC12) resulting in an acylation at the R3 position of the six-membered ring (Fan et al., 2016)(Figure 1.3). That product is then a substrate for Sl-ASAT3, which utilizes iC5-CoA to conjugate an acyl chain to the R3' position (furanose ring) (Schilmiller et al., 2015)(Figure 1.3). Finally, Sl-ASAT4 adds an acetyl group to the R2 position of the triacylsucrose (pyranose ring) resulting in a tetra-acylated sucrose (Schilmiller et al., 2012)(Figure 1.3). Variation in substrate specificity and acyl-CoA availability are two factors that shape the acylsugars that accumulate in cultivated tomato (Ning et al., 2015; Fan et al., 2016).

Comparisons between cultivated tomato and *S. pennellii* reveal differences in ASAT substrate specificity that have emerged in the 3-5 million years since the last common ancestor. Biochemical analysis revealed that a small number of amino acid changes are responsible for altered substrate specificities in the Sl-ASAT2 ortholog – SpASAT2 – and the Sl-ASAT3 ortholog – SpASAT3 (Fan et al., 2017). Changes in substrate specificity resulted in reversed roles in the pathway, in which SpASAT2 and SpASAT3 catalyze the **third** and **second** step, respectively (Figure 1.3). SpASAT3 experienced an acylation position change relative to Sl-ASAT3, now conjugating the acyl chain at the R2 position of mono-acyl sucrose, while SpASAT2 still acylates at the R3 position. The major consequence of this flipped pathway is the production of acylsucroses with acyl chains exclusively on the six-membered ring to produce 'P-type' acylsucroses (Fan et al., 2017). This contrasts with F-type acylsucroses in cultivated tomato, which are acylated on the R3' position of sucrose. In *S. pennellii*, SpASAT3 acylates at the R2 position, in contrast to the acetylation catalyzed by Sl-ASAT4 (Figure 1.3).

A separate, hypothetical biosynthetic pathway describing acylglucose biosynthesis was proposed for *S. pennellii* in a series of papers. Through the action of multiple UDP-glycosyltransferases, different lengths of acyl chains were conjugated to glucose to form different length 1-*O*-acyl-β-D-glucose molecules used for further biosynthesis (Kuai et al., 1997) (Figure 1.3). In addition to purified enzymes, trichome extracts from *S. pennellii* were used to generate these 1-*O*-acyl-β-glucose molecules. The hypothesized acylglucose biosynthetic pathway was demonstrated to utilize two 1-*O*-acyl-β-D-glucose molecules in a disproportionation reaction that generates a single diacylglucose – R1 and R2 acylated –

molecule and free glucose (Li et al., 1999)(Figure 1.3). The acyl chain position on these *in vitro* products are not consistent with those found on *in planta* structures This reaction was proposed to be catalyzed through the action of a serine carboxypeptidase-like enzyme (Figure 1.3). This mechanism required further steps that they hypothesized could be catalyzed by other enzymes.

The biosynthesis of coenzyme A precursors of these pathways play a role in determining acylsugar diversity in addition to sugar-conjugating enzymes. For example, allelic variation in a chain elongation pathway in *S. pennellii* led to discovery of a modified enzyme recruited from the leucine biosynthetic pathway. This neofunctionalized enzyme is responsible for producing isovaleryl coenzyme A (isoC5-CoA), an acyl donor substrate for ASATs (Ning et al., 2015). What sets this isopropylmalate synthase apart is not the reaction it catalyzes – carboxylation of its substrate, the same one carbon elongation reaction as the canonical microbial and plant amino acid biosynthetic enzymes – but instead the unique regulation of enzymatic activity and gene expression of isopropylmalate synthase-like 3 (IPMS3). IPMS3 lacks the inhibitory Leu-binding allosteric C-terminal domain, resulting in a feedback insensitive enzyme. Second, unlike the broad expression consistent with amino acid biosynthetic enzymes, IPMS3 expression is limited to Type I/IV glandular trichome tip cells. In *S. pennellii* LA0716, IPMS3 is further truncated at the C-terminus, resulting in an enzyme with no detectable *in vitro* activity (Fig. 2b). This results in accumulation of acylsugars containing an increased abundance of isobutyryl (isoC4) chains in this and other *S. pennellii* accessions homozygous for this non-functional allele (Ning et al., 2015).
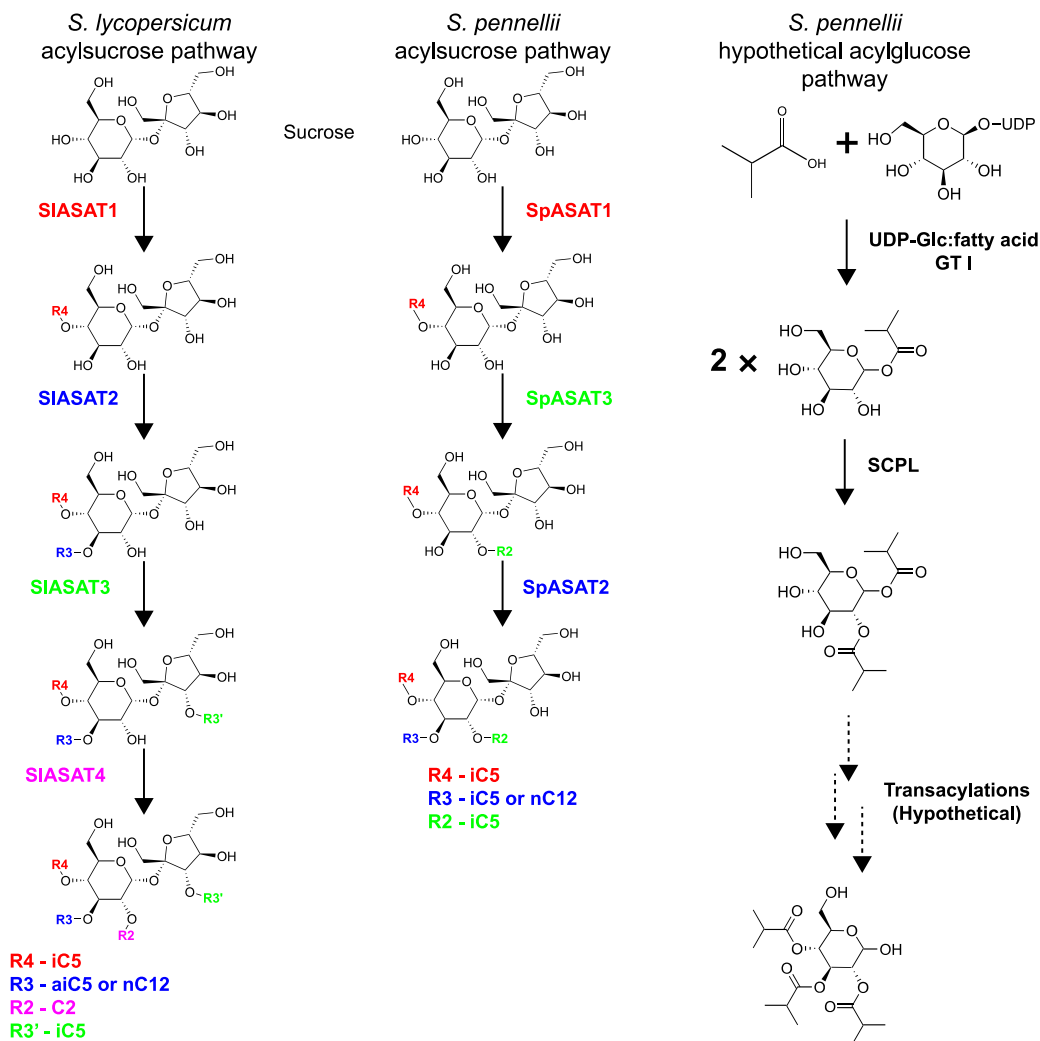
Figure 1.3. Current understanding of acylsugar biosynthesis in the *Solanum*.

Acylsucrose biosynthetic pathways in cultivated tomato (left) and *S. pennellii* (center). A hypothetical acylglucose biosynthetic pathway is also presented (right). Acylsugar biosynthesis has variable acylation patterns and enzymes catalyzing the acylations. The color corresponds to orthologous enzymes and the position of acylation.

Transcriptomic and proteomic analysis of cultivated tomato and a wild relative, *Solanum habrochaites*, revealed a number of other characteristics that influence acylsugar metabolism in glandular trichomes (Balcke et al., 2017). It appears that glandular trichomes obtain metabolites from other tissues such as the underlying leaf. Labeled $^{13}CO_2$ suggests that trichomes are a sink for sucrose, which is transported from the leaf – incorporation of heavy isotope labels into trichome metabolites trails leaf metabolites (Balcke et al., 2017). Expression of amino acid metabolism genes is enriched in the trichomes of both species, but to a greater degree in *S. habrochaites* (Balcke et al., 2017). Long fatty acid modification and lipid degradation genes were also enriched. Both amino acid and fatty acid metabolism enzymes can have a direct effect on the incorporation of these different acyl chains into acylsugars. As more acylsugar biosynthetic genes were uncovered and pieced together to assemble biosynthetic pathways (Schilmiller et al., 2012; Ning et al., 2015; Schilmiller et al., 2015; Fan et al., 2016; Fan et al., 2017), it became possible to study how various evolutionary processes contributed to the evolution of a metabolic network.

Acylsugar biosynthesis and evolution in the broader Solanaceae family

There is variation in the acylsucrose biosynthetic pathways represented in some of the earliest diverging species in the Solanaceae family. *P. axillaris* and *S. sinuata* – which represent some of the earliest diverging lineages in the Solanaceae family – produce acylated sucrose also present in other glandular trichome-bearing species in the nightshade family. There are both similarities and differences between acylsucrose biosynthesis in these two species. For both species, the first step is catalyzed by an enzyme that is absent from *S. lycopersicum* and its wild relatives. Both *P. axillaris* and *S. sinuata* catalyze the first three steps of acylsucrose biosynthesis using orthologous enzymes that acylate at the R2, R4, and R3 positions of sucrose, respectively.

The biosynthetic pathways diverge at this point: PaASAT4 catalyzes the acylation of the R6 position of sucrose, while the R1', R3', and/or R6' positions of sucrose are further acylated to generate the products accumulating in *S. sinuata*.

The identification and characterization of these biosynthetic genes in some of the earliest diverging lineages in the Solanaceae facilitates analysis of the broader evolutionary patterns involved in acylsugar biosynthesis. Phylogenetic analysis and comparison from several species in the Solanaceae to several Convolvulaceae spp. suggest that the emergence of the acylsugar biosynthetic genes between 30-80 ma (mega-annum)(Moghe et al., 2017). This analysis revealed that these biosynthetic genes most likely were derived from alkaloid biosynthetic genes (Moghe et al., 2017). This hypothesis is supported by two enzymes – involved in monoterpenoid indole alkaloid biosynthesis in *C. roseus* – that are closely related to ASATs. Interestingly, it appears that ASATs have shifted in their roles in the respective pathways. The second and third steps in *P. axillaris* and *S. sinuata* are catalyzed by enzymes that are orthologs to those that catalyze the first and second steps in *S. lycopersicum* acylsucrose biosynthesis (Figure 1.4). Based on phylogenetics combined with enzymology, it appears this shift in substrate specificity emerged sometime between the last common ancestor between *H. niger* and the *Solanum* clade and that of *S. nigrum* and *S. lycopersicum* (Figure 1.4).
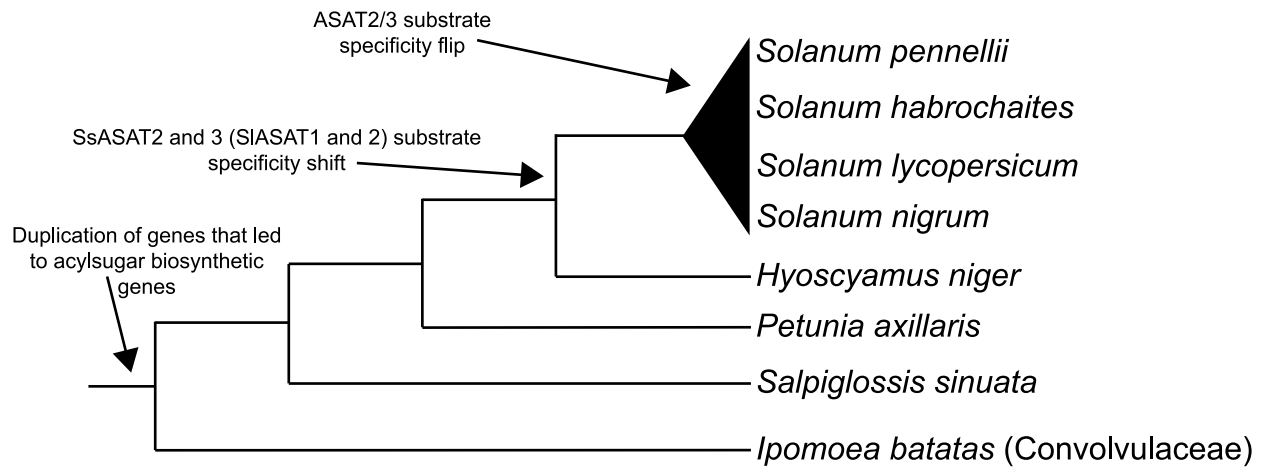
Figure 1.4. Evolution of acylsugar biosynthesis in the Solanaceae.

Evolutionary events that occurred in the Solanaceae family. The duplication event that led to acylsugar biosynthetic genes appeared to have occurred before the divergence of the Solanaceae and Convolvulaceae family. The ancestral ASAT2 and ASAT3 shifted enzymatic activities between the divergence of *H. niger* and the Solanum clade and the last common ancestor between tomato and *Solanum nigrum*. Additionally, the modern (or tomato equivalents) of ASAT2 and ASAT3 flipped substrate specificity in *S. pennellii* and *S. habrochaites*, presumably after the divergence of these lineages from tomato.

Ecological functions of acylsugars

Acylsugars play a role in plant defense – whether it is dependent the stickiness or perhaps their amphipathic nature is not known. For example, acylsugars have been shown to deter or repel *Liriomyza trifolii*, *Macrosiphum euphorbiae*, and *Spodoptera exigua* in different Solanaceae species (Goffreda et al., 1989; Hawthorne et al., 1992; Juvik et al., 1994; Puterka et al., 2003). Additional work demonstrated that acylsugars can adhere to arthropod cuticles, which ultimately immobilizes or suffocates them (Puterka et al., 2003; Wagner et al., 2004). Other studies demonstrated that acylsugar mixtures can delay the development of *Manduca sexta* larvae (Van Dam and Hare, 1998) and serve as a feeding deterrent for *Epitrix hirtipennis* and *Trichobaris compacta* (Hare, 2005). It appears that certain specialized acylsugars have a role to deter, repel, or kill insects that depends on the acylsugars and the insect.

Another interesting set of studies surrounds the wild tobacco, *Nicotiana attenuata*. *N. attenuata* trichomes, and the acylsucroses produced in them, serve as 'dangerous lollipops' that tag neonate *Manduca sexta* larvae when eaten (Weinhold and Baldwin, 2011). This acylsugar consumption releases volatile branched chain fatty acids that are detected by ground hunting ants, *Pogonomyrmex rugosus*. These ants can detect the hydrolyzed branched chain fatty acids present in the frass of the larvae and use that as a guide to find the *M. sexta* larvae. Further analysis revealed that *N. attenuata* acylsugars can inhibit the growth of *M. sexta*, elucidated through both removal of acylsugars from the leaves, and addition of acylsugars to artificial diets (Luu et al., 2017). Additionally, acylsugars – or lack thereof – correlated with resistance to different fungal pathogens in wild tobacco, while directly reducing the germination rate of fungal spores (Luu et al., 2017).

There are some indications that acylsugar composition influences their efficacy. Puterka et al. indicate that both sugar moiety and acyl chain length can affect acylsugar efficacy (Puterka et al., 2003). The study found that synthetically-generated sucrose octanoate had the highest insecticidal activity compared to xylitol and sorbitol (>60% single acyl chain). However, xylitol decanoate appears to have similar efficacy on pear psylla mortality as sucrose octanoate. This suggests that characterization of acylsugar diversity could be valuable in a few ways: First, understanding the variety of acyl chains and sugars incorporated into acylsugars can hone our knowledge of the possible makeup of acylsugars. Second, it allows us to test the efficacies of different acylsugars on insects and determine what factors are important in acylsugar function.

Aims of this research

The goal of this research was to improve our understanding of how sugar core diversity is generated in acylsugars in the Solanaceae family. The first aim involved characterization acylglucose biosynthesis in *S. pennellii*, a wild relative of tomato. This aim involved the characterization of an invertase co-opted from general metabolism to generate acylglucoses *in planta*. A combination of molecular biology, genetics, biochemistry, and analytical chemistry was used to elucidate the last step in acylglucose biosynthesis. The second aim involved characterization of part of the acylinositol biosynthetic pathway in the Andean fruit crop, *S. quitoense*. This aim utilized genetics and biochemistry to investigate the underpinnings of acylinositol biosynthesis.

REFERENCES

REFERENCES

Ahkami A, Johnson SR, Srividya N, Lange BM (2015) Multiple levels of regulation determine monoterpenoid essential oil compositional variation in the mint family. Mol Plant **8**: 188–91.

Asai T, Fujimoto Y (2011) 2-Acety-1-(3-glycosyloxyoctadecanoyl)glycerol and dammarane triterpenes in the exudates from glandular trichome-like secretory organs on the stipules and leaves of *Cerasus yedoensis*. Phytochem Lett **4**: 38–42.

Asai T, Hara N, Fujimoto Y (2010) Fatty acid derivatives and dammarane triterpenes from the glandular trichome exudates of *Ibicella lutea* and *Proboscidea louisiana*. Phytochemistry **71**: 877–894.

Asai T, Nakamura Y, Hirayama Y, Ohyama K, Fujimoto Y (2012) Cyclic glycolipids from glandular trichome exudates of *Cerastium glomeratum*. Phytochemistry **82**: 149–157.

Asai T, Sakai T, Ohyama K, Fujimoto Y (2011) *n*-Octyl α-L-rhamnopyranosyl-(1→2)-β-D-glucopyranoside derivatives from the glandular trichome exudate of *Geranium carolinianum*. Chem Pharm Bull (Tokyo) **59**: 747–52.

Balcke GU, Bennewitz S, Bergau N, Athmer B, Henning A, Majovsky P, Jiménez-Gómez JM, Hoehenwarter W, Tissier A (2017) Multi-omics of tomato glandular trichomes reveals distinct features of central carbon metabolism supporting high productivity of specialized metabolites. Plant Cell **29**: 960–983.

Burke B, Goldsby G, Brian Mudd J (1987) Polar epicuticular lipids of *Lycopersicon pennellii*. Phytochemistry **26**: 2567–2571.

Chortyk OT, Kays SJ, Teng Q (1997) Characterization of insecticidal sugar esters of *Petunia*. J Agric Food Chem **45**: 270–275.

Courdavault V, Papon N, Clastre M, Giglioli-Guivarc'h N, St-Pierre B, Burlat V (2014) A look inside an alkaloid multisite plant: the Catharanthus logistics. Curr Opin Plant Biol **19**: 43–50.

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol **9**: 331–340.

Van Dam NM, Hare JD (1998) Biological activity of *Datura wrightii* glandular trichome exudate against *Manduca sexta* larvae. J Chem Ecol **24**: 1529–1549.

Dudareva N, D'Auria JC, Nam KH, Raguso RA, Pichersky E (1998) Acetyl-CoA:benzylalcohol acetyltransferase – an enzyme involved in floral scent production in *Clarkia breweri*. Plant J **14**: 297–304.

Fan P, Miller AM, Liu X, Jones AD, Last RL (2017) Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. Nat Commun **8**: 1–13.

Fan P, Miller AM, Schilmiller AL, Liu X, Ofner I, Jones AD, Zamir D, Last RL (2016) In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. Proc Natl Acad Sci U S A **113**: E239-48.

Fridman E, Wang J, Iijima Y, Froehlich JE, Gang DR, Ohlrogge J, Pichersky E (2005) Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. Plant Cell **17**: 1252–67.

Fujiwara H, Tanaka Y, Fukui Y, Ashikari T, Yamaguchi M, Kusumi T (1998) Purification and characterization of anthocyanin 3-aromatic acyltransferase from *Perilla frutescens*. Plant Sci **137**: 87–94.

Ghosh B, Westbrook TC, Jones AD (2014) 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.

Goffreda JC, Mutschler MA, Avé DA, Tingey WM, Steffens JC (1989) Aphid deterrence by glucose esters in glandular trichome exudate of the wild tomato, *Lycopersicon pennellii*. J Chem Ecol **15**: 2135–2147.

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol **57**: 303–333.

Hare JD (2005) Biological activity of acyl glucose esters from *Datura wrightii* glandular trichomes against three native insect herbivores. J Chem Ecol **31**: 1475–1491.

Hawthorne DJ, Shapiro JA, Tingey WM, Mutschler MA (1992) Trichome-borne and artificially applied acylsugars of wild tomato deter feeding and oviposition of the leafminer *Liriomyza trifolii*. Entomol Exp Appl **65**: 65–73.

Herrera-Salgado Y, Garduño-Ramírez ML, Vázquez L, Rios MY, Alvarez L (2005) *Myo*-inositol-derived glycolipids with anti-inflammatory activity from *Solanum lanceolatum*. J Nat Prod **68**: 1031–1036.

Hurney SM (2018) Strategies for profiling and discovery of acylsugars. Michigan State University.

Juvik JA, Shapiro JA, Young TE, Mutschler MA (1994) Acylglucoses from wild tomatoes alter behavior and reduce growth and survival of *Helicoverpa zea* and *Spodoptera exigua*. J Econ Ent **87**: 482–492.

King RR, Calhoun LA (1988) 6 2,3-Di-*O*- and 1,2,3-tri-*O*-acylated glucose esters from the glandular trichomes of *Datura metel*. Phytochemistry **27**: 3761–3763.

King RR, Pelletier Y, Singh RP, Calhoun LA (1986) 3,4-Di-*O*-isobutyryl-6-*O*-caprylsucrose: The major component of a novel sucrose ester complex from the type B glandular trichomes of *Solanum berthaultii* Hawkes (Pl 473340). J Chem Soc Chem Commun 1078–1079.

Kuai JP, Ghangas GS, Steffens JC (1997) Regulation of triacylglucose fatty acid composition. Plant Physiol **115**: 1581–1587.

Leong BJ, Lybrand DB, Lou Y-R, Fan P, Schilmiller AL, Last RL (2019) Evolution of metabolic novelty: A trichome-expressed invertase creates specialized metabolic diversity in wild tomato. Sci Adv **5**: eaaw3754.

Li AX, Eannetta N, Ghangas GS, Steffens JC (1999) Glucose polyester biosynthesis. Purification and characterization of a glucose acyltransferase. Plant Physiol **121**: 453–60.

Liu X, Enright M, Barry CS, Jones AD (2017) Profiling, isolation and structure elucidation of specialized acylsucrose metabolites accumulating in trichomes of *Petunia* species. Metabolomics. doi: 10.1007/s11306-017-1224-9.

Liu Y, Jing S-X, Luo S-H, Li S-H (2019) Non-volatile natural products in plant glandular trichomes: chemistry, biological activities and biosynthesis. Nat Prod Rep. doi: 10.1039/C8NP00077H.

Luu VT, Weinhold A, Ullah C, Dressel S, Schoettner M, Gase K, Gaquerel E, Xu S, Baldwin IT (2017) *O*-acyl sugars protect a wild tobacco from both native fungal pathogens and a specialist herbivore. Plant Physiol **174**: 370–386.

Maldonado E, Torres FR, Martínez M, Pérez-Castorena AL (2006) Sucrose esters from the fruits of *Physalis nicandroides* var. attenuata. J Nat Prod **69**: 1511–1513.

Matsuzaki T, Shinozaki Y, Hagimori M, Tobita T, Shigematsu H, Koiwai A (1992) Novel glycerolipids and glycolipids from the surface lipids of *Nicotiana benthamiana*. Biosci Biotech Biochem **56**: 1565–1569.

Matsuzaki T, Shinozaki Y, Suhara S, Ninomiya M, Shigematsu H, Koiwai A (1989) Isolation of glycolipids from the surface lipids of *Nicotiana bigelovii* and their distribution in *Nicotiana* species. Agric Biol Chem **53**: 3079–3082.

Matsuzaki T, Shinozaki Y, Suhara S, Tobita T, Shigematsu H, Koiwai A (1991) Leaf surface glycolipids from *Nicotiana acuminata* and *Nicotiana pauciflora*. Agric Biol Chem **55**: 1417–1419.

Moghe GD, Leong BJ, Hurney SM, Jones AD, Last RL (2017) Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. Elife **6**: 1–33.

Negruk V, Yang P, Subramanian M, McNevin JP, Lemieux B (1996) Molecular cloning and characterization of the CER2 gene of *Arabidopsis thaliana*. Plant J **9**: 137–145.

Ning J, Moghe GD, Leong B, Kim J, Ofner I, Wang Z, Adams C, Jones AD, Zamir D, Last RL (2015) A feedback-insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. Plant Physiol **169**: 1821–35.

Pichersky E, Lewinsohn E (2011) Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol **62**: 549–566.

Puterka GJ, Farone W, Palmer T, Barrington A (2003) Structure-function relationships affecting the insecticidal and miticidal activity of sugar esters. J Econ Entomol **96**: 636–644.

Sakai T, Tanemura Y, Itoh S, Fujimoto Y (2013) Dodecyl α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranoside derivatives from the glandular trichome exudate of *Erodium pelargoniflorum*. Chem Biodivers **10**: 1099–1108.

Schilmiller AL, Charbonneau AL, Last RL (2012) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. Proc Natl Acad Sci **109**: 16377–16382.

Schilmiller AL, Moghe GD, Fan P, Ghosh B, Ning J, Jones AD, Last RL (2015) Functionally divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. Plant Cell **27**: 1002–1017.

Shapiro JA, Steffens JC, Mutschler MA (1994) Acylsugars of the wild tomato *Lycopersicon pennellii* in relation to geographic distribution of the species. Biochem Syst Ecol **22**: 545–561.

Shinozaki Y, Matsuzaki T, Suhara S, Tobita T, Shigematsu H, Koiwai A (1991) New types of glycolipids from the surface lipids of *Nicotiana umbratica*. Agric Biol Chem **55**: 751–756.

St-Pierre B, Laflamme P, Alarco A-M, D V, Luca E (1998) The terminal *O*-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer. Plant J **14**: 703–713.

Wagner GJ, Wang E, Shepherd RW (2004) New approaches for studying and exploiting an old protuberance, the plant trichome. Ann Bot **93**: 3–11.

Weinhold A, Baldwin IT (2011) Trichome-derived *O*-acyl sugars are a first meal for caterpillars that tags them for predation. Proc Natl Acad Sci **108**: 7855–7859.

Wu Q, Cho J-G, Lee D-S, Lee D-Y, Song N-Y, Kim Y-C, Lee K-T, Chung H-G, Choi M-S, Jeong T-S, et al (2013) Carbohydrate derivatives from the roots of *Brassica rapa* ssp. *campestris* and their effects on ROS production and glutamate-induced cell death in HT-22 cells. Carbohydr Res **372**: 9–14.

Zhang CR, Khan W, Bakht J, Nair MG (2016) New anti-inflammatory sucrose esters in the natural sticky coating of tomatillo (*Physalis philadelphica*), an important culinary fruit. Food Chem **196**: 726–732.

Chapter 2. Promiscuity, impersonation, and accommodation: evolution of plant specialized metabolism

Abstract

Specialized metabolic enzymes and metabolite diversity evolve through a variety of mechanisms including promiscuity, changes in substrate specificity, modifications of gene expression and gene duplication. For example, gene duplication and substrate binding site changes led to the evolution of the glucosinolate biosynthetic enzyme, AtIPMDH1, from a Leu biosynthetic enzyme. BAHD acyltransferases illustrate how enzymatic promiscuity leads to metabolite diversity. The examples 4-coumarate:CoA ligase and aromatic acid methyltransferases illustrate how promiscuity can potentiate the evolution of these specialized metabolic enzymes.

Introduction

Plant specialized metabolites are lineage-specific compounds, many of which are involved in ecological interactions, such as herbivore defense or pollinator attraction (Knudsen et al., 1993; Mithöfer and Boland, 2012). The number of specialized metabolites produced across all plant species is estimated to be in the hundreds of thousands (Dixon and Strack, 2003). Specialized metabolic enzymes tend to have lower catalytic efficiency (Milo and Last, 2012) and greater substrate promiscuity (Weng et al., 2012) than primary metabolic enzymes. This review explores factors involved in enzyme evolution and discusses how these result in metabolite diversity.

We focus on mechanisms that play roles in "potentiation" (Figure 2.1A); metabolic examples of what was more generally described by Blount et al. (Blount et al., 2012) as factors that allow for the realization of a new trait. In recent years, enzymatic promiscuity – the ability of an enzyme to catalyze reaction(s) in addition to its primary reaction – has been documented to "potentiate" the evolution of new specialized metabolic activities (Figure 2.1A) (Blount et al., 2012). Substrate promiscuity is documented to play a central role in the evolution of specialized

27

metabolic enzymes (Figure 2.1B) (Weng et al., 2012). Changes in substrate specificity also can result in emergence of novel activities and chemical diversity. Such a shift in substrate specificity can change the primary substrate of an enzyme from an intermediate in an existing biosynthetic pathway to a new substrate, which – in turn – can potentiate novel enzymatic reactions (Figure 2.1C). Gene duplication and divergence in gene expression patterns or enzyme activities also potentiate the evolution of specialized metabolic enzymes (Figure 2.1D) (Moghe and Last, 2015). We highlight examples from the past five years in which promiscuity, changes in substrate specificity, gene duplication, and changes in gene expression were shown to play prominent roles in evolution of specialized metabolic enzymes and generation of chemical diversity. These examples illustrate the power of structural analysis – especially in a comparative evolutionary context – to reveal constraints and opportunities to facilitate the modification or engineering of these enzymes.

Body

Gene duplication and changes in substrate specificity in the evolution of a glucosinolate biosynthetic enzyme

Glucosinolates are a group of structurally diverse, amino-acid derived plant specialized metabolites that mediate interactions between crucifers and insects or pathogens (Halkier and Gershenzon, 2006). The biosynthesis of methionine-derived glucosinolates involves a repeated three step elongation process similar to Leu biosynthesis: condensation with acetyl-CoA, isomerization, and oxidative decarboxylation to successively add one carbon units to the aliphatic side chain (Sønderby et al., 2010). The glucosinolate oxidative decarboxylation step is catalyzed by the *A. thaliana* isopropylmalate dehydrogenase (AtIPMDH1), while two other *A. thaliana* IPMDH enzymes catalyze the same reaction in Leu biosynthesis (He et al., 2009; He et al., 2011a). The Leu biosynthetic substrate (3-isopropylmalate) and glucosinolate substrate (3-

(2'-methylthio)-ethylmalate), have the same carboxyl and hydroxyl group configuration but differ in side chain length and composition, suggesting similarities in the binding dynamics between the enzymes and substrates in the two pathways (Figure 2.2A, side chains are in color).

The AtIPMDH2 Leu biosynthetic enzyme crystal structure with 3-isopropylmalate (3-IPM) revealed several binding interactions with the polar portion of the substrate (Lee et al., 2016). The structure revealed that residues interacting with the polar groups of 3-IPM are conserved between all IPMDH enzymes (Figure 2.2B) (He et al., 2011b). This, combined with the similarity of the polar groups of 3-IPM and (3-(2'-methylthio)-ethylmalate, suggested that recognition of the side chain is responsible for substrate discrimination. There are no specific substrate-enzyme interactions between the 3-IPM aliphatic isopropyl side chain and the residues
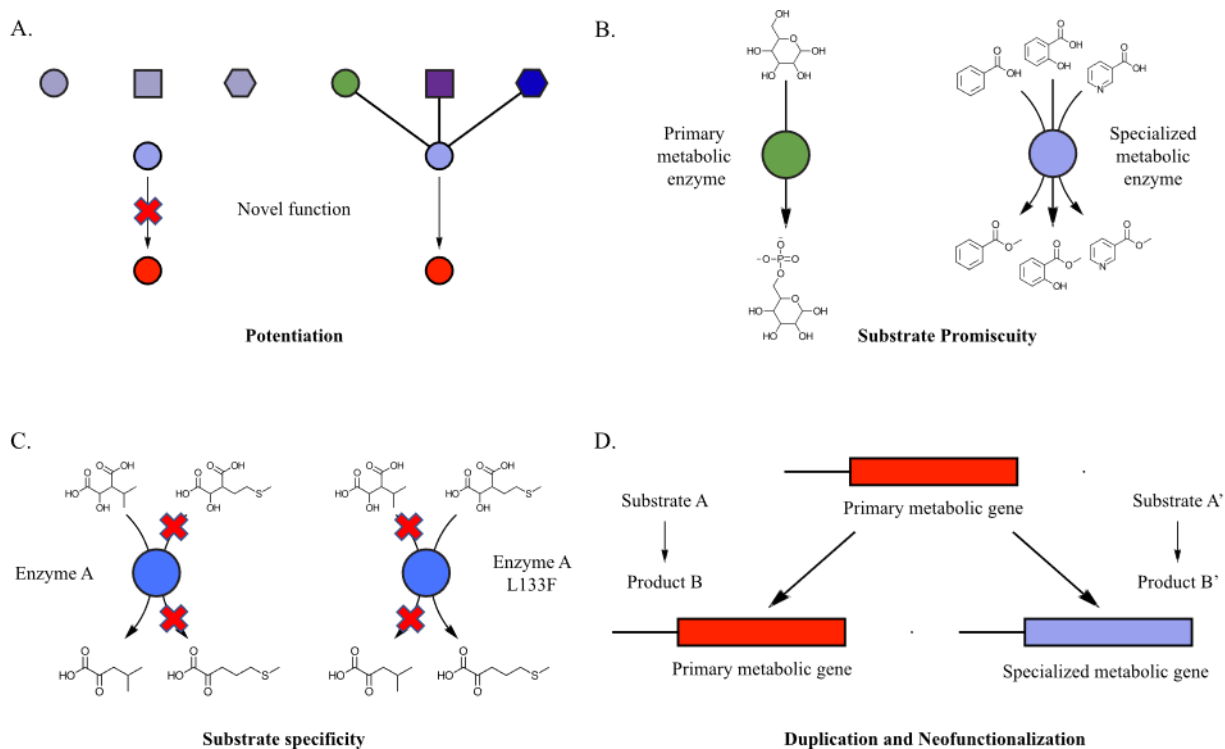
Figure 2.1. Mechanisms leading to evolution of chemical diversity and enzymatic novelty.

(A) Potentiation – The different shapes and colors represent factors involved in the evolution of a novel function or enzymatic activity. These factors enable the realization of novel functions or enzymatic activities. In this example, the green circle represents suitable localization of gene expression, the purple square represents substrate availability in the tissue of interest, and the blue hexagon represents the ability of the enzyme to utilize the substrate . (B) Substrate promiscuity – primary metabolic enzymes typically catalyze a specific reaction, while specialized metabolic enzymes tend to be promiscuous and catalyze reactions using multiple substrates. (C) Substrate specificity – specific amino acid changes result in alteration of enzyme substrate specificity, resulting in a new enzymatic activity or function. (D) Gene duplication and neofunctionalization –a primary metabolic gene is duplicated, facilitating diversification of one isoform into a specialized metabolic function.

in the largely hydrophobic pocket in the active site (Figure 2.2B) (He et al., 2011b; Lee et al., 2016). Thus, the differences between the glucosinolate biosynthetic enzyme AtIPMDH1 and Leu IPMDH enzymes presumably are responsible for the difference in substrate specificity.

Sequence alignments of Leu IPMDH enzymes with AtIPMDH1 revealed a key feature that affects the ability of the enzyme to discriminate between Leu and glucosinolate substrates. AtIPMDH1 carries a Leu:Phe change at a site that is invariant in the Leu enzymes (He et al., 2009; He et al., 2011b). This residue is in the hydrophobic pocket near the isopropyl side chain of the substrate. Reciprocal substitutions of residues at this site led to a decrease in the *in vitro* catalytic efficiency with the native substrate – e.g. 3-IPM for AtIPMDH2 and AtIPMDH3 and 3-(2'-methylthio)-ethylmalate for AtIPMDH1– and an increase with the non-native substrate (He et al., 2011b). Because the chemical structures of 3-(2'-methylthio)-ethylmalate and 3-IPM differ only by the length and structure of the side chain, this result demonstrates that the substitution of Leu by Phe is sufficient to facilitate the accommodation of the 3-(2'-methylthio)-ethylmalate side chain in the enzyme (Lee et al., 2016).

This example illustrates many themes found throughout the evolution of specialized metabolic enzymes (Moghe and Last, 2015). A model for the emergence of the glucosinolate IPMDH can be constructed (He et al., 2011b), starting with a simple gene duplication that led to the opportunity for 'recruitment' of the highly conserved Leu biosynthetic enzyme, IPMDH. This was followed by the L137F change in the amino acid sequence, leading to alteration of substrate specificity. In this model, gene duplication (Figure 2.1D) was critical for the divergence in structure and function of AtIPMDH1, allowing maintenance of the Leu pathway, while facilitating the incorporation of IPMDH activity into the glucosinolate pathway. The residue
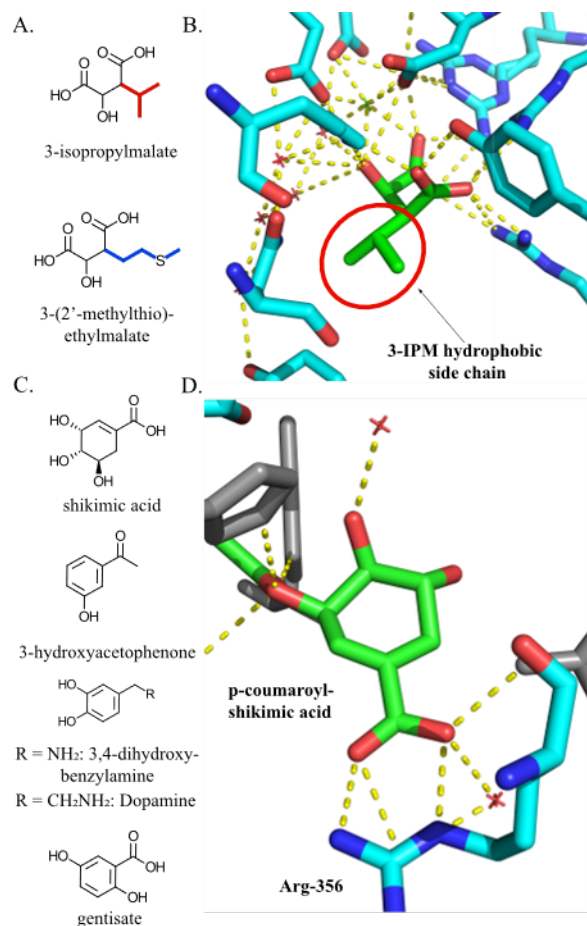
Figure 2.2. Structural analysis of AtIPMDH2 and AtHCT

(A) Structure of substrates for IPMDH enzyme variants in Arabidopsis ─ 3-IPM is the leucine biosynthetic intermediate, and 3-(2'-methylthio)-ethylmalate participates in aliphatic glucosinolate biosynthesis. (B) Interactions of AtIPMDH2 active site residues with 3-IPM demonstrate extensive interactions with the polar moieties, while not specifically interacting with the 3-IPM hydrophobic side chain. 3-IPM structural backbone is green. Active site residues are cyan. Blue represents nitrogen atoms, and red represents oxygen atoms. Green and red crosses are magnesium and water respectively [PDB ID: 5J32]. (C) Substrates used for AtHCT analysis: Shikimic acid is the native substrate; 3-hydroxyacetophenone is a neutral non-native substrate of AtHCT; 3,4-dihydroxybenzylamine and dopamine are substrates that are positively charged under physiological pH. AtHCT has greater *in vitro* activity with gentisate, a non-native substrate, than with shikimate. D) Three-dimensional representation of Arg-356 interaction with the carboxyl group of *p*-coumaroyl-shikimic acid substrate in the active site of AtHCT. *p*-coumaroyl shikimic acid is colored green. Active site residues are colored cyan and red crosses represent water [PDB ID: 5KJU].

change (Figure 2.1C) facilitated the alteration of substrate specificity resulting in incorporation of the IPMDH activity into the glucosinolate pathway and increased chemical diversity through the production of elongated glucosinolates.

Acyltransferase promiscuity modulates metabolite diversity

Acylsucrose acyltransferases (ASATs) are BAHD (BEAT, AHCT, HCT, and DAT) (St-Pierre and De Luca, 2000) acyltransferases involved in biosynthesis of protective glandular trichome acylsugars in cultivated tomato and wild relatives (Weinhold and Baldwin, 2011; Leckie et al., 2016; Luu et al., 2017). The large BAHD family includes enzymes that perform *O*- and *N*- acylation of structurally diverse acceptor substrates such as alkaloids, phenylpropanoids, terpenoids and acylsugars (Blount et al., 2012). Acylsucroses are produced through consecutive ASAT-mediated conjugation of acyl-CoAs, starting with sucrose to form mono-, di-, tri- and tetra-acylated sucroses (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016a). These enzymes contribute to intraspecific and interspecific diversity in the profiles of insecticidal acylsugars of tomato and its wild relatives.

The second enzyme of acylsugar biosynthesis in cultivated and wild tomato species (ASAT2) exhibits an interesting pattern of substrate preference for iso- (3-methylbutyrate) and anteiso branched (2-methylbutyrate) C5 as well as nC12 acyl chains (Fan et al., 2016a). ASAT2 in cultivated tomato shows higher specificity, using aiC5-CoA but not the structurally similar iC5-CoA as a donor substrate. In contrast, the ASAT2 orthologs in other wild tomatoes are promiscuous, using both aiC5-CoA and iC5-CoA. (Fan et al., 2016a). A combination of primary sequence comparisons and homology modeling of the isozymes using the crystal structure of a distantly related BAHD enzyme led to identification of the amino acid responsible for the

difference in promiscuity (Fan et al., 2016a; Fan et al., 2016b). Specifically, Phe408 restricts and Val408 enables iso-branched C5 acyl chain utilization. This simple change contributes to accumulation of more diverse acylsugars in the wild tomato, *Solanum habrochaites* LA1777, compared to the cultivated species (Ghosh et al., 2014). This example illustrates that enzymatic promiscuity can arise due to simple changes and facilitate the generation of product diversity in specialized metabolism (Figure 2.1B).

Hydroxycinnamoyltransferases illustrate how the interface of enzyme and substrate structure shape promiscuity

In addition to sequence changes, structural similarities of the substrates and how they interact with structural elements of the enzymes can also influence enzyme promiscuity. This theme is exemplified by the hydroxycinnamoyltransferases (HCTs), which play a role in phenylpropanoid biosynthesis (Hoffmann et al., 2003; Sander and Petersen, 2011; Eudes et al., 2016b; Levsh et al., 2016). HCTs are involved in the conjugation of *p*-coumaroyl CoA to shikimate to form *p*-coumaroyl shikimate, and can be highly promiscuous (Landmann et al., 2011; Sander and Petersen, 2011; Eudes et al., 2016a; Levsh et al., 2016). For example, the *A. thaliana* HCT (AtHCT) uses nine substrates besides the native substrate shikimate – some better than shikimate (Figure 2.2C) – to produce a diversity of products *in vitro* (Eudes et al., 2016b).

Crystallography and molecular dynamics analysis of AtHCT highlighted two mechanisms by which HCT maintains its *in vivo* function while being such a promiscuous enzyme (Levsh et al., 2016). First, comparison of AtHCT apoenzyme:substrate crystal structure and molecular dynamics revealed that this enzyme undergoes a conformational change upon *p*-coumaroyl-CoA or *p*-coumaroylshikimate binding: this reduces the volume of the active site and

induces changes in the position of residues involved in substrate binding and catalysis (Levsh et al., 2016). Second, a conserved Arg 'handle' in the active site forms electrostatic interactions with the shikimate carboxyl group, orienting the 5-hydroxyl towards the catalytic center (Figure 2.2D) (Levsh et al., 2016). An HCT structure from *Coleus blumei* (CbHCT) was used to corroborate both features (Levsh et al., 2016). CbHCT was crystallized as a ternary complex with *p*-coumaroyl-CoA and 3-hydroxyacetophenone (Figure 2.2C), an uncharged non-native substrate (Levsh et al., 2016). The CbHCT structure revealed that the active site does not shrink when bound to 3-hydroxyacetophenone (Levsh et al., 2016). Furthermore, mutation of the Arg to Ala, Asp, or Glu caused a drastic decrease in activity with shikimate (Levsh et al., 2016). Conversely, those mutations increased activity with positively-charged non-native substrates such as 3,4-dihydroxybenzylamine and dopamine (Figure 2.2C), demonstrating that the Arg helps discriminate against some non-native substrates (Levsh et al., 2016).

These two structural features play roles in substrate specificity with 3-hydroxyacetophenone or positively-charged substrates, but may not discriminate against substrates that are more structurally similar to shikimate. The Arg residue was shown to form electrostatic interactions with the shikimate carboxyl, discriminating against substrates lacking the carboxyl group (Levsh et al., 2016). Perhaps more interesting are the promiscuous activities of AtHCT (Eudes et al., 2016b). AtHCT has activity with the non-native substrates gentisate and 3-hydroxyanthranilate, which share the negatively charged carboxyl group with shikimate (Fig 2). We hypothesize that the Arg handle – in combination with the carboxyl moiety in the promiscuous substrates – may facilitate activity with non-native substrates (Eudes et al., 2016b). This is an example of how multiple structural elements of an enzyme – in this case the Arg handle and a dynamic conformation change – help maintain the primary activity despite

numerous alternative reactions. These studies also highlight how the interplay of structural similarities of substrates and enzymatic features facilitate promiscuity. Discovery of more mechanisms that enable enzymes to maintain a primary metabolic activity despite being promiscuous will lead to general principles that can guide efforts to engineer specialized metabolic enzymes.

How promiscuity allows 4-coumarate:CoA ligase to catalyze multiple steps in phenylpropanoid biosynthesis

4-coumarate:CoA ligase (4CL) catalyzes multiple steps in the biosynthesis of phenylpropanoids including structural lignins and defensive anti-microbial compounds (Figure 2.3D). Promiscuity plays a central role in the ability of the enzyme to conjugate at least three different hydroxylated and methylated forms of hydroxycinnamic acids to coenzyme A in multiple plant species (Figure 2.3A) (Lee and Douglas, 1996; Hu et al., 1998; Ehlting et al., 1999). Two 4CL crystal structures inform our understanding of substrate promiscuity (Hu et al., 2010; Li and Nair, 2015). The *Populus tomentosa* 4CL1 structure revealed a largely hydrophobic hydroxycinnamate binding pocket (Hu et al., 2010). Studies of *Nicotiana tabacum* 4CL2 complexed with feruloyl-adenylate confirmed the existence of this hydrophobic binding pocket, and revealed how it accommodates hydroxycinnamic acids of varying structure (Li and Nair, 2015). This binding flexibility is mediated by interaction with a tyrosine residue responsible for
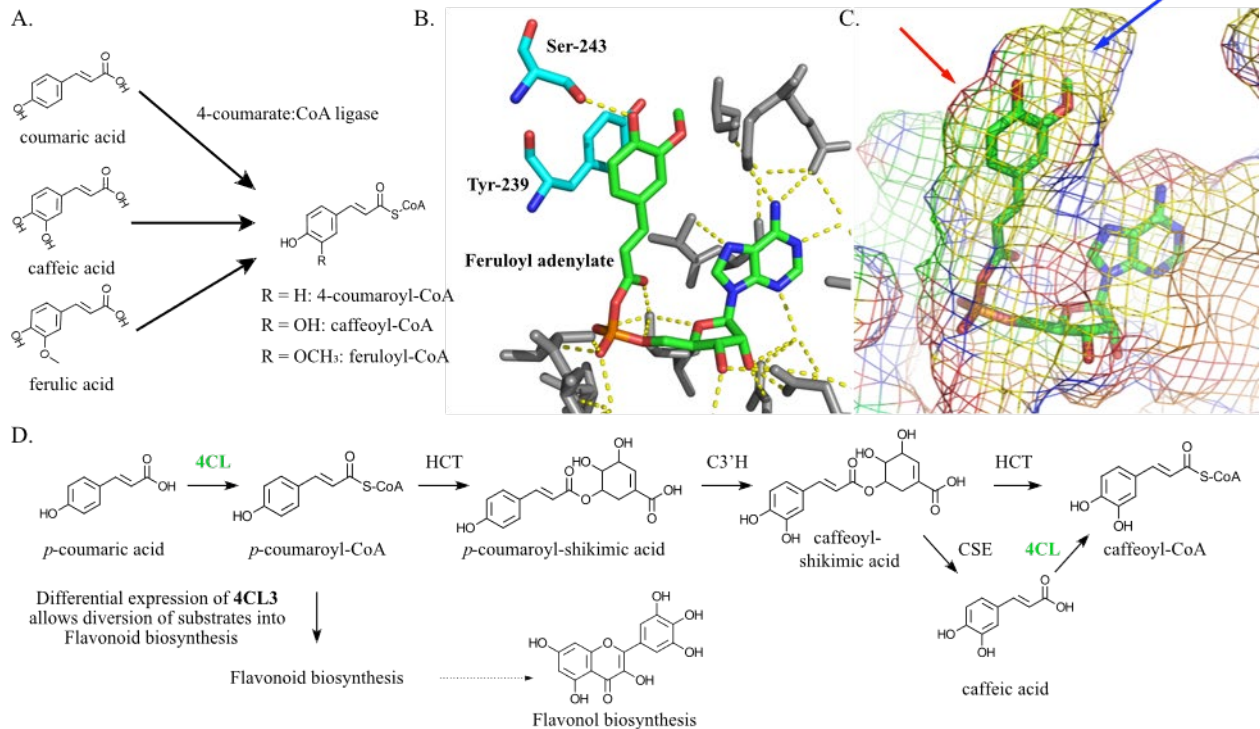
Figure 2.3. **4-coumarate ligase characteristics that impact product diversity.**

(A) At4CL utilizes substrates with hydrogen or hydroxyl and *O*-methyl substitution at the *meta* position of the aromatic ring. (B) Three-dimensional representation of the active site residue interactions with feruloyl-adenylate from the *N. tabacum* 4CL2 co-crystal structure. Feruloyl-adenylate backbone is in green, while residues that interact with the feruloyl moiety are cyan. The red color represents oxygen atoms, the blue color represents nitrogen atoms, and the orange represents phosphorus atoms [PDB ID: 5BSV]. (C) Mesh model of active site around feruloyl-adenylate. Presences or absence of steric hindrance in the active site allows substitution at one of the *meta* positions on the ring, but not the other. Red arrow indicates the *meta* position that is blocked by steric hindrance, while the blue arrow points to position at which substitutions are permitted by the enzyme [PDB ID: 5BSV]. (D) Representation of the phenylpropanoid biosynthetic pathway impacted by 4CL activity. 4CL is involved in biosynthetic steps colored green (Li et al., 2015).

ring stacking with the substrate phenyl ring (Figure 2.3B). The enzyme binding pocket accommodating the hydroxycinnamoyl moiety is spacious at one of the *meta* positions on the phenyl ring and constrained at the other, which accommodates the hydroxy and methoxy groups in caffeic and ferulic acid, respectively (Figure 2.3C, red arrow shows constrained region and blue arrow identifies the spacious area). A serine residue in the active site hydrogen bonds with the *para* hydroxyl present in many monolignols (Figure 2.3B). These combined structural features accommodate diverse modified substrates allowing 4CL to catalyze multiple steps in phenylpropanoid metabolism.

