

CHARACTERIZATION OF PROMISCUOUS TERPENE BIOSYNTHETIC ENZYMES TO
IDENTIFY THE SYNTHESSES OF NOVEL TERPENOIDS

By

Nicholas Schlecht

A DISSERTATION

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

Biochemistry & Molecular Biology – Doctor of Philosophy
Molecular Plant Sciences – Dual Major

2025

ABSTRACT

The growing and aging population coupled with the demand for environmentally sustainable solutions necessitates innovative developments in pharmaceuticals and agrochemicals. Terpenoids, a diverse group of specialized metabolites, have the potential to address these mounting challenges. Chapter 1 reviews terpene biosynthesis, while the remaining chapters cover research advances that improve our understanding of plant terpene biosynthesis and its applications. These research chapters cover the characterization of multiple CYP76BK1 orthologs involved in furanoclerodane metabolism, the discovery of the notably promiscuous *Ajuga reptans* CYP736A358, and exploration of novel enzyme-synthetic substrate analog combinations to yield semi-synthetic products. The CYP76BK1 orthologs were shown to catalyze the production of furan and one of two distinct lactone ring moieties on primarily clerodane backbones. In contrast, CYP736A358 exhibited limited product promiscuity, yet remarkable substrate promiscuity towards diterpenes, suggesting its place as a biotechnology multi-tool. In the final chapter, evolutionarily distinct sesquiterpene synthases natively producing similar products demonstrated different preferences towards synthetic analogs of their native substrate, the products of which had varied anti-fungal activities. These findings not only underscore the enzymatic flexibility within terpene specialized metabolism but provide new tools to produce commercially relevant terpenoids.

Copyright by
NICHOLAS SCHLECHT
2025

ACKNOWLEDGEMENTS

Graduate school is a challenging journey, made even more so by COVID-19. Fortunately, I am surrounded by an incredible community I can lean on for support. First, I want to express my gratitude to my advisor, Björn Hamberger for his advising and mentorship. Björn encouraged me pursue opportunities I might have otherwise overlooked and skillfully balanced providing guidance with independence. His efforts to foster a collaborative lab environment created a space where we could both mentor and be mentored by others. I'd like to particularly thank our post-doc Trine Andersen, whose mentorship and friendship were instrumental within the lab. I also appreciate the insights and guidance from my committee, which not only helped me benchmark my PhD progress, but substantially improved my presentation skills. I'd also like to thank the staff at the mass spec core and NMR facility at MSU as they were essential to my projects and aided in developing my expertise. I had the privilege of teaching recitation sections for several courses, and I'd like to thank the various professors I worked under for granting me teaching opportunities and evaluating my skills.

Beyond that I would like to thank the various graduate student groups, training programs, and associated staff that I have been involved with, including PBHS, GRIT, The BMS program, BMB program, and the Association for Molecular Plant Science Students. These communities, particularly AMPSS, provided me with new opportunities, friendships, and support network. Finally, I'd like to thank my various other friends and family outside of the university. Their encouragement and support were invaluable in helping me navigate the toughest moments of this journey.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	vi
Chapter 1: A review of terpenoid evolution, biosynthesis, and applications	1
REFERENCES.....	13
Chapter 2: CYP76BK1 orthologs catalyze furan and lactone ring formation in clerodane diterpenoids across the mint family	21
REFERENCES.....	51
APPENDIX	58
Chapter 3: An opportunistic cytochrome P450 in <i>Ajuga reptans</i> oxidizes diverse labdane-derived diterpenes	99
REFERENCES.....	117
APPENDIX	122
Chapter 4: Exploiting diverse viridiflorol synthase promiscuity to produce novel semi- synthetic anti-fungal terpenes	139
REFERENCES.....	159
APPENDIX	165
Chapter 5: Additional Research Contributions and Perspectives:	203
REFERENCES.....	207
Future Directions:	208

LIST OF ABBREVIATIONS

TPS	terpene synthase
CYP	cytochrome P450
2OGD	2-oxoglutarate dependent dioxygenase
DOX	dioxygenase
MVA	mevalonate
MEP	methylerythritol phosphate
IDP	isopentenyl diphosphate
DMADP	dimethylallyl diphosphate
FDP	farnesyl diphosphate
GDP	geranyl diphosphate
GGDP	geranylgeranyl diphosphate
PT	prenyltransferase
BAHD	(family named for first 4 members) BEAT , AHCT , HCBT , DAT
DNP	dictionary of natural products
KPP	kolavenyl diphosphate
DXS	1-deoxy-D-xylulose-5-phosphate synthase
BSTFA	<i>N,O</i> -Bis(trimethylsilyl)trifluoroacetamide
TMCS	Trimethylchlorosilane
NMR	nuclear magnetic resonance
GC(LC)-MS	gas chromatography (liquid chromatography) mass spectrometry
FID	flame ionization detection
APCI	atmospheric pressure chemical ionization

Chapter 1: A review of terpenoid evolution, biosynthesis, and applications

Central metabolism vs plant specialized terpene metabolism

The fitness of an organism relies on its ability to respond to environmental stimuli, inter/intraspecific communication, and managing biotic and abiotic stresses. These factors vary by ecological niche, leading to diverse response strategies. While animals often use dynamic and mobile responses such as fight-or-flight, many sessile organisms like plants developed diverse specialized metabolism to improve their fitness within their respective environments.¹

Central metabolism encompasses the biosynthesis of metabolites necessary for survival, therefore it is ubiquitous throughout life. In contrast specialized metabolism, while not required for survival, enhances fitness by providing adaptive advantages in response to unique environmental stimuli. The diversity of ecological environments led to the evolution of phylogenetically distinct response strategies, including specialized metabolic pathways.² For example, glucosinolates are a group of specialized metabolites generally unique to the Brassicales order. These specialized metabolites emerged, in part, due to an arms race with insect herbivores such as the butterflies of the subfamily Pierinae.³ The unique evolutionary responses developed by different phylogenetic lineages has resulted in the incredible diversity of specialized metabolism we have today.

Specialized metabolites are categorized by their structural and biosynthetic similarities, including phenylpropanoids, alkaloids, glucosinolates, polyketides, and terpenoids. To track these data, natural product databases with distinct curation approaches such as the Dictionary of Natural Products or Terokit have been designed

to centralize the growing knowledge base.^{4,5} Amongst natural products, terpenoids are the most diverse, with over 110,000 unique compounds reported across >14,000 species in the TeroKit database, with plants representing over 75% of the biological sources.⁴ In biochemistry, structure determines function. Therefore, terpenoids' structural diversity makes them a valuable resource for accessing diverse chemical properties. In plants, terpenoids play roles in photosynthesis, respiration, development, communication, and defense.⁶ The enrichment of terpenoids in plants reflects the need for sessile organisms like plants to have unique means to respond to their environment.^{7,8}

To grasp the diversity of terpenoids, it is important to understand the structural features that define them and the processes by which they are synthesized. Terpenes are a class of organic compounds derived from prenyl diphosphates, containing one or more characteristic 5-carbon 'isoprene units'. Terpene biosynthesis is initiated by either protonation or diphosphate cleavage of a given length prenyl diphosphate followed by stereospecific cycloisomerizations. Terpenoids are a broader classification—including additional decorations, cleavages, or rearrangements from a terpene backbone.

Prenyl diphosphate biosynthesis

To synthesize terpenes, first prenyl diphosphates must be produced through one of two different routes. Two non-homologous pathways synthesize the compounds isopentyl diphosphate (IDP) and dimethyl allyl diphosphate (DMADP), which serve as the 'isoprene unit' building blocks essential to produce larger prenyl diphosphates. The mevalonate (MVA) pathway is generally conserved throughout archaea, eukaryotes and select bacteria, while the methylerythritol phosphate (MEP) pathway is only found in

bacteria and the plastids of plants.⁹ IDP and DMADP can be condensed into the varied length prenyl diphosphate substrates for terpene synthases (TPSs).¹⁰

Prenyltransferases (PTs) condense DMADP and IDP by cleaving the diphosphate of a prenyl diphosphate thus facilitating a nucleophilic attack from IDP.¹¹ This short chain elongation within PTs is universally conserved, strongly supporting the presence of terpenes in the last universal common ancestor.¹²

Plants uniquely possess both the MVA and MEP pathways, which are cytosolically and plastidially located, respectively. The separation of two different pathways enables complex differential regulation and specialization, allowing for more dynamic responses.¹³ Farnesyl diphosphate (FDP) and squalene synthases produce the C15 and C30 precursors in the cytosol via the MVA pathway (Figure 1.1).¹⁰ In contrast, geranyl diphosphate (GDP) and geranylgeranyl diphosphate (GGDP) synthases have N-terminal plastid targeting motifs localizing them to the plastid where they produce their respective 10 and 20 carbon chain prenyl diphosphates.^{10,14} These prenyl diphosphates are the precursors to mono- (C10), sesqui- (C15), di- (C20), and triterpenes (C30). While the MVA and MEP pathways are conventionally considered isolated from one another, recent evidence suggests there is some interconnectivity between the pathways.^{15,16}

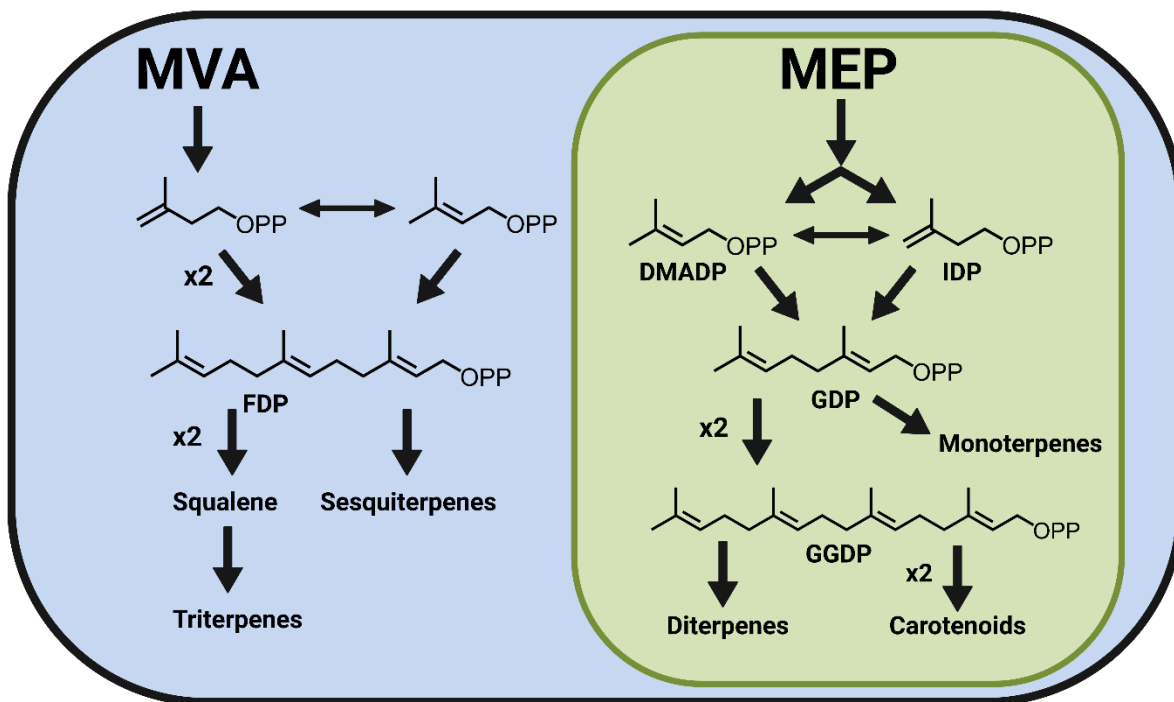


Figure 1.1: Overview of plant prenyl diphosphate and terpene biosynthesis. In both the MVA and MEP pathways IPP and DMAPP are generated, but specific PTs and TPSs limit if a given prenyl diphosphate is cytosolically or plastidially located.

Terpene synthase biochemistry

Terpene synthases generally take achiral prenyl diphosphate substrates and ionize them to facilitate unique carbocation cycloisomerizations. In plants, these enzymes are co-localized with their respective substrates: mono- and diterpene synthases are found in plastids, while sesquiterpene and triterpene synthases are cytosolically located. There are functionally two classes of TPSs distinguished by their mechanism: a class II TPS, which evolved from an ancient triterpene cyclase, and a class I TPS, which evolved from an ancient PT.¹⁷ Class II TPSs use protonation to initiate carbocation rearrangements and generally use a conserved DXDD motif, whereas class I TPSs coordinate magnesium ions with the conserved DDXXD and NSE/DTE motifs to abstract the diphosphate, facilitating carbocation

cycloisomerizations.^{18,19} Class I TPSs often utilize a prenyl diphosphate as their substrate, but others use the product of a class II TPS as their substrate. While there are distinct domains attributed to the different classes, plant TPSs evolved from a fusion of a bacterial class II and class I TPS, linking the $\beta\gamma$ and α domains of each respective class (Figure 1.2a).^{17,20} Consequently, some modern TPSs have a bifunctional class II/class I activity, illustrated in Figure 1.2b. At least two subsequent duplications in plant lineages gave rise to the present TPS subfamilies (Figure 1.2a). The TPS-d and TPS-h subfamilies retain bifunctional class II/class I TPSs. An ancient duplication and divergence produced the TPS-c and TPS-e/f subfamilies, which are diTPS subfamilies that have lost the class I and class II activities respectively. The TPS-e/f subfamily diverged again, forming the remaining mono-, sesqui-, and diTPSs subfamilies.²¹

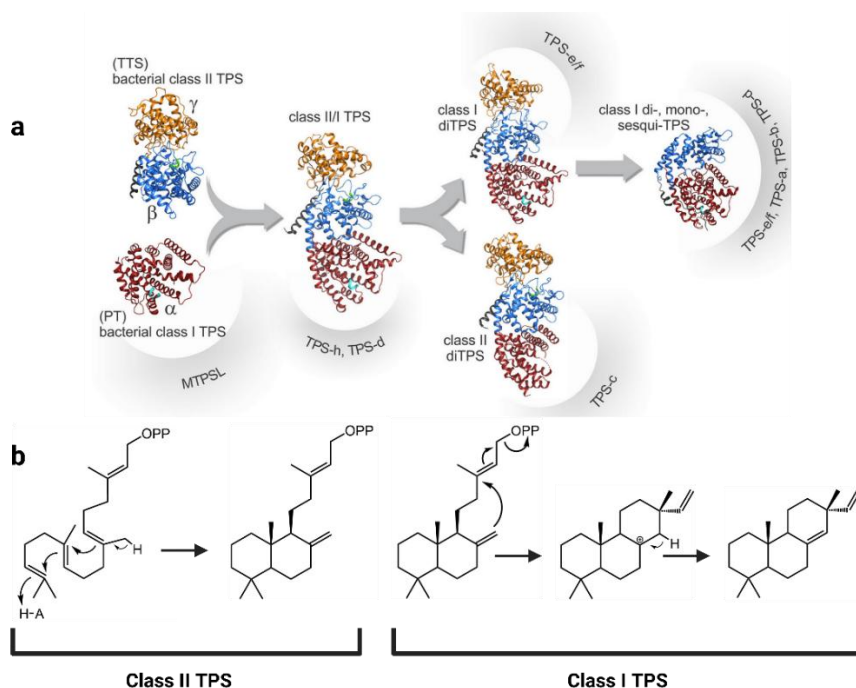


Figure 1.2: Plant TPS structure, evolution, and mechanisms. **a.** The illustration from Karunanithi and Zerbe 2019 depicts the evolution of plant TPSs, portraying key aspects such as structural changes, substrate preferences, and TPS class.¹⁷ **b.** The class II TPS utilizes a folded GGDP to facilitate cycloisomerization and catalyzes a protonation necessary for rearrangement and quenching. The class II product often serves as the

Figure 1.2 (cont'd)

substrate for the class I TPS, which cleaves the diphosphate allowing for additional carbocation rearrangements and a final quenching. This mechanism could be performed by a bifunctional class II/I or two separate enzymes. The decalin core is the characteristic bicyclic structure found as the class II product and is a key feature of labdane diterpenoids.

TPSs can exhibit impressive plasticity and promiscuity. Several studies show TPS catalytic landscapes can have accessible alternative catalytic specificities, making them evolutionarily plastic enzymes capable of neo- and subfunctionalization.^{22–24} The promiscuity can vary widely: some TPSs are highly specific, some catalyze similar reactions across diverse substrates, and others act on a single substrate yet produce over a dozen distinct products.^{25–27} Enzyme promiscuity results in branched and interconnected biosynthetic pathways. This variability facilitates the formation of ‘metabolic grids’, where complex metabolic networks are generated with a minimal number of genes.²⁸

Monoterpenes and sesquiterpenes are generally produced by class I TPSs and triterpenes are produced by class II TPSs, but diterpenes can employ both. Class II diTPSs vary in stereochemistry, quenching, and various methyl and hydride shifts, but they all produce labdane-derived diterpenes recognizable by the bicyclic decalin core (Figure 1.2). In contrast, Class I diTPSs that utilize GGDP are less restricted in their product profiles. Although class II diTPS products are generally limited to labdane-derived structures, they are often further cyclized by a subsequent class I diTPS, expanding labdane-derived diterpene diversity. Given their frequent substrate promiscuity, diTPS pathways often have modular combinations of class II and class I TPSs, allowing additional terpene diversity.^{29–31}

Observed promiscuity and plasticity of TPSs is not only a result of evolutionary

pressures but also a result of their substrate and mechanism. Prenyl diphosphates are linear, electron-rich, and contain multiple tertiary carbons. Their structure permits diverse conformations while providing multiple nucleophiles, and hyperconjugated positions to stabilize carbocations. TPSs initiate carbocation-cascades, yet they often play a passive role, aiming to control intrinsically reactive carbocation rearrangements. This passive role is best exemplified by common *ab initio* approaches, such as gas-phase quantum chemistry and density functional theory, which have accurately described the chemistry without consideration of enzyme structural information.^{32–34} Nonetheless, TPSs do play crucial roles beyond initiation, including enabling ranged proton transfers, promoting stereospecificity, stabilizing carbocation intermediates, and supervising carbocation quenching.³² TPSs are therefore able to employ rich carbocation chemistry that yields the expansive terpene diversity. While most terpenes are produced by TPSs, it is important to note that many unconventional enzymes have been reported to catalyze TPS-like reactions to produce terpenes or related products.³⁵

Expanding terpene complexity (CYPs and other decorating enzymes)

The structural diversity of terpenoids substantially expands due to extensive enzymatic decorations on terpene backbones. Terpenes can undergo a range of oxidations to produce various functional groups including alcohols, aldehydes, ketones, carboxylic acids, as well as oxygen-bridged functional groups like ethers, esters, and epoxides.³⁶ In addition to these functional groups, oxidative enzymes can cause large structural changes (Figure 1.3). Oxidative cleavage has been reported, producing a truncated version of the terpene known as norterpenoids, such as the novel trisnorsesquiterpene discodiene in *Dictyostelium discoideum*.³⁷ It is not unusual for

oxidations to lead to other structural rearrangements, such as the unique ring closures found in the macrocyclic diterpenoids throughout Euphorbiaceae or the unique 18(4→3) methyl shift that generates the abeo-abietanes found in *Tripterygium wilfordii*.^{38,39}

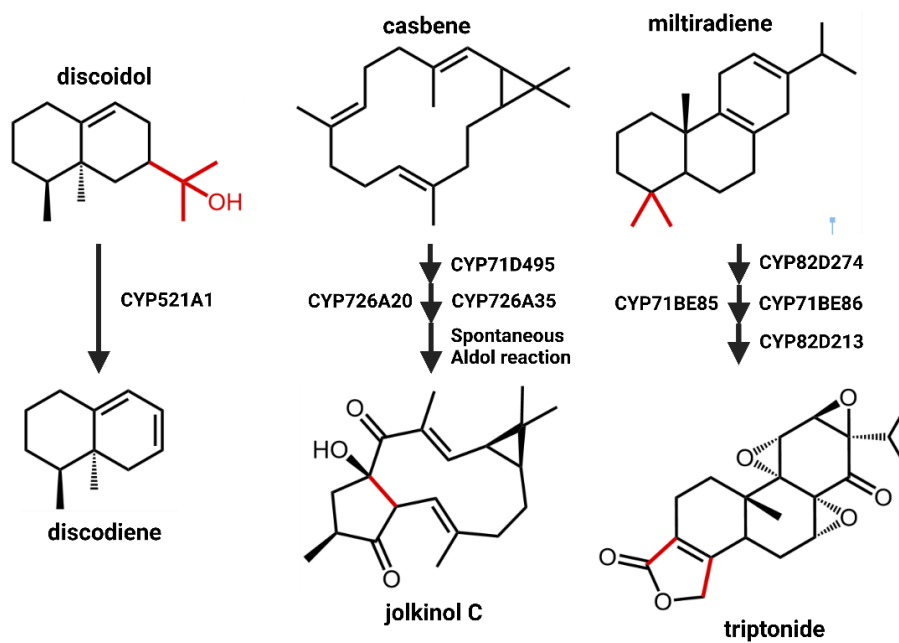


Figure 1.3. Complex oxidative terpene modifications. The examples shown include oxidations such as alcohols, ketones, and epoxides commonly catalyzed by oxidoreductases. Additionally, there are examples of oxidative cleavages, ring closures and methyl shifts highlighted in red.

The examples in Figure 1.3 are all enzymatically oxidized by cytochrome P450 monooxygenases (CYPs). CYPs are a superfamily of enzymes responsible for degradation of xenobiotics and biosynthesis of central and specialized metabolites, including terpenoids.⁴⁰ CYPs are heme-thiolate proteins that are typically membrane bound in eukaryotes, where they receive electrons through cytochrome P450 reductase, which in turn receives its reducing equivalents from NADPH.⁴¹ The CYP superfamily within the Viridiplantae are further arranged by placing families into clans. There are 11 conserved clans within vascular plants and several additional algae-specific clans.⁴²

Many plant CYP families emerged or expanded during terrestrialization to address the unique biotic and abiotic stresses experienced by early plants.⁴³ This expansion was partly a result of plants undergoing whole genome duplication events, which allowed duplicated CYPs to be recruited into specialized metabolism and to neo- or subfunctionalize.^{44,45} The CYP51, CYP71, CYP72, and CYP85 clans are all involved in terpenoid metabolism, with the CYP71 clan playing a particularly significant role.^{36,46}

While CYPs may be heavily involved in terpenoid biosynthesis, other enzymes including additional oxidoreductases and transferases are important in some terpenoid biosynthetic pathways. The 2-oxoglutarate-dependent dioxygenase (2OGD) superfamily, and more particularly the DOXC class lineage, play pivotal roles in both specialized and central metabolism in plants.^{47,48} Short chain dehydrogenases, a family with low sequence similarity, are another large oxidoreductase superfamily involved in terpenoid metabolism.^{49–51} Terpenoid biosynthesis often extends past the oxidoreductases as transferases can utilize the introduced oxidations for further modification. Often those modifications are some form of acylation or glycosylation. Acylations in terpene metabolism are typically catalyzed by the acyl-CoA dependent BAHD family and glycosylations are from the notoriously promiscuous GT1 family of glycosyltransferases.^{52,53}

Oxidations, acylations, glycosylations, and other decorations play important roles in bioactivity. Complexity with distinct steric and electrostatic interactions is necessary for strong and specific binding that grants a compound its bioactivity.^{54,55} Terpene backbones already demonstrate some distinct geometries for steric interactions but are still simpler than decorated terpenoids and the lack of oxidations on terpene backbones

limits their solubility. Decorated terpenoids have better solubility, increased structural complexity, and introduce distinctive polarity necessary for specific binding.

Natural product advancements as an interdisciplinary field

Leveraging specialized metabolism is important in combating complex global issues like the growing antibiotic resistance, climate change and food security. Developing a commercial product requires cross-disciplinary efforts, spanning from metabolite discovery to product development, throughout various research stages and fields. Some of the major hurdles are the discovery and characterization of new specialized metabolites. Ethnobotany helps overcome these hurdles by exploring the relationships between people and plants, leading to the identification of thousands of plants used in traditional medicines.⁵⁶ These discoveries serve as a foundation for future experimentation to evaluate bioactivities and isolate unique compounds.⁵⁷

Identifying bioactive specialized metabolites like terpenoids creates avenues for commercialization of relevant compounds. A simple approach to mass producing natural products is direct extraction from the plant, but some natural products are only found in very low abundances (<0.1% dry weight).^{58,59} In cases like this, other approaches are desirable. While total chemical synthesis is an alternative, the numerous stereogenic centers found in chiral natural products like terpenoids make the total synthesis particularly challenging. For example, a 21-step synthetic route towards the anti-cancer diterpenoid paclitaxel was considered a landmark, yet only had a net yield of 0.118%, illustrating the impracticality of this method to scale commercially.⁶⁰ A more feasible method than total synthesis is a semi-synthetic approach, where a biosynthetic intermediate is produced naturally and derivatized to the desired product by

chemists. Semi-synthesis requires an abundant intermediate that is also not always accessible. Thankfully, a third approach is available: utilizing metabolic engineering to produce terpenoids. Metabolic engineering requires substantial upfront research, including elucidating the biosynthesis of relevant terpenoids. With the advent of modern sequencing, mass spectrometry, and computational tools, a wide array of new bio- and chemoinformatic approaches are now available for identifying candidate biosynthesis genes. These new tools are complemented by the robust heterologous expression systems, such as *Agrobacterium*-mediated transient expression in *Nicotiana benthamiana*, that are valuable for characterizing candidates.⁶¹ This explosion of genomic, transcriptomic, proteomic, metabolomic, and chemoinformatic resources and data has unveiled details of terpene metabolism, regulation, and function.^{62–64}

Terpenoid enrichment within the Lamiaceae family

The Lamiaceae (mint) family of plants may only be the sixth largest family of angiosperms, but it boasts exceptional terpenoid diversity, representing roughly a quarter of all diterpenoids reported in the Dictionary of Natural Products.^{5,20} Compounding evidence suggests that this is, in part, due to multiple ancient whole genome duplications enabling opportunities for neofunctionalization of genes.^{65–67,67} The range of terpenoids within Lamiaceae has attracted the attention of pharmaceutical, agricultural, and cosmetic industries.

The Lamiaceae family produces a wide array of terpenoids, with a notable enrichment of clerodane diterpenoids, particularly for the Ajugoideae and Scutellarioideae subfamilies.⁵ Clerodanes are produced by class II diTPSs and are distinguished from other labdane-derived diterpenes by the series of methyl and hydride

shifts that result from blocking the prototypical carbocation quenching site, yielding a distinct structure while maintaining the bicyclic decalin core of other labdane-derived diterpenes.²³ Interestingly, clerodane synthases tend to not have a corresponding class I TPS partner, with the only known example found in one report of *Scutellaria barbata* diTPSs.^{67–72} Clerodanes nonetheless remain one of the most diverse diterpene backbones because of the wide range of oxidations observed, including many heterocyclic structures.

Summary

This thesis aims to provide further insight into terpenoid metabolism and offer resources for future pathway discovery and engineering. The second chapter has been published and covers the discovery of CYP76BK1 orthologs across Lamiaceae species, their functional characterization, and the evaluation of their role. The third chapter also elucidates a Lamiaceae CYP, CYP736A358, but instead evaluates its ability to oxidize an array of diterpene backbones. Chapter 4 explores an alternative route to producing semi-synthetic products and their anti-fungal effects. The appendices cover additional projects that evaluate the roles of TPSs using genomic data as well as a collaboration utilizing new chemoinformatic approaches to better understand terpenoid biosynthesis. Lastly, the thesis outlines some future directions. Together, this work covers enzyme promiscuity within terpene metabolism, both how it is found naturally and how it can be manipulated as a tool.

REFERENCES

- (1) Rieseberg, T. P.; Dadras, A.; Fürst-Jansen, J. M. R.; Dhabalia Ashok, A.; Darienko, T.; de Vries, S.; Irisarri, I.; de Vries, J. Crossroads in the Evolution of Plant Specialized Metabolism. *Semin. Cell Dev. Biol.* **2023**, *134*, 37–58. <https://doi.org/10.1016/j.semcdb.2022.03.004>.
- (2) Weng, J.-K.; Lynch, J. H.; Matos, J. O.; Dudareva, N. Adaptive Mechanisms of Plant Specialized Metabolism Connecting Chemistry to Function. *Nat. Chem. Biol.* **2021**, *17* (10), 1037–1045. <https://doi.org/10.1038/s41589-021-00822-6>.
- (3) Edger, P. P.; Heidel-Fischer, H. M.; Bekaert, M.; Rota, J.; Glöckner, G.; Platts, A. E.; Heckel, D. G.; Der, J. P.; Wafula, E. K.; Tang, M.; Hofberger, J. A.; Smithson, A.; Hall, J. C.; Blanchette, M.; Bureau, T. E.; Wright, S. I.; dePamphilis, C. W.; Eric Schranz, M.; Barker, M. S.; Conant, G. C.; Wahlberg, N.; Vogel, H.; Pires, J. C.; Wheat, C. W. The Butterfly Plant Arms-Race Escalated by Gene and Genome Duplications. *Proc. Natl. Acad. Sci.* **2015**, *112* (27), 8362–8366. <https://doi.org/10.1073/pnas.1503926112>.
- (4) Zeng, T.; Liu, Z.; Zhuang, J.; Jiang, Y.; He, W.; Diao, H.; Lv, N.; Jian, Y.; Liang, D.; Qiu, Y.; Zhang, R.; Zhang, F.; Tang, X.; Wu, R. TeroKit: A Database-Driven Web Server for Terpenome Research. *J. Chem. Inf. Model.* **2020**, *60* (4), 2082–2090. <https://doi.org/10.1021/acs.jcim.0c00141>.
- (5) Taylor & Francis Group. *Dictionary of Natural Products* (v33.1). <https://dnp.chemnetbase.com/chemical/ChemicalSearch.xhtml?dswid=6473> (accessed 2024-11-22).
- (6) Tholl, D.; Lee, S. Terpene Specialized Metabolism in Arabidopsis Thaliana. *Arab. Book* **2011**, *2011* (9). <https://doi.org/10.1199/tab.0143>.
- (7) Howe, G. A.; Jander, G. Plant Immunity to Insect Herbivores. *Annu. Rev. Plant Biol.* **2008**, *59* (Volume 59, 2008), 41–66. <https://doi.org/10.1146/annurev.arplant.59.032607.092825>.
- (8) Singh, A.; Dwivedi, P. Methyl-Jasmonate and Salicylic Acid as Potent Elicitors for Secondary Metabolite Production in Medicinal Plants: A Review. *J. Pharmacogn. Phytochem.* **2018**, *7* (1), 750–757.
- (9) Lombard, J.; Moreira, D. Origins and Early Evolution of the Mevalonate Pathway of Isoprenoid Biosynthesis in the Three Domains of Life. *Mol. Biol. Evol.* **2011**, *28* (1), 87–99. <https://doi.org/10.1093/molbev/msq177>.
- (10) Vranová, E.; Coman, D.; Gruissem, W. Network Analysis of the MVA and MEP Pathways for Isoprenoid Synthesis. *Annu. Rev. Plant Biol.* **2013**, *64* (1), 665–700. <https://doi.org/10.1146/annurev-arplant-050312-120116>.

- (11) Chang, H.-Y.; Cheng, T.-H.; Wang, A. H.-J. Structure, Catalysis, and Inhibition Mechanism of Prenyltransferase. *IUBMB Life* **2021**, *73* (1), 40–63. <https://doi.org/10.1002/iub.2418>.
- (12) Hoshino, Y.; Villanueva, L. Four Billion Years of Microbial Terpenome Evolution. *FEMS Microbiol. Rev.* **2023**, *47* (2), fuad008. <https://doi.org/10.1093/femsre/fuad008>.
- (13) Pu, X.; Dong, X.; Li, Q.; Chen, Z.; Liu, L. An Update on the Function and Regulation of Methylerythritol Phosphate and Mevalonate Pathways and Their Evolutionary Dynamics. *J. Integr. Plant Biol.* **2021**, *63* (7), 1211–1226. <https://doi.org/10.1111/jipb.13076>.
- (14) Patron, N. J.; Waller, R. F. Transit Peptide Diversity and Divergence: A Global Analysis of Plastid Targeting Signals. *BioEssays* **2007**, *29* (10), 1048–1058. <https://doi.org/10.1002/bies.20638>.
- (15) Opitz, S.; Nes, W. D.; Gershenzon, J. Both Methylerythritol Phosphate and Mevalonate Pathways Contribute to Biosynthesis of Each of the Major Isoprenoid Classes in Young Cotton Seedlings. *Phytochemistry* **2014**, *98*, 110–119. <https://doi.org/10.1016/j.phytochem.2013.11.010>.
- (16) Lipko, A.; Pączkowski, C.; Perez-Fons, L.; Fraser, P. D.; Kania, M.; Hoffman-Sommer, M.; Danikiewicz, W.; Rohmer, M.; Poznanski, J.; Swiezewska, E. Divergent Contribution of the MVA and MEP Pathways to the Formation of Polyprenols and Dolichols in Arabidopsis. *Biochem. J.* **2023**, *480* (8), 495–520. <https://doi.org/10.1042/BCJ20220578>.
- (17) Karunanithi, P. S.; Zerbe, P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Front. Plant Sci.* **2019**, *10*. <https://doi.org/10.3389/fpls.2019.01166>.
- (18) Prisic, S.; Xu, J.; Coates, R. M.; Peters, R. J. Probing the Role of the DXDD Motif in Class II Diterpene Cyclases. *ChemBioChem* **2007**, *8* (8), 869–874. <https://doi.org/10.1002/cbic.200700045>.
- (19) Zhou, K.; Peters, R. J. Investigating the Conservation Pattern of a Putative Second Terpene Synthase Divalent Metal Binding Motif in Plants. *Phytochemistry* **2009**, *70* (3), 366–369. <https://doi.org/10.1016/j.phytochem.2008.12.022>.
- (20) Boachon, B.; Buell, C. R.; Crisovan, E.; Dudareva, N.; Garcia, N.; Godden, G.; Henry, L.; Kamileen, M. O.; Kates, H. R.; Kilgore, M. B.; Lichman, B. R.; Mavrodiev, E. V.; Newton, L.; Rodriguez-Lopez, C.; O'Connor, S. E.; Soltis, D.; Soltis, P.; Vaillancourt, B.; Wiegert-Rininger, K.; Zhao, D. Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. *Mol. Plant* **2018**, *11* (8), 1084–1096. <https://doi.org/10.1016/j.molp.2018.06.002>.

- (21) Jia, Q.; Brown, R.; Köllner, T. G.; Fu, J.; Chen, X.; Wong, G. K.-S.; Gershenzon, J.; Peters, R. J.; Chen, F. Origin and Early Evolution of the Plant Terpene Synthase Family. *Proceedings of the National Academy of Sciences* **2022**, *119* (15), e2100361119. <https://doi.org/10.1073/pnas.2100361119>.
- (22) O'Maille, P. E.; Malone, A.; Dellas, N.; Andes Hess, B.; Smentek, L.; Sheehan, I.; Greenhagen, B. T.; Chappell, J.; Manning, G.; Noel, J. P. Quantitative Exploration of the Catalytic Landscape Separating Divergent Plant Sesquiterpene Synthases. *Nat. Chem. Biol.* **2008**, *4* (10), 617–623. <https://doi.org/10.1038/nchembio.113>.
- (23) Potter, K. C.; Zi, J.; Hong, Y. J.; Schulte, S.; Malchow, B.; Tantillo, D. J.; Peters, R. J. Blocking Deprotonation with Retention of Aromaticity in a Plant Ent-Copalyl Diphosphate Synthase Leads to Product Rearrangement. *Angew. Chem. Int. Ed.* **2016**, *55* (2), 634–638. <https://doi.org/10.1002/anie.201509060>.
- (24) Whitehead, J. N.; Leferink, N. G. H.; Johannissen, L. O.; Hay, S.; Scrutton, N. S. Decoding Catalysis by Terpene Synthases. *ACS Catal.* **2023**, *13* (19), 12774–12802. <https://doi.org/10.1021/acscatal.3c03047>.
- (25) Garms, S.; Köllner, T. G.; Boland, W. A Multiproduct Terpene Synthase from *Medicago truncatula* Generates Cadalane Sesquiterpenes via Two Different Mechanisms. *J. Org. Chem.* **2010**, *75* (16), 5590–5600. <https://doi.org/10.1021/jo100917c>.
- (26) Caniard, A.; Zerbe, P.; Legrand, S.; Cohade, A.; Valot, N.; Magnard, J.-L.; Bohlmann, J.; Legendre, L. Discovery and Functional Characterization of Two Diterpene Synthases for Sclareol Biosynthesis in *Salvia sclarea*(L.) and Their Relevance for Perfume Manufacture. *BMC Plant Biol.* **2012**, *12* (1), 119. <https://doi.org/10.1186/1471-2229-12-119>.
- (27) Johnson, S. R.; Bhat, W. W.; Sadre, R.; Miller, G. P.; Garcia, A. S.; Hamberger, B. Promiscuous Terpene Synthases from *Prunella vulgaris* Highlight the Importance of Substrate and Compartment Switching in Terpene Synthase Evolution. *New Phytol.* **2019**, *223* (1), 323–335. <https://doi.org/10.1111/nph.15778>.
- (28) Lanier, E. R.; Andersen, T. B.; Hamberger, B. Plant Terpene Specialized Metabolism: Complex Networks or Simple Linear Pathways? *Plant J.* **2023**, *114* (5), 1178–1201. <https://doi.org/10.1111/tpj.16177>.
- (29) Zerbe, P.; Bohlmann, J. Plant Diterpene Synthases: Exploring Modularity and Metabolic Diversity for Bioengineering. *Trends Biotechnol.* **2015**, *33* (7), 419–428. <https://doi.org/10.1016/j.tibtech.2015.04.006>.

- (30) Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Møller, B. L.; Hamberger, B. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angew. Chem. Int. Ed.* **2016**, *55* (6), 2142–2146. <https://doi.org/10.1002/anie.201510650>.
- (31) Jia, M.; Mishra, S. K.; Tufts, S.; Jernigan, R. L.; Peters, R. J. Combinatorial Biosynthesis and the Basis for Substrate Promiscuity in Class I Diterpene Synthases. *Metab. Eng.* **2019**, *55*, 44–58. <https://doi.org/10.1016/j.ymben.2019.06.008>.
- (32) Raz, K.; Levi, S.; Gupta, P. K.; Major, D. T. Enzymatic Control of Product Distribution in Terpene Synthases: Insights from Multiscale Simulations. *Curr. Opin. Biotechnol.* **2020**, *65*, 248–258. <https://doi.org/10.1016/j.copbio.2020.06.002>.
- (33) J. Tantillo, D. Biosynthesis via Carbocations: Theoretical Studies on Terpene Formation. *Nat. Prod. Rep.* **2011**, *28* (6), 1035–1053. <https://doi.org/10.1039/C1NP00006C>.
- (34) Tantillo, D. J. The Carbocation Continuum in Terpene Biosynthesis—Where Are the Secondary Cations? *Chem. Soc. Rev.* **2010**, *39* (8), 2847. <https://doi.org/10.1039/b917107j>.
- (35) Rudolf, J. D.; Chang, C.-Y. Terpene Synthases in Disguise: Enzymology, Structure, and Opportunities of Non-Canonical Terpene Synthases. *Nat. Prod. Rep.* **2020**, *37* (3), 425–463. <https://doi.org/10.1039/C9NP00051H>.
- (36) Bathe, U.; Tissier, A. Cytochrome P450 Enzymes: A Driving Force of Plant Diterpene Diversity. *Phytochemistry* **2019**, *161*, 149–162. <https://doi.org/10.1016/j.phytochem.2018.12.003>.
- (37) Chen, X.; Luck, K.; Rabe, P.; Dinh, C. Q.; Shaulsky, G.; Nelson, D. R.; Gershenzon, J.; Dickschat, J. S.; Köllner, T. G.; Chen, F. A Terpene Synthase-Cytochrome P450 Cluster in *Dictyostelium Discoideum* Produces a Novel Trisnorsesquiterpene. *eLife* *8*, e44352. <https://doi.org/10.7554/eLife.44352>.
- (38) King, A. J.; Brown, G. D.; Gilday, A. D.; Forestier, E.; Larson, T. R.; Graham, I. A. A Cytochrome P450-Mediated Intramolecular Carbon–Carbon Ring Closure in the Biosynthesis of Multidrug-Resistance-Reversing Lathyrane Diterpenoids. *Chembiochem* **2016**, *17* (17), 1593–1597. <https://doi.org/10.1002/cbic.201600316>.
- (39) Hansen, N. L.; Kjaerulff, L.; Heck, Q. K.; Forman, V.; Staerk, D.; Møller, B. L.; Andersen-Ranberg, J. Tripterygium Wilfordii Cytochrome P450s Catalyze the Methyl Shift and Epoxidations in the Biosynthesis of Triptonide. *Nat. Commun.* **2022**, *13* (1), 5011. <https://doi.org/10.1038/s41467-022-32667-5>.

- (40) Xu, J.; Wang, X.; Guo, W. The Cytochrome P450 Superfamily: Key Players in Plant Development and Defense. *J. Integr. Agric.* **2015**, *14* (9), 1673–1686. [https://doi.org/10.1016/S2095-3119\(14\)60980-1](https://doi.org/10.1016/S2095-3119(14)60980-1).
- (41) Jensen, K.; Møller, B. L. Plant NADPH-Cytochrome P450 Oxidoreductases. *Phytochemistry* **2010**, *71* (2), 132–141. <https://doi.org/10.1016/j.phytochem.2009.10.017>.
- (42) Hansen, C. C.; Nelson, D. R.; Møller, B. L.; Werck-Reichhart, D. Plant Cytochrome P450 Plasticity and Evolution. *Mol. Plant* **2021**, *14* (8), 1244–1265. <https://doi.org/10.1016/j.molp.2021.06.028>.
- (43) Werck-Reichhart, D.; Nelson, D. R.; Renault, H. Cytochromes P450 Evolution in the Plant Terrestrialization Context. *Philos. Trans. R. Soc. B Biol. Sci.* **2024**, 379 (1914), 20230363. <https://doi.org/10.1098/rstb.2023.0363>.
- (44) Hamberger, B.; Bak, S. Plant P450s as Versatile Drivers for Evolution of Species-Specific Chemical Diversity. *Philos. Trans. R. Soc. B Biol. Sci.* **2013**, 368 (1612), 20120426. <https://doi.org/10.1098/rstb.2012.0426>.
- (45) Kawai, Y.; Ono, E.; Mizutani, M. Expansion of Specialized Metabolism-Related Superfamily Genes via Whole Genome Duplications during Angiosperm Evolution. *Plant Biotechnol.* **2014**, *31* (5), 579–584. <https://doi.org/10.5511/plantbiotechnology.14.0901a>.
- (46) Weitzel, C.; Simonsen, H. T. Cytochrome P450-Enzymes Involved in the Biosynthesis of Mono- and Sesquiterpenes. *Phytochem. Rev.* **2015**, *14* (1), 7–24. <https://doi.org/10.1007/s11101-013-9280-x>.
- (47) Kawai, Y.; Ono, E.; Mizutani, M. Evolution and Diversity of the 2–Oxoglutarate-Dependent Dioxygenase Superfamily in Plants. *Plant J.* **2014**, *78* (2), 328–343. <https://doi.org/10.1111/tpj.12479>.
- (48) Hu, Z.; Ren, L.; Bu, J.; Liu, X.; Li, Q.; Guo, W.; Ma, Y.; Wang, J.; Chen, T.; Wang, L.; Jin, B.; Tang, J.; Cui, G.; Guo, J.; Huang, L. Functional Characterization of a 2OGD Involved in Abietane-Type Diterpenoids Biosynthetic Pathway in *Salvia Miltiorrhiza*. *Front. Plant Sci.* **2022**, *13*. <https://doi.org/10.3389/fpls.2022.947674>.
- (49) Krause, S. T.; Liao, P.; Crocoll, C.; Boachon, B.; Förster, C.; Leidecker, F.; Wiese, N.; Zhao, D.; Wood, J. C.; Buell, C. R.; Gershenzon, J.; Dudareva, N.; Degenhardt, J. The Biosynthesis of Thymol, Carvacrol, and Thymohydroquinone in Lamiaceae Proceeds via Cytochrome P450s and a Short-Chain Dehydrogenase. *Proc. Natl. Acad. Sci.* **2021**, *118* (52), e2110092118. <https://doi.org/10.1073/pnas.2110092118>.

- (50) Okamoto, S.; Yu, F.; Harada, H.; Okajima, T.; Hattan, J.; Misawa, N.; Utsumi, R. A Short-Chain Dehydrogenase Involved in Terpene Metabolism from Zingiber Zerumbet. *FEBS J.* **2011**, *278* (16), 2892–2900. <https://doi.org/10.1111/j.1742-4658.2011.08211.x>.
- (51) Shimura, K.; Okada, A.; Okada, K.; Jikumaru, Y.; Ko, K.-W.; Toyomasu, T.; Sassa, T.; Hasegawa, M.; Kodama, O.; Shibuya, N.; Koga, J.; Nojiri, H.; Yamane, H. Identification of a Biosynthetic Gene Cluster in Rice for Momilactones. *J. Biol. Chem.* **2007**, *282* (47), 34013–34018. <https://doi.org/10.1074/jbc.M703344200>.
- (52) Moghe, G.; Kruse, L. H.; Petersen, M.; Scossa, F.; Fernie, A. R.; Gaquerel, E.; D'Auria, J. C. BAHD Company: The Ever-Expanding Roles of the BAHD Acyltransferase Gene Family in Plants. *Annu. Rev. Plant Biol.* **2023**, *74* (Volume 74, 2023), 165–194. <https://doi.org/10.1146/annurev-arplant-062922-050122>.
- (53) Tiwari, P.; Sangwan, R. S.; Sangwan, N. S. Plant Secondary Metabolism Linked Glycosyltransferases: An Update on Expanding Knowledge and Scopes. *Biotechnol. Adv.* **2016**, *34* (5), 714–739. <https://doi.org/10.1016/j.biotechadv.2016.03.006>.
- (54) Clemons, P. A.; Bodycombe, N. E.; Carrinski, H. A.; Wilson, J. A.; Shamji, A. F.; Wagner, B. K.; Koehler, A. N.; Schreiber, S. L. Small Molecules of Different Origins Have Distinct Distributions of Structural Complexity That Correlate with Protein-Binding Profiles. *Proc. Natl. Acad. Sci.* **2010**, *107* (44), 18787–18792. <https://doi.org/10.1073/pnas.1012741107>.
- (55) Schaeffer, L. Chapter 21 - The Role of Functional Groups in Drug–Receptor Interactions. In *The Practice of Medicinal Chemistry (Third Edition)*; Wermuth, C. G., Ed.; Academic Press: New York, 2008; pp 464–480. <https://doi.org/10.1016/B978-0-12-374194-3.00021-4>.
- (56) Rahman, I. U.; Afzal, A.; Iqbal, Z.; Ijaz, F.; Ali, N.; Shah, M.; Ullah, S.; Bussmann, R. W. Historical Perspectives of Ethnobotany. *Clin. Dermatol.* **2019**, *37* (4), 382–388. <https://doi.org/10.1016/j.clindermatol.2018.03.018>.
- (57) Wang, B.; Deng, J.; Gao, Y.; Zhu, L.; He, R.; Xu, Y. The Screening Toolbox of Bioactive Substances from Natural Products: A Review. *Fitoterapia* **2011**, *82* (8), 1141–1151. <https://doi.org/10.1016/j.fitote.2011.08.007>.
- (58) Zu, Y.; Wang, Y.; Fu, Y.; Li, S.; Sun, R.; Liu, W.; Luo, H. Enzyme-Assisted Extraction of Paclitaxel and Related Taxanes from Needles of *Taxus Chinensis*. *Sep. Purif. Technol.* **2009**, *68* (2), 238–243. <https://doi.org/10.1016/j.seppur.2009.05.009>.
- (59) Jeong, W. T.; Lim, H. B. A UPLC-ESI-Q-TOF Method for Rapid and Reliable Identification and Quantification of Major Indole Alkaloids in *Catharanthus Roseus*. *J. Chromatogr. B* **2018**, *1080*, 27–36. <https://doi.org/10.1016/j.jchromb.2018.02.018>.

- (60) Zhang, S.; Ye, T.; Liu, Y.; Hou, G.; Wang, Q.; Zhao, F.; Li, F.; Meng, Q. Research Advances in Clinical Applications, Anticancer Mechanism, Total Chemical Synthesis, Semi-Synthesis and Biosynthesis of Paclitaxel. *Molecules* **2023**, *28* (22), 7517. <https://doi.org/10.3390/molecules28227517>.
- (61) Norkunas, K.; Harding, R.; Dale, J.; Dugdale, B. Improving Agroinfiltration-Based Transient Gene Expression in *Nicotiana Benthamiana*. *Plant Methods* **2018**, *14* (1), 71. <https://doi.org/10.1186/s13007-018-0343-2>.
- (62) Saurabh Singh, K.; Hooft, J. J. J. van der; Wees, S. C. M. van; H. Medema, M. Integrative Omics Approaches for Biosynthetic Pathway Discovery in Plants. *Nat. Prod. Rep.* **2022**, *39* (9), 1876–1896. <https://doi.org/10.1039/D2NP00032F>.
- (63) Rodríguez-López, C. E.; Jiang, Y.; Kamileen, M. O.; Lichman, B. R.; Hong, B.; Vaillancourt, B.; Buell, C. R.; O'Connor, S. E. Phylogeny-Aware Chemoinformatic Analysis of Chemical Diversity in Lamiaceae Enables Iridoid Pathway Assembly and Discovery of Aucubin Synthase. *Mol. Biol. Evol.* **2022**, *39* (4), msac057. <https://doi.org/10.1093/molbev/msac057>.
- (64) Bryson, A. E.; Lanier, E. R.; Lau, K. H.; Hamilton, J. P.; Vaillancourt, B.; Mathieu, D.; Yocca, A. E.; Miller, G. P.; Edger, P. P.; Buell, C. R.; Hamberger, B. Uncovering a Miltiradiene Biosynthetic Gene Cluster in the Lamiaceae Reveals a Dynamic Evolutionary Trajectory. *Nat. Commun.* **2023**, *14* (1), 343. <https://doi.org/10.1038/s41467-023-35845-1>.
- (65) Godden, G. T.; Kinser, T. J.; Soltis, P. S.; Soltis, D. E. Phylotranscriptomic Analyses Reveal Asymmetrical Gene Duplication Dynamics and Signatures of Ancient Polyploidy in Mints. *Genome Biol. Evol.* **2019**, *11* (12), 3393–3408. <https://doi.org/10.1093/gbe/evz239>.
- (66) Lichman, B. R.; Godden, G. T.; Buell, C. R. Gene and Genome Duplications in the Evolution of Chemodiversity: Perspectives from Studies of Lamiaceae. *Curr. Opin. Plant Biol.* **2020**, *55*, 74–83. <https://doi.org/10.1016/j.pbi.2020.03.005>.
- (67) Hamilton, J. P.; Godden, G. T.; Lanier, E.; Bhat, W. W.; Kinser, T. J.; Vaillancourt, B.; Wang, H.; Wood, J. C.; Jiang, J.; Soltis, P. S.; Soltis, D. E.; Hamberger, B.; Buell, C. R. Generation of a Chromosome-Scale Genome Assembly of the Insect-Repellent Terpenoid-Producing Lamiaceae Species, *Callicarpa Americana*. *GigaScience* **2020**, *9* (9), giaa093. <https://doi.org/10.1093/gigascience/giaa093>.
- (68) Johnson, S. R.; Bhat, W. W.; Bibik, J.; Turmo, A.; Hamberger, B.; Consortium, E. M. G.; Hamberger, B. A Database-Driven Approach Identifies Additional Diterpene Synthase Activities in the Mint Family (Lamiaceae). *J. Biol. Chem.* **2019**, *294* (4), 1349–1362. <https://doi.org/10.1074/jbc.RA118.006025>.

- (69) Pelot, K. A.; Mitchell, R.; Kwon, M.; Hagelthorn, L. M.; Wardman, J. F.; Chiang, A.; Bohlmann, J.; Ro, D.-K.; Zerbe, P. Biosynthesis of the Psychotropic Plant Diterpene Salvinorin A: Discovery and Characterization of the *Salvia Divinorum* Clerodienyl Diphosphate Synthase. *Plant J.* **2017**, *89* (5), 885–897. <https://doi.org/10.1111/tpj.13427>.
- (70) Pelot, K. A.; Chen, R.; Hagelthorn, D. M.; Young, C. A.; Addison, J. B.; Muchlinski, A.; Tholl, D.; Zerbe, P. Functional Diversity of Diterpene Synthases in the Biofuel Crop Switchgrass. *Plant Physiol.* **2018**, *178* (1), 54–71. <https://doi.org/10.1104/pp.18.00590>.
- (71) Qiu, T.; Li, Y.; Wu, H.; Yang, H.; Peng, Z.; Du, Z.; Wu, Q.; Wang, H.; Shen, Y.; Huang, L. Tandem Duplication and Sub-Functionalization of Clerodane Diterpene Synthase Originate the Blooming of Clerodane Diterpenoids in *Scutellaria Barbata*. *Plant J.* **2023**, *116* (2), 375–388. <https://doi.org/10.1111/tpj.16377>.
- (72) Li, H.; Wu, S.; Lin, R.; Xiao, Y.; Morotti, A. L. M.; Wang, Y.; Galilee, M.; Qin, H.; Huang, T.; Zhao, Y.; Zhou, X.; Yang, J.; Zhao, Q.; Kanellis, A. K.; Martin, C.; Tatsis, E. C. The Genomes of Medicinal Skullcaps Reveal the Polyphyletic Origins of Clerodane Diterpene Biosynthesis in the Family Lamiaceae. *Mol. Plant* **2023**, *16* (3), 549–570. <https://doi.org/10.1016/j.molp.2023.01.006>.

Chapter 2: CYP76BK1 orthologs catalyze furan and lactone ring formation in clerodane diterpenoids across the mint family

This chapter is adapted from its original publication in *The Plant Journal*

Nicholas J. Schlecht , Emily R. Lanier¹, Trine B. Andersen, Julia Brose, Daniel Holmes, and Björn R. Hamberger CYP76BK1 orthologs catalyze furan and lactone ring formation in clerodane diterpenoids across the mint family, *The Plant Journal*, Volume 120, Issue 3, September 2024, <https://doi.org/10.1111/tpj.17031>

SUMMARY

The Lamiaceae (mint family) is the largest known source of furanoclerodanes, a subset of clerodane diterpenoids with broad bioactivities including insect antifeedant properties. The *Ajugoideae* subfamily, in particular, accumulates significant numbers of structurally related furanoclerodanes. The biosynthetic capacity for formation of these diterpenoids is retained across most Lamiaceae subfamilies, including the early-diverging *Callicarpoideae* which forms a sister clade to the rest of Lamiaceae. *VacCYP76BK1*, a cytochrome P450 monooxygenase from *Vitex agnus-castus*, was previously found to catalyze the formation of the proposed precursor to furan and lactone-containing labdane diterpenoids. Through transcriptome-guided pathway exploration, we identified orthologs of *VacCYP76BK1* in *Ajuga reptans* and *Callicarpa americana*. Functional characterization demonstrated that both could catalyze the oxidative cyclization of clerodane backbones to yield a furan ring. Subsequent investigation revealed a total of ten *CYP76BK1* orthologs across six Lamiaceae subfamilies. Through analysis of available chromosome-scale genomes, we identified four *CYP76BK1* members as syntelogs within a conserved syntenic block across divergent subfamilies. This suggests an evolutionary lineage that predates the speciation of the Lamiaceae. Functional characterization of the *CYP76BK1* orthologs affirmed conservation of function, as all catalyzed furan ring formation. Additionally, some orthologs yielded two novel lactone ring moieties. The presence of the *CYP76BK1* orthologs across Lamiaceae subfamilies closely overlaps with the distribution of reported furanoclerodanes. Together, the activities and distribution of the *CYP76BK1* orthologs identified here support their central role in furanoclerodane biosynthesis within the

Lamiaceae family. Our findings lay the groundwork for biotechnological applications to harness the economic potential of this promising class of compounds.

Significance Statement: The discovery and functional characterization of *CYP76BK1* orthologs across diverse Lamiaceae subfamilies revealed novel chemistry and their central role in furanoclerodane biosynthesis, providing insights into the metabolic landscape and dynamic evolution of this plant family over approximately 50 million years. These findings pave the way for targeted biosynthetic engineering efforts and the sustainable production of furanoclerodane compounds, offering promising prospects for agricultural and pharmaceutical applications.

INTRODUCTION

The past century has seen a tremendous explosion in the use of chemicals to treat biological threats, both in medicine and in agriculture. However, evolving challenges such as climate change and chemical resistance underscore a critical need for continued development of novel and effective biological control agents.^{1–3} Modern advances in genomics and metabolomics have enabled new opportunities to leverage plant natural products and their semi-synthetic derivatives to address these challenges.⁴

Diterpenoids are a structurally diverse class of specialized metabolites prevalent in plants. A limited set of diterpene backbones gives rise to thousands of unique oxidatively decorated diterpenoids with a broad spectrum of bioactivities. The modular nature of the biosynthetic route through action of diterpene synthases (diTPSs) and cytochrome P450 monooxygenases (P450s) is attractive from both a natural and bioengineering perspective, enabling rapid diversification of chemistries from just a few initial structures. From an evolutionary perspective, diterpenoids can serve as molecular

signatures across taxa, reflecting the history of plant families, speciation, and adaptation. They are key contributors to abiotic stress responses, microbial symbioses, developmental signaling, and defense against herbivores and pathogens.⁵ For humans, their economic importance is based on applications such as therapeutics, nutraceuticals, and in agriculture. One of the most prolific sources of diterpenoids are the Lamiaceae, also known as the mint family, recognized for highly aromatic plants. Species within this family contribute roughly 16% of the over 25,000 distinct diterpenoids reported in the Dictionary of Natural Products (DNP).⁶

Clerodanes, one of the most prevalent classes of diterpenoids, are of considerable economic interest due to their bioactivities. Within the Lamiaceae, numerous clerodane diterpenoids have been reported with a broad spectrum of activities including powerful opioid receptor agonists from the psychedelic plant *Salvia divinorum*, potent insect antifeedant and insecticidal compounds from the Scutellarioideae and Ajugoideae subfamilies, and MRSA-active antibacterial clerodanes in *Callicarpa americana*.^{7–9} Additional reported bioactivities of clerodanes include anti-cancer and antimicrobial therapeutics.¹⁰ There are seven common skeletal configurations of clerodanes with modifications of carbons 11-16 typically featuring either a furan or lactone moiety—collectively called “furanoclerodanes” (Figure 2.1).¹⁰ In rare cases, other heteroatoms are incorporated in sidechains or as pyrrolidine rings.¹¹ Based on available structure-activity relationship data, the widespread presence of furan and lactone moieties and their variations appear to be the main drivers of the biological activities of furanoclerodanes.^{7,10,12} While the furan and lactone sidechain modifications are often observed with clerodane backbones, other labdane backbones

with these same features (hereafter “furanolabdanes”) have also been reported from diverse plant taxa.^{10,13–16}

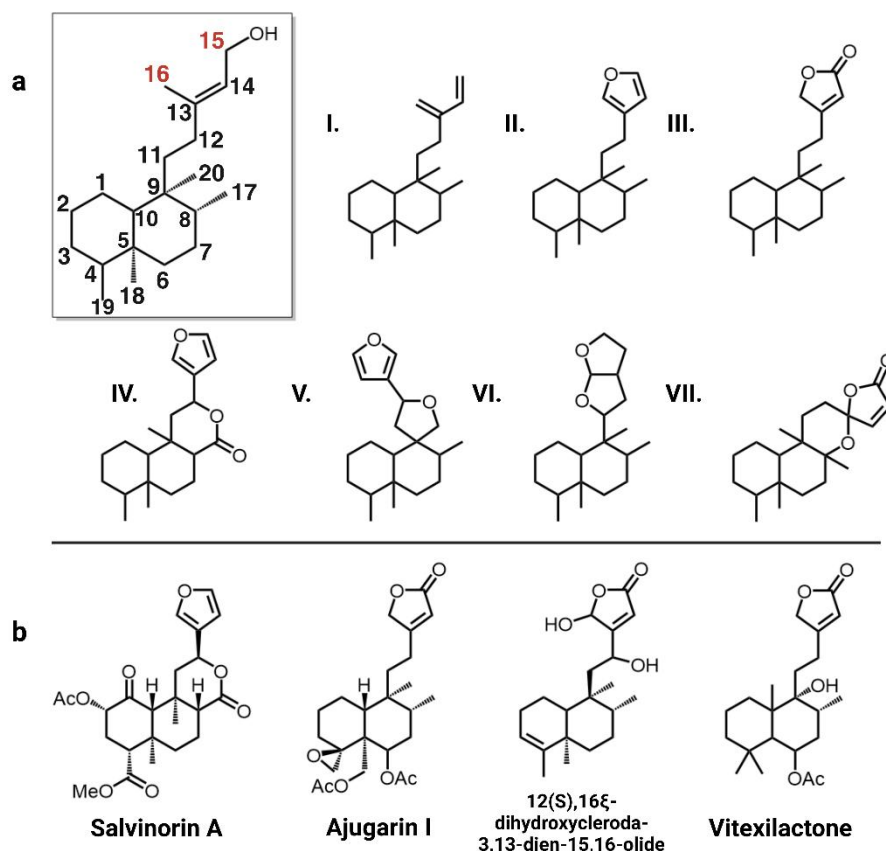


Figure 2.1. Structures of relevant clerodanes and furanoditerpenoids. A. Inset, neo-clerodane backbone with C15 and C16 highlighted for their role in furan and lactone cyclization and lactone carbonyl positioning. **I.-VII.**, seven types of clerodane backbones.¹⁰ B. Salvinorin A from *S. divinorum*; Ajugarin I from *A. reptans*; 12(S),16ξ-dihydroxycleroda-3,13-dien-15,16-olide is a furanoclerodane from *C. americana*; Vitexilactone, example of a furanolabdane from *V. agnus-castus*.^{8,17,18}

Recent investigations have begun to elucidate the enzymes involved in furanoclerodane biosynthesis in select species. Several diTPSs have been identified that form the key diterpene diphosphate precursors to various clerodane backbones. Within the Lamiaceae, it has been reported that *S. divinorum* (SdKPS1), *Vitex agnus-castus* (VacTPS5), *C. americana* (CamTPS2), *Scutellaria baicalensis* (SbdiTPS2.8) and

Salvia splendens (SspdiTPS2.1) all have (-)-kolavenyl diphosphate (KPP) synthases.^{19–}

²² The double bond isomer of KPP, isokolavenyl diphosphate (IKPP), is the major product of ArTPS2 from *Ajuga reptans* and two diTPSs from *Scutellaria* species (SbdiTPS2.7 and SbbdiTPS2.1, 2.3).^{22–24} More recently, a few P450s catalyzing formation of the furan moiety have been identified as well. In *S. divinorum*, CYP76AH39 was found to modify kolavenol (dephosphorylated KPP) with a dihydrofuran moiety.²⁵ In switchgrass (*Panicum virgatum*), several P450s in the monocot-specific CYP71Z family were shown to catalyze furan ring formation on both clerodane and labdane backbones.²⁶ In *V. agnus-castus*, CYP76BK1 was shown to hydroxylate C16 of the labdane peregrinol, which was suggested to be the first step in the formation of the furan ring apparent in vitexilactone (Figure 2.1).²⁰

In this work, we sought to further elucidate biosynthetic routes leading to bioactive furanoclerodanes within the Lamiaceae. We initially focused on two distinct and potentially economically significant compound classes: the Ajugarins, potent insect antifeedant furanoclerodanes produced by Ajugoideae species, and the MRSA-active furanoclerodane isolated from *C. americana* (12(S),16ξ-dihydroxycleroda-3,13-dien-15,16-olide).^{7,9} These pathways, originating from evolutionarily distinct subfamilies, provided a strategic foundation for discovering a broadly connected landscape of furanoclerodane and furanolabdane biosynthesis across the Lamiaceae.

RESULTS

Lamiaceae species produce most plant furanoclerodanes

To gain understanding of furanoclerodane distribution patterns across all organisms, the DNP was mined for reported diterpenoids. Compounds annotated as

labdanes or clerodanes bearing sidechain moieties characteristic of furanoclerodanes (i.e., structures II, III, and VI in Figure 2.1) were extracted. This analysis revealed that furanoclerodanes and furanolabdanes occur across 16 distinct clades across plants and some marine organisms (Figure 2.2). These include 10 dicotyledon families, 2 monocotyledon families, a family of ferns, a family of liverworts, and two different orders of demosponges. The majority of furanoclerodanes are found in the Lamiaceae family. Notably, there are over five times more furanoclerodanes than furanolabdanes across all families, revealing a strong bias towards the clerodane backbone.

Within Lamiaceae, furanoclerodane/labdanes are found in eight of eleven subfamilies but vary in the abundances of uniquely decorated structures (Figures S2.1 and S2). Callicarpoideae contains roughly a dozen unique furanoclerodanes while the Ajugoideae has several hundred. While the Nepetoideae subfamily contains approximately half of all Lamiaceae species, its *Salvia* genus is the only one with reported furanoclerodanes. The asymmetric diversity paired with presence of furanoclerodanes in evolutionarily distant subfamilies suggests there may be a basal, conserved metabolic pathway within the Lamiaceae.

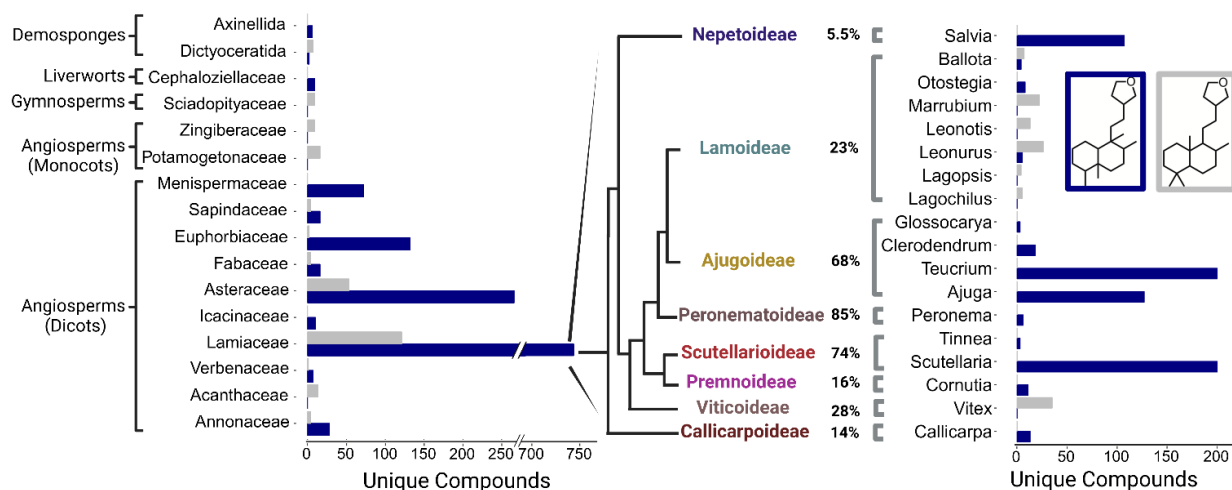


Figure 2.2. Furanoclerodane and furanolabdane abundances. Furanoclerodanes, blue; furanolabdanes, gray. Bars represent the number of unique compounds in the respective clade. The panel on the right shows the enrichment in the Lamiaceae. The percentages next to subfamilies compare the furanoclerodanes/furanolabdanes found in that subfamily to all recorded diterpenoids of that clade. The genus level bar plot was filtered to 3 and above compounds.

Identification of two P450s catalyzing production of furanoclerodanes in *A. reptans* and *C. americana*

To investigate whether a furanoclerodane biosynthetic pathway is conserved across Lamiaceae, we analyzed two taxonomically divergent species, *C. americana* from the early-diverging Callicarpoideae subfamily, as well as *A. reptans*, member of the Ajugoideae. Cytochrome P450 monooxygenases, particularly members of the CYP71 clan, can catalyze oxidative tailoring reactions in plant diterpenoid biosynthesis. Protein blast homology with a set of reference CYP71s was used to identify candidates from transcriptomic (*A. reptans*) and genomic (*C. americana*) data, yielding roughly 250 candidates. Candidate genes in *C. americana* were selected using tissue-specific expression data to correlate expression of candidates with that of the KPP synthase *CamTPS2* (Figure S2.3). In *A. reptans* only expression data from leaf tissue was

available, so candidates were chosen based on strength of expression and clustering with the CYP76 subfamily, which is prominent in Lamiaceae diterpenoid metabolism.²⁷ This approach yielded eight candidates from each species for functional characterization.

