

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Investigations into the biosynthesis of terpenoids from marine sponges

Permalink

<https://escholarship.org/uc/item/2088w5b4>

Author

Wilson, Kayla Ann

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Investigations into the biosynthesis of terpenoids from marine sponges

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Marine Biology

by

Kayla Wilson

Committee in charge:

Professor Bradley S. Moore, Chair
Professor Eric E. Allen
Professor Lihini I. Aluwihare
Professor Pieter C. Dorrestein
Professor William H. Gerwick

2024

Copyright

Kayla Wilson, 2024

All rights reserved.

The Dissertation of Kayla Wilson is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

*To all the teachers and mentors who saw my passion for science
and fiercely protected it. To my parents for always believing in me.
To the ocean for reminding me who I am. To Oliver for being
all I could ask for and more. And finally, to Captain,
for being by my side – always and forever.*

TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE	iii
DEDICATION	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	viii
LIST OF SCHEMES	xiii
LIST OF TABLES	xiv
ACKNOWLEDGEMENTS	xv
VITA.....	xvii
ABSTRACT OF THE DISSERTATION	xviii
CHAPTER 1. Terpenes across the tree of life	1
1.1 Introduction to sponge natural products	1
1.2 Isoprenoid biosynthesis.....	4
1.3 Marine nitrogenous terpenes.....	8
1.4 Introduction to “Specialized terpenoid biosynthesis across domains of life”.....	13
1.5 References for Sections 1.1-1.4	14
1.6 “Specialized terpenoid biosynthesis across domains of life	18
1.6.1 Abstract	18
1.6.2 Introduction	18
1.6.3 Basic biosynthetic principles.....	21
1.6.4 Terpenoid distribution and genome mining	24
1.6.5 Plants	26
1.6.6 Fungi.....	33
1.6.7 Bacteria.....	38

1.6.8	Algae	41
1.6.9	Insects.....	47
1.6.10	Amoeba	51
1.6.11	Sponges	52
1.6.12	Discussion	54
1.6.13	References for Section 1.6	57
1.7	Acknowledgements	69
CHAPTER 2. Structural elucidation of sponge isonitrile terpenes using a chlorooxime probe.....		70
2.1.	Introduction	70
2.2.	References for section 2.1	75
2.3.	Reprint of “Identification of Isonitrile-Containing Natural Products in Complex Biological Matrices Through Ligation with Chlorooximes”	77
2.4.	Acknowledgements.....	119
CHAPTER 3. Investigating the biosynthesis of sponge nitrogenous terpenes		120
3.1.	Enzyme Discovery	120
3.2.	Biocatalysis	123
3.3.	Terpene synthases in biocatalysis	127
3.4.	Introduction to “Terpene biosynthesis in marine sponge animals”	129
3.5.	Reprint of “Terpene biosynthesis in marine sponge animals”	133
3.6.	Taxonomic identification of San Diego sponges	227
3.7.	Synthesis of terpenoid precursors	250
3.8.	Cell free extract experiments	256
3.9.	Sponge terpene synthase point mutations	266

3.10. Sponge cyanide detection experiments	272
3.11. Cloning of the hcnABC cluster	281
3.12. References	284
3.13. Acknowledgements	290
DISCUSSION AND CONCLUSION	291
Discussion and conclusion references	294

LIST OF FIGURES

Figure 1.1.1. Body plan and internal structure of porifera.....	3
Figure 1.3.1. Common marine nitrogenous terpene scaffolds	8
Figure 1.6.1. Structural overview of TS enzymes	23
Figure 1.6.2. Class 1 terpene counts	24
Figure 1.6.3. Phylogeny of class 1 terpene synthases.....	26
Figure 1.6.4. Function and architecture of plant TS enzymes	28
Figure 1.6.5. Diterpene synthase gene clusters.....	29
Figure 1.6.6. Sesqui- and monoterpenoids.....	31
Figure 1.6.7. <i>Nicotiana suaveolens</i> monoterpenes	33
Figure 1.6.8. Fungal TS scaffold generation.....	36
Figure 1.6.9. Cyclooctatin biosynthesis.....	38
Figure 1.6.10. Geosmin and 2-methylisoborneol biosynthesis.....	40
Figure 1.6.11. Algal TS scaffold generation.....	43
Figure 1.6.12. Crystal structure of DabA.....	44
Figure 1.6.13. Cladogram of red algae	46
Figure 1.6.14. Insect terpenoid scaffolds	49
Figure 1.6.15. Discodiene biosynthesis	51
Figure 1.6.16. Sponge terpenoids	53
Figure 2.3.1. Natural isonitriles, the isonitrile-chlorooxime ligation, and reactivity-based screening	78
Figure 2.3.2. Reactivity-based screening for <i>F. ambigua</i> natural products.....	79
Figure 2.3.3. Reactivity-based screening for Bubarida natural products.....	80
Figure 2.3.S1. Ambiguine A isonitrile and conjugate	89

Figure 2.3.S2. Ambiguine C isonitrile and conjugate.....	90
Figure 2.3.S3. Ambiguine D or O isonitrile and conjugate	91
Figure 2.3.S4. Ambiguine E isonitrile and conjugate	92
Figure 2.3.S5. Ambiguine H isonitrile and conjugate	93
Figure 2.3.S6. Ambiguine I isonitrile and conjugate	94
Figure 2.3.S7. Hapalindole H isonitrile and conjugate	95
Figure 2.3.S8. Ambiguine K isonitrile and conjugate	96
Figure 2.3.S9. Ambiguine L isonitrile and conjugate	97
Figure 2.3.S10. GCMS chromatogram of Bubarida crude extract	99
Figure 2.3.S11. LCMS of chlorooxime probe 1 negative control	101
Figure 2.3.S12. MS detectability of isonitrile terpenoids and conjugates	102
Figure 2.3.S13. NMR analysis of sponge-derived sesquiterpene conjugate 2	103
Figure 2.3.S14. NMR analysis of sponge-derived sesquiterpene conjugate 4	104
Figure 2.3.S15. Competition experiment LC-UV chromatograms.....	106
Figure 2.3.S16. Competition experiment extracted ion chromatograms	106
Figure 2.3.S17. MS/MS analysis of conjugated sponge natural products	108
Figure 2.3.S18. MS/MS analysis of probe conjugate S3	109
Figure 3.2.1. Directed evolution	124
Figure 3.2.2. pharmaceutical molecules produced using biocatalysis	127
Figure 3.3.1. Terpenoid biocatalytic production targets	128
Figure 3.5.1. Sponge terpenoids and Bubarida sponge morphology	135
Figure 3.5.2. Eukaryotic features of terpene synthase-containing contigs	137
Figure 3.5.3. Type I terpene synthase phylogeny	138

Figure 3.5.4. Active site residue alignment	138
Figure 3.5.5. Proposed nitrogenous terpene biosynthesis mechanisms	139
Figure 3.5.S1. GCMS traces of sponge crude extract.....	161
Figure 3.5.S2. Mass spectra of sponge crude extract.....	162
Figure 3.5.S3. Undescribed Bubarida morphology	163
Figure 3.5.S4. BlobTools snail plot	164
Figure 3.5.S5. GCMS trace of IDS assays: FPPS1	165
Figure 3.5.S6 GCMS trace of IDS assays: FPPS2.....	165
Figure 3.5.S6. GCMS trace of IDS assays: GGPPS	166
Figure 3.5.S8. GCMS trace of IDS assays: negative control 1 FPPS.....	166
Figure 3.5.S9 GCMS trace of IDS assays: negative control 2 FPPS.....	167
Figure 3.5.S10. GCMS trace of IDS assays: negative control 1 GGPPS	167
Figure 3.5.S11. GCMS trace of IDS assays: negative control 2 GGPPS	168
Figure 3.5.S12. Structure homology models	169
Figure 3.5.S13. uBuTS-10 transcripts.....	170
Figure 3.5.S14. StTS-1 sequence alignment.....	170
Figure 3.5.S15. GCMS traces of FPP TS assays: uBuTS-1	171
Figure 3.5.S16. GCMS traces of FPP TS assays: uBuTS-2	172
Figure 3.5.S17. GCMS traces of FPP TS assays: uBuTS-3	173
Figure 3.5.S18. GCMS traces of FPP TS assays: uBuTS-4	174
Figure 3.5.S19. GCMS traces of FPP TS assays: uBuTS-5	175
Figure 3.5.S20. GCMS traces of FPP TS assays: PhTS-1	176
Figure 3.5.S21. GCMS traces of FPP TS assays: AgTS-1	177

Figure 3.5.S22. GCMS traces of GPP and GGPP TS assays: negative control.....	178
Figure 3.5.S23. GCMS traces of GPP and GGPP TS assays: uBuTS-1	179
Figure 3.5.S24. GCMS traces of GPP and GGPP TS assays: uBuTS-2.....	180
Figure 3.5.S25. GCMS traces of GPP and GGPP TS assays: uBuTS-3.....	181
Figure 3.5.S26. GCMS traces of GPP and GGPP TS assays: uBuTS-4.....	182
Figure 3.5.S27. GCMS traces of GPP and GGPP TS assays: uBuTS-5.....	183
Figure 3.5.S28. GCMS traces of GPP and GGPP TS assays: uBuTS-6.....	184
Figure 3.5.S29. GCMS traces of GPP and GGPP TS assays: uBuTS-7.....	185
Figure 3.5.S30. GCMS traces of GPP and GGPP TS assays: uBuTS-8.....	186
Figure 3.5.S31. GCMS traces of GPP and GGPP TS assays: uBuTS-9.....	187
Figure 3.5.S32. GCMS traces of GPP and GGPP TS assays: AgTS-1.....	188
Figure 3.5.S33. GCMS traces of GPP and GGPP TS assays: AgTS-2.....	189
Figure 3.5.S34. GCMS traces of GPP and GGPP TS assays: PhTS-1	190
Figure 3.5.S35. GCMS traces of GPP and GGPP TS assays: StTS-1	191
Figure 3.5.S36. GCMS traces of FPP + KCN assays: negative control	192
Figure 3.5.S37. GCMS traces of FPP + KCN assays: uBuTS-1	193
Figure 3.5.S38. GCMS traces of FPP + KCN assays: uBuTS-2	194
Figure 3.5.S39. GCMS traces of FPP + KCN assays: uBuTS-3	195
Figure 3.5.S40. GCMS traces of FPP + KCN assays: uBuTS-4	196
Figure 3.5.S41. GCMS traces of FPP + KCN assays: uBuTS-5	197
Figure 3.5.S42. GCMS traces of FPP + KCN assays: uBuTS-8	198
Figure 3.5.S43. GCMS traces of FPP + KCN assays: uBuTS-9	199
Figure 3.5.S44. GCMS traces of FPP + KCN assays: PhTS-1	200

Figure 3.5.S45. GCMS traces of FPP + KCN assays: AgTS-1	201
Figure 3.5.S46. GCMS traces of FPP + KCN assays: StTS-1	202
Figure 3.5.S47. GCMS traces of FPP + KCN assays: AgTS-2	203
Figure 3.5.S48. GCMS traces of FPP + KCN assays: uBuTS-1 to uBuTS-9.....	204
Figure 3.5.S49. Sponge cyanide detection assays.....	205
Figure 3.5.S50. LCMS analyses of undescribed Bubarida extract	206
Figure 3.5.S51. SDS-PAGE gel of TSs	207
Figure 3.6.1. SCS0001-SCS0012 photos.....	230
Figure 3.6.2. SCS0013-SCS0024 photos.....	231
Figure 3.8.1. GCMS extracted ion chromatograms for CFE experiment B part 1	261
Figure 3.8.2. GCMS extracted ion chromatograms for CFE experiment B part 2	262
Figure 3.9.1. Robetta structure models of sponge TSs	267
Figure 3.9.2. GCMS traces of mutant TSs.....	271
Figure 3.10.1. GCMS traces of DMTMM cyanide standard curve	276
Figure 3.10.2. GCMS traces of DMTMM cyanide detection in seawater.....	277
Figure 3.10.3. GCMS traces and mass spectra for DMTMM cyanide detection in seawater	278
Figure 3.10.4. GCMS traces of DMTMM cyanide detection in sponge tissue	279
Figure 3.10.5. GCMS traces of DMTMM cyanide detection for time series experiment.....	282

LIST OF SCHEMES

Scheme 1.2.1. Isoprenoid biosynthesis pathway	4
Scheme 1.2.2. The mevalonate (MEV) pathway	6
Scheme 1.2.3. The methylerythritol 4-phosphate (MEP) pathway.....	7
Scheme 1.3.1. Interconnectedness of nitrogenous terpenes and their precursors.....	11
Scheme 1.6.1.1. Isoprenoid biosynthesis	20
Scheme 2.1.1. Isonitrile-tetrazine ligation	72
Scheme 2.1.2. Isonitrile-chlorooxime ligation.....	73
Scheme 2.3.S1. Synthesis of chlorooxime probe.....	85
Scheme 2.3.S2. Competition experiments with probe 1	105
Scheme 2.3.S3. Proposed fragmentation mechanism for S3	109
Scheme 3.10.1. DMTMM cyanide coupling reaction.....	273
Scheme 3.11.1. Formation of hydrogen cyanide from glycine by HcnABC	282

LIST OF TABLES

Table 1.3.1. Summary of nitrogenous terpene isotope labeling experiments.....	10
Table 3.5.S1. NMR data for sesquiterpenoid-conjugate 2	103
Table 3.5.S2. NMR data for sesquiterpenoid-conjugate 4	104
Table 3.5.S1. Natural product biosynthetic gene clusters from the sponge microbiome	150
Table 3.5.S2. Sequencing data summary	150
Table 3.5.S3. TSHMM1 and terpene synthase sequences	151
Table 3.5.S4. Terpene synthase summary	156
Table 3.5.S5. Terpene synthase gene neighborhood annotations	157
Table 3.5.S6. Accession numbers	158
Table 3.5.S7. Compound summary.....	160
Table 3.6.1. 28S DNA barcoding results	232
Table 3.6.2. CO1 DNA barcoding results.....	239
Table 3.8.1. Cell free extract experiment A sample composition.....	259
Table 3.8.2. Cell free extract experiment B sample composition.....	260
Table 3.8.3. Cell free extract experiment C sample composition.....	263
Table 3.8.4. Cell free extract experiment D sample composition.....	265
Table 3.9.1. Mutagenesis oligonucleotide primers	269
Table 3.10.1. DMTMM detection assay sample composition	275
Table 3.10.2. DMTMM detection time series sample composition	280

ACKNOWLEDGEMENTS

I would like to thank my advisor, Bradley S. Moore, for his support, guidance, and encouragement during my graduate career. He has curated a collaborative, curiosity-driven environment and has assembled a team of brilliant minds who have helped me grow as scientist and a person. Thank you for being the ultimate example of a life filled with learning, compassion, and stoke. I would also like to thank Tristan de Rond, for instilling in me a love for terpenes and for believing in this project from day one. I would also like to acknowledge all the Moore lab members who have been part of this journey with me and have blown me away with their brilliance and generosity, time and time again. I would like to thank my collaborators at Scripps and elsewhere for being welcoming and supportive of my academic endeavors. I would also like to acknowledge the Scripps community for showing me a million ways to live a meaningful, awe-inspiring life. Thank you to Christian McDonald, Richard Walsh, Phil Zerofski, Ashleigh Palinkas, Melissa Torres, and everyone else who helped me get underwater as much as possible. I would also like to thank my undergraduate advisor Larryn Peterson, as well as all the other teachers and mentors in my life, whose guidance led me here today. I also extend my gratitude and appreciation to Mary J. Garson, Paul J. Scheuer, and John D. Faulkner, on whose academic shoulders I humbly stand.

I would like to thank my family for their endless encouragement and for always nurturing my love for science. To my mother, thank you for helping me discover my passion for research and for always reminding me to be myself. To my father, thank you for being a consistent, stable source of love and guidance in my life. I would also like to thank my friends, both near and far, who have filled my graduate school experience with laughter, adventure, and community. I would

also like to thank Oliver, for being the most generous, supportive, and compassionate partner. Your love has seen me through the brightest days and the darkest nights, and I can't express how grateful I am to have you in my life. And finally, I would like to thank Captain, whose companionship and loyalty are unmatched.

Chapter 1, in part is currently being prepared for submission for publication in *Natural Product Reports* by Taylor S. Steele, Kayla Wilson, Immo Burkhardt, and Bradley S. Moore. The dissertation author is the co-primary author of this paper with Taylor S. Steele.

Chapter 2, in part, is a reprint of the material as it appears in *Chemistry – A European Journal* Schäfer, R.J.B., Wilson, K., Biederman, M., Moore, B.S., Sieber, S., and Wennemers, H., 2022.

Chapter 3, in part, is a reprint of the material as it appears in *Proceedings of the National Academy of Sciences*, Wilson, K., de Rond, T., Burkhardt, I., Steele, T.S., Schäfer, R.J.B., Podell, S., Allen, E.E., and Moore, B.S., 2023. The dissertation author was the primary researcher and author of this paper.

VITA

- 2016 Bachelor of Science in Chemistry, Rhodes College
- 2018-2021 National Science Foundation Graduate Research Fellow
- 2018 Master of Science in Oceanography, Scripps Institution of Oceanography,
University of California San Diego
- 2024 Doctor of Philosophy in Marine Biology, Scripps Institution of Oceanography,
University of California San Diego

PUBLICATIONS

J. M. Gauglitz, C. M. Aceves, A. A. Aksenov, G. Aleti, J. Almalitie, A. Bouslimani, E. A. Brown, A. Campeau, A. M. Caraballo-Rodríguez, R. Chaar, R. R. da Silva, A. M. Demko, F. Di Ottavio, E. Elijah, M. Ernst, L. P. Ferguson, X. Holmes, A. K. Jarmusch, L. Jiang, K. B. Kang, I. Koester, B. Kwan, J. Li, Y. Li, A. V. Melnik, C. Molina-Santiago, B. Ni, A. L. Oom, M. W. Panitchpakdi, D. Petras, R. Quinn, N. Sikora, K. Spengler, B. Teke, A. Tripathi, S. Ul-Hasan, J. J.J. van der Hooft, F. Vargas, A. Vrbanac, A. Q. Vu, S. C. Wang, K. Weldon, K. Wilson, J. M. Wozniak, M. Yoon, N. Bandeira, P. C. Dorrestein, Untargeted mass spectrometry-based metabolomics approach unveils molecular changes in raw and processed foods and beverages. *Food Chem.* **302**, 125290 (2020).

R. Schäfer, K. Wilson, M. Biedermann, B. Moore, S. Sieber, H. Wennemers, Identification of Isonitrile-Containing Natural Products in Complex Biological Matrices Through Ligation with Chlorooximes. *Chem. – A Eur. J.*, **29**, e202203277 (2022).

K. Wilson, T. De Rond, Terpene biosynthesis in marine sponge animals. *Proc. Natl. Acad. Sci.* **120**, e2220934120 (2023).

ABSTRACT OF THE DISSERTATION

Investigations into the biosynthesis of terpenoids from marine sponges

by

Kayla Wilson

Doctor of Philosophy in Marine Biology

University of California San Diego, 2024

Professor Bradley S. Moore, Chair

Nature communicates using a vast array of chemical compounds known secondary metabolites or natural products. These diverse molecules mediate many critical ecological processes including predation, reproduction, and defense. For millennia, humans have harnessed the bioactivity of these natural products to improve human health. Many natural products have become or inspired ground-breaking drugs, such as the antiviral acyclovir or the antimalarial artemisinin. The instructions for making these natural products are encoded within an organism's genome. Biosynthetic genes encode the production of enzymes, which act as nature's chemists to synthesize the compounds found in nature. We have harnessed these biosynthetic enzymes to

produce a vast array of desirable molecules, including the antivirals islatravir and molnupiravir. This process is known as biocatalysis, and it is a sustainable way to perform stereo- and regio-selective reactions in mild conditions with minimal byproducts.

Terpenoids are the largest class of natural products to date, with approximately 100,000 terpenoids characterized across all domains of life. Despite the massive structural diversity seen in terpenoids, all terpenoids derive from a handful of linear polyprenyl diphosphate precursors, which are acted upon by terpene synthase enzymes. Marine sponges are well-known producers of terpenoids, with more than 4000 sponge terpenoids reported. An unusual group of terpenoids only found in sponges and nudibranchs are the nitrogenous terpenes, which feature a suite of nitrogen-containing functional groups such as isonitriles, isothiocyanates, isocyanates, formamides, and amines. For decades, the biosynthesis of these nitrogenous terpenes remained a mystery. Herein, I report my investigations into the biosynthesis of nitrogenous terpenes in marine sponges. Chapter 1 is an introduction and includes background information on sponges, marine nitrogenous terpenes, and terpene biosynthesis. Chapter 2 covers the characterization of isonitrile terpenes from a San Diego Bubarida sponge using a chemoselective chlorooxime probe. Investigations into the biosynthesis of nitrogenous terpenes from sponges are detailed in Chapter 3. I report the discovery of the first terpene synthases from sponges and present evidence that the terpene synthase genes are encoded within the sponge animal itself, not a microbial symbiont. Chapter 3 also includes investigations into the origin of the nitrogenous functional groups of nitrogenous terpenes.

CHAPTER 1 – TERPENES ACROSS THE TREE OF LIFE

1.1. Introduction to sponge natural products

Nature is home to a vast array of chemical compounds, and the study of nature's chemistry is known as the field of natural products. In nature, small molecules serve important roles in reproduction, signaling, defense against predation and disease, and many other processes that help an organism survive and thrive in its environment (1). We as humans can harness the bioactive properties of natural products, also known as secondary metabolites, to improve human health. Throughout history, ancient humans have relied on natural remedies, such as ancient Mesopotamians who used medicinal plant oils, or ancient Chinese herbalists who maintained rich historical documentation of natural medicines (2). Modern humans continue to pull chemical inspiration from nature, with more than half of all new approved drugs from 1981 to 2019 being derived from or inspired by nature in some way (3). These natural product-derived drugs have been used to treat cancer, bacterial, fungal, viral, and parasitic infections, inflammation, and more (3).

The marine environment has proven to be a rich source of natural product discovery, with more than 32,000 marine natural products reported to date (4). Ancient humans have relied on the chemistry of the marine environment for thousands of years. Red algae have been used as a traditional medicine in the British Isles, ancient Greeks and Romans used marine snails to dye royal garments, and toxins from a sea hare were even reportedly used to poison Titus, the Roman emperor from 79 to 81 CE (2, 5). In modern history, natural products from marine sources were not studied scientifically on a large scale until the advent of SCUBA (6). As dive technology increased access to the marine environment, early marine natural product chemists discovered

pioneering molecules such as sponge nucleosides in the 1950s, or marine bacterial bromopyrroles or coral prostaglandins in the 1960s (7–9). To date, 24 marine drugs have been approved to treat various cancers, hypertriglyceridemia, chronic pain, and viral infections (10).

Marine sponges have been a particularly rich source of marine natural products, with over 9000 unique sponge-derived molecules to date (4). Some of these sponge-derived molecules have gone on to become or inspire impactful medicines such as the antivirals vidarabine, cytarabine, and acyclovir or the chemotherapeutic eribulin (3). Sponges are incredibly ancient animals; there is much debate about when early ancestors of sponges first appeared, with some fossil records and steroid biomarkers suggesting they appeared roughly 600-660 million years ago, and one fossil record even suggesting sponges may have existed up to 890 million years ago (11–14).

Like other animals, sponges are multicellular and heterotrophic, but unlike most other animals they do not have any tissues, organs, or nervous, digestive, or circulatory systems. Sponges are filter feeders and have pores called ostia, where water enters the sponge and oscula, where water exits the sponge (Fig 1.1.1). The inner feeding chamber of a sponge is lined with choanocytes, flagellar cells that pump water into and out of the sponge and capture food particles such as bacteria, phytoplankton, and other organic matter, which are later ingested via phagocytosis by archaeocyte cells in the mesohyl (15, 16). Interestingly, the mesohyl is also home to a complex community of microorganisms such as bacteria, fungi, and archaea, which are not digested but persist as apparent symbionts. The exact mechanisms underlying the selection and cultivation of this microbiome is unknown, though studies indicate the sponge does selectively differentiate food from symbionts (17, 18). Sponges acquire their microbiomes through a combination of vertical transmission (the transfer of microbial symbionts from a sponge parent to

its offspring), and horizontal transmission (the acquisition of microbial symbionts from the surrounding environment) (19). The sponge microbiome confers a variety of benefits to the sponge host, including carbon metabolism (i.e. glycolysis, TCA cycle, 3-hydroxypropionate CO₂ fixation), nitrogen metabolism (i.e. nitrification, denitrification, ammonia oxidation), and vitamin production (i.e. cobalamin, riboflavin, thiamine, biotin) (15).

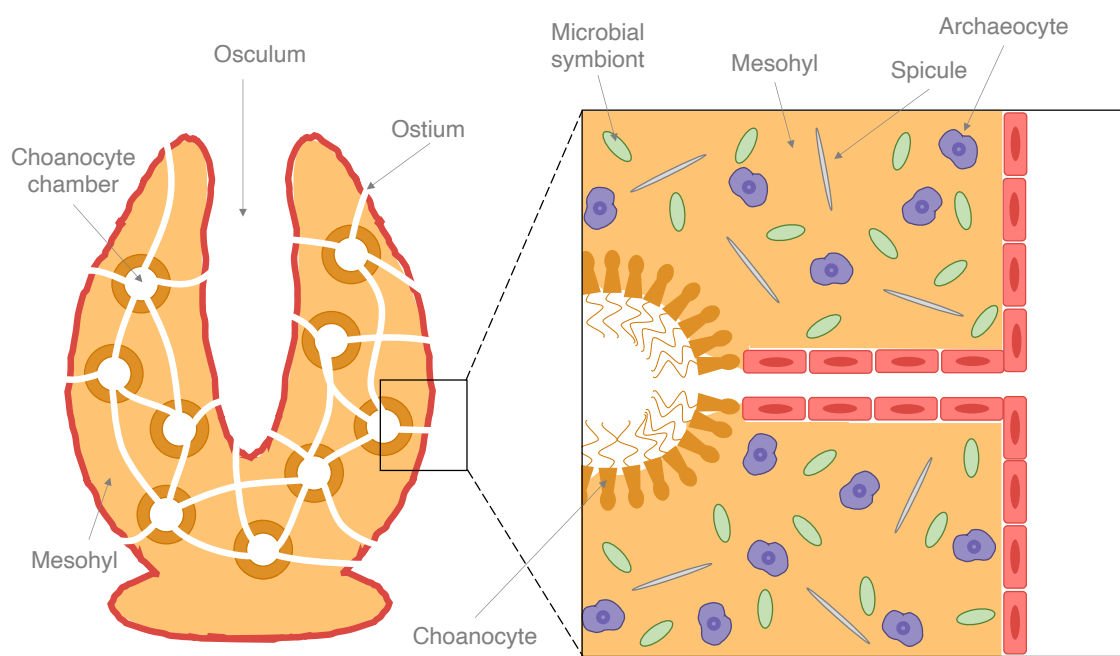


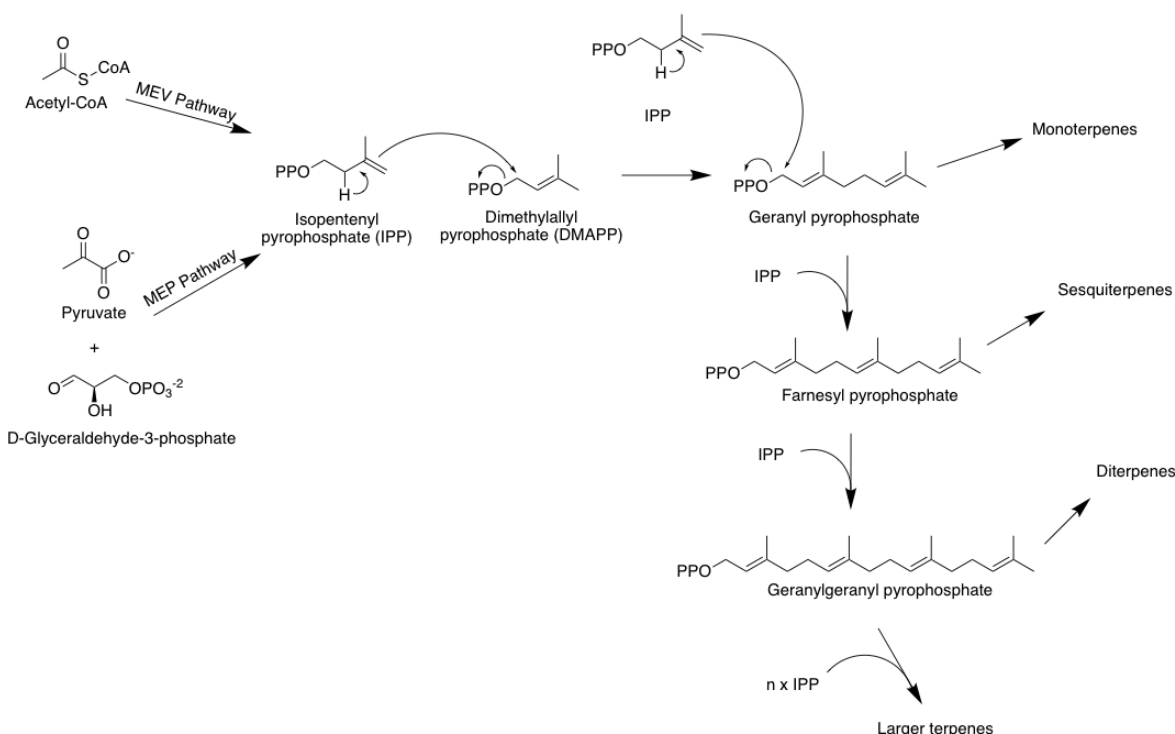
Figure 1.1.1. Body plan and internal structure of porifera (Adapted from (15)).

The sponge microbiome serves as a valuable source of chemical defense and much of that chemical defense comes in the form of bioactive polyketides, non-ribosomal peptides, and other molecules with anti-feeding and antimicrobial properties that protect the sponge animal host from predation and infection. Work in the past 20+ years has uncovered the incredible biosynthetic capacity of sponge microbes, such as studies on the polybrominated diphenylethers and chlorinated peptides produced by the Lamellodysidea symbiont *Hormoscilla spongelliae* (20), or the

onnamides produced by an unclassified bacterial symbiont of *Theonella swinhoei* (21). Today, the role of the sponge microbiome in the production of secondary metabolites is well established, so much so that most sponge-derived natural products are assumed to be produced by the sponge microbiome.

Section 1.2. Isoprenoid biosynthesis

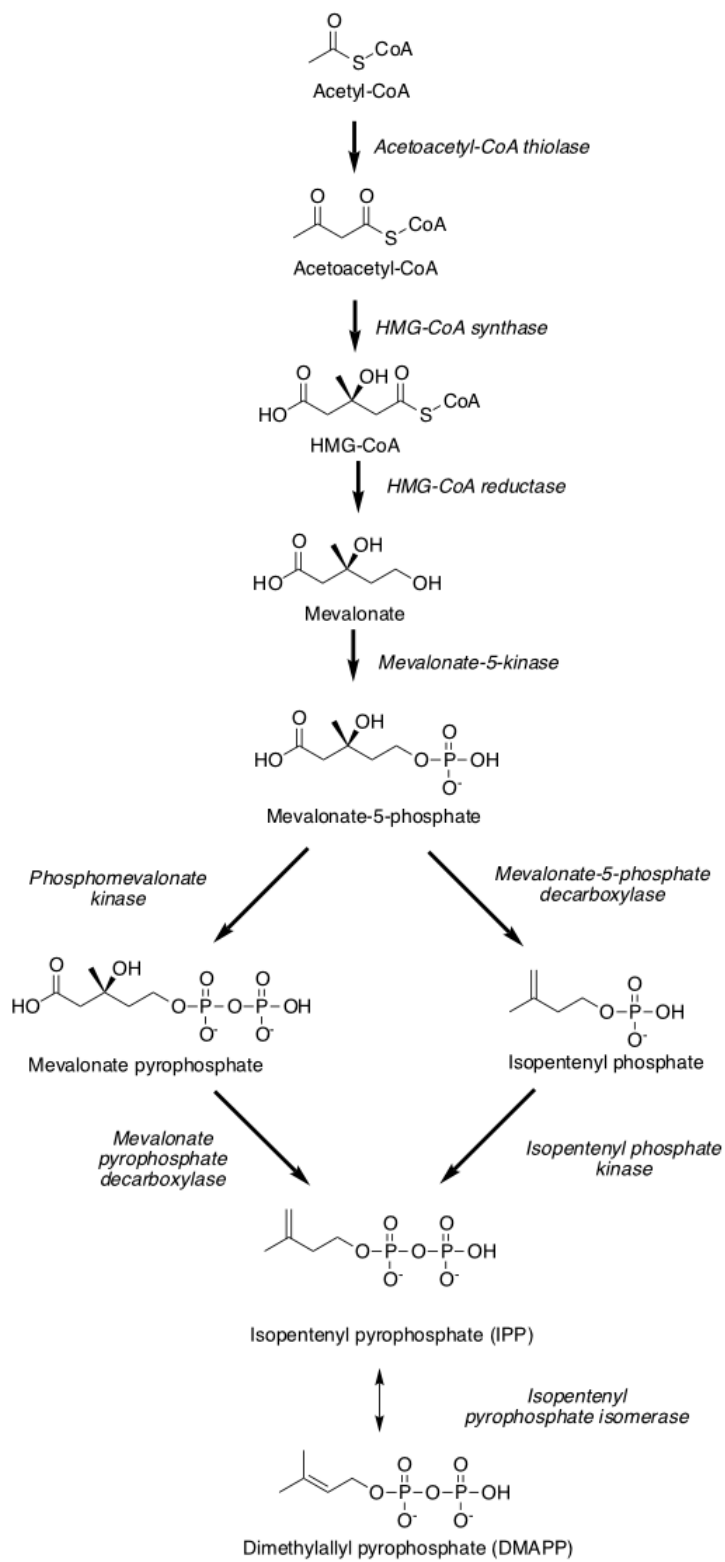
Terpenoids, which constitute the largest class of natural products, are composed of 5-carbon isoprene units and are classified by the number of isoprene units incorporated into the molecule (22). Monoterpenes are composed of two isoprene units, sesquiterpenes have three isoprene units, diterpenes have four isoprene units, with further categorization based on isoprene units, up until polyterpenes, which are terpenes that contain nine or more isoprene units.



Scheme 1.2.1. Isoprenoid biosynthesis pathways.

The isoprene units that make up terpenes come from isopentenyl pyrophosphate (IPP) and dimethylallyl diphosphate (DMAPP) which originate from either the mevalonate (MEV) pathway in eukaryotes, archaea, and some bacteria, or the methylerythritol 4-phosphate (MEP) pathway in plants, bacteria, and protozoa (Scheme 1.2.1)(23). An initial prenylation of IPP and DMAPP forms geranyl diphosphate (GPP), which can be further prenylated with molecules of IPP to yield various prenyl diphosphate species such as farnesyl diphosphate (FPP), which is the biosynthetic precursor to sesquiterpenes, or geranylgeranyl diphosphate (GGPP), which is the biosynthetic precursor to diterpenes.

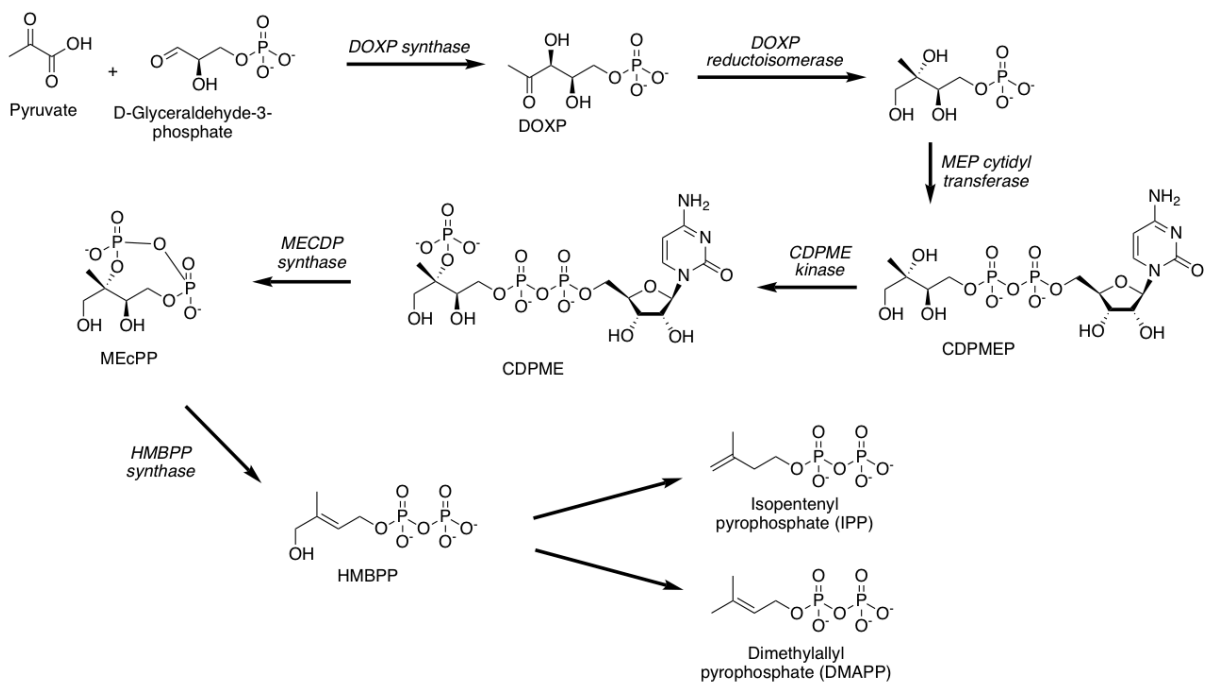
The MEV pathway, which was the first terpene biosynthesis pathway discovered, begins with a Claisen condensation of two acetyl-CoA molecules to yield acetoacetyl-CoA (Scheme 1.2.2)(24). HMG-CoA synthase then catalyzes the adol-like addition of acetyl-CoA to acetoacetyl-CoA, followed by the reduction of HMG-CoA to mevalonate by an HMG-CoA reductase. Mevalonate-5-kinase phosphorylates the hydroxyl group on the C5 position to yield mevalonate-5-phosphate. From there, one of two reactions occur with mevalonate-5-phosphate: an additional phosphorylation by phosphomevalonate kinase to yield mevalonate pyrophosphate (usually occurs in eukaryotes) or a decarboxylation by mevalonate-5-phosphate decarboxylase to yield isopentenyl phosphate (usually in archaea). Following those reactions, isopentenyl diphosphate (IPP) is generated from mevalonate diphosphate via mevalonate diphosphate decarboxylase or IPP is generated from isopentenyl phosphate via isopentenyl phosphate kinase. IPP can then undergo an isomerization catalyzed by IPP isomerase to give DMAPP, which can also be isomerized back to IPP via the same isomerase.



Scheme 1.2.2. The mevalonate (MEV) pathway.

The MEP pathway begins with the coupling of pyruvate and D-glyceraldehyde-3-phosphate to yield 1-deoxy-D-xylulose 5-phosphate (DOXP) (Scheme 1.2.3). DOXP-reductoisomerase is responsible for the carbon rearrangement and carbonyl reduction which gives 2C-methyl-D-erythritol 4-phosphate (MEP) from DOXP. MEP then reacts with cytidine triphosphate to yield diphosphocytidyl-2C-methyl-D-erythritol (CDPMEP), which is then phosphorylated by the ATP dependent CDPME kinase to yield 2-phospho-4-diphosphocytidyl-2C-methyl-D-erythritol (CDPME). 2C-methyl-D-erythritol 2,4-cyclodiphosphate (MECDP) synthase then catalyzes a cyclization to yield the cyclic diphosphate 2C-methyl-D-erythritol 2,4-cyclodiphosphate. A reductive deoxygenation catalyzed by HMBPP synthase opens the diphosphate cycle, giving (*E*)-4-hydroxy-3-methyl-2-butenyl diphosphate. A final reductive deoxygenation step yields a mixture of IPP and DMAPP, which later combine to yield terpenoid precursors.

For further information on terpenoid biosynthesis, see section 1.5.



Scheme 1.2.3. The methylerythritol 4-phosphate pathway.

Section 1.3. Marine nitrogenous terpenes

First discovered in 1973 in an Italian *Axinella cannabina* sponge, nitrogenous terpenes have fascinated and puzzled natural products scientists for more than forty years (25). Nitrogenous terpenes feature nitrogen-containing functional groups such as isonitriles, isothiocyanates, isocyanates, formamates, and amines. Approximately 360 nitrogenous terpenes have been reported from sponges, and approximately 50 from nudibranchs, though they are likely of dietary origin in the nudibranchs (26). These nitrogenous terpenes are unique to the marine environment and are only found in marine sponges and the nudibranchs that eat them. The vast majority of nitrogenous terpenes are sesquiterpenoids and diterpenoids, and while the nitrogenous functional groups remain fairly consistent, there is quite a wide variety in the isoprenoid backbone as well as the placement of the nitrogenous functional groups on the terpenoid scaffold (26)(Fig 1.3.1).

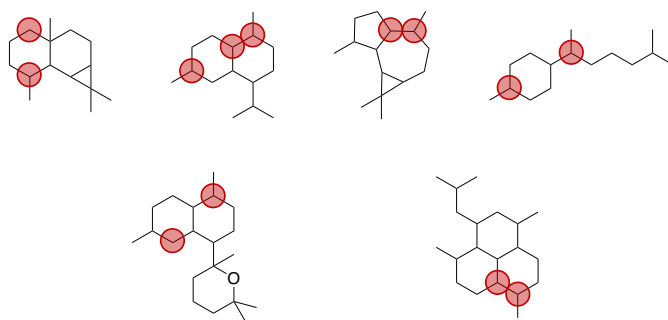


Figure 1.3.1. Common marine nitrogenous terpene scaffolds. Red circles represent common areas of nitrogenous functional groups (isonitriles, amines, isothiocyanates, formamides, isocyanates).

A large amount of work on the biosynthetic origin of these nitrogenous terpenes has been done using radiolabeled precursors. By feeding precursors that feature radioisotopes such as ^{14}C or ^3H , or stable isotopes such as ^{13}C or ^{15}N , to a natural product-producing organism, researchers

can track the isotopically labeled atoms through to the final natural products using a mass spectrometer, scintillation counter, or NMR. Isotopically labeled feeding experiments investigating the biosynthesis of nitrogenous terpenes was largely led by the Garson and Scheuer labs in the 1980s and 1990s, who both published multiple papers on the topic (Table 1.3.1).

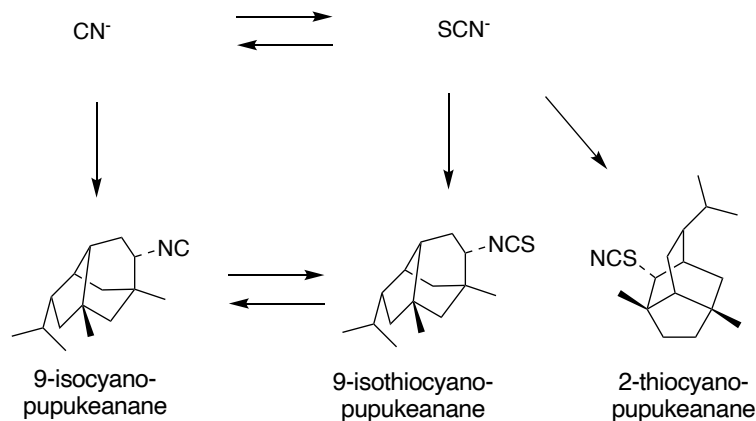
Pioneering work by the Garson group demonstrated that free inorganic cyanide and thiocyanate are the precursors to the nitrogen-containing functional groups of marine nitrogenous terpenes (27, 28). The incorporation of cyanide into the isonitrile group was first demonstrated by Garson in 1986 with the incorporation of [^{14}C]-cyanide into diisocyanoadociane (27), and was later confirmed by Karuso and Scheuer in 1989 with the incorporation of [^{14}C]-cyanide into 2-isocyanoneopupukeanane and kalihinol F (29). The same study by Karuso and Scheuer also confirmed that both the carbon and the nitrogen atoms of cyanide were incorporated into the isonitrile group of 9-isocyanoneopupukeanane, based on feeding experiments with [^{13}C , ^{15}N]-cyanide (29). Dumdei et al demonstrated that [^{14}C]-cyanide and [^{14}C]-thiocyanate were both incorporated into both axisonitrile-3 and axisothiocyanate-3, suggesting that cyanide and thiocyanate can be interconverted by the sponge (28). Further work in an Australian *Axinyssa* sponge showed that [^{14}C]-thiocyanate can be incorporated into the isothiocyano moiety of 9-isothiocyanatopupukeanane and the thiocyano moiety of 2-thiocyanatoneopupukeanane (30). In the same *Axinyssa* sponge, [^{14}C]-cyanide was also incorporated into the isothiocyano moiety of 9-isothiocyanatopupukeanane and the thiocyano moiety of 2-thiocyanatoneopupukeanane (30).

Several of these isotope labeling studies point to the interconnectedness of the nitrogenous functional groups (Scheme 1.3.1). Garson and Simpson demonstrated the interconversion of the isonitrile and isothiocyanate groups of 9-isocyanopupukeanane and 9-

Table 1.3.1. Summary of isotope label feeding experiments investigating nitrogenous terpenes in marine sponges.

Starting material	Labeled product	Labeled product functional group	Sponge	Citation
¹⁴ C]-cyanide	disocyanododecane	isonitriles	<i>Amphimedon terpenensis</i>	M. J. Garson, J. Chem. Soc., Chem. Commun., 1986, 35
¹⁴ C]-disothiocyanatododecane	disocyanododecane	isonitriles	<i>Amphimedon terpenensis</i>	M. J. Garson, J. Chem. Soc., Chem. Commun., 1986, 35
¹⁴ C]-cyanide	2-isocyanopropylketone	isonitriles	<i>Cyclalypta</i> sp.	P. Karuso and P. J. Scheuer, J. Org. Chem., 1989, 54, 2092
¹⁴ C]-cyanide	calthanol F	isonitriles	<i>Acanthella</i> sp.	P. Karuso and P. J. Scheuer, J. Org. Chem., 1989, 54, 2092
¹³ C, ¹⁵ N]-cyanide	9-isocyanononopropylketone	isonitriles	<i>Cyclalypta</i> sp.	P. Karuso and P. J. Scheuer, J. Org. Chem., 1989, 54, 2092
¹⁴ C]-cyanide	axisonitrile-3	isonitrile	<i>Acanthella cavernosa</i>	E. J. Dumdei, A. E. Flowers, M. J. Garson and C. J. Moore, Comp. Biochem. Physiol., 1997, 118A, 1385
¹⁴ C]-thiocyanate	axisonitrile-3	isonitrile	<i>Acanthella cavernosa</i>	E. J. Dumdei, A. E. Flowers, M. J. Garson and C. J. Moore, Comp. Biochem. Physiol., 1997, 118A, 1385
¹⁴ C]-cyanide	axiothiocyanate-3	isothiocyanate	<i>Acanthella cavernosa</i>	E. J. Dumdei, A. E. Flowers, M. J. Garson and C. J. Moore, Comp. Biochem. Physiol., 1997, 118A, 1385
¹⁴ C]-thiocyanate	axiothiocyanate-3	isothiocyanate	<i>Acanthella cavernosa</i>	E. J. Dumdei, A. E. Flowers, M. J. Garson and C. J. Moore, Comp. Biochem. Physiol., 1997, 118A, 1385
¹⁴ C]-thiocyanate	2-thiocyanatopropylketone	thiocyanate	<i>Azirisia</i> n.sp.	J. S. Simpson and M. J. Garson, Tetrahedron Lett., 1998, 39, 5819
¹⁴ C]-thiocyanate	9-isothiocyanatopropylketone	thiocyanate	<i>Azirisia</i> n.sp.	J. S. Simpson and M. J. Garson, Tetrahedron Lett., 1998, 39, 5819
¹⁴ C]-cyanide	2-thiocyanatopropylketone	thiocyanate	<i>Azirisia</i> n.sp.	J. S. Simpson and M. J. Garson, Tetrahedron Lett., 1998, 39, 5819
¹⁴ C]-cyanide	2-thiocyanatopropylketone	thiocyanate	<i>Azirisia</i> n.sp.	J. S. Simpson and M. J. Garson, Tetrahedron Lett., 1998, 39, 5819
¹⁴ C]-isocyanopropylketone	9-isothiocyanatopropylketone	isothiocyanate	<i>Azirisia</i> n.sp.	J. S. Simpson and M. J. Garson, Tetrahedron Lett., 2001, 42, 4267
¹⁴ C]-9-isothiocyanopropylketone	9-isocyanopropylketone	isonitrile	<i>Azirisia</i> n.sp.	J. S. Simpson and M. J. Garson, Tetrahedron Lett., 2001, 42, 4267
¹⁴ C]-axamide	none		<i>Azirisia</i> n.sp.	A. Tingo, C. Santacrose and G. Sodano, Experientia, 1979, 35, 10
¹³ C]-2-formamidopropylketone	none		<i>Hymeniacidon</i> sp.	M. R. Hagadone, P. J. Scheuer and A. Holm, J. Am. Chem. Soc., 1984, 106, 2447
¹³ C]-2-isocyanopropylketone	2-isothiocyanatopropylketone	isothiocyanate	<i>Hymeniacidon</i> sp.	M. R. Hagadone, P. J. Scheuer and A. Holm, J. Am. Chem. Soc., 1984, 106, 2447
¹³ C]-2-isothiocyanopropylketone	no label in 2-isocyanopropylketone		<i>Hymeniacidon</i> sp.	M. R. Hagadone, P. J. Scheuer and A. Holm, J. Am. Chem. Soc., 1984, 106, 2447
¹⁴ C]-cyanide	farnesyl isothiocyanate	isothiocyanate	<i>Syrtocella aurantium</i>	J. S. Simpson, P. Raniga and M. J. Garson, Tetrahedron Lett., 1997, 38, 7947
¹⁴ C]-thiocyanate	farnesyl isothiocyanate	isothiocyanate	<i>Syrtocella aurantium</i>	J. S. Simpson, P. Raniga and M. J. Garson, Tetrahedron Lett., 1997, 38, 7947
¹⁴ C]-cyanide	styrolane A	dichloroimine	<i>Syrtocella aurantium</i>	J. S. Simpson, P. Raniga and M. J. Garson, Tetrahedron Lett., 1997, 38, 7947
¹⁴ C]-cyanide	styrolane B	dichloroimine	<i>Syrtocella aurantium</i>	J. S. Simpson, P. Raniga and M. J. Garson, Tetrahedron Lett., 1997, 38, 7947
¹⁴ C]-thiocyanate	styrolane A	dichloroimine	<i>Syrtocella aurantium</i>	J. S. Simpson, P. Raniga and M. J. Garson, Tetrahedron Lett., 1997, 38, 7947
¹⁴ C]-thiocyanate	styrolane B	dichloroimine	<i>Syrtocella aurantium</i>	J. S. Simpson, P. Raniga and M. J. Garson, Tetrahedron Lett., 1997, 38, 7947
¹⁴ C]-farnesyl isonitrile	farnesyl isothiocyanate	isothiocyanate	<i>Syrtocella aurantium</i>	A. Brust and M. J. Garson, Tetrahedron Lett., 2003, 44, 327
¹⁴ C]-farnesyl isonitrile	styrolane B	dichloroimine	<i>Syrtocella aurantium</i>	A. Brust and M. J. Garson, Tetrahedron Lett., 2003, 44, 327
¹⁴ C]-farnesyl isonitrile	styrolane A	dichloroimine	<i>Syrtocella aurantium</i>	A. Brust and M. J. Garson, Tetrahedron Lett., 2003, 44, 327
¹⁴ C]-farnesyl isonitrile	styrolane B	dichloroimine	<i>Syrtocella aurantium</i>	A. Brust and M. J. Garson, Tetrahedron Lett., 2003, 44, 327
¹⁴ C]-farnesyl formamide	no label in dichloroimine styrolanes		<i>Syrtocella aurantium</i>	A. Brust and M. J. Garson, Tetrahedron Lett., 2003, 44, 327
¹³ C]-formate	none		<i>Hymeniacidon</i> sp.	M. R. Hagadone, P. J. Scheuer and A. Holm, J. Am. Chem. Soc., 1984, 106, 2447
¹⁴ C]-acetate	none		<i>Amphimedon terpenensis</i>	M. J. Garson, J. Chem. Soc., Chem. Commun., 1986, 35
¹⁴ C]-mevalonate	none		<i>Amphimedon</i> sp.	M. J. Garson, V. Partali, S. Liaen-Jensen and I. L. Stoltov, Comp. Biochem. Physiol., 1988, 91B, 293
¹⁴ C]-acetate	β-β-Carotene		<i>Amphimedon</i> sp.	M. J. Garson, V. Partali, S. Liaen-Jensen and I. L. Stoltov, Comp. Biochem. Physiol., 1988, 91B, 293
¹⁴ C]-acetate	Zeaxanthin		<i>Amphimedon</i> sp.	M. J. Garson, V. Partali, S. Liaen-Jensen and I. L. Stoltov, Comp. Biochem. Physiol., 1988, 91B, 293
¹⁴ C]-alanine	none		<i>Amphimedon</i> sp.	C. J. R. Fookes, M. J. Garson, J. K. MacLeod, B. W. Skelton and A. H. White, J. Chem. Soc., Perkin Trans. 1, 1988, 1003
¹⁴ C]-glycine	none		<i>Amphimedon</i> sp.	C. J. R. Fookes, M. J. Garson, J. K. MacLeod, B. W. Skelton and A. H. White, J. Chem. Soc., Perkin Trans. 1, 1988, 1003
¹⁴ C]-leucine	none		<i>Amphimedon</i> sp.	C. J. R. Fookes, M. J. Garson, J. K. MacLeod, B. W. Skelton and A. H. White, J. Chem. Soc., Perkin Trans. 1, 1988, 1003
¹⁴ C]-arginine	none		<i>Amphimedon</i> sp.	C. J. R. Fookes, M. J. Garson, J. K. MacLeod, B. W. Skelton and A. H. White, J. Chem. Soc., Perkin Trans. 1, 1988, 1003
¹⁴ C]-formate	none		<i>Amphimedon</i> sp.	C. J. R. Fookes, M. J. Garson, J. K. MacLeod, B. W. Skelton and A. H. White, J. Chem. Soc., Perkin Trans. 1, 1988, 1003

isothiocyanopupukeanane with ^{14}C isotope feeding experiments in an *Axinyssa* sponge (31). When Garson and Simpson fed ^{14}C -9-isocyanopupukeanane to the *Axinyssa* sponge, they saw the production of ^{14}C -9-isothiocyanopupukeanane and vice versa (31). Iengo, Santacroce, and Sodano hypothesized that the formamide group of axamide-1 was the precursor to the related isonitrile axisonitrile-1 and to test this, they fed an *A. cannabina* sponge a ^{14}C -labeled axisonitrile-1 (32). They observed no incorporation of the radiolabel into any of the other related functionalized sesquiterpenes, which suggested the formamide was not a precursor to the other nitrogenous sesquiterpenes. Hagedone, Scheuer, and Holm investigated the origin of the isonitrile group of 2-isocyanopupukeanane by feeding a *Hymeniacidon* sp. sponge the corresponding formamido- and isothiocyano pupukeananes, both ^{13}C -labeled (33). There was no reported incorporation of either the formamido or isothiocyano label into 2-isocyanopupukeanane.



Scheme 1.3.1. Interconnectedness of nitrogenous terpenes and their precursors based on isotope feeding studies in *Axinyssa* sp. (29, 31, 33).

Labeling studies in *Stylotella aurantium* have also shown that cyanide and thiocyanide are involved in the production of dichloroimine sesquiterpenoid natural products (34, 35). Simpson et al. saw the incorporation of ^{14}C -cyanide and ^{14}C -thiocyanate into the dichlorimine

sesquiterpenoids stylotellane A and B, as well as into farnesyl isothiocyanate, a natural product which often co-occurs with dichloroimine sesquiterpenoids (34). In a 2003 *S. aurantium* study, Brust and Garson showed that [^{14}C]-farnesyl isonitrile was converted into [^{14}C]-stylotellane A and B and [^{14}C]-farnesyl isothiocyanate (35). [^{14}C]-Farnesyl isothiocyanate was also incorporated into stylotellane A (35).

Despite many efforts to elucidate the origin of the terpene backbone or cyanide, feeding studies failed to yield conclusive results on these matters. Feeding isotopically labeled carbon sources, such as formate, mevalonate, and acetate did not yield any isotopically labeled terpenoids (27, 33, 36, 37), though [^{14}C]-acetate was incorporated into the isoprenoids β,β -carotene and zeaxanthin in *Amphimedon* sp. (36). In other biological systems such as plants and microbes, cyanide is generated from amino acids (38), which prompted radio labeled amino acid feeding experiments in *Amphimedon* sp. No radio label incorporation was seen when *Amphimedon* sp. was fed radiolabeled leucine, alanine, or arginine (37).

1.4. Introduction to “Specialized terpenoid biosynthesis across domains of life”

Terpenoids are the largest class of natural products and have impressive ecological and medicinal activities. Recently, several new classes of terpene synthases have been described from taxonomic groups not previously known to contain specialized terpene synthases. I am currently the co-first author, along with Taylor Steele, of a review in preparation for submission to Natural Product Reports. Immo Burkhardt and Bradley Moore are also co-authors on this review in preparation. Taylor, Immo, and I have all recently discovered new terpene synthases from marine eukaryotes and have expertise in terpene synthases from marine eukaryotes. As we were reviewing the terpene literature, we noticed that the field is lacking a taxonomy-based perspective on terpene

synthases and their evolution. While several reviews exist that are specific to one organism or one organismal class, no review has covered terpene synthases across all domains of life. We decided to prepare this review to showcase the diversity of type I terpene synthases across the tree of life and speak to the evolution and ecological roles of specialized terpenoids. We provide an overview of class 1 terpene synthases from plants, fungi, bacteria, algae, insects, amoeba, and sponges. For each organism class, we cover the types of terpenoids produced by those organisms, what terpene synthases have been characterized from those organisms, what is known about the structure and function of those terpene synthases, and the genomic context of those terpene synthases. This review is in preparation for submission at Natural Product Reports later this year.

Section 1.5 References for Sections 1.1-1.4

1. A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, C. T. Supuran, Natural products in drug discovery : advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).
2. D. A. Dias, S. Urban, U. Roessner, A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2**, 303–336 (2012).
3. D. J. Newman, G. M. Cragg, Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
4. C. Lyu, T. Chen, B. Qiang, N. Liu, H. Wang, L. Zhang, Z. Liu, CMNPD: A comprehensive marine natural products database towards facilitating drug discovery from the ocean. *Nucleic Acids Res.* **49**, D509–D515 (2021).
5. L. Cilliers, F. Retief, *Poisoning in Ancient Rome* (Elsevier Inc., 2019) <https://doi.org/10.1016/B978-0-12-815339-0.00015-9>.
6. J. R. Pawlik, C. D. Amsler, R. Ritson-, J. B. McClintock, B. J. Baker, V. J. Paul, “Marine Chemical Ecology : A Science Born of Scuba” in *Research and Discoveries: The Revolution of Science through Scuba*, M. A. Lang, R. L. Marinelli, S. J. Roberts, P. R. Taylor, Eds. (Smithsonian Institution Scholarly Press, 2013), pp. 53–69.
7. W. Bergmann, R. J. Feeney, Contributions to the study of marine products. xxxii. The nucleosides of sponges. *J. Org. Chem.* **16**, 981–987 (1951).
8. P. R. Burkholder, R. M. Pfister, F. H. Leitz, R. M. Pfister, F. H. Leitz, Production of a Pyrrole Antibiotic by a Marine Bacterium. *Appl. Microbiol.* **14**, 649–653 (1966).
9. A. J. Weinheimer, R. L. Spraggins, The occurrence of two new prostaglandin derivatives (15-epi-PGA2 and its acetate, methyl ester) in the Gorgonian *Plexaura homomalla* Chemistry of Coelenterates. XV. *Tetrahedron Lett.* **10**, 5185–5188 (1969).
10. L. E. Fleming, L. B. A. Creencia, W. H. Gerwick, H. C. Goh, M. O. Gribble, B. Maycock, H. Solo-Gabriele, Eds. *Oceans and Human Health: Opportunities and Impacts*, 2nd Edition (Elsevier Science, 2023)
11. C. Nielsen, Early animal evolution : a morphologist ’ s view. *R. Soc. Open Sci.* **6**, 1–10 (2019).
12. B. J. Nettersheim, J. J. Brocks, A. Schwelm, J. M. Hope, F. Not, M. Lomas, C. Schmidt, R. Schiebel, E. C. M. Nowack, P. De Deckker, J. Pawlowski, S. S. Bowser, I. Bobrovskiy, K. Zonneveld, M. Kucera, M. Stuhr, C. Hallmann, Putative sponge biomarkers in unicellular Rhizaria question an early rise of animals. *Nat. Ecol. Evol.* **3**, 577–581 (2019).
13. Z. Yin, M. Zhu, E. H. Davidson, D. J. Bottjer, F. Zhao, P. Tafforeau, Sponge grade body

- fossil with cellular resolution dating 60 Myr before the Cambrian. *Proc. Natl. Acad. Sci.* **112**, 1453–1460 (2015).
14. E. C. Turner, Possible poriferan body fossils in early Neoproterozoic microbial reefs. *Nature* **596**, 87–91 (2021).
 15. U. Hentschel, J. Piel, S. M. Degnan, M. W. Taylor, Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* **10**, 641–654 (2012).
 16. M. Maldonado, Sponge waste that fuels marine oligotrophic food webs : a re-assessment of its origin and nature. *Mar. Ecol.* **37**, 477–491 (2016).
 17. C. R. Wilkinson, R. Garrone, J. Vacelet, Marine sponges discriminate between food bacteria and bacterial symbionts : electron microscope radioautography and situ evidence *. *Proc. R. Soc. London* **528**, 519–528 (1984).
 18. M. Wehrl, M. Steinert, U. Hentschel, Microbial Ecology Bacterial Uptake by the Marine Sponge *Aplysina aerophoba*. *Microb. Ecol.* **53**, 355–365 (2007).
 19. S. Schmitt, H. Angermeier, R. Schiller, N. Lindquist, U. Hentschel, Molecular Microbial Diversity Survey of Sponge Reproductive Stages and Mechanistic Insights into Vertical Transmission of Microbial Symbionts. *Appl. Environ. Microbiol.* **74**, 7694–7708 (2008).
 20. V. Agarwal, J. M. Blanton, S. Podell, A. Taton, M. A. Schorn, J. Busch, Z. Lin, E. W. Schmidt, P. R. Jensen, V. J. Paul, J. S. Biggs, J. W. Golden, E. E. Allen, B. S. Moore, Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat. Chem. Biol.* **13**, 537–543 (2017).
 21. J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, S. Matsunaga, Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl. Acad. Sci.* **101**, 16222–16227 (2004).
 22. D. W. Christianson, Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **117**, 11570–11648 (2017).
 23. W. Eisenreich, M. Schwarz, A. Cartayrade, D. Arigoni, M. H. Zenk, A. Bacher, The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms. *Chem. Biol.* **5**, R221–R233 (1998).
 24. J. McMurry, T. P. Begley, *The Organic Chemistry of Biological Pathways*, 2nd Ed. (Roberts and Company Publishers, 2016).
 25. F. Cafieri, E. Fattorusso, S. Magno, C. Santacroce, D. Sica, Isolation and Structure of Axisonitrile-1 and axisothiocyanate-1: Two Unusual Sesquiterpenoids from the Marine Sponge *Axinella cannabina*. *Tetrahedron* **29**, 4259–4262 (1973).

26. J. Emsermann, U. Kahl, T. Opatz, Marine isonitriles and their related compounds. *Mar. Drugs* **14** (2016).
27. M. J. Garson, Biosynthesis of the Novel Diterpene Isonitrile Diisocyanoadociane by a Marine Sponge of the Amphimedon Genus: Incorporation Studies with Sodium [¹⁴C]Cyanide and Sodium [2-¹⁴C]Acetate. *J. Chem. Soc. Chem. Commun.*, 35–36 (1986).
28. E. J. Dumdei, E. Flowers, M. Garson, M. June, The Biosynthesis of Sesquiterpene Isocyanides and Isothiocyanates in the Marine Sponge *Acanthella cavernosa* (Dendy); Evidence for Dietary Transfer to the Dorid Nudibranch *Phyllidiella pustulosa*. *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* **118**, 1385–1392 (1997).
29. P. Karuso, P. J. Scheuer, Biosynthesis of Isocyanoterpenes in Sponges. *J. Org. Chem.* **54**, 2092–2095 (1989).
30. J. S. Simpson, M. J. Garson, Thiocyanate Biosynthesis in the Tropical Marine Sponge *Axinyssa* n. sp. *Tetrahedron Lett.* **39**, 5819–5822 (1998).
31. J. S. Simpson, M. J. Garson, Advanced precursors in marine biosynthetic study. Part 2: The biosynthesis of isocyanides and isothiocyanates in the tropical marine sponge *Axinyssa* n. sp. **42**, 4267–4269 (2001).
32. A. Iengo, C. Santacroce, G. Sodano, Metabolism in Porifera. X. On the intermediary of a formamide moiety in the biosynthesis of isonitrile terpenoids in sponges. *Experientia* **35**, 10–11 (1979).
33. M. R. Hagadone, P. J. Scheuer, A. Holm, On the Origin of the Isocyano Function in Marine Sponges. *J. Am. Chem. Soc.* **106**, 2447–2448 (1984).
34. J. S. Simpson, P. Raniga, M. J. Garson, Biosynthesis of dichloroimines in the tropical marine sponge *Stylotella aurantium*. *Tetrahedron Lett.* **38**, 7947–7950 (1997).
35. A. Brust, M. J. Garson, Advanced precursors in marine biosynthetic study. Part 3: The biosynthesis of dichloroimines in the tropical marine sponge *Stylotella aurantium*. *Tetrahedron Lett.* **44**, 327–330 (2003).
36. M. J. Garson, V. Partali, S. Liaaen-Jensen, I. L. Stoilov, Isoprenoid biosynthesis in a marine sponge of the Amphimedon genus: Incorporation studies with [1-¹⁴C]acetate, [4-¹⁴C]cholesterol and [2-¹⁴C]mevalonate. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **91B**, 293–300 (1988).
37. C. J. R. Fookes, M. J. Garson, J. K. MacLeod, B. W. Skelton, A. H. White, Biosynthesis of Diisocyanoadociane, a Novel Diterpene from the Marine Sponge *Amphimedon* Sp. Crystal Structure of a Monoamide Derivative. *J. Chem. Soc. Perkin Trans. 1*, 1003–1011 (1988).
38. B. Vennesland, E. E. Conn, C. J. Knowles, J. Westley, F. Wissling, Eds., *Cyanide in Biology*

(Academic Press (London), 1981).

Section 1.6. Specialized terpenoid biosynthesis across domains of life

Section 1.6.1. Abstract

Terpenoids are widely distributed and considered the largest class of natural product chemicals. Found across all domains of life, terpenoids encompass both primary and secondary, or specialized metabolic products. The general biosynthetic machinery for terpenoid precursor supply is ubiquitous across all organisms, though only certain groups are known to produce specialized terpenoids. While initial work on terpenoid biosynthesis focused on plants, the field has rapidly expanded in recent decades, first with bacterial and fungal terpene synthases, and later with those found in insects, algae, and marine animals. The rapid expansion of terpenoid biosynthetic knowledge has been fueled by DNA sequencing and genomic advances that reveal genes encoding terpenoid biosynthesis are often co-clustered, from microbes to plants and animals. These organisms construct terpene scaffolds primarily using class 1 terpene synthases, which lay the foundation for the vast structural diversity of terpenoids. In this review we aim to highlight terpenoid diversity in secondary metabolism across different branches of the tree of life and discuss the apparent evolutionary success of underlying enzymatic principles.

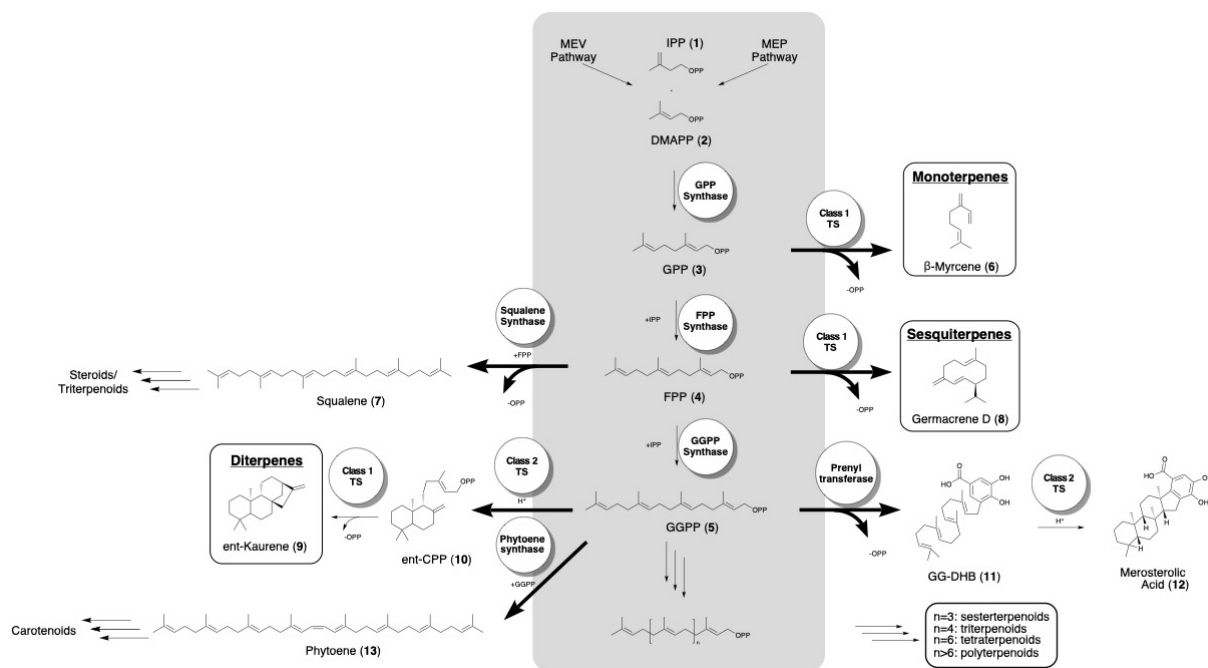
Section 1.6.2. Introduction

Terpenoids are widely considered to be the largest and most diverse class of naturally occurring compounds. Estimates of characterized terpenoids currently number more than 100,000, excluding steroids, and encompass both primary and secondary metabolism. In this review, we defer to work by Chevrette and co-authors to define primary and secondary metabolism (1). Primary, or central metabolism, refers to pathways with a wide taxonomic distribution and whose products are required for autonomous growth, sometimes supported by redundant, more

promiscuous pathways. Secondary metabolism refers to pathways with limited taxonomic distribution and are often the result of enzyme promiscuity and recruitment. Secondary metabolism evolution is typically dynamic, undergoing diversifying or purifying selection to offer a competitive advantage in a given environment, which may have a profound effect on survival. Within secondary metabolism, specialized metabolism is defined as a secondary metabolic pathway in which the metabolite is experimentally described to contribute to niche-specific functions. In primary metabolism, terpenoids, such as steroids and carotenoids, perform various essential physiological roles like mediating growth and reproduction (2). In secondary metabolism, terpenoids help an organism adapt to changes in their environment and serve key functions, such as signaling, defense, and attraction. Initial work on secondary terpenoids often focused on terrestrial plants, which are well known for production of terpenoid-rich essential oils (3), however, these ‘secondary’ terpenoids are now understood to be ubiquitous across the tree of life (4).

Beyond terrestrial plants, terpenoids outside of primary metabolism have been characterized in fungi (5), bacteria (6), and algae (7) as well as many marine animals, like octocorals (8), and sponges (9). Terpenoids also serve many human purposes, for example as key components in the production of widely used fragrances, flavorings, and resins. The medicinal properties of terpenoids have been recognized by humans for centuries in traditional medicines, as well as in modern treatment. Many notable pharmaceuticals are, or are inspired by terpenoid natural products, such as taxol (10) and artemisinin (11). Similarly, because of their role in plant defense mechanisms, many terpenoids are toxic to herbivores and other pests, making them a valuable tool for natural pest control. Overall, terpenoids play a crucial role in the natural world, contributing to the survival and reproduction of many species. While many aspects of terpenoid

chemistry, biosynthesis, evolution and ecology have been reviewed separately, the recent discoveries of class 1 terpene synthases in many different types of organisms with distinct lifestyles, producing diverse chemical structures with a virtually identical set of enzymes, has raised the questions: 1. What renders the terpene biosynthetic logic and underlying enzymology suitable for these different lifeforms? and 2. What could explain the apparent interkingdom success of terpenoids compared to other classes of secondary metabolites? Thus, in this review we will first define class 1 terpenoid diversity and distribution around the tree of life and discuss the enzymatic principles and biosynthetic logic for the assembly of these compounds. Then, we will highlight elucidated biosynthetic pathways, bioactivities, and ecological aspects of class 1 terpenoids produced in taxa with identified TS genes. Finally, we will discuss similarities, differences and overarching principles that might indicate the apparent success of this secondary metabolite machinery.



Scheme 1.6.1. Overview of isoprenoid biosynthesis pathways from across the tree of life.

Section 1.6.3. Basic Biosynthetic Principles

Terpenoids, terpenes, and isoprenoids are all synonymous for natural products constructed from two basic 5-carbon (C5) building blocks: isopentenyl diphosphate (IPP)(1) and dimethylallyl diphosphate (DMAPP)(2). DMAPP and IPP can derive from two independent nonhomologous metabolic pathways, the mevalonate pathway (MEV) or the methylerythritol 4-phosphate (MEP) pathway. The MEV pathway, which utilizes acetyl CoA and is found in most eukaryotes and archaea, was the first and (for several decades) only known pathway for isoprenoid biosynthesis (12). Originally described in yeast and animals in the 1950s, it was assumed to be the only route for IPP and DMAPP biosynthesis until the discovery of the MEP pathway in the 1990s (13). The MEP pathway, which utilizes 3-glyceral phosphate and pyruvate, is found in most bacteria, as well as in plants, which contain both the MEV and MEP pathways. Interestingly, the uncharacterized Candidate Phyla Radiation (CPR) bacterial phylum seems to have a MEV pathway equivalent (14, 15). In plants and other photosynthetic eukaryotes like algae, the MEP pathway is located exclusively in the plastid, suggesting the pathway dates to the ancestral cyanobacterial endosymbiont from which the plastid originated, though some phylogenetic evidence contradicts this theory (12). Overall, the evolutionary history and conservation of these two isoprenoid biosynthesis pathways reinforces the importance of their biological functions.

DMAPP units are extended with IPP units by an isopentenyl diphosphate synthase (IDS) in a 'head-to-tail' reaction to yield the linear isoprenoid diphosphate precursors, beginning with C10 geranyl diphosphate (GPP)(3) and then C15 farnesyl diphosphate (FPP)(4). Further extension of FPP with IPP yields C20 geranylgeranyl diphosphate (GGPP)(5) and finally C25 geranylfarnesyl diphosphate (GFPP)(Scheme 1.6.1). Every known organism possesses the

biosynthetic machinery to produce these linear oligoprenyl precursor and there are numerous classes of enzymes that utilize isoprenyl diphosphate substrates and can be transformed into a variety of terpenoid compounds. These terpenoid compounds are typically named according to the number of C5 isoprene units incorporated into their structure, beginning with hemiterpenes (one), monoterpenes (two), sesquiterpenes (three), diterpenes (four), and sesterterpenes (five), triterpenes (six) and lastly tetraterpenes (eight). In a majority of organisms, two equivalents of FPP are converted by a squalene synthase in a head-to-head reaction to squalene (**7**)(Scheme 1.6.1), which is the precursor to many primary metabolites like steroids, and hopanoids. GGPP can be dimerized head-to-head as well to produce phytoene (**13**), the universal precursor to all phytoenoids and carotenoids. Isoprenyl diphosphate substrates can be acted upon by class 2 TSs, which catalyze a protonation-induced cyclization, and often leaves the pyrophosphate group intact for further biosynthetic steps. Isoprenyl diphosphates are also utilized by prenyltransferases, which transfer the allylic prenyl group onto a receptor molecule. This prenyl chain can be further cyclized by terpene synthases, often containing multiple rings and stereocenters. The resulting terpenoid products from these described reactions are key structural scaffolds to the immense and remarkable structural diversity of terpenoids. The enzymes that convert isoprenyl diphosphates to secondary/specialized, terpenoids directly are known as class 1 terpene cyclases (TSs) and will be the focus of this review. Unlike class 2 TS enzymes, which generate carbocations via protonation induced cyclization, class 1 TSs rely on dephosphorylation to catalyze allylic cation formation. Consequently, class 1 TSs can act either directly on linear IDS products like GPP and FPP or on the products of class 2 TSs that still have an intact pyrophosphate group, like ent-CPP (**10**)(Scheme 1.6.1).

The structure of a class 1 TS stabilizes the allylic cation and facilitates the many intramolecular reactions notable in terpenoid biochemistry, including hydride and methyl shifts, cyclization, and carbocation quenching. These cationic cascades often result in polycyclic and complex hydrocarbon products. Structurally, all class 1 TS enzymes are composed of a metal-binding alpha-helical bundle, a feature that is also seen in, and likely arose from IDS enzymes (Fig. 1.6.1A) (16). Most IDSs and non-plant TSs are single domain enzymes consisting of one alpha domain (Fig. 1.6.1B), while plant terpene cyclases are di-, or tri-domain enzymes (Fig. 1.6.1C). Class 1 plant TSs always contain an alpha domain, which shares the same fold as

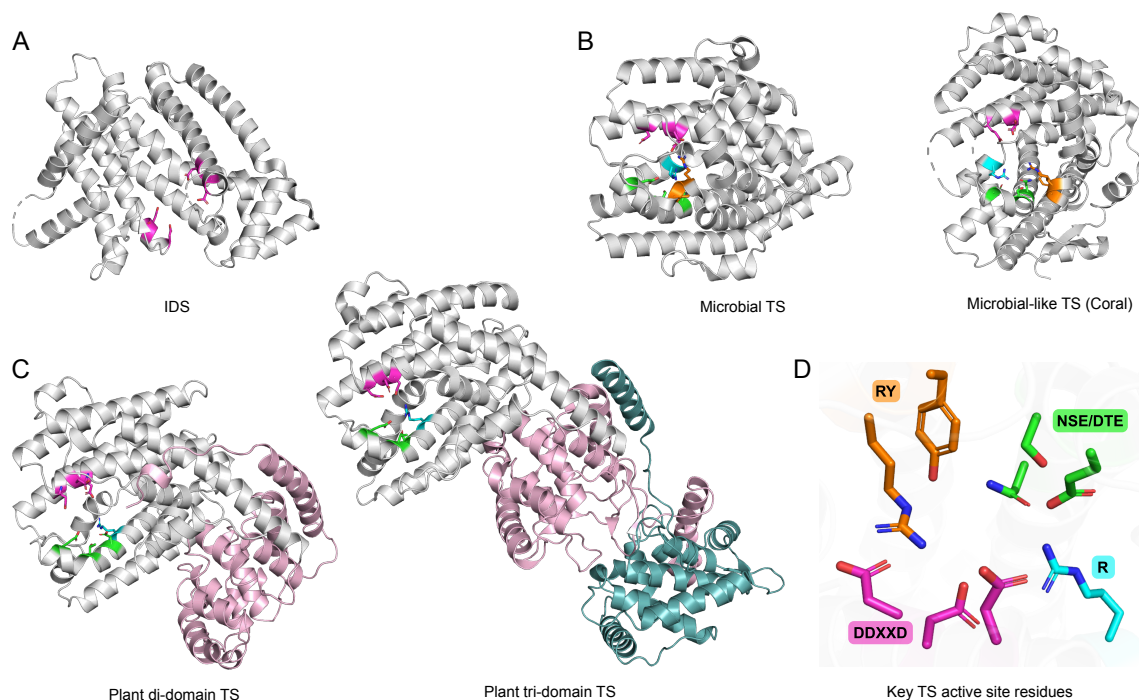


Figure 1.6.1. Structural overview of TS enzymes A. Isopentenyl diphosphate synthase (IDS) (FPP synthase from *Homo sapiens* PDB ID: 5JA0) with first and second aspartate-rich motifs highlighted (FARM and SARM, respectively) B. Microbial TS (epi-isozizaene synthase from *Streptomyces coelicolor*; PDB ID: 4LZ0) and microbial-type TS (cembrane A synthase from *Eleutherobia rubra*; PDB ID: 7S5L) C. Plant di-domain TS (5-epi-aristolochene synthase from *Nicotiana tabacum*; PDB ID: 5EAS) and plant tri-domain TS (taxadiene synthase from *Taxus brevifolia*; PDB ID: 3P5R) Color code: gray = alpha-domain, light pink = beta-domain, teal blue = gamma-domain. D. Expansion of TS active site from the microbial TS, showing DDXXD (neon pink), NSE/DTE (green), R (neon blue), and RY (orange) motifs. Key active site motifs highlighted in each structure according to this color scheme.

monodomain enzymes. Didomain enzymes consist of this alpha domain and an additional beta domain, which shares the conserved fold of squalene-cyclase like class 2 TSs but is often

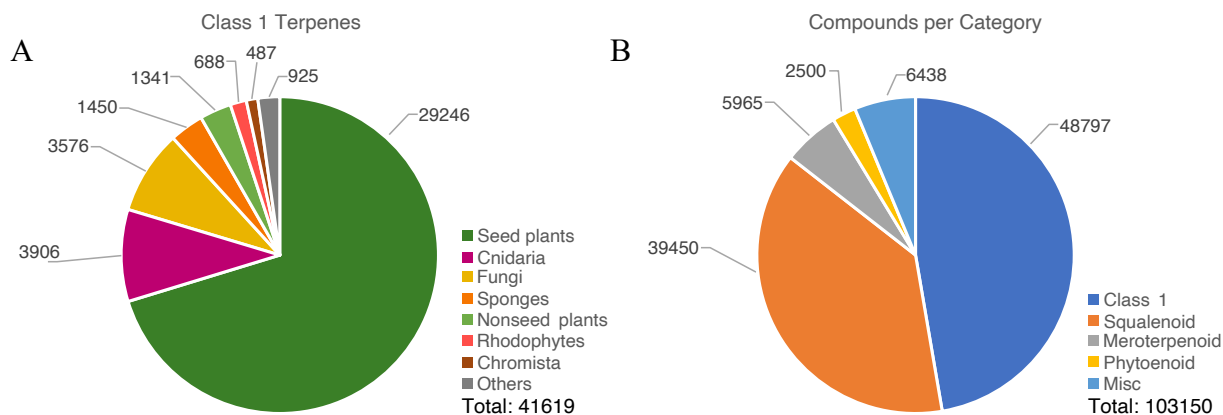


Figure 1.6.2. A) Counts of class 1 terpenoids by taxonomy B) Number of terpenoids per category.

catalytically inactive. Three domain enzymes consist of this alpha- and beta-domain assembly and contain an additional gamma domain, which does not participate in catalysis but most likely has a structural role in these proteins. While the amino acid sequences of class 1 TSs are often quite variable, the overall structure and key active site motifs are conserved. The dephosphorylation reaction is induced by a magnesium cluster, which is bound by highly conserved active site residues, namely the aspartate rich DDXXD motif and NSD/DTE motif (Fig. 1.6.1D). In microbial-type class 1 TSs, a highly conserved RY dimer near the C-terminus also plays a key role in substrate recognition, though this dimer is missing in plant TSs. Many TSs from bacteria, plants, fungi, coral, and sponges also have a conserved arginine pyrophosphate sensor which forms a hydrogen bond with the pyrophosphate moiety of the substrate (Fig. 1.6.1D).

Section 1.6.4. Terpenoid Distribution and Genome Mining

Despite overall low sequence similarity between class 1 TSs, the conservation of key active site motifs has made pattern-based genome mining efforts much more accessible. Conserved motifs are particularly amenable to discovery via a profile hidden Markov model (profile-HMM) (17), which is a statistical model and search method that ultimately relies on conserved residues or domains in a biological sequence to classify and annotate new TSs. With the advent of cheaper and more accessible sequencing, as well as huge growth in reference databases, we can readily access sequences to train profile-HMMs and discover TS sequences in previously unstudied organisms. For example, TSs were initially characterized in plants, bacteria, and fungi, and now, with the incredible growth of available sequencing data, in insects, algae, and marine animals. In a phylogenetic analysis (Fig. 1.6.3) these enzymes typically clade by phylum rather than by function, for example, octocoral terpene cyclases form a distinct monophyletic clade. Microbial-type class 1 TSs are not as clearly separated from one another, forming smaller subclades containing bacteria, fungal, amoeba, and a handful of microbial-like plant terpene cyclases within in a large polyphyletic group, possibly resulting from more recent horizontal gene transfer (HGT) events. Other microbial-type TSs are found in sponges and corals, but display a distinct monophyly, which implies a more ancient HGT event. Finally, recently discovered insect terpene cyclases form a clade within the larger IDS family, suggesting recent neofunctionalization of an IDS to a TS. The expansion and evolution of TSs across phyla is possible because of the ubiquitous nature of the underlying general biosynthetic machinery for terpenoid precursor supply. In this review we aim to highlight terpenoid diversity across different branches of the tree of life and discuss the apparent evolutionary success of underlying enzymatic principles. In this review, we use a taxonomy guided approach to highlight terpenoid diversity in secondary metabolism across

different branches of the tree of life. Taxa to be discussed are highlighted in the phylogeny of class 1 TSs (Fig. 1.6.3).

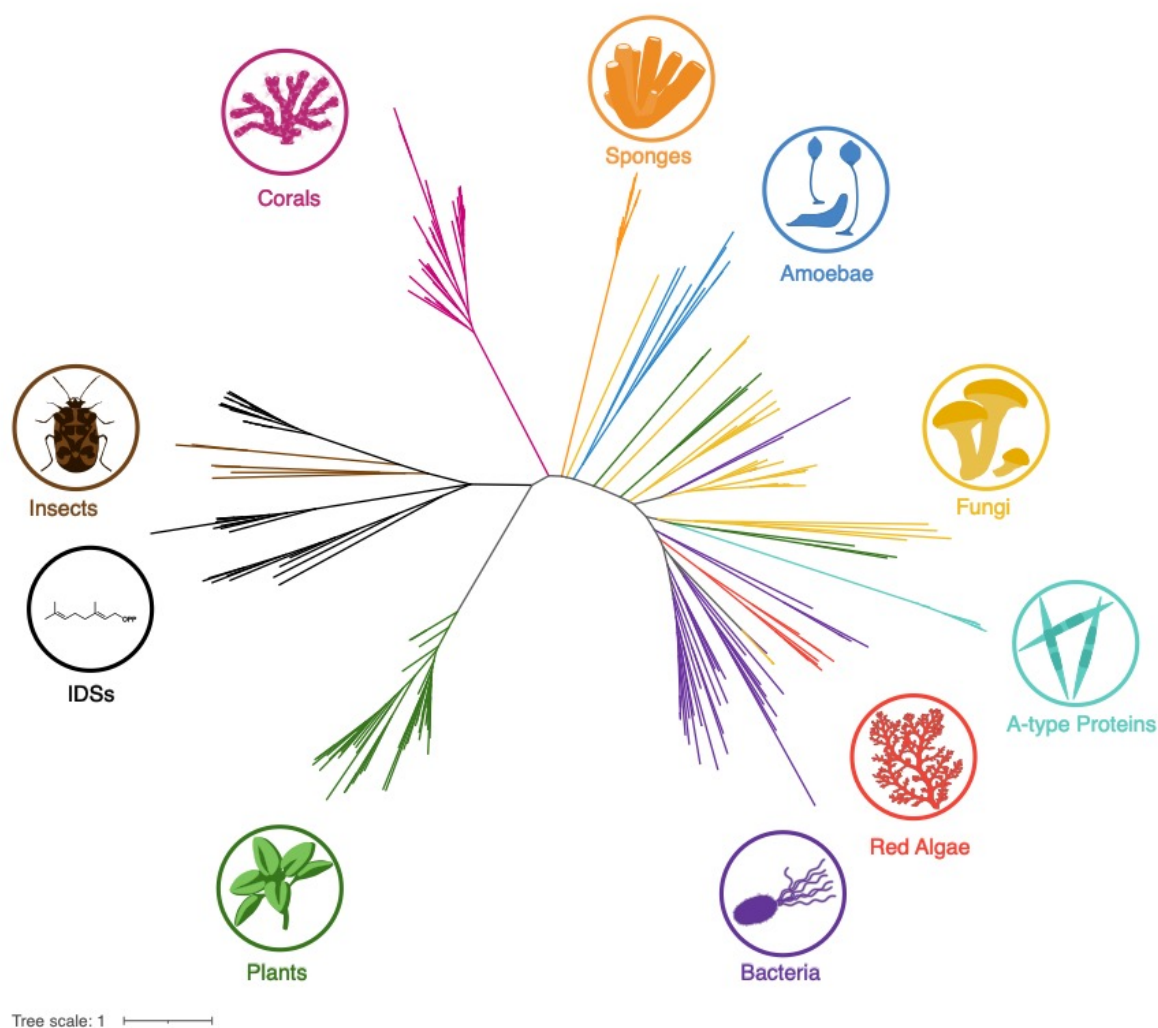


Figure 1.6.3. Phylogeny of class 1 terpene synthases

Section 1.6.5. Plants

Plants are by far the richest known source of terpenoid natural products, being the source of approximately 70% of all known structures (18). One reason for this could be seen in the high number of plant species (>300.000) (19), but also the extreme biosynthetic diversity within some

of these species, some coding for more than 50 TS genes in their genomes (20). Typical plant-type TSs form a monophyletic group (Fig. 1.6.3) and within this clade show more clustering by phylum than catalytic activity (3), reflective of the whole phylogenetic analysis of TSs. Plant TSs have a modular architecture and are either two-or three-domain enzymes, consisting of an alpha-domain, which is responsible for class 1 TS activity, a beta domain, which shows the fold of archetypical squalene-hopene cyclases harboring an active site for class 2 cyclase activity, and the gamma domain, which plays a structural role but does not participate in catalysis (16). Class 1 TS from plants can either have an alpha beta (21) or an alpha-beta-gamma structure (Fig. 1.6.4C)(22). While didomain enzymes are exclusively class 1 TSs, three-domain enzymes can be bifunctional (class 1 and class 2 activity) or monofunctional, where either the alpha- or beta-domain has lost their respective class 1 or class 2 activity. This fact is illustrated by the production of *ent*-kaurene (**9**), the terpene precursor for different important plant hormones regulating growth and germination, like the gibberellins in vascular plants(lycophytes, ferns, gymnosperms and angiosperms)(23) and mosses (24). **9** is synthesized in a two step reaction (Fig. 1.6.4A) in which first a class 2 cyclization takes place to form *ent*-copalyl diphosphate (**10**) and in the second step a class 1 cyclization converts **10** to *ent*-kaurene (**9**). While all plants seem to produce **9** the enzymes employed are different based in lineage. It is typical for vascular plants to employ two discrete, monofunctional three domain enzymes in which the **10**-synthase has a catalytically inactive alpha-domain and the **9**-synthase has an inactive beta-domain (Fig. 1.6.4B)(25). In mosses, which diverged relatively early in the evolutionary history of plants, one enzyme with an active beta and alpha-domain is responsible for the direct conversion of GGPP to **9** (26). The moss enzyme could be seen as a preservation of the ancestral state of an *ent*-kaurene synthase, while vascular plants, which diverged later, evolved to distinct enzymes, most likely by duplication and loss of function of the

unused domain respectively (3). The characterization of several three-domain TSs from non-vascular plants (mosses, liverworts and hornworts) followed by a phylogenetic analysis of the whole plant TS clade showed two early gene duplication events in the last common plant ancestor resulting in three principle clades (27). One of these contains three domain enzymes that are bifunctional *ent*-kaurene (**9**) synthases for non-vascular plants (mosses, hornworts, liverworts) and monofunctional class 2 *ent*-copalyl diphosphate (**10**) synthases in vascular plants, pointing to a bifunctional *ent*-aurene synthase as the ancestral plant TS (27).

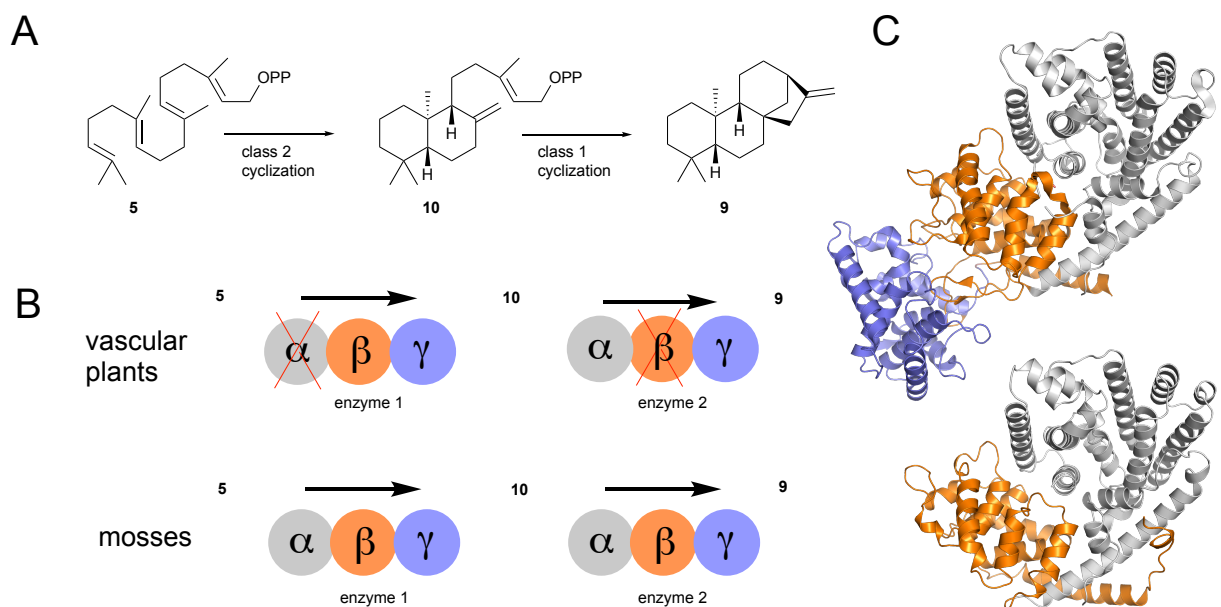


Figure 1.6.4: Function and architecture of plant TS enzymes. **A:** The two-step reaction of GGPP (**5**) to ent-kaurene (**9**), **B:** Different biosynthetic strategies for **9** production: vascular plants employ two monofunctional enzymes that only contain one active domain each, while mosses use one bifunctional TS with two catalytically active domains that act consecutively. **C:** representative crystal structures of three- and di-domain plant TS, taxadiene synthase (up, PDB code 3p5r), *epi*-aristolochene synthase (down, PDB code: 5eas). Color code: gray = alpha-domain, orange = beta-domain, blue = gamma-domain.

Duplication-neofunctionalization events similar to the diversification of *ent*-kaurene production have given rise to a large number of TSs that are involved in secondary metabolism and produce the vast number of structurally distinct terpene carbon scaffolds known from plants.

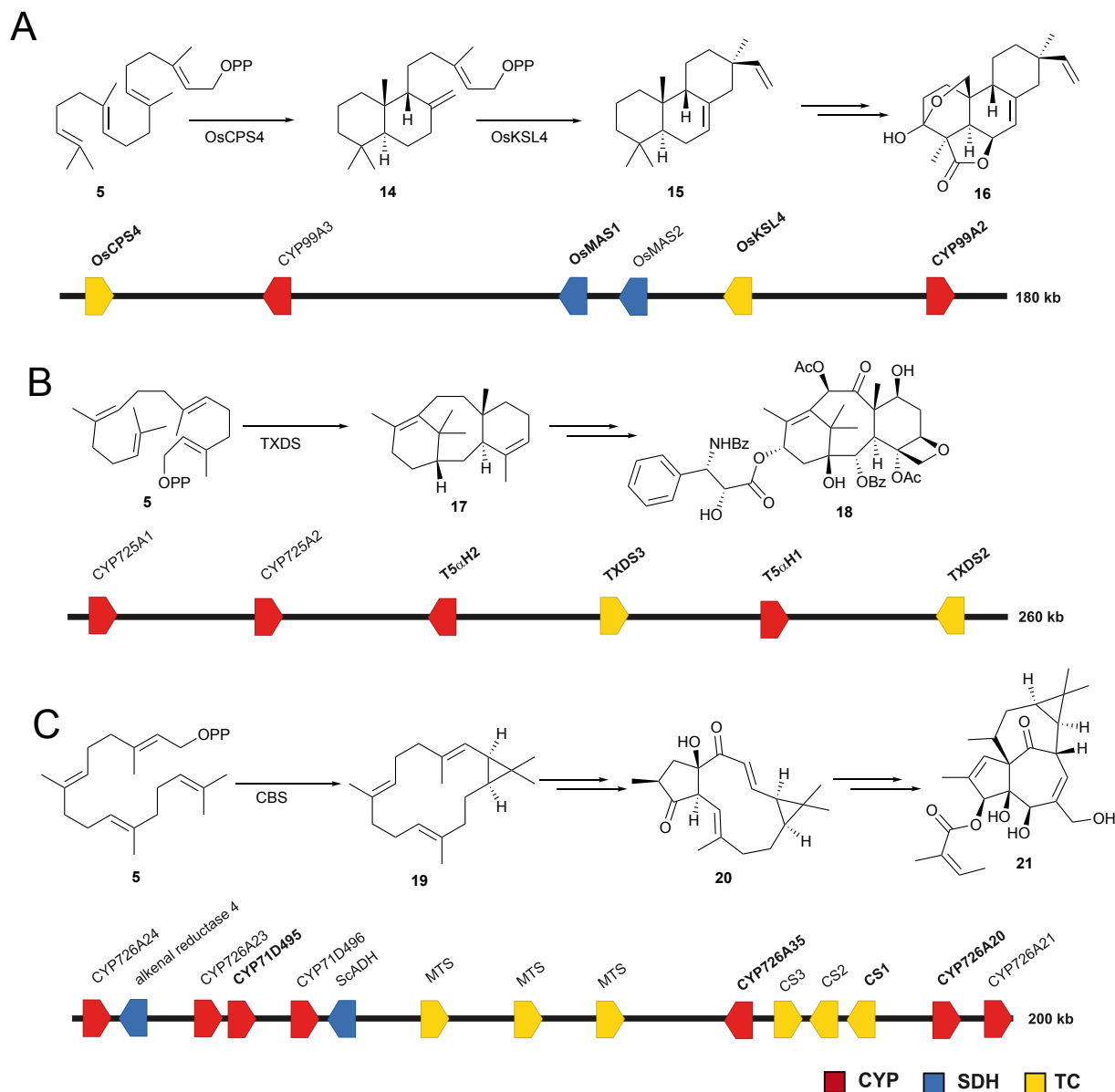


Figure 1.6.5. Diterpene-scaffold generation by tri-domain TSs and their connection with oxidized diterpenoids. **A:** syn-pimaradiene (**15**), the terpene precursor for momilactone B (**16**) is produced in two steps by a monofunctional class 2 and a monofunctional class 1 TS. All pathway enzymes are known, four of the seven genes are co-localized in a gene cluster in *O. sativa*. **B:** Taxadiene (**17**) is produced by taxadiene synthase, a monofunctional class 1 TS. Taxadiene synthase is colocalized with taxadiene-5-a-hydroxylase, a characterized CYP that oxidized **17**, on the genome of *Taxus chinensis*. **C:** A monofunctional class 1 TS converts GGPP to casbene (**19**), the precursor to a plethora of structurally diverse diterpenoids like ingenol (**21**). Casbene synthase is co-localized with enzymes responsible for the production of key intermediate jolkinol C (**20**) and further, uncharacterized CYP, SDH and TS genes on the genome of *Jatropha curcas*. Genes with names printed in bold have been biochemically verified.

One such example is represented in the momilactones from *Oryza sativa* (**28**). Momilactones are

secreted by *O. sativa* into the soil, where they function as allelochemicals exhibiting phytotoxicity (29, 30). carbon scaffold is assembled by a sequence like for the kaurene phytohormones, where the class 2 enzyme OsCPS4 produces *syn*-copalyl diphosphate (**14**) (31), which is converted by the class 1 cyclase OsKSL4 to form *syn*-pimaradiene (**15**) (32). *Syn*-pimaradiene is further functionalized by four cytochrome p450 monooxygenases (CYP) and one short chain dehydrogenase (SDH) resulting in the main phytotoxic compound momilactone B (**16**) (Figure 1.6.5A)(33). A part of the pathway is colocalized on *O. sativa* chromosome 4 as a biosynthetic gene cluster (34), while 3 of the CYP genes are located elsewhere in the genome (33). Intriguingly the moss *Calohypnum plumiforme* also produces momilactones via **15** and encodes for a gene cluster responsible for their production(35). The clusters from *C. plumiforme* and *O. sativa* are not homologous and the *syn*-pimaradiene synthase in *C. plumiforme* is a bifunctional class 2 / class 1 enzyme (36), which shows the same difference as *ent*-kaurene (**9**) production in mosses and vascular plants.

Monofunctional three-domain TSs also utilize linear precursors like GGPP to produce terpene carbon scaffold. Two examples are taxadiene (**17**) synthase from *Taxus* spp. (37) and casbene (**19**) synthase, which is widespread in *Euphorbiaceae* (38).