Four isoforms of 4CL in *A. thaliana* use multiple hydroxycinnamate substrates *in vitro* and have distinct – but overlapping – roles (Ehlting et al., 1999; Hamberger and Hahlbrock, 2004; Li et al., 2015). Recent work suggests that one or more of the At4CL isoforms – in addition to ligating 4-coumaric acid to CoA – catalyzes the formation of caffeoyl-CoA, and this appears to be relevant *in vivo*, representing a later biosynthetic step towards guaiacyl and syringyl lignin (Figure 2.3D) (Vanholme et al., 2013; Li et al., 2015). The action of 4CL in multiple biosynthetic steps is conferred by the catalytic flexibility provided by promiscuous activities. In addition, At4CL3 is involved in flavonoids biosynthesis, which shares the common intermediate – 4-coumaroyl-CoA – with the lignin pathway, but uses different enzymes to synthesize their downstream products (Figure 2.3D). A change in At4CL3 expression diverts metabolites away from the lignin pathway, towards flavonoid biosynthesis (Dobritsa et al., 2011; Li et al., 2015). 4CL illustrates how a combination of promiscuity, gene duplication, and divergence in gene expression potentiate the diverse roles that 4CL plays in phenylpropanoids biosynthesis.

Promiscuity in Salicylic and Benzoic acid methyltransferase evolution

Salicylic Acid (SA) and Benzoic Acid (BA) methyltransferases provide another example where promiscuity plays a role in enzyme function and evolution. These enzymes are in the SABATH (SAMT, BAMT, and Theobromine synthase) methyltransferase family, involved in the methylation of structurally diverse specialized metabolites (Zhao et al., 2008). BA and SA differ by the presence of a 2-hydroxyl group on the aromatic ring (Figure 2.4A), and the enzymes catalyzing methylation of BA and SA are promiscuous (Pott et al., 2004; Huang et al., 2012). The volatile methylated benzenoids are produced in differing ratios across characterized plant groups (Altenburger and Matile, 1988; Loughrin et al., 1990; Pott et al., 2002), and have ecological roles including insect pollinator attraction (Knudsen et al., 1993). The enzyme responsible for SA methylation in *Clarkia breweri* is promiscuous, possessing activities with BA, 3-hydroxybenzoic acid, and cinnamic acid substrates (69%, 2%, and 2% of SA activity, respectively) (Ross et al., 1999).

The structural determinants that facilitate SA carboxyl methyltransferase promiscuity ─ including those involved in the binding of *C. breweri* SAMT (CbSAMT) to SA ─ were explored by X-ray crystallographic analysis of CbSAMT (Zubieta et al., 2003). Methionine residues at positions 150 and 308 form a clamp on both faces of the SA benzyl ring, and the side chain nitrogens of Gln-25 and Trp-151 form hydrogen bonds with the SA carboxylate moiety (Figure 2.4B) (Zubieta et al., 2003). The SA 2-hydroxyl moiety is not close enough to any active site residues or water molecules to form hydrogen bonds; thus there are no obvious enzyme structural elements that would discriminate between BA and SA (Zubieta et al., 2003). However, the 2-hydroxyl forms an intramolecular hydrogen bond with the carboxylate group of SA, which
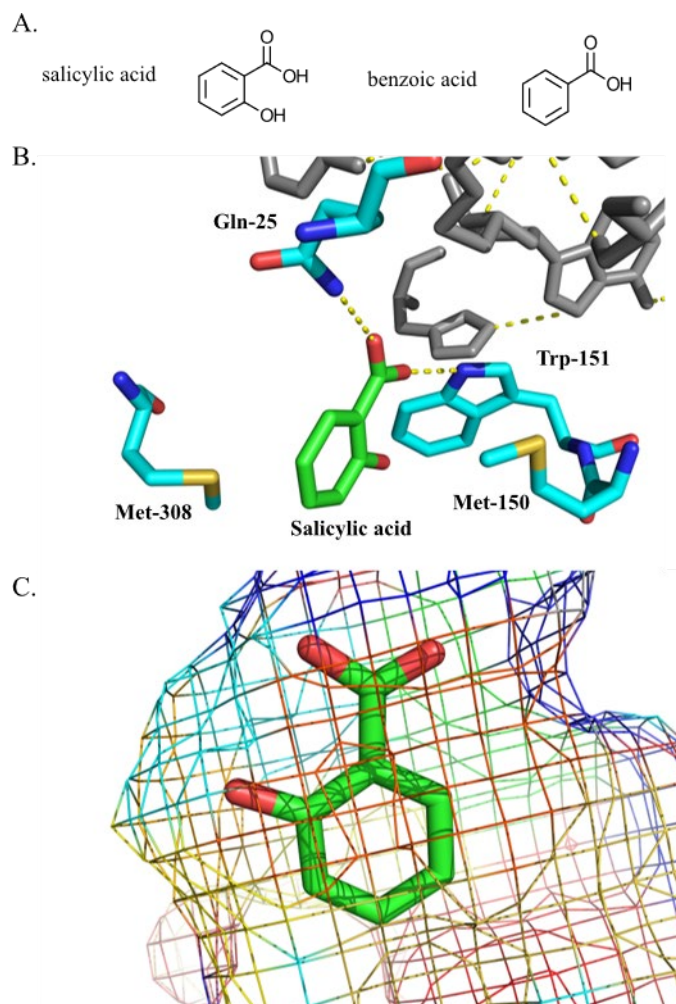
Figure 2.4. Promiscuity in SAMTs and BSMTs is facilitated by substrate structural similarity.

(A) SA and BA have structures that differ by a single hydroxyl at the 2 position. (B) Three-dimensional representation of *C. breweri* SAMT active site interactions with SA. Salicylic acid is colored green, while the interacting active site residues are colored cyan. The red color represents oxygen atoms, the blue represent nitrogen atoms, and the yellow represent sulfur atoms [PDB ID: 1M6E]. (C) Mesh representation of the *C. breweri* SAMT active site demonstrating the steric environment around SA [PDB ID: 1M6E].

restricts its movement, potentially influencing SA and BA activities (Zubieta et al., 2003).

Finally, the remaining active site residues sterically restrain the SA substrate, allowing for

methylation of the carboxylate to occur (Figure 2.4C) (Zubieta et al., 2003). This example

illustrates how similarity of the favored substrate and structural variants can influence

promiscuity. These binding site structural elements can accommodate both SA and BA, and this

promiscuity appears to have influenced the evolutionary history of a clade of SABATH

methyltransferases (Huang et al., 2012).

A combination of phylogenetic analysis, ancestral sequence reconstruction, and enzyme

assays was used to infer that the major activity of the enzyme switched between BA and SA

methylation two times during evolution of the clade that includes SAMTs and BA/SA carboxyl

methyltransferases from plants in the Apocynaceae and Solanaceae families (Huang et al., 2012).

A unique feature of this integrative study is that ancestral sequence reconstruction was used to

predict and test the enzymatic activity of extinct specialized metabolic enzymes at different

nodes in the evolutionary history of this methyltransferase clade (Huang et al., 2012). All

predicted ancestral enzymes tested possessed some level of activity with both BA and SA

(Huang et al., 2012). These results led to the hypothesis that two shifts in substrate preference

occurred over time, with minor promiscuous activities of ancestral enzymes emerging as the

primary activity of descendants (Huang et al., 2012). In this example, the promiscuity of these

different ancestral enzymes potentiated the shifts in enzymatic activity. This work elegantly

demonstrates evolution acting on minor activities of promiscuous enzymes, leading to new major

enzyme activities (Huang et al., 2012).

Conclusion

The examples in this review illustrate multiple ways in which plant specialized metabolic enzymes generate product diversity and metabolic flexibility. Substrate promiscuity, changes in enzyme activity, gene duplication as well as changes in gene expression all provide potentiating environments (Blount et al., 2012) for the evolution of metabolic novelty. In addition to its importance in understanding evolution of protein form and function, a deeper appreciation of the structural basis of simple changes in substrate specificity (Figure 2.1C) and how promiscuous enzymes maintain a primary activity despite the ability to utilize one or more alternate substrates should inform protein engineering and synthetic biology approaches.

REFERENCES

REFERENCES

Altenburger R, Matile P (1988) Circadian rhythmicity of fragrance emission in flowers of *Hoya carnosa* R. Br. Planta **174**: 248–252.

Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature **489**: 513–8.

Dixon RA, Strack D (2003) Phytochemistry meets genome analysis, and beyond... Phytochemistry **62**: 815–816.

Dobritsa AA, Geanconteri A, Shrestha J, Carlson A, Kooyers N, Coerper D, Urbanczyk-Wochniak E, Bench BJ, Sumner LW, Swanson R, et al (2011) A large-scale genetic screen in Arabidopsis to identify genes involved in pollen exine production. Plant Physiol **157**: 947–970.

Ehlting J, Büttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E (1999) Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. Plant J **19**: 9–20.

Eudes A, Mouille M, Robinson DS, Benites VT, Wang G, Roux L, Tsai Y-L, Baidoo EEK, Chiu T-Y, Heazlewood JL, et al (2016a) Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. Microb Cell Fact **15**: 198.

Eudes A, Pereira JH, Yogiswara S, Wang G, Teixeira Benites V, Baidoo EEK, Lee TS, Adams PD, Keasling JD, Loqué D (2016b) Exploiting the substrate promiscuity of hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase to reduce lignin. Plant Cell Physiol **57**: 568–579.

Fan P, Miller AM, Schilmiller AL, Liu X, Ofner I, Jones AD, Zamir D, Last RL (2016a) In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. Proc Natl Acad Sci U S A **113**: E239-48.

Fan P, Moghe GD, Last RL (2016b) Comparative Biochemistry and In Vitro Pathway Reconstruction as Powerful Partners in Studies of Metabolic Diversity. Methods Enzymol. doi: 10.1016/bs.mie.2016.02.023.

Ghosh B, Westbrook TC, Jones AD (2014) 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.

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol **57**: 303–333.

Hamberger B, Hahlbrock K (2004) The 4-coumarate:CoA ligase gene family in *Arabidopsis*

thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc Natl Acad Sci **101**: 2209–2214.

He Y, Chen L, Zhou Y, Mawhinney TP, Chen B, Kang B-H, Hauser BA, Chen S (2011a) Functional characterization of *Arabidopsis thaliana* isopropylmalate dehydrogenases reveals their important roles in gametophyte development. New Phytol **189**: 160–175.

He Y, Galant A, Pang Q, Strul JM, Balogun SF, Jez JM, Chen S (2011b) Structural and functional evolution of isopropylmalate dehydrogenases in the Leucine and glucosinolate pathways of *Arabidopsis thaliana*. J Biol Chem **286**: 28794–28801.

He Y, Mawhinney TP, Preuss ML, Schroeder AC, Chen B, Abraham L, Jez JM, Chen S (2009) A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in Arabidopsis. Plant J **60**: 679–690.

Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. J Biol Chem **278**: 95–103.

Hu WJ, Kawaoka A, Tsai CJ, Lung J, Osakabe K, Ebinuma H, Chiang VL (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). Proc Natl Acad Sci U S A **95**: 5407–12.

Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, Li D, Jiang X-N, Wang D-C (2010) Crystal structures of a *Populus tomentosa* 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. Plant Cell **22**: 3093–3104.

Huang R, Hippauf F, Rohrbeck D, Haustein M, Wenke K, Feike J, Sorrelle N, Piechulla B, Barkman TJ (2012) Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. Proc Natl Acad Sci **109**: 2966–2971.

Knudsen JT, Tollsten L, Bergström LG (1993) Floral scents—a checklist of volatile compounds isolated by head-space techniques. Phytochemistry **33**: 253–280.

Landmann C, Hücherig S, Fink B, Hoffmann T, Dittlein D, Coiner HA, Schwab W (2011) Substrate promiscuity of a rosmarinic acid synthase from lavender (*Lavandula angustifolia* L.). Planta **234**: 305–20.

Leckie BM, D'Ambrosio DA, Chappell TM, Halitschke R, De Jong DM, Kessler A, Kennedy GG, Mutschler MA (2016) Differential and synergistic functionality of acylsugars in suppressing oviposition by insect herbivores. PLoS One **11**: e0153345.

Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family (cDNA structure, gene inheritance and expression, and properties of recombinant proteins). Plant Physiol **112**: 193–205.

Lee SG, Nwumeh R, Jez JM (2016) Structure and mechanism of isopropylmalate

dehydrogenase from *Arabidopsis thaliana*. J Biol Chem **291**: 13421–13430.

Levsh O, Chiang Y-C, Tung CF, Noel JP, Wang Y, Weng J-K (2016) Dynamic conformational states dictate selectivity toward the native substrate in a substrate-permissive acyltransferase. Biochemistry **55**: 6314–6326.

Li Y, Kim JI, Pysh L, Chapple C (2015) Four isoforms of *Arabidopsis thaliana* 4-coumarate: CoA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism. Plant Physiol pp.00838.2015.

Li Z, Nair SK (2015) Structural basis for specificity and flexibility in a plant 4-coumarate:CoA ligase. Structure **23**: 2032–2042.

Loughrin JH, Hamilton-Kemp TR, Andersen RA, Hildebrand DF (1990) Headspace compounds from flowers of *Nicotiana tabacum* and related species. J Agric Food Chem **38**: 455–460.

Luu VT, Weinhold A, Ullah C, Dressel S, Schoettner M, Gase K, Gaquerel E, Xu S, Baldwin IT (2017) *O*-acyl sugars protect a wild tobacco from both native fungal pathogens and a specialist herbivore. Plant Physiol **174**: 370–386.

Milo R, Last RL (2012) Achieving diversity in the face of constraints: lessons from metabolism. Science (80-) **336**: 1663–1667.

Mithöfer A, Boland W (2012) Plant defense against herbivores: chemical aspects. Annu Rev Plant Biol **63**: 431–450.

Moghe G, Last RL (2015) Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. Plant Physiol pp.00994.2015.

Pott MB, Hippauf F, Saschenbrecker S, Chen F, Ross J, Kiefer I, Slusarenko A, Noel JP, Pichersky E, Effmert U, et al (2004) Biochemical and structural characterization of benzenoid carboxyl methyltransferases involved in floral scent production in *Stephanotis floribunda* and *Nicotiana suaveolens*. Plant Physiol **135**: 1946–1955.

Pott MB, Pichersky E, Piechulla B (2002) Evening specific oscillations of scent emission, SAMT enzyme activity, and SAMT mRNA in flowers of *Stephanotis floribunda*. J Plant Physiol **159**: 925–934.

Ross JR, Nam KH, D'Auria JC, Pichersky E (1999) *S*-adenosyl-l-methionine:salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant Methyltransferases. Arch Biochem Biophys **367**: 9–16.

Sander M, Petersen M (2011) Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl-transferase, from *Coleus blumei* Benth. Planta **233**: 1157–1171.

Schilmiller AL, Charbonneau AL, Last RL (2012) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. Proc Natl Acad Sci **109**: 16377–16382.

Schilmiller AL, Moghe GD, Fan P, Ghosh B, Ning J, Jones AD, Last RL (2015) Functionally divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. Plant Cell **27**: 1002–1017.

Sønderby IE, Geu-Flores F, Halkier BA (2010) Biosynthesis of glucosinolates – gene discovery and beyond. Trends Plant Sci **15**: 283–290.

St-Pierre B, De Luca V (2000) Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. Recent Adv. Phytochem. pp 285–315.

Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P, et al (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. Science (80-) **341**: 1103–1106.

Weinhold A, Baldwin IT (2011) Trichome-derived *O*-acyl sugars are a first meal for caterpillars that tags them for predation. Proc Natl Acad Sci **108**: 7855–7859.

Weng JK, Philippe RN, Noel JP (2012) The rise of chemodiversity in plants. Science (80-) **336**: 1667–1670.

Zhao N, Ferrer J-L, Ross J, Guan J, Yang Y, Pichersky E, Noel JP, Chen F (2008) Structural, biochemical, and phylogenetic analyses suggest that indole-3-acetic acid methyltransferase is an evolutionarily ancient member of the SABATH family. Plant Physiol **146**: 455–467.

Zubieta C, Ross JR, Koscheski P, Yang Y, Pichersky E, Noel JP (2003) Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. Plant Cell **15**: 1704–1716.

Chapter 3. Acylglucose biosynthesis in *Solanum pennellii*

Contributions: I contributed to data collection and analysis for Fig. 3.1, 3.2, 3.3, 3.4, 3.5, S3.1, S3.3-S3.9, Table S3.1. I also contributed toward assembling and formatting all the figures, in addition to writing parts of the manuscript.

Abstract

Plants produce a myriad of taxonomically restricted specialized metabolites. This diversity—and our ability to correlate genotype with phenotype—makes the evolution of these ecologically and medicinally important compounds interesting and experimentally tractable. Trichomes of tomato and other nightshade family plants produce structurally diverse protective compounds termed acylsugars. While cultivated tomato (*Solanum lycopersicum*) strictly accumulates acylsucroses, the South American wild relative *Solanum pennellii* produces copious amounts of acylglucoses. Genetic, transgenic, and biochemical dissection of the *S. pennellii* acylglucose biosynthetic pathway identified a trichome gland cell–expressed invertase-like enzyme that hydrolyzes acylsucroses (Sopen03g040490). This enzyme acts on the pyranose ring–acylated acylsucroses found in the wild tomato but not on the furanose ring–decorated acylsucroses of cultivated tomato. These results show that modification of the core acylsucrose biosynthetic pathway leading to loss of furanose ring acylation set the stage for co-option of a general metabolic enzyme to produce a new class of protective compounds.

Introduction

The cultivated tomato biosynthetic network is well characterized, with four ASATs—SlASAT1 to SlASAT4—catalyzing consecutive reactions to produce tri- and tetra-acylated sucroses. SlASAT1 acts first by transferring an acyl chain to the R4 hydroxyl of the pyranose ring of sucrose, and SlASAT2 transfers an acyl chain to the R3 position of the monoacylated sucrose (Fan et al., 2016). Next, SlASAT3 acylates the diacylated sucroses at the furanose ring R3′ position (Schilmiller et al., 2015). SlASAT4 completes the pathway by transferring an acetyl group to the pyranose ring R2 position of a triacylsucrose (Kim et al., 2012; Schilmiller et al.,

2012). Enzyme promiscuity and the presence of an array of acyl-CoAs result in the production of a diverse group of acylsucroses in *Solanum lycopersicum* (Ning et al., 2015; Fan et al., 2017).

The metabolic diversity in acylsugars is even greater in the broader Solanum genus. The wild relative of tomato, *Solanum pennellii* LA0716, is a prime example, producing a mixture of abundant acylsucroses that are distinct from those in *S. lycopersicum*. While *S. lycopersicum* accumulates acylsucroses with two or three acylations on the pyranose ring and a single acylation at the furanose ring R3′ position (termed F-type acylsucroses), *S. pennellii* accumulates distinct P-type triacylsucroses acylated only on the pyranose R2, R3, and R4 positions (Fan et al., 2017). P-type acylsucroses are synthesized by *S. pennellii* orthologs of the *S. lycopersicum* ASAT1, ASAT2, and ASAT3 enzymes. The different acylation pattern observed in *S. pennellii* results from altered substrate specificity and acylation position of SpASAT2 and SpASAT3 relative to their *S. lycopersicum* counterparts (Fan et al., 2017).

S. pennellii LA0716 has other acylsugar characteristics that differentiate it from cultivated tomato (Fig. 3.1A,B). First, it produces copious amounts of acylsugars that render the plant sticky, representing up to ~20% of leaf dry weight (Fobes et al., 1985; Burke et al., 1987). Second, the vast majority of *S. pennellii* LA0716 acylsugars are glucose molecules with three acyl chains (termed "acylglucoses") (Fig. 3.1A and B), while only 7 to 16% of total acylsugars are acylsucroses (Shapiro et al., 1994). In contrast to the well-characterized *S. pennellii* acylsucrose biosynthetic enzymes (Schilmiller et al., 2015; Fan et al., 2017), no complete acylglucose metabolic pathway has yet been described. This is despite the fact that acylglucoses were also characterized in several additional Solanaceae species (King and Calhoun, 1988; Matsuzaki et al., 1989).

A previously proposed partial *S. pennellii* pathway invoked two glucosyltransferases capable of creating 1-O-acyl-d-glucose from uridine diphosphate (UDP)–glucose and free fatty acids of differing structures (Kuai et al., 1997). This mechanism proposed a second step in which a serine carboxypeptidase-like (SCPL) acyltransferase catalyzed disproportionation of two 1-O-isobutyryl- D-glucose molecules to yield one 1,2-O-di-isobutyryl-D-glucose (Li et al., 1999; Li and Steffens, 2000). However, this pathway is unlikely to function in vivo as the 1,2,-*O*-diacylglucoses obtained in vitro differ from the 2,3,4-*O*-tri-acylglucoses observed in *S. pennellii* both in the number (two instead of three) and in the position of acyl chains: *S. pennellii* acylglucoses bear chains at the R2, R3, and R4 positions rather than at the R1 position (Burke et al., 1987). In contrast to the unsubstantiated published biosynthetic path- way, compelling quantitative trait locus (QTL) and biochemical results implicate multiple genetic loci in acylglucose accumulation in *S. pennellii* LA0716. The combination of three *S. pennellii* regions on chromosomes 3, 4, and 11 causes *S. lycopersicum* breeding line CU071026 to accumulate
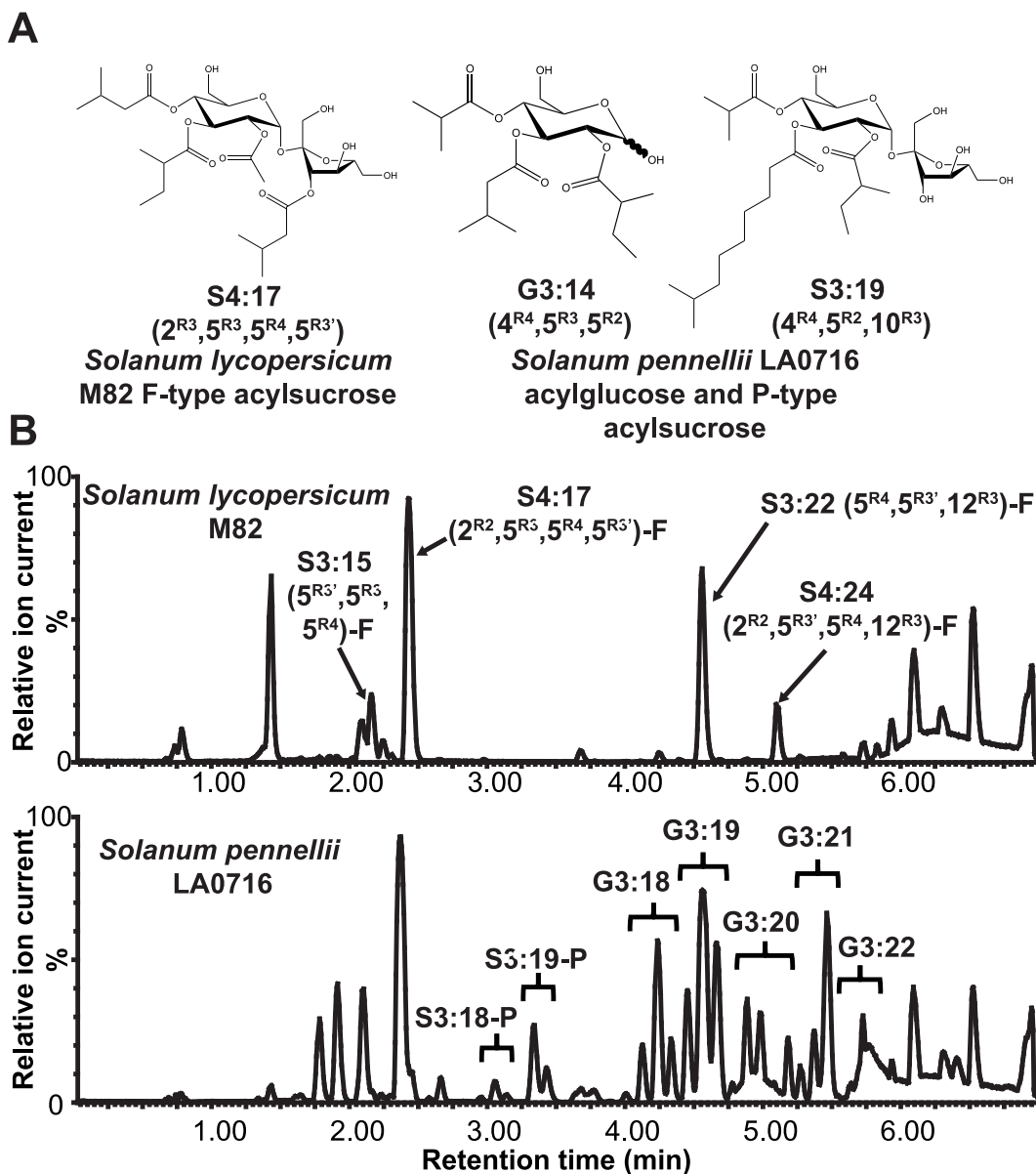
Figure 3.1. Acylsugars from *Solanum lycopersicum* and *Solanum pennellii*.

(A) Examples of NMR-characterized *S. lycopersicum* and *S. pennellii* acylsugar structures. (B) Acylsugar ESI- mode LC-QToF MS profiles. Top: *S. lycopersicum* M82 with acylsucroses S3:15 $(5^{R3'},5^{R3},5^{R4})$-F, S4:17 $(2^{R2},5^{R3'},5^{R3},5^{R4})$-F, S3:22 $(5^{R3'},5^{R4},12^{R3})$-F, and S4:24 $(2^{R2},5^{R3'},5^{R4},12^{R3})$-F annotated. Bottom: *S. pennellii* LA0716 acylsucroses and acylglucoses. Note: S3:15 (5,5,5)-F denotes a sucrose acylsugar with 3 acyl chains of 5 carbons each. The 'F' denotes that at least one of the acyl chains is present on the 5-membered (furanose) ring. The superscript depicts acylation position of the acyl chains. When NMR-derived structural information is available, superscripts indicate acyl chain positions with R representing the pyranose ring, and R' representing the furanose ring.

acylsugars comprising up to 89% acylglucoses (Leckie et al., 2013). The presence of QTLs on both chromosomes 3 and 11 yields detectable acylglucoses, while addition of the chromosome 4 locus leads to elevated accumulation. Notably, chromosome 4 and 11 QTLs respectively include the SpASAT2 and SpASAT3 genes responsible for accumulation of P-type acylsucroses in *S. pennellii* (Schilmiller et al., 2012; Fan et al., 2016). The agreement of data from QTL and biochemical analyses is consistent with the hypothesis that SpASAT2 and SpASAT3 produce P-type acylsucroses that are substrates for a chromosome 3 factor that then synthesizes triacylglucoses.

Results

Published QTL mapping studies indicate that introgression of *S. pennellii* LA0716 loci on chromosomes 3, 4, and 11 leads to accumulation of acylglucoses in a cultivated tomato *S. lycopersicum* background (Leckie et al., 2013; Smeda et al., 2018). Three introgression lines harboring individual acylglucose QTLs in the *S. lycopersicum* background were screened, but none of the single introgressions in lines IL3-5, IL4-1, or IL11-3 (Eshed and Zamir, 1995) yielded detectable leaf acylglucoses (Fig. S3.1). These observations are consistent with the hypothesis that multiple *S. pennellii* loci are needed for *S. lycopersicum* acylglucose accumulation. Indeed, there are low but detectable levels of acylglucoses (87% of total acylsugars; Fig. S3.2) in backcross inbred line BIL6521 (Ofner et al., 2016), which contains *S. pennellii* LA0716 introgressions from chromosomes 1, 3, and 11. This BIL accumulates four acylglucoses (Table S1), with the major one, G3:22 (5,5,12) (Fig. S3.3 and Fig. S3.4), resembling the pyranose ring of the P-type acylsucrose S3:22 (5, 5, 12)-P detected at low levels in trichomes of the single chromosome 11 introgression line, IL11-3 (Schilmiller et al., 2015). In fact, BIL6521 accumulates a P-type acylsucrose, S3:22 (Fig. S3.3 and Fig. S3.4). These results

are consistent with the hypothesis that the chromosome 3 region is necessary for acylglucose production, but only when P-type acylsucroses are produced. Note that in our nomenclature, 'S' and 'G' refer to sucrose or glucose core, respectively, and 3:22 (5,5,12) indicates that there are three ester-linked acyl chains of 5, 5 and 12 carbons, for a total of 22 chain carbons (Schilmiller et al., 2015). When NMR-derived structural information is available, superscripts indicate acyl chain positions with R representing the pyranose ring, and R' representing the furanose ring (Fig. 3.1A).

Because the acylsugar levels in BIL6521, which lack a chromosome 4 introgression, were much lower than most other lines, we tested the impact of adding a chromosome 4 introgression carrying the *SpASAT2* locus. A cross was made between BIL6521 and BIL6180, a recombinant line harboring introgressions on chromosomes 4, 5 and 11, which includes both the *SpASAT2* and *SpASAT3* loci (Fig. 3.2). BIL6180 was previously found to produce only P-type acylsucroses as a result of the chromosome 4 and 11 introgressions (Fig. 3.2), however, it accumulated significantly higher overall levels of acylsucroses compared to BIL6521 (Fig. S3.2) as well as other short-chain containing P-type acylsucroses not present in BIL6521. If all P-type acylsucroses are substrates for a *S. pennellii* LA0716 factor on chromosome 3, we predicted that both of the corresponding acylglucoses G3:15 (5,5,5) and G3:22 (5,5,12) would accumulate in a line harboring the chromosome 3, 4, and 11 introgressions. Indeed, the F2 progeny of BIL6521 × BIL6180, genotyped as heterozygous for the *S. pennellii* chromosome 3 and 4 introgressions, and homozygous for the *S. pennellii* chromosome 11 region, produced these two predicted acylglucoses (Fig. 3.1C, Fig. S3.3 and Fig. S3.4). These findings – in combination with the published QTL results – indicate that the *S. pennellii* chromosome 3 introgression is necessary

54
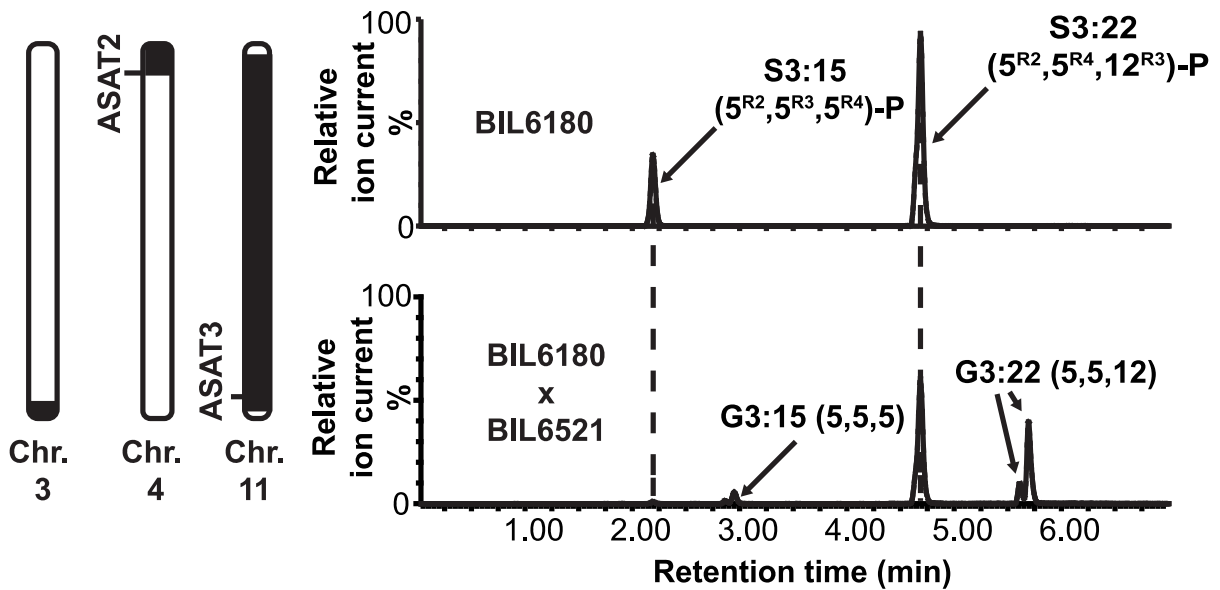
Figure 3.2. Comparison of acylsugars between two backcrossed lines demonstrates the combination of loci is responsible for acylglucose biosynthesis.

Left: Representation of *S. pennellii* chromosomal introgressions in BIL6521 x BIL6180 F2 progeny that contain QTLs affecting acylglucose biosynthesis (Leckie et al., 2013). The black portions of the chromosomes correspond to *S. pennellii* introgressions, while the white portions correspond to the chromosomal regions in the M82 background. Right: ESI- mode LC-MS analysis of BIL6180 compared with the BIL6180 x BIL6521 F2 progeny reveals acylglucose (G-labeled products) accumulation in the hybrid, but not in BIL6180. All ESI- mode acylsugars were identified as their formate adducts. When NMR-derived structural information is available, superscripts indicate acyl chain positions with R representing the pyranose ring, and R' representing the furanose ring. Mass window: 0.05Da.

for acylglucose biosynthesis and suggests that P-type acylsucroses are acylglucose biosynthetic precursors.

The chromosome 3 locus encodes a glandular trichome-expressed β-fructofuranosidase

We sought candidate glycoside hydrolase genes in the 1.7-Mb QTL AG3.2, the acylglucose-associated region from *S. pennellii* LA0716 previously mapped to the bottom of chromosome 3 (Leckie et al., 2013) (Fig. 3.3A). Three of the 238 genes in this region of the *S. lycopersicum* Heinz 1706 genome assembly SL2.50 annotation (Consortium, 2012) are predicted as encoding glycoside hydrolases (members of the GH32, GH35, and GH47 families; Table S2). We focused on the GH32 family *Sopen03g040490* gene because all previously characterized members of the family have β-fructofuranosidase or fructosyltransferase activity (Van Den Ende et al., 2009). As acylsucroses are β-fructofuranosides, we hypothesized that the GH32 enzyme cleaves the glycosidic bond of P-type acylsucroses to generate acylglucoses. Based on the full results of this study, we designate this gene <u>A</u>CYL<u>S</u>UCROSE <u>F</u>RUCTO<u>F</u>URANOSIDASE 1 (*ASFF1*).

S. lycopersicum acylsugars accumulate in type I/IV glandular trichome tip cells (Nakashima et al., 2016) and trichome tip cell-specific gene expression is a hallmark of all characterized acylsugar biosynthetic genes (*e.g., ASAT1/2/3/4*, *IPMS3*) (Schilmiller et al., 2012; Ning et al., 2015; Schilmiller et al., 2015; Fan et al., 2016)**.** We used a reporter gene approach to ask whether *SpASFF1* exhibits trichome-specific expression. The 1.8-kb region immediately upstream of the *SpASFF1* ORF in the *S. pennellii* LA0716 genome drove expression of a green fluorescent protein-β-glucuronidase fusion protein (GFP-GUS) in *S. lycopersicum* M82 plants.
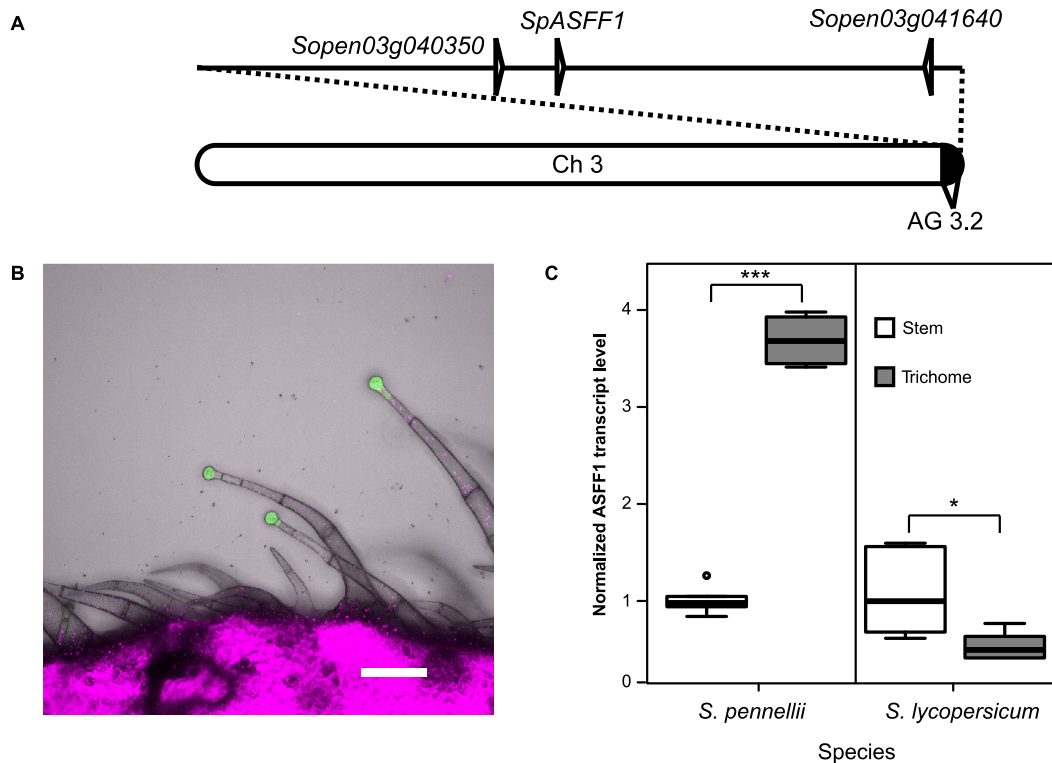
Figure 3.3. Gene *SpASFF1* from QTL AG 3.2 of the glycoside hydrolase 32 family shows trichome-specific expression.

(A) Chromosome 3 with the AG3.2 introgression (*30*). Positions of three glycoside hydrolase genes (*Sopen03g040350, Sopen03g041640*, and *SpASFF1*) on QTL AG 3.2 are indicated. (B) Expression of GFP-GUS under control of the native *ASFF1* promoter from *S. pennellii* LA0716 yields a GFP signal in *S. lycopersicum* M82 type I/IV trichome tip cells, but not in stalk cells or stem tissue. Green channel indicates the GFP signal; magenta channel shows chlorophyll fluorescence. Scale bar = 100 µm. (C) qRT-PCR analysis of transcript abundance indicates that ASFF1 transcripts are higher in *S. pennellii* LA0716 trichomes than in underlying stem tissue but lower in *S. lycopersicum* M82 trichomes than in underlying stem tissue. Whiskers represent minimum and maximum values less than 1.5 times the interquartile range from the 1st and 3rd quartiles, respectively. Values outside this range are represented as circles. Asterisks (*) indicates $p < 0.05$, (***) indicates $p < 0.001$ (Welch two sample *t*-test); $n = 6$ for all species and tissue types.

Indeed, GFP signal in transformed plants was observed in the tip cells of type I/IV trichomes but not in the trichome stalk cells or underlying stem epidermis (Fig. 3.3B). This result is consistent with a role of SpASFF1 enzyme in type I/IV trichome metabolism.

We cross-validated the trichome enriched expression pattern of *ASFF1* in *S. pennellii* LA0716 using Quantitative Reverse Transcriptase PCR (qRT-PCR). ASFF1 transcript levels were 3.7-fold higher in trichomes of *S. pennellii* LA0716 stems than in underlying shaved stem tissue ($p < 0.001$; Welch two sample *t*-test) (Fig. 3.3C). The observed enrichment of transcripts in trichome samples is similar to analysis of previously identified acylsugar biosynthetic genes from tomato, petunia and tobacco (Schilmiller et al., 2012; Ning et al., 2015; Fan et al., 2016; Luu et al., 2017; Nadakuduti et al., 2017). Together, transcript enrichment in trichomes and the restriction of gene expression to trichome tip cells support the hypothesis that *SpASFF1* acts in acylsugar biosynthesis.

Acylglucoses accumulate in *S. pennellii* LA0716 but not in *S. lycopersicum* M82 (Burke et al., 1987; Ghosh et al., 2014). However, *ASFF1* is predicted to encode a full open reading frame in both the *S. lycopersicum* Heinz 1706 and the *S. pennellii* LA0716 genomes (Consortium, 2012; Bolger et al., 2014). While ASFF1 transcripts are enriched in trichomes of *S. pennellii* LA0716, ASFF1 transcripts are half as abundant in *S. lycopersicum* M82 trichomes as in underlying stem tissue ($p < 0.05$; Welch two sample *t*-test) (Fig. 3.3C). Together, the tissue- and species-level specificity of *ASFF1* expression is consistent with a role for the gene in acylglucose biosynthesis.

Gene editing reveals that *SpASFF1* is necessary for *S. pennellii* LA0716 acylglucose accumulation

We used CRISPR/Cas9-mediated gene editing in *S. pennellii* LA0716 to test whether *SpASFF1* is necessary for acylglucose accumulation. Two small guide RNAs (sgRNAs) targeting the third *SpASFF1* exon were used to promote site specific DNA cleavage by hCas9 in the stably transformed plants (Fig. 3.4 and Fig. S3.5). Three homozygous T_1 mutants were obtained with different site-specific mutations, each of which is predicted to cause complete loss of function through translational frame-shifts and premature protein termination. Two of them (*spasff1-1-1* and *spasff1-1-2*), which carry 228 bp and 276 bp insertion-deletions, respectively, are derived from segregation of one heteroallelic T_0 plant. The third mutant (*spasff1-2*) with a 1 bp insertion is the descendant of a homozygous T_0 mutant.

Results from LC-MS analysis of leaf surface metabolites from these lines were consistent with the hypothesis that *SpASFF1* is necessary for acylglucose biosynthesis. All *spasff1* lines failed to accumulate detectable acylglucoses (Fig. 3.4A and B and Fig. S3.6), but produced acylsucroses at levels comparable to total acylsugars in wild-type *S. pennellii* plants (Table 3.1).
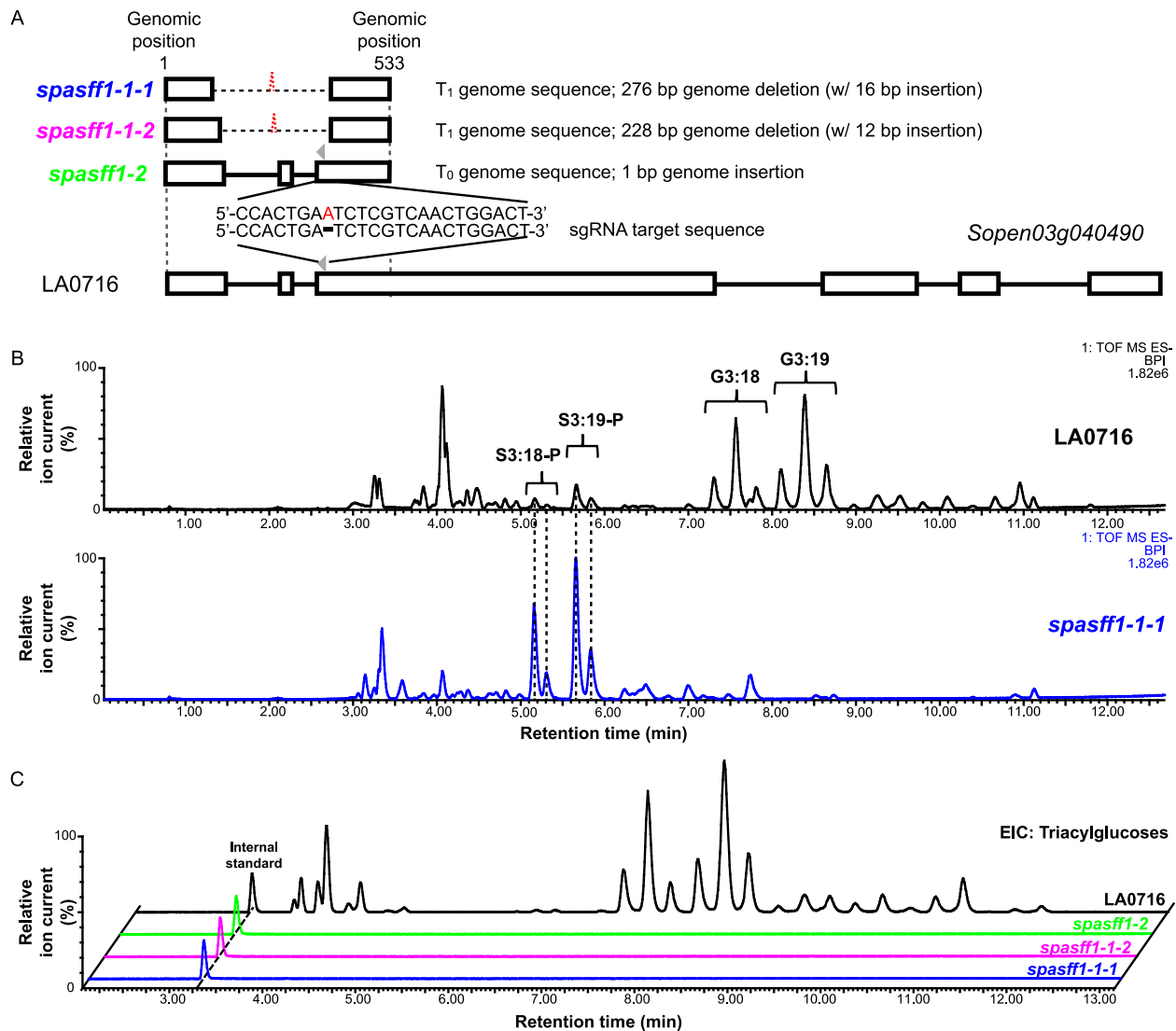
Figure 3.4. CRISPR/Cas9-mediated *S. pennellii* LA0716 *spasff1* knockouts eliminate detectable acylglucoses.

(A) Schematic representation of mutagenesis strategy with two sgRNAs (grey arrowheads – only one sgRNA shown) targeting the *SpASFF1* ORF that result in three homozygous knockout lines. White boxes indicate exons; horizontal bars indicate introns; dotted lines indicate deletions and red letter nucleotides indicates insertion. Mutant allele DNA sequences are found in Fig. S5. (B) Mutant line *spasff1-1-1* accumulates abundant acylsucroses but not acylglucoses. Base peak ESI-mode LC-MS chromatograms are shown for knockout mutant *spasff1-1-1* and LA0716. (C) Extraction ion chromatograms of ESI- mode analysis show the formate adducts of triacylglucose from trichome extracts of *S. pennellii* LA0716 and three *spasff1* mutant plants show that homozygous *asff1* lines produce undetectable levels of triacylglucose. Extracted ion chromatogram values displayed: G3:12 (*m/z*: 435.19), G3:13 (*m/z*: 449.2), G3:14 (*m/z*: 463.22), G3:15 (*m/z*: 477.23), G3:16 (*m/z*: 491.28), G3:17 (*m/z*: 505.26), G3:18 (*m/z*: 519.28), G3:19 (*m/z*: 533.30), G3:20 (*m/z*: 547.31), G3:21 (*m/z*: 561.33), G3:22 (*m/z*: 575.34), and telmisartan

Figure 3.4 (cont'd)

(internal standard) (*m/z*:513.23). The relative base peak intensity (BPI) LC-MS chromatograms for all plant lines are shown in Fig. S6. Mass window: 0.05Da. Note: For panel B and C, *spasff1-1-1/1-1-2* are homozygous T_2 lines, while *spasff1-2* are homozygous T_1 lines that were all grown together. *spasff1* lines were diluted 100 fold before LC-MS analysis to avoid saturation of the LC-MS detector. This is due to higher ionization efficiency of acylsucroses relative to acylglucoses in ESI- mode.