Candidate P450 transcripts were cloned and transiently expressed in *Nicotiana benthamiana* along with upstream terpene precursor genes D-xylulose-5-phosphate synthase (DXS) and geranylgeranyl diphosphate synthase (GGPPS) along with either CamTPS2 or ArTPS2. Consistent with previous studies, endogenous non-specific phosphatase activities were sufficient to provide the substrate for P450 oxidation^{19,23,25,26}. Most of the initial candidates did not convert the clerodane substrates. However, the orthologs *ArCYP76BK1* and *CamCYP76BK1* were found to catalyze formation of two new products (Figure S2.4), identified by NMR as 15,16-epoxy-4,18,(16),14-clerodatriene (**1**) and 15,16-epoxy-3,13(16),14-clerodatriene (**2**). These result from furan ring cyclization of the isokolavenol and kolavenol backbone, respectively (Figures S2.5-S2.8). These compounds are likely intermediates in the pathway toward the bioactive clerodanes (Figure 2.1). This transformation echoes the activity of previously identified CYP71Zs from switchgrass and *CYP76AH39* in *S. divinorum*.²⁵ *ArCYP76BK1* and *CamCYP76BK1* exhibit high protein sequence identity (71-74%) as well as functional similarity to *VacCYP76BK1*, despite approximately 50M years of evolutionary distance between the Callicarpoideae, Viticoideae, and Ajugoideae subfamilies.²⁸ This discovery prompted further investigation into the significance of *CYP76BK1* orthologs in furanoclerodane production across other Lamiaceae subfamilies.

Exploration of the CYP76BK family across the mint family

The set of 48 transcriptomes generated by the Mint Genome Project, which provides widespread coverage across subfamilies, was investigated for the presence of CYP76 family members (Figure 2.3a).²⁹ An additional transcriptome was generated for *Teucrium chamaedrys*. Phylogenetic analysis identified a monophyletic CYP76BK clade containing sequences from species in the Premnoideae, Ajudoideae, Peronematoideae, Scutellarioideae, Viticoideae, and Callicarpoideae (Figure S2.9, Figure 2.3b). This revealed an additional seven species with CYP76BK1 orthologs, named according to species *CpCYP76BK1* (*Cornutia pyramidata*), *PbCYP76BK1* (*Petraeovitex bambusetorum*), *HsCYP76BK1* (*Holmskioldia sanguinea*), *SbCYP76BK1* (*S. baicalensis*), *CbCYP76BK1* (*Clerodendrum bungei*), *Teucrium canadense* (*TcaCYP76BK1*) and *T. chamaedrys* (*TchCYP76BK1*).

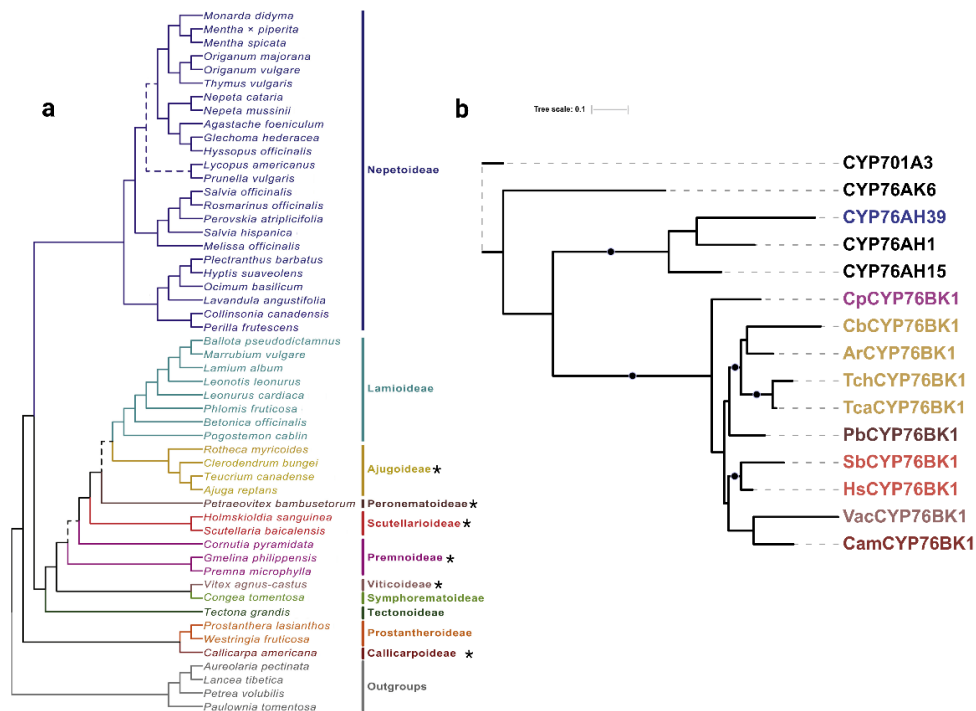


Figure 2.3. Identification of CYP76BK1 orthologs from the Lamiaceae family. (a) Phylogeny of the 48 Lamiaceae representatives screened for CYP76BK1 orthologs.

Figure 2.3 (cont'd)

(*) Indicates subfamilies with at least one CYP76BK1 ortholog (Figure adapted from Boachon et al., 2018.²⁹ (b) Maximum likelihood tree of CYP76BK1 orthologs along with reference sequences from the CYP76 family (in black). Black dots indicate a bootstrap value of 70% or greater (1000 bootstraps). CYP76AH39 is implicated in furanoclerodane biosynthesis in *Salvia*. Reference P450s can be found in Table S2.1.

Functional characterization of CYP76BK1 orthologs reveals alternative cyclization products

Newly identified *CYP76BK1* transcripts were cloned from leaf tissue cDNA and evaluated using transient expression in *N. benthamiana*. Each ortholog was co-expressed with a diTPS to assess activity with both kolavenol and isokolavenol (Figure 2.4). GC-MS analysis of the resulting enzyme products revealed that all CYP76BK1 enzymes exhibited similar activity as ArCYP76BK1 and CamCYP76BK1, generating peaks corresponding to retention times and fragmentation patterns of **1** and **2**. CamCYP76BK1 and ArCYP76BK1 product profiles revealed additional small products insufficient for purification and structural elucidation. These peaks were far more abundant with expression of other orthologs, allowing more complete analysis. Compounds **3**, **4**, and **5** were elucidated by NMR analysis (Figures S2.5-S2.8, S2.10-S2.15), revealing a C16 lactone moiety on **3** and **4** while **5** presents a C15 lactone arrangement. These were formally assigned as 4(18)-clerodadien-16,15-olide (**3**), 3,13-clerodadiene-16,15-olide (**4**) and 4(18)-clerodadien-15,16-olide (**5**). Compound **6** was tentatively characterized as 3,13-clerodadien-15,16-olide based on analogous mass fragmentation patterns and retention time to **5**.

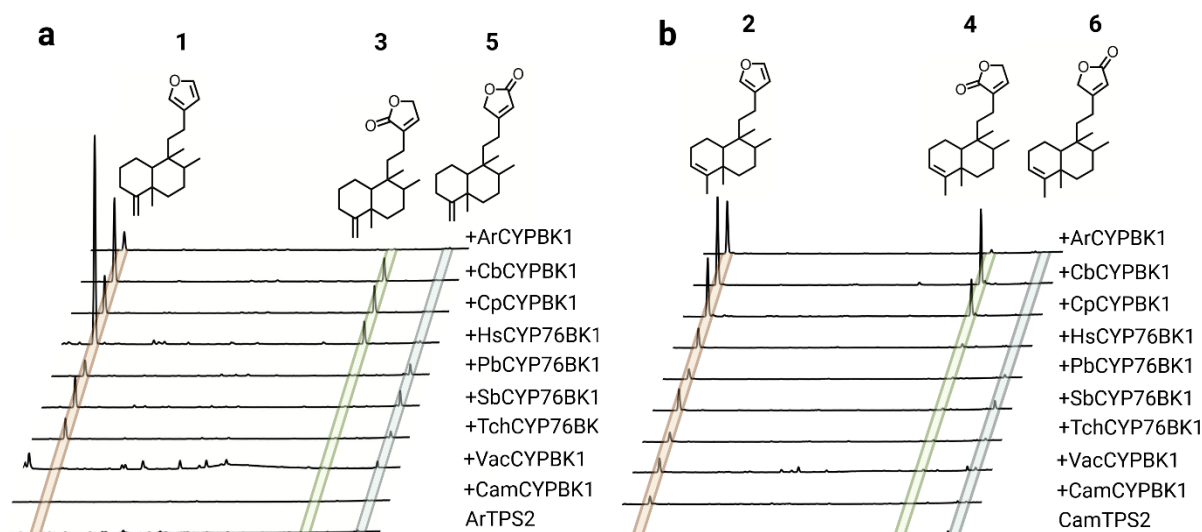


Figure 2.4. Functional characterization of CYP76BK1 orthologs. Total ion chromatograms with absolute intensities from extracts of *N. benthamiana* leaves transiently expressing each P450 candidate with the two different clerodane diTPS. Chromatograms are shown from minutes 12-18, offset by 0.1 minutes and are scaled to the highest peak in each set. Each infiltration also included a separate construct for DXS + GGPPS to increase yields. Mass spectra of compounds **1-6** can be found in Figures S16 and S17. (a) ArTPS2, an isokolavenol synthase, expressed with each ortholog. (b) CamTPS2, a kolavenol synthase, expressed with each ortholog. Compounds **1-5** have NMR data to support their structures (Figures S2.5-S2.8, S2.10-S2.15). The structure of compound **6** is proposed based on similar spectra and retention time to **5**.

VacCYP76BK1 was earlier shown to generate the C16 hydroxylation of peregrinol. This was suggested as precursor to the furanolabdanes rotundifuran and vitexilactone found in *V. agnus-castus* g. We co-expressed *VacCYP76BK1* and *CYP76BK1* orthologs with the peregrinol diphosphate synthase LITPS1 from *Leonotis leonurus*.²³ and could detect traces of products with mass spectra consistent with the furan and lactone derivatives of peregrinol (Figure S2.18), but products were not in sufficient quantities for NMR analysis. The putative peregrinol furan was also detectable with all orthologs except *CamCYP76BK1*, while the putative C15 and C16 lactones were only detectable with expression of *VacCYP76BK1*, *HsCYP76BK1*, *CbCYP76BK1*,

and *ArCYP76BK1*.

Notable differences in activity can be observed among the orthologs. HsCYP76BK1 afforded the best conversion of kolavenol to **1**, while CbCYP76BK1 was most active against isokolavenol with the best yield of **4** and SbCYP76BK1 afforded the highest yield of **6**. There is apparently some bias among orthologs in which lactone variation dominates the product profile, with CbCYP76BK1, CpCYP76BK1, and HsCYP76BK1 preferentially catalyzing formation of the C16 lactone while expression of PbCYP76BK1, SbCYP76BK1, and TchCYP76BK1 led to higher accumulation of the C15 lactone. Relative activities of *ArCYP76BK1* and *CamCYP76BK1* were more difficult to distinguish due to lower overall activity in lactone formation. Overall, there appears to be flexibility in the acceptance of both kolavenol and isokolavenol substrates (Figure S2.19). Neither *CamCYP76BK1* nor *VacCYP76BK1* appear capable of converting isokolavenol to either the furan or lactone products, but both show moderate activity with kolavenol.

Analysis of plant extracts

We analyzed leaf extracts of species with *CYP76BK1* orthologs by GC-MS for evidence of the furan and lactone containing intermediates. However, only the leaf extract of *C. americana* had peaks with the same mass spectrum and retention time as **2**, **4**, and **6** (Figure 2.5). Other extracts showed peaks which did not match the CYP76BK1 products but may represent further modified furanoclerodane products, based on mass fragmentation patterns (Figure S2.20).

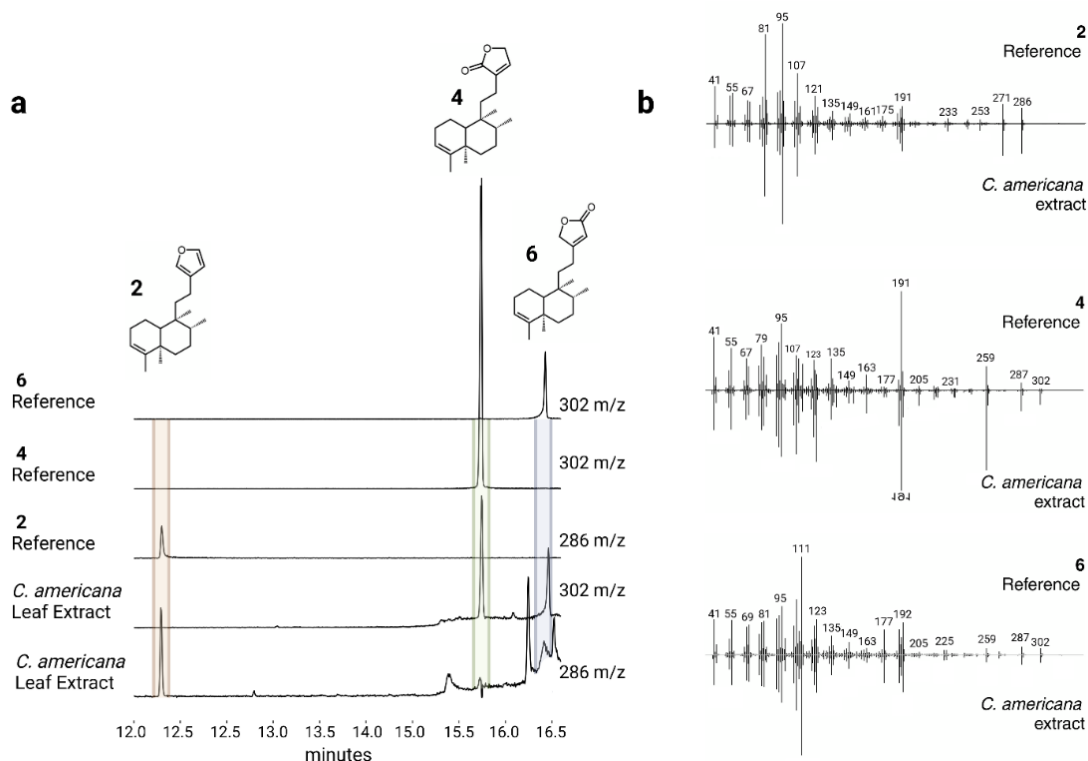


Figure 2.5. GC-MS analysis of plant extracts. (a) EIC showing the presence of the furan (m/z 286) and both lactone (m/z 302) moieties on the kolavenol backbone in the *C. americana* leaf extract. Purified products derived from *N. benthamiana* leaf extracts expressing either CamTPS2 and CbCYP76BK1 (**2** and **4**) or CamTPS2 and SbCYP76BK1 (**6**) were used for reference. (b) Deconvoluted mass spectra of **2**, **4**, and **6**. Top (black) from purified products, bottom (gray) corresponding peak from the *C. americana* leaf extract.

Evaluating syntenic orthologs of CYP76BKs

To further understand the context and evolutionary origin of the *CYP76BK1*s, we examined available chromosome-scale genomes to determine syntenic relationships of *CYP76BK1*s in the Lamiaceae family. The genomes of *C. americana*²¹, *Clerodendrum inerme*³⁰, *Pogostemon cablin*³¹, *Salvia hispanica*³², *Salvia miltiorrhiza*³³, *Salvia officinalis*³⁴, *Salvia splendens*³⁵, *S. baicalensis*³⁶, *Scutellaria barbata*³⁷, *Tectona grandis*³⁸, *Thymus quinquecostatus*³⁹, and *Lavandula angustifolia*⁴⁰ were all selected based on their quality and subfamily membership: one Callicarpoideae, one Ajugoideae,

one Lamioideae, two Scutellarioideae, one Tectonoideae, and six Nepetoideae.

Genomic investigations supported the transcriptomic analyses, as we found *CYP76BK1* members present in the Ajugoideae, Callicarpoideae, and Scutellarioideae subfamilies while absent in the Lamioideae, Nepetoideae, and Tectonoideae representatives. The *CYP76BK1*s in *C. americana*, *C. inerme*, *S. baicalensis*, and *S. barbata* were syntenic orthologs (syntelogs) (Figure 2.6). While the syntenic block is found throughout all the available Lamiaceae genomes, *CYP76BK1* homologs are not represented within the Lamioideae, Nepetoideae, and Tectonoideae.

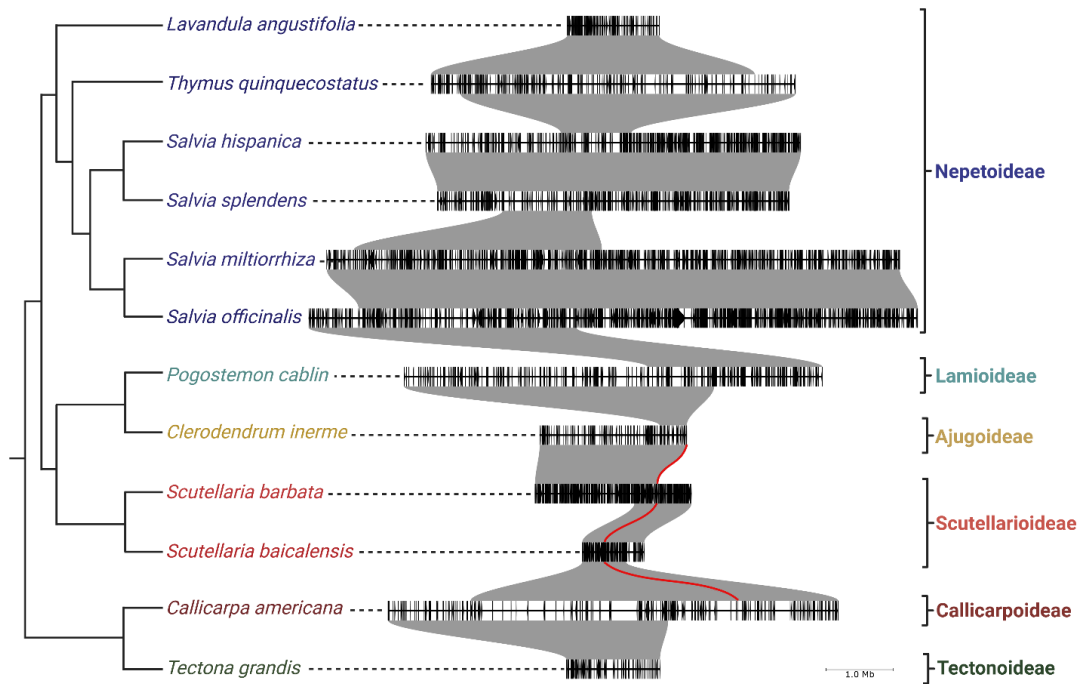


Figure 2.6. *CYP76BK1* Syntenic Block Through the Lamiaceae. Species tree phylogeny of representative species of subfamilies across the Lamiaceae generated in Brose et al. 2024.³² The red lines indicate syntenic orthologs of *CYP76BK1* present in *C. americana*, *S. baicalensis*, *S. barbata*, and *C. inerme*. Large syntenic blocks between the species are colored grey. Individual genes are colored black along the segments. The brackets highlight which syntenic blocks corresponding to each subfamily.

Discussion

Furanoclerodanes have garnered particular interest due to their potent insect-antifeedant, antimicrobial, and other bioactivities, which likely arise from characteristic furan and lactone ring systems. Our findings establish CYP76BK1 as a key enzyme in the biosynthesis of furanoclerodanes across many distinct subfamilies of the Lamiaceae family, opening new avenues for bioengineering and biosynthetic use of furanoclerodanes.

According to compounds reported in the DNP, furanoclerodanes are found in 16 families including some sessile marine organisms as well as liverworts, monocots, and dicots. Despite the widespread occurrence and diversity, approximately 40% of all reported furanoclerodanes are from species that carry a *CYP76BK1* ortholog. The density of unique furanoclerodanes in a narrow set of species, particularly in the Ajugoideae and Scutellarioideae, highlights a rapid expansion into this metabolic niche. Across the Lamiaceae family, the distribution of *CYP76BK1* orthologs significantly overlaps with the occurrence of reported furanoclerodanes. We found orthologs in the Scutellarioideae, Ajugoideae, Callicarpoideae, Premnoideae, Peronematoideae, and Viticoideae subfamilies, all of which have species with reported furanoclerodanes. *H. sanguinea* (*HsCYP76BK1*) and *P. bambusetorum* (*PbCYP76BK1*) have no reported diterpenoids. However, with only limited phytochemical studies available in these species^{41,42}, the presence of *CYP76BK1* could indicate a potential source for new furanoclerodane structures.

In contrast, only a small proportion of reported furanoclerodanes from the Lamiaceae come from subfamilies without a *CYP76BK1* ortholog. Most notably, *Salvia*

is the sole genus enriched in furanoclerodanes within the Nepetoideae. This can be explained by the presence of *CYP76AH39*, a dihydrofuran synthase from *S. divinorum* which has convergently evolved to perform a similar function to the *CYP76BK1* clade. We found in our phylogenetic analysis that putative *CYP76AH39* orthologs are limited to *Salvia* species (Figure S2.9). Moreover, recent work found through genomic and phylogenetic analysis of diterpene synthases at least two lineages leading to the appearance of clerodane biosynthesis in Lamiaceae, with the Nepetoideae lineage evolutionarily distinct from other subfamilies.²² Together, the evolution of clerodane diterpene synthases and furan/lactone yielding P450s highlight convergent evolution of these multi-step biosynthetic pathways. Outside of *Salvia*, the Lamioideae subfamily also contains a small number of furanoclerodanes despite lacking *CYP76BK1* orthologs in the transcriptomes we examined. It is plausible that the Lamioideae may have another enzyme or set of enzymes responsible for the furan ring formation. The multiple emergences of these compounds within and outside the Lamiaceae appears to underscore a selective advantage.^{10,43}

Genomic analysis provided further context for the evolutionary history of *CYP76BK1*. The syntelogs are found in a syntenic block conserved across all available chromosome-scale genomes in the Lamiaceae, however *CYP76BK1* was lost multiple times including in the Tectonoideae, the Lamioideae and Nepetoideae subfamilies. This suggests that *CYP76BK1* emerged prior to the divergence of the major Lamiaceae subfamilies, thus constituting a foundational part of the diterpenoid diversity in Lamiaceae. Subsequent loss can be attributed to localized genomic deletions rather than larger rearrangements. Analysis of the syntenic block with PlantiSmash⁴⁴ found no

additional terpene biosynthetic genes, in contrast to our previous discovery of an ancestral multiradiene biosynthetic gene cluster present throughout the Lamiaceae.⁴⁵

The functional capabilities of the *CYP76BK1* orthologs demonstrated here expand on the previous discovery of the founding member *VacCYP76BK1*. Initially, this ortholog was shown to catalyze hydroxylation of C16 of a labdane diterpenoid, peregrinol.²⁰ Utilizing GC-MS we found that this ortholog also can install both lactone and furan rings on clerodane and labdane substrates, albeit with low activity (Figure S2.18). The putative peregrinol furan was also detectable with all orthologs. We propose an updated mechanism for *CYP76BK1* that accounts for the formation of the C15 and C16 lactone rings as well as the originally proposed furan ring (Figure 2.7). While a single oxidation results in the hydroxylated C15 product, a second oxidation allows spontaneous formation of the furan. A third oxidation then enables formation of the two lactones, depending on the equilibrium state. This mechanism aligns with the product profiles we observed, where the furan is the dominant product. It can also explain how different orthologs might quickly evolve the ability to synthesize one lactone over another by favoring a specific intermediate for the final oxidation.

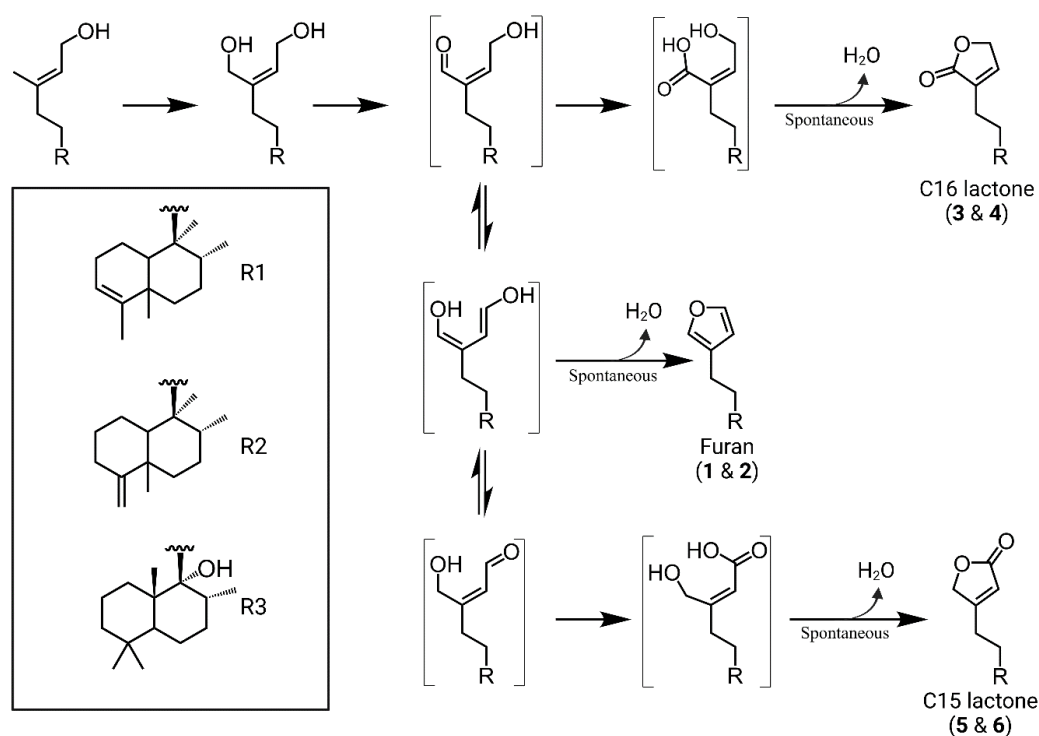


Figure 2.7. Proposed enzyme mechanism. The initial oxidation of C16 was observed with VacCYP76BK1.²⁰ Bracketed structures represent intermediates. Formation of a carbonyl group helps delocalize the electrons, making further rearrangements feasible and permitting an opportunity for cyclization. A nucleophilic attack and dehydration can generate the furan or a subsequent oxidation on either aldehyde intermediate followed by a nucleophilic attack and dehydration will yield the respective lactone ring, leading to either the C15 or C16 lactone. (Solid arrows indicate enzymatic oxidation and dashed arrows indicate spontaneous rearrangement)

Comparison of the product profiles observed here with reported metabolites suggests that additional enzymes beyond CYP76BK1 likely govern final product profile *in planta*. The transient expression of *CYP76BK1* orthologs generally showed the furan as the dominant product, followed by the C16 lactone or the C15 lactone. Most of the structures reported in the Ajugoideae, Scutellaroideae, and Callicarpoideae contain the C15 lactone or furan, as seen in the Ajugarins, the MRSA-active clerodane in *C. americana*, and vitexilactone from *V. agnus-castus*. Type III and type VIII clerodanes appear to be derived from this C15 lactone intermediate. In contrast, only a very few

reported structures contain the C16 lactone arrangement. However, the furanofuran or type VI clerodanes, which are highly abundant in *Clerodendrum*, could plausibly arise from the C16 lactone as an intermediate (Figure S2.1). The remaining furanoclerodane scaffolds, including types II, IV, and V appear to be derived from the furan intermediate. In metabolomic analysis of the plant extracts, we identified **2**, **4**, and **6** only in extracts from *C. americana*. This supports the biological relevance of these CYP76BK1 products, although it is interesting to note that CamCYP76BK1 itself had very low activity in the conversion to both **4** and **6**. Given the suggested multi-oxidation mechanism of CYP76BK1 enzymes, it is possible that other oxidases may be present *in planta* which are capable of steering product outcome towards the lactones. The lack of detectable CYP76BK1 products in the leaf extracts of other species may indicate that other tissues favor accumulation of these products, or that external stress is needed to initiate their biosynthesis. Alternatively, these compounds may represent intermediates in more complex pathways, potentially involving additional tailoring oxidases, shifting the product outcome towards more highly decorated species-specific diterpenoids.

The range of catalytic activities in the CYP76BK1 clade builds on recent discoveries of CYP76AH39 in *S. divinorum* and the CYP71Z clade in switchgrass.^{25,26} While CYP76AH39 can add a dihydrofuran to the kolavenol backbone, the CYP71Zs catalyze furan formation on multiple labdane and clerodane backbones. The CYP76BK1s further expand this biosynthetic toolbox with the ability to generate two lactone variants in addition to a furan on kolavenol, isokolavenol, and to a lesser degree some labdane substrates (Figure S2.18, Figure S2.21). Together these findings represent three distinct evolutionary trajectories towards furanoclerodane and

furanolabdane biosynthesis. The substrate promiscuity demonstrated here among kolavenol, isokolavenol, and peregrinol substrates further supports the shared origin of the CYP76BK1 orthologs. For some species, such as those in the Ajugoideae, the reported furanoclerodanes appear largely isokolavenol-derived. Since *A. reptans* has only an iso-KPP synthase, while others such as *C. americana* have only a KPP synthase, we can surmise that the availability of these substrates likely drives product outcome more than the enzyme selectivity.^{21,23}

A further observation from the DNP analysis is that across all plant families, furanoclerodanes outnumbered their labdane counterparts by over five-fold. Moreover, very few clerodanes are reported which lack a furan-containing moiety, while furanolabdanes comprise only a small proportion of all labdane structures reported. The pronounced skew towards furanoclerodanes may indicate selective pressures favoring the clerodane scaffold for furan and lactone ring installation, which would be consistent with the diverse bioactivities found in furanoclerodanes, most notably their insect antifeedants.^{7,46,47} We speculate that in addition to potential selective pressures from their corresponding bioactivities, the dominance of the furanoclerodanes could relate to the apparent lack of a clerodane-specific class I diTPS in most instances. Despite wide availability of transcriptomic and genomic resources, and numerous reports of class II diTPSs which form clerodienyl diphosphates, there is a general lack of corresponding class I diTPSs to facilitate diphosphate removal.^{19,21,23,48} It remains unclear why this pattern holds true across divergent species, but class I diTPSs typically introduce additional cyclizations in labdane pathways. In the absence of such a class I partner, plants may utilize alternative enzymes such as phosphatases or nudix hydrolases to

dephosphorylate the diphosphate backbone. After dephosphorylation, P450s may evolve to oxidize these uncyclized sidechains, spontaneously forming furan and lactone scaffolds which confer a biological advantage.

Collectively, our work provides a genomic, functional, and metabolomic perspective on the central role of CYP76BK1 in furanoclerodane metabolism across Lamiaceae. These enzymes catalyze key oxidative cyclizations en route to a vast array of bioactive diterpenoid natural products. The conservation and phylogenetic distribution pattern of CYP76BK1 implicates it as an ancestral enzyme lineage facilitating the proliferation of furanoclerodanes throughout the evolution of the Lamiaceae family. Our findings establish a robust molecular toolbox for targeted engineering of furanoclerodane biosynthetic pathways, enabling sustainable production of these high-value compounds for pharmaceutical and agricultural applications

Experimental procedures

Survey of diterpenoids from the DNP

The Dictionary of Natural Products (v30.2) was mined for relevant diterpenoids using the following search criteria. The search category 'Type of Compound' with either V.S.55000 or V.S.54000, which correspond to clerodane and labdane diterpenoids respectively. Multiple subsets of the data were extracted by including specific substructures to isolate particular categories of furanoclerodanes. The substructures consisted of the side chain of the C15/C16 lactones (with and without the C14-C15 double bond), furan, and furanofuran, the commonly found VI substructure in Figure 2.1. CSV files were semi-automatically extracted to include the following data: Chemical Name, Molecular formula, Accurate mass, Type of Compound, Type of Organism, and

Biological Source. A final control subset of data was included where the only search parameter was 'Type of Compound words' and the value was '*diterpen*' to extract all compounds annotated as diterpene or diterpenoid.

The CSVs were imported into R (v4.2.2) and tidyverse 2.0.0 was used for varying analyses. Each dataset had 2 new columns appended to represent their backbone (furanoclerodane, furanolabdane, clerodane, labdane, or other diterpenoid) as well as which modification they have (furan, C15 lactone, C16 lactone, furanofuran, and non-furanoditerpenoid). The files were then concatenated. Duplicate lines were removed along with compounds lacking biological source or type of organism data. Duplicate chemical names placed in 'other diterpenoid' were also removed.

The biological source data was manually curated to remove tissue related data due to its nonuniform categorizations. The 'Type of Organism' category was divided into Kingdom, Phylum, Order, Class, and Family. Due to some DNP entries predating their species' reclassification, various DNP entries were reclassified to contemporary clades, primarily moving various Verbenaceae to Lamiaceae. Data was grouped by either family or genus, with duplicate chemical name found within groups removed to ensure only one unique entry per group. The sum unique compounds were then plotted using different categorical compound descriptors respective of groupings. This includes comparing backbones in Lamiaceae (figure 2.2), which furano-moiety modifications are in Lamiaceae (Figure S2.1), and clerodane, labdane, furanoclerodanes/labdanes, and other diterpenoids that can be found in Lamiaceae (Figure S2.2).

Candidate gene selection

Previously assembled genomic and tissue-specific expression data²¹ were used to identify candidate genes in *C. americana*. The heatmap of candidate gene expression was generated using Heatmapper.⁴⁹ For all other species, previously assembled transcriptomic data were used (Table S2.2).⁵⁰ Candidate P450s (Table S2.3) are initially filtered to based on 45% identity and an E value greater than 1E-5 using BLASTP against a set of reference sequences (Table S2.1). For *A. reptans* the candidates were further narrowed down to CYP76s exclusively. Finding orthologs of the other CYP76BKs followed the same process of *A. reptans* and were then confirmed via phylogenetic relationships.

Phylogenetic trees

Reference sequences used in all protein phylogenies were obtained from GenBank (Table S2.2). Full-length peptide sequences were used. Multiple sequence alignments were generated using ClustalOmega (version 1.2.4; default parameters) and phylogenetic trees were generated by RAxML (version 8.2.12; Model = protgammaauto; Algorithm = a) with support from 1000 bootstrap replicates.^{51,52} Tree graphics were rendered using the Interactive Tree of Life (version 6.5.2).⁵³

Plant material and cloning

Plants were obtained from commercial nurseries or botanical gardens (Table S2.4) and grown in a greenhouse under ambient photoperiod and 24 °C day/17 °C night temperatures.

Synthetic oligonucleotides for all enzymes used in this study are given in Table S2.5. Candidate enzymes were PCR-amplified from leaf cDNA, and coding sequences

were cloned and sequence-verified with respective gene models. Constructs were then cloned into the plant expression vector pEAQ-HT and used in transient expression assays in *N. benthamiana*. One construct, VacCYP76BK1, was synthesized by Twist Bioscience before cloning into pEAQ-HT.

Transcriptomic sequencing and assembly of T. chamaedrys

100 mg young leaf tissue of young leaves were harvested and frozen in liquid nitrogen. RNA was isolated using Spectrum Plant Total RNA kit (Sigma) with on-column DNase digest. TruSeq stranded mRNA (polyA mRNA) libraries were constructed and sequenced on an Illumina Novaseq 6000 to 150 nt in paired-end mode. Sequencing was performed at the Research Technology Support Facility at Michigan State University.

Raw reads were evaluated with fastqc (v0.11.2)⁵⁴, trimmed and corrected using trimmomatic (v0.39)⁵⁵ and subsequently assembled using trinity (v 2.1.1).⁵⁶ Peptide sequences were predicted using Transdecoder (v. 5.5.0).⁵⁷ The predicted peptides were blasted against *CamCYP76BK1*, *ArCYP76BK1*, and *VacCYP76BK1*. *TchCYP76BK1* was found to be 94.7% identical to *TcaCYP76BK1* identified and included in producing the final phylogenies and for cloning.

Transient expression for functional characterization in N. benthamiana

N. benthamiana plants were grown for 5 weeks in a controlled growth room under 12 h light and 12 h dark cycle at (22°C) before infiltration. Constructs for co-expression were separately transformed into *Agrobacterium tumefaciens* strain LBA4404. 20 mL cultures were grown overnight at 30 °C in LB with 50 µg/mL kanamycin and 50 µg/mL rifampicin. Cultures were collected by centrifugation and washed twice

with 10 mL water. Cells were resuspended and diluted to an OD₆₀₀ of 1.0 in 200 µM acetosyringone/water and incubated at 30 °C for 1–2 h. Separate cultures were mixed in a 1:1 ratio for each combination of enzymes, and 4- or 5-week-old plants were infiltrated with a 1 mL syringe into the underside (abaxial side) of *N. benthamiana* leaves. All gene constructs were co-infiltrated with two genes encoding rate-limiting steps in the upstream (MEP) pathway: *P. barbatus* 1-deoxy-D-xylulose-5-phosphate synthase (*PbDXS*) and GGPP synthase (*PbGGPPS*) to boost production of the diterpene precursor GGPP.⁵⁸ Plants were returned to the controlled growth room (22 °C, 12 h diurnal cycle) for 5 days. Approximately 200 mg fresh weight from infiltrated leaves was extracted with 1 mL hexane overnight at room temperature. Plant material was collected by centrifugation, and the organic phase was removed for GC-MS analysis.

Plant extract metabolomics

Leaves from *C. pyramidata*, *P. bambusetorum*, *H. sanguinea*, *S. baicalensis*, *C. bungei*, *T. chamaedrys*, *A. reptans*, and *C. americana* were harvested for metabolite analysis. The leaves were frozen in liquid nitrogen, crushed, and extracted for three hours in ethyl acetate. Leaf material was collected by centrifugation and the organic phase was removed and concentrated for GC-MS analysis.

GC-MS analysis

All GC-MS analyses were performed in Michigan State University's Mass Spectrometry and Metabolomics Core Facility on an Agilent 7890 A GC with an Agilent VF-5ms column (30 m × 250 µm × 0.25 µm, with 10 m EZ-Guard) and an Agilent 5975 C detector. The inlet was set to 250 °C splitless injection of 1 µL and He carrier gas (1 mL/min), and the detector was activated following a 4 min solvent delay. All assays

and tissue analysis used the following method: temperature ramp start 40 °C, hold 1 min, 40 °C/min to 200 °C, hold 4.5 min, 20 °C/min to 240 °C, 10 °C/min to 280 °C, 40 °C/min to 320 °C, and hold 5 min. MS scan range was set to 40–400. Deconvolution of spectra was done utilizing AMDIS.⁵⁹

Product scale-up and NMR

For NMR analysis, production in the *N. benthamiana* system was scaled up to 1 L infection culture. A vacuum-infiltration system was used to infiltrate *A. tumefaciens* strains into whole *N. benthamiana* plants, with approximately 40 plants used for each enzyme combination. The furan and lactone derivatives of CamTPS2 were identified from the combination of CamTPS2 and CbCYP76BK1. The furan derivative of ArTPS2 was identified from the combination of ArTPS2 and ArCYP76BK1, while the C16 lactone derivative was identified from ArTPS2 with HsCYP76BK1 and the C15 lactone derivative utilized SbarbCYP76BK1. After 5 days, all leaf tissue was harvested and extracted overnight in 600 mL hexane at room temperature. The extract was concentrated by rotary evaporator. Each product was purified by silica gel flash column chromatography with a mobile phase of 98% hexane/2% ethyl acetate. NMR spectra were measured in Michigan State University's Max T. Rogers NMR Facility on a Bruker Avance NEO 800 MHz or 600 MHz spectrometer equipped with a helium cooled TCI cryoprobe or a Prodigy TCI cryoprobe, respectively, using CDCl₃ as the solvent. CDCl₃ peaks were referenced to 7.26 and 77.00 ppm for ¹H and ¹³C spectra, respectively.

Syntenic Analysis of Lamiaceae

Publicly available chromosome-scale genomes of *C. armericana*²¹, *Cleorodendrum inerme*³⁰, *Pogostemon cablin*³¹, *Salvia hispanica*³², *Salvia miltiorrhiza*³³, *Salvia officinalis*³⁴, *Salvia splendens*³⁵, *Scutellaria baicalensis*³⁶, *Scutellaria barbata*³⁷, *Tectona grandis*³⁸, *Thymus quinquecostatus*³⁹, and *Lavandula angustifolia*⁴⁰ were obtained and quality assessed using BUSCO (v5.5.0)⁶⁰ embryophyta_odb10. In order to identify syntenic orthologs, only chromosome-scale assemblies with BUSCO (Basic Universal Single Copy Orthologs) scores for the genome greater than 90% and annotation scores greater than 80% (Table S2.6). Syntelogs through the Lamiaceae were obtained for the chromosome scale assemblies within the Lamiaceae with GENESPACE (v.1.1.10).⁶¹ The regions were then visualized using pyGenomeViz.⁶² The phylogeny of the Lamiaceae species was generated by Brose et al. 2024.³² The final figure was edited in BioRender.

Accession numbers:

Relevant accession numbers can be found on Table S2.1 and Table S2.3

Data availability Statement

The raw sequence reads for the *Teucrium chamaedrys* transcriptome are available in the National Center for Biotechnology Information Sequence Read Archive under BioProject PRJNA1124528.

Author contributions

NS, EL, TA and BH conceived the study. NS and EL wrote the manuscript with contributions from JB and TA. Pathways and constructs were designed by NS and EL. Terpene analysis in transient assays was performed by NS and EL. JB performed all

genomic and syntenic analysis. DH assisted critically with NMR analysis.

Acknowledgements

We would like to thank Britta Hamberger for assistance in maintaining plant material. P450 annotation was kindly provided by David Nelson (University of Tennessee). We would like to thank Drs. Cassandra Johnny and Anthony Schillmiller of Michigan State University's Mass Spectrometry and Metabolomics Core Facility for their help in obtaining and interpreting GC-MS data, and the Max T. Rogers NMR Facility for their help in obtaining NMR data. We would also like to thank Dr. David Nelson for naming all CYP sequences presented in this work.

We collectively acknowledge that Michigan State University occupies the ancestral, traditional, and contemporary Lands of the Anishinaabeg – Three Fires Confederacy of Ojibwe, Odawa, and Potawatomi peoples. In particular, the University resides on Land ceded in the 1819 Treaty of Saginaw. We recognize, support, and advocate for the sovereignty of Michigan's twelve federally-recognized Indian nations, for historic Indigenous communities in Michigan, for Indigenous individuals and communities who live here now, and for those who were forcibly removed from their Homelands. By offering this Land Acknowledgement, we affirm Indigenous sovereignty and will work to hold Michigan State University more accountable to the needs of American Indian and Indigenous peoples.

Funding

This work was supported in part through computational resources and services provided by the Institute for Cyber-Enabled Research at Michigan State University, the US Department of Energy Great Lakes Bioenergy Research Center Cooperative

Agreement DE-SC0018409, startup funding from the Department of Biochemistry and Molecular Biology, and support from AgBioResearch (MICL02454). B.H. gratefully acknowledges a generous endowment from James K. Billman, Jr. N.S. is supported by a fellowship from Michigan State University under the predoctoral Training Program in Plant Biotechnology for Health and Sustainability (T32-GM110523) from the National Institute of General Medical Sciences of the National Institutes of Health, E.L. was supported by the NSF Graduate Research Fellowship Program (DGE-1848739). B.H. is in part supported by the National Science Foundation under Grant Number 1737898. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- (1) Garcia-Solache, M. A.; Casadevall, A. Global Warming Will Bring New Fungal Diseases for Mammals. *mBio* **2010**, 1 (1). <https://doi.org/10.1128/mbio.00061-10>.
- (2) Yi, H.; Devkota, B. R.; Yu, J. seung; Oh, K. cheol; Kim, J.; Kim, H. J. Effects of Global Warming on Mosquitoes & Mosquito-Borne Diseases and the New Strategies for Mosquito Control. *Entomol Res* **2014**, 44 (6), 215–235. <https://doi.org/10.1111/1748-5967.12084>.
- (3) MacGowan, A.; Macnaughton, E. Antibiotic Resistance. *Medicine* **2017**, 45 (10), 622–628. <https://doi.org/10.1016/J.MPMED.2017.07.006>.
- (4) Han, J.; Miller, E. P.; Li, S. Cutting-Edge Plant Natural Product Pathway Elucidation. *Curr Opin Biotechnol* **2024**, 87, 103137. <https://doi.org/10.1016/J.COPBIO.2024.103137>.
- (5) Cheng, A. X.; Lou, Y. G.; Mao, Y. B.; Lu, S.; Wang, L. J.; Chen, X. Y. Plant Terpenoids: Biosynthesis and Ecological Functions. *J Integr Plant Biol* **2007**, 49 (2), 179–186. <https://doi.org/10.1111/J.1744-7909.2007.00395.X>.
- (6) *Dictionary of Natural Products 31.1 Chemical Search*. <https://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml;jsessionid=76CBC23017A9120E74C646E51D14C422>.
- (7) Gebbinck, E. A. K.; Jansen, B. J. M. M.; De Groot, A.; Klein Gebbinck, E. A.; Jansen, B. J. M. M.; De Groot, A.; Gebbinck, E. A. K.; Jansen, B. J. M. M.; De Groot, A. Insect Antifeedant Activity of Clerodane Diterpenes and Related Model Compounds. *Phytochemistry* **2002**, 61 (7), 737–770. [https://doi.org/10.1016/S0031-9422\(02\)00174-7](https://doi.org/10.1016/S0031-9422(02)00174-7).
- (8) Roth, B. L.; Baner, K.; Westkaemper, R.; Siebert, D.; Rice, K. C.; Steinberg, S. A.; Ernsberger, P.; Rothman, R. B. Salvinorin A: A Potent Naturally Occurring Nonnitrogenous κ Opioid Selective Agonist. *Proc Natl Acad Sci U S A* **2002**, 99 (18), 11934–11939. <https://doi.org/10.1073/pnas.182234399>.
- (9) Dettweiler, M.; Melander, R. J.; Porras, G.; Risener, C.; Marquez, L.; Samarakoon, T.; Melander, C.; Quave, C. L. A Clerodane Diterpene from *Callicarpa Americana* Resensitizes Methicillin-Resistant *Staphylococcus Aureus* to β -Lactam Antibiotics. *ACS Infect Dis* **2020**, 6 (7), 1667–1673. <https://doi.org/10.1021/acsinfecdis.0c00307>.
- (10) Li, R.; Morris-Natschke, S. L.; Lee, K. H. Clerodane Diterpenes: Sources, Structures, and Biological Activities. *Natural Product Reports*. Royal Society of Chemistry October 2016, pp 1166–1226. <https://doi.org/10.1039/c5np00137d>.

- (11) Kobayashi, J.; Sekiguchi, M.; Shigemori, H.; Ohsaki, A. Echinophyllins A and B, Novel Nitrogen-Containing Clerodane Diterpenoids from *Echinodorus Macrophyllus*. *Tetrahedron Lett* **2000**, 41 (16), 2939–2943. [https://doi.org/10.1016/S0040-4039\(00\)00314-2](https://doi.org/10.1016/S0040-4039(00)00314-2).
- (12) Enriz, R. D.; Baldoni, H. A.; Zamora, M. A.; Jáuregui, E. A.; Sosa, M. E.; Tonn, C. E.; Luco, J. M.; Gordaliza, M. Structure-Antifeedant Activity Relationship of Clerodane Diterpenoids. Comparative Study with Withanolides and Azadirachtin. *J Agric Food Chem* **2000**, 48 (4), 1384–1392. <https://doi.org/10.1021/jf990006b>.
- (13) Kiuchi, F.; Matsuo, K.; Ito, M.; Qui, T. K.; Honda, G. New Norditerpenoids with Trypanocidal Activity from *Vitex Trifolia*. *Chem Pharm Bull (Tokyo)* **2004**, 52 (12), 1492–1494. <https://doi.org/10.1248/CPB.52.1492>.
- (14) Wu, H.; Li, J.; Fronczek, F. R.; Ferreira, D.; Burandt, C. L.; Setola, V.; Roth, B. L.; Zjawiony, J. K. Labdane Diterpenoids from *Leonotis Leonurus*. *Phytochemistry* **2013**, 91, 229–235. <https://doi.org/10.1016/J.PHYTOCHEM.2012.02.021>.
- (15) Wu, X. De; Wang, S. Y.; Wang, L.; He, J.; Li, G. T.; Ding, L. F.; Gong, X.; Dong, L. Bin; Song, L. D.; Li, Y.; Zhao, Q. S. Labdane Diterpenoids and Lignans from *Calocedrus Macrolepis*. *Fitoterapia* **2013**, 85 (1), 154–160. <https://doi.org/10.1016/J.FITOTE.2013.01.011>.
- (16) Zhou, M.; Li, T.; Zeng, C.; Pan, D. bo; Li, H. bo; Yu, Y. Two New Diterpenoids from the Rhizomes of *Zingiber Officinale*. *Nat Prod Res* **2023**, 37 (13), 2255–2262. <https://doi.org/10.1080/14786419.2022.2038595>.
- (17) Nishina, A.; Itagaki, M.; Sato, D.; Kimura, H.; Hirai, Y.; Phay, N.; Makishima, M. The Rosiglitazone-Like Effects of Vitexilactone, a Constituent from *Vitex Trifolia* L. in 3T3-L1 Preadipocytes. *Molecules* **2017**, Vol. 22, Page 2030 **2017**, 22 (11), 2030. <https://doi.org/10.3390/MOLECULES22112030>.
- (18) Khan, A.; Shal, B.; Khan, A. U.; Bibi, T.; Zeeshan, S.; Zahra, S. S.; Crews, P.; Haq, I. ul; Din, F. ud; Ali, H.; Khan, S. Suppression of MAPK/NF-KB and Activation of Nrf2 Signaling by Ajugarin-I in EAE Model of Multiple Sclerosis. *Phytotherapy Research* **2023**, 37 (6), 2326–2343. <https://doi.org/10.1002/PTR.7751>.
- (19) Pelot, K. A.; Mitchell, R.; Kwon, M.; Hagelthorn, D. M.; Wardman, J. F.; Chiang, A.; Bohlmann, J.; Ro, D. K.; Zerbe, P. Biosynthesis of the Psychotropic Plant Diterpene Salvinorin A: Discovery and Characterization of the *Salvia Divinorum* Clerodienyl Diphosphate Synthase. *The Plant Journal* **2017**, 89 (5), 885–897. <https://doi.org/10.1111/TPJ.13427>.

- (20) Heskes, A. M.; Sundram, T. C. M. M.; Boughton, B. A.; Jensen, N. B.; Hansen, N. L.; Crocoll, C.; Cozzi, F.; Rasmussen, S.; Hamberger, B. B.; Hamberger, B. B.; Staerk, D.; Møller, B. L.; Pateraki, I. Biosynthesis of Bioactive Diterpenoids in the Medicinal Plant *Vitex Agnus-Castus*. *Plant Journal* **2018**, 93 (5), 943–958. <https://doi.org/10.1111/tpj.13822>.
- (21) Hamilton, J. P.; Godden, G. T.; Lanier, E.; Bhat, W. W.; Kinser, T. J.; Vaillancourt, B.; Wang, H.; Wood, J. C.; Jiang, J.; Soltis, P. S.; Soltis, D. E.; Hamberger, B.; Robin Buell, C. Generation of a Chromosome-Scale Genome Assembly of the Insect-Repellent Terpenoid-Producing Lamiaceae Species, *Callicarpa Americana*. *Gigascience* **2020**, 9 (9), 1–11. <https://doi.org/10.1093/GIGASCIENCE/GIAA093>.
- (22) Li, H.; Wu, S.; Lin, R.; Xiao, Y.; Luisa Malaco Morotti, A.; Wang, Y.; Galilee, M.; Qin, H.; Huang, T.; Zhao, Y.; Zhou, X.; Yang, J.; Zhao, Q.; Kanellis, A. K.; Martin, C.; Tatsis, E. C.; Morotti, M. AL. The Genomes of Medicinal Skullcaps Reveal the Polyphyletic Origins of Clerodane Diterpene Biosynthesis in the Family Lamiaceae. **2023**. <https://doi.org/10.1016/j.molp.2023.01.006>.
- (23) Johnson, S. R.; Bhat, W. W.; Bibik, J.; Turmo, A.; Hamberger, B. B.; Consortium, E. M. G.; Hamberger, B. B. A Database-Driven Approach Identifies Additional Diterpene Synthase Activities in the Mint Family (Lamiaceae). *Journal of Biological Chemistry* **2019**, 294 (4), 1349–1362. <https://doi.org/10.1074/jbc.RA118.006025>.
- (24) Qiu, T.; Li, Y. Y.; Wu, H.; Yang, H.; Peng, Z.; Du, Z.; Wu, Q.; Wang, H.; Shen, Y.; Huang, L. Tandem Duplication and Sub-Functionalization of Clerodane Diterpene Synthase Originate the Blooming of Clerodane Diterpenoids in *Scutellaria Barbata*. *The Plant Journal* **2023**, 116 (2), 375–388. <https://doi.org/10.1111/TPJ.16377>.
- (25) Kwon, M.; Utomo, J. C.; Park, K.; Pascoe, C. A.; Chiorean, S.; Ngo, I.; Pelot, K. A.; Pan, C. H.; Kim, S. W.; Zerbe, P.; Vederas, J. C.; Ro, D. K. Cytochrome P450-Catalyzed Biosynthesis of a Dihydrofuran Neoclerodane in Magic Mint (*Salvia Divinorum*). *ACS Catal* **2021**, 12 (1), 777–782. <https://doi.org/10.1021/acscatal.1c03691>.
- (26) Muchlinski, A.; Jia, M.; Tiedge, K.; Fell, J. S.; Pelot, K. A.; Chew, L.; Davisson, D.; Chen, Y.; Siegel, J.; Lovell, J. T.; Zerbe, P. Cytochrome P450-Catalyzed Biosynthesis of Furanoditerpenoids in the Bioenergy Crop Switchgrass (*Panicum Virgatum* L.). *The Plant Journal* **2021**, 108 (4), 1053–1068. <https://doi.org/10.1111/TPJ.15492>.
- (27) Bathe, U.; Tissier, A. Cytochrome P450 Enzymes: A Driving Force of Plant Diterpene Diversity. *Phytochemistry*. Elsevier Ltd May 2019, pp 149–162. <https://doi.org/10.1016/j.phytochem.2018.12.003>.

- (28) Kumar, S.; Suleski, M.; Craig, J. M.; Kasprowicz, A. E.; Sanderford, M.; Li, M.; Stecher, G.; Hedges, S. B. TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol Biol Evol* **2022**, 39 (8). <https://doi.org/10.1093/MOLBEV/MSAC174>.
- (29) Boachon, B.; Buell, C. R.; Crisovan, E.; Dudareva, N.; Garcia, N.; Godden, G.; Henry, L.; Kamileen, M. O.; Kates, H. R.; Kilgore, M. B.; Lichman, B. R.; Mavrodiev, E. V.; Newton, L.; Rodriguez-Lopez, C.; O'Connor, S. E.; Soltis, D.; Soltis, P.; Vaillancourt, B.; Wiegert-Rininger, K.; Zhao, D. Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. *Mol Plant* **2018**, 11 (8), 1084–1096. <https://doi.org/10.1016/j.molp.2018.06.002>.
- (30) He, Z.; Feng, X.; Chen, Q.; Li, L.; Li, S.; Han, K.; Guo, Z.; Wang, J.; Liu, M.; Shi, C.; Xu, S.; Shao, S.; Liu, X.; Mao, X.; Xie, W.; Wang, X.; Zhang, R.; Li, G.; Wu, W.; Zheng, Z.; Zhong, C.; Duke, N. C.; Boufford, D. E.; Fan, G.; Wu, C. I.; Ricklefs, R. E.; Shi, S. Evolution of Coastal Forests Based on a Full Set of Mangrove Genomes. *Nat Ecol Evol* **2022**, 6 (6), 738–749. <https://doi.org/10.1038/s41559-022-01744-9>.
- (31) Shen, Y.; Li, W.; Zeng, Y.; Li, Z.; Chen, Y.; Zhang, J.; Zhao, H.; Feng, L.; Ma, D.; Mo, X.; Ouyang, P.; Huang, L.; Wang, Z.; Jiao, Y.; Wang, H. bin. Chromosome-Level and Haplotype-Resolved Genome Provides Insight into the Tetraploid Hybrid Origin of Patchouli. *Nat Commun* **2022**, 13 (1), 1–15. <https://doi.org/10.1038/s41467-022-31121-w>.
- (32) Brose, J.; Hamilton, J. P.; Schlecht, N.; Zhao, D.; Mejía-Ponce, P. M.; Pérez, A. C.; Vaillancourt, B.; Wood, J. C.; Edger, P. P.; Montes-Hernandez, S.; Rosas, G. O. de; Hamberger, B.; Jaramillo, A. C.; Buell, C. R. Chromosome-Scale *Salvia Hispanica* L. (Chia) Genome Assembly Reveals Rampant *Salvia* Interspecies Introgression. *Plant Genome* **2024**, e20494. <https://doi.org/10.1002/TPG2.20494>.
- (33) Pan, X.; Chang, Y.; Li, C.; Qiu, X.; Cui, X.; Meng, F.; Zhang, S.; Li, X.; Lu, S. Chromosome-Level Genome Assembly of *Salvia Miltiorrhiza* with Orange Roots Uncovers the Role of Sm2OGD3 in Catalyzing 15,16-Dehydrogenation of Tanshinones. *Hortic Res* **2023**, 10 (6). <https://doi.org/10.1093/HR/UHAD069>.
- (34) Li, C. Y.; Yang, L.; Liu, Y.; Xu, Z. G.; Gao, J.; Huang, Y. B.; Xu, J. J.; Fan, H.; Kong, Y.; Wei, Y. K.; Hu, W. L.; Wang, L. J.; Zhao, Q.; Hu, Y. H.; Zhang, Y. J.; Martin, C.; Chen, X. Y. The Sage Genome Provides Insight into the Evolutionary Dynamics of Diterpene Biosynthesis Gene Cluster in Plants. *Cell Rep* **2022**, 40 (7), 111236. <https://doi.org/10.1016/j.celrep.2022.111236>.
- (35) Dong, A. X.; Xin, H. B.; Li, Z. J.; Liu, H.; Sun, Y. Q.; Nie, S.; Zhao, Z. N.; Cui, R. F.; Zhang, R. G.; Yun, Q. Z.; Wang, X. N.; Maghuly, F.; Porth, I.; Cong, R. C.; Mao, J. F. High-Quality Assembly of the Reference Genome for Scarlet Sage, *Salvia Splendens*, an Economically Important Ornamental Plant. *Gigascience* **2018**, 7 (7), 1–10. <https://doi.org/10.1093/gigascience/giy068>.

- (36) Zhao, Q.; Yang, J.; Cui, M. Y.; Liu, J.; Fang, Y.; Yan, M.; Qiu, W.; Shang, H.; Xu, Z.; Yidiresi, R.; Weng, J. K.; Pluskal, T.; Vigouroux, M.; Steuernagel, B.; Wei, Y.; Yang, L.; Hu, Y.; Chen, X. Y.; Martin, C. The Reference Genome Sequence of *Scutellaria Baicalensis* Provides Insights into the Evolution of Wogonin Biosynthesis. *Mol Plant* **2019**, *12* (7), 935–950. <https://doi.org/10.1016/J.MOLP.2019.04.002>.
- (37) Xu, Z.; Gao, R.; Pu, X.; Xu, R.; Wang, J.; Zheng, S.; Zeng, Y.; Chen, J.; He, C.; Song, J. Comparative Genome Analysis of *Scutellaria Baicalensis* and *Scutellaria Barbata* Reveals the Evolution of Active Flavonoid Biosynthesis. *Genomics Proteomics Bioinformatics* **2020**, *18* (3), 230–240. <https://doi.org/10.1016/j.gpb.2020.06.002>.
- (38) Zhao, D.; Hamilton, J. P.; Bhat, W. W.; Johnson, S. R.; Godden, G. T.; Kinser, T. J.; Boachon, B.; Dudareva, N.; Soltis, D. E.; Soltis, P. S.; Hamberger, B.; Robin Buell, C. A Chromosomal-Scale Genome Assembly of *Tectona Grandis* Reveals the Importance of Tandem Gene Duplication and Enables Discovery of Genes in Natural Product Biosynthetic Pathways. *Gigascience* **2019**, *8* (3), 1–10. <https://doi.org/10.1093/gigascience/giz005>.
- (39) Sun, M.; Zhang, Y.; Zhu, L.; Liu, N.; Bai, H.; Sun, G.; Zhang, J.; Shi, L. Chromosome-Level Assembly and Analysis of the *Thymus* Genome Provide Insights into Glandular Secretory Trichome Formation and Monoterpenoid Biosynthesis in Thyme. *Plant Commun* **2022**, *3* (6), 100413. <https://doi.org/10.1016/j.xplc.2022.100413>.
- (40) Hamilton, J. P.; Vaillancourt, B.; Wood, J. C.; Wang, H.; Jiang, J.; Soltis, D. E.; Buell, C. R.; Soltis, P. S. Chromosome-Scale Genome Assembly of the ‘Munstead’ Cultivar of *Lavandula Angustifolia*. *BMC Genom Data* **2023**, *24* (1), 1–6. <https://doi.org/10.1186/s12863-023-01181-y>.
- (41) Helfrich, E.; Rimpler, H. Iridoid Glycosides and Phenolic Glycosides from *Holmskioldia Sanguinea*. *Phytochemistry* **1999**, *50* (4), 619–627. [https://doi.org/10.1016/S0031-9422\(98\)00559-7](https://doi.org/10.1016/S0031-9422(98)00559-7).
- (42) Chaudhuri, P. K.; Srivastava, R.; Kumar, S.; Kumar, S. Phytotoxic and Antimicrobial Constituents of *Bacopa Monnieri* and *Holmskioldia Sanguinea*. *Phytotherapy Research* **2004**, *18* (2), 114–117. <https://doi.org/10.1002/PTR.1278>.
- (43) Tiedge, K.; Li, X.; Merrill, A. T.; Davisson, D.; Chen, Y.; Yu, P.; Tantillo, D. J.; Last, R. L.; Zerbe, P. Comparative Transcriptomics and Metabolomics Reveal Specialized Metabolite Drought Stress Responses in Switchgrass (*Panicum Virgatum*). *New Phytologist* **2022**, *236* (4), 1393–1408. <https://doi.org/10.1111/NPH.18443>.
- (44) Kautsar, S. A.; Suarez Duran, H. G.; Blin, K.; Osbourn, A.; Medema, M. H. PlantSMASH: Automated Identification, Annotation and Expression Analysis of Plant Biosynthetic Gene Clusters. *Nucleic Acids Res* **2017**, *45* (W1), W55–W63. <https://doi.org/10.1093/NAR/GKX305>.

- (45) Bryson, A. E.; Lanier, E. R.; Lau, K. H.; Hamilton, J. P.; Vaillancourt, B.; Mathieu, D.; Yocca, A. E.; Miller, G. P.; Edger, P. P.; Buell, C. R.; Hamberger, B. Uncovering a Miltiradiene Biosynthetic Gene Cluster in the Lamiaceae Reveals a Dynamic Evolutionary Trajectory. *Nature Communications* **2023** *14*:1 **2023**, *14* (1), 1–14. <https://doi.org/10.1038/s41467-023-35845-1>.
- (46) Sosa, M. E.; Tonn, C. E.; Giordano, O. S. INSECT ANTIFEEDANT ACTIVITY OF CLERODANE DITERPENOID. *J Nat Prod* **1994**, *57* (9), 2022. <https://doi.org/10.1021/np50111a012>
- (47) Koul, O. Phytochemicals and Insect Control: An Antifeedant Approach. *CRC Crit Rev Plant Sci* **2008**, *27* (1), 1–24. <https://doi.org/10.1080/07352680802053908>.
- (48) Chen, X.; Berim, A.; Dayan, F. E.; Gang, D. R. A (-)-Kolavenyl Diphosphate Synthase Catalyzes the First Step of Salvinorin A Biosynthesis in *Salvia Divinorum*. *J Exp Bot* **2017**, *68* (5), 1109–1122. <https://doi.org/10.1093/jxb/erw493>.
- (49) Babicki, S.; Arndt, D.; Marcu, A.; Liang, Y.; Grant, J. R.; Maciejewski, A.; Wishart, D. S. Heatmapper: Web-Enabled Heat Mapping for All. *Nucleic Acids Res* **2016**, *44* (W1), W147–W153. <https://doi.org/10.1093/NAR/GKW419>.
- (50) Evolutionary Genomics Consortium, M.; Boachon, B.; Robin Buell, C.; Crisovan, E.; Dudareva, N.; Garcia, N.; Godden, G.; Henry, L.; Kamileen, M. O.; Rose Kates, H.; Kilgore, M. B.; Lichman, B. R.; Mavrodiev, E. V.; Newton, L.; Rodriguez-Lopez, C.; O, S. E.; Soltis, D.; Soltis, P.; Vaillancourt, B.; Wiegert-Rininger, K.; Zhao, D. Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae Mint Evolutionary Genomics Consortium*. *Mol Plant* **2018**, *11*, 1084–1096. <https://doi.org/10.1016/j.molp.2018.06.002>.
- (51) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol Syst Biol* **2011**, *7*, 539. <https://doi.org/10.1038/msb.2011.75>.
- (52) Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30* (9), 1312–1313. <https://doi.org/10.1093/BIOINFORMATICS/BTU033>.
- (53) Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v6: Recent Updates to the Phylogenetic Tree Display and Annotation Tool. *Nucleic Acids Res* **2024**, *2024*, 1–5. <https://doi.org/10.1093/NAR/GKAE268>.
- (54) Andrews, S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 2024-05-18).

- (55) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. <https://doi.org/10.1093/BIOINFORMATICS/BTU170>.
- (56) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; Di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A. Trinity: Reconstructing a Full-Length Transcriptome without a Genome from RNA-Seq Data. *Nat Biotechnol* **2011**, *29* (7), 644. <https://doi.org/10.1038/NBT.1883>.
- (57) Haas, B. *Home · TransDecoder/TransDecoder Wiki · GitHub*. <https://github.com/TransDecoder/TransDecoder/wiki> (accessed 2024-05-18).
- (58) Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Møller, B. L.; Hamberger, B. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angewandte Chemie* **2016**, *128* (6), 2182–2186. <https://doi.org/10.1002/ANGE.201510650>.
- (59) NIST. *Automated Mass Spectral Deconvolution and Identification System* [Software]. National Institute of Standards and Technology. <https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdis> (accessed 2024-08-22).
- (60) Simao, F. A.; Waterhouse, R. M.; Ioannidis, P.; Kriventseva, E. V.; Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **2015**, *31* (19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- (61) Lovell, J. T.; Sreedasyam, A.; Schranz, M. E.; Wilson, M.; Carlson, J. W.; Harkess, A.; Emms, D.; Goodstein, D. M.; Schmutz, J. GENESPACE Tracks Regions of Interest and Gene Copy Number Variation across Multiple Genomes. *Elife* **2022**, *11*, 1–20. <https://doi.org/10.7554/ELIFE.78526>.
- (62) Shimoyama, Y. PyGenomeViz: A Genome Visualization Python Package for Comparative Genomics. 2022. <https://github.com/moshi4/pyGenomeViz>

APPENDIX

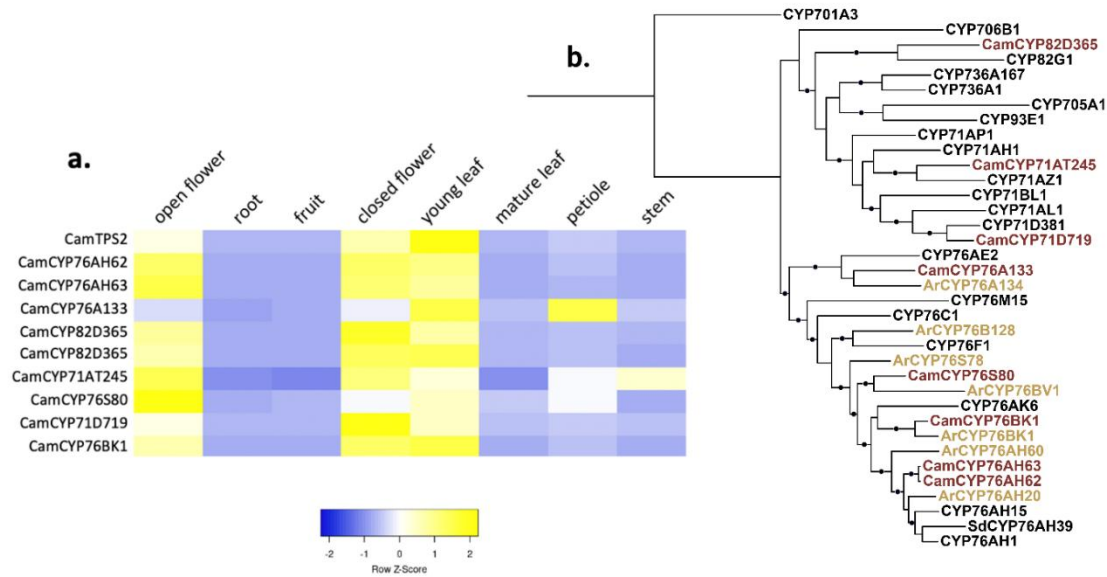


Figure S2.1. P450 candidates selected for functional characterization. (a) Heatmap of *C. americana* candidates showing comparison of expression profile with CamTPS2, the KPP synthase. (b) Maximum-likelihood tree of all cloned candidates from *C. americana* (Cam) and *A. reptans* (Ar). Black dots indicate bootstrap support of 70% or greater (1000 replicates). Sequence names in black are previously reported sequences for reference. The yellow highlights *A. reptans* genes and red highlights *C. americana* genes. Table S2.2 contains the accession numbers for the candidate and references.

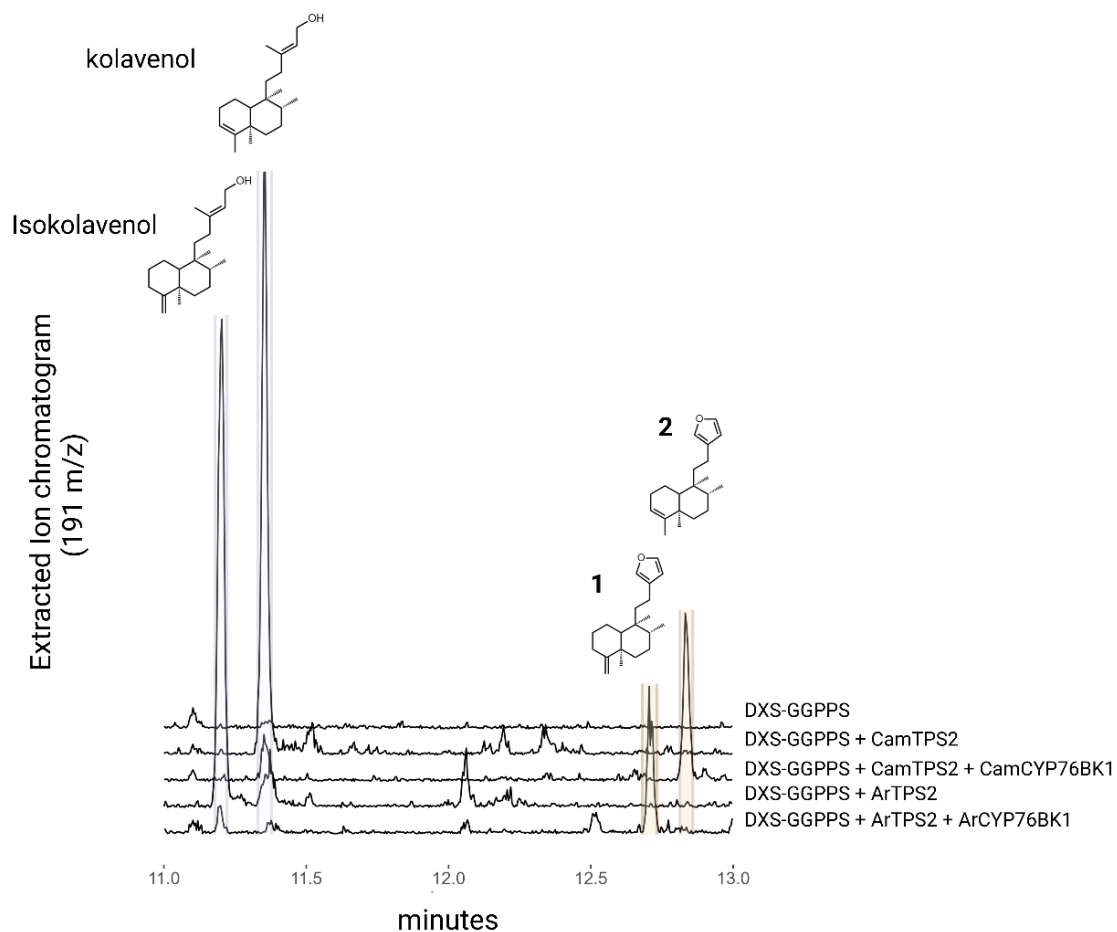


Figure S2.2. *C. americana* and *A. reptans* CYP76BK1 ortholog activity. A 191 m/z extracted ion chromatogram of DXS-GGPPS control as well as the kolavenol and isokolavenol synthases from *C. americana* and *A. reptans* with and without their respective CYP76BK1 ortholog.

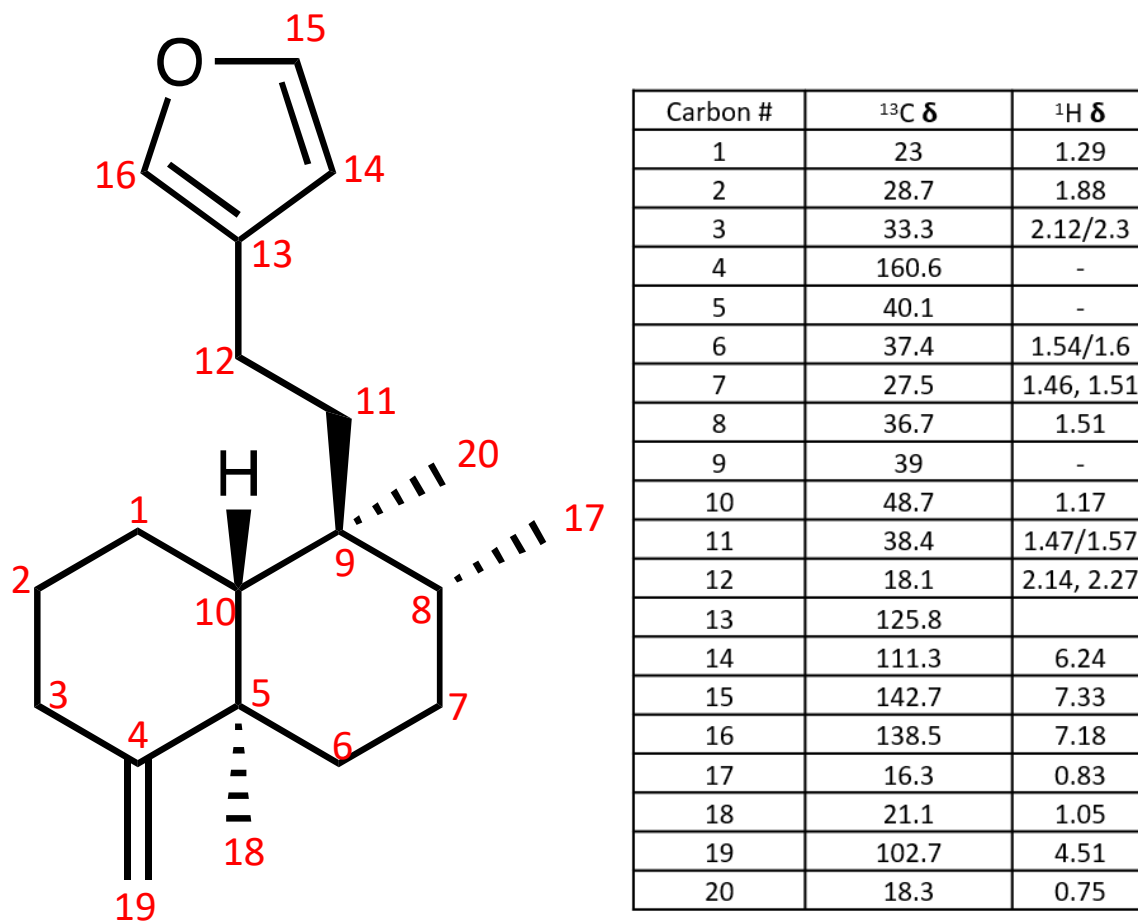


Figure S2.3. NMR chemical shift assignments of compound 1. Connectivity was deduced from ^1H , ^{13}C , HSQC, HMBC, and COSY correlations (Fig. S2.4). CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry was assigned based on the configuration of the precursor isokolavenyl diphosphate.

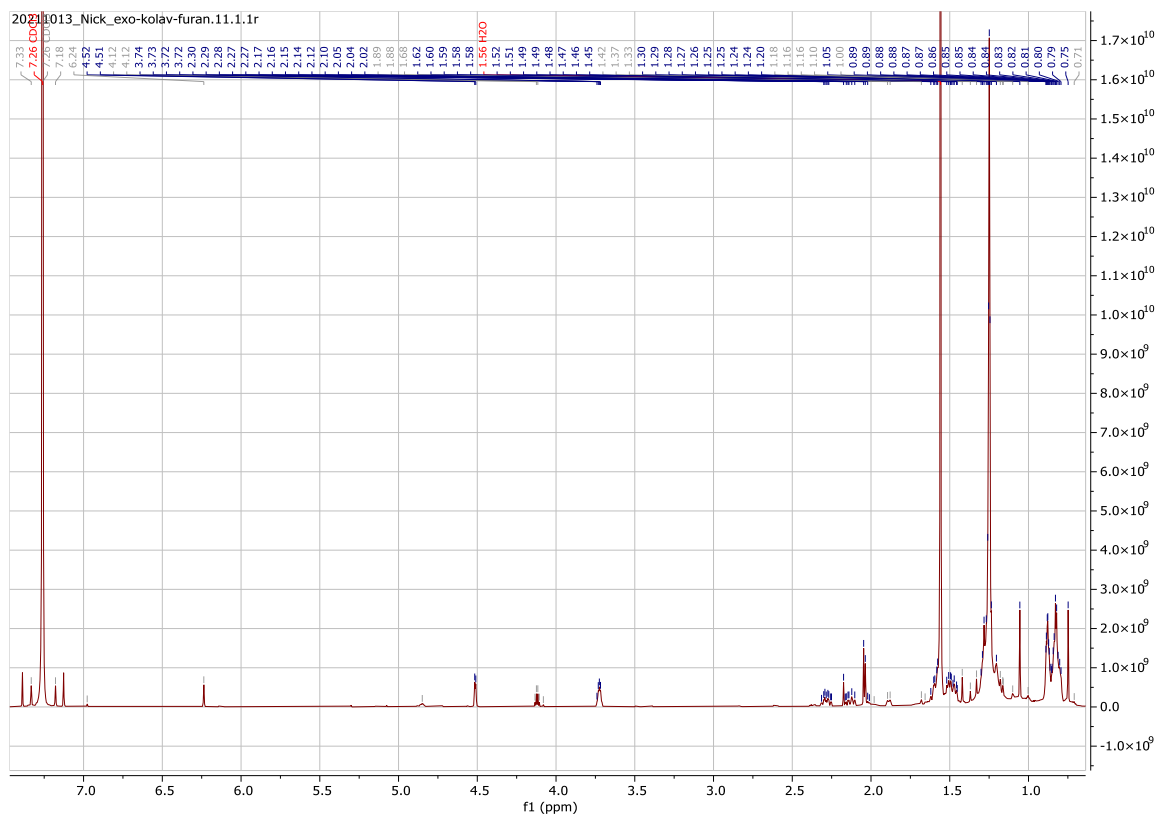


Figure S2.4a. NMR spectra of Compound 1. CDCl₃ was used as the solvent, and CDCl₃ peaks were referenced to 7.26 and 77.00 ppm for ¹H and ¹³C spectra, respectively. (a) ¹H NMR spectrum.

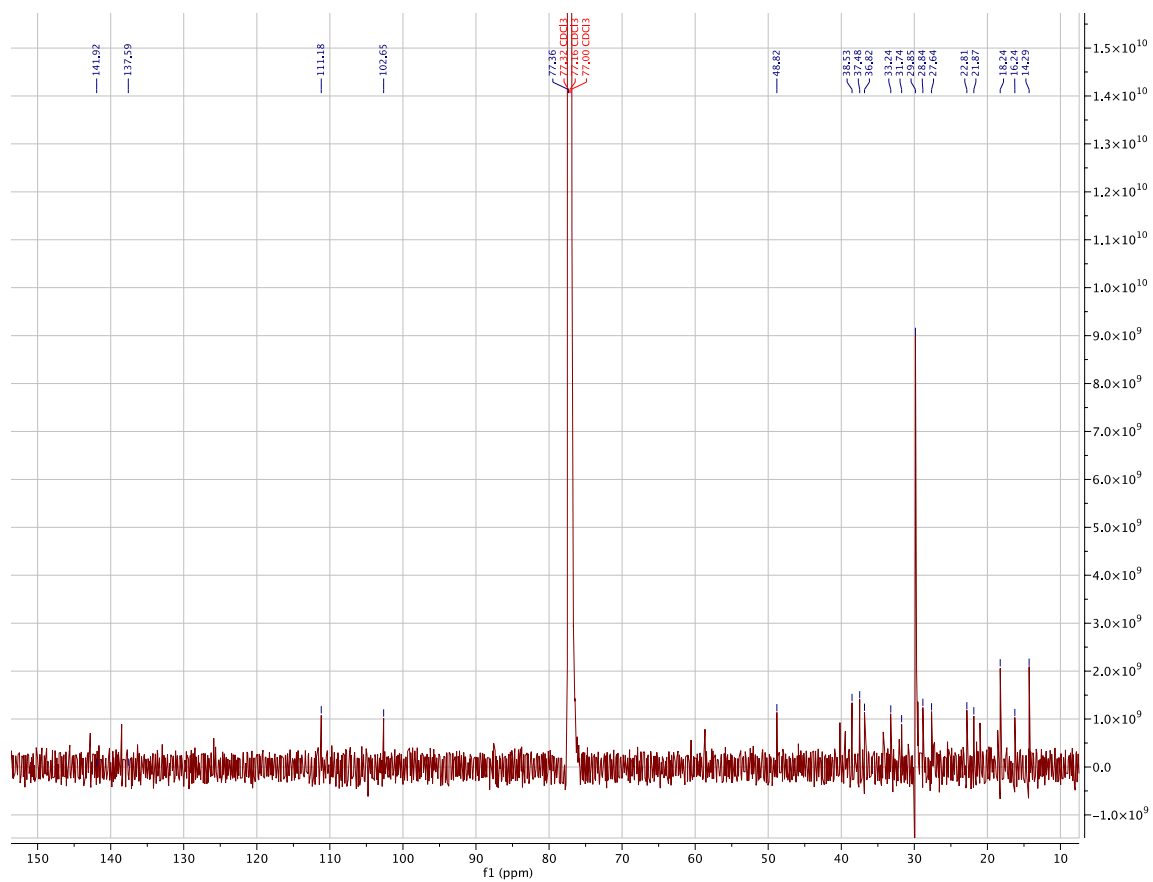


Figure S2.4b. ¹³C NMR spectrum.



Figure S2.4c. HSQC spectrum.



Figure S2.4d. HMBC spectrum.

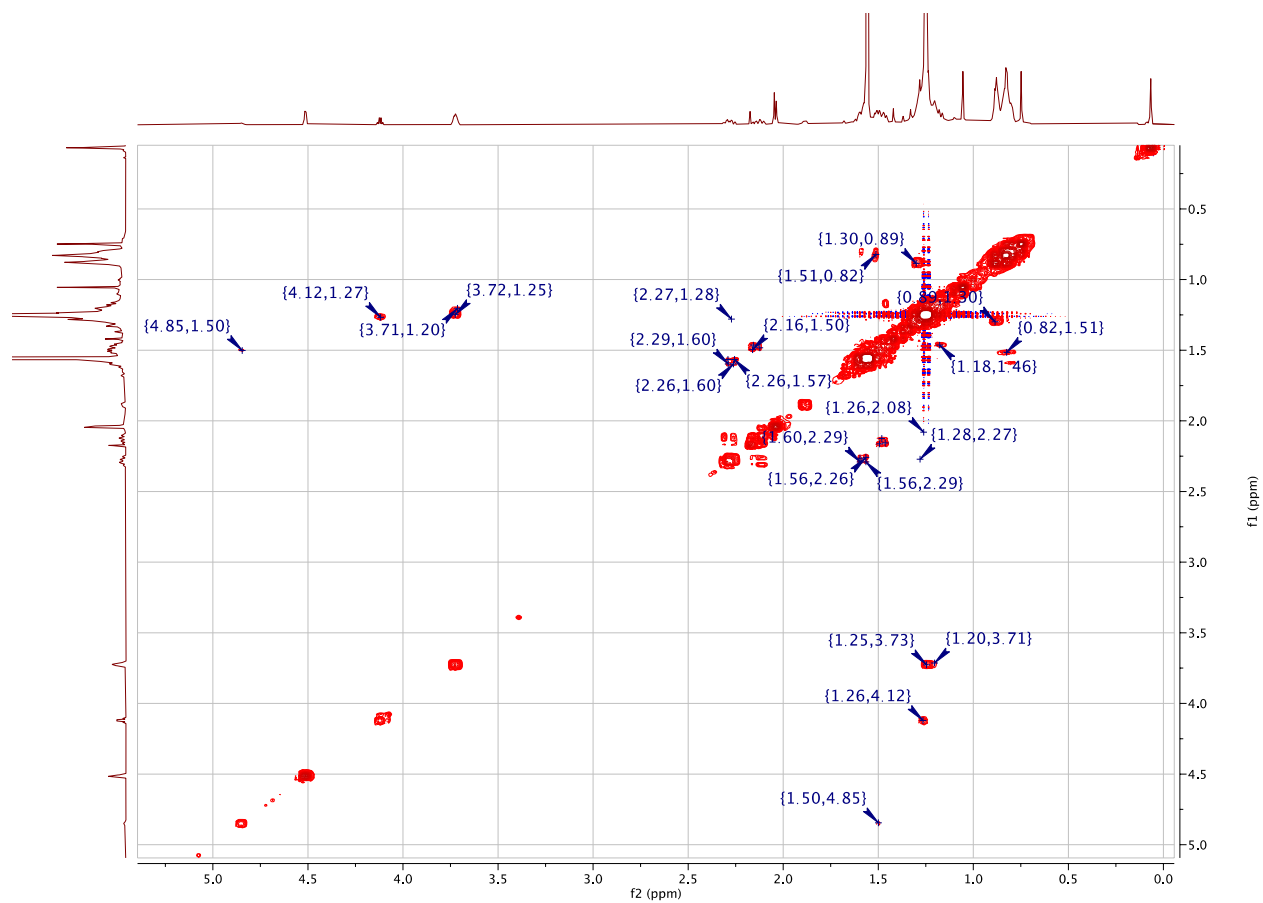
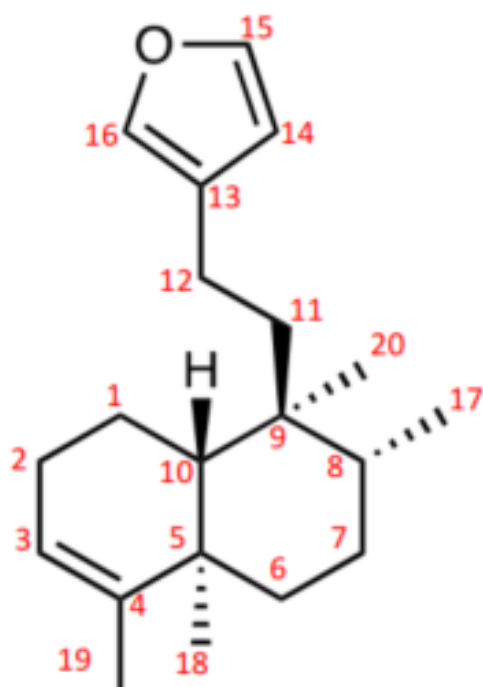


Figure S2.4e. COSY spectrum.



Carbon #	^{13}C δ	^1H δ
1	18.35	1.62/1.43
2	26.97	2.05
3	120.67	5.2
4	144.67	-
5	38.34	-
7	27.53	1.43
8	36.39	1.57
9	38.87	-
10	46.55	1.44
12	18.27	2.32/2.23
13	126.02	-
14	111.16	6.26
15	142.74	7.34
16	138.55	7.2
17	16.09	0.83
18	19.96	1.01
19	18.02	1.59
20	18.35	0.74

Figure S2.5. NMR chemical shift assignments of compound 2. Connectivity was deduced from ^1H , ^{13}C , HSQC, HMBC, and COSY correlations (Fig. S2.6). CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry was assigned based on the configuration of the precursor (-)-kolavenyl diphosphate.

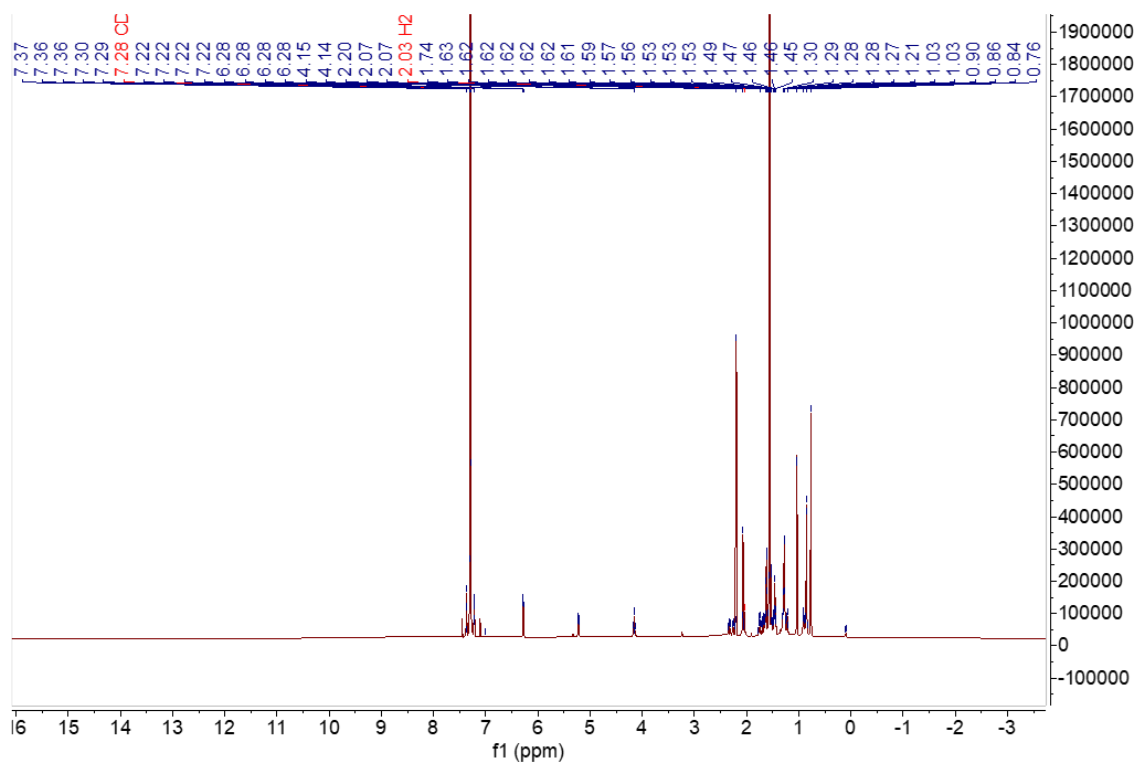


Figure S2.6a. NMR spectra of Compound 2. CDCl₃ was used as the solvent, and CDCl₃ peaks were referenced to 7.26 and 77.00 ppm for ¹H and ¹³C spectra, respectively. (a) ¹H NMR spectrum.

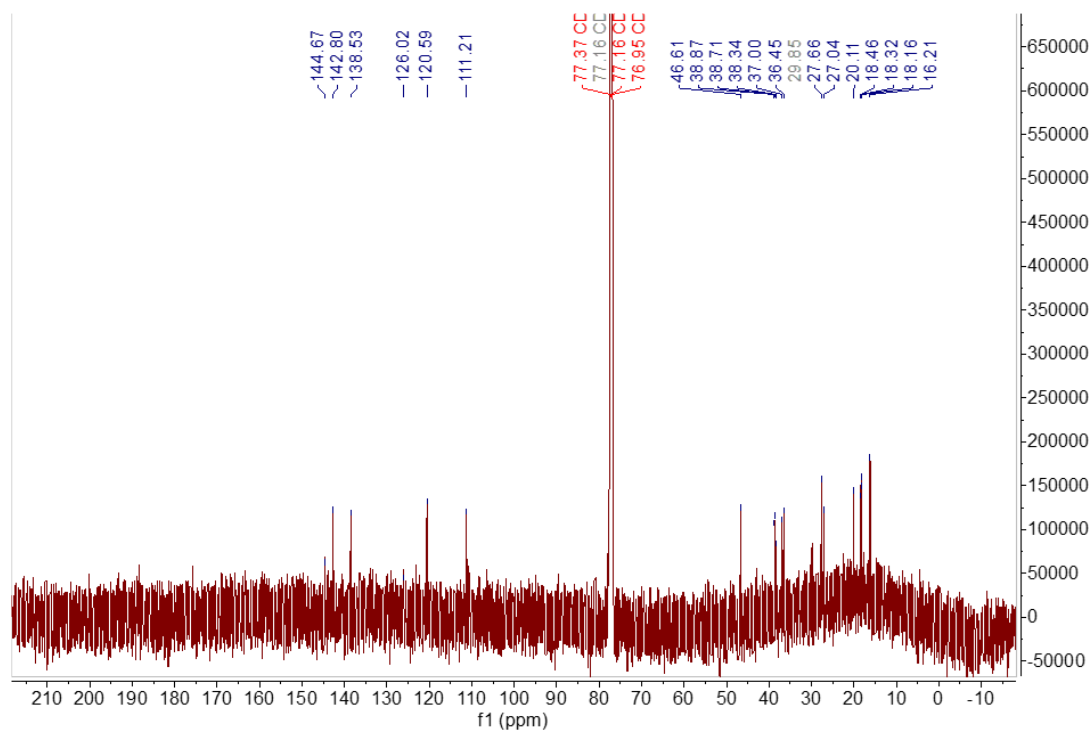


Figure S2.6b. ¹³C NMR spectrum.

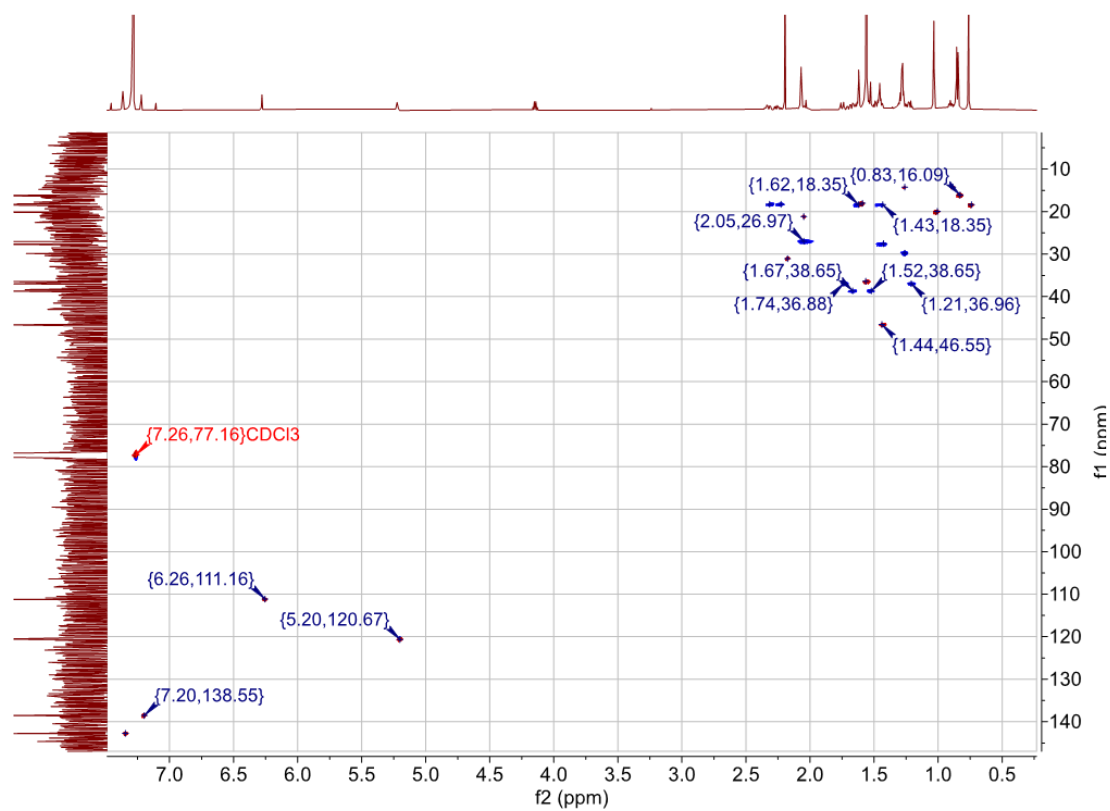


Figure S2.6c. HSQC spectrum.

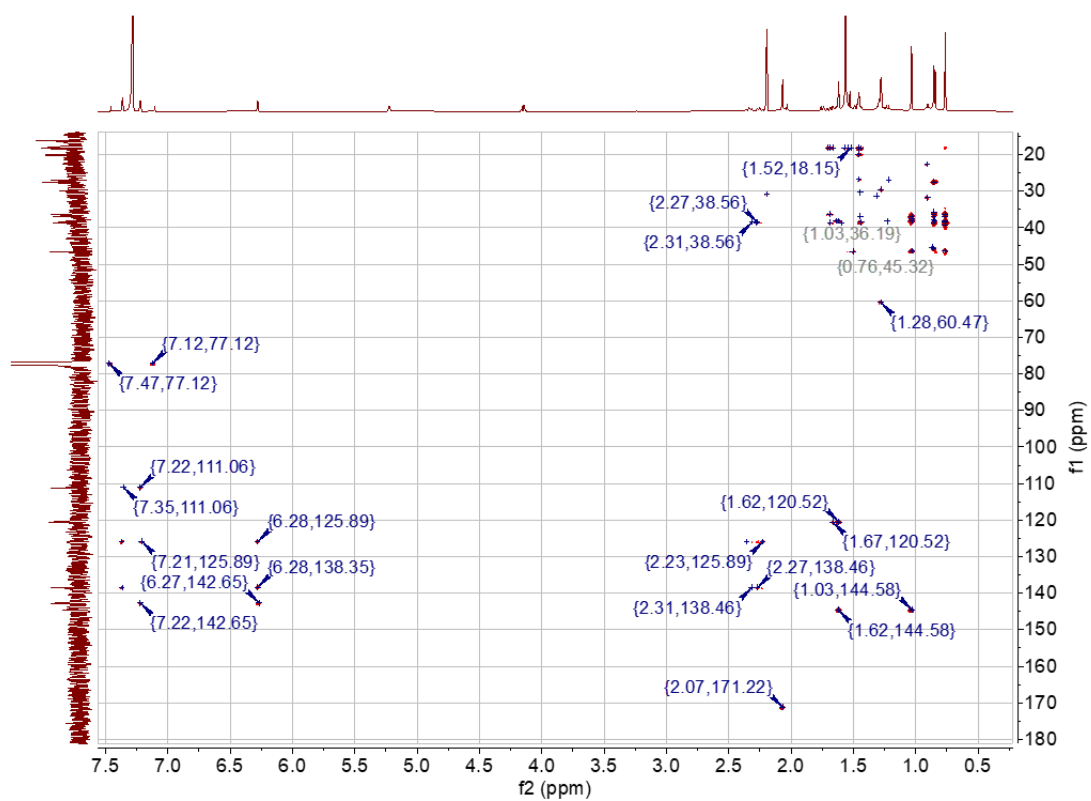


Figure S2.6d. HMBC spectrum.

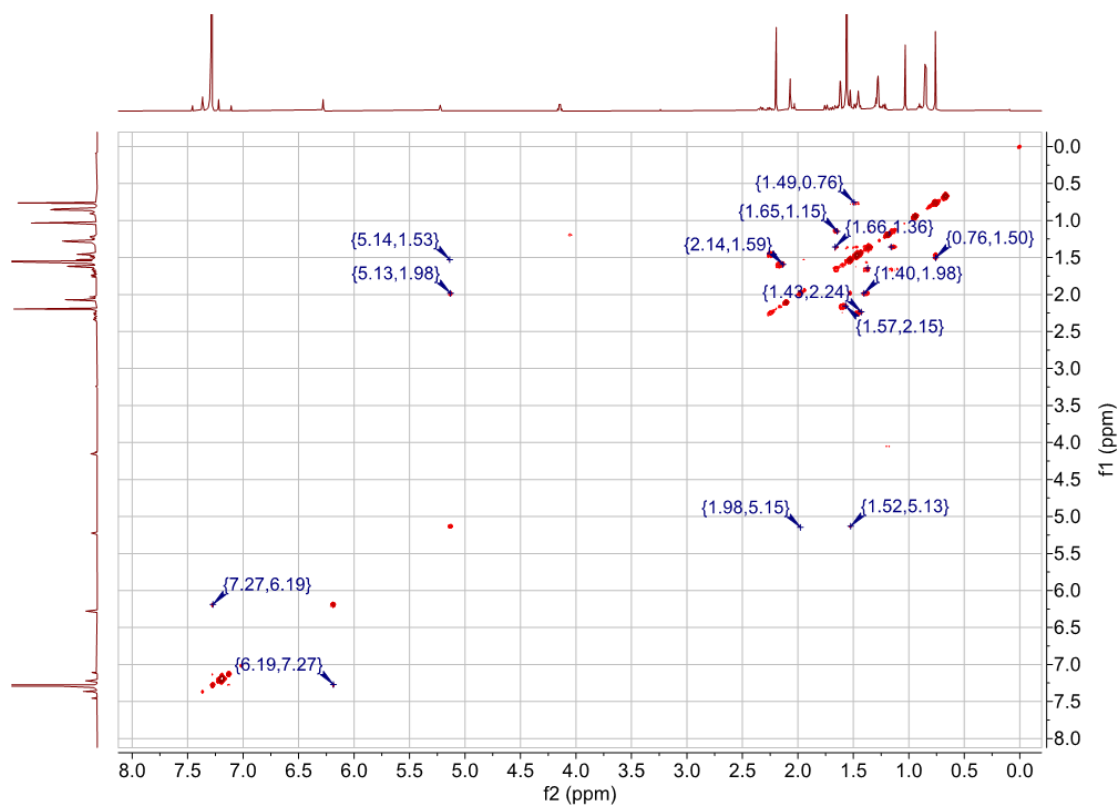


Figure S2.6e. COSY spectrum.

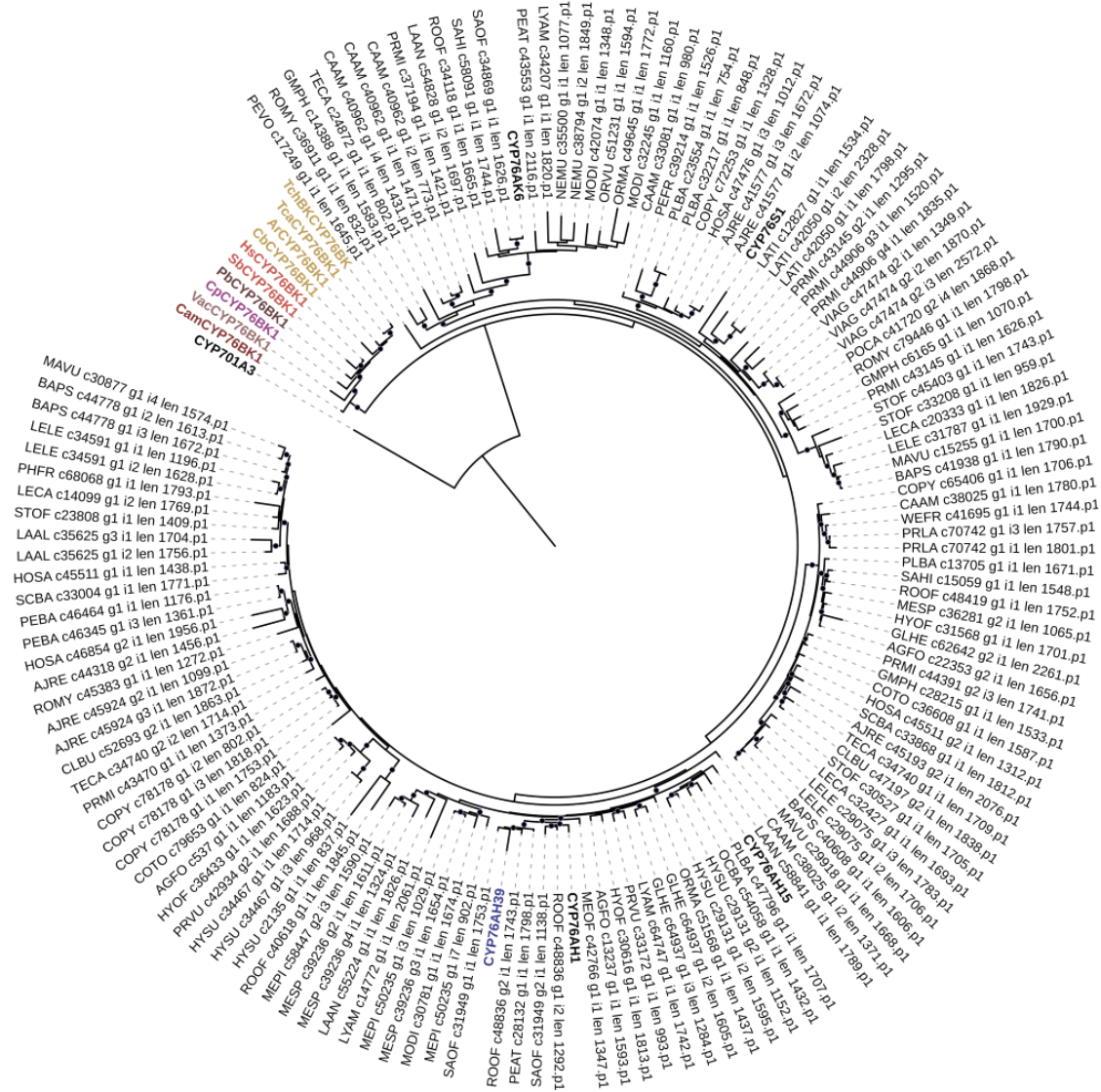
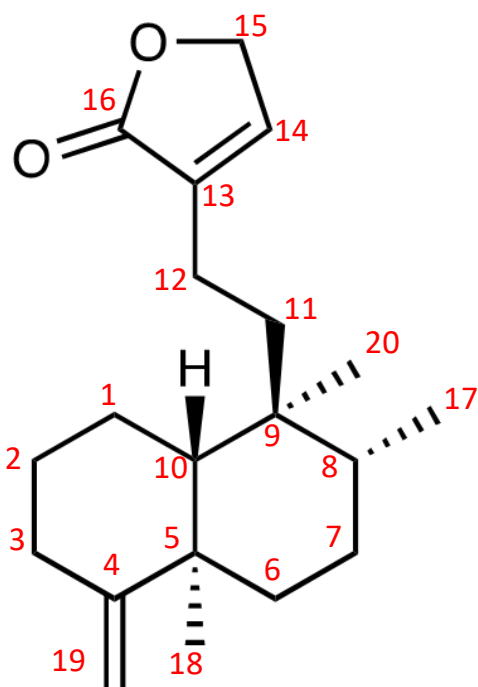


Figure S2.7 CYP76 family phylogeny from the Mint Genome Project. Transcripts with >45% protein identity to either VacYCP76BK, CamCYP76BK1, or ArCYP76BK1 were extracted from this collection of transcriptomes and was further filtered to the CYP76 family alone. Dots on phylogeny represent >70% bootstraps.



Carbon #	^{13}C δ	^1H δ
1	21.77	1.55/1.45
2	28.58	1.32/1.86
3	33.23	2.11/2.27
4	160.4	-
5	40.2	-
6	37.4	1.6
7	27.45	1.48
8	36.7	1.5
9	39.2	-
10	48.6	1.5
11	35.7	1.46/1.56
12	19	2.01/2.15
13	135.3	-
14	143.2	-
15	70.17	4.75
16	174.7	
17	16.2	0.82
18	20.85	1.06
19	102.5	4.5
20	18.2	0.76

Figure S2.8. NMR chemical shift assignments of compound 3. Connectivity was deduced from ^1H , ^{13}C , HSQC, HMBC, and COSY correlations (Fig. S2.9). CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry was assigned based on the configuration of the precursor isokolavenyl diphosphate.

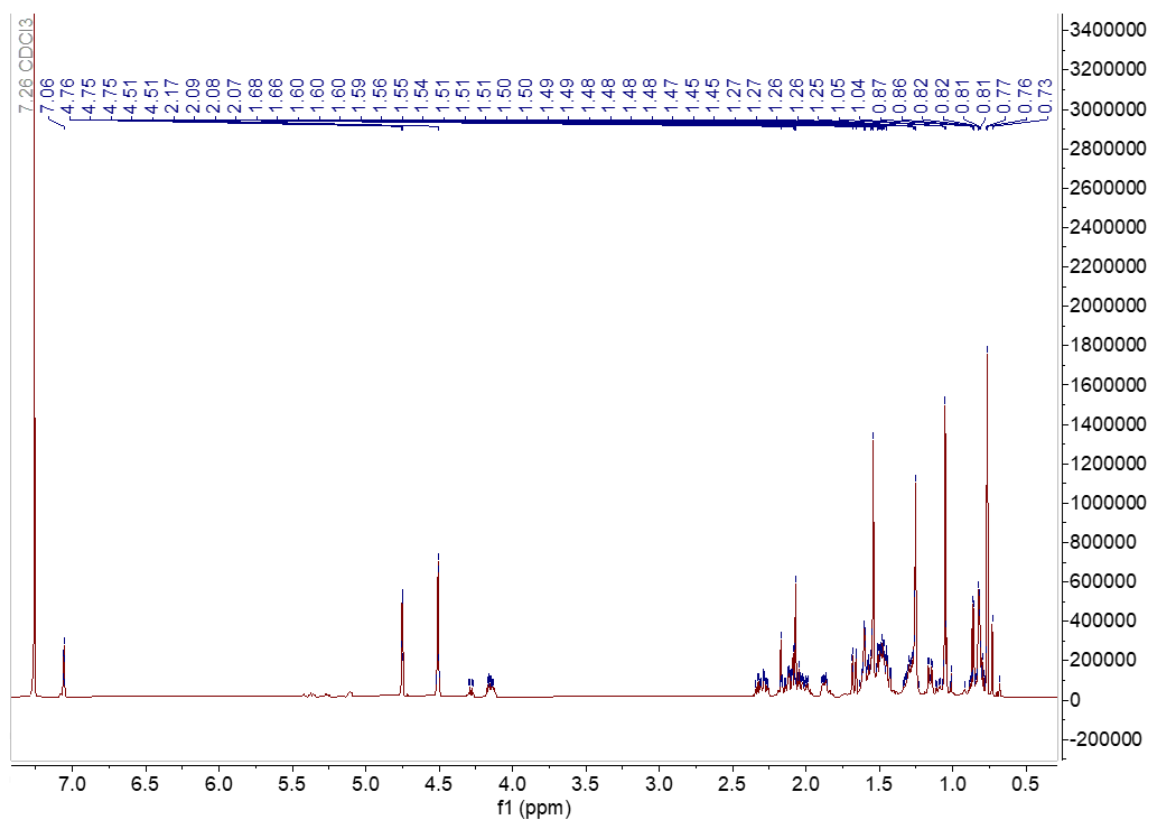


Figure S2.9a. NMR spectra of 3. CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. (a) ^1H NMR spectrum.

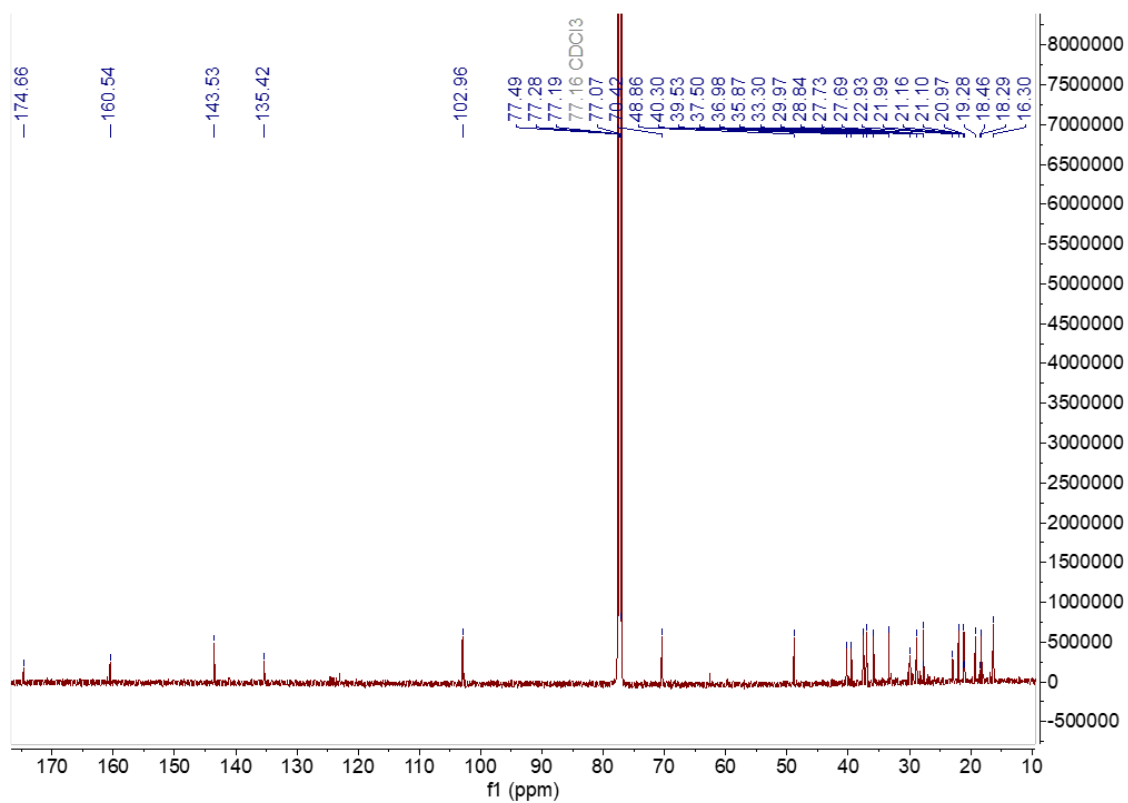


Figure S2.9b. ¹³C NMR spectrum.

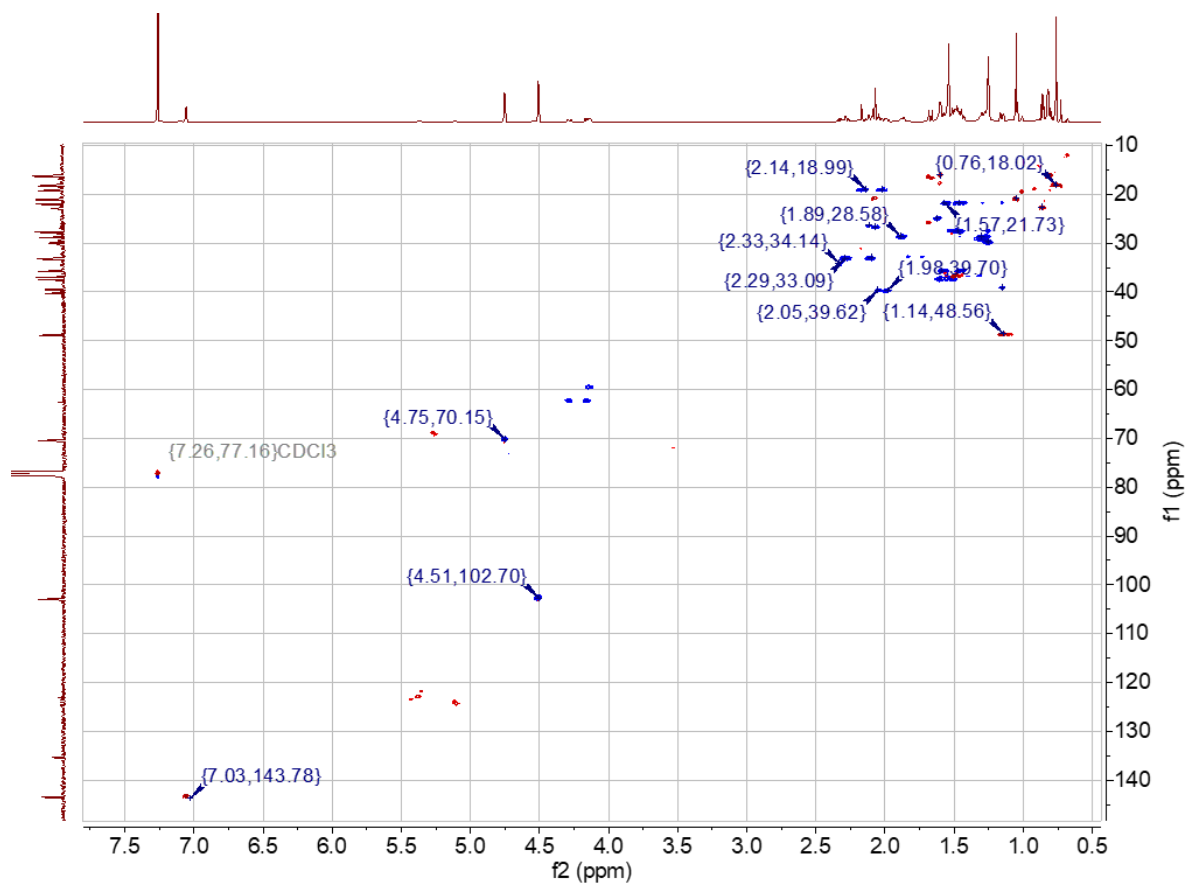


Figure S2.9c. HSQC spectrum.



Figure S2.9d. HMBC spectrum.

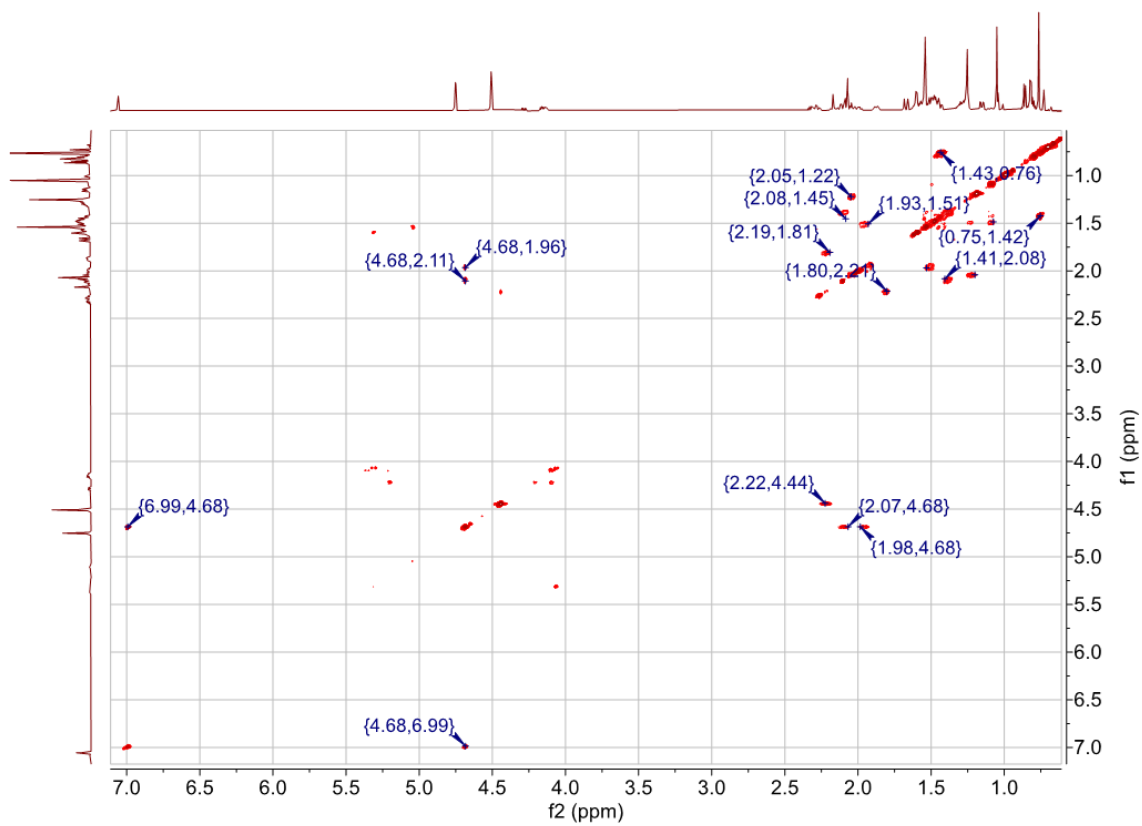
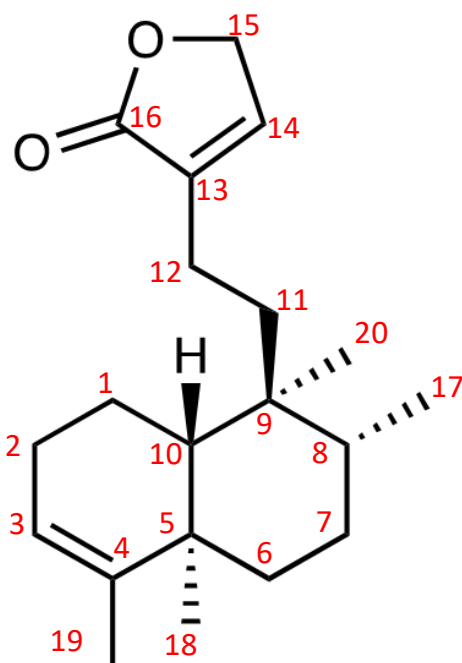


Figure S2.9e. COSY spectrum.



Carbon #	^{13}C δ	^1H δ
1	18.26	1.39/1.45, 1.57
2	19.04	2.1, 2.2
3	120.42	5.19
4	144.25	-
5	38.15	-
6	36.72	1.72, 1.2
7	27.43	
8	36.29	1.51
9	38.65	-
10	46.35	1.41
11	35.74	1.65, 1.46/1.49
12	26.79	1.98/2.07
13	135.22	-
14	143.26	7.08
15	70.13	4.76
16	174.38	-
17	17.97	1.59
18	19.91	1
19	18.19	0.75
20	15.98	1.6, 0.82

Figure S2.10. NMR chemical shift assignments of compound 4. Connectivity was deduced from ^1H , ^{13}C , HSQC, HMBC, and COSY correlations (Supplementary Fig. 6). CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry was assigned based on the configuration of the precursor (-)-kolavenyl diphosphate.

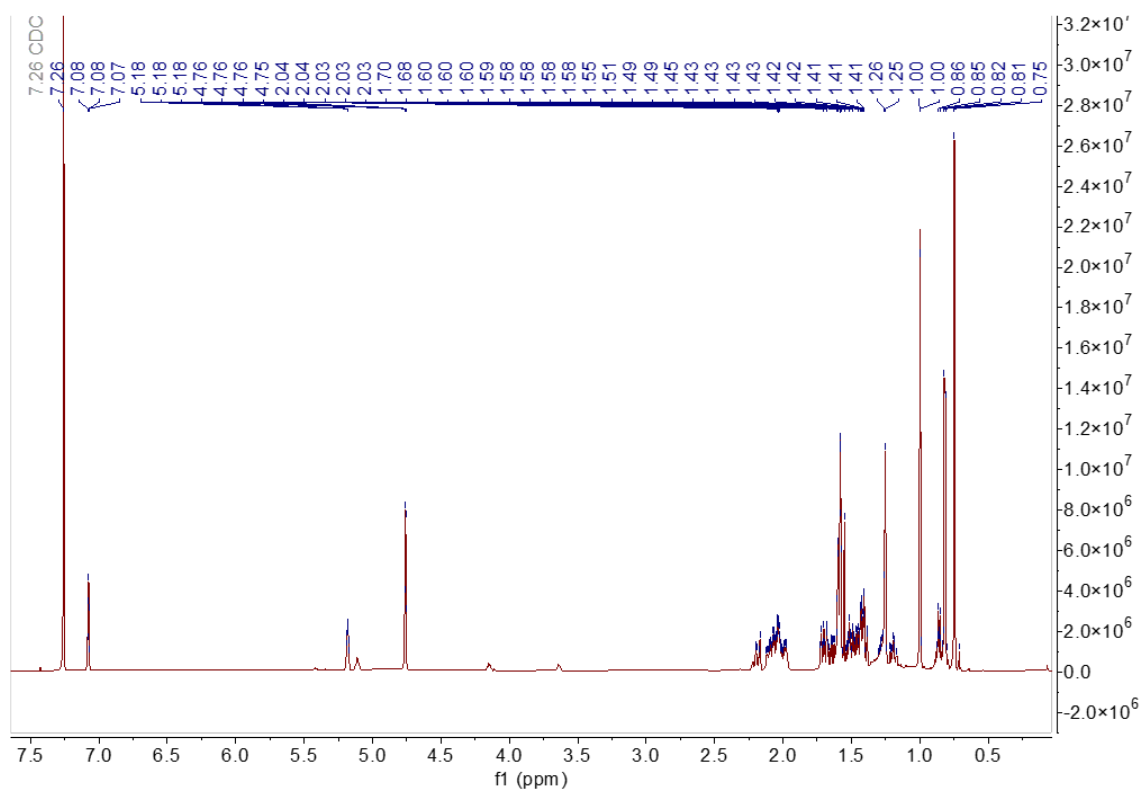


Figure S2.11a. NMR spectra of compound 4. CDCl₃ was used as the solvent, and CDCl₃ peaks were referenced to 7.26 and 77.00 ppm for ¹H and ¹³C spectra, respectively. (a) ¹H NMR spectrum.

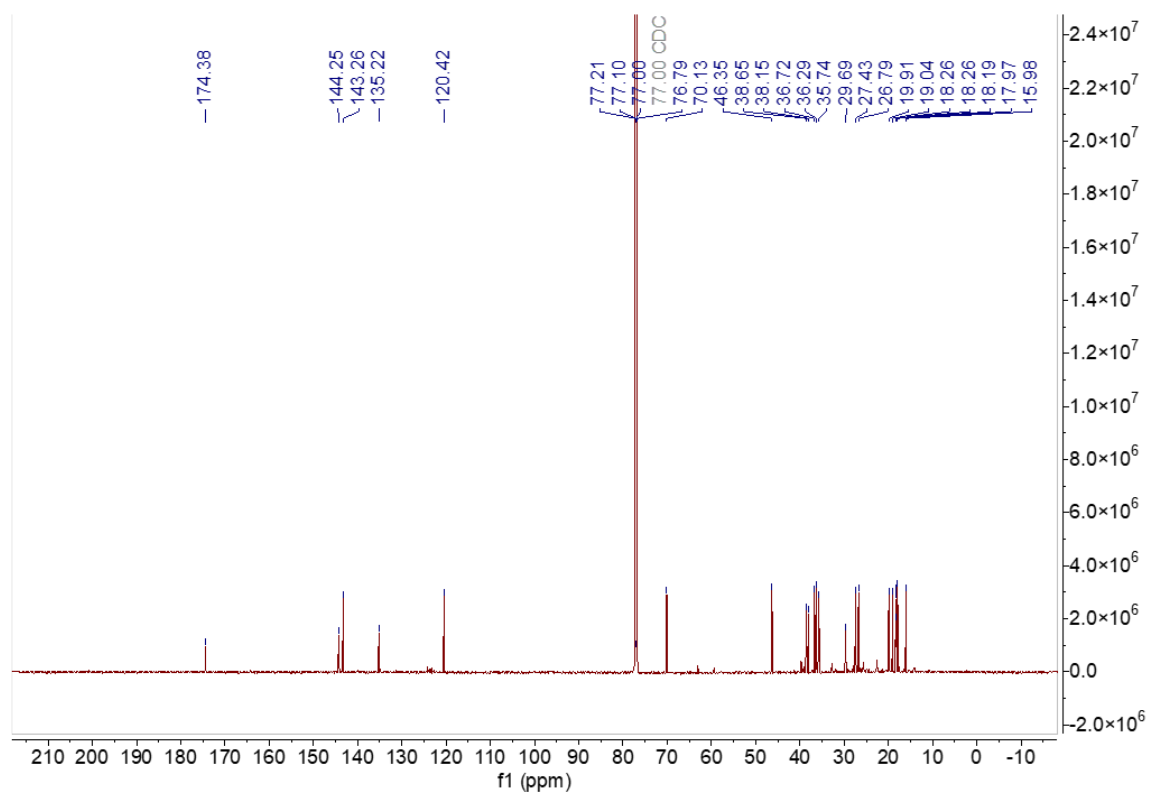


Figure S2.11b. ¹³C NMR spectrum.

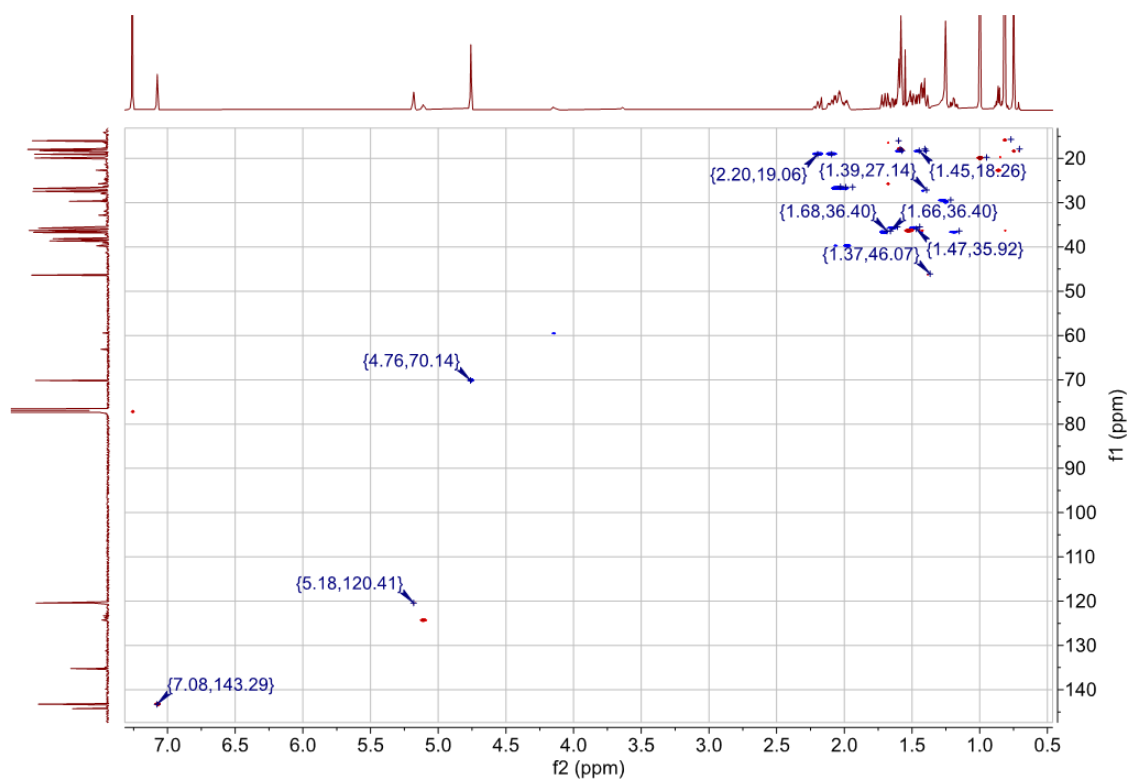


Figure S2.11c. HSQC spectrum.

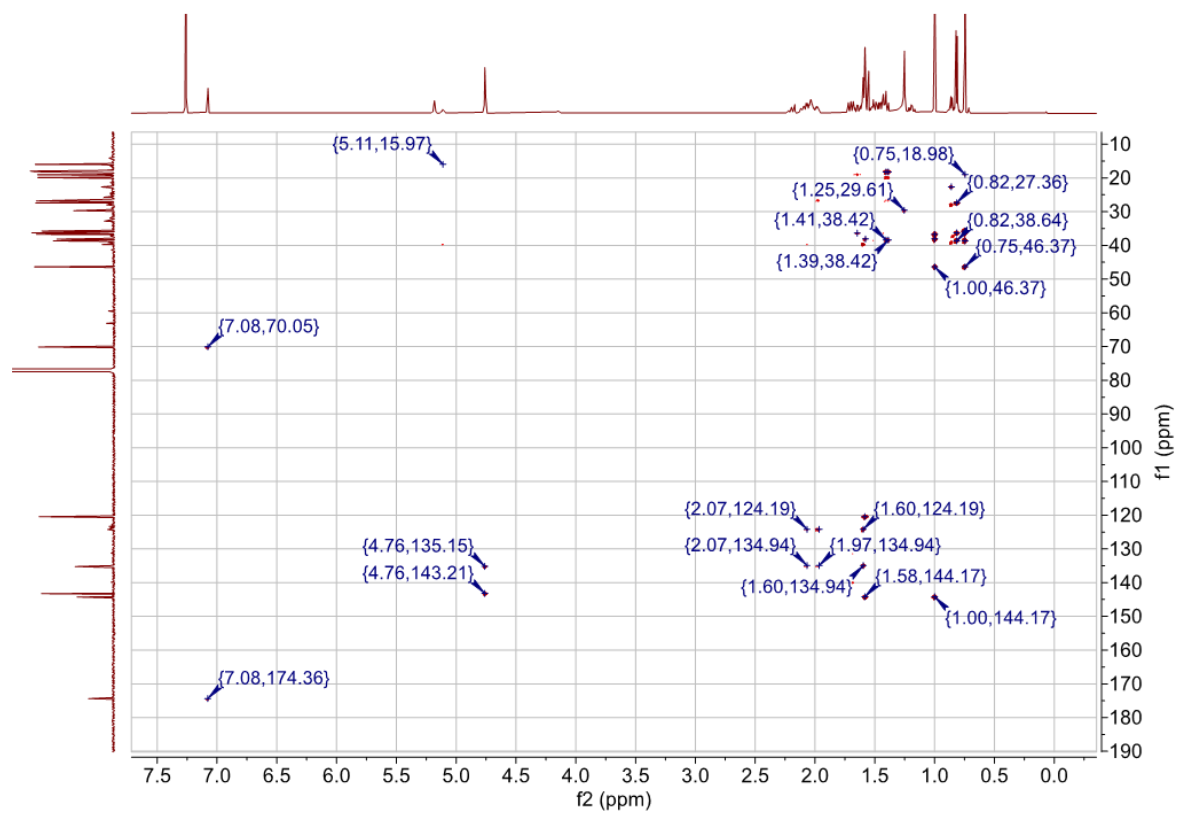


Figure S2.11d. HMBC spectrum.

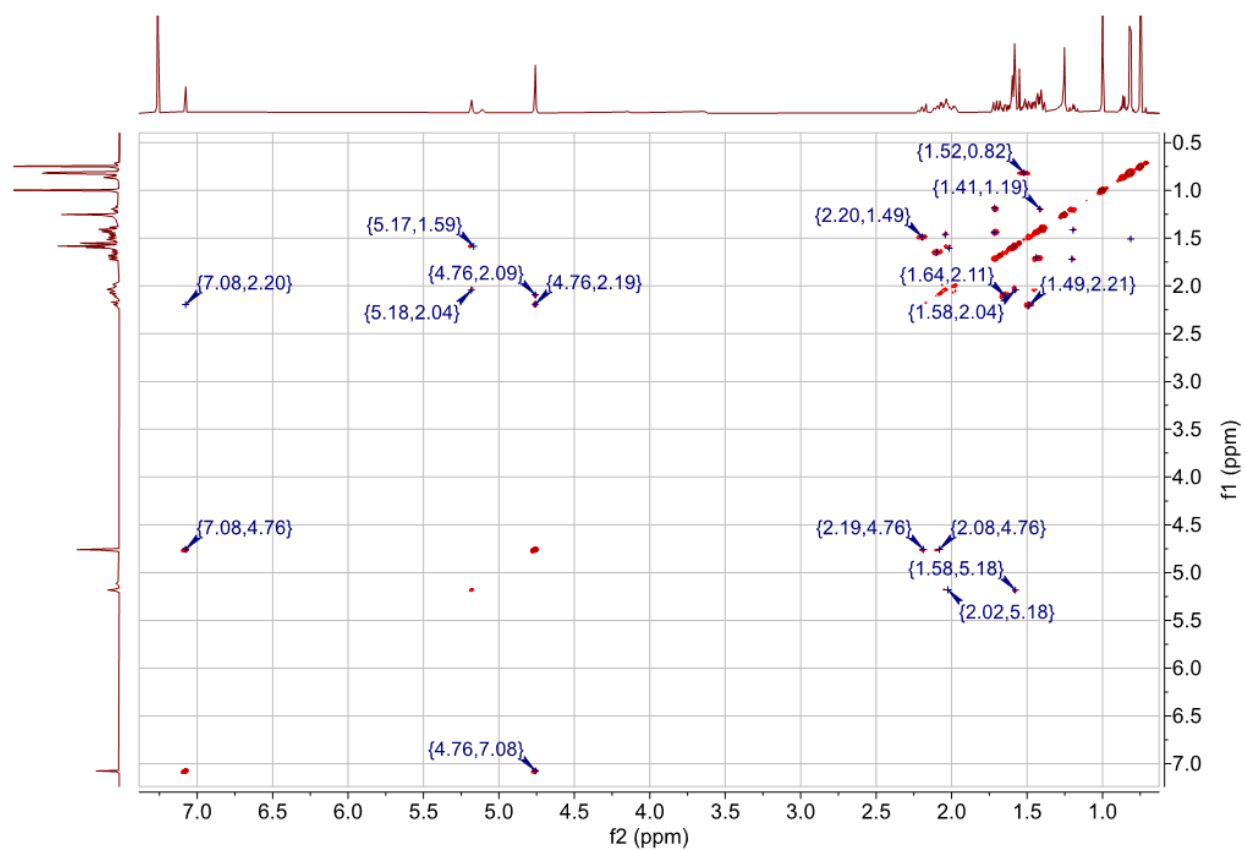
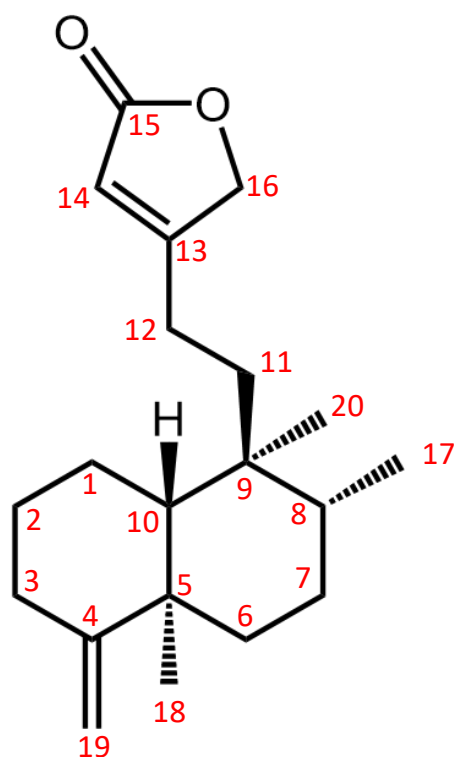


Figure S2.11e. COSY spectrum.



Carbon #	^{13}C δ	^1H δ
1	21.93	1.47
2	28.9	1.25, 1.91
3	33.31, 33.27	2.12, 2.29
4	160.39	NA
5	40.17	NA
6	37.37	1.53, 1.64
7	27.69	1.5
8	36.91	1.41
9	39.42	NA
10	49.12	1.06
11	35.36	1.52, 1.6
12	22.38	2.25, 2.12
13	171.18	NA
14	115.32	5.81
15	174.19	NA
16	73.19	4.71
17	16.14	0.81
18	20.86	1.05
19	103.03	4.52
20	18.06	0.79

Figure S2.12. NMR chemical shift assignments of compound 5. Connectivity was deduced from ^1H , ^{13}C , HSQC, HMBC, and COSY correlations (Supplementary Fig. 8). CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry was assigned based on the configuration of the precursor isokolavenyl diphosphate.

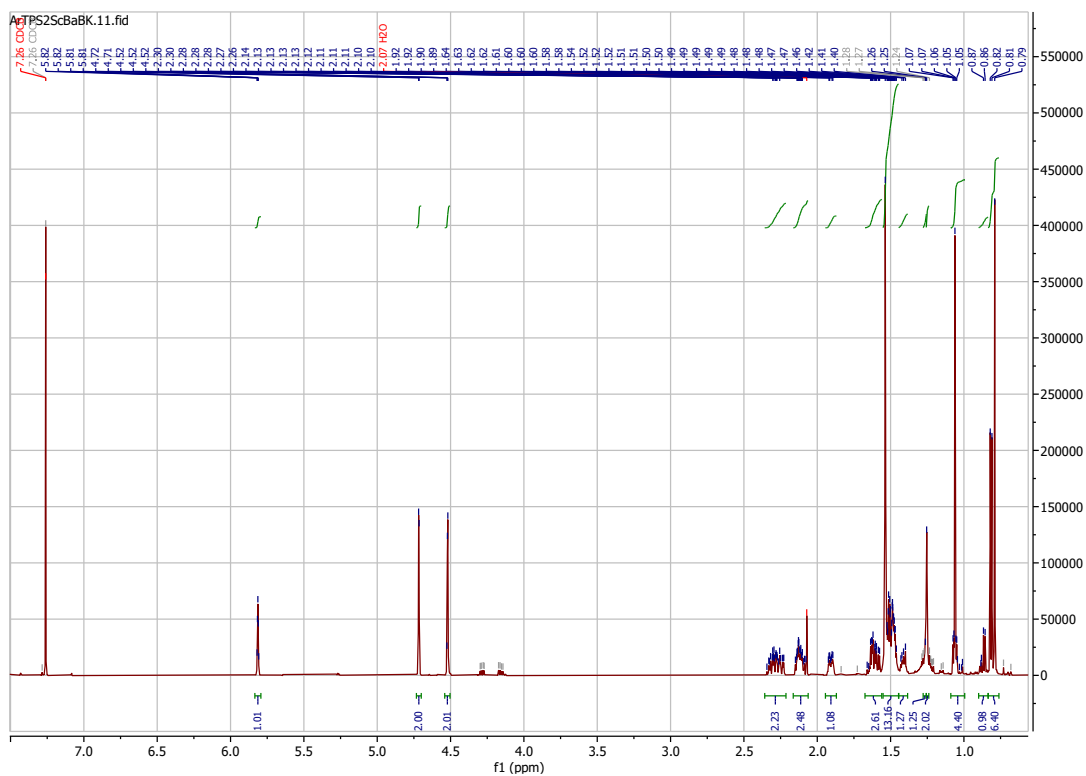


Figure S2.13a. NMR spectra of Compound 5. CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. (a) ^1H NMR spectrum.

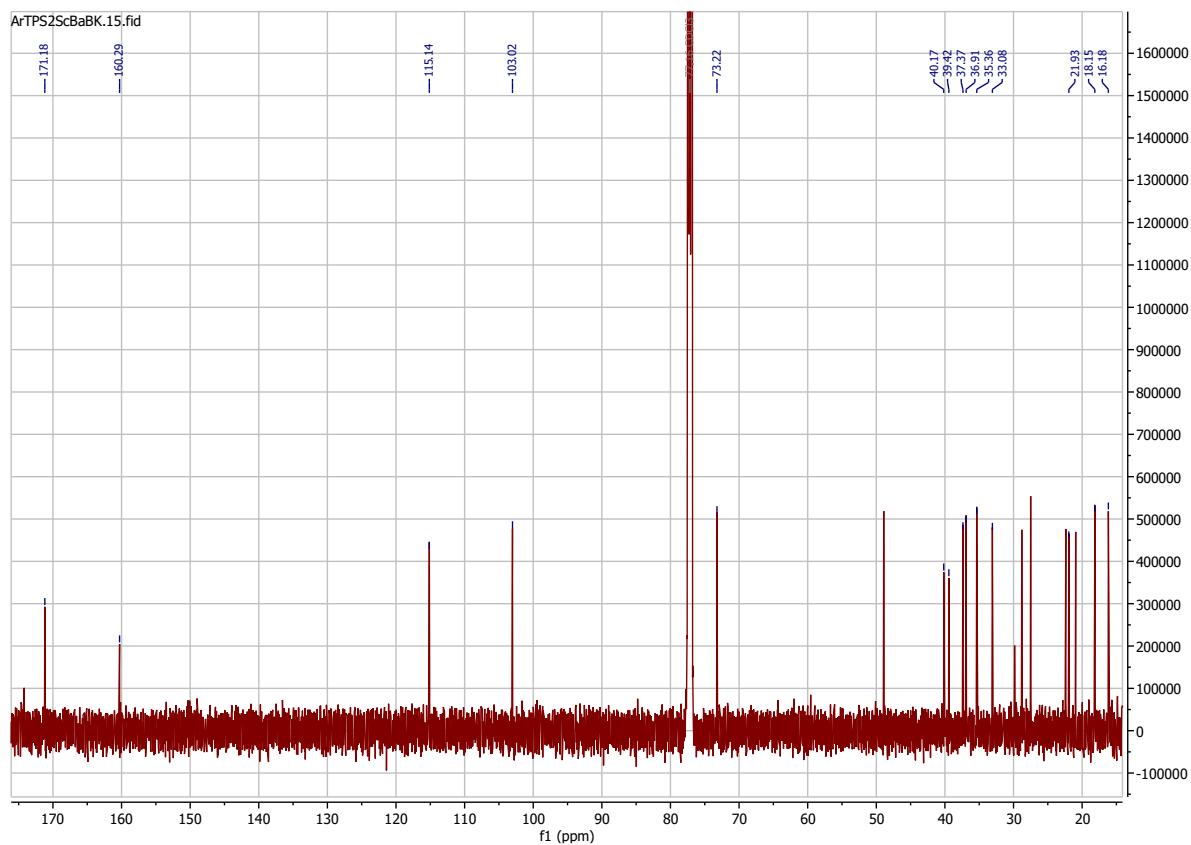


Figure S2.13b. ^{13}C NMR spectrum

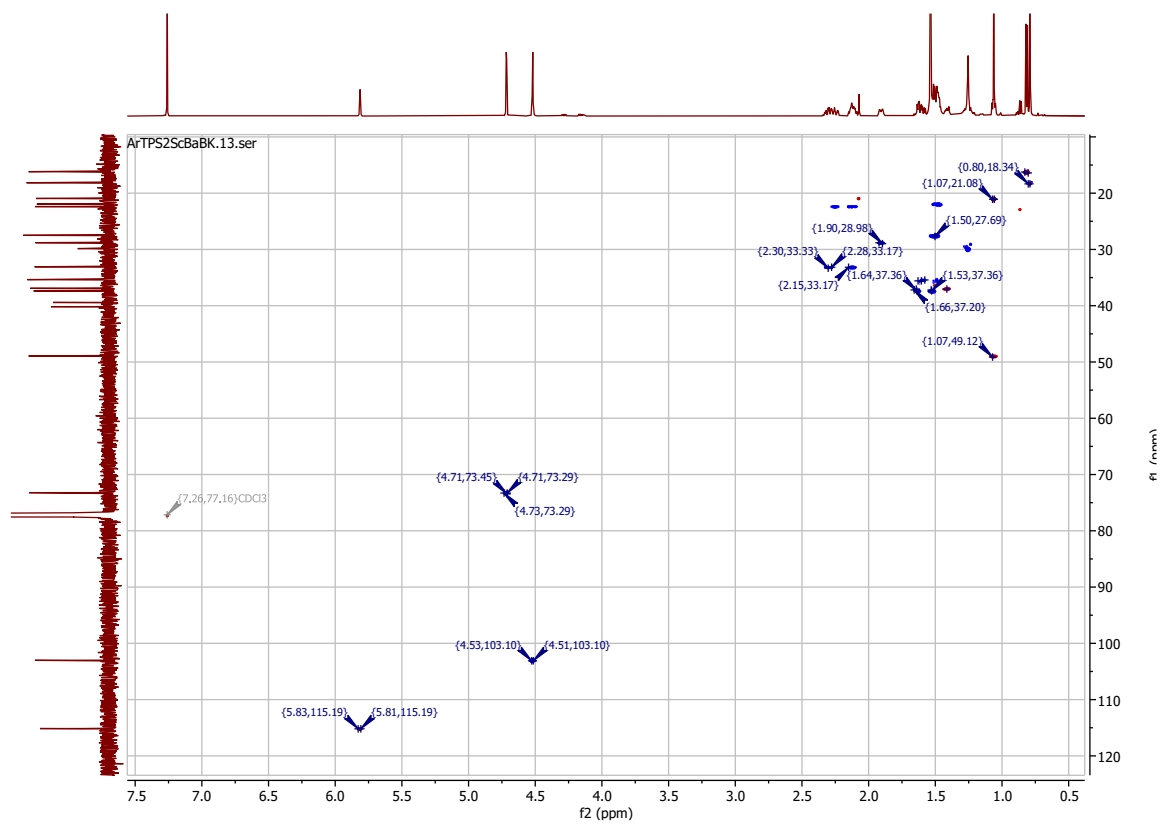


Figure S2.13c. HSQC spectrum.

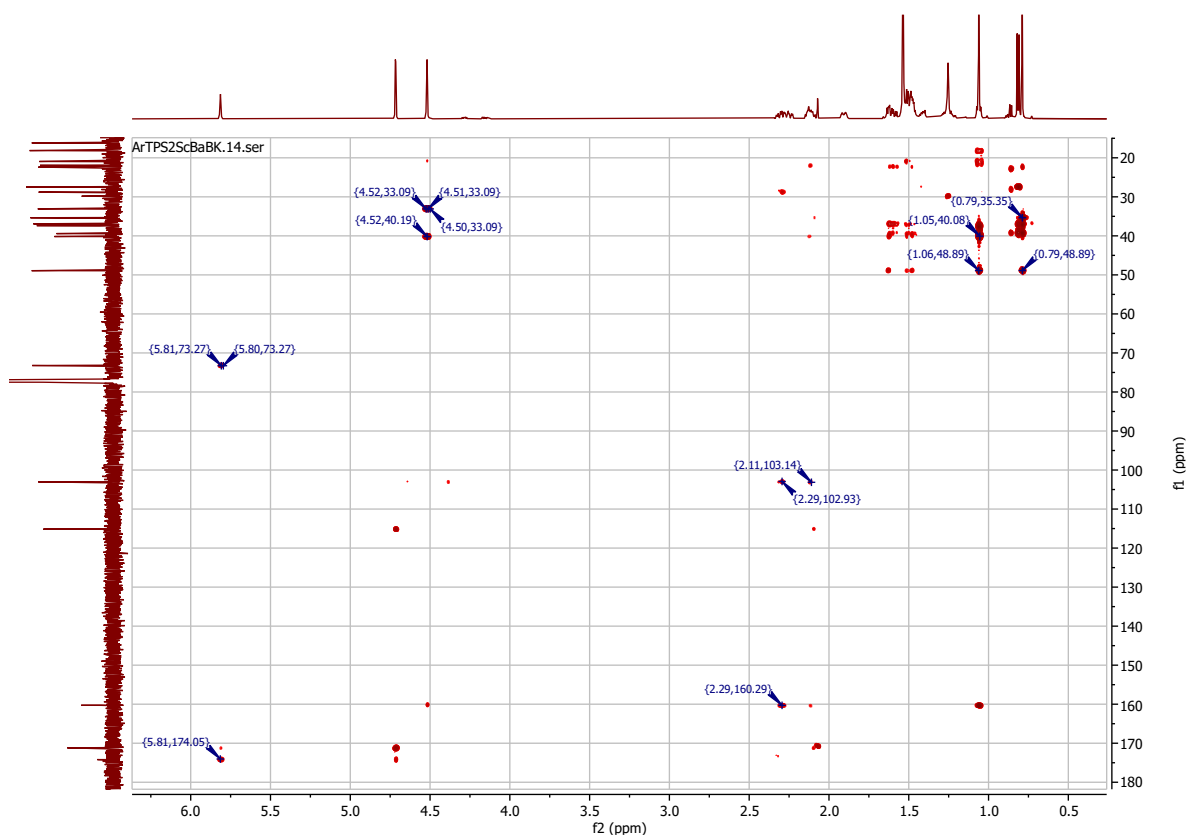


Figure S2.13d. HMBC spectrum.

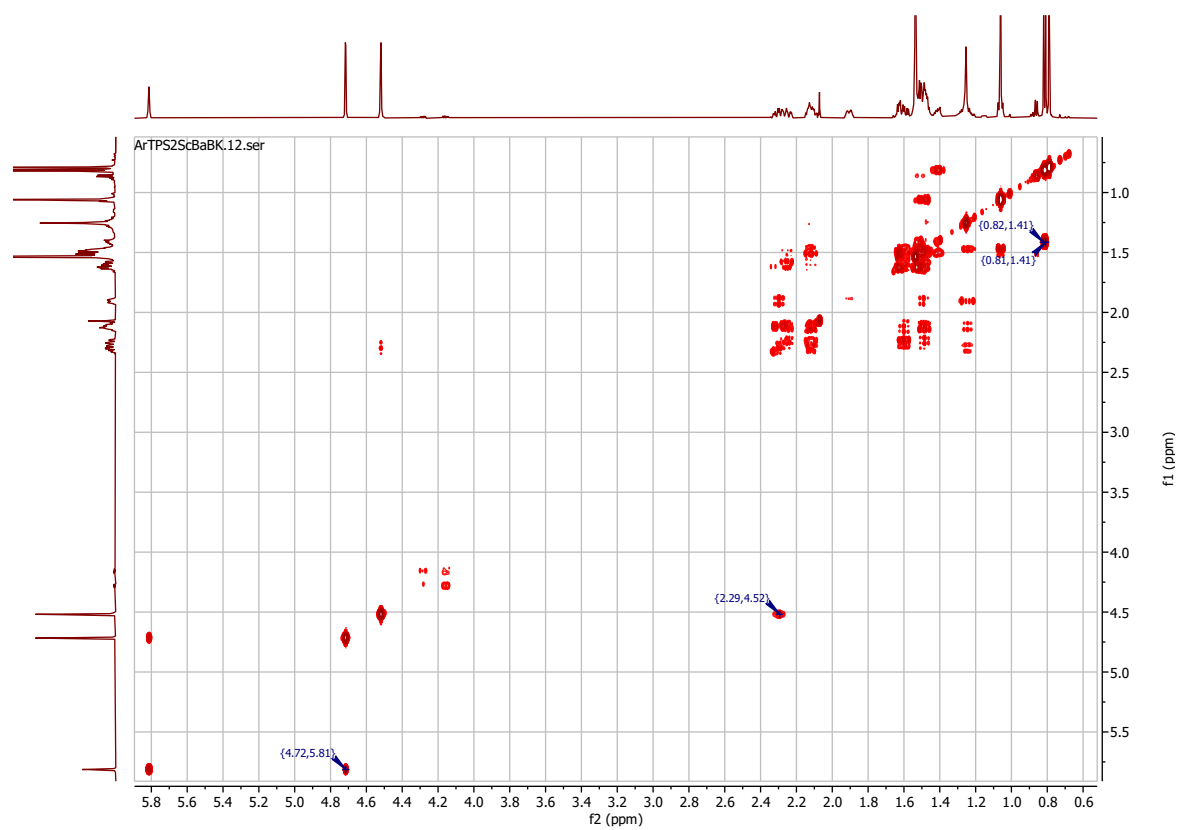


Figure S2.13e. COSY spectrum.

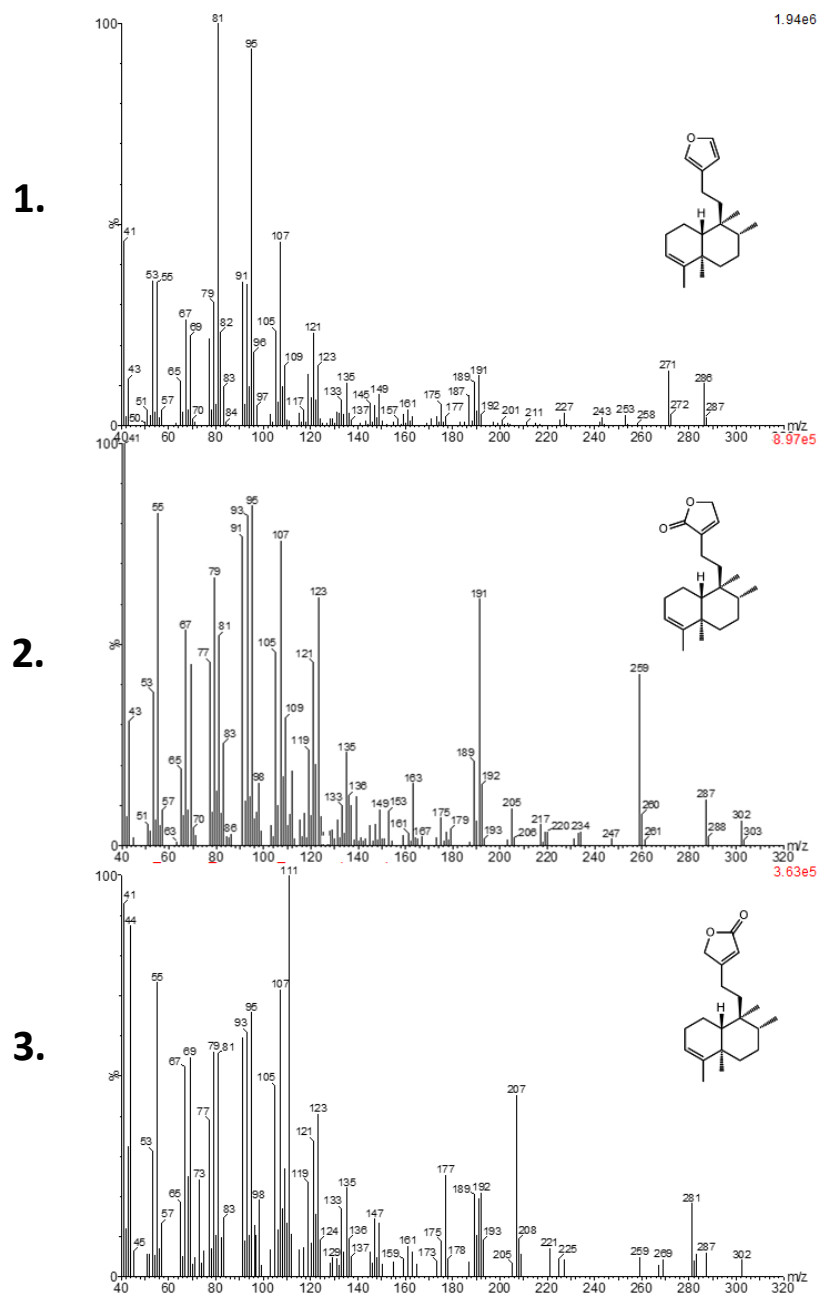


Figure S2.14. (-)-kolavenol CYP76BK1 products. All infiltrations here included DXS-GGPPS + ArTPS2. Spectra 1. had a retention time of 12.44 min and included CbCYP76BK1. Spectra 2. had a retention time of 15.84 min and also included CbCYP76BK1 in the infiltration. Spectra 3. had a retention time of 16.52 min and included SbCYP76BK1 in the infiltration. These spectra align with the numbered compounds on Figure 2.7 of main text.

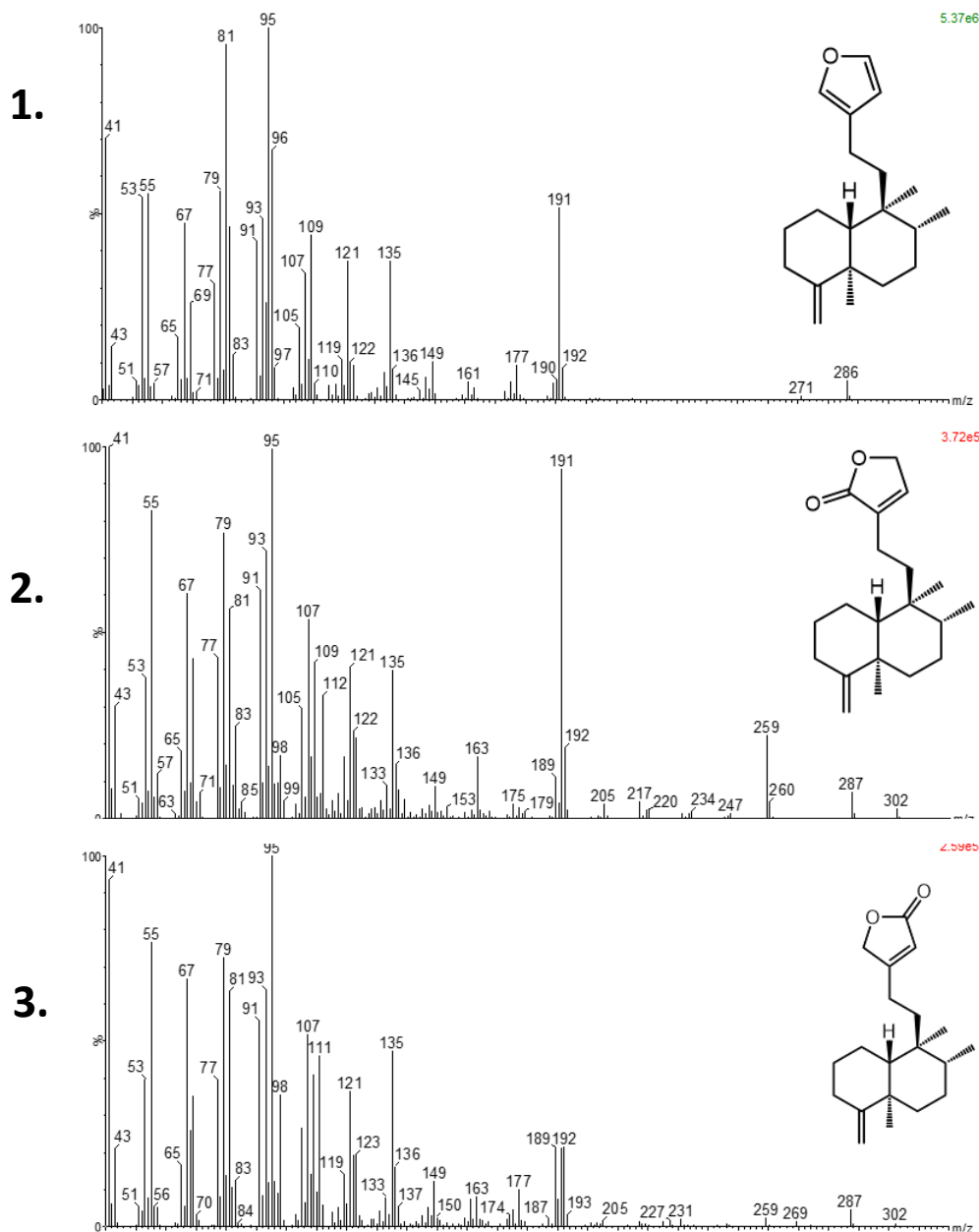


Figure S2.15. isokolavenol CYP76BK1 products. All infiltrations here included DXS-GGPPS + ArTPS2. Spectra 1. had a retention time of 12.35 min and included CbCYP76BK1. Spectra 2. had a retention time of 15.77 min and included CbCYP76BK1 in the infiltration. Spectra 3. had a retention time of 16.47 min and included SbCYP76BK1 in the infiltration.

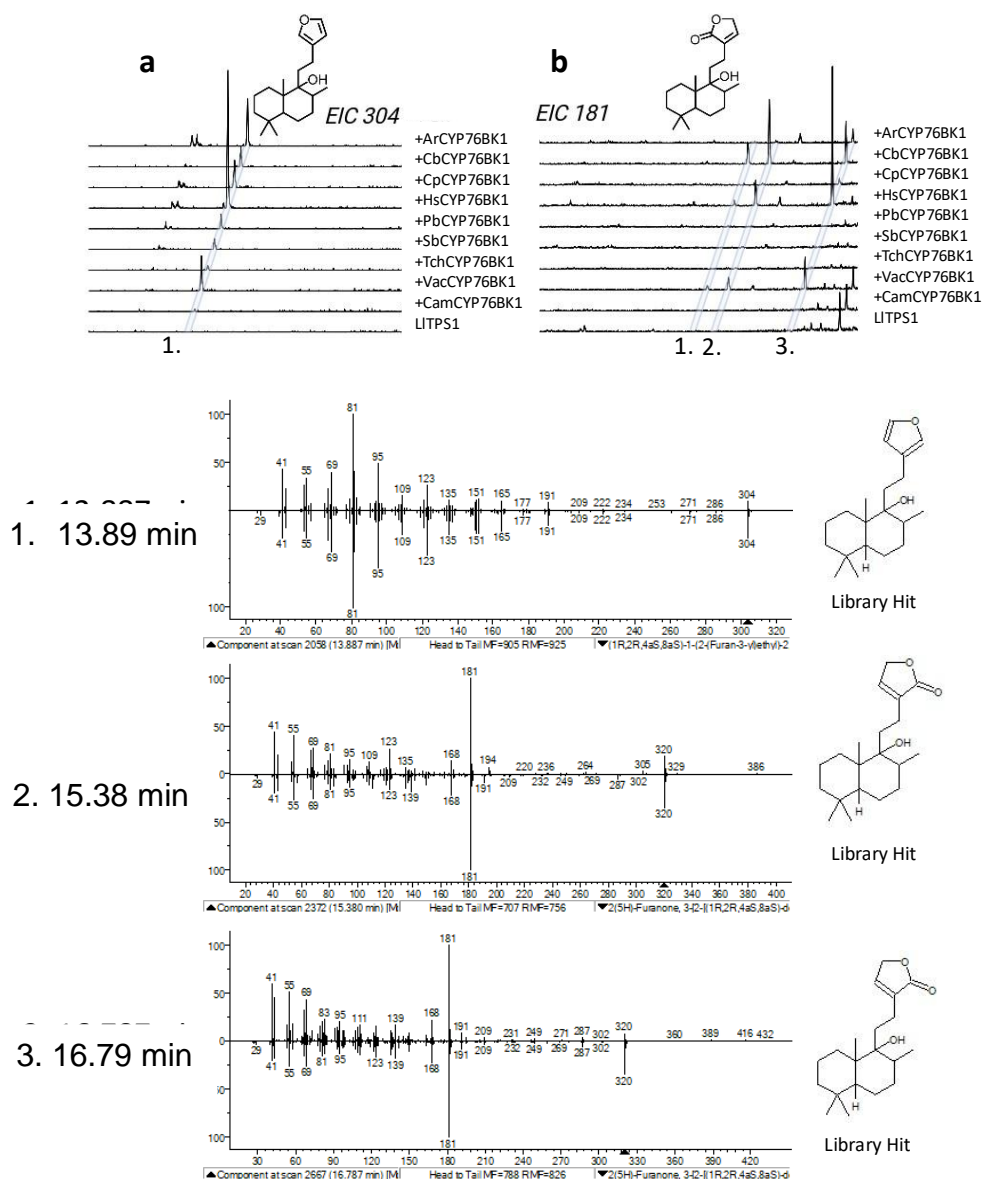


Figure S2.16. Extracted ion chromatograms of the peregrinol backbone (DXS-GGPPS + LITPS1) paired with the various CYP76BK1s. a is an EIC of 304 m/z , which highlights the parent ion of the furan backbone. Lactone parent ions were small, therefore 181 m/z was used for the lactone EIC. Compound 1 was found at 13.89 min, Compound 2 at 15.38 min, and compound 3 at 16.79 min. Library hits had a 70+ percent match to a furan or lactone on the peregrinol backbone, supporting the likely activity.

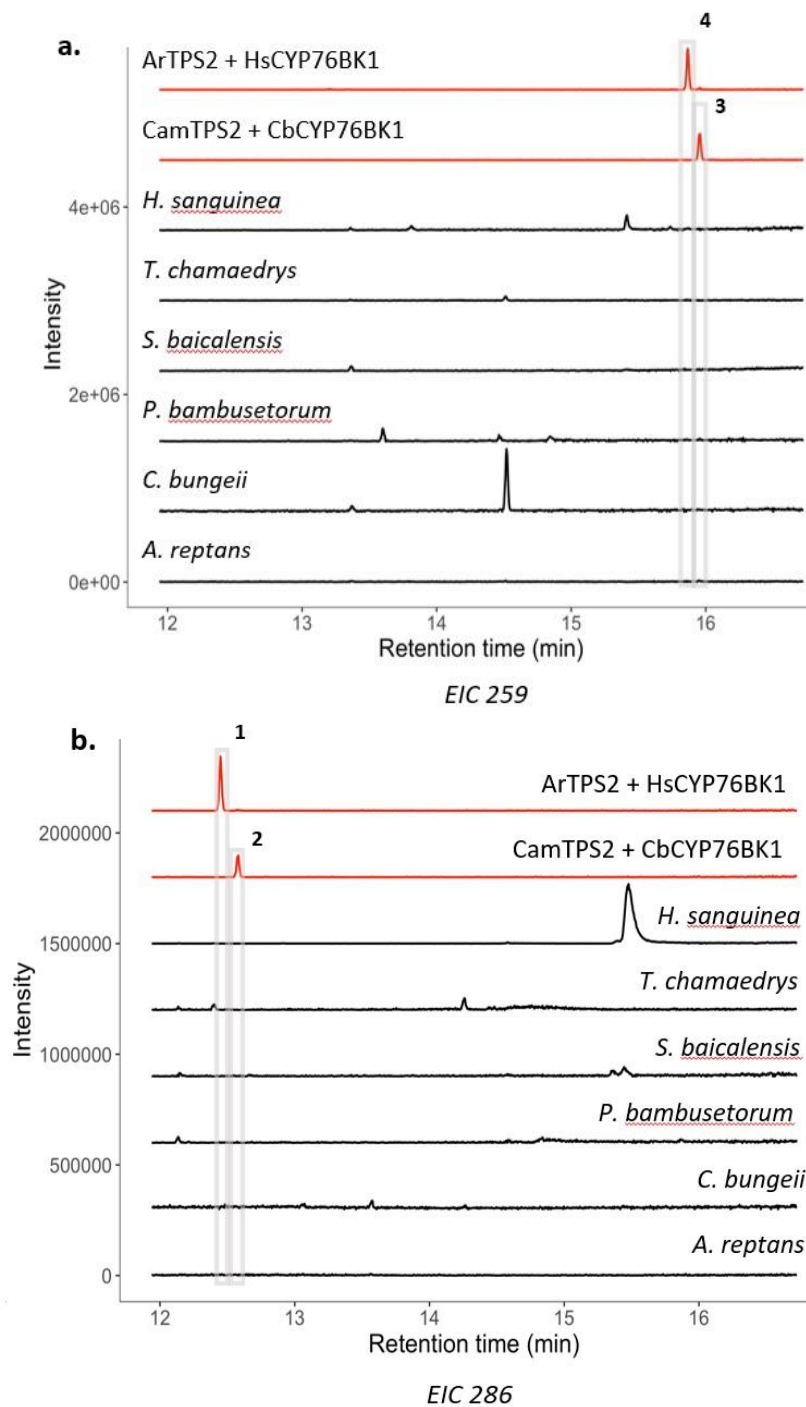


Figure S2.17. GC-MS analysis of plant extracts. None of these extracts have peaks aligning with enzyme products 1, 2, 3, or 4. However, some of the visible peaks in these EIC chromatograms could be unidentified terpenoid structures.

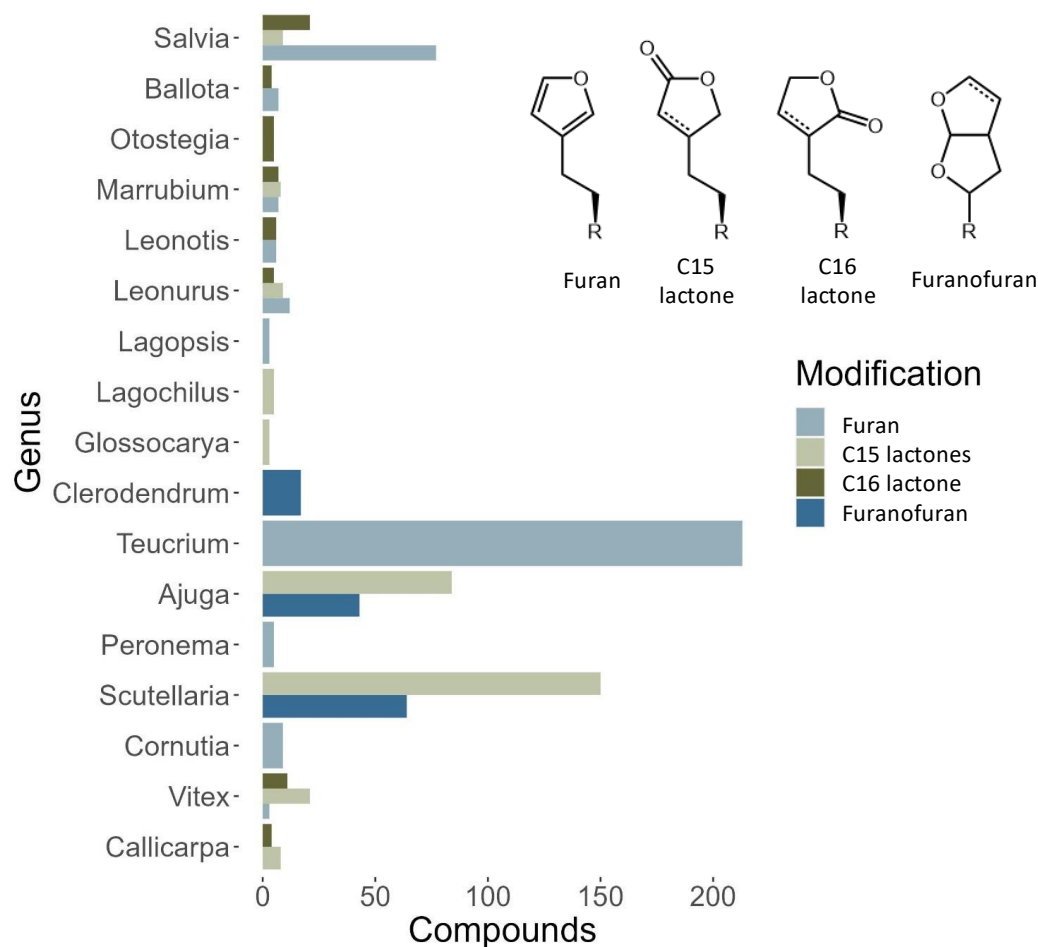


Figure S2.18. Lamiaceae clerodane and labdane furan, lactone, and furanofuran moiety distribution. Dashed lines on structures represent single or double bonds were used in the DNP search. Plotted is the combined clerodane and labdanes containing the drawn substructures. Of note, there is a noticeable lack of c16 lactones, however they are a presumed precursor to furanofurans.

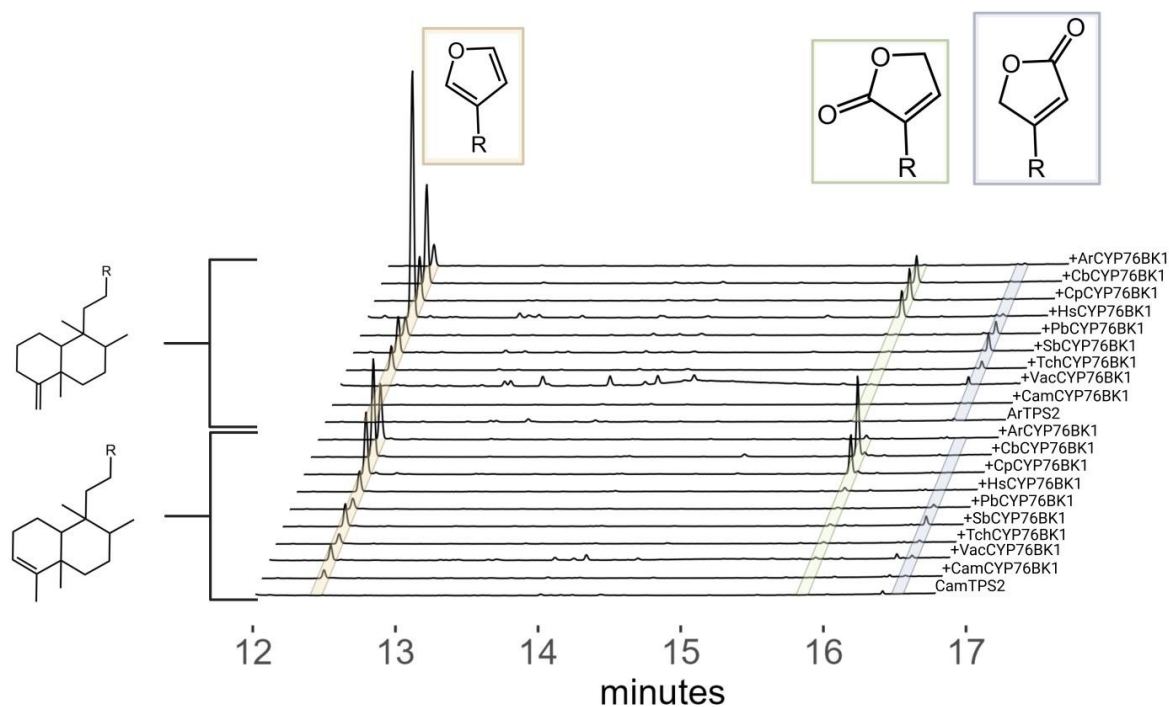


Figure S2.19. Combined kolavenol and isokolavenol CYP76BK1 total ion chromatograms. The bottom 10 chromatograms are the (-)-kolavenyl diphosphate backbone paired with the respective CYP76BK1 orthologs and the top 10 are the isokolavenyl diphosphate backbone paired with the respective CYP76BK1 orthologs. Combining these chromatograms allows direct comparisons of abundances.

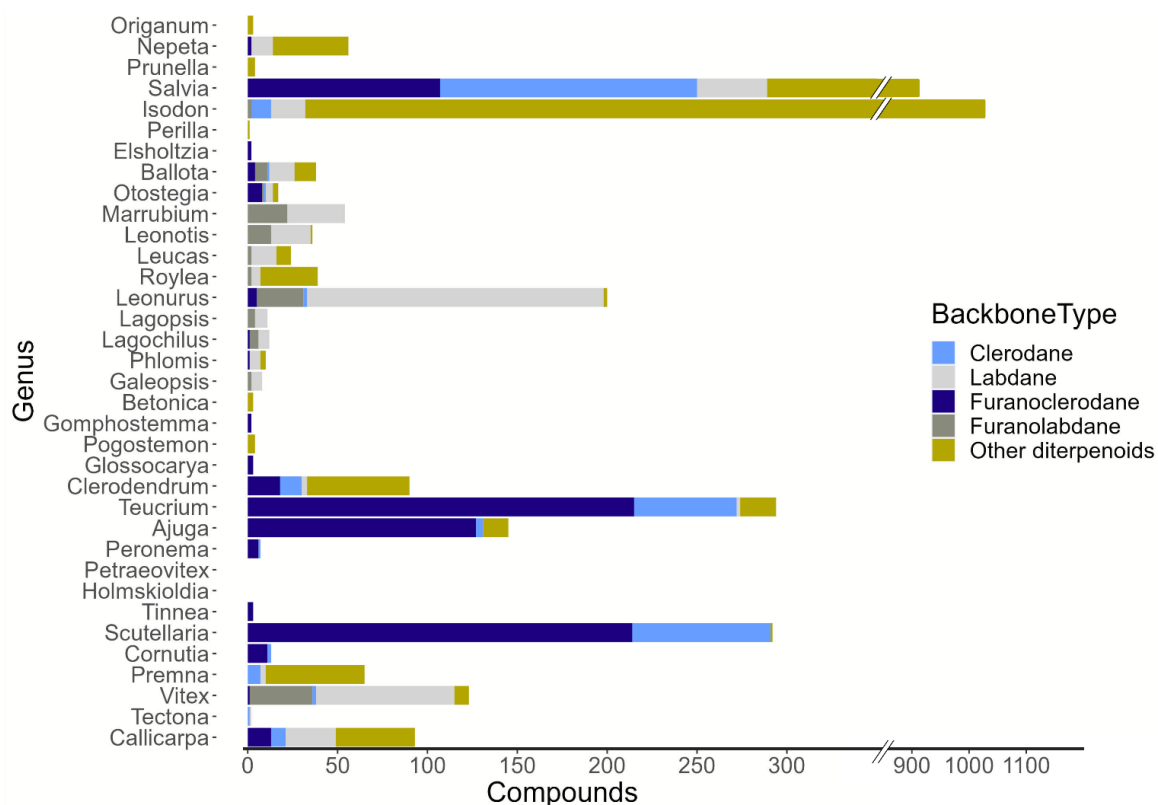


Figure S2.20. Lamiaceae diterpenoids by backbone type. Clerodane and labdanes were further split into furanoclerodanes/furanolabdanes vs clerodanes and labdanes lacking a furan, lactone, or furanofuran moiety. To additionally, gauge diterpenoid diversity in a given species all other diterpenoids were also included as the ‘other’ category.

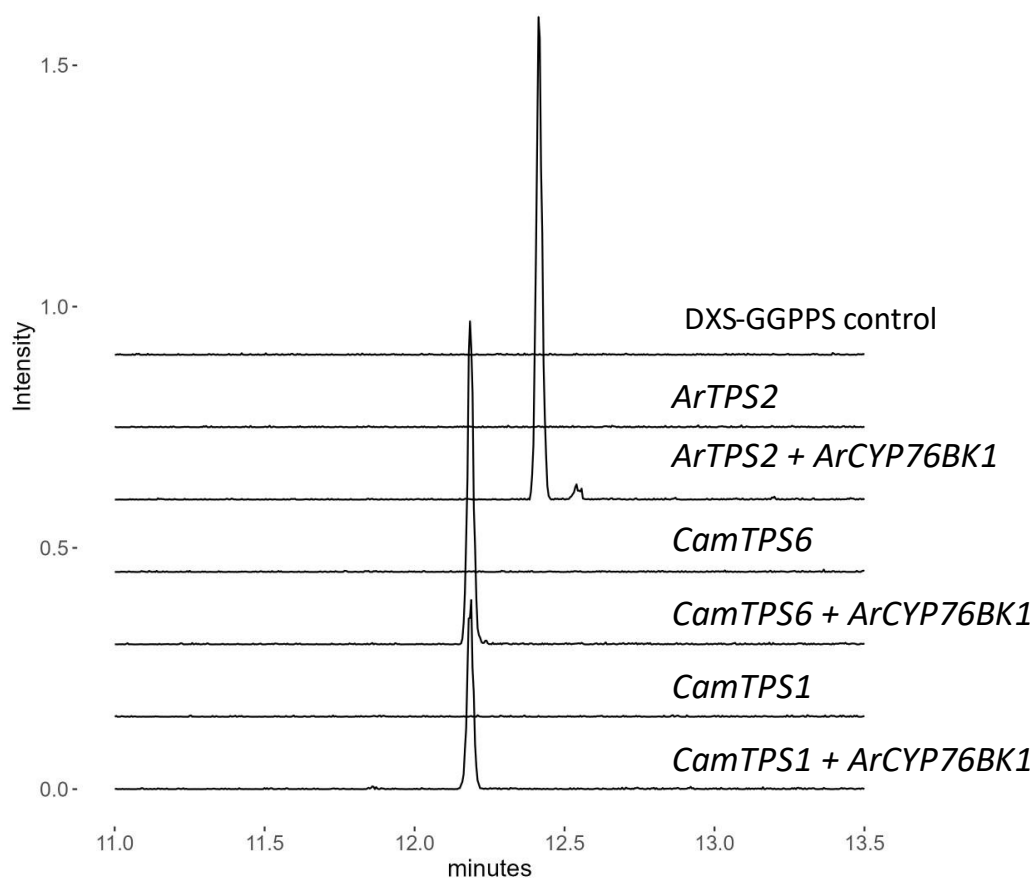


Figure S2.21. Extracted ion chromatograms (286 m/z) of different *N. benthamiana* infiltration extractions. 286 m/z is the parent ion of furan containing products for a (+)-CPP *ent*-CPP and isokolavenol product. *ArTPS2* is a TPS for isokolavenol and the TPS native to where the *ArCYP76BK1* product is from acting as a positive control. *CamTPS1* produces *ent*-CPP and *CamTPS6* produces (+)-CPP. All samples with the *ArCYP76BK1* produced a new 286 m/z product regardless of backbone. The *ent*- and (+)-CPP likely overlapped in elution due to only differing by the decalin core's stereochemistry

Chapter 3: An opportunistic cytochrome P450 in *Ajuga reptans* oxidizes diverse labdane-derived diterpenes

(Manuscript in preparation)

Nicholas J. Schlecht^{1,2}, Trine B. Andersen^{1,2}, Ryn Van Winkle¹, Daniel Holmes³, Björn R. Hamberger^{1,2}

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

²DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, USA

³Department of Chemistry, Michigan State University, East Lansing MI, USA

Abstract

Advancing our understanding of specialized metabolism and improving metabolic engineering tools are critical to developing new agrochemicals and pharmaceuticals. Promiscuous cytochromes P450 (CYPs) offer valuable tools for metabolic engineers by enabling chemical modifications of diverse metabolites, including non-native pathways. *Ajuga reptans*, a Lamiaceae (mint) family species, produces diverse oxidized clerodane diterpenoids, indicating a potential enrichment in diterpene-active CYPs. This observation led to a transcriptome-guided investigation of CYPs led to the discovery of ArCYP736A358, the first identified member of the CYP736 family with activity towards cyclized diterpenes. Specifically, ArCYP736A358 was found to oxidize the diterpenes isokolavelool, (-)-kolavelool, (+)-manool, sclareol, manoyl oxide, sandaracopimaradiene, and *ent*-kaurene. At present, we have elucidated the products as epoxy-isokolavelool, epoxy-manoyl oxide, and 3-hydroxy-*ent*-kaurene. The broad substrate range, and regioselectivity add ArCYP736A358 as a valuable biocatalyst to our growing synthetic biology toolbox.

Introduction

Approximately two thirds of agrochemicals and pharmaceuticals are natural products, natural product derived, mimics, or carry a natural product pharmacore.^{1,2} Effective natural product synthesis is crucial for both direct use and derivatization. Extraction from natural sources is possible but often limited by low concentrations or specific growth requirements, making alternative approaches preferable.³⁻⁵ While great strides have been made in organic chemistry to synthetically produce chiral structures, there remain limitations, largely due to the numerous stereogenic centers.^{6,7} Therefore,

a practical approach would be to use biotechnology to generate products or precursors for semisynthesis, such as commercial artemisinin.⁸

Terpenoids, the largest group of specialized metabolites, comprise over 90,000 unique structures with diverse bioactivities.⁹ Terpene synthases modularly cyclize substrates to generate terpene backbones.^{10,11} These backbones become terpenoids upon oxidation, typically by cytochromes P450 (CYPs), a step that substantially increases terpenoid diversity and functionality while enabling other modifications by additional enzyme families.¹²

A growing emphasis has been placed on enzyme promiscuity, as it can efficiently generate complex metabolite pools.¹¹ Enzyme promiscuity can be divided into various forms including conditional, catalytic, substrate and product promiscuity.¹³ Many of these characteristics have been reported in CYPs. For example, CYP76M8 in *Oryza sativa* oxidized several diterpenes.¹⁴ In *Salvia miltiorrhiza*, CYP76AH3 yields three separate products from ferruginol, and *Ajuga* CYP76BK1 orthologs produced a mixture of a furan and lactone moieties on multiple labdane derived backbones.^{15,16}

To expand available tools for metabolic engineering, we pursued CYPs in *Ajuga reptans*, a mint species with oxidized clerodane diterpenes reported in leaf tissue. These clerodanes demonstrate insect anti-feedant properties that are reportedly enhanced by their epoxides.^{17–19} The clerodane synthase ArTPS2, necessary for clerodane metabolism, is expressed in leaves where oxidized clerodanes are found, suggesting that other biosynthetic genes might also be leaf-expressed.²⁰ We screened the leaf transcriptome for candidate P450s to oxidize clerodanes. Of the candidates, ArCYP736A358 was found to be active with the substrate isokolavelool and a range of

structurally related diterpenes. None of these substrates are known to be natively found in *Ajuga* nor could be identified from plant extracts. While the native function of ArCYP736A358 remains unknown, our findings have uncovered a new biotechnology multi-tool.

Results & Discussion:

Phylogenetic analyses identifies diverse CYP candidates.

A previous study of the *Ajuga reptans* CYP76 family identified the promiscuous enzyme CYP76BK1.¹⁶ In this study we broadened the search to a more diverse subset of CYPs. Full length CYPs identified in the *A. reptans* transcriptome generated in the mint genome project were extracted and used to generate a phylogenetic tree with 147 candidates (Figure S3.1).²¹ A subset of CYPs was chosen based on known terpene metabolism while selecting diverse representatives.^{12,22} In total, 22 of the candidate CYPs were cloned and tested. Nineteen belonged to the CYP71 clan, with the remaining belonging to the CYP51, CYP72, and CYP86 clans as seen in Figure 3.1.

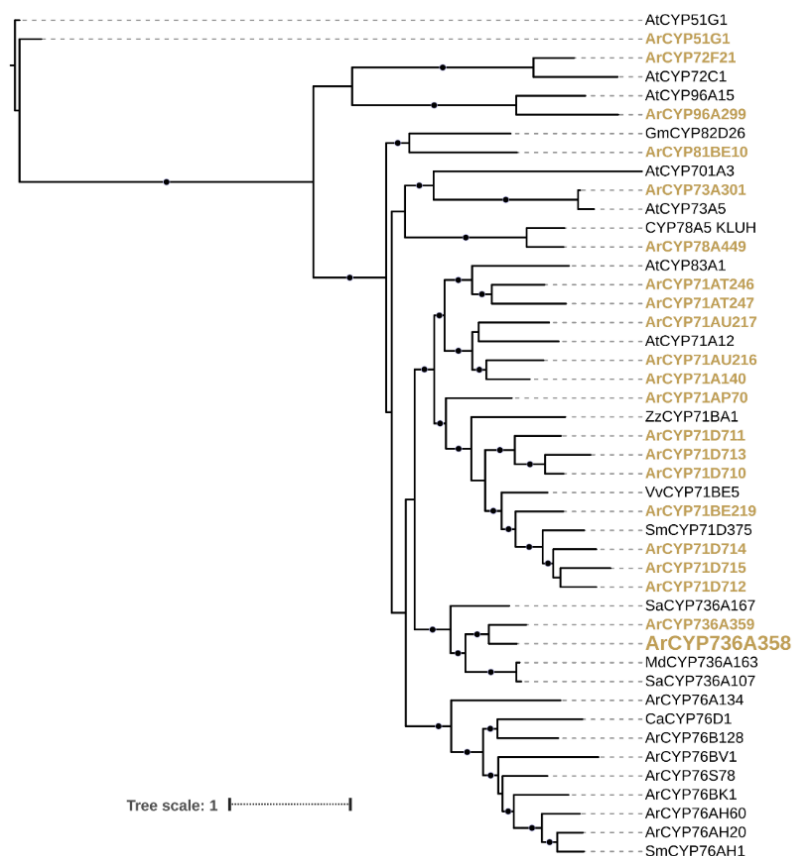


Figure 3.1. Maximum likelihood tree of *A. reptans* CYP candidates and references. Gold CYPs represent *A. reptans* cloned candidates assayed against the isokolavenol and isokolavelool backbone. Black labels are functionally characterized CYPs. Circles depict a bootstrap value of 70% or greater (1000 bootstraps). Candidates' gene IDs and corresponding references can be found in Table S3.1 and S3.2.

Screening candidate CYPs for activity against isokolavenol and isokolavelool

Candidate CYPs were transiently expressed in *Nicotiana benthamiana* via *Agrobacterium tumefaciens* infiltrations. All infiltrations used strains of *Agrobacterium* carrying coding sequences for 1-deoxy-D-xylulose 5-phosphate synthase (PbDXS) and geranylgeranyl diphosphate synthase (PbGGPPS) to enhance terpene yields.²³ Isokolavenyl diphosphate synthase (ArTPS2) was co-infiltrated with and without the CYPs to assess their role in clerodane metabolism. The infiltrations depended on

endogenous *N. benthamiana* enzymes to cleave the diphosphate—a process supported by previous studies.^{16,20,24,25} Additional enzyme combinations also included sclareol synthase, which paired with ArTPS2 generates isokolavelool, to evaluate potential substrate promiscuity amongst the CYPs.

Leaf extracts were analyzed by gas chromatography-mass spectrometry (GC-MS) to identify new products. Among the CYP candidates, only ArCYP736A358 exhibited activity (Figure 3.2). ArCYP736A358 did not generate a new product when paired with ArTPS2. However, when co-expressed with ArTPS2 and the class I diTPS sclareol synthase (SsSS), a new peak with a 306 *m/z* parent ion (Figure S3.2) appeared. This result demonstrated that ArCYP736A358 utilizes isokolavelool, not isokolavenol, as a substrate resulting in oxidation of the backbone. The product was scaled-up by bulk vacuum infiltration and purified by flash chromatography. Subsequent NMR analysis identified this product as epoxy-isokolavelool (**1**) (Figure S3.3). The epoxide replaces the terminal alkene of isokolavelool, potentially explaining why the isokolavelool was an available substrate and isokolavenol was not. No SsSS ortholog has been found in *A. reptans*, implying that this oxidation reaction may not be the native function of ArCYP736A358.

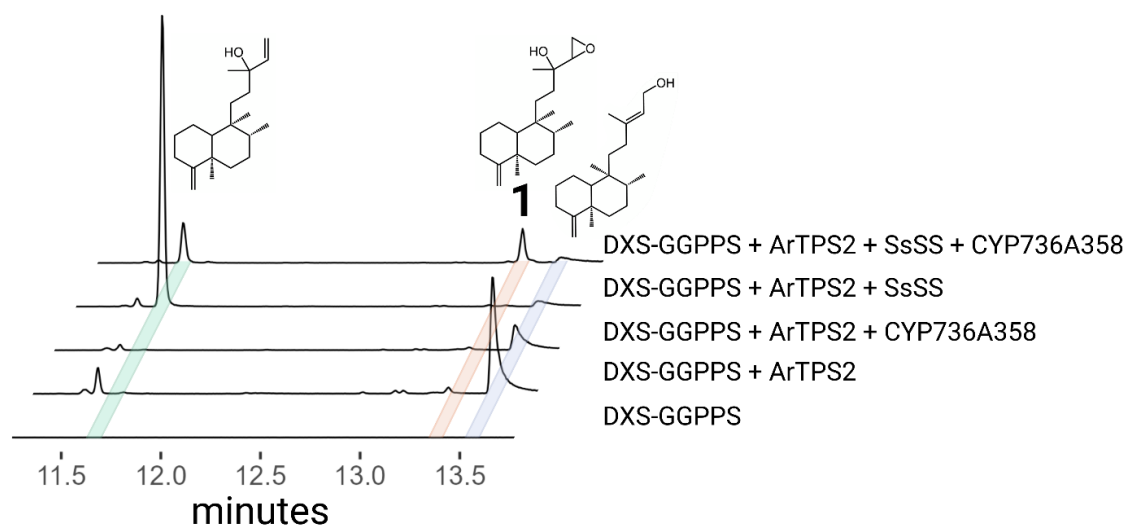


Figure 3.2. GCMS demonstrates oxidation of isokolavelool: Aligned total ion chromatogram showing isokolavenol, isokolavelool and the oxidation of isokolavelool into **1**. Mass spectrum for compound **1** can be found in Figure S3.2. Blue, green, and red highlight isokolavenol, isokolavelool, and product **1**, respectively.

Assaying the promiscuity of ArCYP736A358

To explore the promiscuity of ArCYP736A358, additional assays were conducted with a broader set of labdane-derived diterpenes (Figure 3.3). The new set of assays mirrored the design of the previous experiment, with each class II TPS control evaluated alone, with a class I TPS partner, and both combinations paired with ArCYP736A358. Four additional class II TPSs were selected to assess potential substrate promiscuity within ArCYP736A358: CamTPS2 produces (-)-kolavenyl diphosphate, CamTPS1 *ent*-copalyl diphosphate, CamTPS6 (+)-copalyl diphosphate, and CfTPS2 labda-13-en-8-ol diphosphate. Again, we relied on the endogenous *N. benthamiana* enzymes to dephosphorylate these diphosphate products for detection on GC-MS.

SsSS was paired with each class II diTPS, yielding the same modified moiety as isokolavelool on the different backbones. Four additional class I diTPSs were partnered with a class II TPS to assess if ArCYP736A358 could oxidize diterpenes with broader

structural variation. The additional class I diTPSs were: ArTPS3, a manoyl oxide synthase, CamTPS12, an ent-kaurene synthase, LITPS4a, a sandaracopimaradiene synthase, and CamTPS9, a miltiradiene synthase. The diterpenes produced by these TPSs have an overlapping but varied structural architecture, providing an opportunity to glean insight on important features. Comparisons include stereochemistry variation, ring count, and type of alkenes available.

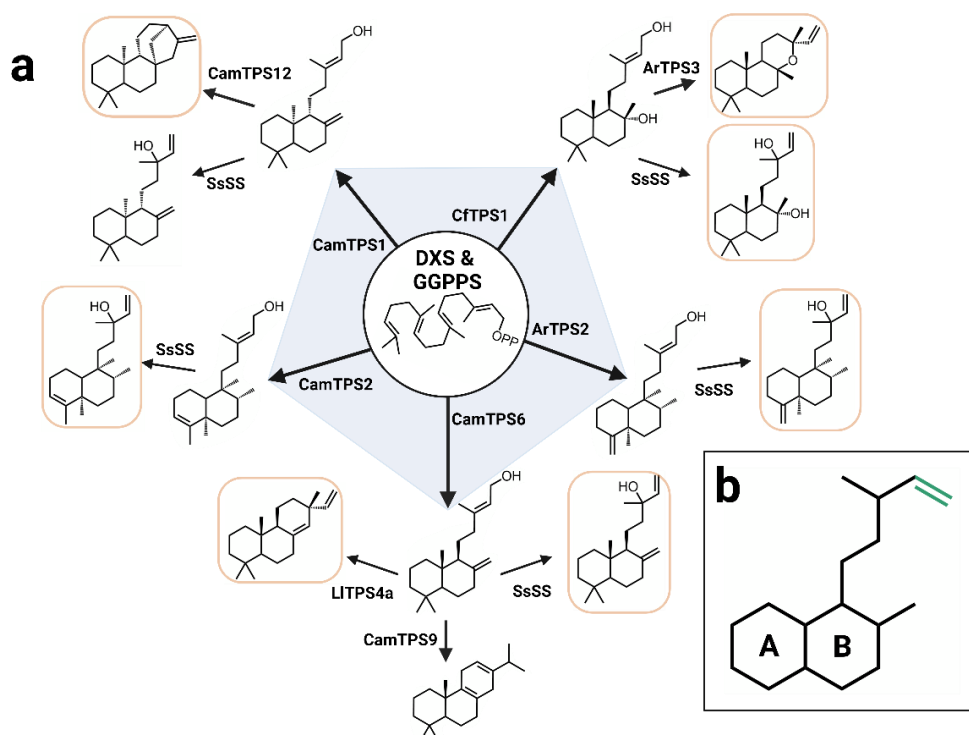


Figure 3.3. Metabolic network of labdane-derived terpenoids assayed with ArCYP736A358. **a.** The central node represents geranylgeranyl diphosphate (GGDP). All structures shown were assayed with ArCYP736A358. Arrows indicate the enzymes used to generate the various products. The orange circled structures were substrates for ArCYP736A358. **b.** highlights a core structure common to CYP736A358 substrates save for *ent*-kaurene which differs in the alkene placement due to structural rearrangements. The bicyclic decalin core has the rings labeled for further description.

ArCYP736A358 did not oxidize any dephosphorylated class II TPS product. In contrast, it generated products for six of the eight class II + class I TPS combinations.

(Figure 3.4). In addition to isokolavelool (ArTPS2+SsSS), *ent*-kaurene (CamTPS1+CamTPS12), manoyl oxide (CfTPS2+ArTPS3), (-)-kolavelool (CamTPS2+SsSS), (+)-manool (CamTPS6+SsSS), sclareol (CfTPS2+SsSS), and sandaracopimaradiene (CamTPS6+LITPS4a) were oxidized by ArCYP736A358. The extracted compounds from each combination were also derivatized to potentially identify non-volatile diterpenes. The only new product detected following derivatization was the oxidized sclareol, which is too polar for GC-MS (Figure 3.4e).

The *ent*-kaurene product, *ent*-kaurene-3-ol, (**2**) was confirmed by an available standard (Figure S3.4), while the manoyl oxide product was verified to be epoxy-manoyl oxide (**3**) by NMR spectroscopy (Figure S3.5). Based on mass spectra and the overlapping structural features of manoyl oxide and isokolavelool, the other products were tentatively assigned as epoxy-(-)-kolavelool (**4**), epoxy-(+)-manool (**5**), epoxy-sclareol (**6**), and epoxy-sandaracopimarene (**7**). Structures **5** and **7** are pending NMR analysis. Compounds **4** and **1** are structurally similar and **6** is challenging to purify due to its polarity limiting GC analyses. Miltiradiene (CamTPS6+CamTPS9) and *ent*-manool (CamTPS1+SsSS) were the class I products that were not substrates for ArCYP736A358. Among the different substrates for ArCYP736A358, *Ajuga* only presumably has the complete TPS pathway to produce *ent*-kaurene, precursor of ubiquitous gibberellin phytohormones. These results also suggest that ArCYP736A358 can utilize a wide range of diterpenes as reactants but prefers monosubstituted alkenes (Figure 3.3b).

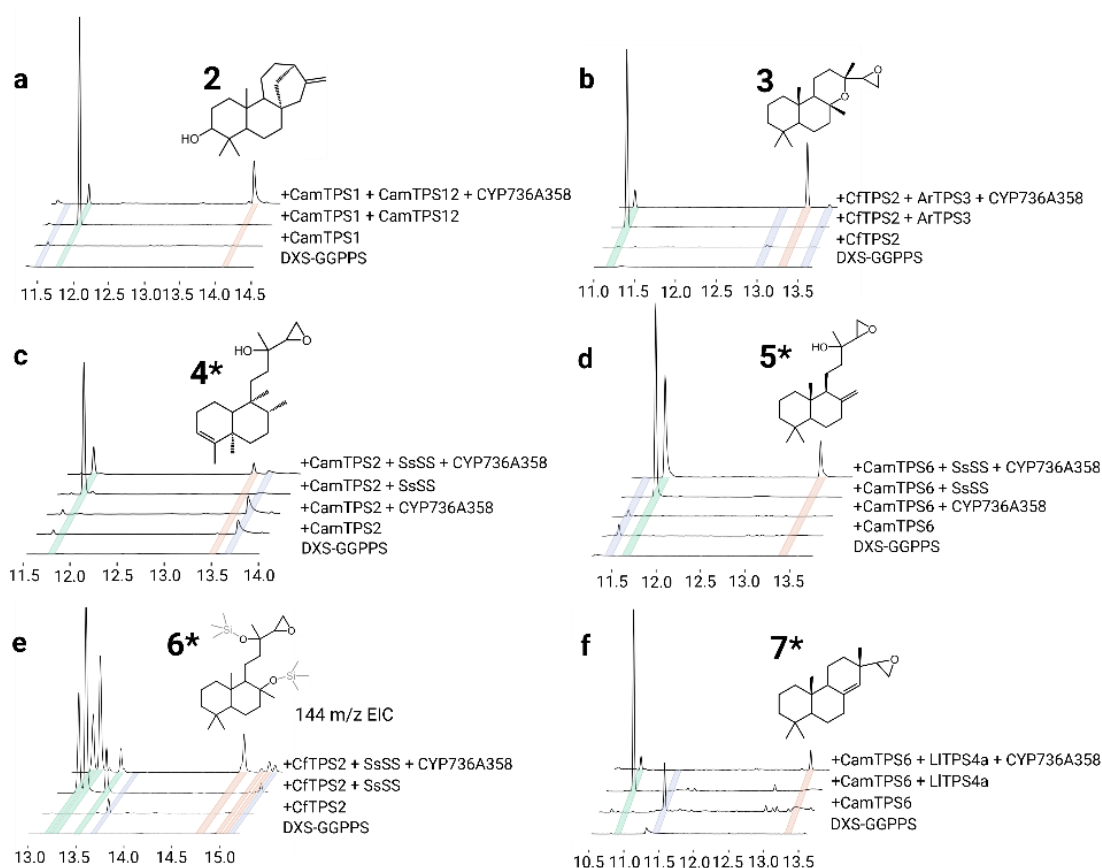


Figure 3.4. Aligned chromatograms of ArCYP736A358 products. All infiltrations included DXS-GGPPS, as a lone control or in combination with other enzymes. * Indicates a tentative structure. 4e was derivatized so that the trimethylsilylated product could be detected. Orange lines highlight the CYP product. Blue is the dephosphorylated class II TPS product. Green is the class II + class I TPS product. CYP736A358 product mass spectra are available in Figure S3.6.

