

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Wandering in genomes and probing biosynthetic machinery in pursuit of the evolutionary origins and diversification of octocoral terpenoids

Permalink

<https://escholarship.org/uc/item/64j5c63k>

ISBN

9798273309111

Author

Grayson, Natalie Eve

Publication Date

2025-01-09

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Wandering in genomes and probing biosynthetic machinery in pursuit of the evolutionary
origins and diversification of octocoral terpenoids

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

in

Marine Biology

by

Natalie Eve Grayson

Committee in charge:

Professor Bradley S. Moore, Chair
Professor William Fenical
Professor Amro Hamdoun
Professor Paul Jensen
Professor Todd P. Michael

2025

Copyright

Natalie Eve Grayson, 2025

All rights reserved.

The dissertation of Natalie Eve Grayson is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

DEDICATION

To Papa, a brilliant scientist and an unapologetic eccentric.

To Bubby, who carried Papa's spark,

and helped to pass the torch to me.

To the tidepools.

To my mom and dad,

for listening with genuine curiosity,

supporting unconditionally,

believing I was on the path, even when the fog was so thick I couldn't see where I was going.

EPIGRAPH

“you do not just wake up and become the butterfly

— *growth is a process.*”

Rupi Kaur

TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE	iii
DEDICATION	iv
EPIGRAPH.....	v
TABLE OF CONTENTS.....	vi
LIST OF FIGURES	viii
LIST OF TABLES	xi
ACKNOWLEDGEMENTS	xii
VITA.....	xv
PUBLICATIONS.....	xv
ABSTRACT OF DISSERTATION	xvi
CHAPTER 1: Introduction	1
1.1 References.....	6
CHAPTER 2: Leveraging Deep Sea Octocoral to Explore Diversification of Diterpene Biosynthesis	8
2.1 Introduction for Chapter 2	8
2.3 Abstract	11
2.4 Introduction	12
2.5 Results	14
2.6 Discussion	23
2.8 Main Text References	26
2.9 Supplementary Information	30
2.10 Supplementary Information References	101
2.11 Acknowledgements.....	105
CHAPTER 3: Uncovering a Widespread Metabolic Gene Cluster Family in Octocoral.....	106
3.1 Introduction for Chapter 3	106
3.2 Chapter 3 Introduction References	107
3.3 Reprint of “A widespread metabolic gene cluster family in metazoans”	108
3.4 Acknowledgements.....	199
CHAPTER 4: A Case Study of Allelic Heterozygosity and Gene Duplication in Relation to Octocoral Terpenoid Evolution	200
4.1 Introduction for Chapter 4.....	200
4.2 Chapter 4 Introduction References	201
4.3 Abstract	202
4.4 Introduction	202
4.5 Results	203
4.6 Discussion	213
4.7 Main Text References	217
4.8 Supplementary Information	221

4.9 Supplementary Information References	293
4.10 Acknowledgements.....	295

LIST OF FIGURES

Figure 2-1: Examples of bioactive octocoral diterpenoids from the five largest structural classes and their biosynthetic precursors produced by octocoral terpene cyclases.	13
Figure 2-2: Taxonomic classification of deep-sea octocoral samples and distribution of diterpenoid compound types throughout octocorals.	16
Figure 2-3: Phylogeny and enzymatic activity of octocoral terpene cyclases (TCs).....	21
Figure 2-4: Codon usage patterns and synonymous codon biases in octocoral terpene cyclase (TC) genes. .	23
Figure 2-5: Schematic of possible early TC evolution in octocorals.....	26
Figure 4-1: GCMS chromatogram of <i>Leptogorgia chilensis</i> crude extract.....	204
Figure 4-2: Phylogenetic relationship and genomic context of L. chilensis TCs.....	207
Figure 4-3: Products generated in enzyme assays.....	209
Figure 4-4: Protein models and comparison of compound production of LcTC7 paralogs.	211
Figure 4-5: Distribution of different repeat classes in L. chilensis assembly.	212
Figure 4-6: Suggested outcome of duplicated genes and comparison of paralog functions.....	217

LIST OF SUPPLEMENTARY FIGURES

Figure 2-S 1: Phylogenomic tree constructed with octocoral mitochondrial genes.	49
Figure 2-S 2: GCMS analysis of hexane extract derived from <i>Acanthogorgia</i> sp. NA124-045.	50
Figure 2-S 3: GCMS analysis of hexane extract derived from <i>Callistephanus simplex</i> NA124-047.....	50
Figure 2-S 4: GCMS analysis of hexane extract derived from <i>Callistephanus simplex</i> NA124-141.	51
Figure 2-S 5: GCMS analysis of hexane extract derived from <i>Paragorgia</i> cf. <i>arborea</i> NA124-195.....	51
Figure 2-S 6: GCMS analysis of hexane extract derived from <i>Balticina</i> cf. <i>californicum</i> NA124-218.....	52
Figure 2-S 7: GCMS analysis of hexane extract derived from <i>Victorgorgia</i> sp. SCB-084.	52
Figure 2-S 8: GCMS analysis of hexane extract derived from <i>Chrysogorgia</i> sp. SCB-144.....	53
Figure 2-S 9: GCMS analysis of hexane extract derived from <i>Chrysogorgia</i> sp. SCB-157.....	53
Figure 2-S 10: GCMS analysis of hexane extract derived from <i>Paragorgia jamesi</i> SCB-216.	54
Figure 2-S 11: Phylogenetic tree of octocoral TC amino acid sequences.	55
Figure 2-S 12: SDS-PAGE analysis of heterologously produced enzymes.	58
Figure 2-S 13: GCMS total ion chromatograms of NA124-045-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	59
Figure 2-S 14: GCMS total ion chromatograms of NA124-045-TC-3 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	60
Figure 2-S 15: GCMS total ion chromatograms of NA124-045-TC-5 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	61
Figure 2-S 16: GCMS total ion chromatograms of NA124-047-TC-3 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	62
Figure 2-S 17: GCMS total ion chromatograms of NA124-047-TC-4 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	63
Figure 2-S 18: GCMS total ion chromatograms of NA124-141-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	64
Figure 2-S 19: GCMS total ion chromatograms of NA124-141-TC-4 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	65
Figure 2-S 20: GCMS total ion chromatograms of NA124-15-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	66

Figure 2-S 21: GCMS total ion chromatograms of NA124-195-TC-3 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	67
Figure 2-S 22: GCMS total ion chromatograms of NA124-195-TC-5 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	68
Figure 2-S 23: GCMS total ion chromatograms of SCB-084-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	69
Figure 2-S 24: GCMS total ion chromatograms of SCB-084-TC-2 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	70
Figure 2-S 25: GCMS total ion chromatograms of SCB-084-TC-5 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	71
Figure 2-S 26: GCMS total ion chromatograms of SCB-084-TC-6 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	72
Figure 2-S 27: GCMS total ion chromatograms of SCB-144-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	73
Figure 2-S 28: GCMS total ion chromatograms of SCB-144-TC-3 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	74
Figure 2-S 29: GCMS total ion chromatograms of SCB-144-TC-4 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	75
Figure 2-S 30: GCMS total ion chromatograms of SCB-144-TC-7 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	76
Figure 2-S 31: GCMS total ion chromatograms of SCB-157-TC-2 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	77
Figure 2-S 32: GCMS total ion chromatograms of SCB-157-TC-5 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	78
Figure 2-S 33: GCMS total ion chromatograms of SCB-218-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	79
Figure 2-S 34: GCMS total ion chromatograms of SCB-218-TC-2 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).	80
Figure 2-S 35: GCMS total ion chromatograms of in vitro incubations of the clade 1-I ancestor with GPP (top), FPP (middle), and GGPP (bottom).	81
Figure 2-S 36: GCMS total ion chromatograms of in vitro incubations of the clade 1-II ancestor with GPP (top), FPP (middle), and GGPP (bottom).	82
Figure 2-S 37: GCMS total ion chromatograms of in vitro incubations of the clade 1-III ancestor with GPP (top), FPP (middle), and GGPP (bottom).	83
Figure 2-S 38: GCMS total ion chromatograms of in vitro incubations of the clade 1-IV ancestor with GPP (top), FPP (middle), and GGPP (bottom).	84
Figure 2-S 39: GCMS total ion chromatograms of in vitro incubations of the clade 1-V ancestor with GPP (top), FPP (middle), and GGPP (bottom).	85
Figure 2-S 40: Determination of the absolute configuration of enzymatically obtained 9.	86
Figure 2-S 41: Evolutionary divergence and population genetic patterns of octocoral terpene cyclase (TC) genes.	87
Figure 2-S 42: Codon composition and codon bias metrics in octocoral terpene cyclase (TC) genes.	88
 Figure 4-S 1: Annotated chromatogram of crude extract	243
Figure 4-S 2: GenomeScope visualization of heterogeneity	244

Figure 4-S 3: Heat map of sequence similarity across all <i>L. chilensis</i> TCs	245
Figure 4-S 4: Full synteny plot of HapA and HapB	245
Figure 4-S 5: Semi-scaffolding of TC5 contigs.....	246
Figure 4-S 6: Semi-scaffolding of TC8 contigs.....	246
Figure 4-S 7: Semi-scaffolding of TC6 contigs.....	247
Figure 4-S 8: Semi-scaffolding of TC9 contigs.....	247
Figure 4-S 9: SDS gel of all proteins	248
Figure 4-S 10: GCMS total ion chromatograms of HapA_TC1.....	249
Figure 4-S 11: GCMS total ion chromatograms of HapB_TC1.....	249
Figure 4-S 12: GCMS total ion chromatograms of HapA_TC2a.....	250
Figure 4-S 13: GCMS total ion chromatograms of HapA_TC2b.....	250
Figure 4-S 14: GCMS total ion chromatograms of HapB_TC3.....	251
Figure 4-S 15: GCMS total ion chromatograms of HapA_TC4.....	251
Figure 4-S 16: GCMS total ion chromatograms of HapA_TC5a.....	252
Figure 4-S 17: GCMS total ion chromatograms of HapA_TC5b.....	252
Figure 4-S 18: GCMS total ion chromatograms of HapA_TC5c.....	253
Figure 4-S 19: GCMS total ion chromatograms of HapA_TC7a.....	253
Figure 4-S 20: GCMS total ion chromatograms of HapB_TC7a.....	254
Figure 4-S 21: GCMS total ion chromatograms of HapB_TC7b.....	254
Figure 4-S 22: GCMS total ion chromatograms of HapA_TC8a.....	255
Figure 4-S 23: GCMS total ion chromatograms of HapB_TC9.....	255
Figure 4-S 24: Multiple sequence alignment of TC7s.....	256
Figure 4-S 25: GCMS total ion chromatograms of TC7a and TC7b site swap mutagenesis with FPP	257
Figure 4-S 26: GCMS total ion chromatograms of TC7a and TC7b site swap mutagenesis with GGPP	258
Figure 4-S 27: GCMS total ion chromatograms of TC7a and TC7b residue 91 mutagenesis with FPP.....	259
Figure 4-S 28: GCMS total ion chromatograms of TC7a and TC7b residue 91 mutagenesis with GGPP	260
Figure 4-S 29: Proposed mechanisms of TC2a and TC2b reactions.....	261

LIST OF TABLES

Table 4-1: Statistics of phased genome assemblies	205
Table 4-2: TC distribution across haplotypes and syntenic blocks	206

LIST OF SUPPLEMENTARY TABLES

Table 2-S 1: Biological samples used in this study.	42
Table 2-S 2: Literature reports of XBECK-type diterpenoids in various octocorals.	43
Table 2-S 3: Amino acid percent identity matrix for TCs from clade 1-I	44
Table 2-S 4: Amino acid percent identity matrix for TCs from clade 1-II	44
Table 2-S 5: Amino acid percent identity matrix for TCs from clade 1-III	44
Table 2-S 6: Amino acid percent identity matrix for TCs from clade 1-IV	45
Table 2-S 7: Amino acid percent identity matrix for TCs from clade 1-V	45
Table 2-S 8: Comparison of identified diterpene type and terpene cyclase clades present in sequencing data if both are available from the same species.	45
Table 2-S 9: Population genetic metrics of coral TCs and mitochondrial genes from investigated octocoral species.	46
Table 2-S 10: Primers used in this study	47
Table 4-S 1: Sequencing statistics	236
Table 4-S 2: Data availability sources.....	236
Table 4-S 3: Assembly and gene model statistics	237
Table 4-S 4: NTlink (v 1.3.11) ³⁷ Scaffolding for all HapA (using ONT reads with size 30 kbp and longer.	237
Table 4-S 5: PCR conditions	238
Table 4-S 6: Primers used.....	238
Table 4-S 7: TC activities	239
Table 4-S 8: TPMs for TCs	240
Table 4-S 9: Mutant TC enzyme activities	241
Table 4-S 10: Complete TE stats	241
Table 4-S 11: Bioinformatics tools used and versions	242

ACKNOWLEDGEMENTS

I would like to thank my advisor and committee chair, Bradley S. Moore, for adopting me into your lab, setting me up for success with excellent mentorship under Immo, and for giving me an amazing project centered around beautiful and charismatic animals. At a time when I felt small and intimidated, you treated me as someone capable of growing into a version of myself I had not yet imagined. By creating such an incredible community of scientists and believing I could swim through their knowledge, collecting pieces along the way, you gave me the confidence to grow into a scientist fluent across all the languages of the Moore lab. Just as important was your encouragement outside the lab, making sure I kept getting in the water, whether by dragging me along for Windansea swims or laughing from your office window as I snuck out for some midday Scripps waves. You taught me that the best science comes when I am salt-crustured.

To the rest of my committee members, thank you for your encouragement and feedback. Paul, you have been a mentor for me since my first days setting foot on this campus, you created an environment where I could imagine a future in this field. Bill, your enthusiasm and deep passion for this marine natural products chemistry is truly contagious, especially when you are parading around in your coral shirt. Your presence reminds me that curiosity and excitement belong at the heart of science. Todd, thank you for letting me read the genetic code of the natural world around me. From spooling a coral's DNA in my own hands to assembling its genome, you gifted me experiences that filled me with awe and wonder.

I would also like to acknowledge the team at Sirenas, specifically Eduardo Esquenazi, for opening my eyes to the chemical language of marine invertebrates. Thank you for letting me spend my days playing with stinky sponges and corals, and for introducing me to the teeming scientific community at Scripps.

I would also like to thank the Moore lab in their entirety; you gave me a home here. Specifically, thank you Hannah Bone for teaching me the foundations of the Moore lab. We were ducklings together, stumbling and laughing our way through late nights and countless failed protein expressions. Those memories will always be with me. Immo, your ideas pushed me to think in new ways, and your feedback always made me sharper. This thesis and my completion of this PhD would not have been possible without you. I will always be grateful for the time you have spent patiently teaching me, never handing me easy answers and instead challenging me to learn. Taylor, you became my PhD mother, reminding me gently, with your calming voice that it was okay to not be okay and that the storm would pass.

Finally, I would like to acknowledge my friends and family. To my parents, thank you feels too small. You have always been my most patient and curious listeners. Whenever the world gets too loud or my thoughts start to whirl, you have been there. Knowing you are proud of me for who I am at my core has given me with the sense of security to be brave enough to “just get on the boat,” time and time again. Malia, I am forever grateful that I was able to do science by your side. Dr. Barioni, thank you for single-handedly inventing science and graciously allowing the rest of us to participate. Soon-to-be Dr. Alksne, from the shreds- at Scripps and IB- the pinpoint work breaks, to our earliest days studying together, you have been alongside me making me laugh and pushing me to send it. And to the rest of the sisters I found along the way, Isabel, Marina, and many more, thank you for reminding me that I am more than my work, for seeing every side of me, and for giving me the kind of deep belly laughter that makes life full.

Chapter 2 contains material currently under revision at the *Proceedings of the National Academy of Sciences*, which has been accepted pending minor revisions. The dissertation author contributed extensively to data generation for this work, serving as a co-researcher of this chapter with Bone, H.K.

Chapter 3, in full, is a reprint of the material as it appears in *Nature Chemical Biology*, Grayson, N. E., Scesa, P. D., Moore, M. L., Ledoux, J.-B., Gomez-Garrido, J., Alioto, T., Michael, T. P., Burkhardt, I., Schmidt, E. W., & Moore, B. S. (2025). The dissertation author was the primary co-researcher and co-author of this chapter with Scesa, P. D.

Chapter 4 contains material currently being prepared for publication, co-authored with Minich, J.J., Allsing, N., Michael, T.P., Moore, B.S., and Burkhardt, I. The dissertation author was a primary researcher and author of this chapter.

VITA

2020 Bachelor of Arts in Ecology and Evolutionary Biology, Princeton University

2022-2024 National Oceanic and Atmospheric Administration Margret A Davidson Fellowship

2025 Doctor of Philosophy in Marine Biology, Scripps Institution of Oceanography

PUBLICATIONS

Grayson, N.E., Minich, J.J., Alsing, N., Michael, T.P., Moore, and B.S., Burkhardt, I. “Allelic Heterozygosity and Gene Duplications Drive Terpene Diversity in *Leptogorgia chilensis*” *In prep.* (2025).

Grayson, N.E., Scesa, P.D., Moore, M.L. Ledoux, J.B., Gomez-Garrido, J., Alioto, T., Michael, T.P., Burkhardt, I., Schmidt E.W., and Moore, B.S. “A widespread metabolic gene cluster family in metazoans” in *Nat Chem Biol* (2025).

Burkhardt I, Durr L, **Grayson N. E.**, and Moore B. “Methods for the Discovery and Characterization of Octocoral Terpene Cyclases” in *Methods in Enzymology* (Academic Press, 2024).

Ledoux, J.-B., Gomez-Garrido, J., Cruz, F., Camara Ferreira, F., Matos, A., Sarropoulou, X., Ramirez-Calero, S., Aurelle, D., Lopez-Sendino, P., **Grayson, N. E.**, Moore, B. S., Antunes, A., Aguilera, L., Gut, M., Salces-Ortiz, J., Fernández, R., Linares, C., Garrabou, J., & Alioto, T. “Chromosome-level genome assembly and annotation of *Corallium rubrum*: a Mediterranean coral threatened by overharvesting and climate change” in *Genome Biology and Evolution*:17-2. (Oxford University Press UK, 2025).

Zdouc, M. M., Blin, K., Louwen, N. L. L., Navarro, J., Loureiro, C., Bader, C. D., Bailey, C. B., Barra, L., Booth, T. J., Bozhüyük, K. A. J., Cediél-Becerra, J. D. D., Charlop-Powers, Z., Chevrette, M. G., Chooi, Y. H., D’Agostino, P. M., de Rond, T., Del Pup, E., Duncan, K. R., Gu, W., ... **Grayson, N.E.**, ... Medema, M. H. “MIBiG 4.0: advancing biosynthetic gene cluster curation through global collaboration” in *Nucleic acids research*: 53-D1 (Oxford University Press, 2025).

Burkhardt, I., **Grayson, N.E.**, Bone, H.K., Leucke, H.A., Gutleben, J., Jensen, P.R., Quattrini, A.M., Chase, A.B., and Moore, B.S. “Diversification of Diterpene Biosynthesis Occurred Early in Octocoral Evolution” *Provisionally Accepted in Proc. Natl. Acad. Sci. U.S.A.* (2025).

ABSTRACT OF DISSERTATION

Wandering in genomes and probing biosynthetic machinery in pursuit of the evolutionary origins and diversification of octocoral terpenoids

by

Natalie Eve Grayson

Doctor of Philosophy in Marine Biology

University of California San Diego, 2025

Professor Bradley S. Moore, Chair

Octocorals are metazoans that prolifically produce terpenoid natural products, rivaling the chemical diversity of plants and microbes. However, until recently the genomic basis of this chemistry was unresolved. With the discovery of terpene cyclase (TC) genes encoded in octocoral genomes, a framework has emerged to trace the evolution of specialized metabolism in these animals. This thesis follows the genetic twists and turns that enable octocorals to generate novel chemistry and shows how functional characterization both validates genomic trends and provides mechanistic insight into enzyme evolution. Chapter 2 explores deep-sea octocorals, sequencing and profiling specimens of both major taxonomic orders collected at depths of ~1000 m. Chemical

analyses revealed broad sesquiterpene diversity, while diterpenes were largely confined to five families. Phylogenetic and functional analyses of deep-sea coral TCs, alongside mined public data sets, revealed rich sesquiterpene synthase diversity across deeply rooted clades, as well as monophyletic groups encoding isofunctional diterpene synthases that give rise to XBECK scaffolds. These results suggest that terpenoid biosynthesis diversified early in octocoral evolution, with the last common ancestor already possessing multiple functionally distinct TC genes, highlighting the power of genomics to illuminate the history of animal specialized metabolism.

Chapter 3 builds on our understanding of octocoral terpene biosynthesis evolution, identifying and characterizing a widespread biosynthetic gene cluster (BGC) family responsible for briarane diterpenoid biosynthesis. High-quality genomic data from corals spanning five families, combined with functional validation of enzyme-encoding genes, this work demonstrates that the evolutionary persistence of briarane chemistry across diverse octocoral lineages stems from a conserved BGC. This finding establishes that octocorals, like plants and microbes, rely on BGCs to encode and preserve specialized metabolism. Finally, chapter 4 focuses on a single prolific species, *Leptogorgia chilensis*, where a phased genome revealed extensive allelic heterozygosity and TC paralog expansions. Functional assays demonstrated how subtle active-site mutations toggle product specificity, linking allelic variation and paralog diversification to lineage-specific scaffold innovation. Collectively, these findings position octocorals as powerful models for exploring the genomic foundations of animal natural product biosynthesis and underscore their potential as reservoirs of bioactive molecules.

CHAPTER 1: Introduction

Scientists are a lot like coral polyps, maybe not at first glance, but let me paint the picture of their surprising similarities. As individuals, we conduct our experiments and share them with the community- polyps, on their own accomplish incremental levels of production. But as a collective, both the colonial marine invertebrates and us lab dwelling inquirers, pioneer new frontiers. Polyps and colonies of gorgonians work together to form entire reef ecosystems; scientists build off each other's work discovering more and more about the world around us. I can't speak for the motives of the coral, but as scientists I think this is a main component of what drives us: the desire to discover something that no-one else has yet understood, to share it with our community, and to watch while others use our findings to stride into the unknown. We are contributing to something bigger than ourselves, leaving our mark in ink, or rather, etched into the virtual cloud. As I introduce the body of work I've accomplished during my PhD, I aim to weave together the stories of the "polyps" who came before me, scientists whose individual contributions provided the foundation from which I have built, enabling me to explore and colonize new depths of discovery. Just as a coral colony grows without knowing how the passage of time and chemistry of the sea will shape its emergent reef, be it branching or encrusting, scientists, too, cannot foresee how their work will be interpreted, expanded, or transformed. They can only hope that their contributions will serve as steppingstones for the community to build upon.

This story begins with Dr. Bill Fenical, an emeritus professor who started his academic career at the Scripps Institution of Oceanography at UC San Diego in 1973. Dr. Fenical was an early colonizer in the field of marine natural products whose nascent curiosity about the ocean's hidden chemistry helped define the field of marine natural products. His primitive research on red abalone pigments sparked a new line of inquiry: if marine organisms could produce pigments, what other specialized metabolites might they be capable of making? For hundreds of years, plants had been recognized as rich sources of bioactive molecules,¹ but at the time, little was understood about the latent biochemical potential of marine organisms. In parallel, marine ecologists were just beginning to scratch the surface in their understanding of foundational ecological and biotic interactions, such as competition and predation.² Few considered that chemical cues, let alone volatile metabolites, might be mediating the dynamics of these complex marine communities. By merging natural products chemistry and marine ecology, Dr. Fenical unveiled chemical mechanisms underlying the ecological

interactions and, in the process, discovered a teeming wealth of novel scaffolds with potent human bioactivities.

When Dr. Fenical joined the Scripps Institution of Oceanography (SIO) in 1973, he stepped into this uncharted space, beginning with the discovery of unique natural products in temperate and tropical seaweeds in the 1970s,³ and expanding to study the chemistry of gorgonian corals in the early 1980s.⁴ He claims that this pivot into the marine environment was motivated by his willingness “*to jump into the ocean and look at what’s there,*” at a time when few were ready to take such a plunge. He, of course, was not alone in these early explorations. Others, including Paul Scheuer, Dick Moore, Jon Clardy and John Faulkner, helped establish marine natural products as a field. Unlocking a world of chemistry hidden beneath the waves, their work demonstrated that marine organisms were capable of producing complex specialized metabolites, much like their terrestrial counterparts. The chemistry they discovered was strikingly novel, including a marked abundance of halogenated compounds rarely found on land. Dr. Fenical and other early marine natural products scientists built on the foundations laid by ecologists and terrestrial natural products chemists. In turn, their discoveries of novel molecules inspired generations of new scientists.

Marine natural products chemists explored an extraordinarily diverse suite of organisms and structural classes, from polyketides produced by marine bacteria⁵ to marine sponge alkaloids.⁶ Pioneer chemical ecologists such as Valerie Paul and Mark Hay, among others, were intrigued by the novelty of the chemistry being produced, the emerging patterns of which organisms made it and when, and began exploring the functional relevance of these compounds within their ecological contexts. Their work forged a new scientific path that integrated natural products chemistry with ecological function. For example, Dr. Paul conducted projects linking chemical-defense compounds from marine cyanobacteria and macroalgae to ecological outcomes such as larval-settlement inhibition and reef-community shifts,^{7,8} and Dr. Hay, through manipulative field experiments, showed how chemically mediated interactions among seaweeds, herbivores, and corals control population structure, community dynamics, and reef resilience.^{9–11}

In addition to piquing the interest of ecologists, these specialized molecules also drew attention from pharmacology labs interested in their therapeutic potential. From these synergistic collaborations, several

hallmark bioactive compounds stand out. Dr. Fenical, working with pharmacologist Dr. Bob Jacobs from UC Santa Barbara, showed that the pseudopterosins produced by sea whips had potent anti-inflammatory activities.¹² These molecules have since been incorporated by the cosmetic industry as skin-soothing ingredients. The soft coral, *Eleutherobia*, was found to produce eleutherobin, which was able to stabilize polymerized tubulin at levels akin to the anticancer drug taxol.¹³ Further, lophotoxin, produced by red gorgonians of the genus *Leptogorgia*, was identified as a neurotoxin that inhibits nicotinic acetylcholine receptors.^{4,14} And so the colony grows, each scientist's contribution a new branching fragment reaching into the vast blue beyond, rooted in the encrusting foundation beneath it: field ecology informing bioactivity screens, chemical characterization enabling pharmacological insights, and each step leaving a lasting mark for the next generation to build upon.

While the ocean has yielded many other notable bioactive molecules with pharmaceutical applications, the examples are deliberately curated, intentionally restricted to highlight only octocoral terpenoids. These molecules and the corals that produce them have served as the bedrock upon which my work was built. To follow the next steps in this story, it's critical to understand the basics of terpene biosynthesis. Entire careers have been devoted to unraveling these pathways, and while I won't retrace every step here, I'll provide some of the essentials needed for the narrative ahead.

Terpenes, and the related terpenoids, are the most chemically and structurally diverse family of natural product molecules in nature.¹⁵ They are built from the C5 isoprene units isopentenyl diphosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) which are produced through one of two independent pathways: the mevalonate (MEV) pathway, found in most eukaryotes and archaea, or the methylerythritol 4-phosphate (MEP) pathway, found in bacteria and plants.¹⁶ These building blocks are combined in a head-to-tail condensation reaction to form longer precursor chains such as geranyl diphosphate (GPP, C10), farnesyl diphosphate (FPP, C15), and geranylgeranyl diphosphate (GGPP, C20). Terpene synthases or cyclases (TCs) then catalyze cyclization of these precursors to produce an enormous diversity of structures often containing multiple fused rings and stereocenters.¹⁷ Specifically, class I TCs are metal-dependent enzymes that initiate cyclization by cleaving the diphosphate group, generating a highly reactive carbocation intermediate. This reaction is driven by a cluster of magnesium ions coordinated by conserved active site motifs such as the DDXXD and NSD/DTE

sequences. Despite their vast sequence diversity, class I TCs share a common α -helical fold, allowing them to guide substrates through precise rearrangements that yield complex terpene scaffolds.

Despite octocorals harboring a wealth of bioactive terpenoid chemistry, the source of these molecules had long been the subject of intense debate. Some argued it came from associated bacteria, others from *Symbiodinium spp.*, the dinoflagellate photosynthetic symbionts of many corals, and others still from the animals themselves. Yet, without genomic evidence to contextualize terpene biosynthesis, researchers were limited to enzymological approaches to investigate its source. Specifically, early work by Russell G. Kerr from Florida Atlantic University sought to resolve this question through radiolabeled precursor-feeding experiments with purified symbiont isolates from *Pseudopterogorgia elisabethae*. He reported that these isolates produced pseudopterogens and that boiling the cells abolished ^{14}C incorporation, suggesting enzymatic synthesis by the symbionts.¹⁸ Kerr pushed the boundaries of what was technically possible at the time, though later genomic insights would overturn this model, revealing coral encoded terpene cyclases through high-quality long-read assemblies.^{19,20} Just as separate coral colonies expand until their edges meet, sequencing technologies were advancing in parallel to natural products chemistry, each field growing independently before their convergence in the genomic era of discovery.

A major breakthrough, and essentially the founding of the field of octocoral terpene biosynthesis, came in 2022 when the then postdocs Immo Burkhardt¹⁹ in Brad Moore's lab at SIO and Paul Scesa²⁰ in Eric Schmidt's lab at the University of Utah, independently leveraged octocoral sequencing data to reveal that the corals themselves encode TCs within their genomes. This discovery built upon decades of work by individuals: natural product chemists cataloging the chemistry of the sea, biochemists unraveling how TCs catalyze the biosynthesis of elaborate, seemingly impossible terpene scaffolds, and genomic pioneers striving to read the code of life itself. Each followed their own curiosity, working in isolation like individual coral polyps, unaware that together they were building a vast reef of knowledge, layer by layer. Guided by their own passions, Immo and Paul became the next polyps to add to this living structure.

While Dr. Fenical's breakthrough began with a literal plunge into the ocean, bridging the worlds of marine biology and chemistry, Immo was simply trying to keep his head above water during the global

pandemic. In the quiet isolation of lockdown, a time when many felt anything but inspired, he spent his days playing video games and attending virtual seminars. One online talk he joined happened to be about Plasmodium-like parasites that infect corals. Afterwards, a colleague casually mentioned to him that many coral terpenes have anti-malarial activity. This sparked an idea: maybe these compounds served as part of a coral immune defense. Out of curiosity, Immo, with a background in fungal terpene biosynthesis and enzyme cyclization mechanisms, queried publicly available octocoral genomes for hits related to TC encoding genes. He sifted through the low scoring matches, scanning for the essential common TC motifs. He identified several promising candidates and, once able to return to the lab, experimentally validated their function. Structural analysis went on to confirm his findings: a crystal structure of a cembrene A synthase from *Eleutherobia rubra* had a typical α -helical fold of type 1 TCs, with spatially conserved catalytic motifs nearly identical to those of microbial TCs.¹⁹

The discovery of octocoral terpene synthases set the stage for a new radiation of scientific exploration, one surrounding the evolutionary forces shaping terpene cyclase diversification. A framework for understanding these processes has long been established by plant biochemists.^{21,22} In plants, terpene cyclases have diversified through lineage-specific expansions to serve diverse ecological roles, from deterring herbivores and combating pathogens to mediating communication and development. Notably, both plant and octocoral terpene cyclases possess a conserved type-I fold and catalytic motifs that underlie parallel mechanisms of cyclization. Despite evolving independently for hundreds of millions of years, many of the terpenoids produced by these organisms share the same core carbon scaffolds, reflecting convergent evolution toward ecologically active terpene frameworks. Further, plant TCs frequently occur within biosynthetic gene clusters (BGCs) and have been found to have tightly regulated tissue specific expression patterns. Although our understanding of octocoral terpenoid biosynthesis is still emerging, early discoveries reveal striking parallels with the well-characterized terpene pathways of plants.

With high-quality octocoral genomes now in hand, we can begin to ask the next set of questions: How have terpene synthases diversified across octocorals? What evolutionary processes drive their lineage-specific expansion and functional divergence? And how are these enzymes organized within emerging biosynthetic gene

clusters? Together, these questions frame the chapters that follow, tracing the diversification, evolution, and genomic architecture of octocoral terpene biosynthesis.

Building on the discoveries of those before me, the work that follows represents my contribution to the growing body of knowledge, adding my voice and work to the growing reef of knowledge surrounding the genetic and biochemical foundations of coral natural products. Growing up in La Jolla, CA, I had always found my rhythm with the tides. My early curiosities drifted from watching the algae and limpets battle for space in the shoreline to wondering how the rockfish knew to slip behind the giant fronds of kelp for protection. From these beginnings bloomed a lasting reverence for the ocean and its vibrant, dynamic world hidden below the surface. In this thesis, I share the questions I have asked, and some that, with the help of excellent mentors and collaborators, I have begun to answer. My hope is that as you read this, your curiosity will be tickled, you may dare to begin to think inquisitively, and perhaps even take the next step, carrying this body of work forward in a new direction.

1.1 References

- (1) Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, 2 (2), 303–336. <https://doi.org/10.3390/metabo2020303>.
- (2) Connell, J. H. Community Interactions on Marine Rocky Intertidal Shores. *Annual Review of Ecology and Systematics* **1972**, 3, 169–192.
- (3) Sims, J. J.; Fenical, W.; Wing, R. M.; Radlick, P. Marine Natural Products. I. Pacifenol, a Rare Sesquiterpene Containing Bromine and Chlorine from the Red Alga, *Laurencia Pacifica*. *J. Am. Chem. Soc.* **1971**, 93 (15), 3774–3775. <https://doi.org/10.1021/ja00744a041>.
- (4) Fenical, W.; Okuda, R. K.; Bandurraga, M. M.; Culver, P.; Jacobs, R. S. Lophotoxin: A Novel Neuromuscular Toxin from Pacific Sea Whips of the Genus *Lophogorgia*. *Science* **1981**, 212 (4502), 1512–1514. <https://doi.org/10.1126/science.6112796>.
- (5) Feling, R. H.; Buchanan, G. O.; Mincer, T. J.; Kauffman, C. A.; Jensen, P. R.; Fenical, W. Salinosporamide A: A Highly Cytotoxic Proteasome Inhibitor from a Novel Microbial Source, a Marine Bacterium of the New Genus *Salinospora**.
- (6) Hirata, Y.; Uemura, D. Halichondrins - Antitumor Polyether Macrolides from a Marine Sponge. *Pure and Applied Chemistry* **1986**, 58 (5), 701–710. <https://doi.org/10.1351/pac198658050701>.
- (7) J. Paul, V.; P. Puglisi, M.; Ritson-Williams, R. Marine Chemical Ecology. *Natural Product Reports* **2006**, 23 (2), 153–180. <https://doi.org/10.1039/B404735B>.
- (8) Paul, V. J.; Puglisi, M. P. Chemical Mediation of Interactions among Marine Organisms. *Nat. Prod. Rep.* **2004**, 21 (1), 189–209. <https://doi.org/10.1039/B302334F>.

- (9) Hay, M. E. Marine Chemical Ecology: What's Known and What's Next? *Journal of Experimental Marine Biology and Ecology* **1996**, 200 (1), 103–134. [https://doi.org/10.1016/S0022-0981\(96\)02659-7](https://doi.org/10.1016/S0022-0981(96)02659-7).
- (10) Hay, M. E. Marine Chemical Ecology: Chemical Signals and Cues Structure Marine Populations, Communities, and Ecosystems. *Ann Rev Mar Sci* **2009**, 1, 193–212. <https://doi.org/10.1146/annurev.marine.010908.163708>.
- (11) Hay, M. E.; Fenical, W. Marine Plant-Herbivore Interactions: The Ecology of Chemical Defense. *Annual Review of Ecology and Systematics* **1988**, 19, 111–145.
- (12) Look, S. A.; Fenical, W.; Jacobs, R. S.; Clardy, J. The Pseudopterosins: Anti-Inflammatory and Analgesic Natural Products from the Sea Whip Pseudopterogorgia Elisabethae. *Proc Natl Acad Sci U S A* **1986**, 83 (17), 6238–6240.
- (13) Long, B. H.; Carboni, J. M.; Wasserman, A. J.; Cornell, L. A.; Casazza, A. M.; Jensen, P. R.; Lindel, T.; Fenical, W.; Fairchild, C. R. Eleutherobin, a Novel Cytotoxic Agent That Induces Tubulin Polymerization, Is Similar to Paclitaxel (Taxol). *Cancer Res* **1998**, 58 (6), 1111–1115.
- (14) Culver, P.; Fenical, W.; Taylor, P. Lophotoxin Irreversibly Inactivates the Nicotinic Acetylcholine Receptor by Preferential Association at One of the Two Primary Agonist Sites. *J Biol Chem* **1984**, 259 (6), 3763–3770.
- (15) Tetali, S. D. Terpenes and Isoprenoids: A Wealth of Compounds for Global Use. *Planta* **2019**, 249 (1), 1–8. <https://doi.org/10.1007/s00425-018-3056-x>.
- (16) Hoshino, Y.; Gaucher, E. A. On the Origin of Isoprenoid Biosynthesis. *Mol Biol Evol* **2018**, 35 (9), 2185–2197. <https://doi.org/10.1093/molbev/msy120>.
- (17) Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **2017**, 117 (17), 11570–11648. <https://doi.org/10.1021/acs.chemrev.7b00287>.
- (18) Mydlarz, L. D.; Jacobs, R. S.; Boehnlein, J.; Kerr, R. G. Pseudopterosin Biosynthesis in Symbiodinium Sp., the Dinoflagellate Symbiont of Pseudopterogorgia Elisabethae. *Chemistry & Biology* **2003**, 10 (11), 1051–1056. <https://doi.org/10.1016/j.chembiol.2003.10.012>.
- (19) Burkhardt, I.; de Rond, T.; Chen, P. Y.-T.; Moore, B. S. Ancient Plant-like Terpene Biosynthesis in Corals. *Nat Chem Biol* **2022**, 18 (6), 664–669. <https://doi.org/10.1038/s41589-022-01026-2>.
- (20) Scesa, P. D.; Lin, Z.; Schmidt, E. W. Ancient Defensive Terpene Biosynthetic Gene Clusters in the Soft Corals. *Nat Chem Biol* **2022**, 18 (6), 659–663. <https://doi.org/10.1038/s41589-022-01027-1>.
- (21) Jia, Q.; Brown, R.; Köllner, T. G.; Fu, J.; Chen, X.; Wong, G. K.-S.; Gershenzon, J.; Peters, R. J.; Chen, F. Origin and Early Evolution of the Plant Terpene Synthase Family. *Proceedings of the National Academy of Sciences* **2022**, 119 (15), e2100361119. <https://doi.org/10.1073/pnas.2100361119>.
- (22) Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The Family of Terpene Synthases in Plants: A Mid-Size Family of Genes for Specialized Metabolism That Is Highly Diversified throughout the Kingdom. *The Plant Journal* **2011**, 66 (1), 212–229. <https://doi.org/10.1111/j.1365-313X.2011.04520.x>.

CHAPTER 2: Leveraging Deep Sea Octocoral to Explore Diversification of Diterpene Biosynthesis

2.1 Introduction for Chapter 2

Octocoral diverged from their sister lineage, the hexacorals, ca. 770 million years ago.¹ Since that split, these classes have followed starkly different evolutionary paths. Corals in the class hexacoral (order *Scleractinia*) formed tight partnerships with photosynthetic symbionts, largely tying them to shallow tropical waters where sunlight fuels their symbionts' productivity.²⁻⁴ Octocorals, on the other hand, have radiated into nearly every oceanic habitat, from coastal reefs to the deep sea.⁵ A main driver of the distinct evolutionary trajectories of reef building corals vs octocorals has been their predation defense strategies. Nematocytes, or stinging cells, are a defining feature of the phylum *Cnidaria*, yet the capacity to produce toxic venoms in these cell types seems restricted to the subphyla *Medusozoa* and *Anthozoa*.^{6,7} Within *Anthozoa* hexacorals retain venomous nematocysts, octocorals do not. Instead, octocorals have evolved the capacity to produce terpenoid specialized chemistry,^{8,9} which act as feeding deterrents against generalist predators.^{10,11} In contrast, hexacorals rely primarily on physical defenses, most notably the evolution of calcified skeletons that provide both structural support and protection from grazers.¹²

Across kingdoms, the evolution of novel metabolic pathways correlates with rapid niche diversification and habitat expansion.^{13,14} Octocorals encompass a staggering diversity of morphologies, reproductive modes, and habitats, raising the fundamental questions about the role of chemistry in shaping their evolutionary trajectory. By exploring the evolution of octocoral terpenoid production, and linking these patterns to ecological and evolutionary context, we can begin to understand whether the production of unique specialized metabolites has actively driven species diversification.

Octocorals are particularly rich in diterpenoids, most of which fall into five broad structural classes: xenicanes, briaranes, pseudopterosins, lophotoxins, and eleutherobins. Despite this diversity, our current understanding of octocoral chemistry is heavily skewed toward shallow-water species. Given the relative ease of accessing shallow reefs, and the appeal of trips to beautiful tropical locations, it is not surprising that these species have been the focus of most research efforts. In contrast, deep-sea species remain largely unexplored due to the logistical and financial challenges of sampling at great depths. However, with the rapid expansion of deep-sea mining and other anthropogenic pressures, there is a growing urgency to catalog the chemical

diversity of deep-sea octocorals before it is lost. To address this gap, a research expedition led by the Jensen Lab deployed remotely operated vehicles (ROVs) to sample corals from a series of offshore seamounts and ridges in the Southern California Borderland, at depths ranging from 110 to 2,400 meters. We worked to assess both the genomic potential and chemical repertoire for terpene production in the samples, uncovering hidden diversity in deep-sea octocorals and providing a foundation for understanding how these compounds fit into the broader ecological and evolutionary context of the group.

In this chapter, we pair genomic and chemical approaches to investigate the terpenoid biosynthetic potential of deep-sea octocorals. We use Illumina short-read sequencing to generate genomic data for nine specimens spanning both major octocoral orders, including previously unstudied genera such as *Chrysogorgia*, *Victorgorgia*, *Callistephanus*, and *Balticina*. We biochemically characterize the terpene cyclases (TCs) present in their genomes and profile the terpene chemistry present in the coral tissue. This work was also a formative stage in my development as a researcher: through these efforts, I learned molecular cloning techniques to experimentally link genes to their products. During this cloning process, I also began to see firsthand the limitations of relying solely on short-read sequencing data. While we could find TCs in our assemblies, we could not always determine how many distinct copies existed, whether they were in genomic neighborhoods with other biosynthetic genes, or how they were organized across haplotypes. Functional annotation of these biosynthetic genes revealed divergent evolutionary trajectories, suggesting that the capacity for diterpene biosynthesis emerged early in octocoral evolution and has been conserved yet diversified across clades. In this work, we begin to uncover how specialized chemistry has shaped octocoral ecology and contributed to their remarkable evolutionary success.

The work presented in this chapter was carried out in collaboration with Dr. Immo Burkhardt, Hannah K. Bone, Helena A. Leucke, Johanna Gutleben, Paul R. Jensen, Andrea M. Quattrini, Alexander B. Chase, and Bradley S. Moore. I am presenting the work for this chapter in its entirety, given the synergy of the analyses and the cohesiveness of the manuscript but wish to state forthrightly the portions to which I directly contributed. My primary contributions included molecular cloning and heterologous expression of TC enzymes on preparative and analytical scales, as well as structural characterization and elucidation of the resulting terpene products. Hannah K. Bone and I also contributed to figure design and overall manuscript organization, while Dr.

Immo Burkhardt served as the primary author of the text. The DNA extraction was conducted by Johanna Gutleben, and the codon usage analysis was performed by Alexander B. Chase.

2.2 Chapter 2 Introduction References

- (1) McFadden, C. S.; Ofwegen, L. P. van; Quattrini, A. M. Revisionary Systematics of Octocorallia (Cnidaria: Anthozoa) Guided by Phylogenomics. *Bull. Soc. Syst. Biol.* **2022**, *1* (3). <https://doi.org/10.18061/bssb.v1i3.8735>.
- (2) McIntyre, A. *Life in the World's Oceans: Diversity, Distribution, and Abundance*; John Wiley & Sons, 2010.
- (3) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J.-M.; Hartmann, G.; Banner, D. W.; Haap, W.; Diederich, F. Systematic Investigation of Halogen Bonding in Protein–Ligand Interactions. *Angew. Chem. Int. Ed.* **2011**, *50* (1), 314–318. <https://doi.org/10.1002/anie.201006781>.
- (4) Cairns, S. D. Deep-Water Corals: An Overview with Special Reference to Diversity and Distribution of Deep-Water Scleractinian Corals. *Bull. Mar. Sci.* **2007**, *81* (3).
- (5) Quintanilla, E.; Rodrigues, C. F.; Henriques, I.; Hilário, A. Microbial Associations of Abyssal Gorgonians and Anemones (>4,000 m Depth) at the Clarion-Clipperton Fracture Zone. *Front. Microbiol.* **2022**, *13*. <https://doi.org/10.3389/fmicb.2022.828469>.
- (6) Fautin, D. G. Structural Diversity, Systematics, and Evolution of Cnidae. *Toxicon* **2009**, *54* (8), 1054–1064. <https://doi.org/10.1016/j.toxicon.2009.02.024>.
- (7) Schmidt, C. A.; Daly, N. L.; Wilson, D. T. Coral Venom Toxins. *Front. Ecol. Evol.* **2019**, *7*. <https://doi.org/10.3389/fevo.2019.00320>.
- (8) Scesa, P. D.; Lin, Z.; Schmidt, E. W. Ancient Defensive Terpene Biosynthetic Gene Clusters in the Soft Corals. *Nat. Chem. Biol.* **2022**, *18* (6), 659–663. <https://doi.org/10.1038/s41589-022-01027-1>.
- (9) Burkhardt, I.; de Rond, T.; Chen, P. Y.-T.; Moore, B. S. Ancient Plant-like Terpene Biosynthesis in Corals. *Nat. Chem. Biol.* **2022**, *18* (6), 664–669. <https://doi.org/10.1038/s41589-022-01026-2>.
- (10) Barsby, T.; Kubanek, J. Isolation and Structure Elucidation of Feeding Deterrent Diterpenoids from the Sea Pansy, *Renilla reniformis*. *J. Nat. Prod.* **2005**, *68* (4), 511–516. <https://doi.org/10.1021/np049609u>.
- (11) Giordano, G.; Carbone, M.; Ciavatta, M. L.; Silvano, E.; Gavagnin, M.; Garson, M. J.; Cheney, K. L.; Mudianta, I. W.; Russo, G. F.; Villani, G.; Magliozzi, L.; Polese, G.; Zidorn, C.; Cutignano, A.; Fontana, A.; Ghiselin, M. T.; Mollo, E. Volatile Secondary Metabolites as Aposematic Olfactory Signals and Defensive Weapons in Aquatic Environments. *Proc. Natl. Acad. Sci.* **2017**, *114* (13), 3451–3456. <https://doi.org/10.1073/pnas.1614655114>.
- (12) Stolarski, J.; Kitahara, M. V.; Miller, D. J.; Cairns, S. D.; Mazur, M.; Meibom, A. The Ancient Evolutionary Origins of Scleractinia Revealed by Azooxanthellate Corals. *BMC Evol. Biol.* **2011**, *11* (1), 316. <https://doi.org/10.1186/1471-2148-11-316>.

- (13) Kitchen, S. A.; Naragon, T. H.; Brückner, A.; Ladinsky, M. S.; Quinodoz, S. A.; Badroos, J. M.; Viliunas, J. W.; Kishi, Y.; Wagner, J. M.; Miller, D. R.; Yousefelahiyeh, M.; Antoshechkin, I. A.; Eldredge, K. T.; Pirro, S.; Guttman, M.; Davis, S. R.; Aardema, M. L.; Parker, J. The Genomic and Cellular Basis of Biosynthetic Innovation in Rove Beetles. *Cell* **2024**, *187* (14), 3563-3584.e26. <https://doi.org/10.1016/j.cell.2024.05.012>.
- (14) Weng, J.-K.; Lynch, J. H.; Matos, J. O.; Dudareva, N. Adaptive Mechanisms of Plant Specialized Metabolism Connecting Chemistry to Function. *Nat. Chem. Biol.* **2021**, *17* (10), 1037–1045. <https://doi.org/10.1038/s41589-021-00822-6>.

2.3 Abstract

Octocorals are the major source of marine-derived bioactive terpenoids. However, the vast majority of explored chemistry is known from shallow-water species, leaving deep-sea octocorals largely unexplored. Recent genomic work uncovered terpene biosynthetic pathways encoded in octocoral genomes, enabling deeper investigation into the evolution and ecological distribution of these compounds. Here, we collected nine deep-sea octocoral specimens representing both taxonomic orders and profiled their terpenoid chemistry. These samples revealed extensive diversity in sesquiterpene scaffolds, along with repeated detection of five widespread structural families of diterpenes (xeniaphyllene-, cembrene B-, elisabetatriene-, cembrene A-, and klysimplexin R; collectively, “XBECK-type” diterpenoids). Phylogenetic analysis of terpene cyclase (TC) sequences from these samples, combined with publicly available sequences and functional characterization of selected genes, indicated that most TC enzymes encode functionally diverse sesquiterpene cyclases that corresponded to deeply-rooted taxonomic origins, rather than biochemical function. Nonetheless, we identified five monophyletic clades, each comprising isofunctional enzymes that produce precursors for distinct XBECK-type diterpenoids, each showing broad representation across octocoral taxa. These findings suggest that diverse diterpenoid biosynthesis evolved early in coral evolution, with the last common ancestor already possessing multiple functionally distinct TC enzymes. Our results establish terpenoid production as an ancient, central trait of octocorals, irrespective of habitat, thereby highlighting the tremendous potential of deep-sea corals as sources for bioactive terpenoids. Finally, these findings raise intriguing questions about the evolutionary origins of these pathways within early cnidarians and across animal phyla more broadly.

2.4 Introduction

Octocorals represent half of all coral species worldwide, totaling approximately 3500 species.¹ Since the split from their sister lineage ca. 770 million years ago, octocorals have colonized nearly all oceanic environments, from shallow tropical reefs to the abyssal zone below 4000 m,² with 75% of species living deeper than 50 m.^{3,4} In contrast to hexacorals (stony corals, black corals, sea anemones), octocorals are known as a prolific source of specialized metabolites, constituting the largest known source of marine-derived terpenoids.^{5,6} Over 4000 octocoral-derived terpenoids have been described, many of which are of considerable medicinal and ecological interest due to their potent biological properties.^{7,8} Prominent examples include cristaxenicin (**1**), with nanomolar antiplasmodial activity,⁹ renillafulin A (**2**), which inhibits barnacle larval settlement,¹⁰ the anti-inflammatory pseudopterosin A (**3**),¹¹ the neuromuscular toxin lophotoxin (**4**),¹² and eleutherobin (**5**), known for pharmacological activity similar to the anti-cancer drug Taxol (Figure 1).¹³ A large number of the octocoral terpenoids identified to date are diterpenoids that are structurally related to compounds **1-5**.¹⁴ However, characterization of coral-derived terpenoids has disproportionately focused on shallow-water tropical species, with fewer than 100 compounds (~2.5%) known from cold- and deep-water species.¹⁵⁻¹⁷ This disparity contrasts markedly with the numerous deep-sea-specific species.^{3,18} The primary reasons for this imbalance include difficulty accessing cold- and deep-water corals, which typically require specialized and costly equipment, such as remotely operated vehicles (ROVs). Moreover, the slow growth of deep-sea corals further restricts the collection of sufficient biomass for chemical studies.^{19,20}

Recently, we¹⁴ and Scesa *et al.*²¹ demonstrated that octocoral genomes harbor genes coding for terpenoid biosynthetic pathways, dispelling earlier hypotheses of microbial symbiont origins. Octocoral genomes encode multiple enzyme classes essential for terpenoid biosynthesis, such as short chain dehydrogenases, cytochrome P450s, and a phylogenetically distinct lineage of class I terpene cyclases (TCs). Class I TCs catalyze the first step in terpenoid biosynthesis, converting readily available oligoprenyl diphosphate precursors from central metabolism into specialized hydrocarbon structures.²² Coral TCs are monophyletic and broadly distributed across both octocoral orders, the Scleralcyonacea and Malacalcyonacea, which diverged from their last common ancestor around 540 million years ago.^{14,23,24} Notably, some of the initially characterized coral TCs produce diterpene (C₂₀-scaffolds) or mono-alcohols closely

resembling the principal carbon scaffolds of compounds **1**, **3**, **4**, and **5**: xeniaphyllene (**6**), elisabethatriene (**8**), cembrene A (**9**), and klysimplexin R (**10**), respectively (Figure 1).^{14,21} These terpene hydrocarbons had previously been hypothesized^{25–27} or experimentally linked²⁸ to higher oxidized terpenoids akin to compounds **1**, **2**, **3**, and **5** prior to genetic studies. We recently identified cembrene B (**7**) as the first biosynthetic intermediate for the briarane diterpenoids, including compound **2**.²⁹ Cembrene B synthases are found within conserved biosynthetic gene clusters co-localized with genes encoding oxidative enzymes that construct the distinctive g-lactone structural motif conserved in almost all briarane terpenoids. The cembrene B synthases were monophyletic and only detected within the order Scleractyonia. Beyond this recent finding, limited information exists on the evolutionary history and taxonomic distribution of coral TC genes, primarily due to sparse sequencing data at the time of initial studies.^{14,21}

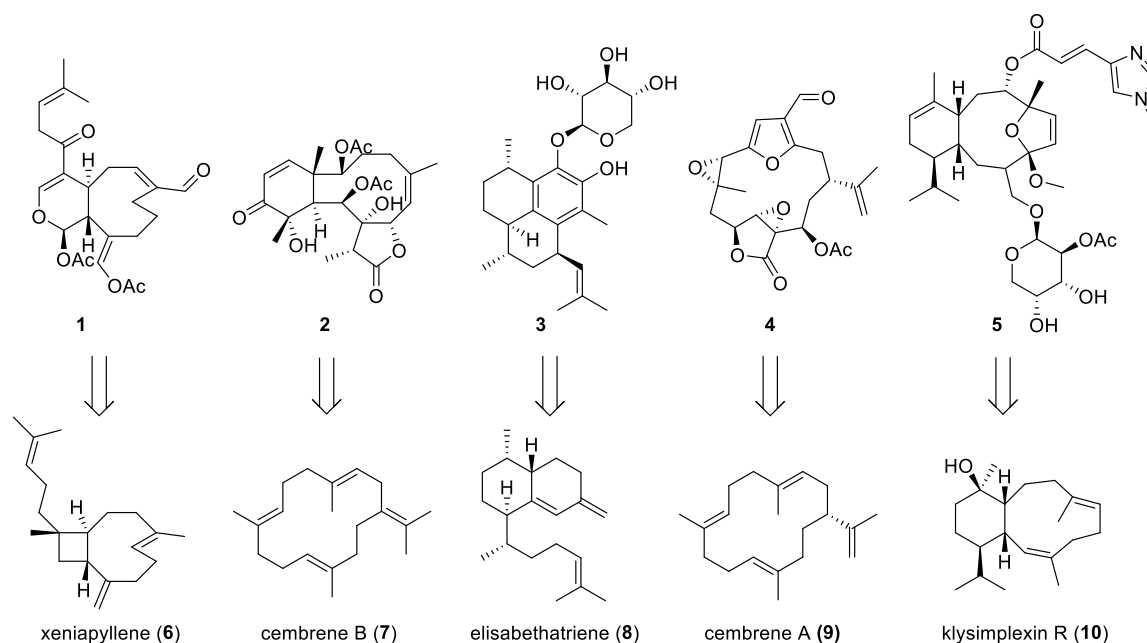


Figure 2-1: Examples of bioactive octocoral diterpenoids from the five largest structural classes and their biosynthetic precursors produced by octocoral terpene cyclases.

Leveraging new capabilities to connect genomic sequences with biosynthetic potential,³⁰ we investigated the chemical and genetic diversity of terpenoid production in deep-sea octocorals. Our results show that deep-sea octocorals harbor a rich diversity of terpenoids comparable to shallow-water species. All but one investigated deep-sea species contained detectable terpenes within their tissues, and all possessed class I TC genes within their genomes. We performed the first comprehensive phylogenetic analysis for octocoral TC genes, including 40 novel sequences obtained from nine deep-sea species and 137 sequences

obtained from publicly available datasets spanning 39 species. Functional annotation of the enzymes revealed divergent evolutionary trajectories, suggesting the early emergence and conservation of diterpene biosynthesis across octocoral lineages.

2.5 Results

Sample collection and taxonomic identification

Deep sea octocoral samples were collected using remotely operated vehicles (ROV) from a series of offshore seamounts and ridges in the Southern California Borderland (depths range from 110 to 2400 meters; Table S1) during two scientific cruises in 2020 and 2021. Nine specimens were initially characterized by morphological traits (Figure 2A) and validated using phylogenomic analyses (Figure 2B). Resulting taxonomic assignments spanned both orders Scleralcyonacea and Malacalcyonacea and included two *Chrysogorgia* spp., two *Paragorgia* spp., two *Callistephanus simplex*, a *Victorgorgia* sp., an *Acanthogorgia* sp., and a *Balticina* sp. (Figure 2A, Figure S1). Notably, *Chrysogorgia*, *Victorgorgia*, *Callistephanus*, and *Balticina* were previously unexplored with regards to their chemistry (Table S1).

Metabolomic exploration of samples and distribution of terpenoid types within octocorals

Metabolomic analysis of the nine deep-sea octocoral samples using gas chromatography-mass spectrometry (GC-MS) revealed substantial variability in sesquiterpene and diterpene hydrocarbon production. In one case only, *Balticina* sp. NA124-218, no terpenes were detected. In contrast, *Chrysogorgia* sp. SCB-144 produced a notably rich profile, with the identification of 22 sesquiterpenes and two diterpenes (Figs S2-S10). Overall, terpene production appeared to be driven more by fine-scale taxonomic relatedness than geographic location. For instance, the *C. simplex* specimens, despite being sampled from different collection sites (Table S1), exhibited similar terpene profiles, consistent with species-specific chemical signatures previously observed in shallow-water corals such as *Sarcophyton*.³¹ However, these species-specific patterns were not conserved among closely related species (i.e., within genus). Minimal overlap in terpene profiles was observed between *Paragorgia jamesi* and *Paragorgia arborea*, as well as between the two distinct, *Chrysogorgia* species.

Although overall terpene profiles were divergent across specimens, certain diterpenes exhibited broad taxonomic distribution. In particular, xeniaphyllene (**6**) was detected in distantly related corals from both

orders, Malacalcyonacea (*Acanthogorgia* sp., *Victorgorgia* sp.) and Scleralcyonacea (*Paragorgia* sp., *Chrysogorgia* spp.) (Figure 2B), suggesting that some diterpenes may be conserved or convergently evolved across octocoral lineages. The detected diterpenes were largely restricted to known coral compounds, xeniaphyllene (**6**), elisabethatriene (**8**) and cembrene A (**9**) (all confirmed by identical mass spectra to standards;¹⁴ Fig S2-S10). Only *Chrysogorgia* sp SCB-157 produced previously uncharacterized diterpenes (Fig. S9). In contrast, we observed far greater structural diversity among sesquiterpenes, including numerous uncharacterized compounds (Figs. S2–S10). These patterns prompted us to examine whether terpenoid structural diversity more broadly reflects phylogenetic relationships within octocorals.

To investigate this, we expanded our analysis by incorporating published reports of diterpenoid production across octocorals, focusing on biosynthetic precursor relationships to compare diterpenoid types in a phylogenetic context. The structural diversity of octocoral diterpenoids is immense, and numerous principal carbon scaffolds have been described.⁸ However, many structurally distinct diterpenoids likely arise from common biosynthetic intermediates, with scaffold diversification occurring in late-stage tailoring steps, as suggested by the frequent co-isolation of structurally related compounds or by chemical interconversion.^{26,27,32} Accordingly, we categorized diterpenoids into five types based on their presumed biosynthetic precursor rather than the final structure, which corresponded to major known coral diterpenoid scaffolds (Figure 1, Figure 2C; see figure legend for details). Together, these five groups, referred to here as the XBECK-structure types, account for more than half of the over 4000 reported octocoral terpenoids, including the examples in Figure 1.³³ Four of the five diterpene types were broadly distributed across both octocoral orders, while the B-type was restricted to Scleralcyonacea (Figure 2B, Table S2), consistent with our recent finding that cembrene B synthases form a monophyletic clade within this order.²⁹ The widespread taxonomic distribution of the other diterpene types suggested a distinct evolutionary history, prompting phylogenetic and functional analysis of the corresponding terpene synthases to investigate their biosynthetic origins.

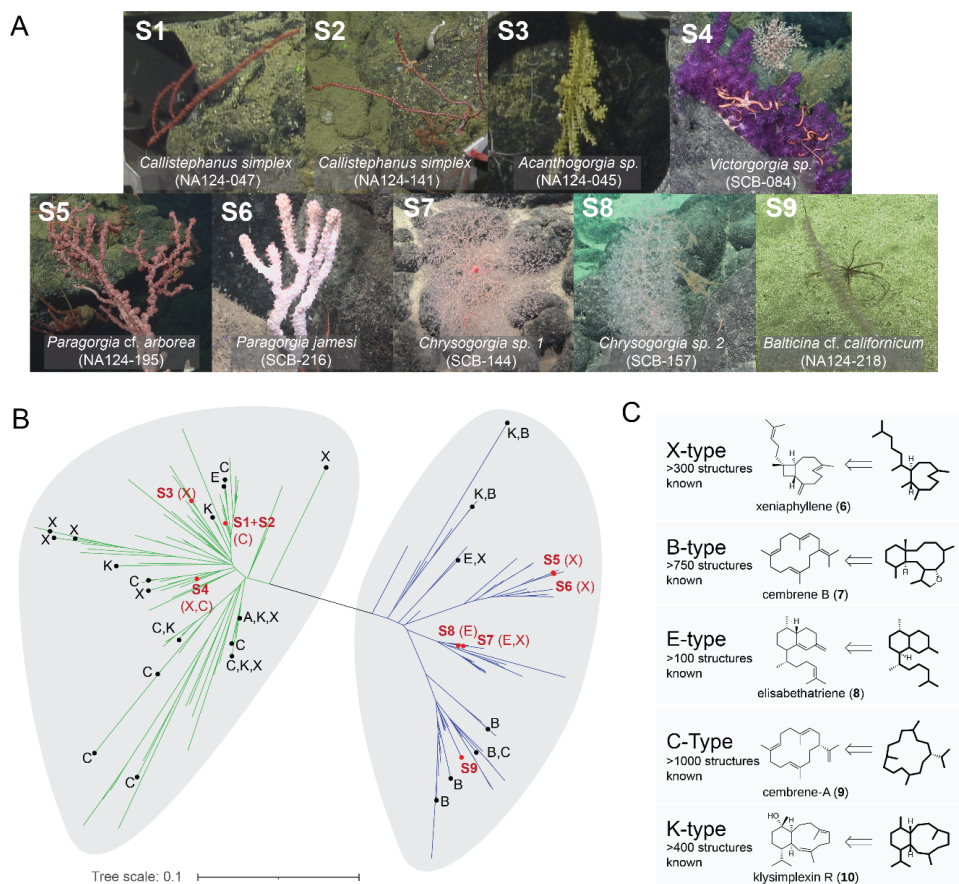


Figure 2-2: Taxonomic classification of deep-sea octocoral samples and distribution of diterpenoid compound types throughout octocorals.

A: In situ photographs of the deep-sea octocoral specimens analyzed in this study (S1-S9), labeled with their determined species identification; **B:** Phylogenomic reconstruction of the nine octocoral mitogenomes sequenced in this work (highlighted in red labels corresponding to Panel A) alongside reference sequences from a published database⁴⁷. Branches are colored by taxonomic order: green for Malacalcyonacea, blue for Scleralcyonacea. Where applicable, detected diterpenoids are annotated by type (see Panel C). For a full representation of the Maximum Likelihood tree including bootstraps, see Figure S2; red font indicates diterpenes characterized from the deep-sea specimens in this study, black font denotes compounds previously reported in the literature; **C:** Biosynthetic classification of octocoral diterpenoids based on their X-type (xeniaphyllene-derived), B-type (cembrene B-derived), E-type (elisabethatriene-derived), C-type (cembrene A-derived), and K-type (kysimplexin R-derived) structures, collectively referred to as XBECK-structure types.

Terpene synthase phylogeny and distribution

All deep-sea coral genome assemblies generated in this work contained TC genes, their number ranging from one to seven per dataset. Phylogenetic analysis of these and publicly available octocoral TC genes revealed four major clades, with distinct taxonomic patterns across clades (Figure 3A, Figure S11, Supplementary file 1). Clades 1 and 3 contained TCs from both octocoral orders, whereas clade 2 was

restricted to Scleralcyonacea and clade 4 to Malacalcyonacea. Clade 1 was further divided into five monophyletic subclades (I-V), with subclades 1-I, 1-II and 1-IV containing representatives from both orders. In contrast, clades 2, 3 and 4 were constrained in their taxonomic distributions, with closely related sequences typically originating from the same family or genus (Figure S11). TC sequences derived from the deep-sea octocoral genomes generated in this study were distributed across all four clades as well as all five subclades in clade 1.

Functional annotation of previously characterized enzymes mapped onto the phylogenetic tree revealed apparent functional conservatism within clade 1 (Figure 3B). This clade exclusively harbored diterpene synthases, including elisabethatriene (**8**) synthases in subclade 1-I, klysimplexin R (**10**) synthases in 1-II, cembrene A (**9**) synthases in 1-III, and cembrene B (**7**) synthases in 1-V.^{14,21,29,34,35} Subclade 1-IV contained no yet characterized enzymes. The remaining clades (2, 3, and 4) primarily contained sesquiterpene synthases, but also included a xeniaphyllene (**6**) synthase in clade 4, and two cembrene A (**9**) synthases in clades 2 and 3. However, no clear relationship between TC product type and phylogenetic placement was initially observed within these clades.

***In vitro* characterization of TCs from deep-sea samples**

To investigate whether the phylogenetic structure of octocoral TCs reflects enzyme function, we characterized additional TCs derived from the nine deep-sea specimens. Targeted genes were heterologously expressed in *Escherichia coli*, and purified recombinant enzymes (Figure S12) were then incubated with the terpene precursors geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) to test for mono-, sesqui- and diterpene production, respectively. We sampled two additional enzymes in each of the clades 1-I through 1-V to assess if these enzymes will predictably generate the diterpene products previously described from members of these clades. The only exception was clade 1-II, where we only identified one new sequence. Upon functional characterization, the additional enzymes from clade 1-I produced **8**; the one additional enzyme from clade 1-II produced **10**; the two additional enzymes from clade 1-III produced **9**; and two additional enzymes from clade 1-V produced **7** (Figure 3B, Figures S13-34). Clade 1-IV contained no previously characterized enzymes, but both representatives characterized here produced **6**. We also observed

two paralogous sequences from the *Balticina* cf. *californicum* genome, that were closest to subclade 1-III, but separated by a fairly long branch and were hence not included in one of the subclades. We targeted both enzymes for characterization, one enzyme selectively producing hinesene (**11**) (Fig S34), a compound known from plants but not previously reported in corals. This enzyme represents the first known hinesene synthase. The second *Balticina* cf. *californicum* TC failed to overexpress in *E. coli*. These results suggest that clade 1, excluding the two *Balticina* cf. *californicum* TCs, consists of broadly distributed, functionally conserved diterpene synthases that likely contribute to the widespread occurrence of common diterpenoids across octocorals. Since the sampling density varied for each subclade, we generated ancestor sequences for each of the ancestral nodes from subclades 1-I to 1-V to see if the ancestor of all identified extant sequences generates the same products as the individually tested enzyme representatives. Indeed, each ancestor reliably produced the same diterpene as the characterized extant enzymes from each clade, the products being elisabethatriene (**8**), klysimplexin R (**10**), cembrene A (**9**), xeniaphyllene (**6**), and cembrene B (**7**) for the ancestors of clades 1-I, 1-II, 1-III, 1-IV, and 1-V, respectively (Figures S35-S39).

Unlike clade 1, clades 2, 3 and 4 showed greater variation in product diversity, with most previously characterized enzymes functioning as sesquiterpene synthases (Figure 3B, 3C and S11) and not showing obvious functional trends matching their phylogenetic relationship. We sampled across these three clades and picked 15 additional enzymes identified in the deep sea octocorals. We aimed to cover a large phylogenetic space while also picking some closely related enzymes for functional characterization. Adding these new 15 enzymes to the 9 previously characterized (Figure 3B, Figure S11) did not reveal significant functional conservation corresponding to phylogeny. Even enzymes with high sequence identity produced different products, for example the **12** and **13** synthases in clade 2 (79% amino acid identity) and the **18**, **19**, **21**, and **13** synthases in clade 4 (57-79% amino acid identity) (Tables S3-S7). Nine out of the total 24 characterized enzymes from these clades produced multiple major products, highlighting increased enzymatic promiscuity compared to clade 1. We also identified an xeniaphyllene (**6**) synthase in clade 3, indicating that **6** synthases occur in three of the four major TC clades. Together, these results fit the observed coral terpene profiles, with clade 1 enzymes showing conserved, predictable diterpene activity, and enzymes in clades 2, 3, and 4 displaying increased product promiscuity and functional divergence, mostly producing sesquiterpenes.

A few diterpene synthases were not located within clade 1. This included two previously characterized cembrene A (**9**) synthases from clades 2 and 3 and two xeniaphyllene (**6**) synthases one previously described, one discovered here, from clades 3 and 4. These enzymes clustered in distant, non-homologous lineages relative to the corresponding clade 1 synthases. To test whether these enzymes produced truly identical products, we determined their absolute configurations. The products of cembrene A (**9**) synthases were compared to authentic standards of (–)-**9**¹⁴ and (+)-**9**³⁶ by GC-MS on a homochiral stationary phase. This analysis showed that all cembrene A (**9**) synthases from clade 1-III produced (–)-**9**, whereas the two enzymes from clades 2 and 3 produced (+)-**9**, proving their products to be distinct from enzymes in clade 1-III (Figure 3B and C, Figure S40). Conversely, optical-rotation measurements of preparative amounts of xeniaphyllene (**6**) obtained from enzymes in clade 1-IV and clade 3 identified them as (–)-**6**, and hence proved them identical to the product of the previously characterized **6** synthase from clade 4.¹⁴ These results revealed that all **6** synthases identified so far form identical products, irrespective of their phylogeny or sequence identity.

Lastly, we analyzed the chemical extraction data for the presence of compounds we identified in the in vitro characterization of enzymes. Whenever we characterized an elisabethatriene (**8**) synthase, a cembrene A (**9**) synthase, or a xeniaphyllene (**6**) synthase, we found the respective terpene product in the chemical extract of the specimen, with the exception of *Acanthogorgia* sp. NA124-045, where a **9** synthase was characterized, but the product was not observed in the extract (Table S8). We did not observe klysimplexin R (**10**) or cembrene B (**7**) in any sample, even when we characterized an active diterpene cyclase from that specimen. Generally, the presence/absence correlation between detected diterpenes and a diterpene synthase in the predicted diterpene subclades 1-I through 1-V was strong and could be extended to species whose sequencing information was analyzed in this work and that had chemical data published in the literature (Table S8). The only exception for this trend was the xeniaphyllene (**6**) production in *Chrysogorgia* sp. SCB-144, where we could not identify a sequence that fell into subclade 1-IV, but identified a phylogenetically distant **6** synthase with identical activity as the clade 1-IV enzymes. As many diterpene cyclases showed some degree of sesquiterpene synthase activity (Figure S13-S34) we also analyzed the sesquiterpene content of the samples. Only in a few cases could we identify sesquiterpenes from in vitro assays in the chemical extracts of the same specimen, the examples being germacrene C (**14**) in *Paragorgia arborea* NA124-195 (Figure S5), germacrene D (**13**) in

Chrysogorgia sp. SCB-144 (Figure S8), and aristolene (**12**) production in *Chrysogorgia* sp. SCB-157 (Figure S9). While germacrene D (**13**) was detected as one of the products generated by elisabethatriene (**8**) synthases, we also identified a selective **13** synthase in *Chrysogorgia* sp. SCB-144. Generally, the diterpene synthases produced multiple sesquiterpene products when incubated with FPP, while showing selective diterpene cyclase activity (Figure S13-S39). We did not observe a co-occurrence of the *in vitro* sesquiterpene blends with the diterpene products of enzymes in subclades 1-I to 1-V within the extracts of the deep-sea coral samples, pointing to their selectivity for diterpene production *in vivo*.

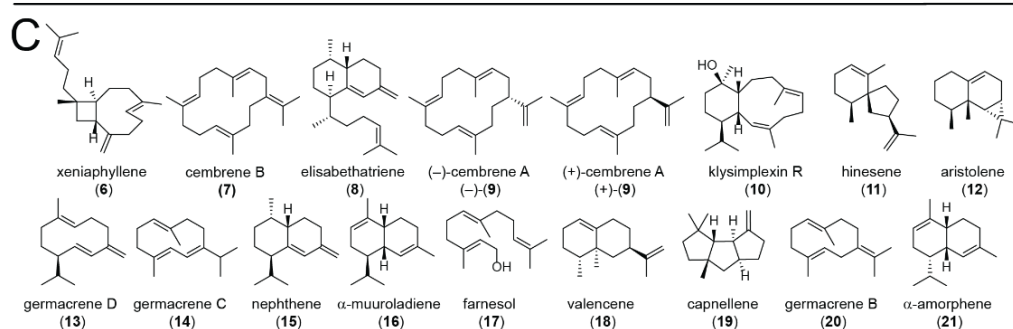
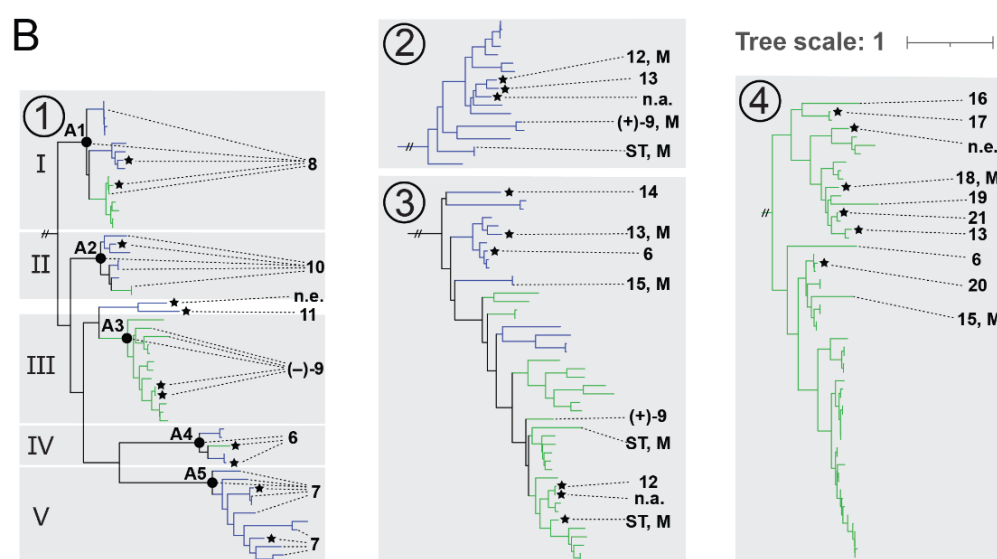
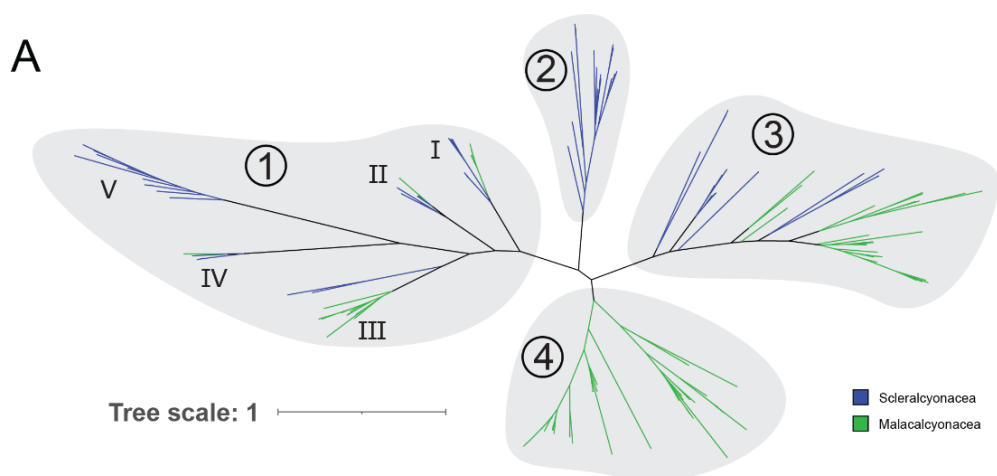


Figure 2-3: Phylogeny and enzymatic activity of octocoral terpene cyclases (TCs).

A: Unrooted phylogenetic tree showing the overall topology of octocoral TCs, with clades 1-4 and subclades I-V within clade 1. **B:** Rectangular view of the phylogenetic tree. Labeled tips correspond to functionally characterized enzymes, annotated with their respective terpene products. Stars indicate enzymes from deep-sea octocorals newly characterized in this study. Filled circles mark ancestral nodes that have been biochemically characterized for each of the subclades 1-I through 1-V, labeled as A1 through A5, respectively. Abbreviations: M, multiple products; n.a., inactive protein; n.e., not expressed protein; ST, unknown sesquiterpene. **C:** Structures of terpene products generated by characterized octocoral TCs.

Lineage-Specific Divergence and Selection in Terpene Cyclase Clades

Next, we aimed to investigate if the observed functional differences in clade 1 and clades 2, 3 and 4 occur due to distinct evolutionary mechanisms. We compared patterns of nucleotide divergence and synonymous site evolution in octocoral TC genes among the distinct clades and the subclades in clade 1 relative to mitochondrial genes. Although mitogenomes in octocorals show evidence of purifying selection and nuclear-mitogenome discordance,^{37,38} they remain widely used as phylogenetic markers (e.g., mutS) and serve here as a conserved reference for comparative analyses. Functionally conserved diterpene synthase subclades within TC clade 1 partially exhibited constrained evolution, with subclades 1-II and 1-IV showing low divergence, while subclade 1-III and the more functionally diverse clades 2 and 4 diverged three- to four-fold more rapidly (Figure S41A). These trends were broadly consistent with population genetic metrics (Table S9), including lower dN/dS ratios in some conserved subclades, but overall, no single metric showed a uniform pattern across clades. Together, these results indicated that evolutionary rates and constraints vary among TC clades but do not follow a simple relationship with enzyme function or phylogenetic grouping.

Taxonomic differences further contributed to variation in TC gene evolution. While mitochondrial genes showed no significant differences in nucleotide diversity (π) or allele frequency distributions (Tajima's D) between the two octocoral orders (π : $p = 0.8$; Tajima's D: $p = 0.4$), TC genes differed significantly in both metrics (π : $p < 0.05$; Tajima's D: $p = 0.06$; Figure S41B), suggesting that terpene biosynthetic genes have been shaped, at least in part, by order-specific demographic histories or selection. Codon usage patterns were consistent with these broader trends. TC genes exhibited greater third-position nucleotide asymmetries than mitochondrial genes, with a significant order-specific bias in A vs. T usage (PR2_AT: $p = 0.03$) but no difference in G vs. C usage (PR2_GC: $p = 0.22$; Figure 4A). Both gene sets showed a strong linear relationship between GC₃ and overall GC content, but TC genes were systematically elevated in both, reflecting gene-specific mutational pressures (Figure S42A).

Codon bias in TC genes was modest overall, with most genes clustering at high effective codon number (N_c) values, consistent with weak codon usage bias (Figure 4B). Nonetheless, TC genes exhibited broader codon repertoires and increased absolute codon frequencies for specific codons relative to mitochondrial

genes (Figure S42C), highlighting compositional differences between gene groups. At finer scales, relative synonymous codon usage (RSCU) analyses identified a subset of codons with significant inter-clade differences ($FDR < 0.05$; RSCU range > 0.7) across TC genes, particularly for alanine, arginine, glycine, histidine, serine, and threonine (Figure 4C). However, these patterns did not consistently align with enzyme function or evolutionary constraint, indicating that synonymous site evolution in TC genes is shaped by a combination of lineage-specific mutational biases, relaxed selection, and demographic history. The distinct codon usage patterns observed among clades are consistent with deep, independent evolutionary trajectories following the early diversification of multiple ancestral TC genes in octocorals.

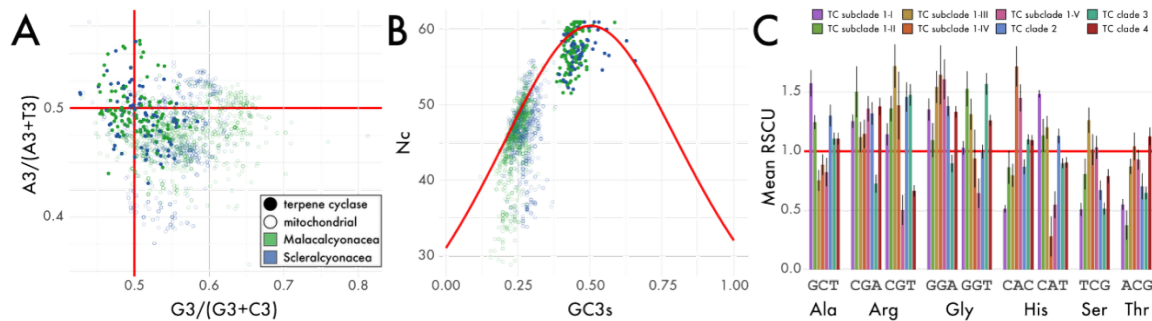


Figure 2-4: Codon usage patterns and synonymous codon biases in octocoral terpene cyclase (TC) genes.

A: PR2 plot showing third-position nucleotide asymmetries in TC and mitochondrial genes. Each point represents an individual gene, colored by taxonomic order (Malacalcyonacea in green, Scleralcyonacea in blue). Red lines at 0.5 indicate expected equilibrium values under unbiased synonymous site evolution. TC genes exhibit greater variability in A/T and G/C asymmetries compared to mitochondrial genes, consistent with relaxed constraints or lineage-specific mutational biases. **B:** Wright's plot of effective codon number (N_c) versus third-position GC content (GC_3) for TC and mitochondrial genes, colored by taxonomic order as in panel A. TC genes cluster at high N_c values, consistent with weak codon usage bias. The red curve shows the expected relationship between N_c and GC_3 under neutral evolution. Mitochondrial genes are shown as a conserved comparative reference. **C:** TC clade variation in relative synonymous codon usage (RSCU) for a subset of codons showing significant differences among TC clades ($FDR < 0.05$; RSCU range > 0.7). Codons are grouped by encoded amino acid, with bars colored by TC clade. These patterns highlight clade-specific synonymous site evolution in TC genes, suggesting lineage-specific mutational biases or weak selection on codon preference.

2.6 Discussion

The recent discovery of terpenoid biosynthetic pathways in octocorals prompted us to investigate their distribution and evolution across taxa and habitats. Here, we show that deep-sea octocorals, spanning six genera and both octocoral orders, produce species-specific sesquiterpenes and a restricted set of diterpenes. The latter corresponded to the precursors of XBECK-type diterpenoids, which encompass five structurally

diverse groups originally reported from shallow-water species. Our findings highlight deep-sea corals as an untapped source of terpenoids, including likely novel XBECK-type diterpenoids, many of which exhibit promising bioactivities (Figure 2B). Furthermore, the taxonomic distribution of these diterpenoid types reveals that those derived from **6**, **7**, **8**, and **10** are found in both octocoral orders, consistent with an ancient evolutionary origin for their biosynthetic pathways.

This evolutionary pattern is mirrored in the phylogeny of octocoral terpene cyclases (TCs). Clade 1 comprises five monophyletic, isofunctional subclades of diterpene synthases that produce the precursors of the five major diterpenoid types, which was established by characterizing extant representatives as well as the reconstructed ancestors for all five subclades (Figure 3B; Figure S11). These subclades show broad taxonomic representation, including sequences from both octocoral orders, consistent with ancient, conserved diterpene biosynthesis. In contrast, clades 2, 3, and 4 contain primarily sesquiterpene cyclases that exhibit rapid functional diversification and more limited taxonomic distribution (Figure S11), which corresponds to the species-specific sesquiterpene profiles observed via GCMS analysis. A similar evolutionary pattern was observed in plant terpene cyclases, where the ancestral function to produce *ent*-kaurene is preserved in all plants, while a vast array of additional TC copies producing a plethora of different products has evolved in different plant species³⁹. The long-term preservation of XBECK-type diterpene synthase activity, coupled with diversification of downstream tailoring enzymes, likely underpins the vast structural diversity of coral diterpenoids belonging to these five types. Indeed, functional diversification at early steps in these pathways^{29,40} would disrupt precursor production and associated biological functions, whereas diversification of late-stage redox enzymes can yield new derivative structures with novel or enhanced activity. While oxidized di- and sesquiterpenoids from octocorals have been linked to defensive roles,^{41,42} the ecological roles of terpene hydrocarbons, specifically sesquiterpenes that might not represent pathway intermediates, remain unclear.⁴³ Instead, they may serve functions like inter- or intraspecific communication, in which species specific blends of compounds could be advantageous, analogous to volatiles in plants.⁴⁴

Comparative evolutionary analyses revealed distinct patterns across TC clades that partially align with their functional roles. The elisabethatriene (**8**)-, klysimplexin R (**10**)-, and xeniaphyllene (**6**)-producing diterpene synthase subclades (1-I, 1-II, 1-IV) exhibited slower divergence and elevated nucleotide diversity, consistent

with functional constraint. In contrast, the cembrene A (**9**)- and cembrene B (**7**)-producing subclades (1-III, 1-V) showed divergence rates comparable to the more recently diversified, functionally heterogeneous sesquiterpene cyclases of clades 2-4. These patterns are further supported by synonymous site evolution and codon usage, which varied among TC clades in a manner consistent with their distinct evolutionary histories. A clear trend that differentiates the functionally conserved groups in clade 1 from clades 2,3 and 4 is lacking. However, the clade-specific variation likely reflects a combination of lineage-specific mutational biases and relaxed selection, superimposed on deep, independent diversification of each TC lineage. In addition, significant differences in nucleotide diversity, synonymous site composition, and codon usage between octocoral orders across all TC clades indicate that demographic history and selection have further shaped TC gene evolution beyond phylogenetic divergence alone.

The presence of multiple, functionally distinct diterpene cyclase clades with broad taxonomic representation, including three subclades comprising sequences from both octocoral orders, suggests that the last common octocoral ancestor possessed at least four terpene cyclase genes: three producing **6**, **8**, and **10**, and a fourth ancestral gene that diversified into clades 2, 3 and 4 (Figure 5A). Whether clades 1-III and 1-V and their observed **7**- and **9**-synthase activity originated in the last common ancestor and was subsequently lost in one order, or evolved independently post-divergence, remains unclear. We previously postulated that a TC gene was acquired in octocorals via horizontal gene transfer from a microbial organism ¹⁴, a hypothesis that was supported by the monophyly of octocoral TCs and the absence of terpene biosynthesis in hexacorals and all other cnidarians (Figure 5B, top). However, the existence of multiple monophyletic, functionally distinct TC clades sharing moderate sequence identity raises the possibility that terpene biosynthetic capability originated earlier. One plausible scenario is that the last common anthozoan, or even cnidarian ancestor, possessed terpene biosynthesis, which was lost independently in medusozoans, hexacorals, and other lineages (Figure 5B, bottom). Similar patterns of widespread ancestral biosynthetic capability followed by multiple independent losses have been observed for enzymes connected to sterol biosynthesis in metazoans.⁴⁵ The deep evolutionary origins of animal terpene biosynthesis, potentially predating modern cnidarian diversification, remain an intriguing possibility. The recent discovery of phylogenetically distinct sponge terpene cyclases⁴⁶

further supports the hypothesis that terpene biosynthesis was widespread in early animals and that further biosynthetic diversity remains to be discovered across the animal kingdom.

Together, these findings represent the first comprehensive evolutionary analysis of octocoral terpene biosynthesis, revealing both ancient, conserved pathways and ongoing gene diversification. As this is a first survey on this matter, future research will expand on linked genomic and chemical datasets, which will further refine our understanding of coral evolution and facilitate the exploration and biotechnological potential of their abundant biosynthetic capacity.

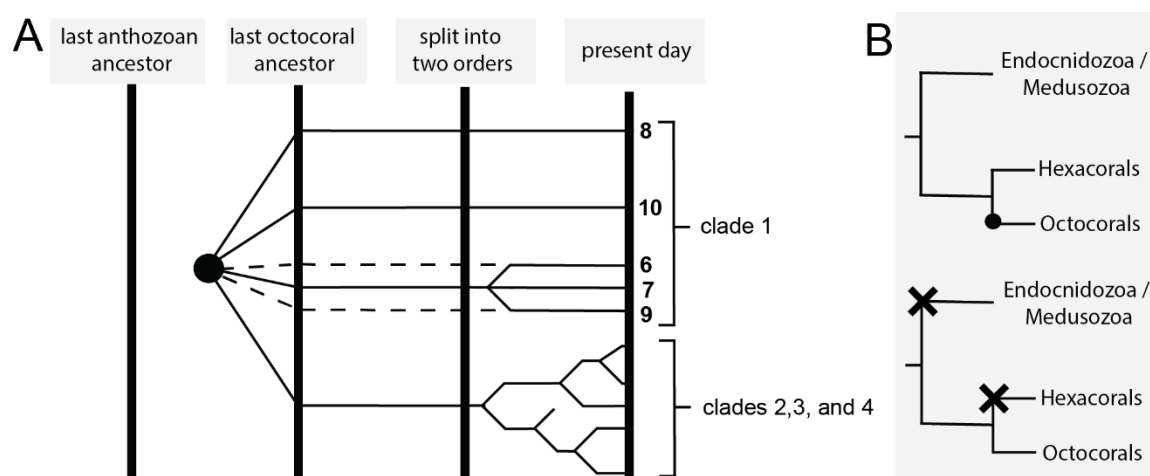


Figure 2-5: Schematic of possible early TC evolution in octocorals.

A: Diversification of different terpene cyclase lineages throughout four timed events, marked by vertical bars. The latest plausible acquisition of the ancestor gene to all octocoral terpene cyclases is marked by a black circle. Diterpene lineages are marked with their generated products. The diversification of clades 2, 3, and 4 is shown schematically to illustrate the prolific duplication-neofunctionalizations observed in the phylogenetic data. **B:** Two competing hypotheses for the emergence of octocoral terpene cyclases. Top: Acquisition of the ancestral terpene cyclase by the last common octocoral ancestor. The acquisition event is marked by a black circle. Bottom: The last common cnidarian ancestor had the capability to produce terpenes, but it was lost in the emerging lineages of the endocnidozoans and medusozoans and later by the hexacorals, after their split from octocorals. Loss events are marked as X.

2.8 Main Text References

- (1) McFadden, C. S.; Van Ofwegen, L. P.; Quattrini, A. M. Revisionary Systematics of Octocorallia (Cnidaria: Anthozoa) Guided by Phylogenomics. *Bull. Soc. Syst. Biol.* **2022**, *1* (3). <https://doi.org/10.18061/bssb.v1i3.8735>.
- (2) Quintanilla, E.; Rodrigues, C. F.; Henriques, I.; Hilário, A. Microbial Associations of Abyssal Gorgonians and Anemones (>4,000 m Depth) at the Clarion-Clipperton Fracture Zone. *Front. Microbiol.* **2022**, *13*. <https://doi.org/10.3389/fmicb.2022.828469>.

- (3) Cairns, S. D. Deep-Water Corals: An Overview with Special Reference to Diversity and Distribution of Deep-Water Scleractinian Corals. *Bull. Mar. Sci.* **2007**, 81 (3), 311–322.
- (4) McFadden, C. S.; Quattrini, A. M.; Brugler, M. R.; Cowman, P. F.; Dueñas, L. F.; Kitahara, M. V.; Paz-García, D. A.; Reimer, J. D.; Rodríguez, E. Phylogenomics, Origin, and Diversification of Anthozoans (Phylum Cnidaria). *Syst. Biol.* **2021**, 70 (4), 635–647. <https://doi.org/10.1093/sysbio/syaa103>.
- (5) Kornprobst, J.-M. Cnidaria and Ctenophora–1. In *Encyclopedia of Marine Natural Products*; John Wiley & Sons, Ltd, 2014; pp 1087–1208. <https://doi.org/10.1002/9783527335855.marprod020>.
- (6) Avila, C. Chemical War in Marine Animal Forests: Natural Products and Chemical Interactions. In *Perspectives on the Marine Animal Forests of the World*; Rossi, S., Bramanti, L., Eds.; Springer International Publishing: Cham, 2020; pp 239–307. https://doi.org/10.1007/978-3-030-57054-5_9.
- (7) Rodríguez, A. D. The Natural Products Chemistry of West Indian Gorgonian Octocorals. *Tetrahedron* **1995**, 51 (16), 4571–4618. [https://doi.org/10.1016/0040-4020\(95\)00216-U](https://doi.org/10.1016/0040-4020(95)00216-U).
- (8) Berrue, F.; Kerr, R. G. Diterpenes from Gorgonian Corals. *Nat. Prod. Rep.* **2009**, 26 (5), 681–710. <https://doi.org/10.1039/B821918B>.
- (9) Ishigami, S.-T.; Goto, Y.; Inoue, N.; Kawazu, S.-I.; Matsumoto, Y.; Imahara, Y.; Tarumi, M.; Nakai, H.; Fusetani, N.; Nakao, Y. Cristaxenicin A, an Antiprotozoal Xenicane Diterpenoid from the Deep Sea Gorgonian *Acanthoprimnoa Cristata*. *J. Org. Chem.* **2012**, 77 (23), 10962–10966. <https://doi.org/10.1021/jo302109g>.
- (10) Keifer, P. A.; Rinehart, K. L.; Hooper, I. R. Renillafulins, Antifouling Diterpenes from the Sea Pansy *Renilla reniformis* (Octocorallia). *J. Org. Chem.* **1986**, 51 (23), 4450–4454. <https://doi.org/10.1021/jo00373a020>.
- (11) Look, S. A.; Fenical, W.; Jacobs, R. S.; Clardy, J. The Pseudopterogens: Anti-Inflammatory and Analgesic Natural Products from the Sea Whip Pseudopterogorgia *Elisabethae*. *Proc. Natl. Acad. Sci USA* **1986**, 83 (17), 6238–6240. <https://doi.org/10.1073/pnas.83.17.6238>.
- (12) Fenical, W.; Okuda, R. K.; Bandurraga, M. M.; Culver, P.; Jacobs, R. S. Lophotoxin: A Novel Neuromuscular Toxin from Pacific Sea Whips of the Genus *Lophogorgia*. *Science* **1981**, 212 (4502), 1512–1514. <https://doi.org/10.1126/science.6112796>.
- (13) Lindel, T.; Jensen, P. R.; Fenical, W.; Long, B. H.; Casazza, A. M.; Carboni, J.; Fairchild, C. R. Eleutherobin, a New Cytotoxin That Mimics Paclitaxel (Taxol) by Stabilizing Microtubules. *J. Am. Chem. Soc.* **1997**, 119 (37), 8744–8745. <https://doi.org/10.1021/ja9717828>.
- (14) Burkhardt, I.; de Rond, T.; Chen, P. Y.-T.; Moore, B. S. Ancient Plant-like Terpene Biosynthesis in Corals. *Nat. Chem. Biol.* **2022**, 18 (6), 664–669. <https://doi.org/10.1038/s41589-022-01026-2>.
- (15) Skropeta, D.; Wei, L. Recent Advances in Deep-Sea Natural Products. *Nat. Prod. Rep.* **2014**, 31 (8), 999–1025. <https://doi.org/10.1039/C3NP70118B>.
- (16) Lebar, M. D.; Heimbegner, J. L.; Baker, B. J. Cold-Water Marine Natural Products. *Nat. Prod. Rep.* **2007**, 24 (4), 774–797. <https://doi.org/10.1039/B516240H>.
- (17) Soldatou, S.; Baker, B. J. Cold-Water Marine Natural Products, 2006 to 2016. *Nat. Prod. Rep.* **2017**, 34 (6), 585–626. <https://doi.org/10.1039/C6NP00127K>.
- (18) Baena, P.; Martell, L.; Soto-Angel, J. J.; Ambroso, S.; López-González, P. J. A New Deep-Sea Species of Golden Gorgonian (Octocorallia: Scleractyonia: Chrysogorgiidae) from Antarctic Waters. *Deep-Sea Res. I: Oceanogr. Res. Pap.* **2024**, 204, 104234. <https://doi.org/10.1016/j.dsr.2024.104234>.

- (19) Sherwood, O. A.; Edinger, E. N. Ages and Growth Rates of Some Deep-Sea Gorgonian and Antipatharian Corals of Newfoundland and Labrador. *Can. J. Fish. Aquat. Sci.* **2009**, *66* (1), 142–152. <https://doi.org/10.1139/F08-195>.
- (20) Prouty, N. G.; Fisher, C. R.; Demopoulos, A. W. J.; Druffel, E. R. M. Growth Rates and Ages of Deep-Sea Corals Impacted by the Deepwater Horizon Oil Spill. *Deep-Sea Res. II: Top. Stud. Oceanogr.* **2016**, *129*, 196–212. <https://doi.org/10.1016/j.dsr2.2014.10.021>.
- (21) Scesa, P. D.; Lin, Z.; Schmidt, E. W. Ancient Defensive Terpene Biosynthetic Gene Clusters in the Soft Corals. *Nat. Chem. Biol.* **2022**, *18* (6), 659–663. <https://doi.org/10.1038/s41589-022-01027-1>.
- (22) Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **2017**, *117* (17), 11570–11648. <https://doi.org/10.1021/acs.chemrev.7b00287>.
- (23) McFadden, C. S.; Quattrini, A. M.; Brugler, M. R.; Cowman, P. F.; Dueñas, L. F.; Kitahara, M. V.; Paz-García, D. A.; Reimer, J. D.; Rodríguez, E. Phylogenomics, Origin, and Diversification of Anthozoans (Phylum Cnidaria). *Systematic Biology* **2021**, *70* (4), 635–647. <https://doi.org/10.1093/sysbio/syaa103>.
- (24) DeLeo, D. M.; Bessho-Uehara, M.; Haddock, S. H. D.; McFadden, C. S.; Quattrini, A. M. Evolution of Bioluminescence in Anthozoa with Emphasis on Octocorallia. *Proc. R. Soc. Lond B. Biol. Sci.* **2024**, *291* (2021), 20232626. <https://doi.org/10.1098/rspb.2023.2626>.
- (25) Li, G.; Dickschat, J. S.; Guo, Y.-W. Diving into the World of Marine 2,11-Cyclized Cembranoids: A Summary of New Compounds and Their Biological Activities. *Nat. Prod. Rep.* **2020**, *37* (10), 1367–1383. <https://doi.org/10.1039/D0NP00016G>.
- (26) Li, Y.; Pattenden, G. Perspectives on the Structural and Biosynthetic Interrelationships between Oxygenated Furanocembranoids and Their Polycyclic Congeners Found in Corals. *Nat. Prod. Rep.* **2011**, *28* (7), 1269–1310. <https://doi.org/10.1039/C1NP00023C>.
- (27) Huber, T.; Weisheit, L.; Magauer, T. Synthesis of Xenia Diterpenoids and Related Metabolites Isolated from Marine Organisms. *Beilstein J. Org. Chem.* **2015**, *11* (1), 2521–2539. <https://doi.org/10.3762/bjoc.11.273>.
- (28) Kohl, A. C.; Ata, A.; Kerr, R. G. Pseudopterosin Biosynthesis—Pathway Elucidation, Enzymology, and a Proposed Production Method for Anti-Inflammatory Metabolites from Pseudopterogorgia Elisabethae. *J. Ind. Microbiol. Biotechnol.* **2003**, *30* (8), 495–499. <https://doi.org/10.1007/s10295-003-0076-7>.
- (29) Grayson, N. E.; Scesa, P. D.; Moore, M. L.; Ledoux, J.-B.; Gomez-Garrido, J.; Alioto, T.; Michael, T. P.; Burkhardt, I.; Schmidt, E. W.; Moore, B. S. A Widespread Metabolic Gene Cluster Family in Metazoans. *Nat. Chem. Biol.* **2025**, 1–10. <https://doi.org/10.1038/s41589-025-01927-y>.
- (30) Medema, M. H.; de Rond, T.; Moore, B. S. Mining Genomes to Illuminate the Specialized Chemistry of Life. *Nat. Rev. Genet.* **2021**, *22* (9), 553–571. <https://doi.org/10.1038/s41576-021-00363-7>.
- (31) Kashman, Y.; Loya, Y.; Bodner, M.; Groweiss, A.; Benayahu, Y.; Naveh, N. Gas-Liquid Chromatograms of Sesquiterpenes as Finger Prints for Soft-Coral Identification. *Mar. Biol.* **1980**, *55* (4), 255–259. <https://doi.org/10.1007/BF00393777>.
- (32) Palframan, M. J.; Pattenden, G. Biosynthetic Interrelationships within Polycyclic Cembranoids Isolated from Corals: Conjecture, Biomimetic Synthesis and Reality. *Eur. J. Org. Chem.* **2020**, *2020* (16), 2330–2349. <https://doi.org/10.1002/ejoc.201901438>.

- (33) Lyu, C.; Chen, T.; Qiang, B.; Liu, N.; Wang, H.; Zhang, L.; Liu, Z. CMNPD: A Comprehensive Marine Natural Products Database towards Facilitating Drug Discovery from the Ocean. *Nucleic Acids Res.* **2021**, *49* (D1), D509–D515. <https://doi.org/10.1093/nar/gkaa763>.
- (34) Scesa, P. D.; Schmidt, E. W. Chapter Fourteen - Brewing Coral Terpenes—A Yeast Based Approach to Soft Coral Terpene Cyclases. In *Methods in Enzymology*; Rudolf, J., Ed.; Terpene Synthases; Academic Press, 2024; Vol. 699, pp 373–394. <https://doi.org/10.1016/bs.mie.2024.03.023>.
- (35) Chen, B.; Mao, J.; Xu, K.; Liu, L.; Lin, W.; Guo, Y.-W.; Wu, R.; Wang, C.; Xu, B. Mining Coral-Derived Terpene Synthases and Mechanistic Studies of the Coral Biflorane Synthase. *Sci. Adv.* **2025**, *11* (9), eadv0805. <https://doi.org/10.1126/sciadv.adv0805>.
- (36) Rinkel, J.; Lauterbach, L.; Rabe, P.; Dickschat, J. S. Two Diterpene Synthases for Spiroalbatene and Cembrene A from *Allokutzneria Albata*. *Angew. Chem. Int. Ed.* **2018**, *57* (12), 3238–3241. <https://doi.org/10.1002/anie.201800385>.
- (37) Ramos, N. I.; DeLeo, D. M.; Horowitz, J.; McFadden, C. S.; Quattrini, A. M. Selection in Coral Mitogenomes, with Insights into Adaptations in the Deep Sea. *Sci. Rep.* **2023**, *13* (1), 6016. <https://doi.org/10.1038/s41598-023-31243-1>.
- (38) Quattrini, A. M.; Snyder, K. E.; Purow-Ruderman, R.; Seiblit, I. G. L.; Hoang, J.; Floerke, N.; Ramos, N. I.; Wirshing, H. H.; Rodriguez, E.; McFadden, C. S. Mito-Nuclear Discordance within Anthozoa, with Notes on Unique Properties of Their Mitochondrial Genomes. *Sci. Rep.* **2023**, *13* (1), 7443. <https://doi.org/10.1038/s41598-023-34059-1>.
- (39) Jia, Q.; Brown, R.; Köllner, T. G.; Fu, J.; Chen, X.; Wong, G. K.-S.; Gershenzon, J.; Peters, R. J.; Chen, F. Origin and Early Evolution of the Plant Terpene Synthase Family. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (15), e2100361119. <https://doi.org/10.1073/pnas.2100361119>.
- (40) Scesa, P. D.; Schmidt, E. W. Pseudopterosin Biosynthesis: Unravelling a Decades Old Problem in Animal Specialized Metabolism. *J. Am. Chem. Soc.* **2025**, *147* (4), 3072–3079. <https://doi.org/10.1021/jacs.4c09925>.
- (41) Barsby, T.; Kubanek, J. Isolation and Structure Elucidation of Feeding Deterrent Diterpenoids from the Sea Pansy, *Renilla Reniformis*. *J. Nat. Prod.* **2005**, *68* (4), 511–516. <https://doi.org/10.1021/np049609u>.
- (42) Giordano, G.; Carbone, M.; Ciavatta, M. L.; Silvano, E.; Gavagnin, M.; Garson, M. J.; Cheney, K. L.; Mudianta, I. W.; Russo, G. F.; Villani, G.; Magliozzi, L.; Polese, G.; Zidorn, C.; Cutignano, A.; Fontana, A.; Ghiselin, M. T.; Mollo, E. Volatile Secondary Metabolites as Aposematic Olfactory Signals and Defensive Weapons in Aquatic Environments. *Proc. Natl. Acad. Sci. USA* **2017**, *114* (13), 3451–3456. <https://doi.org/10.1073/pnas.1614655114>.
- (43) Fenical, W.; Pawlik, J. R. Defensive Properties of Secondary Metabolites from the Caribbean Gorgonian Coral *Erythropodium Caribaeorum*. *Mar. Ecol. Prog. Ser.* **1991**, *75* (1), 1–8.
- (44) Schuman, M. C. Where, When, and Why Do Plant Volatiles Mediate Ecological Signaling? The Answer Is Blowing in the Wind. *Annu. Rev. Plant Biol.* **2023**, *74* (Volume 74, 2023), 609–633. <https://doi.org/10.1146/annurev-arplant-040121-114908>.
- (45) Brunoir, T.; Mulligan, C.; Sistiaga, A.; Vu, K. M.; Shih, P. M.; O'Reilly, S. S.; Summons, R. E.; Gold, D. A. Common Origin of Sterol Biosynthesis Points to a Feeding Strategy Shift in Neoproterozoic Animals. *Nat. Commun.* **2023**, *14* (1), 7941. <https://doi.org/10.1038/s41467-023-43545-z>.

- (46) Wilson, K.; de Rond, T.; Burkhardt, I.; Steele, T. S.; Schäfer, R. J. B.; Podell, S.; Allen, E. E.; Moore, B. S. Terpene Biosynthesis in Marine Sponge Animals. *Proc. Natl. Acad. Sci USA* **2023**, *120* (9), e2220934120. <https://doi.org/10.1073/pnas.2220934120>.
- (47) Muthye, V.; Mackereth, C. D.; Stewart, J. B.; Lavrov, D. V. Large Dataset of Octocoral Mitochondrial Genomes Provides New Insights into Mt-mutS Evolution and Function. *DNA Repair* **2022**, *110*, 103273. <https://doi.org/10.1016/j.dnarep.2022.103273>.

2.9 Supplementary Information

General Methods

All chemicals and solvents were used as received from the commercial supplier (Sigma-Aldrich, Fisher or TCI). The substrates GPP, FPP, and GGPP were synthesized as reported previously ^{1,2}. Authentic terpene standard for **20** and **18** were purchased from Sigma Aldrich and Cayman chemicals, respectively, the residual standards were obtained from our in-house library. Thin Layer chromatography was performed with TLC plates purchased from Merck (silica gel 60 F254, glass plates). Column chromatography was carried out using a Flash EZ prep combiflash system (Teledyne ISCO). If not specified otherwise, GCMS analysis was carried out on an Agilent 7890A gas chromatograph with Agilent 5975C mass spectrometer, using a Rtx-5Sil 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 3min. Flow rate: 1 mL/min, Injection volume: 1 μL, splitless injection, Inlet temperature: 270 °C, MS transfer line temperature: 280 °C, MS source temperature 230 °C, MS quad temperature: 150 °C. NMR spectra were recorded on a JEOL ECZ spectrometer (500 MHz) in C₆D₆ and were referenced against C₆D₆ (δ = 7.16 ppm) for ¹H-NMR and C₆D₆ (δ = 128.06 ppm) for ¹³C-NMR. Data for NMR spectra are reported as follows: shift (δ) in ppm, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet or unresolved, br = broad signal, J = coupling constant(s) in Hz. Optical rotation measurements were carried out using a Jasco P-2000 polarimeter.

Sample collection

Samples for this work were collected on two deep sea exploration expeditions, covering an area of approximately 39 000 km² of the “Southern California Borderland” (SCB), an area off the southwest coast of the US. The SCB extends from Point Conception to the border with Mexico and out to 300 km offshore. Sampling sites were selected with a focus on the presence of potentially commercially interesting mineral-rich hardgrounds, which are prevalent at slopes and ridges, as well as on off-shore seamounts. The first expedition

on board the E/V Nautilus in November 2020 collected biological samples and oceanographic data at eight dive sites using the remotely operated underwater vehicle (ROV) Hercules, supported by the deep-tow system Argus. The second expedition on board Schmidt Ocean Institute's RV Falkor in August 2021 used ROV SuBastian to collect biological samples and oceanographic data at complementing eight dive sites. Upon retrieval from the ROV, animal tissues were subsampled for DNA extraction (immediately frozen at -80 °C) and metabolomics (immediately frozen at -20°C). For exact coordinates of samples investigated in this work see Table S1.

Taxonomic assignment of coral samples

Nine octocorals were identified using genomic data and morphological characters. To help identify octocorals, skeletal axis types and sclerite shapes were examined³. mtMutS sequences, which are a mitochondrial informative DNA barcode,⁴ were then extracted from the assemblies. The data were aligned with muscle in AliView v 1.26⁵ against NCBI GenBank data.

Terpene extraction and analysis

Between 300 and 1000 mg of frozen coral tissue was separated from the samples, thawed and then macerated in 2 mL MeOH/DCM 2:1. The mixture was stirred for 12 hours in a closed 20 mL scintillation vial at room temperature. The resulting suspension was diluted with water (10 mL), 1 mL of hexane was added, and the sample mixed by vortexing for 10 seconds. After layer separation, the top layer was carefully pipetted off into GCMS vials and directly subjected to GCMS analysis. Chromatograms are shown in Figures S2-S10.

DNA extraction and sequencing

DNA was extracted using a bead beating and phenol:chloroform:isoamyl alcohol protocol as previously described⁶. Subsampling and preparations were performed under sterile conditions inside a flow cabinet. Between 0.1 g and 0.5 g of tissue was thawed, cut into small pieces and used for DNA extraction. DNA was eluted into 20-80 µL of EB buffer. All DNA samples were additionally cleaned using a DNA clean & concentrator-100 kit (Zymo) and quantified using Nanodrop. DNA was submitted to the UCSD Institute for Genomic Medicine for library prep using the KAPA HyperPrep Kit (Roche) and metagenomic sequencing using 1 lane of Illumina NovaSeq S4 PE150 technology, targeting a read depth of 75M read pairs per sample.

Genome assembly and phylogenomics

We assembled high-quality draft genomes for nine deep-sea octocoral specimens using the BBMap toolkit v38.96 ⁷ for read trimming and quality filtering. Adapter sequences and low-quality regions were removed, and paired-end reads were merged prior to assembly with IDBA-UD v1.1.3 ⁸ using a k-mer size range of 30–200 with error correction enabled. Only contigs >2 kb in length were retained for further analyses. To exclude potential bacterial or archaeal symbiont sequences, we generated taxon-annotated GC-coverage (TAGC) plots, based on contig coverage (Bowtie2 v2.4.5) ⁹, GC content, and blastn v2.12.0+ searches against the NCBI non-redundant database (downloaded September 2022).

In addition to the deep-sea genomes, we included publicly available genomic and transcriptomic datasets from 34 representative octocoral species (Supplementary File 1). For datasets lacking pre-assembled genomes (i.e., *Capnella imbricata*; SRA accession SRR15733495), raw reads were retrieved from the SRA database and processed identically to the deep-sea samples.

All assemblies (N=43) were screened for the presence of 14 conserved mitochondrial marker genes (atp6, atp8, cob, cox1, cox2, cox3, mutS, nad1, nad2, nad3, nad4, nad4L, nad5, nad6) using diamond blastx (v2.1.6) ¹⁰ against a curated reference database of 184 octocoral mitochondrial genomes ¹¹. To account for the incomplete nature of draft assemblies, we focused on a subset of seven mitochondrial markers (atp6, atp8, cox2, mutS, nad4, nad4L, nad6) that were consistently recovered across samples and exhibited more stable phylogenetic patterns in preliminary analyses. Samples lacking these seven markers or showing evidence of multiple divergent copies were excluded, resulting in a final dataset of 213 taxa for mitochondrial phylogenetic analyses.

Individual mitochondrial genes were aligned using Clustal Omega v1.2.4 ¹² and concatenated into a multi-gene alignment. Phylogenetic relationships were reconstructed using RAxML-NG v1.2.0 ¹³ under a general time reversible (GTR) model with gamma-distributed rate heterogeneity, starting from 20 randomized parsimony trees and assessed with 100 non-parametric bootstrap replicates.

Terpene Cyclase Phylogenetic Analyses

The nine deep-sea draft genome assemblies, along with the publicly available octocoral sequencing datasets, were queried for terpene cyclases using hidden Markov models (HMMs), as previously described ¹⁴. Briefly, open reading frames (ORFs) were translated using getorf (EMBOSS v6.6.0.0) and screened with hmmsearch (HMMER v3.3.2) ¹⁵ against custom TC HMM profiles. Putative hits were identified by manual checking for the conserved motifs DDXXD, NSE, and RY ¹⁴, and their corresponding nucleotide sequences were also extracted. In total, we identified 40 TC sequences from the deep-sea genomes and 137 TC sequences from public datasets, including previously characterized enzymes (Supplementary File 1). Both nucleotide and amino acid sequences were aligned separately. Nucleotide sequences were processed and phylogenies reconstructed as described above and used for all downstream evolutionary and population-level analyses. For visualization purposes, amino acid sequences of the TCs were aligned using Kalign v2.044, and a maximum-likelihood tree was generated in IQ-TREE v2.3.35 ¹⁶ using the Q.plant+I+R5 model selected by ModelFinder ¹⁷. Trees were visualized using iTOL v6 ¹⁸.

Evolutionary analyses: divergence, population genetics, and codon usage

Divergence. To assess evolutionary congruence between terpene cyclase and mitochondrial gene divergence, we compared pairwise distances derived from TC phylogenies to corresponding mitochondrial phylogenetic distances for each genome where both data types were available. Analyses were performed separately for each major TC clade, with genome pairs excluded if mitochondrial data were unavailable or incomplete. For each clade-specific dataset, linear regressions were performed to assess the relationship between TC and mitochondrial divergence, and slope and adjusted R^2 values were reported.

Population Metrics. Population genetic statistics were calculated for TC genes and mitochondrial genes to assess evolutionary constraint and demographic patterns. Prior to alignment, TC nucleotide sequences were screened for complete open reading frames using a custom Python script to exclude sequences with frameshifts or premature stop codons. Genes passing this filter were aligned using Clustal Omega v1.2.4 and used for downstream analyses.

We estimated nucleotide diversity (π), Tajima's D, Watterson's theta (θ_w), and dN/dS ratios using the PopGenome¹⁹, seqinr²⁰, and ape²¹ packages in R. Z-scores and adjusted p-values for dN/dS were calculated

based on variance estimates, and genes with unreliable estimates (e.g., negative or infinite dS) were excluded. Population genetic metrics were first calculated across all samples, then partitioned by octocoral order (Malacalcyonacea and Scleralcyonacea) based on mitochondrial phylogenetic assignment (see above). As a comparison, the same analyses were performed on the seven mitochondrial marker genes (atp6, atp8, cox2, mutS, nad4, nad4L, nad6) from their corresponding genomes. Because these analyses compare divergent octocoral lineages rather than populations within species, the application of dN/dS ratios is appropriate, consistent with prior expectations²².

Codon Bias. Codon usage patterns were assessed for TC and mitochondrial genes to examine potential lineage-specific evolutionary pressures. Basic codon composition statistics, including third-position GC content (GC₃) and absolute codon frequencies, were calculated for each gene to examine correlations between overall GC content, codon preferences, and codon usage bias. Relative synonymous codon usage (RSCU) was also calculated for each gene using the seqinr package. RSCU values above 1.0 indicate preferential codon usage (with values >1.6 suggesting strong bias), while values below 1.0 indicate underrepresentation (with values <0.6 reflecting strong avoidance)²³. To test for differences in RSCU among TC clades, we performed statistical comparisons for each codon, applying false discovery rate (FDR) correction to account for multiple testing. Codons with FDR-adjusted p-values below 0.05 and an RSCU range exceeding 0.7 across clades were considered to exhibit biologically meaningful inter-clade variation.

To further investigate codon usage, calibrated CodonW v1.4.4²⁴ analyses were performed following similar methodologies to investigate codon usage patterns in Demospongiae²⁵. As a calibration step, the mitochondrial gene atp8 was analyzed using the options -c_type 2 -f_type 7 -code 4, and sequences with internal stop codons or frameshifts were excluded based on log file filtering. All TC genes passing alignment quality filtering were subsequently analyzed with CodonW using the same options to generate consistent codon usage indices. Metrics extracted included the effective number of codons (Nc), where lower values indicate stronger codon bias (with Nc $\approx$ 20 representing maximal bias and Nc $\approx$ 61 indicating no bias), and the codon adaptation index (CAI), a measure of codon usage similarity to a reference set, with higher values reflecting greater adaptation to presumed translational preferences.

In addition to standard CodonW metrics, third-position nucleotide counts (A3s, T3s, G3s, C3s) were extracted from CodonW outputs. These were used to calculate PR2 bias metrics (PR2_AT and PR2_GC) in R, summarizing A/T and G/C asymmetries at third codon positions. Deviations from the expected PR2 equilibrium (PR2_AT = 0.5, PR2_GC = 0.5) can indicate strand-specific mutational biases or weak selection on codon composition ²⁶.

Statistical Analyses. Comparisons of nucleotide diversity, Watterson's theta, allele frequency distributions, and codon usage metrics between TC and mitochondrial genes and across taxonomic orders were performed using Wilcoxon rank-sum. All statistical analyses and data visualization were performed in the R software environment v4.2.2.

Molecular cloning

The protein encoding sequences of selected terpene cyclases were either cloned out of genomic DNA or ordered as original nucleotide sequences (Twist Biosciences). This was done using standard operations using commercially available kits. In brief, for genes that were cloned out of genomic DNA, (Supplementary File 1), primers specific to the gene of interest were designed (Table S9). PCR reactions were set up according to a standard protocol (Phusion High-Fidelity DNA Polymerase Master Mix, Thermo Scientific) based on the template DNA size and primer design. An example of the PCR conditions are as follows:

Repetitions	Temperature	Time
1x	98 °C	0:30 min
30x	98 °C	0:10 min
	56-70 °C	0:30 min
	72 °C	0:50 min
1x	72 °C	5:00 min

The resulting products were purified via a gel purification kit (ZymoClean Gel DNA Recovery Kit (ZymoResearch)). Next, the purified terpene cyclase encoding genes, or the ordered genes were inserted into the *E. coli* expression vector pET28a(+) by cutting the empty vector with the restriction enzymes NdeI (NEB) and XhoI (NEB) in CutSmart® buffer (10x, NEB) at 37 °C for 4 h, and combining the open vector with the gene of interest using Gibson assembly (NEBuilder HiFi DNA Assembly Master Mix, New England Biolabs) according to

the manufacturer's protocol. Subsequently, the plasmid is transformed into *E. coli* DH5 α using chemically competent cells using the heat shock transformation method. Successful cloning and transformation of the template DNA was checked by sequencing the mini prepped plasmids (Quick-DNA Miniprep Kit, ZymoResearch). These TC-containing plasmids were transformed into *E. coli* BL21 (DE3) strains for heterologous expression of proteins.

Heterologous expression for analytical scale

Strains of *E. coli* BL21 (DE3) containing terpene cyclase encoding genes were grown from cryostocks in 10 mL of LB medium (Fisher) with 50 mg/mL kanamycin for 18 hours at 37 °C and 200 rpm in a shaking incubator. One mL of the starting culture was used to inoculate 100 mL of TB (Fisher) supplemented with 50 mg/mL kanamycin in a 250 mL Erlenmeyer flask. The culture was incubated at 37 °C and 200 RPM until the culture reached an OD of 0.6-0.8. Cultures were subsequently cooled to 18 °C, induced with 400 μ M IPTG, and then shaken at 200 RPM for 18 hours. Cells were pelleted by centrifugation (8,000 g for 10 minutes at 4 °C) and resuspended in 5 mL of cold TC Lysis/Binding buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 20 mM imidazole, 1 mM MgCl₂ at pH 7.4). Cells were lysed by sonication (2 seconds on, 2 seconds off for 1 minute, 30 seconds rest, repeated 2 times), and then the cell debris was pelleted using centrifugation (11,000g for 30 minutes at 4 °C). The supernatant was filtered with a 0.45 μ m syringe filter and loaded onto a Ni-NTA column pre-equilibrated with TC binding buffer. The resin was washed twice with 2 mL of binding buffer each time, and the protein was eluted with 2.7 mL of TC elution buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 0.5 M imidazole, 1 mM MgCl₂ at pH 7.4) into a 15 mL Falcon tube. Elution fractions were loaded onto a PD-10 column (Sephadex G-25 resin, Cytiva) pre-equilibrated with 25 mL of TC incubation buffer (50 mM Hepes, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol at pH 7.8). The protein was then eluted with 3.5 mL of TC incubation buffer into a 15 mL conical tube. The protein was used in enzyme assays directly or flash frozen in liquid nitrogen and stored at -80 °C for later use.

Heterologous expression for preparative scale

3 L of TB medium were inoculated with 1% vol of an *E. coli* BL21 (DE3) culture containing an expression plasmid with the respective terpene cyclase. Cultures were grown to an O.D. of 0.6-0.8 at 37°C with shaking. Once the O.D. was reached, the incubator was cooled to 18 °C, then induced with 400 μ M IPTG and shaken at

200 RPM for 18 hours. Cells were pelleted using centrifugation (8,000 g for 15 min at 4 °C) and then resuspended in 60 mL of TC binding buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 20 mM imidazole, 1 mM MgCl₂ at pH 7.4). The cells were lysed via sonication (15 seconds on, 45 seconds off, for a total of 7 minutes on-time). Cell debris was removed through centrifugation (11,000 g for 20 mins at 4 °C), and the supernatant was transferred to new tubes with the addition of DNase (Roche Diagnostics, 1 mg per 20 mL of supernatant) and centrifuged again (11,000 g for 20 mins at 4 °C). The supernatant was filtered through a 0.45 µm syringe filter and loaded onto a 5 mL His-Trap FF-NI column (GE). The column was washed with 40 mL of TC binding buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 20 mM imidazole, 1 mM MgCl₂ at pH 7.4) and eluted with TS elution buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 0.5 M imidazole, 1 mM MgCl₂ at pH 7.4), collecting in 2 mL fractions. Fractions were pooled and loaded onto a PD-10 column (Sephadex G-25 resin, Cytiva), pre-equilibrated with 25 mL TC incubation buffer (50 mM Hepes, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol at pH 7.8) in 2.5 mL batches. Each column was eluted with 3.5 mL of TC incubation buffer. The protein was then used in enzyme assays or flash-frozen in liquid nitrogen and stored at -80°C.

Analytical scale assays

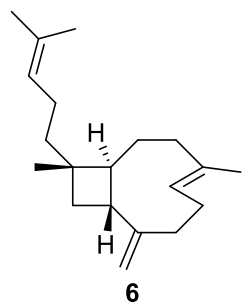
Enzyme assays on an analytical scale were conducted in 1.5 mL of TC incubation buffer (50 mM Hepes, 5 mM MgCl₂, 250 mM NaCl, and 10% glycerol at pH 7.8) with 3 µM enzyme and 0.25 mg mL⁻¹ of substrate (GPP, FPP, or GGPP) to test for mono, sesquiterpene, and diterpene synthase activity. Assays ran for 18 hours at room temperature. 200 µL of hexane were added to the cultures, the mixtures were vortexed, and then centrifuged in a microcentrifuge at maximum speed for 1 minute at room temperature. 150 µL of the extract was carefully pipetted off and analyzed by GC-MS. For assay results see Figures S13-S33.

Preparative scale assays and structure elucidation of TC products

Preparative scale assays were carried out to isolate products for Na124-TC-2 and SCB84-TC-6 and elucidate their structures. FPP (80 mg) was dissolved in ammonium bicarbonate solution (50 mM in 15 mL) and added dropwise to 600 mL of TC incubation buffer (50 mM Hepes, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol at pH 7.8) with 3 µM enzyme concentration over 3 h. The assay was run for 18 h at room temperature, and then extracted with pentane (3 x 200 mL). The organic layer was dried over MgSO₄ and filtered. The resulting filtrate

was then concentrated in vacuo to ~1 mL. The crude fraction was then purified using silica gel chromatography (100% pentane), with individual fractions checked with thin layer chromatography (silica plate, 100% hexane mobile phase) and stained with phosphomolybdic acid. Fractions containing the product were pooled, concentrated to dryness, and analyzed with NMR.

Analytical data for NA124-045-TC-5 product (xeniaphyllene **6**)

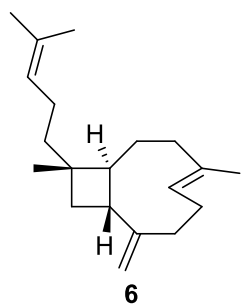


Optical rotation: $[\alpha]_D^{25.0} = -13.2$ (c 0.17, CHCl_3). **$^1\text{H-NMR}$** (500 MHz, C_6D_6): δ = 5.38 (br. s, 1H), 5.28-5.19 (m, 1H), 5.04 (s, 1H), 4.87 (s, 1H), 2.44-2.29 (m, 2H), 2.17-2.10 (m, 1H), 2.09-1.90 (m, 5H,), 1.85-1.77 (m, 1H), 1.77-1.70 (m, 1H), 1.70 (s, 3H), 1.60 (s, 3H), 1.58 (s, 3H), 1.51-1.44 (m, 2H) 1.42-1.33 (m, 2H) 0.98 (s, 3H) ppm. **$^{13}\text{C-NMR}$** (125 MHz, CDCl_3):

154.7, 135.1, 130.8, 125.8, 124.9, 112.4, 52.6, 49.0, 44.6, 40.3, 39.7, 36.3, 35.0, 30.3, 29.0, 25.9, 23.4, 20.1, 17.7, 16.5 ppm.

NMR-data and optical rotation are in agreement with data for **6** produced by XsTC-1 ¹⁴.

Analytical data for SC144-TC-1 product (xeniaphyllene **6**)

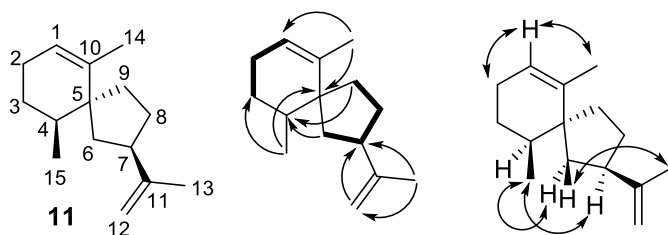


Optical rotation: $[\alpha]_D^{25.0} = -9.3$ (c 0.03, CHCl_3). **$^1\text{H-NMR}$** (500 MHz, C_6D_6): δ = 5.38 (br. s, 1H), 5.28-5.19 (m, 1H), 5.04 (s, 1H), 4.87 (s, 1H), 2.44-2.29 (m, 2H), 2.17-2.10 (m, 1H), 2.09-1.90 (m, 5H,), 1.85-1.77 (m, 1H), 1.77-1.70 (m, 1H), 1.70 (s, 3H), 1.60 (s, 3H), 1.58 (s, 3H), 1.51-1.44 (m, 2H) 1.42-1.33 (m, 2H) 0.98 (s, 3H) ppm. **$^{13}\text{C-NMR}$** (125 MHz, CDCl_3):

154.7, 135.1, 130.8, 125.8, 124.9, 112.4, 52.6, 49.0, 44.6, 40.3, 39.7, 36.3, 35.0, 30.3, 29.0, 25.9, 23.4, 20.1, 17.7, 16.5 ppm.

NMR-data and optical rotation are in agreement with data for **6** produced by XsTC-1 ¹⁴.

Analytical data for NA124-218-TC-2 product (hinesene, **11**):



^aCarbon numberings as shown above table; ^bchemical shifts are given in ppm, coupling constants are given in

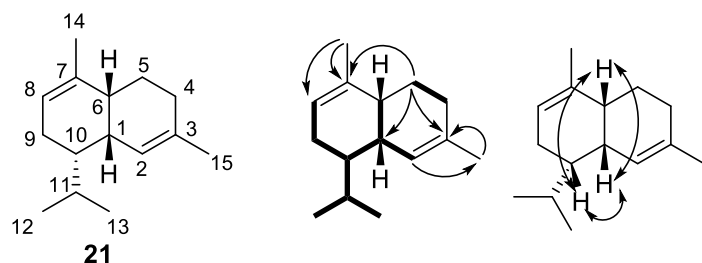
C ^a	¹³ C ^b	¹ H ^b
1	122.1	5.35-5.33 (m, 1H)
2	25.0	1.96-1.91 (m, 2H)
3	28.5	1.54-1.48 (m, 1H) 1.34-1.29 (m, 1H)
4	37.8	1.53-1.50 (m, 1H)
5	49.0	-
6	37.5	1.77-1.72 (m, 1H)
7	48.0	2.43 (qt, J=12.3, 6.5, 1H)
8	32.5	1.81-1.77 (m, 1H) 1.55-1.49 (m, 1H)
9	36.2	1.73-1.70 (m, 1H) 1.59-1.53 (m, 1H)
10	140.1	-
11	148.4	-
12	108.7	4.87 (s, 1H, H _Z) 4.81 (s, 1H, H _E)
13	21.4	1.70 (s, 3H)
14	20.1	1.67 (q, J = 1.9, 3H)
15	16.6	0.88 (d, J = 6.8, 3H)

Hertz, multiplicity: m = multiplet, s = singlet, d = doublet, t = triplet, q = quartet, br = broad.

NMR: See Table. **Optical rotation:** $[\alpha]_{\text{D}}^{28.0} = -11.8$ (c 0.27, CHCl₃).

The absolute configuration was set as shown above by comparing the optical rotation to the literature value of enantioselectively synthesized (–)-**11**²⁷.

Analytical data for SCB84-TC-6 product (a-amorphene **21**):



C ^a	¹³ C ^b	¹ H ^b
1	c	2.69 (br s, 1H)
2	120.0	5.30 (s, 1H)
3	136.0	-
4	26.8	1.95-1.85 (m, 1H) 1.64-1.59 (m, 1H)
5	25.7	1.97-1.91 (m, 1H) 1.65-1.58 (m, 1H)
6	39.0	2.27 (br s, 1H)
7	133.4	-
8	124.5	5.51-5.47 (m, 1H)
9	27.6	2.04-1.96 (m, 1H) 1.73-1.64 (m, 1H)
10	45.4	1.19-1.13 (m, 1H)
11	29.1	1.61-1.54 (m, 1H)
12	21.0	0.95 (d, <i>J</i> = 6.6 Hz, 3H)
13	21.7	0.90 (d, <i>J</i> = 6.6 Hz, 3H)
14	20.9	1.65 (s, 3H)
15	24.3	1.64 (s, 3H)

^aCarbon numberings as shown above; ^bchemical shifts are given in ppm, coupling constants are given in Hertz, multiplicity: m = multiplet, s = singlet, d = doublet, t = triplet, q = quartet, br = broad.

NMR: See Table. **Optical rotation:** $[\alpha]_{\text{D}}^{28.0} = -41.6$ (c 0.045, CH₂Cl₂).

¹H-NMR data and optical rotation are in agreement with reported values from the literature ²⁸.

Determination of the absolute configuration of cembrene A (**9**), produced by different TCs

For determination of the absolute configuration of cembrene A (**9**) from the enzymes DgTC-2, NA124-47-TC-3, NA124-141-TC-4, BaTC-1, and TmTC-1, the extracts of incubation experiments with GGPP and these enzymes were subjected to GC/MS analysis on the same gas chromatograph and mass detector as described in the general methods using a Cyclosil-B column (Agilent, 30 m length, 0.25 mm diameter, 0.25 μ m film). A suitable GC Program for separating (+)- and (–)-**9** (increasing from 120°C at 0.9°C min^{–1} to 180°C and then increasing at 30°C min^{–1} to 250°C, operated in splitless mode) was determined by separating a racemic cembrene A (**9**) standard generated by mixing equal amounts of (+)- and (–)-**9**. Authentic standards for both enantiomers were obtained from in vitro assays using **9**-synthases that had been established as (+)-**9**²⁹ and (–)-**9**¹⁴ synthases previously. The results are shown in Figure S40.

Ancestral sequence reconstruction

Amino acid sequences for ancestral nodes for the subclades 1-I to 1-V were determined using FireProt^{ASR}³⁰. However, using the full phylogeny as presented in the manuscript caused the software to crash. To reduce the sequence count, the terpene cyclase multiple sequence alignment and phylogeny were recreated containing only the sequences from clades 1 and 2 (as outgroup) using kalign2 and IQ-tree with the exact same parameters as described above. The topology of all subclades in clade 1 remained exactly the same with respect to the full phylogeny. The multiple sequence alignment and the phylogeny were used as input files for FireProt-ASR using the ancestral sequence reconstruction mode with default settings. The amino acid sequences from the ancestral nodes for subclades 1-I, 1-II, 1-III, 1-IV and 1-V were exported and ordered as synthetic genes codon-optimized for *E. coli* expression from Twist. The genes were heterologously expressed, purified and characterized using the same methods as described above.

Supplementary Tables

Table 2-S 1: Biological samples used in this study.

Sample number	Taxonomic ID	Depth [m]	Latitude	Longitude	Cruise	Terpenoids reported from genus
NA124-045	<i>Acanthogorgia</i> sp.	884.2	33.036	-121.000	Nautilus 2020	yes (31, 32)
NA124-047	<i>Callistephanus simplex</i>	692.1	33.033	-120.998	Nautilus 2020	no
NA124-141	<i>Callistephanus simplex</i>	562.7	32.738	-120.023	Nautilus 2020	no
NA124-195	<i>Paragorgia arborea</i> cf.	1121.8	32.679	-118.122	Nautilus 2020	yes (33, 34)
NA124-218	<i>Balticina californica</i> cf.	109.0	33.353	-119.045	Nautilus 2020	no
SCB-084	<i>Victogorgia</i> sp.	1269.0	33.146	-120.941	Falkor 2021	no
SCB-144	<i>Chrysogorgia</i> sp.	2477.3	31.901	-120.036	Falkor 2021	no
SCB-157	<i>Chrysogorgia</i> sp.	2415.6	31.899	-120.039	Falkor 2021	no
SCB-216	<i>Paragorgia jamesi</i>	459.8	33.101	-117.881	Falkor 2021	yes (33, 34)

Table 2-S 2: Literature reports of XBECK-type diterpenoids in various octocorals.

Name	Diterpene type
<i>Clavularia inflata</i>	X-type (35)
<i>Erythropodium caribaeorum</i>	K-type, B-type (36, 37)
<i>Briareum asbestinum</i>	K-type, B-type (38, 39)
<i>Heliopora coerulea</i>	E-type, X-type (38, 39)
<i>Junceella fragilis</i>	B-type (40)
<i>Dichotella gemmacea</i>	B-type (40)
<i>Veretillum cf cynomorium</i>	B-type, C-type (41, 42)
<i>Stylatula elongata</i>	B-type (43)
<i>Pennatula aculeata</i>	B-type (44)
<i>Sarcophyton trocheliophorum</i>	C-type (45)
<i>Sarcophyton glaucum</i>	C-type (46)
<i>Sinularia flexibilis</i>	C-type, K-type (47, 48)
<i>Sinularia lochmodes</i>	C-type (49)
<i>Sinularia maxima</i>	C-type (50)
<i>Sinularia brassica</i>	E-type (51)
<i>Eunephthya thyrsoidea</i>	X-type (52)
<i>Cladiella pachyclados</i>	K-type (53)
<i>Anthelia glauca</i>	X-type (54)
<i>Protodendron repens</i>	X-type (55)
<i>Ovabunda macrospiculata</i>	X-type (56)
<i>Eunicella cavolini</i>	K-type (57)
<i>Antillogorgia elisabethae</i>	E-type (58)
<i>Antillogorgia bipinnata</i>	E-type, C-type (59)
<i>Leptogorgia alba</i>	C-Type (60)

Table 2-S 6: Amino acid percent identity matrix for TCs from clade 1-IV

HcTC-5-2	100				
HcTC-5g-1	93.69	100			
NA124_045_TC5	61.36	61.36	100		
SCB_216_TC1	64.14	65.66	68.94	100	
NA124_195_TC3	63.64	64.9	68.43	97.98	100

Table 2-S 7: Amino acid percent identity matrix for TCs from clade 1-V

RmTC-1	100													
Rkoe_cbTC	82.49	100												
Basb_cbTC	49.24	47.45	100											
ErycarTC-6	51.65	49.87	59.6	100										
VgTC-1	48.86	48.35	52.14	50.76	100									
Selo_cbTC	52.79	52.3	55.42	55.7	75.06	100								
NA195_TC-5	50.38	48.59	56.09	61.83	53.81	55.73	100							
CrTC-1	50.9	49.1	57.61	62.6	53.81	55.98	93.91	100						
Crub_cbTC	50.9	49.1	57.61	62.6	53.81	55.98	93.91	100	100					
Dgem_cbTC	55.7	53.05	60.71	62.37	55.53	58.44	65.48	65.74	65.74	100				
PcaTC-1	54.43	51.52	52.9	56.31	60.05	61.71	55.33	55.84	55.84	58.25	100			
NA124_218_TC-1	56.96	53.94	57.18	59.85	62.81	65.24	59.9	60.66	60.66	65.41	67.92	100		

Table 2-S 8: Comparison of identified diterpene type and terpene cyclase clades present in sequencing data if both are available from the same species.

Name	Diterpene type	Represented TC clades
<i>Erythropodium caribaeorum</i> ¹	K-type, B-type (36, 37)	1-II, 1-V, 2, 3
<i>Briareum asbestinum</i> ¹	K-type, B-type (38, 39)	1-II, 1-V, 2
<i>Heliopora coerulea</i> ¹	E-type, X-type (38, 39)	1-I, 1-IV, 2, 3
<i>Dichotella gemmacea</i> ¹	B-type (40)	1-V
<i>Stylatula elongata</i> ¹	B-type (43)	1-V
<i>Sarcophyton trocheliophorum</i> ¹	C-type (45)	1-III
<i>Antillogorgia elisabethae</i> ¹	E-type (58)	1-I, 3
<i>Antillogorgia bipinnata</i> ¹	E-type, C-type (59)	1-I, 1-III, 3
<i>Callistephanus simplex</i> NA124-047 ²	C-type	1-III, 3, 4
<i>Callistephanus simplex</i> NA124-141 ²	C-type	1-III, 3, 4
<i>Acanthogorgia</i> sp. NA124-045 ²	X-type	1-I, 1-III, 1-IV, 3, 4
<i>Victorgorgia</i> sp. SCB-084 ²	C-type, X-type	1-III, 4
<i>Paragorgia</i> cf <i>arborea</i> NA124-195 ²	X-type	1-IV, 1-V, 3
<i>Paragorgia jamesi</i> SCB-216 ²	X-type	1-IV, 3
<i>Chrysogorgia</i> sp. 1 SCB-144 ²	E-type, X-type	1-I, 1-II, 2, 3
<i>Chrysogorgia</i> sp. 2 SCB-157 ²	E-type	1-I, 2, 3
<i>Balticina</i> cf. <i>californica</i> NA124-218 ²	-	1-V

1: Data taken from the literature – see cited reference and supplementary file 1 for the source of chemical data and sequencing data, respectively; 2: Data taken from this work.

Table 2-S 9: Population genetic metrics of coral TCs and mitochondrial genes from investigated octocoral species.

		Number sites	trans/ tranv	Watterson (θ^w)	nucleotide diversity (π)	Tajima's D	dN/dS
TC clade	1I	1263	1.925	141.039091	167.7583333	0.82965799	0.227922406
	1II	1236	1.617647	181.632653	183.9047619	0.073648255	0.665153741
	1III	1320	1.676301	142.393088	121.4666667	-0.653433697	0.212881189
	1IV	1188	1.651899	201.12	219.4	0.694513241	0.163847161
	1V	1227	1.992063	124.839507	134.6060606	0.368956268	0.181396701
	2	1284	1.441558	109.316546	104.4771242	-0.188735319	0.572548683
	3	1359	1.216667	29.828672	26.31802721	-0.422610959	0.291384175
	4	1374	1.887755	61.852105	63.11784512	0.073388753	0.136862777
mito	atp6	720	3.619048	16.353529	8.354019494	-1.505534658	0 (dN=0)
	atp8	297	1.941176	8.429654	2.596306894	-2.046994629	undefined (dS=0)
	cox2	801	4.111111	31.021126	18.96262184	-1.222045977	0.08306654
	mutS	3675	Na	NA	NA	NA	1
	nad4	1524	Na	NA	NA	NA	1
	nad4L	312	2.714286	8.773851	3.707966599	-1.713828086	0.101407718
	nad6	780	4.142857	18.222613	9.500789889	-1.480291239	0.072105138

Table 2-S 10: Primers used in this study

Gene Name	Primer Sequence	Annealing Temperature
NA 124_045 TC1	FWD: ccgcgcggcagccatATGTCGTGTAGCAACGAAGT REV: cctgggtgccgcgcggcagccatATGCCTCATGTAATAACG	72
NA124_045 TC2	FWD: gcctgggtgccgcgcggcagccatgATGTCTTGCATAACGATGT REV: gccggatctcagtggtgggtgggtggtgctcgagCTAATCACCTAAAATGTATTCAATT	57
NA 124_045 TC3	FWD: tgccgcgcggcagccatgATGTCATGCAGTAAAGAAATTC REV: gatctcagtggtgggtgggtggtgctcgagTTAGCTGATGTCCAATTGTG	58
NA 124_045 TC4	FWD: tgggtgccgcgcggcagccatATGGCGTGCAGTAAAGAATT REV: gtcgacggagctcgaattcggatccTATTCTAGCAAAAGATTTCATGC	72
NA 124_045 TC5	FWD: tgccgcgcggcagccatgATGGCTTGCAGTAAACATATTCG REV: gatctcagtggtgggtgggtggtgctcgagTCAAAATGCGTAAAGCATTTCATC	62
NA 124_47 TC2	FWD: cgccctgggtgccgcgcggcagccatATGTCCTGTAGCAAGGTAGTTCGATATCCAAATG REV: tcgacggagctcgaattcggatccCTATGTAAACTGTCCAGATGCTGGAATCACGCCGG	72
NA 124_47 TC3	FWD: gcctgggtgccgcgcggcagccatgATGTCCTCATGTAATAACGC REV: cagccggatctcagtggtgggtgggtggtgctcgagTCAACTGCACAGCATATCAA	59
NA 124_47 TC4	FWD: tgggtgccgcgcggcagccatATGGCGTGCTCCAAAGAATTG REV: cggagctcgaattcggatccTCAGAGATTCCATGCATATTGACATGCCCGC	72
NA 124_141 TC1	FWD: tgggtgccgcgcggcagccatATGGCGTGCTCCAAAGAACT REV: gcttgctgacggagctcgaattcggatccTCAGGCTCTCAAAAGATTGCAC	72
NA 124_141 TC4	FWD: cctgggtgccgcgcggcagccatATGTCCTCATGTAATAACGCCGTACGTGTCCC REV: aagcttgctgacggagctcgaattcggatccTCAACTGCACAGCATATCAACGACGTTGG	72
NA 124_141 TC5	FWD: tgggtgccgcgcggcagccatATGGCGTGCTCCAAAGAATT REV: cgcaagcttgctgacggagctcgaattcggatccTCAGAGATTCCATGCATATTGAC	72
NA 124_195 TC1	FWD: tgccgcgcggcagccatATGTCGTGCAGCAAGAAAT REV: cggagctcgaattcggatccTTATACTGGAATGAAATGTCCACGTGG	72
NA 124_195 TC3	FWD: tgccgcgcggcagccatgATGGCTTGCAGTAAACACGTTCCG REV: gatctcagtggtgggtgggtggtgctcgagTCAAAATGCATAAAGCCTTTTCATC	62
NA 124_195 TC5 Exon 1	FWD: gcctgggtgccgcgcggcagccatgATGTGATGGACAGGAAATCTACGTTCC REV: GTAACGGAATTCAGGAATTGTACGACCTTTGTTGCTGCATCGTCGAAATCAACTG	69
NA 124_195 TC5 Exon 2	FWD: GTTGATTTCGACGATGCAGCAACAAAGGTCGTACAATTCCTGAATTCC REV: ggatctcagtggtgggtgggtggtgctcgagTTAATCTAAGCCATAAAGAAAAACCATC	65
NA 124_218 TC1	FWD: tgccgcgcggcagccatgATGTCAAGCCTGCACAC REV: gatctcagtggtgggtgggtggtgctcgagTTATTACTCAATCCATAAAGAGCAATC	60
NA 124_218 TC2	FWD: tgccgcgcggcagccatgATGTGGCGCACCGATGTT REV: gatctcagtggtgggtgggtggtgctcgagTCATGGGATACTGAATTGTCTATCTAC	63
SCB 84 TC1	FWD: tgccgcgcggcagccatATGTCGTGTAGCAAGGAAGTTC REV: cggagctcgaattcggatccCTATTACTAAGTAACTAAGTAACTGTCCACAGACCGA	72
SCB 84 TC2	FWD: gcgcggcagccatATGTCGTGTAGCAAGGCAGT REV: cggagctcgaattcggatccCTATTCATTAAGTAACTAAGTACACTGTCCACAGAC	72

Table 2-S 10: Primers used in this study (Continued)

SCB 84 TC3	FWD: gcctggtgccgcgcggcagccatgATGCGAGTGACAATTCATG REV: cagccggatctcagtggtggtggtggtgctcgagCTAAATTATACTGCAACCGTAC	58
SCB 84 TC5	FWD: tgggtgccgcgcggcagccatATGCATAAGCACAACTTGAT REV: cttgtcgacggagctcgaattcggatccTTAATTAAGTAAATTGTCAACTGCACG	72
SCB216 TC1	FWD: tgccgcgcggcagccatgATGGCTTGACAGTAAACACGTTTCG REV: gatctcagtggtggtggtggtgctcgagTCAAAATGCATAAAGCCTTTTCATC	62
SCB 144 TC1	FWD: tgggtgccgcgcggcagccatATGGCGTGACAGTAAACCGGA REV: gcaagcttgctgacggagctcgaattcggatccTCATGTTGGTACAAAATGTCCACACG	72
SCB 144 TC3	FWD: tgccgcgcggcagccatgATGTCGTGCAGCAAGCC REV: gatctcagtggtggtggtggtgctcgagTTATTTATCCAGTGTTTTATTAAGTGTCC	62
SCB 144 TC4	FWD: ccgcgcggcagccatATGGCGTACAGTAAACCAAT REV: cggagctcgaattcggatccTCATACAACGTTATATAATCCATCCTC	72
SCB 144 TC7 Exon 1	FWD: tgggtgccgcgcggcagccatATGCTGTCCACGTATGTTGT REV: AACAGCGCCTCCAAAGCATGTCACCTTCAGTCACATTGGC	72
SCB 144 TC7 Exon 2	FWD: GCCAATGTGACTGAAGGTGACATGCTTTGGAGGCGCTGTTT REV: cggagctcgaattcggatccCTAATTTTTAAACAGACCTTCTCCATCAGT	72
SCB 157 TC2	FWD: cgccgcggcagccatATGGCGTACAGTAAGCCAGA REV: cggagctcgaattcggatccTCATGTTGGCACAAAATGTC	72
SCB 157 TC3	FWD: tgccgcgcggcagccatgATGTCGTGCAGTAAACCG REV: gatctcagtggtggtggtggtgctcgagTTATTCATCTAACACAAGTGTTTTATTAA	61
SCB 157 TC5	FWD: ggcagccatATGGCGTATAGTAAACCAATACATGCTCCAAGTGATGGGTAAA REV: gtggtggtgctcgagTCATACAACGTTATACAATCCATCTTCTTTGTATGCAGATATTTT	72
SCB 157 TC6	FWD: gcctggtgccgcgcggcagccatgATGGCATAAGTAAACCAATACAT REV: ggatctcagtggtggtggtggtgctcgagTTATATGTCATACAATCCATCTTTTTTGG	61
SCB 157 TC8	FWD: gcctggtgccgcgcggcagcgcctggATGGCATAAGTATCCCAG REV: gccggtctcagtggtggtggtggtgctcgagTTATATGTCATAAAATCCATCATCC	59

Scleralcyonacea Malacalcyonacea



49

Acanthogorgia sp. - NA124-045

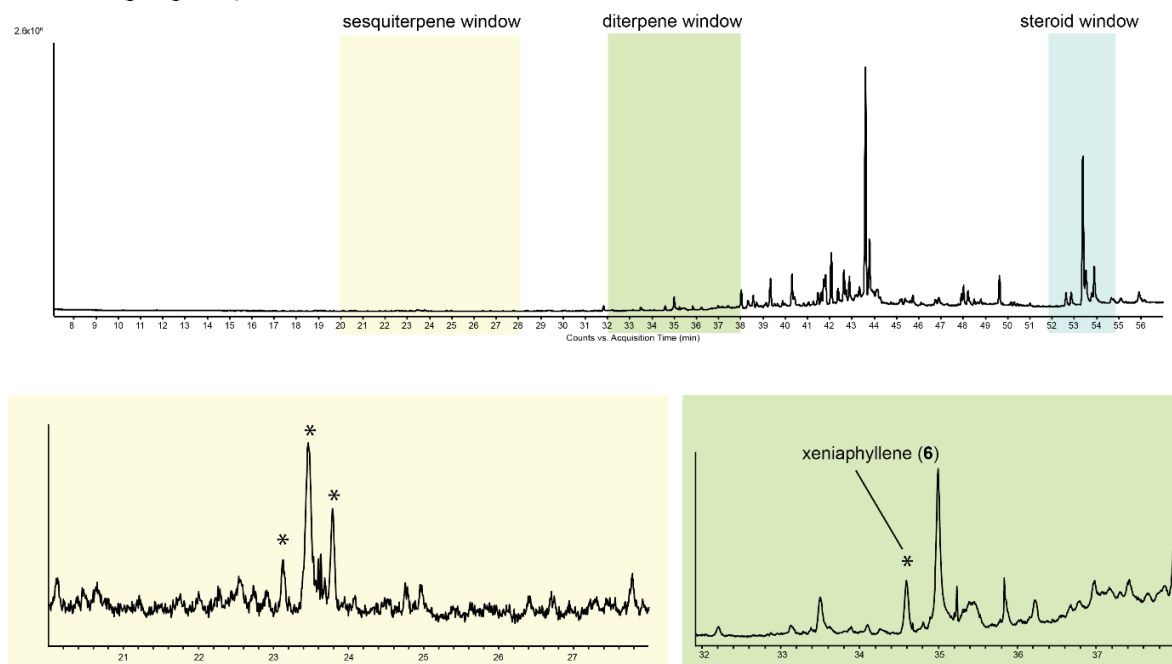


Figure 2-S 2: GCMS analysis of hexane extract derived from *Acanthogorgia sp.* NA124-045.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection. Signals labeled with asterisks represent terpene hydrocarbons.

Callistephanus simplex - NA124-047

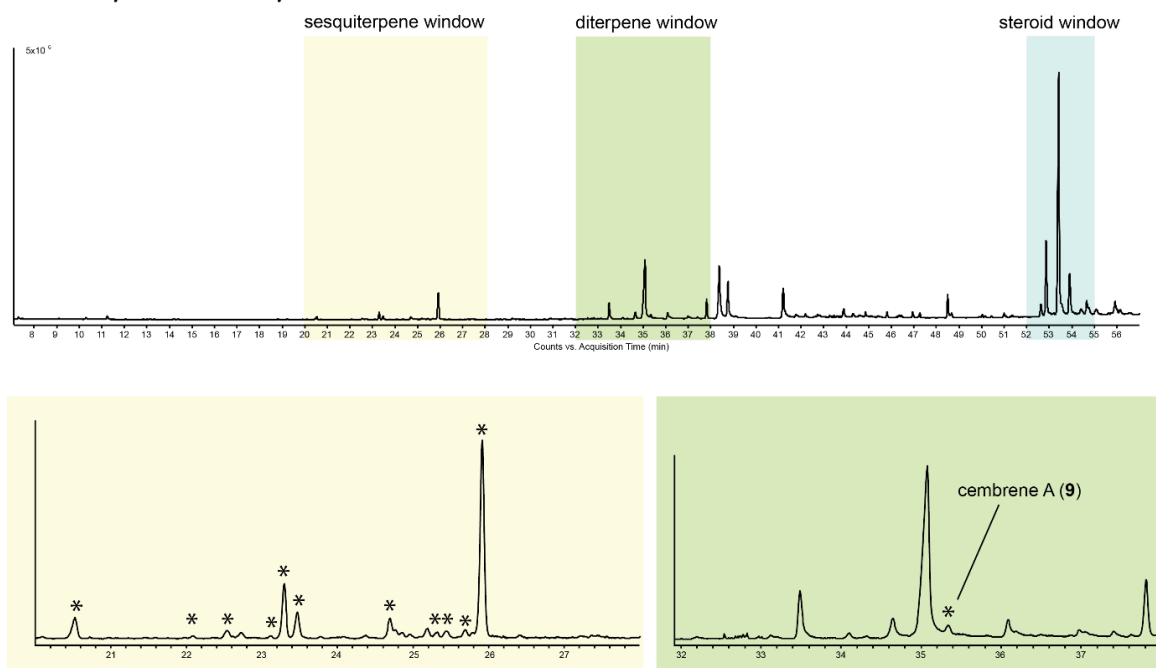


Figure 2-S 3: GCMS analysis of hexane extract derived from *Callistephanus simplex* NA124-047.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection. Signals labeled with asterisks represent terpene hydrocarbons.

Callistephanus simplex - NA124-141

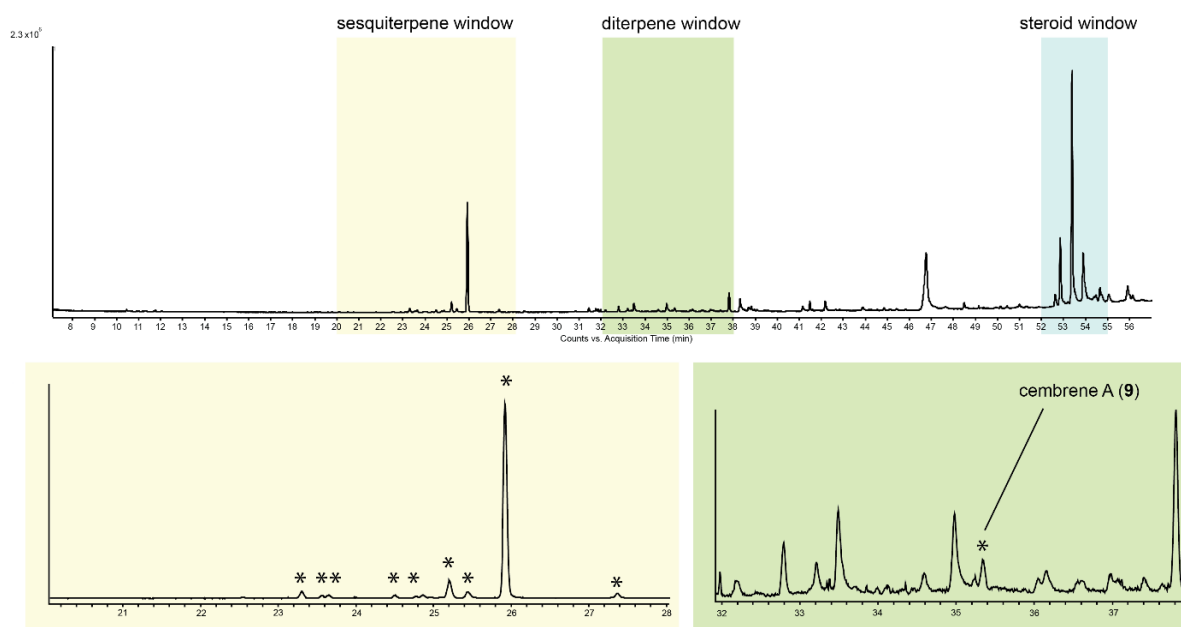


Figure 2-S 4: GCMS analysis of hexane extract derived from *Callistephanus simplex* NA124-141.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection. Signals labeled with asterisks represent terpene hydrocarbons.

Paragorgia cf. arborea - NA124-195

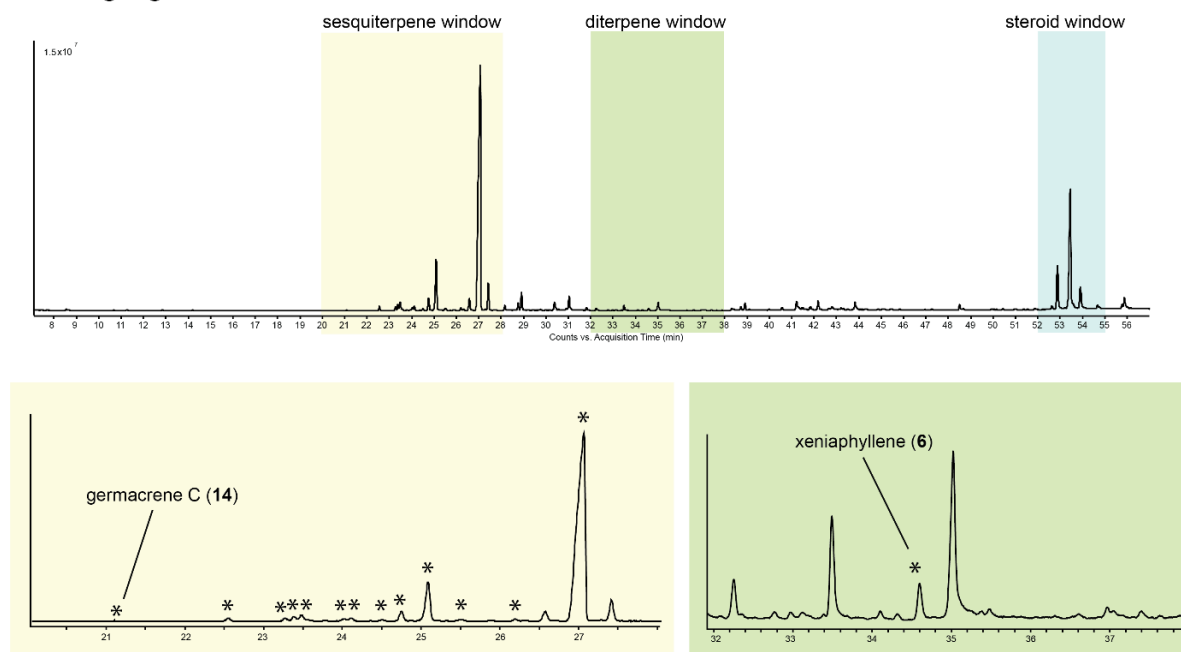


Figure 2-S 5: GCMS analysis of hexane extract derived from *Paragorgia cf. arborea* NA124-195.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection. Signals labeled with asterisks represent terpene hydrocarbons.

Balticina cf. californicum - NA124-218

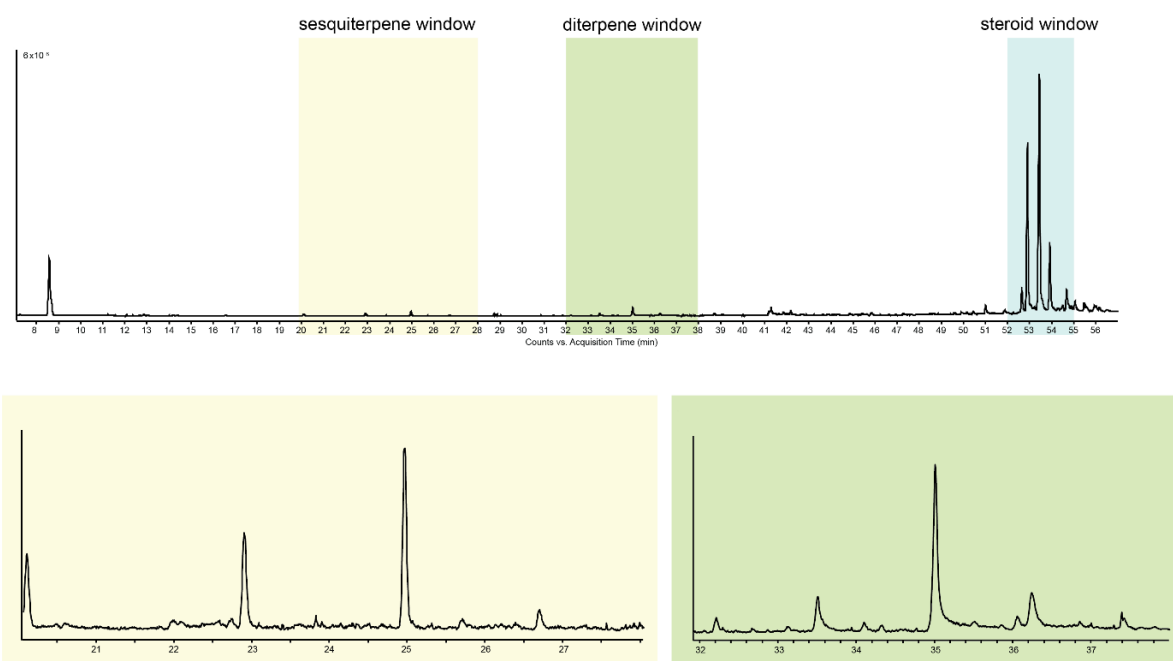


Figure 2-S 6: GCMS analysis of hexane extract derived from *Balticina cf. californicum* NA124-218.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection.

Victorgorgia sp. - SCB-84

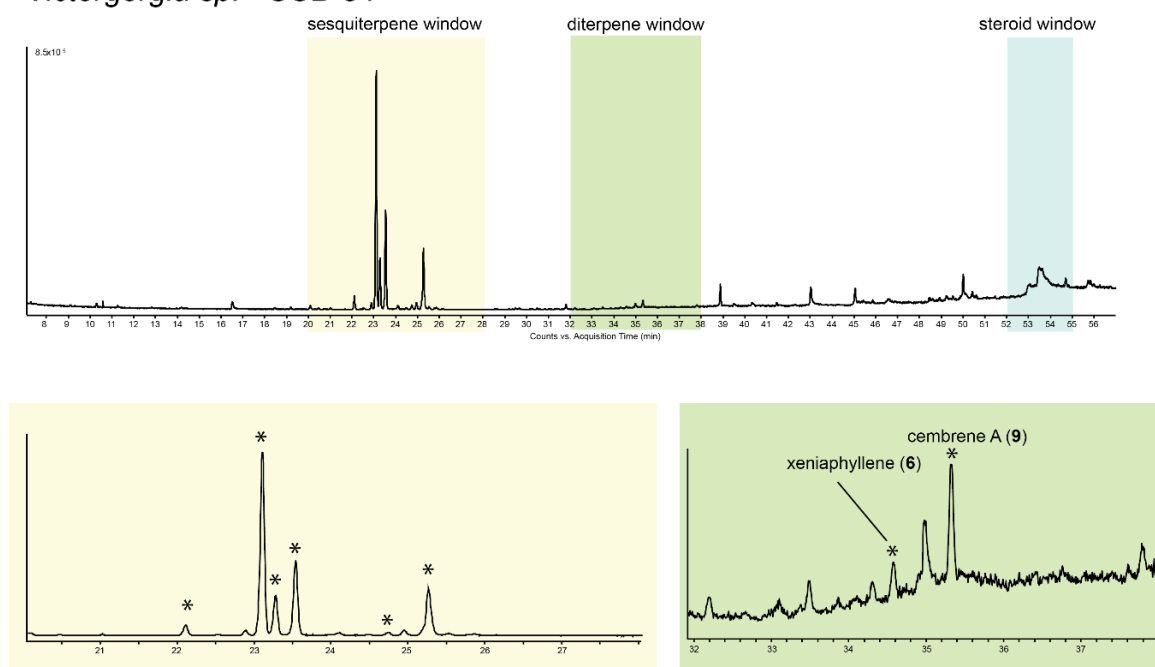


Figure 2-S 7: GCMS analysis of hexane extract derived from *Victorgorgia sp.* SCB-84.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection. Signals labeled with asterisks represent terpene hydrocarbons.

Chrysogorgia sp. - SCB-144

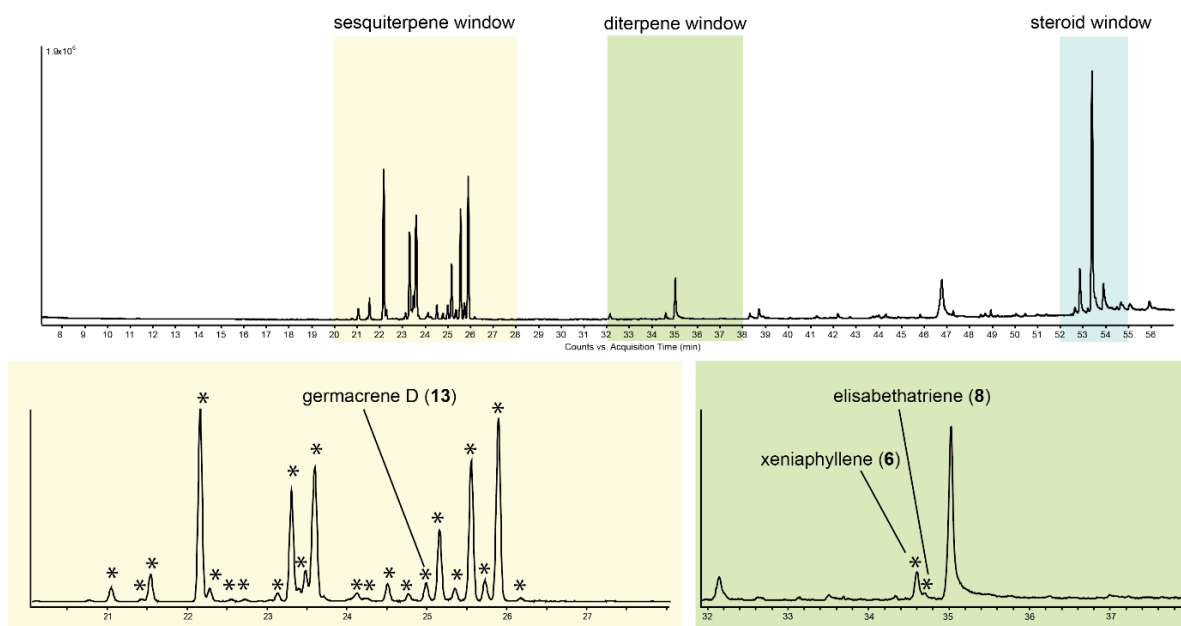


Figure 2-S 8: GCMS analysis of hexane extract derived from *Chrysogorgia* sp. SCB-144.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection. Signals labeled with asterisks represent terpene hydrocarbons.

Chrysogorgia sp. - SCB-157

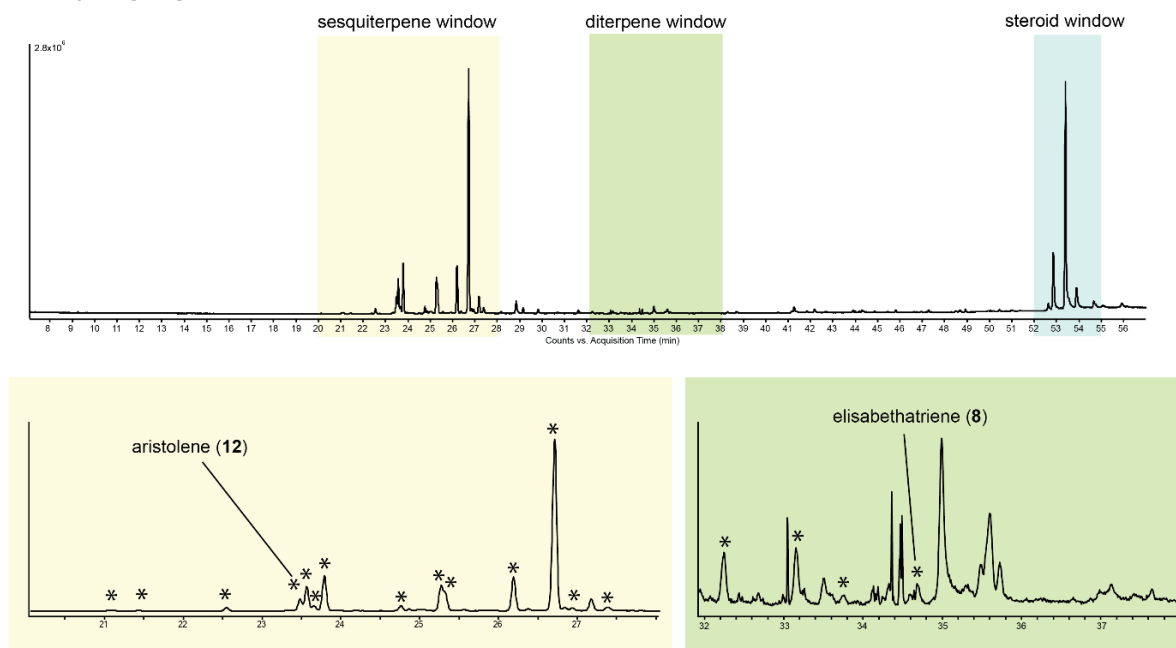


Figure 2-S 9: GCMS analysis of hexane extract derived from *Chrysogorgia* sp. SCB-157.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection. Signals labeled with asterisks represent terpene hydrocarbons.

Paragorgia jamesi - SCB-216

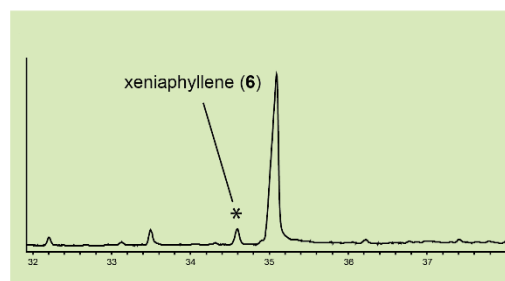
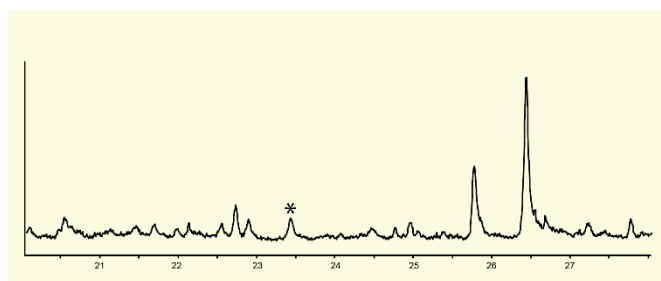
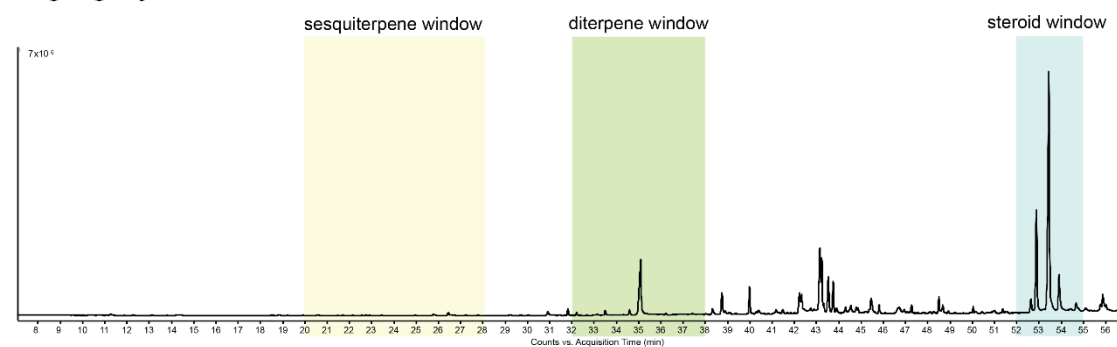


Figure 2-S 10: GCMS analysis of hexane extract derived from *Paragorgia jamesi* SCB-216.

Top: Complete TIC. Bottom: Magnified time windows for sesqui- and diterpene detection. Signals labeled with asterisks represent terpene hydrocarbons.

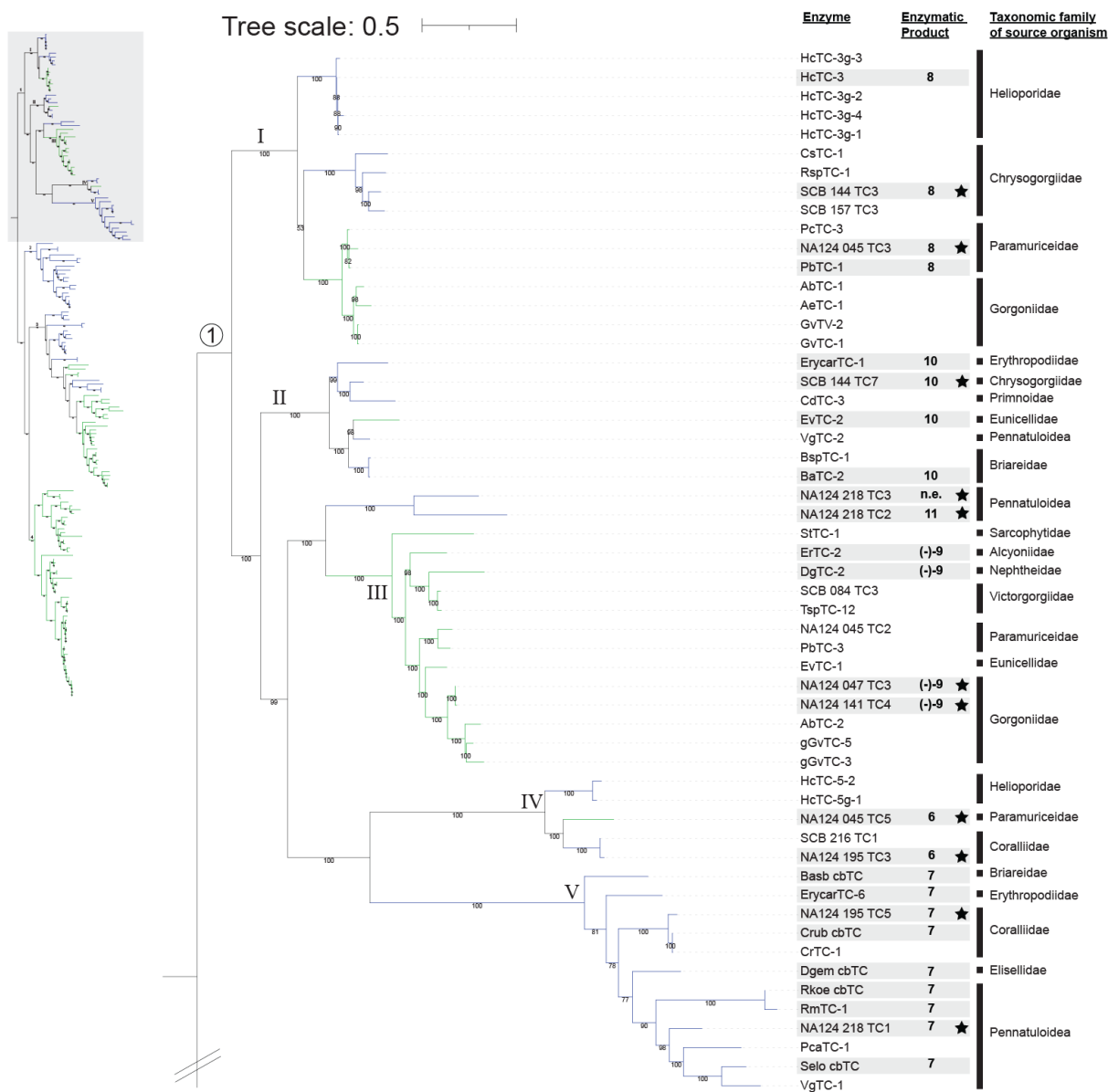


Figure 2-S 11: Phylogenetic tree of octocoral TC amino acid sequences.

The four parts of the tree are labeled by numbers in circles. The subclades of part 1 are labeled with roman numbers. Green = sequences from malacalcyonaceans; blue = sequences from scleralcyonaceans. Bootstrap values are shown. The taxonomic family of the source organism is designated on the right. Enzymes characterized in this work are marked with a star. Abbreviations: m, multiple products; n.a., inactive protein; n.e., not expressed protein; ST, unknown sesquiterpene. The figure is continued on the next two pages.

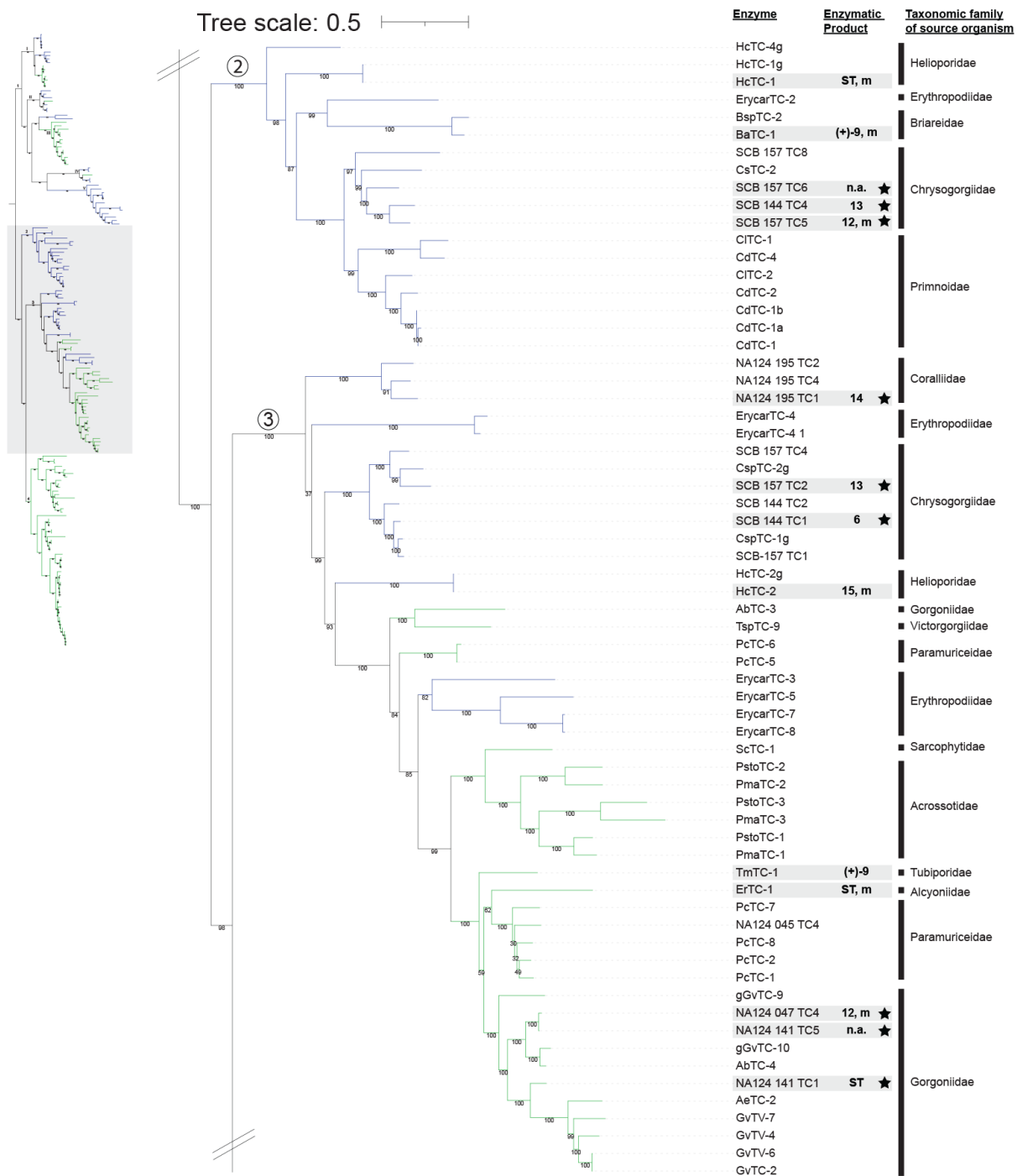


Figure 2-S 11 continued.

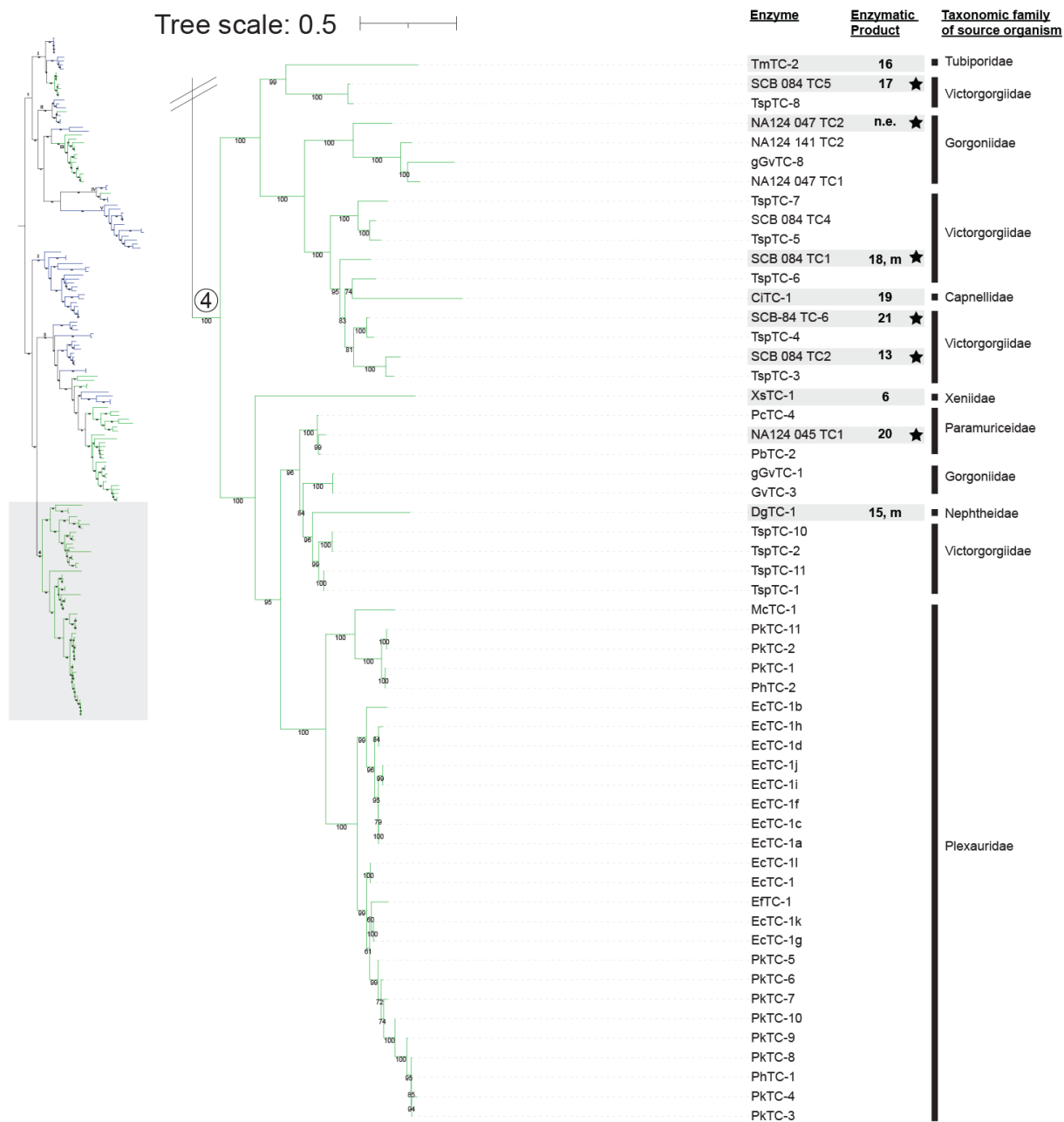


Figure 2-S 11 continued.

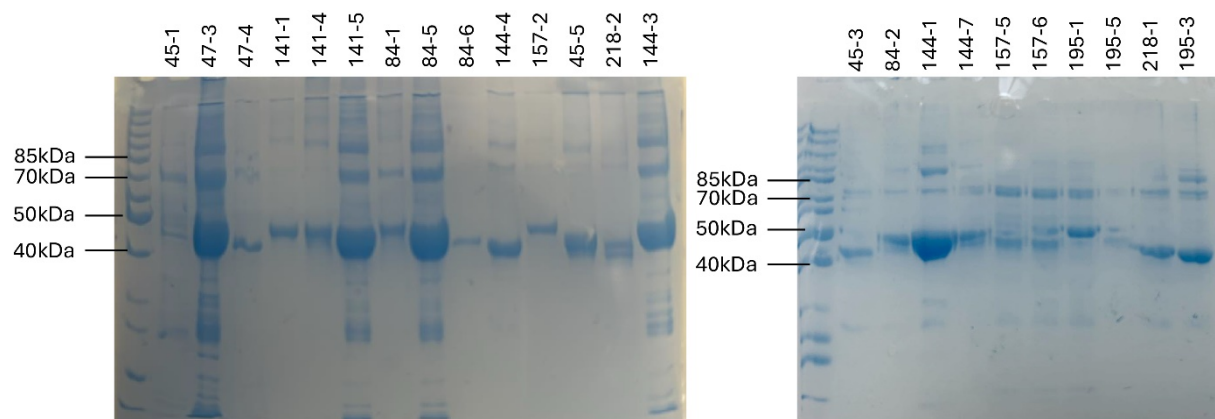


Figure 2-S 12: SDS-PAGE analysis of heterologously produced enzymes.

NA124-045-TC-1

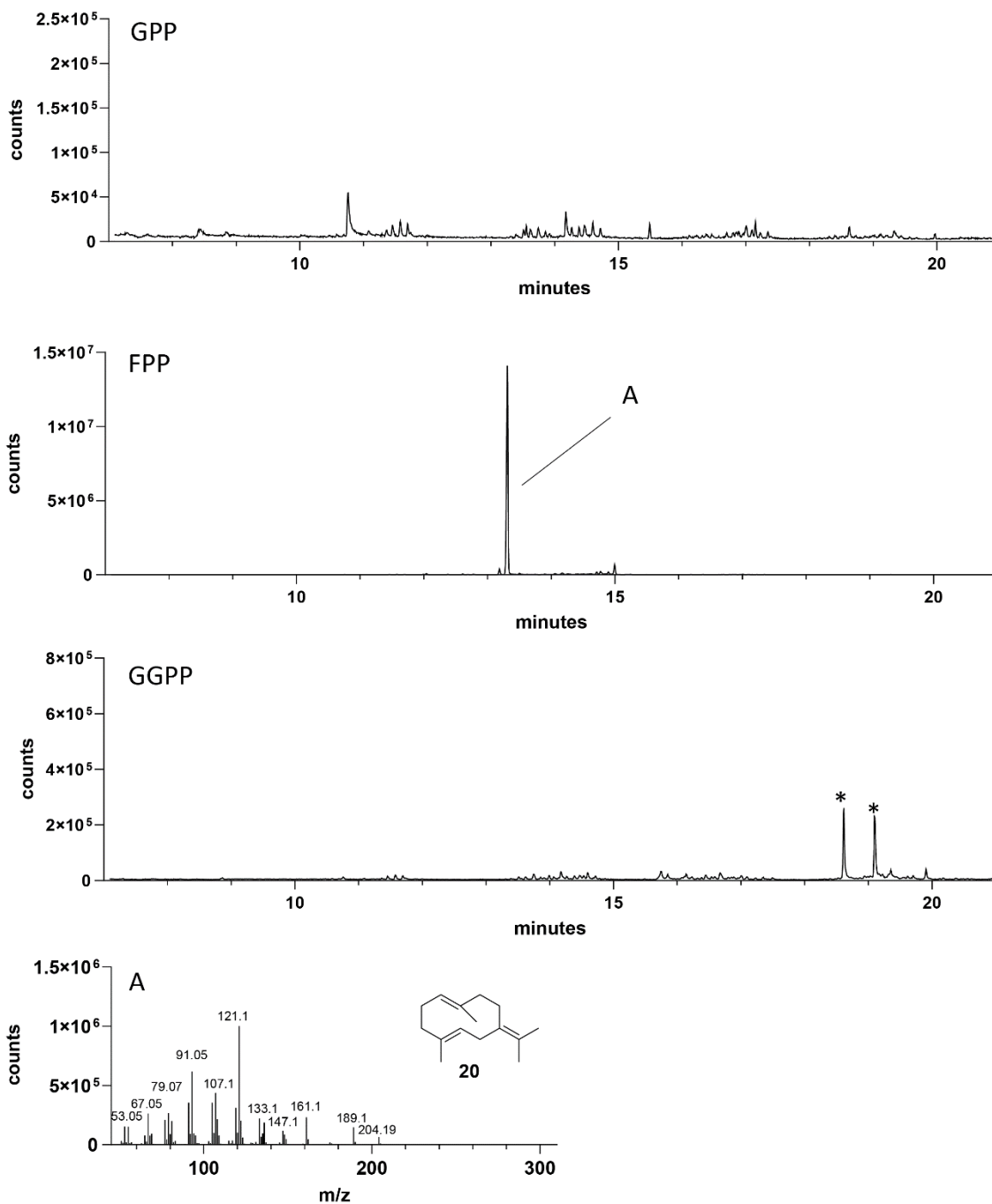


Figure 2-S 13: GCMS total ion chromatograms of NA124-045-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The identity of **20** has been verified using an authentic standard.

NA124-45-TC-3

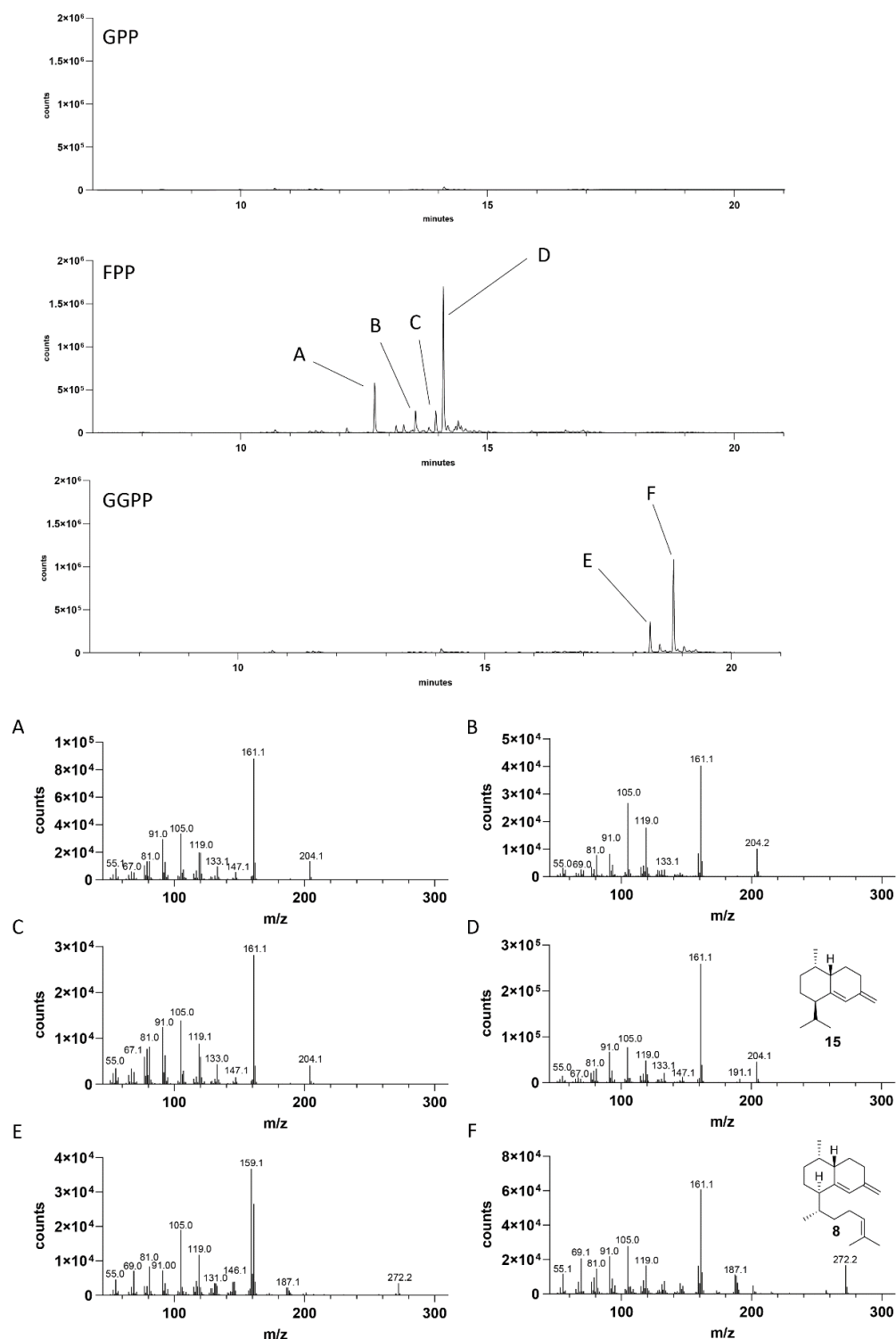


Figure 2-S 14: GCMS total ion chromatograms of NA124-045-TC-3 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identities of **8** and **15** have been verified using authentic standards.

NA124-045-TC-5

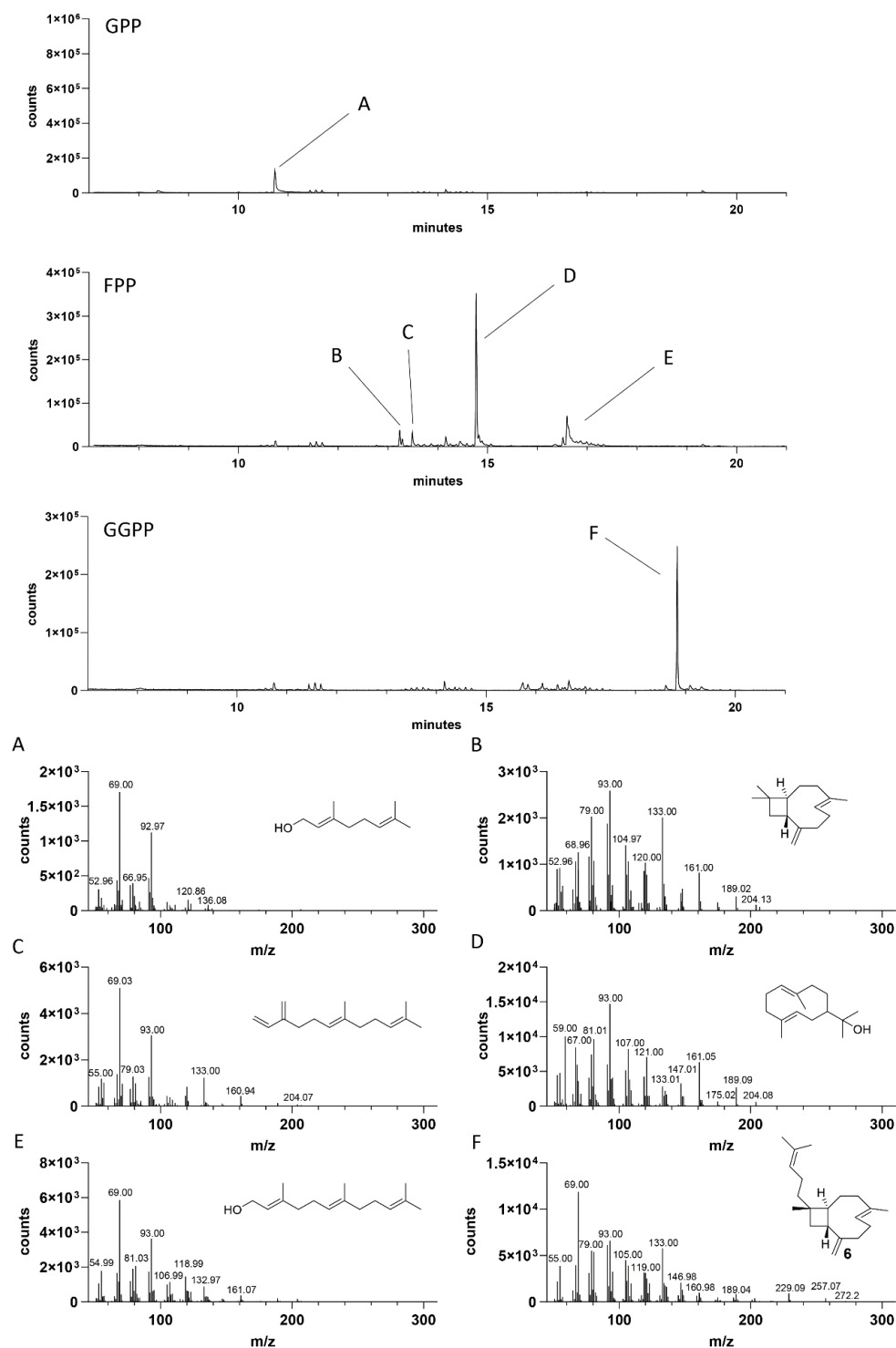


Figure 2-S 15: GCMS total ion chromatograms of NA124-045-TC-5 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. **6** was verified using an authentic standard, the other shown structures are best hits from the NIST database.

NA124-047-TC-3

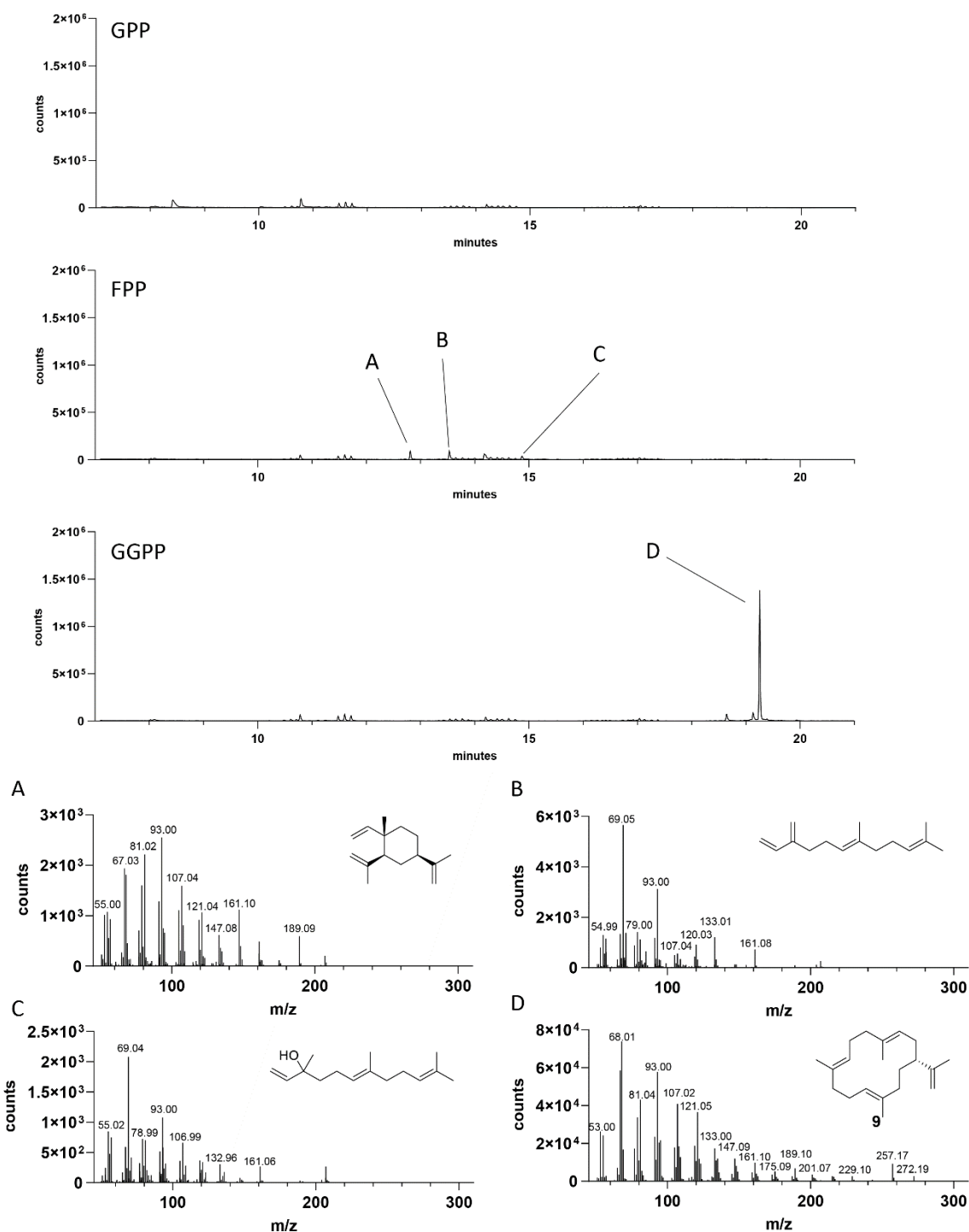


Figure 2-S 16: GCMS total ion chromatograms of NA124-047-TC-3 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. **9** was verified using an authentic standard, the other shown structures are best hits from the NIST database.

NA124-47-TC-4

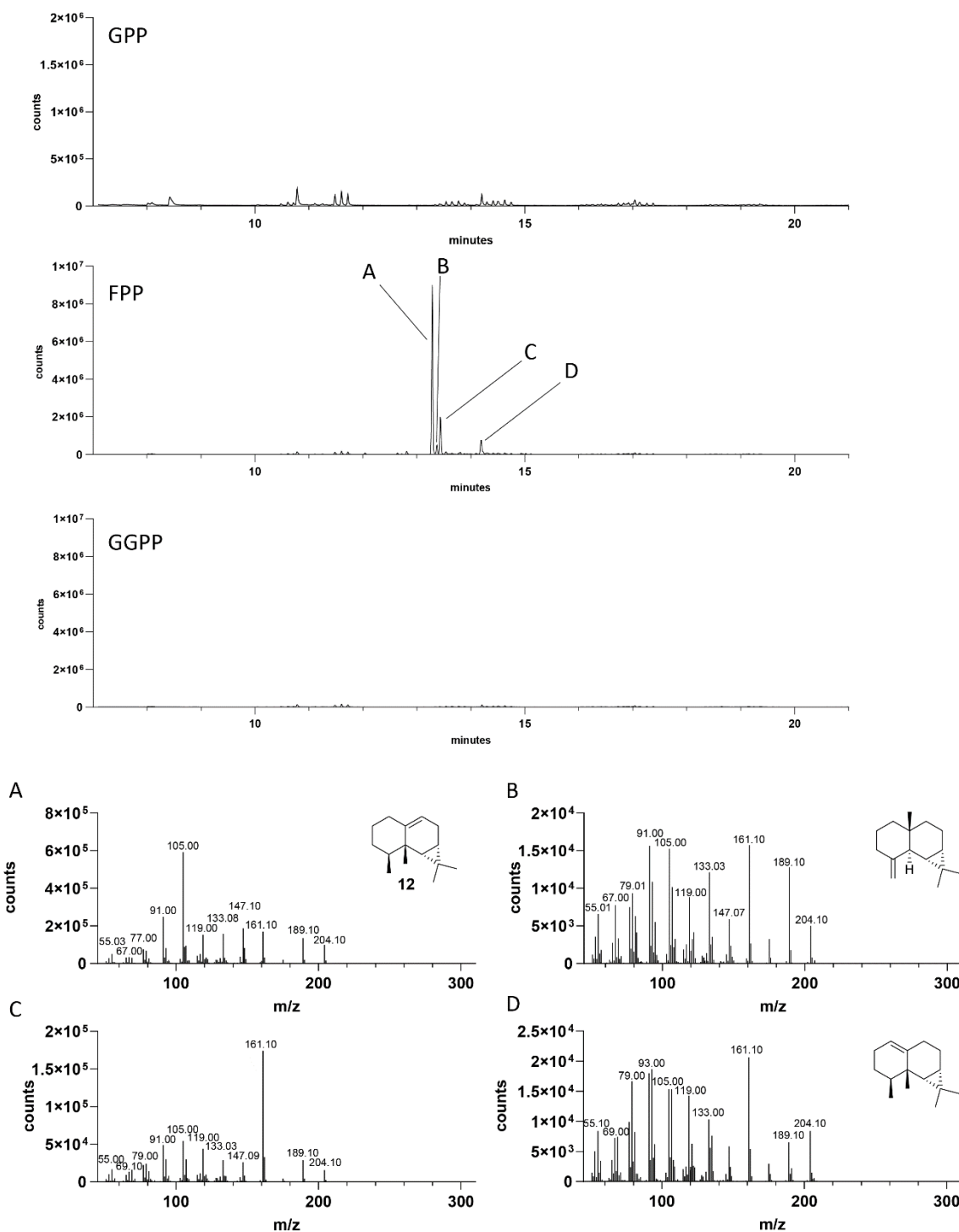


Figure 2-S 17: GCMS total ion chromatograms of NA124-047-TC-4 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. **12** was verified using an authentic standard, the other shown structures are best hits from the NIST database.

NA124-141-TC-1

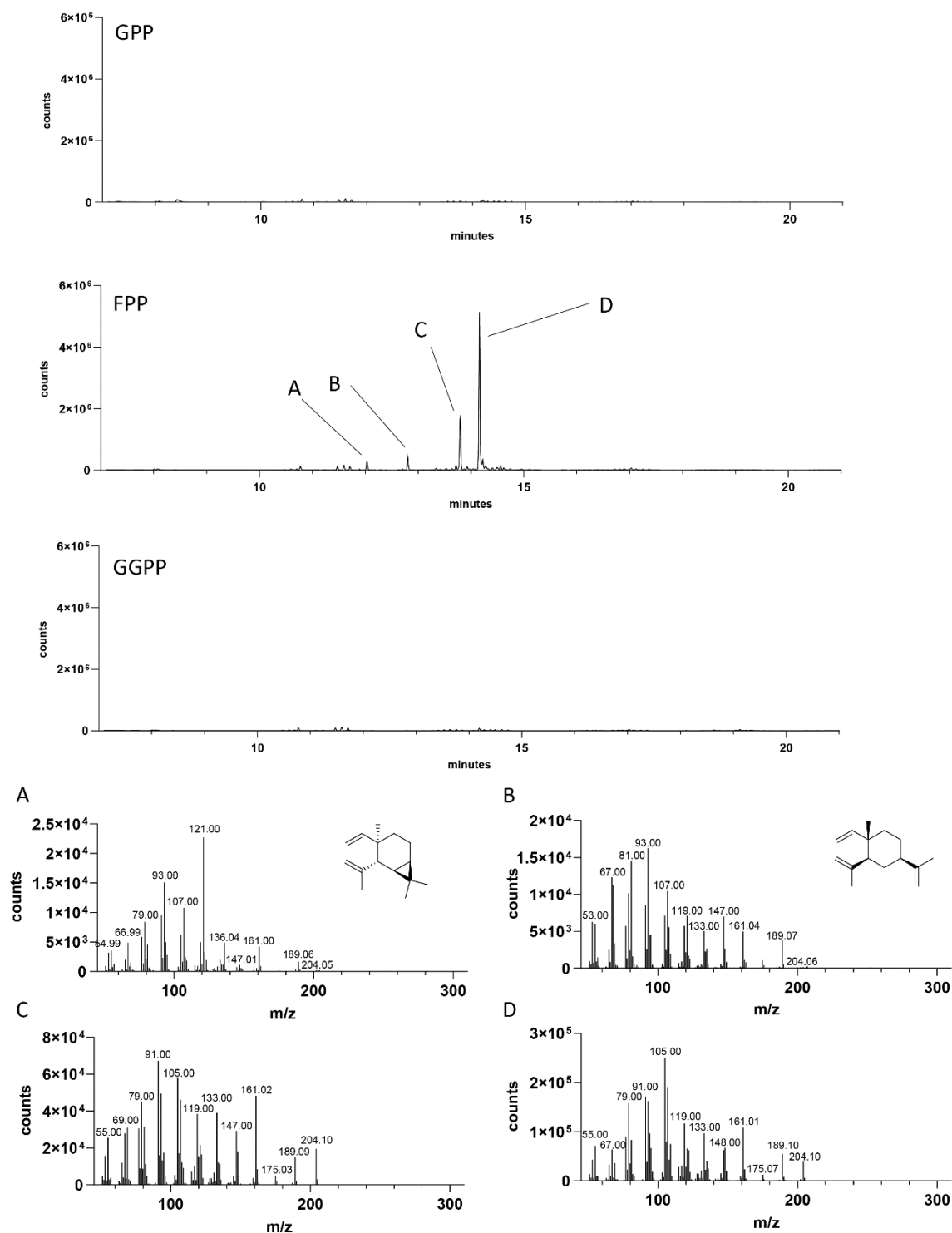


Figure 2-S 18: GCMS total ion chromatograms of NA124-141-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The shown structures are best hits from the NIST database.

NA124-141-TC-4

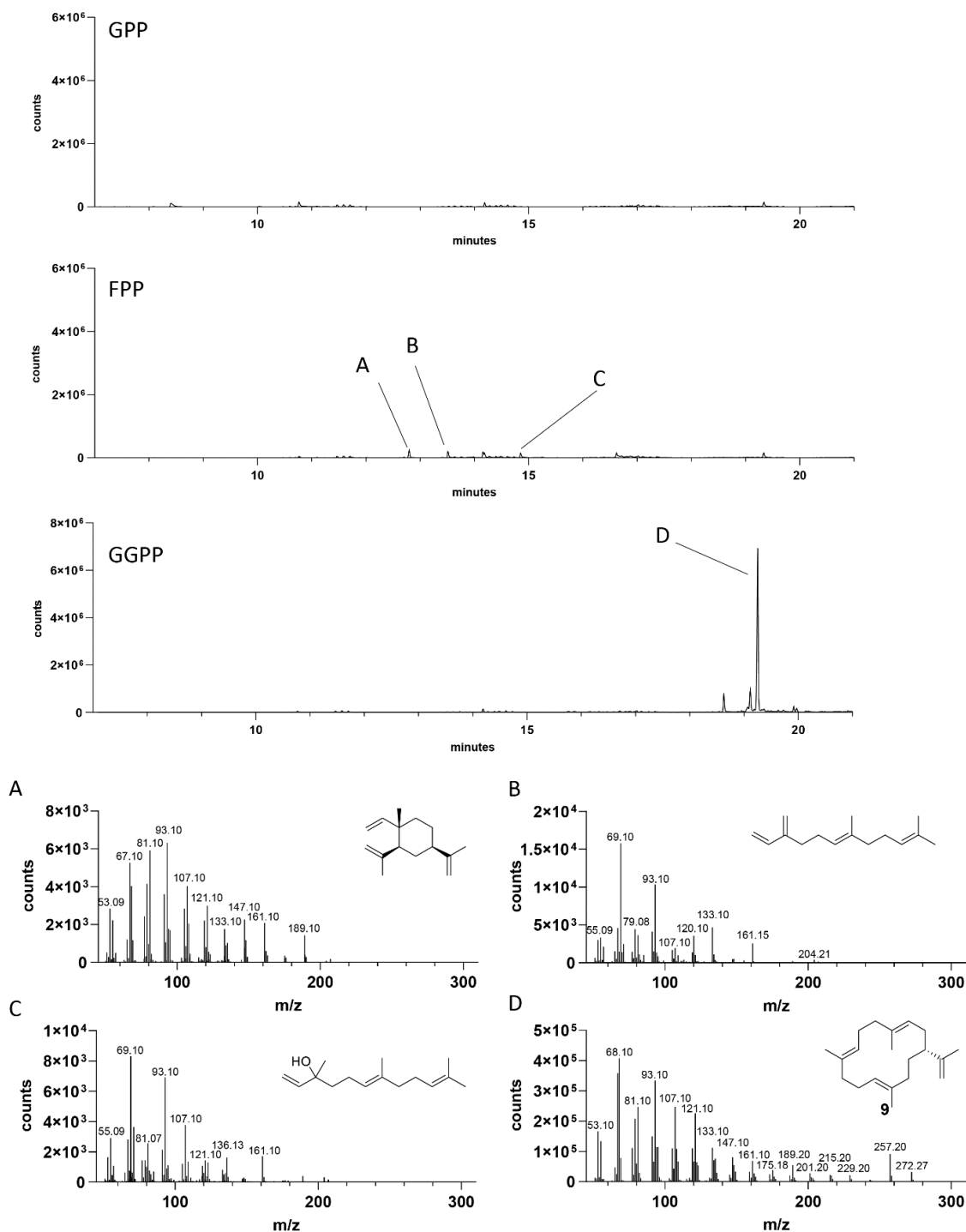


Figure 2-S 19: GCMS total ion chromatograms of NA124-141-TC-4 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. Mass spectra of enzymatically produced terpenes are shown. **9** was verified using an authentic standard, the other shown structures are best hits from the NIST database.

NA124-195-TC-1

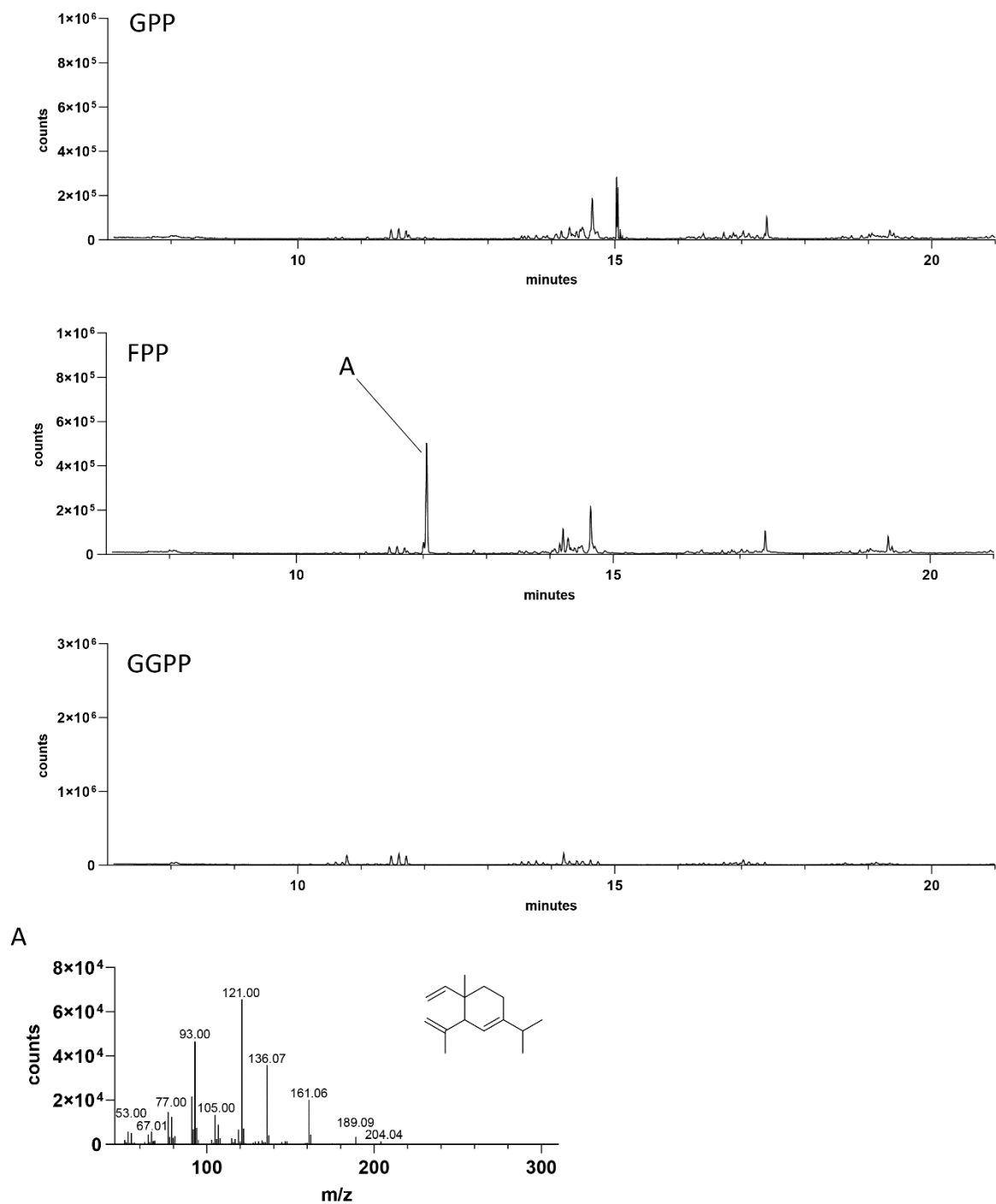


Figure 2-S 20: GCMS total ion chromatograms of NA124-15-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The structure of compound A (d-elemene) was determined by a NIST database hit and is the thermal cope rearrangement product of the enzymatically produced compound germacrene C (**14**).

NA124-195-TC-3

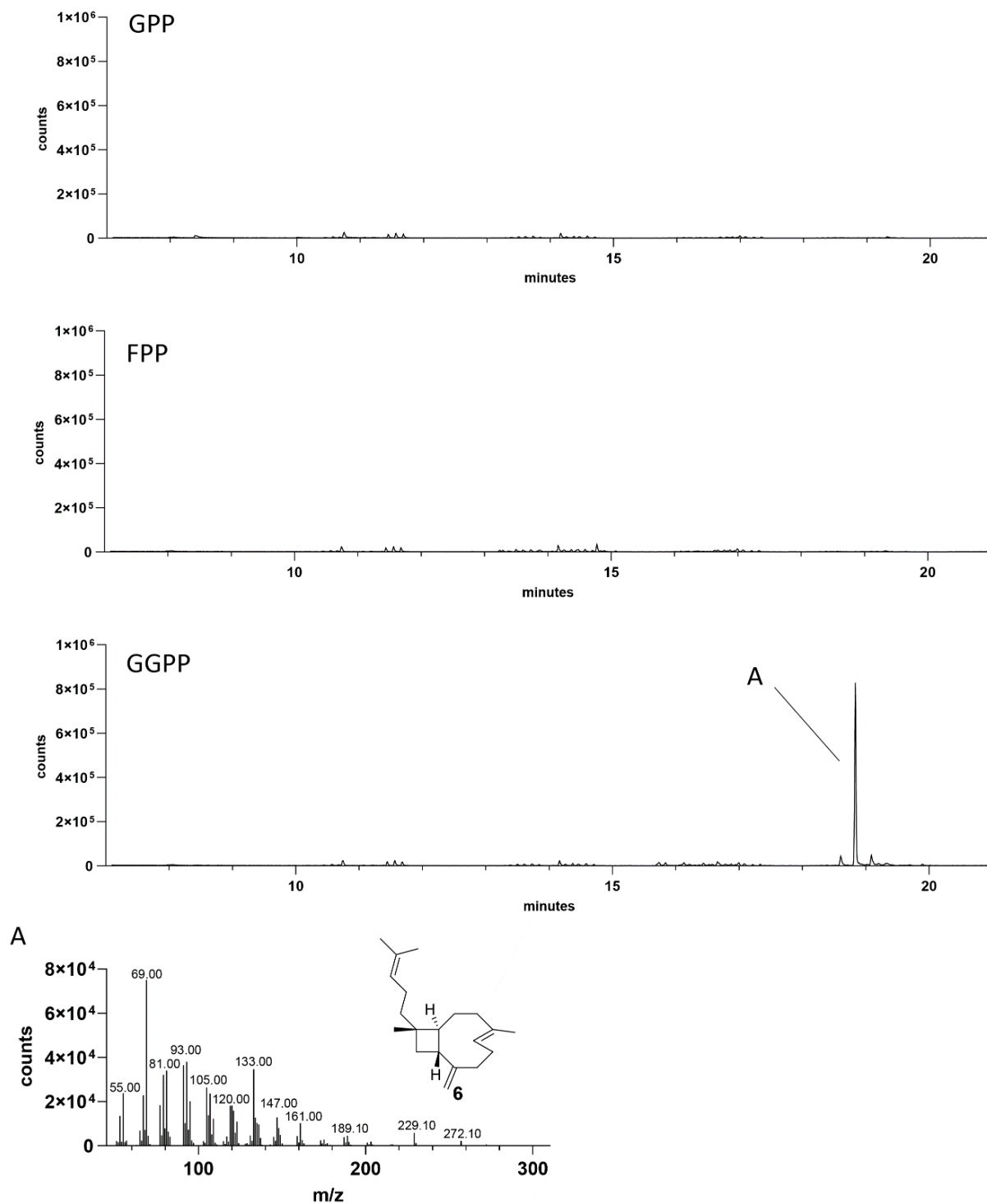
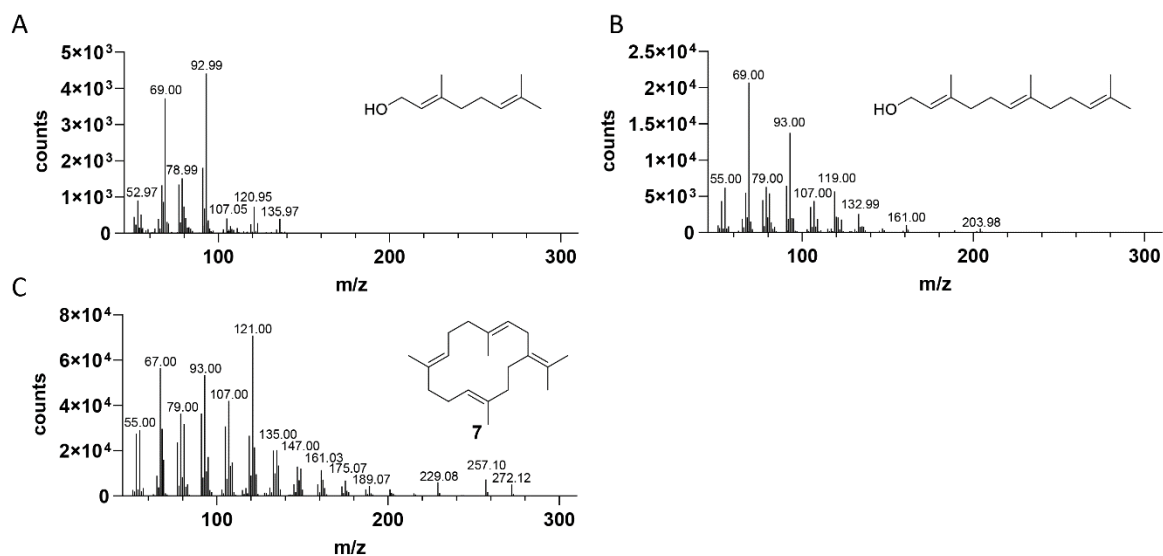


Figure 2-S 21: GCMS total ion chromatograms of NA124-195-TC-3 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. **6** was verified using an authentic standard.

The figure displays three stacked chromatograms, each representing the separation of a different lipid species: GPP (top), FPP (middle), and GGPP (bottom). The x-axis for all plots is 'minutes' (0 to 20), and the y-axis is 'counts'.

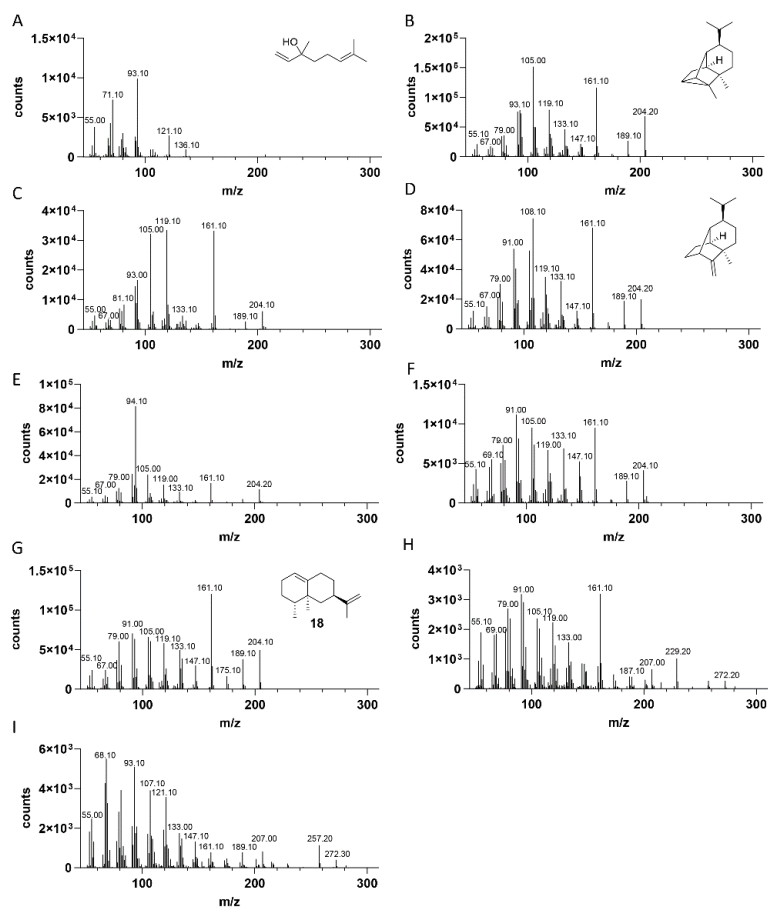
- GPP Chromatogram:** The y-axis scale is 0 to 1×10^6 . A single prominent peak labeled 'A' is observed at approximately 11.5 minutes, reaching a count of about 1.5×10^5 .
- FPP Chromatogram:** The y-axis scale is 0 to 1×10^6 . A single prominent peak labeled 'B' is observed at approximately 17.5 minutes, reaching a count of about 2.5×10^5 .
- GGPP Chromatogram:** The y-axis scale is 0 to 2×10^6 . A single prominent peak labeled 'C' is observed at approximately 19.5 minutes, reaching a count of about 1.7×10^6 .



Mass spectra of enzymatically produced terpenes are shown. **7** was verified using an authentic standard, the other shown structures are best hits from the NIST database.

The figure displays three stacked chromatograms showing the separation of GPP, FPP, and GGPP over a 20-minute period. The y-axis represents 'counts' (0 to 2×10^6) and the x-axis represents 'minutes' (0 to 20).

- GPP Chromatogram:** Shows a single prominent peak labeled **A** at approximately 8 minutes.
- FPP Chromatogram:** Shows multiple peaks labeled **B**, **C**, **D**, **E**, **F**, and **G** between 10 and 15 minutes.
- GGPP Chromatogram:** Shows two late peaks labeled **H** and **I** at approximately 19 and 20 minutes, respectively.



Mass spectra of enzymatically produced terpenes are shown. **18** was verified using an authentic standard, the other shown structures are best hits from the NIST database.

SCB-84-TC-2

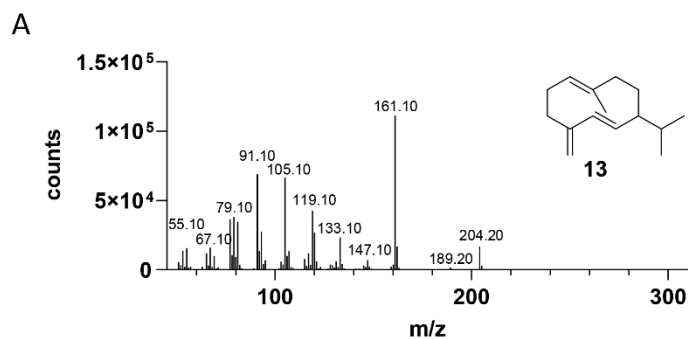
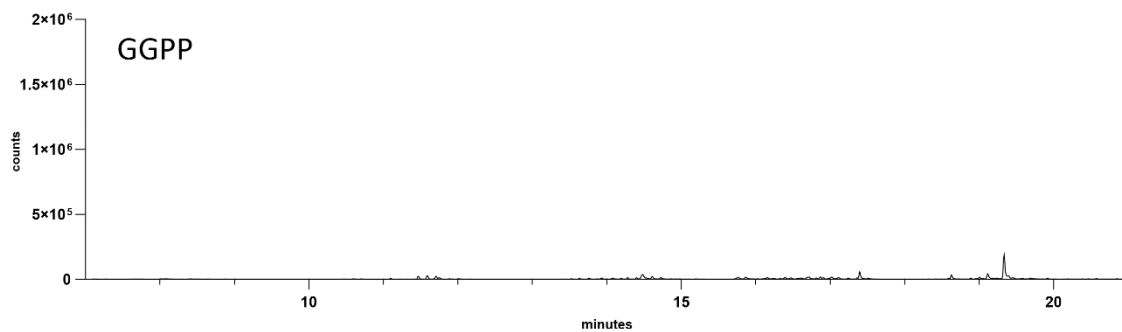
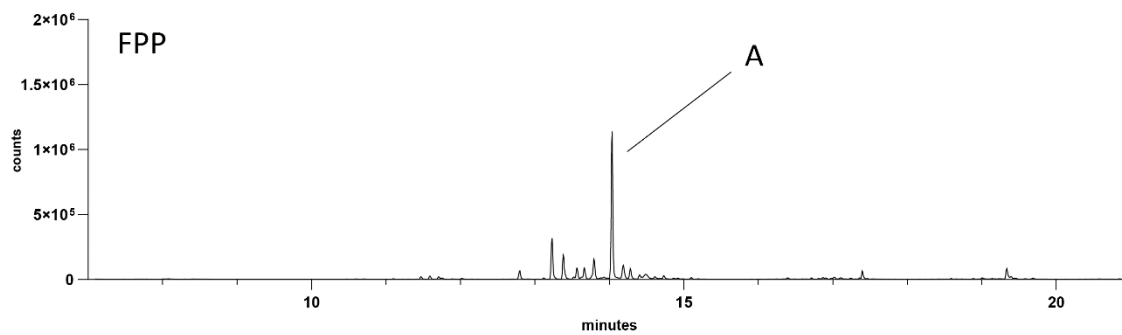
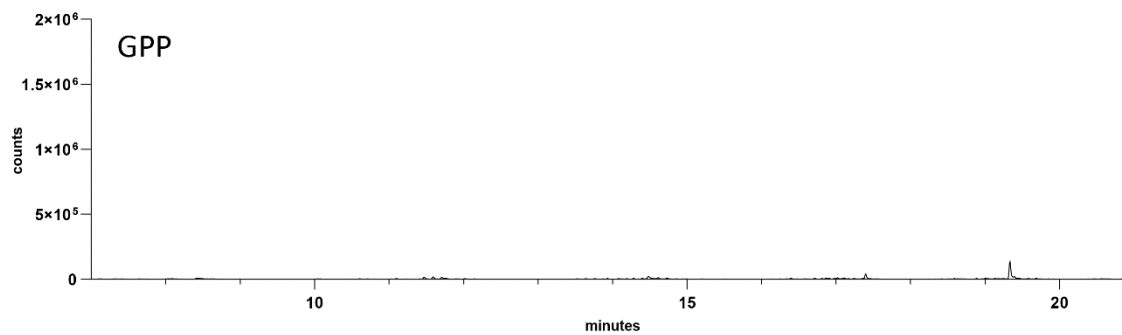


Figure 2-S 24: GCMS total ion chromatograms of SCB-084-TC-2 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. **13** was verified using an authentic standard.

SCB-84-TC-5

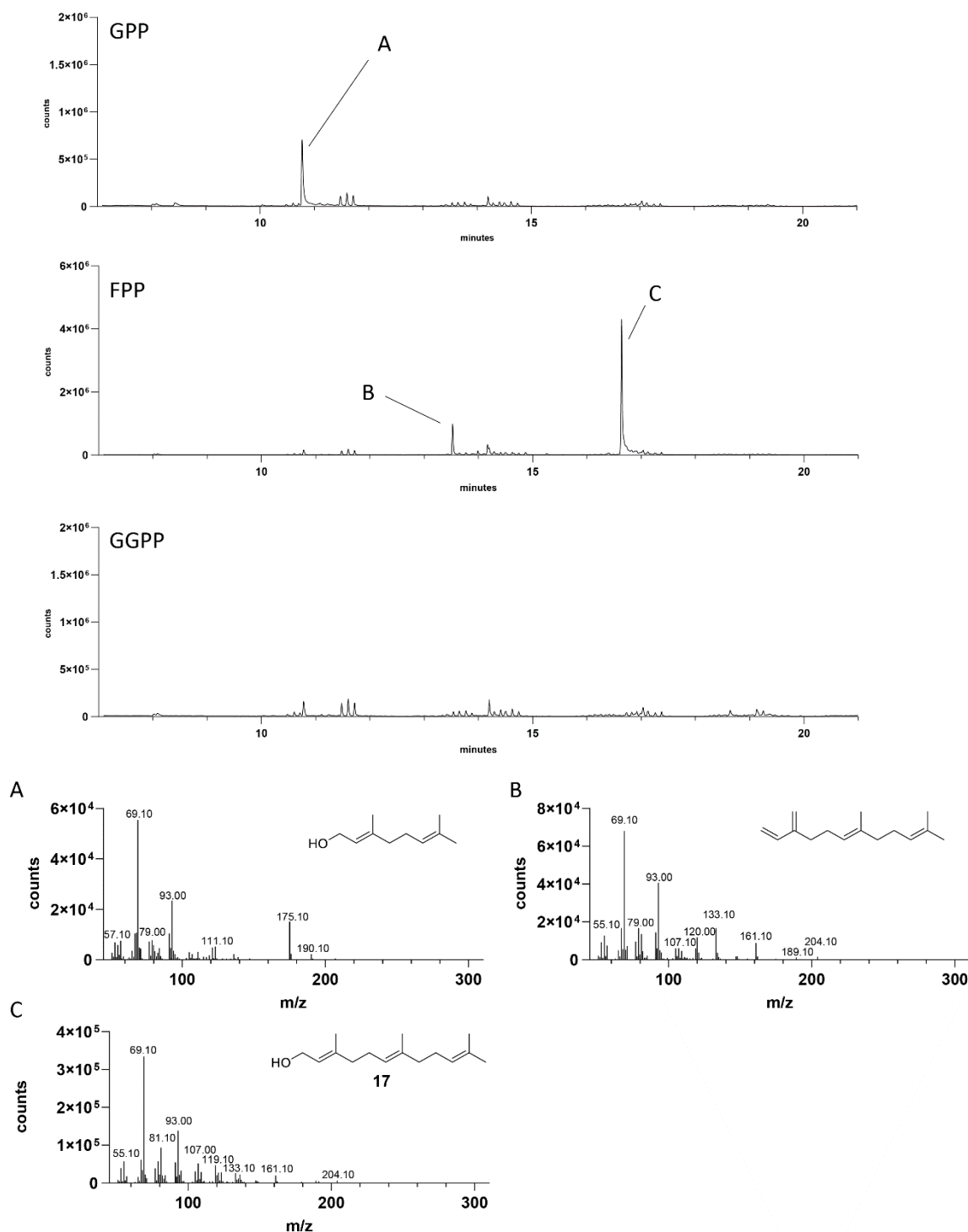
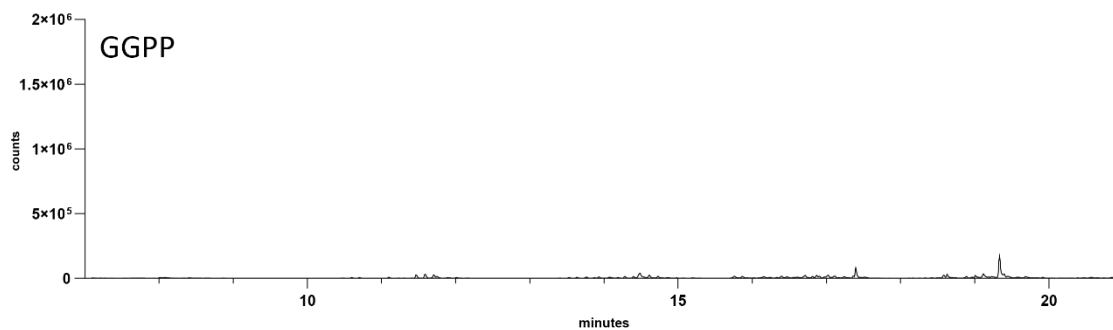
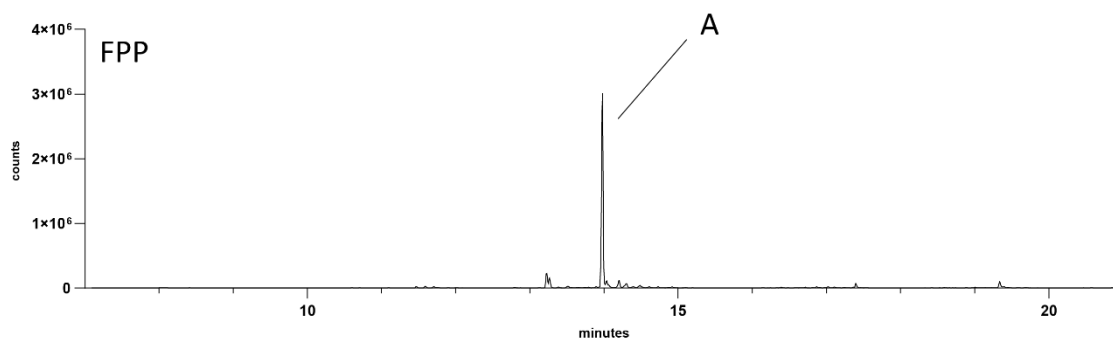
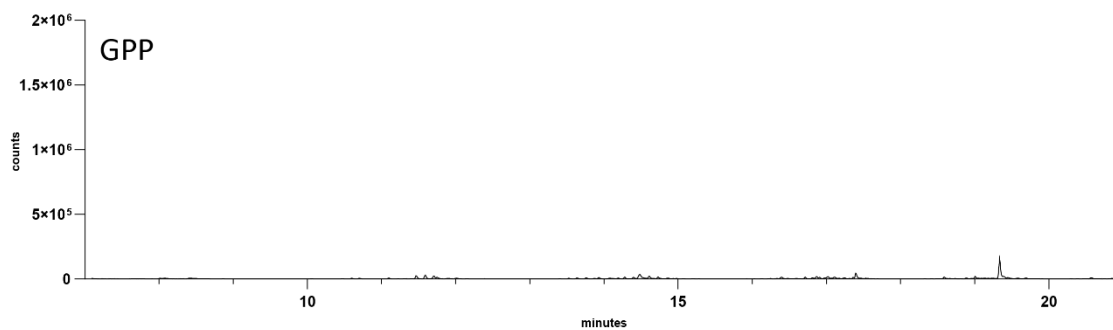


Figure 2-S 25: GCMS total ion chromatograms of SCB-084-TC-5 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. **17** and the other shown structures are best hits from the NIST database.

SCB-84-TC-6



A

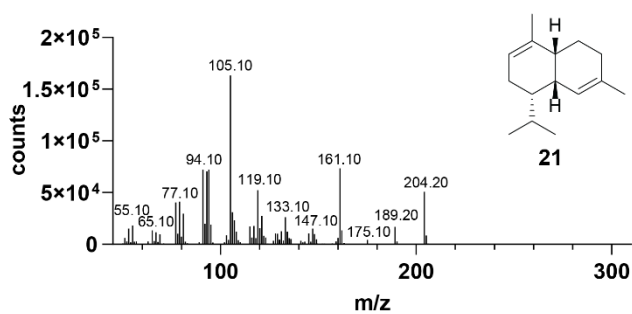


Figure 2-S 26: GCMS total ion chromatograms of SCB-084-TC-6 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The structure of **21** was verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

SCB-144-TC-1

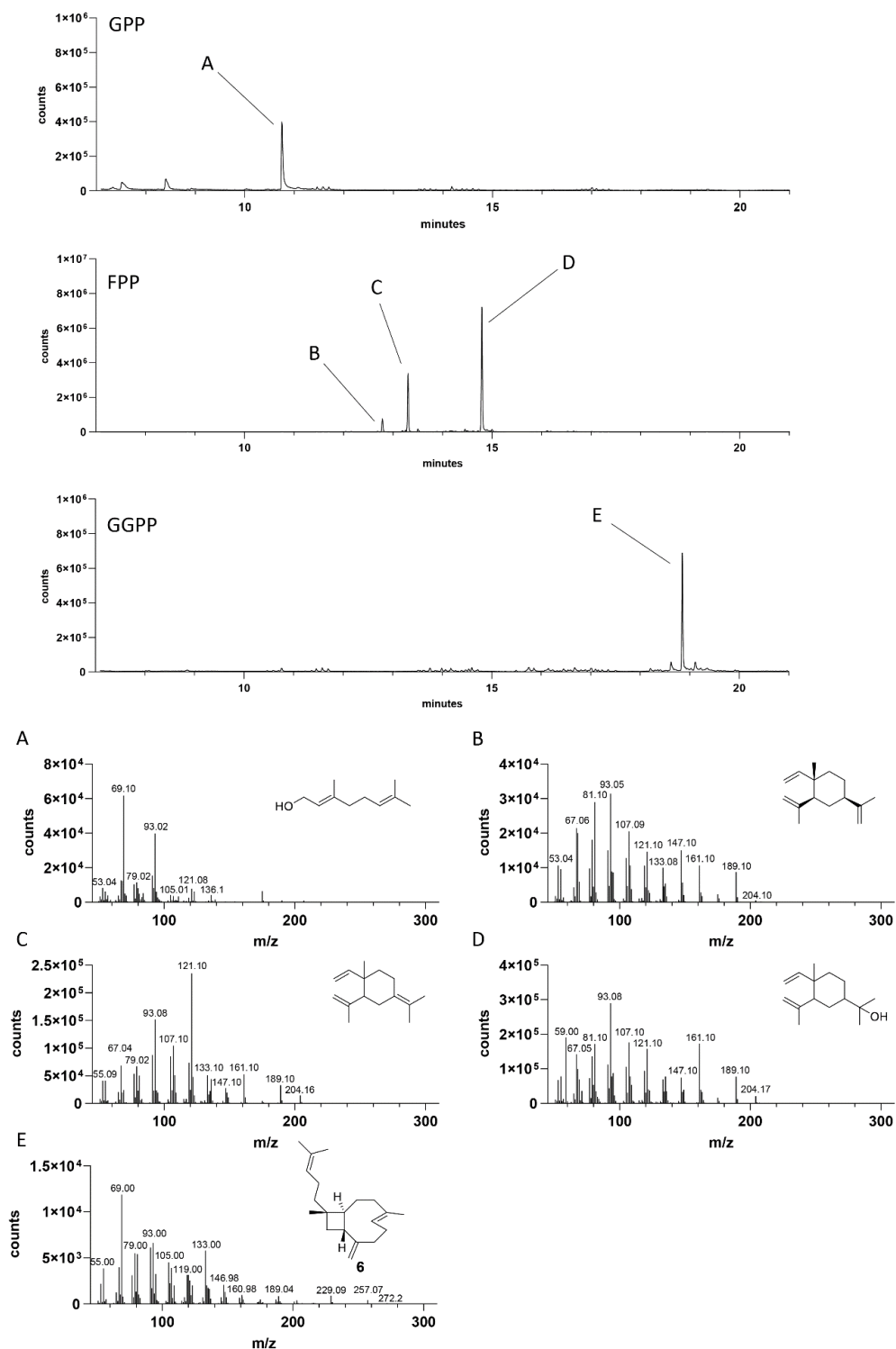


Figure 2-S 27: GCMS total ion chromatograms of SCB-144-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. **6** was verified using an authentic standard, the other shown structures are best hits from the NIST database.