Table 3.1. Acylsugar quantification from *spasff1* lines and WT lines of *S. pennellii* LA0716.

Acyl chains were saponified from the acylsugars, and the resulting sugar cores were analyzed by UPLC-ESI-Multiple Reaction Monitoring. Data are shown from individual T1 homozygous plants grown together but independently from those in Fig. 3.4. These extracts include other glycosylated compounds such as flavonoids, which could be responsible for the nonzero values for glucose measurements in plants lacking detectable acylglucoses (Schilmiller et al., 2010). DW, dry weight.

Line	Plant number	Sucrose (%)	Glucose (%)	Total sugar core (nmol/mg DW)
LA0716	1	1%	99%	41.93
	2	3%	97%	15.78
	3	1%	99%	33.70
	4	2%	98%	55.89
	5	1%	99%	138.80
	6	1%	99%	29.34
spasff1-1-1	1	98%	2%	25.82
	2	99%	1%	23.23
	3	98%	2%	26.79
	4	99%	1%	21.49
spasff1-1-2	1	99%	1%	18.74
	2	99%	1%	24.55
	3	99%	1%	28.72
	4	99%	1%	18.35
spasff1-2	1	97%	3%	80.01
	2	98%	2%	37.51
	3	95%	5%	84.44

SpASFF1 converts pyranose ring-acylated P-type acylsucroses to acylglucoses both *in vivo* and *in vitro*

The results described above strongly suggest that SpASFF1 converts pyranose ring-acylated P-type sucroses to acylglucoses. In addition, IL3-5 does not accumulate detectable acylglucoses despite possessing the *S. pennellii ASFF1* genomic region, suggesting that F-type acylsucroses are not substrates for SpASFF1 (Fig. 3.5A and Fig. S3.7).

We took a transgenic approach to ask whether *SpASFF1* alone is sufficient to confer acylglucose accumulation in a P-type acylsucrose-accumulating background. We transformed the P-type acylsucrose-producing *S. lycopersicum* double introgression BIL6180 (Fig. 3.5B) with a T-DNA containing the *SpASFF1* open reading frame and the 1.8 kb promoter region immediately upstream of its start codon (Fig. 3.5C). In addition to the P-type acylsucroses in the parental BIL6180, the *SpASFF1* transgenics accumulated major hexose acylsugars with MS characteristics consistent with G3:15 (5,5,5) and G3:22 (5,5,12) (Fig. 3.5C, Fig. S3.8). The acyl chain composition of these acylglucoses matches the S3:15 ($5^{R2},5^{R3},5^{R4}$) and S3:22 ($5^{R2},5^{R4},12^{R3}$) P-type acylsucroses detected in BIL6180 (Fig. S3.4). Acylglucoses in the transgenic lines are also identical to those seen in BIL6521 × BIL6180 based on LC retention time and MS fragmentation. This confirms that SpASFF1 converts *S. pennellii* P-type acylsucroses produced by SpASAT2 and SpASAT3 to acylglucoses. Taken together, these *in vivo* results show that *SpASFF1* is sufficient to make acylglucoses *in vivo* when P-type acylsucroses are present, but not in plants accumulating only F-type acylsucroses, where acyl chains are not on the furanose ring.
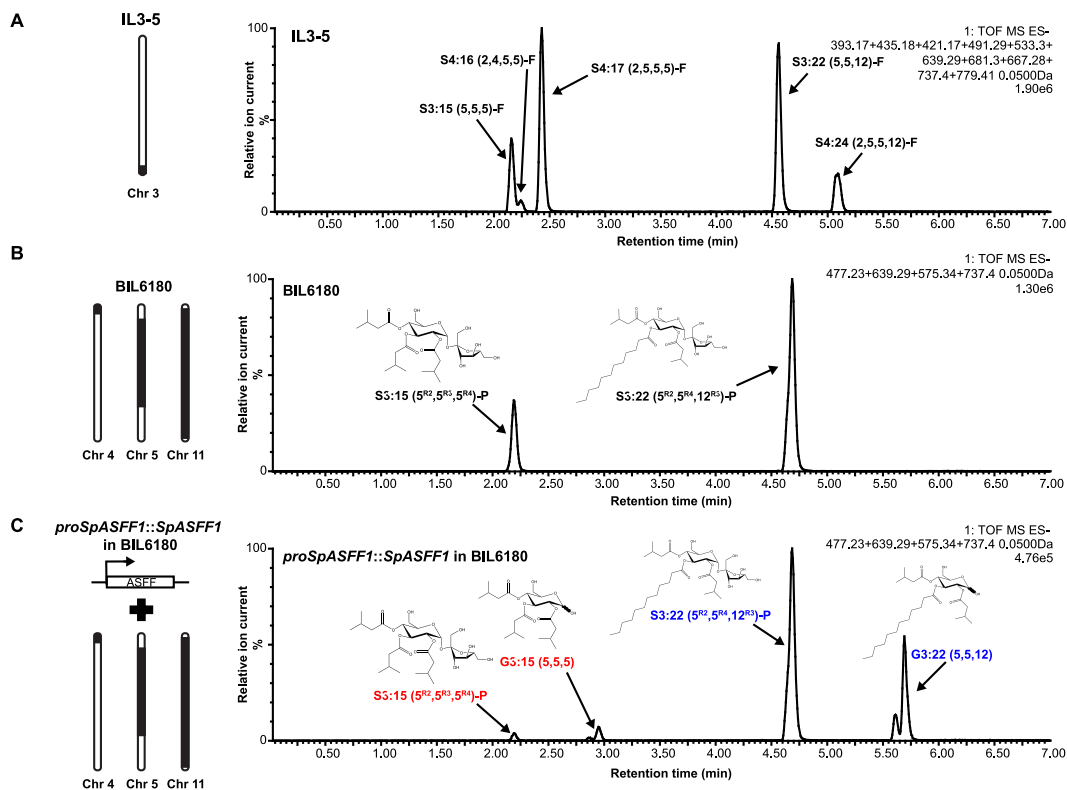
Figure 3.5. Expression of *SpASFF1* in P-type acylsucrose producing BIL6180 trichomes results in accumulation of acylglucoses in surface extracts.

(A) IL3-5 accumulates F-type acylsucroses without detectable acylglucoses. ESI- mode LC-MS analysis of trichome extracts of IL3-5 are shown. Extracted ion chromatograms of S3:15 (*m/z*: 639.29), S4:16 (*m/z*: 667.28), S4:17 (*m/z*: 681.30), S3:22 (*m/z*: 737.40), and S4:24 (*m/z*: 779.41) and their glucose cognates (missing a C5 chain present on the furanose ring), G2:10 (*m/z*: 393.17), G3:11 (*m/z*: 421.17), G3:12 (*m/z*: 435.18), G2:17 (*m/z*: 491.29), and G3:19 (*m/z*: 533.30) are shown. (B) BIL6180 accumulates P-type acylsucroses with no detectable acylglucoses. ESI- mode LC-MS analysis of trichome extracts of BIL6180 are shown. Extracted ion chromatograms of S3:15 (*m/z*: 639.29), S3:22 (*m/z*:737.40), G3:15 (*m/z*: 477.23), and G3:22 (*m/z*: 575.34) are shown. (C) Introduction of *SpASFF1* driven by its endogenous promoter from chromosome 3 into BIL6180 is sufficient to cause accumulation of detectable G3:15 and G3:22 acylglucoses. ESI- mode LC-MS analysis of trichome extracts of a *proSpASFF1::SpASFF1* in a BIL6180 T$_2$ line is shown. Extracted ion chromatograms of S3:15 (*m/z*: 639.29), S3:22 (*m/z*:737.40), G3:15 (*m/z*: 477.23), and G3:22 (*m/z*: 575.34) are shown. Note: All *m/z* values correspond to the formate adducts of those acylsugars. c Acylglucose structure is inferred from collision induced dissociation-mediated fragmentation (Fig S3.8). All ESI- mode acylsugars were identified as formate adducts. When NMR-derived structural information is available, superscripts indicate acyl chain positions with R representing the pyranose ring, and R' representing the furanose ring.

In vitro assays supported the hypothesis that SpASFF1 accepts P-type, but not F-type acylsucroses as substrates. Initial attempts to express SpASFF1 fusion proteins in *E. coli* did not produce soluble protein. For this reason, recombinant His-tagged SpASFF1 was expressed using the *Nicotiana benthamiana* transient expression system (Peyret and Lomonossoff, 2013). The enzyme was tested with both P-type and F-type acylsucrose substrates purified from *S. pennellii asff1* and *S. lycopersicum* M82, respectively. Consistent with *in vivo* observations, SpASFF1 demonstrated hydrolytic activity with purified P-type S3:19 (4^{R4},5^{R2},10^{R3}) (*42*), yielding a compound with *m/z* consistent with a G3:19 (4,5,10) structure (Fig. 3.6A and Fig. S3.9). In contrast, SpASFF1 demonstrated no hydrolytic activity with F-type S3:22 (5^{R4},$5^{R3'}$,12^{R3}) (Ghosh et al., 2014) (Fig. 3.6B), suggesting that the presence of an acyl chain on the sucrose furanose ring prevents enzymatic hydrolysis. We further observed that SpASFF1 activity was undetectable with unmodified sucrose, while a commercially available yeast invertase hydrolyzed sucrose, but not S3:19 (Fig. S3.10). This SpASFF1 *in vitro* substrate specificity corroborates the *in vivo* results showing that acylglucoses only accumulate in lines containing P-type acylsucroses.
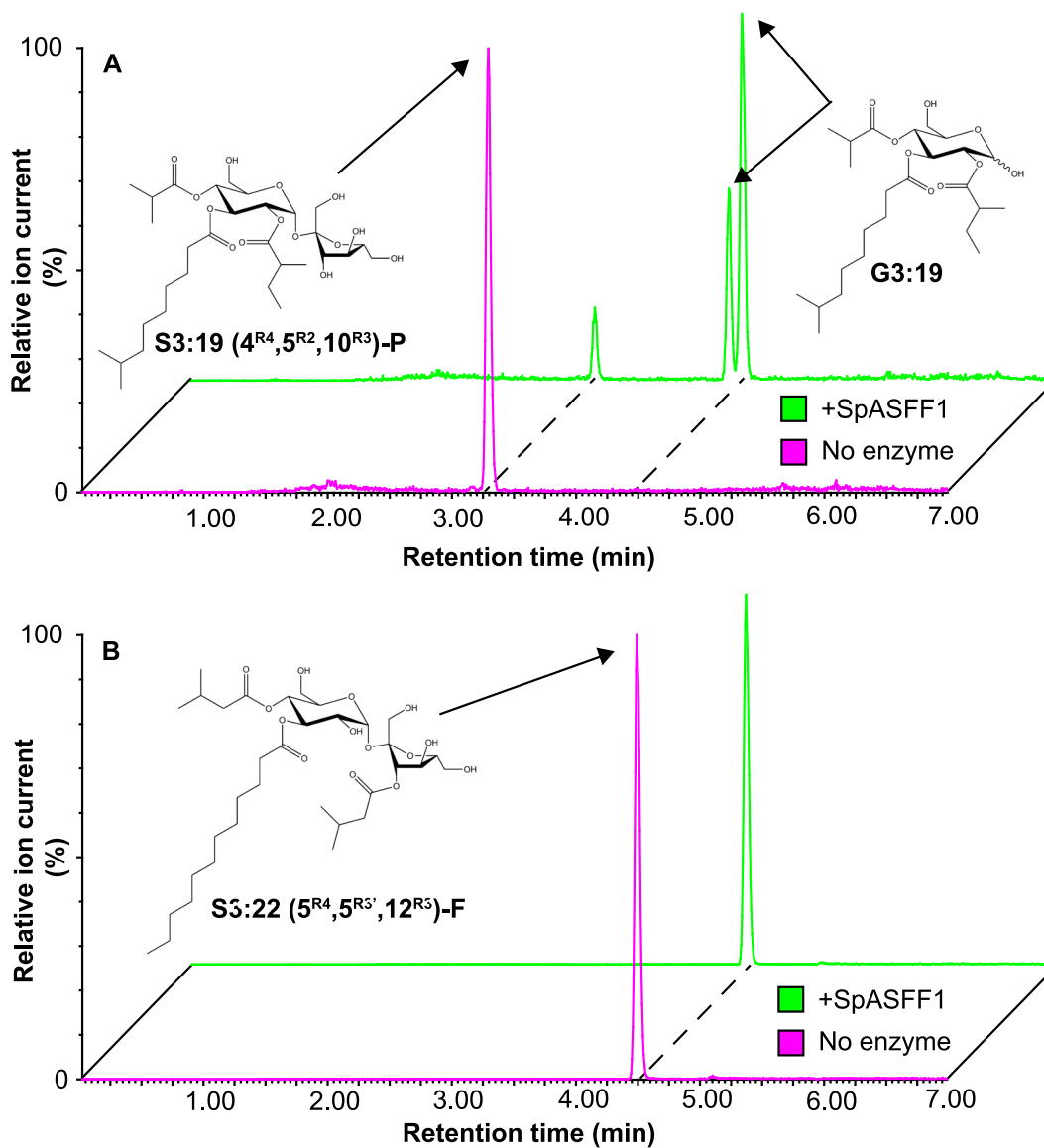
Figure 3.6. SpASFF1 cleaves a P-type S3:19 acylsucrose but not F-type S3:22 acylsucrose.

(A) LC-MS analysis of *in vitro* enzyme assay products indicates that SpASFF1 hydrolyzed P-type S3:19 (5^{R2},10^{R3},4^{R4}) acylsucrose yielding two compounds with $m/z = 533.3$. The two new products are consistent with an acylglucose with a G3:19 (4,5,10) configuration, which represent the α and β anomers of the acylglucose. (B) LC-MS analysis of *in vitro* assays with F-type S3:22 (12^{R3},5^{R4},$5^{R3'}$) acylsucrose indicate no hydrolysis products are generated by SpASFF1. Note: Acylglucose structure is inferred from collision induced dissociation-mediated fragmentation (Fig. S3.9). All ESI- mode acylsugars were identified as formate adducts. When NMR-derived structural information is available, superscripts indicate acyl chain positions with R representing the pyranose ring, and R' representing the furanose ring.

Discussion

The results described above show that *S. pennellii* LA0716 biocatalyzes acylglucoses from P-type acylsucroses via a previously uncharacterized trichome invertase (Fig. 3.6A), a homolog of the most venerable enzyme in the history of biochemistry, yeast invertase. The canonical GH32 enzyme was first characterized in the 1840's through studies of 'optical inversion' of cane sugar (sucrose) into a mixture of glucose and fructose. The enzyme was assayed two decades later, and its study by Maud Leonora Menten and Leonor Michaelis led to the theory of enzyme kinetics early in the 20th Century (Johnson and Goody, 2011). Since that time, other general metabolic activities were identified for diverse GH32 β-fructofuranosidases, including plant glycan biosynthesis, cell wall modification, and hormone metabolism (De Coninck et al., 2005; Minic, 2008).

Our results contrast the previously proposed direct biosynthesis of acylglucoses from UDP-glucose and free fatty acids (Ghangas and Steffens, 1993; Kuai et al., 1997; Li et al., 1999; Li and Steffens, 2000). Steffens and co-workers identified two glucosyltransferases and an SCPL acyltransferase from *S. pennellii* capable of generating 1-*O*-mono- and 1,2-*O*-di-acylglucoses *in vitro* (Ghangas and Steffens, 1993; Kuai et al., 1997; Li et al., 1999; Li and Steffens, 2000). The glycosyltransferases possessed differing specificity for coupling medium versus short acyl chains to glucose, proposed to be responsible for the different acyl chains conjugated to the acylglucose molecule *in vivo* (Kuai et al., 1997). This hypothetical biosynthetic route seems promising at first.

Multiple lines of evidence suggest that these enzymes are not involved in *S. pennellii* acylglucose biosynthesis. First, acylglucoses generated *in vitro* by these enzymes are structurally distinct from the 2,3,4-*O*-tri-acylglucoses detected from *S. pennellii* (Burke et al., 1987); the *in*

vitro products possessed two acyl chains instead of three and were acylated at position R1 of glucose. No further demonstrations were made using an increasingly acylated glucose, nor were triacylglucoses ever synthesized *in vitro*.

Next, comparative transcriptomic data suggest that the SCPL acyltransferase shows similar expression levels in *S. lycopersicum* M82 and *S. pennellii* LA0716, yet there are no acylglucoses detected in M82 (Koenig et al., 2013). Additionally, the SCPL acyltransferase described by Li et al. (2000) is encoded on chromosome 10 (*Solyc10g049210*), in a region not implicated in acylglucose accumulation in QTL mapping studies (Leckie et al., 2013). In contrast, QTLs linked to acylglucose accumulation in *S. pennellii* on chromosomes 4 and 11 include *SpASAT2* and *SpASAT3*, suggesting a connection between acylsucrose and acylglucose biosynthesis (Leckie et al., 2013; Fan et al., 2016).

Finally, transcriptomic analysis of high and low acylsugar-producing *S. pennellii* accessions found that expression of several acylsugar biosynthetic genes were positively correlated with acylsugar accumulation (Mandal et al., 2018). ASAT1-3, IPMS3, SpASFF1 (at the time, uncharacterized) were determined to be upregulated in the high acylsugar-producing lines. In that same comparison, one of the glycosyltransferases and the SCPL enzyme are not enriched in the trichomes of high acylsugar-producing accessions. All of this collectively suggests that the glycosyltransferase and SCPL are not involved in acylglucose biosynthesis.

As the acylsucroses and acylglucoses in *S. pennellii* differ only by the presence or absence of a furanose ring (Fig. 3.1A), we hypothesized that a glycoside hydrolase converts the *S. pennellii* acylsucroses (Schilmiller et al., 2016) to acylglucoses. Three glycoside hydrolase genes were identified in the third acylglucose-linked QTL on chromosome 3. These genes

represent members of glycoside hydrolase (GH) families 32, 35, and 47 (Table S2). Most characterized plant GH35 enzymes act as β-galactosidases while GH47 enzymes function as α-mannosidases in post-translational protein modification (Herscovics, 2001; Tanthanuch et al., 2008). Thus, these were not compelling candidates for cleavage of acylated sucrose substrates. Conversely, GH32 enzymes act on a variety of β-fructofuranosides in plants, including sucrose and fructans (De Coninck et al., 2005; Van Den Ende et al., 2009). Our results indicate that SpASFF1 is a 'derived' β-fructofuranoside, with an active site that can accommodate pyranose- but not furanose-acylated sucrose esters. Understanding the structural features that allow SpASFF1 to hydrolyze P-type acylsucroses could inform engineering of novel specialized metabolites in plants and microbes.

We identified the GH32 SpASFF1 β-fructofuranosidase as being necessary and sufficient for conversion of P-type acylsucroses into acylglucoses. The most direct evidence is that ablation of the *SpASFF1* gene using CRISPR gene editing led to acylsucrose-accumulating wild tomato *S. pennellii* LA0716 mutants with undetectable acylglucoses, showing that the enzyme is necessary for production of acylglucoses (Fig. 3.4). Multiple lines of genetic and biochemical evidence support the hypothesis that SpASFF1 uses P-type acylsucrose substrates. For example, no acylglucoses were detected in the F-type acylsucrose-producing introgression line IL3-5, having *SpASFF1* in the introgressed region (Fig. 3.5A, and Fig. S3.1). In contrast, transgenic trichome expression of the SpASFF1 invertase in the P-type acylsucrose-producing SpASAT2 and SpASAT3 double introgression line *S. lycopersicum* BIL6180 resulted in acylglucose accumulation (Fig. 3.5). Our *in vitro* assay results support the *in vivo* evidence that P-type acylsucroses are SpASFF1 substrates. *In vitro* assays with recombinant SpASFF1 demonstrated conversion of the purified P-type S3:19 ($5^{R2},10^{R3},4^{R4}$) to the cognate acylglucose G3:19 (4,5,10)

(Fig. 3.6A). In contrast, the enzyme was inactive against F-type S3:22 (12^{R3},5^{R4},$5^{R3'}$) (Fig. 3.6B) and did not hydrolyze unacylated sucrose (Fig. S3.10).

Taken together, these data indicate that *S. pennellii* acylglucose metabolism results from evolution of a three-gene epistatic system, where the innovation of P-type acylsucrose synthesis by modification of the core BAHD acyltransferases potentiated evolution of SpASFF1 to produce acylglucoses. Our results reveal that a member of the GH32 β-fructofuranosidase enzyme family acquired expression in the trichome glandular tip cell (Fig. 3.3) and the ability to cleave acylated sucrose (Fig. 3.6), which lead to an increase in the diversity of Solanaceae trichome specialized metabolites. This is a remarkable evolutionary innovation, where a member of an enzyme family long recognized as important in general metabolism was co-opted into specialized metabolism by the 'blind watchmaker' of evolution.

Acylsugar accumulation is widespread throughout the Solanaceae with occurrences in genera as distantly related as *Salpiglossis* and *Solanum*, sharing a last common ancestor > 30 Mya (Särkinen et al., 2013; Ghosh et al., 2014; Moghe et al., 2017). While acylsugars show wide structural variation in the number and length of acyl chains throughout the family, sucrose is the most prominent sugar core. Acylsucroses accumulate in genera whose lineages diverged < 20 Mya, such as *Solanum* and *Physalis* (Maldonado et al., 2006; Ghosh et al., 2014) but also accumulate in species representing earlier diverging lineages, including *Salpiglossis* and *Petunia* (Moghe et al., 2017; Nadakuduti et al., 2017). In addition, acyl chains are present on the furanose ring in at least some members of each of these genera, suggesting that accumulation of F-type acylsucroses evolved long ago.

Though apparently limited in distribution relative to acylsucroses, acylglucoses occur in diverse genera including *Solanum*, *Datura*, and *Nicotiana* (Burke et al., 1987; King and Calhoun, 1988; Matsuzaki et al., 1989). While acylglucose accumulation is common to species in both *Solanum* and *Nicotiana* – which diverged approximately 24 Mya – the differences in *SpASFF1* gene expression and SpASAT substrate specificity that facilitated acylglucose accumulation in *S. pennellii* arose in the ~7 million years since divergence from the last ancestor in common with *S. lycopersicum* (Nesbitt and Tanksley, 2002; Särkinen et al., 2013; Fan et al., 2017). This supports independent evolutionary origins of acylglucoses in distinct lineages. In the *Solanum* genus, P-type acylsucroses are a prerequisite for acylglucose accumulation. The predominance of F-type acylsucroses within the Solanaceae may explain the relative rarity of acylglucoses in the family. However, characterization of the ASAT enzymes responsible for acylsucrose biosynthesis in *Salpiglossis*, *Petunia*, and *Solanum* demonstrates multiple changes in enzyme substrate specificity throughout the evolutionary history of the acylsucrose pathway (Fan et al., 2016; Moghe et al., 2017; Nadakuduti et al., 2017). Plasticity of the acylsugar pathway may have caused the occurrence of P-type acylsucroses multiple times throughout evolutionary history. If so, this would provide independent opportunities for co-option of glycoside hydrolases into acylsugar pathways to produce acylglucoses. Are the enzymes that hydrolyze acylsucroses to yield acylglucoses restricted to the GH32 family or have other enzyme families evolved in different acylglucose-accumulating lineages? To determine whether and to what extent multiple origins of acylglucose biosynthesis share common features remains to be explored.

Over the past decade, discovery of pathways and enzymes of plant specialized metabolism has improved at an increasing rate. Before this time, taxonomic restriction of specialized metabolism biased deep analysis towards pathways found in model organisms: for

instance, glucosinolates in Arabidopsis, cyclic hydroxamic acids in maize and other well-studied grasses and isoflavonoids in Medicago and soybean. Dramatic improvements in the sensitivity and selectivity of MS- and NMR-based analytical methods helped broaden our knowledge of well-studied metabolic networks (Mathew and Padmanaban, 2013; Nagana Gowda and Raftery, 2017). In parallel, development of species-agnostic DNA sequencing and functional genomics screening tools (such as virus-induced gene silencing and genetic transformation), permitted rigorous correlation of *in vitro* activities and *in vivo* phenotypes. The rapid advancement of gene editing techniques using CRISPR-Cas on agriculturally-important and undomesticated species dramatically expands the specialized metabolism functional genomics toolkit. Not only do these methods allow direct tests of *in vivo* function, but also allows elimination of the T-DNA by simple genetic crossing. The removal of the T-DNA permits growing edited mutants in agricultural fields or common gardens with lower regulatory barriers. For example, the *spasff1* mutant lines help dissect the impacts of acylsucroses versus acylglucoses on the fitness of *S. pennellii* both in the greenhouse and field. Such studies could lead to crops with novel natural pesticides, broaden our understanding of the roles of specialized metabolites in mediating environmental interactions, and inform our understanding of the mechanisms underpinning specialized metabolic evolution.

Materials and Methods

Plant material

Seeds of *S. lycopersicum* M82 were obtained from the C.M. Rick Tomato Genetics Resource Center (TGRC; University of California, Davis, CA); seeds of IL3-5, BIL6180, and BIL6521 were obtained from Dr. Dani Zamir (Hebrew University of Jerusalem, Rehovot, Israel) (Ofner et al., 2016); seeds of *S. pennellii* LA0716 were generously provided by Dr. Martha Mutschler (Cornell University, Ithaca, NY). Seeds were treated with half-strength bleach for 30 minutes and rinsed three times in de-ionized water for 5 minutes prior to sowing on moist filter paper in Petri dishes. Upon germination, seedlings were transferred to soil. Young plants were grown in 9-cm pots in peat-based propagation mix (SunGro, Agawam, MA). *S. lycopersicum* and introgression lines were watered four times weekly with de-ionized water and supplemented once weekly with ½ strength Hoaglands solution; *S. pennellii* was watered once weekly with de-ionized water and supplemented once weekly with ½ strength Hoaglands solution. Plants used for analysis were grown in a growth chamber under a 16-h photoperiod (190 µmol m^{-2} s^{-1} photosynthetic photon flux density (PPFD)) with 28 °C day and 22 °C night temperatures set to 50% relative humidity. BIL lines used for crosses were grown in a soil mix consisting of four parts SureMix (Michigan Grower Products, Inc., Galesburg, MI) to one part sand in a greenhouse with a daytime maximum temperature of 30 °C and a nighttime minimum temperature of 16 °C; sunlight was supplemented with high pressure sodium bulbs on a 16/8 light/dark cycle. For seed production, *S. pennellii asff1* T$_0$ plants were grown in soil containing one part Canadian sphagnum (Mosser Lee Co., Millston, WI), one part coarse sand (Quikrete, Atlanta, GA), one part white pumice (Everwood Farm, Brooks, OR), and one part redwood bark (Sequoia Bark Sales, Reedley, CA) supplemented with 1.8 kg crushed oyster shell (Down to Earth Distributors

Inc., Eugene, OR), 1.8 kg hydrated lime (Bonide Products, Inc., Oriskany, NY), and 0.6 kg triple super phosphate (T and N Inc., Foristell, MO) per cubic meter.

Acylsugar analysis

Note: The acylsugar extraction interactive protocol is available in Protocols.io at

http://dx.doi.org/10.17504/protocols.io.xj2fkqe

Leaf surface acylsugars were extracted from single leaflets with 1 mL of a mixture of isopropanol (J.T. Baker, Phillipsburg, NJ):acetonitrile (Sigma-Aldrich, St. Louis, MO):water (3:3:2) with 0.1% formic acid and 1 µM telmisartan (Sigma-Aldrich, St. Louis, MO) as an HPLC standard. The leaf tissue was gently agitated on a rocker in this extraction solvent for 2 minutes. The extraction solvent was collected and stored in 2 mL LC-MS vials at -80 °C.

LC-MS samples (both enzyme assays and plant samples) were run on a Waters Acquity UPLC coupled to a Waters Xevo G2-XS QToF mass spectrometer. 10 µL of the acylsugar extracts were injected into an Ascentis Express C18 HPLC column (10 cm x 2.1 mm, 2.7 µm) (Sigma-Aldrich, St. Louis, MO), which was maintained at 40 °C. The LC-MS methods used these solvents: 10 mM ammonium formate, pH 2.8 as solvent A, and 100% acetonitrile as solvent B. Compounds were eluted using one of two gradients.

A 7-minute linear elution gradient consisted of 5% B at 0 minutes, 60% B at 1 minute, 100% B at 5 minutes, held at 100% B until 6 minutes, 5% B at 6.01 minutes and held at 5% until 7 minutes. A 21-minute linear elution gradient consisted of 5% B at 0 minutes, 60% B at 3 minutes, 100% B at 15 minutes, held at 100% B until 18 minutes, 5% B at 18.01 minutes and held at 5% B until 21 minutes.

The MS settings were as follows for negative ion-mode electrospray ionization: 2.00 kV capillary voltage, 100 °C source temperature, 350 °C desolvation temperature, 600 liters/h desolvation nitrogen gas flow rate, 35V cone voltage, mass range of m/z 50 to 1000 with spectra accumulated at 0.1 seconds/function. Three separate acquisition functions were set up to test different collision energies (0V, 15V, 35V).

The MS settings were as follows for positive ion-mode electrospray ionization: 3.00 kV capillary voltage, 100 °C source temperature, 350 °C desolvation temperature, 600 liters/h desolvation nitrogen gas flow rate, 35V cone voltage, mass range of m/z 50 to 1000 with spectra accumulated at 0.1 seconds/function. Three separate acquisition functions were set up to test different collision energies (0V, 15V, 45V). Lockmass correction was performed using leucine enkephalin as the reference compound for data acquired in both negative and positive ion mode.

Acylsugar quantification

To accurately quantify total acylsugars, samples were saponified before LC-MS analysis and sugar cores quantified with authentic isotopically labelled standards. A leaflet was immersed in 2 mL dichloromethane (VWR International, Radnor, PA) and 500 μL water with 30 s vortexing. After phase separation, 1 mL of the dichloromethane layer was removed to a borosilicate glass vial and evaporated to dryness under flowing air. Dried samples were dissolved in 1 mL acetonitrile with 0.1% formic acid for storage. 20 μL aliquots of acylsugar extracts were dried in 1.7 mL microcentrifuge tube using a SpeedVac and dissolved in 100 μL methanol. An equal volume of 3 N aqueous ammonia solution (Sigma-Aldrich, St. Louis, MO) was added, and the reaction was incubated in a sealed 1.5 mL microcentrifuge tube for 48 hrs in a fume hood. Before LC-MS analysis, samples were dissolved in 200 μL ammonium bicarbonate (pH 7-8) in

90% acetonitrile containing 0.5 µM $^{13}C_{12}$-sucrose and 0.5 µM $^{13}C_6$-glucose as internal standards and transferred to a 2 mL LC-MS vials. Compounds were analyzed on a *Waters ACQUITY TQD* Tandem Quadrupole UPLC/MS/MS system (Waters, Milford, MA). Ten microliters of the acylsugar extracts were injected into a Waters ACQUITY UPLC BEH amide column (2.1x100 mm, 1.7 µM), in a column oven with temperature of 40 °C with flow rate of 0.5 mL/minute. The LC-MS methods used 10 mM ammonium bicarbonate pH 8 in 50% acetonitrile as Solvent A and 10 mM ammonium bicarbonate pH 8 in 90% acetonitrile as Solvent B. The chromatography gradient was: 100% B at 0 minutes, 0% B at 5 minutes, 100% B at 5.01 minutes, held at 100% B until 10 minutes. Multiple-reaction monitoring (MRM) mode was operated to detect each sugar. For glucose: precursor ion, *m/z* 179; product ion, *m/z* 89; cone voltage, 16 V; collision energy, 10 V. For $^{13}C_6$-glucose: precursor ion, *m/z* 185; product ion, *m/z* 92; cone voltage, 16 V; collision energy, 10 V. For sucrose: precursor ion, *m/z* 341; product ion, *m/z* 89; cone voltage, 40 V; collision energy, 22 V. For $^{13}C_{12}$-sucrose: precursor ion, *m/z* 353; product ion, *m/z* 92; cone voltage, 40 V; collision energy, 22 V. Quantification of glucose and sucrose were conducted by standard curves with authentic glucose and sucrose standards (Sigma-Aldrich, St. Louis, MO). Glucose and sucrose standard solutions of 31.25, 62.5, 125, 250, and 500 µM in water were prepared and processed using the same protocol described for acylsugars above.

Acylsucrose purification

All purifications were performed using a Waters 2795 Separations Module (Waters, Milford, MA) and an Acclaim 120 C18 HPLC column (4.6 x 150 mm, 5 µm; Thermo Scientific, Waltham, MA) with a column oven temperature of 30 °C and flow rate of 1 mL/minute. The mobile phase consisted of water (Solvent A) and acetonitrile (Solvent B). Fractions were collected using a 2211 Superrac fraction collector (LKB Bromma, Stockholm, Sweden).

For purification of acylsucroses from *S. pennellii* LA0716, approximately 75 g fresh above-ground tissue of mature *S. pennellii asff1-1* was harvested into a 1-L glass beaker to which 500 mL 100% methanol was added. Tissue was stirred for 2 minutes and filtered through Miracloth (EMD Millipore, Billerica, MA) pre-wetted with methanol into a 1-L round bottom flask. Solvent was removed with a rotary evaporator in a water bath held between 35 and 40 °C and the residue dissolved in 5 mL acetonitrile. A 5-µL aliquot of this solution was diluted 1000-fold in 9:1 water/acetonitrile with 0.1% formic acid for chromatographic purification. The S3:19 compound was purified from 20 injections of 100 µL each using a linear elution gradient of 1% B at 0 minutes, 63% B at 10 minutes, 65% B at 30 minutes, 100% B at 35 minutes brought back to 1% B at 35.01 minutes and held at 1% B until 40 minutes. Eluted compounds were collected in 10-second fractions. Fraction collection tubes contained 333 µL 0.1% formic acid in water, and the S3:19 product eluted at 18-19 minutes.

For purification of the S3:22 acylsucrose from *S. lycopersicum* M82, approximately 75 g fresh above-ground tissue was harvested from mature plants into a 500-mL glass beaker to which 250 mL 100% methanol was added. The tissue was stirred for 2 minutes and filtered through Miracloth into a 1 L round-bottom flask. Methanol was removed with a rotary evaporator in a water bath held between 35 and 40 °C and residue dissolved in 5 mL acetonitrile. This solution was diluted 50-fold in 9:1 water/acetonitrile with 0.1% formic acid for further processing. The S3:22 compound was purified from 10 injections of 100 µL using a linear elution gradient of 1% B at 0 minutes, 50% B at 5 minutes, 70% B at 30 minutes, 100% B at 32 minutes and held at 100% until 35 minutes brought back to 1% B at 35.01 minutes and held at 1% B until 40 minutes. Eluted compounds were collected in 1 minute fractions. Fraction collection tubes contained 333 µL 0.1% formic acid in water, and the S3:22 product eluted at 7-8 minutes.

qPCR analysis

Tissue of 10-week-old *S. pennellii* LA0716 and *S. lycopersicum* M82 were harvested as follows: stems flash-frozen in liquid nitrogen and trichomes shaved into 1.5-mL microcentrifuge tubes with a clean razor blade. Trichomes and denuded stems were kept in liquid nitrogen and ground with plastic micropestles in 1.5-mL microcentrifuge tubes. RNA was extracted from ground trichomes and stems (six biological replicates for each species and tissue type) using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. For each sample, 250 ng of RNA as quantified using a Nanodrop 2000c (Thermo Fisher Scientific, Waltham, MA) was used to synthesize cDNA using SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA). qRT-PCR was carried out using SYBR Green PCR Master Mix on a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems, Warrington, UK) using the following cycling conditions: 48 °C for 30 minutes, 95 °C for 10 minutes, 40 cycles of 95 °C for 15 s and 60 °C for 1 minute followed by melt curve analysis. RT_ASFF_F and RT_ASFF_R primers were used to detect ASFF1 transcript; RT_EF-1a_F/R, RT_actin_F/R, and RT_ ubiquitin_F/R primers were used to detect transcripts of the *EF-1α*, *actin,* and *ubiquitin* genes, respectively (Table S3). For each biological replicate, relative levels of ASFF1 transcript were determined using the $\Delta\Delta C_t$ method (Pfaffl, 2001) and normalized to the geometric mean of EF-1α, actin, and ubiquitin transcript levels.

Genotyping of progeny of BIL6521 x BIL6180

DNA from the progeny of the cross between BIL6521 and BIL6180 was extracted from leaf material that were archived on FTA PlantSaver cards (GE Healthcare, Uppsala, Sweden) and purified according to the manufacturer's specifications. Extracted DNA from the FTA cards

were used for PCR amplification with GoTaq green mastermix to genotype the sample using

04g011460_Marker_Indel-F/R and ASFF_Chr3_Indel_002_F/R (Table S3).

DNA construct assembly

All Sanger DNA sequencing confirmation in this study was performed with the indicated

sequencing primers at the Research Technology Support Facility Genomics Core, Michigan State

University, East Lansing, MI. For *proSpASFF1*::*SpASFF1* ORF – (pK7WG), a 1.8 kb region of

the upstream region and open reading frame of *SpASFF1* was split into four amplicons using

four sets of primers: ASFF_001_F/R, ASFF 002_F/R, ASFF_003_F/R, and ASFF_004_F/R

(Table S3). The first and fourth amplicon contained adapters for assembly into pENTR-D-TOPO

that has been digested with NotI/AscI respectively. The construct was assembled using NEB

Gibson assembly according to manufacturer specifications (NEB, Ipswich, MA). The construct

was verified by Sanger sequencing using M13 Forward, T7 promoter primers, and cloning

primers. The insert was subcloned into pK7WG (Karimi et al., 2002) using LR clonase II

enzyme mix (Thermo Scientific, Waltham, MA) according to manufacturer instructions.

Presence of the insert was determined by colony PCR using ASFF_001F/R. Completed vectors

were transformed into *Agrobacterium* strain AGL0. Leaf material from recovered plants were

archived on FTA PlantSaver cards (GE Healthcare, Uppsala, Sweden) and genotyped by PCR

amplification with GoTaq green mastermix and pK7WG-Kan-F/R primers (Table S3)

For *proSpASFF1*::GFP/GUS – (pKGWFS7), a 1.8 kb region of the upstream region of

ASFF1 was amplified from *S. pennellii* LA0716 genomic DNA using the primers

ASFF_promoter_F1/R1 (Table S3). pENTR-D-TOPO was digested with NotI/AscI to linearize

the vector and create overhangs compatible for Gibson assembly. The amplicon also contained

adapters for insertion into pENTR-D-TOPO digested with NotI/AscI. Constructs were Sanger

sequenced using M13F/R primers in addition to the ASFF_promoter_F1/R1 primers. LR clonase

II mix was used to subclone the fragment into pKGWFS7 (Karimi et al., 2002). Construct was

transformed into Agl0 for plant transformation using the described protocol.

The CRISPR–ASFF1 vector was constructed as follows. CRISPR sgRNAs were designed

using the site finder toolset in Geneious® v10 (www.geneious.com). Two target sequences

located on the exon were selected for their high on-target activity scores, based on a published

algorithm (Doench et al., 2016), and low off-target scores against published *S. pennellii* genome

database (Hsu et al., 2013). Each sgRNA was obtained as a gBlock synthesized *in vitro* by the

method described by the manufacturer, IDT (www.idtdna.com) (Table S3) and subsequently

assembled with pICH47742::2x35S-5'UTR-hCas9(STOP)-NOST (Addgene #49771, kindly

provided by Dr. Sophien Kamoun, Sainsbury Lab, Norwich, UK) (Belhaj et al., 2013),

pICH41780 (Addgene plasmid # 48019) and pAGM4723 (Addgene plasmid # 48015, both gifts

from Dr. Sylvestre Marillonnet) (Weber et al., 2011) and pICSL11024 (Addgene plasmid #

51144, a gift from Dr. Jonathan D. Jones, Sainsbury Lab, Norwich, UK) using Golden Gate

Assembly. In short, the restriction–ligation reactions (20 µL) were set up by mixing 15 ng of

synthesized sgRNAs with 1.5 µL T4 ligase buffer (NEB), 320 U of T4 DNA ligase (NEB), 1.5

µL BSA (0.1 mg/mL, NEB), 8 U of *Bpi*I (Thermoscientific) and 100–200 ng of the intact

plasmids. The reactions were incubated at 37 °C for 30s, followed by 26 cycles (37 °C, 3

minutes; 16 °C, 4 minutes) and then incubated at 50 °C for 5 minutes and 65 °C for 5 minutes.

The ligated products were directly used to transform *E. coli* competent cells. Positive clones

were chosen based on colony PCR and sequenced at the MSU RTSF facility using the

pAGM4723_SeqF1, pAGM4723_SeqR1, pICSL11024_SeqF1, pICH47742CAS9_SeqF2,

pICH47742_SeqF1, pICH41780_SeqR1, and ASFF_SeqR primers (Table S3). The construct was

transformed into *S. pennellii* LA0716 using the plant transformation protocol described below.

Leaf material from recovered plants were archived on FTA PlantSaver cards (GE Healthcare,

Uppsala, Sweden) and genotyped by PCR amplification with ASFF_F/R, followed by Sanger

sequencing with ASFF_SeqR (Table S3).

For *spasff1* line transcript analysis, RNA was extracted from *spasff1-1-1//1-1-2* lines using

RNeasy plant mini kit according to the kit specifications (Qiagen, Venlo, Netherlands). RNA

was quantified using a Nanodrop 2000c (Thermofisher, Waltham, MA). 1 µg of RNA was used

for cDNA synthesis using Superscript II Reverse Transcriptase according to the manufacturer's

specifications. The primers, ASFF1_transcript_amp_01F/R (final concentration: 0.5 µM), were

used to amplify the region within the *ASFF1* CDS, which was cloned into pMINI-T 2.0 (NEB,

Ipswich, MA). T7 and SP6 promoter primers were used for Sanger sequence confirmation of the

inserts, and ClustalW was used for alignment of the transcripts

(https://www.ebi.ac.uk/Tools/msa/clustalo/).

Competent cell preparation and transformation of constructs into Agrobacterium

A single colony of AGL0 or LBA4404 *Agrobacterium* was inoculated into two 5 mL

cultures of YEP media (10 g yeast extract, 10 g Bacto peptone, and 5 g NaCl per liter, pH 7) with

Rifampicin (50 µg/mL). Cultures were incubated in 18x150 mm borosilicate glass test tubes with

foam plugs overnight at 30 °C, shaking at 200 rpm. 190 mL of LB was inoculated with 10 mL of

the overnight cultures in an autoclaved 500 mL Erlenmeyer flask. Cultures were grown in a

shaking incubator (200 rpm) at 30 °C to $OD_{600} = 1.0$, incubated on ice for 10 minutes, and

centrifuged at 4 °C in 50 mL conical tubes at 3,200*g* for 5 minutes. Pellets were resuspended in 1

mL of sterile 20 mM CaCl$_2$ and 100 µL aliquots, dispensed into sterile, pre-chilled 1.7 mL microcentrifuge tubes, snap frozen using liquid nitrogen and stored at -80 °C

For *Agrobacterium* transformation, 1 µg of construct DNA purified using an Omega EZNA plasmid DNA mini kit I (Omega Bio-Tek, Norcross, GA) as added to the frozen *Agrobacterium* aliquots on ice. Cells were thawed in a 37 °C water bath for 5 minutes, mixed well by flicking and snap frozen in liquid nitrogen. Cells were thawed, and 1 mL of YEP added to the tube. The transformations were incubated at 28 °C and 200 rpm for 4 hours. Cells were centrifuged at 17,000g for 30 seconds, the supernatant decanted and the cell pellet resuspended in 100 µL of fresh YEP. The cell pellet was resuspended and the entire suspension was plated onto an LB plate with Spectinomycin (100 µg/mL).

Presence of the insert containing vector was verified by colony PCR. Colonies were collected with a pipette tip and resuspended in 20 µL of sterile water. 2 µL of the cell suspension was added to a PCR tube with a reaction with a final volume of 25 µL. GoTaq green mastermix (2x) was used for the colony PCR according to the manufacturer's specifications (Promega, Madison, WI). Primers (0.4 µM final concentration) pertaining to the insert were used for amplification.

Plant transformation

In all cases, Petri plates containing plant tissue were sealed with a single layer of micropore paper tape (3M, Maplewood, MN). Transformation of *S. lycopersicum* and *S. pennellii* LA0716 was performed using AGL0 using a modification of published protocols (*59, 60*). 50-60 seeds were incubated in 40% bleach, agitating for 5 minutes. Seeds were rinsed six times, each with 40 mL of sterile ddH$_2$O with 5 minutes of rocking and decanting of wash

solution. A flame-sterilized spatula was used to distribute the seeds onto the surface of 1/2x

MSO medium (McCormick, 1991) in a PhytaTray II (Sigma-Aldrich, St. Louis, MO). Containers

were incubated at 25 °C on a 16/8 light/dark cycle with a light intensity of 70 µmol m^{-2} s^{-2} PPFD.

At day eight for *S. lycopersicum* or day 11 for *S. pennellii* LA0716, the seedlings were

removed from the 1/2x MSO medium jar. The hypocotyl and radicle were excised and discarded.

The cotyledon explant was placed on a sterile Petri dish. 1-2 mm was removed from the base and

tip of the cotyledon. An autoclaved piece of Whatman #1 filter paper (GE Healthcare, Uppsala,

Sweden) was placed on the surface of a sterile D1 media plate (McCormick, 1991) on which the

cotyledons were placed adaxial side up. Approximately 100 explants were added per plate. The

plates were placed in the same conditions for two days until day 10.

For co-cultivation, the Agrobacterium containing the construct was streaked out onto an

LB plate containing the appropriate antibiotic. A single colony was inoculated into a 25 mL LB

culture with the same antibiotic plus Rifampicin (50 mg/L) in a 250 mL Erlenmeyer flask. The

culture was incubated at 30 °C in a shaking incubator (225 rpm) for 2 days. The culture was

transferred to a sterile 50 mL conical tube and centrifuged at 3,200*g* for 10 minutes at 20 °C. The

supernatant fluid was decanted and 10 mL of MSO media (McCormick, 1991) was added to the

tube (with no pellet resuspension). The cell pellet was centrifuged at 2,000*g* for 5 minutes and

this washing step was then repeated. The cell pellet was re-suspended in 10-20 mL of MSO

liquid media. Absorbance of the culture was measured at 600 nm. The suspension was diluted

with MSO to OD$_{600}$ = 0.5. Acetosyringone dissolved in DMSO was added at a final

concentration of 375 µM and 5 mL of the Agrobacterium suspension pipetted onto the

cotyledons on the plate and incubated with swirling at room temperature for 10 minutes, at which

point the excess culture was pipetted off. Using a scalpel, cotyledons were transferred to a fresh

D1 medium plate containing autoclaved Whatman paper. Approximately 50 cotyledons per plate were placed abaxial side up. Plates were incubated at 24 °C for 2 days with a 16/8 day-night cycle at 70 μmol m^{-2} s^{-2} PPFD.

For transgenic callus selection, two days after cocultivation, the cotyledons were transferred directly onto sterile 2Z media plates (Fillatti et al., 1987) containing 100 μg/mL kanamycin and 200 μg/mL timentin (no filter paper). Explants were placed abaxial side up with 20-30 cotyledons per plate. Plates were incubated at the same growth conditions for 10 days. Cotyledons were then transferred to a sterile Petri dish and, using a scalpel, calluses were cut and then placed onto fresh 2Z media plates with the same selection. Subsequently, explants were transferred to new 2Z plates every two weeks. Throughout the process, dying tissue was removed, and growing tissue was placed on the media. Five to eight weeks after cocultivation, shoots were harvested from the explants and placed into Phytatray II (Sigma-Aldrich, St. Louis, MO) containing 100 mL of MSSV media (Fillatti et al., 1987) supplemented with Timentin (100 μg/mL), Kanamycin (50 μg/mL), and Indole-3 butyric acid (1 μg/mL). MSSV-containing Phytatrays were incubated at the same growth conditions (16/8 at 70 μmol m^{-2} s^{-2} PPFD). Shoots were monitored for leaf and root production, and shoots with roots and leaves were placed into pots containing RediEarth soil. Flats were covered with a plastic dome in the same growth conditions. Domes were removed from flats after three to four days.

Transient expression and purification of SpASFF1 protein

The ASFF1 CDS was amplified from *S. pennellii* LA0716 trichome cDNA using the ASFF_F and ASFF_R primers (Table S3) and cloned into the pGEM backbone using the pGEM-T Easy cloning kit (Promega, Madison, WI). The ASFF1 CDS was subsequently re-amplified

with the (pEAQ-HT)-ASFF-His_F and (pEAQ-HT)-ASFF-His_R primers (Table S3) to add adapters for Gibson assembly. The resulting PCR product was transferred to pEAQ-HT vector (Peyret and Lomonossoff, 2013) previously digested with NruI-HF and SmaI restriction enzymes (New England Biolabs, Ipswich, MA) using 2x Gibson Assembly master mix (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions to create an expression clone coding for the full-length protein with a C-terminal 6x His tag (ASFF1-HT-pEAQ). The completed vector was subsequently transformed into LBA4404 cells as described above. For transient expression, an *A. tumefaciens* LBA4404 strain carrying the ASFF1-HT-pEAQ construct was streaked onto LB agar containing 50 µg/mL rifampicin and 50 µg/mL kanamycin and incubated for 3 days at 28 °C. Single colonies were used to inoculate 250-mL Erlenmeyer flasks containing 50 mL YEP medium with 50 µg/mL rifampicin and 50 µg/mL kanamycin; cultures were incubated at 28 °C and 300 rpm overnight. Cultures were harvested by centrifugation at 800g and 20 °C for 20 minutes. Supernatant was discarded and the resulting loose pellet resuspended in 50 mL of buffer A (10 mM 2-ethanesulfonic acid (MES; Sigma-Aldrich, St. Louis, MO) pH 5.6, 10 mM MgCl$_2$). This cell suspension was centrifuged at 800g and 20 °C for 20 minutes and the resulting pellet resuspended to a final OD$_{600}$ = 1.0 with buffer A. A 200 mM solution of acetosyringone (Sigma-Aldrich, St. Louis, MO) dissolved in DMSO was added to the suspension at a final concentration of 200 µM and the suspension incubated at room temperature with gentle rocking for 4 h. This suspension was infiltrated into fully expanded leaves of six-week-old *Nicotiana benthamiana* plants using a needle-less 1-mL tuberculin syringe. Plants were grown under 16-h photoperiod (70 µmol m^{-2} s^{-1} PPFD) and constant 22 °C set to 70% relative humidity. At 8 days post-infiltration, 28 g infiltrated leaves were harvested, de-veined, and flash-frozen in liquid nitrogen. Tissue was powdered under liquid nitrogen with mortar and pestle and

added to 140 mL ice-cold buffer B (25 mM 3-[4-(2-hydroxyethyl)piperazin-1-yl]propane-1-sulfonic acid (EPPS) pH 8.0, 1.5 M NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA) with 2 mM dithiothreitol (DTT), 1 mM benzamidine, 0.1 mM phenylmethansesulfonylfluoride (PMSF), 10 µM *trans*-epoxysuccinyl-L-leucylamido(4-guanidino)butane (E-64), and 5% (w/v) polyvinylpolypyrrolidone (PVPP); all reagents were obtained from Sigma-Aldrich, St. Louis, MO except DTT obtained from Roche Diagnostics, Risch-Rotkreuz, Switzerland). The mixture was stirred for 4 h at 4 °C, filtered through six layers of Miracloth and centrifuged at 27,000g, 4 °C for 30 minutes. The supernatant was decanted and passed through a 0.22-µm polyethersulfone filter (EMD Millipore, Billerica, MA) before being loaded onto a HisTrap HP 1-mL affinity column and eluted using a gradient of 10 to 500 mM imidazole in buffer B using an ÄKTA start FPLC module (GE Healthcare, Uppsala, Sweden). Fractions of eluant from the column were analyzed by SDS-PAGE and the presence of ASFF1-HT confirmed by immunoblot using the BMG-His-1 monoclonal antibody (Roche, Mannheim, Germany) to detect His-tagged proteins. Purified ASFF1-HT was subsequently transferred to 100 mM sodium acetate pH 4.5, 50% glycerol using a 10DG desalting column (Bio-Rad, Hercules, CA). Protein was quantified against a standard curve of bovine serum albumin (Thermo Fisher Scientific, Waltham, MA) using a modified Bradford reagent (Bio-Rad, Hercules, CA) according to the manufacturer's instructions.

Enzyme assays

For activity assays, 100 ng ASFF1-HT or 1 µg *Saccharomyces cerevisiae* invertase (Cat. No. I4504, Grade VII, Sigma-Aldrich, St. Louis, MO) and 0.1 nmol F- or P-type acylsucrose or 10 nmol sucrose (Sigma-Aldrich, St. Louis, MO) were added to 30 µL 50 mM sodium acetate, pH 4.5 in 250-µL thin-wall PCR tubes. Reactions were incubated for 1 h at 30 °C and stopped by

addition of 60 µL 1:1 acetonitrile/isopropanol containing 1.5 µM telmisartan as internal standard and centrifuged 10 minutes at 16,000*g* to remove precipitated protein. The supernatant was transferred to 2-mL autosampler vials with 250-µL glass inserts and analyzed by LC-MS as described above.

Statistical Analysis

All statistical analysis was performed using the stats R package (R Core Team, 2017). One-way analysis of variance (ANOVA) was executed on acylsugar data using the "aov" command. Between- and within-group variances were determined using the sum-of-squares values obtained from ANOVA; these values were subsequently used to determine the power of the ANOVA using the "power.anova.test" function. Post-hoc analysis by Tukeys honestly significant difference (HSD) mean-separation test was executed using the "TukeyHSD" command with the results of one-way ANOVA as input. Welch two sample *t*-tests were executed on transcript abundance data using the "t.test" command. The power of these analyses were determined using the "power.t.test" function.

APPENDIX
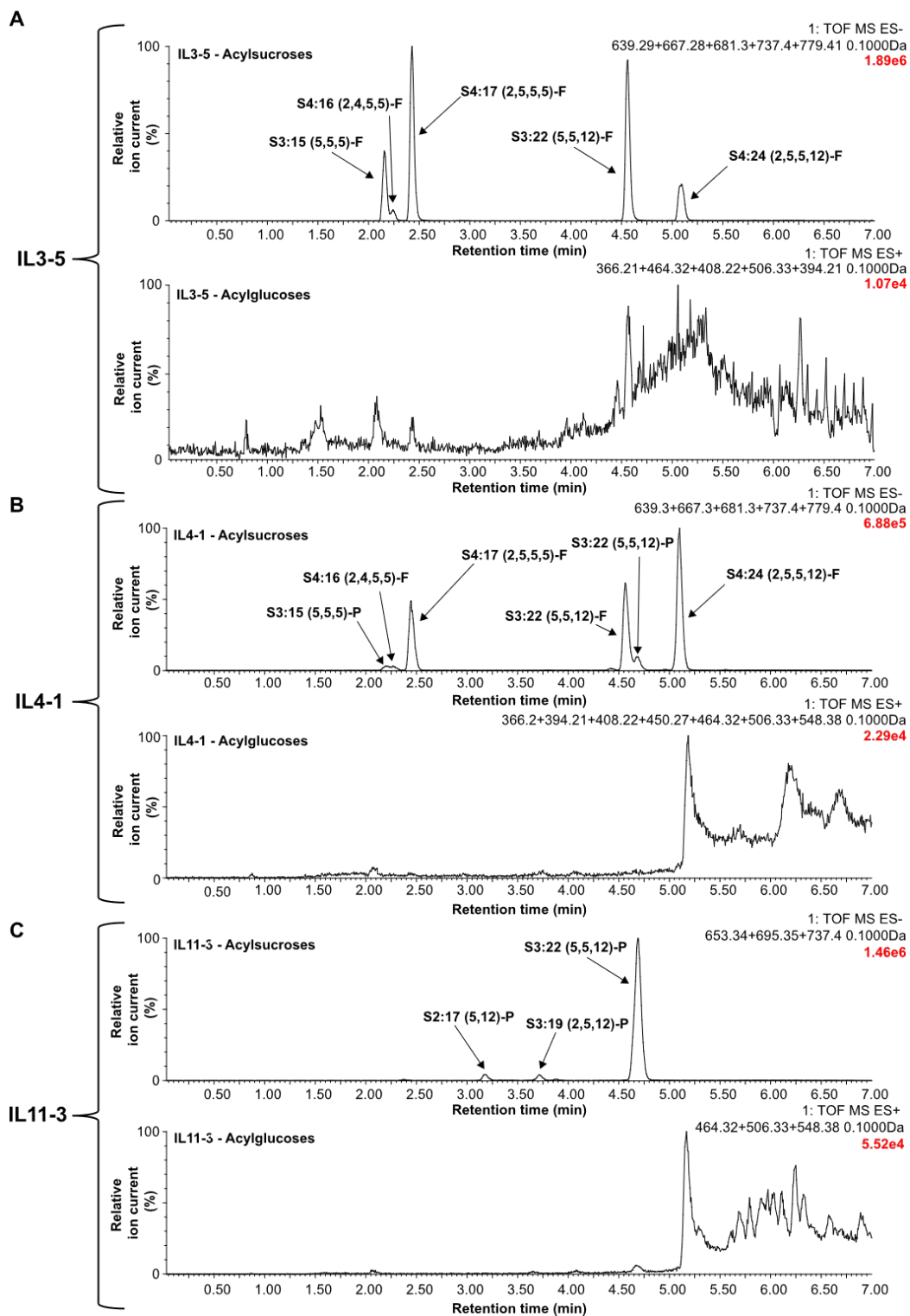
Figure S3.1. Acylglucoses are not detected in trichome extracts of IL3-5, 4-1, or IL11-3.

Figure S3.1 (cont'd)

LC-MS analysis using ESI- and ESI+ mode was used to detect acylsucroses and acylglucoses,respectively. Extracted ion chromatograms for ammonium adducts of expected possible acylglucoses showed no detectable peaks in any of the tested introgression lines (IL3-5, IL4-1, or IL11-3). The major acylsucroses identified were as follows: (A) For IL3-5, the major acylsugars identified were formate adducts of S3:15 (*m/z*: 639.28), S4:16 (*m/z*: 667.28), S4:17 (*m/z*: 681.30), S3:22 (*m/z*: 737.40), and S4:24 (*m/z*: 779.41). (B) For IL4-1, the major acylsugars present that were previously identified are: S3:15 (5,5,5)-P (*m/z*: 639.28), S4:16 (2,4,5,5) (*m/z*: 667.28), S4:17 (2,5,5,5) (*m/z*: 681.30), S3:22 (5,5,12) (*m/z*: 737.40), S3:22 (5,5,12)-P (*m/z*:737.40), and S4:24 (2,5,5,12) (*m/z*: 779.41). (C) For IL11-3, the major acylsugars present that were previously identified are: S2:17 (5,12) (*m/z*: 653.34), S3:19 (2,5,12) (*m/z*: 695.35), and S3:22 (5,5,12) (*m/z*:737.40). Note: The possible acylglucoses masses depended on the acylsucroses present, but all acylglucoses searched for were: G2:10 (5,5) (*m/z*: 366.21), G3:11 (2,4,5) (*m/z*: 394.21), G3:12 (2,5,5) (*m/z*: 408.22), G2:17 (5,12) (*m/z*: 464.32), G3:19 (2,5,12) (*m/z*: 506.33), G3:15 (5,5,5) (*m/z*: 450.27), G3:22 (5,5,12) (*m/z*: 548.38). The 7 minute method was used for this LC-MS analysis, which is described in the Methods (mass window: 0.1Da). Chromatograms are scaled as 0-100% with 100% representing the ion current value listed in the upper right hand corner of the chromatograph (i.e., 1.89e6 for IL3-5 acylsucroses, panel A). All ESI- mode acylsugars were identified as formate adducts, while all ESI+ mode acylsugars were identified as ammonium adducts.
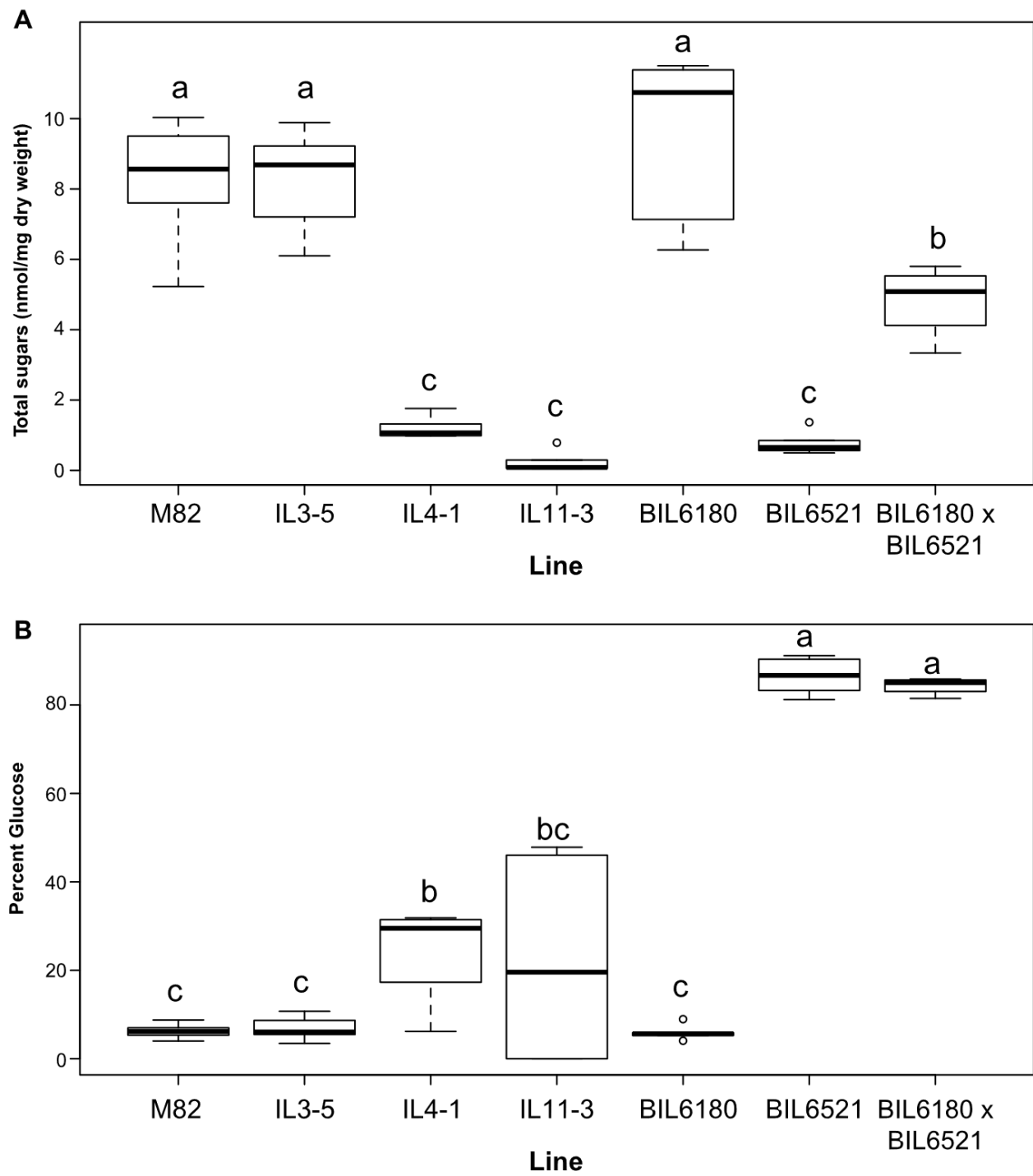
Figure S3.2. Quantification of acylsugars in *S. lycopersicum* M82 and breeding lines containing *S. pennellii* LA0716 introgressions.

Figure S3.2 (cont'd)

(A) Total acylsugar accumulation quantified as the sum of sucrose and glucose from saponified acylsugar extracts. (B) Percentage of saponified sugars in acylsugar extracts detected as glucose. Treatments that do not share a letter are significantly different from one another ($p < 0.05$; one-way ANOVA, Tukey's Honestly Significant Difference mean-separation test). Whiskers represent minimum and maximum values less than 1.5 times the interquartile range from the 1st and 3rd quartiles, respectively. Values outside this range are represented as circles; $n = 6$ for all lines except BIL6180 × BIL6521 F2 plants: $n = 4$.
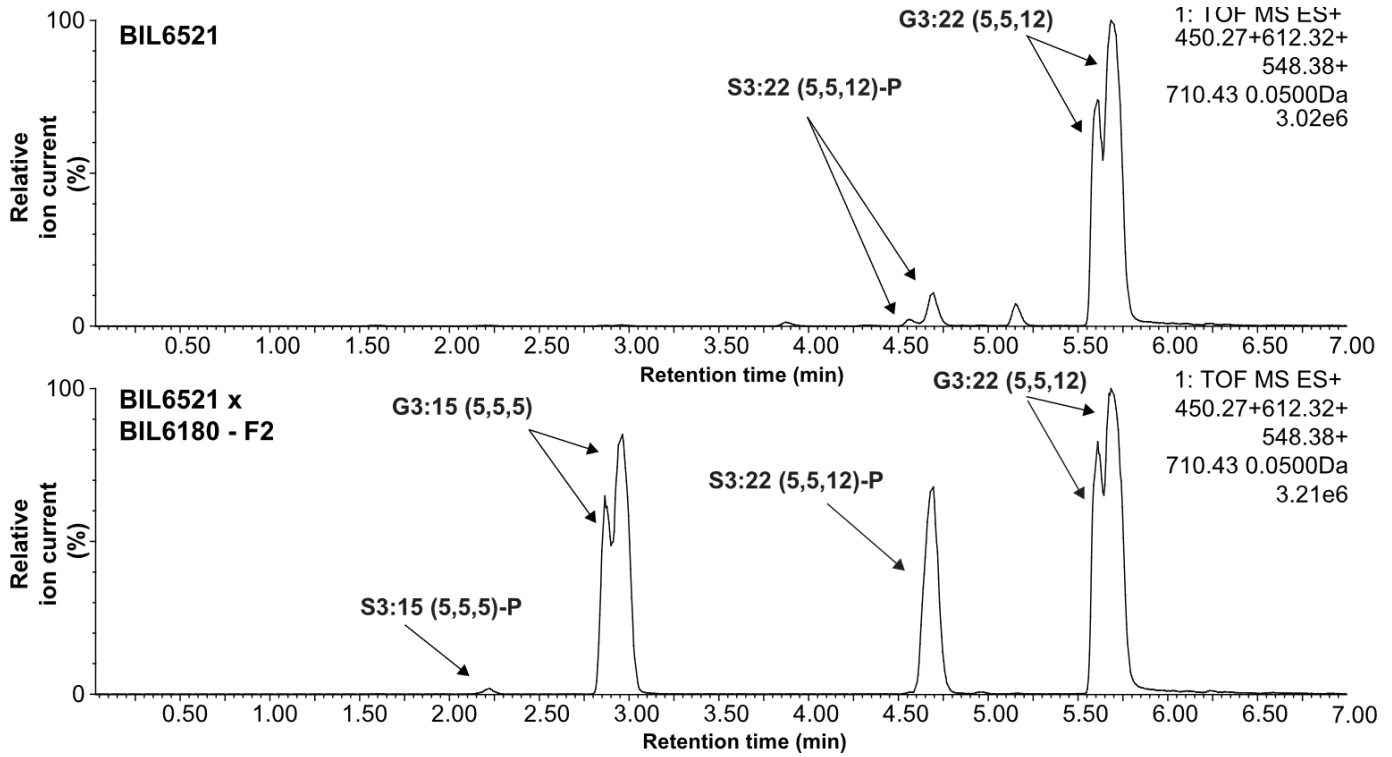
Figure S3.3. Comparison of major acylsugars in BIL6521 and BIL6521 x BIL6180 F2 progeny using LC-MS.

ESI+ mode LC-MS of trichome extracts reveals that the F2 progeny of BIL6521 crossed to BIL6180 (genotyped as heterozygous for Chr. 3 and 4 introgression regions) shows an increase in the short chain acylglucose, G3:15 compared to BIL6521 alone. Extracted ion chromatograms of NH_4^+ adducts of S3:15 (5,5,5) (*m/z*: 612.32) , G3:15 (5,5,5) (*m/z*: 450.27), S3:22 (5,5,12) (*m/z*: 710.43), and G3:22 (5,5,12) (*m/z*:548.38) in BIL6521 and BIL6521 x BIL6180 lines (mass window: 0.05 Da). Samples were run on the 7 minute method described in the method section. The progeny were genotyped as described in the Methods section. All ESI+ mode acylsugars were identified as ammonium adducts.
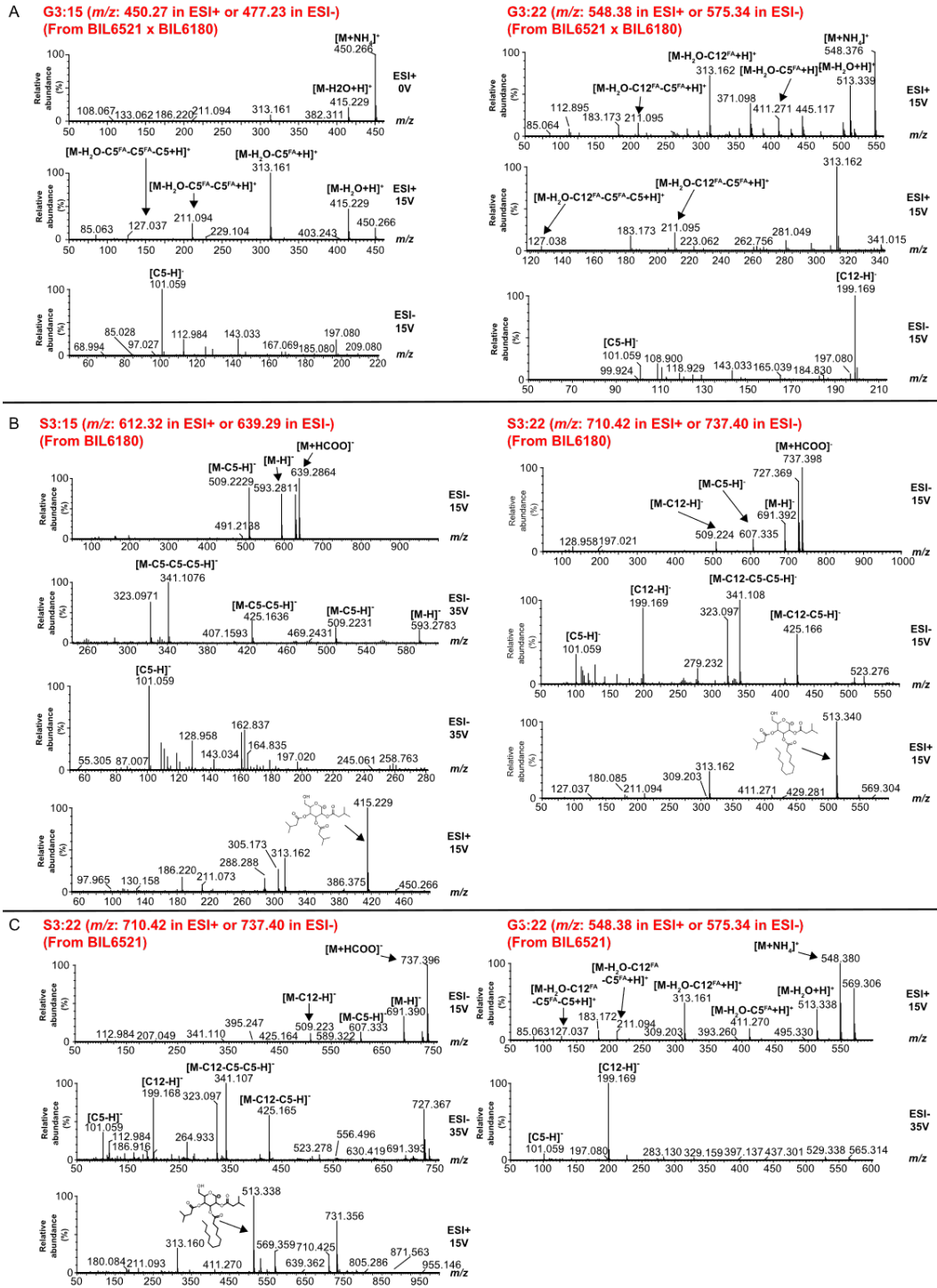
Figure S3.4. Mass spectra of major acylsugars S3:15, S3:22, G3:15, and G3:22 from BIL6521 x BIL6180 - F2 lines, BIL6180, and BIL6521.

Figure S3.4 (cont'd)

Triacylglucoses fragment in either positive or negative ion mode by losing the first two acyl chains as neutral fatty acids followed by the third acyl chain lost as aliphatic ketene (R=C=O). In negative ion mode, fatty acid anions are also seen when the charge stays with the fatty acid fragment. Together these losses allow for determination of the length of acyl chains attached to the acylglucose. Acylsucroses fragment in negative ion mode with neutral losses of aliphatic ketenes. In positive ion mode, fragmentation of ammonium adducts of acylsucroses results in cleavage of the glycosidic linkage with the most stable (and most abundant) ion fragment coming from the charge staying on the furanose ring fragment. When no acyl chains are present on the furanose ring, the most abundant fragment ions are from the pyranose ring and further fragmentation results from neutral loss of fatty acids. (A) Fragmentation of G3:15 and G3:22 from BIL6521 x BIL6180. Fragmentation of G3:15 in ESI+ mode (0 and 15V) results in the loss of two C5 fatty acids, followed by loss of a C5 ketene. Fragmentation of G3:22 in ESI+ mode (15V) results in the loss of C12 and C5 fatty acid followed by loss of a C5 ketene. The higher collision energy at 15V ESI- reveals the presence of a C5 (*m/z:* 101.06*)*, or C5 (*m/z:* 101.06) and C12 (*m/z*: 199.17) fatty acids for G3:15 and G3:22 respectively. (B) Fragmentation of S3:15 and S3:22 from BIL6180. Fragmentation of S3:15 in ESI- mode (15 and 35V) is characterized by the loss of 3 C5 ketenes. Fragmentation of S3:22 in ESI- mode (15V) is characterized by the loss of one C12 ketene, and two C5 ketenes. C5 (*m/z*: 101.06) or C5 (*m/z*: 101.06) and C12 (*m/z*: 199.17) fatty acids are present in ESI- mode for S3:15 and S3:22 respectively. ESI+ mode (15V) of these two acylsugars reveals the presence of three C5 chains or two C5 chains and one C12 on the pyranose ring of sucrose of S3:15 and S3:22 respectively. (C) Fragmentation of S3:22 and G3:22 from BIL6521. The fragmentation of S3:22 in ESI- mode (15 and 35V) is characterized by the loss of one C12 and two C5 ketenes. ESI- mode (35V) also reveals the presence of C5 (*m/z*: 101.06) and C12 (*m/z*: 199.17) fatty acids. ESI+ mode (15V) fragmentation reveals all three acyl chains are present on the pyranose ring. Fragmentation of G3:22 in ESI+ mode (15V) is characterized by the loss of a C12 and C5 fatty acid, followed by the loss of a C5 ketene. ESI- mode (35V) reveals the presence of C5 (*m/z*: 101.06) and C12 (*m/z*: 199.17) fatty acids. All fragmentation in this figure was obtained using collision-induced dissociation of the acylsugars. All ESI- mode acylsugars were identified as formate adducts, while all ESI+ mode acylsugars were identified as ammonium adducts.

```
CLUSTAL O(1.2.4) multiple sequence alignment

                                                                    ↓
LA0716        ATATATCGCACTGCTTATCATTTTCAACCTGCCAAGTATTGGATGAATGgtaatatcagt
spasff1-1-1   ATATATCGCACTGCTTATCATTTTCAACCTGCCAAGTATTGGATGAATGGTAATATCAGT

LA0716        cattgattaatcaaaaagagaagaaaataaatagatgatcatttctatcttgtgctatat
spasff1-1-1   CATTGATTAATCAAAAAGAGAAGAAAATAAATAGATGATCATTTCTATCTTGTGCTATAT

LA0716        aatcctaattctttttgttg-----ttttgt-gttgtcaatggtgatcaaatttacaatt
spasff1-1-1   AATCCTAATT------GTTGCTAGATTTTGTA-----------------------------
                                        ↑

LA0716        gttggttttcgttccgcagATCCTAATGGtaagtttctcaagttgctttgtgtttggtac
spasff1-1-1   ------------------------------------------------------------

LA0716        gacaaaagtcattcatgaaattcgttatgcatgcaggACCAGTGATATACATGGGGATTT
spasff1-1-1   ------------------------------------------------------------

LA0716        ACCATTTGTTCTATCAGTACAATCCTTACGATGCACAAACAGGTAACATTGTATGGGGCC
spasff1-1-1   ------------------------------------------------------------

LA0716        ATTCCACATCCACTGATCTCGTCAACTGGACTCCTCAGCCGCCGGCGCTTCTCCCATCAG
spasff1-1-1   ----------------------------------------------------CCCATCAG

LA0716        AGCCTTACGATTTCAAAGGCTGTTTTACCGGTTCTACAACTCTTCTCTCTGGCGGAAAAC
spasff1-1-1   AGCCTTACGATTTCAAAGGCTGTTTTACCGGTTCTACAACTCTTCTCTCTGGCGGAAAAC

LA0716        CGGCAG
spasff1-1-1   CGGCAG


                                                                    ↓
LA0716        ATATATCGCACTGCTTATCATTTTCAACCTGCCAAGTATTGGATGAATGgtaatatcagt
spasff1-1-2   ATATATCGCACTGCTTATCATTTTCAACCTGCCAAGTATTGGATGAATGGTAATATCAGT

LA0716        cattgattaatcaaaaagagaagaaaataaatagatgatcatttctatcttgtgctatat
spasff1-1-2   CATTGATTAATCAAAAAGAGAAGAAAATAAATAGATGATCATTTCTATCTTGTGCTATAT

LA0716        aatcctaattcttttttgttgttttgtgttgtcaatggtgatcaaatttacaattgttggt
spasff1-1-2   AATCCTAATTCTTTTTGTTGTTTTGTGTTGTCAATGGTGATCAAATTTACAATTGTTGCT

LA0716        ----------tttcgttccgcagATCCTAATGGtaagtttctcaagttgctttgtgtttg
spasff1-1-2   AGATTTTGTA--------------------------------------------------
                        ↑

LA0716        gtacgacaaaagtcattcatgaaattcgttatgcatgcaggACCAGTGATATACATGGGG
spasff1-1-2   ------------------------------------------------------------

LA0716        ATTTACCATTTGTTCTATCAGTACAATCCTTACGATGCACAAACAGGTAACATTGTATGG
spasff1-1-2   ------------------------------------------------------------

LA0716        GGCCATTCCACATCCACTGATCTCGTCAACTGGACTCCTCAGCCGCCGGCGCTTCTCCCA
spasff1-1-2   --------------------------------------------------------CCCA

LA0716        TCAGAGCCTTACGATTTCAAAGGCTGTTTTACCGGTTCTACAACTCTTCTCTCTGGCGGA
spasff1-1-2   TCAGAGCCTTACGATTTCAAAGGCTGTTTTACCGGTTCTACAACTCTTCTCTCTGGCGGA

LA0716        AAACCGGCAG
spasff1-1-2   AAACCGGCAG
```

Figure S3.5. Mutated genomic sequence of three homozygous *spasff1* CRISPR/Cas9 lines.

Figure S3.5 (cont'd)

```
LA0716      TTCTATCAGTACAATCCTTACGATGCACAAACAGGTAACATTGTATGGGGCCATTCCACA
spasff1-2   TTCTATCAGTACAATCCTTACGATGCACAAACAGGTAACATTGTATGGGGCCATTCCACA

LA0716      TCCACTGA-TCTCGTCAACTGGACTCCTCAGCCGCCGGCGCTTCTCCCATCAGAGCCTTA
spasff1-2   TCCACTGAATCTCGTCAACTGGACTCCTCAGCCGCCGGCGCTTCTCCCATCAGAGCCTTA

LA0716      CGATTTCAAAGGCTGTTTTACCGGTTCTACAACTCTTCTCTCTGGCGGAAAACCGGCAGT
spasff1-2   CGATTTCAAAGGCTGTTTTACCGGTTCTACAACTCTTCTCTCTGGCGGAAAACCGGCAGT

LA0716      TCTCTACACCGGCTTAGATTTCTCCGGTATCCAGGTACAAAATCTAGCAGTCCCCAAAAA
spasff1-2   TCTCTACACCGGCTTAGATTTCTCCGGTATCCAGGTACAAAATCTAGCAGTCCCCAAAAA

LA0716      TGTAGCTGA
spasff1-2   TGTAGCTGA
            *********
```

Large insertion-deletions on *spasff1-1-1* and *spasff1-1-2* are shown by blue dashes and letters. Both mutations expand between exons (sequences shown in upper case and introns in lower case). We observed incorrect splicing events in transcripts in both mutant lines: arrows indicate splicing positions and extended exons in mutant lines are highlighted in grey. The mis-splicing events are predicted to result in premature stop codons, which are highlighted in black. Mutant *spasff1-2* contains a single base pair insertion (blue letter) at one of the CRISPR/Cas9 target region (both regions are highlighted in yellow). This is predicted to cause a frame shift with resultant premature stop codon 180 nucleotides downstream (highlighted in black). The LA0716 wild-type allele reading frame is illustrated by a codon in bold and underlined.
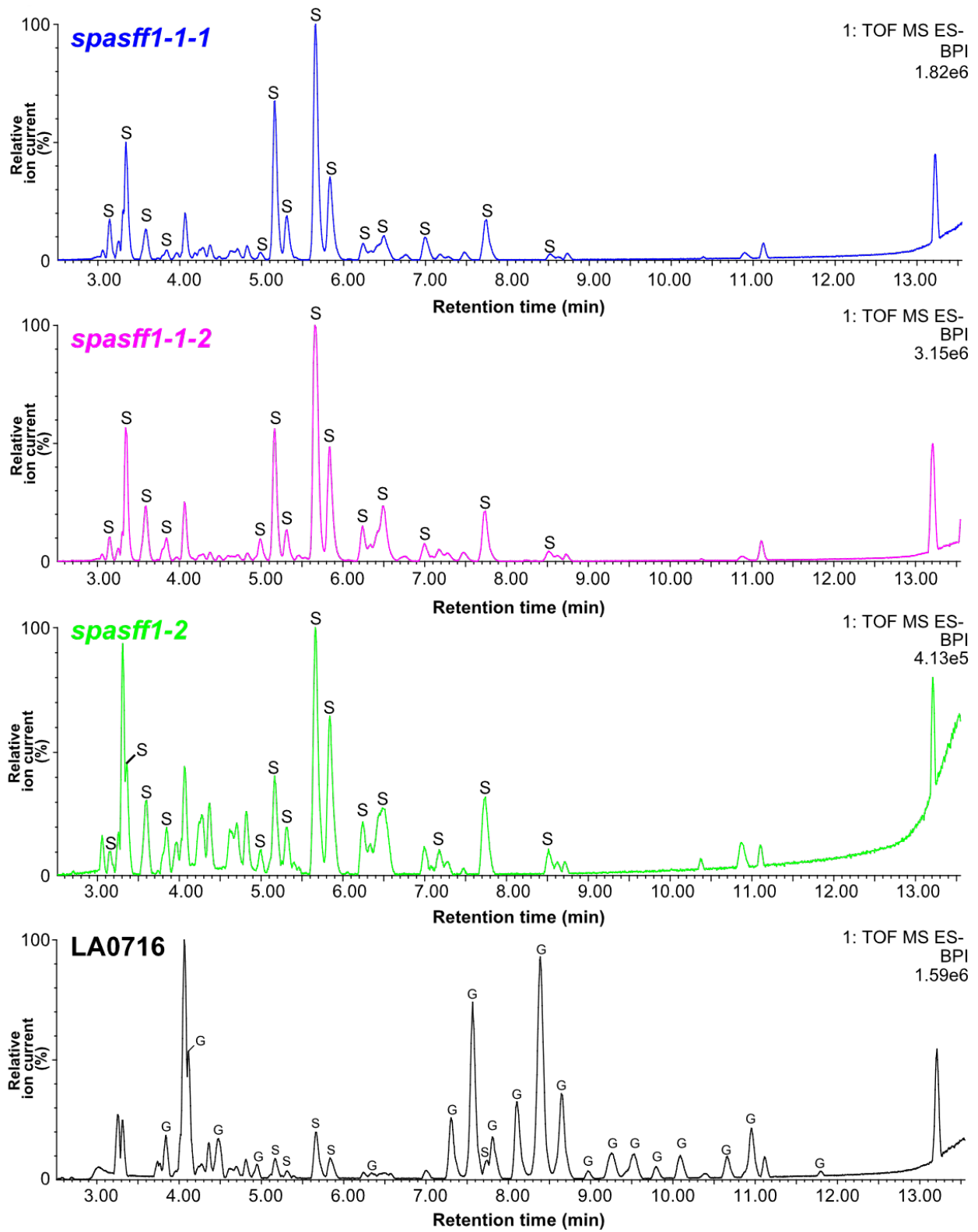
Figure S3.6. Acylsugars in BPI chromatograms of *spasff1* and LA0716 plants.

Figure S3.6 (cont'd)

Base peak intensity (BPI) chromatograms of trichome extracts from *S. pennellii* LA0716 and the three *spasff1* lines shown in Fig 3C. The 21 minute method and ESI- mode LC-MS was used for this analysis, as described in the Methods section. 2.5 to 13.5 minutes is presented because acylsucroses and acylglucoses elute in this time period. Note: *spasff1* lines were diluted 100 fold before LC-MS analysis to avoid saturation of the LC-MS detector. This is due to differences in ionization between acylsucroses and acylglucoses in ESI- mode. Note: *spasff1-1-1/1-1-2* are homozygous T_2 lines, while *spasff1-2* are homozygous T_1 lines that were grown together (the same used in Fig. 3.4B and C).
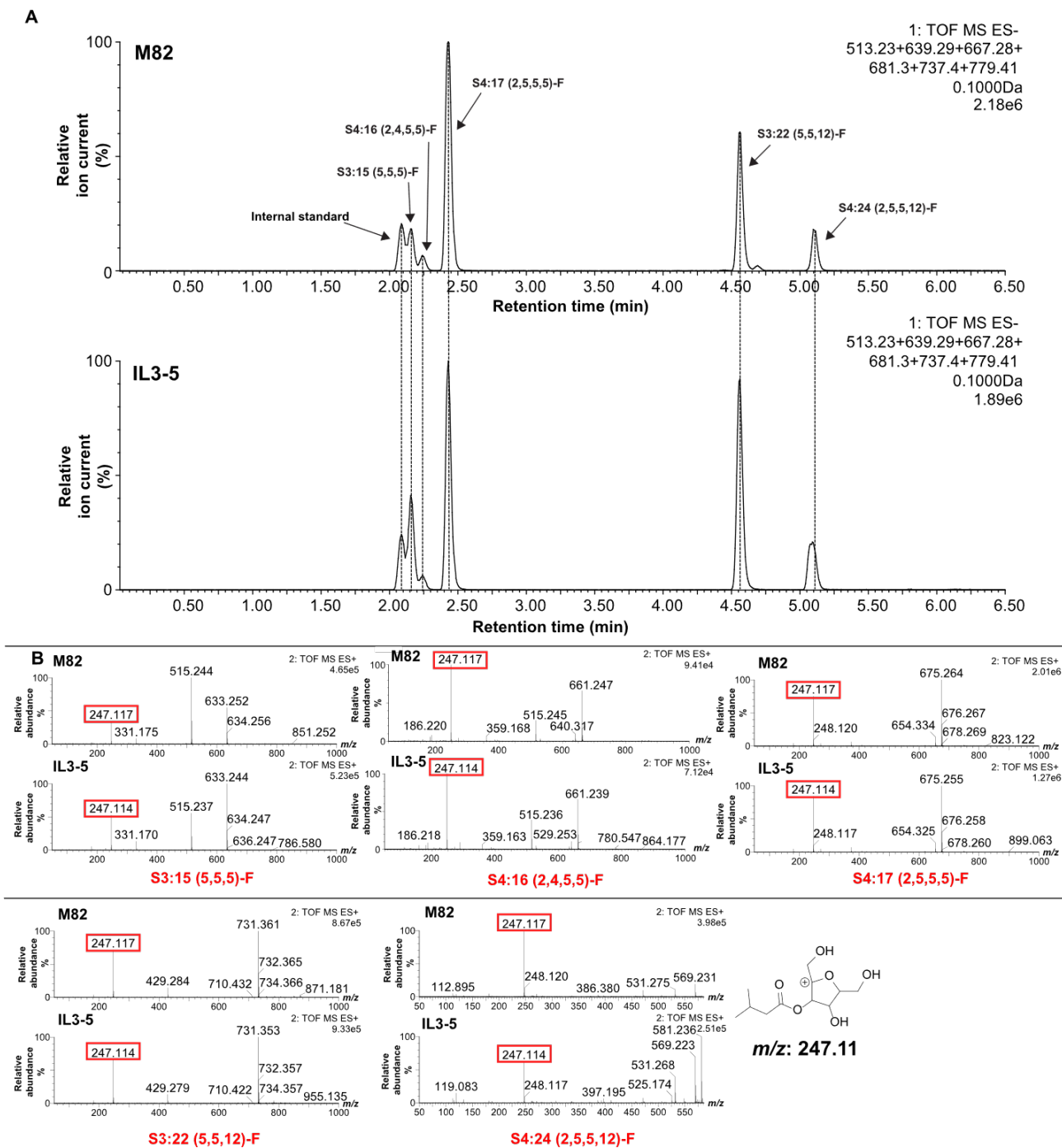
Figure S3.7. Comparison of acylsugars from IL3-5 and parental M82 using LC-MS.

(A) LC-MS analysis using ESI- mode to compare acylsugars of M82 and IL3-5. The major acylsugars in IL3-5 co-elute with acylsugars from M82. Extracted ion chromatograms of S3:15 (*m/z*: 639.28), S4:16 (*m/z*: 667.28), S4:17 (*m/z*: 681.30), S3:22 (*m/z*: 737.40), S4:24 (*m/z*: 779.41) and Telmisartan (*m/z*: 513.23) are shown. (B) ESI+ mode fragmentation of acylsucroses results in the cleavage of the glycosidic linkage (Ghosh et al., 2014). Fragment analysis of the

Figure S3.7 (cont'd)

acylsucroses using collision-induced dissociation from IL3-5 and M82 reveals a fragment ion (*m/z*: 247.11) consistent with the furanose ring of sucrose conjugated to a C5 acyl chain. This indicates that these acylsugars all possess a C5 acyl chain on the furanose ring. Samples were run on the 7 minute method detailed in the Methods section (mass window: 0.1Da). All ESI- mode acylsugars were identified as formate adduct.
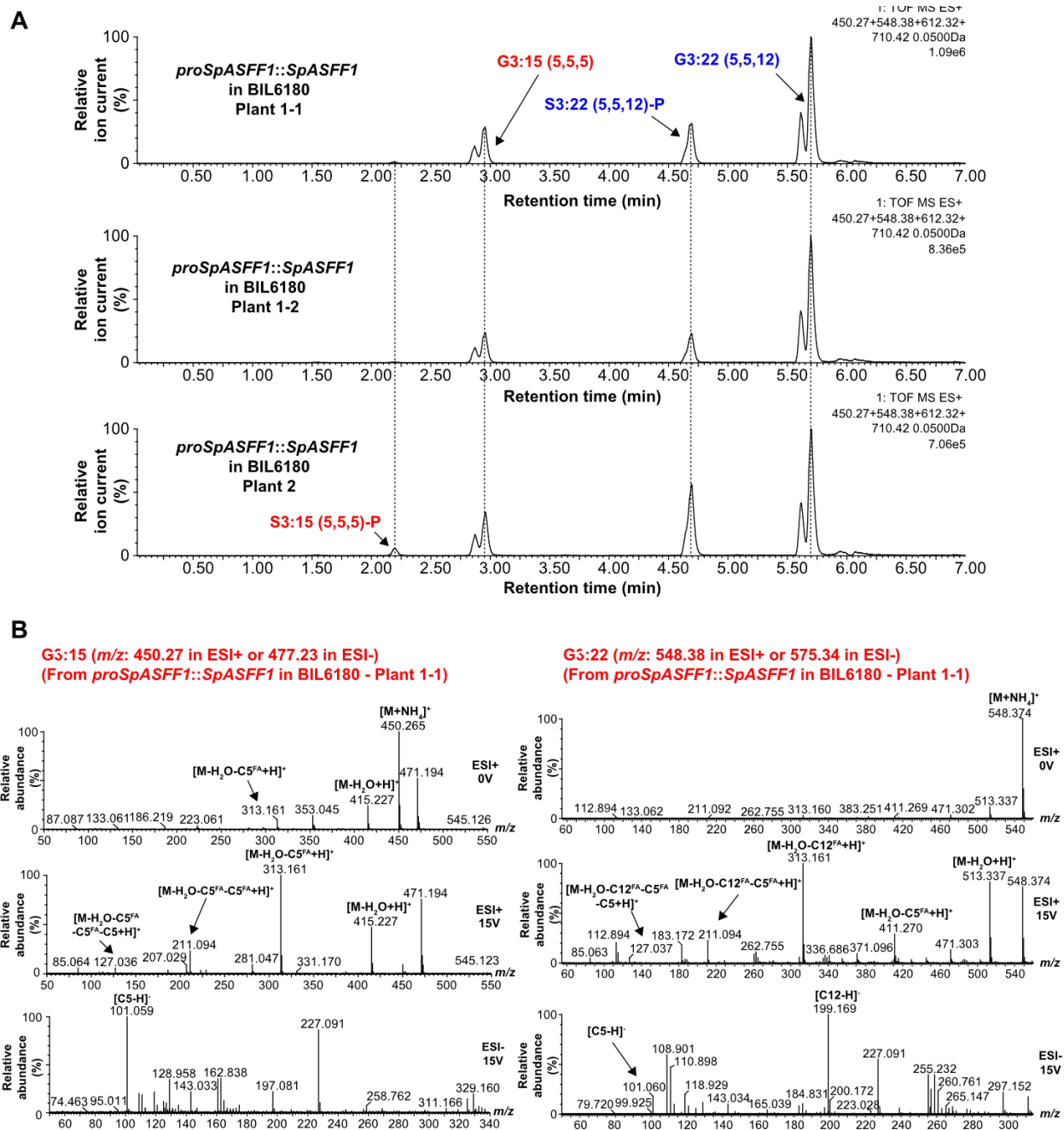
Figure S3.8. LC-MS analysis of P-type acylsucrose-producing *S. lycopersicum* BIL6180 stably transformed with *proSpASFF1*::*SpASFF1*.

(A) ESI+ mode LC-MS of trichome extracts from three *proSpASFF1*::*SpASFF1* T_2 lines originating from 2 T_0 lines. Extracted ion chromatograms of G3:15 (*m/z*:450.27), G3:22 (*m/z*:548.38), S3:15 (*m/z*: 612.32), and S3:22 (*m/z*: 710.42) are shown (mass window: 0.05 Da). Samples were run on the 7 minute method described in the Methods section. Each pair of

Figure S3.8 (cont'd)

acylglucose peaks fragment similarly, consistent with the existence of alpha/beta acylglucose anomers. (B) Fragmentation of G3:15 and G3:22 from *proSpASFF1*::*SpASFF1* in BIL6180 - Plant 1-1 using collision-induced dissociation. Fragmentation of G3:15 in ESI+ mode (0 and 15V) results in the loss of two C5 fatty acids, followed by loss of a C5 ketene. Fragmentation of G3:22 in ESI+ mode (15V) results in the loss of C12 and C5 fatty acid followed by loss of a C5 ketene. The higher collision energy at 15V ESI- reveals the presence of a C5 (*m/z:* 101.06*)*, or C5 (*m/z:* 101.06) and C12 (*m/z*: 199.17) fatty acids for G3:15 and G3:22 respectively. Please reference Fig. S3.4 legend for further detail on fragmentation. All ESI+ mode acylsugars were identified as ammonium adducts.
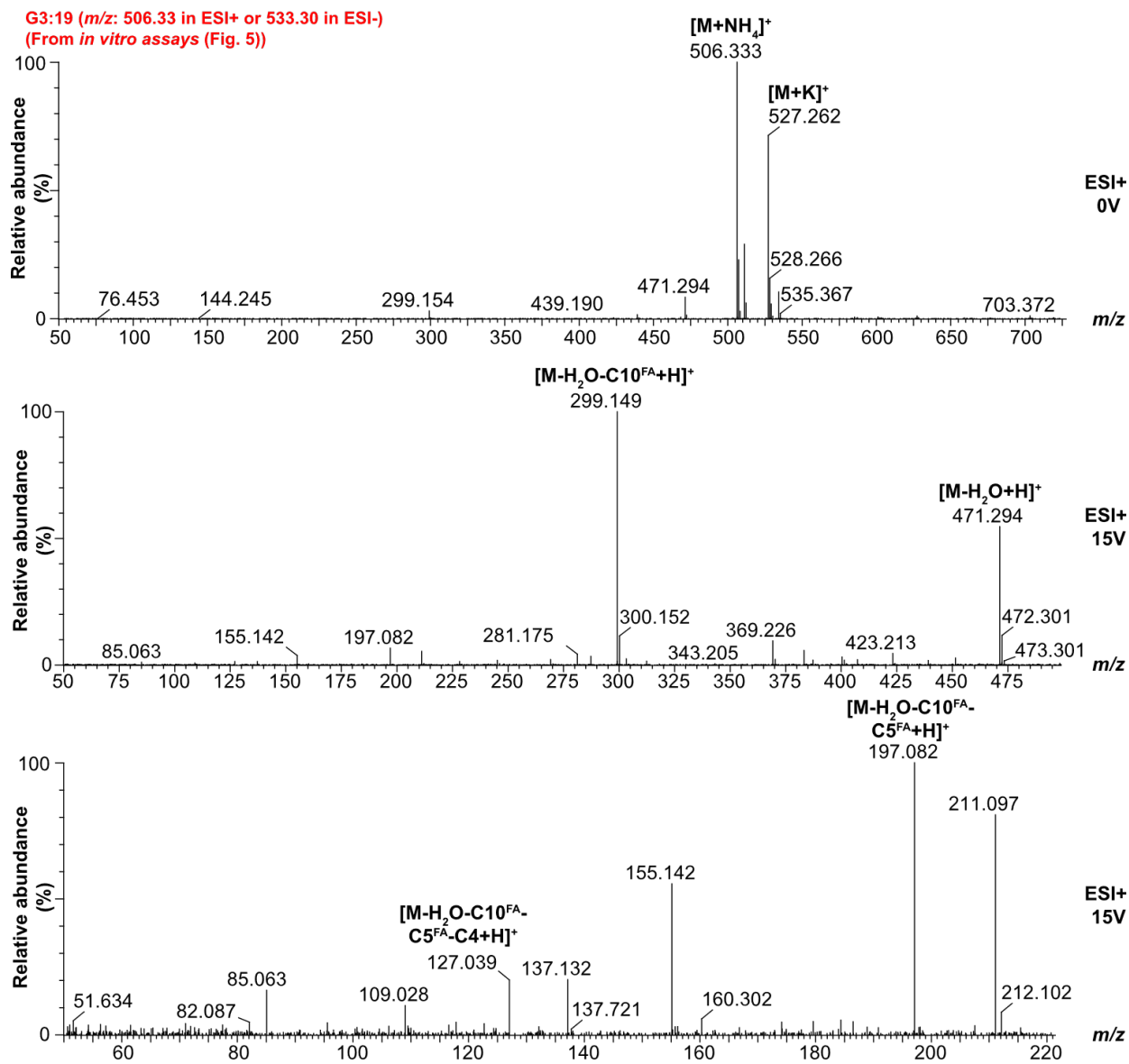
Figure S3.9. Mass spectra of G3:19-derived from SpASFF1 *in vitro* assay.

Fragmentation of the S3:19 + SpASFF1 reaction product, G3:19, in ESI+ mode (0V and 15V) using collision induced dissociation (from Fig. 3.6). The fragmentation is characterized by the loss of a C10 and C5 fatty acid from the triacylglucose, followed by loss of a C4 ketene. These results are consistent with the product cognate, S3:19. Further information of collision induced dissociation is present in Fig. S3.4.
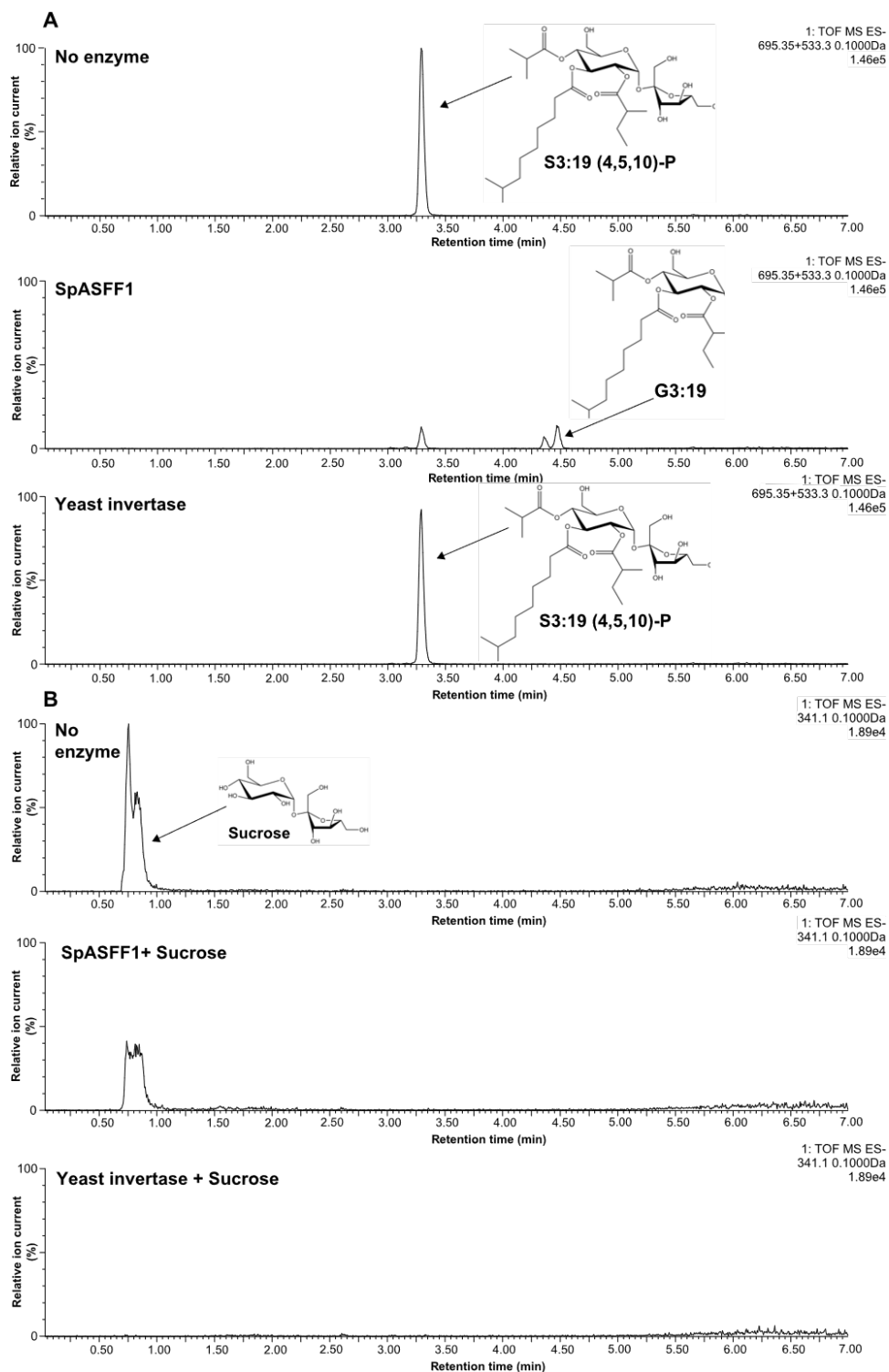
Figure S3.10. SpASFF1 cleaves a purified P-type triacylsucrose but not unmodified sucrose while yeast invertase cleaves unmodified sucrose but not triacylsucrose.

(A) ESI- mode LC-MS analysis of *in vitro* enzyme assay products indicates that SpASFF1 cleaves a P-type S3:19 ($5^{R2},10^{R3},4^{R4}$) acylsucrose to yield a product with *m/z* = 533.3 (middle

Figure S3.10. (cont'd)

chromatograph) while yeast invertase yields no hydrolysis products (lower chromatograph). (B) LC-MS analysis of *in vitro* assays with unacylated sucrose indicates complete hydrolysis of sucrose by yeast invertase (lower chromatograph) but abundant sucrose remaining when incubated with SpASFF1 (middle chromatograph); disappearance of the sucrose substrate was monitored rather than appearance of the glucose or fructose products due to poor detection of the monosaccharides resulting from low ionization efficiency. Note: Acylglucose structure is inferred from collision induced dissociation-mediated fragmentation (Figure S3.9). All ESI-mode acylsugars were identified as formate adducts, while sucrose was identified as an $[M-H]^-$ ion.

Table S3.1. Annotation of acylsugars identified in BIL6521 and BIL6521 x BIL6180 F2 using LC-MS and collision-induced dissociation.

Listed acylsugars were identified from trichome extracts of BIL6521 and BIL6521 x BIL6180 F2 plants. Acylsugars were annotated using LC-qToF mass spectrometry in ESI-/+ mode using varying collision energies as described in Methods. Acylsugar fragmentation was analyzed as shown in Fig. S3.4 using collision-induced dissociation.

BIL6521 acylsugars				
Mass/Charge ratio	ESI-/+	Acylsugar annotation	Retention time (min)	Adduct
681.30	–	S4:17(2,5,5,5)	2.44	HCOO$^-$
737.40	–	S3:22 (5,5,12)	4.56,4.69	HCOO$^-$
779.41	–	S4:24 (2,5,5,12)	5.10	HCOO$^-$
506.33	+	G3:19 (5,5,9)	4.84	NH$_4^+$
520.34	+	G3:20 (5,5,10)	4.97	NH$_4^+$
534.36	+	G3:21 (5,5,11)	5.23	NH$_4^+$
548.38	+	G3:22 (5,5,12)	5.68	NH$_4^+$
BIL6521 x BIL6180 acylsugars				
Mass/Charge ratio	ESI-/+	Acylsugar annotation	Retention time (min)	Adduct
639.29	–	S3:15 (5,5,5)	2.21	HCOO$^-$
681.30	–	S4:17 (2,5,5,5)	2.44	HCOO$^-$
723.38	–	S3:21 (4,5,12)	4.33	HCOO$^-$
737.40	–	S3:22 (5,5,12)	4.69	HCOO$^-$
751.41	–	S3:23 (5,6,12)	4.99	HCOO$^-$
436.25	+	G3:14 (4,5,5)	2.68	NH$_4^+$
450.27	+	G3:15(5,5,5)	2.96	NH$_4^+$
464.28	+	G3:16 (5,5,6)	3.25	NH$_4^+$
520.34	+	G3:20 (5,5,10)	4.95	NH$_4^+$
534.32	+	G3:21 (5,5,11)	4.56	NH$_4^+$
548.38	+	G3:22 (5,5,12)	5.68	NH$_4^+$

Table S3.2. Annotation of three GH candidates for *SpASFF1* identified in the AG 3.2.

This table provides the nominal or canonical activities of the three glycosyl hydrolase enzymes found in AG3.2 including members from the GH32, GH35, and GH47 families.

SGN ID (*S. lycopersicum*/ *S. pennellii*)	Enzyme family	Nominal activity	Typical substrates
Solyc03g121540/ Sopen03g040350	GH35	β-galactosidase	β-D-galactosides; α-L-arabinosides; oligo-galactosides (*47*)
Solyc03g121680/ Sopen03g040490	GH32	β-fructofuranosidase	β-fructofuranosides (*38*)
Solyc03g123900/ Sopen03g041640	GH47	α-mannosidase	Terminal α-D-mannose residues of oligo-mannose oligosaccharides (*48*)

Table S3.3. Primers/gBlocks/sgRNAs used in this study.

Red color indicates the NGG site in the sgRNAs.

Sequence name	Nucleotide sequence
ASFF_F	ATGGGATATGTTAGAAGTGTTTGG
ASFF_R	TCAATTGATTTGAGCTGTTTTCA
(pEAQ-HT)-ASFF-His_F	GTATATTCTGCCCAAATTCGATGGGATATGTTAGAAGTGT
(pEAQ-HT)-ASFF-His_R	TGATGGTGATGGTGATGCCCATTGATTTGAGCTGTTTTCA
RT_actin_F	GGTCGTACCACTGGTATTGT
RT_actin_R	AAACGAAGAATGGCATGTGG
RT_ASFF_F	CTACGCAGGCAGATGTAGAAA
RT_ASFF_R	ATCACTAGAAGGCAAGTGTAAGG
RT_EF-1a_F	TGCTGCTGTAACAAGATGGA
RT_EF-1a_R	AGGGGATTTTGTCAGGGTTG
RT_ubiquitin_F	TCGTAAGGAGTGCCCTAATGCTGA
RT_ubiquitin_R	CAATCGCCTCCAGCCTTGTTGTAA
ASFF_001_F	CAAAAAAGCAGGCTCCGCCTGATAGTTATGCCAATGTACCACA
ASFF_001_R	ACGATCCTCTTAGTTGGTCCAC
ASFF_002_F	GTGGACCAACTAAGAGGATCGT
ASFF_002_R	TGGCCCCATACAATGTTACCTG
ASFF_003_F	CAGGTAACATTGTATGGGGCCA
ASFF_003_R	TTTGGGAAGTTCTGGCTCGG
ASFF_004_F	CCGAGCCAGAACTTCCCAAA
ASFF_004_R	GGGTCGGCGCGCCCACCCTTAAAGTGTTCGACTGACCATTCT
ASFF_Promoter_F1	CAAAAAAGCAGGCTCCGCTGCCAATGTACCACAATTAGTATT
ASFF_Promoter_R1	AGAAAGCTGGGTCGGCTTTTAGTTGAAGATGGCAACTACATTTCA
ASFF_Chr3_Indel_002_F	CAAAAAGAAGAAAAGGAAAACAGACA
ASFF_Chr3_Indel_002_R	GTGGGACTAAAACTTTGTAGTTCC
04g011460_Marker_Indel-F	TAACAAAGCTTATGCACTCTTAG
04g011460_Marker_Indel-R	ATCTACTACCTTCATATGCACAT
ASFF1_transcript_amp_01F	TCATTTCCATTCATAGCTATGGCA
ASFF1_transcript_amp_01R	GTTGCACCATCCACTGCTAA
sgRNA 1 target sequence w/ PAM site	AGTCCAGTTGACGAGATCAG**TGG**
sgRNA 2 target sequence w/ PAM site	TCTTCTCTCTGGCGGAAAAC**CGG**
sgRNA1 gBLOCK	TTTTAATGTACTGGGGTGGATGCAGTGGGCCCCACTCTGTGAAGACAA ACTAGAATTCGAGCTCGGAGGGAGTGATCAAAAGTCCCACATCGATCA

109

	GGTGATATATAGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TTGAGTCCAGTTGACGAGATCAGGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTCGCT TTACTTGTCTTCTGCACGAAGTGGTTTAAACTATCAGTGTTTGACAGGAT ATATT
sgRNA2 gBLOCK	TTTAATGTACTGGGGTGGATGCAGTGGGCCCCACTCTGTGAAGACAAT TACGAATTCCCATGGGGAGGGAGTGATCAAAAGTCCCACATCGATCAG GTGATATATAGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGATT GTCTTCTCTCTGGCGGAAAACGTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT TTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTCGCTCG CTCAGATTGTCTTCTGCACGAAGTGGTTTAAACTATCAGTGTTTGACAG GATATATT
pAGM4723_SeqF1	TTTCGCCACCTCTGACTTGAG
pAGM4723_SeqR1	CAGCTTGGCATCAGACAAACC
pICSL11024_SeqF1	CGGCGAACTAATAACGCTCA
pICH47742CAS9_SeqF2	ACTTACGCTCATCTCTTCGAC
pICH47742_SeqF1	CTGTTGAATTACGTTAAGCATG
pICH41780_SeqR1	TCGGTCACATGTGCATCCTC
ASFF_SeqR	GTTGATTTCCAACTACCATTCTCC
pK7WG-Kan-F	TTGACTCTAGCTAGAGTCCGAA
pK7WG-Kan-R	ATTGAACAAGATGGATTGCACGCA

REFERENCES

REFERENCES

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods **9**: 39.

Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S, Sørensen I, Lichtenstein G, et al (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nat Genet **46**: 1034–1038.

Burke B, Goldsby G, Brian Mudd J (1987) Polar epicuticular lipids of *Lycopersicon pennellii*. Phytochemistry **26**: 2567–2571.

De Coninck B, Le Roy K, Francis I, Clerens S, Vergauwen R, Halliday AM, Smith SM, Van Laere A, Van Den Ende W (2005) Arabidopsis AtcwINV3 and 6 are not invertases but are fructan exohydrolases (FEHs) with different substrate specificities. Plant, Cell Environ. doi: 10.1111/j.1365-3040.2004.01281.x.

Consortium TG (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–641.

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol **34**: 184–191.

Van Den Ende W, Lammens W, Van Laere A, Schroeven L, Le Roy K (2009) Donor and acceptor substrate selectivity among plant glycoside hydrolase family 32 enzymes. FEBS J **276**: 5788–5798.

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics **141**: 1147–62.

Fan P, Miller AM, Liu X, Jones AD, Last RL (2017) Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. Nat Commun **8**: 1–13.

Fan P, Miller AM, Schilmiller AL, Liu X, Ofner I, Jones AD, Zamir D, Last RL (2016) In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. Proc Natl Acad Sci **113**: E239–E248.

Fillatti JJ, Kiser J, Rose R, Comai L (1987) Efficient transfer of a glyphosate tolerance gene into tomato using a binary *Agrobacterium tumefaciens* vector. Nat Biotechnol **5**: 726–730.

Fobes JF, Mudd JB, Marsden MP (1985) Epicuticular lipid accumulation on the leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. Plant Physiol **77**:

567–70.

Ghangas GS, Steffens JC (1993) UDP glucose: fatty acid transglucosylation and transacylation in triacylglucose biosynthesis. Proc Natl Acad Sci **90**: 9911–9915.

Ghosh B, Westbrook TC, Jones AD (2014) 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.

Herscovics A (2001) Structure and function of class I alpha 1,2-mannosidases involved in glycoprotein synthesis and endoplasmic reticulum quality control. Biochimie **83**: 757–62.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol **31**: 827–832.

Johnson KA, Goody RS (2011) The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. Biochemistry **50**: 8264–8269.

Karimi M, Inzé D, Depicker A (2002) GATEWAY™ vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci **7**: 193–195.

Kim J, Kang K, Gonzales-Vigil E, Shi F, Jones AD, Barry CS, Last RL (2012) Striking natural diversity in glandular trichome acylsugar composition is shaped by variation at the Acyltransferase2 locus in the wild tomato *Solanum habrochaites*. Plant Physiol **160**: 1854–70.

King RR, Calhoun LA (1988) 6 2,3-Di-*O*- and 1,2,3-tri-*O*-acylated glucose esters from the glandular trichomes of *Datura metel*. Phytochemistry **27**: 3761–3763.

Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat A V., et al (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci U S A **110**: E2655-62.

Kuai JP, Ghangas GS, Steffens JC (1997) Regulation of triacylglucose fatty acid composition. Plant Physiol **115**: 1581–1587.

Leckie BM, De Jong DM, Mutschler MA (2013) Quantitative trait loci regulating sugar moiety of acylsugars in tomato. Mol Breed **31**: 957–970.

Li AX, Eannetta N, Ghangas GS, Steffens JC (1999) Glucose polyester biosynthesis. Purification and characterization of a glucose acyltransferase. Plant Physiol **121**: 453–60.

Li AX, Steffens JC (2000) An acyltransferase catalyzing the formation of diacylglucose is a serine carboxypeptidase-like protein. Proc Natl Acad Sci **97**: 6902–6907.

Luu VT, Weinhold A, Ullah C, Dressel S, Schoettner M, Gase K, Gaquerel E, Xu S,

Baldwin IT (2017) *O*-acyl sugars protect a wild tobacco from both native fungal pathogens and a specialist herbivore. Plant Physiol **174**: 370–386.

Maldonado E, Torres FR, Martínez M, Pérez-Castorena AL (2006) Sucrose esters from the fruits of *Physalis nicandroides* var. attenuata. J Nat Prod **69**: 1511–1513.

Mandal S, Ji W, Mcknight TD (2018) Candidate gene networks for acylsugar metabolism and plant defense in wild tomato *Solanum pennellii*. bioRxiv. doi: 10.1101/294306.

Mathew AK, Padmanaban VC (2013) Metabolomics: the apogee of the omics trilogy. Int. J. Pharm. Pharm. Sci.

Matsuzaki T, Shinozaki Y, Suhara S, Ninomiya M, Shigematsu H, Koiwai A (1989) Isolation of glycolipids from the surface lipids of *Nicotiana bigelovii* and their distribution in *Nicotiana* species. Agric Biol Chem **53**: 3079–3082.

McCormick S (1991) Transformation of tomato with *Agrobacterium tumefaciens*. Plant Tissue Cult. Man. Springer Netherlands, Dordrecht, pp 311–319.

Minic Z (2008) Physiological roles of plant glycoside hydrolases. Planta **227**: 723–40.

Moghe GD, Leong BJ, Hurney SM, Jones AD, Last RL (2017) Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. Elife **6**: 1–33.

Nadakuduti SS, Uebler JB, Liu X, Jones AD, Barry CS (2017) Characterization of trichome-expressed BAHD acyltransferases in *Petunia axillaris* reveals distinct acylsugar assembly mechanisms within the Solanaceae. Plant Physiol **175**: 36–50.

Nagana Gowda GA, Raftery D (2017) Recent advances in NMR-based metabolomics. Anal Chem **89**: 490–510.

Nakashima T, Wada H, Morita S, Erra-Balsells R, Hiraoka K, Nonami H (2016) Single-cell metabolite profiling of stalk and glandular cells of intact trichomes with internal electrode capillary pressure probe electrospray ionization mass spectrometry. Anal Chem **88**: 3049–3057.

Nesbitt TC, Tanksley SD (2002) Comparative sequencing in the genus Lycopersicon. implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics **162**: 365–79.

Ning J, Moghe GD, Leong B, Kim J, Ofner I, Wang Z, Adams C, Jones AD, Zamir D, Last RL (2015) A feedback-insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. Plant Physiol **169**: 1821–35.

Ofner I, Lashbrooke J, Pleban T, Aharoni A, Zamir D (2016) *Solanum pennellii* backcross inbred lines (BILs) link small genomic bins with tomato traits. Plant J **87**: 151–160.

Peyret H, Lomonossoff GP (2013) The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. Plant Mol Biol **83**: 51–58.

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res **29**: e45.

Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol **13**: 1–15.

Schilmiller A, Shi F, Kim J, Charbonneau AL, Holmes D, Daniel Jones A, Last RL (2010) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. Plant J **62**: 391–403.

Schilmiller AL, Charbonneau AL, Last RL (2012) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. Proc Natl Acad Sci **109**: 16377–16382.

Schilmiller AL, Gilgallon K, Ghosh B, Jones AD, Last RL (2016) Acylsugar acylhydrolases: carboxylesterase catalyzed hydrolysis of acylsugars in tomato trichomes. Plant Physiol **170**: 1331–1344.

Schilmiller AL, Moghe GD, Fan P, Ghosh B, Ning J, Jones AD, Last RL (2015) Functionally divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. Plant Cell **27**: 1002–1017.

Shapiro JA, Steffens JC, Mutschler MA (1994) Acylsugars of the wild tomato *Lycopersicon pennellii* in relation to geographic distribution of the species. Biochem Syst Ecol **22**: 545–561.

Smeda JR, Schilmiller AL, Anderson T, Ben-Mahmoud S, Ullman DE, Chappell TM, Kessler A, Mutschler MA (2018) Combination of acylglucose QTL reveals additive and epistatic genetic interactions and impacts insect oviposition and virus infection. Mol Breed **38**: 3.

Tanthanuch W, Chantarangsee M, Maneesan J, Ketudat-Cairns J (2008) Genomic and expression analysis of glycosyl hydrolase family 35 genes from rice (*Oryza sativa* L.). BMC Plant Biol **8**: 84.

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One **6**: e16765.

Chapter 4. Acylinositol biosynthesis in *Solanum quitoense*

Abstract

Plants synthesize an array of structurally diverse compounds that are restricted in phylogenetic distribution called specialized metabolites. Acylsugars are a group of these metabolites produced in the trichomes of species across the Solanaceae family. There is a tremendous amount of acylsugar diversity – one example of this is in *Solanum quitoense,* a South American fruit crop. This Andean plant produces acylsugars that contain a *myo*-inositol core. VIGS analysis in *S. quitoense* identified two trichome-expressed BAHD acyltransferases involved in acylinositol biosynthesis *in planta*. One acyltransferase catalyzed acetylation of triacylinositols *in vitro*, while the other appears to be involved in early steps in the biosynthetic pathway. These results provide a foundation to understand how a broader diversity of acylsugars is produced across the Solanaceae, expanding beyond our current knowledge of acylsucrose and acylglucose biosynthesis.

Introduction

Published studies indicated that differences in acylsugar core can affect the activity of these protective compounds. For example, Puterka and co-workers reported variation in insecticidal activities for pear sucker (*Cacopsylla pyricola*) that differed depending on the sugar and the conjugated acyl chains generated synthetically (Puterka et al., 2003). A similar effect was found with tobacco aphid (*Myzus nicotianae*) mortality rates, with sucrose octanoate being 2-3 times more effective at $1/6^{th}$ of the concentration of xylitol acylsugars. Leckie et al. examined the effect of acylsucrose and acylglucose mixtures on western flower thrips (*Frankliniella occidentalis)*, and tobacco thrips (*Frankliniella fusca*) oviposition (Leckie et al., 2016). The study found that a mixture of acylsucroses and glucoses has a synergistic effect on reduction of insect oviposition. Acylsugars also affect microbial fitness. For instance, Zhao and

colleagues examined the antimicrobial effects of acylsugars containing different sugar cores and acyl chains against several common Gram positive bacteria (Zhao et al., 2015). They determined that sucrose monocaprate (sucrose with one nC10 acyl ester chain) exhibited a stronger effect than other acylsugars tested. These results indicate that a better understanding of acylsugar biosynthetic pathways is desirable to enable us to perform more rigorous analysis of the structure-function relationship of acylsugars.

Exploring natural chemical diversity can help identify novel acylsugars, and much of the Solanaceae family remains unexplored. One such clade is the spiny Solanums or subgenus Leptostenonum. *Solanum quitoense* represents an interesting species – a member of section Lasiocarpa within the Leptostenonum subgenus (Särkinen et al., 2013). My interest in this species was piqued due to previous work that elucidated structures of a handful of the most abundant acylsugars produced by *S. quitoense* (Hurney, 2018). These acylsugars are represented by a group of metabolites with a *myo*-inositol core conjugated to acetyl groups and medium length acyl chains (nC10 or nC12) (Hurney, 2018) (Figure 4.1). All characterized acylinositols contain two medium length acyl chains, and at least one acetyl group with some being acetylated twice, based on the NMR analysis (Hurney, 2018). Deep profiling of *S. quitoense* samples by LC-MS revealed a matrix of different acylinositols, with more than 30 chromatographically separable peaks present in some extracts (Hurney, 2018). These metabolite structures can help identify biosynthetic enzymes and facilitate future study of acylinositol effect on insects.

Knowledge of characteristics of previously analyzed acylsugar biosynthetic enzymes can facilitate identification of the *S. quitoense* acylinositol biosynthetic pathway. All characterized acylsugar-related genes are enriched in the trichomes relative to the underlying tissue, consistent with tissue-specific expression and biosynthesis
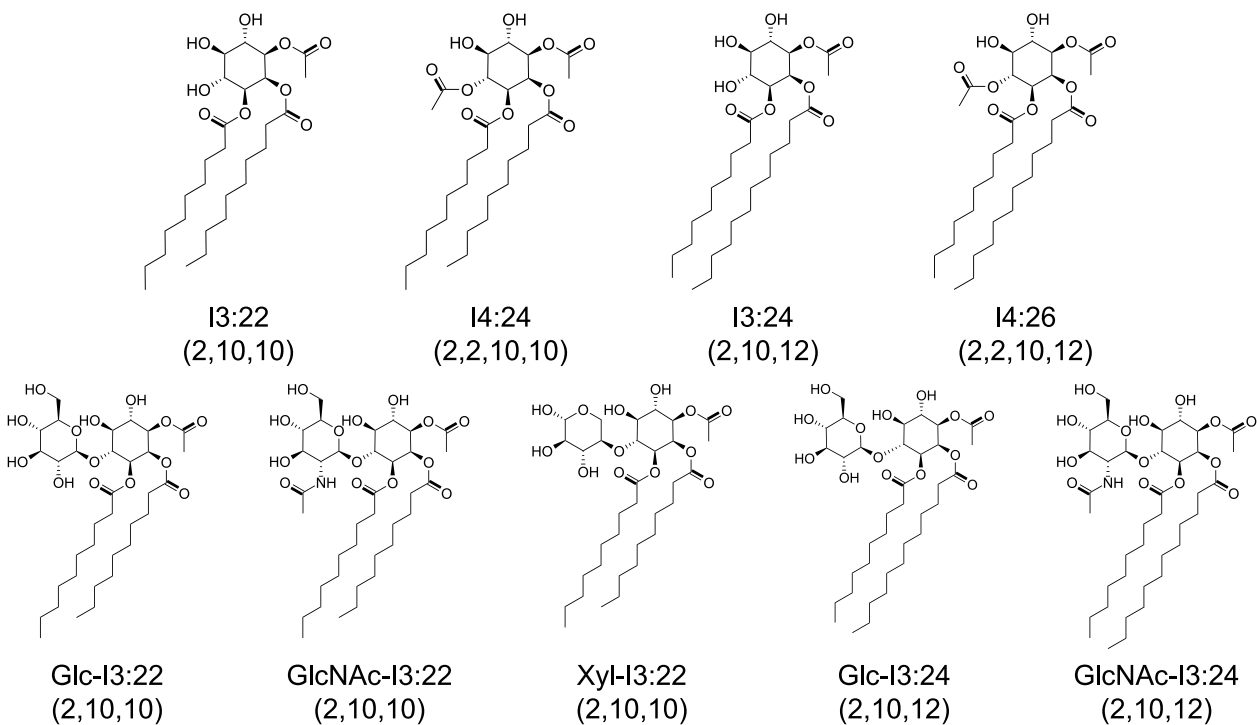
Figure 4.1. Structurally characterized acylsugars produced by *S. quitoense*.

The structures were previously characterized by NMR and reproduced from Hurney et. al. (2018). Acylsugars produced by *S. quitoense* contain acyl esters of straight-chain C10 or C12, and C2. There are four structurally characterized acylinositol compounds, and five glycosylated-acylinositol compounds, modified by glucose, xylose, and *N*-acetylglucosamine (Hurney, 2018).

(Ning et al., 2015; Moghe et al., 2017; Nadakuduti et al., 2017). This characteristic was used previously to drastically reduce the number of suitable transcripts for a biosynthetic pathway. Another feature of known ASAT enzymes is that they are evolutionarily related, forming a subclade within the larger BAHD acyltransferase phylogeny (Moghe et al., 2017). I used these two characteristics to inform my search for *S. quitoense* acylinositol biosynthetic genes.

Here we report identification of candidates for acylinositol biosynthesis using a combination of existing RNAseq data and phylogenetic analysis. An *S. quitoense* virus-induced gene silencing (VIGS) protocol was developed and used to analyze two candidate genes in acylinositol biosynthesis. This *in planta* method successfully identified a triacylinositol acyltransferase (TAIAT). Further VIGS analysis identified an inositol acyltransferase (IAT) that acylates *myo*-inositol using medium-length acyl chains at a position inconsistent with *in planta* acylsugars. Additional *in vitro* analyses were attempted to elucidate how the enzyme functions in acylinositol biosynthesis.

Results

Identification of ASAT candidates in *S. quitoense*

I sought BAHD family candidates for acylinositol biosynthesis in *S. quitoense* and employed a broader approach to identify potential ASATs. The presence of the HXXXD motif – characteristic of BAHD acyltransferases – was used to filter the protein sequences, and the resulting set was screened for the characteristic BAHD enzyme DFGWG motif to further narrow the list. Another criterion was that these two motifs were present in the correct relative positions; this involved selecting proteins in which the HXXXD motif is between the N-terminus and middle of the protein and upstream of the DFGWG motif, which is generally near the C-terminus

(D'Auria, 2006). Candidates were further selected by restricting protein sequence length to 400-500 amino acids. These factors narrowed the list of BAHD candidates to 42 putative proteins. Previously characterized acylsugar acyltransferases (ASATs) are expressed and enriched in glandular trichomes relative to non-acylsugar accumulating tissues (Schilmiller et al., 2012; Ning et al., 2015; Fan et al., 2016; Leong et al., 2019). Thus, we used existing RNAseq data to select candidates with > 500 reads in the trichome samples for further analysis (Moghe et al., 2017). Previously characterized ASATs were found in a clade III representing a subset of BAHD acyltransferases. Multiple sequence alignments and phylogenetic analysis were performed using MEGA-X and the maximum likelihood algorithm to elucidate the phylogenetic relationships between those BAHD sequences and several previously identified BAHD acyltransferases (D'Auria, 2006), including ASATs (Figure 4.2) (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016; Moghe et al., 2017; Nadakuduti et al., 2017). We focused our analysis on proteins closely related to existing ASATs.
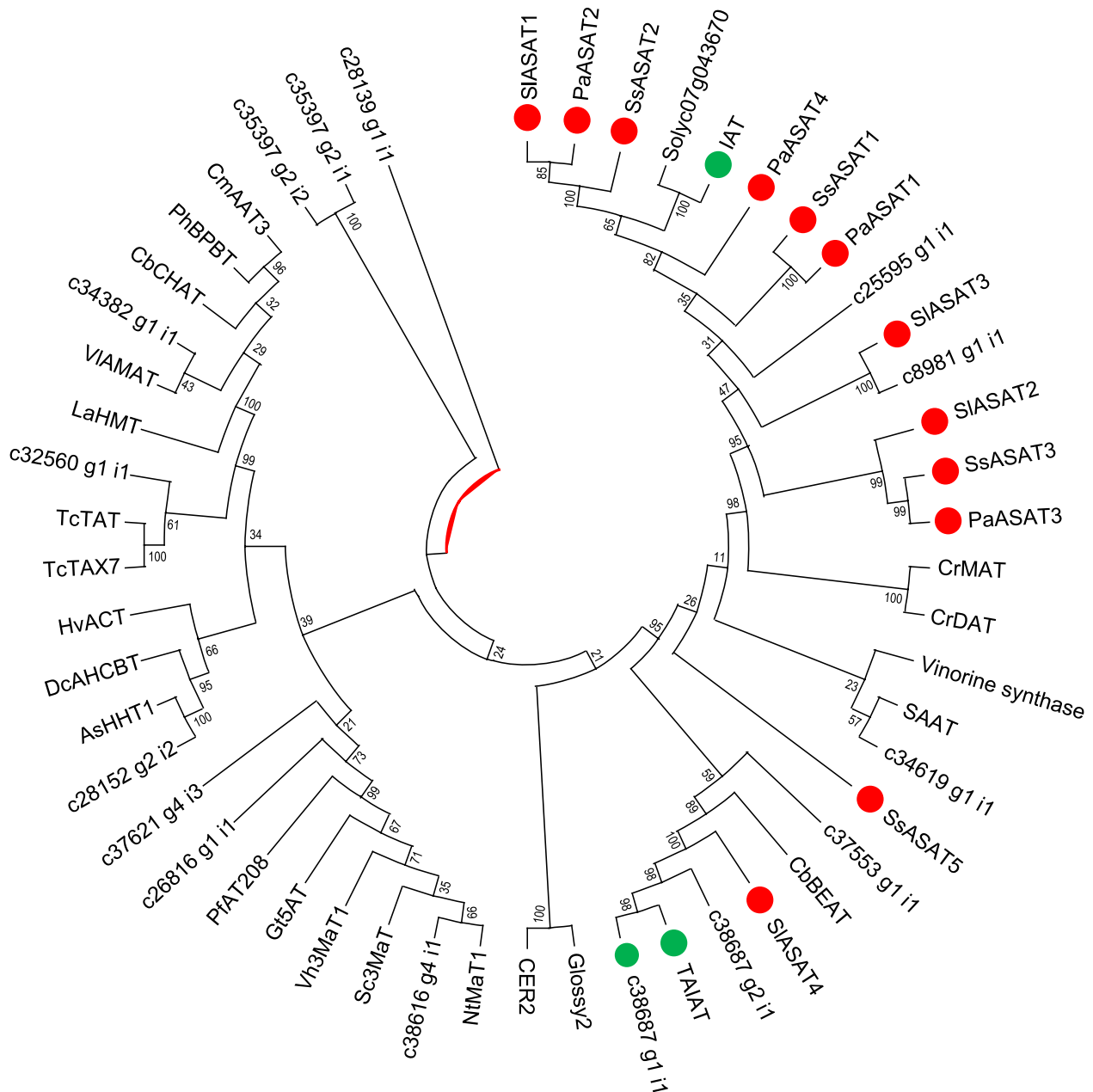
Figure 4.2. Phylogenetic analysis of putative BAHD acyltransferases from *S. quitoense* and previously identified BAHD acyltransferases and ASATs.

The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. The tree with the highest log likelihood (-41895.65) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood values. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.9154)). This analysis involved 52 amino acid sequences. All positions with less than 30% site coverage were eliminated, i.e., fewer than

Figure 4.2. (cont'd)

70% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were 469 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. Bootstrap values are identified by the nodes (1000 replicates). Note: red dots indicate previously characterized ASATs, green dots are BAHD acyltransferases characterized in this study. Proteins that start with 'c#####_g#' represent putative BAHD acyltransferases identified from *Solanum quitoense* RNAseq data

I attempted to identify genes responsible for early steps in the acylinositol biosynthetic pathway. I hypothesized that the BAHD enzymes involved in *S. quitoense* acylinositol biosynthesis would be evolutionarily related to other acylsugar biosynthetic pathways. In addition to the phylogenetic analysis, BLASTn analysis with Sl-ASAT1 as query against the dataset revealed a transcript annotated c39979_g2. This mRNA is predicted to encode a BAHD acyltransferase with the hallmarks of a functional protein, including the canonical HXXXD and DFGWG BAHD motifs (Figure S4.1). Detailed analysis revealed that a *Solanum lycopersicum* chromosome 7 gene – rather than the BLAST query Sl-ASAT1 sequence – is the BAHD acyltransferase most closely related to c39979_g2 (Figure 4.2 and Table S4.1). Previous work indicated that this cultivated tomato gene, Solyc07g043670, is a paralog of Sl-ASAT1 (Moghe et al., 2017). Analysis of RNAseq data revealed that this transcript, c39979_g2, is strongly enriched in *S. quitoense* trichomes compared to the trichome-shaved petiole tissue (Table 4.1). This matched ASATs and other acylsugar biosynthetic genes that show enrichment in the trichomes compared to the underlying tissue in three different species (Schilmiller et al., 2012; Ning et al., 2015; Fan et al., 2016; Fan et al., 2017; Moghe et al., 2017; Nadakuduti et al., 2017; Leong et al., 2019). I hypothesized that this enzyme could be an inositol acyltransferase (IAT) based on the homology to Sl-ASAT1, and enrichment of expression in trichomes.

Two BAHD acyltransferases with trichome-enriched expression, c38687_g1_i4 and c38687_g1_i1, were identified as close homologs of Sl-ASAT4 through BLASTn analysis (Figure S4.2 and Table 4.1). Phylogenetic analysis of the inferred protein sequences revealed that – beyond the relationship to each other – they are most closely related to

Table 4.1. RNAseq reads from trichome and petiole samples for ASAT candidates.

RNAseq reads from experiments described by Moghe et al. (2017). T represents trichome; P represents shaved petiole as previously described.

Gene	T1 reads	T2 reads	P1 reads	P2 reads
c39979_g2_i1	4223	5166	13	10
c38687_g1_i1	1689	1578	5	2
c38687_g1_i4	45109	49756	155	131

Sl-ASAT4, an acetyltransferase (Figure 4.2). We focused on c38687_g1_i4 because it has 30-40x more trichome reads than c38687_g1_i1; we hypothesized that this enzyme (c38687_g1_i4) could function in acetylation based on its phylogenetic relationship to Sl-ASAT4, and the fact that acylinositols in *S. quitoense* contain 1-2 acetyl groups. We refer to this gene as triacylinositol acetyltransferase (TAIAT) based on the *in planta* and biochemical analyses described below.

In vivo analysis of IAT

I sought *in planta* evidence to determine if IAT is involved in *S. quitoense* acylinositol biosynthesis and thus used virus-induced gene silencing (VIGS) to reduce expression of target transcripts. This technique was used because there were no transformation methods available for *S. quitoense*. Two distinct fragments targeting IAT was used for VIGS because the use of two fragments per transcript to achieve a similar phenotype reduces the chance that the result is due to off-target effects. I found no reported instances of VIGS in *S. quitoense* in the literature, so I developed a protocol for VIGS in this species, targeting phytoene desaturase as a positive control. The resulting photobleaching phenotype was present in multiple leaves and used as a positive control to demonstrate the success of the silencing and guide plant sampling (Figure 4.3). Acylsugars were extracted from candidate-silenced plants and compared to empty vector (pTRV2-LIC) negative controls by LC-MS analysis.

Silencing of IAT expression yielded reductions in the four major *S. quitoense* acylinositols, as expected for silencing an early enzyme in the pathway (Figure 4.4)

Figure 4.3. Plant phenotypes of representative *S. quitoense* VIGS plants silenced using PDS fragment cloned into pTRV2-LIC.

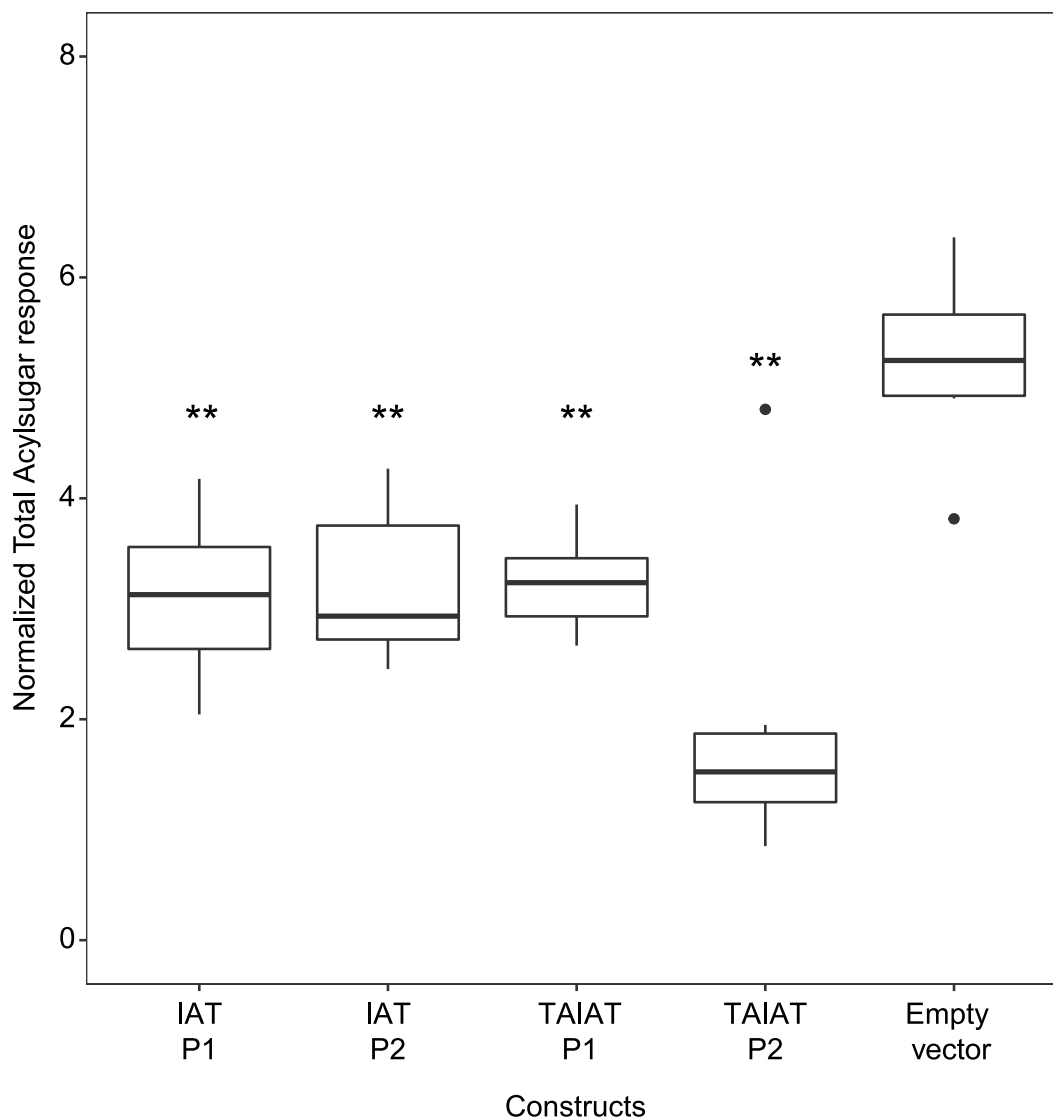Pot length and width is uniform at 2.875 inches each.

Figure 4.4. Analysis of total acylsugars accumulating in leaf surface extracts of VIGS plants

Quantification of acylsugars in plants inoculated with constructs targeting different fragments compared to empty vector controls. Values represent the normalized response of four major acylsugars described (I3:22, I4:24, I3:24, and I4:26). Normalized response is the peak area (m+2 natural isotopes of formate adducts) normalized to both the internal standard and dry weight of the extracted leaf. Welch's two sample t-test was performed between the test samples and the empty vector control plants to determine the p-value. (* - p value < 0.05; ** - p value < 0.01; *** - p value < 0.001) (n=6).

(p value < 0.01). This is similar to phenotypes seen following reduction of early steps in *S. lycopersicum* and *Salpiglossis sinuata* acylsucrose biosynthesis, ie. RNAi for Sl-ASAT1, Sl-ASAT2, Sl-ASAT3, and VIGS for SsASAT1, and SsASAT2 (Schilmiller et al., 2015; Fan et al., 2016; Moghe et al., 2017). This phenotype is consistent with a role for IAT in acylinositol biosynthesis in *S. quitoense*. qPCR analysis revealed reductions in IAT expression in the target plants relative to the controls (Figure S4.3). It was not clear from these data which step this enzyme catalyzes; the *in planta* data for IAT indicated involvement in acylinositol biosynthesis, but did not resolve placement in the pathway.

In vivo analysis of TAIAT

The acylinositol structures led us to hypothesize how TAIAT fits into *S. quitoense* acylinositol biosynthesis. *S. quitoense* acylinositols contain 2 medium-length acyl chains and 1-2 acetyl groups (Hurney, 2018). The timing of the second acetylation can be inferred from existing NMR structures. Four earlier identified non-glycosylated acylinositols contain the following makeup: I3:22 (2,10,10), I4:24 (2,2,10,10), I3:24 (2,10,12), and I4:26 (2,2,10,12). Both I3:22/I4:24 and I3:24/I4:26 differ by a single acetylation, which led to the hypothesis that I3:22 and I3:24 are intermediates in the biosynthesis of I4:24 and I4:26, respectively.

We tested the hypothesis that TAIAT catalyzes the acetylation of triacylinositols by asking whether the VIGS lines had altered levels of tri- and tetraacylinositols *in planta*. There was a statistically significant increase in the ratio of monoacetylated I3:22 (2,10,10) compared to diacetylated I4:24 (2,2,10,10) in plants inoculated with either construct targeting TAIAT relative to the controls (Figure 4.5), with a similar effect for I3:24 (2,10,12) compared to I4:26
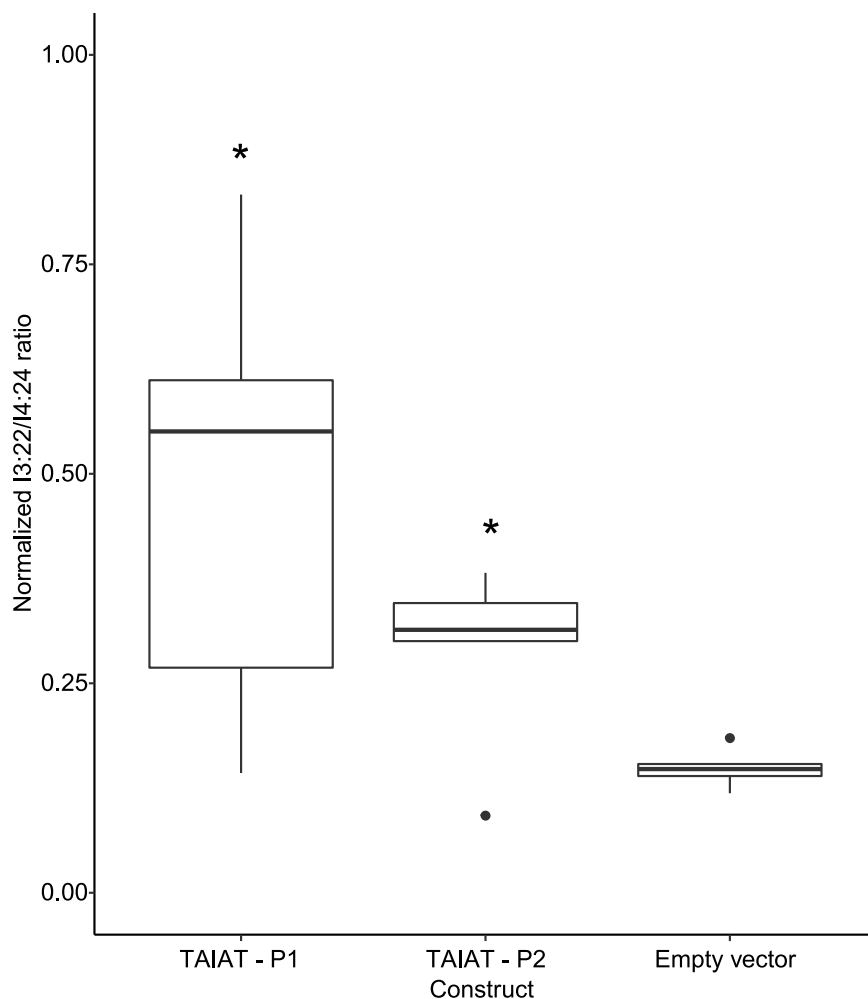
Figure 4.5. Analysis of TAIAT substrate to product ratio in TAIAT knockdown lines (I3:22/I4:24).

Ratio of I3:22 (2,10,10) and I4:24 (2,2,10,10) acylsugars in TAIAT targeted plants compared to empty vector controls. Values represent the ratio of normalized response of I3:22 to I4:24. Normalized response represents peak area (m+2 natural isotopes of formate adducts) normalized to both the internal standard (telmisartan) and dry weight of the extracted leaf. Welch's two-sample t-test was performed between the test samples and the empty vector control plants to determine the p-value. (* - p value < 0.05; ** - p value < 0.01; *** - p value < 0.001) (n=6).
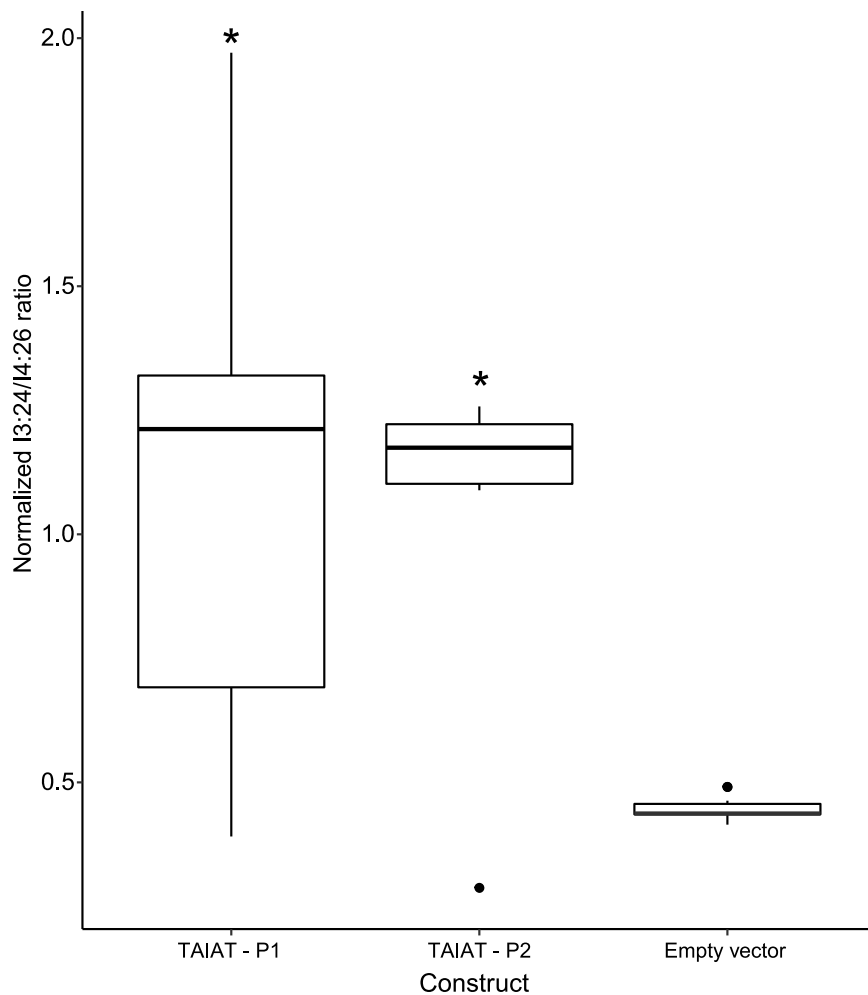
Figure 4.6. Analysis of TAIAT substrate and product ratio in TAIAT knockdown lines (I3:24/I4:26).

Ratio of I3:24 (2,10,12) and I4:26 (2,2,10,12) acylsugars in TAIAT targeted plants compared to empty vector controls. Values represent the ratio of normalized response of I3:24 to I4:26. Normalized response represents peak areas (m+2 natural isotopes of formate adducts) normalized to both the internal standard (telmisartan) and dry weight of the extracted leaf. Welch's two-sample t-test was performed between the test samples and the empty vector control plants to determine the p-value. (* - p value < 0.05; ** - p value < 0.01; *** - p value < 0.001) (n=6).

(2,2,10,12) (Figure 4.6). qPCR analysis revealed reductions in TAIAT expression in the target plants relative to the controls (Figure S4.4). This phenotype supported our hypothesis that I3:22 and I3:24 are direct precursors of the more acetylated products and that TAIAT catalyzes the acetylation of these triacylinositols.

In vitro analysis of TAIAT

We used *in vitro* assays of TAIAT as a complement to the *in planta* phenotypes that suggested TAIAT catalyzes acetylation of I3:22 (2,10,10) and I3:24 (2,10,12). Therefore, His-tagged TAIAT was heterologously expressed in *E. coli* and purified using Ni-Nitrilotriacetic acid (Ni-NTA). *S. quitoense* acylsugar substrates for *in vitro* assays were enriched using semi-preparative liquid chromatography. TAIAT catalyzed the formation of I4:24 (2,2,10,10) *in vitro*, utilizing acetyl-CoA and I3:22 (Figure 4.7). Comparisons between the two substrate peaks present in the (-) and (+) enzyme assays suggested that TAIAT primarily catalyzes acetylation of the major isomer of I3:22 (Figure 4.7). This suggested that TAIAT shows selectivity for the different tri-acylated inositols present in the purified substrates. As predicted, the reaction products fragmented similarly to I4:24 from *S. quitoense* in their collision-induced dissociation mass spectra (Figure S4.5) (Hurney, 2018) and co-eluted with metabolites in the plant extracts using two different chromatographic columns (Figure S4.6). TAIAT also catalyzed the acetylation of I3:24 to I4:26 (2,2,10,12) (Figure 4.8). The reaction product – I4:26 (2,2,10,12) – appeared to be identical to the plant-produced I4:26 based on chromatographic co-elution and mass spectrometric fragmentation (Figure S4.7 and Figure S4.9).
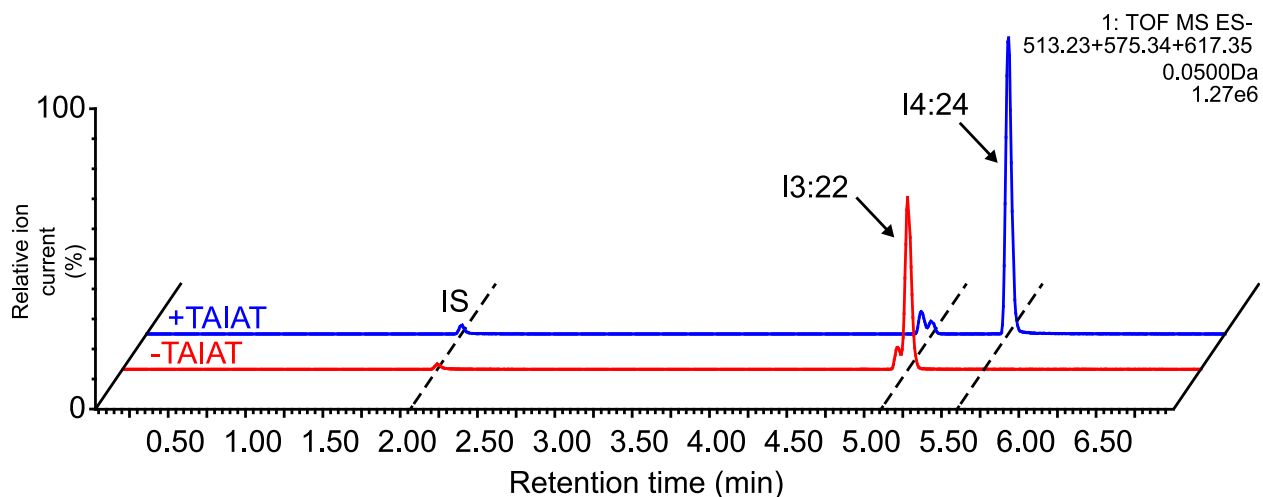
Figure 4.7. TAIAT catalyzes the acetylation of the acylinositol, I3:22.

Enzyme assays testing the ability of TAIAT to catalyze the forward reaction, combining TAIAT with acetyl-CoA and I3:22 (2,10,10) purified from *S. quitoense* results in the formation of I4:24 (2,2,10,10). The red trace represents the no enzyme control, whereas the blue trace represents the assay with enzyme included. ESI- mode was used for the analysis. The internal standard (IS) is telmisartan. The following *m/z* values were used to detect the substrate (*m/z*: 575.34), product (*m/z*: 617.36), and internal standard (*m/z*: 513.23). The samples were run on the 7 minute method described in the LC-MS section.
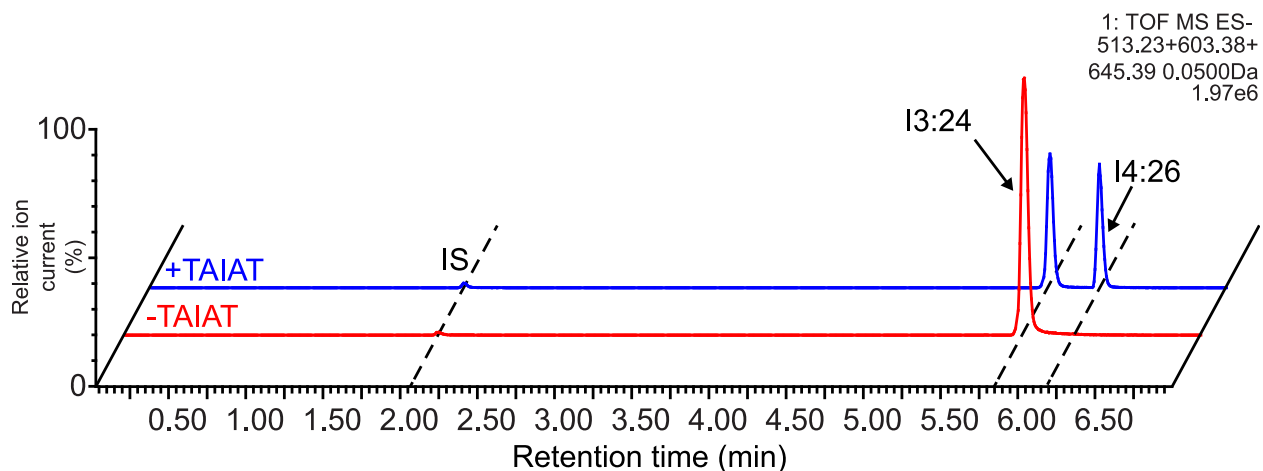
Figure 4.8. TAIAT catalyzes the acetylation of the acylinositol, I3:24.

Enzyme assays testing the ability of TAIAT to catalyze the forward reaction, combining TAIAT with acetyl-CoA and I3:24 (2,10,12) purified from *S. quitoense* results in the formation of I4:26 (2,2,10,12). The red trace represents the no enzyme control, whereas the blue trace represents the assay with enzyme included. ESI- mode was used for the analysis. LC-MS trace depicts formate adduct of I3:24 (*m/z*: 603.38), formate adduct of I4:26 (*m/z*: 645.39), and telmisartan (*m/z*: 513.23). The internal standard (IS) is telmisartan. The samples were run on the 7 minute chromatographic separation method described in LC-MS section.
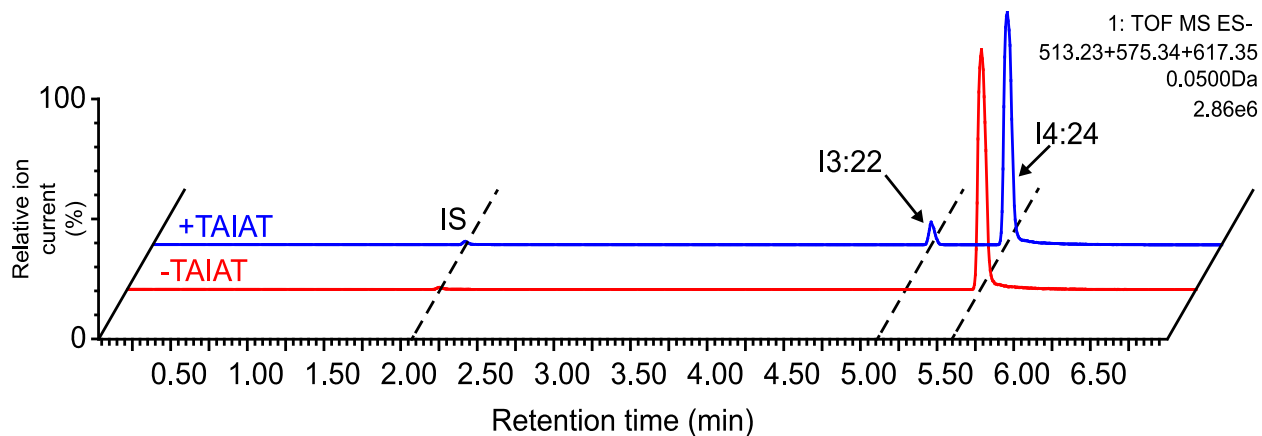
Figure 4.9. Reverse reaction with TAIAT catalyzes the de-acetylation of the diacetyl acylinositol, I4:24.

Enzyme assays testing the ability of TAIAT to catalyze the reverse reaction, combining TAIAT with free CoA and I4:24 (2,2,10,10) purified from *S. quitoense* results in the formation of I3:22 (2,10,10). The red trace represents the no enzyme control, whereas the blue trace represents the assay with enzyme included. ESI- mode was used for the analysis. The LC-MS traces depict the formate adduct of I3:22 (*m/z*: 575.34), formate adduct of I4:24 (*m/z*: 617.36), and telmisartan (*m/z*: 513.23). The internal standard (IS) is telmisartan. The samples were run on the 7 minute method described in LC-MS section.
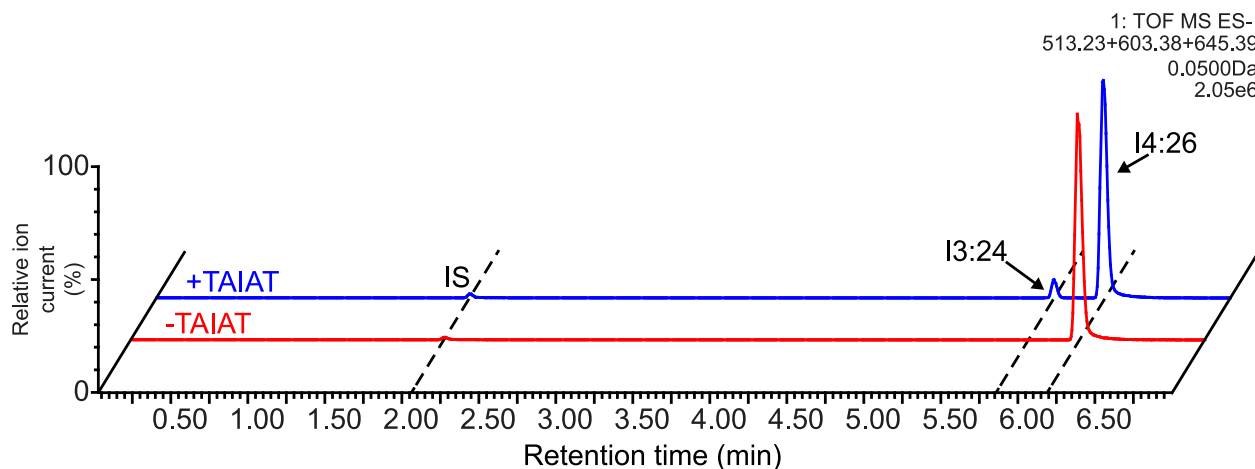
Figure 4.10. Reverse reaction with TAIAT catalyzes the de-acetylation of the diacetyl acylinositol, I4:26.

Enzyme assays testing the ability of TAIAT to catalyze the reverse reaction, combining TAIAT with free CoA and I4:26 (2,2,10,12) purified from *S. quitoense* result in the formation of I3:24 (2,10,12). The red trace represents the no enzyme control, whereas the blue trace represents the assay with enzyme included. ESI- mode was used for the analysis. The LC-MS traces depict the formate adduct of I3:24 (*m/z*: 603.38), formate adduct of I4:26 (*m/z*: 645.39), and telmisartan (*m/z*: 513.23). The internal standard (IS) is telmisartan. The samples were run on the 7 minute method described in LC-MS section.

TAIAT acetylated triacylinositols *in vitro*, however, because BAHD acyltransferases often use a variety of acyl CoAs, so we sought a non-biased approach to assay enzymatic activity of TAIAT (Landmann et al., 2011; Eudes et al., 2016; Fan et al., 2017). Many BAHD acyltransferases can catalyze the reverse reaction, conjugating an acyl chain to free CoA. These reverse reactions represent a neutral method to test the enzyme activity because the coenzyme A substrate is the same regardless of which acyl chain is removed. A BAHD acyltransferase would remove the acyl chain at the catalytically preferred position and chain length. Consistent with our previous hypothesis, TAIAT catalyzed the deacetylation of I4:24 to generate I3:22 in the presence of free CoA, making a product that co-eluted with and fragmented the same as I3:22 from *S. quitoense* (Figure 4.9, Figure S4.7, Figure S4.5). TAIAT also deacetylated the *in vivo* metabolite I4:26 to generate an I3:24 molecule that matched the *S. quitoense* acylsugar (Figure 4.10, Figure S4.7, Figure S4.5, and Figure S4.8). The combination of *in vitro* enzymatic activities and *in vivo* phenotypes provided compelling evidence that TAIAT catalyzes acetylation of triacylinositols in *S. quitoense*.

In vitro analysis of IAT

We hypothesized that IAT catalyzes the first step in acylinositol biosynthesis based on two factors. First, the reduction in total acylsugars in the VIGS results suggested that IAT is involved early in acylinositol biosynthesis. Second, the phylogenetic relationship of IAT to ASAT1 – an enzyme that catalyzes the first step in acylsucrose biosynthesis in cultivated tomato. I used *in vitro* biochemistry to determine whether this enzyme is an inositol acyltransferase. IAT encodes a predicted 435 amino acid protein, which was expressed in *E. coli* as a His-tagged fusion protein and purified using Ni-NTA. It was tested with *myo*-inositol plus decanoyl-CoA (nC10) or lauroyl-CoA (nC12). These substrates were chosen because they were the components

of acylsugars in *S. quitoense.* The fusion protein catalyzed the formation of mono-acylated

inositol (I1:10) *in vitro*, using decanoyl-CoA (nC10) and *myo*-inositol substrates (Figure 4.11).

The collision-induced dissociation of the molecule resulted in release of an nC10 acyl chain – a

fragment with an *m/z*: 171.14 (Figure S4.10), which represents the negatively charged

carboxylate of the nC10 acyl chain. IAT also catalyzed the formation of I1:12 using lauroyl-CoA

(nC12) and *myo*-inositol *in vitro* (Figure 4.12). This product contains an *m/z*: 199.17 fragment,

which matched the carboxylate ion of the nC12 acyl chain (Figure S4.11). IAT also possessed

minor enzymatic activity with iso (0.7%) and anteiso (1.2%) C5-CoA using *myo*-inositol (Note:

These percentages represent enzyme activity relative to IAT activity using nC10-CoA and *myo*-

inositol) (Figure S4.12 and Table S4.2). IAT accepted *myo*-inositol and medium length acyl-

CoAs as substrates, consistent with the NMR resolved structures of purified *in planta*
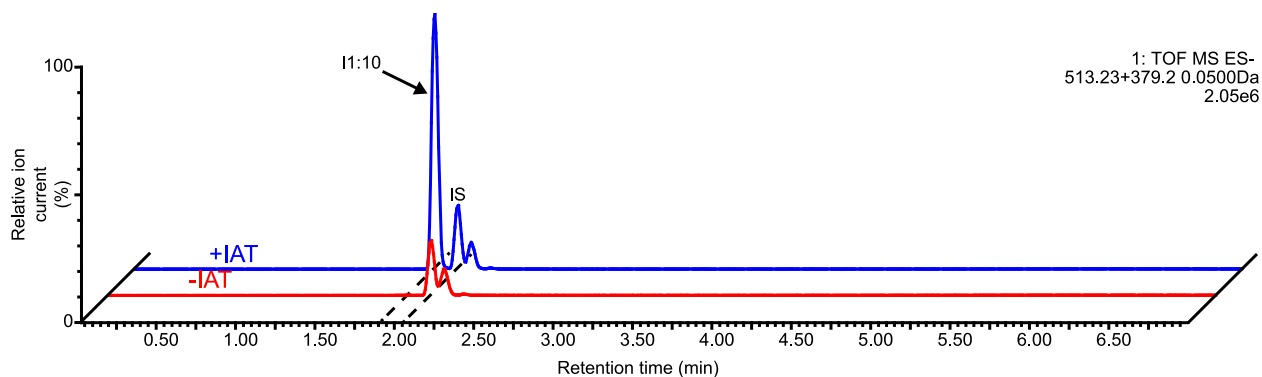
acylinositols.

Figure 4.11. IAT catalyzes the formation of I1:10 with *myo*-inositol and nC10-CoA *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions. nC10-CoA and *myo*-inositol were incubated with the enzyme. LC-MS trace shows the extracted ion chromatograms of telmisartan (*m/z*: 513.23) and the formate adduct of I1:10 (*m/z*: 379.20) (mass window: 0.05 Da). The red trace represents the no enzyme control, whereas the blue trace represents the complete assay with enzyme included. The internal standard is telmisartan. The samples were run on the 7 minute method described in LC-MS section.
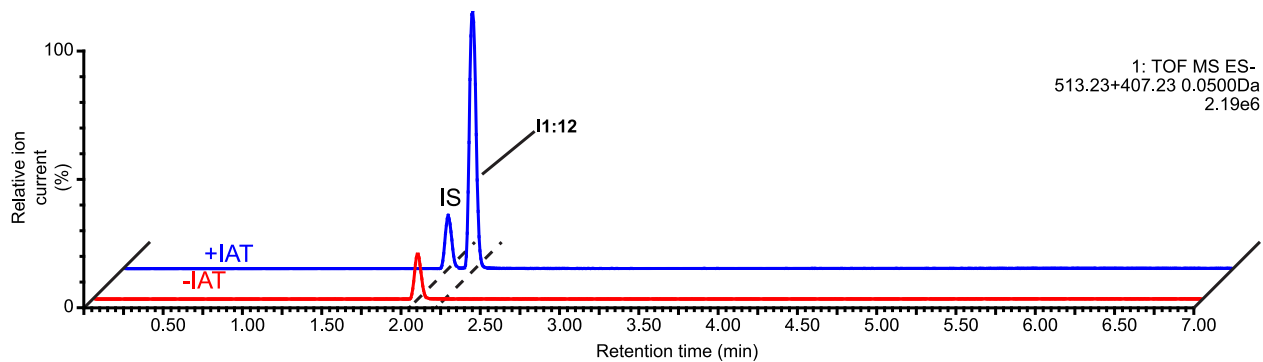
Figure 4.12. IAT catalyzes the formation of I1:12 using *myo*-inositol and nC12-CoA *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions (Putative ortholog to Solyc07g043670). nC12-CoA and *myo*-inositol were incubated with the enzyme. LC-MS trace shows the extracted ion chromatograms of telmisartan (*m/z*: 513.23) and the formate adduct of I1:12 (*m/z*: 407.23) (mass window: 0.05 Da). The red trace represents the no enzyme control, whereas the blue trace represents the complete assay with enzyme included. The internal standard is telmisartan. The samples were run on the 7 minute method described in LC-MS section.

Another characteristic of some ASAT1 orthologs was the ability to catalyze consecutive acylations. Previous work demonstrated both *Solanum nigrum* and *S. lycopersicum* ASAT1 can acylate sucrose twice using the same acyl-CoA (unpublished). IAT also possesses this ability when fed nC10-CoA and *myo*-inositol, generating a small di-acylated inositol peak (I2:20) (Figure 4.13), producing the same characteristic nC10 fragment ion by mass spectrometry (*m/z*: 171.14). The diacylation is also catalyzed with nC12-CoA as a substrate. (Figure S4.17). Interestingly, when IAT and TAIAT are incubated with nC10 and C2-CoA, they generate a peak with an *m/z* consistent with I3:22 (Figure S4.18). However, the product did not co-elute with I3:22 from *S. quitoense*.

If IAT catalyzes the first step in acylinositol biosynthesis, we posited that it would display similar donor and acceptor substrate affinities as biochemically characterized acylsucrose acyltransferases. K_m values for nC10-CoA, nC12-CoA and *myo*-inositol were determined to be 9.5 ± 2.4 µM, 16.6 ± 6.6 µM, and 4.4 ± 1.1 mM, respectively (95% confidence interval) (Figure S4.13 and Figure S4.14). These findings were consistent with previously published results on acylsucrose acyltransferases. For example, the apparent K_m of Sl-ASAT1 for sucrose was determined to be 2.3 mM (Fan et al., 2016), similar to the 4.4 mM observed for IAT (Figure S4.13). ASAT3 enzymes from multiple species had acyl-CoA K_m values in the low µM levels (2-10 µM), while Sl-ASAT1 and Sl-ASAT2 possessed slightly higher K_m values at 20-50 µM (Schilmiller et al., 2015; Fan et al., 2016). These values are similar to the 9.5 and 17 µM values for IAT with nC10-CoA and nC12-CoA, respectively. Taken together, the similarities of the IAT K_m results with characterized ASATs are consistent with a role in acylinositol biosynthesis.
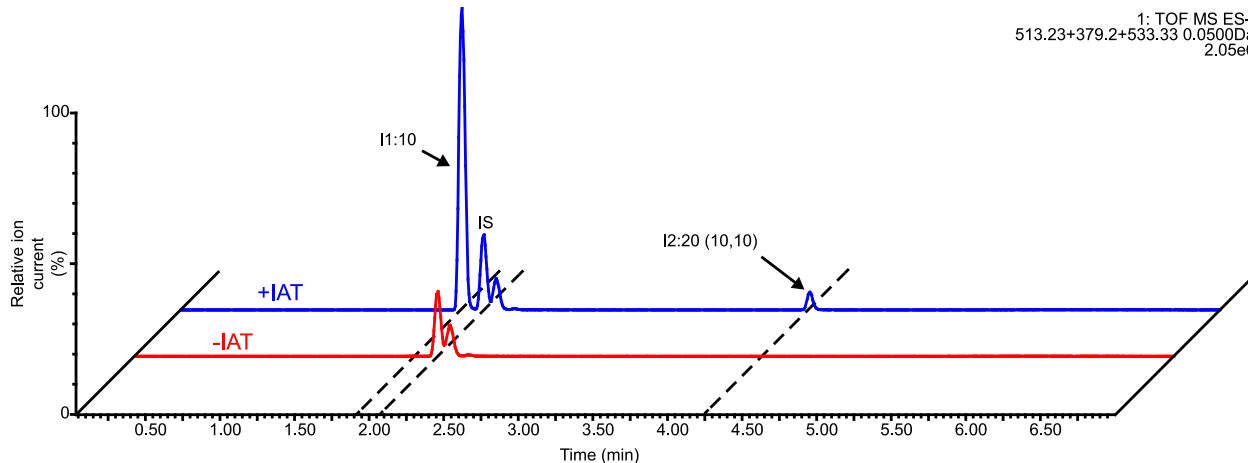
Figure 4.13. IAT can catalyze consecutive acylations of *myo*-inositol using nC10-CoA to generate I2:20 (10,10) *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions with *myo*-inositol and nC10-CoA. Extracted ion chromatograms of telmisartan (*m/z*: 513.23), formate adduct of I1:10 (*m/z*: 379.20), and formate adduct of I2:20 (10,10) (*m/z*: 533.33) (mass window: 0.05 Da). The red trace represents the no enzyme control, whereas the blue trace represents the complete assay with enzyme included. The internal standard is telmisartan. The samples were run on a 7 minute method described in LC-MS section.

We hypothesized that IAT is a promiscuous enzyme and may retain the ability to use some substrates of previously identified ASATs. Previously characterized ASAT1 orthologs primarily use sucrose as a substrate, so we tested IAT activity with sucrose. While IAT weakly catalyzed the acylation of sucrose with nC10-CoA (1.9% of *myo*-inositol activity), it possessed no detectable activity with glucose (Figure 4.14, Figure S4.15, Table S4.2). I also tested four stereoisomers of inositol – the naturally occurring *myo*-, *scyllo*- and D-*chiro*-inositol and synthetic *epi*-inositol. There was measurable enzymatic activity with D-*chiro*-inositol (~1.2% of *myo*-inositol activity) (Fig S4.16 and Table S4.2), but no detectable activity with *scyllo*- or *epi*-inositol. Taken together, these results revealed that IAT accepts a handful of acceptor substrates including *myo*-inositol, D-*chiro*-inositol, and sucrose – a trait is consistent with many previously characterized BAHD acyltransferases.

The cultivated tomato *S. lycopersicum* contains an IAT ortholog: this chromosome 7 gene, Solyc07g043670, is a paralog of the sucrose acylating enzyme, Sl-ASAT1, but not known to be involved in acylsucrose biosynthesis. The tomato gene encodes a full-length open reading frame and other characteristics consistent with other BAHD acyltransferases (Figure S4.1). Given the evolutionary relationship and 78% amino acid identity to IAT, we hypothesized that Solyc07g043670 might acylate *myo*-inositol (Table S4.1). Indeed, it showed minute *in vitro* enzymatic activity with *myo*-inositol and nC10-CoA (Figure S4.19).
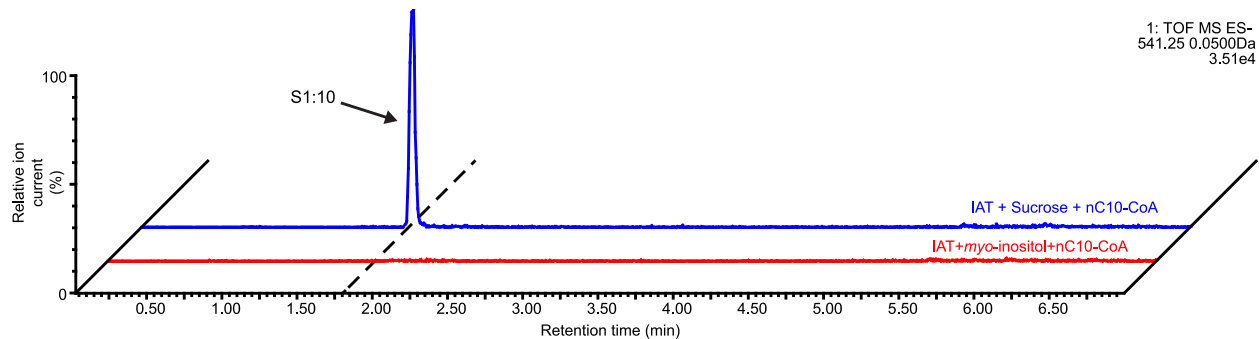
Figure 4.14. IAT catalyzes the formation of S1:10, utilizing nC10-CoA and sucrose *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions. nC10-CoA and sucrose were incubated with the enzyme. LC-MS trace shows the extracted ion chromatograms of the formate adduct of S1:10 (*m/z*: 541.25). The red trace represents the no sucrose control, whereas the blue traces represent replicates of the complete assays with enzyme included. The samples were run on the 7 minute method described in LC-MS section.

If IAT is involved in acylinositol biosynthesis, the simplest hypothesis is that the acyl chain position of the *in vitro* monoacylinositol should match one of the C10 or C12 chains of *S. quitoense* acylsugars. IAT enzyme assays were scaled up to generate enough metabolite for NMR analysis. The NMR analysis was done in collaboration with Steven Hurney and Thilani Anthony of the laboratory of Dr. A. Daniel Jones at Michigan State University. Based upon experience with monoacylsucrose chain migration at pH 7, I ran the reaction at pH 6 to stabilize the assay product but ended up with a mixture of monoacylinositol isomers upon processing and analyzing the sample despite acidification. Four chromatographically distinct monoacylated products were observed in approximately 2:15:22:61 peak area ratios. ^{1}H-NMR analysis revealed four positional isomers at relative peak integration of 2:15:19:64, which agreed well with the LC-MS peak area ratios. Based on the correlation of NMR and MS data, we proposed that the enzyme assay-generated monoacylinositol isomers at position 2 eluted first, followed by that at position 5, position 1/3 and finally position 4/6 (note that the acylations at the 1 and 3 positions (and 4/6 positions) yield enantiomers (mirror images) that are not distinguished by NMR owing to the mirror plane symmetry). Serial dilutions of monoacylinositol LC-MS samples revealed similar proportions of the isomers to those calculated by NMR (Figure S4.20 and 4.15). The strong correlation between the NMR integrations and LC-MS peak areas were used to infer the elution order of these different monoacylinositol isomers.

To determine the structure of the most abundant enzyme assay product, I separated it from the other isomers using a different semipreparative HPLC method designed to resolve the monoacylinositol isomers. We analyzed the product using 1H, 13C, gCOSY, gHSQC, gHMBC and NOESY NMR experiments.
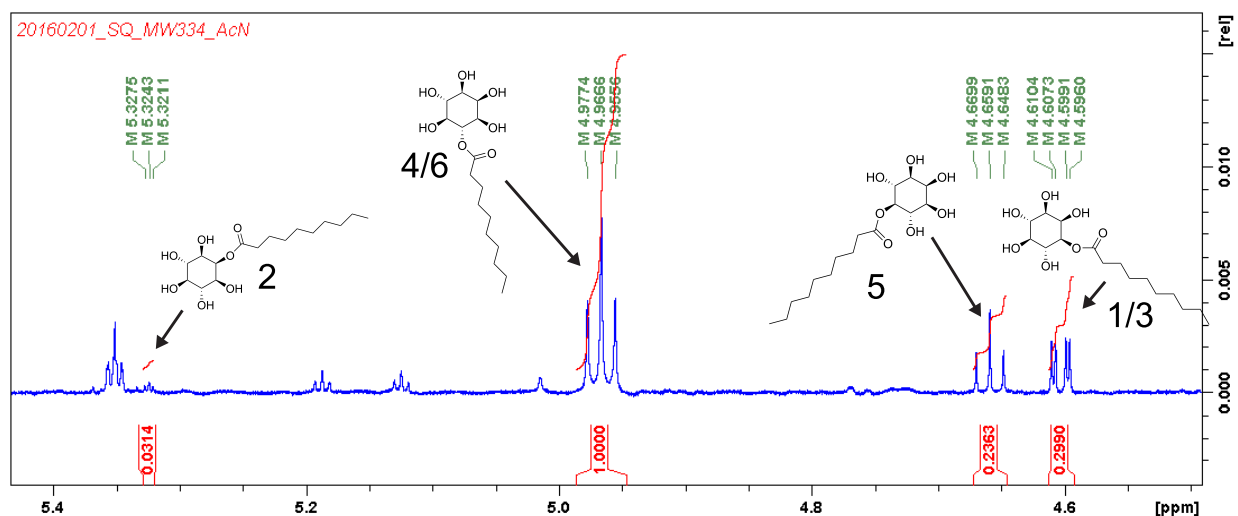
Figure 4.15. ^{1}H-NMR spectrum of ring hydrogen region for C10 mono-acylated myo-inositol.

Relative integrations of the NMR signals correspond to the following acylation positions: 2 position, 2% of the product at a chemical shift of 5.32 ppm.; the 4/6 positions of *myo*-inositol, 64% of the product at a chemical shift of 4.97 ppm; the 5 position of *myo*-inositol, 15% of the product at a chemical shift of 4.66 ppm; finally, the 1/3 position of *myo*-inositol, 19% of the product at a chemical shift of 4.60 ppm. Numbers depict the acylation position on *myo*-inositol.

Table 4.2. NMR analysis of purified sample of I1:10 (position 4/6).
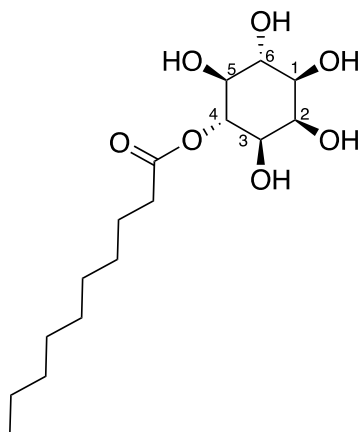
C10 mono-acylated myo-inositol compound

Chemical Formula: $C_{16}H_{30}O_7$

Exact Mass: 334.1992

NMR (900 MHz, Acetonitrile-d_3)

Carbon # (group)	^1H (ppm)	^{13}C (ppm) (from HSQC and HMBC)
1 (CH)	3.30 (dd, J = 9.6, 2.7 Hz, 1H)	72.57
2 (CH)	3.90 (t, J = 2.6 Hz, 1H)	73.44
3 (CH)	3.45 (dd, J = 9.6, 2.7 Hz, 1H)	71.20
4 (CH)	4.97 (t, J = 9.8 Hz, 1H)	75.82
- 1 (CO)	-	174.44
- 2 (CH_2)	2.33 (t, J = 7.5 Hz, 1H)	34.96
- 3 (CH_2)	1.59 (dt, J = 15.1, 7.5 Hz, 4H)	25.64
- 4 (CH_2)	1.29 (tdd, J = 22.9, 16.6, 9.1 Hz, 14H)	30.11
- 5 (CH_2)		30.11
- 6 (CH_2)		30.11
- 7 (CH_2)		30.11
- 8 (CH_2)		32.60
- 9 (CH_2)		23.29
- 10 (CH_3)	0.88 (t, J = 7.0 Hz, 3H)	14.30
5 (CH)	3.22 (t, J = 9.5 Hz, 1H)	72.67
6 (CH)	3.50 (t, J = 9.6 Hz, 1H)	71.02

NMR analysis revealed the major *in vitro* product of IAT possesses an nC10 acyl chain at the R4 position of *myo*-inositol (Table 4.2). This contrasts with *S. quitoense* acylsugars, which contain medium-length chains (nC10 or nC12) at the R2 and R1/3 positions of *myo*-inositol. The discrepancy between the *in vitro* product and *in vivo* metabolite results suggested that the major IAT R4 acylated product does not represent the substrate for the next biosynthetic step.

An alternative hypothesis is that a modified *myo*-inositol is the *in vivo* substrate for IAT. Inositol phosphates are commonly observed in eukaryotes, including as substrates for metabolic enzymes. I tested whether IAT could acylate *myo*-inositol phosphorylated at the 1 or 3 position using nC10-CoA; however, no acylation products were detected (Figure 4.16.A). A positive control for enzymatic activity using nC10-CoA with *myo*-inositol made the I1:10 product, indicating that the IAT enzyme preparation was active (Figure 4.16.B). *In vitro* assays failed to support the hypothesis that *myo*-inositols phosphorylated at the 1 or 3 positions are IAT substrates.

The mixture of monoacylinositols in the NMR samples and rearrangement of enzyme assay products led us to ask if pH can affect the monoacylinositol products that accumulate. Assays at pH 6 analyzed by LC-MS show a single I1:10 peak that increases in intensity over time (10, 15, 30, and 60 minutes)(Figure 4.17). Enzyme assays at pH 6 vs. pH 8 show differing product profiles *in vitro*. The *in vitro* reactions at pH 8 produced a mixture of mono-acylated inositol isomers, with position 4/6 isomers the most abundant (Figure 4.18). The abundance of the isomers with acyl chains at R1/3, 2, and 5 positions of mono-acylated inositol increased with time. The single I1:10 product from Table 4.2 was also incubated at pH 8, where rearrangement of the monoacylinositol occurred without an enzyme present (Figure 4.19).
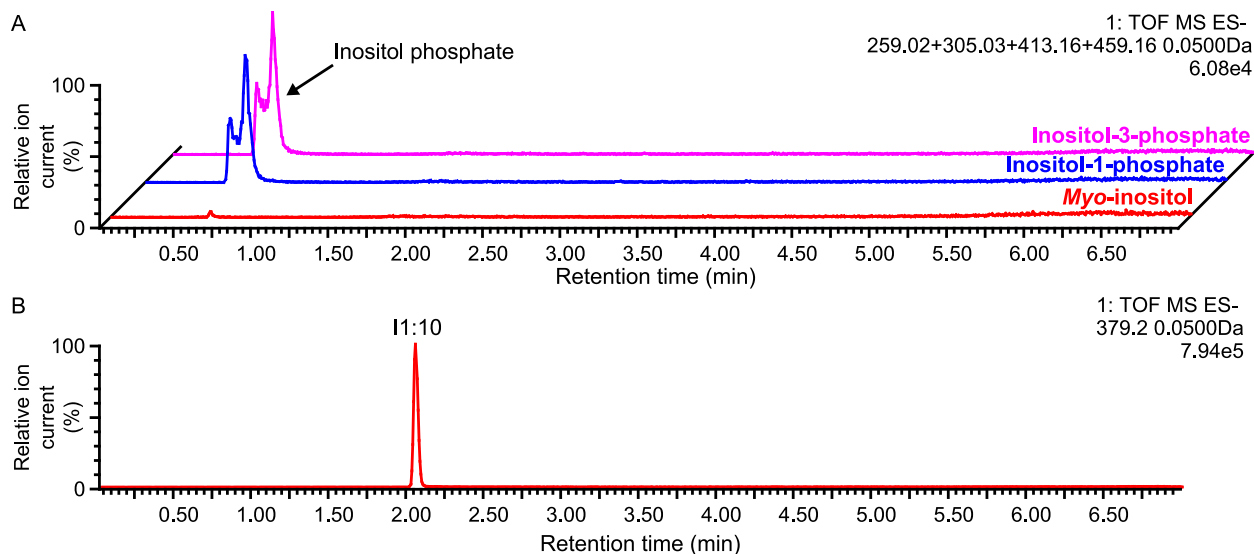
Figure 4.16. Enzyme assays testing IAT activity with inositol phosphate substrates.

(A) LC-MS analysis of assay testing enzymatic activity of inositol phosphate and nC10-CoA as substrates with IAT. Inositol-1-phosphate and Inositol-3-phosphate were used as substrates. Extracted ion chromatograms depict inositol phosphate substrates and putative acylated products: Inositol phosphate (*m/z*: 259.0 and 305.02), and I1:10-P (*m/z*: 412.16, 413.16, and 459.17). (B) LC-MS analysis of positive control reaction run simultaneously with the inositol phosphate reactions. The samples were run on the 7 minute method described in LC-MS section.
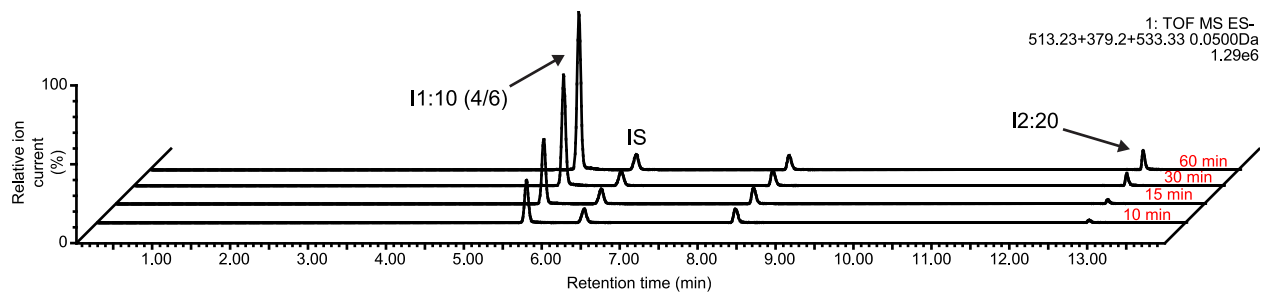
Figure 4.17. IAT produces a single I1:10 peak at pH 6 *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assay reactions run at pH 6. nC10-CoA and *myo*-inositol were incubated with the enzyme. Enzyme assays were run from 10-60 minutes from front to back. The LC-MS trace shows the extracted ion chromatogram for telmisartan (*m/z*: 513.23), I1:10 formate adduct (*m/z*: 379.20), and I2:20 formate adduct (*m/z*: 533.33). The samples were run on 14 minute I1:10 method described in LC-MS section.
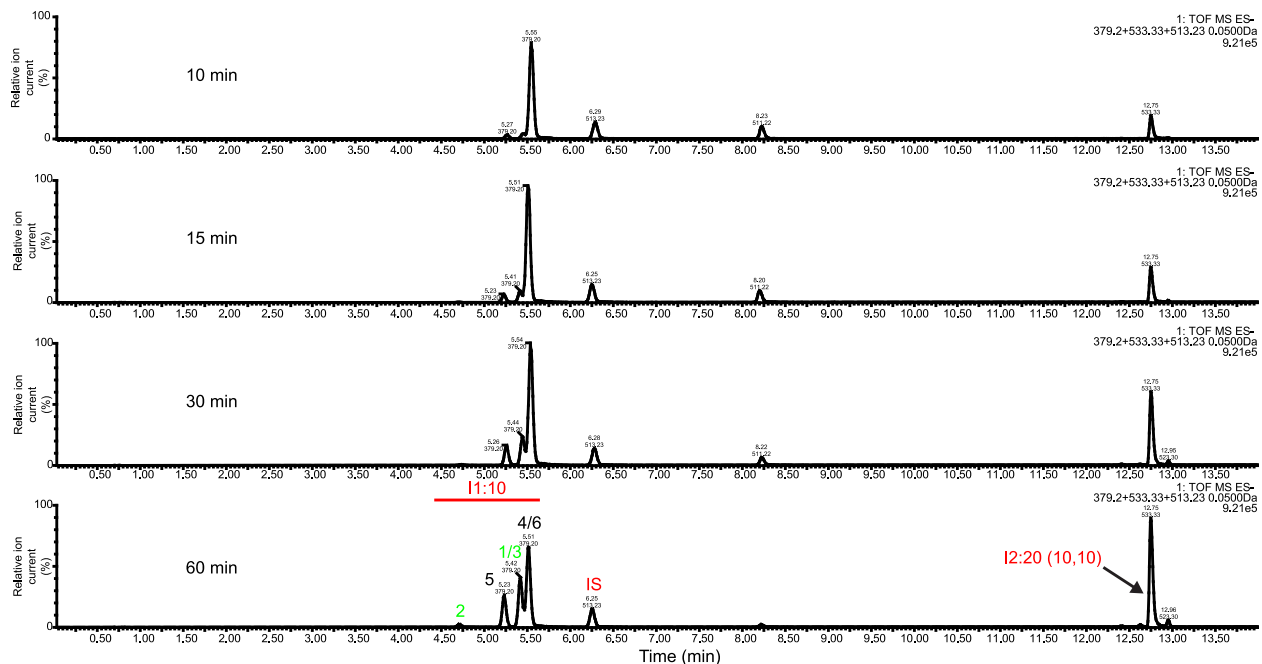
Figure 4.18. IAT produces multiple I1:10 peaks at pH 8.0 *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions run at pH 8. nC10-CoA and *myo*-inositol were incubated with the enzyme. Assay length of the samples ranges from 10-60 minutes from top to bottom. The LC-MS trace shows the extracted ion chromatogram for telmisartan (*m/z*: 513.23), I1:10 formate adduct (*m/z*: 379.20), and I2:20 formate adduct (*m/z*: 533.33). The samples were run on 14 minute I1:10 method described in LC-MS section. Monoacylinositol isomers are labeled; isomers with medium acyl chains present at the correct *in vivo* position are highlighted in green.
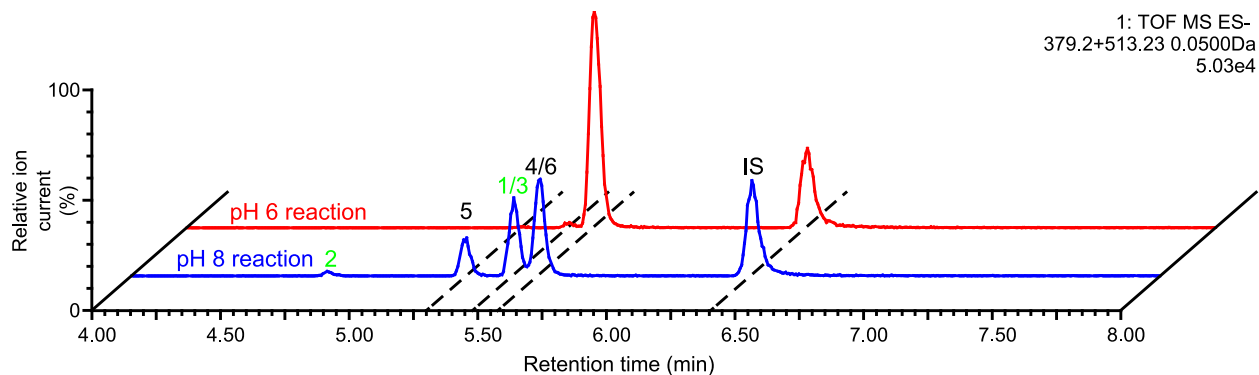
Figure 4.19. IAT-produced I1:10 rearranges non-enzymatically at pH 8, but not pH 6

ESI- mode LC-MS analysis of NMR product, I1:10 (4/6), after incubation at pH 6 or 8 for 60 minutes. The LC-MS trace shows the extracted ion chromatogram for telmisartan (*m/z*: 513.23), and I1:10 formate adduct (*m/z*: 379.20). Monoacylinositol isomers are labeled; isomers with medium acyl chains present at the correct *in vivo* position are highlighted in green. Figure displays from 4-8 minutes of 14 minute I1:10 LC method. The red trace represents the monoacylinositol reaction run at pH 6, while the blue trace represents the monoacylinositol reaction run at pH 8. The internal standard was telmisartan.

Another interesting observation in IAT enzymatic reactions was the synthesis of the I2:20 product from nC10-CoA and *myo*-inositol. The amount of di-acylated I2:20 product increased in the pH 8 reactions compared to the pH 6 reactions (Figure 4.20). More interesting is that a single major diacylated peak accumulated rather than equal mixtures of diacylated inositols, suggesting some level of specificity for the diacylation activity of IAT. This may be related to the original activity of the *S. sinuata* ASAT1 homolog that catalyzes acylation of S1:6 to S2:12 *in vitro* and *in planta* (Moghe et al., 2017).

In vitro analysis of c38687_g1_i1

I previously identified another acetyltransferase candidate related to TAIAT, c38687_g1_i1. Many *S. quitoense* acylinositols possess two acetyl groups, which suggests that more than one acetyltransferase is necessary for acylinositol biosynthesis. We expressed c38687_g1_i1 and tested its enzymatic activity *in vitro*. c38687_g1_i1 and IAT catalyzed the formation of I2:12 when incubated together *in vitro* (Figure 4.21). The same product is not catalyzed by the combination of IAT and TAIAT (Figure 4.21). A sequential assay revealed that c38687_g1_i1 catalyzed the acetylation of the I1:10 *in vitro* (Figure 4.22). This preliminary data suggested a possible route forward in acylinositol biosynthesis but requires further analysis.
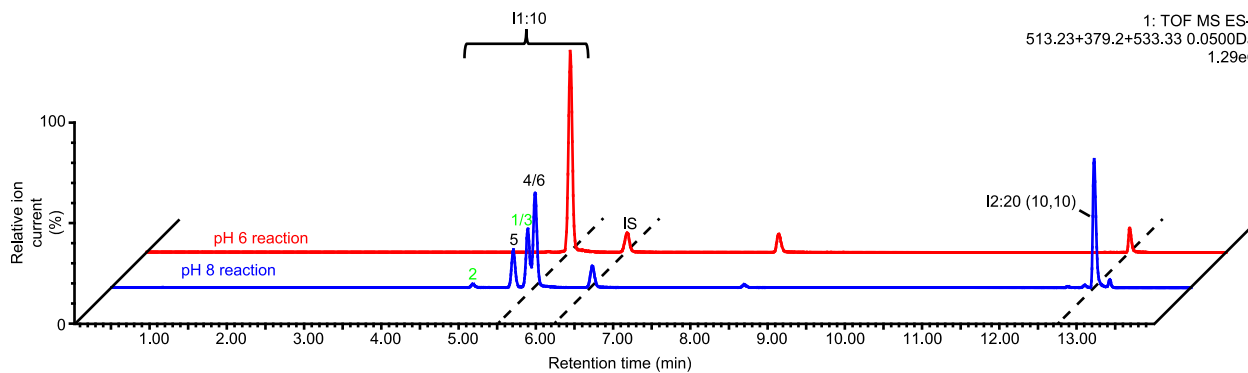
Figure 4.20. Comparison of IAT *in vitro* assays at pH 6 and 8

ESI- mode LC-MS analysis of *in vitro* IAT assays reactions incubated nC10-CoA and *myo*-inositol at pH 6 or 8 for 60 minutes. The LC-MS trace shows the extracted ion chromatogram for telmisartan (*m/z*: 513.23), I1:10 formate adduct (*m/z*: 379.20), and I2:20 formate adduct (*m/z*: 533.33). Monoacylinositol isomers and I2:20 product are labeled: isomers with medium acyl chains present at the correct *in vivo* position are highlighted in green. The samples were run on 14 minute I1:10 method described in LC-MS section.
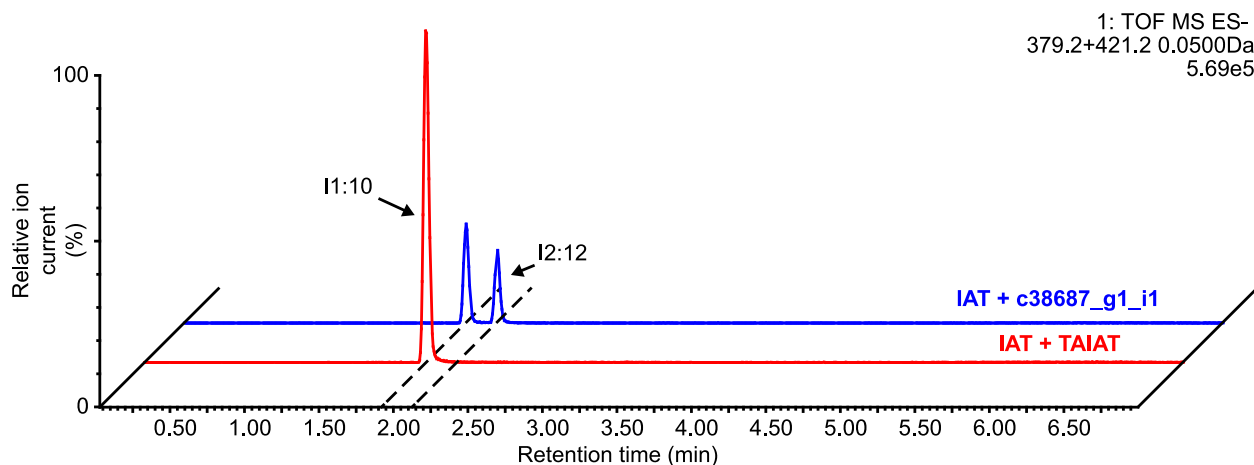
Figure 4.21. Combined assay with c38687_g1_i1 and IAT together catalyze the formation of I2:12

LC-MS analysis (ESI-) of enzyme assays containing IAT and c38687_g1_i1 or TAIAT after incubation with acetyl-CoA and nC10-CoA. The LC-MS trace shows the extracted ion chromatogram for I1:10 formate adduct (*m/z*: 379.20), and I2:12 formate adduct (*m/z*: 421.20). The red trace represents the reaction containing IAT and TAIAT, while the blue trace represents the reaction containing IAT and c38687_g1_i1. The samples were run on the 7 minute method described in LC-MS section.
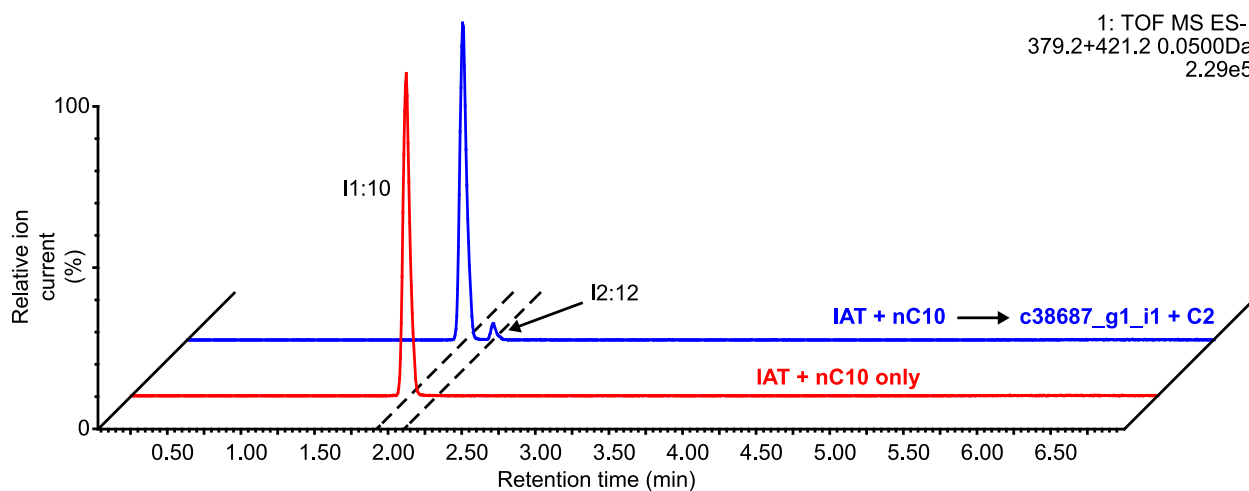
Figure 4.22. c38687_g1_i1 catalyzes the acetylation of the I1:10 product of IAT

LC-MS analysis (ESI-) of consecutive enzyme assays containing IAT, followed by inactivation with heat, then c38687_g1_i1 with nC10-CoA and C2 CoA, respectively. The LC-MS trace shows the extracted ion chromatogram for I1:10 formate adduct (*m/z*: 379.20), and I2:12 formate adduct (*m/z*: 421.20). The red trace represents a consecutive reactions with water added instead of c38687_g1_i1, while the blue trace represents the test assay. The samples were run on the 7 minute method described in LC-MS section.

Discussion

I used a combination of existing transcriptomic data, analytical chemistry, *in vitro* and *in planta* analyses to identify and test functions of two BAHD acyltransferases involved in *S. quitoense* acylinositol biosynthesis. We established a successful VIGS protocol using PDS in *S. quitoense,* which was used for analysis of candidates in acylinositol biosynthesis. VIGS silencing of trichome-expressed transcripts led to phenotypes for both IAT and TAIAT. These results are consistent with the hypothesis that these BAHD acyltransferases are involved in acylinositol biosynthesis

Two lines of evidence indicated that TAIAT catalyzes acetylation of triacylinositols in *S. quitoense*. Silencing of TAIAT caused a change in the ratios of tri- to tetraacylated inositols, which is consistent with the hypothesis that TAIAT acetylates I3:22 and I3:24. *In vitro* enzyme assays revealed that TAIAT acetylated I3:22 and I3:24 purified from *S. quitoense* leaf surface extracts to form I4:24 and I4:26. This enzymatic activity seems similar to that of the closest non-paralog of TAIAT – Sl-ASAT4 – an enzyme responsible for acetylation of triacylsucroses. Analysis of *S. quitoense* RNAseq data revealed multiple paralogs of TAIAT that could play a role in acylinositol biosynthesis – two of which are expressed in the trichomes: c38687_g1_i1 and c38687_g2_i1 (c#####_g#_i# represents distinct transcripts in the RNAseq assembly). The trichome expression in combination with *in vitro* acetylation activity of c38687_g1_i1 suggests this enzyme could fit somewhere in acylinositol biosynthesis.

Analysis of IAT led to a more complicated story. A reduction in total acylsugars in VIGS plants suggested that this enzyme is involved in acylinositol metabolism. This is a result expected for silencing a gene that functions early in an acylsugar biosynthetic pathway. Indeed, we previously observed low acylsucrose phenotypes in silencing of ASAT1 and ASAT2 in

157

cultivated tomato (RNAi) and *S. sinuata* plants (VIGS). The same phenotype was observed in introgression lines that replaced Sl-ASAT2 and Sl-ASAT3 with SpASAT2 and SpASAT3 into a cultivated tomato background, respectively (Schilmiller et al., 2015; Fan et al., 2016). My *in vitro* enzyme assays provided less clear results. On the positive side, I showed that IAT acylates *myo*-inositol using nC10 and nC12-CoAs and possesses K_m values for acyl-CoAs and *myo*-inositol that match previously characterized ASATs (Schilmiller et al., 2015; Fan et al., 2016). NMR analysis revealed the reaction product was acylated at the R4 position of *myo*-inositol – a position that does not contain a medium acyl chain in acylinositols purified from *S. quitoense.* IAT does not use *myo*-inositol modified by a single phosphate at the R1 or R3 position as a substrate. However, IAT assays resulted in a mixture of monoacylinositol isomers at pH 8, which contrasts with the single isomer produced at pH 6. The mixture of monoacylinositol isomers is also produced at pH 8 from a purified single I1:10 NMR product in the absence of IAT.

Multiple hypotheses are consistent with the IAT findings. One hypothesis is that VIGS in *S. quitoense* cross-silenced another transcript. This hypothesis seems unlikely for two reasons. First, there are no other significant BLAST hits when using Sl-ASAT1 or even IAT as a query for the *S. quitoense* trichome and petiole RNAseq assembly. Second, I found reduction in total acylinositol levels with two distinct IAT VIGS constructs, reducing the likelihood of artifactual off-target silencing leading to the phenotype. Genetic evidence points to IAT involvement in acylinositol biosynthesis. I suggest a few hypotheses to explain the discrepancy between *in vitro* and *in planta* data.

One hypothesis is that IAT acts later in the pathway rather than the first step in acylinositol biosynthesis. This is not without some precedent: the activity of the ancestor of Sl-ASAT1 was acylation of monoacylsucrose to generate diacylsucrose. Since the last common

ancestor between *Hyoscyamus niger* and cultivated tomato, the ancestor of Sl-ASAT1 has shifted

from using monoacylsucrose to sucrose, but retains the ability to utilize the mono-acylated

substrate (Moghe et al., 2017). We have not yet identified other enzymes able to use *myo*-

inositol. This analysis is complicated further by the absence of putative orthologs to SsASAT1 or

Sl-ASAT1. A plausible hypothesis is that another enzyme catalyzes the first step in acylinositol

biosynthesis, and it may be unrelated to acylsugar biosynthesis in other Solanaceae species. This

would require further mining of the BAHDs mentioned previously to identify the true enzyme.

An alternative hypothesis is that molecular rearrangement of I1:10 plays a role in

acylinositol biosynthesis. Monoacylinositol rearrangement occurs under basic conditions,

consistent with pH values determined for the cytoplasm of plant cells (Smith and Raven, 1979;

Roberts et al., 1980). We hypothesize that if the enzyme that catalyzes the second step in

acylinositol biosynthesis is specific for certain monoacylinositol isomers, it could exhibit

substrate specificity for one isomer over the others. Another factor that could affect substrate

availability are the previously characterized acylsugar acylhydrolase enzymes (Schilmiller et al.,

2016). ASHs are hydrolases able to remove specific acyl chains from acylsugars that have been

characterized in cultivated tomato and *Solanum pennellii* (Schilmiller et al., 2016). My

preliminary VIGS data indicated that silencing a putative ortholog of ASH1 results in the

accumulation of acylinositol intermediates, which could be dead-end products. Unpublished

analysis performed in our lab suggests that some tomato and *S. pennellii* ASHs can use

monoacylsucroses as substrates. Turnover of acylinositol intermediates in combination with

accumulation of different monoacylinositol isomers could affect the substrates available to later

biosynthetic enzymes. Some combination of these factors could be the key to unlocking the

entire biosynthetic pathway.

Materials and Methods

Heterologous protein expression and purification from *Escherichia coli*

Heterologous protein expression was achieved using pET28b(+) (EMD Millipore, Burlington, MA), in which open reading frames for the enzymes were cloned into using doubly digested vectors of either BamHI/XhoI (TAIAT/c38687_g1_i1), NheI/XhoI (IAT), or NheI/NotI (Solyc07g043670). The doubly digested vectors were assembled with a single fragment containing the ORF containing 5' and 3' adapters for Gibson assembly using 2x NEB Hifi Mastermix (NEB, Ipswich, MA) or ligated into pET28b using DNA ligase (NEB, Ipswich, MA) in the case of Solyc07g043670. The finished constructs were transformed into BL21 Rosetta (DE3) cells (EMD Millipore, Burlington, MA) and verified using colony PCR and Sanger sequenced using T7 promoter and terminator primers.

LB overnight cultures with kanamycin (50 µg/mL) and chloramphenicol (33 µg/mL) were inoculated with a single colony of the bacterial strain containing the desired construct and incubated at 37 °C, 225 rpm, overnight. Larger cultures were inoculated 500:1 with the same antibiotics and incubated at the same temperature and speed. OD600 of the cultures was monitored until between 0.5 and 0.8. Cultures were chilled on ice for 15 minutes, at which IPTG was added to a final concentration of 50 µM for all BAHD sequences except for IAT, which was incubated with 300 µM IPTG. Cultures were incubated at 16 °C, 180 rpm for 16 hours.

Note:All of the following steps are on ice.

Cultures were centrifuged at 4,000g for 10 minutes to collect the cells and repeated until all the culture was processed (4 °C). The cell pellets were resuspended in 25 mL of extraction buffer (50 mM NaPO$_4$, 300 mM NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, pH 8.0) by

vortexing. The cell suspension was sonicated for 8 cycles (30 seconds on, intensity 4, 30 seconds on ice). The cellular extracts were centrifuged at 30,000g for 10 minutes. The supernatant was transferred into another tube and centrifuged again at the same speed and duration. Ni-NTA resin (Qiagen, Hilden, Germany) was centrifuged at 1,000g for 1 minute, resuspended in 1 mL of extraction buffer. The slurry was centrifuged again at 1,000g for 1 minute and the supernatant was decanted. The resin was resuspended using the crude extract and incubated at 4 °C, nutating for 1 hour. The slurry was centrifuged at 3,200g for 5 minutes, and supernatant decanted. The resin was resuspended in 5 mL of extraction buffer and transferred to a gravity flow column (Biorad, Hercules, CA). After loading, the resin was washed with 3 column volumes of extraction buffer (~30 mL). The resin was further washed with 1 column volume of wash buffer (extraction buffer with 40 mM imidazole). The remaining protein was eluted and collected using 2 mL of elution buffer after a 1 minute incubation with the resin. The elution was diluted into 15 mL of storage buffer (extraction buffer, but no imidazole). This elution was concentrated using 10 kDa centrifugal filter units (EMD Millipore, Burlington, MA), and repeated until diluted 1,000 fold. An equal volume of 80% glycerol was added to the elution, mixed, and stored at -20 °C.

General enzyme assays

Assays were run in 100 mM sodium phosphate, pH 6 or pH 8 at a total volume of 60 µL with pH 6 as the default unless otherwise stated. Acyl-CoAs were added to a final concentration of 100 µM. Non-acylated acceptors were added at a final concentration of 1 mM, while acylsugar acceptors – such as I3:22 – were dried down using a speed vac and resuspended in an ethanol:water mixture (1:1) with 1 µL added to the reaction. 6 µL of enzyme was added to each reaction. The assays were incubated at 30 °C for 30 minutes unless otherwise stated. After the

incubation, 2 volumes of stop solution – composed of a 1:1 of acetonitrile and isopropanol with 0.1% formic acid and 1 μM telmisartan as internal standard (Sigma-Aldrich, St. Louis, MO) – were added to the assays and mixed by pipetting. Reactions were stored in the -20 °C freezer for 20 minutes and centrifuged at 17,000g for 5 minutes. The supernatant was transferred to LC-MS tubes and stored at -20 °C.

For kinetic analysis, conditions were used such that the enzyme amount and reaction time were in the linear range. For each substrate (nC10, nC12, and *myo*-inositol), the other substrate was held at saturating concentrations. The reactions were run for 20 minutes, performed in triplicate, and stopped by the addition of 2 volumes of stop solution. Samples were analyzed as described in LC-MS section. Nonlinear regression was performed using standard Michaelis-Menten kinetics model in Graphpad Prism8 (Graphpad Software).

Mono-acylated enzyme assays for NMR

Assays for both NMR analyses were run in 100 mM Ammonium acetate, pH 6.0 at a total volume of 60 mL. nC10-CoA and *myo*-inositol were added to a final concentration of 400 μM and 30 mM, respectively. 6 mL of enzyme solution was added to the reactions purified from 6L of *E.coli* culture. Reactions were incubated at 30 °C for 3 hours. After incubation, 2 volumes of stop solution – 1:1 acetonitrile:isopropanol containing 0.1% formic acid – were added to the assays and mixed by pipetting. Reactions were dried down using the speed vac. In both cases, the products were purified using a Waters 2795 Separations module equipped with LKB Bromma 2211 Superrac fraction collector with automated fraction collection.

NMR analysis of monoacylinositols

For the single mono-acylated NMR product analysis: the residue was resuspended in a 6 mL mixture of water and acetonitrile (6:4) with 0.1% formic acid by vortexing and transferred to LC-MS vials for semipreparative purification. Samples were purified on the semipreparative LC using 200 µL injections at a flow rate of 1.5 mL/minute using a C18 semipreparative column at 30 °C (Acclaim C18 5 µm 120A, 4.6x150 mm). Solvents A and B were: Water with 0.1% formic acid, and acetonitrile, respectively. The 24 minute LC gradient was as follows: 5% B at 0 minutes; 25% B at 1.00 minute; 35% B at 20.00 minutes; 100% B at 21.00 minutes; 100% B until 22.00 minutes; 5% B at 22.01 minutes; 5% B until 24.00 minutes. The duration of collection of each fraction was 15 seconds. The 14 minute I1-10 method described in the method section was used to analyze fractions. Fraction 72 was dried down in a speed vac. The residue was resuspended in a 1 mL mixture of water:acetonitrile (6:4) with 0.1% formic acid. The sample was dried in a speed vac and resuspended in deuterated acetonitrile before being transferred to Shigemi tubes for NMR analysis. 1H, 13C, gCOSY, gHSQC, gHMBC and NOESY NMR experiments were performed at the Max T. Rogers NMR Facility at Michigan State University using a Bruker Avance 900 spectrometer equipped with a TCI triple resonance probe. All spectra were referenced to non-deuterated CD_3CN solvent signals ($\delta H = 1.94$ and $\delta C = 1.32$, 118.26 ppm). The ^1H spectra were recorded at 900 MHz, while the ^{13}C spectra were recorded at 225 MHz.

For the NMR sample containing multiple monoacylinositol isomers: the residue was resuspended in a 6 mL mixture of water and acetonitrile (6:4) with 0.1% formic acid by vortexing and the solvent was evaporated under vacuum using a Thermo Savant SPD 131 DDA Speedvac concentrator with BOC Edwards XDS dry pump. The residue was reconstituted in acetonitrile:isopropanol (ACN:IPA) (1:1) with sonication, combined to a single 18x150 mm tube

and concentrated to dryness under vacuum using the Speedvac. Approximately 300 µL of ACN:IPA (1:1) was added to the tube, sonicated, and centrifuged. An additional 150 µL of ACN:IPA (1:1) was added to the tube, sonicated, and centrifuged to collect any remaining material. The supernatant was transferred to an LC autosampler vial and purified via two 200 µL injections. Samples were purified on the semipreparative LC using 200 µL injections at a flow rate of 1.5 mL/minute using a C18 semipreparative column at 50 °C (Acclaim C18 5 µm 120A, 4.6x150 mm). Solvents A, B, and C were: Water with 0.15% formic acid, acetonitrile, and a mixture of dichloromethane:acetone:methanol (1:1:1 v/v), respectively.

The 50 minute LC gradient was using solvents A and B for 0-26 minutes, and B and C for 26-50 minutes. The gradient began with a hold at 5% B from 0-1 minute, ramp from 5-20% B from 1-2 minutes, ramp from 20-40% B from 2-25 minutes, and ramp from 40-100% B from 25-26 minutes. The solvent profile continued with a hold at 100% B from 26-28 minutes, ramp from 100% B to 100% C from 28-29 minutes, hold at 100% C from 29-39 minutes, and ramp from 100% C to 100% B from 39-40 minutes. The final stages of the solvent gradient were a ramp from 100-5% B from 40-41 minutes, and hold at 5% B for 41-50 minutes. 1 minute fractions were collected (18x150 mm tubes) and fractions were tested for purity using an LC-MS. Fractions 16-21 contained the monoacylinositols, were combined, and concentrated to dryness under vacuum with the Speedvac. The residue was dissolved in ~250 µL of D_3-acetonitrile and vortexed. The sample was transferred to a solvent matched Shigemi tube and analyzed by 1H-NMR.

NMR spectra were recorded using a Bruker Avance 900 MHz NMR spectrometer equipped with a TCI triple-resonance inverse detection cryoprobe at the Max T. Rogers NMR facility at Michigan State University. NMR experiments measured ^1H for structural elucidation.

For the LC-MS analysis of monoacylinositol isomers (Figure S4.20), the following LC gradient used solvents A (10 mM ammonium formate, pH 2.8 in water), and B (acetonitrile). Samples were run on a Waters Acquity UPLC coupled to a Waters Xevo G2-XS QToF mass spectrometer with an Acquity UPLC HSS C18 column (2.1 mm x 100 mm x 1.8 µm). The gradient with a flow rate of 0.5 mL/minutes is as follows: Hold at 99% A from 0-1 minute, ramp to 50% A from 1-5 minutes, ramp to 100% A from 10-10.01 minutes, hold at 100% A from 10.01-12.50 minutes, ramp to 99% A from 12.50-12.51 minutes, and hold at 99% A from 12.51-15.00 minutes.

LC-MS analysis

LC-MS samples (both enzyme assays and plant samples) were run on a Waters Acquity UPLC coupled to a Waters Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA). 10 µL of the samples were injected into an Ascentis Express C18 HPLC column (10 cm x 2.1 mm, 2.7 µm) (Sigma-Aldrich, St. Louis, MO), which was maintained at 40 °C. The LC-MS methods used the following solvents: 10 mM ammonium formate, pH 2.8 as solvent A, and 100% acetonitrile as solvent B. A flow rate of 0.3 mL/minute was used unless otherwise specified. Note: Inositol phosphate assays were run using a Shimadzu LC-20AD HPLC system at a flow rate of 0.4 mL/minute instead of a Waters Acquity UPLC.

A 7-minute linear elution gradient consisted of 5% B at 0 minutes, 60% B at 1 minute, 100% B at 5 minutes, held at 100% B until 6 minutes, 5% B at 6.01 minutes and held at 5% until 7 minutes.

A 14 minute I1-10 linear elution gradient consisted of 5% B at 0 minutes; 25% B at 1 minute; 50% B at 10 minutes; 100% B at 12 minutes; 5% B at 12.01 minutes, 5% B at 14.00 minutes. This method is the default 14 minute method unless otherwise stated.

An alternative 14 minute linear elution gradient was used for the comparison of I3:22 metabolites extracted from *S. quitoense* or generated in an enzyme assay containing IAT and TAIAT. The 14 minute linear elution gradient consisted of 5% B at 0 minutes; 60% B at 1 minute; 100% B at 12 minutes; 100% B at 13 minutes; 5% B at 13.01 minutes, 5% B at 14.00 minutes.

A 21 minute linear elution gradient of 5% B at 0 minutes, 60% B at 3 minutes, 100% B at 15 minutes, held at 100% B until 18 minutes, 5% B at 18.01 minutes and held at 5% until 21 minutes.

For ESI- MS settings: capillary voltage, 2.00 kV; source temperature, 100 °C; desolvation temperature, 350 °C; desolvation nitrogen gas flow rate, 600 liters/hour; cone voltage, 40 V; mass range, *m/z*: 50-1000 (with spectra accumulated at 0.1s per function). Three acquisition functions were used to acquire spectra at different collision energies (0, 15, and 35 V). Lock mass correction was performed using leucine enkephalin as the reference for data acquisition.

For ESI+ MS settings: capillary voltage, 3.00 kV; source temperature, 100 °C; desolvation temperature, 350 °C; desolvation nitrogen gas flow rate, 600 liters/hour; cone voltage, 35 V; mass range, *m/z*: 50-1000 (with spectra accumulated at 0.1s per function). Two acquisition functions were used to acquire spectra at different collision energy settings (0, 10-60 V). Lock mass correction was performed using leucine enkephalin as the reference for data acquisition.

For K_m measurements, the same column, 7 minute LC method, and solvents were used for the analysis on the Waters TQD triple quadrupole mass spectrometer coupled to a Waters

Acquity UPLC. The parameters used for the mass spectrometer are as follows: Capillary voltage, 2.5 kV; Cone voltage, 30 v; Source temperature, 130ºC; Desolvation temperature, 350ºC; Cone gas flow, 20 L/hr, Desolvation gas flow, 800 L/hr. The mass pairs used are as follows: Telmisartan, *m/z*: 513 > 287; I1:10, *m/z*: 379 > 171; *m/z*: I1:12, 407 > 199.

Acylsugar purification

Acylsugars were extracted from 20 mature *S. quitoense* leaves into 500 mL of methanol w/ 0.1% formic acid with gentle agitation in a 1L beaker. The methanol was transferred to a boiling flask and dried down using a rotary evaporator with a warm water bath (~40 °C). The dried residue was resuspended in 2 mL of acetonitrile and stored in an LC-MS tube. 500 µL of the acylsugar solution was transferred into a microcentrifuge tube and dried down using the speed vac. The acylsugars were resuspended in 550 µL of 4:1 acetonitrile:water w/ 0.1% formic acid. 5 – 100 µL injections were made onto a C18 semipreparative column (Acclaim C18 5 µm 120A, 4.6x150 mm) using a 63-minute chromatographic method to separate the acylsugars with a flow rate of 1.5 mL/minute and column temperature of 30 °C. Solvent A was water w/ 0.1% formic acid, and Solvent B was acetonitrile.

The chromatographic gradient is as follows: 5% B at 0.00 minutes, 60% B at 1.00 minute, 100% B at 50.00 minutes, hold at 100% B until 60.00 minutes, 5% B at 60.01 minutes, and hold at 5% B until 63 minutes. 1 minute fractions were collected for a total of 63 fractions. Fractions were screened for the presence of different acylsugars, and corresponding fractions pooled and dried down in the speed vac. Each acylsugar was resuspended in 100% acetonitrile, transferred to LC-MS tubes, and stored in the -20 °C freezer. 200 µL aliquots of the acylsugars were transferred and dried down in microcentrifuge tubes with glass inserts before resuspension in 1:1 ethanol:water mixtures for use in enzyme assays.

Gene identification and phylogenetic analysis

All transcripts identified in Moghe et al. (2017) were analyzed using Geneious R8.1.9. Sequences were selected by HXXXD motifs, which were detected using the 'Search for Motif' function (0 mismatches). Those sequences were further parsed to include only those that contain the DFGWG motif (1 mismatch) using the same function. The remaining sequences were further screened to a length of 400-500 amino acids, and by the relative positions of the two motifs (D'Auria, 2006). Sequences were aligned against several other BAHD sequences from D'Auria (2006), and several characterized ASATs (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016; Moghe et al., 2017). Initial candidate BAHD sequences were identified using OrthoMCL and BLASTn analysis (Moghe et al., 2017).

Phylogenetic reconstructions were performed using MEGA X (Kumar et al., 2018). For phylogenetic reconstructions, amino acid sequences were aligned using MUSCLE under default parameters. A maximum likelihood method was used to generate the phylogenetic tree. The model selection feature in MEGA X was used to determine the best evolutionary model for the maximum likelihood method – Jones-Taylor-Thornton (JTT)+G with five rate categories. 1000 bootstrap replicates were performed using partial deletion (30% gaps) for tree reconstruction.

For multiple sequence alignments and amino acid identities present in figures, alignments were performed in Geneious R8.1.9 using the MUSCLE algorithm under default settings.

VIGS analysis

pTRV2-LIC was digested using PstI-HF to generate the linearized vector. The linearized vector was purified using a 1% agarose gel and gel extracted using an Omega EZNA gel extraction kit. Fragments were amplified using PCR with adapters for ligation into pTRV2-LIC.

Both the PCR fragment and the linearized vector were incubated in separate 5 µL reactions using NEB 2.1 as buffer with T4 DNA polymerase and 5 mM dATP or dTTP (PCR insert/Vector). The reactions were incubated at 22 °C for 30 minutes, subsequently incubated at 70 °C for 20 minutes. The reactions were then stored on ice. 1 µL of the pTRV2-LIC reaction and 2 µL of the PCR-LIC reaction were mixed by pipetting. Reactions were incubated at 65 °C for 2 minutes, then 22 °C for 10 minutes. After which the constructs were transformed into chemically competent *E.coli* cells.

Constructs were tested for the presence of the insert using colony PCR and BL-pTRV2-LIC-seq-F/R primers showing a 300 bp insertion. Positive constructs were miniprepped (Qiagen, Hilden, Germany) and sanger sequenced using the same primers. Sequenced constructs and pTRV1 were transformed into agrobacterium strain, GV3101, using the protocol described previously except on LB plates with kanamycin (50 µg/mL), rifampicin (50 µg/mL), and gentamycin (10 µg/mL). Colonies were assayed for the presence of the insert using the colony PCR and the BL-pTRV2-LIC-seq-F/R primers previously described. The presence of the pTRV1 vector in GV3101 was assayed using colony PCR primers, pTRV1-F/R.

Protocol adapted from: (Velásquez et al., 2009). Seeds were germinated using an incubation in 10% bleach for 30 minutes, followed by 5-6 washes with water. Seeds were transferred to a petri dish with Whatman paper and water in the bottom of the dish. Seeds were stored in a lab drawer until hypocotyls emerge, at which point they were moved to a window sill. Once, cotyledons have emerged, seedlings were transferred to peat pots and grown for approximately 1 week under 16/8 day/night cycle at 24 °C. At 2 days pre-inoculation, LB cultures (Kan/Rif/Gent) were inoculated with the cultures used for leaf inoculation. The strains have constructs containing the gene of interest (GOI) in pTRV2-LIC, an empty vector pTRV2-

LIC, and pTRV1. Cultures were grown overnight at 30 °C with shaking at 225 rpm. Larger

cultures composed of induction media (4.88g MES, 2.5g glucose, 0.12g sodium phosphate

monobasic monohydrate in 500 mL, pH 5.6, 200 µM acetosyringone), were inoculated using a

25:1 dilution of the overnight culture (50 mL total). The larger culture was incubated at 30 °C,

225 rpm, overnight. Cells were harvested by centrifugation at 3,200g for 10 minutes. Cell pellets

were resuspended in 1 volume of 10 mM MES, pH 5.6, 10 mM $MgCl_2$. Cells were gently

vortexed to resuspend the pellet. Cell suspensions were centrifuged at 3,200g for 10 minutes.

Cell pellets were resuspended in 10 mL of 10 mM MES, pH 5.6, 10 mM $MgCl_2$. The OD600

values were measured for each of the cultures. Cell suspensions were diluted using the same

buffer to an OD600 of 1. Acetosyringone was added to the pTRV1 cell suspension to a final

concentration of 400 µM. The different pTRV2-LIC constructs were mixed into 50 mL conical

tubes with an equal volume of pTRV1 suspension, resulting in a final acetosyringone

concentration of 200 µM. Individual seedlings were inoculated through the abaxial side of the

cotyledon. Plants were incubated at 22 °C and shaded for 24 hours. After 24 hours, the plants

were returned to 16/8h day-night cycles at the same temperature. Approximately 3 weeks later,

the plants were sampled for acylsugars and RNA using a bisected leaf for each experiment.

Note: Inoculation timing is very important; the cotyledons should be inoculated after they have

expanded, but before the first two true leaves have fully emerged.

qPCR analysis

RNA extraction was performed using Plant RNeasy kits from Qiagen according to the

kits instructions including using the optional DNase digestion kit (Qiagen, Hilden, Germany).

The concentration of RNA was determined using a Nanodrop 2000c. 1 µg of RNA was used to

generate first strand cDNA using the Superscript II reverse transcriptase kit (Thermofisher

Scientific, Waltham, MA) and oligo dT. The cDNA was subsequently diluted 10x into water, which was used as 1x in the qPCR analysis. Gene-specific primers were designed using Primerquest (IDT, Coralville, IA) and blasted against the *S. quitoense* RNAseq assembly to ensure specificity. The efficiency of the primers was determined using cDNA from *S. quitoense*. A 20-fold further dilution of the cDNA was used for qPCR analysis. qRT-PCR reactions were carried out in 10 μL reactions using 200 nM primers. The reactions were carried out with at least 3 technical replicates per sample. qPCR reactions were run on a Quantstudio 7 Flex real-time PCR system. Data were analyzed using Quantstudio real-time PCR software. The delta Ct values for the empty vector control plants were averaged and used for data analysis. The ΔΔCt method was used to determine changes in transcript abundance in the assayed samples. EF1α was used as a control for transcript normalization.

Acylsugar analysis

The interactive protocol for acylsugar extracts is available at Protocols.io at:

https://dx.doi.org/10.17504/protocols.io.xj2fkqe

The acylsugar extraction protocol was described in (Leong et al., 2019). LC-MS conditions used for acylsugar analysis were described the LC-MS analysis section.

APPENDICES

APPENDIX A

Supplemental material

Figure S4.1. Multiple sequence alignment (MSA) of IAT and other members of ASAT1 monophylletic clade from *S. lycopersicum*, and *S. sinuata*.

Sequence alignments using the MUSCLE algorithm. Residues are colored based on similarity. Location of HXXXD catalytic and DFGWG structural motifs are depicted with red arrows.
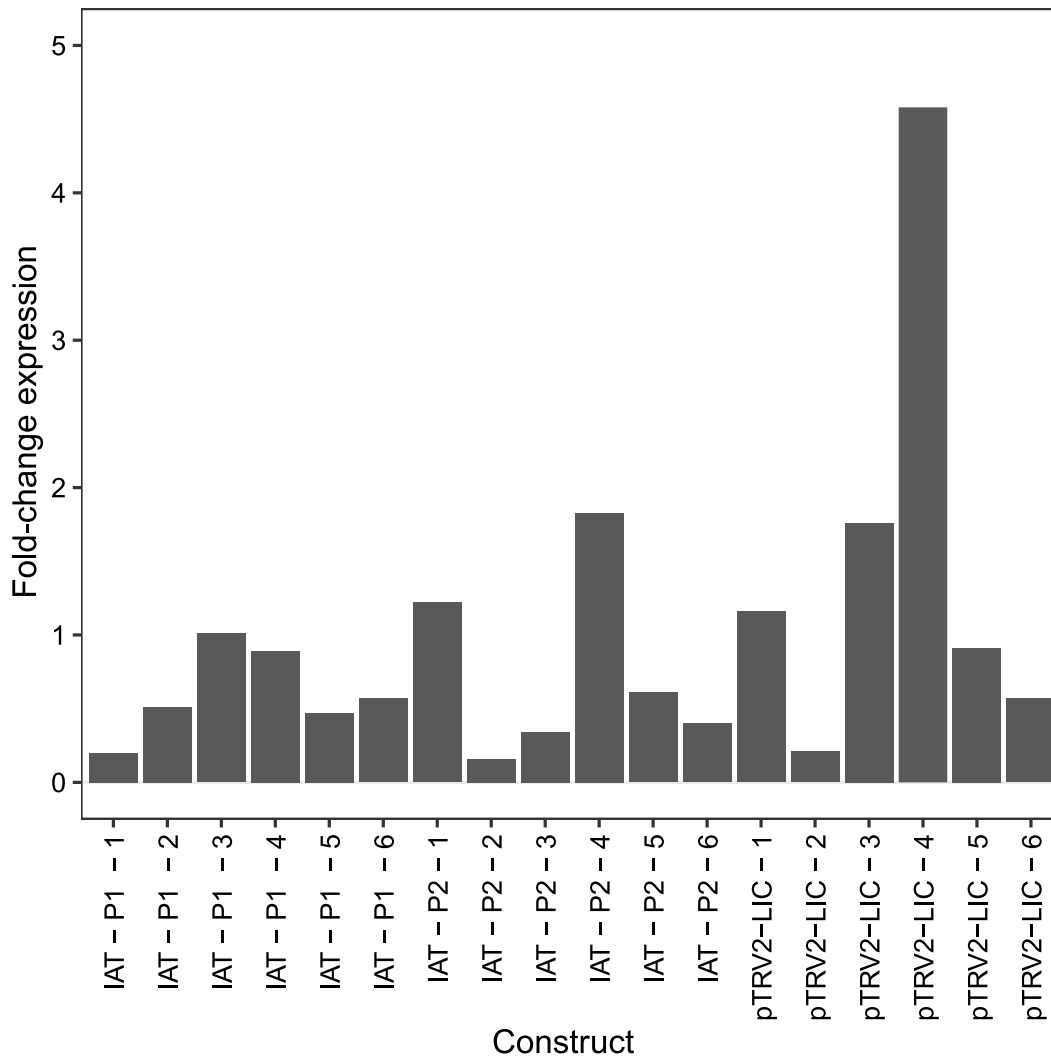
Figure S4.2. Sequence analysis of ASAT4 homologs derived from *S. quitoense* and Sl-ASAT4 and SsASAT5.

Multiple sequence alignment of ASAT4 homologs using MUSCLE. Coloring is based on residue similarity. Location of HXXXD catalytic and DFGWG structural motifs are depicted with red arrows.

Figure S4.3. Expression levels of *IAT* transcripts in VIGS plants.

Relative transcript levels of IAT in VIGS plants targeting IAT compared to mean expression of control plants (ASAT1 homolog) determined by qPCR. RNA for analysis was extracted from whole leaf samples. Samples were normalized to EF1α using the ΔΔCt method. Fold-change expression of the plants is compared to the mean value for all the control plants. The bar graph represents transcript levels of individual VIGS plants.
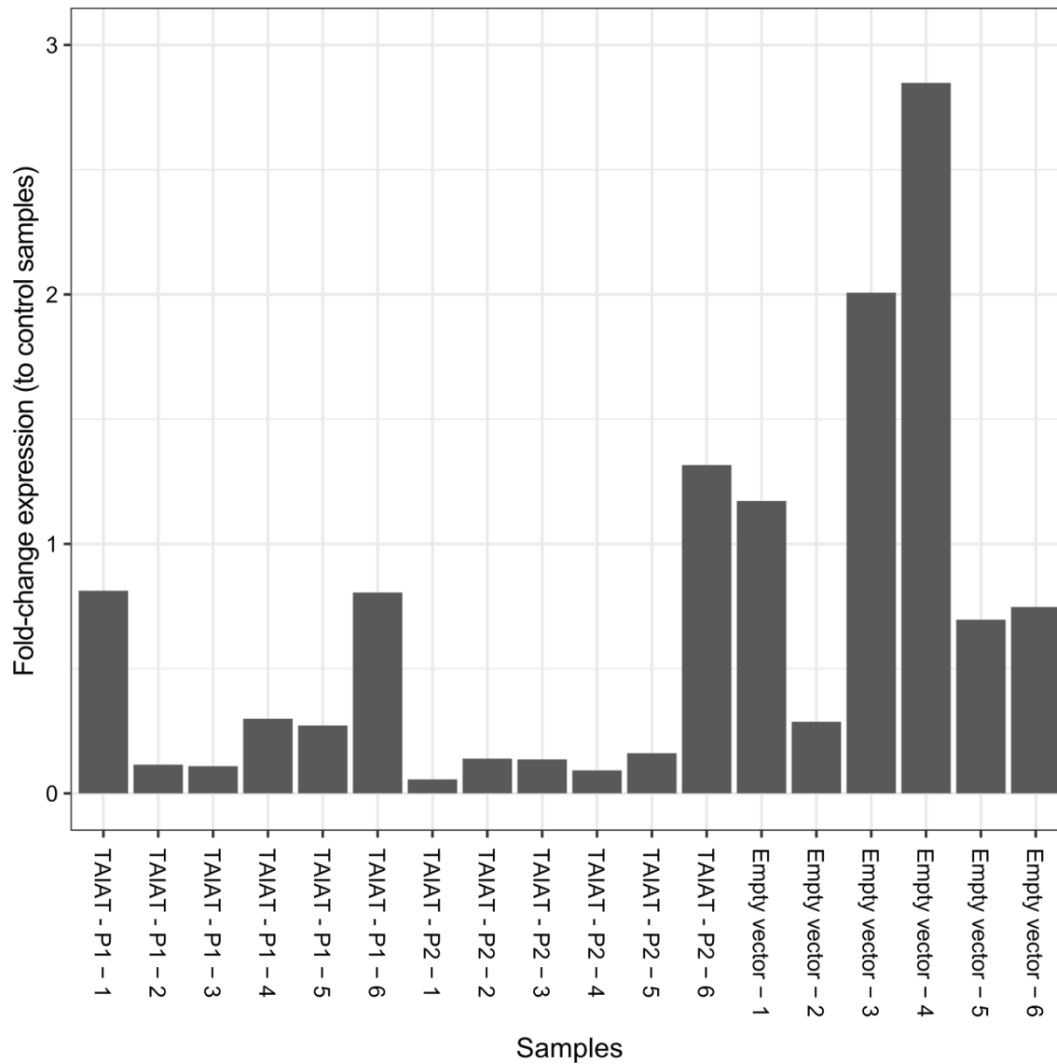
Figure S4.4. Expression levels of *TAIAT* transcripts in VIGS plants.

Relative transcript levels of TAIAT in VIGS plants targeting TAIAT compared to mean expression of control plants (ASAT4 homolog) determined by qPCR. RNA for analysis was extracted from whole leaf samples. Samples were normalized using the $\Delta\Delta Ct$ method to EF1α. Fold-change expression of the plants is compared to the mean value for all the control plants. The bar graph represents transcript levels of individual VIGS plants.
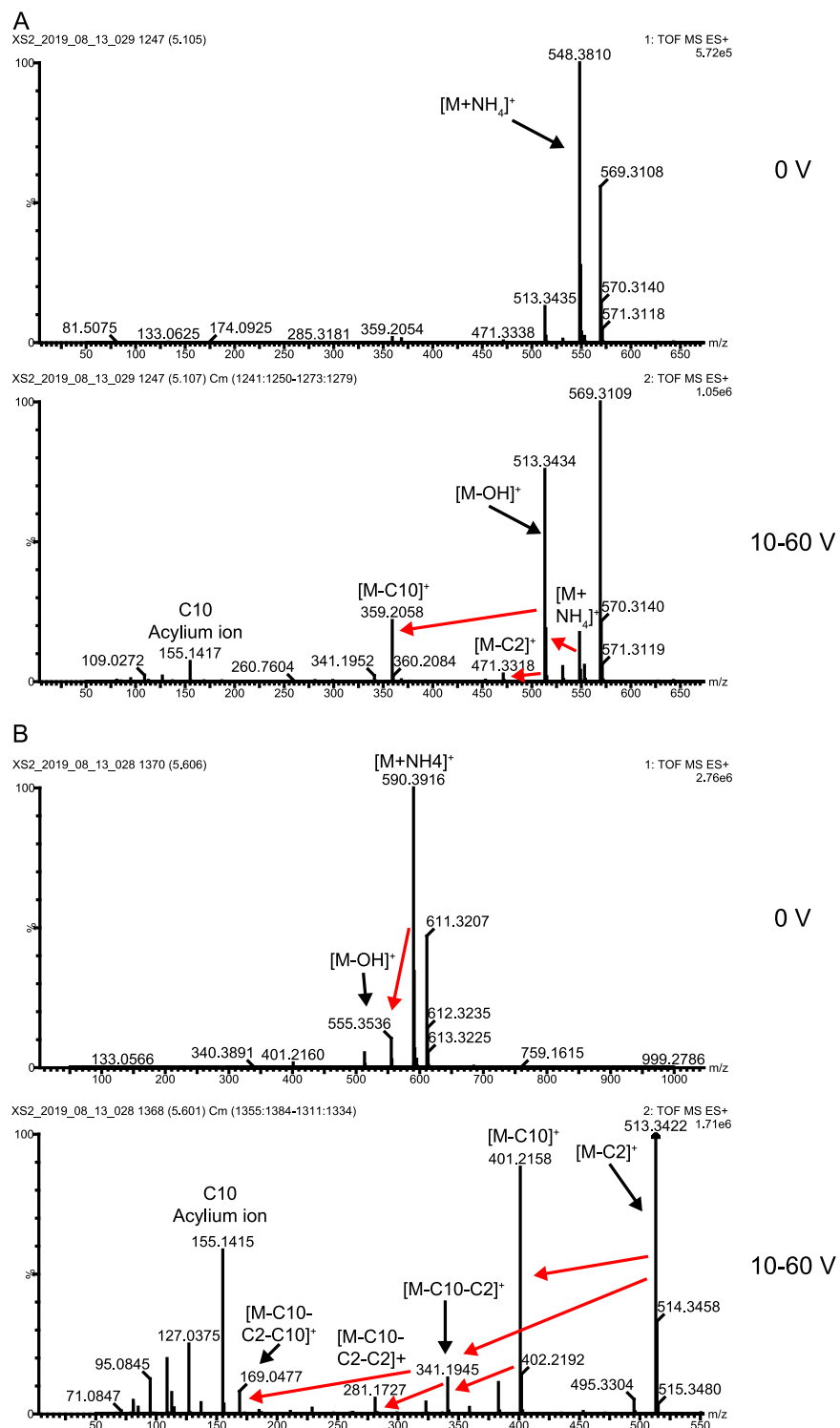
Figure S4.5. Collision-induced dissociation of I3:22 (2,10,10) and I4:24 (2,2,10,10) in ESI+ mode.

Collision-induced dissociation using a ramp of 10-60 volts in ESI+ mode. The fragments represent loss of ketenes from the acylinositols, while the acylium ions are present for C10 (*m/z*: 155.14) acyl chains. Red arrows indicate transitions between ions. (A) Fragmentation of I3:22

Figure S4.5. (cont'd)

product derived from TAIAT reverse reactions. The parent ion (m/z: 548.38) represents the ammonium adduct of I3:22 (2,10,10). (B) Fragmentation of I4:24 derived from TAIAT forward reactions. The parent ion (m/z: 590.39) represents the ammonium adduct of I4:24 (2,2,10,10).
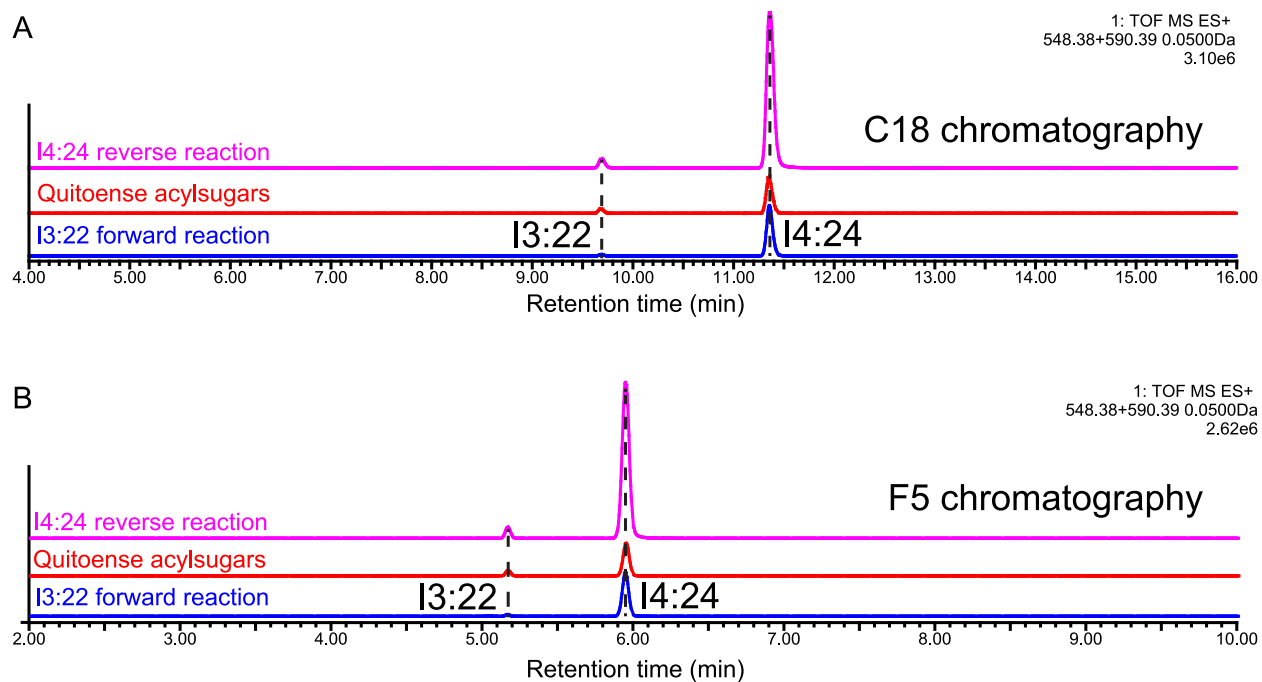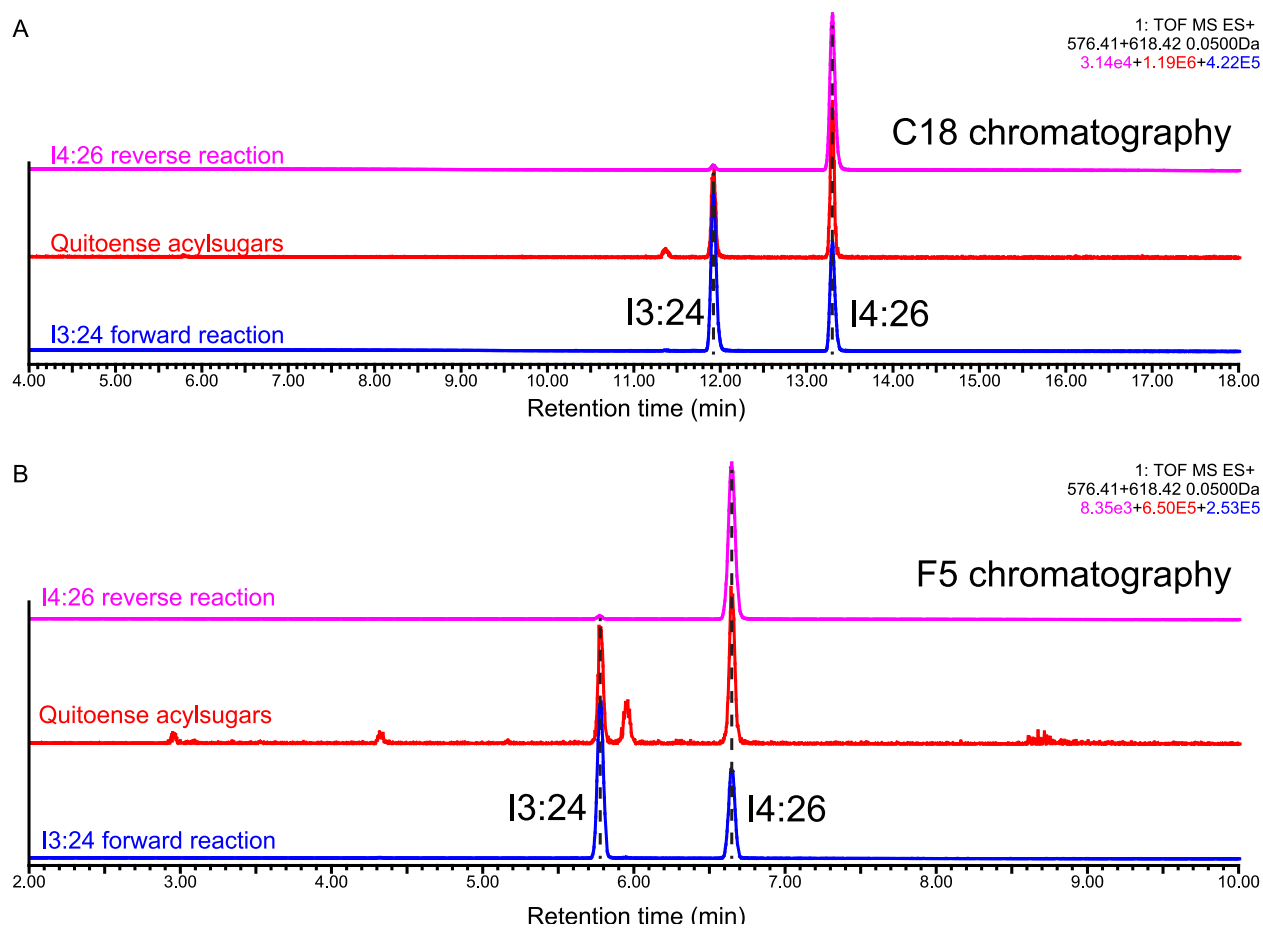
Figure S4.6. I3:22 and I4:24 products co-elute with *S. quitoense* acylsugars.

Comparison of retention time of enzyme assay products with *S. quitoense*-derived products. Products were run on a 21 minute LC-MS method described in the method section. Comparison of plant-derived acylsugars from *S. quitoense* with *in vitro* enzyme assay products utilizing I3:22 (2,10,10) or I4:24 (2,2,10,10) as substrates. The assays were run on a (A) C18 column, (B) F5 column.

Figure S4.7. I3:24 and I4:26 products co-elute with *S. quitoense* acylsugars.

Comparison of retention time of enzyme assay products with *S. quitoense*-derived products. Products were run on a 21 minute LC-MS method described in the method section. Comparison of plant-derived acylsugars from *S. quitoense* with *in vitro* enzyme assay products utilizing I3:24 (2,10,12) or I4:26 (2,2,10,12) as substrates. The assays were run on a (A) C18 column, (B) F5 column.
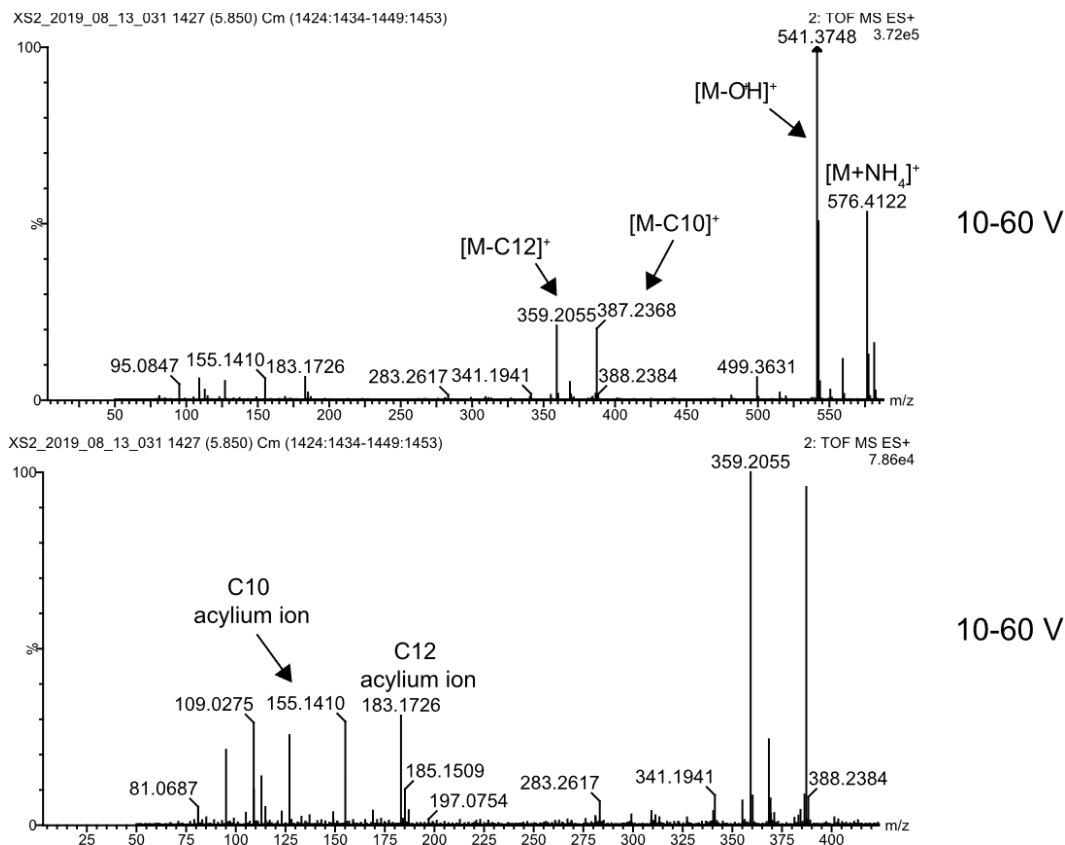
Figure S4.8. Collision-induced dissociation of I3:24 (2,10,12) in ESI+ mode.

Collision-induced dissociation using a ramp of 10-60 volts in ESI+ mode. Fragmentation of I3:24 product derived from TAIAT reverse reactions. The parent ion (*m/z*: 576.41) represents the ammonium adduct of I3:24 (2,10,12). The fragments represent loss of ketenes from the acylinositols, while the acylium ions are present for C10 (*m/z*: 155.14) and C12 (*m/z*: 183.17) acyl chains.
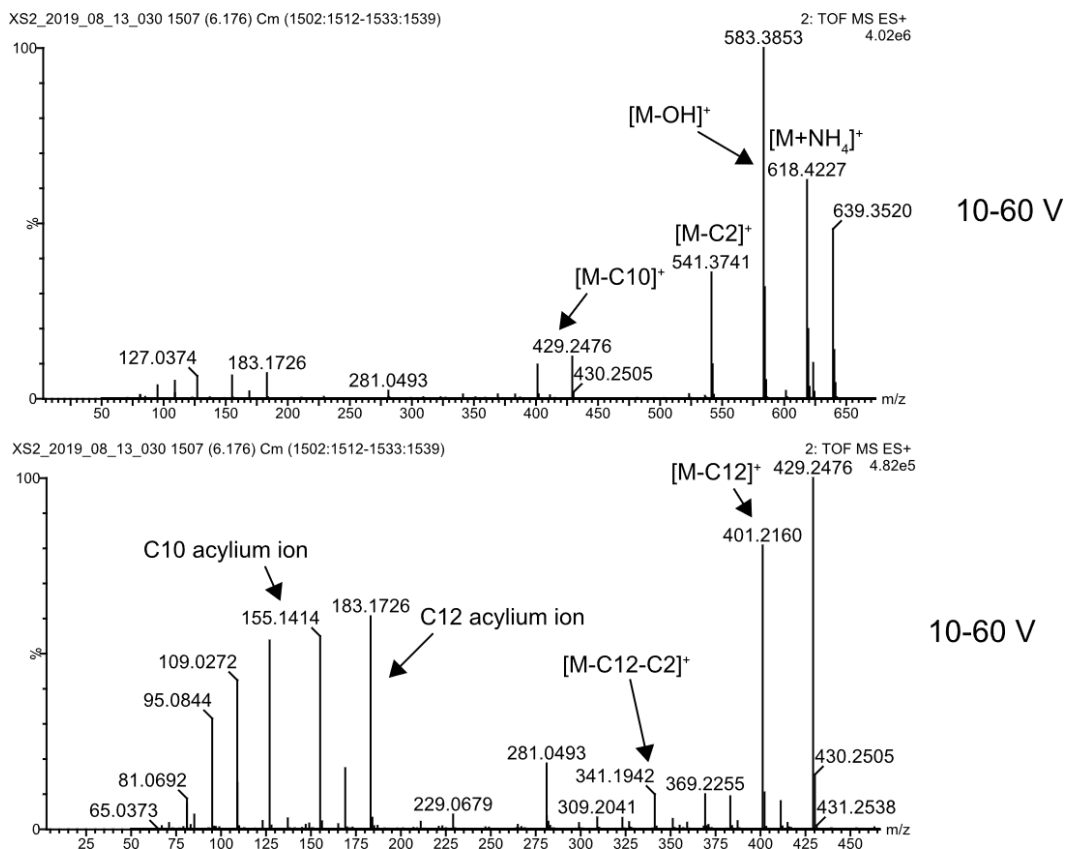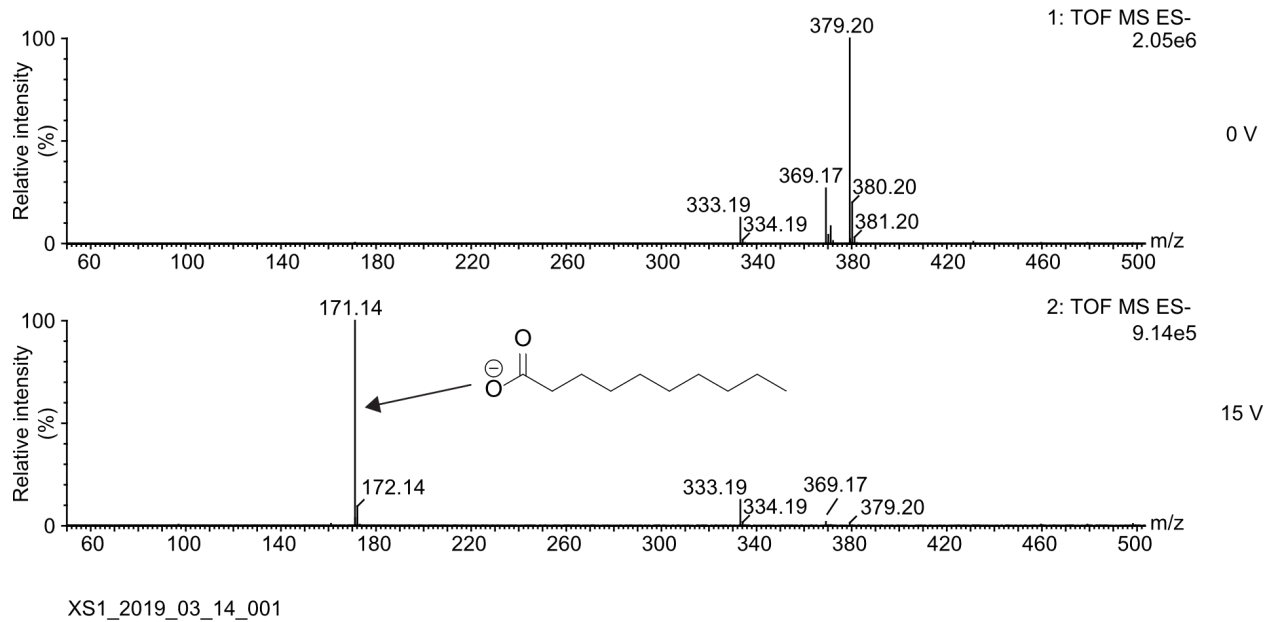
Figure S4.9. Collision-induced dissociation of I4:26 (2,2,10,12) in ESI+ mode.

Collision-induced dissociation using a ramp of 10-60 volts in ESI+ mode. Fragmentation of I4:26 product derived from TAIAT forward reactions. The parent ion (*m/z*: 618.42) represents the ammonium adduct of I4:26 (2,2,10,12). The fragments represent loss of ketenes from the acylinositols, while the acylium ions are present for C10 (*m/z*: 155.14) and C12 (*m/z*: 183.17) acyl chains.

Figure S4.10. Collision-induced dissociation of mono-acylated *myo*-inositol (nC10).

Collision-induced dissociation of I1:10 molecule results in the presence of the nC10 carboxylate ion (*m/z*: 171.14), as shown in the spectra. Collision energy of 15 v was used for this analysis.
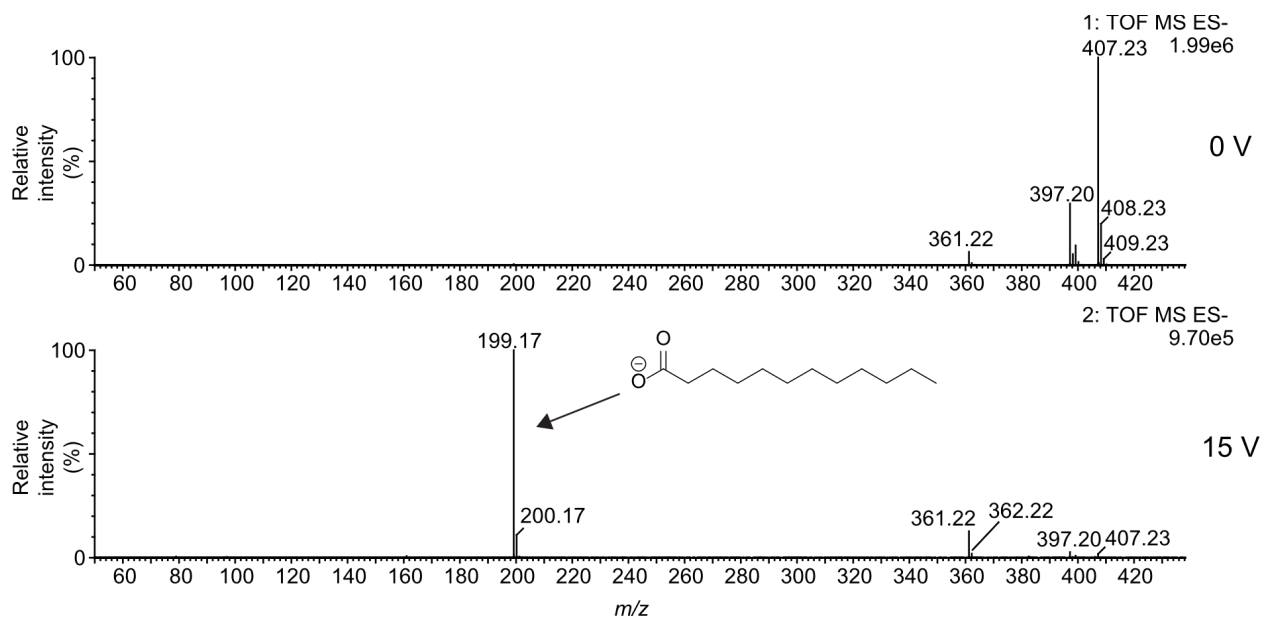
Figure S4.11. Collision-induced dissociation of mono-acylated *myo*-inositol (nC12).

Collision-induced dissociation of I1:12 molecule results in the presence of the nC12 carboxylate ion, as shown in the spectra. Collision energy of 15 volts was used for this analysis.
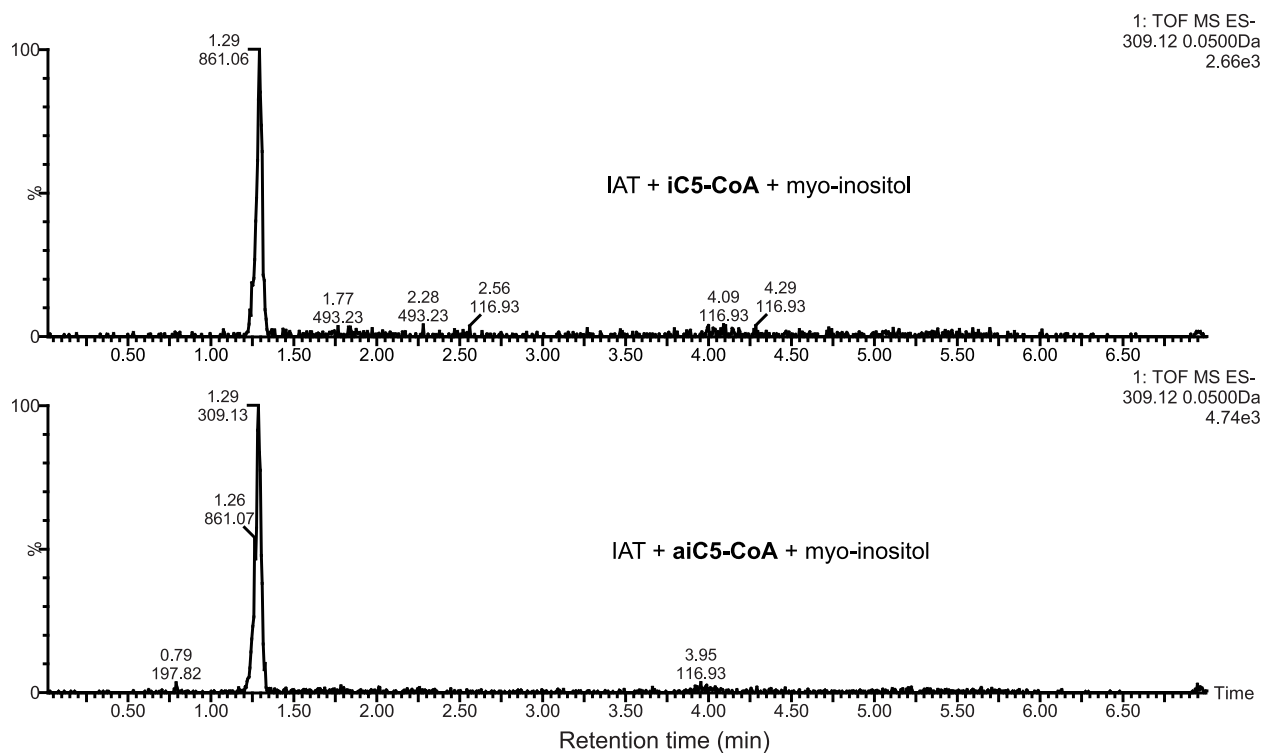
Figure S4.12. IAT catalyzes the formation of I1:5 using *myo*-inositol and iC5-CoA or aiC5-CoA *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions. iC5 (top) or aiC5-CoA (bottom) and *myo*-inositol were incubated with IAT. LC-MS trace shows the extracted ion chromatograms of formate adduct of I1:5 (*m/z*: 309.12) (mass window: 0.05 Da). Sample was run on the 7 minute method described in LC-MS section.

nC10 kinetics for IAT
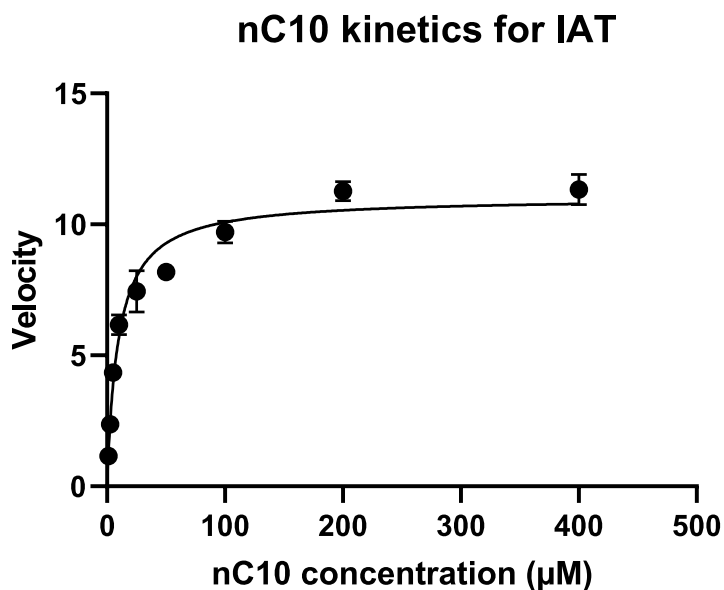
Michaelis-Menten	nC10-CoA
Best-fit values	
Vmax	11.0
Km	9.5 µM
95% CI (profile likelihood)	
Vmax	10.5 to 11.6
Km	7.5 to 11.9 µM
Goodness of Fit	
Degrees of Freedom	25
R squared	0.9635
Sum of Squares	12.05
Sy.x	0.6943
Constraints	
Km	Km > 0
Number of points	
# of X values	27
# Y values analyzed	27

Michaelis-Menten	myo-inositol
Best-fit values	
Vmax	41.6
Km	4.4 mM
95% CI (profile likelihood)	
Vmax	39.0 to 44.6
Km	3.5 to 5.5 mM
Goodness of Fit	
Degrees of Freedom	28
R squared	0.9766
Sum of Squares	100.2
Sy.x	1.891
Constraints	
Km	Km > 0
Number of points	
# of X values	30
# Y values analyzed	30

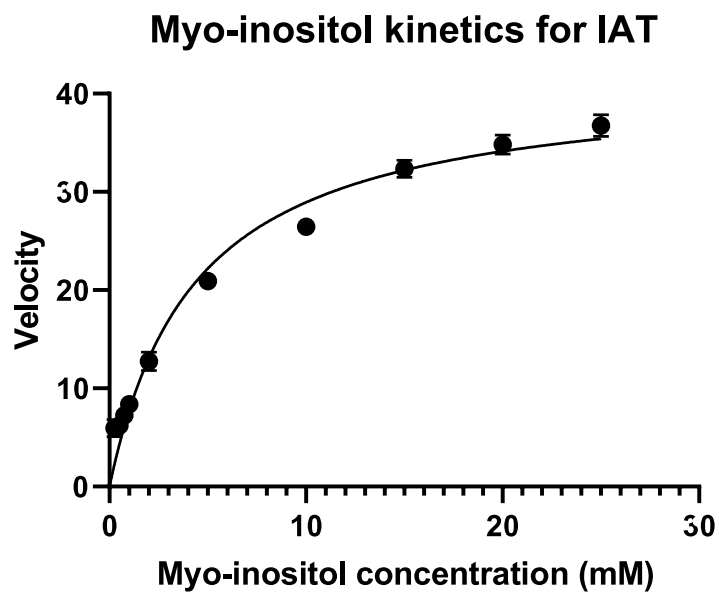# Myo-inositol kinetics for IAT

Figure S4.13. K_m measurements for nC10-CoA and *myo*-inositol with IAT.

Figure S4.13. (cont'd)

Measurements of K_m values of nC10-CoA and *myo*-inositol for IAT. The other substrate was held constant at a saturating concentration. Enzyme velocity was measured as a function of the I1:10 peak area normalized to the internal standard. Samples were analyzed using LC-MS using multiple reaction monitoring on a Waters TQD.

Michaelis-Menten	nC12-CoA
Best-fit values	
Vmax	36.4
Km	16.6 μM
95% CI (profile likelihood)	
Vmax	32.5 to 41.0
Km	11.9 to 23.2
Goodness of Fit	
Degrees of Freedom	19
R squared	0.9492
Sum of Squares	140.4
Sy.x	2.718
Constraints	
Km	Km > 0
Number of points	
# of X values	21
# Y values analyzed	21

nC12-CoA kinetics for IAT
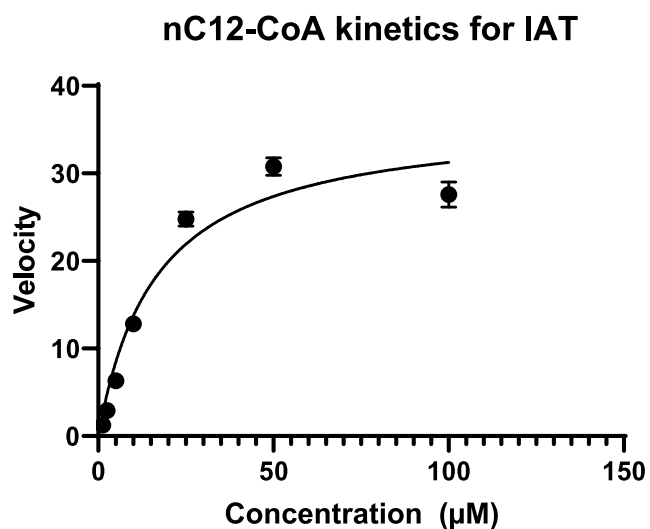
Figure S4.14. K_m measurements for nC12-CoA with IAT.

Measurements of K_m values of nC12-CoA for IAT. The other substrate (*myo*-inositol) was held constant at a saturating concentration. Enzyme velocity was measured as a function of the I1:12 peak area normalized to the internal standard. Samples were analyzed using LC-MS using multiple reaction monitoring on a Waters TQD.
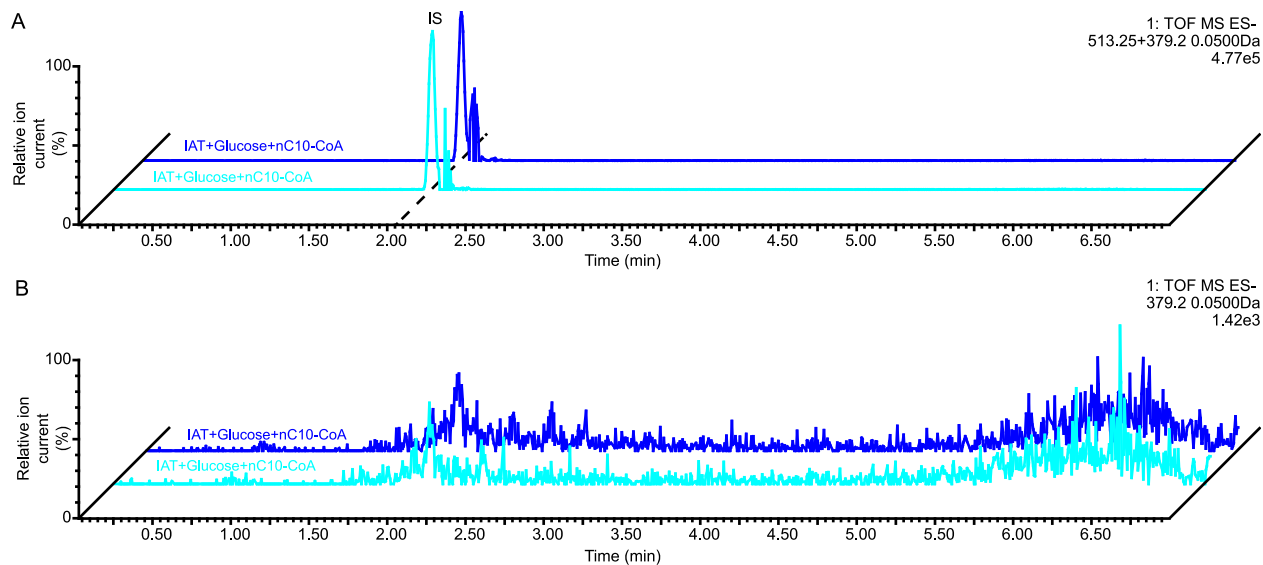
Figure S4.15. IAT does not catalyze the formation of G1:10, utilizing nC10-CoA and glucose *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions. nC10-CoA and glucose were incubated with the enzyme. (A) LC-MS trace shows the extracted ion chromatograms of the formate adduct of G1:10 (*m/z*: 379.20) and telmisartan (*m/z*: 513.25). (B) The extracted ion chromatogram for G1:10 only. The blue traces represent the complete enzyme assays with glucose, nC10-CoA and enzyme. The internal standard was telmisartan. Sample was run on the 7 minute method described in LC-MS section.
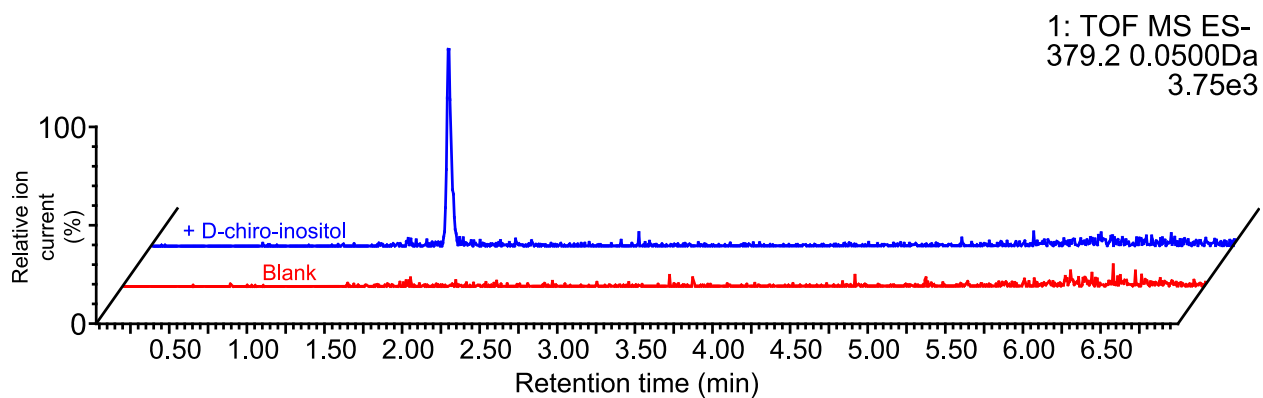
Figure S4.16. IAT can catalyze formation of I1:10 using D-*chiro*-inositol and nC10-CoA.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions. nC10-CoA and D-*chiro*-inositol were incubated with the enzyme. LC-MS trace shows the extracted ion chromatograms of the formate adduct of I1:10 (*m/z*: 379.20). The blue traces represent the complete enzyme assays with D-chiro-inositol, nC10-CoA and enzyme. Sample was run on the 7 minute method described in LC-MS section.
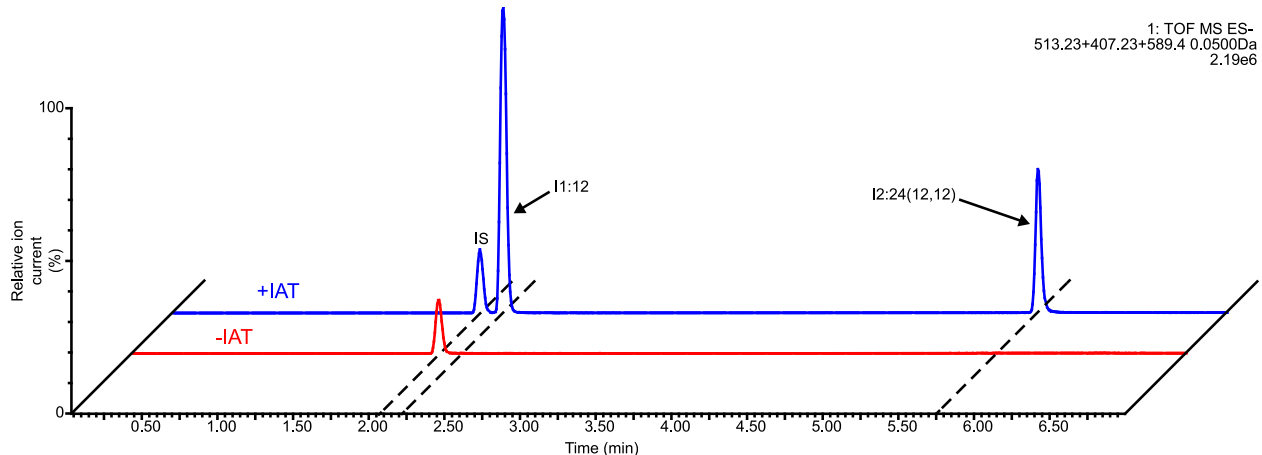
Figure S4.17. IAT catalyzes consecutive acylation of *myo*-inositol using nC12-CoA to generate a di-acylated inositol *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions with *myo*-inositol and nC12-CoA. Extracted ion chromatograms of telmisartan (*m/z*: 513.23), formate adduct of I1:12 (*m/z*: 407.23), and formate adduct of I2:24 (12,12) (*m/z*: 589.40) (mass window: 0.05 Da). The red trace represents the no enzyme control, whereas the blue trace represents the complete assay with enzyme included. The internal standard is telmisartan. Sample was run on the 7 minute method described in LC-MS section.
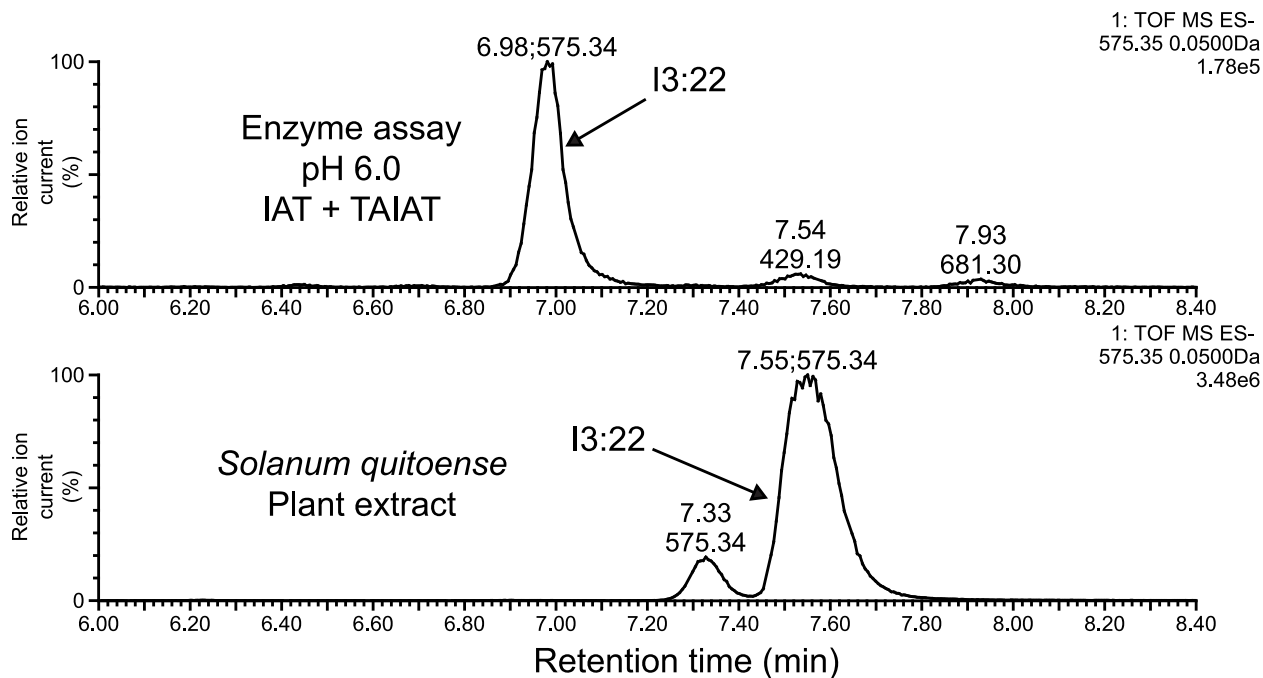
Figure S4.18. IAT and TAIAT catalyze the formation of a chromatographically distinct peak matching the *m/z*: of I3:22, *in vitro*.

LC-MS analysis (ESI-) of enzyme reactions incubating IAT and TAIAT with *myo*-inositol, nC10-CoA, and C2-CoA. Extracted ion chromatogram of I3:22 (*m/z*: 575.35) (mass window: 0.05 Da). The top trace represents the enzymatic assay, whereas the bottom trace represents the plant extract. Sample was run on alternate 14 minute method described in LC-MS section.
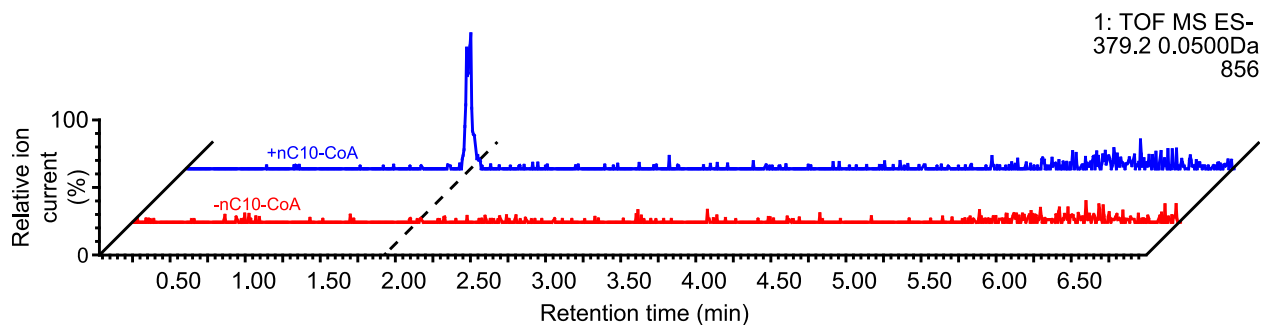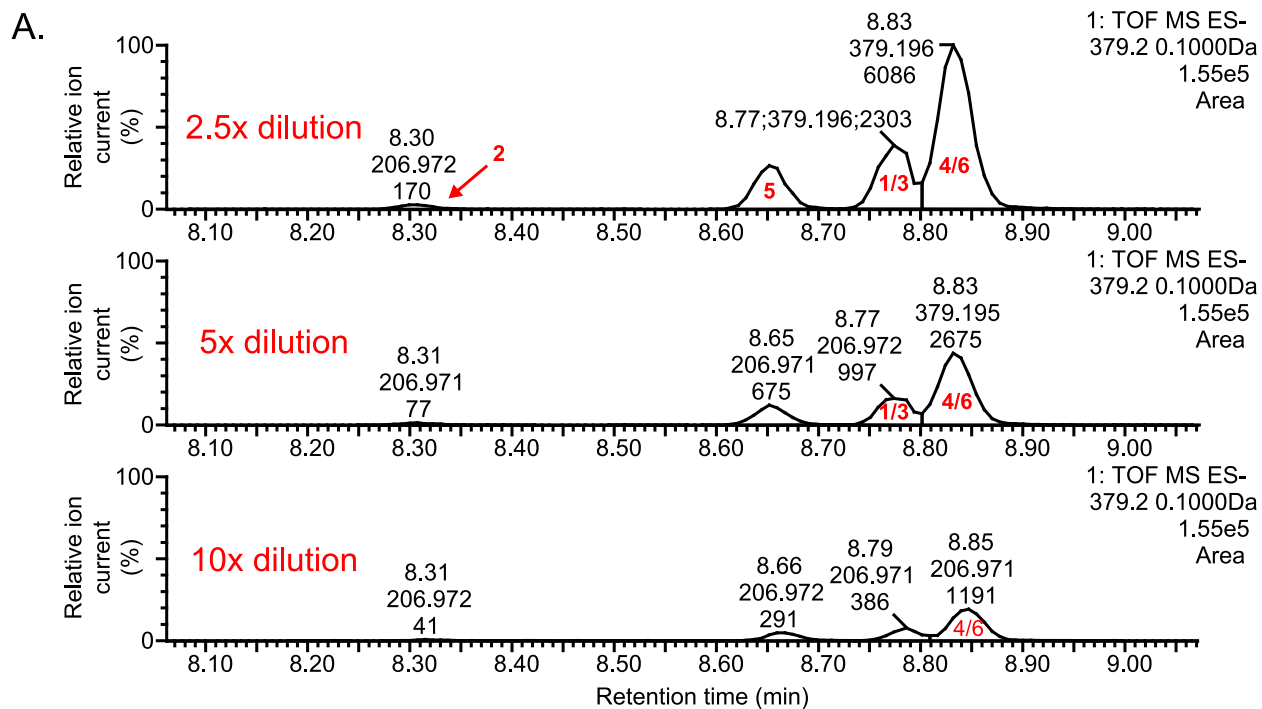
Figure S4.19. Solyc07g043670 can catalyze formation of I1:10 using *myo*-inositol and nC10-CoA *in vitro*.

LC-MS analysis (ESI-) of IAT *in vitro* assays reactions with *myo*-inositol and nC10-CoA. Extracted ion chromatograms of formate adduct of I1:10 (*m/z*: 407.23), (mass window: 0.05 Da). The red trace represents the no CoA control, whereas the blue trace represents the complete assay with CoA included. Sample was run on the 7 minute method described in LC-MS section.

Figure S4.20. LC-MS analysis of monoacylinositol isomers.

(A) LC-MS analysis of I1:10 isomers characterized by NMR using serial dilutions (2.5-, 5-, and 10-fold) of samples. Elution was performed using an Acquity UPLC HSS C18 column (2.1 mm x 100 mm x 1.8 µm). Isomer peaks are labeled in red. (B) Integration of peak area from LC-MS analysis of serially diluted samples compared to blank samples. The percentage of total I1:10 peak area is also calculated.

Table S4.1. Sequence identity of the MSA of the mASAT1 homologs

Percent amino acid sequence identity of the previously identified ASATs and enzymes analyzed in this study.

	IAT	Solyc07g043670	SlASAT1	SsASAT2	SsASAT5	SlASAT4	c38687_g1_i1	TAIAT
IAT		75.7	42.6	43.7	29.5	25.1	26.5	25.2
Solyc07g043670	75.7		41.8	43.3	27.5	25.8	24.7	22.9
SlASAT1	42.6	41.8		65.0	28.6	24.7	26.7	25.9
SsASAT2	43.7	43.3	65.0		30.7	26.6	29.9	28.6
SsASAT5	29.5	27.5	28.6	30.7		28.4	31.9	31.7
SlASAT4	25.1	25.8	24.7	26.6	28.4		55.0	55.2
c38687_g1_i1	26.5	24.7	26.7	29.9	31.9	55.0		85.6
TAIAT	25.2	22.9	25.9	28.6	31.7	55.2	85.6	

Table S4.2. Relative quantification of IAT activities with different substrates.

Enzymatic activity was compared to IAT activity with nC10-CoA and *myo*-inositol. Percent activity represents normalized tested substrate response divided by normalized I1:10 response. Normalized response is the product peak area divided by the internal standard.

Substrate tested	% activity of I1:10
iC5-CoA	0.7%
aiC5-CoA	1.2%
Sucrose	1.9%
D-*chiro*-inositol	1.2%

APPENDIX B

Coauthor summaries

I contributed to the brainstorming, writing, and editing of this manuscript:

Fan, P., Leong, B. J., & Last, R. L. (2019). Tip of the trichome: evolution of acylsugar metabolic diversity in Solanaceae. *Current opinion in plant biology*, *49*, 8-16.

I contributed to manual gene annotation, and writing of this manuscript:

Moore, B. M., Wang, P., Fan, P., Leong, B., Schenck, C. A., Lloyd, J. P., ... & Shiu, S. H. (2019). Robust predictions of specialized metabolism genes through machine learning. *Proceedings of the National Academy of Sciences*, *116*(6), 2344-2353.

I contributed to data collection, analysis, writing, and editing of this manuscript:

Moghe, G. D., Leong, B. J., Hurney, S. M., Jones, A. D., & Last, R. L. (2017). Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife*, *6*, e28468.

I contributed to data collection, analysis, and editing of this manuscript:

Ning, J., Moghe, G. D., Leong, B., Kim, J., Ofner, I., Wang, Z., ... & Last, R. L. (2015). A feedback-insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. *Plant physiology*, *169*(3), 1821-1835.

REFERENCES

REFERENCES

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol **9**: 331–340.

Eudes A, Mouille M, Robinson DS, Benites VT, Wang G, Roux L, Tsai Y-L, Baidoo EEK, Chiu T-Y, Heazlewood JL, et al (2016) Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. Microb Cell Fact **15**: 198.

Fan P, Miller AM, Liu X, Jones AD, Last RL (2017) Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. Nat Commun **8**: 1–13.

Fan P, Miller AM, Schilmiller AL, Liu X, Ofner I, Jones AD, Zamir D, Last RL (2016) In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. Proc Natl Acad Sci U S A **113**: E239-48.

Hurney SM (2018) Strategies for profiling and discovery of acylsugars. Michigan State University.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol **35**: 1547–1549.

Landmann C, Hücherig S, Fink B, Hoffmann T, Dittlein D, Coiner HA, Schwab W (2011) Substrate promiscuity of a rosmarinic acid synthase from lavender (*Lavandula angustifolia* L.). Planta **234**: 305–20.

Leckie BM, D'Ambrosio DA, Chappell TM, Halitschke R, De Jong DM, Kessler A, Kennedy GG, Mutschler MA (2016) Differential and synergistic functionality of acylsugars in suppressing oviposition by insect herbivores. PLoS One **11**: e0153345.

Leong BJ, Lybrand DB, Lou Y-R, Fan P, Schilmiller AL, Last RL (2019) Evolution of metabolic novelty: A trichome-expressed invertase creates specialized metabolic diversity in wild tomato. Sci Adv **5**: eaaw3754.

Moghe GD, Leong BJ, Hurney SM, Jones AD, Last RL (2017) Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. Elife **6**: 1–33.

Nadakuduti SS, Uebler JB, Liu X, Jones AD, Barry CS (2017) Characterization of trichome-expressed BAHD acyltransferases in *Petunia axillaris* reveals distinct acylsugar assembly mechanisms within the Solanaceae. Plant Physiol **175**: 36–50.

Ning J, Moghe GD, Leong B, Kim J, Ofner I, Wang Z, Adams C, Jones AD, Zamir D, Last RL (2015) A feedback-insensitive isopropylmalate synthase affects acylsugar composition

in cultivated and wild tomato. Plant Physiol **169**: 1821–35.

Puterka GJ, Farone W, Palmer T, Barrington A (2003) Structure-function relationships affecting the insecticidal and miticidal activity of sugar esters. J Econ Entomol **96**: 636–644.

Roberts JKM, Ray PM, Wade-Jardetzky N, Jardetzky O (1980) Estimation of cytoplasmic and vacuolar pH in higher plant cells by 31P NMR. Nature **283**: 870–872.

Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol **13**: 1–15.

Schilmiller AL, Charbonneau AL, Last RL (2012) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. Proc Natl Acad Sci **109**: 16377–16382.

Schilmiller AL, Gilgallon K, Ghosh B, Jones AD, Last RL (2016) Acylsugar acylhydrolases: carboxylesterase catalyzed hydrolysis of acylsugars in tomato trichomes. Plant Physiol **170**: 1331–1344.

Schilmiller AL, Moghe GD, Fan P, Ghosh B, Ning J, Jones AD, Last RL (2015) Functionally divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. Plant Cell **27**: 1002–1017.

Smith FA, Raven JA (1979) Intracellular pH and its regulation. Annu Rev Plant Physiol **30**: 289–311.

Velásquez AC, Chakravarthy S, Martin GB (2009) Virus-induced gene silencing (VIGS) in *Nicotiana benthamiana* and tomato. J Vis Exp e1292.

Zhao L, Zhang H, Hao T, Li S (2015) In vitro antibacterial activities and mechanism of sugar fatty acid esters against five food-related bacteria. Food Chem **187**: 370–7.