Taxadiene serves as the carbon scaffold in the taxane diterpenes including the anti-cancer medicine Taxol (38). While not all biochemical conversions from taxadiene (**17**) to taxol (**18**) are known to this date, many enzymes have been identified by transcriptomic approaches, including the second committed step in taxol biosynthesis the hydroxylation of taxadiene by taxadiene 5a-hydroxylase (T5aH)(39). Genome sequencing of two different species of *Taxus* showed that taxadiene synthase and T5aH are colocalized within larger gene clusters containing uncharacterized CYP genes (Fig. 1.6.5B)(10, 40).

Casbene (**19**) is the precursor to a wide array of structurally complex diterpenoids in the plant family of euphorbiaceae (41). The biosynthetic pathways to higher functionalized diterpenoids like ingenol mebutate (**21**) are unknown, but the enzymes for the generation of the bicyclic jolkinol C (**20**) are characterized (42). **19** is oxidized by three CYPs and cyclized by a short chain dehydrogenase (SDH) (42). These genes were shown to be co-localized in larger gene clusters in

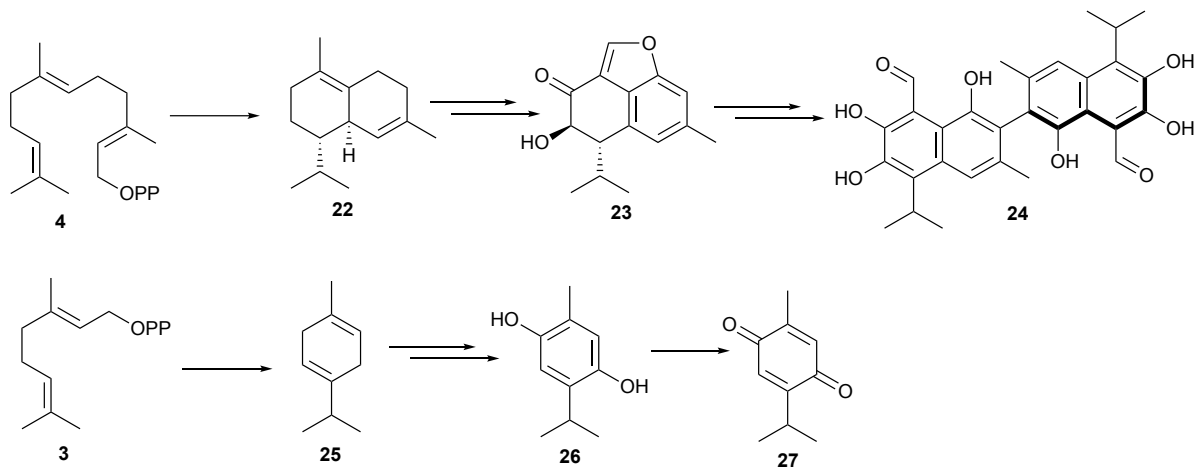


Figure 1.6.6. Sesqui- and monoterpene hydrocarbons and their connected hydrocarbons produced by sesqui- and mono-TS.

several species from the euphorbiaceae group including *Ricinus communis* (43), *Jatropha curcas* (44), and *Euphorbia peplus* (45). The last study identified two gene clusters containing casbene synthase and a multitude of CYPs and SDHs containing a total of 68 gene candidates. Gene silencing experiments determined jolkinol C as a branching point between the co-occurring groups of bicyclic jatrophone diterpenoids and tetracyclic ingenane diterpenoids like ingenol (45).

All discussed examples so far are diterpenoids, but the evolution of the putative ancestral bifunctional diterpene synthase has also brought forth a huge repertoire of monofunctional sesqui- and mono-TS enzymes, that use farnesyl diphosphate (FPP, **4**) and geranyl diphosphate (GPP, **3**) as substrates respectively (3).

The dimeric sesquiterpenoid gossypol (**24**) is a defense compound produced by cotton plants (*Gossypium* spp.) in the presence of phytopathogenic fungi (46). The first biosynthetic step is carried out by a didomain sesquiterpene synthase that converts **4** to delta cadinene (**22**) (47). Several biosynthetic steps towards the intermediate **23** are characterized and carried out by CYPs and SDHs (48). The final step is facilitated by a dirigent protein, that orients two achiral monomeric compounds so the biaryl coupling produces atroposelectively (+)-**24** (48).

Thymohydroquinone (**26**) is a characteristic flavor compound in thyme and oregano and produced via the monoterpene gamma-terpineol (**25**) which is oxidized to thymoquinone (**27**) in three steps carried out by two CYP and one SDH in *Thymus vulgaris* (49).

Due to their hydrophobicity and relatively small molecular mass **27** and pathway intermediates are volatile, some of them contributing to the characteristic odor of their producing plants (49). In fact a big portion of fragrant plant volatiles consists of mono and sesquiterpenes, often as unmodified hydrocarbons produced by TS enzymes (50). Volatile terpenes can act as direct or as indirect defense. The sesquiterpene hydrocarbon was shown to act as both, direct and indirect defense. Infestation of the maize by the larvae of the beetle *Diabrotica virgifera* leads to the belowground release of beta-caryophyllene through the roots, which attracts the entomopathogenic nematode *Heterorhabditis megidis* (51). In *Arabidopsis thaliana* infected by *Pseudomonas syringae* beta-caryophyllene production is strongly increased and inhibits microbial growth (52). Volatiles emitted by floral parts also act as attractants to pollinators (53). *Nicotiana suaveolens* is a night-blooming plant mainly pollinated by hawkmoths (54). Different studies found the concerted higher production of seven monoterpenes at night, the main compound being 1,8-cineole (55, 56). Investigation of the origins of these terpenes showed a single monoterpene synthase to be responsible for the production of all of these compounds, highlighting the ability of class 1 TSs to

produce a set of structurally different products, which simplifies regulation of a complex bouquet of volatiles together being responsible for a scent signal (54).

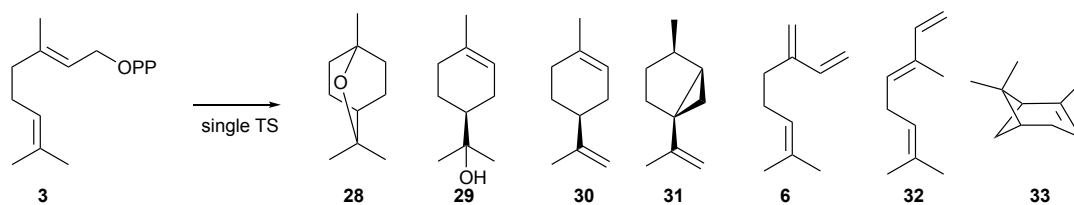


Figure 1.6.7. Several monoterpenes produced by a *Nicotiana suaveolens* mono-TS.

Besides the discussed two- and three-domain TSs, which were thought to exclusively responsible for all terpene TS activity in plants, several lineages of single-domain enzymes were discovered recently in the clubmoss *Selaginella moellendorffii*, which encodes 48 of these enzymes in its genome next to 18 classic two- and three- domain plant TSs (57). Since these monodomain TSs are primarily known from microbes, they were called microbial-type TSs and a later study found them to be abundant in almost all non-seed plants, but in no representative of a seed plant (58). Phylogenetic analyses showed them to cluster within actual bacterial and fungal TS sequences, which, additional to their domain structure points towards horizontal gene transfer from a microbe to non-seed plants. Since these microbial-type plant TSs are not forming a monophyletic clade, but rather occur as 4 monophyletic groups a potential horizontal gene transfer has occurred at least four times (59). The ecological role of these synthases is not known, but all characterized members produce mono- and sesquiterpenes, often complex mixtures, which suggests involvement in volatile production (58).

Section 1.6.6. Fungi

Recent advances in fungal genome sequencing, metabolomics, and the development of synthetic biology approaches to characterize fungal natural products have accelerated our

understanding of natural product biosynthesis in this kingdom. Fungal terpenoids make up almost 8% of known terpenoid natural products, and 32% of non-plant derived terpenoid natural products, making them the second largest contributor to this class of compounds(18)(Fig. 1.6.3). Terpenoid natural products in fungi are primarily known from the two major fungal divisions. Ascomycota, or filamentous fungi, which account for ~60% of described species, and Basidiomycota, including mushroom forming fungi, which account for ~30% of described species(60). Although there are many differences between the terpenomes of the two major fungal phyla, class 1 fungal TSs have been found to be broadly distributed in both phyla (60). A comparative analysis of the two phyla found terpenoids to be the major class of natural products in Basidiomycota, with approximately four times more TS sequences identified per genome than in publicly available Ascomycota genomes (60). Identified TS sequences consist of typical class 1 architecture, consisting of an alpha-domain subunits responsible for class 1 TS activity, as well as unique chimeric fungal bifunctional TSs, which harbor a prenyl transferase (PT) domain and a class 1 TS domain, forming a bifunctional alpha-alpha-domain assembly (20). Fungal terpene synthases are dispersed throughout the microbial-type TS clade (Fig. 1.6.3) forming a polyphyletic group with clustering according to both function and taxonomy. Notably, many fungal TSs are associated with biosynthetic gene clusters which facilitates genome mining and pathway engineering efforts (62). Fungal terpenoids derive from the mevalonate pathway and are typically sesquiterpenes (C15), diterpenes (C20), sesterterpenes (25), and triterpenes (C30), with no fungal monoterpene synthases known to date (60). Fungal terpenoids are functionally diverse and play numerous ecological roles ranging from defense and attraction, to mediating complex interkingdom chemical interactions, such as host plant and endophytic fungi signaling (63). Additionally, many fungal terpenoids are highly bioactive with anticancer, antimicrobial, and antiviral activities. For example, the illudin

family of sesquiterpenoids are effective against several tumor cell-lines, including metastatic prostate cancers (64) and the highly toxic trichothecene mycotoxins have a significant impact on global cereal crops and are known to contaminate livestock feed (65).

The first characterized class 1 fungal TS comes from the Ascomycota plant pathogen *Fusarium sporotrichioides* and was found to catalyze the formation of trichodiene (**34**), which is the first committed step towards the trichothecene mycotoxins (Fig. 1.6.8A). Detailed mechanistic characterization via site directed mutagenesis developed a cyclization cascade proposal which was later rationalized by a crystal structure (66) resulting in a structure-based model of the trichodiene synthase reaction. One of the most interesting observations from this early work is the startling conservation of the terpene synthase fold, which, much like in other organisms, consists of an alpha helical bundle with relatively low sequence similarity to other fungal TS sequences. Fungal TSs were found to maintain the two conserved aspartate-rich motifs, DDXXD/E and NSE/DTE (66) (Fig. 1.6.8B). Trichodiene (**34**) is then modified in *Fusarium* strains into the highly toxic trichothecenes, which share a common tetracyclic ring and includes T-2 toxin (**35**) produced by *F. langsethiae*, *F. poae*, and *F. sporotrichioides* (67, 68)(gene cluster AF359360.3)

Some of the most well-known terpenoid natural products in Basidiomycota belong to the illudin family of sesquiterpenoids, which derive from the protoilludane scaffold (Fig. 1.6.8C)(69). Early isolation and structural characterization work identified *Omphalotus olearius*, or the Jack O'Latern mushroom, as a major producer of bioactive illudins (64). Follow-up studies report the draft genome of *O. olearius* strain VT-653.13 and identified 11 TS sequences (Omp1-10, with 5a and 5b) (70). Biochemical characterization of each TS found Omp6 and 7 (62% AA % identity) to be highly specific Δ -6 protoilludene synthases. Both TSs are located within biosynthetic gene clusters, with Omp6 located within a 25 kb, 18 gene cluster that is proposed to produce illudin

compounds (Fig. 1.6.8C). Omp7, is located within a smaller cluster, with only one P450 and a FAD binding protein. Using newly identified TS sequences as genetic hooks, genome mining of 40 publicly available Basidiomycota genomes revealed 531 putative TS sequences. Phylogenetic analysis highlights five putative functional clades, which were supported with biochemical validation of Fompi1 from *Fomitopsis pinicola*. This enzyme was validated as an α -cuprenene sesquiterpene synthase which matched the predicted product according to the growing Basidiomycota TS phylogeny.

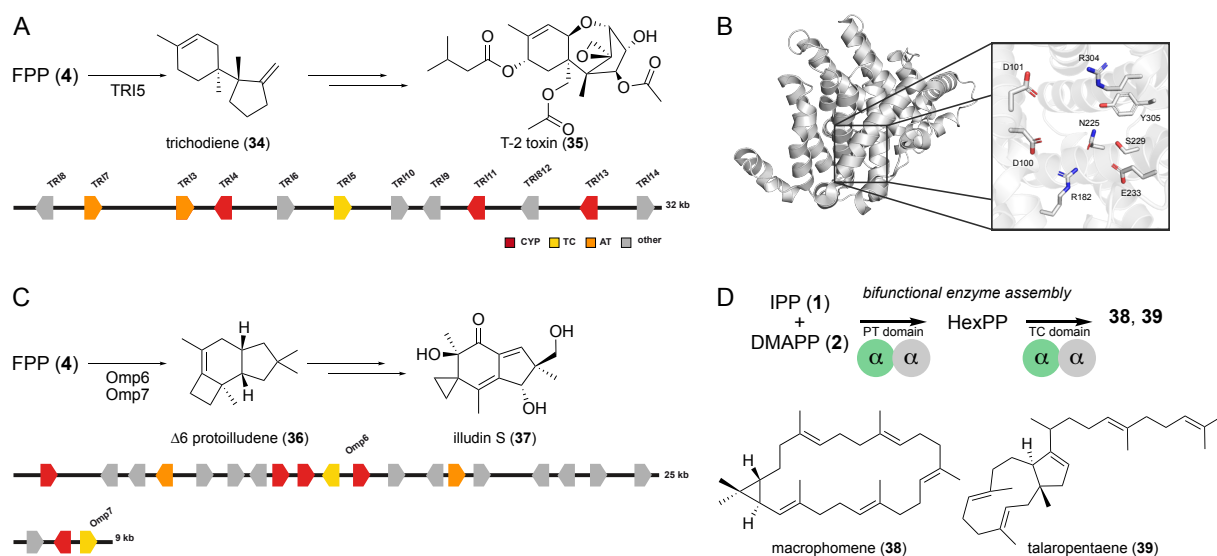


Figure 1.6.8. Fungal TS scaffold generation. **A.** Trichodiene (**34**), the terpene precursor for T-2 toxin (**35**), and the producing gene cluster from *Fusarium sporotrichioides*. **B.** Monomer of trichodiene synthase and highlighted active site residues. **C.** $\Delta 6$ protoilludene (**36**), the terpene precursor for the illuden family of sesquiterpenes, including the antitumor agent, illudin S (**37**). **D.** Bifunctional triterpenoid synthase with, talaropentaene (**38**) and macrophomene (**39**)

Chimeric TSs were initially characterized in fusicoccin biosynthesis. Thea tricyclic hydrocarbon precursor for the fusicoccins is biosynthesized by a fusicoccadiene synthase, which contains both prenyltransferase and terpene cyclase activities. Notably, this enzyme is encoded as a part of a larger gene cluster (71). Bifunctional TSs are also common in sesterterpene biosynthesis in fungi, where catalytically independent domains work in tandem to generate C25 sesterterpene

products (72). Recently, a new family of fungal chimeric class 1 triterpene synthases were discovered and functionally characterized (73). Typically, triterpenoids are biosynthesized via the dimerization of FPP into squalene by class 2 TSs known as squalene synthases. Subsequent cyclizations are catalyzed by triterpene synthases, such as squalene hopene cyclases (SHC), resulting in products like pentacyclic hopene, the precursor of hopanoids. A recent study by Tao and coauthors, reports the discovery and biochemical characterization of two non-squalene-dependent triterpene synthases TvTS and MpMS from *Talaromyces verruculosus* and *Macrophomina phaseolina*, respectively. Much like chimeric fungal sesterterpene synthases, this new family of class 1 triterpene synthases harbor both PT and TS domains (Fig. 1.6.8D). Each PT domain was found to first convert DMAPP and IPP into hexaprenyl diphosphate or HexPP (C30), which is subsequently cyclized by the TS domain into the final triterpenoid product. Expression of full-length TvTS and MpMS in *Saccharomyces cerevisiae* resulted in two different triterpenoid products, talaropentaene (38) and macrophomene (39), respectively (Fig. 1.6.8D). Notably, the 22-membered ring in macrophomene (39) represents the largest macrocycle terpenoid known to-date and only structurally mirrored by the 22-membered rings in nostocyclophanes, a class of oxidatively dimerized polyketides from cyanobacteria (74). Additional protein structure-supported genome mining using AlphaFold2-predicted models and the crystal structure of TvTS-TC identified six new sequences with putative triterpenoid activity. Biochemical characterization of putative triterpenoid bifunctional synthases resulted in the discovery of CgCS from the host *Colletotrichum gloeosporioides*. CgCS is structural similar to TvTS and makes a related product, colleterpenol. Overall, this work identifies a new enzymatic mechanism for the biosynthesis of triterpenoids and expands the product boundaries for class 1 TSs.

Section 1.6.7. Bacteria

While the group has expanded in recent years, bacterial terpenoids still make up less than 0.5% of all terpenoid natural products and approximately 2% of all non-plant terpenoid natural products (6)(Fig. 1.6.2). While many bacteria encode stand-alone terpene synthases, a growing number of bacterial terpene synthases are situated within biosynthetic gene clusters that encode functionalized molecules like albaflavenone (75), pentalenolactone (76), cyclooctatin (**41**) (77), and (+)-kolavenol (78). While bacterial terpenoids like albaflavenone and ent-kaurene (79) are the result of class 1 and class 2 cyclizations, these TSs are usually separate monofunctional enzymes which are clustered together, unlike the bifunctional alpha-beta-gamma domain plant TSs that produce similar terpenoids. Bacterial terpenoids, many of which are released as volatiles, are important ecological mediators of microbe-microbe interactions. These molecules play a particularly large role in soil ecosystems, where volatile terpenes have been shown to mediate interactions between bacteria, plants, fungi, and other microbes across long distances (80). Bacterial terpenes such as albaflavenone and B-pinene are known to inhibit other bacteria and fungi (81–83), while bacterial d-cadinene and B-linalool can stimulate motility in protists (83).

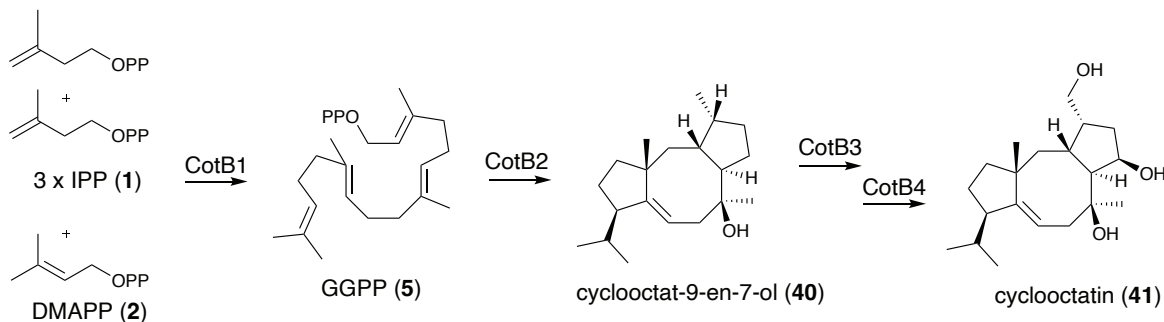


Figure 1.6.9. Cyclooctatin biosynthesis. GGPP synthase CotB1 catalyzes the production of GGPP (**5**) from IPP (**1**) and DMAPP (**2**). Terpene synthase CotB2 produces cyclooctat-9-en-7-ol (**40**), which is converted to cyclooctatin (**41**) by two CYP450s, CotB3 and CotB4.

Several bacterial terpene synthases have been extensively characterized and a thorough summary of this work can be found elsewhere (84). Most of the TSs characterized from bacteria are sesquiterpene synthases, though a handful of monoterpene and diterpene synthases have also been characterized. Perhaps the best studied bacterial TS is the diterpene synthase CotB2, which forms the characteristic 5-8-5 fused ring system of the potent lysophospholipase inhibitor cyclooctatin (**41**)(Fig. 1.6.9). The cyclooctatin gene cluster consists of a GGPP synthase (CotB1), the CotB2 terpene synthase, and two CYP450s (CotB3 and CotB4). CotB2 has been the subject of numerous structural and mechanistic studies, which has yielded unprecedented insight into the inner workings of this enzyme (85–87). Several crystal structures of CotB2 have revealed that following the binding of GGPP and 3 Mg²⁺ ions, the C terminus “lid” of the enzyme closes on the active site, transitioning the enzyme from an open to a closed state. Several mutagenesis studies have shown that single point mutations to active site residues change the product scope drastically from fusicocanes to dolabellanes or cembranoids. This makes CotB2 a promising candidate for biocatalysis since small mutations provide access to a wide range of scaffolds with antibiotic (88), anticancer (89), and antiparasitic activity (90, 91).

Geosmin (**43**) and the related compound 2-methylisoborneol (2-MIB)(**45**)(Fig. 1.6.10) are widespread bacterial terpenoids responsible for the earthy odor of soil. While humans and many animals can sense these compounds (92), their ecological role is still poorly understood (93). Geosmin (**43**) and 2-MIB (**45**) were recently reported as predation deterrents for *Caenorhabditis elegans* (94). Geosmin (**43**) also has a clear effect on certain insects including fruit flies, in which geosmin impacts feeding, oviposition, and flight (95, 96), and mosquitos, in which geosmin acts as an attractant for oviposition sites (97). Becher et al also showed that geosmin (**43**) and 2-MIB (**45**) mediate symbiosis between *Streptomyces* and springtails, by acting as feeding attractants for

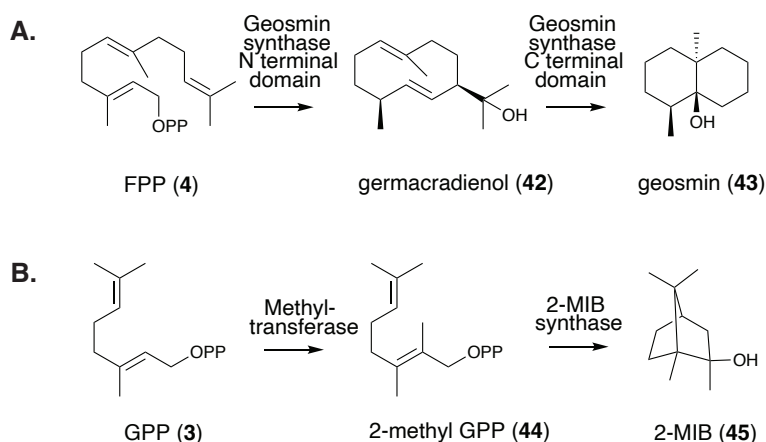


Figure 1.6.10A .Geosmin (**43**) and B. 2-methylisoborneol (2-MIB)(**45**) biosynthesis.

springtails, which facilitates *Streptomyces* spore dispersal through fecal pellets. It was recently reported that relatively low levels (50-5000ng/L) of geosmin (**43**) disrupted several processes involved in embryonic development in zebrafish (98).

Geosmin (**43**) is biosynthesized by an unusual didomain enzyme. The N terminal domain of geosmin synthase converts FPP to germacradienol, while the C terminal domain converts germacrandienol (**42**) to geosmin (**43**). Both the N terminal and C terminal domains contain the usual type I TS active site moieties including the aspartate-rich DDXXD, the NSE triad, and the RY dimer. In some cases, the N terminal domain features an additional NSE triad approximately 38 amino acids downstream of the canonical NSE triad, but mutagenesis of the downstream NSE motif in *Streptomyces coelicolor* geosmin synthase did not significantly affect product composition or enzyme kinetics, suggesting this additional NSE may not play a role in catalysis (99).

2-MIB (**45**) is the product of a 2 gene biosynthetic gene cluster consisting of a SAM-dependent methyl transferase and a monoterpene synthase. The methyl transferase starts by

methyating GPP (**3**) at the C2 position to form (E)-2-methyl GPP (**44**), the substrate of the monoterpene synthase. Similar to many other monoterpene synthases, which isomerize GPP to linalyl diphosphate (LPP), 2-MIB synthase isomerizes 2-methyl GPP (**44**) to 2-methyl LPP in the first step of catalysis, before catalyzing the rest of the cyclization to 2-MIB (**45**). Many 2-MIB synthases found across several bacterial species are unusually large bacterial terpene synthases, with approximately 400-500 amino acid residues. Most 2-MIB synthases features a proline-rich (>25% proline) 13kDa N-terminal domain which is not seen in other terpene synthases, however this domain is entirely disordered and its function is unknown (100). Crystal structures of 2MIB synthase from *Streptomyces coelicolor* reveal an “incompletely closed” active site and an unusual proline-rich domain of unknown function, which is only found in 2-MIB synthase homologues, but not other terpene synthases (100).

Section 1.6.8. Algae

Algae, Chlorophytes, Rhodophytes, and for the purpose of this review, Chromista, contribute ~2% of total known terpenoids and ~19% of non-plant derived terpenoid natural products. Of the known terpenoids from algae, 48% are expected to be produced by class 1 TS enzymology. Notable features include oxidative scaffold tailoring with oxygen and halogens. The incorporation of halogens, like bromine and chlorine, is a hallmark of marine chemistry and rarely seen in terrestrial compounds. Red macroalgae, or seaweeds, are particularly gifted producers of halogenated terpenoids, or haloterpenoids. Algal haloterpenoids are hypothesized to play several ecological roles, such as, deterring herbivory (101) and preventing the formation of biofilms and biofouling via quorum sensing disruption (102). The biosynthesis of haloterpenoids has even been suggested to play a role in reactive oxygen species (ROS) scavenging (103).

Some red algae are known to accumulate haloterpenoids in specialized refractile inclusion bodies, or corps en cerise. Additionally, there is evidence for the subcellular compartmentalization of terpenoid biosynthetic machinery in these unique cell structures. Paradas et al describe a ‘mevalonosome’, which is a vacuolar organelle containing enzymes involved in the mevalonate pathway in *Plocamium Brasiliense* (104). Cytochemical analysis supports the presence of active 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) in the *P. brasiliense* mevalonosome, a key enzyme of isoprenoid biosynthesis in the mevalonate pathway. Further work to establish the subcellular localization of other terpenoid enzymes is needed.

The first biochemically characterized macroalgal sesquiterpene synthase comes from the red alga, *Laurencia pacifica* (105). Kersten et al. identified and functionally characterized a bourbonane-producing sesquiterpene synthase, LphTS-A, using transcriptomics and heterologous expression in yeast (Fig. 1.6.11A). LphTS-A is one of three terpene synthases identified in the transcriptome of *L. pacifica*. To resolve the absolute stereochemistry of prespatane (47), the major bourbonane sesquiterpene product of LphTS-A, Kersten et al used a microgram-scale nuclear magnetic resonance-coupled crystalline sponge X-ray diffraction analysis. Three more TS transcripts were identified from *Laurencia dendroidea* from a publicly available dataset in the NCBI Sequence Read Archive (SRA). Phylogenetic analysis revealed a well-supported monophyletic clade comprising the six red algal sequences, separate from sesquiterpene synthases from other domains of life. Interestingly, this clade is nested between bacterial and fungal TS sequences, suggesting a recent horizontal gene transfer event, possibly from a microbial source.

Systematic sequence analysis of publicly available red algal genomes and transcriptomes by Wei et al further expands the small but growing number of algal TSs (7). In addition to the six already known sequences from *Laurencia* spp, three new TSs were identified from *Erythrolobus*

australicus and *Porphyridium purpureum*. No typical plant TS sequences were found in any of the algal genomes analyzed. In-depth genomic analysis of *P. purpureum* validated red algal TS sequences are not from microbial contamination and are part of the nuclear genome. Upstream and downstream neighboring genes relative to the TS sequence were determined to be most closely related to eukaryotic genes. Notably, the assembled region containing the TS sequence had high

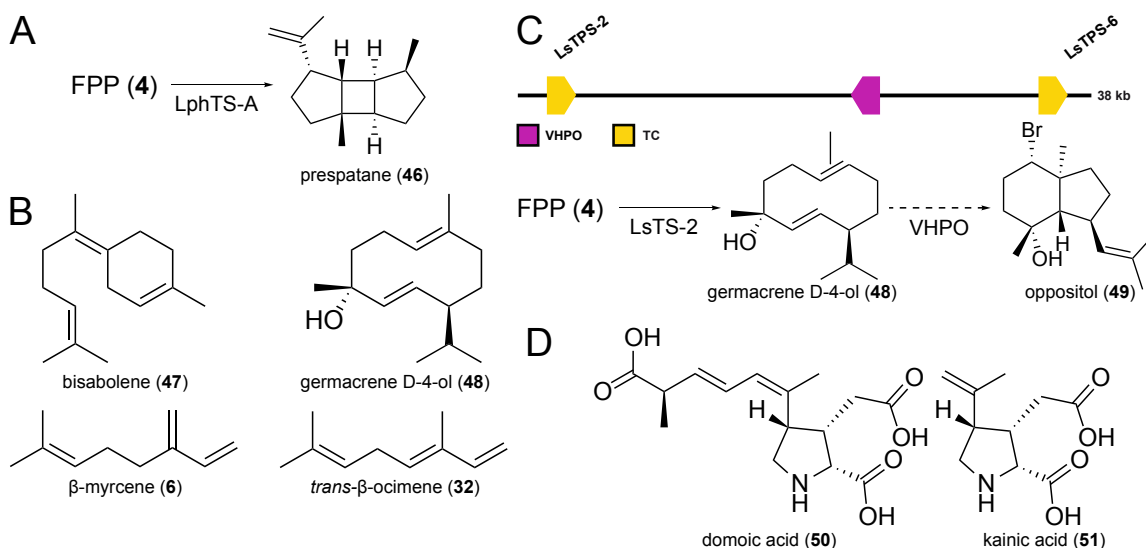


Figure 1.6.11. Algal TS scaffold generation. **A.** Prespatane (46), the product of LphTS-A **B.** Key haloterpenoid scaffolds, myrcene (6), trans- β -ocimene (32), bisabolene (47), and germacrene D-4-ol (48) **C.** Proposed pathway to oppositol (49) the putative product of the haloterpenoid associated gene cluster from *Laurencia subopposita*. **D.** Structures of domoic acid (50) and kainic acid (51)

contiguous coverage, which provides confidence in assembly accuracy and assignment of neighboring genes.

Work by Steele et al reports ten new red algal TSs from the key haloterpenoid producing red algae *Portieria hornemannii*, *Plocamium pacificum*, and *Laurencia subopposita* (in preparation). This work supports the existence of microbial-type class 1 TSs, through whole genome sequencing and phylogenetic analysis. In all four species and the authors demonstrate a clear delineation between algal genes and those from associated microbes. *In vitro* reconstitution

and biochemical characterization revealed the selective production of key haloterpenoid precursors, including the well-known plant terpenes myrcene (**6**), *trans*- β -ocimene (**32**), bisabolene (**47**), and germacrene D-4-ol (**48**) described for the first time from algal enzymes (Fig. 1.6.11B). As sequencing data of algae improves, we are rapidly expanding our understanding of gene clustering in red seaweeds. This is exemplified by the first known haloterpenoid associated

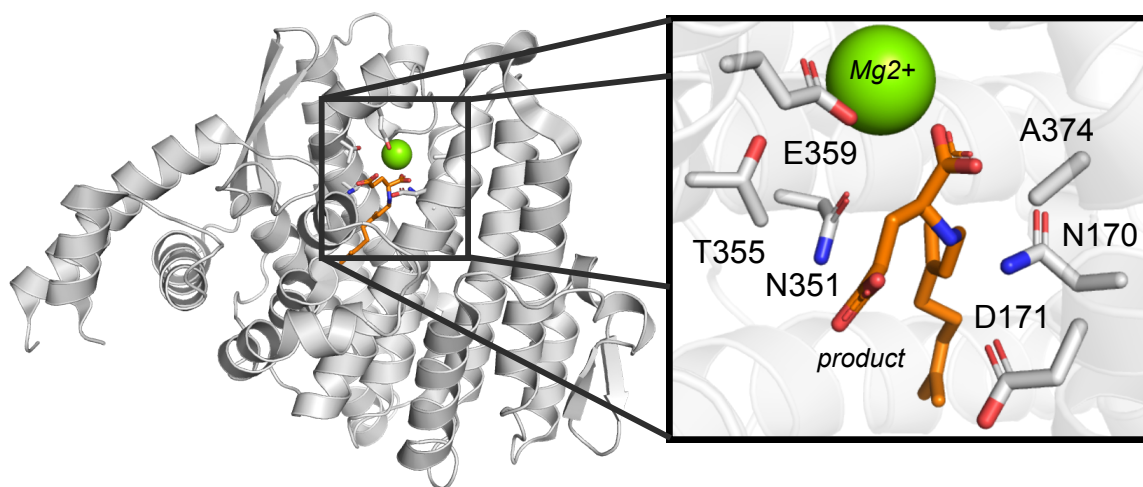


Figure 1.6.12. Crystal structure of DabA (6VL1). Panel highlights modified active site residues DDXXD is NDXXA and NSE is NTE. Orange is the DabA product *N*-geranyl-L-glutamate.

gene cluster in *L. subopposita*, where the sequence lsTPS-2 was found to be co-clustered with a vanadium-dependent haloperoxidases, the predicted enzyme responsible for the halogenation of algal haloterpenoids (Fig. 1.6.11C). The expansion of the red algal TS clade by Steele et al supports our understanding of terpenoid biosynthesis in algae, however further studies of downstream red algal tailoring reactions and brown macroalgae are still required to fill gaps in algal terpenoid chemistry.

Kainoids are a family of potent marine neurotoxins produced by a variety of marine and terrestrial organisms. Oceanic kainoid production is primarily associated with the harmful algal bloom-forming diatom genus *Pseudonitzschia*, however, domoic acid (DA) (**50**) and kainic acid (KA)(**51**) are produced by both macro- and micro- algae. DA (**50**) and KA (**51**) both share a tri-substituted pyrrolidine ring, the hallmark of the kainoid family, and differ at the C4 substituent in prenyl chain length (Fig. 1.6.11D). Early stable isotope experiments suggested marine kainoids are composed of glutamic acid and either DMAPP or GPP. Work by Brunson et al. identified the four-gene cluster responsible for DA (**50**) production using transcriptome and genome sequencing (106). Initially, one of the gene candidates, *dabA*, was annotated as a terpene cyclase, and structural prediction suggested a terpene cyclase-like fold. However, DabA was found to repurpose TS enzymology to catalyze an intermolecular *N*-prenylation reaction of a primary amine, rather than the canonical intramolecular cyclization reaction. Follow-up work by Chekan et. al. identified homologous biosynthetic genes using whole genome sequencing of the red seaweeds *Digenea simplex* and *Palmaria palmata* (107). Here they identified the DabA homolog, KabA, which also catalyzes the *N*-prenylation of l-Glu but accepts DMAPP rather than GPP, to form the lower isoprene homolog of DA, KA. A later study by Chekan et al. reports the 2.1 Å crystal structure of DabA (108). Like most terpene cyclases, DabA consists of a mostly alpha-helical core, however DabA has several key differences from canonical terpene cyclases. First, DabA contains N- and C- terminal extensions and an internal insertion, which connect to the main alpha-helical core via a beta-hairpin extension. Next, DabA has an elongated hydrophobic tunnel which accommodates the unique binding of the pyrophosphate substrate. Similarly, DabA has distinct active site binding

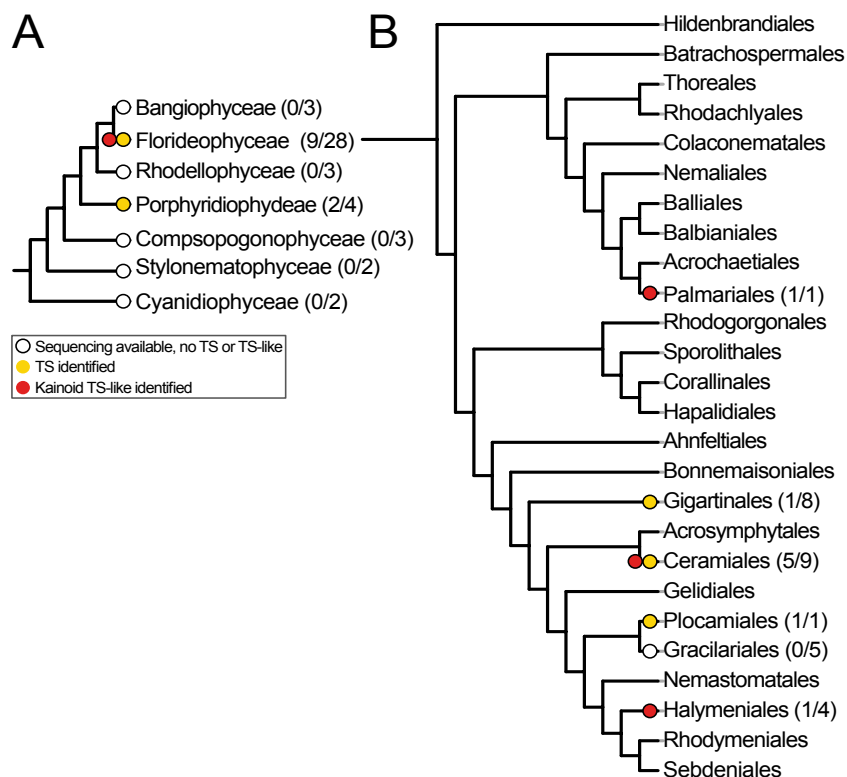


Figure 1.6.13. Cladogram of red algae by Order (A) by Class (B) within Florideophyceae. Blue circles denote sequencing data is available, but no TS or TS-like sequences have been identified. Yellow circles denote a typical microbial-type TS has been identified in the group. Red circles denote a TS-like modifications that remodel the canonical magnesium (Mg^{2+})-binding motif found in terpene cyclases and enable the intermolecular prenylation reaction, where the canonical DDXXD motif is now NDXXA and NSE is NTE (Fig. 1.6.12). The authors also identify key residues that dictate prenyl group chain length specificity, and thus, catalytically differentiates DabA and KabA.

Both *dabA* and *kabA* have remarkably low sequence similarity to any publicly available protein sequence and thus serve as a key genetic hook to query for kainoid biosynthetic genes. This is demonstrated in work by Steele and Brunson et al. where the authors identify the *rad* biosynthetic gene cluster and biochemically characterize RadA, which despite having higher amino acid sequence similarity to the red algal kainic acid biosynthesis enzyme KabA, catalytically behaves more like the homologous diatom enzyme DabA (109). The consistent

monophyly of the A proteins from these distantly related taxa suggests a possible horizontal gene transfer event. Given the limited distribution of kainoid biosynthetic enzymes in modern diatom genera and red macroalgae, as well as the lack of clear homology of the glutamate *N*-prenyltransferase (RadA, KabA, and DabA) to known algal proteins suggests HGT from a potential third host as a mechanism for the acquisition of these kainoid biosynthetic enzymes.

In summary, algal TS and repurposed TS-like sequences are not widespread across red algae (Fig. 1.6.13). With the majority of known TS and TS-like sequences found within the class Florideophyceae. Both canonical terpenoid and kainoid producing sequences clade within the broader microbial-type class 1 family of TSs. This in addition to the sparse distribution of TS and TS-like sequences implies acquisition from an unknown microbial source, possibly via HGT. Algal terpenoids are typically highly functionalized and there remains a gap in knowledge of the downstream tailoring reactions, particularly in haloterpenoids. Similarly, no brown or green algal TS sequences are yet to be identified, opening further questions around TS evolution and function in these distinct algal lineages.

Section 1.6.9. Insects

In the chemical ecology of insects, terpenoids play important roles in defense and communication. For example (6R,7S)-himachala-9,11-diene (**54**) in the crucifer flea beetle *Phyllotreta striolata*, is a major male-produced aggregation pheromone (110). Terpenoid biosynthesis in insects is an instance of striking chemical convergence. While insects contain typical IDS and mevalonate pathway enzymes, no homologs of plant or microbial-type TS genes have been found in the genomes of insects. Interestingly, many insect terpenoids are similar or even identical to plants, with many putative downstream oxidative tailoring reactions yet to be characterized. For example, (*E*)- β -caryophyllene is a sesquiterpene produced by plants such as

maize, cannabis, and *Arabidopsis* (111) and in insects is produced by the Asian lady beetle (112). Despite this shared chemistry, it is unlikely terpenoid biosynthesis arose in insects due to a common ancestral biosynthetic pathway with plants or other kingdoms, which highlights the fundamental question of whether insect terpenoids are synthesized *de novo*, derived from plant precursors, or produced by symbionts. Comprehensive analysis of insect genomes and transcriptomes suggests TS evolution occurred via gene duplication and neofunctionalization of insect IDS sequences. Interestingly, in several instances where high quality genome data is available, insect IDS-like TS sequences are found to be co-clustered with typical insect IDS sequences (4). In the class of Insecta, all insect lineages sampled (Blattodea, Hemiptera, Phthiraptera, Hymenoptera, Coleoptera, Lepidoptera, and Diptera), except for the Phthiraptera (parasitic lice), contain characterized or predicted TS proteins (113).

Early work by Martin et al. utilized cell-free assays of the terpenoid producing pine engraver beetle, *Ips pini* (114). Cell-free assays using [1-3H] GPP demonstrated GPP is converted to the acyclic monoterpene myrcene (**6**) which is a precursor to the *I. pini* aggregation pheromone ipsdienol (**52**) (Fig. 1.6.14A). Historically, the biosynthesis of the aggregation pheromone components in *I. pini* had thought to occur only through hydroxylation of monoterpene olefins, such as myrcene (**6**) or α -pinene (**33**), derived from their conifer hosts (115). However, the sex-specific and inducible enzymatic activity in extracts of male *I. pini* shifted the focus from myrcene (**6**) as a host-provided precursor to myrcene (**6**) as an insect-derived biosynthetic intermediate (116). Subsequent work successfully identified a geranyl diphosphate synthase (GPPS) with myrcene synthase activity in *I. pini* (117). Recombinant enzyme produced both GPP and myrcene.

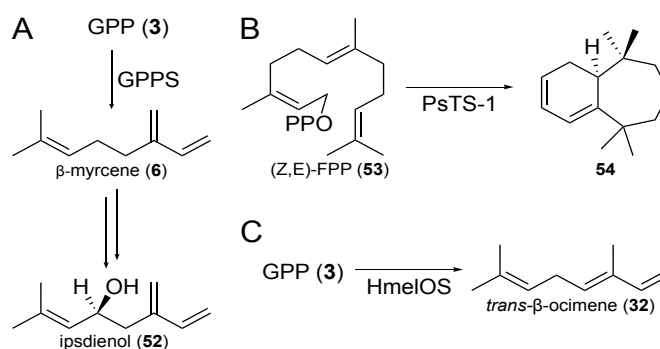


Figure 1.6.14: Insect terpenoid scaffolds **A.** Myrcene (**6**) conversion to ipsdienol (**52**) in *Ips pini* **B.** Conversion of (Z,E)-FPP (**53**) to (6R,7S)-himachala-9,11-diene (**54**) by PsTPS1 in *Phyllotreta striolata* **C.** Production of (E)- β -ocimene (**32**) in *Heliconius melpomene* by HmelOS.

Taken together this is the first evidence of genome encoded insect TS enzymatic machinery and the first example of a monoterpene synthase in the Metazoa.

More recent studies utilize transcriptome and genome sequencing to greatly expanded our understanding of TS enzymes in insects. Beran et al. investigate the biosynthesis of sesquiterpene aggregation pheromones in the flea beetle, *Phyllotreta striolata*, an important pest of crucifer crops in North America and Southeast Asia (118). Here, the authors describe an emerging class of new, evolutionarily distinct, TS enzymes in *P. striolata*. Initially, no homologs to typical plant or microbial terpene synthases were identified, however, *P. striolata* was found to express a remarkably high number of trans-IDS sequences. From nine different IDS-type transcripts in the *P. striolata* transcriptome, two were found to encode bona fide trans- or cis-IDS enzymes, while four transcripts encode TSs, suggesting an evolutionary origin of these enzymes from IDS progenitors. Using a combination of cell-free assays, transcriptomics, and heterologous expression, at least four different IDS-like TS enzymes were successfully identified and biochemically characterized, where the authors found PsTPS1 to convert (Z,E)-FPP (**53**) to (6R,7S)-himachala-9,11-diene (**54**) (Fig. 1.6.14B) and other sesquiterpenes observed in beetle extracts. Within the *P. striolata* transcriptome, two expressed enzymes had genuine IDS activity,

with PsIDS1 synthesizing (E,E)-FPP (**4**), whereas PsIDS3 produced neryl diphosphate, (Z,Z)-FPP, and (Z,E)-FPP (**53**). Phylogenetic analysis of the new IDS-like TS enzymes, including several related species from the order Coleoptera, supports the evolutionary conversion of IDS to TS function. Interestingly, exon–intron structures of IDS and TS genes in *P. striolata* are conserved, further supporting this TS gene family evolved from IDS ancestors.

Works by Lancaster et al, identify and characterize IDS-like TS enzymes from the Southern green stink bug, *Nezara viridula* and the harlequin bug, *Murgantia histrionica*. Phylogenetic analysis suggests that insect TS activity might have emerged more recently from IDS proteins and this event has occurred independently in these and other insect taxa. Mining of brown marmorated stink bug, *Halyomorpha halys* genome and transcriptome revealed a family of six IDS-like TS genes (**4**). This new family of IDS-like TS sequences was found to be organized in two separate clusters, each of which most likely emerged by gene duplication, where putative TS sequences were found to be co-clustered with typical insect IDS sequences. This model of identifying candidate TS enzymes by searching for IDS gene expansion was successful in the search for an (E)- β -ocimene synthases in the butterfly genus *Heliconius*. Darragh et al. investigate the recent divergence of two butterfly species, *Heliconius melpomene* and *Heliconius cydno*, which differ in the presence and absence of (E)- β -ocimene (**32**)(119). Using linkage mapping, gene expression, and functional analyses, *H. melpomene* was found to have two new TS genes, one of which HmelOS, can produce (E)- β -ocimene (**32**) (Fig. 1.6.14C) Furthermore, the *Heliconius sp.* TS enzymes are unrelated to previously described Coleopteran and Hemipteran sequences, supporting their independent evolutionary origins.

Although insect TS sequences are phylogenetically distant, the shared chemistry across kingdoms once again highlights the utility of terpenoids as an expansion of primary metabolic pathways across domains of life.

Section 1.6.10. Amoeba.

Amoeba are microeukaryotes that produce a variety of mono-, sesqui-, and diterpenoids (120, 121). Dictyostelids, a subgroup of amoebae known for their unique life cycle consisting of both unicellular and multicellular phases, emit a complex mixture of terpenoids during the multicellular phases of their life cycles. The production of those dictyostelid terpenoids has been linked to several recently characterized terpene synthases, most of which are stand-alone terpene synthases, however a terpene synthase-CYP450 gene cluster was recently reported from an amoeba (122)(Fig. 1.6.15). The first amoebal terpene synthases, DdTS1-9, were reported from *Dictyostelium discoideum* by Chen et al. in 2016 (123). Heterologously expressed DdTS1-9 produced a variety of sesquiterpenes, including those found in the headspace extracts of *D. discoideum*, such as (*E,E*)- α -farnesene, protoillud-7-ene, (3*S*)-(+)-asterisca-2(9),6-diene, and (-)-b-barbatene. (121, 123, 124). Interestingly, the expression of these terpene synthase genes in *D. discoideum* was linked to the amoeba's stages of development, with the highest expression during

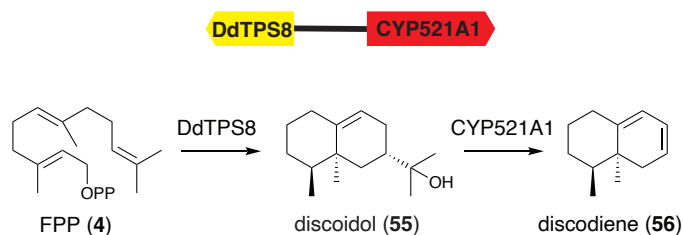


Figure 1.6.15. Discodiene biosynthesis in *Dictyostelium discoideum*. FPP (**4**) is converted to discoidol (**55**) by the class 1 TS DdTPS8, which is then converted to discodiene (**56**) by the CYP450 CYP521A1.

the latter half of the multicellular stage (123). Further work revealed the presence of terpene synthases in other dictyostelid species, including *Dictyostelium purpureum*, *Polysphondylium pallidum*, and *Dictyostelium lacteum*.(121, 125).

While little is known about the ecological role of terpenes in amoeba, (*E,E*)- α -farnesene has been shown to have insect deterrent activity in plants (126), which supports the hypothesis proposed by Chen et al. 2016 that (*E,E*)- α -farnesene could protect amoeba from predation by nematodes (123). It's also hypothesized that the production of β -barbatene in the fruiting bodies stage of the amoebal life cycle could indicate its role in attracting insects for spore dispersal, an activity that has been documented for β -barbatene in fruiting body-producing fungi (127). It was recently reported that some of these dictyostelid terpene synthases cluster with cytochrome P450 enzymes, the first reported case of biosynthetic gene clustering in amoeba (122). DdTS8, which produces the sesquiterpene alcohol discoidol, clusters with CYP521A1, a cytochrome P450 that acts on discoidol (**55**) to produce discodiene (**56**), which is also found in the headspace of *D. discoideum* (122)(Fig. 1.6.15). Similar TS-P450 clustering was also seen with three other *D. discoideum* TSs, as well as three TSs from *D. purpureum*, though the products of those clusters were not determined. These amoebal clusters add to the growing list of terpenoid-encoding clusters from non-plant eukaryotes.

Section 1.6.11. Sponges

Marine sponges are the most prolific source of marine natural products to date, with more than 9400 reported sponge molecules on CMNPD (128). Sponges contain a rich array of terpenoids, including bioactive terpenoids such as the anti-malarial molecule kalihinol A, the anti-

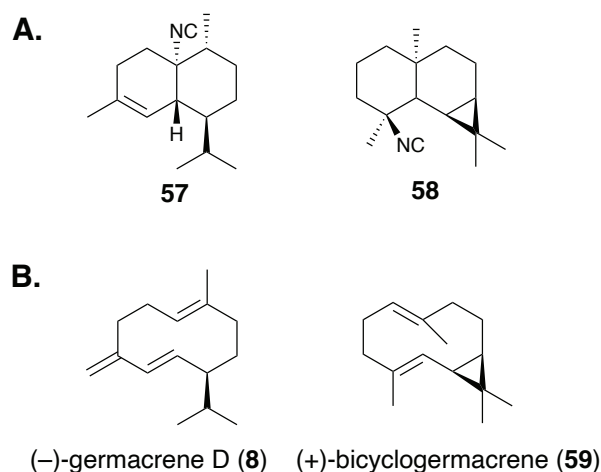


Figure 1.6.16. Sponge terpenoids. **A.** Sponge nitrogenous terpenes. **B.** Products of sponge terpene synthases.

tumor compound avarol, the anti-inflammatory agent manoalide. Approximately 4300 sponge natural products are isoprenoids, comprising nearly half of all known sponge derived molecules. The majority of sponge isoprenoids likely originate from class 2 TSs, such as steroids and meroterpenoids. Scaffolds that originate from class 1 TSs comprise a smaller portion of sponge terpenoids, with approximately 1450 reported class 1 sponge terpenoids (Fig. 1.6.2). Many of the sponge class 1 terpenoids feature nitrogen-containing functional groups like isonitriles, isocyanates, and isothiocyantes. These molecules, also known as nitrogenous terpenes, are only found in marine sponges and the nudibranchs that feed on them.

Very little is known about the biosynthesis of sponge derived terpenoids, however recently, a novel group of TSs were reported for the first time from marine sponges (9)(Fig. 1.6.16). These TSs were reported from multiple species of marine sponges, several of which are reported to produce nitrogenous terpenes. They contain the usual DDXXD and RY motifs, as well as a slightly unusual NGD moiety and they clade distinctly from other known TSs from other domains of life.

The genes that encode these TSs do not appear to cluster with other biosynthetic genes. Interestingly, these TSs were encoded by the sponge animal itself, which helped to shed light on a question that had long puzzled marine natural products scientists: Are sponge derived terpenoids produced by the sponge animal or its microbiome? Marine sponges feature a rich microbiome, which has made it difficult to determine whether a molecule derived from a sponge originates from the sponge animal host or a microbial symbiont. It has been well established that sponge associated microbes are responsible for the production of sponge derived molecules from other natural product classes, like PKSs, NRPSs, and PDBEs. Early cell sorting experiments, however, have found that terpenoids are often stored in the sponge animal host cells, which further supports the importance of the sponge host animal in the production of terpenoids (129–132).

Section 1.6.12. Discussion

Terpenoids persist across taxonomic kingdoms, geographic locations, and environmental niches because life is uniquely primed to evolve and maintain specialized terpenoid biosynthesis. Primary isoprenoid biosynthesis pathways are ubiquitous throughout nature, providing the necessary precursors for secondary terpenoid biosynthesis. Similarly, the plasticity of terpene synthases raises the possibility that an inherent lack of substrate specificity in some TSs may play a role in their evolution. Additionally, enzymes, like CYP450s, often implicated in terpenoid tailoring, are widely distributed and often maintain enough substrate promiscuity to act on a variety of terpenoid products and facilitate the evolution of new pathways. Together, this suggests the cost of evolving a new terpenoid biosynthesis pathway is relatively low as the precursor production and tailoring machinery is already included. All that is required is to acquire a terpene synthase,

either through horizontal gene transfer or neofunctionalization of an existing IDS, to produce new specialized terpenoids.

The small size (300-500aa) and lack of specialized redox partners or complex cofactors make TSs excellent candidates for horizontal gene transfer and easy incorporation into an organism's metabolism, driving the emergence of new terpenoid pathways. As a group, TSs have a rugged evolutionary landscape and high plasticity, meaning small mutations often affect large changes to product promiscuity. Aside from a few conserved active site motifs, TSs typically have a high degree of sequence variability and small mutations can greatly affect product production. Coupled with the potent bioactivity of terpenoid products, positive evolutionary pressure to maintain and expand bioactive and ecologically relevant terpenoids can push a TS towards new product specificity. This high plasticity of TSs allows for many different sequence possibilities to produce a functional TS.

Terpenoids play a crucial role in mediating interspecies interactions. For example, terpenoid chemoattractants facilitate spore and gamete dispersal and antibacterial terpenoids that protect plants from pathogens (133–135). These ecological responses are triggered by terpenoids activating receptors, which in turn activate various signaling pathways. Terpenoids are well documented to activate olfactory receptors in insects, mammals, and other animals, which may orchestrate processes like development, growth, and reproduction. Recent research suggests that nuclear receptors, a group of ligand-activated transcription factors involved in reproduction and development across all metazoan groups, evolved from ancient terpene synthases and retained the same ligand-binding domain core of an ancestral terpene synthase shared with a metazoan ancestor (136). These nuclear receptors are basal for metazoans, indicating that the evolutionary history of

metazoans is closely linked to that of isoprenoids. Isoprenoid precursors are ubiquitous across all animals, as are the receptors that sense isoprenoids.

Terpene synthases can be acquired by organisms through two routes: neofunctionalization and horizontal gene transfer. In insects, neofunctionalization is the dominant route, as their terpene synthases closely resemble their own native IDSs, more than those of other bacterial or fungal terpene synthases. IDSs and TSs share an alpha helical fold and similar mechanisms, indicating that they may have a shared common ancestor. Horizontal gene transfer is currently the most common mode of TS evolution, as seen in fungi (137), red algae (108, 109), and plants (59). The more distinct groups of two- and three-domain plant terpene synthases, coral (8), and sponge (9) terpene synthases are less clear, but due to the structural similarity to microbial TS enzymes an ancient horizontal gene transfer event in both lineages is a possibility. The polyphyletic nature of the microbial-type TSs clade also supports the hypothesis of dynamic horizontal gene transfer events throughout these phyla. Similarly, in red algae, phylogenetic analysis of the TS-like prenyltransferases (A proteins) suggests a horizontal gene transfer event happened recently.

As we sequence a greater number and diversity of species, we continue to discover new specialized terpene biosynthesis genes across distinct evolutionary lineages. Many recent terpene synthase discoveries have come from non-model organisms like corals, sponges, algae, and insects. As such, there is an ongoing need to sequence unusual organisms and to deposit the resulting sequences into publicly available databases, such as NCBI and ENA. Additionally, bioinformatic toolsets need to be developed with terpene synthase discovery in mind. Unlike many other natural product biosynthesis genes such as PKSs or NRPSs, which have high sequence homology and are highly predictable by homology-based analysis tools, terpene synthases are often missed by standard genome mining programs. It is quite likely novel terpene synthases are

waiting to be discovered in organisms that have yet to be sequenced. As shown in the discoveries of insect, coral, and sponge TSs, all of which were initially missed by traditional enzyme discovery tools. To continue to discover and expand our understanding of TS enzymology and evolution, a combinatorial approach of sequencing, homology-independent search tools, and biochemical validation is needed to discover truly novel terpene synthases.

Section 1.6.13. References

1. M. G. Chevrette, K. Gutiérrez-García, N. Selem-Mojica, C. Aguilar-Martínez, A. Yañez-Olvera, H. E. Ramos-Aboites, P. A. Hoskisson, F. Barona-Gómez, Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat. Prod. Rep.* **37**, 566–599 (2020).
2. E. Pichersky, R. A. Raguso, Why do plants produce so many terpenoid compounds? *New Phytol.* **220**, 692–702 (2018).
3. F. Chen, D. Tholl, J. Bohlmann, E. Pichersky, The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J* **66**, 212–229 (2011).
4. D. Tholl, Z. Rebholz, A. V. Morozov, P. E. O’Maille, Terpene synthases and pathways in animals: enzymology and structural evolution in the biosynthesis of volatile infochemicals. *Nat. Prod. Rep.* (2023) <https://doi.org/10.1039/D2NP00076H> (April 13, 2023).
5. M. B. Quin, C. M. Flynn, C. Schmidt-Dannert, Traversing the fungal terpenome. *Nat. Prod. Rep.* **31**, 1449–1473 (2014).
6. J. D. Rudolf, T. A. Alsup, B. Xu, Z. Li, Bacterial Terpenome. *Natural Product Reports* **23**, 9–10 (2006).
7. G. Wei, Q. Jia, X. Chen, T. G. Köllner, D. Bhattacharya, G. K.-S. Wong, J. Gershenzon, F. Chen, Terpene Biosynthesis in Red Algae Is Catalyzed by Microbial Type But Not Typical Plant Terpene Synthases. *Plant Physiol* **179**, 382–390 (2019).
8. I. Burkhardt, T. de Rond, P. Y.-T. Chen, B. S. Moore, Ancient plant-like terpene biosynthesis in corals. *Nat Chem Biol* **18**, 664–669 (2022).
9. K. Wilson, T. de Rond, I. Burkhardt, T. S. Steele, R. J. B. Schäfer, S. Podell, E. E. Allen, B. S. Moore, Terpene biosynthesis in marine sponge animals. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2220934120 (2023).

10. X. Xiong, J. Gou, Q. Liao, Y. Li, Q. Zhou, G. Bi, C. Li, R. Du, X. Wang, T. Sun, L. Guo, H. Liang, P. Lu, Y. Wu, Z. Zhang, D.-K. Ro, Y. Shang, S. Huang, J. Yan, The *Taxus* genome provides insights into paclitaxel biosynthesis. *Nat. Plants* **7**, 1026–1036 (2021).
11. C. J. Paddon, J. D. Keasling, Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat Rev Microbiol* **12**, 355–367 (2014).
12. J. Lombard, D. Moreira, Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol* **28**, 87–99 (2011).
13. G. Flesch, M. Rohmer, Prokaryotic hopanoids: the biosynthesis of the bacteriohopane skeleton. *Eur. J. Biochem.* **175**, 405–411 (1988).
14. A. L. Jaffe, C. J. Castelle, P. B. Matheus Carnevali, S. Gribaldo, J. F. Banfield, The rise of diversity in metabolic platforms across the Candidate Phyla Radiation. *BMC Biology* **18**, 69 (2020).
15. Y. Hoshino, E. A. Gaucher, On the Origin of Isoprenoid Biosynthesis. *Molecular Biology and Evolution* **35**, 2185–2197 (2018).
16. R. Cao, Y. Zhang, F. M. Mann, C. Huang, D. Mukkamala, M. P. Hudock, M. E. Mead, S. Prisic, K. Wang, F.-Y. Lin, T.-K. Chang, R. J. Peters, E. Oldfield, Diterpene cyclases and the nature of the isoprene fold. *Proteins: Structure, Function, and Bioinformatics* **78**, 2417–2432 (2010).
17. B.-J. Yoon, Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr Genomics* **10**, 402–415 (2009).
18. , Dictionary of Natural Products (April 13, 2023).
19. M. J. M. Christenhusz, J. W. Byng, The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201–217 (2016).
20. K. D. Allen, K. McKernan, C. Pauli, J. Roe, A. Torres, R. Gaudino, Genomic characterization of the complete terpene synthase gene family from *Cannabis sativa*. *PLOS ONE* **14**, e0222363 (2019).
21. C. M. Starks, K. Back, J. Chappell, J. P. Noel, Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase. *Science* **277**, 1815–1820 (1997).
22. M. Köksal, Y. Jin, R. M. Coates, R. Croteau, D. W. Christianson, Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* **469**, 116–120 (2011).
23. P. Hedden, The Current Status of Research on Gibberellin Biosynthesis. *Plant Cell Physiol* **61**, 1832–1849 (2020).

24. C. Schwechheimer, Gibberellin Signaling in Plants – The Extended Version. *Front Plant Sci* **2** (2012).
25. C. I. Keeling, H. K. Dullat, M. Yuen, S. G. Ralph, S. Jancsik, J. Bohlmann, Identification and Functional Characterization of Monofunctional ent-Copalyl Diphosphate and ent-Kaurene Synthases in White Spruce Reveal Different Patterns for Diterpene Synthase Evolution for Primary and Secondary Metabolism in Gymnosperms. *Plant Physiol.* **152**, 1197–1208 (2010).
26. K. Hayashi, H. Kawaide, M. Notomi, Y. Sakigi, A. Matsuo, H. Nozaki, Identification and functional analysis of bifunctional ent-kaurene synthase from the moss *Physcomitrella patens*. *FEBS Letters* **580**, 6175–6181 (2006).
27. Q. Jia, R. Brown, T. G. Köllner, J. Fu, X. Chen, G. K.-S. Wong, J. Gershenzon, R. J. Peters, F. Chen, Origin and early evolution of the plant terpene synthase family. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2100361119 (2022).
28. T. Kato, C. Kabuto, N. Sasaki, M. Tsunagawa, H. Aizawa, K. Fujita, Y. Kato, Y. Kitahara, N. Takahashi, Momilactones, growth inhibitors from rice, *oryza sativa* L. *Tetrahedron Letters* **14**, 3861–3864 (1973).
29. T. Toyomasu, M. Usui, C. Sugawara, K. Otomo, Y. Hirose, A. Miyao, H. Hirochika, K. Okada, T. Shimizu, J. Koga, M. Hasegawa, M. Chuba, Y. Kawana, M. Kuroda, E. Minami, W. Mitsuhashi, H. Yamane, Reverse-genetic approach to verify physiological roles of rice phytoalexins: characterization of a knockdown mutant of OsCPS4 phytoalexin biosynthetic gene in rice. *Physiologia Plantarum* **150**, 55–62 (2014).
30. H. Kato-Noguchi, M. Hasegawa, T. Ino, K. Ota, H. Kujime, Contribution of momilactone A and B to rice allelopathy. *J Plant Physiol.* **167**, 787–791 (2010).
31. K. Otomo, Y. Kanno, A. Motegi, H. Kenmoku, H. Yamane, W. Mitshuhachi, H. Oikawa, H. Toshima, H. Itoh, M. Matsuoko, T. Sassa, T. Toyomasu, Diterpene Cyclases Responsible for the Biosynthesis of Phytoalexins, Momilactones A, B, and Oryzaalexins A–F in Rice. *Biosci. Biotechnol. Biochem.* **68**, 2001–2006 (2004).
32. P. R. Wilderman, M. Xu, Y. Jin, R. M. Coates, R. J. Peters, Identification of Syn-Pimara-7,15-Diene Synthase Reveals Functional Clustering of Terpene Synthases Involved in Rice Phytoalexin/Allelochemical Biosynthesis. *Plant Physiol.* **135**, 2098–2105 (2004).
33. R. De La Peña, E. S. Sattely, Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. *Nat Chem Biol* **17**, 205–212 (2021).
34. K. Shimura, A. Okada, K. Okada, Y. Jikumaru, K.-W. Ko, T. Toyomasu, T. Sassa, M. Hasegawa, O. Kodama, N. Shibuya, J. Koga, H. Nojiri, H. Yamane, Identification of a Biosynthetic Gene Cluster in Rice for Momilactones. *J Biol Chem* **282**, 34013–34018 (2007).

35. L. Mao, H. Kawaide, T. Higuchi, M. Chen, K. Miyamoto, Y. Hirata, H. Kimura, S. Miyazaki, M. Teruya, K. Fujiwara, K. Tomita, H. Yamane, K. Hayashi, H. Nojiri, L. Jia, J. Qiu, C. Ye, M. P. Timko, L. Fan, K. Okada, Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12472–12480 (2020).
36. K. Okada, H. Kawaide, K. Miyamoto, S. Miyazaki, R. Kainuma, H. Kimura, K. Fujiwara, M. Natsume, H. Nojiri, M. Nakajima, H. Yamane, Y. Hatano, H. Nozaki, K. Hayashi, HpDTC1, a Stress-Inducible Bifunctional Diterpene Cyclase Involved in Momilactone Biosynthesis, Functions in Chemical Defence in the Moss *Hypnum plumaeforme*. *Sci Rep* **6**, 25316 (2016).
37. M. R. Wildung, R. Croteau, A cDNA Clone for Taxadiene Synthase, the Diterpene Cyclase That Catalyzes the Committed Step of Taxol Biosynthesis. *Journal of Biological Chemistry* **271**, 9201–9204 (1996).
38. M. T. Dueber, W. Adolf, C. A. West, Biosynthesis of the Diterpene Phytoalexin Casbene: Partial Purification and Characterization of Casbene Synthetase from *Ricinis communis* 1. *Plant Physiol.* **62**, 598–603 (1978).
39. S. Jennewein, R. M. Long, R. M. Williams, R. Croteau, Cytochrome P450 Taxadiene 5 α -Hydroxylase, a Mechanistically Unusual Monooxygenase Catalyzing the First Oxygenation Step of Taxol Biosynthesis. *Chem. Biol.* **11**, 379–387 (2004).
40. J. Cheng, X. Wang, X. Liu, X. Zhu, Z. Li, H. Chu, Q. Wang, Q. Lou, B. Cai, Y. Yang, X. Lu, K. Peng, D. Liu, Y. Liu, L. Lu, H. Liu, T. Yang, Q. Ge, C. Shi, G. Liu, Z. Dong, X. Xu, W. Wang, H. Jiang, Y. Ma, Chromosome-level genome of Himalayan yew provides insights into the origin and evolution of the paclitaxel biosynthetic pathway. *Molecular Plant* **14**, 1199–1209 (2021).
41. M. Fattahian, M. Ghanadian, Z. Ali, I. A. Khan, Jatrophone and rearranged jatrophone-type diterpenes: biogenesis, structure, isolation, biological activity and SARs (1984–2019). *Phytochem Rev* **19**, 265–336 (2020).
42. D. Luo, R. Callari, B. Hamberger, S. G. Wubshet, M. T. Nielsen, J. Andersen-Ranberg, B. M. Hallström, F. Cozzi, H. Heider, B. Lindberg Møller, D. Staerk, B. Hamberger, Oxidation and cyclization of casbene in the biosynthesis of Euphorbia factors from mature seeds of *Euphorbia lathyris* L. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5082–E5089 (2016).
43. A. J. King, G. D. Brown, A. D. Gilday, T. R. Larson, I. A. Graham, Production of Bioactive Diterpenoids in the Euphorbiaceae Depends on Evolutionarily Conserved Gene Clusters. *The Plant Cell* **26**, 3286–3298 (2014).
44. A. J. King, G. D. Brown, A. D. Gilday, E. Forestier, T. R. Larson, I. A. Graham, A Cytochrome P450-Mediated Intramolecular Carbon–Carbon Ring Closure in the Biosynthesis of Multidrug-Resistance-Reversing Lathyrane Diterpenoids. *ChemBioChem* **17**, 1593–1597 (2016).

45. T. Czechowski, E. Forestier, S. H. Swamidatta, A. D. Gilday, A. Cording, T. R. Larson, D. Harvey, Y. Li, Z. He, A. J. King, G. D. Brown, I. A. Graham, Gene discovery and virus-induced gene silencing reveal branched pathways to major classes of bioactive diterpenoids in *Euphorbia peplus*. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2203890119 (2022).
46. P. Heinsteins, Stimulation of Sesquiterpene Aldehyde Formation in *Gossypium arboreum* Cell Suspension Cultures by Conidia of *Verticillium dahliae*. *J. Nat. Prod.* **48**, 907–915 (1985).
47. A. A. Bell, R. D. Stipanovic, C. R. Howell, P. A. Fryxell, Antimicrobial terpenoids of *Gossypium*: Hemigossypol, 6-methoxyhemigossypol and 6-deoxyhemigossypol. *Phytochem* **14**, 225–231 (1975).
48. X. Tian, J.-X. Ruan, J.-Q. Huang, C.-Q. Yang, X. Fang, Z.-W. Chen, H. Hong, L.-J. Wang, Y.-B. Mao, S. Lu, T.-Z. Zhang, X.-Y. Chen, Characterization of gossypol biosynthetic pathway. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5410–E5418 (2018).
49. S. T. Krause, P. Liao, C. Crocoll, B. Boachon, C. Förster, F. Leidecker, N. Wiese, D. Zhao, J. C. Wood, C. R. Buell, J. Gershenzon, N. Dudareva, J. Degenhardt, The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2110092118 (2021).
50. M. Rosenkranz, Y. Chen, P. Zhu, A. C. Vlot, Volatile terpenes – mediators of plant-to-plant communication. *The Plant Journal* **108**, 617–631 (2021).
51. S. Rasmann, T. G. Köllner, J. Degenhardt, I. Hiltbold, S. Toepfer, U. Kuhlmann, J. Gershenzon, T. C. J. Turlings, Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature* **434**, 732–737 (2005).
52. M. Huang, A. M. Sanchez-Moreiras, C. Abel, R. Sohrabi, S. Lee, J. Gershenzon, D. Tholl, The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (E)- β -caryophyllene, is a defense against a bacterial pathogen. *New Phytol.* **193**, 997–1008 (2012).
53. J. T. Knudsen, R. Eriksson, J. Gershenzon, B. Ståhl, Diversity and distribution of floral scent. *Bot. Rev.* **72**, 1 (2006).
54. S. Roeder, A.-M. Hartmann, U. Effmert, B. Piechulla, Regulation of simultaneous synthesis of floral scent terpenoids by the 1,8-cineole synthase of *Nicotiana suaveolens*. *Plant Mol Biol* **65**, 107–124 (2007).
55. J. H. Loughrin, T. R. Hamilton-Kemp, H. R. Burton, R. A. Andersen, Effect of diurnal sampling on the headspace composition of detached *Nicotiana suaveolens* flowers. *Phytochemistry* **32**, 1417–1419 (1993).

56. R. A. Raguso, R. A. Levin, S. E. Foose, M. W. Holmberg, L. A. McDade, Fragrance chemistry, nocturnal rhythms and pollination “syndromes” in *Nicotiana*. *Phytochemistry* **63**, 265–284 (2003).
57. G. Li, T. G. Köllner, Y. Yin, Y. Jiang, H. Chen, Y. Xu, J. Gershenzon, E. Pichersky, F. Chen, Nonseed plant *Selaginella moellendorffii* has both seed plant and microbial types of terpene synthases. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14711–14715 (2012).
58. Q. Jia, G. Li, T. G. Köllner, J. Fu, X. Chen, W. Xiong, B. J. Crandall-Stotler, J. L. Bowman, D. J. Weston, Y. Zhang, L. Chen, Y. Xie, F.-W. Li, C. J. Rothfels, A. Larsson, S. W. Graham, D. W. Stevenson, G. K.-S. Wong, J. Gershenzon, F. Chen, Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12328–12333 (2016).
59. Q. Jia, T. G. Köllner, J. Gershenzon, F. Chen, MTPSLs: New Terpene Synthases in Nonseed Plants. *Trends Plant Sci* **23**, 121–128 (2018).
60. C. Schmidt-Dannert, Biosynthesis of terpenoid natural products in fungi. *Adv Biochem Eng Biotechnol* **148**, 19–61 (2015).
61. P. Zhang, G. Wu, S. C. Heard, C. Niu, S. A. Bell, F. Li, Y. Ye, Y. Zhang, J. M. Winter, Identification and Characterization of a Cryptic Bifunctional Type I Diterpene Synthase Involved in Talaronoid Biosynthesis from a Marine-Derived Fungus. *Org. Lett.* **24**, 7037–7041 (2022).
62. Y. Yuan, S. Cheng, G. Bian, P. Yan, Z. Ma, W. Dai, R. Chen, S. Fu, H. Huang, H. Chi, Y. Cai, Z. Deng, T. Liu, Efficient exploration of terpenoid biosynthetic gene clusters in filamentous fungi. *Nat Catal* **5**, 277–287 (2022).
63. J. M. Galindo-Solís, F. J. Fernández, Endophytic Fungal Terpenoids: Natural Role and Bioactivities. *Microorganisms* **10**, 339 (2022).
64. R. Schobert, S. Knauer, S. Seibt, B. Biersack, Anticancer active illudins: recent developments of a potent alkylating compound class. *Curr Med Chem* **18**, 790–807 (2011).
65. M. Kimura, T. Tokai, N. Takahashi-Ando, S. Ohsato, M. Fujimura, Molecular and Genetic Studies of *Fusarium* Trichothecene Biosynthesis: Pathways, Genes, and Evolution. *Biosci. Biotechnol. Biochem.* **71**, 2105–2123 (2007).
66. M. J. Rynkiewicz, D. E. Cane, D. W. Christianson, Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13543–13548 (2001).
67. A. V. Nathanail, E. Varga, J. Meng-Reiterer, C. Bueschl, H. Michlmayr, A. Malachova, P. Fruhmann, M. Jestoi, K. Peltonen, G. Adam, M. Lemmens, R. Schuhmacher, F. Berthiller, Metabolism of the *Fusarium* Mycotoxins T-2 Toxin and HT-2 Toxin in Wheat. *J. Agric. Food Chem.* **63**, 7862–7872 (2015).

68. , Tri6 encodes an unusual zinc finger protein involved in regulation of trichothecene biosynthesis in *Fusarium sporotrichioides* | Applied and Environmental Microbiology (April 13, 2023).
69. M. M. Cadelis, B. R. Copp, S. Wiles, A Review of Fungal Protoilludane Sesquiterpenoid Natural Products. *Antibiotics* **9**, 928 (2020).
70. G. T. Wawrzyn, M. B. Quin, S. Choudhary, F. López-Gallego, C. Schmidt-Dannert, Draft Genome of *Omphalotus olearius* Provides a Predictive Framework for Sesquiterpenoid Natural Product Biosynthesis in Basidiomycota. *Chem. Biol.* **19**, 772–783 (2012).
71. M. Chen, W. K. W. Chou, T. Toyomasu, D. E. Cane, D. W. Christianson, Structure and Function of Fusicoccadiene Synthase, a Hexameric Bifunctional Diterpene Synthase. *ACS Chem. Biol.* **11**, 889–899 (2016).
72. R. Chiba, A. Minami, K. Gomi, H. Oikawa, Identification of Ophiobolin F Synthase by a Genome Mining Approach: A Sesterterpene Synthase from *Aspergillus clavatus*. *Org. Lett.* **15**, 594–597 (2013).
73. H. Tao, L. Lauterbach, G. Bian, R. Chen, A. Hou, T. Mori, S. Cheng, B. Hu, L. Lu, X. Mu, M. Li, N. Adachi, M. Kawasaki, T. Moriya, T. Senda, X. Wang, Z. Deng, I. Abe, J. S. Dickschat, T. Liu, Discovery of non-squalene triterpenes. *Nature* **606**, 414–419 (2022).
74. J. L. Chen, R. E. Moore, G. M. L. Patterson, Structures of nostocyclophanes A-D. *J. Org. Chem.* **56**, 4360–4364 (1991).
75. B. Zhao, X. Lin, L. Lei, D. C. Lamb, S. L. Kelly, M. R. Waterman, D. E. Cane, Biosynthesis of the Sesquiterpene Antibiotic Albaflavenone in *Streptomyces coelicolor* A3(2). *J Biol Chem* **283**, 8183–8189 (2008).
76. C. N. Tetzlaff, Z. You, D. E. Cane, S. Takamatsu, S. Omura, H. Ikeda, A Gene Cluster for Biosynthesis of the Sesquiterpenoid Antibiotic Pentalenolactone in *Streptomyces avermitilis*. *Biochem* **45**, 6179–6186 (2006).
77. S.-Y. Kim, P. Zhao, M. Igarashi, R. Sawa, T. Tomita, M. Nishiyama, T. Kuzuyama, Cloning and Heterologous Expression of the Cyclooctatin Biosynthetic Gene Cluster Afford a Diterpene Cyclase and Two P450 Hydroxylases. *Chem. Biol.* **16**, 736–743 (2009).
78. C. Nakano, M. Oshima, N. Kurashima, T. Hoshino, Identification of a New Diterpene Biosynthetic Gene Cluster that Produces O-Methylkolavelool in *Herpetosiphon aurantiacus*. *ChemBioChem* **16**, 772–781 (2015).
79. W. Liu, X. Feng, Y. Zheng, C.-H. Huang, C. Nakano, T. Hoshino, S. Bogue, T.-P. Ko, C.-C. Chen, Y. Cui, J. Li, I. Wang, S.-T. D. Hsu, E. Oldfield, R.-T. Guo, Structure, function and inhibition of ent-kaurene synthase from *Bradyrhizobium japonicum*. *Sci Rep* **4**, 6214 (2014).

80. K. Schulz-Bohm, L. Martín-Sánchez, P. Garbeva, Microbial Volatiles: Small Molecules with an Important Role in Intra- and Inter-Kingdom Interactions. *Front Microbiol* **8**, 2484 (2017).
81. H. Gürtler, R. Pedersen, U. Anthoni, C. Christophersen, P. H. Nielsen, E. M. H. Wellington, C. Pedersen, K. Bock, Alabaflavenone, a sesquiterpene ketone with a zizaene skeleton produced by a Streptomyces with a new rope morphology. *J. Antibiot.* **47**, 434–439 (1994).
82. P. Garbeva, C. Hordijk, S. Gerards, W. de Boer, Volatile-mediated interactions between phylogenetically different soil bacteria. *Front Microbiol* **5**, 289 (2014).
83. C. Song, R. Schmidt, V. de Jager, D. Krzyzanowska, E. Jongedijk, K. Cankar, J. Beekwilder, A. van Veen, W. de Boer, J. A. van Veen, P. Garbeva, Exploring the genomic traits of fungus-feeding bacterial genus Collimonas. *BMC Genomics* **16**, 1103 (2015).
84. J. S. Dickschat, Bacterial terpene cyclases. *Nat. Prod. Rep.* **33**, 87–110 (2015).
85. T. Tomita, S.-Y. Kim, K. Teramoto, A. Meguro, T. Ozaki, A. Yoshida, Y. Motoyoshi, N. Mori, K. Ishigami, H. Watanabe, M. Nishiyama, T. Kuzuyama, Structural Insights into the CotB2-Catalyzed Cyclization of Geranylgeranyl Diphosphate to the Diterpene Cyclooctat-9-en-7-ol. *ACS Chem. Biol.* **12**, 1621–1628 (2017).
86. R. Driller, S. Janke, M. Fuchs, E. Warner, A. R. Mhashal, D. T. Major, M. Christmann, T. Brück, B. Loll, Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis. *Nat Commun* **9**, 3971 (2018).
87. R. Janke, C. Görner, M. Hirte, T. Brück, B. Loll, The first structure of a bacterial diterpene cyclase: CotB2. *Acta Cryst D* **70**, 1528–1537 (2014).
88. F. Aqil, M. Zahin, K. A. El Sayed, I. Ahmad, K. Y. Orabi, J. M. Arif, Antimicrobial, antioxidant, and antimutagenic activities of selected marine natural products and tobacco cembranoids. *Drug Chem Toxicol* **34**, 167–179 (2011).
89. A. H. de Boer, I. J. de V. Leeuwen, Fusicocanes: diterpenes with surprising biological functions. *Trends Plant Sci* **17**, 360–368 (2012).
90. N. P. Thao, B. T. T. Luyen, R. Brun, M. Kaiser, P. Van Kiem, C. Van Minh, T. J. Schmidt, J. S. Kang, Y. H. Kim, Anti-Protozoal Activities of Cembrane-Type Diterpenes from Vietnamese Soft Corals. *Molecules* **20**, 12459–12468 (2015).
91. R. Driller, D. Garbe, N. Mehlmer, M. Fuchs, K. Raz, D. T. Major, T. Brück, B. Loll, Current understanding and biotechnological application of the bacterial diterpene synthase CotB2. *Beilstein J. Org. Chem.* **15**, 2355–2368 (2019).
92. V. Liato, M. Aïder, Geosmin as a source of the earthy-musty smell in fruits, vegetables and water: Origins, impact on foods and water, and review of the removing techniques. *Chemosphere* **181**, 9–18 (2017).

93. P. Garbeva, M. Avalos, D. Ulanova, G. P. van Wezel, J. S. Dickschat, Volatile sensation: The chemical ecology of the earthy odorant geosmin. *Environ. Microbiol.*, 1–10.
94. L. Zaroubi, I. Ozugergin, K. Mastronardi, A. Imfeld, C. Law, Y. G  linas, A. Piekny, B. L. Findlay, The Ubiquitous Soil Terpene Geosmin Acts as a Warning Chemical. *Appl. Environ. Microbiol.* **88**, e00093-22 (2022).
95. M. C. Stensmyr, H. K. M. Dweck, A. Farhan, I. Ibba, A. Strutz, L. Mukunda, J. Linz, V. Grabe, K. Steck, S. Lavista-Llanos, D. Wicher, S. Sachse, M. Knaden, P. G. Becher, Y. Seki, B. S. Hansson, A Conserved Dedicated Olfactory Circuit for Detecting Harmful Microbes in *Drosophila*. *Cell* **151**, 1345–1357 (2012).
96. P. G. Becher, M. Bengtsson, B. S. Hansson, P. Witzgall, Flying the Fly: Long-range Flight Behavior of *Drosophila melanogaster* to Attractive Odors. *J Chem Ecol* **36**, 599–607 (2010).
97. N. Melo, G. H. Wolff, A. L. Costa-da-Silva, R. Arribas, M. F. Triana, M. Gugger, J. A. Riffell, M. DeGennaro, M. C. Stensmyr, Geosmin Attracts *Aedes aegypti* Mosquitoes to Oviposition Sites. *Curr Biol* **30**, 127–134 (2020).
98. W. Zhou, Y. Wang, J. Wang, C. Peng, Z. Wang, H. Qin, G. Li, D. Li, Geosmin disrupts energy metabolism and locomotor behavior of zebrafish in early life stages. *Sci. Total Environ.* **859**, 160222 (2023).
99. J. Jiang, X. He, D. E. Cane, Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat Chem Biol* **3**, 711–715 (2007).
100. M. K  ksal, W. K. W. Chou, D. E. Cane, D. W. Christianson, Structure of 2-Methylisoborneol Synthase from *Streptomyces coelicolor* and Implications for the Cyclization of a Noncanonical C-Methylated Monoterpenoid Substrate. *Biochemistry* **51**, 3011–3020 (2012).
101. V. J. Paul, M. E. Hay, J. E. Duffy, W. Fenical, K. Gustafson, Chemical defense in the seaweed *Ochtodes secundiramea* (Montagne) Howe (Rhodophyta): effects of its monoterpenoid components upon diverse coral-reef herbivores. *J. Exp. Mar. Biol. Ecol.* **114**, 249–260 (1988).
102. M. Sandy, J. N. Carter-Franklin, J. D. Martin, A. Butler, Vanadium bromoperoxidase from *Delisea pulchra*: enzyme-catalyzed formation of bromofuranone and attendant disruption of quorum sensing. *Chem. Commun.* **47**, 12086–12088 (2011).
103. V. Lemesheva, C. Birkemeyer, D. Garbary, E. Tarakhovskaya, Vanadium-dependent haloperoxidase activity and phlorotannin incorporation into the cell wall during early embryogenesis of *Fucus vesiculosus* (Phaeophyceae). *Eur. J. Phycol.* **55**, 275–284 (2020).
104. W. C. Paradas, T. M. Crespo, L. T. Salgado, L. R. de Andrade, A. R. Soares, C. Hellio, R. R. Paranhos, L. J. Hill, G. M. de Souza, A. G. A. C. Kelecom, B. A. P. Da Gama, R. C. Pereira, G. M. Amado-Filho, Mevalonosomes: specific vacuoles containing the

- mevalonate pathway in *Plocamium brasiliense* cortical cells (Rhodophyta). *J Phycol* **51**, 225–235 (2015).
105. R. D. Kersten, S. Lee, D. Fujita, T. Pluskal, S. Kram, J. E. Smith, T. Iwai, J. P. Noel, M. Fujita, J.-K. Weng, A Red Algal Bourbonane Sesquiterpene Synthase Defined by Microgram-Scale NMR-Coupled Crystalline Sponge X-ray Diffraction Analysis. *J. Am. Chem. Soc.* **139**, 16838–16844 (2017).
 106. J. K. Brunson, S. M. K. McKinnie, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. Smith, D. A. Hutchins, A. E. Allen, B. S. Moore, Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018).
 107. J. R. Chekan, S. M. K. McKinnie, M. L. Moore, S. G. Poplawski, T. P. Michael, B. S. Moore, Scalable Biosynthesis of the Seaweed Neurochemical, Kainic Acid. *Angewandte Chemie International Edition* **58**, 8454–8457 (2019).
 108. J. R. Chekan, S. M. K. McKinnie, J. P. Noel, B. S. Moore, Algal neurotoxin biosynthesis repurposes the terpene cyclase structural fold into an N-prenyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12799–12805 (2020).
 109. T. S. Steele, J. K. Brunson, Y. Maeno, R. Terada, A. E. Allen, M. Yotsu-Yamashita, J. R. Chekan, B. S. Moore, Domoic acid biosynthesis in the red alga *Chondria armata* suggests a complex evolutionary history for toxin production. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117407119 (2022).
 110. R. J. Bartelt, B. W. Zilkowski, A. A. Cossé, U. Schnupf, K. Vermillion, F. A. Momany, Male-Specific Sesquiterpenes from *Phyllotreta* Flea Beetles. *J. Nat. Prod.* **74**, 585–595 (2011).
 111. F. Beran, T. G. Köllner, J. Gershenzon, D. Tholl, Chemical convergence between plants and insects: biosynthetic origins and functions of common secondary metabolites. *New Phytol.* **223**, 52–67 (2019).
 112. A. E. Brown, E. W. Riddick, J. R. Aldrich, W. E. Holmes, Identification of (–)- β -Caryophyllene as a Gender-Specific Terpene Produced by the Multicolored Asian Lady Beetle. *J Chem Ecol* **32**, 2489–2499 (2006).
 113. Z. Rebholz, L. Shewade, K. Kaler, H. Larose, F. Schubot, D. Tholl, A. V. Morozov, P. E. O’Maille, Emergence of terpene chemical communication in insects: Evolutionary recruitment of isoprenoid metabolism. *Protein Sci* **n/a**, e4634.
 114. D. Martin, J. Bohlmann, J. Gershenzon, W. Francke, S. J. Seybold, A novel sex-specific and inducible monoterpene synthase activity associated with a pine bark beetle, the pine engraver, *Ips pini*. *Naturwissenschaften* **90**, 173–179 (2003).
 115. P. R. Hughes, Myrcene: A precursor of pheromones in *Ips* beetles. *J. Insect Physiol.* **20**, 1271–1275 (1974).

116. S. J. Seybold, D. R. Quilici, J. A. Tillman, D. Vanderwel, D. L. Wood, G. J. Blomquist, De novo biosynthesis of the aggregation pheromone components ipsenol and ipsdienol by the pine bark beetles *Ips paraconfusus* Lanier and *Ips pini* (Say) (Coleoptera: Scolytidae). *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8393–8397 (1995).
117. A. B. Gilg, C. Tittiger, G. J. Blomquist, Unique animal prenyltransferase with monoterpene synthase activity. *Naturwissenschaften* **96**, 731–735 (2009).
118. F. Beran, P. Rahfeld, K. Luck, R. Nagel, H. Vogel, N. Wielsch, S. Irmisch, S. Ramasamy, J. Gershenzon, D. G. Heckel, T. G. Köllner, Novel family of terpene synthases evolved from trans-isoprenyl diphosphate synthases in a flea beetle. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2922–2927 (2016).
119. K. Darragh, A. Orteu, D. Black, K. J. R. P. Byers, D. Szczerbowski, I. A. Warren, P. Rastas, A. Pinharanda, J. W. Davey, S. F. Garza, D. A. Almeida, R. M. Merrill, W. O. McMillan, S. Schulz, C. D. Jiggins, A novel terpene synthase controls differences in anti-aphrodisiac pheromone production between closely related *Heliconius* butterflies. *PLOS Biology* **19**, e3001022 (2021).
120. X. Chen, T. G. Köllner, W. Xiong, G. Wei, F. Chen, Emission and biosynthesis of volatile terpenoids from the plasmodial slime mold *Physarum polycephalum*. *Beilstein J. Org. Chem.* **15**, 2872–2880 (2019).
121. J. Rinkel, T. G. Köllner, F. Chen, J. S. Dickschat, Characterisation of three terpene synthases for β -barbatene, β -araneosene and nephtenol from social amoebae. *Chem. Commun.* **55**, 13255–13258 (2019).
122. X. Chen, K. Luck, P. Rabe, C. Q. Dinh, G. Shaulsky, D. R. Nelson, J. Gershenzon, J. S. Dickschat, T. G. Köllner, F. Chen, A terpene synthase-cytochrome P450 cluster in *Dictyostelium discoideum* produces a novel trisnorsesquiterpene. *eLife* **8**, e44352 (2019).
123. X. Chen, T. G. Köllner, Q. Jia, A. Norris, B. Santhanam, P. Rabe, J. S. Dickschat, G. Shaulsky, J. Gershenzon, F. Chen, Terpene synthase genes in eukaryotes beyond plants and fungi: Occurrence in social amoebae. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12132–12137 (2016).
124. P. Rabe, J. Rinkel, B. Nubbemeyer, T. G. Köllner, F. Chen, J. S. Dickschat, Terpene Cyclases from Social Amoebae. *Angew. Chem., Int. Ed.* **55**, 15420–15423 (2016).
125. X. Chen, T. G. Köllner, G. Shaulsky, Q. Jia, J. S. Dickschat, J. Gershenzon, F. Chen, Diversity and Functional Evolution of Terpene Synthases in Dictyostelid Social Amoebae. *Sci Rep* **8**, 14361 (2018).
126. M. Huang, C. Abel, R. Sohrabi, J. Petri, I. Haupt, J. Cosimano, J. Gershenzon, D. Tholl, Variation of Herbivore-Induced Volatile Terpenes among *Arabidopsis* Ecotypes Depends on Allelic Differences and Subcellular Targeting of Two Terpene Synthases, TPS02 and TPS03. *Plant Physiol.* **153**, 1293–1310 (2010).

127. J. Fäldt, M. Jonsell, G. Nordlander, A.-K. Borg-Karlson, Volatiles of Bracket Fungi *Fomitopsis pinicola* and *Fomes fomentarius* and Their Functions as Insect Attractants. *J Chem Ecol* **25**, 567–590 (1999).
128. C. Lyu, T. Chen, B. Qiang, N. Liu, H. Wang, L. Zhang, Z. Liu, CMNPD: a comprehensive marine natural products database towards facilitating drug discovery from the ocean. *Nucleic Acids Res* **49**, D509–D515 (2021).
129. M. D. Unson, D. J. Faulkner, Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia* **49**, 349–353 (1993).
130. A. E. Flowers, M. J. Garson, R. I. Webb, E. J. Dumdei, R. D. Charan, Cellular origin of chlorinated diketopiperazines in the dictyoceratid sponge *Dysidea herbacea* (Keller). *Cell Tissue Res* **292**, 597–607 (1998).
131. M. J. Uriz, X. Turon, J. Galera, J. M. Tur, New light on the cell location of avarol within the sponge *Dysidea avara* (Dendroceratida). *Cell Tissue Res* **285**, 519–527 (1996).
132. M. J. Garson, J. E. Thompson, R. M. Larsen, C. N. Battershill, P. T. Murphy, P. R. Bergquist, Terpenes in sponge cell membranes: Cell separation and membrane fractionation studies with the tropical marine sponge *Amphimedon* sp. *Lipids* **27**, 378–388 (1992).
133. P. Becher, V. Verschut, M. Bibb, M. Bush, B. Molnar, E. Barane, M. Al-Bassam, G. Chandra, L. Song, G. Challis, M. Buttner, K. Flärdh, Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to *Streptomyces* bacteria promoting spore dispersal. *Nat Microbiol* **5**, 1–9 (2020).
134. D. A. T. Boncan, S. S. K. Tsang, C. Li, I. H. T. Lee, H.-M. Lam, T.-F. Chan, J. H. L. Hui, Terpenes and Terpenoids in Plants: Interactions with Environment and Insects. *Int J Mol Sci* **21**, 7382 (2020).
135. B. Singh, R. A. Sharma, Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech* **5**, 129–151 (2015).
136. D. R. Houston, J. G. Hanna, J. C. Lathe, S. G. Hillier, R. Lathe, Evidence that nuclear receptors are related to terpene synthases. *J Mol Endocrinol* **68**, 153–166 (2022).
137. Q. Jia, X. Chen, T. G. Köllner, J. Rinkel, J. Fu, J. Labbé, W. Xiong, J. S. Dickschat, J. Gershenzon, F. Chen, Terpene Synthase Genes Originated from Bacteria through Horizontal Gene Transfer Contribute to Terpenoid Diversity in Fungi. *Sci Rep* **9**, 9223 (2019).

Section 1.7. Acknowledgments

Chapter 1, in part is currently being prepared for submission for publication in Natural Product Reports by Taylor S. Steele, Kayla Wilson, Immo Burkhardt, and Bradley S. Moore. The dissertation author is the co-primary author of this paper with Taylor S. Steele.

Section 2.1. Introduction

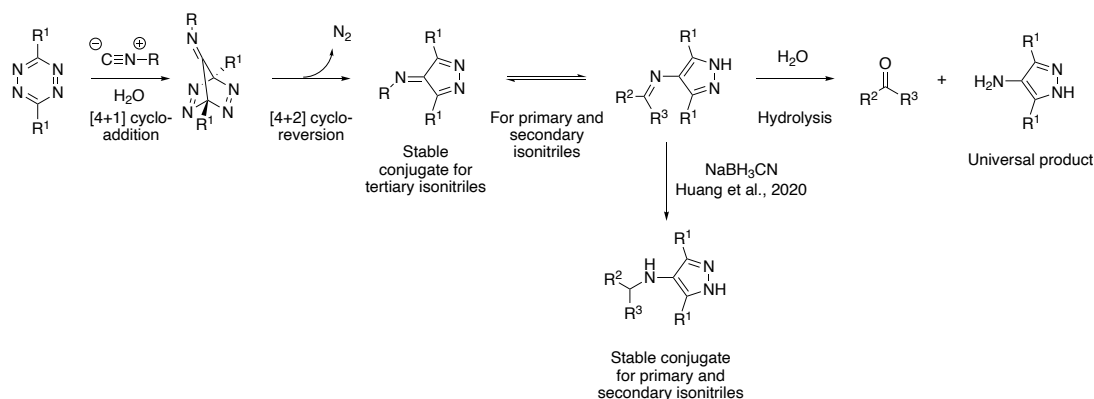
The detection, purification, and structural elucidation of natural products is a critical but sometimes challenging part of natural product research. Early natural products chemists often had to work on large preparative scales to gather enough material to characterize a natural product molecule. Advancements in NMR, chromatography, mass spectrometry, and other analytical technologies have allowed for the detection and purification of smaller and smaller amounts of compound. While many natural product chemists still rely on early purification techniques, like TLC or gravity column chromatography, as a whole the field relies heavily on the use of high-performance liquid chromatography (HPLC), often coupled to ultraviolet-visible spectrophotometers and/or mass spectrometers. HPLC uses a pressurized liquid phase to separate a mixture of compounds through a compressed solid phase. The pressure used in HPLC allows for faster separation of mixtures compared to simple gravity chromatography. Additionally, the stationary phase of HPLC is composed of much smaller particles, which increases the overall surface area for interactions between the sample compounds and the solid phase, allowing for better separation of the sample compounds. The higher resolution of HPLC also allows for the separation of very small amounts of compound. The UV-Vis and MS detectors that are often coupled to HPLCs have become increasingly sensitive, which also aids in the detection and purification of small amounts of compounds.

Even with advances in detection and purification technologies, many natural products are still difficult to characterize for a number of reasons. The “easiest” natural products to purify have a strong UV chromophore, ionize well for MS analysis, and have reasonable retention times on common reversed-phase HPLC columns, but many natural products do not fit these “ideal”

parameters. One strategy to address the challenges in detecting and purifying “non-ideal” natural products is derivatization, where a synthetic derivatizing agent (also referred to as a probe or label) is reacted with a natural product to confer a number of desired attributes, such as increased or decreased polarity, increased UV absorption, or increased MS ionization (1). These probes have been used to aid in the detection of numerous natural products, including domoic acid (2), saxitoxin (3), anatoxin-a (3), and ivermectin (4). Many of these derivatizing agents work by binding to certain functional groups on the target molecules, such as amines, alcohols, carboxylic acids, ketones, and more, with varying degrees of chemoselectivity (1). When working with complex chemical extracts from natural sources, high chemoselectivity is often valuable as it helps to avoid off-target labeling. Chemoselectivity can often be affected by pH and solvent, which must be accounted for when using a probe to target a natural product from a chemical extract (5, 6). Choosing the correct functional group to target for natural product labeling also requires careful consideration, given the abundance of functional groups such as carboxylic acids, carbonyls, and amines in non-target primary metabolites.

Isonitriles are found in a variety of natural products, including sponge isonitrile terpenes, cyanobacterial hapalindoles and ambiguines, and plant isocyalexin A. Isonitriles were first reported to react with tetrazines in a [4+1] cycloaddition in 1982 (7). This isonitrile-tetrazine click reaction has since been used as an important tool for biorthogonal ligation chemistry, which focuses on using chemical reactions to study biology without interfering with biological processes (8). The reaction has also more recently been used to detect and isolate isonitrile natural products (9, 10). The reaction begins with a [4+1] cycloaddition to form an unstable tetraazanorbornadienimine intermediate, followed by a spontaneous [4+2] cycloreversion

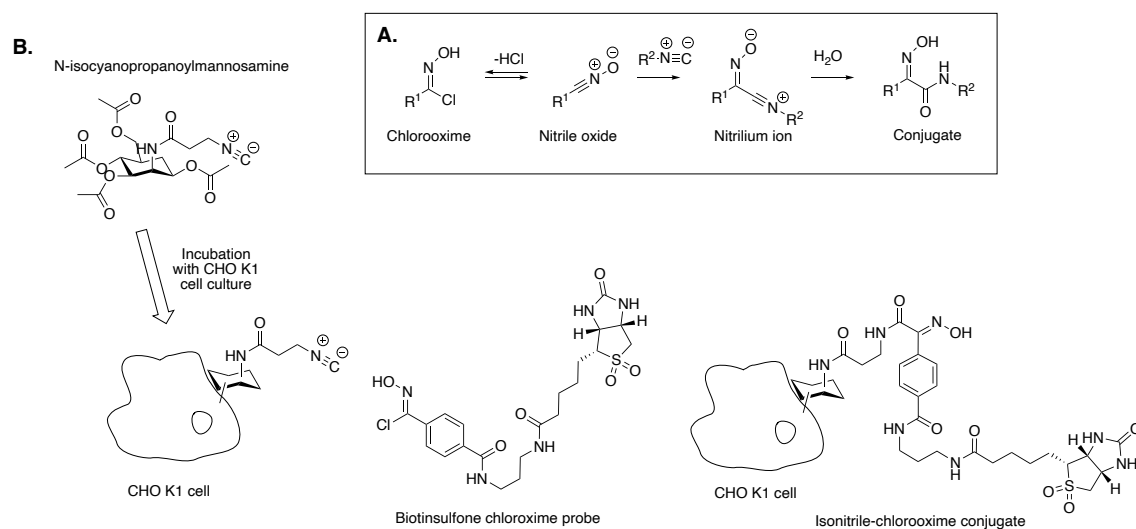
to yield a 4*H*-pyrazol-4-imine product (Scheme 2.1.1). This product is stable for reactions with tertiary isonitriles, however for primary and secondary isonitriles, the product tautomerizes and quickly hydrolyzes in the presence of water to the universal aminopyrazole and carbonyl products. The universal aminopyrazole product is useful for detecting the presence of primary and secondary isonitriles, however it does not allow for direct capture and enrichment of these natural products for structural determination. To address this, Huang et al proposed an additional reduction step using sodium cyanoborohydride, which allows for the formation of a reduced conjugate instead of the hydrolysis product (10)(Scheme 2.1.1). Using this methodology, they were able to detect primary isonitrile lipopeptides from *Streptomyces tsukubensis* NRRL 18488. The isonitrile-tetrazine ligation was also used for untargeted labeling of an *E. coli* strain encoding isonitrile synthesis genes from *Photorhabdus luminescens* and *Vibrio cholerae* which enabled the detection of 4-methoxyphenyl isocyanide (9).



Scheme 2.1.1. Isonitrile-tetrazine ligation. A [4+1] cycloaddition followed by a spontaneous [4+2] cycloreversion yields a 4*H*-pyrazol-4-imine product which is stable for reactions with tertiary isonitriles. For reactions with tetrazines and primary or secondary isonitriles, the 4*H*-pyrazol-4-imine intermediate undergoes an isomerization followed by quick hydrolysis to the universal aminopyrazole product and a carbonyl product. Huang et al., 2020 proposed using NaBH_3CN as a mild reducing agent to prevent hydrolysis and yield the reduced stable conjugate for primary and secondary isonitriles.

