UNIVERSITY OF CALIFORNIA SAN DIEGO

# The Genetic Basis for Domoic Acid Biosynthesis

A dissertation submitted in partial satisfaction of the
Requirements for the Degree Doctor of Philosophy

in

Marine Biology

by

John Kilpatrick Brunson

Committee in charge:

Professor Andrew E. Allen, Co-Chair
Professor Bradley S. Moore, Co-Chair
Professor Eric E. Allen
Professor Lihini I. Aluwihare
Professor James W. Golden

2021

The dissertation of John Kilpatrick Brunson is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

Dedicated to my father, John M. Brunson,
who grew up watching The Undersea World of Jacques Cousteau on TV, dreamed of being a marine biologist, introduced me to SCUBA and to the ocean,
and sacrificed so much to give me the freedom to follow my dreams

WE DID IT!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to extend my sincerest appreciation to my advisors, Bradley S. Moore and Andrew E. Allen. Both of you have supported me endlessly throughout my graduate career and encouraged me to tackle difficult problems, apply myself fully to my academic pursuits, and explore the exciting worlds of chemistry, genomics, phycology, and marine studies. What a fun ride it was!

In the Moore lab, I want to thank two postdoctoral scholars who have since moved on to professorship and new horizons: Shaun McKinnie and Jonathan Chekan. Shaun, thanks for teaching me just about everything I now know about chemistry – it was a pleasure to explore domoic acid biosynthesis with you. Jon, thanks for improving my confidence at the protein biochemistry bench. I have very much enjoyed our ongoing collaborations. Special thanks also go out to Zachary Miles, who helped me purify my first enzymes, and Taylor Steele, with whom I had the pleasure to explore red algae genomics.

In the Allen lab at the J. Craig Venter Institute, I want to extend a huge thank you to Hong Zheng. Hong, you make it possible for all of us to do our best work in the lab and your continued guidance has been extremely valuable. I'd also like to thank Ariel Rabines for all of his help with sequencing prep and teaching me how to set up 96 PCR reactions at once. I want to extend a huge thank you to Vincent Bielinski, who taught me how to clone quickly and with confidence. Lessons like "cloning should be easy" and "the faster you fail, the faster you succeed" will stick with me forever.

Outside of San Diego, I'd like to thank Jason Smith for his never-ending support on all of my *Pseudo-nitzschia* projects. Thanks for being a great host when I came up to Monterey Bay as a highly intimidated first year graduate student. Big thanks to John Ryan at the Monterey Bay

Aquarium Research Institute for helping me to understand oceanographic data – all of those lessons from Physical Oceanography class finally made sense when I could see the data for myself in the context of my own projects! Big thanks to Mari Yotsu-Yamashita and Yukari Maeno at Tohoku University, Japan for the chance to establish worldwide connections and collaborations to explore domoic acid in red algae.

I wouldn't have been able to do any of this work without the unending support of my family. Mom and dad – I know it was hard to have your only child move all the way across the country to do his doctorate degree. Thanks for always believing in me and supporting me – the best is still yet to come. Finally, I want to thank my wonderful wife Sarah Snook Brunson. I met you maybe three weeks after moving to San Diego. You saw me all the way through my doctorate degree, you were here for ALL of it. Thank you for all the love, support and encouragement. I can't wait for our future together. I love you and our silly pets, Zoe and Katie.

The main text and supplemental information sections of Chapter 2 are reprints of the material as it appears in "**J. K. Brunson***, S. M. K. McKinnie*, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. Smith, D. A. Hutchins, A. E. Allen, B. S. Moore. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* 361(6409), 1356-1358 (2018)." The dissertation author was one of two equally contributing primary investigators and authors of this manuscript.

The main text and supplemental information sections of Chapter 3 have been submitted for publication as "T. S. Steele*, **J. K. Brunson***, Y. Maeno, R. Teruda, A. E. Allen, M. Yotsu-Yamashita, J. R. Chekan, B. S. Moore. Domoic acid biosynthesis in red alga *Chondria armata* suggests a complex evolutionary history for toxin production (2021) Manuscript in submission."

The dissertation author was one of two equally contributing primary investigators and authors of this manuscript.

The main text and supplemental information sections of Chapter 4 are currently in preparation for manuscript submission as "**J. K. Brunson**, A. J. Rabines, H. Zheng, P. Venepally, J. P. Ryan, B. S. Moore, G. J. Smith, A. E. Allen. Metatranscriptomic sequencing and barcoding of a major harmful algal bloom event in Monterey Bay, California. (2021) Manuscript in preparation." The dissertation author was the primary investigator and author of this manuscript.

VITA

| | |
|---|---|
| **2014** | Bachelor of Science, Genetics, University of Georgia |
| **2015-2021** | Graduate Student Researcher, Scripps Institution of Oceanography, University of California San Diego |
| **2015-2016** | Dickinson-McCrink Fellow, Scripps Institution of Oceanography, University of California San Diego |
| **2016-2018** | National Institutes of Health (NIH) Chemistry and Biology Interface (CBI) Fellowship Trainee, project site University of California San Diego |
| **2019** | Teaching Assistant, Marine Biochemistry, University of California San Diego |
| **2019** | Student Representative for University of California San Diego at Graduate Student Research Day at Sacramento Capitol (March 19th 2019) |
| **2019** | Edward A. Frieman Prize for Excellence in Graduate Student Research, Scripps Institution of Oceanography, University of California San Diego |
| **2019** | Claude E. Zobell fellowship awardee, Scripps Institution of Oceanography, University of California San Diego |
| **2019-2021** | National Institutes of Health (NIH) Ruth L Kirschstein Predoctoral Award (F31ES030613) |
| **2021** | Doctor of Philosophy, Marine Biology, Scripps Institution of Oceanography, University of California San Diego |

**Publications**

| | |
|---|---|
| **2015** | B. J. Karas, R. E. Diner, S. C. Lefebvre, J. McQuaid, A. P. Phillips, C. M. Noddings, **J. K. Brunson**, R. E. Valas, T. J. Deerinck, J. Jablanovic, J. T. F. Gillard, K. Beeri, M. H. Ellisman, J. I. Glass, C. A. Hutchinson III, H. O. Smith, J. C. Venter, A. E. Allen, C. L. Dupont, P. D. Weyman. Designer diatom episomes delicered by bacterial conjugation. *Nat. Commun*. **6**, 6925 (2015). |
| **2016** | **J. K. Brunson**, J. Griffith, D. Bowles, M. E. Case, J. Arnold. *lac-1* and *lag-1* with *ras-1* affect aging and the biological clock in *Neurospora crassa*. *Ecol. Evol.* **6**, 8341–8351 (2016) |
| **2018** | **J. K. Brunson***, S. M. K. McKinnie*, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. |

Smith, D. A. Hutchins, A. E. Allen, B. S. Moore. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018)

**2021** J. Turnšek, **J. K. Brunson**, M. D. Viedma, T. J. Deerinck, A. Horák, M. Oborník, V. A. Bielinski, A. E. Allen. Proximity proteomics in a marine diatom reveals a putative cell surface-to-chloroplast iron trafficking pathway. *Elife* **10**, e52770 (2021)

ABSTRACT OF THE DISSERTATION

The Genetic Basis for Domoic Acid Biosynthesis

by

John Kilpatrick Brunson

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2021

Professor Andrew E. Allen, Co-Chair

Professor Bradley S. Moore, Co-Chair

Domoic acid (DA), the causative agent of amnesic shellfish poisoning in humans, is a potent neurotoxin produced by algal blooms of the diatom genus *Pseudo-nitzschia*. Due to its bioaccumulation in shellfish, DA production must be closely monitored in the marine environment. Despite extensive research on *Pseudo-nitzschia* toxicity, the basic biology underlying DA production has remained unclear. Specifically, a genetically encoded biosynthesis for DA has eluded chemists and algal bloom researchers. Such genetic information

would provide a genetic basis for monitoring toxin production in the environment. This dissertation describes the discovery of the DA biosynthesis (*dab*) genes and subsequent application to environmental monitoring. Chapter 2 details the initial discovery of the *dab* genes in *Pseudo-nitzschia multiseries* using comparative transcriptomics to identify candidate genes upregulated under DA-inducing culturing conditions. Subsequent validation of enzyme activity *in vitro* establishes the chemical transformations required to build the key structural features of DA. Chapter 3 describes the identification of *dab* gene homologues in the red macroalgae *Chondria armata*, a DA-producing seaweed from which the toxin was first isolated. The red algal DA (*rad*) biosynthesis genes uncovered here suggest a biosynthetic pathway consistent with the *Pseudo-nitzschia* pathway with only slight modifications to enzyme activity *in vitro*. Similarities and differences between the *rad* genes and diatom *dab* genes suggest a complex evolutionary history for toxin production in the two distantly related organisms. Finally, Chapter 4 describes a molecular approach to *Pseudo-nitzschia* and DA monitoring, focusing on the 2015 North American West Coast *Pseudo-nitzschia australis* bloom. In this study, RNA barcoding and metatranscriptomics datasets were generated from weekly phytoplankton net tow samples taken from Monterey Bay, California, a hotspot for DA production during the 2015 bloom. Sequence barcoding improved description of *Pseudo-nitzschia* species composition in the context of the larger microbial community. Meanwhile, metatranscriptomics enabled the identification of *dab* transcripts simultaneous with detection of DA and toxic species, and also provided insight into the shifting physiology of the *P. australis* bloom throughout its lifespan. The combination of efforts describes herein establishes a paradigm for the genetic monitoring of DA producing algal blooms.

# Chapter 1: Introduction to the Dissertation

## 1.1 Domoic acid (DA), a Harmful Algal Bloom (HAB) neurotoxin

### 1.1.1 A Brief Overview of HABs and Discovery of their Toxins

Harmful agal booms (HABs) pose a persistent threat to human and environmental health. A HAB occurs when a given type of algae, often a single genus or species, grows far beyond its normal bounds to dominate the local or regional ecosystem. While overgrowths of multicellular seaweeds such as *Sargassum* can certainly wreak environmental havoc in their own right, the vast majority of HAB-forming species are prokaryotic or eukaryotic unicellular algae.[1,2] These microalgal overgrowths, occurring in both freshwater and marine environments, earn their "harmful" title by depleting local ecosystems of key nutrients and creating anoxic zones following their ultimate demise and decay. Many HAB-forming species also produce potent molecular toxins that accumulate in higher trophic levels and can poison fish, birds, and mammals.[3,4] Because of the broad-reaching negative impacts exerted by HABs, significant study has been devoted to the environmental factors contributing to the formation of these frequently toxic biological events.

Although many different types of diverse microalgae cause both freshwater and marine HAB events, both are typically fueled by an influx of nutrients to the aquatic ecosystem. Many freshwater HABs, commonly comprised of prokaryotic "blue-green" cyanobacteria, flourish because of anthropogenic nutrient input from wastewater runoff and fertilizer.[5] Meanwhile, the seasonal marine HABs occurring along the world's coastlines are natural events fueled by wind-driven upwelling of nutrient-rich water from the ocean depths. A plethora of stramenopile (ie: diatoms, *Aureococcus*, *Chattonella*) and alveolate (ie: dinoflagellates) eukaryotic algae make up the diversity of marine HABs.[2] Presently, it remains unclear whether or not seasonal marine HAB events will increase in frequency and severity when faced with the impacts of climate

change on the world's oceans over the coming decades.[6–9] Recent oceanic temperature anomalies have impacted the range and strength of notable bloom events, and as such, it is important to understand emerging sources of HABs and their toxicity in an effort to protect human health worldwide.[10,11]

Humans typically encounter HAB toxins in the form of contaminated seafood. As a result, many of the pathologies caused by HAB toxins are referred to as "shellfish poisonings." Maladies such as amnesic shellfish poisoning, paralytic shellfish poisoning and diarrhetic shellfish poisoning are characterized by acute neurotoxic effects and frequently involve gastrointestinal distress.[12,13] In severe cases, these shellfish poisoning syndromes can impart chronic neurological symptoms and even prove fatal. Regardless of their presentation, each of the different shellfish poisoning syndromes can be linked back to consumption of a discrete molecular toxin produced by algae. The potent bioactivity of these toxins has accordingly inspired their longstanding study in chemistry, biology, and medicinal sciences. For example, various ion channels and receptors have been discovered and further described thanks to studies aiming to elucidate mechanisms of action and molecular targets of HAB toxins.[14–16]

While the fascinating molecules produced by HABs have piqued the curiosity of modern science in recent decades, their toxic effects have been understood by humans for centuries. Indigenous peoples living along the North American West Coast were intimately aware of the link between nighttime bioluminescent displays, now known to be caused by algae, and toxic shellfish that was unsafe to eat.[17] One of the first recorded European encounters with HAB-tainted seafood comes from the logbooks of Captain George Vancouver.[18] Sometime in June of 1793, several of Capt. Vancouver's crew members harvested and consumed mussels from one of the inlets along the coast of modern-day British Columbia, Canada. These crewmen were very

soon “seized with a numbness about their faces and extremities, (and) their whole bodies were very shortly affected in the same manner, attended with sickness and giddiness.” One crew member was so afflicted that he became unable to swallow water and before long had died as a result of severe illness. Capt. Vancouver was convinced that the source of toxicity was in the mussels, writing in his logbook, “There was no doubt that this was occasioned by a poison contained in the mussels he had eaten about eight o'clock in the morning; at nine he first found himself unwell, and died at half past one.”

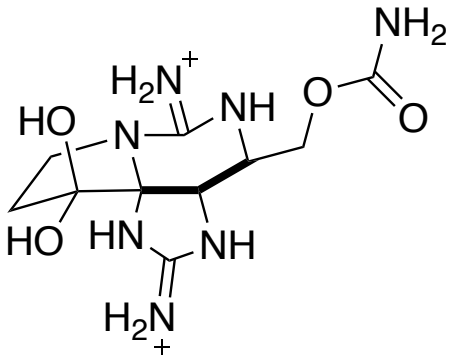



**Figure 1.1** Structure of saxitoxin, paralytic shellfish poison and likely cause of illness among Capt. Vancouver’s crew.

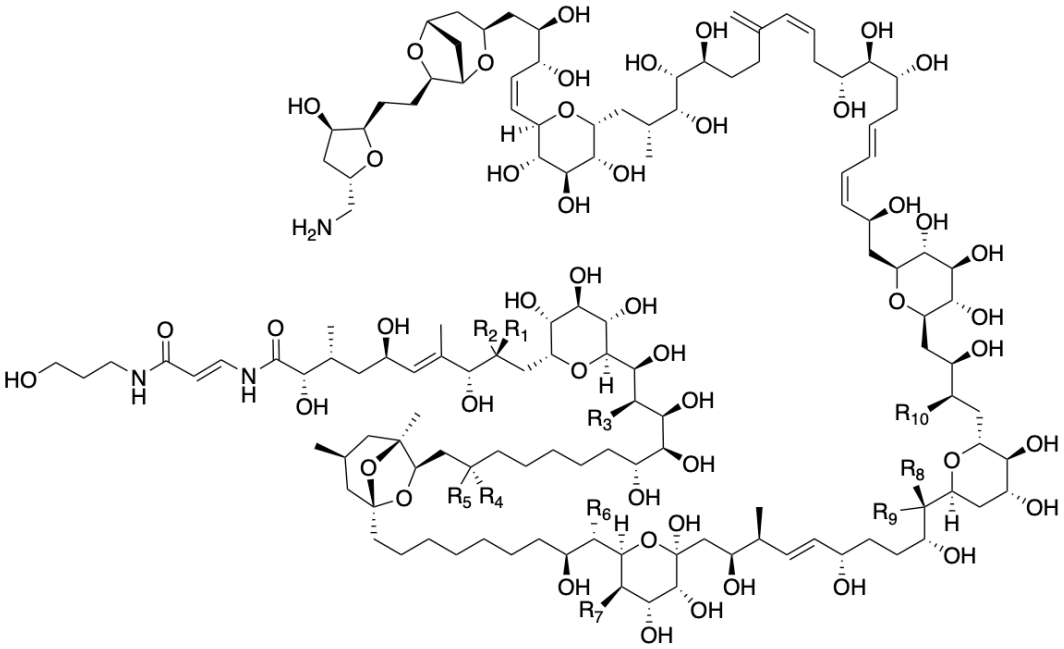
The mussel contaminant that caused the crewmember’s illness and subsequent death was almost certainly a molecule called saxitoxin. Today, saxitoxin is well-understood as a potent neurotoxin and the causative agent of paralytic shellfish poisoning, or PSP. Saxitoxin was first isolated from Alaskan butter clams (*Saxidomus gigangteus*) in 1957.[19] However, it would be nearly two decades before a concrete molecular structure could be elucidated for the toxin. Assigning the correct structure to saxitoxin was challenging due to its highly polar nature and recalcitrance to crystallization, a prerequisite for the x-ray diffraction studies that allow chemists to visualize the three-dimensional structure of a molecule. Nevertheless, the crystallographic structure for saxitoxin was unveiled by two independent groups in 1975, revealing a highly

unusual, nitrogenous molecule with a unique tricyclic core that had not been observed previously.[20,21] Sure enough, this molecule proved exceedingly toxic, with a high propensity for blocking the voltage-gated sodium channels used to generate action potentials in neurons.[22] Two classes of bloom-forming harmful algae are known make saxitoxin: cyanobacteria such as *Dolichospermum* and *Cylindrospermopsis* as well as dinoflagellates including the widespread genus *Alexandrium*.[23]

Another marine toxin with a storied history comes from the Hawaiian island of Maui. According to local legend, fishermen from the village at Hana were plagued by bad luck and frequently lost men at sea. Upon reaching the shore after yet another ill-fated seafaring expedition, the remaining fishermen took out their rage on a local, hunchbacked hermit. Perhaps this was not the best idea, as the hermit turned out to be a shark god, whom the fishermen nevertheless killed and tossed into a local tidepool. Now sufficiently cursed, the tidepool began to grow algae so toxic that the locals named it "limu-make-o-Hana", or "deadly seaweed from Hana."[24]

When chemists finally got their hands on the "limu," they found that it was not a seaweed but was instead a soft coral invertebrate called a zoanthid. Nevertheless, this organism harbored an exceptionally toxic molecule that was first isolated in 1971 by Dr. Richard E. Moore and Dr. Paul J. Scheuer.[25] This molecule was named palytoxin after the scientific name for the zoanthid genus *Palythoa* that "limu-make-o-Hana" belongs to. Over ten years later, the exact structure of palytoxin was elucidated independently and nearly simultaneously by Dr. Moore and by another group led by Dr. Daisuke Uemura in Japan, a monumental achievement given the sheer structural complexity and size of the molecule (Fig 1.2).[26,27] Further toxicity studies demonstrated that this molecular behemoth was just as toxic as the Hawaiian legends claimed, targeting the ATP-

dependent sodium/potassium pumps that create the ionic gradients critical for maintaining cellular homeostasis and supporting life.[28] Although the presence of palytoxin in *Palythoa* zoanthids is well-appreciated, chemists and biologists have suggested that microorganisms, such as bacteria or algae, are likely the true producers of the compound.[24] Notably, dinoflagellates belonging to the genus *Ostreopsis* are known producers of ovatoxin, a molecule with striking structural similarity to palytoxin (Fig 1.2).[29] Blooms of *Ostreopsis* are common in the Mediterranean off the Italian coast, and an ovatoxin-producing bloom of this dinoflagellate may have contributed to cases of respiratory distress in Genoa during the summer of 2005.[24]



| | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_7$ | $R_8$ | $R_9$ | $R_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Palytoxin: | OH | H | OH | $CH_3$ | H | H | OH | OH | H | OH |
| Ovatoxin-*a*: | H | OH | H | H | $CH_3$ | OH | H | H | OH | H |

**Figure 1.2.** Structure of palytoxin, originally isolated from "limu-make-o-Hana", and close structural homologue ovatoxin-a, from *Ostreopsis* dinoflagellates.

### 1.1.2 The Surprising Culprit Behind a New Shellfish Poisoning Syndrome

In 1987, a new type of shellfish poisoning syndrome was identified in individuals from Prince Edward Island (PEI), Canada. After consuming mussels harvested from PEI, many people developed severe neurological symptoms, typical for a bout of shellfish poisoning. However, these neurological symptoms were unique from the well-studied effects of PSP linked to saxitoxin, with a dominant symptom being a notable loss of short-term memory rather than paralysis. This novel syndrome was therefore named Amnesic Shellfish Poisoning (ASP), and a total of 107 people contracted ASP from the PEI mussels in 1987, including 3 fatalities.[30] The United States Food and Drug Administration (FDA) promptly warned consumers to avoid eating mussels sourced from PEI, and the incident left Canadian fisheries scientists scrambling to find a molecular agent that could be linked to the symptoms of ASP.

After convening to solve the problem of the toxic PEI mussels, it took scientists a mere five days to identify the structure of the contaminating molecule in the tainted seafood. The effort, led by chemist Dr. J.C. Wright identified a known molecule, domoic acid (DA), as the probably toxic agent.[31] The relatively rapid identification of DA from the PEI mussels stands in stark contrast to the storied histories surrounding the isolation and structural elucidation of saxitoxin and palytoxin, wherein a familiar source of toxicity led scientists to excitingly potent molecules with never-before-seen structures. This time, when faced with a new variety of shellfish poisoning, scientists narrowed in on a molecule that had already been described in the chemical literature. Before its identification in PEI mussels, DA was a relatively obscure chemical product that had been described approximately 30 years prior from extracts of the red macroalgae *Chondria armata*.[32,33] The discovery of DA in *C. armata* was made by Japanese scientists, including chemist Dr. Koji Daigo, studying the chemical constituents of various

seaweeds. However, the presence of DA in PEI shellfish remained unexplained until researchers noticed an unusually high abundance of diatoms, a variety of stramenopile eukaryotic microalgae characterized by silica-laden exteriors, within the mussels. These diatoms, initially identified as *Nitzschia pungens*, were verified as DA producing microalgae shortly after.[34] After some time, *Nitzschia pungens* was re-classified as a member of a new diatom genus, *Pseudo-nitzschia*, and since then DA-producing species of *Pseudo-nitzschia* have been described around the world.[35,36]

### 1.1.3 From Natural Medicine to Neurotoxin: DA, the Amnesic Shellfish Poison

The apparent toxicity of DA, the causative agent of ASP, was almost as surprising as its presence in PEI mussels as a diatom-produced metabolite. The original source of DA in the macroalga *C. armata*, has its own interesting history, notable not for its toxicity but for its medicinal properties. Seaweeds such as *C. armata*, referred to in Japan as "domoī", were of particular interest to chemists due to their use as natural remedy for parasitic intestinal worms in Japanese history. References to the anthelmintic properties of Japanese seaweeds can be found in recorded history from over a thousand years ago.[37] After DA was purified from *C. armata*, it was indeed found to possess such an anti-parasitic bioactivity.[33] The activity exhibited by DA was similar to that of the structurally related molecule kainic acid, discovered just a few years prior in the red macroalgae *Digenea simplex*, a seaweed called "kaininsō" in Japan, that was particularly well-known as an anthelmintic agent for centuries.[38] Kainic acid (KA) has even been used in combination therapies treating infections implicating parasitic worms of the *Ascaris* genus [39]. However, the acute neurotoxicity of these molecules was not appreciated until the 1990s when DA and KA had been well-characterized as potent agonists of ionotropic glutamate receptors.

Both DA and KA bear strong structural homology to glutamic acid, an amino acid and key neurotransmitter implicated in the proper functioning of the central nervous system in vertebrates. Receptors for glutamate in neurons, known as ionotropic glutamate receptors (iGluRs), are key players in signal transduction across neuronal synapses. Binding of glutamate activates the receptor and opens an ion channel, allowing passage of sodium ions to enable downstream signal transduction and synaptic transmission.[40] Like glutamate, DA and KA are also capable of binding to iGluRs, but both tend to do so at a much lower concentration than the native neurotransmitter and stay bound to the receptor for a longer period of time.[41] This tight association with glutamate receptors and prolonged receptor activation is the mechanism by which DA exerts its toxicity, effectively holding open the associated ion channel and resulting in mass influx of sodium ions into the neuron, a phenomenon that typically results in neuronal swelling and apoptosis. The amnesic properties of DA poisoning are well explained by this mechanism, as iGluR plays a key role in synaptic plasticity and is thought to be implicated in learning and memory.[40] While DA is the more potent of the two molecules, KA has been used extensively in neuroscience research to study neurodegenerative diseases such as Alzheimer's in biological models.[42]

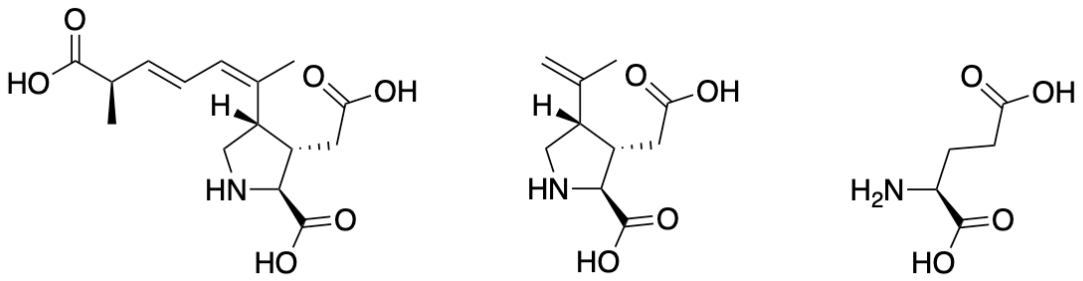



**Figure 1.3.** Structures of domoic acid (DA, left), kainic acid (KA, center) and L-glutamate (right).

The impacts of DA toxicity reach far beyond the acute poisoning exhibited by cases of ASP in humans. Chronic consumption of seafood laden with DA has been linked to decreased short term memory and cognitive deficits.[43,44] Additional studies have also suggested that DA may also impair proper kidney functioning and neonatal development in mammals.[45,46] Acute DA toxicity is also not limited to humans, with marine mammals being one of the most commonly impacted organisms. Poisoning with DA has been linked to an increased sea lion mortality events, and the amnesic effects of the neurotoxin create an increased incidence in stranding marine mammals, including whales.[47,48] Sea birds are also affected by DA poisoning, with many pelicans and Brandt's cormorants falling victim to the toxin in Monterey Bay in the early 1990s.[49] As such, algal bloom producing high quantities of DA have a wide range of potential ecological and human health impacts.

## 1.2 Production of DA by a cosmopolitan, HAB-forming diatom genus

### 1.2.1 *Pseudo-nitzschia* in our world ocean

The discovery of DA production by the diatom *Pseudo-nitzschia* came as a bit of a surprise. Up until that point, most HAB species described belonged to either cyanobacteria or dinoflagellates. As such, diatoms were not recognized as a potentially toxic organism. Diatoms are ubiquitous eukaryotic phytoplankton in the world's oceans with one of their most recognizable features being an external cell wall comprised of biomineralized silica, a cellular structure called a "frustule."[50] These microalgae are also key players in carbon fixation in marine environments, contributing an estimated twenty percent of biological primary productivity in the ocean.[51] Diatoms are especially prolific in nutrient rich, upwelling regimes where they form the base of the marine food web, providing key biomass to drive major fisheries

around the world.[52] The "bloom-and-bust" cycles of diatom activity fuels major geochemical cycles wherein sinking diatom frustules export carbonate and silica to the seafloor, an effective method of natural carbon sequestration.[53,54] Diatoms inhabit a wide variety of marine environments, ranging from nutrient-rich coastal regions, the nutrient-poor open ocean and even in the Arctic and Antarctic.[51]

**Figure 1.4.** Overlapping chains of *Pseudo-nitzschia spp.* diatoms, photo courtesy of Dr. G. Jason Smith (Moss Landing Marine Laboratories).

Diatoms belonging to *Pseudo-nitzschia*, the genus primarily responsible for oceanic DA production and the culprit behind the 1987 PEI incident, are found in all of the world's oceans. In fact, DNA metabarcoding studies performed on samples from the transoceanic Tara Oceans expedition revealed that *Pseudo-nitzschia* was one of the most cosmopolitan and diverse genera of diatoms, being not only present but also abundant in all sequenced samples.[55] Indeed, individual species of *Pseudo-nitzschia* have been isolated from marine samples the world over, with 52 species described as of 2018 and new species being discovered regularly.[36,56] Of these

described species, 26 are known to produce DA in culture. Species capable of producing DA can also be found worldwide, although no toxicity has been demonstrated thus far for the Antarctic species *P. antarctica* and *P. subcurvata*.[36] Notably, *Pseudo-nitzschia* is unique in its ability to produce DA. Only two species outside of the *Pseudo-nitzschia* genus have been described to produce the neurotoxin, with both species currently classified within the genus *Nitzschia*. These species, *N. bizertensis* and *N. navis-varingia*, are not bloom forming diatoms and typically inhabit brackish waters.[57,58] Furthermore, these diatoms don't appear to be closely related to other *Nitzschia* species according to recent molecular phylogenetics studies, making their classification a bit of a mystery.[59] No other diatom genus has been reported to produce DA in culture or in the broader marine environment.

Ever since *Pseudo-nitzschia* was identified as a toxin producing HAB organism, substantial research effort has gone into understanding the oceanographic conditions that cause *Pseudo-nitzschia* to form large blooms and produce high quantities of DA. As it turns out, it has been rather difficult to connect active DA production to a discrete environmental stimulus. Culturing conditions ranging from light irradiance and nutritional limitation to growth phase and inclusion of specific associated bacteria have all been connected to increased DA production in some form or another, although heightened cellular stress seems to be a common theme among those conditions that increase DA production.[35,60] Generally, *Pseudo-nitzschia* appears to thrive in coastal environments following nutrient addition via seasonal upwelling. Recent studies have demonstrated that *Pseudo-nitzschia* has genus-specific adaptations for dealing with chronic, low levels of iron in the marine environment and can outcompete other bloom forming diatoms once iron is pulsed into the system.[61] Indeed, *Pseudo-nitzschia* diatoms exhibit the strongest response

and growth following open ocean iron fertilization experiments in high-nitrate, low-chlorophyll regions where iron is the key limiting micronutrient.[62,63]

### 1.2.2 The 2015 *Pseudo-nitzschia* HAB spans the North American West Coast

While blooms of DA-producing *Pseudo-nitzschia* have been reported from around the world, perhaps none have been as catastrophic as the HAB event that took place on the North American West Coast during the spring and summer of 2015.[11] Beginning in April 2015, a nearly monospecific bloom of *P. australis* came to dominate the West Coast, spanning the coastline from Point Conception, just north of Los Angeles, and northward to the Aleutian Islands of Alaska. The bloom persisted throughout the summer for most of the coastline, with some locations still experiencing a toxic HAB event during fall and winter of 2015. This gigantic *P. australis* HAB produced record-setting levels of DA, prompting the closure of various fisheries in California, Oregon, Washington State, and British Colombia. Affected fisheries included Dungeness crab, razor clams, mussels and other bivalves, and many of these fisheries were not able to open until early 2016.[11] The economic losses as a result of this event were humongous and difficult to fully estimate, with United States Dungeness crab fishery alone accounting for nearly $100 million in economic damages. Fishing communities were hit especially hard and economic damages rippled throughout these communities, extending from fishing operations to the hospitality industry.[64]

The 2015 *P. australis* bloom was unprecedented in its range and toxicity. But what factors led to the extreme distribution of this toxic HAB event? While blooms of *P. australis* are common on the coast of California, HAB scientists believe that a persistent oceanographic anomaly in the northeast Pacific increased the habitable range of *P. australis* northward along

the North American coastline.[11] This warm-water anomaly, termed "the Blob", persisted from late 2013 through the end of 2015 off the coast of Alaska, warming the North Pacific and allowing the highly-toxic *P. australis* to migrate into regions beyond its normal range.[10,65] Additionally, strong seasonal upwelling began nearly simultaneously along the entire coastline, bringing nutrient rich waters up from the ocean depths to the surface.[11] Local nutrient anomalies may have further impacted the toxicity of the *P. australis* bloom. In Monterey Bay, California, exceedingly high DA concentrations resulting from the 2015 bloom were coincident with historically low silica concentrations, one of the many nutritional limitations known to stimulate DA production in culture.[66] While many explanations now exist for the range and toxicity of this monstruous bloom event, predicting the scope of the 2015 HAB ahead of time would have been exceedingly difficult. However, predictive monitoring of toxic bloom events remains one of the biggest goals in the HAB community.

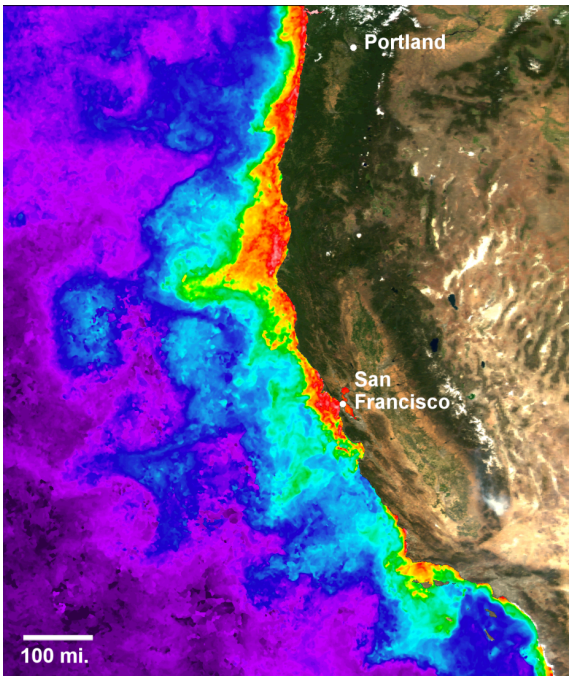



**Figure 1.5.** Extent of the Pacific Northwest *Pseudo-nitzschia* bloom in late-spring 2015. Image from Brunson et al 2018.

### 1.3 Next-generation HAB monitoring using key toxin biosynthesis genes

#### 1.3.1 Current Monitoring Approaches

The initiation and development of HAB events is currently monitored in a variety of different ways. Local ecosystem monitoring involves the analysis of coastal water samples for the proliferation of potentially toxic species and the presence of various HAB toxins. In California, these activities are accomplished by the California Harmful Algal Bloom Monitoring and Alert Program (CalHABMAP), wherein weekly water samples are taken from piers across the state.[67] In this sampling approach, HAB-causing algae are enumerated by light microscopy and toxin analysis is performed using analytical chemistry methods such as liquid chromatography-mass spectrometry (LC-MS). At the regional scale, wind and nutrient data are extremely useful for detecting the onset of seasonal upwelling events and predicting their impacts on phytoplankton growth.[11] Satellite imagery is commonly used to measure decreased sea surface temperature linked to active upwelling of cold, nutrient rich deep water.[68] Additionally, satellite measurement of chlorophyll concentration can identify active phytoplankton growth on the regional scale using detection modules such as the Visible Infrared Imaging Radiometer Suite (VIIRS).[69] The synthesis of these various data points have informed current predictive monitoring approaches that have proved valuable for forecasting potential HAB events.[70,71] However, further refinement of HAB forecasting models and methods is required to increase the accuracy of predictive monitoring.

Recent years have witnessed the implementation of genetic approaches to enable rapid identification of HAB species. Quantitative PCR (qPCR) is one such approach used to target the amplification of diagnostic, species-specific sequences to enable the detection and quantification of diverse HAB species in the environment.[72] Sandwich hybridization assays (SHAs) represent a

similar, powerful approach for the quantification of a given nucleotide sequence in a complex sample. The detection and quantification of *Pseudo-nitzschia* using SHA probes has been well-developed and can be performed on the benchtop or *in situ* onboard autonomous underwater vehicles (AUVs) and environmental sample processors (ESPs).[73] Both of these methodologies are more rapid and precise than traditional light microscopy-based approaches for the detection of potentially toxic organisms. However, the process of simply identifying HAB species does not directly address the question of active toxin production in the marine environment.

### 1.3.2 Application of HAB toxin biosynthesis to genetic monitoring approaches

In order to use environmental genetics to detect the capacity for HAB toxin production, the genes encoding the biochemical pathway for toxin synthesis must be identified and understood. Such a biosynthesis of a given natural compound is carried out by a cascade of enzymatic reactions to convert central metabolic precursors, such as amino acids or fatty acid building blocks, into the mature specialized compound. