

SECRETORY GLANDULAR TRICHOMES:
ANALYSIS AND ENGINEERING OF PLANT SPECIALIZED METABOLISM

By

Daniel Benjamin Lybrand

A DISSERTATION

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

Biochemistry and Molecular Biology—Doctor of Philosophy

2020

ABSTRACT

SECRETORY GLANDULAR TRICHOMES: ANALYSIS AND ENGINEERING OF PLANT SPECIALIZED METABOLISM

By

Daniel Benjamin Lybrand

Plants produce a wide variety of specialized metabolites that mediate interactions between plants and other organisms or the environment. Specialized metabolic pathways are often unique to a subset of plant species and augment the core metabolic pathways that are nearly universal to all plants. In addition to phylogenetic restriction, specialized metabolic pathways also exhibit compartmentalization within a plant, either at the tissue or cell type level. Many plants possess secretory glandular trichomes, structures consisting of one or a few cells that protrude from the surfaces of aboveground plant tissues like leaves, stems, and flowers. These trichomes accumulate or secrete specialized metabolites. This work describes the biochemical pathways behind two classes of trichome-localized specialized metabolites: pyrethrins and acylsugars. Metabolic engineering of the pyrethrin pathway in plant hosts and geographic variation of acylsugar phenotypes are also discussed.

ACKNOWLEDGEMENTS

The author thanks Drs. Robert L. Last, Gregg A. Howe, A. Daniel Jones III, Eran Pichersky, and Kevin D. Walker for mentorship and advice over the last six years. The author also thanks Dr. Timothy A. Whitehead for previously serving on the author's advisory committee. Thanks go also to Drs. Thilani M. Anthony and Craig A. Schenck for frequent feedback on abstracts, manuscripts, and this dissertation.

Special thanks go to the author's wife, Van Quynh Duong, for putting up with abundant profanity during the formatting of this document, and to the author's dog, Meiji, for providing the author with an excuse to leave the house while compiling this document.

This research was made possible by the Michigan State University College of Natural Science Recruiting Fellowship, the Jack Throck Watson Graduate Fellowship, the National Institutes of Health predoctoral training grant T32-GM110523, the National Science Foundation (NSF) collaborative research grant CBET-1565355 (to Dr. Eran Pichersky), the NSF collaborative research grant CBET-1565232 (to Dr. Robert L. Last), and the NSF research PGR grant IOS-1546617 (to Drs. Robert L. Last, A. Daniel Jones III, Eran Pichersky, Shin-Han Shiu, and Cornelius S. Barry).

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	x
KEY TO ABBREVIATIONS	xv
CHAPTER 1 – BIOSYNTHESIS AND REGULATION OF PYRETHRIN INSECTICIDES	1
TRICHOME-LOCALIZED PLANT SPECIALIZED METABOLISM.....	2
EARLY USE AND DEVELOPMENT OF PYRETHRUM-BASED INSECTICIDES.....	4
INDUSTRIAL PYRETHRINS PRODUCTION	7
PYRETHRIN BIOSYNTHESIS	9
<i>Monoterpenoid biosynthesis</i>	9
<i>Rethrolone biosynthesis</i>	11
<i>Condensation of monoterpenoids and rethrolones</i>	12
REGULATION OF PYRETHRIN BIOSYNTHESIS	13
<i>Spatiotemporal specificity of pyrethrin biosynthesis</i>	13
<i>Pyrethrin production in response to wounding and volatiles</i>	16
CONCLUDING REMARKS AND FUTURE PERSPECTIVES	17
REFERENCES	19
CHAPTER 2 – METABOLIC ENGINEERING OF THE PYRETHRIN PATHWAY IN PLANTS	29
METABOLIC ENGINEERING IN PLANTS.....	30
MATERIALS AND METHODS.....	37
<i>Plant growth conditions</i>	37
<i>Co-expression analysis</i>	37
<i>Vector construction and generation of tomato transgenic plants</i>	37

<i>RNA isolation and qRT-PCR analysis</i>	39
<i>Subcellular localization</i>	39
<i>Enzymatic assays of recombinant TcCDS, ShADH, ShALDH, and TcNudix1</i>	40
<i>Extraction and analysis of trans-chrysanthemic acid and related compounds from tomato fruits</i>	42
<i>GC-MS analysis and LC-MS analysis</i>	43
<i>Lycopene measurements</i>	44
RESULTS.....	45
<i>Identification of TcCDS transgenic tomato lines producing trans-chrysanthemic acid</i>	45
<i>Reconstruction of the complete pathway to trans-chrysanthemic acid in tomato fruit</i>	45
<i>Production of trans-chrysanthemic acid and related compounds in transgenic tomato fruits expressing CDS, ADH and ALDH</i>	49
<i>Identification of TcNudix1 and subcellular localization of the protein</i>	55
<i>Tissue-specific expression of TcNudix1</i>	59
<i>Characterization of the hydrolysis activity of TcNudix1</i>	59
<i>Heterologous co-expression of TcNudix1 and TcCDS in tomato</i>	61
<i>Expression of TcCDS in type VI trichomes of S. lycopersicum</i>	61
DISCUSSION	65
APPENDIX	73
REFERENCES	83
CHAPTER 3 – ACYLGLUCOSE BIOSYNTHESIS IN <i>SOLANUM PENNELLII</i>	88
ACYLSUGAR BIOSYNTHESIS IN SOLANACEAE.....	89
MATERIALS AND METHODS.....	92
<i>Plant material</i>	92
<i>Acylsugar analysis</i>	93

<i>Acylsugar quantification</i>	94
<i>Acylsucrose purification</i>	96
<i>qPCR analysis</i>	97
<i>Genotyping of progeny of BIL6521 x BIL6180</i>	98
<i>DNA construct assembly</i>	98
<i>Competent cell preparation and transformation of constructs into Agrobacterium</i>	101
<i>Plant transformation</i>	102
<i>Transient expression and purification of SpASFF1 protein</i>	104
<i>Enzyme assays</i>	106
<i>Statistical Analysis</i>	106
RESULTS.....	107
<i>An S. pennellii chromosome 3 locus is necessary for acylglucose production from P-type acylsucroses</i>	107
<i>The chromosome 3 locus encodes a glandular trichome-expressed β-fructofuranosidase</i>	111
<i>Gene editing reveals that SpASFF1 is necessary for S. pennellii LA0716 acylglucose accumulation</i>	113
<i>SpASFF1 converts P-type acylsucroses to acylglucoses both in vivo and in vitro</i>	114
DISCUSSION	119
APPENDIX	126
REFERENCES	148
CHAPTER 4 – INTRASPECIFIC VARIATION IN <i>SOLANUM PENNELLII</i> TRICHOME METABOLISM..	153
CHEMICAL DIVERSITY IN <i>SOLANUM PENNELLII</i> TRICHOMES	154
MATERIALS AND METHODS.....	155
<i>Plant material</i>	155
<i>Acylsugar extraction</i>	156

<i>Metabolomic analysis by LC-MS</i>	157
<i>Untargeted metabolomics data processing</i>	158
<i>Acylsugar quantification</i>	159
<i>RNA extraction, cDNA synthesis, and qPCR</i>	161
<i>Acylsugar purification</i>	162
<i>NMR spectroscopy</i>	164
RESULTS.....	164
<i>Untargeted metabolomics reveals acylsugars and flavonoids in trichomes</i>	164
<i>Acylsugar core composition varies across the range of S. pennellii</i>	172
<i>Variable acyl chain and sugar composition yield acylsugar diversity</i>	174
<i>NMR spectroscopy shows structural relationships among compounds with similar mass spectra</i>	176
<i>Flavonoids vary by core and degree of methylation</i>	178
<i>Multivariate analysis implicates short branched acyl chains in North-South acylsugar variation</i>	179
<i>Variation in medium-length acyl chains drives variation within the South range</i>	182
<i>LA2963 segregates from other Atico region accessions due to high acylsucrose content</i> ..	186
DISCUSSION	190
APPENDIX	197
REFERENCES	279

LIST OF TABLES

Table S2.1.....	74
Table S2.2.....	81
Table S2.3.....	82
Table 3.1.....	115
Table S3.1.....	144
Table S3.2.....	146
Table S3.3.....	147
Table 4.1.....	168
Table 4.2.....	171
Table 4.3.....	173
Table 4.4.....	183
Table S4.1.....	198
Table S4.2.....	203
Table S4.3.....	210
Table S4.4.....	217
Table S4.5.....	224
Table S4.6.....	231
Table S4.7.....	238
Table S4.8.....	245
Table S4.9.....	252
Table S4.10.....	259

Table S4.11.....	266
Table S4.12.....	273
Table S4.13.....	275
Table S4.14.....	277

LIST OF FIGURES

Figure 1.1	5
Figure 1.2	14
Figure 2.1	36
Figure 2.2	46
Figure 2.3	48
Figure 2.4	50
Figure 2.5	52
Figure 2.6	54
Figure 2.7	56
Figure 2.8	58
Figure 2.9	60
Figure 2.10	62
Figure 2.11	64
Figure 2.12	66
Figure 2.13	67
Figure S2.1.....	76
Figure S2.2.....	77
Figure S2.3.....	78
Figure S2.4.....	79
Figure S2.5.....	80
Figure 3.1	109

Figure 3.2	112
Figure 3.3	116
Figure 3.4	118
Figure 3.5	120
Figure S3.1.....	127
Figure S3.2.....	129
Figure S3.3.....	130
Figure S3.4.....	131
Figure S3.5.....	133
Figure S3.6.....	135
Figure S3.7.....	137
Figure S3.8.....	139
Figure S3.9.....	141
Figure S3.10.....	142
Figure 4.1	166
Figure 4.2	177
Figure 4.3	181
Figure 4.4	185
Figure 4.5	187
Figure 4.6	189
Figure S4.1.....	201
Figure S4.2.....	202

Figure S4.3.....	204
Figure S4.4.....	205
Figure S4.5.....	206
Figure S4.6.....	207
Figure S4.7.....	208
Figure S4.8.....	209
Figure S4.9.....	211
Figure S4.10.....	212
Figure S4.11.....	213
Figure S3.12.....	214
Figure S4.13.....	215
Figure S4.14.....	216
Figure S4.15.....	218
Figure S4.16.....	219
Figure S4.17.....	220
Figure S4.18.....	221
Figure S4.19.....	222
Figure S4.20.....	223
Figure S4.21.....	225
Figure S4.22.....	226
Figure S4.23.....	227
Figure S4.24.....	228

Figure S4.25.....	229
Figure S4.26.....	230
Figure S4.27.....	232
Figure S4.28.....	233
Figure S4.29.....	234
Figure S4.30.....	235
Figure S4.31.....	236
Figure S4.32.....	237
Figure S4.33.....	239
Figure S4.34.....	240
Figure S4.35.....	241
Figure S4.36.....	242
Figure S4.37.....	243
Figure S4.38.....	244
Figure S4.39.....	246
Figure S4.40.....	247
Figure S4.41.....	248
Figure S4.42.....	249
Figure S4.43.....	250
Figure S4.44.....	251
Figure S4.45.....	253
Figure S4.46.....	254

Figure S4.47.....	255
Figure S4.48.....	256
Figure S4.49.....	257
Figure S4.50.....	258
Figure S4.51.....	260
Figure S4.52.....	261
Figure S4.53.....	262
Figure S4.54.....	263
Figure S4.55.....	264
Figure S4.56.....	265
Figure S4.57.....	267
Figure S4.58.....	268
Figure S4.59.....	269
Figure S4.60.....	270
Figure S4.61.....	271
Figure S4.62.....	272

KEY TO ABBREVIATIONS

AAE	acyl activating enzyme
ACS	acyl CoA synthetase
ADH	alcohol dehydrogenase
ALDH	aldehyde dehydrogenase
ANOVA	analysis of variance
AOC	allene oxide cyclase
AOS	allene oxide synthase
ASAT	acylsugar acyltransferase
ASFF	acylsucrose fructofuranosidase
ASH	acylsugar hydrolase
BIL	backcrossed introgression line
BT	<i>Bacillus thuringiensis</i>
CA	chrysanthemic acid
CCMT	10-carboxychrysanthemic acid methyltransferase
CDP	chrysanthemyl diphosphate
CDSase	chrysanthemyl diphosphate synthase
CHH	chrysanthemol 10-hydroxylase
CMP	chrysanthemyl monophosphate
CoA	coenzyme A
DMADP	dimethylallyl diphosphate
DXP	1-deoxy-D-xylulose-5-phosphate

ECH	enoyl CoA hydratase
ER	endoplasmic reticulum
FAO	Food and Agriculture Organization of the United Nations
GC-MS	gas chromatography-mass spectrometry
GGPPS	geranyl-geranyl diphosphate synthase
GH	glycoside hydrolase
GLIP	GDSL lipase/acyltransferase
HPLC	high performance liquid chromatography
IDP	isopentenyl diphosphate
IL	introgression line
IPMS	isopropylmalate synthase
JMH	jasmone hydroxylase
LC-MS	liquid chromatography-mass spectrometry
LOX	lipoxygenase
MCPI	metallocarboxypeptidase inhibitor
OPDA	12-oxophytodienoic acid
OPLS-DA	orthogonal partial least squares discriminant analysis
ORF	open reading frame
PA	pyrethric acid
PCA	principal component analysis
PG	polygalacturonase
PPFD	photosynthetic photon flux density

PYS	pyrethrolone synthase
QC	quality control
QTL	quantitative trait locus
SCPL	serine carboxypeptidase-like acyltransferase
SGT	secretory glandular trichome
TGRC	Tomato Genetics Resource Center
ToF	time-of-flight
UDP	uridine diphosphate
UPLC-HR-MS	ultraperformance liquid chromatography high resolution-mass spectrometry
VOC	volatile organic compound

CHAPTER 1 – BIOSYNTHESIS AND REGULATION OF PYRETHRIN INSECTICIDES

Portions of this chapter are adapted from a review article submitted to *Trends in Plant Science*:

Lybrand, D.B, Xu, H., Last, R.L., and Pichersky, E. How plants synthesize pyrethrins – safe and biodegradable insecticides.

TRICHOME-LOCALIZED PLANT SPECIALIZED METABOLISM

Small molecules mediate many interactions between plants and other organisms. The metabolites that a plant produces can alter growth, behavior, or metabolism in microbes, animals, or other plants. Many plant metabolites increase the fitness of the producing plant by eliciting positive interactions with other organisms or mitigating negative ones. *Medicago truncatula* produces flavonoids that elicit symbiosis with nitrogen-fixing *Sinorhizobium meliloti* bacteria by initiating nodulation while *Nicotiana* and *Capsicum* species produce the phytoalexin capsidiol to inhibit growth of fungal pathogens such as *Alternaria alternata* [1,2]. *Nicotiana attenuata* flowers emit volatile benzyl acetone, simultaneously attracting the moth pollinator *Manduca sexta* while deterring herbivorous *Diabrotica undecimpunctata* [3,4]. *Ailanthus altissima* secretes ailanthone and other allelopathic chemicals into the soil thereby preventing germination of competing plants while leaves of the tea plant (*Camellia sinensis*) emit nerolidol triggering a cold stress response in neighboring conspecifics [5,6]. Many of the molecules that effect these interactions come from lineage-specific biosynthetic pathways and are termed specialized metabolites [7]. Unlike the highly conserved core pathways common to nearly all plants, specialized metabolic pathways evolve rapidly. In some cases, conserved specialized metabolic pathways in closely related species may differ in their complement of enzymatic steps or contain orthologous enzymes with divergent substrate specificity or product profiles [8,9]. This leads to chemical diversity among related plant species even when similar specialized metabolic pathways are present in all species considered. These unique pathways and their myriad products provide a chemical palette that allows plants to contend with different stressors from diverse environments.

To mediate interactions with other organisms, plants may secrete specialized metabolites into the soil, emit them into the air, or accumulate them within plant tissues. While some specialized metabolites accumulate in or are emitted from basic structures of plant anatomy (e.g., formylated phloroglucinol compounds in leaves of various *Eucalyptus* species, taxoids in stems and bark of many *Taxus* species, benzyl acetone in *Nicotiana* flowers, and tropane alkaloids in roots of Solanaceae [10–13]), other specialized metabolites are produced by or accumulate in specialized structures such as laticifers and trichomes [14–17]. Trichomes are structures that protrude from plant epidermal tissue. They can be unicellular or multicellular, glandular or non-glandular, and are abundant on above-ground tissues of many plants [14,16,18,19]. Secretory glandular trichomes (SGTs) often accumulate high concentrations of a wide array of specialized metabolites [14,16]. Their presence on the surface of aerial plant tissues allows the chemicals they produce to act as a first line of defense against organisms that might infect or consume plant tissues.

In addition to their role in protecting the plants that produce them, many SGT-localized specialized metabolites have found their way into human culture. Humans have used plant specialized metabolites, including those produced in SGTs, for thousands of years. In human hands, the uses for these compounds extend far beyond deterring herbivores and pathogens. For example, the sesquiterpene lactone artemisinin, which accumulates in SGTs of *Artemisia annua*, is used to treat malaria-causing *Plasmodium falciparum* and other human parasites [20,21]; the meroterpenoids cannabidiol and tetrahydrocannabinol produced by SGTs of *Cannabis* spp. have a long history of human recreational use [22,23]; and, ironically, numerous SGT-localized volatile monoterpenoids and phenylpropanoids that deter some herbivores from

consuming plants like *Ocimum basilicum* increase their culinary appeal to humans [24,25]. However, while humans have co-opted numerous SGT-localized compounds for purposes unrelated to their *in planta* roles, plants and humans have a common use for other specialized metabolites. In particular, the pyrethrins produced in SGTs of *Tanacetum cinerariifolium* protect both the plant and humans against insect pests.

EARLY USE AND DEVELOPMENT OF PYRETHRUM-BASED INSECTICIDES

Pyrethrins constitute a small class of specialized metabolites produced in Dalmatian pyrethrum (*Tanacetum cinerariifolium*) and provide the plant with an effective endogenous chemical defense against insect herbivores and fungal pathogens [26–29]. The insecticidal properties of pyrethrum have been known in western Europe and the United States since the 1840s but were likely discovered in eastern Europe as early as the late 17th century [30,31]. Pyrethrum products were used primarily as household insecticides in the 19th century [32]. By the early 20th century, they became tools for prevention of insect-borne diseases (e.g., malaria and yellow fever) [33,34] and alternatives to widely used agricultural pesticides with high mammalian toxicity (e.g., arsenic and cyanide) [35,36]. Identification of pyrethrins as the active components of pyrethrum in the early 1920s spurred research on their structure and mode of action [37]. This work eventually led to full structural elucidation of the six naturally occurring pyrethrins (Fig. 1.1A) and total syntheses for all six of these compounds [38–42] as well as the development of synthetic pyrethroid insecticides which exhibit increased environmental stability and toxicity toward insects relative to their natural counterparts (reviewed in [43]).

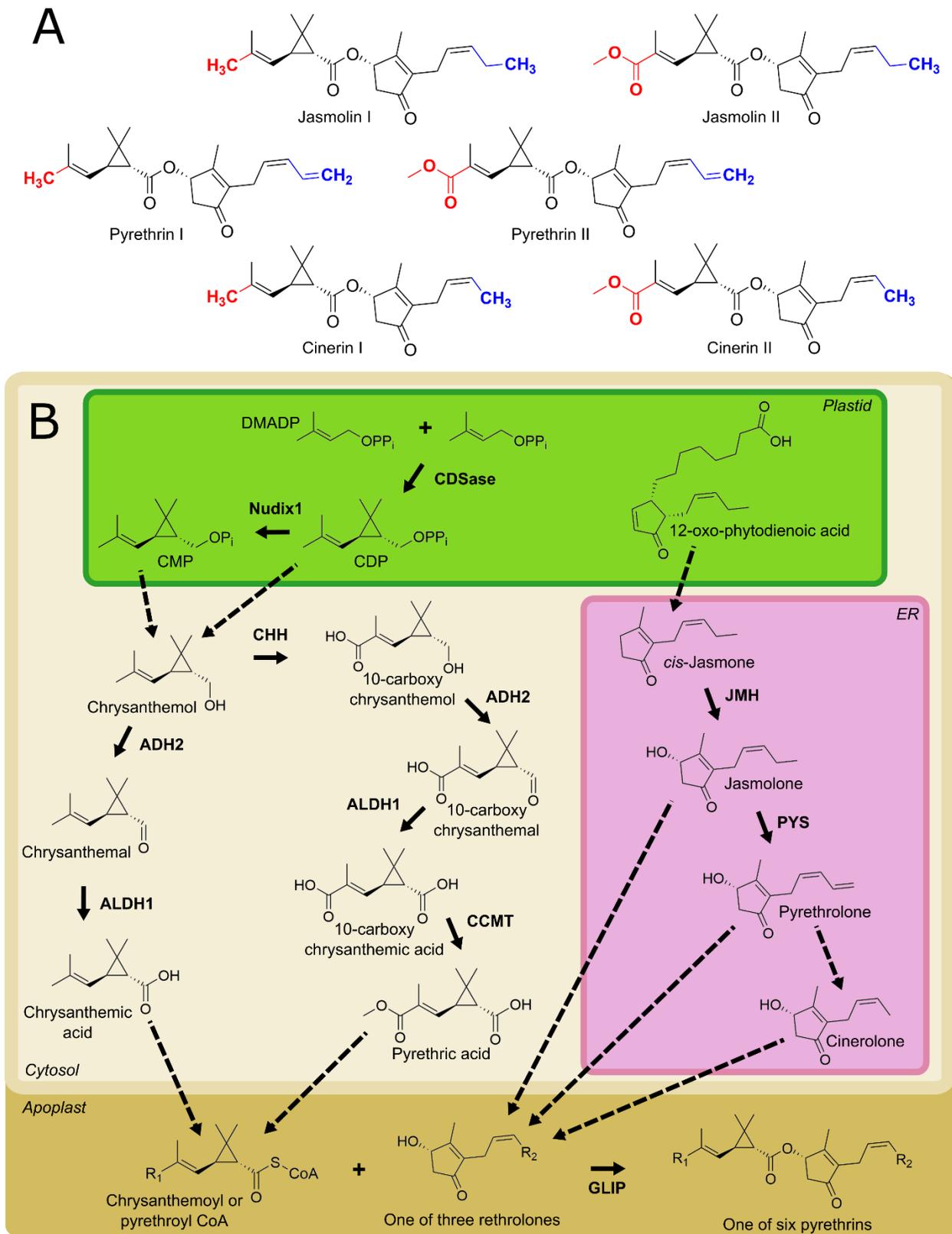


Figure 1.1 Pyrethrins and their biosynthesis.

Figure 1.1 (cont'd) (A) Structures of the six natural pyrethrins found in *Tanacetum cinerariifolium*. Variable portions of the monoterpene moiety are highlighted in red, while variable portions of the rethrolone moiety are highlighted in blue. (B) The monoterpene skeleton of pyrethrins is plastidially derived while subsequent modifications of this skeleton such as oxidation, methylation, and conjugation to Coenzyme A occur in the cytosol and apoplast. The oxylipin skeleton of pyrethrins is produced in the plastid while conversion to rethrolones occurs in the endoplasmic reticulum. Final conjugation of the monoterpene and rethrolone moieties of pyrethrins occurs in the apoplast. Enzymes are indicated in bold text; solid arrows indicate known steps in the pathway; dashed arrows indicate steps not yet elucidated. Abbreviations: ADH2 – alcohol dehydrogenase 2; ALDH1 – aldehyde dehydrogenase 1; CCMT – 10-carboxychrysanthemyl acid methyltransferase; CDP – chrysanthemyl diphosphate; CDSase – chrysanthemyl diphosphate synthase; CMP – chrysanthemyl monophosphate; CoA – coenzyme A; DMADP – dimethylallyl diphosphate; ER – endoplasmic reticulum; GLIP – GDSL lipase-like protein; JMH – jasmone hydroxylase; PYS – pyrethrolone synthase. R groups: R₁ – CH₃ or COOCH₃; R₂ – CH₃, CH=CH₂, or CH₂CH₃.

When applied as a spray or powder, both natural pyrethrins and their synthetic pyrethroid derivatives cause knockdown and death of insects by binding to voltage-gated sodium channels in the insect nervous system, resulting in persistent channel activity [44–46].

Nearly 100 years since their initial identification, scientists are again exploring natural pyrethrins as viable agricultural insecticides [47–57]. This renewed interest stems from unintended consequences of high pyrethroid stability. Pyrethroids were initially developed to solve the problem of rapid pyrethrin photodegradation [58]. While the natural pyrethrins exhibit half-lives of two hours to two days in agricultural settings [59–61], synthetic pyrethroids exhibit half-lives of weeks to months [58,62]. The long half-lives of synthetic pyrethroids led to environmental persistence and ecological harm [63–65], as well as the development of knockdown resistance toward synthetic pyrethroids in agricultural insect pests and insect disease vectors [66–68]. In contrast, the natural pyrethrins remain effective against some insect pests that developed resistance to synthetic pyrethroids [69–72]. In the last several years, our understanding of the pyrethrin biosynthetic pathway expanded greatly with the number of characterized pyrethrin-specific biosynthetic genes increasing from two to nine [73–77]. Additionally, genetic engineering experiments demonstrated the promise of increasing commercial pyrethrin production via heterologous hosts or engineering crop plants with endogenous pyrethrin defenses [78,79]. These advances portend the beginning of a new era for use of pyrethrins as agricultural insecticides.

INDUSTRIAL PYRETHRINS PRODUCTION

Industrial production of the pyrethrin insecticides currently requires large-scale cultivation of Dalmatian pyrethrum. Pyrethrins accumulate to 1-2% of dry mass in the mature

flower heads which are then harvested, dried, and powdered [26,80]. The powdered material may then be marketed directly or extracted with organic solvents for formulation into insecticidal soaps and sprays [81]. In addition to dry powders and liquid sprays used for small-scale ground-level treatment [82] or large-scale aerial application [83], pyrethrins are also formulated into other products such as lotions and mosquito coils for personal insect protection [84,85].

While *T. cinerariifolium* was originally harvested in its native Dalmatia in present-day Croatia, the crop was introduced into Japan in the late 19th century and by the 1930s, Japan produced most of the world's supply [28,86,87]. More than a dozen countries have participated in industrial pyrethrum production with significant sources of pyrethrum coming from Africa, Asia, Europe, and South America (see: <http://www.fao.org/faostat/>). However, Japan maintained a virtual monopoly on the pyrethrum market until World War II, after which east African nations took over most production. By the mid-1980s, Japanese pyrethrum production was negligible. Kenya dominated the pyrethrum market for the second half of the 20th century but production dropped sharply in the mid-2000s. Currently, the major commercial pyrethrum producers are Rwanda, Tanzania, and the Australian state of Tasmania [88,89]. Total levels of global industrial pyrethrum production have also fluctuated widely during the past half-century. Since the Food and Agriculture Organization of the United Nations (FAO) began keeping records of pyrethrum production in 1961, production has ranged from a record high of more than 30,000 metric tons in 1983 to an apparent low of less than 5,000 tons in 2007, which marked the end of major production in Kenya. Production in Tanzania and Rwanda increased after this point, stabilizing global production. However, Australian production is not reported to the FAO

and is thought to account for more than half of all global pyrethrum production [89], making reliable estimation of current global production difficult. FAO data indicates that global production excluding Australia totaled nearly 14,000 metric tons in 2017, the last year for which data are available. It is therefore likely that world production of pyrethrum approaches or exceeds the former 1983 record of 30,000 metric tons.

PYRETHRIN BIOSYNTHESIS

Natural pyrethrins comprise six esters, each consisting of a monoterpenoid acid moiety conjugated to a rethrolone-type oxylipin alcohol (Fig 1A) [39,40,90]. Their biosynthesis is known exclusively from *T. cinerariifolium*. Early feeding studies demonstrated that the pyrethrin biosynthetic pathway draws from two core plant metabolic pathways: the two monoterpenoids (chrysanthemic acid and pyrethric acid) are derived from the plastidial 1-deoxy-D-xylulose-5-phosphate (DXP) terpenoid pathway, while the three rethrolones (pyrethrolone, jasmolone, and cinerolone) are derived from the octadecanoid pathway (Fig. 1.1B) [91]. Pyrethrins containing chrysanthemic acid are termed 'type I' while those containing pyrethric acid are 'type II'. The three rethrolones termed pyrethrolone, jasmolone, and cinerolone are found in pyrethrins I and II, jasmolins I and II, and cinerins I and II, respectively. Pyrethrin I is the most abundant of the six pyrethrins in *T. cinerariifolium* while pyrethrin II is the second most abundant (relative abundances of jasmolins and cinerins vary among *T. cinerariifolium* varieties) [26].

Monoterpenoid biosynthesis

The full monoterpenoid pathway to chrysanthemic and pyrethric acids was recently elucidated with the aid of *T. cinerariifolium* transcriptomic and genomic resources (Fig. 1.1B)

[73,74,92–95]. Nearly two decades ago, it was shown that chrysanthemyl diphosphate synthase (CDSase) catalyzes the first step in biosynthesis of these monoterpenoids via an unusual head-to-middle condensation of two dimethylallyl diphosphate (DMADP) units in plastids (Fig. 1.1B) [29,93]. Further conversion of chrysanthemyl diphosphate (CDP) to the downstream acids requires dephosphorylation and oxidation; the action of one or more plastidial phosphatases was predicted to occur before oxidation of chrysanthemol in the cytosol. However, while initial characterization of CDSase identified CDP as the reaction product, results of subsequent investigations suggested that CDSase might perform both the condensation of the two DMADP molecule precursors to give the cyclic monoterpene skeleton *and* cleavage of the diphosphate group, yielding chrysanthemol [92]. This ‘chrysanthemol synthase’ model allowed for diffusion of the first monoterpene product directly into the cytosol without invoking as-of-yet undiscovered phosphatases. Recently, a Nudix-family phosphatase from *T. cinerariifolium*, Nudix1, was characterized that specifically dephosphorylates CDP yielding chrysanthemyl monophosphate (Fig. 1.1B) [77]. Therefore, the involvement of this and other phosphatases, perhaps in addition to the dephosphorylating activity of CDS, cannot yet be ruled out [74,77].

Regardless of how chrysanthemol is generated from CDP, all evidence shows this compound as the branch point for the synthesis of both chrysanthemic acid and pyrethric acid. Once in the cytosol, the chrysanthemol hydroxyl can be directly modified by two oxidoreductases validated *in vitro* and *in planta*, alcohol dehydrogenase 2 (ADH2) and aldehyde dehydrogenase 1 (ALDH1), which catalyze sequential oxidation of chrysanthemol to produce chrysanthemic acid (Fig. 1.1B) [74]. A portion of the chrysanthemol pool can be hydroxylated at the C10 position by the cytochrome P450 chrysanthemol 10-hydroxylase (CHH; CYP71BZ1),

yielding the dihydroxylated compound 10-hydroxychrysanthemol [73]. The 10-hydroxyl group of this compound is converted to a carboxylic acid group by two additional oxidation steps catalyzed by CHH, yielding 10-carboxychrysanthemol, while the C1 hydroxyl group is oxidized to a carboxylic acid group by ADH2 and ALDH1, as in the biosynthesis of chrysanthemic acid described above. Transient expression of CDSase, ADH2, ALDH1, and CHH in *Nicotiana benthamiana* leaves indicated that oxidation of the 10-hydroxy group to the carboxylic acid by CHH precedes oxidation of the C1 hydroxyl by ADH2 and ALDH1. The combined result of the actions of these three enzymes is 10-carboxychrysanthemic acid [73,74]. The 10-carboxychrysanthemic acid molecule is then methylated by the SABATH-family 10-carboxychrysanthemic acid 10-methyltransferase (CCMT), yielding pyrethric acid (Fig. 1.1B) [73].

Current understanding of the pathway requires that both chrysanthemic and pyrethric acids are conjugated to coenzyme A (CoA) prior to incorporation of the monoterpenoid moiety into pyrethrins [96]. A chrysanthemic acid:CoA ligase from *T. cinerariifolium* (acyl activating enzyme 1; AAE1), was described (T. Yang, PhD Thesis, Wageningen University, 2013). However, this enzyme was plastid-localized while formation of chrysanthemic acid is likely cytosolic. Additionally, testing AAE1 enzyme activity *in vitro* or via transient expression in *N. benthamiana* with other pyrethrin pathway genes failed to demonstrate CoA-ligating activity with chrysanthemic acid (H. Xu and E. Pichersky, unpublished).

Rethrolone biosynthesis

In contrast to biosynthesis of pyrethrin monoterpenoid moieties, biosynthesis of rethrolone moieties is still less well understood. Structural similarities between rethrolones and

octadecanoids such as jasmonic acid suggested a common origin. This hypothesis was supported by feeding of pyrethrin-producing *T. cinerariifolium* flowers with [1-¹³C]-D-glucose, which yielded a pyrethrin ¹³C labelling pattern consistent with a linolenic acid precursor for rethrolones [91,97]. Additionally, transcripts of octadecanoid pathway genes *lipoxygenase 1* (*LOX1*), *allene oxide synthase* (*AOS*), and *allene oxide cyclase* (*AOC*) show co-expression with *CDSase* in *T. cinerariifolium* flowers [76,98]. However, enzymatic activities have not been confirmed for the *AOS* or *AOC* gene products. Recent labelling studies demonstrated that both 12-oxophytodienoic acid (OPDA) and *cis*-jasmane are precursors of rethrolones (Fig. 1.1B) while jasmonic acid is not [99]. This not only illuminates pyrethrin biosynthesis but also provides insight into the thus far incompletely resolved origins of *cis*-jasmane in plants [99–101].

While no enzymes from *T. cinerariifolium* have been identified that catalyze formation of the *cis*-jasmane intermediate, the cytochrome P450 jasmane hydroxylase (JMh; CYP71AT148) hydroxylates the 3-position of *cis*-jasmane, yielding jasmolone which may then be incorporated into jasmolins I and II (Fig. 1.1B) [76]. Alternatively, another cytochrome P450, pyrethrolone synthase (PYS; CYP82Q3), can desaturate the terminal bond of the jasmolone pentenyl tail, yielding pyrethrolone which may then be incorporated into pyrethrins I and II (Fig. 1.1B) [75]. The biosynthetic route to cinerolone, the third rethrolone and precursor for cinerins I and II, is still unknown but may proceed by further oxidation and decarboxylation of the terminal carbon of the side chain.

Condensation of monoterpenoids and rethrolones

Following formation of rethrolones and monoterpenoid acid CoA esters, the monoterpenoid moiety must be transferred from CoA to the rethrolone. This is catalyzed by the

GDSL lipase-like protein (GLIP), which can accept either chrysanthemoyl-CoA or pyrethroyl-CoA as an acyl donor and any of jasmolone, pyrethrolone, or cinerolone as an acyl acceptor, thereby generating any of the six natural pyrethrins (Fig. 1.1A,B) [96]. In contrast to all other known steps in pyrethrin biosynthesis, this step occurs in the apoplast.

REGULATION OF PYRETHRIN BIOSYNTHESIS

Spatiotemporal specificity of pyrethrin biosynthesis

As with many specialized metabolites [102–104], production of pyrethrins is spatially restricted in *T. cinerariifolium*. Pyrethrins accumulate throughout above-ground parts of the plants but their levels in tissues are correlated with the presence of secretory glandular trichomes which are more abundant on the surface of disc floret ovaries of *T. cinerariifolium* than in any other plant tissue (Fig. 2A,B) [105]. Trichome density and pyrethrin levels increase as the flower buds develop, reaching their maximum in senescing flower heads [105,106]. Pyrethrin biosynthetic gene expression parallels pyrethrin accumulation: transcript analysis indicated that all genes associated with pyrethrin biosynthesis were expressed primarily in disc florets with relatively low levels of expression observed in the larger ray florets [73–77]. Just as overall levels of pyrethrin accumulation correlated with trichome density on the surface of disc florets [107], seven of nine presently known pyrethrin biosynthetic genes (*CDSase*, *Nudix1*, *ADH2*, *ALDH1*, *CHH*, *JMH*, and *PYS*) were primarily expressed in trichomes [73,75–77]. The *CDSase* promoter was also shown to drive expression of reporter genes specifically in trichome tip cells of *Chrysanthemum x morifolium*, a hybrid chrysanthemum closely related to *T. cinerariifolium*, and in the more phylogenetically distant *Nicotiana tabacum* [108]. *CDSase* enzyme activity was found at high levels in trichomes of developing seeds but at low levels in

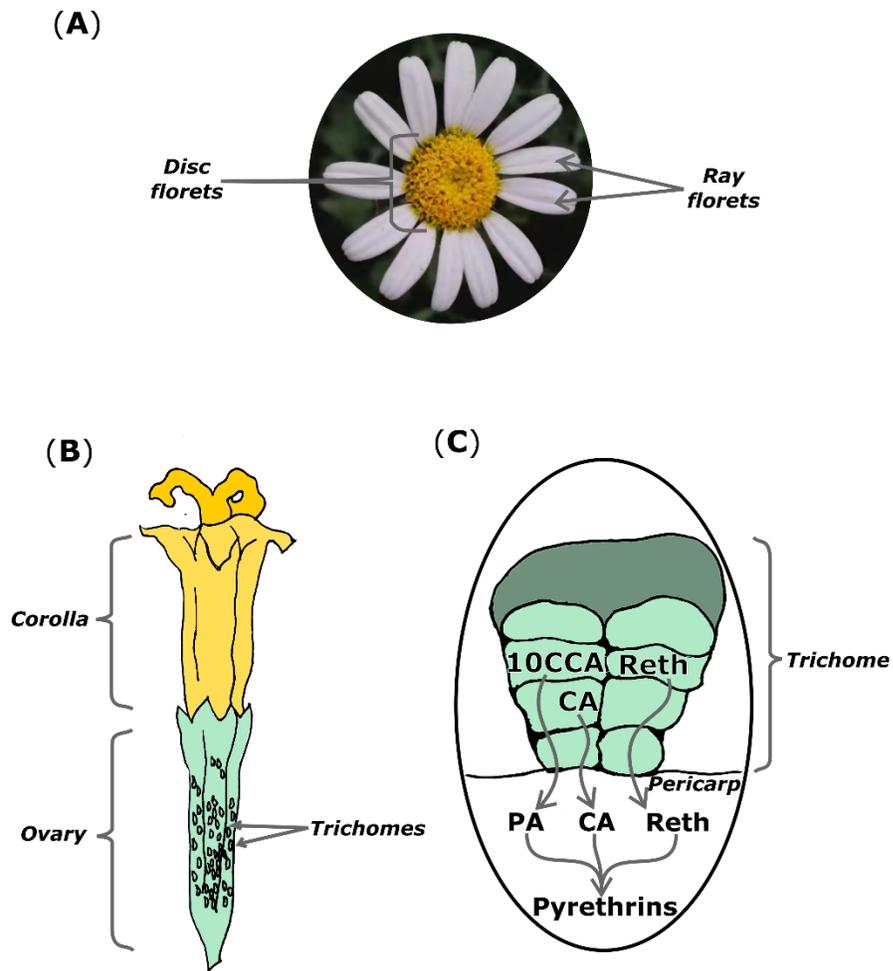


Figure 1.2 Production of pyrethrins in flowers of *Tanacetum cinerariifolium*. (A) Pyrethrins accumulate primarily in disc florets of mature flowers while levels in ray florets are relatively low. (B) Pyrethrins production occurs in the trichome-laden ovaries of disc florets. (C) The rethrolone and chrysanthemic acid precursors of pyrethrins are biosynthesized in trichomes covering the disc floret ovary while pyrethric acid is produced in the ovary pericarp; mature pyrethrins accumulate in the ovary pericarp. Abbreviations: CA – chrysanthemic acid; PA – pyrethric acid; reth – rethrolone; 10CCA – 10-carboxychrysanthemic acid. Photograph in (A) by H. Xu.

the ovaries and vegetative portions of the plant [29]. Expression of the oxylipin biosynthetic genes *LOX1*, *AOS*, and *AOC* were also primarily expressed in trichomes [76,98]. As the actions of *LOX1*, *AOS*, and *AOC* enzymes are also involved in jasmonic acid biosynthesis, restriction of the pyrethrin pathway to trichomes may facilitate high-level pyrethrin production without interfering with jasmonate signaling in adjacent tissues.

In contrast to the trichome-localized expression of most pyrethrin biosynthetic genes, the genes encoding the final enzyme in pyrethric acid biosynthesis (*CCMT*) and the enzyme catalyzing the final esterification step in biosynthesis of all six pyrethrins (*GLIP*) are primarily expressed in ovary tissue [73,76]. This expression pattern, combined with the apoplastic localization of the *GLIP* enzyme, suggests that biosynthesis of pyrethrin monoterpenoid and rethrolone precursors occurs in the trichome followed by export of these components into the apoplast of underlying tissue and subsequent esterification to form the final pyrethrin products (Fig. 2C) [29,73].

Analysis of pyrethrin biosynthetic gene expression throughout floral development indicates that, in addition to tissue-level restriction, pyrethrin biosynthesis is also temporally restricted with the bulk of biosynthesis occurring in young developing flower buds. Accumulation of mRNAs from all pyrethrin biosynthetic genes was high during early stages of development but decreased as the flowers opened and matured, consistent with the gradual increase in pyrethrin content in early development that slows at later stages [73–77,106]. Paradoxically, young *T. cinerariifolium* seedlings also showed significant levels of pyrethrins but relatively low levels of *CDSase* transcript and possessed no *CDSase* activity [29], suggesting that transport of pyrethrins from surrounding ovary tissue into developing seeds occurs. Analysis of

pyrethrin localization within the ovaries of mature flowers revealed that pyrethrin levels were high in pericarp tissue of developing achenes at late stages of floral development while pyrethrin levels in the developing embryo itself were low; however, ripe achenes contained little pyrethrin in the pericarp, instead accumulating the bulk of pyrethrins in the mature embryo [29]. Transfer of pyrethrins from floral tissue into developing achenes and the subsequent accumulation of pyrethrins in the embryo evidently provides young *T. cinerariifolium* seedlings with pre-synthesized chemical defenses, thus protecting the otherwise vulnerable seedlings from insect attack and fungal pathogens in the absence of pyrethrin biosynthetic gene expression.

Pyrethrin production in response to wounding and volatiles

While production of some plant defensive specialized metabolites is constitutive or developmentally regulated, other defensive compounds are induced (*e.g.*, by herbivory, wounding, or abiotic stress) [109–111]. Pyrethrin production in floral tissues of *T. cinerariifolium* was constitutive but pathway gene expression and pyrethrin accumulation in vegetative tissues was inducible by mechanical wounding, application of methyl jasmonate, and volatile organic compounds (VOCs) [112–115]. While mechanical wounding of *T. cinerariifolium* plants grown under field conditions had no effect on total pyrethrin levels [112], wounding led to increased levels of pyrethrins in *T. cinerariifolium* plants grown in growth chambers [111,114]. There is evidence that VOCs emitted in response to wounding stimulate pyrethrin production in vegetative tissues. When *T. cinerariifolium* seedlings were grown near wounded conspecifics, levels of pyrethrins I and II increased [113]. In contrast, when portions of wounded plants were protected from exposure to their own VOCs, levels of pyrethrin II in protected

tissues increased while levels of pyrethrin I remained unchanged, suggesting that some portions of the pyrethrin biosynthetic pathway were triggered via a systemic response while others responded only to an external volatile signal [111].

Analysis of volatile organic compounds (VOCs) emitted by wounded *T. cinerariifolium* plants revealed five major components: the green leaf volatiles (*Z*)-3-hexenal, (*E*)-2-hexenal, (*Z*)-3-hexen-1-ol, and (*Z*)-3-hexen-1-yl acetate, and the sesquiterpene (*E*)- β -farnesene which protects *T. cinerariifolium* from aphids by mimicking their natural alarm pheromones [116,117]. Consistent with reports from other species, exposure of *T. cinerariifolium* seedlings to a cocktail of VOCs similar to those observed in wounded conspecifics induced expression of several genes involved in specialized metabolism including genes involved in pyrethrin biosynthesis [113,117,118]. Specifically, genes from the DXP and oxylipin pathways as well as the pyrethrin-specific genes *CDSase* and *GLIP* showed increased expression at various time-points ranging from three to 12 hours after seedling exposure to VOCs [113,117].

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Pyrethrins and their derivatives have remained essential tools for pest control for over 100 years. With increased concerns about ecological harm caused by synthetic pesticides, natural solutions for pest control in agricultural and household settings are in greater demand than ever. The recent expansion of knowledge about biosynthesis and regulation of pyrethrins in *T. cinerariifolium* is timely and will help to improve industrial pyrethrin yields. Combined with modern advances in genetic engineering and genome editing, we can also look towards production of pyrethrins in heterologous hosts or innovative crop protection strategies such as designer crops containing pyrethrin pathway genes to improve endogenous chemical defenses

against insect herbivores. Finally, there is the possibility of combining genetic engineering *in planta* with an *in vitro* process to optimize the commercial production of natural pyrethrins. The two moieties – the monoterpeneoid acids and the rethrolones – are stereochemicals that are best produced in genetically engineered organisms rather than by organic synthesis. If production of pyrethrins in an engineered system falls short of projections based on precursor accumulation, precursors could be biosynthesized in different organisms, purified, and processed to remove non-specific conjugations such as glycosylation. Once stereospecific intermediates are obtained in large quantities by purification from the host organisms, the two moieties could be combined *in vitro*, possibly using GLIP or similar lipases.

While we are already well on our way to elucidating the full pyrethrin biosynthetic pathway, a few questions remain. The route to the major pyrethrin constituents, pyrethrins and jasmolins I and II, has been elucidated but we lack an understanding of how cinerolone, a component of the minor pyrethrins cinerin I and II, is formed. Additionally, engineering of the full pathway into crop plants may benefit from a better understanding of pathway regulation in *T. cinerariifolium*. In the native producer, pyrethrin precursors are synthesized in trichomes but the mature pyrethrins are assembled in underlying tissue. This spatial separation of biosynthetic steps may be necessary for effective pyrethrin formation and may require transporters to shuttle pyrethrin precursors between tissues. If so, steps must be taken to ensure proper gene regulation or tissue-specific transporter expression when introducing the pathway into heterologous hosts.

REFERENCES

REFERENCES

- 1 Ng, J.L.P. *et al.* (2015) Flavonoids and Auxin Transport Inhibitors Rescue Symbiotic Nodulation in the *Medicago truncatula* Cytokinin Perception Mutant cre1. *Plant Cell* 27, 2210–2226
- 2 Song, N. *et al.* (2019) An ERF2-like transcription factor regulates production of the defense sesquiterpene capsidiol upon *Alternaria alternata* infection. *J. Exp. Bot.* 70, 5895–5908
- 3 Haverkamp, A. *et al.* (2016) Hawkmoths evaluate scenting flowers with the tip of their proboscis. *eLife* 5, e15039
- 4 Kessler, D. *et al.* (2019) The defensive function of a pollinator-attracting floral volatile. *Funct. Ecol.* 33, 1223–1232
- 5 Demasi, S. *et al.* (2019) Ailanthone from *Ailanthus altissima* (Mill.) Swingle as potential natural herbicide. *Sci. Hortic.* 257, 108702
- 6 Zhao, M. *et al.* (2020) Sesquiterpene glucosylation mediated by glucosyltransferase UGT91Q2 is involved in the modulation of cold stress tolerance in tea plants. *New Phytol.* 226, 362–372
- 7 Pichersky, E. and Lewinsohn, E. (2011) Convergent Evolution in Plant Specialized Metabolism. *Annu. Rev. Plant Biol.* 62, 549–566
- 8 Fan, P. *et al.* (2017) Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. *Nat. Commun.* 8, 2080
- 9 Moghe, G.D. *et al.* (2017) Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife* 6, e28468
- 10 dos Santos, B.M. *et al.* (2019) Quantification and Localization of Formylated Phloroglucinol Compounds (FPCs) in *Eucalyptus* Species. *Front. Plant Sci.* 10, 186
- 11 Yu, C. *et al.* Tissue-specific study across the stem of *Taxus media* identifies a phloem-specific TmMYB3 involved in the transcriptional regulation of paclitaxel biosynthesis. *Plant J.* DOI: 10.1111/tpj.14710
- 12 Guo, H. *et al.* (2020) Evolution of a Novel and Adaptive Floral Scent in Wild Tobacco. *Mol. Biol. Evol.* 37, 1090–1099
- 13 Qiu, F. *et al.* (2020) Functional genomics analysis reveals two novel genes required for littorine biosynthesis. *New Phytol.* 225, 1906–1914

- 14 Huchelmann, A. *et al.* (2017) Plant Glandular Trichomes: Natural Cell Factories of High Biotechnological Interest. *Plant Physiol.* 175, 6–22
- 15 Gorpenchenko, T.Y. *et al.* (2019) Tempo-Spatial Pattern of Stepharine Accumulation in *Stephania glabra* Morphogenic Tissues. *Int. J. Mol. Sci.* 20, 808
- 16 Liu, Y. *et al.* (2019) Non-volatile natural products in plant glandular trichomes: chemistry, biological activities and biosynthesis. *Nat. Prod. Rep.* 36, 626–665
- 17 Benninghaus, V.A. *et al.* (2020) Comparative proteome and metabolome analyses of latex-exuding and non-exuding *Taraxacum koksaghyz* roots provide insights into laticifer biology. *J. Exp. Bot.* 71, 1278–1293
- 18 Bar, M. and Shtein, I. (2019) Plant trichomes and the biomechanics of defense in various systems, with Solanaceae as a model. *Botany* 97, 651–660
- 19 Karabourniotis, G. *et al.* (2020) Protective and defensive roles of non-glandular trichomes against multiple stresses: structure-function coordination. *J. For. Res.* 31, 1–12
- 20 Lommen, W.J.M. *et al.* (2006) Trichome dynamics and artemisinin accumulation during development and senescence of *Artemisia annua* leaves. *Planta Med.* 72, 336–345
- 21 Wang, J. *et al.* (2015) Haem-activated promiscuous targeting of artemisinin in *Plasmodium falciparum*. *Nat. Commun.* 6, 10111
- 22 McKenna, T.K. (1993) *Food of the Gods: the search for the original tree of knowledge: a radical history of plants, drugs, and human evolution*, Bantam Books.
- 23 Happyana, N. *et al.* (2013) Analysis of cannabinoids in laser-microdissected trichomes of medicinal *Cannabis sativa* using LCMS and cryogenic NMR. *Phytochemistry* 87, 51–59
- 24 Carvalho, S.D. *et al.* (2016) Light Quality Dependent Changes in Morphology, Antioxidant Capacity, and Volatile Production in Sweet Basil (*Ocimum basilicum*). *Front. Plant Sci.* 7, 1328
- 25 Litvin, A. *et al.* (2020) Effects of Supplemental Light Source on Basil, Dill, and Parsley Growth, Morphology, Aroma, and Flavor. *J. Am. Soc. Hortic. Sci.* 145, 18–29
- 26 Grdisa, M. *et al.* (2013) Chemical Diversity of the Natural Populations of Dalmatian Pyrethrum (*Tanacetum cinerariifolium* (Trevir.) Sch.Bip.) in Croatia. *Chem. Biodivers.* 10, 460–472
- 27 Yang, T. *et al.* (2012) Pyrethrins Protect Pyrethrum Leaves Against Attack by Western Flower Thrips, *Frankliniella occidentalis*. *J. Chem. Ecol.* 38, 370–377

- 28 Pares, B. *et al.* (1925) Economic Survey: The Economic Situation in Yugoslavia. *Slav. Rev.* 4, 491–505
- 29 Ramirez, A.M. *et al.* (2012) Bidirectional Secretions from Glandular Trichomes of Pyrethrum Enable Immunization of Seedlings. *Plant Cell* 24, 4252–4265
- 30 McLaughlin, G.A. (1973) History of Pyrethrum. In *Pyrethrum: The Natural Insecticide* (Casida, J. E., ed), pp. 3–15, Academic Press
- 31 Clark, J.F.M. (2001) Bugs in the system: Insects, agricultural science, and professional aspirations in Britain, 1890-1920. *Agric. Hist.* 75, 83–114
- 32 Lange, H.W. and Akesson, N.B. (1973) Pyrethrum for Control of Agricultural Insects. In *Pyrethrum: The Natural Insecticide* (Casida, J. E., ed), pp. 261–279, Academic Press
- 33 Orenstein, A.J. (1913) Mosquito Catching in Dwellings in the Prophylaxis of Malaria. *Am. J. Public Health* 3, 106–110
- 34 Reed, W. and Carroll, J. (1901) The Prevention of Yellow Fever. *Public Health Pap. Rep.* 27, 113–129
- 35 Fryer, J.C.F. *et al.* (1928) English-grown pyrethrum as an insecticide. I. *Ann. Appl. Biol.* 15, 423–445
- 36 Richardson, C.H. *et al.* (1937) The toxicity of certain insecticides to the chinch bug. *J. Agric. Res.* 54, 0059–0078
- 37 Staudinger, H. and Ruzicka, L. (1924) Substances for killing insects I. The isolation and constitution of effective parts of dalmatian insect powder. *Helv. Chim. Acta* 7, 177–201
- 38 Laforge, F.B. and Barthel, W.F. (1944) Constituents of pyrethrum flowers XVI Heterogeneous nature of pyrethrolone. *J. Org. Chem.* 9, 242–249
- 39 Laforge, F. and Barthel, W. (1947) Constituents of Pyrethrum Flowers .20. the Partial Synthesis of Pyrethrins and Cinerins and Their Relative Toxicities. *J. Org. Chem.* 12, 199–202
- 40 Godin, P. *et al.* (1965) Insecticidal Activity of Jasmolin 2 and Its Isolation from Pyrethrum (*Chrysanthemum cinerariaefolium* Vis). *J. Econ. Entomol.* 58, 548-
- 41 Rugutt, J.K. *et al.* (1999) NMR and Molecular Mechanics Study of Pyrethrins I and II. *J. Agric. Food Chem.* 47, 3402–3410
- 42 Kawamoto, M. *et al.* (2020) Total Syntheses of All Six Chiral Natural Pyrethrins: Accurate Determination of the Physical Properties, Their Insecticidal Activities, and Evaluation of Synthetic Methods. *J. Org. Chem.* 85, 2984–2999

- 43 Matsuo, N. (2019) Discovery and development of pyrethroid insecticides. *Proc. Jpn. Acad. Ser. B-Phys. Biol. Sci.* 95, 378–400
- 44 Amar, M. *et al.* (1992) Patch-Clamp Analysis of the Effects of the Insecticide Deltamethrin on Insect Neurons. *J. Exp. Biol.* 163, 65–84
- 45 McCavera, S.J. and Soderlund, D.M. (2012) Differential state-dependent modification of inactivation-deficient Na(v)1.6 sodium channels by the pyrethroid insecticides S-bioallethrin, tefluthrin and deltamethrin. *Neurotoxicology* 33, 384–390
- 46 Chen, M. *et al.* (2018) Action of six pyrethrins purified from the botanical insecticide pyrethrum on cockroach sodium channels expressed in *Xenopus* oocytes. *Pest. Biochem. Physiol.* 151, 82–89
- 47 Van Timmeren, S. and Isaacs, R. (2013) Control of spotted wing drosophila, *Drosophila suzukii*, by specific insecticides and by conventional and organic crop protection programs. *Crop Prot.* 54, 126–133
- 48 Wyss, E. and Daniel, C. (2004) Effects of autumn kaolin and pyrethrin treatments on the spring population of *Dysaphis plantaginea* in apple orchards. *J. Appl. Entomol.* 128, 147–149
- 49 Sial, A.A. *et al.* (2019) Evaluation of organic insecticides for management of spotted-wing drosophila (*Drosophila suzukii*) in berry crops. *J. Appl. Entomol.* 143, 593–608
- 50 Joseph, S.V. (2018) Lethal and Sublethal Effects of Organically-Approved Insecticides against *Bagrada hilaris* (Hemiptera: Pentatomidae). *J. Entomol. Sci.* 53, 307–324
- 51 Tacoli, F. *et al.* (2017) Control of *Scaphoideus titanus* with Natural Products in Organic Vineyards. *Insects* 8, 129
- 52 Razze, J.M. *et al.* (2016) Evaluation of Bioinsecticides for Management of *Bemisia tabaci* (Hemiptera: Aleyrodidae) and the Effect on the Whitefly Predator *Delphastus catalinae* (Coleoptera: Coccinellidae) in Organic Squash. *J. Econ. Entomol.* 109, 1766–1771
- 53 Oliveira, C.R. *et al.* (2019) Nanopesticide based on botanical insecticide pyrethrum and its potential effects on honeybees. *Chemosphere* 236, 124282
- 54 Shrestha, G. *et al.* (2020) Spinosad and Mixtures of an Entomopathogenic Fungus and Pyrethrins for Control of *Sitona lineatus* (Coleoptera: Curculionidae) in Field Peas. *J. Econ. Entomol.* 113, 669–678
- 55 Yao, J. *et al.* (2019) Differential susceptibilities of two closely-related stored product pests, the red flour beetle (*Tribolium castaneum*) and the confused flour beetle (*Tribolium confusum*), to five selected insecticides. *J. Stored Prod. Res.* 84, 101524

- 56 Fernandez-Grandon, G.M. *et al.* (2020) Additive Effect of Botanical Insecticide and Entomopathogenic Fungi on Pest Mortality and the Behavioral Response of Its Natural Enemy. *Plants* 9, 173
- 57 Korunić, Z. *et al.* (2020) Evaluation of diatomaceous earth formulations enhanced with natural products against stored product insects. *J. Stored Prod. Res.* 86, 101565
- 58 Demoute, J.-P. (1989) A brief review of the environmental fate and metabolism of pyrethroids. *Pestic. Sci.* 27, 375–385
- 59 Antonious, G.F. (2004) Residues and Half-Lives of Pyrethrins on Field-Grown Pepper and Tomato. *J. Environ. Sci. Heal. B* 39, 491–503
- 60 Pan, L. *et al.* (2017) Dissipation and Residues of Pyrethrins in Leaf Lettuce under Greenhouse and Open Field Conditions. *Int. J. Environ. Res. Public Health* 14, 822
- 61 Feng, X. *et al.* (2018) Residue analysis and risk assessment of pyrethrins in open field and greenhouse turnips. *Environ. Sci. Pollut. Res.* 25, 877–886
- 62 Katagi, T. (1991) Photodegradation of the pyrethroid insecticide esfenvalerate on soil, clay minerals, and humic acid surfaces. *J. Agric. Food Chem.* 39, 1351–1356
- 63 Hartz, K.E.H. *et al.* (2019) Survey of bioaccessible pyrethroid insecticides and sediment toxicity in urban streams of the northeast United States. *Environ. Pollut.* 254, UNSP 112931
- 64 Li, H. *et al.* (2017) Global occurrence of pyrethroid insecticides in sediment and the associated toxicological effects on benthic invertebrates: An overview. *J. Hazard. Mater.* 324, 258–271
- 65 Stehle, S. and Schulz, R. (2015) Agricultural insecticides threaten surface waters at the global scale. *Proc. Natl. Acad. Sci.* 112, 5750–5755
- 66 Maestre-Serrano, R. *et al.* (2019) Co-occurrence of V1016I and F1534C mutations in the voltage-gated sodium channel and resistance to pyrethroids in *Aedes aegypti* (L.) from the Colombian Caribbean region. *Pest Manag. Sci.* 75, 1681–1688
- 67 Davila-Barboza, J. *et al.* (2019) Novel Kdr mutations (K964R and A943V) in pyrethroid-resistant populations of *Triatoma mazzottii* and *Triatoma longipennis* from Mexico and detoxifying enzymes. *Insect Sci.* 26, 809–820
- 68 Cheng, X. *et al.* (2019) Pyrethroid resistance in the pest mite, *Halotydeus destructor*: Dominance patterns and a new method for resistance screening. *Pest. Biochem. Physiol.* 159, 9–16

- 69 Duchon, S. *et al.* (2009) Pyrethrum: A Mixture of Natural Pyrethrins Has Potential for Malaria Vector Control. *J. Med. Entomol.* 46, 516–522
- 70 Anderson, J.F. and Cowles, R.S. (2012) Susceptibility of *Cimex lectularius* (Hemiptera: Cimicidae) to Pyrethroid Insecticides and to Insecticidal Dusts With or Without Pyrethroid Insecticides. *J. Econ. Entomol.* 105, 1789–1795
- 71 Scott, J.G. *et al.* (2013) Insecticide resistance in house flies from the United States: Resistance levels and frequency of pyrethroid resistance alleles. *Pest. Biochem. Physiol.* 107, 377–384
- 72 Kgoroebutswe, T.K. *et al.* (2020) Distribution of *Anopheles* mosquito species, their vectorial role and profiling of knock-down resistance mutations in Botswana. *Parasitol. Res.* DOI: 10.1007/s00436-020-06614-6
- 73 Xu, H. *et al.* (2019) Pyrethric acid of natural pyrethrin insecticide: complete pathway elucidation and reconstitution in *Nicotiana benthamiana*. *New Phytol.* 223, 751–765
- 74 Xu, H. *et al.* (2018) Coexpression Analysis Identifies Two Oxidoreductases Involved in the Biosynthesis of the Monoterpene Acid Moiety of Natural Pyrethrin Insecticides in *Tanacetum cinerariifolium*. *Plant Physiol.* 176, 524–537
- 75 Li, W. *et al.* (2019) Pyrethrin Biosynthesis: The Cytochrome P450 Oxidoreductase CYP82Q3 Converts Jasmolone To Pyrethrolone. *Plant Physiol.* 181, 934–944
- 76 Li, W. *et al.* (2018) Jasmone Hydroxylase, a Key Enzyme in the Synthesis of the Alcohol Moiety of Pyrethrin Insecticides. *Plant Physiol.* 177, 1498–1509
- 77 Li, W. *et al.* (2020) A Trichome-Specific, Plastid-Localized *Tanacetum cinerariifolium* Nudix Protein Hydrolyzes the Natural Pyrethrin Pesticide Biosynthetic Intermediate trans-Chrysanthemyl Diphosphate. *Front. Plant Sci.* 11,
- 78 Hu, H. *et al.* (2018) Modification of chrysanthemum odour and taste with chrysanthemol synthase induces strong dual resistance against cotton aphids. *Plant Biotechnol. J.* 16, 1434–1445
- 79 Xu, H. *et al.* (2018) Production of trans-chrysanthemic acid, the monoterpene acid moiety of natural pyrethrin insecticides, in tomato fruit. *Metab. Eng.* 47, 271–278
- 80 Li, J. *et al.* (2014) Comparative analysis of pyrethrin content improvement by mass selection, family selection and polycross in pyrethrum [*Tanacetum cinerariifolium* (Trevir.) Sch.Bip.] populations. *Ind. Crop. Prod.* 53, 268–273
- 81 Ginsburg, J.M. and Kent, C. (1937) The Effect of Soap Sprays on Plants. *J. N. Y. Entomol. Soc.* 45, 109–113

- 82 Cilek, J.E. *et al.* (2008) Evaluation of an Automatic-Timed Insecticide Application System for Backyard Mosquito Control. *J. Am. Mosq. Control Assoc.* 24, 560–565
- 83 Elnaiem, D.-E.A. *et al.* (2008) Impact of aerial spraying of pyrethrin insecticide on *Culex pipiens* and *Culex tarsalis* (Diptera: Culicidae) abundance and West Nile virus infection rates in an urban/suburban area of Sacramento County, California. *J. Med. Entomol.* 45, 751–757
- 84 John, N.A. and John, J. (2015) Prolonged use of mosquito coil, mats, and liquidators: A review of its health implications. *Int. J. Clin. Exp. Physiol.* 2, 209–213
- 85 Barker, S.C. and Altman, P.M. (2010) A randomised, assessor blind, parallel group comparative efficacy trial of three products for the treatment of head lice in children - melaleuca oil and lavender oil, pyrethrins and piperonyl butoxide, and a “suffocation” product. *BMC Dermatol.* 10, 6
- 86 Glassford, J. (1930) The economics of pyrethrum. *J. Econ. Entomol.* 23, 874–877
- 87 Grunge, W.H. (1939) Japan’s Pyrethrum Position Threatened. *Far Eastern Survey* 8, 109–110
- 88 Pethybridge, S.J. *et al.* (2008) Diseases of pyrethrum in tasmania: Challenges and prospects for management. *Plant Dis.* 92, 1260–1272
- 89 Ryan, R.F. *et al.* (2015) Pyrethrum: the Natural Choice in Pest Control. In *1st International Symposium on Pyrethrum, the Natural Insecticide: Scientific and Industrial Developments in the Renewal of a Traditional Industry* 1073 (Chung, B., ed), pp. 131–135, Int. Soc. Horticultural Science
- 90 Crombie, L. and Holloway, S. (1985) Biosynthesis of the Pyrethrins - Unsaturated Fatty-Acids and the Origins of the Rethrolone Segment. *J. Chem. Soc.-Perkin Trans. 1* DOI: 10.1039/p19850001393
- 91 Matsuda, K. *et al.* (2005) Biosynthesis of pyrethrin I in seedlings of *Chrysanthemum cinerariaefolium*. *Phytochemistry* 66, 1529–1535
- 92 Yang, T. *et al.* (2014) Chrysanthemyl Diphosphate Synthase Operates *in Planta* as a Bifunctional Enzyme with Chrysanthemol Synthase Activity. *J. Biol. Chem.* 289, 36325–36335
- 93 Rivera, S.B. *et al.* (2001) Chrysanthemyl diphosphate synthase: Isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from *Chrysanthemum cinerariaefolium*. *Proc. Natl. Acad. Sci.* 98, 4373–4378
- 94 Yamashiro, T. *et al.* (2019) Draft genome of *Tanacetum cinerariifolium*, the natural source of mosquito coil. *Sci. Rep.* 9, 18249

- 95 Khan, S. *et al.* (2017) Comparative transcriptome analysis reveals candidate genes for the biosynthesis of natural insecticide in *Tanacetum cinerariifolium*. *BMC Genomics* 18, 54
- 96 Kikuta, Y. *et al.* (2012) Identification and characterization of a GDSL lipase-like protein that catalyzes the ester-forming reaction for pyrethrin biosynthesis in *Tanacetum cinerariifolium*- a new target for plant protection. *Plant J.* 71, 183–193
- 97 Schaller, A. and Stintzi, A. (2009) Enzymes in jasmonate biosynthesis - Structure, function, regulation. *Phytochemistry* 70, 1532–1538
- 98 Ramirez, A.M. *et al.* (2013) A Trichome-Specific Linoleate Lipoxygenase Expressed During Pyrethrin Biosynthesis in *Pyrethrum*. *Lipids* 48, 1005–1015
- 99 Matsui, R. *et al.* (2020) Jasmonic acid is not a biosynthetic intermediate to produce the pyrethrolone moiety in pyrethrin II. *Sci Rep* 10, 6366
- 100 Wasternack, C. and Strnad, M. (2018) Jasmonates: News on Occurrence, Biosynthesis, Metabolism and Action of an Ancient Group of Signaling Compounds. *Int. J. Mol. Sci.* 19, 2539
- 101 Matsui, R. *et al.* (2019) Feeding experiment using uniformly C-13-labeled alpha-linolenic acid supports the involvement of the decarboxylation mechanism to produce cis-jasmone in *Lasiodiplodia theobromae*. *Biosci. Biotechnol. Biochem.* 83, 2190–2193
- 102 Dastmalchi, M. *et al.* (2019) Purine Permease-Type Benzylisoquinoline Alkaloid Transporters in Opium Poppy. *Plant Physiol.* 181, 916–933
- 103 Nakashima, T. *et al.* (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
- 104 Yamamoto, K. *et al.* (2019) The complexity of intercellular localisation of alkaloids revealed by single-cell metabolomics. *New Phytol.* 224, 848–859
- 105 Zito, S. *et al.* (1983) Distribution of Pyrethrins in Oil Glands and Leaf Tissue of *Chrysanthemum cinerariaefolium*. *Planta Med.* 47, 205–207
- 106 Head, S.W. (1973) Composition of *Pyrethrum* Extract and Analysis of Pyrethrins. In *Pyrethrum: The Natural Insecticide* (Casida, J. E., ed), pp. 25–53, Academic Press
- 107 Suraweera, D.D. *et al.* (2017) Dynamics of flower, achene and trichome development governs the accumulation of pyrethrins in *pyrethrum* (*Tanacetum cinerariifolium*) under irrigated and dryland conditions. *Ind. Crops Prod.* 109, 123–133

- 108 Sultana, S. *et al.* (2015) Molecular cloning and characterization of the trichome specific chrysanthemyl diphosphate/chrysanthemol synthase promoter from *Tanacetum cinerariifolium*. *Sci. Hortic.* 185, 193–199
- 109 Van Geem, M. *et al.* (2015) Interactions Between a Belowground Herbivore and Primary and Secondary Root Metabolites in Wild Cabbage. *J. Chem. Ecol.* 41, 696–707
- 110 Li, D. *et al.* (2015) Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. *Proc. Natl. Acad. Sci.* 112, E4147–E4155
- 111 Ueda, H. and Matsuda, K. (2011) VOC-mediated within-plant communications and nonvolatile systemic signals upregulate pyrethrin biosynthesis in wounded seedlings of *Chrysanthemum cinerariaefolium*. *J. Plant Interact.* 6, 89–91
- 112 Baldwin, I.T. *et al.* (1993) Foliar and floral pyrethrins of *Chrysanthemum cinerariaefolium* are not induced by leaf damage. *J. Chem. Ecol.* 19, 2081–2087
- 113 Kikuta, Y. *et al.* (2011) Specific Regulation of Pyrethrin Biosynthesis in *Chrysanthemum cinerariaefolium* by a Blend of Volatiles Emitted from Artificially Damaged Conspecific Plants. *Plant Cell Physiol.* 52, 588–596
- 114 Ueda, H. *et al.* (2012) Plant communication: Mediated by individual or blended VOCs? *Plant Signal. Behav.* 7, 222–226
- 115 Dabiri, M. *et al.* (2020) Partial sequence isolation of DXS and AOS genes and gene expression analysis of terpenoids and pyrethrin biosynthetic pathway of *Chrysanthemum cinerariaefolium* under abiotic elicitation. *Acta Physiol Plant* 42, 30
- 116 Li, J. *et al.* (2019) Defense of pyrethrum flowers: repelling herbivores and recruiting carnivores by producing aphid alarm pheromone. *New Phytol.* 223, 1607–1620
- 117 Sakamori, K. *et al.* (2016) Selective regulation of pyrethrin biosynthesis by the specific blend of wound induced volatiles in *Tanacetum cinerariifolium*. *Plant Signal. Behav.* 11, e1149675
- 118 Dombrowski, J.E. *et al.* (2019) Transcriptome analysis of the model grass *Lolium temulentum* exposed to green leaf volatiles. *BMC Plant Biol.* 19, 222

CHAPTER 2 – METABOLIC ENGINEERING OF THE PYRETHRIN PATHWAY IN PLANTS

Portions of this chapter were adapted from the following previously published manuscripts:

Xu, H., Lybrand, D., Bennewitz, S., Tissier, A., Last, R.L., and Pichersky, E. (2018) Productions of *trans*-chrysanthemoid acid, the monoterpene acid moiety of natural pyrethrin insecticides, in tomato fruit. *Metab. Eng.* 47, 271-278. DOI: 10.1016/j.ymben.2018.04.004

Li, W., Lybrand, D.B., Xu, H., Zhou, F., Last, R.L., and Pichersky, E. (2019) A trichome-specific, plastid-localized *Tanacetum cinerariifolium* Nudix protein hydrolyzes the natural pyrethrin pesticide biosynthetic intermediate *trans*-chrysanthemyl diphosphate. *Front. Plant. Sci.* 11. DOI: 10.3389/fpls.2020.00482

METABOLIC ENGINEERING IN PLANTS

Specialized metabolic pathways have been the target of many metabolic engineering projects in plants and microbes. While humans find many plant specialized metabolites useful, the plants that produce them may accumulate the compounds of interest at low levels, making the products expensive, or may be threatened, creating a conflict between natural product use and conservation efforts. Furthermore, whether the product of interest is extracted from wild or cultivated plants, the supply of the compound may be disrupted by seasonal variation and weather patterns. Alternative platforms for plant specialized metabolite production can lower the cost of high-value compounds, increase their total production capacity, and stabilize the supply of compounds that could otherwise be disrupted by environmental or economic factors. For example, levels of the antineoplastic taxol in *Taxus brevifolia* are extremely low (0.01% and 0.006% dry weight in bark and needles, respectively) and the trees are slow-growing and relatively rare [1]. Metabolic engineering has provided alternative sources for taxol and its precursors that can achieve higher yields without the need for harvesting *Taxus* plants, including genetically engineered bacterial strains, artificial fungal consortia, and transiently transfected plants [2–4]. Similar strategies were also employed for compounds with relatively high abundance in the native producer: while both artemisinin and pyrethrins accumulate to ~2% dry weight in elite breeding lines of *Artemisia annua* and *T. cinerariifolium*, respectively, products from both pathways were engineered in alternative systems [5,6]. Artemisinic acid, a precursor of artemisinin, has been produced in liquid cultures of *Saccharomyces cerevisiae* and transiently transfected *Nicotiana benthamiana* plants while manipulation of liquid *T. cinerariifolium* cell cultures by phytohormones led to increased pyrethrin production [7–9].

These examples demonstrate the viability of engineering alternative platforms for plant specialized metabolite production and illustrate the utility of metabolic engineering for ensuring a robust and stable supply of useful natural products.

While metabolic engineering can facilitate production of plant specialized metabolites that are intrinsically useful to humans, genetic engineering can also be applied to bolster endogenous plant defenses. The most commercially successful example of this is expression of BT protein toxins originally derived from *Bacillus thuringiensis* in crop plants such as maize and cotton (reviewed in [10]). Increased knowledge of plant metabolism and improved genetic engineering technologies allowed more complex engineering of plant chemical defenses. Recent modifications of plant specialized metabolism through genetic engineering yielded increased plant resistance to insects, microbes, viruses, and abiotic stress. Elucidation of the betalain pathway facilitated production of these red and yellow pigments, typically restricted to members of the Caryophyllaceae, in various ornamental and crop species of the Solanaceae, providing them with resistance to the fungal pathogen *Botrytis cinerea* [11]. Engineered production of astaxanthin, typically absent in angiosperms, in *Arabidopsis* plants improved both their resistance to the bacterial pathogen *Pseudomonas syringae* and their oxidative stress tolerance [12]. Modification of the flavonoid profile of *Glycine max* via CRISPR-mediated gene editing improved resistance of the crop to soya bean mosaic virus without altering total flavonoid levels or introducing new compounds [13]. Thus, engineering of specialized metabolic pathways in plant hosts is both useful for production of valuable compounds and an effective strategy for improving crop resistance to diverse stressors.

Successful examples of plant specialized metabolic engineering for production of valuable compounds and endogenous plant chemical defenses prompted us to pursue engineering of the pyrethrin pathway through stable transformation of a plant host. At the outset of this project, only two steps in the *T. cinerariifolium* pyrethrin biosynthetic pathway were known: chrysanthemyl diphosphate synthase (TcCDS), catalyzing the initial condensation of two units of dimethylallyl diphosphate (DMADP) to yield chrysanthemyl diphosphate (CDP), and a GDSL lipase-like protein (TcGLIP) that catalyzes the final esterification of the monoterpenoid and rethrolone subunits to yield mature pyrethrins [14,15]. While TcGLIP requires substrates that are unique to the pyrethrin pathway, the TcCDS substrate, DMADP, is produced by the mevalonate and DXP pathways found in all green plants [14–16]. We therefore sought to use the *TcCDS* gene from *T. cinerariifolium* to reconstruct the monoterpenoid portion of the pyrethrin pathway in heterologous plant hosts. Previous reports indicate that transient expression of TcCDS in *Nicotiana benthamiana* leaves yielded modest levels of volatile chrysanthemol and lavandulol and oxidized products such as chrysanthemic acid also accumulate [17,18]. We reasoned that stable transformation of a plant host with the *TcCDS* gene could facilitate production of chrysanthemol and other monoterpenoids for extraction and downstream applications or production of CDP-derived metabolites with protective effects to enhance endogenous plant defenses.

We chose the fruit of cultivated tomato (*Solanum lycopersicum*) as a platform requiring minimal alteration of metabolism for our efforts to achieve high-level production of CDP derivatives. Tomato fruits accumulate substantial amounts of terpenoids produced from DMADP and isopentenyl diphosphate (IDP), particularly the carotenoid lycopene which

consistently exceeds 100 mg/kg in fruits of processing tomatoes under field conditions [19,20]. Both the carotenoid pathway that acts in tomato fruit and the TcCDS enzyme in *T. cinerariifolium* are plastidially localized, using IDP and/or DMADP from the plastidial DXP pathway [17,21], and previous efforts to reroute flux from the carotenoid pathway to monoterpenoid biosynthesis by *S. lycopersicum* fruit-specific expression of monoterpenoid biosynthetic genes yielded production of volatile monoterpenoids with a concomitant decrease in carotenoid levels [22,23]. Thus, endogenous tomato fruit metabolism should provide ample DMADP substrate for the TcCDS enzyme. As an added benefit, restricting metabolic modifications to the fruit should prevent growth defects that occur when carotenoid biosynthesis is modified throughout the plant [24]. These features made tomato fruit an ideal tissue for our metabolic engineering experiments.

We began our engineering experiments by expressing the *TcCDS* gene in *S. lycopersicum* under the fruit-specific polygalacturonase (PG) promoter [25]. We found that two dehydrogenases from *Solanum habrochaites*, an alcohol dehydrogenase (ShADH) and an aldehyde dehydrogenase (ShALDH), were capable of oxidizing chrysanthemol, a derivative of CDP, yielding chrysanthemic acid *in vitro*. We further identified a member of the Nudix (nucleoside diphosphate linked to other moieties X) hydrolase family (*TcNudix1*) that shows coexpression with *TcCDS* in the transcriptome of *T. cinerariifolium* [18] and confirmed that the TcNudix1 enzyme dephosphorylates CDP *in vitro*. We crossed tomato plants expressing *TcCDS* under control of the PG promoter with tomato lines expressing *ShADH* and *ShALDH* from *S. habrochaites* as well as *TcNudix1* from *T. cinerariifolium*, also under control of the PG promoter. We measured volatile and non-volatile compounds in fruit extracts of all transgenic lines and

crosses by GC-MS and LC-MS to quantify the effects of *TcCDS* expression and multiple transgene expression on accumulation of CDP derivatives and carotenoids.

We also chose cultivated tomato to test TcCDS-mediated augmentation of endogenous plant chemical defenses. Plant-wide ectopic expression of the *TcCDS* gene in *Chrysanthemum morifolium* resulted in accumulation of several CDP-derived volatile and non-volatile compounds in plant tissue and conferred improved plant resistance to cotton aphids, apparently without major alteration of plant growth phenotype [26]. However, just as we chose tomato fruit as the tissue for high-level CDP production due to its high endogenous accumulation of plastidially derived terpenoids, we decided that expression of *TcCDS* to enhance endogenous plant chemical defenses should also be localized in a tissue that already supports high flux through the DXP pathway. The type VI secretory glandular trichomes of *Solanum* species, including the cultivated *S. lycopersicum* and its wild relative *Solanum habrochaites*, produce abundant terpenoids from the DXP pathway [27,28]. These glandular trichomes and the compounds they produce are a key defense against insects in *Solanum* spp. [29]. To determine the impact of TcCDS activity on the terpenoid composition of type VI trichomes, we expressed the *TcCDS* gene in *S. lycopersicum* using a variety of type VI trichome-specific promoters. We used the *S. lycopersicum* metallocarboxypeptidase inhibitor (MCPI) promoter which drives strong gene expression in type VI trichomes of *S. lycopersicum* [30]. Transgenic lines showing robust production of CDP derivatives were crossed with transgenic lines in which the *Nudix1* gene from *T. cinerariifolium* was expressed under control of the MCPI promoter. We evaluated the effects of *TcCDS* expression on trichome-localized metabolites either alone or in combination with *TcNudix1* using both GC-MS and LC-MS.

Heterologous expression of the *TcCDS* gene in both fruits and type VI trichomes of *S. lycopersicum* yielded accumulation of CDP and a variety of volatile and non-volatile CDP derivatives. Fruits of transgenic lines expressing *TcCDS* alone accumulated primarily volatile chrysanthemol and an abundant chrysanthemol glycoside with low levels of oxidized chrysanthemol, chrysanthemic acid, and chrysanthemyl glycoside. Addition of either the *ShADH* or *ShALDH* transgene increased the relative abundance of oxidized CDP derivatives while a combination of the *TcCDS*, *ShADH*, and *ShALDH* transgenes led to the highest overall abundance of CDP derivatives including an eight-fold increase of free and glycosylated chrysanthemic acid relative to lines expressing the *TcCDS* transgene alone. All lines expressing *TcCDS* with *ShADH*, *ShALDH*, or both dehydrogenases showed a reduction in total carotenoid content. This could be due to diversion of flux from tetraterpenoids to monoterpenoids. However, co-expression of *TcCDS* and the *TcNudix1* gene returned carotenoids to wild-type levels, suggesting that reduced carotenoid levels are the result of inhibition of geranyl-geranyl diphosphate synthase (GGPPS) by CDP that is relieved when CDP is rapidly dephosphorylated. Transgenic lines expressing *TcCDS* in type VI trichomes accumulated a similar complement of volatile and non-volatile CDP derivatives to that seen in transgenic fruits. Unlike transgenic fruits, production of CDP derivatives did not impact levels of native plastidially derived terpenoids (in this case, volatile monoterpenes). Co-expression of *TcCDS* and *TcNudix1* increased levels of both volatile and non-volatile CDP derivatives. These results indicate the feasibility of engineering high-level production of monoterpenoid intermediates from the pyrethrin pathway in a heterologous plant host as well as the possibility of using genes from the pyrethrin pathway to introduce new trichome-localized chemical defenses into tomato (Fig. 2.1).

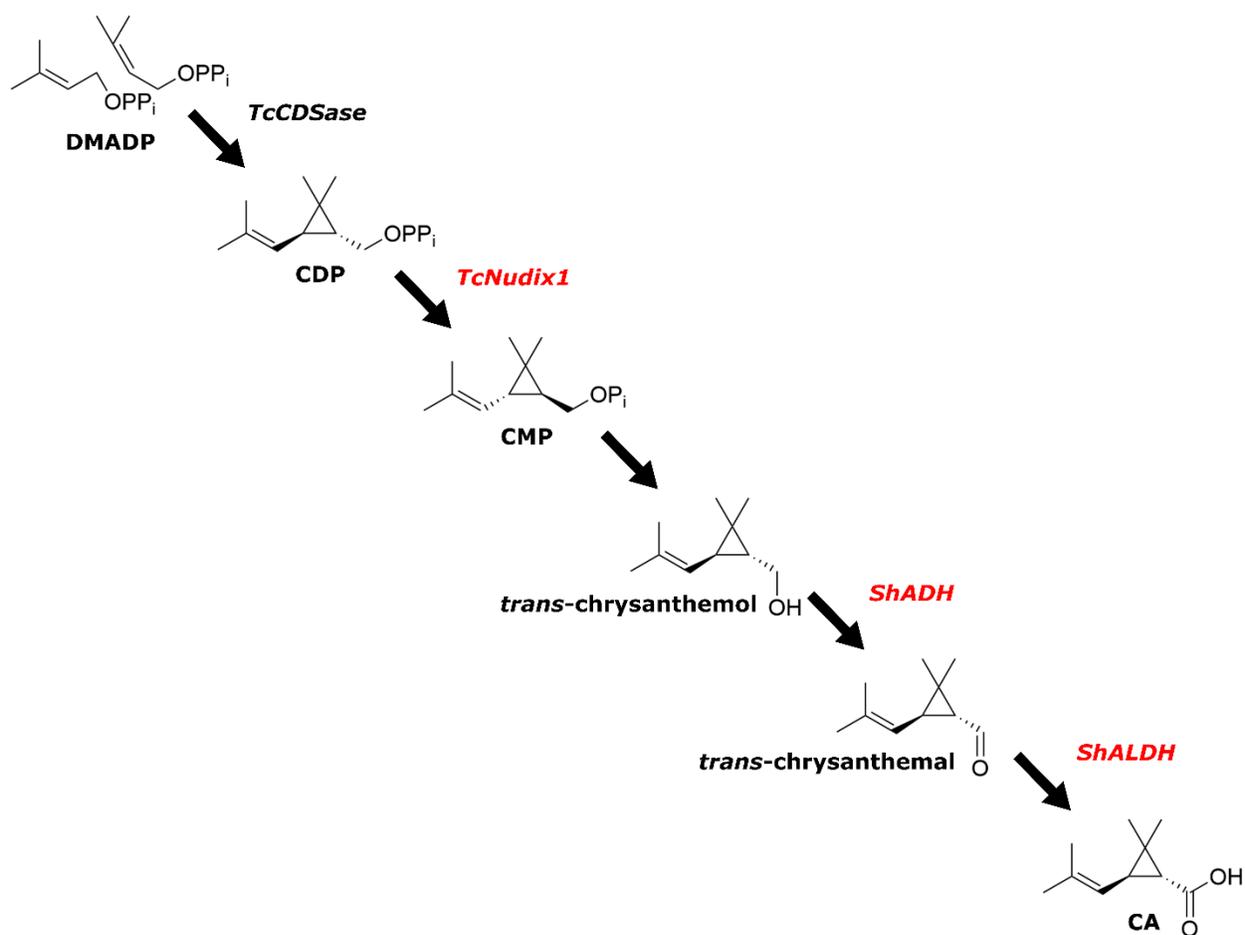


Figure 2.1 The engineered pathway for production of trans-chrysanthemal acid in cultivated tomato (*Solanum lycopersicum*). Activities catalyzed by products of transgenes shown in red can be replaced by endogenous tomato enzymes. Abbreviations: DMADP – dimethylallyl diphosphate; CPP – chrysanthemyl diphosphate; CMP – chrysanthemyl monophosphate; CA – chrysanthemal acid; TcCDS – chrysanthemyl diphosphate synthase from *Tanacetum cinerariifolium*; TcNudix1 – Nudix1 from *T. cinerariifolium*; ShADH – alcohol dehydrogenase from *Solanum habrochaites*; ShALDH – aldehyde dehydrogenase from *S. habrochaites*.

MATERIALS AND METHODS

Plant growth conditions

Pyrethrum (*Tanacetum cinerariifolium*) plants were grown in growth chambers on soil with a 16/8 h light/dark period. Day and night temperatures were 25°C and 20°C, respectively. Tomato (*Solanum lycopersicum*) plants were grown in growth chambers on soil at constant 22°C with a 16/8 h light/dark regime.

Co-expression analysis

Co-expression analysis of candidate phosphatases was performed relative to the *T. cinerariifolium CDSase* gene. Candidates were selected using the previously published transcriptome from *T. cinerariifolium* [18] and Pearson correlation coefficients were determined using the SPSS statistics software package (<https://www.ibm.com/products/spss-statistics>).

Vector construction and generation of tomato transgenic plants

The complete open reading frame (ORF) of *TcCDS* was synthesized using a previously published sequence [14] and spliced into the binary vector pBIN19 carrying the tomato fruit-specific polygalacturonase (PG) promoter and PG terminator as well as the kanamycin resistance marker gene *NPTII* driven by the *CaMV 35S* promoter [25]. The *TcCDS* ORF was spliced between the PG promoter and the PG terminator using the BamHI and SpeI restriction sites. The encoded TcCDS protein sequence [14] is missing the first 4 amino acids of the transit peptide, as determined by Yang and co-workers [17]. Nevertheless, both versions of the *TcCDS* ORF direct the synthesis of identical mature proteins, each containing a transit peptide that causes the proteins to be imported into the plastids (Appendix Fig. S2.1; subcellular localization

was determined by GFP tagging and confocal microscopy according to Falara and co-workers [31]). The coding region of *ShADH* (ortholog of Solyc03g044200 from *Solanum habrochaites* LA1777) was obtained by PCR on cDNA from trichomes of *S. habrochaites* LA1777 using primers SBE634_ADH_F and SBE635_ADH_R (Table S2.1 in Appendix). The product was cloned into vector pICH41308 of the Golden Gate modular cloning system [32]. The resulting plasmid pAGT1217 was then used to construct recombinant binary vectors. Similarly, the coding region of *ShALDH* (ortholog of Solyc06g060250 from *S. habrochaites* LA1777) was amplified from trichome cDNA of *S. habrochaites* LA1777 using primers SBE537_ALDH_F and SBE538_ALDH_R (Table S2.1 in Appendix) and cloned into pICH41308. The resulting plasmid, pAGT918, was then used to construct binary vectors for tomato transformation. The coding regions of *ADH* and *ALDH* were amplified from plasmids pAGT1217 and pAGT918 respectively and introduced into the binary vector pBIN19 carrying PG promoter and PG terminator at *Age*I and *Spe*I restriction sites. The ORF of *TcNudix1* was amplified from cDNA of *T. cinerariifolium* using primers pBIN19-Nudix1-F and pBIN19-Nudix1-R and integrated directly into plasmid pBIN19 with the fruit-specific PG promoter.

Type VI trichome expression constructs were assembled using the metallopeptidase inhibitor (MCPI) promoter. The *MPCI* promoter was amplified using the MCPI-pBIN19-F and MCPI2nd-overlap-(Nudix)-R primers (see Table S2.1 in Appendix) and both the MCPI promoter and either the *TcCDS* ORF or the *TcNudix1* ORF were assembled into the pBIN19 backbone.

For generation of transgenic lines, a binary vector carrying a single gene (*TcCDS*, *ShADH*, *ShALDH*, or *TcNudix1*) was introduced into *S. lycopersicum* cultivar MP1 plants by the University

of Nebraska Plant Transformation Facility (<http://biotech.unl.edu/plant-transformation>). Transgenic plants rooted on kanamycin selection were transferred to soil and grown in a greenhouse with 14/10 h day/night photoperiod at 22°C. Positive transgenic tomato plants were further verified by genomic DNA PCR for the presence of corresponding genes.

Crossing tomato plants was performed according to the method of Kimura and Sinha [33]. The seeds from crossed flowers germinated on kanamycin selection were transferred to soil and grown in the same conditions as described above. Positive transgenic tomato plants were further verified by genomic DNA PCR for the presence of corresponding genes.

RNA isolation and qRT-PCR analysis

Total RNA was isolated from ripening pericarps of tomato fruits using the Total RNA Isolation Kit from Omega Biotek (Norcross, GA) containing a DNA digestion step using the manufacturer's protocol. cDNAs were prepared using the High Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific, Waltham, MA) following the manufacturer's instructions. Real-time PCR was performed using the Stepone Real-time PCR system (Applied Biosystems, Foster City, CA). Assays were done in six independent biological replicates (separate plants growing in the same growth facility). The relative transcript levels for different genes were normalized to that of *elongation factor 2 (EF-2)*.

Subcellular localization

The open reading frame of *TcNudix1* was integrated to plasmid pEZS-NL to create an open reading frame (PRF) fused to GFP. The construct was transformed into Arabidopsis protoplast cells for confocal microscopy examination. This was done using previously described methods [31].

Enzymatic assays of recombinant TcCDS, ShADH, ShALDH, and TcNudix1

The truncated open reading frame of *TcCDS* missing the first 50 codons was synthesized and cloned into the pEXP5-CT/TOPO vector, generating a fusion gene that encodes a tag of six His residues at the C-terminus. The complete ORFs of *ShADH* and *ShALDH* were amplified from plasmids pAGT1217 and pAGT918, respectively, and introduced into the expression vector pET28a+ or pHIS8, in each case generating a fusion gene construct encoding a “tag” of HIS6 residues at the N-terminus. The ORF of *TcNudix1* without the N-terminal 56 amino acid transit peptide and stop codon was integrated into the vector pET28a. Constructs were then shuttled into *E. coli* BL21 (pLySs). The expression and purification of the recombinant proteins were performed as described previously [34]. Briefly, bacteria harboring an expression vector were cultured in 500 ml LB to OD₆₀₀ = 0.6 at 37°C, and protein expression was induced by 0.5 mM IPTG and 16°C shaking. After overnight incubation, the induced bacterial culture was centrifuged, and bacteria were resuspended in wash buffer (50 mM Tris, pH8.0, 300 mM NaCl, 20 mM imidazole and 10 mM 2-mercaptoethanol). Bacteria cells were broken up by ultrasonication (3 seconds sonication with 6 seconds break, 200 times). Qiagen Ni-NTA agarose (0.5 ml) was added to each lysate and incubated for 1 h to let protein bind, then the mixture was poured into Qiagen 1ml polypropylene columns and the agarose was washed with 50 ml wash buffer. Finally, purified protein was eluted with elution buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 250 mM imidazole and 10 mM 2-mercaptoethanol).

For enzymatic assay of *TcCDS*, 30 µg affinity-purified His-tagged enzyme was incubated at 30°C for 3 h with 0.4 mM DMADP, in a final volume of 50 µl of assay buffer containing 50 mM Tris-HCl (pH 7.5), 2 mM DTT, 5 mM MgCl₂. The produced CDP was hydrolyzed by 5 units of

Roche rAPid alkaline phosphatase (Sigma, St. Louis, MO) at 37 °C for 1 h following the manufacturer's instructions or 0.2N HCl at room temperature for 30 min. The reaction products were extracted by 100 µl MTBE and analyzed by GC-MS.

For enzymatic assay of ShADH, 15 µl eluted protein from corresponding recombinant vector was incubated at 30°C for overnight with 1 mM mix of *trans*- and *cis*-chrysanthemol, in a final volume of 50 µl of assay buffer containing 50 mM Tris-HCl (pH 7.5), 2 mM DTT, 1 mM NAD⁺.

For enzymatic assay of ShALDH, a coupled assay was performed by 30°C overnight incubation of reaction mix composed of 15 µl eluted protein of ShADH and 15 µl eluted protein of ShALDH from corresponding recombinant vectors with 1 mM mix of *trans*- and *cis*-chrysanthemol, in a final volume of 50 µl containing 50 mM Tris-HCl (pH 7.5), 2 mM DTT, 1 mM NAD⁺. Reaction products were extracted with 100 µl MTBE and analyzed by GC-MS.

For enzymatic assay of TcNudix1, purified protein was incubated with CPP or other prenyl diphosphate. The 50 µl reactions contained: 100 mM Tris (pH 7.5), 5 mM MgCl₂, 1 µM TcNudix1 protein and various concentration of prenyl diphosphate (0.2-10 µM). The reaction solution was extracted directly for LC-MS or incubated overnight with crude protein extracts from flowers or leaves for GC-MS analysis. When measuring kinetic parameters, purified TcNudix1 was incubated with CPP from 0.1 to 5 µM for 5 min reaction and the decrease in the CPP peak as measured by LC-MS was calculated. The colorimetric assay for monophosphatase activity of TcNudix1 with different prenyl diphosphate substrates was performed by mixing 7 micrograms of purified TcNudix1 with 10 µM each substrate in 50 µl reaction containing 100 mM Tris (pH 7.5) and 5 mM MgCl₂. Each reaction was incubated for 30 min at room

temperature, then 100 μ l BioMol Green Reagent (Enzo[®] Life Science, Ann Arbor, MI) was added and incubated for 30 min for the color to develop. The yellow malachite green molybdate in this reagent binds to free orthophosphate and forms a green complex that absorbs at 620-640 nm.

Extraction and analysis of trans-chrysanthemic acid and related compounds from tomato fruits

The tomato fruit achieves full size while still green, and the point at which the full-size green fruit shows the first sign of red color is called the breaker stage [35]. Pericarps of tomato fruits beyond the breaker stage (as specified) were collected and ground into fine powder in liquid nitrogen. For measurement of volatile compounds, 6 g of fruit homogenate was extracted at room temperature with shaking at 50 rpm overnight with 6.0 ml *tert*-butyl methyl ether (MTBE) containing 0.002 ng/ μ l tetradecane as internal standard. The MTBE layer was transferred to a new tube, followed with dehydration by anhydrous Na₂SO₄, and then concentrated by evaporating the solvent under vacuum for 10-15 min to a final volume of about 0.5 ml. Samples were then analyzed by GC-MS. For analysis of non-volatile compounds, fruit tissue was powdered in liquid nitrogen. To 0.1 g of powdered tissue was added 0.5 ml of 80% acetonitrile, 20% water with 0.1% formic acid and 10 μ M propyl-4-hydroxybenzoate (internal standard). Extracts were centrifuged for 10 min at 10,000 *g* and supernatant transferred to LC vials. Extract supernatant was analyzed by LC-MS.

For analysis of trichome-localized metabolites, single leaflets of tomato plants were harvested and transferred to 1.7-mL polypropylene microfuge tubes. To each leaflet, 1 mL of extraction solvent was added, and the tube was sealed and rocked gently for 1 min. Supernatant was then transferred to autosampler vials for downstream analysis by GC-MS or

LC-MS. For volatile compounds, the extraction solvent used was MTBE containing 0.002 ng/ μ l tetradecane as internal standard. For non-volatile compounds, a 3:3:2 mixture of acetonitrile/isopropanol/H₂O with 10 μ M propyl-4-hydroxybenzoate as internal standard was used.

GC-MS analysis and LC-MS analysis

For GC-MS analysis, a 1 μ l aliquot of sample was injected into a Shimadzu QP-2010 GC-MS system equipped with the Rxi-5Sil column (30 m \times 0.25 mm \times 0.25 μ m film thickness, Restek, USA). Helium (1.4 ml/min) was used as a carrier gas with split mode at a ratio of 1:2. The injection temperature was set at 240°C, and the interface temperature was 280°C. The oven temperature program was as follows: initial temperature, 50°C for 3 min, followed by a ramp from 50 to 110°C at a rate of 10°C min⁻¹ and then from 110°C to 150 °C at a rate of 5°C min⁻¹, held for 3 min, and finally increased to 300°C at a rate of 10°C min⁻¹, held for 3 min. The identification of the volatiles was assigned by comparison of their retention times and mass fragmentations with those of literature and NIST library and by comparison of spectral data with standards.

For LC-MS analysis of transgenic fruit and trichome extracts, samples (10 μ L injection volume) were applied to an Ascentis Express C18 column (100 \times 2.1 mm, 2.7 μ m particle size; Supelco, Bellefonte, PA) at 40°C. Liquid chromatography was carried out using an Acquity UPLC system (Waters). Solvents used for liquid chromatography were water + 0.1% formic acid (solvent A) and acetonitrile (solvent B). A 0.3 ml/min flow rate was used. The LC gradient started at 99% solvent A, 1% solvent B and increased to 99% solvent B over 16 minutes following a linear gradient. Solvent B was held at 99% for 2 minutes followed by a return to 99%

solvent A for 2 minutes. Mass spectrometric analysis was carried out with a Xevo G2-XS Q-ToF instrument (Waters). Electrospray ionization was employed in negative-ion mode using 2.0 kV capillary voltage, 40V sample cone voltage, and 100°C source temperature. Desolvation gas flow of 600 L/hr and temperature of 350°C were used.

Enzymatic reactions of TcNudix1 were analyzed using a Micromass Quattro Premier Mass Spectrometer LC-MS/MS System (Waters Corporation, Milford, MA) with an Ascentis Express C18 column (100 x 2.1 mm, 2.7 µm particle size; Supelco, Bellefonte, PA). The gradient held at 99% mobile phase A (0.05% triethylamine) and 1% mobile phase B (50% acetonitrile, 50% isopropanol and 0.05% triethylamine) for 0.5 min, then increased 99% B in a linear gradient from 0.5 min to 3 min, and held at 99% B for 1 min. Peaks were detected in MRM mode (multiple reaction monitoring). The precursor ion for CPP and GPP is 313.1, for CMP and GMP is 233.1, for FPP is 381.1, for FMP is 301.1, for GGPP is 449.2, for GGMP is 369.2, product ion for all compounds is 78.8.

Lycopene measurements

Measurements of lycopene were performed as previously described (Gutensohn et al., 2014). Fruit tissue was homogenized and extracted with acetone/hexane (4/6), absorbance at 663, 645, 505 and 453 were measured by spectrophotometer. Content of lycopene = $-0.0458 \times A_{663} + 0.204 \times A_{645} + 0.372 \times A_{505} - 0.0806 \times A_{453}$ (mg/100 ml).

RESULTS

Identification of TcCDS transgenic tomato lines producing trans-chrysanthemic acid

In contrast to the non-transgenic parent tomato line, 2.3 to 7.0 $\mu\text{g g}^{-1}$ fresh weight of *trans*-chrysanthemic acid was detected in ripened fruits of 14 independent *TcCDS* transgenic lines (Fig. 2.2). The transgenic line designated as TcCDS32 had the highest concentration of *trans*-chrysanthemic acid (7.0 $\mu\text{g g}^{-1}$ fresh weight) and this line was used for analysis of precursors of *trans*-chrysanthemic acid. We observed that mature fruits of TcCDS32 contained 1.6 $\mu\text{g g}^{-1}$ fresh weight of *trans*-chrysanthemol and 0.35 $\mu\text{g g}^{-1}$ fresh weight of *trans*-chrysanthemal. The concentration of CDP, analyzed by acid hydrolysis and measuring the resulting monoterpenes yomogi alcohol and artemisia alcohol (Appendix Fig. S2.2), was determined to be 3.7 $\mu\text{g g}^{-1}$ fresh weight.

Reconstruction of the complete pathway to trans-chrysanthemic acid in tomato fruit

It was previously shown that *TcCDS* expression in a heterologous plant system also led to appreciable *trans*-chrysanthemol production [17,18]. *trans*-Chrysanthemol production could be the direct outcome of CDS catalysis [17], or the action of endogenous non-specific phosphatases might be responsible [18]. In *T. cinerariifolium*, *trans*-chrysanthemol is enzymatically oxidized to *trans*-chrysanthemal and then to *trans*-chrysanthemic acid by specific dehydrogenases [18], but such oxidation reactions also occur in the leaves of *N. benthamiana* expressing CDS [18] and in tomato fruits (Fig. 2.2). However, the presence of CDP, *trans*-chrysanthemol, and *trans*-chrysanthemal in addition to *trans*-chrysanthemic acid in *TcCDS* transgenic tomato fruits suggested the hypothesis that the *trans*-chrysanthemic acid yield could be increased if the rate of such oxidation reactions were enhanced.

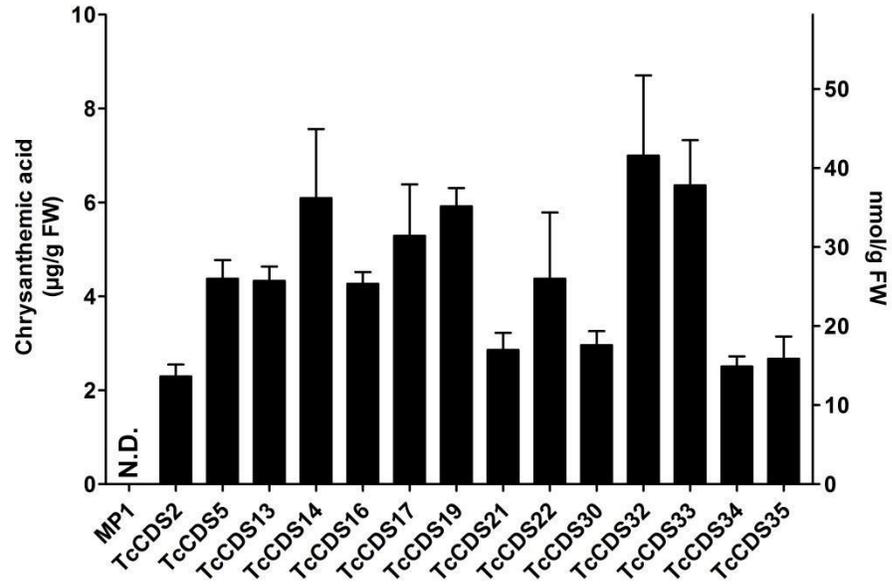


Figure 2.2 Concentration of *trans*-chrysanthemic acid in ripened fruits (at the 15th day after breaker, Br+15) of 14 transgenic tomato lines expressing *TcCDS* under the control of the PG promoter. Values of chrysanthemic acid are shown in micrograms per gram of fresh weight (left-hand Y-axis) and in nanomoles per gram of fresh weight (right-hand Y-axis) (means \pm SD, $n = 6$). N.D., Not detected.

To test this theory, we produced transgenic tomato plants that express a *Solanum habrochaites* gene encoding a sesquiterpene alcohol dehydrogenase (*ShADH*, the ortholog of Solyc03g044200 from *S. habrochaites* LA1777) under the control of the PG promoter. We also generated transgenic tomato plants that express a *Solanum habrochaites* gene encoding a sesquiterpene aldehyde dehydrogenase (*ShALDH*, ortholog of Solyc06g060250 from *S. habrochaites* LA1777) also under the control of the PG promoter. In *S. habrochaites* LA1777, *ShADH* and *ShALDH* catalyze two successive steps in the oxidation of santalenol and bergamotenol to the respective carboxylic acids, santalenoic and bergamotenoic acids, which are the major products of type VI glandular trichomes in this accession (Dr. Alain Tissier, personal communication). *In vitro* enzymatic assays of *ShADH* and *ShALDH* showed that these two enzymes can sequentially oxidize *trans*-chrysanthemol to *trans*-chrysanthemic acid in the presence of NAD⁺ (Fig. 2.3). Transcript levels of four independent *ShADH* transgenic tomato lines and six positive *ShALDH* transgenic tomato lines were examined in ripening fruits (Appendix Fig. S2.3). *ShADH* line 4 and *ShALDH* line 13 had the highest expression level among *ShADH* and *ShALDH* lines transgenic lines, respectively. Therefore, these lines were used in sequential crosses with the *TcCDS32* transgenic to yield progeny plants that harbor *TcCDS* and *ShADH*, *TcCDS* and *ShALDH*, as well as plants that contain all three genes. qRT-PCR analysis of transcript levels in the ripened fruits of these lines showed that in the line expressing all three genes, each of these genes was expressed, although at somewhat lower levels than observed in lines containing each gene alone or lines containing only two

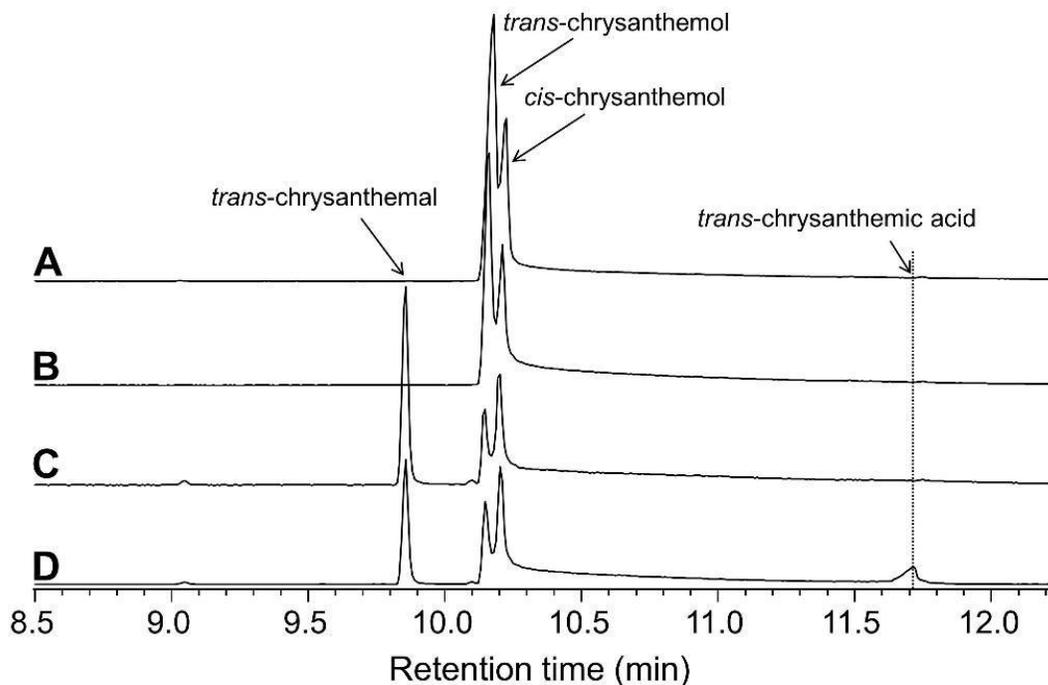


Figure 2.3 GC-MS analysis of MTBE extracts of *in vitro* enzymatic assay of ShADH and ShALDH. All reactions were incubated overnight. **(A)** Products obtained after incubating 15 μ l of eluted protein from an empty-vector prep with a 1 mM mix of *trans*- and *cis*-chrysanthemol in the presence of 1 mM NAD^+ . **(B)** Products obtained after incubating 15 μ l of eluted ShALDH protein with a 1 mM mix of *trans*- and *cis*-chrysanthemol in the presence of 1 mM NAD^+ . **(C)** Products obtained after incubating 15 μ l of eluted ShADH protein with a 1 mM mix of *trans*- and *cis*-chrysanthemol in the presence of 1 mM NAD^+ . **(D)** Products obtained after incubating 15 μ l of eluted ShADH protein and 15 μ l of eluted ShALDH protein with a 1mM mix of *trans*- and *cis*-chrysanthemol in the presence of 1 mM NAD^+ .

heterologous genes (Fig. 2.4). The reduction in transcript level was particularly noticeable for *TcCDS*, with a transcript level in the line harboring all three genes that was only 11% of that in line TcCDS32. We hypothesize that this reduction was due in part to the fact that all three heterologous genes contained the same promoter, PG, and thus the three promoters may compete for the same transcription factors.

Production of trans-chrysanthemol and related compounds in transgenic tomato fruits expressing CDS, ADH and ALDH

While tomato fruits expressing *TcCDS* alone accumulated some *trans*-chrysanthemol and *trans*-chrysanthemol, the concentration of the desired product, *trans*-chrysanthemol, was expected to increase in transgenic tomato fruit expressing *ShADH* and *ShALDH* in addition to *TcCDS*. We therefore measured the concentration of CDP, *trans*-chrysanthemol and related compounds in ripened fruits of different transgenic tomato plants expressing *TcCDS*, *ShADH* and *ShALDH* and combinations thereof. The concentration of CDP, which was 3.7 $\mu\text{g g}^{-1}$ fresh weight in the *TcCDS*-expressing line TcCDS32, decreased to 2.5, 2.9, and 2.2 $\mu\text{g g}^{-1}$ fresh weight respectively in lines TcCDS32 \times *ShADH4*, TcCDS32 \times *ShALDH13* and TcCDS32 \times *ShADH4* \times *ShALDH13* (Fig. 2.5A). Similarly, the concentration of free *trans*-chrysanthemol, present at 1.6 $\mu\text{g g}^{-1}$ fresh weight in TcCDS32, decreased to 0.72, 0.98 and 0.58 $\mu\text{g g}^{-1}$ fresh weight respectively in lines TcCDS32 \times *ShADH4*, TcCDS32 \times *ShALDH13* and TcCDS32 \times *ShADH4* \times *ShALDH13* (Fig. 2.5B). Since previous reports indicated that heterologously produced *trans*-chrysanthemol may be glycosylated [17,18], we treated fruit extracts with β -glucosidase and measured the levels of total free *trans*-chrysanthemol, calculating the amount of glycosylated *trans*-chrysanthemol as the difference between the total and free *trans*-chrysanthemol

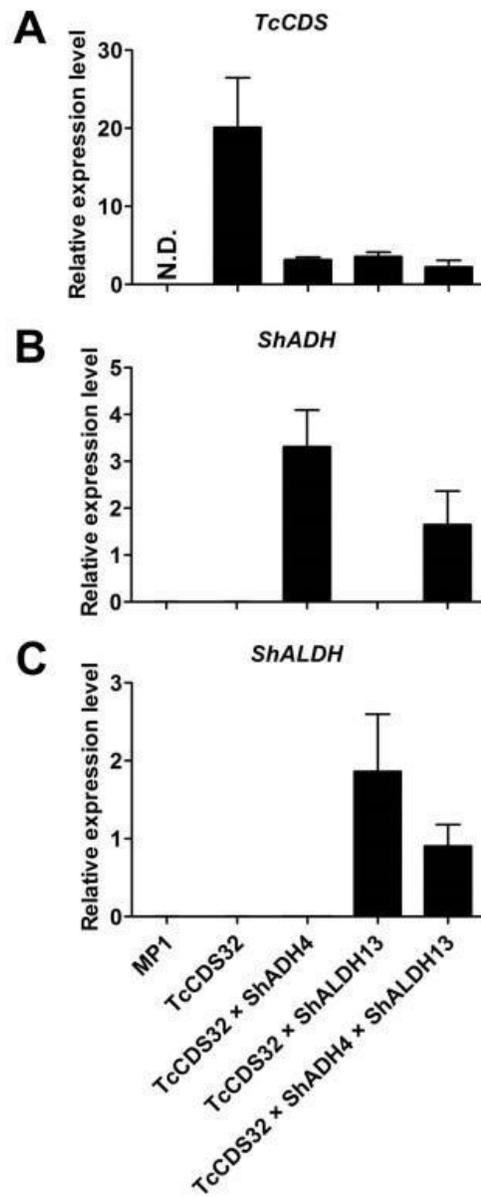


Figure 2.4 qRT-PCR analysis of transcript levels in ripening fruits (at the 3rd day after breaker, Br+3) of non-transgenic tomato and different transgenic lines. **(A)** Levels of TcCDS transcript. **(B)** Levels of ShADH transcript. **(C)** Levels of ShALDH transcript. Results are expressed relative to that of elongation factor 2 (EF2) (means \pm SD, $n = 6$). N.D., Not detected.

levels. Our results indicate that most of the *trans*-chrysanthemol produced in all transgenic lines was found to be glycosylated, with the concentration of total *trans*-chrysanthemol measured at 42.2, 9.6, 10.4 and 6.0 $\mu\text{g g}^{-1}$ fresh weight respectively in lines TcCDS32, TcCDS32 \times ShADH4, TcCDS32 \times ShALDH13 and TcCDS32 \times ShADH4 \times ShALDH13 (Fig. 2.5B). LC-MS analysis indicated that the glycosides of *trans*-chrysanthemol in tomato fruit were malonylated glucose or related hexoses (Appendix Fig. S2.4E,G), similar to those observed in plants of two tobacco species, *N. tabacum* and *N. benthamiana*, expressing *TcCDS* (Appendix Fig. S2.4B) [17].

The *trans*-chrysanthemol concentration, 0.35 $\mu\text{g g}^{-1}$ fresh weight in TcCDS32, increased to 0.86 $\mu\text{g g}^{-1}$ fresh weight in the TcCDS32 \times ShADH4 line but was only 0.079 and 0.27 $\mu\text{g g}^{-1}$ fresh weight in TcCDS32 \times ShALDH13 and TcCDS32 \times ShADH4 \times ShALDH13 (Fig. 2.5C).

Measurements of *trans*-chrysanthemol in fruits of the triple transgenic line TcCDS32 \times ShADH4 \times ShALDH13 showed a large increase in the concentration of this acid. The 67.1 $\mu\text{g g}^{-1}$ fresh weight of free *trans*-chrysanthemol in this line represented an 8.6-fold increase over the concentrations found in the TcCDS32 line (7.00 $\mu\text{g g}^{-1}$ fresh weight). Lines TcCDS32 \times ShADH4 and TcCDS32 \times ShALDH13 had concentrations of free *trans*-chrysanthemol representing 4.3-fold and 2.7-fold increases, respectively, compared with those in TcCDS32 fruits. Heterologously produced *trans*-chrysanthemol in *N. benthamiana* leaves was also reported to be mostly in the form of malonylglycosides [18], and analysis of *trans*-chrysanthemol in ripened tomato fruits of transgenic plants showed that a large portion of it is indeed present as *trans*-chrysanthemol malonylglycosides (Appendix Fig. S2.4F,H), similar to those found in transgenic tobacco plants that produce

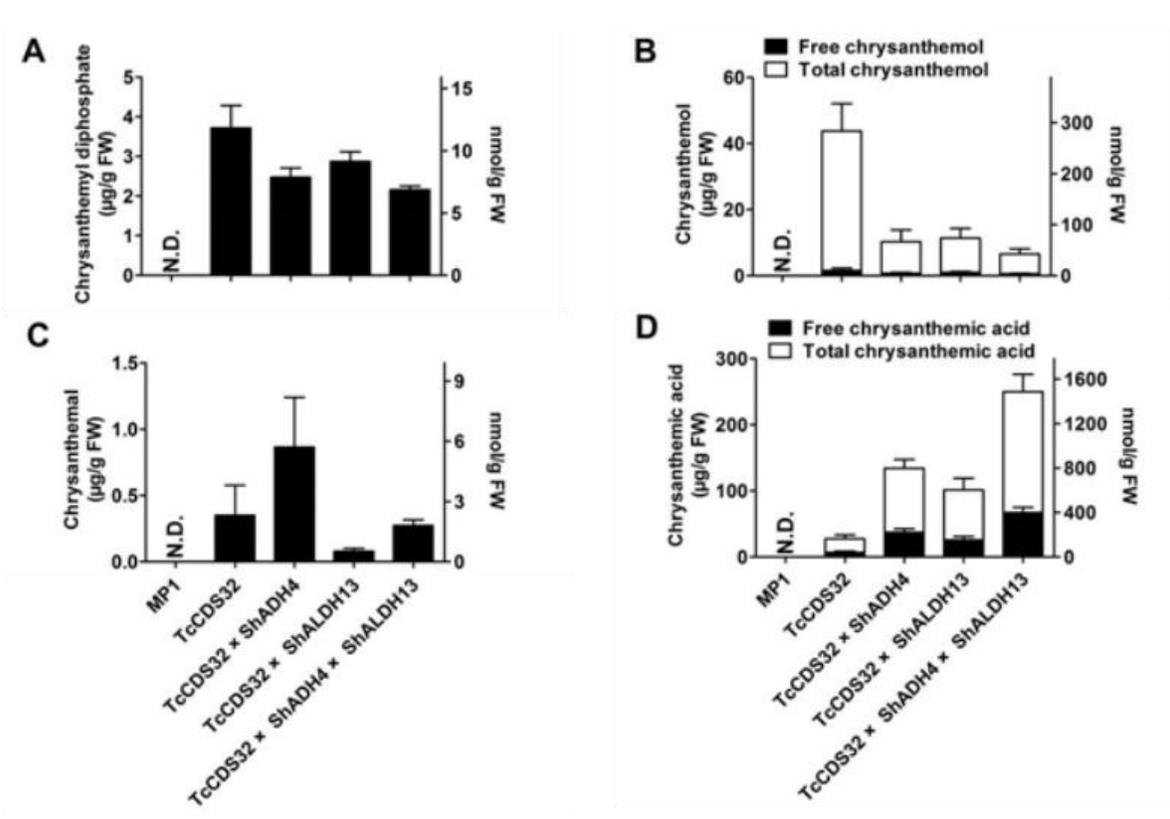


Figure 2.5 Production of *trans*-chrysanthemol and its precursors in ripened transgenic tomato fruits (Br+15) expressing *TcCDS*, *ShADH* and *ShALDH* under the control of PG promoter. **(A)** Concentration of CDP. **(B)** Concentration of *trans*-chrysanthemol. **(C)** Concentration of *trans*-chrysanthemol. **(D)** Concentration of *trans*-chrysanthemol. All the data are means \pm SD ($n \geq 6$). Values are shown in micrograms per gram of fresh weight (left-hand Y-axis) and in nanomoles per gram of fresh weight (right-hand Y-axis). N.D., Not detected.

trans-chrysanthemic acid (Appendix 2.17C) [18]. The concentrations of conjugated *trans*-chrysanthemic acid, calculated by subtracting the total *trans*-chrysanthemic acid detected after base hydrolysis from the measured levels of free *trans*-chrysanthemic acid, were determined to be 13.6, 60.0, 50.3 and 115.8 $\mu\text{g g}^{-1}$ fresh weight in ripe tomato fruits of TcCDS32, TcCDS32 \times ShADH4, TcCDS32 \times ShALDH13 and TcCDS32 \times ShADH4 \times ShALDH13 lines respectively. These levels represent 1.9-, 1.6-, 2.0- and 1.7-fold higher levels compared with the concentrations of free *trans*-chrysanthemic acid in the corresponding transgenic tomato fruits respectively (Fig. 2.5D). Overall, co-expression of *ShADH* and *ShALDH* in fruit of TcCDS32 line enhanced the accumulation of total *trans*-chrysanthemic acid, defined as free plus glycosylated forms, up to 183.0 $\mu\text{g g}^{-1}$ fresh weight – with 62% of it in the glycosylated form – from 20.6 $\mu\text{g g}^{-1}$ fresh weight in tomato fruits expressing TcCDS only, representing an 8-fold increase (Fig. 2.5D).

The use of DMADP by TcCDS in ripening tomato fruits was predicted to compete with the synthesis of the red pigment lycopene, the most abundant terpenoid synthesized during fruit maturation in the tomato cultivar MPI used in this study and in other red fruited varieties. Indeed, ripe tomato fruit expressing *TcCDS* alone were substantially less red than corresponding nontransgenic fruits (Fig. 2.6A), with lycopene accumulation decreasing by 95%, at 5.0 $\mu\text{g g}^{-1}$ fresh weight compared with 104.0 $\mu\text{g g}^{-1}$ fresh weight in non-transgenic fruits (Fig. 2.6B). Surprisingly, lycopene in ripened fruits of TcCDS32 \times ShADH4, TcCDS32 \times ShALDH13 and TcCDS32 \times ShADH4 \times ShALDH13 plants was higher than in TcCDS plants, with 17.6, 15.0, and 33.6 $\mu\text{g g}^{-1}$ fresh weight, respectively. This indicated that the reduction in lycopene content was not directly correlated with diversion of plastidial DMADP to

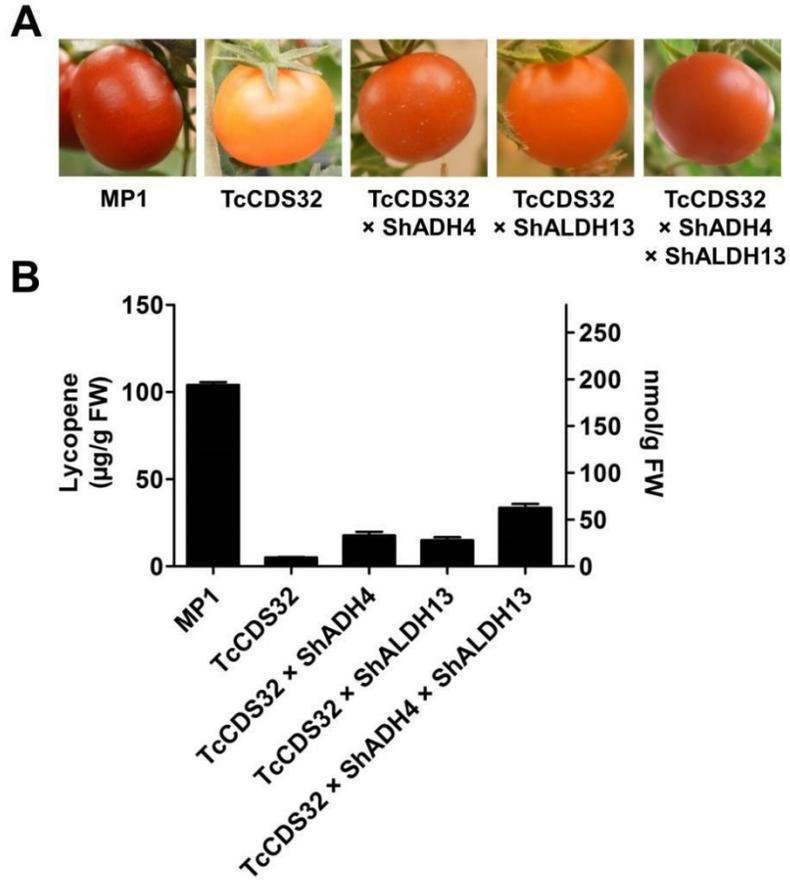


Figure 2.6 Coloration and lycopene contents of ripened fruits (Br+15) of control and transgenic tomato plants. **(A)** Phenotype of *S. lycopersicum* non-transgenic control, TcCDS32, TcCDS32 × ShADH4, TcCDS32 × ShALDH13 and TcCDS32 × ShADH4 × ShALDH13 plants. **(B)** Lycopene concentrations of ripened fruits of non-transgenic control, TcCDS32, TcCDS32 × ShADH4, TcCDS32 × ShALDH13 and TcCDS32 × ShADH4 × ShALDH13 plants. All data points are means \pm SD ($n \geq 6$). Values of lycopene are shown in micrograms per gram of fresh weight (left-hand Y-axis) and in nanomoles per gram of fresh weight (right-hand Y-axis).

production of *trans*-chrysanthemic acid and other monoterpenoid products. Previous experiments in which monoterpenoid biosynthesis was engineered into tomato fruits showed similar reduction in lycopene and in one case, this was shown to be the result of inhibition of geranylgeranyl diphosphate synthase (GGPPS), an upstream enzyme in the lycopene biosynthetic pathway by one of the introduced compounds, neryl diphosphate [22,23]. We hypothesized that reduction in lycopene could be caused in part by accumulation of CDP and consequent inhibition of GGPPS.

Identification of TcNudix1 and subcellular localization of the protein

To test our hypothesis that CDP accumulation might interfere with lycopene biosynthesis, we sought a phosphatase that could hydrolyze CDP and facilitate production of volatiles such as *trans*-chrysanthemol with a smaller steady-state pool of pathway intermediates. We previously constructed multiple RNA-seq databases from pyrethrum floral tissues at various developmental stages [18]. Co-expression analysis using *TcCDS*, which had previously been established to be involved in pyrethrin biosynthesis [36,37], identified a member of the Nudix family within the top 20 contigs whose expression was most highly correlated with *TcCDS* (contig #12 with a correlation coefficient = 0.9878; Appendix Table S2.2). We named this gene *TcNudix1*. While there were at least 56 Nudix hydrolase genes found in our RNA-seq data (albeit not all with full-length gene sequences), none of the other showed a correlation coefficient of >0.95 to reference gene *TcCDS*.

TcNudix1 is a small protein of 229 amino acids with a calculated molecular mass of 25.5 kD. Sequence comparisons of *TcNudix1* with *RhNudix1* from Rose (*Rosa × hybrida*), which hydrolyzes the first phosphate of GPP, and *AtNudix1* from Arabidopsis, which hydrolyzes the

first phosphate of IPP and DMADP, as well as AtNudix5, 6, 11 and 15, for which no prenyl diphosphatase activity was demonstrated [3,38,39], indicate that all of these sequences share two short conserved sequences in the center of the amino acid chains (boxed areas in Fig 2.7). However, TcNudix1 shares a much more extensive sequence identity with RhNudix1 and AtNudix1 - the two proteins shown to hydrolyze prenyl diphosphates (63.5% and 50% identities, respectively) - than with the four other Nudix proteins, which have either been shown not to have prenyl diphosphate hydrolyzing activity or have not been biochemically characterized (<20%).

TcNudix1, as well as AtNudix 5, 6, 11 and 15, and a number of other Nudix proteins in sequence databases, have N-terminal sequence extensions not found in RhNudix1 and AtNudix1 (Fig. 2.7). Analysis of TcNudix1 sequence using the ChloroP prediction program suggests that TcNudix1 has a 56 amino acid plastid transit peptide at the N-terminus (<http://www.cbs.dtu.dk/services/ChloroP/>) that targets it to the plastids. To experimentally determine the subcellular localization of TcNudix1, we created a fused gene construct starting with the TcNudix1 ORF and terminating with the GFP ORF, placed it in an expression vector driven by the 35S promoter, and transformed Arabidopsis protoplasts. Confocal microscope images showed that in protoplasts transformed with the TcNudix1-GFP fusion construct, the observed GFP signal overlapped with the chloroplast autofluorescence signal, but in protoplasts transformed with a stand-alone GFP ORF the GFP signal was present in the cytosol (Appendix Fig. S2.5).

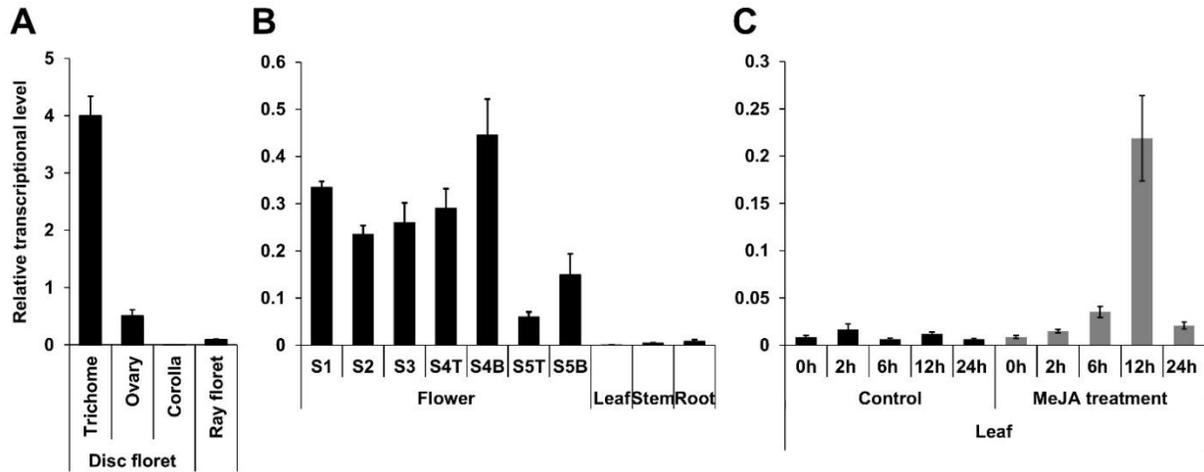


Figure 2.8 Tissue-specific expression of *TcNudix1*. **(A)** RT-qPCR analysis of *TcNudix1* transcript levels in different parts of stage-3 flowers. **(B)** RT-qPCR analysis of *TcNudix1* transcript levels in different developmental stages of flowers, leaf, stem, and root ('T': ray floret, 'B': disk florets). **(C)** RT-qPCR analysis of 2-week-old leaves treated with MeJA. (Data are presented as means \pm SD, $n = 3$ or 4).

Tissue-specific expression of TcNudix1

Genes participating in the early steps of the synthesis of the terpene moiety of pyrethrins are highly expressed in the trichomes of ovary, show increased expression in later stages of flower development, and are induced by MeJA in leaves [37]. Analysis of TcNudix1 gene expression (Fig. 2.8) shows that *TcNudix1* is also highly specific to trichomes (Fig. 2.8A), and its developmental stage expression pattern is similar to other genes involved in the early steps of the synthesis of the terpene moiety of pyrethrins (Fig. 2.8B). Furthermore, *TcNudix1* expression is also induced by MeJA (Fig. 2.8C).

Characterization of the hydrolysis activity of TcNudix1

To characterize TcNudix1 enzymatic activity, we produced a recombinant protein (lacking the part of the ORF encoding the transit peptide and containing a His-tag at the C-terminus) in *E. coli*, purified it by affinity chromatography (Appendix 2.19), and tested its activity with various prenyl diphosphate substrates (Fig. 2.9A). TcNudix1 showed highest levels of activity with the *cis*-prenyl diphosphates NPP and zFPP as well as with CPP. It was less active with IPP, GPP, and eFPP, and showed no activity with DMADP and GGPP (Fig. 2.9A, Appendix Table S2.3). TcNudix1 hydrolyzed CPP to CMP (Fig. 2.9B) with a K_m value of $0.137 \pm 0.05 \mu\text{M}$.

We further tested for the presence of a second phosphatase in pyrethrum flowers and leaves by performing *in vitro* coupled assays that included DMADP and TcCDS (to generate CPP) and TcNudix1 alone or both TcNudix1 and crude protein extracts from pyrethrum flowers or leaves (Fig. 2.10). After overnight incubation, the reaction products were extracted and analyzed by GC-MS (Fig. 2.10). Chrysanthemol was detected when TcCDS, TcNudix1 and crude

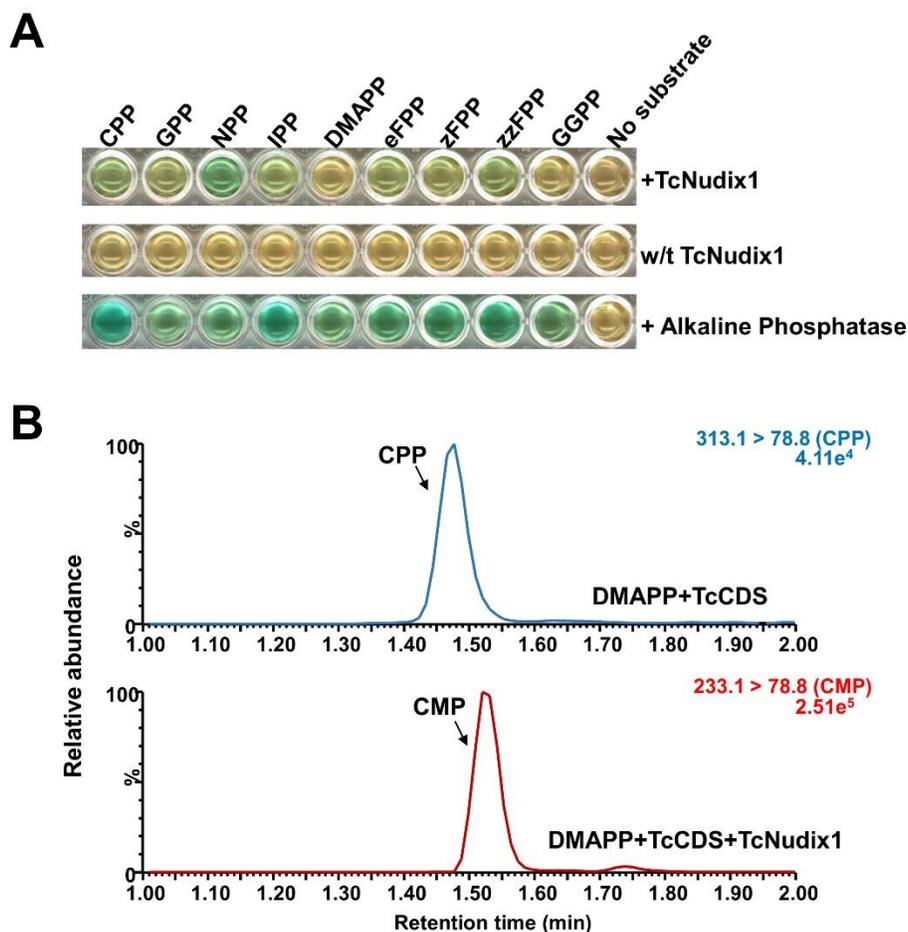


Figure 2.9 Activity of TcNudix1 with prenyl diphosphate substrates. **(A)** A colorimetric assay in which green color indicates production of free phosphate. In each well, 5 μ g of purified TcNudix1 was incubated with 10 μ M of the indicated substrate for 30 min. **(B)** LC-MS analysis of conversion of CPP to CMP catalyzed by TcNudix1. DMADP was incubated with TcCDS to generate CPP as the substrate for TcNudix1, then TcNudix1 was added to produce CMP, which was detected with LC-MS under MRM mode; precursor ion and product ion for CPP are 313.1 and 78.8 respectively (313.1 > 78.8), and the ions for CMP are 233.1 and 78.8.

protein from pyrethrum flowers or leaves were included. No chrysanthemol was detected when CPP was incubated with TcNudix1 alone without leaf or floral crude protein extracts, or with floral or leaf crude protein extracts without added TcNudix1. The most likely explanation for the latter observation is that concentration of TcNudix1 in the crude extract, while high relative to other proteins (assumed from the transcript data), is still much lower than the concentration of the added purified recombinant TcNudix1 protein, so that the two-step reaction of converting the exogenously added CPP to chrysanthemol is too slow to yield a detectable product.

Heterologous co-expression of TcNudix1 and TcCDS in tomato

We observed that expression of TcCDS in the tomato fruit using the PG promoter greatly reduced the accumulation of the red pigment lycopene. We hypothesized that CPP made in these fruits competitively inhibits the activity of geranylgeranyl diphosphate synthase (GGPPS), an enzyme that catalyzes the formation of the lycopene precursor GGPP from DMADP and IPP [22,23]. To test if TcNudix1 can hydrolyze CPP in tomato fruit expressing *TcCDS* and thereby relieve the inhibition of GGPPS, we obtained transgenic tomato plants expressing TcNudix1 under the control of the PG promoter and crossed these lines to the *TcCDS*-overexpressing lines and measured the concentration of lycopene (Fig. 2.11). Plants expressing *TcCDS* alone (Fig. 2.11A, B) indeed showed greatly decreased levels of lycopene biosynthesis (Fig. 2.11C, D). In contrast, plants expressing *TcNudix1* in addition to *TcCDS* (Fig. 2.11A, B) had wild-type levels of lycopene (Fig. 2.11C, D).

Expression of TcCDS in type VI trichomes of S. lycopersicum

To test the feasibility of augmenting plant chemical defenses, we expressed the *TcCDS* transgene in *S. lycopersicum* M82 plants under the control of the type VI trichome-specific

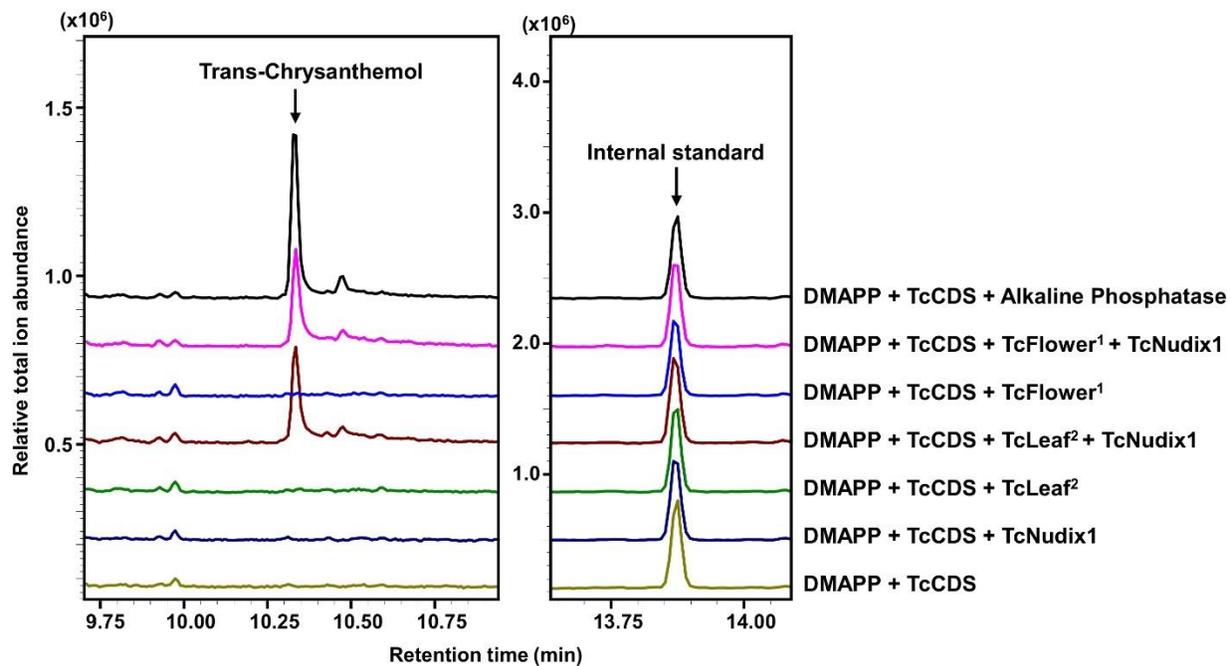


Figure 2.10 Complete conversion of CPP to chrysanthemol by TcNudix1 and additional phosphatase(s) from pyrethrum tissues. DMADP was incubated with TcCDS to generate CPP, followed by the addition of TcNudix1, TcNudix with crude flower protein (TcFlower¹), only crude flower protein extract, TcNudix1 with crude leaf protein (TcLeaf²), and only crude leaf protein. Chrysanthemol was detected by GC-MS. DMADP and TcCDS with alkaline phosphatase treatment was used as positive control. Tetradecane was added as internal standard.

metallocarboxypeptidase inhibitor (MCPI) promoter. Five transgenic lines were obtained that accumulated levels of volatile *trans*-chrysanthemol and *trans*-chrysanthemal ranging from 0.03 $\mu\text{mol/g}$ dry weight in line CDS15 to 0.13 $\mu\text{mol/g}$ dry weight in line CDS2 (Fig. 2.12A). The chrysanthemol malonyl glycoside previously observed in tomato fruit and *N. benthamiana* leaves expressing the *TcCDS* gene [17,18] also accumulated in these lines, reaching a much higher level than the volatile compounds (5.9 $\mu\text{mol/g}$ dry weight in CDS8 to 12.6 $\mu\text{mol/g}$ dry weight in CDS7; Fig. 2.12B). The ratio of volatile to glycosylated CDP derivatives that accumulate in type VI trichomes is similar to that observed when *TcCDS* was expressed in tomato fruit (Fig. 2.5B). In contrast to the transgenic tomato fruits, no *trans*-chrysanthemic acid or chrysanthemyl malonyl glycoside was observed in lines expressing *TcCDS* in trichomes.

To determine whether co-expressing *TcCDS* with additional genes from the chrysanthemic acid pathway could increase the total production of CDP derivatives in type VI trichomes, we crossed transgenic lines CDS8 and CDS15 with lines expressing the *TcNudix1* gene under control of the MCPI promoter. This resulted in a statistically significant increase in levels of volatile *trans*-chrysanthemol and *trans*-chrysanthemal in all hybrid lines tested (Fig. 2.13A, B). Co-expression of *TcCDS* and *TcNudix1* in trichomes had mixed results with respect to accumulation of chrysanthemol malonyl glycoside. When line CDS8 was crossed with *TcNudix1*-expressing lines (either line Nudix1-1 or Nudix1-8), it yielded an apparent increase in glycoside levels (Fig. 2.13C). However, the increases were not statistically significant. Crossing line CDS15 with lines Nudix1-8 or Nudix1-10 resulted in little or no change of glycoside levels (Fig. 2.13C). As chrysanthemol malonyl glycoside accounts for most CDP derivative in this system (Fig. 2.12A, B), this indicates that *TcNudix1* has little impact on accumulation of CDP derivatives.

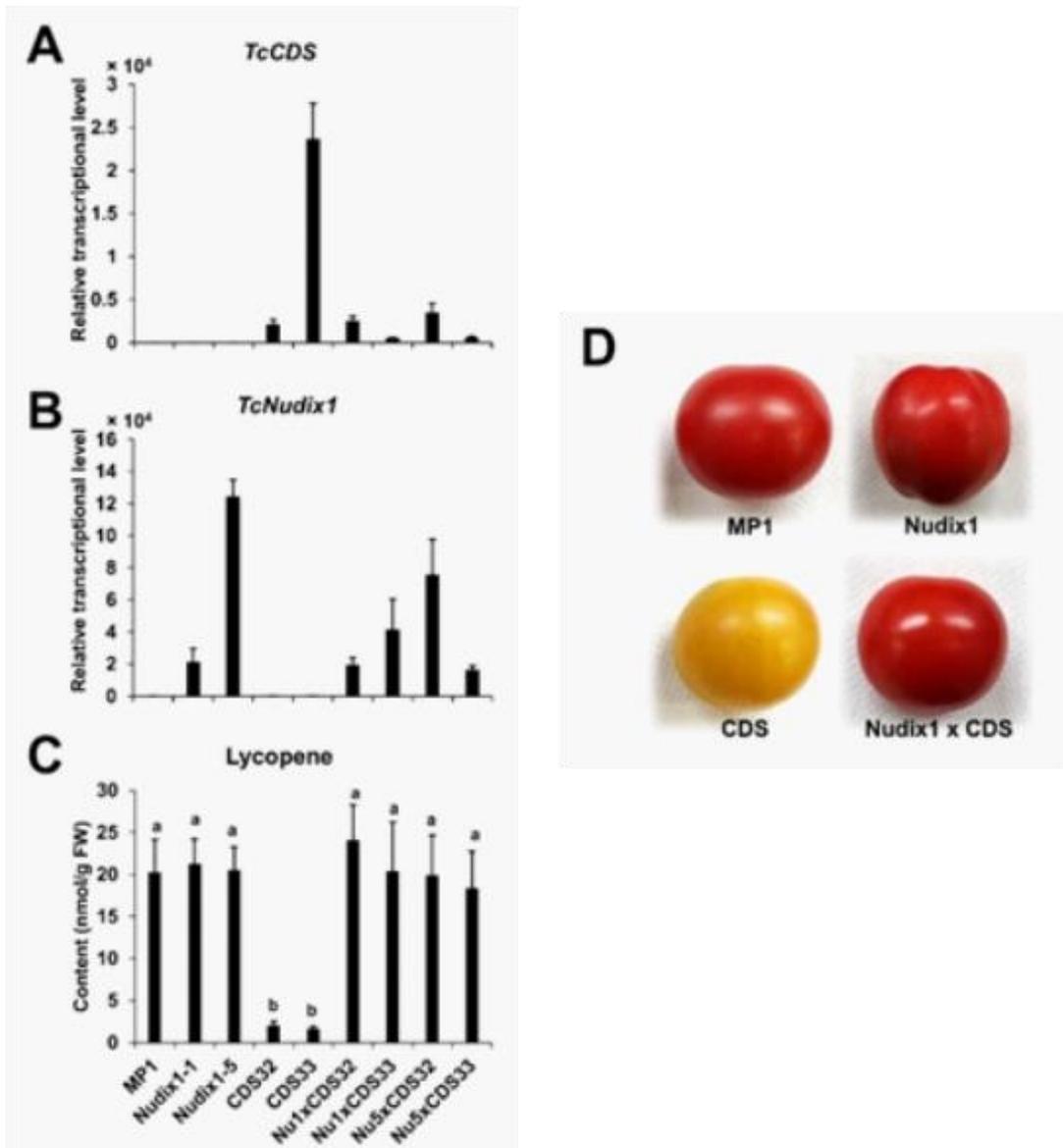


Figure 2.11 Coexpressing *TcNudix1* with *TcCDS* in tomato fruit restores wild-type lycopene levels. **(A)** RT-qPCR analysis of *TcNudix1* transcript levels in fruits of tomato transgenic lines expressing *TcCDS* (CDS32/33), *TcNudix1* (Nudix1-1/5) and coexpression lines (Nu1/5 × CDS32/33). **(B)** RT-qPCR analysis of *TcNudix1* transcript levels in fruits of each lines. **(C)** Lycopene content in fruits of each lines. **(D)** Color of mature fruits from different lines.

DISCUSSION

Metabolic engineering has been used to produce plant-derived compounds in a variety of microbial and plant hosts [3,4,8,9,40]. In most cases, extraction and purification of high-value compounds for medicinal or other purposes is the intention. However, alteration of plant metabolism can also improve nutritional quality or enhance stress tolerance [11,13,41]. Pyrethrin insecticides produced by *T. cinerariifolium* are an endogenous chemical defense for the plant and a valuable pest control tool for humans [42–44]. We explored both aspects of plant metabolic engineering – increasing production of a valuable compound and improving the properties of a crop plant – using the pyrethrin pathway. At the outset of this project, few genes from the *T. cinerariifolium* pyrethrin pathway were known. We pursued production of the monoterpene precursors of pyrethrins, including *trans*-chrysanthemol and *trans*-chrysanthemic acid, using the *TcCDS* gene. These compounds have proven utility both as precursors for synthesis of insect hormones for baiting insect traps in agricultural settings and as an engineered addition to chemical defense of horticultural crops against insect herbivores [26,45–47].

Ripe fruits of cultivated tomato, *S. lycopersicum*, accumulate high levels of the carotenoid lycopene averaging ~100 mg/kg fresh weight under field conditions [20]. This indicates a high natural flux through the plastidial terpenoid pathway and makes tomato fruits an attractive platform for engineering heterologous production of other plastidially derived terpenoids. Several metabolic engineering studies previously investigated tomato fruit as a platform for monoterpene production. Introducing a single monoterpene synthase such as linalool synthase from *Clarkia breweri* or geraniol synthase from *Ocimum basilicum* into

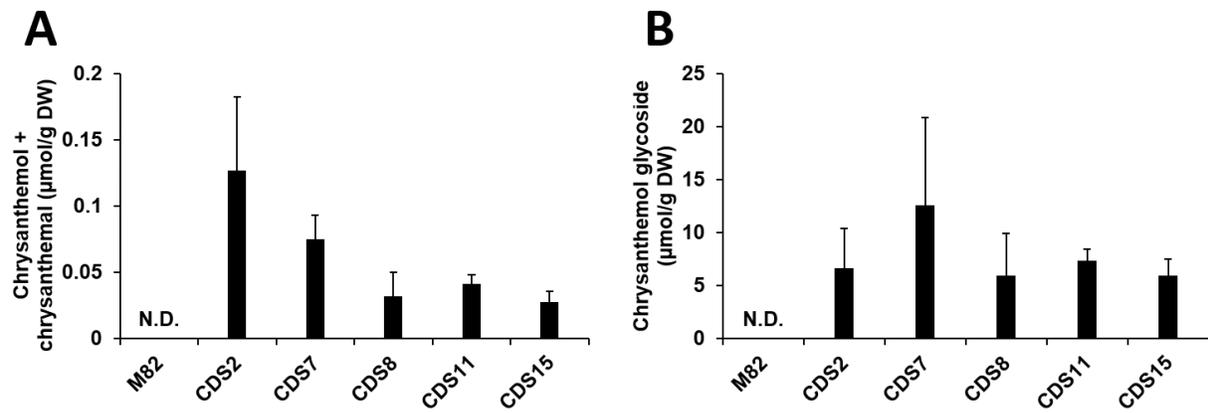


Figure 2.12 Accumulation of CPP derivatives in type VI trichomes of *S. lycopersicum* expressing the *TcCDS* gene. **(A)** Levels of *trans*-chrysanthemol, *trans*-chrysanthemal. **(B)** Levels of chrysanthemol malonyl glycoside. Data are presented as means \pm SD; $n = 3$.

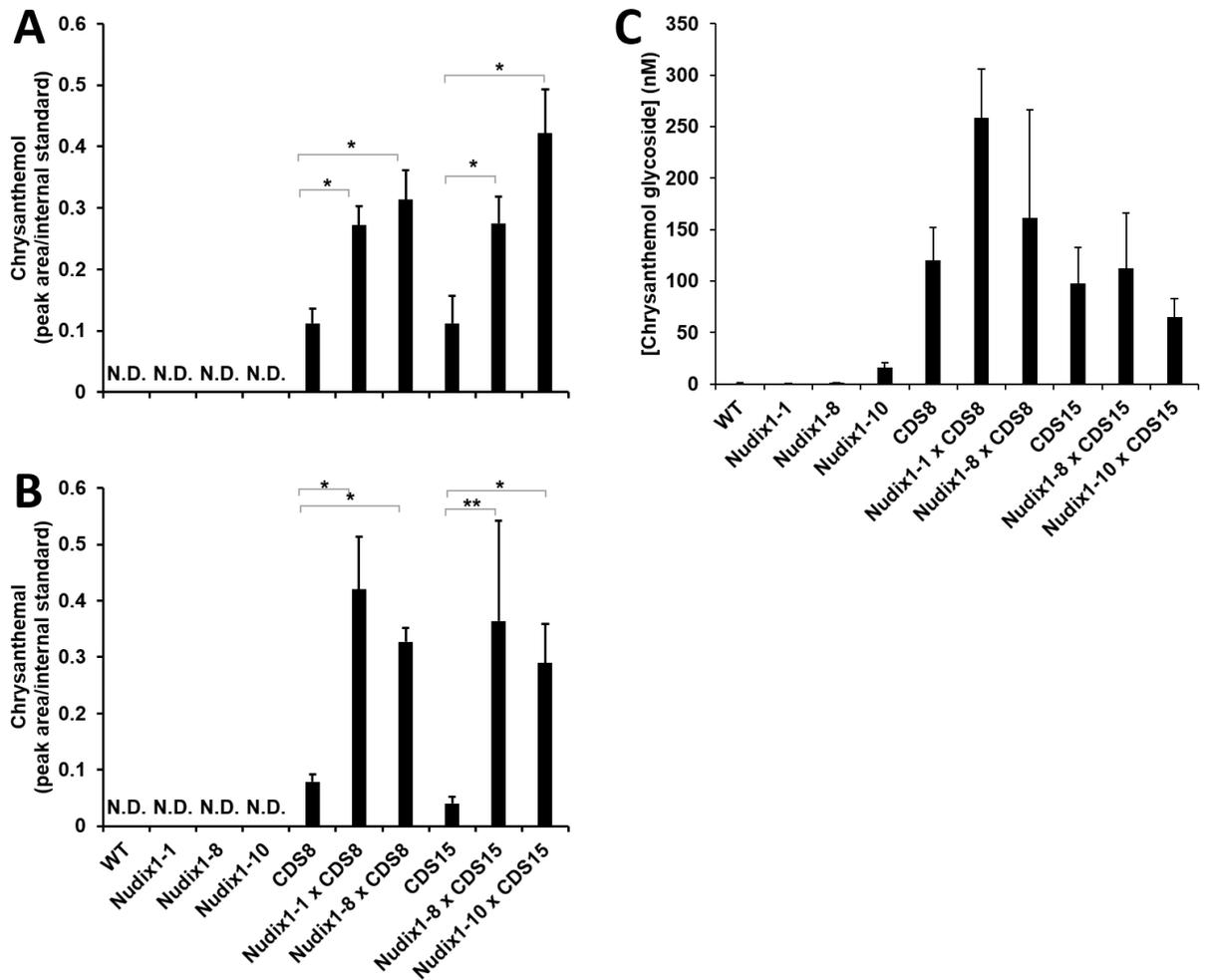


Figure 2.13 Accumulation of CPP derivatives in type VI trichomes of *S. lycopersicum* lines expressing the *TcCDS* gene crossed with lines expressing the *TcNudix1* gene. **(A)** Accumulation of *trans*-chrysanthemol. **(B)** Accumulation of *trans*-chrysanthemol. **(C)** Accumulation of chrysanthemol malonyl glycoside. Data are presented as means \pm SD, $n = 3-5$. ANOVA: * – $p < 0.05$, ** – $p < 0.01$.

S. lycopersicum under a fruit-specific promoter yielded production of the expected volatile compounds linalool and geraniol [48,49]. Fruit expression of the *S. lycopersicum neryl diphosphate synthase (SINDPS1)* gene, typically restricted to tomato trichomes, yielded production of nerol while fruit expression of the tomato *phellandrene synthase (SIPHS1)* gene, also typically restricted to tomato trichomes, led to a mixture of myrcene, ocimene, and geraniol. Combined expression of both *SINDPS1* and *SIPHS1* in tomato fruit caused production of the full suite of volatile monoterpenes typically observed in *S. lycopersicum* type VI trichomes, including the major products β -phellandrene and δ -2-carene, indicating that co-expression of pathway genes can improve the efficacy of metabolic engineering in tomato fruit [23,27]. We also used tomato fruit as a platform to attempt high-level production of monoterpenoids from the pyrethrin pathway.

Expression of *TcCDS* alone in tomato fruit yielded modest accumulation of volatile monoterpenoids, including *trans*-chrysanthemol, *trans*-chrysanthemal, and *trans*-chrysanthemic acid. We also analyzed transgenic tomato fruit for the presence of glycosylated monoterpenoids. We identified multiple monoterpenoid glycosides, including a previously characterized chrysanthemol malonyl glycoside and a novel chrysanthemyl malonyl glycoside [17,18]. As was observed in previous monoterpenoid engineering efforts in tomato fruit, lycopene levels dropped dramatically [22,23,49]. Reduced lycopene levels are to be expected as some of the flux through the DXP pathway is re-routed from carotenoid biosynthesis to monoterpenoid biosynthesis. However, total levels of monoterpenoids did not fully account for the decrease in lycopene observed.

When *TcCDS* was co-expressed with *ShADH* and *ShALDH*, we observed both higher total yields of monoterpenoid production and a greater proportion of oxidized products in the form of volatile *trans*-chrysanthemic acid and chrysanthemyl malonyl glycoside. Interestingly, lycopene levels were also higher in transgenic lines expressing all three genes than in those expressing *TcCDS* alone. It was previously shown that heterologous production of neryl diphosphate in tomato fruit led to inhibition of geranyl-geranyl diphosphate synthase (GGPPS), an early enzyme in the lycopene biosynthetic pathway [23]. We hypothesize that chrysanthemyl diphosphate (CPP), the immediate product of the *TcCDS* enzyme, also inhibits GGPPS, leading to the reduced levels of lycopene in *TcCDS* lines. The observation that combined expression of *TcCDS* enzyme with the dehydrogenases *ShADH* and *ShALDH* leads to a partial restoration of lycopene production supports this theory. These dehydrogenases that accelerate conversion of *trans*-chrysanthemol to *trans*-chrysanthemal and *trans*-chrysanthemic acid may indirectly reduce steady-state pools of CPP in the fruit, thereby relieving inhibition on GGPPS and allowing increased carotenoid production despite high flux from the DXP pathway to monoterpenoids. Co-expression of *TcCDS* with *TcNudix1*, a phosphatase we identified in *T. cinerariifolium* that shows high specificity for CPP, indeed yielded transgenic lines that expressed wild-type levels of lycopene, supporting the hypothesis that reduced lycopene content in lines expressing *TcCDS* results from CPP-mediated inhibition of the carotenoid pathway.

When attempting to alter a complex biological system containing many unknown or underdetermined biochemical processes, some perturbations intended to increase yield of a particular compound may lead instead to unintended effects that detrimentally impact

accumulation of the desired product. Generation of *trans*-chrysanthemol and its oxidized or glycosylated derivatives in tomato fruits expressing *TcCDS* alone or in combination with *ShADH* and *ShALDH* is expected to rely on promiscuous endogenous phosphatases from the tomato fruit. Presence of an additional phosphatase that is specific for CPP, such as TcNudix1, is expected to increase the maximum possible flux from CPP to downstream products. While this could in theory lead to greater accumulation of *trans*-chrysanthemol in transgenic tomato fruits, it also clearly relieves inhibition of the carotenoid pathway in fruit. When lycopene levels are low, indicating significant inhibition of the fruit carotenoid pathway, total production of CPP-derived monoterpenoids in transgenic lines can be nearly double the production of lycopene observed in wild-type tomato fruit. This high-level production of monoterpenoids may in fact be dependent on suppression of carotenoid biosynthesis by a significant pool of inhibitory prenyl diphosphate (here, CPP). If this is the case, the increase in maximum theoretical flux through the monoterpenoid pathway afforded by the activity of TcNudix1 may come at the cost of increased competition between the TcCDS enzyme and endogenous GGPPS for DXP pathway products. Maximum production of CPP-derived compounds in this system may therefore require a balance between allowing high flux through the engineered pathway and maintaining a sufficient steady-state pool of CPP to inhibit carotenoid biosynthesis.

Expression of *TcCDS* in type VI trichomes of *S. lycopersicum* yielded a suite of CPP-derived compounds including *trans*-chrysanthemol, *trans*-chrysanthemol malonyl glycoside, and chrysanthemol malonyl glycoside similar to those found in tomato fruit expressing *TcCDS*. *trans*-Chrysanthemol and the corresponding malonyl glycoside were notably absent from trichomes. Both *trans*-chrysanthemol and chrysanthemol malonyl glycoside deterred cotton aphid (*Aphis gossypii*)

feeding on *Chrysanthemum x morifolium* plants expressing the *TcCDS* gene [26], suggesting that production of these compounds in trichomes of transgenic tomato plants could also have a protective effect. Co-expression of both *TcCDS* and *TcNudix1* in type VI trichomes statistically significantly increased the levels of *trans*-chrysanthemol and *trans*-chrysanthemal, indicating that this combination of genes could provide better protection against aphids or other insect herbivores than *TcCDS* alone. However, the effect of *TcNudix1* on levels of chrysanthemol malonyl glycoside was minor. While some lines showed an apparent increase in glycoside levels, these changes were not statistically significant. Given that the majority of CPP derivatives accumulate in trichomes as chrysanthemol malonyl glycoside, it appears that co-expression of *TcNudix1* with *TcCDS* in type VI trichomes has little impact on monoterpene metabolism. The relative efficacy of *trans*-chrysanthemol versus chrysanthemol malonyl glycoside in deterring insect herbivores is unknown, preventing us from predicting the possible effect of *TcNudix1* expression on transgenic tomato insect resistance.

Results from expression of *TcCDS* alone and in combination with accessory genes (*ShADH*, *ShALDH*, or *TcNudix1*) in both fruit and type VI trichomes of *S. lycopersicum* demonstrate the efficacy of metabolic engineering for production of useful compounds for extraction (fruit) or for enhancement of endogenous plant chemical defenses (trichomes). Levels of *trans*-chrysanthemic acid and its glycoside produced by the activity of *TcCDS*, *ShADH*, *ShALDH*, and endogenous tomato enzymes exceeded the capacity for terpene production predicted based on wild-type levels of lycopene in fruit. It is notable that this reconstructed pathway contains only one enzyme associated with *trans*-chrysanthemic acid biosynthesis in its native host. Dehydrogenases from *S. habrochaites* were used for these experiments but the

alcohol dehydrogenase 2 (TcADH2) and aldehyde dehydrogenase 1 (TcALDH1) enzymes responsible for converting *trans*-chrysanthemol to *trans*-chrysanthemic acid in *T. cinerariifolium* have since been characterized. Future engineering efforts using these enzymes specific to the natural pyrethrin pathway could increase total levels of monoterpenoids or the proportion of oxidized products. Expression of *TcCDS* in type VI trichomes was also successful and resulted in accumulation of two compounds, *trans*-chrysanthemol and chrysanthemol malonyl glycoside, which have protective effects against insect herbivores. However, in this case, combining *TcCDS* with another gene from the pyrethrin pathway, *TcNudix1* did not have a clear positive effect on production of the desired compounds. Together, these two projects demonstrate the effectiveness of metabolic engineering both for high-level production of useful compounds and for improvement of existing crops. However, the apparent action of unidentified endogenous phosphatases, oxidoreductases, and glycosyltransferases in fruit and trichomes, and the unpredictable effects of *TcNudix1* expression drive home the fact that these complex biological systems are underdetermined, requiring extensive testing and empirical evidence for successful engineering.

APPENDIX

Table S2.1 Primers used in this study.

Oligo name	Oligo sequence (5'-3')
TcCDS-BamHI-F	GGATCCATGTCTTGGTGTCTCTTATGCAGTCTTTC
TcCDS-SpeI-R	ACTAGTTTACTTATGTCCCTTATACATCTTTCCAGAC
ShADH-AgeI-F	GGGACCGGTATGGAGTCAAGCAACCCAAAGGTC
ShADH-SpeI-R	GGACTAGTTTAGA AACTTGATAATAATCTTGACACAATGTG
ShALDH-AgeI-F	GGGACCGGTATGGATGCAGAGGCGATTGTGAAG
ShALDH-SpeI-R	GGACTAGTTTACCACCCAATCAAAGCACGGATG
pBin19-Nudix1-F	GGCGGATCCATGGCGATGACAGTTGGTTTAGG
pBin19-Nudix1-R	GGCACTAGTTCAAGAATGAGTAGTGAAAATATTG
MCPI-pBIN19-F	CGCGGATCCATCCTGAGCTAGAAGTTATGACCGTTTTG
MCPI2nd-overlap (Nudix)-R	AACTGTCATCGCCATATATTATGTGATGCTACTTTGATTGG
Nudix F	ATGGCGATGACAGTTGGTTTAGG
Nudix (NOS)-overlap R	AATGTTTGAACGATCTCAAGAATGAGTAGTGAAAATATTG
NOS-F	GATCGTTCAAACATTTGGCAATAAA
NOS-R	GCCGAATTCGATCTAGTAACATAGATGACACCGCG
TcCDS-CT-F	ATGACTACGACATTGAGCAGCAATCTAGAC
TcCDS-CT-R	CTTATGTCCCTTATACATCTTTCCAGAC
ShADH-pHIS8-F	CATGCCATGGAGTCAAGCAACCCAAAGGTC
ShADH-pHIS8-R	ACGCGTCGACTTAGAACTTGATAATAATCTTGACACAATGTG
ShALDH-pHIS8-F	CGGAATTCATGGATGCAGAGGCGATTGTGAAG
ShALDH-pHIS8-R	CCGCTCGAGTTACCACCCAATCAAAGCACGGATG
pET28-TcNudix1 (-56aa)-F	GGCGAATTCAAAACAAGGAACGAGCATTTC
pET28-TcNudix1 (-56aa)-R	GGCGTCGACTCAAGAATGAGTAGTGAAAATATTG
SBE634_ADH_F	TTTGAAGACAAAATGGAGTCAAGCAACCCAAAG
SBE635_ADH_R	TTTGAAGACAAAAGCTTAGAACTTGATAATAATCTTGACAC
SBE537_ALDH_F	TTTGAAGACAAAATGGATGCAGAGGCGATTGTGAAGGAAT TGAGAGGGACGTACGGGAGTGGGAAAC
SBE538_ALDH_R	TTTGAAGACAAAAGCTTACCACCCAATCAAAGCACG
TcCDS ^a -CGFP-F	CCGCTCGAGATGGCTTGCTCTAGTAGTCTTTCTTCC
TcCDS ^b -CGFP-F	CCGCTCGAGATGTCTTGGTGTCTCTTATGCAGTCTTTC
TcCDS-CGFP-R	CCGGAATTCGCTTATGTCCCTTATACATCTTTCCAGAC
pEZY-NL-Nudix1-F	GGCGAATTCATGGCGATGACAGTTGGTTTAGG
pEZY-NL-Nudix1-R	GGCGGATCCCAAGAATGAGTAGTGAAAATATTG
EF-RT-F	GCTCTCCAGGAGGCACTCCCTG
EF-RT-R	CTTGGCTGGGTCATCCTTGGAG
ShADH-RT-F	GGAGCAACATGGAGAGAAGTTCATG
ShADH-RT-R	CCTTGACTTCGAGCTCCCTCAATC
ShALDH-RT-F	CAGTTGTGGTGGATTCAAACATCG
ShALDH-RT-R	ATCAGGAGAGATGCACGTTTGTCC
TcCDS-RT-F	ACGTGCATCTTCTGGACCTCTTC
TcCDS-RT-R	TGAACAATCCGACGGTTAAGAGTC

Table S2.1 (cont'd)

Oligo name	Oligo sequence (5'-3')
Nudix1 RT1-F	CTCGGGGAAGAATGCTAAATCA
Nudix1 RT1-R	ACACCGTAACCCCAACCTCTG
Nudix1 RT2-F	TCGCGGCTCAACTCGT
Nudix1 RT2-R	GGCGTCTGATTTGGGTCTGA

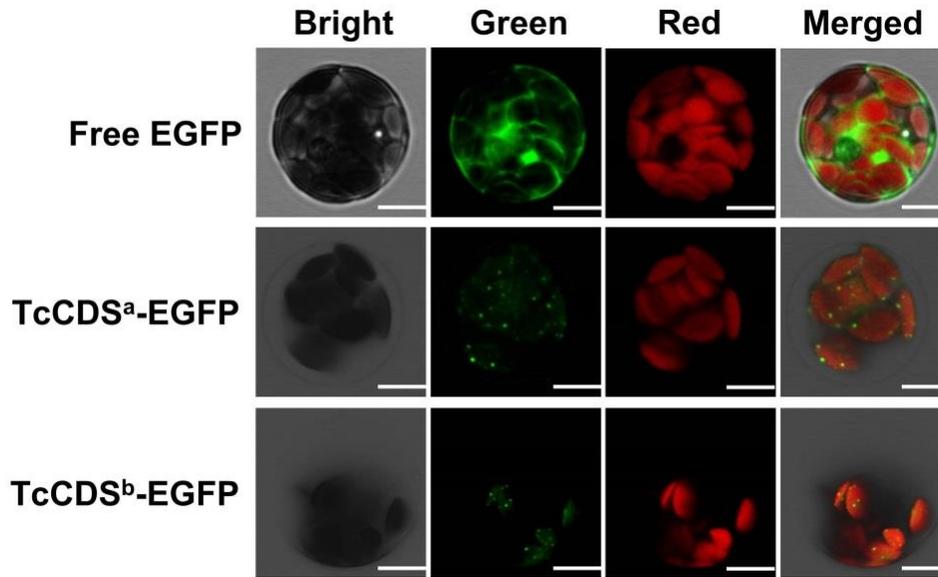


Figure S2.1 Subcellular localization of transiently expressed TcCDS in Arabidopsis leaf mesophyll protoplasts. The complete open reading frame of *TcCDS* was introduced into the pEYL-NL vector and fused at the N-terminal of the Enhanced Green Fluorescence Protein (*EGFP*), then transiently expressed in Arabidopsis leaf mesophyll protoplasts. The signals of the *EGFP*-fused proteins were visualized by laser confocal microscopy (shown in green). Chloroplasts are identified by red chlorophyll autofluorescence (shown in red). The column labeled “Merged” provides a view of all fluorescent signals obtained for this sample. TcCDS^a - the complete open reading frame of *TcCDS* was amplified from cDNA of *T. cinerariifolium* flower by RT-PCR according to the published sequence in [17]. TcCDS^b - the complete open reading frame of *TcCDS* was synthesized according to the published sequence in [14]. Scale bars =10 μ m.

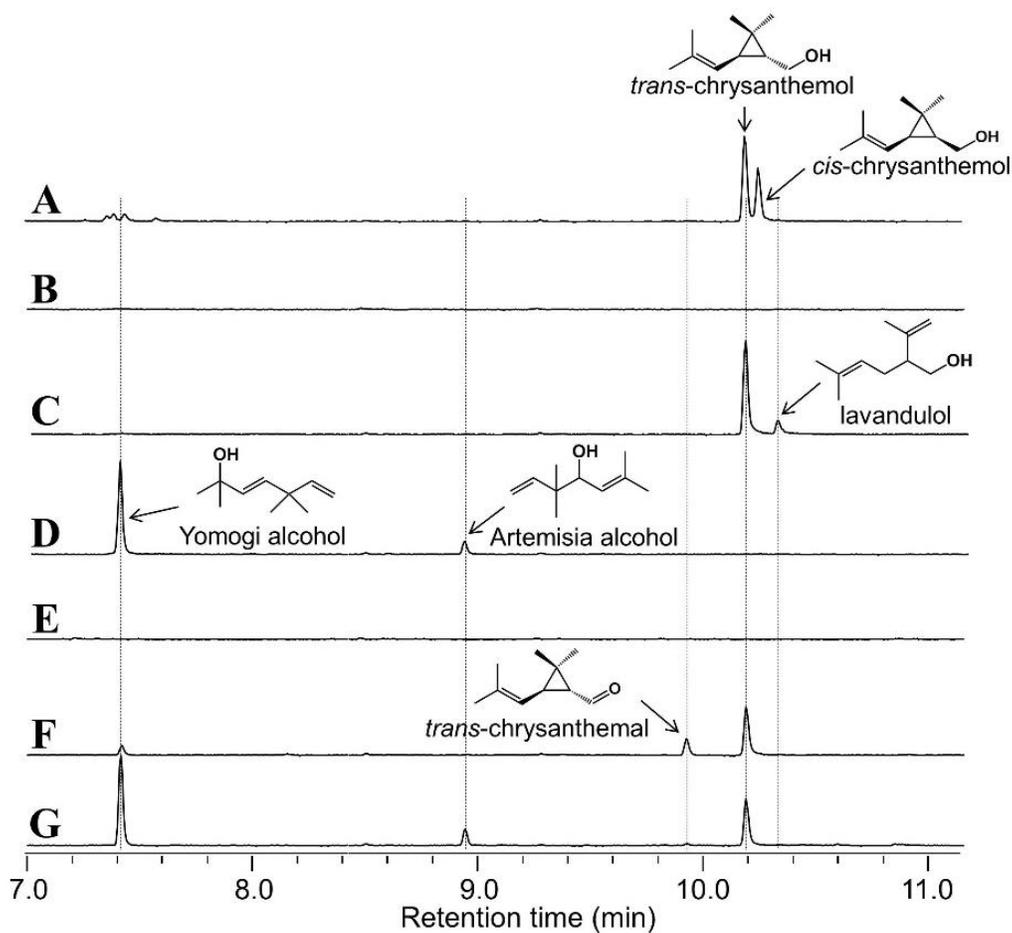


Figure S2.2 GC-MS analysis of hexane extracts of reaction products catalyzed in vitro by TcCDS and by TcCDS32 transgenic tomato fruit. MS detection is by total ion. A, Standard mix of *trans*- and *cis*-chrysanthemol. B, CDS reaction products. The product of the reaction is CDP, which is not soluble in hexane. The chromatograph shows that no free *trans*-chrysanthemol was synthesized in this reaction. C, CDS reaction products first treated with alkaline phosphatase. D, CDS reaction products first treated with 0.2 M HCl. E, Non-transgenic tomato fruit homogenate. F, TcCDS32 tomato fruit homogenate. G, TcCDS32 tomato fruit homogenate first treated with 0.2 M HCl.

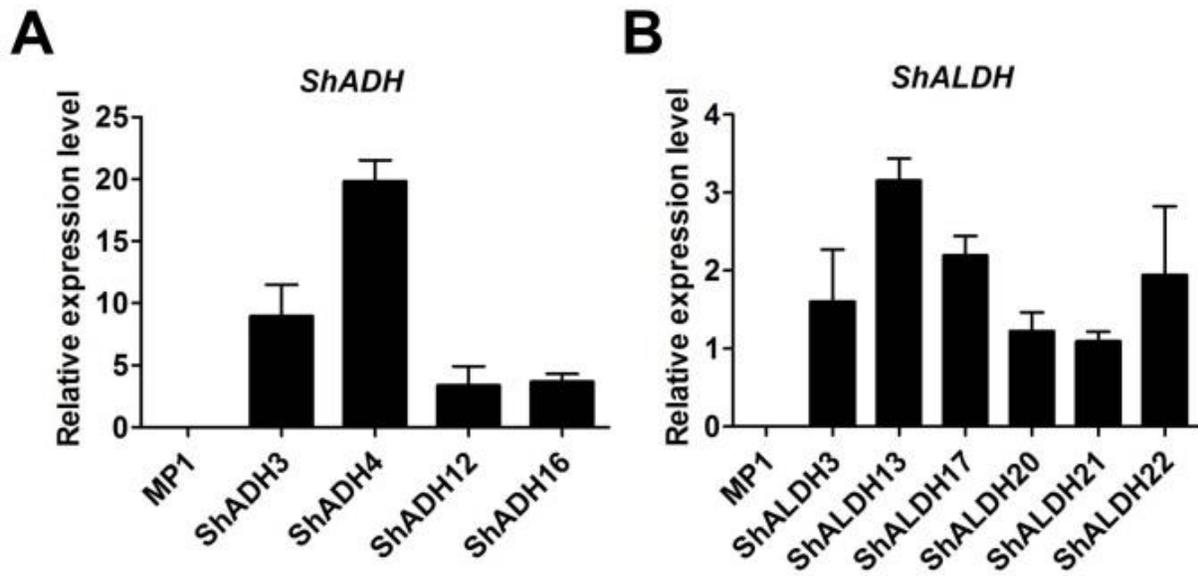


Figure S2.3 qRT-PCR analysis of transcript levels of transgenes in several transgenic lines of *ShADH* (A) and *ShALDH* (B). Results are expressed relative to that of elongation factor 2 (means \pm SD, $n = 3$).

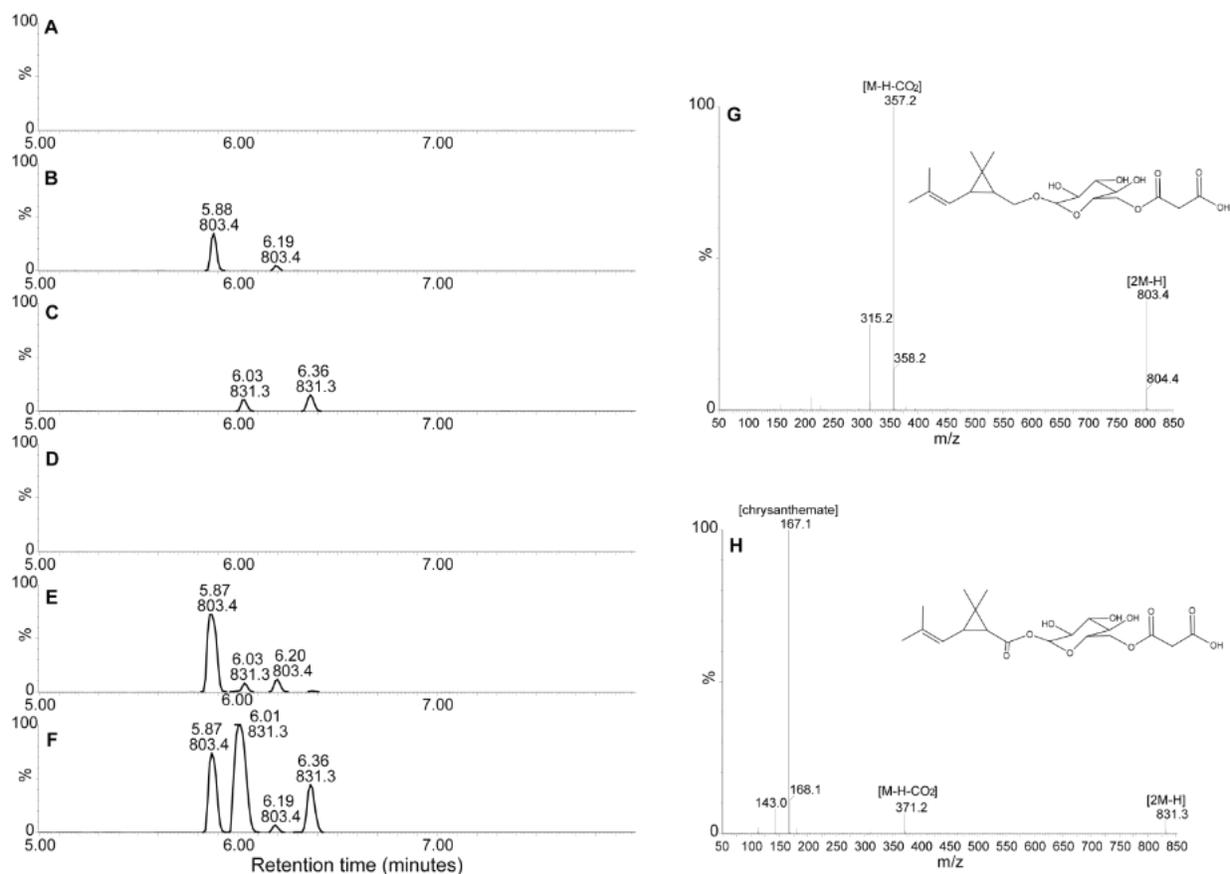


Figure S2.4 LC-MS analysis of plant extracts. A-C: extracted ion chromatograms showing chrysanthemol malonyl glycosides (m/z 803.37) and chrysanthemic acid malonyl glycosides (m/z 831.33) in leaves of *Nicotiana benthamiana*. These extracts were produced in work by Xu et al., (2018) and are shown here as markers for the corresponding glycosides observed in tomato fruit. A, extract from unfiltered leaf; B, extract from leaf transiently expressing *TcCDS*; C, extract from leaf transiently expressing *TcCDS*, *TcADH2*, and *TcALDH1*. D-F: extracted ion chromatograms showing chrysanthemol malonyl glycosides and chrysanthemic acid malonyl glycosides in fruits of *Solanum lycopersicum*: D, extract from fruit of WT *S. lycopersicum*; E, extract from fruit of transgenic line *TcCDS32* expressing *TcCDS*; F, extract from fruit of transgenic line expressing *TcCDS*, *ShADH*, and *ShALDH*. G, mass spectrum of chrysanthemol malonyl glycoside of peak with RT 6.20 min from (E) consistent with spectrum of previously reported chrysanthemol malonyl glycoside (structure depicted) [17]. H, mass spectrum of chrysanthemic acid malonyl glycoside with RT 6.36 min from (F; structure extrapolated from chrysanthemol malonyl glycoside).

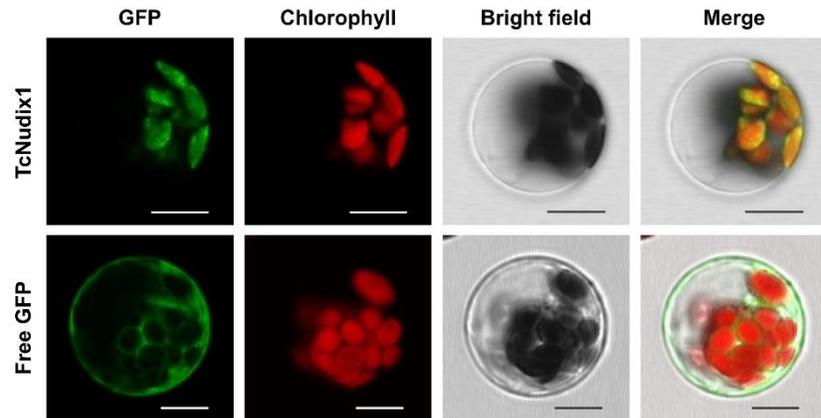


Figure S2.5 Representative results showing subcellular localization of TcNudix1. Images from left to right are: green fluorescence from GFP, autofluorescence from chloroplasts, bright field under visible light, and overlay of those three images. Images for the same gene are of the same cell. Unfused GFP construct was used as control. Scale bars = 10 μm .

Table S2.2 Correlation analysis

Rank	Gene	Pearson Correlation	P value	Predicted function
	CDSase	1		
1	TRINITY_DN91754_c0_g2_i1	0.999098465	1.83E-09	
2	TRINITY_DN158798_c1_g5_i1	0.996393869	1.17E-07	
3	TRINITY_DN149780_c0_g1_i2	0.994524669	4.09E-07	Glutathione S transferase
4	TRINITY_DN159661_c0_g4_i9	0.994044212	5.26E-07	Heat_shock_protein
5	TRINITY_DN147751_c0_g1_i2	0.993810671	5.90E-07	
6	TRINITY_DN129539_c0_g1_i1	0.993336904	7.36E-07	Lipase
7	TRINITY_DN128614_c0_g1_i1	0.992671839	9.78E-07	Acyl-transferase
8	TRINITY_DN75102_c0_g2_i1	0.991266721	1.65E-06	
9	TRINITY_DN135414_c0_g1_i1	0.9895933	2.80E-06	Laccase-22_Multicopper_oxidase
10	TRINITY_DN182299_c0_g1_i1	0.989122858	3.19E-06	
11	TRINITY_DN159654_c0_g1_i3	0.98809509	4.18E-06	
12	TRINITY_DN158987_c4_g2_i1	0.987774188	4.53E-06	Nudix1
13	TRINITY_DN4250_c0_g1_i1	0.987757184	4.55E-06	
14	TRINITY_DN130811_c0_g1_i1	0.987661988	4.65E-06	Pectinesterase
15	TRINITY_DN142284_c0_g2_i7	0.986848332	5.63E-06	Protein_phosphatase
16	TRINITY_DN143344_c0_g1_i1	0.986703478	5.82E-06	Pinoresinollariciresinol reductase
17	TRINITY_DN157112_c1_g1_i1	0.986141549	6.59E-06	
18	TRINITY_DN157112_c1_g1_i1	0.986141549	6.59E-06	
19	TRINITY_DN151384_c1_g2_i1	0.985915932	6.91E-06	aldehyde dehydrogenase
20	TRINITY_DN159641_c5_g2_i3	0.984992529	8.36E-06	

Table S2.3 Substrate specificity of TcNudix1 to CPP, GPP and GGPP.

Substrate^a	Relative activity
Chrysanthemyl diphosphate (CPP)	100% ^b
Geranyl diphosphate (GPP)	11.26 ± 3.08%
Farnesyl diphosphate (FPP)	10.01 ± 0.78%
Geranyl-geranyl diphosphate (GGPP)	Not Detected

a. Concentration of substrate is 0.1 μM

b. 100% represent 2.93 nmol/s/μmol protein

REFERENCES

REFERENCES

- 1 Witherup, K.M. *et al.* (1990) *Taxus spp.* Needles Contain Amounts of Taxol Comparable to the Bark of *Taxus brevifolia*: Analysis and Isolation. *J. Nat. Prod.* 53, 1249–1255
- 2 Abdallah, I.I. *et al.* (2019) Metabolic Engineering of *Bacillus subtilis* Toward Taxadiene Biosynthesis as the First Committed Step for Taxol Production. *Front. Microbiol.* 10, 218
- 3 Li, J. *et al.* (2019) Chloroplastic metabolic engineering coupled with isoprenoid pool enhancement for committed taxanes biosynthesis in *Nicotiana benthamiana*. *Nat. Commun.* 10, 4850
- 4 El-Sayed, E.-S.R. *et al.* (2020) Semi-continuous production of the anticancer drug taxol by *Aspergillus fumigatus* and *Alternaria tenuissima* immobilized in calcium alginate beads. *Bioprocess Biosyst. Eng.* DOI: 10.1007/s00449-020-02295-8
- 5 Graham, I.A. *et al.* (2010) The Genetic Map of *Artemisia annua* L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin. *Science* 327, 328–331
- 6 Li, J. *et al.* (2014) Comparative analysis of pyrethrin content improvement by mass selection, family selection and polycross in pyrethrum [*Tanacetum cinerariifolium* (Trevir.) Sch.Bip.] populations. *Ind. Crop. Prod.* 53, 268–273
- 7 Hitmi, A. *et al.* (2001) Effects of plant growth regulators on the growth and pyrethrin production by cell cultures of *Chrysanthemum cinerariaefolium*. *Aust. J. Bot.* 49, 81–88
- 8 van Herpen, T.W.J.M. *et al.* (2010) *Nicotiana benthamiana* as a Production Platform for Artemisinin Precursors. *PLoS ONE* 5, e14222
- 9 Paddon, C.J. *et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–532
- 10 Tabashnik, B.E. *et al.* (2013) Insect resistance to Bt crops: lessons from the first billion acres. *Nat. Biotechnol.* 31, 510–521
- 11 Polturak, G. *et al.* (2017) Engineered gray mold resistance, antioxidant capacity, and pigmentation in betalain-producing crops and ornamentals. *Proc. Natl. Acad. Sci.* 114, 9062–9067
- 12 Yu, L. *et al.* (2019) *Arabidopsis thaliana* Plants Engineered To Produce Astaxanthin Show Enhanced Oxidative Stress Tolerance and Bacterial Pathogen Resistance. *J. Agric. Food Chem.* 67, 12590–12598

- 13 Zhang, P. *et al.* (2019) Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus. *Plant Biotechnol. J.* DOI: 10.1111/pbi.13302
- 14 Rivera, S.B. *et al.* (2001) Chrysanthemyl diphosphate synthase: Isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from *Chrysanthemum cinerariaefolium*. *Proc. Natl. Acad. Sci.* 98, 4373–4378
- 15 Kikuta, Y. *et al.* (2012) Identification and characterization of a GDSL lipase-like protein that catalyzes the ester-forming reaction for pyrethrin biosynthesis in *Tanacetum cinerariifolium*- a new target for plant protection. *Plant J.* 71, 183–193
- 16 Kuzuyama, T. and Seto, H. (2012) Two distinct pathways for essential metabolic precursors for isoprenoid biosynthesis. *Proc. Jpn. Acad., Ser. B* 88, 41–52
- 17 Yang, T. *et al.* (2014) Chrysanthemyl Diphosphate Synthase Operates *in Planta* as a Bifunctional Enzyme with Chrysanthemol Synthase Activity. *J. Biol. Chem.* 289, 36325–36335
- 18 Xu, H. *et al.* (2018) Coexpression Analysis Identifies Two Oxidoreductases Involved in the Biosynthesis of the Monoterpene Acid Moiety of Natural Pyrethrin Insecticides in *Tanacetum cinerariifolium*. *Plant Physiol.* 176, 524–537
- 19 Fraser, P.D. *et al.* (2000) Application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. *Plant J.* 24, 551–558
- 20 Garcia, E. and Barrett, D.M. (2006) Assessing Lycopene Content in California Processing Tomatoes. *J. Food Process Pres.* 30, 56–70
- 21 Hirschberg, J. (2001) Carotenoid biosynthesis in flowering plants. *Curr. Opin. Plant Biol.* 4, 210–218
- 22 Gutensohn, M. *et al.* (2013) Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *Plant J.* 75, 351–363
- 23 Gutensohn, M. *et al.* (2014) Metabolic engineering of monoterpene biosynthesis in tomato fruits via introduction of the non-canonical substrate neryl diphosphate. *Metab. Eng.* 24, 107–116
- 24 Rossi, L. *et al.* (2017) Overexpression of *Populus × canescens* isoprene synthase gene in *Camelina sativa* leads to alterations in its growth and metabolism. *J. Plant Physiol.* 215, 122–131

- 25 Nicholass, F.J. *et al.* (1995) High levels of ripening-specific reporter gene expression directed by tomato fruit polygalacturonase gene-flanking regions. *Plant. Mol. Biol.* 28, 423–435
- 26 Hu, H. *et al.* (2018) Modification of chrysanthemum odour and taste with chrysanthemol synthase induces strong dual resistance against cotton aphids. *Plant Biotechnol. J.* 16, 1434–1445
- 27 Schillmiller, A.L. *et al.* (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci.* 106, 10865–10870
- 28 Sallaud, C. *et al.* (2009) A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z -Farnesyl Pyrophosphate in the Wild Tomato *Solanum habrochaites*. *Plant Cell* 21, 301–317
- 29 Tian, D. *et al.* (2012) Role of trichomes in defense against herbivores: comparison of herbivore response to woolly and hairless trichome mutants in tomato (*Solanum lycopersicum*). *Planta* 236, 1053–1066
- 30 Schillmiller, A.L. *et al.* (2010) Studies of a Biochemical Factory: Tomato Trichome Deep Expressed Sequence Tag Sequencing and Proteomics. *Plant Physiol.* 153, 1212–1223
- 31 Falara, V. *et al.* (2011) The Tomato Terpene Synthase Gene Family. *Plant Physiol.* 157, 770–789
- 32 Engler, C. *et al.* (2014) A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synth. Biol.* 3, 839–843
- 33 Kimura, S. and Sinha, N. (2008) Crossing Tomato Plants. *Cold Spring Harb. Protoc.* 2008,
- 34 Xu, H. *et al.* (2013) Characterization of the Formation of Branched Short-Chain Fatty Acid:CoAs for Bitter Acid Biosynthesis in Hop Glandular Trichomes. *Mol. Plant* 6, 1301–1317
- 35 Gillasp, G. *et al.* (1993) Fruits - a Developmental Perspective. *Plant Cell* 5, 1439–1451
- 36 Li, W. *et al.* (2019) Pyrethrin Biosynthesis: The Cytochrome P450 Oxidoreductase CYP82Q3 Converts Jasmolone To Pyrethrolone. *Plant Physiol.* 181, 934–944
- 37 Li, W. *et al.* (2018) Jasmone Hydroxylase, a Key Enzyme in the Synthesis of the Alcohol Moiety of Pyrethrin Insecticides. *Plant Physiol.* 177, 1498–1509
- 38 Henry, L.K. *et al.* (2018) Contribution of isopentenyl phosphate to plant terpenoid metabolism. *Nat. Plants* 4, 721–729

- 39 Zhou, F. and Pichersky, E. (2020) The complete functional characterisation of the terpene synthase family in tomato. *New Phytol.* 226, 1341–1360
- 40 Lunn, D. *et al.* (2019) Tri-Hydroxy-Triacylglycerol Is Efficiently Produced by Position-Specific Castor Acyltransferases. *Plant Physiol.* 179, 1050–1063
- 41 Paine, J.A. *et al.* (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* 23, 482–487
- 42 Elnaiem, D.-E.A. *et al.* (2008) Impact of aerial spraying of pyrethrin insecticide on *Culex pipiens* and *Culex tarsalis* (Diptera : Culicidae) abundance and West Nile virus infection rates in an urban/suburban area of Sacramento County, California. *J. Med. Entomol.* 45, 751–757
- 43 Duchon, S. *et al.* (2009) Pyrethrum: A Mixture of Natural Pyrethrins Has Potential for Malaria Vector Control. *J. Med. Entomol.* 46, 516–522
- 44 Yang, T. *et al.* (2012) Pyrethrins Protect Pyrethrum Leaves Against Attack by Western Flower Thrips, *Frankliniella occidentalis*. *J. Chem. Ecol.* 38, 370–377
- 45 Flores, M.F. *et al.* (2015) Monitoring *Pseudococcus calceolariae* (Hemiptera: Pseudococcidae) in Fruit Crops Using Pheromone-Baited Traps. *J. Econ. Entomol.* 108, 2397–2406
- 46 Bergmann, J. *et al.* (2019) Synthesis of citrophilus mealybug sex pheromone using chrysanthemol extracted from Pyrethrum (*Tanacetum cinerariifolium*). *Nat. Prod. Res.* 33, 303–308
- 47 Ho, H.-Y. *et al.* (2009) Identification and Synthesis of the Sex Pheromone of the Madeira Mealybug, *Phenacoccus madeirensis* Green. *J. Chem. Ecol.* 35, 724–732
- 48 Lewinsohn, E. *et al.* (2001) Enhanced Levels of the Aroma and Flavor Compound S-Linalool by Metabolic Engineering of the Terpenoid Pathway in Tomato Fruits. *Plant Physiol.* 127, 1256–1265
- 49 Davidovich-Rikanati, R. *et al.* (2007) Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. *Nat. Biotechnol.* 25, 899–901

CHAPTER 3 – ACYLGLUCOSE BIOSYNTHESIS IN *SOLANUM PENNELLII*

This chapter was adapted from the following previously published manuscript ('*' denotes equal contribution):

Leong, B.J.* , Lybrand, D.B.* , Fan, P., Lou, Y., Schillmiller, A.L., and Last, R.L. (2019) Evolution of metabolic novelty: a trichome-expressed invertase creates specialized metabolic diversity in wild tomato. *Sci. Adv.* 5, eaaw3754. DOI: 10.1126/sciadv.aaw3754

ACYLSUGAR BIOSYNTHESIS IN SOLANACEAE

Glandular trichome-synthesized acylated sugars (“acylsugars”) are structurally diverse specialized metabolites found throughout the Solanaceae [1–7]. These compounds have documented roles in direct and indirect protection against herbivores and microbes [8,9], as well as allelopathic properties [9,10]. Their low toxicity to vertebrates sparks interest in plant breeding strategies for deploying acylsugars in crop protection [11,12]. These metabolites consist of a sugar core—typically sucrose—with aliphatic chains of variable length, structure, and number attached by ester linkages. Acylsugars were reported from genera across the Solanaceae family, including *Datura*, *Nicotiana*, *Petunia*, *Physalis*, *Salpiglossis*, and *Solanum* with single species producing at least three dozen chromatographically distinct acylsugars [6,7,10,13–16].

In recent years, several evolutionarily related enzymes were implicated in the core acylsucrose biosynthetic pathways in species across the family, including the cultivated tomato *Solanum lycopersicum*, *Petunia axillaris*, and *Salpiglossis sinuata* [1,3–5,7,17]. These biosynthetic pathways consist of trichome-expressed BAHD-family (BEAT, AHCT, HCBT, and DAT) acylsugar acyltransferases (ASATs), which sequentially transfer acyl groups from acyl-coenzyme A (acyl-CoA) substrates to specific hydroxyl groups of sucrose [1,3,17].

The cultivated tomato biosynthetic network is well characterized, with four ASATs—SIASAT1 to SIASAT4—catalyzing consecutive reactions to produce tri- and tetra-acylated sucroses. SIASAT1 acts first by transferring an acyl chain to the R4 hydroxyl of the pyranose ring of sucrose, and SIASAT2 transfers an acyl chain to the R3 position of the monoacylated sucrose [17]. Next, SIASAT3 acylates the diacylated sucrose at the furanose ring R3' position [3].

SIASAT4 completes the pathway by transferring an acetyl group to the pyranose ring R2 position of a triacylsucrose [1]. Enzyme promiscuity and the presence of an array of acyl-CoAs result in the production of a diverse group of acylsucroses in *S. lycopersicum* [3,18].

Acylsugar diversity is even greater in the broader *Solanum* genus. *Solanum pennellii* LA0716, a wild relative of tomato, is a prime example and produces a mixture of acylsucroses that are distinct from those found in *S. lycopersicum*. While *S. lycopersicum* accumulates acylsucroses with two or three acylations on the pyranose ring and a single acylation on the furanose ring (termed “F-type” acylsucroses), *S. pennellii* accumulates distinct triacylsucroses acylated only at the pyranose R2, R3, and R4 positions (termed “P-type”) acylsucroses [4]. 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 [4].

S. pennellii LA0716 has other acylsugar characteristics that differentiate it from cultivated tomato. First, it produces copious amounts of acylsugars that render the plant sticky, representing up to ~20% of leaf dry weight [13,19]. Second, the vast majority of *S. pennellii* LA0716 acylsugars are glucose molecules with three acyl chains (termed “acylglucoses”), while only 7 to 16% of total acylsugars are acylsucroses [20]. In contrast to the well-characterized *S. pennellii* acylsucrose biosynthetic enzymes [3,4], no complete acylglucose metabolic pathway has yet been described. This is despite the fact that acylglucoses were also characterized in several additional Solanaceae species [10,14]. 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 [21]. 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*-diisobutyryl-D-glucose and free glucose [22,23]. 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*-triacylglucoses observed in *S. pennellii* in both the number (two instead of three) and position of acyl chains: *S. pennellii* acylglucoses bear chains at the R2, R3, and R4 positions rather than at the R1 position [13].

In contrast to the unsubstantiated published biosynthetic pathway, 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 acylsugars comprising up to 89% acylglucoses [24]. The presence of QTLs on both chromosomes yields detectable acylglucoses, while addition of the chromosome 4 locus leads to elevated accumulation. Notably, chromosome 4 and 11 QTLs include the *SpASAT2* and *SpASAT3* genes, respectively, responsible for accumulation of P-type acylsucroses in *S. pennellii* [3,4]. This prompted us to hypothesize that the triacylsucroses synthesized by ASATs are substrates for a chromosome 3 factor that hydrolyzes them, yielding triacylglucoses.

We report the characterization of the plant specialized metabolic invertase-like enzyme acylsucrose fructofuranosidase (*SpASFF1*; Sopen03g040490), a chromosome 3 β -fructofuranosidase capable of cleaving the glycosidic bond of P-type acylsucroses. Genetic and transgenic plant approaches demonstrate that *S. pennellii* LA0716 acylglucose production

requires *SpASFF1*. This work also corroborates the previously reported three-gene epistatic interaction between loci on chromosomes 3, 4, and 11 containing the *SpASFF1*, *SpASAT2*, and *SpASAT3* genes, respectively, that conditions high-level acylglucose accumulation [24]. While yeast invertase and other variants involved in core metabolism have been studied since the 19th century [25–27], this work documents a new type of role for β -fructofuranosidase-type enzymes in specialized metabolism. These results extend our understanding of evolutionary mechanisms leading to trichome specialized metabolic diversity by demonstrating how neofunctionalization led to co-option of invertase from general metabolism into a cell type-specific specialized metabolic network.

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 D. Zamir (Hebrew University of Jerusalem, Rehovot, Israel) [28]; seeds of *S. pennellii* LA0716 were provided by M. Mutschler (Cornell University, Ithaca, NY). Seeds were treated with half-strength bleach for 30 min and rinsed three times in de-ionized water for 5 min before sowing on moist filter paper in petri dishes. Upon germination, seedlings were transferred to soil. Young plants were grown in 9-cm pots in a 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 half-strength Hoagland's solution; *S. pennellii* was watered once weekly with deionized water and supplemented once weekly with half-strength Hoagland's solution. Plants used for analysis were grown in a growth

chamber under a 16-hour photoperiod [$190 \mu\text{mol m}^{-2} \text{s}^{-1}$ 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-hour light/8-hour dark cycle. For seed production, *S. pennellii asff1* T₀ 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 of crushed oyster shell (Down to Earth Distributors Inc., Eugene, OR), 1.8 kg of hydrated lime (Bonide Products, Inc., Oriskany, NY), and 0.6 kg of triple super phosphate (T and N Inc., Foristell, MO) per cubic meter.

Acylsugar analysis

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 a high-performance liquid chromatography (HPLC) standard. The leaf tissue was gently agitated on a rocker in this extraction solvent for 2 min. 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 an Acquity UPLC coupled to a Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA). Five microliters of the acylsugar extracts were injected onto an Ascentis Express C18 HPLC column

(100 mm 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. Compounds were eluted using one of two gradients.

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

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

The MS settings were as follows for positive ion-mode electrospray ionization (ESI⁺): capillary voltage, 3.00 kV; source temperature, 100°C; desolvation temperature, 350°C; desolvation nitrogen gas flow rate, 600 liters/h; cone voltage, 35V; and mass range, m/z 50 to 1000 (spectra accumulated at 0.1 s per function). Three separate acquisition functions were set up to test different collision energies (0, 15, and 45 V). 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 of dichloromethane (VWR International, Radnor, PA) and 500 μ L of 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. Twenty-microliter aliquots of acylsugar extracts were dried in 1.5-mL microcentrifuge tube using a SpeedVac and dissolved in 100 μ L of 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 hours in a fume hood. The solvent was removed using a SpeedVac. Before LC-MS analysis, samples were dissolved in 200 μ L ammonium bicarbonate (pH 8) in 90% acetonitrile containing 0.5 μ M $^{13}\text{C}_{12}$ -sucrose and 0.5 μ M $^{13}\text{C}_6$ -glucose as internal standards and transferred to 2-mL LC-MS vials. Compounds were analyzed using an Acquity UPLC coupled to an Acquity TQD Tandem Quadrupole mass spectrometer (Waters, Milford, MA). Five microliters of the acylsugar extracts were injected onto an Acquity BEH amide column (100 x 1.7 mm, 1.7 μ M) in a column oven with temperature of 40°C and with a flow rate of 0.5 mL/min. The LC-MS method 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 as follows: 100% B at 0 min, 0% B at 5 min, 100% B at 5.01 min, and held at 100% B until 10 min. 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 $^{13}\text{C}_6$ -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 $^{13}\text{C}_{12}$ -sucrose: precursor ion, m/z 353; product ion, m/z 92; cone

voltage, 40 V; collision energy, 22 V. Quantification of glucose and sucrose was 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 μ M in water were prepared and processed using the same protocol for acylsugars, as described 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 μ m; ThermoFisher Scientific, Waltham, MA) with a column oven temperature of 30°C and flow rate of 1 mL/min. 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 min 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- μ L 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 μ L each using a linear elution gradient of 1% B at 0 min, 63% B at 10 min, 65% B at 30 min, 100% B at 35 min brought back to 1% B at 35.01 min and held at 1% B until 40 min. Eluted compounds were collected in 10-s fractions. Fraction collection tubes contained 333 μ L 0.1% formic acid in water, and the S3:19 product eluted at 18-19 min.

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 min and filtered through Miracloth pre-wetted with methanol 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 min, 50% B at 5 min, 70% B at 30 min, 100% B at 32 min and held at 100% until 35 min brought back to 1% B at 35.01 min and held at 1% B until 40 min. Eluted compounds were collected in 1-min fractions. Fraction collection tubes contained 333 µL 0.1% formic acid in water, and the S3:22 product eluted at 7-8 min.

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 (ThermoFisher 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 min, 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 1 min 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-1 α* , *actin*, and *ubiquitin* genes, respectively (Table S3.1 in Appendix). For each biological replicate, relative levels of ASFF1 transcript were determined using the $\Delta\Delta C_t$ method [29] and normalized to the geometric mean of EF-1 α , 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 S3.1 in Appendix).

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 S3.2 in Appendix). The first and fourth amplicon contained adapters for assembly into pENTR-D-TOPO that has been digested with NotI/Ascl 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 [30] using LR clonase II enzyme mix (ThermoFisher 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 S3.1 in Appendix).

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 S3.1 in Appendix). pENTR-D-TOPO was digested with NotI/Ascl 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/Ascl. 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 [30]. The 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 [31], and low off-target scores against published *S. pennellii* genome database [32]. Each sgRNA was obtained as a gBlock synthesized *in vitro* by Integrated DNA Technologies (www.idtdna.com) (Table S3.1 in Appendix) and subsequently assembled with

pICH47742::2x35S-5'UTR-hCas9(STOP)-NOST [Addgene #49771, provided by S. Kamoun (Sainsbury Lab, Norwich, UK)] [33], pICH41780 (Addgene plasmid # 48019) and pAGM4723 (Addgene plasmid # 48015, both gifts from S. Marillonnet) [34] and pICSL11024 [Addgene plasmid #51144, a gift from J. D. Jones (Sainsbury Lab, Norwich, UK)] using Golden Gate Assembly. In short, the restriction-ligation reactions (20 μ L) were set up by mixing 15 ng of synthesized sgRNAs with 1.5 μ L T4 ligase buffer (NEB), 320 U of T4 DNA ligase (NEB), 1.5 μ L BSA (0.1 mg/mL, NEB), 8 U of *Bpi*I (ThermoFisher Scientific, Waltham, MA) and 100–200 ng of the intact plasmids. The reactions were incubated at 37°C for 30 s, followed by 26 cycles (37°C, 3 min; 16°C, 4 min) and then incubated at 50°C for 5 min and 65°C for 5 min. 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, 80 pICH47742_SeqF1, pICH41780_SeqR1, and ASFF_SeqR primers (Table S3.1 in Appendix). 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 S3.1 in Appendix).

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 Scientific, Waltham, MA). 1 μ g 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 μ M), 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.

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 of yeast extract, 10 g of Bacto peptone, and 5 g of NaCl per liter (pH 7)] with Rifampicin (50 μ g/mL). Cultures were incubated in borosilicate glass test tubes (18 x 150 mm) with foam plugs overnight at 30°C, shaken at 200 rpm. One hundred ninety 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 OD600 = 1.0, incubated on ice for 10 min, and centrifuged at 4°C in 50-mL conical tubes at 3,200g for 5 min. Pellets were resuspended in 1 mL of sterile 20 mM CaCl₂ and 100 μ L 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 μ g of construct DNA purified using an Omega EZNA plasmid DNA mini kit I (Omega Bio-Tek, Norcross, GA) was added to the frozen *Agrobacterium* aliquots on ice. Cells were thawed in a 37°C water bath for 5 min, 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 μ L of fresh YEP. The cell pellet was resuspended and the entire suspension was plated onto an LB plate with spectinomycin (100 μ g/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. Two microliters of the cell suspension was added to a PCR tube with a final reaction 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 dishes 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 [35,36]. Fifty to sixty seeds were incubated in 40% bleach, agitating for 5 min. Seeds were rinsed six times, each with 40 mL of sterile ddH₂O with 5 min 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 [35] in a PhytaTray II (Sigma-Aldrich, St. Louis, MO). Containers were incubated at 25°C on a 16-hour light/8-hour dark cycle with a light intensity of 70 $\mu\text{mol m}^{-2}\text{s}^{-1}$ 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. One to two millimeters were 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 [35] 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 cocultivation, 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 min at 20°C. The supernatant fluid was decanted and 10 mL of MSO media [35] was added to the tube (with no pellet resuspension). The cell pellet was centrifuged at 2,000*g* for 5 min and this washing step was then repeated. The cell pellet was resuspended 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 μ M and 5 mL of the *Agrobacterium* suspension pipetted onto the cotyledons on the plate and incubated with swirling at room temperature for 10 min, 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-hour light/8-hour dark cycle at 70 μ mol m⁻²s⁻¹ PPFD.

For transgenic callus selection, two days after cocultivation, the cotyledons were transferred directly onto sterile 2Z media plates [36] containing 100 μ g/mL kanamycin and 200 μ g/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 [36] 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-hour light/8-hour dark cycle at 70 µmol m⁻²s⁻¹ PPF). Shoots were monitored for leaf and root production, and shoots with roots and leaves were placed into pots containing Redi-Earth 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 S3.1 in Appendix) 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 S3.1 in Appendix) to add adapters for Gibson assembly. The resulting PCR product was transferred to pEAQ-HT vector [37] 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 *Agrobacterium 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 shaken at 300 rpm overnight. Cultures were harvested by centrifugation at 800g and 20°C for 20 min. 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) at pH 5.6 and 10 mM MgCl₂]. This cell suspension was centrifuged at 800g and 20°C for 20 min and the resulting pellet resuspended to a final OD₆₀₀ = 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 hours. This suspension was infiltrated into fully expanded leaves of six-week-old *Nicotiana benthamiana* plants using a needleless 1-mL tuberculin syringe. Plants were grown under 16-hour photoperiod (70 µmol m⁻²s⁻¹ PPF) and constant 22°C set to 70% relative humidity. At 8 days after infiltration, 28 g of infiltrated leaves were harvested, deveined, and flash-frozen in liquid nitrogen. Tissue was powdered under liquid nitrogen with a 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) at pH 8.0, 1.5 M NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA) with 2 mM dithiothreitol (DTT), 1 mM benzamidine, 0.1 mM phenylmethanesulfonylfluoride (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, which was obtained from Roche Diagnostics (Risch-Rotkreuz, Switzerland)]. The mixture was stirred for 4 hours at 4°C, filtered through six layers of Miracloth and centrifuged at 27,000g, 4°C for 30 min. The supernatant was

decanted and passed through a 0.22- μ m 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 were analyzed by SDS-polyacrylamide gel electrophoresis and the presence of ASFF1-HT was 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) with 50% glycerol using a 10DG desalting column (Bio-Rad, Hercules, CA). Protein was quantified against a standard curve of BSA (ThermoFisherScientific, Waltham, MA) using a modified Bradford reagent (Bio-Rad, Hercules, CA) according to the manufacturer's instructions.

Enzyme assays

For activity assays, 100 ng of ASFF1-HT or 1 μ g *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 μ L of 50 mM sodium acetate (pH 4.5) in 250- μ L thin-wall PCR tubes. Reactions were incubated for 1 hour at 30°C and stopped by addition of 60 μ L of 1:1 acetonitrile/isopropanol containing 1.5 μ M telmisartan as internal standard and centrifuged for 10 min at 16,000g to remove precipitated protein. The supernatant was transferred to 2-mL autosampler vials with 250- μ L glass inserts and analyzed by LC-MS as described above.

Statistical Analysis

All statistical analyses were performed using the "stats" R package [38]. 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. Analysis by Tukey's post hoc 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 was determined using the "power.t.test" function.

RESULTS

An S. pennellii chromosome 3 locus is necessary for acylglucose production from P-type acylsucroses

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 [24,39]. 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 [40] yielded detectable leaf acylglucoses (Fig. S3.1 in Appendix). 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 Appendix) in backcross inbred line BIL6521 [28], which contains *S. pennellii* LA0716 introgressions from chromosomes 1, 3, and 11. This BIL accumulates four acylglucoses (Table S3.2 in Appendix), with the major one, G3:22 (5,5,12) (Fig. S3.3 and Fig. S3.4 in Appendix), 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 [3]. In fact, BIL6521 accumulates a P-type acylsucrose, S3:22 (Fig. S3.8 and Fig. S3.9 in

Appendix). 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' indicate a sucrose or glucose core, respectively, and 3:22 (5,5,12) indicates that there are three ester-linked acyl chains, two of 5 carbons and one of 12 carbons, for a total of 22 chain carbons [3]. 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 lacks 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. We crossed BIL6521 with BIL6180, a recombinant line harboring introgressions on chromosomes 4, 5 and 11, including both the *SpASAT2* and *SpASAT3* loci (Fig. 3.1C). BIL6180 was previously found to produce only P-type acylsucroses as a result of the chromosome 4 and 11 introgressions; however, it accumulated significantly higher overall levels of acylsucroses compared to BIL6521 (Fig. S3.1C in Appendix) 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. S3.8 and Fig. S3.9 in Appendix). These findings – in combination with the published QTL results – indicate that the *S. pennellii* chromosome 3 introgression is

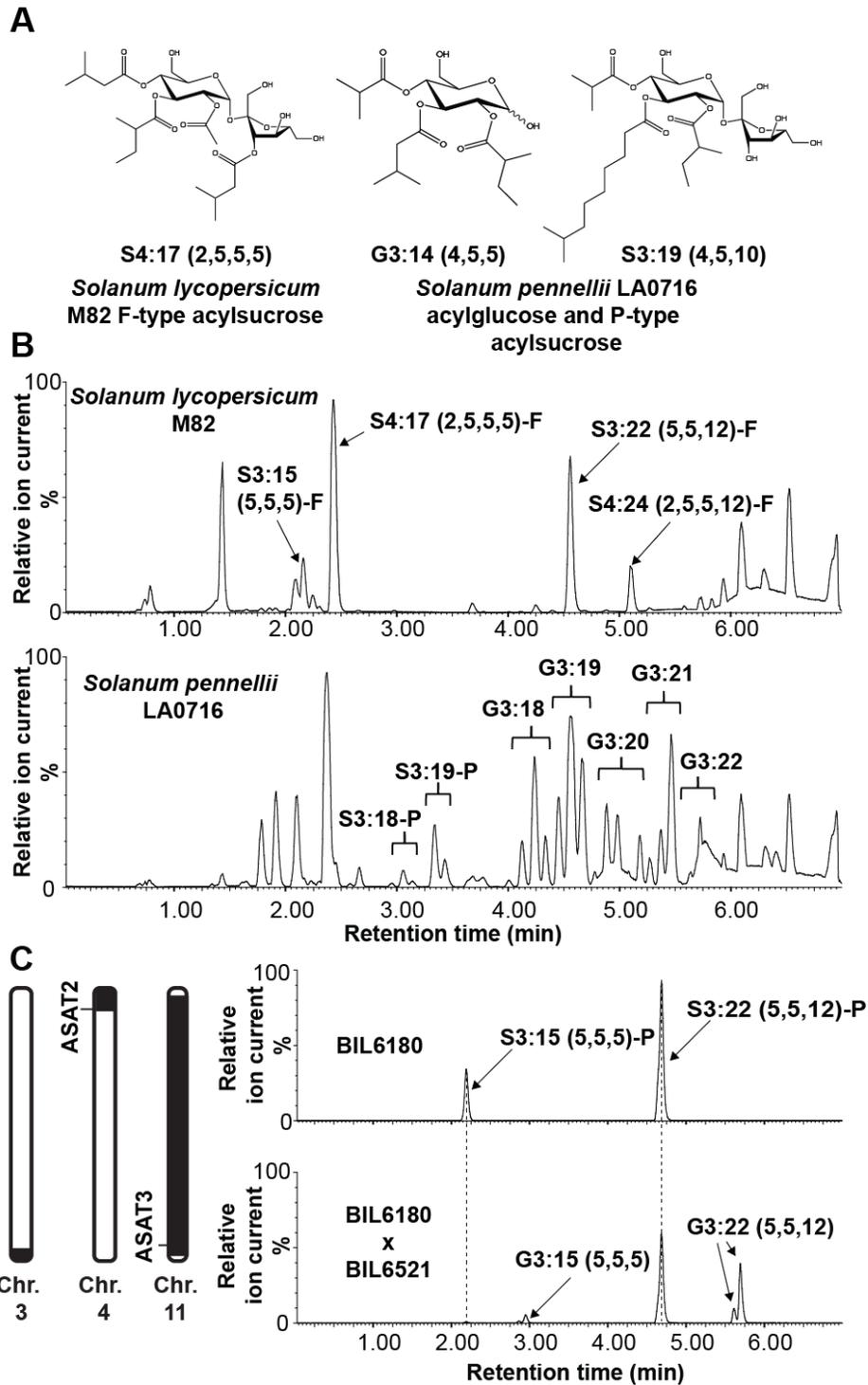


Figure 3.1 Three *S. pennellii* LA0716 regions condition acylglucose accumulation.

Figure 3.1 (cont'd) (A) Examples of NMR-resolved *S. lycopersicum* and *S. pennellii* acylsugar structures. Acylsugars from *S. lycopersicum* are composed of sucrose acylated on both the pyranose and furanose rings ('F-type'). *S. pennellii* acylsugars are a mixture of sucrose ('P-type') and glucose-based compounds with acylation exclusively on the pyranose ring. (B) Acylsugar ESI- mode LC-QToF-MS profiles. Top: *S. lycopersicum* M82 with acylsucroses S3:15 (5,5,5)-F, S4:17 (2,5,5,5)-F, S3:22 (5,5,12)-F, and S4:24 (2,5,5,12)-F annotated. Bottom: *S. pennellii* LA0716 acylsucroses and acylglucoses. (C) Left: Representation of *S. pennellii* chromosomal introgressions in BIL6521 x BIL6180 progeny that contain QTLs affecting acylglucose biosynthesis (30). The black portions of the chromosomes correspond to *S. pennellii* introgressions, while the white portions correspond to the chromosomal regions in the M82 background. ESI- mode LC-MS analysis of BIL6180 compared with the BIL6180 x BIL6521 F2 progeny reveals acylglucose accumulation in the hybrid, but not in BIL6180. All ESI- mode acylsugars were identified as formate adducts.

necessary 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 (Fig. 3.2A) [24]. Three of the 238 genes in this region of the *S. lycopersicum* Heinz 1706 genome assembly SL2.50 annotation [41] are predicted as encoding glycoside hydrolases (members of the GH32, GH35, and GH47 families; Table S3.3 in Appendix). We focused on the GH32 family *Sopen03g040490* gene because all previously characterized members of the family have β -fructofuranosidase or fructosyltransferase activity [42]. 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 *ACYLSUCROSE FRUCTOFURANOSIDASE 1 (ASFF1)*.

S. lycopersicum acylsugars accumulate in type I/IV glandular trichome tip cells [43] and trichome tip cell-specific gene expression is a hallmark of all characterized acylsugar biosynthetic genes (*e.g.*, *ASAT1/2/3/4*, *IPMS3*) [1,3,17,18]. 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. 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. 3.2B). This result is consistent

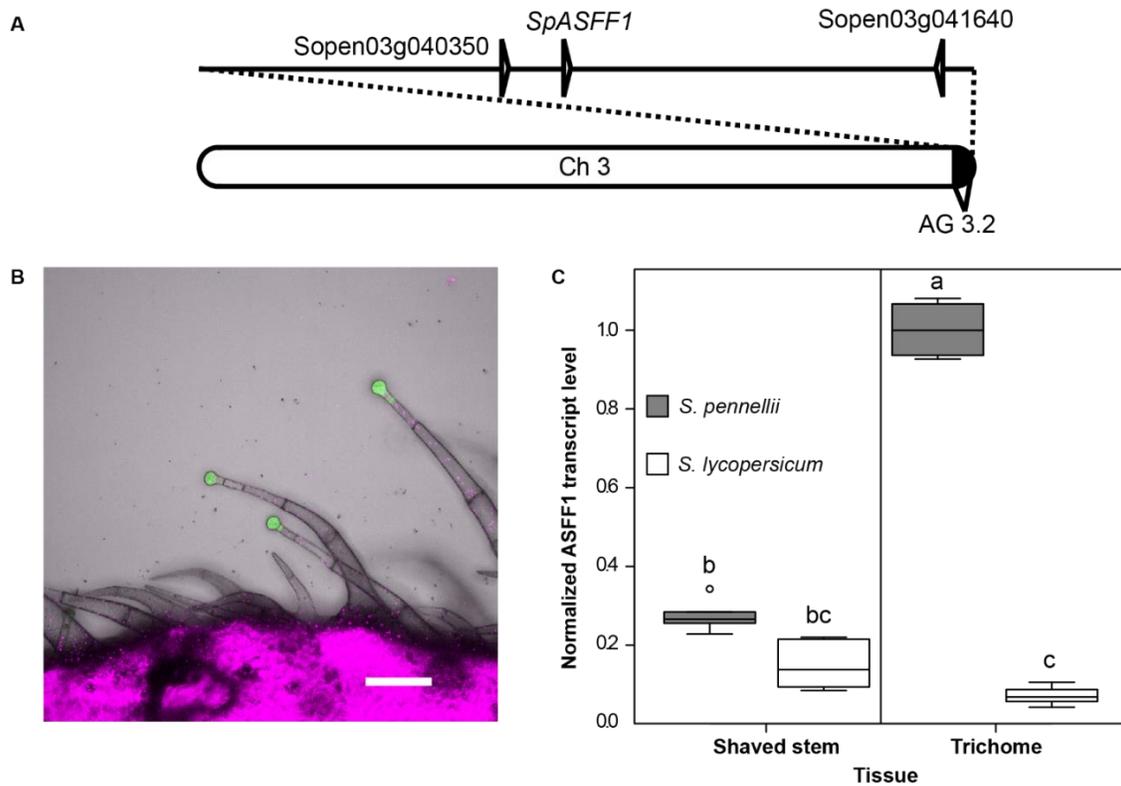


Figure 3.2 Glycoside hydrolase 32 family gene *SpASFF1* from QTL AG 3.2 shows trichome-specific expression. (A) Chromosome 3 with the AG3.2 introgression. Positions of three glycoside hydrolase candidate genes are indicated. (B) Expression of GFP-GUS under control of the native *ASFF1* promoter from *S. pennellii* LA0716 yields GFP signal in *S. lycopersicum* M82 type I/IV trichome tip cells, but not stalk cells or stem tissue. Green channel indicates GFP signal; magenta channel shows chlorophyll fluorescence. Scale bar = 100 μ m. (C) qRT-PCR analysis of *ASFF1* transcripts shows significantly higher levels in *S. pennellii* LA0716 trichomes than in underlying stem tissue or trichomes of *S. lycopersicum* M82. Treatments that do not share a letter are significantly different from one another ($p < 0.001$; one-way ANOVA, Tukey's post-hoc 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 species and tissue types.

with a role of the 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 transcription PCR (qRT-PCR). *ASFF1* transcript levels were 3.7-fold higher in trichomes of *S. pennellii* LA0716 stems than in underlying shaved stem tissue ($p < 0.001$; one-way ANOVA, Tukey's post hoc mean-separation test) (Fig. 3.2C). The observed enrichment of transcripts in trichome samples is similar to analysis of previously identified acylsugar biosynthetic genes from tomato, petunia and tobacco [1,5,9,17,18]. 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 [2,13]. However, *ASFF1* is predicted to encode a full open reading frame in both the *S. lycopersicum* Heinz 1706 and the *S. pennellii* LA0716 genomes [41,44]. While *ASFF1* transcripts are enriched in trichomes of *S. pennellii* LA0716, no significant difference was observed between *ASFF1* transcript levels in *S. lycopersicum* M82 stems and trichomes ($p = 0.063$; one-way ANOVA, Tukey's post hoc mean-separation test) (Fig. 3.2C). Additionally, we found that *ASFF1* transcripts are enriched 14-fold in *S. pennellii* LA0716 trichomes relative to transcripts in trichomes of *S. lycopersicum* M82 ($p < 0.001$, one-way ANOVA, Tukey's post-hoc mean-separation test) (Fig. 3.2C) 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 guide RNAs (sgRNAs) targeting the third *SpASFF1* exon were used to promote site-specific DNA cleavage by hCas9 in stably transformed plants (Fig. 3.3A; Fig. S3.5 in Appendix). Three homozygous T1 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 T₀ plant. The third mutant (*spasff1-2*) with a 1 bp insertion is the descendant of a homozygous T₀ 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. 3.3B, C; Fig. S3.6 in Appendix), but produced acylsucroses at levels comparable to total acylsugars in wild-type *S. pennellii* plants (Table 3.1). *SpASFF1* converts 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. 3.4A; Fig. S3.7 in Appendix). We took a transgenic approach to determine 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. 3.4B) with a T-DNA containing the *SpASFF1* open reading frame and the 1.8 kb promoter region immediately upstream of its start codon (Fig. 3.4C). In addition to the P-type acylsucroses in the parental

Table 3.1 Quantitative analysis of acylsugars from *S. pennellii* LA0716 and three *spasff1* lines. Acyl chains were saponified from the acylsugars and the resulting sugar cores analyzed by UPLC-ESI-Multiple Reaction Monitoring. Data are shown from individual T₁ homozygous plants grown together but independently from those in Fig. 3.3. These extracts include other glycosylated compounds such as flavonoids, which could be responsible for the non-zero values for glucose measurements in plants lacking detectable acylglucoses.

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
<i>spasff1-1-1</i>	1	98%	2%	25.82
	2	99%	1%	23.23
	3	98%	2%	26.79
	4	99%	1%	21.49
<i>spasff1-1-2</i>	1	99%	1%	18.74
	2	99%	1%	24.55
	3	99%	1%	28.72
	4	99%	1%	18.35
<i>spasff1-2</i>	1	97%	3%	80.01
	2	98%	2%	37.51
	3	95%	5%	84.44

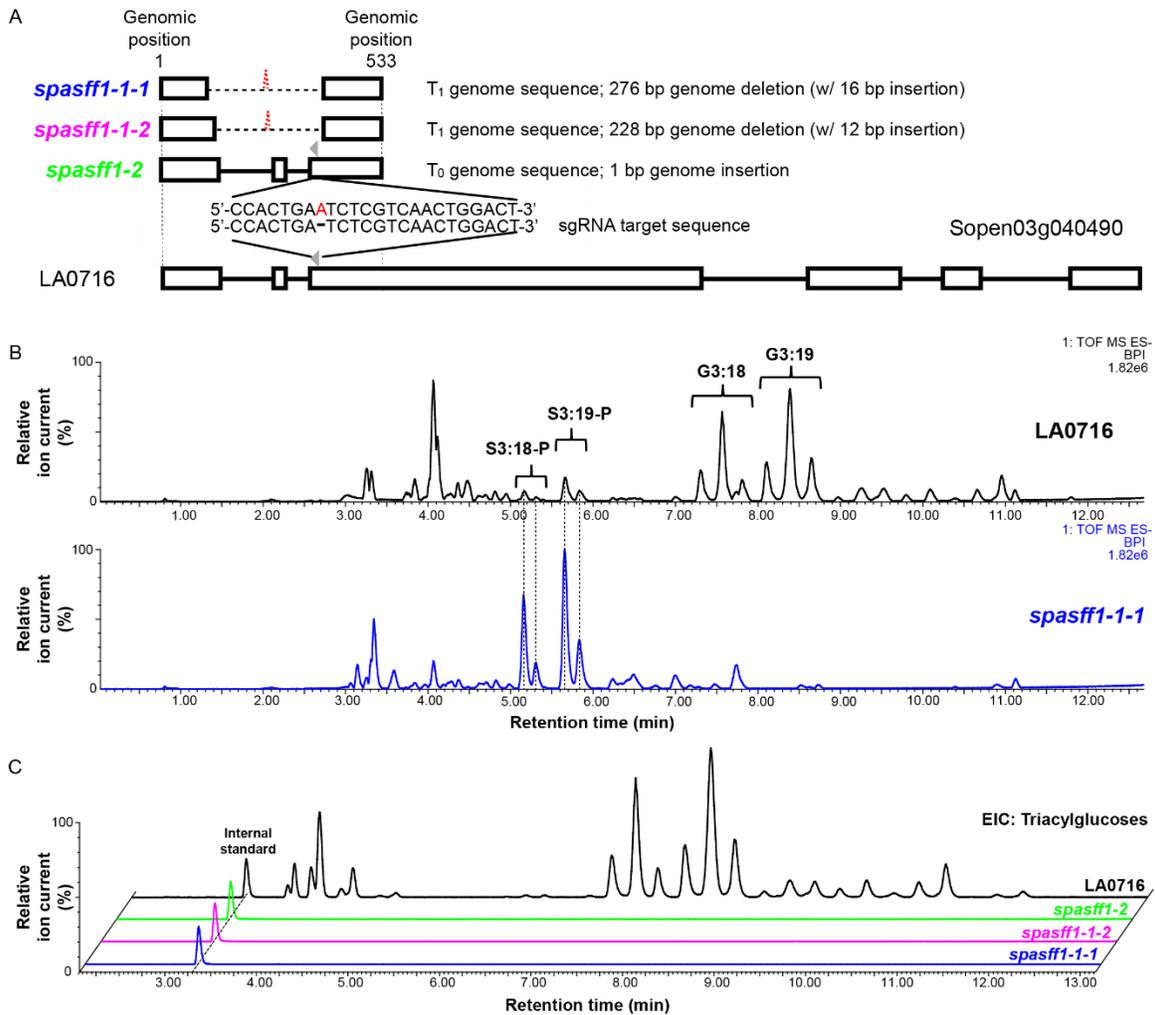


Figure 3.3 CRISPR/Cas9-mediated *S. pennellii* LA0716 *spasff1* knockouts eliminate detectable acylglucosides. (A) Schematic representation of mutagenesis strategy with two sgRNAs (grey arrowheads – only one sgRNA shown) targeting the *SpASFF1* ORF that result in three homozygous knockout lines. White boxes indicate exons, horizontal bars indicate introns, dotted lines indicate deletions and red letter indicates insertion. Mutant allele DNA sequences are found in Fig. S5. (B) Mutant line *spasff1-1-1* accumulates abundant acylsucroses but no detectable acylglucosides. ESI- base-peak intensity LC-MS chromatograms are shown for *spasff1-1-1* and LA0716. (C) ESI- mode analysis of formate adducts of triacylglycerol extracted ion chromatograms of trichome extracts from *S. pennellii* LA0716 and three *spasff1* mutant plants show that homozygous *asff1* lines produce undetectable levels of triacylglycerol. 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 (internal standard) (m/z :513.23). For panel B and C, *spasff1-1-1/1-1-2* are homozygous T₂ lines, while *spasff1-2* are homozygous T₁ lines that were all grown together. All *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 acylglucosides in ESI- mode.

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. 3.4C; Fig. S3.8 in Appendix). 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. S3.9). 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. Together, these *in vivo* results show that *SpASFF1* is sufficient to yield acylglucoses *in vivo* when P-type acylsucroses are present, but not in plants accumulating only F-type acylsucroses.

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 *N. benthamiana* transient expression system [37]. 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. 3.5A and Fig. S3.9 in Appendix). In contrast, *SpASFF1* demonstrated no hydrolytic activity with F-type S3:22 ($5^{R4}, 5^{R3'}, 12^{R3}$) [2] (Fig. 3.5B), 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 in Appendix). This *SpASFF1 in vitro* substrate specificity corroborates the *in vivo* results showing

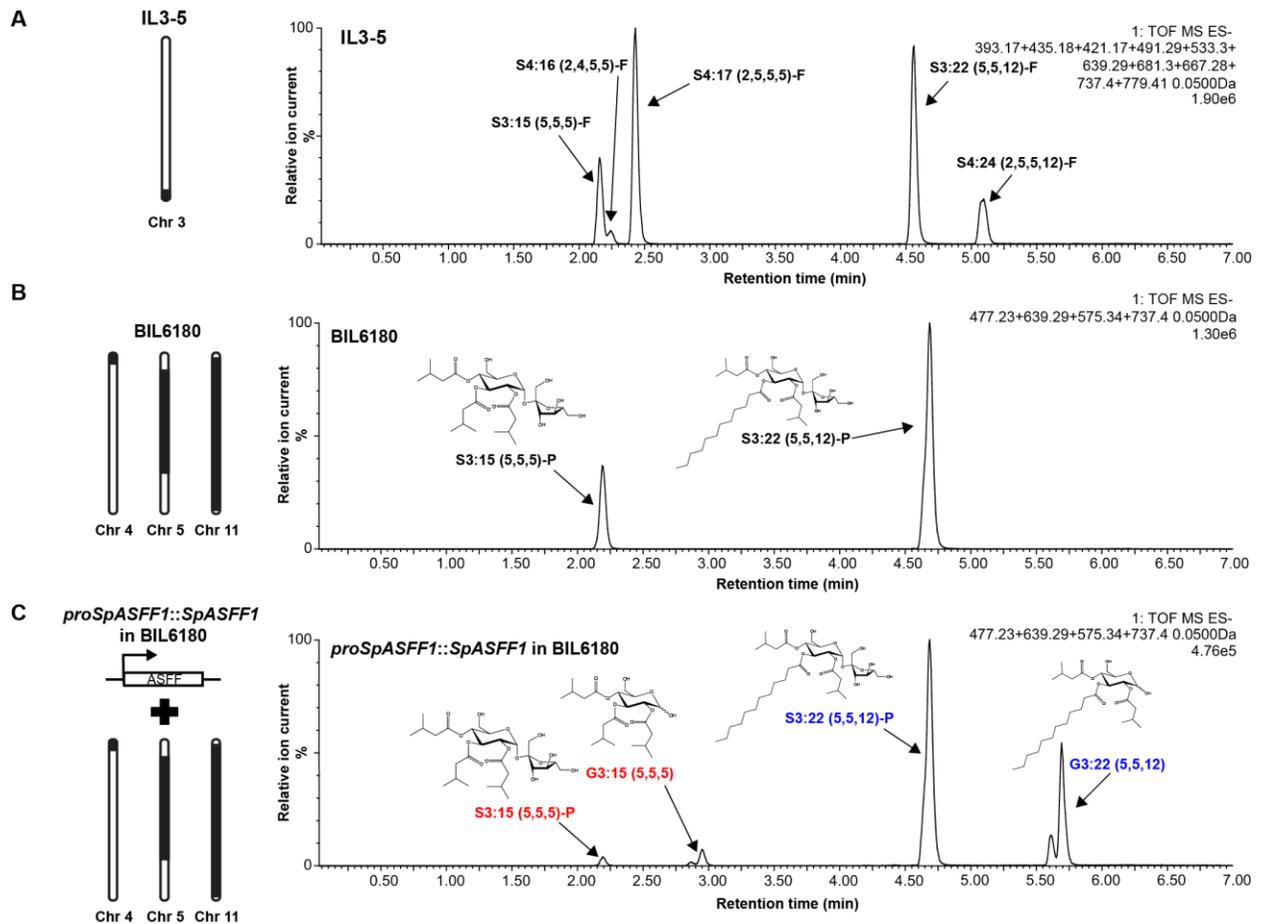


Figure 3.4 Expression of *SpASFF1* in P-type acylsucrose producing BIL6180 trichomes results in accumulation of acylglucosides in surface extracts.

(A) IL3-5 accumulates F-type acylsucroses without detectable acylglucosides. 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) in addition to 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 acylglucosides. 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 in BIL6180 is sufficient to cause accumulation of detectable G3:15 and G3:22 acylglucosides. 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. Mass window: 0.05Da in all experiments. Acylglucose structure is inferred from collision induced dissociation-mediated fragmentation (Fig S8). All ESI- mode acylsugars were identified as formate adducts.

that acylglucoses only accumulate in lines containing P-type acylsucroses.

DISCUSSION

The results described above show that *S. pennellii* LA0716 biocatalyzes acylglucoses from P-type acylsucroses via a previously uncharacterized trichome invertase (Fig. 3.5A), a homolog of the most venerable enzyme in the history of biochemistry, yeast invertase from which Michaelis and Menten derived the theory of enzyme kinetics [45], and a member of an enzyme family that has been important in the study of plant physiology since the 19th century [25]. Since that time, other general metabolic activities were identified for diverse GH32 β -fructofuranosidases, including plant glycan biosynthesis, cell wall modification, and hormone metabolism [46,47].

Our results contrast the previously proposed direct biosynthesis of acylglucoses from UDP-glucose and free fatty acids [21–23,48]. 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* [21,23,48]. The glucosyltransferases 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* [21]. 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*.

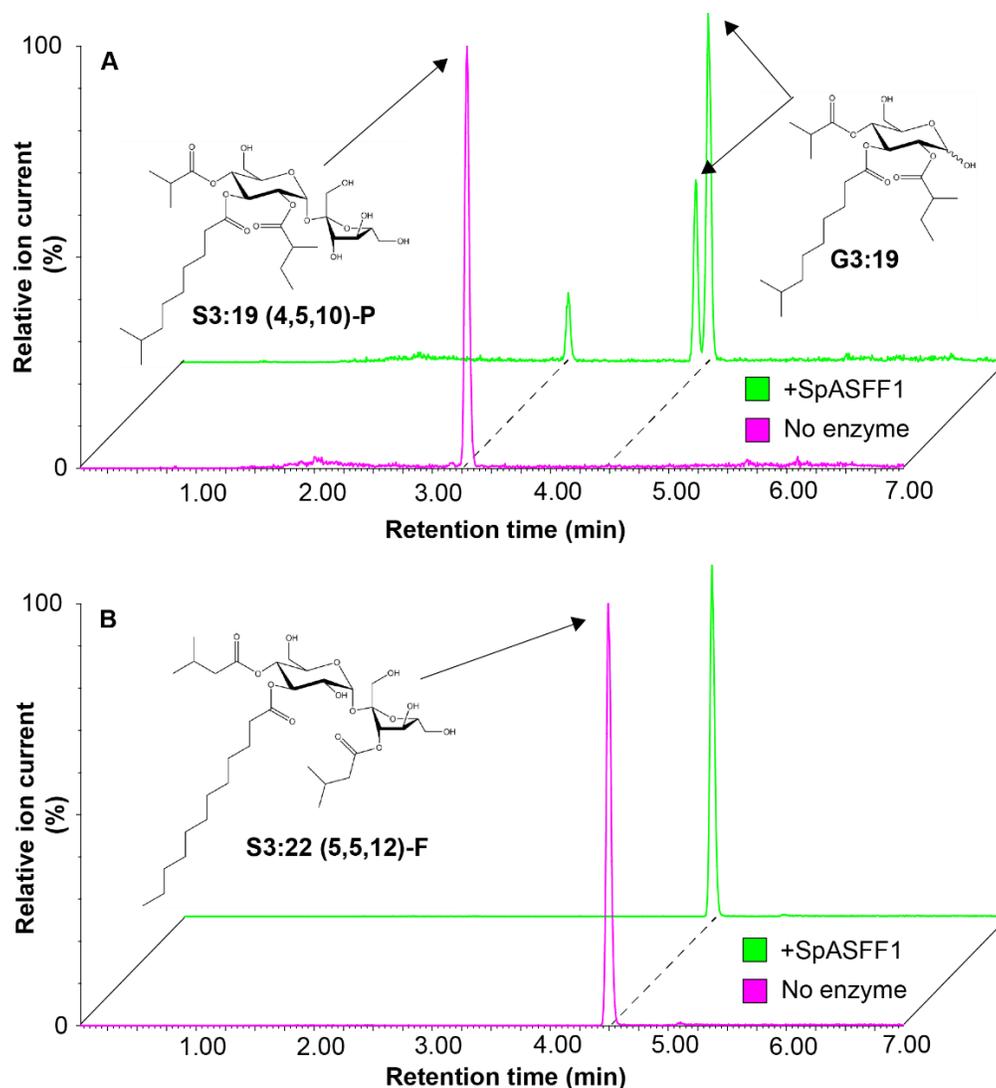


Figure 3.5 SpASFF1 cleaves a P-type S3:19 acylsucrose but not F-type S3:22 acylsucrose. (A) LC-MS analysis of *in vitro* enzyme assay products indicates that SpASFF1 hydrolyzed P-type S3:19 ($5^{R2}, 10^{R3}, 4^{R4}$) acylsucrose yielding two compounds with $m/z = 533.3$. This m/z is consistent with an acylglucose product with a G3:19 (4,5,10) configuration; the two peaks 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 indicates no hydrolysis products with SpASFF1. Acylglucose structure is inferred from collision induced-dissociation spectra (Fig. S9). All ESI- mode acylsugars were identified as formate adducts.

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 [49]. Additionally, the SCPL acyltransferase described by Li and co-workers [23] is encoded on chromosome 10 (*Solyc10g049210*) in a region not implicated in acylglucose accumulation in QTL mapping studies [24]. 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 [17,24].^[1]

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 [50] 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 S3.3 in Appendix). Most characterized plant GH35 enzymes act as β -galactosidases while GH47 enzymes function as α -mannosidases in post-translational protein modification [51,52]. 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 [42,46]. 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. 3.3). 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. 3.4A; Fig. S3.6 in Appendix). 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. 3.4). 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. 3.5A). In contrast, the enzyme was inactive against F-type S3:22 (12^{R3},5^{R4},5^{R3'}) (Fig. 3.5B) and did not hydrolyze unacylated sucrose (Fig. S3.10 in Appendix).

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. 3.2) and the ability to cleave acylated sucrose (Fig. 3.5), which led 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 [2,7,53]. 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* [2,15] but also accumulate in species representing earlier diverging lineages, including *Salpiglossis* and *Petunia* [5,7]. 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 also occur in diverse genera including *Solanum*, *Datura*, and *Nicotiana* [10,13,14]. 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* [4,53,54]. 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 [5,7,17]. 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? 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, the 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 [55,56]. 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.

APPENDIX

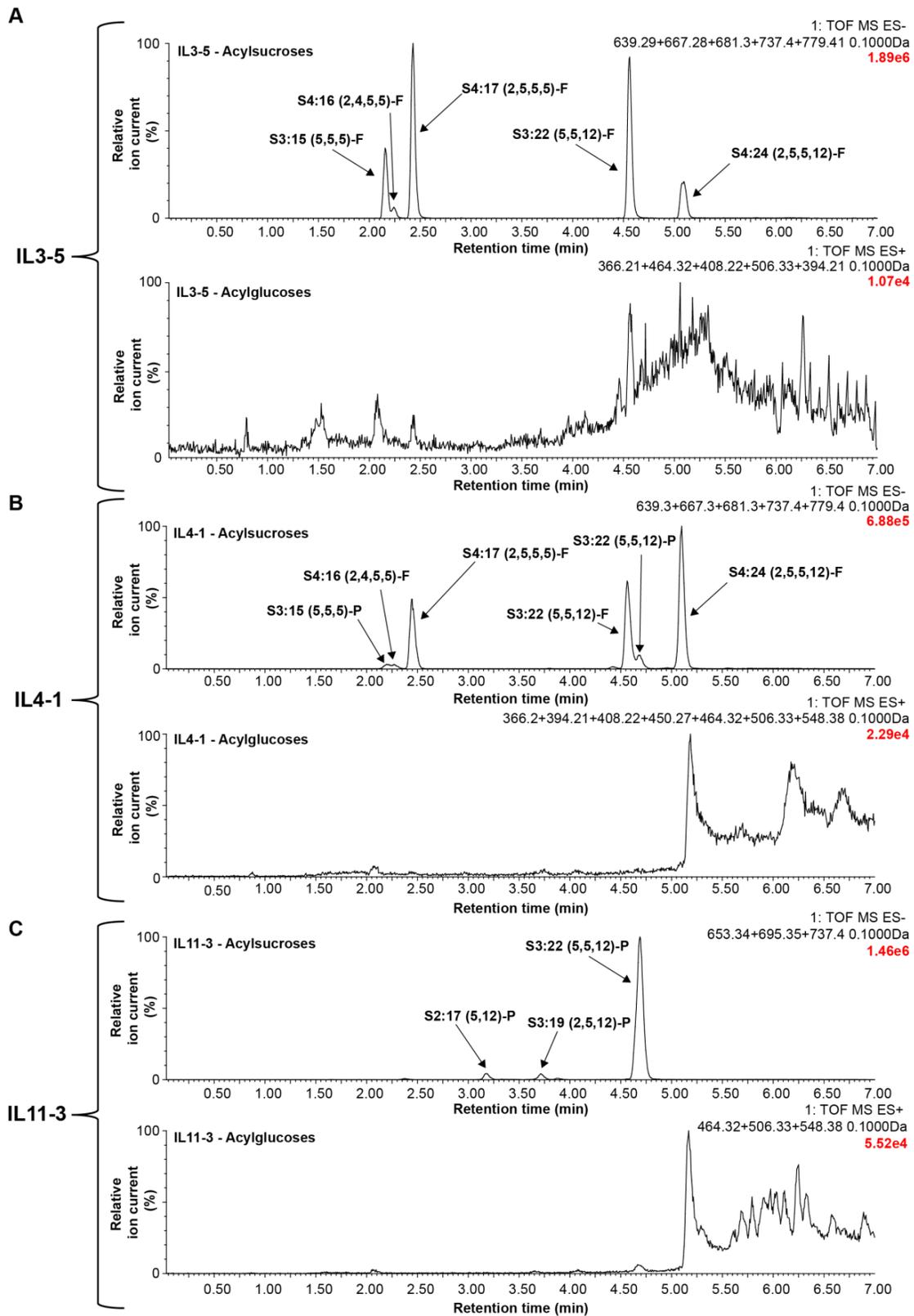


Figure S3.1 Acylglucoses are not detected in trichome extracts of IL3-5, IL4-1, or IL11-3.

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 min 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 righthand 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.

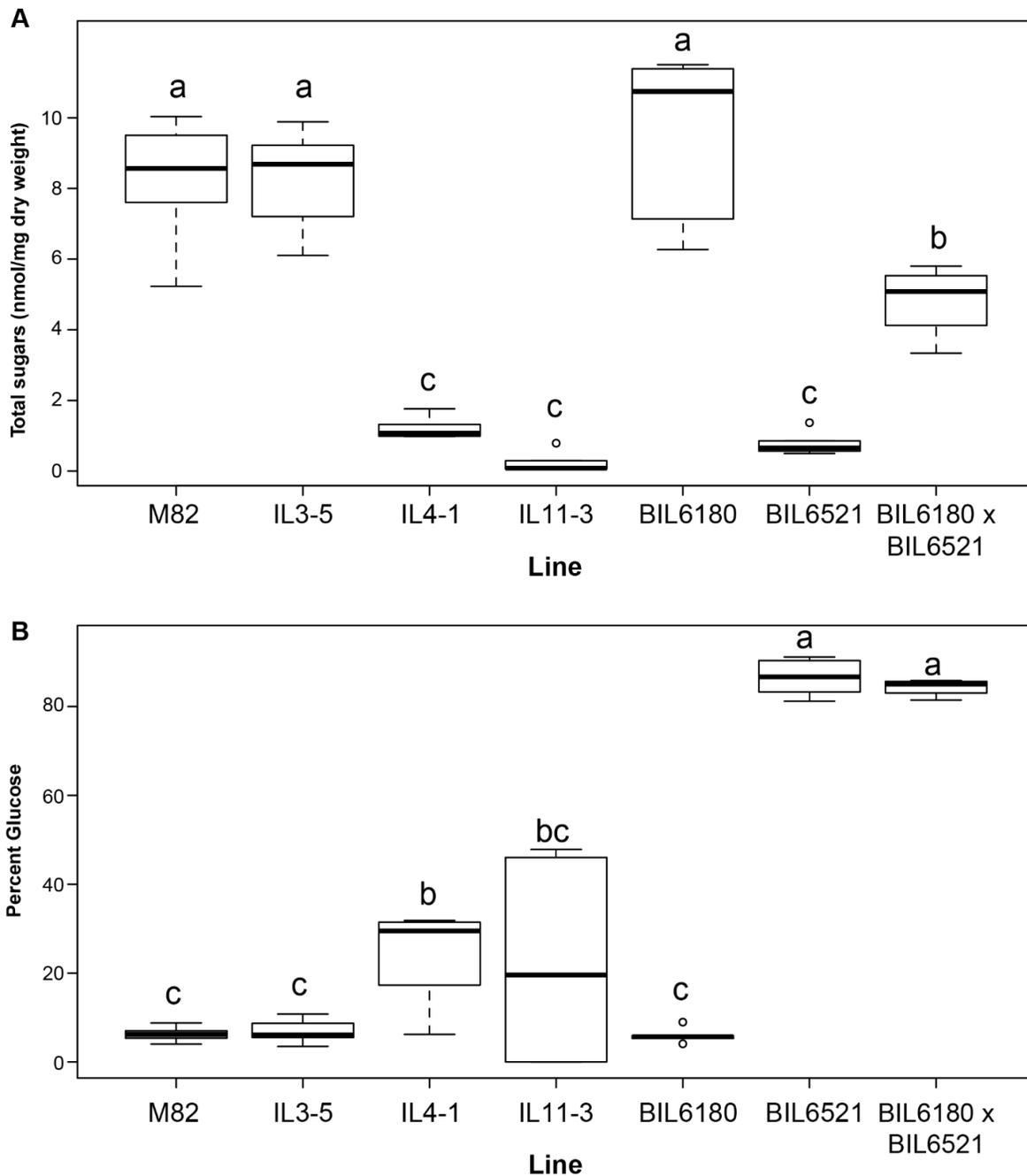


Figure S3.2 Quantification of acylsugars in *S. lycopersicum* M82 and breeding lines containing *S. pennellii* LA0716 introgressions. (A) Total acylsugar accumulation quantified as the sum of sucrose and glucose from saponified acylsugar extracts. (B) Percentage of saponified sugars in acylsugar extracts 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 x BIL6521 F2 plants: $n = 4$.

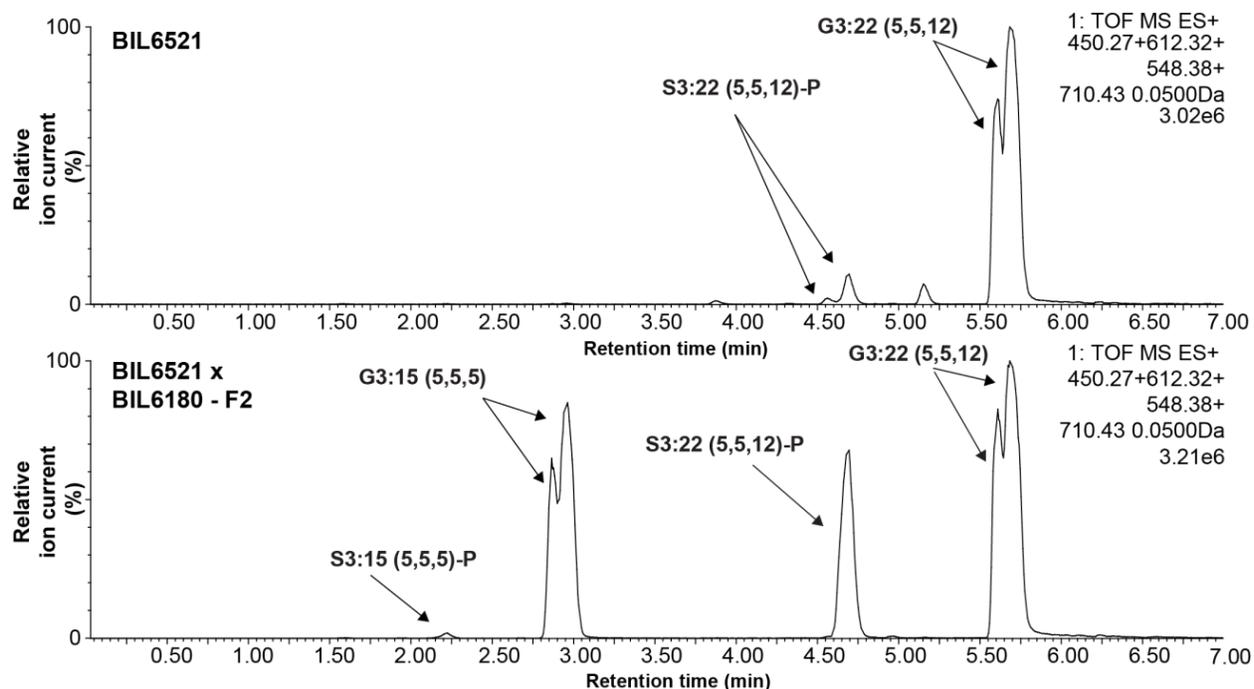


Figure S3.3 Comparison of major acylsugars in BIL6521 and BIL6521 × BIL6180 F2 progeny using LC-MS. ESI+ mode LC-MS of trichome extracts reveals that the F2 progeny of BIL6521 crossed to BIL6180 (genotyped as heterozygous for Chr. 3 and 4 introgression regions) shows an increase in the short chain acylglucose, 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 × BIL6180 lines (mass window: 0.05 Da). Samples were run on the 7 min 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.

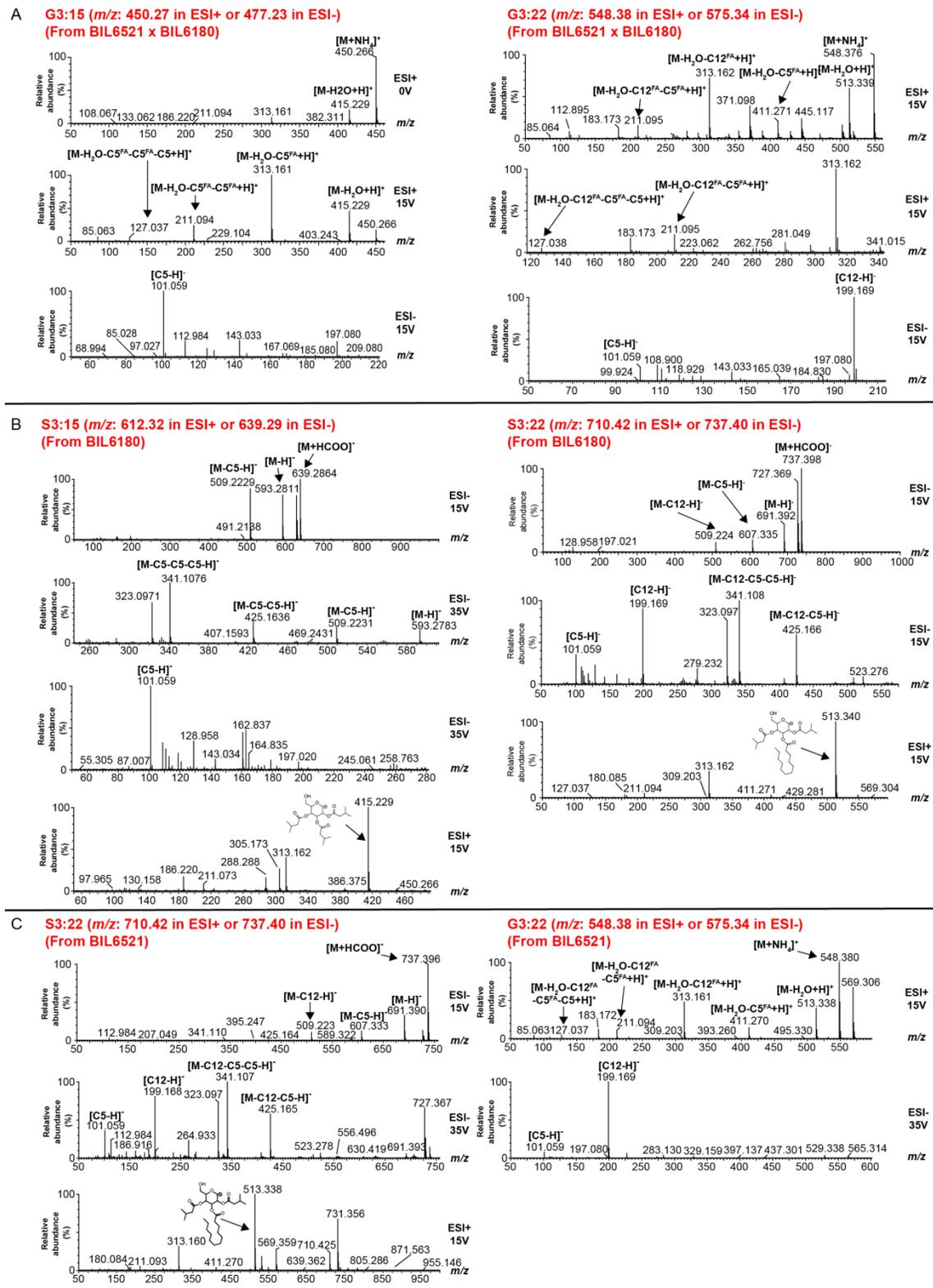


Figure S3.4 Mass spectra of major acylsugars S3:15, S3:22, G3:15, and G3:22 from BIL6521 x BIL6180 F₂ progeny, BIL6180, and BIL6521.

Figure S3.4 (cont'd) Triacylglycerols 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 acylglycerol. 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 CATTGATTAATCAAAAAGAGAAGAAAATAAATAGATGATCATTTCATCTTGTGCTATAT

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

LA0716      gttggttttcgttccgcagATCCTAATGGtaagtttctcaagttgctttgtgttggtac
spasff1-1-1 -----

LA0716      gacaaaagtcatcattcatgaaattcgttatgcatgcaggACCAGTGATATACATGGGGATT
spasff1-1-1 -----

LA0716      ACCATTGTTCTATCAGTACAATCCTTACGATGCACAAACAGGTAACATTGTATGGGGCC
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 CATTGATTAATCAAAAAGAGAAGAAAATAAATAGATGATCATTTCATCTTGTGCTATAT

LA0716      aatcctaattctttttggtgttttgtggttcaatggtgatcaaatttacaattggtggt
spasff1-1-2 AATCCTAATTCTTTTGTGTTTGTGTTGTCATGGTGATCAAATTTACAATTGTTGCT

LA0716      -----tttcgttccgcagATCCTAATGGtaagtttctcaagttgctttgtgttg
spasff1-1-2 AGATTTTGTA-----
                                     ↑

LA0716      gtacgacaaaagtcatcattcatgaaattcgttatgcatgcaggACCAGTGATATACATGGGG
spasff1-1-2 -----

LA0716      ATTTACCATTGTTCTATCAGTACAATCCTTACGATGCACAAACAGGTAACATTGTATGG
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.

```

LA0716      TTCCTATCAGTACAATCCTTACGATGCACAAACAGGTAACATTGTATGGGGCCATTCCACA
spasff1-2  TTCCTATCAGTACAATCCTTACGATGCACAAACAGGTAACATTGTATGGGGCCATTCCACA

LA0716      TCCACTGA-TCTCGTCAACTGGACTCCTCAGCCGCCGGCGCTTCTCCATCAGAGCCTTA
spasff1-2  TCCACTGAATCTCGTCAACTGGACTCCTCAGCCGCCGGCGCTTCTCCATCAGAGCCTTA

LA0716      CGATTTCAAAGGCTGTTTTACCGGTTCTACAAC TCTTCTCTCTGGCGGAAAACCGG CAGT
spasff1-2  CGATTTCAAAGGCTGTTTTACCGGTTCTACAAC TCTTCTCTCTGGCGGAAAACCGG CAGT

LA0716      TCTCTACACCGGCTTAGATTTCTCCGGTATCCAGGTACAAAATCTAGCAGTCCCCAAAAA
spasff1-2  TCTCTACACCGGCTTAGATTTCTCCGGTATCCAGGTACAAAATCTAGCAGTCCCCAAAAA

LA0716      TGTAGCTGA
spasff1-2  TGTAGCTGA
                *****

```

Figure S3.5 (cont'd) 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.

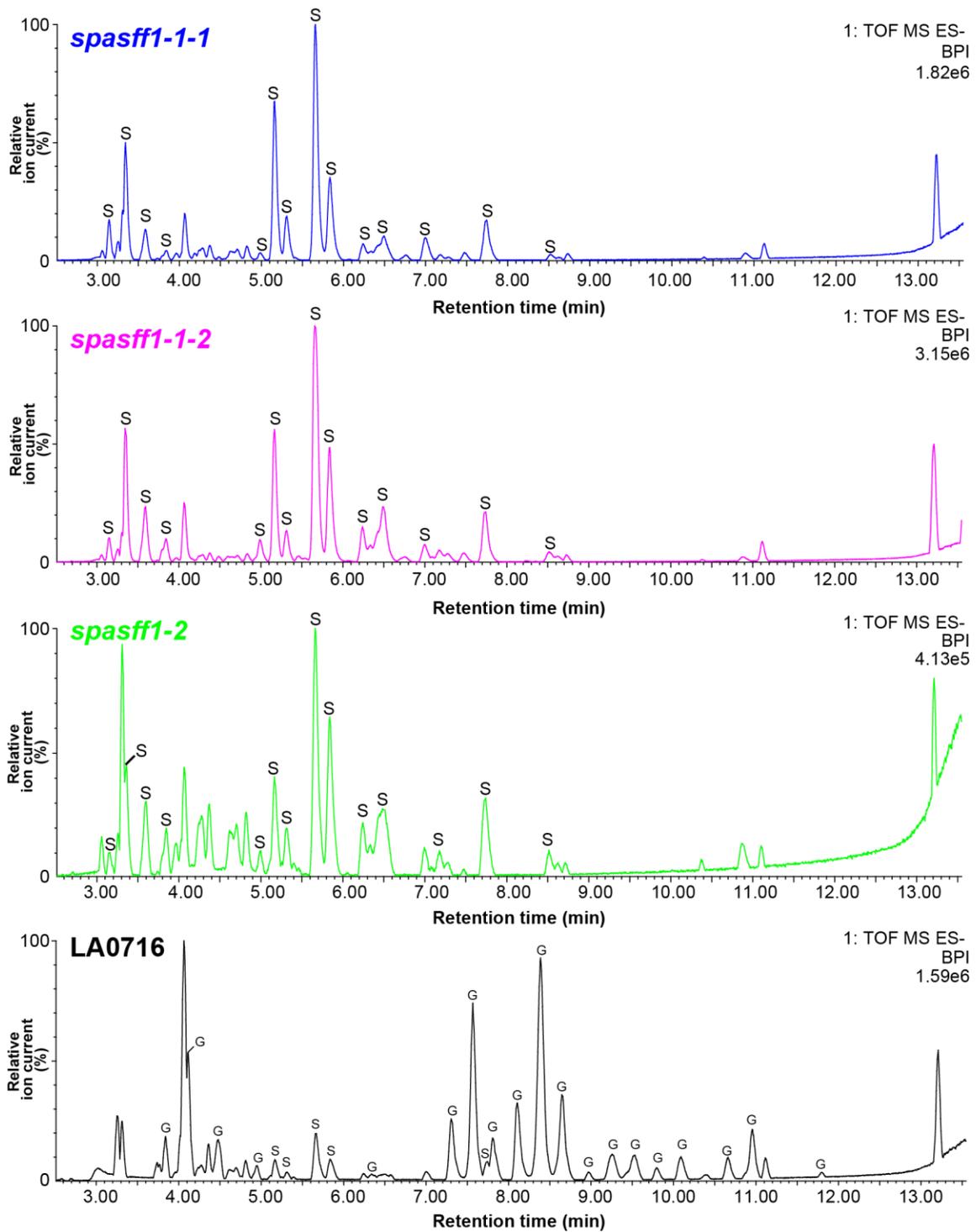


Figure S3.6 Acylsugars in BPI chromatograms of *spasff1* and LA0716 plants.

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. 3.3C. The 21-min method and ESI-mode LC-MS were used for this analysis, as described in the Materials and Methods section. All acylsucroses and acylglucoses elute in the 2.5 to 13.5 min window presented here. The *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. All *spasff1-1-1/1-1-2* are homozygous T₂ lines, while *spasff1-2* are homozygous T₁ lines.

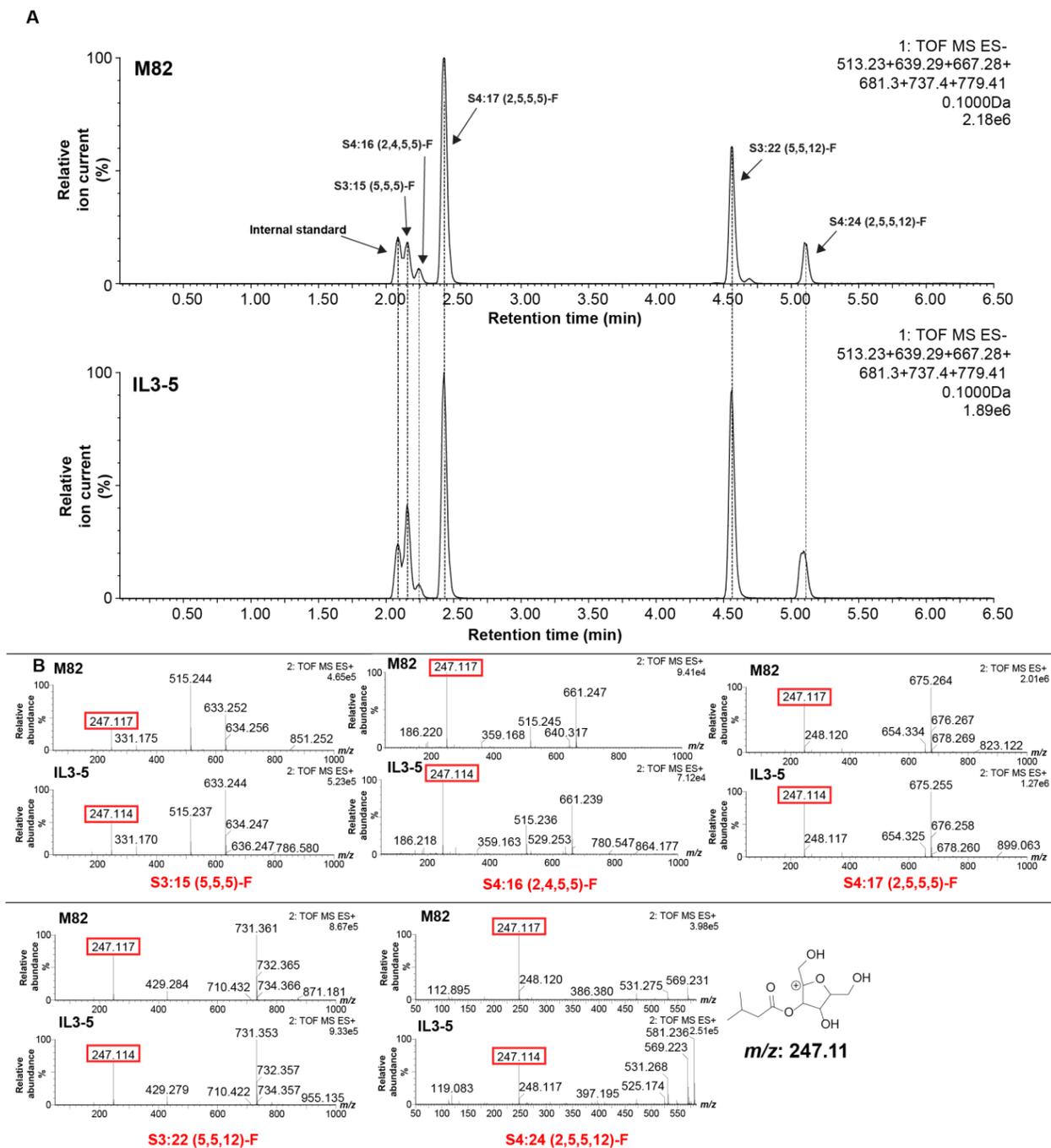


Figure S3.7 Comparison of acylsugars from IL3-5 and parental M82 using LC-MS.

Figure S3.7 (cont'd) (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. Fragment analysis of the 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 using the 7-min method detailed in the Materials and Methods section (mass window: \pm 0.1Da). All ESI- mode acylsugars were identified as formate adducts.

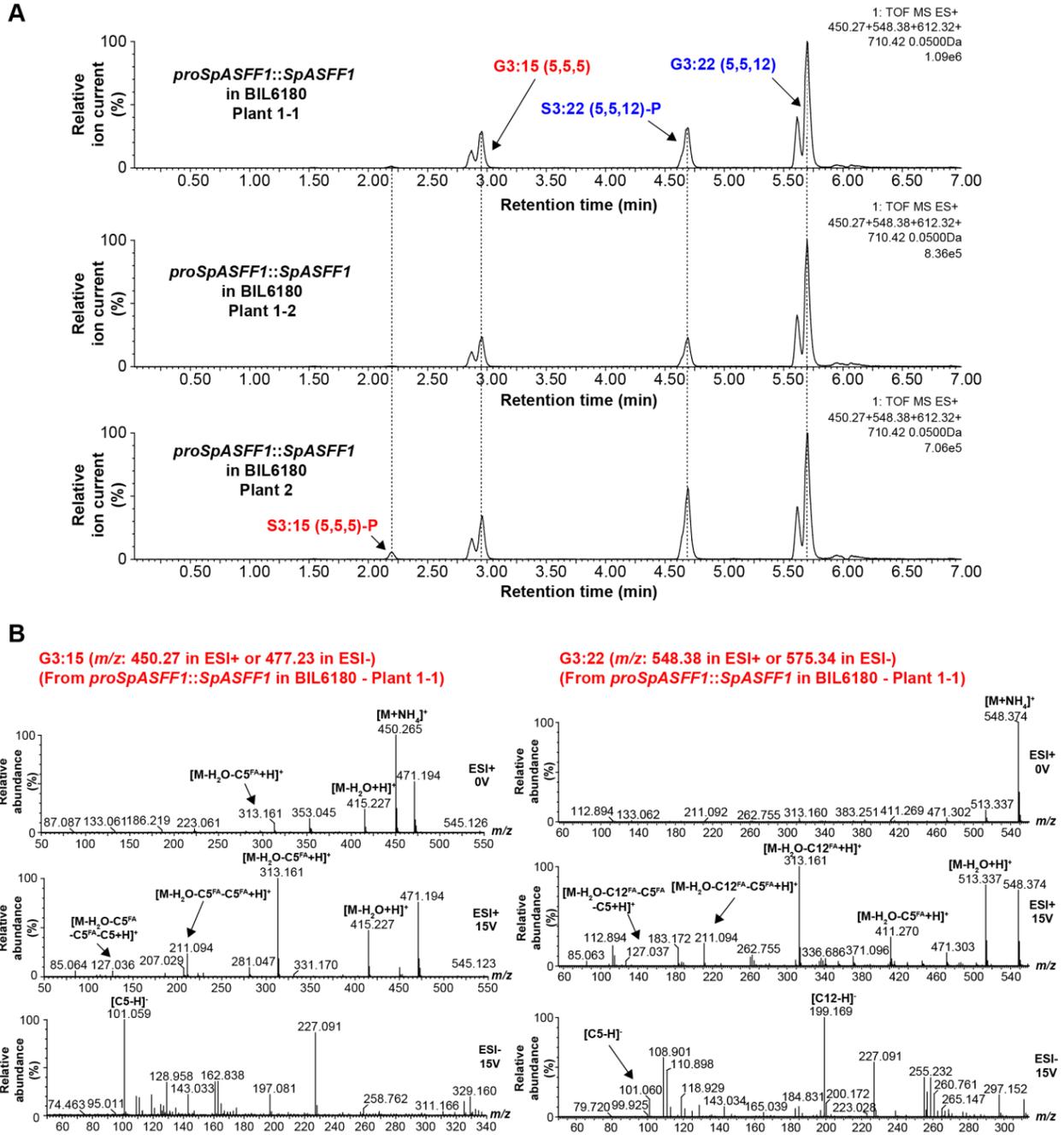


Figure S3.8 LC-MS analysis of P-type acylsucrose-producing *S. lycopersicum* BIL6180 stably transformed with *proSpASFF1::SpASFF1*.

Figure S3.8 (cont'd) (A) ESI+ mode LC-MS of trichome extracts from three *proSpASFF1::SpASFF1* T₂ lines originating from 2 T₀ 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 min method described in the Methods section. Each pair of 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. S4 legend for further detail on fragmentation. All ESI+ mode acylsugars were identified as ammonium adducts.

G3:19 (m/z : 506.33 in ESI+ or 533.30 in ESI-)
 (From *in vitro* assays (Fig. 5))

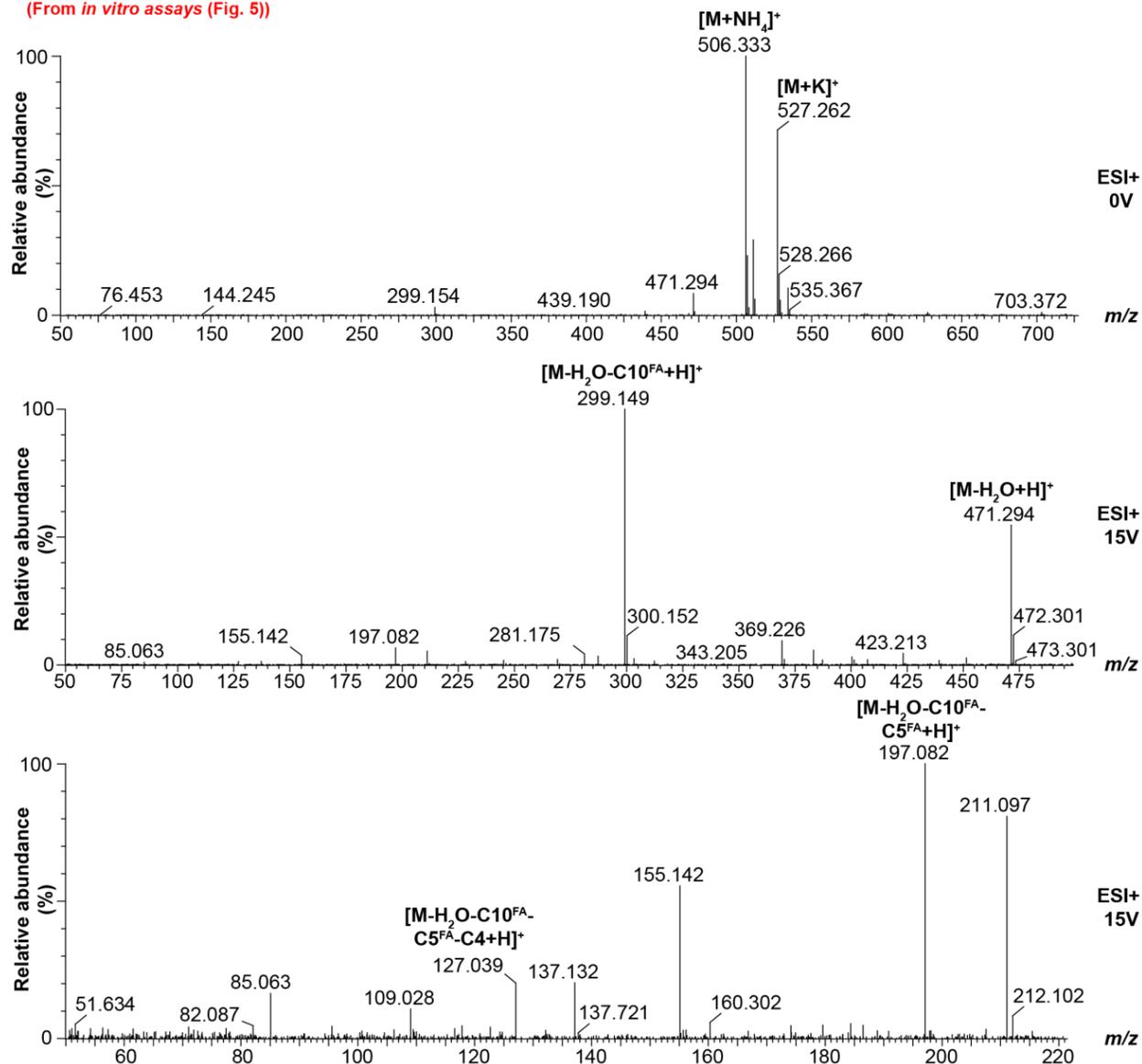


Figure S3.9 Mass spectra of G3:19-derived from SpASFF1 *in vitro* assay. Fragmentation of the S3:19 + SpASFF1 reaction product, G3:19, in ESI+ mode (0V and 15V) using collision induced dissociation (from Fig. 3.5). The fragmentation is characterized by the loss of a C10 and C5 fatty acid from the triacylglycerol, followed by loss of a C4 ketene. These results are consistent with the product cognate, S3:19.

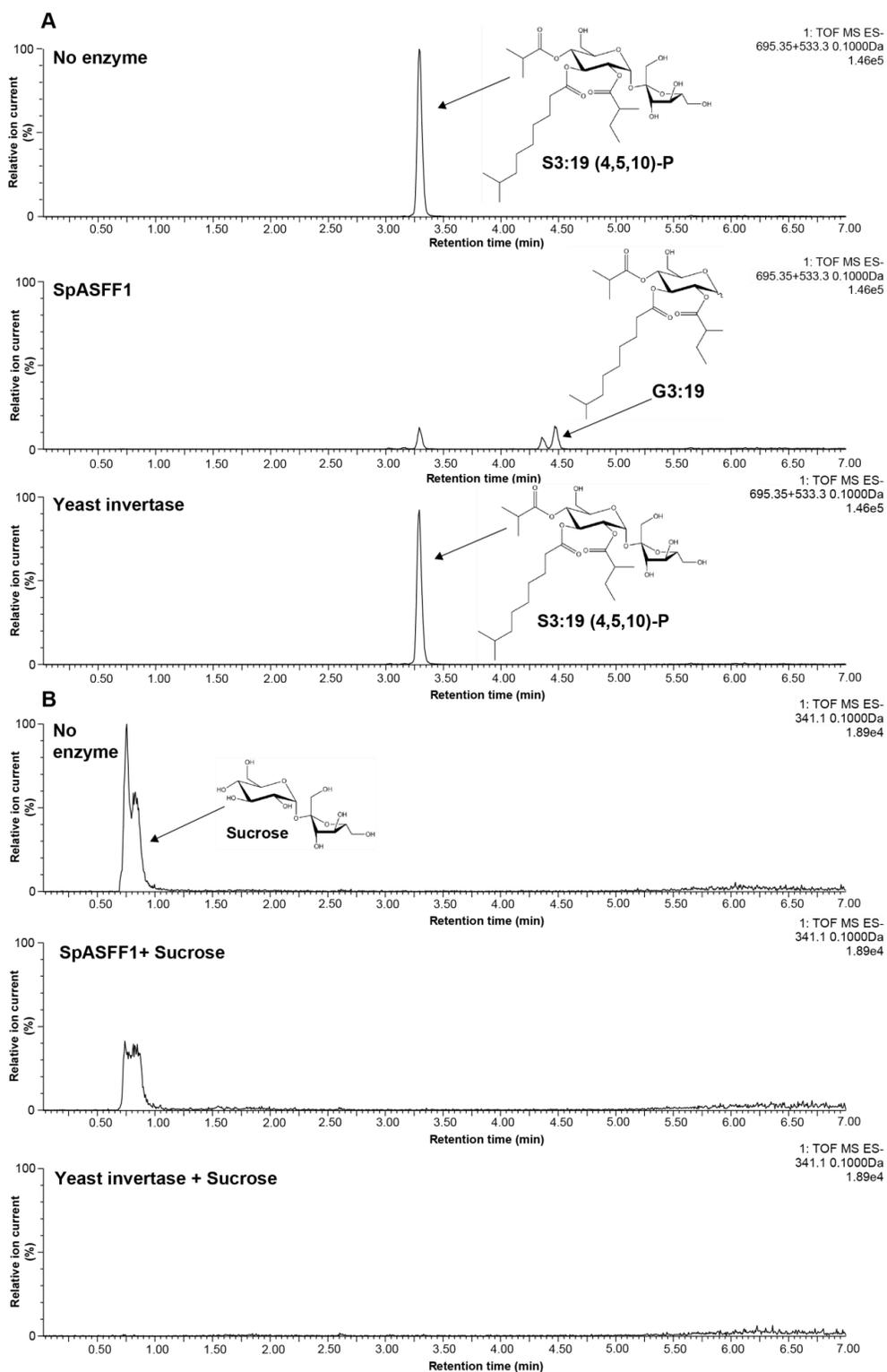


Figure S3.10 SpASFF1 cleaves a purified P-type triaclysucrose but not unmodified sucrose while yeast invertase cleaves unmodified sucrose but not triaclysucrose.

Figure S3.10 (continued) (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 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. The acylglucose structure is inferred from the collision-induced dissociation spectrum (Figure S3.9). All ESI- mode acylsugars were identified as formate adducts, while sucrose was identified as an [M-H]⁻ ion.

Table S3.1 Oligonucleotides used in this study.

Sequence name	Nucleotide sequence
ASFF_F	ATGGGATATGTTAGAAGTGTTTGG
ASFF_R	TCAATTGATTTGAGCTGTTTTCA
(pEAQ-HT)-ASFF-His_F	GTATATTCTGCCCAAATTCGATGGGATATGTTAGAAGTGT
(pEAQ-HT)-ASFF-His_R	TGATGGTGATGGTGATGCCATTGATTTGAGCTGTTTTCA
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	AGGGGATTTTGTGAGGGTTG
RT_ubiquitin_F	TCGTAAGGAGTGCCCTAATGCTGA
RT_ubiquitin_R	CAATCGCCTCCAGCCTTGTTGTAA
ASFF_001_F	CAAAAAAGCAGGCTCCGCTGATAGTTATGCCAATGTACCACA
ASFF_001_R	ACGATCCTCTTAGTTGGTCCAC
ASFF_002_F	GTGGACCAACTAAGAGGATCGT
ASFF_002_R	TGGCCCCATACAATGTTACCTG
ASFF_003_F	CAGGTAACATTGTATGGGGCCA
ASFF_003_R	TTTGGGAAGTTCTGGCTCGG
ASFF_004_F	CCGAGCCAGAACTTCCCAA
ASFF_004_R	GGGTCGGCGCGCCACCCTTAAAGTGTTGACTGACCATTCT
ASFF_Promoter_F1	CAAAAAAGCAGGCTCCGCTGCCAATGTACCACAATTAGTATT
ASFF_Promoter_R1	AGAAAGCTGGGTGCGCTTTTAGTTGAAGATGGCAACTACATTTCA
ASFF_Chr3_Indel_002_F	CAAAAAGAAGAAAAGGAAAACAGACA
ASFF_Chr3_Indel_002_R	GTGGGACTAAACTTTGTAGTTCC
04g011460_MarkerF	TAACAAAGCTTATGCACTCTTAG
04g011460_MarkerR	ATCTACTACCTTCATATGCACAT
ASFF1_transcript_F	TCATTTCCATTCATAGCTATGGCA
ASFF1_transcript_R	GTTGCACCATCCACTGCTAA
sgRNA 1	AGTCCAGTTGACGAGATCAGTGG
sgRNA 2	TCTTCTCTCTGGCGGAAAACCGG
sgRNA1 gBLOCK	TTTTAATGTAAGGGTGGATGCAGTGGGCCCCACTCTGTGAA GACAA ACTAGAATTCGAGCTCGGAGGGAGTGATCAAAAGTCC CACATCGATCAGGTGATATATAGCAGCTTAGTTTATATAATGAT AGAGTCGACATAGCGATTGAGTCCAGTTGACGAGATCAGGTTT TAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA ACTGAAAAAGTGGCACCGAGTCGGTG

Table S3.1 (cont'd)

Sequence name	Nucleotide sequence
sgRNA2 gBLOCK	TTTAATGTA CTGGGGTGGATGCAGTGGGCCCCACTCTGTGAAGA CAATTACGAATTCCCATGGGGAGGGAGTGATCAAAAGTCCCAC ATCGATCAGGTGATATATAGCAGCTTAGTTTATATAATGATAGAG TCGACATAGCGATTGTCTTCTCTCTGGCGGAAAACGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTG TACAAAGTTGGCATTACGCTCGCTCGCTCAGATTGTCTTCTGCACG AAGTGGTTTAAACTATCAGTGTTTGACAG GATATATT
pAGM4723_SeqF1	TTTCGCCACCTCTGACTTGAG
pAGM4723_SeqR1	CAGCTTGGCATCAGACAAACC
pICSL11024_SeqF1	CGGCGAACTAATAACGCTCA
pICH47742CAS9_SeqF2	ACTTACGCTCATCTCTTCGAC
pICH47742_SeqF1	CTGTTGAATTACGTTAAGCATG
pICH41780_SeqR1	TCGGTCACATGTGCATCCTC
ASFF_SeqR	GTTGATTTCCA ACTACCATTCTCC
pK7WG-Kan-F	TTGACTCTAGCTAGAGTCCGAA
pK7WG-Kan-R	ATTGAACAAGATGGATTGCACGCA

Table S3.2 Annotation of acylsugars identified in BIL6521 and BIL6521 × BIL6180 F₂ using LC-MS and 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	NH4 ⁺
520.34	+	G3:20 (5,5,10)	4.97	NH4 ⁺
534.36	+	G3:21 (5,5,11)	5.23	NH4 ⁺
548.38	+	G3:22 (5,5,12)	5.68	NH4 ⁺
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	NH4 ⁺
450.27	+	G3:15(5,5,5)	2.96	NH4 ⁺
464.28	+	G3:16 (5,5,6)	3.25	NH4 ⁺
520.34	+	G3:20 (5,5,10)	4.95	NH4 ⁺
534.32	+	G3:21 (5,5,11)	4.56	NH4 ⁺
548.38	+	G3:22 (5,5,12)	5.68	NH4 ⁺

Table S3.3 Annotation of glycoside hydrolase (GH) candidates identified in the AG3.2 region.

SGN ID (<i>S. lycopersicum</i>/<i>S. pennellii</i>)	Enzyme family	Nominal activity	Typical substrates
Solyc03g121540/ Sopen03g040350	GH35	β -galactosidase	β -D-galactosides; β -L-arabinosides; oligogalactosides [52]
Solyc03g121680/ Sopen03g040490	GH32	β -fructofuranosidase	β -fructofuranosides [42]
Solyc03g123900/ Sopen03g041640	GH47	α -mannosidase	Terminal α -D-mannose residues of oligomannose oligosaccharides [51]

REFERENCES

REFERENCES

- 1 Schillmiller, A.L. *et al.* (2012) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. *Proc. Natl. Acad. Sci.* 109, 16377–16382
- 2 Ghosh, B. *et al.* (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
- 3 Schillmiller, A.L. *et al.* (2015) Functionally Divergent Alleles and Duplicated Loci Encoding an Acyltransferase Contribute to Acylsugar Metabolite Diversity in *Solanum* Trichomes. *Plant Cell* 27, 1002–1017
- 4 Fan, P. *et al.* (2017) Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. *Nat. Commun.* 8, 2080
- 5 Nadakuduti, S.S. *et al.* (2017) Characterization of Trichome-Expressed BAHD Acyltransferases in *Petunia axillaris* Reveals Distinct Acylsugar Assembly Mechanisms within the Solanaceae. *Plant Physiol.* 175, 36–50
- 6 Liu, X. *et al.* (2017) Profiling, isolation and structure elucidation of specialized acylsucrose metabolites accumulating in trichomes of *Petunia* species. *Metabolomics* 13, 85
- 7 Moghe, G.D. *et al.* (2017) Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife* 6, e28468
- 8 Weinhold, A. and Baldwin, I.T. (2011) Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *Proc. Natl. Acad. Sci.* 108, 7855–7859
- 9 Luu, V.T. *et al.* (2017) O-Acyl Sugars Protect a Wild Tobacco from Both Native Fungal Pathogens and a Specialist Herbivore. *Plant Physiol.* 174, 370–386
- 10 Matsuzaki, T. *et al.* (1989) Isolation and Characterization of Tetra- and Triacylglycerol from the Surface Lipids of *Nicotiana glauca*. *Agric. Biol. Chem.* 53, 3343–3345
- 11 Escobar-Bravo, R. *et al.* (2016) A Jasmonate-Inducible Defense Trait Transferred from Wild into Cultivated Tomato Establishes Increased Whitefly Resistance and Reduced Viral Disease Incidence. *Front. Plant Sci.* 7,
- 12 Leckie, B.M. *et al.* (2016) Differential and Synergistic Functionality of Acylsugars in Suppressing Oviposition by Insect Herbivores. *PLoS ONE* 11, e0153345
- 13 A. Burke, B. *et al.* (1987) Polar epicuticular lipids of *Lycopersicon pennellii*. *Phytochemistry* 26, 2567–2571

- 14 King, R.R. and Calhoun, L.A. (1988) 2,3-Di-O- and 1,2,3-tri-O-acylated glucose esters from the glandular trichomes of *Datura metel*. *Phytochemistry* 27, 3761–3763
- 15 Maldonado, E. *et al.* (2006) Sucrose Esters from the Fruits of *Physalis nicandroides* var. *attenuata*. *J. Nat. Prod.* 69, 1511–1513
- 16 Kim, J. *et al.* (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, 1854–1870
- 17 Fan, P. *et al.* (2016) In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. *Proc. Natl. Acad. Sci.* 113, E239–E248
- 18 Ning, J. *et al.* (2015) A feedback insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. *Plant Physiol.* DOI: 10.1104/pp.15.00474
- 19 Fobes, J.F. *et al.* (1985) Epicuticular Lipid Accumulation on the Leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. *Plant Physiol.* 77, 567–570
- 20 Shapiro, J.A. *et al.* (1994) Acylsugars of the wild tomato *Lycopersicon pennellii* in relation to geographic distribution of the species. *Biochem. Syst. Ecol.* 22, 545–561
- 21 Kuai, J.P. *et al.* (1997) Regulation of Triacylglyceride Fatty Acid Composition (Uridine Diphosphate Glucose:Fatty Acid Glucosyltransferases with Overlapping Chain-Length Specificity). *Plant Physiol.* 115, 1581–1587
- 22 Li, A.X. *et al.* (1999) Glucose Polyester Biosynthesis. Purification and Characterization of a Glucose Acyltransferase. *Plant Physiol.* 121, 453–460
- 23 Li, A.X. and Steffens, J.C. (2000) An acyltransferase catalyzing the formation of diacylglyceride is a serine carboxypeptidase-like protein. *Proc. Natl. Acad. Sci.* 97, 6902–6907
- 24 Leckie, B.M. *et al.* (2013) Quantitative trait loci regulating sugar moiety of acylsugars in tomato. *Mol. Breeding* 31, 957–970
- 25 Darwin, S.F. and Acton, E.H. (1894) *Practical Physiology of Plants*, Cambridge University Press.
- 26 Sainz-Polo, M.A. *et al.* (2013) Three-dimensional Structure of *Saccharomyces* Invertase: ROLE OF A NON-CATALYTIC DOMAIN IN OLIGOMERIZATION AND SUBSTRATE SPECIFICITY. *J. Biol. Chem.* 288, 9755–9766
- 27 Wan, H. *et al.* (2018) Evolution of Sucrose Metabolism: The Dichotomy of Invertases and Beyond. *Trends Plant Sci.* 23, 163–177

- 28 Ofner, I. *et al.* (2016) *Solanum pennellii* backcross inbred lines (BILs) link small genomic bins with tomato traits. *Plant J.* 87, 151–160
- 29 Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, 45e–445
- 30 Karimi, M. *et al.* (2002) GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195
- 31 Doench, J.G. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191
- 32 Hsu, P.D. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832
- 33 Belhaj, K. *et al.* (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* 9, 39
- 34 Weber, E. *et al.* (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE* 6, e16765
- 35 McCormick, S. (1991) Transformation of tomato with *Agrobacterium tumefaciens*. In *Plant Tissue Culture Manual* (Lindsey, K., ed), pp. 311–319, Springer Netherlands
- 36 Fillatti, J. *et al.* (1987) Efficient Transfer of a Glyphosate Tolerance Gene into Tomato Using a Binary *Agrobacterium-Tumefaciens* Vector. *Bio-Technology* 5, 726–730
- 37 Peyret, H. and Lomonosoff, G.P. (2013) The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant. Mol. Biol.* 83, 51–58
- 38 R Core Team (2017) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing.
- 39 Smeda, J.R. *et al.* (2018) Combination of Acylglucose QTL reveals additive and epistatic genetic interactions and impacts insect oviposition and virus infection. *Mol. Breeding* 38, 3
- 40 Eshed, Y. and 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, 1147–1162
- 41 The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641
- 42 Van Den Ende, W. *et al.* (2009) Donor and acceptor substrate selectivity among plant glycoside hydrolase family 32 enzymes. *FEBS* 276, 5788–5798

- 43 Nakashima, T. *et al.* (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
- 44 Bolger, A. *et al.* (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* 46, 1034–1038
- 45 Johnson, K.A. and Goody, R.S. (2011) The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* 50, 8264–8269
- 46 De Coninck, B. *et al.* (2005) Arabidopsis AtcwINV3 and 6 are not invertases but are fructan exohydrolases (FEHs) with different substrate specificities. *Plant Cell Environ.* 28, 432–443
- 47 Minic, Z. (2008) Physiological roles of plant glycoside hydrolases. *Planta* 227, 723–740
- 48 Ghangas, G.S. and Steffens, J.C. (1993) UDPglucose: fatty acid transglucosylation and transacylation in triacylglycerol biosynthesis. *Proc. Natl. Acad. Sci.* 90, 9911–9915
- 49 Koenig, D. *et al.* (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci.* 110, E2655–E2662
- 50 Schillmiller, A.L. *et al.* (2016) Acylsugar Acylhydrolases: Carboxylesterase-Catalyzed Hydrolysis of Acylsugars in Tomato Trichomes. *Plant Physiol.* 170, 1331–1344
- 51 Herscovics, A. (2001) Structure and function of Class I α 1,2-mannosidases involved in glycoprotein synthesis and endoplasmic reticulum quality control. *Biochimie* 83, 757–762
- 52 Tanthanuch, W. *et al.* (2008) Genomic and expression analysis of glycosyl hydrolase family 35 genes from rice (*Oryza sativa* L.). *BMC Plant Biol.* 8, 84
- 53 Särkinen, T. *et al.* (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.* 13, 214
- 54 Nesbitt, T.C. and Tanksley, S.D. Comparative Sequencing in the Genus *Lycopersicon*: Implications for the Evolution of Fruit Size in the Domestication of Cultivated Tomatoes.
- 55 Patti, G.J. *et al.* (2012) Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13, 263–269
- 56 Nagana Gowda, G.A. and Raftery, D. (2017) Recent Advances in NMR-Based Metabolomics. *Anal. Chem.* 89, 490–510

CHAPTER 4 – INTRASPECIFIC VARIATION IN *SOLANUM PENNELLII* TRICHOME METABOLISM

Dr. Thilani M. Anthony collected and interpreted all unpublished NMR spectra presented in this chapter.

CHEMICAL DIVERSITY IN *SOLANUM PENNELLII* TRICHOMES

Solanum pennellii secretory glandular trichomes (SGTs) produce abundant and diverse acylsugars. These compounds accumulate to up to 20% leaf dry weight [1], consist of both acylglucoses and acylsucroses [2,3], and collectively possess an array of at least 13 different acyl chains [3–5]. However, the profiles of intact acylsugars that accumulate *in planta* were previously unstudied and the true chemical diversity in *S. pennellii* SGTs unknown. Instead, previous studies of acylsugar metabolism in *S. pennellii* focused on either the chemical substructures of acylsugars (*i.e.*, sugar cores and fatty acids; [3–5]) or the enzymes that synthesize and degrade acylsugars (ASATs, IMPS3, ASHs, ASFF1) [5–8]. In addition, the majority of studies on acylsugar biosynthesis in *S. pennellii* focused on one accession, *S. pennellii* LA0716, collected in southern Peru in 1958 [4,6–13], though most of the roughly two-dozen accessions of *S. pennellii* analyzed thus far from across the 1500-km geographic range of the species are known to produce acylsugars [3,5]. Notable exceptions to these trends are an early report on *S. pennellii* acylsugar metabolism in which Burke and co-workers partially characterized acylglucoses in *S. pennellii* LA0716 by NMR, determining the positions and structures of acyl chains from nine acylglucoses [2], work by Shapiro and co-workers which described abundance of sugar cores and acyl chains in acylsugars from 19 accessions of *S. pennellii* distributed across the range of the species [3], and work by Ning and co-workers who analyzed acylsugar acyl chains in 14 *S. pennellii* accessions to determine the genetic basis for prominent accumulation of 3-methylbutanoate acyl chains in northern accessions of the species in contrast to prominent accumulation of 2-methylpropanoate acyl chains in southern accessions [5].

Characterization of acylsugar phenotype by relative abundance of sugar cores and acyl chains is useful for identifying major differences between biosynthetic pathways in related species or populations by comparative biochemistry. However, subtle variations in acylsugar phenotype overlooked by these analyses may facilitate discovery of additional pathway enzymes but require a more precise approach to detect. To create a more complete picture of acylsugar diversity in *S. pennellii*, we combined untargeted ultraperformance liquid chromatography-high resolution mass spectrometry (UPLC-HR-MS) and NMR spectroscopy to characterize the SGT metabolome of 16 *S. pennellii* accessions from across the plant's geographic range, revealing variation in levels of 43 specialized metabolites including 39 acylsugars. We initially annotated all metabolites based on mass spectra and subsequently purified and resolved structures of selected acylsugars by NMR. We applied multivariate statistical analyses to these profiling data to identify specific compounds that distinguish various *S. pennellii* accessions from one another. Our analyses confirmed previous reports showing that acyl chain complement drives acylsugar variation between *S. pennellii* accessions [3,5], and when compared to expression levels of *ASFF1*, a gene involved in acylsugar turnover, demonstrated a positive correlation between *ASFF1* gene expression and acylglucose accumulation. We also observed tetraacylglucoses and methyl flavonoids, two classes of compounds previously undescribed in *S. pennellii* SGTs.

MATERIALS AND METHODS

Plant material

Seeds of all *Solanum pennellii* accessions were obtained from the C.M. Rick Tomato Genetics Resource Center (TGRC; University of California, Davis, CA). Seeds were treated with

half-strength bleach for 30 min and rinsed three times in de-ionized water for 5 min before sowing on moist filter paper in petri dishes. Seedlings were transferred to peat pots upon germination and grown under a 16-hour photoperiod [$190 \mu\text{mol m}^{-2} \text{s}^{-1}$ photosynthetic photon flux density (PPFD)] at 21°C and 75% relative humidity until the first true leaves developed. Plants in peat pots were then transferred to 9-cm pots in a peat-based propagation mix (SunGro, Agawam, MA). Plants were then grown under a 12-hour photoperiod ($\sim 600 \mu\text{mol m}^{-2} \text{s}^{-1}$ PPFD) with a 28°C daytime temperature, 12°C nighttime temperature, and 50% relative humidity for the remainder of the experiment. All light/dark cycles totaled 24 hours. Plants were watered with deionized water on Mondays and supplemented with half-strength Hoagland's solution on Thursdays.

Acylsugar extraction

Single leaflets from the youngest fully expanded leaves of individual 16-week-old *S. pennellii* plants were harvested and placed into pre-washed 10 x 75 mm borosilicate glass test tubes. Leaflets from six individual plants of each *S. pennellii* accession were collected. An empty test tube was also included as a process blank. To each tube, 1 mL of a 3:3:2 mixture of acetonitrile/isopropanol/water containing 0.1% formic acid and $0.25 \mu\text{M}$ telmisartan was added. Tubes were vortexed for 30 s and solvent decanted into 2-mL glass autosampler vials. Equal volumes of each extract (excluding the process blank) were combined to create a pooled quality control (QC) sample. Vials were sealed with PTFE-lined caps and stored at -20°C for later processing.

Metabolomic analysis by LC-MS

Aliquots of *S. pennellii* acylsugar extracts, process blank, and QC sample were diluted 100-fold in 1:1 methanol/water containing 0.1% formic acid in new 2-mL autosampler vials. Five aliquots of both the diluted process blank and QC samples were prepared and analyzed. Samples were subjected to UPLC-MS analysis using an Acquity UPLC coupled to a G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA). Separations were performed using an Acquity BEH C18 UPLC column (2.1 x 100 mm, 1.7 μ m; Waters Corporation). The mobile phases consisted of 10 mM ammonium formate, pH 2.8 (solvent A) and 10 mM ammonium formate, pH 2.8, in 90% acetonitrile (solvent B). Five-microliter volumes were injected onto the column and eluted with a linear elution gradient of 0% B at 0-1 min, 55% B at 1.01 min, 100% B at 16-18 min, and 0% B at 18.01-20 min. The solvent flow rate was 0.4 mL/min and the column temperature was 40°C. Analyses were performed using positive-ion mode electrospray ionization and sensitivity mode analyzer parameters. Source parameters were as follows: capillary voltage at 3.00 kV, sampling cone voltage at 35 V, source offset at 80 V, source temperature at 100°C, desolvation temperature at 350°C, cone gas flow at 50.0 L/hour, and desolvation gas flow at 600.0 L/hour. Mass spectrum acquisition was performed from 2 to 18 min over an m/z range of 50 to 1500 with a scan time of 0.5 s. Gentle ionization conditions used a collision energy of 6.0 eV; fragment ions were obtained using a collision energy ramp of 15.0 to 40.0 eV. Lockmass reference spectra using leucine enkephalin (m/z 556.2766) were collected without application of spectral correction. Spectra were acquired in continuum format using quasi-simultaneous acquisition of low- and high-energy spectra (MS^E).

Untargeted metabolomics data processing

For untargeted metabolomic analysis, data were initially processed using Progenesis QI v2.4 software (Nonlinear Dynamics Ltd., Newcastle, UK). Leucine enkephalin lockmass correction (m/z 556.2766) was applied during run importation and all runs were aligned to retention times of a bulk pool run automatically selected by the software. Peak picking was carried out using an automatic sensitivity level of 5 (most sensitive) without restriction on minimum chromatographic peak width. Peak picking was restricted to features eluting between 2.15 and 14.5 min. Spectral deconvolution was carried out considering the following possible adduct ions: $M+H-H_2O$, $M+H$, $M+NH_4$, $M+Na$, $M+K$, $M+C_2H_8N$, $2M+H$, $2M+NH_4$, $2M+Na$, $2M+K$, $2M+C_2H_8$.

Further analysis of compounds identified by Progenesis QI software was executed using EZinfo v3.0.2 software (Umetrics, Umeå, Sweden). For principal component analysis (PCA), data were subjected to logarithmic transformation and scaled to unit variance (“autoscaled”). For partial least squares discriminant analysis (PLS-DA) and orthogonal partial least squares/projection to latent structures discriminant analysis (OPLS-DA), no data transformation was applied, and Pareto scaling was implemented. Generation of OPLS-DA models was carried out as follows: for each model, the relevant data files were divided into three subsets, each subset containing data files representing two of six biological replicates from each accession considered by the model. The three data subsets, each representing one-third of the relevant data, were used to generate three independent OPLS-DA models. Each model was then used to classify the remaining two-thirds of the data not used in generation of the model, representing

four of six biological replicates from each accession considered by the model. All OPLS-DA model statistics reported represent averages of the three independent models.

For all metabolic features identified with Progenesis QI and used in downstream analyses with EZinfo, spectra were interpreted using MassLynx v4.2 software (Waters Corporation). Accurate masses of all features in all raw data files were obtained by applying the Continuous Lockmass Correction feature of the Accurate Mass Measure module. All precursor ions (identified as either $M+NH_4$ or $M+H$ adducts) were selected from the low-energy function while all fragment ions were identified in the high-energy function. Observed m/z values for precursor and product ions as well as neutral loss masses were compared to theoretical values generated using ChemDraw v19.0 software (PerkinElmer, Inc., Waltham, MA).

Acylsugar quantification

Acylsugars were quantified from untargeted LC-MS data by integration of extracted ion chromatogram peaks using the QuanLynx module of MassLynx software (Waters Corporation). All acylsucroses and acylglucoses detected in the metabolomics dataset were quantified using a standard curve of two acylsucroses and two acylglucoses [S3:12(4,4,4), S3:18(4,4,10)-1, G3:12(4,4,4), and G3:18(4,4,10)-1] at 0.3125, 0.625, 1.25, 2.5, and 5.0 μ M. Acylsugars containing fewer than 18 carbons in all acyl chains were quantified using the corresponding 12-carbon acylsugar while acylsugars containing 18 or more carbons were quantified using the corresponding 18-carbon acylsugar. All quantifications were performed using extracted ion chromatograms of the m/z value for the relevant $M+NH_4$ adduct (this adduct was chosen because it appeared consistently for all acylsugars analyzed). A mass window of 0.05 Da was used. When multiple acylsugar isomers (including anomers) were present, all acylsugars of a

given molecular formula were quantified using a single extracted ion chromatogram trace. The retention time window was adjusted for each compound based on the number of and retention time differences between isomers. Telmisartan was used as an internal reference for all quantifications.

For quantification of total acylsugar cores (*i.e.*, sucrose and glucose), acylsugar extracts were saponified and analyzed by LC-MS. For each acylsugar extract, a 20- μ L aliquot was evaporated to dryness in a 1.7-mL microfuge tube using a vacuum centrifuge and dissolved in 200 μ L of a 1:1 methanol/3 N aqueous ammonia solution. The saponification reaction was incubated at room temperature for 48 hours at which point solvent was removed by vacuum centrifuge at room temperature. The sample was dissolved in 200 μ L of 10 mM ammonium bicarbonate (pH 8.0) in 90% acetonitrile containing 0.5 μ M $^{13}\text{C}_{12}$ -sucrose and 0.5 μ M $^{13}\text{C}_6$ -glucose as internal standards and transferred to a 2-mL LC-MS vial. Samples were subjected to UPLC-MS-MS analysis using an Acquity UPLC coupled to an Acquity TQD tandem quadrupole mass spectrometer (Waters Corporation). Separations were performed using an Acquity BEH Amide UPLC column (2.1 x 100 mm, 1.7 μ m; Waters Corporation). The mobile phases consisted of 10 mM ammonium bicarbonate, pH 8.0, in 50% acetonitrile (solvent A) and 10 mM ammonium bicarbonate, pH 8.0, in 90% acetonitrile (solvent B). Five-microliter samples were injected onto the column and eluted with a linear elution gradient of 100% B at 0 min, 0% B at 5 min, 100% B at 5.01 min, and held at 100% B until 10 min. The solvent flow rate was 0.5 mL/min and the column temperature was 40°C. Sugars were detected using multiple reaction monitoring (MRM). The following transitions were used: glucose precursor ion ($[\text{M}-\text{H}]^-$) of m/z 179, product ion of m/z 89, cone voltage of 16 V, and collision potential of 10 V; $^{13}\text{C}_6$ -glucose

precursor ion ($[M-H]^-$) of m/z 185, product ion of m/z 92, cone voltage of 16 V, and collision potential of 10 V; sucrose precursor ion ($[M-H]^-$) of m/z 341, product ion of m/z 89, cone voltage of 40 V, and collision potential of 22 V; $^{13}C_{12}$ -sucrose precursor ion ($[M-H]^-$) of m/z 353, product ion of m/z 92, cone voltage of 40 V, and collision potential of 22 V. Levels of sucrose and glucose were quantified using a standard curve of the corresponding sugar at 31.25, 62.5, 125, 250, and 500 μ M.

RNA extraction, cDNA synthesis, and qPCR

Single leaflets from the youngest fully expanded leaf of 12-week-old *S. pennellii* plants were harvested, placed in 1.7-ml microfuge tubes, flash-frozen in liquid nitrogen, and stored at $-80^{\circ}C$. Tissues were later powdered by hand under liquid nitrogen in the original collection tubes using plastic micropestles. RNA was extracted from ground leaflets (three biological replicates for each accession) 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 (ThermoFisher 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: $48^{\circ}C$ for 30 min, $95^{\circ}C$ for 10 min, 40 cycles of $95^{\circ}C$ for 15 s and $60^{\circ}C$ for 1 min 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-1 α* , *actin*, and *ubiquitin* genes, respectively (Table 3.1 in Chapter 3 Appendix). For each biological replicate, relative

levels of ASFF1 transcript were determined using the $\Delta\Delta C_t$ method [14] and normalized to the geometric mean of EF-1 α , actin, and ubiquitin transcript levels.

Acylsugar purification

All purifications were performed using a Waters 2795 Separations Module (Waters Corporation) and an Acclaim 120 C18 HPLC column (4.6 x 150 mm, 5 μ m; ThermoFisher Scientific, Waltham, MA) with a column oven temperature of 30°C and flow rate of 2 mL/min. For acylsucrose purification, the mobile phase consisted of water (solvent A) and acetonitrile (solvent B). For acylglucose purification, methanol was used as solvent B. Fractions were collected using a 2211 Superrac fraction collector (LKB Bromma, Stockholm, Sweden).

Purification of acylsucroses was achieved using a ~150 mM solution of acylsugars extracted from the *S. pennellii* LA0716 *asff1-1* mutant which accumulates exclusively acylsucroses. This extract was diluted 14-fold in 70% acetonitrile containing 0.1% formic acid. Acylsucroses were purified using 10 injections of 50 μ L each. A linear gradient of 45% B at 0 min, 60% B at 30 min, 100% B at 30.01 min held until 35 min, and 45% B at 35.01 min held until 40 min was used. Fractions were collected at 10-s intervals into tubes containing 300 μ L 0.1% formic acid in water. The S3:12(4,4,4) compound eluted between 1 and 2 min; the S3:18(4,4,10)-1 compound eluted between 12 and 14 min; the S3:18(4,4,10)-2 compound eluted between 14 and 16 min; the S3:19(4,5,10)-1 compound eluted between 17 and 19 min; and the S3:19(4,5,10)-2 compound eluted between 19 and 21 min.

Purification of acylglucoses was achieved using a ~500 mM solution of acylsugars extracted from *S. pennellii* LA0716, which accumulates > 90% acylglucoses. This extract was diluted 20-fold in 1:1 methanol/water containing 0.1% formic acid. Acylglucoses were purified

using 20 injections of 50 μ L each. A linear gradient of 5% B at 0-1 min, 60% B at 2 min, 100% B at 32 min held until 35 min, and 5% B at 36 min held until 40 min was used. The G3:12(4,4,4) compound eluted between 6 and 7 min; the G3:18(4,4,10)-1 compound eluted between 17 and 18 min; the G3:18(4,4,10)-2 compound eluted between 18 and 19 min; the G3:19(4,5,10)-1 compound eluted between 20 and 21 min; and the G3:19(4,5,10)-2 compound eluted between 21 and 22 min.

Purity of acylsugar fractions was verified by HPLC-MS using an LC-20AD HPLC (Shimadzu, Kyoto, Japan) coupled to a G2-XS QToF mass spectrometer (Waters Corporation). Separations were performed using an Ascentis Express C18 HPLC column (2.1 x 100 mm, 2.7 μ m; Supelco, Bellefonte, PA). The mobile phases consisted of 100 mM ammonium formate, pH 3.4 (solvent A) and 100 mM ammonium formate, pH 3.4, in 90% methanol (solvent B). Five-microliter samples were injected onto the column and eluted with a linear elution gradient of 5% B at 0-1 min, 60% B at 1.01 min, 100% B at 8 min, and 5% B at 8.01-10 min. The solvent flow rate was 0.4 mL/min and the column temperature was 40°C. Analyses were performed using positive-ion mode electrospray ionization and sensitivity mode analyzer parameters. Source parameters were as follows: capillary voltage at 3.00 kV, sampling cone voltage at 40 V, source offset at 80 V, source temperature at 100°C, desolvation temperature at 350°C, cone gas flow at 50.0 L/hour, and desolvation gas flow at 600.0 L/hour. Mass spectrum acquisition (MS^E) was performed over an m/z range of 50 to 1500 with a scan time of 0.5 s. Adduct ions were obtained using a collision energy of 6.0 eV; fragment ions were obtained using a collision energy ramp of 15.0 to 40.0 eV. Spectra were acquired in centroid format.

Pure acylsugar fractions were pooled and solvent removed using a vacuum centrifuge. Samples were reconstituted in 1 mL 3:3:2 acetonitrile/isopropanol/water with 0.1% formic acid, transferred to 2-mL glass autosampler vials, sealed with PTFE-lined caps, and stored at -20°C. Aliquots of purified acylsugars were quantified using the saponification method described above.

NMR spectroscopy

For structural resolution of acylsugars by NMR, ~2.0 mg of each purified acylsugar was prepared. An aliquot of purified acylsugar was evaporated using a vacuum centrifuge and the resulting pellet reconstituted using 0.5 mL of deuteriochloroform (CDCl₃; 99.8 atom % D; Sigma-Aldrich, St. Louis, MO). Solvent was evaporated under N₂ gas and the pellet reconstituted and dried twice more by the same method. Finally, the pellet was reconstituted using 600 µL CDCl₃ (99.96% atom % D, Sigma-Aldrich) and transferred to an NMR tube for analysis. ¹H, gCOSY, gHSQC, gHMBC, and ¹H-¹H JRES spectra were collected at the Max T. Rogers NMR Facility at Michigan State University using a DDR 500 MHz NMR spectrometer (Agilent, Santa Clara, CA) equipped with a 7600AS 96-sample autosampler running VnmrJ v3.2A. ¹³C spectra were collected on the same instrument at 125 MHz. All spectra were referenced to non-deuterated chloroform solvent signals (δ H = 7.26 (s) and δ C = 77.2 (t) ppm). Additional details of the NMR data collection methods can be found in Appendix Table S4.1.

RESULTS

Untargeted metabolomics reveals acylsugars and flavonoids in trichomes

Previous studies indicated that wild *Solanum* species including *Solanum pennellii* exhibit intraspecific variation in the amount and type of acylsugars produced [3,5,15]. To identify

geographic trends in acylsugar quantity and quality in *S. pennellii*, we extracted compounds from the surface of leaflets using six biological replicates from 16 accessions spanning the 1500-km geographic range of the species (Fig. 4.1). We included eight accessions from the northern portion of Peru (North range) and eight from the southern portion (South range). We further classified two clusters of accessions within the South range by region including the southernmost Atico group and the Pisco group. A group of accessions from the Nazca region, described as *S. pennellii* var. *puberulum*, are trichome deficient and exhibit minimal accumulation of acylsugars and transcripts of genes associated with acylsugar metabolism [3,16]; our pilot experiments confirmed the absence of acylsugars in this group (data not shown) and accessions from this group were therefore excluded from this study. All extracts were analyzed by UPLC-HR-MS using positive-mode electrospray ionization.

To survey the intraspecific chemical diversity in *S. pennellii* trichomes, we extracted metabolic features from our UPLC-HR-MS leaf dip extract data using Progenesis Q1 v2.4 software, resulting in detection of a total of 2361 compound ions. Following automated deconvolution of mass spectra to combine adducts resulting from the same compound, 1559 compound ions remained. These compound ions were filtered to remove features representing contaminants introduced during sample extraction, chromatography, or ionization. Of these compound ions, 551 metabolic features showed the highest mean abundance in the process blank samples, 1363 features had a maximum abundance that was < 0.5% of the most abundant compound in the dataset, and 1258 features showed a coefficient of variation > 20% across QC samples. After removal of compound ions falling into one or more of these categories, 54 metabolic features remained. Manual annotation of spectra obtained by

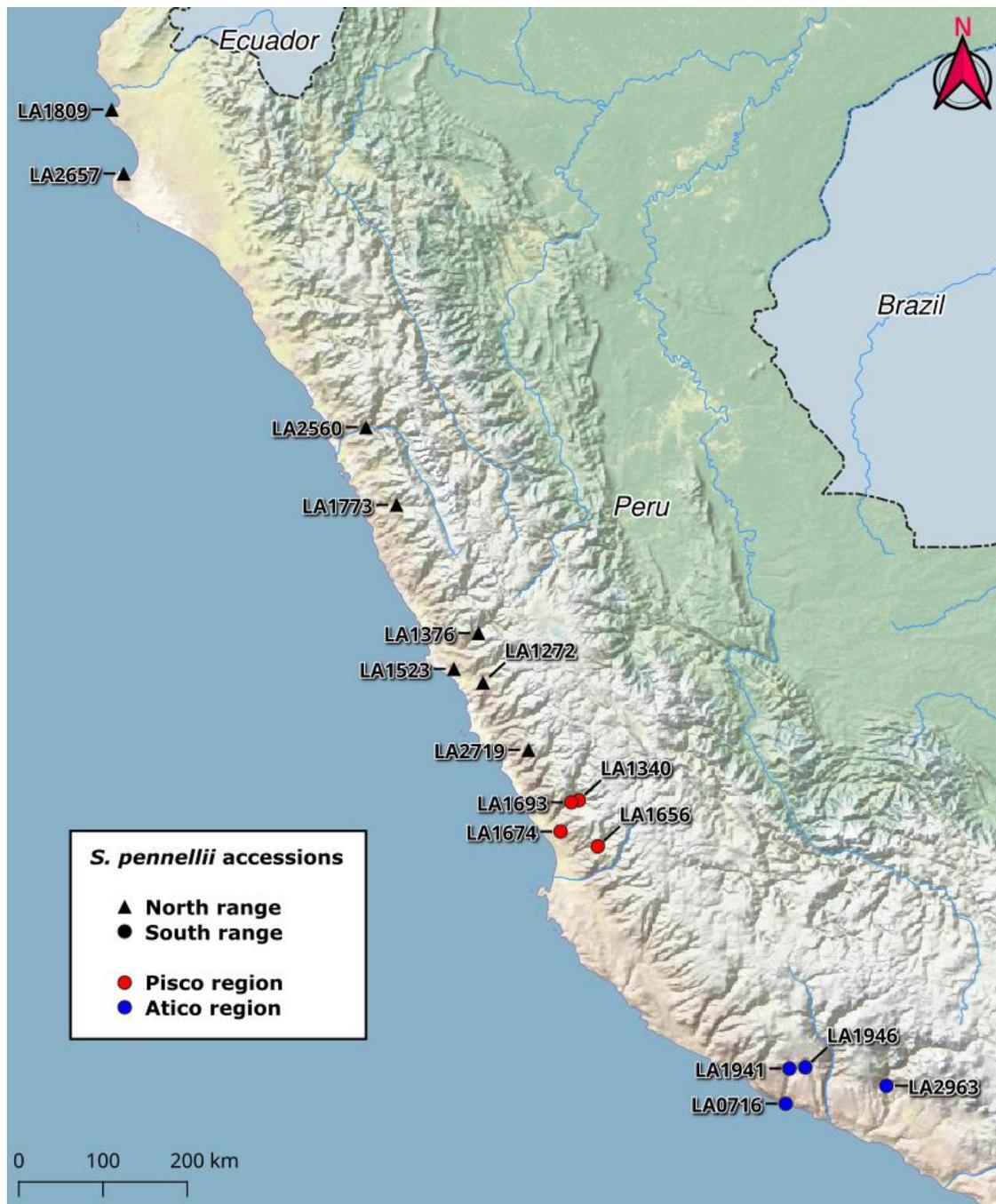


Figure 4.1 Locations of *S. pennellii* accessions used in this study across the geographic range of the species in Peru. Accessions classified as North range are denoted with black triangles, those classified as South range with circles. South range accessions are further classified by region (red for Pisco, blue for Atico).

collision-induced dissociation (CID) and comparison to previously characterized trichome-localized metabolites in *Solanum* spp. [2,6,17] led us to putatively categorize all 54 metabolic features as acylsugars or flavonoids. All acylsugars identified possessed either a six-carbon monosaccharide core or a 12-carbon disaccharide core based on analysis of neutral losses from pseudomolecular ions and m/z of product ions. As all previous studies of *S. pennellii* acylsugars have identified acylglucoses or acylsucroses [2,3,6], we considered these to be the most probable compound classes found in our samples.

Acylglucoses are typically resolved as distinct α and β anomers by reverse-phase UPLC, meaning that two metabolic features may be identified by the *in silico* workflow for each unique acylglucose present in a sample. The two anomers may interconvert under acidic conditions including the mobile phases used for liquid chromatography in this study. In addition, the β anomers of acylglucoses often (but not always) overlap with the α anomers of later-eluting acylglucose structural isomers in UPLC chromatograms, precluding a straightforward arithmetic relationship between the number of acylsugar metabolic features detected and the real number of unique acylglucoses present in a sample. Therefore, chromatograms and associated spectra for all features categorized as acylglucoses were examined to determine the true number of acylglucoses present in the dataset. This resulted in a reduction of the 54 metabolic features identified *in silico* to 43 putatively identified unique compounds. After manual annotation of features and consideration of the chromatographic behavior of acylglucose anomers and structural isomers, we identified 18 triacylsucroses, 19 triacylglucoses, two tetraacylglucoses, and four flavonoids (Tables 4.1 and 4.2). Although our methods are capable of extracting and detecting many core or primary metabolites, these

Table 4.1 Annotations of acylsugars in *S. pennellii*. RT = retention time (min); m/z_{acc} = accurate $[M+NH_4]^+$ mass measured; m/z_{ex} = exact mass calculated from formula; Δm (ppm) = parts per million error between m/z_{ex} and m/z_{acc} ; fragment m/z = ions used for acyl chain determinations.

Name	RT	Formula	m/z_{acc}	m/z_{ex}	Δm (ppm)	Fragment m/z
Triacylsucroses						
S3:12(4,4,4)	2.21	C ₂₄ H ₄₀ O ₁₄	570.2778	570.2756	3.9	373.1872, 285.1326, 197.0809, 127.0395
S3:13(4,4,5)	2.34	C ₂₅ H ₄₂ O ₁₄	584.2926	584.2913	2.2	387.2010, 299.1507, 197.0809, 127.0395
S3:14(4,5,5)	2.59	C ₂₆ H ₄₄ O ₁₄	598.3075	598.3069	1.0	401.2178, 313.1668, 211.0951, 127.0396
S3:15(5,5,5)	3.00	C ₂₇ H ₄₆ O ₁₄	612.3229	612.3226	0.5	415.2348, 313.1661, 211.0974, 127.0395
S3:16(5,5,6)	3.44	C ₂₈ H ₄₈ O ₁₄	626.3392	626.3382	1.6	429.2489, 327.1810, 211.0946, 127.0373
S3:16(4,4,8)	3.67	C ₂₈ H ₄₈ O ₁₄	626.3387	626.3382	0.8	429.2489, 285.1360, 197.0809, 127.0395
S3:17(4,5,8)	4.23	C ₂₉ H ₅₀ O ₁₄	640.3543	640.3539	0.6	443.2710, 299.1541, 211.0974, 127.0395
S3:17(4,4,9)	4.47	C ₂₉ H ₅₀ O ₁₄	640.3536	640.3539	-0.5	443.2646, 285.1341, 197.0773, 127.0396
S3:18(4,4,10)-1	5.45	C ₃₀ H ₅₂ O ₁₄	654.3699	654.3695	0.6	457.2864, 285.1360, 197.0837, 127.0395
S3:18(4,4,10)-2	5.71	C ₃₀ H ₅₂ O ₁₄	654.3699	654.3695	0.6	457.2864, 285.1360, 197.0837, 127.0395
S3:19(4,5,10)-1	6.24	C ₃₁ H ₅₄ O ₁₄	668.3856	668.3852	0.6	471.3013, 299.1507, 211.0974, 127.0395
S3:19(4,5,10)-2	6.52	C ₃₁ H ₅₄ O ₁₄	668.3855	668.3852	0.5	471.3013, 299.1507, 211.1003, 127.0395
S3:20(5,5,10)	7.57	C ₃₂ H ₅₆ O ₁₄	682.4011	682.4008	0.4	485.3103, 313.1661, 211.0974, 127.0395
S3:20(4,4,12)	8.20	C ₃₂ H ₅₆ O ₁₄	682.4009	682.4008	0.2	485.3146, 285.1360, 197.0837, 127.0395
S3:21(5,5,11)	8.56	C ₃₃ H ₅₈ O ₁₄	696.4166	696.4165	0.1	499.3259, 313.1661, 211.0974, 127.0395
S3:21(4,5,12)	9.14	C ₃₃ H ₅₈ O ₁₄	696.4161	696.4165	-0.6	499.3259, 299.1472, 211.0974, 127.0395
S3:22(5,5,12)	10.26	C ₃₄ H ₆₀ O ₁₄	710.4319	710.4321	-0.3	513.3442, 313.1661, 211.0974, 127.0395

Table 4.1 (cont'd)

Name	RT	Formula	m/z_{acc}	m/z_{ex}	Δm (ppm)	Fragment m/z
Triacylsucroses (cont'd)						
S3:23(5,6,12)	11.24	C ₃₅ H ₆₂ O ₁₄	724.4471	724.4478	-1.0	527.3616, 327.1810, 211.0946, 127.0373
Triacylglucoses						
G3:12(4,4,4)	2.76; 2.84	C ₁₈ H ₃₀ O ₉	408.2235	408.2228	1.7	373.1872, 285.1326, 197.0809, 127.0395
G3:13(4,4,5)	3.12; 3.24	C ₁₉ H ₃₂ O ₉	422.2392	422.2385	1.7	387.2014, 299.1501, 197.0801, 127.0396
G3:14(4,5,5)	3.70; 03.83	C ₂₀ H ₃₄ O ₉	436.2547	436.2541	1.4	401.2178, 299.1466, 211.0951, 127.0374
G3:15(5,5,5)	4.42; 4.58	C ₂₁ H ₃₆ O ₉	450.2705	450.2698	1.6	415.2308, 313.1626, 211.0974, 127.0395
G3:16(5,5,6)	5.23; 5.41	C ₂₂ H ₃₈ O ₉	464.2859	464.2854	1.1	429.2489, 327.1810, 211.0946, 127.0395
G3:16(4,4,8)-1	5.56; 5.80	C ₂₂ H ₃₈ O ₉	464.2861	464.2854	1.5	429.2529, 285.1360, 197.0809, 127.0395
G3:16(4,4,8)-2	5.80; 6.04	C ₂₂ H ₃₈ O ₉	464.2857	464.2854	0.7	429.2529, 285.1360, 197.0809, 127.0395
G3:17(4,5,8)-1	6.46; 6.72	C ₂₃ H ₄₀ O ₉	478.3011	478.3011	0.00	443.2628, 299.1472, 211.0946, 127.0395
G3:17(4,5,8)-2	6.71; 6.99	C ₂₃ H ₄₀ O ₉	478.3008	478.3011	-0.6	443.2628, 299.1472, 197.0781, 127.0373
G3:18(4,4,10)-1	8.03; 8.33	C ₂₄ H ₄₂ O ₉	492.3168	492.3167	0.2	457.2779, 285.1326, 197.0809, 127.0395
G3:18(4,4,10)-2	8.33; 8.64	C ₂₄ H ₄₂ O ₉	492.3170	492.3167	0.6	457.2779, 285.1326, 197.0809, 127.0395
G3:19(4,5,10)-1	9.05; 9.34	C ₂₅ H ₄₄ O ₉	506.3328	506.3324	0.8	471.2970, 299.1472, 211.0974, 127.0395
G3:19(4,5,10)-2	9.34; 9.66	C ₂₅ H ₄₄ O ₉	506.3328	506.3324	0.8	471.2970, 299.1507, 211.0946, 127.0395
G3:20(5,5,10)	10.47; 10.72	C ₂₆ H ₄₆ O ₉	520.3486	520.3480	1.2	485.3146, 313.1661, 211.0974, 127.0395
G3:20(4,4,12)	11.10; 11.42	C ₂₆ H ₄₆ O ₉	520.3483	520.3480	0.6	485.3103, 285.1326, 197.0809, 127.0395
G3:21(5,5,11)	11.47; 11.75	C ₂₇ H ₄₈ O ₉	534.3637	534.3637	0.00	499.3215, 313.1626, 211.0974, 127.0373
G3:21(4,5,12)	12.10; 12.40	C ₂₇ H ₄₈ O ₉	534.3636	534.3637	-0.2	499.3290, 299.1507, 211.0974, 127.0395

Table 4.1 (cont'd)

Name	RT	Formula	m/z_{acc}	m/z_{ex}	Δm (ppm)	Fragment m/z
Triacylglycerols (cont'd)						
G3:22(5,5,12)	13.10; 13.36	C ₂₈ H ₅₀ O ₉	548.3794	548.3793	0.2	513.3442, 313.1661, 211.0974, 127.0395
G3:23(5,6,12)	14.02; 14.29	C ₂₉ H ₅₂ O ₉	562.3938	562.3950	-2.1	527.3471, 327.1848, 211.0960, 127.0372
Tetraacylglycerols						
G4:14(2,4,4,4)	3.54; 3.79	C ₂₀ H ₃₂ O ₁₀	450.2343	450.2334	2.0	415.1946, 327.1417, 239.0891, 197.0809, 127.0373
G4:15(2,4,4,5)	4.10; 4.45	C ₂₁ H ₃₄ O ₁₀	464.2502	464.2491	2.4	429.2162, 341.1562, 239.0922, 197.0837, 127.0395

Table 4.2 Annotations of flavonoids in *S. pennellii*. RT = retention time (min); m/z_{acc} = accurate $[M+H]^+$ mass measured; m/z_{ex} = exact mass calculated from formula; Δm (ppm) = parts per million error between m/z_{ex} and m/z_{acc} ; core = putative flavonol core based on molecular formula; # Me = number of methyl groups based on molecular formula and mass spectrum (Appendix Fig. S4.2).

Name	RT	Formula	m/z_{acc}	m/z_{ex}	Δm (ppm)	Core	# Me
Flavonoid A	3.04	C ₁₇ H ₁₄ O ₆	315.0869	315.0863	1.9	kaempferol	2
Flavonoid C	3.17	C ₁₈ H ₁₆ O ₇	345.0980	345.0969	3.2	quercetin	3
Flavonoid D	4.00	C ₁₉ H ₁₈ O ₇	359.1137	359.1125	3.3	quercetin	4
Flavonoid B	4.90	C ₁₈ H ₁₆ O ₆	329.1025	329.102	1.5	kaempferol	3

presumably occur at levels much lower than the acylsugars which accumulate to 10-20% leaf dry weight in our analysis and were therefore not detected following the 100-fold dilution of extracts necessary to accurately measure acylsugars. As such, our dataset consisted entirely of specialized metabolites, most of which were acylsugars.

Acylsugar core composition varies across the range of S. pennellii

To assess the quality of our dataset against previously published analyses of *S. pennellii* acylsugars, we quantified total levels of acylsucroses and acylglucoses. Quantification was performed by comparison of extracted ion chromatograms of m/z values for $M+NH_4$ adducts of all acylsugar features to standard curves of purified acylsucrose and acylglucose compounds. Quantification of acylsucroses and acylglucoses revealed wide variation both in total accumulation of acylsugars (from 133 $\mu\text{mol/g}$ dry weight (DW) in LA2657 to 340 $\mu\text{mol/g}$ DW in LA2560), and in relative abundance of acylglucoses and acylsucroses (from 42% acylglucoses in LA2963 to 95% acylglucoses in LA0716) (Table 4.3; Fig. S4.1A,B). We found no discernable geographic trends in total acylsugar accumulation (Table 4.3; Fig. S4.1A). However, consistent with the results of Shapiro and co-workers [3], we observed a trend towards higher relative abundance of acylglucoses in southern accessions compared with northern accessions. In the northern span of the range, acylglucose composition varied from 56% (LA2657) to 70% (LA2719), while in the southern span, values ranged from 77% (LA1693) to 95% (LA0716) acylglucose; the South range accession LA2963 is a notable exception to this trend, showing a lower acylglucose composition (42%) than any other accession (Table 4.3; Fig. S4.1B).

Table 4.3 Acylsugar accumulation and percent acylglucose in accessions of *S. pennellii* as determined by UPLC-MS-MS. Values are presented as \pm SD. Results of ANOVA and Tukey's mean-separation test are indicated as letters. Accessions that do not have at least one letter in common are significantly different from one another. The range and region of each accession within Peru is also indicated.

Accession	Total acylsugars ($\mu\text{mol/g DW}$)	Tukey's MST	% acylglucose	Tukey's MST	Range	Region
LA1809	136 \pm 27	B	69 \pm 4	CDE	North	
LA2657	133 \pm 28	B	56 \pm 6	EF	North	
LA2560	340 \pm 64	A	65 \pm 6	CDE	North	
LA1773	237 \pm 98	AB	66 \pm 2	CDE	North	
LA1376	261 \pm 105	AB	70 \pm 7	CDE	North	
LA1523	158 \pm 68	B	65 \pm 8	CDE	North	
LA1272	163 \pm 109	B	58 \pm 4	DEF	North	
LA2719	218 \pm 44	AB	70 \pm 3	BCDE	North	
LA1340	166 \pm 39	B	80 \pm 10	ABC	South	Pisco
LA1693	193 \pm 88	AB	77 \pm 7	ABCD	South	Pisco
LA1674	248 \pm 83	AB	90 \pm 4	A	South	Pisco
LA1656	269 \pm 54	AB	90 \pm 4	AB	South	Pisco
LA1946	257 \pm 94	AB	82 \pm 20	ABC	South	Atico
LA1941	244 \pm 66	AB	95 \pm 2	A	South	Atico
LA2963	183 \pm 38	B	42 \pm 6	F	South	Atico
LA0716	238 \pm 75	AB	95 \pm 2	A	South	Atico

Variable acyl chain and sugar composition yield acylsugar diversity

We sought to characterize the compounds constituting acylsugar diversity in *S. pennellii*. Previous studies of acylsugar metabolism in this species reported acylsugar diversity in terms of the total abundance of sugar cores and acyl chains in derivatized acylsugar extracts but did not analyze the intact acylsugars that accumulate *in planta* [3,5,8]. Using CID spectra, we analyzed both the acyl chain composition and the sugar core of each intact acylsugar present in the dataset, creating a more complete picture of acylsugar diversity. Molecular formulas were determined by comparing accurate m/z values of $[M+NH_4]^+$ pseudomolecular ions to theoretical m/z values of hypothetical acylsugar $[M+NH_4]^+$ adducts. The 39 acylsugars present in the dataset were described by 26 unique molecular formulas, indicating the presence of several structural isomers. The molecular formulas of all acyl chain components from individual acylsugars were inferred by a similar process using ketene and fatty acid neutral losses from pseudomolecular precursor ions observed in the high-energy CID function. While acylium product ions representing acyl chains appear in many spectra, their occurrence is inconsistent across compounds, especially in those of low abundance. Therefore, all acyl chain assignments were made using the neutral loss data, which could be unambiguously interpreted for all spectra. Using this method, we annotated all 39 acylsugars identified in our dataset in terms of putative sugar core, number of acyl chains, and number of carbons in each acyl chain (Table 4.1).

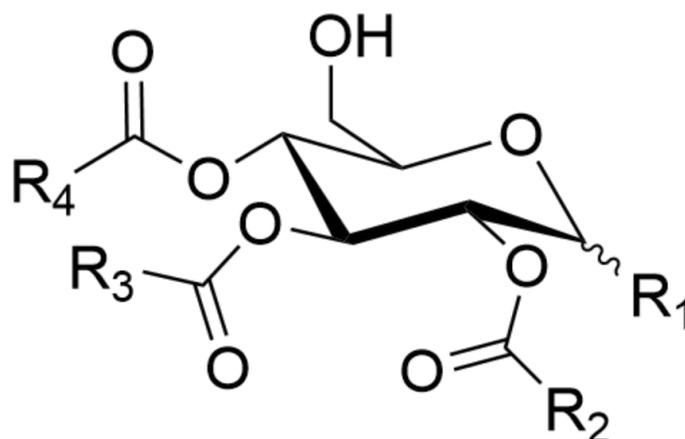
Spectral annotation of all acylsugars present in the dataset revealed three notable characteristics of the acylsugar profile. First, while some pairs of non-anomeric structural isomers clearly differ in the number of carbons in their individual constituent acyl chains (*e.g.*,

S3:16(5,5,6) vs. S3:16(4,4,8) and G3:21(5,5,11) vs. G3:21(4,5,12); see Table 4.1 for additional examples), six pairs of structural isomers that are clearly resolved by reverse-phase HPLC have indistinguishable mass spectra (*e.g.*, S3:18(4,4,10)-1 eluting at 5.45 min and S3:18(4,4,10)-2 eluting at 5.71 min; see Table 4.1 for additional examples). This suggests two additional dimensions of possible acylsugar structural isomerism that are not mutually exclusive: acylsugars with similar complements of acyl chains but differing in terms of the positions at which these acyl chains are attached (positional isomers), and acylsugars bearing acyl chains with identical chemical formulas but different branching patterns. The latter hypothesis is supported by previous reports of unbranched, *iso*-branched, and *anteiso*-branched acyl chains in *S. pennellii* acylsugars [3,5,18]. A second notable finding was detection of two tetraacylglucoses, G4:14(2,4,4,4) and G4:15(2,4,4,5). Tetraacylated sugars were not previously reported in *S. pennellii* and this observation suggests the presence of an additional ASAT activity (either a second acylation activity catalyzed by a previously described ASAT or an entirely new ASAT enzyme) beyond those described in the species to date [7,19]. Third, all but one of the annotated triacylsucroses [S3:17(4,4,9)] show a pattern of acyl group neutral losses that mirror neutral losses observed in at least one triacylglucose. We hypothesized that pairs of acylsucroses and acylglucoses with similar fragmentation patterns possessed identical acyl chain complements, consistent with the current model of *S. pennellii* acylsugar biosynthesis in which the β -fructofuranose rings of acylsucroses are enzymatically cleaved off, yielding acylglucoses [8]. These observations indicate that variation in the identity of acyl chains, number of acyl chains, and identity of sugar core all contribute to the large number of acylsugars found in *S. pennellii*. While the presence of multiple acylsugar structural isomers

with identical mass spectra implies the presence of isomeric acyl chains, and the similarity in neutral loss patterns between acylsucroses and acylglucoses suggests identical acyl chain composition, the soft-ionization mass spectrometry techniques applied here were unable to test these hypotheses.

NMR spectroscopy shows structural relationships among compounds with similar mass spectra

To better understand the structural features underlying acylsugar diversity in *S. pennellii* and test our hypotheses regarding acylsugar isomerism and the structural relationship between acylsucroses and acylglucoses with similar mass spectra, we selected 10 acylsugars for purification and structural resolution by NMR, including five acylsucroses (S3:12(4,4,4), S3:18(4,4,10)-1, S3:18(4,4,10)-2, S3:19(4,5,10)-1, and S3:19(4,5,10)-2) and five acylglucoses (G3:12(4,4,4), G3:18(4,4,10)-1, G3:18(4,4,10)-2, G3:19(4,5,10)-1, and G3:19(4,5,10)-2) (Table 4.1; Fig. 4.2; See Appendix Tables S4.2-11, Figs. S4.3-62 for NMR chemical shifts and spectra). NMR spectroscopy confirmed that all disaccharide-containing acylsugars possess a sucrose core while all monosaccharide-containing acylsugars are based on glucose, consistent with previous analyses of *S. pennellii* acylsugars [2,3,6]. Our analysis further indicated that all acylsugars examined are acylated at the R2, R3, and R4 positions of the pyranose ring, also consistent with previous reports [2,6]. The structures of two compounds, G3:12(4,4,4) and S3:19(4,5,10)-1, matched two previously published acylsugar structures [2,6]. To test the hypotheses that acylsugar isomers with identical mass spectra possess either identical complements of acyl chains attached to different positions of the sugar core or isomeric acyl chains with different branching patterns, we compared the structures of four pairs of structural isomers including two pairs of acylsucrose isomers and two pairs of acylglucose isomers (S3:18(4,4,10)-1/2,



- S3:12(4,4,4)** R₁ = Fru; R₂ = R₃ = R₄ = Me₂CH
- S3:18(4,4,10)-1** R₁ = Fru; R₂ = R₄ = Me₂CH; R₃ = Me₂CH(CH₂)₆
- S3:18(4,4,10)-2** R₁ = Fru; R₂ = R₄ = Me₂CH; R₃ = Me(CH₂)₈
- S3:19(4,5,10)-1** R₁ = Fru; R₂ = MeCH₂MeCH; R₃ = Me₂CH(CH₂)₆; R₄ = Me₂CH
- S3:19(4,5,10)-2** R₁ = Fru; R₂ = MeCH₂MeCH; R₃ = Me(CH₂)₈; R₄ = Me₂CH
- G3:12(4,4,4)** R₁ = OH; R₂ = R₃ = R₄ = Me₂CH
- G:18(4,4,10)-1** R₁ = OH; R₂ = R₄ = Me₂CH; R₃ = Me₂CH(CH₂)₆
- G:18(4,4,10)-2** R₁ = OH; R₂ = R₄ = Me₂CH; R₃ = Me(CH₂)₈
- G:19(4,5,10)-1** R₁ = OH; R₂ = MeCH₂MeCH; R₃ = Me₂CH(CH₂)₆; R₄ = Me₂CH
- G:19(4,5,10)-2** R₁ = OH; R₂ = MeCH₂MeCH; R₃ = Me(CH₂)₈; R₄ = Me₂CH

Figure 4.2 NMR-resolved structures of acylsugars purified from *S. pennellii*. For acylsucroses, position R₁ is found only in the α configuration. For acylglucoses, position R₁ exists in both α and β forms. Fru = β-fructofuranose.

S3:19(4,5,10)-1/2, G3:18(4,4,10)-1/2, G3:19(4,5,10)-1/2). In each case, both isomers had identical configurations of acyl chains at the R2 and R4 positions. However, all four isomeric pairs acyl chains with the same molecular formula but different branching patterns at the R3 position. For all isomer pairs tested, we observed an *iso*-branched 10-carbon acyl chain (R3 = (Me)₂CH(CH₂)₆) in the earlier-eluting isomer and an unbranched 10-carbon acyl chain (R3 = Me(CH₂)₈) in the later-eluting isomer (Fig. 4.2). Combined with annotation of acylsugar mass spectra (Table 4.1), this demonstrates that acylsugar diversity is influenced not only by variation in the molecular formulas of constituent acyl chains but also by variation in acyl chain branching patterns. We also compared the structures of acylsucroses and acylglucoses with similar neutral loss patterns. We found that the acylation pattern of each of the five purified acylsucroses was identical to that of one of the five purified acylglucoses (*e.g.*, S3:12(4,4,4) and G3:12(4,4,4); S3:19(4,5,10)-1 and G3:19(4,5,10)-1; Fig. 4.2). This is consistent with some of the acylsucroses observed in this dataset being intermediates in acylglucose biosynthesis. These NMR-resolved acylsugar structures corroborate findings suggested by UPLC-HR-MS data and indicate that variation in acylsugar branching patterns contributes to the presence of acylsugar structural isomers and that similar neutral loss patterns in acylsucroses and acylglucoses sometimes reflect identical acyl chain complements.

Flavonoids vary by core and degree of methylation

While 39 distinct acylsugars were present in our dataset, we identified only four flavonoids. All four compounds were found at low levels and only a few abundant peaks appear in their mass spectra for interpretation. However, molecular formulas for all four compounds were obtained by comparison of the accurate *m/z* values measured for [M+H]⁺

pseudomolecular ions to hypothetical exact m/z values. These formulas are consistent with di-, tri-, and tetramethylated derivatives of tetra- and pentahydroxylated flavonols (Table 4.2), resembling the methylated myricetins observed in *S. habrochaites* and *S. lycopersicum* [17,20,21]. As *S. lycopersicum* accumulates glycosylated derivatives of the flavonols kaempferol and quercetin (tetra- and pentahydroxylated, respectively) in type VI trichomes [21,22], we hypothesized that the methylated flavonoids observed in *S. pennellii* leaf dips were also kaempferol- and quercetin-derived. Analysis of flavonoid mass spectra indicated loss of methyl radicals, further supporting the presence of methyl groups on these compounds (Fig. S4.2). Two kaempferol-like flavonoids were observed possessing two and three methylations (denoted as flavonoids A and B), while two quercetin-like flavonoids were observed possessing three and four methylations (flavonoids C and D). Few low-mass fragment ions were present in the spectra to aid in assignment of methyl group positions as previously demonstrated with myricetin derivatives [23]. Nevertheless, our results indicated flavonoid diversity in terms of both flavonol core and degree of methylation.

Multivariate analysis implicates short branched acyl chains in North-South acylsugar variation

The complete structural characterization of 10 compounds and annotation of an additional 33 compounds in leaf dip extracts of *S. pennellii* including acylsugars and flavonoids revealed multiple dimensions of chemical diversity, including variation in sugar core and acylation pattern in acylsugars, and variation in degree of hydroxylation and methylation in flavonoids. Variation in sugar core and acyl chain composition of acylsugars from different *S. pennellii* accessions was previously shown to exhibit geographic trends [3,5] (Table 4.3, Fig. S4.1B). While these trends based on aggregate characteristics of a chemical class such as sugar

core and acyl chain abundances provide insight into some of the metabolic underpinnings of specialized metabolite diversity in *S. pennellii* [5] (Table 4.3; Fig. 4.2), we sought to extend this analysis to the level of individual metabolites. We used our full dataset representing 43 specialized metabolites in 16 accessions of *S. pennellii* to identify metabolite-based differences between accessions in different parts of the geographic range. Due to overlapping retention times observed with some acylglucose anomers (described above) and the resulting difficulty in assigning accurate abundances to individual acylglucoses, we used the original dataset containing 54 metabolite features obtained prior to spectral interpretation instead of the dataset containing the 43 unique metabolites.

Unsupervised principal component analysis (PCA) of all accessions revealed clear separation of accessions in the North range from those in the South range with the exception of two outliers (Fig. 4.3). These samples both represent individuals of South range accession LA1946 that cluster with North range samples; we hypothesize that this is due to seed contamination or sample identification error during the metabolite extraction and sample preparation process. Principal component 1 (PC1) accounted for approximately 45% of the variance in the dataset and drove strong separation between North and South accessions, while PC2 accounted for 18% of the variance and associated primarily with variation within the South range accessions with little consistent difference between North and South range samples apparent.

The clear separation of samples from northern and southern accessions prompted us to execute a supervised analysis of the North and South range groups using orthogonal partial least squares/projection to latent structures discriminant analysis (OPLS-DA) to identify the

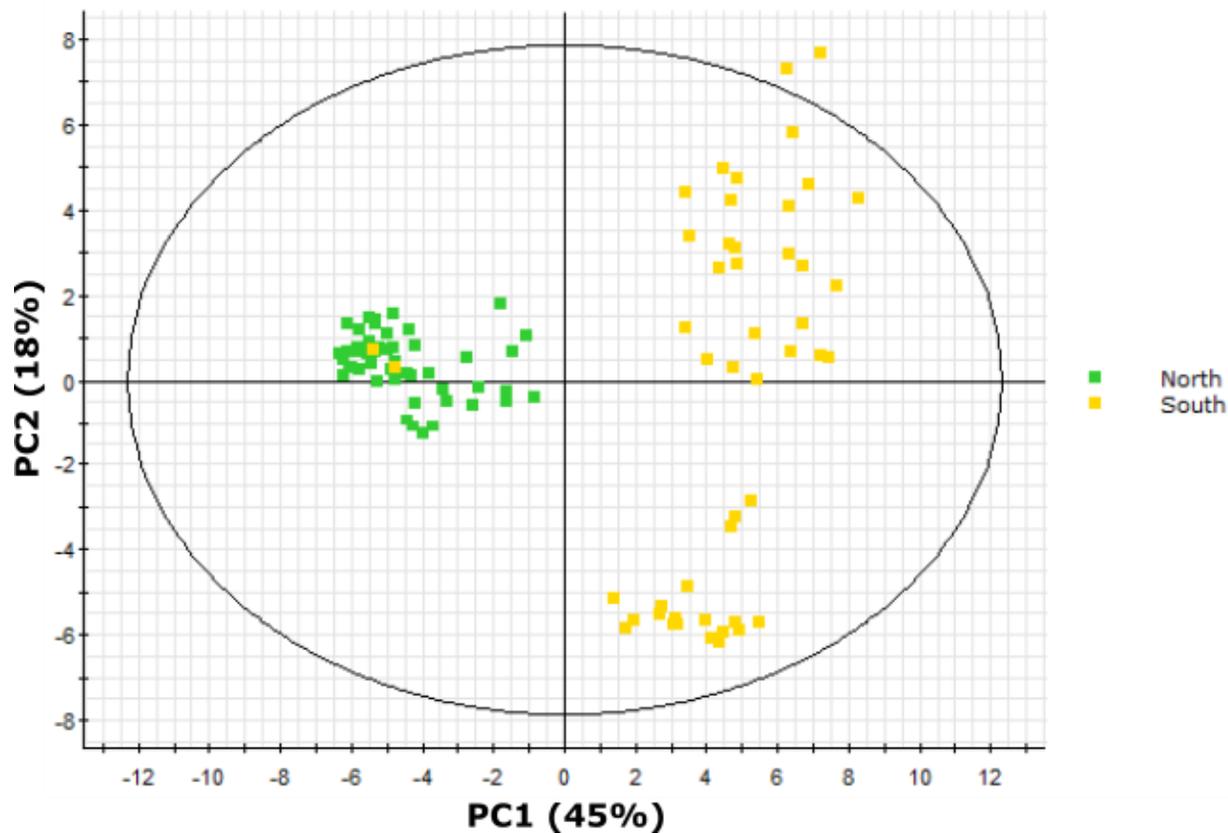


Figure 4.3 PCA scores plot of samples from 16 *S. pennellii* accessions from across Peru separated by abundances of 54 metabolite features identified in trichome extracts by UPLC-HR-MS. Samples from the North range are indicated in green, while samples from the South range are indicated in yellow (See Fig. 4.1 for details on geographic range).

metabolite features that drive this separation. Our model successfully classified 100% of North range samples and 94% of South range samples (Table 4.4), indicating that metabolite features identified by the model were good predictors of a sample's geographic origin. Five metabolites demonstrated the strongest correlation with the North and South range samples (Table S4.12): three acylglucoses [G3:15(5,5,5), G3:16(5,5,6), G3:21(5,5,11)] and two acylsucroses [S3:16(5,5,6), S3:21(5,5,11)] are most characteristic of North range accessions, while four acylglucoses (G3:12(4,4,4), G3:13(4,4,5), G3:18(4,4,10)-2, G3:19(4,5,10)-2) and one acylsucrose (S3:18(4,4,10)-2), are most characteristic of South range accessions. Four acylsugars associated with the South range (G3:12(4,4,4), G3:18(4,4,10)-2, G3:19(4,5,10)-2, S3:18(4,4,10)-2) were structurally characterized by NMR in this study (Fig. 4.2), while a fifth [G3:14(4,5,5)] was characterized in previous work [6]. All four-carbon acyl chains in these acylsugars are 2-methylpropanoate, while only one of the five-carbon chains in the G3:14(4,5,5) compound is 3-methylbutanoate (the other five-carbon acyl chains in G3:14(4,5,5) and G3:19(4,5,10)-2 are 2-methylbutanoate). While we cannot definitively identify the branching pattern of five-carbon acyl chains in the metabolites associated with the North range, our findings agree with previously observed trends in *S. pennellii* favoring accumulation of four-carbon 2-methylpropanoate chains in southern accessions and five-carbon 3-methylbutanoate chains in northern accessions, with five-carbon 2-methylbutanoate chains abundant across the range [3,5].

Variation in medium-length acyl chains drives variation within the South range

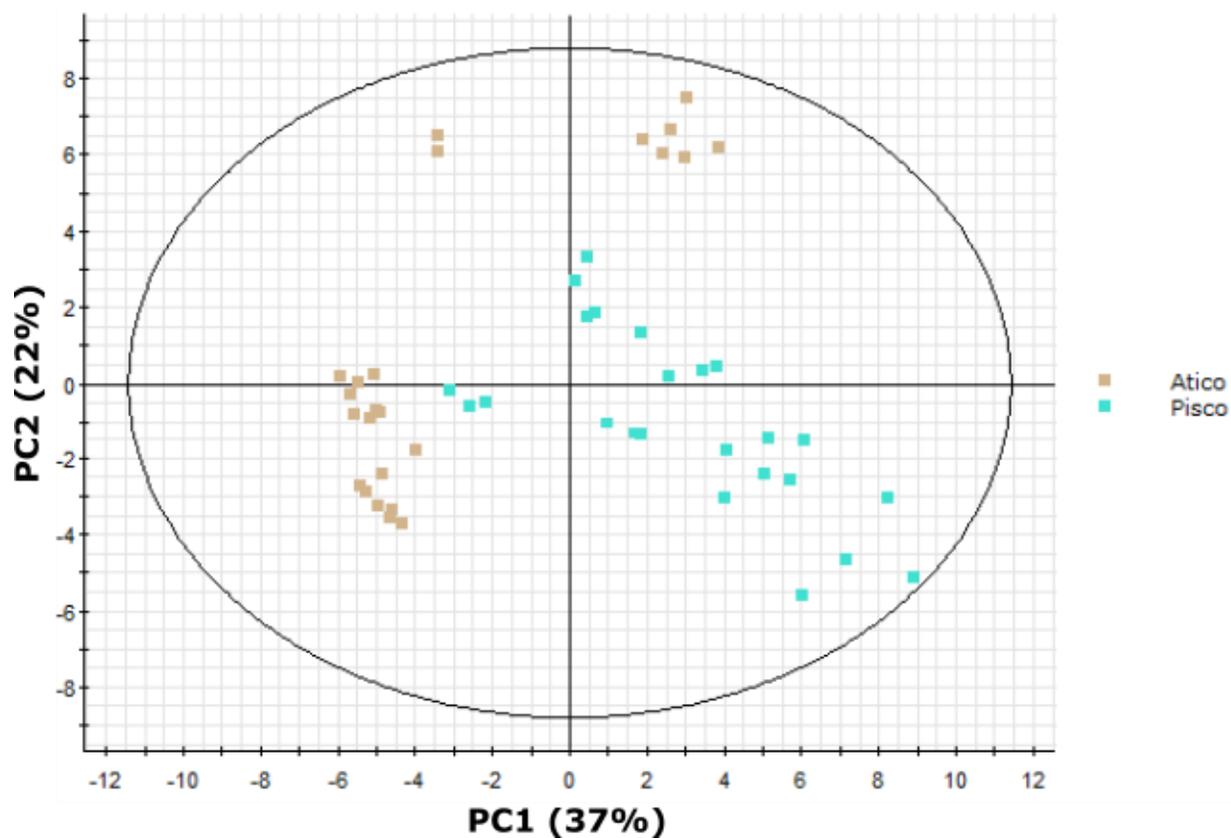
As our PCA also indicated substantial intragroup variation in South range samples (Fig. 4.3), we performed additional multivariate analyses using exclusively South range accessions.

Table 4.4 OPLS-DA model performance. The table indicates the percentage of test samples that each model classified correctly, incorrectly, or was unable to classify.

	% Correct	% Incorrect	% Unknown
Full range			
North	100	0	0
South	94	4	2
South range			
Pisco	67	4	29
Atico	77	4	19
Atico region			
LA0716/LA1941/LA1946	97	0	3
LA2963	100	0	0

These accessions form two distinct geographic clusters denoted by the Pisco and Atico regional assignments (Fig. 4.1). PCA of these accessions indicated separation of Atico and Pisco samples but with an obvious bimodality observed in samples from the Atico region (Fig. 4.4). PC1 accounted for 37% of variance while PC2 accounted for 22%; however, neither PC represented separation of Pisco samples from all Atico samples. Instead, PC1 appeared to drive separation of Pisco samples from the largest cluster of Atico samples while PC2 drove separation of Pisco samples from the smaller Atico cluster.

To identify the metabolite features driving separation of *S. pennellii* accessions from the Pisco and Atico regions, we performed OPLS-DA. Our model successfully classified 67% of Pisco region samples and 77% of Atico region samples but misclassified or was unable to classify 28% of all samples (Table 4.4), indicating that this model performed relatively poorly in terms of separating samples by geographic region when compared to our North/South range OPLS-DA model. This may be due to the large proportion of orthogonal variation in the Atico region samples. However, we were still able to identify metabolites that had strong correlation with either the Pisco or Atico region samples (Table S4.13). The top five compounds demonstrating strong correlation with Pisco region samples comprised four acylglucoses [G3:16(4,4,8)-1, G3:16(4,4,8)-2, G3:17(4,5,8)-1, and G3:17(4,5,8)-2] and one acylsucrose [S3:17(4,4,9)], while the top five metabolites correlating with the Atico region samples consisted of two acylglucoses [G3:20(4,4,12), G3:21(4,5,12)], two acylsucroses [S3:20(4,4,12), S3:21(4,5,12)], and one flavonoid (flavonoid A). To our knowledge, no structures of any compounds extracted from *S. pennellii* and matching these structural annotations have been published to date. However, a clear trend is visible in the acylsugar features. While the distribution of short (and presumably



branched) four- and five-carbon acyl chains is similar between correlative features from both regions, the longer acyl chains show a sharp distinction with four of five acylsugars from the Pisco region bearing an eight-carbon acyl chain and all four acylsugars from the Atico region containing a 12-carbon acyl chain.

LA2963 segregates from other Atico region accessions due to high acylsucrose content

The obvious within-group variation evidenced by the bimodal clustering of Atico region samples in our regional PCA (Fig. 4.4) prompted us to explore the metabolic underpinnings of this variation by multivariate analysis. PCA of the four accessions included in the Atico region (LA0716, LA1941, LA1946, and LA2963) showed two major clusters (Fig. 4.5). One cluster contained all biological replicates of accessions LA0716 and LA1941 along with four samples of LA1946 [both LA1946 samples outside this cluster represent outliers that clustered with North range accessions in our North/South PCA (Fig. 4.3)]. The other major cluster contained all samples of accession LA2963. PC1 accounted for 47% of variance and described most of the variation between accession LA2963 samples and other Atico region accessions, while PC2 accounted for 35% of variance and described primarily variation within the main Atico cluster.

As our initial analysis of *S. pennellii* sugar core abundance indicated accession LA2963 as an outlier in the South range (Table 4.3; Appendix Fig. S4.1B), we hypothesized that this sugar core variation, rather than acyl chain variation, might drive PCA separation of accession LA2963 from the other Atico region accessions. To test this hypothesis, we generated an OPLS-DA model discriminating between LA2963 samples and all other Atico region samples. Our model successfully classified 97% of samples from the main Atico cluster and 100% of LA2963 samples (Table 4.4), indicating that metabolite features identified by the model were good predictors of

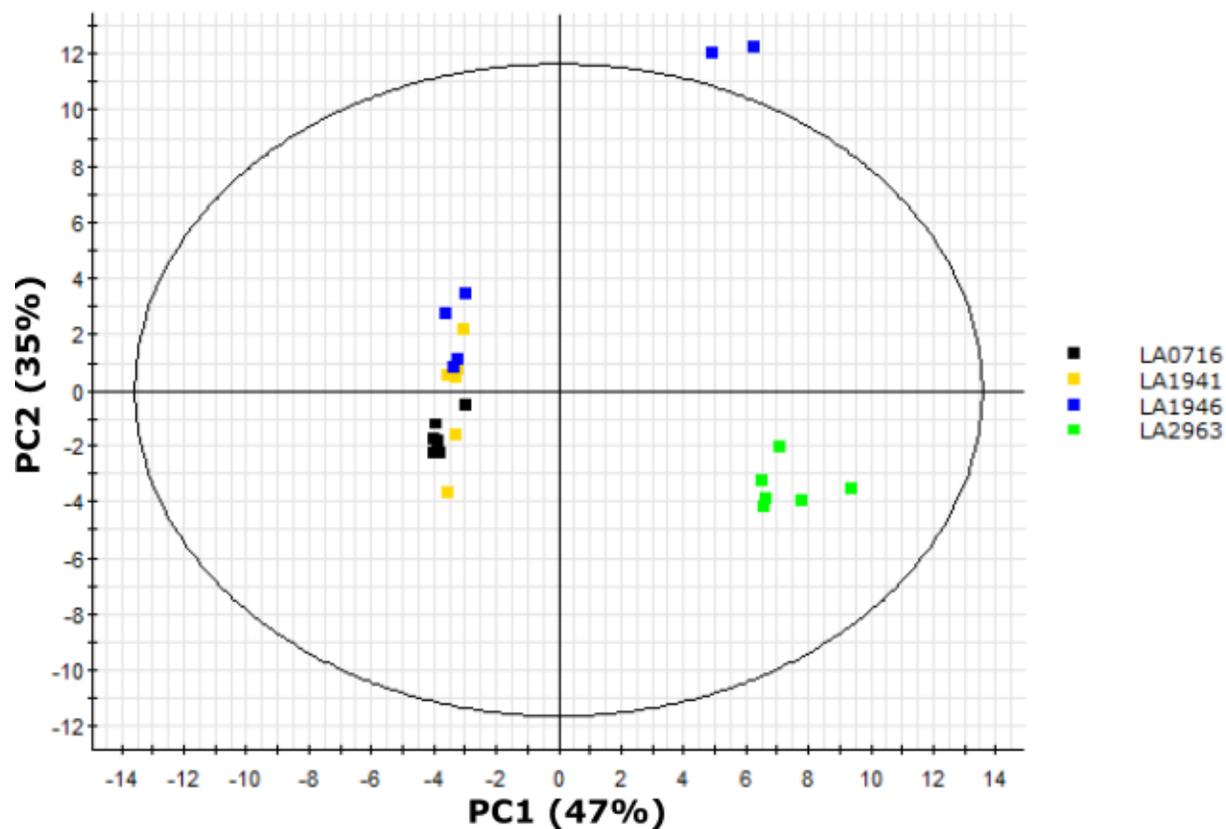


Figure 4.5 PCA scores plot of samples from four *S. pennellii* accessions in the Atico region of Peru separated by abundances of 54 metabolite features identified in trichome extracts by UPLC-HR-MS. Samples from accession LA0716 are indicated in black, samples from accession LA1941 in yellow, samples from accession LA1946 in blue, and samples from accession LA2963 in green.

accession. Examination of the five metabolites possessing the strongest correlation with either the main Atico cluster or the LA2963 cluster provided support for our hypothesis (Table S4.14). Compounds correlated with the main Atico cluster included four acylglucoses [G3:18(4,4,10)-1, G3:18(4,4,10)-2, G3:19(4,5,10)-1, G3:20(4,4,12)] and one flavonoid (flavonoid A) while compounds correlated with accession LA2963 included five acylsucroses [S3:18(4,4,10)-1, S3:19(4,5,10)-1, S3:19(4,5,10)-2, S3:20(4,4,12), S3:21(4,5,12)]. This suggested that relative abundance of acylsucroses and acylglucoses drove the separation of LA2963 samples from other Atico region samples.

The ASFF1 enzyme hydrolyzes acylsucroses to yield acylglucoses in *S. pennellii* LA0716 [8], suggesting a biochemical basis for differential accumulation of acylsucroses and acylglucoses in *S. pennellii* accessions accumulating otherwise similar acylsugars. Two of the acylsucroses correlated with accession LA2963 (S3:18(4,4,10)-1, S3:19(4,5,10)-1) have structures consistent with possible precursors of acylglucoses correlated with the main Atico cluster (G3:18(4,4,10)-1, G3:19(4,5,10)-1; described above; Fig. 4.2), and a third compound correlated with LA2963 [S3:20(4,4,12)] has a fragmentation pattern consistent with a possible precursor of another Atico cluster-correlated acylglucose [G3:20(4,4,12)]. We hypothesized that low ASFF1 activity in plants of accession LA2963 relative to other Atico region accessions contributed to the low accumulation of acylglucoses in this accession and corresponding high accumulation of acylsucroses. To test this hypothesis, we measured relative accumulation of acylglucoses by saponification of acylsugar extracts and LC-MS sugar core quantification, and relative ASFF1 transcript abundance by RT-qPCR in leaflets from three biological replicates of *S. pennellii* LA0716 and LA2963 (Fig. 4.6). Consistent with our previous sugar core quantification

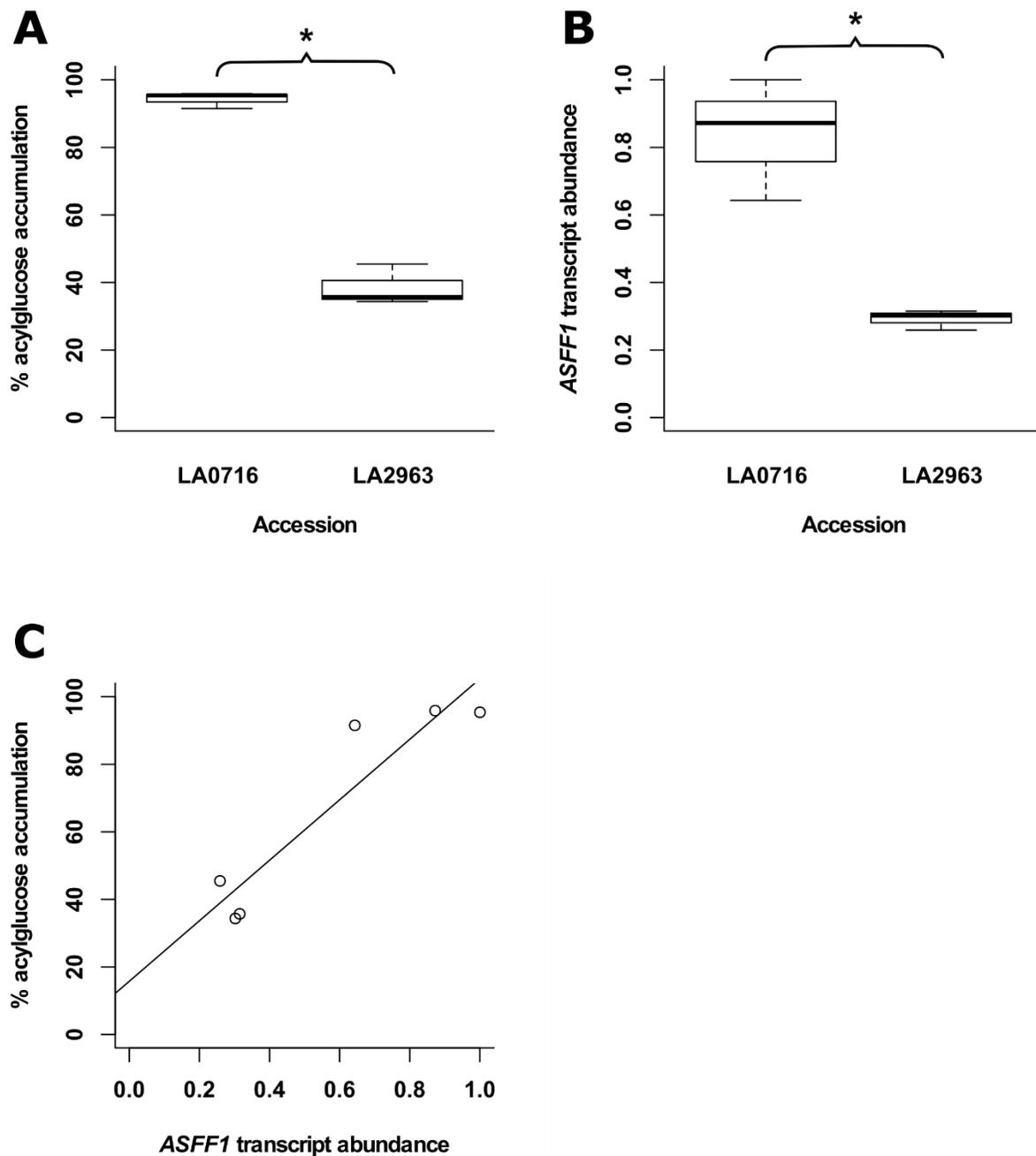


Figure 4.6 Analysis of acylglucose accumulation and *ASFF1* transcript abundance in leaflets of *S. pennellii* accessions LA0716 and LA2963. (A) Percentage of total acylsugars accumulating as acylglucoses. (B) Relative abundance of *ASFF1* transcripts. (C) Linear regression of *ASFF1* transcript abundance and percentage of percentage of acylsugars accumulating as acylglucoses ($R^2 = 0.84$). “*” indicates $p < 0.05$ (ANOVA); $n = 3$ for both accessions.

results (Table 4.3; Appendix Fig. S4.1B), we found that acylglucoses constituted 94% of acylsugars in LA0716 but only 38% of acylsugars in LA2963 (Fig. 4.6A). ASFF1 transcripts were 2.9-fold more abundant in LA0716 than in LA2963 (Fig. 4.6B). Linear regression analysis indicated a positive correlation between ASFF1 transcript abundance and percentage of acylsugars accumulating as acylglucoses ($R^2 = 0.84$; Fig. 4.6C). This correlation supported a role of ASFF1 in acylsugar core variation between accessions of *S. pennellii*. However, this analysis cannot account for post-transcriptional or post-translational regulation of ASFF1 enzyme abundance and differences in enzyme activity. Therefore, while the relationship between ASFF1 transcript levels and acylglucose accumulation is clear, the mechanistic significance is not.

DISCUSSION

Acylsugars constitute an abundant and diverse class of specialized metabolites that accumulate in SGTs and other tissues of Solanaceae species, including *S. pennellii* [2,24–27]. These compounds have demonstrated roles in protecting *Solanum* species against insect pests including silverleaf whitefly (*Bemisia tabaci*), western flower thrips (*Frankliniella occidentalis*), and army beetworm (*Spodoptera exigua*) [28–30]. To capitalize on the protective properties of acylsugars, plant breeders are creating tomato lines with altered acylsugar profiles and increased insect resistance [18,30–32]. Acylsugars accumulate to high levels in *Solanum* species and undergo rapid turnover, indicating that both biosynthetic and degradative processes shape acylsugar phenotype [1,3,8,33]. The core acylsugar biosynthetic pathway comprising the ASAT enzymes was characterized in *S. pennellii* and several other Solanaceae species, while additional enzymes that affect acylsugar phenotype through acylsugar precursor production (*e.g.*, IPMS3, ECH, ACS) or acylsugar turnover (*e.g.*, ASHs, ASFF1) were also characterized in *S. pennellii* and *S.*

lycopersicum [5–8,19,34,35]. Characterization of the acylsugars found in *S. pennellii* is essential for further elucidating elements of acylsugar biosynthesis and degradation pathways. Previously published reports demonstrated how knowledge of variation in acylsugars in *Solanum* species such as sugar core and acyl chain complement between or within species facilitated discovery of pathway genes such as *ASFF1* and *IPMS3* [5,8]. However, these analyses did not capture the intact acylsugar profile present *in planta* and miss intraspecific variations in the levels of individual compounds which may inform pathway elucidation. In contrast, analyses of intact acylsugars in Solanaceae species by UPLC-HR-MS and NMR spectroscopy contributed to characterization of enzymes in the acylsugar pathways of *S. habrochaites* [7,15,24] and *Petunia axillaris* [35,36]. Here, we took a similar approach with *S. pennellii* and characterized specialized metabolites that accumulate on leaf surfaces and in SGTs of the species.

Our analysis of metabolites extracted from the surface of *S. pennellii* leaflets confirmed previously published results describing *S. pennellii* acylsugars and revealed new characteristics of specialized metabolism in the species. We identified a total of 43 specialized metabolites in trichomes consisting of 18 acylsucroses, 21 acylglucoses, and four flavonoids. MS analysis alone indicated the presence of two tetraacylglucoses (Table 4.1), a type of acylsugar previously unknown in *S. pennellii*, as well as four methyl flavonoids (Table 4.2), a class of compounds known from the closely related *S. lycopersicum* and *S. habrochaites* but previously unknown in this species [17,20–23]. A combination of MS and NMR confirmed previously reported geographic trends in accumulation of 2-methylpropanoate and 3-methylbutanoate acyl chains in *S. pennellii* acylsugars (Table 4.1; Fig. 4.2) [3,5]. Multivariate analysis of our full dataset provided additional confirmation of the differential accumulation of short branched acyl chains

in acylsugars from northern and southern *S. pennellii* accessions and revealed the presence of geographic variation between smaller sub-regions within the range of the species (Fig. 4.4; Fig. 4.5). Accessions from the Pisco and Atico regions were distinguished by enrichment of eight-carbon acyl chains in the former and 12-carbon acyl chains in the latter. Finally, we observed that within the Atico region, the acylsugar profile of accession LA2963 shows near one-dimensional variation from those of other nearby accessions, differing primarily by its relatively low abundance of acylglucoses (42%) compared to other Atico region accessions (82-95%). This prompted a comparison of acylglucose accumulation and *ASFF1* transcript levels in accessions LA0716 and LA2963, which indicated a possible role of the ASFF1 enzyme in regulating the relative abundance of sucrose and glucose cores in *S. pennellii* acylsugars. In total, our metabolomic analysis corroborated two known components of intraspecific acylsugar variation in *S. pennellii* (sugar core and short branched acyl chain abundance) and revealed two new types of variation (number of acylations and abundance of eight- and 12-carbon acyl chains). These additional dimensions of acylsugar variation demonstrate that, although the core acylsugar biosynthetic pathway in *S. pennellii* has been elucidated, aspects of acylsugar biosynthesis and degradation remain to be characterized.

The accumulation of tetraacylglucoses in *S. pennellii* was previously unreported and their discovery indicates the presence of a previously unknown ASAT activity. Thus far, three ASATs involved in acylsugar biosynthesis, each performing a single acylation step, have been identified in *S. pennellii* [7]. The presence of tetraacylglucoses in this species requires a fourth acylation step, although structures of these compounds were not resolved and the position of the fourth acylation is unknown. This step could be performed by one of the previously

described acyltransferases from the *S. pennellii* acylsugar pathway (*i.e.*, ASAT1/2/3) or by an acyltransferase not previously implicated in acylsugar biosynthesis. Notably, both tetraacylglucoses identified in this study possess a single acetyl group (Table 4.1). While acetylations are common in acylsugars of *S. habrochaites* and *S. lycopersicum*, they are absent from published analyses of *S. pennellii* acylsugars [3,5,24]. In *S. lycopersicum*, acetylation of triacylsucroses is performed by ASAT4 [37]. The corresponding enzyme in *S. pennellii*, if one exists, is therefore worth investigating as a candidate acylsugar acetyltransferase in this species. It should be noted that tetraacylsucroses were not observed in this study. One possible explanation for this absence is abundance below the detection limit, as acylsucroses exhibit poorer ionization efficiencies than acylglucoses under the positive-mode electrospray ionization employed here. As the tetraacylglucoses reported here are minor components of the acylsugar profile (both compounds have a maximum abundance < 2% of the most abundant acylsugar in the dataset), it is possible that hypothetical corresponding tetraacylsucroses would fall below the detection limit of the present analytical methods. It is also possible that tetraacylsucroses are rapidly degraded (*e.g.*, by ASH or ASFF1 enzymes), preventing their accumulation to detectable levels. Finally, tetraacylglucoses may not be hydrolysis products of tetraacylsucroses, but rather derived via direct acetylation of triacylglucoses. Further characterization of acyltransferases will be required to elucidate this step and test additional hypotheses regarding biosynthesis and turnover of tetraacylated sugars in *S. pennellii*.

While both eight- and 12-carbon acyl chains were previously reported from *S. pennellii* acylsugars, the differential accumulation of these acyl chains in accessions from the Pisco and Atico regions were not identified [3,5]. A variety of mechanisms for differential accumulation of

acylsugars containing these acyl chains are possible, including differences in acyl CoA production, incorporation of acyl chains into acylsugars, or acylsugar turnover. An enoyl CoA hydratase (ECH) and acyl CoA synthetase (ACS) involved in production of 10- and 12- carbon acyl CoA acylsugar precursors were recently identified [34]. Changes in substrate specificity of these enzymes in different *S. pennellii* accessions could contribute to varying composition of medium-length acyl CoA pools and subsequent incorporation into acylsugars. Alternatively, differences in acyl chain composition may be due to previously uncharacterized intraspecific variation in ASAT affinity for acyl CoAs, requiring further acyltransferase characterization. Finally, differential abundance of 8- and 12-carbon acyl chains in *S. pennellii* acylsugars may reflect differential affinity for acylsugars containing these acyl chains by enzymes involved in acylsugar turnover. For example, the ASH carboxylesterase enzymes facilitate acylsugar degradation in *S. lycopersicum* and *S. pennellii* by removing acyl chains from acylsucroses and acylglucoses [6]. ASHs primarily remove acyl chains from the R3 position of acylsugars and NMR spectra of acylsugars in *S. pennellii* consistently show medium-length chains at position R3 with positions R2 and R4 exclusively bearing four- or five-carbon acyl chains (Fig. 4.2) [6], suggesting that 8- and 12-carbon acyl chains could be targets for ASH-mediated acylsugar tailoring. While there are several hypotheses for mechanisms driving differential accumulation of 8- and 12-carbon acyl chains in *S. pennellii* acylsugars, we note that both straight and branched 8- and 12-carbon acyl chains have been observed in *S. pennellii* [3–5] and our NMR analysis did not include characterization of 8- or 12-carbon acyl chain-containing acylsugars. Therefore, further structural characterization of these acylsugars is warranted prior to exploration of their biosynthetic origins.

While the observation of acylglucoses representing a greater proportion of acylsugars in *S. pennellii* accessions from the South range of Peru than in accessions from the North range was previously reported [3], the difference in proportions of sugar cores in acylsugar extracts of accession LA2963 from other accessions from the Atico region was previously undescribed. We hypothesized that the proportion of acylsugars accumulating as acylglucoses in *S. pennellii* could be associated with activity of the ASFF1 enzyme, which hydrolyzes acylsucroses to acylglucoses in accession LA0716 [8]. We determined that the percentage of acylsugars accumulating as acylglucoses correlated with abundance of ASFF1 transcripts in two accessions from the Atico region, LA0716 and LA2963 (Fig. 4.6C). This provides an opportunity to explore acylsugar biosynthetic gene regulation. Potential approaches include characterization of the *ASFF1* promoter region from Atico region accessions or analysis of transcription factors that are enriched in *S. pennellii* accessions that accumulate high levels of acylsugars [16].

The intraspecific variations in *S. pennellii* acylsugar phenotype reported here not only confirm previously characterized aspects of the acylsugar pathway but also provide a starting point for further pathway analysis. The dimensions of acylsugar variation are potentially linked to all known components of acylsugar metabolism including enzymes in auxiliary pathways that generate acylsugar precursors (*i.e.*, IPMS3, ECH, ACS), activities of the core acylsugar biosynthetic pathway (*i.e.*, ASATs), and enzymes that degrade or remodel acylsugars (*i.e.*, ASHs and ASFF1). The mechanisms of variation in acylsugar phenotype may reflect differences in substrate specificity of pathway enzymes or disparity in enzyme activity levels due to differences in pathway regulation within the species or allelic variation. Improved understanding of acylsugar biosynthesis in *S. pennellii* will require further structural analysis of

acylsugars, biochemical characterization of enzymes in the pathway, and an understanding of the genetic regulatory network governing pathway expression.

APPENDIX

Table S4.1 NMR metadata.

Analysis description	
Supervisor	Dr. Daniel Holmes
Operator	Thilani Anthony
Institution	Michigan State University
Data and time of data acquisition	October 2019 - December 2019
Sample description	
Field frequency lock	Chloroform- <i>d</i> ₁
Additional solute	None
Solvent	CDCl ₃ (500 MHz NMR: 600 μL)
Chemical shift standard	CDCl ₃
Concentration standard	None
Instrument description	
Agilent DirectDrive2 500 MHz NMR	
Geographic location of the instrument	42.7288, -84.4745
Magnet	499.70 MHz
Probe	OneNMR Probe with Protune accessory for hands-off tuning
Autosampler	7600AS 96 sample autosamplers
Acquisition software	VnmrJ 3.2A

Table S4.1 (cont'd)

Acquisition parameters	
Agilent DirectDrive2 500 MHz NMR	
a) Acquisition parameters file reference	¹ H: VnmrJ/ Experiment Selector/ Common/ PROTON ¹³ C: VnmrJ/ Experiment Selector/ Common/ CARBON HSQC: VnmrJ/ Experiment Selector/ Common/ (HC)HSQCAD HMBC: VnmrJ/ Experiment Selector/ Common/ (HC)gHMBCAD COSY: VnmrJ/ Experiment Selector/ Common/ (HH)gCOSY J-resolved: VnmrJ/ Experiment Selector/ Liquid/ JSpectra/ HOMO2DJ
b) Sample details	Tube: Kontes NMR tube, 8 in Temperature: 25 °C
c) Instrument operation details	Radiation frequency: ¹ H: 499.90 ¹³ C: 125.71 HSQC: 499.90, 125.71 HMBC: 499.90, 125.71 COSY: 499.90, 499.90 J-resolved: 499.90 Acquisition nucleus: ¹ H: 90° = 7.9 μs, ¹³ C: 90° = 10.20 μs
d) Number of data points acquired	¹ H: 16384 ¹³ C: 32768 HSQC: 1202, 128 HMBC: 1202,200 COSY: 674, 200 J-resolved: 2810, 64
e) Data acquisition details	¹ H: number of scans: 32 ¹³ C: number of scans: 256 HSQC: t1 increments: 400; scan per t1 increment: 4 HMBC: t1 increments: 512; scan per t1 increment: 4 COSY: t1 increments: 512; scan per t1 increment: 4-16 J-resolved: t1 increments: 128; scan per t1 increment: 16

Table S4.1 (cont'd)

Spectral processing parameters	
Agilent DirectDrive2 500 MHz NMR	
a) Software	VnmrJ 3.2 A
b) Process weighting	¹ H: LineBroaden ¹³ C: LineBroaden HSQC: gaussian (F2); gaussian (F1) HMBC: sqsinebell (F2); gaussian (F1) COSY: sqsinebell (F2); sqsinebell (F1) J-resolved: sinebell (F2); sinebell (F1)

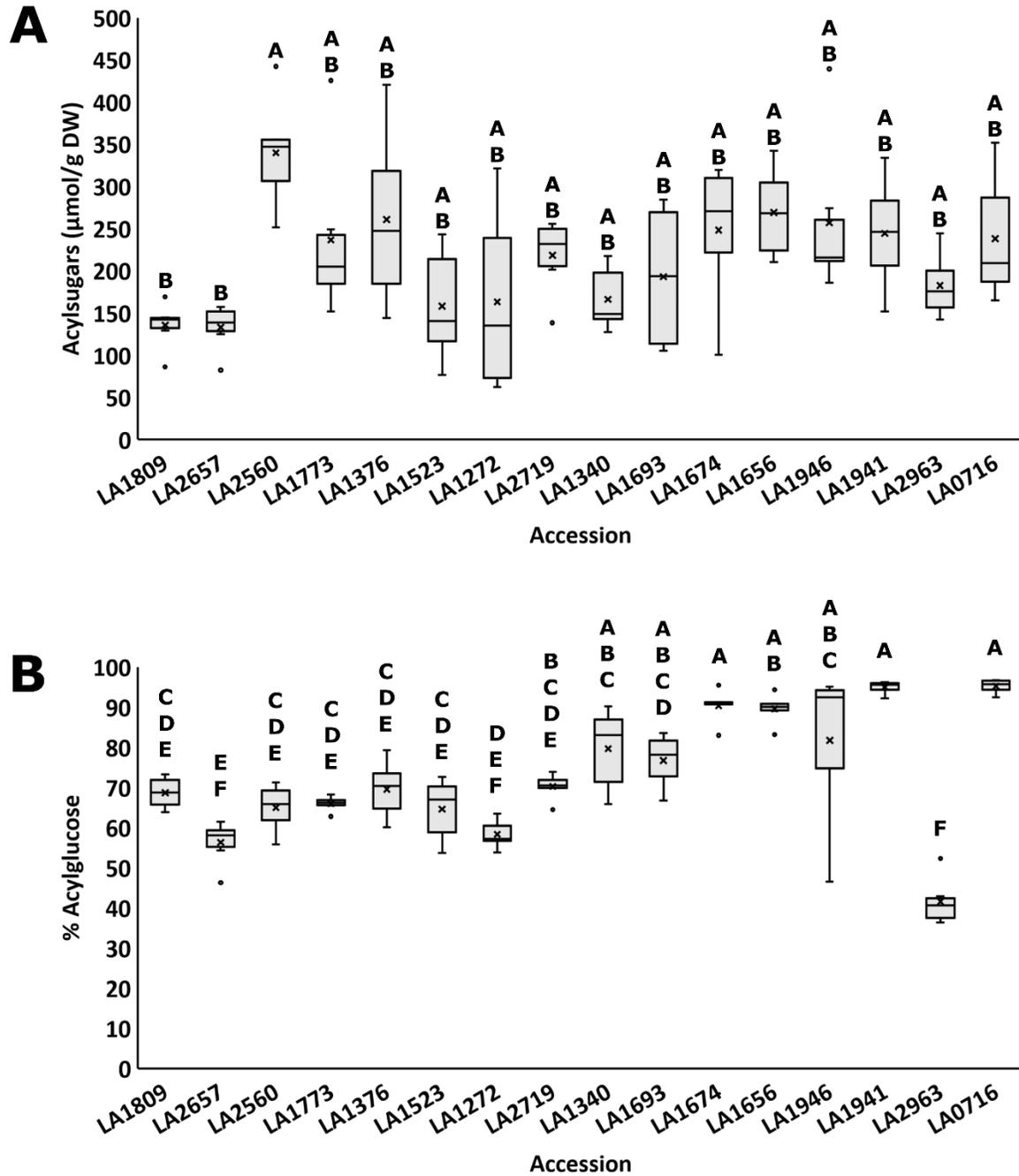


Figure S4.1 Quantification of acylsugars in 16 accessions of *S. pennellii*. Accessions are arranged left to right by latitude from north to south. (A) Total acylsugars. (B) Percent acylglucose accumulation. Results of ANOVA and Tukey’s mean-separation test are indicated by letters; accessions that do not share at least one letter are significantly different from one another ($p < 0.001$, $n = 6$ for all accessions).

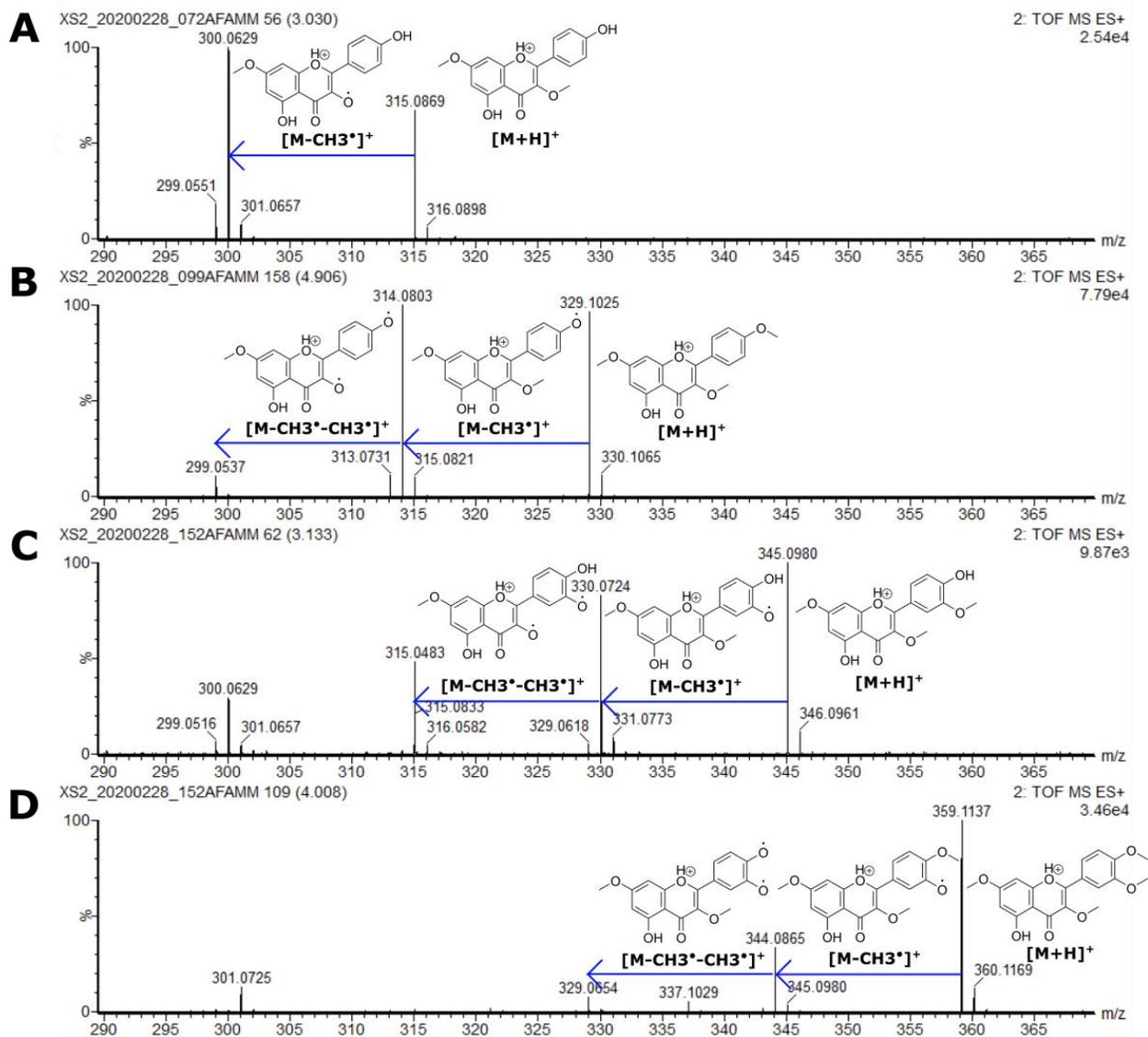


Figure S4.2 CID spectra of flavonoids extracted from *S. pennellii* analyzed by ES+ UPLC-HR-MS. (A) Flavonoid A; (B) flavonoid B; (C) flavonoid C; (D) flavonoid D. See Table 4.3 for additional details.

Table S4.2 NMR chemical shifts for S3:12(4,4,4) Purified from *S. pennellii* LA0716.

Carbon # (group)	¹ H (ppm)	¹³ C (ppm) (from HSQC and HMBC)
<p>S3:12(4,4,4) Purified from <i>S. pennellii</i> LA0716 Chemical Formula: C₂₄H₄₀O₁₄ HRMS: (ESI) <i>m/z</i> calculated for C₂₄H₄₀O₁₄ ([M+NH₄]⁺): 570.2756 Experimental <i>m/z</i>: 570.2778 NMR (500 MHz, CDCl₃) Sample mass: 2 mg</p>		
1 (CH)	5.76 (d, <i>J</i> = 4.0 Hz, 1H)	88.78
2 (CH)	4.86 (dd, <i>J</i> = 10.3, 4.0 Hz, 1H)	70.74
- 1 (CO)	-	176.72
- 2 (CH)	2.46 (hept, <i>J</i> = 7.1 Hz, 1H)	33.88
- 3,4 (CH ₃)	1.09 (d, <i>J</i> = 7.0 Hz, 6H)	18.87
3 (CH)	5.55 (t, <i>J</i> = 9.9 Hz, 1H)	69.10
- 1 (CO)	-	176.56
- 2 (CH)	2.53 (hept, <i>J</i> = 7.0 Hz, 1H)	33.85
- 3,4 (CH ₃)	1.13 (d, <i>J</i> = 7.0 Hz, 6H)	18.86
4 (CH)	4.93 (t, <i>J</i> = 10.0 Hz, 1H)	68.44
- 1 (CO)	-	176.16
- 2 (CH)	2.53 (hept, <i>J</i> = 7.0 Hz, 1H)	33.85
- 3,4 (CH ₃)	1.13 (d, <i>J</i> = 7.0 Hz, 6H)	18.86
5 (CH)	4.23 (m, 1H)	71.86
6 (CH ₂)	3.60 (m, 2H)	61.50
1' (CH ₂)	3.61 (m, 1H), 3.51 (d, <i>J</i> = 11.9 Hz, 1H)	64.41
2' (C)	-	104.44
3' (CH)	4.27 (m, 1H)	77.89
4' (CH)	4.27 (m, 1H)	73.10
5' (CH)	3.76 (m, 1H)	81.34
6' (CH ₂)	3.88 (d, <i>J</i> = 13.0 Hz, 1H), 3.75 (m, 1H)	60.13

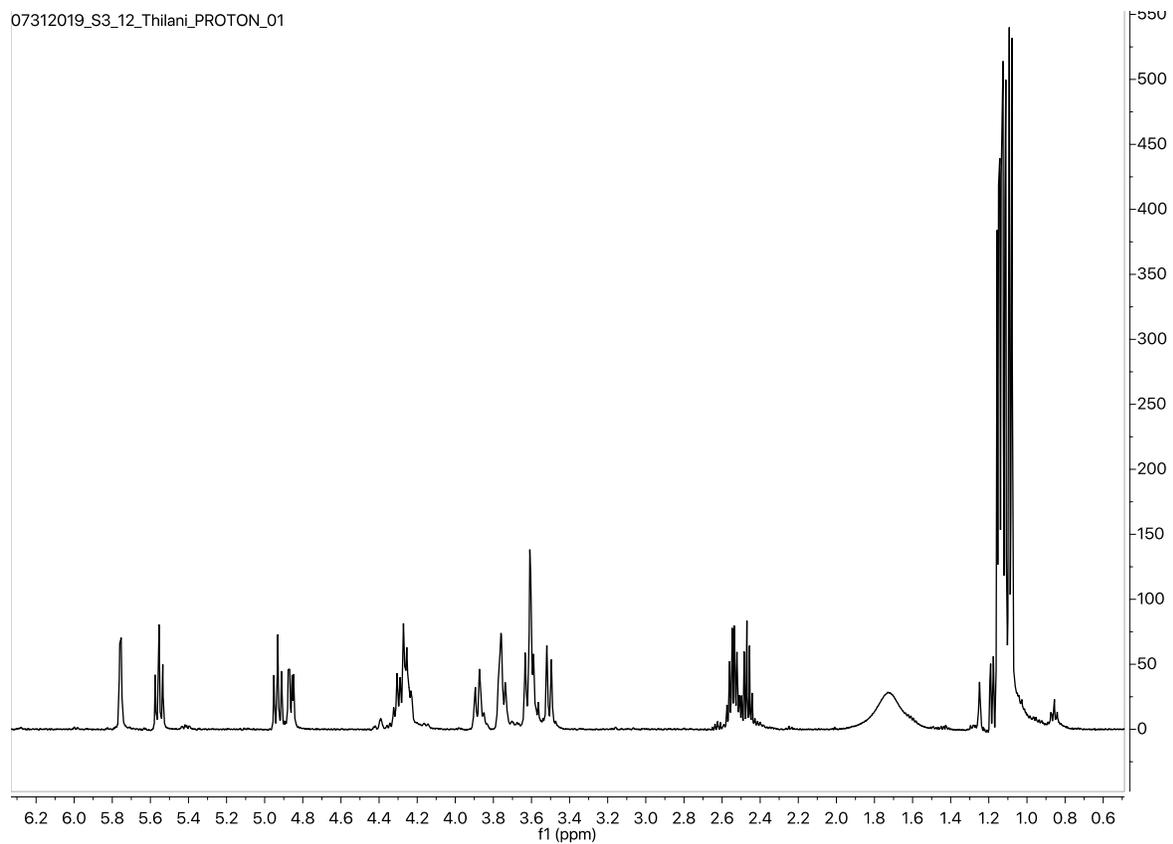


Figure S4.3 ^1H NMR spectrum for S3:12(4,4,4) purified from *S. pennellii* LA0716.

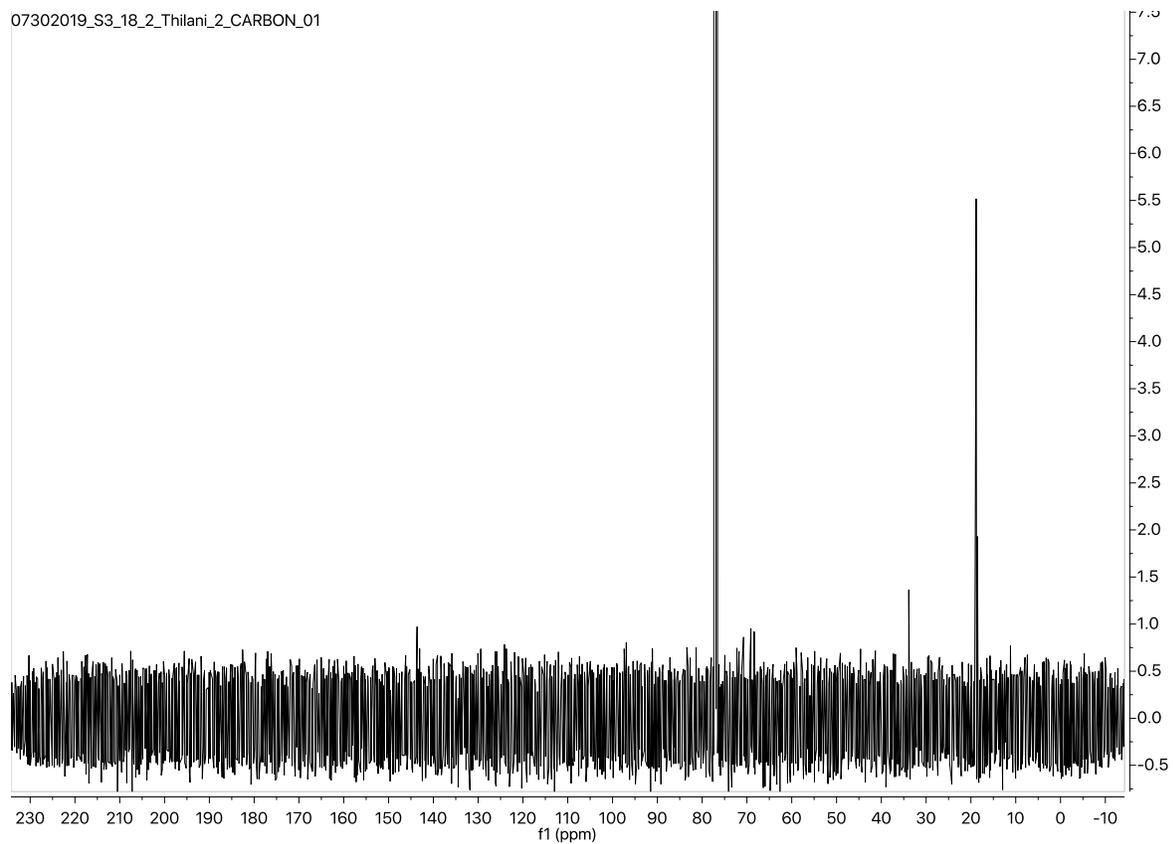


Figure S4.4 ^{13}C NMR spectrum for S3:12(4,4,4) purified from *S. pennellii* LA0716.

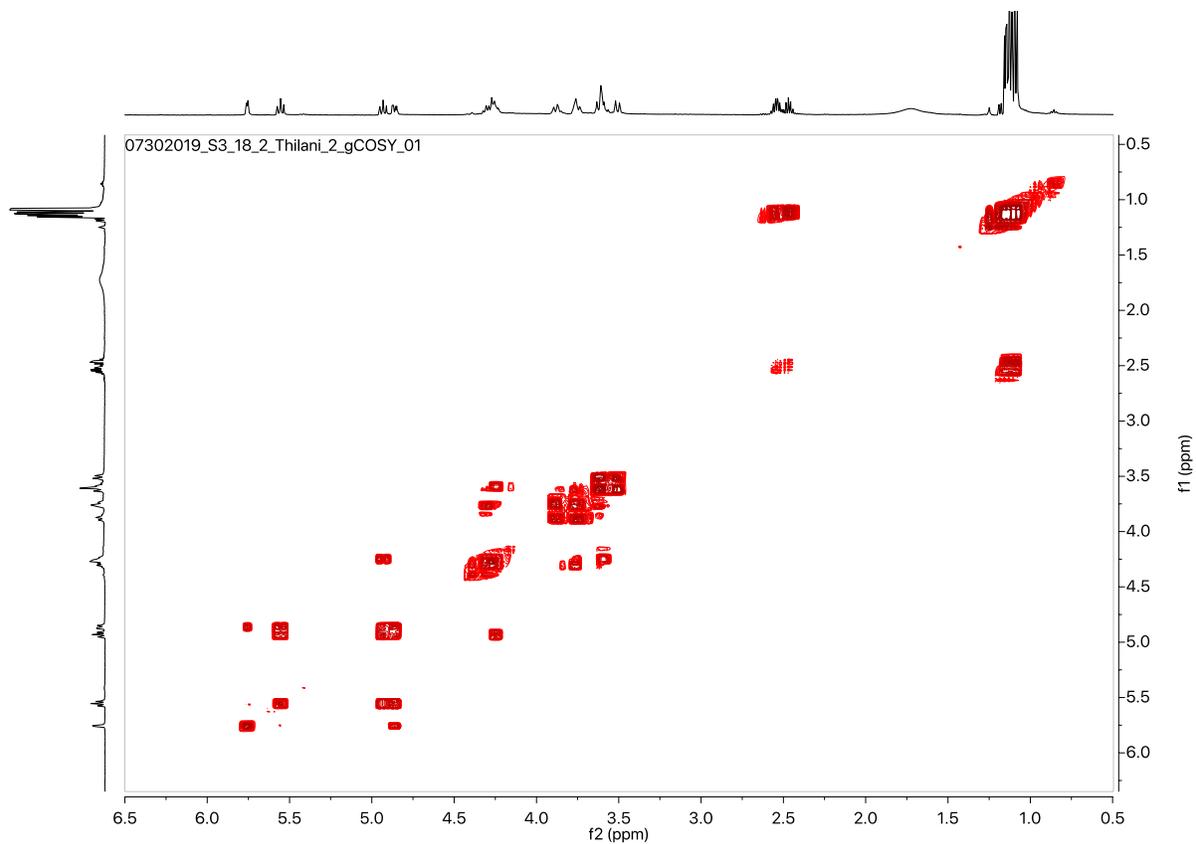


Figure S4.5 gCOSY NMR spectrum for S3:12(4,4,4) purified from *S. pennellii* LA0716.

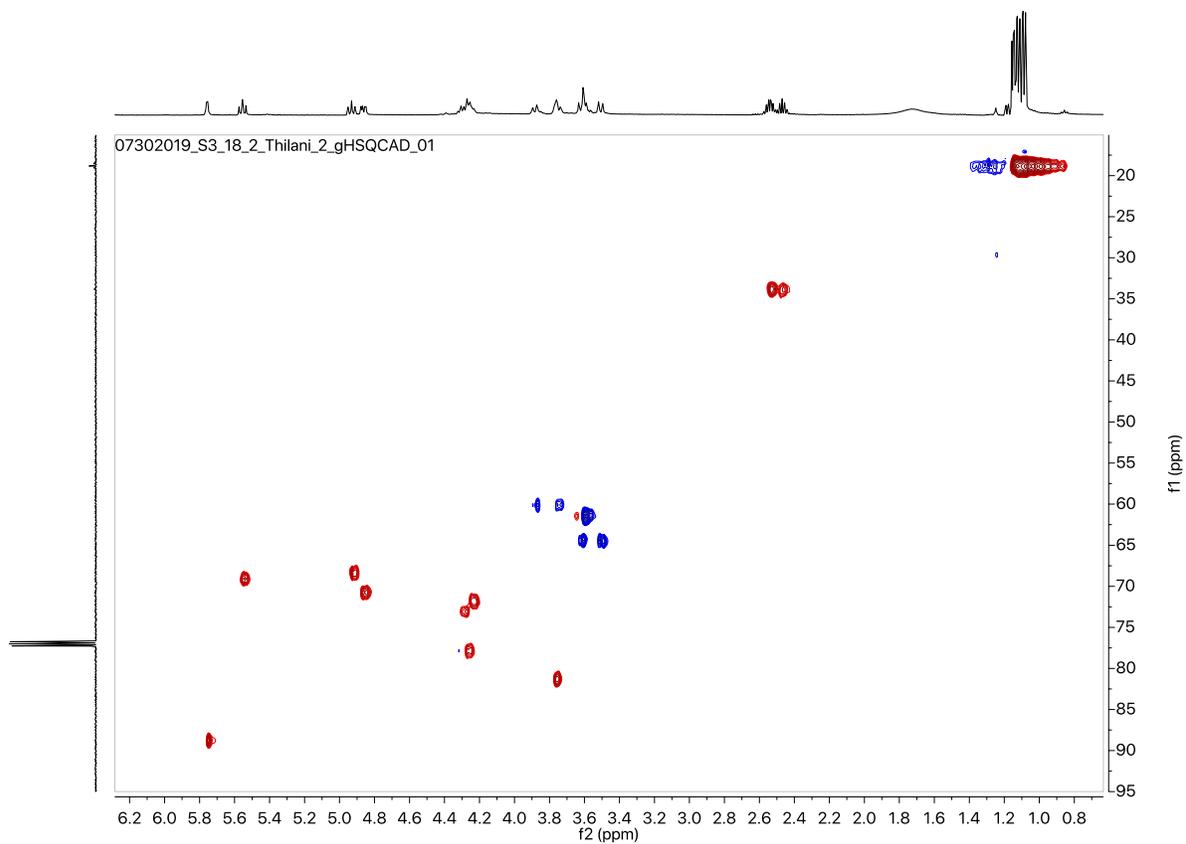


Figure S4.6 gHSQCAD NMR spectrum for S3:12(4,4,4) purified from *S. pennellii* LA0716.

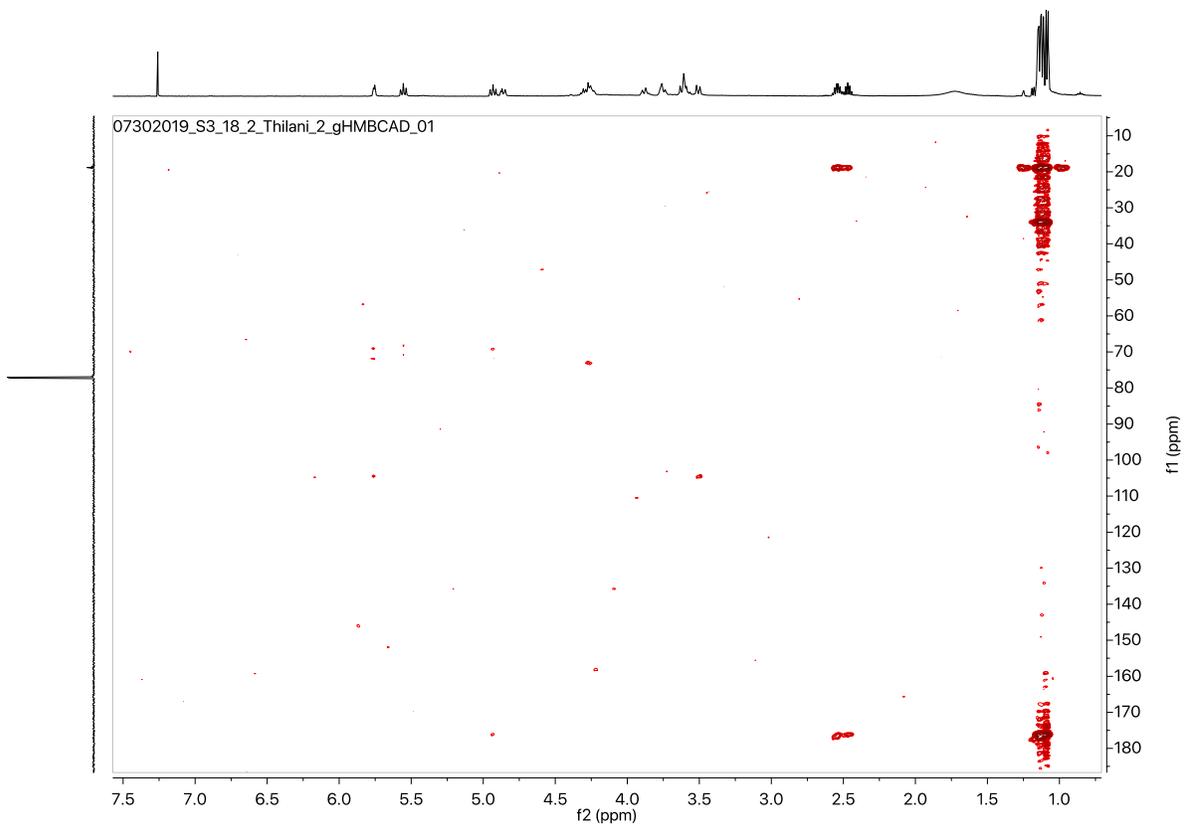


Figure S4.7 gHMBCAD NMR spectrum for S3:12(4,4,4) purified from *S. pennellii* LA0716.

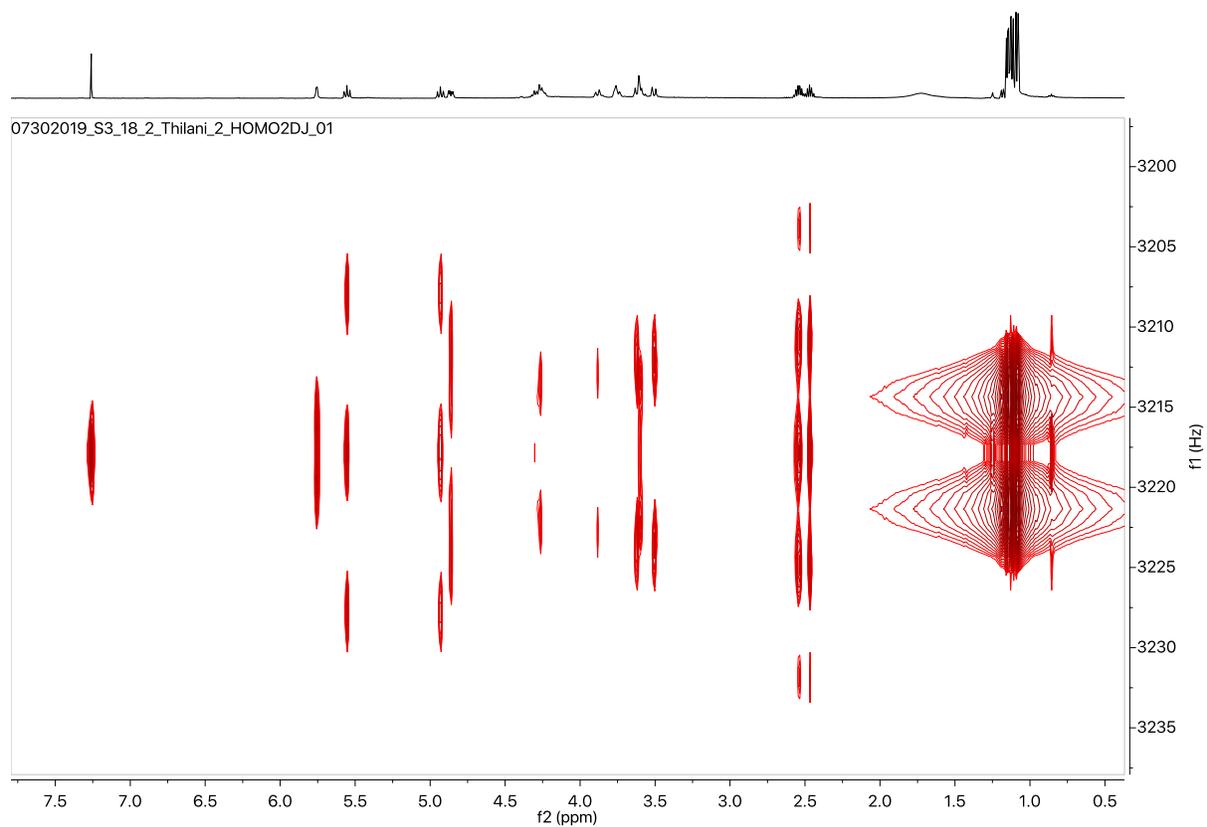
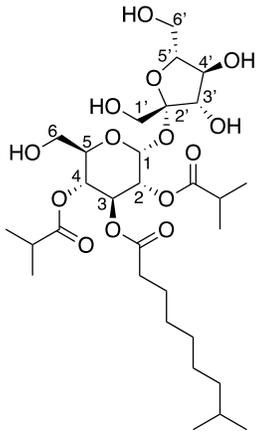


Figure S4.8 ^1H - ^1H HOMO2DJ NMR spectrum for S3:12(4,4,4) purified from *S. pennellii* LA0716.

Table S4.3 NMR chemical shifts for S3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

	<p style="text-align: center;">S3:18(4,4,10)-1</p> <p style="text-align: center;">Purified from <i>S. pennellii</i> LA0716</p> <p style="text-align: center;">Chemical Formula: C₃₀H₅₂O₁₄</p> <p style="text-align: center;">HRMS: (ESI) <i>m/z</i> calculated for C₃₀H₅₂O₁₄ ([M+NH₄]⁺): 654.3695</p> <p style="text-align: center;">Experimental <i>m/z</i>: 654.3699</p> <p style="text-align: center;">NMR (500 MHz, CDCl₃)</p> <p style="text-align: center;">Sample mass: 2 mg</p>	
<p style="text-align: center;">Carbon # (group)</p>	<p style="text-align: center;">¹H (ppm)</p>	<p style="text-align: center;">¹³C (ppm) (from HSQC and HMBC)</p>
<p>1 (CH)</p>	<p>5.77 (d, <i>J</i> = 4.0 Hz, 1H)</p>	<p>88.79</p>
<p>2 (CH)</p> <ul style="list-style-type: none"> - 1 (CO) - 2 (CH) - 3,4 (CH₃) 	<p>4.84 (dd, <i>J</i> = 10.3, 4.0 Hz, 1H)</p> <p>2.55 (hept, <i>J</i> = 7.0 Hz, 1H)</p> <p>1.13 (d, <i>J</i> = 7.0 Hz, 6H)</p>	<p>70.80</p> <p>176.59</p> <p>33.84</p> <p>19.41</p>
<p>3 (CH)</p> <ul style="list-style-type: none"> - 1 (CO) - 2 (CH₂) - 3 (CH₂) - 4,5,6 (CH₂) - 7 (CH₂) - 8 (CH) - 9,10 (CH₃) 	<p>5.56 (dd, <i>J</i> = 10.3 Hz, 1H)</p> <p>2.21 (t, <i>J</i> = 7.8 Hz, 2H)</p> <p>1.52(m, 2H)</p> <p>1.25 (m)</p> <p>1.13 (m)</p> <p>1.49 (m)</p> <p>0.85 (m)</p>	<p>69.06</p> <p>172.84</p> <p>34.28</p> <p>24.83</p> <p>29.39</p> <p>38.93</p> <p>27.98</p> <p>22.65</p>
<p>4 (CH)</p> <ul style="list-style-type: none"> - 1 (CO) - 2 (CH) - 3,4 (CH₃) 	<p>4.91 (t, <i>J</i> = 10.4 Hz, 1H)</p> <p>2.55 (hept, <i>J</i> = 7.0 Hz, 1H)</p> <p>1.13 (d, <i>J</i> = 7.0 Hz, 6H)</p>	<p>68.40</p> <p>176.16</p> <p>33.84</p> <p>19.41</p>
<p>5 (CH)</p>	<p>4.21 (m, 1H)</p>	<p>71.89</p>
<p>6 (CH₂)</p>	<p>3.61 (m, 2H)</p>	<p>61.51</p>
<p>1' (CH₂)</p>	<p>3.60 (m, 1H), 3.52 (d, <i>J</i> = 12.0 Hz, 1H)</p>	<p>64.60</p>
<p>2' (C)</p>	<p>-</p>	<p>104.52</p>
<p>3' (CH)</p>	<p>4.25 (m, 1H)</p>	<p>78.18</p>
<p>4' (CH)</p>	<p>4.31 (t, <i>J</i> = 8.4 Hz, 2H)</p>	<p>72.89</p>
<p>5' (CH)</p>	<p>3.74 (m, 1H)</p>	<p>81.39</p>
<p>6' (CH₂)</p>	<p>3.87 (d, <i>J</i> = 13.0 Hz, 1H), 3.74 (m, 1H)</p> <p>-</p>	<p>60.02</p>

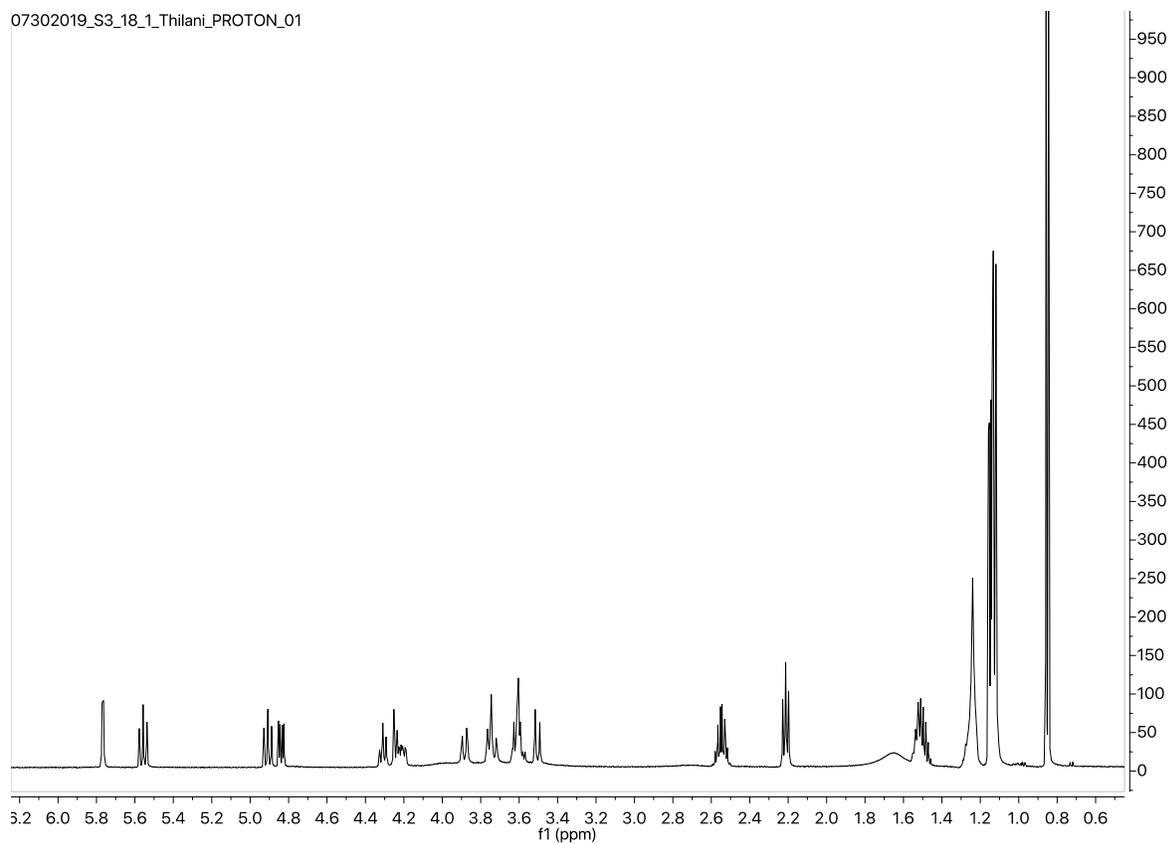


Figure S4.9 ^1H NMR spectrum for S3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

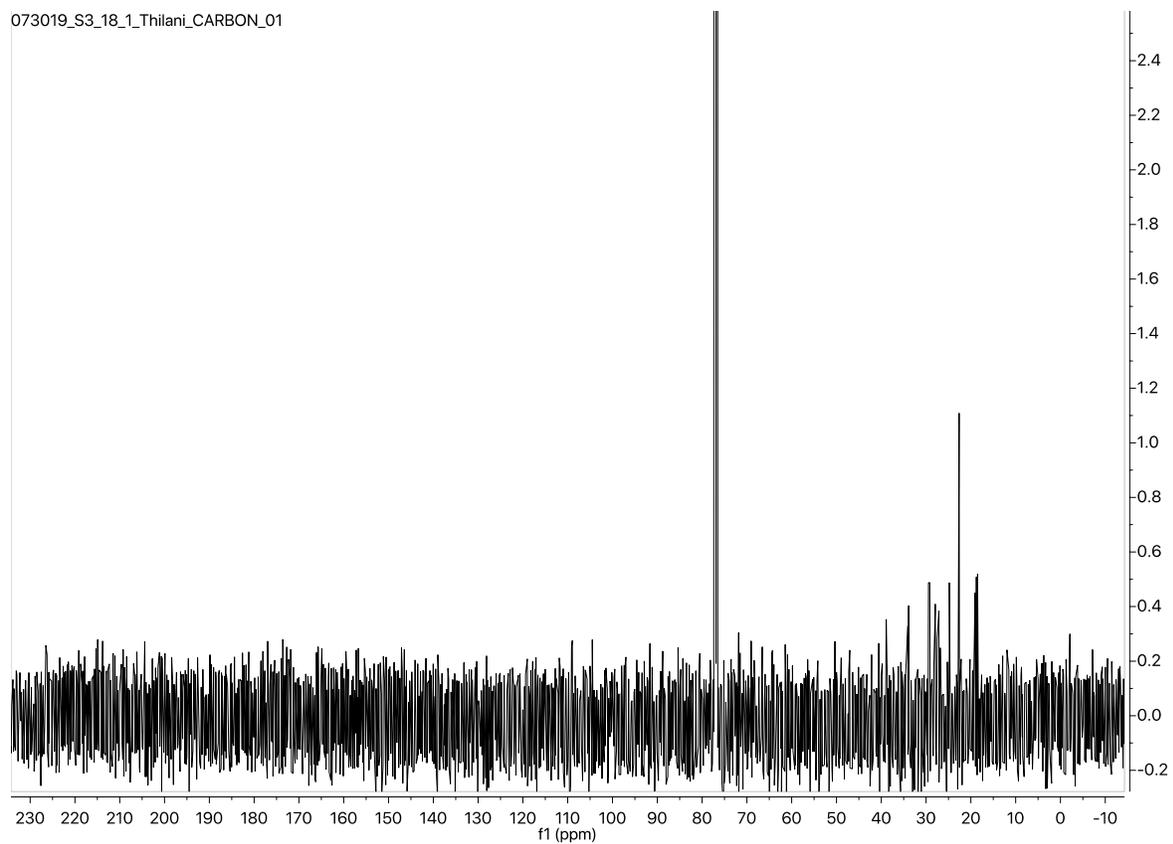


Figure S4.10 ^{13}C NMR spectrum for S3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

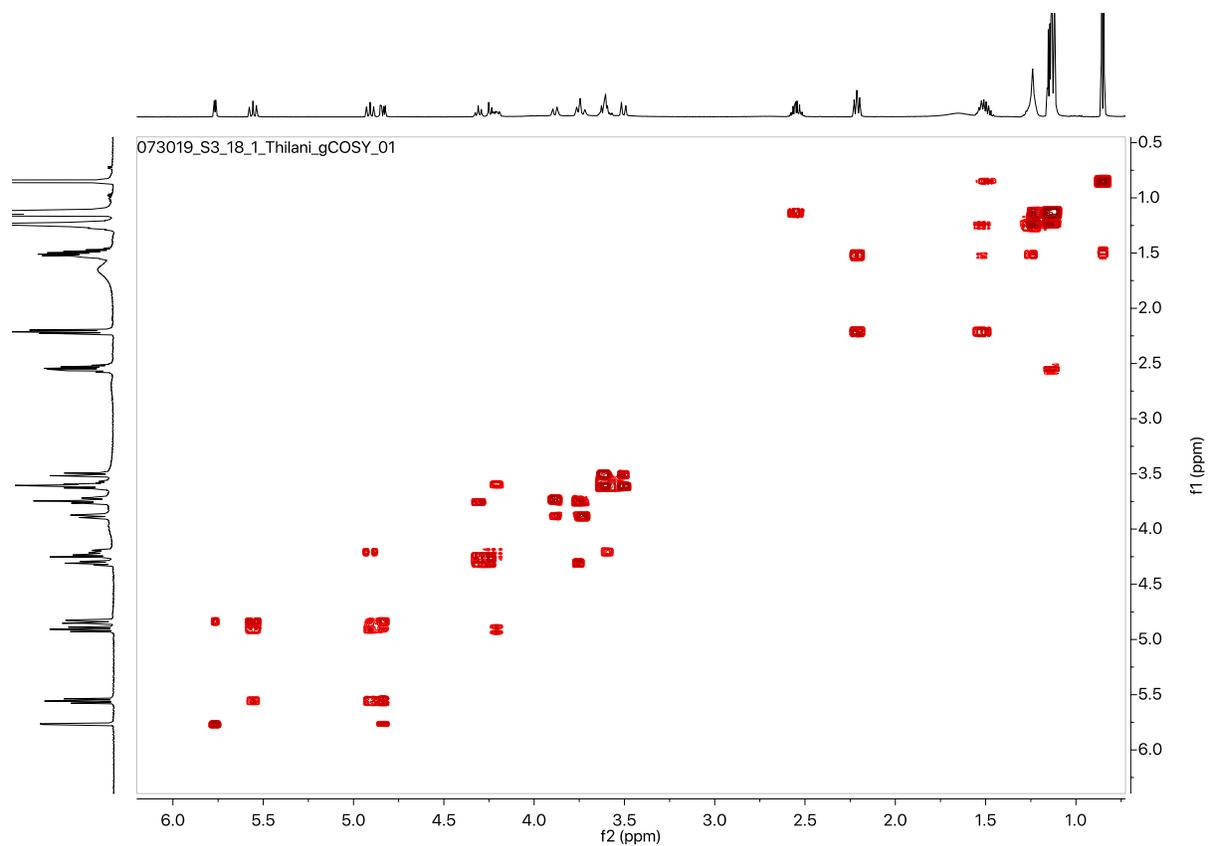


Figure S4.11 gCOSY NMR spectrum for S3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

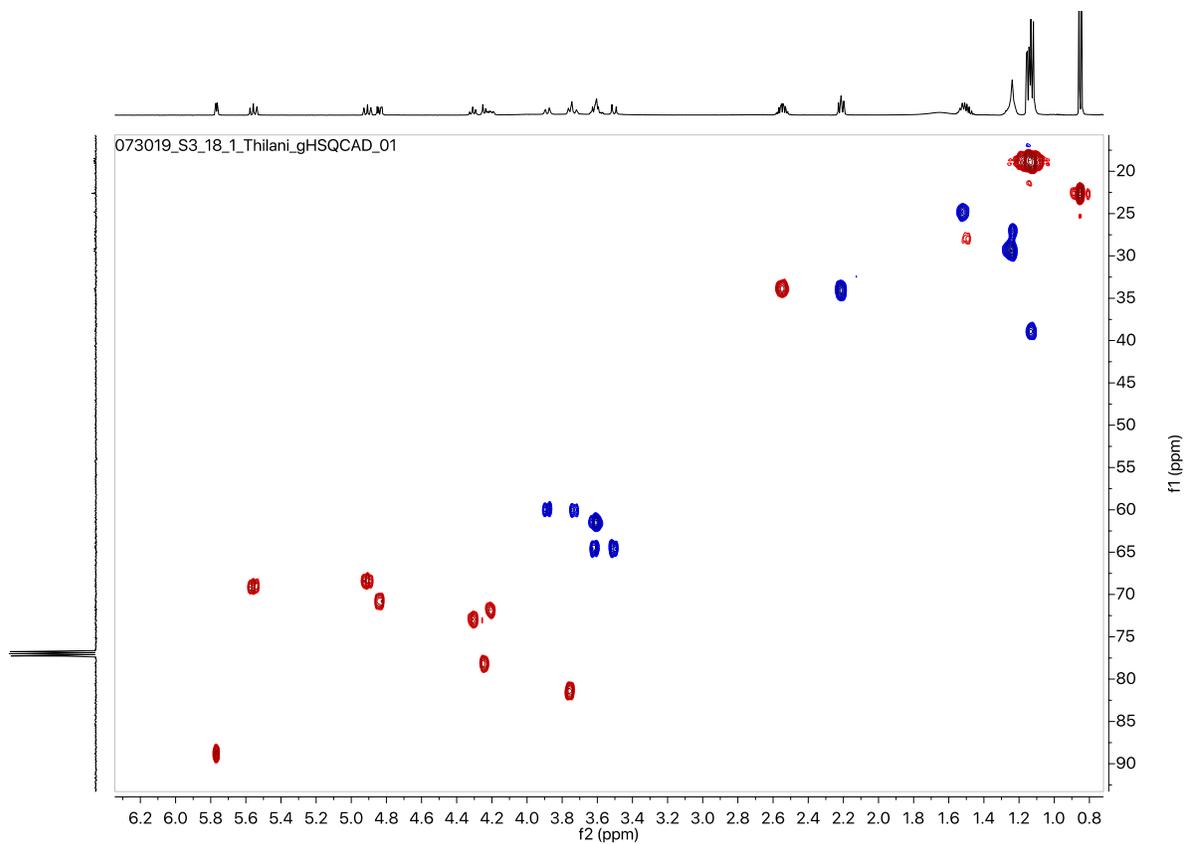


Figure S3.12 gHSQCAD NMR spectrum for S3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

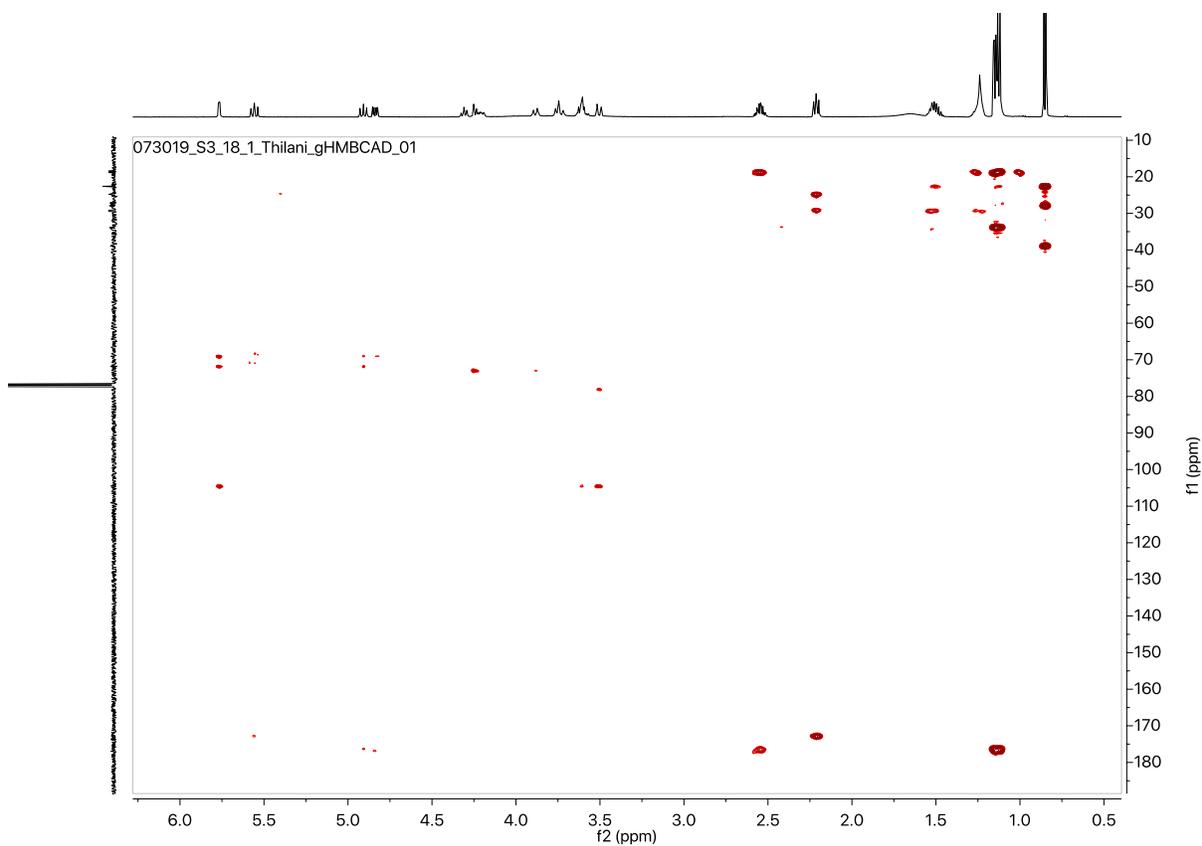


Figure S4.13 gHMBCAD NMR spectrum for S3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

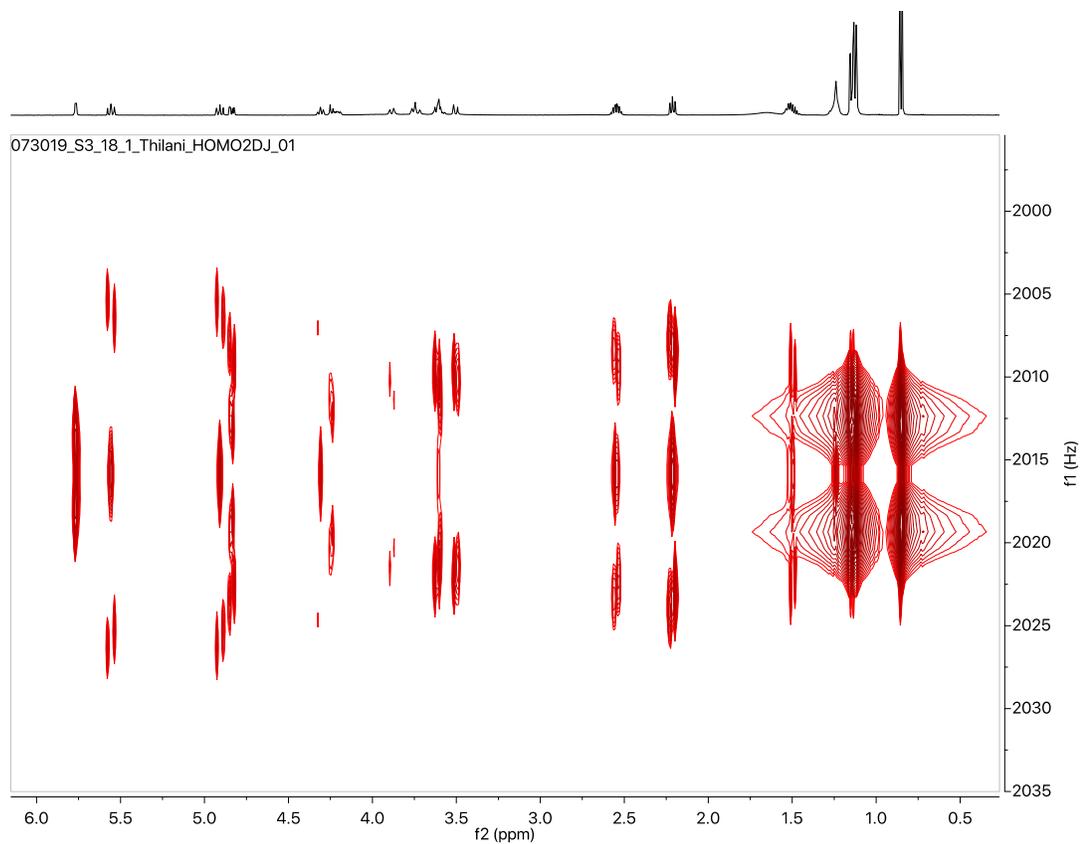
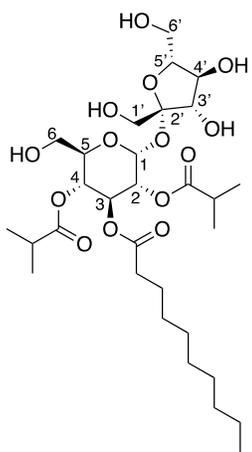


Figure S4.14 ^1H - ^1H HOMO2DJ NMR spectrum for S3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

Table S4.4 NMR chemical shifts for S3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

Carbon # (group)	¹ H (ppm)	¹³ C (ppm) (from HSQC and HMBC)
 <p>S3:18(4,4,10)-2 Purified from <i>S. pennellii</i> LA0716 Chemical Formula: C₃₀H₅₂O₁₄ HRMS: (ESI) <i>m/z</i> calculated for C₃₀H₅₂O₁₄ ([M+NH₄]⁺): 654.3695 Experimental <i>m/z</i>: 654.3699 NMR (500 MHz, CDCl₃) Sample mass: 2 mg</p>		
1 (CH)	5.77 (d, <i>J</i> = 4.0 Hz, 1H)	88.81
2 (CH)	4.83 (dd, <i>J</i> = 10.3, 4.0 Hz, 1H)	70.81
- 1 (CO)	-	176.89
- 2 (CH)	2.55 (hept, <i>J</i> = 7.0 Hz, 1H)	33.85
- 3,4 (CH ₃)	1.13 (d, <i>J</i> = 7.0 Hz, 6H)	18.91
3 (CH)	5.56 (dd, <i>J</i> = 10.0 Hz, 1H)	69.02
- 1 (CO)	-	172.70
- 2 (CH ₂)	2.21 (t, <i>J</i> = 7.6 Hz, 2H)	34.11
- 3 (CH ₂)	1.51 (pent, <i>J</i> = 7.1 Hz, 2H)	24.79
- 4,5,6,7 (CH ₂)	1.24 (m)	29.31
- 8 (CH ₂)	1.24 (m)	38.93
- 9 (CH ₂)	1.27 (m)	22.56
- 10 (CH ₃)	0.87 (t, <i>J</i> = 6.9 Hz, 3H)	14.06
4 (CH)	4.90 (t, <i>J</i> = 10.4 Hz, 1H)	68.44
- 1 (CO)	-	176.16
- 2 (CH)	2.55 (hept, <i>J</i> = 7.0 Hz, 1H)	33.85
- 3,4 (CH ₃)	1.13 (d, <i>J</i> = 7.0 Hz, 6H)	18.91
5 (CH)	4.20 (m, 1H)	71.85
6 (CH ₂)	3.61 (m, 2H)	61.53
1' (CH ₂)	3.60 (m, 1H), 3.51 (d, <i>J</i> = 12.0 Hz, 1H)	64.47
2' (C)	-	104.45
3' (CH)	4.25 (m, 1H)	78.27
4' (CH)	4.30 (t, <i>J</i> = 8.4 Hz, 2H)	72.99
5' (CH)	3.75 (m, 1H)	81.40
6' (CH ₂)	3.89 (d, <i>J</i> = 13.0 Hz, 1H), 3.74 (m, 1H)	60.02

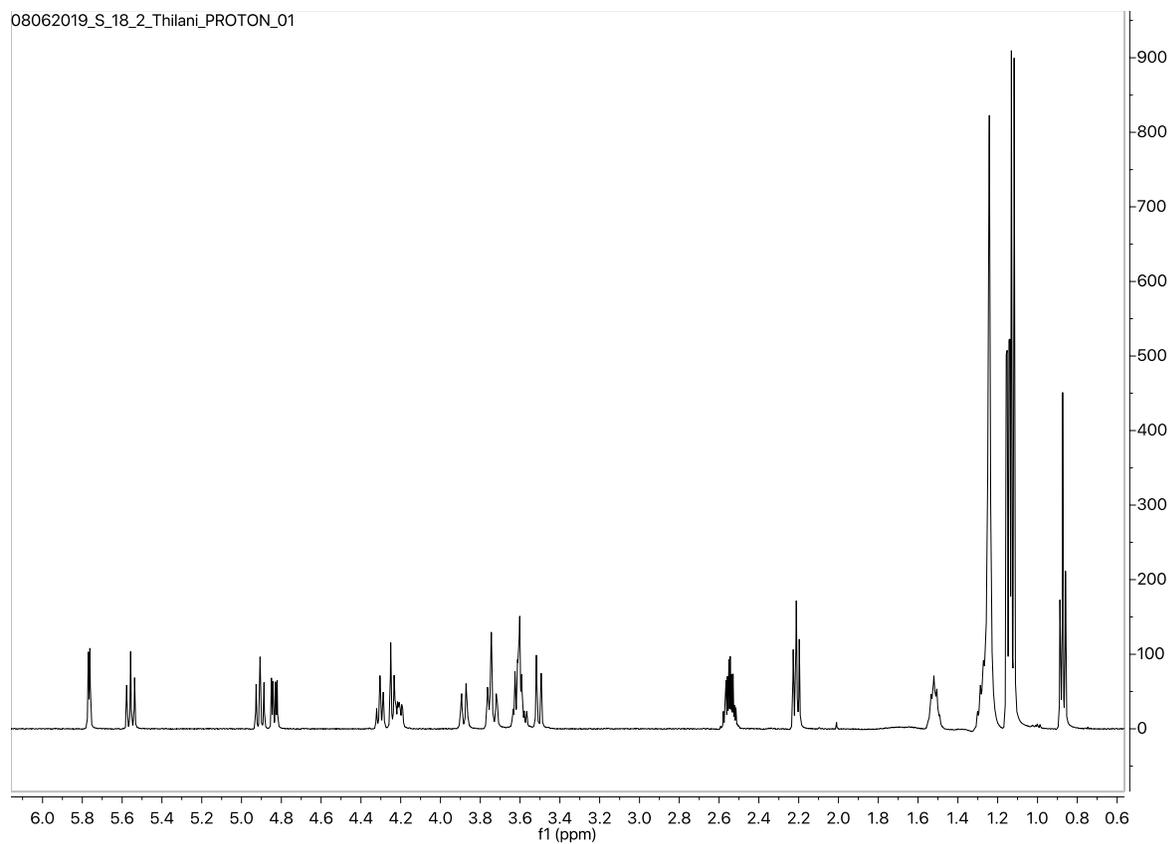


Figure S4.15 ^1H NMR spectrum for S3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

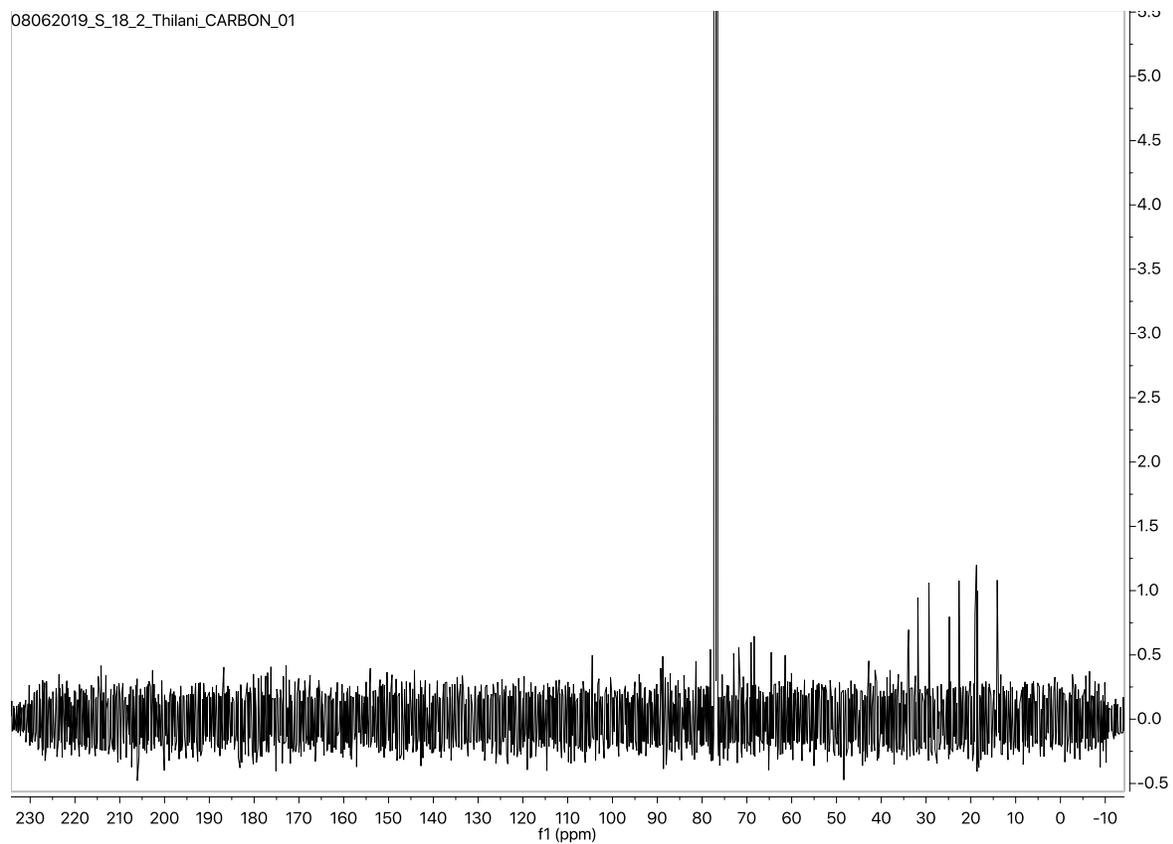


Figure S4.16 ^{13}C NMR spectrum for S3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

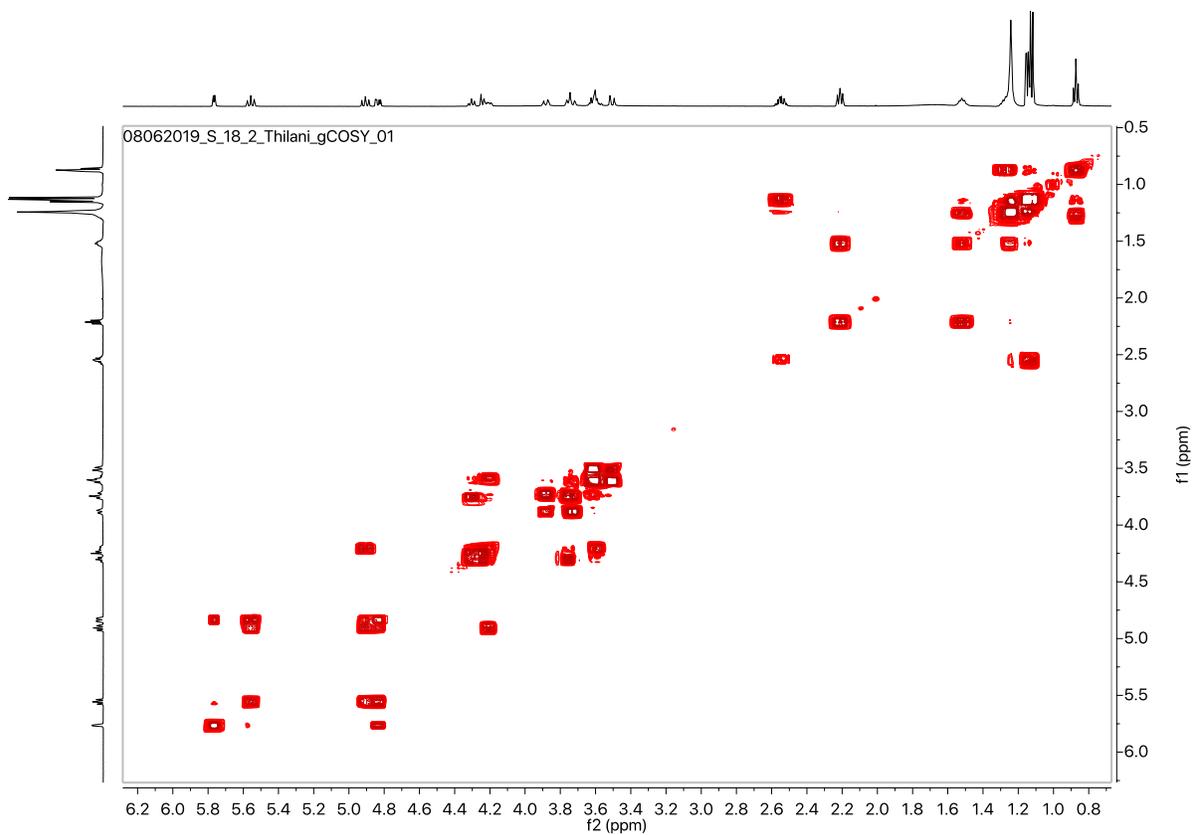


Figure S4.17 gCOSY NMR spectrum for S3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

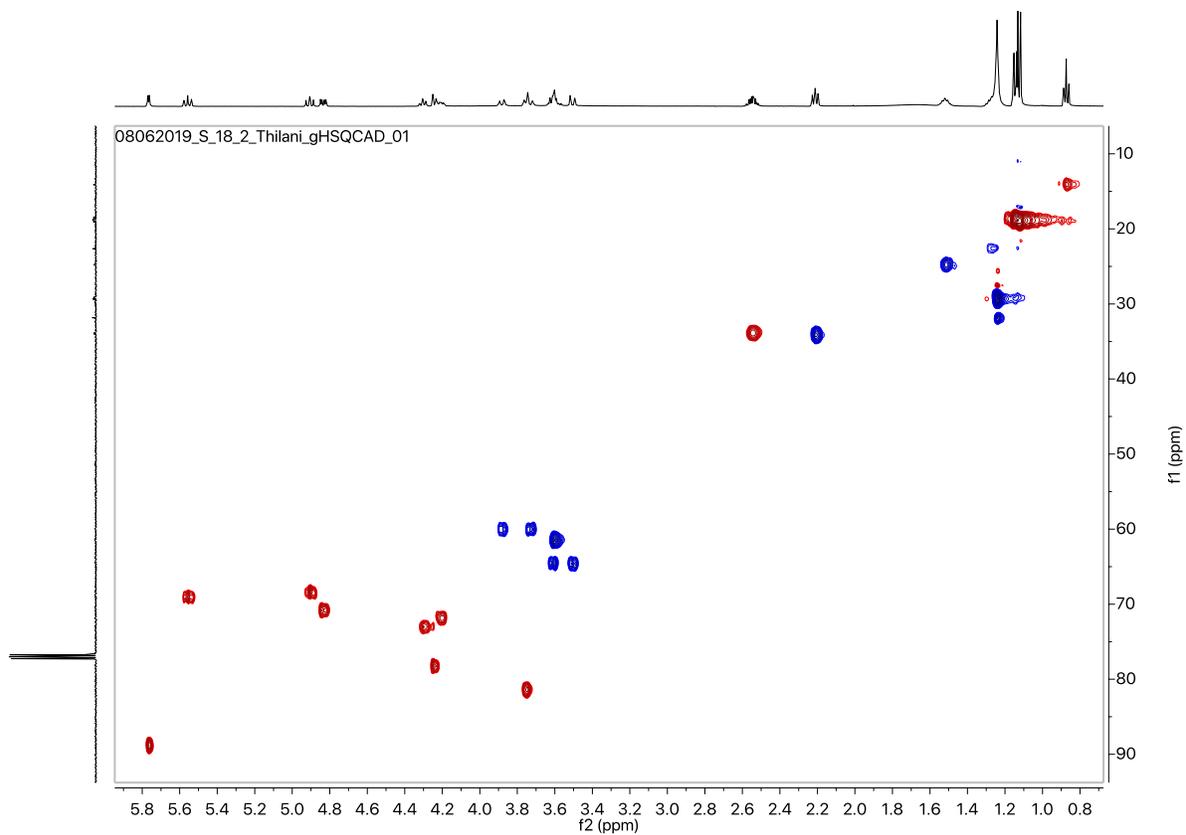


Figure S4.18 gHSQCAD NMR spectrum for S3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

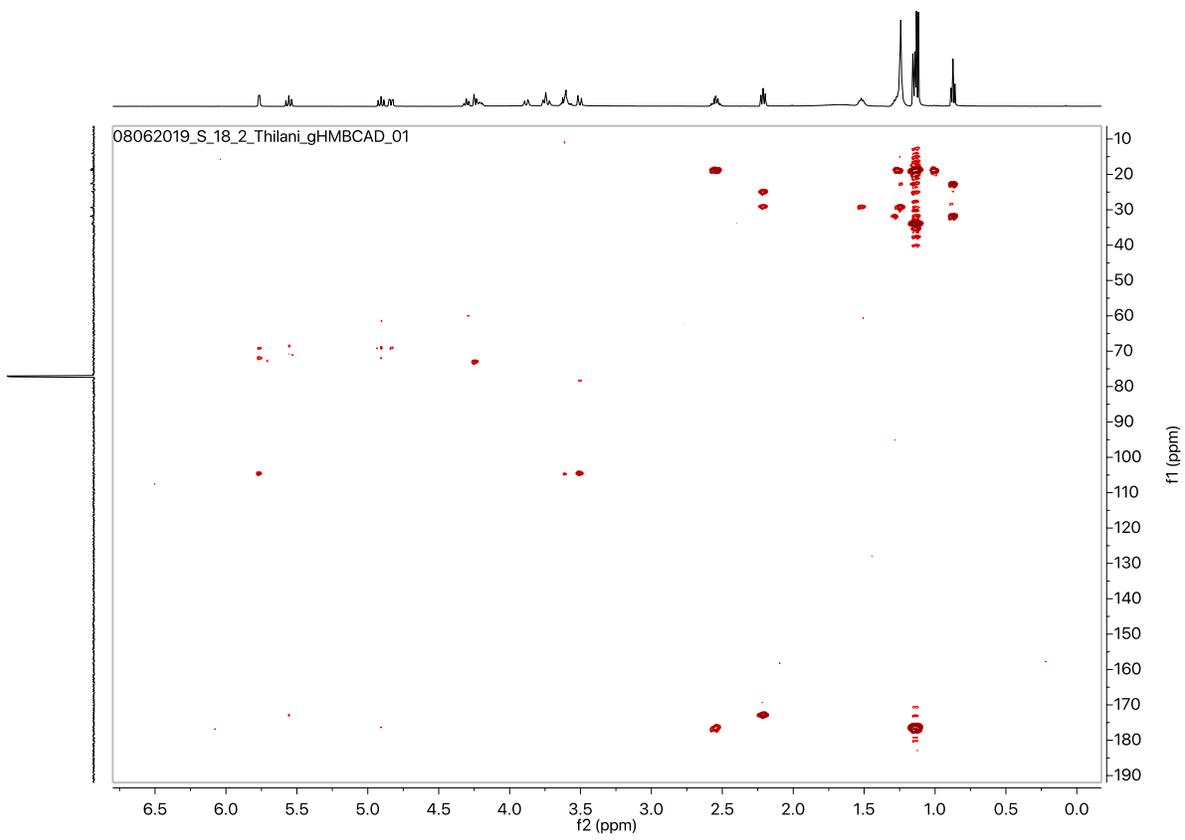


Figure S4.19 gHMBCAD NMR spectrum for S3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

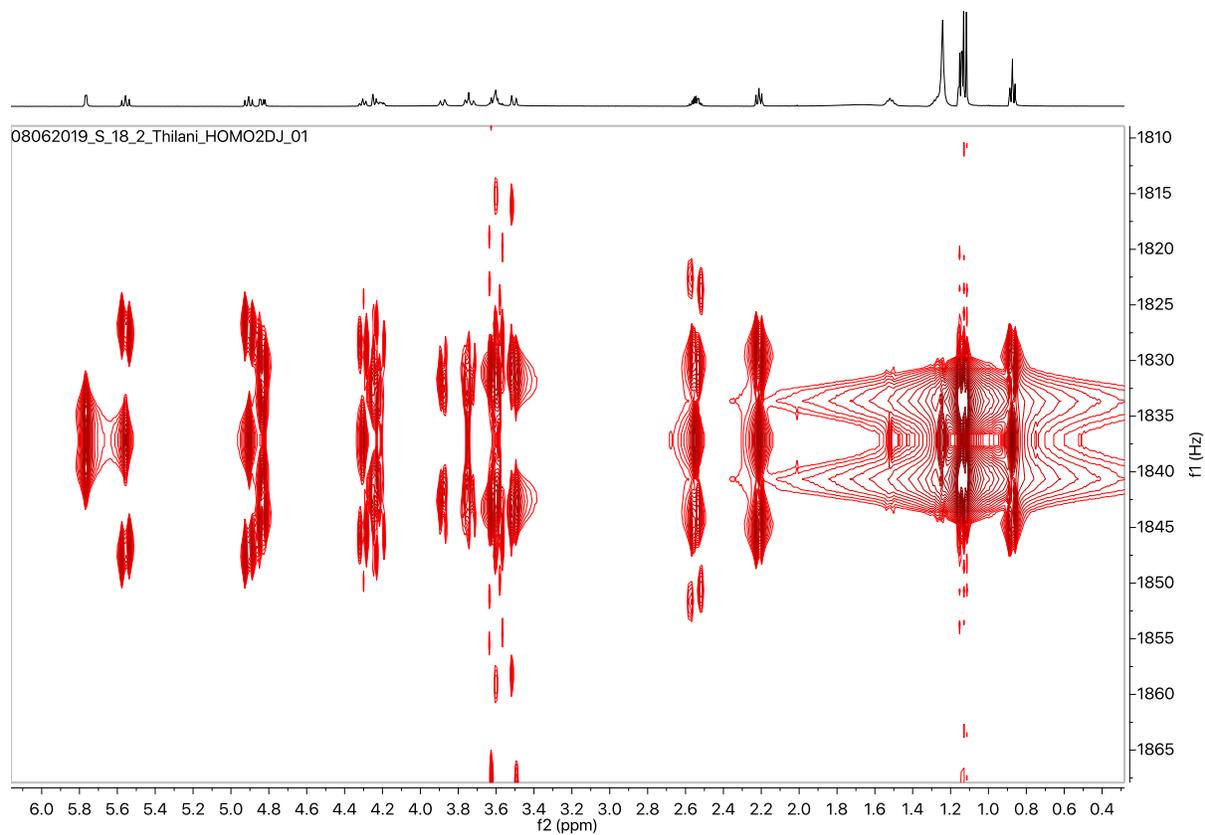
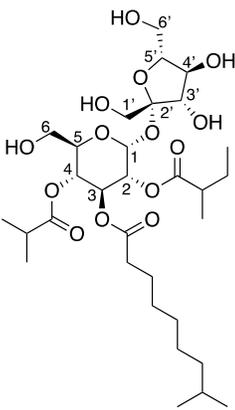


Figure S4.20 ^1H - ^1H HOMO 2DJ NMR spectrum for S3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

Table S4.5 NMR chemical shifts for S3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

	<p align="center">S3:19(4,5,10)-1</p> <p align="center">Purified from <i>S. pennellii</i> LA0716</p> <p align="center">Chemical Formula: C₃₁H₅₄O₁₄</p> <p align="center">HRMS: (ESI) <i>m/z</i> calculated for C₃₁H₅₄O₁₄ ([M+NH₄]⁺): 668.3852</p> <p align="center">Experimental <i>m/z</i>: 668.3856</p> <p align="center">NMR (500 MHz, CDCl₃)</p> <p align="center">Sample mass: 2 mg</p>	
<p align="center">Carbon # (group)</p>	<p align="center">¹H (ppm)</p>	<p align="center">¹³C (ppm) (from HSQC and HMBC)</p>
<p>1 (CH)</p>	<p align="center">5.78 (d, <i>J</i> = 4.0 Hz, 1H)</p>	<p align="center">88.80</p>
<p>2 (CH)</p> <ul style="list-style-type: none"> - 1 (CO) - 2 (CH) - 3 (CH₃) - 4 (CH₂) - 5 (CH₃) 	<p align="center">4.85 (dd, <i>J</i> = 10.4, 4.0 Hz, 1H)</p> <p align="center">-</p> <p align="center">2.40 (sextet, <i>J</i> = 6.8 Hz, 1H)</p> <p align="center">1.12 (m, 3H)</p> <p align="center">1.42, 1.60 (m, 2H)</p> <p align="center">0.86 (t, <i>J</i> = 7.3 Hz, 3H)</p>	<p align="center">70.84</p> <p align="center">176.77</p> <p align="center">40.61</p> <p align="center">16.08</p> <p align="center">26.78</p> <p align="center">11.42</p>
<p>3 (CH)</p> <ul style="list-style-type: none"> - 1 (CO) - 2 (CH₂) - 3 (CH₂) - 4,5,6 (CH₂) - 7 (CH₂) - 8 (CH) - 9,10 (CH₃) 	<p align="center">5.56 (t, <i>J</i> = 10.0 Hz, 1H)</p> <p align="center">-</p> <p align="center">2.20 (t, <i>J</i> = 7.5 Hz, 2H)</p> <p align="center">1.51(m, 2H)</p> <p align="center">1.24 (m)</p> <p align="center">1.13 (m)</p> <p align="center">1.48 (m)</p> <p align="center">0.85 (m)</p>	<p align="center">69.00</p> <p align="center">173.04</p> <p align="center">34.16</p> <p align="center">24.68</p> <p align="center">29.34</p> <p align="center">38.82</p> <p align="center">28.08</p> <p align="center">22.71</p>
<p>4 (CH)</p> <ul style="list-style-type: none"> - 1 (CO) - 2 (CH) - 3,4 (CH₃) 	<p align="center">4.90 (t, <i>J</i> = 10.0 Hz, 1H)</p> <p align="center">-</p> <p align="center">2.53 (hept, <i>J</i> = 7.0 Hz, 1H)</p> <p align="center">1.11 (d, <i>J</i> = 7.0 Hz, 6H)</p>	<p align="center">68.49</p> <p align="center">176.16</p> <p align="center">34.03</p> <p align="center">18.85</p>
<p>5 (CH)</p>	<p align="center">4.20 (m, 1H)</p>	<p align="center">71.85</p>
<p>6 (CH₂)</p>	<p align="center">3.60, 3.60 (m, 2H)</p>	<p align="center">61.55</p>
<p>1' (CH₂)</p>	<p align="center">3.61 (m, 1H), 3.51 (d, <i>J</i> = 11.9 Hz, 2H)</p>	<p align="center">64.60</p>
<p>2' (C)</p>	<p align="center">-</p>	<p align="center">104.52</p>
<p>3' (CH)</p>	<p align="center">4.24 (m, 1H)</p>	<p align="center">78.07</p>
<p>4' (CH)</p>	<p align="center">4.30 (t, <i>J</i> = 8.5 Hz, 1H)</p>	<p align="center">72.96</p>
<p>5' (CH)</p>	<p align="center">3.74 (m, 1H)</p>	<p align="center">81.35</p>
<p>6' (CH₂)</p>	<p align="center">3.88 (d, <i>J</i> = 13.1 Hz, 1H), 3.74 (m, 1H)</p>	<p align="center">60.06</p>

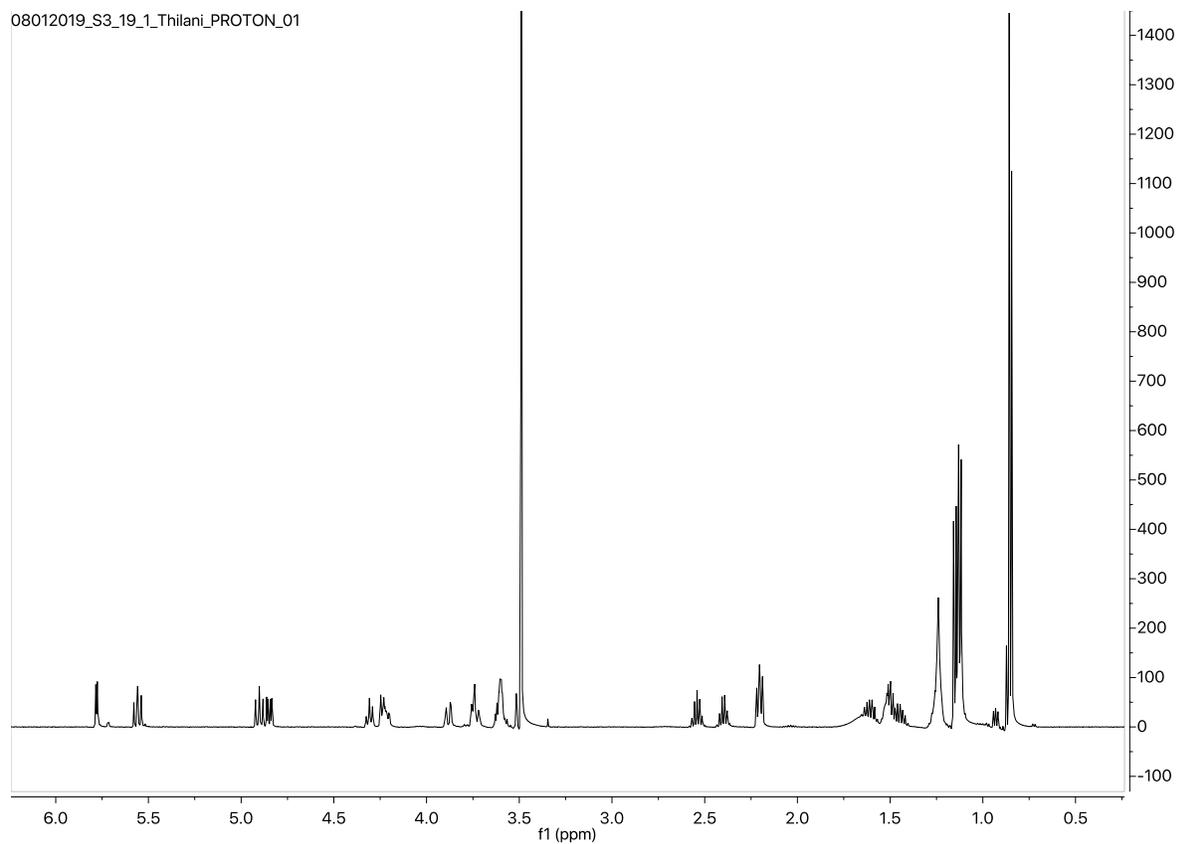


Figure S4.21 ^1H NMR spectrum for S3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

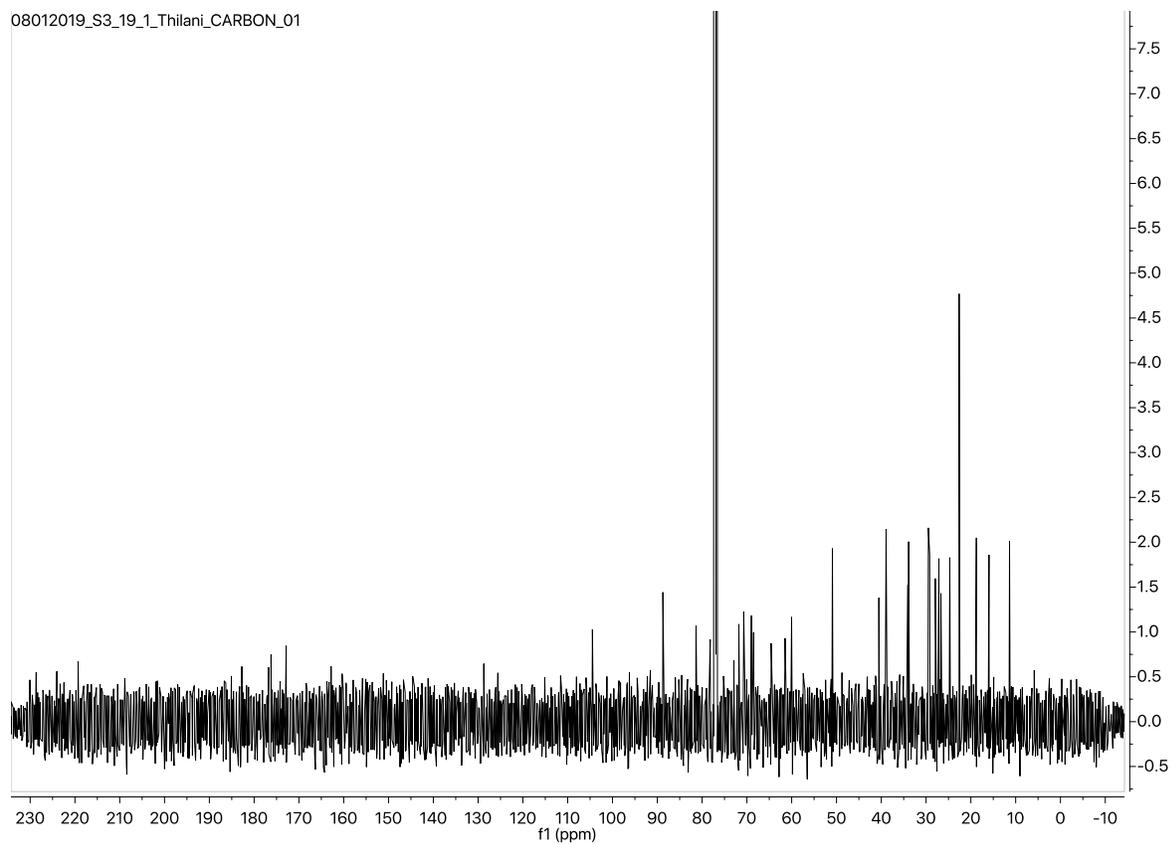


Figure S4.22 ^{13}C NMR spectrum for S3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

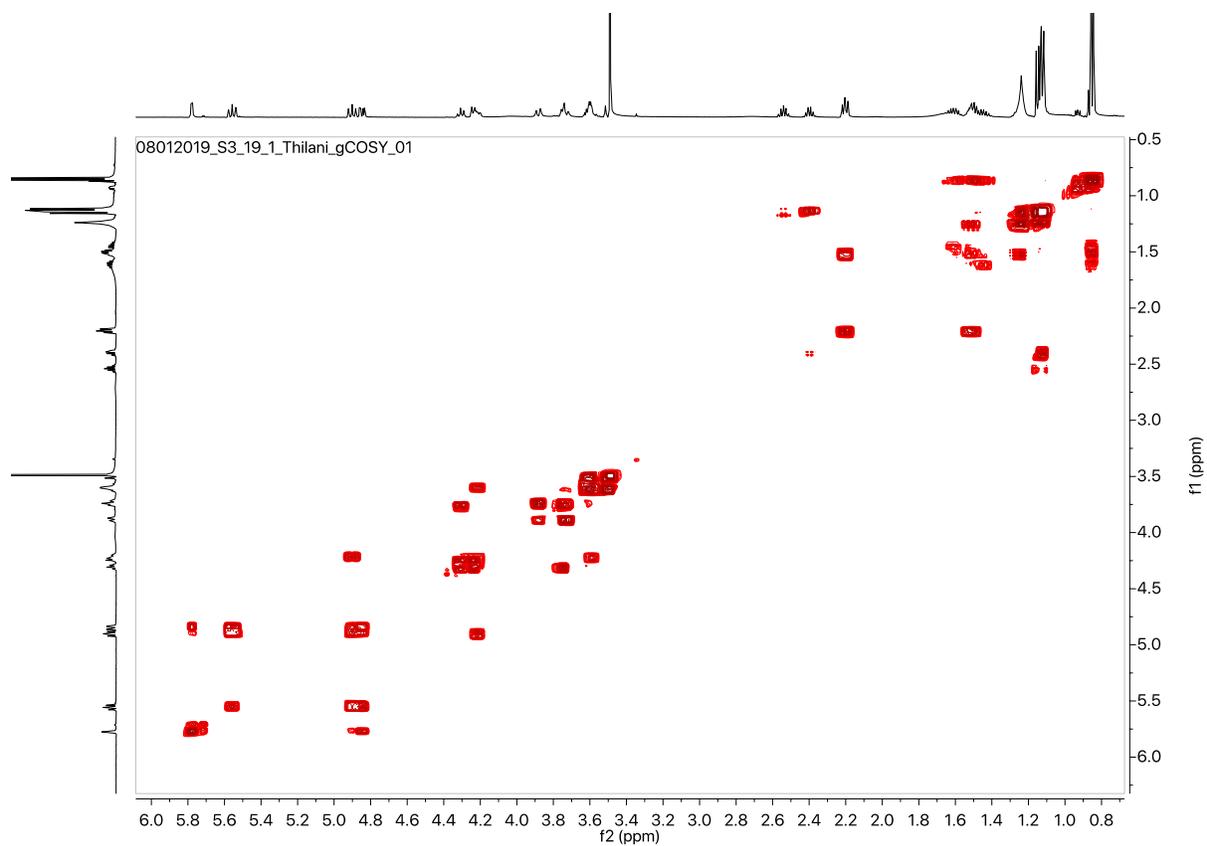


Figure S4.23 gCOSY NMR spectrum for S3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

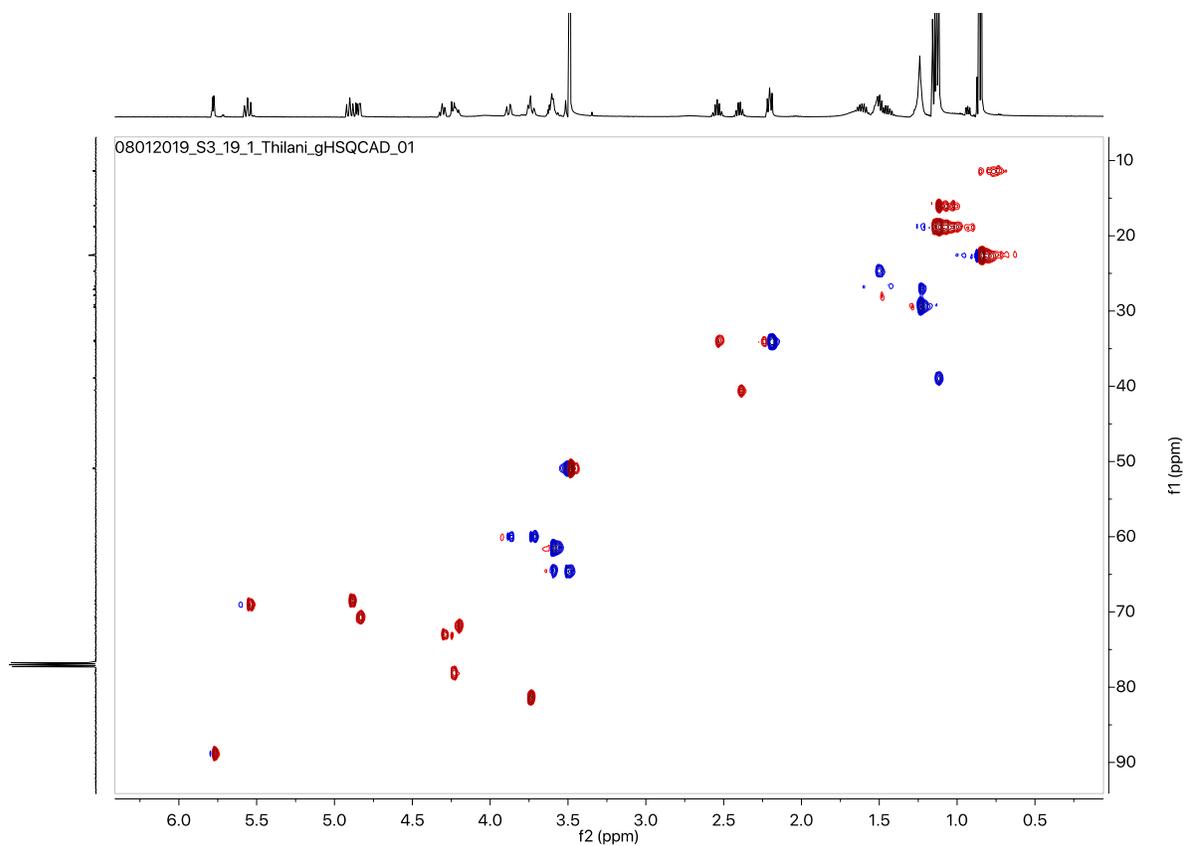


Figure S4.24 gHSQCAD NMR spectrum for S3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

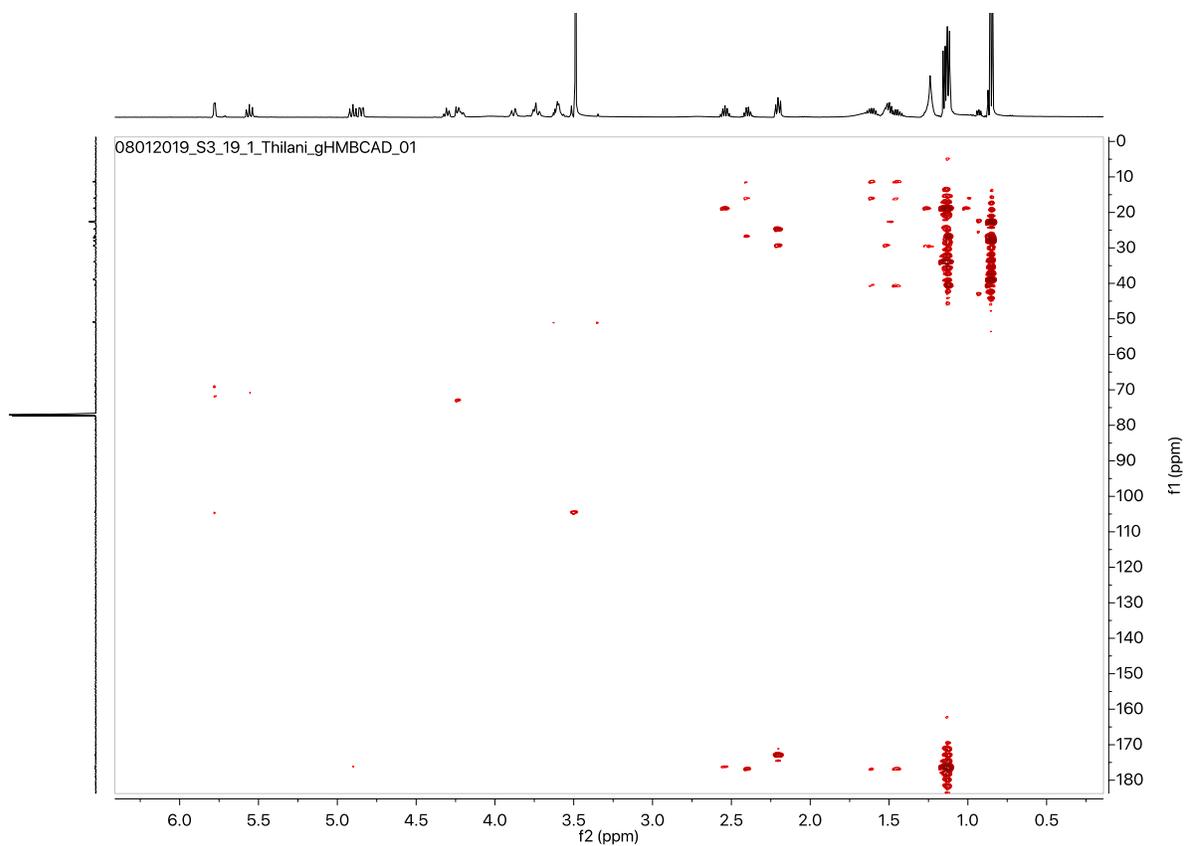


Figure S4.25 gHMBCAD NMR spectrum for S3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

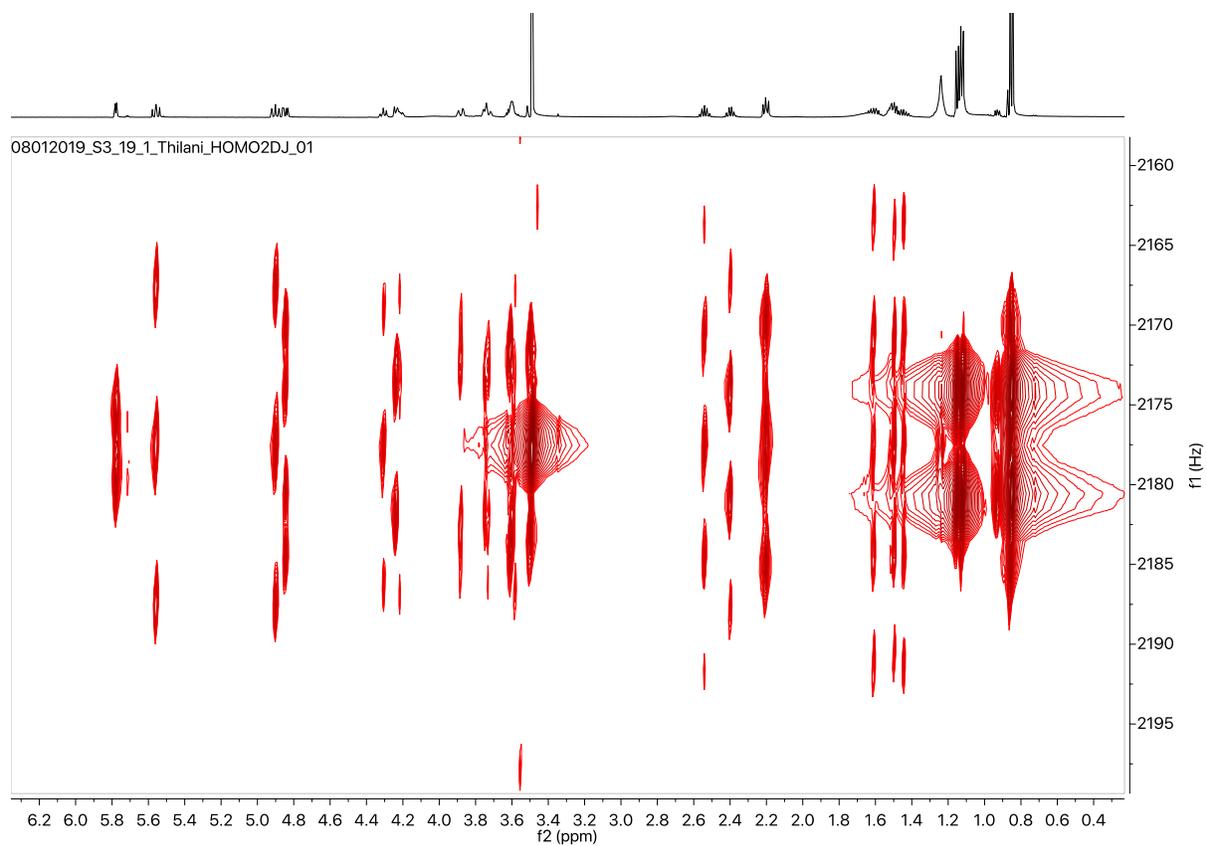
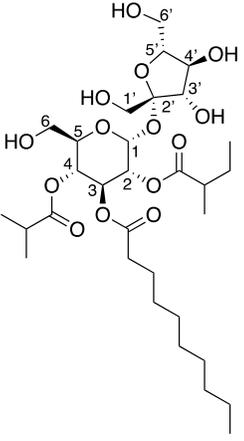


Figure S4.26 ^1H - ^1H HOMO2DJ NMR spectrum for S3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

Table S4.6 NMR chemical shifts for S3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

Carbon # (group)	¹ H (ppm)	¹³ C (ppm) (from HSQC and HMBC)
 <p>S3:19(4,5,10)-2 Purified from <i>S. pennellii</i> LA0716 Chemical Formula: C₃₁H₅₄O₁₄ HRMS: (ESI) <i>m/z</i> calculated for C₃₁H₅₄O₁₄ ([M+NH₄]⁺): 668.3852 Experimental <i>m/z</i>: 668.3855 NMR (500 MHz, CDCl₃) Sample mass: 2 mg</p>		
1 (CH)	5.78 (d, <i>J</i> = 4.0 Hz, 1H)	88.86
2 (CH)	4.85 (dd, <i>J</i> = 10.4, 4.0 Hz, 1H)	70.68
- 1 (CO)	-	176.81
- 2 (CH)	2.40 (sextet, <i>J</i> = 6.8 Hz, 1H)	40.57
- 3 (CH ₃)	1.13 (m, 3H)	16.03
- 4 (CH ₂)	1.45, 1.61 (m, 2H)	26.62
- 5 (CH ₃)	0.87 (t, <i>J</i> = 7.3 Hz, 3H)	11.36
3 (CH)	5.56 (t, <i>J</i> = 10.0 Hz, 1H)	68.48
- 1 (CO)	-	172.64
- 2 (CH ₂)	2.20 (t, <i>J</i> = 7.5 Hz, 2H)	34.07
- 3 (CH ₂)	1.51 (m, 2H)	24.75
- 4,5,6,7 (CH ₂)	1.25 (m)	29.26
- 8 (CH ₂)	1.24 (m)	31.92
- 9 (CH ₂)	1.25 (m)	22.49
- 10 (CH ₃)	0.87 (m)	14.13
4 (CH)	4.91 (t, <i>J</i> = 10.0 Hz, 1H)	68.48
- 1 (CO)	-	176.16
- 2 (CH)	2.54 (hept, <i>J</i> = 6.84 Hz, 1H)	33.86
- 3,4 (CH ₃)	1.13 (d, <i>J</i> = 6.84 Hz, 6H)	18.82
5 (CH)	4.20 (m, 1H)	71.92
6 (CH ₂)	3.60, 3.60 (m, 2H)	61.51
1' (CH ₂)	3.61, 3.51 (d, <i>J</i> = 11.9 Hz, 2H)	64.56
2' (C)	-	104.45
3' (CH)	4.24 (m, 1H)	78.32
4' (CH)	4.31 (t, <i>J</i> = 8.4 Hz, 2H)	72.98
5' (CH)	3.75 (m, 1H)	81.43
6' (CH ₂)	3.87 (d, <i>J</i> = 13.1 Hz, 1H), 3.74 (m, 1H)	60.09

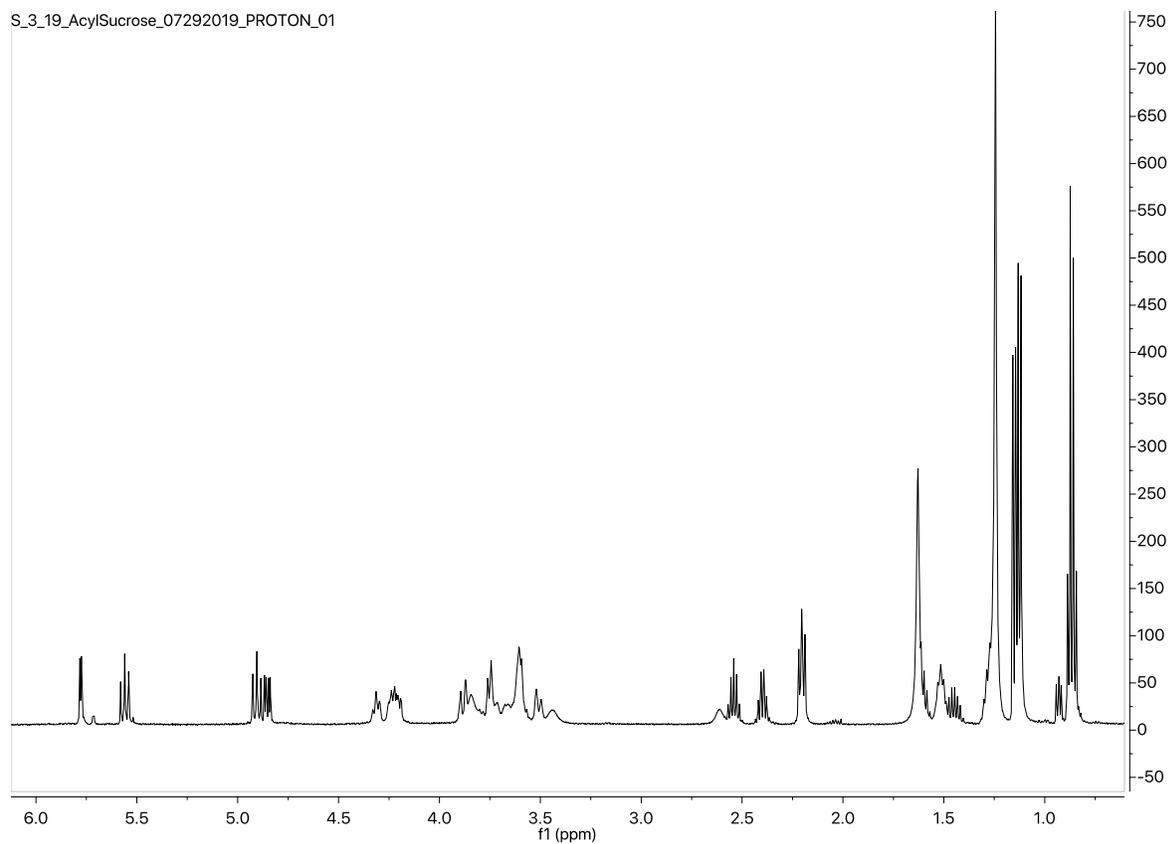


Figure S4.27 ^1H NMR spectrum for S3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

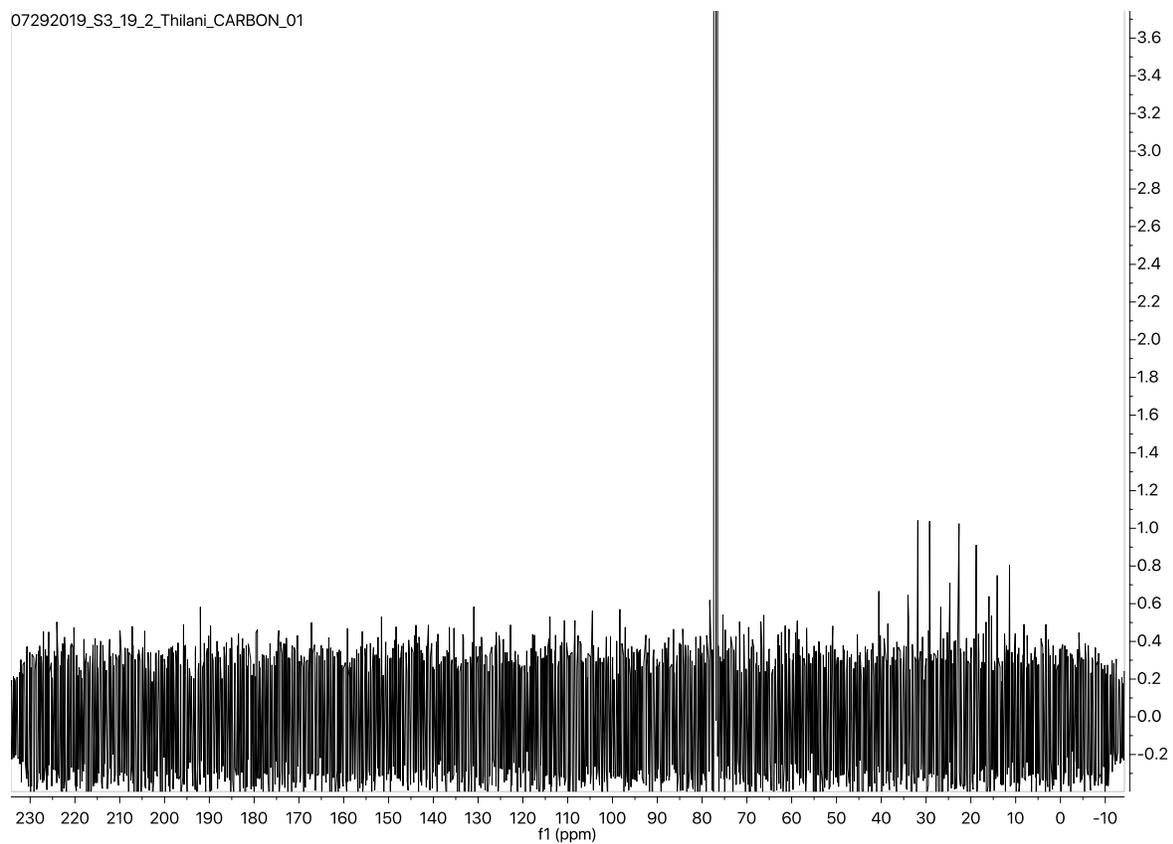


Figure S4.28 ^{13}C NMR spectrum for S3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

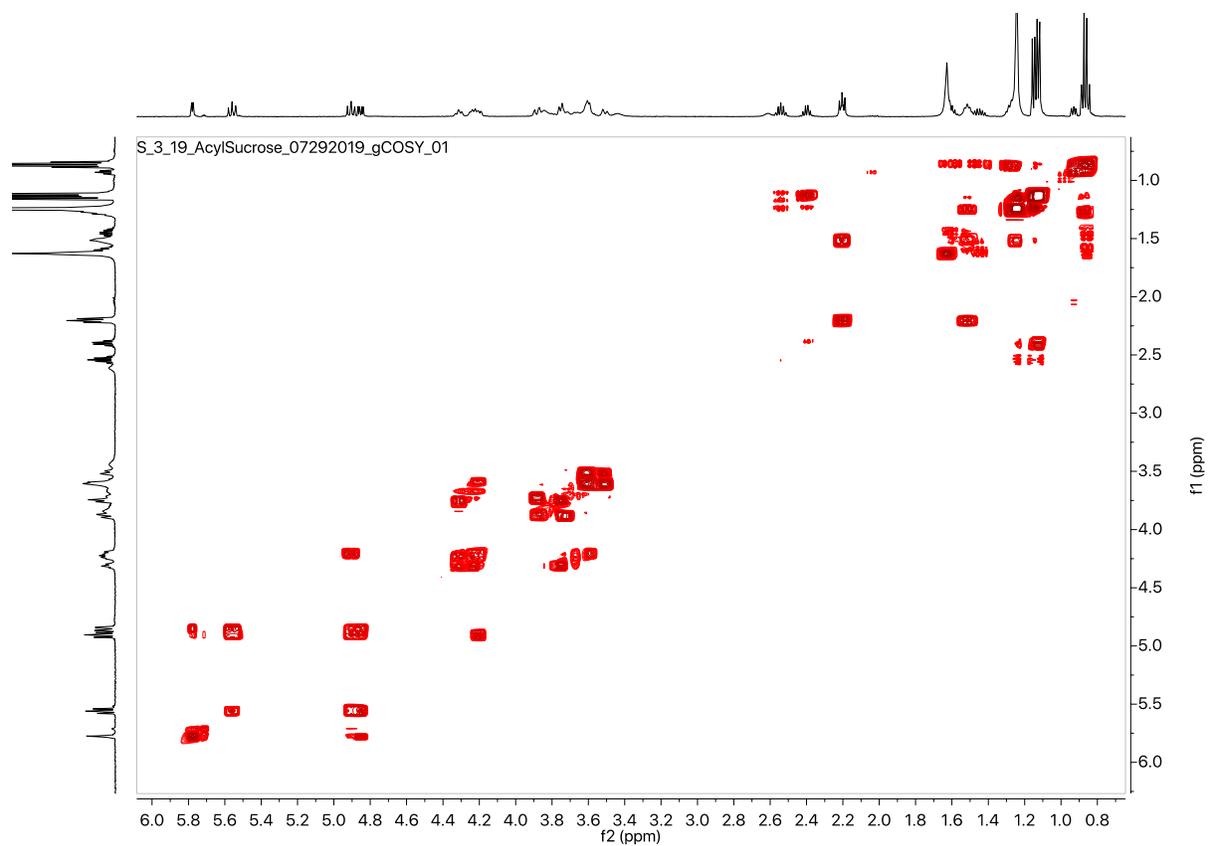


Figure S4.29 gCOSY NMR spectrum for S3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

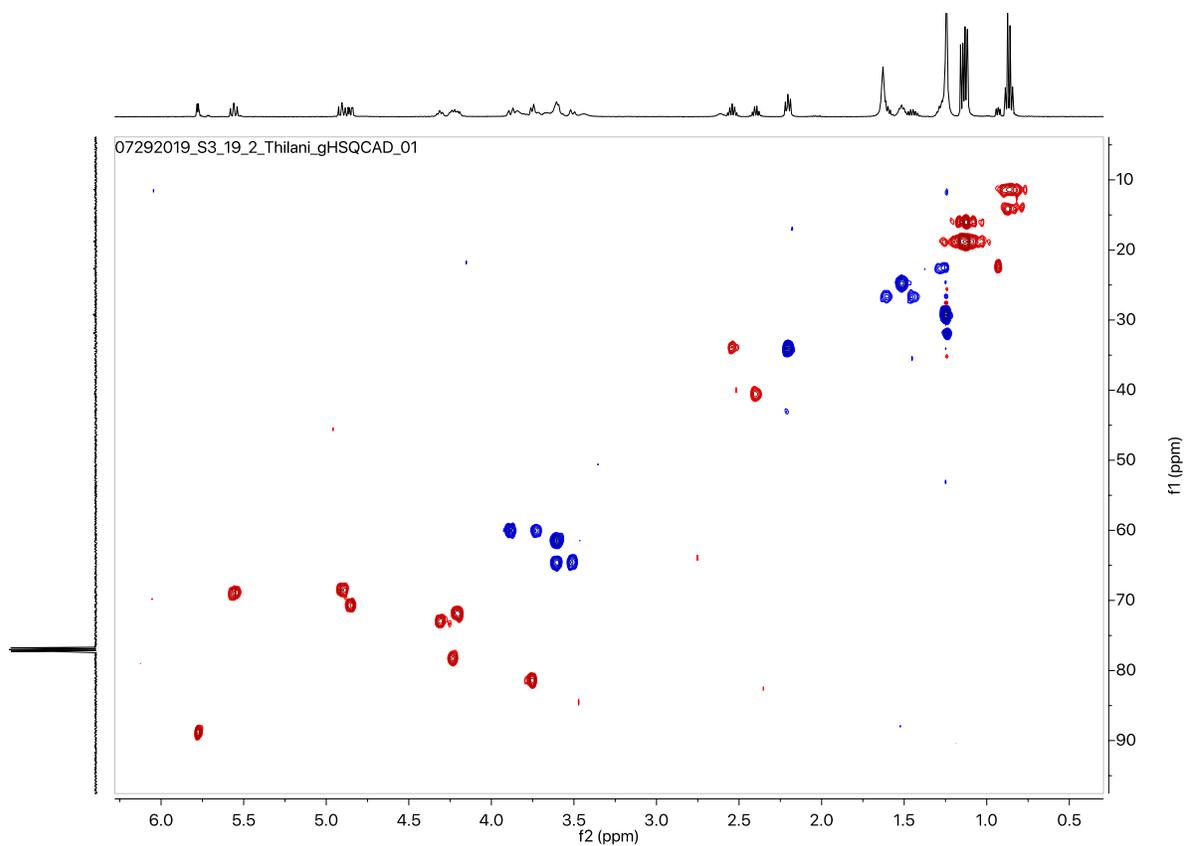


Figure S4.30 gHSQCAD NMR spectrum for S3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

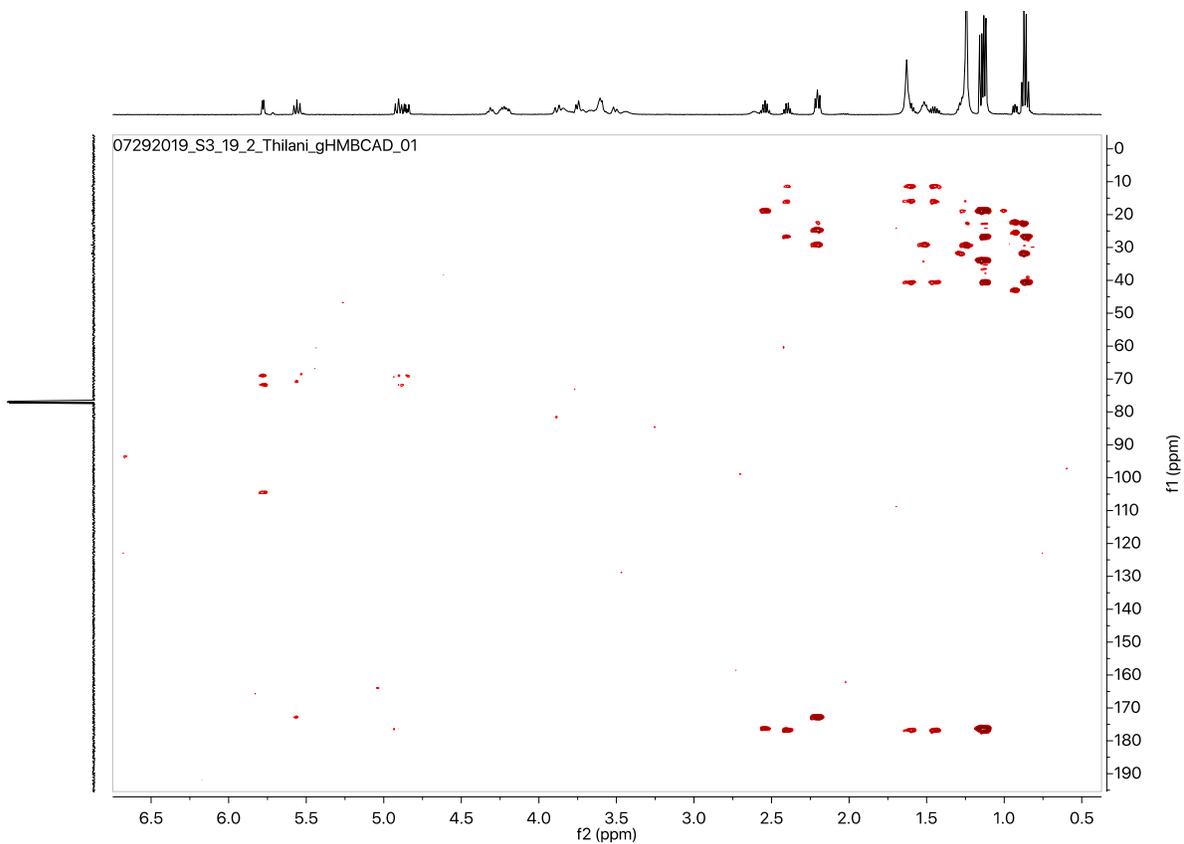


Figure S4.31 gHMBCAD NMR spectrum for S3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

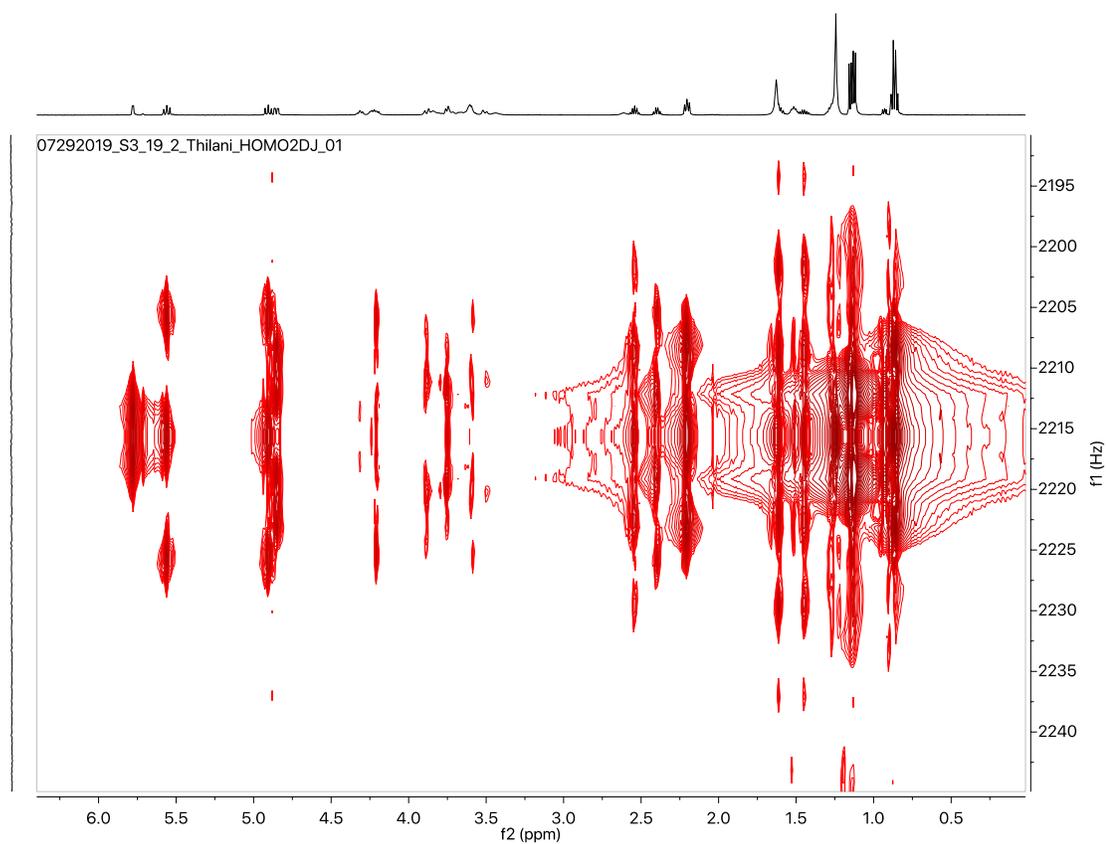


Figure S4.32 ^1H - ^1H HOMO 2DJ NMR spectrum for S3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

Table S4.7 NMR chemical shifts for G3:12(4,4,4) purified from *S. pennellii* LA0716.

Carbon # (group)	¹ H (ppm)		¹³ C (ppm) (from HSQC and HMBC)	
	α	β	α	β
1 (CH)	5.49 (d, <i>J</i> = 3.7 Hz)	4.74 (d, <i>J</i> = 7.6 Hz)	90.28	95.75
2 (CH)	4.91 (m)	4.89 (m)	71.06	73.30
- 1 (CO)	-	-	175.83	175.83
- 2 (CH)	2.49 (hept, <i>J</i> = 7.0 Hz)	2.49 (hept, <i>J</i> = 7.0 Hz)	33.93	33.93
- 3,4 (CH ₃)	1.10 (m)	1.10 (m)	18.87	18.87
3 (CH)	5.67 (t, <i>J</i> = 9.9 Hz)	5.39 (t, <i>J</i> = 9.7 Hz)	68.67	71.21
- 1 (CO)	-	-	175.91	175.91
- 2 (CH)	2.56 (hept, <i>J</i> = 7.0 Hz)	2.56 (hept, <i>J</i> = 7.0 Hz)	33.87	33.87
- 3,4 (CH ₃)	1.14 (m)	1.14 (m)	18.85	18.85
4 (CH)	5.02 (t, <i>J</i> = 9.7 Hz)	5.02 (t, <i>J</i> = 9.7 Hz)	68.53	68.53
- 1 (CO)	-	-	176.83	176.83
- 2 (CH)	2.49 (hept, <i>J</i> = 7.0 Hz)	2.49 (hept, <i>J</i> = 7.0 Hz)	33.93	33.93
- 3,4 (CH ₃)	1.10 (m)	1.10 (m)	18.87	18.87
5 (CH)	4.08 (ddd, <i>J</i> = 10.3, 4.2, 2.3 Hz)	3.58 (m)	69.47	74.54
6 (CH ₂)	3.57, 3.68 (m)	3.57, 3.68 (m)	61.05	61.05

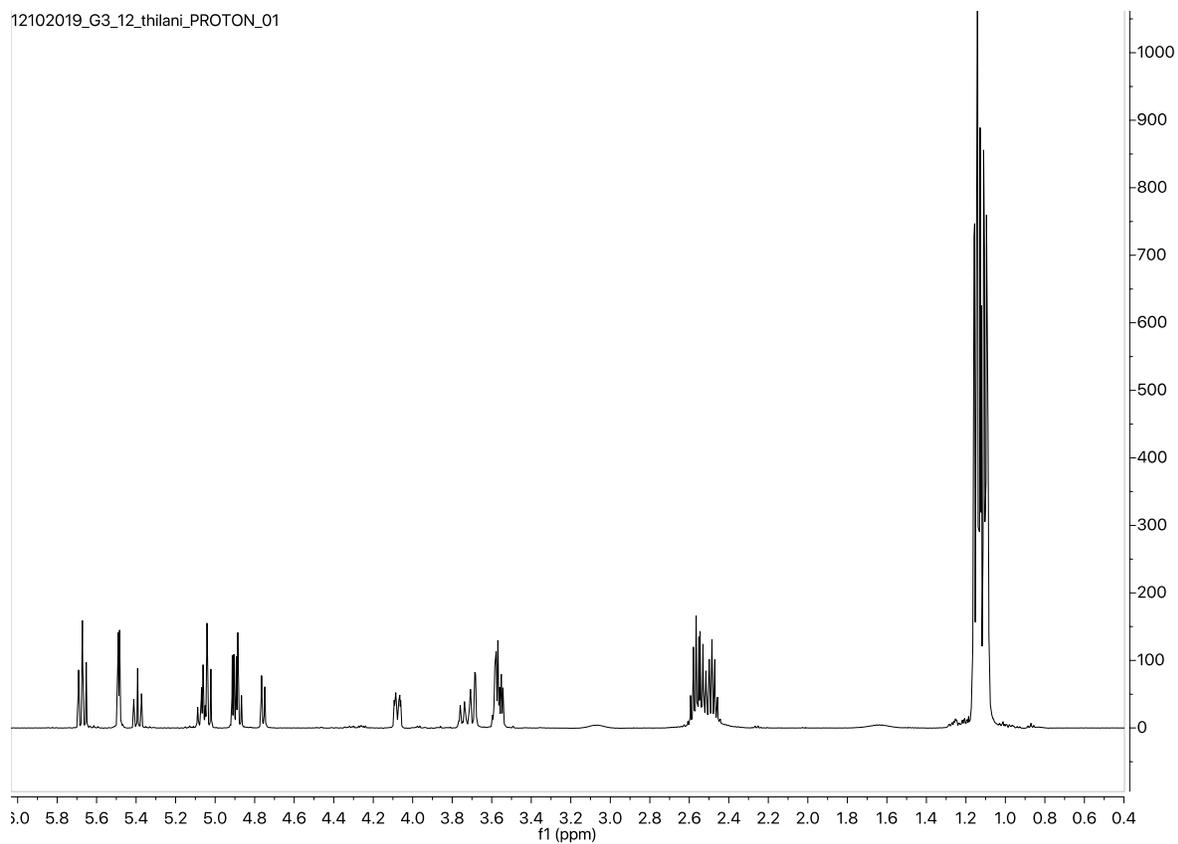


Figure S4.33 ^1H NMR spectrum for G3:12(4,4,4) purified from *S. pennellii* LA0716.

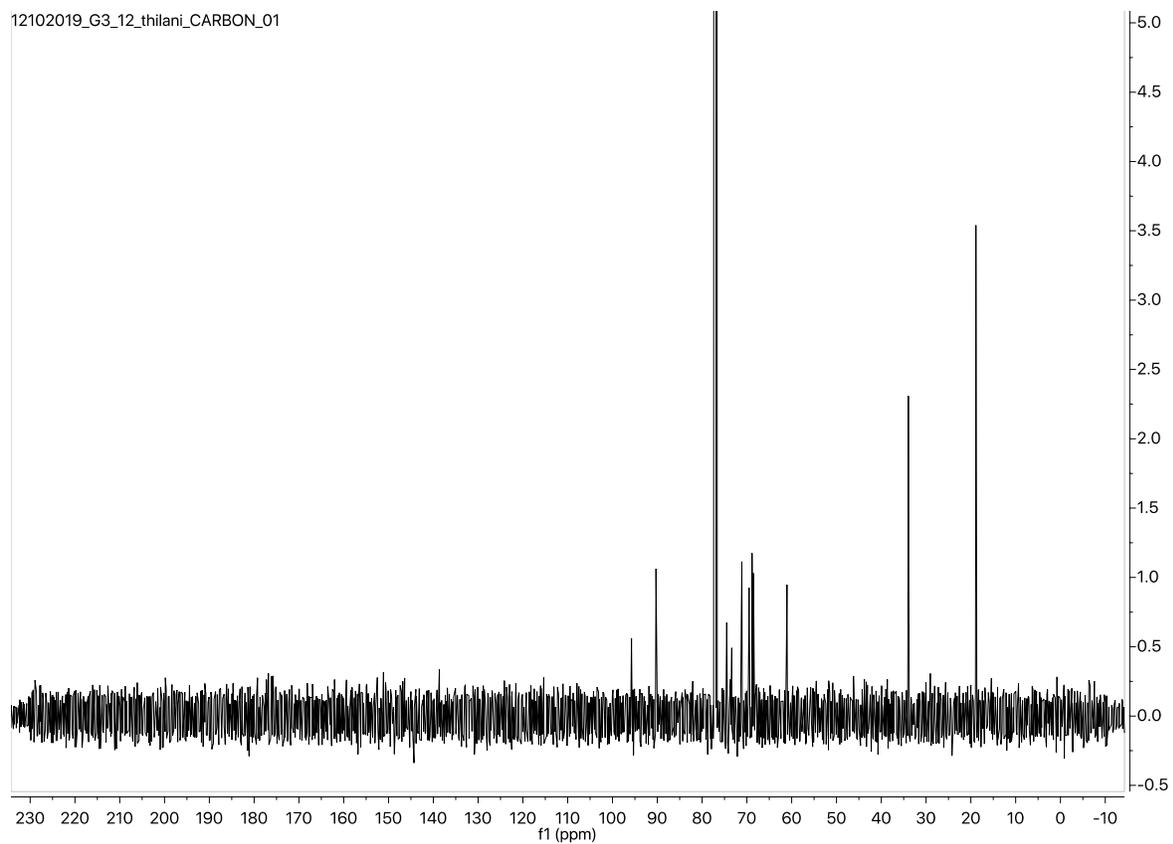


Figure S4.34 ^{13}C NMR spectrum for G3:12(4,4,4) purified from *S. pennellii* LA0716.

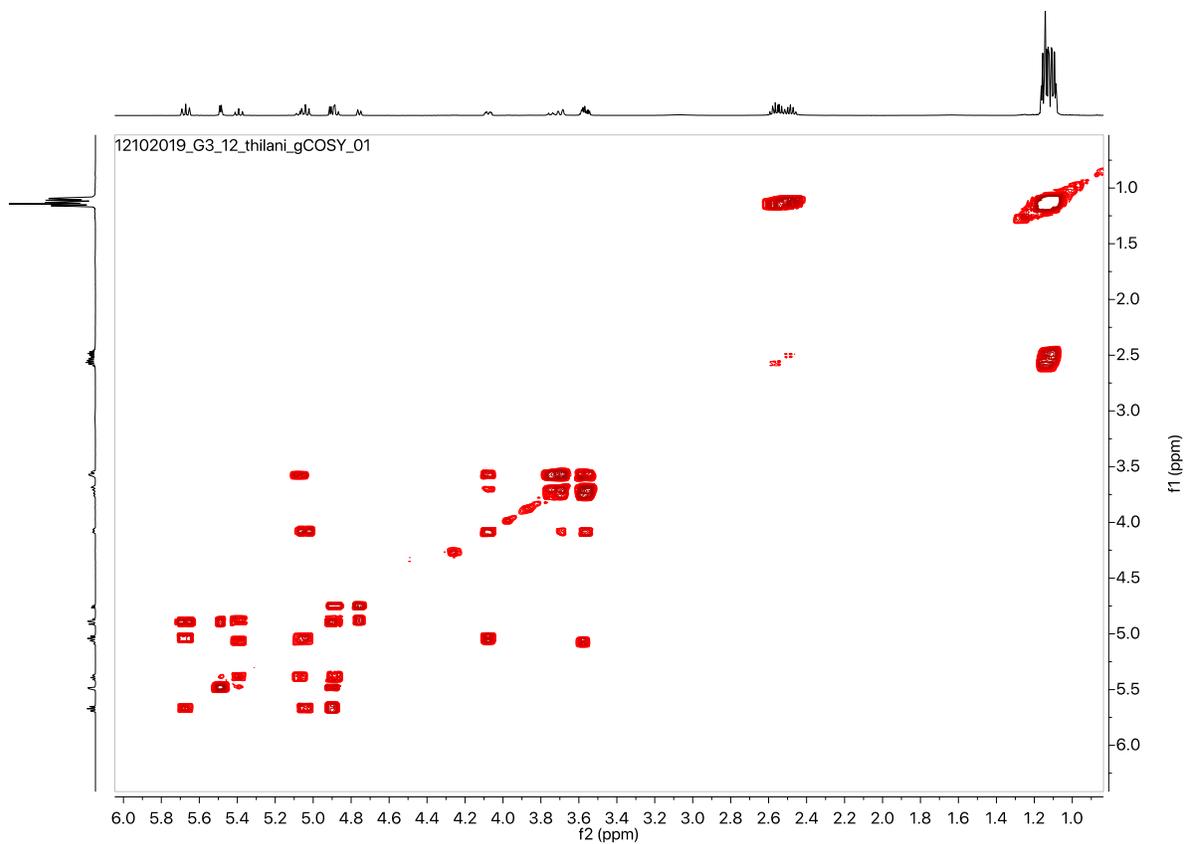


Figure S4.35 gCOSY NMR spectrum for G3:12(4,4,4) purified from *S. pennellii* LA0716.

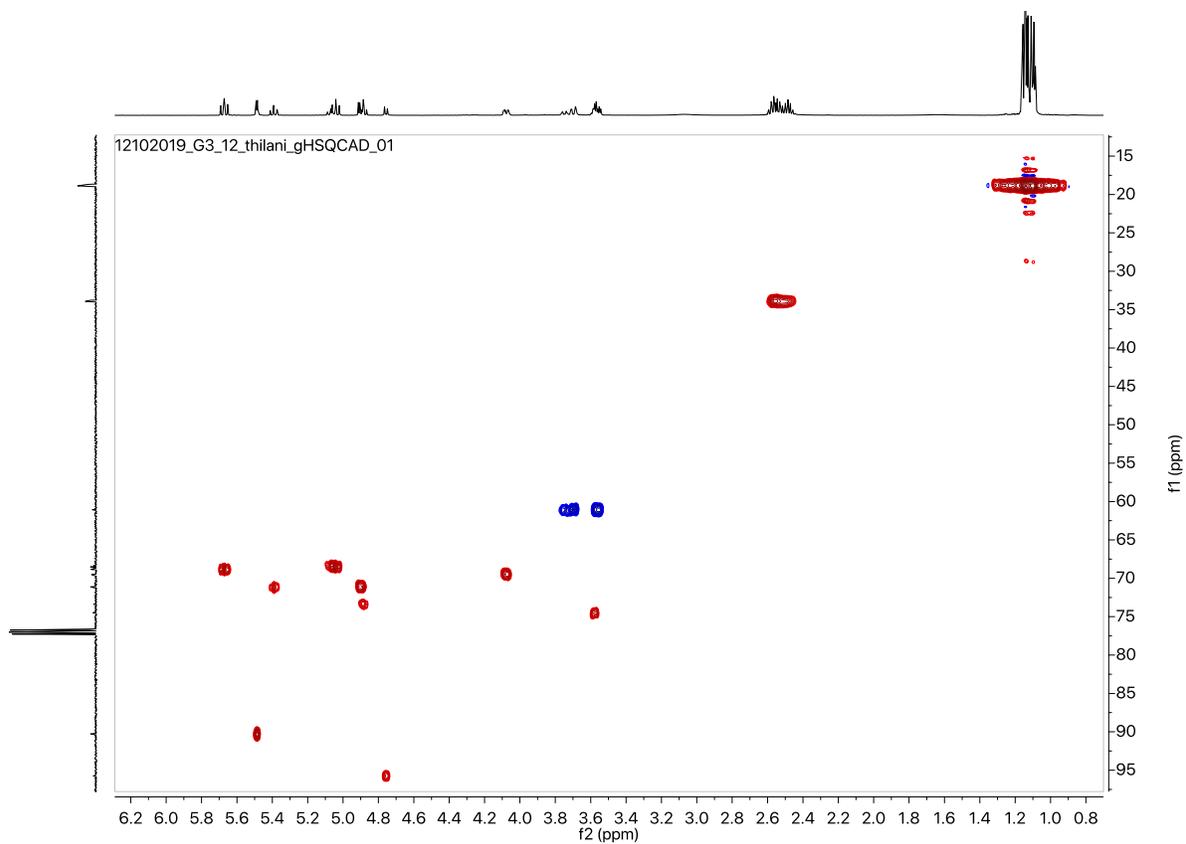


Figure S4.36 gHSQCAD NMR spectrum for G3:12(4,4,4) purified from *S. pennellii* LA0716.

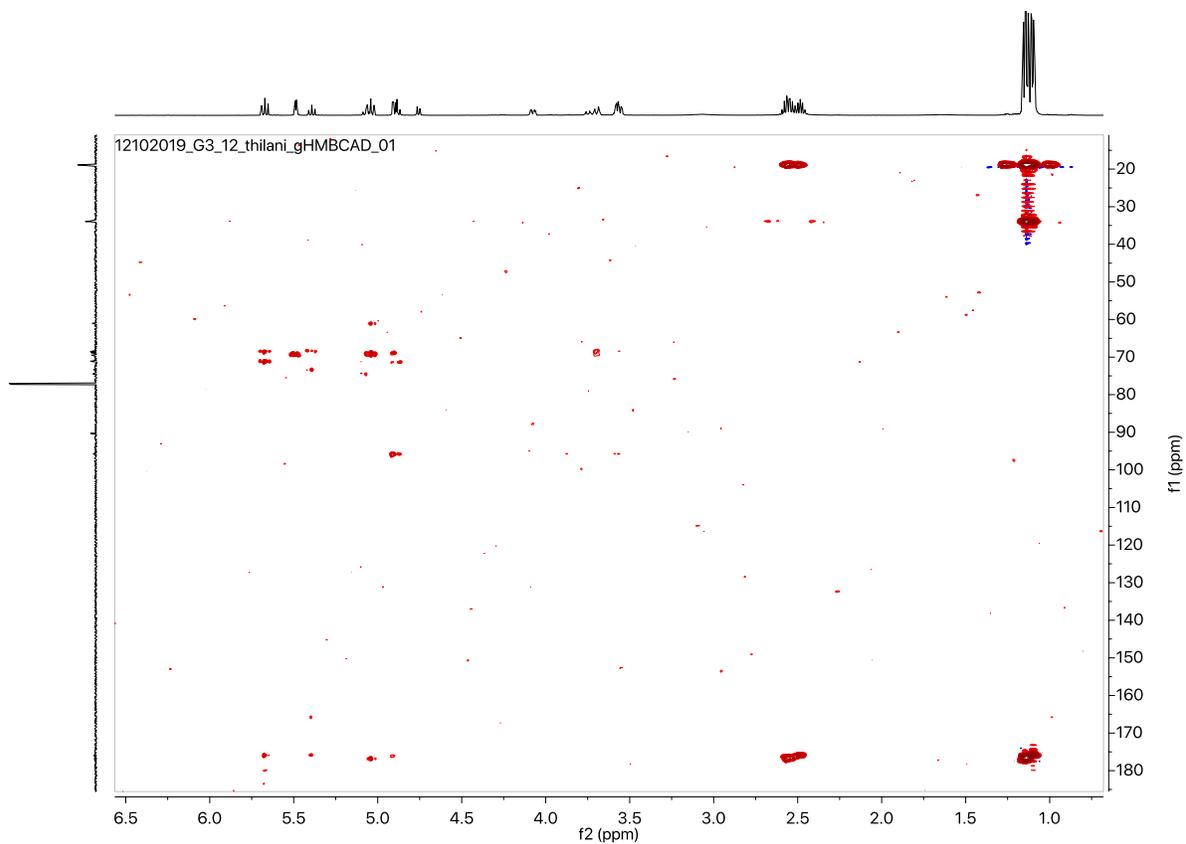


Figure S4.37 gHMBCAD NMR spectrum for G3:12(4,4,4) purified from *S. pennellii* LA0716.

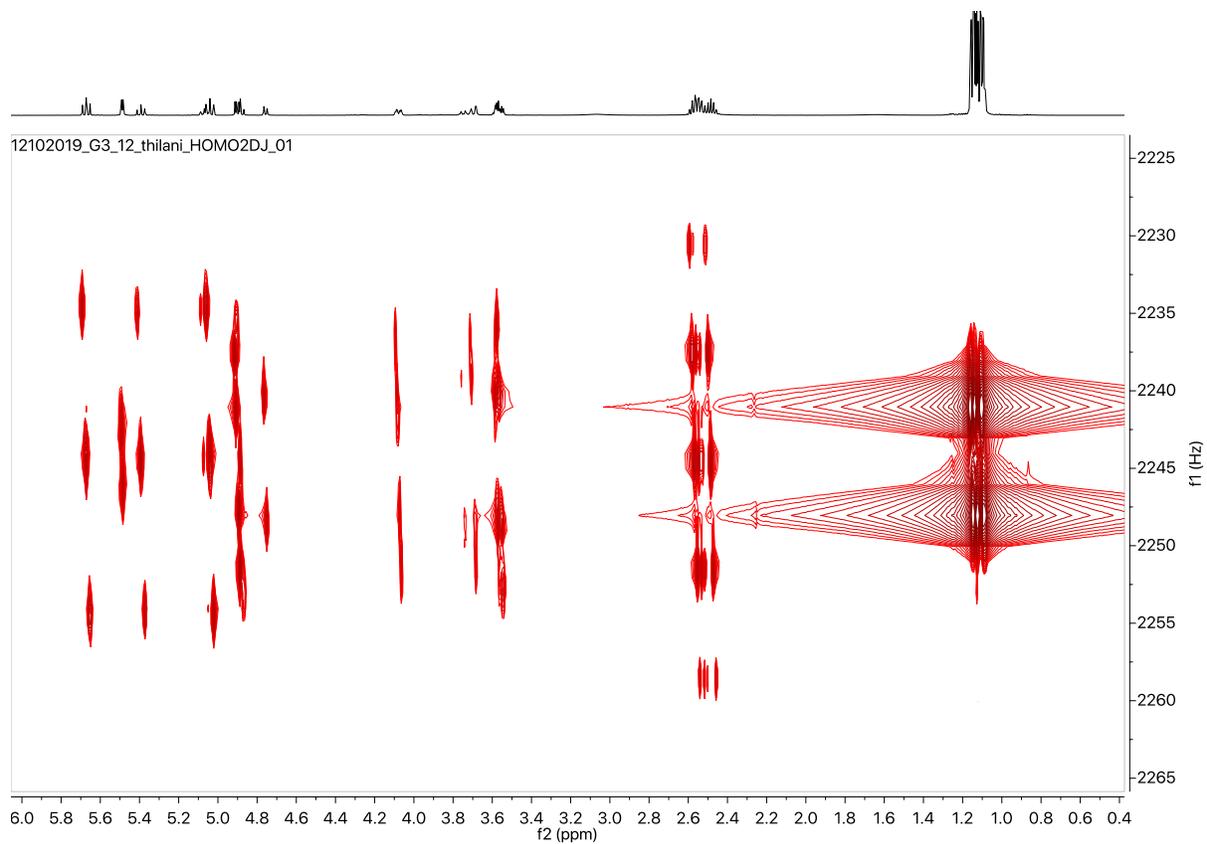


Figure S4.38 ^1H - ^1H HOMO2DJ NMR spectrum for G3:12(4,4,4) purified from *S. pennellii* LA0716.

Table S4.8 NMR chemical shifts for G3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

Carbon # (group)	¹ H (ppm)		¹³ C (ppm) (from HSQC and HMBC)	
	α	β	α	β
1 (CH)	5.50 (d, <i>J</i> = 3.7 Hz, 1H)	4.75 (d, <i>J</i> = 8.1 Hz, 1H)	90.25	95.78
2 (CH) - 1 (CO) - 2 (CH) - 3,4 (CH ₃)	4.88 (dd, <i>J</i> = 9.9, 3.7 Hz) - 2.56 (hept, <i>J</i> = 7.0 Hz) 1.14 (m)	4.85 (m) - 2.56 (hept, <i>J</i> = 7.0 Hz) 1.14 (m)	71.13 176.73 33.92 18.82	73.44 176.73 33.92 18.82
3 (CH) - 1 (CO) - 2 (CH ₂) - 3 (CH ₂) - 4,5,6 (CH ₂) - 7 (CH ₂) - 8 (CH) - 9,10 (CH ₃)	5.69 (t, <i>J</i> = 9.9 Hz) - 2.23 (t, <i>J</i> = 7.4 Hz) 1.54(m) 1.25(m) 1.24 (m) 1.50 (m) 0.85 (d, <i>J</i> = 6.6 Hz, 6H)	5.41 (t, <i>J</i> = 9.6 Hz) - 2.23 (t, <i>J</i> = 7.4 Hz) 1.54(m) 1.25(m) 1.24 (m) 1.50 (m) 0.85 (d, <i>J</i> = 6.6 Hz, 6H)	68.83 172.62 34.09 24.84 29.37 27.15 27.93 22.65	71.15 172.62 34.09 24.84 29.37 27.15 27.93 22.65
4 (CH) - 1 (CO) - 2 (CH) - 3,4 (CH ₃)	5.02 (m) - 2.56 (hept, <i>J</i> = 7.0 Hz) 1.14 (m)	5.02(m) - 2.56 (hept, <i>J</i> = 7.0 Hz) 1.14 (m)	68.53 176.73 33.92 18.82	68.53 176.73 33.92 18.82
5 (CH)	4.06 (ddd, <i>J</i> = 10.2, 4.0, 2.2 Hz)	3.56 (m)	69.53	74.52
6 (CH ₂)	3.53, 3.66 (m)	3.53, 3.66 (m)	61.00	61.00

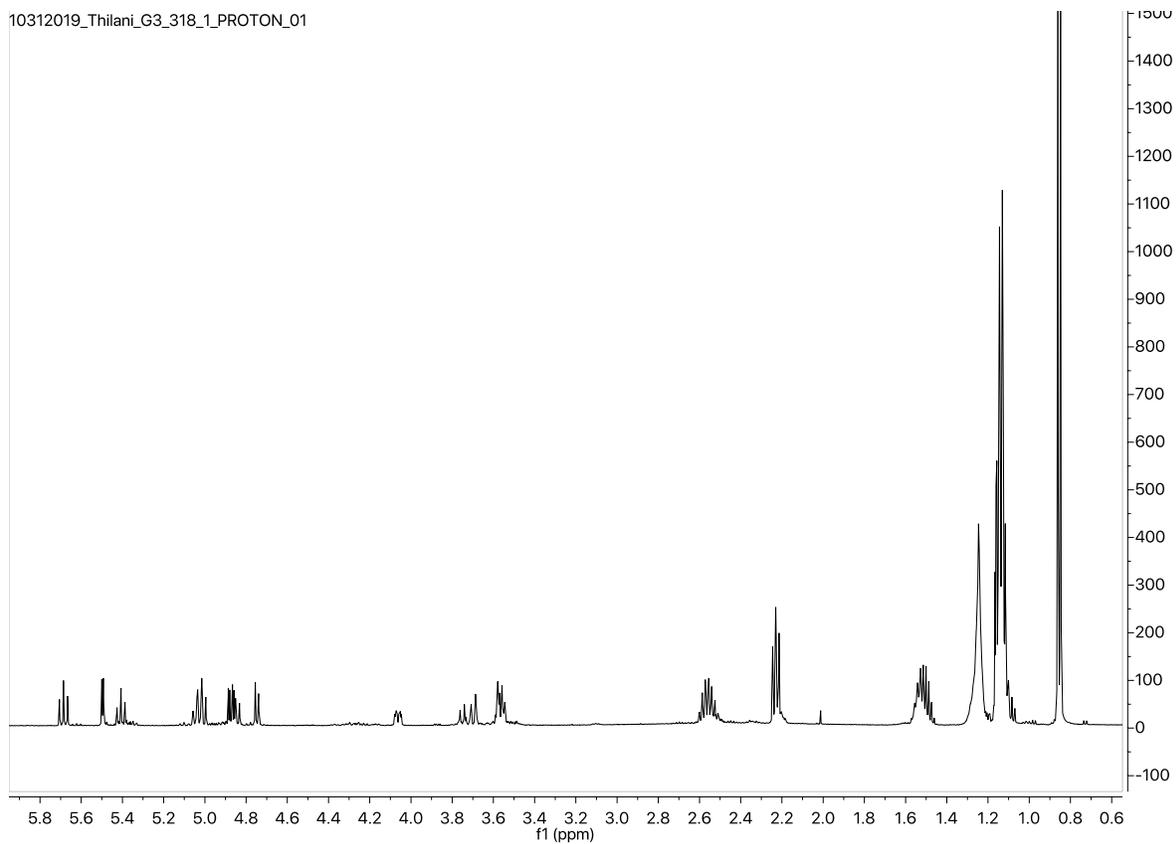


Figure S4.39 ^1H NMR spectrum for G3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

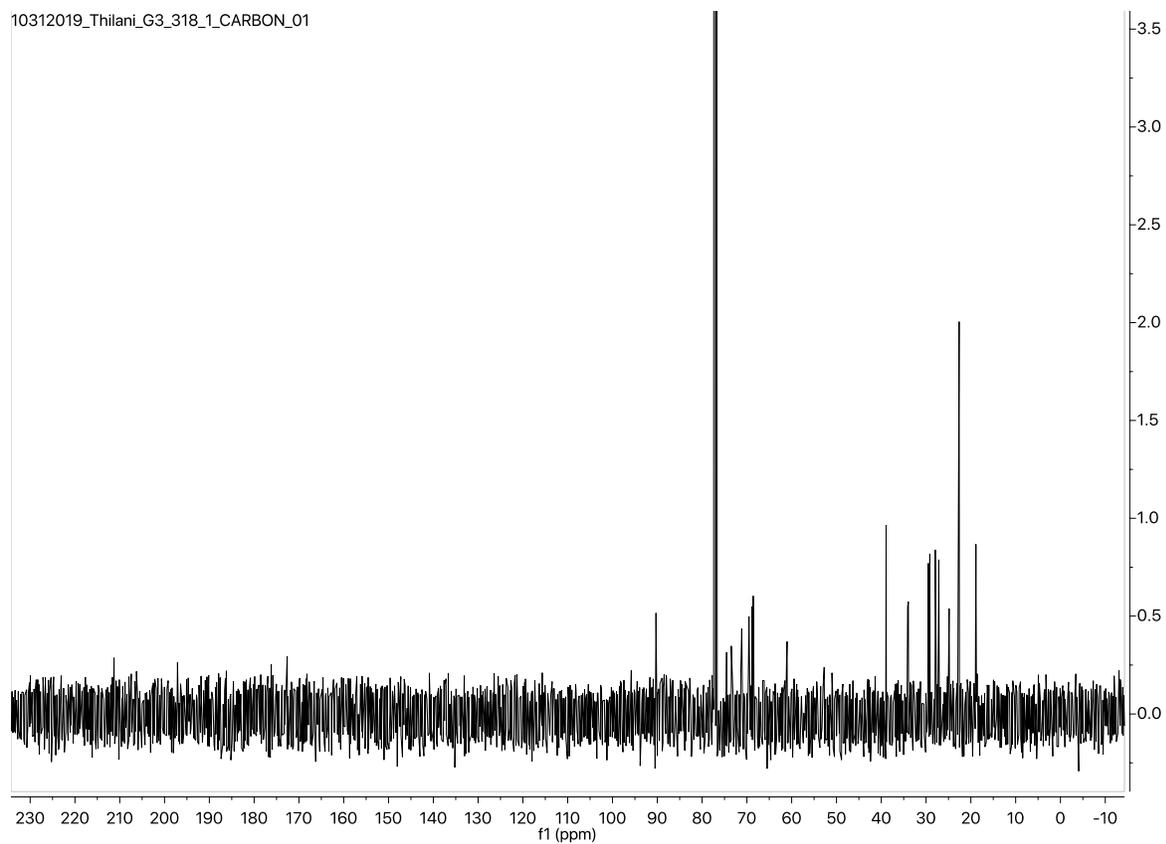


Figure S4.40 ^{13}C NMR spectrum for G3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

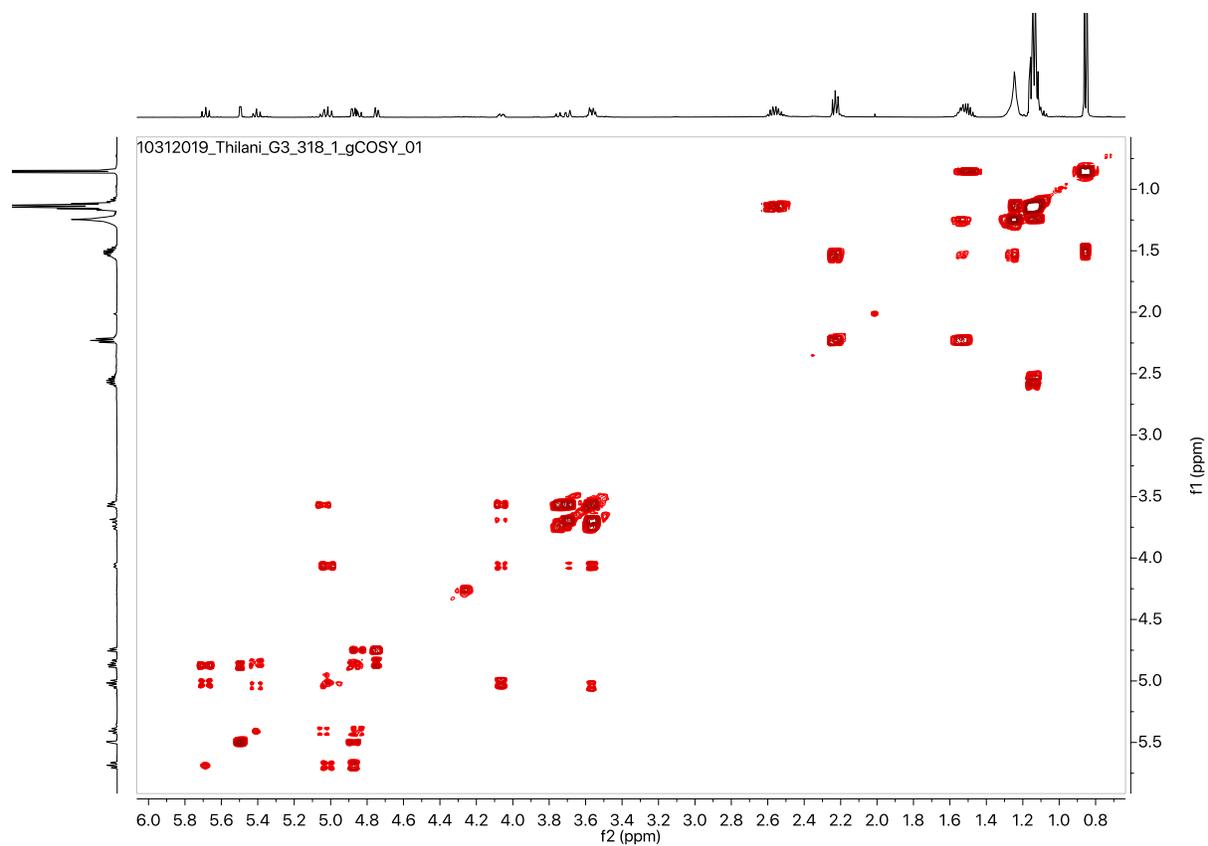


Figure S4.41 gCOSY NMR spectrum for G3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

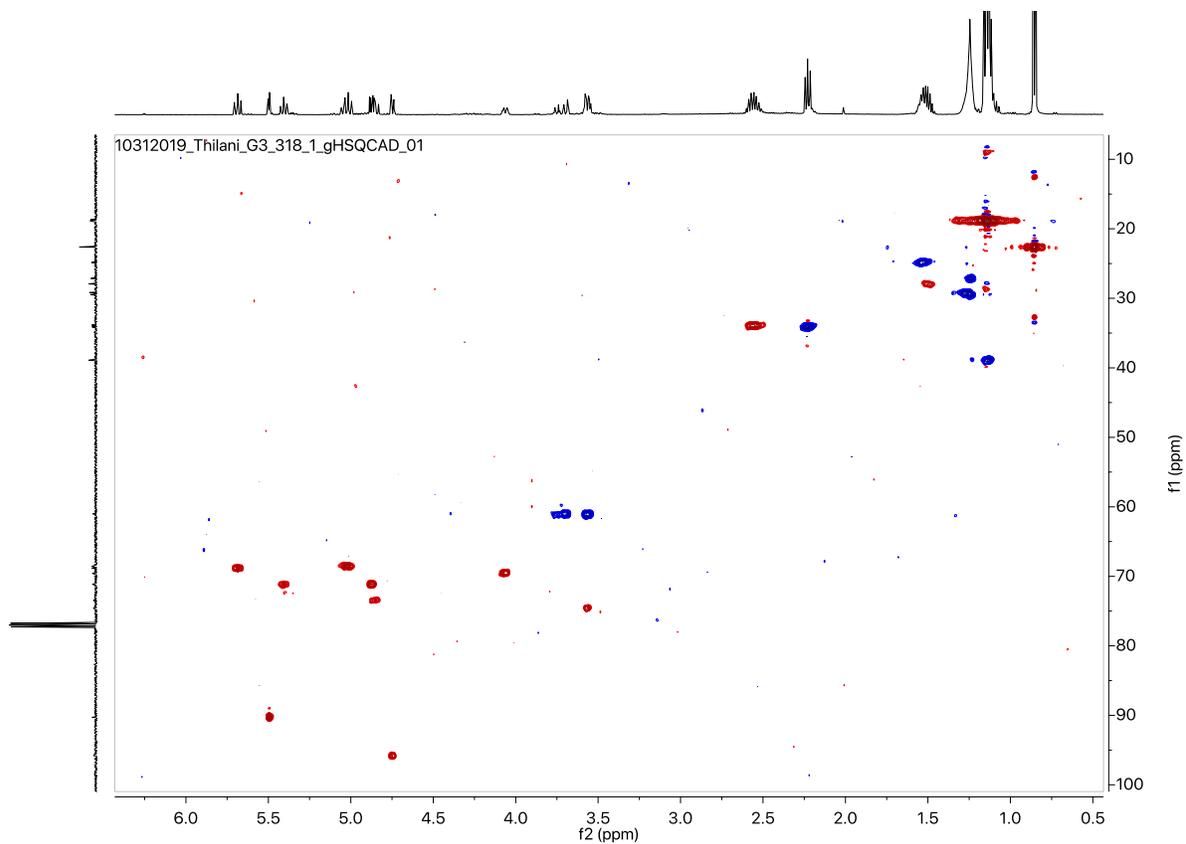


Figure S4.42 gHSQCAD NMR spectrum for G3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

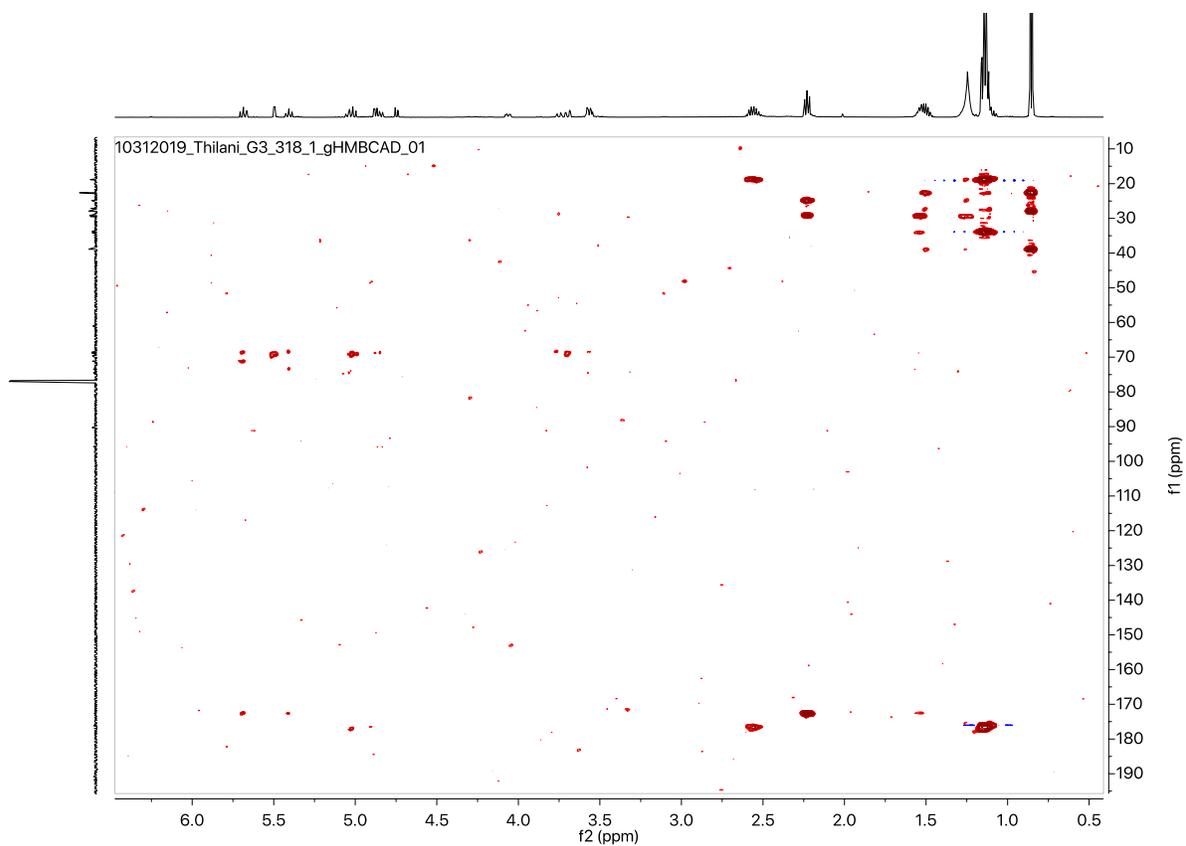


Figure S4.43 gHMBCAD NMR spectrum for G3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

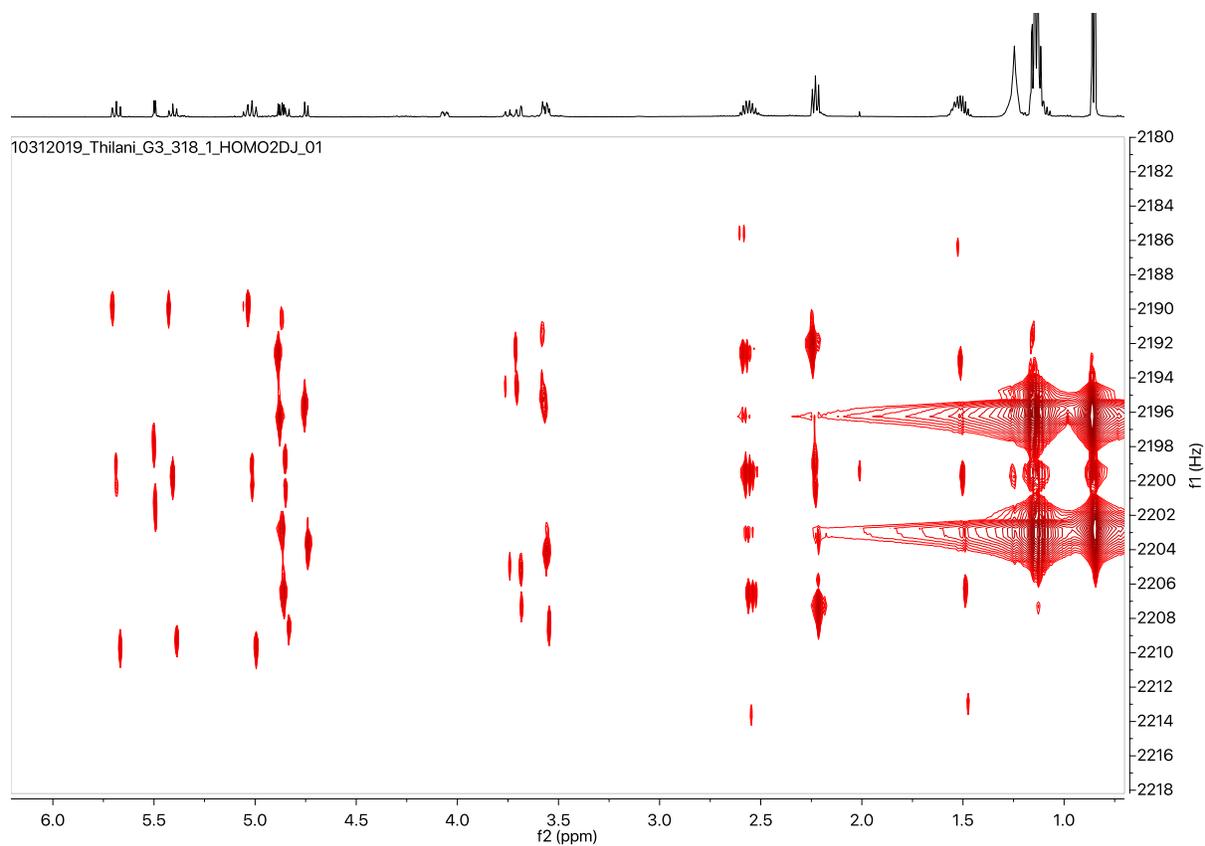


Figure S4.44 ^1H - ^1H HOMO2DJ NMR spectrum for G3:18(4,4,10)-1 purified from *S. pennellii* LA0716.

Table S4.9 NMR chemical shifts for G3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

Carbon # (group)	¹ H (ppm)		¹³ C (ppm) (from HSQC and HMBC)	
	α	β	α	β
1 (CH)	5.50 (d, <i>J</i> = 3.4 Hz)	4.75 (d, <i>J</i> = 8.1 Hz)	90.29	95.77
2 (CH) - 1 (CO) - 2 (CH) - 3,4 (CH ₃)	4.86 (dd, <i>J</i> = 10.0, 3.4 Hz) - 2.56 (hept, <i>J</i> = 7.0 Hz) 1.15 (m)	4.85 (m) - 2.56 (hept, <i>J</i> = 7.0 Hz) 1.15 (m)	71.15 176.83 33.89 18.79	73.48 176.83 33.89 18.79
3 (CH) - 1 (CO) - 2 (CH ₂) - 3 (CH ₂) - 4,5,6,7 (CH ₂) - 8 (CH ₂) - 9 (CH ₂) - 10 (CH ₃)	5.69 (t, <i>J</i> = 10.0 Hz) - 2.23 (t, <i>J</i> = 7.4 Hz) 1.53(m) 1.24(m) 1.24 (m) 1.28 (m) 0.88 (t, <i>J</i> = 7.0 Hz)	5.40 (t, <i>J</i> = 9.7 Hz) - 2.23 (t, <i>J</i> = 7.4 Hz) 1.53(m) 1.24(m) 1.24 (m) 1.28 (m) 0.88 (t, <i>J</i> = 7.0 Hz)	68.83 172.82 34.10 24.82 29.20 31.79 22.70 14.06	71.13 172.82 34.10 24.82 29.20 31.79 22.70 14.06
4 (CH) - 1 (CO) - 2 (CH) - 3,4 (CH ₃)	5.02 (m) - 2.56 (hept, <i>J</i> = 7.0 Hz) 1.15 (m)	5.02(m) - 2.56 (hept, <i>J</i> = 7.0 Hz) 1.15 (m)	68.54 176.83 33.89 18.79	68.54 176.83 33.89 18.79
5 (CH)	4.07 (ddd, <i>J</i> = 10.2, 4.0, 2.3 Hz)	3.56 (m)	69.57	74.53
6 (CH ₂)	3.57, 3.68 (m)	3.57, 3.68 (m)	61.06	61.06

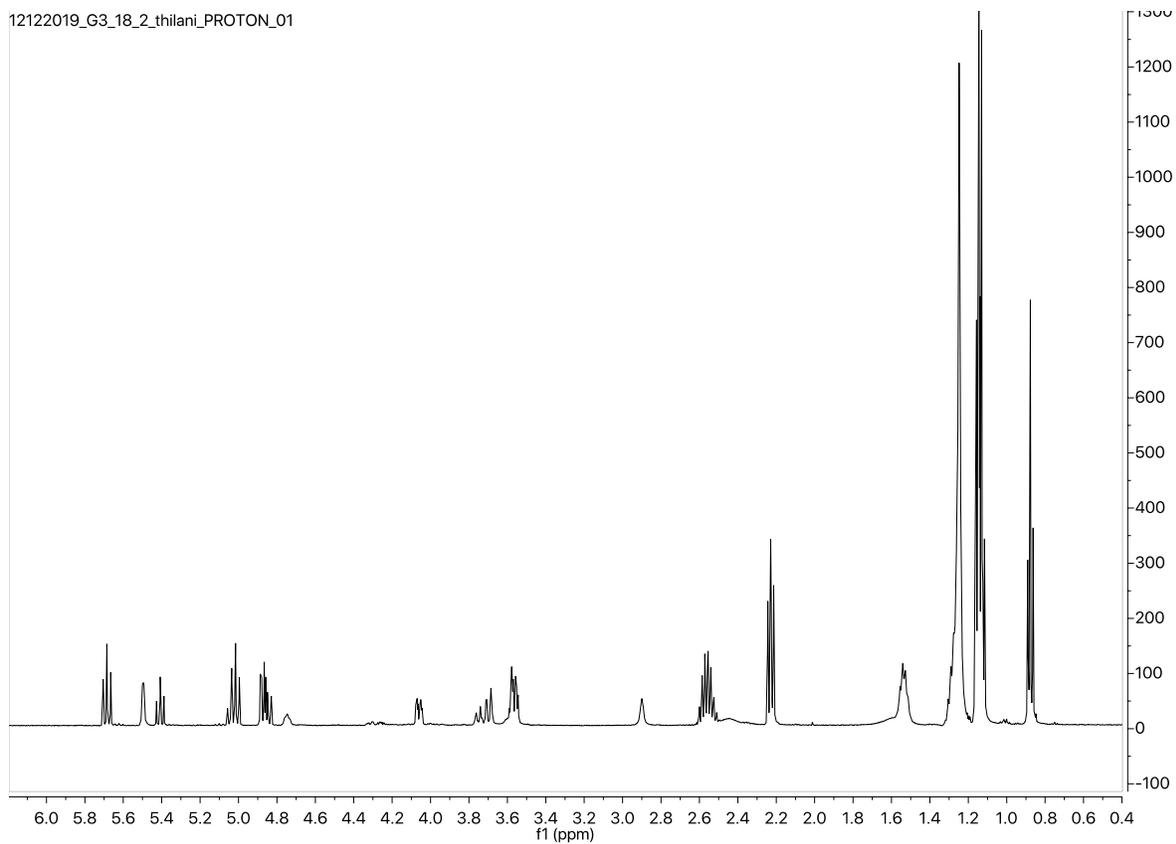


Figure S4.45 ^1H NMR spectrum for G3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

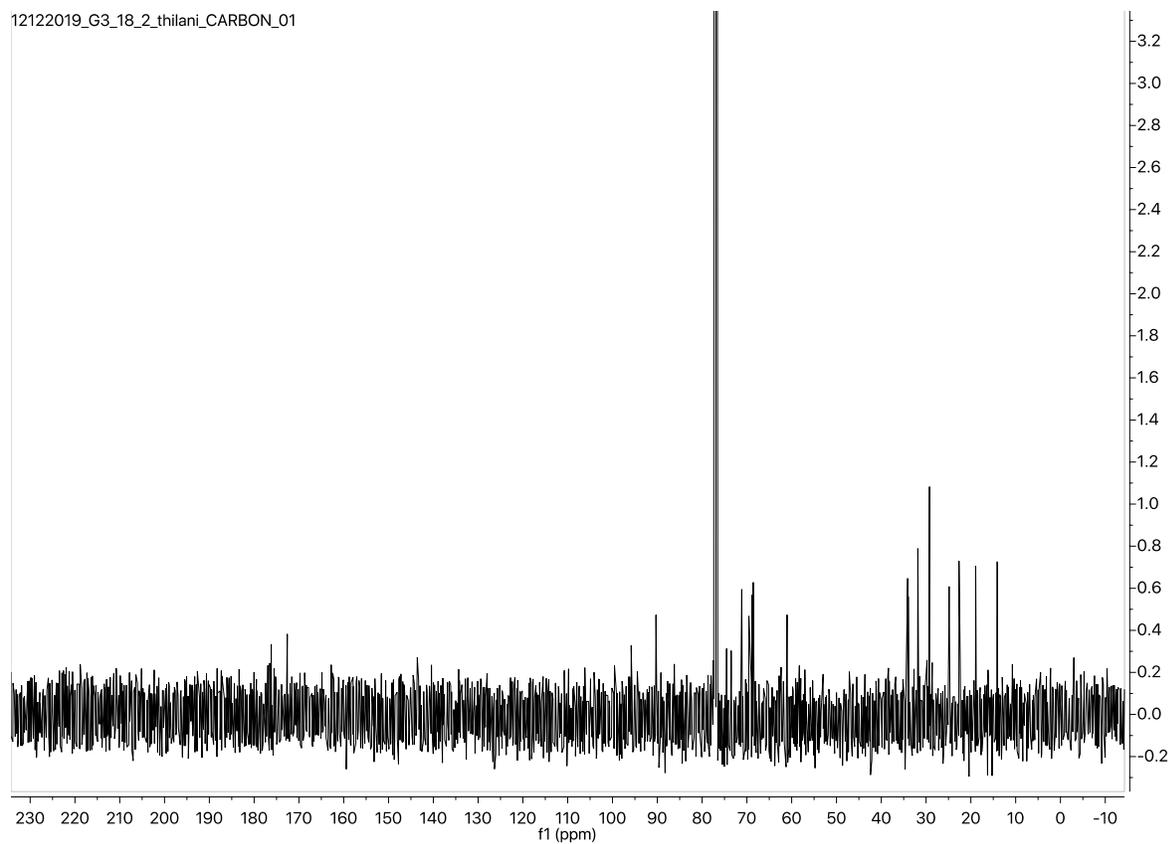


Figure S4.46 ^{13}C NMR spectrum for G3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

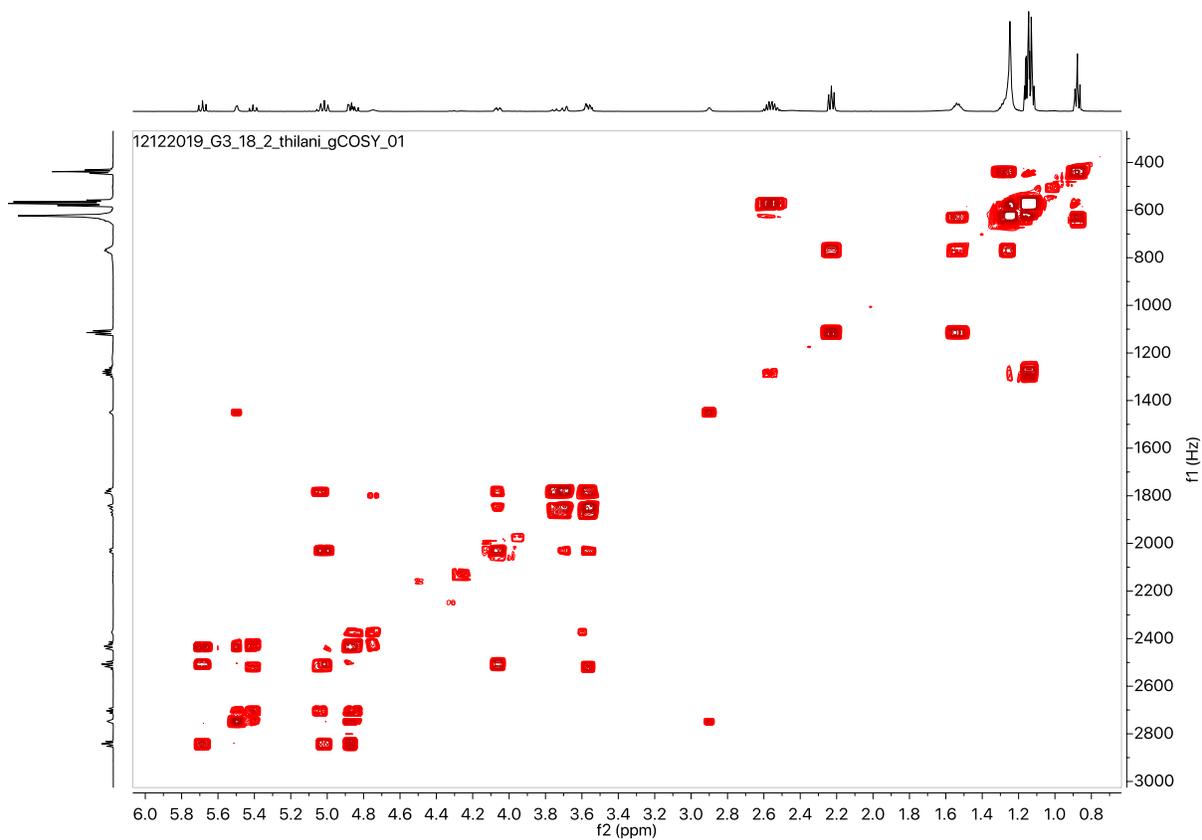


Figure S4.47 gCOSY NMR spectrum for G3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

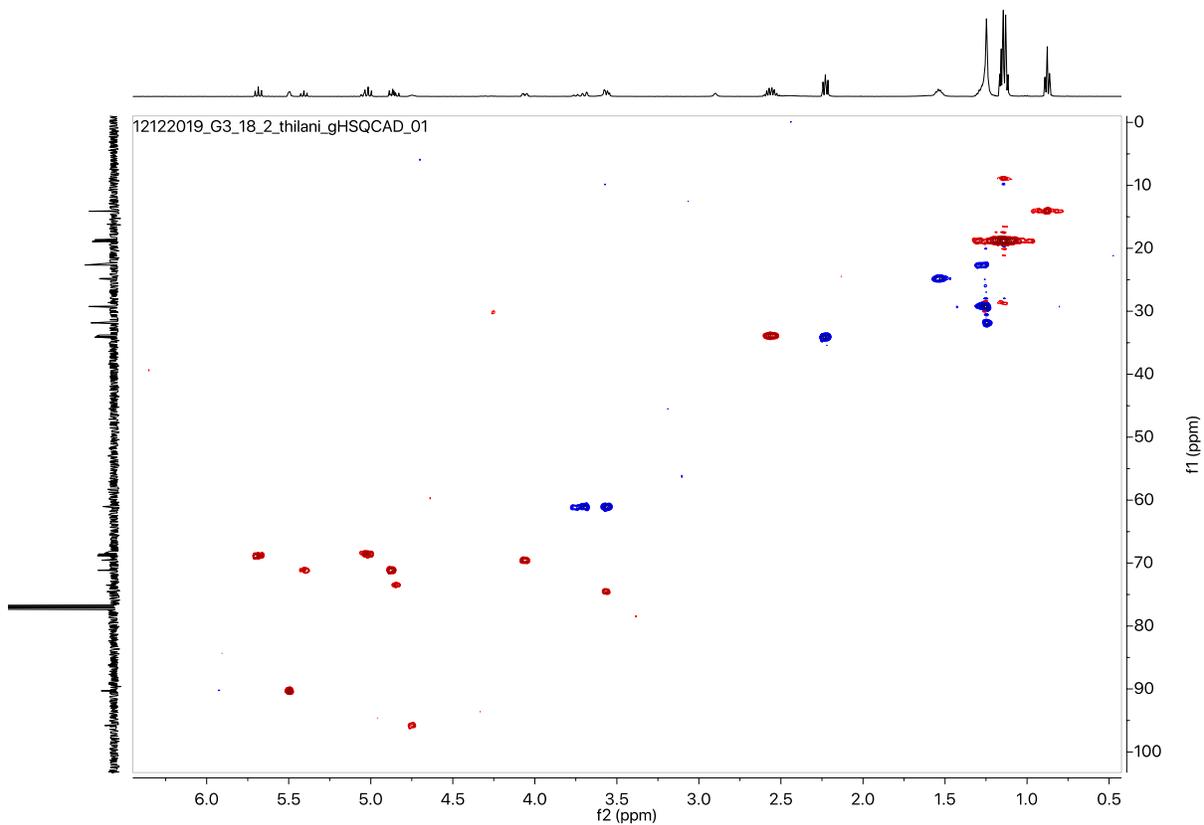


Figure S4.48 gHSQCAD NMR spectrum for G3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

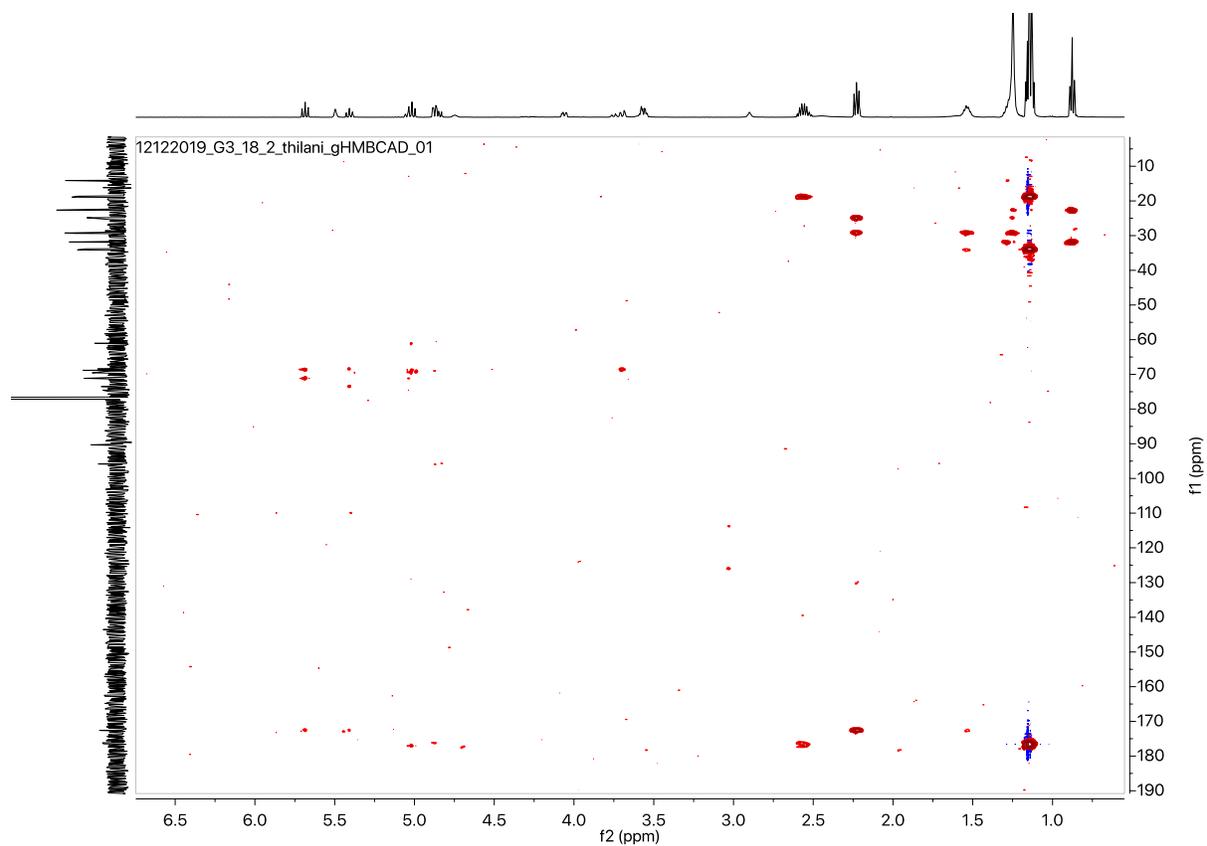


Figure S4.49 gHMBCAD NMR spectrum for G3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

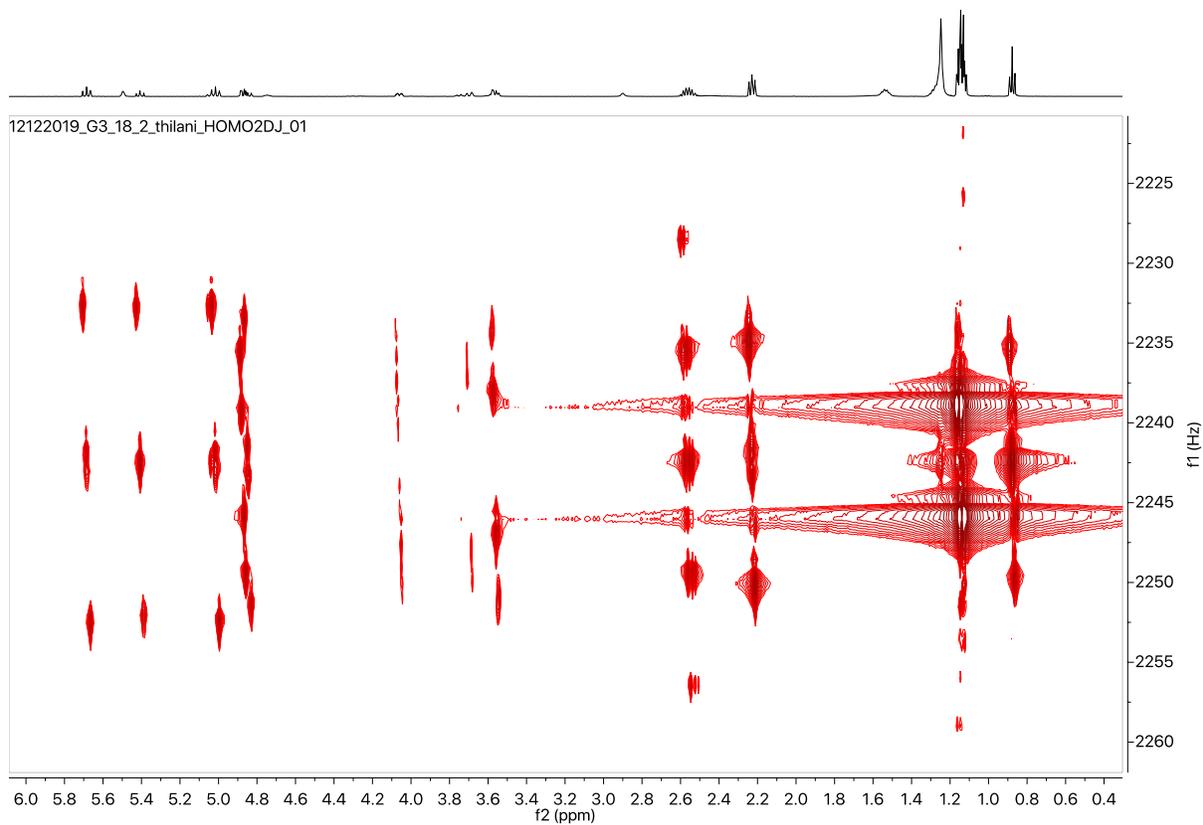


Figure S4.50 ^1H - ^1H HOMO2DJ NMR spectrum for G3:18(4,4,10)-2 purified from *S. pennellii* LA0716.

Table S4.10 NMR chemical shifts for G3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

Carbon # (group)	¹ H (ppm)		¹³ C (ppm) (from HSQC and HMBC)	
	α	β	α	β
1 (CH)	5.51 (d, <i>J</i> = 3.6 Hz)	4.74 (d, <i>J</i> = 8.1 Hz)	90.29	95.86
2 (CH)	4.88 (m)	4.87 (m)	71.17	73.35
- 1 (CO)	-	-	176.22	176.22
- 2 (CH)	2.41 (sextet, <i>J</i> = 6.9 Hz)	2.41 (sextet, <i>J</i> = 6.9 Hz)	40.92	40.92
- 3 (CH ₃)	1.13 (m, 3H)	1.13 (m, 3H)	16.35	16.35
- 4 (CH ₂)	1.45, 1.65 (m, 2H)	1.45, 1.65 (m, 2H)	26.49	26.49
- 5 (CH ₃)	0.88 (t, <i>J</i> = 7.3 Hz, 3H)	0.88 (t, <i>J</i> = 7.3 Hz, 3H)	11.50	11.50
3 (CH)	5.69 (t, <i>J</i> = 9.9 Hz)	5.41 (t, <i>J</i> = 9.6 Hz)	68.69	71.09
- 1 (CO)	-	-	172.71	172.71
- 2 (CH ₂)	2.23 (t, <i>J</i> = 7.4 Hz)	2.23 (t, <i>J</i> = 7.4 Hz)	34.05	34.05
- 3 (CH ₂)	1.54(m)	1.54(m)	24.76	24.76
- 4,5,6 (CH ₂)	1.25(m)	1.25(m)	29.27	29.27
- 7 (CH ₂)	1.24 (m)	1.24 (m)	27.22	27.22
- 8 (CH)	1.50 (m)	1.50 (m)	27.88	27.88
- 9,10 (CH ₃)	0.88 (d, <i>J</i> = 7.0 Hz, 6H)	0.88 (d, <i>J</i> = 7.0 Hz, 6H)	22.48	22.48
4 (CH)	5.04 (m)	5.04(m)	68.55	68.55
- 1 (CO)	-	-	176.73	176.73
- 2 (CH)	2.54 (hept, <i>J</i> = 7.0 Hz)	2.54 (hept, <i>J</i> = 7.0 Hz)	33.98	33.98
- 3,4 (CH ₃)	1.13 (m)	1.13 (m)	18.31	18.31
5 (CH)	4.07 (ddd, <i>J</i> = 10.2, 4.0, 2.2 Hz)	3.58 (m)	69.49	74.58
6 (CH ₂)	3.58, 3.69 (m)	3.58, 3.69 (m)	61.10	61.10

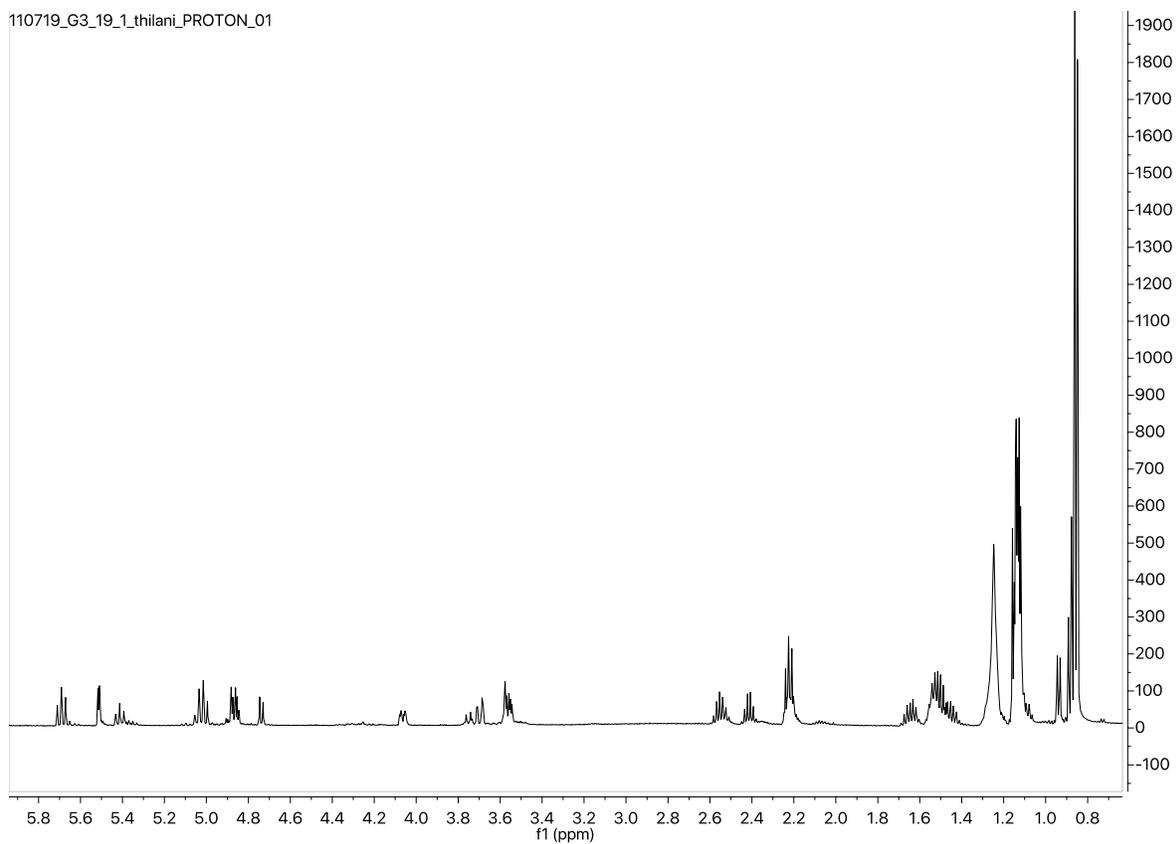


Figure S4.51 ^1H NMR spectrum for G3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

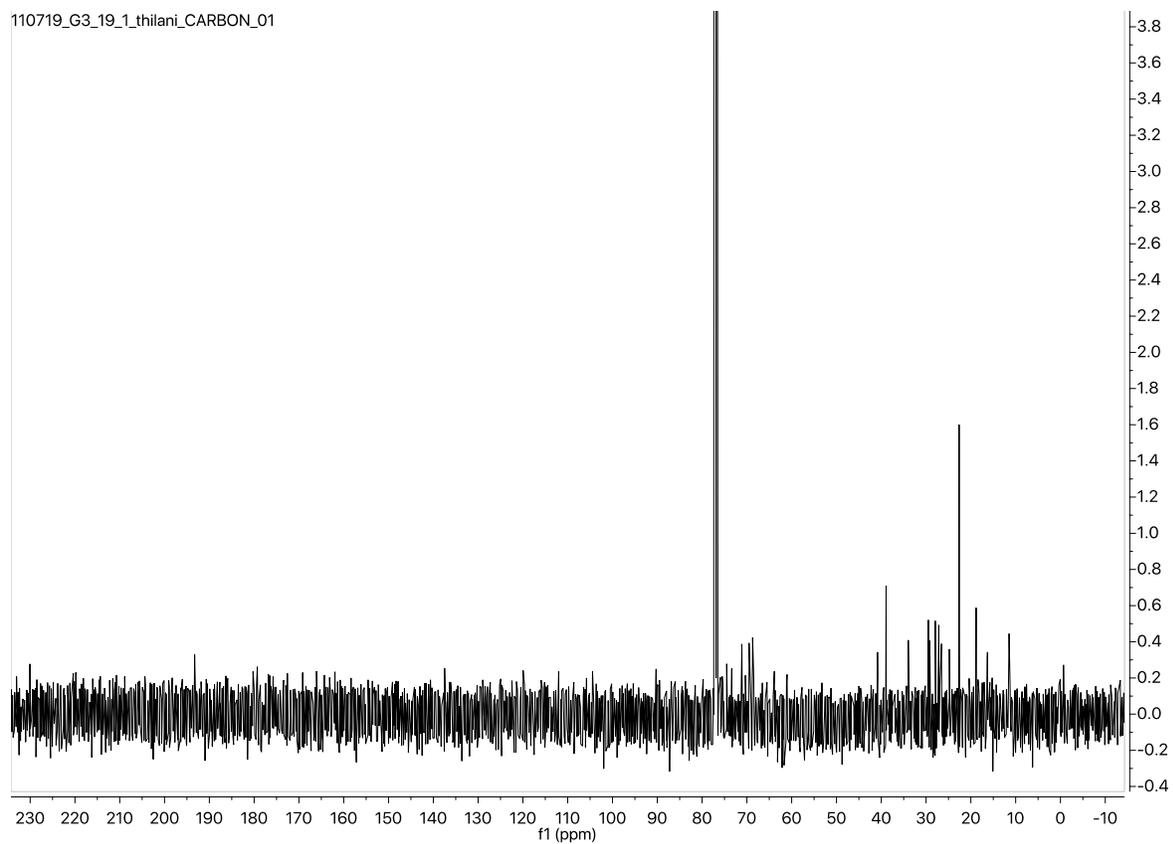


Figure S4.52 ^{13}C NMR spectrum for G3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

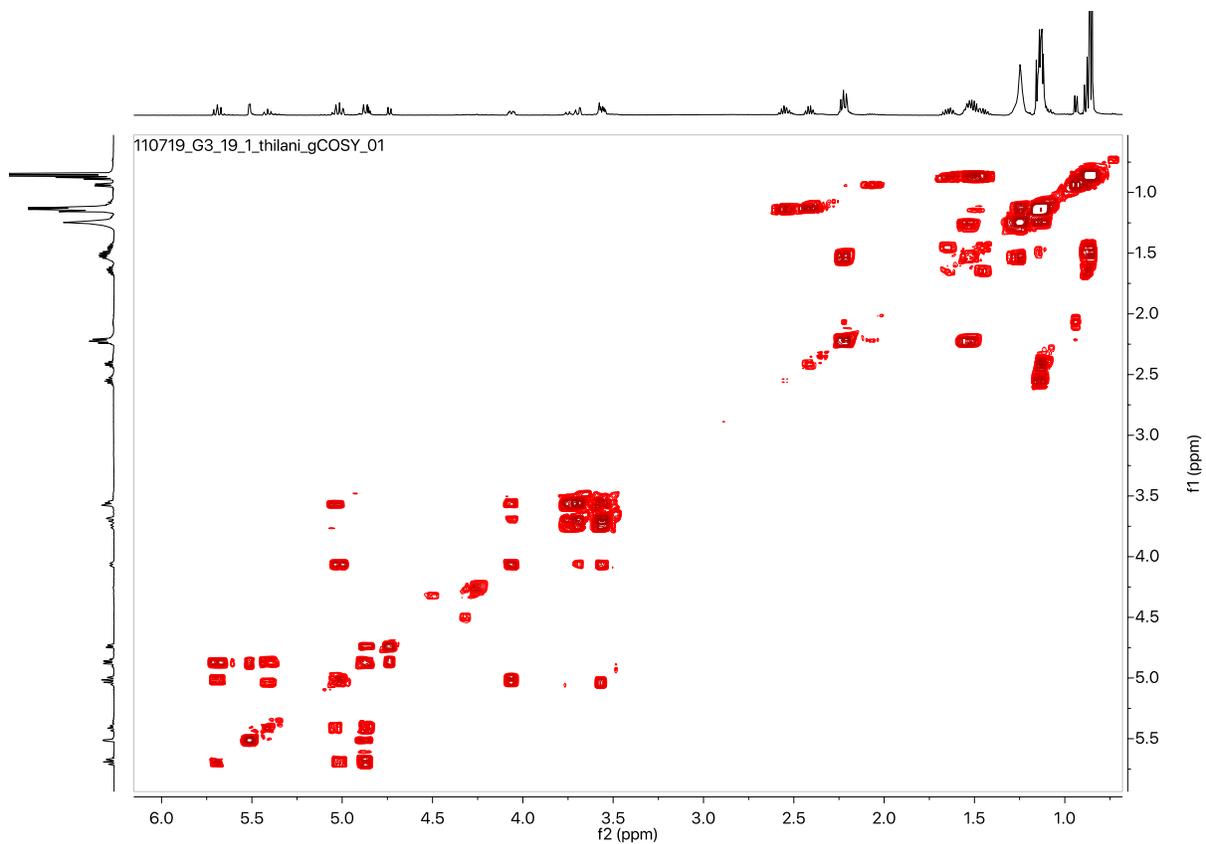


Figure S4.53 gCOSY NMR spectrum for G3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

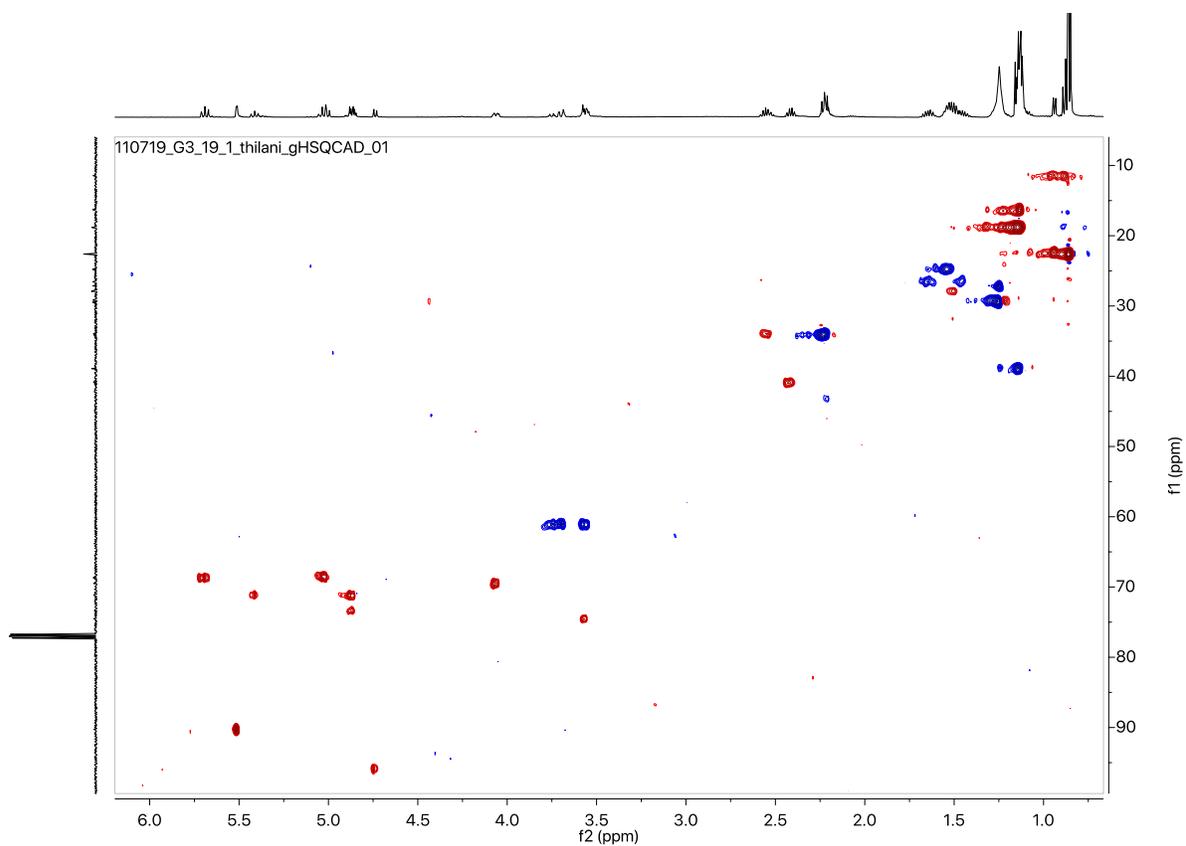


Figure S4.54 gHSQCAD NMR spectrum for G3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

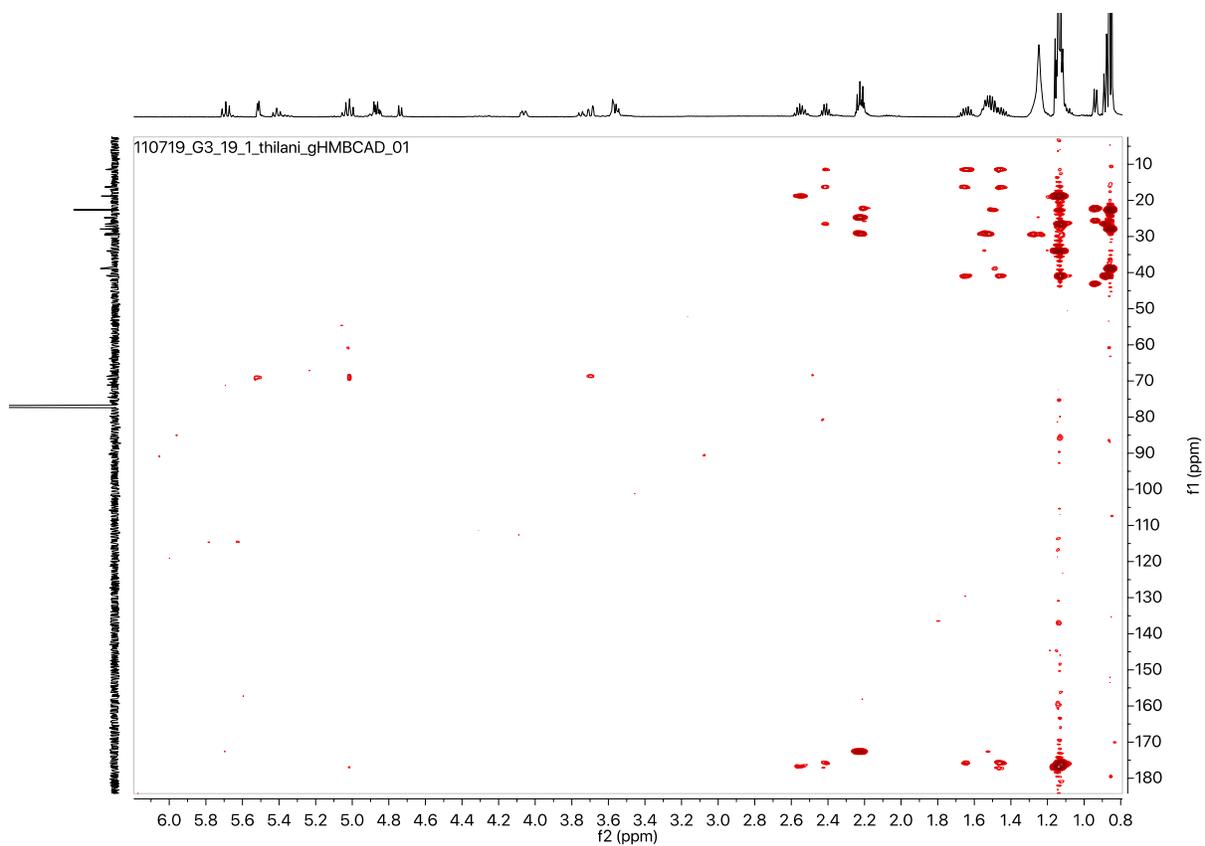


Figure S4.55 gHMBCAD NMR spectrum for G3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

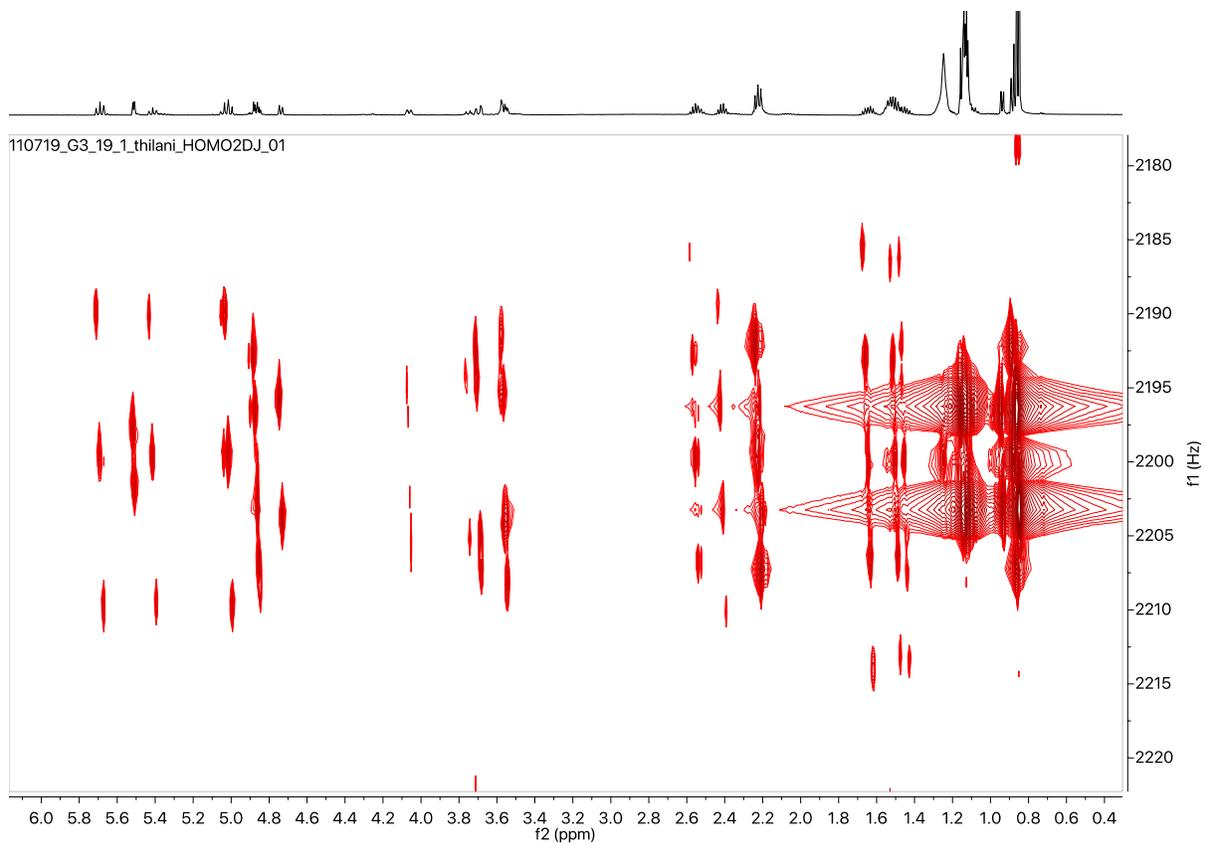
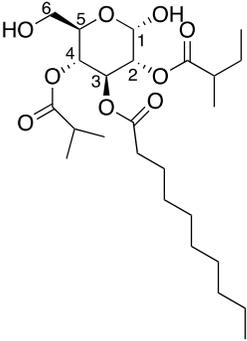


Figure S4.56 ^1H - ^1H HOMO2DJ NMR spectrum for G3:19(4,5,10)-1 purified from *S. pennellii* LA0716.

Table S4.11 NMR chemical shifts for G3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

Carbon # (group)	¹ H (ppm)		¹³ C (ppm) (from HSQC and HMBC)	
	α	β	α	β
 <p>G3:19(4,5,10)-2 Purified from <i>S. pennellii</i> LA0716 Chemical Formula: C₂₄H₄₂O₉ HRMS: (ESI) <i>m/z</i> calculated for C₂₄H₄₂O₉ ([M+NH₄)⁺): 506.3324 Experimental <i>m/z</i>: 506.3328 NMR (500 MHz, CDCl₃) Sample mass: 2 mg</p>				
1 (CH)	5.51 (d, <i>J</i> = 3.1 Hz)	4.74 (d, <i>J</i> = 7.6 Hz)	90.25	95.82
2 (CH)	4.87 (m)	4.86 (m)	71.13	73.31
- 1 (CO)	-	-	176.22	176.22
- 2 (CH)	2.41 (sextet, <i>J</i> = 7.0 Hz)	2.41 (sextet, <i>J</i> = 7.0 Hz)	40.86	40.86
- 3 (CH ₃)	1.13 (m, 3H)	1.13 (m, 3H)	16.35	16.35
- 4 (CH ₂)	1.45, 1.65 (m, 2H)	1.45, 1.65 (m, 2H)	26.53	26.53
- 5 (CH ₃)	0.89 (t, <i>J</i> = 7.3 Hz, 3H)	0.89 (t, <i>J</i> = 7.3 Hz, 3H)	11.76	11.76
3 (CH)	5.68 (t, <i>J</i> = 9.6 Hz)	5.41 (t, <i>J</i> = 9.7 Hz)	68.66	71.01
- 1 (CO)	-	-	172.62	172.62
- 2 (CH ₂)	2.13 (t, <i>J</i> = 7.4 Hz)	2.21 (t, <i>J</i> = 7.4 Hz)	34.08	34.08
- 3 (CH ₂)	1.53(m)	1.53(m)	25.13	25.13
- 4,5,6,7 (CH ₂)	1.25(m)	1.25(m)	29.33	29.33
- 8 (CH ₂)	1.24 (m)	1.24 (m)	32.21	32.21
- 9 (CH ₂)	1.25 (m)	1.25 (m)	22.56	22.56
- 10 (CH ₃)	0.88 (t, <i>J</i> = 7.0 Hz)	0.88 (t, <i>J</i> = 7.0 Hz)	14.07	14.07
4 (CH)	5.02 (t, <i>J</i> = 9.7 Hz)	5.02 (t, <i>J</i> = 9.7 Hz)	68.53	68.53
- 1 (CO)	-	-	176.83	176.83
- 2 (CH)	2.54 (hept, <i>J</i> = 7.0 Hz)	2.54 (hept, <i>J</i> = 7.0 Hz)	33.60	33.60
- 3,4 (CH ₃)	1.14 (m)	1.14 (m)	18.82	18.82
5 (CH)	4.04 (ddd, <i>J</i> = 10.2, 4.0, 2.2 Hz)	3.56 (m)	69.49	74.55
6 (CH ₂)	3.57, 3.68 (m)	3.57, 3.68 (m)	60.95	60.95

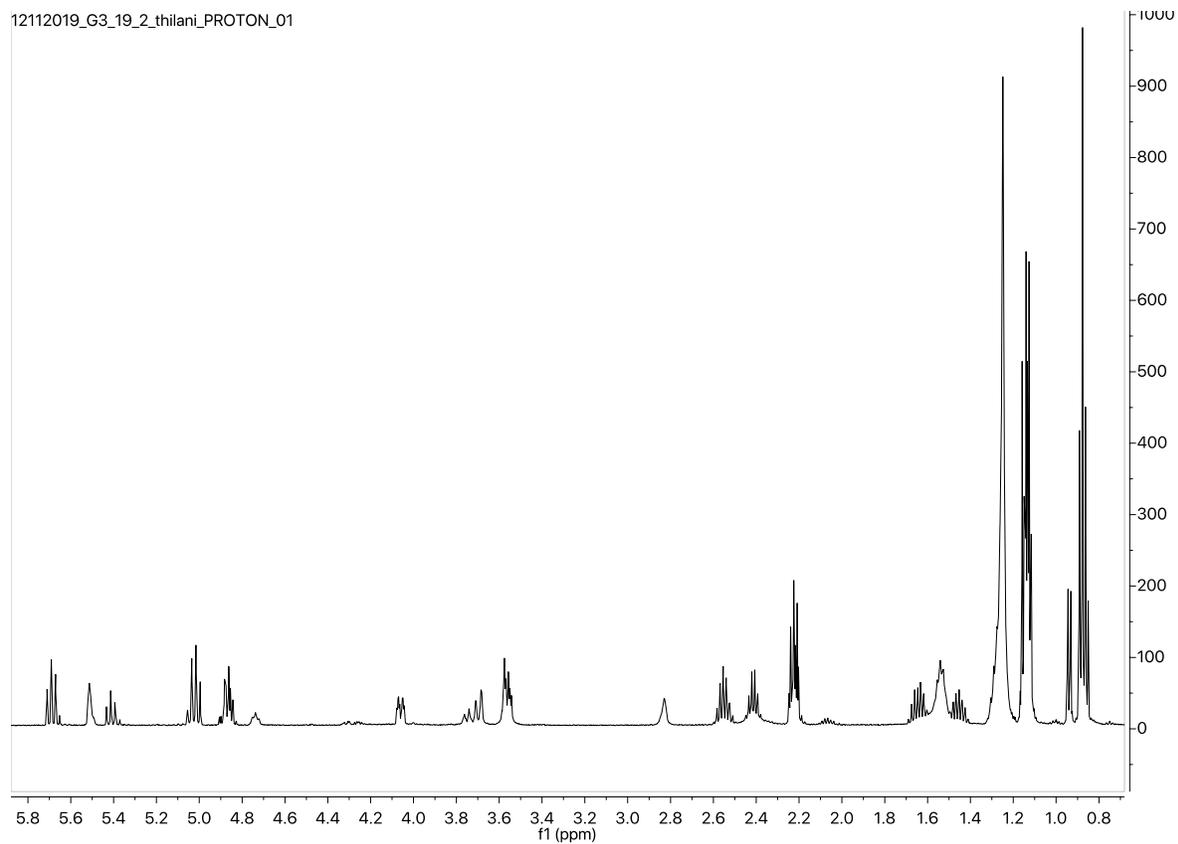


Figure S4.57 ^1H NMR spectrum for G3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

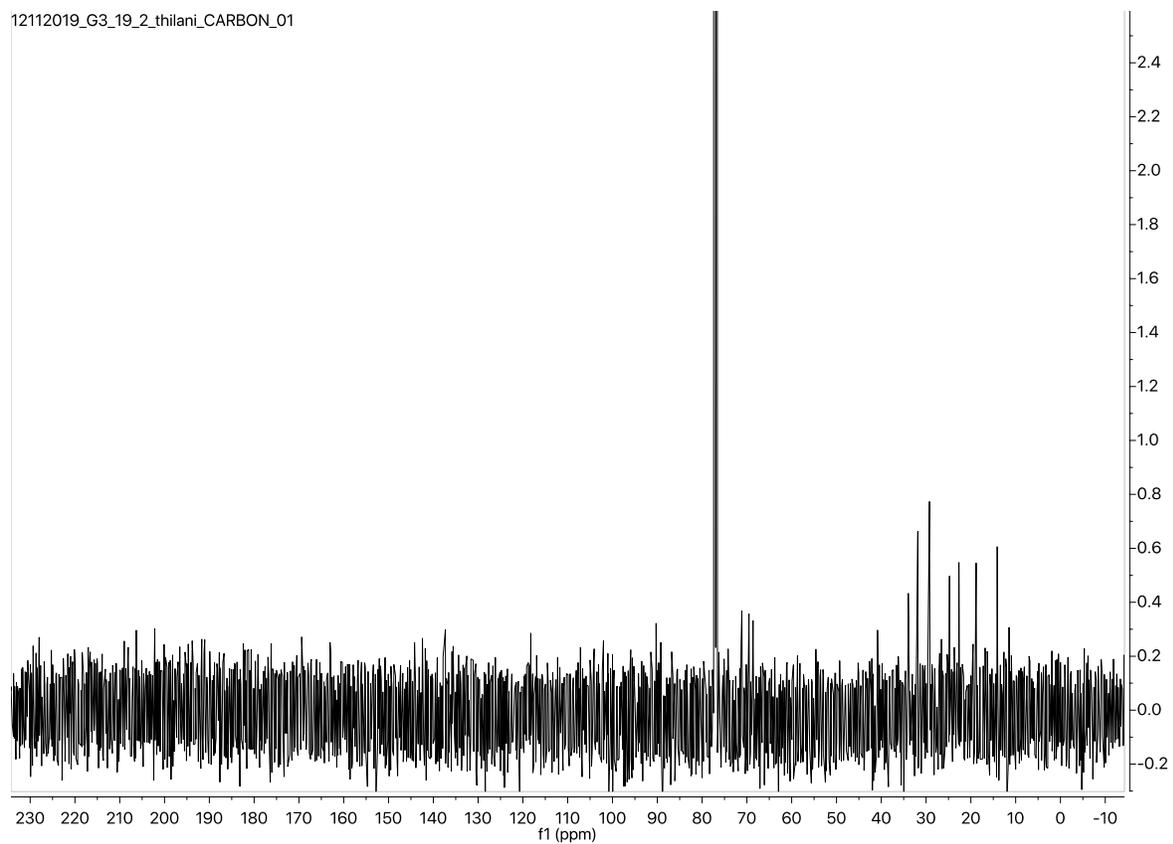


Figure S4.58 ^{13}C NMR spectrum for G3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

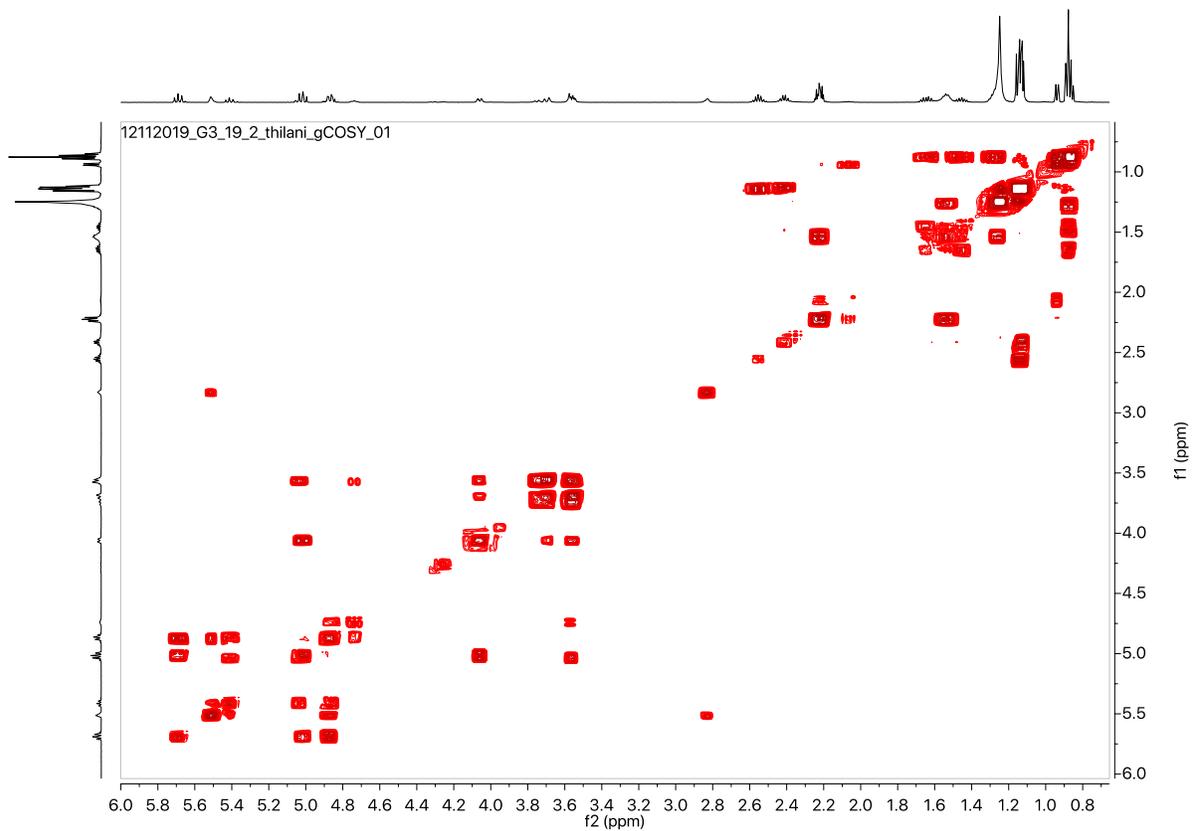


Figure S4.59 gCOSY NMR spectrum for G3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

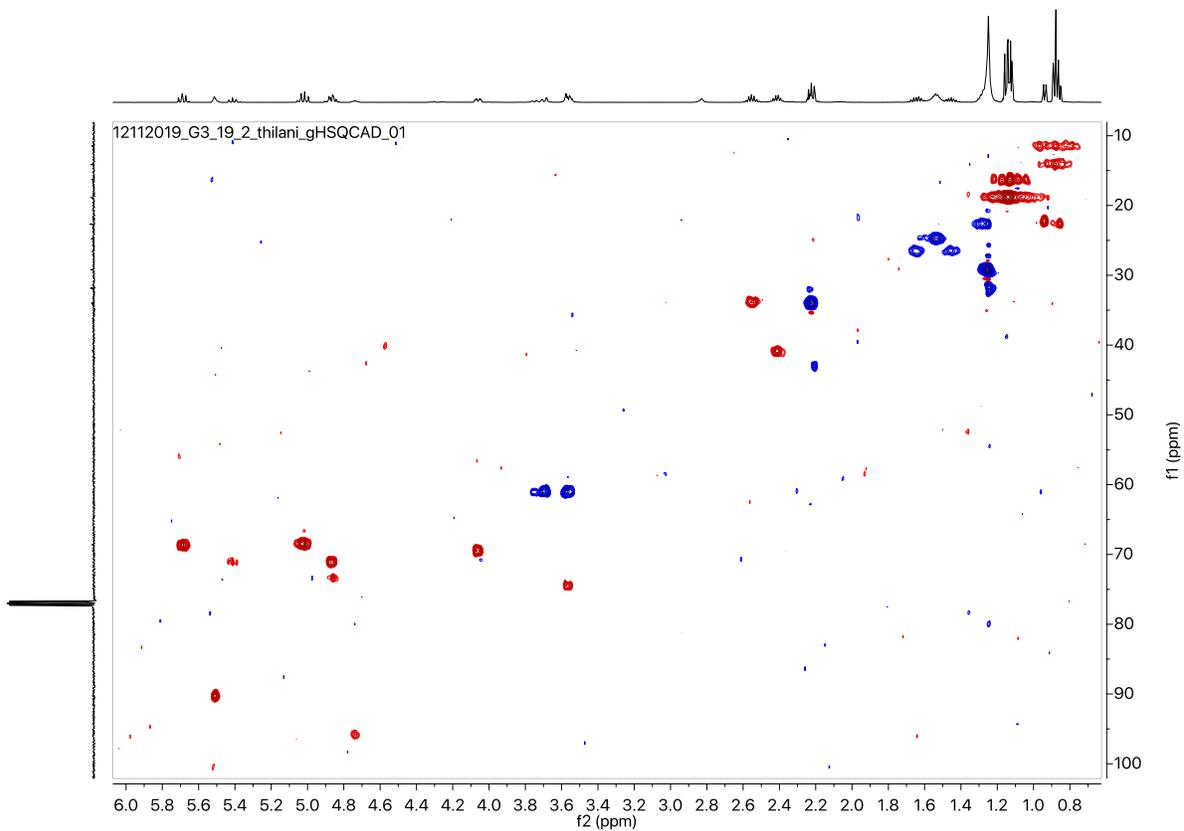


Figure S4.60 gHSQCAD NMR spectrum for G3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

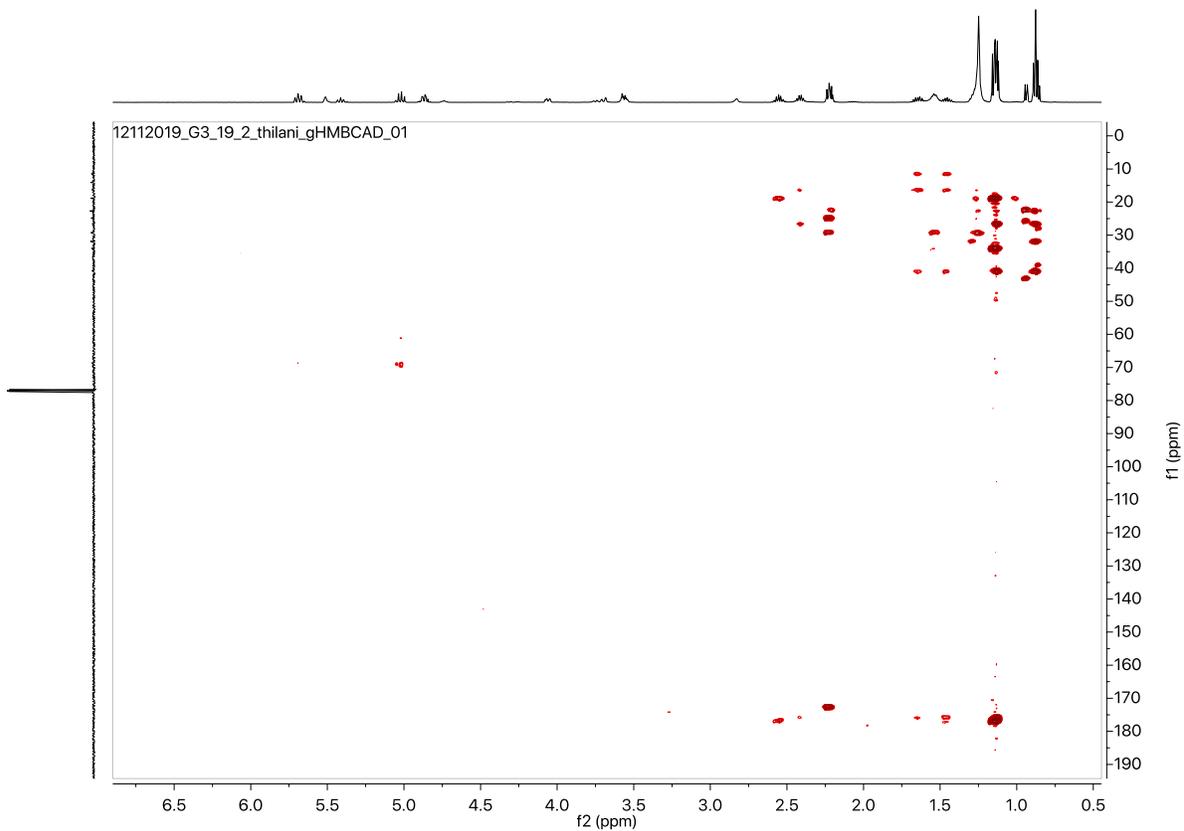


Figure S4.61 gHMBCAD NMR spectrum for G3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

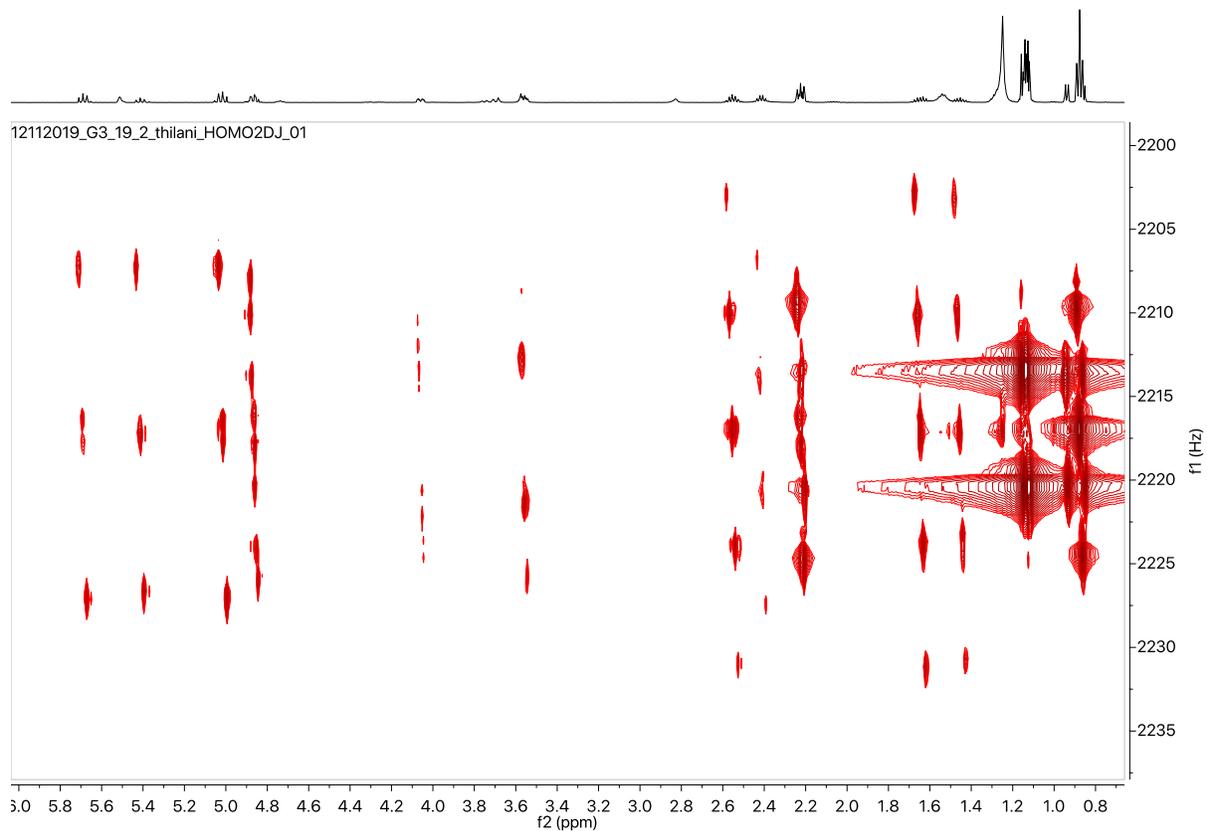


Figure S4.62 ^1H - ^1H HOMO2DJ NMR spectrum for G3:19(4,5,10)-2 purified from *S. pennellii* LA0716.

Table S4.12 Loadings and correlation values for 54 metabolite features from the North range/South range OPLS-DA model.

Compound	North-South load	North-South corr
G3:15(5,5,5)b	-4.20E-01	-8.09E-01
G3:21(5,5,11)a	-7.67E-02	-7.55E-01
G3:15(5,5,5)a	-3.15E-01	-7.51E-01
G3:21(5,5,11)b	-1.15E-01	-7.46E-01
S3:21(5,5,11)	-1.55E-01	-7.33E-01
G3:16(5,5,6)a	-1.04E-01	-6.89E-01
S3:16(5,5,6)	-7.95E-03	-6.09E-01
S3:20(5,5,10)	-1.89E-01	-6.05E-01
G3:23(5,6,12)a	-2.87E-02	-6.04E-01
S3:15(5,5,5)	-1.51E-02	-6.01E-01
G3:23(5,6,12)b	-3.79E-02	-5.90E-01
G3:16(5,5,6)b	-1.24E-01	-5.85E-01
G3:22(5,5,12)a	-7.71E-02	-5.80E-01
G3:22(5,5,12)b	-1.12E-01	-5.53E-01
S3:23(5,6,12)	-7.65E-02	-5.53E-01
S3:22(5,5,12)	-2.16E-01	-5.51E-01
G3:20(5,5,10)a	-1.52E-01	-5.11E-01
G3:20(5,5,10)b	-2.02E-01	-4.73E-01
G3:14(4,5,5)	-4.29E-02	-1.34E-01
flavonoid C	-2.53E-03	-7.77E-02
flavonoid A	2.27E-03	4.50E-02
S3:21(4,5,12)	1.23E-02	8.82E-02
G3:19(4,5,10)-1a	2.44E-02	9.94E-02
S3:14(4,5,5)	7.20E-03	2.26E-01
S3:12(4,4,4)	9.82E-03	2.45E-01
S3:13(4,4,5)	1.81E-02	2.59E-01
G4:14(2,4,4,4)b	1.77E-02	2.70E-01
G4:15(2,4,4,5)	1.32E-02	2.95E-01
G3:17(4,5,8)-1a	4.68E-02	3.01E-01
S3:19(4,5,10)-1	8.45E-02	3.13E-01
G3:18(4,4,10)-1a	1.15E-01	3.22E-01
S3:20(4,4,12)	4.13E-02	3.23E-01
G4:14(2,4,4,4)a	1.87E-02	3.34E-01
G3:21(4,5,12)b	4.16E-02	3.42E-01
S3:18:(4,4,10)-1	1.06E-01	3.50E-01
G3:21(4,5,12)a	2.99E-02	3.59E-01
S3:17(4,5,8)	2.33E-02	3.62E-01

Table S4.12 (cont'd)

Compound	North-South load	North-South corr
G3:16(4,4,8)-1a	8.70E-02	3.65E-01
flavonoid B	3.41E-02	3.74E-01
S3:16(4,4,8)	2.02E-02	3.77E-01
flavonoid D	1.70E-02	3.78E-01
S3:19(4,5,10)-2	4.70E-02	3.90E-01
G3:17(4,5,8)-2b	5.52E-02	4.03E-01
G3:17(4,5,8)-1b/2a	1.18E-01	4.07E-01
S3:17(4,4,9)	1.44E-02	4.13E-01
G3:16(4,4,8)-1b/2a	1.38E-01	4.16E-01
G3:16(4,4,8)-2b	5.74E-02	4.25E-01
G3:20(4,4,12)	6.97E-02	4.83E-01
S3:18(4,4,10)-2	3.86E-02	5.04E-01
G3:18(4,4,10)-2b	1.47E-01	5.52E-01
G3:19(4,5,10)-1b/2a	2.43E-01	5.70E-01
G3:18(4,4,10)-1b/2a	2.66E-01	6.63E-01
G3:13(4,4,5)	3.87E-01	8.66E-01
G3:12(4,4,4)	2.81E-01	8.75E-01

Table S4.13 Loadings and correlation values for 54 metabolite features from the Pisco region/Atico region OPLS-DA model.

Compound	Pisco-Atico load	Pisco-Atico corr
G3:16(4,4,8)-1b/2a	-4.00E-01	-7.11E-01
G3:17(4,5,8)-2b	-1.61E-01	-6.99E-01
G3:17(4,5,8)-1b/2a	-3.41E-01	-6.95E-01
G3:16(4,4,8)-2b	-1.56E-01	-6.86E-01
G3:17(4,5,8)-1a	-1.78E-01	-6.63E-01
G3:16(4,4,8)-1a	-2.59E-01	-6.35E-01
S3:17(4,4,9)	-3.41E-02	-5.81E-01
S3:17(4,5,8)	-6.27E-02	-5.70E-01
G3:12(4,4,4)	-2.43E-01	-5.68E-01
S3:16(4,4,8)	-5.15E-02	-5.62E-01
G3:13(4,4,5)	-3.16E-01	-5.25E-01
G3:16(5,5,6)b	-9.77E-02	-4.50E-01
G4:15(2,4,4,5)	-2.85E-02	-3.75E-01
G4:14(2,4,4,4)a	-3.73E-02	-3.50E-01
G4:14(2,4,4,4)b	-3.76E-02	-3.44E-01
G3:14(4,5,5)	-3.52E-02	-1.45E-01
S3:18(4,4,10)-2	-1.43E-02	-1.12E-01
S3:16(5,5,6)	-7.72E-04	-4.09E-02
G3:16(5,5,6)a	2.47E-02	1.87E-03
G3:15(5,5,5)b	8.78E-02	4.38E-03
S3:19(4,5,10)-2	7.59E-03	4.34E-02
flavonoid C	3.39E-03	5.54E-02
flavonoid D	7.23E-03	7.15E-02
G3:15(5,5,5)a	8.36E-02	1.01E-01
S3:23(5,6,12)	2.67E-02	1.14E-01
flavonoid B	3.11E-02	1.67E-01
S3:22(5,5,12)	8.00E-02	1.68E-01
S3:21(5,5,11)	3.64E-02	1.77E-01
G3:18(4,4,10)-2b	9.00E-02	1.98E-01
G3:23(5,6,12)b	1.40E-02	2.15E-01
G3:23(5,6,12)a	1.06E-02	2.27E-01
G3:18(4,4,10)-1b/2a	1.65E-01	2.61E-01
G3:19(4,5,10)-1b/2a	2.19E-01	3.11E-01
G3:18(4,4,10)-1a	1.81E-01	3.15E-01

Table S4.13 (cont'd)

Compound	Pisco-Atico load	Pisco-Atico corr
G3:19(4,5,10)-1a	1.30E-01	3.27E-01
G3:20(5,5,10)b	8.16E-02	3.32E-01
S3:20(5,5,10)	6.48E-02	3.36E-01
G3:22(5,5,12)b	6.05E-02	3.45E-01
G3:22(5,5,12)a	3.94E-02	3.46E-01
S3:12(4,4,4)	2.64E-02	3.48E-01
S3:15(5,5,5)	6.74E-03	3.54E-01
S3:13(4,4,5)	4.79E-02	3.64E-01
S3:19(4,5,10)-1	1.70E-01	3.71E-01
G3:21(5,5,11)b	3.96E-02	3.73E-01
G3:21(5,5,11)a	3.01E-02	3.77E-01
S3:18:(4,4,10)-1	2.00E-01	3.88E-01
S3:14(4,5,5)	2.16E-02	3.91E-01
G3:20(5,5,10)a	8.39E-02	3.93E-01
S3:20(4,4,12)	8.91E-02	4.04E-01
S3:21(4,5,12)	1.01E-01	4.22E-01
G3:20(4,4,12)	1.13E-01	4.38E-01
flavonoid A	3.69E-02	4.48E-01
G3:21(4,5,12)b	1.01E-01	4.72E-01
G3:21(4,5,12)a	7.07E-02	4.78E-01

Table S4.14 Loadings and correlation values for 54 metabolite features from the intraregion Atico OPLS-DA model.

Compound	Atico-LA2963 load	Atico-LA2963 corr
G3:19(4,5,10)-1b/2a	-3.78E-01	-8.04E-01
G3:18(4,4,10)-2b	-2.17E-01	-7.63E-01
G3:20(4,4,12)	-1.30E-01	-7.49E-01
G3:18(4,4,10)-1b/2a	-3.38E-01	-7.39E-01
flavonoid A	-4.11E-02	-6.69E-01
flavonoid C	-2.22E-02	-6.52E-01
G3:19(4,5,10)-1a	-1.91E-01	-6.52E-01
G3:21(4,5,12)a	-6.74E-02	-6.36E-01
G3:21(4,5,12)b	-9.85E-02	-6.35E-01
G3:17(4,5,8)-2b	-2.37E-02	-5.89E-01
G3:18(4,4,10)-1a	-2.69E-01	-5.86E-01
G3:17(4,5,8)-1b/2a	-1.52E-02	-5.65E-01
flavonoid D	-2.98E-02	-5.21E-01
G3:20(5,5,10)a	-8.81E-02	-4.99E-01
flavonoid B	-5.79E-02	-4.76E-01
G3:16(4,4,8)-2b	-7.17E-03	-4.70E-01
G3:20(5,5,10)b	-8.84E-02	-4.14E-01
G3:16(4,4,8)-1b/2a	-7.23E-03	-3.92E-01
G3:21(5,5,11)a	-2.42E-02	-3.55E-01
G3:21(5,5,11)b	-3.11E-02	-3.48E-01
G3:22(5,5,12)a	-2.61E-02	-3.34E-01
G3:22(5,5,12)b	-3.84E-02	-3.21E-01
G3:16(4,4,8)-1a	-4.87E-03	-2.84E-01
G4:14(2,4,4,4)a	-1.83E-03	-2.78E-01
G3:23(5,6,12)a	-7.70E-03	-2.73E-01
G3:23(5,6,12)b	-1.01E-02	-2.66E-01
G4:14(2,4,4,4)b	-4.50E-03	-2.27E-01
S3:16(5,5,6)	-1.90E-03	-2.22E-01
S3:23(5,6,12)	-1.79E-02	-2.20E-01
G3:16(5,5,6)a	-1.96E-02	-2.15E-01
G3:15(5,5,5)b	-7.07E-02	-2.06E-01
G3:15(5,5,5)a	-5.50E-02	-1.88E-01
G3:14(4,5,5)	-2.56E-02	-1.43E-01
G4:15(2,4,4,5)	-5.00E-04	-9.65E-02

Table S4.14 (cont'd)

Compound	Atico-LA2963 load	Atico-LA2963 corr
G3:16(5,5,6)b	-2.49E-02	7.43E-02
S3:21(5,5,11)	-2.10E-02	1.84E-01
S3:22(5,5,12)	-3.17E-02	1.99E-01
G3:17(4,5,8)-1a	3.77E-03	2.85E-01
G3:12(4,4,4)	1.25E-01	5.20E-01
G3:13(4,4,5)	1.99E-01	5.22E-01
S3:20(5,5,10)	4.53E-02	5.57E-01
S3:15(5,5,5)	9.00E-03	6.36E-01
S3:12(4,4,4)	4.52E-02	7.82E-01
S3:13(4,4,5)	8.35E-02	8.27E-01
S3:16(4,4,8)	1.42E-02	8.70E-01
S3:18(4,4,10)-2	7.11E-02	8.78E-01
S3:14(4,5,5)	3.83E-02	8.82E-01
S3:17(4,5,8)	1.61E-02	9.10E-01
S3:17(4,4,9)	2.07E-02	9.36E-01
S3:20(4,4,12)	1.73E-01	9.41E-01
S3:21(4,5,12)	1.91E-01	9.55E-01
S3:19(4,5,10)-2	1.35E-01	9.59E-01
S3:19(4,5,10)-1	3.85E-01	9.92E-01
S3:18:(4,4,10)-1	4.30E-01	9.97E-01

REFERENCES

REFERENCES

- 1 Fobes, J.F. *et al.* (1985) Epicuticular Lipid Accumulation on the Leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. *Plant Physiol.* 77, 567–570
- 2 A. Burke, B. *et al.* (1987) Polar epicuticular lipids of *Lycopersicon pennellii*. *Phytochemistry* 26, 2567–2571
- 3 Shapiro, J.A. *et al.* (1994) Acylsugars of the wild tomato *Lycopersicon pennellii* in relation to geographic distribution of the species. *Biochem. Syst. Ecol.* 22, 545–561
- 4 Slocombe, S.P. *et al.* (2008) Transcriptomic and Reverse Genetic Analyses of Branched-Chain Fatty Acid and Acyl Sugar Production in *Solanum pennellii* and *Nicotiana benthamiana*. *Plant Physiol.* 148, 1830–1846
- 5 Ning, J. *et al.* (2015) A feedback insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. *Plant Physiol.* DOI: 10.1104/pp.15.00474
- 6 Schillmiller, A.L. *et al.* (2016) Acylsugar Acylhydrolases: Carboxylesterase-Catalyzed Hydrolysis of Acylsugars in Tomato Trichomes. *Plant Physiol.* 170, 1331–1344
- 7 Fan, P. *et al.* (2017) Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. *Nat. Commun.* 8, 2080
- 8 Leong, B.J. *et al.* (2019) Evolution of metabolic novelty: A trichome-expressed invertase creates specialized metabolic diversity in wild tomato. *Sci. Adv.* 5, eaaw3754
- 9 Correll, D.S. (1962) *The potato and its wild relatives.*, Texas Research Foundation.
- 10 Ghangas, G.S. and Steffens, J.C. (1993) UDPglucose: fatty acid transglucosylation and transacylation in triacylglycerol biosynthesis. *Proc. Natl. Acad. Sci.* 90, 9911–9915
- 11 Kuai, J.P. *et al.* (1997) Regulation of Triacylglycerol Fatty Acid Composition (Uridine Diphosphate Glucose:Fatty Acid Glucosyltransferases with Overlapping Chain-Length Specificity). *Plant Physiol.* 115, 1581–1587
- 12 Li, A.X. *et al.* (1999) Glucose Polyester Biosynthesis. Purification and Characterization of a Glucose Acyltransferase. *Plant Physiol.* 121, 453–460
- 13 Li, A.X. and Steffens, J.C. (2000) An acyltransferase catalyzing the formation of diacylglycerol is a serine carboxypeptidase-like protein. *Proc. Natl. Acad. Sci.* 97, 6902–6907
- 14 Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, 45e–445

- 15 Kim, J. *et al.* (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, 1854–1870
- 16 Mandal, S. *et al.* (2020) Candidate Gene Networks for Acylsugar Metabolism and Plant Defense in Wild Tomato *Solanum pennellii*. *Plant Cell* 32, 81–99
- 17 Schmidt, A. *et al.* (2011) Polymethylated Myricetin in Trichomes of the Wild Tomato Species *Solanum habrochaites* and Characterization of Trichome-Specific 3'/5'- and 7/4'-Myricetin O-Methyltransferases. *Plant Physiol.* 155, 1999–2009
- 18 Leckie, B.M. *et al.* (2014) Quantitative trait loci regulating the fatty acid profile of acylsugars in tomato. *Mol. Breeding* 34, 1201–1213
- 19 Schillmiller, A.L. *et al.* (2015) Functionally Divergent Alleles and Duplicated Loci Encoding an Acyltransferase Contribute to Acylsugar Metabolite Diversity in *Solanum* Trichomes. *Plant Cell* 27, 1002–1017
- 20 Schmidt, A. *et al.* (2012) Characterization of a flavonol 3-O-methyltransferase in the trichomes of the wild tomato species *Solanum habrochaites*. *Planta* 236, 839–849
- 21 Kim, J. *et al.* (2014) Analysis of Natural and Induced Variation in Tomato Glandular Trichome Flavonoids Identifies a Gene Not Present in the Reference Genome. *The Plant Cell* 26, 3272–3285
- 22 Kang, J.-H. *et al.* (2010) The Tomato *odorless-2* Mutant Is Defective in Trichome-Based Production of Diverse Specialized Metabolites and Broad-Spectrum Resistance to Insect Herbivores. *Plant Physiol.* 154, 262–272
- 23 Li, C. *et al.* (2013) Identification of methylated flavonoid regioisomeric metabolites using enzymatic semisynthesis and liquid chromatography-tandem mass spectrometry. *Metabolomics* 9, 92–101
- 24 Ghosh, B. *et al.* (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
- 25 Zhang, C.-R. *et al.* (2016) New antiinflammatory sucrose esters in the natural sticky coating of tomatillo (*Physalis philadelphica*), an important culinary fruit. *Food Chemistry* 196, 726–732
- 26 Bernal, C.-A. *et al.* (2018) Peruvioses A to F, sucrose esters from the exudate of *Physalis peruviana* fruit as α -amylase inhibitors. *Carbohydrate Research* 461, 4–10
- 27 Korenblum, E. *et al.* (2020) Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. *Proc Natl Acad Sci USA* 117, 3874–3883

- 28 Mirnezhad, M. *et al.* (2010) Metabolomic analysis of host plant resistance to thrips in wild and cultivated tomatoes. *Phytochem. Analysis* 21, 110–117
- 29 Leckie, B.M. *et al.* (2016) Differential and Synergistic Functionality of Acylsugars in Suppressing Oviposition by Insect Herbivores. *PLoS ONE* 11, e0153345
- 30 Smeda, J.R. *et al.* (2018) Combination of Acylglucose QTL reveals additive and epistatic genetic interactions and impacts insect oviposition and virus infection. *Mol. Breeding* 38, 3
- 31 Leckie, B.M. *et al.* (2012) Quantitative trait loci increasing acylsugars in tomato breeding lines and their impacts on silverleaf whiteflies. *Mol. Breeding* 30, 1621–1634
- 32 Leckie, B.M. *et al.* (2013) Quantitative trait loci regulating sugar moiety of acylsugars in tomato. *Mol. Breeding* 31, 957–970
- 33 Wang, Z. and Jones, A.D. (2014) Profiling of Stable Isotope Enrichment in Specialized Metabolites Using Liquid Chromatography and Multiplexed Nonselective Collision-Induced Dissociation. *Anal. Chem.* 86, 10600–10607
- 34 Fan, P. *et al.* (2020) Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. *Biorxiv*
- 35 Nadakuduti, S.S. *et al.* (2017) Characterization of Trichome-Expressed BAHD Acyltransferases in *Petunia axillaris* Reveals Distinct Acylsugar Assembly Mechanisms within the Solanaceae. *Plant Physiol.* 175, 36–50
- 36 Liu, X. *et al.* (2017) Profiling, isolation and structure elucidation of specialized acylsucrose metabolites accumulating in trichomes of *Petunia* species. *Metabolomics* 13, 85
- 37 Schillmiller, A.L. *et al.* (2012) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. *Proc. Natl. Acad. Sci.* 109, 16377–16382