In 2019, Schäfer et al. proposed a new reaction for biorthogonal ligation with isonitriles: the isonitrile-chlorooxime ligation (11) (Scheme 2.1.2). Unlike the tetrazine ligation, which only

provides stable conjugates for tertiary isonitriles, the chlorooxime ligation yields stable conjugates for primary, secondary, and tertiary isonitriles without the need for an additional reducing agent. This reaction is described for imaging metabolically labeled cell surface glycans via cellular incubation of N-acetylmannosamine derivatives, which are readily incorporated into cell membranes as neuraminic acids. Schaefer et al. prepared peracetylated N-isocyanopropionylmannosamine which was then incubated with Chinese hamster ovary cells (CHO K1 cells) for incorporation into the cell surface. The CHO K1 cells were then incubated with a prepared biotinsulfone chlorooxime probe, which was ligated to the isonitrile-tagged cell surface. Staining with a biotin-binding dye allowed for cell surface visualization by confocal microscopy. Schäfer et al. also demonstrated that the isonitrile-chlorooxime ligation could be used in combination with other ligation reactions for imaging of multiple cellular targets.



Scheme 2.1.2. A. Isonitrile-chlorooxime ligation. B. Use of the isonitrile-chlorooxime conjugation for cell surface glycan labeling as shown in Shaefer et al., 2019. (Figure adapted from (12))

In the following publication, of which Dr. Rebecca Schäfer is the lead author and I am the second author, we describe the use of the isonitrile-chlorooxime ligation to chemoselectively target isonitrile-containing natural products for purification and structural elucidation (13). Building on

Dr. Schäfer's previous work from the Helma Wennemers laboratory on biorthogonal cell surface glycan labeling, this paper demonstrates a new use for this ligation as a natural product probe. When Dr. Schäfer joined the Moore lab as a postdoctoral scholar, I was attempting to purify and structurally characterize several isonitrile sesquiterpenoids from a local marine sponge of the order Bubarida. Several isonitrile sesquiterpenoids had been previously reported from the same Bubarida sponge (14), and I could detect by GC-MS several compounds with the same molecular weight as the reported isonitrile sesquiterpenoids; however, these compounds were highly lipophilic, did not separate well on reversed phase HPLC C18 columns, had no UV chromophore that could be easily detected by UV-Vis, and had weak ionization by HPLC-MS. The molecules are highly volatile, and as such were often lost during concentration by rotary evaporator, hi-vac, and condensation under nitrogen or argon gas. The sponge extracts also contained several compounds with the same molecular weight and retention time, which made separating each molecule for further structural elucidation very challenging. Dr. Schäfer had already been developing a chlorooxime probe for natural product detection with Maurice Bierdmann and Simon Sieber for use in cyanobacterial extracts. This synthetic chlorooxime probe was designed by Dr. Schäfer to have enhanced UV absorbance from an aromatic ring and a distinct 1:1 ^{79}Br : ^{81}Br isotopic pattern for MS-detection from a bromine moiety. Reactions of this chlorooxime probe with *Fischerella ambigua* UTEX 1903 cultures increased the MS detection limit of multiple known isonitrile-containing indole alkaloids and allowed for targeted identification of multiple related ambiguines and hapalindole H. While the probe allowed for facile detection of isonitrile-containing cyanobacterial natural products, no attempts were made to purify and structurally elucidate the chlorooxime-natural product conjugates. Dr. Schäfer and I realized that using the same bromobenzo chlorooxime probe for my Bubarida sponge extracts would not only help my efforts to purify and structurally elucidate

the isonitrile terpene natural products, but it would also demonstrate a novel use for Dr. Schaefer's probe.

For this publication, I provided Bubarida sponge tissue and chemical extracts, and along with Dr. Schäfer, performed the isonitrile chlorooxime ligation reactions, purified the natural-product probe conjugates, and structurally characterized the conjugates. This project helped by strengthen a variety of skills, including small molecule detection by GCMS, LCMS, and LC UV-Vis, small molecule purification by preparative TLC, flash chromatography, and normal phase HPLC, and small molecule structure elucidation by NMR. This work allowed us to confirm that the molecules reported by Thompson et al. from this San Diego sponge in 1985 are the same molecules we report from our San Diego Bubarida sponge (14). The structural elucidation of these Bubarida sponge natural products played a critical role in setting up the work in Chapter 3 describing a new class of terpene synthases from the same Bubarida sponge. Elucidating the structure of those sponge natural products also aided our understanding of how the sponge natural products are connected to the terpene synthase products described in Chapter 3.

Section 2.2. References for Section 2.1

1. C. C. Hughes, Chemical labeling strategies for small molecule natural product detection and isolation. *Nat. Prod. Rep.* **38**, 1684–1705 (2021).
2. D. G. Beach, H. Liu, M. A. Quilliam, Analytical Methods Sensitive determination of domoic acid in mussel tissue using dansyl chloride derivatization and liquid chromatography-mass spectrometry. *Anal. Methods* **7**, 1000–1007 (2015).
3. A. Roy-lachapelle, M. Sollicec, S. Sauvé, Determination of BMAA and three alkaloid cyanotoxins in lake water using dansyl chloride derivatization and high-resolution mass spectrometry. *Anal. Bioanal. Chem.* **407**, 5487–5501 (2015).
4. X. Xie, S. Gong, X. Wang, Y. Wu, L. Zhao, Food Additives & Contaminants : Part A Simplified RP-HPLC method for multi-residue analysis of abamectin , emamectin benzoate and ivermectin in rice. *Food Addit. Contam. Part A* **28**, 19–25 (2011).

5. B. Seiwert, U. Karst, Simultaneous LC / MS / MS Determination of Thiols and Disulfides in Urine Samples Based on Differential Labeling with Ferrocene-Based Maleimides. *Anal. Chem.* **79**, 7131–7138 (2007).
6. C. Avonto, O. Taglialatela-Scafati, F. Pollastro, A. Minassi, V. Di-Marzo, L. De-Petrocellis, G. Appendino, An NMR spectroscopic method to identify and classify thiol-trapping agents: Revival of michael acceptors for drug discovery? *Angew. Chemie - Int. Ed.* **50**, 467–471 (2011).
7. B. P. Imming, R. Mohr, E. Müller, W. Overheu, G. Seitz, [4 + 1] Cycloaddition of Isocyanides to 1,2,4,5-Tetrazines: A Novel Synthesis of Pyrazole. *Angew. Chemie - Int. Ed.* **21**, 20133 (1982).
8. R. E. Bird, S. A. Lemmel, X. Yu, Q. A. Zhou, Bioorthogonal Chemistry and Its Applications. *Bioconjug. Chem.* **32**, 2457–2479 (2021).
9. T. Tørring, S. R. Shames, W. Cho, C. R. Roy, J. M. Crawford, Acyl Histidines : New N-Acyl Amides from *Legionella pneumophila*. *ChemBioChem* **18**, 638–646 (2017).
10. Y.-B. Huang, W. Cai, A. Del Rio Flores, F. F. Twigg, W. Zhang, Facile Discovery and Quantification of Isonitrile Natural Products via Tetrazine-Based Click Reactions. *Anal. Chem.* **92**, 599–602 (2020).
11. R. J. B. Schafer, M. R. Monaco, M. Li, A. Tirla, P. Rivera-Fuentes, H. Wennemers, The Bioorthogonal Isonitrile – Chlorooxime Ligation. *J. Am. Chem. Soc.* **141** (2019).
12. Y. Méndez, A. V. Vasco, A. R. Humpierre, B. Westermann, Isonitriles: Versatile Handles for the Bioorthogonal Functionalization of Proteins'. *ACS Omega* **5**, 25505–25510 (2020).
13. R. Schäfer, K. Wilson, M. Biedermann, B. Moore, S. Sieber, H. Wennemers, Identification of Isonitrile-Containing Natural Products in Complex Biological Matrices Through Ligation with Chlorooximes. *Chem. – A Eur. J.* **29**, e202203277 (2022).
14. J. E. Thompson, R. P. Walker, D. J. Faulkner, Screening and bioassays for biologically-active substances from forty marine sponge species from San Diego, California, USA. *Mar. Biol.* **21**, 11–21 (1985).

Section 2.3. Reprint of “Identification of Isonitrile-Containing Natural Products in Complex Biological Matrices Through Ligation with Chlorooximes”

Identification of Isonitrile-Containing Natural Products in Complex Biological Matrices through Ligation with Chlorooximes

Rebecca J. B. Schäfer,^[a, b] Kayla Wilson,^[b] Maurice Biedermann,^[a] Bradley S. Moore,^{*,[b, c]} Simon Sieber,^{*,[d]} and Helma Wennemers^{*,[a]}

Abstract: Isonitrile-containing natural products have garnered attention for their manifold bioactivities but are difficult to detect and isolate due to the chemical lability of the isonitrile functional group. Here, we used the isonitrile-chlorooxime ligation (INC) in a reactivity-based screening (RBS) protocol for the detection and isolation of alkaloid and terpene isonitriles in the cyanobacterium *Fischerella ambigua* and a

marine sponge of the order Bubarida, respectively. A trifunctional probe bearing a chlorooxime moiety, a UV active aromatic moiety, and a bromine label facilitated the chemoselective reaction with isonitriles, UV-Vis spectroscopic detection, and mass spectrometric analysis. The INC-based RBS allowed for the detection, isolation, and structural elucidation of isonitriles in microgram quantities.

Natural products featuring an isonitrile functional group occur in a wide variety of natural sources including bacteria, fungi, marine sponges, and plants (Figure 1a).^[1–6] Many exhibit potent bioactivity against bacteria, fungi, and parasites.^[4–7] Natural isonitriles are therefore promising targets and lead structures for drug development. The detection and isolation of isonitrile-containing natural products is, however, challenging as the isonitrile functional group is not UV-active and is inherently labile towards hydrolysis under acidic and strongly basic conditions.

Reactivity-based screening (RBS) – the chemical labeling of a specific functional group in a natural product – is an attractive

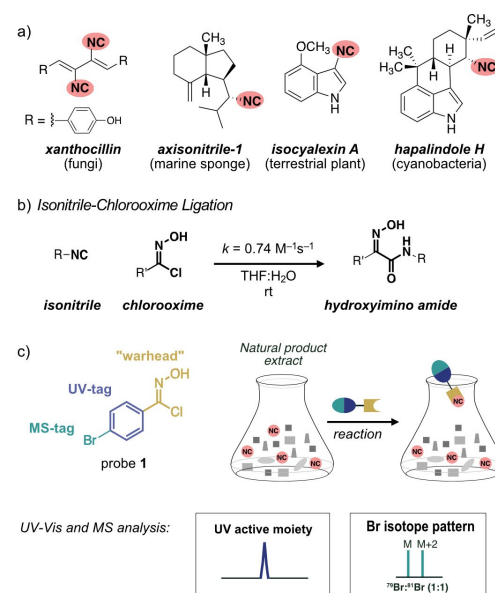


Figure 1. a) Examples of isonitriles and their natural sources. b) Isonitrile-chlorooxime ligation. c) Reactivity-based screening (RBS) for isonitriles with probe 1.

method to overcome challenges arising from poor detectability, low bioavailability, and chemical instability.^[8,9] Ideal RBS reactions are highly chemoselective for the target functional group, form a covalent bond that is stable under isolation conditions, and afford a label that improves detection, for example, by UV-Vis spectroscopy or mass spectrometry (MS).^[10,11] In a recent

[a] Dr. R. J. B. Schäfer, M. Biedermann, Prof. H. Wennemers
 Laboratory of Organic Chemistry, ETH Zürich
 Vladimir-Prelog-Weg 3, 8093 Zürich (Switzerland)
 E-mail: Helma.Wennemers@org.chem.ethz.ch
 Homepage: <https://wennemers.ethz.ch>

[b] Dr. R. J. B. Schäfer, K. Wilson, Prof. B. S. Moore
 Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093 (USA)
 E-mail: bsmoore@ucsd.edu
 Homepage: <https://bsmoore.scrippsprofiles.ucsd.edu>

[c] Prof. B. S. Moore
 Skaggs School of Pharmacy and Pharmaceutical Sciences
 University of California San Diego
 La Jolla, CA 92093 (USA)

[d] Dr. S. Sieber
 University of Zurich
 Winterthurerstrasse 190, 8057 Zurich (Switzerland)
 E-mail: simon.sieber@chem.uzh.ch

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/chem.202203277>

© 2022 The Authors. Chemistry – A European Journal published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

report, a tetrazine-based probe was used for the qualitative detection of isonitrile-containing natural products.^[12] This [4 + 1] tetrazine-isonitrile ligation provides stable products with tertiary isonitriles. In the case of primary and secondary isonitriles the initially formed 4*H*-pyrazol-4-imine product undergoes spontaneous hydrolysis and escapes detection^[13] if not trapped through reduction with NaBH₃CN.^[12]

We recently developed the isonitrile-chlorooxime (INC) ligation (Figure 1b).^[14] This highly chemoselective reaction is fast (0.74 M⁻¹s⁻¹) and allows for the labeling of cell surface glycans in cellulose.^[14] The resulting hydroxyimino amide products are stable, regardless of whether primary, secondary, or tertiary isonitriles serve as reaction partners.^[15] We, therefore, anticipated that appropriately functionalized chlorooximes could serve as powerful probes for the RBS facilitated detection, isolation, and – ideally also – structural elucidation of isonitriles in complex biological matrices. We envisioned brominated chlorooxime **1** as a suitable probe for RBS for isonitriles (Figure 1c). This probe consists of a) a chlorooxime moiety as a “warhead” for chemoselective reaction with isonitriles, b) a bromine substituent as a mass tag to facilitate MS-based analysis by providing a distinct ⁷⁹Br:⁸¹Br (1:1) isotope pattern in the labeled natural products,^[16] and c) an aromatic moiety for enhanced UV absorption. Probe **1** was readily available in two synthetic steps starting from commercial 4-bromo benzaldehyde (Scheme S1 in the Supporting Information).^[17]

Here, we applied the INC ligation for the detection and identification of isonitrile-containing alkaloids and terpenes in the cyanobacterium *Fischerella ambigua* (*F. ambigua*) and the Bubarida marine sponge. We show that INC is a robust tool for RBS and sufficiently sensitive to allow even for the identification and isolation of isonitrile natural products that occur in minute quantities.

We started our RBS for isonitriles by exploring whether probe **1** detects isonitriles in a crude extract of *F. ambigua* (Figure 2a). This culturable cyanobacterium is a well-studied producer of tryptophan-derived isonitrile-containing indole alkaloids, including hapalindoles and ambiguines.^[18,19] Analysis of a crude methanol extract of cultured strain UTEX 1903 by ultrahigh performance liquid chromatography (UHPLC) coupled with high-resolution mass spectrometry (HRMS) detected seven known isonitrile indole alkaloids under our laboratory conditions (Figures S1a–S9a).

Upon incubation of the dried crude extract with probe **1** in a THF/water (1:1) mixture followed by UHPLC–HRMS analysis, we observed nine signals with a bromine isotope pattern (Figures 2b and S1b–S9b). Controls that contained only probe **1** and only the crude extract, respectively, did not show these signals (Figures S1b–S9b). The signals correspond to the masses of the conjugates between probe **1** and ambigaine A, C, H, I, E, D or O, K, and L and hapalindole H, nine isonitrile indole alkaloids produced by the *F. ambigua* UTEX 1903 strain (Figure 2c). Thus, conjugation with probe **1** enhanced the detection limit of isonitriles in the natural product extract. The detected natural products contain secondary and tertiary isonitriles. These results highlight the value of probe **1** for the

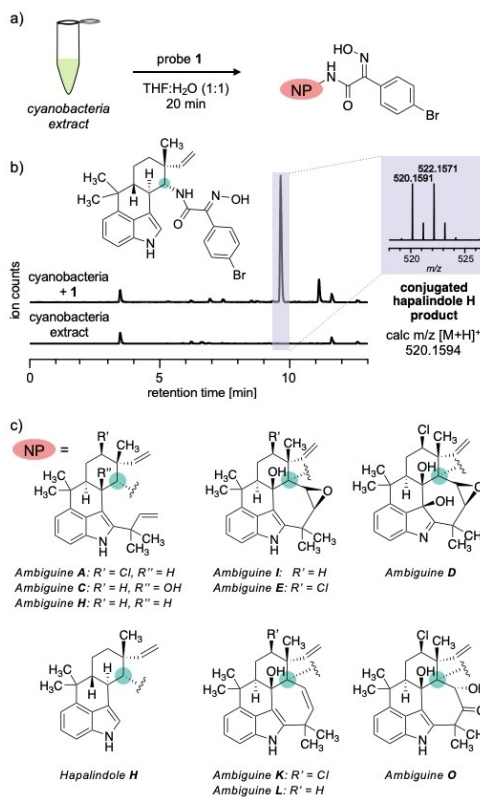


Figure 2. a) Incubation of a crude extract of *F. ambigua* (strain UTEX 1903) with **1**. b) Representative extracted ion chromatogram of the extract with (top) and without (bottom) probe **1**, for the mass of conjugated hapalindole H. c) Labeled isonitrile-containing congeners of natural products (a green dot indicates the location of the isonitrile group and the attachment site of probe **1**).

identification of different types of isonitriles through RBS in complex mixtures.

Next, we tested the value of chlorooxime probe **1** for the identification and isolation of isonitrile-containing natural products in a marine sponge. In contrast to cyanobacteria, which are culturable and therefore allow for increasing the production of natural isonitriles in the laboratory, marine sponges are typically not maintained in the laboratory, least scaled-up. Additionally, many marine sponge environments are under increased threat from ocean warming, terrestrial pollution, overfishing, and other stressors, which makes it increasingly more challenging to find and collect significant quantities of isonitrile-producing sponges. Thus, the detection, isolation, and characterization of such low-abundance sponge natural products require a highly sensitive methodology.

SCUBA collections off the coast of California gave access to 20 g (wet weight) of a marine sponge specimen that is a known producer of isonitrile-containing terpenes.^[20] Extraction with methylene chloride provided ~60 mg (dry weight) of a crude extract. Gas chromatography–mass spectrometry (GC–MS) analysis revealed mass-to-charge ratios corresponding to putative nitrogenous sesquiterpenes, including isonitrile, isocyanate, formamide, and isothiocyanate congeners as judged by comparison with spectral data from the NIST mass spectral library (Figure S10).^[21] Isolation and structural characterization of the isonitriles by conventional means proved, however, challenging due to their low UV activity, high lipophilicity, and volatility. We, therefore, envisioned that labeling the isonitrile-sesquiterpenes with our chlorooxime probe would facilitate MS and UV analysis-supported detection and, ideally, chromatographic purification and elucidation of the constitutive structure of the terpene backbone.

Treatment of the Bubarida sponge crude extract with chlorooxime probe 1 in THF and citrate buffer pH 5 (1:1)^[22] followed by LC–UV analysis revealed two major signals (dark and light blue) and one minor signal (gray; Figure 3a).^[23] The MS spectra associated with these UV signals I, II, and III showed the bromine isotope pattern (Figure 3b). Controls that contained only probe 1 or only the crude extract did not show these signals (Figure 3a) corroborating that these products are derived from labeled isonitriles.^[23] The mass spectra corresponding to the UV signals I and II ($t = 17.5$ and 17.9 min) show the same mass-to-charge ratio of $[M+H]^+ = 447.17$ Da suggesting that the isonitrile precursors are constitutional isomers with a C_{15} terpene-skeleton (Figure 3a). The UV-activity of the labeled compounds allowed for isolation of these two major products (Figure 3a, dark and light blue) in a quantity of less than 500 μ g by preparative thin-layer chromatography (TLC) and normal phase HPLC. This amount sufficed for 1D and 2D NMR spectroscopic analysis including 2D nuclear Overhauser effect (NOE) to assign the relative stereochemistry (Tables S1 and S2, Figures S13 and S14). These analyses revealed that LC UV signal II corresponds to sesquiterpene-probe conjugate 2 and thus the known isonitrile congener (3) of sesquiterpene isothiocyanate epipolasin A (Figure 3b).^[24,25] Signal I corresponds to sesquiterpene-probe conjugate 4 and thus cadinane-type sesquiterpene 5, which is a constitutional isomer of 3 (Figure 3b). 5 is a known sesquiterpene from the Mediterranean sponge *Axinyssa* sp. and *Axinella cannabina* (Figure 3c),^[26,27] but has not been reported from this Bubarida sponge.

The amount of the compound(s) corresponding to the minor LC UV signal III ($t = 14.9$ min, Figure 3a) did not suffice for isolation. Analysis of the corresponding MS spectrum revealed molecular ions at m/z 493.1314 $[M+H]^+$ and 511.1408 $[M+H]^+$. These masses are in agreement with molecular formulas of oxygenated isonitrile sesquiterpenoids (e.g., $[C_{23}H_{30}BrN_2O_5]^+$ and $[C_{23}H_{32}BrN_2O_6]^+$), compounds that are rare^[3] and have not yet been observed in the Bubarida sponge. Thus, the chemo-selective ligation with probe 1 revealed compounds in a marine sponge that could not be detected with conventional analytical methods.

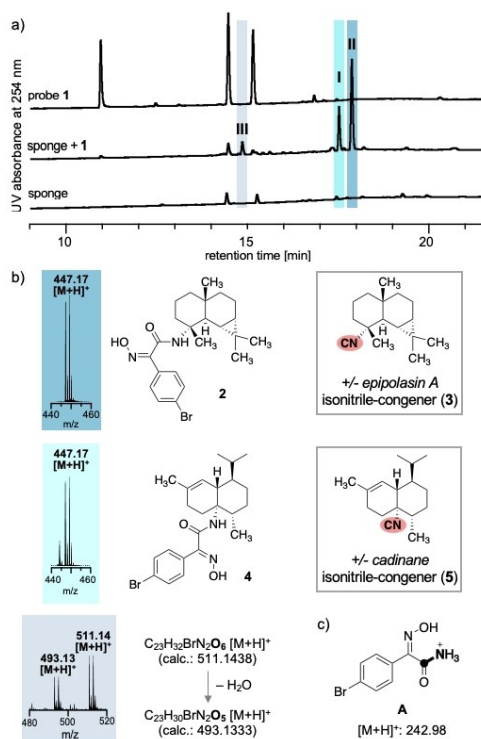


Figure 3. a) Stacked LC–UV chromatograms of probe 1 (top), sponge crude extract with (middle) and without (bottom) probe 1. b) MS spectra of the newly formed products as monitored by UV (blue: I and II; gray: III) 2 and 4 and the corresponding isonitrile-containing epipolasin A (3) and cadinane-type (5) sesquiterpenoids. c) Fragment A formed by MS/MS from ligation products.

These findings also corroborate the exquisite chemoselectivity for isonitriles of the INC ligation. Products from reaction of probe 1 with compounds bearing other functional groups, including closely related isocyanates and isothiocyanates that are produced by the Bubarida sponge,^[26] were not detected. We confirmed this notable chemoselectivity of chlorooxime reactivity further through competition experiments between isocyanates, isothiocyanates and isonitriles for reaction with probe 1 (Scheme S2, Figures S15 and S16).^[28]

Furthermore, MS/MS spectra of each of the three compounds showed upon collision-induced dissociation a common fragment with a mass-to-charge ratio of 242.98 $[M+H]^+$ _{obs} (Figure S17), which corresponds to hydroxyimino amide A (242.9764 $[M+H]^+$ _{calc}, Figure 3c). To ascertain the origin of this MS/MS fragment, we synthesized hydroxyimino amide S3 from probe 1 and 1,1,3,3-tetramethyl butyl isocyanide. MS/MS analysis of S3 showed the formation of the same fragment A (Figure S18). Thus, A arises from MS fragmentation of a conjugate between probe 1 and an isonitrile (Scheme S3). The

occurrence of MS/MS fragment A is therefore diagnostic for isonitriles and a valuable signature for isonitriles.^[29]

In conclusion, reactivity-based screening by using isonitrile-chlorooxime ligation is a valuable tool for the labeling, detection, and isolation of isonitrile-containing natural products at microgram quantities in complex mixtures. This work puts forth a trifunctional probe consisting of a) a chlorooxime for chemoselective reaction with isonitriles in complex mixtures, b) an aromatic moiety for detection by UV-Vis, and c) a bromine mass tag for detection by mass spectrometry. The probe facilitated the identification, isolation, and structure elucidation of alkaloid and terpene isonitriles in cyanobacteria and marine sponges collected in the field. Our findings open exciting prospects for the identification of isonitrile natural products with as-yet unknown bioactivities.

Acknowledgements

H.W., S.S., and R.J.B.S. gratefully acknowledge the Swiss National Science Foundation for funding (200020_18729/1; H.W.) (CRSK-2_190556; Spark grant S.S.), and (P2EZP3_195643; postdoctoral fellowship R.J.B.S.). B.S.M. acknowledges NIH grant R01-GM085770. K.W. graciously acknowledges the support from the National Science Foundation's Graduate Research Fellowship Program and the University of California San Diego Marine Sciences Academic Senate Grant. We thank Reiko Cullum (Prof. W. Fenical, Scripps Institution of Oceanography) for access to normal-phase HPLC analysis and Dr. Constanze Paulus for preliminary experiments with *F. ambigua*. We thank Prof. Thomas Turner (UCSB) for marine sponge taxonomic analysis.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Chemoselective · ligation · chlorooximes · isonitriles · natural products · reactivity-based screening

- [1] M. S. Edenborough, R. B. Herbert, *Nat. Prod. Rep.* **1988**, *5*, 229–245.
- [2] M. J. Garson, J. S. Simpson, *Nat. Prod. Rep.* **2004**, *21*, 164–179.
- [3] J. Emsermann, U. Kaul, T. Opatz, *Mar. Drugs* **2016**, *14*, 16.
- [4] M. Schnermann, R. A. Shen, *Nat. Prod. Rep.* **2015**, *32*, 543–577.
- [5] R. M. Hohlmann, D. H. Sherman, *Nat. Prod. Rep.* **2021**, *38*, 1567–1588.
- [6] A. Massarotti, F. Brunelli, S. Aprile, M. Giustiniano, G. C. Tron, *Chem. Rev.* **2021**, *121*, 10742–10788.
- [7] For an early example, see: W. Rothe, *Dtsch. Med. Wochenschr.* **1949**, *79*, 1080–1081.
- [8] C. L. Cox, J. I. Tietz, K. Sokolowski, J. O. Melby, J. R. Doroghazi, D. A. Mitchell, *ACS Chem. Biol.* **2014**, *9*, 2014–2022.
- [9] For a recent review, see: C. C. Hughes, *Nat. Prod. Rep.* **2021**, *38*, 1684–1705.
- [10] For examples, see: a) C. Ross, K. Scherlach, F. Kloss, C. Hertweck, *Angew. Chem. Int. Ed.* **2014**, *53*, 7794–7798; b) H. Jeon, C. Lim, J. M. Lee, S. Kim,

Chem. Sci. **2015**, *6*, 2806–2811; c) S. D. Schnell, L. V. Hoff, A. Panchagnula, M. H. H. Wurzenberger, T. M. Klapötke, S. Sieber, A. Linden, K. Gademann, *Chem. Sci.* **2020**, *11*, 3042–3047; d) G. C. Rudolf, M. F. Koch, F. A. M. Mandl, S. A. Sieber, *Chem. Eur. J.* **2015**, *21*, 3701–3707.

- [11] For examples, see: a) Q. Jiang, Z. Zhu, P. Shou, F. Teng, Y. Zhu, H. Zhao, B. Yang, *Phytochem. Anal.* **2020**, *31*, 322–332; b) G. Castro-Falcón, D. Hahn, D. Reimer, C. C. Hughes, *ACS Chem. Biol.* **2016**, *11*, 2328–2336; c) T. Maxson, J. I. Tietz, G. A. Hudson, X. R. Guo, H.-C. Tai, D. A. Mitchell, *J. Am. Chem. Soc.* **2016**, *138*, 15157–15166; d) D. Back, B. T. Shaffer, J. E. Loper, B. Philmus, *J. Nat. Prod.* **2022**, *85*, 105–114; e) G. Castro-Falcón, N. Millán-Aguinaga, C. Roullier, P. R. Jensen, C. C. Hughes, *ACS Chem. Biol.* **2018**, *13*, 3097–3106.
- [12] Y.-B. Huang, W. Cai, A. Del Rio Flores, F. F. Twigg, W. Zhang, *Anal. Chem.* **2020**, *92*, 599–602.
- [13] a) H. Stöckmann, A. A. Neves, S. Stairs, K. M. Brindle, F. J. Leeper, *Org. Biomol. Chem.* **2011**, *9*, 7303–7305; b) S. Stairs, A. A. Neves, H. Stöckmann, Y. A. Wainman, H. Ireland-Zecchini, K. M. Brindle, F. J. Leeper *ChemBioChem* **2013**, *14*, 1063–1067. For a recent report introducing a more stable product, see: J. Tu, D. Svatunek, S. Parvez, A. C. Liu, B. J. Levandowski, H. J. Eckvahl, R. T. Peterson, K. N. Houk, R. M. Franzini, *Angew. Chem. Int. Ed.* **2019**, *58*, 9043–9048; *Angew. Chem.* **2019**, *131*, 9141–9146.
- [14] R. J. B. Schäfer, M. R. Monaco, M. Li, A. Tirla, P. Rivera-Fuentes, H. Wennemers, *J. Am. Chem. Soc.* **2019**, *141*, 18644–18648.
- [15] T. Soeta, S. Takashita, Y. Sakata, Y. Ukaji, *Org. Biomol. Chem.* **2016**, *14*, 694–700.
- [16] K. K. Palaniappan, A. A. Pitcher, B. P. Smart, D. R. Spicariach, A. T. lavarone, C. R. Bertozzi, *ACS Chem. Biol.* **2011**, *6*, 829–836.
- [17] For a previous use of probe 1, see: B. C. Sanders, F. Friscourt, P. A. Ledin, N. E. Mbua, S. Arumugam, J. Guo, T. J. Boltje, V. V. Popik, G.-J. Boons, *J. Am. Chem. Soc.* **2011**, *133*, 949–957.
- [18] a) A. Raveh, S. Carmeli, *J. Nat. Prod.* **2007**, *70*, 196–201; b) S. Mo, A. Krunic, G. Chlipala, J. Orjala, *J. Nat. Prod.* **2009**, *72*, 894–899.
- [19] T. A. Smitka, R. Bonjouklian, L. Doolin, N. D. Jones, J. B. Deeter, W. Y. Yoshida, M. R. Prinsep, R. E. Moore, G. M. L. Patterson, *J. Org. Chem.* **1992**, *57*, 857–861.
- [20] J. E. Thompson, R. P. Walker, D. J. Faulkner, *Mariani Found. Paediatr. Neurol. Ser.* **1985**, *21*, 11–21.
- [21] <https://chemdata.nist.gov>.
- [22] Reactions in THF/H₂O (1:1) gave comparable results, see the Supporting Information for details.
- [23] Of note, conjugation with probe 1 increases the detectability markedly (~200-fold; Figure S12). Under the chromatography conditions, degradation products of probe 1 emerge, see also Figure S11.
- [24] H. Tada, F. Yasuda, *Chem. Pharm. Bull.* **1985**, *33*, 1941–1945.
- [25] a) R. P. Walker, *The Chemical Ecology of Some Sponges and Nudibranchs from San Diego*, Dissertation University of California, San Diego, **1981**; b) J. E. Thompson, R. P. Walker, S. J. Wratten, D. J. Faulkner, *Tetrahedron* **1982**, *38*, 1865–1873.
- [26] "Marine Isocyanate Compounds", C. W. J. Chang, P. J. Scheuer in *Studiecent. TNO Scheepsbouw Navig. Rep.* (Ed.: P. J. Scheuer), Springer, Berlin, **1993**, pp. 33–75.
- [27] a) P. Ciminiello, E. Fattorusso, S. Magno, L. Mayol, *Experientia* **1986**, *42*, 625–627; b) R. J. Capon, J. K. MacLeod, *Aust. J. Chem.* **1988**, *41*, 979–983; c) V. Kurnianda, H. Hirade, R. Jansen, J. Tanaka *J. Asian Nat. Prod. Res.* **2022**, *24*, 39–44.
- [28] The INC ligation can be performed in water/acetonitrile mixtures, highlighting the chemoselectivity of chlorooximes for isonitriles over nitriles.
- [29] Note, chlorooximes can also react with thiols (ref. [14]). Thus, this signature can be particularly valuable when isonitriles occur in a mixture with thiols.

Manuscript received: October 20, 2022

Accepted manuscript online: November 4, 2022

Version of record online: December 8, 2022

Chemistry–A European Journal

Supporting Information

Identification of Isonitrile-Containing Natural Products in Complex Biological Matrices through Ligation with Chlorooximes

Rebecca J. B. Schäfer, Kayla Wilson, Maurice Biedermann, Bradley S. Moore,* Simon Sieber,* and Helma Wennemers*

Table of content

1. GENERAL ASPECTS AND MATERIALS.....	S3
2. SYNTHESIS AND ANALYTICAL DATA OF PROBE 1	S4
3. ISONITRILE NATURAL PRODUCT DETECTION IN CYANOBACTERIUM F. AMBIGUA UTEX 1903.....	S7
Cultivation and extraction of <i>F. ambigua</i>	S7
Labeling protocol of cyanobacterial extract.....	S7
UHPLC-MS, HR-ESI-MS and HPLC analysis	S7
MS spectra and extracted ion chromatograms.....	S7
4. ISONITRILE NATURAL PRODUCT DETECTION IN MARINE SPONGE AXINELLA SP.....	S17
Sponge collection and extraction.....	S17
GCMS analysis of sponge <i>Axinella</i> sp. extract.....	S17
Labeling protocol of sponge crude extract for MS analysis	S19
HPLC-MS analysis.....	S19
Degradation products of probe 1	S20
Labeling protocol for isolation of labeled isonitrile natural products	S21
Difference in MS-detectability of non-conjugated isonitrile terpenoids and isonitrile terpenoids ligated with probe 1	S21
NMR spectroscopic analysis of sponge-derived sesquiterpene-conjugate 2	S22
NMR spectroscopic analysis of sponge-derived sesquiterpene-conjugate 4	S23
5. COMPETITION EXPERIMENTS BETWEEN ISONITRILES, ISOCYANATES, AND ISOTHIOCYANATES FOR PROBE 1	S24
6. MS/MS FRAGMENTATION OF HYDROXYIMINO AMIDE PRODUCTS	S26
MS/MS analysis of conjugated sponge natural products	S27
MS/MS analysis of conjugate S3 and proposed MS/MS fragmentation mechanism	S28
7. NMR SPECTRA OF NATURAL PRODUCT CONJUGATES AND COMPOUND S3	S29
Compound 2	S30
Compound 4	S32
Compound S3	S36
8. REFERENCES	S37

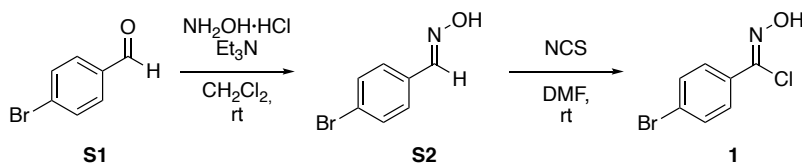
1. General aspects and materials

All chemicals were purchased from Fisher Scientific, Alfa Aesar, or MilliporeSigma and used without further purification. All solvents were of HPLC grade or higher. Preparative flash column chromatography was carried out on a Teledyne ISCO Combiflash® Rf+ Lumen™ system using liquid loading and silica gel (EMD, 40-63µm) for the stationary phase. Normal phase semi-preparative HPLC purification was carried out using a Shimadzu SCL-10A system and a Shimadzu SPD-M10A UV/Vis detector (Shimadzu Corp., Kyoto Japan). Preparative TLC purifications were carried out using TLC Silica gel 60 F254 glass plates (20x20cm, MilliporeSigma™ Supelco™).

NMR spectroscopic data were obtained either on a 500 MHz JEOL NMR spectrometer with a 3.0 mm probe, a 600 MHz Bruker NMR spectrometer with a 1.7 mm cryoprobe, or a 500 MHz Varian NMR spectrometer with a 5 mm, ¹³C optimized cold probe. The values of the chemical shifts are listed in ppm and coupling constants are reported in Hz. NMR chemical shifts were referenced to the residual solvent peaks (*d*_H 7.26 and *d*_C 77.16).

High performance liquid chromatography high resolution mass spectrometry (HPLC-HRMS) analysis was conducted on an Agilent 6530 Accurate-Mass Q-TOF MS (MassHunter software, Agilent) equipped with a dual electrospray ionization (ESI) source and an Agilent 1260 LC system (ChemStation software, Agilent) with a diode array detector. Q-TOF MS settings and LC gradients are indicated in the respective experimental sections of the supplementary information.

2. Synthesis and analytical data of probe 1



Scheme S1. Synthesis of chlorooxime probe 1.

Oxime S2: 4-bromobenzaldehyde (**S1**) (1.0 g, 5.4 mmol, 1.0 equiv.) was dissolved in CH_2Cl_2 (30 mL). Hydroxylamine hydrochloride (751 mg, 10.8 mmol, 2.0 equiv.) was added, followed by NEt_3 (1.51 mL, 10.8 mmol, 2.0 equiv.). The reaction mixture was stirred at room temperature until complete consumption of starting material was observed by TLC. The crude reaction mixture was diluted with one reaction volume of water. Phases were separated and the aqueous phase was extracted three times with CH_2Cl_2 . The combined organic extracts were washed with brine, dried over MgSO_4 , filtered, and concentrated under reduced pressure. The crude product was purified by automated normal phase column chromatography (24 g cartridge, 0–45% EtOAc over 15 min). Oxime **S2** was obtained in 90% yield (980 mg, 4.9 mmol) as a white powder. The analytical data correspond to the reported values in the literature.^[1]

¹H-NMR (500 MHz, CDCl_3) δ [ppm] = 8.09 (s, 1H), 7.91 (s, 1H), 7.54–7.50 (m, 2H), 7.45–7.43 (m, 2H). ¹³C-NMR (121 MHz, CDCl_3) δ [ppm] = 149.6, 132.2, 130.9, 128.6, 124.5.

TLC: Hex:EtOAc (4:1), R_F = 0.5 and 0.25 (*E* and *Z* isomers, respectively)

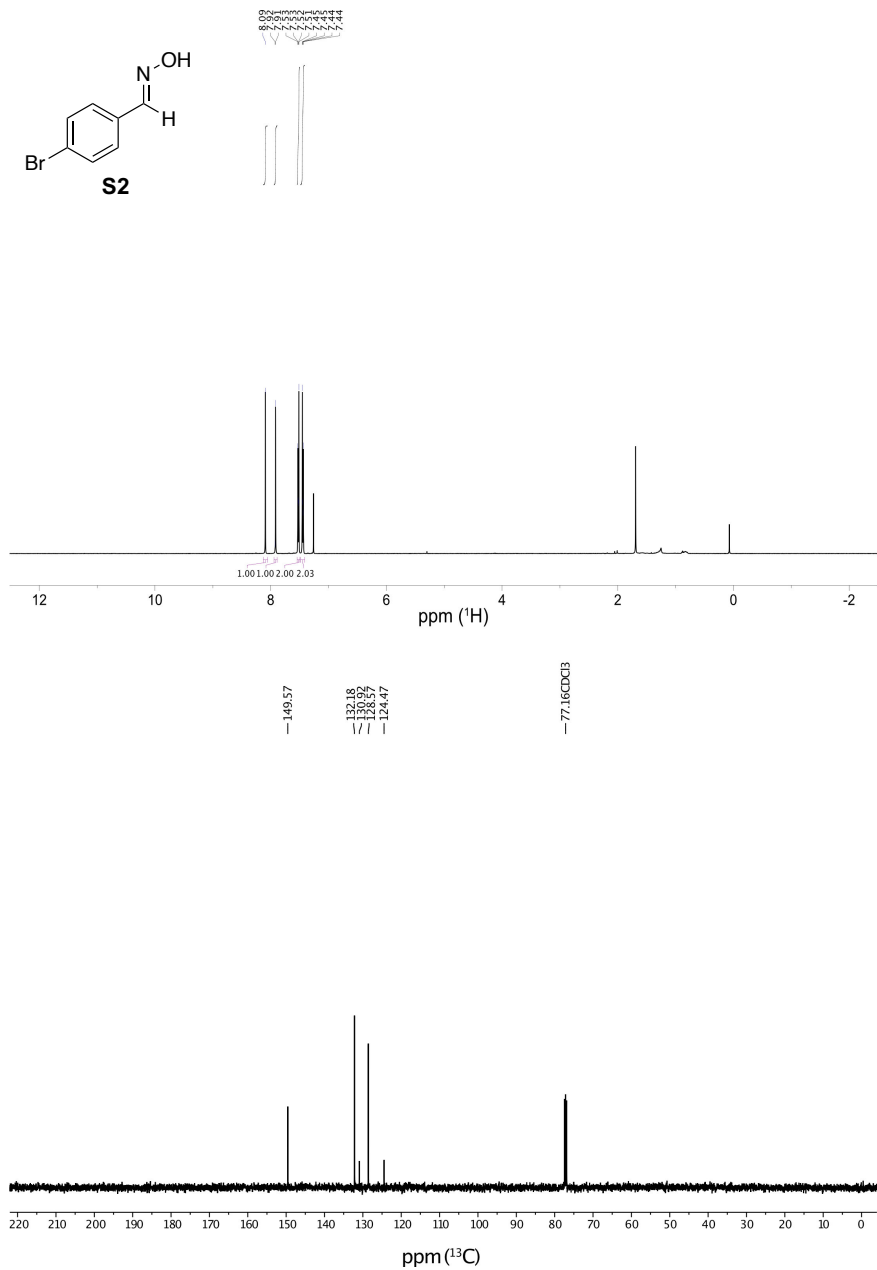
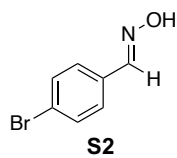
Chlorooxime 1: 4-bromobenzaldoxime (**S2**) (400 mg, 2.0 mmol, 1.0 equiv.) was dissolved in DMF (1.5 mL). *N*-chlorosuccinimide (280 mg, 2.1 mmol, 1.05 equiv.) was added in small portions over 2 min at room temperature. The reaction mixture was stirred at room temperature until complete consumption of starting material was observed by TLC. The crude reaction mixture was diluted with EtOAc (6 mL) and H_2O (8 mL). Phases were separated and the aqueous phase was extracted three times with EtOAc (3 mL). The combined organic extracts were then washed five times with 5% LiCl_{aq} solution (10 mL each). The organic phase was washed with brine, dried over MgSO_4 , filtered, and concentrated under reduced pressure to complete dryness. Product **1** was obtained as a white crystalline powder in 89% yield (420 mg, 1.79 mmol). The analytical data correspond to the reported values in the literature.^[2]

¹H-NMR (500 MHz, CDCl_3) δ [ppm] = 8.04 (s, 1H), 7.74–7.69 (m, 2H), 7.56–7.52 (m, 2H).

¹³C-NMR (121 MHz, CDCl_3) δ [ppm] = 139.5, 131.9, 131.4, 128.8, 125.5.

TLC: Hex:EtOAc (4:1) R_F = 0.71

^1H NMR (CDCl_3 , 500 MHz) and ^{13}C NMR (CDCl_3 , 121 MHz) spectra of oxime S2



1

O=[N]C(Cl)=c1ccc(Br)cc1

8.22
7.70
7.69
7.68
7.67
7.66
7.65
7.64
7.63
7.62
7.61
7.60
7.59
7.58
7.57
7.56
7.55
7.54
7.53
7.52
7.51
7.50
7.49
7.48
7.47
7.46
7.45
7.44
7.43
7.42
7.41
7.40
7.39
7.38
7.37
7.36
7.35
7.34
7.33
7.32
7.31
7.30
7.29
7.28
7.27
7.26
7.25
7.24
7.23
7.22
7.21
7.20
7.19
7.18
7.17
7.16
7.15
7.14
7.13
7.12
7.11
7.10
7.09
7.08
7.07
7.06
7.05
7.04
7.03
7.02
7.01
7.00
6.99
6.98
6.97
6.96
6.95
6.94
6.93
6.92
6.91
6.90
6.89
6.88
6.87
6.86
6.85
6.84
6.83
6.82
6.81
6.80
6.79
6.78
6.77
6.76
6.75
6.74
6.73
6.72
6.71
6.70
6.69
6.68
6.67
6.66
6.65
6.64
6.63
6.62
6.61
6.60
6.59
6.58
6.57
6.56
6.55
6.54
6.53
6.52
6.51
6.50
6.49
6.48
6.47
6.46
6.45
6.44
6.43
6.42
6.41
6.40
6.39
6.38
6.37
6.36
6.35
6.34
6.33
6.32
6.31
6.30
6.29
6.28
6.27
6.26
6.25
6.24
6.23
6.22
6.21
6.20
6.19
6.18
6.17
6.16
6.15
6.14
6.13
6.12
6.11
6.10
6.09
6.08
6.07
6.06
6.05
6.04
6.03
6.02
6.01
6.00
5.99
5.98
5.97
5.96
5.95
5.94
5.93
5.92
5.91
5.90
5.89
5.88
5.87
5.86
5.85
5.84
5.83
5.82
5.81
5.80
5.79
5.78
5.77
5.76
5.75
5.74
5.73
5.72
5.71
5.70
5.69
5.68
5.67
5.66
5.65
5.64
5.63
5.62
5.61
5.60
5.59
5.58
5.57
5.56
5.55
5.54
5.53
5.52
5.51
5.50
5.49
5.48
5.47
5.46
5.45
5.44
5.43
5.42
5.41
5.40
5.39
5.38
5.37
5.36
5.35
5.34
5.33
5.32
5.31
5.30
5.29
5.28
5.27
5.26
5.25
5.24
5.23
5.22
5.21
5.20
5.19
5.18
5.17
5.16
5.15
5.14
5.13
5.12
5.11
5.10
5.09
5.08
5.07
5.06
5.05
5.04
5.03
5.02
5.01
5.00
4.99
4.98
4.97
4.96
4.95
4.94
4.93
4.92
4.91
4.90
4.89
4.88
4.87
4.86
4.85
4.84
4.83
4.82
4.81
4.80
4.79
4.78
4.77
4.76
4.75
4.74
4.73
4.72
4.71
4.70
4.69
4.68
4.67
4.66
4.65
4.64
4.63
4.62
4.61
4.60
4.59
4.58
4.57
4.56
4.55
4.54
4.53
4.52
4.51
4.50
4.49
4.48
4.47
4.46
4.45
4.44
4.43
4.42
4.41
4.40
4.39
4.38
4.37
4.36
4.35
4.34
4.33
4.32
4.31
4.30
4.29
4.28
4.27
4.26
4.25
4.24
4.23
4.22
4.21
4.20
4.19
4.18
4.17
4.16
4.15
4.14
4.13
4.12
4.11
4.10
4.09
4.08
4.07
4.06
4.05
4.04
4.03
4.02
4.01
4.00
3.99
3.98
3.97
3.96
3.95
3.94
3.93
3.92
3.91
3.90
3.89
3.88
3.87
3.86
3.85
3.84
3.83
3.82
3.81
3.80
3.79
3.78
3.77
3.76
3.75
3.74
3.73
3.72
3.71
3.70
3.69
3.68
3.67
3.66
3.65
3.64
3.63
3.62
3.61
3.60
3.59
3.58
3.57
3.56
3.55
3.54
3.53
3.52
3.51
3.50
3.49
3.48
3.47
3.46
3.45
3.44
3.43
3.42
3.41
3.40
3.39
3.38
3.37
3.36
3.35
3.34
3.33
3.32
3.31
3.30
3.29
3.28
3.27
3.26
3.25
3.24
3.23
3.22
3.21
3.20
3.19
3.18
3.17
3.16
3.15
3.14
3.13
3.12
3.11
3.10
3.09
3.08
3.07
3.06
3.05
3.04
3.03
3.02
3.01
3.00
2.99
2.98
2.97
2.96
2.95
2.94
2.93
2.92
2.91
2.90
2.89
2.88
2.87
2.86
2.85
2.84
2.83
2.82
2.81
2.80
2.79
2.78
2.77
2.76
2.75
2.74
2.73
2.72
2.71
2.70
2.69
2.68
2.67
2.66
2.65
2.64
2.63
2.62
2.61
2.60
2.59
2.58
2.57
2.56
2.55
2.54
2.53
2.52
2.51
2.50
2.49
2.48
2.47
2.46
2.45
2.44
2.43
2.42
2.41
2.40
2.39
2.38
2.37
2.36
2.35
2.34
2.33
2.32
2.31
2.30
2.29
2.28
2.27
2.26
2.25
2.24
2.23
2.22
2.21
2.20
2.19
2.18
2.17
2.16
2.15
2.14
2.13
2.12
2.11
2.10
2.09
2.08
2.07
2.06
2.05
2.04
2.03
2.02
2.01
2.00
1.99
1.98
1.97
1.96
1.95
1.94
1.93
1.92
1.91
1.90
1.89
1.88
1.87
1.86
1.85
1.84
1.83
1.82
1.81
1.80
1.79
1.78
1.77
1.76
1.75
1.74
1.73
1.72
1.71
1.70
1.69
1.68
1.67
1.66
1.65
1.64
1.63
1.62
1.61
1.60
1.59
1.58
1.57
1.56
1.55
1.54
1.53
1.52
1.51
1.50
1.49
1.48
1.47
1.46
1.45
1.44
1.43
1.42
1.41
1.40
1.39
1.38
1.37
1.36
1.35
1.34
1.33
1.32
1.31
1.30
1.29
1.28
1.27
1.26
1.25
1.24
1.23
1.22
1.21
1.20
1.19
1.18
1.17
1.16
1.15
1.14
1.13
1.12
1.11
1.10
1.09
1.08
1.07
1.06
1.05
1.04
1.03
1.02
1.01
1.00
0.99
0.98
0.97
0.96
0.95
0.9

3. Isonitrile natural product detection in cyanobacterium *Fischerella ambigua* UTEX 1903

Cultivation and extraction of *F. ambigua*

The cyanobacteria strain *Fischerella ambigua* UTEX 1903 was cultivated using a modified BG11 medium (without addition of NaNO₃ salt) for approximately 8 weeks. The culture (20 mL) was centrifuged (30 min, 4500 g) and the supernatant separated from the biomass. The biomass was extracted with MeOH (50 mL), the mixture was sonicated (3 x 3 min), centrifuged (30 min, 4500 g), the solvent was removed by decantation, and the process was repeated twice. The methanol extracts were combined, and the solvent was removed at a rotary evaporator at 40 °C. The crude extract (70 mg) was dissolved in MeOH (1 mL), the mixture was centrifuged (10 min, 10000 g), and the solution was analyzed by UHPLC-MS (1 µL).

Labeling protocol of cyanobacterial extract

All reactions were performed under stirring conditions in HPLC vials. In the first reaction, the chemoselective probe **1** (10 µL in THF, conc.: 1 mg/100 µL) was mixed with the cyanobacteria crude extract (40 µL, 70 mg/mL), and H₂O (50 µL) was added. The second reaction (control 1) contained only probe **1** (10 µL in THF, conc.: 1 mg/100 µL) and was mixed with THF (40 µL) and H₂O (50 µL). A second negative control (control 2) contained only the cyanobacteria crude extract (40 µL, 70 mg/mL) with THF (10 µL) and H₂O (50 µL). All reactions were stirred for 20 min at room temperature. The solvent was removed under a gentle nitrogen stream, and the residue was dissolved in MeOH (50 µL) and centrifuged (5 min, 10000 g). The samples were analyzed by UHPLC/HRMS. All reactions were performed in triplicates.

UHPLC-MS, HR-ESI-MS and HPLC analysis

HR-ESI-MS data were collected on an Ultimate 3000 UHPLC instrument (Thermo Fisher Scientific, Switzerland) coupled to a QExactive instrument (Thermo Fisher Scientific, Switzerland). A BEH C₁₈ column (150 x 2.1 mm, 1.7 µm, Phenomenex, USA) was used for the separation, the column oven was kept at 40 °C, the flow was set to 0.4 mL/min, and the volume of injection was 1 µL. The solvent system was composed of A (H₂O + 0.1 % FA), B (MeCN + 0.1 % FA) and the gradient starting from 3 % B. The gradient increased to 100% B in 12 min and the column was washed for 3 min.

MS Spectra and Extracted Ion Chromatograms

The following LCMS traces are extracted ion chromatograms (EICs) for either the natural product mass or the natural product probe conjugate mass of isonitriles from *Fischerella ambigua* UTEX 1903 followed by the corresponding HRMS spectrum of the highlighted signal.^[3] The boxes highlight the chemical structure and the corresponding EIC in each figure. To ensure that the detected masses derive from the conjugated products, traces of the control reactions were also extracted for the same mass. The absence of any signal in the respective control reaction corroborates ligation of probe **1** with the isonitrile natural product.

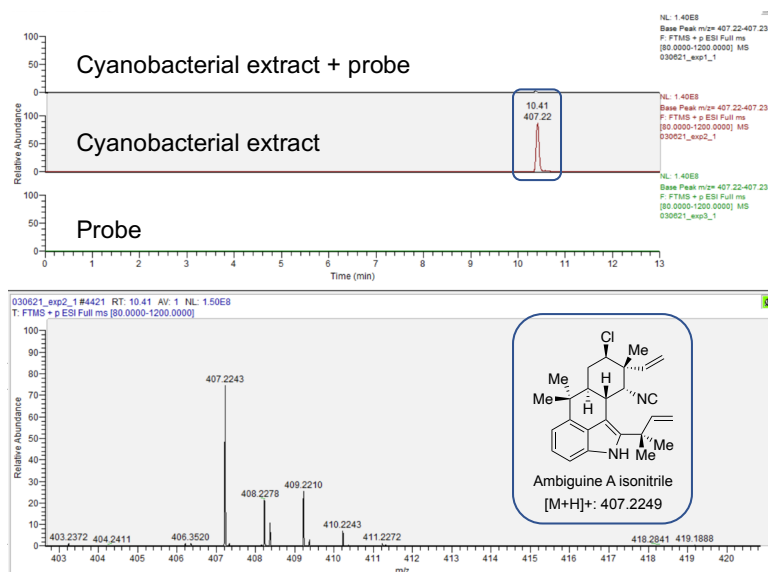


Figure S1a. Ambiguine A isonitrile (EIC m/z 407.22 – 407.23)

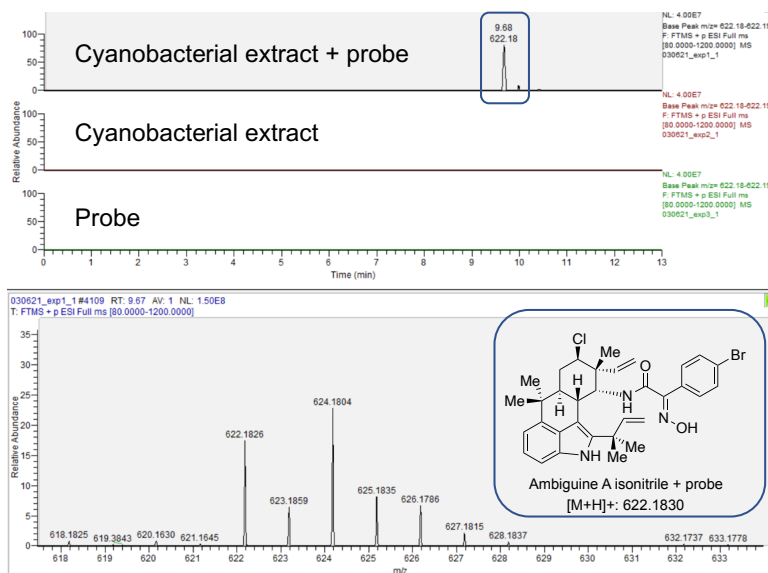


Figure S1b. Ambiguine A-probe 1 conjugate (EIC m/z 622.18 – 622.19)

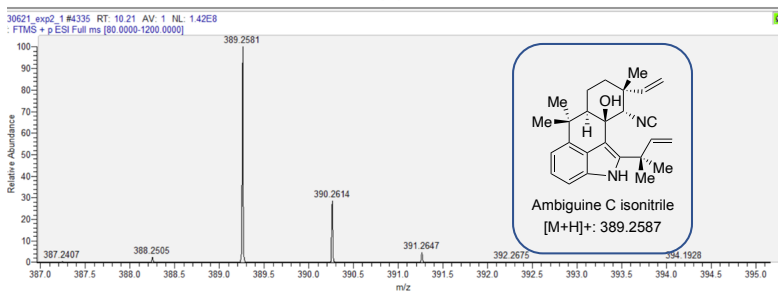
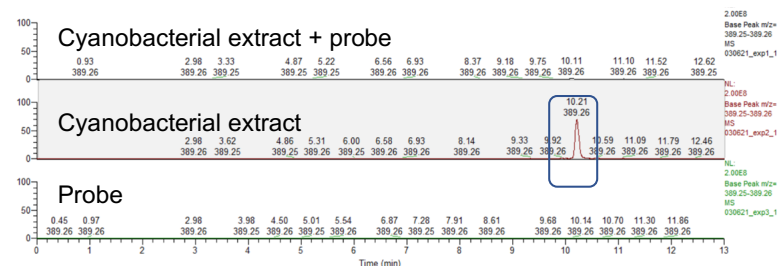


Figure S2a. Ambiguine C (EIC m/z 389.25 – 389.26)

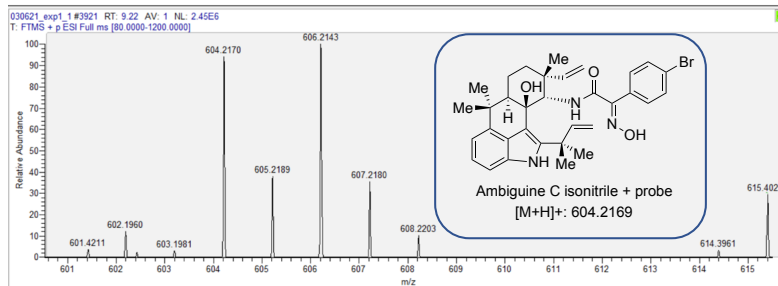
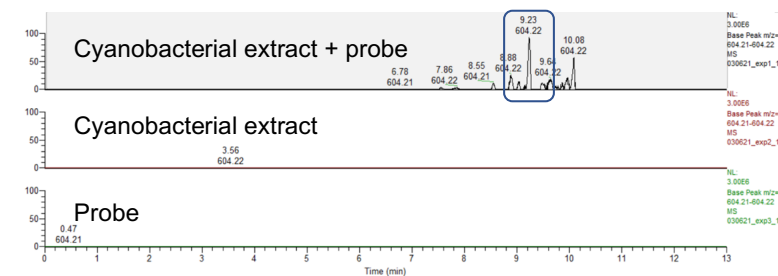


Figure S2b. Ambiguine C–probe 1 conjugate (EIC m/z 604.21 – 604.22)

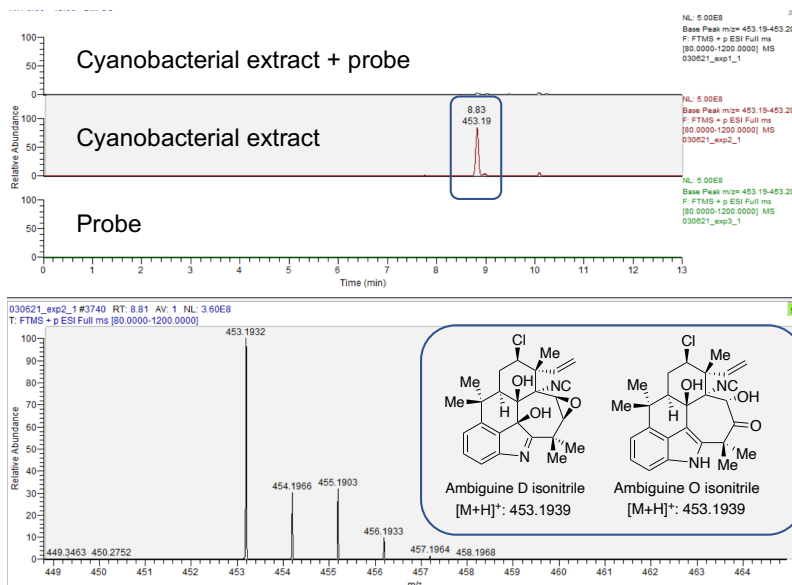


Figure S3a. Ambiguine D or O (EIC m/z 453.19– 453.20)

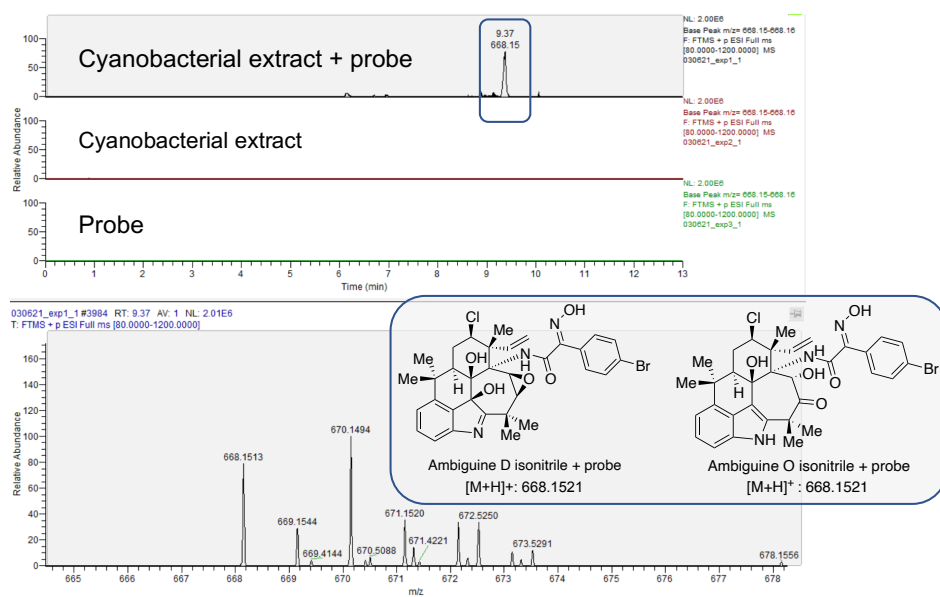


Figure S3b. Ambiguine D- or O- probe 1 conjugate (EIC m/z 668.15 – 668.16)

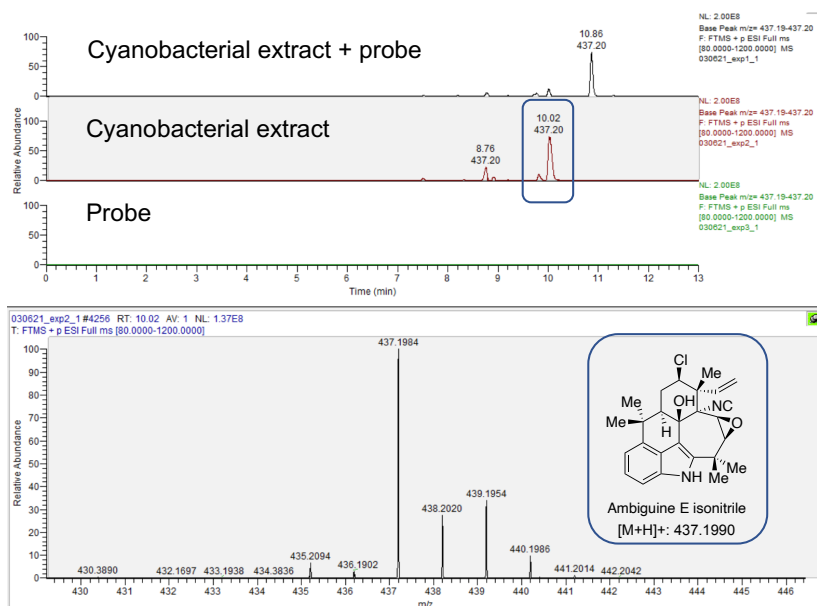


Figure S4a. Ambiguine E (EIC m/z 437.19 – 437.20)

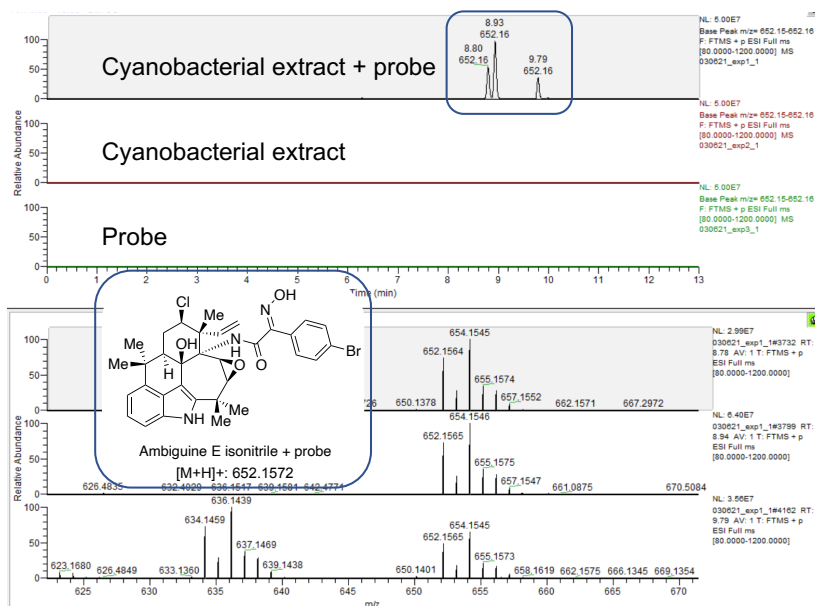


Figure S4b. Ambiguine E-probe 1 conjugate (EIC m/z 652.15 – 652.16)

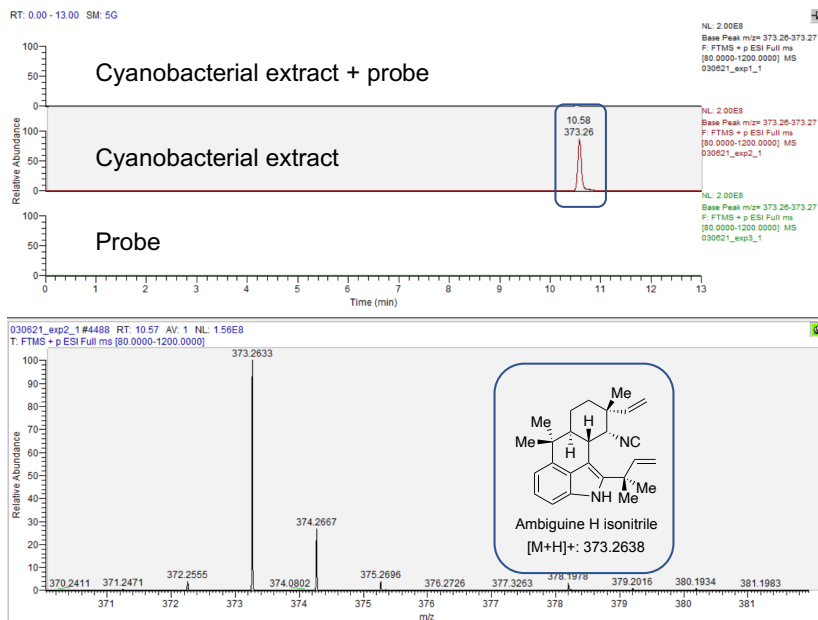


Figure S5a. Ambiguine H (EIC m/z 373.26 – 373.27)

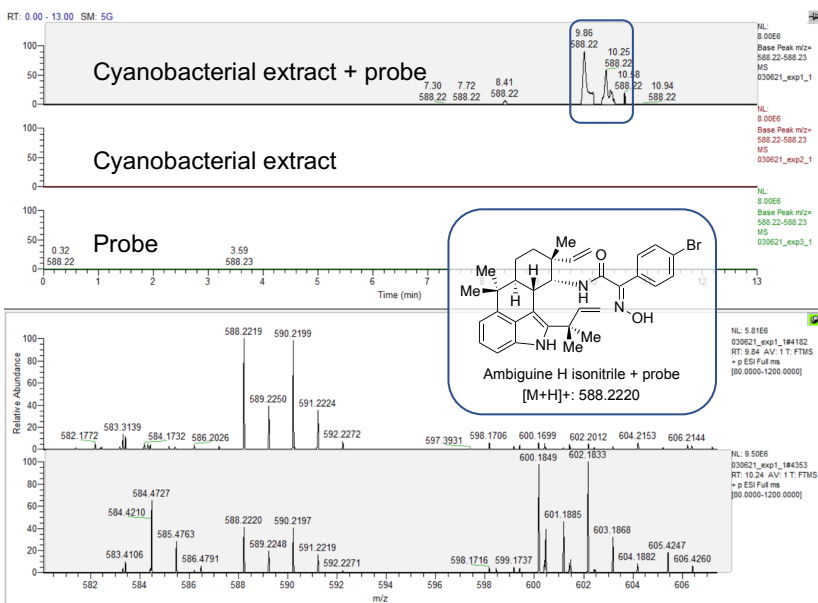
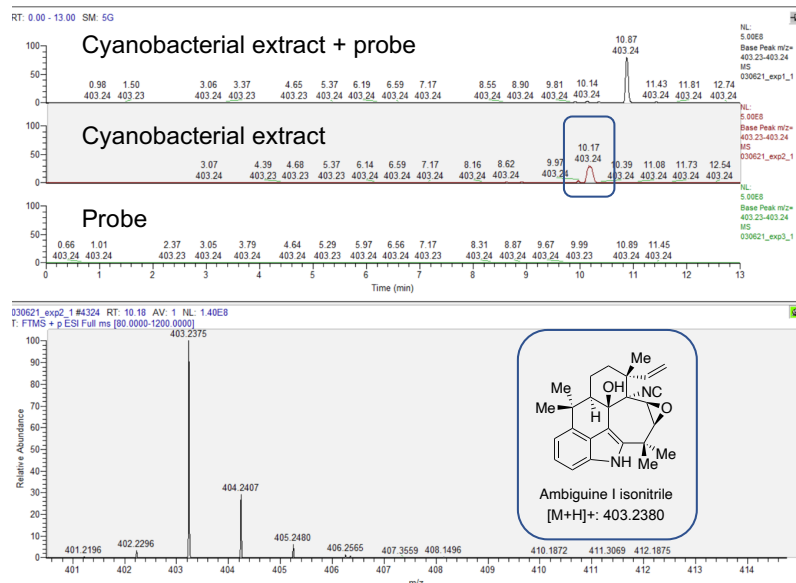


Figure S5b. Ambiguine H-probe 1 conjugate (EIC m/z 588.22 – 588.23)



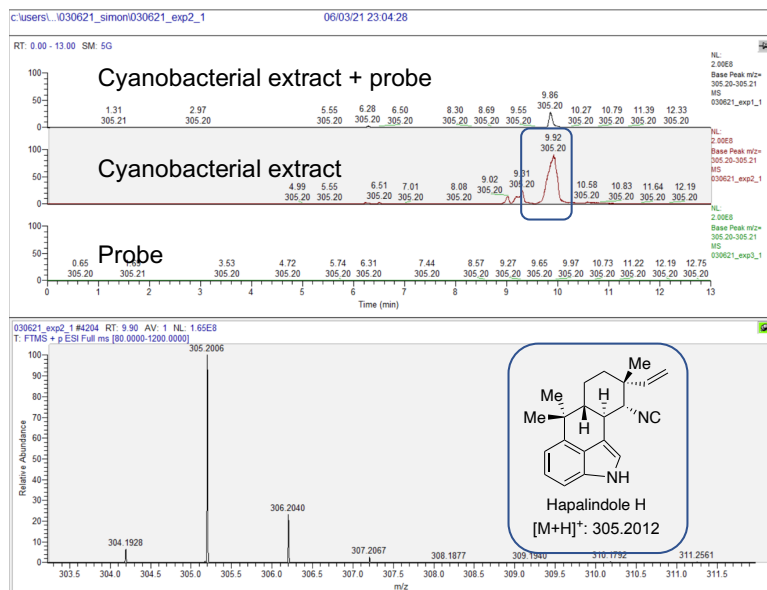


Figure S7a. Hapalindole H (EIC m/z 305.20 – 305.21 [M+H]⁺)

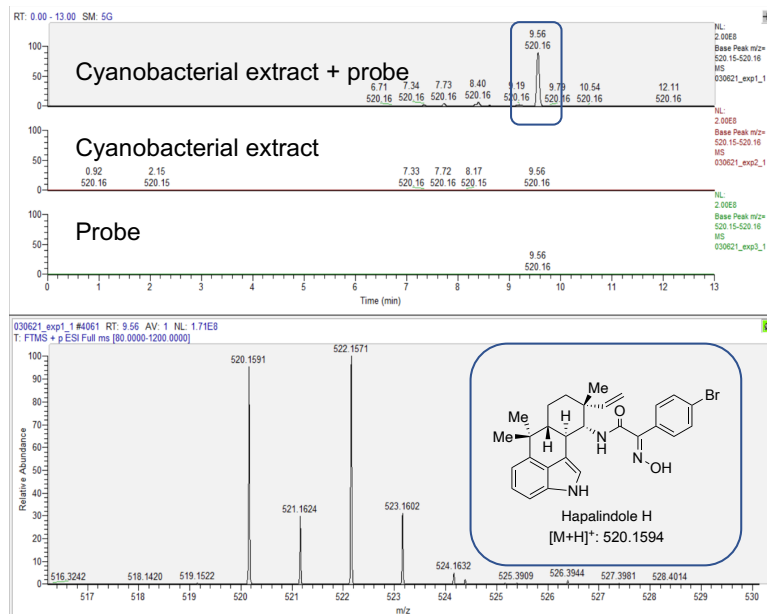


Figure S7b. Hapalindole H-probe 1 conjugate (EIC m/z 520.15 – 520.16)

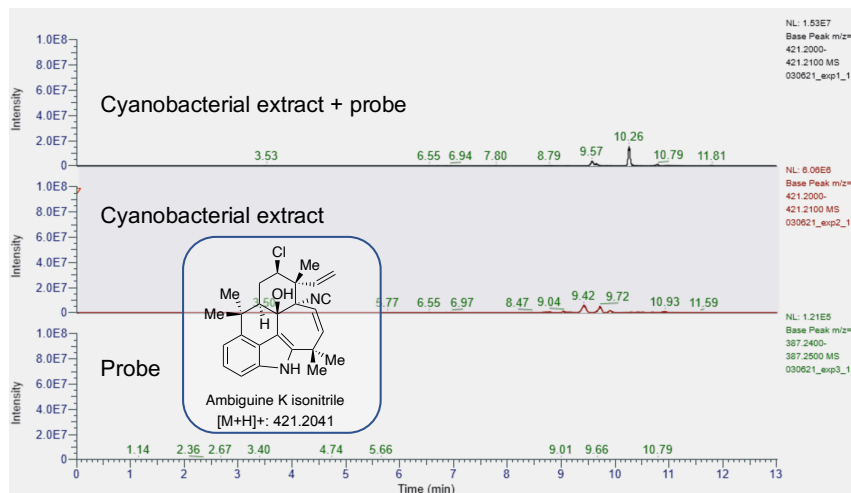


Figure S8a. Ambiguine K (EIC m/z 421.20 – 421.21)

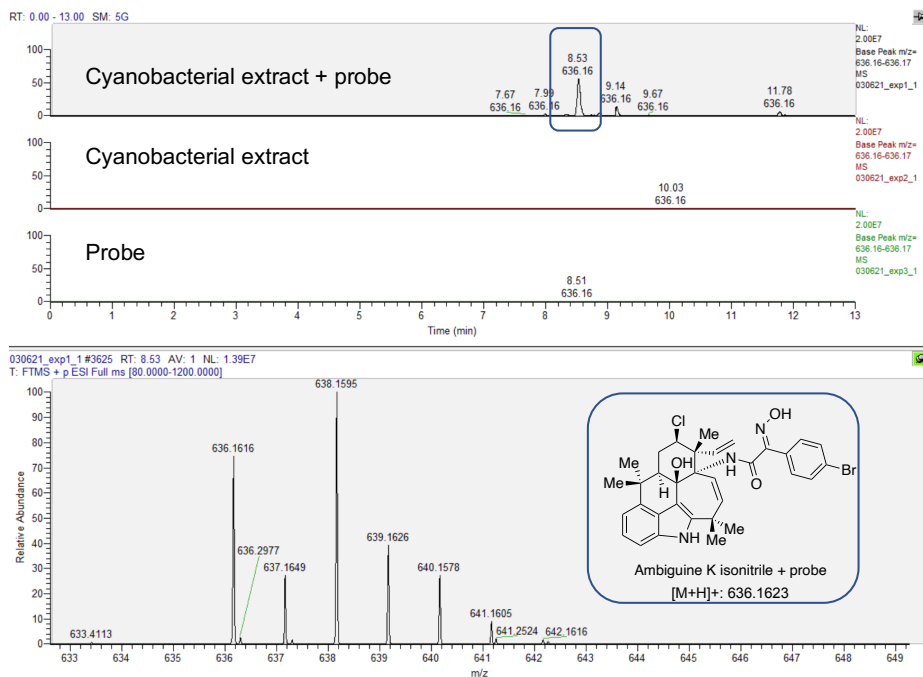


Figure S8b. Ambiguine K–probe 1 conjugate (EIC m/z 636.16 – 636.17)

4. Isonitrile natural product detection in marine sponge *Axinella sp.*

Sponge collection and extraction

Approximately 20 g (wet weight) of fresh *Axinella* sponge was collected by SCUBA from San Diego waters at approximately 15 meters depth in January 2019. The sponge was homogenized with liquid nitrogen and mortar and pestle. The resulting sponge powder was then extracted with CH₂Cl₂ (225 mL), vacuum filtered and carefully concentrated at 30 °C using a rotary evaporator to a final volume of 20 mL. The sponge extract was stored at –20°C.

GCMS analysis of sponge *Axinella sp.* extract

10 µL of the CH₂Cl₂ sponge extract was diluted in hexanes (40 µL) and 1 µL was analyzed on an Agilent 7890A gas chromatograph with Agilent 5975C mass spectrometer, using a HP-5MS 30 m × 0.25 mm, 0.25 µ column. Oven program was as follows: hold 70 °C for 3 min, 10 °C/min to 325 °C, hold at 325 °C for 3 min. Flow rate: 0.44 mL/min, Injection volume: 1 µL, Splitless injection, Inlet temperature: 200 °C, MS transfer line temperature: 250 °C, MS source temperature 230 °C, MS quad temperature: 150 °C

NIST mass spectral library (<https://chemdata.nist.gov>) was used as a resource to assist compound identification of the sponge extract by comparing our raw data with reference mass spectra and GC retention times deposited in the NIST mass database (Structures in Figure S10)

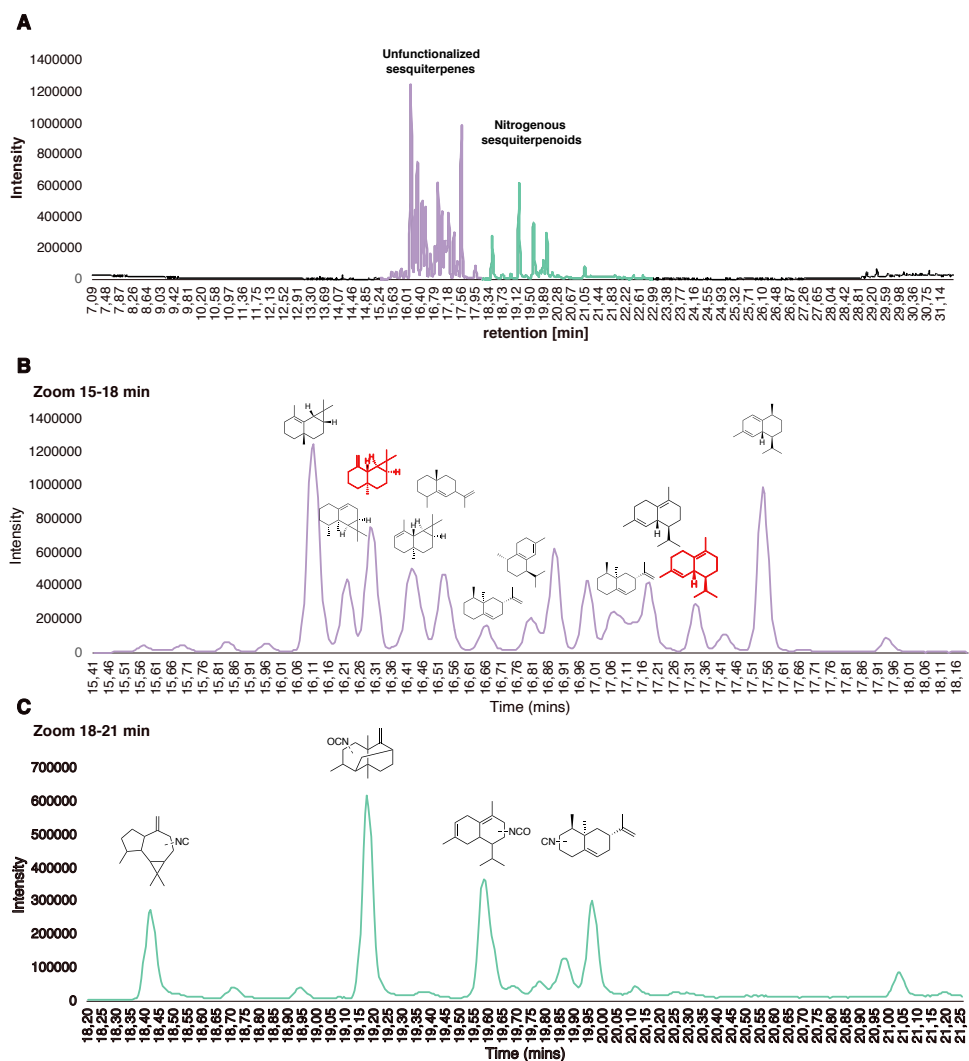


Figure S10. A) Full GC-MS chromatogram of the CH_2Cl_2 -extract of *Axinella* sp. with the region of unfunctionalized sesquiterpenes highlighted in purple and nitrogenous sesquiterpenoids highlighted in green. B) Zoom of unfunctionalized sesquiterpene region ($t = 15\text{--}18$ min). Chemical structures are putative annotations based on NIST GCMS library hits. C) Zoom of nitrogenous sesquiterpene region ($t = 18\text{--}21$ min). Chemical structures are putative annotations based on NIST GCMS library hits.

Labeling protocol of sponge crude extract for MS analysis

Preparation of stock solutions:

Sponge extract: 100 μL of CH_2Cl_2 sponge extract was carefully concentrated to dryness under reduced pressure and redissolved in THF (40 μL).

Probe: 1 mg of **1** was dissolved in THF (100 μL).

Labeling reaction with extract: 20 μL of the sponge extract stock solution was transferred to an HPLC vial and diluted with THF (20 μL). Then citrate buffer (50 μL , 55 mM, pH 5) was added, followed by 10 μL of a stock solution of **1**. The reaction was left standing at room temperature for 1 h.

Control 1 – only sponge extract: For control 1, 20 μL of the sponge extract solution were transferred to an HPLC vial and diluted with THF (30 μL). Then citrate buffer (50 μL , 55 mM, pH 5) was added. The reaction was left standing at room temperature for 1 h.

Control 2 – only probe **1**: For control 2, 20 μL of stock solution of **1** was transferred into an HPLC vial and diluted with THF (30 μL). Then citrate buffer (50 μL , 55 mM, pH 5) was added. The reaction was left standing at room temperature for 1 h.

Of note, the labeling reaction can be performed in THF:H₂O (1:1) (pH 6.5) or under buffered conditions in THF:citrate buffer (1:1) at pH 5 without noticeable differences in labeling efficiency. We observed a higher amount of hydrolyzed probe product under buffered conditions, which did, however, not impact the labeling reaction. In case of thiol-rich natural product extracts, we recommend buffered conditions, as thiols can react with the chlorooxime probe as competitive nucleophiles at neutral pH.^[6] Any undesired labeling with thiols can be easily differentiated through the absence of the diagnostic MS2 fragment **A**.

HPLC-MS analysis

HPLC samples were prepared by diluting 1 μL of each crude reaction mixture with acetonitrile (99 μL).

LC-MS analysis was performed with an Agilent Technologies 1260 Infinity series HPLC equipped with a degasser, binary pump, autosampler, and diode array detector coupled to an Agilent Technologies 6530 Accurate-Mass Q-TOF LC/MS. The instrument was calibrated using the Agilent Reference Calibration Mix®. Separations were performed with a Kinetex™ 5 μm C18 100 Å, 150 x 4.6 mm column. The solvent system contained water + 0.1% formic acid as solvent A, and acetonitrile + 0.1% formic acid as solvent B. Solvents were of mass spectrometry-grade. Data were collected and analyzed using MassHunter Workstation Software version B.05.01.

HPLC-HRMS method for MS1

0.75 mL min⁻¹ flow rate; gradient: 0-3 min 5% B, 3-23 min 5-100% B, 23-30 min 100% B, 30-37 min 5% B. DAD 190-600 nm. Dual ESI ion source, positive polarity. T = 300°C; Drying gas: 11 L/min;

V_{cap}: 3000 V. Acquisition mass range: 100-1700 m/z; rate: 2 spectra/sec, 500 ms/spectrum. Sample was prepared in acetonitrile; V_{inj}: 10 μL .

Degradation products of probe **1**

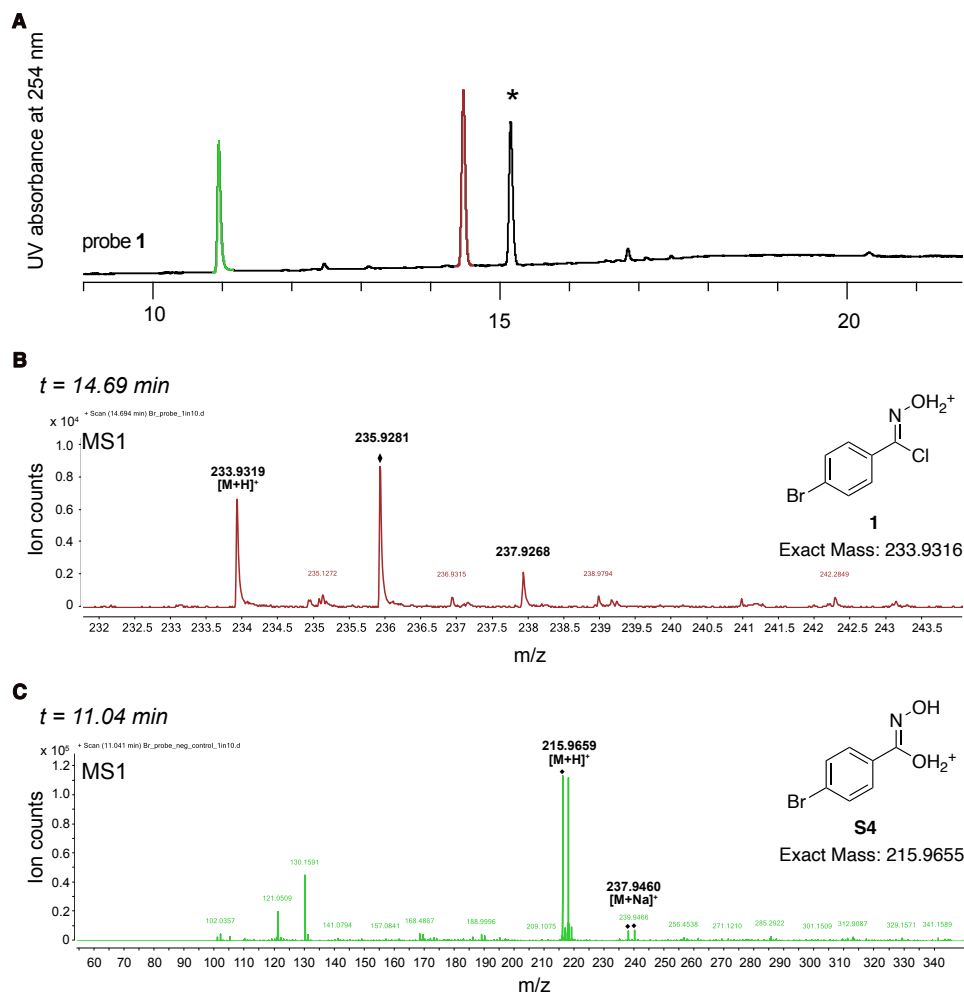


Figure S11. A) LC-UV chromatogram of negative control with probe **1** in THF:H₂O (1:1) shows three UV signals. Middle: MS spectrum derived from UV signal highlighted in red in LC-UV chromatogram shows a mass peak of m/z 233.9319 which corresponds to the mass of the proton-adduct of **1** ($M+H$)_{calc} 233.9316). C) MS spectrum derived from the UV signal highlighted in green in LC-UV chromatogram corresponds to the degradation product **S4** formed through hydrolysis of probe **1**. The origin of the signal indicated by an asterisk is not known.

Labeling protocol for isolation of labeled isonitrile natural products

4 mL of CH_2Cl_2 extract was concentrated to dryness under reduced pressure to give 12.9 mg of crude dried extract. The residual solid was resuspended in THF (940 μL) in a 10 mL screwcap vial. Citrate buffer (1 mL, 50 mM, pH 5) was added to the extract, followed by 180 μL of a freshly prepared stock solution of **1** (10 mg/mL THF). The labeling reaction was followed by TLC (4:1 Hex:EtOAc, R_f (**1**) = 0.71) by monitoring consumption of **1** and formation of new UV active spots. More stock solution was added in 15 min intervals until the probe was not consumed anymore, indicating complete labeling of putative isonitriles in the extract (approx. 1–2h). After completion of the reaction, NaCl_{sat} (2 mL) was added and the two phases were separated. The aqueous layer was extracted three times with EtOAc (3x2mL). The organic phases were combined and dried over MgSO_4 , filtered and concentrated to dryness under reduced pressure.

The crude extract was purified by automated normal phase column chromatography (4 g cartridge, 0–60% EtOAc in hexanes, 12 min). Two major peaks were collected (compound **2** and compound **4**). Compound **4** required further purification by normal phase semi-preparative HPLC (Luna® 5 μm Silica (2) 100 Å, LC column 100 x 10 mm; gradient: 0–60% EtOAc/hexanes, 20 min, elution time = 10.2 min). Compound **2** was used as is for NMR and HRMS analysis.

HRMS analysis of **2**: calc. 447.1647; found: 447.1626 $\Delta m/z$: 4.7 ppm

HRMS analysis of **4**: calc. 447.1647; found: 447.1669 $\Delta m/z$: 4.9 ppm

Difference in MS-detectability of non-conjugated isonitrile terpenoids and isonitrile terpenoids ligated with probe 1

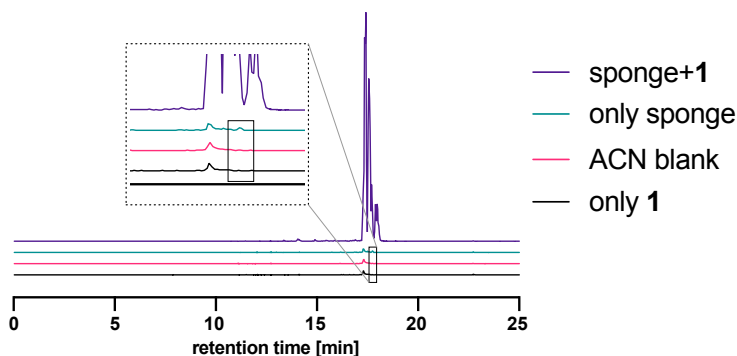


Figure S12. Extracted ion chromatograms of sponge extract + **1** (purple), only sponge extract (green), acetonitrile blank (red), and only probe **1** (black). All samples were extracted for the exact masses of sesquiterpene-probe conjugate **2** (m/z 447.165) and isonitrile sesquiterpenoid **3** (m/z 232.206). The MS ion count signal corresponding to the non-ligated terpene is highlighted in a grey box (green graph). The detectability of the isonitrile terpene natural product after ligation with probe **1** increased by approximately 200-fold.

A

B Key NOE interactions

1

2

Table S1. NMR spectroscopic data for sesquiterpenoid-conjugate **2** in CDCl₃ (500MHz)

Position	δ_c	δ_H (mult, J in Hz)	NOESY	COSY	HMBC
1	131.5	–	–	–	4, 5
2	132.3	7.55 (m)	–	4	1, 3, 6
3	132.3	7.55 (m)	–	5	1, 2, 6
4	129.8	7.43 (m)	–	2	5, 6, 7
5	129.8	7.43 (m)	–	3	4, 6, 7
6	124.6	–	–	–	2, 3, 4, 5
7	150.7	–	–	–	2, 3, 4, 5
8	161.9	–	–	–	9
9	–	5.86 (bs)	11, 21, 23, 24	–	8
10	59.4	–	–	–	11, 12a, 23, 24
11	19.2	1.52 (s)	12a, 16, 18a	–	12, 24, 10
12a	38.7	2.70 (m)	–	12b	11, 10, 24, 14
12b	38.7	1.62 (m)	–	12a, 14b	10, 11
13a	19.1	1.59 (m)	12a	–	12
13b	19.1	1.59 (m)	–	–	12
14a	39.4	1.37 (m)	–	14b	16, 17a, 17b
14b	39.4	1.01 (m)	–	13, 14a	16, 17a, 17b
15	32.9	–	–	–	16, 18, 24
16	19.5	0.93 (s)	11, 14, 18a	–	14, 15, 17, 24
17a	41.7	1.12 (dd, 13.4, 8.4)	14, 16	17b, 18a	15, 18, 19, 24
17b	41.7	0.73 (below 21)	–	17a, 18a, 18b	16, 18b, 18a, 19, 24
18a	15.5	1.79 (m)	16	17a, 17b, 18, 19	20, 23
18b	15.5	1.49 (m)	–	17b, 18a	15, 17, 20, 23
19	19.5	0.63 (dd, 8.9, 8.7)	–	18a, 23	17, 18, 22
20	17.7	–	–	–	18, 21, 22, 24
21	15.3	0.73 (s)	9, 18b, 24	–	19, 20, 22
22	28.7	0.78 (s)	9	–	20, 21, 23
23	19.8	0.37 (dd, 9.3, 6.6)	9, 11	19, 24	10, 22
24	47.0	1.04 (d, 6.6)	9, 21	12b, 14b, 23	10, 15, 17, 19, 20

NMR spectroscopic analysis of sponge-derived sesquiterpene-conjugate **4**

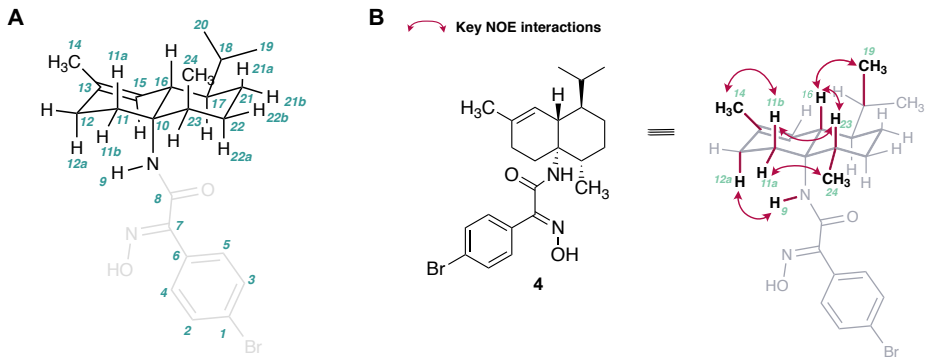


Figure S14. Left: Structure of **4** with numbered atoms. Right: Structure of **2** with relative stereochemistry, the absolute stereochemistry is not known. Red arrows indicate key NOEs.

Table S2. NMR spectroscopic data for sesquiterpenoid-conjugate **4** in CDCl₃ (500 MHz).

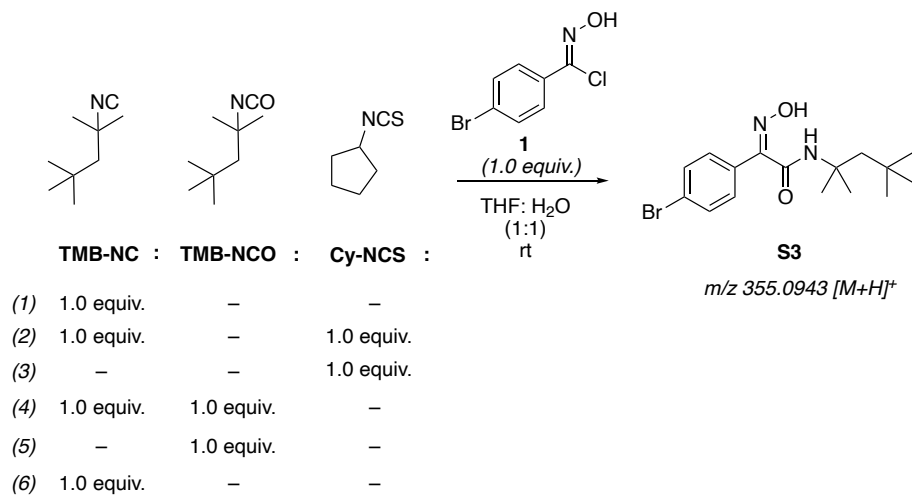
Position	δ_c	δ_H (mult, J in Hz)	NOESY	COSY	H2BC	HMBC
1	128.6	-	-	-	2, 3	4, 5
2	131.2	7.51 (m)	-	4	4	3, 5
3	131.2	7.51 (m)	-	5	5	2, 4
4	128.1	7.40 (m)	-	2	2	5, 1, 7
5	128.1	7.40 (m)	-	3	3	4, 1, 7
6	124.2	-	-	-	-	-
7	151.4	-	-	-	-	-
8	not found	-	-	-	-	-
9	-	5.55 (bs)	12a	-	16	21, 12, 16, 10
10	60.5	-	-	-	-	9, 12, 14, 24
11a	19.7	1.69 (m)	-	11b, 16	12	10, 12, 16
11b	19.7	1.92 (ddd, 13.9, 10.7, 7.4)	-	12b, 11a	12	12
12a	25.9	1.77 (m)	-	12b, 11b	11	11, 13
12b	25.9	2.04 (dd, 18.3, 7.3)	-	12a, 11b, 11a	11	10, 13
13	131.6	-	-	-	-	12a, 12b, 14, 16
14	22.6	1.68 (s)	11b	15	12	12, 13, 15
15	123.0	5.56 (m)	18	14, 16	16, 14	11, 12a, 12b, 16
16	40.7	2.86 (dd, 11.1, 5.9)	19, 23, 11b	14, 15, 17	17	11, 17, 10, 13, 15
17	47.2	1.27 (overlapping grease)	20	21b, 18, 16	16, 18	14, 18, 19, 20, 21a, 22b
18	26.3	1.98 (m)	-	19, 20, 16	19, 20	19, 20
19	14.6	0.85 (d, 6.9)	16	18	18	20
20	20.9	0.89 (d, 6.9)	-	18	18	19
21a	22.9	1.16 (m)	-	21b, 17	23, 17	18
21b	22.9	1.56 (dt, 9.4, 3.4, 3.4)	19	21a, 17	-	18
22a	29.9	1.17 (m)	-	23, 21a	17, 23	24
22b	29.9	1.63 (dt, 9.9, 3.9, 3.9)	-	22a, 23	23	24
23	36.3	2.59 (m)	16	22b, 21a, 24	22, 24	24, 21b, 22a, 22b
24	15.2	0.96 (d, 6.9)	11a	23	23	10, 22, 23

5. Competition experiments between isonitriles, isocyanates, isothiocyanates for probe 1

Isonitrile **TMB-NC** and either isothiocyanate **Cy-NCS** or isocyanate **TMB-NCO** (1.0 equiv.) were dissolved in THF (480 μ L). Then water (500 μ L) was added. Lastly, 10 μ L (2.0 equiv.) of a stock solution of **1** in THF (26 mg/100 μ L THF) was added and the reaction mixture was stirred for 1 h at room temperature (reactions 1–6, Scheme S2). Brine was added and the two phases were separated. The organic phase was dried over MgSO_4 , filtered, and concentrated under reduced pressure.

Comparison of all reactions by HPLC-UV analysis revealed that only reactions containing isonitrile **TMB-NC** and probe **1** (reaction (1), (2) and (4)) showed a new signal (Figure S14, $t = 20$ min). No new signal was observed in the control reactions and competition reactions with **TMB-NCO** and **Cy-NCS** (reactions (2), (3), (5) and (6)). Additionally, HRMS analysis confirmed formation of isonitrile-probe conjugate **S3** in reactions (1), (2) and (3) (Figure S15).

S3: $^1\text{H-NMR}$ (500 MHz, CDCl_3) δ [ppm] : 10.28 (bs, OH, 1H), 7.55-7.52 (m, 2H), 7.46-7.44 (m, 2H), 5.82 (s, NH, 1H), 1.75 (s, 2H), 1.51 (s, 6H), 0.99 (s, 3H). $^{13}\text{C-NMR}$ (121 MHz, CDCl_3) δ [ppm]: 162.01, 151.65, 132.16, 131.12, 129.26, 124.64, 56.97, 52.83, 31.81, 31.62, 28.84.



Scheme S2. Competition experiments between isonitrile **TMB-NC**, isocyanate **TMB-NCO** and isothiocyanate **Cy-NCS** for reaction with probe **1**.

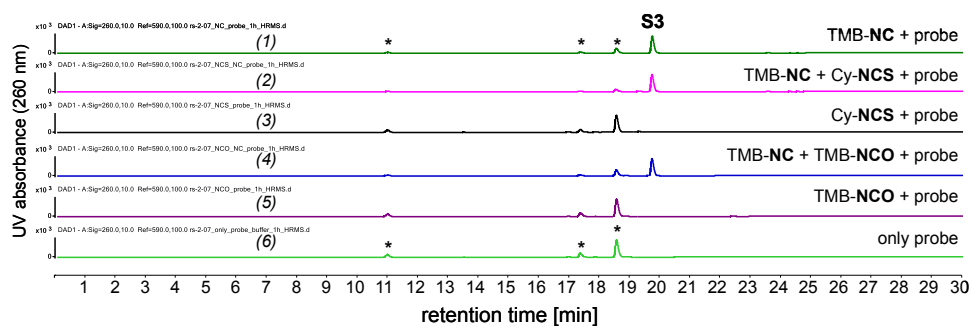


Figure S15. LC-UV chromatograms of reactions (1) to (6) with absorbance measured at 260 nm. **S3** indicated the desired conjugated product at $t = 20$ min. Stars indicate UV signals derived from LC-degradation products of probe 1 at $t = 18.7$, 17.4 min and 11.0 min, as discussed previously.

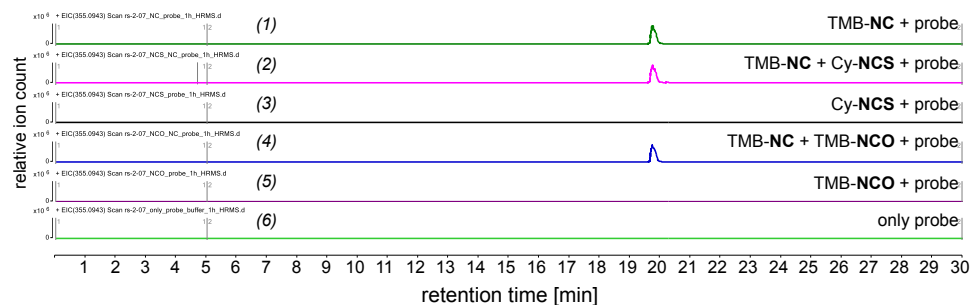


Figure S16. Extracted ion chromatograms (EICs) of reactions (1) to (6) extracted for the mass of conjugated product **S3** (m/z 355.0943 $[M+H]^+$).

6. MS/MS fragmentation of hydroxyimino amide products

HPLC samples were prepared, and LC-MS analysis was performed with the same equipment and solvents as described above (page S18).

HPLC-HRMS method for MS2

0.75 mL min⁻¹ flow rate; gradient: 0-3 min 5% B, 3-23 min 5-100% B, 23-30 min 100% B, 30-37 min 5% B. DAD 190-600 nm. Dual ESI ion source, positive polarity. T = 300°C; Drying gas: 11L/min; V_{cap}: 3000 V.

Acquisition for MS1: mass range: 100-1700 m/z, rate: 10 spectra/s, 100 ms/spectrum

Acquisition for MS2: mass range 50-1700 m/z, rate: 4 spectra/s, 250 ms/spectrum, E_{coll} = 5 V.

Injection volume, V_{inj}: 5 µL.

The MS/MS fragmentation patterns of the parent ions m/z 511.14, 447.16, 447.16 identified as the source of UV signals I, II, III contain a MS/MS signal with a mass-over-charge ratio of 242.98 [M+H] (Figure S16).

We hypothesized that this fragment could be a diagnostic for isonitriles that reacted with probe **1**. To test this hypothesis, we synthesized probe conjugate **S3** (page S23) from commercially available 1,1,4,4-tetramethylbutyl isocyanide (TMB-NC) and probe **1** and compared its MS2 fragmentation pattern with that of the sponge natural product probe conjugates.

S3 fragments into 4 major fragments: 1 (m/z 242.98), 2 (m/z 225.95), 3 (m/z 197.95), and 4 (m/z 181.96) (Figure S17), thus confirming that fragment **A** is a diagnostic MS2 fragment for isonitrile-probe conjugates. Scheme S3 shows a proposed mechanism for the MS/MS fragmentation of a conjugate between an isonitrile and probe **1**.

This diagnostic fragment, in combination with the UV absorbance and the bromine isotope pattern, allows for an unambiguous identification of isonitrile functionalized natural products.

MS/MS analysis of conjugated sponge natural products

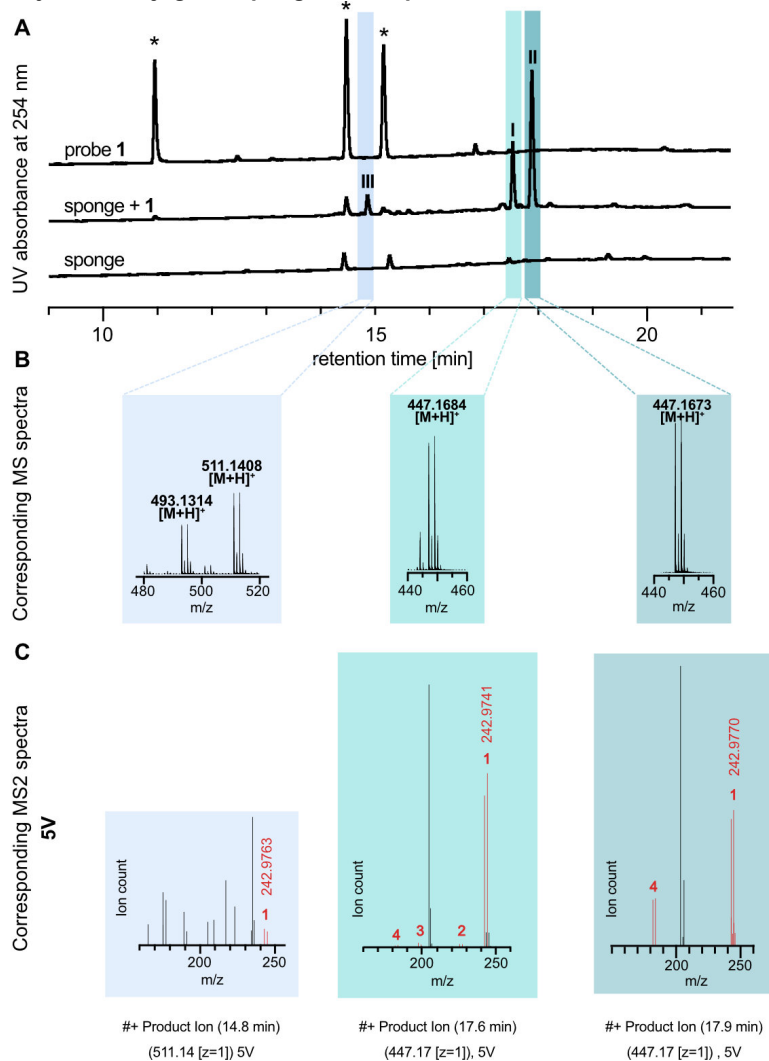


Figure S17. A) LC-UV analysis of only probe control (top), labeling reaction of sponge crude extract with probe 1 (middle), only sponge crude extract (bottom). Boxes highlight new UV signals I, II and III that are only observed in the labeling reaction. Asterisk indicate signals derived from probe 1. B) MS spectra corresponding to newly formed UV signals show the anticipated bromine isotope pattern. C) MS/MS fragmentation of the identified brominated product ions at a collision energy of 5V all produce diagnostic fragment A (indicated as 1). MS2 spectrum of II (light blue box) additionally produces fragment 2, 3 and 4. MS2 spectrum of I (dark blue box) additionally produces fragment 4.

MS/MS analysis of probe conjugate S3 and proposed fragmentation mechanism

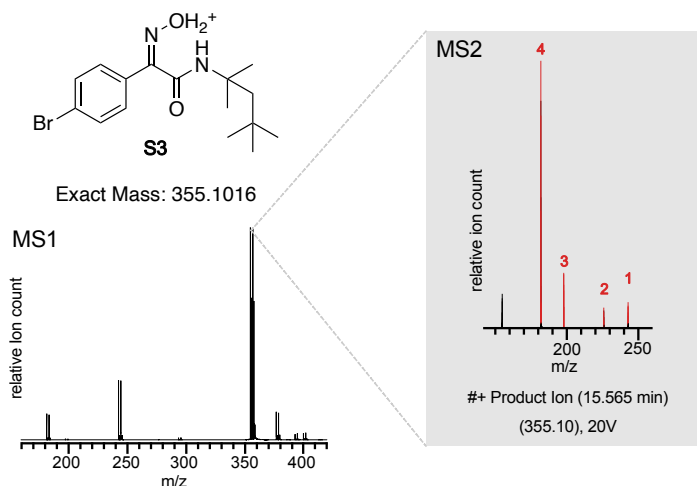
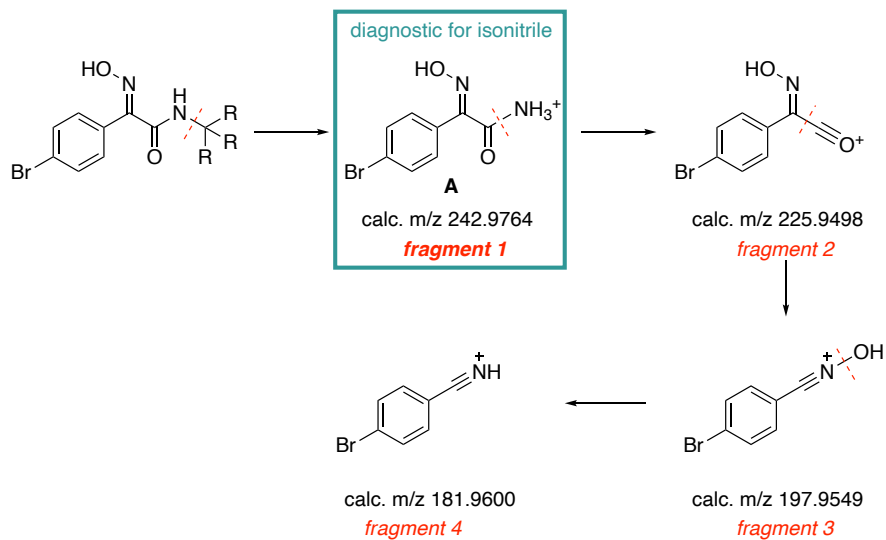


Figure S18. Left: Structure of **S3**, below extract of the MS1 spectrum of synthetic probe conjugate **S3** with the expected product mass peak at m/z 355.10. Right: MS/MS analysis of product ion m/z 355.10 generates MS fragments 1-4 with measured mass-to-charge ratios of 1 (m/z 242.9768), 2 (m/z 225.9504), 3 (m/z 197.9537) and 4 (m/z 181.9608).

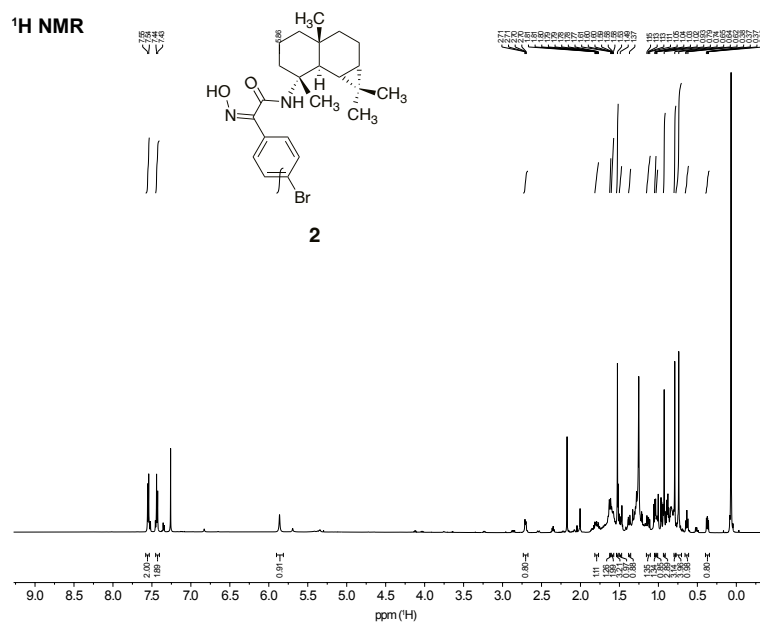


Scheme S3. Proposed fragmentation mechanism of isonitrile-probe conjugates with the calculated expected exact mass-over-charge- ratios (m/z) for each fragment. Formation of **A** (fragment 1, blue box) is diagnostic for the reaction of chlorooxime probe **1** with isonitriles.

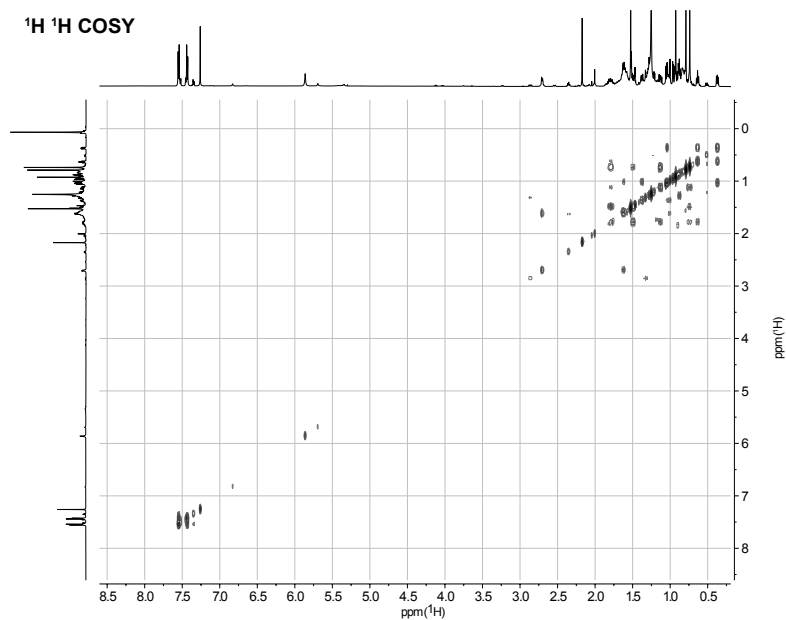
7. NMR spectra of natural product probe conjugates and compound S3

Compound 2

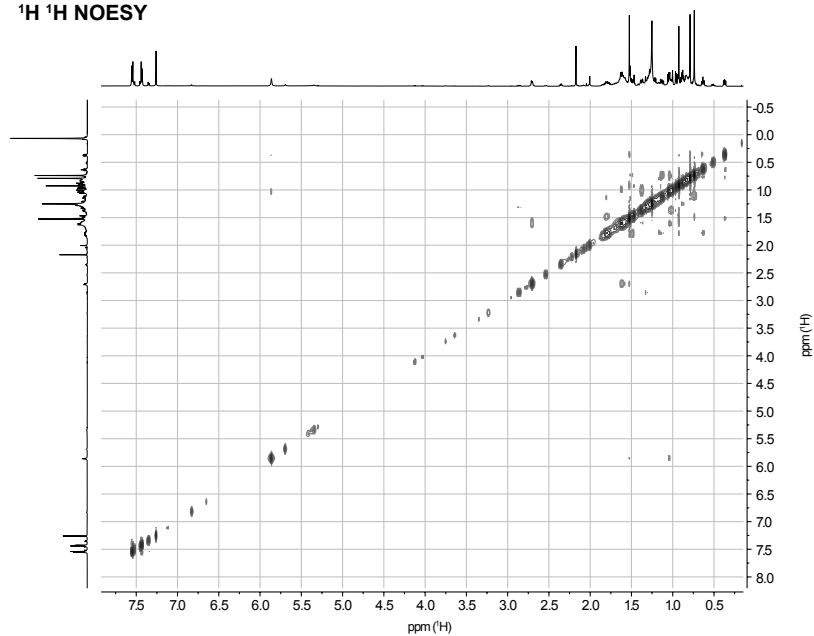
¹H NMR



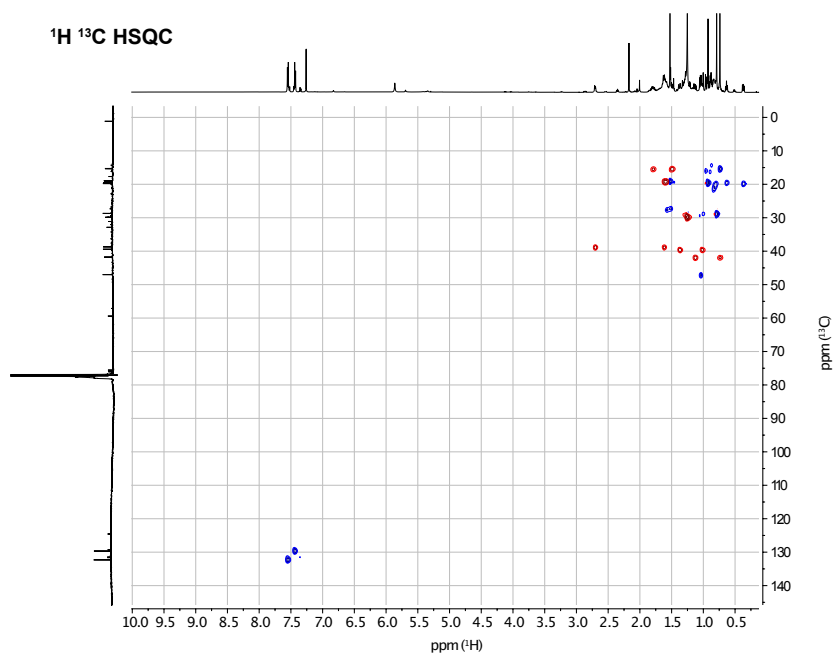
¹H ¹H COSY

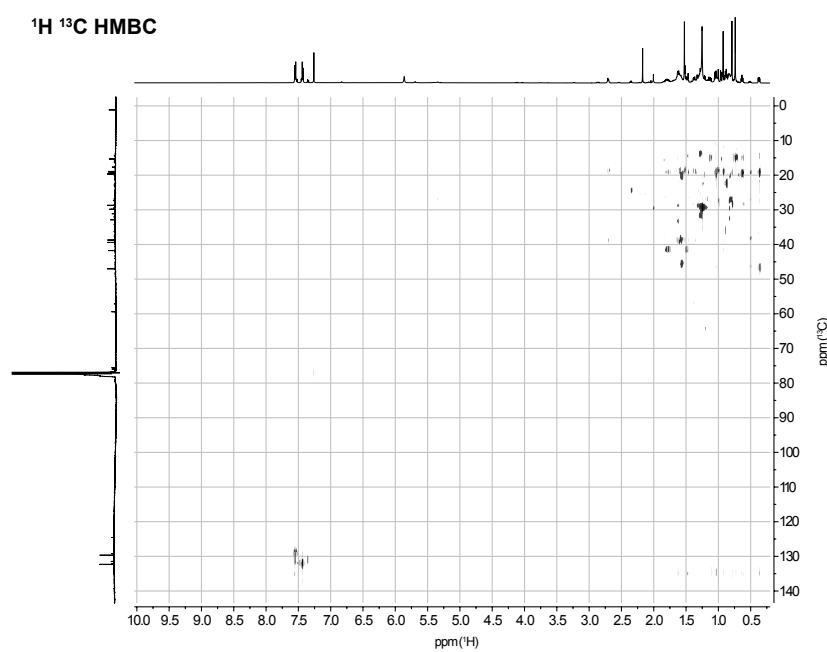
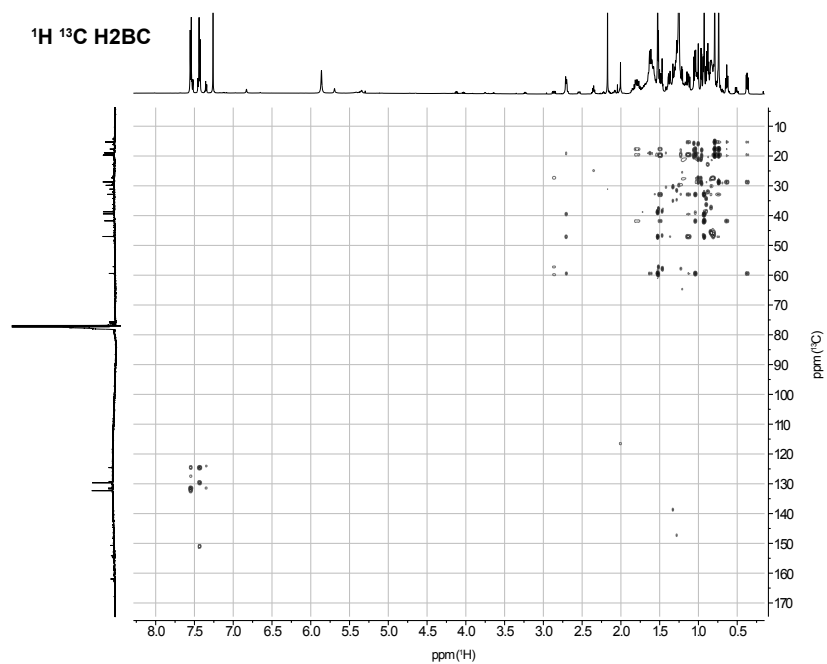


¹H ¹H NOESY

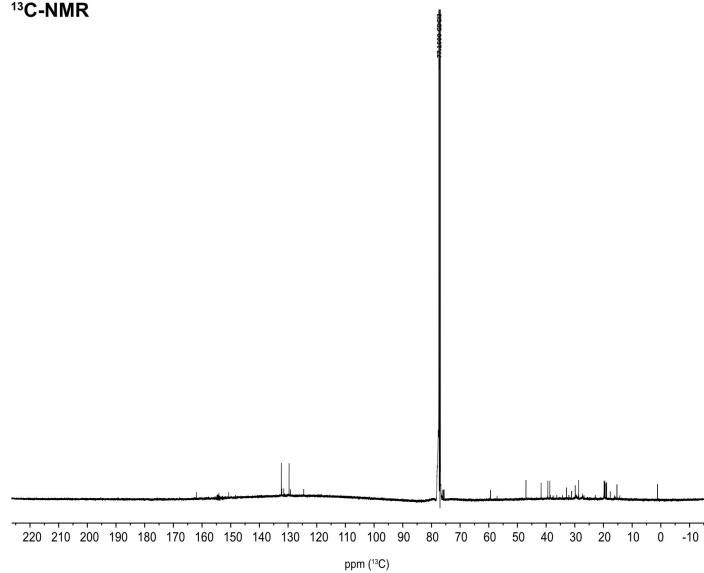


¹H ¹³C HSQC



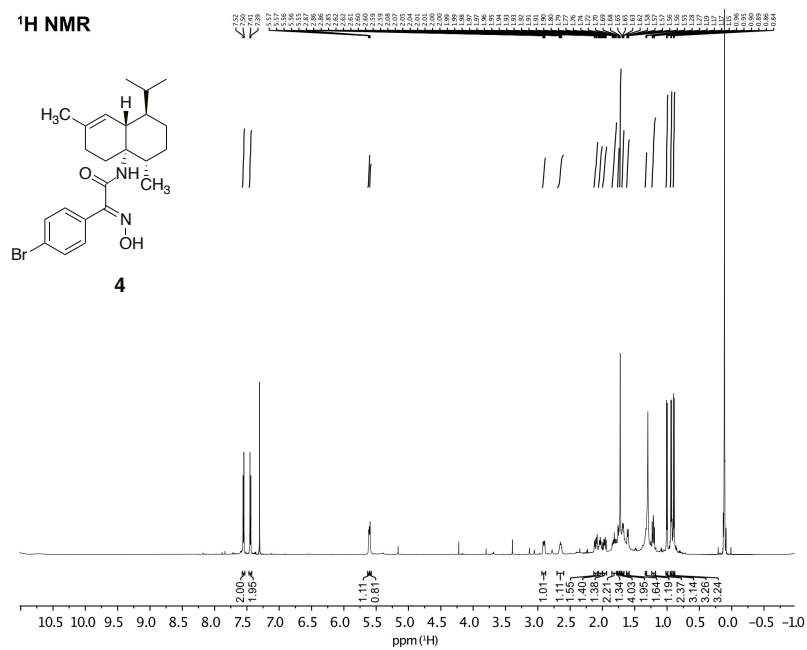


¹³C-NMR

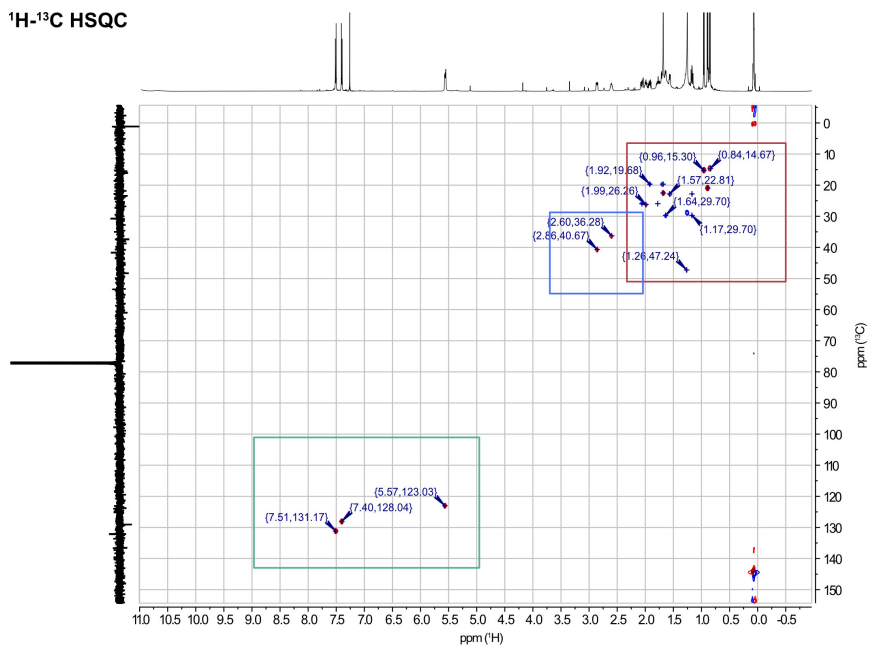


Compound 4

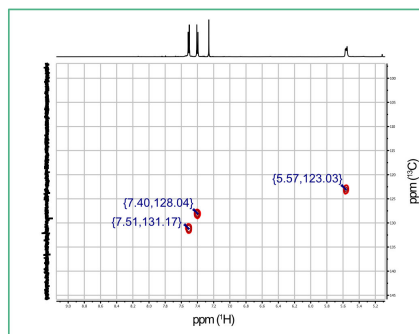
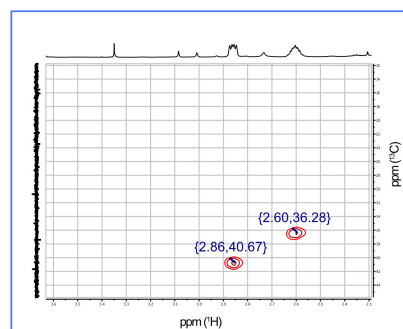
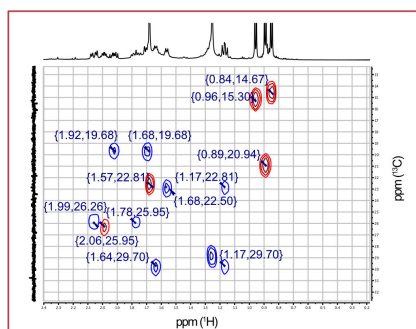
¹H NMR

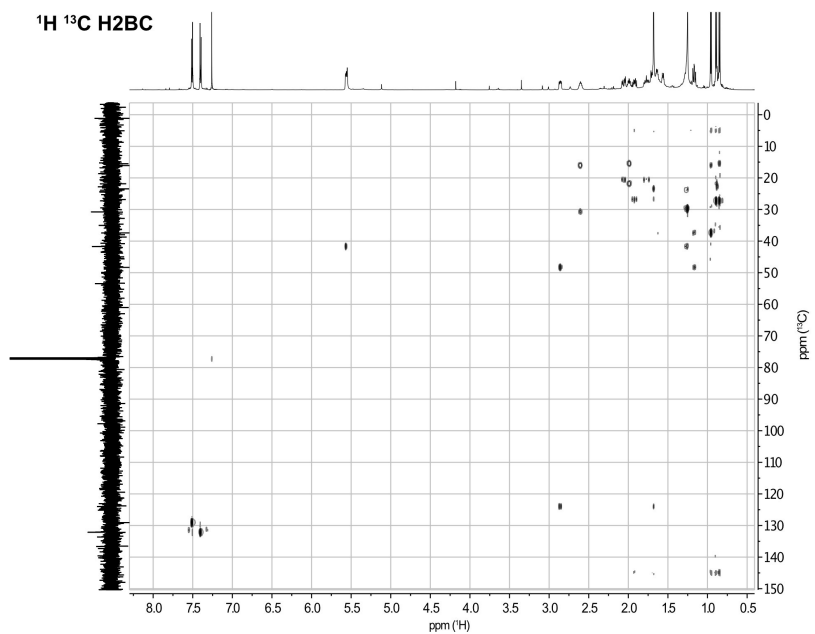
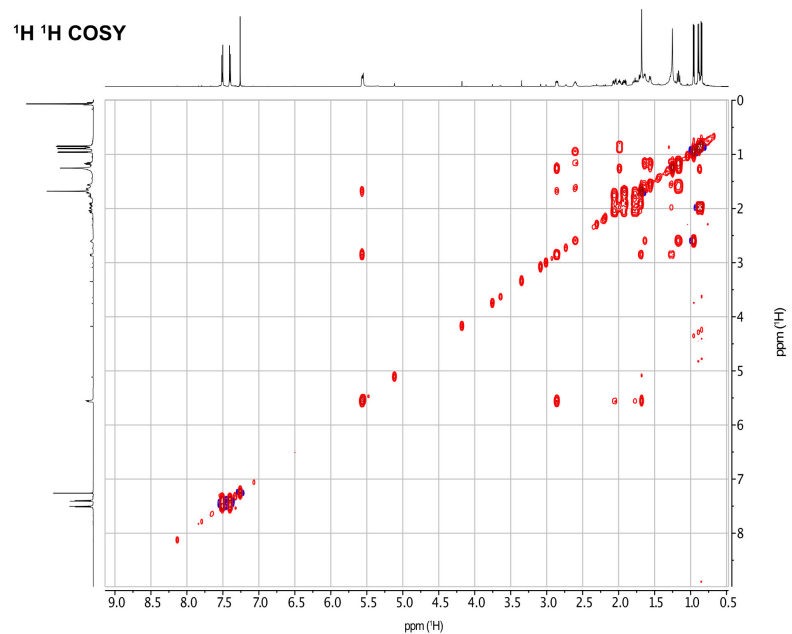


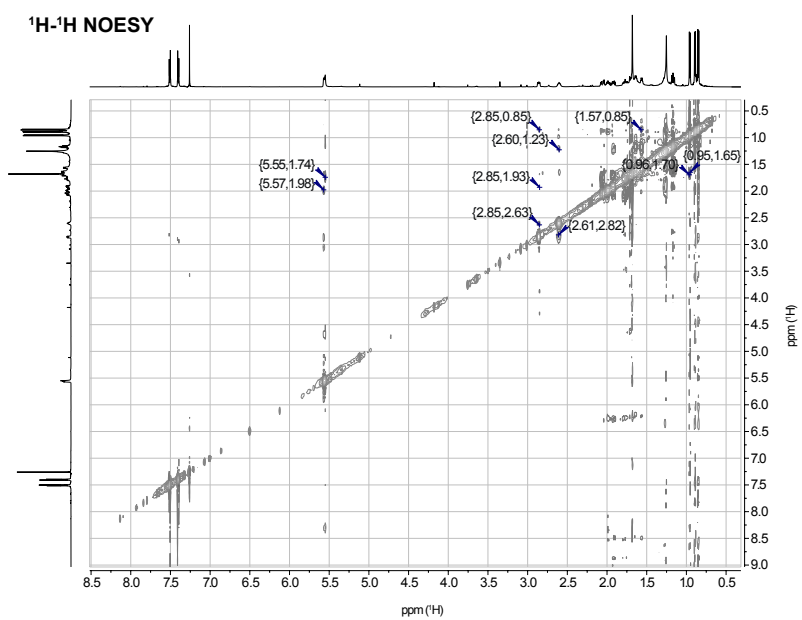
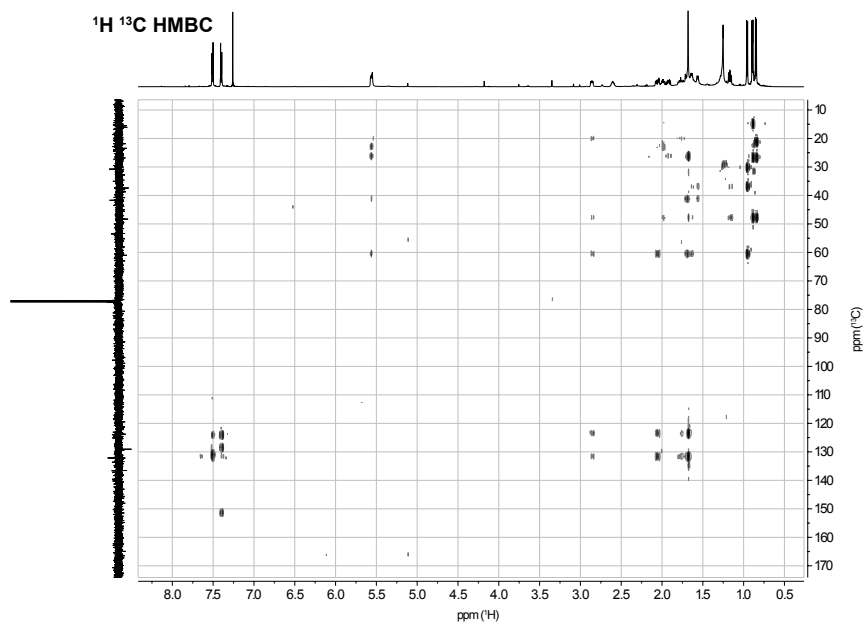
^1H - ^{13}C HSQC



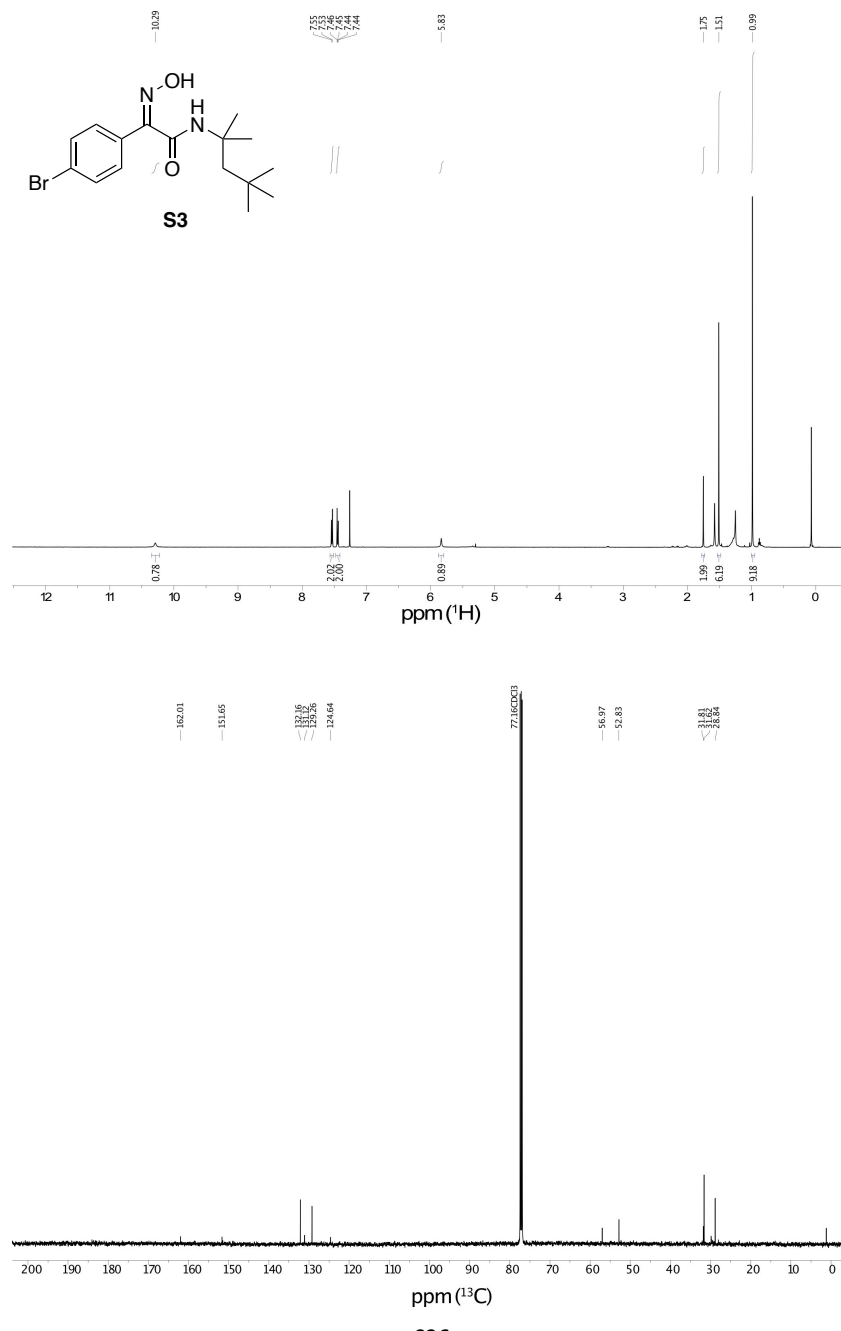
Different zooms of ^1H ^{13}C HSQC







^1H NMR (CDCl_3) and ^{13}C NMR (CDCl_3) spectra of compound **S3**



8. References

- [1] G. Zhang, X. Wen, Y. Wang, W. Mo, C. Ding, *J. Org. Chem.* **2011**, 76, 4665
- [2] B. C. Sanders, F. Friscourt, P. A. Ledin, N. E. Mbua, S. Arumugam, J. Guo, T. J. Boltje, V. V. Popik, G.-J. Boons, *J. Am. Chem. Soc.* **2011**, 133, 949–957.
- [3] S. Mo, A. Kronic, G. Chlipala, J. Orjala, *J. Nat. Prod.* **2009**, 72, 894–899.
- [4] A. Raveh, S. Carmeli, *J. Nat. Prod.* **2007**, 70, 196–201.
- [5] R. J. B. Schäfer, M. R. Monaco, M. Li, A. Tirla, P. Rivera-Fuentes, H. Wennemers, *J. Am. Chem. Soc.* **2019**, 141, 18644–18648.

Section 2.4. Acknowledgements

Chapter 2, in part, is a reprint of the material as it appears in *Chemistry – A European Journal* Schäfer, R.J.B., Wilson, K., Biederman, M., Moore, B.S., Sieber, S., and Wennemers, H., 2022.

Section 3.1 Enzyme discovery

For thousands of years, humans have used enzymes to carry out useful chemical conversions, such as the conversion of grain to beer, milk to cheese, or animal skin to leather. Enzymes are specialized proteins that function as nature's chemists, acting as catalysts in a plethora of biologically-relevant chemical reactions. The process of using enzymes to carry out anthropologically-useful reactions is known as biocatalysis. Several industries rely heavily on biocatalysis, including pharmaceuticals, textiles, food, biofuel, detergents, and paper (1). Industrial use of enzymes is increasing, and the global enzyme market has grown from \$5.01 billion in 2016 to \$6.4 billion in 2021 and is projected to exceed \$8 billion by 2026. (1) As the need for biocatalysis grows, there is an increasing call for enzymes with improved thermal, pH, and solvent tolerance, better regio- and stereoselectivity, improved turnover, and novel activities. While several scientific research efforts are working to address these needs, including enzyme evolution, synthetic biology, strain engineering, enzyme immobilization, process design, and others, the focus of this section will be enzyme discovery efforts (2).

Before the advent of DNA and RNA sequencing, early enzyme discovery efforts often relied on activity-guided fractionation of crude cell free extracts of a source organism. An example of such early enzyme discovery is the 1950 discovery of catechol 1,2 dioxygenase from cell free extracts of *Pseudomonas* sp. (3). Though employed less frequently in recent years, cell free extract fractionation is still a valuable tool for enzyme discovery, particularly for uncharacterized enzyme classes with low sequence homology to known enzymes. More recent applications include the discovery of the OleTJE class of P450s from *Jeotgalicoccus* spp. (4) and e-Poly-L-lysine synthase from *Streptomyces albulus* (5). Most modern enzyme discovery efforts involve some kind of

sequencing, with of DNA or of RNA. Many enzyme discovery efforts have utilized large libraries of pure microbial cultures. As whole genome sequencing of bacterial cultures has gotten cheaper and faster, it has become easier to generate large databases of microbial genomes which can then be mined for biosynthetic genes. Many genome mining tools have been developed to facilitate microbial genome mining, including ant-SMASH, BiG-SCAPE, and BAGEL, which have aided in the discovery of hundreds of novel enzymes including polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), and others (6).

More recently, a variety of methods and tools have been developed to aid enzyme discovery from complex microbial communities, and/or organisms that are difficult or impossible to culture in lab. Single cell genomics is useful tool for enzyme discovery from such samples. Single cell genomics is particularly useful for environmental samples with non-culturable microbes, such as the obligate symbionts found in animal microbiomes. Using a combination of FACS cell sorting, single cell genomics, and clone library generation, Siegl and Hentschel discovered novel NRPS and PKS enzymes from a marine sponge microbiome (7).

Metagenomics, the study of DNA from complex environmental samples often containing multiple organisms, is a powerful and increasingly utilized tool for enzyme discovery (8). The most widely used method for enzyme discovery from metagenomic or metatranscriptomic sequencing is activity-guided functional screening of clone libraries (9). This method involves sequencing a metagenome, creating a library of bacterial clones containing small pieces of those metagenomic sequences, and screening that library for a specific enzymatic activity. Hundreds of enzymes have been discovered this way, including novel polyesterases from a moss metagenome

(10) and several different kinds of hydrolases from a cow gut metagenome (11). This methodology is particularly useful for discovering enzymes like lipases, esterases, chitinases, and cellulases, which act on common, easily accessible substrates like lipids and sugars. Because this method relies on heterologous expression, and therefore faces issues with protein folding, substrate availability, and transcription/translation differences, there has been much work in this space on heterologous host development (12).

Sequence directed metagenomic enzyme discovery involves using *in silico* methods like protein prediction software, sequence similarity networks, BLAST, and hidden Markov models (HMMs) to identify new targets that can then be biochemically tested through heterologous expression. After metagenomic sequencing, assembly, and gene prediction, programs like BLAST, HMMER, or DIAMOND can be used to search for gene candidates. This method's dependence on homology-based gene identification can be a major limitation when trying to identify a gene with low sequence similarity to known proteins. This method has still yielded several new enzymes, such as the heat stable and organic solvent tolerant fatty acid esterase EstDZ2, which was discovered from a hot spring metagenome using a NCBI conserved domain BLAST search (13). Similar methodologies can also be employed using metatranscriptomics, which can be particularly useful for eukaryotic enzyme discovery, since eukaryotic gene splicing can make gene annotation challenging. Metatranscriptome sequencing of a sheep rumen yielded two novel glucoside hydrolases with good thermal and pH tolerance (14).

A combination of several enzyme discovery methods is often necessary for particularly challenging enzyme discovery. One such example is the discovery of the hydroxynitrile lyase

DtHNL from the fern *Davallia tyermannii*. Several hydroxynitrile lyases have been discovered from multiple distinct sources, but none share any sequence homology or conserved motifs, which makes the discovery of new hydroxynitrile lyases very difficult. Lanfranchi et al used a combination of activity assays with *D. tyermannii* tissue, transcriptomics, proteomics, and enzymatic screening to identify the novel hydroxynitrile lyase DtHNL(15).

Section 3.2 Biocatalysis

A number of steps are required to take a wild-type enzyme from discovery to industrial use as a biocatalyst. The first step in the process is cascade design, or identifying retrobiosynthetic steps that can be used to help identify a biocatalytic route to a target compound. RetroBioCat (16) is a program that helps chemists design biocatalytic cascades by entering information about desired starting materials, products, reaction conditions, host organisms, and more. Many databases have also been developed to catalog enzymes for potential use in biocatalysis. Family-specific databases like PrenDB for prenyltransferases (17) or RedoxiBase for oxidoreductases (18) help researchers find enzymes to perform specific biocatalytic reactions. BioCatNet is a larger database specifically designed for biocatalysis and features sequence, structure, substrate, kinetics, and other data for multiple enzyme families including transaminases, CYP450s, short chain dehydrogenases, hydratases, and more (19).

Once a potential biocatalyst has been identified, there is often a need for further optimization to increase parameters like product yield, enantio-, regio- and/or stereo- specificity, enzyme thermostability, and pH tolerance. Directed evolution is a powerful and increasingly popular tool for biocatalyst optimization. (Fig. 3.2.1) Directed evolution begins with a DNA

sequence that encodes for your enzyme of interest, which is then mutated to generate a library of sequence variants. That library is transformed into a desired host organism to produce the enzyme of interest, which is then screened for a parameter of your choosing, such as increased activity, better stereoselectivity, or thermotolerance. The best performing variants are then selected, and their DNA is used as the template for another round of DNA library generation, and the process is repeated. (Fig. 3.2.1).

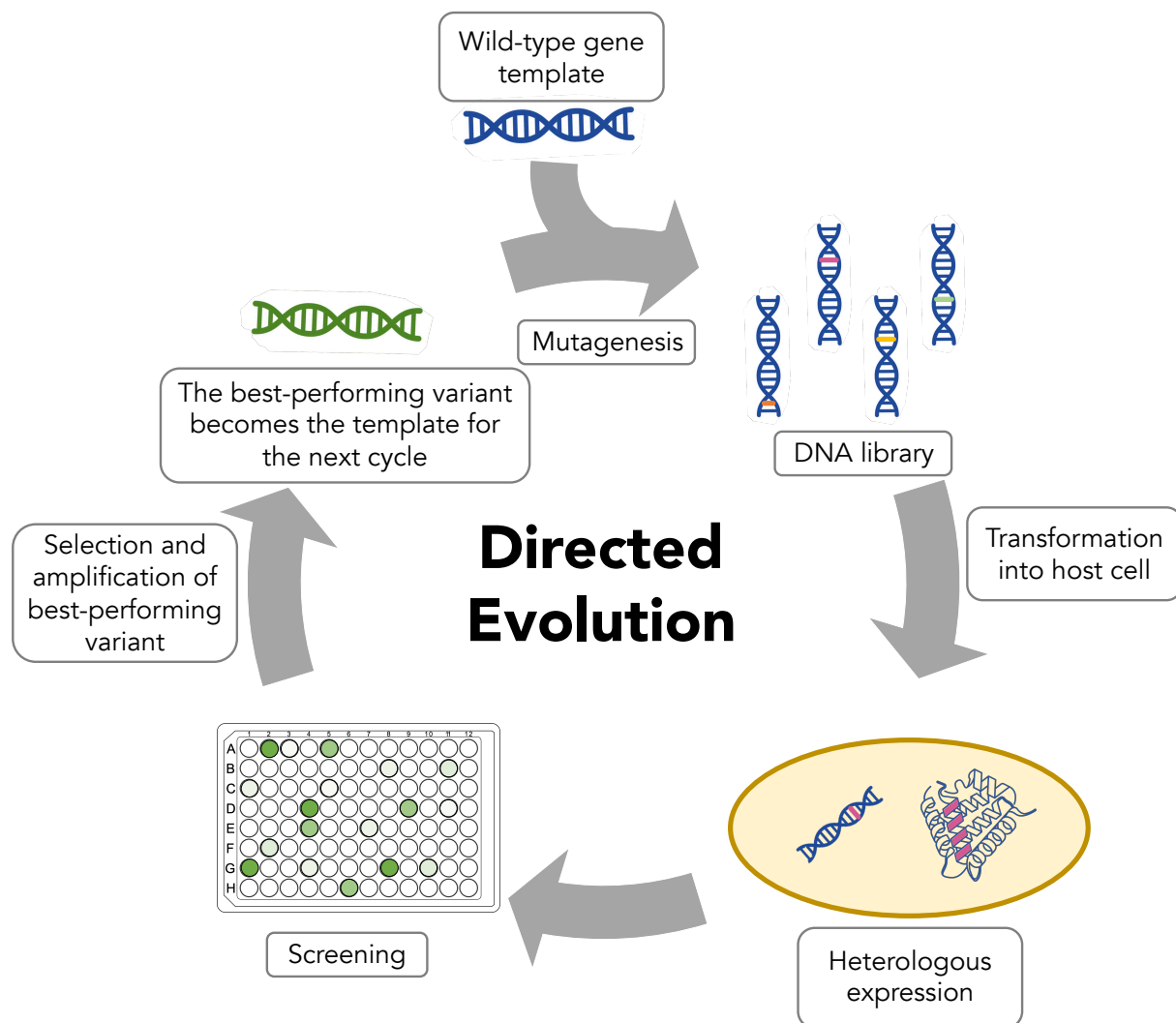


Figure 3.2.1. An overview of the directed evolution workflow.

There are many ways to generate DNA libraries including site saturated mutagenesis, where one residue site is selected as a good target for mutagenesis and then replaced with codons encoding all 20 natural amino acids (20), or combinatorial codon mutagenesis, where multiple degenerate primers are used to introduce mutations at multiple sites within the target gene (21). Both of these methodologies are considered targeted approaches and its helpful to have structural data either from a crystal structure or a structure model to help guide which sites to target for mutagenesis. Other library generation strategies include error prone PCR, an untargeted approach which uses PCR to introduce mutations at random points in the gene (22), or DNA shuffling, in which your target gene is randomly fragmented and assembled using primerless PCR (23). These two approaches are less targeted and don't require structure data, but often generate large libraries that are time consuming to screen. It is also common to combine multiple mutagenesis approaches during a single directed evolution project, depending on the goals of the project and regions of the enzyme being targeted.

Host selection is an important aspect of directed evolution. Frequently issues are encountered when expressing non-native enzymes in a host cell due to differences in codon usage, signal peptides, and posttranslational modifications. While enzymes often express better in heterologous hosts that are relatively closely related to the native host, this is not always the case and sometimes many different heterologous hosts must be screened. While *Escherichia coli* and *Saccharomyces cerevisiae* are by far the most widely used expression hosts for biocatalysis, there are still many different strains to pick from with different transformation efficiency, plasmid maintenance abilities, and growth rates. Downstream screening processes must also be considered when selecting a host. Some hosts allow for easier enzyme purification, while hosts are more

amenable to cell lysate or whole cell culture assays (24). Many host systems allow for cell surface display, which allows your target protein to be anchored to the surface of the host cell, which can eliminate the need for protein purification and facilitate downstream screening assays.

Various screening methods are available for directed evolution efforts, depending on the activity of the enzymes being mutagenized, the parameters being optimized, and the lab equipment available for the experiments. Many screening methodologies rely on the use of colorimetric or fluorometric assays, which yield readouts that can be easily detected by monitoring UV-Vis absorbance or fluorescence. Examples of colorimetric assays include the malachite green assay for detecting phosphate, which has been used to screen terpene synthases (25), or the pH indicator phenol red, which has been used to screen transaminases (26) and hydrolases (27). Systems have also been developed to couple GFP fluorescence with enzyme activity or expression, which allows for screening vis FACS or fluorometer-equipped plate readers (28, 29). HPLC-based screening methods can also be used to measure the consumption of substrates or cofactors, or the production of enzyme products, however this can often be a bottleneck in the pipeline for high throughput projects due to the time required for HPLC, though developments in analytical techniques have made high throughput liquid chromatography more feasible (30).

Directed evolution has proven to be a very effective and increasingly popular tool for biocatalyst optimization. Several pharmaceutical companies have recently used biocatalysis for the full or partial synthesis of clinically used compounds. GSK recently published a biocatalytic method to perform one step in the synthesis of the lysine-specific demethylase-1 inhibitor GSK2879552 (Fig. 3.2.2). Starting with a wild-type imine reductase, Schober et al achieved a

>38,000 fold increase in enzyme activity over 3 rounds of directed evolution (31). The biocatalytic reaction with the engineered imine reductase also had a 84.4% isolated yield with >99.9% purity and >99.7 e.e (31). Merck also recently announced the biocatalytic production of two antivirals, islatravir (32) and molnupiravir (33)(Fig. 3.2.2). The islatravir biocatalytic cascade consists of 5 main enzymatic reactions, and each enzyme in the cascade underwent multiple rounds of directed evolution with impressive increases in activity and enantioselectivity. Directed evolution of a 5-*S*-methylthioribose kinase involved in the biocatalytic production of molnupiravir resulted in a 100-fold increase in activity and >99:1 diastereoselectivity, compared to 7:1 diastereoselectivity for the wild type (33, 34).

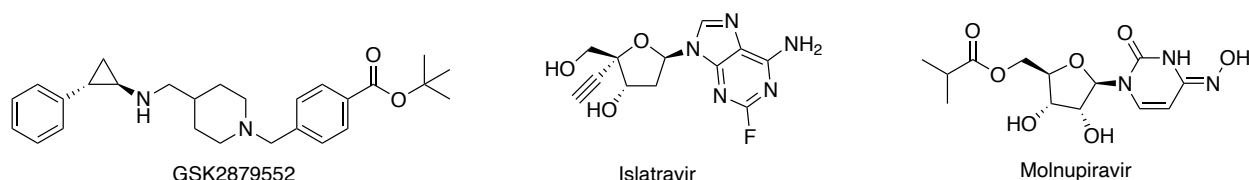


Figure 3.2.2. Pharmaceutical molecules produced using biocatalysis.

Section 3.3 Terpene synthases in biocatalysis

Terpenoids and isoprenoids are widely used across many industries and there are large commercial incentives for industrial production of isoprenoids (35). Small isoprenoids like isoprene have industrial uses in the production of rubber, polymers, adhesives, fuels, and other materials (36). Specialized terpenoids have wide use as flavorings, fragrances, insecticides, and pharmaceuticals. Terpenoids like menthol, limonene, and nootkatone are used as aroma and flavor compounds in food, beverages, cosmetics, personal care products, and medications (Fig. 3.3.1). Nootkatone and menthol are both natural insecticides and citrus oil is used as a solvent in household cleaners and adhesives. Terpenoids are also important precursors in the production of pharmaceutical drugs like artemisinin and taxol (37–39).

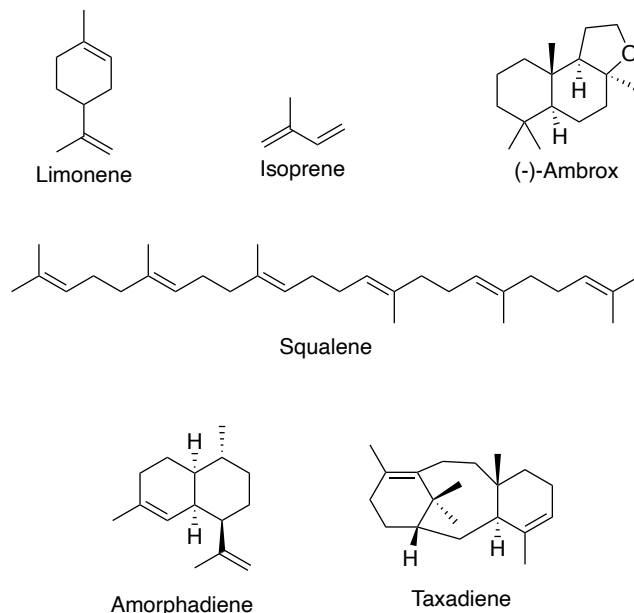


Figure 3.3.1. Recent terpenoids targeted for industrial biocatalytic production.

Because of the widespread commercial appeal of these molecules, much work has been focused on the biocatalytic production of terpenoids and isoprenoids. Bio-based production of isoprenoids has been covered extensively elsewhere (35), but here I will describe a few key examples of successful terpenoid biocatalysis. Limonene is widely used flavor and fragrance compound, typically harvested as a byproduct of the orange juice industry. The production of limonene in *E. coli* was first reported in 2003, where the introduction of a GPP synthase and limonene synthase resulted in a 5mg/L yield of limonene (40). This system relied on the tightly regulated endogenous *E. coli* MEP pathway, but introduction of an exogenous MEV pathway in 2013 increased yields to 400mg/L (41). In 2017, Korman et al. reported a novel cell free biocatalytic cascade which drastically increased limonene yield to >12g/L (42). Using an entirely cell-free system bypassed many of the cell toxicity issues experienced by previous cell-based attempts and allowed for commercial production of limonene by the biotech company Invizyne, which was founded by Korman and Opgenorth. (-)-ambrox is a major fragrance component of

amberggris and is a valuable fragrance compound for perfumes and cosmetics. BASF patented the biocatalytic production of (-)-ambrox in 2021, in which homofarnesylic acid is cyclized to (-)-ambrox by a squalene hopene cyclase from *Zymomonas mobilis* (43). The Swiss chemical company Givaudan also patented the biocatalytic conversion of homofarnesol to crystallized (-)-ambrox (44).

Biocatalysis of isoprenoids has also been utilized by the pharmaceutical industry. The development of a semibiocatalytic production of the life-saving antimalarial drug artemisinin was funded by the Bill and Melinda Gates Foundation in the early 2000s (45). Years of work went into the development of *E. coli*- and *S. cerevisiae*-based strategies for the production of >25g/L of the artemisinin precursors amorphaadiene and artemisinic acid (37, 45). In 2013, the French pharmaceutical company Sanofi announced it would begin producing artemisinin on a multi-tonne scale using a semisynthetic approach in which artemisinic acid would be produced by a fermentative yeast process (46). Squalene is an isoprenoid that is widely used in personal care products and is also used as a vaccine adjuvant in the pharmaceutical sector. In 2011, Novartis patented the bio-based production of squalene via genetically engineered yeast for use in vaccine manufacturing (47). The biotech company Amyris is one of the largest global suppliers for squalene and holds patents for its biobased production of squalene. It uses sugarcane and genetically engineered yeast to produce farnesene, which is then dimerized to squalene using a synthetic catalyst (48, 49).

Section 3.4 Introduction to “Terpene biosynthesis in marine sponge animals”

This chapter encompasses a large part of my PhD research, focusing on the biosynthesis of nitrogenous terpenoids from marine sponges. These nitrogenous terpenes are long-studied members of the marine natural products, and their history is covered extensively in Section 1.3. Though these molecules have been known since 1973, their biosynthetic origins had remained quite mysterious for much of the past 50 years (50). In this chapter, I will describe my discovery of a new class of sponge terpene synthases likely involved in the biosynthesis of marine nitrogenous terpenes, the first terpene synthases ever characterized from the sponge holobiome and the first small molecule biosynthesis genes characterized from the sponge animal host – not a microbial symbiont. This work adds to a growing body of literature on marine animal biosynthesis, which includes pioneering work by Immo Burkhardt (51) and Paul Scesa (52) on coral terpene biosynthesis and the discovery of sea slug polyketide biosynthesis by Torres et al. (53). While marine microbes are well established producers of bioactive and structurally diverse secondary metabolites, marine animals are just beginning to gain recognition for their biosynthetic potential (54), and I’ve enjoyed working at the edge of this developing field. This evidence of animal biosynthesis is particularly significant given that it occurs marine sponges, where decades of precedent had dictated that the sponge microbiome is the most likely source of the molecules described from sponges.

These sponge terpene synthases are highly unusual and form their own distinct clade in the terpene synthase family tree. While their uniqueness made them interesting enzymes to study, it also made them challenging enzymes to discover. Traditional genome mining tools rely heavily on sequence similarity and given that these sponge terpene synthases had low sequence homology

to other known terpene synthases, they were often missed by my early attempts at enzyme discovery. This work highlights the limitations of homology-based genome mining and demonstrates a need for combinatorial genome mining efforts that combine different methodologies and tools to uncover cryptic enzymes. This work required a combination of metagenomic sequencing, transcriptomic sequencing, hidden Markov model (HMM) searches, manual sequence analysis, gene synthesis, heterologous expression, and biochemical characterization to uncover these novel enzymes. This work helps to provide a roadmap for how to approach a difficult enzyme discovery challenge, involving non-model organisms, complex environmental samples, and unique target enzymes.

This work helped me build and strengthen a variety of skills from the fields of chemistry, molecular biology, biochemistry, marine biology, and bioinformatics. A large part of what attracted me to this project was the opportunity it provided to do local, marine field work. As part of this project, I earned multiple SCUBA certificates, including the American Academy of Underwater Sciences scientific diving certification and my National Association of Underwater Instructors rescue diver certification, and I also learned scientific small boat operations. I learned about the local marine ecology of my San Diego dive sites, and I also gained underwater photography and videography skills. In the lab, I gained extensive experience with DNA and RNA extraction and next generation sequencing, both long read and short read sequencing. Especially during the COVID-19 pandemic, I honed my skills in bioinformatics, including metagenomic assembly, genome mining, metagenome annotation, taxonomic analyses, and more. As part of my enzyme discovery efforts, I also learned molecular cloning, enzyme mutagenesis, and heterologous protein expression. I performed many protein purifications and enzyme assays. This project also

gave me the opportunity to do organic synthesis, small molecule purification, and structure elucidation. I also gained valuable experience in science communication, through several talks, poster presentations, discussions, and publications on this project. I also got the opportunity to mentor a high school student for this project, which gave me valuable mentorship experience.

This work is a product of several years of work with a number of talented scientists. Tristan de Rond was an especially valuable influence in the early days of this project, and he provided countless hours of training, mentorship, and discussions on this project. Immo Burkhardt played a crucial role in the discovery of the terpene synthase enzymes and provided extensive help with protocol development. Taylor Steele's expertise in long read sequencing was extremely helpful in establishing the animal origin of these terpene synthases. Rebecca Schäfer's insight was particularly helpful in our efforts to understand the origin of the nitrogenous functional groups and the role of cyanide in the production of nitrogenous terpenes. Sheila Podell and Eric Allen provided valuable sponge samples and sequencing data that helped to establish the occurrence of these terpene synthases across various sponge species. Amelie Hopke also played an important role in the mutagenesis of these terpene synthases, to better understand the role of the active site residues in the product scope of these enzymes. Bradley Moore, my principal investigator, provided incredibly insightful mentorship and direction throughout this work and helped guide me down the winding road of this project. I am incredibly grateful for all of their help and support.

Section 3.5 Reprint of “Terpene biosynthesis in marine sponge animals”



Terpene biosynthesis in marine sponge animals

Kayla Wilson^a, Tristan de Rond^{ab}, Immo Burkhardt^a, Taylor S. Steele^{ac}, Rebecca J. B. Schäfer^a, Sheila Podell^a, Eric E. Allen^a, and Bradley S. Moore^{a,d,1}

Edited by Dean DellaPenna, Michigan State University, East Lansing, MI; received December 11, 2022; accepted January 25, 2023

Sea sponges are the largest marine source of small-molecule natural products described to date. Sponge-derived molecules, such as the chemotherapeutic eribulin, the calcium-channel blocker manoalide, and antimalarial compound kalihinol A, are renowned for their impressive medicinal, chemical, and biological properties. Sponges contain microbiomes that control the production of many natural products isolated from these marine invertebrates. In fact, all genomic studies to date investigating the metabolic origins of sponge-derived small molecules concluded that microbes—not the sponge animal host—are the biosynthetic producers. However, early cell-sorting studies suggested the sponge animal host may play a role particularly in the production of terpenoid molecules. To investigate the genetic underpinnings of sponge terpenoid biosynthesis, we sequenced the metagenome and transcriptome of an isonitrile sesquiterpenoid-containing sponge of the order Bubarida. Using bioinformatic searches and biochemical validation, we identified a group of type I terpene synthases (TSs) from this sponge and multiple other species, the first of this enzyme class characterized from the sponge holobiome. The Bubarida TS-associated contigs consist of intron-containing genes homologous to sponge genes and feature GC percentage and coverage consistent with other eukaryotic sequences. We identified and characterized TS homologs from five different sponge species isolated from geographically distant locations, thereby suggesting a broad distribution amongst sponges. This work sheds light on the role of sponges in secondary metabolite production and speaks to the possibility that other sponge-specific molecules originate from the animal host.

natural products | terpene synthase | marine sponge | biosynthesis

Marine sponges are the richest source of bioactive small molecules from the ocean, accounting for nearly 30% of all known marine natural products (1). Over 9,000 sponge-derived molecules have been characterized to date and represent nearly all biosynthetic molecular classes, such as polyketides, peptides, alkaloids, and terpenoids (1, 2). Not only do sponge-derived compounds serve important ecological roles as feeding deterrents, antioxidants, and antifouling agents, but they also feature impressive medicinal activity, such as the antimalarial activity of kalihinol A, the antitumor activity of avarol, and the anti-inflammatory, antimicrobial, and analgesic activities of manoalide. (Fig. 1A) (3–11). In recent times, the vast majority of investigations into the biosynthetic origins of sponge-derived molecules have focused on the sponge microbiome—the consortium of bacteria, fungi, archaea, and viruses that live in the sponge host—and not on the sponge animal host itself (*SI Appendix, Table S1*) (12–21). To date, all biosynthetic genes responsible for sponge-derived natural product production have been found in microbial symbionts, which has led to a paradigm that the sponge microbiome is the general source of these molecules.

Pregenomic era cell sorting studies determined that certain sponge metabolites are enriched in sponge animal cells, while others are concentrated in microbial symbiont cells. These experiments suggested that both the sponge animal and its microbiome control the production of sponge-derived metabolites (23–26); however, very few recent studies have followed up on the sponge animal's role in specialized metabolite production. Notably, most of the sponge cell-associated metabolites were identified as terpenoids, making this class of sponge-derived molecules an appealing target for the study of sponge animal biosynthesis.

Terpenoids are the largest known class of secondary metabolites (12, 27), and sponges are an especially rich source, with approximately 4,100 of the 9,400 reported sponge natural products belonging to this compound class (2). Biosynthetically, terpenoids are derived from linear oligoprenyl diphosphate precursors, which are assembled from C₅ isoprenoid building blocks by isopentyl diphosphate synthases (IDSs). These linear precursors are cyclized by terpene synthases (TSs) through cationic cascade reactions initiated by either diphosphate elimination (type I) or protonation (type II) (28). Across various lifeforms, type II TSs are generally more common and conserved than type I TSs, since

Significance

Over 9,000 unique molecules have been isolated from marine sponges, including several bioactive molecules such as the calcium-channel blocker manoalide, antimalarial compounds kalihinol A and axisonitrile-3, and anticancer drug eribulin. Until now, all recent work on the origins of bioactive sponge-derived molecules has indicated that they are made by the symbiotic microorganisms that live within the sponge animal host. Herein, we describe type I terpene synthases characterized from the sponge holobiome and report that these enzymes originate from the sponge animal itself and not a microbial symbiont.

Author affiliations: ^aScripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; ^bSchool of Chemical Sciences, University of Auckland, Auckland 1142, New Zealand; ^cDepartment of Chemistry & Biochemistry, University of California, San Diego, La Jolla, CA 92093; and ^dSkaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093

Author contributions: K.W., T.d.R., and B.S.M. designed research; K.W., T.d.R., I.B., T.S.S., and S.P. performed research; K.W., T.d.R., I.B., T.S.S., R.J.B.S., S.P., E.E.A., and B.S.M. analyzed data; and K.W. and B.S.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: bsmoore@ucsd.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2220934120/-DCSupplemental>.

Published February 21, 2023.

type II TSs function in the production of key cellular molecules such as sterol triterpenoids, which are involved in membrane stabilization, hormonal signaling, and other essential processes (29). Accordingly, the majority of (>3,000) sponge terpenoids likely originate from type II terpene synthase chemistry, most of which are steroids, carotenoids, and meroterpenoids. We estimate 6% (~550) of all sponge natural products originate from type I TSs. Most type I TS-derived terpenoids from plants, fungi, and bacteria are usually oxidatively tailored by cytochrome p450s and short-chain dehydrogenases (28, 30, 31). In sponges, notably, roughly 2/3 of probable type I TS-type terpene products belong to the sponge-exclusive class known as nitrogenous terpenes, which feature nitrogen-containing functional groups like isonitriles, isothiocyanates, and formamides that originate from inorganic cyanide (32, 33).

Type I TSs are found in a wide variety of organisms but are scarce in the animal kingdom, where they have, so far, only been described in insects (34–37) and most recently octocorals (38, 39). While all type I TSs feature a conserved IDS-type α -helical fold and are thought to share an ancient evolutionary origin with IDSs (40), insect type I TSs appear to have more recently evolved independently from duplicated IDSs and, therefore, do not bear much sequence similarity to canonical type I TSs (34–37). On the other hand, coral TSs may be the result of ancient horizontal gene transfer from bacteria, given that they share a similar overall structure with key active site residues but still form their own distinct type I TS clade (34). While rare in animals, the presence of type I TSs in insects and corals does suggest that the sponge host animal could potentially house the genes necessary to produce terpenoids. Early radiolabeled feeding work investigated the biosynthesis of sponge nitrogenous terpenoids and, though key information about the origin of the nitrogen-containing functional groups was obtained, the origin of the terpenoid backbone could

not be established (33, 41–48). As far as we know, no studies have successfully characterized a TS from a sponge or a sponge-associated microbe, leaving a gap in our understanding of the origin and enzymology underlying sponge terpene biosynthesis.

To address this fundamental gap, we investigated terpene biosynthesis in an undescribed San Diego sponge in the order Bubarida [formal species description in process (49), previously identified as *Axinella* sp. (50)], which is known to contain isonitrile sesquiterpenoids (Fig. 1 *B* and *C*) (50). By performing metagenomic and transcriptomic sequencing of the sponge and its microbiota, we identified several candidate TS genes resident on sponge genomic contigs. Heterologous expression of these putative TSs and homologs from other *Agelas*, *Stylissa*, and *Phakellia* sponges demonstrated the activity of seven type I sesquiterpene synthases, which are amongst the first biochemically characterized small molecule biosynthetic enzymes of sponge animal origin.

Results

Metagenome Sequencing and Assembly. We first queried publicly available sponge metagenomes for canonical microbial and plant-type terpene synthases. We also queried all four sponge genomes available on National Center for Biotechnology Information (NCBI)—including the *Amphimedon queenslandica* reference genome (51)—as well as genomes from the widespread sponge symbiont genus *Entotheonella* (52). While we readily identified multiple type II TSs, we detected only one putative type I TS gene, from a *Lamellodysidea herbacea* sponge-derived “*Candidatus* Paraprochloron terpios SP5CPC1” Metagenome Associated Genome (accession number GCA_014323965.1) (53, 54). Given the broad distribution of characterized sponge terpenoids, we were surprised we did not observe more candidate type I TSs in sponge

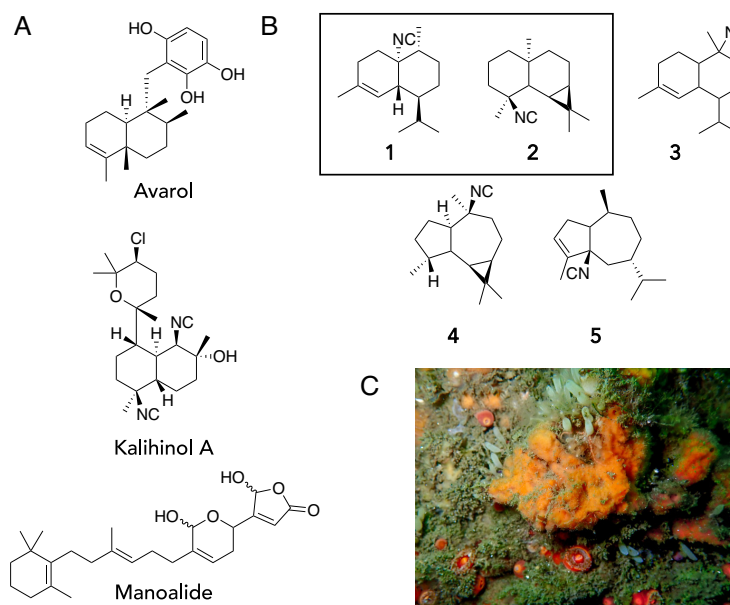


Fig. 1. Terpenoids from marine sponges. (A) Bioactive sponge-derived terpenoids: avarol from *Dysidea avara*, kalihinol A from *Acanthella* sp., and manoalide from *Luffariella variabilis*. (B) Reported isonitrile sesquiterpenes 1 to 5 from Bubarida sponges from San Diego. Boxed compounds are molecules validated from the uncharacterized Bubarida sample that was sequenced in this study and represent relative stereochemistry only (22). (C) In situ uncharacterized Bubarida sponge from San Diego.

metagenomes, which we hypothesized was because of the small number of public sponge sequences and a lack of sequences with paired metabolomics data indicative of terpenoids. To address these challenges, we sequenced the metagenome and metatranscriptome of a sponge known to produce isonitrile sesquiterpenoids **1** to **5** (50), belonging to an undescribed species of the order Bubarida.

We collected the San Diego Bubarida sponge by SCUBA and performed gas chromatography mass spectrometry (GCMS) analysis of crude sponge extract, which indicated the presence of several previously reported isonitrile sesquiterpenoids **1** to **5** (SI Appendix, Figs. S1–S3) (50). We further independently established the structures of **1** and **2** using an isonitrile chemoselective chlorooxime probe (22). Next, we generated a hybrid metagenomic assembly from a combination of Illumina paired-end (PE250) and Oxford Nanopore reads, as well as a metatranscriptomic assembly from Illumina data (SI Appendix, Methods and Table S2). The hybrid assembly features 5.5k contigs larger than 5k bp, totaling 180M bp. The N50 value was 61 kb, and the maximum contig size was 960 kb (SI Appendix, Table S2 and Fig. S4).

Terpene Synthase Genome Mining and Identification. To guide our bioinformatic analysis, we hypothesized that the TSs responsible for sesquiterpene production belonged to one of three previously described groups: 1. Canonical microbial type I TSs from a microbial symbiont; 2. IDS-like TSs akin to those found in insects; or 3. Coral-like type I TSs in the sponge animal. We first searched for canonical microbial type I TSs in our metagenome using BLAST and HMM searches with published TS HMMs, which were solely built from either plant or bacterial TSs (55, 56); however, these initial searches did not uncover any promising candidate TSs from the sponge microbiome. We next investigated the IDSs within the metagenome to search for evidence of neofunctionalization, like that seen in insect TSs. We identified three intron-containing IDS homologs in our metagenomic assembly that we heterologously expressed in *Escherichia coli* and found to perform typical IDS biochemistry. Two functioned as canonical farnesyl diphosphate (FPP) synthases, while the third functioned as a geranylgeranyl diphosphate (GGPP) synthase (SI Appendix, Figs. S5–S11). None had terpene cyclase-like activities.

Next, we queried the sponge metatranscriptome with our custom type I TS HMM (TSHMM1), which we created from a diverse set of characterized fungal, bacterial, plant, and recently discovered coral TSs (SI Appendix, Table S3) (38). We hypothesized this more expansive TS HMM would cast a broader net than established TS HMMs to help uncover unusual TSs that were not revealed in earlier searches. Using TSHMM1, we identified 20 transcripts (representing eight unique ORFs) with relatively weak E-values ($>10E-7$) that contained key terpene synthase amino acid sequence motifs (i.e., DDXXD, NSD/DTE, and RY) and resembled other type I TSs based on homology modeling (SI Appendix, Tables S3 and S4 and Fig. S12). Notably, we detected transcripts derived from five unique putative type I TS genes in the polyadenylation-enriched transcriptome, one of which had a poly-A tail (uBuTS-10), suggestive of eukaryotic origin (SI Appendix, Fig. S13). We mapped four of the putative TSs from the transcriptome to our assembled metagenome, and we additionally identified six unique putative type I TSs only found in the metagenome (SI Appendix, Table S4). After identifying several single-exon candidate gene sequences in the metagenomic dataset, we examined the gene neighborhoods of these putative TSs on assembled metagenomic contigs. We determined that several of the surrounding genes contained introns and had homology to genes from *A. queenslandica* and other publicly

available sponge animal genomes (Fig. 2A and SI Appendix, Table S5). The GC percentages and coverages of these TS-containing contigs were consistent with contigs from our assemblies predicted to belong to the sponge animal host (Fig. 2B). Taking all these features into consideration, we conclude that these genes are part of the sponge animal host genome.

We repeated our TSHMM1 search on several other public and in-house sponge metagenomic datasets. We identified two additional putative TSs from two in-house metagenomes of Caribbean *Agelas clathrodes* and *Agelas tubulata*, sponges which are not known to produce sesquiterpenoids. By searching all publicly available Sequence Read Archive sponge entries, we also identified a complete putative TS from a Norwegian *Phakellia ventrilarum* and two overlapping partial TS sequences from a Guamanian *Stylissa massa* (SI Appendix, Fig. S14) (57, 58). Neither of these sponges had paired metabolomics data available to verify the presence of terpenoids; however, both *S. massa* and several different sponge species from the *Phakellia* and *Stylissa* genera have previously been reported to contain isonitrile terpenoids (59–61). Interestingly, other Norwegian *P. ventrilarum* specimens from the same area as the sequenced TS-containing *P. ventrilarum* specimen have been shown to contain isonitrile sesquiterpenoids (62). *P. ventrilarum* has also recently been reported to form a subclade with the San Diego uncharacterized Bubarida sponge and *Axinella cannabina*, another nitrogenous sesquiterpene-producing sponge (49).

In Vitro Validation of Sponge Terpene Synthase Function.

To investigate the activity of these candidate sponge TSs, we heterologously expressed 13 codon-optimized TS genes from five sponge species in *E. coli* BL21 (DE3). We purified the recombinant proteins and performed in vitro assays with geranyl diphosphate (GPP), FPP, and GGPP. Seven of the 13 TSs showed sesquiterpene synthase activity with selectivity for FPP, five showed low-level monoterpene synthase activity with GPP, and one showed diterpene synthase activity with GGPP (Fig. 3 and SI Appendix, Table S4 and Figs. S15–S35). Five of the active sesquiterpene synthases, which are named according to the taxonomy of their sponge of origin, (uBuTS-1, uBuTS-2, and uBuTS-3 from the undescribed Bubarida, AgTS-1 from *A. clathrodes*, and PhTS-1 from *P. ventrilarum*) displayed high product specificity, while the other two (uBuTS-4 and uBuTS-5) produced a mixture of sesquiterpene products (SI Appendix, Figs. S15–S21). The products of sponge sesquiterpene synthases uBuTS-1, uBuTS-2, and AgTS-1 were isolated from scaled-up in vitro assays and purified, and their structures were characterized by NMR spectroscopy and polarimetry to be (+)-bicyclogermacrene (6), (–)-germacrene D (8), and (+)-alloaromadendrene (7), respectively (Fig. 3 and SI Appendix, Supplementary Note). The products of uBuTS-3 and PhTS-1 were determined via GCMS using purified **6** and **8** from uBuTS-1 and uBuTS-2, respectively, as standards. uBuTS-4 and uBuTS-5 produced several sesquiterpenes distinct from **6** to **8** which were observed by GCMS but not fully characterized (SI Appendix, Figs. S18 and S19). Four of the TSs with low-level monoterpene synthase activity produced only linear myrcene and ocimene type products (uBuTS-2, uBuTS-3, uBuTS-4, and AgTS-1) and one produced both cyclized and linear monoterpenes (PhTS-1) (SI Appendix, Figs. S24–S26, S32, and S34). SrTS-1 from *Stylissa* produced a mixture of diterpenes when incubated with GGPP (SI Appendix, Fig. S35). The remaining five enzymes were not active with the tested substrates or had low levels of expression (SI Appendix, Figs. S22–S35).

To better understand the sponge TSs in the context of other known type I TSs, we collected sequences of type I TSs with characterized activity from several domains of life, including

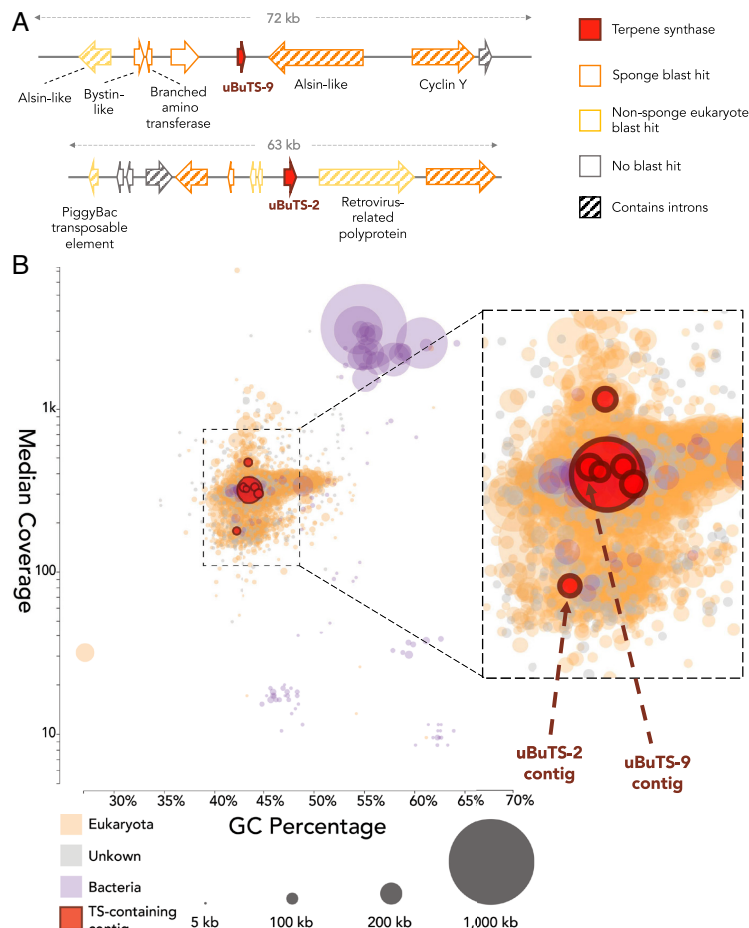


Fig. 2. Eukaryotic features of terpene synthase-containing contigs. (A) TS-containing contigs from the uncharacterized Bubarida sponge. Annotations based on closest blastx hit from the NCBI GenBank nonredundant database (*SI Appendix, Table S5*). (B) GC percentage vs. median coverage plot for all uncharacterized Bubarida metagenomic contigs greater than 5 kb in length. Each circle represents an individual contig. The size of the circle indicates the length of the contig, and for contigs not harboring a TS, the circle's color indicates taxonomic assignment as inferred by homology. Total contig count, sum, contig length, and n50 are as follows: Eukaryota: 2,700, 140 Mb, 74 kb; Unknown: 433, 7.1 Mb, 21 kb; Bacteria: 122, 8.8 Mb, 180 kb. TS-containing contigs: 7, 650 kb, 74 kb. Maximum contig size is 960,860 kb.

sequences of several IDSs as an outgroup, and conducted a phylogenetic analysis. The sponge TSs distinctly clade apart from all known terpene synthases and form a monophyletic group (Fig. 3). Despite clading separately from other TSs, they still share many of the key residues that define canonical TSs (Fig. 4). The sponge TSs have the standard aspartate-rich DDXXD motif, which is characteristic of TSs and IDSs and critical for metal binding. Canonical TSs also contain an additional metal binding region exemplified by (N/D)DXX(S/T)XXX(D/E), commonly known as the NSD/DTE motif, which has been hypothesized to arise from the second aspartate-rich DDXXD motif found in IDSs (40, 63). Twelve of the 13 characterized sponge TSs share an unusual NDXXGXXXD motif which aligns with the traditional NSD/DTE domain. While rare in microbial TSs, this unusual Gly residue in the “middle” Ser/Thr position is seen in plant TSs (64). All of the active sponge TSs we characterized feature an RY dimer near the C terminus that's also present in microbial and octocoral terpene synthases but is not present in known plant terpene synthases (27).

The Role of Sponge TSs in Isonitrile Sesquiterpenoid Biosynthesis.

Given the predominant terpenoid natural products in the Bubarida sponge are isonitrile sesquiterpenoids, including **1** and **2** that resemble uBuTS products **9** and **7**, respectively, we performed exploratory experiments to investigate the role of the newly discovered sponge TSs in the biosynthesis of isonitrile sesquiterpenoids. Garson and colleagues have proposed that inorganic cyanide is the source of the isonitrile and other nitrogenous functional groups unique to sponge terpenoids, by means of radiolabel feeding experiments (33, 41–48, 65). They proposed that a rearranged terpene carbocation intermediate generated by a TS during cyclization might react with the cyanide ion. To test this hypothesis, we performed a series of in vitro TS assays and controls with FPP and potassium cyanide. We tested the 13 sponge TS enzymes, including the five TS enzyme homologs that were not functional in the terpene cyclase assay. In all cases, we did not observe the production of nitrogenous sesquiterpenoids by GCMS in the enzymatic assays or noenzyme controls (*SI Appendix, Figs. S36–S47*). We also combined uBuTS1–9 and ran an in vitro assay with FPP and

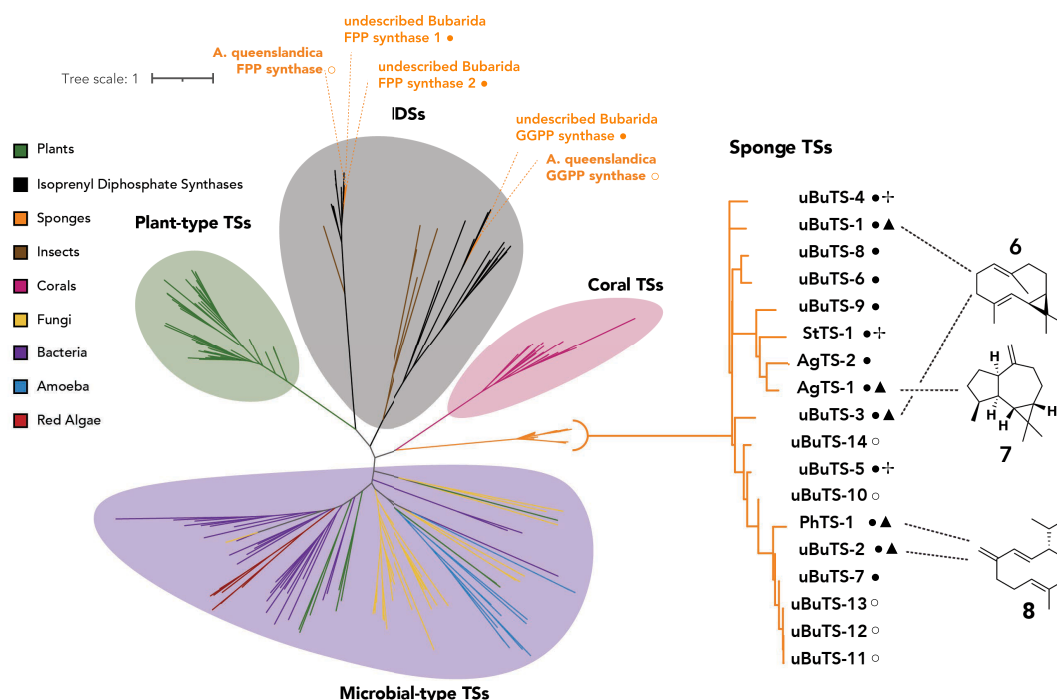


Fig. 3. Phylogenetic tree of selected type I TSs from several domains of life. Isoprenyl diphosphate synthases (IDSs) are included as an outgroup. The first two or three letters of each TS name represent the taxonomy of sponge from which it originated. Enzymes with ● have been biochemically tested, while those with ○ have not been tested. The major sesquiterpene products of enzymes marked with ▲ were fully characterized (*SI Appendix, Supplementary Note*), whereas enzymes with + produce a mixture of terpenes identified only by GC-MS (*SI Appendix, Figs. S18, S19, and S35*). Entries marked with only ● did not produce terpenes under our experimental conditions.

cyanide to test whether there was cross-reactivity between any of the nine uBuTS enzymes, but again we observed only the unmodified sesquiterpenes (*SI Appendix, Fig. S48*). These results suggest that a separate enzyme is required for the regio- and stereoselective installation of the isonitrile. We inspected the gene neighborhoods of the uBuTS genes for clustered genes that may

be involved in cyanide generation or installation; however, no candidates were identified.

To further interrogate the role of cyanide in this system, we tested the sponge tissue for the presence of free cyanide using 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride, a selective reagent that allows for cyanide detection by

uBuTS-1	101	D	D	H	F	D	256	N	D	V	A	G	M	N	R	D	330	R	Y
uBuTS-2	98	D	D	H	F	D	253	N	D	I	L	G	I	P	R	D	325	R	Y
uBuTS-3	103	D	D	H	F	E	260	N	D	I	L	G	K	N	R	D	333	R	Y
uBuTS-4	100	D	D	Y	F	D	255	N	D	I	L	G	M	N	R	D	327	R	Y
uBuTS-5	101	D	D	H	F	D	260	N	D	I	L	G	I	P	R	D	331	R	Y
PhTS-1	98	D	D	H	F	D	253	N	D	V	L	G	I	P	R	D	325	R	Y
AgTS-1	99	D	D	Y	F	D	253	N	D	L	L	G	K	K	K	D	327	R	Y
M. chamomilla germacrene D synthase	86	D	D	T	Y	D	231	D	D	V	V	G	H	K	E	E	-	-	
P. dulcis bicyclogermacrene synthase	95	D	D	T	F	D	240	D	D	I	V	G	H	E	D	E	-	-	
A. terreus aristocholene synthase	89	D	D	L	L	E	219	N	D	I	Y	S	Y	E	K	E	314	R	Y
D. discoideum synthase	81	D	D	F	L	E	231	N	D	C	A	S	Y	D	K	E	326	R	Y
T. musica synthase	94	D	D	V	L	E	259	N	D	L	Y	S	Y	H	R	E	350	R	Y

Fig. 4. Multiple sequence alignment of key active site residues for sponge TSs and selected plant, microbial, and coral TSs.

GCMS (66). Freshly collected Bubarida sponge tissue did not contain a detectable amount of cyanide (*SI Appendix*, Fig. S49). Additional work to understand the biosynthesis of isonitrile sesquiterpenoids is ongoing.

Discussion

We have discovered a new class of type I TSs from marine sponges. The Bubarida TS genes are surrounded by intron-containing genes that closely resemble genes from publicly available sponge genomes and are located on contigs with nucleotide compositions (percent GC) and coverages characteristic of other contigs matching eukaryotic, rather than bacterial sequences. Additionally, the polyadenylation of a TS transcript in our RNA-seq data and the fact we identified type I TSs in multiple sponge species from geographically distant locations strongly support the origin of these TSs sequences in the sponge animal host and not a microbial symbiont. The discovery of sponge animal TSs contradicts the current paradigm that microbes are the only source of natural products isolated from sponges. Besides octocorals and insects, sponges are now the third class of animals recognized to biosynthesize specialized terpenes other than steroids and carotenoids.

Given that marine sponges have been studied as sources of bioactive natural products—including hundreds of terpenoids—for over 70 y, it is notable that TSs were identified in only a few sponges. We believe that this is due to a scarcity of sponge-specific genomic and transcriptomic data, as relatively few species of sponges have any sequencing data associated with them, and even fewer projects focus on the sponge animal as opposed to the associated microbiota. Additionally, nitrogenous terpene production is largely concentrated in a relatively small number of sponges belonging to the order Bubarida (67), for which sequencing data are especially sparse. Furthermore, the order Bubarida is a recently erected and highly debated order that arose because of genetic evidence conflicting with earlier morphology-based classifications of *Axinella*, *Acanthella*, *Phakellia*, *Axinyssa*, and other nitrogenous terpene-producing genera (49, 68). It has even been suggested that nitrogenous terpenes are a chemotaxonomic marker of Bubarida (67), and the monophyletic nature of the newly identified TS sequences points to one common ancestor sequence that was inherited with evolutionary radiation of sponge species. Therefore, our unexpected discovery of functional TSs encoded by sponges outside of Bubarida and not previously reported to produce terpenoids suggests either a need for

further expansion of the Bubarida order or perhaps that this terpene producing capacity may be more widespread than currently appreciated. Unlike coral TSs, which are ubiquitous across all octocorals and are descended from an ancestral TS in the last common octocoral ancestor, sponge TSs are confined to a smaller group of sponge species. Our observation suggests that these TSs may have evolved more recently to serve a specific evolutionary need; however, more research is needed to access the greater distribution and evolutionary history of these genes. As such, further enzyme discovery efforts could greatly benefit from the increased pairing of public sequencing and metabolomics datasets (69).

Many terpenoids found in sponges are uniquely functionalized with isonitrile and related nitrogen-containing functional groups like isothiocyanates, formamides, isocyanates, and amines, suggesting the presence of enzymes that stereoselectively install these functional groups onto terpene hydrocarbons. We hypothesize that the sesquiterpenes produced by the Bubarida TS enzymes are intermediates in the biosynthesis of the isonitrile natural products (22, 50) given their uncanny structural resemblance and the fact *Axinella cannabina*, another nitrogenous terpene-producing sponge closely related to the San Diego Bubarida, is reported to produce bicyclogermacrene **6** (70). The genes and enzymes responsible for these distinctive sponge functionalizations, however, remain unknown, though work from Garson, Scheuer, and others has suggested the origin of these functional groups is inorganic cyanide (44–47). Our observation that the sponge TSs do not install the isonitrile functionality upon the addition of cyanide to TS enzyme assays suggests that there is at least one additional enzyme involved in this biotransformation. We posit that the mechanism to generate **1** from **8** and **2** from **6** proceeds via protonation of an alkene, ring formation, and carbocation quenching by free cyanide to form the isonitrile (Fig. 5), a reaction that may be catalyzed by an enzyme akin to a type II TS, which are known to perform cyclizations on isoprenoid substrates (28). Currently, there is no known precedent for such an “isonitrile synthase” which utilizes cyanide as a substrate; however, the study of enzymes such as FlvF, a repurposed terpene cyclase involved in C–N bond formation, could provide insight into how an isonitrile synthase might evolve from isoprenoid biosynthesis machinery (71). While plant, bacteria, and fungi TS genes often have genes coding for tailoring enzymes like cytochrome P450s clustered near them on the genome (72), the Bubarida sponge TSs do not cluster with obvious candidate tailoring enzyme genes. There is still much to

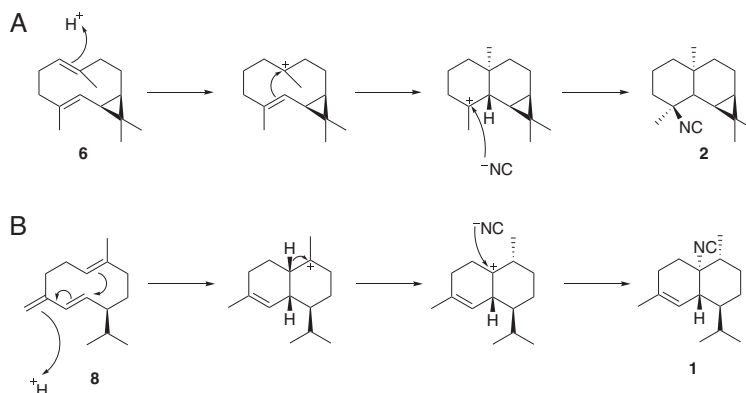


Fig. 5. Two proposed mechanistic routes to nitrogenous sesquiterpenoids from sponge TS products. (A) Proposed mechanism to yield undescribed Bubarida natural product **2** from uBuTS-1 product **6**. (B) Proposed mechanism to yield undescribed Bubarida natural product **1** from uBuTS-2 product **8**.

be learned about this fascinating pathway, of which the late D. John Faulkner once claimed that there is “no more interesting biosynthetic study among the marine natural products” (73).

The discovery of sponge animal biosynthesis genes sheds new light on other classes of sponge natural products that have long been speculated to be sponge animal-derived, such as the pyrrole-imidazole alkaloids, bromotyrosine alkaloids, and other amino acid-derived sponge alkaloids. These sponge-specific alkaloids are also enriched in sponge cell cultures and sponge tissues (74, 75), and recent work has shown that bromotyrosine alkaloids are shared between sponges with vastly different microbiomes (76). Accordingly, the sponge terpene synthase work presented herein not only lends credence to the possibility of sponge animal biosynthesis in other natural product classes but also provides a framework for identifying novel biosynthetic genes from sponge animal hosts.

Methods and Materials

Methods for sponge collection, sequencing, enzyme discovery, and chemical analysis are included in the *SI Appendix, Methods*. The *SI Appendix, Methods* also includes methodologies for

enzyme expression, purification, and activity assays. The structural characterization of enzyme products is detailed in the *SI Appendix, Supplementary Note*. Remaining data and methods are included in the *SI Appendix*. The accession numbers of sequences deposited and used in this study are listed in *SI Appendix, Table S6*.

Data, Materials, and Software Availability. DNA data have been deposited in NCBI Bioproject ([PRJNA907134](#) and [PRJNA824609](#)) (77, 78).

ACKNOWLEDGMENTS. We thank Christian McDonald, Phil Zerofski, and Rich Walsh (Scripps Institution of Oceanography) for their support and assistance in sponge collection, Thomas Turner (UC Santa Barbara) for his assistance with sponge taxonomy, Kristen Jepsen (Institute for Genomic Medicine, UC San Diego) for guidance with sequencing, and Lihini Aluwihare (Scripps Institution of Oceanography) and Brendan Duggan (UC San Diego) for access to GCMS and NMR equipment, respectively. This research was supported by a UCSD Academic Senate Research Grant to B.S.M. and K.W., the NSF Graduate Research Fellowship Program to K.W., NIH grants R01-GM085770 to B.S.M., R01-ES030316 to B.S.M. and E.E.A., F32-GM129960 to T.d., the Marsden Fund, Royal Society Te Apārangi to T.d., a Leopoldina postdoctoral fellowship (LPDS 2019-04) to I.B., and a Swiss NSF postdoctoral fellowship (P2EZP3_195643) to R.J.B.S.

- M. F. Mehlub, J. Lei, C. Franco, W. Zhang, Marine sponge derived natural products between 2001 and 2010: Trends and opportunities for discovery of bioactives. *Mar. Drugs* **12**, 4539–4577 (2014).
- C. Lyu et al., CMNPD: A comprehensive marine natural products database towards facilitating drug discovery from the ocean. *Nucleic Acids Res.* **49**, D509–D515 (2021).
- P. Proksch, Defensive roles for secondary metabolites from marine sponges and sponge-feeding nudibranchs. *Toxicon* **32**, 639–655 (1994).
- S. Takamatsu et al., Marine natural products as novel antioxidant prototypes. *J. Nat. Prod.* **66**, 605–608 (2003).
- N. Fusetani, Antifouling marine natural products. *Nat. Prod. Rep.* **28**, 400–410 (2011).
- E. D. de Silva, P. J. Scheuer, Monoalide, an antibiotic sesterterpenoid from the marine sponge *Luffariella variabilis* (polejaeff). *Tetrahedron Lett.* **21**, 1611–1614 (1980).
- M. L. Ferrández et al., Avarol and avarone, two new anti-inflammatory agents of marine origin. *Eur. J. Pharmacol.* **253**, 75–82 (1994).
- M. Amigó, M. Payá, A. Braza-Bóils, S. De Rosa, M. C. Terencio, Avarol inhibits TNF- α generation and NF- κ B activation in human cells and in animal models. *Life Sci.* **82**, 256–264 (2008).
- J. C. de Freitas, A. B. Blankemeier, R. S. Jacobs, In vitro inactivation of the neurotoxic action of β -bungarotoxin by the marine natural product, monoalide. *Experientia* **40**, 864–865 (1984).
- C. W. J. Chang, A. Patra, D. M. Roli, P. J. Scheuer, Kalihinol-A, a highly functionalized diisocyanate diterpenoid antibiotic from a sponge. *J. Am. Chem. Soc.* **106**, 4644–4646 (1984).
- H. Miyake et al., Antimalarial activity of kalihinol A and new relative diterpenoids from the Okinawan sponge, *Acanthella* sp. *Tetrahedron* **54**, 13467–13474 (1998).
- J. W. Blunt et al., Marine natural products. *Nat. Prod. Rep.* **35**, 8–53 (2018).
- U. Hentschel, J. Piel, S. M. Degan, M. W. Taylor, Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* **10**, 641–654 (2012).
- K. M. Fisch et al., Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* **5**, 494–501 (2009).
- E. W. Schmidt, A. Y. Obratsova, S. K. Davidson, D. J. Faulkner, M. G. Haygood, Identification of the antifungal peptide-containing symbiont of the marine sponge *Theonella swinhoei* as a novel δ -proteobacterium, “*Candidatus Entotheonella palauensis*”. *Mar. Biol.* **136**, 969–977 (2000).
- T. Mori et al., Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1718–1723 (2018).
- V. Agarwal et al., Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat. Chem. Biol.* **13**, 537–543 (2017).
- M. A. Storey et al., Metagenomic exploration of the marine sponge *Mycale hentscheli* uncovers multiple polyketide-producing bacterial symbionts. *MBio* **11**, 1–16 (2020).
- Y. Nakashima, Y. Egami, M. Kimura, T. Wakimoto, I. Abe, Metagenomic analysis of the sponge *Discodermia* reveals the production of the cyanobacterial natural product kasumigamide by “*entothionella*”. *PLoS One* **11**, 1–15 (2016).
- M. F. Freeman, A. L. Vagstad, J. Piel, Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated “*Entotheonella*” bacteria. *Curr. Opin. Chem. Biol.* **31**, 8–14 (2016).
- R. Ueoka et al., Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nat. Chem. Biol.* **11**, 705–712 (2015).
- R. Schäfer et al., Identification of isonitrile-containing natural products in complex biological matrices through ligation with chloroximes. *Chem. Eur. J.* **29**, e202203277 (2022).
- M. D. Unson, D. J. Faulkner, Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia* **49**, 349–353 (1993).
- A. E. Flowers, M. J. Garson, R. I. Webb, E. J. Dumdei, R. D. Charan, Cellular origin of chlorinated diketopiperazines in the dictyoceratid sponge *Dysidea herbacea* (Keller). *Cell Tissue Res.* **292**, 597–607 (1998).
- M. J. Uriz, X. Turon, J. Galera, J. M. Tur, New light on the cell location of avarol within the sponge *Dysidea avara* (Dendroceratida). *Cell Tissue Res.* **285**, 519–527 (1996).
- M. J. Garson et al., Terpenes in sponge cell membranes: Cell separation and membrane fractionation studies with the tropical marine sponge *Amphimedon* sp. *Lipids* **27**, 378–388 (1992).
- J. S. Dickschat, Bacterial terpene cyclases. *Nat. Prod. Rep.* **33**, 87–110 (2016).
- D. W. Christianson, Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* **117**, 11570–11648 (2017).
- J. Schrader, J. Bohlmann, “Biotechnology of Isoprenoids” in *Advances in Biochemical Engineering/Biotechnology*, T. Scheper, Ed. (Springer, 2015), pp. 3–475.
- J. D. Rudolf, T. A. Alsup, B. Xu, Z. Li, Bacterial terpenome. *Nat. Prod. Rep.* **23**, 9–10 (2006).
- M. B. Quin, C. M. Flynn, C. Schmidt-Dannert, Traversing the fungal terpenome. *Nat. Prod. Rep.* **31**, 1449–1473 (2014).
- J. Emsermann, U. Kahl, T. Opatz, Marine isonitriles and their related compounds. *Mar. Drugs* **14**, 16 (2016).
- M. J. Garson, J. S. Simpson, Marine isocyanides and related natural products—Structure, biosynthesis and ecology. *Nat. Prod. Rep.* **21**, 164–179 (2004).
- F. Beran et al., Novel family of terpene synthases evolved from trans-isoprenyl diphosphate synthases in a flea beetle. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2922–2927 (2016).
- J. Lancaster et al., De novo formation of an aggregation pheromone precursor by an isoprenyl diphosphate synthase-related terpene synthase in the harlequin bug. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8634–E8641 (2018).
- A. B. Gilg, J. C. Bearfield, C. Tittiger, W. H. Welch, G. J. Blomquist, Isolation and functional expression of an animal geranyl diphosphate synthase and its role in bark beetle pheromone biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9760–9765 (2005).
- A. B. Gilg, C. Tittiger, G. J. Blomquist, Unique animal prenyltransferase with monoterpene synthase activity. *Naturwissenschaften* **96**, 731–735 (2009).
- I. Burkhardt, T. De Rond, P. Y. Chen, B. S. Moore, Ancient plant-like terpene biosynthesis in corals. *Nat. Chem. Biol.* **18**, 664–669 (2022).
- P. D. Scesa, Z. Lin, E. W. Schmidt, Ancient defensive terpene biosynthetic gene clusters in the soft corals. *Nat. Chem. Biol.* **18**, 659–663 (2022).
- G. Wei et al., Evolution of isoprenyl diphosphate synthase-like terpene synthases in fungi. *Sci. Rep.* **10**, 1–13 (2020).
- E. J. Dumdei, E. Flowers, M. Garson, M. June, The biosynthesis of sesquiterpene isocyanides and isothiocyanates in the marine sponge *Acanthella cavernosa* (Dendy); Evidence for dietary transfer to the dorid nudibranch *Phyllidiella pustulosa*. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **118**, 1385–1392 (1997).
- M. J. Garson, V. Partali, S. Liaaen-Jensen, I. L. Stoilov, Isoprenoid biosynthesis in a marine sponge of the *Amphimedon* genus: Incorporation studies with [14 C]acetate [4- 14 C]cholesterol and [2- 14 C] mevalonate. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **91B**, 293–300 (1988).
- J. S. Simpson, M. J. Garson, Biosynthetic pathways to isocyanides and isothiocyanates; precursor incorporation studies on terpene metabolites in the tropical marine sponges *Amphimedon terpenensis* and *Axinyssa* n.sp. *Org. Biomol. Chem.* **2**, 939–948 (2004).
- M. J. Garson, Biosynthetic studies on marine natural products. *Nat. Prod. Rep.* **6**, 143–170 (1989).
- J. S. Simpson, M. J. Garson, Advanced precursors in marine biosynthetic study: The biosynthesis of diisocyanododecane in *Amphimedon terpenensis*. *Tetrahedron Lett.* **40**, 3909–3912 (1999).
- J. S. Simpson, M. J. Garson, Thiocyanate biosynthesis in the tropical marine sponge *Axinyssa* n.sp. *Tetrahedron Lett.* **39**, 5819–5822 (1998).
- C. W. J. Chang, P. J. Scheuer, Biosynthesis of marine isocyanoterpenoids in sponges. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **97**, 227–233 (1990).
- P. Karuso, P. J. Scheuer, Biosynthesis of isocyanoterpenes in sponges. *J. Org. Chem.* **54**, 2092–2095 (1989).
- T. L. Turner, S. Lonhart, The sponges of the carmel pinnacles marine protected area. *bioRxiv* [Preprint] (2022). <https://doi.org/10.1101/2022.11.02.514922> (Accessed 9 November 2022).
- J. E. Thompson, R. P. Walker, D. J. Faulkner, Screening and bioassays for biologically active substances from forty marine sponge species from San Diego, California, USA. *Mar. Biol.* **21**, 11–21 (1985).
- M. Srivastava et al., The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010).
- M. C. Wilson et al., An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58–62 (2014).

53. S. Podell *et al.*, A genomic view of trophic and metabolic diversity in clade-specific *Lamellodysidea* sponge microbiomes. *Microbiome* **8**, 1–17 (2020).
54. S. A. Agger, F. Lopez-Gallego, T. R. Hoye, C. Schmidt-Dannert, Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120. *J. Bacteriol.* **190**, 6084–6096 (2008).
55. Y. Yamada *et al.*, Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 857–862 (2015).
56. J. Mistry *et al.*, Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
57. T. Ryu *et al.*, Hologenome analysis of two marine sponges with different microbiomes. *BMC Genomics* **17**, 158 (2016).
58. B. Plese *et al.*, Mitochondrial evolution in the Demospongiae (Porifera): Phylogeny, divergence time, and genome biology. *Mol. Phylogenet. Evol.* **155**, 107011 (2021).
59. H. Mitome, N. Shirato, H. Miyaoka, Y. Yamada, R. W. M. Van Soest, Terpene isocyanides, isocyanates, and isothiocyanates from the Okinawan marine sponge *Stylissa* sp. *J. Nat. Prod.* **67**, 833–837 (2004).
60. N. Chanthathamrongsiri, S. Yuenyongsawad, C. Wattanapiromsakul, A. Plubrukarn, Bifunctionalized amphilectane diterpenes from the sponge *Stylissa cf. massa*. *J. Nat. Prod.* **75**, 789–792 (2012).
61. D. Wolf, F. J. Schmitz, New diterpene isonitriles from the sponge *Phakellia pulcherrima*. *J. Nat. Prod.* **61**, 1524–1527 (1998).
62. S. T. Possner, *Isolierung und Strukturaufklärung neuer Naturstoffe aus marinen Invertebraten* (University of Hamburg, Hamburg, Germany, 2005).
63. J. A. Aaron, D. W. Christianson, Trinuclear metal clusters in catalysis by terpene synthases. *Pure Appl. Chem.* **82**, 1585–1597 (2010).
64. K. Zhou, R. J. Peters, Investigating the conservation pattern of a putative second terpene synthase divalent metal binding motif in plants. *Phytochemistry* **70**, 366–369 (2009).
65. C. J. R. Fookes, M. J. Garson, J. K. MacLeod, B. W. Skelton, A. H. White, Biosynthesis of diisocyanodocane, a novel diterpene from the marine sponge. *J. Chem. Soc. Perkin Trans. 1*, 1003–1011 (1988).
66. A. Yamaguchi, H. Miyaguchi, A screening method for cyanide in blood by dimethoxytriazinyl derivatization-GC/MS. *J. Chromatogr. Sci.* **59**, 1–6 (2021).
67. A. Galitz, Y. Nakao, P. J. Schupp, G. Wörheide, A soft spot for chemistry—Current taxonomic and evolutionary implications of sponge secondary metabolite distribution. *Mar. Drugs* **19**, 1–25 (2021).
68. C. Morrow, P. Cárdenas, Proposal for a revised classification of the Demospongiae (Porifera). *Front. Zool.* **12**, 1–27 (2015).
69. M. A. Schorn *et al.*, A community resource for paired genomic and metabolomic data mining. *Nat. Chem. Biol.* **17**, 363–368 (2021).
70. P. Ciminiello, E. Fattorusso, S. Magno, L. Mayol, Sesquiterpenoids based on the epi-maaliene skeleton from the marine sponge *Axinella cannabina*. *J. Nat. Prod.* **48**, 64–68 (1985).
71. D. A. Yee *et al.*, Genome mining of alkaloidal terpenoids from a hybrid terpene and nonribosomal peptide biosynthetic pathway. *J. Am. Chem. Soc.* **142**, 710–714 (2020).
72. M. H. Medema, T. de Rond, B. S. Moore, Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.* **22**, 553–571 (2021).
73. D. J. Faulkner, Interesting aspects of marine natural products chemistry. *Tetrahedron* **33**, 1421–1443 (1977).
74. P. Andrade, R. Willoughby, S. A. Pomponi, R. G. Kerr, Biosynthetic studies of the alkaloid, stevensine, in a cell culture of the marine sponge *Teichaxinella morchella*. *Tetrahedron Lett.* **40**, 4775–4778 (1999).
75. E. Richelle-Maurera *et al.*, Localization and ecological significance of oroidin and sceptrin in the Caribbean sponge *Agelas conifera*. *Biochem. Syst. Ecol.* **31**, 1073–1091 (2003).
76. I. Mohanty *et al.*, Presence of bromotyrosine alkaloids in marine sponges is independent of metabolomic and microbiome architectures. *mSystems* **6**, e01387-20 (2021).
77. Wilson *et al.*, Terpene biosynthesis in marine sponge animals. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/sra/PRJNA907134> Deposited 6 December 2022.
78. S. Podell, Caribbean Agelas sponge metagenomes. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA824609> Deposited 7 April 2022.

Supplementary Information for:

Terpene biosynthesis in marine sponge animals

Kayla Wilson¹, Tristan de Rond^{1,2}, Immo Burkhardt¹, Taylor S. Steele^{1,3}, Rebecca J. B. Schäfer¹, Sheila Podell¹, Eric E. Allen¹, Bradley S. Moore^{1,4,5}

¹Scripps Institution of Oceanography, University of California San Diego; 9500 Gilman Dr., La Jolla, CA 92093, USA.

²School of Chemical Sciences, University of Auckland, Auckland, New Zealand

³Department of Chemistry & Biochemistry, University of California, San Diego, La Jolla, California 92093, United States

⁴Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego; 9500 Gilman Dr., La Jolla, CA 92093, USA.

⁵Corresponding author, email: bsmoore@ucsd.edu

Table of Contents

1. General Methods

2. Supplementary Tables

Table S1 – Natural product biosynthetic gene clusters from the sponge microbiome

Table S2 – Sequencing data summary

Table S3 – TSHMM1 and Fig. 3 TS sequences

Table S4 – Terpene synthase summary

Table S5 – Terpene synthase gene neighborhood annotations

Table S6 – Accession numbers

3. Supplementary Figures

Figure S1-S2 – GCMS traces of sponge crude extract

Figure S3 – Undescribed Bubarida morphology

Figure S4 – BlobTools snail plot

Figures S5-S11 – GCMS traces of IDS assays

Figure S12 – Homology models of uBuTS-1, uBuTS-2, and AgTS-1

Figure S13 – uBuTS-10 Transcript

Figure S14 – StTS-1 Sequence alignment

Figures S15-S21 – GCMS traces of FPP TS assays

Figures S22-S35 – GCMS traces of GPP and GGPP TS assays

Figures S36-S48 – GCMS traces of FPP + KCN assays

Figure S49 – Sponge cyanide detection assays

Figure S50 – LCMS analyses of undescribed Bubarida sponge extract

Figure S51 – SDS-PAGE gel of TSs

4. Supplementary Note

Supplementary note 1: Structure elucidation

5. Supplementary References

Supplementary Methods

Collection of marine sponges

Undescribed Bubarida sponges were collected by SCUBA from the Pacific Ocean near San Diego, California (N32° 51.344', W117° 16.621'), United States of America at approximately 15 meters depth and 15 °C temperature. The undescribed Bubarida sponges were dark orange to brown and lumpy with a shiny appearance at the surface. They tend to be found on rocky substrates, sometimes in kelp forests, often surrounded by other invertebrates and algae (Fig. S3). All specimens were collected under California Fish and Wildlife Scientific Collection Permits SC-13980 and SC-192800001. The freshly collected sponge was cut into approximately 8 cm³ pieces and separated into groups based on sample usage. Sponge tissue for chemical extraction was flash frozen in 50 mL Falcon tubes submerged in liquid nitrogen and stored at -80 °C. Sponge tissue for morphological preservation and spicule analysis was flash frozen with liquid nitrogen in 50 mL Falcon tubes filled with calcium and magnesium free artificial sea water with glycerol (449 mM sodium chloride, 33 mM sodium sulfate, 9 mM potassium chloride, 2.5 mM sodium bicarbonate, 1 mM disodium EDTA, 15% v/v glycerol, 0.2 µM filtered) and stored at -80°C. Sponge tissue for DNA and RNA sequencing was flash frozen in liquid nitrogen in 50mL Falcon tubes filled with RNAlater (25 mM sodium citrate, 10 mM disodium EDTA, 700 g/L ammonium sulfate, pH 5.2) and stored at -80°C.

Agelas sponges were collected by SCUBA divers from Sirenas Marine Discovery (San Diego, CA, USA) off St. Thomas, Virgin Islands (N18°18.571', W64°57.681') at a depth of approximately 12 meters, and processed for metagenomic DNA extraction, Illumina paired-end sequencing, and assembly as previously described (1).

DNA extraction

For Illumina sequencing: Sponge tissue in RNAlater was thawed at room temperature and rinsed with 10 mL sterile calcium- and magnesium-free artificial sea water with glycerol (449 mM sodium chloride, 33 mM sodium sulfate, 9 mM potassium chloride, 2.5 mM sodium bicarbonate, 1 mM disodium EDTA, 15% v/v glycerol, 0.2 µM filtered). The rinsed sponge tissue was thoroughly homogenized with liquid nitrogen and a mortar and pestle. The homogenized sponge powder was transferred to a 2mL microcentrifuge tube and resuspended in 200µL chaos buffer (5M guanidine thiocyanate, 2% Sarkosyl, 50mM EDTA, 0.2µM filtered). The resuspended sponge powder was transferred to a Qiagen 0.1 mm glass bead-filled screw top plastic tube. 5µL of Proteinase K (20mg/mL, Zymo Research cat. No. D3001-2-20) and 50µL 2-mercaptoethanol was added to the tube and the tube was incubated at 55°C for 1 hour, with inversion by hand every 15 minutes. The tube was centrifuged at maximum speed for 3 minutes and the supernatant was pipetted off and saved in a separate microcentrifuge tube. 600µL of solid tissue buffer blue (Zymo Research cat. no. D4068-2-10) was added to the glass bead-filled tube containing the tissue pellet. The glass bead filled tube was homogenized on setting 4.0 for 5 seconds. The glass bead-filled tube was centrifuged at maximum speed for 30 seconds and the reserved supernatant was added back into the glass bead-filled tube. 500µL of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the glass bead-filled tube, rotated by hand very gently for 5 minutes, and centrifuged at maximum speed for 5 minutes. The top aqueous layer was removed to a new microcentrifuge tube. A Zymo Research Quick-DNA Miniprep Kit (cat. no. D3024) was used for the following steps. 0.33 volumes of isopropanol was added to the aqueous layer and loaded onto a Zymo-Spin IICR column. The column was centrifuged at 10,000g for 1 minute. The outside of the column was rinsed with 500µL ethanol and the column was transferred to a new microcentrifuge tube. 200µL of Zymo PreWash Buffer was added to the column and the column was centrifuged at 10,000g for 1 minute. 5µL RNase A (10mg/mL, Zymo Research cat. no. E1008) in 95 µL Zymo Wash Buffer was added to the column. The column in the tube was sealed with Parafilm and incubated at 37°C for 45 minutes. The Parafilm was removed, and the column was centrifuged at 10,000 g for 2 minutes. The column was air dried with the lid open for 5 minutes then eluted with 2 x 50µL of 40°C Zymo Elution Buffer. The concentration and purity of the DNA was evaluated via Qubit and Nanodrop. The DNA was stored at -80°C.

For Oxford Nanopore sequencing: A QIAGEN Genomic Tip 500/G Blood & Cell Culture DNA Maxi kit (cat. no. 13362) was used for high molecular weight DNA extraction. 400mg of frozen sponge stored in RNAlater (25mM sodium citrate, 10mM disodium EDTA, 700g/L ammonium sulfate, pH 5.2) was thawed and removed from excess RNAlater. The sponge tissue was chopped up with a sterile razor blade then

thoroughly homogenized with liquid nitrogen and a mortar and pestle. The QIAGEN Protocol: Preparation of Tissue Samples (page 31-36, June 2015 QIAGEN Genomic DNA Handbook) was used to prepare the sponge tissue for Genomic-tip loading. The Protocol: Isolation of Genomic DNA from Blood, Cultured Cells, Tissue, Yeast, of Bacteria using Genomic-tips was used to isolate DNA from the prepared sponge tissue sample (page 49-52, June 2015 QIAGEN Genomic DNA Handbook). The high molecular weight DNA was size selected using a Sage Science BluePippin and High Pass Plus >15kb Gel Cassette (Sage Science part no. BPLUS10). The size selected DNA was cleaned up using AMPure XP beads (Beckman Coulter, Item no. A63880) following the PCR Purification Process Procedure. The concentration and purity of the DNA was evaluated via Qubit and Nanodrop. The high molecular weight DNA was stored in elution buffer (10mM Tris-Cl, pH 8.5) at 4°C.

28S rRNA Barcoding

PCR amplification of 28S rDNA was performed using the 28S-C2-fwd (GAA AAG AAC TTT GRA RAG AGA GT) and 28S-D2-rev (TCC GTG TTT CAA GAC GGG) primers (2). The PCR mix consisted of 10 µL 5PRIME HotMasterMix (QuantaBio, Beverly, MA, USA), 1 µL each primer (10 µM), 1 µL template DNA, 12 µL water. The thermocycler program was as follows: initial denaturation phase of 94 °C for 2 min followed by 30 cycles of 30 s denaturation at 94 °C, 30 s annealing at 48 °C, 1 min elongation at 65 °C, and a final elongation at 65 °C for 5 min. The 28S sequence is identical to that of a Californian sponge belonging to the order Bubarida soon to be formally described by Turner and Lonhart, 2022 (3). The 28S amplicon sequence from the San Diego undescribed Bubarida is as follows:

```
AACGAATGCAGCCCGATGGGCTTCGGTCGGGTTCAGGGGCGGTGGTGCCGTCGGAGCGGGCGGG
CGGACCCGTGAGGGTCGCCTTTCAGTTCTCGAGCGGTCCATCGCCCTGCACTCCCGGTGCATAGCC
GGCCAACGTCGGTCCGGTTCGGCTCACGCGGATGGTCGGGCAGGTGCCTCAATGCGCCGCTTCGG
CGTGCGCAGGGAGAACTTACAGCCGATTGTTCCGGCAGTCGGACGTTGCGGACCGAGGAGATGTGT
GACGCTTCACACCGGCAGTGCAGGCTCGCCGTTCTCGCGGTGGGCGGGTCTCTTCTGATCGTTG
GGTAGGCGCTACTGGGACTGCCCCGTAGTGCCAGTCGACGTCCCGTCCCTCGGTGGGAGGTTG
GTCACGCTTGCGCTGTAGTCGCTGGCTGAAGGAAGGCTGCATCCGACCCGTCTTGAAACACGGA
```

Illumina Sequencing

For metagenome sequencing, 375 ng of DNA from each sample was fragmented by Adaptive Focused Acoustics (E220 Focused Ultrasonicator, Covaris, Woburn, Massachusetts) to produce an average fragment size of 500 basepairs (bp). Sequencing libraries were generated using the KAPA Hyper Prep Kit (KAPA Biosystems, Wilmington, MA, USA) following manufacturer's instructions using 5 cycles of amplification at the IGM Genomics Center, University of California, San Diego, La Jolla, CA. Resulting libraries were multiplexed with another sponge sequencing project and sequenced with 250 basepair (bp) Paired End reads (PE250). 101.3 Gbp of sequence with was generated using an Illumina NovaSeq Sequencing System (Illumina, San Diego, CA, USA) at the Center for Advanced Technology, University of California, San Francisco, San Francisco, CA. NCBI BioProject number: PRJNA907134. NCBI Sequence Read Archive accession: SRR22527615

For RNA sequencing, total RNA was assessed for quality using an Agilent TapeStation 4200, and RNA sequencing libraries were generated using 500 nanograms of RNA and two methods 1) TruSeq Stranded mRNA Prep (Illumina, San Diego, CA) and 2) TruSeq Stranded Total RNA Library Prep Gold supplemented with Illumina ribodepletion probes for Bacteria. Samples were processed following manufacturer's instructions with the shearing time modified to five minutes. Resulting libraries were multiplexed with another sponge sequencing project and sequenced with 100 basepair (bp) Paired End reads (PE100) on the HiSeq 4000 Sequencing System (Illumina, San Diego, CA, USA) at the IGM Genomics Center, University of California, San Diego, La Jolla, CA. Samples were demultiplexed using bcl2fastq Conversion Software (Illumina, San Diego, CA). 51.3 Gbp and 8.1 Gbp of sequence was generated for the stranded mRNA library and ribodepleted libraries respectively. NCBI BioProject number: PRJNA907134. NCBI Sequence Read Archive accession: SRR22527612 (ribodepleted RNA) and SRR22527613 (mRNA)

Oxford Nanopore Sequencing

Size-selected, high-molecular weight (HMW) DNA was sequenced using the Oxford Nanopore MinION platform (Oxford Nanopore Technologies (ONT), Oxford, UK). A one-dimensional (1d) library was prepared with 600 ng HMW DNA using the Ligation Sequencing Kit (SQK-LSK109) (ONT, Oxford, UK).

This library was loaded onto a R9.4 flowcell and run for 48 hr, resulting in 960 Mb of sequence with a read length N50 of 22 kb. The data from this was further processed and used for assembly as described below. NCBI BioProject number: PRJNA907134. NCBI Sequence Read Archive accession: SRR22527614

Metagenome and transcriptome assembly

Raw nanopore data was basecalled with Guppy v. 5.0.11+2b6dbffa5 using parameters --flowcell FLO-MIN111 and --kit SQK-LSK109, and trimmed with Porechop 0.2.4 using default parameters. Raw Illumina metagenome data was trimmed using Trimmomatic (4) 0.39 using parameters ILLUMINACLIP:adapters.fa:2:30:10:3:true HEADCROP:10 SLIDINGWINDOW:8:20 LEADING:10 TRAILING:10 MINLEN:40. A hybrid genome assembly was performed using Masurca v3.4.2 (5) using parameters NANOPORE= nanopore_data.fastq; PE= ns 410 60 novaseq_data_1.fastq novaseq_data_2.fastq; PE= hs 410 60 hiseq_data_1.fastq hiseq_data_2.fastq; FLYE_ASSEMBLY = 0; CA_PARAMETERS = cgwErrorRate=0.15; LIMIT_JUMP_COVERAGE = 300; MEGA_READS_ONE_PASS=0; USE_LINKING_MATES = 1; EXTEND_JUMP_READS=0. The raw Illumina data were then mapped to this assembly using bwa-mem v2.2.1 using default parameters, the mapped data was used to polish the Masurca assembly with Pilon v1.24 (6) using default parameters, and the resulting polished assembly was polished again in the same fashion. The hybrid assembly features 5.5k contigs larger than 5k bp, totaling 180M bp. Eukaryotic genome assembly completeness as estimated by BUSCO v. 5.1.2 (7) using lineage dataset metazoa_odb10 and Augustus as gene predictor was 73.6% (C:73.6%,S:69.2%,D:4.4%,F:5.0%,M:21.4%,n:954). Genome assemblies using Flye, Metaspades and HASLR were also attempted but yielded lower quality assemblies than Masurca. Transcriptome assembly from raw Illumina RNASeq data was performed using Trinity v2.8.4 using default parameters.

Assembly visualization using Blobtools

Blobtools v2.6.1 (8) was used to generate %GC vs coverage and snail plots as follows: Two Diamond databases, one generated from the NCBI nr dataset and one custom-built from two publicly-available *Halichondria panicea* transcriptomes (HBWD000000000 and GIFJ000000000) (9, 10), were queried for the assembly's contigs using Diamond v2.0.13(11) blastx using parameters --evaluate 1e-25 --long-reads, the results concatenated and used to assign taxonomy to assembly contigs using blobtools. Independently, trimmed reads were mapped to the assembly using bwa-mem2, position-wise coverage calculated using bedtools v2.27.1 genomecov, contig-wise coverage modes and medians calculated using awk 'BEGIN{OFS=" ";current=\$1;median_cov=-1};\$1!=current{print current,bestcov,median_cov;current=\$1;bestcovfreq=-1;bestcov=-1;portion_along=0;median_cov=-1};{portion_along+=5;median_cov=-1}&&portion_along>=0.5{median_cov=\$2};\$3>bestcovfreq{bestcovfreq=\$3;bestcov=\$2}' and added to the blobdir using blobtools.

Chemical Extraction

Methylene Chloride Extractions for GCMS: 5 g of frozen undescribed Bubarida sponge was homogenized with liquid nitrogen and mortar and pestle. The sponge powder was extracted with 200 mL methylene chloride. The extract was vacuum filtered and carefully concentrated using a rotary evaporator to 20 mL final volume. Extract was stored at -20 °C.

Methanol Extractions for LCMS: 500 mg of frozen, dry undescribed Bubarida sponge was homogenized with liquid nitrogen and mortar and pestle. The sponge powder was extracted with 4 mL methanol for 1.5 hrs. The extract was filtered, concentrated to dryness under nitrogen gas, and redissolved in 2:1 methylene chloride:methanol. The extract in methylene chloride and methanol was concentrated to dryness under nitrogen gas, filtered, redissolved in methanol, and stored at -20 °C.

Chemical analysis

LCMS

LC-MS analysis (Fig. S50) was performed with an Agilent Technologies 1260 Infinity series HPLC equipped with a degasser, binary pump, autosampler, and diode array detector coupled to an Agilent

Technologies 6530 Accurate-Mass Q-TOF LC/MS. The instrument was calibrated using the Agilent Reference Calibration Mix®. Separations were performed with a Phenomenex Luna 5 µm C18(2) 100 Å LC Column 250 x 4.6 mm. The solvent system contained water + 0.1% formic acid as solvent A, and acetonitrile + 0.1% formic acid as solvent B. Solvents were of mass spectrometry-grade. Data were collected and analyzed using MassHunter Workstation Software version B.05.01. 0.7 mL min⁻¹ flow rate; gradient: 0-5 min 5% B, 5-15 min 5-100% B, 15-20 min 100% B, 20-21 min 100-5% B, 21-26 min 5% B. Dual ESI ion source, positive polarity. T = 300 °C; Acquisition mass range: 100-1700 m/z; Sample was prepared in MeOH; V_{inj} : 10 µL.

GCMS

GCMS analysis was performed on an Agilent 7890A gas chromatograph with Agilent 5975C mass spectrometer, using a HP-5MS 30 m × 0.25 mm, 0.25 µm column. Oven program was as follows: hold 70 °C for 3 min, 10 °C/min to 325 °C, hold at 325 °C for 3 min. Flow rate: 0.44 mL/min, Injection volume: 1 µL, splitless injection, inlet temperature: 200 °C, MS transfer line temperature: 250 °C, MS source temperature 230 °C, MS quad temperature: 150 °C.

Terpene synthase discovery

TSHMM1, a terpene synthase profile HMM created from biochemically characterized terpene synthase (TS) genes (Table S3) using HMMER 3.1b2, was used to query the undescribed Bubarida transcriptome and the *Agelas clathrodes* and *Agelas tubulata* metagenomes (NCBI BioProject PRJNA824609) for potential TSs, as previously described (12, 13).

Searching for homologs of undescribed Bubarida terpene synthases using NCBI Blast in the publicly available sponge nuclear genomes, only gene fragments could be detected in *Stylissa carteri* (<http://sc.reefgenomics.org/>) and *Stylissa massa* holobiont metagenomes (JGI Integrated Microbial Genomes and Metagenomes Database id numbers Ga0072505, Ga0072501, Ga0070431) (14). The synthesized stylissa was built from two overlapping partial TS sequences from Guamanian 2 stylissa massa sequencing experiments (Ga0070431_10856351 (C-term) and Ga0072505_11956581 (N-term) (Fig. S14)(15). No homologs could be detected in *Amphimedon queenslandica*, *Ephydatia muelleri*, *Halichondria panicea*, or *Xestospongia testudinaria*(15–17). Finally, all raw sequencing data, except for amplicon sequencing data, in the NCBI Sequence Read Archives associated with taxons within Porifera (taxid 6040) and “Sponge metagenome” (taxid 1163772) was downloaded, translated *in silico*, and queried for undescribed Bubarida terpene synthase homologs using the custom HMM described above (TSHMM1). 2469 reads from *Phakellia ventilabrum* RNASeq experiment SRX6964817 were detected, prompting us to assemble this raw data using Trinity with default parameters(18). Upon querying the resulting *P. ventilabrum* transcriptome assembly we identified 1 full-length terpene synthase homolog PhTS-1.

Homology modeling

Amino acid sequences of sponge TSs were submitted to Robetta (<https://rosetta.bakerlab.org/>) for structure prediction using *Eleutherobia rubra* cembrane A synthase (PDB entry 7S5L) as a template. Sponge TS homology models were visualized using PyMOL (<https://pymol.org/>).

Phylogenetic analysis

Sequences of type I TSs were obtained from Burkhardt et al(13). All sequence alignments were generated with Kalign v.2.04 The phylogenetic analysis was performed using IQ-TREE multicore version 1.6.12 substitution model LG+R6 (19, 20). Trees were visualized using IToL and Adobe Illustrator. Plant TS sequences were manually truncated to their alpha domains.

Heterologous expression for analytical scale

Genes for heterologous expression were purchased from TWIST Biosciences as codon optimized (Twist algorithm), N-terminal hexahistidine-tagged sequences cloned into pET 28a (+) vectors (cut sites XhoI and NdeI). The plasmids were chemically transformed into *E. coli* BL21 (DE3). 15 mL LB medium (Fisher) with 50 mg/L kanamycin was inoculated with transformed bacteria and grown for 18 h at 37 °C and 180 rpm. 1 mL of these cultures was transferred to 100 mL terrific broth (Fisher, supplemented with 50 mg/L kanamycin) and grown to OD₆₀₀=0.6-0.8 at 37 °C and 180 rpm. Cultures were subsequently cooled to 18

°C, induced with IPTG (400 μ M) and shaken at 18 °C and 180 rpm for 18h. The cells were pelleted by centrifugation (8000 g, 15 min, 4 °C) and resuspended in TS-binding buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 20 mM imidazole, 1 mM MgCl₂, pH 7.4, 5 mL buffer per 100 mL culture). The cells were lysed by sonication (1 sec on, 1 sec off pulses for 1 min, 30 sec off, 5x) and the debris was removed by centrifugation (11000 g, 20 min, 4 °C). The supernatant was filtered through a 0.22 μ m membrane sterile filter (Merck Millipore) and loaded onto a pre equilibrated (TS-binding buffer) gravity flow column containing Chelating Sepharose Fast Flow resin (GE Healthcare, 0.5 mL) loaded with Ni cations, washed with TS binding buffer (2x 2 mL) and eluted with 2.5 mL TS-elution buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 0.5 M imidazole, 1 mM MgCl₂, pH 7.4). The elution fractions were loaded onto PD-10 columns (GE Healthcare) that were pre equilibrated with 25 mL TS reaction buffer (50 mM HEPES, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol, pH 7.8) and eluted with 3.5 mL TS reaction buffer (Fig. S51). The protein was either used in assays immediately or flash frozen in liquid nitrogen and stored at -80 °C until use.

Heterologous expression for preparative scale

3x 1 L TB media were inoculated with 1 vol% of an *E. coli* overnight culture carrying the expression plasmid for the respective TS. Cultures were grown to OD₆₀₀ = 0.6-0.8 at 37 °C and 180 rpm, cooled to 18 °C, induced with IPTG (400 μ M) and shaken at 18 °C and 180 rpm for 18 h. The cells were pelleted by centrifugation (8000 g, 15 min) and resuspended in TS-binding buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 20 mM imidazole, 1 mM MgCl₂, pH 7.4, 20 mL buffer per 1L culture). The cells were lysed by sonication (15 sec on, 45 sec off pulses for a total on-time of 7 min) and the debris was removed by centrifugation (11000 g, 20 min, 4 °C), transferred to a new tube and centrifuged again (11000 g, 20 min, 4 °C) after DNaseI (Roche Diagnostics, 1 mg per 20mL supernatant) was added. The supernatant was filtered through a 0.22 μ m membrane sterile filter (Merck Millipore) and loaded onto a 5-mL HisTrap-FF Ni-column (GE). The column was washed with TS-binding buffer (40 mL), eluted with TS-elution buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 0.5 M imidazole, 1 mM MgCl₂, pH 7.4) and the elution was collected in 2 mL fractions. Enzyme containing fractions were pooled and the buffer was exchanged in 2.5 mL batches using PD-10 columns equilibrated with TS-reaction buffer (50 mM HEPES, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol, pH 7.8) and eluted with 3.5 mL TS-reaction buffer per column used. The protein was either used in assays immediately or flash frozen in liquid nitrogen and stored at -80 °C until use.

Analytical scale enzyme assays

Analytical scale assays were carried out in 1.5 mL TS reaction buffer (50 mM HEPES, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol, pH 7.8) with an enzyme concentration of 3 μ M and 0.25 mg/mL GPP, FPP or GGPP to test mono-, sesqui-, and diterpene synthase activity. The assays were run for 18 h at room temperature and extracted with hexanes (200 μ L). The extracts were analyzed by GCMS.

Preparative scale enzyme assays

For product isolation, FPP (80 mg, for sesquiterpene production) were dissolved in ammonium bicarbonate solution (15 mL, 50 mM) and added dropwise over 5 h to 600 mL TS reaction buffer (50 mM HEPES, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol, pH 7.8) containing 1-6 μ M enzyme at room temperature. The assay was run for 18 h at room temperature and extracted with pentane (3x 200 mL). The organic layer was dried over MgSO₄, filtered, and concentrated in vacuo to 0.5 mL. The crude product was purified by silica gel column chromatography. Fractions were checked by TLC (stained with phosphomolybdic acid stain) and product containing fractions were pooled, concentrated to dryness, and analyzed by NMR spectroscopy and polarimetry.

Analytical scale enzyme assays with potassium cyanide

Assays were carried out in 1.5mL of either pH5 reaction buffer (100 mM acetate, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol, pH 5.0), pH6 reaction buffer (100 mM Na₂HPO₄, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol, pH 6.0), or pH7 reaction buffer (100 mM HEPES, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol, pH 7.0). Enzyme concentration was 3 μ M, FPP concentration was 0.25 mg/mL, and potassium cyanide concentration was 10mM. The assays were run for 18 h at room temperature and extracted with hexanes (200 μ L). The extracts were analyzed by GCMS.

Analytical scale IDS enzyme assays

Assays were carried out in 1.5mL of TS reaction buffer (50 mM HEPES, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol, pH 7.8). IDS and TS enzymes were added to a final concentration of 3 μ M. DMAPP was added to a final concentration of 0.25 mg/mL and IPP was added to a final concentration of 0.5 mg/mL (2 eq.) or 0.75 mg/mL (3 eq.). The assay was run at room temperature for 18 h and extracted with hexanes (200 μ L). The extracts were analyzed by GCMS.

NMR spectroscopy

NMR spectra were recorded on a JEOL ECZ spectrometer (500 MHz) and were referenced against deuterated chloroform (δ = 7.26 ppm) for ¹H-NMR and deuterated chloroform (δ = 77.2 ppm) for ¹³C-NMR.

Optical Rotation

Optical rotation measurements were taken with a Jasco P-2000 polarimeter.

Supplementary Tables

Molecule	Molecule Class	Gene Cluster	Producer Organism	Host Sponge	Citation
Onnamide	PKS	<i>onn</i>	Uncultured bacterial symbiont	<i>Theonella swinhoei</i>	(21)
Psymbirin	PKS	<i>psy</i>	Uncultured symbiont	<i>Psammocinia aff bulbosa</i>	(22)
Theonellamide	PKS-NRPS	<i>tna</i>	<i>Candidatus Entotheonella sarta</i>	<i>Theonella swinhoei</i>	(23)
Polybrominated diphenyl ethers	PDBE	<i>bmp</i>	<i>Hormoscilla spongelliae</i>	Dysideidae sponges	(24)
Kasumigamide	PKS-NRPS	<i>kas</i>	Uncultured 'Entotheonella' bacteria	<i>Theonella swinhoei</i>	(25)
Polytheonamide	RiPP	<i>poy</i>	Uncultured 'Entotheonella' bacteria	<i>Theonella swinhoei</i>	(26)
Misakinolide	PKS	<i>mis</i>	<i>Candidatus Entotheonella sarta</i> TSWA1	<i>Theonella swinhoei</i>	(27)

Table S1. Natural product biosynthetic gene clusters from the sponge microbiome.

Sequencing dataset	Total sequence	Number of reads
Oxford Nanopore Minion metagenome	960 Mbp	76,640
Illumina Novaseq metagenome	101.3 Gbp	202,618,666
Illumina HiSeq4000 stranded mRNA-seq	51.3 Gbp	256,503,457
Illumina HiSeq4000 ribodepleted RNA-seq	8.1 Gbp	40,717,030

Table S2. Sequencing data summary.

E2E2N7	B6SCF5	I6RAQ6	J7LQ09	Q5SBP3
P59287	P0CJ43	O48935	Q4KSH9	Q84UU4
B3TPQ6	Q41594	B5A434	Q9FXY7	H8ZM70
C7E5V7	O24475	Q6PWU2	A4FVP2	C5YHH7
O49853	Q84LF0	J7LH11	B1B1U3	O81191
O64404	G1JUH1	Q6Q3H2	B9S9Z3	Q94G53
Q8GUE4	B9RXW0	I6QPS5	G5CV45	B0FGA9
E2E2P0	O64961	Q29VN2	C7ASI9	B2KSJ5
D9XDR8	B1W019	D2B747	A0A1L9UKS1.1	A0A1B4XBG5.1
Q96WT2.1	Q9UR08	A0A0P0ZEM1.1	A0A0P0ZD79.1	C9K2Q3
A2PZA5	EW16201	A0A2L0VXR0	P0DL13	Q55012
M1VDX3	B2DBF0	P13513	WP_020663213	WP_011030632
XP_041688448.1	XP_002557473.1	XP_001836356.1	WP_012119179.1	WP_010981512.1
XP_031085895.1	XP_014539580.1	WP_014133196.1	WP_003956090.1	WP_011030119.1
XP_023427124.1	XP_044701492.1	WP_015102836.1	WP_003963519.1	WP_012410187.1
XP_001836356.1	XP_001276070.1	WP_003954606.1	WP_003952918.1	WP_143632835.1
XP_007824857.1	XP_024552383.1	WP_012792334.1	WP_011874125.1	WP_010984429.1
XP_001276070.1	XP_001832573.2	WP_010314578.1	WP_053126184.1	WP_005320742.1
XP_001832925.1	XP_001836556.1	WP_014153723.1	WP_011031839.1	WP_005317515.1
XP_038933630.1	XP_001832925.1	WP_012241161.1	WP_003994861.1	WP_030261827.1
WP_006348711.1	XP_645958.1	XP_644874.1	XP_640697.1	XP_642261.1
XP_642260.1	XP_645125.1	XP_642676.1	XP_638489.1	XP_629084.1
EIM83755.1	KAG5727529.1	BAJ27126.1	KID95099.1	KFG77771.1
EIM91001.1	KAH0583476.1	BAU98235.1	BAX76657.1	AB267396.1
EIM88705.1	KAG5741072.1	AQY56778.1	AQY56777.1	ASV63466.1
EIM82223.1	EAU89322	GU123140.1	QGA30878.1	ASV63465.1
EIM91236.1	LC228601.1	BBF45518	BBD74517.1	AB448947.1
LC484924	QJQ03973.1	LC228602	BBD74518.1	ASV63464.1
>Paramuricea_biscaya_TPS-1 MSCSKEIHAPRRWVDRHKQIPVLPQNAVEKLISMNELIELVIECGLCDKTSINKMYDKINTYQFMWCIVD TVPASQYAEIEFKSSLHFLCALFLVDDAVESYSANEMQDLSRSYDTLEKEVCKTFPNFPSINEMKESLM HLRNPFDRSSITFCMQYVNKITAILLEGNTPHHVYNLRRRTSNAISIAFQAVLIKSKCGSIITSHMLW RRVFDGLVILFYQFGELISGATETAQQHIAVVTELRLGCLYICIVINDLYSYQRDKLASSDNMIKTWLL KTVSSLSEATARCSQILDAIMKMYQRVEQCIQSNPGCPQLESLETTIYTTVGWIRSHTTVVPRYSES QLKVALVEVEEKELPKWLAEKDEYGWNVVEKFVETLNDEKHKGLLDALQGIADGRDQLLKTQLDIS*				
>D_gigantea_TPS-1 MWCGKEVRVPRQWAVLEQEVLEKEPDPELVDIDGMIKWITECDIVDENNVRAYVKLVKPPYFIRLIYPIL PKDKLCVEACKIFLHFVISLYCSDDRTELECDLDDIVKMCDAENLKEPVFEMFPKFTIKEMQSSLKFL SDSKLISPVVLCMDFVNKVTSCVIEYGEYSETAVLDFRRRFLNTIEFYLRGLEMEKKMHRKDTENKTLW RRIFDGGPLFFLPYVEISSFSLGKVEGHIPTISEMYIVSAFVCIIVNDLYSYKRETMESINCDSVIKHWLSD EKISSMEEACEKVSRLNATVQYMFQVQRVKLDYPNSPEAHVLYEYIAHVTIGWLYMHGEGNDRYKD SPWRVSLTEVEEDKIQEWLSCKDSYGFEALNQFLASNPKAKEIIDALGGDEIHHGGLINESS				

>D_gigantea_TPS-2 MFSDNVRVPEKWTISHKKMLAEHEEDKELIALDKLLLVVKETGLTTEAASKSVFDNLNAYRYLRCLFPIL PEDQTGKKLFQNLHFLVMGYVIDDTIEKYSEKEMQELIDGYNFLQNQVSKTLPKLPSSISKVWKLGEA RNKFSRSAIVTVVDYVNRRTMILLEGEICEESVLNRYNRLSNAIAVYFNAILSKTKTGCKISENEMLYRRC FDALSAIVYMSTEVFSKTLVRNQMLPTSLEYKFYLLSTLFCVVINDLYSYERDKLDDTDSVIVKVVYAQNE VTDMKAAATKVSKILDRIQQMYLFFVKEGKSCHPELSEWFERIAAMTVGWYIHKYVVFRYVSSPFQIAV VEVKDEMIPDWLTEKDAYGETIVQNFLDGLLEYPHQKDVLDLSLYGCNRL
>E_rubra-TPS-1 MSISQMVFSKELRVPKEWSKYHFDIVKEPIDPELSEDELFDWVEDLGLTDDKSVVTKYAQRTGGYLF RLHMVFFPKNSLFKKLFLWAHLVPSLFVTDDFLEATSEIEMRQVCDGFEFLAGQIRGQLPQFPPTIAEM KQSLLLQKIDEKLIPHITHLMDFANNVAKSIIQHSSIPESVNEYWRRLLVVCSLYYAGVAFEVKHSVRSY SEDVWTGVFTVAYLLWLPCLILSGAVGKTTEHLSLINELSLATFSARVVNDIYSYNRELVLHKKPVSM VARIAESKEVTTADEALIKNVEILGAIVKVMYQRIEKAKQENQTNKELCKWLDNIGVGTIGVHFVHHYIP RYTSAPCRLSFVGVVEDELKEWRKCEEEPPKELFPVLLDHSQIAKRINDAISGVVPTQVKNVLCDT*
>E_rubra-TPS-2_7S5L_A MSGKIVRAPSNWTTPHKKMLKEDEDQELIAFDKLLSWVSETGVGTTEEQAKIVFKKINAYFYLRCLYPVL PRDPKSMKIFQNLHFLILGYIIDDAIEKYENENEMKELIAGYNLLQSQVSETFPNFPSSIFEMKQLLGNMKN DFSKSAITTFDFYVKNKTTLLLEEGEIAEHNVLNRYKRLSNAVAVYFDALLSKTKTGCEISDGAMLWRRRC FDALAIAYVMLTEVFSKTLVKNHTLPVSEFYKFYLLSILFCVVINDLHSYERDKLDDTDSVVKVWFKEGS VANMEAATSKVSKILDALQQMYLFFVEEGKARHPELSEWFERIAAMTVGWYIHKTVVPRYVSSPFQIEV VEIHENMISNWLKKDAYGQRRVQQFLKNLNDPQKKNIMLYDE*
>B.asbestinum_TPS-1 MVFSKPIHAPKKWVEKGKGLLDEEVDEQLEGLQEIIDTFIQCDLCEDETVRKYRGKLRPYNFTSLNPIL PETKLGREVFKAHAHFTMGIFIADDIMEKQPEEEIRKMCTTFAHLDAKSRQSFPSPFAIDDIKEVVDASS VSKNVAGPIVFLEQFLNKQASLLLSHGDFSLSDVKEFRLRGYNMMAIYFLRVLEETEIQKKPGSIEVIWR RAFGGCVFPFYLLTAEIFSGAIGKLEHYTALTELYVLGGIYVATLNDVFSYYREAHTVADNVVKAMVVSQ KEVETLEEAAANVAEFLDSIMQYLYEKVEELQRKDPNNRELHVCLDYVGRMTVGSVYAHLLPRYKDTT LSYYLEEVEDQLSSWLSSKGEYGTMRMLKNILEVIHQRMADGQMEALAGAFPYCGKFLGVVPTDICCT SSIV*
>B.asbestinum_TPS-2 MTKPFVETAPTRWVKFHARVLSQPPPIENFIRMEQLINWGVACGVTNEAGIRKSQKLNAYVYLRCMYP LVADDAWSAEMFRINLHFTTAGYIVDDRIERYTMEEMNELCDGYDMLEKEVSKLFPKCPSIEEMRCRL QHLKNKFSIAAITMVMDFVFNQSAFLFRQGKTSRDRVDNFRKRLSNAVITYYQAVRNKVKGTGCKITEGE MLWRRCFDVLAVPSYLAPESFTHAVEKQEWPLEMLYELYMLGILYSTVINDLYSYHREKLDDCDNVIKV WMQENSVSIEEANEKICMILDAILQIMYEKIEKAKAQYSKPELQRLDYTGIVSAGWIFVHNTAAPRY LESPYQVVLREIEDKDIQGWLENKNEFGWRVVRHFMETMKSEKGKPVMDAMCGFTDARNILI*
>Xenia_TPS-1_trans MSEKNVVRIPMKWGRIEREILTQNTIPELVDNRLISWVKECNLADEALVTKYMNVIRPYHFSRLVPILP DKDVCREAFVAFIEFLIVLYLCDDLEAKCNLNEIVCSAYDFLDEKLKQCFPRIPSVVEELRSFLPHVKK ERLLALVSLIDSVSRVVSPLLKYSVFPQESVDFRFRFSHSLGYNLKGVLHEKKMLGGVSENKLLWR RIFEGAPITFLMYLEISSLSVTGSNKHIPIATEMYILSSSLCCMVTNDVYSYHRECNEGLKVDNIISLWLHNK QIASVAEGVSRISRLNSAVKYMLAKVKSMKSEYPNNFNVTLEFIALSSVGWLYMHDQGVPRYSDS PWRLNLVDIEESDITSWLAEKDPYADGVIDQFKYSNLDKAKKFIDSLCEKTSVSEEQWND
>Heliopora_coerulea_TPS-1 MAFSKPIYAPKKWAEEKAGVTSMNEEKQLVGMDELINWVAECELAGEATVRKYWDTVRPYHFMRCCL YPVFPDNLCKEFFKVLLHASISYIADDQLEKQSREEAKIACAHAHARIDQQSTKRFPNLPTIQEMKQILST FSTPSIVGPTTMMFADFGNRLAKVLLLHGNWNGNVVADFRLRNSNLVSMYFQAIQAEKTPNKKNTTLET LWRRIFGGVAAPYSSLAIEIASGAIENSKQHIAVITDMHLLSGLFSTTINDLYSYFLEKSSICDNVVAALLHE KTAENVSEAVDKVAQILDAILKLMYKAEQIKLQFPDNRELDFDYVGQVTVGWYYLHECALPRYQG SPWRVHLEEIDEDEIPKWLSEKGYGSKVMQELMELVQERMADGKMDAVHGKFPINEKYVAEKEDGL YKII*

>Heliopora_coerulea_TPS-2 MACSKEMYIPKEWTKYHHDIVKEVKRADLEADDEVAKWVVGLGLSDKPATEKYIDSTRPYHFMRCIAV LVPSNPLCRNIFKTWQIYVTGLFVSDDYLEQTSLEAVVQVSDSFDMLNEQISERFPVIPHQAQIMRSLKI NEKLIPHVLVQIDWLNTIANDLLQHGDSEDDVWDFRCRMSAWIRAYFDGLKSQVGPVKVKTDDDFM WLRIDGGVGPFLLINEILVVELGKCKGHVKLITQLYLSSCVCVLINDCYSYHRETVVITYNTIKVILENKE VSEVPDAVSKLLQTINAIKMYQTIKTKHEYDPNMLNKLDAIGGATAGWFFVHDKVLPQYQATSW RFSLVEVEEMELKEWRKSADEKPSDLVQPLLYRFNAKGKEIDALAYVPKRSRRIVPE*
>Heliopora_coerulea_TPS-3 MSCSKEVRAPRKWVQRQREMOKVKVDDNLISMDIELVVECGLCDEASVHKMYERTNTYQFMWTLI PTVPDQWQFEIFKTSLHFLCALFLVDDAVESYSEKQMKEMSDAYDMLEKQVCENFPFCPSINEMKQ SLKHLKNPFDIASITFCMHYVKNIASVLVNEGSTPQNVVYNFRRRTSNAISISFQAVLIKAKLGSNVTGDE MLWRRVYDGLVILFYQFQFGLVGLTKNTEKYMTVITELRLLGCLYCVINDLYSYHRDKLVSSDNIKTWL LGKAVSNLSEASSRCEILDAMVQYMHVERIRQSHPGCPELEALLETTVYTTVGWIFAHTAVVPRY SESPKVLHVEVEESELHTWFCEKNHYGRSVIEKFLKAVNDEKHKGILDALSGFVDGRGQLLKTQL*
>Renilla_reniformis_TPS-1 MLRTEVKIPEVWCIPCDLKCSTPPQLKKDIIDWCVRTGITADRRTAEKALDRLKPYMYMKILFPKLKPDN PLIASLYELNASVIVTGFTMDVLETYTLAIEELDHTFRETQNWLSPLPPKDYDPLDAILDHIPTSTKPPY SRSVVAMFTDYFNRYCAAHTAYKPSETRLGAFRRRLAAAVTAYLEMAKRKRRRDGNLHEEEFLWQR SADNLSFPVLMMLAEVFTGILADADVSTTLHYHFFYSNLFAVLNLDLSYHRDKDNDNSLVKLWLKLG VAKDFDDAATKIIDFLNSIMVRMYVATKELLESNPENEPLRRFVEAVGYTFNGWILVHTTAVERYKLSPF QTVVAPVARGKEEVWLKGETAFGKRCVAMFEELMGERAEEMNKLYGLDS
>Tubipora_musica_TPS-2 MSCSKEIRIPSDWASVQRKSMQEGPDESLLDFFELANWLRECGVTDDQCKVRKYVDATRPVLSIRTIL RVFPNNTLCRMHFKFLTQYITIGIYIDDDVLESHYPLEVLKEICYEYDQLDQKLLGQFPQIPSRKELKNFLA HLKNEKVISTVTFHMDFVNRVTVNILQNGDFSEEAVLEFRRRLSNILLYLQGVQYEKRGAGGKPPSTVV EALWNRVFGGAAVTWRLFGEVPSGTLGKIGEHITLISELYLSALYCVINDLYSYHREIDAVPTAGENFV RVMFNQKEVANLTVAAKVASILNEITKYTYEQVREFKASYPNSELHQLFDDIGCGTVGWYYFHECT NPRYKESNVRISMKDVDKNEIEDWLSTKDSYGWNIVKQFLASYDAKAKRLNDAVACRIPAYAEFT*
>uBuTS-1_undescribed_Bubarida MSLSVFAKIPVPALKLENPLHYSDSAVSLAKVREPLLSGELKSRLHRHALDLKLPKYIENLNVHDFVAK VHDLGVSNNHPTPGQNRQWLFASLVLVFIFFDDHFDTAUVRTPENIARVSKDMRDVLRSLSRHNLGGL QGSLEDWPAKVPCKEAYHWLLREGEDLREGTAQLIHDTFVDYCYGTEGEVIEWQPDMYRGDFTAWN LDRYCEIRKRSIGVLFVGLPLFIYRNWQKEHITTCIDLLHEAAIIVALANDVAGMNRDLEDNEAIDITS KIAATSSEVVQYHNKKVEVLHKKVLEECNTWRFMEQVEASVGLFLWHLNAQRYAVNSEHYTL
>uBuTS-2_undescribed_Bubarida MSLSVFAKIPVPTLELEHPMHYVDSTVSAKVRISVSAGLNNRLYRRVLDLKLNPYINHMDINDFVAK ACNLGLPTPDQERLWIFTSVNALFIFEFDDHFDKPLLIAPEYVARRSKQMRVLRSLSSHKLSSGLQGS EDWPAEVPCKEAYLWLLREAEDLRKGAELLHDVDFDCNGVEEEVIQWQSDAHRGDFVTWNLDTRYK EARKRSVGYIFTVPVPLFIINKWIKDYFTTCIDLLYEAIIIVALANDILGIPRDHGDDGMTVLRVITSQDK VVEVHNEMVEVLHKKVLEECNKRRFMEEMEAAVGLFLWQLDSNRYAGRLPITIY
>uBuTS-3_undescribed_Bubarida MSLSAFTKIPVPALELEYPMHYADSAVSEAKLRETSVSTELKSKLHRHILNLKKAQYAEHLNIYDFIAKV HSIGVSNHPTVNKSEKRQVLFASSLIVLLFQFDDHFEATDPLHLTLESIAKRSKEMRSLSSLSAHNLSG LQGSLEEWPAAPGKEAYLWLLREADDLREGTAELVHDTFVDYCHGVLMEIIEWQSDMYRGDFTAW NLDTRYKEVRKRSSGVIFGTVGILYITQKWIQKEHITTCIDLLYEAIIIVSFGNDILGKNRDSVDSSVIDITS LKIATSSNIVLKHNTMVKALHSKVLLEEGDTRHFMEEMEAASVGLFLWQLYSKRYL
>uBuTS-4_undescribed_Bubarida MSLSVFAKIPVPTLELEQPVHYVDSASVVKFREPLVSGDVKVRKLHRHVLDLKLNPYIQLHGIYDLVAK VYNLGVSNHPLNQYRLWLLVSTLEILIFEFDYFDKSLRTPENIAKLSKEMRGVLRSLSRHNLGGLQGS LEDWPAEVPCKQAYLWHLREAEDLREGTAQLIHDTFVDYCYGVEVEVTEWQSDVNKGDLTAWNLDLDR YNDARKHSVGLAYAVGPLYANKWLQKEHIITCTDLVYDASILTGLNDILGMNRDRGDSGTMTSLKLL ASSSEVVQYHNELVEVLHKKVIELECNTRHYMEEVEAAVVAFFLWQSRARYAV

>uBuTS-5_undescribed_Bubarida MSLSVFTKIPIPAVELEHLMHYADSAVSEAKLRKHSVSEELKSQMHRHLVDLKLNPYPDNLNIYDFVA KLFNLGVSKHPTFDQNRLWLFSSILILFLEFDDHFDHFDNPLHLTTENIARLSKEMRAVLRSLSRHNL SGLQGSLEDWPAEVPCKEAYRWLLREAKDLKEGAAELLHDIFVDYCFGVEEEVIEWKSDMYRGDLTA WSLDTRYKEVRKRAAGVVFVPLFVTRKWIQKEHVNTCTDLLYEAALLVGLSNDILGIPRDLRDSGTM TVLKIASTSEVVQHHNMMVEVLHKKVLALEGNTMHFMDAVETSVAGVFLWQCHSKRYTL
>uBuTS-6_undescribed_Bubarida MSLGVFAKIPVPALELEHPVHYADSAVSAKVRISVSAGLNNRLYRRVLGLKMNPEYIKHMDINDFVA KACNLGLPTPDQDRLWIFASVTGLFIFEFDDHFDKPLLIAPEYVARRSKQMRVLRSLSSHKL SGLQGS LEDWPAEVPCKQAYLWHLREAEDLREGAAELVQDLFVDYCYGTEEEIIEWQPDAYRGDYTAWNLDY NEVRKRSAGLTFVSVGPLFVAHKWIQKKHVVTCTDLLYEATIVIALANDILGVNRDRDDRLCVTSCLKSS CSEVVQYHNELVEVLHKKVLELEGNTRRFMEEEVASVGLAIPVAVPFKTLWCGLG
>uBuTS-7_undescribed_Bubarida MSLGVFAKIPVPALELEHPVHYADSAVSAKVRISVSAGLNNRLYRRVLDLKLNPYINHMDINDFVAK ACNLGLPTPDQERLWIFTSVNALFIFEFDDHFDKPLLIAPEYVARRSKQMRVLRSLSSHKL SGLQGS EDWPAEVPCKEAYLWLLREAEDLRKGAELLHDVDFIDFCNGVEEEVIQWQSDAHRGDFGTWNLDYK EARKRSVGYIFTVPVPLFIINKWIQDYFTTCIDLLYEAIIIVGLGNDILGMLRDSAEDGSMTVLKIITSHDK IVEVHNKMVEVLHKKVLELECNKRRFMEEMEAAVGLFLWQLDSNRYAGRLPITY
>uBuTS-8_undescribed_Bubarida MSLSTFAKIPVPELELEHQMHYADSIVSVTKLREPSVSKELSRHLRRVLELRNPQNIELNSYDFVAK VCTFGVSNHSTTCSHQNRWWIFASISVIFLFEFDDQFDKPLLLSPENIASLSKKMRGVLRSLSGHKLSG LQGNLEDWPAEVLGKEAYLWLLREAEDLREGAAELVQDLFVDYCYGTEEEIIEWQPDAYRGDYTAWN LDYRNEVRKRSAGLTFVSVGPLFVAHKWIQKKHVVTCTDLLYEATIVIALANDILGVNRDRDDRLCVTS KLSSCSEVVQYHNELVEVLHKKVLELEGNTRRFMEEEVASVGLFLWQCRSKRYAVVFEC
>uBuTS-9_undescribed_Bubarida MSLFAEVIKALPLERPLQYAHSAALNRLRKASESVCSSTKRLQKHCSALKLNQYIEAANFSHFAAK VYNIGVATEPGQDRLGIFVATLALFVFEFDDHFDKPCGTPENVSRLSMEMRILLRALSRLHGLAGLQGV DDWPTGAFGREAYLWLLKEAEDLRKGAELMHYSFLDYCFGVESEILEWAPDAYRGDTTAWNLDY SEVRKRSVGVTITLAPLYIINKWTKKEHFATCNLDLLYNAALIGALANDVLGMKRDQEESSQLGMTALKI VSASEAAQYHNEKVECLKDILDLDGDRRLMEEEVINVAQFLWQCNARRYN
>uBuTS-10_undescribed_Bubarida MSLSVFAKIPVPTLELEHPMHYVDSTVSAKVRISVSAGLNNRLYRRVLDLKLNPYINHMDINDFVAK ACNLGLPTPDQERLWIFTSVNALFIFEFDDHFDKPLLIAPEYVARRSKQMRVLRSLSSHKL SGLQGS EDWPAEVPCKEAYLWLLREAKDLKEGAAELLHDIFVDYCFGVEEEVIEWKSDMYRGDLTAWSLDRYK EVRKRAAGVVFVPLFVTRKWIQKEHVNTCTDLLYEAALLVGLSNDILGIPRDLRDSGTM TVLKIAST SEVVQHHNMMVEVLHKKVLALEGNTMHFMEEVETSVAGVFLWQCHSRRYTL
>uBuTS-11_undescribed_Bubarida FAKIPVPALELEHPVHYADSAVSAKVRISVSAGLNNRLYRRVLGLKMNPEYIKHMDINDFVAKACNL GLPTPDQDRLWIFASVTGLFIFEFDDHFDKPLLIAPEYVARRSKQMRVLRSLSGHKLSGLQGSLEDW PTEVPCKEAYLWLLREAEDQRKGAELLHDVDFIDFCYGVEEEIIEWQSDAHRGDFTAWTFDRYNEAR KHSVGYIFTVPVPLFIINKWIQKEYFTSCIDLLYEAIIIVGLGNDILGMLRDSAEDGSMTVLKIITSHDKIVE VHNEMVEVLHKKVLELECNKRRFMEEMEAAVGLFLWQLDSNRYAGRLPITY
>uBuTS-12_undescribed_Bubarida MSLGVFAKIPVPALELEHPVHYADSAVSAKVRISVSAGLNNRLYRRVLGLKMNPEYIKHMDINDFVA KACNLGLPTPDQDRLWIFASVTGLFIFEFDDHFDKPLLIAPEYVARRSKQMRVLRSLSGHKLSGLQGS LEDWPTEVPCKEAYLWLLREAEDQRKGAELLHDVDFIDFCYGVEEEIIEWQSDAHRGDFTAWTFDRY NEARKHSVGYIFTVPVPLFIINKWIQKEYFTSCIDLLYEAIIIVGLGNDILGMLRDSAEDGSMTVLKIITSHD KIVEVHNKMVEVLHKKVLELECNKRRFMEEMEAAVGLFLWQLDSNRYAGRLPITY

>uBuTS-13_undescribed_Bubarida MLHSQSSLTSSSLAVYKGLQALNSIIACLRPACSCCECKLHCSVIMSLGVFAKIPVPALELEHPVHYADSA VSVAKVRISVSAGLNNRLYRRVLGLKMNPEYIKHMDINDFVAKACNLGLPTDQDRLWIFASVTGLFIF EFDDHFDKPLLIAPEYVARRSKQMRVLRSLSGHKLSSGLQGSMEDWPTEVPCKEAYLWLLREAEDLR KGAAELLHDVFIDFCYGVVEEIIQWQSDAHRGDFTAWTFDRYNEARKHSGYIFTPVVPLFIINKWIKKE YFTSCIDLLYEAIIIVGLGNDILGMLRDSAEDGSMTVLKIITSHDKIVEVHNKMVEVLHKKVLELECNKR FMEEMEAAVGLFLWQLDSNRYAGRLPITY
>uBuTS-14_undescribed_Bubarida MSLSVFTKIPIPAVELEHLMHYADSAVGETKLRKHSVSGELKSLHRHLLDLKLNQYLDHLNIYDCVAK CFNLGVNHHPTLDQNQLWIFASILFLFEFDDHFDNPLHLPENTARLSKEMRAVLRSLSRHNSGLQ GGLEDWPAEVPCKEAYLWLLREAEDLKEGAAELTHGLFVDYCFGVVEEVIEWKSDTYRGDLTAWSLD RYKEVRKRTAAGGFAIVVPLFVTKWIKKEHVHKCTDLLYEASLLALNDILGIPRDRRDSATMTALKI ASTSEVVQHNNMVEVLHKKVLALEGNTKHFMEEEVASVVGFFLWQCHSKRYTL
>PhTS-1_Phakelia_ventilabrum MSLSVFTKIPIPTMKLEPPMYADSAVSVAKLRESLVSEKLNRLHRLVLDLKNPQYIDHMNINDFVAK ACNLGAPTPDQDWLWLFSTVHGIFIFEFDDHFDKPRQIAPECVAKRAKEMRAVLRSLGSHKLSGLQGS LEHWPSEVPCKEAYLWLLRKAELREGAAELLHDVFIDFCNGVEEIIQWQSDAYRGDFTAWNLDHYK EARKRSVGIIYSPVPLFIINKWIKKEYFTTCIDLLYEAIIIVALANDVLGIPRDHGDDGTMVLRITNLCD VIKVHNEMVEELHKKVLELECNAMYFMEEMEAGMVGLFLWQLNSNRYAGRLPITY
>AgTS-1_Agelas_clathrodes MSQFVEVPKISLVLQHPLRYAHSADFSAKLQISSVMSLKDCLHNHCSALKLNPRYIEELNIADYVAKTS NLGVEYHATEEEDRLWIFAATLGIFLFEFDDYFDKPSNTPENVARLSTEMRAVQRLSRHGLSGLQGN DDWPTAVPYREAYCWLLREAEDLRGAEEVHYTFLDYCFGVEMEITEWAPDMYQHDTSSWNLDR NEVRKRSGGTHFAFVGPLYVVKWITKEHFNACNDLLYDAAIVITLANDLLGKKKDLQTENDSISIKTIKI ATTSEIAQHHNEKVECLRKDILQLDGDIRRLMEEEVIVAGLFLWQYNSRRYE
>StTS-1_Stylissa_massa MSRFVEVPKALALQHPFKYAHSALSVAKLREASVSSLLKTRLRNHLSALNLPKYMEALNIPDFIAKIY GIECPSTEDKERQWVFSATFTNCFVFDHDFDEQVGTPEIVAKLSLEMRDILRALSRHTLSGLQGVLDD WPTDVPCKKAYLWLLQEAELRKGAELVHATFNDYCLGVESEIVEWAPDVHRGDMASAWNLDRC VRKRSAAAGNMALAPLYINKWMTREHYRACNDLLYDVSLIILPNDVIQSVKRNHNSISMETTKILSADE IVQCHNKVVECLRKDILELDGDTTRFMEEIEISAVGVFLWQCNCNKRYVD
>AgTS-2_Agelas_tubulata MSQFVEVHIKSLLLQHPLRYAHSALNLAKLNRSSVSSSLKDKFEQHCALNPRSLTDIGYSDFVAKIN NIGVKYHPAEEQDRLWIFAATLFTFLMQFDGHFDKPSITPENISRLTMEIKAILRALSRHGLSGFQGNLD DWPTSLPFREPYLWLLREAEDLRKGAELVHYTFLDYCLGLEMEVTEWAPDVYRHDISSWSLARCYE ARKRSRGQTVDLVAPLYANKWTKKEHFTACIDLFYDVAIITTLANDLLGEKEDLENDHIGMTTLKVANK SKIVQVHNEKVEGLRNDILGLDGDIRRLMEEEVTTAAFLCECTRKKLDYSESDYDS

Table S3. List of TSs included in TSHMM1 and in the TS phylogenetic tree in Figure 3. Grey shaded cells are UniProt IDs, blue shaded cells are NCBI accession numbers, and yellow shaded cells are GenBank accession numbers. White cells are previously published coral TS sequences (13). Orange shaded cells are sponge TSs described in this study, which were included in Figure 3, but not TSHMM1.

Name	Origin	Sesquiterpene synthase activity	Monoterpene synthase activity	Diterpene synthase activity	Major product(s)
uBuTS-1	undescribed Bubarida metagenome and transcriptome	Yes	No	No	(+)-bicyclogermacrene
uBuTS-2	undescribed Bubarida metagenome and transcriptome	Yes	Yes	No	(-)-germacrene D
uBuTS-3	undescribed Bubarida metagenome	Yes	Yes	No	(+)-bicyclogermacrene
uBuTS-4	undescribed Bubarida metagenome	Yes	Yes	No	mixed sesquiterpenes
uBuTS-5	undescribed Bubarida metagenome and transcriptome	Yes	No	No	mixed sesquiterpenes
uBuTS-6	undescribed Bubarida metagenome	No	No	No	None
uBuTS-7	undescribed Bubarida transcriptome	No	No	No	None
uBuTS-8	undescribed Bubarida metagenome and transcriptome	No	No	No	None
uBuTS-9	undescribed Bubarida metagenome	No	No	No	None
uBuTS-10	undescribed Bubarida transcriptome	N/A	N/A	N/A	N/A
uBuTS-11	undescribed Bubarida transcriptome	N/A	N/A	N/A	N/A
uBuTS-12	undescribed Bubarida metagenome	N/A	N/A	N/A	N/A
uBuTS-13	undescribed Bubarida transcriptome	N/A	N/A	N/A	N/A
uBuTS-14	undescribed Bubarida metagenome	N/A	N/A	N/A	N/A
AgTS-1	Agelas clathrodes metagenome	Yes	Yes	No	(+)-alloaromadendrene
AgTS-2	Agelas tubulata metagenome	No	No	No	None
PhTS-1	Phakellia ventilabrum transcriptome	Yes	Yes	No	(-)-germacrene D
StTS-1	Stylissa massa metagenome	No	No	Yes	mixed diterpenes

Table S4. Sponge terpene synthase summary table.

Start	End	Length	Introns?	Blast Annotation	Blast Annotation Organism	Blast Annotation Accession Number	Blast E value	Halichondria panicea hit	H. panicea E value
uBuTS-9 contig 1 (72kb)									
Start	End	Length	Introns?	Blast Annotation	Annotation Organism	Annotation Accession Number	Blast E value	Halichondria panicea hit	H. panicea E value
6851	11435	4585	yes	alsin-like	Stylophora pistillata	XP_022793651.1	2.00E-07	GIFJ01263445.1	3.00E-21
13656	14960	1305	no	bystin-like	Amphimedon queenslandica	XP_00338529.1	2.00E-163	HBWD01518048.1	5.00E-66
15162	15900	739	no	Branched-chain-amino-acid aminotransferase-like	Amphimedon queenslandica	XP_003383850.2	4.00E-37	GIFJ01253076.1	4.00E-16
21281	25316	4037	no	uncharacterized protein	Amphimedon queenslandica	XP_019851048.1	4.00E-83	HBWD01019842.1	7.00E-100
28635	29621	987	no	uBuTS-9					
32769	46513	13744	yes	alsin-like	Amphimedon queenslandica	XP_019854288.1	4.00E-19	HBWD01019722.1	7.00E-43
53000	61934	8934	yes	cyclin-Y-like protein 1	Amphimedon queenslandica	XP_003388010.1	4.00E-15		
64295	66067	1793	yes	no blast hit					
uBuTS-2 contig 1 (63kb)									
3008	1242	1234	yes	piggyBac transposable element derived protein 4-like	Stylophora pistillata	XP_022798311	9.00E-07		
6832	7714	882	no	no blast hit					
8023	8950	927	no	no blast hit				HBWD01105536.1	6.00E-20
11212	15012	3800	yes	no blast hit				HBWD01373978.1	2.00E-13
15451	20081	4630	yes	uncharacterized protein	Amphimedon queenslandica	XP_019855287.1	7.00E-26	HBWD01009091.1	6.00E-69
23055	23823	768	no	uncharacterized protein	Amphimedon queenslandica	XP_019854607.1	2.00E-04	GIFJ01039005.1	5.00E-07
25849	26652	803	no	uncharacterized protein	Ecotocarpus sp. CCAP 13	CAB1113566.1	3.00E-23		
26853	27419	566	no	uncharacterized protein	Ecotocarpus sp. CCAP 13	CAB1116628.1	3.00E-20	GIFJ01283391.1	7.00E-05
31393	32400	1007	no	uBuTS-2					
36476	50257	13781	yes	uncharacterized/retrovirus related pol polyprotein	Stylophora pistillata	PFX30308.1	4.00E-23	GIFJ01196170.1	1.00E-57
52757	63023	10266	yes	uncharacterized protein	Oospasas minuta	KA16647352.1	1.00E-36	GIFJ01515311.1	1.00E-12
uBuTS-1 contig 1 (74kb)									
18935	19957	1022	no	uBuTS-1					
20717	21677	960	no	hypothetical protein LOC99_12348	Oospasas minuta	KA16647352.1	1.00E-32		
25869	29525	3656	no	hypothetical protein LOC99_12348	Oospasas minuta	KA16647352.1	1.00E-37		
33244	34419	1175	yes	pogo transposable element with ZNF domain-like	Saccoglossus kowalevskii	XP_006822101.1	1.00E-54	GIFJ01059185.1	3.00E-05
36033	36819	786	no	hypothetical protein LOC99_12348	Oospasas minuta	KA16647352.1	4.00E-33		
50879	55506	4627	yes	PREDICTED: serine/threonine-protein kinase svkA-like	Amphimedon queenslandica	XP_019855734.1	4.00E-82	HBWD01267899.1	5.00E-144
70866	73178	2312	no	PREDICTED: serine/threonine-protein kinase 10-like	Amphimedon queenslandica	XP_019859200.1	2.00E-65	GIFJ01029571.1	3.00E-134
uBuTS-8 contig 1 (61kb)									
3727	7370	3643	yes	PREDICTED: runt-related transcription factor	Amphimedon queenslandica	XP_019858567.1	2.00E-33	GIFJ01194107.1	8.00E-56
14197	19827	5630	yes	runt-related transcription factor	Amphimedon queenslandica	ACF696957.1	3.00E-49	HBWD01004530.1	5.00E-54
46761	47771	1010	no	uBuTS-8					
50497	60666	10169	yes	uncharacterized protein LOC11340013	Stylophora pistillata	XP_022802513.1	7.00E-22	HBWD0101394.1	4.00E-54
uBuTS-1 contig 2 (57kb)									
1802	2815	1013	no	uBuTS-1					
4622	7297	2675	yes	PREDICTED: 60S ribosomal protein L12-like	Amphimedon queenslandica	XP_003382502.1	2.00E-26	GIFJ01084458.1	1.00E-21
9219	11890	2671	no	Hyp4351	Branchiostoma lanceolatum	CAH1270521.1	3.00E-80	GIFJ01153800.1	2.00E-151
13646	14517	871	yes	Hypothetical predicted protein	Paramunicea clavata	CAB4002751.1	9.00E-45	HBWD01137827.1	5.00E-53
20900	26503	5603	yes	PREDICTED: uncharacterized protein LOC109582994	Amphimedon queenslandica	XP_019853667.1	2.00E-60	HBWD01255793.1	1.00E-49
41880	42758	878	yes	archaemetzincin-2-like	Stylophora pistillata	XP_022791489.1	3.00E-60	HBWD01208581.1	1.00E-31
uBuTS-14 contig 1 (47kb)									
1021	5706	4685	yes	PREDICTED: plectastrin homology domain-containing family F member 2-like isoform X2	Amphimedon queenslandica	XP_019864174.1	2.00E-26	HBWD01207577.1	6.00E-33
13957	15744	1787	yes	PREDICTED: plectastrin homology domain-containing family F member 2-like isoform X3	Amphimedon queenslandica	XP_003382481.1	1.00E-10	HBWD01207582.1	3.00E-15
				PREDICTED: proprotein convertase subtilisin/kexin type 9-like	Amphimedon queenslandica	XP_019849385.1	3.00E-13	GIFJ01293424.1	3.00E-13
28994	30166	1172	yes	PREDICTED: uncharacterized protein LOC100638844	Amphimedon queenslandica	XP_019849384.1	9.00E-57	GIFJ01125311.1	4.00E-102
30301	36391	6090	yes	PREDICTED: uncharacterized protein LOC100638844	Amphimedon queenslandica	XP_019849384.1	9.00E-57	GIFJ01125311.1	4.00E-102
42985	44406	1421	yes	hypothetical protein MCG8623886.1	Proteobacteria bacterium	MCG8623886.1	1.00E-51	HBWD01267899.1	1.30E-43
44681	45673	992	no	uBuTS-14					
uBuTS-2 contig 2 (272 kb)									
219	3532	3313	yes	PREDICTED: vinculin-like	Amphimedon queenslandica	XP_003382439.1	7.00E-11	HBWD01162922.1	2.00E-15
11347	27492	16145	yes	PREDICTED: mucin-17-like	Amphimedon queenslandica	XP_019849308.1	4.00E-06	GIFJ01353008.1	4.00E-21
57307	72065	14758	yes	TPR and ankyrin repeat-containing protein 1-like	Halolitis rufescens	XP_046368821.2	5.00E-145	GIFJ01089742.1	0
73001	77134	4133	no	uncharacterized protein K02A2_6-like	Oreochromis aureus	XP_039455996.1	0	GIFJ01340966.1	0
78527	88037	9510	yes	interferon-stimulated 20 kDa exonuclease-like 2	Chrysomya picta bellii	XP_005310999.1	4.00E-17	GIFJ01335375.1	4.00E-17
88149	99651	11502	yes	PREDICTED: protein asteroid homolog 1-like	Amphimedon queenslandica	XP_011407600.1	3.00E-76	HBWD01172737.1	1.00E-86
109646	121637	11991	yes	TPR and ankyrin repeat-containing 1-like	Paramunicea clavata	CAB4029118.1	8.00E-160	GIFJ01089739.1	0.00E+00
123248	136125	12877	yes	TPR and ankyrin repeat-containing protein 1-like isoform X2	Orbicella faveolata	XP_020607958.1	3.00E-145	GIFJ01089738.1	0.00E+00
141266	143344	2078	yes	Hypothetical predicted protein	Mytilus galloprovincialis	VDI88231.1	3.00E-11	GIFJ01053260.1	8.00E-08
151338	160040	8702	yes	TPR and ankyrin repeat-containing protein 1	Lingula anatina	XP_013405370.1	9.00E-161	GIFJ01089740.1	0.00E+00
170415	180500	10085	yes	PREDICTED: uncharacterized protein LOC100635005	Amphimedon queenslandica	XP_019859701.1	0	HBWD01148209.1	5.00E-117
193727	195100	1373	no	protein O-mannose kinase-like isoform X1	Acropora millepora	XP_029212908.2	1.00E-48	HBWD01506726.1	4.00E-66
196853	203076	6223	yes	PREDICTED: balbiani ring protein 3-like	Amphimedon queenslandica	XP_019849443.1	2.00E-10	HBWD01221237.1	1.00E-29
210324	212035	1711	no	PREDICTED: balbiani ring protein 3-like	Amphimedon queenslandica	XP_019849443.1	2.00E-37	HBWD01261137.1	5.00E-46
219388	221995	2607	yes	hypothetical protein BSL78_12423	Aposichopus japonicus	MRZV01000408.1	9.00E-20	HBWD01045466.1	8.00E-21
237060	240800	3740	no	PREDICTED: balbiani ring protein 3-like	Amphimedon queenslandica	XP_019849443.1	4.00E-109	HBWD01261143.1	1.00E-166
244030	247157	3127	yes	PREDICTED: 40S ribosomal protein S11-like	Amphimedon queenslandica	XP_003382520.1	5.00E-33	HBWD01596638.1	5.00E-32
247280	253239	5959	yes	protein kinase	Proteobacteria bacterium	MCG8623461.1	5.00E-99	GIFJ01138049.1	5.00E-112
266184	267194	1010	no	uBuTS-2					
267920	269747	1827	no	hypothetical protein LOC99_12348	Oospasas minuta	KA16647352.1	6.00E-32		

Table S5. Annotations for predicted genes on TS-containing contigs from the undescribed Bubarida sponge. Gene and organism annotations are based on the closest NCBI blastx hit against the non-redundant protein sequences database. The contigs were also queried against a custom BLAST database generated from two publicly available *Halichondria panicea* transcriptomes (HBWD000000000 and GIFJ000000000) using tblastx, the results of which are shown in the 2 right-most columns.

Repository	Accession	Sequence
BioProject	PRJNA907134	Undescribed Bubarida BioProject
BioProject	PRJNA824609	<i>Agelas clathrodes</i> and <i>Agelas tubulata</i> BioProject
BioSample	SAMN31952758	Undescribed Bubarida specimen SCS0036
BioSample	SAMN31952757	Undescribed Bubarida specimen SCS0044
BioSample	SAMN27408796	<i>Agelas clathrodes</i> specimen SIA219
BioSample	SAMN08501975	<i>Agelas tubulata</i> specimen SBM112
SRA	SRR22527612	Illumina undescribed Bubarida ribodepleted RNA-seq
SRA	SRR22527613	Illumina undescribed Bubarida poly-A enriched RNA-seq
SRA	SRR22527614	Nanopore undescribed Bubarida metagenomic sequencing
SRA	SRR22527615	Illumina undescribed Bubarida metagenomic sequencing
SRA	SRR18681978	<i>Agelas clathrodes</i> metagenomic sequencing
SRA	SRR23375712	<i>Agelas tubulata</i> metagenomic sequencing
GenBank	OP977975-OP977983	uBuTS genes
GenBank	OQ320741-OQ320747	uBuTS contigs

Table S6. Accession numbers of sequences from this study.

Structure	Name	IUPAC name	Molecular Formula	Method of identification	Figures appeared in
	8-Myrcene	7-methyl-3-methylenoocta-1,6-diene	C10H16	NIST spectral library comparison	S5, S24, S25-S27, S32, S34
	Ocimene	3,7-dimethylocta-1,3,6-triene	C10H16	NIST spectral library comparison	S5
	(-)-germacrene D 8	(S,1E,6E)-8-isopropyl-1-methyl-5-methylenecyclodeca-1,6-diene	C15H24	NMR, OR	S5-S6, S16, S19-S20, S38, S41, S44, S48
	(E)-β-Farnesene	(Z)-7,11-dimethyl-3-methylenedodeca-1,6,10-triene	C15H24	NIST spectral library comparison	S7, S46
	farnesol	(Z,E,6E)-3,7,11-trimethyldodeca-2,6,10-trien-1-ol	C15H26O	NIST spectral library comparison	S7, S47
	cembrene A	(1E,5E,9E,12R)-1,5,9-Trimethyl-12-(prop-1-en-2-yl)cyclotetradeca-1,5,9-triene	C20H32	comparison to purified standard	S7
	4-ethoxy ethylbenzoate	4-Ethoxy ethylbenzoate	C11H14	NIST spectral library comparison	S8-S11
	(+)-bicyclogermacrene 6	(1S,2E,6E,10R)-3,7,11,11-tetramethylbicyclo[8.1.0]undeca-2,6-diene	C15H24	NMR, OR	S15, S17, S19, S37-S38, S41, S45, S48
	(+)-alloaromadendrene 7	(1aS,4aR,7S,7aS,7bR)-1,1,7-trimethyl-4-methylenedecahydro-1H-cyclopropa[e]azulene	C15H24	NMR, OR	S21, S45
	linalool	3,7-dimethylocta-1,6-dien-3-ol	C10H18O	NIST spectral library comparison	S22-S35
	1,2-dibromobenzene	1,2-dibromobenzene	C6H4Br2	NIST spectral library comparison	S22-S24, S27- S35
	β-springene	(6E,10E)-7,11,15-trimethyl-3-methylenhexadeca-1,6,10,14-tetraene	C20H32	NIST spectral library comparison	S22, S24-S25, S27-S28, S30-S34
	α-springene	(3E,6E,10E)-3,7,11,15-tetramethylhexadeca-1,3,6,10,14-pentaene	C20H32	NIST spectral library comparison	S22, S24-S25, S27, S30-S34
	geranylinalool	(6E,10E)-3,7,11,15-tetramethylhexadeca-1,6,10,14-tetraen-3-ol	C20H34O	NIST spectral library comparison	S22, S27, S33

Structure	Name	IUPAC name	Molecular Formula	Method of identification	Figures appeared in
	decane	decane	C ₁₀ H ₂₂	NIST spectral library comparison	S23, S28, S29
	cis-β-Ocimene	(Z)-3,7-dimethylocta-1,3,6-triene	C ₁₀ H ₁₆	NIST spectral library comparison	S24, S26
	trans-β-Ocimene	(E)-3,7-dimethylocta-1,3,6-triene	C ₁₀ H ₁₆	NIST spectral library comparison	S24-S26, S32, S34
	geraniol	(E)-3,7-dimethylocta-2,6-dien-1-ol	C ₁₀ H ₁₈ O	NIST spectral library comparison	S24
	α-phellandrene	5-isopropyl-2-methylcyclohexa-1,3-diene	C ₁₀ H ₁₆	NIST spectral library comparison	S34
	γ-terpinene	1-isopropyl-4-methylcyclohexa-1,4-diene	C ₁₀ H ₁₆	NIST spectral library comparison	S34
		1-methyl-5-(prop-1-en-2-yl)cyclohex-1-ene	C ₁₀ H ₁₆	NIST spectral library comparison	S34
	tributylamine	tributylamine	C ₁₂ H ₂₇ N	NIST spectral library comparison	S37, S47
		(1S,2S,3S,6S)-3,7,7-trimethyl-2-(prop-1-en-2-yl)-3-vinylbicyclo[4.1.0]heptane	C ₁₅ H ₂₄	NIST spectral library comparison	S37
		[2aS,2a1R,4aS,7aR]-1,1,2a,5-tetramethyl-2,2a,2a1,3,4,4a,7,7a-octahydro-1H-cyclopenta[cd]indene	C ₁₅ H ₂₄	NIST spectral library comparison	S37
		(1aS,7bR)-3,3,5,7b-tetramethyl-1a,2,3,4,4a,5,6,7b-octahydro-1H-cyclopropa[e]azulene	C ₁₅ H ₂₄	NIST spectral library comparison	S40
	brasila-5,10-diene	(3S,3aR)-3,5,5-trimethyl-7-(prop-1-en-2-yl)-2,3,3a,4,5,6-hexahydro-1H-indene	C ₁₅ H ₂₄	NIST spectral library comparison	S40
	caryophyllene	(1R,9S,Z)-4,11,11-trimethyl-8-methylenebicyclo[7.2.0]undec-4-ene	C ₁₅ H ₂₄	NIST spectral library comparison	S40
	β-copaene	(1R,2S,8S)-8-isopropyl-1-methyl-3-methylenetricyclo[4.4.0.0.2,7]decane	C ₁₅ H ₂₄	NIST spectral library comparison	S41

Table S7. Summary table for compounds shown in the Supplementary Information.

Supplementary Figures

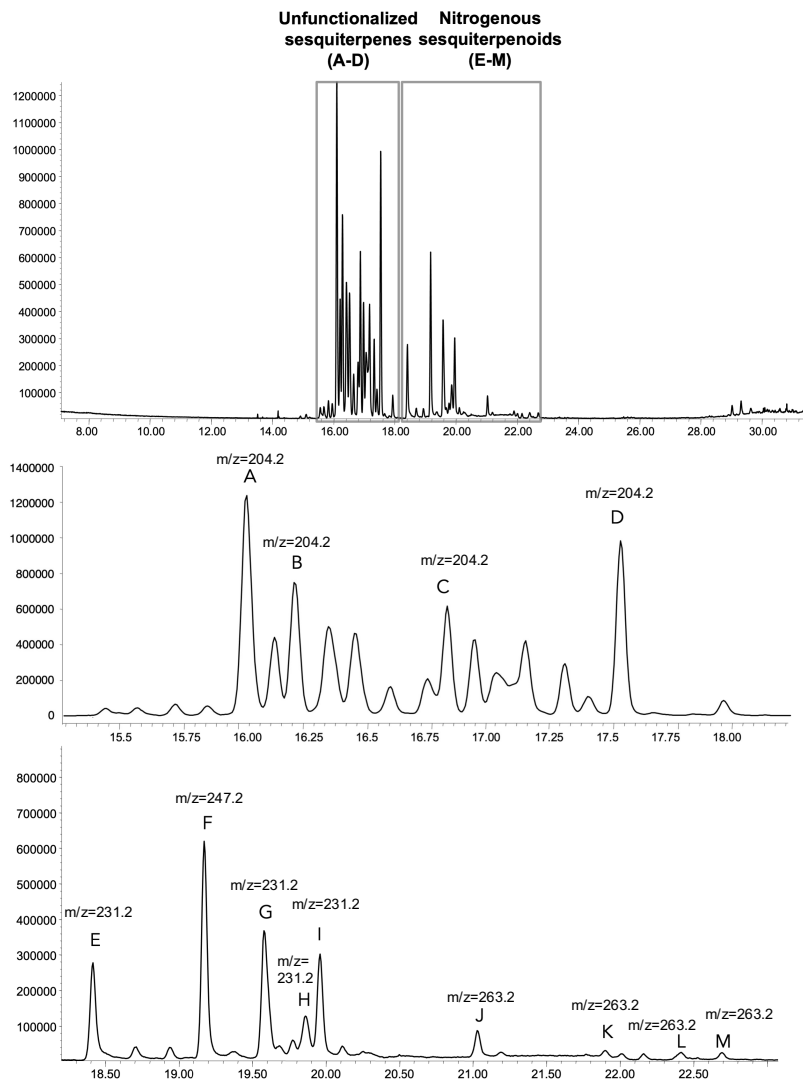


Figure S1. GCMS analysis of crude methylene chloride extracts of undescribed Bubarida sponge. Top: complete GCMS chromatogram; middle: Zoom on signals corresponding to unfunctionalized sesquiterpenes; bottom: Zoom on signals corresponding to isonitrile sesquiterpenoids. Mass spectra for major peaks are shown in Fig. S2 in alphabetized boxes on the next page.

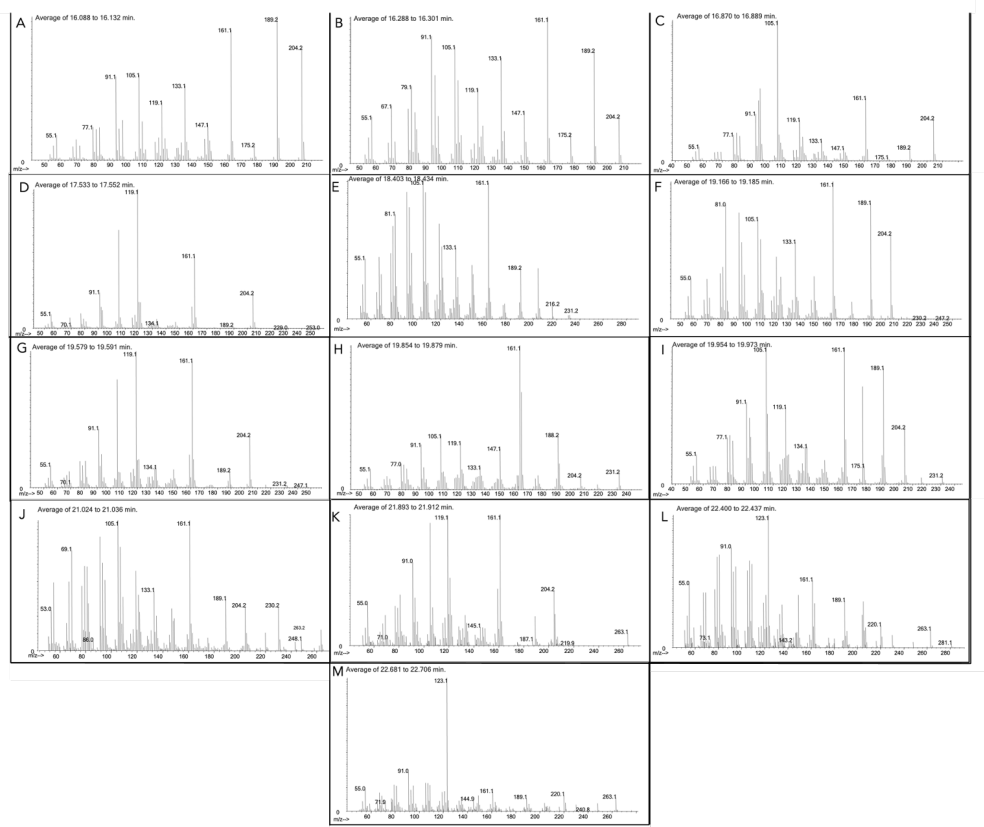


Figure S2. Mass spectra for the alphabetized peaks shown in Fig. S1.

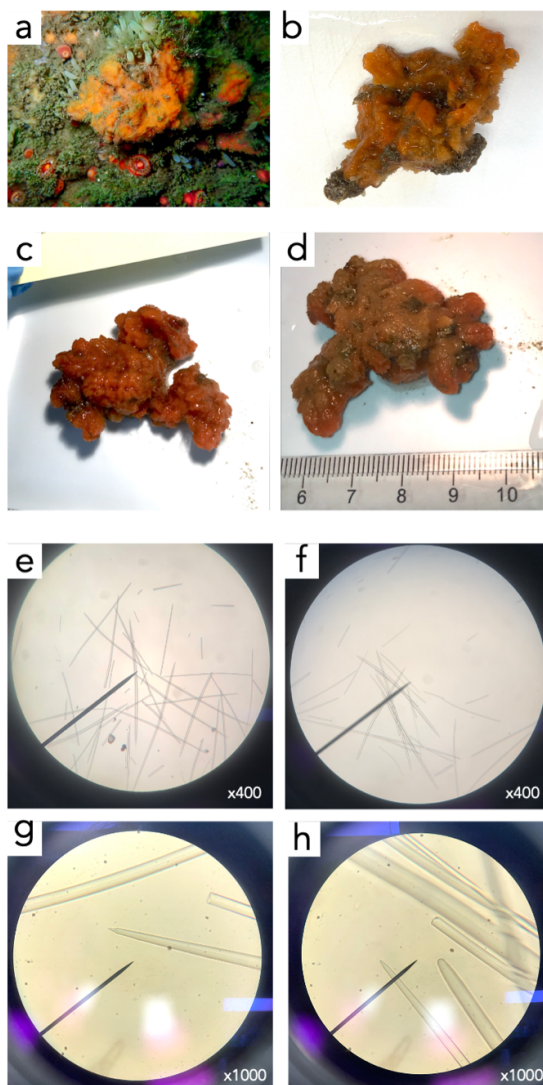
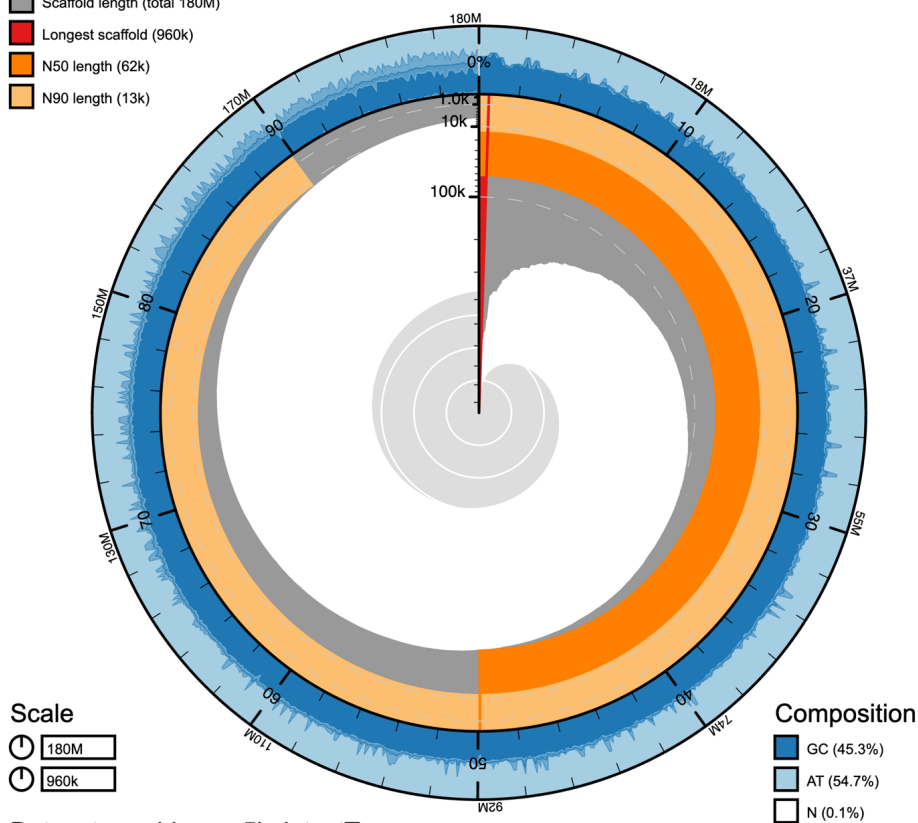


Figure S3. A. Underwater *in situ* photo of the undescribed Bubarida at God's Rock (N32° 51.344', W117° 16.621') in San Diego, California, United States and B. the same undescribed Bubarida specimen in the laboratory post-collection. C. Top view of another undescribed Bubarida specimen and D. the underside of that same undescribed Bubarida specimen with metric ruler for scale. E-F. Light microscopy images of undescribed Bubarida spicules at x400 magnification and G-H. at x1000 magnification.

Scaffold statistics

- Log10 scaffold count (total 5.5k)
- Scaffold length (total 180M)
- Longest scaffold (960k)
- N50 length (62k)
- N90 length (13k)



Dataset: scs44_over5k_latestTaxonomy

Figure S4. BlobTools snail plot of contigs 5kb and longer from the hybrid Illumina short-read and Oxford Nanopore long-read undescribed Bubarida metagenome assembly. The contig lengths are shown in dark grey, with the radius scaled to the length of the longest contig (960kb), which is shown in red. The N50 (62kb) is shown in orange and the N90 (13kb) is shown in pale orange. The entire 180 Mb assembly is represented in the outer ring, with GC/AT percentage represented in dark and light blue.

Undescribed Bubarida FPPS1 + 2 eq. IPP + DMAPP + uBuTS-2

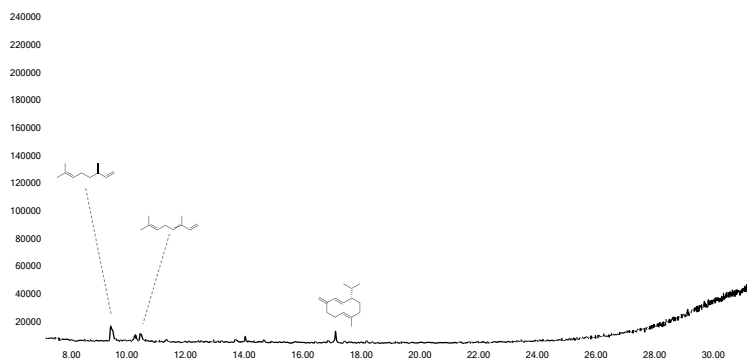


Figure S5. GCMS TIC trace of paired FPP synthase 1/uBuTS-2 assay. Structure annotations are based on NIST spectral library comparison and comparison with (-)-germacrene D **8** standard.

Uncharacterized Bubarida FPPS2 + 2 eq. IPP + DMAPP + uBuTS-2

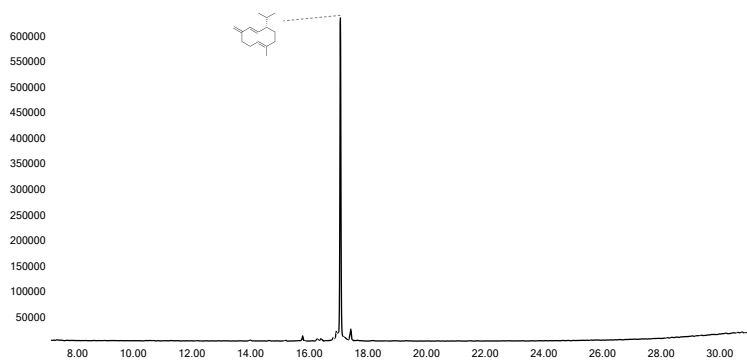


Figure S6. GCMS TIC trace of paired FPP synthase 2/ uBuTS-2 assay. Structure annotations are based on NIST spectral library comparison and comparison with (-)-germacrene D **8** standard.

Undescribed Bubarida GGPPS + 3 eq. IPP + DMAPP + ErTC-2
(cembrene A synthase from Burkhardt et al., 2022)

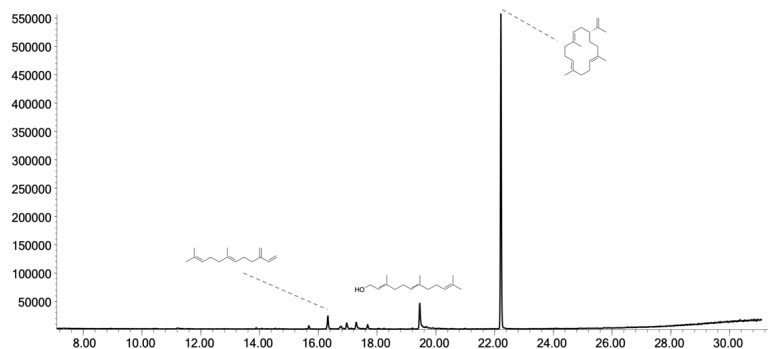


Figure S7. GCMS TIC trace of paired GGPP synthase/ErTC-2 assay(13). Structure annotations are based on NIST spectral library comparison and comparison with cembrene A standard.

2 eq. IPP + DMAPP + uBuTS-2 (negative control 1 for FPPS1 and FPPS2)

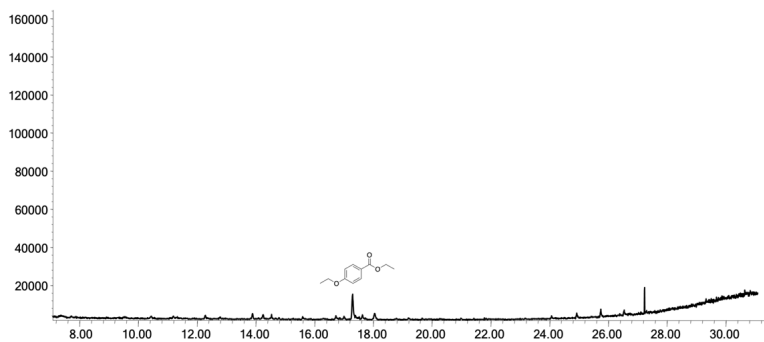


Figure S8. GCMS TIC trace of negative control 1 for FPPS1 and FPPS2. Structure annotations are based on NIST spectral library comparison.

2 eq. IPP + DMAPP (negative control 2 for FPPS1 and FPPS2)

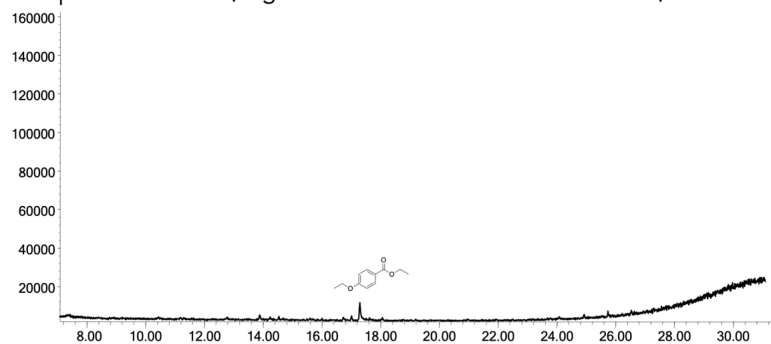


Figure S9. GCMS TIC trace of negative control 2 for FPPS1 and FPPS2. Structure annotations are based on NIST spectral library comparison.

3 eq. IPP + DMAPP + ErTC-2 (cembrene A synthase from Burkhardt et al., 2022) (negative control 1 for GGPPS)

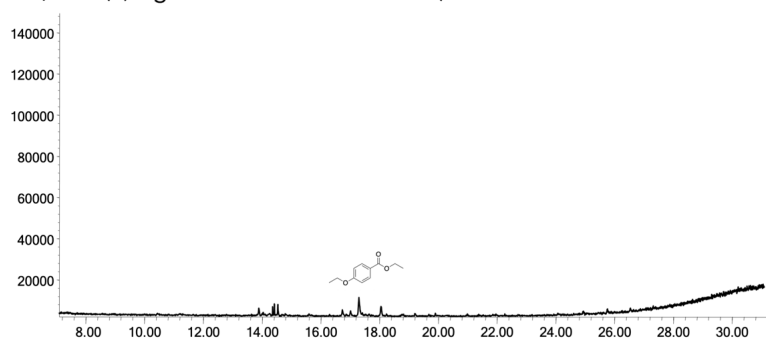


Figure S10. GCMS TIC trace of negative control 1 for GGPPS(13). Structure annotations are based on NIST spectral library comparison.

3 eq. IPP + DMAPP (negative control 2 for GGPPS)

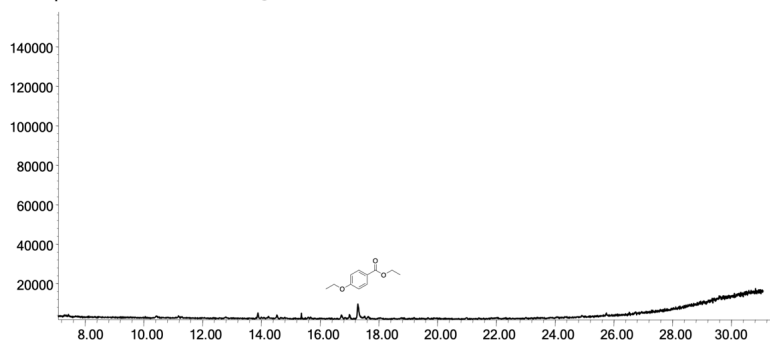


Figure S11. GCMS TIC trace of negative control 2 for GGPPS. Structure annotations are based on NIST spectral library comparison.

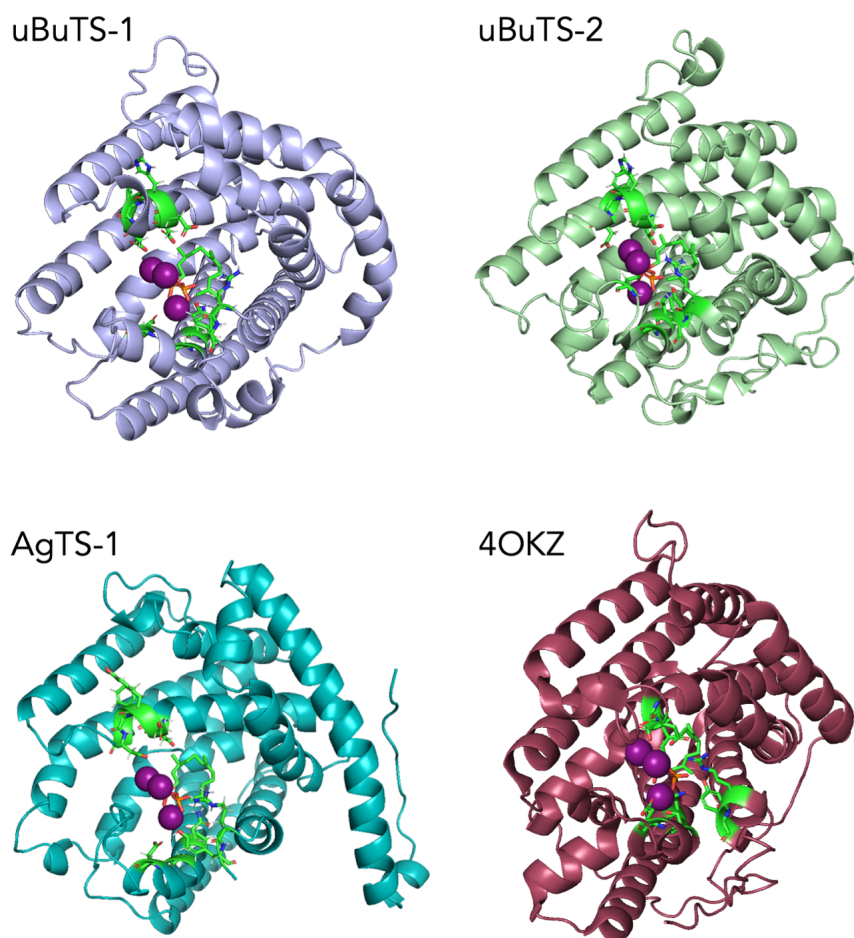


Figure S12. Homology models of uBuTS-1, uBuTS-2, and AgTS-1 and crystal structure of selinadiene synthase (PDB entry 4OKZ) (28). Amino acid sequences of sponge TSs were submitted to Robetta (<https://robetta.bakerlab.org/>) for structure prediction using *Eleutherobia rubra* cembrane A synthase (PDB entry 7S5L) as a template. Sponge TS homology models were visualized using PyMOL (<https://pymol.org/>). FPP and Mg^{2+} locations are based on selinadiene synthase (PDB entry 4OKZ). Typical conserved residues for type I TSs are shown as sticks and highlighted in green.

>TRINITY_DN66204_c0_g1_i4 len=1370 path=[3:0-267 6:268-316 7:317-371 9:372-397 11:398-485 12:486-575 13:576-584 15:585-596 16:597-605 18:606-606 19:607-620 21:621-621 22:622-633 23:634-668 24:669-669 26:670-675 27:676-676 29:677-678 30:679-748 33:749-749 36:750-765 39:766-766 42:767-824 45:825-901 48:902-975 51:976-1000 54:1001-1037 58:1038-1039 59:1040-1045 61:1046-1049 62:1050-1073 64:1074-1077 65:1078-1093 68:1094-1369]

GCAGGTAGGAAGTTAGTTAAAGGCTATTAAGGATGAAATGCTACACAGCCAGCTCAGTAGTCTAACTTCTCTCTAAATGTATTT
GCACTTTGACCTTAGGGCAGTATATAAAGGTGTTCTGCAGGCTCTGAATAATTGCCTGTTTAAGACCTGCATGCAGCTGCAAA
GTAAATTGTATTGTTTCAGTCATCATGTCACTTAGTGTGTTTGGCAAGATTCTGTCCCTACATTGGAGCTTGAGCATCCAATGCA
TTATGTAGACTCTACTGTTAGCGTGGCTAAAGTTCGAATATCATCAGTGAGTGCAGGACTAAACAATAGATTGTATCGCCGTG
TATTGGATTAAACTCAACCCGGAATACATCAACCACATGGATTAATGACTTTGTTGCCAAAGCCTGCAATCTCGGTCTT
CCAACCCCTGACCAGGAACGGCTGTGGATTTCACCTTCAGTTAATGCACTTTTATATTCCGAGTTCGACGATCACTTTGACAA
GCCACTACTCATCGCACCAGAGTATGTTGCAAGGCGATCCAAGCAGATGAGGGCTGTGTTGAGAAGCCTAAGCAGCCACAA
GCTTAGTGGCCTGCAGGGAAGCCTGGAGGATTGGCCTGCTGAAGTTCATGCAAGGAGGCTTACCTTTGGCTTTAAGGGAA
GCTAAAGATCTGAAGGAGGGCGCAGCAGAGCTGCTCATGATATTTTGTGACTATTGTTTCGGTGTGAAGAGGAGGTTAT
CGAGTGGAAATCAGATATGTATCGGGGTGATCTTACTGCCTGGAGTCTTGACCCTACAAGGAGGTTGCGAAACGTGCAGCA
GGTGTAGTTTTTGAATCGTAGTACCCTCTTCGTCACCCGTAAGTGGATACAGAAGGAGCATGTCAACACATGCAGTGACC
TTCTGTATGAGGCTGCACTCCTGGTTGGACTCTCGAACGACATCTAGGAATCCCAAGGGACCTTAGAGACAGCGGAACAT
GCAGTGCTCAAGATAGCAAGCAGTGAAGTGTCCAGCATATAATGATGTTGAAGTACTCCACAAGAAAGTTCTA
GCCTTGAAGGCAACACAATGCACCTTCATGGAAGAGGTGGAGACTAGTGTAGCGGGGGTTTCTGTTGGCAGTGCCATTCA
AGACGTTATACCTTTGTGAATTGTATTTTTCTCTATAGCTGTATAATCTGTGTATTCCTCATGATAATTCAGAACAAATCAAGTT
TTAGTAATATCATCTTTTATAAATATTGTCATGCTACATTAATTGATGAATTAACAAATTTGTAAGTTTTTGAATTTGTCTCTGC
CACTGACTGAATAAAATTTGCAAAATTA

>TRINITY_DN66204_c0_g1_i4.p1 type:complete len:328 gc:universal TRINITY_DN66204_c0_g1_i4:193-1176(+)
MSLSVFVKIPVPTLELEHPMHYVDSTVSVAKVRISVSAGLNNRLYRRVLDLKLNPYINHMDINDFVAKACNLGLPTDQERLWIFTS
VNALFIEFDDHFDKPLLIAPEYVARRSKQMRVLRSLSSHKL SGLQGSLEDWPAEVPCKEAYLWLLREAKDLKEGAAELLHDIFVDY
CFGVEEVIEWKSDMYRGDLTAWSLDRYKEVRKRAAGVVFVAVVPLFVTRKWQKEHVNTCTDLLYEALLVGLSNDILGIPDRDRS
GTMTVLKIASTSEVVQHNNMMVEVLHKKVLALEGNTHMFEEVETSVAGVFLWQCCHSRRYTL*

Figure S13. Polyadenylated transcript with uBuTS-10 sequence bolded. Translated uBuTS-10 amino acid sequence shown below.

CLUSTAL O(1.2.4) multiple sequence alignment

Ga0070431_10856351	-----	0
Ga0072505_11956581	MSRFVEVPKALALQHPFKYAHSAHSVAKLREASVSLKTRLRNHLNALNPKYMEAL	60
StTS-1	MSRFVEVPKALALQHPFKYAHSAHSVAKLREASVSLKTRLRNHLNALNPKYMEAL	60
Ga0070431_10856351	-----CFFVFDHDFDEQVGTPEIVAKLSLEMRD	28
Ga0072505_11956581	NIPDFIAKIYGIECPSTEDKERQWFSATFTNCFFVFDHDFDEQVGTPEIVAKLSLEMRD	120
StTS-1	NIPDFIAKIYGIECPSTEDKERQWFSATFTNCFFVFDHDFDEQVGTPEIVAKLSLEMRD	120

Ga0070431_10856351	ILRALSRLTSLGLQGVLDWPTDVPCKKAYLWLLQEAEGLRKGAELVHATFNDYCLGVE	88
Ga0072505_11956581	ILRALSRLTSLGLQGVLDWPTDVPCKKAYLWLLQEAEGLRKGAELVHATFNDYCLGVE	180
StTS-1	ILRALSRLTSLGLQGVLDWPTDVPCKKAYLWLLQEAEGLRKGAELVHATFNDYCLGVE	180

Ga0070431_10856351	SEIVEWAPDVHRGMSAWNLDRCETVRKRSAGNMALAPLYVINKWMTREHYRACNDLLY	148
Ga0072505_11956581	SEIVEWAPDVHRGMSAWNLDRCETVRKRSAGNMALAPLYVINKWMTREHYRACNDLLY	240
StTS-1	SEIVEWAPDVHRGMSAWNLDRCETVRKRSAGNMALAPLYVINKWMTREHYRACNDLLY	240

Ga0070431_10856351	DVSLIIALPNDVIQSVKRNHNSISMETTKILSADEIVQCHNKKVECLRKDIELDGDTRR	208
Ga0072505_11956581	DVSLIIALPNDVIQSVKRNHNSISME-----	266
StTS-1	DVSLIIALPNDVIQSVKRNHNSISMETTKILSADEIVQCHNKKVECLRKDIELDGDTRR	300

Ga0070431_10856351	FMEIEIEISAVGVFLWQCNCRYVD*	232
Ga0072505_11956581	-----	266
StTS-1	FMEIEIEISAVGVFLWQCNCRYVD-	324

Figure S14. Clustal Omega (1.2.4) multiple sequence alignment of synthesized gene StTS-1 and the *Stylissa massa* TS fragments of which StTS-1 is comprised: Ga0070431_10856351(IMG study ID Gs0114528, project ID Gp0112735) and Ga0072505_11956581 (IMG study ID Gs0114528, project ID Gp0114746). Predicted key active site motifs are highlighted in yellow.

uBuTS-1

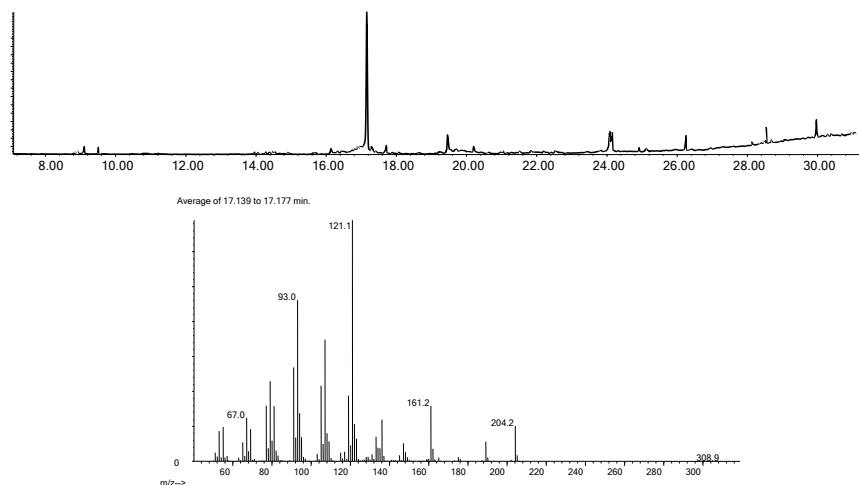


Figure S15. – GCMS TIC trace of the hexanes extract of *in vitro* assay with purified uBuTS-1 and FPP. The mass spectrum of the major peak is shown below. The structure of the main product, (+)-bicyclogermacrene **6**, is based on full structural characterization by NMR and OR (Supplementary Note).

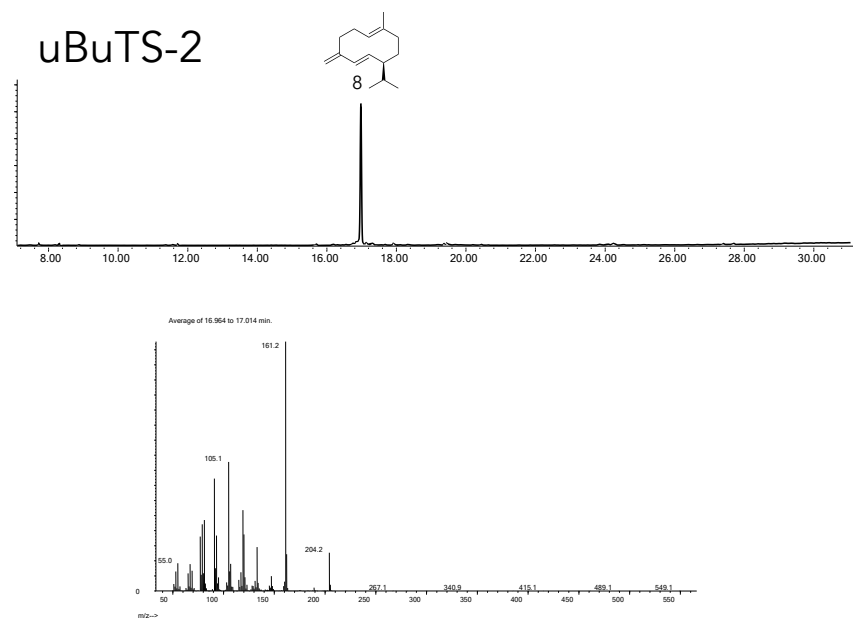


Figure S16. GCMS TIC trace of the hexanes extract of *in vitro* assay with purified uBuTS-2 and FPP. The mass spectrum of the major peak is shown below. The structure of the main product, (-)-germacrene D **8**, is based on full structural characterization by NMR and OR (Supplementary Note).

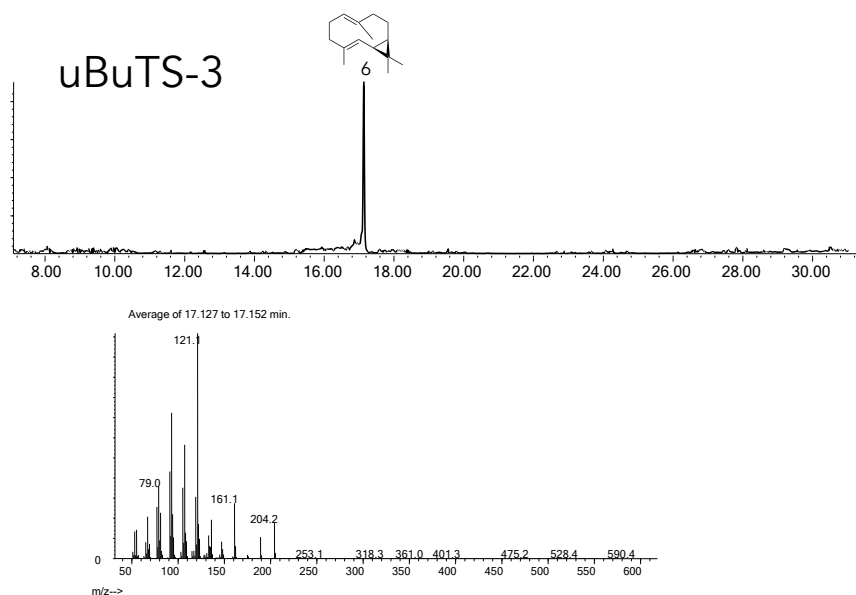
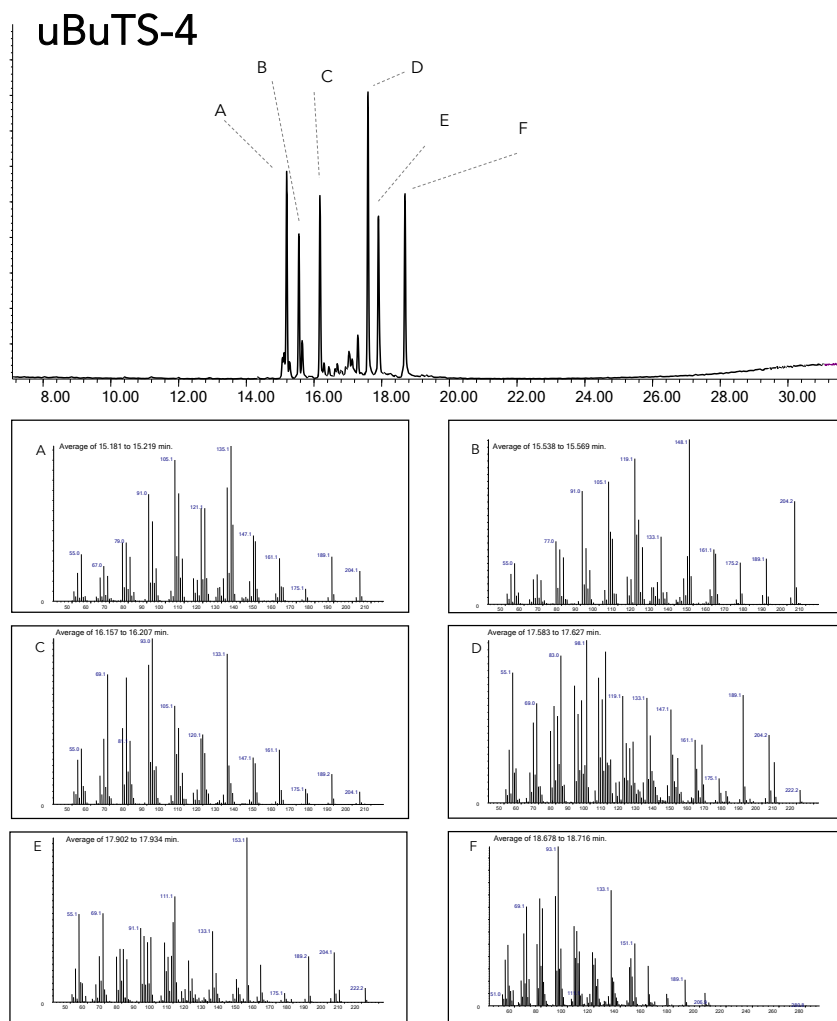
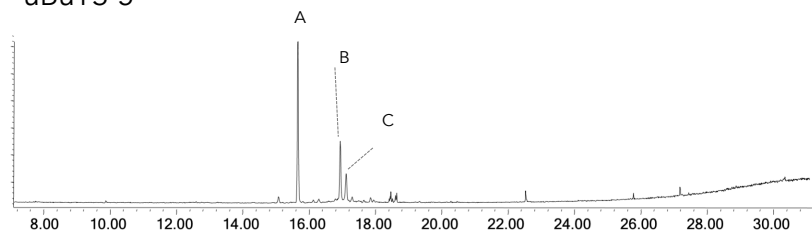


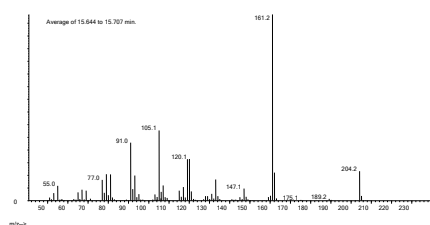
Figure S17. GCMS TIC trace of the hexanes extract of an *in vitro* assay with purified uBuTS-3 and FPP. The mass spectrum of the major peak is shown below. The structure of the main product, (+)-bicyclogermacrene **6**, is based on GCMS with purified standard.



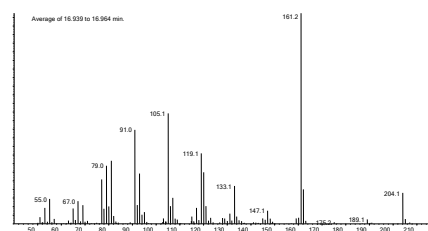
uBuTS-5



A



B



C

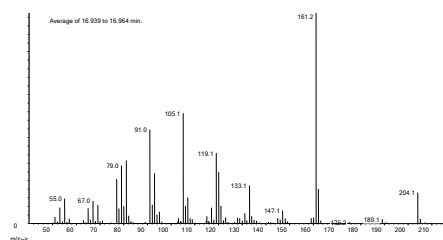


Figure S19. GCMS TIC trace of the hexanes extract of an *in vitro* assay with purified uBuTS-5 and FPP. Structure annotations are based on NIST spectral library comparison and comparison with (+)-bicyclogermacrene **6** and (-)-germacrene D **8** standards. Mass spectra for major peaks are shown in alphabetized boxes below the GCMS trace.

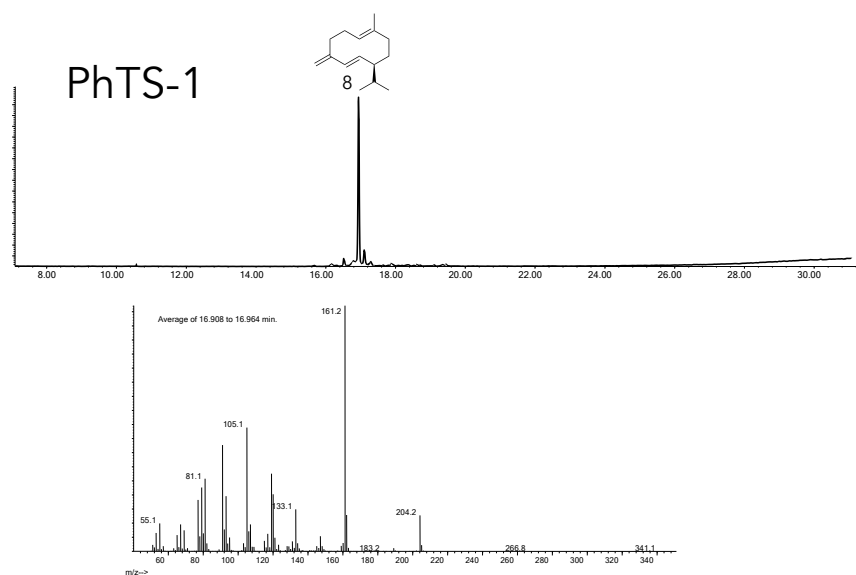


Figure S20. GCMS TIC trace of the hexanes extract of an *in vitro* assay with purified **PhTS-1** and FPP. The mass spectrum of the major peak is shown below. The structure of the main product, (-)-germacrene **D 8**, is based on GCMS with purified standard.

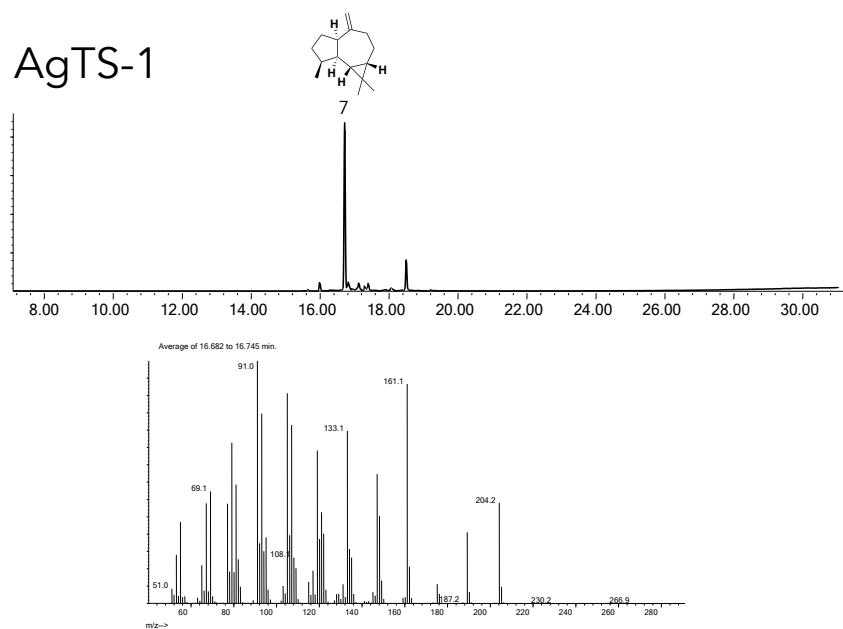


Figure S21. GCMS TIC trace of the hexanes extract of an *in vitro* assay with purified AgTS-1 and FPP. The mass spectrum of the major peak is shown below. The structure of the main product, (+)-allo aromadendrene **7**, is based on full structural characterization by NMR and OR (Supplementary Note).

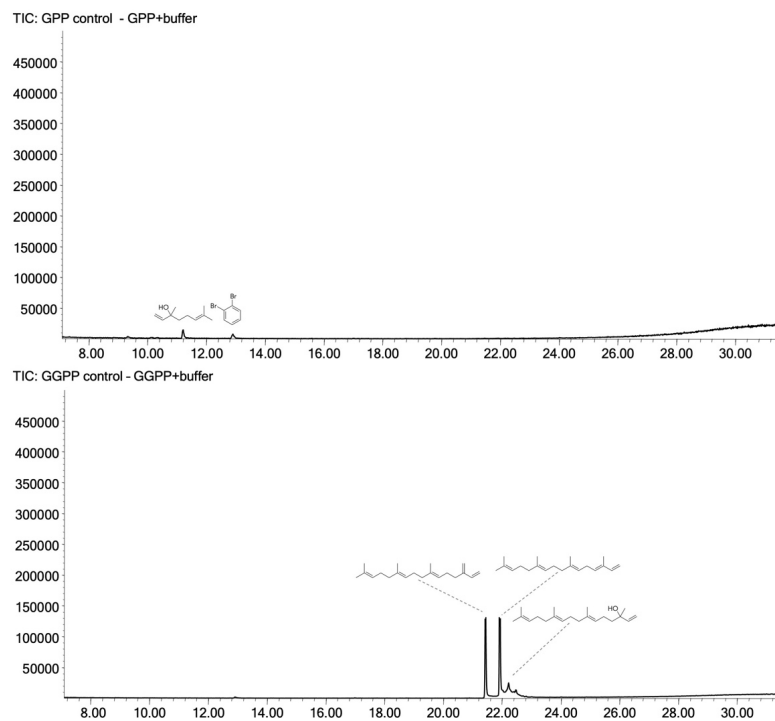


Figure S22. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) controls. Annotations based on NIST spectral library comparison.

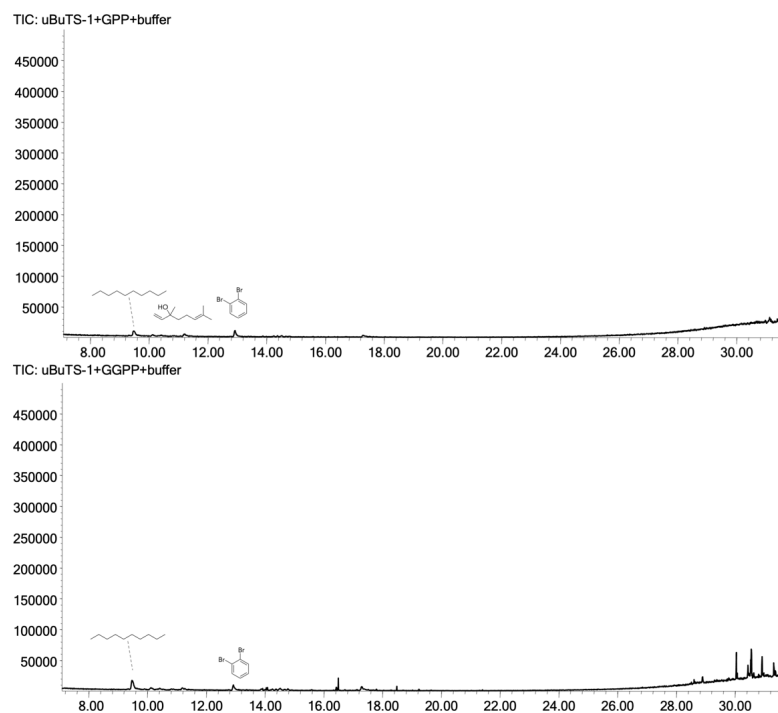


Figure S23. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBuTS-1. Annotations based on NIST spectral library comparison.

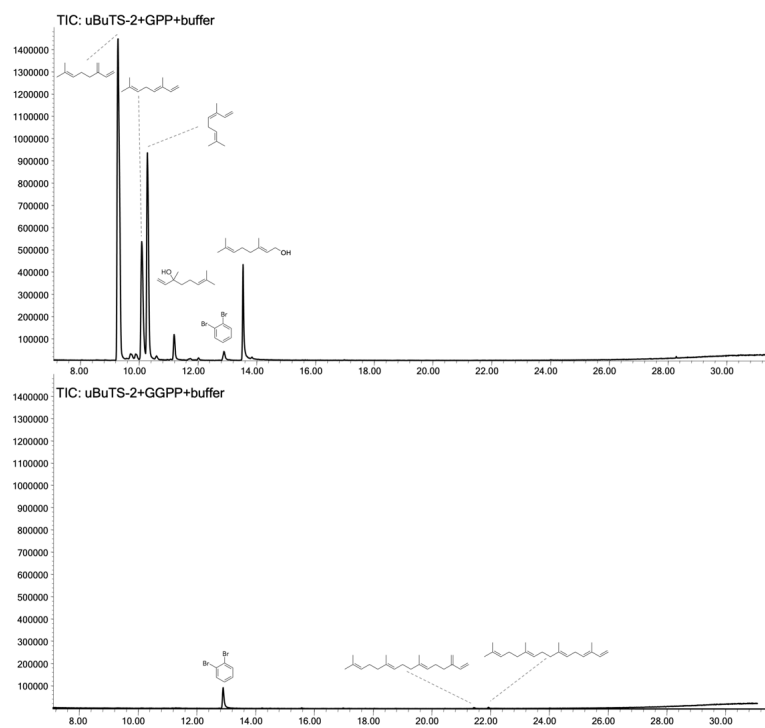


Figure S24. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBuTS-2. Annotations based on NIST spectral library comparison.

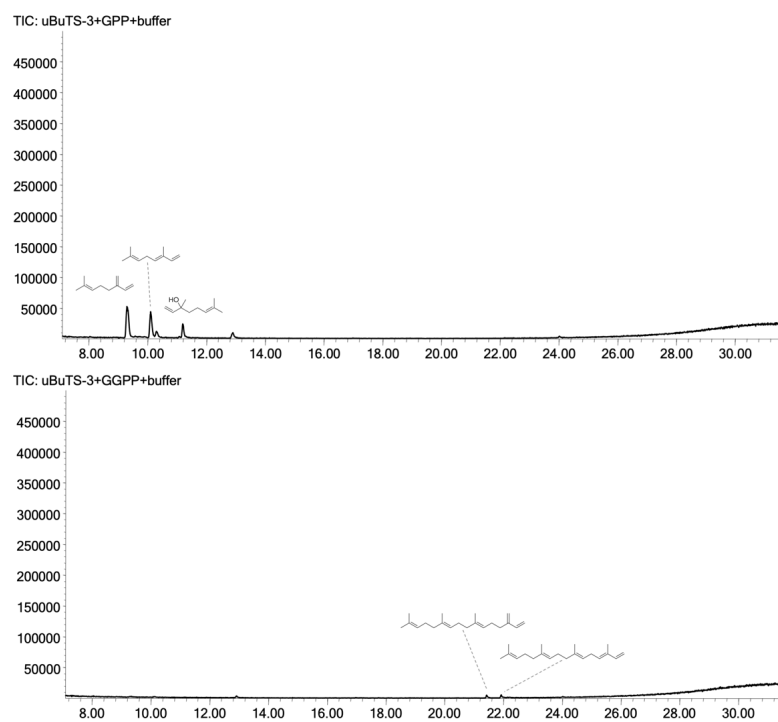


Figure S25. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBuTS-3. Annotations based on NIST spectral library comparison.

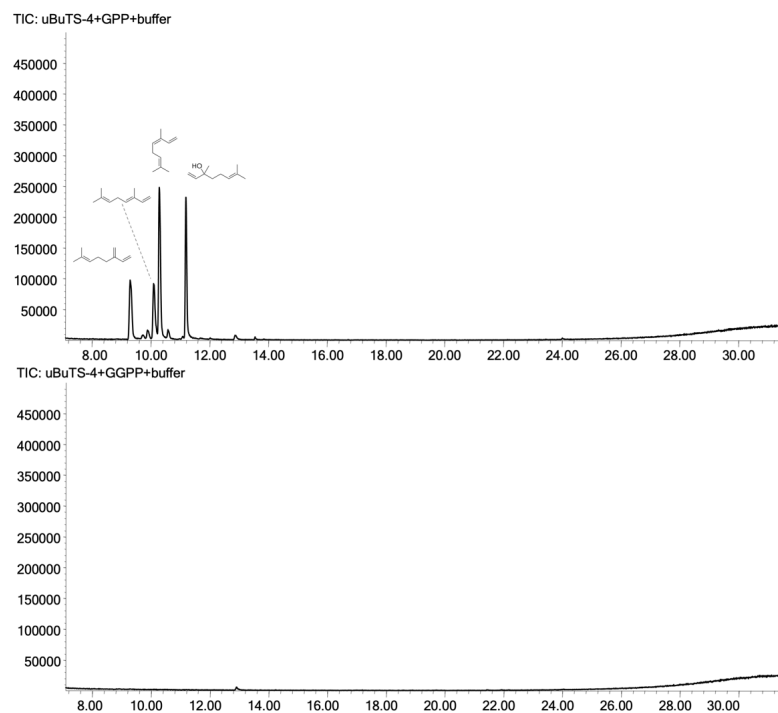


Figure S26. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBuTS-4. Annotations based on NIST spectral library comparison.

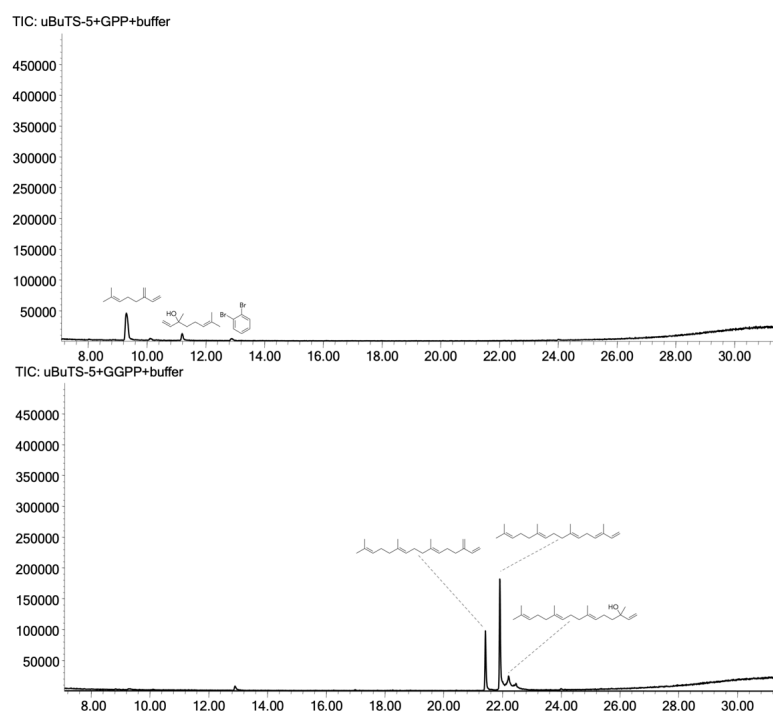


Figure S27. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBuTS-5. Annotations based on NIST spectral library comparison.

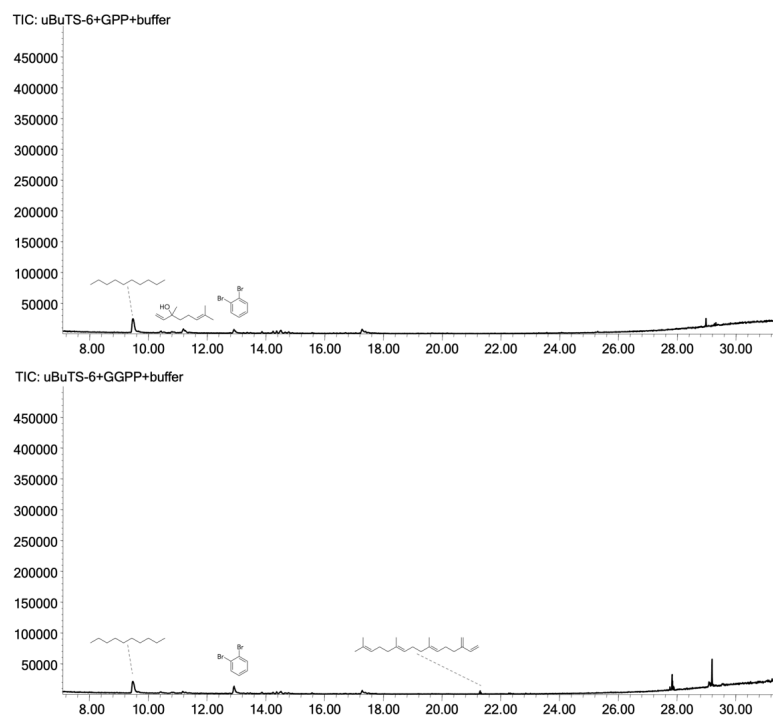


Figure S28. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBu-6. Annotations based on NIST spectral library comparison.

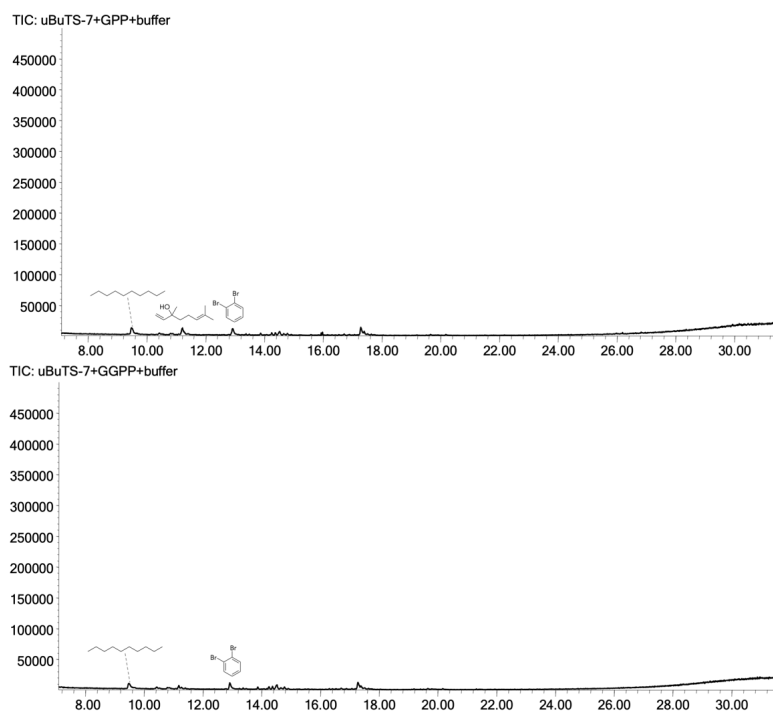


Figure S29. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBuTS-7. Annotations based on NIST spectral library comparison.

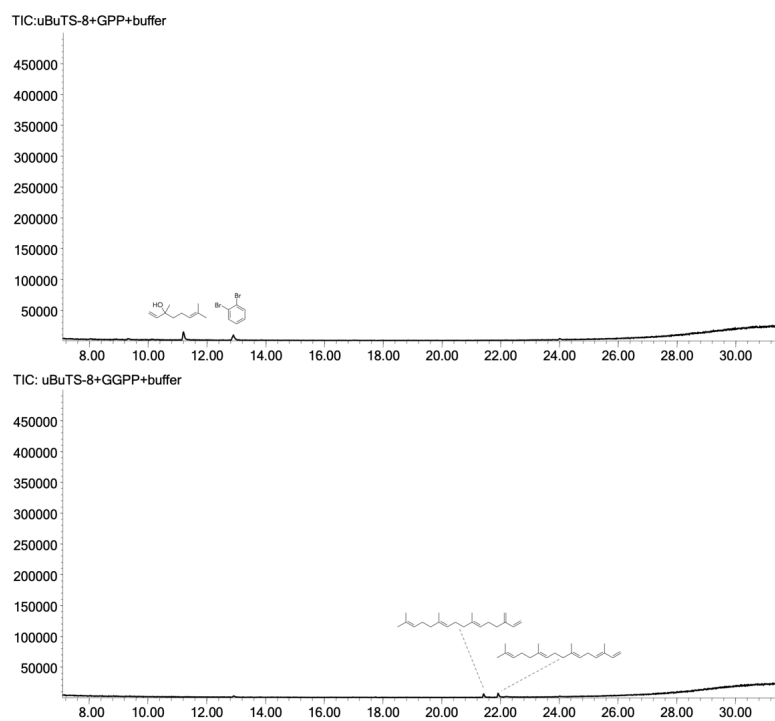


Figure S30. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBuTS-8. Annotations based on NIST spectral library comparison.

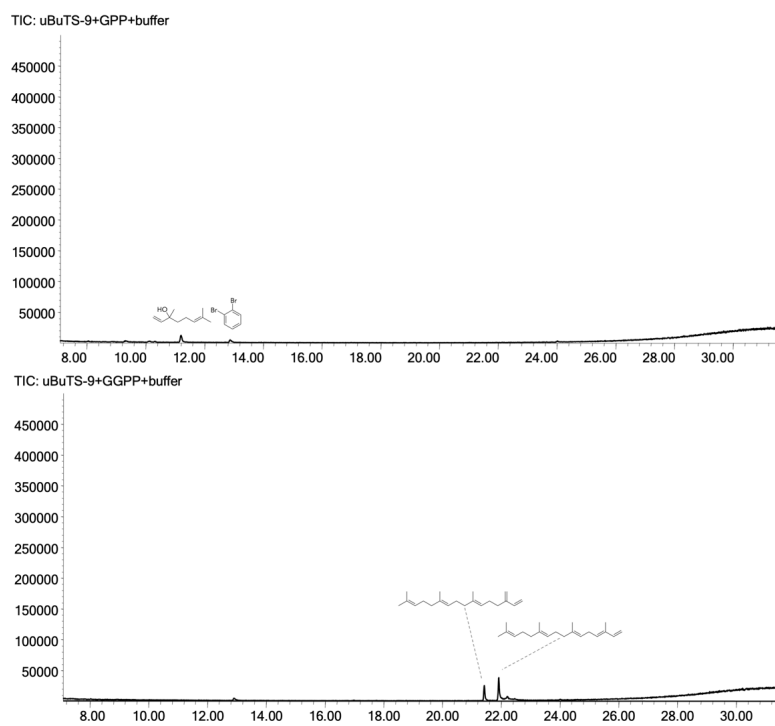


Figure S31. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with uBuTS-9. Annotations based on NIST spectral library comparison.

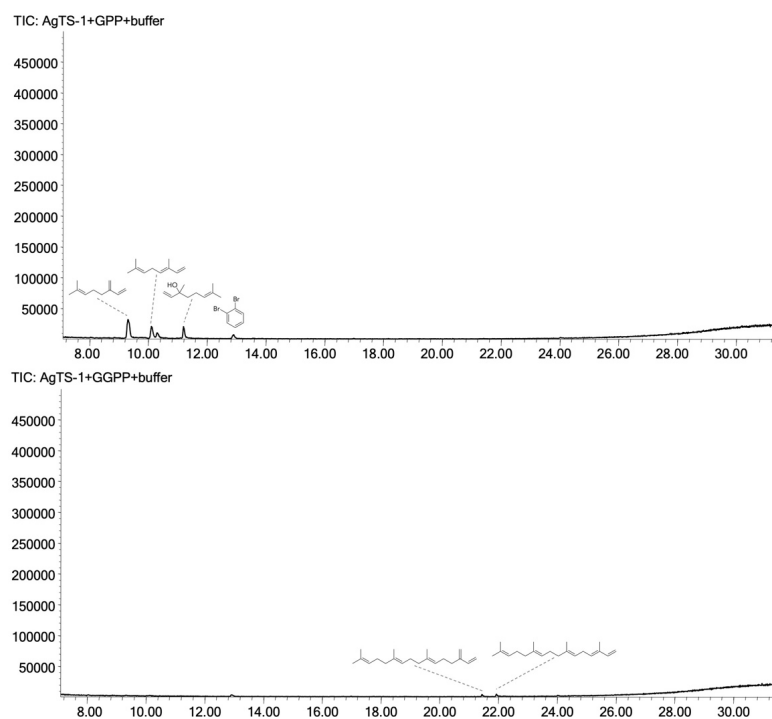


Figure S32. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with AgTS-1. Annotations based on NIST spectral library comparison.

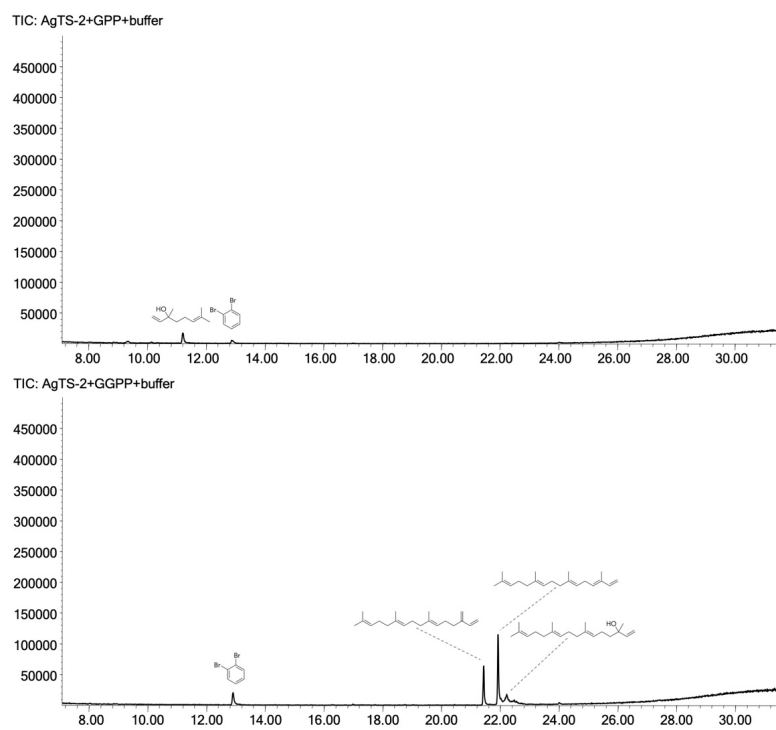


Figure S33. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with AgTS-2. Annotations based on NIST spectral library comparison.

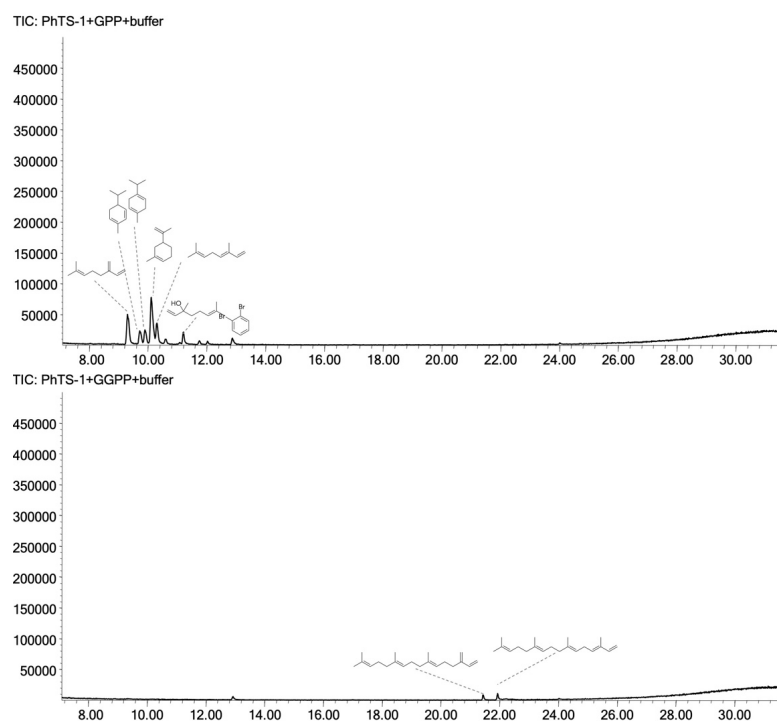


Figure S34. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (lower) assays with PhTS-1. Annotations based on NIST spectral library comparison.

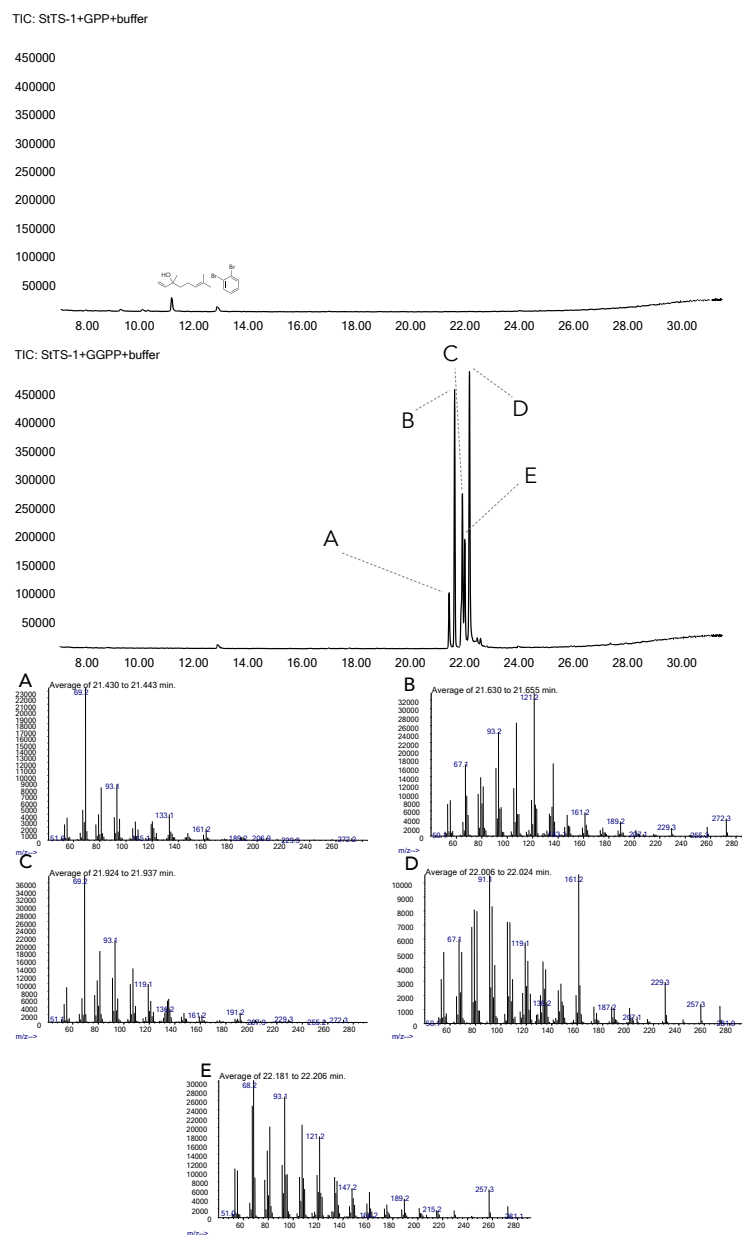


Figure S35. GCMS TIC traces of hexanes extracts of GPP (upper) and GGPP (middle) assays with StTS-1. MS of major GGPP assay peaks are shown in the lower panel. Annotations based on NIST spectral library comparison.

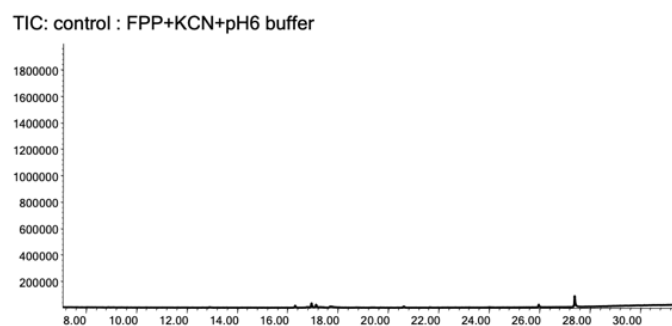
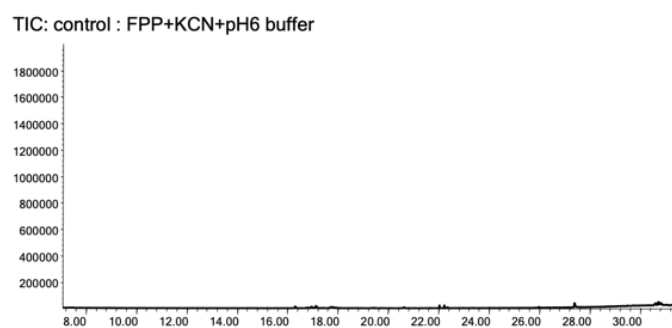
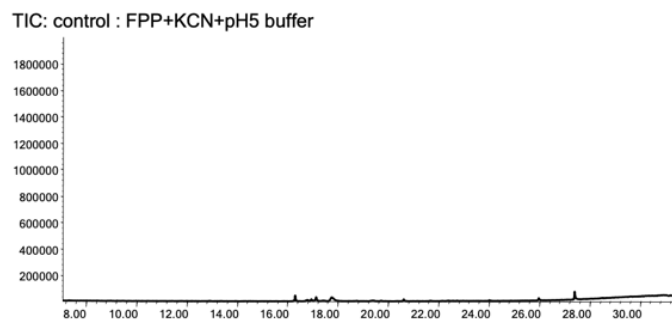


Figure S36. GCMS TIC traces of hexanes extracts of no enzyme FPP and KCN controls at pH5, 6, and 7.

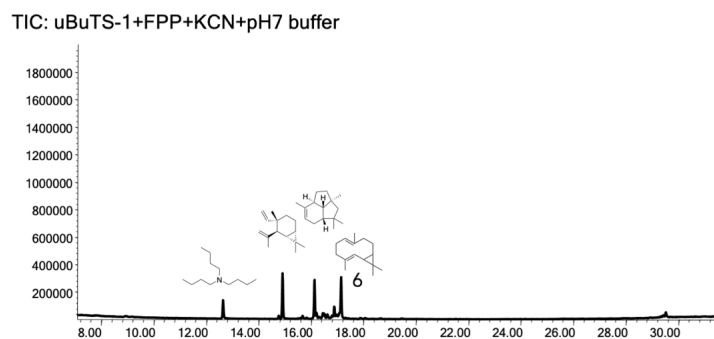
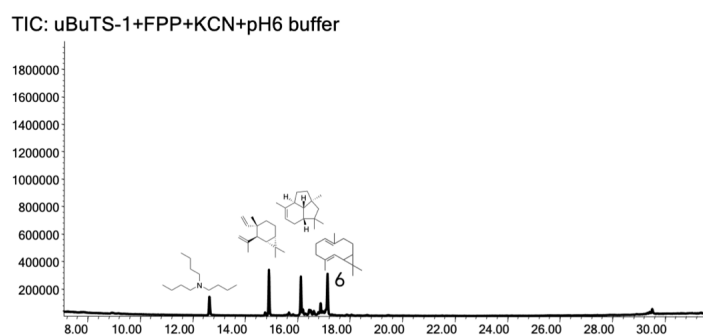
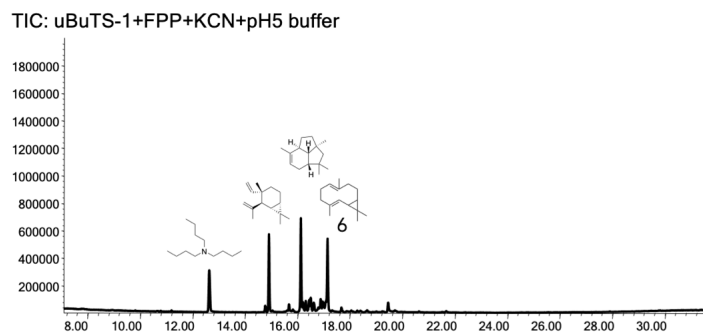


Figure S37. GCMS TIC traces of hexanes extracts of KCN and FPP assays with uBuTS-1 at pH5, 6, and 7. Annotations based on NIST spectral library comparison and comparison with bicyclogermacrene **6** standard.

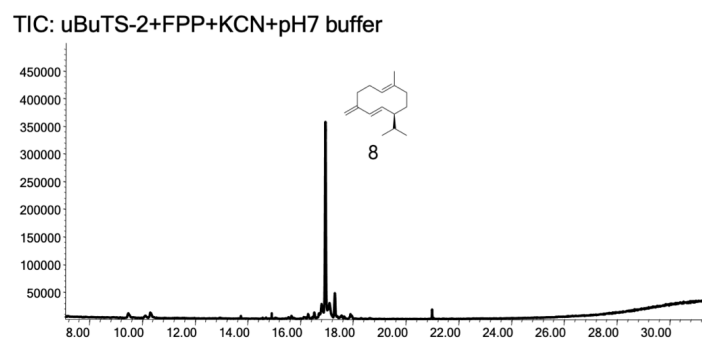
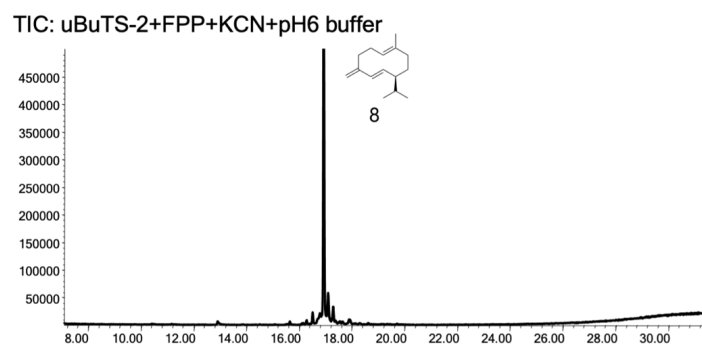
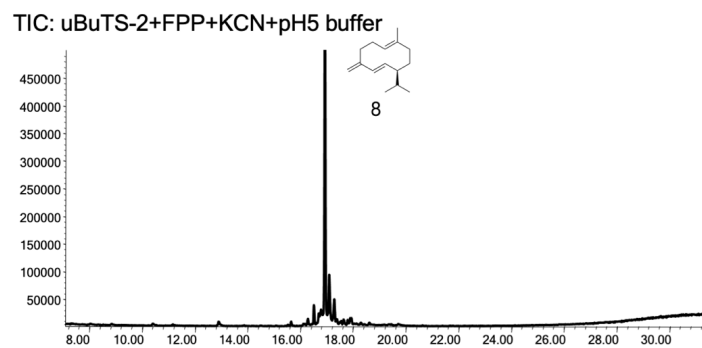
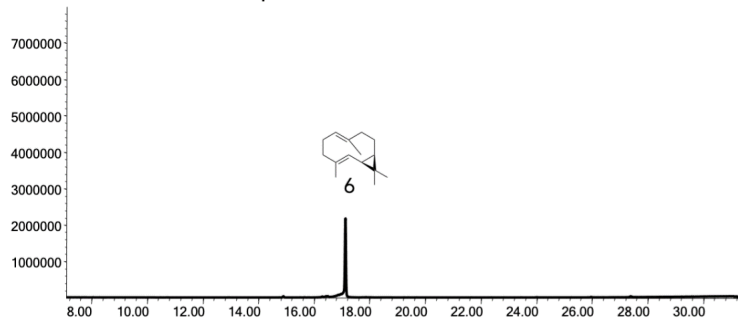
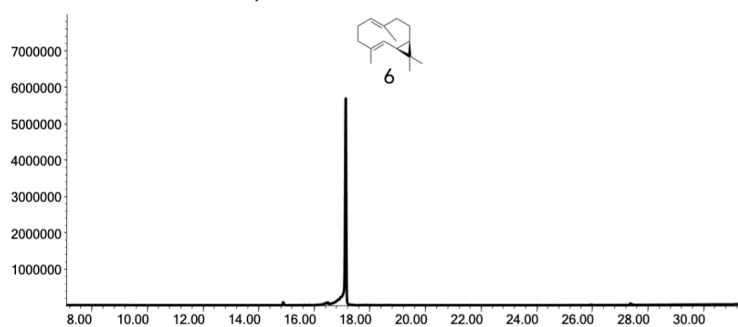


Figure S38. GCMS TIC traces of hexanes extracts of KCN and FPP assays with uBuTS-2 at pH5, 6, and 7. Annotations based on NIST spectral library comparison and comparison with germacrene D 8 standard.

TIC: uBuTS-3+FPP+KCN+pH5 buffer



TIC: uBuTS-3+FPP+KCN+pH6 buffer



TIC: uBuTS-3+FPP+KCN+pH7 buffer

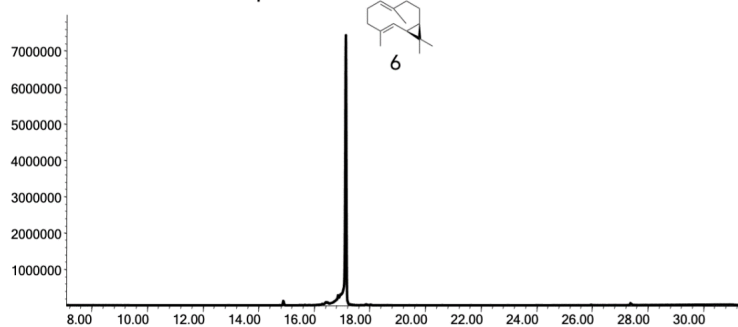


Figure S39. GCMS TIC traces of hexanes extracts of KCN and FPP assays with uBuTS-3 at pH5, 6, and 7. Annotations based on NIST spectral library comparison and comparison with bicyclogermacrene **6** standard.

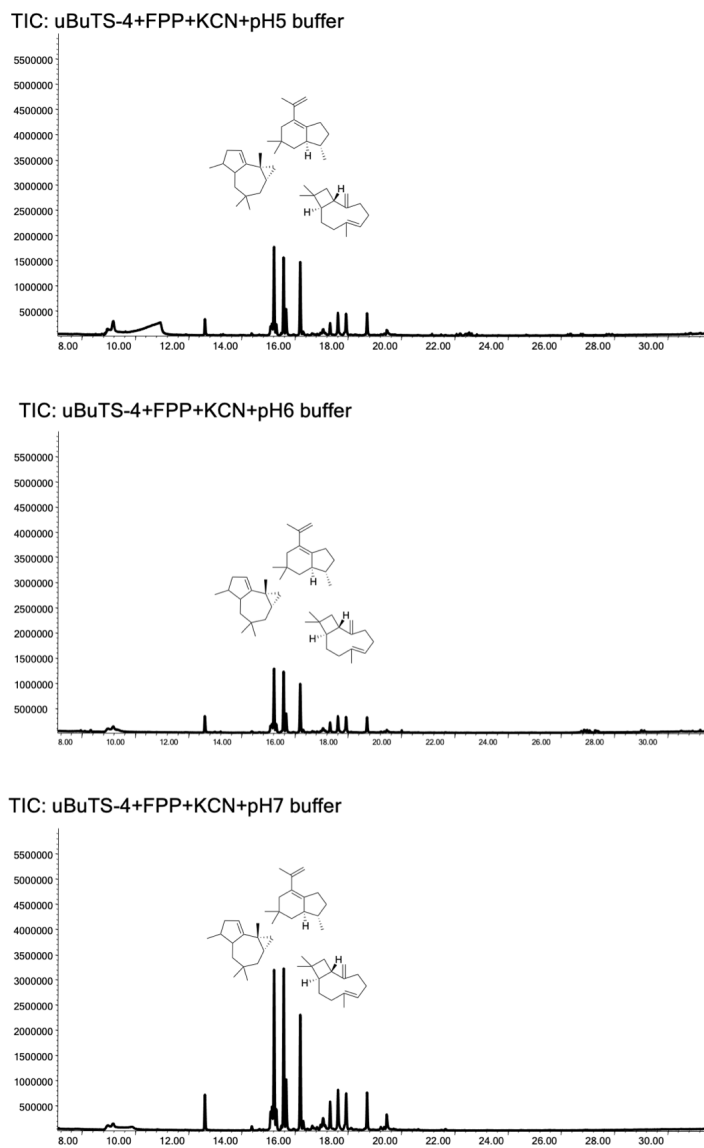
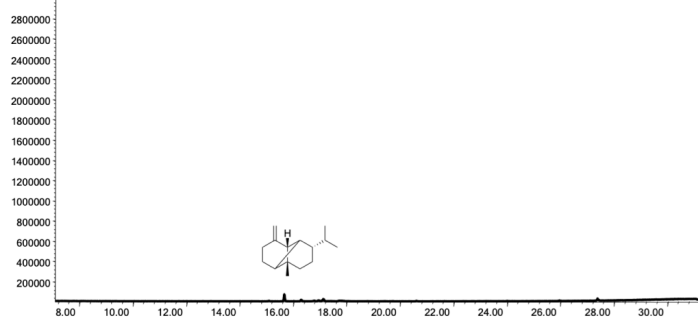
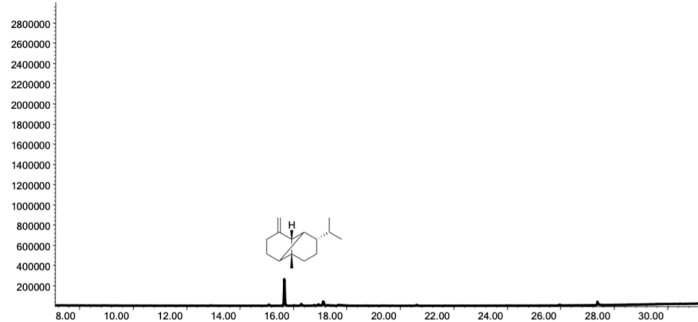


Figure S40. GCMS TIC traces of hexanes extracts of KCN and FPP assays with uBuTS-4 at pH5, 6, and 7. Annotations based on NIST spectral library comparison.

TIC: uBuTS-5+FPP+KCN+pH5 buffer



TIC: uBuTS-5+FPP+KCN+pH6 buffer



TIC: uBuTS-5+FPP+KCN+pH7 buffer

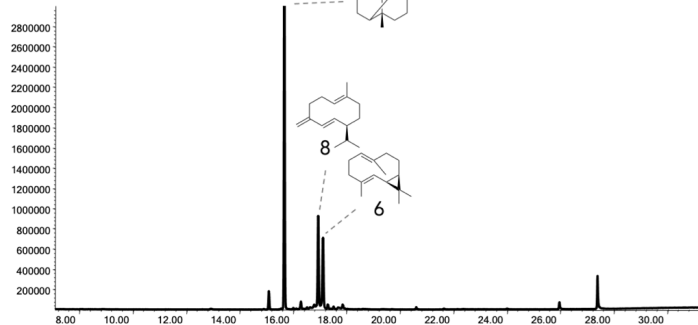


Figure S41. GCMS TIC traces of hexanes extracts of KCN and FPP assays with uBuTS-5 at pH5, 6, and 7. Annotations based on NIST spectral library comparison and comparison with bicyclo[2.2.2]oct-5-ene **6** and germacrene D **8** standards.

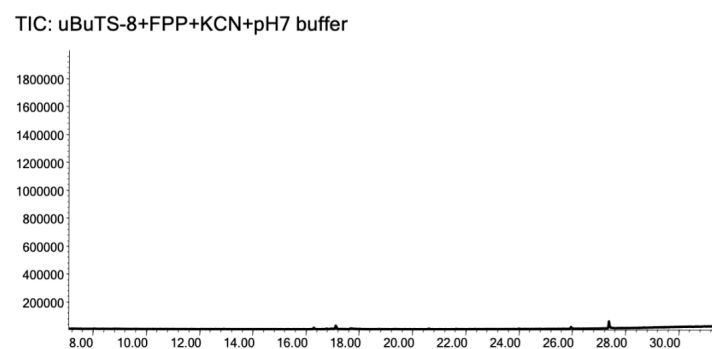
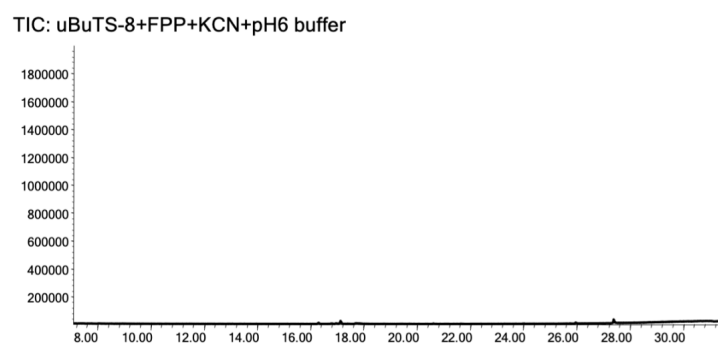
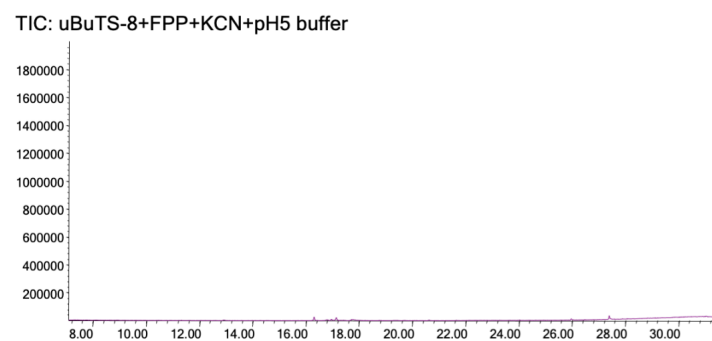


Figure S42. GCMS TIC traces of hexanes extracts of KCN and FPP assays with uBuTS-8 at pH5, 6, and 7.

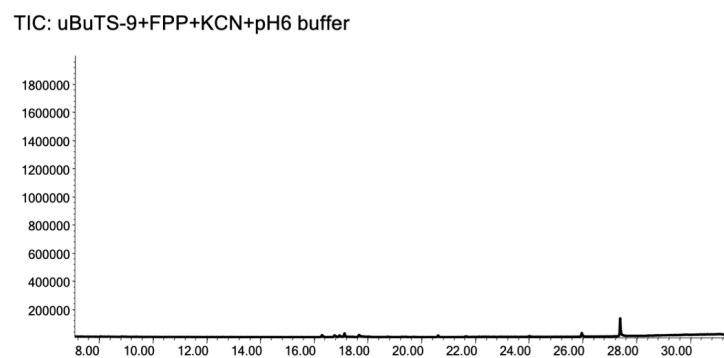
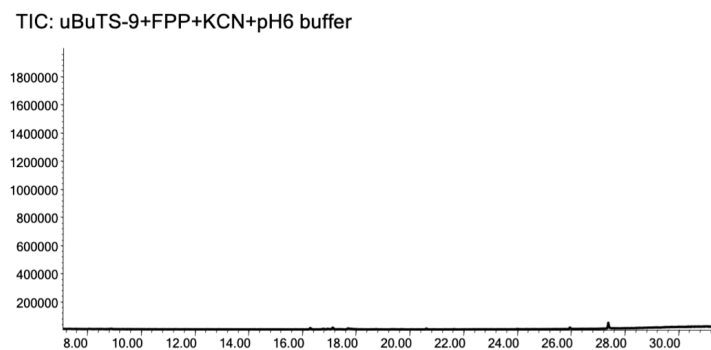
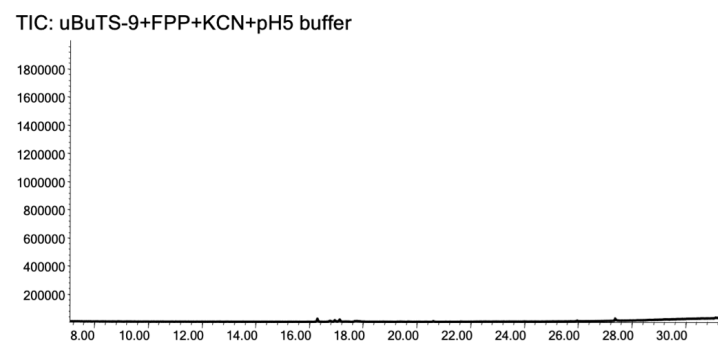


Figure S43. GCMS TIC traces of hexanes extracts of KCN and FPP assays with uBuTS-9 at pH5, 6, and 7.

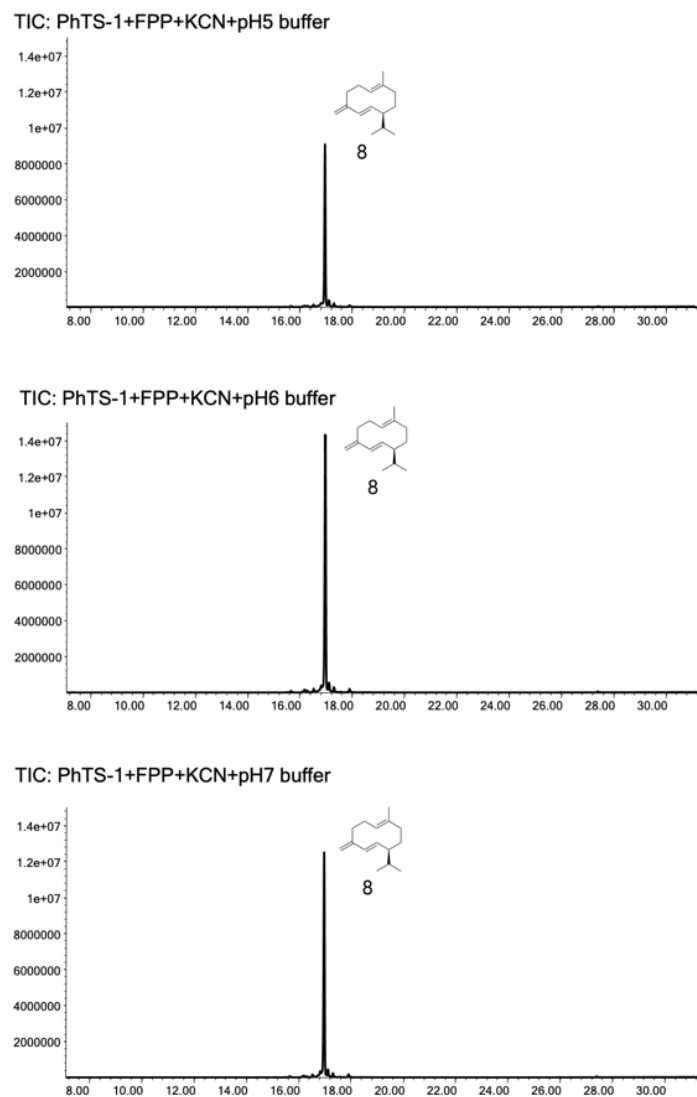


Figure S44. GCMS TIC traces of hexanes extracts of KCN and FPP assays with PhTS-1 at pH5, 6, and 7. Annotations based on NIST spectral library comparison and comparison with germacrene D 8 standard.

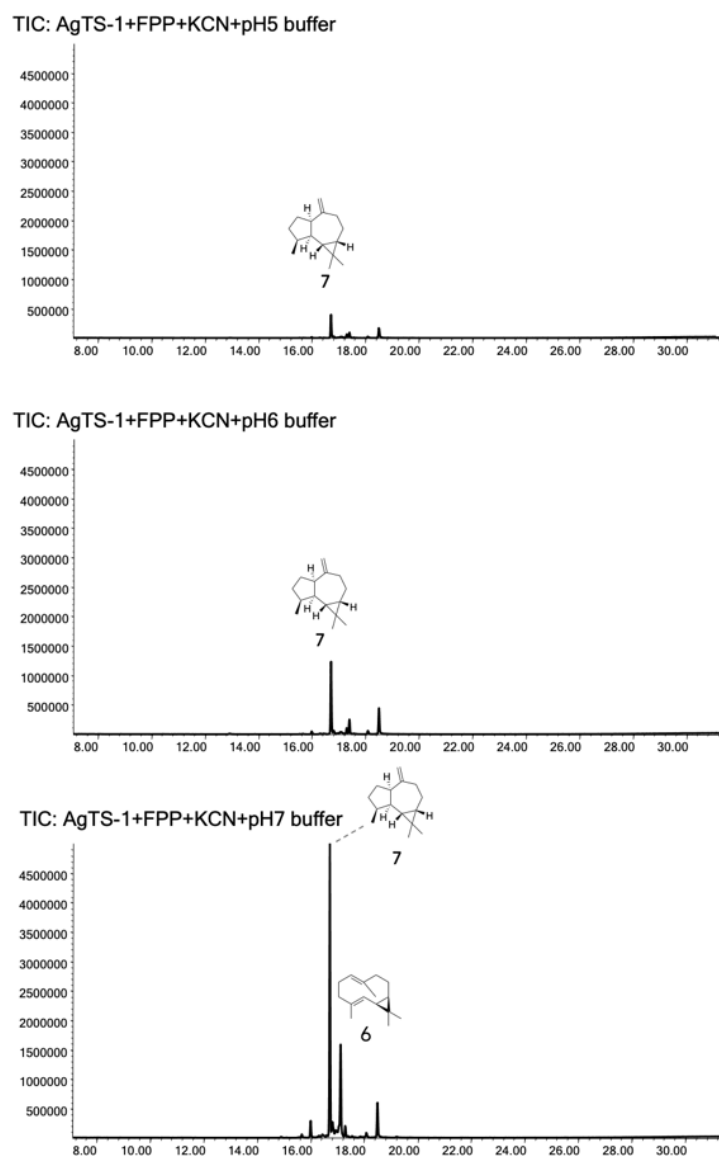


Figure S45. GCMS TIC traces of hexanes extracts of KCN and FPP assays with AgTS-1 at pH5, 6, and 7. Annotations based on NIST spectral library comparison and comparison with alloaromadendrene 7 standard.

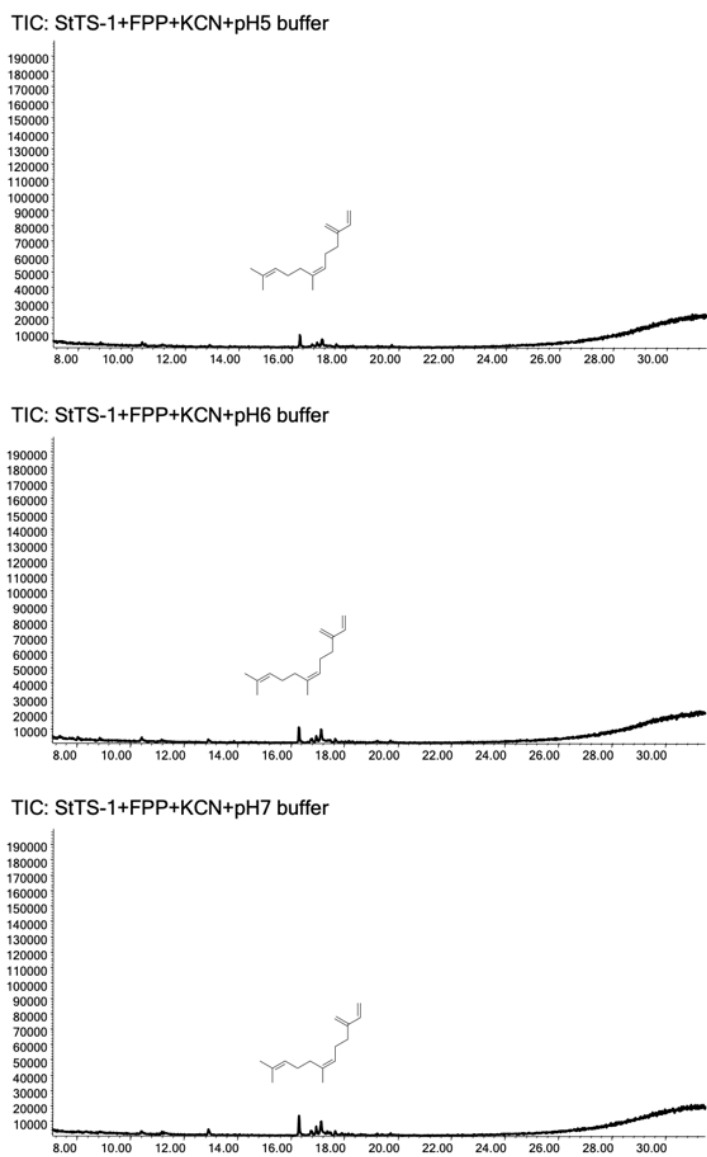


Figure S46. GCMS TIC traces of hexanes extracts of KCN and FPP assays with StTS-1 at pH5, 6, and 7. Annotations based on NIST spectral library comparison.

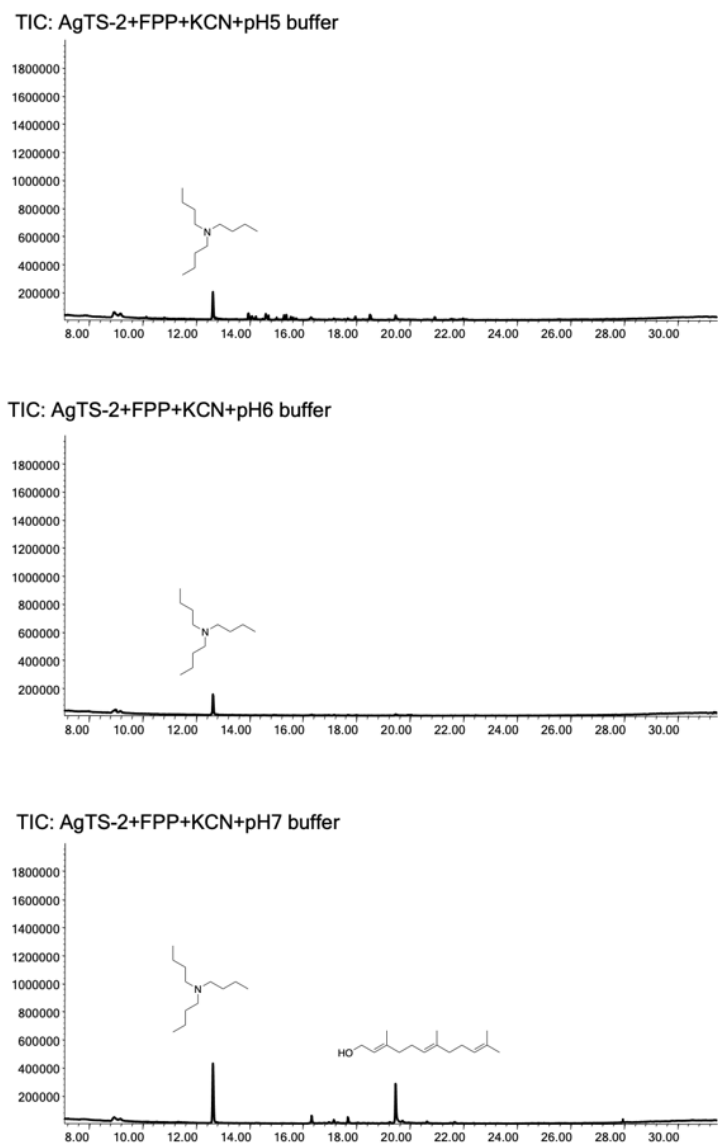


Figure S47. GCMS TIC traces of hexanes extracts of KCN and FPP assays with AgTS-2 at pH5, 6, and 7. Annotations based on NIST spectral library comparison.

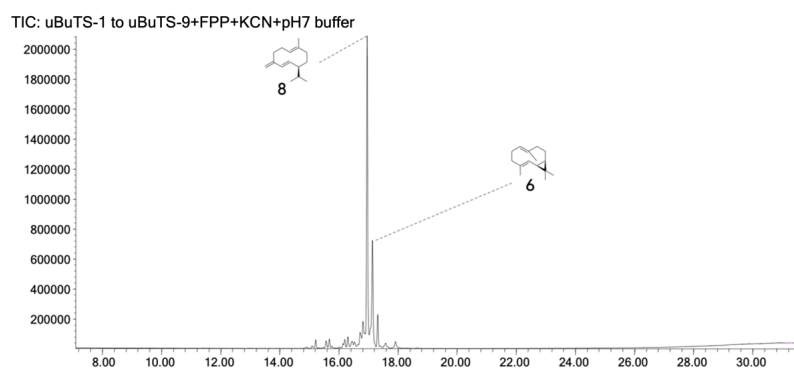


Figure S48. GCMS TIC traces of hexanes extracts of KCN and FPP assays with uBuTS-1 to uBuTS-9 at pH5, 6, and 7. Annotations based on comparison with bicyclogermacrene **6** and germacrene D **8** standards.

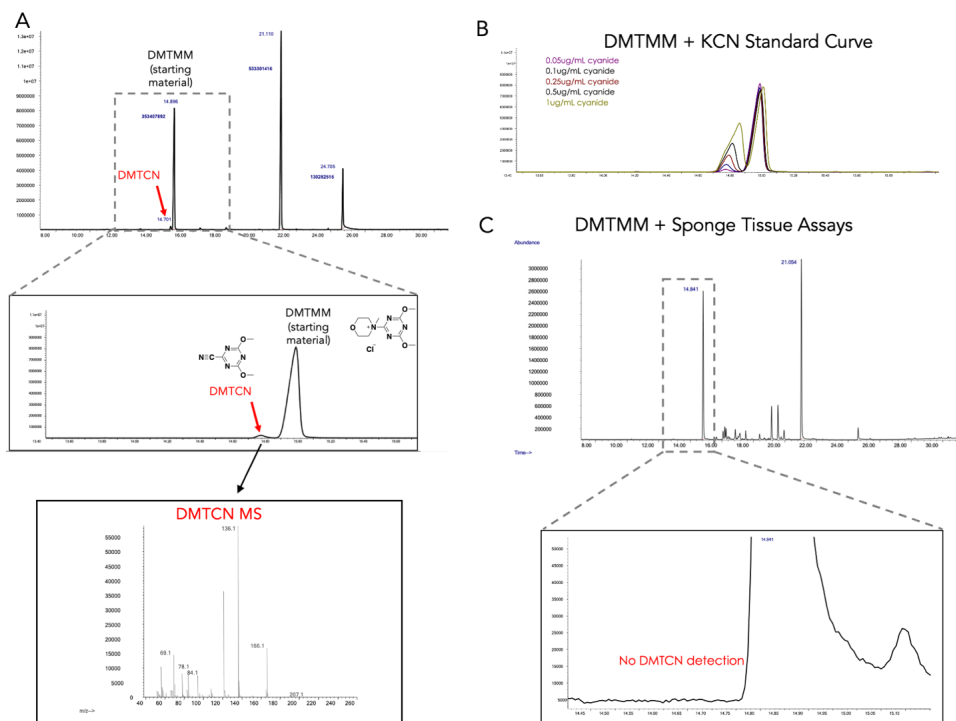
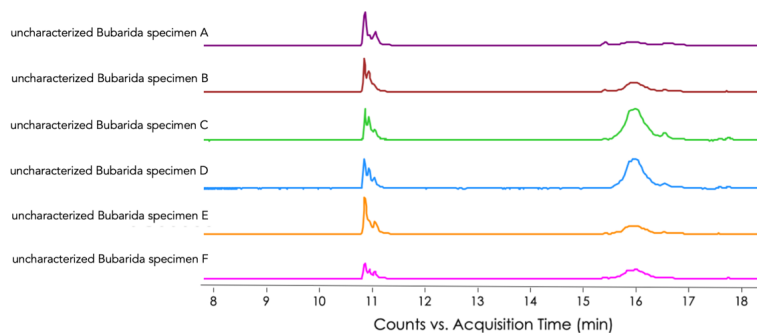


Figure S49. – A. GCMS TIC trace of ethyl acetate extract of a 900µL reaction of 80mM 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) and 0.05µg/mL potassium cyanide as described in Yamaguchi and Miyaguchi, 2020. The product, 2-cyano-4,6-dimethoxy-1,3,5- triazine (DMTCN) is shown in red. B. GCMS TIC trace of the ethyl acetate extracts of reactions with DMTMM and several concentrations of potassium cyanide, from 0.05µg/mL to 1µg/mL. C. GCMS TIC trace of the ethyl acetate extract of a 900uL reaction of 80mM DMTMM with approximately 0.2g homogenized undescribed Bubarida sponge tissue. No DMTCN is detectable.

A. Summed EIC (m/z: 222.2, 232.2, 248.2, 250.2, 264.2)



B. LCMS analysis of uncharacterized Bubarida sponge

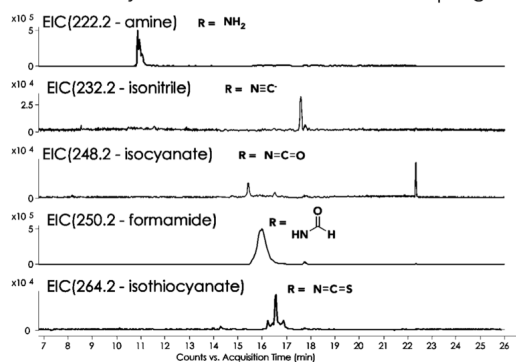


Figure S50. A. Summed LCMS EIC of m/z values representing nitrogenous sesquiterpenes (amine, isonitrile, isocyanate, formamide, and isothiocyanate) for multiple undescribed Bubarida specimen crude extracts. B. EICs of crude undescribed Bubarida extract for m/z values representing the sesquiterpene amine, isonitrile, isocyanate, formamide, and isothiocyanate.

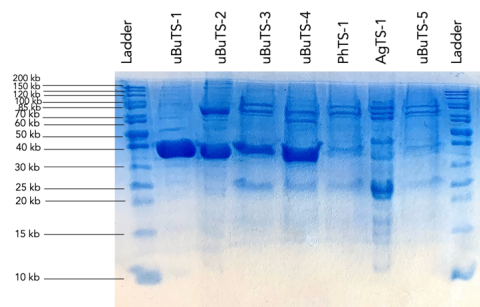


Figure S51. SDS-PAGE gel of active sesquiterpene synthases uBuTS1-5, PhTS-1, and AgTS-1. The ladder is EZ-Run™ Rec Protein Ladder (Fisher BioReagents™, cat. no. BP3602500).

Supplementary Note – Analytical data for isolated terpenes.

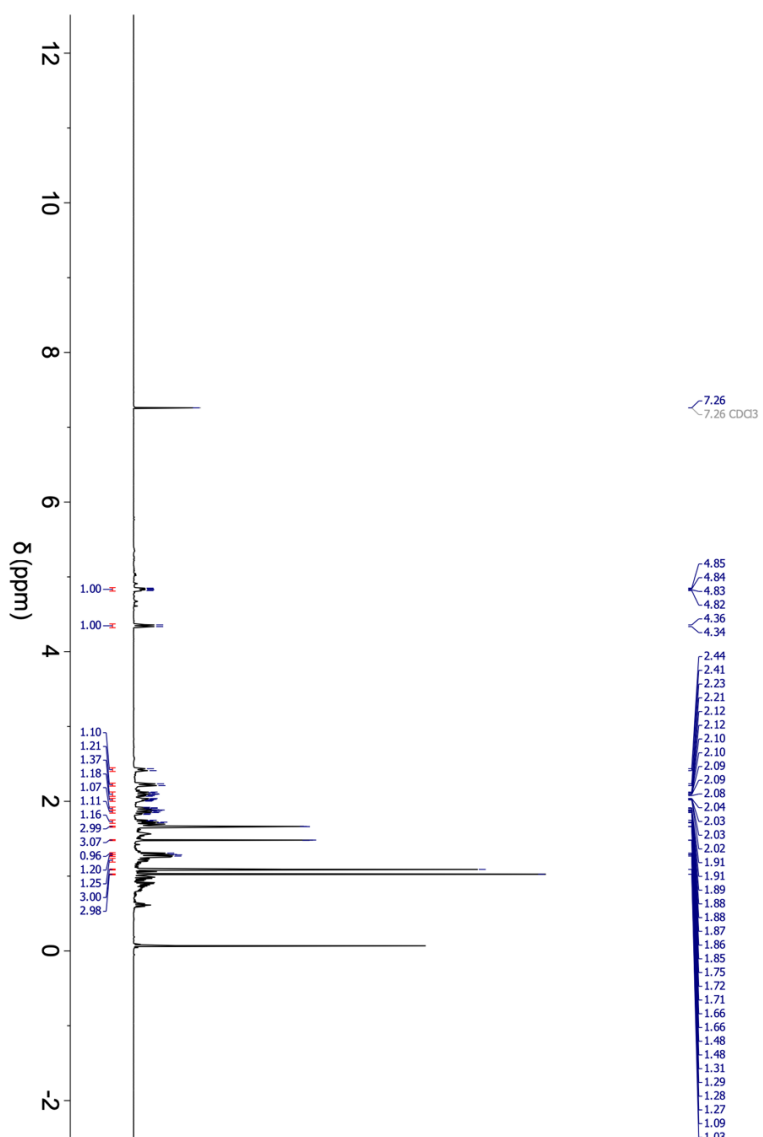
(+)-Bicyclogermacrene **6**: **¹H-NMR** (500 MHz, Chloroform-*d*): δ = 4.83 (dd, *J* = 11.2, 5.3 Hz, 1H), 4.35 (d, *J* = 11.5 Hz, 1H), 2.42 (d, *J* = 13.2 Hz, 1H), 2.22 (d, *J* = 12.0 Hz, 1H), 2.10 (ddd, *J* = 12.5, 11.1, 3.8 Hz, 1H), 2.04 – 2.00 (m, 1H), 1.92 – 1.88 (m, 1H), 1.86 (dd, *J* = 10.5, 3.4 Hz, 1H), 1.73 (dd, *J* = 13.0, 4.1 Hz, 1H), 1.66 (d, *J* = 1.4 Hz, 3H), 1.48 (d, *J* = 1.6 Hz, 3H), 1.30 (d, *J* = 8.8 Hz, 1H), 1.28 (d, *J* = 8.7 Hz, 1H), 1.22 (ddd, *J* = 12.5, 4.8, 1.4 Hz, 1H), 1.09 (s, 3H), 1.03 (s, 3H). **¹³C-NMR** (125 MHz, Chloroform-*d*): δ = 141.00, 128.12, 126.65, 124.97, 41.31, 37.39, 30.21, 29.37, 27.09, 26.16, 21.02, 20.01, 16.71, 15.58 (29). **Optical rotation**: $[\alpha]_D^{26.0} = +16$ (c 0.82, CHCl₃) (30, 31). **EI-MS** (70eV): *m/z* (%) = 204 (15), 189 (8), 162 (5), 161 (23), 148 (4), 147 (6), 136 (17), 135 (5), 134 (5), 133 (9), 123 (9), 122 (15), 121 (100), 120 (6), 119 (26), 117 (3), 109 (8), 108 (11), 107 (50), 106 (6), 105 (29), 95 (9), 94 (20), 93 (65), 92 (9), 91 (37), 82 (3), 81 (20), 80 (8), 79 (31), 78 (5), 77 (22), 69 (6), 68 (3), 67 (17), 65 (7), 55 (12). See NMR spectra below.

(+)-Alloaromadendrene **7**: **¹H-NMR** (500 MHz, Chloroform-*d*): δ = 4.72 (d, *J* = 13.0 Hz, 2H), 2.67 (q, *J* = 8.1 Hz, 1H), 2.34 – 2.27 (m, 2H), 2.07 – 2.04 (m, 1H), 1.88 – 1.82 (m, 3H), 1.72 (d, *J* = 4.2 Hz, 2H), 1.31 (d, *J* = 1.8 Hz, 1H), 1.23 – 1.21 (m, 1H), 1.00 (s, 3H), 0.95 (s, 3H), 0.94 (d, *J* = 6.9 Hz, 3H), 0.55 (ddd, *J* = 10.9, 9.2, 6.2 Hz, 1H), 0.24 (dd, *J* = 10.9, 9.3 Hz, 1H). **¹³C-NMR** (125 MHz, Chloroform-*d*): δ = 152.79, 109.79, 50.91, 42.29, 37.92, 35.84, 31.33, 29.85, 28.76, 28.35, 24.91, 23.63, 22.26, 16.55, 16.01 (32). **Optical rotation**: $[\alpha]_D^{26.0} = +8.58$ (c 0.167, CHCl₃) (33). **EI-MS** (70eV): *m/z* (%) = 205 (7), 204 (45), 190 (5), 189 (32), 176 (4), 175 (9), 162 (16), 161 (96), 149 (10), 148 (37), 147 (56), 146 (3), 145 (5), 136 (4), 135 (20), 134 (23), 133 (73), 132 (3), 131 (8), 129 (4), 128 (4), 123 (6), 122 (30), 121 (38), 120 (27), 119 (65), 118 (4), 117 (14), 116 (4), 115 (9), 109 (15), 108 (19), 107 (75), 106 (28), 105 (90), 104 (4), 103 (7), 96 (6), 95 (27), 94 (21), 93 (77), 92 (24), 91 (100), 83 (7), 82 (17), 81 (47), 80 (12), 79 (64), 78 (13), 77 (40), 69 (42), 68 (5), 67 (39), 66 (5), 65 (15), 55 (30), 53 (18). See NMR spectra below.

(-)-Germacrene D **8**: **¹H-NMR** (500 MHz, Chloroform-*d*): δ = 5.78 (d, *J* = 15.8 Hz, 1H), 5.25 (dd, *J* = 15.8, 10.0 Hz, 1H), 5.16 – 5.08 (m, 1H), 4.81 – 4.72 (m, 2H), 2.44 – 2.34 (m, 2H), 2.24 (dd, *J* = 12.0, 3.2 Hz, 1H), 2.12 – 2.07 (m, 1H), 1.99 (ddd, *J* = 15.3, 7.1, 3.4 Hz, 2H), 1.51 (s, 3H), 1.50 – 1.39 (m, 4H), 0.86 (d, *J* = 6.7 Hz, 3H), 0.81 (d, *J* = 6.8 Hz, 3H). **¹³C-NMR** (125 MHz, Chloroform-*d*): δ = 149.02, 135.63, 134.16, 133.70, 129.79, 109.19, 53.04, 40.84, 34.63, 32.88, 29.38, 26.64, 20.91, 19.47, 16.06 (34). **Optical rotation**: $[\alpha]_D^{26.0} = -159$ (c 0.51, CHCl₃) (35, 36). **EI-MS** (70eV): *m/z* (%) = 204 (15), 162 (15), 161 (100), 160 (4), 147 (6), 134 (3), 133 (18), 131 (4), 121 (5), 120 (23), 119 (33), 117 (8), 115 (4), 107 (11), 106 (8), 105 (52), 103 (3), 95 (5), 93 (22), 92 (9), 91 (45), 81 (29), 80 (7), 79 (27), 78 (7), 77 (22), 69 (8), 67 (11), 65 (7), 55 (11), 53 (8). See NMR spectra below.

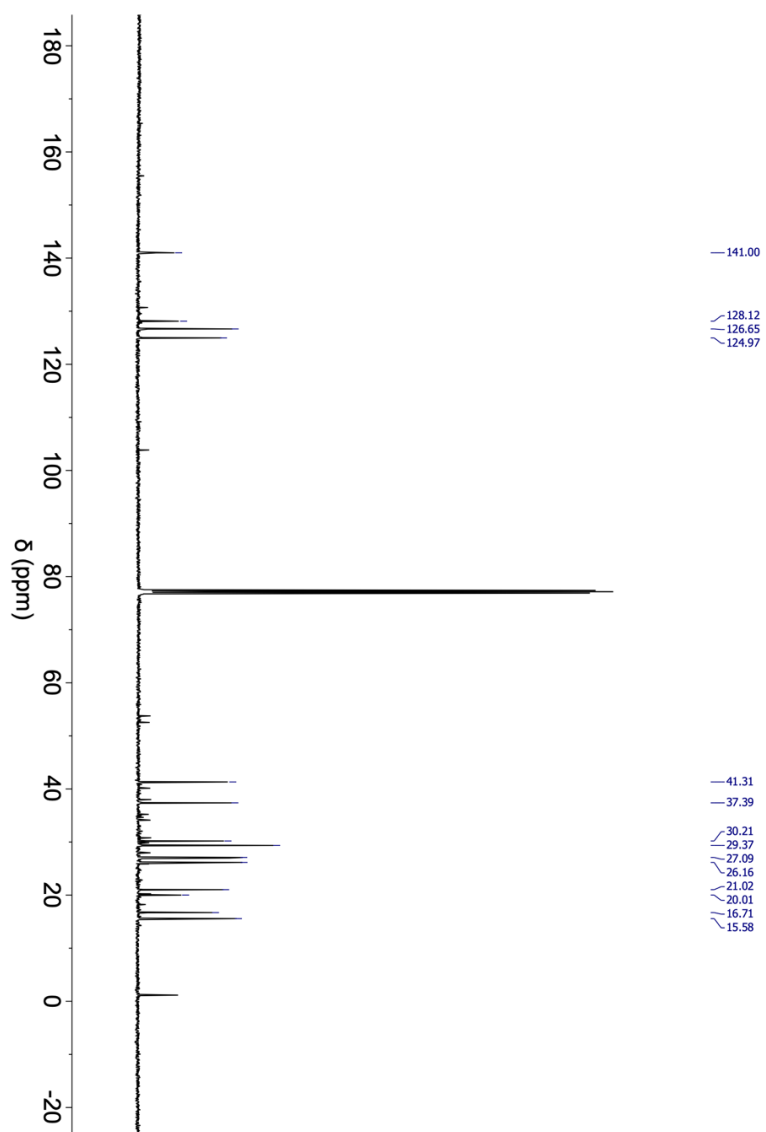


(+)-Bicyclogermacrene ^1H



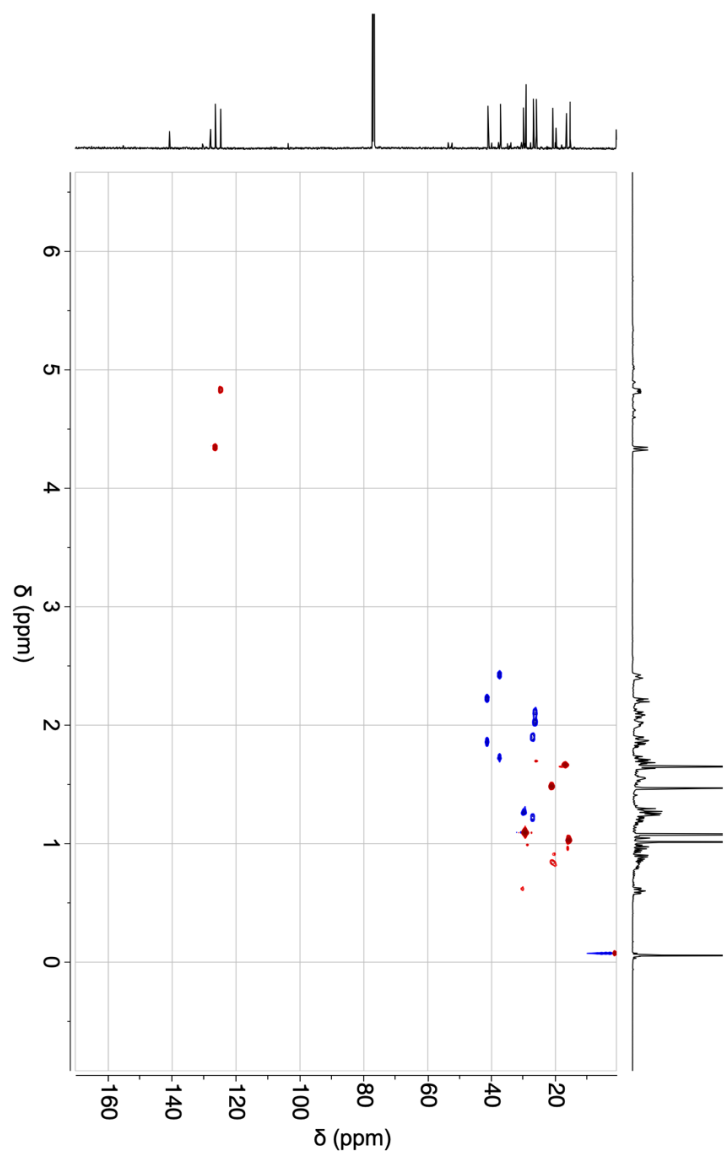


(+)-Bicyclogermacrene ^{13}C



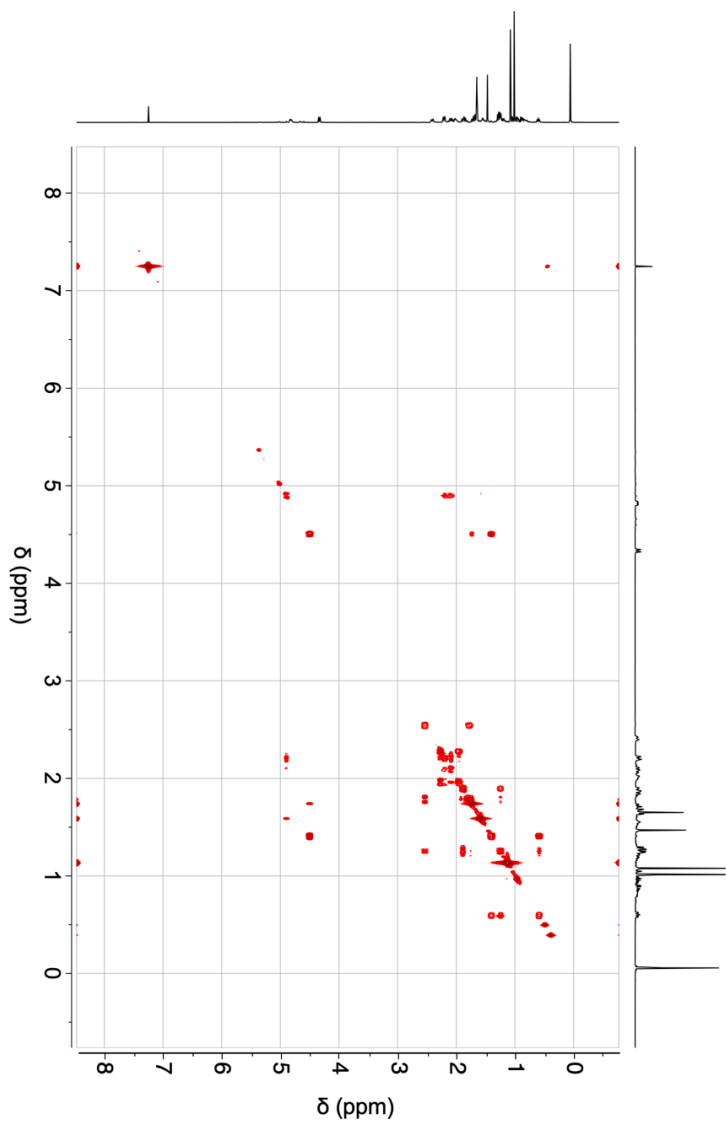


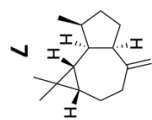
(+)-Bicyclogermacrene HSQC



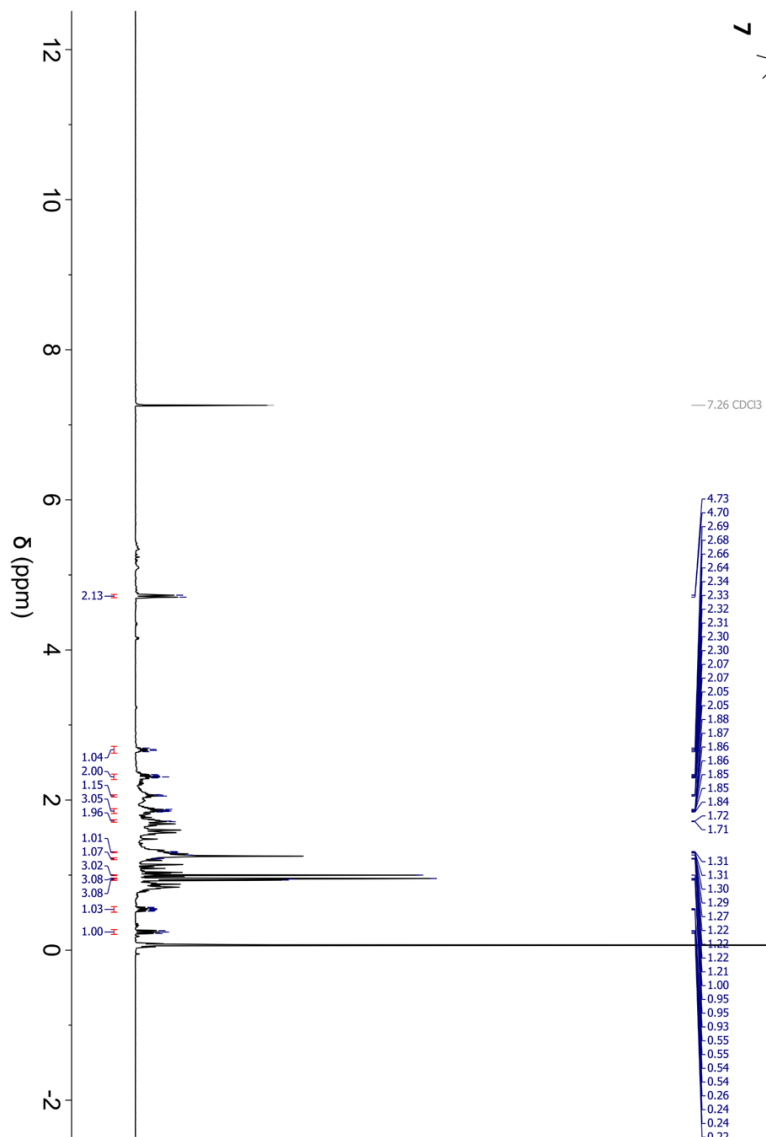


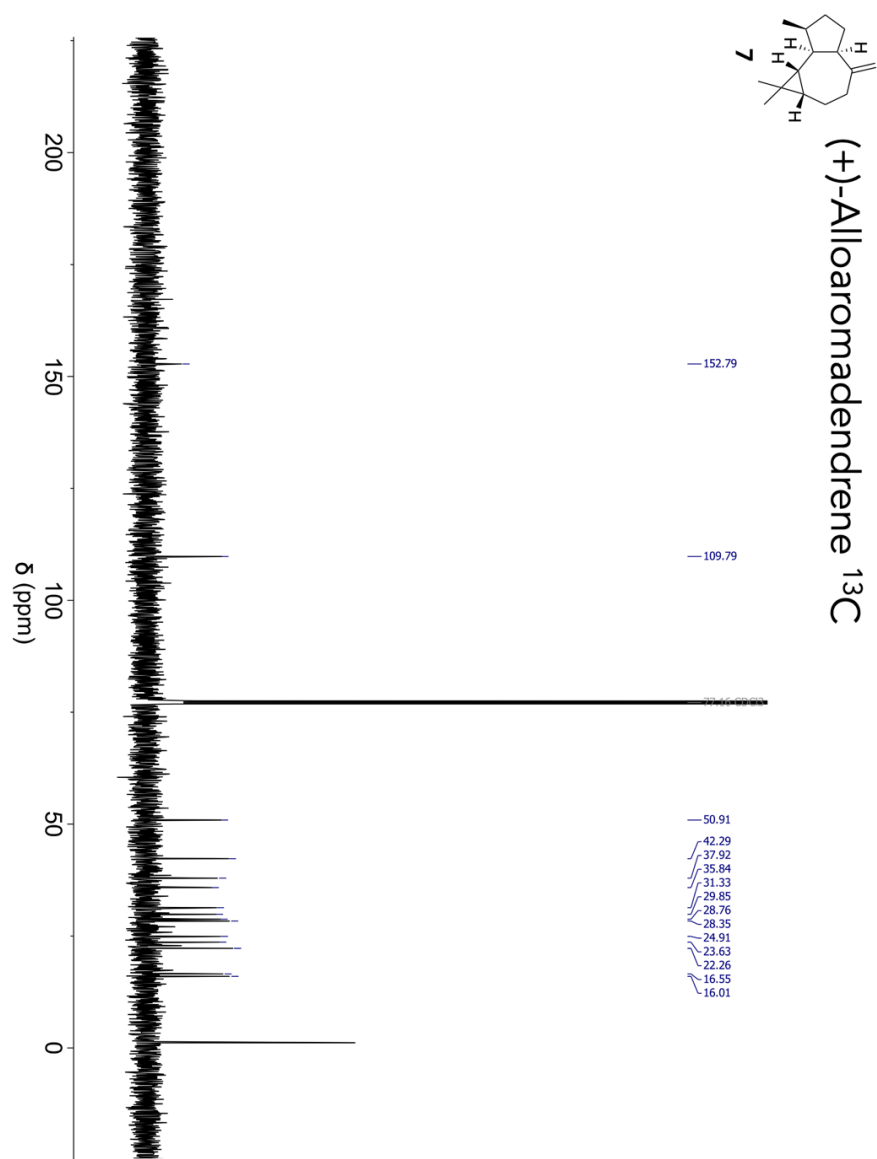
(+)-Bicyclogermacrene COSY

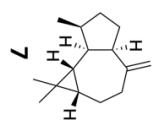




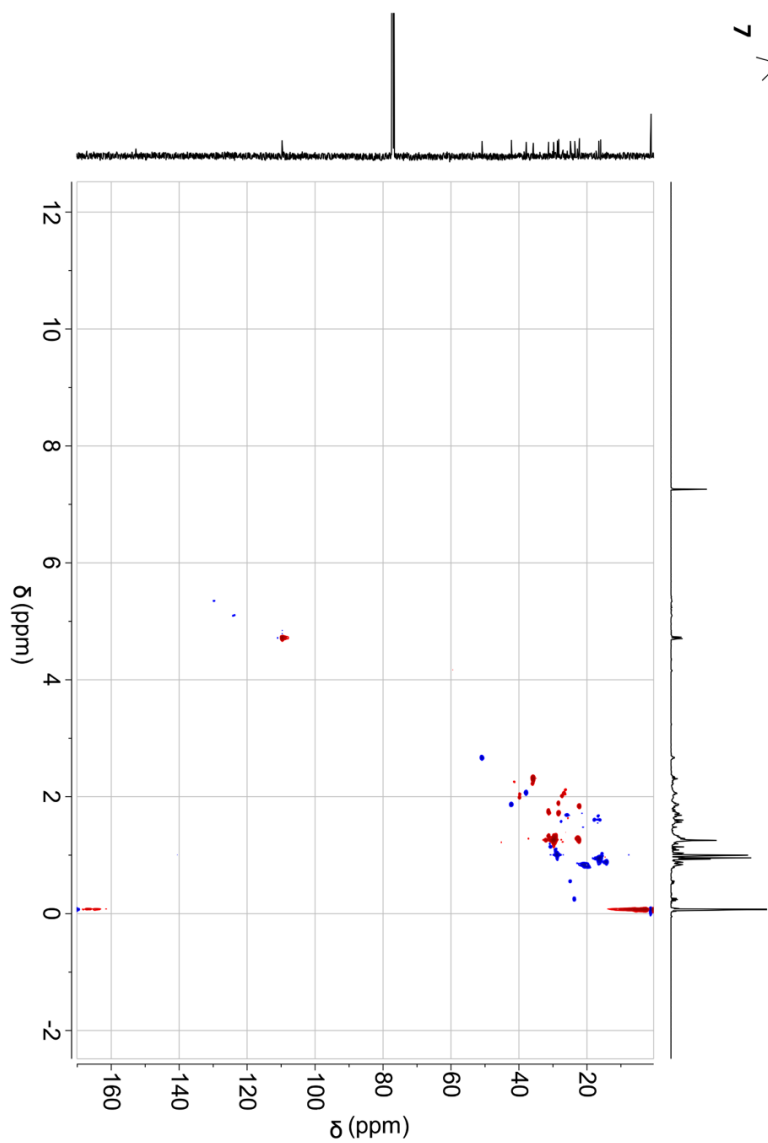
(+)-Alloaromadendrene ^1H

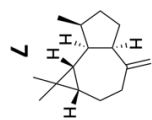




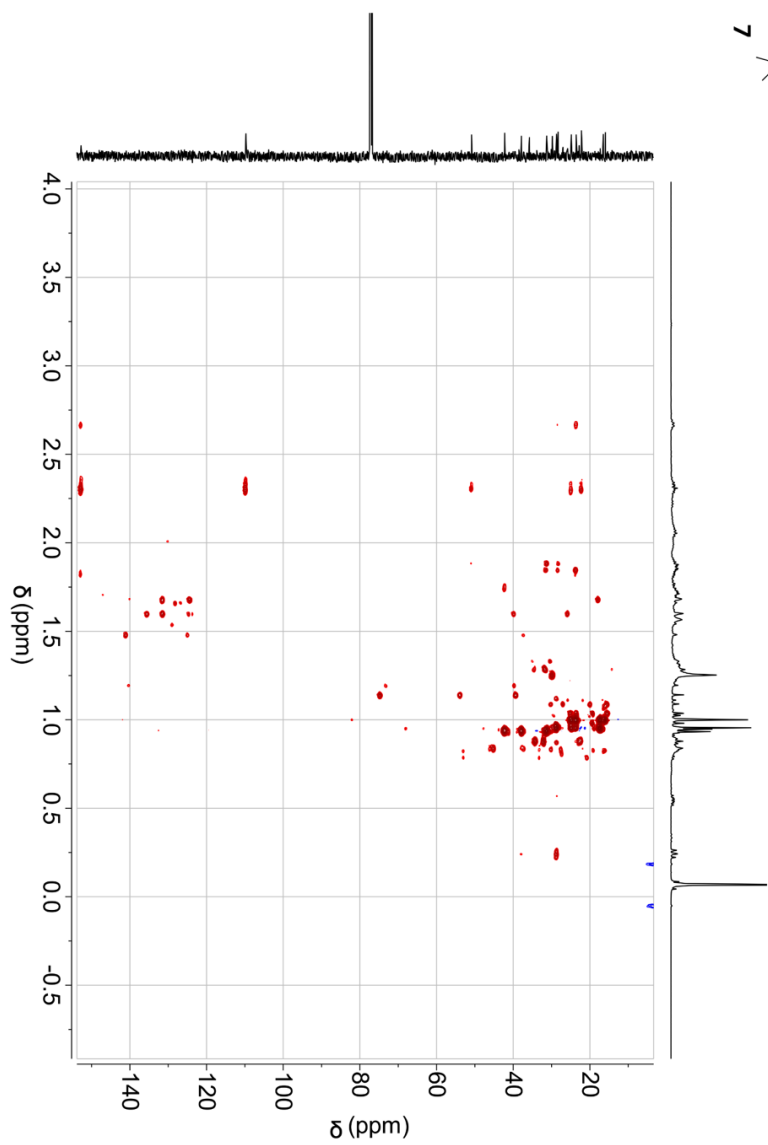


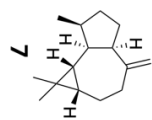
(+)-Alloaromadendrene HSQC



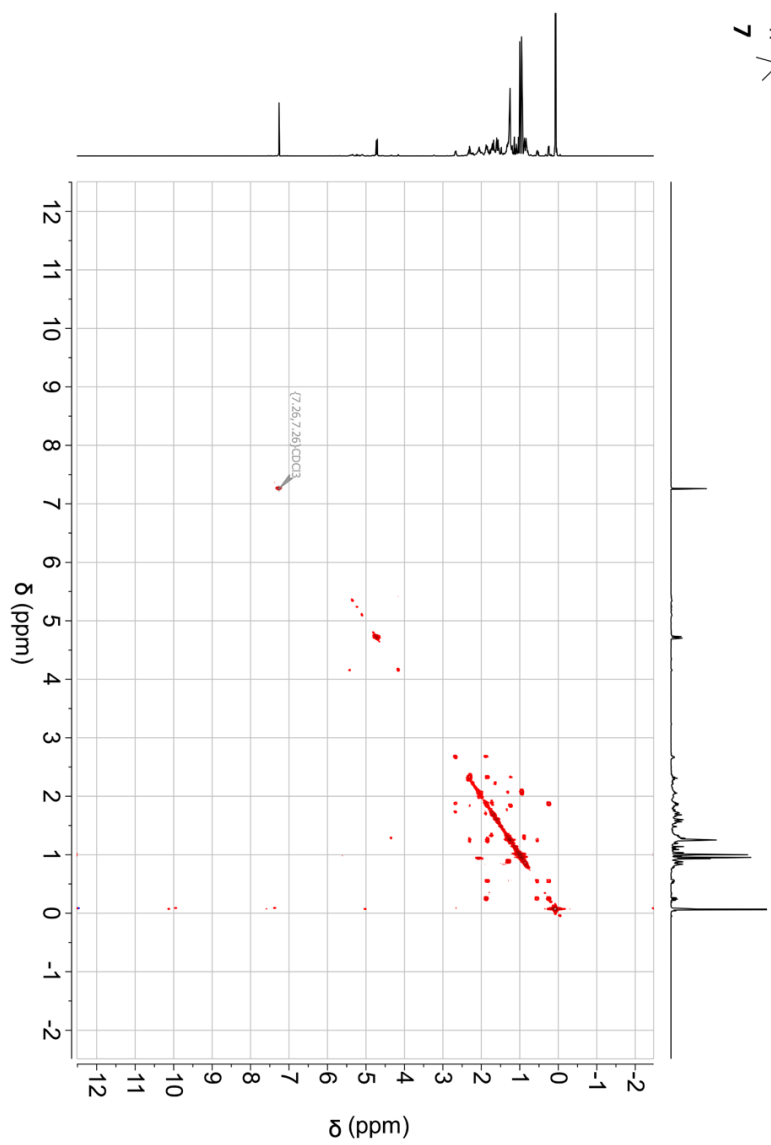


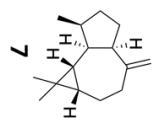
(+)-Alloaromadendrene HMBC



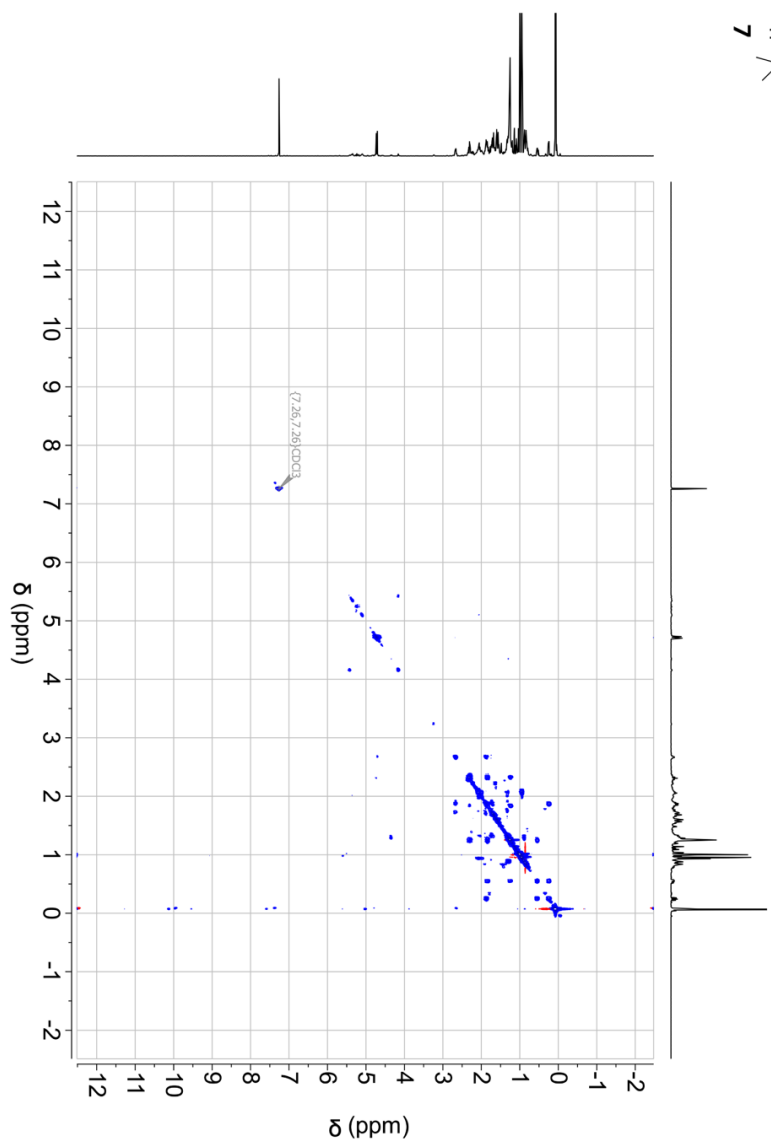


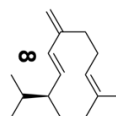
(+)-Alloaromadendrene COSY





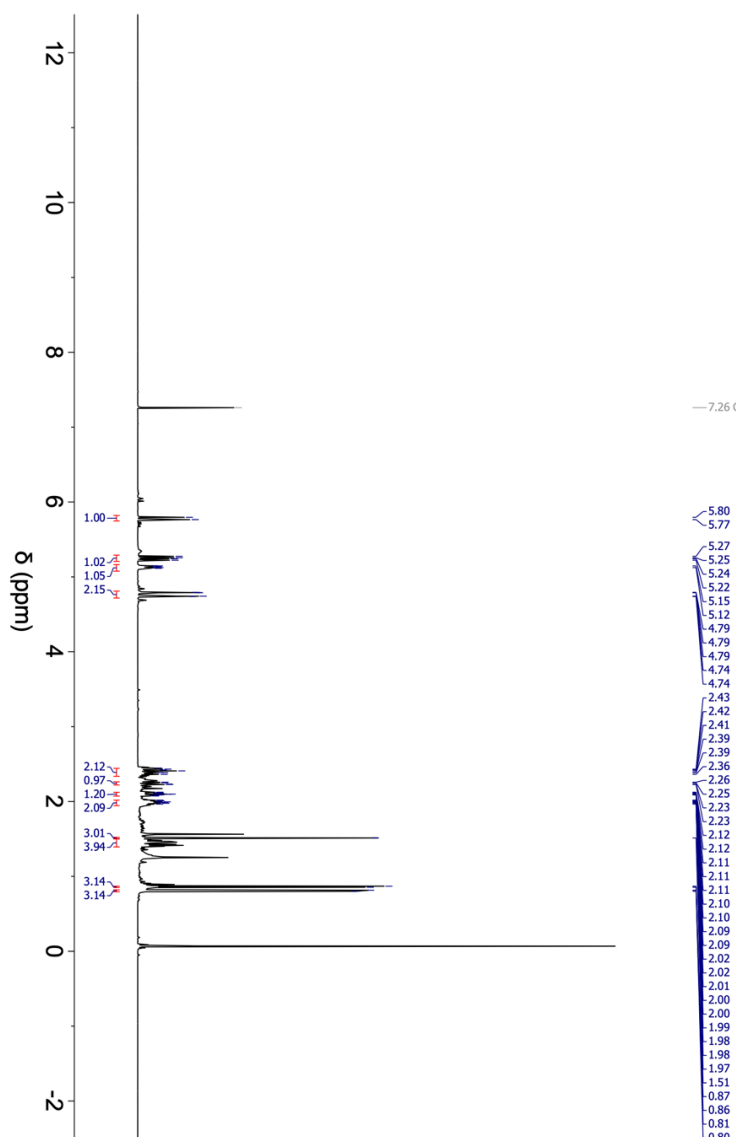
(+)-Alloaromadendrene NOESY

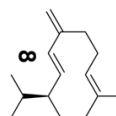




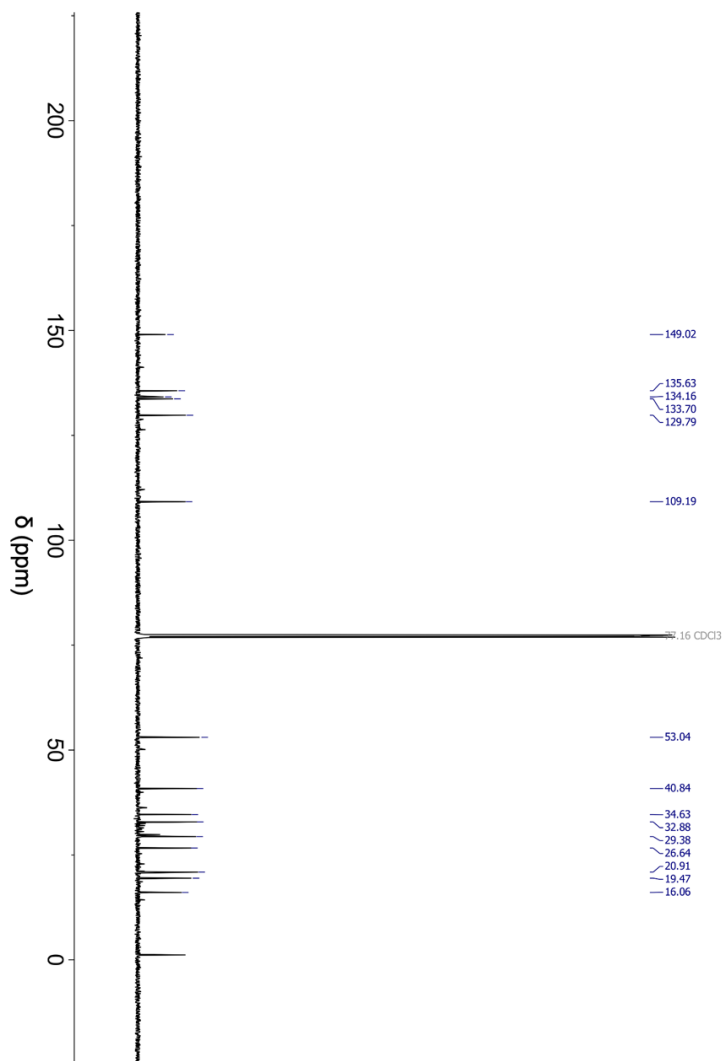
(-)-Germacrene D ^1H

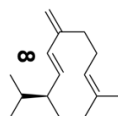
— 7.26 CDCl₃



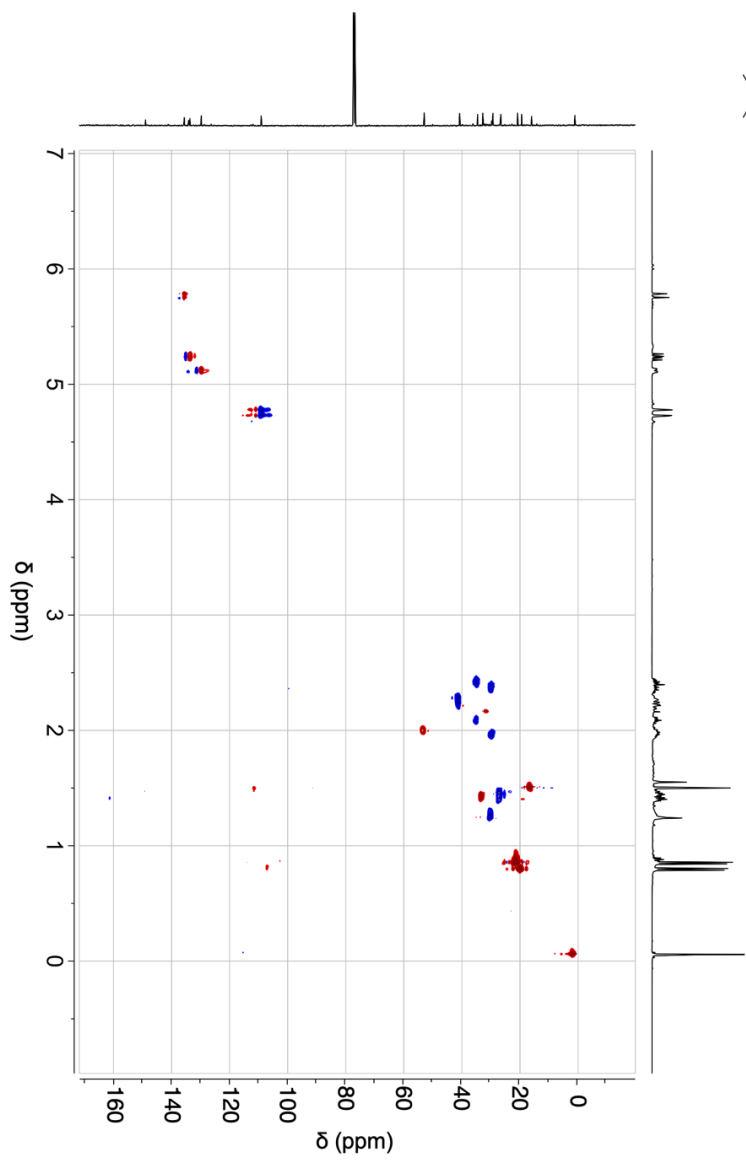


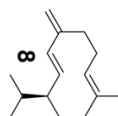
(-)-Germacrene D ^{13}C



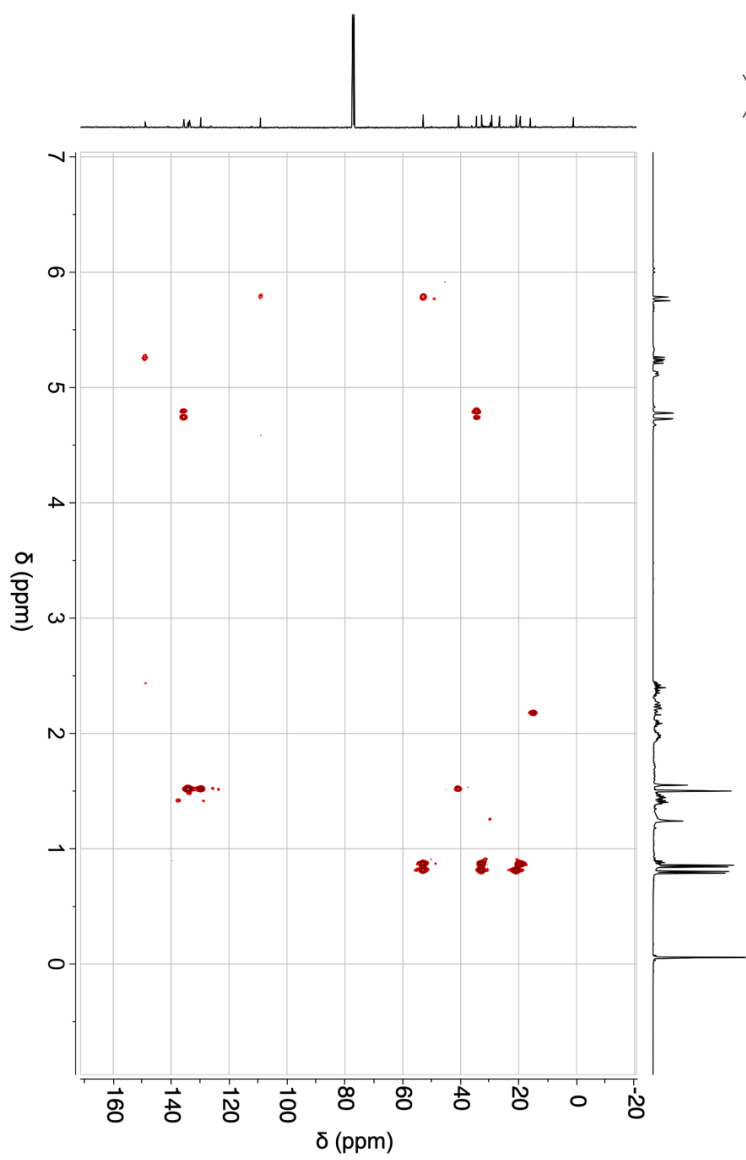


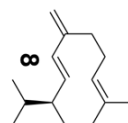
(-)-Germacrene D HSQC



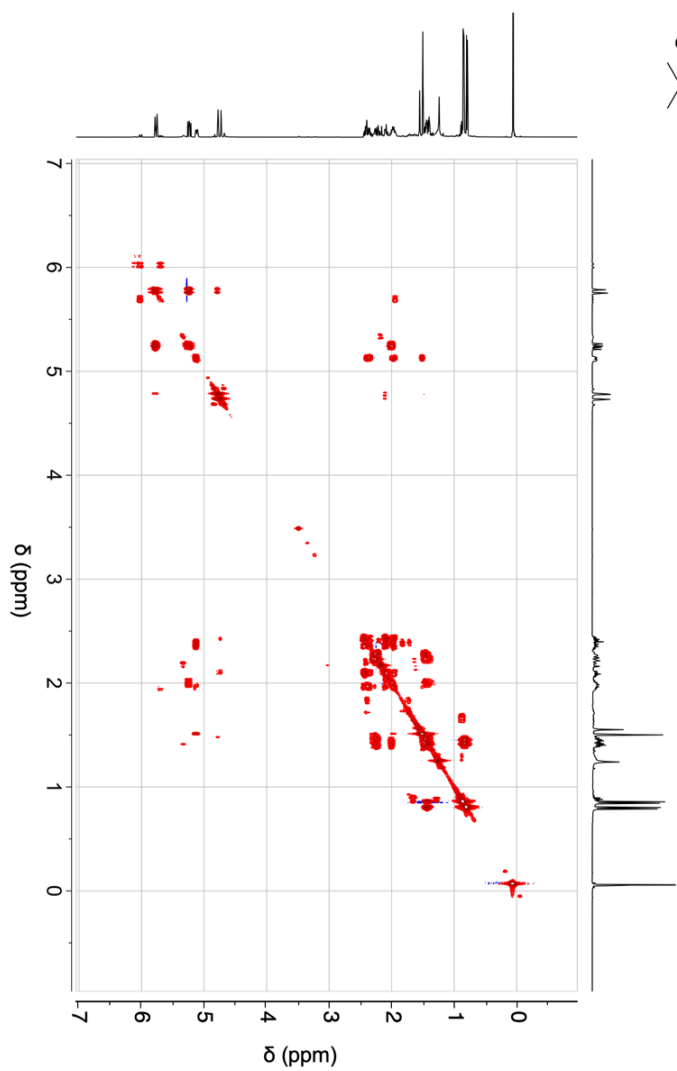


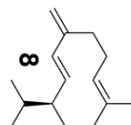
(-)-Germacrene D HMBC



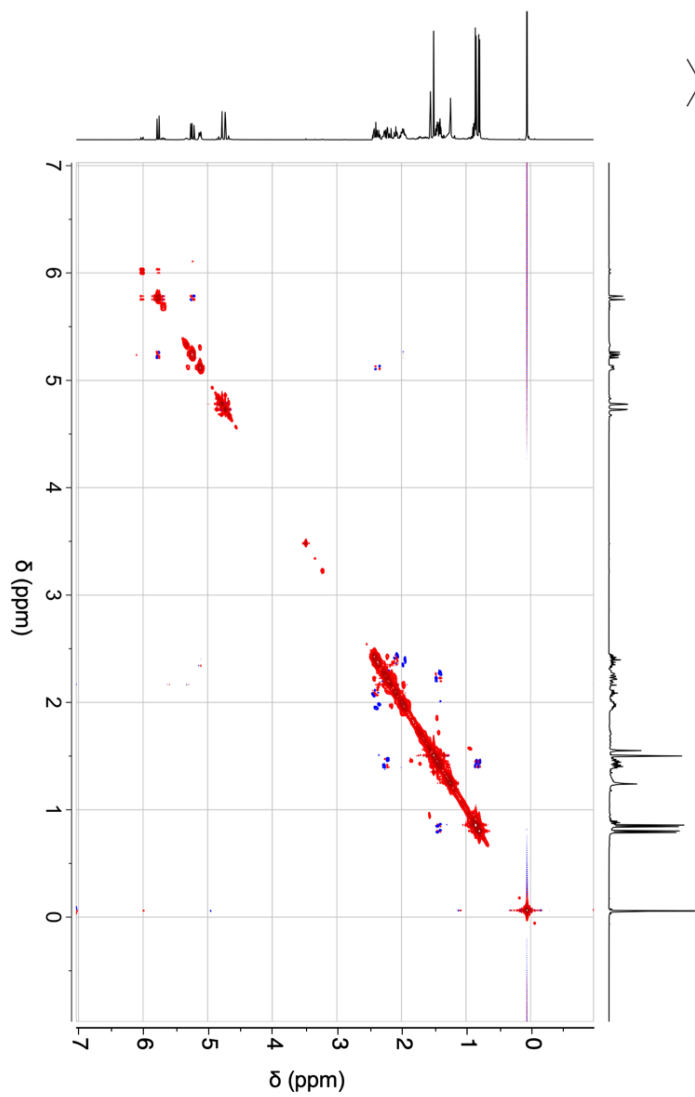


(-)-Germacrene D COSY





(-)-Germacrene D NOESY



Supplementary References

1. S. Podell, *et al.*, Pangenomic comparison of globally distributed *Poribacteria* associated with sponge hosts and marine particles. *ISME J.* **13**, 468–481 (2019).
2. D. Erpenbeck, *et al.*, Molecular biodiversity of Red Sea demosponges. *Mar. Pollut. Bull.* **105**, 507–514 (2016).
3. T. L. Turner, S. Lonhart, The Sponges of the Carmel Pinnacles Marine Protected Area. *bioRxiv* (2022) <https://doi.org/https://doi.org/10.1101/2022.11.02.514922>.
4. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
5. A. V. Zimin, *et al.*, The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677 (2013).
6. B. J. Walker, *et al.*, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, 1–14 (2014).
7. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
8. R. Challis, E. Richards, J. Rajan, G. Cochrane, M. Blaxter, BlobToolKit - interactive quality assessment of genome assemblies. *G3 Genes, Genomes, Genet.* **10**, 1361–1374 (2020).
9. L. Schmittmann, S. Franzenburg, L. Pita, Individuality in the Immune Repertoire and Induced Response of the Sponge *Halichondria panicea*. *Front. Immunol.* **12**, 1–13 (2021).
10. A. D. Finoshin, *et al.*, Iron metabolic pathways in the processes of sponge plasticity. *PLoS One* **15**, 1–25 (2020).
11. B. Buchfink, K. Reuter, H. G. Drost, Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
12. S. R. Eddy, Accelerated profile HMM searches. *PLoS Comput. Biol.* **7** (2011).
13. I. Burkhardt, T. De Rond, P. Y. Chen, B. S. Moore, Ancient plant-like terpene biosynthesis in corals. *Nat. Chem. Biol.* **18**, 664–669 (2022).
14. I. M. A. Chen, *et al.*, The IMG/M data management and analysis system v.6.0: New tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2021).
15. T. Ryu, *et al.*, Hologenome analysis of two marine sponges with different microbiomes. *BMC Genomics* **17**, 1–11 (2016).
16. M. Srivastava, *et al.*, The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010).
17. N. J. Kenny, *et al.*, Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. *Nat. Commun.* **11**, 1–11 (2020).
18. B. Plese, *et al.*, Mitochondrial evolution in the Demospongiae (Porifera): Phylogeny, divergence time, and genome biology. *Mol. Phylogenet. Evol.* **155**, 1–14 (2021).
19. T. Lassmann, O. Frings, E. L. L. Sonhammer, Kalign2: High-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res.* **37**, 858–865 (2009).
20. B. Q. Minh, *et al.*, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
21. J. Piel, *et al.*, Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl. Acad. Sci.* **101**, 16222–16227 (2004).
22. K. M. Fisch, *et al.*, Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* **5**, 494–501 (2009).
23. T. Mori, *et al.*, Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc. Natl. Acad. Sci.* **115**, 1718–1723 (2018).
24. V. Agarwal, *et al.*, Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat. Chem. Biol.* **13**, 537–543 (2017).
25. Y. Nakashima, Y. Egami, M. Kimura, T. Wakimoto, I. Abe, Metagenomic analysis of the sponge *Discodermia* reveals the production of the cyanobacterial natural product kasumigamide by “*Entotheonella*.” *PLoS One* **11**, 1–15 (2016).
26. M. F. Freeman, A. L. Vagstad, J. Piel, Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated “*Entotheonella*” bacteria. *Curr. Opin. Chem. Biol.* **31**, 8–14 (2016).
27. R. Ueoka, *et al.*, Metabolic and evolutionary origin of actin-binding polyketides from diverse

- organisms. *Nat. Chem. Biol.* **11**, 705–712 (2015).
28. P. Baer, *et al.*, Induced-fit mechanism in class I terpene cyclases. *Angew. Chemie - Int. Ed.* **53**, 7652–7656 (2014).
 29. J. E. McMurry, G. K. Bosch, Synthesis of Macrocyclic Terpenoid Hydrocarbons by Intramolecular Carbonyl Coupling: Bicyclgermacrene, Lepidozene, and Casbene. *J. Org. Chem.* **52**, 4885–4893 (1987).
 30. T. J. A. Bruce, *et al.*, Response of economically important aphids to components of *Hemizygia petiolata* essential oil. *Pest Manag. Sci.* **61**, 1115–1121 (2005).
 31. Z. Y. Huang, Q. Y. Wu, C. X. Li, H. L. Yu, J. H. Xu, Facile Production of (+)-Aristolochene and (+)-Bicyclgermacrene in *Escherichia coli* Using Newly Discovered Sesquiterpene Synthases from *Penicillium expansum*. *J. Agric. Food Chem.* **70**, 5860–5868 (2022).
 32. R. Faure, Two-Dimensional Nuclear Magnetic Resonance of Sesquiterpenes. 4*-Application to Complete Assignment of ¹H and ¹³C NMR Spectra of Some Aromadendrane Derivatives. *Magn. Reson. Chem.* **29**, 969–971 (1991).
 33. M. Yasumoto, K. Mada, T. Ooi, T. Kusumi, New terpenoid components from the volatile oils of the soft corals *Clavularia viridis* and *Sarcophyton acutangulum*. *J. Nat. Prod.* **63**, 1534–1536 (2000).
 34. T. Røstelién, A. K. Borg-Karlson, J. Fäldt, U. Jacobsson, H. Mustaparta, The plant sesquiterpene germacrene D specifically activates a major type of antennal receptor neuron of the tobacco budworm moth *Heliothis virescens*. *Chem. Senses* **25**, 141–148 (2000).
 35. H. Itokawa, H. Matsumoto, S. Mihashi, Constituents of *Torilis japonica* D.C. I. Isolation and Optical Purity of Germacra-4 (15), 5 (E), 10 (14)-trien-1 β -ol. *Chem. Pharm. Bull.* **31**, 1743–1745 (1983).
 36. M. Niwa, M. Iguchi, S. Yamamura, Co-occurrence of (-) and (+)-Germacrene-D in *Solidago altissima* L.: Determination of the Optical Rotation of optically Pure Germacrene-D. *Chem. Pharm. Bull.* **28**, 997–999 (1980).

Section 3.6 Taxonomic identification of San Diego sponges

Over the course of this project, we collected several marine sponges from various locations in San Diego. One of the most challenging aspects of the early days of this project was determining which species of sponge we had collected. Many of the local marine invertebrate guidebooks only included a few sponge species and often included outdated taxonomic names. The more detailed books on sponge taxonomy were quite dense and relied on specialized training in spicule analysis, which I lacked. When I read sponge natural product papers, I rarely found information on how the authors determined the taxonomy of the sponge they studied. Most natural products papers did not include a photograph of the sponge and if a written description of the sponge was provided, it was often brief and unhelpful. I noticed that many papers did not explicitly state whether a taxonomist made the species determination or if the natural product chemist made the determination.

As I began reading literature on sponge taxonomy, I realized there were many challenges that made the taxonomic identification of sponges particularly challenging. Many sponges have very similar morphologies, and it is easy to confuse one species for another based on morphology alone. Some sponge species also have huge variation in morphology, meaning two members of the same species can have large differences in color, shape, or texture. Until recently, all sponge taxonomy was based on spicule analysis, which requires specialized microscopy and a highly trained eye. More recently, the field has started to incorporate DNA analysis, including DNA barcoding using 18S rRNA, 28S rRNA, ITS2 rRNA, and CO1 rRNA (55). As taxonomists have started using DNA barcoding to determine sponge taxonomy, there has been major reorganization of the spicule-based taxonomy of marine sponges. Many of the spicule-based phylogenies do not match the DNA-based phylogenies, which has caused much debate about how to organize the

porifera taxonomic tree. Some DNA barcodes are only useful for family or genus-level classification and are not precise enough to make a species-level ID. Additionally, for DNA barcoding to be a helpful tool for de novo taxonomic identification, DNA barcodes must be sequenced for all known sponge species and those sequences must be deposited into publicly accessible databases. Many sponge species have not been DNA barcoded and/or have not had DNA barcode sequences entered into public databases, which makes it challenging to determine the species of an unknown sponge based on DNA barcoding alone.

To address some of these issues, I included photographs (Fig 3.6.1-3.6.2) and DNA barcoding data (Table 3.6.1-3.6.2) for 24 San Diego sponges. These sponges were collected from Quast Rock (32°51'21.8"N 117°16'45.4"W) in San Diego, CA in November 2018 at approximately 23 m depth and approximately 14°C. Whole genomic DNA was extracted from tissue stored in RNAlater using a DNeasy PowerLyzer PowerSoil Kit (Qiagen). Taxonomic identifications were made via PCR amplification of subunits of two mitochondrial genes commonly used for sponge barcoding: cytochrome oxidase subunit 1 (CO1) and 28S nuclear rRNA subunit (28S). PCR primers dgLCO1490 (5'-3') GGT CAA CAA ATC ATA AAG AYA TYG G and dgHCO2198 (5'-3') TAA ACT TCA GGG TGA CCA AAR AAY CA were used for CO1 amplification (56) and PCR primers 28S-C2-fwd (5'-3') GAA AAG AAC TTT GRA RAG AGA GT and 28S-D2-rev (5'-3') TCC GTG TTT CAA GAC GGG were used for 28S amplification (57). For each 25µL reaction, 10uL of 5PRIME HotMasterMix (QuantaBio), 2µL of 10µM each oligonucleotide primers, 1µL sponge template DNA, and 12µL ultrapure water were combined and run through the following thermocycler program: initial denaturation for 2 minutes at 94 °C, 30 cycles of denaturation for 30 seconds at 94 °C, annealing for 30 seconds at 50 °C for CO1 and 48 °C for 28S, extension for 1 minute at 65 °C, and a final extension for 5 minutes at 65 °C. Amplicons were

sent for sequencing to GeneWiz and amplicon sequences were compared against NCBI GenBank entries using BLAST.

Consulting a sponge taxonomist, Prof. Thomas Turner, at UC Santa Barbara was incredibly helpful and I would recommend his guidance in future studies of Southern California sponges. His two recent publications on the sponges of the Carmel Pinnacles (58) and California Axinellida (59) sponges were particularly helpful in the taxonomic identification of the San Diego nitrogenous terpene-producing sponge described above. The sponge had been previously identified as *Axinella* sp. by Thompson, Walker, and Faulkner in 1985. The genus *Axinella* has since been shown to be polyphyletic based on DNA barcoding (60) and many species from the orders Agelasida, Axinellida, Bubarida, and Suberitida have been reassigned in recent years (61, 62). Using 28S barcoding, spicule analysis, and morphological analysis, Turner and Lonhart were able to confirm that our local nitrogenous terpene-producing sponge was not an *Axinella*, but instead belonged to a new genus and species in the order Bubarida that is currently undergoing formal description (58). This valuable insight could have only been provided by a sponge taxonomist, and I would recommend more sponge natural product chemists consult with sponge taxonomists.

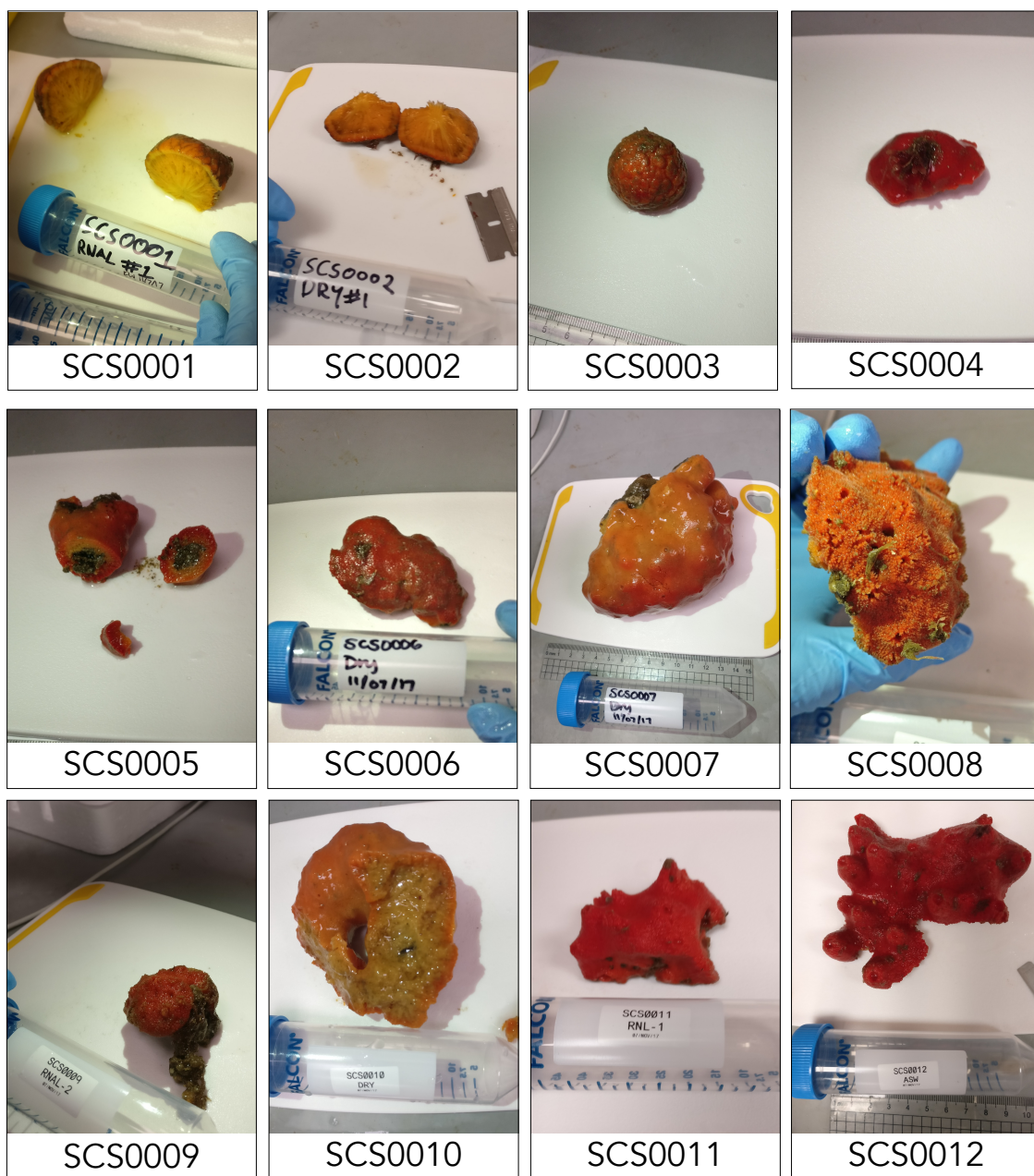


Figure 3.6.1. Photographs of San Diego sponges SCS0001-SCS0012 in the lab approximately 3 hours after collection.

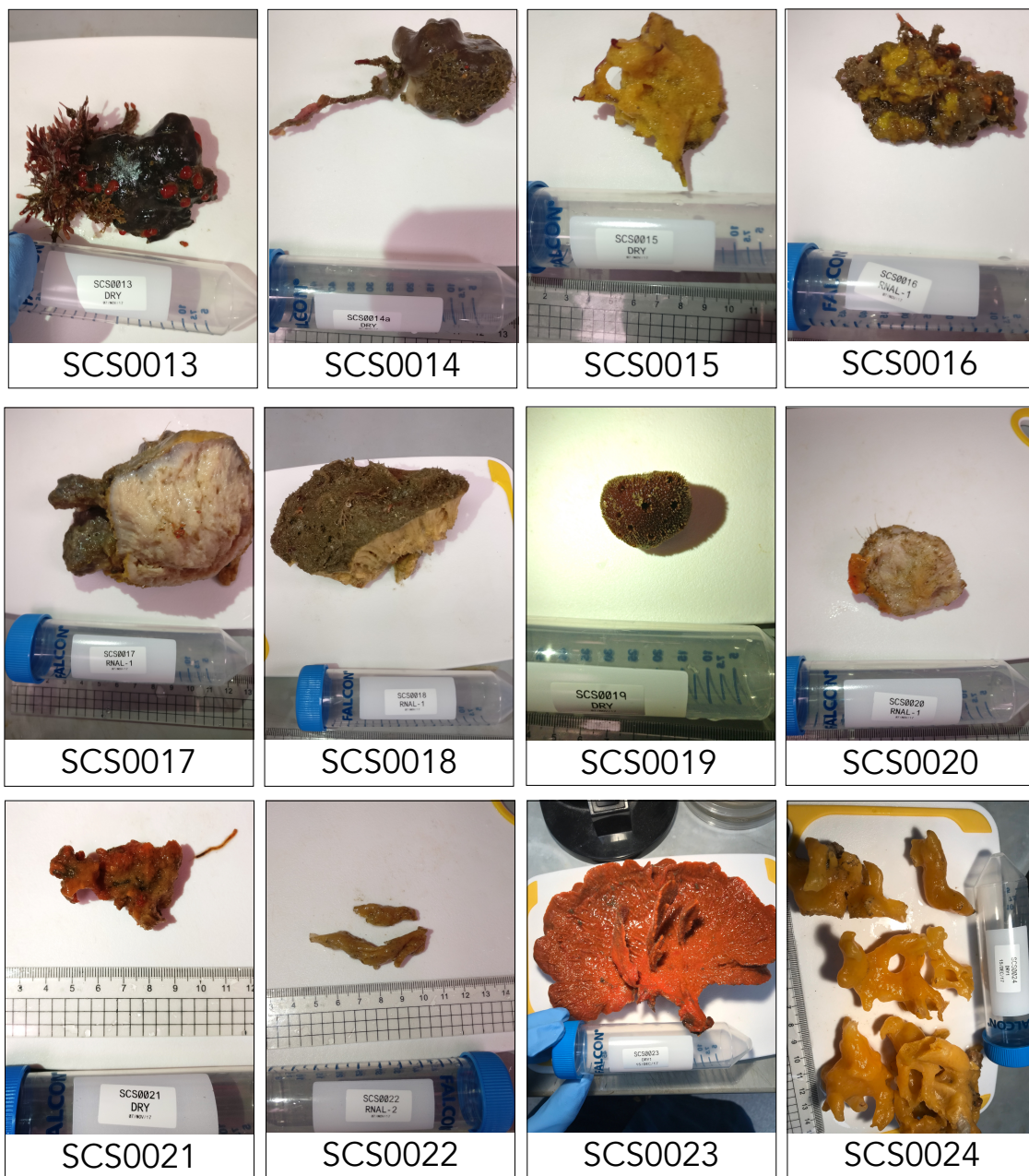


Figure 3.6.2. Photographs of San Diego sponges SCS0013-SCS0024 in the lab approximately 3 hours after collection.

Table 3.6.1. 28S DNA barcoding results table.

Specimen	28S Sequence	28S Blast hit	Sequence ID	E value	Percent Identity
SCS0001	TAGGAGGGAGCGATGCAGCTGAGTAGC TCCCTTTCAGGCTCAGGAAGTGTCGGAT CGGCTTGCTGGCTGGCGGAAGCCGTAA GGCGCCGGCGGTTGGACGGCCGGTCTGA CTGCTCTGCACTCCTGGGGGGAGCTGGC CAACAACAGTTGCCTCTGGCTCACAAG GGCGGTTGGGCAGGTGGCCTCCTCTTCG AGGTGGAACCTACAGCCATCCGTCTGG CAGTCCGGAGGCGACTGAGGAGAGTTG TGAACCTTTCACACTTGCACTGCAGGCC TTCGGGGCCGGGCCGCTCTTTCCGTTT GGGCACCGCTCCGGGGACTGCTTGCA TGTCCGCTGACGAATGTCTGCTCGGTGG GGGGAATGGTCACGCCGTGTGCTGCAG TTGTTGGTGTCCAAATGGCTGCATCCGA CCCGTCTTGAAAACACGGAA	<i>Tethya californiana</i> voucher TLT361 large subunit ribosomal RNA gene, partial sequence	MT137925.1	0	99.06
SCS0002	GCAGCTGAGTAGCTCCCTTTCAGGCTCA GGAAGTGTCGGATCGGCTTGCTGGCTG GCGGAAGCCGTAAGGCGCCGGCGGTTG GACGGCCGGTTCGACTGCTCTGCACTCCT GGGGGGAGCTGGCCAACAACAGTTGCC TCTGGCTCACAAGGGCGGTTGGGCAGG TGGCCTCCTCTTCGAGGTGGAACCTACA GCCATCCGTCTGGCAGTCCGGAGGCGA CTGAGGAGAGTTGTGAACTTTCACACTT GCAGTGCAGGCCCTTCGGGGCCGGGCC GCCTCTTCCGTTTGGGCACCGCTCCGG GGACTGCTTGCACTGTCCGCTGACGAAT GTCTGCTCGGTGGGGGAATGGTCACG CCGTGTGCTGCAGTTGTTGGTGTCCAAA TGGCTGCATCCGACCCGTCTTGAAAACA CGGAA	<i>Tethya californiana</i> voucher TLT361 large subunit ribosomal RNA gene, partial sequence	MT137925.1	0	99.51
SCS0003	CTGAGTAGCTCCCTTTCAGGCTCAGGAA GTGTCGGATCGGCTTGCTGGCTGGCGG AAGCCGTAAGGCGCCGGCGGTTGGACG GCCGGTCGACTGCTCTGCACTCCTGGGG GGAGCTGGCCAACAACAGTTGCCTCTG GCTCACAAGGGCGGTTGGGCAGGTGGC CTCCTCTTCGAGGTGGAACCTACAGCCA TCCGTCTGGCAGTCCGGAGGCGACTGA GGAGAGTTGTGAACTTTCACACTTGCA TGCAGGCCCTTCGGGGCCGGGCCGCT CTTCCGTTTGGGCACCGCTCCGGGGAC TGCTTGCACTGTCCGCTGACGAATGTCT CTCGGTGGGGGAATGGTCACGCCGT GTGCTGCAGTTGTTGGTGTCCAAATGGC TGCATCCGACCCGTCTTGAAAACACGG AA	<i>Tethya californiana</i> voucher TLT361 large subunit ribosomal RNA gene, partial sequence	MT137925.1	0	99.51

Table 3.6.1.(Cont.) 28S DNA barcoding results table.

Specimen	28S Sequence	28S Blast hit	Sequence ID	E value	Percent Identity
SCS0004	ATGCAGCCTGATGGCTCCGTTCTGGGTTTC AGGGAGGGTCAGTCGGTCGGANAGTGC TGGTCGACTCGCGAGAGTGGCCAGTCT CGTTTGGCCGGCGACCCCTCTGCATTCC CGTCTTGGAGTCGACCAACGGGGGTGG TAGGTGACCGGCAGGCAGGAGGGCAGG TGCCTCGAGTGGCCTTTGGGCCGCTGGA GAACTTAAAGCCTCCAGTTGTGCTATGG TTTGCCAGCTGCCGAGGAGGTTGTGCG ACCTTCGCACATGTCAGCGCAGGCTCCG CAGGGGCCGGGTCGCTTCTGTCCGTGTG GAAAGCGCGGCAAGGGACTGCTTGCA TGTCTGCTGACGCTCGCTCCACTCGGGC GGGAGGCGGGCCACGACGTGCGCTGAA GTCGTTGGTTGAGGGATGGCTTCATCCG ACCCGTCTTGAAACACGGAA	<i>Dragmacido n mexicanum</i> voucher TLT489 large subunit ribosomal RNA gene, partial sequence	OM256551.1	0	99.54
SCS0005	GCGATGCAGCCTGATGGCTCCGTTCTGG GTTTCAGGGAGGGTCAGTCGGTCGGAGA GTGCTGGTCGACTCGCGAGAGTGGCCA GTCTCGTTTGGCCGGCGACCCCTCTGCA TTCCCGTCTTGGAGTCGACCAACGGGG GTGGTAGGTGACCGGCAGGCAGGAGGG CAGGTGCCTCGAGTGGCCTTTGGGCCGC TGGAGAACTTAAAGCCTCCAGTTGTGCT ATGGTTTGCCAGCTGCCGAGGAGGTTGT GCGACCTTCCCACATGTCAGCGCAGGCT CCGACGGGGCCGGGTCGCTTCTGTCCGT GTGGAAGCGCGGCAAGGGACTGCTTG CAGTGTCTGCTGACGCTCGCTCCACTCG GGCGGGAGGCGGGCCACGACGTGCGCT GAAGTCGTTGGTTGAGGGATGGCTTCAT CCGACCCGTCTTGAAACACGGAA	<i>Dragmacido n mexicanum</i> voucher TLT489 large subunit ribosomal RNA gene, partial sequence	OM256551.1	0	99.31
SCS0006	GATGCAGCCTGNTGGNTCCGTGTCTGGG TTCAGGGAGGGTCATTCTGGTCGGAGNG TGCTGGTCCACTCNCGACAGTGGCCAAT CTCGTTTGGCCGGCGACCCCTCTGCATT CCCGTCTTGGAGTCTACCAACGGGGGT GGTAGGTGACCGGCAGGCTGGAGGGCA GGTGCCTCCAGTGGCCTTTGGGCCGCTG GATAACTTAAAGCCTCCAGTTGTGCTAT GGTTTGCCANCTGCCGAGGAGGTTGTG CGACCTTCCCACATGTCNCGCAGGCTC CGCAGGGGCCGGGTCGCTTCTGTCCGTG TGGAAGCGCGGCAAGGGACTGCTTGC AGTGTCTGCTGACGCTCGCTCTACTCGG GCGGGAGGCGGGCCNCNACNTGCNCTG AAGTCNTTGGTTGAGGGATGGCTTCNTC CTGACCCNNCNTGAANCACNGAANNAC CCGTCTTGAAACACGGAA	<i>Dragmacido n mexicanum</i> voucher TLT489 large subunit ribosomal RNA gene, partial sequence	OM256551.1	0	93.5

Table 3.6.1.(Cont.) 28S DNA barcoding results table.

Specimen	28S Sequence	28S Blast hit	Sequence ID	E value	Percent Identity
SCS0007	CGATGNAGCCTGATGGCTCCGTNTCNG GTTTCAGGGAGGGTCANTCGGTCGNAGA GTGNTGGTCCACTCNCGAGAGTGGCCA ATCTCGTTTGGCCGGCGACCCCTCTGCN TTCCCGTCTTGAGTCTACCAACGGGGG TGGTAGGTGACCGGCACGCAGGAGGGC AGGTGCCTCCAGTGGCCTTTGGGCCGCT GGATAACTTAAAGCCTCCAGTTGTGCTA TGGTTTGCCAGCTGCCGAAGAGGTTGTG CGACCTTCCCACATGTCANCGCAGGCTC CGCAGGGGGCCGGGTCGCTTCTGTCCGTG TGAAAGCGCGGCNAGGGACTGCTTGC AGTGTCTGCTGACGCTCGCTCCACTCGG GCGGGAGGCGGGCCACGACNTGCGCTG AAGTCNNTGGTTGAGGGATGGCTTCAT CCNACCCNNCTTGAAACACNGAANNAC CCGTCTTGAAACACGNA	<i>Dragmacido n mexicanum</i> voucher TLT489 large subunit ribosomal RNA gene, partial sequence	OM256551.1	0	94.19
SCS0008	n/a				
SCS0009	TCAGGGTGTCGAGGTTGTCGTGCTCCAT GTAGAGTTCAGTTTGTAGTGCCGGGNTG CCGGTGCTGCGGACGCTTCTCGGATGG AGTTGGACCGTCAGTTGGGAAGAGTCT GGTGCCTGTGCAATCAGTGGCCGGGAG GTAAGCCCTGCAGTCCTTGCGGTCTGCG GGGTTACGTATAGCCGTGCCGTTGTGGC TCCAGTCGTCTTTCAGGTGTGTTGGATG TCTGATCGTCCATAAATGGAAGGTGCA GTCTCCCTCGTGGGAGTTCGGGCCGGT TCCGGGGTTGCGTCTGTACGCCGAGTC GTGTTCCGTTCTCTCTGTACGCTTCG GTGTGTGGGGTAGAGTACGGCACGGCC TGACGTGTCAGTGTCCCGTCGGGACTGT GTGCCAAGTTCTGTGCCGAGGTGACTG GCGAAGAGTCGTTACCCTGTGACCCGTC TTGAAACACGGAA	<i>Heterosclero morpha</i> sp. 8 JV-2020 voucher SI0119172 large subunit ribosomal RNA gene, partial sequence	MW016136. 1	8.00E -80	79.35
SCS0010	CGATGCAGCCTGATGGCTCCGTTCCGGT TCAGGGAGGGTCAGTCGGTCGGAGAGT GCTGGTCTGACTCGCGAGAGTGGCCAGT CTCGTTTGGCCGGCGACCCCTCTGCATT CCCGTCTTGAGTTCGACCAACGGGGGT GGTAGGTGACCGGCAGGCAGGAGGGCA GGTGCCTCCAGTGGCCTTTGGGCCGCTG GAGAACTTAAAGCCTCCAGTTGTGCTAT GGTTTGCCAGCTGCCGAAGAGGTTGTG CGACCTTCCCACATGTCAGCGCAGGCTC CGCAGGGGGCCGGGTCGCTTCTGTCCGTG TGAAAGCGCGGCAAGGGACTGCTTGC AGTGTCTGCTGACGCTCGCTCTACTCGG GCGGGAGGCGGGCCACGACGTGCGCTG AAGTCGTTGGTTGAGGGATGGCTTCATC CGACCCGTCTTGAAACACGGAA	<i>Dragmacido n mexicanum</i> voucher TLT489 large subunit ribosomal RNA gene, partial sequence	OM256551.1	0	99.07

Table 3.6.1.(Cont.) 28S DNA barcoding results table.

Specimen	28S Sequence	28S Blast hit	Sequence ID	E value	Percent Identity
SCS0011	NNGCGTGNTNGANTNNGNTCCTNNTCC NNGCTCAAGAGACGTCTCGTGGNTTNN NTCTGCNNCCCGCCGACCACTCTGCACT TCTGCGTTGAGCCGGCCAACGGCAATT GTCCCCGGCCCGGGANGGCGTCGGGGA AGGTGGCTCCCCTCTTGCCTGGANAAC TATAGCCCTACGTCTGGACGTTCCGGTG GNGATTGANGAGAGTTGCNAGTCTTCG CAAGTTGCAACCCACGCCTCCCCCGAG CTGGGCCGCTTCTGTCCGGATGGTTGC CGAGGCTGTCACTGCGTGCAGTGGCGG TCAACCGTCCCGCTTGCCGGGCGGGAG GTCGGTCGCACCTTGCCTGCGNTCGTT GGTGCCTGGATGGCCGCATCCGACCCG TCTTGGAAACGCGGAATTCTGTGCCGAG GTGACTGGCGAAGAGTCGTTACCCTGT GACCCGTCTTGAAACACGGAA	<i>Demospongi</i> <i>ae</i> sp. GW41396 genomic DNA sequence contains 28S rRNA gene, isolate GW41396, specimen voucher 161119-02- ZMSBUK	LR596426.1	4.00E -88	85.13
SCS0012	GNGTGAAGTGNNTCCTGNTCATGTCAA GANACGTTTCGTGGCCTCNNTCTGNGN GCTGCNACCACTTTGTCTTCCGAGGGG AGCCGGCCAATGGGAATAGTCTGCGGC CCGTGANGTCNTCGGGNAAGGTGGCTC CCCNCTTGCCCTTGNNACTTATAGCCCT ACGCTCTGGACGTTCCGGTGGGGATTGA GGAGAGTTTCGAGTCTTCTGNATGTGCG ATCCACGCCTCCCCCAAGGCTGGGCCTC TTCTGTCCGGATGGTTGCCAGGTTGTG GCTGGCTGCTGTGCCGGTCAACCGTGCC CCTTGCTCTCTCTGANGTCTTCCGCACC TTGNGCTGAGNTCNTTGGTGCCTGGATG GCCGCATCCCACCCGNCCTTGGTGTGCCG AATTCTGTGCCGAGGTGACTGGCGAAG AGTCGTTACCCTGTGACCCGTCTTGAAA CACGGAA	<i>Chalinidae</i> sp. 1 JV- 2020 voucher KBOA09271 756 large subunit ribosomal RNA gene, partial sequence	MW016063. 1	5.00E -17	96.77
SCS0013	TGCTCTATGTCCAGGTCCACTTGTGGNG CTGGGGCGNTGGCGCTGCTGACNNTAT CTGGGAGGGCCCTGGANNGTCTCTGG AAANAATGGGGCGGACTGNNNTTGGTG AAATGGNTTTCTTTACCTGNNCTCCTTG CGCTCAANGGGGTACGTATAGCCGTG CCGTTGTGGTCCAGTCGTCTTTCAGGT GTGTTGGATGTCTGATCGTCCATAAATG GAAGGTGCAGTCTCCCCCTCGTGGGAGTT CGGGCCGGTTCCGGGGTTGCGTCTGTAC GCCGGAGTCGTGTTCCGTTCTCTCTCTG TACGCTTCGGTGTGTGGGGTAGAGTAC GGCACGGCCTGACGTGTCAGTGTCCCGT CGGGACTGTGTGCCAAGTTCTGTGCCGA GGTGACTGGCGAAGAGTCGTTACCCTG TGACCCGTCTTGAAACACGGAA	<i>Haliclona</i> sp. GW33419 genomic DNA sequence contains 28S rRNA gene, specimen voucher SNSB- BSPG.GW3 3419	OX421643.1	1.00E -53	81.6

Table 3.6.1.(Cont.) 28S DNA barcoding results table.

Specimen	28S Sequence	28S Blast hit	Sequence ID	E value	Percent Identity
SCS0014	AGGTTGTCGTGCTCCATGTANAGTTCAG TTTGTAGTGCCGGGGTGCCGGTGCTGCG GACGCTTTCTCGGATGGAGTTGGACCGT CAGTTGGGAAGAGTCTGGTGCCTGTGC AATCAGTGGCCGGGAGGTAAGCCCTGC AATCCTTGCGGTCTGCGGGGTACGTAT AGCCGTGCCGTTGTGGCTCCAGTCATCT TTCAGGTGTGTTGGATGTCTGATCGTCC ATAAATGGAAGGTGCAGTCTCCCCTCGT GGGAGTTTCGGGCCCGGTTCCGGGGTTGC GTCTGTACGCCGGAGTCGTGTTCCGTTT TCTCTGTACGCTTCGGTGTGTGGGGT AGAGTACGGCACGGCCTGACGTGTCAG TGTCCCGTCGGGACTGTGTGCCAAGTTC TGTGCCGAGGTGACTGGCGAAGAGTCG TTACCCTGTGACCCGTCTTGAAACACGG AA	<i>Heterosclero morpha</i> sp. 8 JV-2020 voucher SI0119172 large subunit ribosomal RNA gene, partial sequence	MW016136. 1	2.00E -70	78.41
SCS0015	NCGATGCAGCTGATTAGCTCCGCTCAA GTTTCAGGAGCGACGTCGAGCAGTGATG CCGGCCCGCAAGAGCCGATAGGCTTTG CTCTCCGGATAGCTGTGACTCCTCTGC ACTCTTGTGCGAGCTGGCCAACAACAG TTGCCTGCGGCTGACAAGGGGGGACGG GCAGGTGCTCACTCTTCGGAGTGATACT TACAGCCGTTTCGTTTGGAAGTCCGCAGG TGACTGAGGATAACTGGCAGTGACAGGC CTCCGGGTCTTGGGCTGCTCCTGGTTCGT TTGGGCACCGCGTCAGGGTCTGCGTGC AGTGTCTGTGACGGATGACCTCTCGGG CGGGAGACCAGTCACACCTTGCGCTGC AGTTGTTGGTGATCGAATGGCCGCATCC GACCCGTCTTGAAAACACGGAA	<i>Haplosclerid a</i> sp. A KZ- 2019 internal transcribed spacer 2 and large subunit ribosomal RNA gene, partial sequence	MN153806.1	0.00E +00	98.01
SCS0016	n/a				
SCS0017	n/a				
SCS0018	n/a				

Table 3.6.1.(Cont.) 28S DNA barcoding results table.

Specimen	28S Sequence	28S Blast hit	Sequence ID	E value	Percent Identity
SCS0019	GGGTGTCGAGGTTGTCGTGCTCCATGTA GAGTTCAGTTTGTAGTGCCGGGGTGCCG GTGCTGCGGACGCTTTCTCGGATGGAGT TGGACCGTCAGTTGGGAAGAGTCTGGT GCCTGTGCAATCAGTGGCCGGGAGGTA AGCCCTGCAGTCCTTGCGGTCTGCGGGG TTACGTATAGCCGTGCCGTTGTGGCTCC AGTCGTCTTTCAGGTGTGTTGGATGTCT GATCGTCCATAAATGGAAGGTGCAGTC TCCCCTCGTGGGAGTTTCGGGCCGTTCC GGGGTTGCGTCTGTACGCCGGAGTCGT GTTCCGTTCTCTCTGTACGCTTCGGT GTGTGGGGTAGAGTACGGCACGGCCTG ACGTGTCAGTGTCCCGTCGGGACTGTGT GCCAAGTTCTGTGCCGAGGTGACTGGC GAAGAGTCGTTACCCGTGACCCGTCTT GAAAACACGGAA	<i>Heterosclero morpha</i> sp. 8 JV-2020 voucher SI0119172 large subunit ribosomal RNA gene, partial sequence	MW016136. 1	8.00E -80	79.44
SCS0020	n/a				
SCS0021	AACGAATGCAGCCCGATGGGCTTCGGT CGGGTTCAGGGGCGGTGGTGCCGTCGG AGCGGGCGGG CGGACCCGTGAGGGTCGCCTTTCAGTTC TCGAGCGGTCCATCGCCCTGCACTCCCG GTCGATAGCC GGCCAACGTCCGTCCGGTTCGGCTCAC GCGGATGGTCGGGCAGGTGCCTCAATG CGCCGCTTCGG CGTGCGCAGGGAGAACTTACAGCCGAT TGTTCCGGCAGTCGGACGTTGCGGACCG AGGAGATGTGT GACGCTTCACACCGGCAGTGCAGGCTC GCCGTTCTCGCGGTGGGCCGGGTCTCTT CTGATCGTTG GGTAGGCGCTACTGGGACTGCCCGTCA GTGCCAGTCGACGTCCCGTCCCCTCGGT CGGGAGGTTG GTCACGCCTTGCGCTGTAGTCGCTGGCT GAAGGAAGGCTGCATCCGACCCGTCTT GAAACACGGA	<i>Axinella cannabina</i> 28S ribosomal RNA gene, partial sequence	GQ466062.1	0	95.16

Table 3.6.1.(Cont.) 28S DNA barcoding results table.

Specimen	28S Sequence	28S Blast hit	Sequence ID	E value	Percent Identity
SCS0022	AGCTCAGAGTAGTCGAGGTTGTCGTGCT CTGTTGGGGTTCAGCTGGTGGCAGTTGG GTTGCTGGCTGCTGCCGGTGCAGTCTCC GATGGAGTTGGACCGTCAGTTGGGAAG GCGTTGGGACCTGCTTGTCGGTGGTCGG GAGGTAAGTCTCGGTTCCGTATGGTGCT GGGATTACGTACAGCCGGTCCGTCGTG GTTCCGATTAGCCTTTCAGGTGCGTCAG ACGGCTTGTCGTCTGGAATGGGAAGGT GCAGTCTCTCCTCTGAGGGAGGCCAGG CCGGTGTCGAGGTTGTGTCGCAGTGTGG GGTCGTGCTCCGTTCTCTCTTCGATGCG CTCTAGCGCTCGGGGTTGGGTGCGGCG CGTCCCACAGCTTGCTGGCGTCCCGTCG GCATCGTGTAACAGGTTTTCTGTGCCTA GGTGACTGGCGAAGAGTCGTTACTCTGT GCCCCGTCTTGAAAACACGGAA	<i>Cladocroce</i> sp. UPDM SZN-HL117 isolate HL117 large subunit ribosomal RNA gene, partial sequence	MT224122.1	4.00E -73	78.41
SCS0023	n/a				
SCS0024	GATGCGGCCATAATGACTACCTTCGGGT TCAGGTTGTCTCGGTGATGGTGTGCGG TTGGTGGGAACTCGTTTGAGTCTCTCCT GGGCAAGCCTGCTGAGGCTCCCGCACT CCCGTCTGTAGTCCGCCAACGGCGGTCA GATGCGGTCCAACAGGGTCGTGGGGTT AGGTACCTCCTGTGTCGGCATCCTTCGG GTTGTTGGTGCGGGAGAACTTACAGCCT CCGGTCCGGAGATCGCACGTGACCGAG GACCGTTGCGACCATTTTCGCGCGGTGCA GCGCAGGCTCCGTTAGGGGCCGGGTCG CTCCTGGGCCTGGGTGCTTCGCGGCGTG CCACTGCATGCAGTGACCGCTGACGGT AATGCGCGCTCGCCCCGGGAGGTGCGCC ACGATGTGCGCTGTGGTCGTTGGTGTGC GGATGGCTTCATCCGACCCGTCTTGAAA ACACGGAA	<i>Endectyon</i> <i>hyle</i> voucher TLT411 large subunit ribosomal RNA gene, partial sequence	OM256563.1	0	99.78

Table 3.6.2. CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0001	GTTGGGAGATGACCATTGTATAATGTA ATTGTGACAGCTCATGCCTTTGTAATGA TATTTTTTTTAGTAATGCCAGTAATGAT TGGGGGGTTTGGAAATTGATTTGTGCCA TTATATATTGGTGCACCAGATATGGCAT TTCCTAGATTAAATAATATAAGTTTTTG GTTATTGCCTCCTTCGTAACTTTATTAT TGGGTTCTGCTTTTGTGGAACAAGGAGC TGGAAGTGGGTGGACAGTTTATCCACCT TTAGCAAGTATACAGGCTCATTCAAGGA GGATCAGTAGATATGGCAATATTTAGTT TACATTTAGCCGGGATTTCTTCAATTTT AAGTACAATAAATTTTATAACAACAATT TTAAATATGAGGGCACCGGTATAACA ATGGATAGAATGCCACTATTTGTTTGAT CAATTTTAGTTACAGTAGTTTATTATT ATTATCTTTACCTGTTTTAGCGGGTGCT ATTACAATGCTTCTTACAGACAGAAATT TTAATACAGCATTTTTTGATCCTGCTGG AGGAGGGGATCCAATTTTATATCAACA TTT	<i>Tethya californiana</i> cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria 1	KJ620403.1	0	100
SCS0002	TTGGGAGATGACCATTGTATAATGTAA TTGTGACAGCTCATGCCTTTGTAATGAT ATTTTTTTTAGTAATGCCAGTAATGATT GGGGGGTTTGGAAATTGATTTGTGCCAT TATATATTGGTGCACCAGATATGGCATT TCCTAGATTAAATAATATAAGTTTTTG TTATTGCCTCCTTCGTAACTTTATTATT GGGTTCTGCTTTTGTGGAACAAGGAGCT GGAAGTGGGTGGACAGTTTATCCACCTT TAGCAAGTATACAGGCTCATTCAAGGAG GATCAGTAGATATGGCAATATTTAGTTT ACATTTAGCCGGGATTTCTTCAATTTTA AGTACAATAAATTTTATAACAACAATTT TAAATATGAGGGCACCGGTATAACAA TGGATAGAATGCCACTATTTGTTTGATC AATTTTAGTTACAGTAGTTTATTATTA TTATCTTTACCTGTTTTAGCGGGTGCTA TTACAATGCTTCTTACAGACAGAAATTT TAATACAGCATTTTTTGATCCTGCTGGA GGAGGGGATCCAATTTTATATCAACATT TA	<i>Tethya californiana</i> cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria 1	KJ620403.1	0	100

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0003	TTGGGAGATGACCATTTGTATAATGTAA TTGTGACAGCTCATGCCTTTGTAATGAT ATTTTTTTTAGTAATGCCAGTAATGATT GGGGGGTTTGGAAATTGATTTGTGCCAT TATATATTGGTGCACCAGATATGGCATT TCCTAGATTAAATAATATAAGTTTTTGG TTATTGCCTCCTTCGTAACTTTATTATT GGGTTCTGCTTTTGTGGAACAAGGAGCT GGAAGTGGGTGGACAGTTTATCCACCTT TAGCAAGTATACAGGCTCATTGAGGAG GATCAGTAGATATGGCAATATTTAGTTT ACATTTAGCCGGGATTTCTTCAATTTTA AGTACAATAAATTTTATAACAACAATTT TAAATATGAGGGCACCAGGTATAACAA TGGATAGAATGCCACTATTTGTTTGATC AATTTTAGTTACAGTAGTTTATTATTATTA TTATCTTTACCTGTTTTAGCGGGTGCTA TTACAATGCTTCTTACAGACAGAAATTT TAATACAGCATTTTTTGATCCTGCTGGA GGAGGGGATCCAATTTTATA	<i>Tethya californiana</i> cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria l	KJ620403.1	0	100
SCS0004	TACAATGCTAGGTGATGATCATTTATAT AATGTTATAGTAACAGCCCATGCTTTAG TGATGATATTTTTCTTAGTTATGCCAGT AATGATAGGGGGGTTTGGGAATTGATT GGTACCGTTATATATAGGTGCGCCTGAT ATGGCTTTCCCAAGATTAAATAATCTTA GTTTTTGAGTATTACCCCTTCAATTAC TTTATTATTAGGTTCTGCTTTTGTAGAG CAAGGAGCTGGGACAGGATGAACACTT TACCCGCCCTTAGCAGGCATACAAACA CATTGAGGTGGATCTGTAGATATGGCA ATATTTAGTCTTCATTTAGCTGGTTTATC TTCTATATTAGGGTCAATGAACTTTATT ACAACTATTCTAAATATGAGGGCACCA GGGATTACAATGGATAAATTGCCATTAT TTGTGTGATCTATTTTAATTACAACCTTT TTATTAATATTAACCTTACCAGTATTGG CAGGTGCGATTACAATGCTTCTTACAGA TAGAAATTTTAATACAACATTCTTTGAC CCTGCTGGAGGAGGGGATCCGATACTA TATCAACATT	<i>Dragmacidon mexicanum</i> voucher TLT489 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondria l	OM234773.1	0	100

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0005	CGGTACAATGCTAGGTGATGATCATTTA TATAATGTTATAGTAACAGCCCATGCTT TAGTGATGATATTTTTCTTAGTTATGCC AGTAATGATAGGGGGGTTTGGGAATTG ATTGGTACCGTTATATATAGGTGCGCCT GATATGGCTTTCCCAAGATTAAATAATC TTAGTTTTTGAGTATTACCCCTTCAATT ACTTTATTATTAGGTTCTGCTTTTGTAG AGCAAGGAGCTGGGACAGGATGAACAC TTTACCCGCCCTTAGCAGGCATACAAAC ACATTCAAGGTGGATCTGTAGATATGGC AATATTTAGTCTTCATTTAGCTGGTTTA TCTTCTATATTAGGGTCAATGAACTTTA TTACAACCTATTCTAAATATGAGGGCACC AGGGATTACAATGGATAAATTGCCATT ATTTGTGTGATCTATTTTAATTACAAC TTTTTATTAATATTAACCTTACCAGTATT GGCAGGTGCGATTACAATGCTTCTTACA GATAGAAATTTTAATACAACATTCTTTG ACCCTGCTGGAGGAGGGGATCCGATAC TATATCAACATTTATTTTG	<i>Dragmacido n mexicanum</i> voucher TLT489 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondria 1	OM234773.1	0	100
SCS0006	n/a				
SCS0007	GGTACAATGCTAGGTGATGATCATTTAT ATAATGTTATAGTAACAGCCCATGCTTT AGTGATGATATTTTTCTTAGTTATGCCA GTAATGATAGGGGGGTTTGGGAATTGA TTGGTACCGTTATATATAGGTGCGCCTG ATATGGCTTTCCCAAGATTAAATAATCT TAGTTTTTGAGTATTACCCCTTCAATT ACTTTATTATTAGGTTCTGCTTTTGTAG AGCAAGGAGCTGGGACAGGATGAACAC TTTACCCGCCCTTAGCAGGCATACAAAC ACATTCAAGGTGGATCTGTAGATATGGC AATATTTAGTCTTCATTTAGCTGGTTTA TCTTCTATATTAGGGTCAATGAACTTTA TTACAACCTATTCTAAATATGAGGGCACC AGGGATTACAATGGATAAATTGCCATT ATTTGTGTGATCTATTTTAATTACAAC TTTTTATTAATATTAACCTTACCAGTATT GGCAGGTGCGATTACAATGCTTCTTACA GATAGAAATTTTAATACAACATTCTTTG ACCCTGCTGGAGGAGGGGATCCGATAC TATATCAACA	<i>Dragmacido n mexicanum</i> voucher TLT489 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondria 1	OM234773.1	0	100

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0008	GGGAGACGATCATTATATAATGTTATA GTAAGTCTCATGCTTTTGTATGATTTT TTTTTTAGTTATGCCTGTGATGATTGGG GGTTTTGGGAATTGGTTAGTGCCTTTAT ATATTGGTGCTCCGGATATGGCCTTTCC TCGATTAAATAATATAAGTTTTTGATTA TTGCCTCCGGCTTTAAGTTTATTATTAG CTTCTGCTTTTGTGGAGCAAGGGGCAGG AACTGGATGAACAGTTTATCCCCCTTA TCCGGGATTTCAGGCCCATTCGGGGGGTT CAGTTGATTTGGTTATTTTTAGTCTACA TTTGGCAGGAATTTTCGTCGATATTAGGT GCTATGAATTTTATTACAACAATAATAA ATATGAGGGCGCCGGGGATAACTATGG ATAGAATGCCATTATTTGTTTGATCTAT TTTAGTAACTGCTATTTTATTATTATTAT CTTTACCAGTATTAGCTGGGGCGATAAC TATGTTATTAACAGATAGAAATTTTAAT ACTGCA	<i>Clathria rugosa</i> cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria 1	KJ620406.1	0	97.06
SCS0009	AGTATGTTAATTCGGTTAGAGTTATCGG GTCCGGGATCAATGTTGGGAGACGATC ATTTATATAATGTTATAGTAACTGCTCA TGCTTTTGTATGATTTTTTTTTTTAGTTA TGCCTGTGATGATTGGGGGTTTTGGGAA TTGGTTAGTGCCTTTATATATTGGTGCT CCGGATATGGCCTTTCTCGATTAAATA ATATAAGTTTTTGATTATTGCCTCCGGC TTTAAGTTTATTATTAGCTTCTGCTTTTG TGGAGCAAGGGGCAGGAAGTGGATGAA CAGTTTATCCCCCTTTATCCGGGATTCA GGCCCATTCGGGGGGTTTCAGTTGATTG GTTATTTTTAGTCTACATTTGGCAGGAA TTTCGTCGATATTAGGTGCTATGAATTT TATTACAACAATAATAATATGAGGGC GCCGGGGATAACTATGGATAGAATGCC ATTATTTGTTTGATCTATTTTAGTAACTG CTATTTTATTATTATTATCTTTACCAGTA TTAGCTGGGGCGATAACTATGTTATTAA CAGATAGAA	<i>Clathria rugosa</i> cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria 1	KJ620406.1	0	96.49

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0010	CAATGCTAGGTGATGATCATTTATATAA TGTATAGTAACAGCCCATGCTTTAGTG ATGATATTTTTCTTAGTTATGCCAGTAA TGATAGGGGGGTTTGGGAATTGATTGG TACCGTTATATATAGGTGCGCCTGATAT GGCTTTCCCAAGATTAAATAATCTTAGT TTTTGAGTATTACCCCTTCAATTACTTT ATTATTAGGTTCTGCTTTTGTAGAGCAA GGAGCTGGGACAGGATGAACACTTTAC CCGCCCTTAGCAGGCATACAAACACAT TCAGGTGGATCTGTAGATATGGCAATAT TTAGTCTTCATTTAGCTGGTTTATCTTCT ATATTAGGGTCAATGAACTTTATTACAA CTATTCTAAATATGAGGGCACCAGGGA TTACAATGGATAAATTGCCATTATTTGT GTGATCTATTTTAATTACAACCTTTTTAT TAATATTAACCTTTACCAGTATTGGCAGG TGCATTACAATGCTTCTTACAGATAGA AATTTTAATACAACATTCTTTGACCCTG CTGGAGGAGGGGATCCGATACTATATC AAC	<i>Drumacido n mexicanum</i> voucher TLT489 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondria l	OM234773.1	0	100
SCS0011	TGTGGGAGACGATCATTTATATAATGT TATAGTAACTGCTCATGCTTTTGTTATG ATTTTTTTTTTAGTTATGCCTGTAATGAT TGGGGGTTTTGGGAAGTGGTTTGTGCCT TTATATATTGGTGCCCCAAATATGGCCT TTCCTCGATTAAATAATATAATTTTTTG ATTATTGCCTCCGGCTTTAATTTTATTAT TAGCTTCTGCTTTTGTGGAGCAAGGGGC AGGAACTGGATGAACAGTTTATCCGCC TTTATCTGGGATTCAGGCTCATTCCGGG GGTTCAGTAGATTTAGTAATTTTATAGTA TACATTTGGCAGGAATTTTCGTCAATATT AGGTGCTATGAATTTTATTACAACAATA ATAAATATGAGAGCGCCCGGGATAACT ATGGATAGAATGCCCTTAT	<i>Clathria rugosa</i> cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria l	KJ620406.1	0	96.84

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0012	TGTGGGAGACGATCATTATATAATGT TATAGTAACTGCTCATGCTTTTGTATG ATTTTTTTTTTAGTTATGCCTGTAATGAT TGGGGGTTTTGGGAACTGGTTTGTGCCT TTATATATTGGTGCCCCAAATATGGCCT TTCCTCGATTAAATAATATAATTTTTTG ATTATTGCCTCCGGCTTTAATTTTATTAT TAGCTTCTGCTTTTGTGGAGCAAGGGGC AGGAACTGGATGAACAGTTTATCCGCC TTTATCTGGGATTCAGGCTCATTCGGGG GGTTCAGTAAATTTAGTAATTTTAGTA TACATTTGGCAGGAATTCGTCAATATT AGGTGCTATGAATTTTATTACAACAATA ATAAATATGAGAGCGCCCGGGATAACT ATGGATAGAATGCCCTTATTTGTTTGAT CTATTTTAGTAACTGCTGTTTTATTATTA TTATCTTTACCCGTATTAGCTGGGGCGA TAACTATGTTATTAACAGATAGAAATTT TAATAC	<i>Clathria rugosa</i> cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria 1	KJ620406.1	0	96.87
SCS0013	CAATGTTGGGGGATGACCATCTATATA ATGTTATAATTACGGCTCATGGTCTTAT AATGATCTTTTTCTTAGTTATGCCAGTT CTGATAGGGGGGTTTGGTAATTGATTTG TTCCTCTTTATATCGGAGCCCCGGATAT GGCTTTTCCAAGATTAAATAATATTAGT TTTTGGGTTTTACCCCATCTTTAATATT ACTGTTAGGTTGAGCTTTTGTGAACAA GGGGTTGGAGCTGGATGAACTCTTTATC CTCCTTTATCGAGCGTTCAAGCTCATTC AGGGGGTTTCGGTAGATGCAGCAATATT TAGTCTTCATTTAGCAGGGATTCATCA ATTTTAGGGTCTATGAATTTTATTACTA CTATTTTAAATATGCGGGCCCCCTGGTAT TACAATGGATAGATTGCCCTTATTTGTT TGATCTATTTAATAACAACCTTATCTTTT AATATTATCTTTACCTGTATTGGCAGGG GCGATAACTATGCTTTTAAACAGATAGA AATTTTAAACAACCTTTTTTCGATCCAG CAGGTGGTGGGGATCCAATA	<i>Penares</i> sp1 voucher LIT66 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondria 1	ON130542.1	0	96.37

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0014	GATCAATGTTGGGGGATGACCATCTAT ATAATGTTATAATTACGGCTCATGGTCT TATAATGATCTTTTTCTTAGTTATGCCA GTTCTGATAGGGGGGTTTGGTAATTGAT TTGTTCTCTTTATATCGGAGCCCCGGA TATGGCTTTTCCAAGATTAAATAATATT AGTTTTTGGGTTTTACCCCATCTTTAAT ATTACTGTTAGGTTTCAGCTTTTGTGAA CAAGGGGTTGGAGCTGGATGAACTCTT TATCCTCCTTTATCGAGCGTTCAAGCTC ATTCAGGGGGTTTCGGTAGATGCAGCAA TATTTAGTCTTCATTTAGCAGGGATTTC ATCAATTTTAGGGTCTATGAATTTTATT ACTACTATTTTAAATATGCGGGCCCCCTG GTATTACAATGGATAGATTGCCCTTATT TGTTTGATCTATTTTAAATAACAATTAT CTTTTAAATATTATCTTTACCTGTATTGGC AGGGGCGATAACTATGCTTTTAAACAGA TAGAAATTTTAAATAACAATTTTTTCGAT CCAGCAGGTGGTGGGGATCCAATATTA TTTCAGCATTTATT	<i>Penares</i> sp1 voucher LIT66 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondria l	ON130542.1	0	96.32
SCS0015	TGGCGATGATCATTATATAACGTAATA GTAAGTGTACACGCCTTTGTAATGATT TCTTTCTGGTTATGCCTGTGATGATTGG GGGTTTCGGAAATTGATTTGTTCTTTA TATATTGGTGCGCCTGACATGGCTTTTC CTAGATTGAATAATCTTAGTTTTTGTATT ATTACCGCCCGCTTAACCTTATTATTA GGCTCAGCGTTTGTAGAACAAGGAGCT GGTACAGGTTGAACTGTTTATCCTCCTT TATCAAGTATACAAGCTCACTCCGGGG GATCGGTTGATATGGCAATTTTGTAGTCT TCATTTAGCCGGGATATCTTCTATATTG GGAGCAATGAATTTTATTACAACCATTA TAAATATGAGAGCGCCAGGTATTACAA TGGACCGAATGCCATTATTTGTGTGATC TGTTTTAGTAACTGCCATTTTATTATTGT TATCTTTACCGGTATTGGCCGGGGCTAT AACAATGCTTTTGACAGATCGAAATTTT AATACTGCG	<i>Halichondria</i> <i>panicea</i> voucher SB16A cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria l	KY492588.1	0	100

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0016	CAATGTTGGGAGATGATCATTATATATAA CGTTATTGTAAGTCTCATGCTTTTGTT ATGATTTTTTTTTTAGTTATGCCAGTAAT GATTGGGGGGTTTGGAATTGGTTTGTG CCTTTATATATTGGTGCCCCAGATATGG CTTCCCTCGATTAAATAATATAAGTTT TTGATTATTGCCCCCGGCTTTAAGCTTA TTATTGGCTTCTGCTTTTGTGGAGCAAG GGGCAGGAAGTGGGTGGACAGTTTATC CCCCCTTATCTGGTATTCAGGCCCATTC GGGGGGCTCAGTAGATTTAGTAATATT AGCTTACATTTGGCAGGGATTCTTCTA TATTGGGCGCTATGAACTTTATTACTAC AATAATAAACATGAGGGGGCCGGGTGT ATCTTTGGATAGAATGCCATTATTTGTT TGATCTATTTAGTAACTGCGGTTTTAT TATTATTATCTTTACCCGATTAGCTGG GGCCATAACTATGTTATTAACGGATAG AAATTTTAATACAGCTTTTTTTGATCCT GCGGGAGGAGGGGATCCTATTTTATAT CAGC	<i>Clathria pauper</i> mitochondria l partial COI gene for cytochrome oxidase subunit 1, specimen voucher NIWA37073 A	LN850182.1	0	96.98
SCS0017	TGCTAGGGGACGATCATTATATAACGT TCTTGTAAGTGCATGCTTTTGTAATG ATATTTTTTTTTTAGTTATGCCGTAATGA TTGGAGGGTTTGTTAACTGACTGGTGCC CTTGATATTGGAGCGCCTGATATGGCT TTTCCTCGATTAAATAATATAAGTTTTT GATTATTGCCCCCGGCTTTAAGTTTATT GTTGGCGTCTGCTTTTCGTGGAGCAAGGG GCAGGAACAGGATGAACTGTTTATCCC CCTCTCTCGGGTAATCAGGCTCATTCTG GCGGGTCAGTTGATTTAGTAATCTTTAG CCTACATTTAGCCGGAATTTCTTCCATT TTAGGCGCTATGAATTTTATTACTACAA TAGTAAACATGCGGATAACGGGGATGA CTTTTGATAGGTTGCCTTTATTTGTTTGA TCAATTTTAGTTACCGCGGTTTATTGTT ATTGTCCTTACCTGTATTAGCGGGGGCT ATTACCATGTTGTTAACCGACAGGAATT TTAACAC	<i>Timea</i> sp. G303973 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondria l	AY561968.1	0	100
SCS0018	n/a				

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0019	TTAGGGGATGATCATTTATATAATGTAA TAGTAACAGCTCATGCATTCGTAATGAT ATTTTTTTTAGTTATGCCAGTAATGATA GGGGGGTTTGGTAATTGGTTAGTACCAT TATATATTGGTGCTCCTGATATGGCTTT TCCTAGATTAAATAATATTAGTTTTTGG TTGTTACCTCCGGCGCTAACTTTATTGT TAGGTTCTGCTTTTGTGGAACAAGGGGC TGGTACAGGTTGAACGGTTTATCCGCCT TTATCTAGTATTCAAACCTATTCTGGGG GATCTGTGGACATGGCGATATTTAGTCT TCATTTAGCAGGAATCTCTTCTATATTG GGTGCCATGAATTTTATAACAACAATTT TTAATATGAGGGCACCGGGGATAACTA TGGATAGAATACCTTTATTTGTTTGATC GGTTTTAGTAACTGCTTTTTTATTATTAT TATCATTGCCAGTATTAGCCGGTGCTAT AACAATGCTTTTAACCGATCGGAACCTT AATACTA	<i>Haliclona</i> sp. UPDM SZN- HL114 isolate HL114 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondria l	MT226342.1	0	96.67
SCS0020	AATGTTGGGGGATGACCATCTTTATAAT GTTATAATTACTGCTCATGGTCTTATAA TGATTTTTTTCTTGGTTATGCCGGTTATG ATTGGGGGGTTTGGTAATTGATTTGTCC CTCTCTATATTGGGGCCCCAGATATGGC TTTTCCAAGATTAAATAATATTAGTTTT TGAGTGTTACCTCCCTCATTAAATATTAT TATTAGGCTCAGCTTTTGTGTAACAGGG GGTTGGTGCCGGATGAACAATTTATCCT CCTTTATCGAGTGTGCAAGCTCATTCTG GGGGTTCGGTTGATGCGGCGATATTTAG TCTTCACCTGGCCGGGATTTCAATCAATC TTAGGGTCAATGAATTTTATTACTACTA TTTTTAATATGCGGGCACCTGGTATTAC GATGGATAGACTGCCTTTATTTGTTTGG TCAATTTTAGTAACAACCTTATCTTTTAA TATTATCTTTACCTGTTTTGGCAGGTGC TATAACTATGCTTTTAAACAGACAGAAAT TTTAATACAACCTTTTTCGACCCGGCAG GCGGAGGGGACCCAATATTATTTTCAGC A	<i>Stelletta dorsigera</i> voucher MNHN:Porif era, unaccessione d cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondria l	HM592750.1	0	99.11

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0021	GTGGAGCCGGAACCTCTTCAATCAACTC ACCACGATGCACGGCTTGGTGATGGTCT TCGGGGCGATTATGCCGGCGTTTACCGG CCTCGCTAATTGGCAGGTGCCGATGATG ATCGGCGCCGCCGATATGGCTTTTGCAC GAATGAATAATTGGTCGTTTGGCTTTT GCCGCCGGCGGGCGATGCTGCTGATCGG CTCTTTCTTTACGGCAAGCGGAGCGACG GCGGCGGGCTGGACTCTCTACGCTCCTC TTTCCGTGCAGCAGGGAATCGGAATGG ATATGGCTATCTTCGCCATTCAATTTGAT GGGCGCATCATCCATTATGGGGGCGAT TAATATCATCACCACAATTATTAATATG CGCGCGCCCGATATGCCGCTATTAAGA ATGCCGCTGTTCTG	<i>Mycale acerata</i> cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondria	MG763898.1	3.00E-83	82.64
SCS0022	ATGTTAGGGGATGATCATTTATATAATG TGATAGTAACAGCTCATGCAATTTGTAAT GATATTTTTTTTAGTTATGCCAGTAATG ATAGGGGGGTTTGGTAATTGGTTAGTTC CACTATATATTGGTGCTCCTGATATGGC TTTTCCTAGATTAAATAATATTAGTTTTT GGTTGTTACCTCCAGCGCTTACTCTATT GTTAGGTTCTGCTTTTGTGGAACAAGGG GCTGGGACAGGTTGAACGGTTTATCCG CCTTTATCTAGTATTCAAACTCATTCTG GGGGATCTGTGGACATGGTGATATTTA GTCTTCATTTAGCAGGGATATCTTCTAT ATTGGGTGCTATGAACTTTATAACGACA ATCTTTAATATGAGGGCACCGGGGATA ACTATGGATAGAATGCCTTTATTTGTTT GATCTGTTTTAGTGACTGCCTTTTTATTA TTGTTATCATTGCCAGTATTAGCCGGTG CTATAACAATGCTTTTAACCGATCGAAA CTTTAATACTACTTTTTTTGATCCGGCTG GAGGGGGGGATCCAATATTATATCAAC AT	<i>Cladocroce</i> sp. 1 JV-2020 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondria	MT586740.1	0	98.57

Table 3.6.2.(Cont.) CO1 DNA barcoding results table.

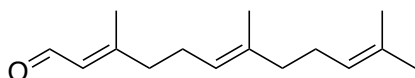
Specimen	CO1 Sequence	CO1 Blast hit	Sequence ID	E value	Percent Identity
SCS0023	TATTAGAAATAGACATTTATATAATGCT ACTGTTACAGCTCATGCTTTGGTTATAA TTTTTTTTTTTGTAATACCAACTATAGTA GGTGGATTTCGGAATGATTGTTACCCT TGATAATTGGAGCCCCTGATATAGCATT TCCTCGATTAAATAATTTAAGTTTTGA TTATTACCTCCTGCATTGATGTTTTAAT TGCTTCAATACTAATAATCCTGGAATA GGTACTAGATGAACTATTTATCCTCCTT TATCTTCTAATATTGCACATTCTAAACC TTCGGTTGATTTAGGTATTTTATCTTTAC ATATTGCGAGAGCATCTAGCATTATGG GTTCTATTAATTTTTTGGTTACAATAAG AAATATAAAATATAAAAGAAGAAAAAT AATAAACTTATCATTATTTTGTTGAGCT ATTTTAGTGACTGCTATATTATTAGTTC TTTCTCTTCCTGTATTAGCTGCAGCTATT ACAATGTTATTATTTGATCGTAATTTTA ATAGT	<i>Lissoclinum fragile</i> clone 07Jun08-1-1 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondria 1	JF506185.1	0	89.8
SCS0024	n/a				

Section 3.7. Synthesis of terpenoid precursors

A few aspects of this project required the synthesis of terpenoid precursors, specifically GPP and FPP. In the beginning of the project, I prepared deuterated FPP for sponge cell free assays (see Section 3.8). I later synthesized GPP to assay the sponge terpene synthases for monoterpene synthase activity. Dr. Shaun McKinnie and Dr. Immo Burkhardt both provided valuable synthetic expertise for the preparation of these precursors. This aspect of the project helped me build synthetic chemistry skills I first learned in my undergraduate research laboratory.

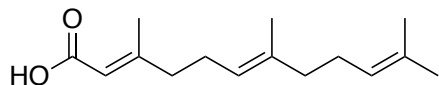
Section 3.7.1. Deuterated farnesyl pyrophosphate synthesis

Farnesal:



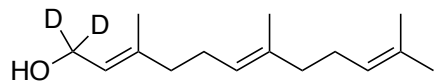
To a stirred mixture of farnesol (2.00 g, 8.99 mmol) in 200 mL CH_2Cl_2 was added Dess-Martin periodinane (5.72 g, 13.49 mmol), and the resulting mixture was stirred at room temperature for 1 hour. The reaction mixture was concentrated under reduced pressure and 20 mL of hexane was added. After precipitation of a white solid, the solution was filtered and transferred to a clean conical flask. The white solid was rinsed with hexane (20 mL x 2) and the resulting solution was filtered and transferred to the same conical flask. The combined hexane solutions were concentrated under reduced pressure. The crude product was purified via silica column chromatography (9:1 hexane:ethyl acetate) to yield farnesal as a clear, slightly yellow oil (1.265 g, 63%).

Farnesoic acid:



Farnesal (1.763 g, 8.00 mmol) was added to stirred tert-butanol (64 mL) and 2-methyl-2-butene (16 mL) at room temperature. Sodium chlorite (6.512 g, 72 mmol) and sodium dihydrogen phosphate (6.720 g, 56 mmol) in 28 mL of water and added dropwise over 15 minutes. The resulting mixture was stirred at room temperature for 44 hours. The reaction mixture was concentrated under reduced pressure to remove volatiles and 80 mL of water was added. The solution was acidified with 2M HCl to pH 3 and extracted with hexane (3 x 75 mL). The combined organic extracts were dried with magnesium sulfate, filtered, and concentrated under reduced pressure. The crude product was purified via silica column chromatography (9:1 hexane:ethyl acetate) to yield farnesoic acid (1.1000 g, 58%).

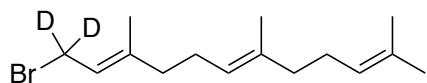
[D₂]-Farnesol:



Farnesoic acid (0.827 g, 3.5 mmol) was added to dry tetrahydrofuran (10 mL) with stirring under argon gas. A slurry of lithium aluminum deuteride (0.294 g, 7 mmol) was added slowly under argon gas at 0 °C and the reaction was stirred for 1 hour. Wet diethyl ether (30 mL) was added slowly and the reaction was stirred at room temperature for 10 minutes. Saturated

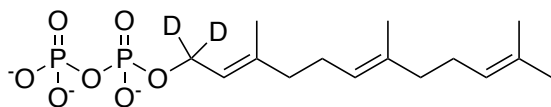
ammonium chloride (30 mL) was added and the resulting mixture was extracted with diethyl ether (30 mL x 3). The combined organic extracts were dried with magnesium sulfate, filtered, and concentrated under reduced pressure. The crude product was purified via silica column chromatography (9:1 hexane:ethyl acetate) to yield [D₂]-farnesol (0.2971 g, 38%).

[D₂]-Farnesyl bromide:



PBr₃ (0.029 mL, 0.3 mmol) was added dropwise to [D₂]-Farnesol (.127 mL, 0.5 mmol) in dry methylene dichloride (5 mL) with stirring, under Argon gas, at -30°C. The reaction was stirred at -30°C for 1 hour. The reaction mixture was then added to a stirring solution of saturated sodium bicarbonate (5 mL) at 0°C. The flask was allowed to come to room temperature and an additional 5 mL of saturated sodium bicarbonate was added. The aqueous layer extracted with ethanol (2 x 10 mL) and all combined organic extracts were dried with magnesium sulfate, filtered, and concentrated under reduced pressure. The crude product was carried forward to the next reaction without further purification.

[D₂]-Farnesyl pyrophosphate:



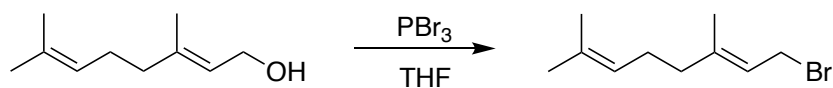
[D₂]-Farnesyl bromide (0.5 mmol) was added dropwise to a stirring solution of tris (tetra-n-butylammonium) hydrogen pyrophosphate (0.958 g, 1.0 mmol) in dry acetonitrile (10 mL) under argon gas at -30°C. The reaction was stirred at -30°C for 10 minutes before it was allowed to warm to room temperature and stir for 2.5 hours. The solvent was removed *in vacuo* and the resulting yellow oil was resuspended in minimal ion exchange buffer (25 mM ammonium bicarbonate in 2% isopropyl alcohol and 98% water) before being loaded onto an ion exchange column (DOWEX AG 50W-X8, NH₄⁺ form, 18 mL column volume). The first two column volumes (36 mL total) were combined and lyophilized for 22 hours to yield a fluffy white solid. The white solid was dissolved in 50 mM ammonium bicarbonate (2 mL), then 1:1 isopropyl alcohol/acetonitrile (8 mL) was added. The solution was vortexed, centrifuged (2000 rpm, 5 min), and the supernatant was transferred to a conical vial. The precipitate was resuspended in 50mM ammonium bicarbonate (2 mL) and 1:1 isopropyl alcohol/acetonitrile (8 mL) and the entire centrifugation, supernatant removal, and precipitate resuspension procedure was repeated 4 times. Volatiles in the pooled supernatants were removed *in vacuo* and the remaining cloudy aqueous liquid was lyophilized for 24 hours, giving a fluffy white solid (0.1604 g, 74%).

Section 3.7.2. Geranyl pyrophosphate synthesis

Preparation of tris(tetrabutylammonium) hydrogen pyrophosphate trihydrate (63):

To an approximately 100mL column of Dowex AG 50W-X8 cation exchange resin was added 1 column volume of 0.25M H₂SO₄. The column was washed with DI H₂O until the eluent had a pH of approximately 5. Disodium dihydrogen pyrophosphate (3.13 g, 14 mmol) was dissolved in 25 mL of DI H₂O containing 1 mL of concentrated ammonium hydroxide. The resulting clear solution was loaded onto the prepared ion exchange column and eluted with DI H₂O until the eluent had a pH of approximately 5 (~150mL eluent collected in a 250mL round bottom flask). A stir bar was added and then the solution was titrated to pH 7.4 with tetrabutylammonium hydroxide. The solution was flash frozen in the round bottom flask by rotating the flask using a rotary evaporator arm over a dewar of liquid nitrogen. The uniformly frozen solution was lyophilized overnight (18 hours) to yield a very hygroscopic fluffy white powder. The powder was immediately transferred to tared glass vials, which were flushed with argon, covered with a plastic screw top lid, parafilm, and stored at -20° C.

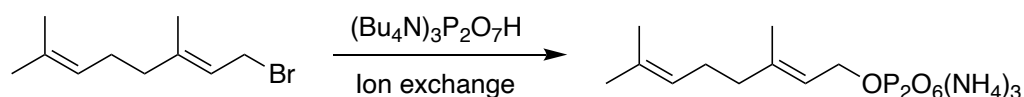
Geranyl bromide:



To a stirred mixture of geraniol (1.00 g, 6.49 mmol, 1.126mL) in 15mL dry THF at -78° C was added PBr₃ (1.23 g, 4.55 mmol, 0.43 mL) dropwise. The resulting mixture was stirred at -78° C for 1 hour, then stirred at room temperature for 20 mins while the reaction was monitored by TLC (4:1 hexanes: ethyl acetate). The reaction was quenched with approximately 30 mL ice water

and extracted with pentane (2 x 25 mL). The mixture was dried with MgSO₄, filtered into a 250 mL round bottom flask, and concentrated under reduced pressure, yielding a pale-yellow oil (1.3 g crude weight, ~4.88 mmol product as estimated by NMR).

Trisammonium geranyl diphosphate (63):



Half (0.65g) of the crude geranyl bromide product (~2.44 mmol) was removed from the 250mL reaction flask. The flask was stopped with a rubber septum and flushed with a continuous stream of argon. Tris(tetrabutylammonium) hydrogen pyrophosphate trihydrate (3.715 g, 3.25 mmol) was dissolved in 10mL dry acetonitrile and was slowly added to the 250 mL flask via syringe. The reaction was stirred at room temperature overnight (17 hours).

Next, the ammonium form of the resin was generated by washing the Dowex AG 50W-X8 resin column with 3 column volumes of 5M ammonium hydroxide. The resin was washed with DI H₂O until the pH of the filtrate dropped to pH 5, then the resin was washed with ~100mL ion-exchange buffer (25mM ammonium bicarbonate in 98% DI H₂O / 2% isopropanol, prepared immediately prior to use). Meanwhile, the reaction mixture was carefully concentrated under reduced pressure to yield a cloudy pale-yellow oil that formed waxy crystals as it cooled to room temperature. The waxy residue was dissolved in 5 mL of ion-exchange buffer, and the resulting solution was loaded onto the prepared ion-exchange column. The flask was rinsed with ion-exchange buffer (2 x 5mL) and both washes were loaded into the column before elution with 1.5 column volumes of ion-exchange buffer. The eluant was collected in a 500-mL round bottom flask

and was flash frozen by rotating the flask using a rotary evaporator arm over a dewar of liquid nitrogen. The uniformly frozen solution was lyophilized overnight (18 hours) to yield a slightly yellow solid.

The solid was dissolved in 5 mL of 0.05 M ammonium bicarbonate and transferred to a 50-mL Falcon tube. Then 20 mL of 1 : 1 (v/v) acetonitrile : isopropyl alcohol was added, the contents were vortexed thoroughly, and a white precipitate formed. The mixture was centrifuged for 5 min at 2000 rpm. The supernatant solution was removed with a pipette, and residue was suspended in 5 mL of 0.05 M ammonium bicarbonate, and the process was repeated. The suspension was transferred to a 250mL round bottom flask, concentrated under reduced pressure, and was redissolved in ~100mL freshly-prepared ion-exchange buffer. The solution was flash frozen by rotating the flask using a rotary evaporator arm over a dewar of liquid nitrogen. The uniformly frozen solution was lyophilized for 48 hours to yield a fluffy white solid. The powder was transferred to a tared screw-cap glass vials, which were flushed with argon, closed with a plastic screw top lid, parafilm, and stored at -20° C.

Section 3.8. Cell free extract experiments

In the beginning of the project, before we sequenced the metagenome and transcriptome of the Bubarida sponge, we conducted cell-free extract experiments to investigate the biosynthesis of nitrogenous terpenes in the sponge. These experiments had some challenges and did not yield much valuable information about the biosynthesis of nitrogenous terpenes, but still provided valuable laboratory skills and experimental design experience. The presence of endogenous nitrogenous terpenes in the cell free extracts made it difficult to determine the

production of new nitrogenous terpene metabolites, which precipitated our synthesis of deuterated FPP to differentiate new nitrogenous terpene production from endogenous terpene presence. All of the CFE experiments used frozen sponge tissue that has been stored at -80°C for at least a month, which may have affected the activity of the enzymes in the samples. We also attempted to supplement our cell free extracts with various metals and cofactors, which we hypothesized could enhance enzyme activity, however no noticeable enhancement was observed. We were also fairly limited with sample quantity, so many of these experiments were performed on small scales which may have been too small to detect minor production of new nitrogenous terpenes. Once the metagenomic and transcriptomic data for the Bubarida sponge was generated, our focus shifted away from the cell free extract experiments, however, the new biosynthetic knowledge gained from our sponge terpene synthase discovery could inform future cell free extract experiments that may yield valuable information about the remaining biosynthetic steps. Cell free extract experiments with isotopically labeled terpene synthase products (bicyclogermacrene, alloaromadendrene, germacrene D) or isotopically labeled cyanide and thiocyanate could be especially valuable.

Frozen dry sponge was pulverized by mortar and pestle in liquid nitrogen. The homogenate was suspended in a lysis buffer solution (see Experiments A-D for buffer composition for each experiment). The suspension was centrifuged at 4 °C for 5 minutes at 2000 rpm. The supernatant was transferred to another tube and centrifuged at 4 °C for 5 minutes at 2000 rpm. 2 mL of crude supernatant was loaded onto a pre-packed PD-10 desalting column equilibrated with an elution buffer (see Experiments A-D for solution composition for each experiment). The first 3 mL of flow through was collected and used as cell free extract (CFE) in all following assays. Trace element solutions containing elements commonly used by biological enzymes were prepared with

the following metal salts at 0.5M concentration: iron(II) citrate, calcium chloride, zinc sulfate, potassium chloride, copper(II) chloride, nickel(II) sulfate, cobalt(II) chloride, and ammonium iron(II) sulfate. 100 μ L of each stock solution was added to 10 mL of water to make the combined trace element solution (TES) used in all following cell free extract experiments. 1 M aqueous stock solutions were prepared of the nitrogenous functional group salts potassium cyanate, potassium cyanide, and potassium thiocyanate. 20 μ L of each stock solution was added to 1mL of water to make the nitrogenous functional group solution (NFGS) used in all cell free extract experiments.

Experiment A: CFE was prepared using a lysis buffer solution of aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, and 5 mM EDTA (pH 8.0) and an elution buffer solution of aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, and 10 mM MgCl₂. Farnesyl pyrophosphate ammonium salt (FPP) was synthesized and graciously donated by Darlene Nguyen and Shaun McKinnie. Samples 1-6 were prepared as described in the table below (Table 3.8.1). After incubating for 30 minutes at room temperature in small glass test tubes, 100 μ L of hexane was added to each tube. The tubes were vortexed, centrifuged, and the hexane layer from each tube was transferred to a glass vial for gas chromatography mass spectrometry (GCMS) analysis.

CFE Experiment A, for which aqueous buffer solutions were used and FPP was the only substrate added, did not show any measurable substrate modification by GCMS analysis. Little to no difference in intensity was seen between the nitrogenous sesquiterpene peaks in the crude sponge homogenate supernatant, which had not been run through a size exclusion column, and those in the size exclusion column-treated CFE; this suggests the size exclusion column failed to

remove the latent nitrogenous sesquiterpenes in the supernatant. To increase the solubility of the latent nitrogenous sesquiterpenes, glycerol or CHAPS detergent was added to the buffer solutions in CFE Experiment B.

Table 3.8.1. Cell free extract experiment A sample composition

Cell Free Extract Experiment A					
Sample Number	CFE (μL)	Crude supernatant (μL)	Elution buffer (μL)	50mM FPP in elution buffer (μL)	12.5 mM FPP in 25% DMSO, 75% elution buffer (μL)
1	400	0	0	0	0
2	400	0	0	8	0
3	400	0	0	0	24
4	0	200	200	0	0
5	0	0	400	8	0
6	0	0	400	0	24

Experiment B: Glycerol CFE was prepared (as described above in the Cell free extract preparations section) using a lysis buffer solution of 20% glycerol/80% aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, and 5 mM EDTA (pH 8.0) and an elution buffer solution of 20% glycerol/80% aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, and 10 mM MgCl₂. CHAPS CFE was prepared using a lysis buffer solution of aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, 5 mM EDTA, and 6.5 mM 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS detergent) (pH 8.0) and an elution buffer solution of aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, 10 mM MgCl₂, and 6.5 mM 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS detergent). Samples 1-6 were prepared as described in the table below (Table 3.8.2). After incubating overnight at room temperature in small glass test tubes, 200μL of hexane was added to

each tube. The tubes were vortexed, centrifuged, and the hexane layer from each tube was transferred to a glass vial for GCMS analysis.

There was a significant reduction in intensity of the nitrogenous sesquiterpene peaks between the glycerol crude supernatant and the Glycerol CFE (Fig. 3.8.1). There was no significant reduction seen in the CHAPS buffer preparations. There were some differences seen between the Experiment B m/z 204 EICs for the Glycerol CFE (GCFE) control with no added substrate and the GCFE incubated overnight with FPP (Figure 3.8.2.). The intensity of the 9.55 minute peak seen in both EICs is higher in the GCFE+FPP sample and there are several peaks between 9.45 and 9.65 minutes that appear in the GCFE+FPP sample that were not present in the GCFE control sample.

Table 3.8.2. Cell free extract experiment B sample composition

Cell Free Extract Experiment B			
Sample Number	Glycerol CFE (μL)	CHAPS CFE (μL)	50mM FPP in elution buffer (μL)
1	400	0	0
2	400	0	8
3	400	0	8
4	0	400	0
5	0	400	8
6	0	400	8

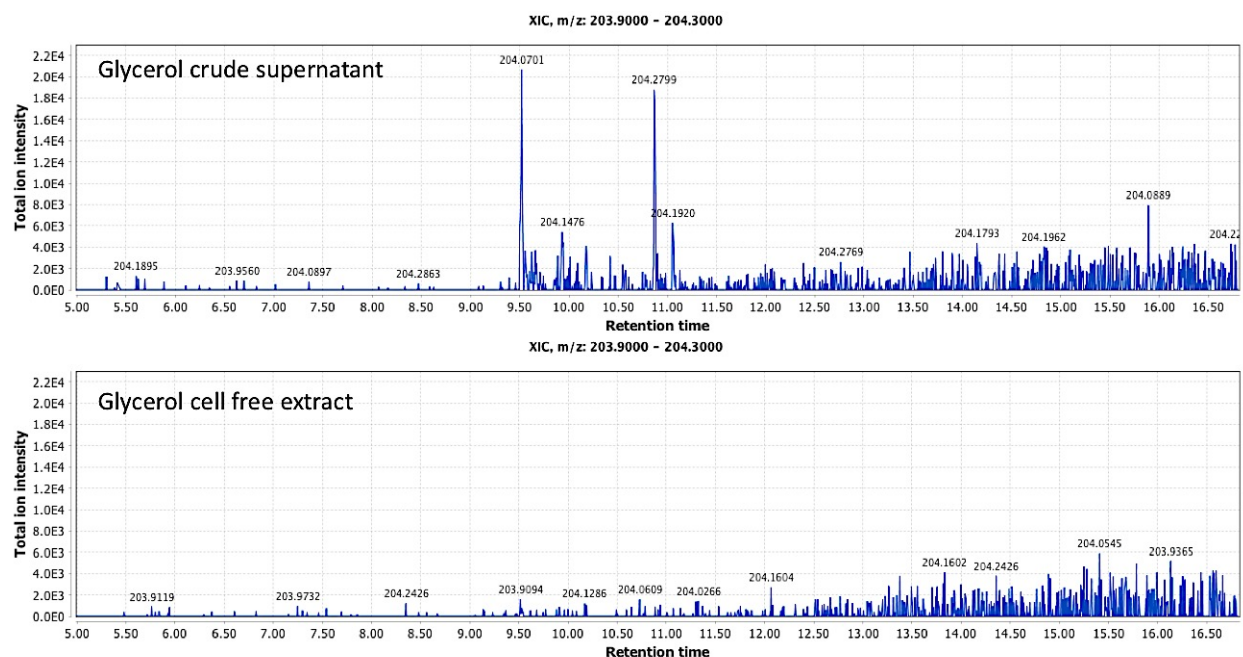


Figure 3.8.1. GCMS $m/z=204$ EICs for CFE experiment B, showing the difference between the crude sponge homogenate supernatant and the size exclusion column-treated CFE, both prepared with 20% glycerol buffers. The difference in intensities between the EICs for the crude and the CFE show that the CFE processing removed much of the signal from the latent sesquiterpenes.

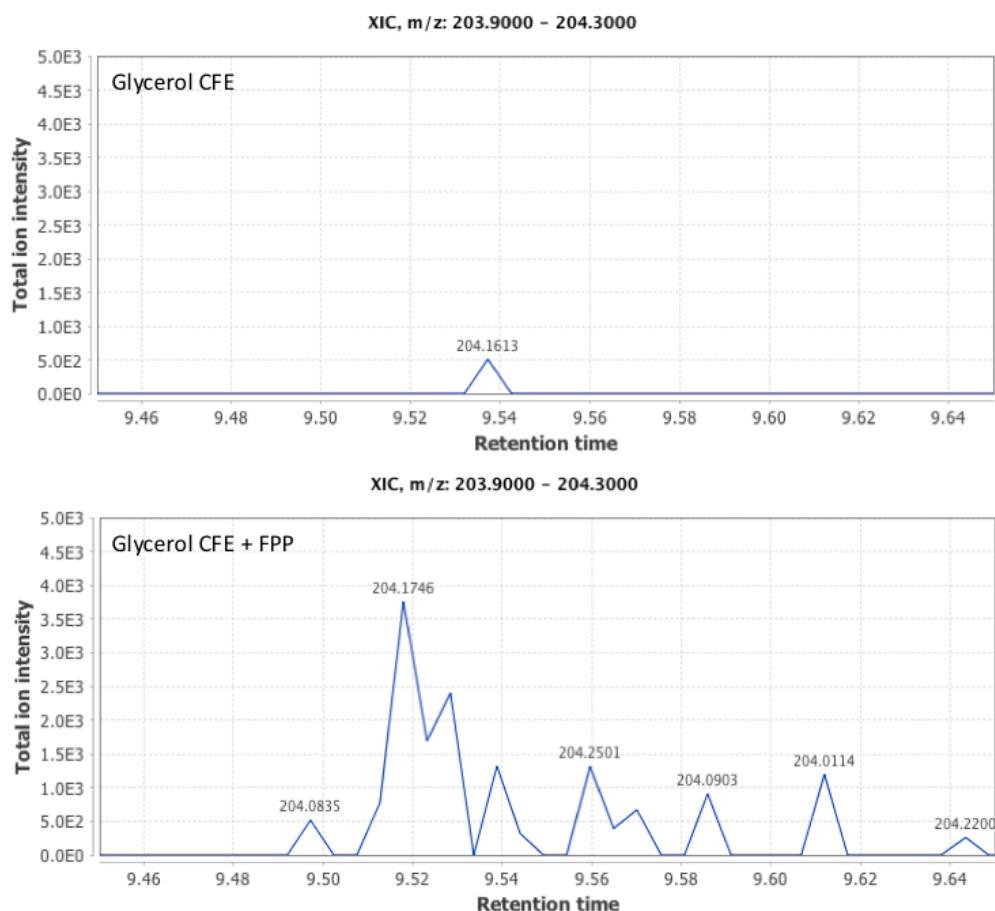


Figure 3.8.2. GCMS $m/z=204$ EICs for CFE experiment B showing the difference between the glycerol CFE without and with FPP added for the retention time 9.45-9.65 mins.

Experiment C: Glycerol CFE was prepared using a lysis buffer solution of 20% glycerol/80% aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, and 5 mM EDTA (pH 8.0) and an elution buffer solution of 20% glycerol/80% aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, and 10 mM $MgCl_2$. CHAPS CFE was prepared using a lysis buffer solution of aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, 5 mM EDTA, and 6.5 mM 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS detergent) (pH 8.0) and an elution buffer solution of aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, 10 mM $MgCl_2$, and 6.5 mM 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate

(CHAPS detergent). Samples 1-6 were prepared as described in the table below (Table 3.8.3). After incubating for 1 hour at 0°C in small glass test tubes, 200µL of hexane was added to each tube. The tubes were vortexed, centrifuged, and the hexane layer from each tube was transferred to a glass vial for GCMS analysis.

Experiment C was run with the same glycerol CFE (GCFE) as experiment B, however the increase in intensity of the 9.45-9.65 minute peaks seen in experiment B was not observed in experiment C. It should be noted that the cell free extracts did sit at 4°C overnight between Experiments B and C and Experiment C was run for 1 hour at 0°C while Experiment B was run overnight at room temperature, which may have contributed to the difference in results.

Table 3.8.3. Cell free extract experiment C sample composition

Cell Free Extract Experiment C			
Sample Number	Glycerol CFE (µL)	CHAPS CFE (µL)	25mM FPP in elution buffer (µL)
1	400	0	0
2	400	0	16
3	400	0	16
4	0	400	0
5	0	400	16
6	0	400	16

Experiment D: Glycerol CFE was prepared using a lysis buffer solution of 20% glycerol/80% aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, and 5 mM EDTA

(pH 8.0) and an elution buffer solution of 20% glycerol/80% aqueous 50 mM Tris buffer, 100 mM NaCl, 1 mM dithiothreitol, and 10 mM MgCl₂. Fractionated sesquiterpenes were prepared by running sponge hexane extraction through a Pasteur pipette silica gel column equilibrated with 100% hexane, collecting the first 3 mL and removing the solvent in vacuo. Commercial β -caryophyllene (Tokyo Chemical Company, Nihonbashi-honcho, Chuo-ku, Tokyo) was added as described in the table below. Purified epi-isozizaene synthase was graciously donated by Christopher Eiben at UC Berkeley. Samples 1-13 were prepared as described in the table below (Table 3.8.4). After incubating for 1.5 hours at room temperature in small glass test tubes, 200 μ L of hexane was added to each tube. The tubes were vortexed, centrifuged, and the hexane layer from each tube was transferred to a glass vial for GCMS analysis.

Experiment D exclusively used 20% glycerol buffer solutions, features FPP as a substrate, and also incorporated a trace element solution to provide any potentially necessary metal cofactors, a solution of nitrogenous functional group derivatives to investigate functionalization activity, and fractionated sesquiterpenes isolated from the host sponge to investigate functionalization of the cyclized sesquiterpene scaffolds. Additionally, to some samples was added a known terpene cyclase to determine if the FPP could be utilized as a substrate and a commercial cyclized sesquiterpene to determine if functionalization of a non-native sesquiterpene scaffold was seen. No significant differences in intensity were seen in any peaks associated with functionalized or unfunctionalized sesquiterpenes. In sample 13, which contained epi-isozizaene (the known terpene cyclase) and FPP, production of epi-isoizazene was seen, which suggested the added FPP was being utilized as a substrate and that the epi-isozizaene enzyme was active.

Table 3.8.4. – Cell free extract experiment D sample composition

Cell Free Extract Experiment D									
Sample Number	Glycerol CFE (μL)	Crude supernatant (μL)	Elution buffer 20% Glycerol (μL)	25mM FPP in elution buffer (μL)	Trace Element Solution (μL)	Nitrogenous Functional Group Solution (μL)	Fractionated Sesquiterpenes (μL)	Caryophyllene (μL)	Purified Epi-isoizaene Synthase (μL)
1	400	0	163	16	6	15	0	0	0
2	400	0	178	16	6	0	0	0	0
3	400	0	149	0	6	15	30	0	0
4	400	0	155	0	0	15	30	0	0
5	400	0	194	0	6	0	0	0	0
6	400	0	200	0	0	0	0	0	0
7	0	200	379	0	6	15	0	0	0
8	0	200	400	0	0	0	0	0	0
9	0	200	349	0	6	15	0	30	0
10	0	0	549	0	6	15	30	0	0
11	0	0	549	0	6	15	0	30	0
12	0	0	563	16	6	15	0	0	0
13	0	0	379	16	0	0	0	0	5

Section 3.9. Sponge terpene synthase point mutations

After discovering sponge TSs, I wanted to investigate the structure-function relationship of these enzymes using mutagenesis. TSs have several conserved active site residues, whose function has been extensively studied in other TSs, so those active site motifs were clear targets for mutagenesis. While our sponge TSs had the typical conserved TS active site motifs, some sponge TS active site residues were slightly different from canonical TS residues. For example, canonical TSs also contain a metal-binding (N/D)DXX(S/T)XXX(D/E) region, commonly known as the NSD/DTE motif, but our sponge TSs feature an unusual NDXXGXXXD motif which aligns with the traditional NSD/DTE domain. While the active site moieties were obvious targets for mutagenesis, we also wanted to target other residues that might interact with the substrate. We have not obtained crystal structures of the sponge TSs so we utilized structure models to identify other target residues. This mutagenesis effort not only helped yield valuable information about the activity of these TSs, but it also served as a valuable training experience for my high school mentee, Amelie Hopke, who worked on this project during a summer research internship in our lab. This mutagenesis campaign helped to teach and reinforce key biological principles while also helping Amelie build important wet-lab skills such as pipetting, aseptic technique, mini-preps, bacterial culturing, PCR, gel electrophoresis, and more. Amelie contributed greatly to these mutagenesis efforts and also helped me gain valuable mentorship and teaching experience.

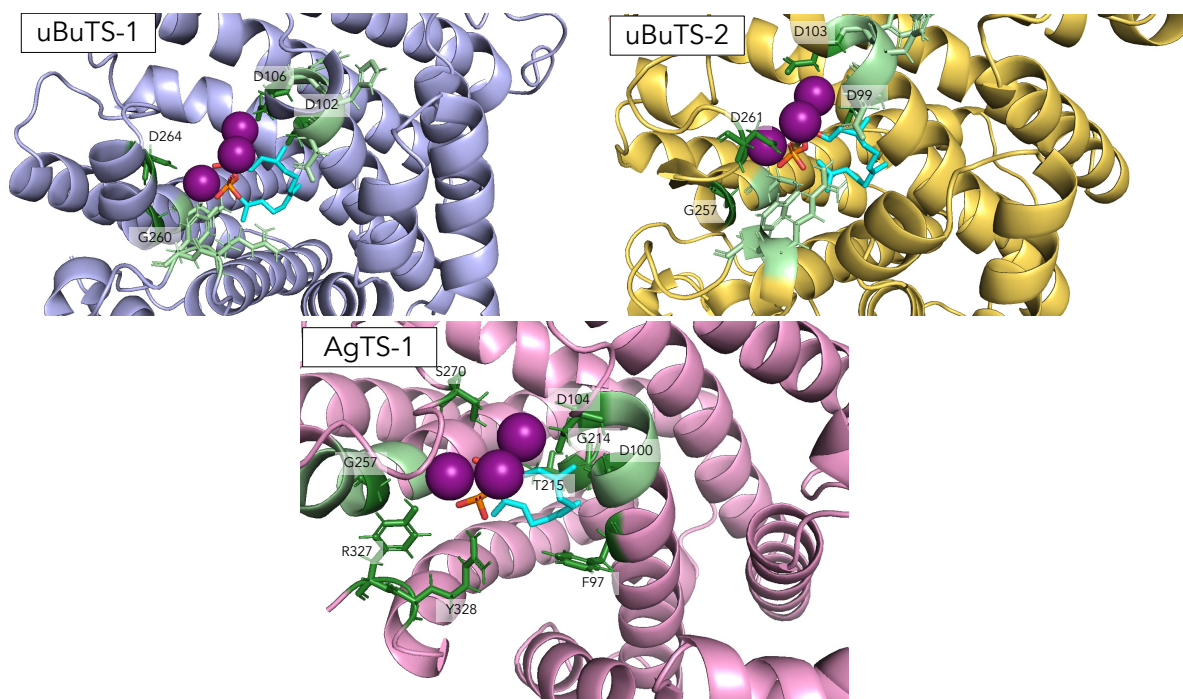


Figure 3.9.1. Robetta structure models of sponge TSs. Active site motifs are shown in light green and mutation target residues are shown in dark green. FPP is shown in turquoise and Mg^{2+} ions are shown in dark purple.

Using the Robetta server (64), we generated models of uBuTS-1, uBuTS-2, and AgTS-1, which we visualized in PyMOL. FPP from the crystal structure of selinadiene synthase (PDB: 4OKZ) was overlaid onto the sponge TS model structure (Fig 3.9.1). Mutagenesis primers were designed in Benchling and synthesized by IDT (Table 3.9.1). The template DNA for the mutagenesis PCRs were the pET-28a(+) clones of uBuTS-1, uBuTS-2, and AgTS-1. For each 20 μ L reaction, 10uL Phusion polymerase (NEB), 2 μ L 10 μ M oligonucleotide primer, 2 μ L template DNA, 0.6 μ L DMSO, and 5.4 μ L ultrapure water were combined. The PCR program for primers uBuTS-1_D102A, uBuTS-1_D106A, uBuTS-1_D264A, uBuTS-2_D261A is as follows: initial denaturation for 30 seconds at 98 °C, 30 cycles of denaturation for 10 seconds at 98°C, annealing for 30 seconds at 65°C, extension for 4 minutes at 72°C, and a final extension for 10 minutes at

72°C. The PCR program for uBuTS-1_D102A, uBuTS-1_D106A, uBuTS-1_D264A, uBuTS-2_D261A is as follows: initial denaturation for 30 seconds at 98°C, 30 cycles of denaturation for 10 seconds at 98°C, annealing for 30 seconds at 65°C, extension for 4 minutes at 72°C, and a final extension for 10 minutes at 72°C. The PCR program for all AgTS-1 mutants is as follows: initial denaturation for 30 seconds at 98°C, 30 cycles of denaturation for 10 seconds at 98°C, annealing for 30 seconds with a gradient from 57°C to 69°C, extension for 4 minutes at 72°C, and a final extension for 10 minutes at 72°C. Following the PCR cycles, 0.4µL of restriction enzyme DpnI (NEB) was added to each sample and allowed to digest for 1hr at 37°C. PCR products were visualized using agarose gel electrophoresis (0.7% w/v agarose gel, 118V 30mins) and immediately transformed into *E. coli* DH5α using the TSS protocol (65). The plasmid DNA was purified using a QIAprep Spin Miniprep kit (Qiagen) and the whole plasmids were sequenced by Primodium Labs (Monrovia, CA). The plasmid sequences were checked for proper mutagenesis, then the miniprep products were transformed into *E. coli* BL21 (DE3) using the TSS protocol (65). Protein purification, enzyme activity assays, and GCMS analysis were carried out as described in Wilson et al., 2023 (66) in the SI Appendix under sections *Heterologous expression for analytical scale*, *Analytical scale enzyme assays*, and *GCMS*.

Table 3.9.1. Mutagenesis oligonucleotide primers.

Primer Name	Primer Sequence	Annealing Temp (°C)	GC%
uBuTS-1_D102A	TTGTTCAATTTTGAATTTCGGACCAATTTTGATACCGCCGTCCGTCIGACT	69.2	41.18%
uBuTS-1_D106A	TTGTTCAATTTTGAATTTCAGCAACCAATTTTCGACCCGCCGTCCGTCIGACT	69.2	41.18%
uBuTS-1_G261S	CTGGCCAATGACGTAGCGAGCATGAATCGCGATCTGGAAAGATAACGAAAGCCATTGATA	73.3	50.00%
uBuTS-1_D264A	CGGGCATGAATCGCGCGCTGGAAAGATAACGAAAGCCATTGATATCAGGAGTCTGAAAAA	71.0	45.61%
uBuTS-2_D99A	GCACTCTTTAATTTTCGAAATTTGCGGATCACTTTTGATAAACCTTTACTGATTGCGCCGGAG	68.3	38.33%
uBuTS-2_D103A	GCACTCTTTAATTTTCGAAATTTGATGATCACTTTGCGAAACCTTTACTGATTGCGCCGGAG	68.3	38.33%
uBuTS-2_G257S	ATCATCGTCGCTTTGGCCAAACGATAATTCGTGAGCAATTCGCGC	72.5	54.76%
uBuTS-2_D261A	TCGCITTTGGCCAAACGATAATTCGTGGCAATTCGCGCGCTCATGGTGATGAT	74.8	54.00%
AgTS-1_F97Y	GCAACGTTGGGCAATCTTCTGTATGAATTCGACGACTACTTCGATAAAACCG	69.5	45.10%
AgTS-1_D100A	GCACTCTTCTGTTTGAATTCGCGGACTACTTCGATAAAACCGTCCAAACACGCC	70.0	46.15%
AgTS-1_D104A	GCACTCTTCTGTTTGAATTCGACGACTACTTCGCGAAACCGTCCAAACACGCC	70.0	46.15%
AgTS-1_G214A	AACGAGGTGCGTAAACGGTCCGGTGCCACCCCAATTCGCCCTTTGTGGTCCGTTATAT	75.3	52.63%
AgTS-1_T215S	AACGAGGTGCGTAAACGGTCCGGTGCGTCCCAATTCGCCCTTTGTGGTCCGTTATAT	75.3	52.63%
AgTS-1_G257S	GTAATCACGTTAGCTAACGACTTACTGAGCAAAAAAGACCTGCAAACGGAAAAATGAC	69.2	40.00%
AgTS-1_S270D	GAAAAATGACAGTATTGACATTAATAAAACAATCAAAATCGCAACCAACGTCAGA	65.3	34.00%
AgTS-1_327RYdel	GGTCTTTTCCCTGTGGCAATAATAATTCGTGAGTAACCTGAGCACCAC	67.8	43.75%

Due to time constraints, we selected one mutant from each enzyme to purify and assay: uBuTS-1_D106A and uBuTS-2_D261A. uBuTS-1_D106A targeted the first aspartate residue in the DDXXD metal-binding motif. This mutation was expected to kill catalytic activity based on the key catalytic role of this motif and previous TS mutagenesis studies that showed inactivation of similar mutants (67, 68). As expected, this mutation inactivated the enzyme and we saw no sesquiterpene production from the uBuTS-1_D106A mutant (Fig 3.9.2). uBuTS-2_D261A targeted the final aspartate residue in the metal-binding NDXXGXXD motif. Lin and Cane targeted the final residue of the NSE motif in epi-isozizaene synthase and they reported reduced catalytic activity and changes in product composition (67). We similarly observed reduced catalytic activity and changes in product composition compared to wildtype (Fig 3.9.2). The wildtype uBuTS-2 selectively produced germacrene D, while uBuTS-2_D261A produced germacrene D and another sesquiterpene product putatively identified as beta-cadinene based on comparison with the NIST GCMS library.

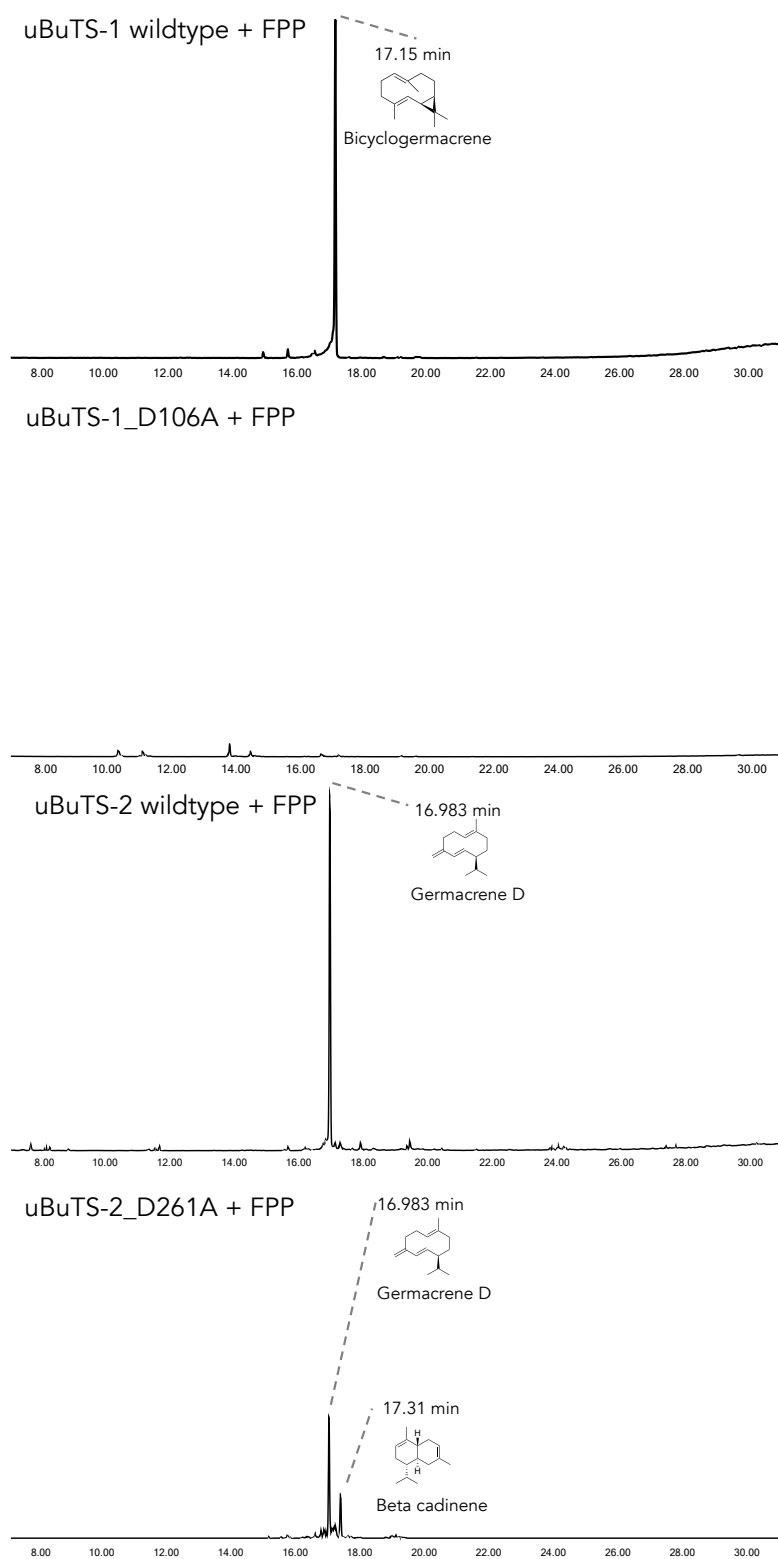
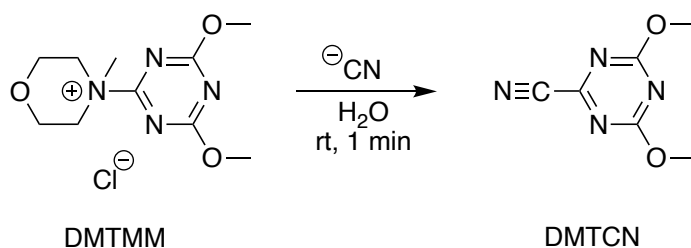


Figure 3.9.2. GCMS traces of wildtype and mutant TS enzyme assays.

Section 3.10. Sponge cyanide detection experiments

Given the demonstrated role of cyanide in nitrogenous terpene production, we decide to investigate the presence of cyanide in the nitrogenous terpene-producing San Diego Bubarida sponge. Mary Garson and Paul Scheuer's work showed that when nitrogenous terpene-producing sponges are fed cyanide, the sponges readily incorporate that cyanide into nitrogenous terpenes, however, to the best of our knowledge, only one publication reported the presence of cyanide in a sponge. Hamann, Scheuer, and Kelly-Borges reported a Hawaiian sponge from the family Aplysiniellidae emitted a strong cyanide odor when broken apart, and the presence of cyanide in this sponge was confirmed with commercial cyanide test strips (69). This sponge contained unusual meroterpenoids which featured cyano functional groups, as opposed to isocyano groups which are much more common in sponge terpenoids (69–71). These meroterpenoids seem quite biosynthetically distinct from most other sponge isonitrile terpenoids, which seem to originate from type I TSs and usually do not contain additional quinone-type moieties. This sponge is also geographically and phylogenetically quite distant from other common nitrogenous terpene producing sponges.

To test for the presence of cyanide in San Diego Bubarida sponge, we collected four sponges from God's Rock in La Jolla, CA and tested their tissue and the surrounding seawater for the presence of cyanide. Because of cyanide's small molecular weight, volatility, and lack of UV signal, we decided to detect cyanide by GCMS using the commercial coupling agent 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) using protocols described by Yamaguchi and Miyaguchi 2020 (72) (Scheme 3.10.1).



Scheme 3.10.1. DMTMM cyanide coupling reaction

The sponge collection was carried out at God's Rock (32°51'20.6"N 117°16'37.3"W) on April 25th, 2022 at approximately 23m depth and 14°C. Four sponges were collected in one plastic snap-top container at depth, then brought to the surface in the same container with the same seawater from the collection site. The sponges were transported to the laboratory in the same closed container and were processed approximately three hours after collection. The four sponges collected were all morphologically very similar, however upon closer visual inspection at the surface and GCMS analysis, it was determined that only one of the four sponges was the nitrogenous terpene-producing *Bubarida* sponge, and the other three sponges were another unidentified sponge that did not produce nitrogenous terpenes. The sponges were processed for the initial round of DMTMM experiments, then the remaining sponge tissue was placed in a small continuous-flow aquarium with filtered seawater from the Scripps Pier for 48 hours.

For DMTMM cyanide detection sample preparation, Borax (200 μ L saturated aqueous solution, for removal of hydration water of cyanide and salting out of DMTCN) and the sample (200 μ L) were added to a glass LCMS vial and then a DMTMM solution (100 μ L, 20 w/v %, $\sim$ 723 mM) was added. The mixture was vortexed for 1 min at 23°C, and then DBB solution (400 μ L, 10 μ M in EtOAc, internal standard) was added and the mixture was vortexed again. The mixture was centrifuged (3000g, 10 s at 23°C), the organic layer was carefully pipetted into a new glass LCMS

vial with a vial insert, and immediately analyzed by GCMS. A detailed sample description is shown below (Table 3.10.1).

GCMS analysis of the KCN standards (Samples 9-13) showed that this method was suitable for quantitative detection of cyanide in seawater (Fig. 3.10.1) It should be noted that the DMTMM-cyanide conjugate, DMTCN, was also detectible by GC-ECD (Fig. 3.10.1). DMTCN was detected in the surrounding seawater from the sponge collection site (Fig. 3.10.2). Based on peak area estimations from our standard curve, we estimate 0.005 $\mu\text{g/mL}$ cyanide was present in the seawater. The retention time and mass spectra for the seawater DMTCN peak matches that of the DMTCN peaks from our standard curve. The mass spectra also matches the mass spectra reported in the original Yamaguchi and Miyaguchi paper (ref) (Fig. 3.10.3)

Surprisingly, DMTCN was not detected in the homogenized sponge tissue sample. (Fig. 3.10.4). We believe there are several possible reasons for this. Perhaps the sponge had already released all available cyanide into the seawater in the 3 hours between sample collection and sample processing. It is also possible that there is another source of cyanide in the sponge's environment and the cyanide we saw in the surrounding seawater was produced by an organism other than the sponge. Additionally, given cyanide is a highly volatile gas, it is possible any cyanide in the sponge tissue was released into the air during the tissue homogenization process. Because we only found one Bubarida sponge at our collection site, we only had one biological replicate for these experiments so these results should be considered preliminary.

Table 3.10.1. Sample composition for the initial round of DMTMM cyanide detection experiments.

Sample name	Surrounding Seawater	Homogenized sponge tissue	DMTMM (20% w/v)	DBB (10umol in EtOAc)	EtOAc	Borax (saturated aqueous solution)	KCN (50mM)	KCN (5mM)	MQ water
Seawater sample with DMTMM coupling system	200uL	0uL	100uL	400uL	0uL	200uL	0uL	0uL	0uL
Seawater no DMTMM negative control	200uL	0uL	0uL	400uL	0uL	200uL	0uL	0uL	0uL
Seawater no DBB negative control	200uL	0uL	0uL	0uL	400uL	200uL	0uL	0uL	0uL
Seawater no DMTMM or DBB negative control	200uL	0uL	0uL	0uL	400uL	200uL	0uL	0uL	0uL
KCN-spiked seawater positive control	193.8uL	0uL	100uL	400uL	0uL	200uL	6.16uL	0uL	0uL
Sponge tissue sample with DMTMM coupling system	0uL	200uL	100uL	400uL	0uL	200uL	0uL	0uL	0uL
Sponge tissue no DMTMM or DBB negative control	0uL	200uL	0uL	0uL	400uL	200uL	0uL	0uL	0uL
KCN-spiked sponge tissue positive control	0uL	193.8uL	100uL	400uL	0uL	200uL	6.16uL	0uL	0uL
KCN standard curve 0.05mg/mL	0uL	0uL	100uL	400uL	0uL	200uL	0uL	153.6uL	846.4uL
KCN standard curve 0.1mg/mL	0uL	0uL	100uL	400uL	0uL	200uL	0uL	307.2uL	692.8uL
KCN standard curve 0.25mg/mL	0uL	0uL	100uL	400uL	0uL	200uL	0uL	768uL	232uL
KCN standard curve 0.5mg/mL	0uL	0uL	100uL	400uL	0uL	200uL	153.6uL	0uL	846.4uL
KCN standard curve 1mg/mL	0uL	0uL	100uL	400uL	0uL	200uL	307.2uL	0uL	692.8uL

MS Standard Curve

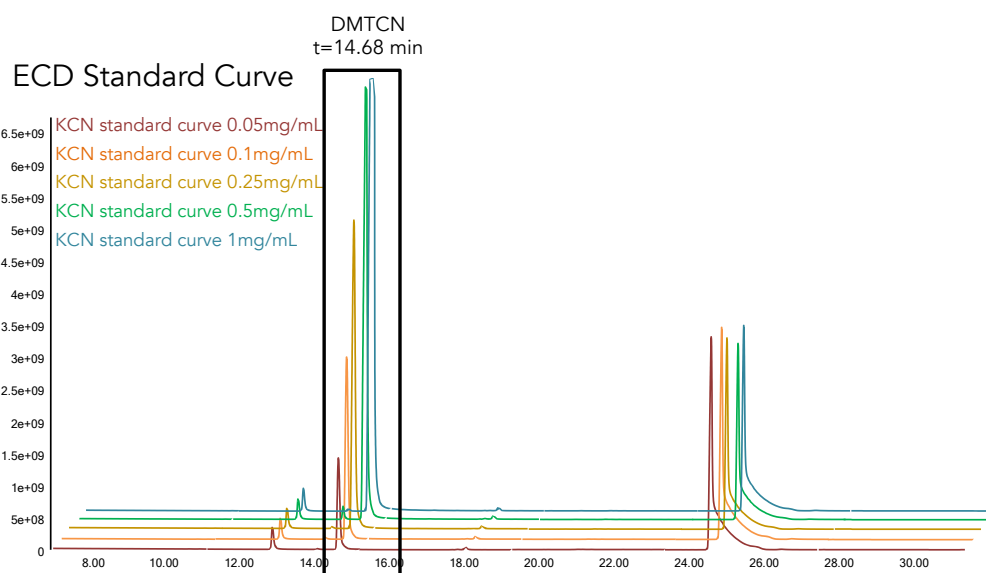
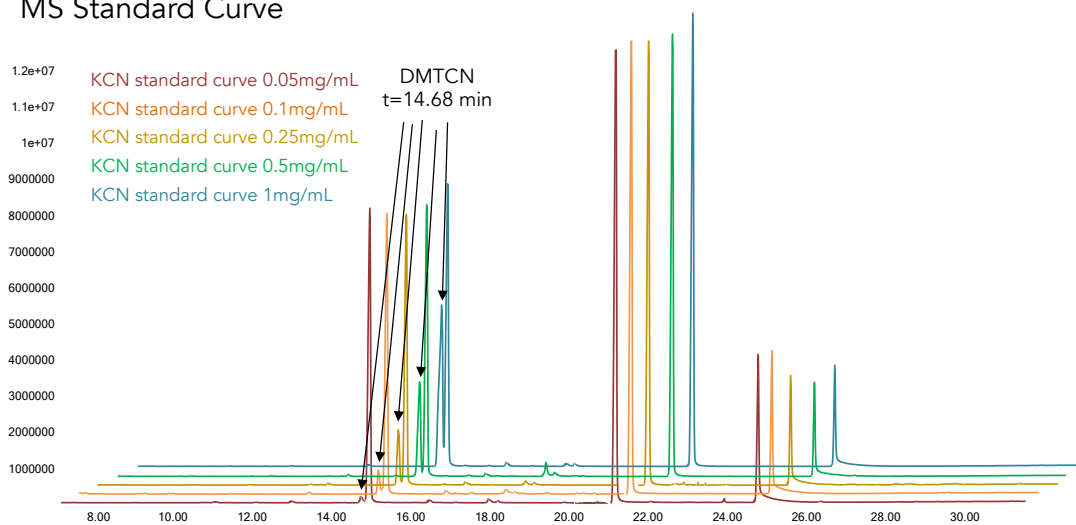


Figure 3.10.1. Standard curve of DMTMM detection of potassium cyanide (KCN) standard solutions from concentrations 0.05 mg/mL to 1 mg/mL.

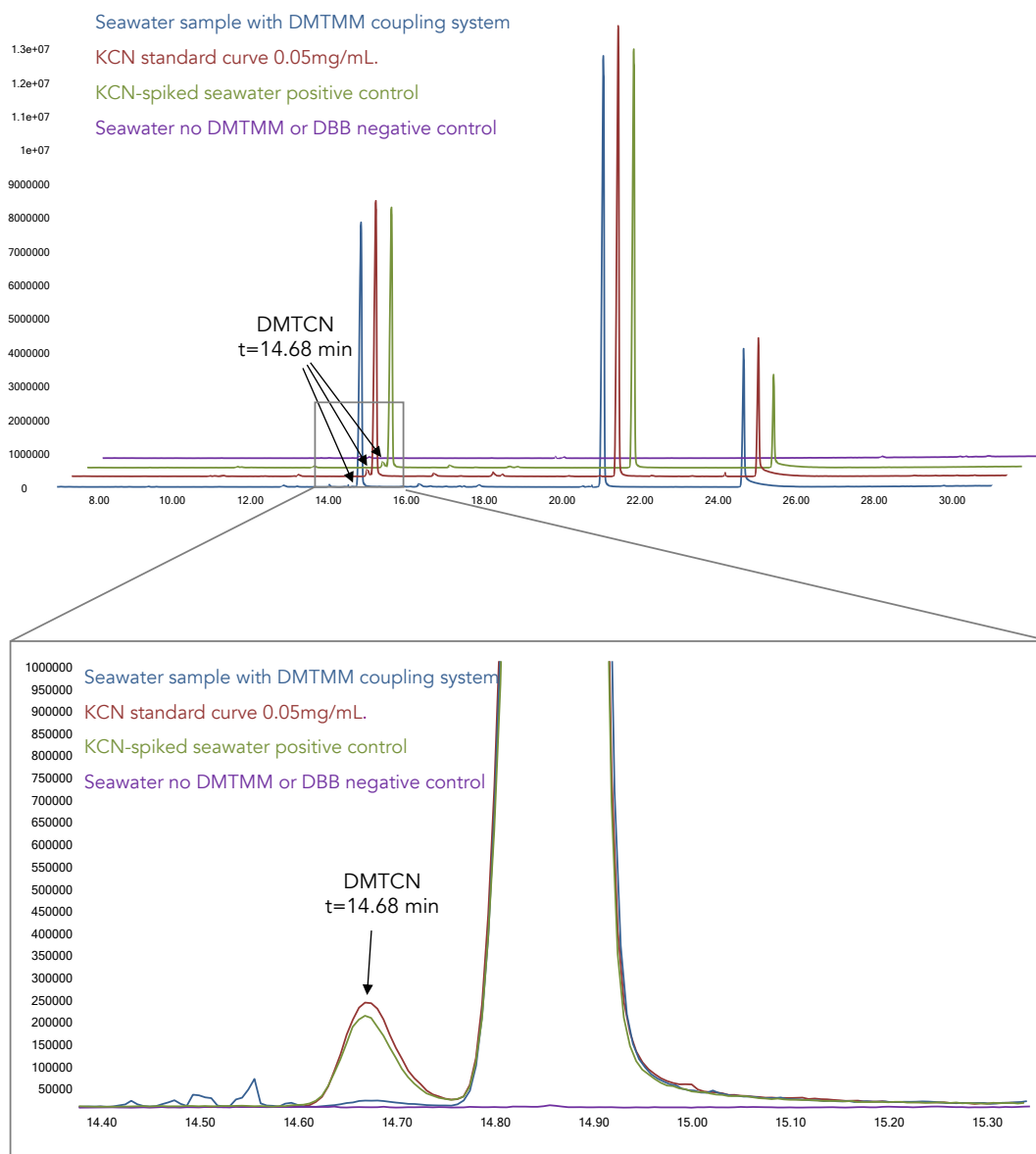


Figure 3.10.2. GCMS traces for DMTMM detection of cyanide in seawater. The peak at $t=14.68$ min for the sponge seawater sample (blue) is associated with the DMTMM-cyanide coupling product DMTCN and indicates the presence of cyanide in the sample. Peaks at the same retention time are shown in the KCN standard and positive control.

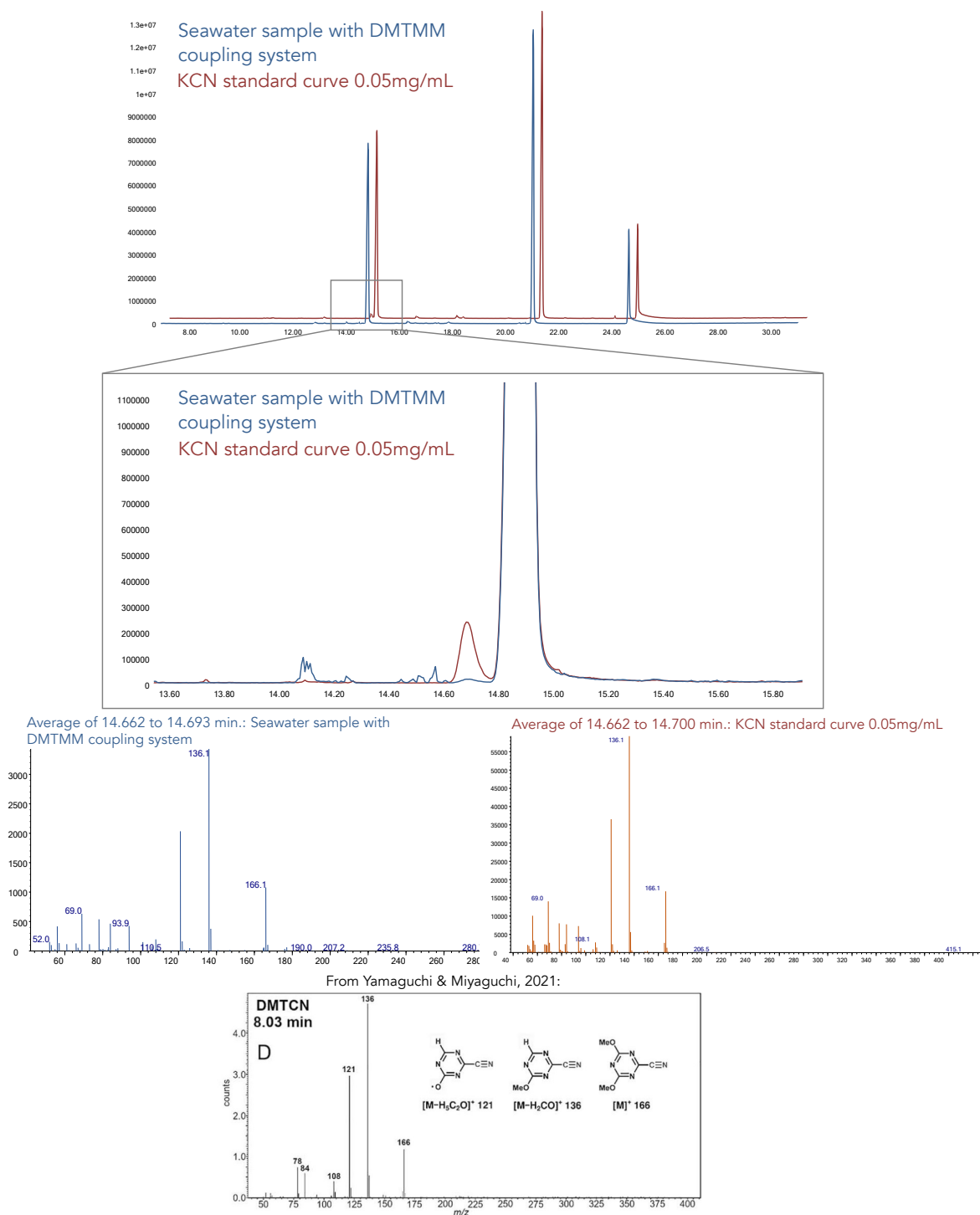


Figure 3.10.3. GCMS traces and MS spectra for DMTMM cyanide detection in seawater. The peak at $t=14.68$ min for the sponge seawater sample (blue) is associated with the DMTMM-cyanide coupling product DMTCN and indicates the presence of cyanide in the sample. A peak with the same retention time and MS spectra is shown in the KCN standard reported (red). The MS spectra for both the sponge seawater sample and the KCN standard match the spectra reported in the original DMTMM detection paper (72).

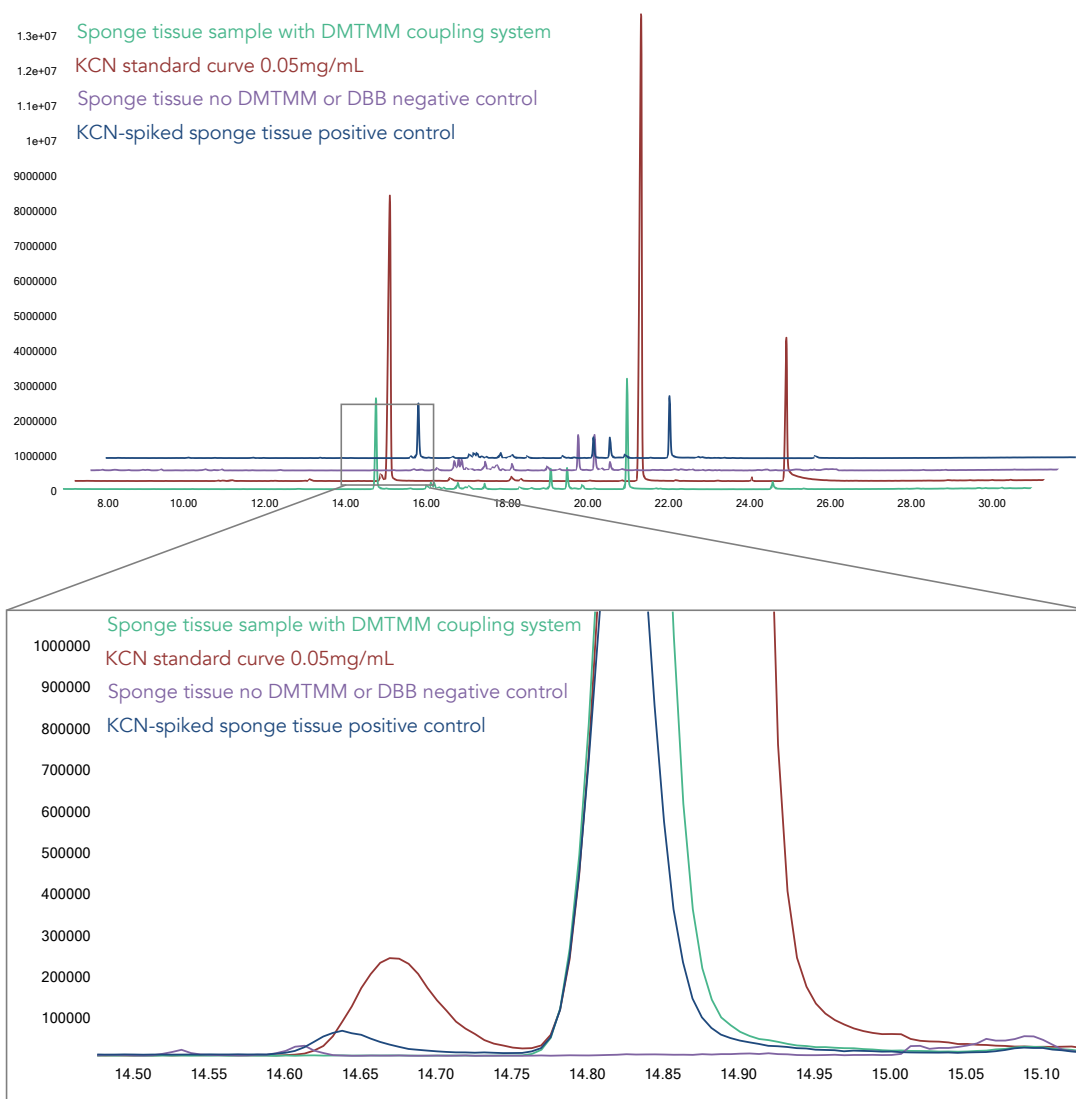


Figure 3.10.4. GCMS traces for DMTMM detection of cyanide in sponge tissue. No cyanide is detected in the sponge tissue sample (green).

Table 3.10.2. Sample composition for DMTMM cyanide detection time series experiments

Sample name	Surrounding Seawater	DMTMM (20% w/v)	DBB (10umol in EtOAc)
initial no sponge seawater sample	200uL	100uL	400uL
t=0hr sponge seawater sample	200uL	100uL	400uL
t=1hr sponge seawater sample	200uL	100uL	400uL
t=2hr sponge seawater sample	200uL	100uL	400uL
t=3hr sponge seawater sample	200uL	100uL	400uL

We were surprised that the sponge tissue did not contain detectable levels of cyanide, despite the surrounding seawater containing cyanide, so we decided to do another round of experiments with the remaining sponge tissue, which had been in a continuous-flow aquarium for 48 hours between the first and second rounds of DMTMM experiments. The remaining Bubarida sponge tissue was placed in a sealed 300 mL plastic screw-top jar filled with filtered seawater from the Scripps Pier. This container was not continuous-flow and the seawater was not changed out during the time course study. Immediately prior to the sponge being placed in the jar, an aliquot of the seawater was taken to evaluate base-line cyanide levels in the seawater. Aliquots were taken immediately after the addition of the sponge, and at 1 hour, 2 hours and 3 hour time points. The aliquots were then processed according to the Yamaguchi and Miyaguchi protocol as described above and analyzed by GCMS (72) (Table 3.10.2).

No DMTCN was detected in any of the time series samples (Fig 3.10.5). Given the sponge was stored in a continuous flow aquarium for 48 hours prior to this experiment, it is possible that any cyanide released by the sponge was flushed out of the aquarium before the start of the time series. The stress of collection, transportation, and tissue sampling may have caused the sponge to stop producing cyanide. It is also possible that the sponge is not the true producer of cyanide, and

that another organism from its natural environment is responsible for the cyanide detected in the seawater from the collection site. Again, it is noted that this experiment was only carried out with one sponge specimen and this experiment should be repeated with multiple biological replicates to make any conclusions about the source of cyanide.

Additional work should be carried out to investigate the presence of cyanide in the Bubarida sponge environment. Additional seawater samples should be taken *in situ* at the Bubarida collection site, at multiple distances from the sponge to see if there is a cyanide concentration gradient in the environment. Collection vessels should be sealed to prevent the release of cyanide gas and cyanide detection assays should be carried out immediately after sponge collection.

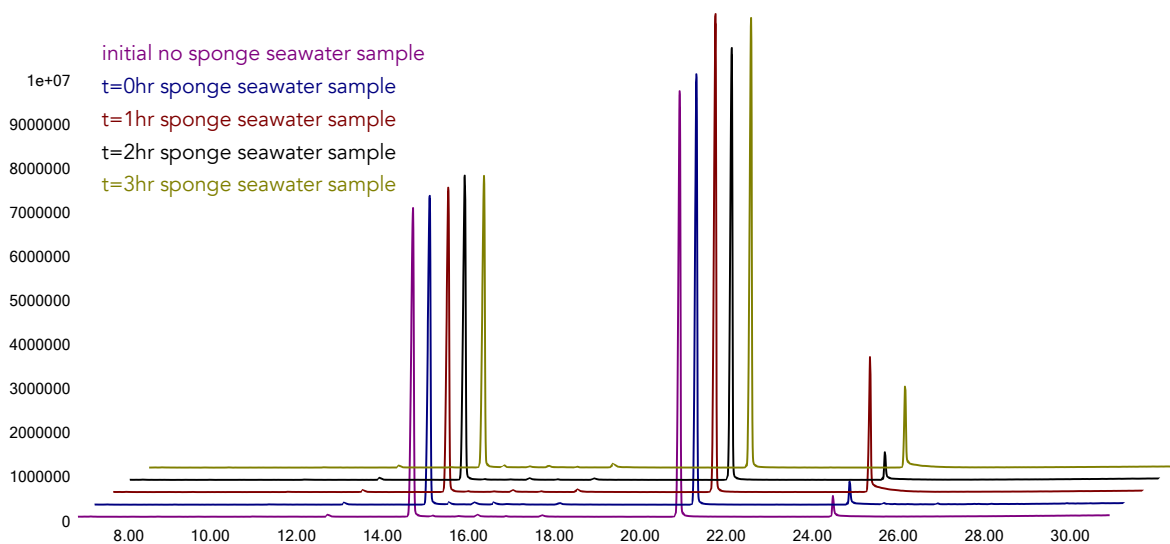
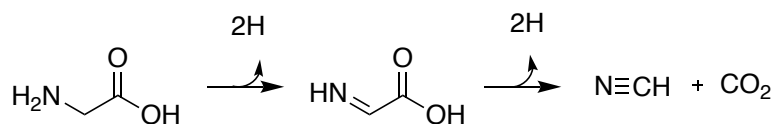


Figure 3.10.5. GCMS traces for DMTMM detection of cyanide in sponge seawater time series experiment. No cyanide is detected in any sample.

Section 3.11. Cloning of the hcnABC cluster



Scheme 3.11. 1. Formation of hydrogen cyanide from glycine by HcnABC.

Hydrogen cyanide is produced from glycine in bacteria by hydrogen cyanide (HCN) synthase, which is encoded by the gene cluster hcnABC (73, 74) (Scheme 3.11.1). The cluster seems to have evolved from amino acid dehydrogenases/oxidases: HcnA is related to formate dehydrogenases, HcnB is related to sarcosine (N-methylglycine) oxidases, and HcnC is related to D-amino acid oxidases. This cluster was first characterized in *Pseudomonas fluorescens* CHA0, but is found in many bacteria and is especially common in soil and rhizome-associated bacteria (73, 75, 76).

Given the role of cyanide in the production of nitrogenous terpenes, we queried our Bubarida sponge metagenome for hcnABC. Blasting for each of the genes individually returned many amino acid reductase/oxidase genes unrelated to HCN production, however, sorting the blast results by the contig on which they appeared allowed us to identify a three-gene cluster of hcnABC homologues. The genes appeared in a metagenome associated genome (MAG) which is predicted to belong to the *Burkholderia cepacia* complex (*B. cepacia* is no longer considered a single species and taxonomy of this group is debated) (77). *B. cepacia* is a common human commensal and pathogen and strains isolated from humans have been reported to produce HCN (78, 79).

To further investigate this putative hcnABC cluster, I cloned the cluster into a TOPO vector and heterologously expressed the cluster in *E. coli*. We did not detect HCN production from the

hcnABC-containing cultures, however, we only tested the cultures on a small scale using commercial HCN test strips, which are less sensitive than the DMTMM method (72), which was published after these HCN experiments were conducted. Because of concerns about the safety issues associated with hydrogen cyanide, we did not attempt to induce further expression of the hcnABC cluster or scale up our culturing. We were especially concerned that the shaking and heat required for large scale culturing in open flasks could release cyanide into the air. Some of these concerns could be mediated if we had access to a shaking incubator that could fit in a fume hood.

The hcnABC cluster was PCR amplified from metagenomic Bubarida sponge DNA (DNA extraction methods described in Wilson et al., 2023 (66)) using primers KWP014: GATCGGGGCAATACGATAGA and KWP017: CTGACCACCCCCGGCTCGAA. For each 25 μ L reaction, 12.5 μ L Q5 polymerase master mix, 9.75 μ L ultrapure water, 0.25 μ L template DNA, and 2.5 μ L oligonucleotide primers (10uM each) were combined and run through the following PCR cycle: initial denaturation for 1 minute at 98 °C, 35 cycles of denaturation for 15 seconds at 98 °C, annealing for 30 seconds at 69 °C, extension for 4 minutes at 72°C, and a final extension for 2 minutes at 72 °C. The ~6kb band was gel purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research). The gel purified product was cloned into the pCR®-Blunt II-TOPO® vector using an Invitrogen Zero Blunt™ TOPO™ PCR Cloning Kit (Thermo Fisher). The construct was transformed into C2987H *E. coli* DH5 α competent cells (NEB) according to the included transformation protocol. The construct was isolated with a QIAprep Spin Miniprep ki, digested with restriction enzyme XbaI (NEB), and sequenced by GeneWiz. 5 mL Construct-containing *E. coli* DH5 α cultures were incubated at 37°C with shaking overnight in sealed 15mL Falcon tubes

then tested for the presence of cyanide using the Quantofix cyanide 1-30 µg/L kit (Macherey-Nagel). No change in color was seen in the test strips compared to the empty vector control.

Section 3.12. References

1. J. Chapman, A. E. Ismail, C. Z. Dinu, Industrial Applications of Enzymes : Recent Advances , Techniques , and Outlooks. *Catalysts* **8**, 20–29 (2018).
2. B. Wiltschi, T. Cernava, A. Dennig, M. Galindo, M. Geier, S. Gruber, M. Haberbauer, P. Heidinger, E. Herrero, R. Kratzer, C. Luley-goedl, C. A. Müller, J. Pitzer, D. Ribitsch, M. Sauer, K. Schmölzer, W. Schnitzhofer, C. W. Sensen, J. Soh, K. Steiner, C.K. Winker, M. Winkler, T. Wriessnegger, Enzymes revolutionize the bioproduction of value-added compounds : From enzyme discovery to special applications. *Biotechnol. Adv.* **40**, 107520 (2020).
3. O. Hayaishi, K. Hashimoto, Pyrocatecase: A new enzyme catalizing oxidative breakdown of pyrocatechin. *J. Biochem.* **37**, 371–374 (1950).
4. M. A. Rude, T. S. Baron, S. Brubaker, M. Alibhai, S. B. Del Cardayre, A. Schirmer, Terminal Olefin (1-Alkene) Biosynthesis by a Novel P450 Fatty Acid Decarboxylase from *Jeotgalicoccus* Species. *Appl. Environ. Microbiol.* **77**, 1718–1727 (2011).
5. K. Yamanaka, C. Maruyama, H. Takagi, Y. Hamano, e -Poly- L -lysine dispersity is controlled by a highly unusual nonribosomal peptide synthetase. *Nat. Chem. Biol.* **4** (2008).
6. N. Ziemert, M. Alanjary, T. Weber, Natural Product Reports The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* **33**, 988–1005 (2016).
7. A. Siegl, U. Hentschel, PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environ. Microbiol. Rep.* **2**, 507–513 (2010).
8. S. L. Robinson, J. Piel, S. Sunagawa, A roadmap for metagenomic enzyme discovery. *Nat. Prod. Rep.* **38**, 1994–2023 (2021).
9. F. Berini, C. Casciello, G. L. Marcone, F. Marinelli, Metagenomics : novel enzymes from non-culturable microbes. *FEMS Microbiol. Lett.* **364**, 1–19 (2017).
10. C. A. Muller, V. Perz, C. Provasnek, F. Quartinella, G. M. Guebitz, G. Berg, Discovery of Polyesterases from Moss- Associated Microorganisms. *Appl. Environ. Microbiol.* **83**, 1–13 (2017).
11. M. Ferrer, O. V Golyshina, N. Tatyana, A. N. Khachane, D. Reyes-, V. A. P. Martins, D. Santos, K. Elborough, G. Jarvis, A. Neef, M. M. Yakimov, N. Kenneth, P. N. Golyshin, Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora.

- Environ. Microbiol.* **7**, 1996–2010 (2005).
12. W. Liebl, A. Angelov, J. Juergensen, J. Chow, A. Loeschcke, T. Drepper, T. Classen, J. Pietruzska, A. Ehrenreich, W. R. Streit, K. Jaeger, Alternative hosts for functional (meta)genome analysis. *Appl. Microbiol. Biotechnol.* **98**, 8099–8109 (2014).
 13. D. Zarafeta, D. Moschidi, E. Ladoukakis, S. Gavrilov, Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. *Nat. Publ. Gr.*, 1–16 (2016).
 14. B. He, S. Jin, J. Cao, L. Mi, J. Wang, Biotechnology for Biofuels Metatranscriptomics of the Hu sheep rumen microbiome reveals novel cellulases. 1–15 (2019).
 15. E. Lanfranchi, T. Pavkov-keller, E. Koehler, M. Diepold, H. Joosten, M. Gruber-khadjawi, G. G. Thallinger, Enzyme discovery beyond homology : a unique hydroxynitrile lyase in the Bet v1 superfamily. *Sci. Rep.* **7**, 1–15 (2017).
 16. W. Finnigan, L. J. Hepworth, S. L. Flitsch, N. J. Turner, RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades. *Nat. Catal.* **4**, 98–104 (2021).
 17. J. Gunera, F. Kindinger, S. M. Li, P. Kolb, PrenDB, a substrate prediction database to enable biocatalytic use of prenyltransferases. *J. Biol. Chem.* **292**, 4003–4021 (2017).
 18. B. Savelli, Q. Li, M. Webber, A. M. Jemmat, A. Robitaille, M. Zamocky, C. Mathé, C. Dunand, RedoxiBase: A database for ROS homeostasis regulated proteins. *Redox Biol.* **26**, 1–5 (2019).
 19. P. C. F. Buchholz, C. Vogel, W. Reusch, M. Pohl, D. Rother, A. C. Spieß, J. Pleiss, BioCatNet: A Database System for the Integration of Enzyme Sequences and Biocatalytic Experiments. *ChemBioChem* **17**, 2093–2098 (2016).
 20. R. M. P. Siloto, R. J. Weselake, Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatal. Agric. Biotechnol.* **1**, 181–189 (2012).
 21. K. D. Belsare, M. C. Andorfer, F. S. Cardenas, J. R. Chael, H. J. Park, J. C. Lewis, A Simple Combinatorial Codon Mutagenesis Method for Targeted Protein Engineering. *ACS Synth. Biol.* **6**, 416–420 (2017).
 22. E. O. McCullum, B. A. R. Williams, J. Zhang, J. C. Chaput, “Random Mutagenesis by Error-Prone PCR” in *In Vitro Mutagenesis Protocols*, (2010), pp. 103–109.
 23. W.P. Stemmer, Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).
 24. A. Pourmir, T. W. Johannes, Directed evolution: Selection of the host organism. *Comput.*

- Struct. Biotechnol. J.* **2**, e201209012 (2012).
25. M. Vardakou, M. Salmon, J. A. Faraldos, P. E. O'Maille, Comparative analysis and validation of the malachite green assay for the high throughput biochemical characterization of terpene synthases. *MethodsX* **1**, e187–e196 (2014).
 26. M. D. Truppo, J. D. Rozzell, J. C. Moore, N. J. Turner, Rapid screening and scale-up of transaminase catalysed reactions. *Org. Biomol. Chem.* **7**, 395–398 (2009).
 27. F. Moris-Varas, A. Shah, J. Aikens, N. P. Nadkarni, J. D. Rozzell, D. C. Demirjian, Visualization of enzyme-catalyzed reactions using pH indicators: Rapid screening of hydrolase libraries and estimation of the enantioselectivity. *Bioorganic Med. Chem.* **7**, 2183–2188 (1999).
 28. S. Lindman, A. Hernandez-Garcia, O. Szczepankiewicz, B. Frohm, S. Linse, In vivo protein stabilization based on fragment complementation and a split GFP system. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19826–19831 (2010).
 29. A. W. Golinski, K. M. Mischler, S. Laxminarayan, N. L. Neurock, M. Fossing, H. Pichman, S. Martiniani, B. J. Hackel, High-throughput developability assays enable library-scale identification of producible protein scaffold variants. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–11 (2021).
 30. N. Vervoort, K. Goossens, M. Baeten, Q. Chen, Recent advances in analytical techniques for high throughput experimentation. *Anal. Sci. Adv.* **2**, 109–127 (2021).
 31. M. Schober, C. MacDermaid, A. A. Ollis, S. Chang, D. Khan, J. Hosford, J. Latham, L. A. F. Ihnken, M. J. B. Brown, D. Fuerst, M. J. Sanganee, G.-D. Roiban, Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nat. Catal.* **2**, 909–915 (2019).
 32. M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra-Garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. McLaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel, A. Rodriguez-Granillo, S. A. Robaire, E. C. Sherer, M.D. Truppo, A. M. Whittaker, D. Verma, L. Xiao, Y. Xu, H. Yang, Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science (80-.)*. **366**, 1255–1259 (2019).
 33. T. Benkovics, J. A. McIntosh, S. M. Silverman, J. Kong, P. Maligres, T. Itoh, H. Yang, M. A. Huffman, D. Verma, W. Pan, H.-I. Ho, J. Vroom, A. Knight, J. Hurtak, W. Morris, N. A. Strotman, G. Murphy, K. M. Maloney, P. S. Fier, Evolving to an Ideal Synthesis of Molnupiravir, an Investigational Treatment for COVID-19. *ChemRxiv* (2020) <https://doi.org/10.26434/chemrxiv.13472373.v1>.
 34. J. A. McIntosh, T. Benkovics, S. M. Silverman, M. A. Huffman, J. Kong, P. E. Maligres,

- T. Itoh, H. Yang, D. Verma, W. Pan, H. I. Ho, J. Vroom, A. M. Knight, J. A. Hurtak, A. Klapars, A. Fryszkowska, W. J. Morris, N. A. Strotman, G. S. Murphy, K. M. Maloney, P. S. Fier, Engineered Ribosyl-1-Kinase Enables Concise Synthesis of Molnupiravir, an Antiviral for COVID-19. *ACS Cent. Sci.* **7**, 1980–1985 (2021).
35. J. Schrader, J. Bohlmann, Eds., *Biotechnology of Isoprenoids* (Springer Cham, 2015).
 36. M. Li, R. Nian, M. Xian, H. Zhang, Metabolic engineering for the production of isoprene and isopentenol by *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **102**, 7725–7738 (2018).
 37. C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, H. Jiang, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, H. Jiang, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger, J. D. Newman, High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
 38. P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, S. Fickes, D. Diola, K. R. Benjamin, J. D. Keasling, M. D. Leavell, D. J. McPhee, N. S. Renninger, J. D. Newman, C. J. Paddon, Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 111–118 (2012).
 39. S. Malik, R. M. Cusidó, M. H. Mirjalili, E. Moyano, J. Palazón, M. Bonfill, Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: A review. *Process Biochem.* **46**, 23–34 (2011).
 40. O. A. Carter, R. J. Peters, R. Croteau, Monoterpene biosynthesis pathway construction in *Escherichia coli*. *Phytochemistry* **64**, 425–433 (2003).
 41. J. Alonso-Gutierrez, R. Chan, T. S. Bath, P. D. Adams, J. D. Keasling, C. J. Petzold, T. S. Lee, Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. *Metab. Eng.* **19**, 33–41 (2013).
 42. T. P. Korman, P. H. Opgenorth, J. U. Bowie, A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nat. Commun.* **8**, 1–8 (2017).
 43. M. Breuer, W. Siegel, S. Rüdenauer, R. Pelzer, Enzymatic cyclization of homofarnesyllic acid (2021).
 44. G. Sanhaji, A. Rousseaux, S. Noel, E. Eichhorn, A solid form of (-)-ambrox formed by a bioconversion of homofarnesol in the presence of a biocatalyst. 1–55 (2017).
 45. C. J. Paddon, J. D. Keasling, Semi-synthetic artemisinin: A model for the use of synthetic

- biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12**, 355–367 (2014).
46. M. Peplow, Sanofi launches malaria drug production. *Chem. world*, 1–4 (2013).
 47. M. Broecker, Production of squalene from hyper-producing yeasts. 1–14 (2011).
 48. J. A. Ubersax, D. M. Platt, Genetically modified microbes producing isoprenoids (2010).
 49. K. Fisher, S. J. Schofer, D. B. Kanne, Squalane and isosqualane compositions and methods for preparing the same. 1–31 (2011).
 50. F. Cafieri, E. Fattorusso, S. Magno, C. Santacroce, D. Sica, Isolation and Structure of Axisonitrile-1 and axisothiocyanate-1: Two Unusual Sesquiterpenoids from the Marine Sponge *Axinella cannabina*. *Tetrahedron* **29**, 4259–4262 (1973).
 51. I. Burkhardt, T. De Rond, P. Y. Chen, B. S. Moore, Ancient plant-like terpene biosynthesis in corals. *Nat. Chem. Biol.* **18**, 664–669 (2022).
 52. P. D. Scesa, Z. Lin, E. W. Schmidt, Ancient defensive terpene biosynthetic gene clusters in the soft corals. *Nat. Chem. Biol.* **18**, 659–663 (2022).
 53. J. P. Torres, Z. Lin, J. M. Winter, P. J. Krug, E. W. Schmidt, Animal biosynthesis of complex polyketides in a photosynthetic partnership. *Nat. Commun.* **11**, 1–12 (2020).
 54. J. P. Torres, E. W. Schmidt, The biosynthetic diversity of the animal world. *J. Biol. Chem.* **294**, 17684–17692 (2019).
 55. S. Vargas, A. Schuster, K. Sacher, G. Büttner, S. Schätzle, B. Lächli, K. Hall, J. N. A. Hooper, D. Erpenbeck, G. Wörheide, Barcoding sponges: An overview based on comprehensive sampling. *PLoS One* **7**, e39345 (2012).
 56. C. P. Meyer, Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. *Biol. J. Linn. Soc.* **79**, 401–459 (2003).
 57. D. Erpenbeck, O. Voigt, A. M. Al-Aidaros, M. L. Berumen, G. Büttner, D. Catania, A. N. Guirguis, G. Paulay, S. Schätzle, G. Wörheide, Molecular biodiversity of Red Sea demosponges. *Mar. Pollut. Bull.* **105**, 507–514 (2016).
 58. T. L. Turner, S. Lonhart, The Sponges of the Carmel Pinnacles Marine Protected Area. *bioRxiv*, 1–59 (2022).
 59. T. L. Turner, M. S. Pankey, The order Axinellida (Porifera : Demospongiae) in California. *bioRxiv*, 1–54 (2022).
 60. E. Gazave, S. Carteron, A. Chenuil, E. Richelle-Maurer, N. Boury-Esnault, C. Borchellini, Polyphyly of the genus *Axinella* and of the family Axinellidae (Porifera: Demospongiae).

- Mol. Phylogenet. Evol.* **57**, 35–47 (2010).
61. A. Galitz, Y. Nakao, P. J. Schupp, G. Wörheide, A Soft Spot for Chemistry – Current Taxonomic and Evolutionary Implications of Sponge Secondary Metabolite Distribution. *Mar. Drugs* **19**, 1–25 (2021).
 62. C. Morrow, P. Cárdenas, Proposal for a revised classification of the Demospongiae (Porifera). *Front. Zool.* **12**, 1–27 (2015).
 63. A. B. Woodside, Z. Huang, C. D. Poulter, Trisammonium Geranyl Diphosphate. *Org. Synth.* **8**, 616 (1993).
 64. D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, 526–531 (2004).
 65. B. C. T. Chung, R. H. Miller, “Preparation and storage of competent *Escherichia coli* cells” in *Methods in Enzymology*, (1993), pp. 621–627.
 66. K. Wilson, T. de Rond, I. Burkhardt, T. S. Steele, R. J. B. Schäfer, S. Podell, E. E. Allen, B. S. Moore, Terpene biosynthesis in marine sponge animals. *Proc. Natl. Acad. Sci.* **120**, 1–8 (2023).
 67. X. Lin, D. E. Cane, Biosynthesis of the Sesquiterpene Antibiotic Albaflavenone in *Streptomyces coelicolor*. Mechanism and Stereochemistry of the Enzymatic Formation of Epi-isozizaene. *J. Am. Chem. Soc.* **131**, 6332–6333 (2009).
 68. D. B. Little, R. B. Croteau, Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases d -selinene synthase and c -humulene synthase q. *Arch. Biochem. Biophys.* **402**, 120–135 (2002).
 69. M. T. Hamann, P. J. Scheuer, M. Kelly-Borges, Biogenetically Diverse, Bioactive Constituents of a Sponge, Order Verongida: Bromotyramines and Sesquiterpene-Shikimate Derived Metabolites. *J. Org. Chem.* **58**, 6565–6569 (1993).
 70. M. T. Hamann, P. J. Scheuer, Cyanopuupehenol, an antiviral metabolite of a sponge of the order Verongida. *Tetrahedron Lett.* **32**, 5671–5672 (1991).
 71. M. T. Hamann, “Biologically active constituents of some marine invertebrates.” (1992).
 72. A. Yamaguchi, H. Miyaguchi, A screening method for cyanide in blood by dimethoxytriazinyl derivatization-GC/MS. *J. Chromatogr. Sci.* **59**, 1–6 (2021).
 73. J. Laville, C. Blumer, C. VON Schroetter, V. Gaia, G. Defago, C. Keel, D. Haas, Characterization of the hcnABC Gene Cluster Encoding Hydrogen Cyanide Synthase and Anaerobic Regulation by ANR in the Strictly Aerobic Biocontrol Agent *Pseudomonas fluorescens* CHA0. *J. Bacteriol.* **180**, 3187–3196 (1998).

74. M. C. Blumer, D. Haas, Mechanism, regulation, and ecological role of bacterial cyanide biosynthesis. *Arch. Microbiol.* **173**, 170–177 (2000).
75. M. Bruto, C. Prigent-combaret, D. Muller, Y. Moe, Analysis of genes contributing to plant-beneficial functions in plant related Proteobacteria. *Sci. Rep.* **4**, 1–10 (2014).
76. A. Sehwat, S. S. Sindhu, B. R. Glick, Hydrogen cyanide production by soil bacteria : Biological control of pests and promotion of plant growth in sustainable agriculture. *Pedosph. An Int. J.* **32**, 15–38 (2022).
77. Y. Jin, J. Zhou, J. Zhou, M. Hu, Q. Zhang, N. Kong, H. Ren, Genome-based classification of Burkholderia cepacia complex provides new insight into its taxonomic status. *Biol. Direct* **15**, 1–14 (2020).
78. B. Ryall, X. Lee, J. E. A. Zlosnik, S. Hoshino, H. D. Williams, Bacteria of the Burkholderia cepacia complex are cyanogenic under biofilm and colonial growth conditions. *BMC Microbiol.* **8**, 1–9 (2008).
79. F. J. Gilchrist, H. Sims, A. Alcock, A. M. Jones, R. J. Bright-thomas, D. Smith, P. Špane, A. K. Webb, W. Lenney, Is Hydrogen Cyanide a Marker of Burkholderia cepacia Complex ? **51**, 3849–3851 (2013).

Section 3.13. Acknowledgements

Chapter 3, in part, is a reprint of the material as it appears in *Proceedings of the National Academy of Sciences*, Wilson, K., de Rond, T., Burkhardt, I., Steele, T.S., Schäfer, R.J.B., Podell, S., Allen, E.E., and Moore, B.S., 2023. The dissertation author was the primary researcher and author of this paper.

DISCUSSION AND CONCLUSION

In summary, terpenoids are a vast and widespread class of natural products that feature impressive structural diversity and high bioactivity. Terpene synthases are powerful enzymes that produce tens of thousands of terpenoid products from just a handful of linear substrates. There is an increased interest in terpene synthases as biocatalysts, which has driven terpene synthase discovery efforts. As we sequence more organisms, we find more and more terpene synthases, and it becomes increasingly apparent that these enzymes are widespread throughout nature. The discovery of terpene synthases from marine sponges has added a new branch to the terpene synthase family tree. For decades, it has been assumed that the sponge microbiome is the source of the chemical diversity isolated from sponges, however this discovery of terpene synthases in the sponge animal host calls that paradigm into question. Moving forward, I hope that this work inspires researchers to investigate the role of the sponge animal in the biosynthesis of sponge derived natural products from other classes, such as pyrrole indole alkaloids.

While the discovery of sponge terpene synthases sheds light onto one step of nitrogenous terpene biosynthesis, the full pathway to these molecules has yet to be elucidated. The Bubarida nitrogenous terpene natural products feature additional cyclization compared to the terpene synthase products. Perhaps this additional cyclization could be carried out by a type 2 cyclase, or another unusual terpene synthase, such as an UbiA-type terpene synthase. It is worth investigating other isoprenoid biosynthesis genes within the Bubarida sponge genome and transcriptome. It is also possible that the sponge microbiome may play a role in the production of nitrogenous terpenes and should also be investigated for further isoprenoid biosynthesis capacity. Discovering the enzyme responsible for nitrogenous functional group incorporation is undoubtedly challenging, as

no precedent exists for this kind of enzymatic reaction. The sponge terpene synthases are located on contigs with other unannotated genes, which should be investigated as potential biosynthesis enzymes, especially since biosynthetic gene clustering has recently been found in other animals. RNA-seq co-expression analysis could also be a valuable tool to narrow down the potential candidate genes for nitrogenous functional group incorporation. Other biosynthetic enzymes have been shown to form isonitriles (1), and though none utilize cyanide, these enzymes may still serve as helpful bioinformatic hooks and may provide insight into the types of enzymatic chemistry that may underly the production of nitrogenous terpenes.

For the study of San Diego Bubarida sponges specifically, there are a number of directions in which this work could continue. While these sponges were occasionally stored in aquaria for short periods of time, I never made a concerted effort to culture these sponges in aquaria. If these Bubarida sponges can be maintained in aquaria with retention of nitrogenous terpene production, there are a number of experiments that should be carried out. Stable isotope feeding experiments with whole sponges could yield vital information about the order of biosynthetic steps. ^{13}C and /or deuterium-labeled FPP could be fed to the sponge animal directly, or be used as a substrate for the terpene synthases to produce isotopically-labeled terpene synthase products, which could also be fed to the sponge animal. Isotopically-labeled cyanide and thiocyanate could also be used to replicate experiments carried out by Garson and Scheuer and confirm their findings in the Bubarida sponge (2). Imaging mass spectrometry might also be a helpful tool for determining where the biosynthetic steps take place and whether a bacterium may play a role in that process. The preliminary finding of cyanide in the seawater around the Bubarida sponge should also be further investigated. In the paper where Turner and Lonhart describe this Bubarida sponge (3), they also

note that this sponge has different types which vary by location and morphology. This raises the question of how the nitrogenous terpenes are distributed across these various species types.

The sponge terpene synthases in and of themselves have potential for further study. A brief attempt was made by Percival Chen to crystalize uBuTS-1 and uBuTS-2, however we did not follow up on these crystallization efforts. A crystal structure of a sponge terpene synthase could provide valuable information about how the active site residues interact with the prenyl substrates and how the structure of the enzyme fold affects the structure of the final products. With the acquisition of high throughput screening equipment by the Center for Marine Biotechnology and Biomedicine, enzyme engineering efforts with these sponge terpene synthases are also quite appealing. A crystal structure could help guide future engineering and mutagenesis efforts, though enough is known about the conserved terpene synthase active site residues that a crystal structure may not be needed to start directed evolution efforts. Directed evolution could yield terpene synthases with characteristics that are desired for biocatalysis, such as increased activity, greater substrate scope, different product profiles, greater stability, and others.

This work started with a desire to explore the ocean and a curiosity about a lumpy orange sponge that lives therein. The path I followed during my graduate career was not linear, and there were many challenges and rabbit holes along the way. Throughout it all, I let curiosity guide my journey and I would encourage the students who come after me to do the same. These molecules have fascinated me, and I hope they may fascinate many other scientists in the future. John Faulkner was absolutely correct when he said there is “no more interesting biosynthetic study

among the marine natural products.”(4) Nature holds countless mysteries and I feel grateful to have answered just one of them.

References

1. X. Zhang, L. Evanno, E. Poupon, Biosynthetic Routes to Natural Isocyanides. *European J. Org. Chem.* **2020**, 1919–1929 (2020).
2. M. J. Garson, J. S. Simpson, Marine isocyanides and related natural products - Structure, biosynthesis and ecology. *Nat. Prod. Rep.* **21**, 164–179 (2004).
3. T. L. Turner, S. Lonhart, The Sponges of the Carmel Pinnacles Marine Protected Area. *bioRxiv*, 1–59 (2022).
4. D. J. Faulkner, Interesting aspects of marine natural products chemistry. *Tetrahedron* **33**, 1421–1443 (1977).