Exploring the CYP736A358 products and the CYP736 family

Leaf tissue extracts from *A. reptans* showed no detectable ArCYP736A358 products, aligning with the lack of reported diTPSs for those pathways in *A. reptans*. These products could require certain stress responses to induce metabolite production or may not be found in our cultivar. To ascertain if these have been reported elsewhere, the Dictionary of Natural Products (DNP) was investigated (Table S3.3). No substructures corresponding to this study's reported products were identified in *Ajuga* species, further corroborating the absence *in planta*. Investigating the entire DNP

revealed *ent*-kaurene-3-ol has >500 substructures whereas the epoxide products ranged from 0-10. C3 hydroxylation of *ent*-kaurene and similar labdanoids is reported for other CYPs across diverse species, suggesting that this activity has emerged multiple times throughout plant evolution.^{26,27} In contrast, reports of epoxide-containing products were sparse in the DNP, with no substructures matching epoxy-kolavelool, epoxy-isokolavelool, or epoxy-(+)-manool and only a few reports for epoxy-manoyl oxide, epoxy-sclareol, and epoxy-sandaracopimarene substructures. Of those, two epoxy-sandaracopimarenes and one epoxy-sclareol are reported in the Nepetoideae subfamily of mints. These seven biosynthetic pathways are new-to-nature and three products were previously not accessible.

The CYP736 family emerged early in land plant's evolution, which is reflected in the range of substrates utilized by the few characterized enzymes.²⁸ Members of the CYP736A subfamily are involved in the biosynthesis of multiple phenolics, cyanogenic glucosides, the monoterpene thymoquinone, the sesquiterpene santalol, and hydroxylation of linear diterpene geranylinalool.^{29–34} These diverse activities indicate that the subfamily evolved a variety of functions, driving chemical diversification across plants. To investigate potential homologs of ArCYP736A358, transcriptomes from the mint genome project were mined to generate a phylogeny of the family, revealing 139 CYP736As (Figure S3.7), and illustrating an intriguing expansion in the subfamily.

ArCYP736A358 was found in a clade of nine other proteins and included representatives of six Lamiaceae subfamilies including the early diverging subfamily Callicarpoideae.^{21,25} This distribution suggests that an ancestral CYP736A358 existed in the earliest members of the family.

Conclusion:

This study identified ArCYP736A358, the first CYP736 to oxidize cyclized diterpenes. This enzyme exhibits remarkable substrate promiscuity, capable of oxidizing at least two clerodane backbones, two labdane backbones, an isopimarane, manoyl oxide, as well as ent-kaurene. While the biosynthesis of these products is novel, many of the products are fully new-to-nature. The native role of ArCYP736A358 remains unclear, but its promiscuity may facilitate the evolution of new chemistries.

Beyond plants, ArCYP736A358 will likely have other applications. The epoxide product's function is currently unknown, but the functional group is known for its reactivity and other epoxide-containing diterpenes are often noted for having unique bioactivities.^{35–37} Though we showed that ArCYP736A358 has a general preference towards mono-substituted alkenes, we only assayed a fraction of known diterpenoids with this moiety. There are at least 24 distinct diterpene backbones with monosubstituted alkenes in Lamiaceae alone and over 2500 labdane-derived diterpenoids with mono-substituted alkenes in the DNP (figure 3.3b).^{9,20} While we cannot ascertain the complete extent of ArCYP736A358s promiscuity, this enzyme will likely have biotechnology impacts in metabolic engineering in the future.

Experimental procedures

Mining Cytochromes P450 from mint transcriptomes

For all species, previously assembled transcriptomic data were used from the mint genome consortium (Table S3.4).²¹ Peptide sequences were predicted using Transdecoder (v. 5.5.0).³⁸ Peptide sequences for *A. reptans* Candidate P450s were initially filtered based on 45% identity and an E value greater than 1E-5 using BLASTP

against a set of reference sequences (Table S3.2) and a minimum length of 450 amino acids. When identifying orthologs of ArCYP736A358 specifically, the various mint transcriptomes were blasted (blastp) against ArCYP736A358, and a more stringent filter for blast hits was applied, requiring 50% identity to ArCYP736A358, a minimum length of 450 amino acids, and the longest isoforms only.

Phylogenetic trees

Reference sequences used in all protein phylogenies were obtained from GenBank and can be found in Table S3.2. Individual phylogenies varied in reference and candidates chosen, but followed the same construction. ClustalOmega (version 1.2.4; default parameters) was used to build the multiple sequence alignment that was subsequently input into RAxML (version 8.2.12; Model = protgammaauto; Algorithm = a) to generate a phylogenetic tree with 1000 bootstraps.^{39,40} The phylogenetic trees were rendered using the Interactive Tree of Life (version 6.5.2).⁴¹

Plant material and cloning

Ajuga reptans was obtained from the commercial nursery Pixies Garden and grown in a greenhouse under ambient photoperiod, grow lights for the winter, and 24 °C day/17 °C night temperatures. Candidate genes were PCR-amplified from leaf cDNA, inserted into the cloning vector pJet v1.2, and coding sequences were sequence-verified against the respective gene models. Constructs were subsequently cloned into the binary expression vector pEAQ-HT and once again sequence verified.

Transient expression for functional characterization in N. benthamiana

N. benthamiana plants were grown for 5 weeks in a controlled growth room under 12h light and 12h dark cycle at (22°C) before infiltration. Constructs for co-expression were separately transformed into the *Agrobacterium tumefaciens* strain, GV3101. Twenty mL cultures were grown overnight at 30 °C in LB with 50 µg/mL kanamycin, 50 µg/mL rifampicin, and 30 µg/mL gentamicin. Cultures were collected by centrifugation and washed twice with 10 mL water. Cells were resuspended and diluted to an OD₆₀₀ of 1.0 in 200 µM acetosyringone/water and incubated at 30 °C for 1–2 h. Separate cultures were mixed in a 1:1 ratio for each combination of enzymes, and 4- or 5-week-old plants were infiltrated with a 1 mL syringe into the underside (abaxial side) of *N. benthamiana* leaves. All gene constructs were co-infiltrated with two genes encoding rate-limiting steps in the upstream (MEP) pathway: *P. barbatulus* 1-deoxy-D-xylulose-5-phosphate synthase (*PbDXS*) and GGPP synthase (*PbGGPPS*) to boost production of the diterpene precursor GGDP.²³ Plants were returned to the controlled growth room (22 °C, 12 h diurnal cycle) for 5 days. Approximately 200 mg of fresh weight, infiltrated leaves was extracted with 1 mL hexane overnight at room temperature. Plant material was collected by centrifugation, and the organic phase was removed for GC-MS analysis.

Extracting Natural Product information from the Dictionary of Natural Products

Various parameters were used for datamining the DNP (v33.1). Initially, searches were limited to “*Ajuga**” using the ‘Biological Source’ category. To search specifically for the substrates, the SMILES of the various substrates were loaded into the drawn structure field. To capture the most diversity and reduce false negatives, the

compounds were flattened to remove stereochemistry. Lastly, as the aim was to also capture if *A. reptans* had the available substrates as well as products, alkenes were altered to be single bonds, as some product derivatives will have lost a given double bond. The chosen information to export was the chemical name, molecular formula, biological source, type of organism, type of compound, and SMILES. This information was transferred into Table S3.4 in addition to the image of the substructure searched for and the Biological source information. This process was repeated when expanding to search all organisms with some modifications. The 'Biological Source' category was changed to indicate that any source is valid. Substrates were no longer of interest, so the ArCYP736A358 product structures were included instead and were input such that double bonds could be any bond type (single or double, any stereochemistry) to account for potential oxidative changes of downstream products. This information was similarly exported and transferred into Table S3.4.

Plant metabolite extractions

Leaves from *A. reptans* were harvested for metabolite analysis. The leaves were frozen in liquid nitrogen, crushed, and extracted for three hours in ethyl acetate. Leaf material was collected by centrifugation and the organic phase was removed and concentrated for GC-MS analysis.

GC-MS analysis

All GC-MS analyses were performed at Michigan State University's (MSU) Mass Spectrometry and Metabolomics Core Facility utilizing an Agilent 7890 A GC with an Agilent VF-5ms column (30 m × 250 µm × 0.25 µm, with 10 m EZ-Guard) and an Agilent 5975 C detector. The inlet was set to 250 °C splitless injection of 1 µL and He carrier

gas (1 mL/min), and the detector was given a 4 min solvent delay. All non-derivatized assays and tissue analyses used the following method: temperature ramp start 40 °C, hold 1 min, 40 °C/min to 200 °C, hold 4.5 min, 20 °C/min to 240 °C, 10 °C/min to 280 °C, 40 °C/min to 320 °C, and hold 5 min. MS scan range was set to 40–400 *m/z*.

Deconvolution of spectra was done utilizing AMDIS.⁴²

To derivatize samples 100 µl of an extract was dried down and resuspended in 90 µl *N,O*-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) with 1% Trimethylchlorosilane (TMCS). Samples were immediately capped and incubated at 60 °C overnight (~12 h) before being run on GC-MS the following day. The GC-MS acquisition method had the same settings as non-derivatized assays but had a 4.9 min solvent delay.

Product scale-up and NMR spectroscopy

For NMR analysis, production in the *N. benthamiana* system was scaled up to 1 L mixtures of the requisite *A. tumefaciens* strains mixed to equal ODs. A vacuum-infiltration system was used to infiltrate *A. tumefaciens* strains into whole *N. benthamiana* plants, with approximately 30-60 plants used for each enzyme combination, which followed the same combinations used when screening. After 5 days, all leaf tissue was harvested and extracted overnight in approximately 600 mL hexane at room temperature. The extract was concentrated by a rotary evaporator. Each product was purified by silica gel flash column chromatography with a mobile phase of 98% hexane/2% ethyl acetate. Samples required a follow up Pasteur pipette column running pure hexane as a final clean up. NMR spectra were measured in Michigan State University's Max T. Rogers NMR Facility on a Bruker Avance NEO 800 MHz or 600 MHz spectrometer equipped with a helium cooled TCI cryoprobe or a Prodigy TCI

cryoprobe, respectively, using CDCl₃ as the solvent. CDCl₃ peaks were referenced to 7.26 and 77.00 ppm for ¹H and ¹³C spectra, respectively.

Accession numbers:

Relevant accession numbers can be found on Table S3.1 and Table S3.2

Author contributions

NS, TA and BH conceived the study. NS wrote the manuscript with contributions from TA and BH. Pathways and constructs were designed by NS. Terpene analysis in transient assays was performed by NS and TA. DH assisted critically with NMR analysis.

Acknowledgements

We would like to thank Britta Hamberger for assistance in maintaining plant material. We would also like to thank Emily Lanier and Abby Bryson who provided some necessary plasmids for this study. P450 annotation was kindly provided by David Nelson (University of Tennessee). We would like to thank Drs. Cassandra Johnny and Anthony Schillmiller of Michigan State University's Mass Spectrometry and Metabolomics Core Facility for their help in obtaining and interpreting GC-MS data, and the Max T. Rogers NMR Facility for their help in obtaining NMR data.

We collectively acknowledge that Michigan State University occupies the ancestral, traditional, and contemporary Lands of the Anishinaabeg – Three Fires Confederacy of Ojibwe, Odawa, and Potawatomi peoples. In particular, the University resides on Land ceded in the 1819 Treaty of Saginaw. We recognize, support, and advocate for the sovereignty of Michigan's twelve federally-recognized Indian nations, for historic Indigenous communities in Michigan, for Indigenous individuals and

communities who live here now, and for those who were forcibly removed from their Homelands. By offering this Land Acknowledgement, we affirm Indigenous sovereignty and will work to hold Michigan State University more accountable to the needs of American Indian and Indigenous peoples.

Funding

This work was supported in part through computational resources and services provided by the Institute for Cyber-Enabled Research at Michigan State University, the US Department of Energy Great Lakes Bioenergy Research Center Cooperative Agreement DE-SC0018409, startup funding from the Department of Biochemistry and Molecular Biology, and support from AgBioResearch (MICL02454). B.H. gratefully acknowledges a generous endowment from James K. Billman, Jr. N.S. is supported by a fellowship from Michigan State University under the predoctoral Training Program in Plant Biotechnology for Health and Sustainability (T32-GM110523) from the National Institute of General Medical Sciences of the National Institutes of Health. B.H. is in part supported by the National Science Foundation under Grant Number 1737898. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- (1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Last 25 Years. *J. Nat. Prod.* **2007**, *70* (3), 461–477. <https://doi.org/10.1021/np068054v>.
- (2) Sparks, T. C.; Hahn, D. R.; Garizi, N. V. Natural Products, Their Derivatives, Mimics and Synthetic Equivalents: Role in Agrochemical Discovery. *Pest Management Science* **2017**, *73* (4), 700–715. <https://doi.org/10.1002/ps.4458>.
- (3) Liu, G.-M.; Fang, W.-S.; Qian, J.-F.; Lan, H. Distribution of Paclitaxel and Its Congeners in *Taxus Mairei*. *Fitoterapia* **2001**, *72* (7), 743–746. [https://doi.org/10.1016/S0367-326X\(01\)00302-1](https://doi.org/10.1016/S0367-326X(01)00302-1).
- (4) Larson, T. R.; Branigan, C.; Harvey, D.; Penfield, T.; Bowles, D.; Graham, I. A. A Survey of Artemisinic and Dihydroartemisinic Acid Contents in Glasshouse and Global Field-Grown Populations of the Artemisinin-Producing Plant *Artemisia Annua* L. *Industrial Crops and Products* **2013**, *45*, 1–6. <https://doi.org/10.1016/j.indcrop.2012.12.004>.
- (5) Jeong, W. T.; Lim, H. B. A UPLC-ESI-Q-TOF Method for Rapid and Reliable Identification and Quantification of Major Indole Alkaloids in *Catharanthus Roseus*. *Journal of Chromatography B* **2018**, *1080*, 27–36. <https://doi.org/10.1016/j.jchromb.2018.02.018>.
- (6) Quílez del Moral, J. F.; Pérez, Á.; Barrero, A. F. Chemical Synthesis of Terpenoids with Participation of Cyclizations plus Rearrangements of Carbocations: A Current Overview. *Phytochem Rev* **2020**, *19* (3), 559–576. <https://doi.org/10.1007/s11101-019-09646-8>.
- (7) Thomas, W. P.; Pronin, S. V. New Methods and Strategies in the Synthesis of Terpenoid Natural Products. *Acc. Chem. Res.* **2021**, *54* (6), 1347–1359. <https://doi.org/10.1021/acs.accounts.0c00809>.
- (8) Kung, S. H.; Lund, S.; Murarka, A.; McPhee, D.; Paddon, C. J. Approaches and Recent Developments for the Commercial Production of Semi-Synthetic Artemisinin. *Front. Plant Sci.* **2018**, *9*. <https://doi.org/10.3389/fpls.2018.00087>.
- (9) Taylor & Francis Group. *Dictionary of Natural Products* (v33.1). <https://dnp.chemnetbase.com/chemical/ChemicalSearch.xhtml?dswid=6473> (accessed 2024-11-22).
- (10) Zerbe, P.; Bohlmann, J. Plant Diterpene Synthases: Exploring Modularity and Metabolic Diversity for Bioengineering. *Trends in Biotechnology* **2015**, *33* (7), 419–428. <https://doi.org/10.1016/j.tibtech.2015.04.006>.

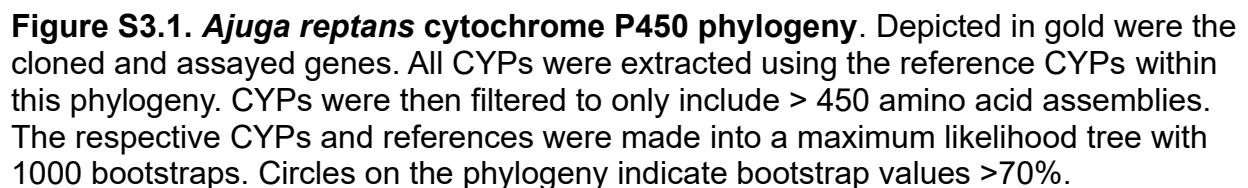
- (11) Lanier, E. R.; Andersen, T. B.; Hamberger, B. Plant Terpene Specialized Metabolism: Complex Networks or Simple Linear Pathways? *The Plant Journal* **2023**, *114* (5), 1178–1201. <https://doi.org/10.1111/tpj.16177>.
- (12) Pateraki, I.; Heskes, A. M.; Hamberger, B. Cytochromes P450 for Terpene Functionalisation and Metabolic Engineering. In *Biotechnology of Isoprenoids*; Schrader, J., Bohlmann, J., Eds.; Springer International Publishing: Cham, 2015; pp 107–139. https://doi.org/10.1007/10_2014_301.
- (13) Kreis, W.; Munkert, J. Exploiting Enzyme Promiscuity to Shape Plant Specialized Metabolism. *Journal of Experimental Botany* **2019**, *70* (5), 1435–1445. <https://doi.org/10.1093/jxb/erz025>.
- (14) Wang, Q.; Hillwig, M. L.; Okada, K.; Yamazaki, K.; Wu, Y.; Swaminathan, S.; Yamane, H.; Peters, R. J. Characterization of CYP76M5–8 Indicates Metabolic Plasticity within a Plant Biosynthetic Gene Cluster *. *Journal of Biological Chemistry* **2012**, *287* (9), 6159–6168. <https://doi.org/10.1074/jbc.M111.305599>.
- (15) Guo, J.; Ma, X.; Cai, Y.; Ma, Y.; Zhan, Z.; Zhou, Y. J.; Liu, W.; Guan, M.; Yang, J.; Cui, G.; Kang, L.; Yang, L.; Shen, Y.; Tang, J.; Lin, H.; Ma, X.; Jin, B.; Liu, Z.; Peters, R. J.; Zhao, Z. K.; Huang, L. Cytochrome P450 Promiscuity Leads to a Bifurcating Biosynthetic Pathway for Tanshinones. *New Phytologist* **2016**, *210* (2), 525–534. <https://doi.org/10.1111/nph.13790>.
- (16) Schlecht, N. J.; Lanier, E. R.; Andersen, T. B.; Brose, J.; Holmes, D.; Hamberger, B. R. CYP76BK1 Orthologs Catalyze Furan and Lactone Ring Formation in Clerodane Diterpenoids across the Mint Family. *The Plant Journal* **2024**, *120* (3), 984–997. <https://doi.org/10.1111/tpj.17031>.
- (17) Klein Gebbinck, E. A.; Jansen, B. J. M.; de Groot, A. Insect Antifeedant Activity of Clerodane Diterpenes and Related Model Compounds. *Phytochemistry* **2002**, *61* (7), 737–770. [https://doi.org/10.1016/S0031-9422\(02\)00174-7](https://doi.org/10.1016/S0031-9422(02)00174-7).
- (18) Coll, J.; Tandrón, Y. A. Neo-Clerodane Diterpenoids from Ajuga: Structural Elucidation and Biological Activity. *Phytochem Rev* **2008**, *7* (1), 25–49. <https://doi.org/10.1007/s11101-006-9023-3>.
- (19) Li, R.; Morris-Natschke, S. L.; Lee, K.-H. Clerodane Diterpenes: Sources, Structures, and Biological Activities. *Nat. Prod. Rep.* **2016**, *33* (10), 1166–1226. <https://doi.org/10.1039/C5NP00137D>.
- (20) Johnson, S. R.; Bhat, W. W.; Bibik, J.; Turmo, A.; Hamberger, B.; Consortium, E. M. G.; Hamberger, B. A Database-Driven Approach Identifies Additional Diterpene Synthase Activities in the Mint Family (Lamiaceae). *Journal of Biological Chemistry* **2019**, *294* (4), 1349–1362. <https://doi.org/10.1074/jbc.RA118.006025>.

- (21) Boachon, B.; Buell, C. R.; Crisovan, E.; Dudareva, N.; Garcia, N.; Godden, G.; Henry, L.; Kamileen, M. O.; Kates, H. R.; Kilgore, M. B.; Lichman, B. R.; Mavrodiev, E. V.; Newton, L.; Rodriguez-Lopez, C.; O'Connor, S. E.; Soltis, D.; Soltis, P.; Vaillancourt, B.; Wiegert-Rininger, K.; Zhao, D. Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. *Molecular Plant* **2018**, *11* (8), 1084–1096. <https://doi.org/10.1016/j.molp.2018.06.002>.
- (22) Xu, J.; Wang, X.; Guo, W. The Cytochrome P450 Superfamily: Key Players in Plant Development and Defense. *Journal of Integrative Agriculture* **2015**, *14* (9), 1673–1686. [https://doi.org/10.1016/S2095-3119\(14\)60980-1](https://doi.org/10.1016/S2095-3119(14)60980-1).
- (23) Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Møller, B. L.; Hamberger, B. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angewandte Chemie International Edition* **2016**, *55* (6), 2142–2146. <https://doi.org/10.1002/anie.201510650>.
- (24) Chen, X.; Berim, A.; Dayan, F. E.; Gang, D. R. A (–)-Kolavenyl Diphosphate Synthase Catalyzes the First Step of Salvinorin A Biosynthesis in *Salvia Divinorum*. *Journal of Experimental Botany* **2017**, *68* (5), 1109–1122. <https://doi.org/10.1093/jxb/erw493>.
- (25) Hamilton, J. P.; Godden, G. T.; Lanier, E.; Bhat, W. W.; Kinser, T. J.; Vaillancourt, B.; Wang, H.; Wood, J. C.; Jiang, J.; Soltis, P. S.; Soltis, D. E.; Hamberger, B.; Buell, C. R. Generation of a Chromosome-Scale Genome Assembly of the Insect-Repellent Terpenoid-Producing Lamiaceae Species, *Callicarpa Americana*. *GigaScience* **2020**, *9* (9), g1aa093. <https://doi.org/10.1093/gigascience/g1aa093>.
- (26) Wang, Q.; Hillwig, M. L.; Wu, Y.; Peters, R. J. CYP701A8: A Rice Ent-Kaurene Oxidase Paralog Diverted to More Specialized Diterpenoid Metabolism. *Plant Physiology* **2012**, *158* (3), 1418–1425. <https://doi.org/10.1104/pp.111.187518>.
- (27) Mafu, S.; Jia, M.; Zi, J.; Morrone, D.; Wu, Y.; Xu, M.; Hillwig, M. L.; Peters, R. J. Probing the Promiscuity of Ent-Kaurene Oxidases via Combinatorial Biosynthesis. *Proceedings of the National Academy of Sciences* **2016**, *113* (9), 2526–2531. <https://doi.org/10.1073/pnas.1512096113>.
- (28) Nelson, D. R.; Ming, R.; Alam, M.; Schuler, M. A. Comparison of Cytochrome P450 Genes from Six Plant Genomes. *Tropical Plant Biol.* **2008**, *1* (3), 216–235. <https://doi.org/10.1007/s12042-008-9022-1>.
- (29) Mai, T. D.; Kim, H. M.; Park, S. Y.; Ma, S. H.; Do, J. H.; Choi, W.; Jang, H. M.; Hwang, H. B.; Song, E. G.; Shim, J. S.; Joung, Y. H. Metabolism of Phenolic Compounds Catalyzed by Tomato CYP736A61. *Enzyme and Microbial Technology* **2024**, *176*, 110425. <https://doi.org/10.1016/j.enzmictec.2024.110425>.

- (30) Sircar, D.; Gaid, M. M.; Chizzali, C.; Reckwell, D.; Kaufholdt, D.; Beuerle, T.; Broggini, G. A. L.; Flachowsky, H.; Liu, B.; Hänsch, R.; Beerhues, L. Biphenyl 4-Hydroxylases Involved in Aucuparin Biosynthesis in Rowan and Apple Are Cytochrome P450 736A Proteins. *Plant Physiology* **2015**, *168* (2), 428–442. <https://doi.org/10.1104/pp.15.00074>.
- (31) Takos, A. M.; Knudsen, C.; Lai, D.; Kannangara, R.; Mikkelsen, L.; Motawia, M. S.; Olsen, C. E.; Sato, S.; Tabata, S.; Jørgensen, K.; Møller, B. L.; Rook, F. Genomic Clustering of Cyanogenic Glucoside Biosynthetic Genes Aids Their Identification in *Lotus Japonicus* and Suggests the Repeated Evolution of This Chemical Defence Pathway. *The Plant Journal* **2011**, *68* (2), 273–286. <https://doi.org/10.1111/j.1365-313X.2011.04685.x>.
- (32) Kim, E.; Kim, M.; Oh, M.-K. Whole-Cell Bioconversion for Producing Thymoquinone by Engineered *Saccharomyces Cerevisiae*. *Enzyme and Microbial Technology* **2024**, *178*, 110455. <https://doi.org/10.1016/j.enzmictec.2024.110455>.
- (33) Wang, Y.; Gong, X.; Li, F.; Zuo, S.; Li, M.; Zhao, J.; Han, X.; Wen, M. Optimized Biosynthesis of Santalenes and Santalols in *Saccharomyces Cerevisiae*. *Appl Microbiol Biotechnol* **2021**, *105* (23), 8795–8804. <https://doi.org/10.1007/s00253-021-11661-9>.
- (34) Li, J.; Halitschke, R.; Li, D.; Paetz, C.; Su, H.; Heiling, S.; Xu, S.; Baldwin, I. T. Controlled Hydroxylations of Diterpenoids Allow for Plant Chemical Defense without Autotoxicity. *Science* **2021**, *371* (6526), 255–260. <https://doi.org/10.1126/science.abe4713>.
- (35) Pineschi, M. Asymmetric Ring-Opening of Epoxides and Aziridines with Carbon Nucleophiles. *European Journal of Organic Chemistry* **2006**, *2006* (22), 4979–4988. <https://doi.org/10.1002/ejoc.200600384>.
- (36) Gomes, A. R.; Varela, C. L.; Tavares-da-Silva, E. J.; Roleira, F. M. F. Epoxide Containing Molecules: A Good or a Bad Drug Design Approach. *European Journal of Medicinal Chemistry* **2020**, *201*, 112327. <https://doi.org/10.1016/j.ejmech.2020.112327>.
- (37) Kaur, B.; Singh, P. Epoxides: Developability as Active Pharmaceutical Ingredients and Biochemical Probes. *Bioorganic Chemistry* **2022**, *125*, 105862. <https://doi.org/10.1016/j.bioorg.2022.105862>.
- (38) Haas, B. *Transdecoder (v. 5.5.0)*. GitHub. <https://github.com/TransDecoder/TransDecoder/wiki/Home> (accessed 2024-11-22).
- (39) *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega* | *Molecular Systems Biology*. <https://www.embopress.org/doi/full/10.1038/msb.2011.75> (accessed 2024-11-22).

- (40) Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, 30 (9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- (41) Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v6: Recent Updates to the Phylogenetic Tree Display and Annotation Tool. *Nucleic Acids Research* **2024**, 52 (W1), W78–W82. <https://doi.org/10.1093/nar/gkae268>.
- (42) NIST. *Automated Mass Spectral Deconvolution and Identification System [Software]*. National Institute of Standards and Technology. <https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdis> (accessed 2024-11-22).

Supplemental tables are provided with an external folder containing the various tables.



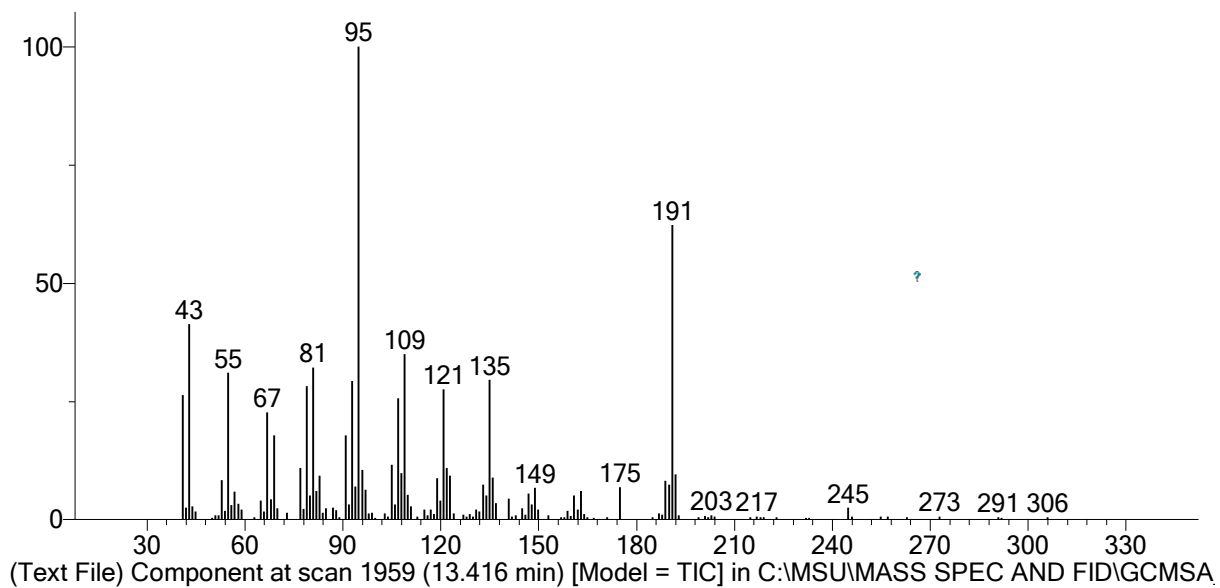
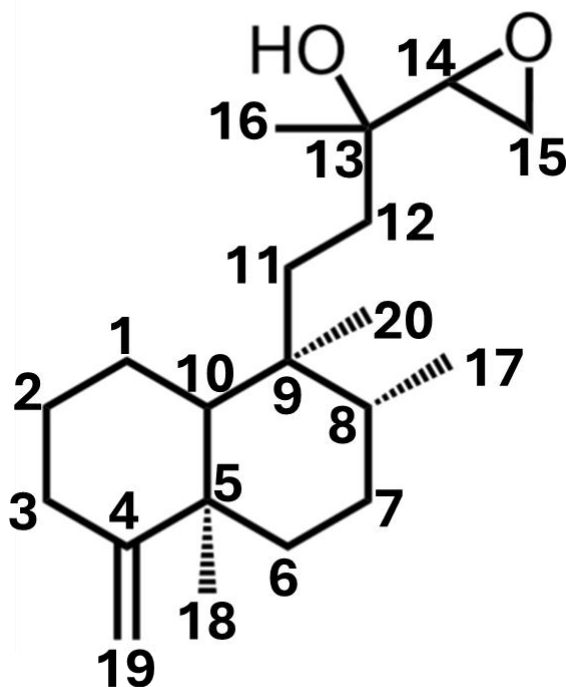


Figure S3.2. mass spectrum of epoxy-isokolavelool. The *N. benthamiana* infiltration including DXS-GGPPS, ArTPS2, SsSS, and ArCYP43 produced a peak at 13.41 minutes with a parent ion of 306 m/z. The mass spectrum was deconvoluted via AMDIS.



Atom	δ (ppm)	Atom	δ (ppm)
1 C	21.67	11 C	30.85
H'	1.54	H'	1.45
H''	1.46	H''	1.34
2 C	28.83	12 C	31.6
H'	1.89	H'	1.43
H''	1.32	H''	1.26
3 C	33.09	13 C	69.28
H'	2.11	14 C	57.92
H''	2.31	H	2.91
4 C	160.74	15 C	44.27
5 C	40.04	H'	2.75
6 C	37.35	H''	2.86
H'	1.62	16 C	26.17
H''	1.55	H3	1.3
7 C	27.46	17 C	15.93
H'	1.49	H3	0.82
H''	1.48	18 C	20.85
H	1.44	19 C	102.47
9 C	38.92	H'	4.53
10 C	48.51	H''	4.53
H	1.07	20 C	18.35
		H3	0.77

Figure S3.3a. Epoxy-isokolavelool chemical shift assignments. The structure for epoxy-isokolavelool (1) was deduced from (b) ^1H NMR spectrum, (c) COSY spectrum, (d) HSQC spectrum, (e) HMBC spectrum, (f) ^{13}C NMR spectrum. CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry was assigned based on the configuration of the precursor isokolavenyl diphosphate.

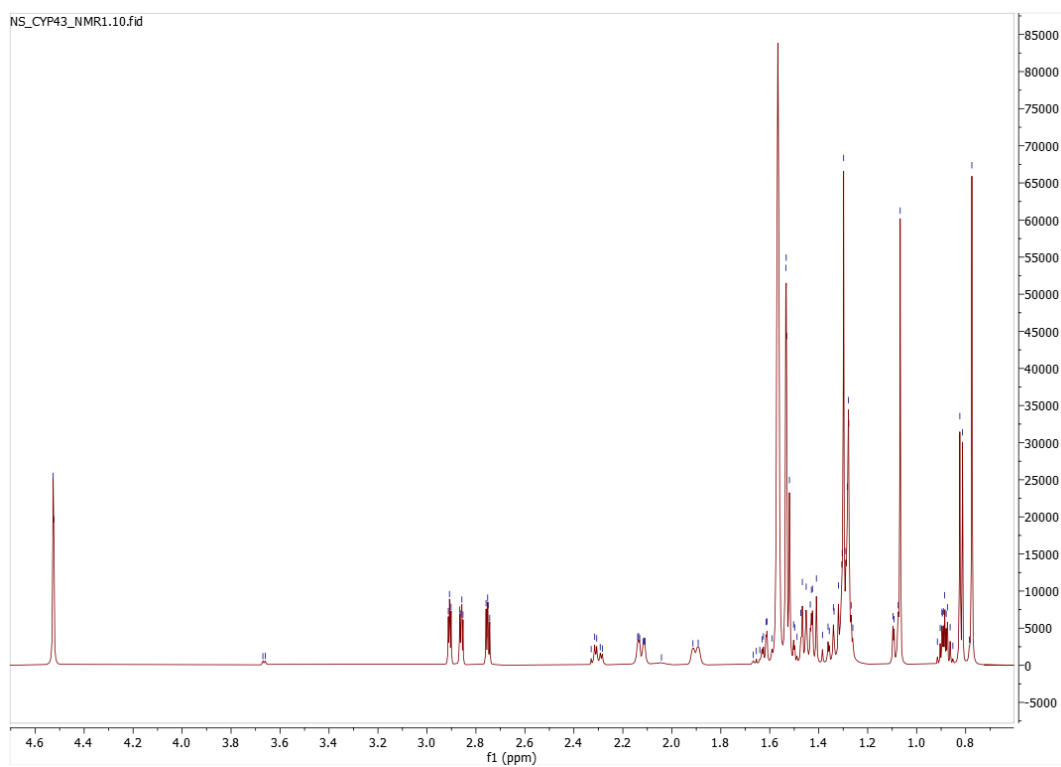


Figure S3.3b. ^1H NMR spectrum.

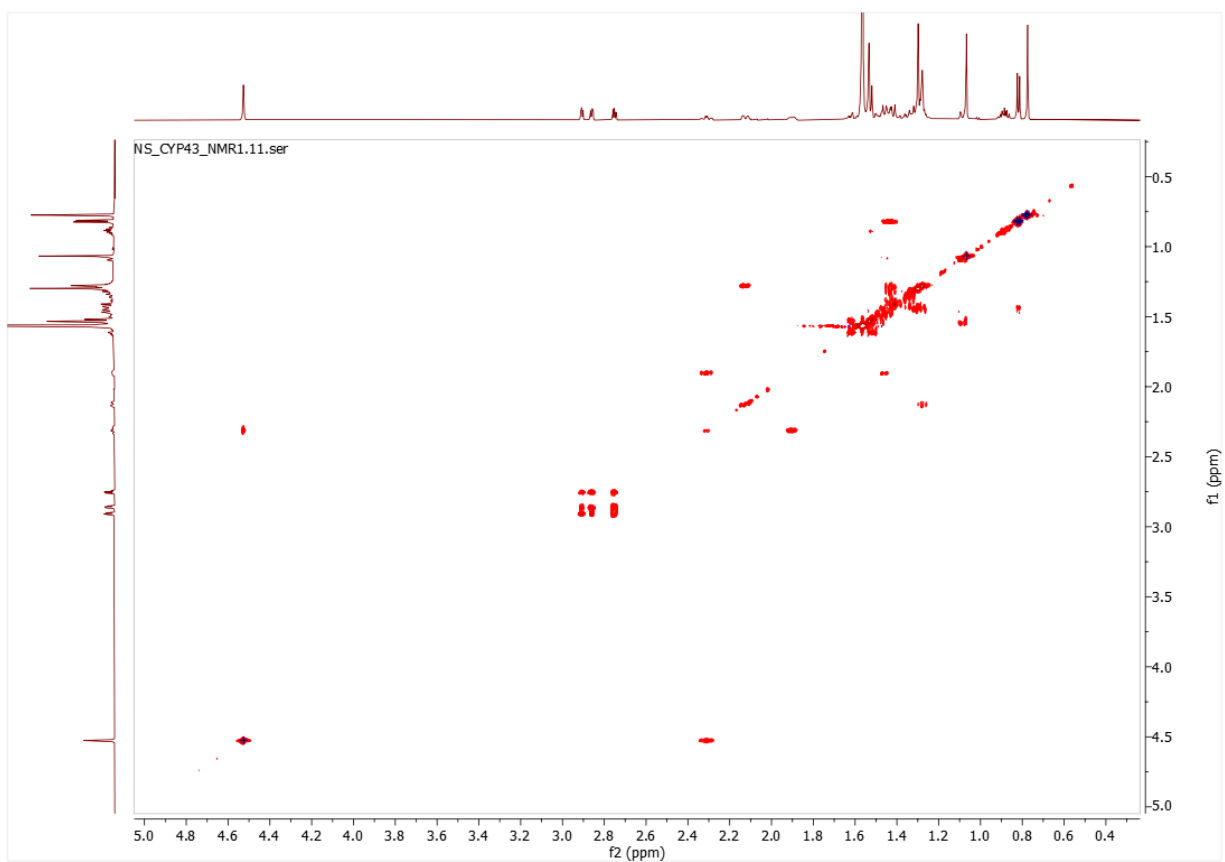


Figure S3.3c. COSY spectrum.

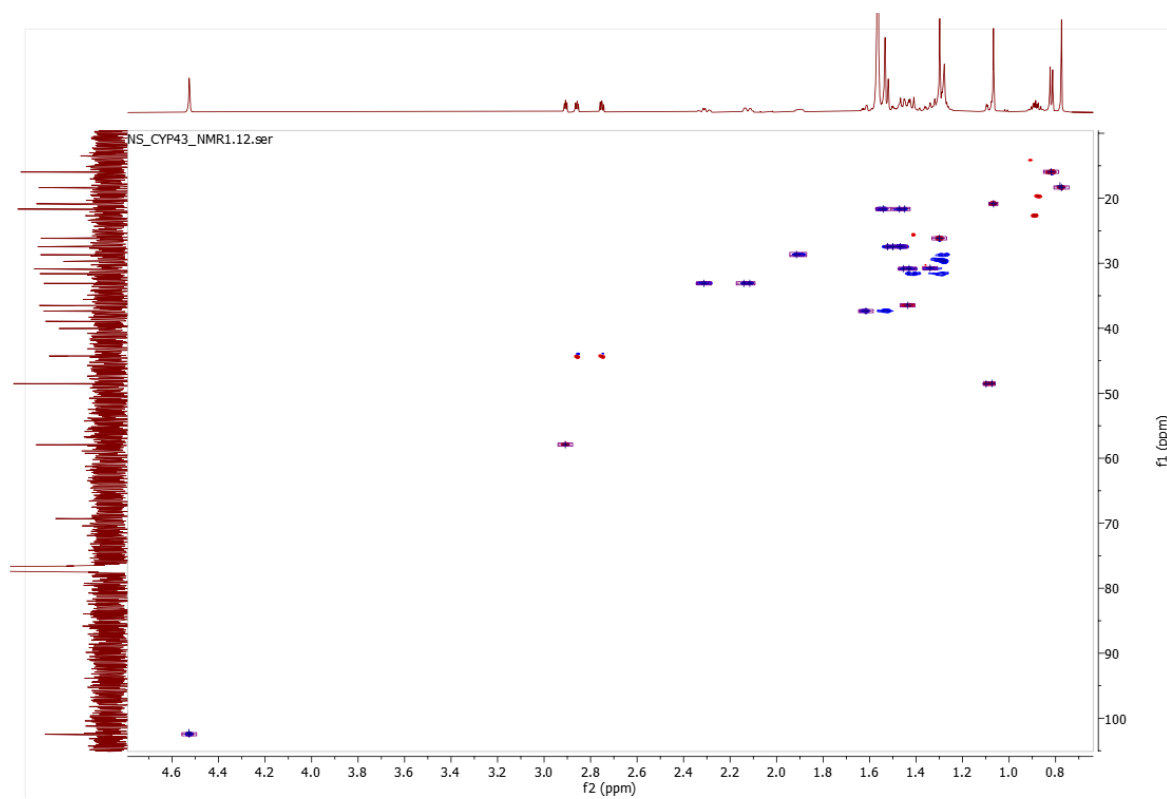


Figure S3.3d. HSQC spectrum.

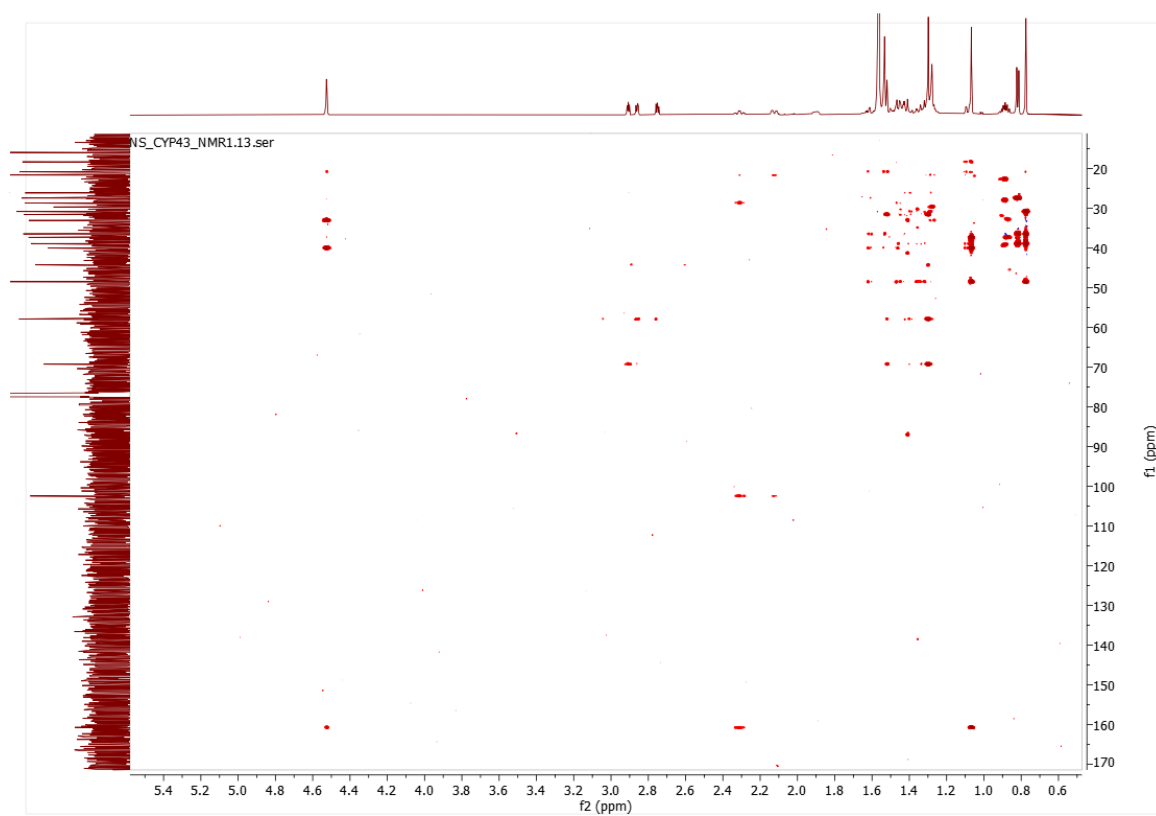


Figure S3.3e. HMBC spectrum.

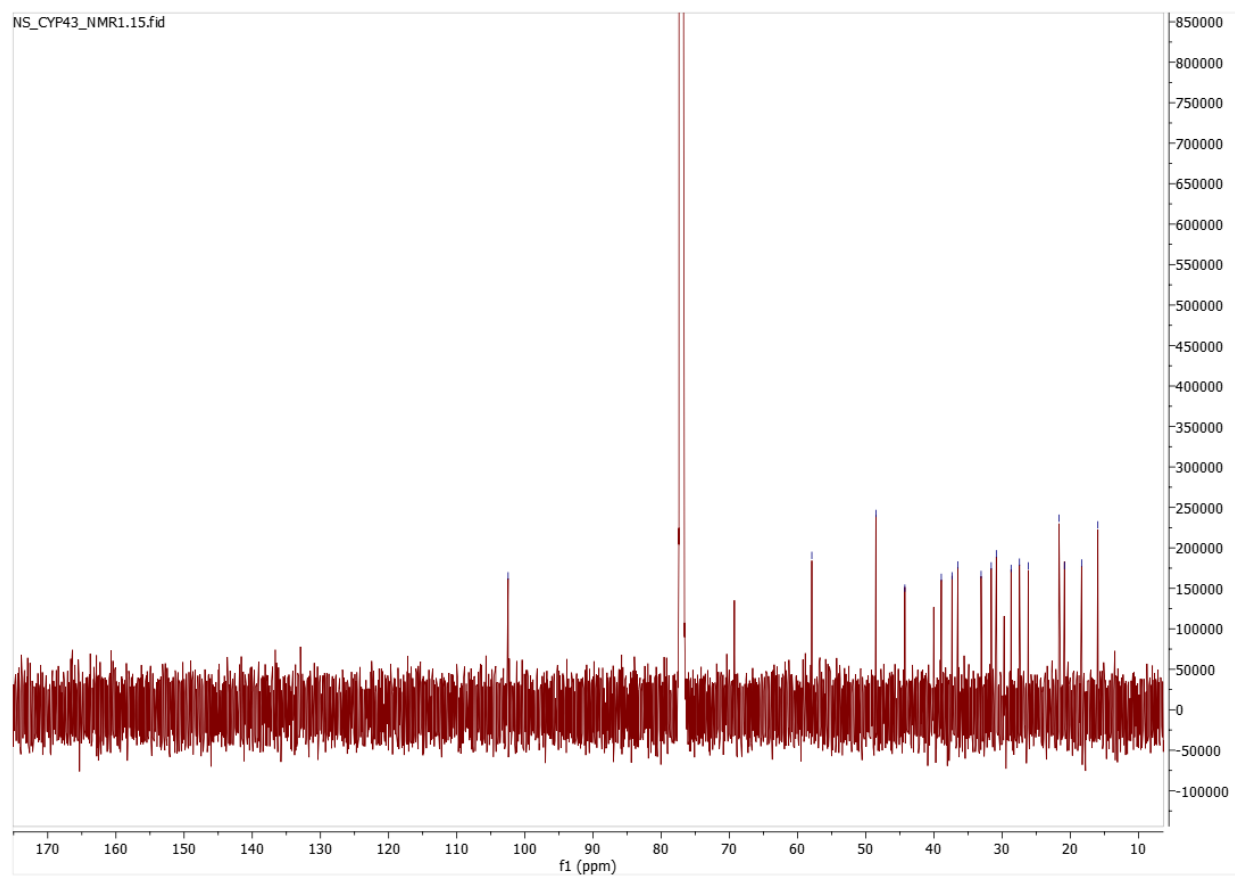


Figure S3.3f. ^{13}C NMR spectrum.

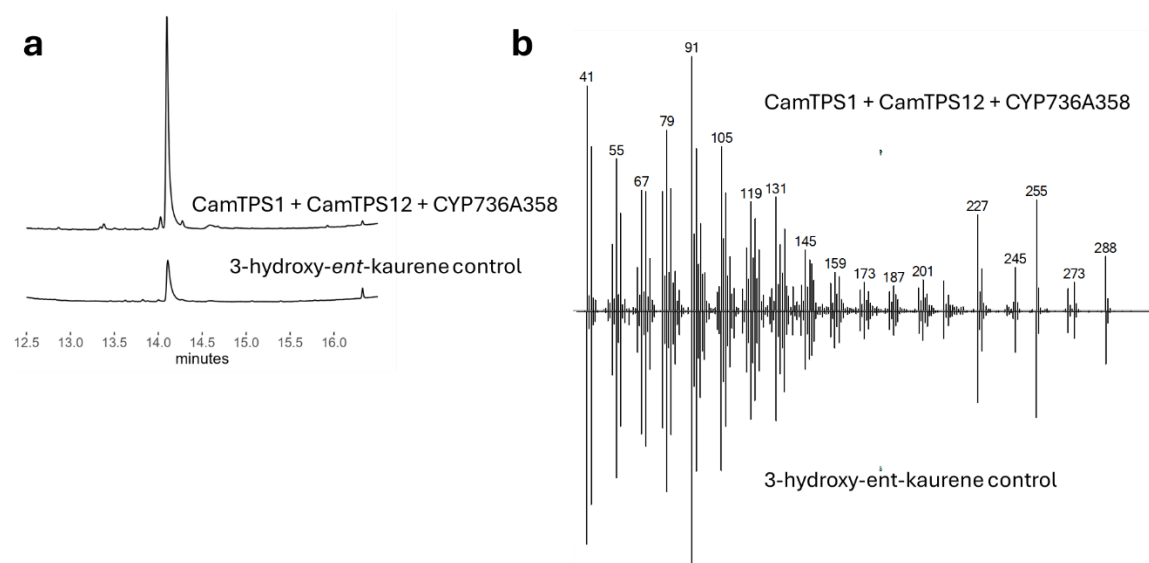
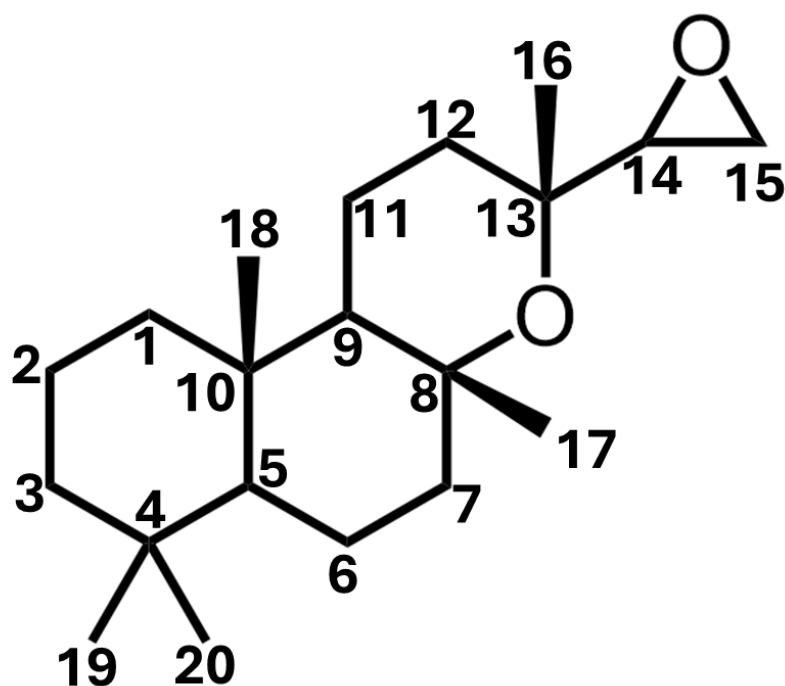


Figure S3.4. GC-MS verification of 3-hydroxy-ent-kaurene by an available standard. An NMR-verified 3-hydroxy-ent-kaurene sample was obtained from a purified sorghum CYP product (data not shown) and served as standard for validation. a. Shows the aligned chromatograms whereas b. shows the corresponding mass spectra of both the 3-hydroxy-ent-kaurene control and the CYP736A358 product.



Atom	δ (ppm)	Atom	δ (ppm)
1 C	39.04	11 C	14.98
H'	0.91	H'	1.52
H''	1.62	H''	1.62
2 C	18.58	12 C	33.53
H'	1.45	H'	1.67
H''	1.62	H''	1.52
3 C	42.12	13 C	70.49
H'	1.15	14 C	59.78
H''	1.41	H	2.83
4 C	33.26	15 C	43.84
5 C	56.31	H'	2.77
H	0.98	H''	2.67
6 C	19.8	16 C	24.57
H'	1.27	H3	1.23
H''	1.66	17 C	24.94
7 C	42.79	H3	1.29
H'	1.4	18 C	15.6
H''	1.75	H3	0.79
8 C	74.98	19 C	33.33
9 C	57.03	H3	0.87
H	1.23	20 C	21.28
10 C	36.87	H3	0.81

Figure S3.5a. Epoxy-manoyl oxide chemical shift assignments. The structure for epoxy-isokolavelool (1) was deduced from (b) ^1H NMR spectrum, (c) COSY spectrum, (d) HSQC spectrum, (e) HMBC spectrum, and (f) ^{13}C NMR spectrum. CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry was assigned based on the configuration the precursor manoyl oxide

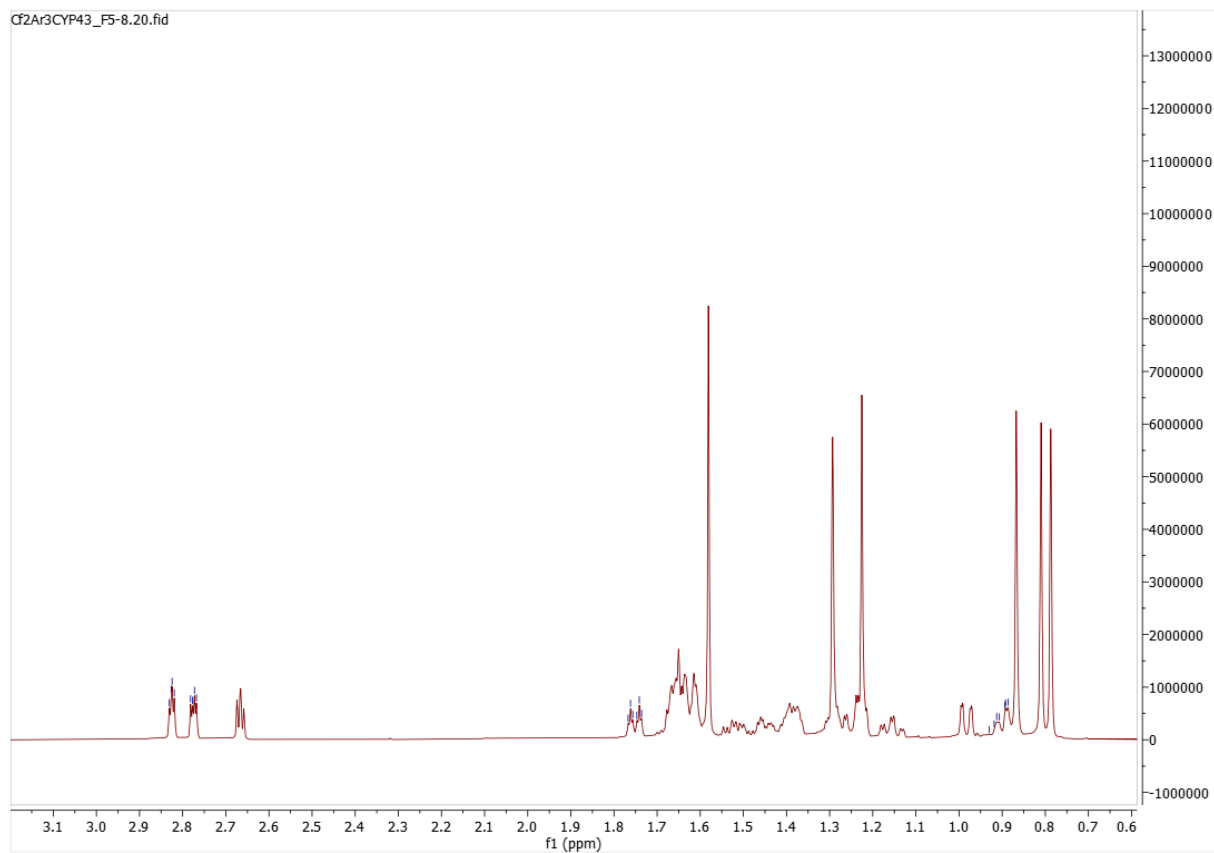


Figure S3.5b. ^1H NMR spectrum.

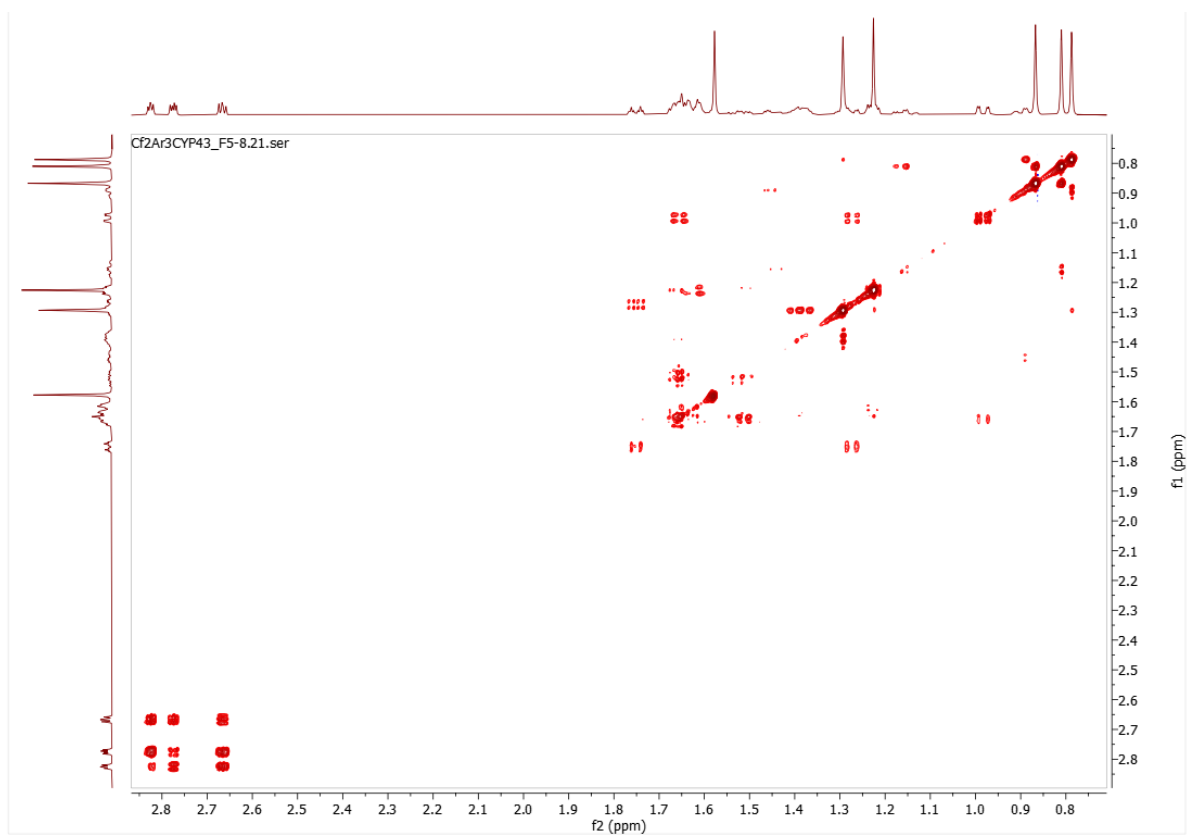


Figure S3.5c. COSY spectrum.

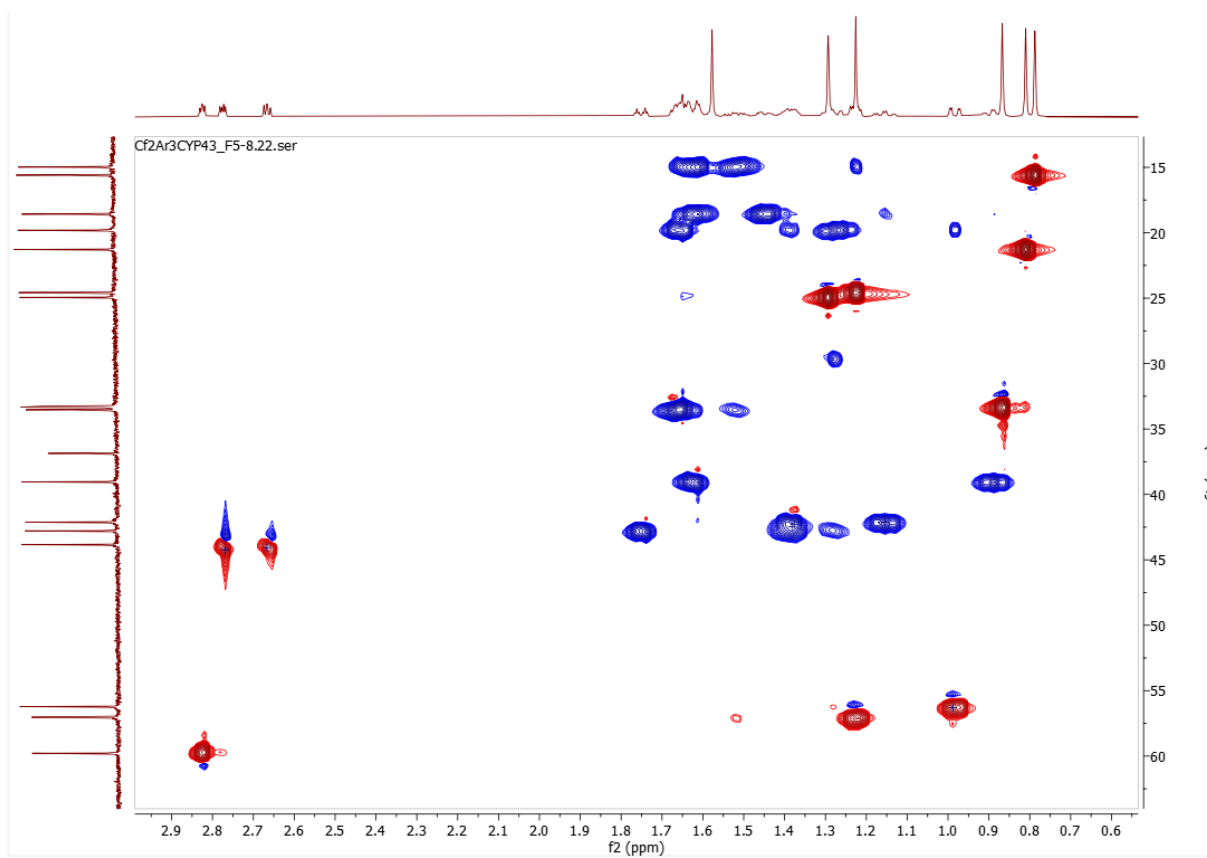


Figure S3.5d. HSQC spectrum.

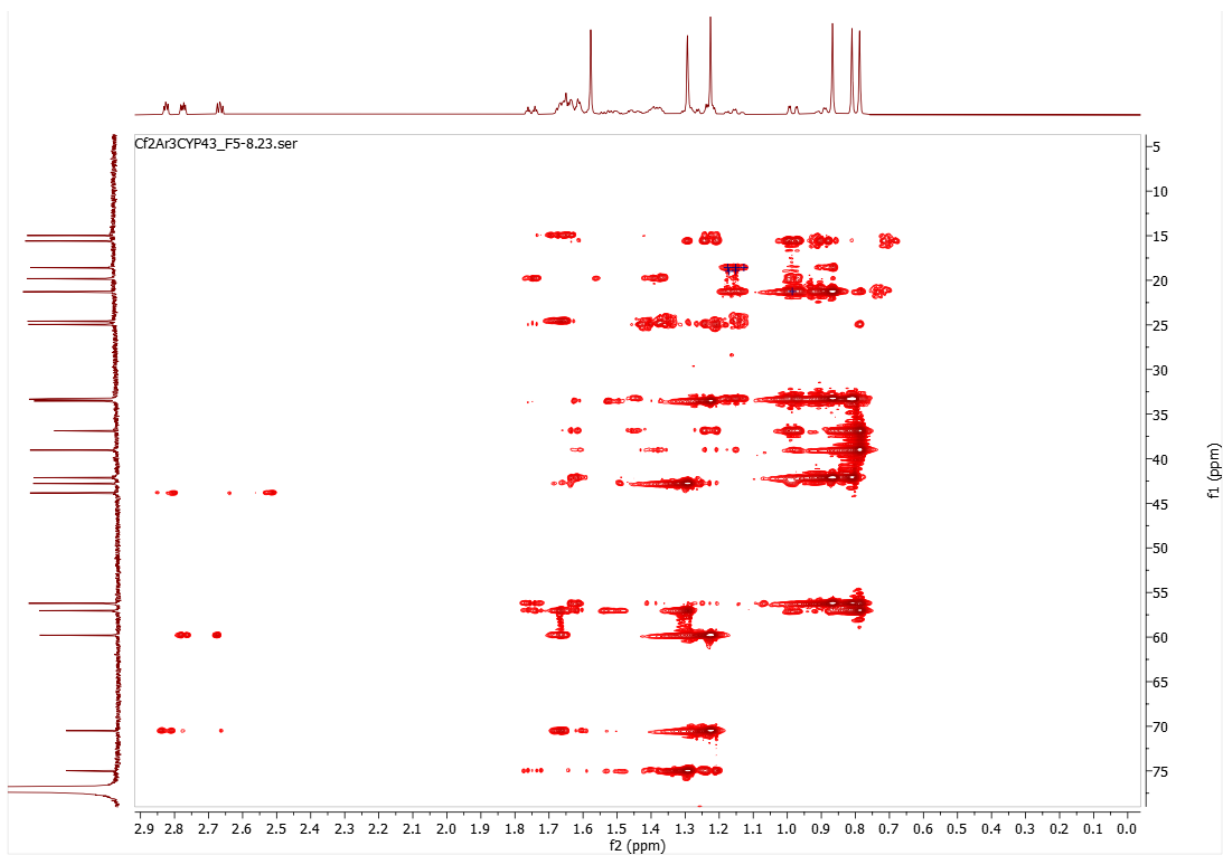


Figure S3.5e. HMBC spectrum.

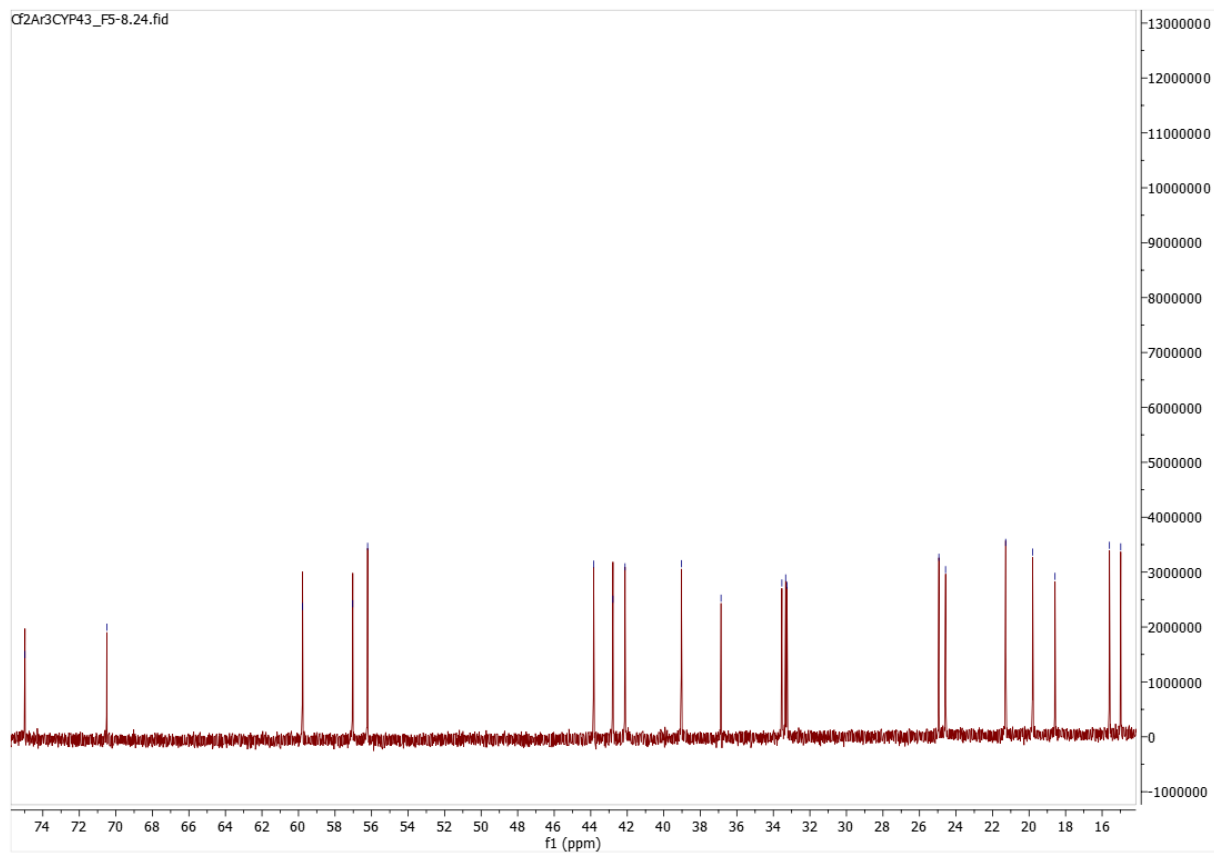


Figure S3.5f. ^{13}C NMR spectrum.

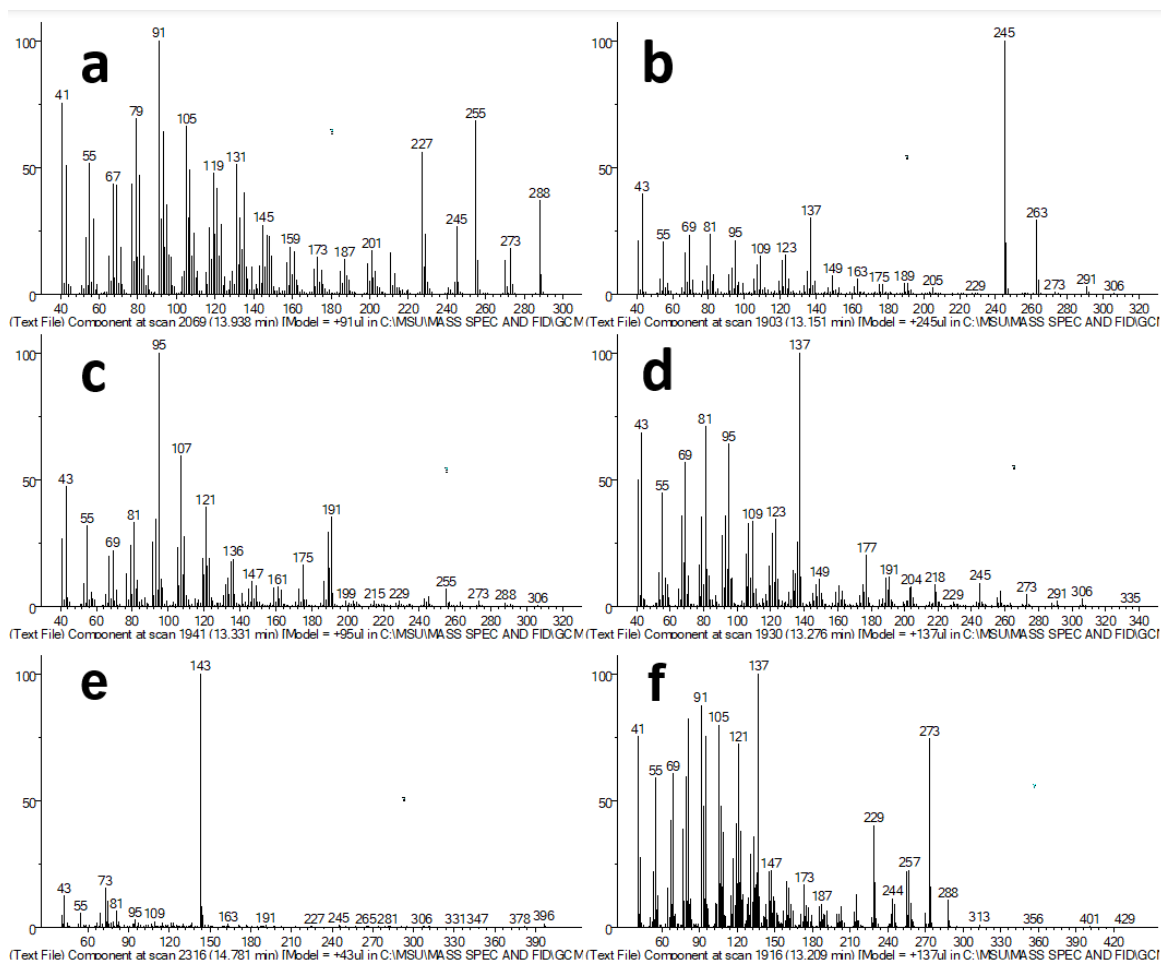


Figure S3.6. mass spectra of the ArCYP736A358 products. a. 3-hydroxy-*ent*-kaurene, b epoxy-manoyl oxide, c epoxy-(-)-kolavelool*, d epoxy-(+)-manool*, e silylated epoxy-sclareol*, f epoxy-sandaracopimarene*, or compounds 2, 3, 4*, 5*, 6*, 7* respectively. The derivatization of epoxy-sclareol had other minor products

Chapter 4: Exploiting diverse viridiflorol synthase promiscuity to produce novel semi-synthetic anti-fungal terpenes

(manuscript in prep)

Authors: Nick Schlecht^{1,2}, Trine Andersen^{1,2}, Matthew Giletto⁴, Navreet Singh¹, Daniel

Holmes⁴ Edmund Ellsworth⁴, Björn Hamberger^{1,2}

Affiliations

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

²DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, USA

³Department of Pharmacology and Toxicology, Michigan State University

⁴Department of Chemistry, Michigan State University, East Lansing MI, USA

Abstract:

Chiral compounds are effective for agricultural and public health applications but performing stereospecific modification through organic chemistry remains challenging. Terpenes are structurally complex natural products generated by terpene synthases (TPSs), which cyclize their achiral precursors, prenyl diphosphates. The mechanism of TPSs relies on the intrinsic reactivity of carbocations to cause rearrangements, which can result in promiscuous activities. Viridiflorol is an anti-fungal sesquiterpene. It is produced by a range of phylogenetically diverse TPSs. We leveraged a panel of synthetic prenyl diphosphates to screen the activity of viridiflorol and related TPSs. This approach revealed that sixteen TPS-substrate combinations produced a variety of sesquiterpene derivatives and truncated putative monoterpenes. Disc diffusion assays demonstrated that viridiflorol inhibited diverse soilborne phytopathogens, including *Rhizoctonia solani*, a species that affects several agriculturally relevant crops. Sesquiterpene derivatives tested against *R. solani* exhibited varying levels of inhibition. Most notably, the fluorinated sesquiterpene derivatives from one combination were twice as effective as viridiflorol, showing 41% less growth than the solvent control. These results suggest that partnering substrate analogs with promiscuous enzymes may be a promising strategy for producing new sesquiterpene derivatives with altered bioactivities.

Introduction

Identifying and improving anti-fungal compounds is important due to the substantial crop losses caused by soilborne pathogens. Roughly 90% of major plant diseases are soilborne, causing up to 4 billion dollars a year lost in the US alone.^{1,2}

Adapting to the rapidly changing agricultural needs, public health challenges, and the necessity for more ecologically friendly compounds warrants the development of new chemistries.³⁻⁵ Fortunately, natural products have evolved over millions of years to address a wide range of biotic and abiotic stimuli, making them a diverse and valuable resource. The development of commercial chemicals involves not only increasing natural product yields but also modifying these products to optimize their bioactivity. This trend is evident in commercial drugs and agrochemicals, where over 25 years, natural products account for only 5% of new chemicals while another 47% of new chemicals were modified natural products, pharmacophore or mimics.⁶

Although great advances have been made in enantioselective chemistry, total synthesis is often an insurmountable task, requiring other approaches.^{7,8} Semi-synthetic compounds are often derivatized from natural products to either optimize or transform the native bioactivities.⁹ These modifications can alter characteristics such as half-lives and binding affinities, enabling tailored properties.^{10,11} While derivatization can be beneficial, stereospecific modification remains challenging for many natural products.⁷ One recent alternative involves synthesizing or derivatizing a structurally similar achiral analog to the native substrate. When paired with promiscuous enzymes, this approach can utilize native stereospecific mechanisms to synthesize unnatural semi-synthetic products.¹²⁻¹⁵ Terpenoids are well-suited for this approach, as they are heavily chiral and their precursors, prenyl diphosphates, are easily modifiable due to their linear, achiral structure made of repeated units.

The varying length prenyl diphosphates are created by sequential condensation of five carbon isoprene units yielding geranyl diphosphate (C10), farnesyl diphosphate

(FDP) (C15), geranylgeranyl diphosphate (C20) and longer chains. Sesquiterpenes are characterized by having three isoprene units (15-carbons) and are derived from farnesyl diphosphate via the mevalonate pathway.¹⁶ Sesquiterpenoids comprise over 24,000 structures in the Dictionary of Natural Products and are found in all domains of life.¹⁷ This diversification stems from complex chiral products generated by class I terpene synthases (TPSs).¹⁸ Class I TPSs initiate cyclization by cleaving off the diphosphate from the substrate, followed by a series of carbocation-mediated cycloisomerizations.¹⁹ TPS reactions largely rely on the intrinsic reactivity of carbocations, mainly stabilizing specific carbocations and using structural features to favor certain conformations.^{19,20} This property enabled some TPSs to evolve different forms of promiscuity, including substrate promiscuity in *Prunella vulgaris* (PvTPS5) and *Antirrhinum majus* (AmNES/LIS-1).^{21,22} promiscuous sesquiterpene synthases can likely be exploited to generate terpene derivatives by hijacking of native machinery with alternative substrates.

Convergent evolution has repeatedly driven sesquiterpene synthases from diverse evolutionary lineages to catalyze similar reactions. For example, viridiflorol and related tricyclic sesquiterpenes have corresponding TPSs that have emerged in bacteria, fungi, and plants. Some examples include *Serendipita indica* (SiVS), *Agrocybe aegerita* (Agr5), *Streptomyces avermitilis* (SaAMS) and, *Clitopius pseudopinsitus* (CpSTS9) (Figure 4.1).^{22–27} While these TPSs can generate viridiflorol, some also produce additional products in different ratios, suggesting different catalytic environments. Viridiflorol exhibits several advantageous bioactivities relevant to the cosmetic, pharmaceutical, and agrochemical industries—justifying an exploration into

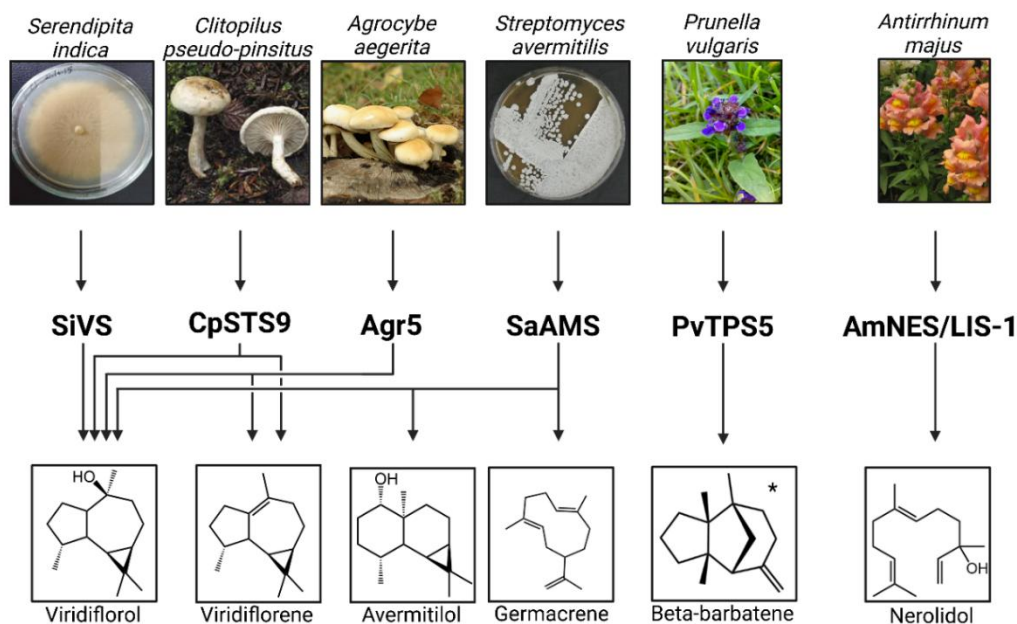


Figure 4.1. Sesquiterpene synthases, their sources and products. TPSs used in this study are shown, highlighting their major sesquiterpene products as well as biological sources. The pictures generally illustrate the species, save *Clitopilus pseudo-pinsitus*, which is an image of *Clitopilus umbilicatus*, a close relative.^{32–36} * tentatively identified by GC-MS library hits.²²

We leveraged the viridiflorol synthase diversity represented by SiVS, CpSTS9, Agr5, SaAMS, PvTPS5, and AmNES/LIS-1 by pairing them with semi-synthetic Farnesyl diphosphate (FDP) analogs to produce unnatural sesquiterpenes. In addition to discovering novel products, this study enabled activity comparisons amongst viridiflorol synthases and the substrates. To contextualize the applications and value of viridiflorol and the semi-synthetic sesquiterpene products, we compared their anti-fungal activities against known phytopathogens. Our results expanded the known activity of viridiflorol and found that the semi-synthetic terpene products exhibited varied activity.

Materials and Methods

Sourcing and Cloning Genes

The genes encoding viridiflorol synthases SiVS (*Serendipita indica*), Agr5 (*Agrocybe aegerita*), CpSTS9 (*Clitopilus pseudopinsitus*), and SaAMS (*Streptomyces avermitilis*) as well as promiscuous TPSs PvTPS5 and AmNES/LIS-1 were codon-optimized for *E. coli*, had their stop codons temporarily removed, and attached homology arms for the pET28-b(+) vector digested with NcoI and XhoI (Table S4.1).^{21–25,27} In-Fusion cloning (Takara Bio) was used to recombine the genes into the bacterial expression vector pET28-b(+), introducing a C-terminal 6x-HIS tag and a new stop codon. The pET28-b(+) TPS vectors were transformed into chemically competent OverExpress C41TM cells (Sigma).

Enzyme expression and purification

Transformed *E. coli* strains were inoculated in Terrific Broth (TB) with 50 µg/mL kanamycin and grown overnight. The starter culture was diluted in fresh TB + kanamycin to 0.02 OD₆₀₀ and incubated at 37 °C and 200 rotations per minute (RPM). At 0.6 OD₆₀₀, 10 µl of 1 M isopropyl 1-thio-β-d-galactopyranoside was added, then the temperature was lowered to 17 °C and rotations were decreased to 140 RPM. Expression continued overnight and cells were harvested by centrifugation at 5,000 x *g* at 4 °C for 15 min. Cell pellets were resuspended in lysis buffer (20 mM HEPES, pH 7.5, 0.5 M NaCl, 25 mM imidazole, 5% v/v glycerol, an EDTA-free protease inhibitor mixture tablet (Sigma-Aldrich), 50 U/µl benzonase, and 0.1 mg/mL lysozyme) and lysed by sonication. The lysate was centrifuged for 25 min at 14,000 x *g*, and the supernatant was loaded onto equilibrated 1-mL His SpinTrap columns (GE Healthcare). Proteins

were washed twice with buffer A (20 mM HEPES, pH 7.5, 0.5 M NaCl, 25 mM imidazole, 5% v/v glycerol) and eluted with buffer B (20 mM HEPES, pH 7.5, 0.5 M NaCl, 350 mM imidazole, 5% v/v glycerol). The eluted product was desalted on a PD Minitrap G-25 columns (GE Healthcare). Purified proteins were quantified using a Nanodrop (A280), followed by either *in vitro* assays or storage at -80 °C.

Screening In vitro assays

For each screening assay, 50 µg of purified enzyme, 1 Unit of CIP phosphatase (Promega), or a no-protein control was added to a 2 mL amber vial, with 750 µl reaction buffer (50 mM HEPES, 7.5 mM MgCl₂, 5% glycerol, pH 7.2). Then 10 µl of synthetic FDP analog (2 mg/mL dissolved in 70% methanol : 30% water) or FDP was added. Modified GGPP substrates were synthesized by Matthew Giletto and Edmund Ellsworth (Michigan State University Medicinal Chemistry Facility) and obtained at approximately 50% purity by weight (phosphate impurities) and were resuspended in 70% methanol 30% water to 2 mg/ml (or saturation). Figure S4.1 shows the general scheme for their synthesis. 500 µl hexane was overlaid on the aqueous phase, then the vial was capped and shaken at 95 RPM and 30 °C overnight. The reaction mixture was then vortexed and centrifuged before transferring the hexane layer to a new vial for analysis by a Gas Chromatograph-Flame Ionization Detection (GC-FID). If unique peaks compared to the phosphatase controls were identified, the sample was reanalyzed by a Gas Chromatograph-Mass Spectrometry (GC-MS). The deconvoluted mass spectrum generated using the software AMDIS was then used to search against a library of known compounds.³⁷ Peaks that had mass spec library hits suggesting they were terpenes and had a parent ion matching the molecular weight of a dephosphorylated

FDP analog quenched by water or deprotonated were considered terpene derivatives.

Purification and identification of semisynthetic products

The *in vitro* assay protocol described above was slightly modified for product scale-up. Several 10x scaled-up reactions were prepared for each successful combination, with 275 μ L of FDP analogs added instead of 100 μ L. Reactions were incubated for 2-3 days, sampled by GC-FID, and additional enzyme was added for low-yielding combinations and incubated 2 additional days. Products were concentrated by gentle air flow to 1.5 mL. Five hundred mg prepacked silica columns (Waters) were equilibrated with hexane, then products were loaded onto the column. The typical elution sequence typically followed: 10 mL hexane, 5 mL 1% ethyl acetate, 5 mL 2.5% ethyl acetate, 1 mL 5% ethyl acetate, 1 mL 10% ethyl acetate, 3 mL 100% ethyl acetate, collecting 1 mL fractions. Fractions were screened with GC-FID and verified with GC-MS. Additional columns were often necessary, where a silica slurry was manually loaded into a Pasteur pipette to create a longer column for better separation. Product yields were determined by a comparison to a standard curve of 1-eicosene generated by a GC-FID. Purified samples were dried and resuspended in CDCl_3 for Nuclear Magnetic Resonance (NMR) analysis. NMR spectra were recorded in Michigan State University's Max T. Rogers NMR Facility on a Bruker Avance NEO 800 or 600 MHz spectrometer equipped with a helium cooled TCI cryo-probe or a Prodigy TCI cryoprobe, respectively. CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively.

GC-FID, GC-MS, and LC-MS analyses

Initial screenings were conducted using an Agilent 7890 flame ionization detector

with an Agilent HP-5ms (30 m x 0.25 mm x 0.25 μ m) column. Product validation was performed on an Agilent 7890A + 5975C GC-MS with an Agilent VF-5ms column (30 m x 250 μ m x 0.25 μ m, with 10m EZ-Guard). Both instruments used the following method: temperature ramp start at 40°C, hold 1 min, 40 °C/min to 200 °C, hold 4.5 min, 20 °C/min to 240 °C, 10 °C/min to 280 °C, 40 °C/min to 320 °C, and hold 5 min. For the 5975C MS the scan range was set from 40-450 *m/z*. Spectra deconvolution was done utilizing AMDIS.³⁷ The dephosphorylated products of substrates **1** and **2** could not be detected by GC, so liquid chromatography atmospheric pressure chemical ionization mass spectrometry was necessary. Liquid chromatography mass spectrometry (LC-MS) was necessary for combinations with substrates 1 and 2.

LC-MS analyses were performed on a Waters Xevo G2-XS quadrupole time-of-flight mass spectrometer with a Waters ACQUITY column manager and Waters ACQUITY Premier BEH C18 column (2.1 x 50 mm 1.7 μ m). Injection volume for each sample was 1 μ L, and flow rate was set to 0.3 mL/min with a column temperature of 40 °C. The mobile phase consisted of 10 mM ammonium formate (pH 2.8) (Solvent A) and acetonitrile (Solvent B) with the following method: initial 99% A : 1 % B , continuous gradient to 2% A : 98% B over 4 min, hold for 1 min, continuous gradient to 99% A : 1% B over 0.1 min, hold 1.5 min. Mass spectra were generated through atmospheric pressure chemical ionization (APCI) in positive-ion mode with leucine enkephalin as a lockmass, and continuum peak acquisition were collected with a mass range of *m/z* 50-1500 and a scan duration of 0.2 s. Capillary and cone voltage were 20 V and 80 V, respectively. Cone and desolvation gas flow rates were 25 and 600 L/h, respectively. Source and desolvation temperatures were 100°C and 350°C, respectively. High-energy

spectra were generated with argon as the collision gas and a voltage ramp from 20 to 80 V.

Fungal Bioassays

Agriculturally prevalent soilborne phytopathogens: *Macrophomina phaseolina*, *Rhizoctonia solani*, *Colletotrichum acutatum*, and *Pythium irregulare* species were received from Professors Martin Chilvers and Timothy Miles, Department of Plant, Soil, and Microbial Sciences at Michigan State University. Initial assays for each pathogen included two fungal plugs as biological replicates on a 5 cm radius Petri dish containing 0.1 x PDA (potato dextrose broth mix and agar). 80 µg of viridiflorol or corresponding volume of the solvent hexane (8 µl) was loaded onto a 0.6 cm radius round filter paper discs (Cytiva Whatman grade 4 qualitative filter papers). A 0.6 cm fungal plug was placed on the plate and the loaded discs were placed approximately 1.25 cm away from the plug on both sides. Plates were photographed daily with an EPSON Perfection V700 Photo until fungal cultures in control plates reached the edge of the plate or one week had passed. Plates were photographed such that the fungal plugs horizontally aligned with the discs surrounding them. All images were analyzed in the program ImageJ, measuring the radial mycelium growth from the fungal plug towards the loaded discs.³⁸

This experiment was repeated with decreasing quantities of viridiflorol (80 µg, 20 µg, 8 µg, 4 µg, and 0.8 µg) to generate a dose-response curve (Figure 4.3). Growth inhibition was calculated as one minus the average growth rate at a given viridiflorol concentration divided by the average growth rate of the control (equation 4.1). The final anti-fungal experiment only used *R. solani*, had three biological replicates per condition when possible, and all samples had 20 µg of viridiflorol or a semi-synthetic product

loaded except **SiVS + 10 A** and **B** which only had 17.92 µg and 9.8 µg respectively. An analysis of variance (ANOVA) comparing each sample and the average growth of each sample was performed after excluding the SiVS samples due to being single replicates of different concentrations. A Tukey's t-test was used to assess pairwise significance between samples.

$$\text{Growth inhibition} = 1 - \frac{G_{[x]}}{G_{\text{control}}}$$

Equation 4.1: Fungal Growth Inhibition Equation: This equation was used to determine the effect of a given concentration of viridiflorol on each soilborne pathogen. $G_{[x]}$ is the average radial growth of a given viridiflorol concentration whereas G_{control} was the average radial growth for the hexane control.

Results

Screening TPS activity with synthetic substrate analogs

The synthetic FDP analog variation enabled us to explore how altered polarity and structural diversity affect the catalytic landscape of the TPSs and the anti-fungal activity of their products. To systematically explore synthetic substrate cyclization and TPS promiscuity towards unnatural substrates, each TPS—FDP analog combination was assayed and screened by GC-FID. Controls without protein, with phosphatases (instead of TPSs), or with FDP were used to identify non-TPS products and confirm TPS activity. The panel of six TPSs and 12 synthesized FDP analogs, along with their respective controls, resulted in 108 unique enzyme—substrate combinations assayed. Phosphatase controls yielded multiple products for substrates 3, 4, 5, 9, 10, and 11 implying that minor rearrangements can occur with phosphatases for half of the substrates (Figure S4.2). Although many TPS-FDP analog combinations did not create novel products, TPSs frequently generated smaller but identical peaks to phosphatase

controls (Figure S4.3-S4.18). This observation suggests that TPS diphosphate cleavage may not be fully inhibited, even if specific cycloisomerizations are fully inhibited. TPS-FDP analog combinations that generated distinct peaks were validated by GC-MS, resulting in 16 combinations that yielded terpenes (Figure 4.2b).

Overall, results showed a spectrum of enzyme promiscuity and substrate permissibility. SaAMS had the broadest substrate range. However, SaAMS and Agr5 frequently caused internal cleavages of the substrates, producing monoterpenes instead of derivatized sesquiterpenes. CpSTS9 had a narrower substrate range but typically yielded more individual sesquiterpene derivatives in a given combination and were more abundant than identical products in other combinations (Figure S4.19 & S4.20). Substrates **10** and **7** exhibited the broadest utility, while substrates **1**, **2**, **6**, **8**, and **12** generated no terpene products with any TPS tested. The fluorinated substrate **11** was modified by PvTPS5 and CpSTS9. The products generated from substrates **4**, **5**, and **9** had retention times and mass spectra characteristic of monoterpenes and monoterpene-like products, suggesting that internal cleavages occurred during carbocation rearrangements. While 16 combinations generated terpene derivatives, many of these combinations had multiple products. Most notably, substrate **10** had eight separate products and substrate **7** had four terpene derivatives between their different combinations. (Figure S4.19-S4.20).

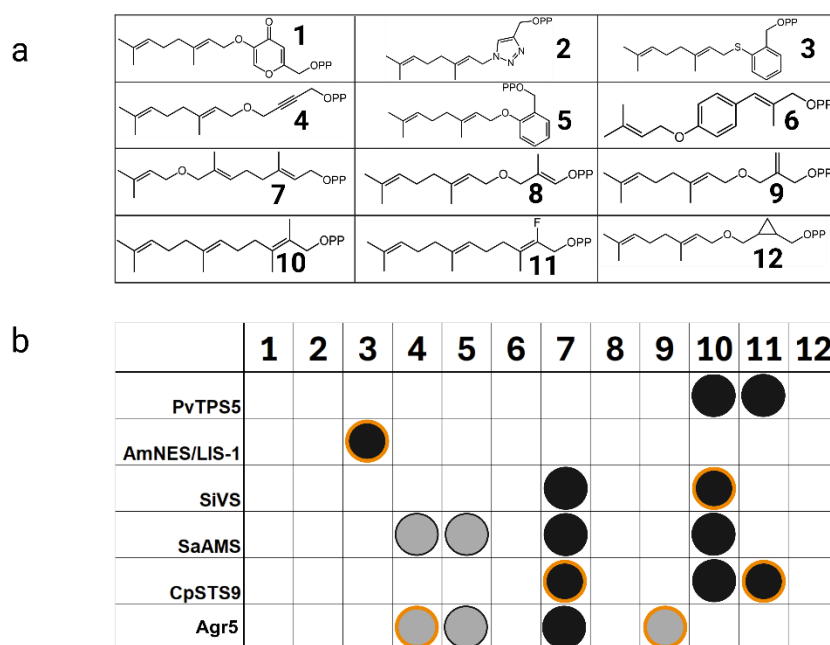


Figure 4.2. Functional enzyme: FDP analog-TPS combinations. **a.** Synthetic farnesyl diphosphate analog structures tested as substrates in this study **b.** Summary of TPS-FDP analog combination results. columns show the numbered structure as pictured in a. The rows show the TPS along with the phylogenetic relationships between the TPSs. Black discs indicate an assay that produced a putative sesquiterpene analog based on GC-MS results. Gray discs are for assays that yielded monoterpene-like products from an internal cleavage based on mass spec library hits and retention times. Orange outlines indicate products scaled up for anti-fungal assays. The chromatograms and mass spectra for the respective products can be found in Figure S4.3-S4.18.

Six combinations (highlighted in orange in Figure 4.2) were scaled up for product purification and anti-fungal assays. Two of those products which we gleaned structural information on via NMR spectroscopy. **Agr5 + 9** produced both a truncated monoterpene-like product and a linear compound identical to one of the phosphatase products. The linear compound generated by both phosphatase and Agr5 was identified by NMR analysis to be (*E*)-4-(((*E*)-3,7-dimethylocta-2,6-dien-1-yl)oxy)but-2-en-1-ol (Figure S4.21). Despite also being a phosphatase product, the unnatural terminal methylene appears to shift upon dephosphorylation (Figure S4.21). The crude NMR data for the **CpSTS9 + 7** product had a mixture of two major compounds, with the

predominant component having NMR data consistent with (5*E*,9*E*)-6,10-dimethyl-3-(prop-1-en-2-yl)oxacycloundeca-5,9-diene, a compound with a single heterocycle reported in a 2018 study where synthetic substrates were also provided to TPSs (Figure S4.22).¹⁴

Evaluating the anti-fungal capacity of viridiflorol

Exploration of the anti-fungal capabilities of viridiflorol has been limited so far.^{23,31,39} To better measure the extent of this activity we tested viridiflorol against a collection of four phytopathogens from diverse lineages. The collection included: *Colletotrichum acutatum*, *Macrophomina phaseolina*, *Rhizoctonia solani*, and *Pythium irregulare*. The first three species belong to the fungal classes Dothideomycetes, Agaricomycetes, and Sordariomycetes while *P. irregulare* is a fungal-like oomycete. This collection therefore represents phylogenetically diverse lineages of soilborne phytopathogens. The initial disc diffusion assay used a concentrated dose (80 µg) of viridiflorol. Significant differences in growth rate were observed for each species, demonstrating that viridiflorol has at least weak/moderate anti-fungal activity across phylogenetically distinct soilborne pathogens (Figure 4.3).

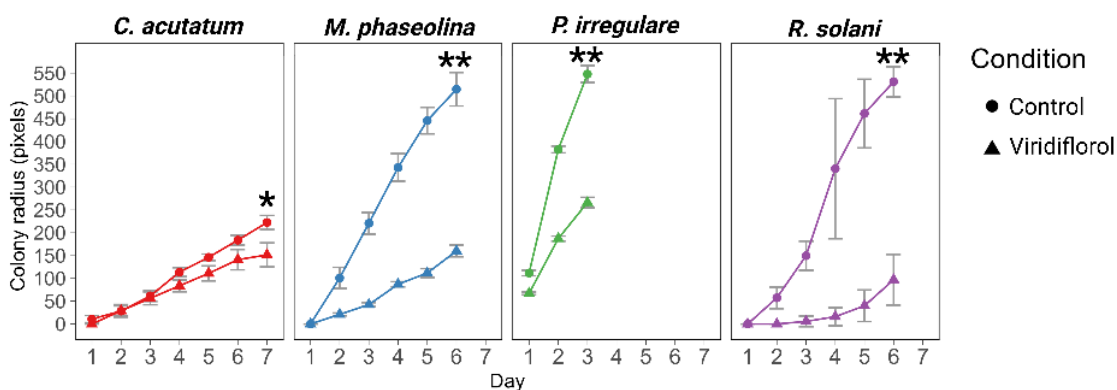


Figure 4.3. Growth rates of phytopathogenic fungi with and without 80 ug viridiflorol discs. Horizontal growth rates from the plug towards the loaded discs were recorded with ImageJ daily for both viridiflorol and the solvent control. Collection of time-

Figure 4.3 (cont'd)

points stopped once controls reached the edge of the disc or after 7 days. The final time point had a Welch's T test to compare growth with and without viridiflorol where * indicates $p \leq 0.01$ and ** indicates $p \leq 0.0001$.

Assays with varying quantities of viridiflorol (80 μg , 20 μg , 8 μg , 4 μg , 0.8 μg) were used to assess dose-response against different phytopathogens (Figure 4.4). *P. irregularis* was statistically the most sensitive, significantly inhibited at 0.8 μg viridiflorol ($p = 3.43\text{E-}4$), though this only represented an 11.9% reduction in growth. *M. phaseolina* followed, inhibited at as low as 8 μg ($p = 1.05\text{E-}4$), representing a 24.9% inhibition. *R. solani* showed moderate sensitivity, showing significant inhibition of 29.1% at 20 μg viridiflorol ($p = 1.44\text{E-}3$). *C. acutatum* was the least sensitive, statistically significant only at 80 μg viridiflorol where it displayed a 31.2% inhibition ($p = 1.3\text{E-}3$). The uniform growth of *P. irregularis* contributed to higher statistical power potentially inflating its apparent sensitivity. Nonetheless, these results indicate that viridiflorol inhibits growth of a broad spectrum of phytopathogens.

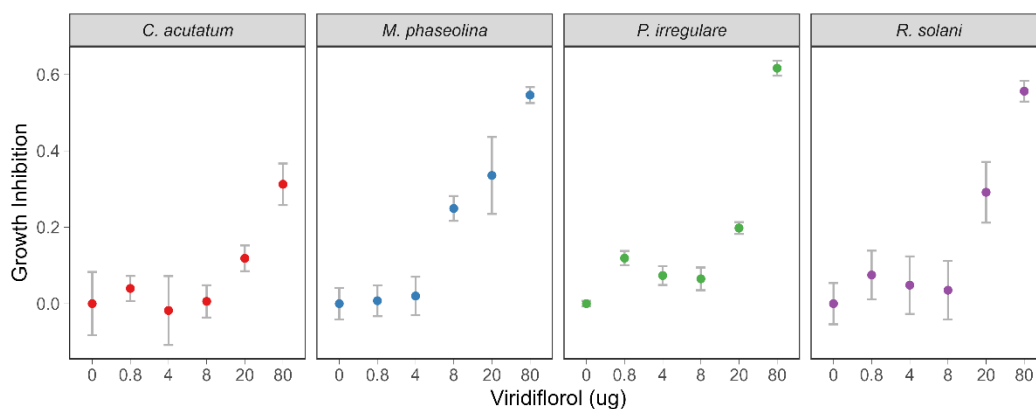


Figure 4.4. Dose-response effect of viridiflorol on different phytopathogens. Disc diffusion assays measured across varying viridiflorol concentrations were used to approximate percent inhibition. Analysis used the last day of growth prior to controls reaching the edge of plates. *P. irregularis* used measurements after the second day, *R. solani* used measurements after the third day, and both *C. acutatum* and *M. phaseolina* used measurements after seven days growth.

Evaluating the anti-fungal properties of semi-synthetic sesquiterpenes

The semi-synthetic terpene products were assayed against *R. solani* due to its sensitivity to viridiflorol and its agricultural relevance (Figure 4.5).^{40–44} The low purification yields of **SiVS + 10** products prevented replicates but they were still included to serve as an exploratory qualitative comparison. 20 µg of viridiflorol or synthetic substrates was loaded onto discs for comparison against *R. solani*. After 32 hours—at which point the fungal growth reached the discs—statistical analyses were applied to evaluate activity. An ANOVA test confirmed variance in growth rates between samples ($p = 3.21\text{E-}13$) leading to pairwise statistical analyses (Tukey's t test) of the samples (Table S4.2).

All synthetic substrates significantly inhibited *R. solani*'s growth compared to the solvent control, with average inhibition ranging from 11.5% to 40% and viridiflorol displaying 21% inhibition. The **CpSTS9 + 7** products exhibited the lowest inhibition (11.5%) and was the only sample that was significantly less effective than viridiflorol. **Agr5 + 4**, **AmNES/LIS1 + 3 fractions A** and **B** varied from 14 to 28% growth inhibition, but none were statistically different from viridiflorol. Only the products of **Agr5 + 4** and **CpSTS9 + 11** significantly outperformed viridiflorol, inhibiting growth by 32% and 40.2% respectively. Two other samples, **SiVS + 10 fractions A** and **B**, reported a 30.2% and 18.9% decrease in growth rates, respectively, but are only provided as qualitative references. In addition to being single replicates, only 17.9 µg from fraction **A** and 9.8 µg from fraction **B** were loaded in the discs compared to the standard 20 µg due to the low purification yields. Given these limited testing conditions, the inhibitory effects of fractions A and B is promising but warrants further investigation.

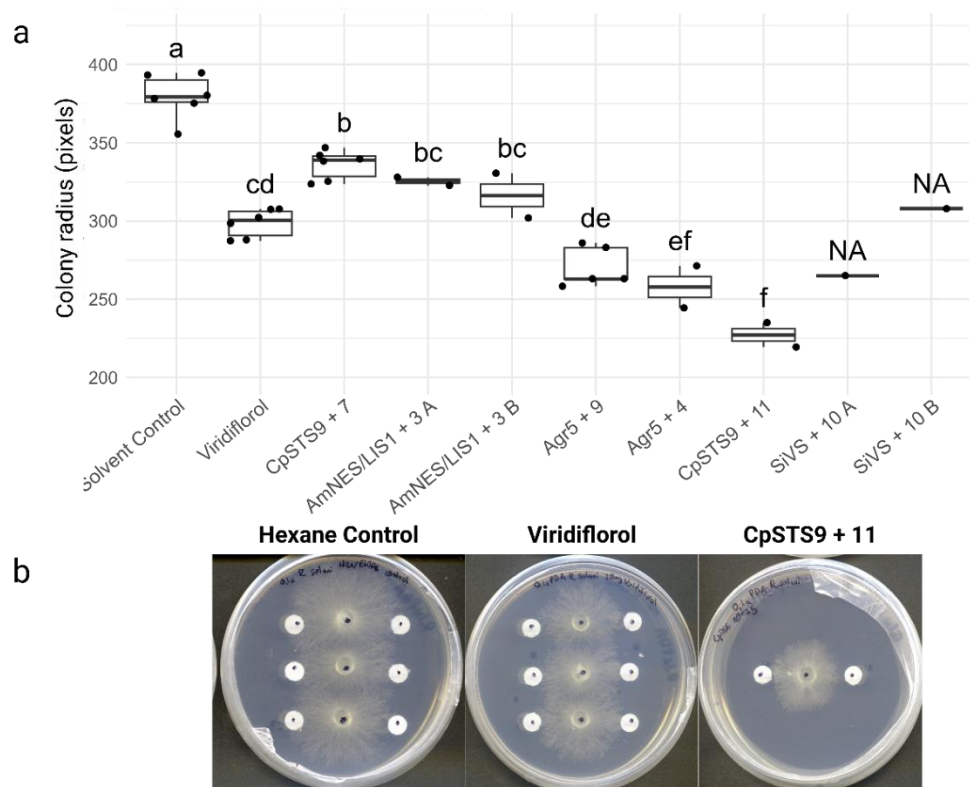


Figure 4.5. *R. solani* growth inhibition by semi-synthetic terpene products after 32 hours. a Boxplot of *R. solani* growth on plates with distinct compounds loaded on each disc. Each disc had 20 μ g loaded save **SiVS + 10 fractions A and B**, which had 17.92 **Figure 4.5. (cont'd)**

μ g and 9.8 μ g, respectively. Dots represent individual horizontal measurements towards a disc. Letters represent the statistical groupings resulted from a Tukey's T test. **b** Images of three of the plates for the solvent control, viridiflorol, and **CpSTS9 + 11**, illustrating the increasing inhibition from **CpSTS9 + 11**.

Discussion

Climate change poses several challenges to agriculture, including an anticipated increase in phytopathogenic fungi.⁴⁵ Expanding the available chemical space via our semi-synthetic approach may aid in combating the growing agricultural challenges. The synthetic substrate screenings revealed sixteen TPS-substrate combinations that produced terpenes. Five of the twelve prenyl diphosphate derivatives were not used as substrates with the tested TPSs. The remaining seven showed unique product patterns

depending on both enzymes and substrates. Consistent with their native function, many TPSs yielded multiple products. For instance, the PvTPS5 + 10 combination generated at least eleven different sesquiterpene derivatives alone. At least eight of these products are shared across other TPS combinations yet at different ratios (Figure S4.20).

Phosphatase activity resulted in dephosphorylation and some rearrangements. The NMR structure elucidated for the product of **Agr5 + 4** also corresponded to a phosphatase product (Figure S4.21). The linear product shows a methylene shift upon diphosphate cleavage. The other identified structure was one of the two products from **CpSTS9 + 7**. The crude NMR data show a mixture of compounds with the major component having NMR data consistent with (5E,9E)-6,10-dimethyl-3-(prop-1-en-2-yl)oxacycloundeca-5,9-diene, a structure resolved in previous work (Figure S4.22).¹⁴ Other structures could not be resolved, however their mass spectra and corresponding library hits suggest various cyclizations. Tentative identification is given in figures S4.3-S4.18.

We demonstrated that viridiflorol has broad anti-fungal activity, expanding previous work.^{23,39,46,47} The phytopathogen collection used here not only included phylogenetically diverse classes of fungi and an oomycete, but it also represents a wide array of diseases. This includes various root rots, blights, and damping off of seedlings found in various fruits, vegetables, tubers, wheat, barley, maize, sorghum and other grains.^{40,48–51} Viridiflorol was found to weakly inhibit *C. acutatum* and had stronger activity against *M. phaseolina*, *R. solani*, and *P. irregulare*.

R. solani was selected for further assays due to its viridiflorol sensitivity and its ability to infect a broad range of food and bioenergy crops.^{40–44} All sesquiterpene

derivatives produced in this study significantly inhibited *R. solani*'s growth to varying extents. This result aligns with published work that demonstrated that volatile compounds can inhibit *R. solani* growth.⁵² **CpSTS9 + 11** was the best performing, nearly twice as effective at inhibition than viridiflorol. This combination is a mixture containing putative cyclized and fluorinated sesquiterpenes as well as peaks identical to several phosphatase products. The fluorinated sesquiterpenes may have been highly effective because fluorination tends to affect binding affinities, metabolic stability, and membrane permeability.⁵³

Conclusion and Future directions

This study identified over a dozen enzyme—substrate analog combinations that could produce sesquiterpene derivatives. NMR data supports the structures of two products. We characterized the anti-fungal activity of viridiflorol and derivatives that were scalable. These assays revealed that viridiflorol has broader applications as an anti-fungal agent than previously shown. Furthermore, additional anti-fungal assays demonstrated that the sesquiterpene derivatives varied in activity, including the CpSTS9 + 11 combination which was twice as effective at inhibiting *R. solani* growth than viridiflorol.

Producing more of the active FDP analogs could enable better optimization for the scale-up and purification, as there were considerable yield losses during these processes. Optimizing or investigating alternative extraction methods, such as hydrophobic adsorbers, and altering the substrate solution to limit protein denaturation can improve the initial extraction yields. Additionally, utilizing automated flash chromatography or fractional distillation for better purification and retention of volatile

synthetic products could improve purification efficiency.¹⁴ Beyond this initial optimization, the various products could be properly purified, structurally elucidated, and evaluated for their anti-fungal activity. Overall, this work serves as a foundation for future development of semi-synthetic anti-fungal compounds.

REFERENCES

- (1) Pimentel, D.; McLaughlin, L.; Zepp, A.; Lakitan, B.; Kraus, T.; Kleinman, P.; Vancini, F.; Roach, W. J.; Graap, E.; Keeton, W. S.; Selig, G. Environmental and Economic Effects of Reducing Pesticide Use: A Substantial Reduction in Pesticides Might Increase Food Costs Only Slightly. *BioScience* **1991**, *41* (6), 402–409. <https://doi.org/10.2307/1311747>.
- (2) Mazzola, M. Management of Resident Soil Microbial Community Structure and Function to Suppress Soilborne Disease Development. In *Climate change and crop production*; CAB International: Wallingford, UK, 2010; pp 200–218. <https://doi.org/10.1079/9781845936334.0200>.
- (3) Fisher, M. C.; Henk, D. A.; Briggs, C. J.; Brownstein, J. S.; Madoff, L. C.; McCraw, S. L.; Gurr, S. J. Emerging Fungal Threats to Animal, Plant and Ecosystem Health. *Nature* **2012**, *484* (7393), 186–194. <https://doi.org/10.1038/nature10947>.
- (4) Bebber, D. P.; Ramotowski, M. A. T.; Gurr, S. J. Crop Pests and Pathogens Move Polewards in a Warming World. *Nature Clim Change* **2013**, *3* (11), 985–988. <https://doi.org/10.1038/nclimate1990>.
- (5) Bajželj, B.; Richards, K. S.; Allwood, J. M.; Smith, P.; Dennis, J. S.; Curmi, E.; Gilligan, C. A. Importance of Food-Demand Management for Climate Mitigation. *Nature Clim Change* **2014**, *4* (10), 924–929. <https://doi.org/10.1038/nclimate2353>.
- (6) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Last 25 Years. *J. Nat. Prod.* **2007**, *70* (3), 461–477. <https://doi.org/10.1021/np068054v>.
- (7) Gellman, A. J. Chiral Surfaces: Accomplishments and Challenges. *ACS Nano* **2010**, *4* (1), 5–10. <https://doi.org/10.1021/nn901885n>.
- (8) Quasdorf, K. W.; Overman, L. E. Catalytic Enantioselective Synthesis of Quaternary Carbon Stereocentres. *Nature* **2014**, *516* (7530), 181–191. <https://doi.org/10.1038/nature14007>.
- (9) Jansen, D. J.; Shenvi, R. A. Synthesis of Medicinally Relevant Terpenes: Reducing the Cost and Time of Drug Discovery. *Future Medicinal Chemistry* **2014**, *6* (10), 1127–1146. <https://doi.org/10.4155/fmc.14.71>.
- (10) Yu, Z.; Chen, Z.; Li, Q.; Yang, K.; Huang, Z.; Wang, W.; Zhao, S.; Hu, H. What Dominates the Changeable Pharmacokinetics of Natural Sesquiterpene Lactones and Diterpene Lactones: A Review Focusing on Absorption and Metabolism. *Drug Metabolism Reviews* **2021**, *53* (1), 122–140. <https://doi.org/10.1080/03602532.2020.1853151>.

- (11) Sati, P.; Sharma, E.; Dhyani, P.; Attri, D. C.; Rana, R.; Kiyekbayeva, L.; Büsselberg, D.; Samuel, S. M.; Sharifi-Rad, J. Paclitaxel and Its Semi-Synthetic Derivatives: Comprehensive Insights into Chemical Structure, Mechanisms of Action, and Anticancer Properties. *European Journal of Medical Research* **2024**, 29 (1), 90. <https://doi.org/10.1186/s40001-024-01657-2>.
- (12) Hammer, S. C.; Syrén, P.-O.; Seitz, M.; Nestl, B. M.; Hauer, B. Squalene Hopene Cyclases: Highly Promiscuous and Evolvable Catalysts for Stereoselective CC and CX Bond Formation. *Current Opinion in Chemical Biology* **2013**, 17 (2), 293–300. <https://doi.org/10.1016/j.cbpa.2013.01.016>.
- (13) Seitz, M.; Syrén, P.-O.; Steiner, L.; Klebensberger, J.; Nestl, B. M.; Hauer, B. Synthesis of Heterocyclic Terpenoids by Promiscuous Squalene–Hopene Cyclases. *ChemBioChem* **2013**, 14 (4), 436–439. <https://doi.org/10.1002/cbic.201300018>.
- (14) Oberhauser, C.; Harms, V.; Seidel, K.; Schröder, B.; Ekramzadeh, K.; Beutel, S.; Winkler, S.; Lauterbach, L.; Dickschat, J. S.; Kirschning, A. Exploiting the Synthetic Potential of Sesquiterpene Cyclases for Generating Unnatural Terpenoids. *Angewandte Chemie International Edition* **2018**, 57 (36), 11802–11806. <https://doi.org/10.1002/anie.201805526>.
- (15) Hou, A.; Lauterbach, L.; Dickschat, J. S. Enzymatic Synthesis of Methylated Terpene Analogues Using the Plasticity of Bacterial Terpene Synthases. *Chemistry – A European Journal* **2020**, 26 (10), 2178–2182. <https://doi.org/10.1002/chem.201905827>.
- (16) Vranová, E.; Coman, D.; Gruissem, W. Network Analysis of the MVA and MEP Pathways for Isoprenoid Synthesis. *Annu. Rev. Plant Biol.* **2013**, 64 (1), 665–700. <https://doi.org/10.1146/annurev-arplant-050312-120116>.
- (17) Taylor & Francis Group. *Dictionary of Natural Products* (v33.1). <https://dnp.chemnetbase.com/chemical/ChemicalSearch.xhtml?dswid=6473> (accessed 2024-11-22).
- (18) Degenhardt, J.; Köllner, T. G.; Gershenzon, J. Monoterpene and Sesquiterpene Synthases and the Origin of Terpene Skeletal Diversity in Plants. *Phytochemistry* **2009**, 70 (15), 1621–1637. <https://doi.org/10.1016/j.phytochem.2009.07.030>.
- (19) Miller, D. J.; Allemann, R. K. Sesquiterpene Synthases: Passive Catalysts or Active Players? *Nat. Prod. Rep.* **2012**, 29 (1), 60–71. <https://doi.org/10.1039/C1NP00060H>.
- (20) J. Tantillo, D. Biosynthesis via Carbocations: Theoretical Studies on Terpene Formation. *Natural Product Reports* **2011**, 28 (6), 1035–1053. <https://doi.org/10.1039/C1NP00006C>.

- (21) Nagegowda, D. A.; Gutensohn, M.; Wilkerson, C. G.; Dudareva, N. Two Nearly Identical Terpene Synthases Catalyze the Formation of Nerolidol and Linalool in Snapdragon Flowers. *The Plant Journal* **2008**, *55* (2), 224–239. <https://doi.org/10.1111/j.1365-313X.2008.03496.x>.
- (22) Johnson, S. R.; Bhat, W. W.; Sadre, R.; Miller, G. P.; Garcia, A. S.; Hamberger, B. Promiscuous Terpene Synthases from *Prunella Vulgaris* Highlight the Importance of Substrate and Compartment Switching in Terpene Synthase Evolution. *New Phytologist* **2019**, *223* (1), 323–335. <https://doi.org/10.1111/nph.15778>.
- (23) Ntana, F.; Bhat, W. W.; Johnson, S. R.; Jørgensen, H. J. L.; Collinge, D. B.; Jensen, B.; Hamberger, B. A Sesquiterpene Synthase from the Endophytic Fungus *Serendipita Indica* Catalyzes Formation of Viridiflorol. *Biomolecules* **2021**, *11* (6), 898. <https://doi.org/10.3390/biom11060898>.
- (24) Zhang, C.; Chen, X.; Orban, A.; Shukal, S.; Birk, F.; Too, H.-P.; Rühl, M. *Agrocybe Aegerita* Serves As a Gateway for Identifying Sesquiterpene Biosynthetic Enzymes in Higher Fungi. *ACS Chem. Biol.* **2020**, *15* (5), 1268–1277. <https://doi.org/10.1021/acscchembio.0c00155>.
- (25) Chou, W. K. W.; Fanizza, I.; Uchiyama, T.; Komatsu, M.; Ikeda, H.; Cane, D. E. Genome Mining in *Streptomyces Avermitilis*: Cloning and Characterization of SAV_76, the Synthase for a New Sesquiterpene, Avermitilol. *J. Am. Chem. Soc.* **2010**, *132* (26), 8850–8851. <https://doi.org/10.1021/ja103087w>.
- (26) Padovan, A.; Keszei, A.; Köllner, T. G.; Degenhardt, J.; Foley, W. J. The Molecular Basis of Host Plant Selection in *Melaleuca Quinquenervia* by a Successful Biological Control Agent. *Phytochemistry* **2010**, *71* (11), 1237–1244. <https://doi.org/10.1016/j.phytochem.2010.05.013>.
- (27) Nagamine, S.; Liu, C.; Nishishita, J.; Kozaki, T.; Sogahata, K.; Sato, Y.; Minami, A.; Ozaki, T.; Schmidt-Dannert, C.; Maruyama, J.; Oikawa, H. Ascomycete *Aspergillus Oryzae* Is an Efficient Expression Host for Production of Basidiomycete Terpenes by Using Genomic DNA Sequences. *Applied and Environmental Microbiology* **2019**, *85* (15), e00409-19. <https://doi.org/10.1128/AEM.00409-19>.
- (28) D'Ambrosio, M.; Ciocarlan, A.; Colombo, E.; Guerriero, A.; Pizza, C.; Sangiovanni, E.; Dell'Agli, M. Structure and Cytotoxic Activity of Sesquiterpene Glycoside Esters from *Calendula Officinalis* L.: Studies on the Conformation of Viridiflorol. *Phytochemistry* **2015**, *117*, 1–9. <https://doi.org/10.1016/j.phytochem.2015.05.005>.

- (29) de Matos Balsalobre, N.; dos Santos, E.; Mariano dos Santos, S.; Arena, A. C.; Konkiewitz, E. C.; Ziff, E. B.; Nazari Formagio, A. S.; Leite Kassuya, C. A. Potential Anti-Arthritic and Analgesic Properties of Essential Oil and Viridiflorol Obtained from *Allophylus Edulis* Leaves in Mice. *Journal of Ethnopharmacology* **2023**, *301*, 115785. <https://doi.org/10.1016/j.jep.2022.115785>.
- (30) Gilabert, M.; Marcinkevicius, K.; Andujar, S.; Schiavone, M.; Arena, M. E.; Bardón, A. Sesqui- and Triterpenoids from the Liverwort *Lepidozia Chordulifera* Inhibitors of Bacterial Biofilm and Elastase Activity of Human Pathogenic Bacteria. *Phytomedicine* **2015**, *22* (1), 77–85. <https://doi.org/10.1016/j.phymed.2014.10.006>.
- (31) Hulley, I. M.; van Vuuren, S. F.; Sadgrove, N. J.; van Wyk, B.-E. Antimicrobial Activity of *Elytropappus Rhinocerotis* (Asteraceae) against Micro-Organisms Associated with Foot Odour and Skin Ailments. *Journal of Ethnopharmacology* **2019**, *228*, 92–98. <https://doi.org/10.1016/j.jep.2018.09.014>.
- (32) Youssef, N. H.; Al-Huqail, A. A.; Ali, H. M.; Abdelsalam, N. R.; Sabra, M. A. The Role of Serendipita Indica and Lactobacilli Mixtures on Mitigating Mycotoxins and Heavy Metals' Risks of Contaminated Sewage Sludge and Its Composts. *Sci Rep* **2020**, *10* (1), 15159. <https://doi.org/10.1038/s41598-020-71917-8>.
- (33) Reimer, L. C.; Sardà Carbasse, J.; Koblitiz, J.; Ebeling, C.; Podstawka, A.; Overmann, J. BacDive in 2022: The Knowledge Base for Standardized Bacterial and Archaeal Data. *Nucleic Acids Research* **2022**, *50* (D1), D741–D746. <https://doi.org/10.1093/nar/gkab961>.
- (34) Jian, S.-P.; Bau, T.; Zhu, X.-T.; Deng, W.-Q.; Yang, Z. L.; Zhao, Z.-W. Clitopilus, Clitocella, and Clitopilopsis in China. *Mycologia* **2020**, *112* (2), 371–399. <https://doi.org/10.1080/00275514.2019.1703089>.
- (35) *Prunella vulgaris* subsp. *lanceolata* (Heal All, Selfheal) | North Carolina Extension Gardener Plant Toolbox. <https://plants.ces.ncsu.edu/plants/prunella-vulgaris-subsp-lanceolata/> (accessed 2024-11-28).
- (36) *Antirrhinum majus* (Common Snapdragon, Garden Snapdragon, Snapdragon) | North Carolina Extension Gardener Plant Toolbox. <https://plants.ces.ncsu.edu/plants/antirrhinum-majus/> (accessed 2024-11-28).
- (37) NIST. *Automated Mass Spectral Deconvolution and Identification System [Software]*. National Institute of Standards and Technology. <https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdis> (accessed 2024-11-22).
- (38) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 Years of Image Analysis. *Nat Methods* **2012**, *9* (7), 671–675. <https://doi.org/10.1038/nmeth.2089>.

- (39) Najar, B.; Mecacci, G.; Nardi, V.; Cervelli, C.; Nardoni, S.; Mancianti, F.; Ebani, V. V.; Giannecchini, S.; Pistelli, L. Volatiles and Antifungal-Antibacterial-Antiviral Activity of South African *Salvia* Spp. Essential Oils Cultivated in Uniform Conditions. *Molecules* **2021**, *26* (9), 2826. <https://doi.org/10.3390/molecules26092826>.
- (40) Ajayi-Oyetunde, O. O.; Bradley, C. A. *Rhizoctonia Solani*: Taxonomy, Population Biology and Management of *Rhizoctonia* Seedling Disease of Soybean. *Plant Pathology* **2018**, *67* (1), 3–17. <https://doi.org/10.1111/ppa.12733>.
- (41) Tsrar, L. Biology, Epidemiology and Management of *Rhizoctonia Solani* on Potato. *Journal of Phytopathology* **2010**, *158* (10), 649–658. <https://doi.org/10.1111/j.1439-0434.2010.01671.x>.
- (42) El-Tarabily, K. A. Suppression of *Rhizoctonia Solani* Diseases of Sugar Beet by Antagonistic and Plant Growth-promoting Yeasts. *Journal of Applied Microbiology* **2004**, *96* (1), 69–75. <https://doi.org/10.1046/j.1365-2672.2003.02043.x>.
- (43) Senapati, M.; Tiwari, A.; Sharma, N.; Chandra, P.; Bashyal, B. M.; Ellur, R. K.; Bhowmick, P. K.; Bollinedi, H.; Vinod, K. K.; Singh, A. K.; Krishnan, S. G. *Rhizoctonia Solani* Kühn Pathophysiology: Status and Prospects of Sheath Blight Disease Management in Rice. *Front. Plant Sci.* **2022**, *13*. <https://doi.org/10.3389/fpls.2022.881116>.
- (44) Pascual, C. B.; RAYMUNDO, A. D.; HYAKUMACHI, M. Resistance of Sorghum Line CS 621 to *Rhizoctonia Solani* AG1-IA and Other Sorghum Pathogens. *J Gen Plant Pathol* **2000**, *66* (1), 23–29. <https://doi.org/10.1007/PL00012918>.
- (45) Li, P.; Tedersoo, L.; Crowther, T. W.; Wang, B.; Shi, Y.; Kuang, L.; Li, T.; Wu, M.; Liu, M.; Luan, L.; Liu, J.; Li, D.; Li, Y.; Wang, S.; Saleem, M.; Dumbrell, A. J.; Li, Z.; Jiang, J. Global Diversity and Biogeography of Potential Phytopathogenic Fungi in a Changing World. *Nat Commun* **2023**, *14* (1), 6482. <https://doi.org/10.1038/s41467-023-42142-4>.
- (46) Pripdeevech, P.; Chukeatirote, E. Chemical Compositions, Antifungal and Antioxidant Activities of Essential Oil and Various Extracts of *Melodorum Fruticosum* L. Flowers. *Food and Chemical Toxicology* **2010**, *48* (10), 2754–2758. <https://doi.org/10.1016/j.fct.2010.07.002>.
- (47) Scher, J. M.; Speakman, J.-B.; Zapp, J.; Becker, H. Bioactivity Guided Isolation of Antifungal Compounds from the Liverwort *Bazzania Trilobata* (L.) S.F. Gray. *Phytochemistry* **2004**, *65* (18), 2583–2588. <https://doi.org/10.1016/j.phytochem.2004.05.013>.

- (48) De Silva, D. D.; Crous, P. W.; Ades, P. K.; Hyde, K. D.; Taylor, P. W. J. Life Styles of *Colletotrichum* Species and Implications for Plant Biosecurity. *Fungal Biology Reviews* **2017**, 31 (3), 155–168. <https://doi.org/10.1016/j.fbr.2017.05.001>.
- (49) Aoki, T.; O'Donnell, K.; Geiser, D. M. Systematics of Key Phytopathogenic Fusarium Species: Current Status and Future Challenges. *J Gen Plant Pathol* **2014**, 80 (3), 189–201. <https://doi.org/10.1007/s10327-014-0509-3>.
- (50) Marquez, N.; Giachero, M. L.; Declerck, S.; Ducasse, D. A. Macrophomina Phaseolina: General Characteristics of Pathogenicity and Methods of Control. *Front. Plant Sci.* **2021**, 12. <https://doi.org/10.3389/fpls.2021.634397>.
- (51) Whipps, J. M.; Lumsden, R. D. Biological Control of *Pythium* Species. *Biocontrol Science and Technology* **1991**, 1 (2), 75–90. <https://doi.org/10.1080/09583159109355188>.
- (52) Kai, M.; Effmert, U.; Berg, G.; Piechulla, B. Volatiles of Bacterial Antagonists Inhibit Mycelial Growth of the Plant Pathogen *Rhizoctonia Solani*. *Arch Microbiol* **2007**, 187 (5), 351–360. <https://doi.org/10.1007/s00203-006-0199-0>.
- (53) Zhang, X.-J.; Lai, T.-B.; Kong, R. Y.-C. Biology of Fluoro-Organic Compounds. In *Fluorous Chemistry*; Horváth, I. T., Ed.; Springer: Berlin, Heidelberg, 2012; pp 365–404. https://doi.org/10.1007/128_2011_270.

APPENDIX

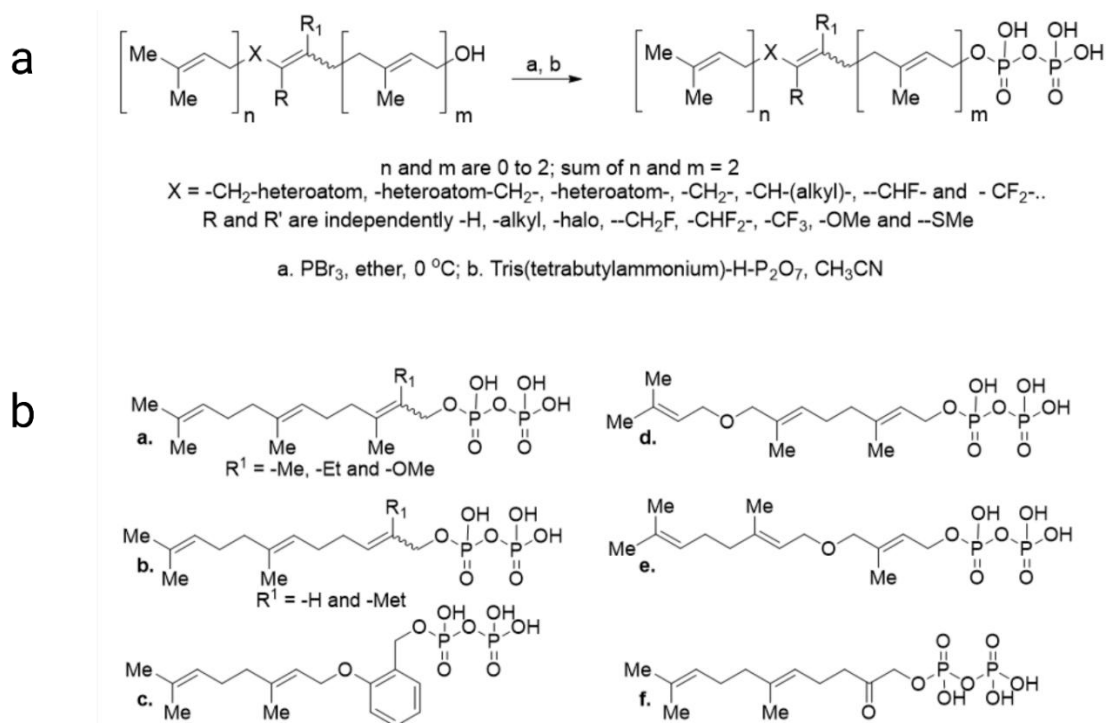


Figure S4.1. Schematic of FDP analog synthesis. a illustrates broadly substitutable modified prenyl groups that can be combinatorially synthesized followed by the addition of the diphosphate. b illustrates example products

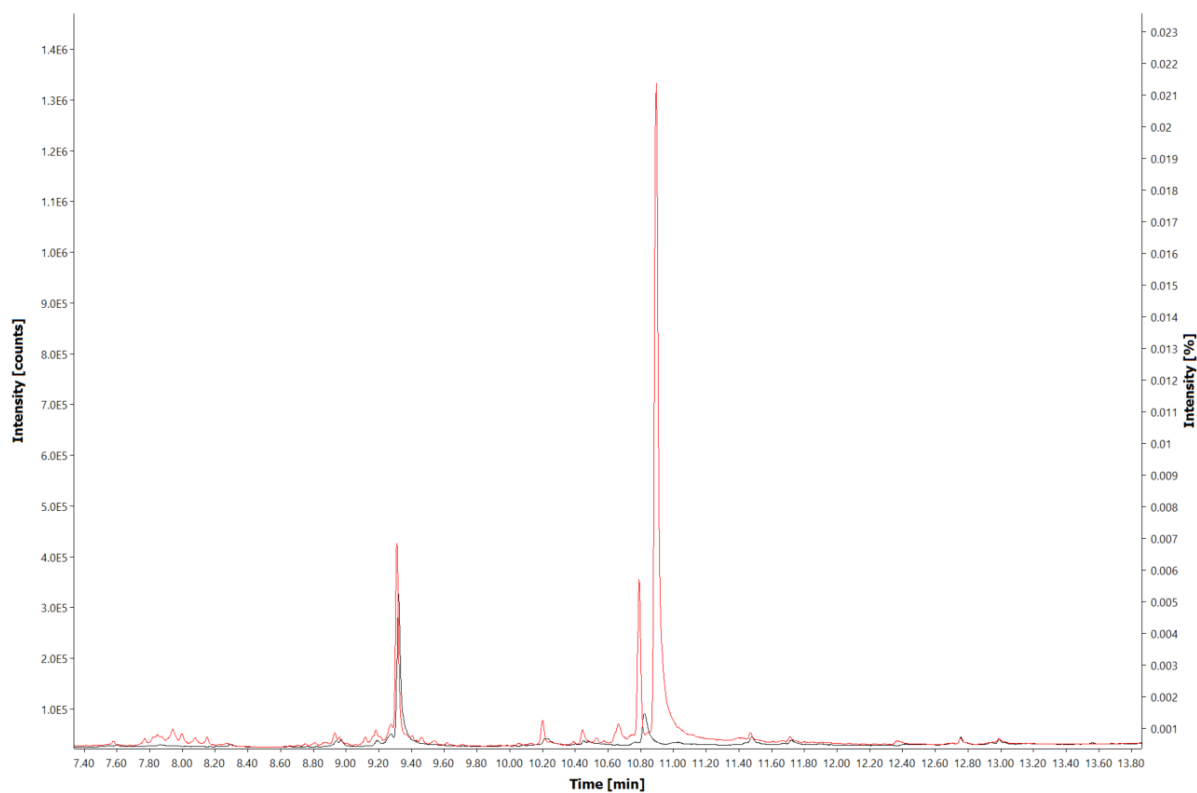


Figure S4.2a. multiproduct synthetic FDP analog—phosphatase combinations. The black trace shows the no protein control while red trace is the phosphatase treatment. Several phosphatase products were found with substrates 3 (S4.2a), 4 (4.2b), 5 (4.2c), 9 (4.2d), 10 (4.2e), and 11 (4.2f).

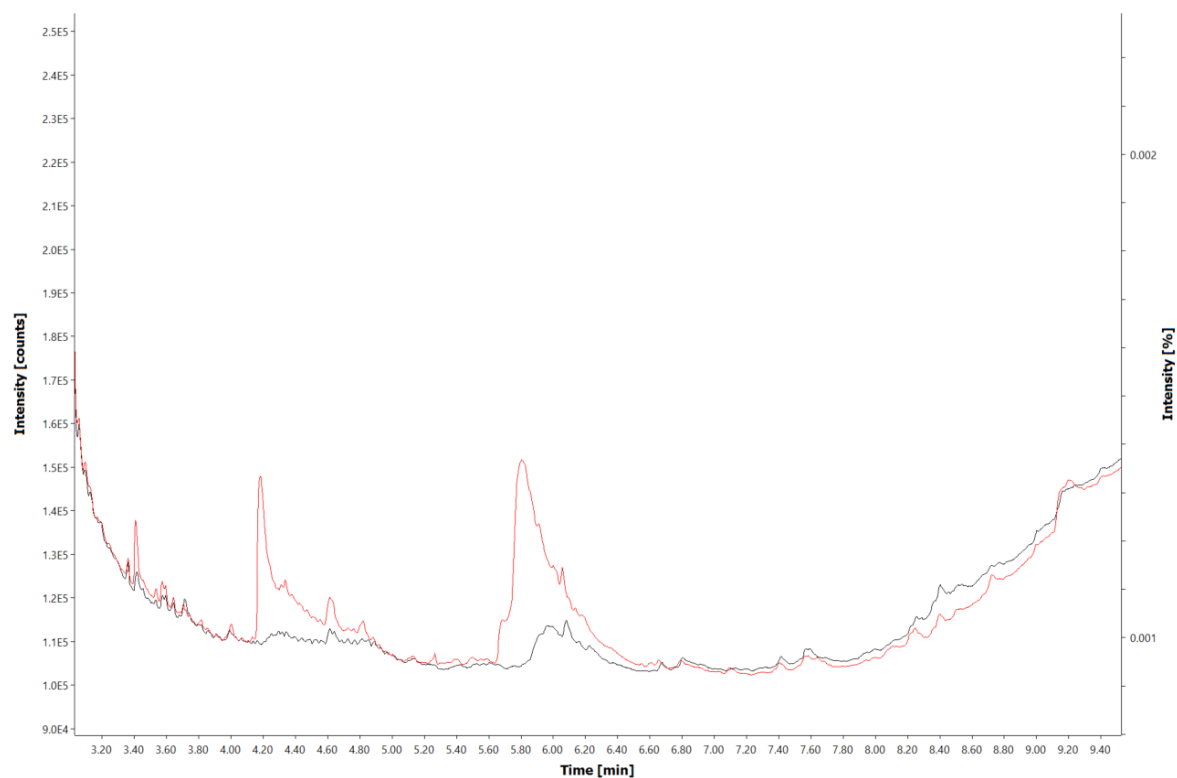


Figure S4.2b. multiproduct synthetic FDP analog—phosphatase combinations. The black trace shows the no protein control while red trace is the phosphatase treatment. Several phosphatase products were found with substrates 3 (S4.2a), 4 (4.2b), 5 (4.2c), 9 (4.2d), 10 (4.2e), and 11 (4.2f).

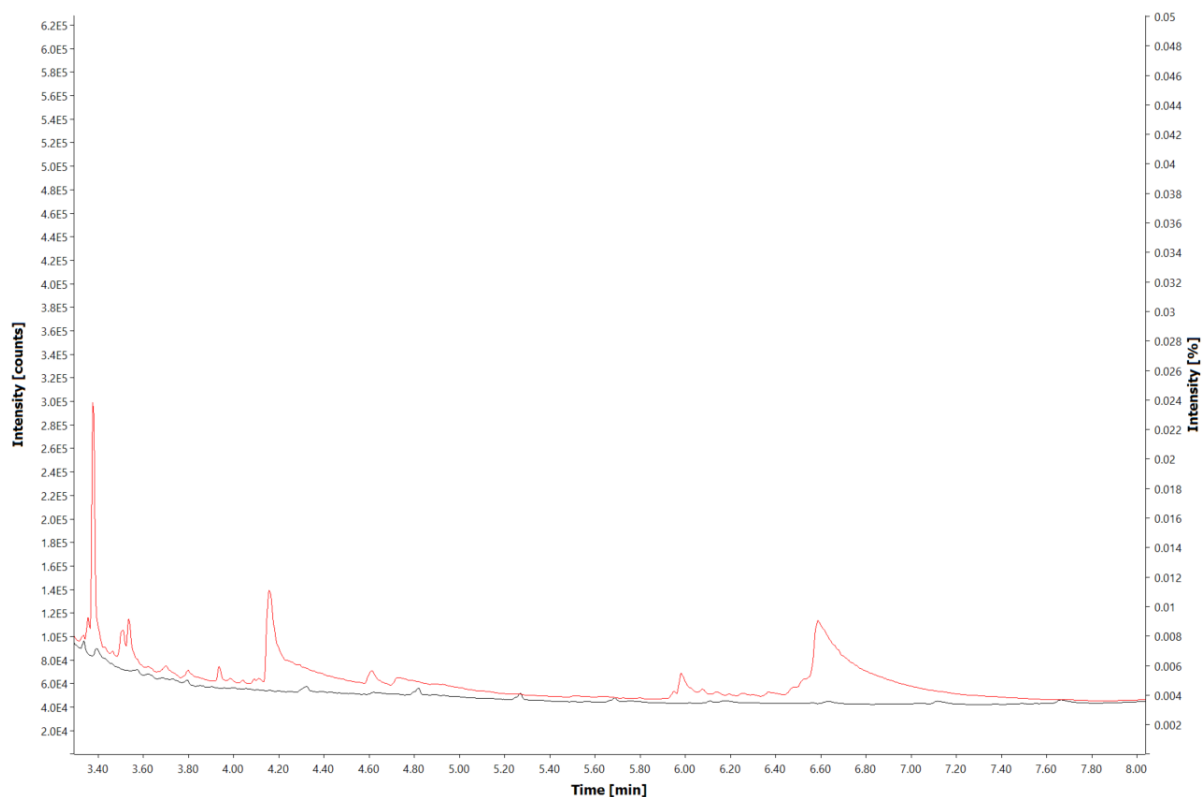


Figure S4.2c. multiproduct synthetic FDP analog—phosphatase combinations. The black trace shows the no protein control while red trace is the phosphatase treatment. Several phosphatase products were found with substrates 3 (S4.2a), 4 (4.2b), 5 (4.2c), 9 (4.2d), 10 (4.2e), and 11 (4.2f).

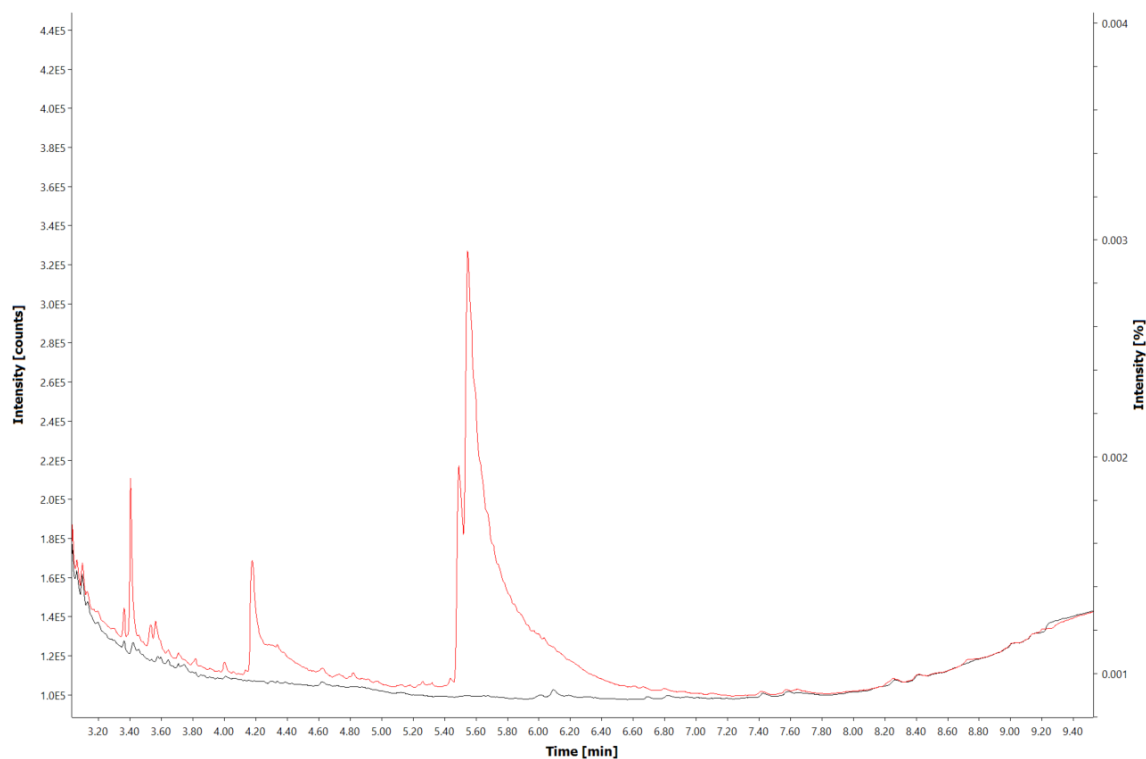


Figure S4.2d. multiproduct synthetic FDP analog—phosphatase combinations. The black trace shows the no protein control while red trace is the phosphatase treatment. Several phosphatase products were found with substrates 3 (S4.2a), 4 (4.2b), 5 (4.2c), 9 (4.2d), 10 (4.2e), and 11 (4.2f).

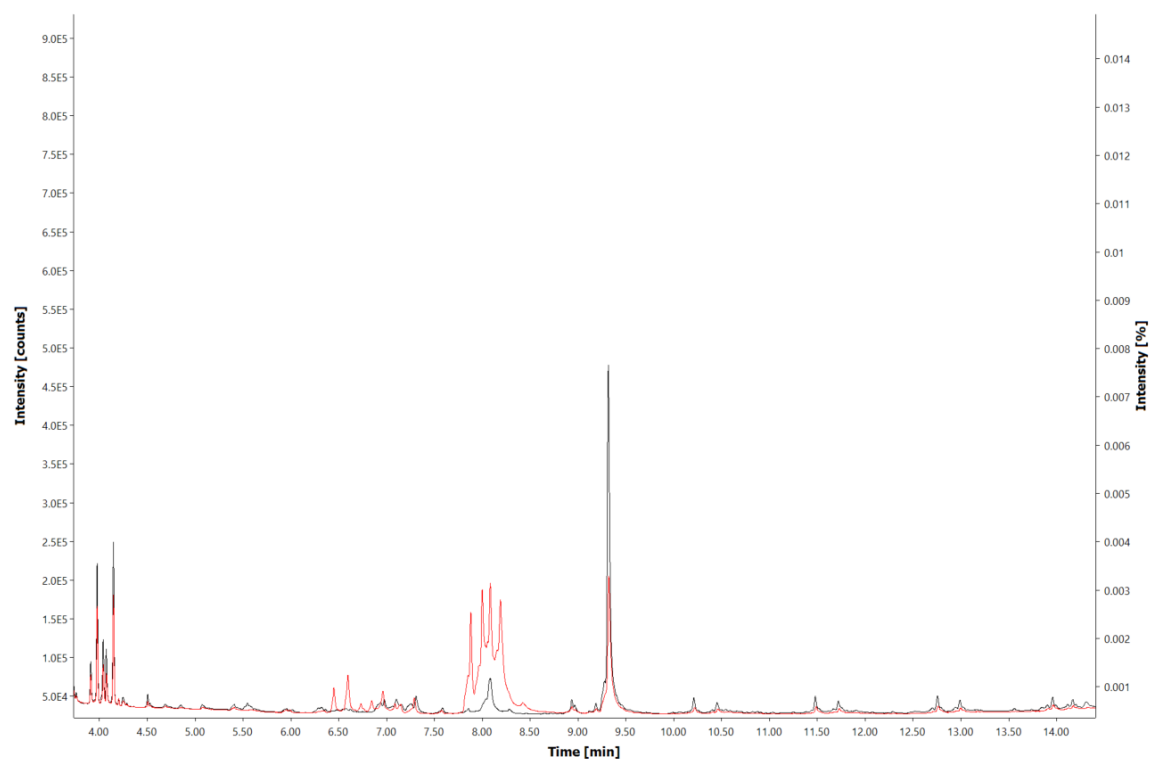


Figure S4.2e. multiproduct synthetic FDP analog—phosphatase combinations. The black trace shows the no protein control while red trace is the phosphatase treatment. Several phosphatase products were found with substrates 3 (S4.2a), 4 (4.2b), 5 (4.2c), 9 (4.2d), 10 (4.2e), and 11 (4.2f).

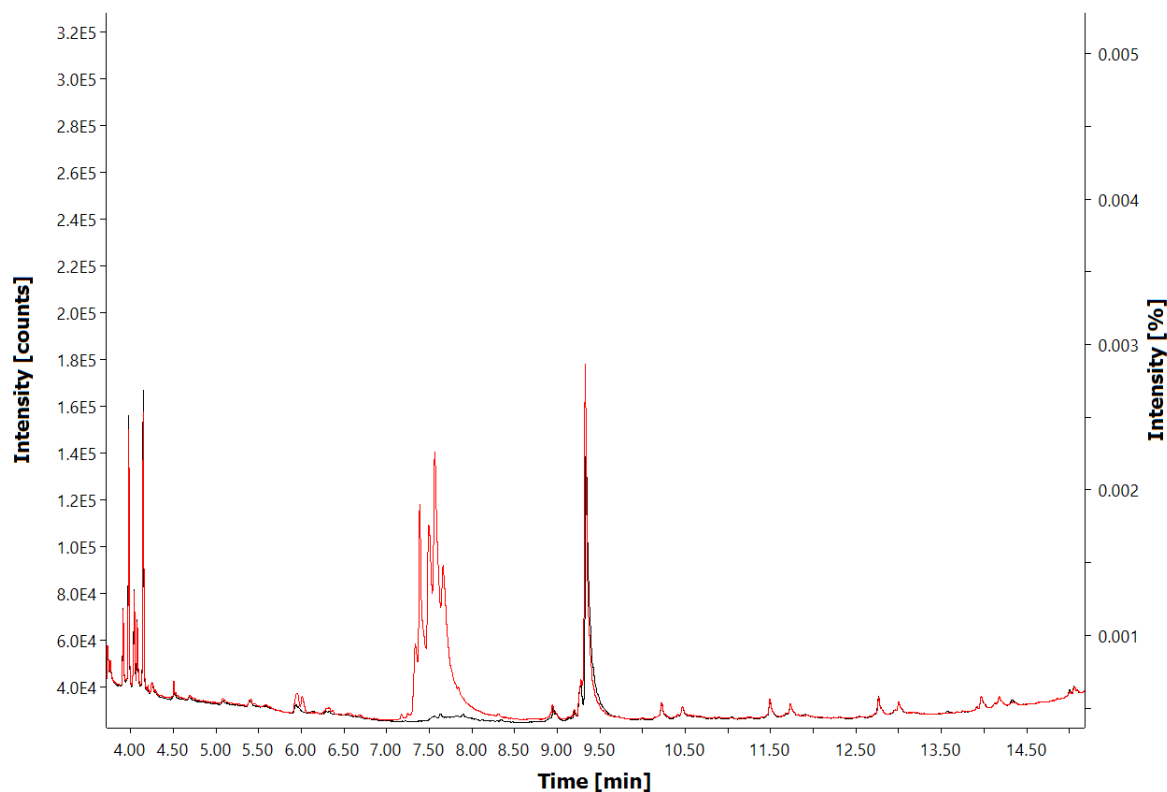


Figure S4.2f. multiproduct synthetic FDP analog—phosphatase combinations. The black trace shows the no protein control while red trace is the phosphatase treatment. Several phosphatase products were found with substrates 3 (S4.2a), 4 (4.2b), 5 (4.2c), 9 (4.2d), 10 (4.2e), and 11 (4.2f).

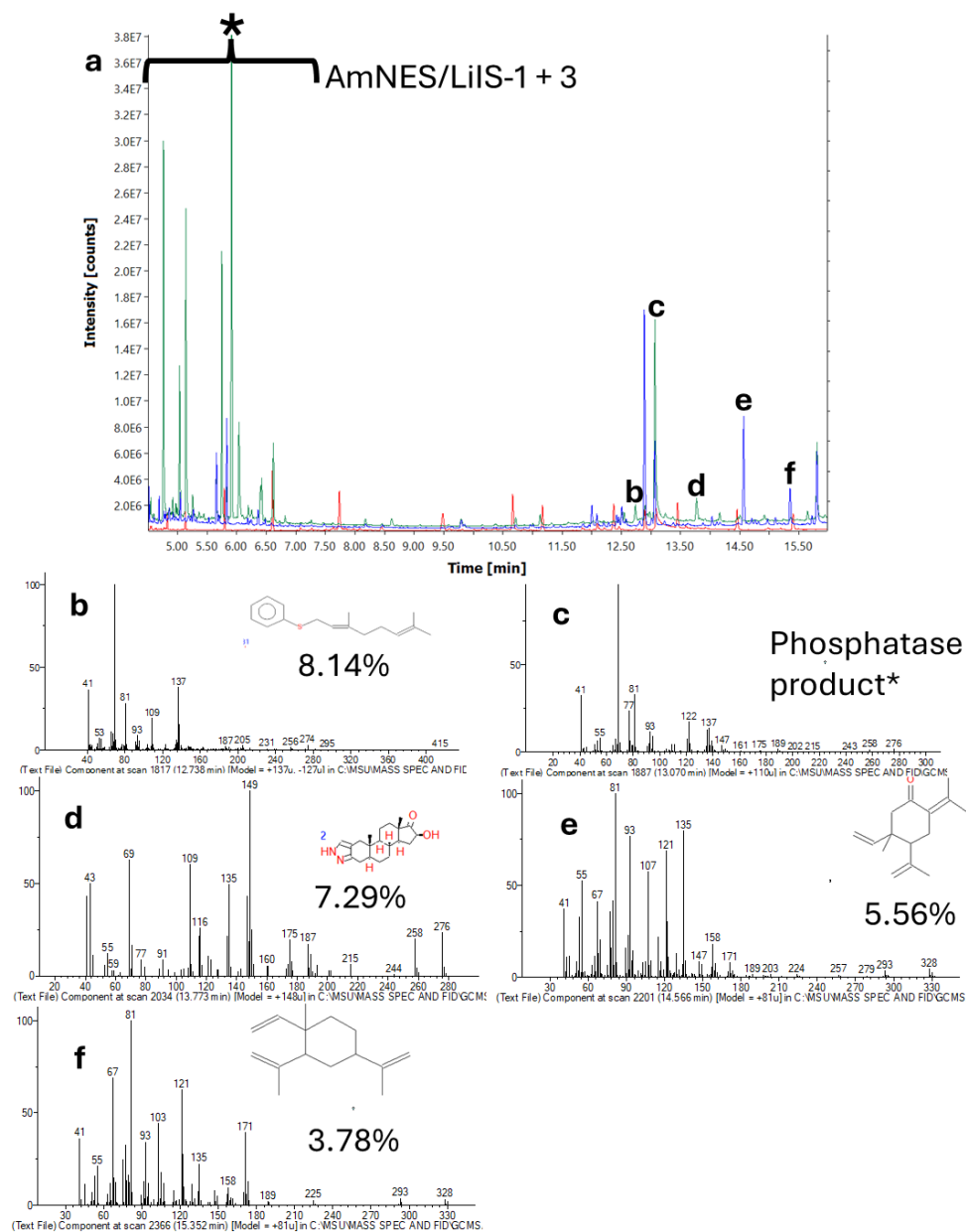


Figure S4.3. AmNES/LIS-1 + Substrate 3 combinations chromatograms and mass spectra. a. GC-MS of the specific fractions used in figure 5. Red is a phosphatase control. Blue is AmNES/LIS-1 Fraction A while green is Fraction B used in the antifungal assays in figure 5. b-f have mass spectra present in figure 3A. * **bracket** shows ~C10 cleaved products.

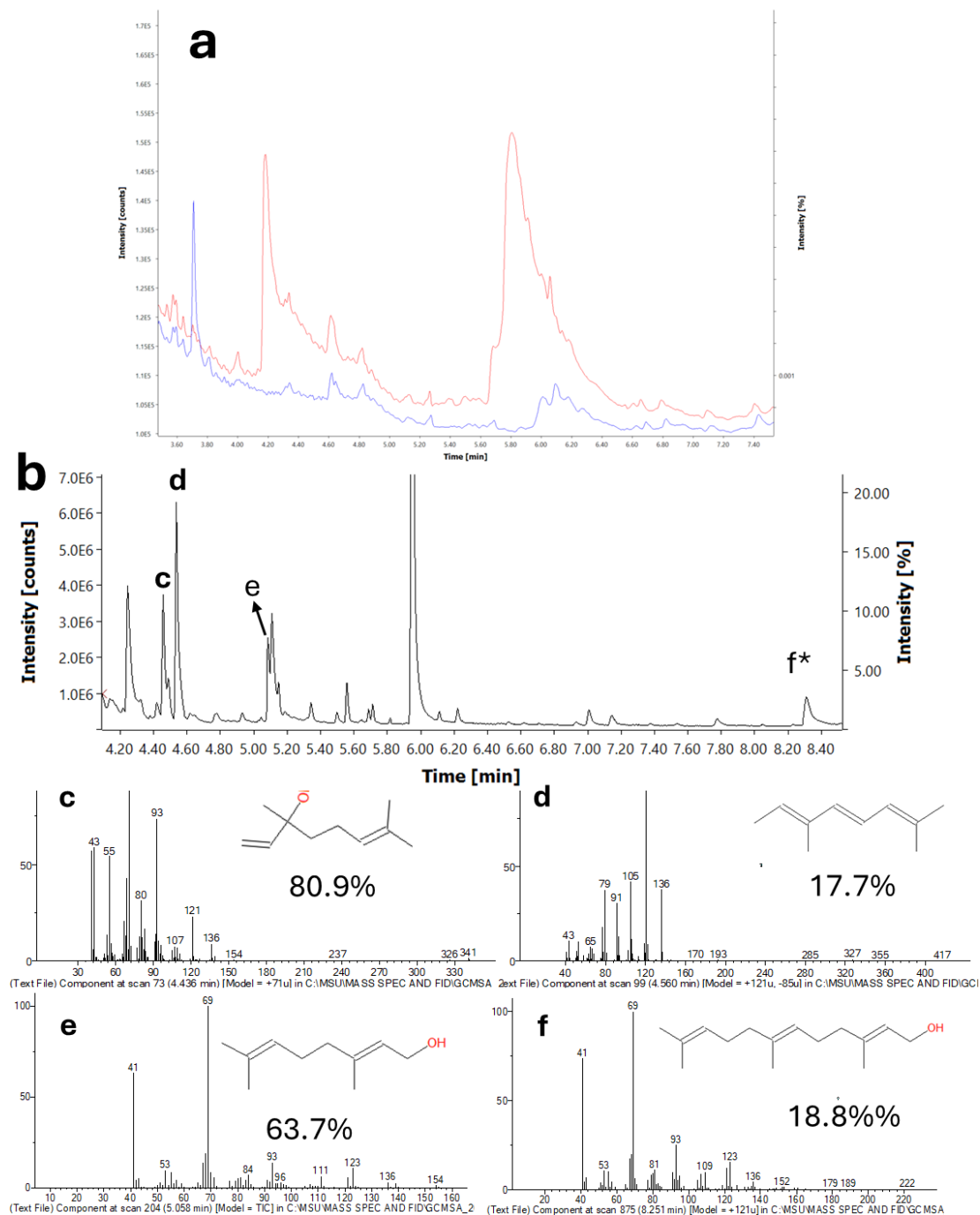


Figure S4.4. Agr5 + substrate 4 chromatograms. A. shows GC-FID of phosphatase (red) and Agr5 (blue). B. GC-MS of the fraction used for anti-fungal assays in figure 5. c-f show relevant mass spectra.

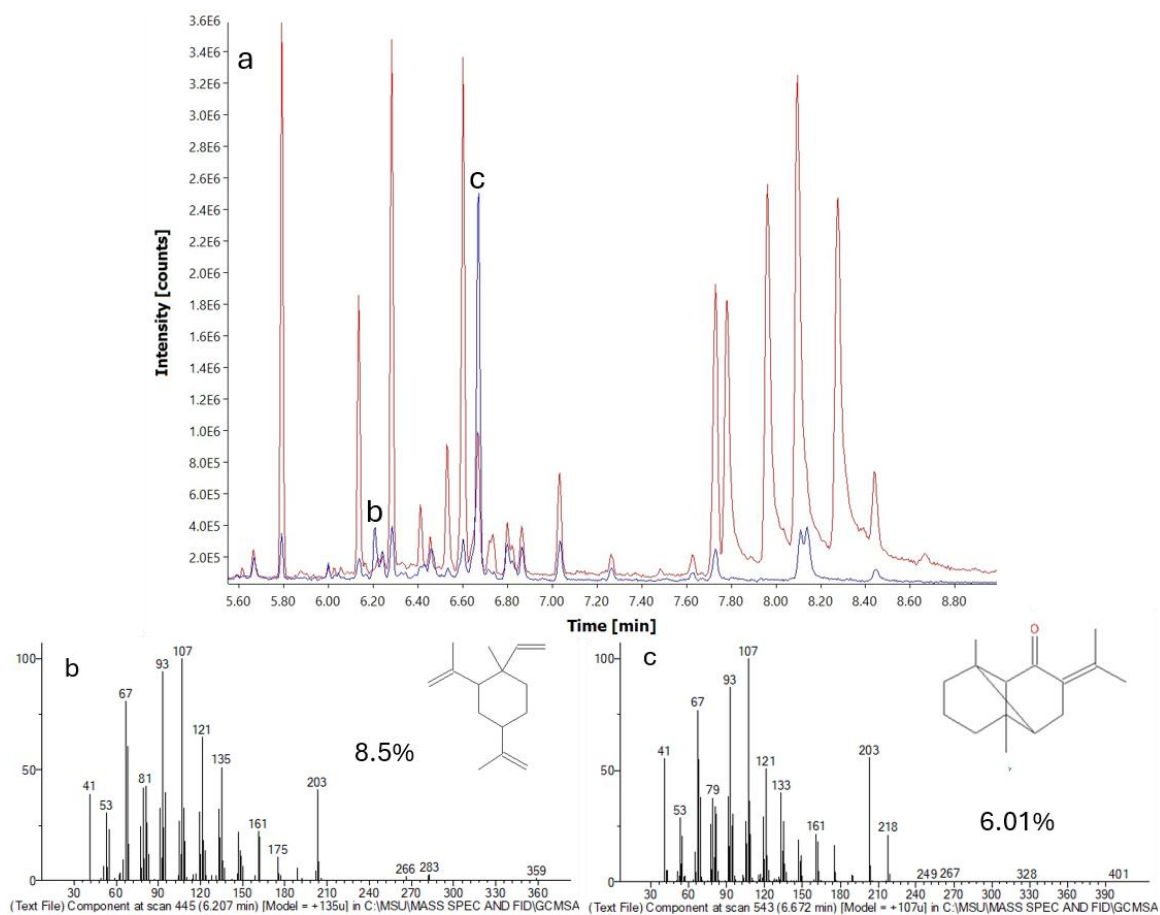


Figure S4.5. SaAMS + Substrate 4 chromatograms. **a.** shows GC-MS of phosphatase (red) and SaAMS (blue) paired with substrate 4. **b-c** are spectra of peaks novel to SaAMS.

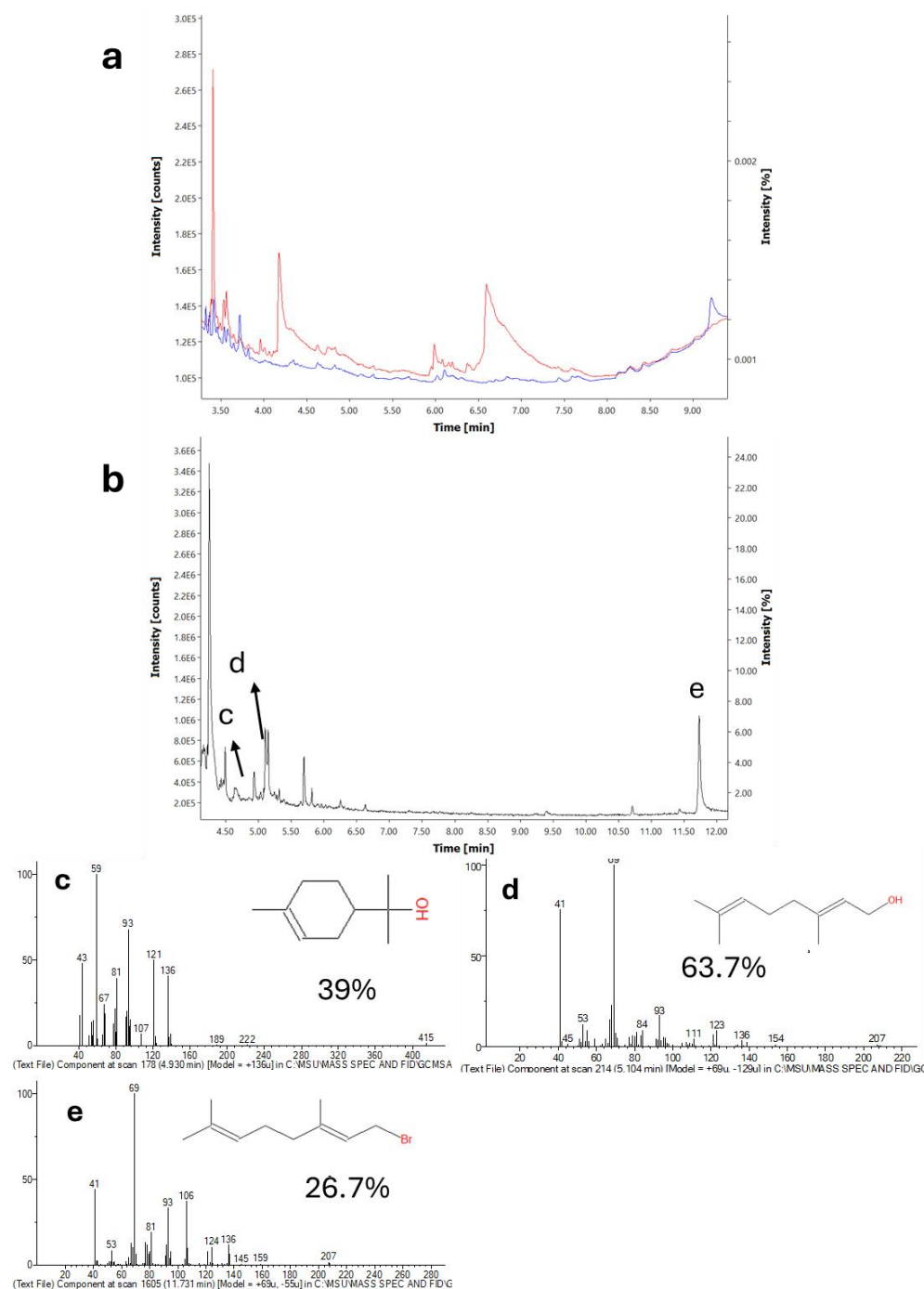


Figure S4.6. SaAMS + substrate 5 chromatograms. A. shows GC-FID of phosphatase (red) and SaAMS (blue). B. GC-MS semi-pure fractions of the combination. C-E show mass spectra of relevant peaks found in b

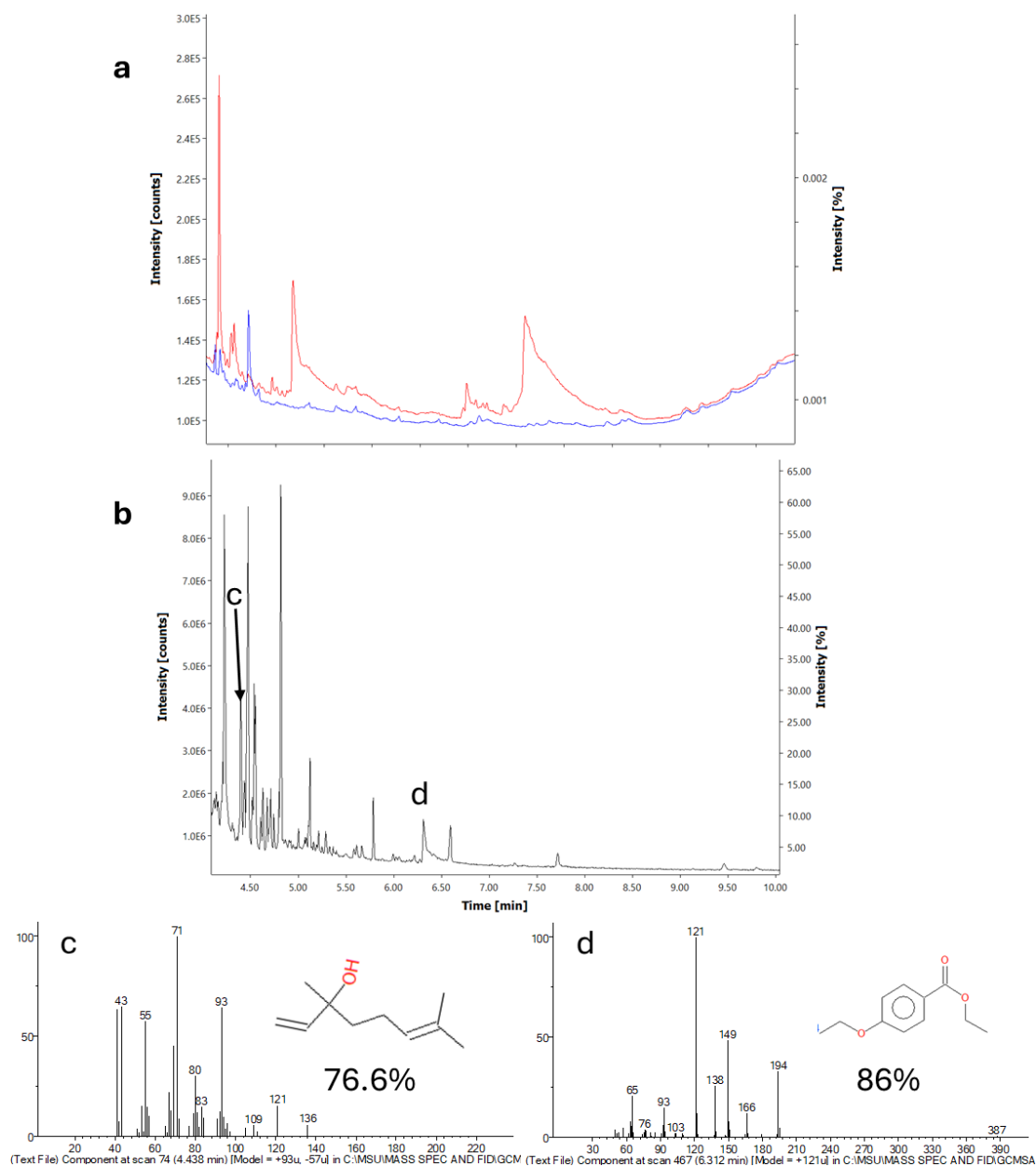


Figure S4.7. Agr5 + substrate 5 chromatograms. A. shows GC-FID of phosphatase (red) and Agr5 (blue) paired with substrate 5. B. GC-MS of these assays. C. and d. shows relevant mass spectra from b.

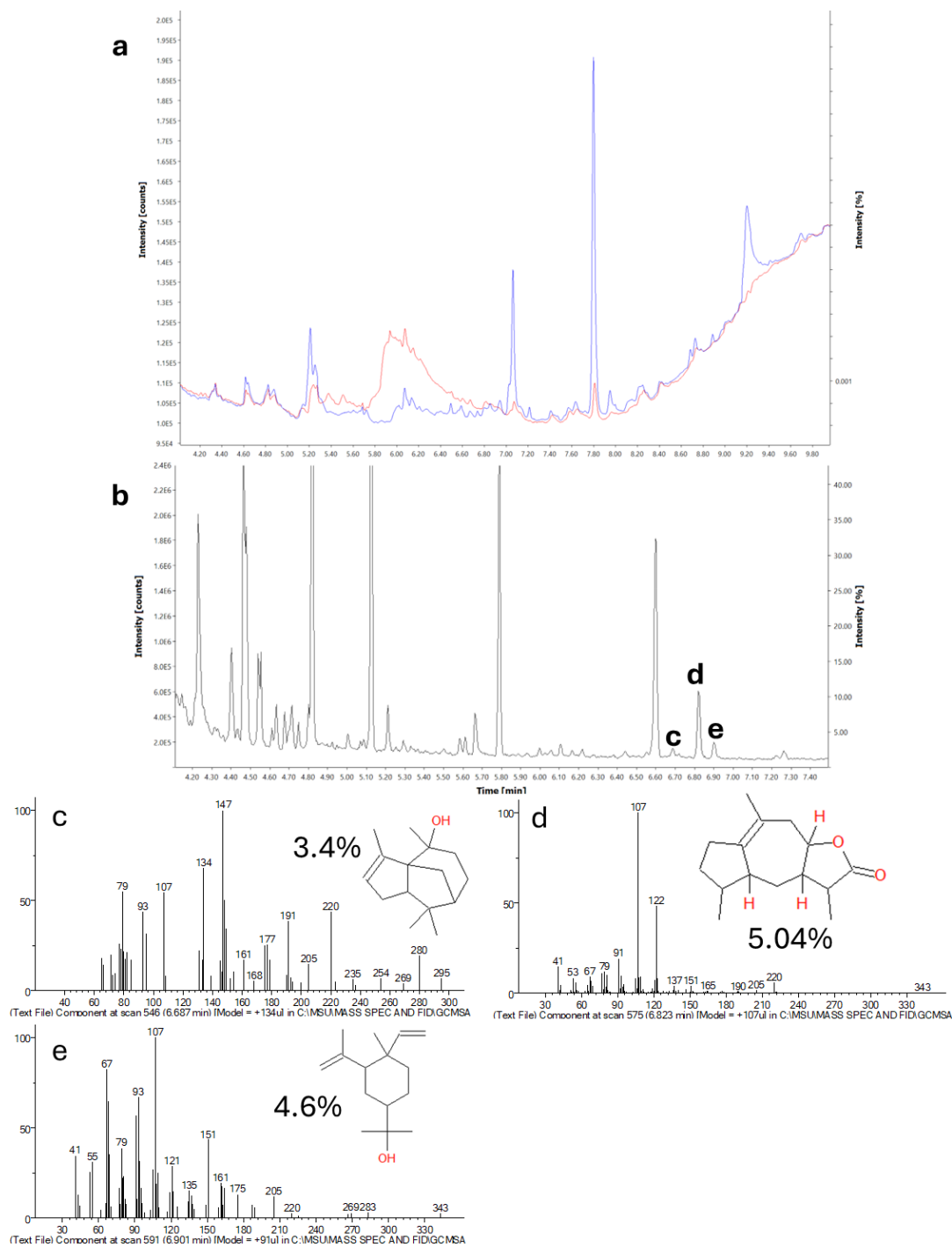


Figure S4.8. SaAMS + substrate 7 chromatograms and mass spectra. a. GC-FID chromatogram. Blue is a TPS combination. Red is phosphatase. b. GC-MS chromatogram. c-e are mass spectra labeled on b.

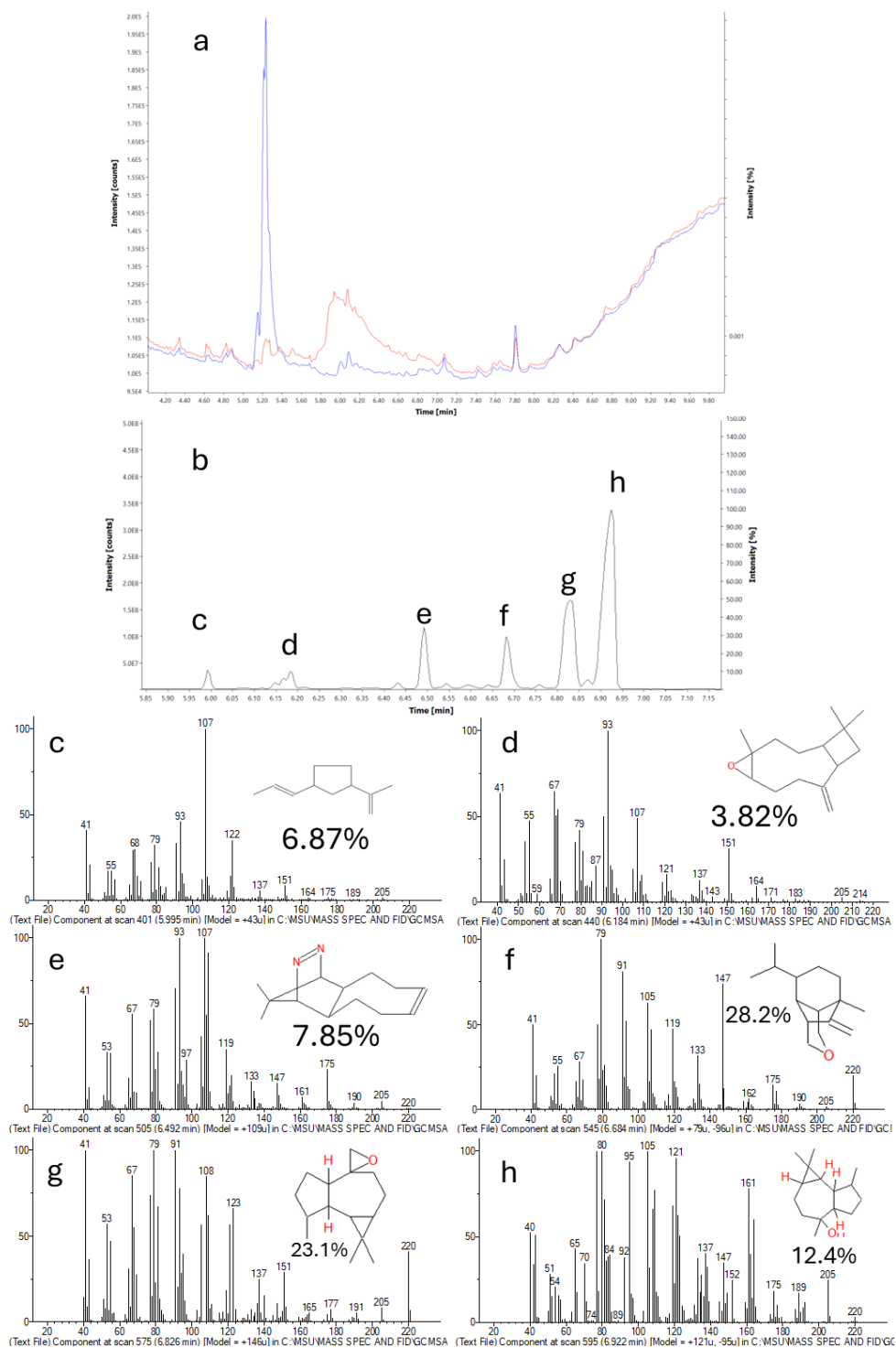


Figure S4.9. CpSTS9 + 7 chromatograms and mass spectra. a. GC-FID chromatograms. Blue is the CpSTS9 + 7 and red is phosphatase control. b. is the GC-MS chromatogram of the fraction used in figure 5. c-h is mass spectra from figure b.

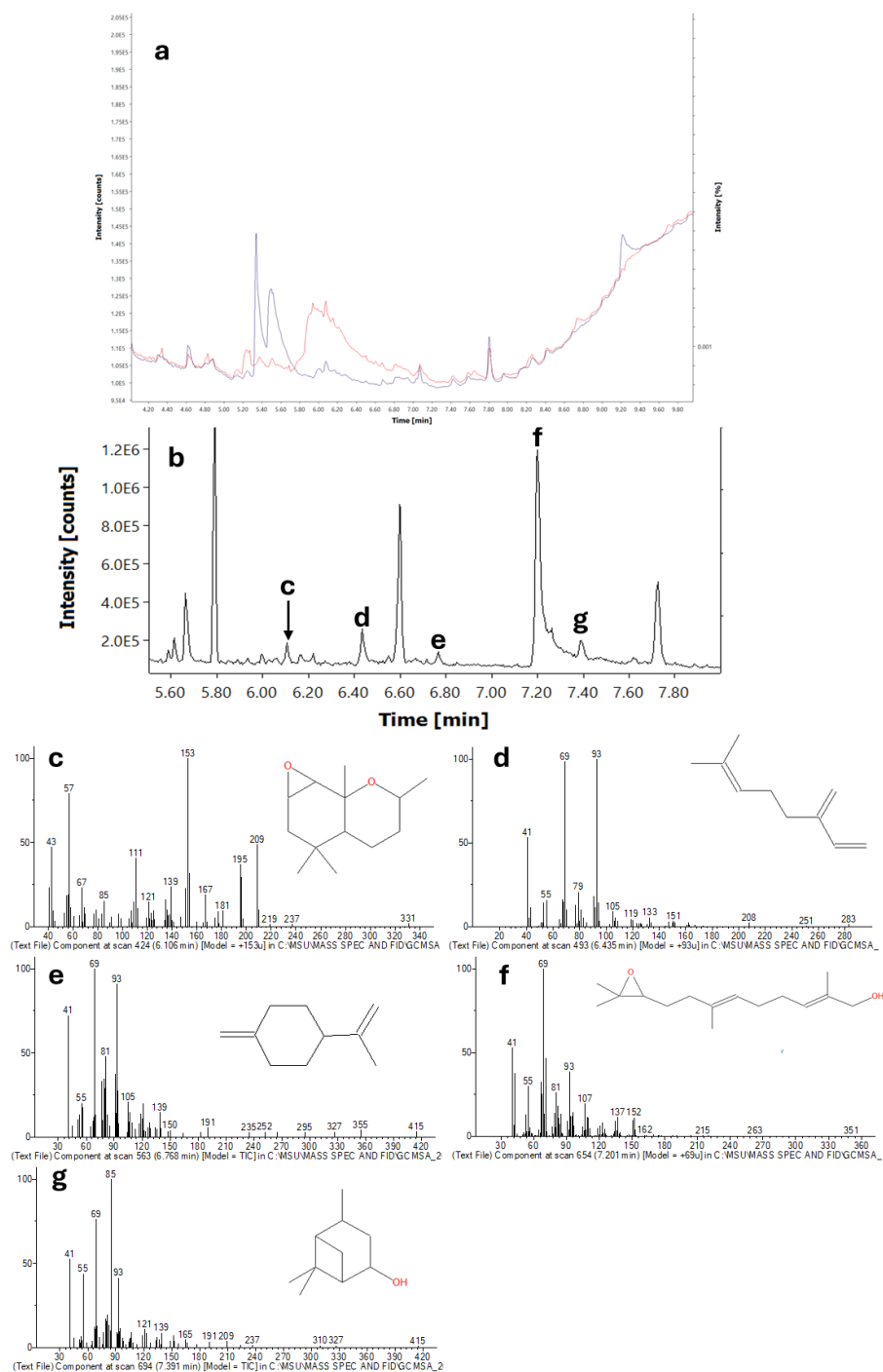


Figure S4.10. Agr + 7 chromatograms and mass spectra. a. GC-FID chromatograms. Blue is the Agr5 + 7 and red is phosphatase control. b. is the GC-MS chromatogram of the fraction used in figure 5. c-g is mass spectra from figure b.

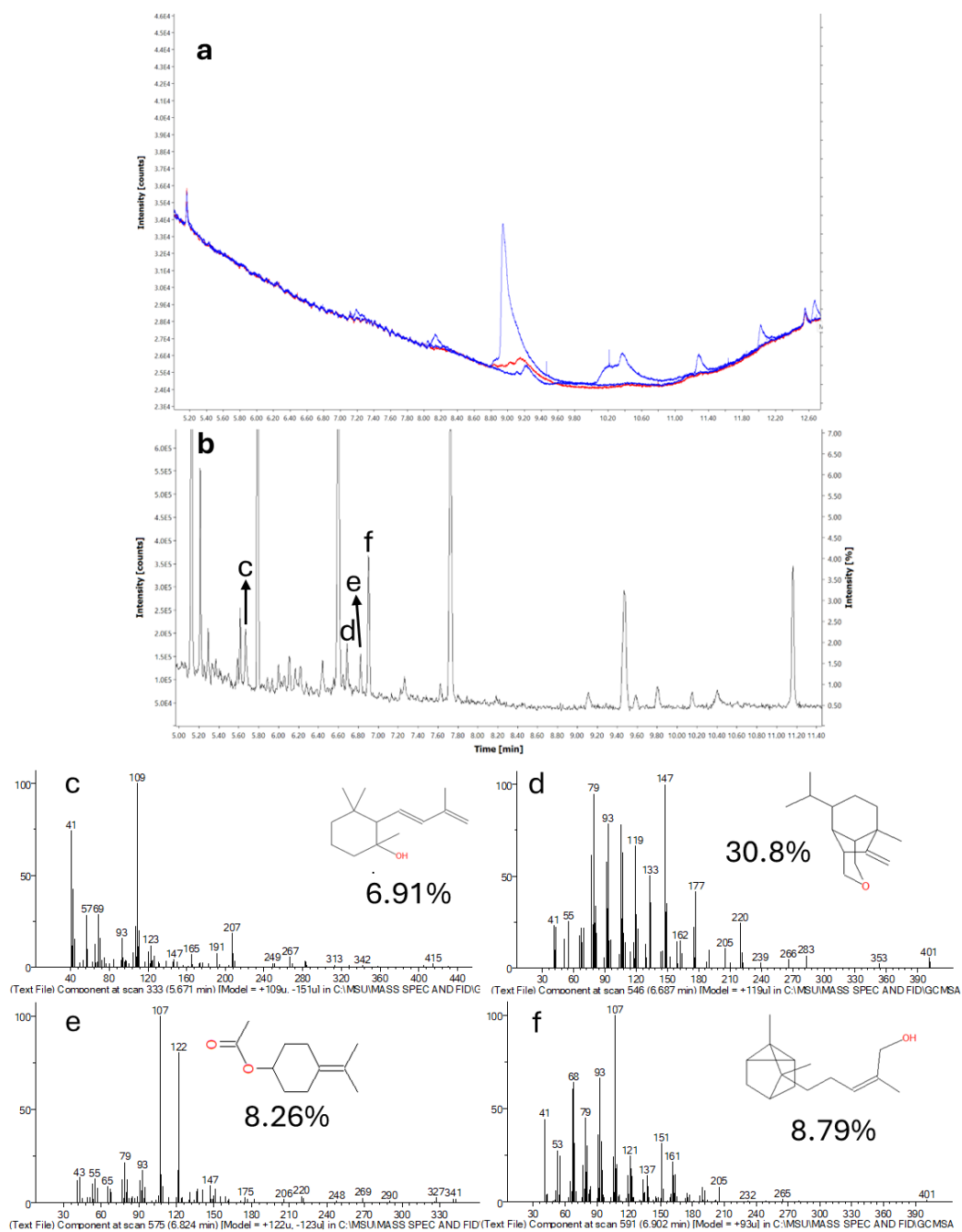


Figure S4.11. SiVS + 7 chromatograms and mass spectra. a. GC-FID chromatograms. Blue is the SiVS + 7 and red is phosphatase control. b. is the GC-MS chromatogram of the fraction used in figure 5. c-f is mass spectra from figure b.

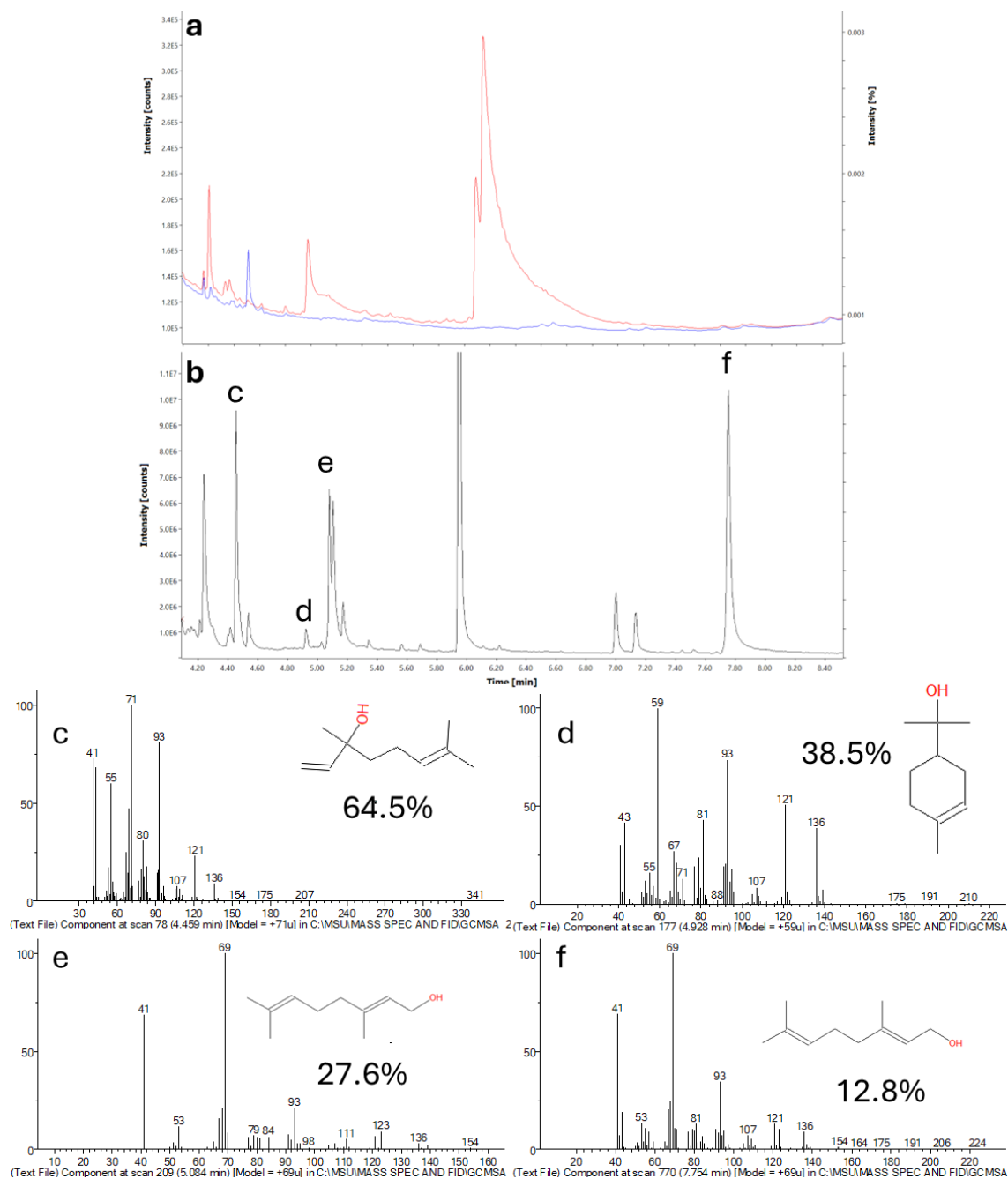


Figure S4.12. Agr5 + 9 chromatograms and mass spectra. a. GC-FID chromatograms. Blue is the Agr5 + 9 and red is phosphatase control. b. is the GC-MS chromatogram of the fraction used in figure 5. c-f is mass spectra from figure b.

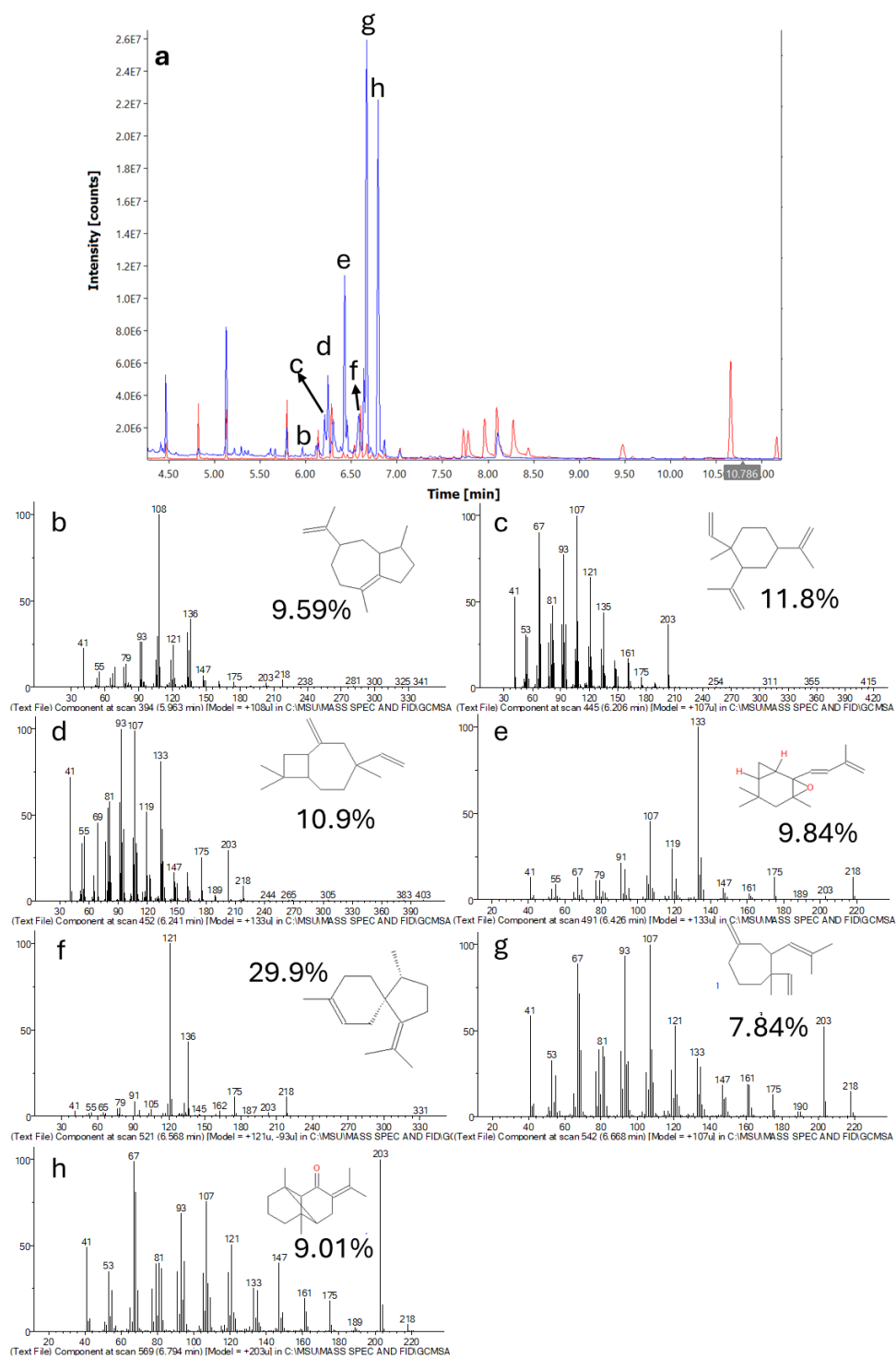
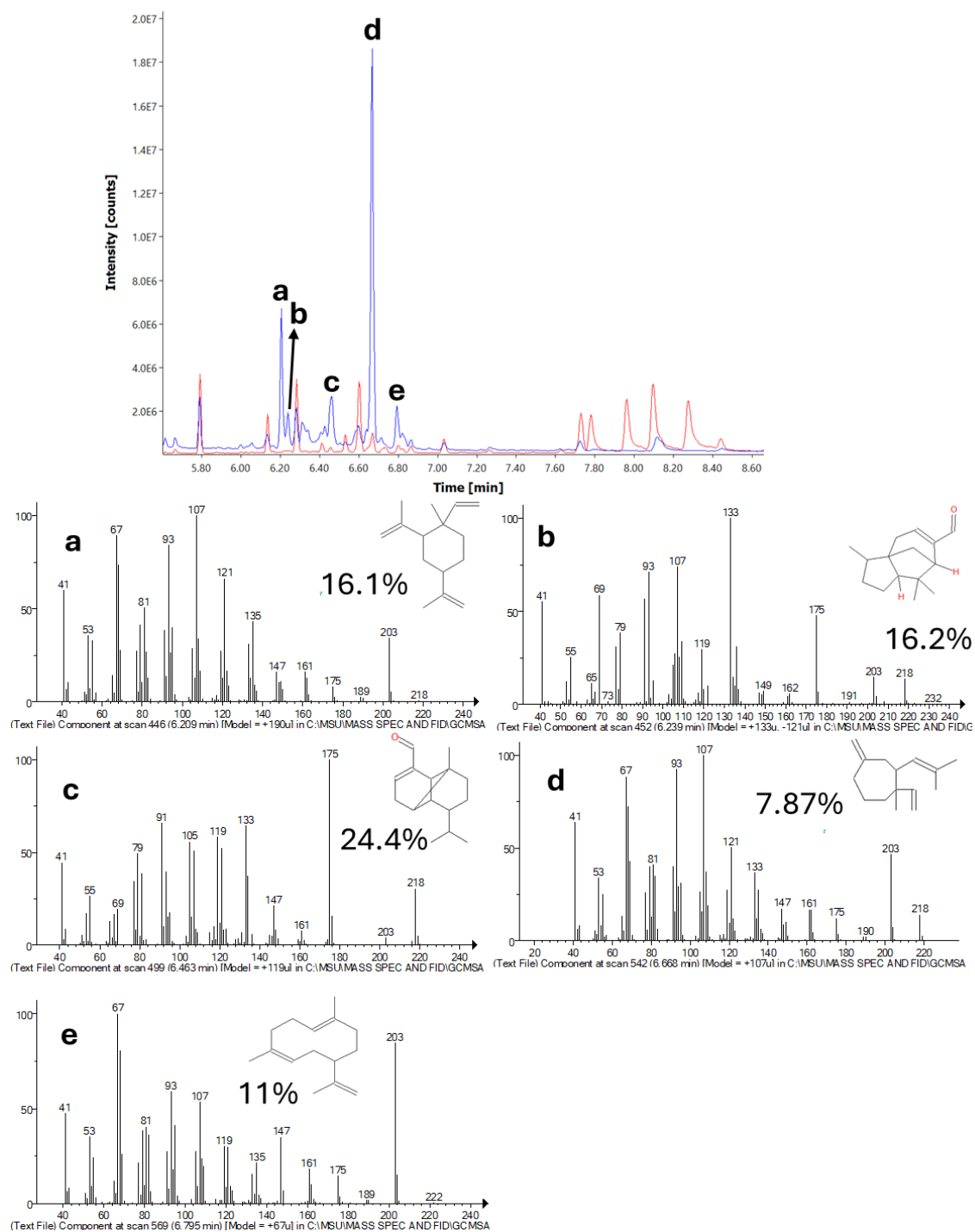


Figure S4.13. SiVS + 10 chromatograms and mass spectra. a. GC-FID chromatograms. Blue is the SiVS + 10 and red is phosphatase control. b. is the GC-MS chromatogram of the fraction used in figure 5. c-h is mass spectra from figure b.



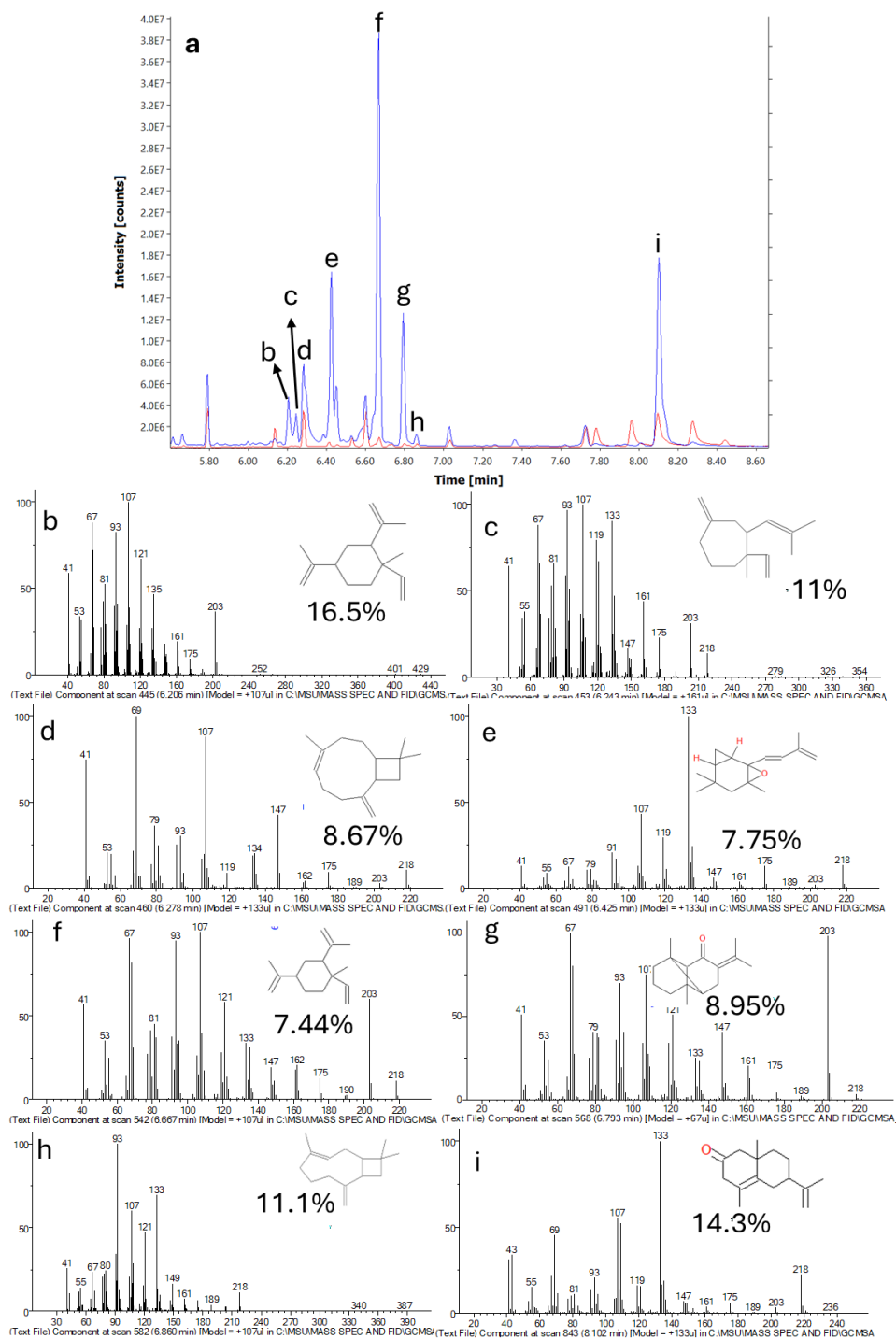


Figure S4.15. CpSTS9 + 10 chromatograms and mass spectra. a. GC-MS data showing blue as CpSTS9 +10 and red is phosphatase control. b-k is mass spectra from figure b.

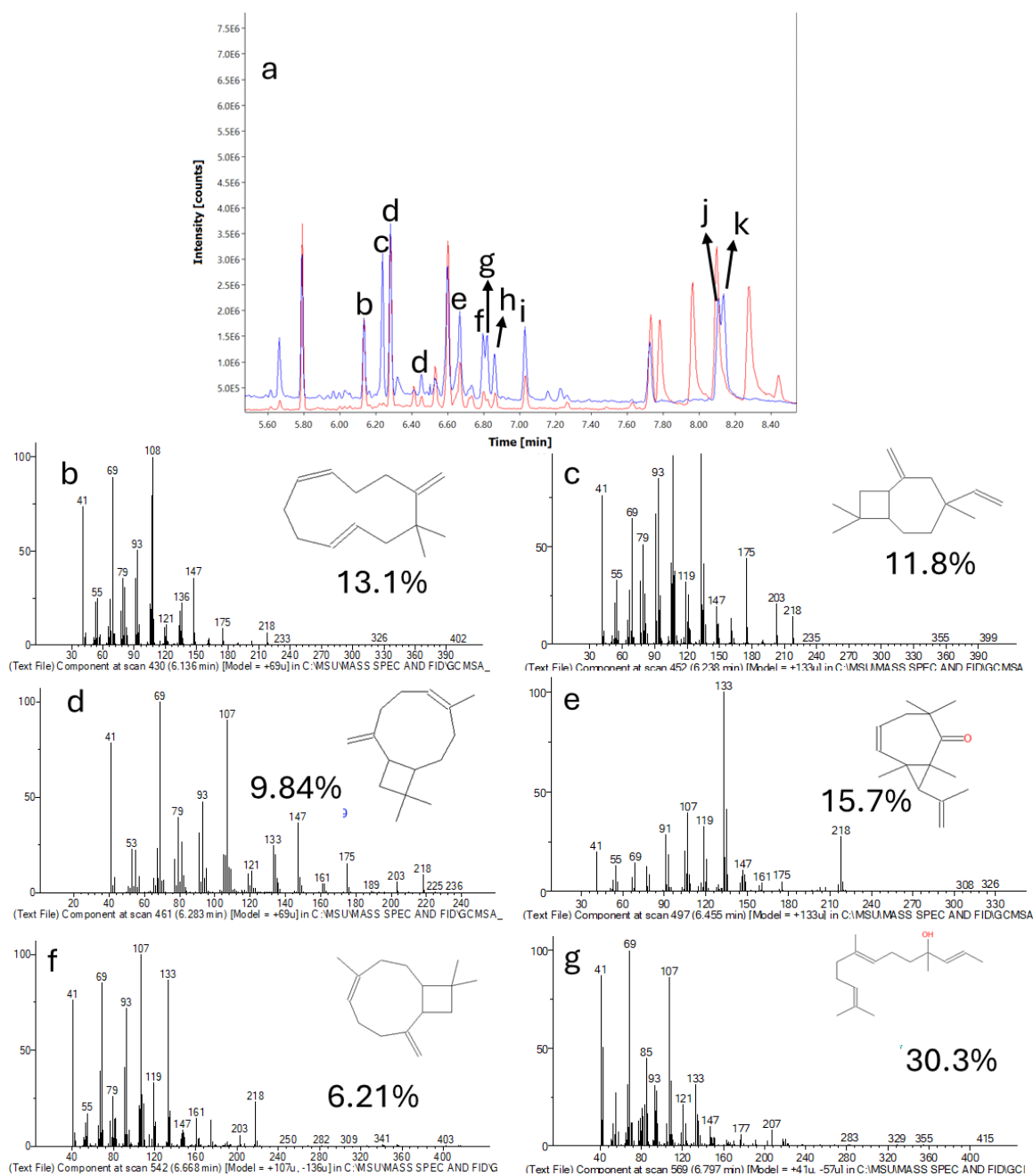


Figure S4.16. PvTPS5 + 10 chromatograms and mass spectra. a. GC-MS data showing blue as CpSTS9 +10 and red is phosphatase control. b-k is mass spectra from figure b.

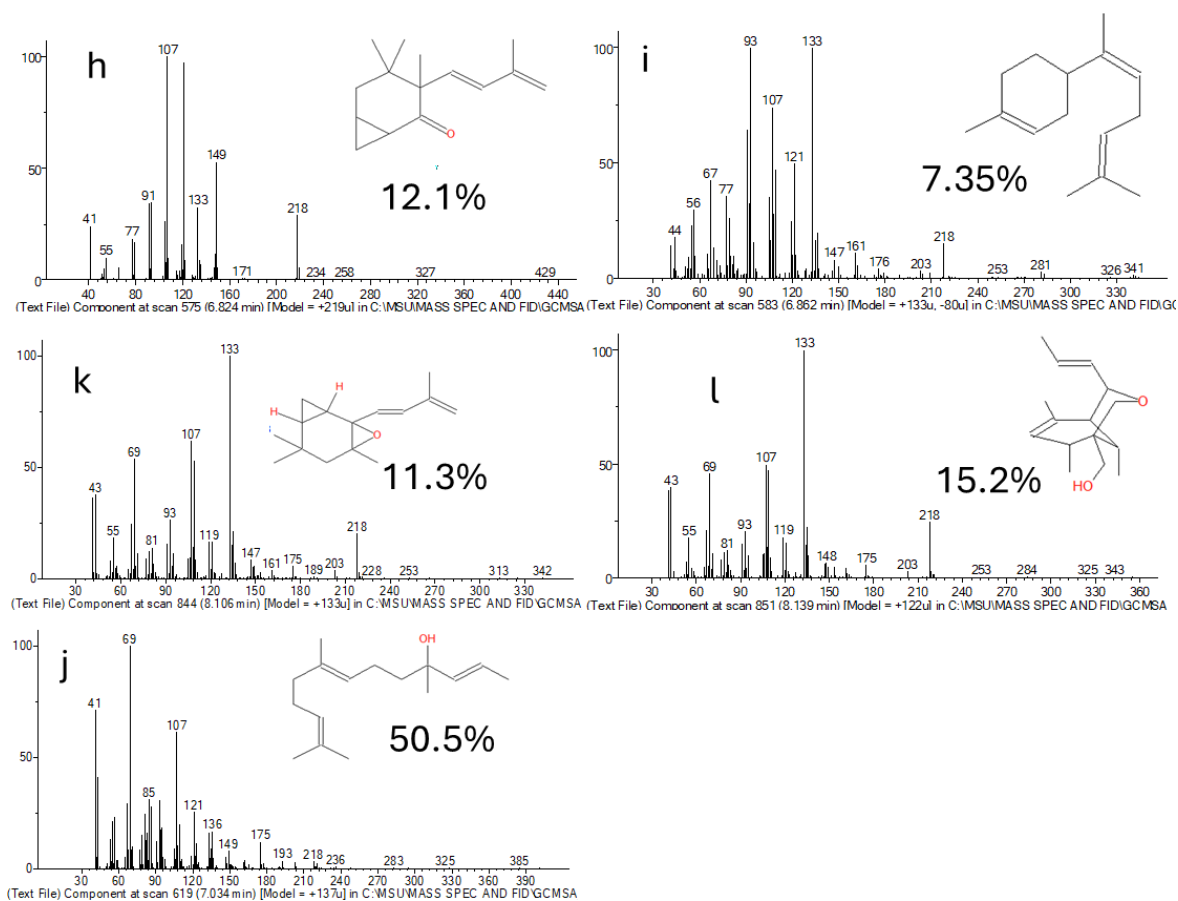


Figure S4.16. PvTPS5 + 10 chromatograms and mass spectra. a. GC-MS data showing blue as CpSTS9 +10 and red is phosphatase control. b-k is mass spectra from figure b.

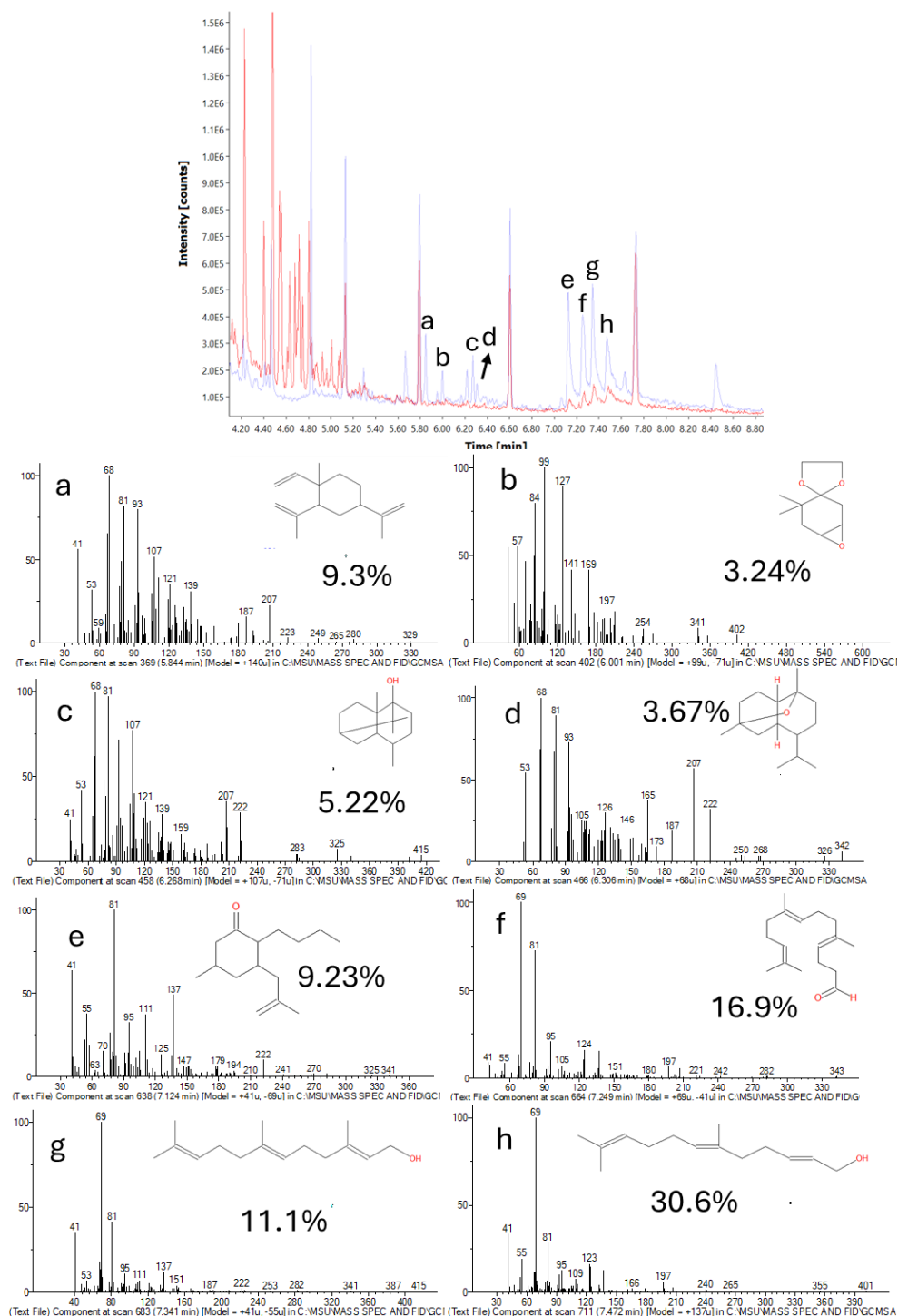


Figure S4.17. CpSTS9 + 11 chromatograms and mass spectra. a. GC-FID chromatograms. Blue is the CpSTS9 + 11, and red is phosphatase control. b. is the GC-MS chromatogram of the fraction used in figure 5. c-h is mass spectra from figure b.

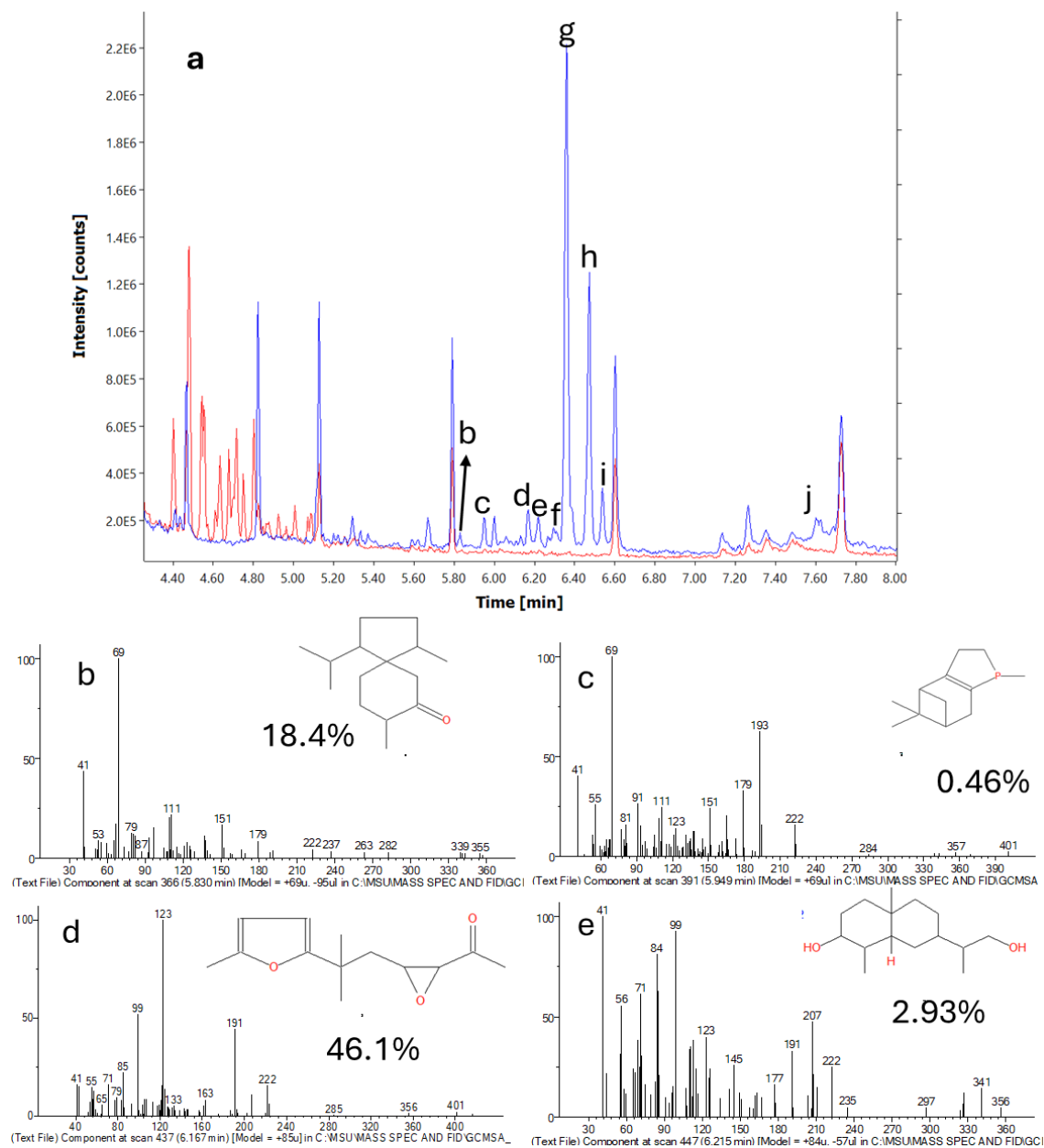


Figure S4.18. PvTPS5 + 11 chromatograms and mass spectra. a. GC-MS chromatograms. Blue is the PvTPS5 + 11 and red is phosphatase control. b. c-j is mass spectra from figure a.

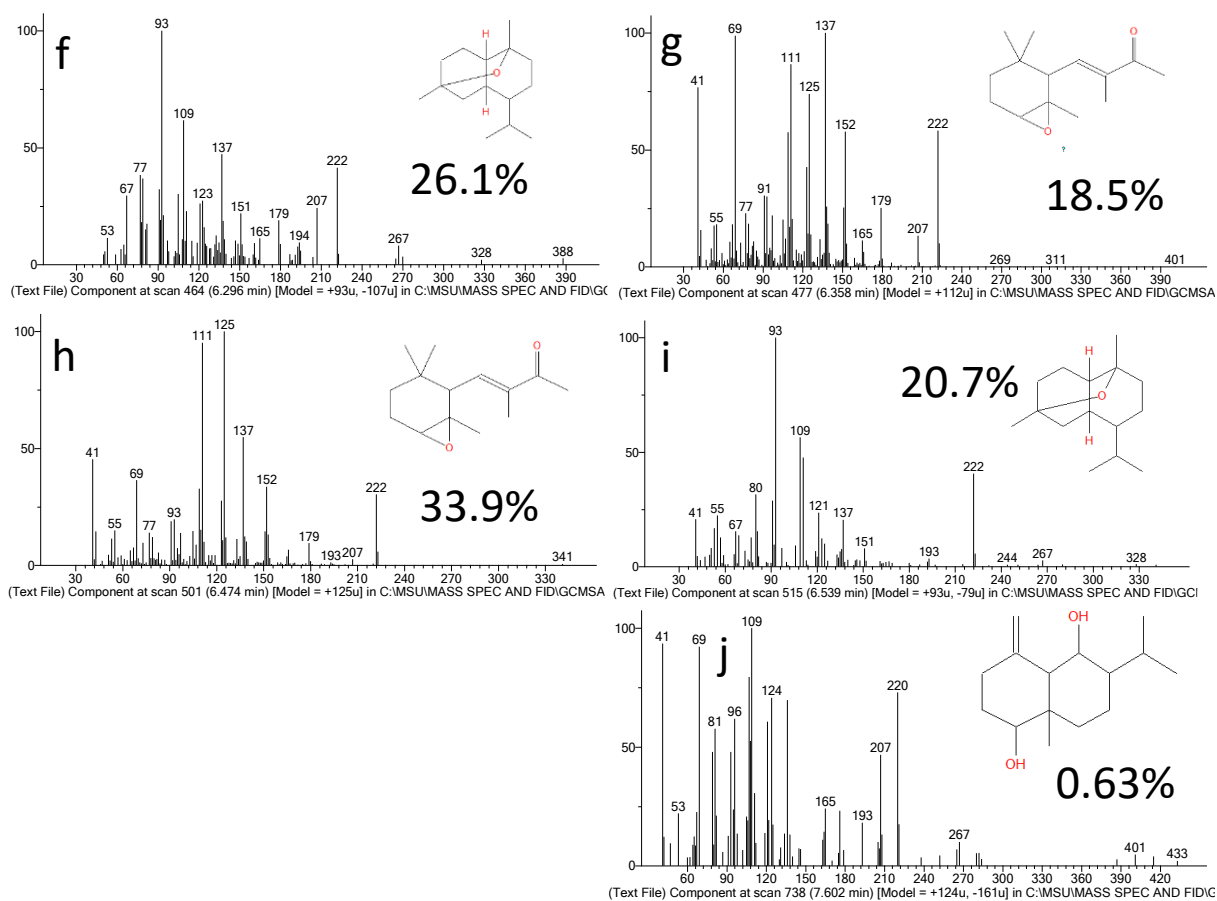


Figure S4.18. PvTPS5 + 11 chromatograms and mass spectra. a. GC-MS chromatograms. Blue is the PvTPS5 +11 and red is phosphatase control. b. c-j is mass spectra from figure a.

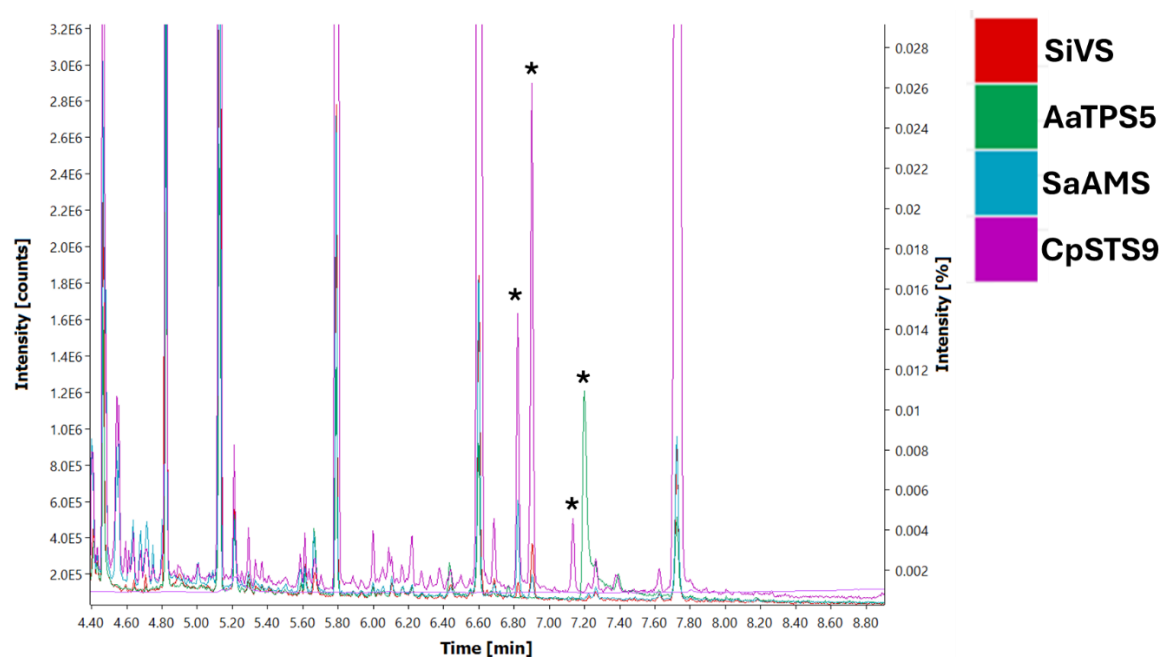


Figure S4.19. TPS-FDP analog combinations chromatograms overlaying all TPS combinations that worked with substrate 7. Asterisks labeled over peaks indicate terpene derivatives.

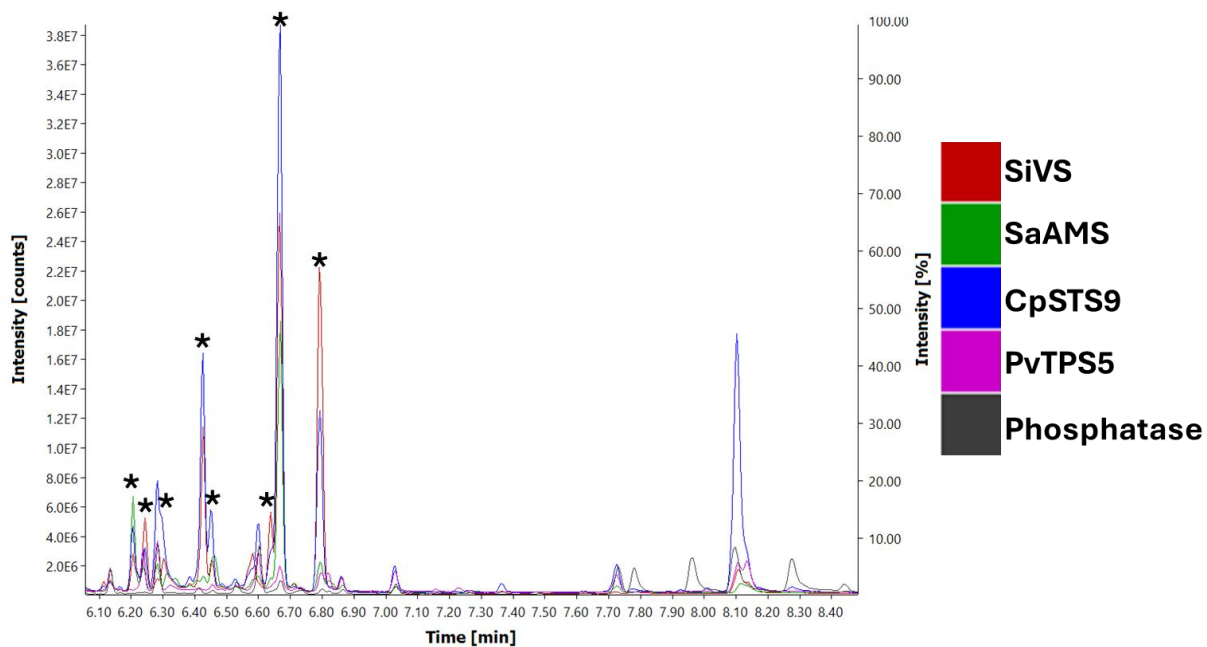
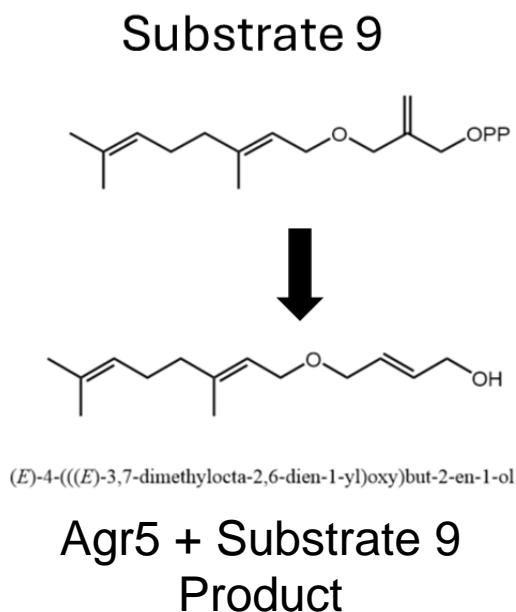


Figure S4.20. TPS—FDP analog combinations chromatograms overlaying all TPS combinations that worked with substrate 10. Asterisks labeled over peaks indicate terpene derivatives.



Atom	δ (ppm)	Min..Max (ppm)	Quality
1 C	25.7	25.62..25.78	0.79
H3	1.7	1.70..1.71	
2 C	131.75	131.47..132.02	
3 C	123.92	123.86..123.99	0.81
H	5.11	5.08..5.15	0.7
4 C	26.36	26.29..26.44	0.82
H2	2.12	2.09..2.16	0.6
5 C	39.6	39.53..39.68	0.81
H2	2.07	2.04..2.09	0.79
6 C	140.85	140.82..140.87	
7 C	120.41	120.35..120.48	0.81
H	5.38	5.35..5.41	0.55
8 C	66.9	66.83..66.97	0.81
H2	4.03	4.01..4.05	0.54
9 O			
10 C	65.48	65.41..65.55	0.78
H2	4.07	4.06..4.08	
11 C	132.16	132.10..132.22	0.8
H	5.85	5.83..5.87	
12 C	128.75	128.69..128.82	0.79
H	5.75	5.72..5.79	
13 C	16.48	16.41..16.55	0.81
H3	1.69	1.68..1.70	
14 C	17.7	17.62..17.77	0.79
H3	1.63	1.62..1.63	
15 C	58.89	58.79..58.98	0.77
H2	4.23	4.21..4.25	0.43
16 O			
H			

Figure S4.21a. chemical shift assignment of Agr5 + 9. a. The structure for (*E*)-4-(((*E*)-3,7-dimethylocta-2,6-dien-1-yl)oxy)but-2-en-1-ol (**1**) and corresponding assignments deduced from (b) ^1H , (c) COSY, (d) HSQC, (e) HMBC, and (f) ^{13}C correlations. CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry was assigned based on the configuration of the substrate.

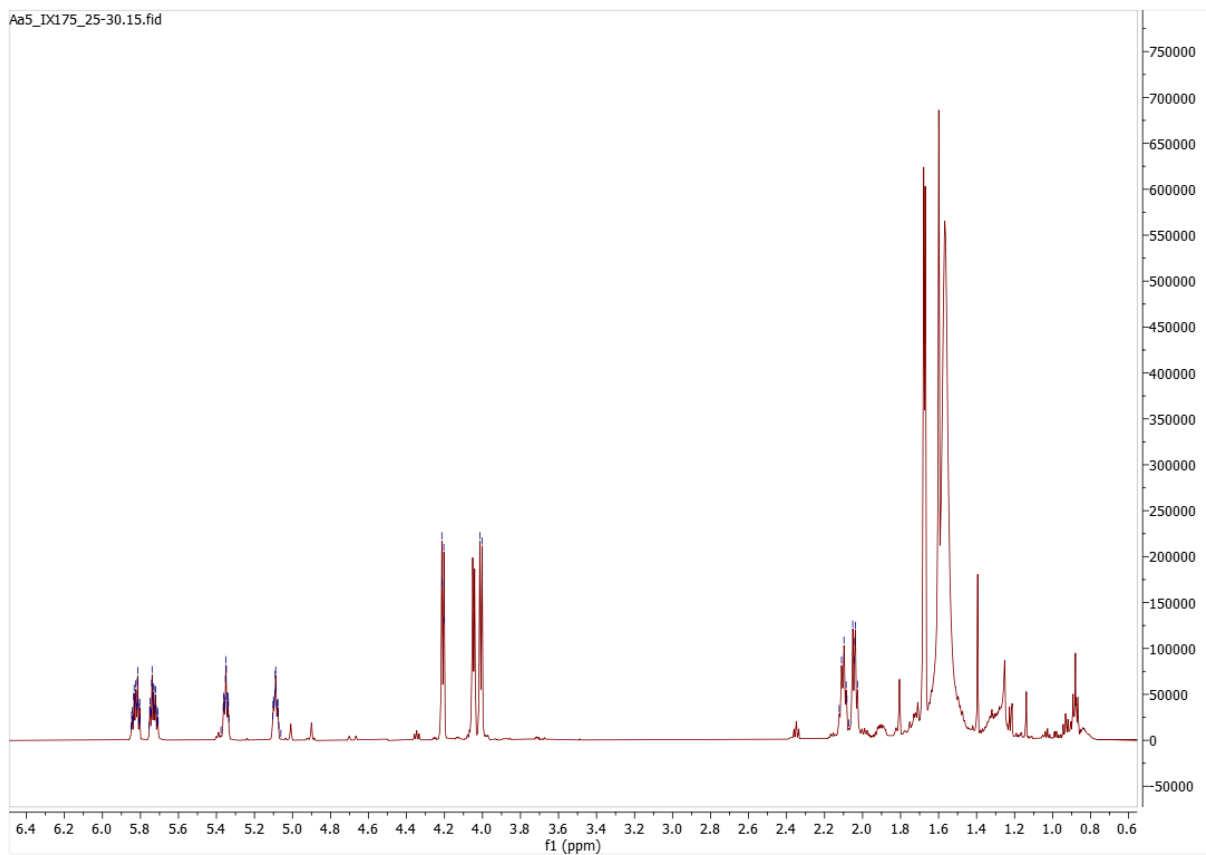


Figure S4.21b. ^1H spectrum.

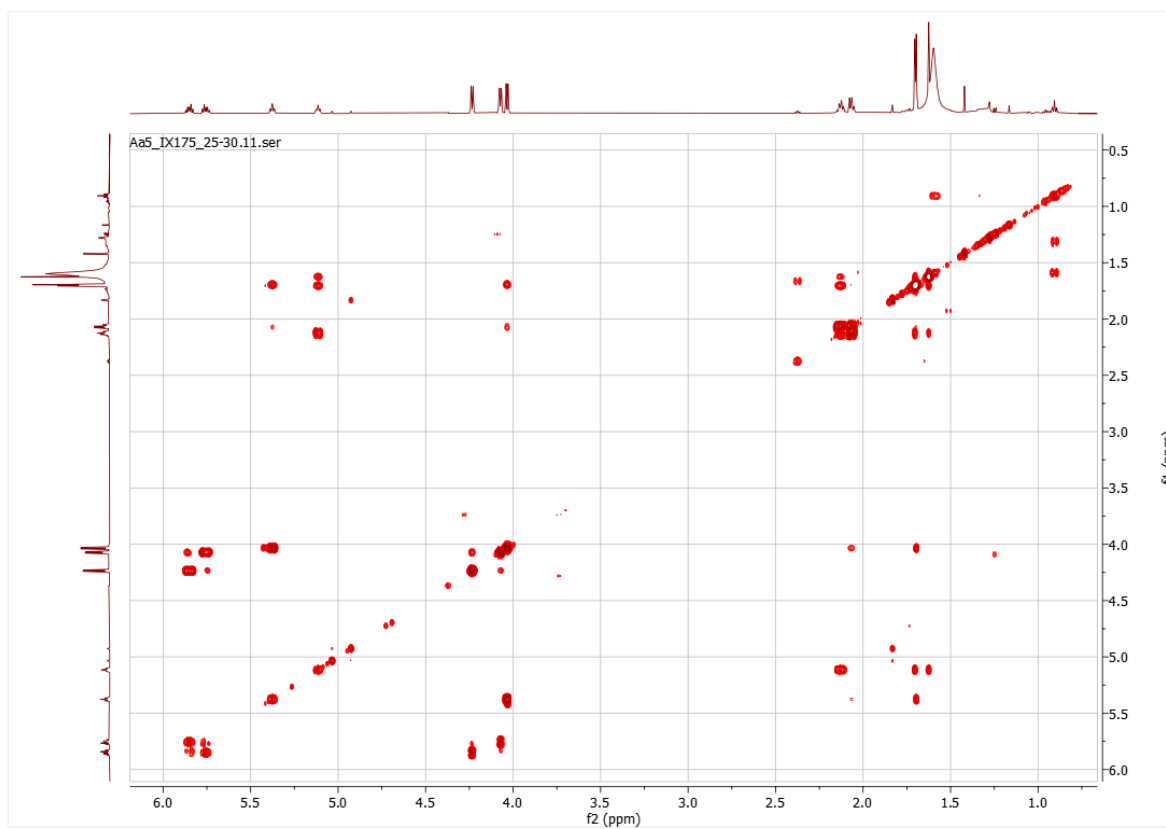


Figure S4.21c. COSY spectrum.

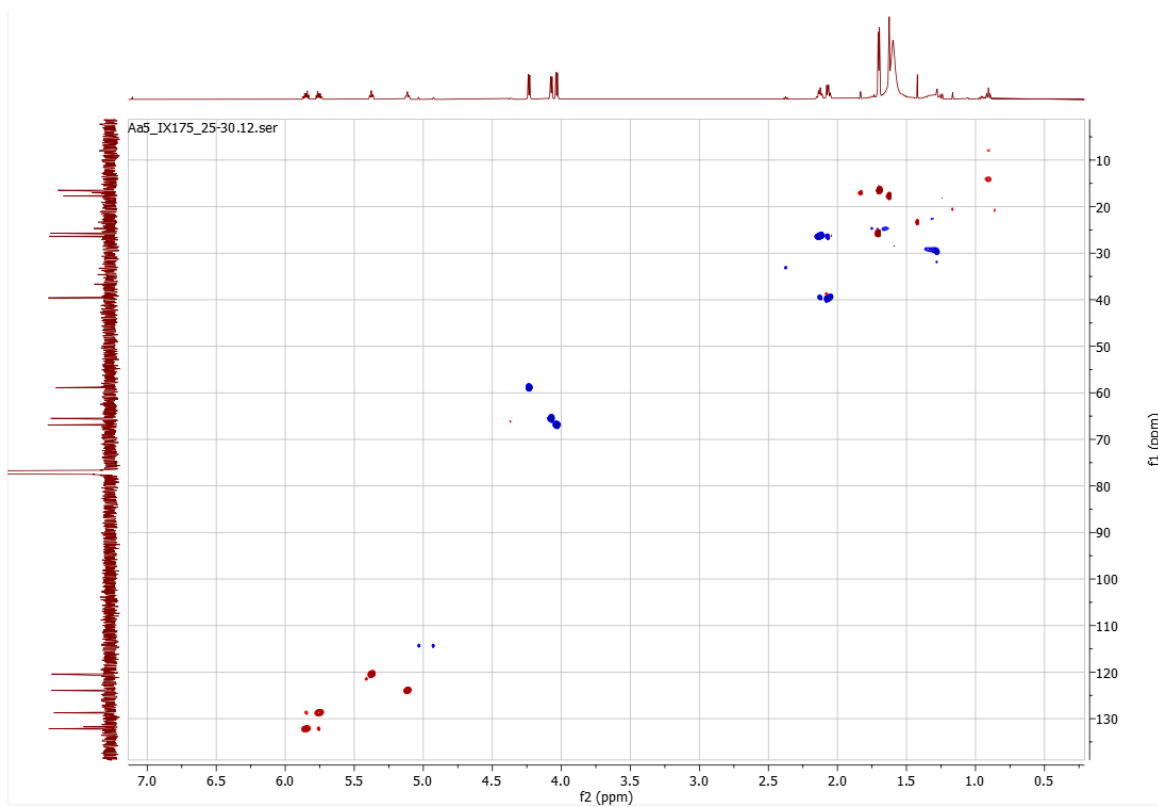


Figure S4.21d. HSQC spectrum.

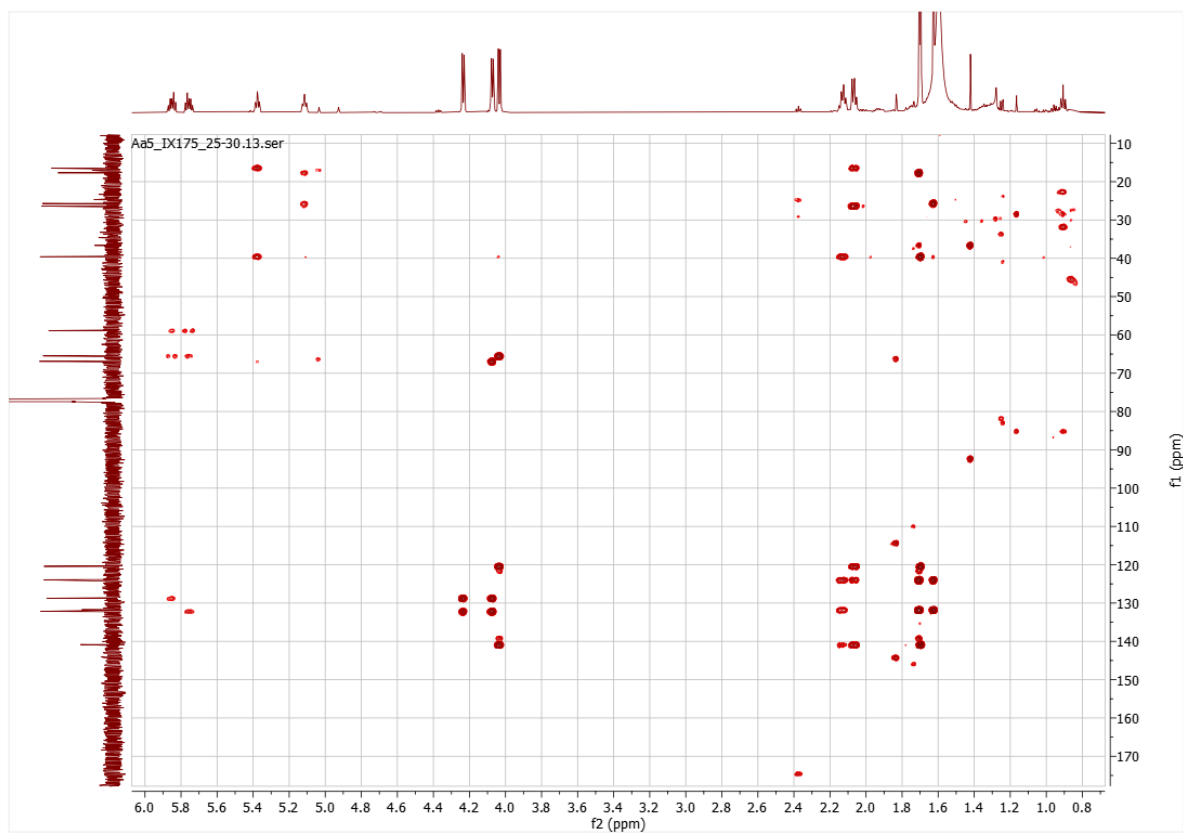


Figure S4.21e. HMBC spectrum.

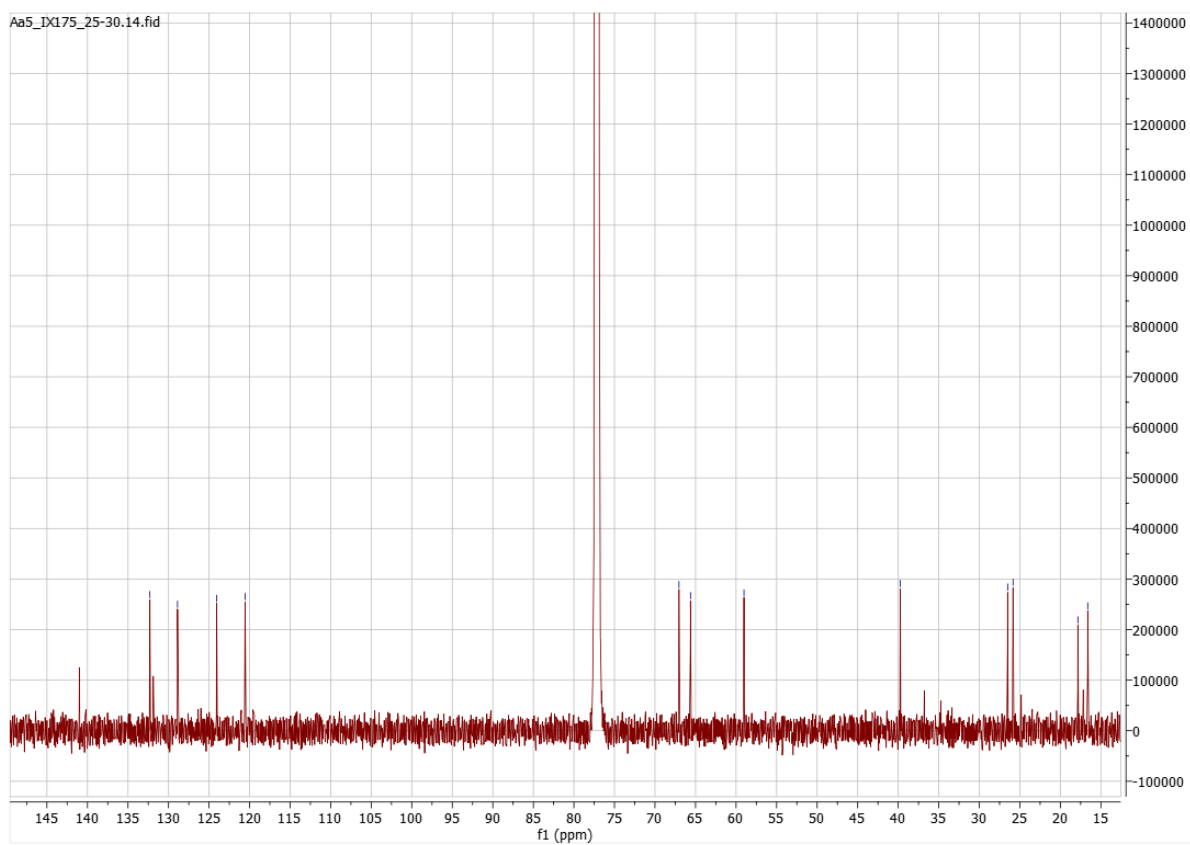


Figure S4.21f. ^{13}C spectrum.

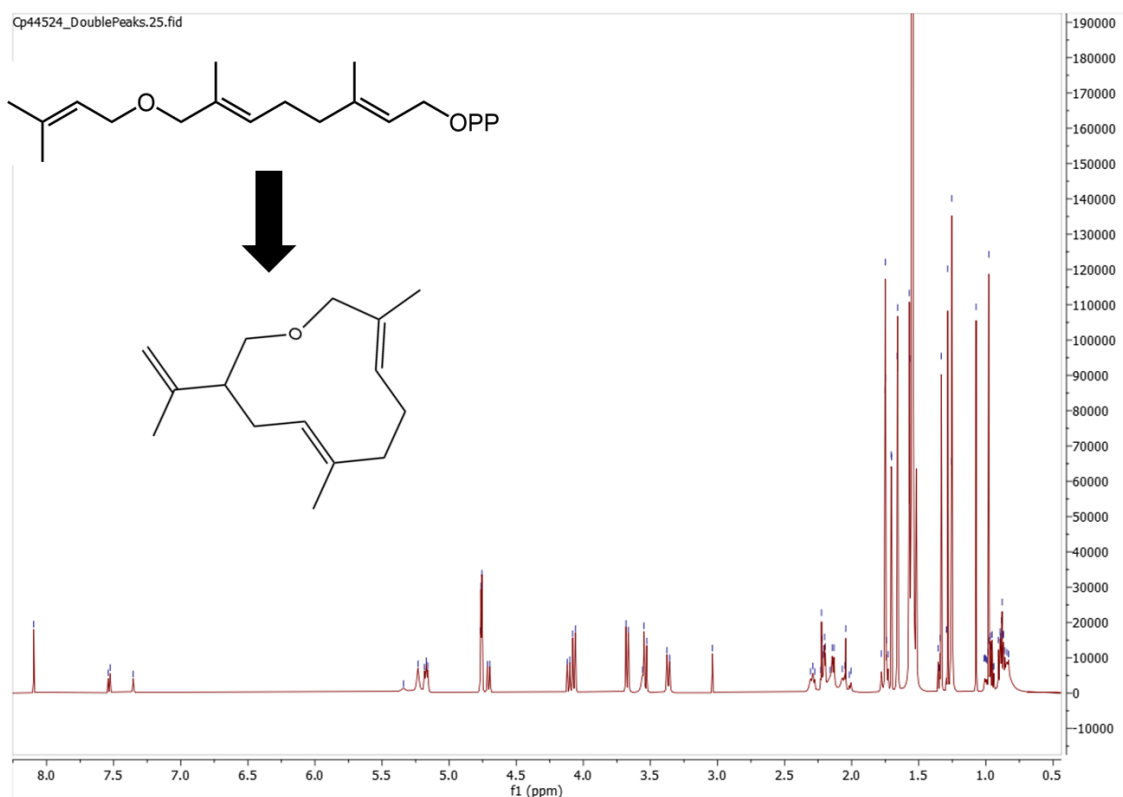


Figure S4.22a. NMR spectroscopy of CpSTS9 + 7 product. The structure for (5E,9E)-6,10-dimethyl-3-(prop-1-en-2-yl)-oxacycloundeca-5,9-diene was identified in Oberhauser et al. 2019¹⁴ and their corresponding NMR similarly corresponds to our respective (a) ¹H, (b) COSY, (c) HSQC, (d) HMBC, and (e) ¹³C NMR spectroscopy data. CDCl₃ was used as the solvent, and CDCl₃ peaks were referenced to 7.26 and 77.00 ppm for ¹H and ¹³C spectra, respectively. Absolute stereochemistry was assigned based on the configuration of the substrate.

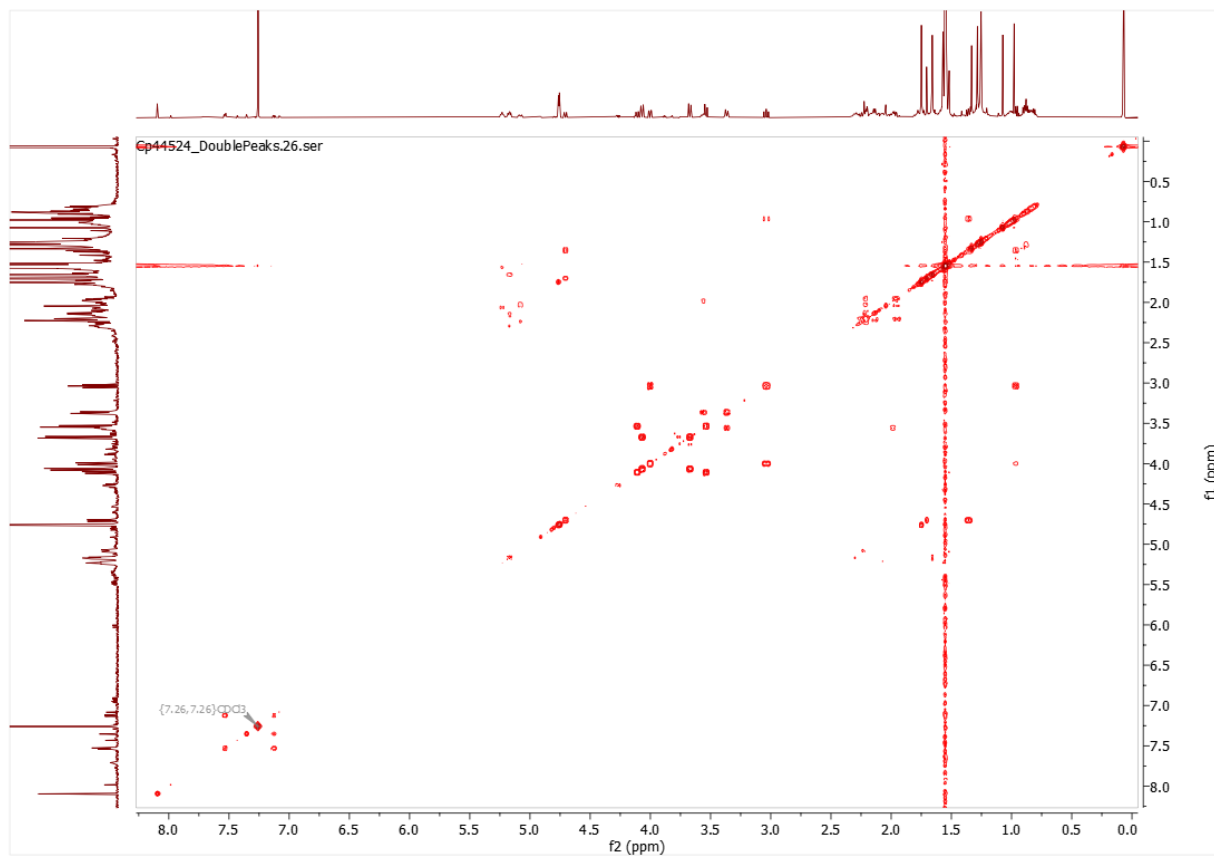


Figure S4.22b. COSY NMR spectrum.

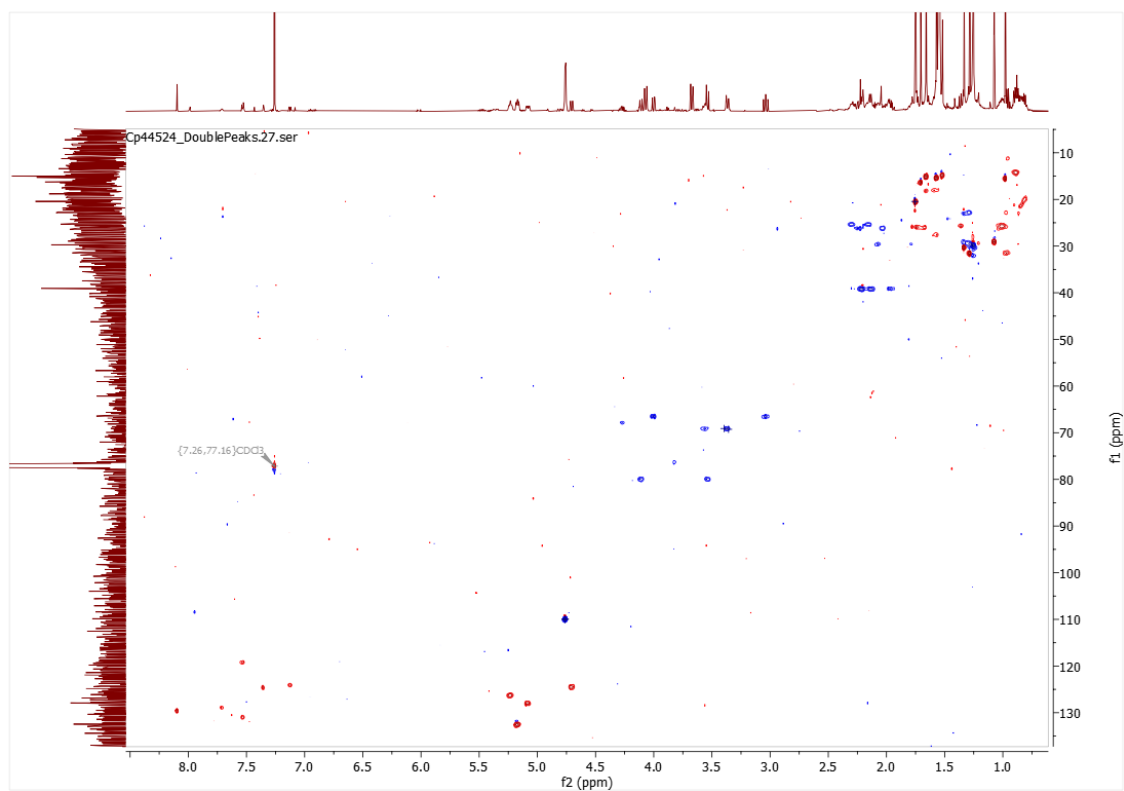


Figure S4.22c. HSQC NMR spectrum.

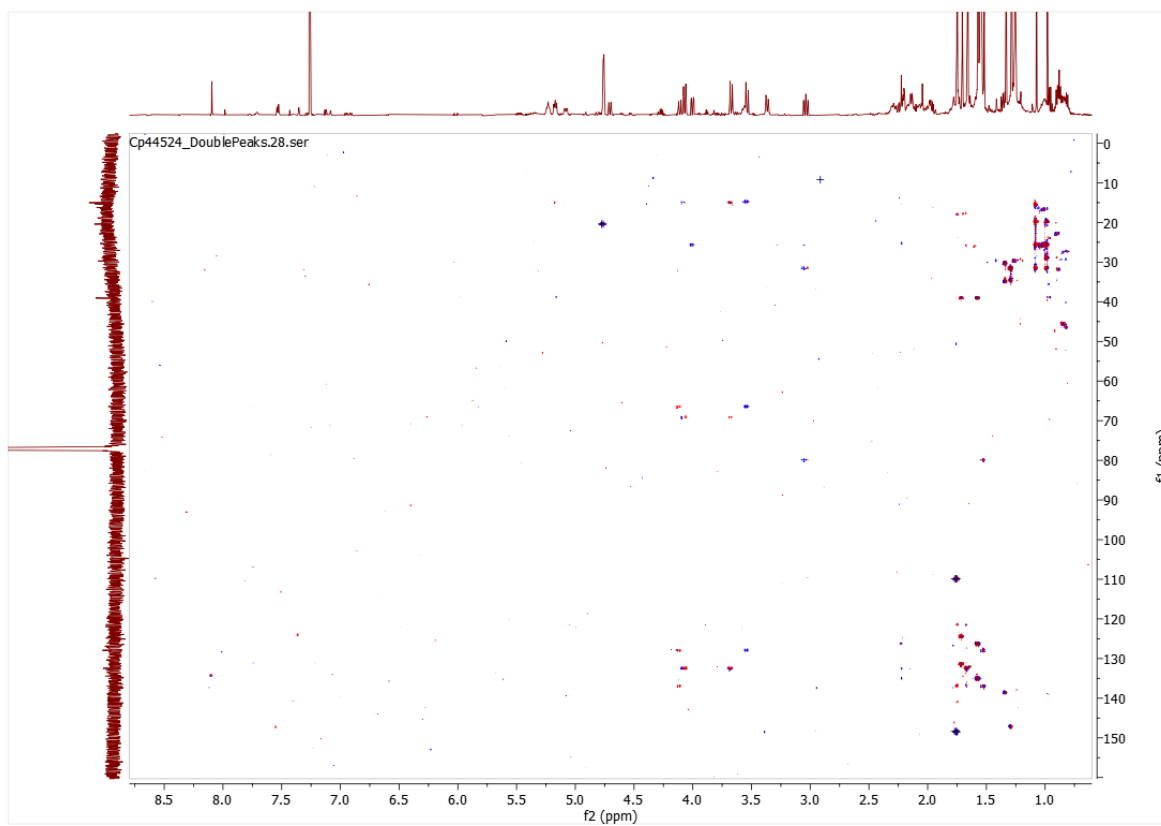


Figure S4.22d. HMBC NMR spectrum.

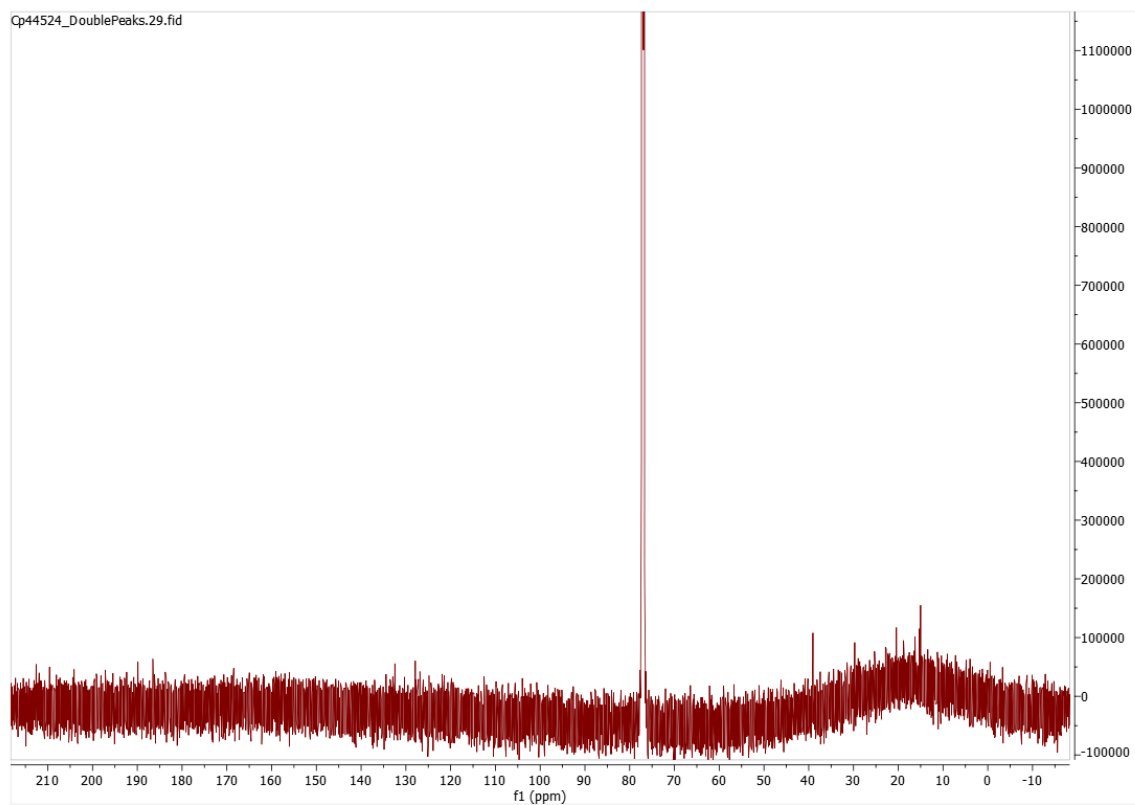


Figure S4.22e. ^{13}C NMR spectrum.

Chapter 5: Additional Research Contributions and Perspectives:

A Reconstructive and Deconstructive Approach for Unravelling the Complete Diterpene Library

In addition to the chapters presented in my thesis, one of my most substantial contributions was to a chemoinformatic project. This project was led by Dr. Davis Mathieu, a former Hamberger Lab graduate student, who formulated the project with his internship advisor, Oliver Ebenhöf from Heinrich-Heine-Universität. This project utilized the chemoinformatic tool Pickaxe to explore diterpene chemistry by applying reaction rules to generate reaction networks.¹ In the beginning we took diterpenoids from the Dictionary of Natural Products (DNP) and deconstructed them back to their diterpene backbone by sequentially removing functional groups. This process, paired with DNP metadata, enabled us to evaluate details about discrete diterpene backbones, including oxidation patterns and the phylogenetic relationship of the different backbones. My contributions involved sharing mined data from the DNP and providing expertise on terpene biochemistry for the reconstruction of diTPS pathways. We then leveraged Pickaxe to develop reaction rules describing terpene synthase carbocation chemistry. This project consumed much of my focus and time in 2023 as I delved into chemoinformatic software, natural product databases, and diTPS mechanism literature. Here I would review existing TPS mechanisms and break down the complex carbocation rearrangements into individual steps. I could translate these steps as a reaction rule where I would declare the substructure of the compound that needs to be identified and declare what reaction should occur if it sees the specific substructure. Breaking complex reactions down in this fashion makes them robust rules that may

apply to several distinct carbocation rearrangements. These rules could also be performed iteratively, which leads to a reaction network that illustrates known carbocation rearrangements and predicts latent diterpene chemistry. This manuscript is in preparation. In addition to collaborating with Heinrich-Heine-Universität, we ended up collaborating heavily with Luke Busta, a professor at University of Minnesota-Duluth, whom helped identify approaches for data analysis. This collaboration also earned us authorship on a similar project in the Busta lab analyzing the connectivity and modification within triterpenes in cuticle waxes.²

Deciphering the tetraploid genome and diterpenoid metabolism of *Teucrium chamaedrys*

I collaborated closely with fellow graduate student Dr. Abby Bryson, also in the Hamberger lab, to co-apply for and secure a \$30,000 award from the Neogen Land Grant to investigate *Teucrium chamaedrys*, a key species in the Ajugoideae mint subfamily. At the time of the award, no genomes from the Ajugoideae subfamily were available, and we anticipated that this project would substantially advance our understanding of furanoclerodane metabolism. Shortly after, Robin Buell from University of Georgia and Benjamin Lichman from University of York published a *Teucrium marum* genome to elucidate iridoid metabolism and featured RNA-seq from diverse tissue types.³ While this publication prevented our genome from being the first in the Ajugoideae subfamily, we still had the first *T. chamaedrys* genome. Thanks to the close phylogenetic relationship between the two species we were able to expand our study into comparative genomics as well as gene expression across additional tissues.

Despite the close relationship, comparisons between the *T. marum* and *T.*

chamaedrys genomes revealed that *T. marum* is diploid while *T. chamaedrys* is a tetraploid. This observation paired with additional evidence indicated that a recent whole genome duplication has occurred within the genus. Our analysis of TPSs revealed multiple orthologs of other clerodane synthases, which we functionally characterized. Most orthologs were isokolavenyl diphosphate synthases and the others no longer yielded products. To our surprise, we also discovered that ~70% of diTPSs in both species are located on the same genomic locus, corresponding to the Lamiaceae-wide miltiradiene biosynthetic gene cluster (BGC).⁴ This cluster now represents the largest diTPS BGC to date and appears to have recruited other putative diTPSs. Phylogenetic analysis of the recruited putative diTPSs suggests that they are involved in other diterpenoid metabolism that ‘branches’ from a similar starting point. This manuscript is under review at *Plant Communication* and the putative diTPSs are being cloned and functionally characterized by other members of the Hamberger laboratory.

While not part of the main manuscript, I conducted coexpression analyses from the *T. marum* data, identifying a coexpression module containing orthologs of isoKPP synthase and CYP76BK1 (Chapter 2), suggesting a relationship to furanoclerodane metabolism. With the help of a summer REU student, we cloned and assayed ~75% of the candidate cytochromes P450 (CYPs), 2-oxoglutarate dioxygenase (2OGDs), and other oxidoreductases. We found three phylogenetically distinct CYPs, currently referred to as TeMaCYP3, TeMaCYP4B, and TeChCYP7, that all produce furans similar to the products of CYP76BK1 (Chapter 2). While phylogenetically unrelated, the presence of multiple enzymes performing the same function is reminiscent of a gene-dosing effect and may improve flux through the respective pathway. The same

coexpression module also contained putative sesquiterpene synthases whose products we have yet to identify. I hypothesized that the sesquiterpene produced has synergistic or complementary effects to furanoclerodanes, necessitating coexpression without being in the same biosynthetic pathway but additional analyses are necessary to fully resolve this relationship.

Chromosome-scale *Salvia hispanica* L. (Chia) genome assembly reveals rampant *Salvia* interspecies introgression

Dr. Julia Brose, a former graduate student from Robin Buell's lab located at University of Georgia, sought my expertise for work on a recently published paper evaluating the genome of *Salvia hispanica*.⁵ Their comparison of available genomes revealed an enrichment in TPSs within the Chia pinta variety. I phylogenetically placed their genes of interest relative to functionally validated TPSs. This work revealed that a bloom of TPS-a subfamily sesquiterpene synthases was found exclusively in Chia pinta. Deeper analysis revealed that these TPS genes were collectively a part of 6 separate biosynthetic gene clusters present in the Chia pinta variety and not the Chia negra variety, where Chia negra fully lacked these genes. Genomic analyses of other *Salvia* suggested that these genes were likely recruited as a result of interspecies introgressions from other *Salvia* species.

REFERENCES

- (1) Shebek, K. M.; Strutz, J.; Broadbelt, L. J.; Tyo, K. E. J. Pickaxe: A Python Library for the Prediction of Novel Metabolic Reactions. *BMC Bioinformatics* **2023**, *24* (1), 106. <https://doi.org/10.1186/s12859-023-05149-8>.
- (2) Babineau, N.; Nguyen, L. T. D.; Mathieu, D.; McCue, C.; Schlecht, N.; Abrahamson, T.; Hamberger, B.; Busta, L. A Molecular Representation System with a Common Reference Frame for Natural Products Pathway Discovery and Structural Diversity Tasks. *bioRxiv* October 1, 2024, p 2024.10.01.616173. <https://doi.org/10.1101/2024.10.01.616173>.
- (3) Smit, S. J.; Ayten, S.; Radzikowska, B. A.; Hamilton, J. P.; Langer, S.; Unsworth, W. P.; Larson, T. R.; Buell, C. R.; Lichman, B. R. The Genomic and Enzymatic Basis for Iridoid Biosynthesis in Cat Thyme (*Teucrium Marum*). *The Plant Journal* **2024**, *118* (5), 1589–1602. <https://doi.org/10.1111/tpj.16698>.
- (4) Bryson, A. E.; Lanier, E. R.; Lau, K. H.; Hamilton, J. P.; Vaillancourt, B.; Mathieu, D.; Yocca, A. E.; Miller, G. P.; Edger, P. P.; Buell, C. R.; Hamberger, B. Uncovering a Miltiradiene Biosynthetic Gene Cluster in the Lamiaceae Reveals a Dynamic Evolutionary Trajectory. *Nat Commun* **2023**, *14* (1), 343. <https://doi.org/10.1038/s41467-023-35845-1>.
- (5) Brose, J.; Hamilton, J. P.; Schlecht, N.; Zhao, D.; Mejía-Ponce, P. M.; Cruz-Pérez, A.; Vaillancourt, B.; Wood, J. C.; Edger, P. P.; Montes-Hernandez, S.; de Rosas, G. O.; Hamberger, B.; Cibrian-Jaramillo, A.; Buell, C. R. Chromosome-Scale *Salvia Hispanica* L. (Chia) Genome Assembly Reveals Rampant *Salvia* Interspecies Introgression. *The Plant Genome* **2024**, *17* (3), e20494. <https://doi.org/10.1002/tpg2.20494>.

Future Directions:

Ajuga reptans has proven a valuable source for identifying CYPs capable of modifying clerodanes, revealing an important step in the production of furanoclerodanes with ArCYP76BK1 and the promiscuous ArCYP736A358. The direct products of these enzymes have not been functionally characterized before. Knowing many furanoclerodanes are insect anti-feedants and are cytotoxic, we have partnered with the chemical company BASF to assay the insect antifeedant properties of the different CYP76BK1 products and the Michigan State University professor Jamie Bernard's group to assay the cytotoxicity of our compounds against multiple myeloma cancer cell lines. Hopefully a similar characterization can be done with the ArCYP736A358 products as well. Although our investigation into *A. reptans* has been fruitful, we were limited to a single transcriptome, limiting the available bioinformatic analyses. Assembling a chromosome-scale genome or a comprehensive set of transcriptomes could facilitate future studies.

Similarly, we could continue to use the coexpression data from the *T. marum* analysis to clone the remaining candidates that have not been investigated yet. Given that we already have functioning sesquiterpene synthases from *Teucrium*, verifying these TPS products could be relatively straightforward. We also identified phylogenetically distinct CYPs that all produce the furan moiety on isokolavenol. This redundancy appearing from unrelated enzymes in the same species was a surprise. There are a few possible explanations for their presence. They may be differentially regulated, their native activity is promiscuous, and furan formation is relatively easy, or they all independently evolved to improve metabolic flux towards furanoclerodanes that

are abundant in *Teucrium*. Future work to characterize putative diTPSs found on the largest BGC is necessary to understand the ramification of a highly conserved BGC. Given the recruitment of phylogenetically distinct TPSs, we hypothesize that this BGC is possibly itself a network, where branching class I diTPSs can build scaffolds off the same class II diTPS products.

I have not elucidated all of ArCYP736A358's products identified in Chapter 3, however many of the remaining structures are being purified for NMR. Beyond the publication, the epoxy-isokolavelool may also be cytotoxic against multiple myeloma, as isokolavelool itself has been shown to be by Jamie Bernard's research group. Additionally, epoxides tend to be bioactive. It is valuable to test the limits of ArCYP736A358 (Chapter 3), by assaying it against some other natural products with terminal alkenes.

In Chapter 4, I showed the potential for feeding semi-synthetic substrates to TPSs, but scaling up, purification, and NMR spectroscopy was a larger hurdle in the project than we anticipated, culminating in my inability to do NMR analyses on most structures and hindering the initially planned anti-fungal assays. To address the scale-up challenges, synthesizing new products would enable different optimization approaches, including optimizing the substrate solvents to reduce protein denaturation, identifying if batch additions of enzyme throughout the assays improves yields, and optimizing extraction techniques (nonpolar adsorbents vs various nonpolar solvents) and concentration techniques. Additionally, identifying a better method to remove and replace the solvent with CDCl_3 without product loss would enhance our ability to do NMR spectroscopy. With extra material, we could conduct additional anti-fungal assays

on our products for better statistical power and could be assayed against a broader range of phytopathogens. Based on the substrates that were functional, expanding our analogs to include halogens, other methylation sites, and thiol linkages would be a worthwhile continuation. Lastly, structure activity relationships of the unnatural substrates and TPSs could be explored using methods like homology modeling, molecular docking, and mutational studies.