SCB-144-TC-3

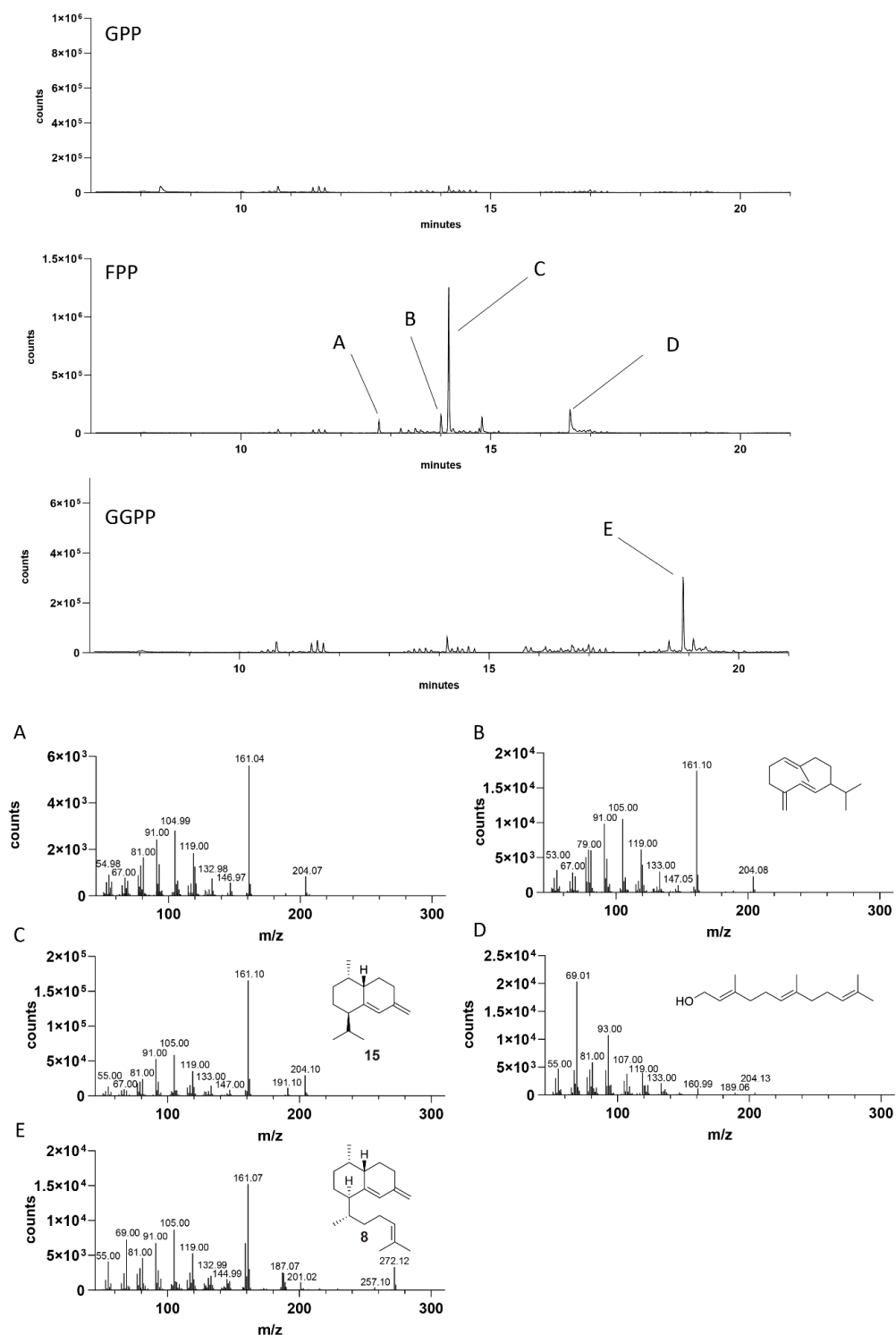


Figure 2-S 28: GCMS total ion chromatograms of SCB-144-TC-3 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identity of **8** and **15** has been verified using authentic standards.

SCB-144-TC-4

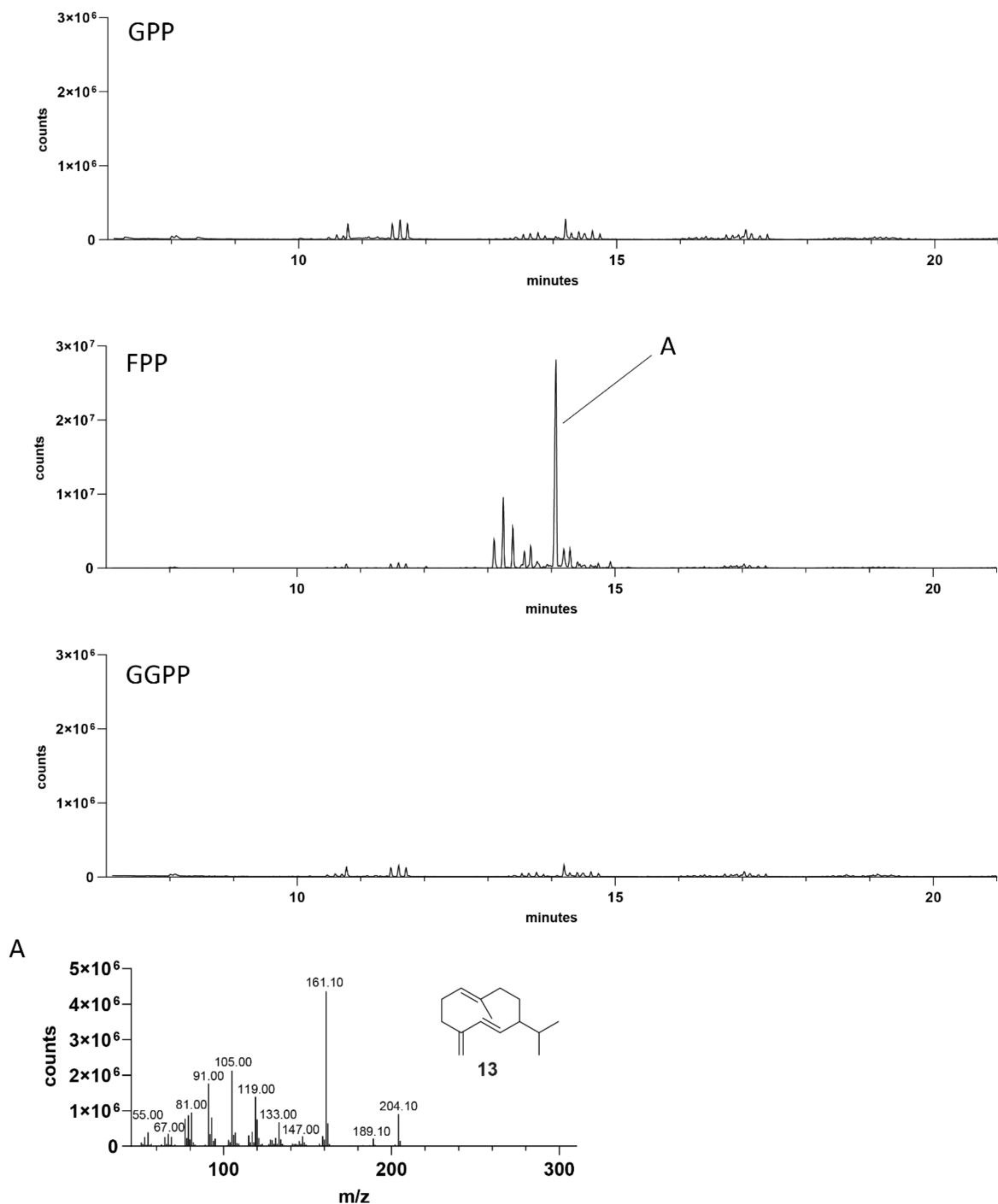


Figure 2-S 29: GCMS total ion chromatograms of SCB-144-TC-4 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The structure of **13** has been verified using an authentic standard.

SCB-144-TC-7

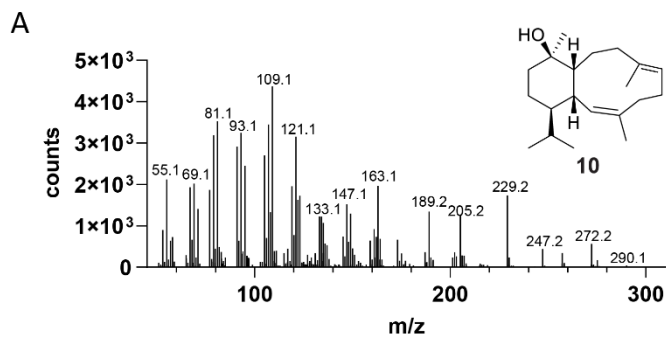
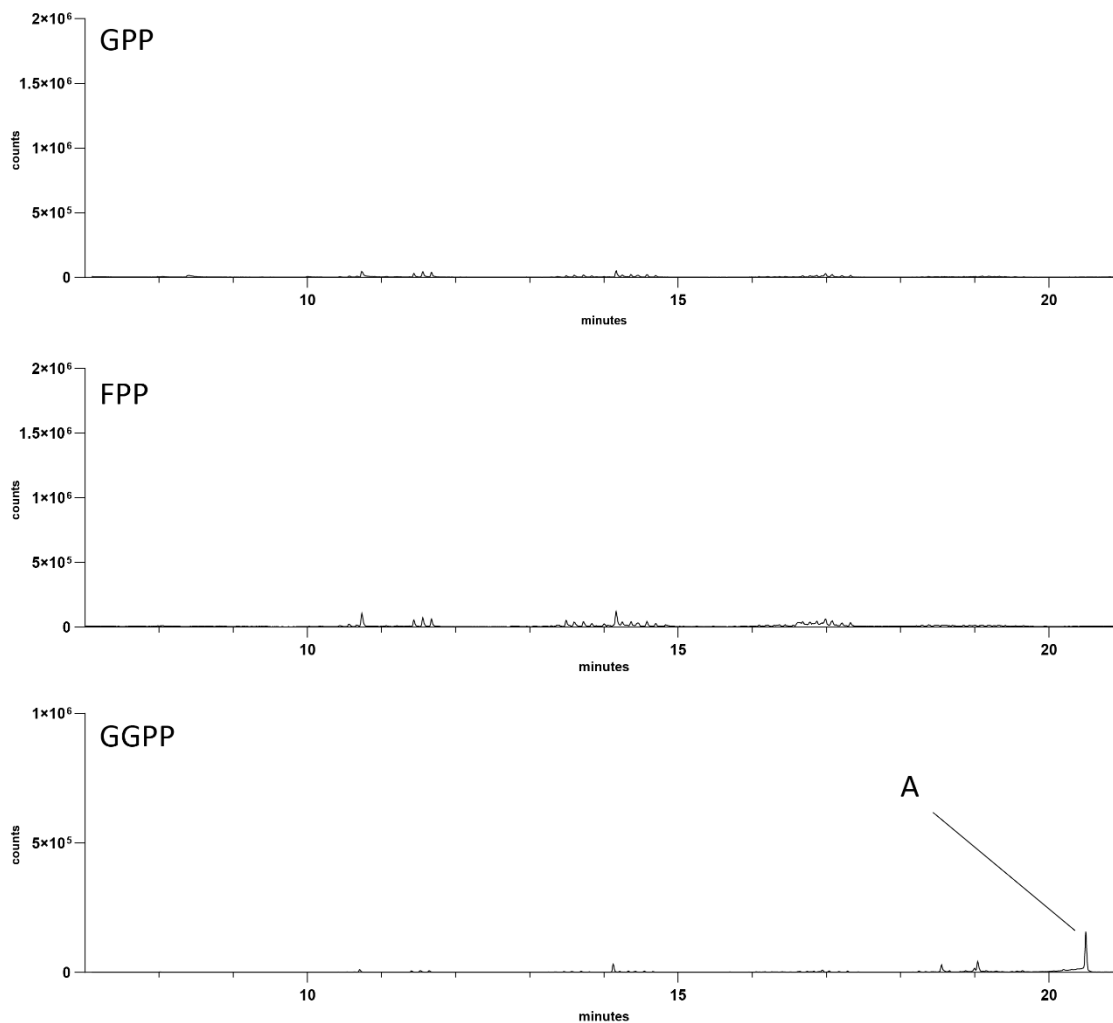


Figure 2-S 30: GCMS total ion chromatograms of SCB-144-TC-7 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identity of **10** was verified using an authentic standard.

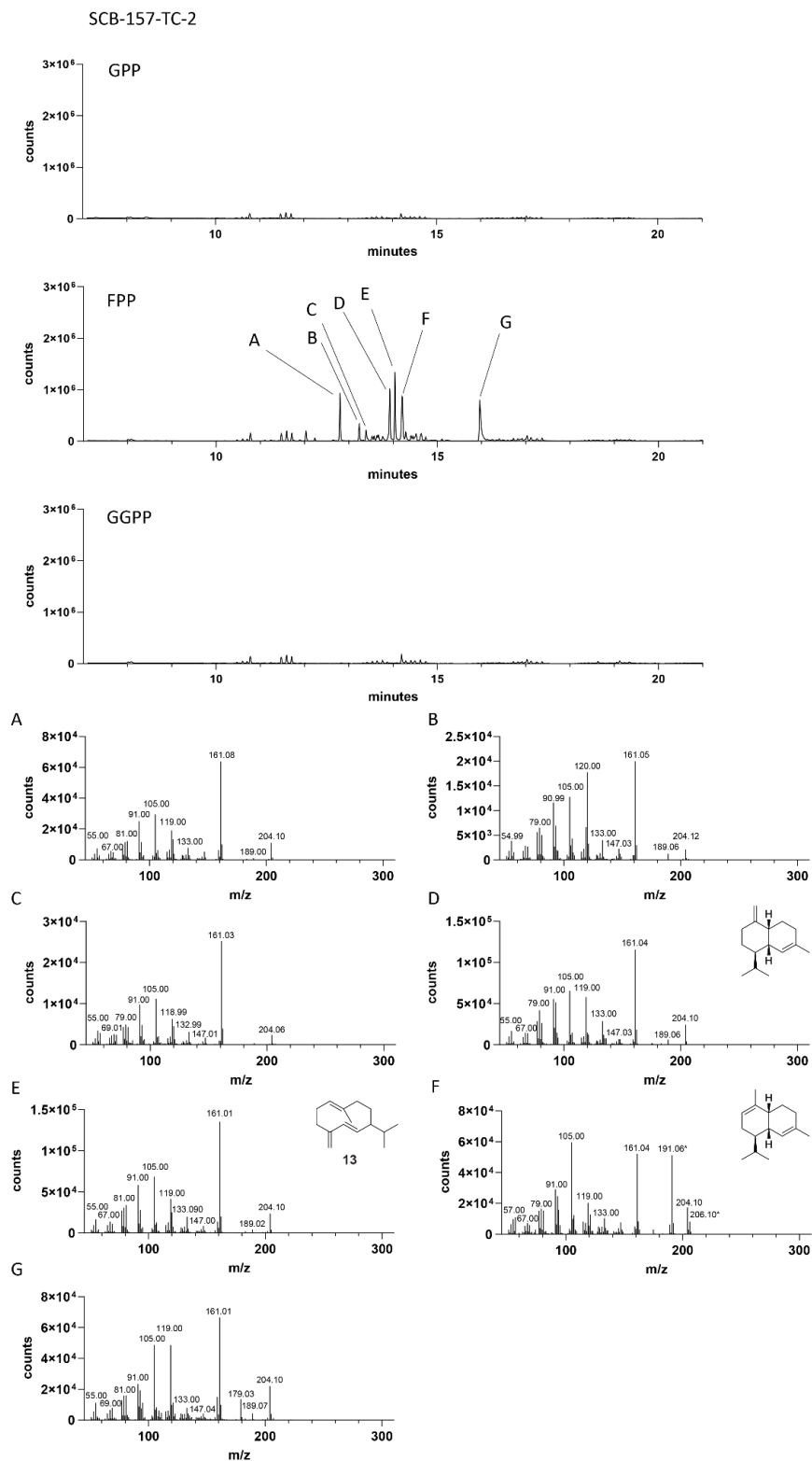


Figure 2-S 31: GCMS total ion chromatograms of SCB-157-TC-2 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identity of **13** was proven using an authentic standard, the other shown structures are best hits from the NIST database.

SCB-157-TC-5

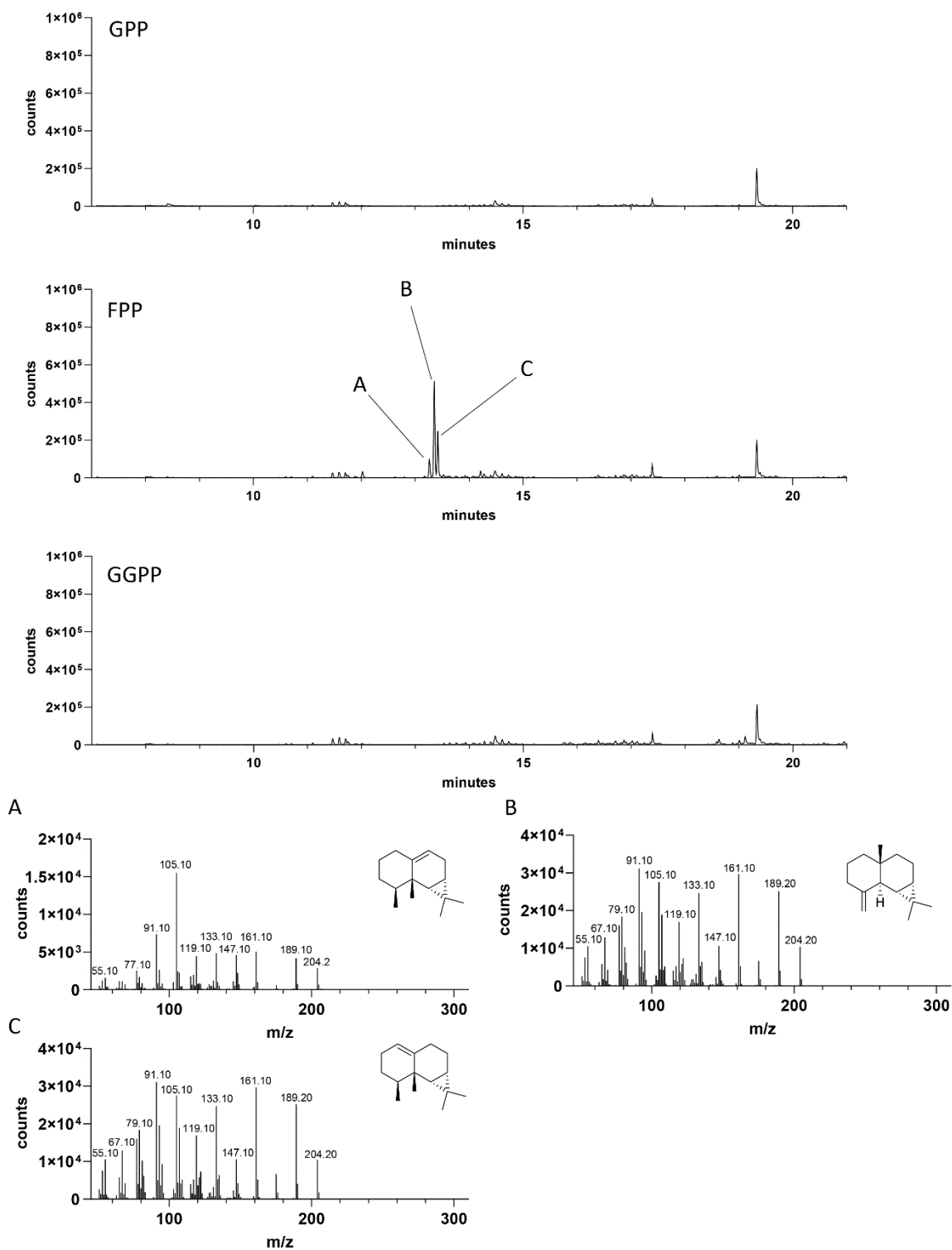


Figure 2-S 32: GCMS total ion chromatograms of SCB-157-TC-5 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The shown structures are best hits from the NIST database.

SCB-218-TC-1

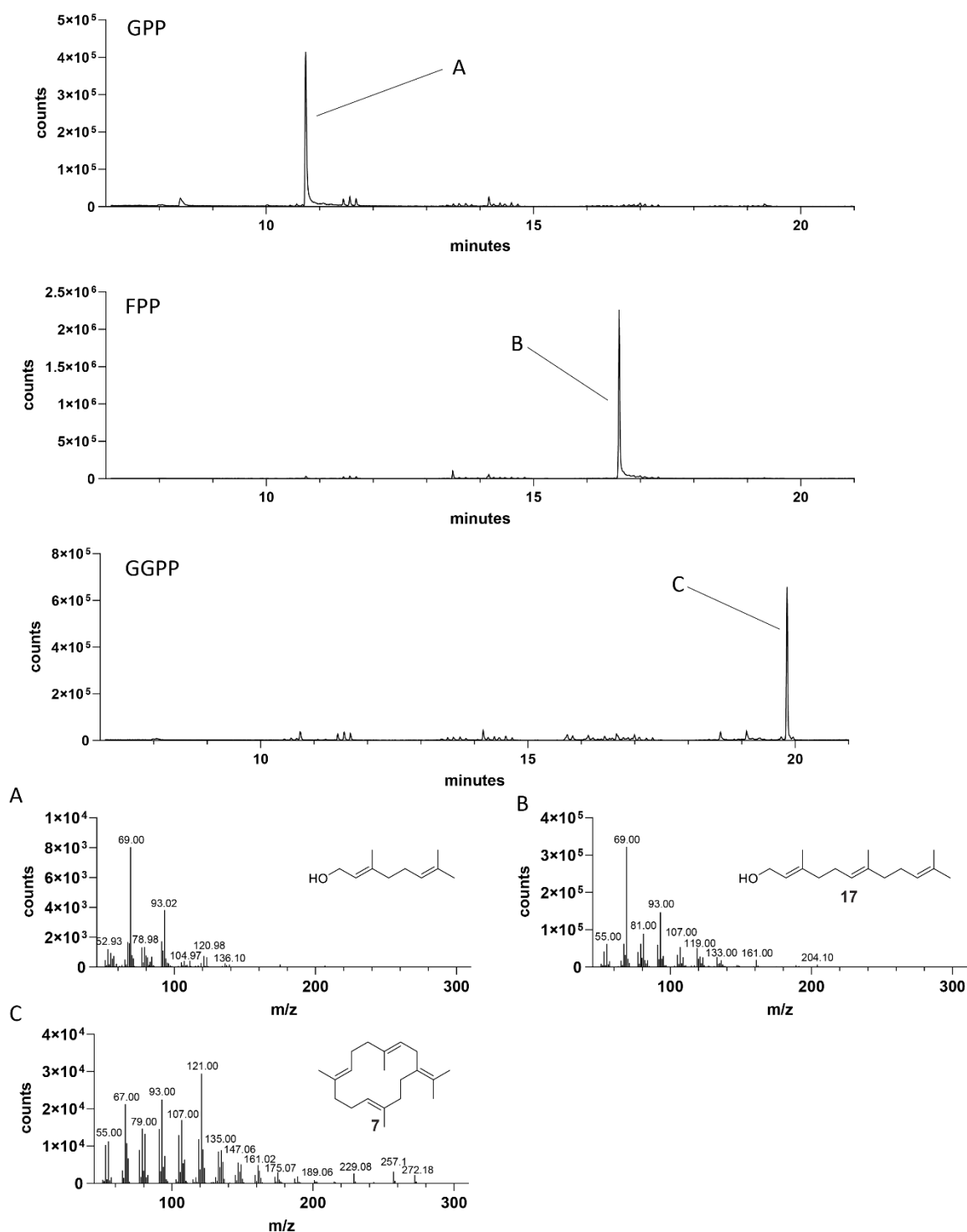


Figure 2-S 33: GCMS total ion chromatograms of SCB-218-TC-1 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identity of **7** was proven using an authentic standard, the other shown structures are best hits from the NIST database.

SCB-218-TC-2

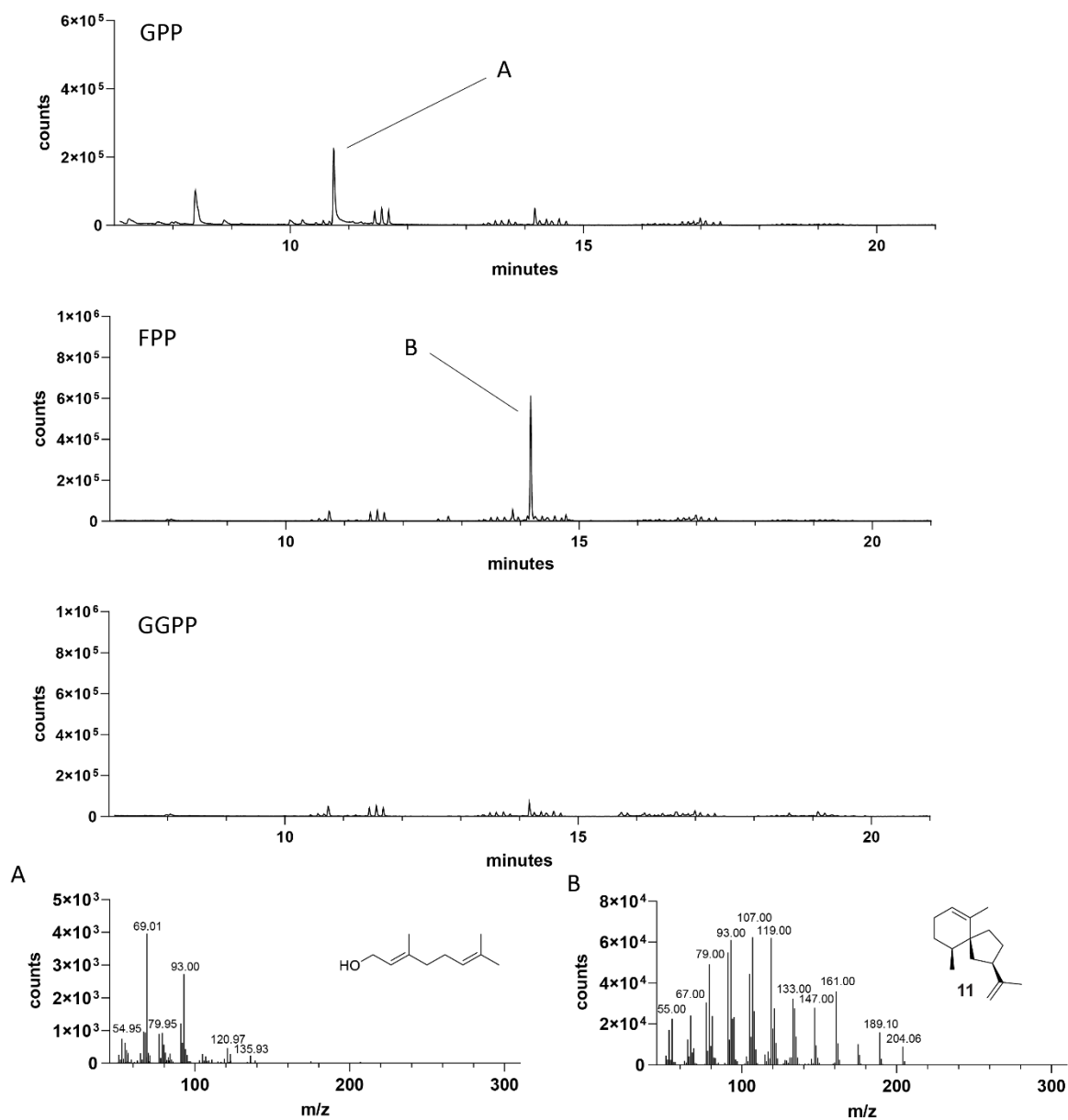


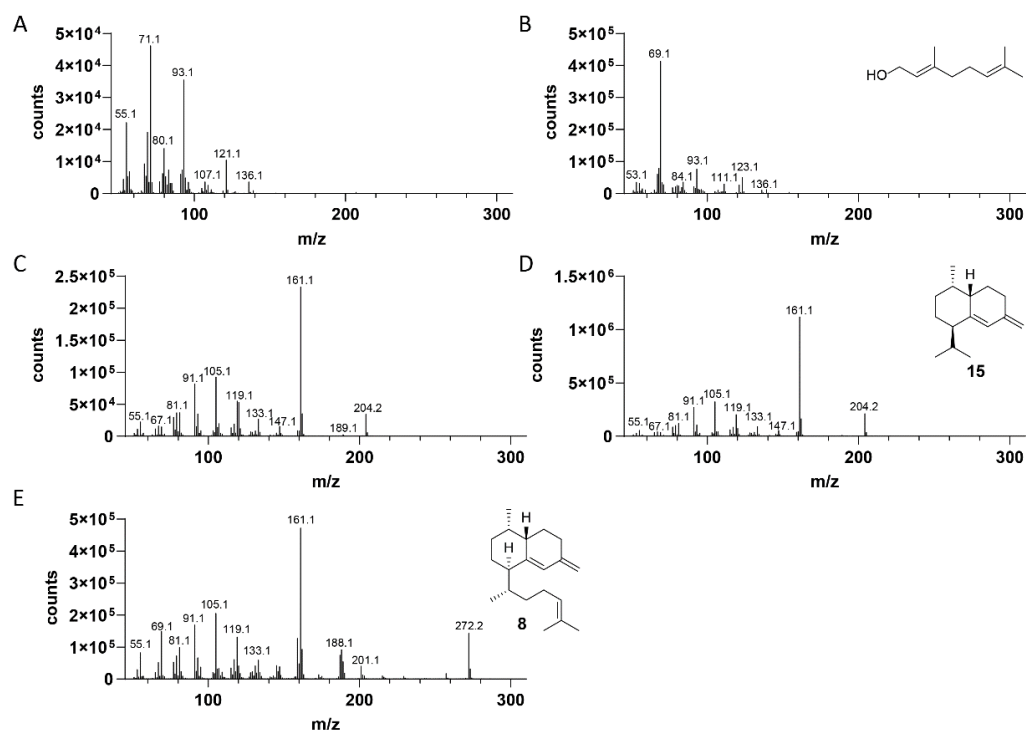
Figure 2-S 34: GCMS total ion chromatograms of SCB-218-TC-2 in vitro incubations with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The structure of **11** was elucidated by NMR spectroscopy of purified material from a large-scale in vitro assay.

The figure displays three stacked chromatograms showing the separation of GPP, FPP, and GGPP over a 20-minute period. The y-axis for all plots is 'counts'.

- GPP Chromatogram:** The y-axis ranges from 0 to 1.5×10^7 . Two peaks are labeled: 'A' at approximately 8 minutes and 'B' at approximately 11 minutes.
- FPP Chromatogram:** The y-axis ranges from 0 to 1.5×10^7 . Two peaks are labeled: 'C' at approximately 13 minutes and 'D' at approximately 14 minutes.
- GGPP Chromatogram:** The y-axis ranges from 0 to 1×10^7 . One peak is labeled: 'E' at approximately 18 minutes.

The x-axis for all plots is 'minutes', ranging from 0 to 20.



Mass spectra of enzymatically produced terpenes are shown. The identity of **8** and **15** was proven using an authentic standard, the other shown structures are best hits from the NIST database.

Ancestor clade 1-II

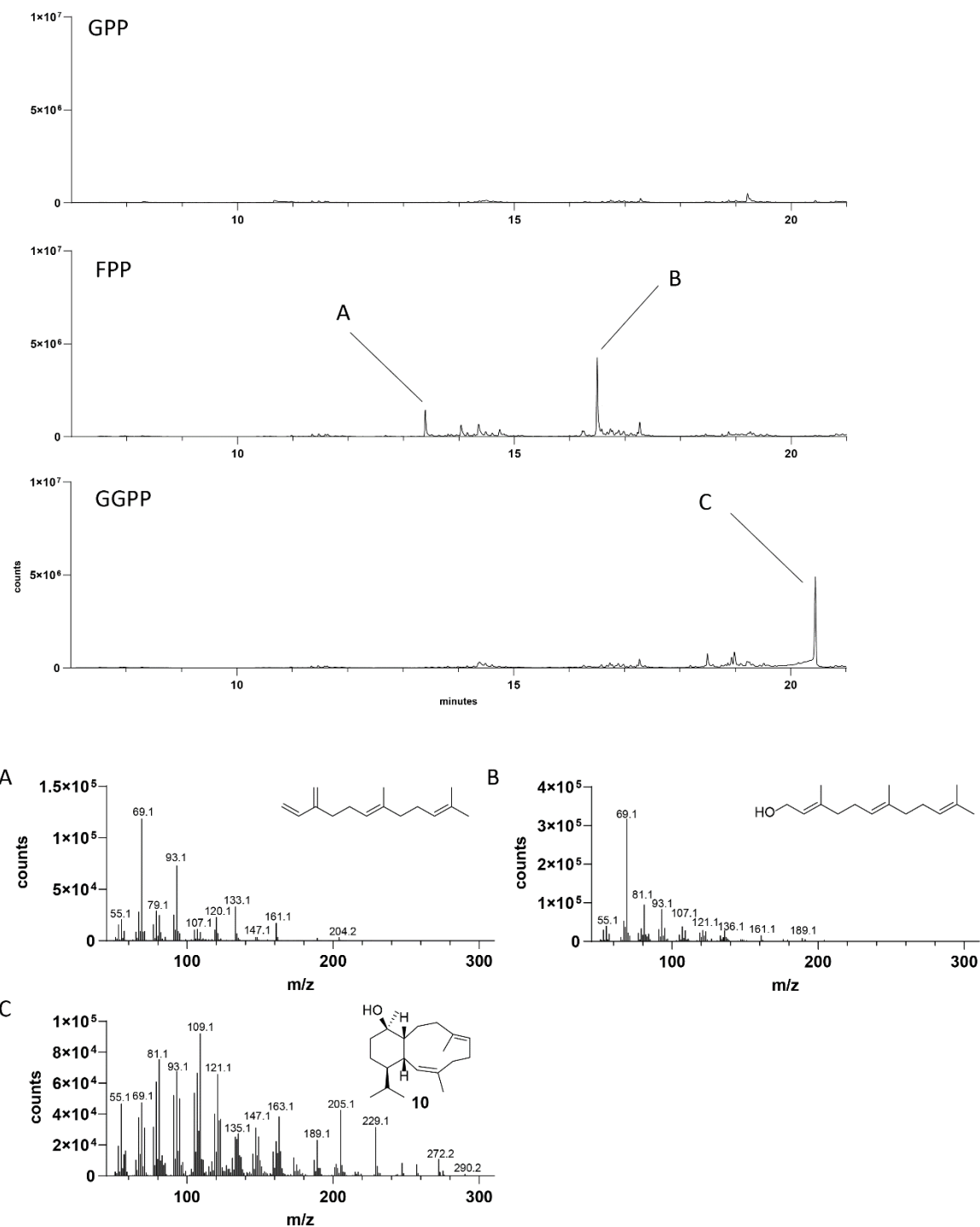


Figure 2-S 36: GCMS total ion chromatograms of in vitro incubations of the clade 1-II ancestor with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identity of **10** was proven using an authentic standard, the other shown structures are best hits from the NIST database.

Ancestor clade 1-III

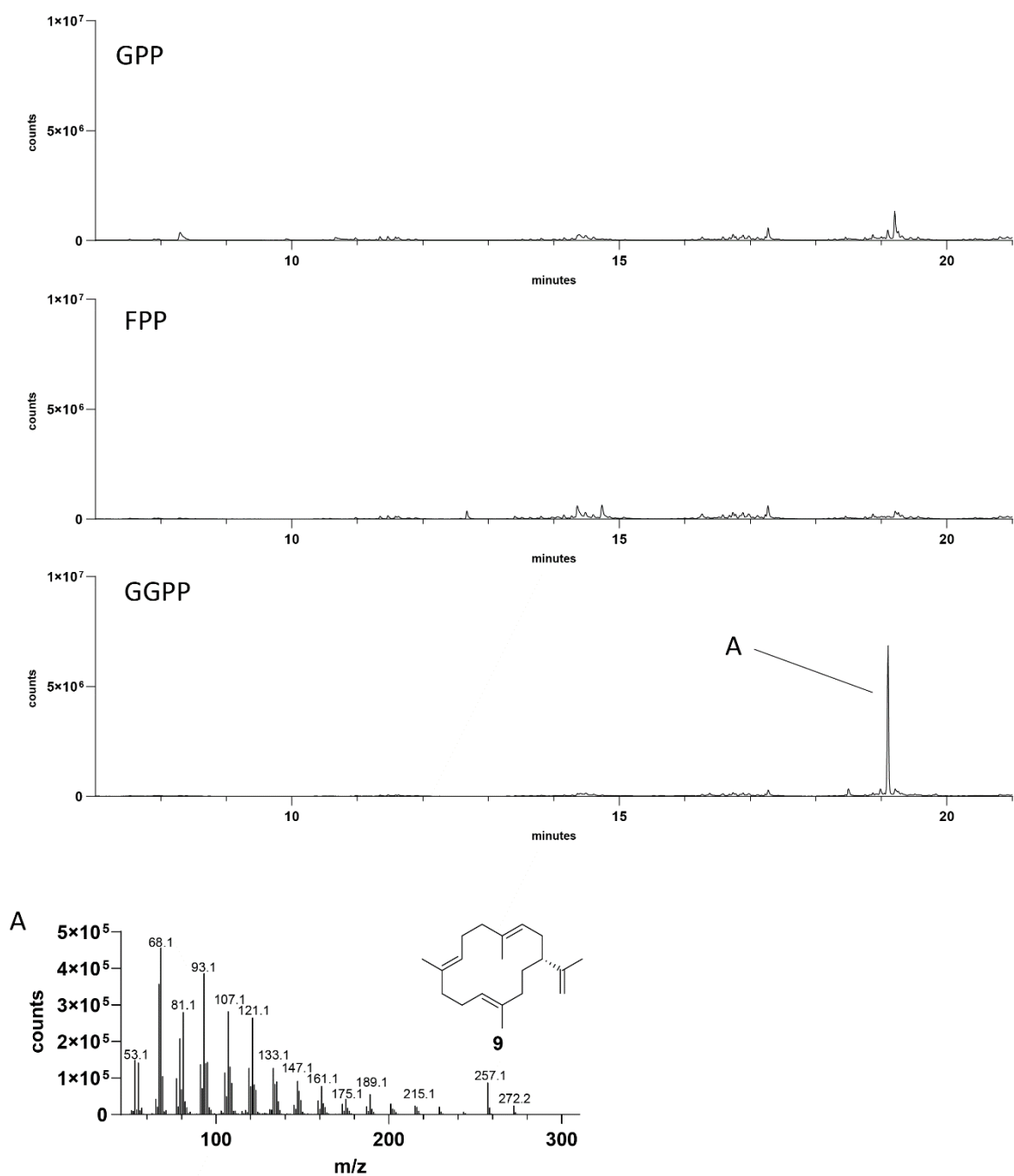


Figure 2-S 37: GCMS total ion chromatograms of in vitro incubations of the clade 1-III ancestor with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identity of **9** was proven using an authentic standard.

Ancestor clade 1-IV

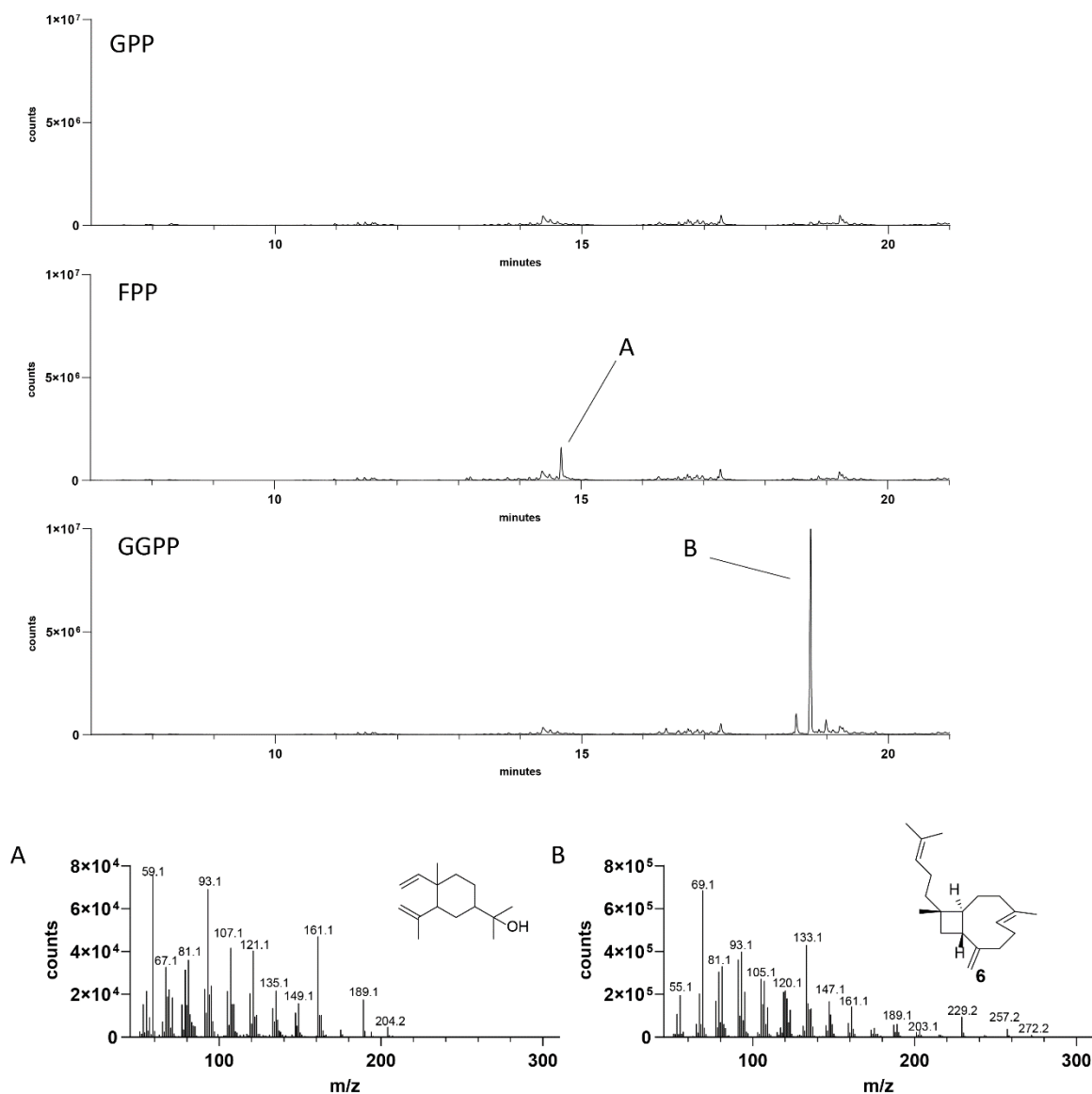


Figure 2-S 38: GCMS total ion chromatograms of in vitro incubations of the clade 1-IV ancestor with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identity of **6** was proven using an authentic standard, the other shown structures are best hits from the NIST database.

Ancestor clade 1-V

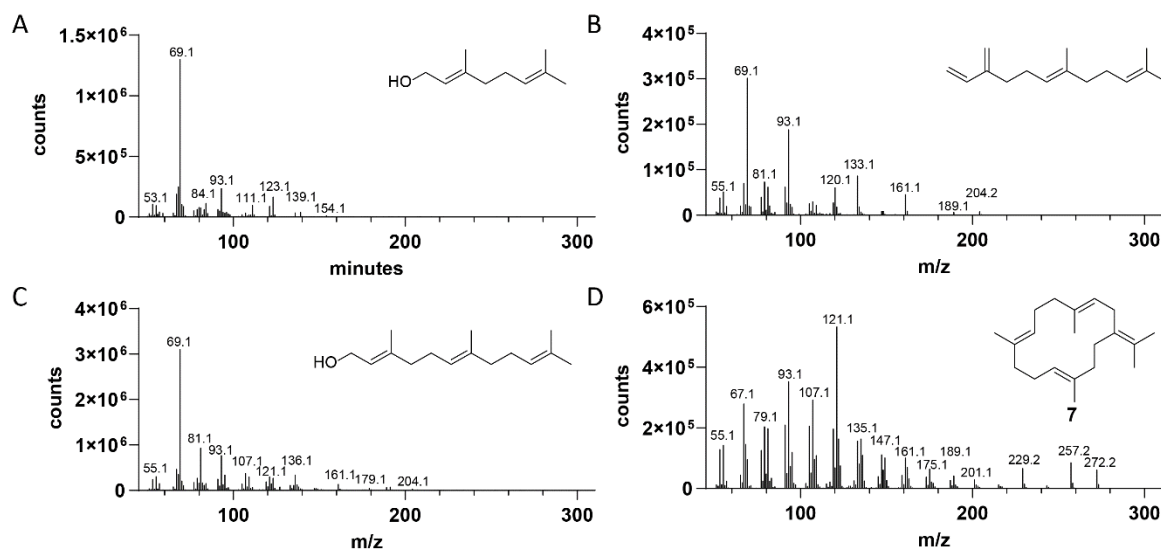
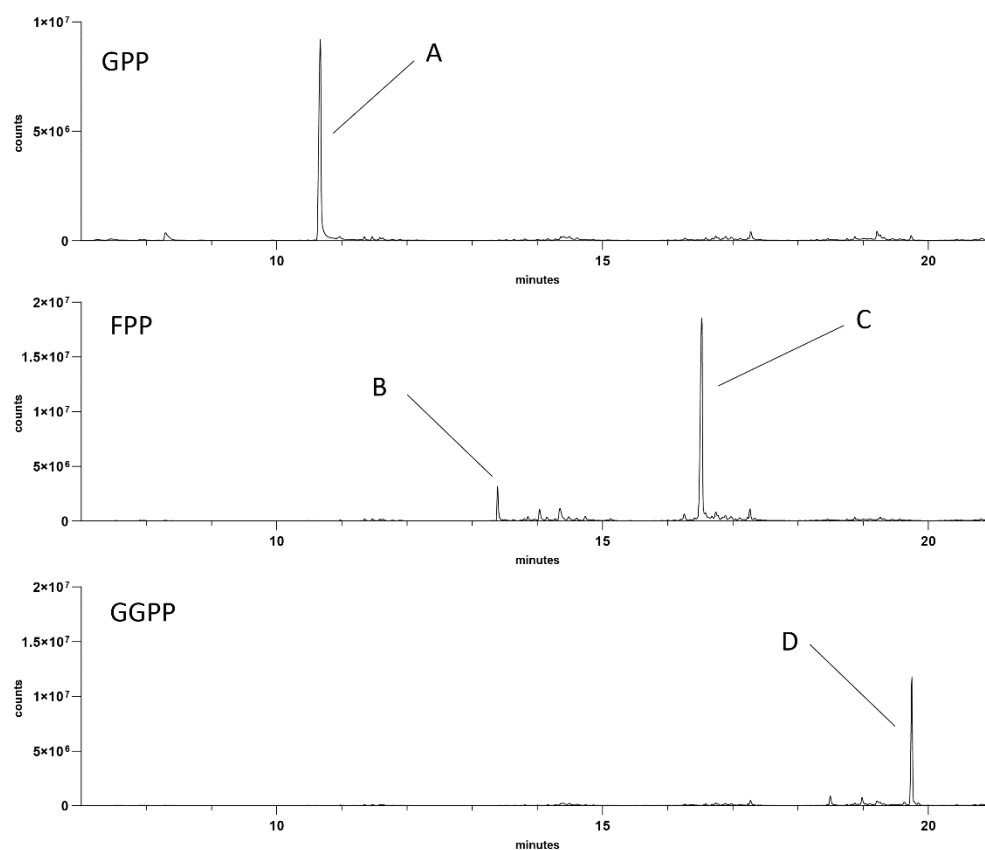


Figure 2-S 39: GCMS total ion chromatograms of in vitro incubations of the clade 1-V ancestor with GPP (top), FPP (middle), and GGPP (bottom).

Mass spectra of enzymatically produced terpenes are shown. The identity of **7** was proven using an authentic standard, the other shown structures are best hits from the NIST database.

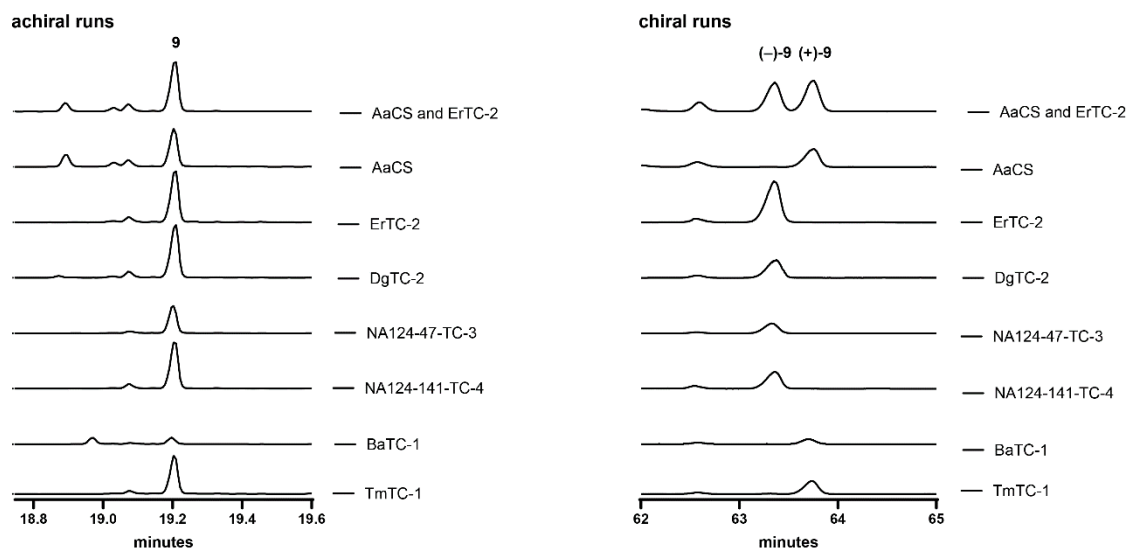


Figure 2-S 40: Determination of the absolute configuration of enzymatically obtained 9.

Left: Runs on achiral stationary phase; right: the same samples analyzed on a homochiral stationary phase. The absolute configuration of cembrene A produced by ErTC-2 and AaCS was elucidated previously as (-)-**9** and (+)-**9**, respectively.(14, 29)

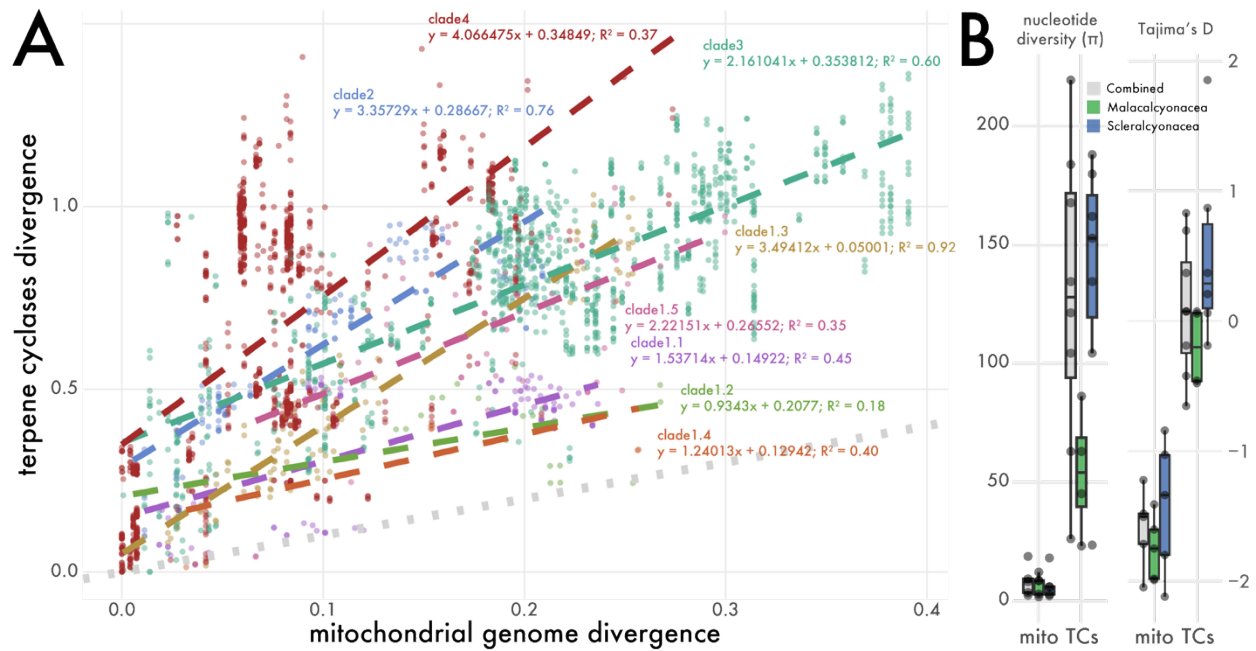


Figure 2-S 41: Evolutionary divergence and population genetic patterns of octocoral terpene cyclase (TC) genes.

A) Pairwise nucleotide divergence of TC genes relative to mitochondrial genomes across octocoral taxa. Points are colored by TC clade (clades 1–4), with clade 1 subdivided into functionally characterized subclades (1-I to 1-V). Dashed lines show linear regressions for each clade; slope and adjusted R^2 values indicate relative evolutionary rates. The grey dashed line (slope = 1.0) reflects the expected divergence based on mitochondrial phylogeny. **B)** Nucleotide diversity (π) and Tajima's D for TC genes (clades 1–4 and subclades 1-I to 1-V) and mitochondrial genes, shown for all taxa and separated by order. Mitochondrial metrics provide a neutral baseline for comparison.

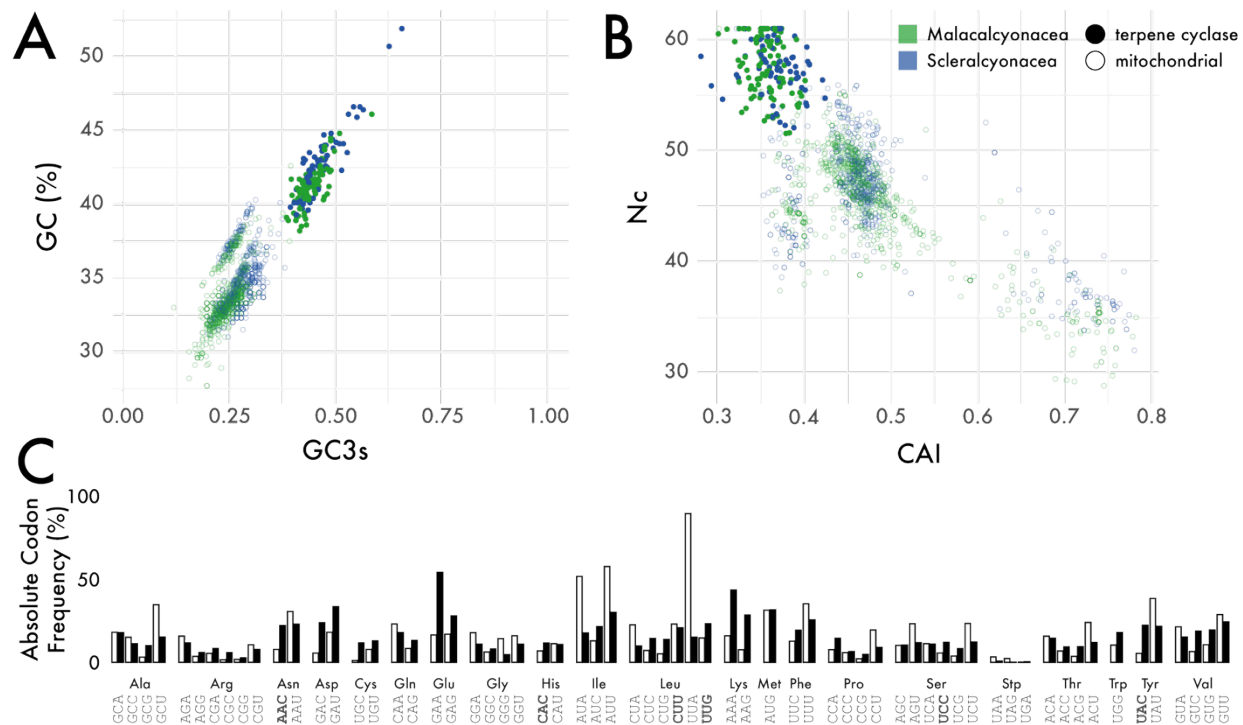
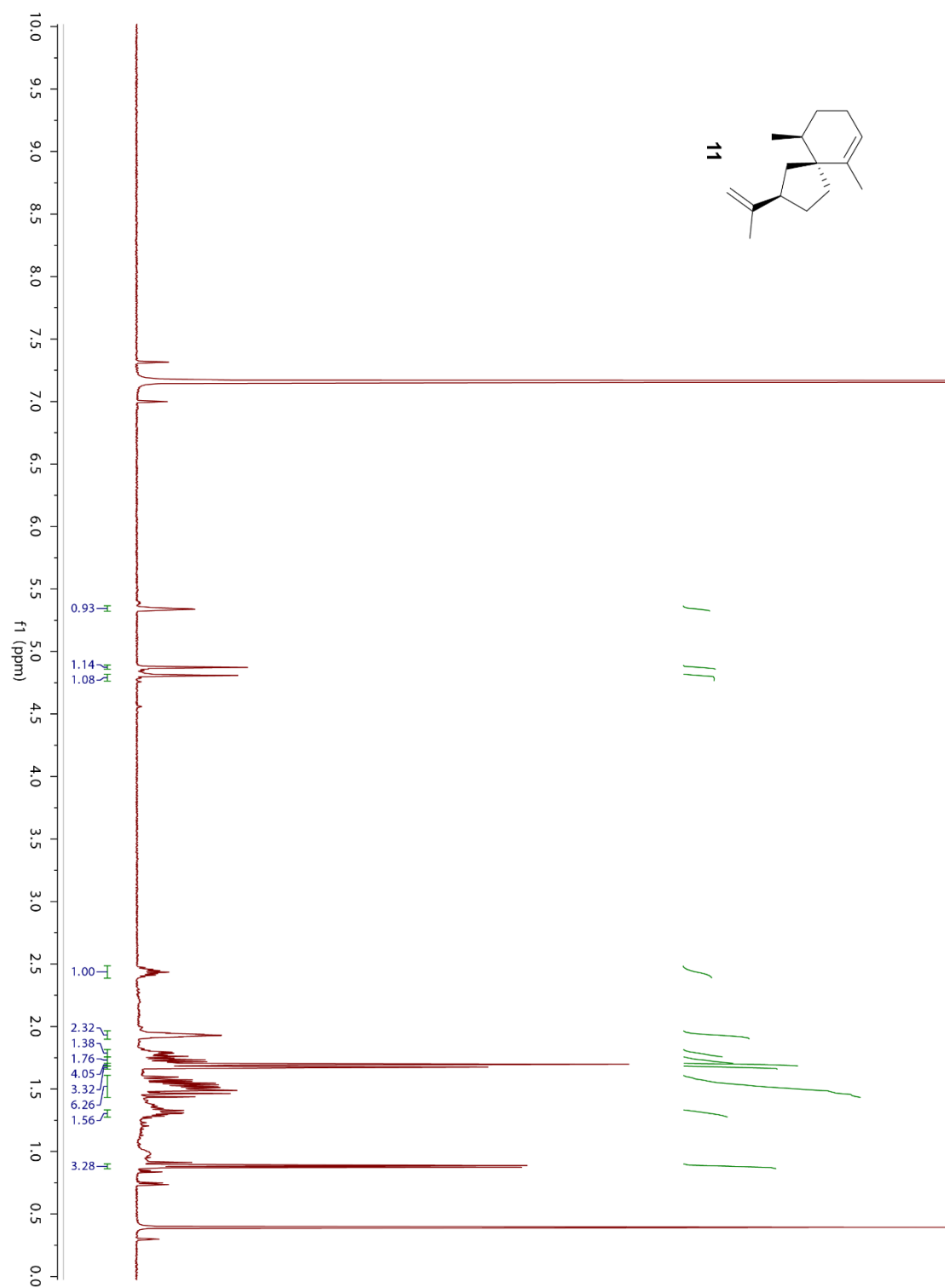


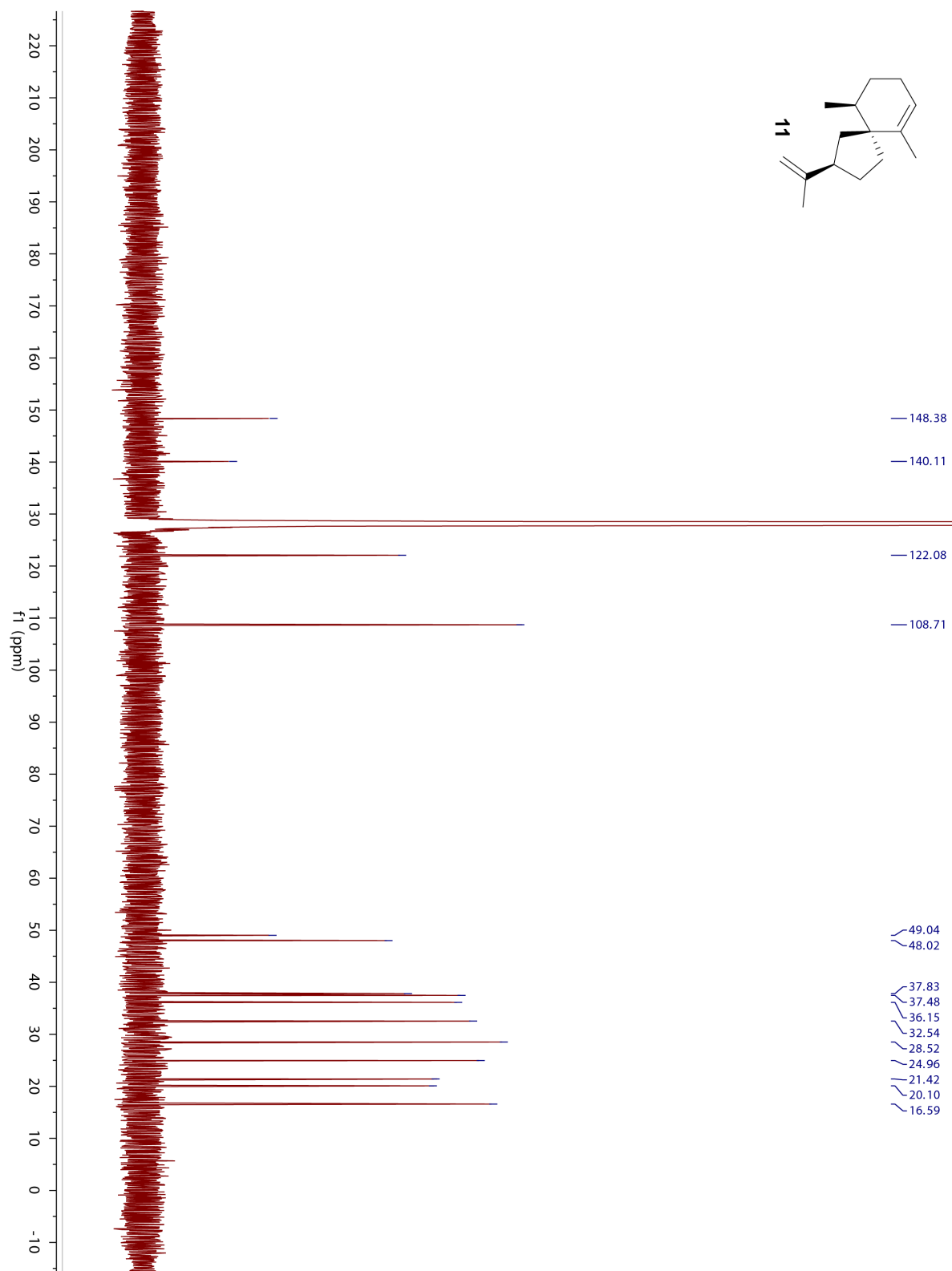
Figure 2-S 42: Codon composition and codon bias metrics in octocoral terpene cyclase (TC) genes.

A) Neutrality plot showing the relationship between third-position GC content (GC_3) and overall GC content. Points are colored by taxonomic order (Malacalcyonacea in green, Scleralcyonacea in blue). TC genes exhibit a strong positive linear relationship between GC_3 and overall GC content, consistent with mutational bias influencing codon composition. However, these genes are systematically elevated in both overall GC and GC_3 relative to typical mitochondrial values, suggesting gene-specific differences in mutational pressures or base composition constraints. **B)** Relationship between effective codon number (N_c) and codon adaptation index (CAI). Lower N_c values correspond to stronger codon bias, while higher CAI values indicate increased codon usage bias. TC genes cluster in the high N_c , low CAI region, consistent with weak codon usage bias and relaxed synonymous site constraints. This pattern is expected for genes evolving under lineage-specific mutational biases or with codon usage patterns not strongly shaped by translational selection. **C)** Absolute codon frequencies for TC and mitochondrial genes, shown for all codons and grouped by encoded amino acid. Bars are colored by gene group (TC or mitochondrial). Mitochondrial genes serve as a comparison, highlighting differences in synonymous codon usage and overall codon composition between gene groups.

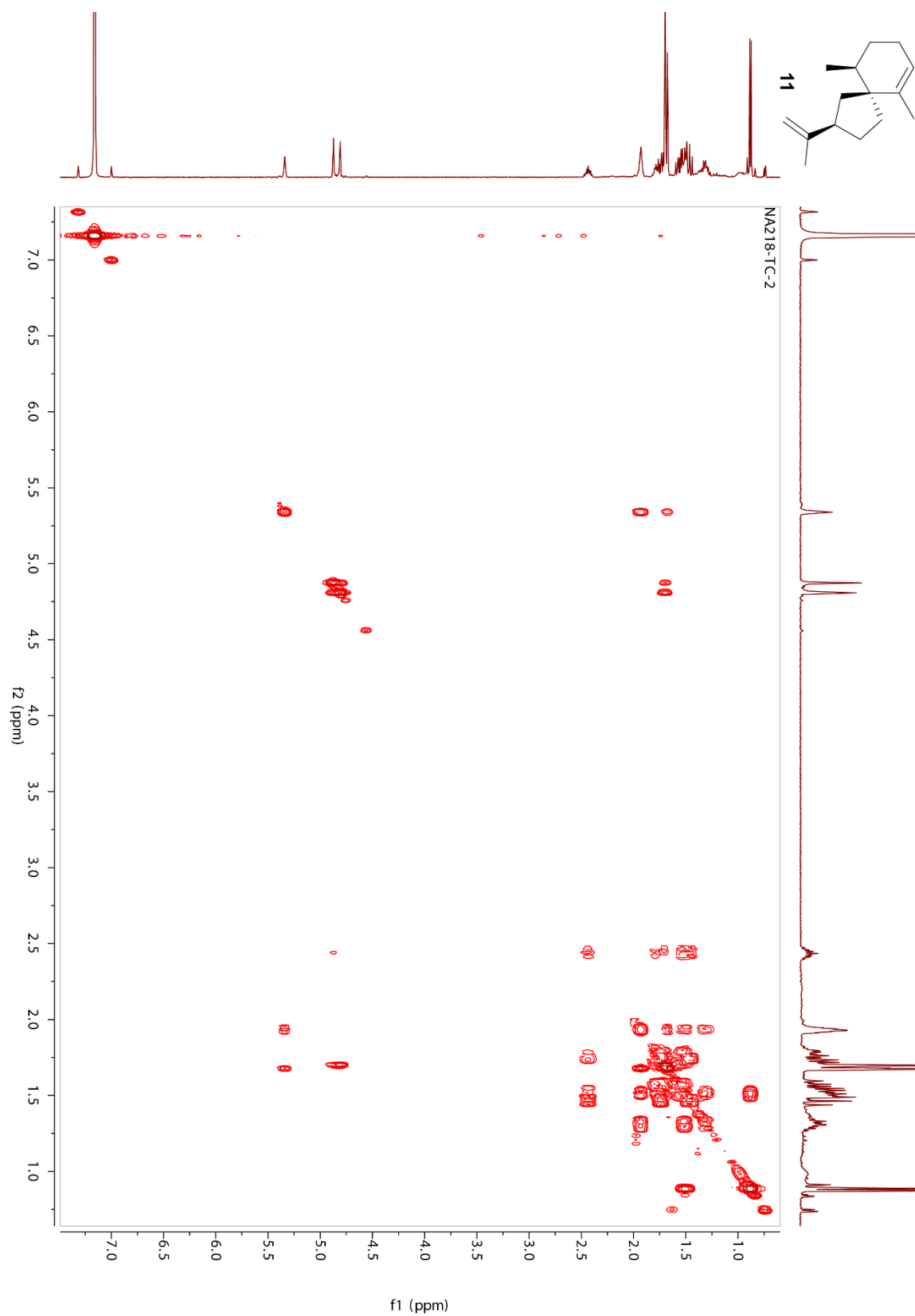
NMR spectra

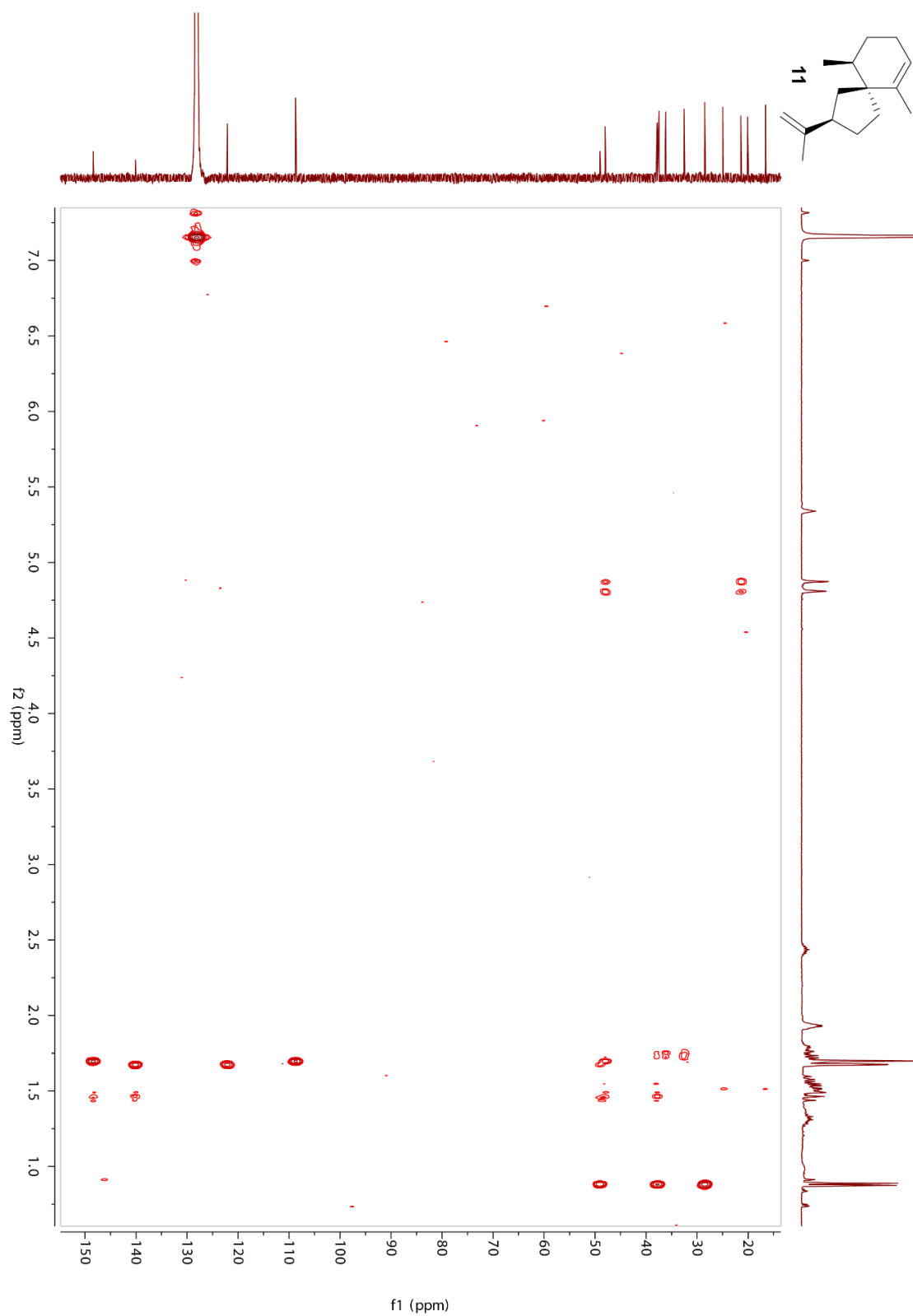


¹H-NMR spectrum (500 MHz) of 11 in C₆D₆.

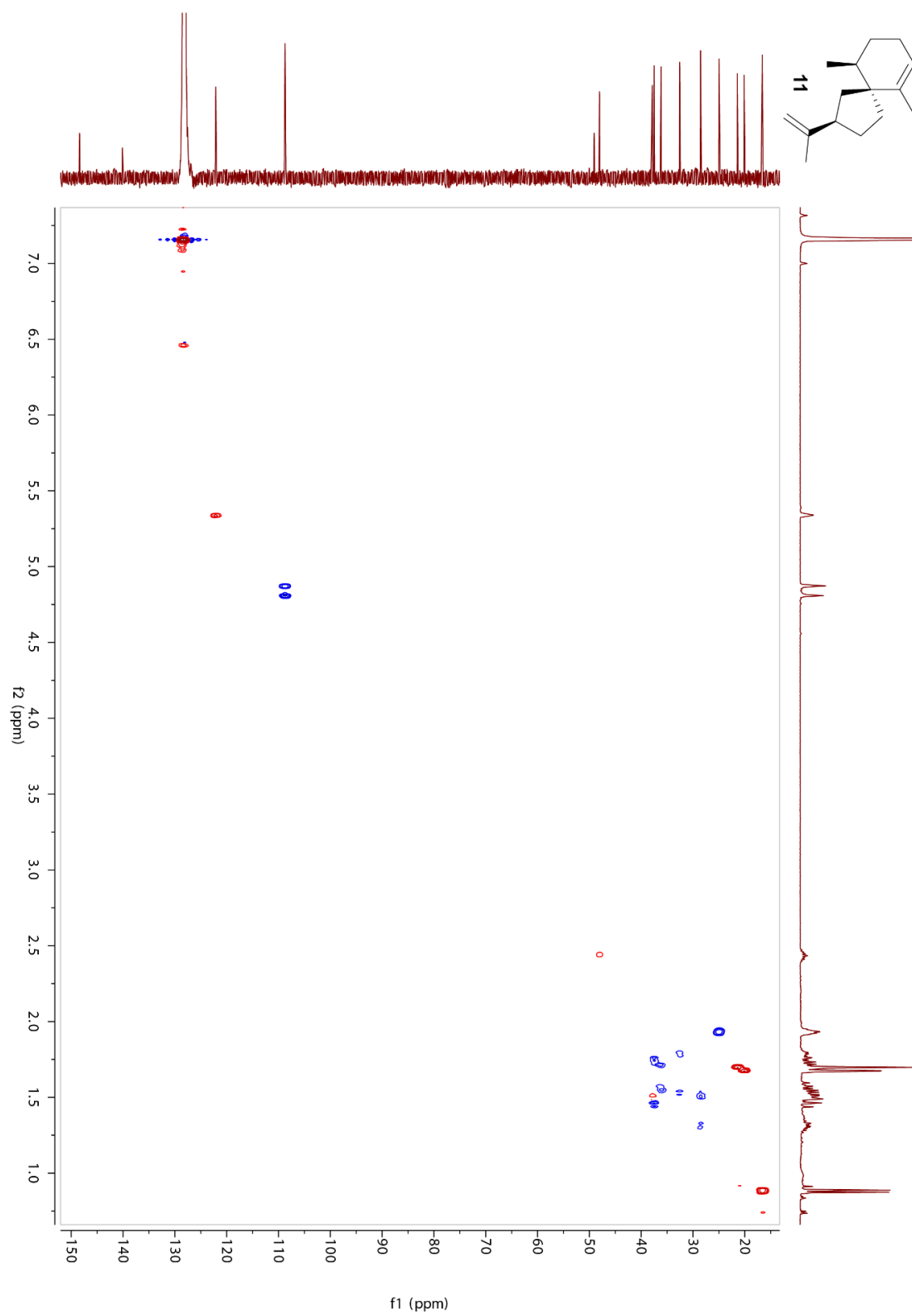


¹³C-NMR spectrum (125 MHz) of 11 in C₆D₆.

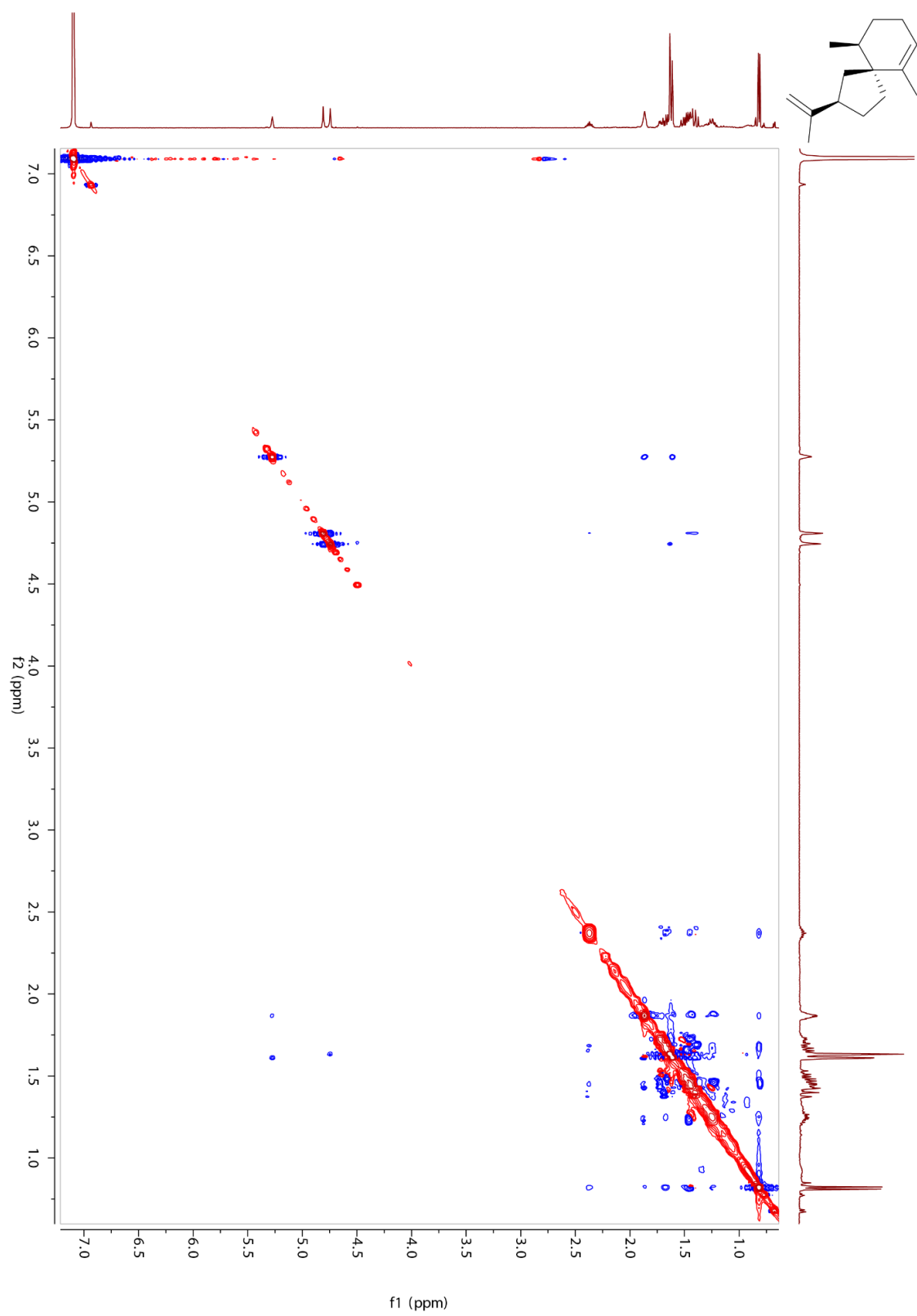




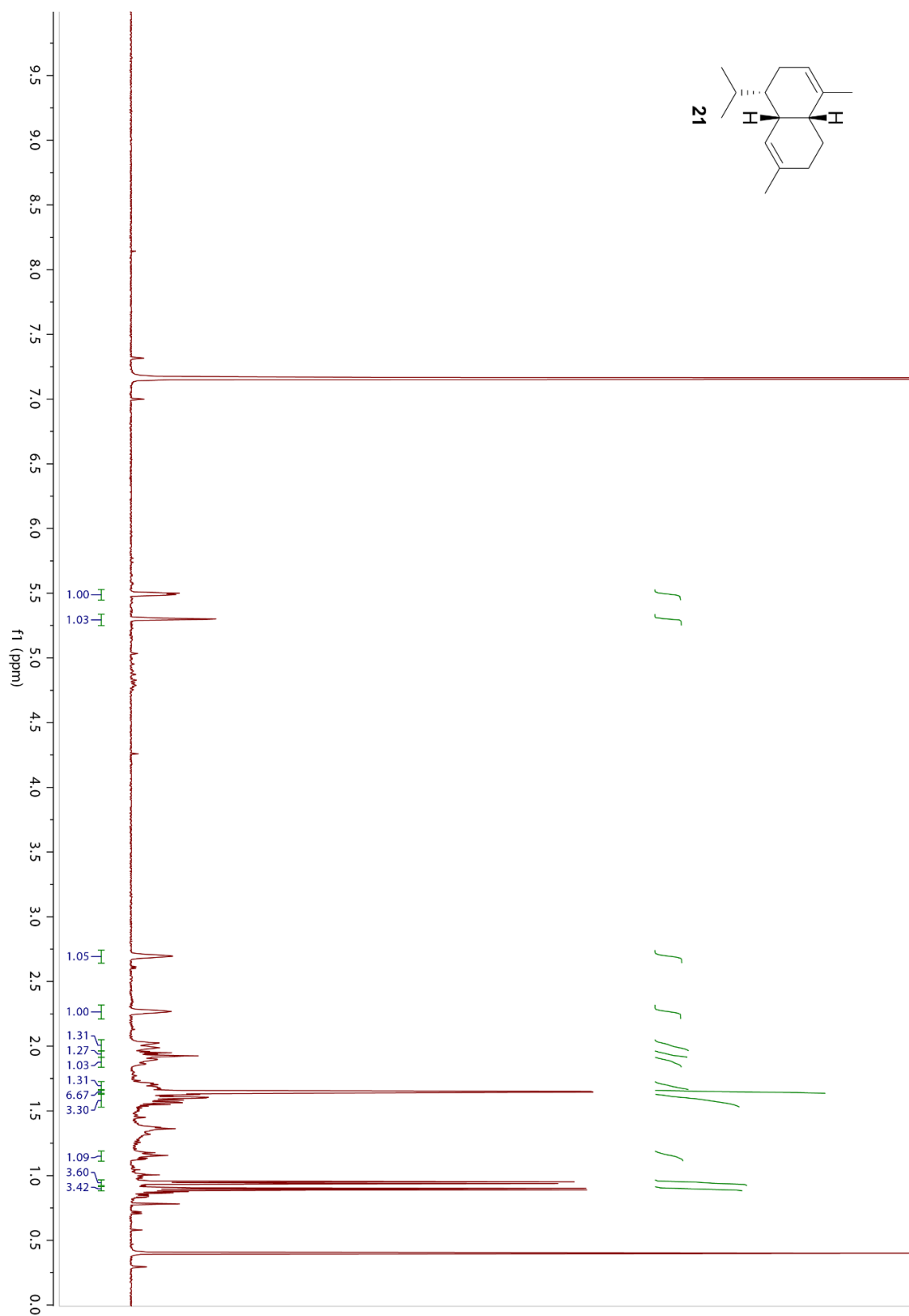
^1H , ^{13}C -HMBC spectrum (500 MHz) of 11 in C_6D_6 .



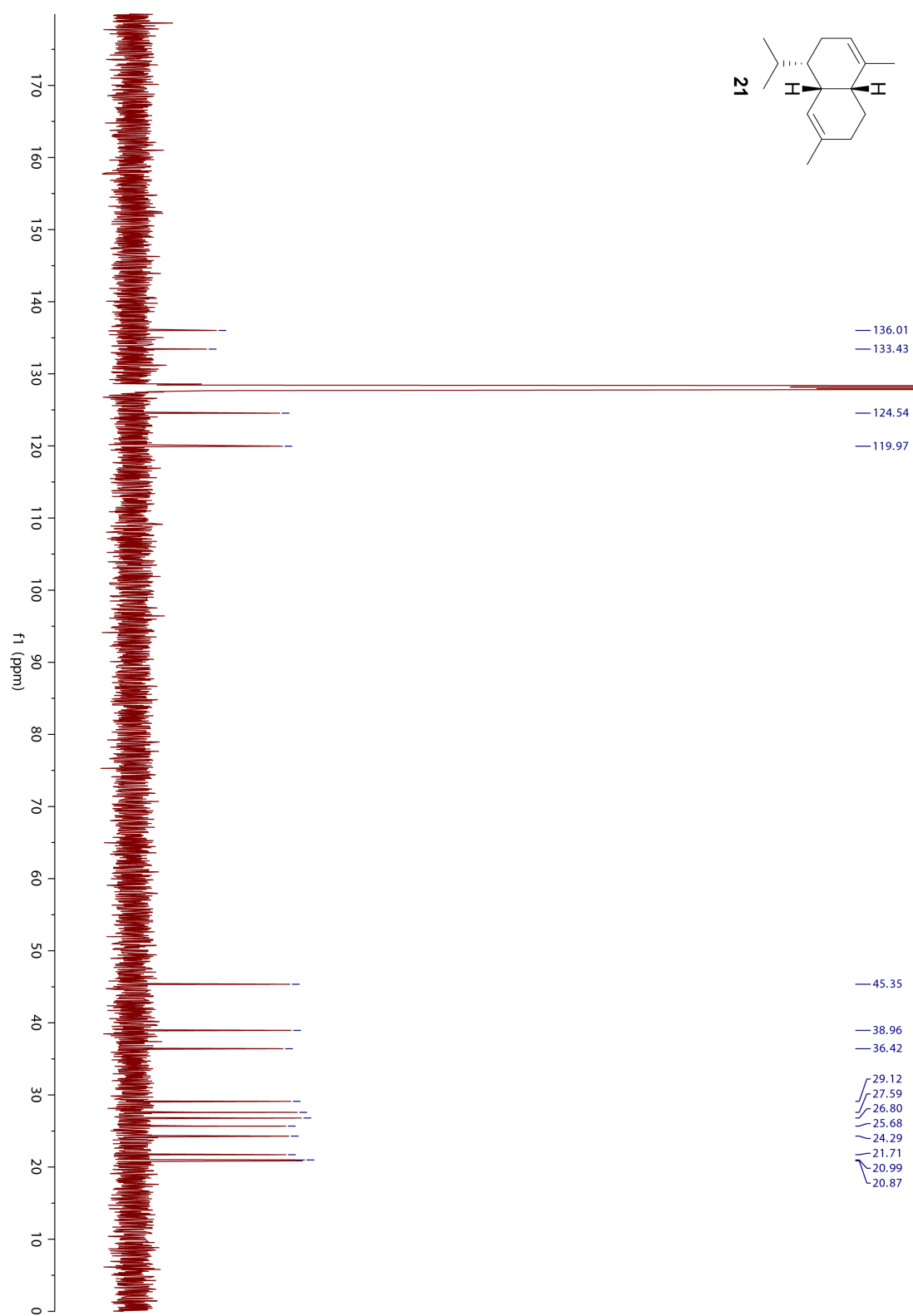
^1H , ^{13}C -HSQC spectrum (500 MHz) of 11 in C_6D_6 .

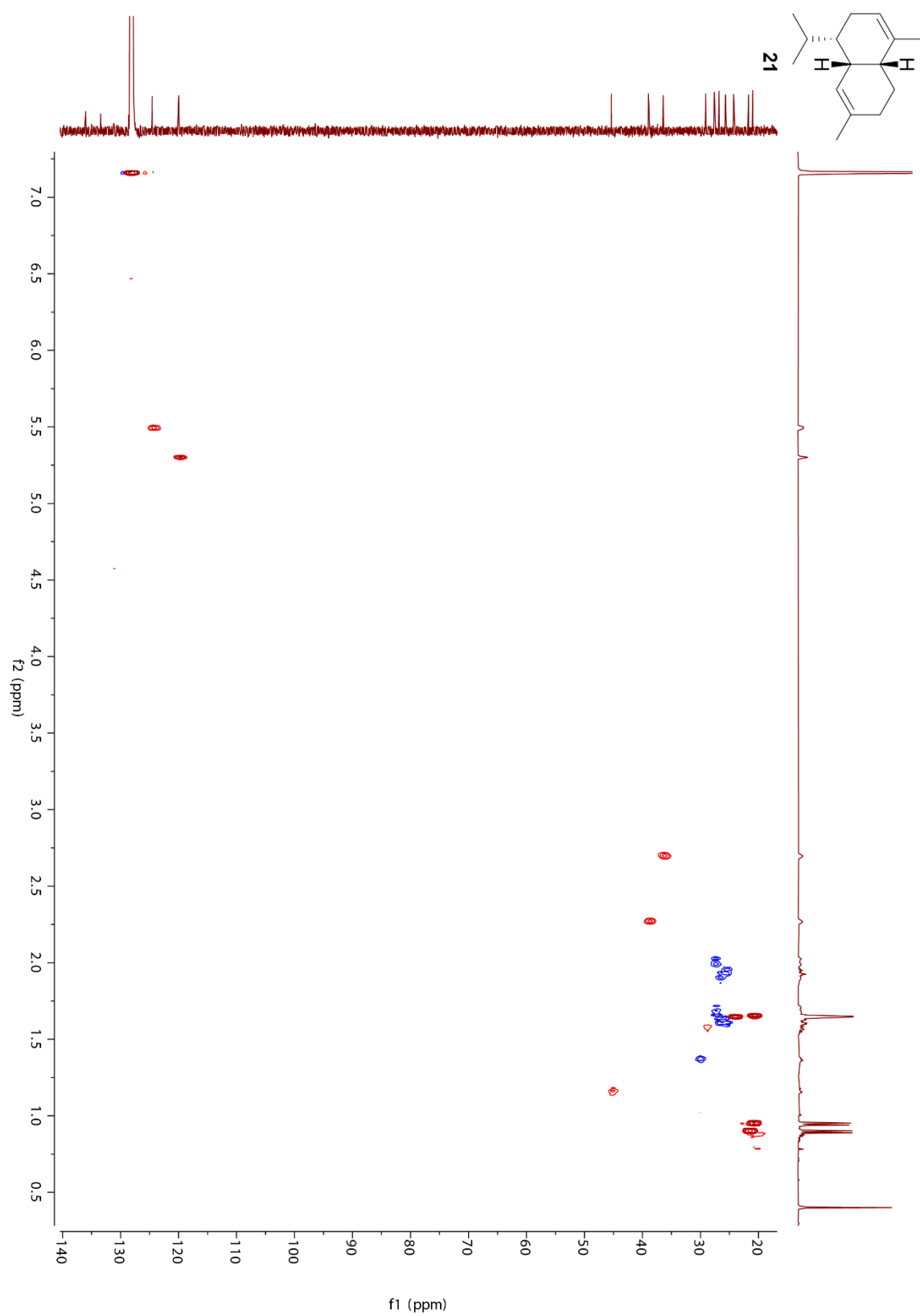


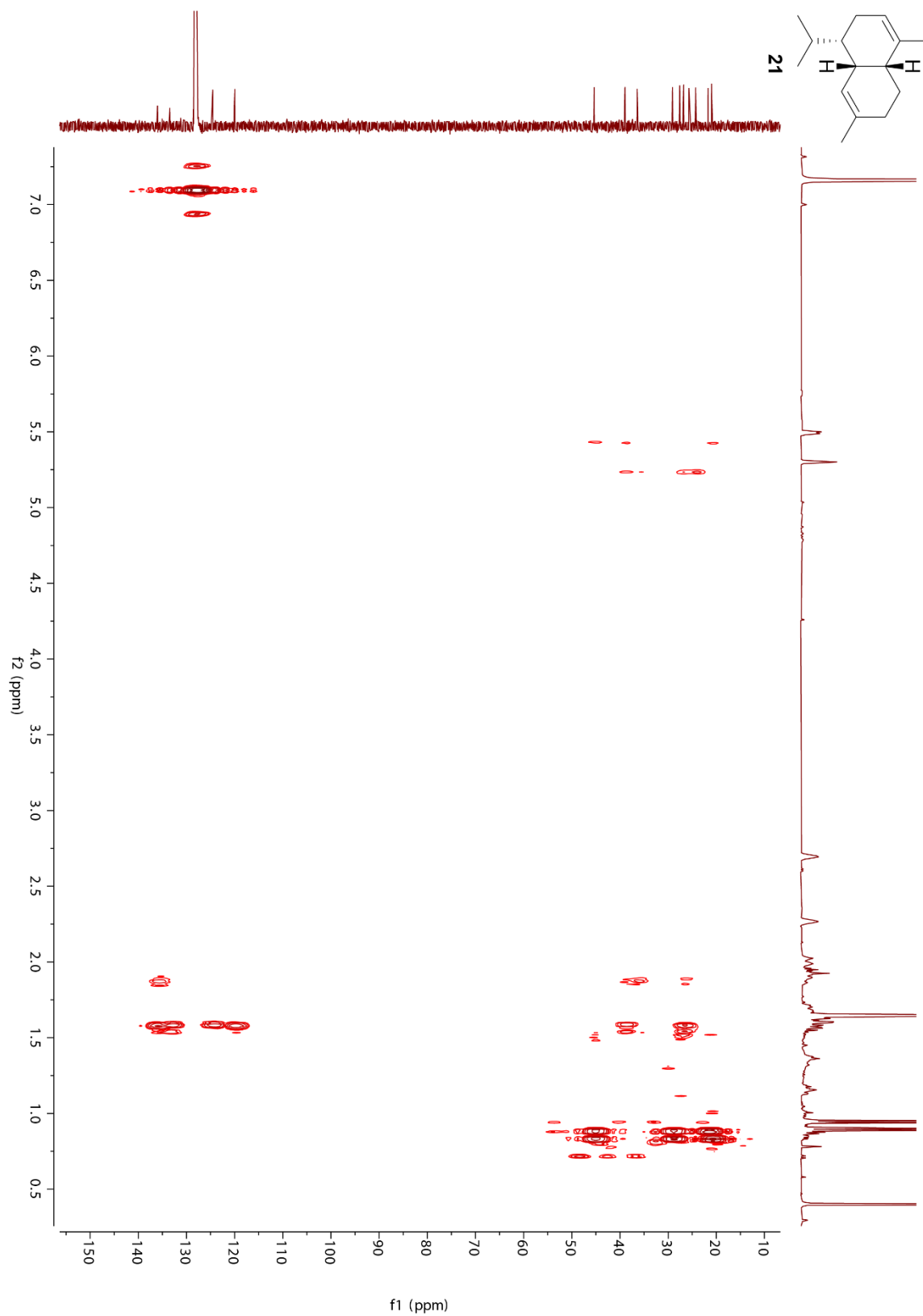
^1H , ^1H -NOESY spectrum (500 MHz) of 11 in C_6D_6 .

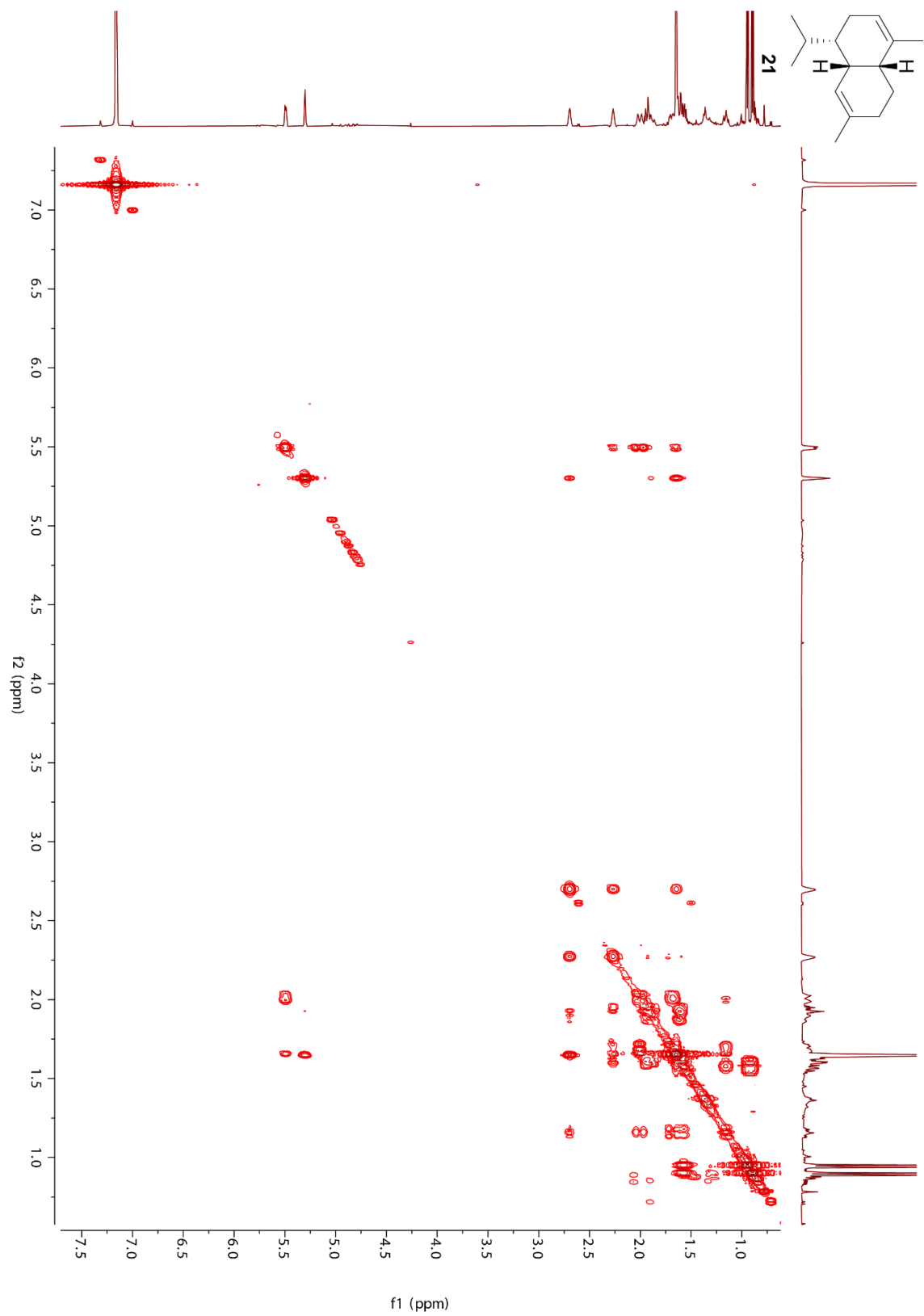


¹³C-NMR spectrum (125 MHz) of 21 in C₆D₆.

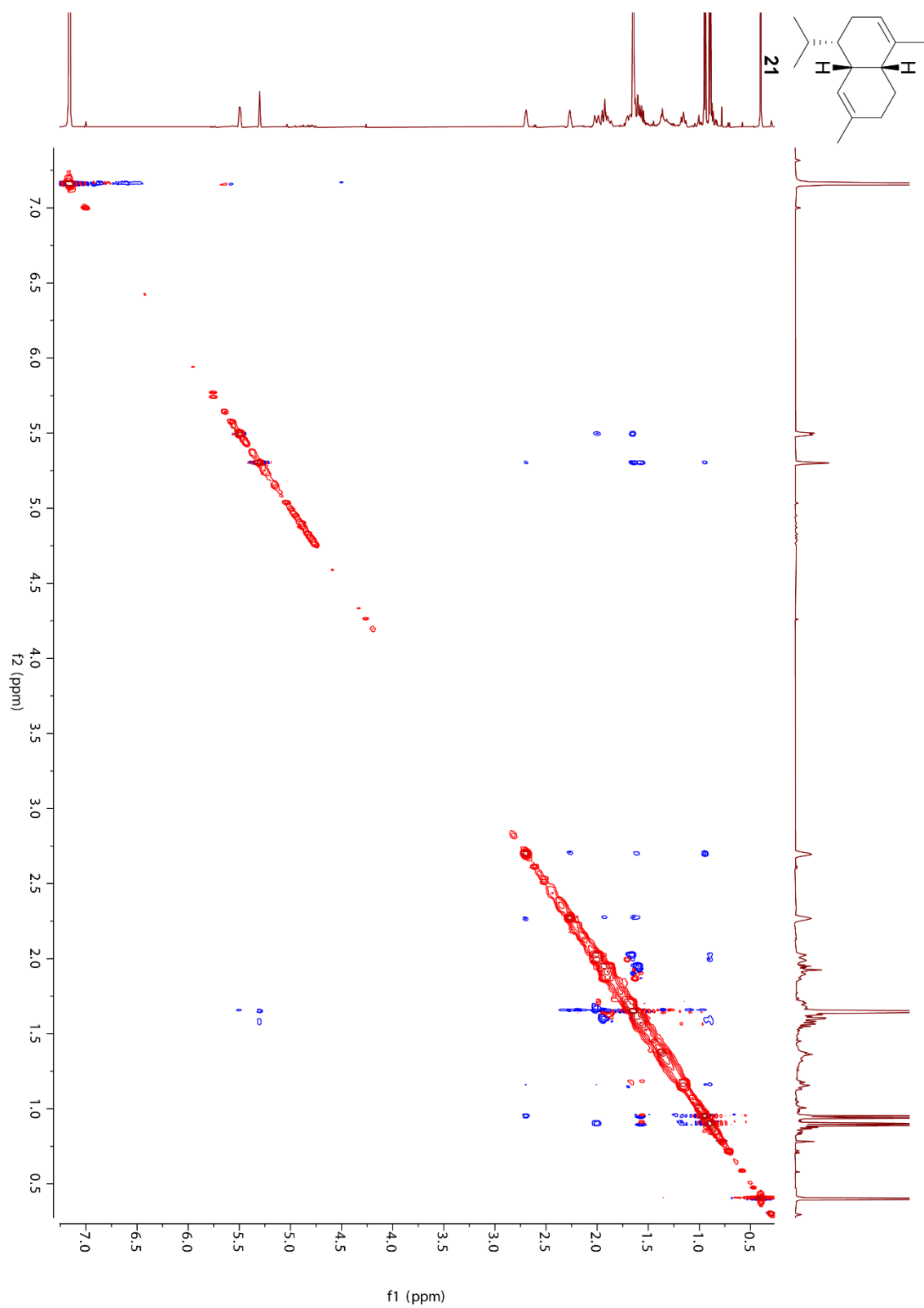








$^1\text{H}, ^1\text{H}$ -COSY spectrum (500 MHz) of 21 in C_6D_6 .



^1H , ^1H -NOESY spectrum (500 MHz) of **21** in C_6D_6 .

2.10 Supplementary Information References

- (1) Woodside, A. B.; Huang, Z.; Poulter, C. D. Trisammonium Geranyl Diphosphate. *Org. Synth.* **1988**, 66, 211. <https://doi.org/10.15227/orgsyn.066.0211>.
- (2) Davisson, V. J.; Woodside, A. B.; Poulter, C. D. Synthesis of Allylic and Homoallylic Isoprenoid Pyrophosphates. In *Methods in Enzymology*; Steroids and Isoprenoids, Part A; Academic Press, 1985; Vol. 110, pp 130–144. [https://doi.org/10.1016/S0076-6879\(85\)10068-6](https://doi.org/10.1016/S0076-6879(85)10068-6).
- (3) *Illustrated Trilingual Glossary of Morphological and Anatomical Terms Applied to Octocorallia*; Bayer, F. M., Grasshoff, M., Verseveldt, J., Eds.; Brill, 2023.
- (4) McFadden, C. S.; Benayahu, Y.; Pante, E.; Thoma, J. N.; Nevarez, P. A.; France, S. C. Limitations of Mitochondrial Gene Barcoding in Octocorallia. *Mol. Ecol. Resour.* **2011**, 11 (1), 19–31. <https://doi.org/10.1111/j.1755-0998.2010.02875.x>.
- (5) Larsson, A. AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets. *Bioinformatics* **2014**, 30 (22), 3276–3278. <https://doi.org/10.1093/bioinformatics/btu531>.
- (6) Patin, N. V.; Kunin, V.; Lidström, U.; Ashby, M. N. Effects of OTU Clustering and PCR Artifacts on Microbial Diversity Estimates. *Microb. Ecol.* **2013**, 65 (3), 709–719. <https://doi.org/10.1007/s00248-012-0145-4>.
- (7) Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. **2014**.
- (8) Peng, Y.; Leung, H. C. M.; Yiu, S. M.; Chin, F. Y. L. IDBA-UD: A de Novo Assembler for Single-Cell and Metagenomic Sequencing Data with Highly Uneven Depth. *Bioinformatics* **2012**, 28 (11), 1420–1428. <https://doi.org/10.1093/bioinformatics/bts174>.
- (9) Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, 9 (4), 357–359. <https://doi.org/10.1038/nmeth.1923>.
- (10) Buchfink, B.; Reuter, K.; Drost, H.-G. Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND. *Nat. Methods* **2021**, 18 (4), 366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
- (11) Muthye, V.; Mackereth, C. D.; Stewart, J. B.; Lavrov, D. V. Large Dataset of Octocoral Mitochondrial Genomes Provides New Insights into *Mt-mutS* Evolution and Function. *DNA Repair* **2022**, 110, 103273. <https://doi.org/10.1016/j.dnarep.2022.103273>.
- (12) Sievers, F.; Higgins, D. G. Clustal Omega. *Curr. Prot. Bioinf.* **2014**, 48 (1), 3.13.1–3.13.16. <https://doi.org/10.1002/0471250953.bi0313s48>.
- (13) Kozlov, A. M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference. *Bioinformatics* **2019**, 35 (21), 4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>.
- (14) Burkhardt, I.; de Rond, T.; Chen, P. Y.-T.; Moore, B. S. Ancient Plant-like Terpene Biosynthesis in Corals. *Nat Chem Biol* **2022**, 18 (6), 664–669. <https://doi.org/10.1038/s41589-022-01026-2>.
- (15) Mistry, J.; Finn, R. D.; Eddy, S. R.; Bateman, A.; Punta, M. Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions. *Nucleic Acids Res.* **2013**, 41 (12), e121. <https://doi.org/10.1093/nar/gkt263>.

- (16) Minh, B. Q.; Schmidt, H. A.; Chernomor, O.; Schrempf, D.; Woodhams, M. D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, 37 (5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- (17) Kalyaanamoorthy, S.; Minh, B. Q.; Wong, T. K. F.; von Haeseler, A.; Jermiin, L. S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, 14 (6), 587–589. <https://doi.org/10.1038/nmeth.4285>.
- (18) Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v6: Recent Updates to the Phylogenetic Tree Display and Annotation Tool. *Nucleic Acids Res.* **2024**, 52 (W1), W78–W82. <https://doi.org/10.1093/nar/gkae268>.
- (19) Pfeifer, B.; Wittelsbürger, U.; Ramos-Onsins, S. E.; Lercher, M. J. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Mol. Biol. Evol.* **2014**, 31 (7), 1929–1936. <https://doi.org/10.1093/molbev/msu136>.
- (20) Charif, D.; Lobry, J. R. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*; Bastolla, U., Porto, M., Roman, H. E., Vendruscolo, M., Eds.; Springer: Berlin, Heidelberg, 2007; pp 207–232. https://doi.org/10.1007/978-3-540-35306-5_10.
- (21) Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* **2004**, 20 (2), 289–290. <https://doi.org/10.1093/bioinformatics/btg412>.
- (22) Kryazhimskiy, S.; Plotkin, J. B. The Population Genetics of dN/dS. *PLOS Genet.* **2008**, 4 (12), e1000304. <https://doi.org/10.1371/journal.pgen.1000304>.
- (23) Khandia, R.; Singhal, S.; Kumar, U.; Ansari, A.; Tiwari, R.; Dhama, K.; Das, J.; Munjal, A.; Singh, R. K. Analysis of Nipah Virus Codon Usage and Adaptation to Hosts. *Front. Microbiol.* **2019**, 10. <https://doi.org/10.3389/fmicb.2019.00886>.
- (24) *Correspondence Analysis of Codon Usage*. <https://codonw.sourceforge.net/> (accessed 2025-07-10).
- (25) Plese, B.; Kenny, N. J.; Rossi, M. E.; Cárdenas, P.; Schuster, A.; Taboada, S.; Koutsouveli, V.; Riesgo, A. Mitochondrial Evolution in the Demospongiae (Porifera): Phylogeny, Divergence Time, and Genome Biology. *Mol. Phylogenet. Evol.* **2021**, 155, 107011. <https://doi.org/10.1016/j.ympev.2020.107011>.
- (26) Sueoka, N. Intrastrand Parity Rules of DNA Base Composition and Usage Biases of Synonymous Codons. *J. Mol. Evol.* **1995**, 40 (3), 318–325. <https://doi.org/10.1007/BF00163236>.
- (27) Blay, G.; Cardona, L.; Collado, A. M.; García, B.; Morcillo, V.; Pedro, J. R. Synthesis of Spirovetivane Sesquiterpenes from Santonin. Synthesis of (+)-Anhydro- β -Rotunol and All Diastereomers of 6,11-Spirovetivadiene. *J. Org. Chem.* **2004**, 69 (21), 7294–7302. <https://doi.org/10.1021/jo040189n>.
- (28) Rabe, P.; Schmitz, T.; Dickschat, J. S. Mechanistic Investigations on Six Bacterial Terpene Cyclases. *Beilstein J. Org. Chem.* **2016**, 12 (1), 1839–1850. <https://doi.org/10.3762/bjoc.12.173>.
- (29) Rinkel, J.; Lauterbach, L.; Rabe, P.; Dickschat, J. S. Two Diterpene Synthases for Spiroalbatene and Cembrene A from *Allokutzneria Albata*. *Angew. Chem. Int. Ed.* **2018**, 57 (12), 3238–3241. <https://doi.org/10.1002/anie.201800385>.
- (30) Musil, M.; Khan, R. T.; Beier, A.; Stourac, J.; Konegger, H.; Damborsky, J.; Bednar, D. FireProtASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. *Brief. Bioinform.* **2021**, 22 (4), bbaa337. <https://doi.org/10.1093/bib/bbaa337>.

- (31) Patiño Cano, L. P.; Quintana Manfredi, R.; Pérez, M.; García, M.; Blustein, G.; Cordeiro, R.; Pérez, C. D.; Schejter, L.; Palermo, J. A. Isolation and Antifouling Activity of Azulene Derivatives from the Antarctic Gorgonian *Acanthogorgia Laxa*. *Chem. Biodivers.* **2018**, *15* (1), e1700425. <https://doi.org/10.1002/cbdv.201700425>.
- (32) Manzo, E.; Ciavatta, M. L.; Gavagnin, M.; Villani, G.; Quaranta, C.; D'Souza, L.; Cimino, G. A New Xenicane Norditerpene from the Indian Marine Gorgonian *Acanthogorgia Turgida*. *Nat. Prod. Res.* **2009**, *23* (18), 1664–1670. <https://doi.org/10.1080/14786410802144941>.
- (33) Stonik, V. A.; Makar'eva, T. N.; Dmitrenok, A. S. New Diterpenoid of the Xeniane Series from the gorgonian *Paragorgia Arborea*. *Chem. Nat. Compd.* **1990**, *26* (1), 103–104. <https://doi.org/10.1007/BF00605222>.
- (34) D'Ambrosio, M.; Guerriero, A.; Pietra, F. Arboxeniolide-1, a New, Naturally Occurring Xeniolide Diterpenoid from the Gorgonian *Paragorgia Arborea* of the Crozet Is. (S. Indian Ocean). *Z. Naturforsch. C* **1984**, *39* (11–12), 1180–1183. <https://doi.org/10.1515/znc-1984-11-1232>.
- (35) Wang, S.-K.; Huang, M.-J.; Duh, C.-Y. Cytotoxic Constituents from the Formosan Soft Coral *Clavularia Inflata* Var. *Luzoniana*. *J. Nat. Prod.* **2006**, *69* (10), 1411–1416. <https://doi.org/10.1021/np0601253>.
- (36) Banjoo, D.; Mootoo, B. S.; Ramsewak, R. S.; Sharma, R.; Lough, A. J.; McLean, S.; Reynolds, W. F. New Erythrolides from the Caribbean Gorgonian Octocoral *Erythropodium Caribaeorum*. *J. Nat. Prod.* **2002**, *65* (3), 314–318. <https://doi.org/10.1021/np010333p>.
- (37) Cinel, B.; Roberge, M.; Behrisch, H.; van Ofwegen, L.; Castro, C. B.; Andersen, R. J. Antimitotic Diterpenes from *Erythropodium Caribaeorum* Test Pharmacophore Models for Microtubule Stabilization. *Org. Lett.* **2000**, *2* (3), 257–260. <https://doi.org/10.1021/ol9912027>.
- (38) Rodríguez, A. D.; Cobar, O. M. The Briarellins, New Eunicellin-Based Diterpenoids from a Caribbean Gorgonian, *Briareum Asbestinum*. *Tetrahedron* **1995**, *51* (25), 6869–6880. [https://doi.org/10.1016/0040-4020\(95\)00339-A](https://doi.org/10.1016/0040-4020(95)00339-A).
- (39) Rodríguez, A. D.; Ramírez, C.; Cobar, O. M. Briareins C–L, 10 New Briarane Diterpenoids from the Common Caribbean Gorgonian *Briareum Asbestinum*. *J. Nat. Prod.* **1996**, *59* (1), 15–22. <https://doi.org/10.1021/np960001y>.
- (40) Liaw, C.-C.; Shen, Y.-C.; Lin, Y.-S.; Hwang, T.-L.; Kuo, Y.-H.; Khalil, A. T. Frajunolides E–K, Briarane Diterpenes from *Junceella Fragilis*. *J. Nat. Prod.* **2008**, *71* (9), 1551–1556. <https://doi.org/10.1021/np800126f>.
- (41) Guerriero, A.; D'Ambrosio, M.; Pietra, F. Verecynarmin A, a Novel Briarane Diterpenoid Isolated from Both the Mediterranean Nudibranch Mollusc *Armina Maculata* and Its Prey, the Pennatulacean Octocoral *Veretillum Cynomorium*. *Helv. Chim. Acta* **1987**, *70* (4), 984–991. <https://doi.org/10.1002/hlca.19870700408>.
- (42) Guerriero, A.; D'Ambrosio, M.; Pietra, F. Isolation of the Cembranoid Preverecynarmin Alongside Some Briaranes, the Verecynarmins, from Both the Nudibranch Mollusc *Armina Maculata* and the Octocoral *Veretillum Cynomorium* of the East Pyrenean Mediterranean Sea. *Helv. Chim. Acta* **1990**, *73* (2), 277–283. <https://doi.org/10.1002/hlca.19900730206>.
- (43) Wratten, S. J.; Faulkner, D. J.; Hirotsu, K.; Clardy, J. Stylatulide, a Sea Pen Toxin. *J. Am. Chem. Soc.* **1977**, *99* (8), 2824–2825. <https://doi.org/10.1021/ja00450a083>.

- (44) Bahl, A.; Jachak, S. M.; Palaniveloo, K.; Ramachandram, T.; Vairappan, C. S.; Chopra, H. K. 2-Acetoxyverecynarmin C, a New Briarane COX Inhibitory Diterpenoid from *Pennatula Aculeata*. *Nat. Prod. Commun.* **2014**, 9 (8), 1934578X1400900820. <https://doi.org/10.1177/1934578X1400900820>.
- (45) Duh, C.-Y.; Hou, R.-S. Cytotoxic Cembranoids from the Soft Corals *Sinularia Gibberosa* and *Sarcophyton Trocheliophorum*. *J. Nat. Prod.* **1996**, 59 (6), 595–598. <https://doi.org/10.1021/np960174n>.
- (46) Bernstein, J.; Shmeuli, U.; Zadock, E.; Kashman, Y.; Néeman, I. Sarcophine, a New Epoxy Cembranolid from Marine Origin. *Tetrahedron* **1974**, 30 (16), 2817–2824. [https://doi.org/10.1016/S0040-4020\(01\)97451-4](https://doi.org/10.1016/S0040-4020(01)97451-4).
- (47) Wen, T.; Ding, Y.; Deng, Z.; van Ofwegen, L.; Proksch, P.; Lin, W. Sinulaflexiolides A–K, Cembrane-Type Diterpenoids from the Chinese Soft Coral *Sinularia Flexibilis*. *J. Nat. Prod.* **2008**, 71 (7), 1133–1140. <https://doi.org/10.1021/np070640g>.
- (48) Kusumi, T.; Uchida, H.; Ishitsuka, M. O.; Yamamoto, H.; Kakisawa, H. Alcyonin, a New Cladiellane Diterpene from the Soft Coral *Sinularia Flexibilis*. *Chem. Lett.* **1988**, 17 (6), 1077–1078. <https://doi.org/10.1246/cl.1988.1077>.
- (49) Tseng, Y.-J.; Ahmed, A. F.; Dai, C.-F.; Chiang, M. Y.; Sheu, J.-H. Sinulochmodins A–C, Three Novel Terpenoids from the Soft Coral *Sinularia Lochmodes*. *Org. Lett.* **2005**, 7 (17), 3813–3816. <https://doi.org/10.1021/ol051513j>.
- (50) Thao, N. P.; Nam, N. H.; Cuong, N. X.; Quang, T. H.; Tung, P. T.; Tai, B. H.; Luyen, B. T. T.; Chae, D.; Kim, S.; Koh, Y.-S.; Kiem, P. V.; Minh, C. V.; Kim, Y. H. Diterpenoids from the Soft Coral *Sinularia Maxima* and Their Inhibitory Effects on Lipopolysaccharide-Stimulated Production of Pro-Inflammatory Cytokines in Bone Marrow-Derived Dendritic Cells. *Chem. Pharm. Bull.* **2012**, 60 (12), 1581–1589. <https://doi.org/10.1248/cpb.c12-00756>.
- (51) Radwan, M. M.; Manly, S. P.; El Sayed, K. A.; Wali, V. B.; Sylvester, P. W.; Awate, B.; Shah, G.; Ross, S. A. Sinulodurins A and B, Antiproliferative and Anti-Invasive Diterpenes from the Soft Coral *Sinularia Dura*. *J. Nat. Prod.* **2008**, 71 (8), 1468–1471. <https://doi.org/10.1021/np800208k>.
- (52) Hooper, G. J.; Davies-Coleman, M. T. New Metabolites from the South African Soft Coral *Capnella Thyroidea*. *Tetrahedron* **1995**, 51 (36), 9973–9984. [https://doi.org/10.1016/0040-4020\(95\)00570-X](https://doi.org/10.1016/0040-4020(95)00570-X).
- (53) Hassan, H. M.; Khanfar, M. A.; Elnagar, A. Y.; Mohammed, R.; Shaala, L. A.; Youssef, D. T. A.; Hifnawy, M. S.; El Sayed, K. A. Pachycladins A–E, Prostate Cancer Invasion and Migration Inhibitory Eunicellin-Based Diterpenoids from the Red Sea Soft Coral *Cladiella Pachyclados*. *J. Nat. Prod.* **2010**, 73 (5), 848–853. <https://doi.org/10.1021/np900787p>.
- (54) Rudi, A.; Ketzinel, S.; Goldberg, I.; Stein, Z.; Kashman, Y.; Benayahu, Y.; Schleyer, M. Antheliatin and Zahavins A and B, Three New Cytotoxic Xenicane Diterpenes from Two Soft Corals. *J. Nat. Prod.* **1995**, 58 (10), 1581–1586. <https://doi.org/10.1021/np50124a016>.
- (55) Urda, C.; Fernández, R.; Pérez, M.; Rodríguez, J.; Jiménez, C.; Cuevas, C. Protoxicinins A and B, Cytotoxic Long-Chain Acylated Xenicanes from the Soft Coral *Protodendron Repens*. *J. Nat. Prod.* **2017**, 80 (3), 713–719. <https://doi.org/10.1021/acs.jnatprod.7b00046>.
- (56) Groweiss, A.; Kashman, Y. Xeniculin, Xeniphyllenol and Xeniphyllenol Oxide, New Diterpenoids from the Soft-Coral *Xenia Macroscopiculata*. *Tetrahedron Lett.* **1978**, 19 (25), 2205–2208. [https://doi.org/10.1016/S0040-4039\(01\)86846-5](https://doi.org/10.1016/S0040-4039(01)86846-5).

- (57) Rosa, S. D.; Cimino, G.; Giulio, A. D.; Milone, A.; Crispino, A.; Iodice, C. A New Bioactive Eunicellin-Type Diterpene from the Gorgonian Eunicella Cavolini. *Nat. Prod. Lett.* **1995**, *7* (4), 259–265. <https://doi.org/10.1080/10575639508043220>.
- (58) Look, S. A.; Fenical, W.; Matsumoto, G. K.; Clardy, J. The Pseudopterosins: A New Class of Antiinflammatory and Analgesic Diterpene Pentosides from the Marine Sea Whip Pseudopterogorgia Elisabethae (Octocorallia). *J. Org. Chem.* **1986**, *51* (26), 5140–5145. <https://doi.org/10.1021/jo00376a016>.
- (59) Rodríguez, A. D.; Shi, J.-G.; Huang, S. D. Highly Oxygenated Pseudopterane and Cembranolide Diterpenes from the Caribbean Sea Feather Pseudopterogorgia Bipinnata. *J. Nat. Prod.* **1999**, *62* (9), 1228–1237. <https://doi.org/10.1021/np990064r>.
- (60) Gutiérrez, M.; Capson, T. L.; Guzmán, H. M.; González, J.; Ortega-Barría, E.; Quiñoá, E.; Riguera, R. Leptolide, a New Furanocembranolide Diterpene from Leptogorgia Alba. *J. Nat. Prod.* **2005**, *68* (4), 614–616. <https://doi.org/10.1021/np049745z>.
- (61) Muthye, V.; Mackereth, C. D.; Stewart, J. B.; Lavrov, D. V. Large Dataset of Octocoral Mitochondrial Genomes Provides New Insights into Mt-mutS Evolution and Function. *DNA Repair* **2022**, *110*, 103273. <https://doi.org/10.1016/j.dnarep.2022.103273>.

2.11 Acknowledgements

Chapter 2 contains material currently under revision at the Proceedings of the National Academy of Sciences that has been accepted with minor revisions. The dissertation author was a primary co-researcher of this chapter with Bone, H.K.B.

CHAPTER 3: Uncovering a Widespread Metabolic Gene Cluster Family in Octocoral

3.1 Introduction for Chapter 3

As scientists, we're often driven by genuine curiosity: the pursuit of discovery and a desire to push the boundaries of our understanding of the world around us. But despite that shared motivation, beyond collaborations between labs with different specialties, much of what we do often remains siloed.

At its core, the academic system isn't built to encourage collaboration. Funding is finite and increasingly competitive. Recognition often goes to the lab that publishes first, not the one that builds on shared knowledge. Add to that the fact that groups working on similar problems are often scattered across institutions and countries, and it's easy to see why many labs end up advancing in isolation. But this redundancy comes at a cost: parallel experiments drain resources, duplicate effort, and sometimes lead to rushed or lower-quality science under the pressure to be first. When research funding is tight and the stakes of discovery are high, we can't afford to work in isolation.

During COVID, when most labs had paused non-COVID research and in-person conferences were off the table, collaboration became even harder. Around that time, our lab and Eric Schmidt's lab independently characterized and published, in the same issue of *Nature Chemical Biology*, the discovery of genomically encoded octocoral terpene synthases. Both papers also revealed the discovery of putative biosynthetic gene clusters encoded in octocoral genomes. Dr. Immo Burkhardt, from the Moore Lab, identified a cluster in a sea pansy,¹ while Paul, at the time a postdoc in the Schmidt lab, uncovered a homologous region in an encrusting *Erythropodium gorgonian* soft coral.² That moment could have triggered a race between the groups to validate the first potential metazoan biosynthetic gene clusters. Instead, we extended an olive branch. The unexpected genomic homology in the distantly related coral's genomes didn't just suggest a deeper evolutionary story, it set the stage for collaboration. In a way, the shared DNA led us down a yellow brick road of sorts, paved with A, C, G, and T, guiding two labs toward each other rather than into competition.

We started slowly- Zoom calls, sharing data, discussing our approaches and soon realized the power of our combined strengths. The Moore Lab's expertise in genomics and phylogenetics was a perfect complement to the Schmidt Lab's biochemical platform. From here we built a plan. I was mentored by Dr.

Immo Burkhardt, as I took the reins on sequencing, assembling, and genome mining efforts. While I identified genes related to terpenoid biosynthesis from phylogenetically distinct octocorals, Paul characterized the encoded enzymes through a combination of heterologous expression systems. Together, our approaches informed and reinforced one another, accelerating discovery and expanding the scope of what either lab could have achieved alone.

Through this collaboration, we uncovered a conserved five gene biosynthetic cluster responsible for the production of briarane type diterpenoids. We sequenced five octocoral species from evolutionarily distinct families, all known producers of briaranes, and found that each harbored this same gene cluster. Using heterologous expression, we reconstituted the biosynthesis of cembrene B γ -lactone, a molecule bearing the conserved lactone ring of briarane type diterpenes. This discovery not only validated the function of terpene cyclases, cytochrome P450s, and a short chain dehydrogenase acting in concert, but also demonstrated that animals, like microbes and plants, can use gene clusters to encode complex chemical pathways. It expands the known boundaries of biosynthetic logic in metazoans and reinforces the evolutionary significance of gene clustering for specialized metabolite production.

The research described in this chapter arose from a close collaboration among Paul D. Scesa, Malia L. Moore, Jean-Baptiste Ledoux, Jessica Gomez-Garrido, Tyler Alioto, Todd P. Michael, Immo Burkhardt, Eric W. Schmidt, and Bradley S. Moore. As co-first author with Paul D. Scesa, I played a major role in the data generation, analysis, and writing of the manuscript. Specifically, I sourced and extracted high-quality genomic DNA from the corals, assembled the genome, and mined the data for terpenoid biosynthetic genes. Genomic sequencing was conducted in collaboration with Malia L. Moore. Paul D. Scesa subsequently led the enzymatic work. While the writing of the manuscript was a highly collaborative effort, I prepared the initial draft and coordinated its integration across all contributors.

3.2 Chapter 3 Introduction References

- (1) Burkhardt, I.; de Rond, T.; Chen, P. Y.-T.; Moore, B. S. Ancient Plant-like Terpene Biosynthesis in Corals. *Nat. Chem. Biol.* **2022**, *18* (6), 664–669. <https://doi.org/10.1038/s41589-022-01026-2>.
- (2) Scesa, P. D.; Lin, Z.; Schmidt, E. W. Ancient Defensive Terpene Biosynthetic Gene Clusters in the Soft Corals. *Nat. Chem. Biol.* **2022**, *18* (6), 659–663. <https://doi.org/10.1038/s41589-022-01027-1>.

3.3 Reprint of “A widespread metabolic gene cluster family in metazoans”

A widespread metabolic gene cluster family in metazoans

Received: 15 August 2024

Accepted: 30 April 2025

Published online: 13 June 2025

 Check for updates

Natalie E. Grayson^{1,8}, Paul D. Scesa^{2,8}, Malia L. Moore^{1,3},
Jean-Baptiste Ledoux⁴, Jessica Gomez-Garrido^{5,6}, Tyler Alioto^{5,6},
Todd P. Michael^{1,3}, Immo Burkhardt¹✉, Eric W. Schmidt²✉ &
Bradley S. Moore^{1,7}✉

Octocorals are metazoans that prolifically produce terpenoid natural products, rivaling the chemical diversity of plants and microbes. We recently established that these cnidarians uniformly express terpene cyclases and that their encoding genes often reside within putative biosynthetic gene clusters (BGCs). Here we report the discovery and characterization of a widespread gene cluster family for briarane diterpenoid biosynthesis. We sequence five genomes from evolutionarily distinct families of briarane-producing octocorals, revealing a conserved five-gene cluster. Expressing these genes in heterologous hosts, we reconstitute the biosynthesis of cembrene B γ -lactone, an established molecule that contains the lactone structural feature distinctive of briarane diterpenoids. The discovery of the genomic basis of briarane biosynthesis establishes that animals also use gene cluster families to produce specialized metabolites. Furthermore, the presence of BGCs in octocorals proves that the formation and maintenance of BGCs related to specialized metabolite biosynthesis is a more widespread phenomenon than previously realized.

Specialized metabolites or natural products are small molecules that, while not vital for primary growth, enable a producing organism to survive in an ecological niche by mediating time-dependent, tissue-dependent or stimulus-dependent functions¹. Although most of these compounds are known from plants and microbes, they are widespread in all kingdoms of life, including animals, where they too serve diverse purposes including communication, development and defense². Examples include volatile sex pheromones from frogs³ and spiders⁴, polyketide pigments in parrots⁵ and sea urchins⁶, sugar-derived hormones in nematodes⁷, hydroquinones used by rove beetles to fend off predators⁸ and peptide toxins used by cone snails to stun their prey⁹.

The overwhelming majority of research on specialized metabolite biosynthesis has been conducted in bacteria and fungi, where genes coding for a biosynthetic pathway are typically arranged in physically colocalized, biosynthetic gene clusters (BGCs)^{10,11}. In contrast, the few characterized biosynthetic pathways in animals are not encoded in BGCs but have their encoding genes dispersed throughout the genome. This is currently regarded as the general rule for biosynthetic pathways in nonfungal eukaryotes¹². However, improved genomic resources in other eukaryotes such as plants and algae have revealed the colocalization of nonparalogous genes into BGCs, where they code for specialized metabolic pathways (Fig. 1)^{13,14}.

¹Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA.

²Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT, USA. ³The Plant Molecular and Cellular Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA. ⁴CIIMAR/CIMAR LA, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Porto, Portugal.

⁵Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain. ⁶Universitat de Barcelona (UB), Gran Via de les Corts Catalanes, Barcelona, Spain.

⁷Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA. ⁸These authors contributed equally:

Natalie E. Grayson, Paul D. Scesa. ✉e-mail: iburkhardt@ucsd.edu; ewst@utah.edu; bsmoore@ucsd.edu

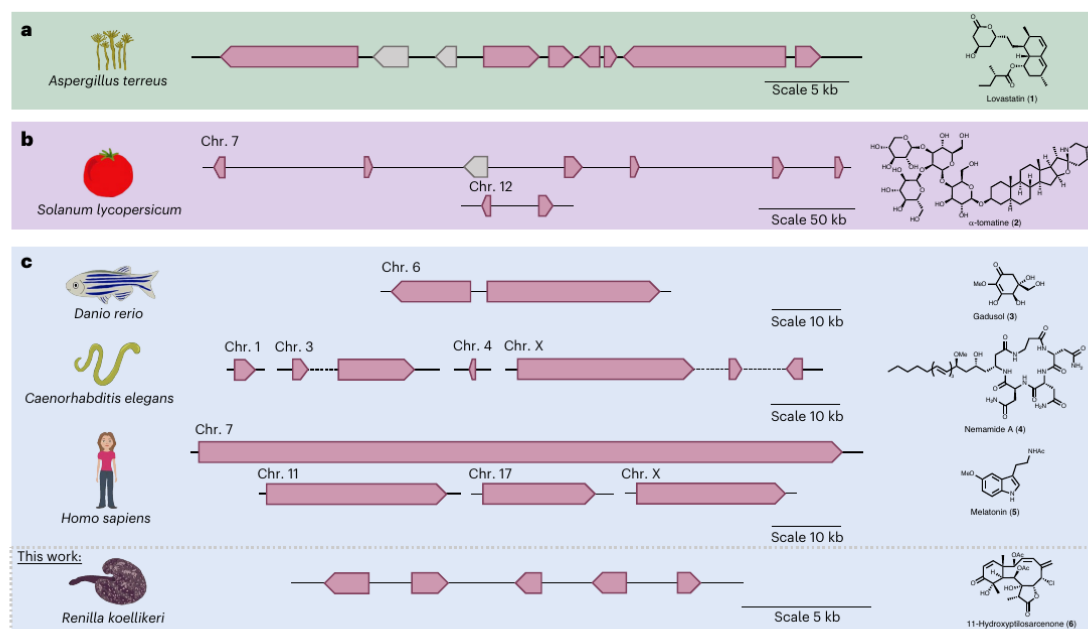


Fig. 1 | Gene clustering trends across eukaryotes. **a**, Example fungal BGC. Fungal gene clusters are often relatively compact and encode a complete biosynthetic pathway, as exemplified in the lovastatin (1) BGC from *A. terreus*. **b**, Example plant BGC. Plant BGCs often encode only partial pathways and are more extended across the chromosomes compared to fungal BGCs, as shown for the two partial tomatine (2) BGCs from *Solanum lycopersicum*. **c**, Example animal BGCs. Genes coding for specialized metabolite biosynthesis in animals sometimes occur in clusters such as the genes for gadusol (3) production, but they show high

similarity and synteny to microbial genes and are, therefore, likely a result of horizontal gene transfer. Other animal genes that are not HGT products are usually not colocalized such as for nemamide A (4) or melatonin (5) production in *C. elegans* and *Homo sapiens*, respectively. This work focuses on the investigation and characterization of a relatively tight BGC occurring in numerous octocorals such as *R. koellikeri*, putatively encoding enzymes to produce briaranes such as 11-hydroxyptilosarcenone (6).

Because of their abundance, eukaryotic BGCs were first recognized in fungi, where they are now regarded as the norm in specialized metabolite biosynthesis. A prominent example is the lovastatin (1) BGC, which codes for nine nonhomologous genes within a compact genomic region of 40 kb and encodes the complete set of enzymes needed to produce 1, a resistance gene, a transporter and a transcription factor (Fig. 1)¹⁵. The more recently discovered BGCs in plant genomes show clear differences from fungal BGCs. As generally the case in plant genomes, they show larger intergenic regions and do not regularly code for the complete biosynthetic pathway, although exceptions exist such as for the 12-gene product avenacin¹⁶. A typical example for partial clustering is the tomatine (2) BGC, which codes for six steps in the pathway starting from cholesterol¹⁷. However, the first enzymatic step of the pathway is encoded by a gene residing over 7,000 kb away from this cluster, while two additional genes are coded as a gene pair on a different chromosome (Fig. 1)¹⁸. There are only a few examples of small-molecule BGCs in animals that have since been identified. However, their origins link to microbial gene sets acquired through horizontal gene transfer (HGT) as in the case of mycosporine-like amino acids such as gadusol (3) that widely occur in vertebrates¹⁹ (Fig. 1) and a penicillin biosynthesis gene pair from the springtail *Folsomia candida* that shows high homology to fungal sequences^{20,21}. In contrast, the genes involved in the biosynthesis of the hybrid polyketide and nonribosomal peptide hormone nemamide A (4) from the nematode *Caenorhabditis elegans* do not show clear sequence homology with microbial genes and, thus, may have evolved de novo in nematodes²². However, none of these genes are colocalized

in a cluster, but dispersed over large genomic regions across four separate chromosomes.

Although uncommon, a few gene clusters in animal genomes that are not the result of horizontal gene transfer are known, such as the *Hox*²³ and β -globin-encoding²⁴ clusters in humans, as well as a laccase gene cluster recently discovered in rove beetles²⁵. However, these gene clusters are made up of paralogous genes and have arisen from successive tandem duplications from one ancestor gene; only the last example is involved in small-molecule biosynthesis. Generally, the few characterized biosynthetic pathways in animals, including the extraordinarily well-investigated examples of steroid hormones²⁶ or melatonin (5)²⁷ in humans, are not encoded in BGCs but have their encoding genes dispersed throughout the genome.

It was, therefore, surprising when we recently and independently identified potential terpenoid BGCs in octocorals, consisting of up to 13 colocalized genes spanning up to 500 kb that did not result from HGT or pure tandem duplications^{28,29}. Instead, over evolutionary time, coral gene paralogs from different protein families appear to have been recruited to single genomic scaffolds, forming potential BGCs. The capacity to produce terpenoids is exclusive to octocorals and absent in the sister class of hexacorals (which includes sea anemones and stony corals) and other cnidarians.

Within our analyzed datasets, we observed related putative gene clusters from *Renilla muelleri* and *Erythropodium caribaeorum*, corals that belong to different families (Pennatuloidae versus Erythropodiidae) and differ in their lifestyle and morphology. Yet, both corals are known sources of briarane-type diterpenoids, a Scleractinia-specific

structural family of diterpenoids, numbering over 700 representatives and showing a wide array of potent biological properties³⁰.

Here, we test the hypothesis that briarane diterpenoids are encoded by broadly distributed and ancient BGCs, which is unprecedented across metazoans. We assemble and analyze high-quality genome sequences of six key briarane-producing octocorals from different taxonomic families, revealing a conserved gene cluster family. Using a combination of protein biochemistry and pathway engineering, we reconstitute the biosynthesis of cembrene B γ -lactone, which contains the γ -lactone structural feature distinctive of all briarane diterpenoids. These results provide compelling evidence that corals use BGCs to synthesize their abundant bioactive terpenoids, opening the possibility that other metazoans may have similarly evolved BGCs in their own specialized metabolism.

Results

Chemical validation of briarane production

We previously characterized homologous cembrene B synthases identified from the genomes of *R. muelleri* and *E. caribaeorum*^{28,29}. Both encoding genes were colocalized with cytochrome P450 (CYP) genes, which also shared homology with their counterparts on the two distinct genomes. The tantalizing possibility that coral animals evolved gene cluster families for the biosynthesis of structurally related specialized metabolites motivated us to evaluate whether diverse corals known to contain briarane diterpenoids also encode cembrene B synthases in their genomes. As such, we first set out to confirm literature reports of briarane-producing octocoral families. Previously, Pennatulioidea, Ellisellidae, Coralliidae, Briareidae and Erythropodiidae were established as briarane-containing octocorals (Fig. 2a)^{31–37}. As each of these five coral families belong to the Scleractinia order, we further examined the remaining 14 families within this taxonomic order and found that five others reported terpenoid chemistry, yet not with briarane scaffolds (Fig. 2b and Supplementary Table 1) (<https://marinlit.rsc.org>). For the remaining nine families, no studies on terpenoids were found.

We, thus, collected coral tissue from representative species across five known briarane-producing families, including *Renilla koellikeri* (family Pennatulioidea), *Stylatula elongata* (family Pennatulioidea)³⁸, *Dichotella gemmacea* (family Ellisellidae), the branching and encrusting morphologies of *Briareum asbestinum* (family Briareidae), *E. caribaeorum* (family Erythropodiidae) and *Corallium rubrum* (family Coralliidae) (Extended Data Fig. 1). *R. koellikeri*, collected by SCUBA in San Diego, is not a reported source for briaranes but is related to the Atlantic *R. muelleri* that produces briaranes³⁹ and whose genome encodes cembrene B synthase as part of a putative BGC^{28,29}. We analyzed the tissue of *R. koellikeri* and isolated 11-hydroxyptilosarcenone (**6**), whose chemical structure was confirmed by nuclear magnetic resonance (NMR) spectroscopy (Supplementary Note 2, Supplementary Table 2 and Supplementary Fig. 1). This compound was previously identified from the sea pen *Ptilosarcus gurneyi* but not from a *Renilla* sea pansy^{40,41}.

Limitations in accessible biomass for the other octocorals targeted in this study led us to use molecular networking analyses of crude extracts and semipure fractions to confirm briarane production. However, none of our features were spectral matches to diterpenes in the GNPS library. We, therefore, used the *R. koellikeri* 11-hydroxyptilosarcenone (**6**) as a standard in the molecular network to identify a cluster containing 54 nodes representing unique masses of briarane compounds (Fig. 2c, Extended Data Fig. 2 and Supplementary Table 3). This analysis not only expanded the ptilosarcenone network from *R. koellikeri* (Pennatulioidea) but also established related briaranes in *D. gemmacea* (Ellisellidae), *E. caribaeorum* (Erythropodiidae) and *B. asbestinum* (Briareidae). Two nodes corresponding to compounds in these corals had exact masses consistent with literature values: gemmacolide R (**7**) from *D. gemmacea* and erythrolide D (**8**) from *E. caribaeorum* (Fig. 2c and Supplementary Table 1)^{42,43}. However, we did not

detect briarane nodes from the extracts of *C. rubrum* (Coralliidae) or *S. elongata* (Pennatulioidea). While there is literature precedent of *S. elongata* and corals in the family Coralliidae, albeit not specifically *C. rubrum*, making briaranes, we included these corals in our study assuming that they should have the genomic capacity to produce briaranes even though the samples we analyzed were devoid at the time of collection^{34,35,44}.

Genome assembly and syntenic analysis of briarane BGC family

At the start of this study, a few briarane-containing coral draft genomes with very low continuity (contig N50) were available: *E. caribaeorum* (N50 = 2 kb)²⁹, *R. muelleri* (N50 = 70.5 kb)⁴⁵ and *Pteroeides caledonicum* (N50 = 4 kb) (assembly accessions and statistics in Supplementary Table 4). To capture intact BGCs, we proceeded to generate higher-quality assemblies of corals across four briarane-producing families, obtaining long-read sequencing data for *R. koellikeri* (family Pennatulioidea), *S. elongata* (family Pennatulioidea), *D. gemmacea* (family Ellisellidae), the branching and encrusting morphologies of *B. asbestinum* (family Briareidae) and *E. caribaeorum* (family Erythropodiidae). We assembled 26–78 Gb of sequencing reads into genomes ranging in size from 231 Mb (*R. koellikeri*) to 1,237 Mb (*B. asbestinum*) (Table 1). To assess the completeness of our assemblies, we estimated genome sizes with *k*-mer-based analyses using Illumina short reads (Supplementary Fig. 2), producing genomes ranging from 144 Mb (*R. koellikeri*) to 1,155 Mb (*B. asbestinum*) (Table 1). Illumina short-read coverage was too shallow for *E. caribaeorum* to allow for *k*-mer-based analysis (sequencing data in Extended Data Table 1). Additionally, we included our recently reported chromosomally resolved *C. rubrum* (family Coralliidae) genome at 475 Mb (20 chromosomes, 951 scaffolds)^{46,47} in our analysis to enable BGC exploration across all five known briarane-producing coral families. The genome assemblies are now some of the most contiguous for octocorals with contig N50 values above 196 kb (Table 1, Fig. 3a and Extended Data Table 2). The chromosomal resolution assembly of *C. rubrum* had a scaffold N50 of 16,290 kb, on par with the current best scaffolded octocoral assembly from *Xenia* sp. with a scaffold N50 of 14,832 kb (ref. 48). Benchmarking universal single-copy ortholog (BUSCO) completeness scores ranged from 86.4% to 88% (Table 1). BUSCO scores are highly dependent on the database used for the analysis and the low scores could reflect the use of the metazoan database, which was the closest available. The lower-than-expected completeness is on par with the BUSCO for the *Xenia* assembly, which was 88.1% complete (Fig. 3a)⁴⁸. We further assembled transcriptomes to guide our genomic analyses by identifying actively transcribed genes and determining intron bounds (Extended Data Table 2).

Using the putative briarane BGC gene sequences from *R. muelleri* and *E. caribaeorum* as queries, we analyzed the five newly established genomes of *D. gemmacea*, *B. asbestinum*, *R. koellikeri*, *C. rubrum* and *S. elongata* and the public *P. caledonicum* for homologous genes. In total, the genomes from these eight octocoral species span five of the 20 octocoral families in the order Scleractinia. To determine whether non-briarane-producing octocoral families also have syntenic BGCs, we also explored the publicly available *Heliopora coerulea* (family Helioporidae) genome⁴⁹. We identified genes encoding homologs of the cembrene B terpene cyclase (cbTC) in all eight octocorals shown or suspected to produce briaranes, whereas the *H. coerulea* genome did not contain a homologous sequence, although it did encode nine other TCs. In the eight octocoral species containing a homologous cbTC gene, seven contained iterations of the briarane BGC (Fig. 3b). Given that the *P. caledonicum* genome had poor contiguity, we were not surprised to find the cbTC homolog on a 3-kb contig and expect that it too is part of a BGC.

We observed one conserved short-chain dehydrogenase (*cbSDH*) and up to three CYPs (*cbCYPa*, *cbCYPb* and *cbCYPc*) per briarane BGC

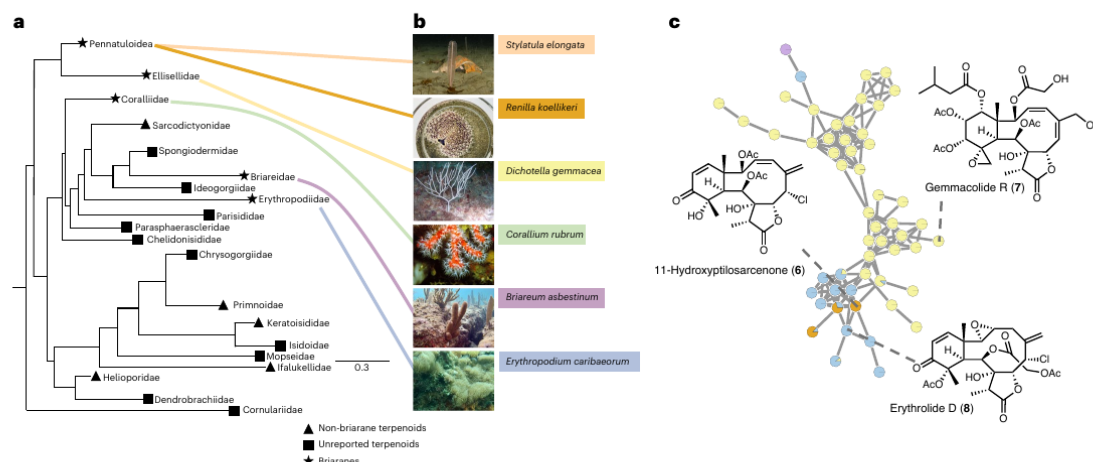


Fig. 2 | Distribution of briarane diterpenoids in the phylum Octocorallia. **a**, Phylogeny of octocoral rooted at Scleractyonacea. Families containing known briarane producer species are denoted with a star. Families with reported other terpene chemistry are marked with triangles, while families with no reported chemical data are marked with squares. **b**, Organisms used in the molecular

network. **c**, The briarane diterpenoid-containing molecular network. Highlighted nodes include 11-Hydroxyptilosarconone (6) from *R. koellikeri*, erythrolide D (8) from *E. caribaeorum* and gemmacolide R (7) from *D. gemmacea*. The full molecular network is shown in Extended Data Fig. 2.

Table 1 | Assembly statistics and BUSCO assessment of six octocoral genomes

Organism	Contigs (n)	Assembly size (Mb)	k-mer-based genome size (Mb)	Contig N50 (kb)	Longest (kb)	BUSCO (%C/%D)
<i>B. asbestinum</i>	22,526	1,254	1,155	332	4,126	87.3/4.6
<i>D. gemmacea</i>	8,171	550	401	196	2,036	87.5/15.6
<i>R. koellikeri</i>	557	231	144	4,895	22,942	86.9/2.6
<i>S. elongata</i>	392	360	269	5,139	39,131	88.0/1.0
<i>E. caribaeorum</i>	8,289	300	-	233	1,576	86.4/0.6
<i>C. rubrum</i> ⁴⁶	20 chromosomes (951 scaffolds)	545	540	1,600 (scaffold N50=18,521 kb)	115,000	88.5/1.2

(Fig. 3b and Extended Data Fig. 3). In the eight octocorals with briarane BGCs, four maintained the full five-gene briarane BGC spanning 22 kb in *R. muelleri* and up to 40 kb in *C. rubrum*, albeit with different gene organization. Notably, in the cases of *S. elongata*, *E. caribaeorum* and *B. asbestinum*, the orthologous genes were split between two BGC loci (Supplementary Table 5). For *S. elongata* and *B. asbestinum*, their respective *cbTC* and *cbCYPc* genes reside on one contig, while *cbCYPa*, *cbCYPb* and at least two *cbSDHs* are on a second contig. In the case of *E. caribaeorum*, *cbCYPc* resides outside the BGC contig. Furthermore, in the case of *B. asbestinum*, the orthologous genes have much longer gaps, resulting in substantially larger BGCs spanning ~150 kb. *S. elongata* and *B. asbestinum* further have multiple copies of the *SDH* gene at three (85%, 89% and 90% sequence similarity) and two (46% sequence similarity) copies, respectively. The duplications of the *SDHs* and gene rearrangements suggest lineage-specific independent evolution across octocoral species.

Because we identified the same set of homologous CYPs in a wide range of octocoral species, we conducted a phylogenetic analysis of these sequences in the background of broader cnidarian CYPs to shed light on their evolutionary origin. Our analysis included CYPs from four nonoctocoral cnidarians as a reference set: *Hydra vulgaris* (24 CYPs), *Acropora digitifera* (24 CYPs), *Aurelia aurita* (37 CYPs) and *Nematostella vectensis* (70 CYPs)⁵⁰. From octocorals, we included

CYPs from six briarane-producing species, *D. gemmacea*, *B. asbestinum*, *R. koellikeri*, *S. elongata*, *C. rubrum* and *E. caribaeorum*, as well as the non-briarane-producing *Xenia* sp. and *H. coerulea*, yielding 318 octocoral CYP sequences at ~40 encoded per genome (Supplementary Table 6). We combined the annotated reference cnidarian CYPs with the annotated octocoral CYPs to build a phylogenetic tree of protein sequences (Fig. 3c and Supplementary Fig. 3). The CYPs from the four reference cnidarians were previously associated with accepted CYP clans and families⁵⁰, thereby enabling us to examine the evolutionary context of the octocoral BGC CYPs. Notably, some octocoral CYPs fell into the clades associated with the cnidarian CYP clans, suggesting that they fulfill primary, widespread functions generally present in cnidarians. Others, including the BGC-associated CYPs, formed new monophyletic clades that were octocoral specific. The three briarane CYPs, *cbCYPa*–*cbCYPc*, form monophyletic clades consisting of *cbCYPa*, *cbCYPb* and *cbCYPc* homologs from different species (Fig. 3c). While *cbCYPb* and *cbCYPc* share a common ancestor that may suggest related enzymatic functions, *cbCYPa* is distantly related and instead lies closer to clan 46 with predicted functions associated with steroid biosynthesis (Fig. 3c). Interestingly, *H. coerulea*, while also from the order Scleractyonacea, has CYPs closely related to the *cbCYPb* and *cbCYPc* clades but the genes encoding these proteins are not colocalized, hinting at the loss of the briarane BGC in this lineage.

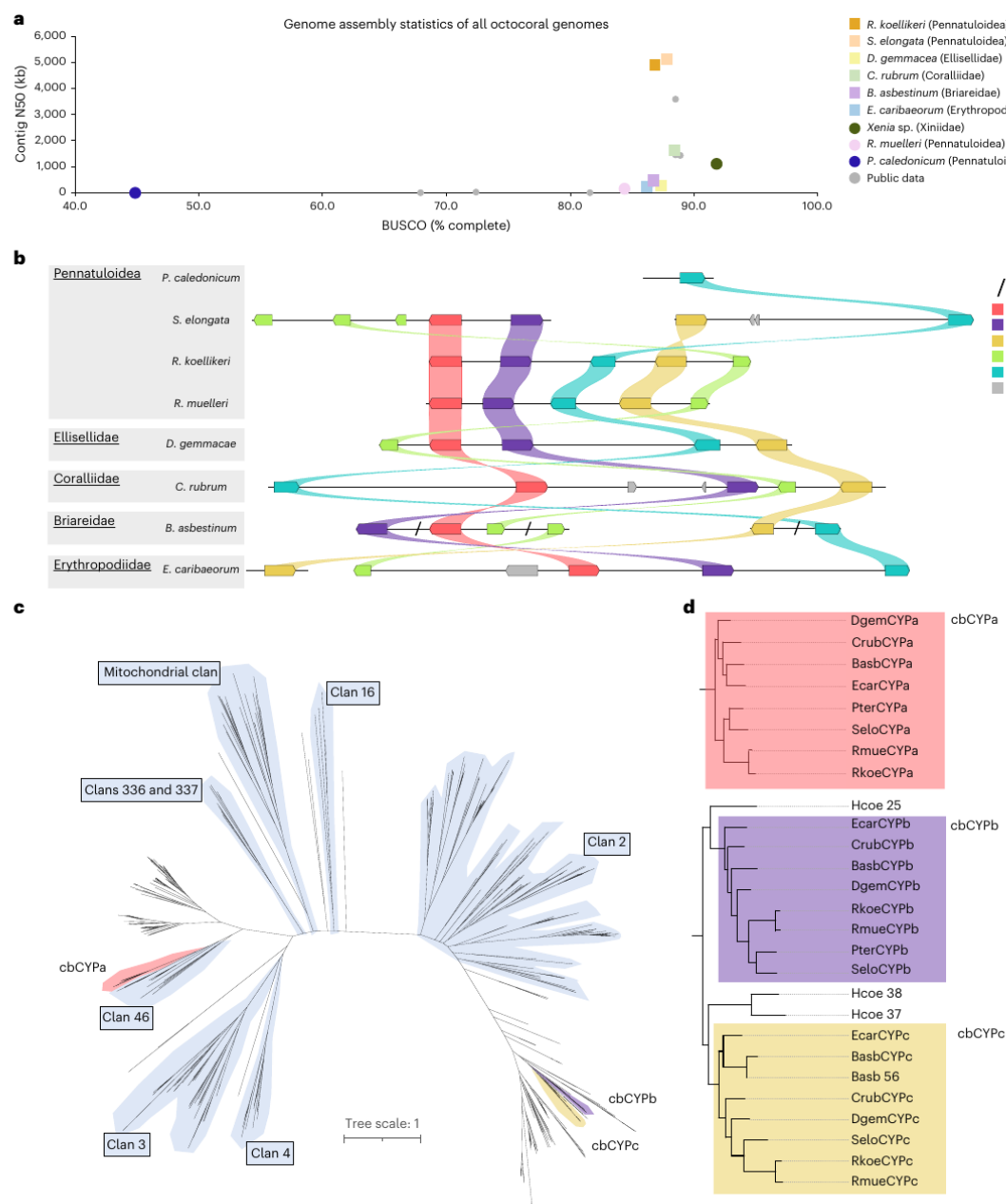


Fig. 3 | Genome-guided discovery of a conserved octocoral BGC family. **a**, Comparison of BUSCO and contig N50 data from 17 octocoral genomes. **b**, Synteny of the briarane BGCs in Scleractyonacea corals. Intron sequences were removed for ease of syntenic visualization (depiction of full BGC to scale in Extended Data Fig. 3). **c**, Phylogeny of cnidarian CYPs. The tree includes all CYPs from eight octocorals (*H. coerulea*, *D. gemmacea*, *B. asbestinum*, *R. koellikeri*, *S. elongata*, *C. rubrum*, *E. caribaeorum* and *Xenia* sp.) and four nonoctocoral

cnidarians (*H. vulgaris*, *A. digitifera*, *A. aurita* and *N. vectensis*) as reference. Previously established CYP clans⁵⁰ are marked with blue and contain octocoral and other cnidarian CYPs. Unmarked clades contain only octocoral CYP sequences. **d**, Zoomed-in branches of the distinct briarane BGC clades of cbCYPs, highlighted coral (cbCYPa), lavender (cbCYPb) and mustard (cbCYPc), labeled accordingly with octocoral species initials followed by enzyme type.

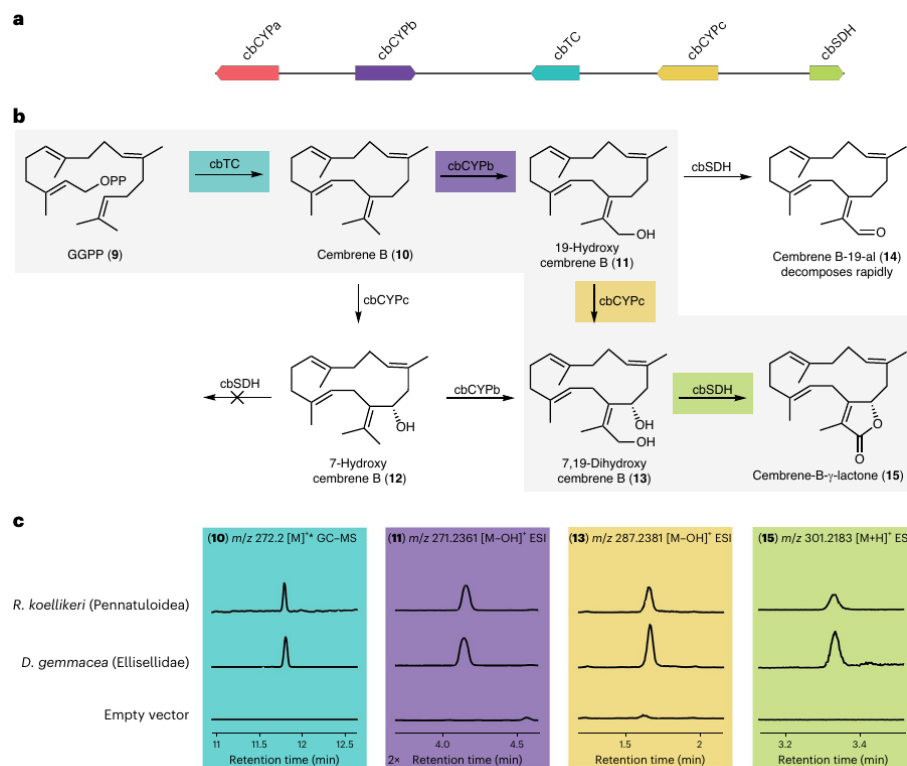


Fig. 4 | Biochemical validation of clustered biosynthesis genes across syntentic homologs. **a**, Annotation of the *R. koellikeri* briarane BGC. This prototypical example includes the terpene cyclase cbTC, the three different CYP450 subtypes cbCYPa, cbCYPb and cbCYPc and the short-chain dehydrogenase cbSDH. **b**, Scheme depicting the reactions tested and theorized overall pathway to cembrene B-γ-lactone **15**. This pathway is mediated stepwise by initial cyclization of **9** to **10** by cbTC, followed by mono-oxygenation to **11** by cbCYPb or to **12** by cbCYPc. Mono-oxygenation of **11** or **12** to **13** is followed by two consecutive dehydrogenations by cbSDH to γ-lactone **15**. Alternatively, a putative shunt product **14** can also be formed from **11** by cbSDH (Supplementary Fig. 14). **c**, Chromatographic data for main pathway enzymatic assays demonstrating

the depicted reactions. From left to right, GC-MS extracted ion chromatogram (EIC) trace at 272.1 m/z showing the formation of **10**, LC-MS EIC trace at 271.2 m/z ([M - OH]⁺ ion) showing **11**; LC-MS EIC trace at 287.2 m/z ([M - OH]⁺ ion) showing **13** and LC-MS EIC trace at 301.2 m/z ([M + H]⁺ ion) showing the formation of **15**. All traces were compared to the relevant purified reference compound and negative control (bottom trace). The negative control for the SDH assay was a reaction mixture without enzyme; for all others, the negative control was the crude extract of yeast harboring empty vector. The full four gene pathways from *R. koellikeri* and *D. gemmacea* were characterized along with additional representative homologs identified by synteny (Supplementary Figs. 4, 5, 7–9, 12 and 13). All tested enzyme homologs showed identical reactivity.

To test whether our phylogenetic approach could be used for genome mining in octocorals with lower-quality assemblies, we next included CYP sequences from the poor-contiguity genome assembly of *P. caledonicum*. We found corresponding sequences present in two of the three clades, thereby identifying cbCYPa and cbCYPb orthologs encoded in the genome (Fig. 3d).

Biochemical validation of the briarane BGC

To provide support that the discovered BGC is involved in the early stages of briarane biosynthesis, we set out to biochemically validate the clustered biosynthesis genes (Supplementary Table 7). We first tested whether all the homologous cbTC terpene synthases (TPSs) functioned as cembrene B cyclases as we originally established in *R. muelleri* and *E. caribaeorum*^{28,29}. We expressed each, six in total, in *Saccharomyces cerevisiae* and evaluated their production profiles by gas chromatography–mass spectrometry (GC–MS) analysis (Fig. 4 and Supplementary Fig. 4). In all cases, we confirmed that geranylgeranyl pyrophosphate (GGPP; **9**) was cyclized into cembrene B (**10**) as the sole product.

For the characterization of the tailoring enzymes, we tested the functions of all cluster genes from the *R. koellikeri* BGC and the full cluster genes of *D. gemmacea* to showcase the functions of homologs from different octocoral families. Additional enzymes from other identified clusters were also tested to show that association by synteny serves as functional prediction even with low sequence identity to their homolog in the *R. koellikeri* cluster (*C. rubrum* CYPb, 64.8%; *E. caribaeorum* CYPb, 62.9%; *E. caribaeorum* SDH, 46.6%) or when originating from a split gene cluster (*B. asbestinum* CYPb, *S. elongata* CYPc, *B. asbestinum* SDH and *S. elongata* SDH) (Figs. 3b and 4 and Extended Data Fig. 4). Despite these differences, all tested enzyme homologs showed identical reactivity to their *R. koellikeri* or *D. gemmacea* counterpart.

Because CYPs can oxidize unactivated carbon centers, we assumed that one of the three colocated CYP subtypes likely introduces the first functionalization. We evaluated their functions in pairwise fashion by coexpression with cognate cbTC genes using previously established methods (Fig. 4b)^{51–53}. We consistently measured the production of 19-hydroxycembrene B (**11**) in yeast upon coexpression with cbCYPb

homologs. Its formation proceeded in a regiospecific manner with no other detectable monooxygenated products (Supplementary Fig. 5 and Supplementary Note 3). Scale-up fermentation and chromatographic purification provided **11** in pure form, which we characterized by NMR spectroscopic analyses (Supplementary Tables 8 and 9, Supplementary Note 4 and Supplementary Fig. 6). When we instead coexpressed *cbCYPc* homologs with *cbTC* genes, we observed formation of 7-hydroxycembrene (**12**) (Fig. 4b), which could be detected by LC–MS (Supplementary Fig. 7) and characterized by NMR (Supplementary Tables 8 and 9, Supplementary Note 5 and Supplementary Fig. 6). Because *cbCYPb* and *cbCYPc* were found to oxidize complimentary positions on **10** with respect to the briarane pathway, we tested the *cbCYPb* product **11** by feeding it to yeast expressing *cbCYPc* homologs (Fig. 4b and Supplementary Fig. 8). We clearly observed a single oxidized product that, upon purification and combined spectroscopic (Supplementary Tables 8 and 9, Supplementary Note 6 and Supplementary Fig. 9) and X-ray crystallographic characterization (Supplementary Fig. 10 and Supplementary Table 10), was established as (7S)-7,19-dihydroxycembrene B (**13**). We also tested the opposite case by feeding **12** to yeast harboring *cbCYPb* from *E. caribaeorum* and similarly obtained **13** (Fig. 4b and Supplementary Fig. 9). When comparing the preparative biotransformation conversion rates of these two enzymes, *cbCYPc* converted nearly 50% of **11** to **13** (11.3 mg from 22 mg of **11**) while *cbCYPb* converted less than 5% of **12** to **13** (1.6 mg from 44 mg of **12**). The cooperative and related functions of the *cbCYPb* and *cbCYPc* enzymes correlates well with their shared evolution (Fig. 3c). Because of the low conversion rate that the tested *cbCYPb* showed with **12**, we propose that this enzyme most likely acts first on **10** followed by *cbCYPc* to produce **13** (Fig. 4b). Despite numerous attempts with different substrate and gene pairs, we have yet to observe a function for the third CYP, *cbCYPa*. It is a possibility of course that *cbCYPa* is nonfunctional in vivo, explaining our inability to successfully characterize it; however, we predict that it too is related to the biosynthesis of briarane diterpenes.

As a variety of cembrene double-bond isomers are produced by corals, we considered the possibility that the described CYP450 enzymes could act nonspecifically on numerous cembrene isomers. We, thus, expressed *cbCYPb* homologs with cembrene A or C as the substrate; however, no oxidized products were produced (Extended Data Fig. 5), showing high specificity of *cbCYPb* for cembrene B.

We next evaluated the colocalized *cbSDH* homologs for their possible involvement in further transforming the diol **13**. We expressed *cbSDHs* in *Escherichia coli* to conduct a series of in vitro incubation experiments (Supplementary Fig. 11). Incubation of purified *cbSDH* enzymes with nicotinamide adenine dinucleotide phosphate (NADP) and **11** produced aldehyde (**14**; Supplementary Fig. 12) but incubation with **12** yielded no detectable product. Notably, **14** could not be isolated from large-scale incubations because of its apparent instability but its identity was proven by LC–MS coelution with a **14** standard, obtained by chemical oxidation of **11** (Supplementary Fig. 12). Incubation with the diol **13** produced cembrene B γ -lactone (**15**), a structure feature diagnostic of all briarane diterpenoids (Fig. 4b and Supplementary Fig. 13). Two consecutive dehydrogenation steps were detected by liquid chromatography (LC)–MS through the formation of a product ion at m/z 301.2183 ($[M + H]^+$, calculated m/z 301.2163 for $C_{20}H_{30}O_2^+$). We scaled up the in vitro reaction of **13** and NADP with *cbSDH* to provide lactone **15**, which was purified and characterized spectroscopically. The structure was supported by NMR (Supplementary Tables 8 and 9, Supplementary Note 7 and Supplementary Fig. 6) and the diagnostic γ -lactone carbonyl stretch in the infrared (IR) spectrum ($\nu_{\max} = 1,757\text{ cm}^{-1}$). A proposed mechanism for the generation of **15** from **13** could be the enzymatic formation of an aldehyde in analogy to the oxidation of **11**, followed by the formation of a five-membered hemiacetal intermediate, which gets oxidized by *cbSDH* again (Supplementary Fig. 14). Experiments to test whether **14** can act as an intermediate by feeding it to yeast strains harboring *cbCYPb* or *cbCYPc* failed because of the instability of **14**.

Discussion

In summary, we discovered a broadly conserved BGC family in corals and biochemically validated its function in the early biosynthetic steps toward briarane diterpenoids. The distinctive family of bioactive molecules constitutes over 700 characterized members that are exclusively found in corals. Our work established the consistent presence of the **15**-forming core cluster in briarane-producing coral families. This includes *R. koellikeri*, which we establish as a briarane producer in this work that encodes the coclustered cembrene B synthase as the only TC gene on its genome. The biosynthetic logic of briarane's diagnostic γ -lactone core structure, including its conserved 7S stereocenter, is generated by four of the coclustered gene products, *cbTC*, *cbCYPb*, *cbCYPc* and *cbSDH*, confirming the clustered function of the briarane locus. The enzymatic transformations begin with cembrene B (**10**) formation by *cbTC* and then proceed with subsequent 19-hydroxylations and 7-hydroxylations carried out by *cbCYPb* and *cbCYPc*, respectively. Finally, *cbSDH* transforms diol **13** directly into lactone **15**.

We propose that *cbSDH* uses a unique mechanism of lactone biosynthesis (Fig. 5a), possibly through aldehyde formation at the 19 position, hemiacetal formation and a second dehydrogenation to produce the diagnostic lactone (Supplementary Fig. 14). Furthermore, while we have not yet characterized the enzyme(s) catalyzing bond formation between carbons 1 and 10 (ref. 30), we have two potential hypotheses for how this bond could be formed. A noncanonical class 2 TC acting on lactone **15** could introduce this feature to produce **16** (Fig. 5a)³⁴. Alternatively, a $\Delta^{1,14}$ epoxide intermediate **19**, potentially installed by a CYP, might precede a transannular cyclization reaction to **20** (Fig. 5b).

In animals, previous reports indicate that genes can be found in clusters when they result from successive tandem duplications or HGT. However, such factors are not at play here in the octocorals. Following the divergence of octocorals from hexacorals (>500 million years ago), genes linked to terpenoid biosynthesis evolved, as indicated by the widespread conservation of briarane CYP and TC gene homologs across Octocorallia. The briarane BGC itself is likely evolutionarily ancient and evolved soon after the division of the octocoral orders Scleractyonacea and Malacalcyonacea. It is only found in briarane-containing corals from Scleractyonacea and has similar synteny in distantly related corals such as *R. koellikeri* (Pennatulioidea) and *Erythropodium caribaeorum* (Erythropodiidae) (Fig. 5c)³⁵. A key evolutionary event in the formation of the briarane BGC was the duplication of the related *cbCYPb* and *cbCYPc* genes early in the evolution of Scleractyonacea. Supporting this timing, the *cbCYPb*–*cbCYPc* pair is found across the Scleractyonacea coral families Pennatulioidea, Ellisellidae, Coralliidae, Briareidae and Erythropodiidae but not elsewhere in sequenced octocorals. Subsequently, we speculate that *cbTC*, *cbSDH* and *cbCYPa* were recruited and incorporated adjacent to the duplicated *cbCYPb*–*cbCYPc* gene pair, also before speciation in the Scleractyonacea, through chromosomal rearrangements to form the BGC—a phenomenon not previously reported in metazoans. Following the formation of the briarane BGC, all or parts of it were lost in a group of octocorals, as no briaranes have been reported from several Scleractyonacea families (Fig. 5c). Supporting the gene loss scenario, *H. coerulea* (Helioporidae) does not harbor a cembrene B synthase gene in its genome but retains *cbCYPb* and *cbCYPc* relatives, suggesting that these genes were once part of a briarane BGC. More complex evolutionary scenarios are also observable in sequenced genomes, such as the BGCs from *S. elongata* and *B. asbestinum*, which contain additional *SDH* gene copies. Their appearance coincides with a split of the gene cluster into two parts. The observation of metazoan BGC formation and decay suggests potential mechanisms of chromosomal evolution not previously observed in animals.

In contrast to a well-preserved central briarane BGC, genes for specialized, late-stage tailoring processes are not colocalized. The core briarane skeleton is modified in many ways to create the >700 known briarane derivatives (Fig. 5a). For example, diterpenoids such as 11-hydroxyptilosarconone (**6**), gemmacolide R (**7**), bathyptilone

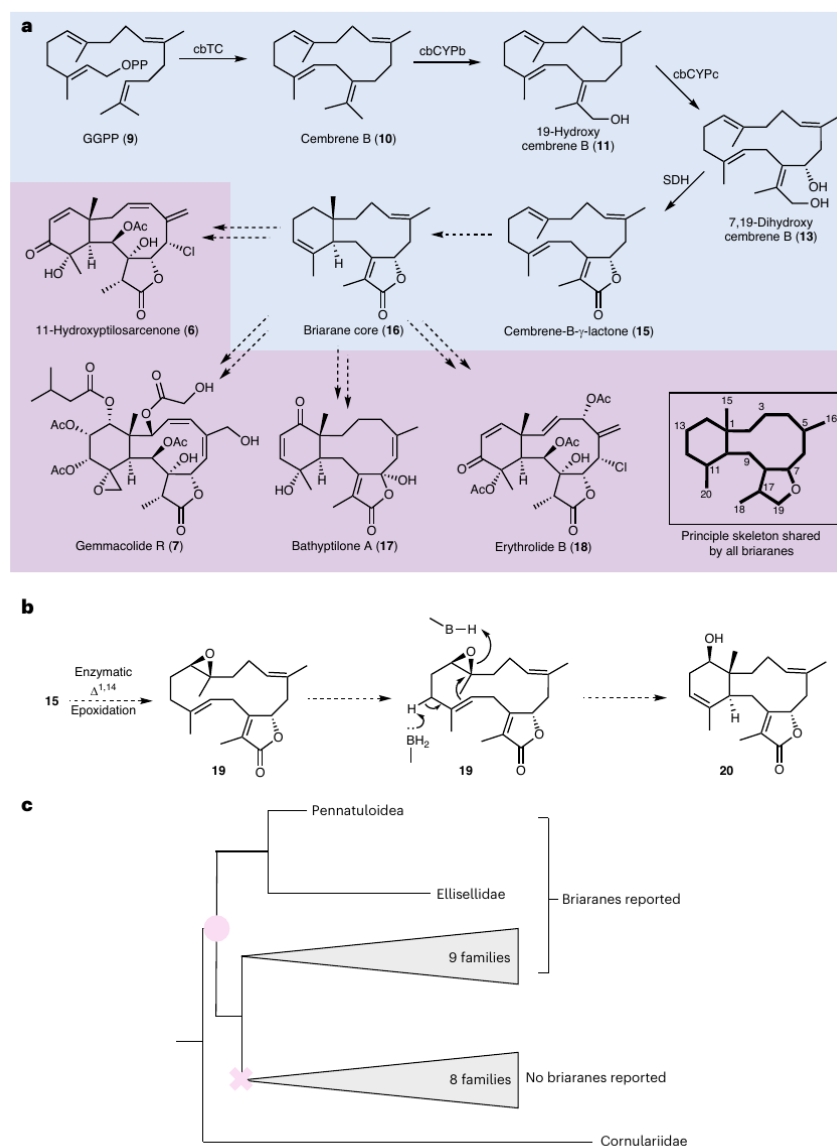


Fig. 5 | Proposed biosynthetic pathway for briarane production and estimated time points for the evolution of the briarane BGC. a, Clustered enzymes characterized in this work produce the lactone **15**, a central intermediate in briarane biosynthesis, by subsequent terpene cyclization, dual hydroxylation and two dehydrogenations. A missing central transformation is a 1,10-cyclization, the last fundamental step to obtain a compound containing the complete briarane scaffold (**16**). These steps must be conserved in all briarane producers (blue background). Other modifications of the briarane core are most likely introduced later and are lineage and species specific (purple background). These biosynthetic steps are evolutionarily younger and might not be present in gene clusters. **b,** A putative CYP-mediated 1,10-ring closing reaction, initiated

by epoxidation of **15** to **19**, which could be opened to produce alcohol **20** bearing the briarane scaffold. The resulting oxidized 14 position occurs as either oxygenated or olefinic in most briaranes. **c,** Simplified octocoral phylogeny, rooted at Scleractyonia. Estimated time scale of BGC evolution. The circular mark in purple denotes the latest acquisition time of the briarane BGC, given that Ellisellidae, Pennatuloidae, Briareidae and Erythropodiidae (members of the top nine families) are shown to contain the briarane BGC and are known briarane producers. The purple-colored X denotes a likely BGC deletion event, in which the whole gene cluster or parts of it were lost. None of the eight families in this clade are reported to produce briaranes and the only genome-sequenced species, *H. coerulea*, does not harbor a cembrene B synthase.

A (17) and erythrolide B (18), found in different coral species, have been modified by diverse enzymatic reactions, including oxidation, chlorination and acylation, by genes found in different chromosomal locations. This may represent a general rule in metazoan BGCs, in which the required core genes are found in the BGC, while those for variable and species-specific tailoring steps are distributed through the genomes.

While the innate functions of coral terpenoids are largely unknown outside of a few examples having feeding-deterrent properties³², we suspect that the conservation of their BGCs indicates central, unrealized biological roles pertaining to some of the more common coral terpenoid classes³⁶. With the characterization of a coral terpenoid BGC, future genetic experiments are positioned to address their function once genome-editing methods have been established in octocorals, as recently accomplished in the hexacoral *Acropora millepora*⁵⁷.

The stage is set for genome-mining and bioengineering studies to delineate biosynthetic cascade reactions to rare, bioactive coral molecules for their biosustainable production and application as therapeutic compounds. The briarane gene cluster family is likely the first of many to be described. We previously reported that some corals contain 70-kb and 500-kb putative BGCs with upward of 13 terpenoid biosynthesis genes^{28,29}. With over 4,000 terpenoids characterized from corals (determined by searching the CMNPD database)⁵⁸, gene clustering is likely common in this ancient phylum. Looking ahead, we have yet to see whether the active assembly of BGCs by evolutionary forces is an exceedingly rare event that is limited to octocorals or whether this is a more common phenomenon in animals and eukaryotes in general that has been overlooked so far because of limited genomic information in the eukaryotic kingdom.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-025-01927-y>.

References

1. Medema, M. H., de Rond, T. & Moore, B. S. Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.* **22**, 553–571 (2021).
2. Torres, J. P. & Schmidt, E. W. The biosynthetic diversity of the animal world. *J. Biol. Chem.* **294**, 17684–17692 (2019).
3. Poth, D., Wollenberg, K. C., Vences, M. & Schulz, S. Volatile amphibian pheromones: macrolides from mantellid frogs from Madagascar. *Angew. Chem. Int. Ed. Engl.* **51**, 2187–2190 (2012).
4. Chinta, S. P. et al. The sex pheromone of the wasp spider *Argiope bruennichi*. *Angew. Chem. Int. Ed. Engl.* **49**, 2033–2036 (2010).
5. Masello, J. F., Lubjuhn, T. & Quillfeldt, P. Is the structural and psittacofulvin-based coloration of wild burrowing parrots *Cyanoliseus patagonus* condition dependent? *J. Avian Biol.* **39**, 653–662 (2008).
6. Anderson, H. A., Mathieson, J. W. & Thomson, R. H. Distribution of spinochrome pigments in echinoids. *Comp. Biochem. Physiol.* **28**, 333–345 (1969).
7. Butcher, R. A., Fujita, M., Schroeder, F. C. & Clardy, J. Small-molecule pheromones that control dauer development in *Caenorhabditis elegans*. *Nat. Chem. Biol.* **3**, 420–422 (2007).
8. Dettner, K. & Schwinger, G. Defensive secretions of three oxytelinae rove beetles (Coleoptera: Staphylinidae). *J. Chem. Ecol.* **8**, 1411–1420 (1982).
9. Olivera, B. M. et al. Peptide neurotoxins from fish-hunting cone snails. *Science* **230**, 1338–1343 (1985).
10. Chevrete, M. G. et al. Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat. Prod. Rep.* **37**, 566–599 (2020).
11. Rokas, A., Mead, M. E., Steenwyk, J. L., Raja, H. A. & Oberlies, N. H. Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Nat. Prod. Rep.* **37**, 868–878 (2020).
12. Nützmann, H.-W., Scazzocchio, C. & Osbourn, A. Metabolic gene clusters in eukaryotes. *Annu. Rev. Genet.* **52**, 159–183 (2018).
13. Steele, T. S. et al. Domoic acid biosynthesis in the red alga *Chondria armata* suggests a complex evolutionary history for toxin production. *Proc. Natl Acad. Sci. USA* **119**, e2117407119 (2022).
14. Smit, S. J. & Lichman, B. R. Plant biosynthetic gene clusters in the context of metabolic evolution. *Nat. Prod. Rep.* **39**, 1465–1482 (2022).
15. Ryngajto, M., Boruta, T. & Bizukojć, M. Complete genome sequence of lovastatin producer *Aspergillus terreus* ATCC 20542 and evaluation of genomic diversity among *A. terreus* strains. *Appl. Microbiol. Biotechnol.* **105**, 1615–1627 (2021).
16. Li, Y. et al. Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat. Commun.* **12**, 2563 (2021).
17. Itkin, M. et al. Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* **341**, 175–179 (2013).
18. Nützmann, H.-W., Huang, A. & Osbourn, A. Plant metabolic clusters—from genetics to genomics. *N. Phytol.* **211**, 771–789 (2016).
19. Osborn, A. R. et al. De novo synthesis of a sunscreen compound in vertebrates. *eLife* **4**, e05919 (2015).
20. Roelofs, D. et al. A functional isopenicillin N synthase in an animal genome. *Mol. Biol. Evol.* **30**, 541–548 (2013).
21. Faddeeva-Vakhrusheva, A. et al. Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia candida*. *BMC Genom.* **18**, 493 (2017).
22. Feng, L., Gordon, M. T., Liu, Y., Basso, K. B. & Butcher, R. A. Mapping the biosynthetic pathway of a hybrid polyketide–nonribosomal peptide in a metazoan. *Nat. Commun.* **12**, 4912 (2021).
23. de Rosa, R. et al. Hox genes in brachiopods and priapulids and protostome evolution. *Nature* **399**, 772–776 (1999).
24. Storz, J. F., Opazo, J. C. & Hoffmann, F. G. Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Mol. Phylogenet. Evol.* **66**, 469–478 (2013).
25. Kitchen, S. A. et al. The genomic and cellular basis of biosynthetic innovation in rove beetles. *Cell* **187**, 3563–3584 (2024).
26. Turcu, A. F., Rege, J., Auchus, R. J. & Rainey, W. E. 11-Oxygenated androgens in health and disease. *Nat. Rev. Endocrinol.* **16**, 284–296 (2020).
27. Tan, D.-X. et al. On the significance of an alternate pathway of melatonin synthesis via 5-methoxytryptamine: comparisons across species. *J. Pineal Res.* **61**, 27–40 (2016).
28. Burkhardt, I., de Rond, T., Chen, P.-Y. T. & Moore, B. S. Ancient plant-like terpene biosynthesis in corals. *Nat. Chem. Biol.* **18**, 664–669 (2022).
29. Scesa, P. D., Lin, Z. & Schmidt, E. W. Ancient defensive terpene biosynthetic gene clusters in the soft corals. *Nat. Chem. Biol.* **18**, 659–663 (2022).
30. Moon, N. G. & Harned, A. M. Synthetic explorations of the briarane jungle: progress in developing a synthetic route to a common family of diterpenoid natural products. *R. Soc. Open Sci.* **5**, 172280 (2018).
31. La, M.-P. et al. Briarane diterpenoids from the gorgonian *Dichotella gemmacea*. *Mar. Drugs* **12**, 6178–6189 (2014).
32. Barsby, T. & Kubanek, J. Isolation and structure elucidation of feeding deterrent diterpenoids from the sea pansy, *Renilla reniformis*. *J. Nat. Prod.* **68**, 511–516 (2005).

33. Fenical, W. & Pawlik, J. R. Defensive properties of secondary metabolites from the Caribbean gorgonian coral *Erythropodium caribaeorum*. *Mar. Ecol. Prog. Ser.* **75**, 1–8 (1991).
34. Harvell, C. D. et al. Local and geographic variation in the defensive chemistry of a West Indian gorgonian coral (*Briareum asbestinum*). *Mar. Ecol. Prog. Ser.* **93**, 165–173 (1993).
35. Ksebati, M. B. & Schmitz, F. J. Diterpenes from a soft coral, *Minabea* sp., from Truk Lagoon. *Bull. Soc. Chim. Belg.* **95**, 835–851 (1986).
36. Wratten, S. J. & Faulkner, D. J. Some diterpenes from the sea pen *Stylatula* sp. *Tetrahedron* **35**, 1907–1912 (1979).
37. Yang, J. et al. Briarane-type diterpenoids from the China gorgonian coral *Subergorgia reticulata*. *Biochem. Syst. Ecol.* **35**, 770–773 (2007).
38. Williams, G. C. Living genera of sea pens (Coelenterata: Octocorallia: Pennatulacea): illustrated key and synopses. *Zool. J. Linn. Soc.* **113**, 93–140 (1995).
39. Keifer, P. A., Rinehart, K. L. & Hooper, I. R. Renilla-foulins, antifouling diterpenes from the sea pansy *Renilla reniformis* (Octocorallia). *J. Org. Chem.* **51**, 4450–4454 (1986).
40. Hendrickson, R. L. & Cardellina, J. H. Structure and stereochemistry of insecticidal diterpenes from the sea pen. *Tetrahedron* **42**, 6565–6570 (1986).
41. Nurco, D. J., Conklin, D. E., Shapiro, N. S. & Tran, E. The absolute structure of ptilosarcenone 2.5-hydrate, a diterpenoid briarane from the orange sea pen *Ptilosarcus gurneyi* (Gray). *Acta Crystallogr. E Struct. Rep. Online* **67**, o181–o182 (2010).
42. Li, C. et al. Bioactive (3Z,5E)-11,20-epoxybriara-3,5-dien-7,18-olide diterpenoids from the South China Sea gorgonian *Dichotella gemmacea*. *Mar. Drugs* **9**, 1403–1418 (2011).
43. Pordesimo, E. O., Schmitz, F. J., Ciereszko, L. S., Hossain, M. B. & Van der Helm, D. New briarane diterpenes from the Caribbean gorgonians *Erythropodium caribaeorum* and *Briareum* sp. *J. Org. Chem.* **56**, 2344–2357 (1991).
44. Molina, S. L. et al. Metabolic profiling of the soft coral *Erythropodium caribaeorum* (Alcyonacea: Anthothelidae) from the Colombian Caribbean reveals different chemotypes. *Mar. Drugs* **18**, 4 (2020).
45. Jiang, J. B. et al. A hybrid de novo assembly of the sea pansy (*Renilla muelleri*) genome. *GigaScience* **8**, giz026 (2019).
46. Ledoux, J.-B. et al. Chromosome-level genome assembly and annotation of *Corallium rubrum*: a Mediterranean coral threatened by overharvesting and climate change. *Genome Biol. Evol.* **17**, evae253 (2025).
47. Corominas, M. et al. The Catalan initiative for the Earth BioGenome Project: contributing local data to global biodiversity genomics. *NAR Genom. Bioinform.* **6**, lqae075 (2024).
48. Hu, M., Zheng, X., Fan, C.-M. & Zheng, Y. Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia*. *Nature* **582**, 534–538 (2020).
49. Ip, J. C.-H., Ho, M.-H., Chan, B. K. K. & Qiu, J.-W. A draft genome assembly of reef-building octocoral *Heliopora coerulea*. *Sci. Data* **10**, 381 (2023).
50. Pankov, K. V. et al. The cytochrome P450 (CYP) superfamily in cnidarians. *Sci. Rep.* **11**, 9834 (2021).
51. DeJong, J. M. et al. Genetic engineering of taxol biosynthetic genes in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **93**, 212–224 (2006).
52. Pompon, D., Louerat, B., Bronine, A. and Urban, P. in *Methods in Enzymology* Vol. 272 (eds Johnson, E. F. & Waterman, M. R.) 51–64 (Academic, 1996).
53. Ro, D.-K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
54. Rudolf, J. D. & Chang, C.-Y. Terpene synthases in disguise: enzymology, structure, and opportunities of non-canonical terpene synthases. *Nat. Prod. Rep.* **37**, 425–463 (2020).
55. McFadden, C. S., van Ofwegen, L. P. & Quattrini, A. M. Revisionary systematics of Octocorallia (Cnidaria: Anthozoa) guided by phylogenomics. *Bull. Soc. Syst. Biol.* **1**, 8735 (2022).
56. He, J., Xin, P., Ma, X., Chu, J. & Wang, G. Gibberellin metabolism in flowering plants: an update and perspectives. *Front. Plant Sci.* **11**, 532 (2020).
57. Cleves, P. A., Strader, M. E., Bay, L. K., Pringle, J. R. & Matz, M. V. CRISPR/Cas9-mediated genome editing in a reef-building coral. *Proc. Natl Acad. Sci. USA* **115**, 5235–5240 (2018).
58. Lyu, C. et al. CMNPD: a comprehensive marine natural products database towards facilitating drug discovery from the ocean. *Nucleic Acids Res.* **49**, D509–D515 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

Methods

Sample collection

Divers collected *R. koellikeri* and *S. elongata* (both in the Pennatulioidea superfamily) from La Jolla, CA, *B. asbestinum* (Briareidae) and *E. caribaeorum* (Erythropodiidae) coral from Broward County, FL, and *C. rubrum* (Coralliidae) coral from Cap Castel, Illes Medes, Catalunya. We also obtained a *D. gemmacea* (Ellisellidae) specimen from Two Little Fishies. All animals were collected by hand in accordance with federal and state laws with the required license using noninvasive methods (sampling whole individuals in the cases of *R. koellikeri*, *D. gemmacea* and *S. elongata*; sampling small portions of larger colonies using a knife for the *B. asbestinum*, *C. rubrum* and *E. caribaeorum*). We transported either tissues or whole organisms back to the lab. The samples were washed three times for 5 min each with Ca²⁺-free and Mg²⁺-free artificial seawater within 24 h of collection and then flash-frozen with liquid nitrogen as live organisms. At the time of flash-freezing, aliquots of each organism were preserved in RNAlater. All biologic material was and kept at -80 °C for less than 2 weeks until DNA extraction. In the case of *S. elongata*, taxonomy was uncertain; thus, an aliquot of tissue from the base of the polyp leaf was collected, dissolved in bleach and analyzed under a microscope. Given the absence of slender rods with hollow centers that those characterized by *Stylatula*²⁴, with help from C. S. McFadden (Harvey Mudd College), we were able to discern with certainty that the organism was *S. elongata* (Extended Data Fig. 1)²⁸.

High-molecular-weight DNA extraction and quantification

High-molecular-weight (HMW) DNA was isolated from tissue from *R. koellikeri*, *S. elongata*, *B. asbestinum* and *D. gemmacea* as previously described²⁹. In brief, tissue was homogenized by a mortar and pestle while frozen with liquid nitrogen. The Qiagen genomic tip kit 500/G9 (10262) was used, following the protocol for the 'preparation of tissue samples' with the recommendations of tissue amounts for mouse tail. DNA purity was determined using a BioDrop μ LITE (BioDrop), confirming that the optical density ratio at 260 to 280 nm ($OD_{260/280}$) was 1.8 and $OD_{260/230}$ was 2.0–2.2. DNA concentrations were assessed using the double-stranded DNA high-sensitivity assay on a Qubit fluorometer version 3.0 (Thermo Fisher Scientific). DNA size distribution was assessed with a TapeStation (Agilent 4150). On the basis of the degree of purity and size distribution of the DNA, samples were either purified and size-selected further using the Oxford Nanopore Technology (ONT) short fragment eliminator kit (EXP-SFE001) and Beckman-Coulter AMPure RNA clean XP beads (A63987) or directly used to build sequencing libraries. Depending on the concentration and size distribution after quality control, samples were used to prepare libraries for Pacific Biosciences (PacBio) or ONT sequencing.

For all four organisms, an aliquot of HMW DNA was also used to obtain Illumina short-read data.

E. caribaeorum DNA extraction and sequencing

HMW DNA was extracted from frozen, RNAlater-preserved tissue using chloroform–phenol extraction as previously described²⁹. Crude DNA precipitate from chloroform–phenol extraction was dissolved in nuclease-free water (200 μ l) and loaded onto an equilibrated Qiagen genomic tip (20/G, 10223) column. The column was washed and eluted and the DNA was precipitated from the eluent according to the genomic tip kit protocol. DNA quality and quantity were assessed by Nanodrop and gel electrophoresis. Illumina reads were previously obtained and are publicly available for this sample (Supplementary Table 4)²⁹.

RNA extraction

RNA was extracted from the tissues of *R. koellikeri*, *S. elongata* and *D. gemmacea* preserved in RNAlater as previously described²⁹. In brief, the RNeasy plant mini kit was used (Qiagen, 74904), including the optional step of an on-column DNase digestion. Quality and quantity were both assessed using Nanodrop spectrophotometric and

Qubit fluorometric methods as described above. For *E. caribaeorum* and *B. asbestinum*, RNA was extracted using methods previously described²⁹.

Sequencing methods

The *R. koellikeri* and *S. elongata* genomes were sequenced with PacBio. PacBio Sequel IIe running SMRT Link version 11.1 with SMARTbell prep kit 3.0 (102-182-700) produced 28.7 Gb with a mean length of 11,734 bp and 85.4% HiFi reads ($Q > 20$) for *R. koellikeri* and 26.2 Gb with a mean length of 12,845 bp and 77.9% HiFi reads for *S. elongata*. Because of lower yields and quality of DNA, *D. gemmacea* and *B. asbestinum* genomes were sequenced with ONT. Following library prep with ligation chemistry (SQK-LSK114), an ONT Promethion running MinKNOW version 23.04.6 produced 77.8 Gb for *D. gemmacea* and 59.4 Gb for *B. asbestinum*, quantified by NanoPlot version 1.40.0. *E. caribaeorum* HMW DNA (3 mg) was used to prepare an ONT library using the ligation sequencing kit (SQK-LSK114) according to the manufacturer's protocol with size selection using long fragment buffer from the kit. The library was loaded onto a Promethion flow cell (FLO-PRO114M) and data were collected using MinKNOW (version 24.02.10) with real-time, high-accuracy base calling. The run yielded 17.46 Gb of called bases with an N50 of 3.09 kb.

We generated short reads for genome size estimates and assembly polishing using an Illumina NovaSeq6000 with an S4 300 cycle run configuration, following Illumina DNA library prep (20060059), which produced 150 million reads for *R. koellikeri*, 206 million reads for *S. elongata*, 146 million reads for *D. gemmacea* and 114 million reads for *B. asbestinum*. For *R. koellikeri*, *S. elongata* and *D. gemmacea*, the same Illumina run yielded RNA sequencing data for transcriptome assembly and genome annotations. Complementary DNA (cDNA) libraries were prepared from total RNA with an Illumina TruSeq stranded total RNA kit (20020599) with ribosomal RNA depletion steps, barcoded with Integrated DNA Technologies TruSeq RNA indices (UDI-20022371). This yielded 25.8 million reads for *R. koellikeri* peduncle tissue, 54.8 million reads for *R. koellikeri* body tissue, 68.8 million reads for *S. elongata* and 47.7 million reads for *D. gemmacea*. Given insufficient depth for TPS transcript mapping across several samples, an additional sequencing run with an S2 300 cycle configuration yielded 114 million reads for *R. koellikeri* peduncle tissue, 64 million reads for *R. koellikeri* body tissue, 226 million reads for *S. elongata* and 618 million reads for *D. gemmacea*. In the case of *E. caribaeorum*, genomic short-read sequencing data from the same individual were generated in a previous study and are already publicly available (Supplementary Table 4)²⁹.

Genome size estimation

Genome size was estimated by k -mer frequency distribution analysis (genome size = k -mer number/peak depth). First, the short reads were filtered using fastp (versions 0.19.4 and 23.2)⁶⁰ with default parameters and quality checked with FastQC (version 0.11.9). Subsequently, the k -mers ($k = 17$ bp) were counted using Jellyfish (version 2.2.10)⁶¹ with the parameters '-C-m 51-s 10000000000-t 50'. The output histogram file was then analyzed with the following equation: sum of the k -mer number $\times k$ -mer count - first peak/ k -mer peak. This equation assumes that all k -mers that occurred exactly once in the read set are erroneous and should be ignored, computes the depth of nonerroneous k -mers and estimates diploid and haploid genome sizes using the two different peak coverages. GenomeScope was also used to estimate the genome size with default parameters⁶². The frequency distribution of 17-mers is based on the genome characteristics and pattern of Poisson distribution (Supplementary Fig. 2 and Extended Data Table 2).

Genome assembly

NanoPlot (version 1.41.0)⁶³ was used to determine the quality of the sequencing data obtained for all five organisms.

ONT. While we used the entirety of ONT-generated reads for *B. asbestinum* and *E. caribaeorum*, in the case of *D. gemmacea*, seqtk (version 1.3) was used to remove reads that were shorter than 10 kb. We used flye (version 2.9.1)⁶⁴ with default parameters to assemble the genomes of *B. asbestinum*, *E. caribaeorum* and *D. gemmacea*. We then used Minimap2 (version 2.24)⁶⁵ to align the long reads to the assembly and, with these aligned reads, we generated a consensus assembly running racon (version 1.5.0)⁶⁶ with long reads three times. We next trimmed and removed adaptor sequences from Illumina short reads. FastQC was used to assess quality and fastp was used for cleaning. Then, we used Minimap2 to map the short reads against the consensus assembly. Lastly, we used pilon (version 1.24)⁶⁷ to polish the ONT assemblies with our mapped short-read data.

PacBio. *R. koellikeri* and *S. elongata* were assembled using hifiiasm (version 0.19.5)⁶⁸ with default parameters; however, in the case of *R. koellikeri*, hifiiasm incorrectly calculated the homozygous coverage peak. The hom-cov parameter was adjusted to 165, where the homozygous peak appeared in the *k*-mer-based plot (Supplementary Fig. 2). All assembled genomes were checked for duplication (D%) rate and completeness (C%) with BUSCO (version 5.4.7)⁶⁹ using a metazoan database as reference. Assembly contiguity was determined by QUAST (version 5.2.0)⁷⁰.

Transcriptome assembly

The trimmed and adaptor-removed (FastQC and FastP with default parameters) cDNA sequences from *R. koellikeri*, *S. elongata*, *B. asbestinum*, and *D. gemmacea* were assembled with trinity (version 2.8.5)⁷¹ using the stranded parameter (`--SS_lib_type RF`). The *E. caribaeorum* transcriptome was assembled as described previously using Trimmomatic⁷² for adaptor removal and SPAdes⁷³ for assembly⁷⁹. The processed cDNA sequences were also mapped to the assembled genomes using HiSat2⁷⁴ (Extended Data Table 2). A poor alignment score for *B. asbestinum* (40.36%) hints at potential contamination or poor extraction of RNA from the sample; however, the alignment was sufficient for our genome-mining pipeline.

Syntenic analysis

We found the terpene cyclases in our genomic and transcriptomic assemblies using traditional basic local alignment search tool (BLAST, version 2.6.0) searches with the *R. muelleri* cembrene B cyclase previously characterized by Burkhardt et al.²⁸ We then used the Conserved Domain Database (National Institutes of Health (NIH) web-based search tool)⁷⁵ to annotate other functional proteins in the genomic neighborhood of the terpene cyclases. We determined intron bounds of all genes with their hits in the respective transcriptomes. In the case that the gene of interest was not transcriptionally active and could not be found by BLASTing into the transcriptome, we used fgenesh, a hidden Markov model (HMM)-based gene structure prediction tool in the softberry suite⁷⁶, with *N. vectensis* as a protein model. We further validated the predicted genes with hmmscan⁷⁷ against the profile HMM database to confirm the presence of the correct conserved motifs (Supplementary Fig. 15). We subsequently annotated the contigs in Benchling and exported them as a GenBank file. To simplify the syntenic analysis, we removed all introns from gene-coding regions and shortened intergenic regions. The nonmodified BGCs are depicted in Supplementary Fig. 4. We then used Clinker (version 0.26)⁷⁸ to calculate the sequence similarity and visualize the genetic synteny (synteny in Fig. 3b; percentage identity matrices in Extended Data Fig. 3).

Phylogenetic analysis

For phylogenetic analysis of the CYPs, we followed the method outlined by Pankov et al.⁵⁰ for genome-wide annotation of CYP genes. In short, we used TransDecoder (version 5.7.1) to identify candidate coding regions within transcript sequences of *R. koellikeri*, *S. elongata*, *B. asbestinum*, *D. gemmacea*, *H. coerulea*, *C. rubrum* and *E. caribaeorum*. BLAST⁷⁹ was

used to identify local alignments between the coding regions of the seven octocoral and a query that consisted of all annotated CYPs from *H. vulgaris*, *A. digitifera*, *A. aurita* and *N. vectensis* in a BLASTp search (protein query against a proteome). Given that not all CYPs are transcriptionally active and cDNA sequencing only offers a snapshot of an organism's biosynthetic potential, we used tBLASTn (protein query against a genome nucleotide) searches to identify CYP alignments between the original cnidarian CYP list amended with the CYPs discovered in the seven octocoral transcriptomes and the respective genomes of those octocoral (list of CYPs included and sources in Supplementary Table 6). To pull out the scoring pairs from the genomic blast searches, we used BEDTools (version 2.26.0)⁸⁰. Furthermore, we used transeq (version 5.7.1) to translate the nucleotide hits in the genomes. To ensure that all protein hits were in fact CYPs, we used hmmscan to ensure they had the correct motifs. Sequences with alignment scores and coverage of 80% or lower and percentage identity of 20% or lower were discarded. We then used cd-hit (version 4.8.1)⁸¹ with a 99% cutoff to remove duplicates.

To pull out candidate *cbCYPa*, *cbCYPb* and *cbCYPc* sequences from the poor-contiguity genome of the *P. caledonicum*, we used BLAST with the homologous *R. muelleri* sequences. Finally, we used Kalign (version 2.04)⁸² to generate alignments of the CYP amino acid sequences found in the octocoral genomes and transcriptomes and the *H. vulgaris*, *A. digitifera*, *A. aurita* and *N. vectensis* CYPs annotated by Pankov et al.⁵⁰ and the putative briarane biosynthesis CYPs in *P. caledonicum*. The phylogenetic analysis was performed using IQ-TREE multicore (version 2.0.3) substitution model Q.pfam+F + R10 (determined by ModelFinder as included in IQ-TREE)⁸³ and 1,000 bootstrap replicates. The resulting tree was visualized with iTOL (version 6.9.1)⁸⁴ (Fig. 3c and Supplementary Fig. 3).

Chemical isolation and elucidation of the *R. koellikeri* briarane diterpene

R. koellikeri (approximately 1 kg wet weight) was extracted with methanol–dichloromethane 2:1 (600 ml) overnight three times with stirring. The solvent was evaporated and the residue was partitioned between water and hexane (100 ml each) and then water and ethyl acetate (100 ml each). The hexane, ethyl acetate and aqueous fractions were all analyzed by LC–MS for the presence of chlorinated compounds. While the organic fractions both had chlorinated compounds, the ethyl acetate fraction was much less complex and was used for the following stages of isolation and purification. After evaporating the solvent, 1.30 g of crude extract remained. The residue was then resuspended in 100 ml of 40% methanol and loaded onto a 20-g C₁₈ (reversed-phase silica) column and fractionated with a gradient of water–methanol from 40% to 100% stepwise with 20 ml per fraction. Fractions were analyzed by LC–MS and those that contained chlorinated compounds with the fragmentation patterns of terpenes in the retention time window matching the expected polarity of a functionalized briarane were pooled. Further purifications were carried out with an Agilent Technologies 1260 Infinity series HPLC equipped with a degasser, quaternary pump, autosamplers, diode array and fraction collector. A Kinetex 5-μm C18 (100 Å, 250 × 10.0 mm) column was used with an isocratic hold of 47% acetonitrile–water at a flow rate of 2 ml min⁻¹. A semipure fraction containing a diterpenoid was then characterized by NMR (600-MHz Bruker Avance III spectrometer using a 1.7-mm inverse-detection triple-resonance (H-C/N/D) cryoprobe and 500-MHz JEOL ECZ spectrometer using a 3-mm probe) (Supplementary Note 2, Supplementary Table 2 and Supplementary Fig. 1).

NMR analysis of *R. koellikeri* compound

We identified 11-hydroxyptilosarcenone (**6**) as the major terpenoid present in *R. koellikeri* tissue (Supplementary Note 2, Supplementary Table 2 and Supplementary Fig. 1). The identity of this compound, which was previously isolated from the sea pen *Ptilosarcus gurneyi* but

not from *Renilla* species, was supported by high-resolution MS and NMR data. We collected full one-dimensional and two-dimensional (2D) NMR datasets on the mixture, including 2D correlated spectroscopy (COSY), nuclear Overhauser effect spectroscopy, heteronuclear single quantum coherence spectroscopy (HSQC) and heteronuclear multiple bond correlation (HMBC) spectra, in CDCl₃ to determine the main compound as **6**, representing 75% of the mixture. We found inconsistencies in the reported positional assignments of ¹H-NMR and ¹³C-NMR chemical shifts. The structure and chemical shifts of **6** were assigned using the 2D data as follows: COSY correlations between H-13 and H-14 along with an HMBC correlation from H-14 to C-12 indicated the presence of an enone fragment. HMBC correlations from H₃-15 to C-1, C-2, C-10 and C-14 along with HMBC correlations from H₃-20 to C-10, C-11 and C-12 confirmed the presence of the cyclohexanone fragment and the positions of the angular methyl and tertiary alcohol. An HMBC correlation from H-2 to an acetate carbonyl (δ_c 169.3) and a COSY correlation from H-2 to H-3 and from H-3 to H-4 established the position of the allylic acetate, which in turn was connected to the 1,1-disubstituted olefin according to an HMBC correlation from H-16_a and H-16_b to C-4. HMBC correlations from H-16_a and H-16_b to the conspicuous C-6 (δ_H 5.05 and δ_c 60.9 in the HSQC) suggested the likely position of a chloromethine. In turn, H-6 showed a COSY correlation to H-7 of the butyrlactone oxymethine (δ_H 4.98 and δ_c 78.8 in the HSQC). A COSY correlation between H₃-18 and H-17 along with HMBC correlations from H₃-18 to C-8, C-17 and C-19 allowed full assignment of the methylated butyrlactone. A COSY correlation between the H-9 acetoxymethine (δ_H 5.81 and δ_c 69.4 in the HSQC) and H-10 suggested the closure of the decane ring. The decane ring closure and lactone position were confirmed by a wealth of HMBC correlations from H-9. This included an HMBC correlation from H-9 to the acetate carbonyl (δ_c 170.0) and to C-7, C-8, C-10, C-11 and C-17 (full NMR data and 2D correlations in Supplementary Note 2 and Supplementary Fig. 1). This led us to corroborate the structure of **6** as previously described, with the newly assigned positions of NMR chemical shifts presented in Supplementary Table 2.

Crude chemical extraction and molecular networking

D. gemmacea, *C. rubrum*, *E. caribaeorum*, *S. elongata*, *B. asbestinum*, and *R. koellikeri* were used as samples for this work. Approximately, 1 g of each frozen coral sample was crushed and extracted overnight in 2 ml of a 2:1 methanol–dichloromethane mixture at room temperature while stirring. For all corals, we sampled tissues across different parts of the organism and homogenized them for the extraction. All extracts were first passed through a cotton-filled Pasteur pipette to remove large remaining fragments and then evaporated at 40 °C using a rotary evaporator. The remaining residue was dissolved in methanol (HPLC-grade, Fisher Scientific) and prepared for LC–MS analysis by filtering through a Mini Spin filter column (American Chromatography Supplies). The semipurified 11-hydroxyptilosarcenone (**6**) from *R. koellikeri*, in addition to the tissue from **6** crude extract samples were analyzed by LC–MS method A. The obtained LC–MS/MS data were converted to .mzML format using msConvert (ProteoWizard)⁸⁵ and analyzed using MzMine (version 3.9.0)⁸⁶. The feature-based molecular network parameters used and raw data are available through the MASSESIVE dataset (MSV000094792). The processed spectra were uploaded to GNPS⁸⁷ and visualized with cytoscape⁸⁸.

LC–MS method A. LC–MS/MS analysis was conducted on a 6530 Accurate-Mass quadrupole time-of-flight (Q-TOF) mass spectrometer (Agilent Technologies) coupled to the 1290 Infinity leveled-up 1260 Infinity LC system (Agilent Technologies) with a Kinetex LC column (150 × 4.6 mm, 5 μ m, C18 100 Å; Phenomenex). Dual electrospray ionization (ESI) was used as the ion source in positive ionization mode with a carrier gas temperature of 300 °C and the ionization voltage cap set to 3,000 V. The applied collision energy voltages can be described by the following formula $2.6 \times m/z + 14.75$. The LC gradient consisted of a 2-min

isocratic period at 20% acetonitrile in water at 1.0 ml min⁻¹ flow rate followed by a linear increase to 95% acetonitrile after 18 min total, an additional linear increase from 95% to 100% acetonitrile after 2 min total and then a 3-min isocratic period at 100% acetonitrile (25 min total).

Plasmids, strains and media

Candidate terpene cyclase genes were codon-optimized for yeast and purchased from Twist Biosciences in the pESC-leu2d vector backbone (Addgene, 20120) along with a copy of the GGPP synthase gene *XdCrtE*^{53,89,90}. Candidate CYP genes codon-optimized for yeast were purchased from Twist Biosciences in the pESC-URA vector backbone (Agilent Technologies) along with a copy of the *H. coelestis* CPR gene (*HcCPR*). Candidate SDH genes were purchased from Twist Biosciences in the pET-28b+ vector with N-terminal His-tags at the NdeI/XhoI site. Haploid *S. cerevisiae* YPH499 and YPH500 were purchased from Agilent Technologies (included with pESC-URA plasmid as a kit). Plasmids were transformed into yeast cells using a lithium acetate method and into *E. coli* BL21(DE3) using chemical transformation⁹¹. Synthetic complete (SC) medium⁹² was prepared using yeast nitrogen base (Gibco) and SC medium supplement (Aldrich). The appropriate nutrient (uracil and/or leucine) was omitted from the recipe for auxotrophic selection. SC liquid production medium was made in a similar manner but containing 20 ml L⁻¹ glycerol (2% final concentration), 20 ml L⁻¹ ethanol (2% final concentration) and 40 g L⁻¹ galactose (4% final concentration) with the appropriate nutrient dropped out. Yeast peptone (YP) production medium was prepared using 15 g L⁻¹ peptone, 15 g L⁻¹ yeast extract, 2 g L⁻¹ glucose (0.2% final concentration), 20 ml L⁻¹ glycerol (2% final concentration), 20 ml L⁻¹ ethanol (2% final concentration) and 40 g L⁻¹ galactose (4% final concentration). Luria–Bertani (LB) medium was purchased from Fisher. Terrific Broth (TB) medium was purchased from Sigma and supplemented with 8 ml L⁻¹ glycerol (0.8% final). Kanamycin was added to LB and TB media at 50 mg L⁻¹ final concentration.

Biochemical characterization of terpene cyclase genes

Terpene cyclase plasmids were transformed into *S. cerevisiae* strain YPH499 and then grown on SC (–Leu) dropout agar plates for 3 days. Colonies (3–5 pooled) were seeded into 10 ml of SC (–Leu) liquid medium and grown overnight at 30 °C. Overnight seed cultures were centrifuged then resuspended in YP liquid production medium and used to inoculate 50 ml total of YP liquid production medium (OD₆₀₀ = 1.50 ± 0.06) in 250-ml baffled flasks with a vented metal cap. Cultures were grown at 22 °C for 5 days and then whole culture broths were extracted with 1:1 diethyl ether–hexane (50 ml) by shaking followed by centrifugation. The separated organic layer was collected and dried by shaking with a sodium sulfate and silica gel mixture. The volume was then reduced to about 3 ml by evaporating under a stream of air in a test tube and filtered with a PTFE syringe filter. A portion (2 ml) was evaporated, dissolved in hexane (0.2 ml) and analyzed by GC–MS.

GC–MS was performed on an Agilent 7890B GC equipped with a 5977B MSD and 7693 autosampler and an Agilent DP5 MS + DG column. The oven program consisted of a 1-min period at 75 °C, followed by a linear increase to 250 °C after 11 min, then an increase to 300 °C after 1 min and finally a 1-min hold at 300 °C (14 min total) with a helium carrier gas flow rate of 1 ml min⁻¹. A 6-min solvent delay with 1:50 split was used.

Biochemical screening of CYP genes

CYP plasmids were transformed into *S. cerevisiae* strain YPH500 and then grown on SC (–Ura) dropout agar plates for 2 days. Colonies were seeded into 3 ml of SC (–Ura) liquid medium and grown overnight at 30 °C. Overnight cultures (1-ml portions each) containing CYP plasmids were added to a 24-well plate and centrifuged, washed with SC (–Leu, –Ura) medium, centrifuged again and resuspended with 4 ml of SC (–Leu, –Ura) medium. Yeast cells harboring the DgemTC plasmid were harvested from agar plates grown as described above by scraping up several colonies with an inoculating loop and then added to each

well containing CYP-harboring cell suspensions. Cell mixtures were incubated overnight at 30 °C and then 0.2 ml of each cell suspension was plated onto SC (–Leu, –Ura) dropout agar in six-well dishes and grown for 2 days, after which colonies formed. Colonies were seeded into 5 ml of SC (–Leu, –Ura) liquid medium and grown overnight at 30 °C. Overnight seed cultures were used to inoculate SC (–Leu, –Ura) liquid production medium (25 ml) with HP20 resin added (1 g) in 125-ml flasks to an OD₆₀₀ of about 1.0 and then grown at 20 °C for 3 days. The whole culture broths were filtered through cheese cloth loaded into a syringe barrel. The separated resin was eluted with acetonitrile (15 ml). Aliquots (1 ml) were centrifuged and then analyzed by LC–MS method B.

LC–MS method B. UPLC analysis was performed and high-resolution mass spectra was collected using a Waters Acquity UPLC linked to a Waters Xevo G2-XS Q-TOF. A Waters BEH C₁₈ column (2.1 × 50 mm, 1.7 μm) was used. The LC gradient consisted of a 1-min isocratic period with 50% acetonitrile in water at 0.6 ml min^{−1} followed by a linear increase to 75% acetonitrile after 5 min total, an increase to 100% acetonitrile after 6 min total and then a 2-min isocratic period at 100% acetonitrile (8 min total).

Scale-up fermentation to produce 10, 11 and 12

Dgem cbTC, Ecar cbCYPb and Selo cbCYPc showed relatively high levels of production and were used as examples for scale-up fermentation. Yeast cultures harboring either TC plasmid alone or a hybrid strain harboring a TC and a CYP plasmid were prepared by inoculating colonies into SC dropout liquid media (–Leu for TC strain and –Leu, –Ura for TC/CYP strains, 75 ml each) followed by growth overnight at 30 °C. The overnight cultures were split into 30-ml aliquots, seeded into a 2.8-L baffled Fernbach flask containing YP production medium without added galactose (800 ml) and grown overnight at 20 °C to an OD₆₀₀ of about 1.0. Cultures were induced by addition of 200 ml of 20% galactose to a final concentration of 40 g L^{−1} and then approximately 20 ml of HP20 was added as a slurry in ethanol. The culture was grown at 20 °C for 3 days and then harvested in the manner described above by collecting the resin and eluting with acetone (100 ml each). Second and third elutions were performed using 50 ml each of acetone–hexane 1:1 and 50 ml of diethyl ether, respectively. The first and second elutions were pooled and evaporated and then combined with the third (ether) elution and 25 ml of water was added. A separatory funnel was used to separate the top organic layer, which was then evaporated on a rotary evaporator. The syrupy residue was extracted with hexane and the hexane layer was collected with a glass pipette. Aliquots of these crude extracts were evaporated and checked by thin-layer chromatography (TLC) and LC–MS method C. Extracts in hexane were fractionated by flash chromatography using a Teledyne Isco CombiFlash system on a column of silica gel (24 g, 230–400-mesh grade 60) using hexane–isopropanol (0–20% isopropanol gradient over 12 min at 35 ml min^{−1} flow rate) with ultraviolet (UV) monitoring at 215 nm. Fractions were checked by TLC and evaporated to produce the following (in duplicate): Dgem cbTC strain, 4.2 ± 2.3 mg of cembrene B (**10**); Dgem cbTC/Ecar cbCYPb strain, 2.7 ± 0.1 mg of cembrene B (**10**) and 11.1 ± 3.5 mg of 19-hydroxycembrene B (**11**); Dgem cbTC/Selo cbCYPc strain, 2.8 ± 0.1 mg of cembrene B (**10**) and 22.1 ± 4.2 mg of 7-hydroxycembrene B (**12**).

LC–MS method C. UPLC analysis was performed and high-resolution mass spectra were collected using a Waters Acquity UPLC linked to a Waters Xevo G2-XS Q-TOF. A Waters HSS T3 C₁₈ column (2.1 × 100 mm, 1.8 μm) was used. The LC gradient consisted of a 1-min isocratic period at 60% acetonitrile in water at 0.4 ml min^{−1} flow rate, followed by a linear increase to 100% acetonitrile after 8 min total and then a 2.5-min isocratic period at 100% acetonitrile (10.5 min total).

UV spectra were recorded using a Thermo Scientific NanoDrop in cuvette mode with quartz cuvettes. IR spectra were recorded on a

Nicolet iSS Fourier-transform IR spectrometer operating in transmittance mode using polyethylene cards (Thermo Scientific). NMR data for **1**, **2**, **3**, **4**, **5**, and **6** were collected and processed using the following instruments, software and parameters: ¹H-NMR and ¹³C-NMR spectra were recorded on a Varian iNOVA 500 (¹H, 500 MHz) NMR spectrometer equipped with a 3-mm Nalorac MDBG probe or Varian iNOVA 600 (¹H, 600 MHz) NMR spectrometer equipped with a 5-mm Nalorac HCN probe operated using OpenVnmrJ. Data were processed and analyzed using MestreNova (version 14.3.1). Chemical shifts were referenced to the solvent residual proton for ¹H-NMR (δ 7.26 for CDCl₃) and the ¹³C signal for ¹³C-NMR (δ 77.2 for CDCl₃).

Cembrene B (10). Clear oil; ¹H-NMR (500 MHz, C₆D₆) δ 5.15 (m, 2H), 5.03 (t, *J* = 6.7 Hz, 1H), 2.89 (d, *J* = 6.9 Hz, 2H), 2.28 (t, *J* = 7.6 Hz, 2H), 2.16 (m, 2H), 2.12 (m, 2H), 2.10–2.00 (m, 6H), 1.69 (s, 3H), 1.66 (s, 3H), 1.60 (s, 3H), 1.55 (s, 3H), 1.52 (s, 3H). ¹³C-NMR (125 MHz, C₆D₆) δ 134.8, 133.7, 133.0, 131.7, 126.8, 125.4, 124.7, 124.7, 40.4, 39.3, 37.9, 32.0, 31.0, 25.4, 24.3, 20.9, 20.7, 16.0, 15.4, 15.1. ESI-MS [*M*]⁺ *m/z* 272.1.

19-Hydroxycembrene B (11). Clear oil; UV (acetonitrile) λ_{max} (log ε) 216 (3.3); IR (polyethylene film) ν_{max} 3,337, 2,962, 2,922, 1,436, 1,383, 1,001 cm^{−1}; NMR (Supplementary Tables 8 and 9); high-resolution ESI-MS [*M* + *H*]⁺ *m/z* 289.2540 (calculated for C₂₀H₃₃O⁺, 289.2531).

7-Hydroxycembrene B (12). Clear oil; [α]_D²⁰ −8.73° (c 0.275, methanol); UV (methanol) λ_{max} (log ε) 215 (3.2); IR (polyethylene film) ν_{max} 3,334, 2,916, 2,852, 1,462, 1,383, 994 cm^{−1}; NMR (Supplementary Tables 8 and 9); high-resolution ESI-MS [*M* − H₂O + *H*]⁺ *m/z* 271.2426 (calculated for C₂₀H₃₁⁺, 271.2426).

Cembrene 19-hydroxylase substrate selectivity assay

Sarcophyton trocheliophorum transcriptome reads (SRR12573941) were downloaded from the National Center for Biotechnology Information (NCBI) and assembled using SPAdes. Potential TPS genes were identified by BLAST using known octocoral TPS genes as a query. The cembrene C synthase was the sole TPS found in the *S. trocheliophorum* transcriptome. The cembrene A synthase was identified from our *B. asbestinum* transcriptome data in a similar manner. The respective TPS genes (cembrene A synthase from *B. asbestinum* and cembrene C synthase from *S. trocheliophorum*) were cloned into pESC-URA and expressed in yeast. The cembrene 19-hydroxylase Ecar cbCYPb was coexpressed with Basb cbTC, Basb caTC and Stro ccTC in the manner described above using SC liquid production medium and HP20 resin. The terpene products (cembrene A and cembrene C) were identified by extraction, purification using silica gel chromatography and characterization by ¹H-NMR and ¹³C-NMR spectroscopy with comparison to data from the literature. Only coexpression of Ecar cbCYPb with Basb cbTC (cembrene B) showed the production of oxidized products, while coexpression with Basb caTC and Stro ccTC (cembrene A and cembrene C, respectively) did not show any detectable oxidized products.

Biotransformation screen of cembrene alcohols 11 and 12 using yeast harboring CYP genes

YPH500 yeast cells harboring cbCYPb, cbCYPc or empty pESC-URA were seeded into SC (–Ura) liquid medium containing dextrose and grown overnight at 30 °C. An aliquot (0.5 ml) of each culture was seeded into SC (–Ura) liquid medium containing dextrose (10 ml). Cells were grown 4 h at 30 °C then centrifuged and resuspended in SC (–Ura) liquid medium containing galactose. To each culture was added BSA at 0.7 mg ml^{−1} followed by either **11** (to cbCYPc and empty vector) or **12** (to cbCYPb and empty vector) as a solution in ethanol to a final concentration of 100 μg ml^{−1}. Cultures were grown at 20 °C and 180 rpm in shake flasks for 3 days and then extracted with diethyl ether. The ether was evaporated and the residue was dissolved in methanol and analyzed by LC–MS method B.

Scaled-up biotransformation to produce diol 13

YPH500 yeast cells harboring either cbCYPb or cbCYPc were seeded into SC (–Ura) liquid medium containing dextrose (10 ml) and grown overnight at 30 °C. The seed cultures were added to SC (–Ura) liquid medium containing galactose (250 ml) followed by albumin (200 µg ml^{−1}) and then **11** (22 mg in ethanol) to cbCYPc and **12** (44 mg in ethanol) to cbCYPb. The cultures were shaken at 20 °C and 180 rpm for 4 days and then each was extracted with ethyl acetate (100 ml). The organic layer was evaporated and the residue purified by combiflash using a C₁₈ column and water–acetonitrile. The fractions were evaporated to produce **13** (11.3 mg from **11** and cbCYPc and 1.6 mg from **12** and cbCYPb). The compound was crystallized by evaporation from heptane and dichloromethane, and then further characterized by X-ray diffraction analysis (Cambridge Crystallographic Data Center (CCDC) 2371467).

7,19-Dihydroxycembrene B (13). Colorless crystals; [α]_D²⁰ +3.54° (c 0.226, methanol); UV (methanol) λ_{max} (log ϵ) 215 (3.1); IR (polyethylene film) ν_{max} 3,357, 2,919, 2,852, 1,711, 1,462, 1,383, 994 cm^{−1}; NMR (Supplementary Tables 8 and 9); high-resolution ESI-MS [M – H₂O + H]⁺ m/z 287.2381 (calculated for C₂₀H₃₁O⁺, 287.2370).

Biochemical characterization of SDH genes in vitro

Plasmids bearing SDH genes were transformed into *E. coli* BL21(DE3) strain and grown on LB kanamycin plates overnight at 37 °C. Colonies were inoculated into LB liquid medium (20 ml) with 50 mg L^{−1} kanamycin and 0.2% dextrose and grown overnight at 30 °C. Cultures were used to inoculate TB liquid medium (1 L) with 50 mg L^{−1} kanamycin and grown at 28 °C for 5 h resulting in varying OD₆₀₀ values between 0.5 and 1.5. Cultures were cooled to 16 °C and induced with 0.5 mM IPTG; then, an additional 50 mg L^{−1} kanamycin was added and the cells were grown overnight at 16 °C. Some strains grew noticeably more slowly, including those harboring *E. caribaeorum* cbSDH, *R. koellikeri* cbSDH and *S. elongata* cbSDH, correlating with higher relative SDH activity. These strains were grown for an additional 24 h to improve protein yields. Cells were then harvested by centrifugation and stored at −80 °C until use.

Cell pellets were suspended in 25 ml of lysis buffer (HEPES 25 mM pH 8.2, NaCl 100 mM, glycerol 5% v/v, TCEP 100 µM, lysozyme 0.6 mg ml^{−1} and DNase 0.05 mg ml^{−1}) and incubated at 4 °C for 1 h. The cells were disrupted using sonication and then centrifuged at 52,000g for 20 min at 4 °C. The supernatants were transferred to 50-ml conical tubes and about 0.25 ml of Ni-NTA resin was added; then, the mixture was incubated on ice for 1 h with periodic mixing. The suspensions were centrifuged at 5,000g for 5 min and the supernatant was pipetted off. The pellets of Ni-NTA resin were suspended in 2 ml of wash buffer (HEPES 25 mM pH 8.2, NaCl 100 mM, glycerol 5% v/v, TCEP 100 µM and imidazole 25 mM) and transferred to 2-ml tubes before centrifuging at 13,000g. This wash step was repeated once and then the Ni-NTA resin was eluted by adding 2 ml of elution buffer (HEPES 25 mM pH 8.2, NaCl 100 mM, glycerol 5% v/v, TCEP 100 µM and imidazole 250 mM). After centrifugation, the supernatant was pipetted into a separate tube and the elution was repeated. Using a 2-ml centrifugal concentrator with a membrane (3-kDa molecular weight cutoff), the eluents were pooled for each sample, concentrated and buffer-exchanged twice in assay buffer (HEPES 25 mM pH 8.2, NaCl 100 mM, glycerol 5% v/v and TCEP 100 µM). Typical protein titers were determined by UV absorption at 280 nm to be between 3 and 6 mg from 1 L of culture (Supplementary Fig. 11). Assays were performed by dissolving NADP (200 µM) and BSA (1 mg ml^{−1}) in 0.5 ml of assay buffer and then adding 100 µg of cembrene alcohol (**11**, **12** or **13**) as a 10-µl aliquot in ethanol followed by addition of the SDH enzyme to a final volume of 1 ml (final enzyme concentrations between 5 and 9 µM). The reaction was shaken at 200 rpm and 20 °C for 16 h and then extracted with diethyl ether (three times with 0.5 ml) with vortexing followed by centrifugation at 13,000g for 3 min. The pooled organic layers were evaporated under a stream of air and the

organic residue dissolved in 2 ml of methanol. Enzyme reactions were analyzed by LC–MS method C.

Semisynthesis of 14

To a solution of **11** (25 mg, 0.086 mmol, one equivalent) in dichloromethane (2 ml) was added Dess–Martin periodinane (183.2 mg, 0.432 mmol, five equivalents) and the suspension was stirred at room temperature overnight. The reaction was quenched by addition of 200 µl of isopropanol followed by addition of saturated sodium chloride solution. The dichloromethane layer was separated, dried on a rotary evaporator and purified by silica gel chromatography to produce **14** (20 mg, 0.070 mmol, 81% yield).

Aldehyde 14. Clear oil; UV (acetonitrile) λ_{max} (log ϵ) 215 (3.8), 252 (3.7); IR (polyethylene film) ν_{max} 2,964, 2,918, 2,852, 1,668, 1,261, 1,020 cm^{−1}; NMR (Supplementary Tables 8 and 9); high-resolution ESI-MS [M – H₂O + H]⁺ m/z 269.2264 (calculated for C₂₀H₂₉⁺, 269.2264).

Preparative-scale in vitro reaction with SDH and diol 13

E. coli strain BL21(DE2) harboring *S. elongata* cbSDH on pET-28b(+) was grown in six separate 1-L cultures using TB medium, as described above. A cell pellet (100 ml in volume) was suspended in HEPES lysis buffer (50 mM HEPES, 150 mM NaCl and 10 mM imidazole, pH 8.0) and treated with lysozyme (0.6 mg ml^{−1}) and DNase (0.1 mg ml^{−1}). No detergents or reducing agents were added, as these were found to inhibit the reaction in previous experiments. The suspension was sonicated and then centrifuged at 52,000g for 20 min. The clarified lysate was incubated over 20 ml of Ni-NTA in a beaker with stirring and then collected in a glass column. The resin bed was washed with HEPES lysis buffer (50 ml), HEPES wash buffer (50 ml; 50 mM HEPES, 150 mM NaCl and 50 mM imidazole, pH 8.0) and then eluted with HEPES elution buffer (five times with 15 ml; 50 mM HEPES, 150 mM NaCl and 250 mM imidazole, pH 8.0). Fractions 2, 3, 4 and 5 were diluted up to 200 ml with HEPES reaction buffer (50 mM HEPES and 150 mM NaCl, pH 8.0) in an Erlenmeyer flask with gentle magnetic stirring at room temperature before adding BSA (0.1 mg ml^{−1}), NADP (2 mM) and **13** (15 mg in 1,000 µl of DMSO). The reaction mixture was stirred for 2 days. The aqueous reaction was extracted twice with 100 ml of dichloromethane, using centrifugation to break the emulsion, and the organic extract was dried over sodium sulfate and evaporated. The residue was filtered through a short bed of silica gel using hexane–ethyl acetate 1:1 and then purified by preparative HPLC using a phenylhexyl column and water–acetonitrile as the mobile phase to provide lactone **15** (1.5 mg).

Cembrene B lactone 15. Clear oil; [α]_D²⁰ +42.4° (c 0.250, methanol); UV (methanol) λ_{max} (log ϵ) 230 (4.1), 260 (shoulder); IR (polyethylene film) ν_{max} 2,919, 2,850, 1,757, 1,710, 1,461, 1,385, 1,289, 1,089, 1,029 cm^{−1}; NMR (Supplementary Tables 8 and 9); high-resolution ESI-MS [M + H]⁺ m/z 301.2183 (calculated for C₂₀H₂₉O₂⁺, 301.2163).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The feature-based molecular network parameters used and raw data are available through MASSIVE ([MSV000094792](https://www.ebi.ac.uk/MSV000094792)). NCBI BioSample accession numbers for the genomes generated in this paper are as follows: *R. koellikeri* (SAMN40621396), *B. asbestinum* (SAMN40621398), *D. gemmacea* (SAMN40621399), *E. caribaeorum* (SAMN41659149) and *S. elongata* (SAMN40621397). NMR and EI-MS spectra of characterized compounds are available in Supplementary Information. Sequence Read Archive accession numbers for raw sequencing reads generated in this study are available in Supplementary Table 4. Genbank accession numbers for annotated BGC-containing contigs are available in

Supplementary Table S. A list of all protein sequences used in the study is available in Supplementary Note 1. The sources of all CYPs used for phylogenetic analysis are available in Supplementary Table 6. X-ray diffraction data are available from the CCDC ([2371467](https://www.ccdc.ac.uk)). The conserved domain database webtool used to explore genomic neighborhoods can be accessed online (<https://www.ncbi.nlm.nih.gov/cdd/>).

References

59. Burkhardt, I., Dürr, L., Grayson, N. E. & Moore, B. S. in *Methods in Enzymology* Vol. 699 (ed. Rudolf, J.) 343–371 (Academic, 2024).
60. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, e107 (2023).
61. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
62. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
63. De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**, btad311 (2023).
64. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
65. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
66. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
67. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
68. Cheng, H. et al. Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **40**, 1332–1335 (2022).
69. Manni, M., Berkeley, M. R., Seppely, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
70. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
71. Grabherr, M. G. et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. *Nat. Biotechnol.* **29**, 644–652 (2011).
72. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
73. Pribelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes de novo assembler. *Curr. Protoc. Bioinformatics* **70**, e102 (2020).
74. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
75. Wang, J. et al. The conserved domain database in 2023. *Nucleic Acids Res.* **51**, D384–D388 (2022).
76. Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **7**, S10 (2006).
77. Potter, S. C. et al. HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
78. Gilchrist, C. L. M. & Chooi, Y.-H. Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
79. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
80. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
81. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
82. Lassmann, T., Frings, O. & Sonhammer, E. L. L. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res.* **37**, 858–865 (2009).
83. Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
84. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
85. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
86. Schmid, R. et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* **41**, 447–449 (2023).
87. Nothias, L.-F. et al. Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **17**, 905–908 (2020).
88. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
89. Scesa, P. D. & Schmidt, E. W. Biomimetic approach to diverse coral diterpenes from a biosynthetic scaffold. *Angew. Chem. Int. Ed. Engl.* **62**, e202311406 (2023).
90. Alcaino, J. et al. Functional characterization of the *Xanthophyllomyces dendrorhous* farnesyl pyrophosphate synthase and geranylgeranyl pyrophosphate synthase encoding genes that are involved in the synthesis of isoprenoid precursors. *PLoS ONE* **9**, e96626 (2014).
91. Sambrook, J. & Russell, D. W. Preparation and transformation of competent *E. coli* using calcium chloride. *Cold Spring Harb. Protoc.* **2006**, pdb.prot3932 (2006).
92. Synthetic Complete (SC) medium. *Cold Spring Harb. Protoc.* <https://doi.org/10.1101/pdb.rec090589> (2016).

Acknowledgements

We thank P. Zerofski (Scripps Institution of Oceanography) for collecting *R. koellikeri* and *S. elongata*, J. Sprung (Two Little Fishes) for the *D. gemmacea* tissue and J. Garrahou for the *C. rubrum* tissue. For her help in species identification, we would like to recognize C. S. McFadden (Harvey Mudd College). For coral images, we thank C. S. McFadden for *D. gemmacea*, F. Zuberer (CNRS) for *C. rubrum*, P. Webster (@underwaterpat) for *S. elongata* and J. Simpson (Herbert Wertheim UF Scripps Institute for Biomedical Innovation and Technology) for *B. asbestinum* and *E. caribaeorum*. We thank T. Damiani (IOCB Prague) and A. M. C. Rodriguez (UC San Diego) for their assistance with feature-based molecular networking. Chemical isolation and elucidation of the *R. koellikeri* briarane diterpene was aided by A. Bogdanov (Scripps Institution of Oceanography). We thank N. M. Lacerna (University of Utah) for aid with gel electrophoresis. This work was supported by the NIH (R01-GM146224 to B.S.M., R35-GM148283 to E.W.S. and K99-GM148783 to P.D.S.), a Margaret A. Davidson Graduate Fellowship to N.E.G (NERRS NA22NOS4200050), a National Science Foundation Graduate Research Fellowship to M.L.M., Tang Genomics Fund to T.P.M. and Fundação para a Ciência e a Tecnologia funds (UIDB/04423/2020, UIDP/04423/2020 and 2021.00855.CEECIND) to J.-B.L. X-ray diffraction research reported in this publication was performed by R. T. Vanderlinden (U. Utah) and supported by the Office of the Director, NIH under award S10OD030326. NMR data collection at UC San Diego (the Scripps

Institution of Oceanography and Biomolecular NMR Facilities) was assisted by B. Duggan. Additional NMR data were obtained at the University of Utah Health Sciences NMR Core. GC–MS data were obtained at the University of Utah Health Sciences Proteomics Core by Q. Pearce.

Author contributions

Conceptualization, P.D.S., I.B., E.W.S. and B.S.M. Data creation, N.E.G., P.D.S., M.L.M. and I.B. Formal analysis and validation, N.E.G., P.D.S., I.B., J.G.-G. and T.A. Funding acquisition, P.D.S., J.-B.L., T.P.M., E.W.S. and B.S.M. Investigation and methodology, N.E.G., P.D.S., M.L.M. and I.B. Project administration, P.D.S., I.B., E.W.S. and B.S.M. Resources, J.-B.L., T.P.M., E.W.S. and B.S.M. Writing—original draft, N.E.G., P.D.S., I.B. and B.S.M. Writing—review and editing, all authors.

Competing interests

The authors declare no competing interests.

Additional information

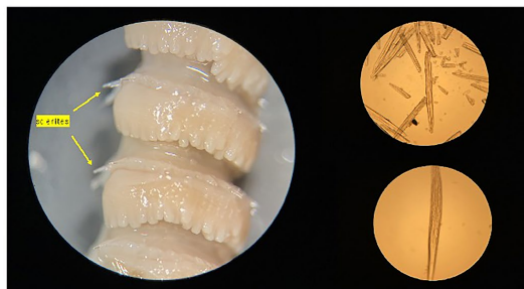
Extended data is available for this paper at <https://doi.org/10.1038/s41589-025-01927-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-025-01927-y>.

Correspondence and requests for materials should be addressed to Immo Burkhardt, Eric W. Schmidt or Bradley S. Moore.

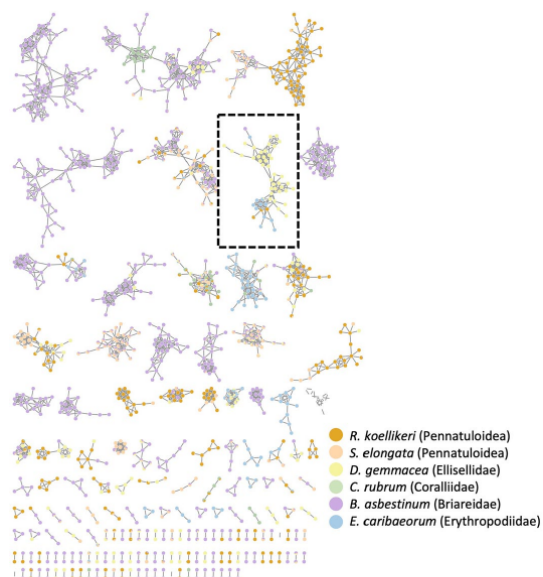
Peer review information *Nature Chemical Biology* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



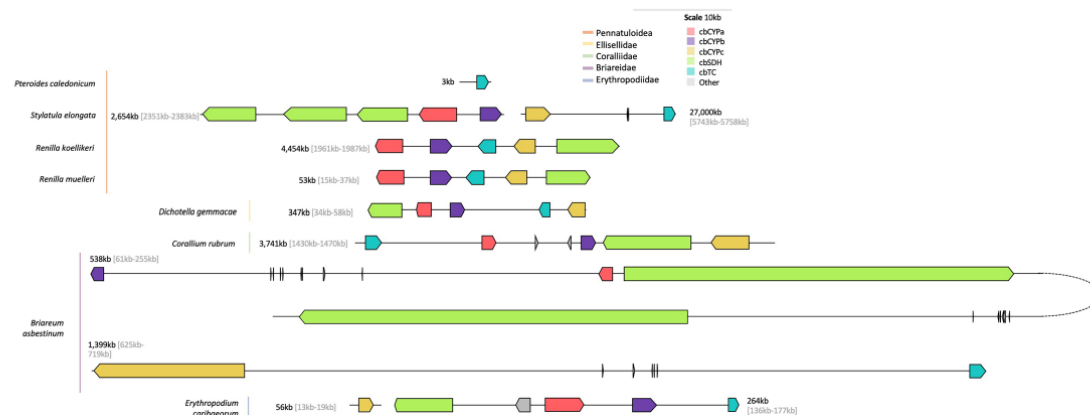
Extended Data Fig. 1 | *Stylatula elongata* species identification. To accurately differentiate between two visually similar southern California sea pen species, *Stylatula elongata* and *Virgularia* sp., we examined the presence or absence of sclerites at the base of the polyp leaves. Sclerites are microscopic carbonate structures that provide structural support and serve as important taxonomic markers in octocorals. We imaged these sclerites at 10x magnification to

assess their morphology and distribution. The observed sclerites, measuring approximately 0.5–1.5 mm in length, confirmed the identification of the specimen as *Stylatula elongata*, as this species is known to possess these structures, whereas *Virgularia* lacks them. The figure illustrates the presence of the sclerites in our sample tissue, providing critical diagnostic evidence for species determination.



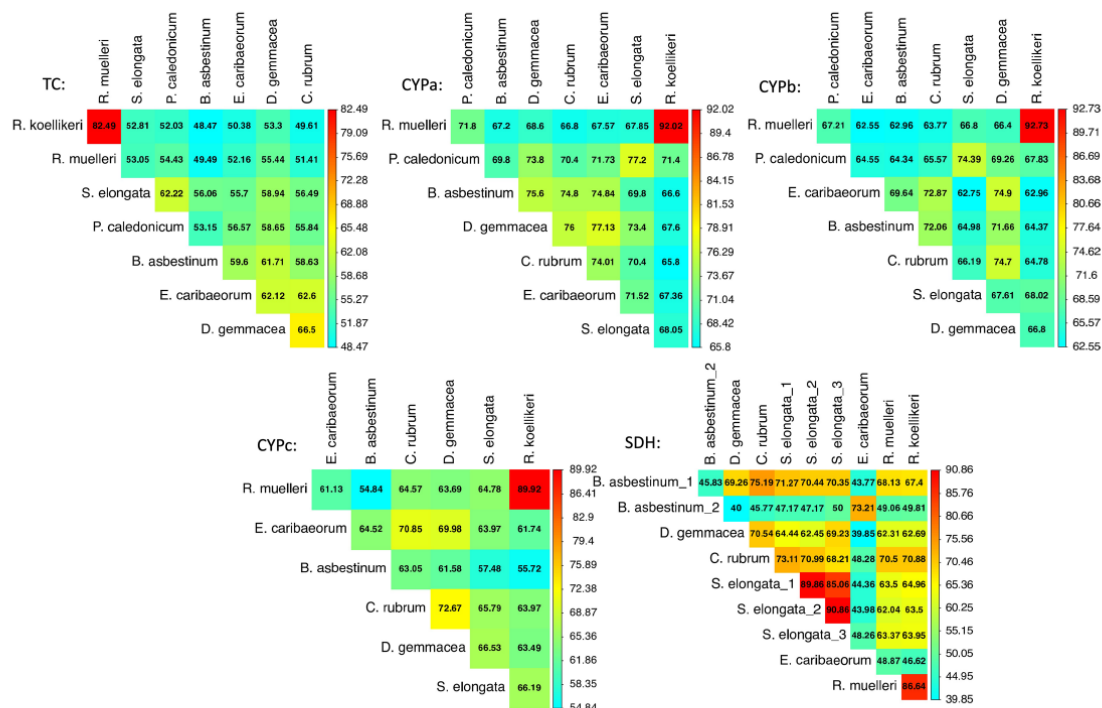
Extended Data Fig. 2 | The briarane terpenoid-containing complete molecular network. Feature-based molecular networking (FBMN) was used to visualize relationships among metabolites detected in liquid chromatography-tandem mass spectrometry (LC-MS/MS) data from crude extracts of the six focal coral species analyzed in this study. Each node in the network represents a molecular

feature, while edges indicate spectral similarity, grouping structurally related metabolites into clusters. The boxed region specifically highlights the briarane diterpene network, as shown in Fig. 2c, distinguishing this family of metabolites within the broader chemical space.



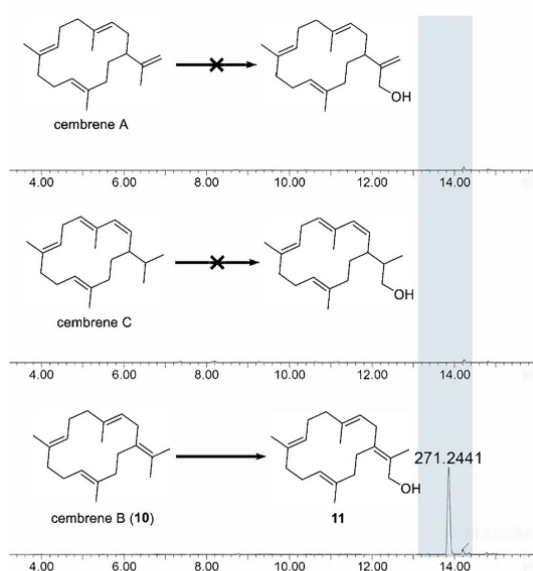
Extended Data Fig. 3 | Full to scale synteny of the briarane BGCs in Scleractyonacea corals. Full, to-scale micro synteny representation of the briarane biosynthetic gene clusters (BGCs) identified in briarane producing octocoral. Annotated arrows indicate genes within the BGCs, with intron sequences included. Gene lengths and intergenic gaps are shown to scale, providing an accurate depiction of cluster organization. The total length of the

contig containing the BGC is displayed in black adjacent to each contig, while the specific BGC-containing region depicted in the figure is indicated in parentheses. This visualization highlights the conserved and variable elements of briarane BGCs across coral species, offering insights into their genetic architecture and potential biosynthetic capabilities.



Extended Data Fig. 4 | %ID matrices for all syntenic briarane BGC genes. Heat maps illustrating the percent identity (% ID) across five biosynthetic gene cluster (BGC) genes identified in the focal *Scleractyonacea* corals. Each row represents a gene from a specific coral species analyzed in this study. Percent identity is color-coded, with red indicating the highest sequence similarity and blue

representing the lowest. This visualization provides a comparative overview of sequence conservation and divergence among BGC genes, offering insights into the evolutionary relationships and potential functional variations in briarane biosynthesis across coral species.



Extended Data Fig. 5 | LCMS for co-expression of cembrene A, B, and C synthases with cembrene 19-hydroxylase Ecar cbCYPb. EIC trace monitored at m/z 271.2. Co-expressed TC genes are as follows: Top) Basb caTC; Middle) Stro ccTC; Bottom) Basb cbTC. Oxidized products are only produced in the presence of cembrene B.

Extended Data Table 1 | Basic assembly statistics for octocoral genomes

Organism	Source	Accession	Genome Size (Mb)	Busco C (%) metazoan DB	contig N50 (kb)
<i>R. koellikeri</i>	This paper	SAMN40621396	144	86.9	4,895.0
<i>B. asbestinum</i>	This paper	SAMN40621398	1,155	87.3	331.8
<i>D. gemmacea</i>	This paper	SAMN40621399	401	87.5	195.9
<i>S. elongata</i>	This paper	SAMN40621397	269	88.0	5,139.0
<i>C. rubrum</i>	DOI: 10.1101/2024.07.13.603384	PRJEB54017	475	88.5	1,637.0
<i>E. caribaeorum</i>	This paper	SAMN41659149	300	86.4	233.0
<i>R. muelleri</i>	DOI: 10.1093/gigascience/giz026	http://rmue.reefgenomics.org	172	84.5	70.5
<i>P. caledonicum</i>	NA	SAMN14247561	292	72.4	19.2
<i>H. coerulea</i>	DOI: 10.1038/s41597-023-02291-z	SAMN33360050	429	88.9	1,428.0
<i>D. gigantea</i>	DOI: 10.1093/gbe/evz043	SAMN10516674	286	88.5	1,446.0
<i>Xenia</i> sp.	DOI: 10.1038/s41586-020-2385-7	SAMN23741995	227	92.0	1,100.0
<i>P. clavata</i>	DOI: 10.1534/g3.120.401371	SAMEA6085997	607	72.4	19.2
<i>Trachythela</i> sp.	DOI: 10.1093/gbe/evaa265	SAMN16072316	578	88.5	3,564.0
<i>R. reniformis</i>	NA	SAMEA103980763	132	39.8	2.0
<i>P. stokvisi</i>	NA	SAMN16093713	689	81.6	1.4
<i>P. marumi</i>	NA	SAMN16114184	583	67.9	1.3

Octocoral genomes generated in this study (5) and public assemblies (11). DOI's of papers associated with public genomes are provided where relevant. Accession numbers are provided for all.

Extended Data Table 2 | Long and short read sequencing statistics across all in house genomes

	<i>B. asbestinum</i>	<i>D. gemmacea</i>	<i>R. koellikeri</i>	<i>S. elongata</i>	<i>E. caribeorum</i>
Short Read Genomic Sequencing					
Total Sequences (M)	114	147	150	207	56
Total Bases (Gb)	17	22	23	31	5
Sequences flagged as poor quality	0	0	0	0	0
Sequence length	151	151	151	151	100
%GC	38	39	39	40	42
Kmer Based Genome size (Mb)	1,155	401	144	269	-
Genome Coverage	14	41	185	141	18
Long Read Genomic Sequencing					
Raw bases from WGS-Nanopore (Gb)	59	78	NA	NA	17
Raw bases from WGS-Pacbio (Gb)	NA	NA	29	26	NA
Mean Read length (kb)	2	4	12	13	2
%Reads Q>20	NA	NA	85	78	NA
Transcriptomic Short Read Sequencing					
Total Sequences (M)	184	618	178	226	NA
Coverage (x)	47	250	268	165	NA
HiSat Alignment	40.36%	92.43%	90.51%	60.42%	NA

For the genomes generated for this study, the statistics for the raw short and long reads are provided. Statistics for transcriptomic reads generated and their alignment rate to their respective assemblies.

A widespread metabolic gene cluster family in metazoans

In the format provided by the
authors and unedited

Contents

Supplementary Table 1. CMNPD results for Scleralcyonacea coral.	3
Supplementary Table 2. ¹ H and ¹³ C-NMR data for 11-hydroxyptilosarcenone (6) in CDCl ₃ ...	4
Supplementary Table 3. Samples used for molecular networking analyses.	5
Supplementary Table 4. Sequencing Accession numbers.	5
Supplementary Table 5. Genbank Accessions for annotated BGC contigs.	6
Supplementary Table 6. All CYPs used in phylogenetic analysis.	6
Supplementary Table 7. Recombinant enzymes biochemically characterized in this study. 7	
Supplementary Table 8. ¹ H-NMR data for 11 – 15 measured in CDCl ₃ at 500 MHz.	8
Supplementary Table 9. ¹³ C-NMR data for 11 – 15 measured in CDCl ₃ at 125 MHz.	9
Supplementary Table 10. X-ray crystallographic data for diol 13 (CCDC# 2371467).	10
Supplementary Figures	11
Supplementary Figure 1. 2D NMR correlations of 11-hydroxyptilosarcenone (6).	11
Supplementary Figure 2. K-mer plots for genome size predictions.....	11
Supplementary Figure 3. Phylogeny of all cnidarian CYPs.	12
Supplementary Figure 4. GCMS traces of cbTC harboring yeast.	13
Supplementary Figure 5. LCMS extracted ion chromatogram (EIC) traces for cbCYPb in vivo assays.....	14
Supplementary Figure 6. 2D NMR correlations for biosynthetic products.....	15
Supplementary Figure 7. LCMS EIC traces for cbCYPc in vivo assays.....	16
Supplementary Figure 8. LCMS EIC traces for cbCYPc enzyme in vivo biotransformation assays.....	17
Supplementary Figure 9. LCMS EIC traces for cbCYPb enzyme in vivo biotransformation assay.	18
Supplementary Figure 10. X-ray crystal structure of diol 13.....	19
Supplementary Figure 11. SDS-PAGE analysis of SDH proteins.	19
Supplementary Figure 12. LCMS EIC traces for cbSDH in vitro assays using added 19-hydroxy cembrene B (11).....	20
Supplementary Figure 13. LCMS EIC traces for cbSDH in vitro assays using added 7,19-dihydroxy cembrene B (13).....	21
Supplementary Figure 14. Proposed lactonization mechanism.	22
Supplementary Figure 15. Pipeline of BGC annotation.	22
Supplementary Note 1. Sequences of proteins expressed and analyzed in this study.	23

Supplementary Note 2. NMR spectra of <i>R. koellikeri</i> terpene mixture.	28
Supplementary Note 3. NMR spectra of cembrene B (10).	32
Supplementary Note 4. NMR spectra of 19-hydroxycembrene B (11).	34
Supplementary Note 5. NMR spectra of 17-hydroxycembrene (12).	40
Supplementary Note 6. NMR spectra of dihydroxycembrene B (13).	46
Supplementary Note 7. NMR spectra of cembrene B γ -lactone (15).	51
Supplementary Note 8. NMR spectra of cembrene B aldehyde (14).	56
Supplementary Note 9. NMR spectra of cembrene A.	62
Supplementary Note 10. NMR spectra of cembrene C.	64

Supplementary Table 1. CMNPD results for Scleralcyonacea coral.

Current Family Name	Terpenoids reported	Briaranes reported	Producer organism	Example briarane
Sarcodictyonidae	yes	no	N/A	N/A
Spongiodermidae	no	no	N/A	N/A
Ideogorgiidae	no	no	N/A	N/A
Parisididae	no	no	N/A	N/A
Parasphaerascleridae	yes	no	N/A	N/A
Chelidonisididae	no	no	N/A	N/A
Chrysogorgiidae	no	no	N/A	N/A
Primnoidae	yes	no	N/A	N/A
Keratoisididae	yes	no	N/A	N/A
Isidoidae	no	no	N/A	N/A
Mopseidae	no	no	N/A	N/A
Pleurogorgiidae	no	no	N/A	N/A
Ifalukellidae	yes	no	N/A	N/A
Helioporidae	yes	no	N/A	N/A
Cornulariidae	no	no	N/A	N/A
Pennatuloidae	yes	stylatulide	<i>Stylatula elongata</i>	stylatulide ¹
Briareidae	yes	briareolate ester J	<i>Briareum asbestinum</i>	briareolate ester J ²
Erythropodiidae	yes	eythrolide D	<i>Erythropodium caribaeorum</i>	eythrolide D ³
Coralliidae	yes	minabein-1	<i>Minabea</i> sp.	minabein ¹⁴
Ellisellidae	yes	yes	<i>Dichotella gemmacea</i>	gemmacolide R ⁵

Supplementary Table 2. ^1H and ^{13}C -NMR data for 11-hydroxyptilosarcenone (**6**) in CDCl_3 .

Position	Reported ^{6,7}		This study	
	δ_{C}	δ_{H}	$\delta_{\text{C}}^{\text{e, f}}$	δ_{H}
1	43.1 ^a	-	29.6	-
2	76.2	5.89 (d, 9.0)	76.2	5.93 (d, 7.2)
3	130.3 ^b	5.54 (dd, 11.7, 9.0)	127.4	5.57 (dd, 11.9, 10.2)
4	127.3 ^b	5.98 (d, 11.7)	130.5	6.00 (d, 11.9)
5	136.5	-	^g	-
6	60.9	4.96 (br d, 3.2) ^d	60.9	5.05 (br d)
7	69.2 ^c	5.02 (br s) ^d	78.8	4.98 (d, 3.7)
8	81.6	-	81.9	-
9	78.9 ^c	5.77 (d, 5.6)	69.4	5.81 (d, 6.1)
10	30.9 ^a	2.92 (d, 5.6)	43.2	2.95 (d, 6.1)
11	76.1	-	76.2	-
12	199.2	-	199.5	-
13	121.8	6.08 (d, 10.5)	121.7	6.13 (d, 10.8)
14	156.7	6.70 (d, 10.5)	156.7	6.73 (d, 10.8)
15	13.7	1.23 (s)	13.8	1.22 (s)
16	117.3	6.00 (br s)	117.5	6.03 (br s)
	-	5.88 (br s)	-	5.91 (br s)
17	45.8	2.35 (q, 7.1)	45.8	2.39 (q, 7.4)
18	6.7	1.21 (d, 7.1)	6.8	1.31 (d, 7.6)
19	174.8	-	175.0	-
20	25.2	1.46 (s)	25.2	1.49 (s)
CH₃COO-	170.0, 169.3	-	169.5/169.1	-
CH₃COO-	21.9, 21.1	2.20, 2.12	21.8, 21.1	2.24, 2.16
-OH	-	4.88 (s)	-	^g

^a Interchanged. ^b Interchanged. ^c Interchanged. ^d Interchanged. ^e Multiplicity determined by HSQC. ^f Detected by HSQC or HMBC. ^g Not detected.

Supplementary Table 3. Samples used for molecular networking analyses.

Organism	Sample Type	Exact mass of network briarane	Corresponding literature exact mass
<i>Stylatula elongata</i>	crude extract	NA	NA
<i>Briareum asbestinum</i>	encrusting morphology crude extract	NA	NA
	branching morphology crude extract	629.2659	NA
<i>Erythropodium caribaeorum</i>	fraction	554.1547 (erythrolide D)	554.1555 ⁸
<i>Renilla koellikeri</i>	crude extract	496.1449 (11-hydroxyptilosarcenone)	496.1500 ⁶
	pure compound	496.1449 (11-hydroxyptilosarcenone)	496.1500 ⁶
<i>Dichotella gemmacea</i>	crude extract	696.2207 (gemmacolide R)	696.2629 ⁵
<i>Corallium rubrum</i>	crude extract	NA	NA

Supplementary Table 4. Sequencing Accession numbers.

	Raw illumina RNA reads		Raw ONT reads		Raw PacBio reads		Raw illumina gDNA reads	
	Source	Accession	Source	Accession	Source	Accession	Source	Accession
<i>R. koellikeri</i>	This paper	SRR28507355	NA	NA	This paper	SRR28507351	NA	NA
<i>B. asbestinum</i>	This paper	SRR28507352	This paper	SRR28507348	NA	NA	This paper	SRR28507346
<i>D. gemmacea</i>	This paper	SRR28507353	This paper	SRR28507349	NA	NA	This paper	SRR28507347
<i>E. caribaeorum</i>	DOI: 10.1038/s41589-023-01352-z	SRR15783032	This paper	SRR29682103	NA	NA	DOI: 10.1038/s41589-023-01352-z	ERR4660993
<i>S. elongata</i>	This paper	SRR28507354	NA	NA	This paper	SRR28507350	NA	NA

Supplementary Table 5. Genbank Accessions for annotated BGC contigs.

Organism	Accession
<i>B. asbestinum</i>	PP584625.1 , PP584626
<i>D. gemmacea</i>	PP584627
<i>S. elongata</i>	PP584628 , PP584629
<i>R. koellikeri</i>	PP584630
<i>C. rubrum</i>	PP584631
<i>E. caribaeorum</i>	PP584632 , PP584633

Supplementary Table 6. All CYPs used in phylogenetic analysis.

Organism	# Annotated CYPs	Reason Included	CYP Sequence Source
<i>Briareum asbestinum</i>	65	Briarane Producer	This paper
<i>Corallium rubrum</i>	40	Briarane Producer	This paper
<i>Dichotella gemmacea</i>	42	Briarane Producer	This paper
<i>Erythropodium caribaeorum</i>	41	Briarane Producer	This paper
<i>Renilla koellikeri</i>	31	Briarane Producer	This paper
<i>Stylatula elongata</i>	40	Briarane Producer	This paper
<i>Heliopora coerulea</i>	40	Other octocoral	This paper
<i>Xenia sp.</i>	19	Other octocoral	This paper
<i>Hydra vulgaris</i>	24	Model Cnidarian	Pankov, K. V. et al. ⁹
<i>Acropora digitifera</i>	24	Model Cnidarian	Pankov, K. V. et al. ⁹
<i>Aurelia aurita</i>	37	Model Cnidarian	Pankov, K. V. et al. ⁹
<i>Nematostella vectensis</i>	70	Model Cnidarian	Pankov, K. V. et al. ⁹

Supplementary Table 7. Recombinant enzymes biochemically characterized in this study.

Entry	Enzyme	Organism	Classification	Function
1	Basb_cbTC	<i>Briareum asbestinum</i>	terpene cyclase	cembrene B synthase
2	Rkoe_cbTC	<i>Renilla koellikeri</i>	terpene cyclase	cembrene B synthase
3	Selo_cbTC	<i>Stylatula elongata</i>	terpene cyclase	cembrene B synthase
4	Dgem_cbTC	<i>Dichotella gemmacea</i>	terpene cyclase	cembrene B synthase
5	Crub_cbTC	<i>Corallium rubrum</i>	terpene cyclase	cembrene B synthase
6	Dgem_cbCYPa	<i>Dichotella gemmacea</i>	clan 46 cytochrome P450	unknown
7	Dgem_cbCYPb	<i>Dichotella gemmacea</i>	octocoral specific cytochrome P450	cembrene B 19-hydroxylase
8	Dgem_cbCYPc	<i>Dichotella gemmacea</i>	octocoral specific cytochrome P450	cembrene B 7-hydroxylase
9	Selo_cbCYPc	<i>Stylatula elongata</i>	octocoral specific cytochrome P450	cembrene B 7-hydroxylase
10	Rkoe_cbCYPa	<i>Renilla koellikeri</i>	clan 46 cytochrome P450	unknown
11	Rkoe_cbCYPb	<i>Renilla koellikeri</i>	octocoral specific cytochrome P450	cembrene B 19-hydroxylase
12	Rkoe_cbCYPc	<i>Renilla koellikeri</i>	octocoral specific cytochrome P450	cembrene B 7-hydroxylase
13	Crub_cbCYPa	<i>Corallium rubrum</i>	clan 46 cytochrome P450	unknown
14	Crub_cbCYPb	<i>Corallium rubrum</i>	octocoral specific cytochrome P450	cembrene B 19-hydroxylase
15	Ecar_cbCYPa	<i>Erythropodium caribaeorum</i>	clan 46 cytochrome P450	unknown
16	Ecar_cbCYPb	<i>Erythropodium caribaeorum</i>	octocoral specific cytochrome P450	cembrene B 19-hydroxylase
17	Basb_cbCYPa	<i>Briareum asbestinum</i>	clan 46 cytochrome P450	unknown
18	Basb_cbCYPb	<i>Briareum asbestinum</i>	octocoral specific cytochrome P450	cembrene B 19-hydroxylase
19	Basb_cbSDH1	<i>Briareum asbestinum</i>	short chain dehydrogenase	unknown
20	Basb_cbSDH2	<i>Briareum asbestinum</i>	short chain dehydrogenase	cembrene B alcohol dehydrogenase
21	Dgem_cbSDH1	<i>Dichotella gemmacea</i>	short chain dehydrogenase	cembrene B alcohol dehydrogenase
22	Ecar_cbSDH1	<i>Erythropodium caribaeorum</i>	short chain dehydrogenase	cembrene B alcohol dehydrogenase
23	Rkoe_cbSDH1	<i>Renilla koellikeri</i>	short chain dehydrogenase	cembrene B alcohol dehydrogenase
24	Rkoe_cbSDH2	<i>Renilla koellikeri</i>	short chain dehydrogenase	unknown
25	Selo_cbSDH1	<i>Stylatula elongata</i>	short chain dehydrogenase	cembrene B alcohol dehydrogenase
26	Basb_caTC	<i>Briareum asbestinum</i>	Terpene synthase	Cembrene A synthase
27	Stro_ccTC	<i>Sarcophyton trocheliophorum</i>	Terpene synthase	Cembrene C synthase
28	HcCPR	<i>Heliopora coelurea</i>	CYP450 reductase	CYPb and CYPc redox partner
29	EcCYTb5	<i>Erythropodium caribaeorum</i>	Cytochrome b5	Potential CYP redox partner

Supplementary Table 8. ¹H-NMR data for **11** – **15** measured in CDCl₃ at 500 MHz.

Position	11	12	13	14	15
2	2.05 (m)	2.04 (m)	2.05 (m) ^a	2.02 (m)	1.98 (m) 2.15 (m)
3	2.08 (m)	2.08 (m)	2.07 (m) ^a	2.06 (t, 6.5)	2.04 (m) 2.17 (m)
4	5.03 (t, 7.4)	4.98 (m) ^a	5.07 (t, 5.1)	5.05 (t, 7.4)	5.08 (t, 6.4)
6	2.00 (t, 7.9)	2.26 (dd, 10.6, 1.0)	2.37 (t, 13.2)	2.10 (t, 8.8)	2.21 (dd, 14.6, 7.0) 2.38 (dd, 14.6, 4.4)
		2.34 (dd, 10.8, 10.7)	2.23 (m) ^a		
7	2.28 (t, 7.9)	4.87 (dd, 10.9, 3.8)	4.86 (dd, 11.0, 4.4)	2.73 (t, 7.9)	4.75 (m)
9	2.84 (d, 7.2)	2.73 (dd, 16.8, 8.1) 2.87 (dd, 16.8, 8.1)	2.98 (dd, 16.1, 5.1) 2.75 (dd, 16.1, 6.6)	3.07 (d, 7.0)	3.06 (d, 6.8)
10	5.05 (t, 7.4)	4.98 (m) ^a	4.97 (m) ^a	5.08 (t, 7.4)	5.00 (t, 6.4)
12	2.16 (m)	2.08 (m)	2.09 (m) ^a 2.15 (m) ^a	2.18 (m)	2.14 (m)
13	2.20 (m)	2.13 (m) 2.21 (m)	2.25 (m) ^a 2.17 (m) ^a	2.19 (m)	2.27 (m) 2.10 (m)
14	4.94 (t, 6.6)	4.98 (m)	5.00 (m) ^a	4.91 (t, 6.6)	4.86 (t, 6.4)
15	1.56 (s)	1.52 (s)	1.53 (s)	1.54 (s)	1.53 (s)
16	1.60 (s)	1.62 (s)	1.64 (s)	1.60 (s)	1.72 (s)
18	1.76 (s)	1.70 (s)	1.79 (s)	1.74 (s)	1.80 (s)
19	4.09 (s)	1.66 (s)	4.16 (d, 11.7) 4.01 (d, 11.7)	10.08 (s)	-
20	1.60 (s)	1.66 (s)	1.68 (s)	1.62 (s)	1.61 (s)

^aOverlapping signal

Supplementary Table 9. ^{13}C -NMR data for **11** – **15** measured in CDCl_3 at 125 MHz.

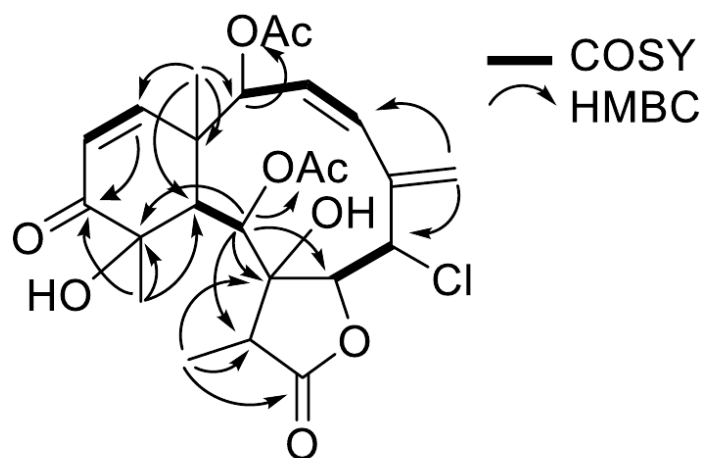
Position	11	12	13	14	15
1	133.2 (C)	133.2 (C)	133.0 (C)	133.2 (C)	133.9 (C)
2	40.0 (CH_2)	39.2 (CH_2)	39.0 (CH_2)	39.7 (CH_2)	39.5 (CH_2)
3	23.8 (CH_2)	24.3 (CH_2)	23.9 (CH_2)	23.7 (CH_2)	23.9 (CH_2)
4	125.1 (CH)	127.1 (CH)	127.8 (CH)	125.5 (CH)	127.5 (CH)
5	134.4 (C)	130.3 (C)	130.2 (C)	133.4 (C)	131.3 (C)
6	38.3 (CH_2)	44.8 (CH_2)	44.8 (CH_2)	39.0 (CH_2)	42.2 (CH_2)
7	31.4 (CH_2)	69.5 (CH)	69.1 (CH)	29.7 (CH_2)	83.9 (CH)
8	136.7 (C)	130.3 (C)	137.9 (C)	161.9 (C)	163.1 (C)
9	31.3 (CH_2)	25.4 (CH_2)	25.9 (CH_2)	33.6 (CH_2)	25.7 (CH_2)
10	123.7 (CH)	124.6 (CH)	123.3 (CH)	120.6 (CH)	119.1 (CH)
11	134.5 (C)	132.3 (C)	133.4 (C)	136.9 (C)	137.5 (C)
12	39.1 (CH_2)	38.1 (CH_2)	38.1 (CH_2)	39.1 (CH_2)	38.7 (CH_2)
13	25.0 (CH_2)	25.0 (CH_2)	24.8 (CH_2)	24.8 (CH_2)	24.8 (CH_2)
14	126.7 (CH)	127.3 (CH)	127.3 (CH)	126.1 (CH)	125.6 (CH)
15	15.0 (CH_3)	15.1 (CH_3)	14.9 (CH_3)	14.9 (CH_3)	15.1 (CH_3)
16	15.7 (CH_3)	16.9 (CH_3)	16.7 (CH_3)	15.8 (CH_3)	16.5 (CH_3)
17	129.3 (C)	133.2 (C)	133.8 (C)	133.0 (C)	123.1 (C)
18	16.9 (CH_3)	20.3 (CH_3)	18.0 (CH_3)	10.8 (CH_3)	8.6 (CH_3)
19	64.2 (CH_2)	21.6 (CH_3)	62.9 (CH_2)	191.1 (CH)	174.7 (C)
20	16.0 (CH_3)	17.2 (CH_3)	17.0 (CH_3)	15.6 (CH_3)	15.6 (CH_3)

^c Overlapping signal.

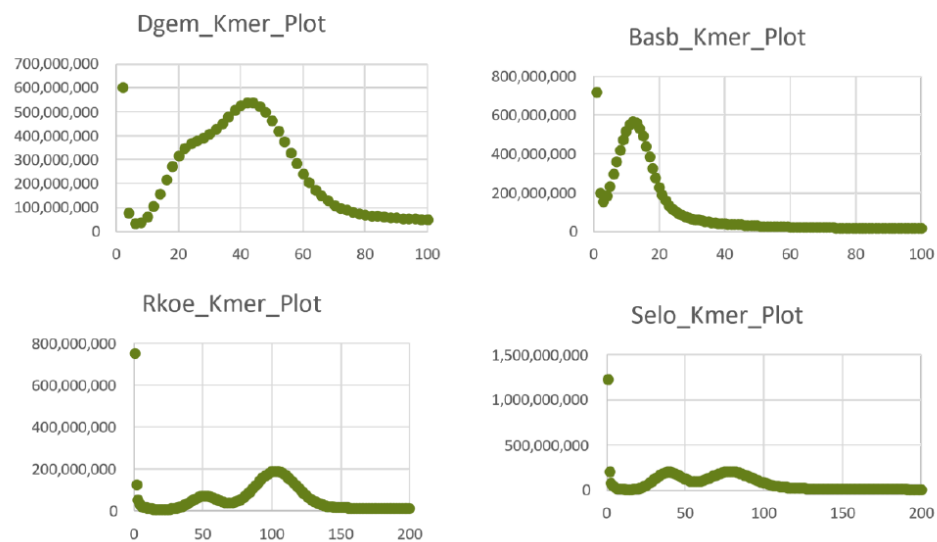
Supplementary Table 10. X-ray crystallographic data for diol **13** (CCDC# 2371467).

Formula	C20H32O2
<i>D</i> calc. (g cm ⁻³)	1.068
μ (mm ⁻¹)	0.513
Formula Weight	304.45
Colour	colourless
Shape	needle
Size (mm ³)	0.15×0.08×0.03
<i>T</i> (K)	100.00(10)
Crystal System	orthorhombic
Flack Parameter	0.00(9)
Hooft Parameter	0.01(6)
Space Group	<i>P</i> 212121
<i>a</i> (Å)	7.36900(10)
<i>b</i> (Å)	14.6718(2)
<i>c</i> (Å)	17.5054(3)
α (°)	90
β (°)	90
γ (°)	90
<i>V</i> (Å ³)	1892.62(5)
<i>Z</i>	4
<i>Z'</i>	1
Wavelength (Å)	1.54184
Radiation type	Cu K α
Θ min (°)	3.931
Θ max (°)	74.501
Measured Refl.	27879
Independent Refl.	3857
Reflections with $I > 2(I)$	3705
<i>R</i> int	0.0467
Parameters	203
Restraints	0
Largest Peak	0.149
Deepest Hole	-0.135
GooF	1.078
<i>wR</i> 2 (<i>all data</i>)	0.0893
<i>wR</i> 2	0.0885
<i>R</i> 1 (<i>all data</i>)	0.0350
<i>R</i> 1	0.0336

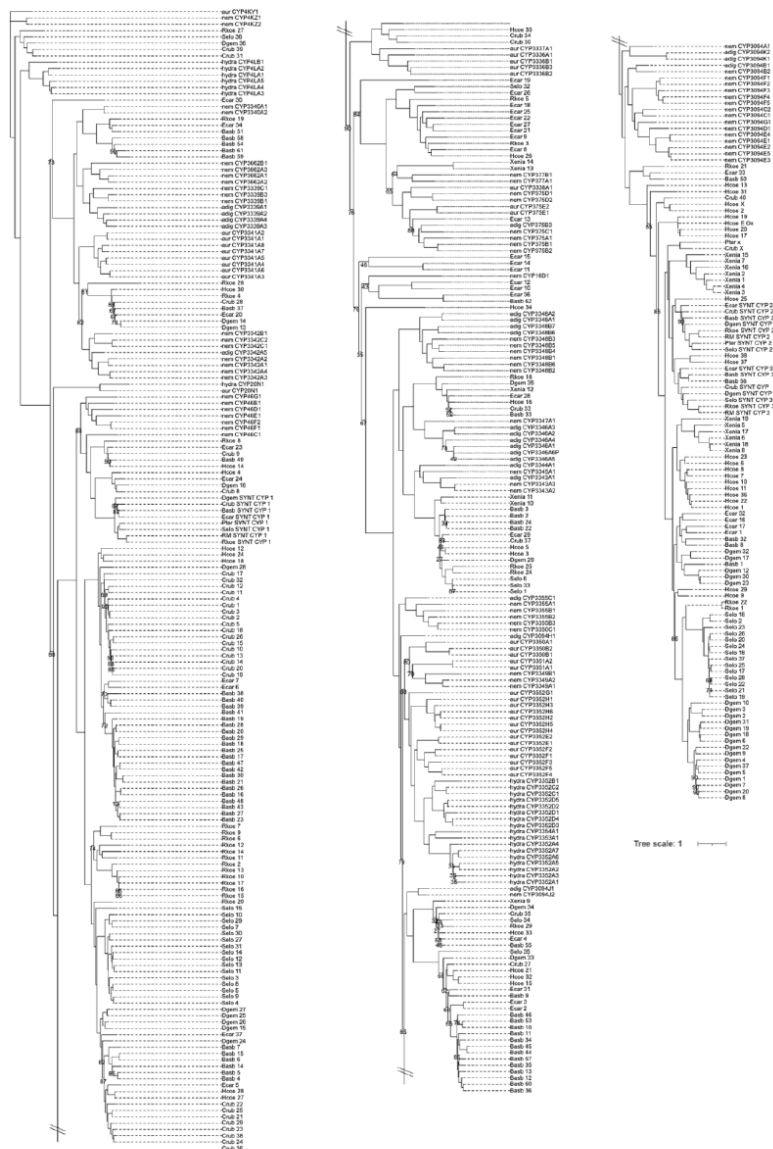
Supplementary Figures



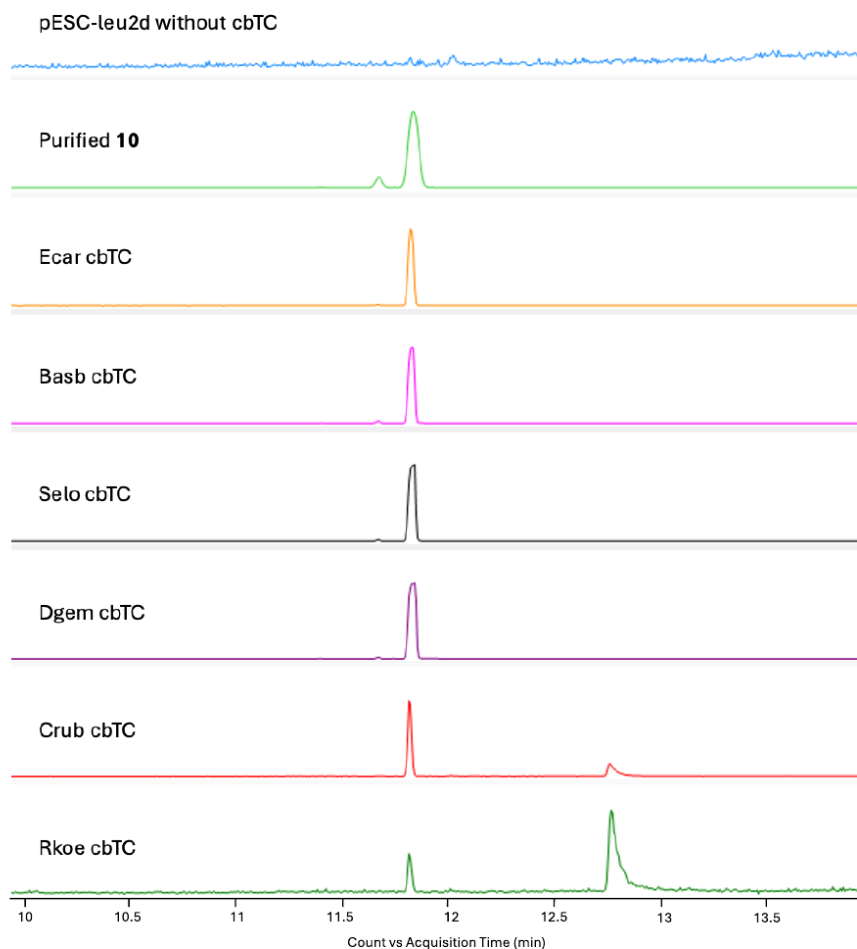
Supplementary Figure 1. 2D NMR correlations of 11-hydroxyptilosarcenone (6).



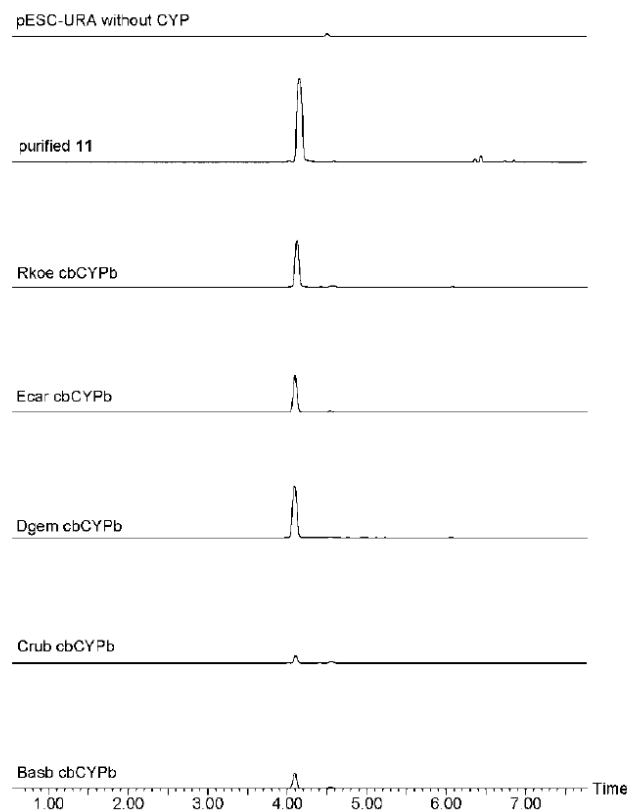
Supplementary Figure 2. K-mer plots for genome size predictions.



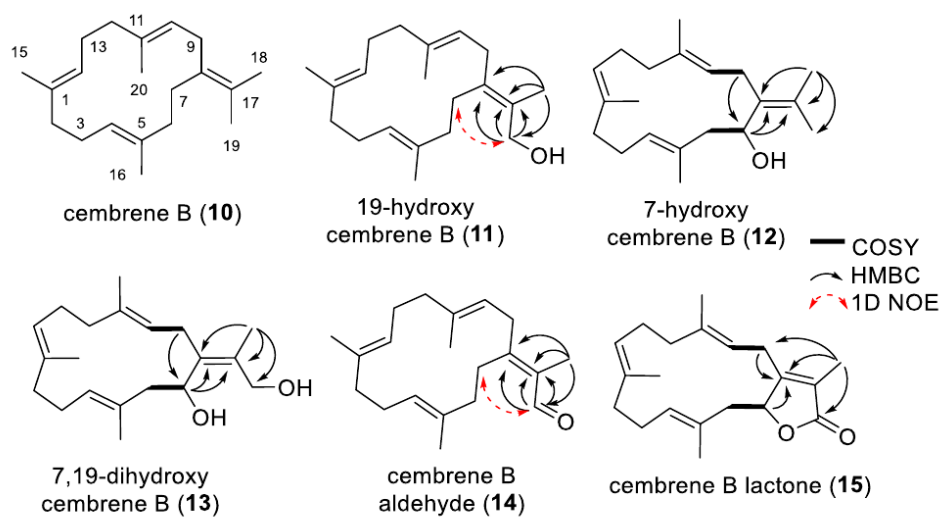
Supplementary Figure 3. Phylogeny of all cnidarian CYPs. Bootstraps lower than 90 are labeled.



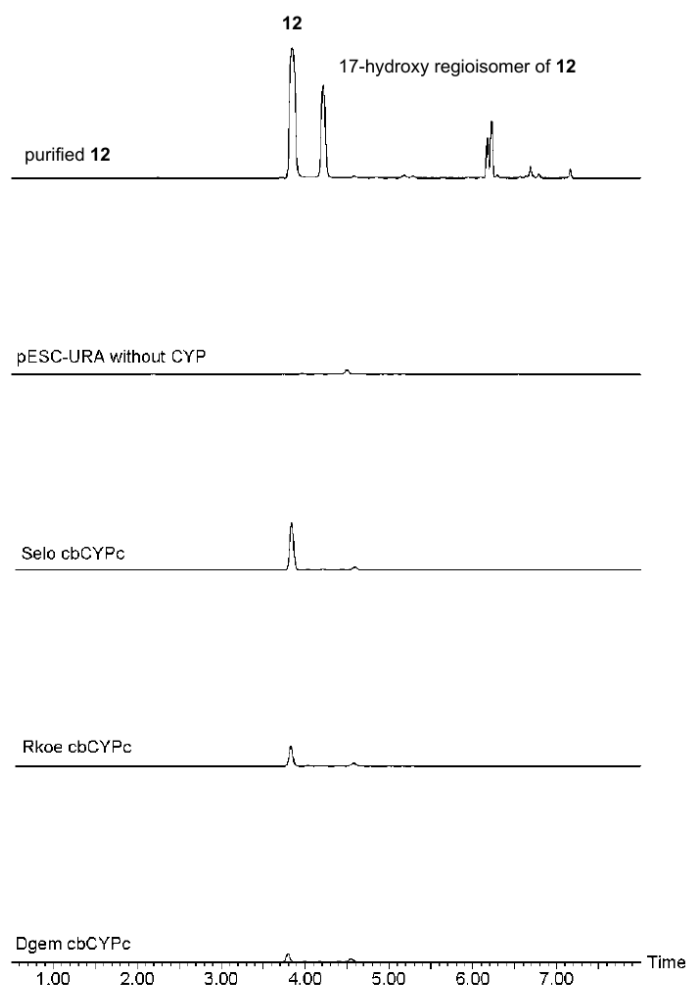
Supplementary Figure 4. GCMS traces of cbTC harboring yeast. GCMS EIC traces at 272.2 m/z for crude extracts of yeast harboring a cbTC or pESC-leu2d without cbTC. A trace of purified **10** is shown for comparison. The peak for **10** is present at 11.8 min. The peak at 12.8 min in Crub and Rkoe samples is GGOH. Experiments were repeated in triplicate with independent grow outs of the yeast.



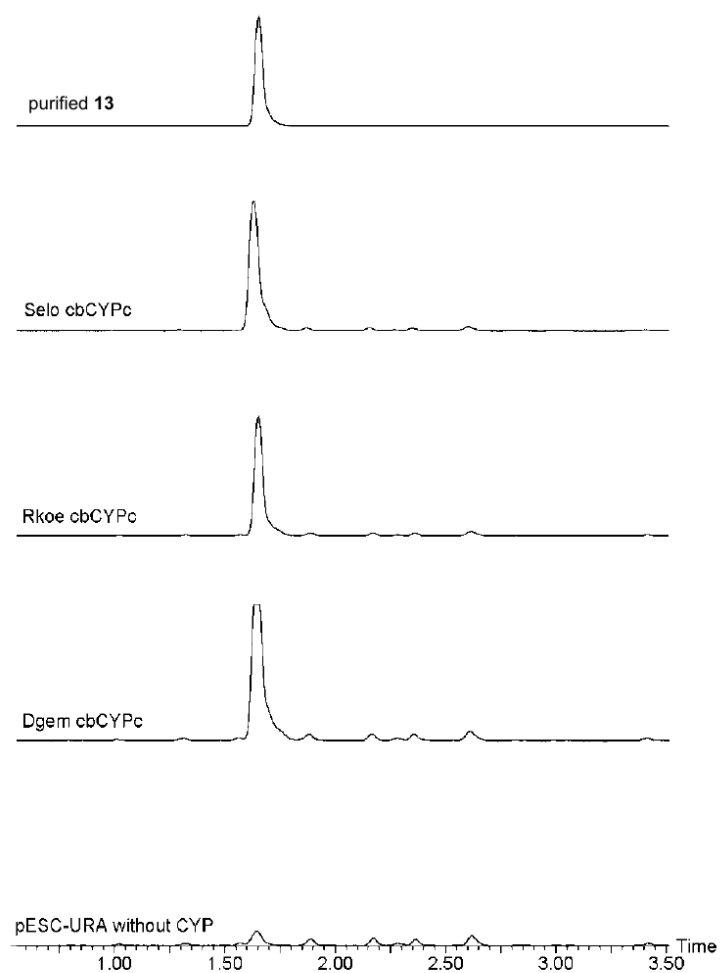
Supplementary Figure 5. LCMS extracted ion chromatogram (EIC) traces for cbCYPb in vivo assays. Using co-expression with a cbTC at m/z 271.2 in positive mode electrospray ionization (ESI). Experiments were repeated in triplicate with independent grow outs of the yeast.



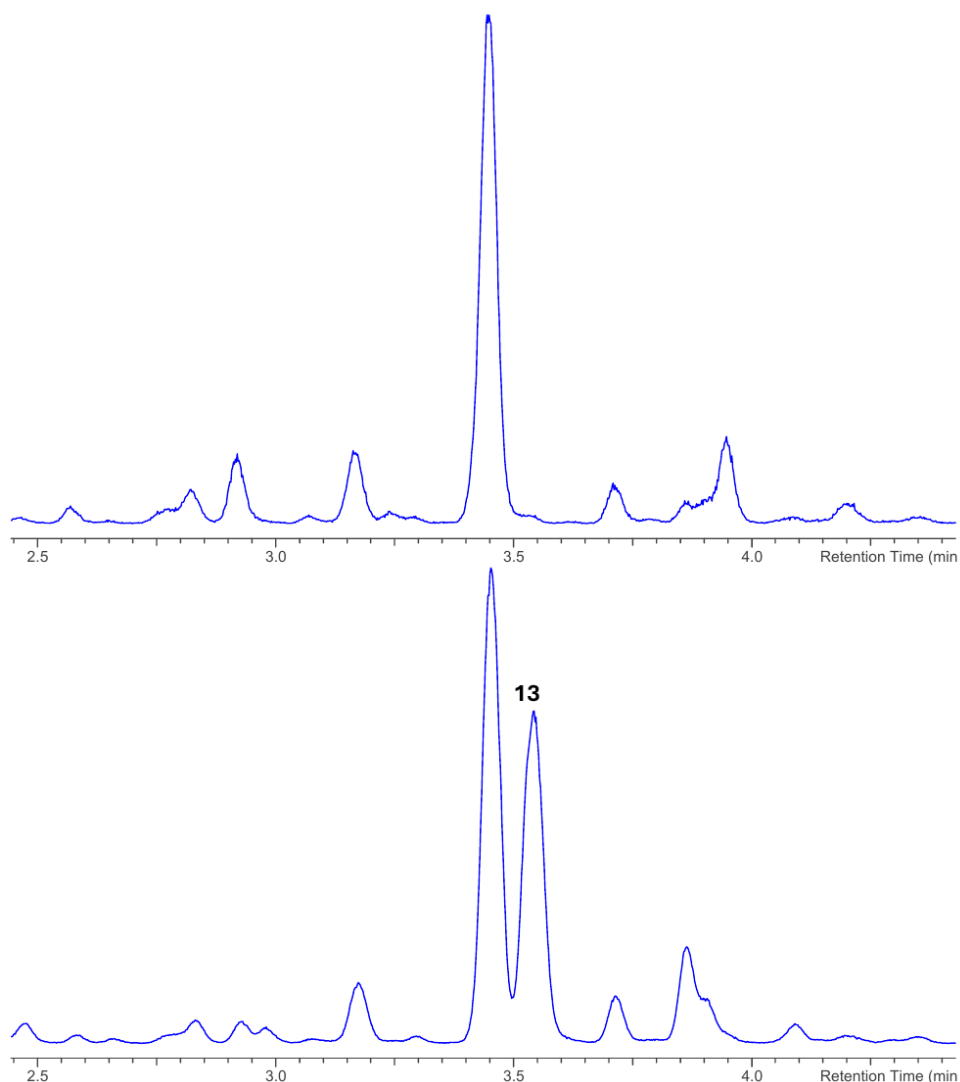
Supplementary Figure 6. 2D NMR correlations for biosynthetic products.



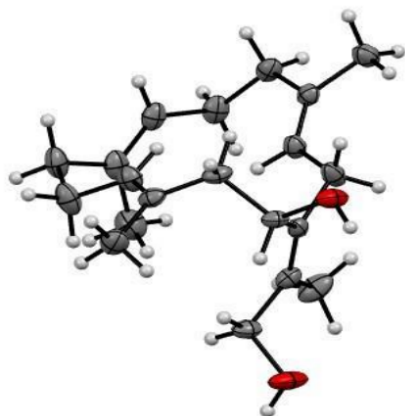
Supplementary Figure 7. LCMS EIC traces for cbCYPc in vivo assays. Using co-expression with a cbTC at m/z 271.2 in positive mode ESI. The peak after **12** in the trace is the 17-hydroxy regioisomer of **12**. Experiments were repeated in triplicate with independent grow outs of the yeast.



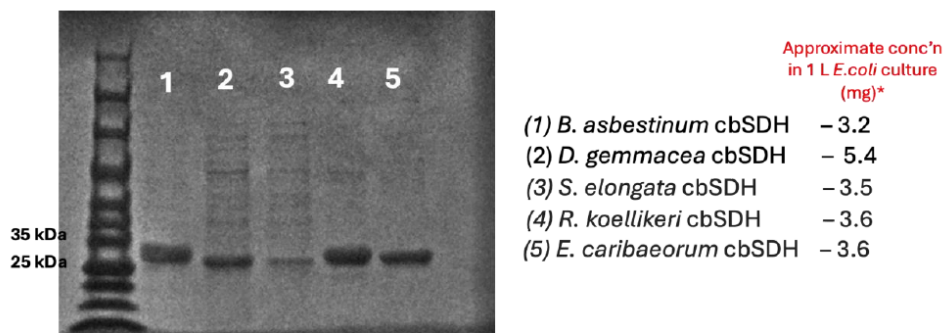
Supplementary Figure 8. LCMS EIC traces for cbCYPc enzyme in vivo biotransformation assays. Using added 19-hydroxy cembrene B (**11**) monitored at m/z 287.2 in positive mode ESI. Experiments were repeated in triplicate with independent grow outs of the yeast.



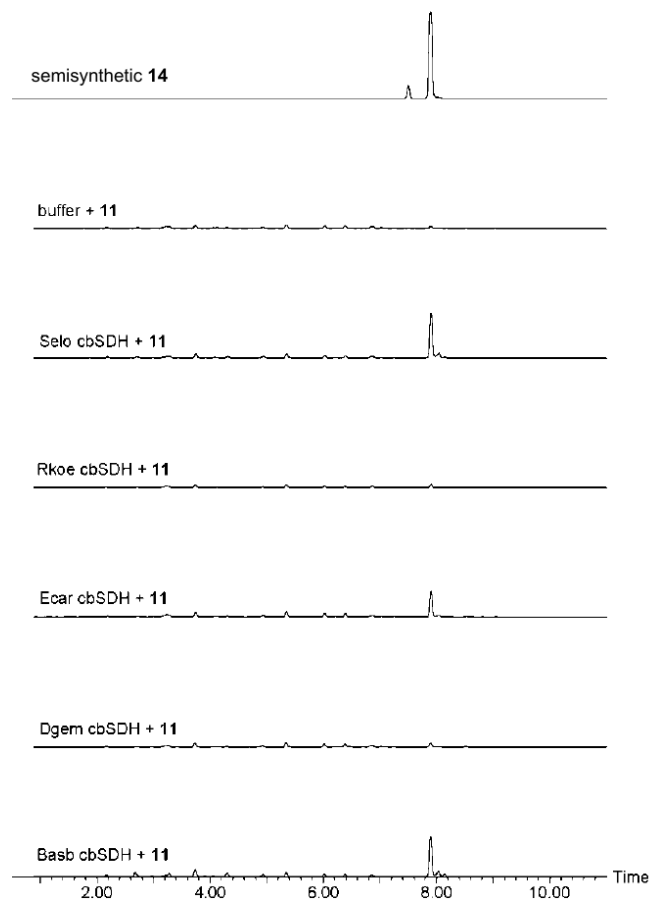
Supplementary Figure 9. LCMS EIC traces for cbCYPb enzyme in vivo biotransformation assay. Using added 7-hydroxy cembrene B (**12**) monitored at m/z 287.2 in positive mode ESI. Top) Empty vector control. Bottom) Extract of yeast harboring Ecar CYPb. Diol **13** is marker. Experiments were repeated in triplicate with independent grow outs of the yeast.



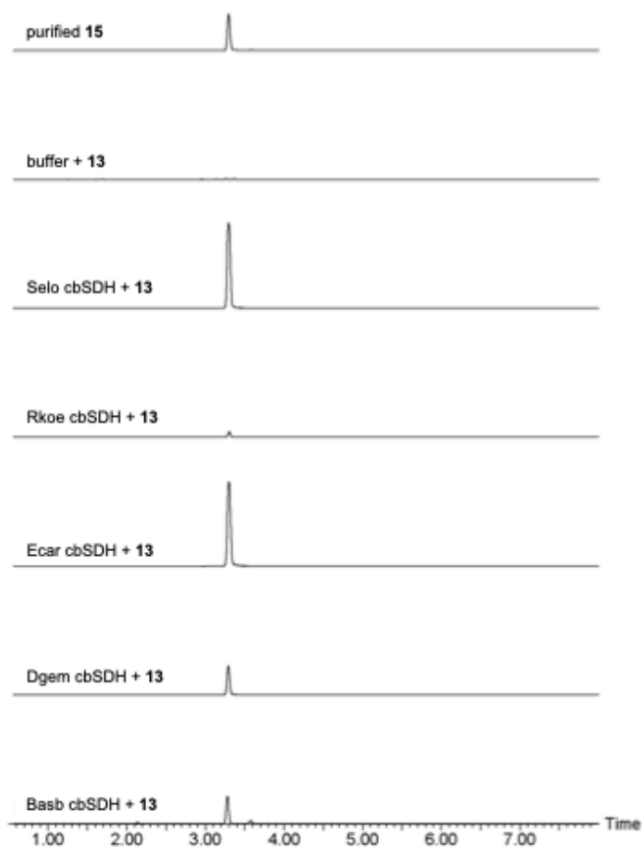
Supplementary Figure 10. ORTEP structure of diol **13** at the 50% thermal ellipsoid probability level, determined from X-ray crystallography (CCDC# 2371467).



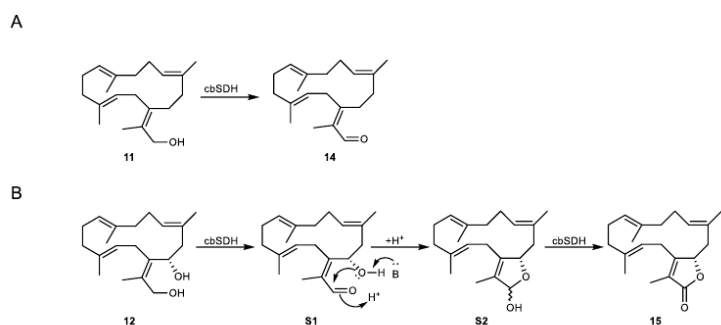
Supplementary Figure 11. SDS-PAGE analysis of SDH proteins. Proteins were expressed in *E. coli*, purified on a Ni-NTA column, and analyzed using a 4% to 20% gradient gel. Proteins in the following marked lanes are: 1) Basb_cbSDH2 2) Dgem_cbSDH1 3) Selo_cbSDH1 4) Rkoe_cbSDH1 5) Ecar_cbSDH1. The lane at left contains reference markers (Biorad Precision Plus Protein Kaleidoscope). The SDH proteins are expected to be approximately 29 kDa. The approximate concentrations according to the absorbance at 280 nm are also reported.



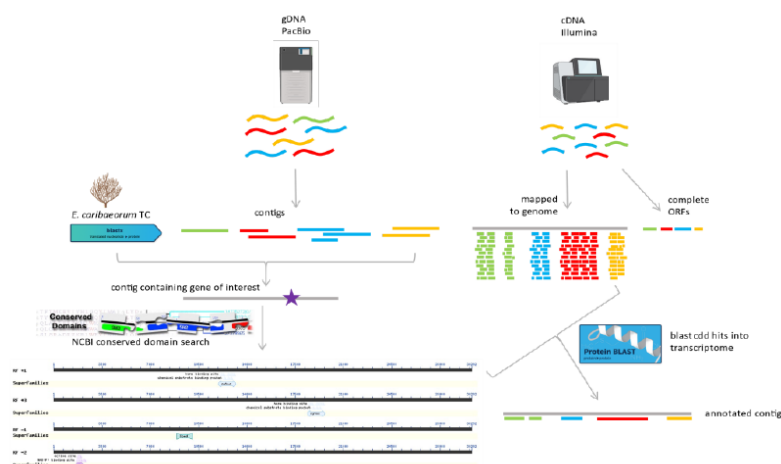
Supplementary Figure 12. LCMS EIC traces for cbSDH in vitro assays using added 19-hydroxy cembrene B (**11**). Monitored at m/z 287.2 in positive mode ESI. Enzyme assay experiments were repeated in triplicate with independent grow outs of the *E. coli* for each replicate.



Supplementary Figure 13. LCMS EIC traces for cbSDH in vitro assays using added 7,19-dihydroxy cembrene B (**13**). Monitored at m/z 323.2 ($[M+Na]^+$) in positive mode ESI. Enzyme assay experiments were repeated in triplicate with independent grow outs of the *E.coli* for each replicate.



Supplementary Figure 14. Proposed lactonization mechanism. A: formation of shunt product S1 by C19 alcohol oxidation by cbSDH. B: stepwise formation of the lactone by double oxidation at position 19 carried out by cbSDH.



Supplementary Figure 15. Pipeline of BGC annotation. High molecular weight genomic DNA is sequenced with either PacBio or Oxford Nanopore Technologies long read sequencing instruments. Those reads are then assembled into contigs which are used as the database for terpene synthase tblastn searches. When contigs containing TC genes are found, the surrounding genomic neighborhoods are then explored with NCBI conserved domain searches. In parallel, tissue from the same octocoral individuals is also used to generate RNAseq reads via short read sequencing platforms such as Illumina. These reads are then used to de novo assemble transcriptomes and are also mapped as raw reads to the genome. Based on conserved domain hits and mapping results, using blast searches complete coding sequences are then pulled out of the transcriptome.

Supplementary Note 1. Sequences of proteins expressed and analyzed in this study.

Cembrene synthase (TC) protein sequences tested in this study

(* = stop codon)

>Rkoe cbTC

MLRKEVKIPEQWCIPCEVEAVTPPKLRKDIIDWCVCTGITADRRTAEKAVDRLNPYMYMKILFPKLPDEPLIASQYELNA
SVIVTGMDDVLETYTLAIEELDHTFKETENWLLSLPSKDYDPLDAILDHIPTTKPTYSRSVAMFTDYFNKYCAAIRAGYK
PPDDLRLRRFRARLAAAVTAYLEMAKRKRDRDGLDEEEFLWQRSADNLSFPVLMLEAVFTGVLANGEIPNTTLHYHYFY
SNLFAIVLNDLNSYHRDRDNDNSLVKLWLKLGVAEKFEDEAAAKTVDFLNSIIVRMVYAMTENSESHPKSESFRFRFREA
VGYTFNGWILVHTTAVERYKSSPFQTVLSPVARGAEAWLKSETAFGKRCVTIFEELISRRVEEMNKLYGLDST*

>Basb cbTC

MSCKNLVRFPVNWCVPLSECHKSEEARKKHVDWCVNTGVATKLKAEKMSVQLNPFYMATLFPKVNPDAAAAEALY
LINASLIVGGYIVDDVLETYSVSGLEELDRFFRFTGWWVSKLLPDDYPKLDEILCSLSEIKVPYSKAVICMFIDNFNKYCYEI
FHKYTSSVQTVGNFRKRLSSSVTAYFEMALAKGRKSKVDEKEFLWQRSADNLGFPVLMLEVISGLLKQESTKIPQVITI
HYFHMYSNLFAHVLNDLNSYHRDVATNHNHNLVKLWLQNGIATDFDDAAIKIVEFLNSVIINIHKYCQKLRKREFSNSTETD
NFLESIKYFTTGWVHVHTTAVERYRLSPHQITLAKVESGEVNWQLQDESEFGKRCVQLFDKLLKERQAEYTVIYGME*

>Crub cbTC

MSTDREIYVPSWCVPLYNKPSIKLRERIADWMTKTGLASKTLAQKAVEKLNPFQYMNILFPKVDLSDSSAARLYEINAAAL
IVGGYTADDVIETYSLEALKELDYFFRSTEAWVAKLEPSDYPTVKEIAESLSEFEMEMSYKAVICMFIDSFNQYCYEIWRA
HSVSFSKMSQFRRRLSASVTAYLEIAMETRRGDANIDEKEFLWQRSADNLGFPVLMLEVLGSLNNECEIPLVTLHYFH
MYSNLFAHVLNDLNSYHRDIHTDHSIVKLWKKKGIAIDFDDAATKVKVQFLNSVTKHMQSFKKVVQDYPNCLALR
KFLNISYFTTGWVHVHTTAVERYKLSYQANLVKVEPNGEEEWLEDDSEFGNRCVQMFQDQLLCERENEMVFLYGLE*

>Selo cbTC

MSSEWTVSIPVSWYVPTNTTSLNLFRIKIVSWCTRTGIAGRSVAEKAVSRLNPAHYMMTLFPKVDPLQPTLSKLYEMNA
AIIVGGYIMDDVLETYTLALVELDAFFRSTEAWLSQLQPSEYPKIEIVEQLQEITVPFSRAVVCMTDYFNQCCALSKI
DEISPSNVHRHYRRLSASVTAYLEMAEKQRDTAVKLDEKEFLWQRSADNLGFPVLMFIEVFSNLLQEAELAKIPTTIFHY
FHMYSNLFAHVLNDLNSFHRDIDTDHNNLIKWLQINGLASDFEDAACKVVQFLSSAIIRMYKAFAECSKQFPNSPALER
FLESIGYSFNGVIVVHTTTVERYRLSPYQISLTIVQDQKQDEWLKNESNFGKRCVQLFEELMKEREEMIEIYGLSN*

>Dgem cbTC

MSKNFNIGAVVYPVDWCSP LTHKSSDTTELKKGIVWCVSTGIASKDVAEKAVVQLNPFQYMNMLFPKVDPSDSNSAE
LYKINASLIVAGYIMDDVLETYSLEALEELDKFFRATEEWWSKLNPVDYPNLGRISSETLSEIEMTYSRAVICMFIDNKNYSF
EIWRNYKVSLSKVSQFRRRLSASVTAYLEMAKRKRVRSENIDEKEFLWRRSADNLGFPVLMLEVLNSVLNKENIDENIPST
TFHYFHMYSNLFAHVLNDLNSYHRDIETNHNHSLVKLWMLKDGVAVDFNDAATKIVQFLNSVIIRMHKAFQETERRRFECP
PLQAFLRNTAYTFAGWILVHTTAVERYKLSYQTTLAKVGASRFGEWLESESDFGNRCVQLFDELMQEREEMVKIYGM
DGLN*

>Briareum_cembreneA_synthase

KTMFSKPIHAPKKWVEKGKGLLDEEVDEQLEGLQEIIDTFIQCDLCEDETTVRKYRGKLRPNFTRSLNPIPETKLGREVFK
ALAHFTMGIFIAADDIMEKQPEEEIRKMCTTFAHLDAKSQSFPSFPAIDDIKEVVDASSVSKNVVGPVIFLEQFLNKQASLL
LSHGDFSLSDVKEFRLRGYNMMAIYFLRVLEETEIQKKPGSIEVIWRRAFGGCAPFYLLTAEIFSGAIGKLKEHYALTLEY
VLGGIYVATLNDVFSYREAHTVADNVVKAMVSQKEVETLEEAANVAEFLDSIMQYLYEKVEELQRKDPNNRELHVCL
DYVGRMTVGSVYTHLLPRYKDTTLYYLEEVKEDQLSSWLSKKGEYGTMLKNILEVIHQRMADGQMEALAGAFPYCG
KFLGQN*

>Sarcophyton_cembreneC_synthase

KTMKQQTQKVKMSSDHNLRFPSDWTAPHKKILATEECDELIALNELLRWVNETGVTEETTKIAFKKINAYFYLRCLYPVL
PDDPKSKKLFQVNLQFLVVGYYIDDVLEKYNENEMKELISSYNSLKNQASEKFPKYPSLSEMKGLVSVYQKEYTKKAVVA
YVDYLNKATILLEGGELTEDVMNFRTRLSNVIVLYFDALLSKTKTGCLVSERELLWRRCFDGIPIPVYMFTEVYSKILEKN
YVLPDLVFKLYLLGSFFSVINDLYSYERDKSDKTGETIIVWFHEKTAEDITAATLKISKILNATLQKMYLLVEEGKTSHP
ISEWFHGWVLLTVGWIHIIHNAVPRYACSSQITPVEIEDDMISDWLLEKDVYGGQVQVQEFLENIDDP SRKTTMDFLCGS
AFDLAC*

CYP450 protein sequences tested in this study

>Dgem cbCYPa

MELCDVLLYFATFFLILVSILSTYIAYAFGYRRLKYNHIPPPLLFHFIKGHAGKILKVIDDGYPLDEYLLDCHKIYGNITLWVL
FHEPLISVTSPELVKQTLVVENFPKSSKLFACGVRLFGERLAGHGLITNFDHENWVKRREFMNPFAFHRRYLRLNMPVFN
SCSKQLERKLESIAADGKTKVKLLEELGPLTLDVICQAGYGANFDIVNNPNSQFPTAFNLTMQGMEMFVFNPFHRIDFRTF
GFQRKVIDACRLLRKTGKDLIEQRRRDIQIGKDVNDILSFILKSADKDMFLTDEDLIDEFVTFYAGHETTARTLAFCLME
LAFHPEVKAKLVEEITSVLGKNDDVSFDDLAKLEYTWMCIRETRLRLYPAAAGIRRFIDKPTVLDGISIPGDDLYVSSYSAR
HPNHFDDAEEFKPERWSSEEVVTSRLTAFPFVSGPRQCIGQTFAQIESKILLSRLLRKVIDLVPGHKKITQRIQPRD
GIECVLTERM*

>Dgem cbCYPb

MFFEVAASLLVWVWVMTTYQSRRNMPPGPYPLPIGNTHQLGPDPPFTMDKLWEKYGDVYQVRFVPGTFVIVNSF
EAAREALVTKDDFAGRPIDLMYPVDIIEAKSVASVDFGVKLMFRKKIMKSALRVFGDGLTQVQDRVNSAVKELLEIEK
TNGKPPFVKHFVSAALTTQLWGWISSQRYPGHRILNALIEFQEKMIYLMRQGSYFQLLPMKFFPTKFMRTLKEVLQM
REDIFGAEEHLRTYNKETCRDIIDSLISSYIEKSKTTTKDIGSIDDIKFLLDVAVAGADTSITAVSWFIFYVLNKLQEK
LHAELDDVVGDRDLPCWSDIKNLSYLQAVVCEVMRHTGFLPWPMPHKTIRDTTIQGYHVPKDPVFINFYRIHRDPPEEYH
EPLKFNPDRFLNSDGTGTVGWTAVSAFLPFGVGRRACLGQDLGKMQVFSVISCLLHQFVFLADNESKPSLEDGMLGS
LRYPDYIEIAKKRM*

>Dgem cbCYPc

MIVEIITSLIAVWFAYFFIKTYRERKCKPPGPFPLPFIGNLHQLGSDPPFTMDNLWLKYGDVYQVTFPVGTFVIVNSGEAAR
EALVTRKDDFAGRPIDLMYPVDIIEAKSVASVDFGVKLMFRKKIMKSALRVFGDGLTQVQDRVNSAVKELLEIEK
NGEAFYLQDQVAAAIASQLWEWLSSKKCSFDDKSLKEIVQFNENITSLARQGNYYQMLPLLKYLPFKMKNIQDVTVKLR
DKIFSRLEEHRRTYEGVIRDTDALICSYEMEKSKASRKDIGQIEDLKFLMIDVILAGADTSRSLVWCILFLSLRKEQLT
KLQKEIDEVVGDRDLPRWKDVKNLPKLQSFVCEVLRYSIFPWPMPHKTIRNTTNGYLIPIKYTPVFLNFYRIHRDPKEWN
DPSEFEPDRFLDSAGNFIGWSTVKAFLPFGIGRRSCVGEDLGKMQVFGILSSLIHQFSLPDNHRPSLLETVSGFIHQ
PKNHKIIANKRE*

>Ecar cbCYPa

MLILFLVAGLMLFVYIGYLLIWRRLKYRHIPGPPLAHVIKGHAGDMLKKIADGFPLDEYLLDCHKSYGNVLIWLYHEPLL
SVTSPAMVKETLVLNRFPKSSKLFQCGVRLFGARLAGHGLITNFDHENWVKRRELNMNPAFHRRYLRLNLPFINSCKKR
LERKLESSANGVTKVRFDELGLVTLDMVCLAGYKDFDIVNNPNSQFPKAFNLTMEGMEMFAFNPFHQLDVRTFKFQ
REVIEACRFLRKTNEMITQRRQDLLNGVDVPNDILTFLKSADKATDLSDEDLIDEFVTFYAGHETTARTLAFCLMELAF
YPEVKAKLLEEIANVLGERDDVTENLPKLEYTWFICIRETLRLYPAAAGIRRYIDKPSILNGIKIPGKTDLVSSYSARHPD
HFDHPPEEFIPERWASHDIVTSSGLTSFPFVSGYRTICIGQTFAQIESKILLARILRKFSIQLVPGHKKITQRIQPREGIECSL
KRIHHHHHHH*

>Ecar cbCYPb

MRKNAHSWQTMFLEVCLSLTVWVFLWYMMTTYHSRRLTPPGPFPLPLIGNTHLIGSNPPFTMDKLWETYGDVYQVKFP
VGTFVIVNSFEAAREALITRKDDFAGRPIDLMYPVDIIEGKSVASADFGKLMFRKKITKSARVFGDNVHVQGRMTNA
VKELLSEIADENPFVSYSIAITTLWWEISSRQYRHGNKTLRSLVEFQEKMIYLMRQGSYFQMLPFMKFLPTKFMRL
TLQEVLEMRETIFGAEEHRRTYTTGVTRDIIDSMIASEIEAKTTDKDVGTVEDIKFLLLDAVAGADTSITVSWFILHM
ILDKDLQTRLQNELDDVVGDRDLPCWGDVKNLHYLQATVCEVMRHTGFLPWPMPHRTIRDTSIKGYHVPKHTSVFINFY
RVHRDPQEFCEPMLFDPGRFVNSDGTGKWTAVSAFLPFGVGRRACLGQDLGKMQLSVLSCLLYQFTFHEVESER
PSLEEGDTGSLYHPTDYKVIARRDHHHHHHH*

>Selo cbCYPc

MLLEVIAIATTVLWYLGKTYQARKWGPPGPFPLPGIGNLHQLGASPPFTLDHLWQTYGDVYQITFPVGTFFVIVNSGR
AAREALVTRKDDFAGRPIDLMYPVDIIEGKSVASADFGKLMFRKKITKSARVFGDNVHVQGRMTNA
KRKEAFYIQDISSAIAQLWEWLTSKTSFDDASLHEVIAFNENVTSLARQGNYYQMLPLTYLPTAFVKQLQNTLRMR
DDIYACELDEHRLTYVDGCIIRDVTDALISSYHVEKAKCSDREVGPIDELKFLMIDVVLGADTSRTVVAWCVFLLVQKE
LQNELQEQEIDRAADEHYLLWEQVKNLPKLQSFVCEVLRYSIAIFPWPMPHKTIRDTTLNGYHIAKNTPIFINFYRIHRDINE
WDEPEVFKPHRFLDSGDNFVGWAKKGLPFGIGRRFCLGQELGKMQVFGMVANLLHQFTLELPAGDPIPSILDSKKS
FVHQPPSHRVIATKRR*

>Crub cbCYPb

MFLEVAASLLVWVWVMTTYQSRRNMPPGPYPLPIGNTHQLGSDPPFTMDNLWEKFGDVYQVKFPIGTFFVIVNSFEA
AREALVTKDDFAGRPIDLMYPVDIIEAKSVASCDYGLKLMFRKKIISALHVFQDNLQVVEERVNSAVMELLTEIQEMN

GDAPFIKYFISAAVTSQLWGWISSERYRFGDKRLHSLIEFQENMIFLLRQGSYFQMLPLMKYLPTKFMKTLQEVLVMREEI
FGAELEKHCRTYNEDTSRDIIDSLISSYNKEKSKTSARDVGTVEDIMFLLLDVAVGADTSITVISWFILHMILKKDLQTKLQ
KELDEAVGRNRLPRWDDIKSLPRLQATVCEVMRHTGFLPWPMPHRAIRDTTIQGYHVPKNTSVFINFYRIHRDAKEFDEP
SLFNPDRFLNSDGTGTFVGSWISAFLPFGVGRRACLGADLGKMQVFSVISSLLHQFTFDVNDQEPTPNLVDSISGSLRHP
KDFKVIANKRL*

> Crub cbCYPa

MEFSLVLSYSITCFLIALSLLTYVGYTFGWRRMQYRHIPSPPLSHFFKGHAGEMFKYMKNGYPLDEYLLDCHKKYGNILI
WYIHEPLLSVTSPALVKETLVRNFPKSSKLFQCGVKLFGARLAGHGLITNFDHENWLKRRREFMNPFAHRRYLRNMVPT
FNRCCERLERKLQRIANGVTKVRLDELGSLTLDVICEAGYGKSFDIVNNPKSQFPTAFNRTMEGMELSVFHLHRIDFR
TFPFQRKVIEACHFLRKTGKDLITERRRDIQNGYDVPNDILTFILKSADKGLLSDDDLIDEFVTFYAGHETTARTLAFCL
MELAFHPECKVKLLEEISRNLGKGDVDSFEDLPKLQYTGMCILESLRLYPAAAGIRRVIKTPTVLRGLHIPENDLYVSSYV
SARLPEHFADADQFKPERWSSEEVVLASRLTAPFSTGPRTICIGQTFANMETKILLSRLLRKFSIELVPGHTKKITQRIQIP
RDGIECTLTNKV*

>Rkoe cbCYPc

MFLEVVVALMAVWLIHFAVKTWRSRLGDPGPIPLPLIGNLHQLGSNPPYIDHLWRRYGDVYQITFPVGTFFVINS
KTAREALVERKDDFAGRPTIEMYPIDVILEGKNIGNGDYRPLLLFQRKVLRSALHVVWEGAKRAEARNRAVVTLLRETE
AKNGAPFHIQDYVVSIAIASQLWEWLTSKRTDVRQESLKEVIRFNEISITSLARQGNYYQMLPLTYLPTFTKKIRETVELRD
KIYARELDEHLASYVEGRIRDVTDALIAAYKAETAKGLDKDMGTIEDLKFLMIDVMLAGADTSRTVVSWSILFVSDTELQ
KLQEQIDRVFGGEERLLAEDSKCLDRLQSFVCEVLRVFSLPFWMPHRTIRDTEVGGYKIAKDPVFVNYRIHRDPEEF
EAPDLFLAERFLDTGGHFVGTAVSGFIPFGAGRRSCSQDLGRMQVFAVLANLIRQFRLESPPGEPAPSIDDTVSGFI
NQPKSHRVAIKR*

>Rkoe cbCYPb

KTMILEVFLALSVMWLTWYYITVFYSRRGNAPPGPIPLPFIGNALQLGQHPPFTMDALWRAYGDVYQIPVGTFFVINS
YQATKEALVTKRDDFAGRPIDIMYPVDVITQAKSVASADYGVKFMFRKLLKSAFRVFGESLERVEERVNSAVVELIGVVE
KSGGEPLYIKHHISSAITAQLWDWISSQRCRFGDETLSLIEFQDKMIFLLMQGSLHQMLPFMKYLPTKFNRYLADVLR
MREEIFGTTELHQHLQTYNNDPRDIIDNLIASIEAEVDKGTTRDIGSIEDIKYLLLDIVVAGADTSITSVSWFVLYMVLYRDR
QKTLQAELEDEVGVDRLPWSDIKNLHRLQSAVCEVMRITGFLPWPMPHRAIRDTTVQGYRIPKGTSVFINFYRTHRDPK
EFAAPEEFRPNRFLDADGKFGVGTGVAAFMPFGLGRRACIGQDLGKMQTFSVISSLLHQFSLALPELAPTSLDEGKLG
SLRHPVDYHVIATRRH*

>Rkoe cbCYPa

MEYTTILWLLALFCCLVLVVMGYCYAFIYRRYLRHIPGPPITHFIKGNVEMLMKIANGYPLDEYLLNCHKAYGPVVC
WLFHEPLSVTTPELVRETAVRNFPKSSKLFQCGVRLFGARLAGHGLITNFDHENWLVRRTELNPAFHRRYLRKNMIPF
NECCAAEFENKLARYADGRTRVKLLDEICPLTDLVLSRAAYGTDQIVNNPDSFPPTAFNLTMGSMEMFVFNPLHWLDF
RTFPFQRKVVEACKFLRKTGNDMLVQRRRDIEAGKEVPSDLLTFIVKSGEKLSDILTDVDLDELVTFFYAGHETTARTLAF
TFMELGFHPEVMEKLVSEITSVLPPTGDVTMEHLNKLQYTWLCLRETLRLYPAAAGLRRYINTPCELGGYQIPDKTDLYVS
SYVSARMPGNFEDAKEYRPERWLSEEVVNSRLTQFPFSGYRTICIGQTFAMMETKLLARILRKYVIDLVPGHSSKITQ
RITIQRDMVECFVSKREWPPQTREPNEKSP*

>Basb cbCYPa

MEFSNALYFVILLVIVVTTSLTYLGYLFGWRKLKYRHIPGPPLAHFIKGHAGEIFKVIDDGYPLDEYLLTCHKTYGNVVLW
LFHEPFLSVTSPELVKETLGARRFPKSSKLFQCGVKLFGARLAGHGLITNFDHENWLKRRREFMNPFAHRRYLRKNMIPF
GCCESLERKLASSVNCTTKVRLDELGYLTLDVICQAGYGKDFDIVNNPSSKFPTAFNLTMQGMEMFAFNPLHWIDFRT
FPFQRKVVDACHFLRQTGKELITERRELLKGGADIPNDILTFILKSADKNMILTDEDLIDEFVTFYAGHETTARTLGFCML
ELAFHPEVKEKLEELSVLPETGDISFDDLPKLEYTSMCIRETLRLYPAAAGIRRFIDKPTVLGIFIPENTDLYVSSYASARH
PNHFEDAEEFKPERWKSQDVLSSRLTAPFSGYRTICIGETFANMEAKILLSRLLRKFEHLVPGHTKKISQSITIRPRDG
IECRISKRA*

>Basb cbCYPb

MFLEVVVALVWTLWYLIHTYQCRNMPPGPFPLLLGNTHQIGTDPFMSMDHLWKKYGDVYQVTFPIGTFFIINGFD
AAREALVTRKDDFAGRPIDLMYPVDIITQGKSVASADYGLKFIFRKKIVKSALHIFGNLSQVEERMRSVAVKDLLTEIDETQ
GRPFSLKYYISAALTSQLWGWISSQYRHHGNETLQALIQFQEKMIFLMRQGSYFQMLPFMKFFPTKFTKTLKDVLMKREE
IFGQELNEHLRTYNNEETRDIIISLIFSYTEKKKTTVKDVGTVDDIKYLLLDVAVGADTSITSVSWFMLYLVSKNIDQSKI
HDELDKVVGPDRPPRWEDVKLHYLQASVCEVMRITGFLPWPMPHRTIRDTTVQGYHVPKNTSVFINFYRIHRDPKVYS
EPSVFNPDRLNVDGAFNGWTLPAFVFPFGVGRRACLGQDLGKLQVLSVSSLLHQFTFDADNQHPLPSLDDGTGSG
LRHPIDFKVIARRQ*

SDH protein sequences identified in this study

>Basb cbSDH2

MVQYLGVVTGGAQGIGLAIVEEILQSGGKAYFFDVKEEGKQTTEAELQSKYGKNRVFLKCDVANAEDEFKGAIDKAKKE
MGGLNVFCNNAGFARPFYDISRWEQEIDVNLKASIRGSLVALDLMSAEKRGLGGVIVNTSSSFVALNPVPLFPTYCAAKA
GVIAFMRSMAPLARAANVRINCPCPDATDTPQYQKTLPELAATFSQEYKSFVESMASGGGKQAISPAQKPKDVAVG
VMKIIEDETKNGEALVVSKEHGTYHYHIFK*

>Selo cbSDH1

MACTEYLAIVTGAAGGIGRAIVSALLEAGGKVHFLDVKVDQGNQTQKEFDKQFGIGRAKFVVCDSVRTQLQKAFQDA
KLDLGLNVVCNNAGIFSPVFDRTWRWKVAVNLTATIHGTLGLDMMSTNKGKGGVINTSSVEGLRPFPPIGPVYSA
TKYGVIGFSRSLAPVVAENVRINCICPDGTNTLVANSVAELLPVFSIDEIQSLLTVGRGSNWADAKLEPKENAPALAS
SVVGEEVLKLVKDETKNGEALVSKVLGTYFQKFE*

>Dgem cbSDH1

MTTTYLAIVTGAAGGIGRAITSALLEAGGKVHFLDVNIEEGKAAQNDFDKKYGAGQSKFVFCDDVSDKDQLEEFREAKS
ELEGLNVVCNNAGITCPVFDRTWRWPRVVEVNLSTATHGTLGLDLMGTSGGGTGGVINTSAIAGLRPVVPVAPVYSATK
HGVIIGFSRSLAPVVAENVRINCICPDGTNTLVANSVAELLPVFSIDEIQSLLAHVYGMDSAGQGPRVSADQSDTLPL
SAIGEQLVLELVKDESKNGEALVSKVLGTYFQKFN*

>Basb cbSDH1

MADPYLAIVTGAAGGIGYAITVALLETGAKIHFLDVNSEEKAAQKTFDKKYGAGKAKFVVCDSVDRALQKAFKDAKS
QLGGLNVVCNNAGITCPVFDRTWRWPKVSVNLSAAIHGTLGLDLMGTSGGGTGGVINTSAIAGLRPVPIAPVYSATK
HGIIGFSRSLAPIASDNVRINCICPDGTNTLVANSVAELLPVFSVAEIQSLLAHVYGIDTNNKQSTQSSNDMSSSLLPL
SVAAEVLKLIQDETQGAIVVSKVLGTYSHDFK*

>Rk cbSDH1

MATYLAIVTGAAGGIGHAITSAILQSGGKVHFLDIEEELGIKAQAEFEKYGVGKSKFVMCDVSNREELKEVFIAAKREM
GLNVVCNNAGIHPVFDRTWRWPKVVDVNLTAIQGTLLGLDLMGTSGGGTGGVINTSAMAAGLRPVPLAPVYSATKH
GIIGLSRSLAPMSIADKVRINCPCPDGTNTLVANSMAELLPVFSVEEIQKILVKHIYQLDLNEETMQAGPDLKGPADAL
LRPSEVGEEVFKLIQDDSKNGEALVSKQLGTYWQFN*

>Ecar cbSDH

MVEYLAVVTGGAQGIGLAIVMELLEAGGKVYFIDMDEEGKKAEDLQHRYQDQVRFKICDVSNEKEFEASIRKAREE
MRGLNVCMNNAGFARPFYDMSRWELEIDVNLKASIRGSLIALELMNPENNGSGGVIVNTSSYTLNPAPLFPTYCAAK
AGVIAFARGFAPVAVTSNVRINCPLDATMTPMYQKTLPELASTLSQEFQSFMTMTSSQSGSKQAISPTQEPRDVALG
VIKILQDETNGEALVVSKEHGTYHYHIFK*

>Rk cbSDH2

MAEKRVELERLASSLVAIKWTIVGIALLMLLKNLLSTKRQTRRKLPKGKVVITGANSIGIFATAMEFARRRRARVVLACRR
DGQRTAARIRRAADNNDVWRQLDLSCCDSVRRFARDVLTSERRHLVNVNAGVMTPPDRHTRDGRDQLAVNYLG
HFLLTNLLNLLKKSDGKILNLTSHYMRATHVDLTDPRPCDPWVAHHRSLAVVMSTRELSRRLRGNTVANVIDPGT
NTSLRRTHWLGFPLRYFLVPVTWALWQTEPHVQGSIANLLASRALDEHSGQYFADGKIQHIAEHALNENAAKQLWE
NSVKLVGLQENAVL*

Other relevant protein sequences

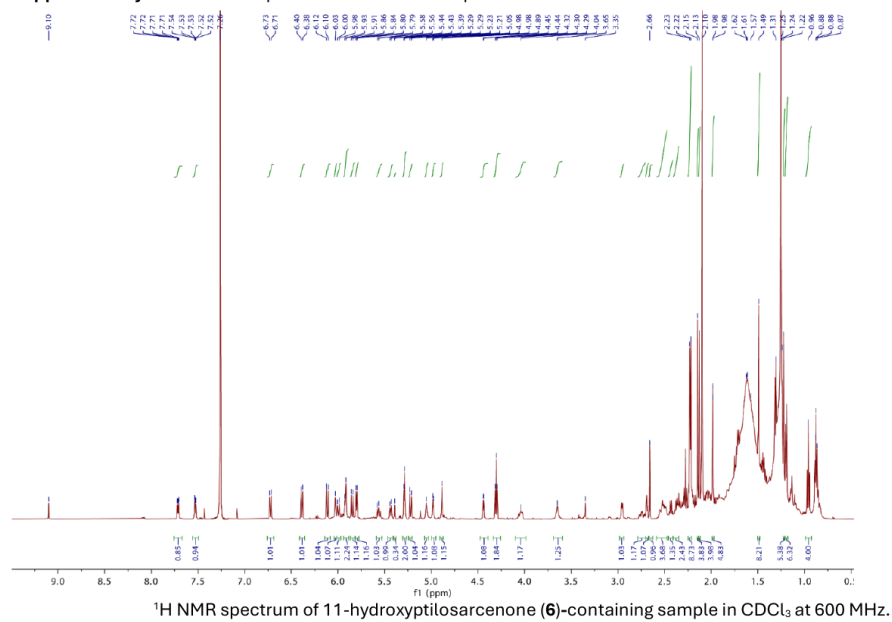
>HcCPR

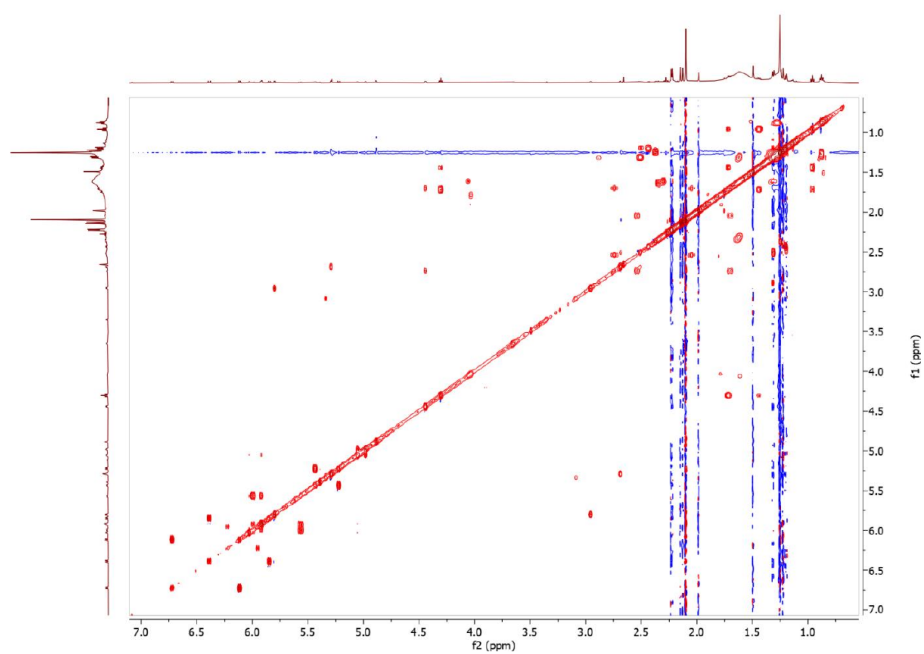
MSSSTYMDDES VGSM SAYPIGYVDIFVSSFIGLVLYWFVVRKRKSGEIQQLPAVTANGNAKLDDSSSIEKMKKGKQV
AIFYGSQTGTAE EFASHLAKDA AKYGMKASVFDPEEFDMENLCRMDEIQDSFAIFVMATYGE G DPTDNAREFFDWLED
DREDLECLKYAVFGLGNKTYENYNTASKFVDRRLAELGCTRLFERGEGDDGDIENDFVTWKEFWPAVLEHYGIDVS
SLTLSKSLIRQYDLVHTDMPQDKVFSGEMARLHSLHLQPPYDQKNPYLAPISENRELHKSGDRSCMHIELDISDSKL
RYDAGDHVAIYPTNDPELVERLGKILNVDLDVVFLVNKDEESNKKNPFCPTTYSALLHYVDIAGKAGNVLIELIDYA
QDPKDKEFEKLTNNNEEGKHLHQEWIMEDHRTIDVLEDLPSVRPIDLLEFLPRLNCRYYSISSPKLHRYHIHVTAVL
VDYNTRLGRHVKGVA TTWLAGKKPKEDNSPRVPIFVRKTQFRLPFKPTTPVIMIGPGTGFAFRGFIQERHSSKKKEGTIG
DTLFFGCRHKAEDYIYEEIEQYHAEGTISHLFVAFSRDQAEKRYVQHLIKENGIVVWNLKNEGHVYVCGDAKHMAK
DVHDALVGVVETHGEMNDSQATDFIKLHSGKYSVDVWVSHHHHHH*

>EcCYtb5

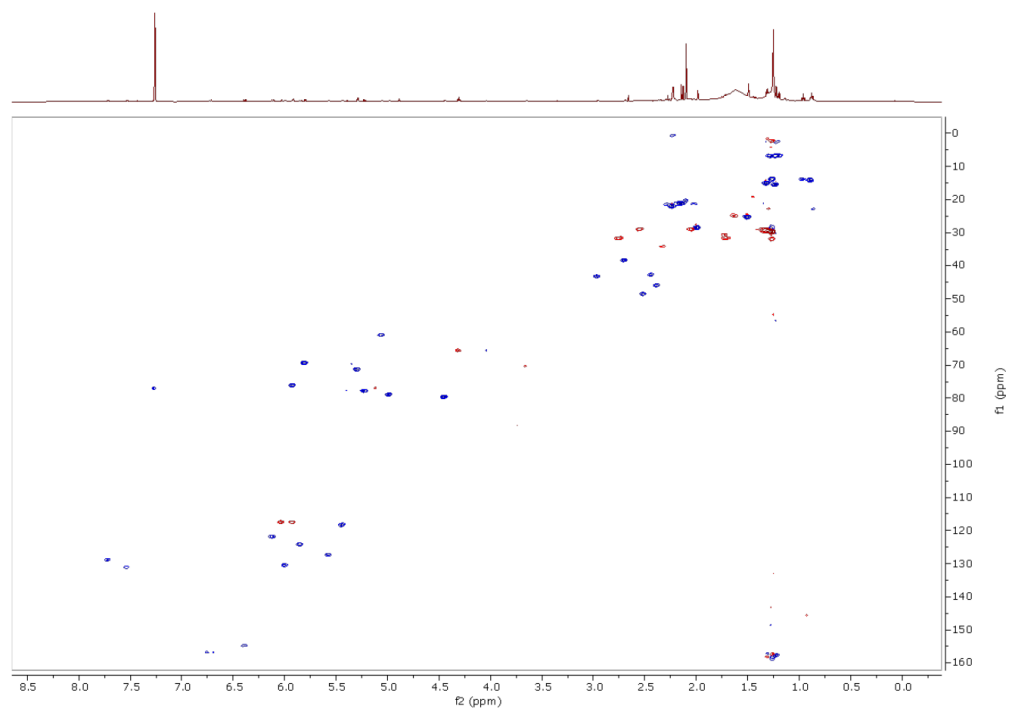
MSTGEEAKKVYTLKEVGEHNSSKSLWIVISNRVYDVTDFIDEHPGGEEVLEQAGHIATEAFEDVGHSEDARELMKDYCI
GDLCEEDKEPEYKKYDGPAAKDDPKSMLSWLIPAAISVGIIAVALYKTISS
>XdCrtE
MDYANILTAIPLEFTPQDDIVLLEPYHYLGKNPGKEIRSQIEAFNYWLDVKKEDLEVIQNVVGMMLHTASLLMDDVEDSS
VLRRGSPVAHLIYGIPQTINTANYVYFLAYQEIFKLRPTPIPMPIPPSSASLQSSVSSASSSSSASSENGGTSTPNSQIPFS
KDTYLDKVITDEILSLHRGQGLELFWRDSLTCPSEEEYVKMVLGKTGGLFRIAVRLMMAKSECIDFVQLVNLISIFYQIR
DDYMNLQSSEYAHNKNFAEDLTEGKFSFPTIHSIHANPSSRLVINTLQKKSTSPEILHHCNVNMRTEHSFEYTQEVNLNL
SGALERELGRLQGEFAEANSRMDLGDVDSEGRTGKNVKLEAILKKLADIPL*

Supplementary Note 2. NMR spectra of *R. koellikeri* terpene mixture.

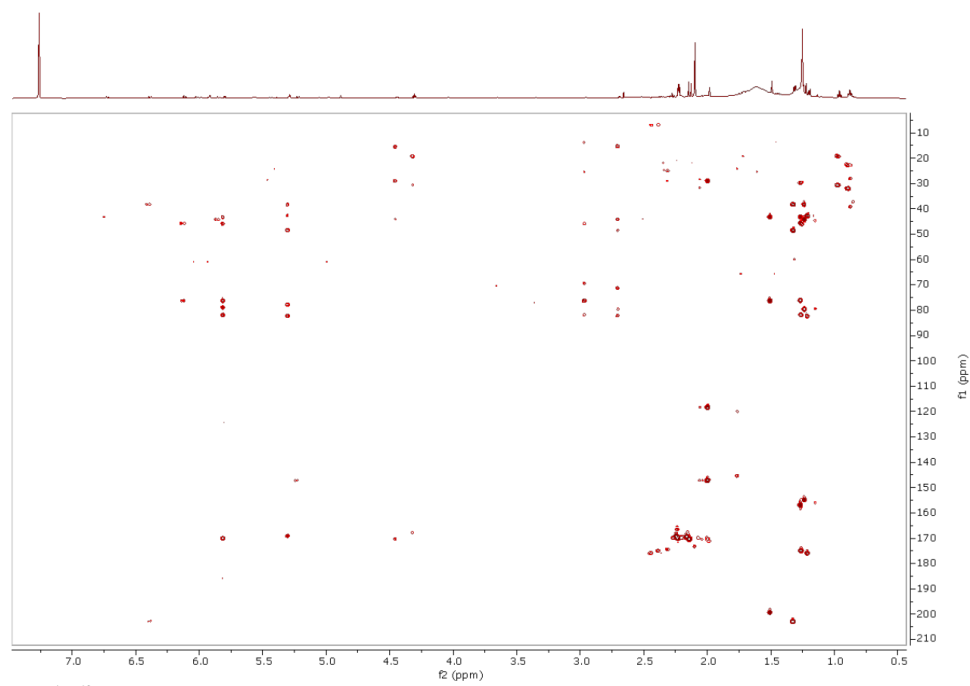




^1H - ^1H COSY spectrum of 11-hydroxyptilosarcenone (**6**)-containing sample in CDCl_3 at 600 MHz.

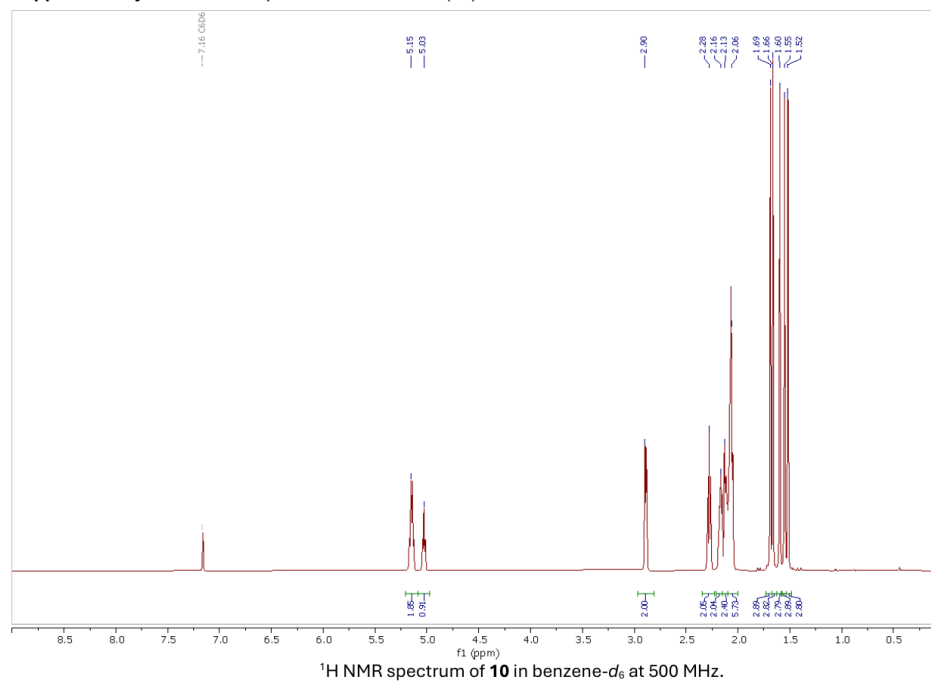


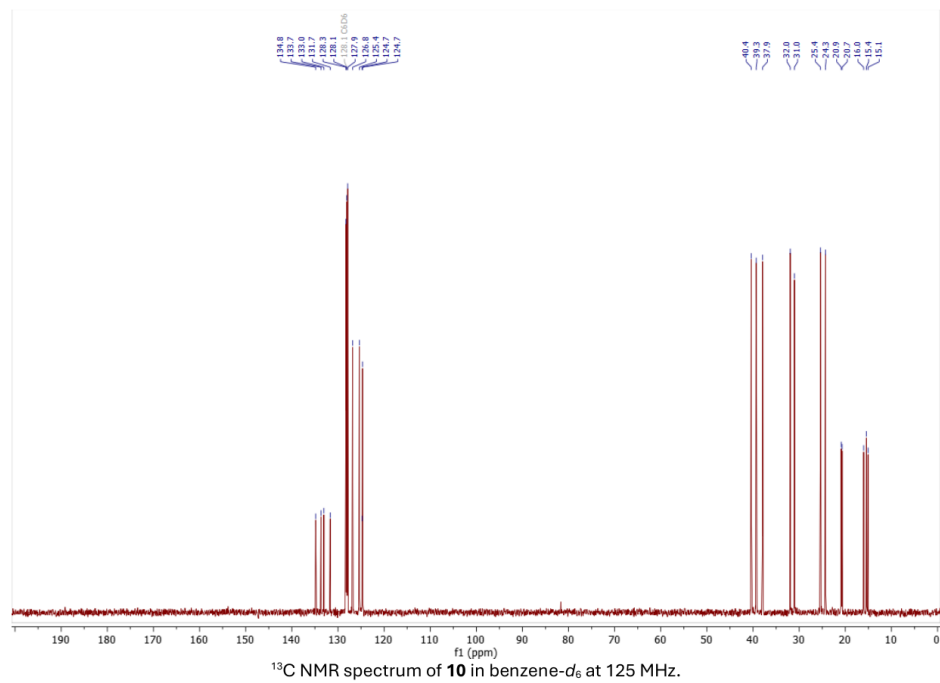
^1H - ^{13}C HSQC spectrum of 11-hydroxyptilosarcenone (**6**)-containing sample in CDCl_3 at 600 MHz.



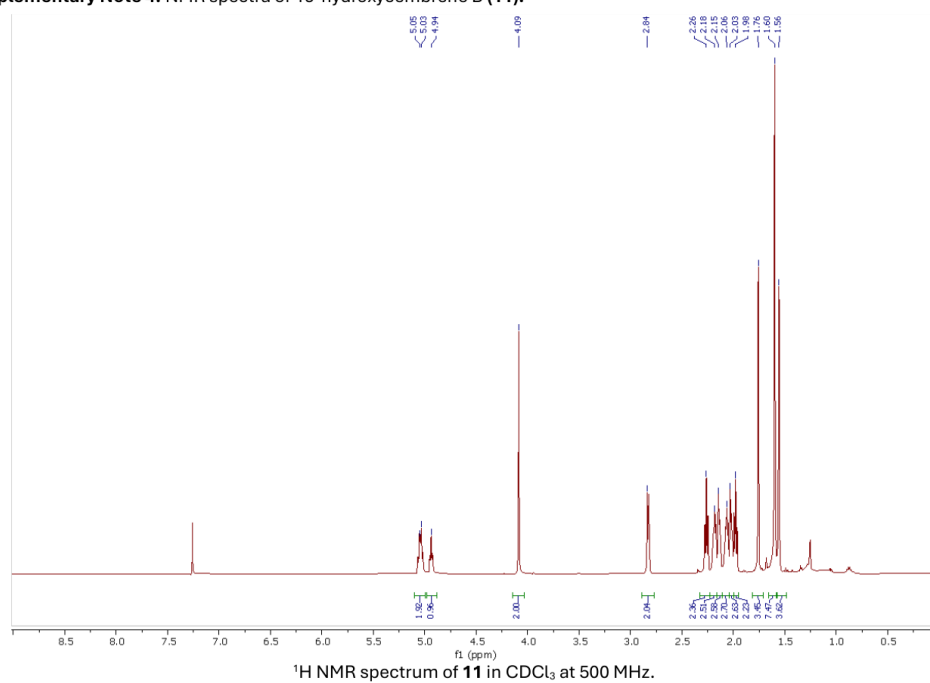
^1H - ^{13}C HMBC spectrum of 11-hydroxyptilosarcenone (**6**)-containing sample in CDCl_3 at 600 MHz.

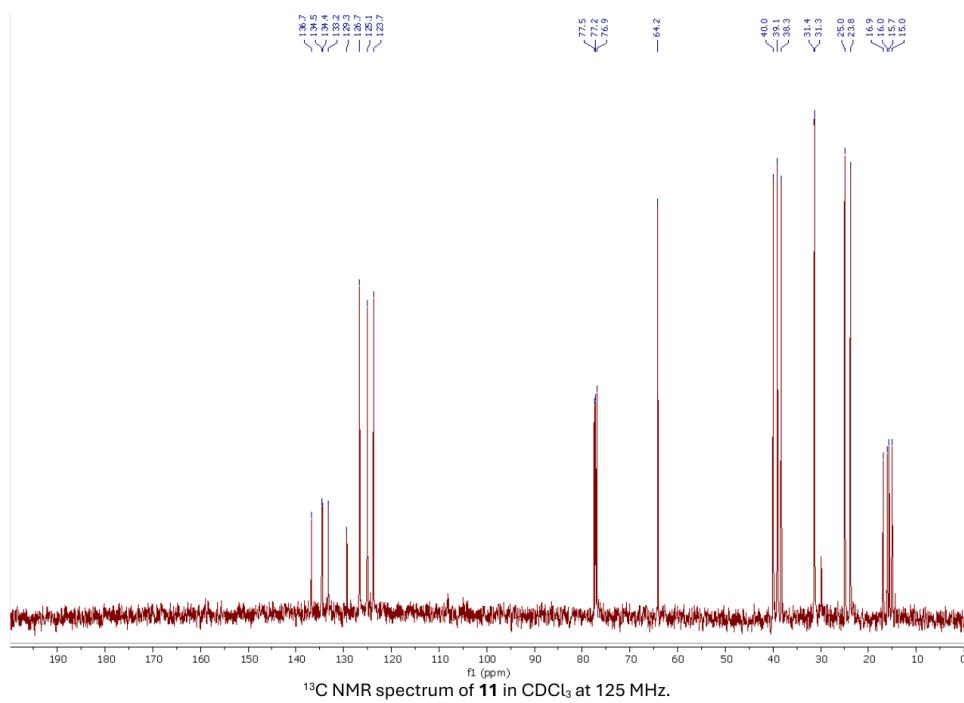
Supplementary Note 3. NMR spectra of cembrene B (10**).**

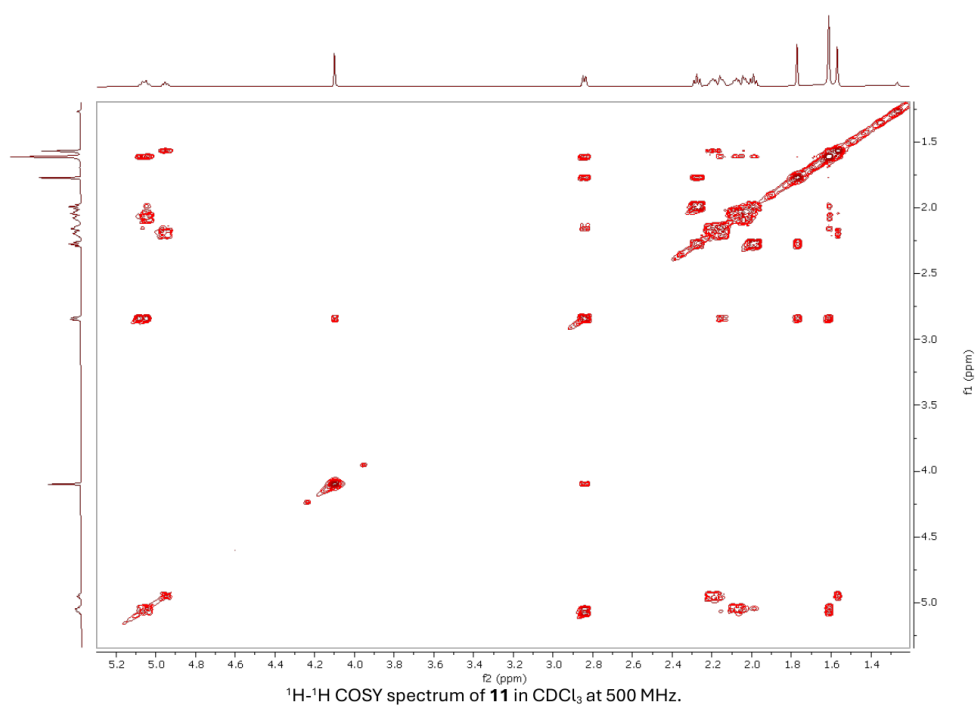


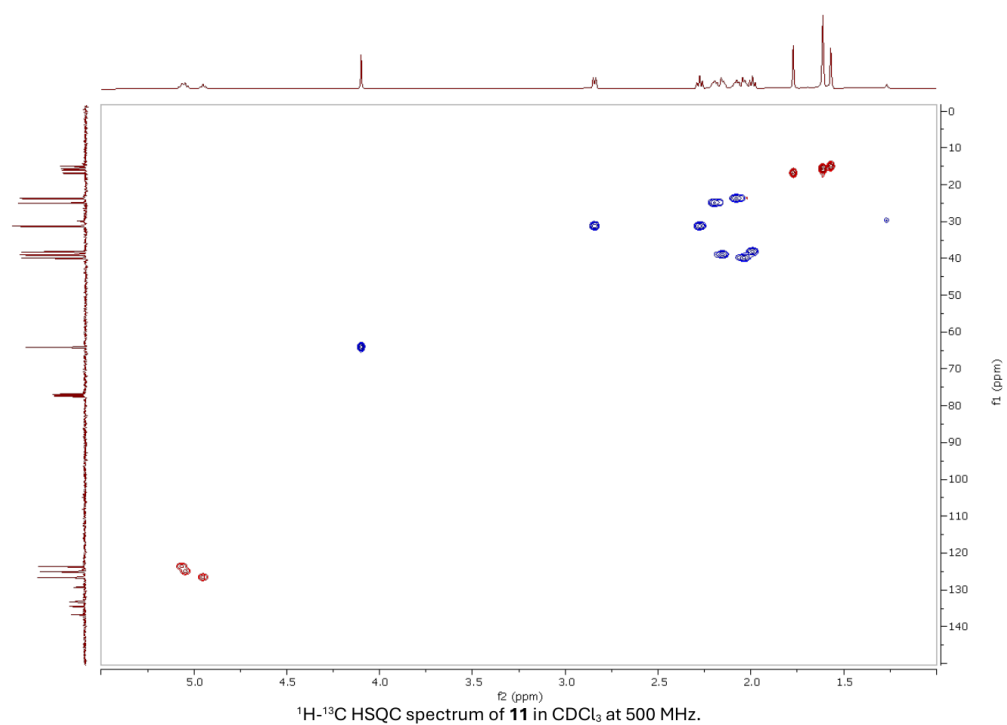


Supplementary Note 4. NMR spectra of 19-hydroxycembrene B (**11**).

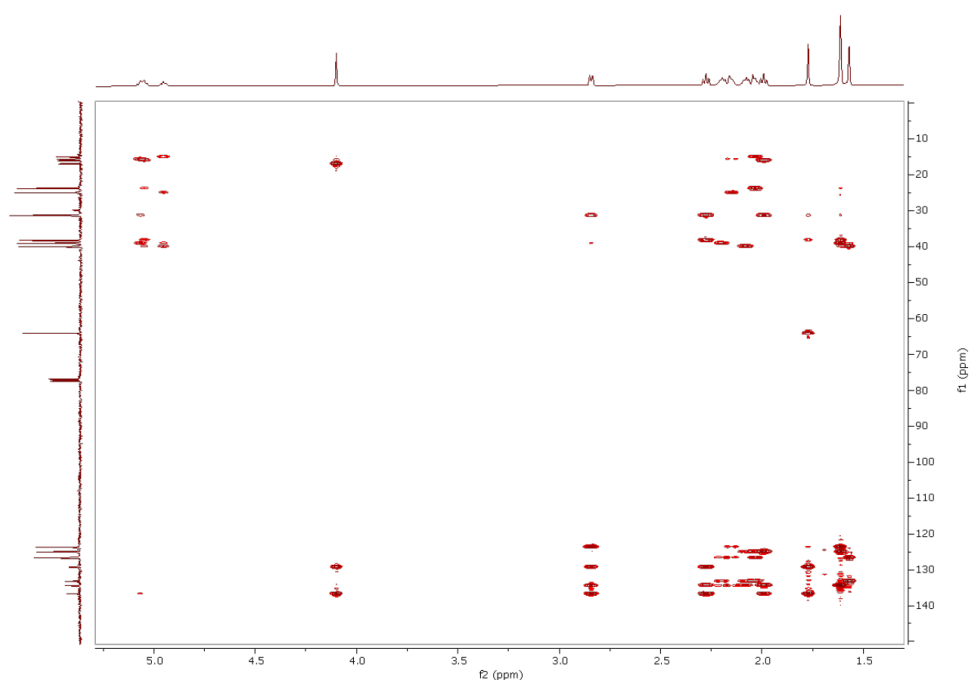




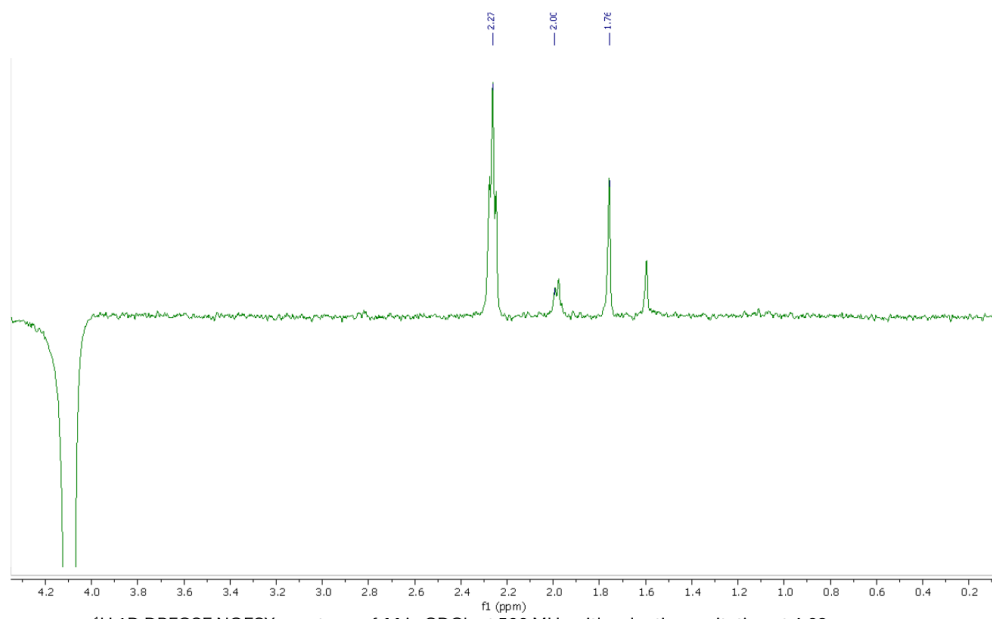




37

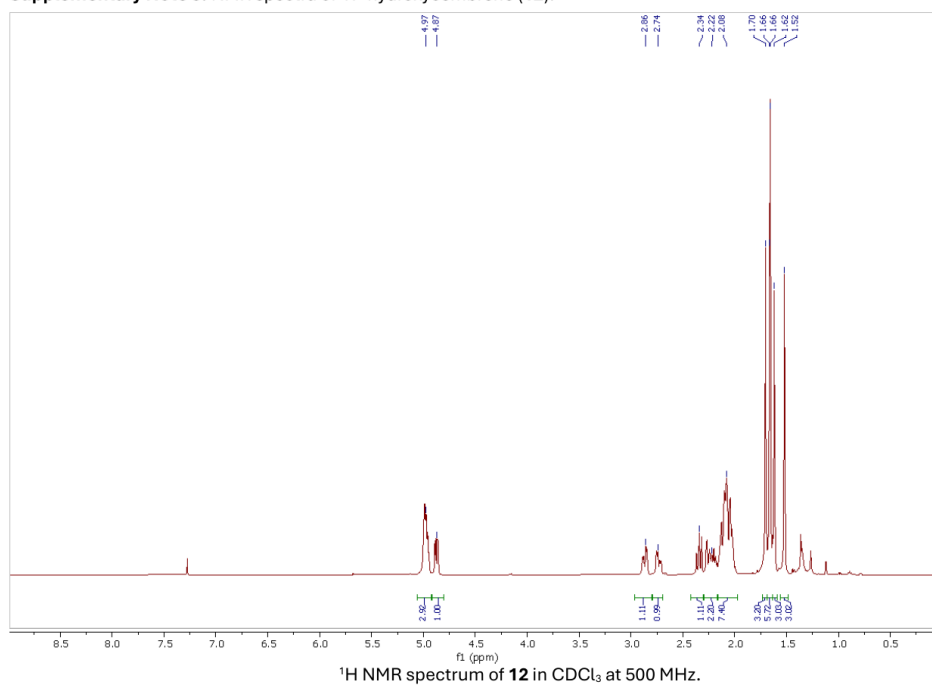


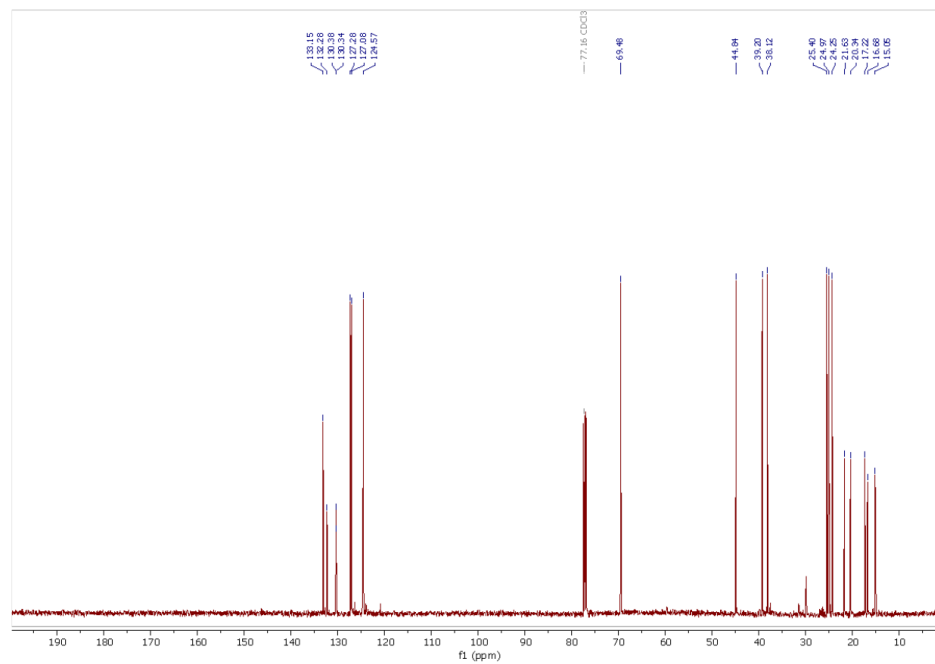
^1H - ^{13}C HMBC spectrum of **11** in CDCl_3 at 500 MHz.



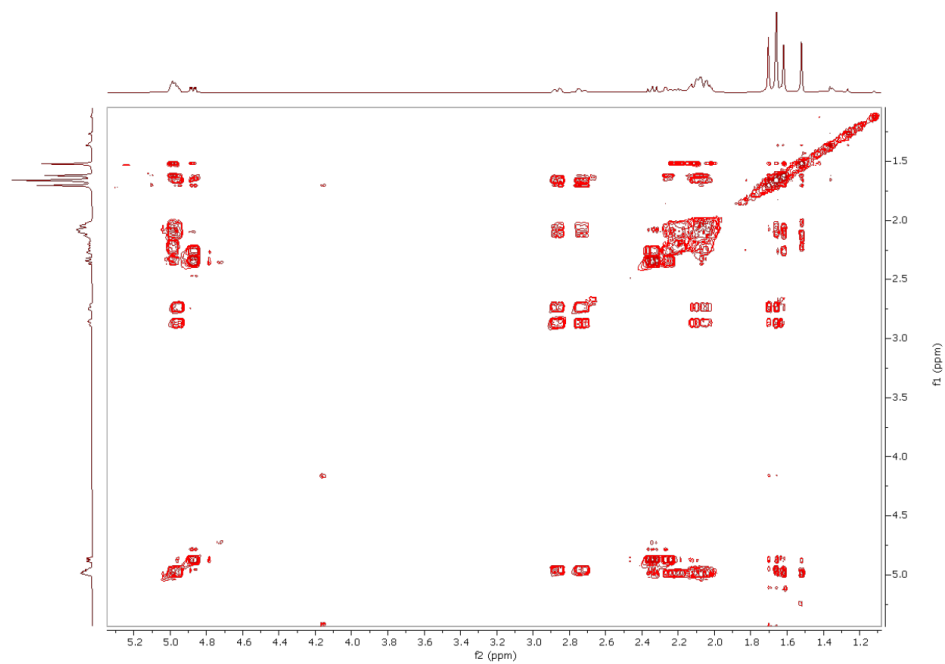
^1H 1D DPGSE NOESY spectrum of **11** in CDCl_3 at 500 MHz with selective excitation at 4.09 ppm.

Supplementary Note 5. NMR spectra of 17-hydroxycembrene (12).

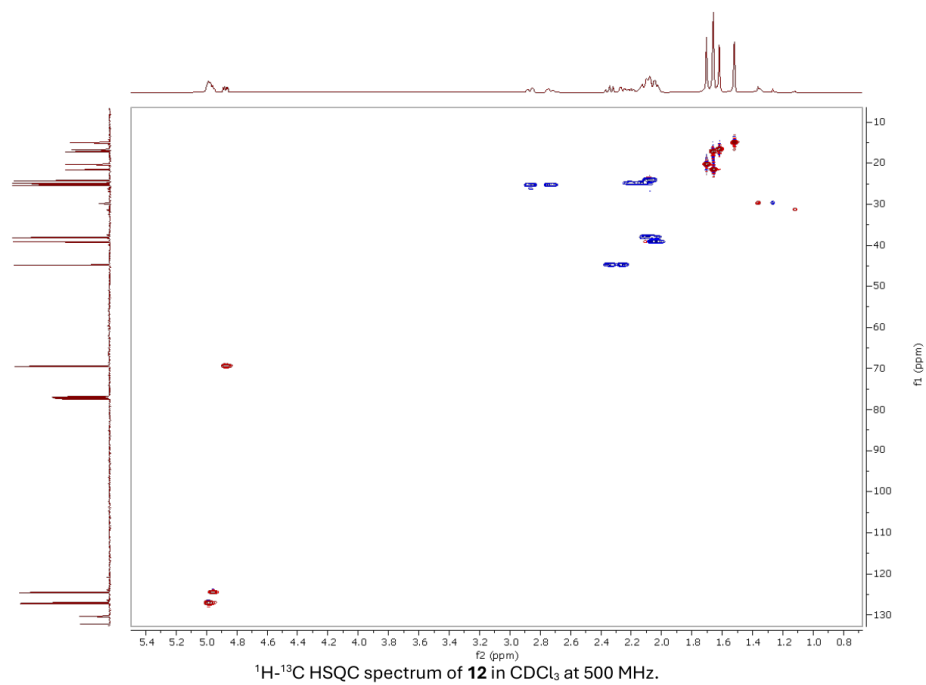


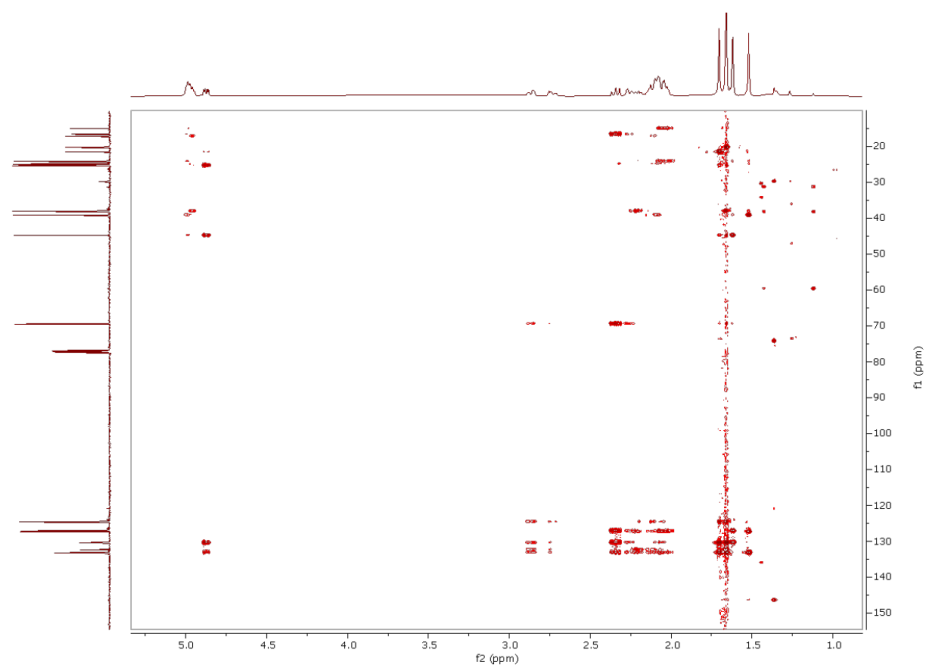


^{13}C NMR spectrum of **12** in CDCl_3 at 125 MHz.

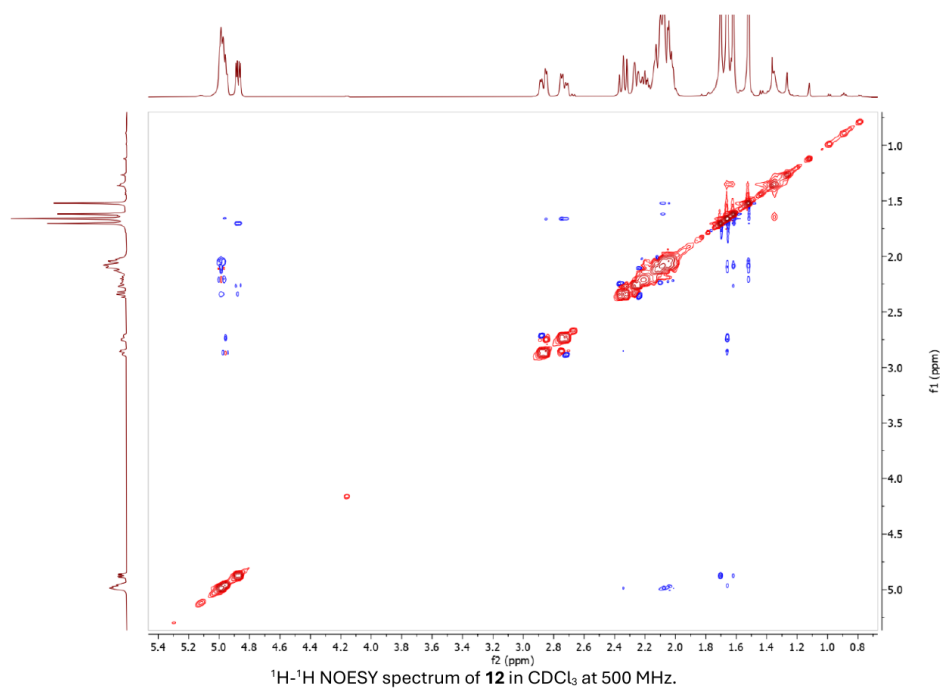


^1H - ^1H COSY spectrum of **12** in CDCl_3 at 500 MHz.



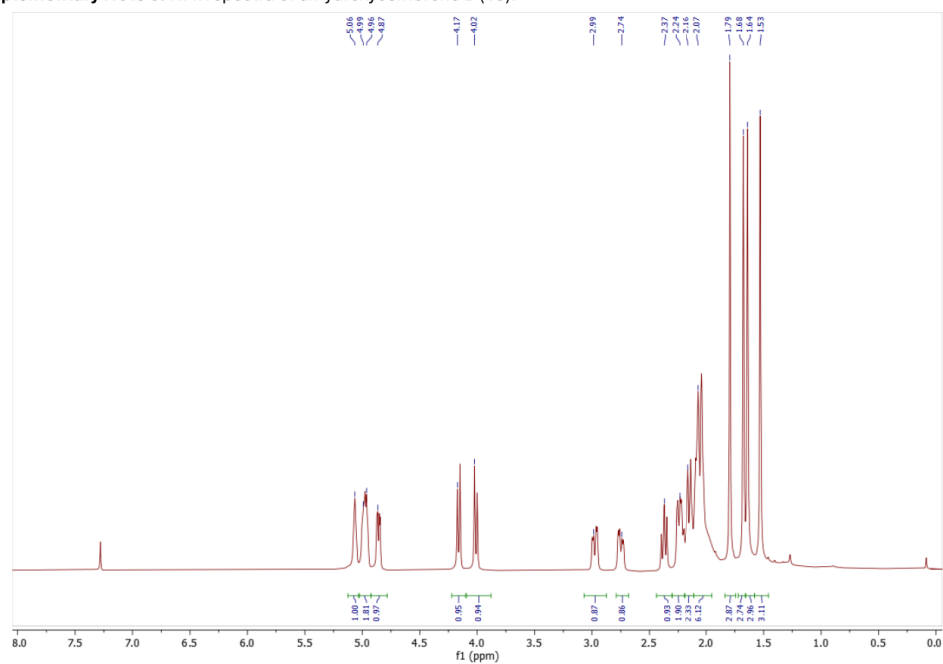


^1H - ^{13}C HMBC spectrum of **12** in CDCl_3 at 500 MHz.

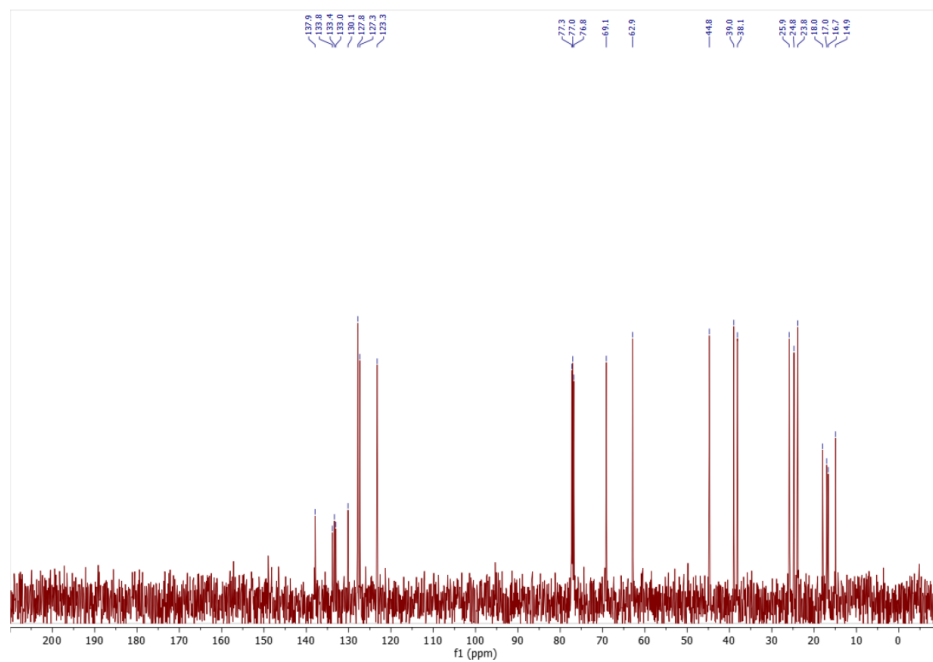


^1H - ^1H NOESY spectrum of **12** in CDCl_3 at 500 MHz.

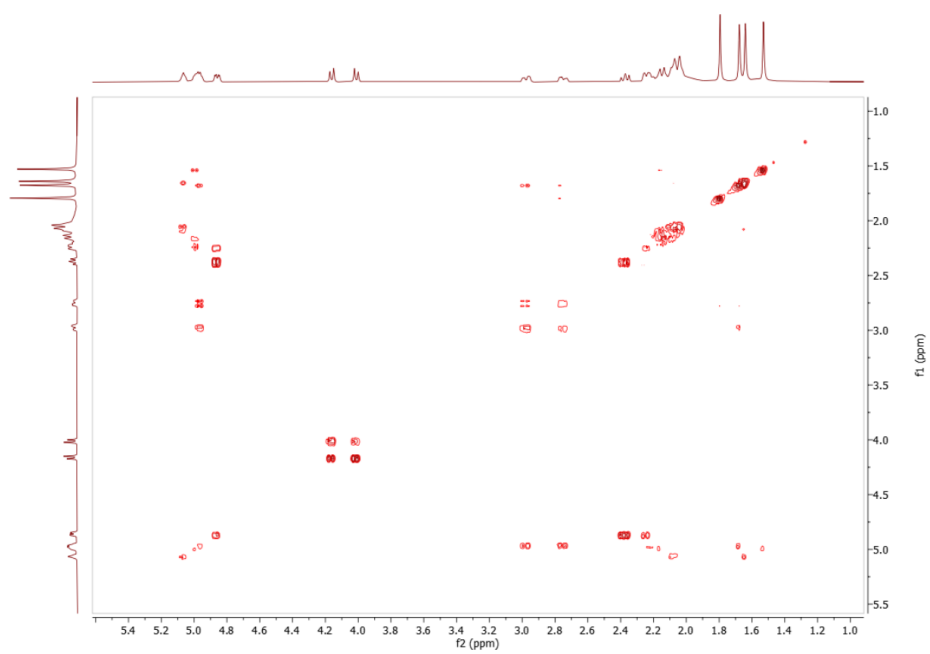
Supplementary Note 6. NMR spectra of dihydroxycembrene B (**13**).



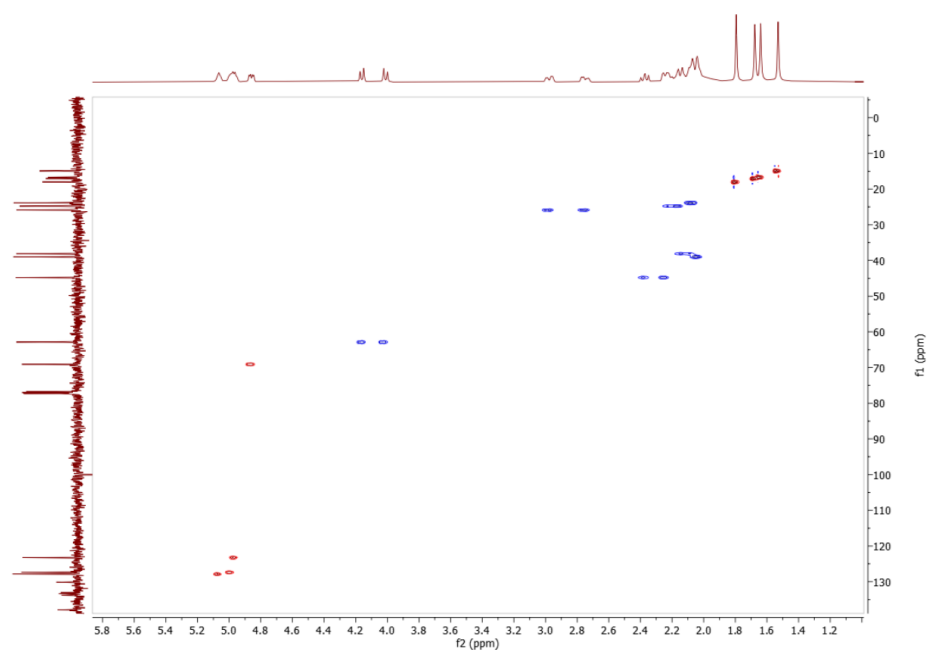
^1H NMR spectrum of **13** in CDCl_3 at 500 MHz.



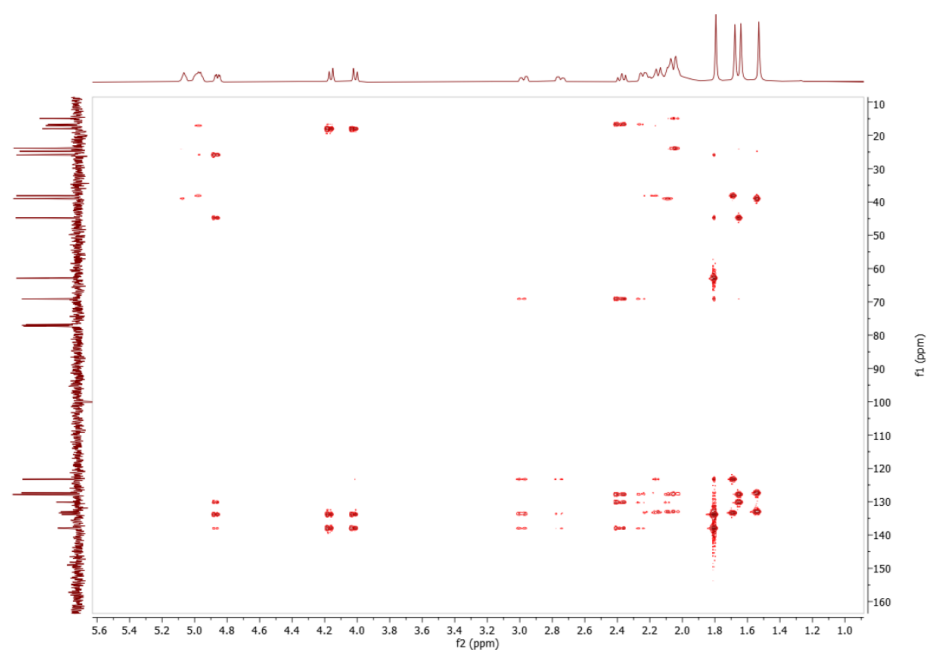
¹³C NMR spectrum of **13** in CDCl₃ at 125 MHz.



^1H - ^1H COSY spectrum of **13** in CDCl_3 at 500 MHz.

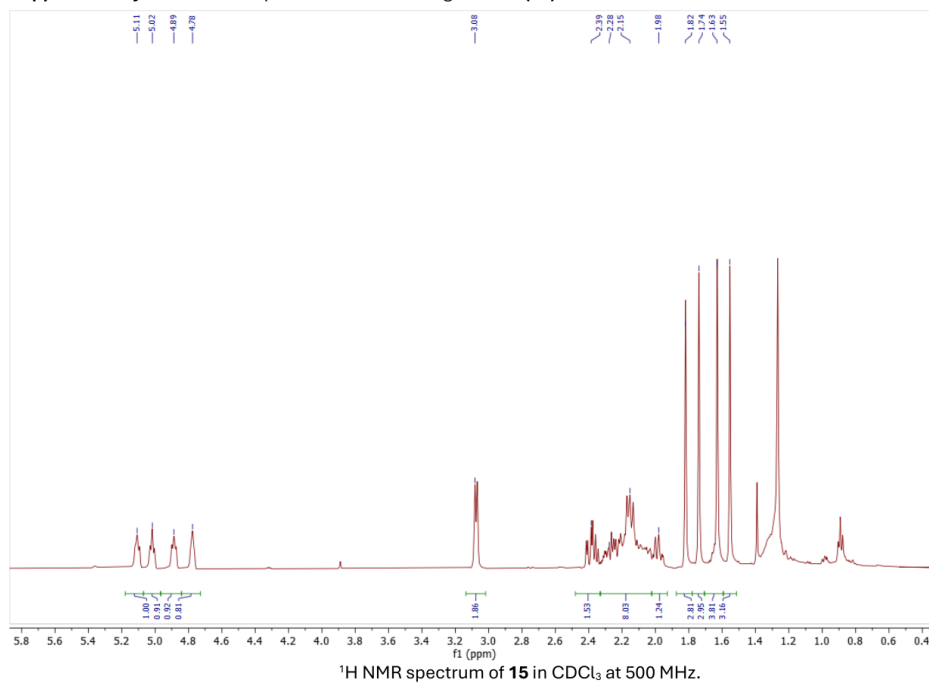


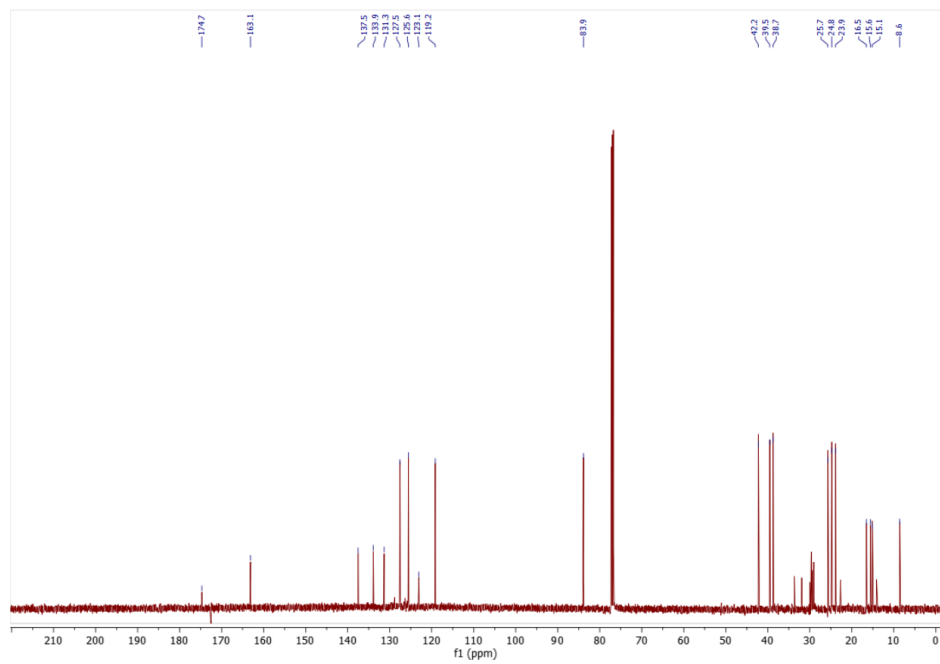
^1H - ^{13}C HSQC spectrum of **13** in CDCl_3 at 500 MHz.



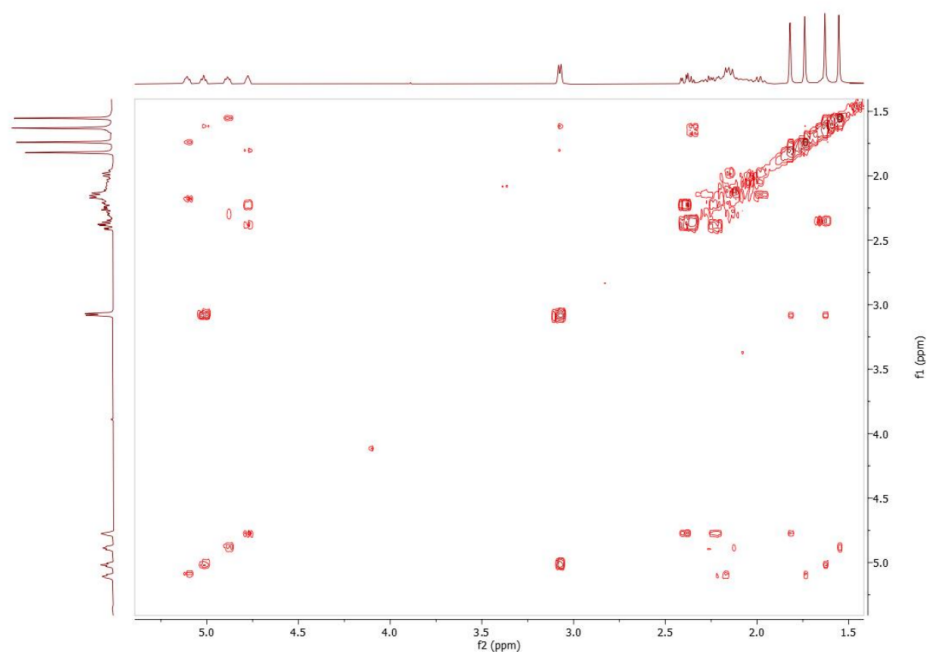
^1H - ^{13}C HMBC spectrum of **13** in CDCl_3 at 500 MHz.

Supplementary Note 7. NMR spectra of cembrene B γ -lactone (15**).**

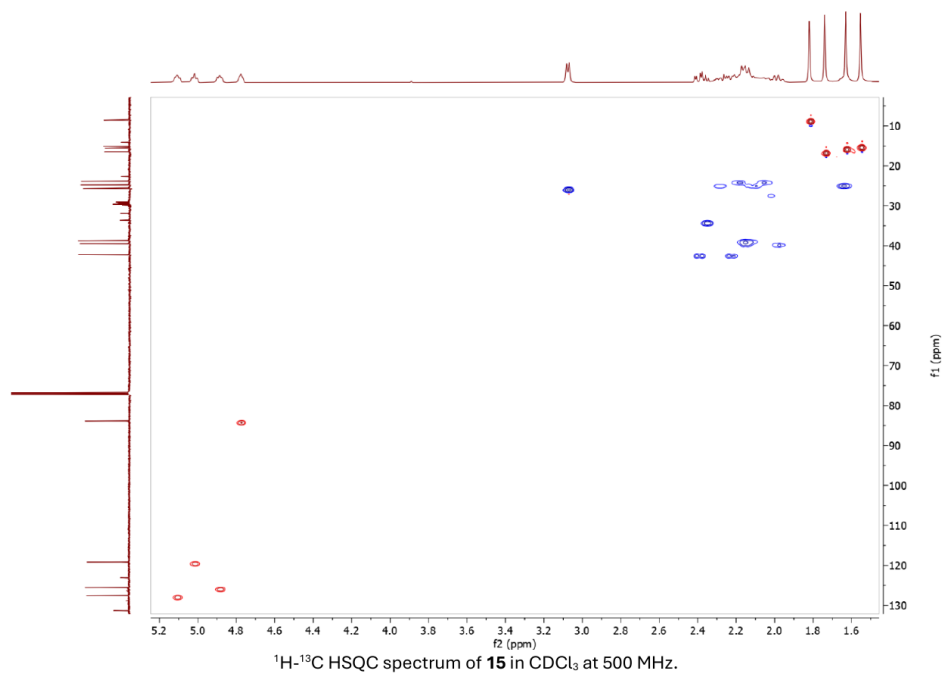




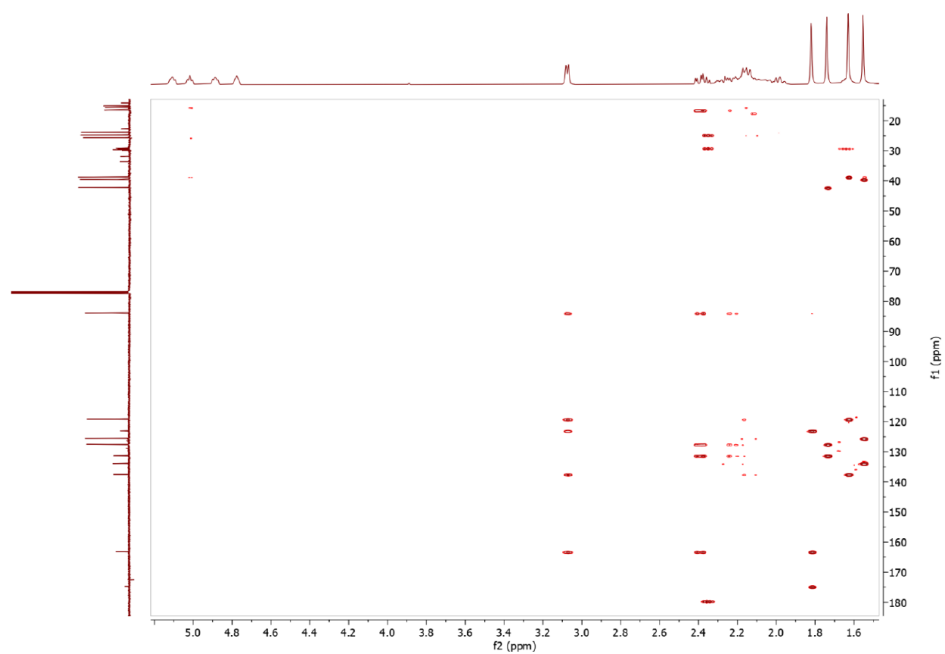
^{13}C NMR spectrum of **15** in CDCl_3 at 125 MHz.



^1H - ^1H COSY spectrum of **15** in CDCl_3 at 500 MHz.

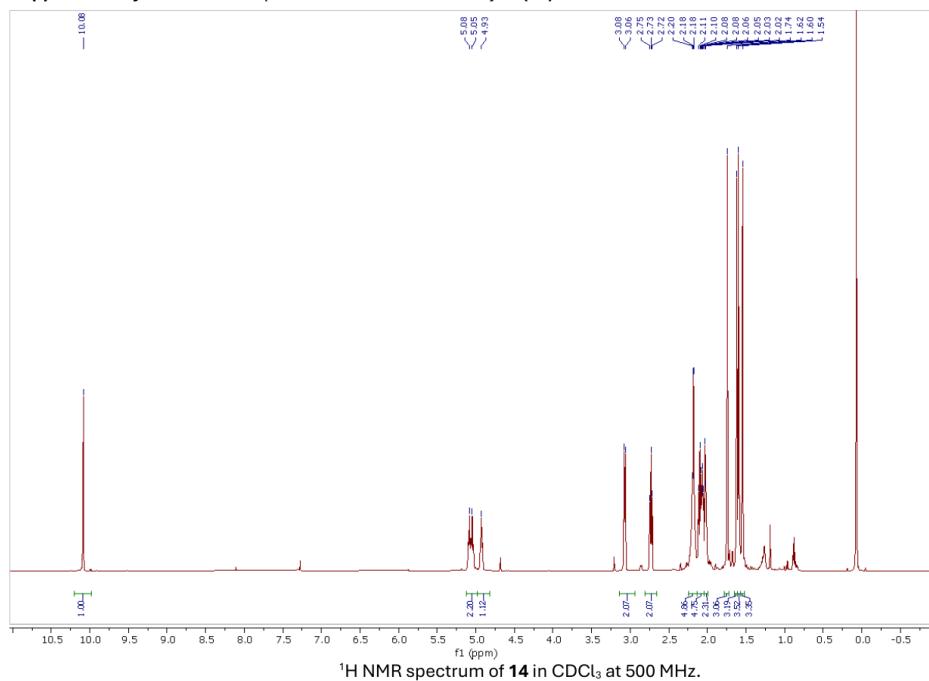


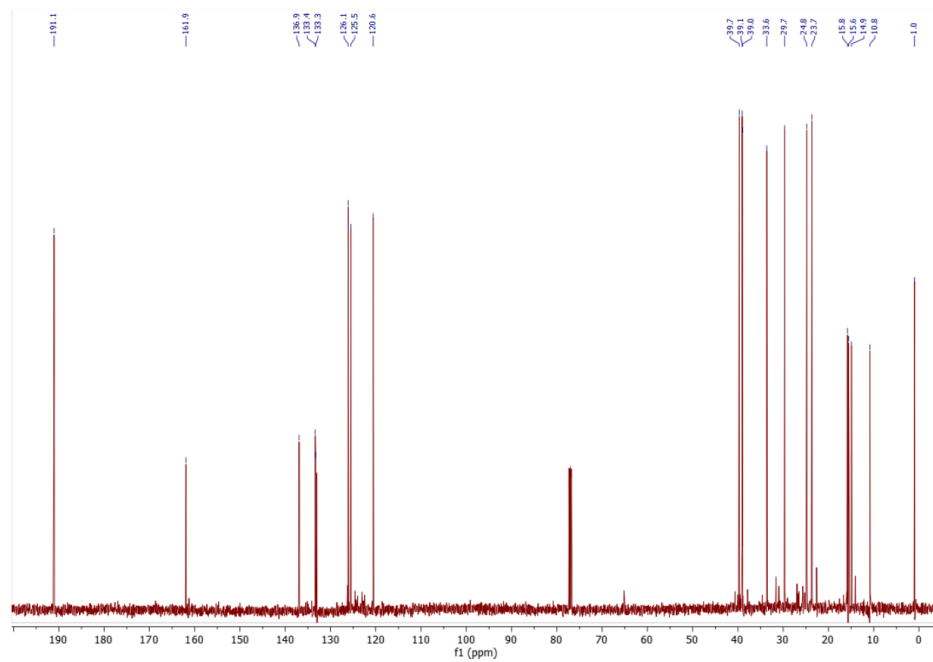
^1H - ^{13}C HSQC spectrum of **15** in CDCl_3 at 500 MHz.



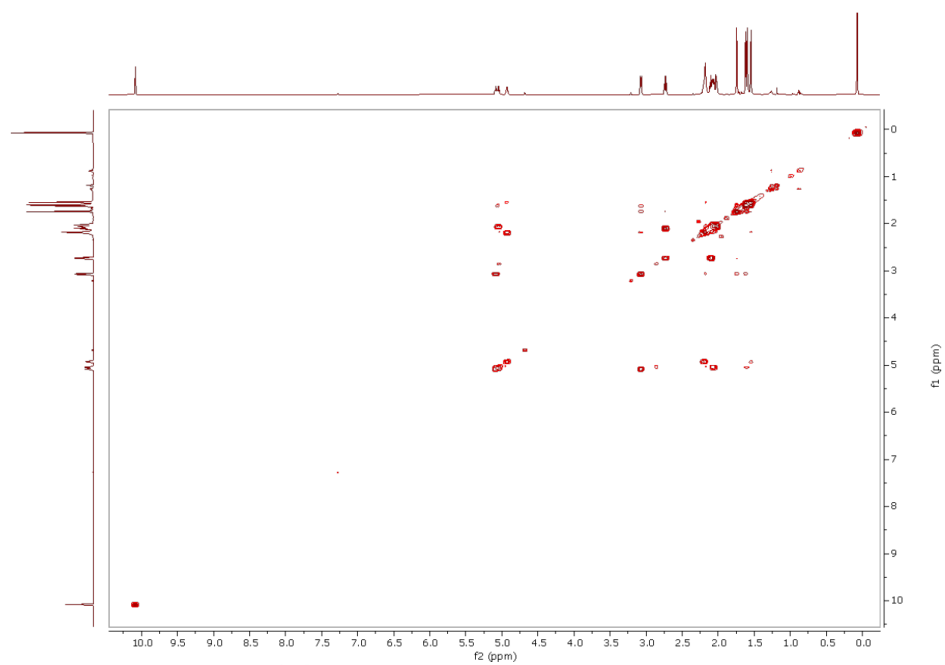
^1H - ^{13}C HMBC spectrum of **15** in CDCl_3 at 500 MHz.

Supplementary Note 8. NMR spectra of cembrene B aldehyde (**14**).

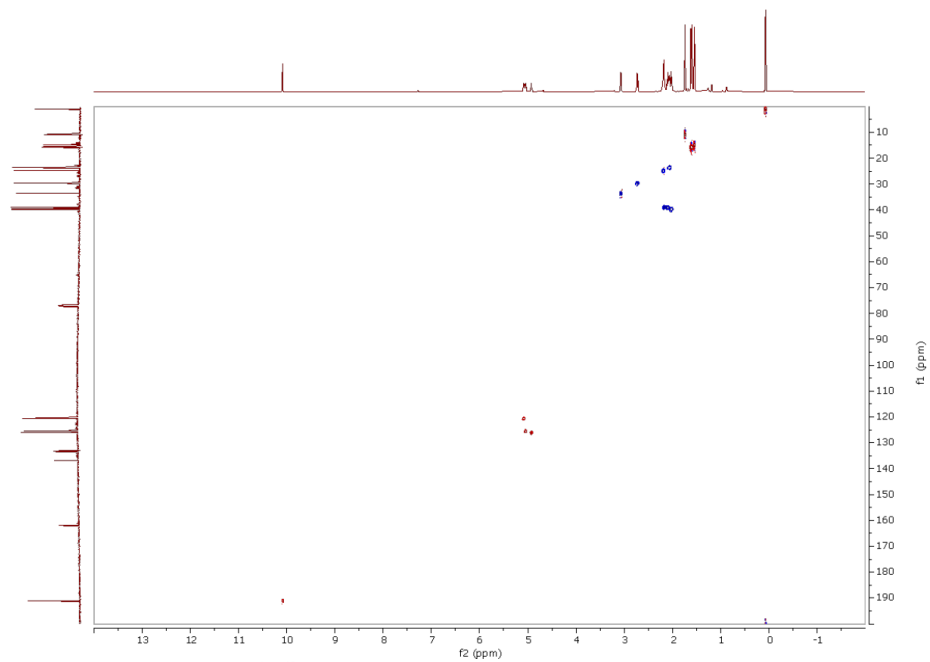




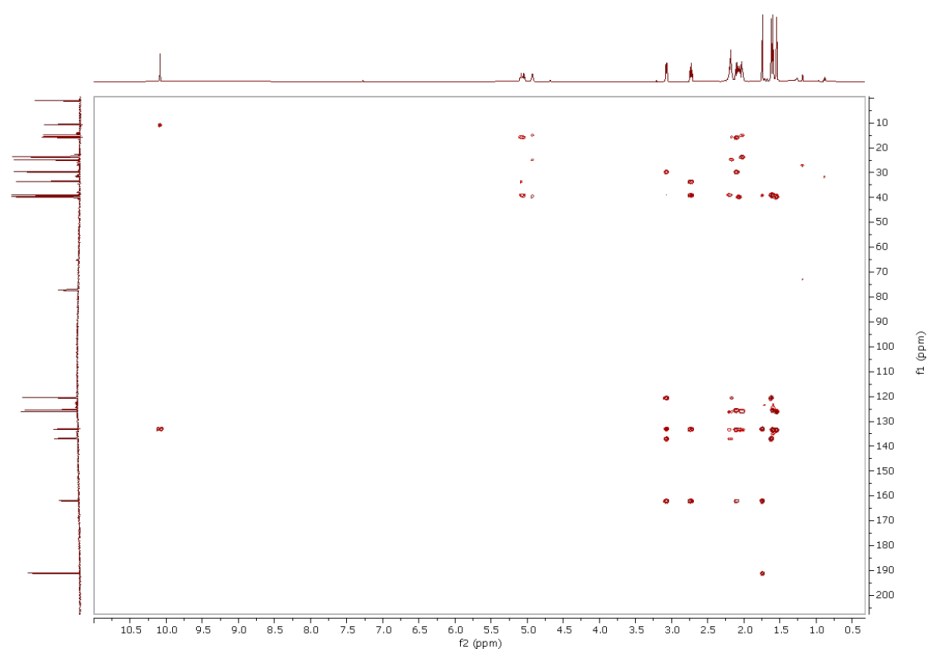
¹³C NMR spectrum of **14** in CDCl₃ at 125 MHz.



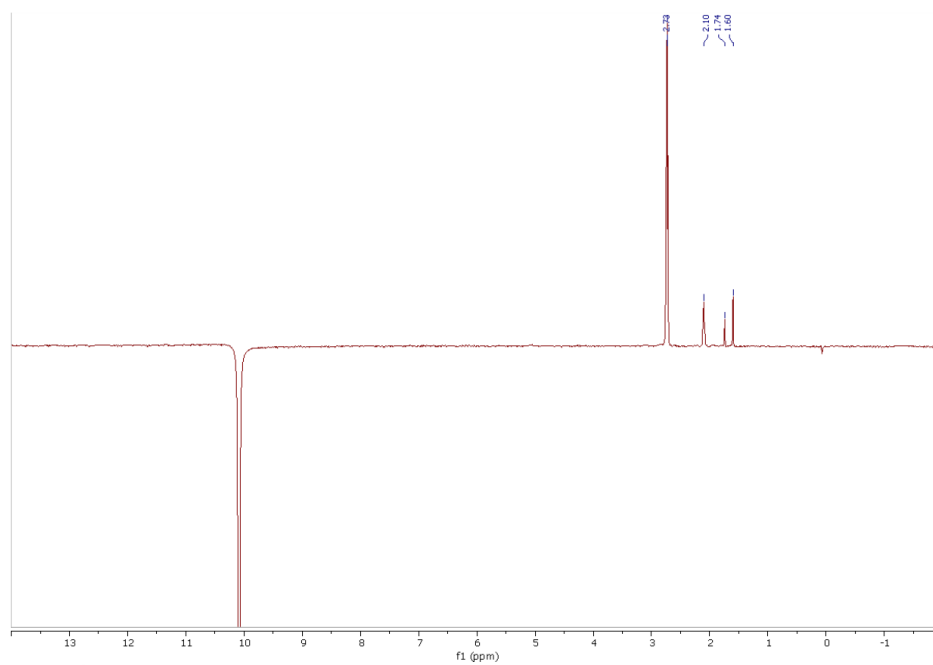
^1H - ^1H COSY spectrum of **14** in CDCl_3 at 500 MHz.



^1H - ^{13}C HSQC spectrum of **14** in CDCl_3 at 500 MHz.



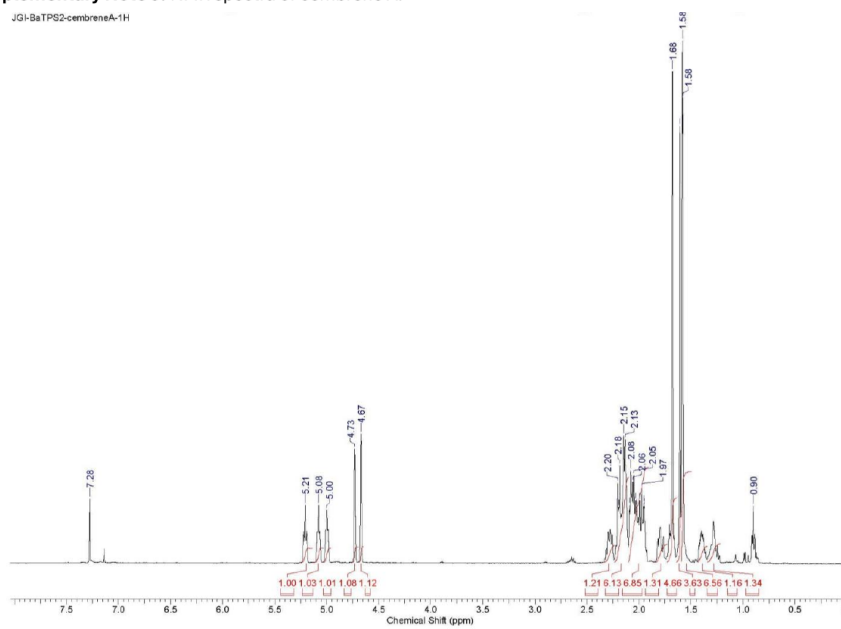
^1H - ^{13}C HMBC spectrum of **14** in CDCl_3 at 500 MHz.



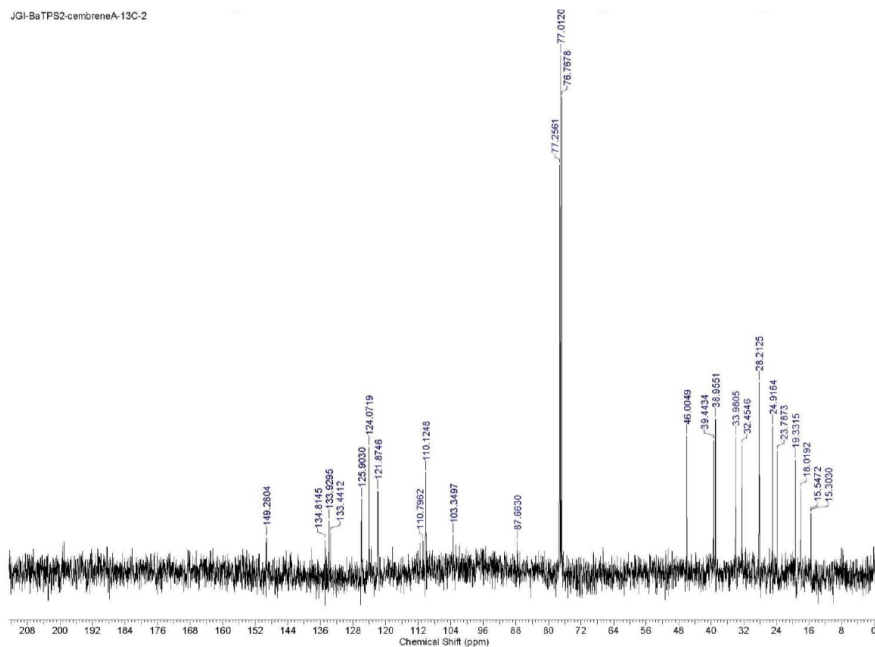
¹H 1D DPGSE NOESY spectrum of **14** in CDCl₃ at 500 MHz with selective excitation at 10.08 ppm.

Supplementary Note 9. NMR spectra of cembrene A.

JG-BaTPS2-cembreneA-1H



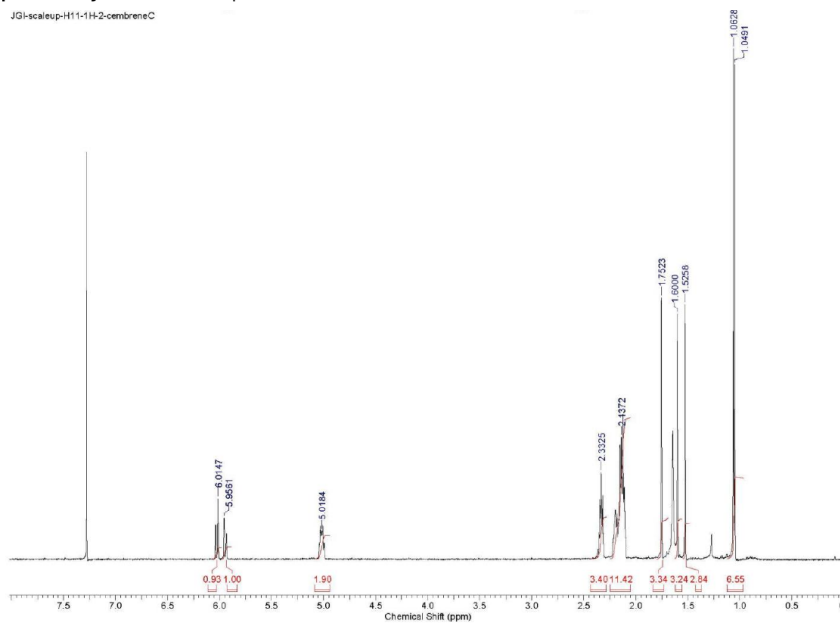
¹H NMR spectrum of cembrene A in CDCl₃ at 500 MHz



^{13}C NMR spectrum of cembrene A in CDCl_3 at 500 MHz

Supplementary Note 10. NMR spectra of cembrene C.

JIGI-scaleup-H11-1H-2-cembreneC



¹H NMR spectrum of cembrene C in CDCl₃ at 500 MHz

1. Wratten, S. J., Faulkner, D. J., Hirotsu, K. & Clardy, J. Stylatulide, a sea pen toxin. *J. Am. Chem. Soc.* **99**, 2824–2825 (1977).
2. Meginley, R. J. *et al.* Briareolate Esters from the Gorgonian *Briareum asbestinum*. *Mar. Drugs* **10**, 1662–1670 (2012).
3. Fenical, W. & Pawlik, J. R. Defensive properties of secondary metabolites from the Caribbean gorgonian coral *Erythropodium caribaeorum*. *Mar. Ecol. Prog. Ser.* **75**, 1–8 (1991).
4. Ksebati, M. B. & Schmitz, F. J. Diterpenes from a Soft Coral, *Minabea* SP., from Truk Lagoon. *Bull. Sociétés Chim. Belg.* **95**, 835–851 (1986).
5. Li, C. *et al.* Bioactive 11,20-Epoxy-3,5(16)-diene Briarane Diterpenoids from the South China Sea Gorgonian *Dichotella gemmacea*. *J. Nat. Prod.* **74**, 1658–1662 (2011).
6. Nurco, D. J., Conklin, D. E., Shapiro, N. S. & Tran, E. The absolute structure of ptilosarcenone 2.5-hydrate, a diterpenoid briarane from the orange sea pen *Ptilosarcus gurneyi* (Gray). *Acta Crystallogr. Sect. E Struct. Rep. Online* **67**, o181–o182 (2010).
7. Hendrickson, R. L. & Cardellina, J. H. Structure and stereochemistry of insecticidal diterpenes from the sea penXXX. *Tetrahedron* **42**, 6565–6570 (1986).
8. Pordesimo, E. O., Schmitz, F. J., Ciereszko, L. S., Hossain, M. B. & Van der Helm, D. New briarein diterpenes from the Caribbean gorgonians *Erythropodium caribaeorum* and *Briareum* sp. *J. Org. Chem.* **56**, 2344–2357 (1991).
9. Pankov, K. V. *et al.* The cytochrome P450 (CYP) superfamily in cnidarians. *Sci. Rep.* **11**, 9834 (2021).

3.4 Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in *Nature Chemical Biology*, Grayson, N. E., Scesa, P. D., Moore, M. L., Ledoux, J.-B., Gomez-Garrido, J., Alioto, T., Michael, T. P., Burkhardt, I., Schmidt, E. W., & Moore, B. S. (2025). The dissertation author was the primary co-researcher and co-author of this chapter with Scesa, P. D.

CHAPTER 4: A Case Study of Allelic Heterozygosity and Gene Duplication in Relation to Octocoral Terpenoid Evolution

4.1 Introduction for Chapter 4

Plants are among the earliest recognized producers of terpenes and account for around 70% of all described compounds (Dictionary of Natural Products).¹ They have long served as the textbook examples of terpene biosynthetic diversity,² with some species, such as *Eucalyptus grandis* and *Eucalyptus globulus* encoding 113 and 106 terpene cyclases (TCs), respectively.³ However, this richness creates its own challenges. With so many TCs present, it becomes difficult to track the evolutionary trajectories of individual genes or to untangle the mechanisms driving lineage-specific expansions. High ploidy in many plants add another layer of complexity, masking heterozygosity and obscuring how new TC variants emerge, diversify, and eventually become fixed within populations. As a result, while plants showcase the enormous potential of terpene chemistry, they are often too complex to serve as clean models for studying the fine-scale evolutionary processes underlying these gene families.

In this chapter, I focus on a single octocoral species as a more tractable window into terpene synthase evolution. With its comparatively small TC repertoire and diploid genome, this system allows us to clearly see how gene duplication and sequence divergence interact to produce functional diversity. Here, we can trace cases of neofunctionalization, where new enzymatic activities arise, as well as subfunctionalization, where the duplicated gene gains enzymatic activity in addition to maintaining the ancestral function. The diploid state provides a unique snapshot of these dynamics in action, revealing how the genetic capacity to produce a novel terpene scaffold can either become fixed or fade away in a population. This perspective highlights the dynamic evolutionary “tug of war” between innovation and loss that underlies the chemical diversity of octocorals.

This work also reflects my own growth as a scientist. Through studying this non-model organism, I developed a deeper understanding of genomics and genome assembly, learning to interpret patterns of duplication, divergence, and selection from complex datasets. Beyond the computational side, I also moved closer to the enzymes themselves through biochemical validation. By functionally testing a gene that we initially predicted to be subfunctionalized, I was able to connect the abstract patterns of sequence evolution

to tangible chemical outputs. This process sharpened my appreciation for the delicate balance of enzyme evolution: the challenge of acquiring a new function while maintaining strict selectivity.

One case in particular drove this home. We identified a single amino acid substitution that completely shifted the product profile of a terpene synthase, essentially allowing the enzyme to “step” into a new chemical space. Through mutagenesis experiments, I was able to watch this process unfold, demonstrating how a single random mutation, at just the right place and time, can push nature’s boundaries and open the door to entirely new metabolic possibilities. This journey, from genome to enzyme, from mutation to new chemistry, underscores both the beauty and the power of evolutionary innovation, setting the stage for a deeper exploration of how octocorals have leveraged terpene synthases to diversify their specialized chemistry.

The work described in this chapter was conducted under my lead as first author. Yanzhe Liu and Immo Burkhardt carried out the initial discovery and early characterization of the TC enzymes, and I subsequently completed the enzyme characterization and expanded the project with additional experimental work. My contributions included performing enzyme mutagenesis, conducting repeat analyses, and carrying out most of the structural elucidation of enzyme products, with portions of the work and significant guidance from Immo Burkhardt. Jeremiah J. Minich generated the genomic sequencing data, while I assembled the reads into genomes and, with the support of Nicholas Allsing, generated gene models to facilitate bioinformatic analyses. I served as the primary author of the manuscript, preparing the main text and figures with input and revisions from all collaborators.

4.2 Chapter 4 Introduction References

- (1) Jia, Q.; Brown, R.; Köllner, T. G.; Fu, J.; Chen, X.; Wong, G. K.-S.; Gershenzon, J.; Peters, R. J.; Chen, F. Origin and Early Evolution of the Plant Terpene Synthase Family. *Proc. Natl. Acad. Sci.* **2022**, *119* (15), e2100361119. <https://doi.org/10.1073/pnas.2100361119>.
- (2) Jiang, S.-Y.; Jin, J.; Sarojam, R.; Ramachandran, S. A Comprehensive Survey on the Terpene Synthase Gene Family Provides New Insight into Its Evolutionary Patterns. *Genome Biol. Evol.* **2019**, *11* (8), 2078–2098. <https://doi.org/10.1093/gbe/evz142>.
- (3) Külheim, C.; Padovan, A.; Hefer, C.; Krause, S. T.; Köllner, T. G.; Myburg, A. A.; Degenhardt, J.; Foley, W. J. The Eucalyptus Terpene Synthase Gene Family. *BMC Genomics* **2015**, *16* (1), 450. <https://doi.org/10.1186/s12864-015-1598-x>.

4.3 Abstract

Octocorals are prolific sources of bioactive terpenoids, yet the genetic mechanisms that generate this chemical diversity remain largely unresolved. The red gorgonian *Leptogorgia chilensis*, known for its neurotoxic diterpenoid lophotoxin, also harbors a multitude of other terpenoids spanning distinct scaffolds, providing a compelling model for linking genome architecture to natural product biosynthesis. Here, we assembled a phased, high-quality genome and transcriptome, identifying 24 terpene cyclase (TC) genes falling into nine phylogenetic clades. The assembly possessed extensive TC allelic variation, including haplotype-specific gene duplications as well as loss-of-function mutations. Biochemical characterization of 14 of the 24 TCs uncovered the production of ten distinct terpene scaffolds and highlighted several cases of functional diversification of paralogs. Exploring a specific case of gene neofunctionalization, structural modeling and site-directed mutagenesis showed that a single amino acid substitution can toggle product specificity between sesquiterpene and diterpene synthase activity, underscoring how previously unrealized subtle molecular changes elicit in vivo diversification. Together, these findings demonstrate the extent to which gene duplication and allelic divergence drive terpene biosynthetic adaptation.

4.4 Introduction

Terpenes are the most structurally diverse classes of natural compounds, fulfilling ecological functions that range from chemical defense to interspecies signaling across diverse taxa.^{1,2} While the terpenoid chemistry produced by octocorals has been studied for decades,^{3,4} their ecological functions and the biochemical pathways for their production remain incompletely understood. Preliminary bioassays have demonstrated that octocoral terpenoids can act as antifouling agents, feeding deterrents, and antipredator compounds.⁵⁻⁷ These bioactivities have also attracted interest for pharmaceutical and commercial applications.^{8,9} Strikingly, a vast majority of biologically active terpenoids produced by octocorals have remained largely inaccessible due to biomass limitations and the historical absence of genetic context. One such bioactive compound, lophotoxin (**1**), a neuromuscular toxin,^{10,11} is produced by the red gorgonian *Leptogorgia chilensis*. Reports on this pacific gorgonian are limited, with both the biosynthetic origin of **1** and the coral's broader metabolic capacity largely unexplored.

While plant¹² and microbial¹³ terpene biosynthesis has been extensively studied, the genes encoding terpene cyclases (TCs) have only recently been discovered in sessile marine invertebrates like octocorals,^{14,15} and sponges.¹⁶ Analogous to their plant and microbial counterparts, metazoan TCs catalyze the cyclization of polyprenyl diphosphate substrates of varying chain lengths. However, these enzymes have evolved independently from other TC lineages over hundreds of millions of years, acquiring distinct sequence motifs and resulting in production of lineage-specific terpene scaffolds. Across all octocoral genomes sequenced to date, spanning diverse ecological niches, terpene synthases are universally present and diverse in sequence and function.^{14,17–20} Due to our limited understanding of TC biochemistry, our capacity to predict an encoding gene's function from sequence identity and phylogenetic relatedness alone is unreliable. Even closely related enzymes with only a few nucleotide differences can catalyze entirely different cyclization reactions.²¹ Conversely, enzymes with no significant sequence homology can produce identical products,^{22,23} illustrating an extraordinary evolutionary flexibility that is not yet mechanistically rationalized.

Despite evidence that octocorals produce structurally diverse terpenoid compounds, the complete biosynthetic capacity encoded by single coral specimens, as well as how these animals evolve the capacity to generate terpenoid diversity remains unknown. The molecular basis potentiating terpenoid chemical innovation in plants often arises from the expansion and diversification of biosynthetic genes^{12,24} through processes such as whole-genome²⁵ and sub-genomic duplications, tandem duplication driven by unequal crossing over,²⁶ and regulatory evolution.²⁷ Whether similar processes occur in octocorals, and how variation in TC gene sequence, copy number, and expression drive terpene biosynthetic adaptation has yet to be explored. To address these gaps, we selected *L. chilensis* as our focal system. In this work, we uncover a rich mosaic of terpene hydrocarbons in the extract of the coral, and through a haplotype-resolved genome assembly paired with biochemical characterization of TCs, we investigate how this diversity connects to genomic potential and what evolutionary mechanisms might underlie it.

4.5 Results

Leptogorgia species have long been studied for the defensive chemistry present in their tissues. According to MarinLit, 69 compounds have been reported from the genus *Leptogorgia*, of which 26 are non-

steroidal terpenoids. Although many reported compounds are furanocembranolides like lophotoxin (**1**), five distinct sesquiterpene scaffolds (**2–6**) have also been identified across different *Leptogorgia* species.^{10,28–33} This breadth of chemistry among closely related corals led us to investigate whether the diversity of terpenoid scaffolds in the genus arises because different species each produce distinct sets of compounds, or whether individual species are capable of synthesizing the wide range of scaffolds reported collectively across the genus. To establish a model *Leptogorgia* system for paired omics, we collected *L. chilensis* from the Scripps Reserve Underwater Canyon in San Diego by SCUBA at approximately 25 meters depth. Gas chromatography–mass spectrometry (GC-MS) analysis of a hexane extract of its tissue revealed the presence of **1**, which was the only previously described compound from this species, along with its biosynthetic precursor, cembrene A (**7**), both confirmed via comparisons with authentic standards (**Fig 1**, **SI Fig 1**). Additionally, the extract also contained numerous additional terpene features, including both sesquiterpenes and diterpenes. This larger, previously unreported chemical diversity in *L. chilensis* prompted us to investigate the genetic capacity underpinning its production.

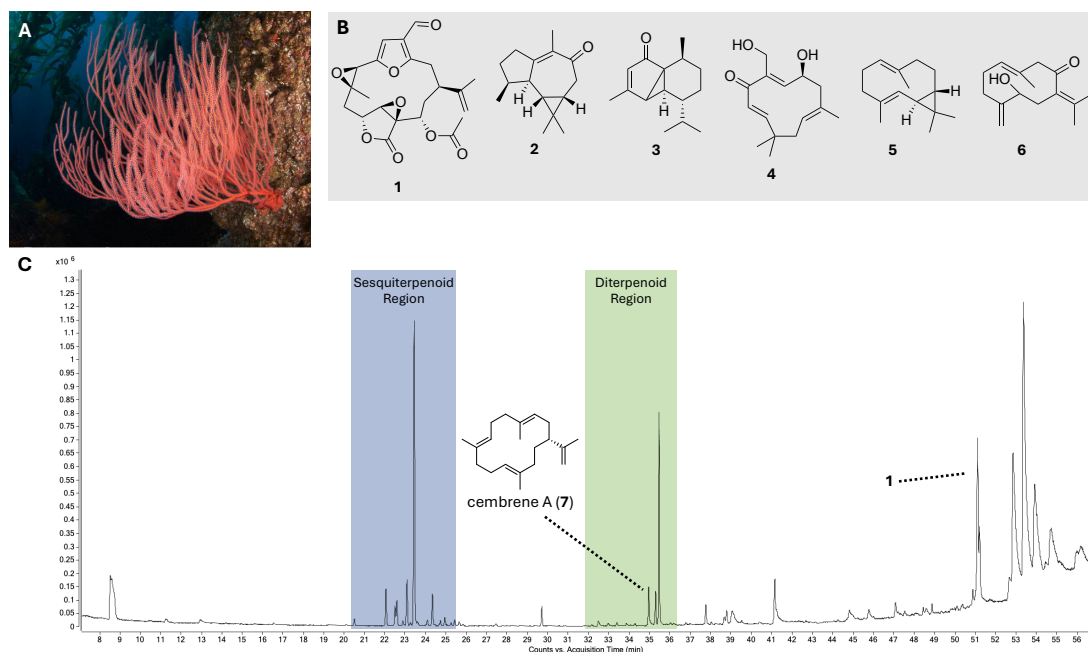


Figure 4-1: GCMS chromatogram of *Leptogorgia chilensis* crude extract.

A: Image of *L. chilensis* in La Jolla, CA. **B:** Reported terpenoids **2–6** found in tissues of different *Leptogorgia* species. **C:** GC chromatogram of an organic extract of *L. chilensis*. Annotated sesquiterpene and diterpene regions are based on retention times. Compounds with authentic standards, lophotoxin (**1**) and cembrene A (**7**) are shown. For complete annotated chromatogram, see **SI Fig 1**.

Genomic Assembly and Mining

The genome size of the *L. chilensis* genome was estimated by K-mer frequency analysis with Illumina short reads to be 194 Mb with 0.85% heterozygosity (**SI Fig 2**). Using a combination of Pacific Bioscience (PacBio) and Oxford Nanopore Technology (ONT) long read sequencing data, we generated a phased, high-quality genome assembly of *L. chilensis* (**SI Table 1**). Each haplotype assembly had approximate genome sizes of 245 magabases (Mb) and 226 Mb and contig N50 lengths of 3.3 Mb and 2.5 Mb, respectively (**Table 1, SI Tables 2-3**). We confirmed the differences between the haplotypes by mapping the Illumina short reads used to estimate the genome size and heterozygosity (270x coverage) onto each haplotype assembly and visualizing with IGV (**SI Table 3**).

Table 4-1: Statistics of phased genome assemblies

	Haplotype A	Haplotype B
Assembly Size	235 Mb	225 Mb
Contig N50	3.3 Mb	2.5 Mb
# of Contigs	379	276
Protein BUSCO Complete (%)	94.6	92.3
Genes	23,443	23,001
Repeat (%)	30	40
TC Genes	14	10

We sequenced the transcriptome of the same sample used for genome sequencing with Illumina short reads. We aligned the raw reads to our assembled genome with an overall alignment rate of 92.88% and assembled a *de novo* transcriptome of 141,509 genes with a contig N50 length of 3.5 kb and 95.9% BUSCO completeness (**SI Table 3**). The transcriptomic reads were used in conjunction with publicly available octocoral proteomes to generate protein predictions for the two haplotype assemblies. The proteomes for hapA and hapB had BUSCO scores of 94.6% and 92.3%, respectively. HapA had 23,443 genes annotated and a repeat content of 30%, while hapB had 23,001 genes predicted and a repeat content of 40% (**Table 1, SI Table 3**).

Mining the *L. chilensis* haplotypes for TC encoding genes, we found 24 genes containing the canonical octocoral TC motifs.^{14,15} This number includes allelic copies across haplotypes and given the low overall heterozygosity of the genome (**SI Fig 2**), it was intriguing to find that the distribution of TC alleles was

asymmetric: 14 were on hapA and 10 were on hapB (**Table 2**). A multiple sequence alignment of the 24 TC protein sequences followed by phylogenetic analysis revealed that these genes fell into nine distinct, homologous groups (**SI Fig 3**). To place the *L. chilensis* TCs in an evolutionary context, we constructed a broader phylogeny, incorporating the *L. chilensis* TC protein sequences with all publicly available octocoral TC genes. The resulting tree showed that while the *L. chilensis* TCs are interspersed among TCs from other genera, the LcTC groups exhibit lineage specific diversification (**Fig 2a**). One of the LcTCs fell into TC clade 1-III with previously characterized C-type TCs (see chapter 2), in line with the detection of cembrene A (7) in the tissue extract of the coral. Based on sequence identity analysis, out of the 14 TCs on hapA, 8 were clearly distinct, and the other 6 were recently duplicated paralogs (**SI Fig 3**). On hapB there were 7 unique TC sequences and the remaining 3 were duplicate copies. Across the 9 groups of TC sequences, group 2 was the most divergent, with an amino acid sequence identity of 89% between the paralogs yet 100% identity of alleles (**SI Fig 3**).

Table 4-2: TC distribution across haplotypes and syntenic blocks

TC	Syntenic Block	# Copies Haplotype A	# Copies Haplotype B
TC1	C	1	1
TC2	B	2	2
TC3	A	1	1
TC4	A	2	2
TC5	B	3	0
TC6	B	0	1
TC7	D	2	2
TC8	C	2	1
TC9	A	1	0

Next, we conducted a syntenic analysis to explore the genomic context of the identified TC genes. All TCs were in four distinct contiguous regions in our assembly, syntenic blocks A-D (**Fig. 2b, Table 2**). To gain better resolution and confidently assign TC-containing contigs to their respective haplotypes, we scaffolded our assembly using ultra-long ONT reads and the NtLink³⁴ pipeline, which combines minimizer-based mapping, overlap detection, gap-filling, and iterative scaffolding to order, orient, and merge fragmented contigs (**SI Table 4**). Upon closer inspections of TC containing syntenic blocks, we found haplotype specific TC duplications and

pseudogenizations for TC5, TC6, TC8 and TC9 (**Fig 3**). TC5 had three copies on hapA (TC5a, TC5b and TC5c), but hapB only had one unfunctional homolog, lacking a start codon. When examining the syntenic block containing TC5b and TC5c on hapA, we found that although several genes within this region aligned with the contig containing the non-functional TC5a on haplotype B, no homologs of those genes were detected (**SI Fig 5**). TC6 was only present on hapB as a single copy, and TC9 was only present as a single copy on hapA, both having a pseudogenized allele on the other haplotype (**SI Figs 6,7**). TC8 had two complete copies on hapA (TC8a and TC8b) and just one on hapB; however, additional copies of TC8 were present in both hapA and hapB, but frameshift mutations resulted in these enzymes being incomplete (**SI Fig 8**).

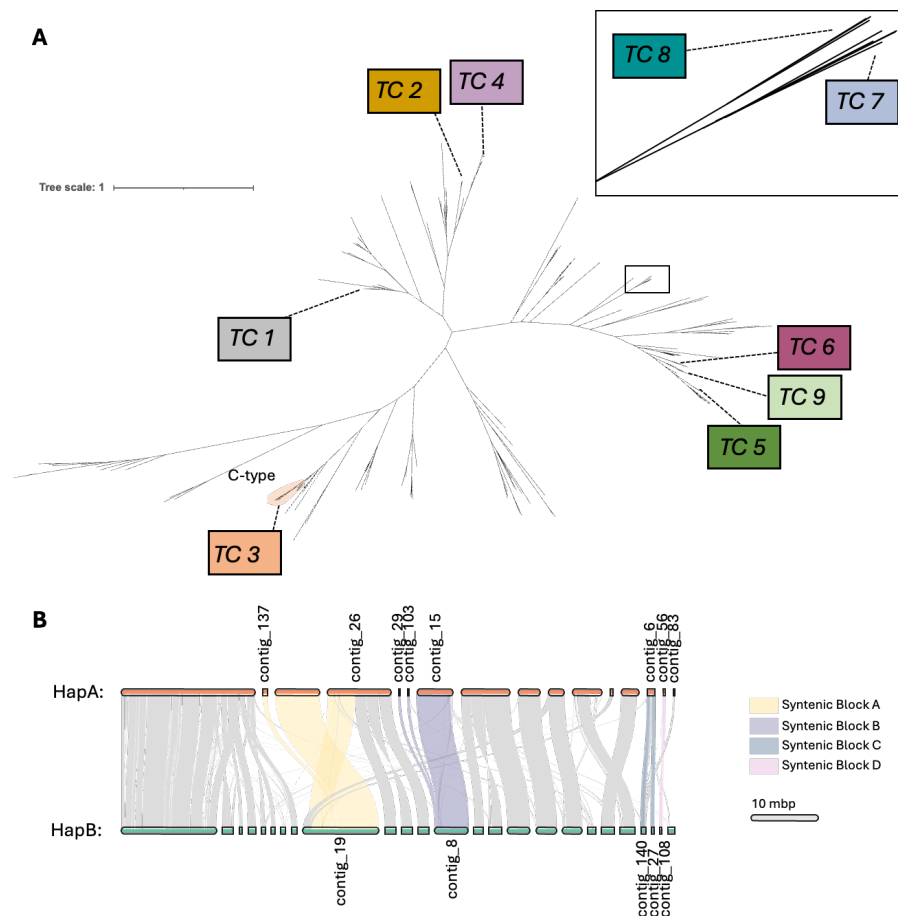


Figure 4-2: Phylogenetic relationship and genomic context of *L. chilensis* TCs.

A: Unrooted phylogenetic tree showing all publicly available octocoral TCs. Boxed are regions that *L. chilensis* TCs fall. Zoomed in branches containing these regions with TC clades annotated are depicted on the right. TC3 clades with C-type TCs. **B:** Select macro-synteny of phased assembly, only longest contigs and contigs containing TCs and their syntenic partners are shown. Full synteny see **SI Fig 4**. Syntenic blocks containing TC encoding genes are highlighted and their respective contigs are labelled.

Biochemical and Transcriptomic Exploration of TCs

To link the observed terpenes in the extract to the observed genomic potential we biochemically characterized representative *L. chilensis* TCs from each of the nine groups (**Fig 3**). We heterologously expressed targeted genes in *Escherichia coli*, purified the recombinant enzymes, and tested all TCs with the linear precursors farnesyl diphosphate, FPP (**18**) and geranylgeranyl diphosphate, GGPP (**19**).³⁵ and the product profiles were analyzed by GC–MS and NMR spectrometry (**SI Tables 5-6, SI Figs 9-23, NMR Results Section, Structural Elucidation Methods**). All enzymes from different TC groups displayed unique functions. TC1 functioned as a germacrene B (**8**) cyclase, TC2a, made (+)- β -cubebene (**9**), TC3 functioned as a (–)-cembrene A (**7**) cyclase, TC4a functioned as a (+)- β -caryophyllene (**10**) cyclase, TC5a generated guaia-11(12),6(14)-diene (**11**), a previously unreported compound, TC6 was inactive, TC7a made (+)- γ -curcumene (**12**), TC8 produced (Z)- α -farnesene (**13**), and TC9 functioned as a (+)-maaliene (**14**) cyclase (**Fig 3**).

When referencing our *L. chilensis* crude extract (**Fig 1**), we matched the products of TC2a, TC3, TC4a, TC7a, and TC9 with five of the compounds identified in the coral's tissue (**SI Fig 1**). However, we could not find all the detected terpene hydrocarbons from the extraction from *L. chilensis* tissue among the products of our in vitro experiments, and we did not find the products produced by TC1, TC5a, or TC8a in the crude extract. To determine if sequence divergent gene paralogs have neofunctionalized and hence could be responsible for the unaccounted terpenes in the *L. chilensis* extract, we systematically expressed TCs that had duplicate copies: TC2b, 4b, 5b and 5c, and 7b. While TC8 was duplicated in hapA the sequences of the two copies were so similar (98.31%) that we did not test the second copy. We found that while TC2a produced **9**, the second paralog, TC2b, made germacrene A (**15**), which was not present in the crude extract in detectable quantities. While TC5b had the same function as TC5a, generating **11**, which was not present in the crude extract, the third copy, TC5c, formed (+)- α -gurjunene (**16**) from FPP, which was found in the coral's tissue. Lastly, TC7a's paralog, TC7b, made a diterpene related to TC7a's **7**, forming (+)-prenyl- β -curcumene (**17**) from GGPP, which was also present in the *L. chilensis* crude extract. Attempts to characterize TC4b failed due to apparent toxicity of the protein to *E. coli*. However, by heterologously expressing 14 of the 24 TCs present in the *L. chilensis* genome, we were able to reconstitute the complete suite of terpenes present in the coral's crude extract (**SI Fig 1, SI Table 7**).

While our enzyme assays demonstrated the capacity of *L. chilensis* to produce 11 distinct terpene

scaffolds, only 7 of these were detected in the crude coral extract. To better understand this discrepancy, we examined the transcriptional activity of all *L. chilensis* terpene cyclases using our paired transcriptomic data. All TCs that produced compounds detected in the crude extract, TC2a, TC3, TC4a, TC5c, TC7a, TC7b, and TC9 were transcriptionally active, and the six TCs with little to no expression, TC1, TC2b, TC4b, TC5a, TC5b, and TC8a, produced compounds not observed in the organism's tissue (**SI Tables 7-8**).

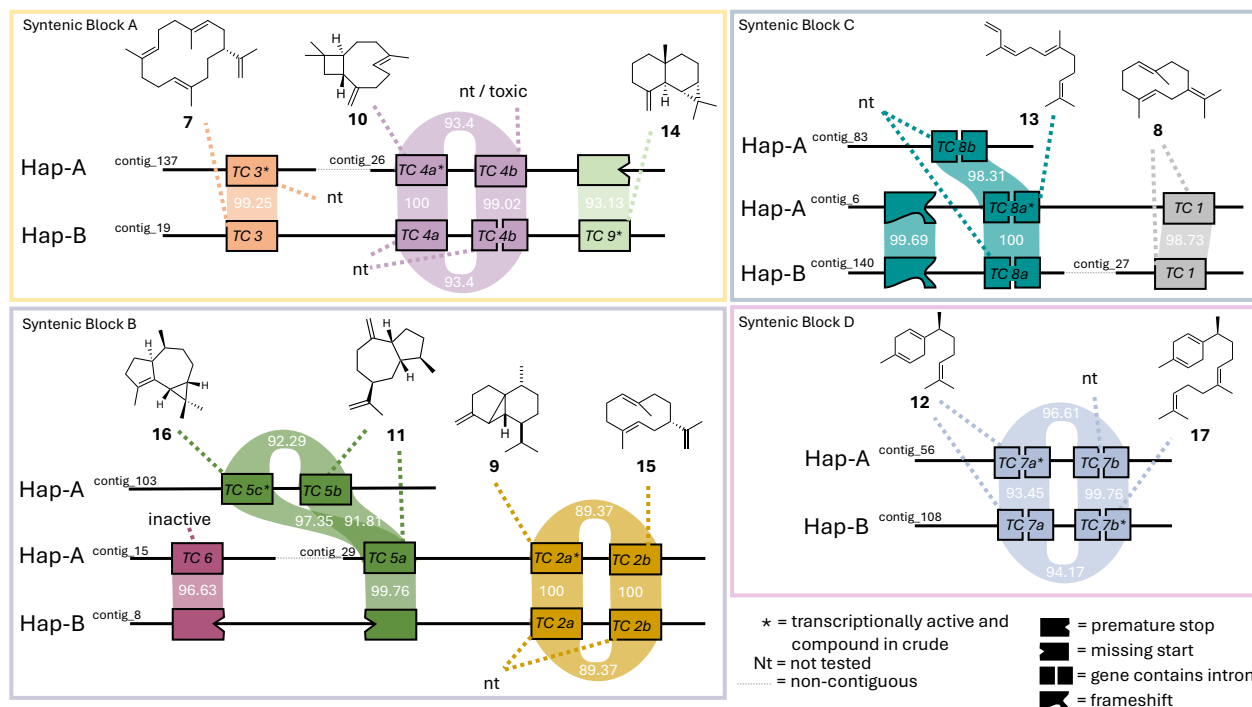


Figure 4-3: Products generated in enzyme assays.

All gene paralogs were tested. Due to high degree of identity across alleles, several were not tested as annotated with nt. The color of the genes and annotation were assigned via phylogeny and PID (**Fig 2, SI Fig 3**). For structure elucidation and chromatographic data see **SI Methods** and **SI Figs 10-23**. While genes and contig lengths are not to scale, the order and relation to other genes on the contigs are depicted. Several pseudogenes are also annotated along with information about their PID and degradation status based on the emoji used to show the gene. Transcription activity of alleles was not differentiated, and the asterisk is indicating at least one of the alleles was transcriptionally active for that gene.

Enzyme Mutagenesis

Among the four sets of TC paralogs in *L. chilensis*, as a pair, 7a and 7b were the most similar, with the HapA alleles sharing 96.6% amino acid identity and the HapB alleles sharing 94.17%. However, biochemical characterization revealed clear functional divergence: 7a acted exclusively as a sesquiterpene cyclase

producing **12**, whereas 7b primarily exhibited diterpene cyclase activity producing **17**, with a diminished capacity to produce **12** from FPP. Hence, we set out to pinpoint the molecular switch responsible for enabling or restricting diterpene cyclase activity. A multiple sequence alignment with both alleles of TC7a and TC7b revealed numerous cases that both TC7b's contained a distinct AA at a given position compared to one of the TC7a homologs. A four-way comparison of both 7a and 7b alleles allowed us to further restrict the number of potentially responsible amino acid positions as there were only 6 AA differences that both TC7b's exhibited compared to both TC7a's (**SI Fig 24**).

To investigate the functional relevance of the differing amino acids, we introduced five individual residue swaps into the diterpene synthase, substituting each position with the corresponding amino acid from the sesquiterpene synthase. One pair of adjacent residues at positions 140 and 141 was not tested independently due to their proximity. Of the five substitutions tested, only the swap at position 91 in which an isoleucine replaced a serine in TC7b (TC7b-S91I), disrupted diterpene synthase activity (**SI Figs 25-28**). We also observed that TC7b-S91I showed higher sesquiterpene synthase activity compared to 7b-WT but exhibited still lower activity than the native 7a's FPP activity (**Fig 4c**). To explore whether this residue serves as a bidirectional switch, we introduced the reciprocal mutation into the sesquiterpene synthase, replacing isoleucine with serine at the same position (TC7a-I91S). Indeed, TC7a-I91S was able to act on GGPP to form **12** (**Fig 4c**). However, while TC7a-I91S showed a similar relative FPP turnover as TC7b-S91I, it showed a significantly lower GGPP turnover compared to TC7b-WT (**Fig. 4C**). Given the pronounced effect of the swap at residue 91 compared to the limited impact of the other six residue swaps, we examined 3D protein models of the sesquiterpene synthase TC7a and the diterpene synthase TC7b. Notably, only the amino acid at position 91 was predicted to be in the active site (**Fig 4a**). Further, the residues at this position differed in steric hindrance: TC7a had an isoleucine, whereas TC7b had a serine (**Fig 4b**). Therefore, the difference in substrate preference between 7a and 7b can be rationalized by a difference in active site volume, a phenomenon that has been described previously in bacterial TCs.^{36,37}

To further probe the steric importance of residue 91, we explored how substitutions with different sizes affect diterpene synthase activity. We introduced glycine, alanine, and valine substitutions at position 91 into 7b, where glycine, alanine and valine substitutions maintained catalytic activity with GGPP, resulting in

continued production of **17** (SI Figs 25-28, SI Table 9). As the valine sidechain was the largest, which still permitted turnover of GGPP in LcTC-7b, we introduced a valine at position 91 in LcTC-7a. However, LcTC-7a-S91V was not able to produce **17** from GGPP (SI Figs 25-28, SI Table 9). This suggests that while the size of residue 91 plays a critical role in the functional divergence between LcTC-7a and 7b, there are additional, non-active site mutations that are contributing through more subtle, long-range conformational shifts resulting in slight changes in the contour of the active site or overall protein dynamics.

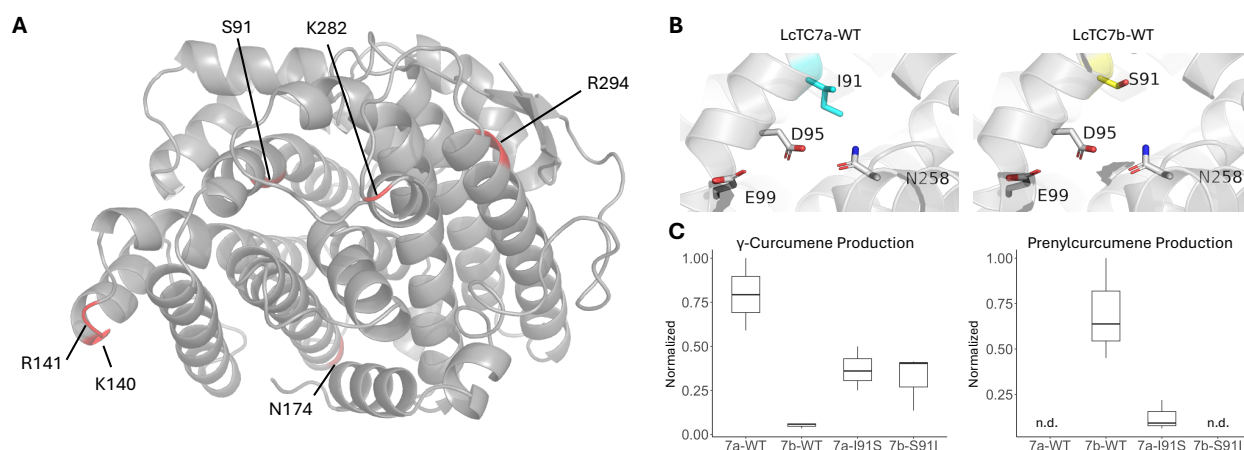


Figure 4-4: Protein models and comparison of compound production of LcTC7 paralogs.

A: 3D model of the diterpene synthase with the 6 positions being different in the sesquiterpene synthase highlighted in red. **B:** Active site of wild type sesquiterpene synthase (left) and diterpene synthase (right). Select highly conserved active site residues have been included in grey and the active site residue distinguishing the two enzymes at position 91 are highlighted in blue (7A) and yellow (7b). **C:** Quantitative comparison of compound production across wild type and reciprocal residue 91 variants. Left: Box plots of all four enzymes, acting on FPP to produce **12**, while 7b-WT had the lowest production, there were still detectable amounts of product formed. Right: Box plots of four enzymes, 7b-WT and 7a-I91S produced **17** from GGPP, but 7a-WT and 7b-S91I had no diterpene activity. All assays were conducted in triplicate. GCMS integrated peaks from the resulting enzyme assays were all normalized to the wildtype producing strain (7a-WT in the case of FPP catalysis and 7b-WT in the case of GGPP catalysis). N.d. represents no detectable amount of compound. See SI Figs 25-28 for all relevant chromatograms.

TCs are in TE-Rich Genomic Neighborhoods

After establishing that the terpene diversity in *L. chilensis* partially results from recent gene duplications we investigated the mechanisms underpinning this observation. Transposable elements (TEs) are generally thought to play a key role in genomic evolution of plants and animals.³⁸ In plants specifically, where secondary metabolism is better-investigated, different TE classes show elevated genomic abundance and

display distinct spatial organization near genes involved in specialized metabolite biosynthesis. These elements have been implicated in gene duplication, chromosomal rearrangements, biosynthetic gene cluster (BGC) formation, and coordinated expression of biosynthetic genes.^{39,40} To determine if a similar scenario was occurring in corals, we conducted a repetitive element analysis. We annotated TE classes in our phased assemblies and calculated their genome-wide base pair contributions. We then examined 100 kbp windows surrounding each TC gene to determine the local percentage of base pairs associated with each TE class and compared these values to those from randomly sampled 100 kbp genomic regions. While terminal inverted repeats (TIRs) accounted for an average of 9.61% of the *L. chilensis* genome, their abundance in TC neighborhoods was 14.50%, reaching as high as 21.78% around TC2b (**Fig 5**), representing a significant enrichment (**SI Table 10**). A more detailed analysis of the TIR category revealed that CACTA, hAT, Mutator, and PIF-Harbinger elements were particularly enriched near TC genes (**SI Table 10**).⁴¹

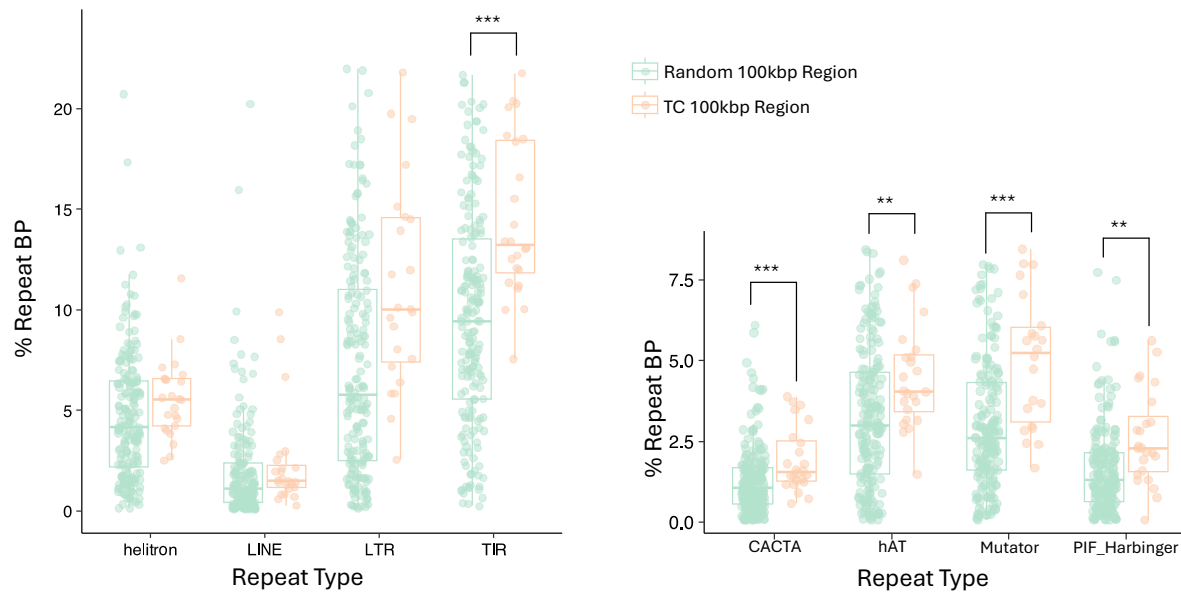


Figure 4-5: Distribution of different repeat classes in *L. chilensis* assembly.

In orange is the % repeat base pairs in the genomic neighborhood of all 24 TC genes defined as 100 kbp. In green is the % repeat base pairs in randomly selected genomic neighborhoods (100 per haplotype) defined as 100 kbp regions. Each unique dot represents a genomic neighborhood. In general, TIRs were enriched in TC regions, and more specifically, there was also significant local enrichment of CACTA, hAT, Mutator, and PIF-Harbinger elements near TC genes (**SI Table 10**).

4.6 Discussion

This study provides a high-resolution view into TC evolution through the lens of a single, chemically prolific octocoral species. By combining haplotype-resolved genomics with biochemical characterization, we reveal how gene duplication, allelic heterozygosity, and amino acid-level divergence contribute to the emergence of terpene scaffold diversity in *L. chilensis*. Our results demonstrate that even subtle sequence changes can produce marked shifts in enzyme function, highlighting the high degree of evolvability within this enzyme class.

Research on octocoral terpenoids is extensive, with over 4,000 sesqui- and diterpenoids reported to date;⁴² however, much taxonomic identification supporting this work was based on morphological traits, which often failed to distinguish cryptic species and has led to stark chemical variability and inconsistencies. Molecular phylogenetics coupled with chemo-typing has since revealed that coral species produce idiosyncratic terpenoid profiles.⁴³ While diverse hydrocarbons have been reported from multiple species within the genus *Leptogorgia*,^{10,28–33} yet before this work it was unclear whether this was due to species-specific profiles or simply an artifact of historical taxonomic identification limitations. Our analysis of *Leptogorgia chilensis* revealed that compounds previously reported only in other species across the genus were also present in this species, indicating they are not species-specific but instead conserved at the genus level. This suggests that the genus maintains a shared core chemical repertoire, while individual species may also evolve unique compounds, leading to both conserved scaffolds across species and lineage-specific chemical diversification. We expect that other species within the genus likewise maintain this ancestral chemistry while contributing their own innovations.

Our genomics work centers on how one species expanded and rewired terpene-biosynthetic pathways to generate chemical diversity. Via our pseudo-phased genomic assembly, we were able to discover a surprisingly high degree of TC heterozygosity. The relatively low overall heterozygosity (0.845%) in the coral suggests distinct selective pressures acting on TC diversification. Further, our phylogenetic analysis of TCs revealed numerous gene pairs that, given their high degree of sequence identity and physical proximity, appeared to result from gene duplications (**Fig 3**). Duplicated enzyme-encoding genes can follow one of three fates (**Fig 6a**).⁴⁴ In the first, one copy retains the original function while the other remains unchanged through

gene conservation. In an alternative scenario, deemed subfunctionalization, one copy acquires a partial new role alongside its original activity, or lastly, one copy evolves a completely novel function, a process called neofunctionalization. However, because we cannot yet predict octocoral TC functions from sequences alone, to determine which evolutionary trajectories the TC clades followed, biochemical characterization of the closely related enzymes was necessary.

Our biochemical characterization of the diverse suite of TCs present in *L. chilensis* revealed neofunctionalization and subfunctionalization arising from gene duplication events, each of which lends subtle distinct insights to the mechanisms by which TCs evolve new selectivity. For example, TC2a and TC2b, which share 89.37% protein identity, encode enzymes that produce **15** and **9**, respectively. Despite their sequence identity, these products arise via distinct reaction paths, that differ at the first step of their multi-step cascade mechanisms (**Fig 6b, SI Fig 29**).^{45,46} In another case, TC5 exhibits notable haplotype asymmetry, with three paralogs present on one haplotype and only a single, start-codon-lacking copy on the other. Two paralogs on haplotype A produce the previously undescribed guaiane-sesquiterpene **11**, while the third cyclase, TC5c, produces the aromadrendane-type sesquiterpene **16**, despite its 92.29% identity to its neighboring paralog TC5b. In both TC2 and TC5, neofunctionalization produced cyclases that not only generate novel products but also maintain a high level of product specificity, a striking outcome given that evolutionary changes might be expected to relax selectivity, leading to more promiscuous cyclization outcomes. Further, TC7a and TC7b share 94.17% of their amino acid sequence identity, but act on C15 and C20 precursors respectively to produce sesqui- and diterpene isoprenologs of b-curcumene (**Fig 6b**).

Exploring the idea that octocoral TCs are easily impacted by subtle changes in active site architecture, particularly residues that impose steric constraints within the binding pocket, we conducted enzyme mutagenesis with the two TC7 gene paralogs. To pinpoint which amino acid changes drove the shift in product specificity between the sesqui- and diterpene-producing paralogs, we individually swapped five of the six differing residues between TC7a and TC7b. This revealed that changes at position 91 had the greatest impact on function. We then introduced a range of alternative amino acids at this position and observed that the two enzymes did not respond identically to these substitutions. This pattern suggests epistatic effects: while the

other five residues did not independently alter function, they likely influence the enzyme's behavior by shaping the active site pocket, thereby modulating how changes at position 91 affect specificity.

Haplotype-specific variation in TC genes (**Fig 3**) underlies transcriptional regulation as a primary mechanism shaping the terpene profile of *L. chilensis*. Specifically, TC2a produced **9** in vitro and in vivo, while TC2b was transcriptionally silent and its characterized product, **15**, was not detected in the crude extract. However, given that the alleles of TC2a and TC2b had 100% sequence identity, it is probable that both functions of these paralogs have undergone positive selection and are selectively expressed based on biotic and abiotic factors. Additionally, in the case of TC5, hapA has three complete copies while hapB has none, and only one of the paralogs, TC5c, is transcribed. The capacity to produce **16** therefore is limited to one haplotype. Alternatively, TC7a and TC7b are both expressed in *L. chilensis* and their distinct products, the sesquiterpene (**12**) and diterpene (**17**) analogs of curcumene, are present in the crude extract. Also of note, while TC9 on hapA has a premature stop codon, TC9 on hapB is transcribed and its product **14** is detected in the crude extract. These findings suggest that *L. chilensis* diversifies its terpene repertoire not only by duplicating genes, but also by selectively expressing specific genes and alleles with distinct functions. However, because we sampled a single individual from one location at one time point for transcriptional analysis, we presume that TC transcriptional activities might vary depending on biotic and abiotic stimuli.

Whether through whole genome duplications (WGD) or TE-mediated sub-genomic duplications, polyploid organisms leverage their extra set of genomic information to generate metabolic diversity. Sub-genomic duplications can arise through two primary mechanisms. The first involves reverse transcription and reintegration, an RNA-mediated process typically associated with LTR and LINE transposable elements.^{38,40,47,48} The second is DNA-based duplication, which can result from unequal crossover, often producing clusters of two or more paralogous sequences. The role of transposable elements in this DNA-based mechanism remains less well understood.^{49,50} While plants have been found to utilize both WGDs and TEs to drive genomic innovation,^{44,51} invertebrates are known to maintain chromosomal stability over evolutionary timescales.⁵² Further, specifically in *L. chilensis*, we identified numerous TC gene pairs in TE dense regions that, given their high degree of sequence identity, appeared to result from gene duplications. Given the duplicated TC genes in *L. chilensis* retain introns and are flanked by CACTA and Mutator transposable elements, neither of which are

associated with RNA-mediated retrotransposition, an RNA-based mechanism is unlikely. While the exact role of the TEs remains unclear, the evidence points to a DNA-based duplication process. Whether this TE-associated mechanism is widespread across octocorals remains unknown, but it offers a compelling hypothesis for how invertebrates may generate chemical novelty without large-scale genomic expansion.

We have previously reported the presence of BGCs related to terpenoid biosynthesis in octocorals, which contain colocalized TCs, cytochrome p450s, and short chain dehydrogenases.²⁰ However, in the case of *L. chilensis*, 22 of the 24 TCs are not located within genomic neighborhoods that include other biosynthesis related genes, and only homologs of TC1 were found in putative BGCs. The absence of co-localized tailoring enzymes near these rapidly evolving TCs suggests that cyclase diversification may be outpacing the evolution of scaffold-specific tailoring machinery. It remains unclear whether octocoral terpenoid modifying enzymes are broadly promiscuous and capable of acting on a range of nascent scaffolds, or whether more tightly integrated pathways evolve only after a scaffold becomes ecologically or functionally entrenched.

Given the limited number of octocoral genomes currently available, comprehensive patterns of biosynthetic gene evolution remain largely unresolved. The diverse set of TCs identified in *L. chilensis* represents one of the first examples to be explored in detail and provides a foundation for understanding the origins and diversification of specialized metabolism in animals. A diverse terpenoid repertoire in these corals may serve functions beyond predator deterrence, potentially functioning as indicators of individual identity, reproductive status, or other social cues unique to their life history. Future work at the population level will be essential to uncover whether terpene profiles correlate with ecological or biological traits such as sex, colony maturity, or spatial structure. Together, our findings shed light on the genomic and evolutionary strategies by which a single octocoral lineage generates extraordinary chemical diversity. They also raise important questions about the pace and mode of metabolic innovation in animals. As additional octocoral genomes are sequenced and their biosynthetic capacity explored, these systems may reveal fundamental principles of enzyme evolution and open new doors for sustainable discovery and engineering of bioactive natural products.

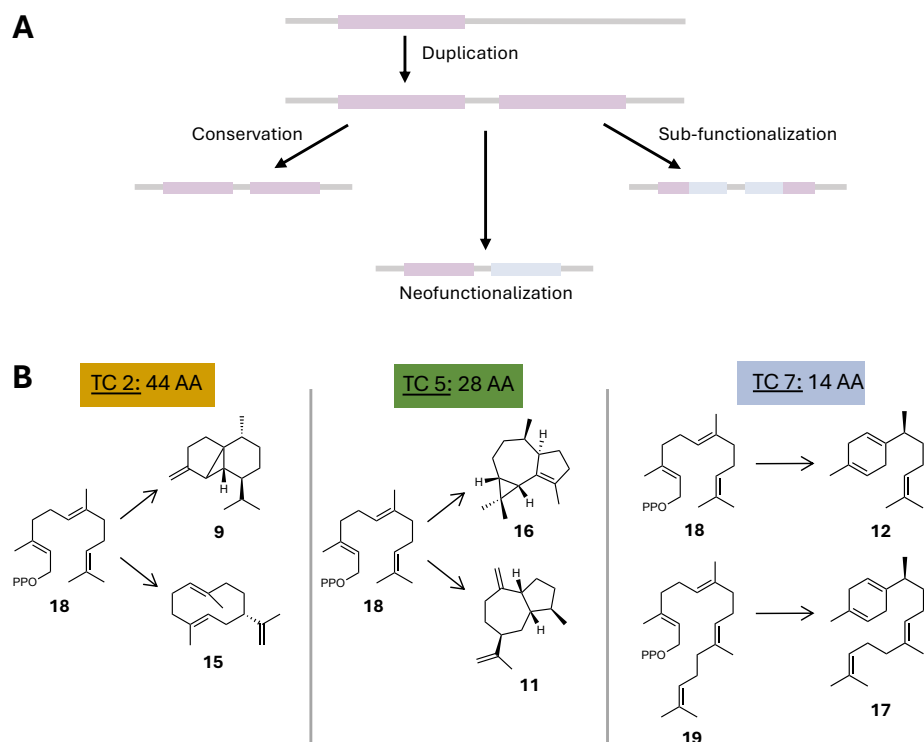


Figure 4-6: Suggested outcome of duplicated genes and comparison of paralog functions.

A- Outcomes of a duplicated gene. **B-** Product differentiation of three cases of neofunctionalized LTCs. Each case providing unique insights into potential evolved specificity. For proposed mechanism of TC2 cyclizations see **SI Fig 29**.

4.7 Main Text References

- (1) Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **2017**, 117 (17), 11570–11648. <https://doi.org/10.1021/acs.chemrev.7b00287>.
- (2) Gershenzon, J.; Dudareva, N. The Function of Terpene Natural Products in the Natural World. *Nat. Chem. Biol.* **2007**, 3 (7), 408–414. <https://doi.org/10.1038/nchembio.2007.5>.
- (3) Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R. Marine Natural Products. *Nat. Prod. Rep.* **2023**, 40 (2), 275–325. <https://doi.org/10.1039/D2NP00083K>.
- (4) Coll, J. C. The Chemistry and Chemical Ecology of Octocorals (Coelenterata, Anthozoa, Octocorallia). *Chem. Rev.* **1992**, 92 (4), 613–631. <https://doi.org/10.1021/cr00012a006>.
- (5) Barsby, T.; Kubanek, J. Isolation and Structure Elucidation of Feeding Deterrent Diterpenoids from the Sea Pansy, *Renilla Reniformis*. *J. Nat. Prod.* **2005**, 68 (4), 511–516. <https://doi.org/10.1021/np049609u>.
- (6) Keifer, P. A.; Rinehart, K. L. Jr.; Hooper, I. R. Renillafoulins, Antifouling Diterpenes from the Sea Pansy *Renilla Reniformis* (Octocorallia). *J. Org. Chem.* **1986**, 51 (23), 4450–4454. <https://doi.org/10.1021/jo00373a020>.

- (7) Changyun, W.; Haiyan, L.; Changlun, S.; Yanan, W.; Liang, L.; Huashi, G. Chemical Defensive Substances of Soft Corals and Gorgonians. *Acta Ecol. Sin.* **2008**, *28* (5), 2320–2328. [https://doi.org/10.1016/S1872-2032\(08\)60048-7](https://doi.org/10.1016/S1872-2032(08)60048-7).
- (8) Look, S. A.; Fenical, W.; Jacobs, R. S.; Clardy, J. The Pseudopterosins: Anti-Inflammatory and Analgesic Natural Products from the Sea Whip Pseudopterogorgia Elisabethae. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83* (17), 6238–6240.
- (9) Rocha, J.; Peixe, L.; Gomes, N. C. M.; Calado, R. Cnidarians as a Source of New Marine Bioactive Compounds—An Overview of the Last Decade and Future Steps for Bioprospecting. *Mar. Drugs* **2011**, *9* (10), 1860–1886. <https://doi.org/10.3390/md9101860>.
- (10) Fenical, W.; Okuda, R. K.; Bandurraga, M. M.; Culver, P.; Jacobs, R. S. Lophotoxin: A Novel Neuromuscular Toxin from Pacific Sea Whips of the Genus Lophogorgia. *Science* **1981**, *212* (4502), 1512–1514. <https://doi.org/10.1126/science.6112796>.
- (11) Culver, P.; Fenical, W.; Taylor, P. Lophotoxin Irreversibly Inactivates the Nicotinic Acetylcholine Receptor by Preferential Association at One of the Two Primary Agonist Sites. *J. Biol. Chem.* **1984**, *259* (6), 3763–3770.
- (12) Jia, Q.; Brown, R.; Köllner, T. G.; Fu, J.; Chen, X.; Wong, G. K.-S.; Gershenzon, J.; Peters, R. J.; Chen, F. Origin and Early Evolution of the Plant Terpene Synthase Family. *Proc. Natl. Acad. Sci.* **2022**, *119* (15), e2100361119. <https://doi.org/10.1073/pnas.2100361119>.
- (13) Dickschat, J. S. Bacterial Diterpene Biosynthesis. *Angew. Chem. Int. Ed.* **2019**, *58* (45), 15964–15976. <https://doi.org/10.1002/anie.201905312>.
- (14) Burkhardt, I.; de Rond, T.; Chen, P. Y.-T.; Moore, B. S. Ancient Plant-like Terpene Biosynthesis in Corals. *Nat. Chem. Biol.* **2022**, *18* (6), 664–669. <https://doi.org/10.1038/s41589-022-01026-2>.
- (15) Scesa, P. D.; Lin, Z.; Schmidt, E. W. Ancient Defensive Terpene Biosynthetic Gene Clusters in the Soft Corals. *Nat. Chem. Biol.* **2022**, *18* (6), 659–663. <https://doi.org/10.1038/s41589-022-01027-1>.
- (16) Wilson, K.; de Rond, T.; Burkhardt, I.; Steele, T. S.; Schäfer, R. J. B.; Podell, S.; Allen, E. E.; Moore, B. S. Terpene Biosynthesis in Marine Sponge Animals. *Proc. Natl. Acad. Sci.* **2023**, *120* (9), e2220934120. <https://doi.org/10.1073/pnas.2220934120>.
- (17) Chen, B.; Mao, J.; Xu, K.; Liu, L.; Lin, W.; Guo, Y.-W.; Wu, R.; Wang, C.; Xu, B. Mining Coral-Derived Terpene Synthases and Mechanistic Studies of the Coral Biflorane Synthase. *Sci. Adv.* **2025**, *11* (9), eadv0805. <https://doi.org/10.1126/sciadv.adv0805>.
- (18) Scesa, P. D.; Schmidt, E. W. Biomimetic Approach to Diverse Coral Diterpenes from a Biosynthetic Scaffold. *Angew. Chem. Int. Ed.* **2023**, *62* (39), e202311406. <https://doi.org/10.1002/anie.202311406>.
- (19) Scesa, P. D.; Schmidt, E. W. Pseudopterosin Biosynthesis: Unravelling a Decades Old Problem in Animal Specialized Metabolism. *J. Am. Chem. Soc.* **2025**, *147* (4), 3072–3079. <https://doi.org/10.1021/jacs.4c09925>.
- (20) Grayson, N. E.; Scesa, P. D.; Moore, M. L.; Ledoux, J.-B.; Gomez-Garrido, J.; Alioto, T.; Michael, T. P.; Burkhardt, I.; Schmidt, E. W.; Moore, B. S. A Widespread Metabolic Gene Cluster Family in Metazoans. *Nat. Chem. Biol.* **2025**, 1–10. <https://doi.org/10.1038/s41589-025-01927-y>.

- (21) Leferink, N. G. H.; Scrutton, N. S. Predictive Engineering of Class I Terpene Synthases Using Experimental and Computational Approaches. *Chembiochem* **2022**, *23* (5), e202100484. <https://doi.org/10.1002/cbic.202100484>.
- (22) Burkhardt, I.; Kreuzenbeck, N. B.; Beemelmans, C.; Dickschat, J. S. Mechanistic Characterization of Three Sesquiterpene Synthases from the Termite-Associated Fungus *Termitomyces*. *Org. Biomol. Chem.* **2019**, *17* (13), 3348–3355. <https://doi.org/10.1039/C8OB02744G>.
- (23) Wei, X.; Ning, W.; McCadden, C. A.; Alsup, T. A.; Li, Z.; Łomowska-Keehner, D. P.; Nafie, J.; Qu, T.; Opoku, M. O.; Gillia, G. R.; Xu, B.; Icenhour, D. G.; Rudolf, J. D. Exploring and Expanding the Natural Chemical Space of Bacterial Diterpenes. *Nat. Commun.* **2025**, *16* (1), 3721. <https://doi.org/10.1038/s41467-025-57145-6>.
- (24) Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. 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.* **2011**, *66* (1), 212–229. <https://doi.org/10.1111/j.1365-313X.2011.04520.x>.
- (25) Rai, A.; Hirakawa, H.; Nakabayashi, R.; Kikuchi, S.; Hayashi, K.; Rai, M.; Tsugawa, H.; Nakaya, T.; Mori, T.; Nagasaki, H.; Fukushima, R.; Kusuya, Y.; Takahashi, H.; Uchiyama, H.; Toyoda, A.; Hikosaka, S.; Goto, E.; Saito, K.; Yamazaki, M. Chromosome-Level Genome Assembly of *Ophiorrhiza Pumila* Reveals the Evolution of Camptothecin Biosynthesis. *Nat. Commun.* **2021**, *12* (1), 405. <https://doi.org/10.1038/s41467-020-20508-2>.
- (26) Zhao, D.; Hamilton, J. P.; Bhat, W. W.; Johnson, S. R.; Godden, G. T.; Kinser, T. J.; Boachon, B.; Dudareva, N.; Soltis, D. E.; Soltis, P. S.; Hamberger, B.; Buell, C. R. A Chromosomal-Scale Genome Assembly of *Tectona Grandis* Reveals the Importance of Tandem Gene Duplication and Enables Discovery of Genes in Natural Product Biosynthetic Pathways. *GigaScience* **2019**, *8* (3), giz005. <https://doi.org/10.1093/gigascience/giz005>.
- (27) Van Moerkercke, A.; Steensma, P.; Gariboldi, I.; Espoz, J.; Purnama, P. C.; Schweizer, F.; Miettinen, K.; Vanden Bossche, R.; De Clercq, R.; Memelink, J.; Goossens, A. The Basic Helix-Loop-Helix Transcription Factor BIS2 Is Essential for Monoterpenoid Indole Alkaloid Production in the Medicinal Plant *Catharanthus Roseus*. *Plant J.* **2016**, *88* (1), 3–12. <https://doi.org/10.1111/tpj.13230>.
- (28) Kapustina, I. I.; Makarieva, T. N.; Guzii, A. G.; Kalinovskiy, A. I.; Popov, R. S.; Dyshlovoy, S. A.; Grebnev, B. B.; von Amsberg, G.; Stonik, V. A. Leptogorgins A–C, Humulane Sesquiterpenoids from the Vietnamese Gorgonian *Leptogorgia* Sp. *Mar. Drugs* **2020**, *18* (6), 310. <https://doi.org/10.3390/md18060310>.
- (29) Ortega, M. J.; Zubía, E.; Sánchez, M. C.; Carballo, J. L. Cembrane Diterpenes from the Gorgonian *Leptogorgia Laxa*. *J. Nat. Prod.* **2008**, *71* (9), 1637–1639. <https://doi.org/10.1021/np8002639>.
- (30) Roussis, V.; Fenical, W.; Miralles, J.; Kornprobst, J. M. Piccolamine, an Unusual N,N-Dimethylaminomethylfuranosesquiterpene and Related Terpenoids from the Gorgonian *Leptogorgia Piccola*. *New J. Chem.* **1991**, No. 15, 959–961.
- (31) Martins, D. L.; Epifanio, R. de A. A New Germacrane Sesquiterpene from the Brazilian Endemic Gorgonian *Phyllogorgia Dilatata* Esper. *J. Braz. Chem. Soc.* **1998**, *9*, 586–590. <https://doi.org/10.1590/S0103-50531998000600012>.
- (32) McPhail, K. L.; Davies-Coleman, M. T.; Starmer, J. Sequestered Chemistry of the Arminacean Nudibranch *Leminda Millecra* in Algoa Bay, South Africa. *J. Nat. Prod.* **2001**, *64* (9), 1183–1190. <https://doi.org/10.1021/np010085x>.

- (33) Norte, M.; Fernández, José J.; González, R.; Rodríguez, M. L.; Ruiz-Pérez, C.; Rodríguez-Romero, V. (+)-Lepidozene a New Bicyclic Sesquiterpene from the Gorgonia *Lophogorgia Ruberrima*. *Tetrahedron* **1990**, *46* (24), 8237–8242. [https://doi.org/10.1016/S0040-4020\(01\)81479-4](https://doi.org/10.1016/S0040-4020(01)81479-4).
- (34) Coombe, L.; Warren, R. L.; Wong, J.; Nikolic, V.; Birol, I. ntLink: A Toolkit for De Novo Genome Assembly Scaffolding and Mapping Using Long Reads. *Curr. Protoc.* **2023**, *3* (4), e733. <https://doi.org/10.1002/cpz1.733>.
- (35) Burkhardt, I.; Dürr, L.; Grayson, N. E.; Moore, B. S. Methods for the Discovery and Characterization of Octocoral Terpene Cyclases. In *Methods in Enzymology*; Academic Press, 2024. <https://doi.org/10.1016/bs.mie.2024.02.011>.
- (36) Xu, H.; Dickschat, J. S. Turning a Sesquiterpene Synthase into a Di- and Sesterterpene Synthase. *ACS Catal.* **2023**, *13* (19), 12723–12729. <https://doi.org/10.1021/acscatal.3c03740>.
- (37) Hou, A.; Goldfuss, B.; Dickschat, J. S. Functional Switch and Ethyl Group Formation in the Bacterial Polytrichastrene Synthase from *Chryseobacterium Polytrichastri*. *Angew. Chem. Int. Ed.* **2021**, *60* (38), 20781–20785. <https://doi.org/10.1002/anie.202109465>.
- (38) Schrader, L.; Schmitz, J. The Impact of Transposable Elements in Adaptive Evolution. *Mol. Ecol.* **2019**, *28* (6), 1537–1549. <https://doi.org/10.1111/mec.14794>.
- (39) Boutanaev, A. M.; Osbourn, A. E. Multigenome Analysis Implicates Miniature Inverted-Repeat Transposable Elements (MITEs) in Metabolic Diversification in Eudicots. *Proc. Natl. Acad. Sci.* **2018**, *115* (28), E6650–E6658. <https://doi.org/10.1073/pnas.1721318115>.
- (40) Lynch, R. C.; Padgitt-Cobb, L. K.; Garfinkel, A. R.; Knaus, B. J.; Hartwick, N. T.; Allsing, N.; Aylward, A.; Bentz, P. C.; Carey, S. B.; Mamerto, A.; Kitony, J. K.; Colt, K.; Murray, E. R.; Duong, T.; Chen, H. I.; Trippe, A.; Harkess, A.; Crawford, S.; Vining, K.; Michael, T. P. Domesticated Cannabinoid Synthases amid a Wild Mosaic Cannabis Pangenome. *Nature* **2025**, 1–10. <https://doi.org/10.1038/s41586-025-09065-0>.
- (41) Ou, S.; Su, W.; Liao, Y.; Chougule, K.; Agda, J. R. A.; Hellinga, A. J.; Lugo, C. S. B.; Elliott, T. A.; Ware, D.; Peterson, T.; Jiang, N.; Hirsch, C. N.; Hufford, M. B. Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline. *Genome Biol.* **2019**, *20* (1), 275. <https://doi.org/10.1186/s13059-019-1905-y>.
- (42) Lyu, C.; Chen, T.; Qiang, B.; Liu, N.; Wang, H.; Zhang, L.; Liu, Z. CMNPD: A Comprehensive Marine Natural Products Database towards Facilitating Drug Discovery from the Ocean. *Nucleic Acids Res.* **2021**, *49* (D1), D509–D515. <https://doi.org/10.1093/nar/gkaa763>.
- (43) Maloney, K. N.; Botts, R. T.; Davis, T. S.; Okada, B. K.; Maloney, E. M.; Leber, C. A.; Alvarado, O.; Brayton, C.; Caraballo-Rodríguez, A. M.; Chari, J. V.; Chicoine, B.; Crompton, J. C.; Davis, S. R.; Gromek, S. M.; Kurnianda, V.; Quach, K.; Samples, R. M.; Shieh, V.; Sultana, C. M.; Tanaka, J.; Dorrestein, P. C.; Balunas, M. J.; McFadden, C. S. Cryptic Species Account for the Seemingly Idiosyncratic Secondary Metabolism of Sarcophyton Glaucum Specimens Collected in Palau. *J. Nat. Prod.* **2020**, *83* (3), 693–705. <https://doi.org/10.1021/acs.jnatprod.9b01128>.
- (44) Panchy, N.; Lehti-Shiu, M.; Shiu, S.-H. Evolution of Gene Duplication in Plants.
- (45) Lopez-Gallego, F.; Agger, S. A.; Abate-Pella, D.; Distefano, M. D.; Schmidt-Dannert, C. Sesquiterpene Synthases Cop4 and Cop6 from *Coprinus Cinereus*: Catalytic Promiscuity and Cyclization of Farnesyl Pyrophosphate Geometric Isomers. *ChemBioChem* **2010**, *11* (8), 1093–1106. <https://doi.org/10.1002/cbic.200900671>.

- (46) Ye, D.; Shao, Y.-Z.; Li, W.-R.; Cui, Z.-J.; Gong, T.; Yang, J.-L.; Wang, H.-Q.; Dai, J.-G.; Feng, K.-P.; Ma, M.; Ma, S.-G.; Liu, Y.-B.; Zhu, P.; Yu, S.-S. Characterization and Engineering of Two Highly Paralogous Sesquiterpene Synthases Reveal a Regioselective Reprotonation Switch. *Angew. Chem. Int. Ed.* **2024**, 63 (13), e202315674. <https://doi.org/10.1002/anie.202315674>.
- (47) Parker, H. G.; VonHoldt, B. M.; Quignon, P.; Margulies, E. H.; Shao, S.; Mosher, D. S.; Spady, T. C.; Elkahouloun, A.; Cargill, M.; Jones, P. G.; Maslen, C. L.; Acland, G. M.; Sutter, N. B.; Kuroki, K.; Bustamante, C. D.; Wayne, R. K.; Ostrander, E. A. An Expressed Fgf4 Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs. *Science* **2009**, 325 (5943), 995–998. <https://doi.org/10.1126/science.1173275>.
- (48) Zhang, J.; Dean, A. M.; Brunet, F.; Long, M. Evolving Protein Functional Diversity in New Genes of *Drosophila*. *Proc. Natl. Acad. Sci.* **2004**, 101 (46), 16246–16250. <https://doi.org/10.1073/pnas.0407066101>.
- (49) McKenzie, S. K.; Kronauer, D. J. C. The Genomic Architecture and Molecular Evolution of Ant Odorant Receptors. *Genome Res.* **2018**, 28 (11), 1757–1765. <https://doi.org/10.1101/gr.237123.118>.
- (50) Schrader, L.; Kim, J. W.; Ence, D.; Zimin, A.; Klein, A.; Wyschetzki, K.; Weichselgartner, T.; Kemena, C.; Stökl, J.; Schultner, E.; Wurm, Y.; Smith, C. D.; Yandell, M.; Heinze, J.; Gadau, J.; Oettler, J. Transposable Element Islands Facilitate Adaptation to Novel Environments in an Invasive Species. *Nat. Commun.* **2014**, 5, 5495. <https://doi.org/10.1038/ncomms6495>.
- (51) Clark, J. W.; Donoghue, P. C. J. Whole-Genome Duplication and Plant Macroevolution. *Trends Plant Sci.* **2018**, 23 (10), 933–945. <https://doi.org/10.1016/j.tplants.2018.07.006>.
- (52) Simakov, O.; Bredeson, J.; Berkoff, K.; Marletaz, F.; Mitros, T.; Schultz, D. T.; O’Connell, B. L.; Dear, P.; Martinez, D. E.; Steele, R. E.; Green, R. E.; David, C. N.; Rokhsar, D. S. Deeply Conserved Synteny and the Evolution of Metazoan Chromosomes. *Sci. Adv.* **2022**, 8 (5), eabi5884. <https://doi.org/10.1126/sciadv.abi5884>.

4.8 Supplementary Information

Methods

Sample collection.

Divers collected *Leptogorgia chilensis* specimen from La Jolla, CA by hand in accordance with federal and state laws with the required license using non-invasive methods. We sampled small portions of a larger colony using a knife. We then transported the sample back to the lab where it was washed 3 times for 5 min each with Ca²⁺- and Mg²⁺-free artificial seawater within 24 hours of collection then flash frozen with liquid nitrogen as a live organism. At the time of flash freezing, an aliquot was preserved in RNA-later. All biologic material was and kept at -80 °C for less than 2 weeks until DNA extraction and RNA extraction.

Genomic DNA Extraction and Sequencing.

High molecular weight (HMW) DNA was isolated from tissue from *L. chilensis* as previously described.¹ In brief, tissue was homogenized by mortar and pestle while frozen with liquid nitrogen. The Qiagen genomic-

tip kit 500/G (Cat# 10262) was used, following the protocol for the “Preparation of Tissue Samples” with the recommendations of tissue amounts for mouse tail. DNA purity was determined using a BioDrop μ LITE (BioDrop, Holliston, MA, USA), confirming that the OD 260/280 was 1.8 and the OD 260/230 was between 2.0–2.2. DNA concentrations were assessed using the dsDNA HS assay on a Qubit fluorometer v3.0 (Thermo Fisher Scientific, Singapore). DNA fragment size distribution was assessed with a TapeStation (Agilent 4150, Santa Clara, CA, USA). The high molecular weight (HMW) DNA was sequenced using three technologies: Illumina (San Diego, CA, USA), Oxford Nanopore Technologies (ONT; Oxford, United Kingdom), and Pacific Biosciences (PacBio) HiFi (Menlo Park, CA, USA). For the Illumina libraries, DNA was acoustically sheared with a Covaris, a library processed with the Illumina DNA library prep (20060059) and sequenced on a NovaSeq6000. Specifically, the sample was sequenced on a S4 300 cycle run which generated 448 million reads (Salk Institute for Biological Studies). For long read sequencing, the DNA was further cleaned and short fragments removed by using a 0.4x Ampure cleanup along with the Short Read Eliminator XS kit (Cat# 102-208-200, Circulomics-PacBio, Menlo Park, CA, USA). The cleaned HMW DNA was then sequenced with ONT LSK-112 chemistry (SQK-LSK112) on an R10.4.1 flow cell. For PacBio sequencing, the cleaned HMW DNA was sheared at “Speed 31”) using Megaruptor 3 (Diagenode, Seraing, Belgium, Cat# B06010003) and processed for library preparation (SMRTbell 3.0, Cat #. (In addition, the cleaned HMW DNA was library also sequenced using PacBio HiFi using SMRTbell 3.0 (Cat# 102-182-700) on a Revio instrument.

Transcriptomic Extraction and Sequencing.

RNA was extracted from the tissues of *L. chilensis* preserved in RNA-later as previously described.^{2,3} In brief, the RNeasy Plant Mini Kit was used (Qiagen, Cat # 74904) including the optional step of an on column DNase digestion. Quality and quantity were both assessed via Nanodrop spectrophotometric and Qubit fluorometric methods as described in DNA extraction. cDNA libraries were prepared from total RNA with Illumina TruSeq Stranded Total RNA kit (Cat # 20020599) with rRNA depletion steps, barcoded with IDT TruSeq RNA indexes (Cat # UDI-20022371).

Genome size estimation.

Genome size was estimated by K-mer frequency distribution analysis. First, the short reads were filtered using FastP (v0.19.4 & v23.2)⁴ with default parameters and quality checked with FastQC (v 0.11.9).⁵

Subsequently, the K-mers (k=21 bp) were counted using Jellyfish (v 2.2.10)⁶ with the parameters “-C -m 21 -s 1G -t 50”. The output histogram file was then analyzed with GenomeScope⁷ to estimate the genome size with parameters k=21, ploidy= 2. The K-mer frequency distribution displayed a characteristic bimodal pattern, consistent with a diploid genome exhibiting low heterozygosity (see **SI Table 11** for all bioinformatic tools and version used). GenomeScope estimated the haploid genome size to be ~194.3 Mb, with approximately 0.86% heterozygosity and ~34% of the genome composed of repetitive elements. The model achieved a fit between 70.1% and 91.1%, and the estimated sequencing error rate was 0.35% (**SI Table 1, SI Fig 2**).

Genome Assembly.

Nanoplot (v 1.41.0)⁸ was used to determine the quality of the long read sequencing data obtained for *L. chilensis*. In the case of the ONT reads generated, filtlong (v 0.2.1) was used to remove reads that were shorter than 30kb and had an average phred quality score of at lower than 10 (maintaining 27K reads of the original 7M reads- less than 0.5%). The ultralong reads from ONT and the PacBio reads were assembled with Hifiasm (v 0.16.1)⁹ with the -ul parameter. The assembled genome was checked for duplication (D%) rate and completeness (C%) with Benchmarking Universal Single-Copy Orthologs (BUSCO v 5.4.7)¹⁰ using a metazoan database as reference. Assembly contiguity was determined by QUAST (v 5.2.0).¹¹

Gene Annotation and Collinearity.

We first curated a repeat library using RepeatModeler (v 2.0.7)¹² on each haplotype of the genome. Next, we used RepeatModeler using our genomic database and created a soft masked genome with RepeatMasker (v 4.1.2). We predicted gene models with the TSEBRA pipeline (using Braker v 2.1.6).^{13–15} We used an existing developed Snakemake workflow for running TSEBRA, available here: https://gitlab.com/salk-tm/snake_tsebra. The specific protein library we used contained proteomes from 15 different cnidaria: *Acropora digitifera*, *Exaiptasia diaphana*, *Orbicella faveolate*, *Stylophora pistillata*, *Pocillopora damicornis*, *Dendronephthya gigantea*, *Actinia tenebrosa*, *Acropora millepora*, *Xenia* sp. Carnegie-2017, *Acropora muricata*, *Pocillopora verrucosa*, *Montipora capricornis*, *Montipora foliosa*, *Nematostella vectensis*, *Porites lutea*. We also used the RNA-seq data generated from the *L. chilensis* tissue to enhance our gene model. The RNA-seq reads were aligned to the genome with Hisat2 (v 2.2.1).¹⁶ Putative functional annotations of gene models were assigned using eggno-mapper (v2.0.1).^{17,18} Overall gene model quality and completeness was

assessed by proteome BUSCO (v 5.4.7)¹⁰ scores. With the gene models from each haplotype, we next ran Orthofinder (v 2.5.4),¹⁹ then found and visualized syntenic blocks with jcvr-tools (v 1.5.6).²⁰

Transcriptomic *De Novo* Assembly and Alignment Based.

The trimmed and adapter removed (FastQC and FastP with default parameters) cDNA sequences from *L. chilensis* were assembled with Trinity (v 2.8.5)²¹ using default parameters. The processed cDNA sequences were also mapped to the assembled genomes using HiSat2 (v 2.2.1).¹⁶

TC Mining and Haplotype Difference Validation.

To identify terpene cyclases in the *L. chilensis* genome assembly, we used the program getorf (EMBOSS v 6.6.0.0)²² to find and translate ORFs. The translated AA sequences were then searched against the curated octocoral TC hmmfile provided by Burkhardt *et al.*³, utilizing the HMMER hmmsearch program (version: HMMER 3.3.2 Nov 2020)²³ to identify amino acid sequences assigned to terpene cyclases. Results with an E-value below the threshold of 2.3×10^{-8} were considered for further analysis.

Heterologous strain growth conditions and expression on analytical scale.

For all TC characterization we used *E. coli* BL21 strains transformed with the targeted gene for expression in a pET-28a vector following the methods in Burkhardt *et al.* All TC encoding sequences were either cloned out of genomic DNA or ordered as original nucleotide sequences (Twist Biosciences). In brief, we grew up our target gene containing strains in TB media and induced protein production with IPTG (400 mM final concentration). After 12-18 hours of protein production at 18 °C we pelleted the cells with centrifugation and subsequently resuspended the lysis with cold TC binding buffer. We then disrupted the cells with sonification and pelleted the cell debris with centrifuge. We then purified the proteins using Ni-NTA columns with TC binding and elution buffers. Lastly, we used PD-10 columns to solubilize our TCs in TC incubation buffer.

***In vivo* TC enzyme assays on analytical scale.**

We determined the concentration of all purified proteins by Bradford assay using manufacturing standard protocols. TC enzyme assays were set up as previously established in Burkhardt *et al.* However, because no monoterpenes were present in the crude extract, we tested each enzymes ability to act on FPP and GGPP as substrates. In brief, both FPP and GGPP were suspended in 50 mM ammonium bicarbonate at a concentration of 3.75 mg/mL. Reactions for each enzyme and substrate were prepared each containing a 1 mM

enzyme solution in 1.490 mL TC reaction buffer and 10 mL of each substrate stock solution (0.25mg/mL end concentration). After a 12-18 hour incubation at room temperature reactions were extracted with 200 mL of hexane and analyzed by GCMS (Agilent 7890A gas chromatograph with an Agilent 5975C mass spectrometer using an Rtx-5Sil 30 m × 0.25 mm, 0.25- μ m column). The used temperature oven program was as follows: Hold 70 °C for 3 min, 10 °C min⁻¹ to 325 °C and hold at 325 °C for 3 min. The flow rate was 1 mL min⁻¹, and the injection volume was 1 μ L. The splitless injection inlet temperature was 270 °C, the MS transfer line temperature was 280 °C, the MS source temperature was 230 °C and the MS quad temperature was 150 °C.

Heterologous expression preparative scale.

In the cases that a standard was not readily available, or a reasonable database hit could not be found, we set up large scale enzyme assays and conducted subsequent product purification. Full detailed methods can be found in Burkhardt et al. In short, for all large-scale TC characterizations we used the original *E. coli* BL21 strains transformed with the targeted gene for expression as mentioned in the above preparative scale assays. 3x 1L of TB medium were inoculated with 1% *E. coli* BL21 culture. Cultures were grown to an O.D. of 0.6-0.8 at 37 °C with shaking. We then induced protein production with IPTG (400 mM final concentration) and after 12-18 hours of protein production at 18 °C we pelleted the cells with centrifugation (8,000 g for 15 min at 4 °C) and subsequently resuspended the lysis with cold TC binding buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 20 mM imidazole, 1 mM MgCl₂ at pH 7.4). We then disrupted the cells with sonification (15 seconds on, 45 seconds off, for a total of 7 minutes on-time) and pelleted the cell debris with centrifugation (11,000 g for 20 mins at 4 °C). The cell pelleting was repeated a second round with the addition of 1-2 mg of DNase I (Roche Diagnostics). We then filtered the proteins with a 0.45 μ m syringe and loaded them onto a 5 mL HisTrapFF-Ni columns (GE) for purification. We used a peristaltic pump at a flow rate of ~4 mL per min in conjunction with the column. We washed the protein with TC binding buffer (40 mL) and eluted with TC elution buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 0.5 M imidazole, 1 mM MgCl₂ at pH 7.4), collecting in 2 mL fractions. Fractions were pooled and loaded onto a PD-10 column, pre-equilibrated with 25 mL TC incubation buffer (50 mM Hepes, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol at pH 7.8) in 2.5 mL batches. Each column was eluted with 3.5 mL of TC incubation buffer

(50 mM Hepes, 5 mM MgCl₂, 250 mM NaCl, 10% glycerol at pH 7.8). The protein was then used in enzyme assays or flash-frozen in liquid nitrogen and stored at -80 °C.

TC enzyme assays on preparative scale.

Given that we had at this point already characterized all TCs with both FPP and GGPP, we set up the enzyme assay with our substrate of interest to generate the necessary unknown products. FPP (80 mg) was dissolved in ammonium bicarbonate solution (50 mM in 15 mL) and added dropwise to 600 mL of TC incubation buffer with 3 µM enzyme concentration over 5 hours. We determined the concentration of all purified proteins by Bradford assay using manufacturing standard protocols. After a 12–18 hour incubation at room temperature, the reactions were extracted with pentane (3 x 200 mL). Organic fractions containing terpene were dried over MgSO₄ and filtered. The resulting filtrate was then concentrated to ~1 mL. The crude fraction was then purified using silica gel chromatography (100% pentane). Individual fractions were checked with thin layer chromatography (silica plate, 100% hexane mobile phase) and stained with phosphomolybdic acid. Fractions containing the product were pooled, concentrated to dryness, and analyzed with NMR.

Heterologous expression using co-expression plasmid.

In the case of LcTC7b-HapB, we conducted our scale up protein expression using a co-expression plasmid. The plasmid used and subsequent methods have been adapted from Xu *et al.*²⁴ In brief, we used an *E. coli* BL21 strain co-transformed with the plasmid PCDF-MK14, which encodes for the overproduction of GGPP, and the pET-28a plasmid containing LcTC7b-HapB. 6x 1L medium were inoculated with 1% of this *E. coli* BL21 culture. Cultures were grown to an O.D. of 0.6-0.8 at 37 °C with shaking (~4 hours). We then induced protein production with IPTG (400 mM final concentration per culture) and isoprenol (4 mL per culture) and incubated the cultures for 24 hours with shaking at 18 °C for sufficient protein production. Cells were pelleted by centrifugation (15,000g, 10 min, 4 °C) and either frozen and stored at -80 °C or extracted immediately.

Purification of terpene product from co-expression.

Cell pellets from *in vivo* scale up terpene production were extracted with 50mL acetone with stirring at room temperature overnight. The extract was filtered and concentrated to dryness. The filtrate was then resuspended in 10 mL of water and partitioned with 2 mL of hexanes. Liquid partitioning was repeated three times in total.

The organic terpene containing fraction was then purified by silica gel chromatography with the same methods as out *in vitro* scale up purification process. Fractions containing products were pooled, concentrated to dryness, and analyzed using NMR spectroscopy.

GCMS analysis of *L. chilensis* crude extract.

300 mg of frozen *L. chilensis* tissue was thawed and then macerated in 2 mL MeOH/CH₂Cl₂ 2:1. The mixture was stirred for 12 hours in a closed 20 mL scintillation vial at room temperature. The resulting suspension was diluted with water (10 mL), 1 mL of hexane was added, and the sample mixed by vortexing for 10 seconds. After layer separation, the top layer was carefully pipetted off into GCMS vials and directly subjected to GCMS analysis (Agilent 7890A gas chromatograph with an Agilent 5975C mass spectrometer using an Rtx-5Sil 30 m × 0.25 mm, 0.25-μm column). The used temperature oven program was as follows: Hold 70 °C for 3 min, 10 °C min⁻¹ to 325 °C and hold at 325 °C for 3 min. The flow rate was 1 mL min⁻¹, and the injection volume was 1 μL. The splitless injection inlet temperature was 270 °C, the MS transfer line temperature was 280 °C, the MS source temperature was 230 °C and the MS quad temperature was 150 °C.

Comparison of TC7 paralogs.

Using all four LcTC 7s (HapA-7a, HapA-7b, HapB-7a, HapB7b), we performed a multiple sequence alignment with kalign2 (v 2.04).²⁵ We manually inspected the output alignment to identify point differences between all 4 enzymes. We also made protein models of a representative forming (+)-prenyl-β-curcumenone (**17**) synthase, LcTC7b-HapB, and a (+)-γ-curcumenone (**12**) synthase, LcTC7a-HapB. The web-based Swiss-model²⁶ interface with the crystal structure of ErTC1³ as the template was used to generate the two models. The resulting pdb files were visualized with pymol (v 3.1.0).

Enzyme mutagenesis.

The point mutants of the forming **17** synthase, LcTC7b-HapB, and γ-curcumenone synthase, LcTC7a-HapB were made using standard molecular cloning techniques. In short, template DNA of each of the two enzymes was used as the starting material for PCR based mutant generation. Primers specific to the DNA encoding the AA residues of interest were designed (**SI Table 5**). Two rounds of PCR reactions were set up according to a standard protocol (Phusion High-Fidelity DNA Polymerase Master Mix, Thermo Scientific) based

on the template DNA size and primer design. The first round was to insert the point mutants at a site of interest, FWD and REV primers contained the desired mutation and amplified out from the mutant position. The second round of PCR, the products of the first rounds were used as the template and the primers were the start and ends of the entire TC sequence, joining the two partial sequences generated in the first round. An example of the PCR conditions is reported in **SI Table 6**.

The resulting products were purified via a gel purification kit (Zymoclean Gel DNA Recovery Kit (ZymoResearch)). Next, the purified mutant terpene cyclase encoding genes were inserted into the *E. coli* expression vector pET28a(+). Subsequently, the plasmid was transformed into *E. coli* DH5α using chemically competent cells using the heat shock transformation method. Successful cloning and transformation of the template DNA was checked by sequencing the mini prepped plasmids (Quick-DNA Miniprep Kit, ZymoResearch). The plasmids were then transformed into *E. coli* BL21 (DE3) strains for heterologous expression of proteins.

TC Specific TE Analysis

EDTA (v 1.9.697)²⁷ was used to identify TEs in the *L. chilensis* genome using the following command: `EDTA.pl --genome {inputFastaFile} --anno 1 --threads 20`. Bedtools (v 2.26.0)²⁸ was used to pull out 100kbp regions surrounding all TC coding genes and 100 kbp random regions across both haplotypes. The intersect feature was then used to find all regions in the 100 kbp windows that contained TEs. The output from this command was formatted and processed in Microsoft excel and visualized with R studios ggplot2 package.²⁹

Structure elucidation of TC products

General methods

All chemicals and solvents were used as received from the commercial supplier (SigmaAldrich, Fisher, or TCI). The substrates GPP, FPP, and GGPP were synthesized as reported previously. Thin layer chromatography was performed with thin layer chromatography plates purchased from Merck (silica gel 60 F254, glass plates). Column chromatography and normal phase HPLC was performed using a Flash EZ prep combiflash system (Teledyne ISCO). NMR spectra were recorded on a Bruker Avance III spectrometer (600 MHz) using a 1.7-mm inverse detection triple resonance (H-C/N/D) cryoprobe or a JEOL ECZ spectrometer

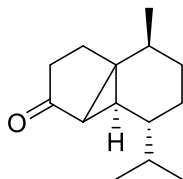
(500 MHz) and were referenced against deuterated benzene ($\delta = 7.16$ ppm) or deuterated chloroform ($\delta = 7.26$ ppm) for ^1H -NMR and deuterated benzene ($\delta = 128.06$ ppm) or deuterated chloroform ($\delta = 77.16$ ppm) for ^{13}C -NMR. Data for NMR spectra are reported as follows: shift (δ) in ppm; s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet or unresolved; br, broad signal; J, coupling constant(s) in Hz. Data were analyzed using Mestrenova 12.0.0. Optical rotation measurements were performed using a Jasco P-2000 polarimeter.

Derivatization of cubebene-containing mixture and isolation of derivative S1.

The main product of TC2a (likely β -cubebene according to its MS spectrum) was unstable on silica and could not be separated from the side products. We derivatized it to its nor-oxo derivative by oxidative double bond cleavage³⁰ to assure its identity and determine its absolute configuration.

For this, the terpene hydrocarbon mixture obtained from the preparative scale TC2a reaction (6 mg, 0.029 mmol, 1 eq) was dissolved in 1.5 mL acetone/ water (10:1). Subsequently, 2,6-lutidine (6.3 mg, 0.058 mmol, 2.0 eq), 4-methylmorpholine *N*-oxide (5.2 mg, 0.044 mmol, 1.5 eq) and osmium tetroxide (4% in water, 3.75 μL , 0.6 μmol , 0.02 eq) were added and the resulting mixture was stirred at room temperature until no more hydrocarbons were detected in the mixture by TLC (2 hours reaction time). Next, (Diacetoxyiodo)benzene (20 mg, 0.059 mmol, 2.0 eq) was added and stirring was continued for 18 hours before the reaction was quenched with aqueous sodium thiosulfate (saturated, 3 mL). The resulting mixture was diluted with water (10 mL) and extracted with EtOAc (3x 3 mL). The organic layer was dried over MgSO_4 , concentrated *in vacuo* and purified by flash column chromatography (hexane/EtOAc 95:5). Compound **S1** was obtained as a colorless oil (1.5 mg, 0.007 mmol, 25%).

Analytical data for synthetic derivative of the TC2a product (**S1**):

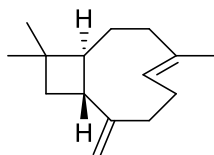


NMR: $^1\text{H-NMR}$ (500 MHz, CDCl_3): δ = 2.18-2.12 (m, 2H), 2.08-2.01 (m, 1H), 1.86-1.79 (m, 2H), 1.69-1.60 (m, 2H), 1.51-1.45 (m, 1H), 1.48 (d, J = 2.2 Hz, 1H), 1.26 (t, J = 2.7, 1H), 1.21-1.14 (m, 1H), 0.99-0.90 (m, 1H), 0.98 (d, J = 6.5 Hz, 3H), 0.93 (d, J = 6.7 Hz, 3H), 0.90 (d, J = 6.8 Hz, 3H), 0.58 (tdd, J = 13.6, 11.6, 2.4 Hz, 1H) ppm. **$^{13}\text{C-NMR}$** (125 MHz, CDCl_3): 214.8, 43.5, 40.5, 39.9, 33.5, 33.4, 32.7, 31.5, 30.9, 26.8, 26.1, 20.1, 19.6, 19.1 ppm.

Optical rotation: $[\alpha]_D^{28.0} = +12.7$ (c 0.05, CHCl_3).

NMR data agree with reported values for the stereo-selectively synthesized compound. The absolute configuration of compound **S1**, and therefore (+)- β -cubebene (**9**), was set comparing the determined optical rotation of the compound to the value from the same report.^{31,32}

Analytical data for TC4 ((+)- β -caryophyllene **10**):

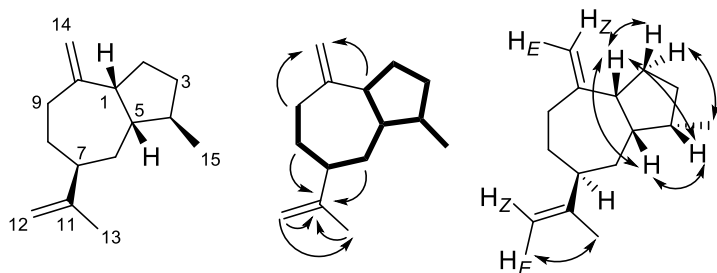


NMR: $^1\text{H-NMR}$ (500 MHz, C_6D_6): δ = 5.40-5.33 (m, 1H), 5.03 (s, 1H), 4.87 (s, 1H), 2.42-2.24 (m, 2H), 2.21-2.11 (m, 1H), 2.08-1.80 (m, 4H), 1.75-1.66 (m, 2H), 1.60-1.45 (m, 1H), 1.56 (s, 3H), 1.42-1.28 (m, 2H), 1.00 (s, 3H), 0.94 (s, 3H) ppm. **$^{13}\text{C-NMR}$** (125 MHz, CDCl_3): 154.6, 135.3, 124.8, 112.2, 53.8, 48.9, 40.7, 40.3, 35.1, 33.1, 30.2, 29.7, 28.8, 22.8, 16.5 ppm.

Optical rotation: $[\alpha]_D^{28.0} = +5.7$ (c 0.16, CHCl_3).

$^1\text{H-NMR}$ data are in agreement with reported values from the literature.³¹ The absolute configuration was assigned by comparing the determined optical rotation to the literature.³¹

Analytical data for TC5a and TC5b product (guaia-11(12),6(14)-diene **11**):

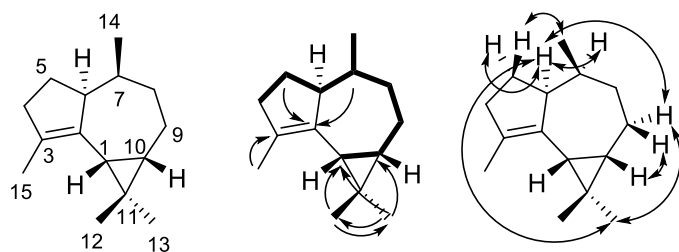


C ^a	¹³ C ^b	¹ H ^b
1	48.7	2.69-2.62 (m, 1H)
2	28.3	1.81-1.70 (m, 2H)
3	32.6	1.65-1.57 (m) 1.26 (ddt, ² J = 12.4, ³ J = 8.4, ³ J = 5.6)
4	38.9	1.93-1.86 (m)
5	40.8	2.03 (dddd, ³ J = 10.7, ³ J = 8.4, ³ J = 6.6, ³ J = 2.2)
6	27.3	1.48 (ddd, ² J = 14.4, ³ J = 4.5, ³ J = 2.4) 1.38 (ddd, ² J = 14.3, ³ J = 11.0, ³ J = 5.4)
7	42.6	2.27-2.20 (m, 1H)
8	31.8	1.91-1.82 (m, 1H) 1.57-1.51 (m, 1H)
9	34.4	2.27-2.20 (m, 1H) 2.37-2.29 (m, 1H)
10	153.4	-
11	149.7	-
12	109.7	4.86 (brs, 1H) 4.839 (brs, 1H)
13	22.1	1.66 (s, 3H)
14	109.5	4.91 (brs) 4.842 (brs)
15	15.9	0.86 (d, ³ J = 5.4, 3H)

^aCarbon numberings as shown above table; ^bchemical shifts are given in ppm, coupling constants are given in Hertz, multiplicity: m = multiplet, s = singlet, d = doublet, t = triplet, q = quartet, br = broad.

NMR: See Table. **Optical rotation:** $[\alpha]_D^{28.0} = -0.57$ (c 0.37, CHCl₃).

Analytical data for TC5c product ((+)- α -gurjunene **16**):



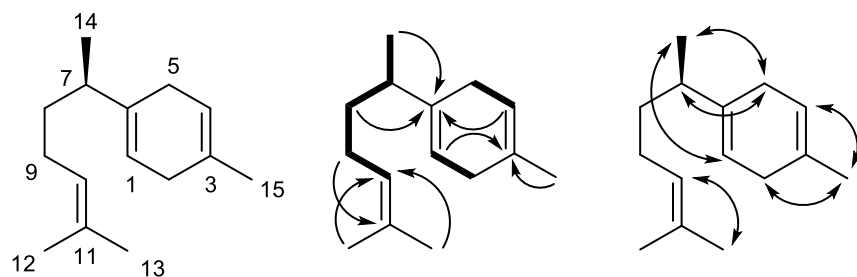
C ^a	¹³ C ^b	¹ H ^b
1	26.1	1.13-1.09 (m, 1 H)
2	136.4	-
3	137.6	-
4	36.7	2.35-2.26 (m, 1 H) 2.21-2.14 (m, 1 H)
5	29.4	1.88-1.81 (m, 1 H) 1.63-1.53 (m, 1 H)
6	49.1	2.76-2.70 (m, 1 H)
7	32.9	1.91-1.83 (m, 1 H) 1.42-1.33 (m, 1 H)
8	32.6	1.80-1.73 (m, 1 H)
9	21.9	1.83-1.75 (m, 1 H) 1.40-1.28 (m, 1 H)
10	28.7	0.79-0.73 (m, 1 H)
11	22.0	-
12	29.0	1.11 (s, 3 H)
13	16.7	0.92 (s, 3 H)
14	18.7	1.07 (d, $J = 7.1$ Hz, 3H)
15	14.8	1.69 (m, 3 H)

^aCarbon numberings as shown above table; ^bchemical shifts are given in ppm, coupling constants are given in Hertz, multiplicity: m = multiplet, s = singlet, d = doublet, t = triplet, q = quartet, br = broad.

NMR: See Table. **Optical rotation:** $[\alpha]_D^{28.0} = +77.9$ (c 0.018, CHCl₃).

The absolute configuration was assigned by comparing the determined optical rotation to the literature.³³

Analytical data for HapB TC7a product ((+)- β -curcumene **12**):



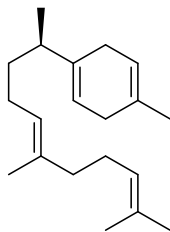
C ^a	¹³ C ^b	¹ H ^b
1	118.4	5.46 (s, 1H)
2	32.0	2.57-2.51 (m, 2H)
3	131.2	-
4	119.3	5.44 (s, 1H)
5	26.9	2.60-2.54 (m, 2H)
6	139.1	-
7	40.7	2.18-2.09 (m, 1H)
8	35.4	1.54-1.45 (m, 1H) 1.38-1.29 (m, 1H)
9	26.6	2.07-1.97 (m, 2H)
10	125.4	5.25-5.20 (m, 1H)
11	131.0	-
12	25.9	1.69 (s, 3H)
13	17.8	1.56 (s, 3H)
14	19.9	1.02 (d, ³ J = 10.7, 1H)
15	23.2	1.61 (s, 3H)

^aCarbon numberings as shown above table; ^bchemical shifts are given in ppm, coupling constants are given in Hertz, multiplicity: m = multiplet, s = singlet, d = doublet, t = triplet, q = quartet, br = broad.

NMR: See Table. **Optical rotation:** $[\alpha]_D^{28.0} = +11.6$ (c 0.25, CHCl₃).

The absolute configuration was assigned by comparing the determined optical rotation to the literature.³⁴

Analytical data for HapB TC7b product ((+)-prenyl- β -curcumene **17**):



NMR: **¹H-NMR** (500 MHz, C₆D₆): δ = 5.48-5.45 (m, 1H), 5.45-5.42 (m, 1H), 5.30-5.22 (m, 2H), 2.62-2.49 (m, 4H), 2.23-2.07 (m, 5H), 2.04 (q, J = 7.5, 2H), 1.69 br. s, 3H), 1.61 (s, 3H), 1.60 (s, 3H), 1.57 (s, 3H), 1.55-1.46 (m, 1H), 1.38-1.20 (m, 1H), 1.03 (d, J = 6.9, 3H) ppm. **¹³C-NMR** (125 MHz, C₆D₆): 138.8, 134.6, 130.89, 130.86, 125.0, 124.7, 119.0, 118.1, 40.4, 40.0, 35.1, 31.7, 26.9, 26.6, 26.2, 25.6, 22.9, 19.6, 17.5, 15.8 ppm.

Optical rotation: $[\alpha]_D^{28.0} = +12.1$ (c 0.018, CH₂Cl₂).

NMR data are in agreement with reported values from the literature.³⁵ The absolute configuration was assigned by comparing the determined optical rotation to the literature.³⁵

Analytical data for TC9 product ((+)-maaliene **14**):



C ^a	¹³ C ^b	¹ H ^b
1	34.7	-
2	41.0	1.37-1.30 (m, 1H) 1.09 (dt, <i>J</i> = 13.4, 4.4 Hz, 1H)
3	24.2	1.65 (qt, <i>J</i> = 13.4, 4.1 Hz, 1H) 1.53-1.47 (m, 1H)
4	38.3	2.39 (ddt, <i>J</i> = 13.0, 4.3, 2.1 Hz, 1H) 2.00 (dt, <i>J</i> = 12.8, 4.7 Hz, 1H)
5	150.2	-
6	44.5	1.56 (br d, <i>J</i> = 6.1 Hz, 1H)
7	20.5	0.87-0.84 (m, 1H)
8	19.9	0.58-0.53 (m, 1H)
9	15.9	1.80 (ddt, <i>J</i> = 13.0, 4.3, 2.1 Hz, 1H) 1.56-1.50 (m, 1H)
10	38.9	1.19 (ddd, <i>J</i> = 13.5, 8.3, 1.7 Hz, 1H) 0.94-0.86 (m, 1H)
11	17.5	-
12	29.4	1.04 (s, 3H)
13	15.6	0.92 (s, 3H)
14	17.4	0.85 (s, 3H)
15	106.3	4.99-4.97 (m, 2H)

^aCarbon numberings as shown above table; ^bchemical shifts are given in ppm, coupling constants are given in Hertz, multiplicity: m = multiplet, s = singlet, d = doublet, t = triplet, q = quartet, br = broad.

NMR: See Table. **Optical rotation:** $[\alpha]_D^{28.0} = +9.8$ (c 0.32, hexane).

The absolute configuration was assigned by comparing the determined optical rotation to the literature.³⁶

Supplemental Tables

Table 4-S 1: Sequencing statistics

	Total Sequences (M)	Total Bases (Gb)	Sequences flagged as poor quality	Sequence length	%GC	Coverage (X)	Read Length N50 (kb)	%Reads Q>15
Short Read Genomic Sequencing	450	68	0	151	38	276	-	-
ONT Sequencing	7.1	15.8	-	-	-	64.4	4.66	47.2
Filtered ONT Sequencing (Ultra Long)	0.027	1.13	-	-	-	4.6	40.4	52.6
PacBio Sequencing	2.7	21.6	-	-	-	88	9.4	100
Short Read Transcriptomic Sequencing	1,100	112.8	0	101	41	460	-	-

Table 4-S 2: Data availability sources

Data Type	Platform	BioProject	BioSample	Accession	Repository / DOI
Illumina RNA reads	NCBI SRA	PRJNA1338264	SAMN52403426	pending	-
ONT reads	NCBI SRA	PRJNA1338264	SAMN52403426	pending	-
PacBio reads	NCBI SRA	PRJNA1338264	SAMN52403426	pending	-
Genome assembly	Figshare	-	-	-	10.6084/m9.figshare.30296743
Gene model	Figshare	-	-	-	10.6084/m9.figshare.30296743

Table 4-S 3: Assembly and gene model statistics

	N50 (Mbp)	GC (%)	Total length (Mbp)	# Contigs	TC genes	BUSCO (%Complete/ %Duplicated)	# of Genes Predicted (k)	Total % Repeats	Protein BUSCO (%Complete)	HiSat RNA Read Alignment
Collapsed	5.4	37.49	245	282	15	89.3/ 4.6	-	-	-	-
HapA	3.3	37.44	236	379	14	89.6/ 2.9	23.4	30	94.6	92.88
HapB	2.5	37.43	226	276	11	88.1/ 2.7	23	40	92.3	91.82

Table 4-S 4: NTlink (v 1.3.11)³⁷ Scaffolding for all HapA (using ONT reads with size 30 kbp and longer.

LINK	Contig 1	Gap size	Contig 2	Gap size	Contig 3	Gap size	Contig 4
ntLink_0	h1tg000145l-	21N	h1tg000112l+	21N	h1tg000089l-	21N	h1tg000303l+
ntLink_1	h1tg000078l+	21N	h1tg000310l+	21N	h1tg000054l-	10025N	h1tg000098l-
ntLink_2	h1tg000057l-	21N	h1tg000111l-	21N	h1tg000084l+	21N	h1tg000127l-
ntLink_3	h1tg000008l-	21N	h1tg000016l+	966N	h1tg000065l+	21N	h1tg000010l-
ntLink_4	h1tg000120l-	21N	h1tg000263l-	2892N	h1tg000219l+		
ntLink_5	h1tg000032l+	21N	h1tg000257l-	21N	h1tg000067l+		
ntLink_6	h1tg000027l+	21N	h1tg000206l+	3270N	h1tg000134l-		
ntLink_7	h1tg000025l+	21N	h1tg000072l+	21N	h1tg000051l-		
ntLink_8	h1tg000018l-	21N	h1tg000122l-	21N	h1tg000030l+		
ntLink_9	h1tg000013l-	21N	h1tg000191l+	12058N	h1tg000058l-		
ntLink_10	h1tg000009l-	21N	h1tg000176l+	21N	h1tg000320l-		
ntLink_11	h1tg000275l-	21N	h1tg000289l-				
ntLink_12	h1tg000245l+	21N	h1tg000352l-				
ntLink_13	h1tg000164c-	21N	h1tg000281l-				
ntLink_14	h1tg000159l-	21N	h1tg000286l-				
ntLink_15	h1tg000139l-	21N	h1tg000168l+				
ntLink_16	h1tg000108l-	21N	h1tg000110l+				
ntLink_17	h1tg000091l-	21N	h1tg000094l-				
ntLink_18	h1tg000079l+	21N	h1tg000152l+				
ntLink_19	h1tg000077l+	1025N	h1tg000279l-				
ntLink_20	h1tg000048l+	2074N	h1tg000135l-				
ntLink_21	h1tg000042l+	21N	h1tg000085l+				
ntLink_22	h1tg000036l-	21N	h1tg000231l-				
ntLink_23	h1tg000035l-	21N	h1tg000055l-				
ntLink_24	h1tg000026l-	21N	h1tg000031l-				
ntLink_25	h1tg000023l-	2067N	h1tg000090l-				
ntLink_26	h1tg000017l+	21N	h1tg000095l-				
ntLink_27	h1tg000015l+	21N	h1tg000103l+				
ntLink_28	h1tg000012l-	21N	h1tg000118l-				
ntLink_29	h1tg000006l+	21N	h1tg000056l-				

Table 4-S 5: PCR conditions

Repetitions	Temperature	Time
1x	98°C	0:30 min
3x	98°C	0:10 min
	56-70°C	0:30 min
	72°C	0:50 min
1x	72°C	5:00 min

Table 4-S 6: Primers used

Gene	Fwd Primer	Rev Primer
7b_S91I	TACTTGTGCTTAACCATTTTTATGAACGACGA	TCGTCGTTTCATAAAAAATGGTTAAGCACAAAGTA
7b_R141K	AGCAGTCATTAAAAGTGAAAAAGGTTAATGAAAAGATC	GATCTTTTCATTAACCTTTTTCACCTTTTAATGACTGCT
7b_N174D	AGGAAATTTCTCGCAAGAAGATGTGAGGCATTACTGGAG	CTCCAGTAATGCCTCACATCTTCTTGCGAGAAATTCCT
7b_K282E	TACTGTCAGAGATAAACAGGAAAACAAGGAATGTTCTG	CAGAACATTCCTTGTTTTCCTGTTTATCTCTGACAGTA
7b_R294Q	ATGATTTTGAAGCTTTTCAAAAATCTCTTGAATTACC	GGTAATTCGAAGAGATTTTGAAGGCTTCGAAATCAT
7b_S91A	TACTTGTGCTTAACCGCTTTTATGAACGACGA	TCGTCGTTTCATAAAAGCGGTTAAGCACAAAGTA
7b_S91V	TACTTGTGCTTAACCGTTTTTATGAACGACGA	TCGTCGTTTCATAAAACGGTTAAGCACAAAGTA
7b_S91G	TACTTGTGCTTAACCNKTTTATGAACGACGA	TCGTCGTTTCATAAMNNGGTTAAGCACAAAGTA
7a_I91S	TACTTGTGCTTAACCATTTTTATGAACGACGA	TACTTGTGCTTAACCAAGTTTTATGAACGACGA
7a_I91V	TACTTGTGCTTAACCGTTTTTATGAACGACGA	TCGTCGTTTCATAAAACGGTTAAGCACAAAGTA

Table 4-S 7: TC activities

	Transcribed	Compound Produced	Compound in <i>L. chilensis</i> Crude
TC1a	N	8	N
TC2a	Y	9	Y
TC2b	N	15	N
TC3a	Y	7	Y
TC4a	Y	10	Y
TC4b	N	NA	N
TC5a	N	11	N
TC5b	N	11	N
TC5c	Y	16	Y
TC6	N	NA	N
TC7a	Y	12	Y
TC7b	Y	17	Y
TC8a	Y	13	Y
TC8b	Y	nt	NA
TC9	Y	14	Y

Table 4-S 8: TPMs for TCs

TPM	Hap	Observation	TC
0	hap1	not_found	TC1
0	hap2	not_found	TC1
0.701464	hap1	in_crude	TC2a
0.160385	hap2	in_crude	TC2a
0	hap1	not_found	TC2b
0	hap2	not_found	TC2b
0.24868	hap1	in_crude	TC3
0.166739	hap2	in_crude	TC3
4.467161	hap1	in_crude	TC4a
3.138854	hap2	in_crude	TC4a
0	hap1	not_found	TC4b
0	hap2	not_found	TC4b
0	hap1	not_found	TC5a
0.462848	hap1	in_crude	TC5b
0	hap1	not_found	TC5c
0	hap1	not_found	TC6
2.990755	hap1	in_crude	Tc7a
12.477757	hap2	in_crude	TC7a
34.637796	hap1	in_crude	TC7b
5.426189	hap2	in_crude	TC7b
0	hap1	not_found	TC8a
0	hap2	not_found	TC8a
0	hap1	not_found	TC8b
3.628644	hap2	in_crude	TC9

Table 4-S 9: Mutant TC enzyme activities

TC	Product From FPP	Product From GGPP
7a-WT	12	NA
7b-WT	12	17
7b_S91I	12	NA
7b_R141K	12	17
7b_N174D	12	17
7b_K282E	12	17
7b_R294Q	12	17
7b_S91A	12	17
7b_S91V	12	17
7b_S91G	12	17
7a_I91S	12	17
7a_I91V	12	NA

Table 4-S 10: Complete TE stats

	HapA	HapB	Average	TC Region	WILCOXON-TEST p -value
LINE	2.12%	1.81%	1.97%	2.40%	0.05
LTR	10.98%	7.32%	9.15%	11.20%	0.001
TIR (DNA)	7.99%	9.59%	8.79%	14.50%	2x10 ⁻⁵
Helitron	4.45%	5.58%	5.02%	5.59%	6x10 ⁻³
CACTA	0.93%	1.42%	1.18%	1.93%	1x10 ⁻³
hAT	2.82%	3.07%	2.95%	4.51%	3x10 ⁻³
PIF_Harbinger	1.16%	1.19%	1.18%	2.61%	5x10 ⁻⁴
Mutator	2.93%	3.82%	3.38%	7.97%	6x10 ⁻⁵

Table 4-S 11: Bioinformatics tools used and versions

Tool	Version
fastp	v 0.19.4 & v 23.2
Fastqc	v 0.11.9
Jellyfish	v 2.2.10
GenomeScope	v 2.0.1
NanoPlot	v 1.41.0
Filtlong	v 0.2.1
Hifiasm	v 0.16.1
BUSCO	v 5.4.7
QUAST	v 5.2.0
RepeatModeler	v 2.0.7
RepeatMasker	v 4.1.2
TSEBRA	Braker v 2.1.6
hisat2	v 2.2.1
egg-nog-mapper	v 2.0.1
orthofinder	v 2.5.4
jcvi-tools	v 1.5.6
EDTA	v 1.9.697
trinity	v 2.8.5
Emboss	v 6.6.0.0
HMMER	v 3.3.2
kalign2	v 2.04
pymol	v 3.1.0
swiss-model	NA
bedtools	v 2.26.0
samtools	v 1.3.1
ggplot2	v 3.5.2
minimap2	v 2.17
salmon	v 0.14.2
lqtree	v 2.1.4
rstudio	v 2023.12.1+402
msa	v 1.34.0
corrplot	v 0.95
ntLink	v 1.3.11

Supplemental Figures

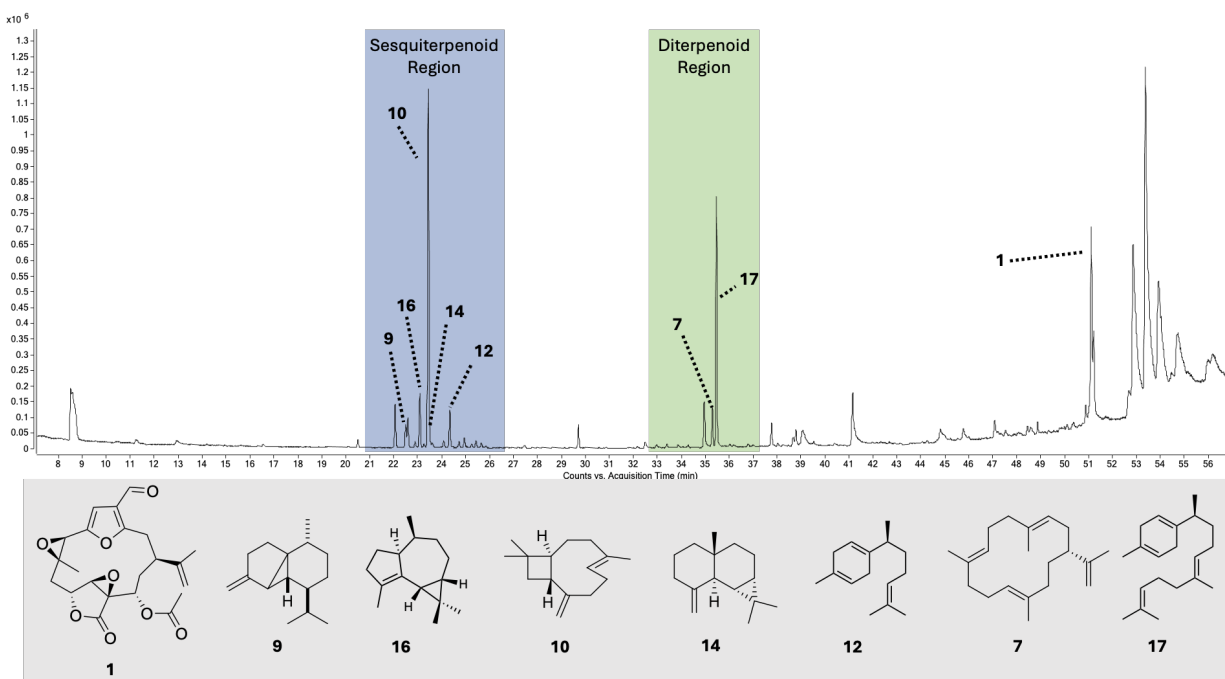


Figure 4-S 1: Annotated chromatogram of crude extract

GC-MS chromatogram of *L. chilensis* tissue extracted with CH_2Cl_2 :MeOH (2:1). All terpene and terpenoid peaks as validated either by database hits, ordered chemical standards, in house chemical standards, or scaled up enzyme assays and structural elucidation are annotated and structures provided in panel below. Sesquiterpene and diterpene regions, based on common retention times for the respective masses of these classes, are highlighted.

p = 2		
k = 21		
property	min	max
Homozygous (aa)	99.12%	99.15%
Heterozygous (ab)	0.85%	0.88%
Genome Haploid Length	194,217,992 bp	194,411,649 bp
Genome Repeat Length	66,547,052 bp	66,613,407 bp
Genome Unique Length	127,670,940 bp	127,798,242 bp
Model Fit	70.12%	91.11%
Read Error Rate	0.35%	0.35%

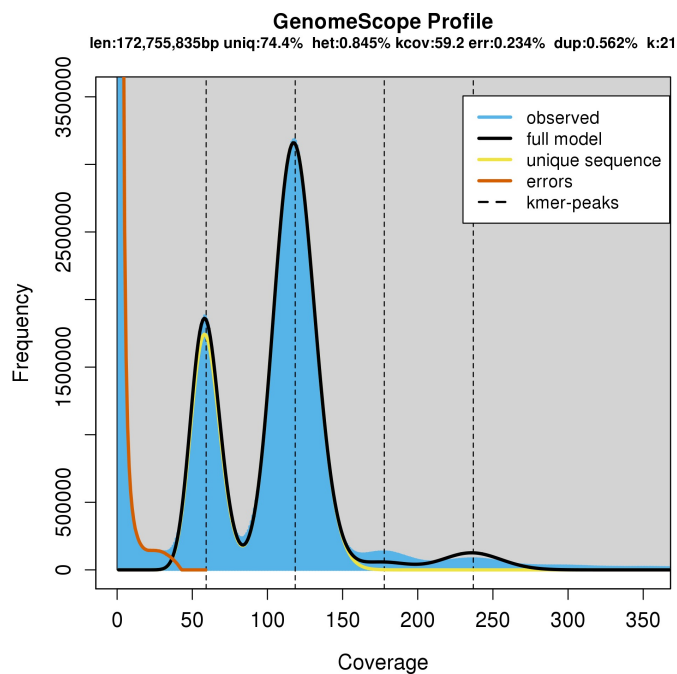


Figure 4-S 2: GenomeScope visualization of heterogeneity

Jellyfish (v 2.2.10) was used to calculate K-mer (21-mers) count distribution. The resulting heterogeneity was 0.845%. The peak at 50 coverage provides evidence of the genomes haploid state.

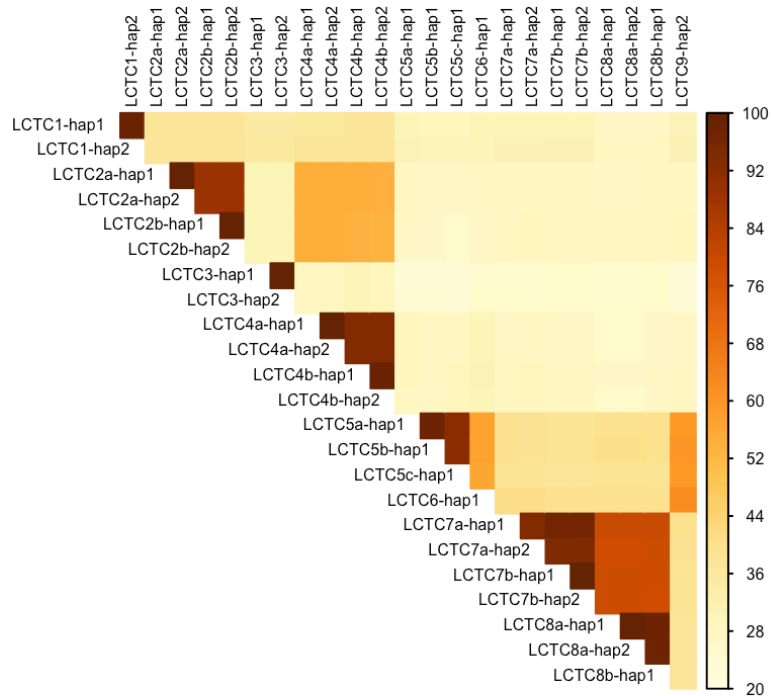


Figure 4-S 3: Heat map of sequence similarity across all *L. chilensis* TCs

Sequence similarity matrix of all complete TCs from both *L. chilensis* haplotypes. Deep red color indicates a high degree of similarity, up to 100%, while pale yellow denotes a lower degree of similarity, as low as 20%.

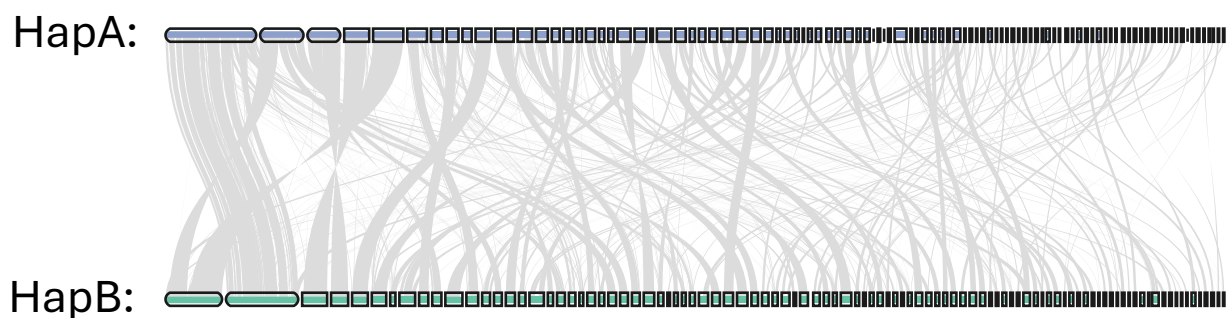


Figure 4-S 4: Full synteny plot of HapA and HapB

All contigs present in each phased assembly are depicted. Synteny lines in grey highlight genomic regions conserved across the two haplotypes.

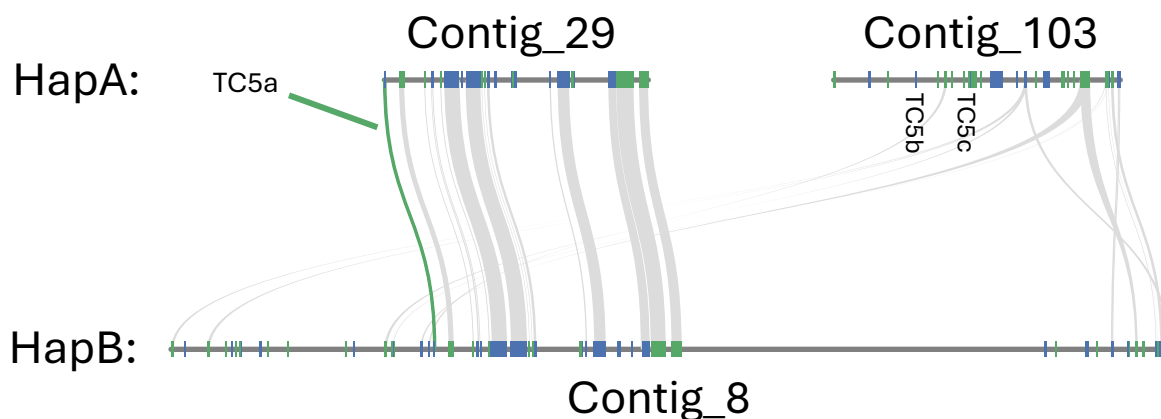


Figure 4-S 5: Semi-scaffolding of TC5 contigs

Micro-synteny of TC5 containing contigs. HapA contig 29 contains TC5a, HapA contig 103 contains TC5b and TC5c. Syntenic analysis have found homologous genes on HapA contig 29 and HapB 8 and HapA contig 103 and HapB 8. While the analysis was able to find a homolog to HapA TC5a, the degraded copy of HapB TC5a, no homologs were found for TCs 5b or 5c.

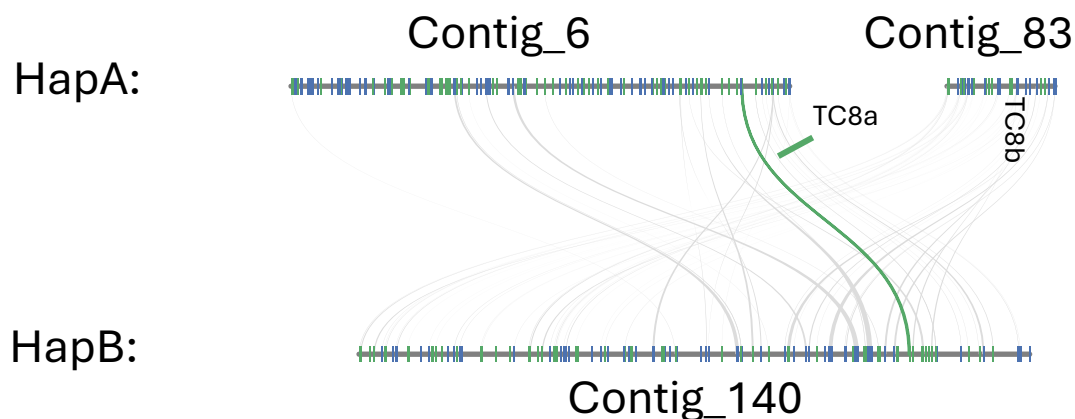


Figure 4-S 6: Semi-scaffolding of TC8 contigs

Micro-synteny of TC58 containing contigs. HapA contig 6 contains TC8a, HapA contig 83 contains TC8b. Syntenic analysis have found homologous genes on HapA contig 6 and HapB 140 and HapA contig 83 and HapB 140. While the analysis was able to find a homolog to HapA TC8a, the HapB TC8a, no homologs were found for HapA TC8b.

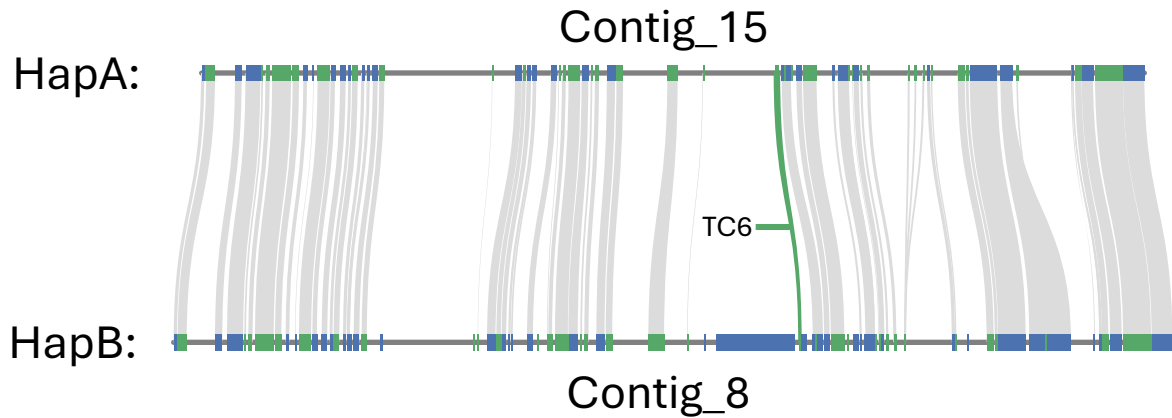


Figure 4-S 7: Semi-scaffolding of TC6 contigs

Micro-synteny of TC6 containing contig. HapA contig 15 contains TC6, HapB contig 8 contains pseudogenized TC6. Notably, syntenic analysis was able to find homologous genes surrounding TC6, and the pseudogenized copy on HapB.

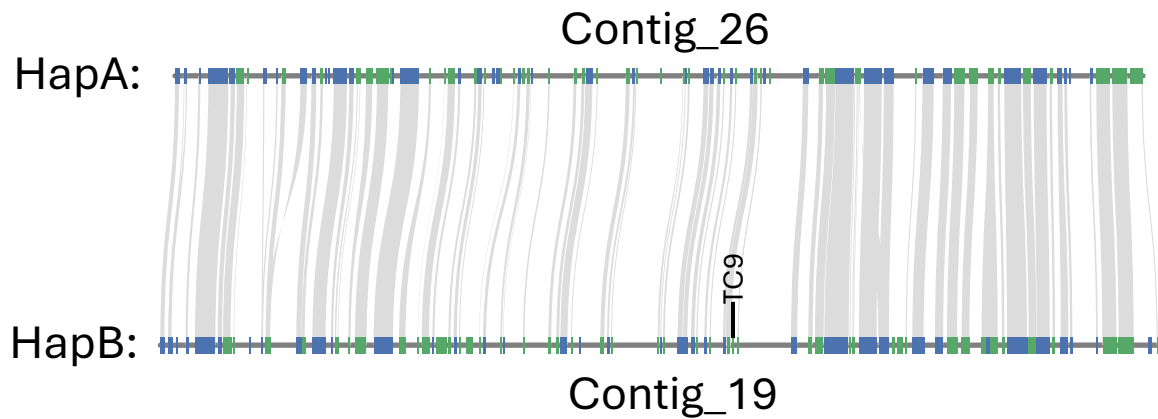


Figure 4-S 8: Semi-scaffolding of TC9 contigs

Micro-synteny of TC9 containing contig. HapA contig 26 contains pseudogenized TC9, HapB contig 19 contains TC9. While syntenic analysis were able to find homologous genes surrounding TC9, no HapA TC9 homologs was found.

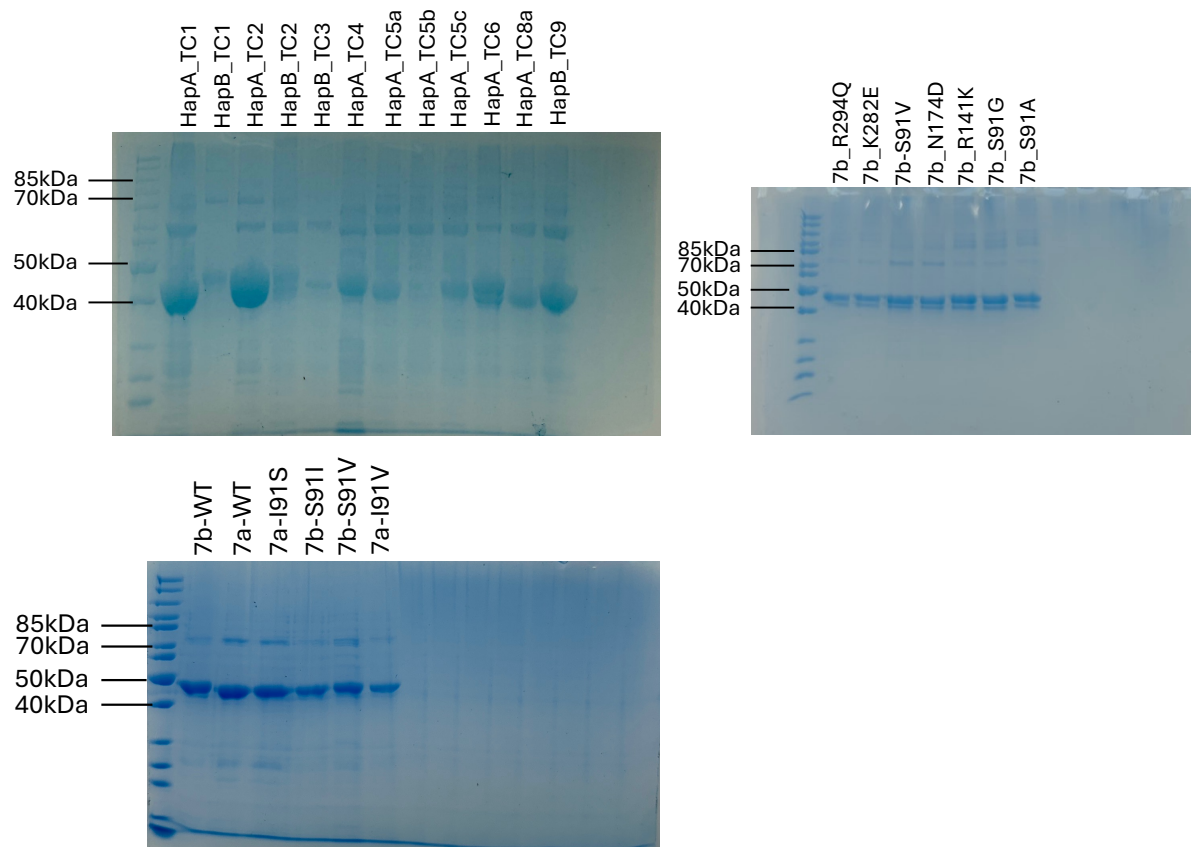


Figure 4-S 9: SDS gel of all proteins

SDS gel of all enzymes tested in this study. Associated ladder is depicted. Both 7b-S91V lanes shown in gels 2 and 3 correspond to the same purified enzyme preparation.

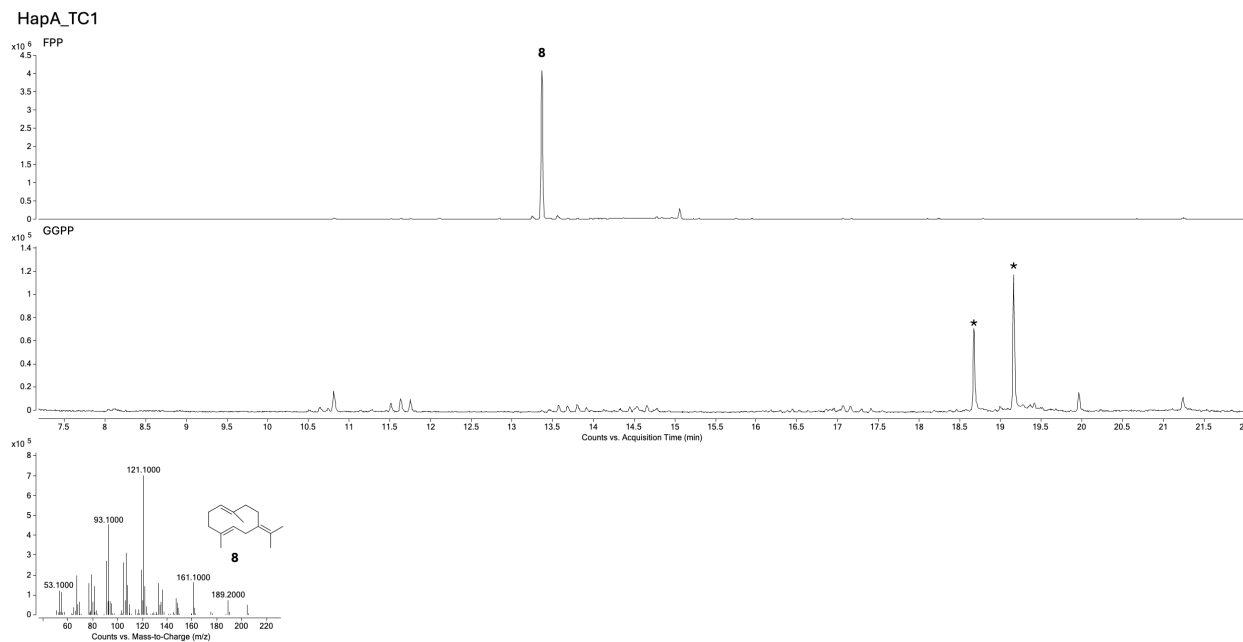


Figure 4-S 10: GCMS total ion chromatograms of HapA_TC1

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The identity of **8** has been verified using a commercially available authentic standard.

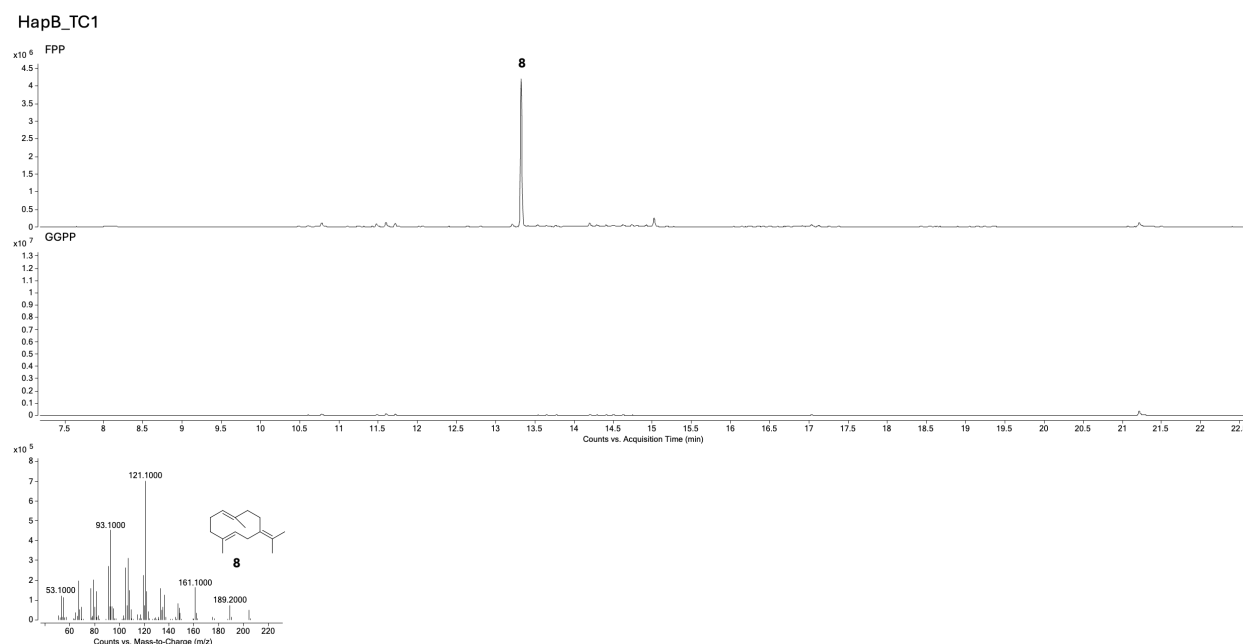


Figure 4-S 11: GCMS total ion chromatograms of HapB_TC1

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The identity of **8** has been verified using a commercially available authentic standard.

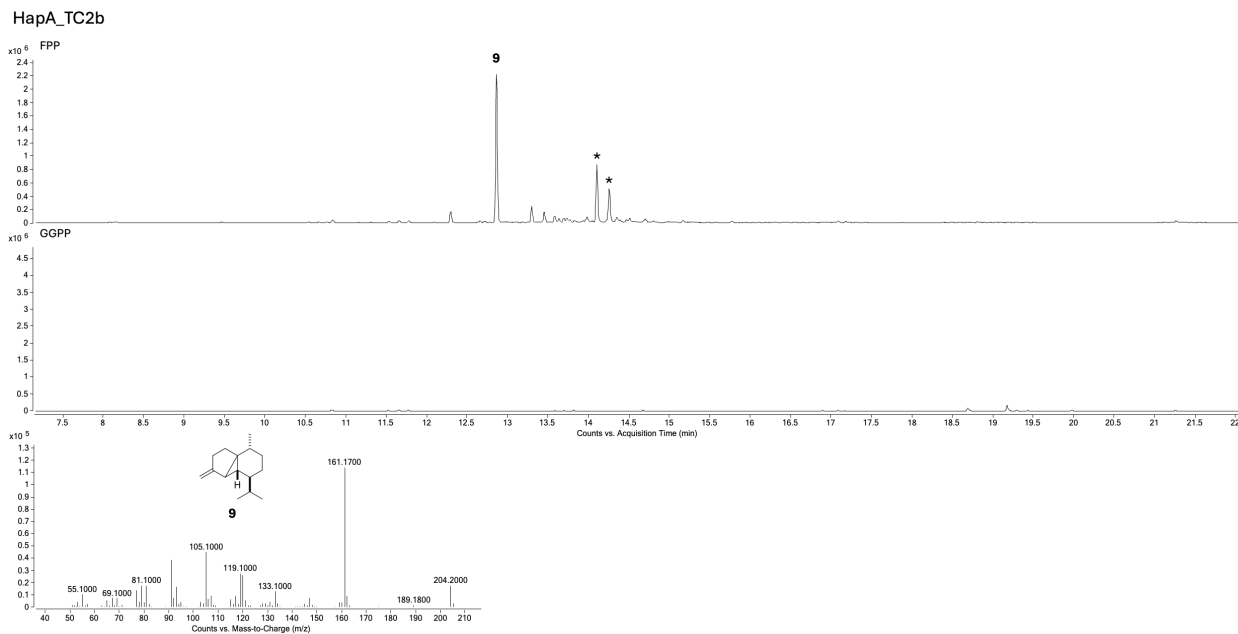


Figure 4-S 12: GCMS total ion chromatograms of HapA_TC2a

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The structure of **9** was verified by NMR structure elucidation of purified, derivatized compound from a large-scale in-vitro assay.

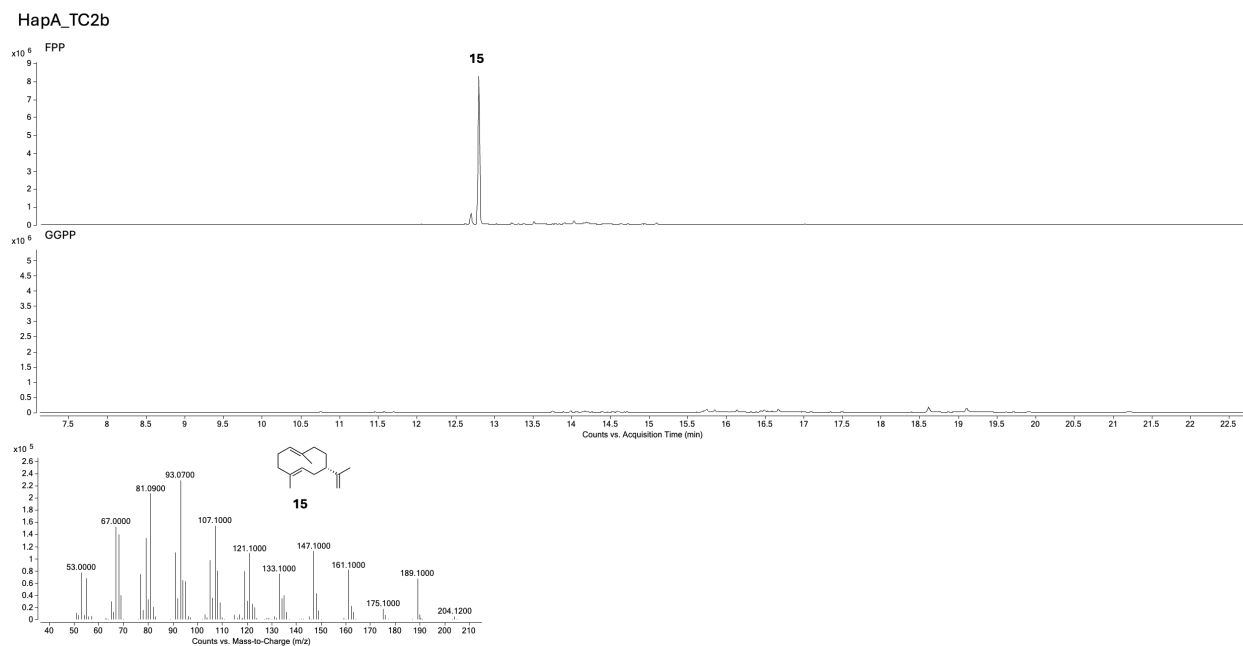


Figure 4-S 13: GCMS total ion chromatograms of HapA_TC2b

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The identity of **15** has been verified using a commercially available authentic standard.

HapB_TC3

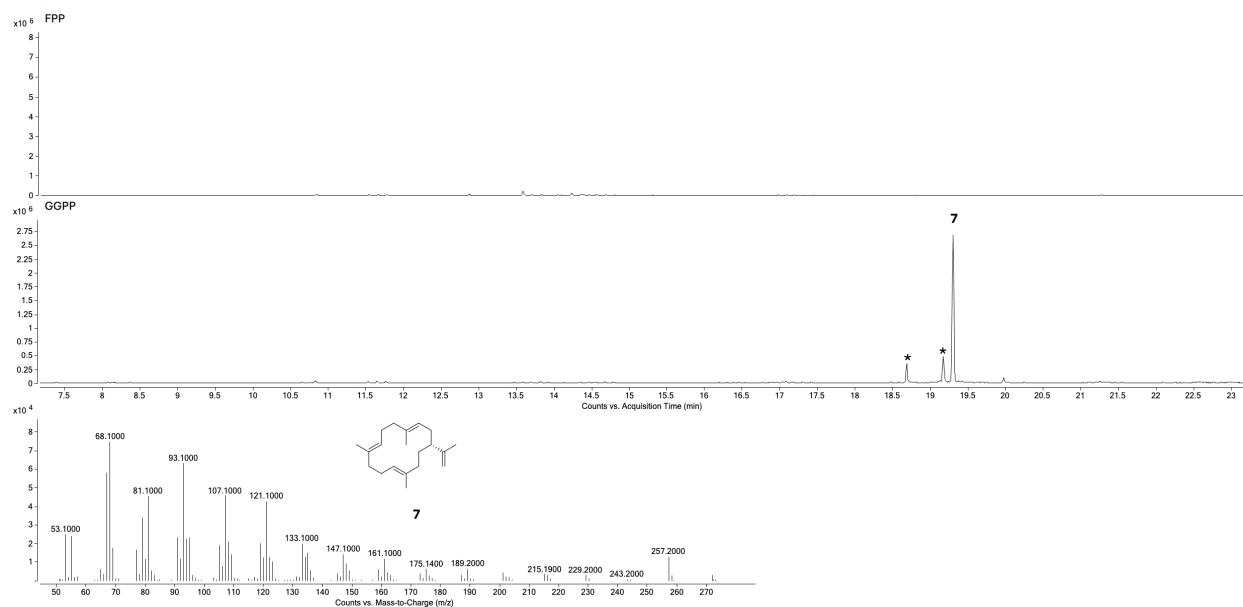


Figure 4-S 14: GCMS total ion chromatograms of HapB_TC3

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The identity of compound **7** was confirmed by comparison to the *ErTC-2* enzymatic product, which we previously isolated and structurally characterized.³

HapA_TC4

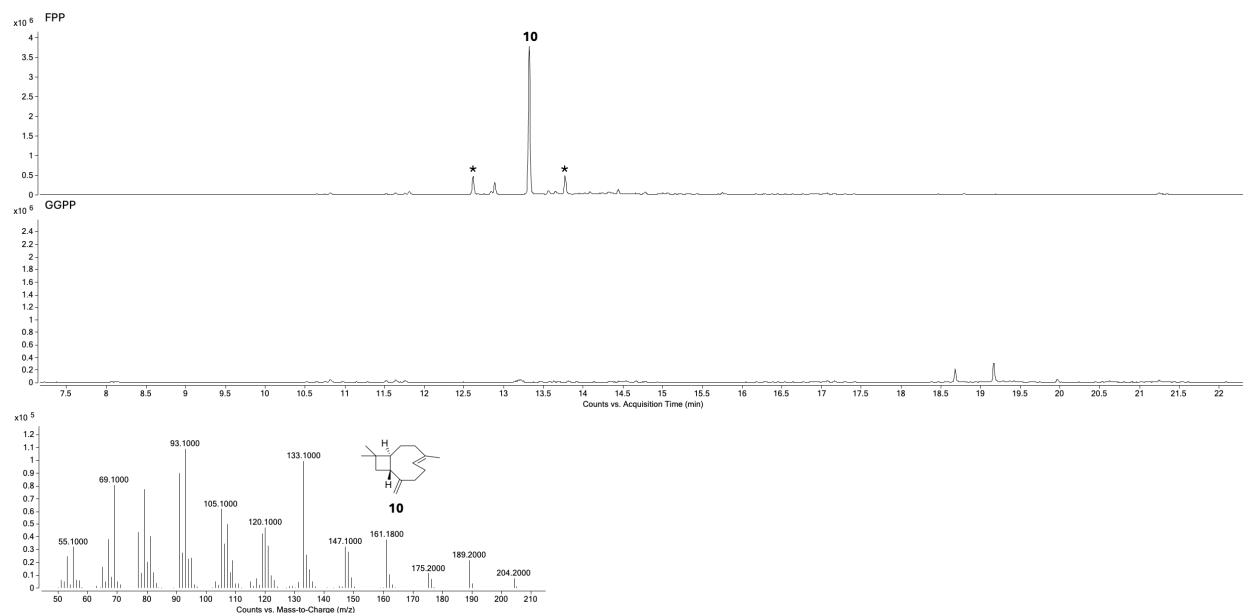


Figure 4-S 15: GCMS total ion chromatograms of HapA_TC4

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The identity of **10** has been verified using a commercially available authentic standard.

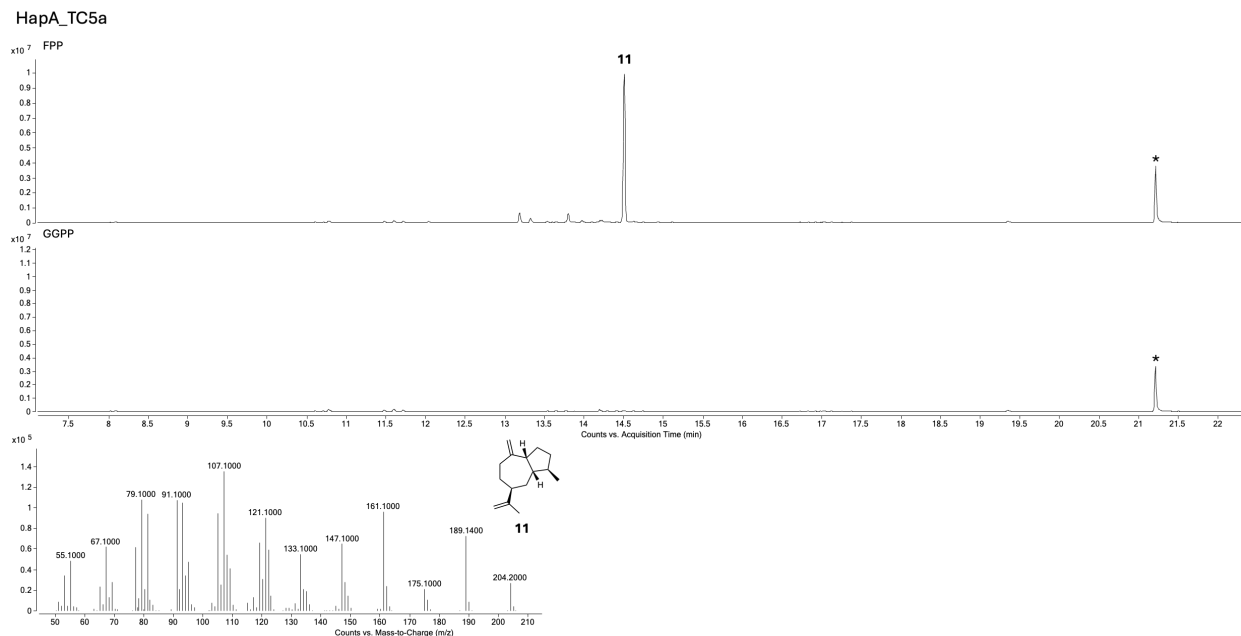


Figure 4-S 16: GCMS total ion chromatograms of HapA_TC5a

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The structure of **11** was verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

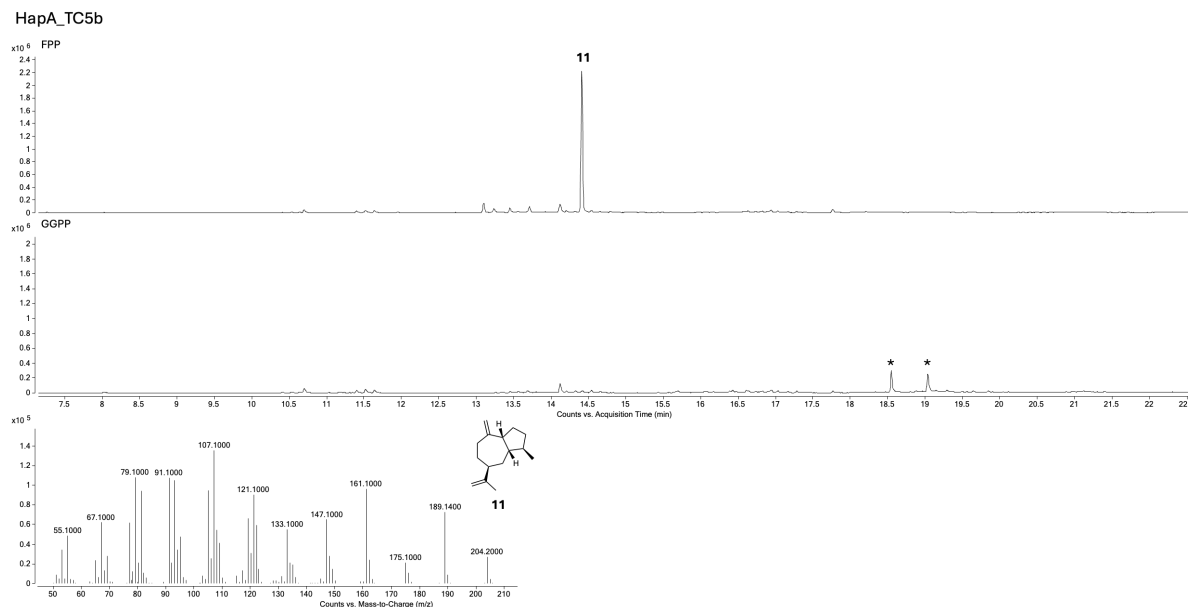


Figure 4-S 17: GCMS total ion chromatograms of HapA_TC5b

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The structure of **11** was verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

HapA_TC5c

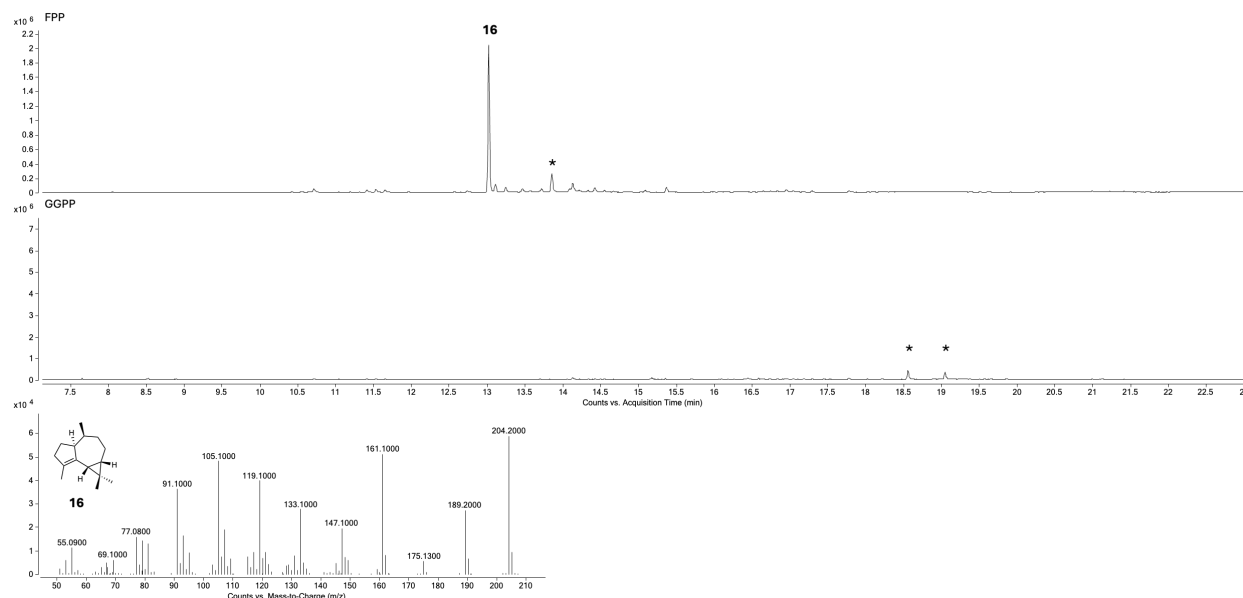


Figure 4-S 18: GCMS total ion chromatograms of HapA_TC5c

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. Signals marked with asterisks are non-enzymatic degradation products of substrates. The structure of **16** was verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

HapA_TC7a

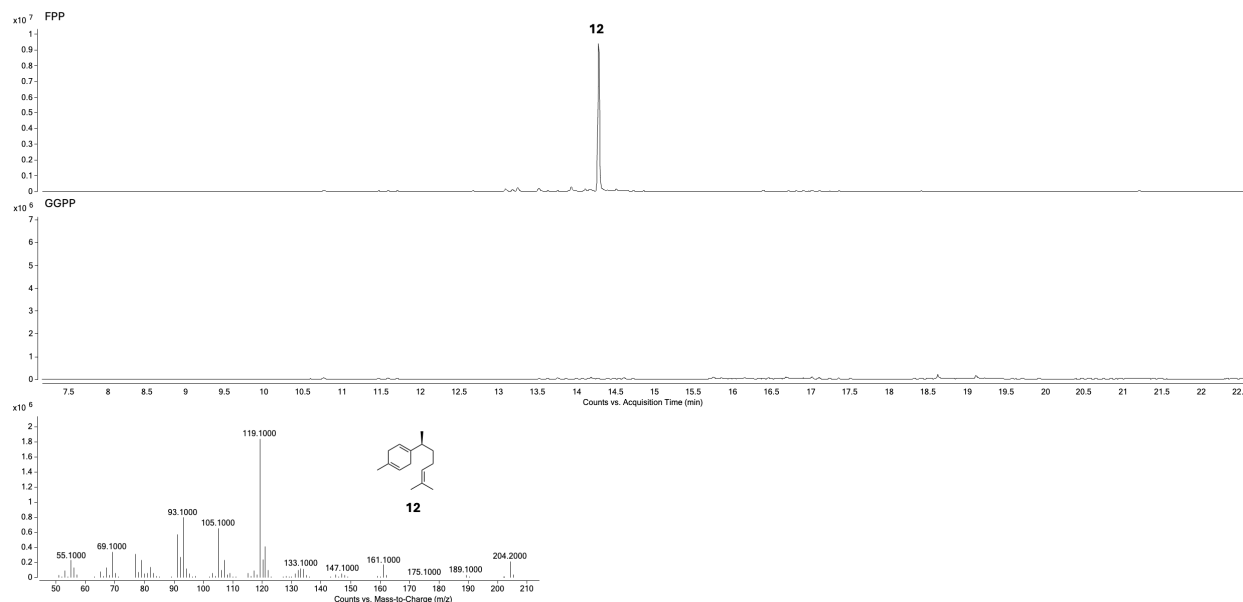


Figure 4-S 19: GCMS total ion chromatograms of HapA_TC7a

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. The structure of **12** was verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

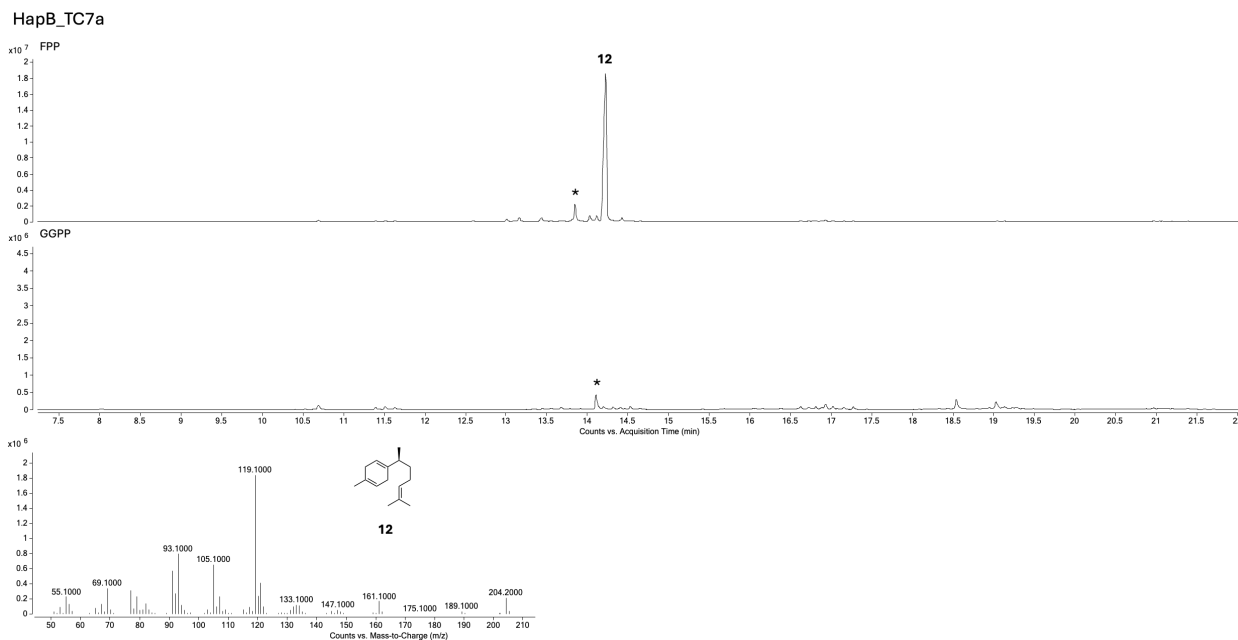


Figure 4-S 20: GCMS total ion chromatograms of HapB_TC7a

In vitro incubations with FPP (top) and GGPP (bottom). Signals marked with asterisks are non-enzymatic degradation products of substrates. Mass spectrum of enzymatically produced terpene is shown. The structure of **12** was verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

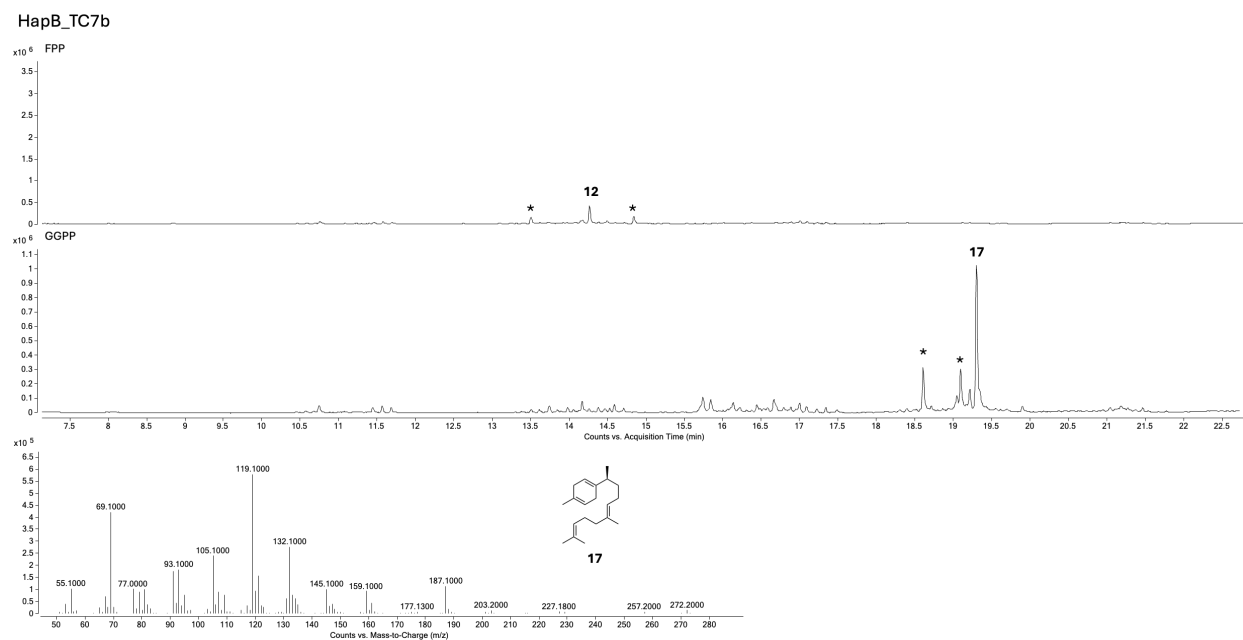


Figure 4-S 21: GCMS total ion chromatograms of HapB_TC7b

In vitro incubations with FPP (top) and GGPP (bottom). Signals marked with asterisks are non-enzymatic degradation products of substrates. Mass spectrum of enzymatically produced terpene **17** is shown. Mass spectra of **12** matched HapB_TC7a's spectra. The structure of **17** was verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

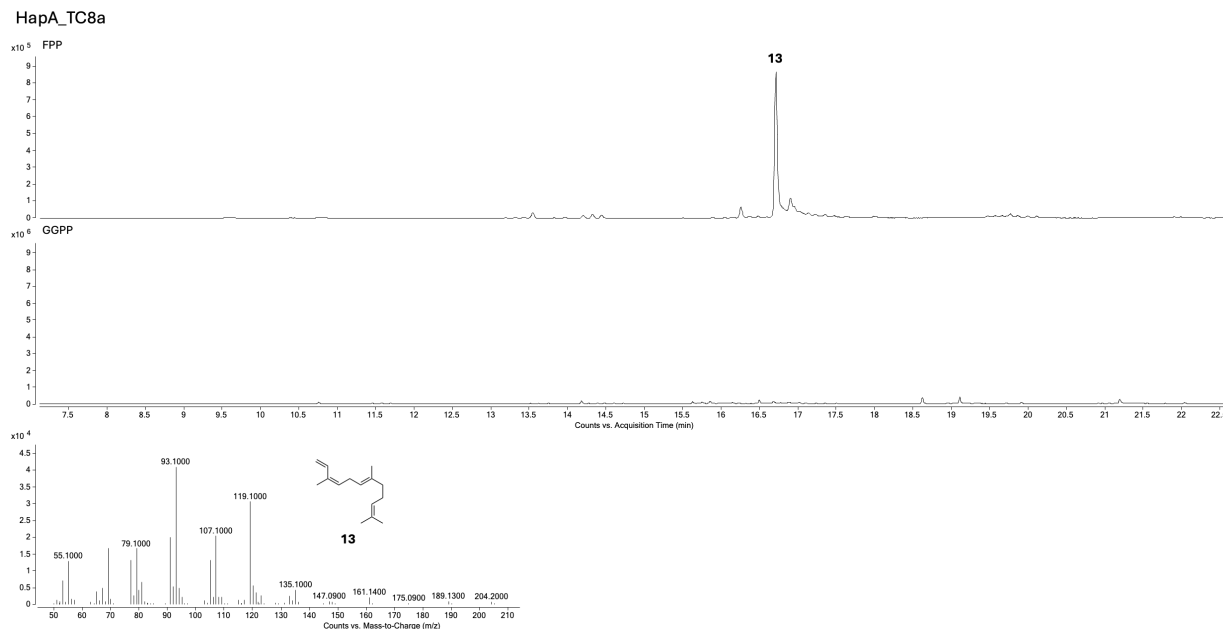


Figure 4-S 22: GCMS total ion chromatograms of HapA_TC8a

In vitro incubations with FPP (top) and GGPP (bottom). Mass spectrum of enzymatically produced terpene is shown. The identity of compound **13** was confirmed by comparison to the LsTPS-1 enzymatic product, which we previously isolated and structurally characterized.³⁸

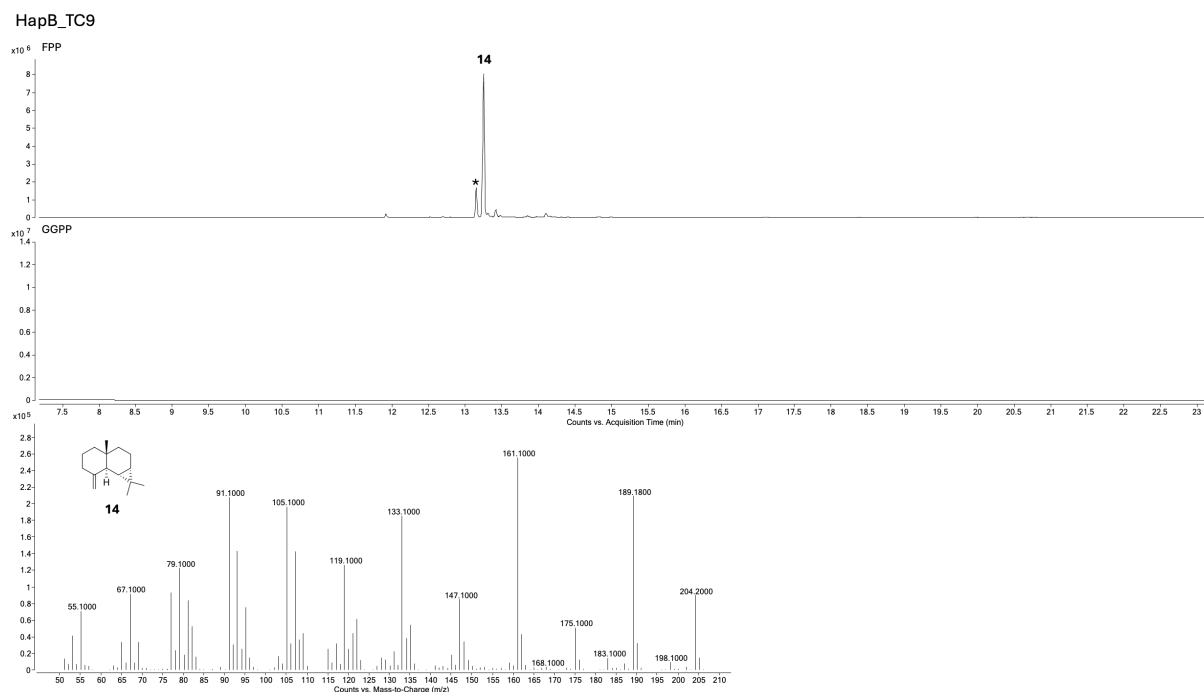


Figure 4-S 23: GCMS total ion chromatograms of HapB_TC9

In vitro incubations with FPP (top) and GGPP (bottom). Signals marked with asterisks are non-enzymatic degradation products of substrates. Mass spectrum of enzymatically produced terpene is shown. The identity of **13** has been verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

```

HapA_7b  MACSKELRVFKESKYHHDLVNEAINPELFSDELINFAIALGLSDNKAAPKKYLYTCRPHYMNRYYVVFIPDNPMKSNIFKRFAYLCLTSFMNDDLMEAFSESEMRQVC 110
HapB_7b  MACSKELRVFKESKYHHDLVNEAINPELFSDELINFAIALGLSDNKAAPKKYLYTCRPHYMNRYYVVFIPDNPMKSNIFKRFAYLCLTSFMNDDLMEAFSESEMRQVC 110
HapA_7a  MACSKELRVFKESKYHHDLVNEAINPELFSDELINFAIALGLSDNKAAPKKYHYTCRPHYMNRYYVVFIPDNPMKSNIFKRFAYLCLTIFMNDLMEAFSESEMRQVC 110
HapB_7a  MAYSKELRVFKESKYHHDLVNEAISPELFSDELINFAIALGLSDNKAAPKPYLNTCRPHYMNRYYVVFIPDNLSKNIFKRFAYLCLTIFMNDLMEAFSESEMRQVC 110

HapA_7b  DAFQKMDVEICEQFPFRPTLAQMKQSLKVKRVNEKIIPNVLNLEFTNSVAKDLIEEGNFSQEDVRHYWRRLLVMVVGFFYVEGAKSESIQGVTTTTDEYLWRRVMSGAIIV 220
HapB_7b  DAFQKMDVEICEQFPFRPTLAQMKQSLKVKRVNEKIIPNVLNLEFTNSVAKDLIEEGNFSQENVRHYWRRLLVMVVGFFYVEGAKSESIQGVTTTTDEYLWRRVMSGAIIV 220
HapA_7a  DAFQKMDVEICEQFPFRPTLAQMKQSLKVKRVNEKVIIPNVLNLEFTNSVAKDLIEEGNFSQEDVRYYWRRLLVMVVGFFYVEGAKSESIQGVTTTTDEYLWRRVMSGAIIL 220
HapB_7a  DAFQKMDGEICEQFPFRPTLAEMKQSLKVKRVNEKIIPNVLNLEFTNSVAKDLIEEGNFSQEDVRHYWRRLLVMVVGFFYVEGAKSESIQGVTTTTDEYLWRRVMSGAIIV 220

HapA_7b  WMYGHEVMSEMLGKLDGTQISFIHELYGLTAIFCLIVNDIYSYEREKIIGKRVMTVRDKQKNKECSDDFEAFRKSLRITNAIHKVMYQKIESAKQESPGDAEFHKLLDN 330
HapB_7b  WMYGHEVMSEMLGKLDGTQISFIHELYGLTAIFCLIVNDIYSYEREKIIGKRVMTVRDKQKNKECSDDFEAFRKSLRITNAIHKVMYQKIESAKQESPGDAEFHKLLDN 330
HapA_7a  WMYGHEVMSEMLGKLDGTQISFIHELYGLTAIFCLIVNDIYSYEREKIIGKRVMTVRDKQKNKECSDDFEAFQKSLRITNAIHKVMYQKIESAKQESPGDAEFHKLLDN 330
HapB_7a  WMYGHELMSEMLGKLEGTQISFIHELYGLTAIFCLIVNDIYSYEREKIIGKRVMTVRDKQKNKECSDDFEAFQKSLRITNAIHKVMYQKIEHAKQESPGNAEFHKLLDN 330

HapA_7b  IGKGTVGTVFFFDKCSRYNETKWRLTLVEADGMELEEWKSEVVEEFDTVQHFLQDASPKAKRMDEAIKRCVLDNMNANVLP. 412
HapB_7b  IGKGTVGTVFFFDKCSRYNETKWRLTLVEADGMELEEWKSEVVEEFDTVQHFLQDASPKAKRMDEAIKRCVLDNMNANVLP. 412
HapA_7a  IGKGTVGTMFFFDKCSRYNETKWRLTLVEVDGMELEEWKSKVVEEFDTVQHFLQDASPKAKRMDEAIKRCVLDNMNANVLP. 412
HapB_7a  IGKGTVGTVFFFDKCSRYNETKWRLTLVEADGMELEEWKSEVVEEFDTVRHFLQDASPKAKRMDEAIKRCVLDNMNANVLP. 412

```

Figure 4-S 24: Multiple sequence alignment of TC7s

Alignment generated by kalign2 (v 2.04)²⁵ of all four TC7s.

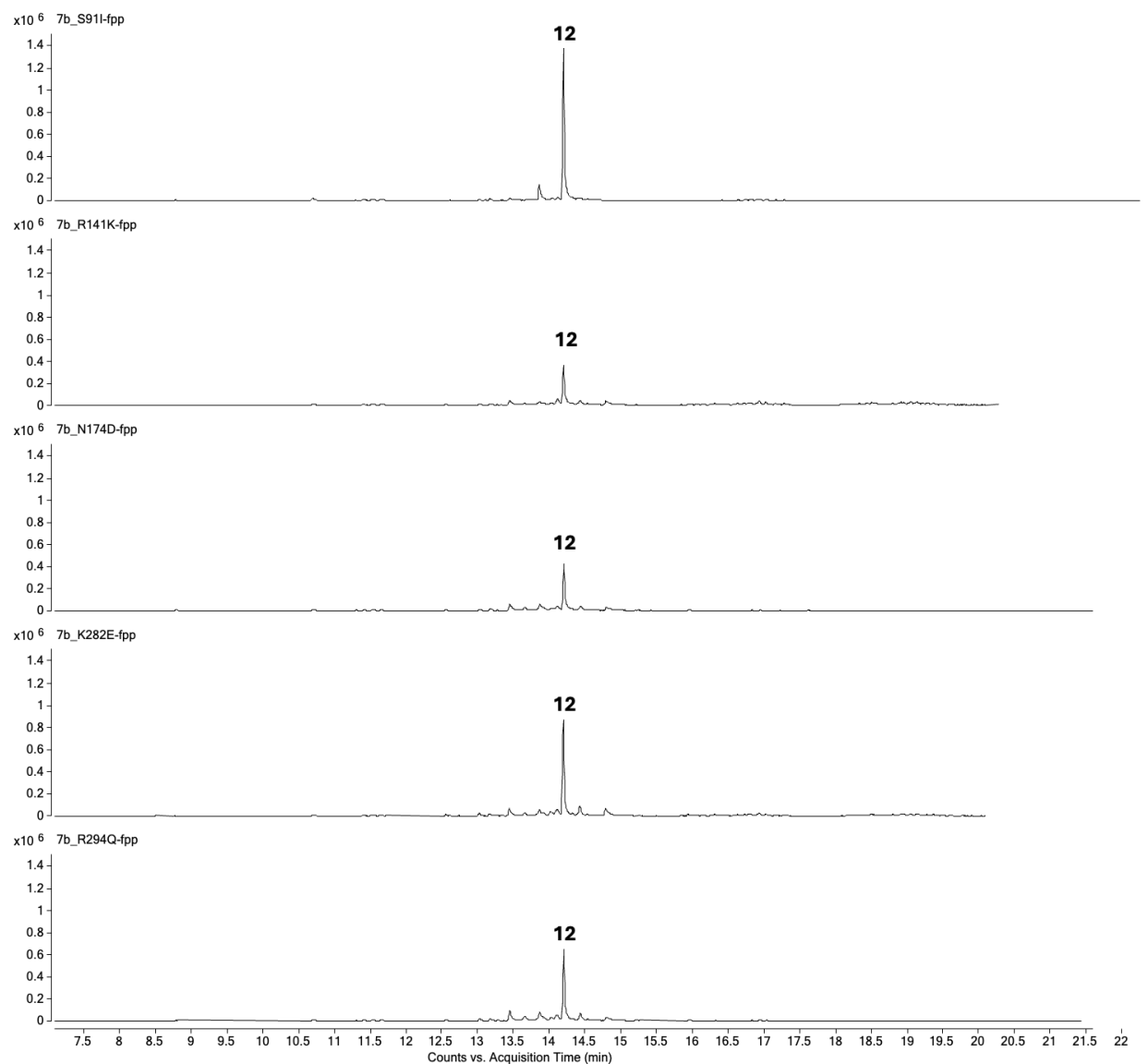


Figure 4-S 25: GCMS total ion chromatograms of TC7a and TC7b site swap mutagenesis with FPP

GCMS total ion chromatograms of TC7a and TC7b site swap mutant in vitro incubations with FPP. The enzyme used in each assay is denoted on the upper left corner of the chromatogram. All product producing assay have respective peaks annotated. The identity of **12** has been verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

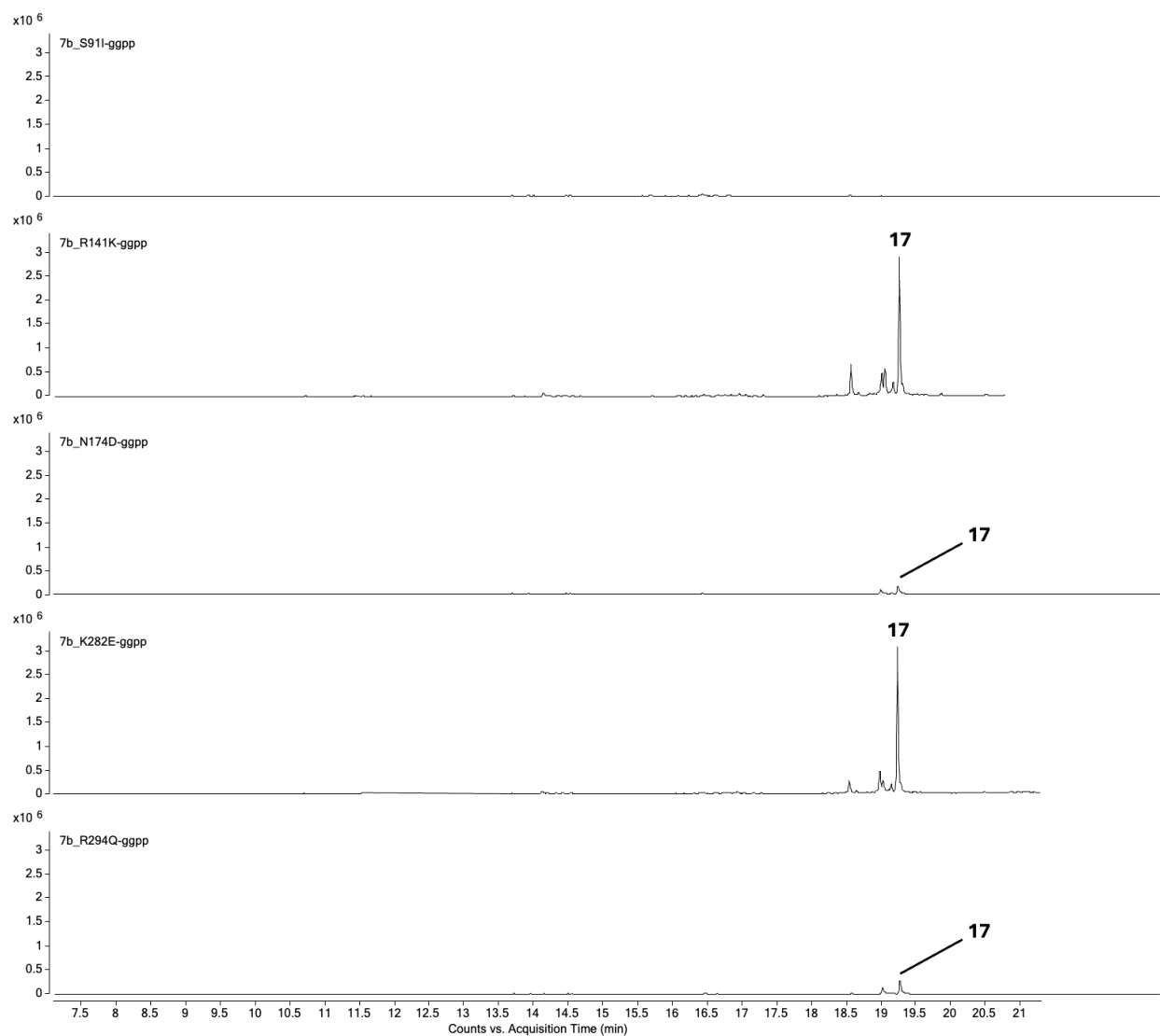


Figure 4-S 26: GCMS total ion chromatograms of TC7a and TC7b site swap mutagenesis with GGPP

GCMS total ion chromatograms of TC7a and TC7b site swap mutant in vitro incubations with GGPP. The enzyme used in each assay is denoted on the upper left corner of the chromatogram. All product producing assay have respective peaks annotated. The identity of **17** has been verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

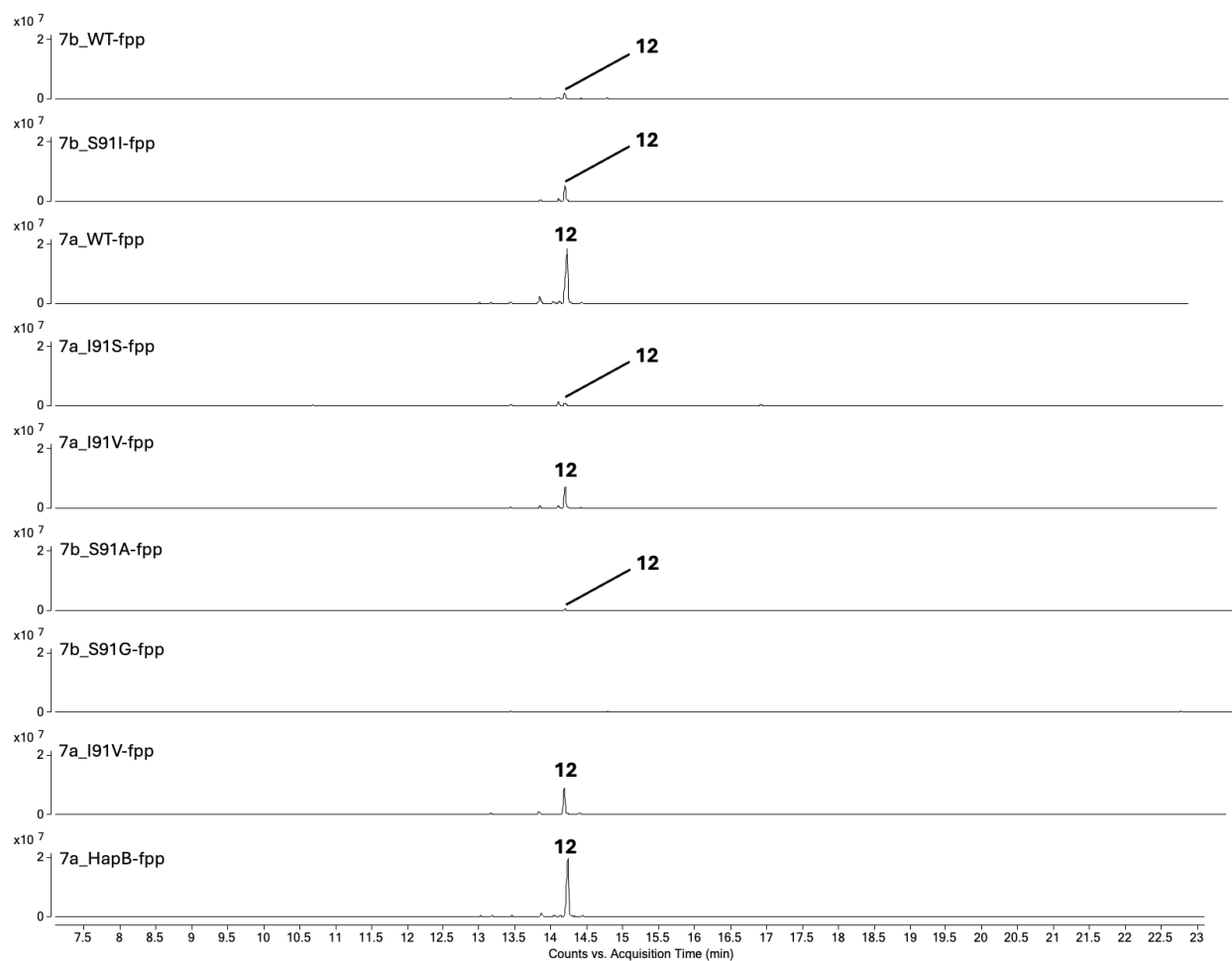


Figure 4-S 27: GCMS total ion chromatograms of TC7a and TC7b residue 91 mutagenesis with FPP

GCMS total ion chromatograms of TC7a and TC7b residue 91 mutant in vitro incubations with FPP. The enzyme used in each assay is denoted on the upper left corner of the chromatogram. All product producing assay have respective peaks annotated. The identity of **12** has been verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

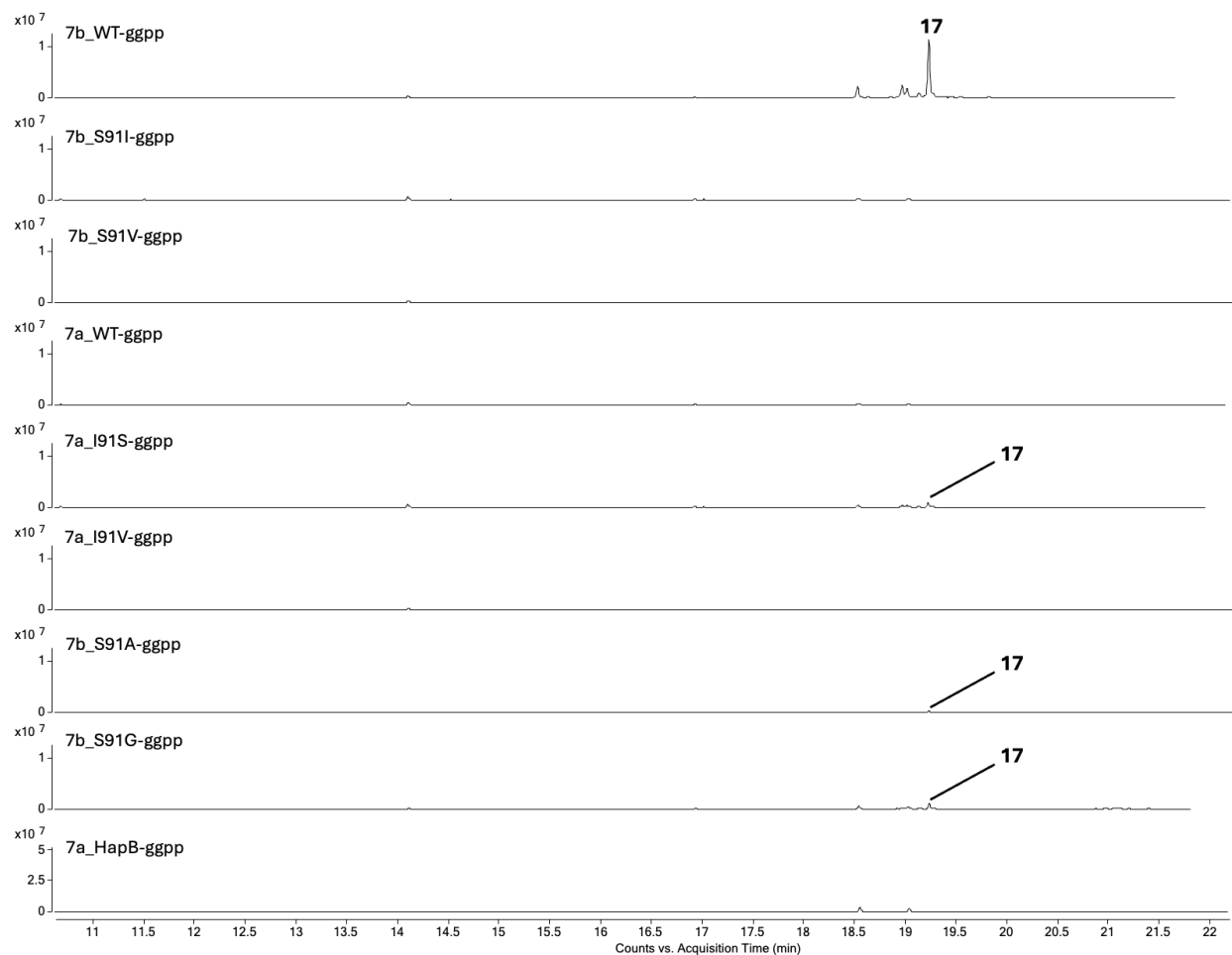


Figure 4-S 28: GCMS total ion chromatograms of TC7a and TC7b residue 91 mutagenesis with GGPP

GCMS total ion chromatograms of TC7a and TC7b residue 91 mutant in vitro incubations with GGPP. The enzyme used in each assay is denoted on the upper left corner of the chromatogram. All product producing assay have respective peaks annotated. The identity of **17** has been verified by NMR structure elucidation of purified compound from a large-scale in-vitro assay.

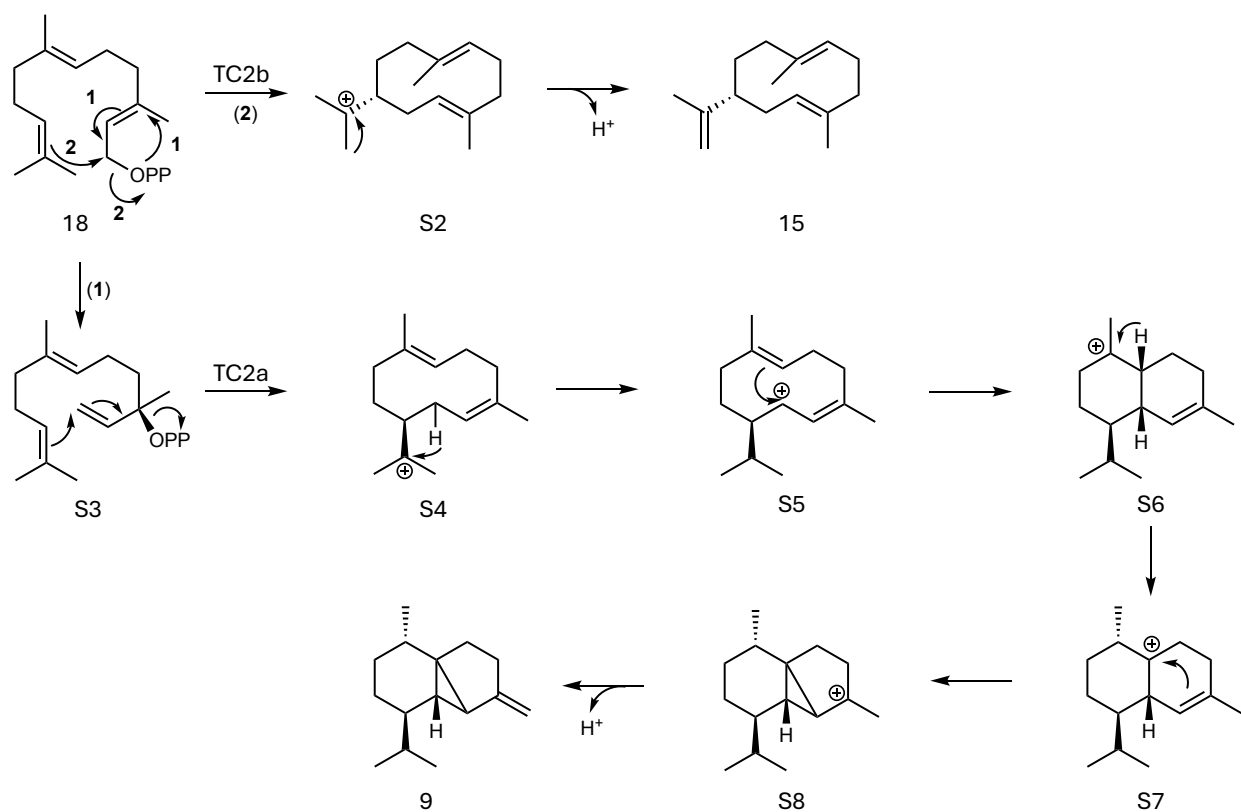
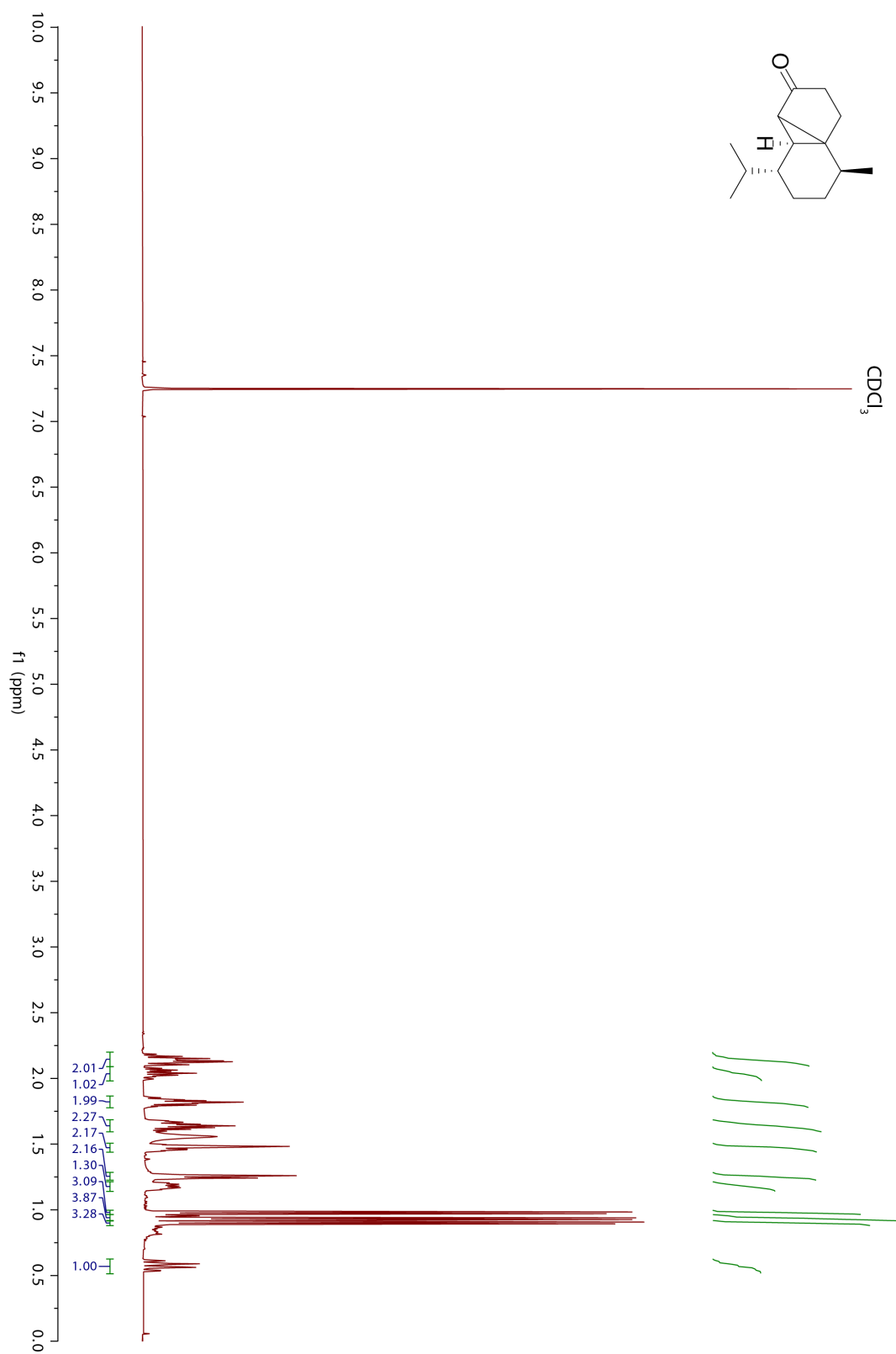


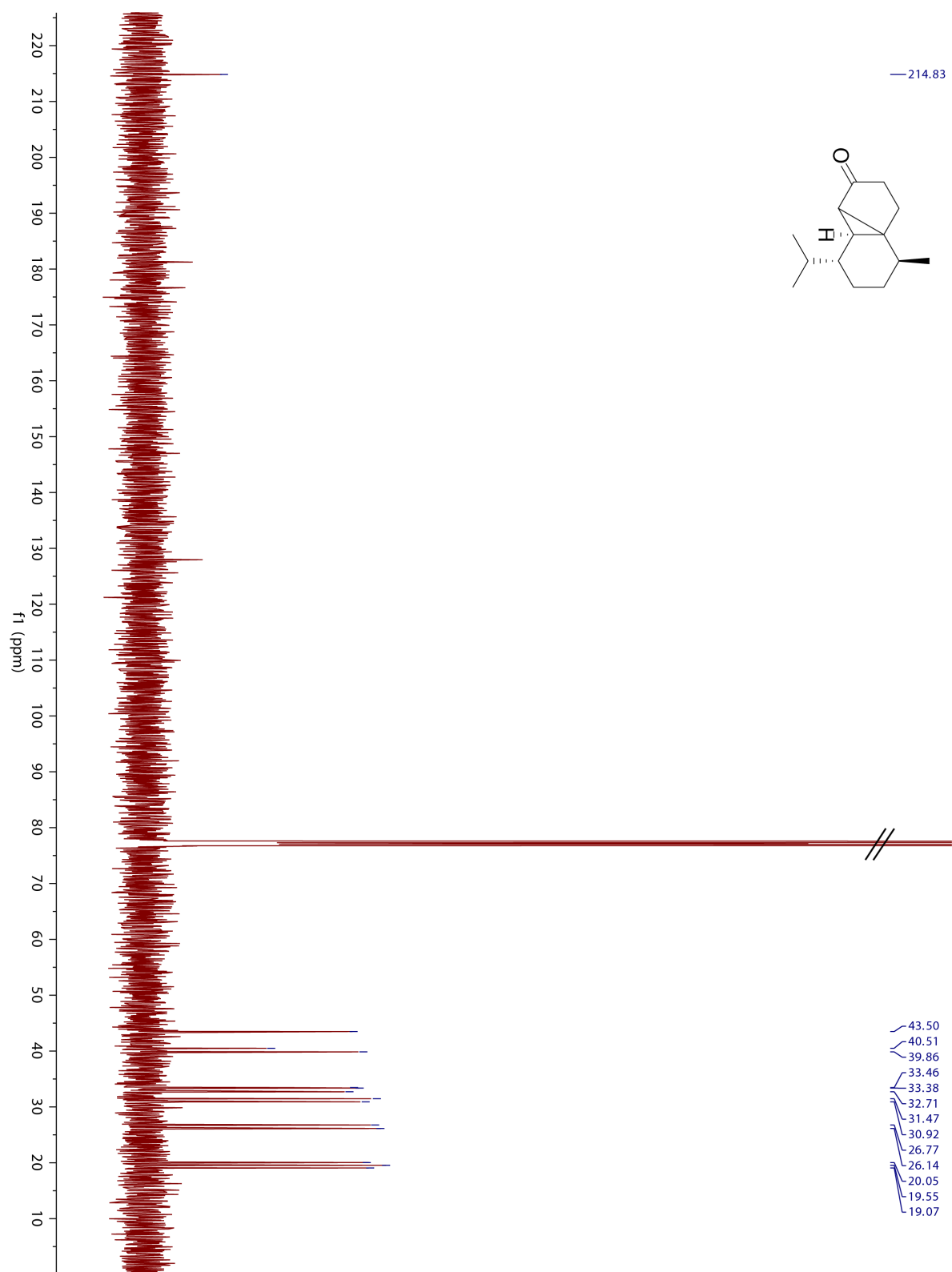
Figure 4-S 29: Proposed mechanisms of TC2a and TC2b reactions

Proposed cyclization mechanisms for paralogs of TC2. TC2a acts on **S3** rather than **18** to generate the carbocation **S4**, and via a series of steps ultimately generated **9**. TC2b instead acts on **18** to form **S2** which is quenched via cyclization into **15**.

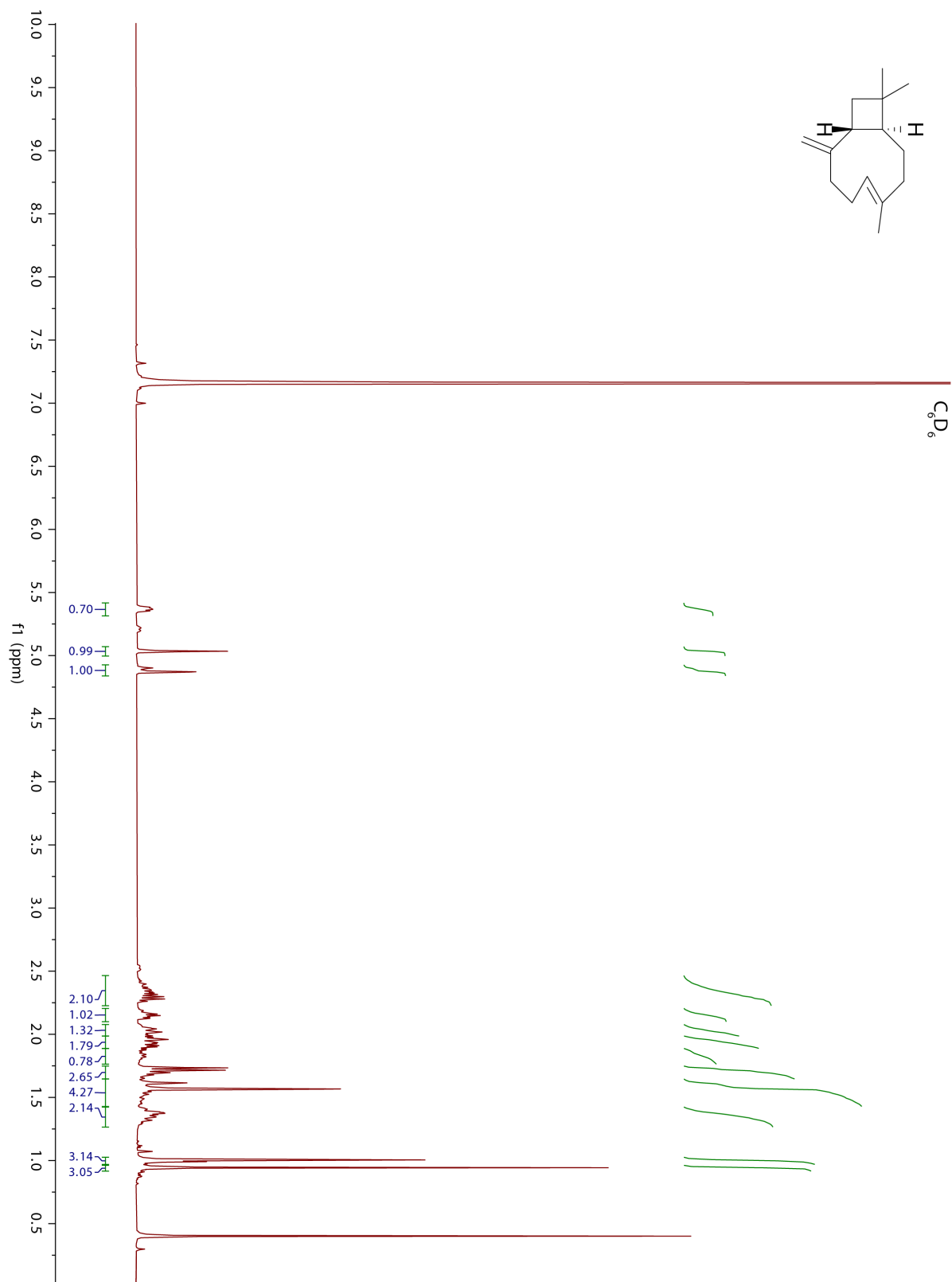
NMR Spectra



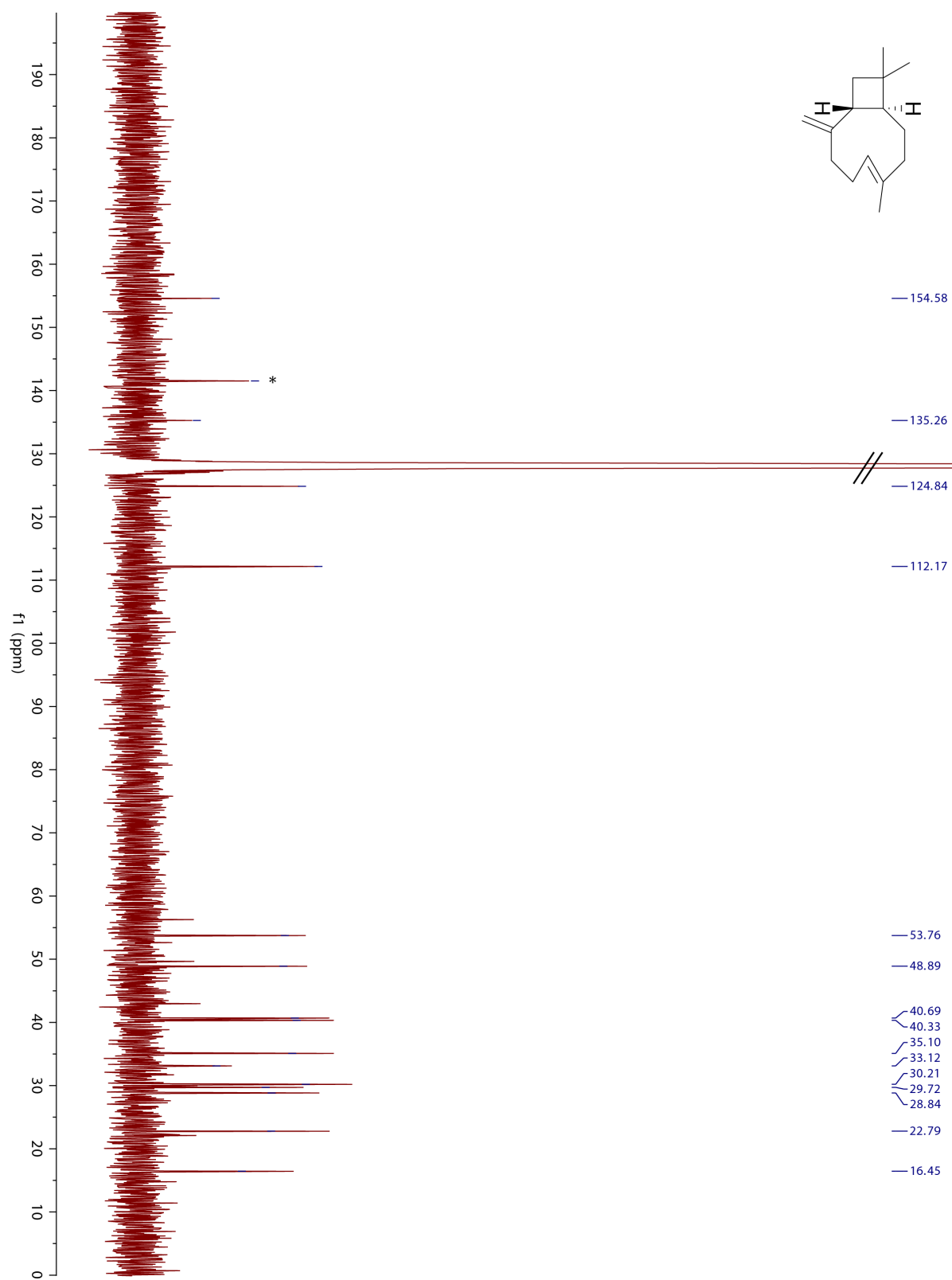
¹H-NMR spectrum of **S1** in CDCl₃.



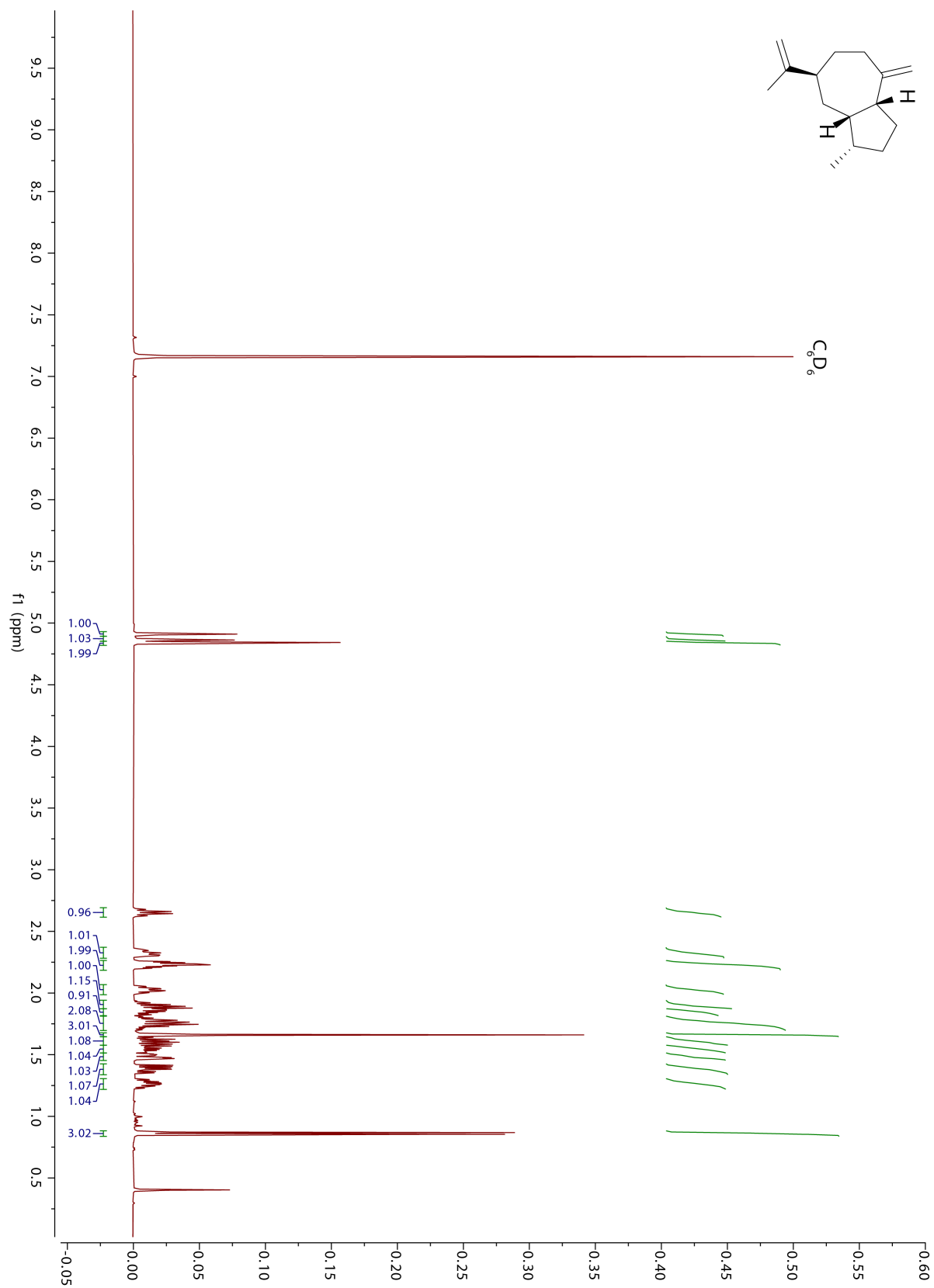
¹³C-NMR spectrum of **S1** in CDCl₃.



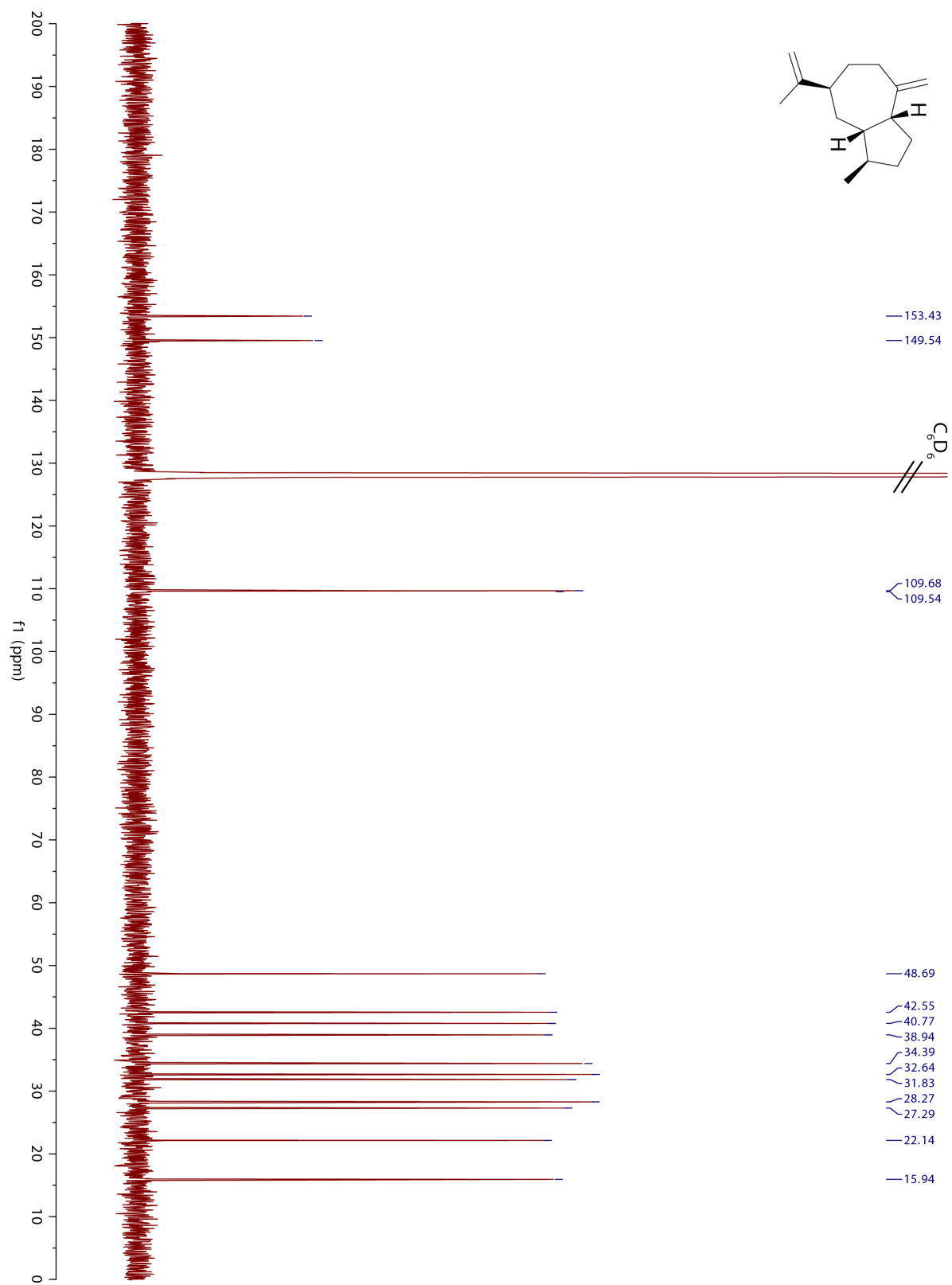
1H -NMR spectrum of **10** in C_6D_6 .

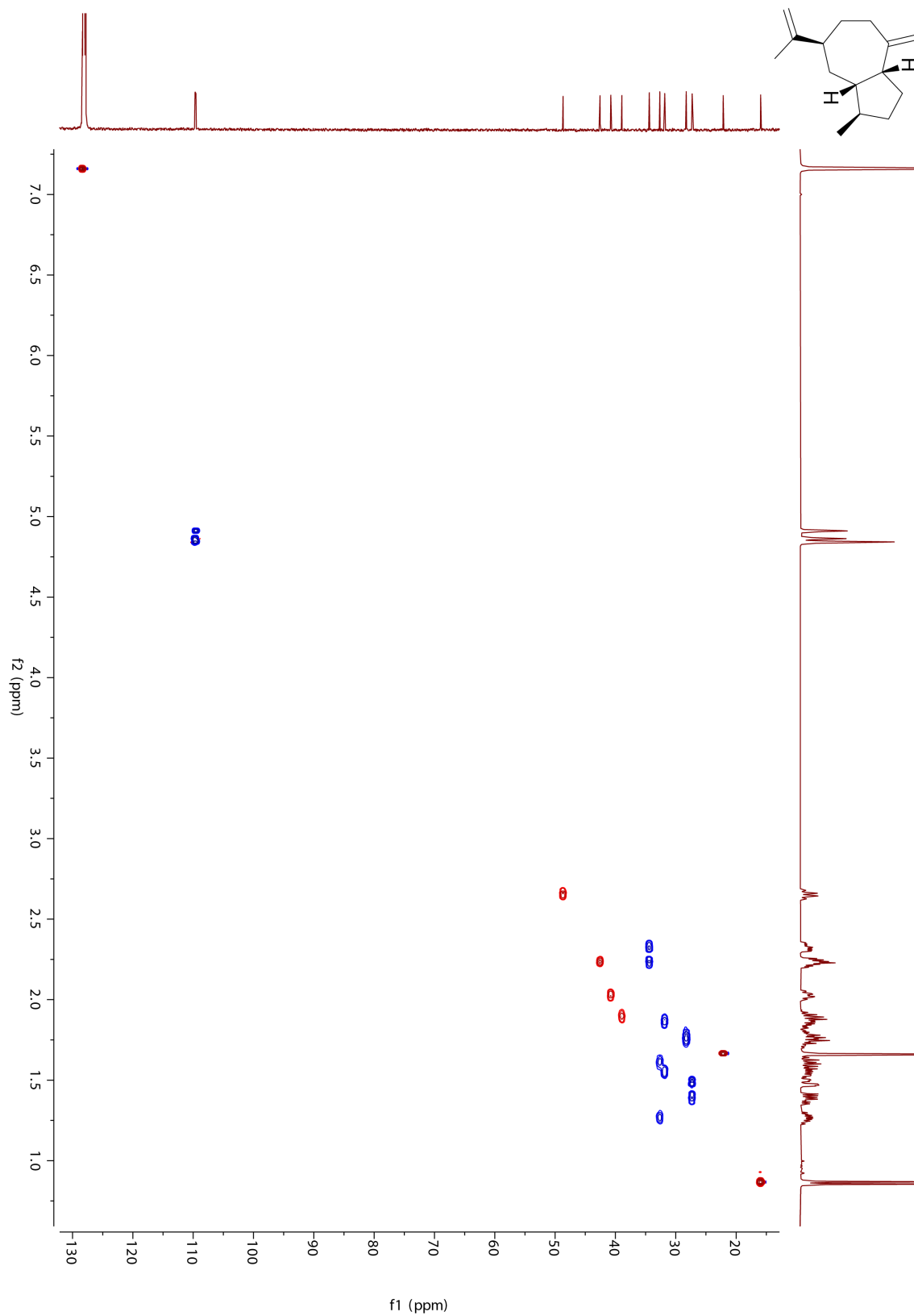


¹³C-NMR spectrum of **10** in C₆D₆. Signal marked with asterisk is a contamination in the used batch of C₆D₆.

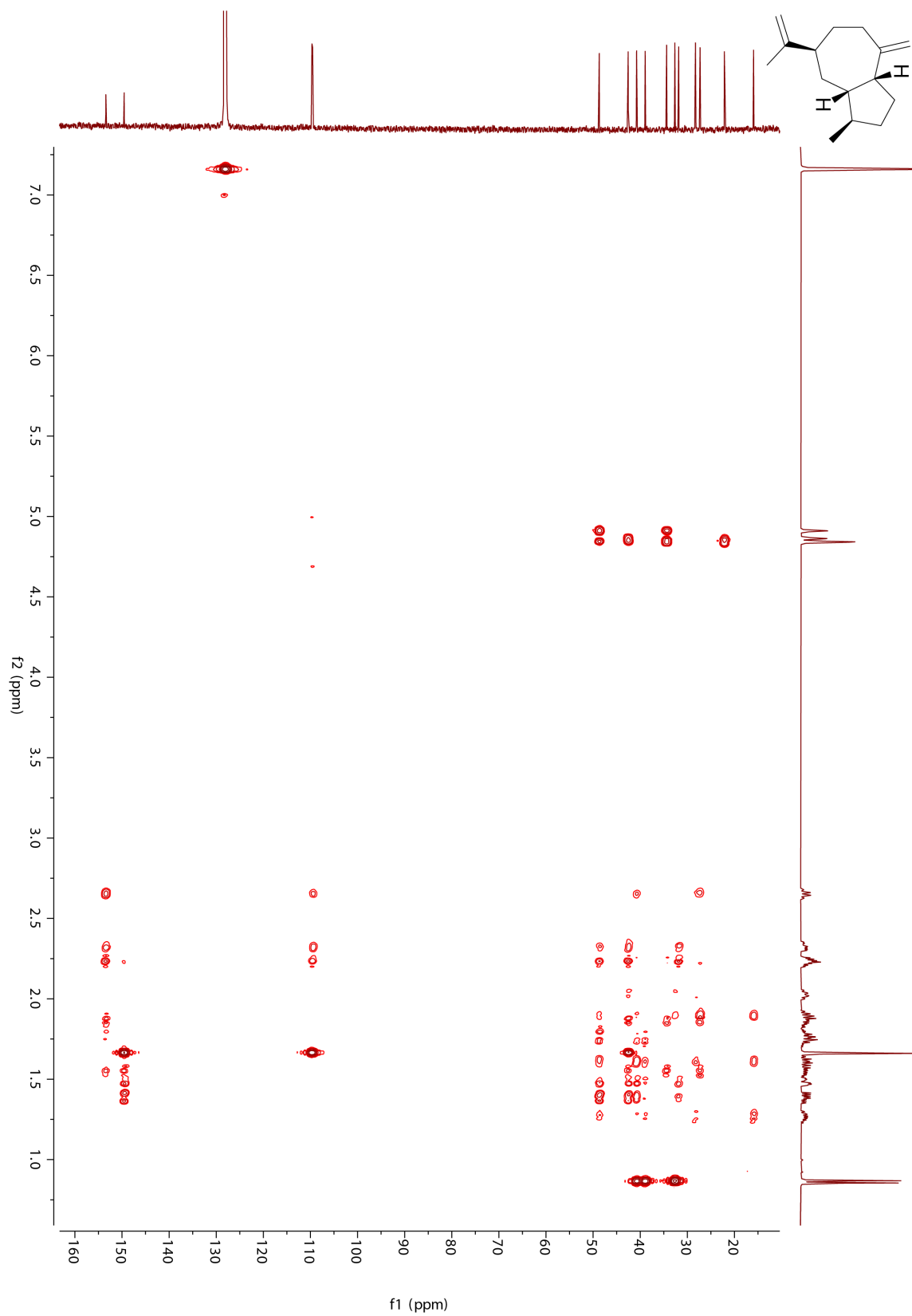


¹H-NMR spectrum of **11** in C₆D₆.

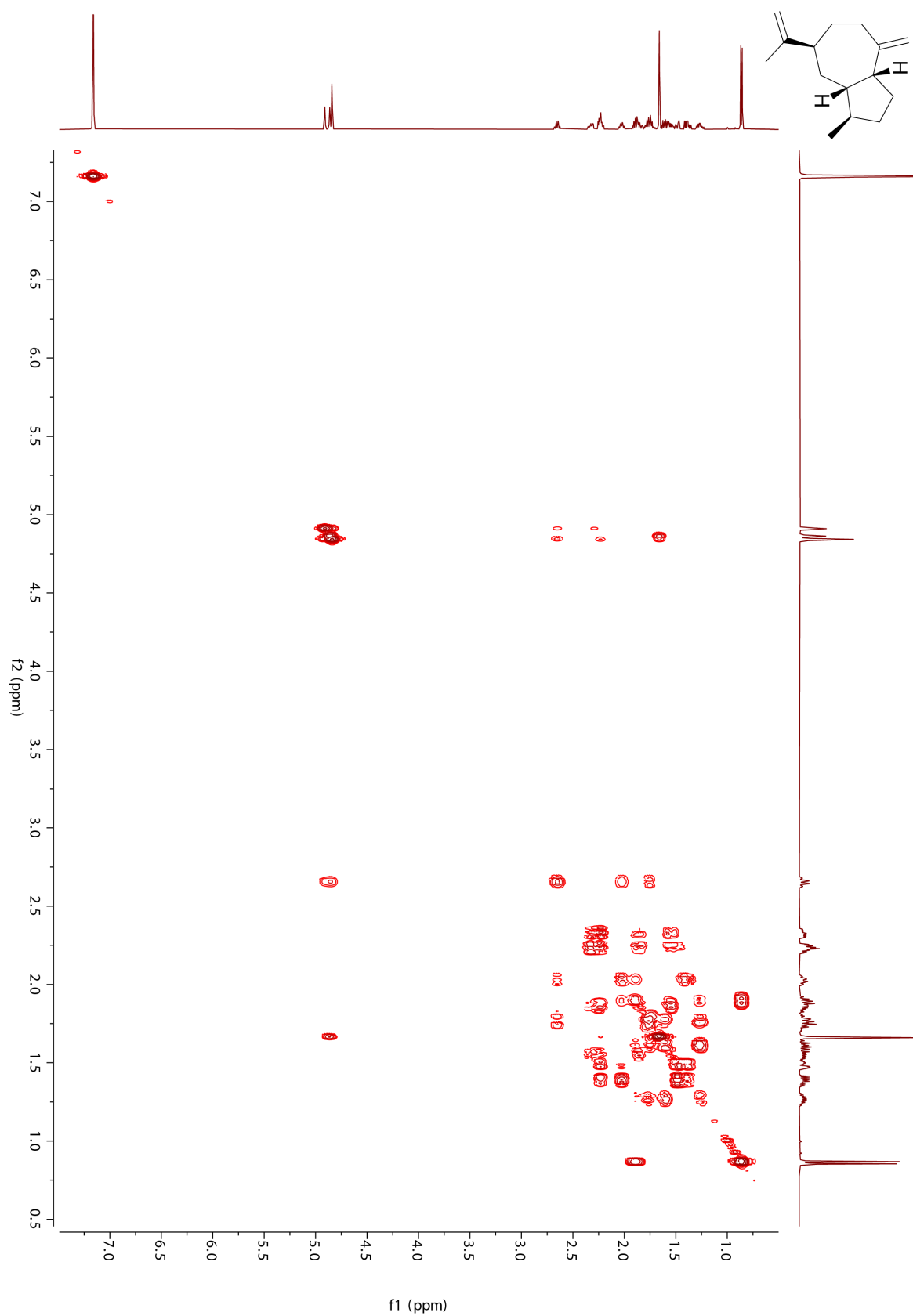




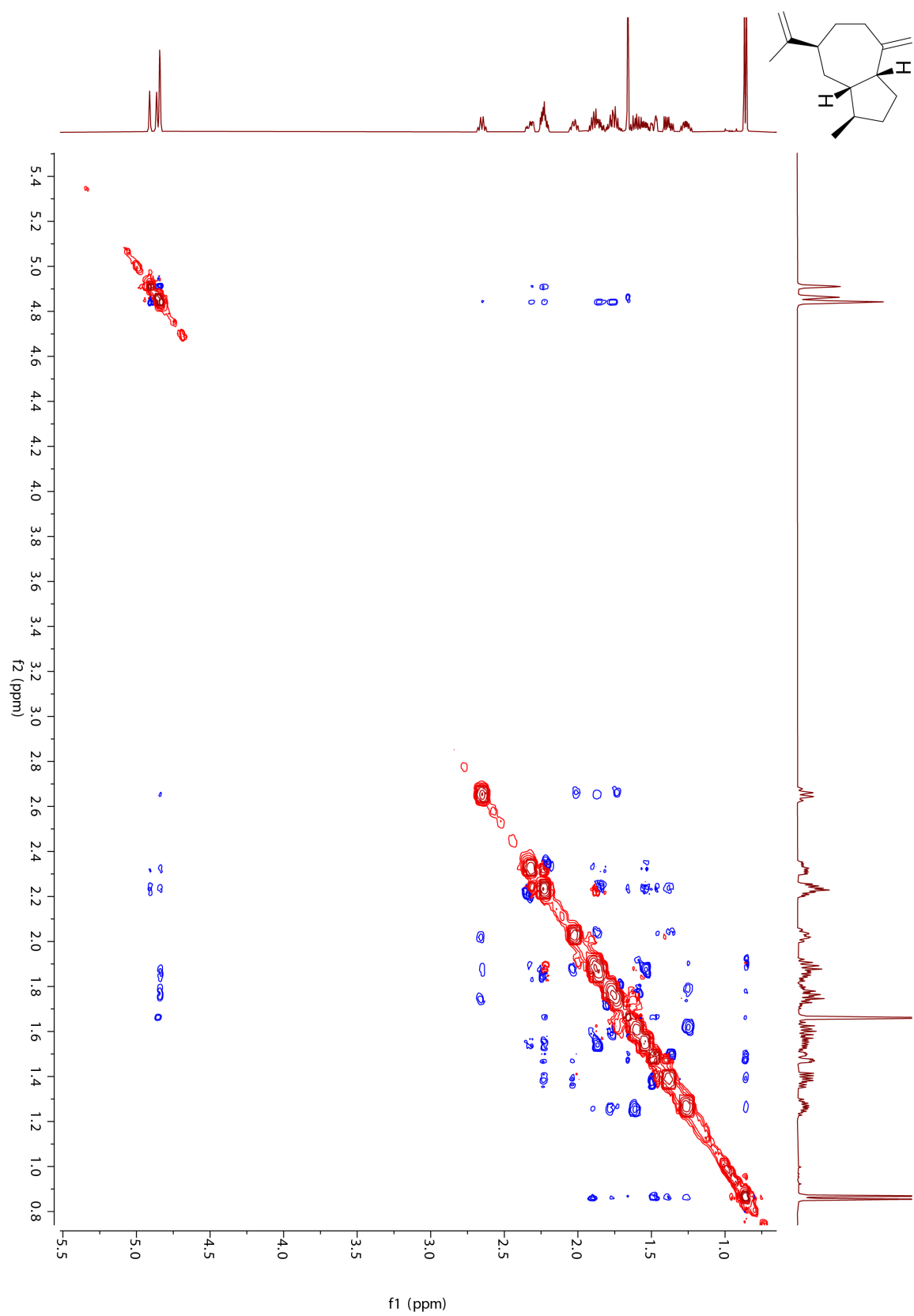
^1H , ^{13}C -HSQC spectrum **11** in C_6D_6 .



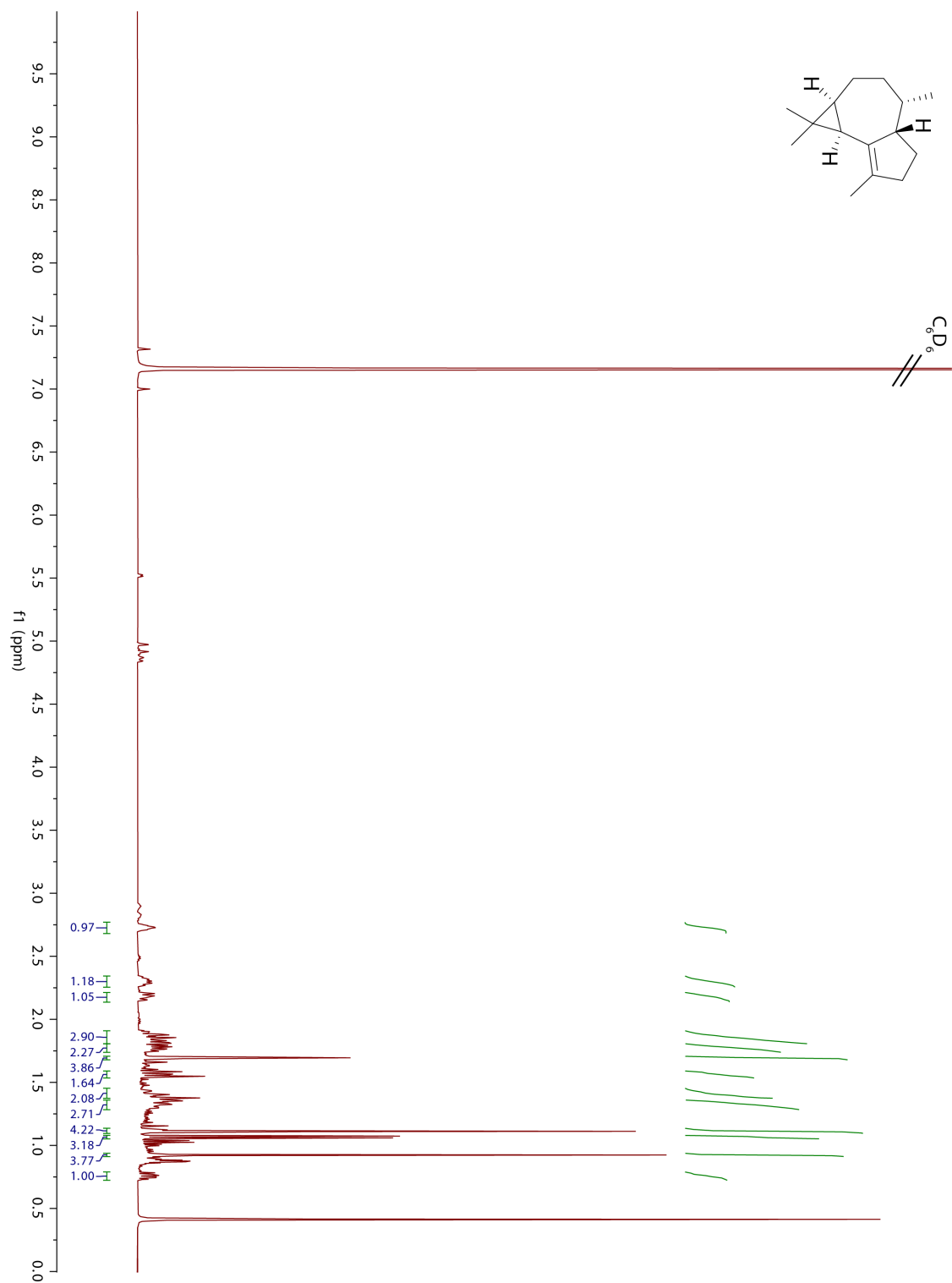
$^1\text{H}, ^{13}\text{C}$ -HMBC spectrum **11** in C_6D_6 .



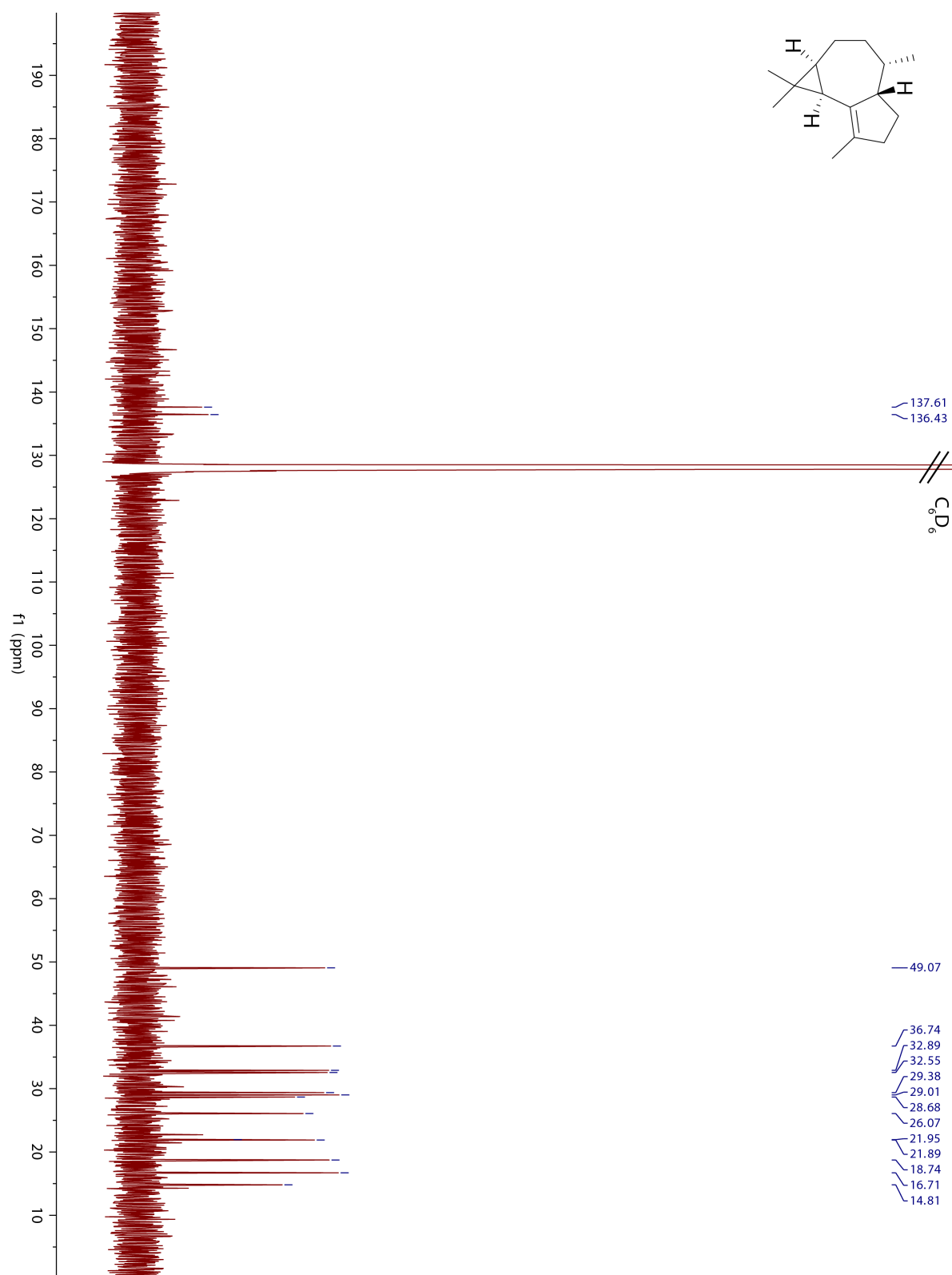
$^1\text{H}, ^1\text{H}$ -COSY spectrum **11** in C_6D_6 .



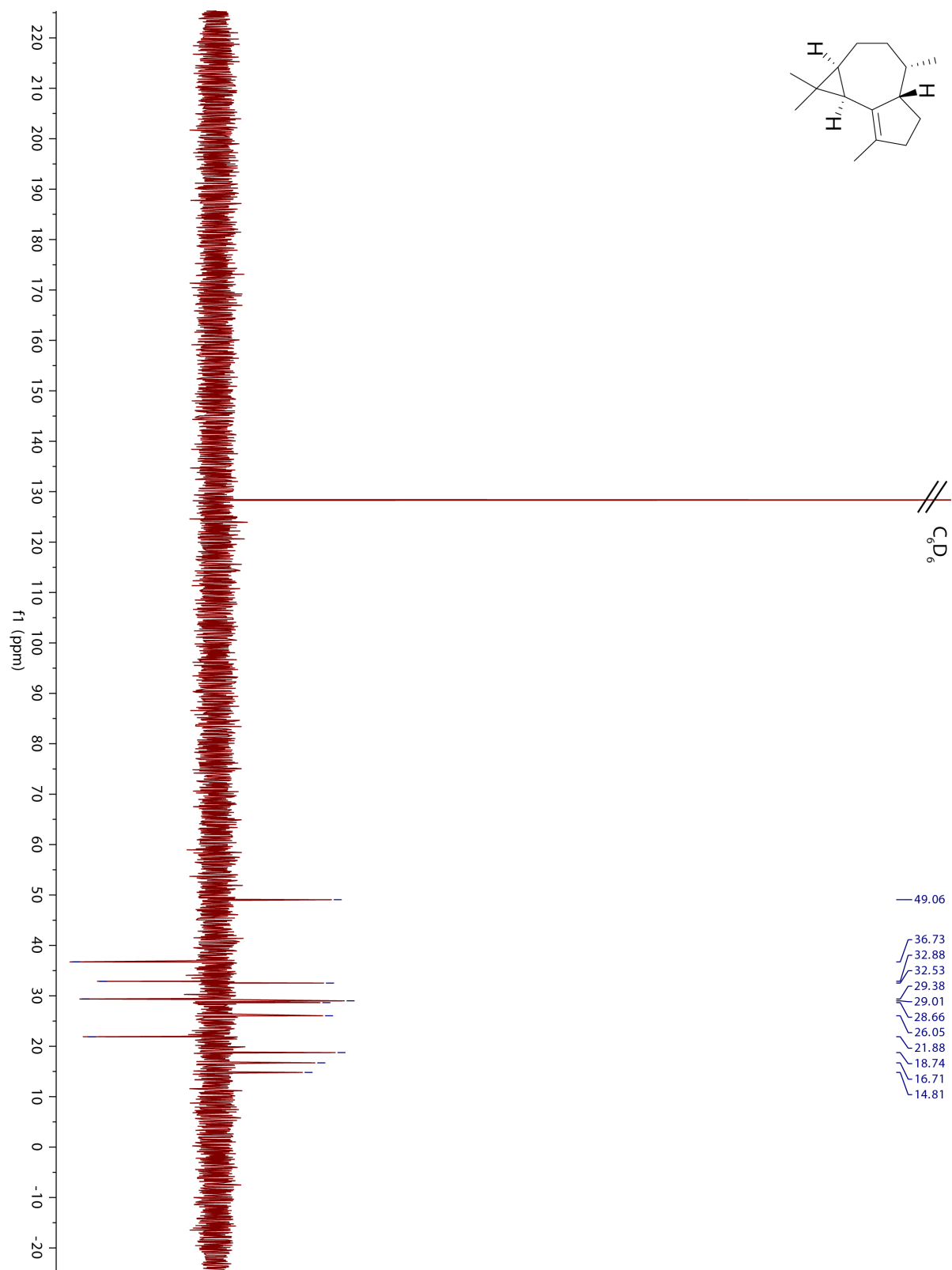
^1H , ^1H -NOESY spectrum **11** in C_6D_6 .

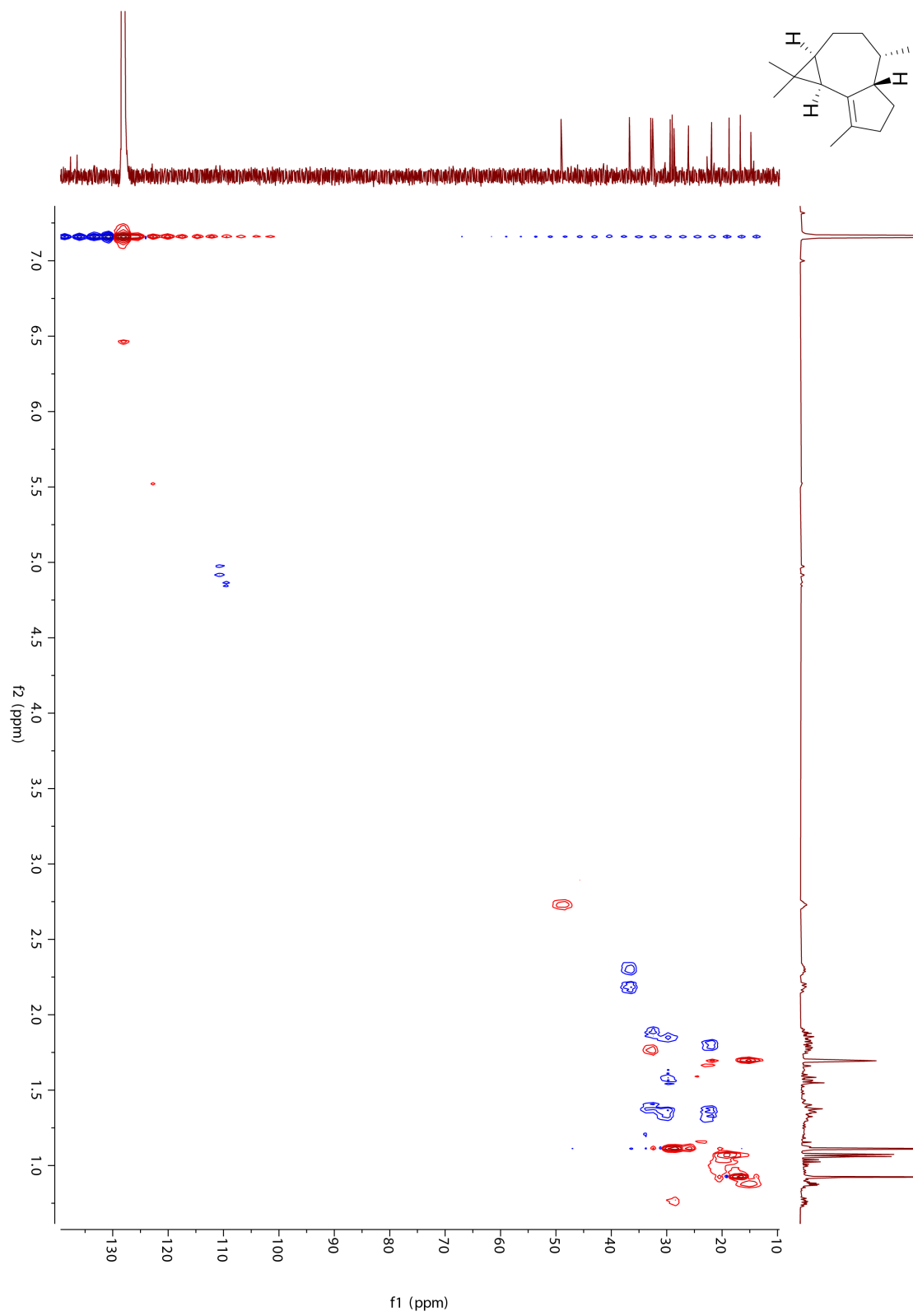


^1H -NMR spectrum of **16** in C_6D_6 .

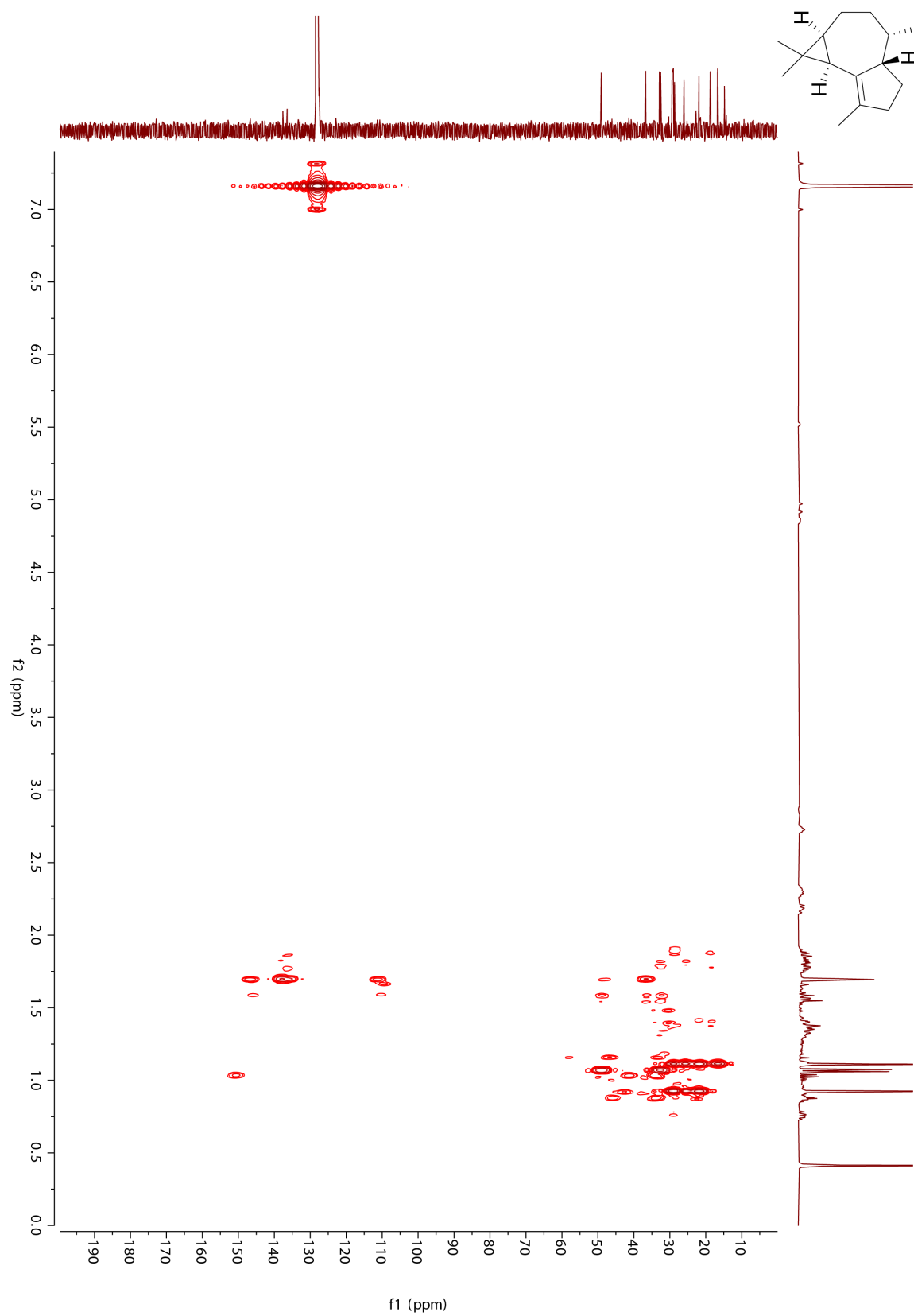


¹³C-NMR spectrum of **16** in C₆D₆.

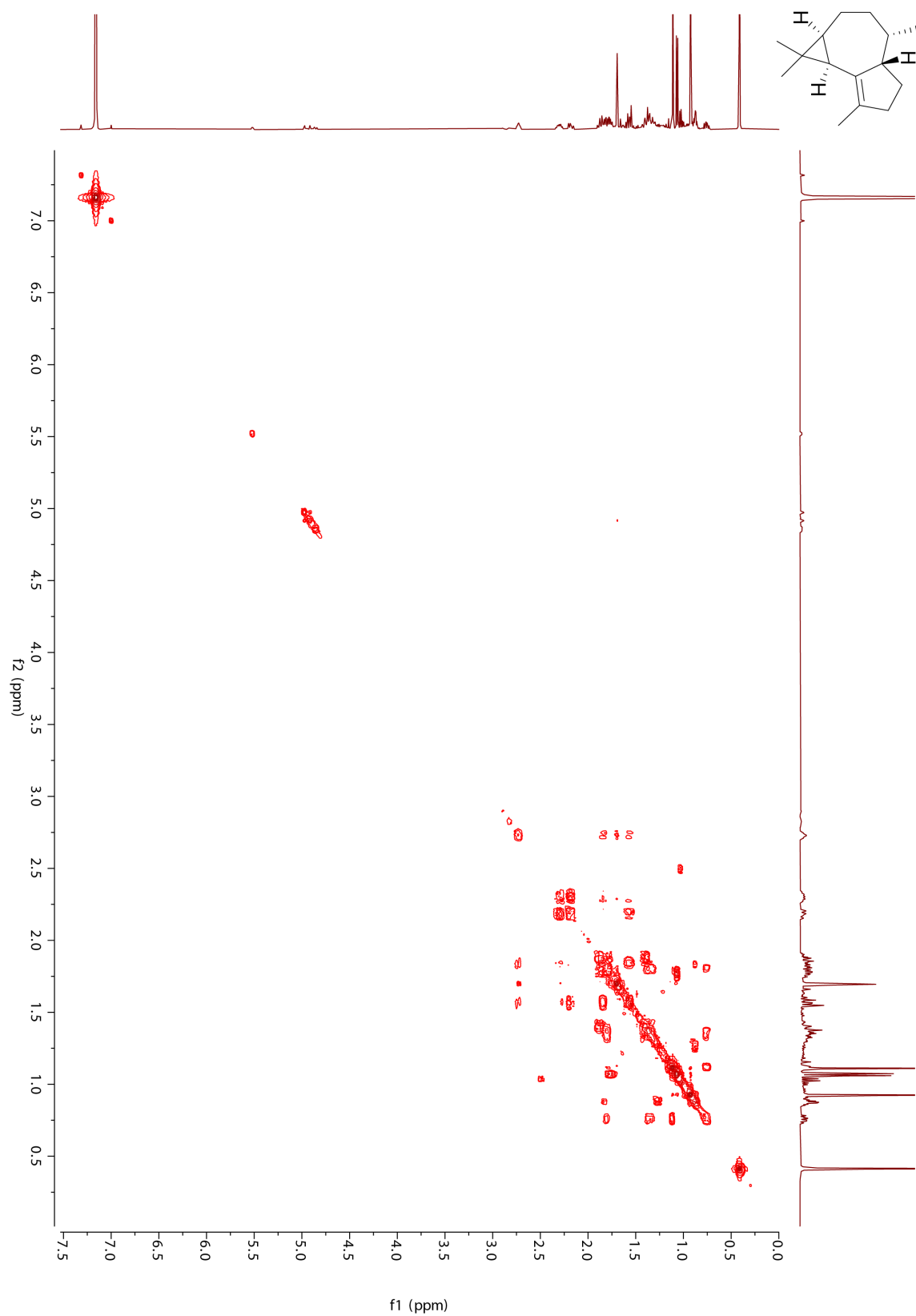




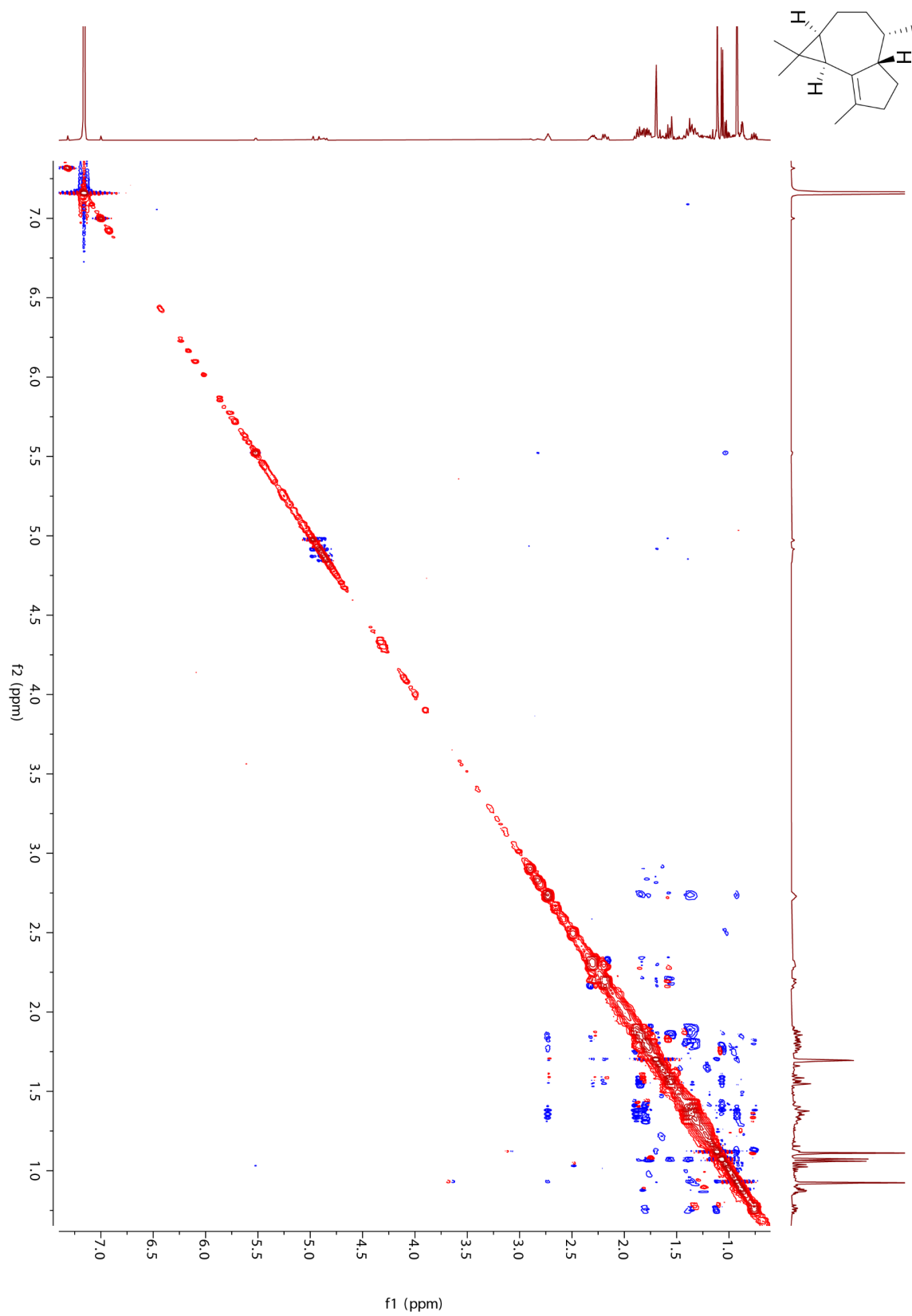
^1H , ^{13}C -HSQC spectrum **16** in C_6D_6 .



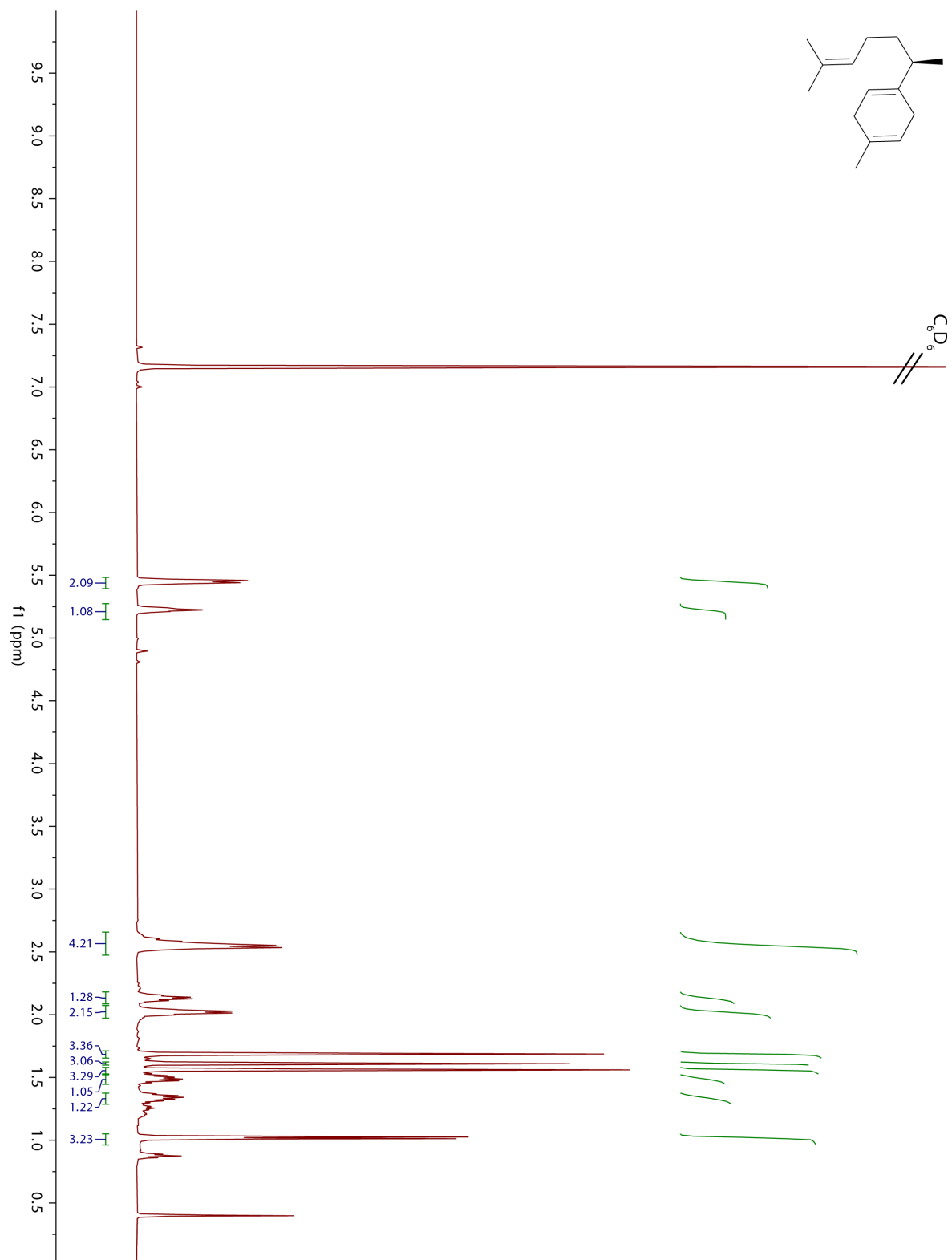
^1H , ^{13}C -HMBC spectrum **16** in C_6D_6 .



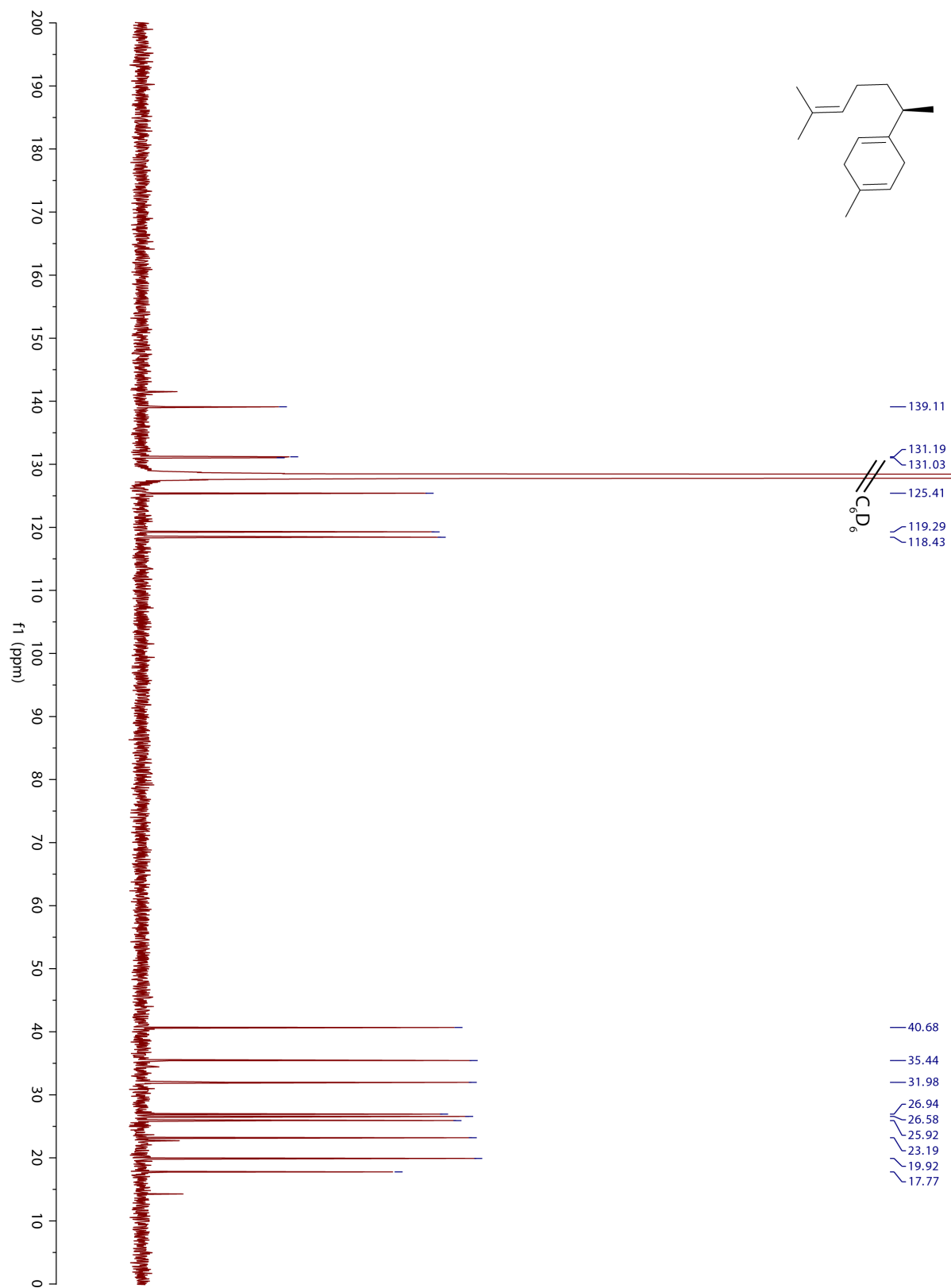
$^1\text{H}, ^1\text{H}$ -COSY spectrum **16** in C_6D_6 .



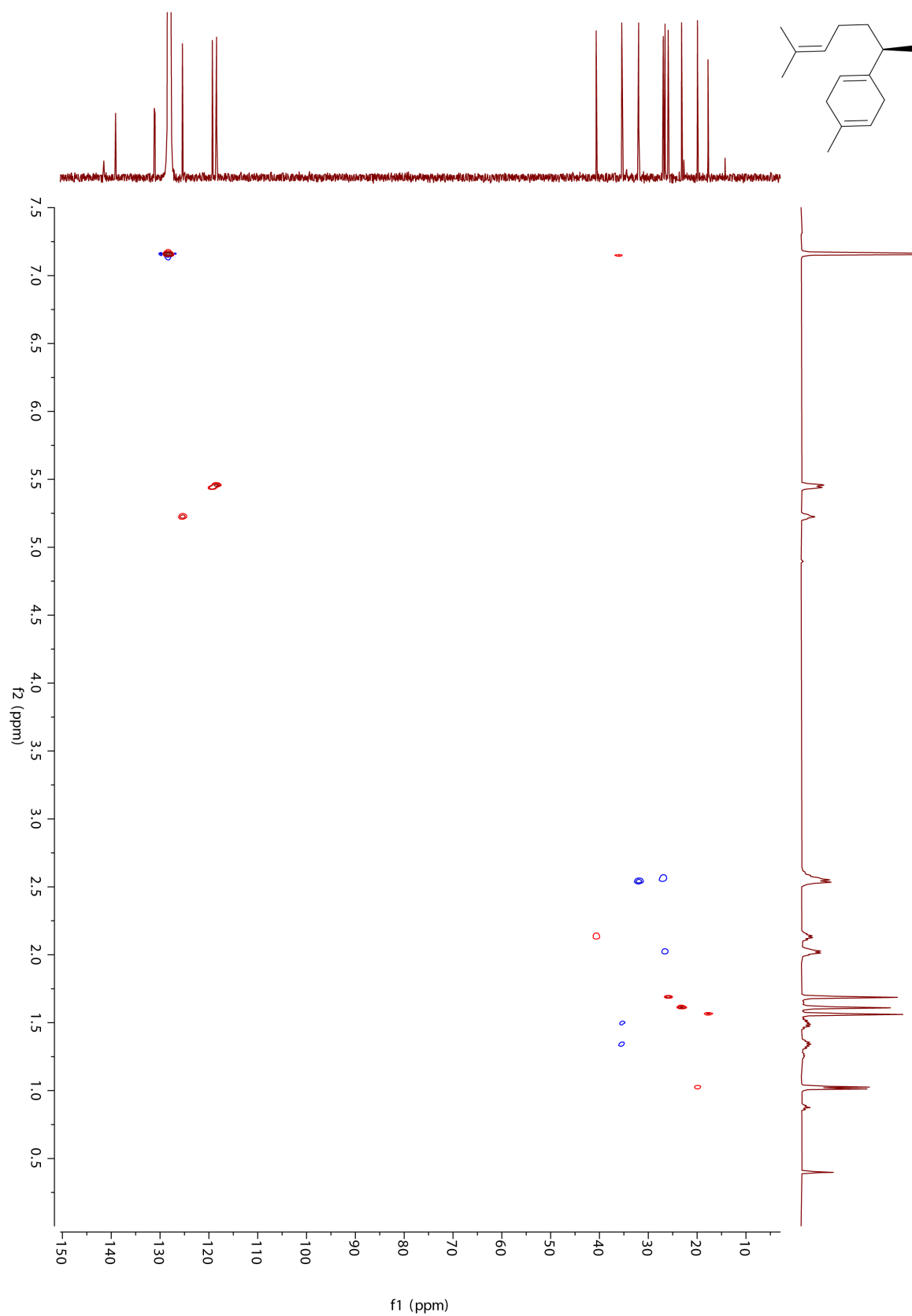
$^1\text{H}, ^1\text{H}$ -NOESY spectrum **16** in C_6D_6 .



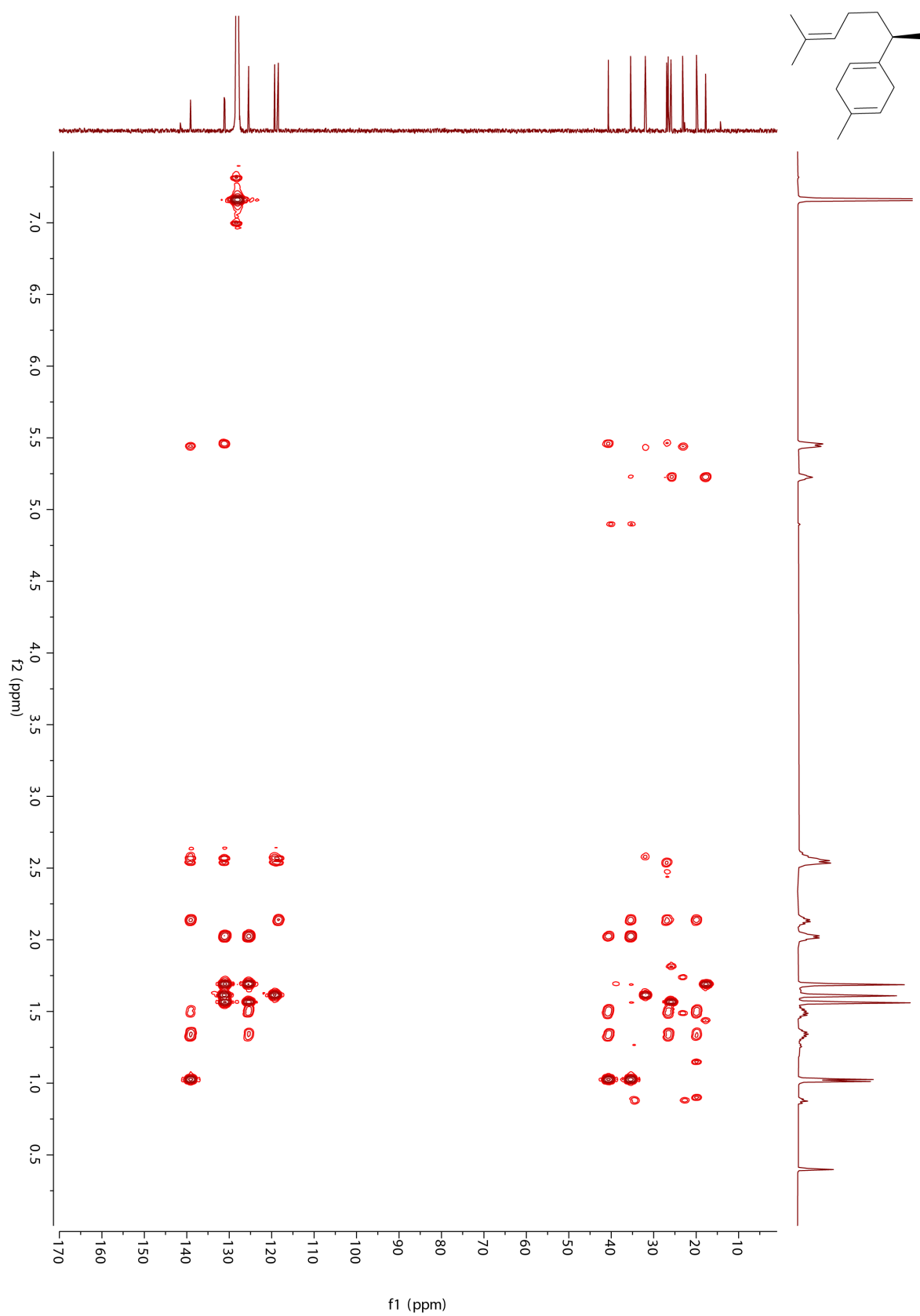
^1H -NMR spectrum of **12** in C_6D_6 .

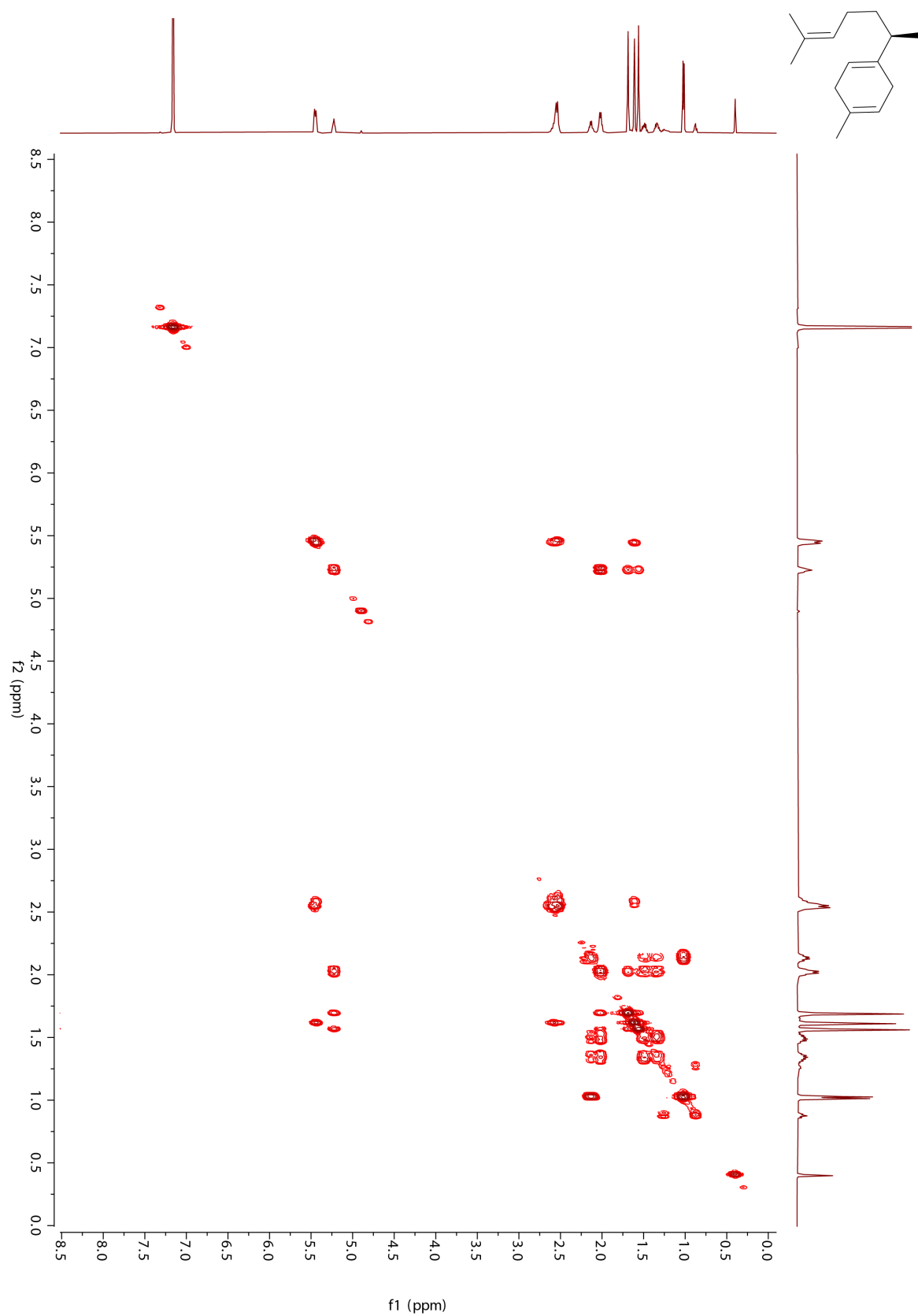


¹³C-NMR spectrum of **12** in C₆D₆.

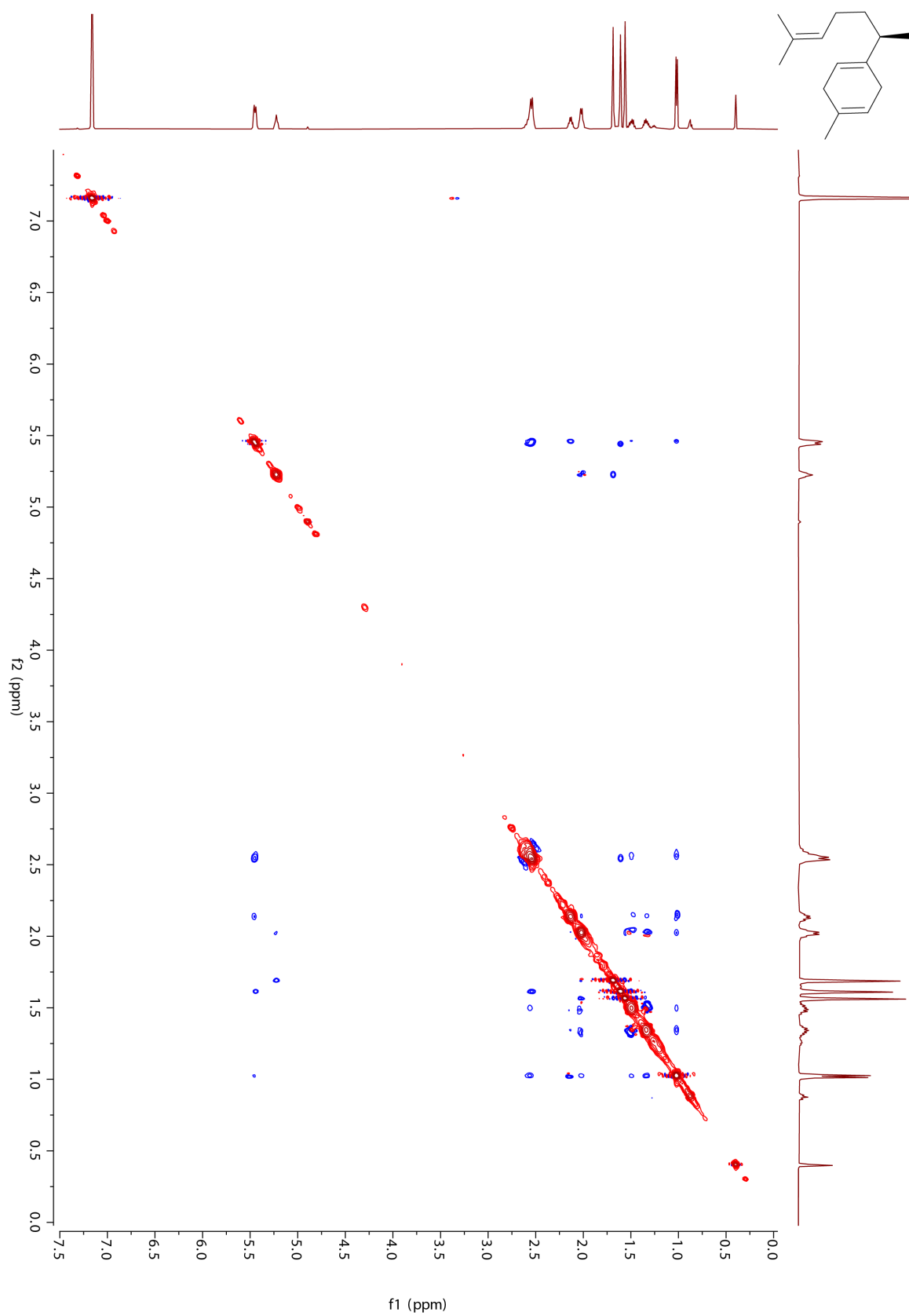


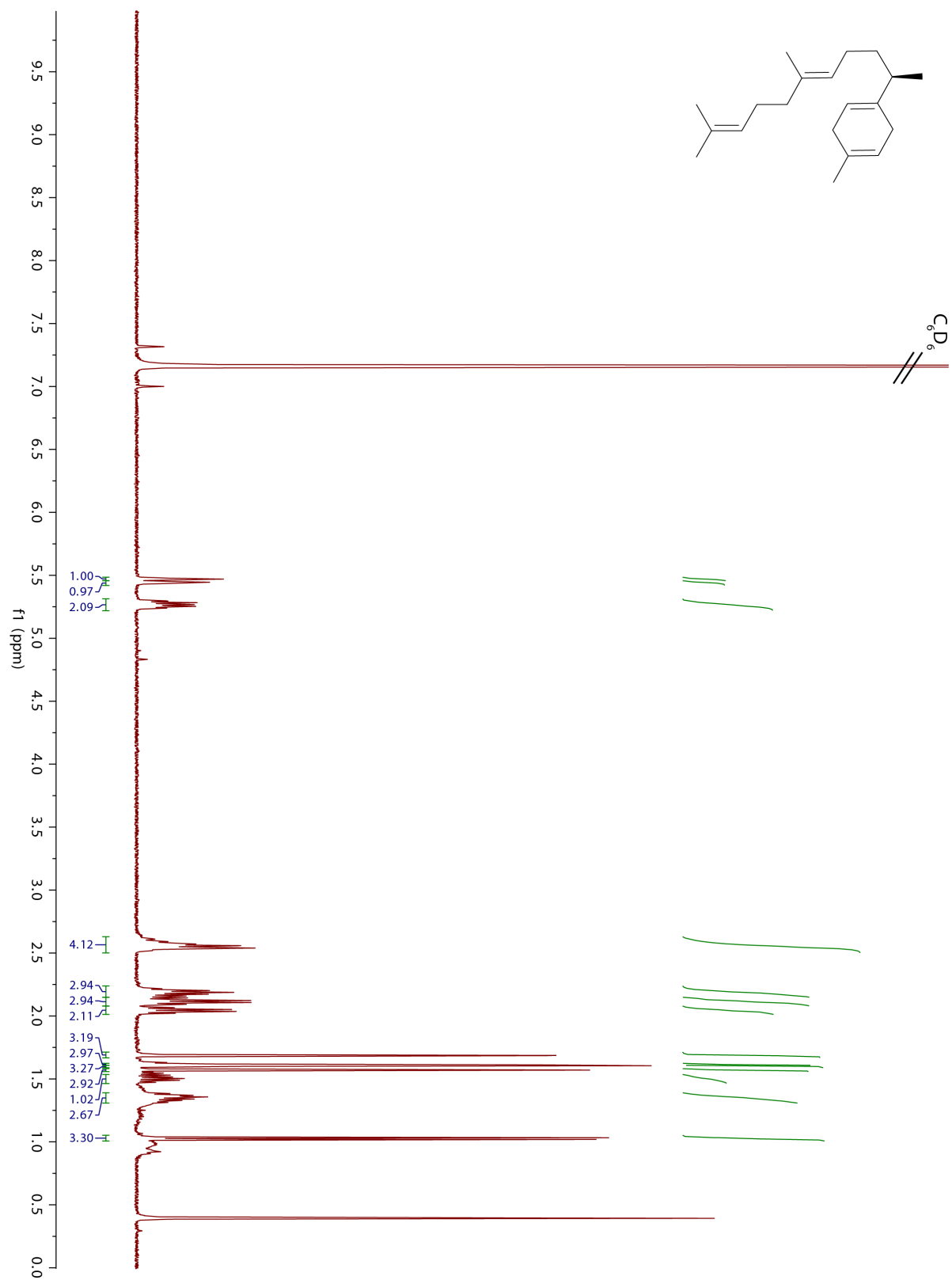
^1H , ^{13}C -HSQC spectrum **12** in C_6D_6 .



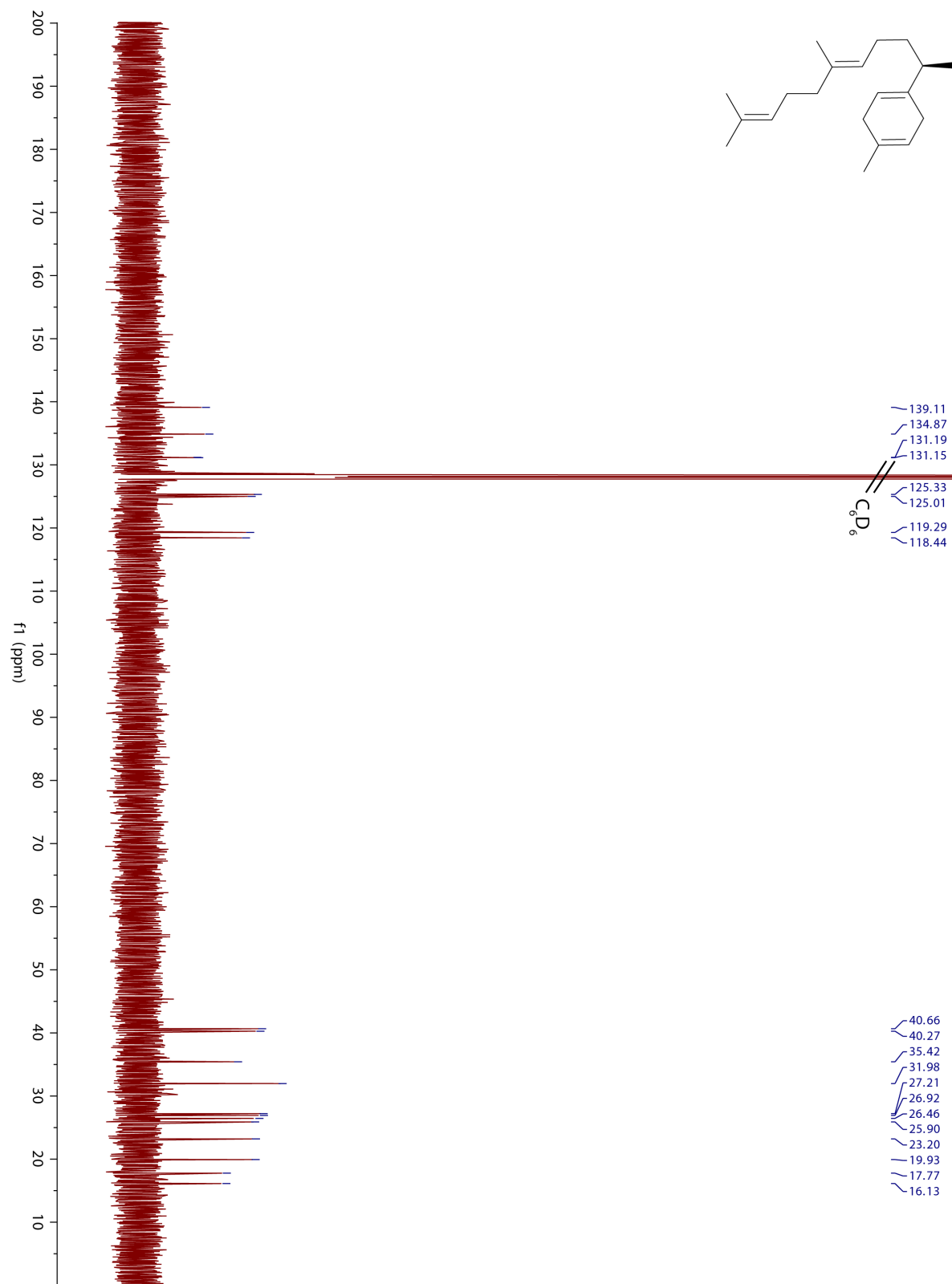


^1H , ^1H -COSY spectrum **12** in C_6D_6 .

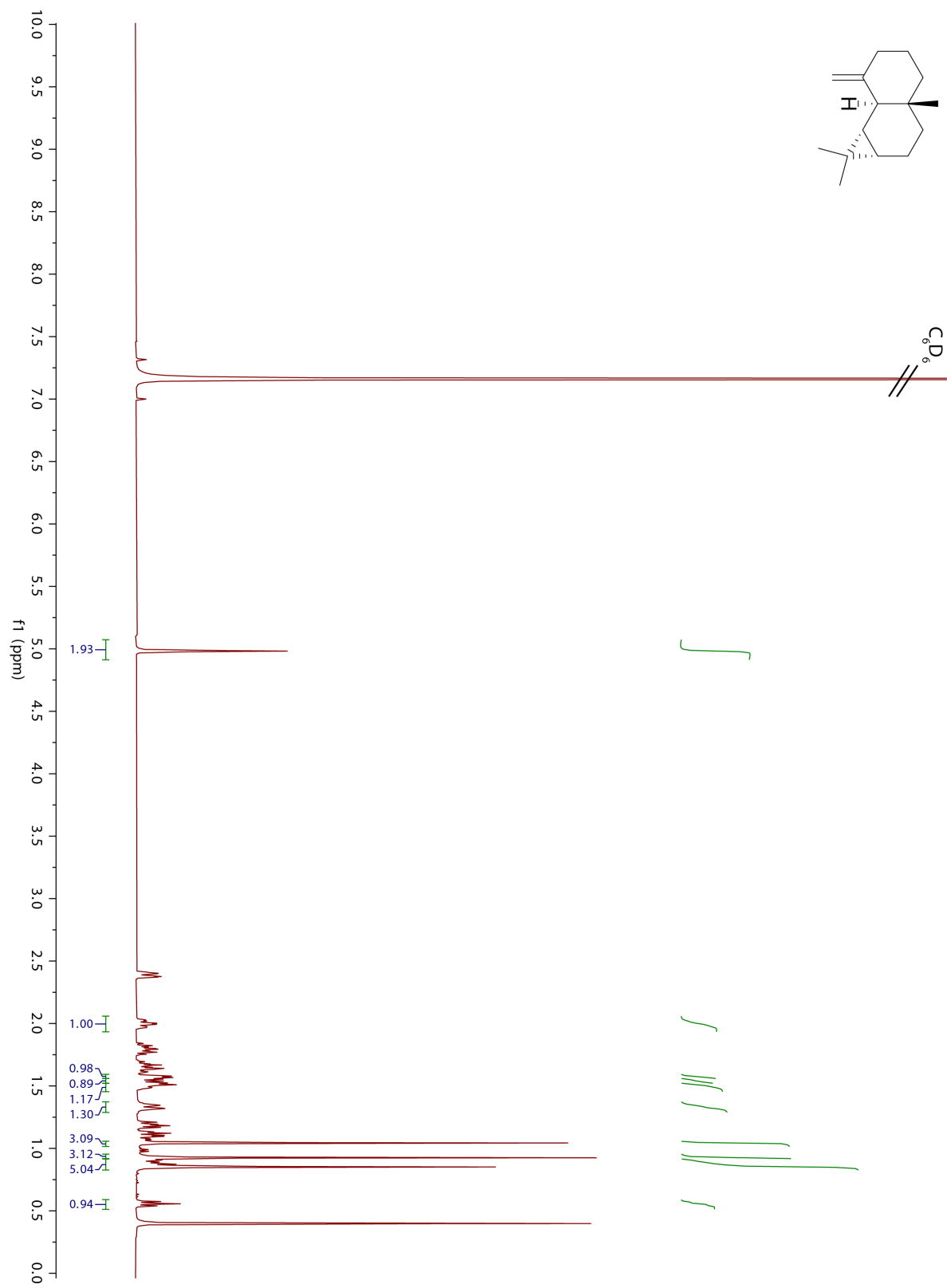




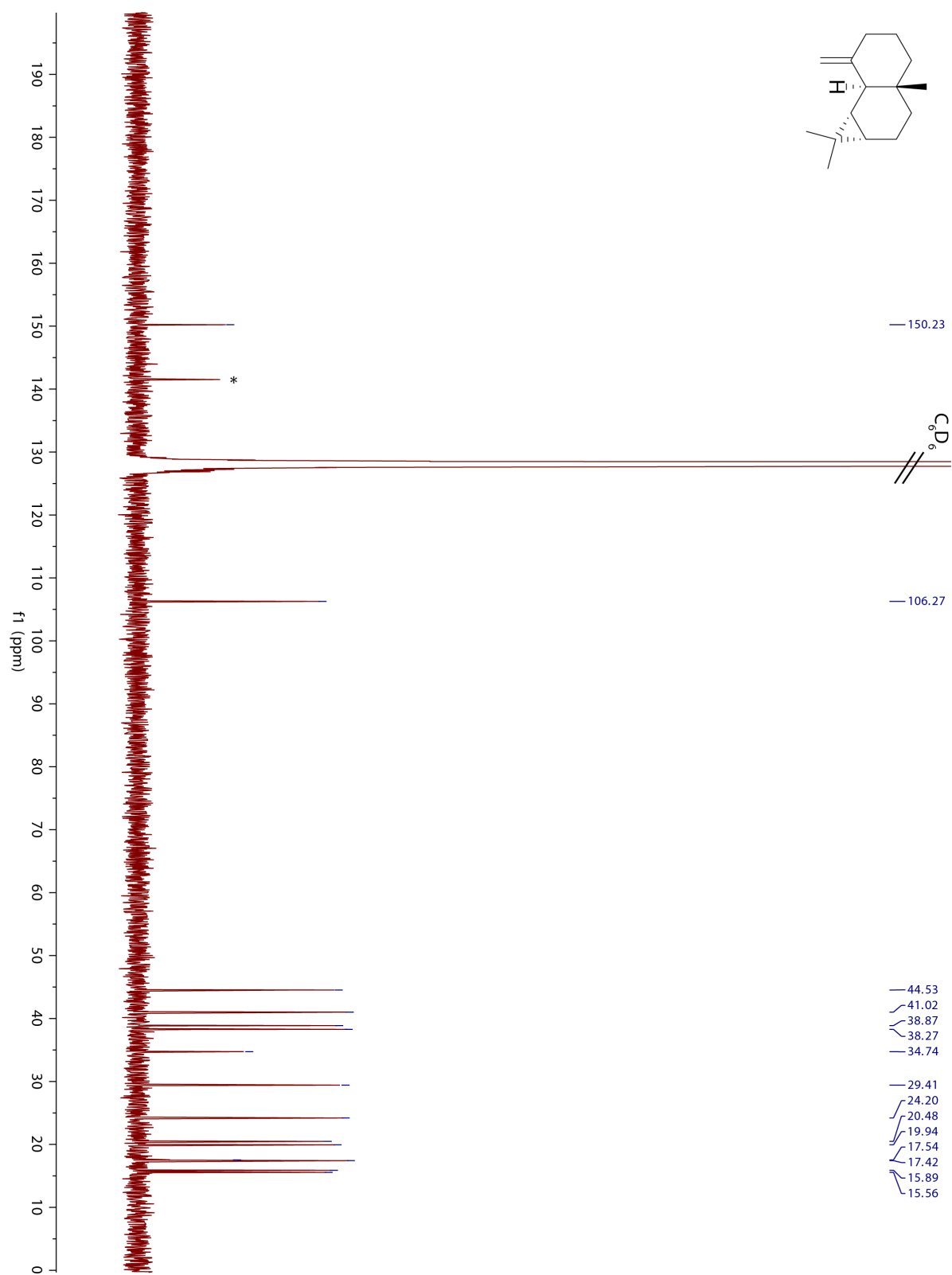
^1H -NMR spectrum of **17** in C_6D_6 .



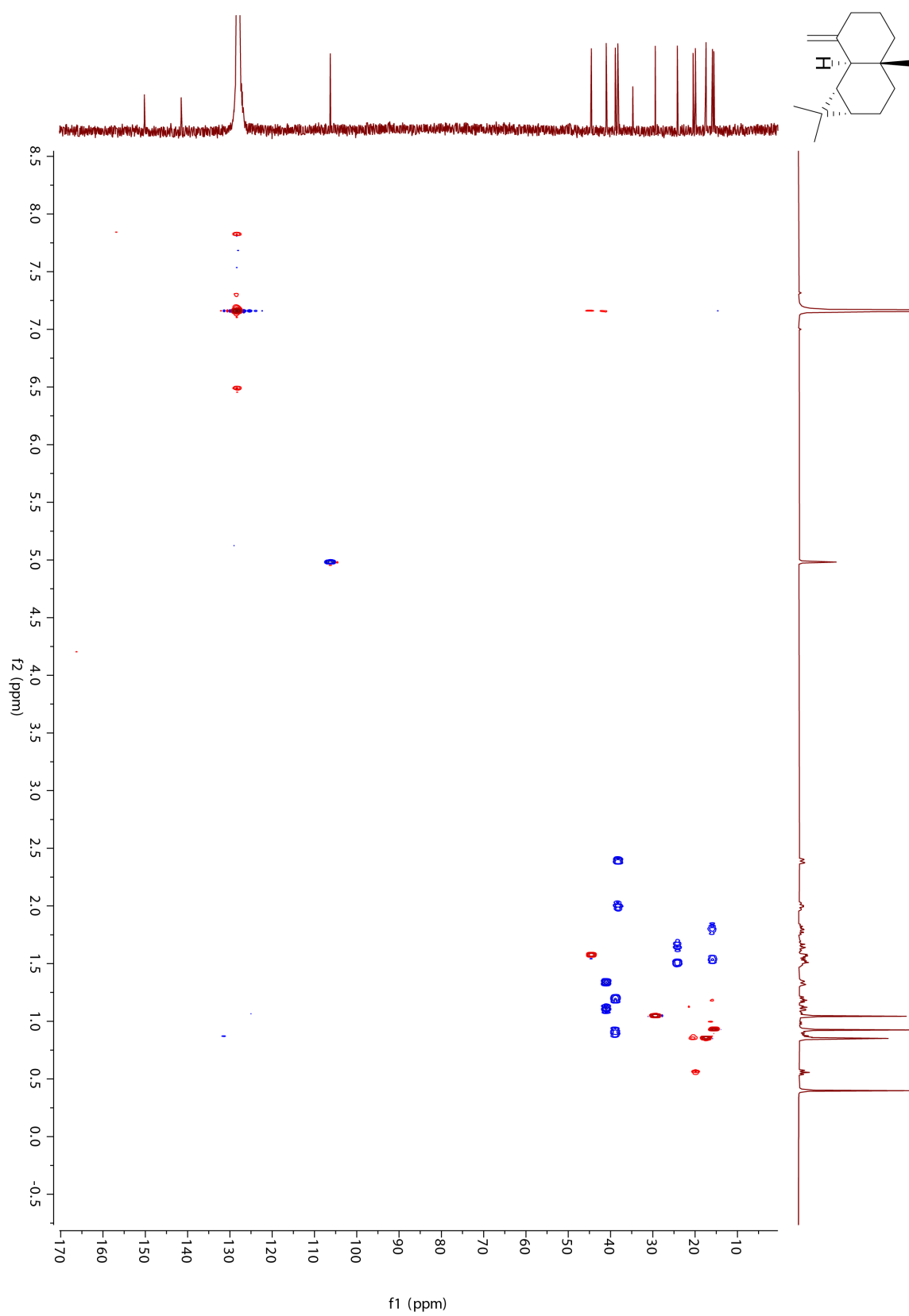
^{13}C -NMR spectrum of **17** in C_6D_6 .



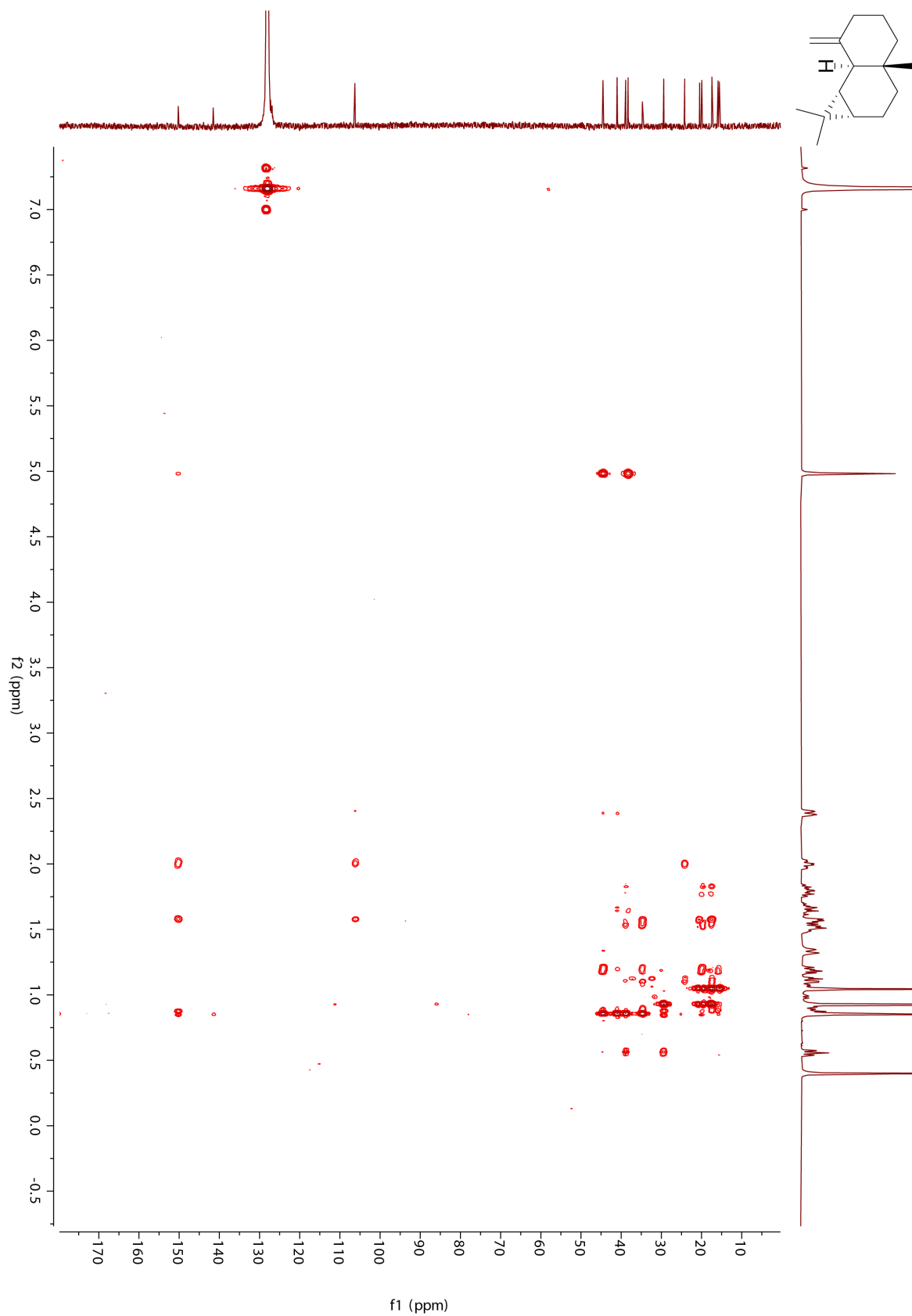
¹H-NMR spectrum of **14** in C₆D₆.



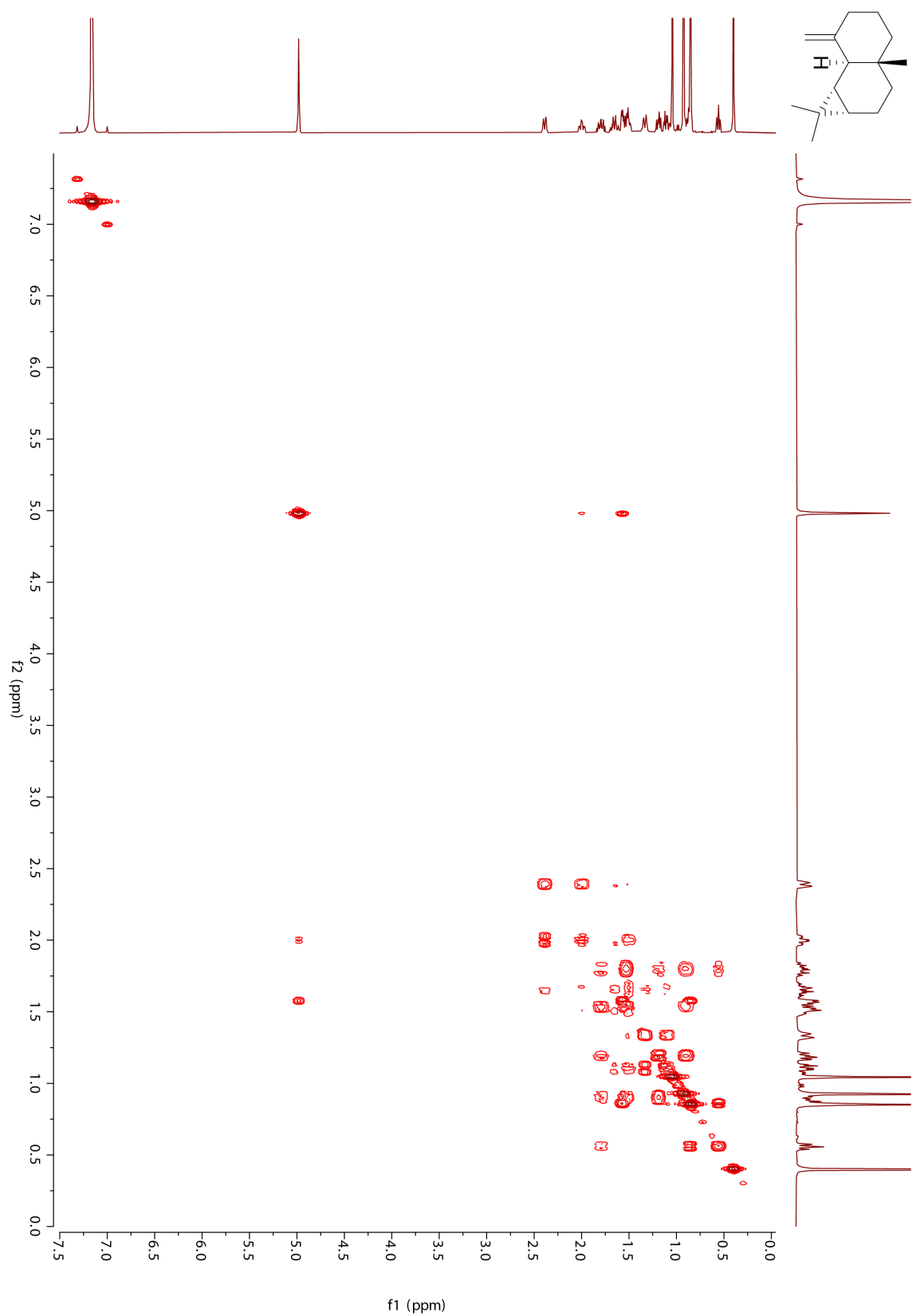
^{13}C -NMR spectrum of **14** in C_6D_6 . Signal marked with asterisk is a contamination in the used batch of C_6D_6 .



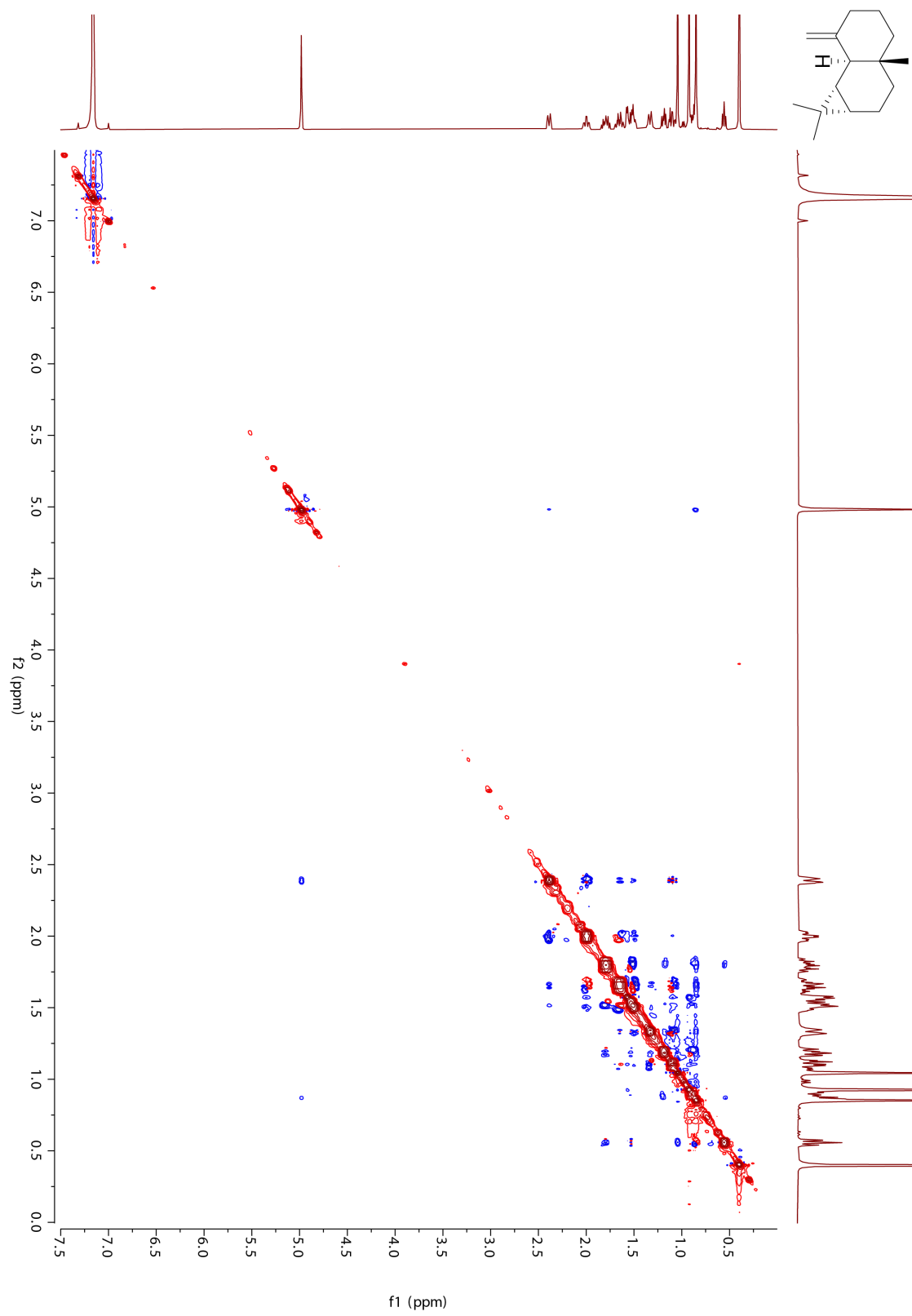
^1H , ^{13}C -HSQC spectrum **14** in C_6D_6 .



^1H , ^{13}C -HMBC spectrum **14** in C_6D_6 .



^1H , ^1H -COSY spectrum **14** in C_6D_6 .



^1H , ^1H -NOESY spectrum **14** in C_6D_6 .

4.9 Supplementary Information References

- (1) Burkhardt, I.; Dürr, L.; Grayson, N. E.; Moore, B. S. Methods for the Discovery and Characterization of Octocoral Terpene Cyclases. In *Methods in Enzymology*; Academic Press, 2024. <https://doi.org/10.1016/bs.mie.2024.02.011>.
- (2) Grayson, N. E.; Scesa, P. D.; Moore, M. L.; Ledoux, J.-B.; Gomez-Garrido, J.; Alioto, T.; Michael, T. P.; Burkhardt, I.; Schmidt, E. W.; Moore, B. S. A Widespread Metabolic Gene Cluster Family in Metazoans. *Nat. Chem. Biol.* **2025**, 1–10. <https://doi.org/10.1038/s41589-025-01927-y>.
- (3) Burkhardt, I.; de Rond, T.; Chen, P. Y.-T.; Moore, B. S. Ancient Plant-like Terpene Biosynthesis in Corals. *Nat. Chem. Biol.* **2022**, 18 (6), 664–669. <https://doi.org/10.1038/s41589-022-01026-2>.
- (4) Chen, S. Ultrafast One-Pass FASTQ Data Preprocessing, Quality Control, and Deduplication Using Fastp. *iMeta* **2023**, 2 (2), e107. <https://doi.org/10.1002/imt2.107>.
- (5) Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 2024-07-30).
- (6) Marçais, G.; Kingsford, C. A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-Mers. *Bioinformatics* **2011**, 27 (6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>.
- (7) Vurture, G. W.; Sedlazeck, F. J.; Nattestad, M.; Underwood, C. J.; Fang, H.; Gurtowski, J.; Schatz, M. C. GenomeScope: Fast Reference-Free Genome Profiling from Short Reads. *Bioinformatics* **2017**, 33 (14), 2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>.
- (8) De Coster, W.; Rademakers, R. NanoPack2: Population-Scale Evaluation of Long-Read Sequencing Data. *Bioinformatics* **2023**, 39 (5), btad311. <https://doi.org/10.1093/bioinformatics/btad311>.
- (9) Cheng, H.; Jarvis, E. D.; Fedrigo, O.; Koepfli, K.-P.; Urban, L.; Gemmell, N. J.; Li, H. Haplotype-Resolved Assembly of Diploid Genomes without Parental Data. *Nat. Biotechnol.* **2022**, 40 (9), 1332–1335. <https://doi.org/10.1038/s41587-022-01261-x>.
- (10) Tegenfeldt, F.; Kuznetsov, D.; Manni, M.; Berkeley, M.; Zdobnov, E. M.; Kriventseva, E. V. OrthoDB and BUSCO Update: Annotation of Orthologs with Wider Sampling of Genomes. *Nucleic Acids Res.* **2025**, 53 (D1), D516–D522. <https://doi.org/10.1093/nar/gkae987>.
- (11) Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, 29 (8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- (12) Flynn, J. M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A. G.; Feschotte, C.; Smit, A. F. RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families. *Proc. Natl. Acad. Sci.* **2020**, 117 (17), 9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
- (13) Stanke, M.; Diekhans, M.; Baertsch, R.; Haussler, D. Using Native and Syntenically Mapped cDNA Alignments to Improve de Novo Gene Finding. *Bioinformatics* **2008**, 24 (5), 637–644. <https://doi.org/10.1093/bioinformatics/btn013>.
- (14) Stanke, M.; Schöffmann, O.; Morgenstern, B.; Waack, S. Gene Prediction in Eukaryotes with a Generalized Hidden Markov Model That Uses Hints from External Sources. *BMC Bioinformatics* **2006**, 7 (1), 62. <https://doi.org/10.1186/1471-2105-7-62>.
- (15) Gabriel, L.; Hoff, K. J.; Brůna, T.; Borodovsky, M.; Stanke, M. TSEBRA: Transcript Selector for BRAKER. *BMC Bioinformatics* **2021**, 22, 566. <https://doi.org/10.1186/s12859-021-04482-0>.
- (16) Kim, D.; Paggi, J. M.; Park, C.; Bennett, C.; Salzberg, S. L. Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* **2019**, 37 (8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.

- (17) Cantalapiedra, C. P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-Mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38* (12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>.
- (18) Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S. K.; Cook, H.; Mende, D. R.; Letunic, I.; Rattei, T.; Jensen, L. J.; von Mering, C.; Bork, P. eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res.* **2019**, *47* (D1), D309–D314. <https://doi.org/10.1093/nar/gky1085>.
- (19) Emms, D. M.; Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* **2019**, *20* (1), 238. <https://doi.org/10.1186/s13059-019-1832-y>.
- (20) Tang, H.; Krishnakumar, V.; Zeng, X.; Xu, Z.; Taranto, A.; Lomas, J. S.; Zhang, Y.; Huang, Y.; Wang, Y.; Yim, W. C.; Zhang, J.; Zhang, X. JCVI: A Versatile Toolkit for Comparative Genomics Analysis. *iMeta* **2024**, *3* (4), e211. <https://doi.org/10.1002/imt2.211>.
- (21) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A. Trinity: Reconstructing a Full-Length Transcriptome without a Genome from RNA-Seq Data. *Nat. Biotechnol.* **2011**, *29* (7), 644–652. <https://doi.org/10.1038/nbt.1883>.
- (22) Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16* (6), 276–277. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2).
- (23) Potter, S. C.; Luciani, A.; Eddy, S. R.; Park, Y.; Lopez, R.; Finn, R. D. HMMER Web Server: 2018 Update. *Nucleic Acids Res.* **2018**, *46* (W1), W200–W204. <https://doi.org/10.1093/nar/gky448>.
- (24) Xu, B.; Ning, W.; Wei, X.; Rudolf, J. D. Mutation of the Eunicellane Synthase Bnd4 Alters Its Product Profile and Expands Its Prenylation Ability. *Org. Biomol. Chem.* **2022**, *20* (45), 8833–8837. <https://doi.org/10.1039/D2OB01931K>.
- (25) Lassmann, T.; Frings, O.; Sonnhammer, E. L. L. Kalign2: High-Performance Multiple Alignment of Protein and Nucleotide Sequences Allowing External Features. *Nucleic Acids Res.* **2009**, *37* (3), 858. <https://doi.org/10.1093/nar/gkn1006>.
- (26) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303. <https://doi.org/10.1093/nar/gky427>.
- (27) Ou, S.; Su, W.; Liao, Y.; Chougule, K.; Agda, J. R. A.; Hellings, A. J.; Lugo, C. S. B.; Elliott, T. A.; Ware, D.; Peterson, T.; Jiang, N.; Hirsch, C. N.; Hufford, M. B. Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline. *Genome Biol.* **2019**, *20* (1), 275. <https://doi.org/10.1186/s13059-019-1905-y>.
- (28) Quinlan, A. R.; Hall, I. M. BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* **2010**, *26* (6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- (29) Wickham, H. Introduction. In *ggplot2: Elegant Graphics for Data Analysis*; Wickham, H., Ed.; Springer: New York, NY, 2009; pp 1–7. https://doi.org/10.1007/978-0-387-98141-3_1.
- (30) Nicolaou, K. C.; Adsool, V. A.; Hale, C. R. H. An Expedient Procedure for the Oxidative Cleavage of Olefinic Bonds with PhI(OAc)₂, NMO, and Catalytic OsO₄. *Org. Lett.* **2010**, *12* (7), 1552–1555. <https://doi.org/10.1021/ol100290a>.

- (31) Rabe, P.; Schmitz, T.; Dickschat, J. S. Mechanistic Investigations on Six Bacterial Terpene Cyclases. *Beilstein J. Org. Chem.* **2016**, *12* (1), 1839–1850. <https://doi.org/10.3762/bjoc.12.173>.
- (32) Fehr, C.; Winter, B.; Magpantay, I. Synthesis of (–)-Cubebol by Face-Selective Platinum-, Gold-, or Copper-Catalyzed Cycloisomerization: Evidence of Chirality Transfer and Mechanistic Insights. *Chem. – Eur. J.* **2009**, *15* (38), 9773–9784. <https://doi.org/10.1002/chem.200901292>.
- (33) Friedel, H. D.; Matusch, R. Neue Aromadendran-Derivate Aus Tolu-Balsam. *Helv. Chim. Acta* **1987**, *70* (7), 1753–1759. <https://doi.org/10.1002/hlca.19870700711>.
- (34) Faraldos, J. A.; O’Maille, P. E.; Dellas, N.; Noel, J. P.; Coates, R. M. Bisabolyl-Derived Sesquiterpenes from Tobacco 5-Epi-Aristolochene Synthase-Catalyzed Cyclization of (2Z,6E)-Farnesyl Diphosphate. *J. Am. Chem. Soc.* **2010**, *132* (12), 4281–4289. <https://doi.org/10.1021/ja909886q>.
- (35) Wei, X.; Ning, W.; McCadden, C. A.; Alsup, T. A.; Li, Z.; Łomowska-Keehner, D. P.; Nafie, J.; Qu, T.; Opoku, M. O.; Gillia, G. R.; Xu, B.; Icenhour, D. G.; Rudolf, J. D. Exploring and Expanding the Natural Chemical Space of Bacterial Diterpenes. *Nat. Commun.* **2025**, *16* (1), 3721. <https://doi.org/10.1038/s41467-025-57145-6>.
- (36) Weinheimer, A. J.; Washecheck, P. H.; Helm, D. van der; Hossain, M. B. The Sesquiterpene Hydrocarbons of the Gorgonian, *Pseudopterogorgia Americana*, the Nonisoprenoid β -Gorgonene. *Chem. Commun. Lond.* **1968**, No. 18, 1070–1071. <https://doi.org/10.1039/C19680001070>.
- (37) Coombe, L.; Warren, R. L.; Wong, J.; Nikolic, V.; Birol, I. ntLink: A Toolkit for De Novo Genome Assembly Scaffolding and Mapping Using Long Reads. *Curr. Protoc.* **2023**, *3* (4), e733. <https://doi.org/10.1002/cpz1.733>.
- (38) Steele, T. S.; Burkhardt, I.; Moore, M. L.; de Rond, T.; Bone, H. K.; Barry, K.; Bunting, V. M.; Grimwood, J.; Handley, L. H.; Rajasekar, S.; Talag, J.; Michael, T. P.; Moore, B. S. Biosynthesis of Haloterpenoids in Red Algae via Microbial-like Type I Terpene Synthases. *ACS Chem. Biol.* **2024**, *19* (1), 185–192. <https://doi.org/10.1021/acscchembio.3c00627>.

4.10 Acknowledgements

Chapter 4 contains material currently being prepared for publication, co-authored with Minich, J.J., Allsing, N., Michael, T.P., Moore, B.S., Burkhardt, I. The dissertation author was a primary researcher and author of this chapter.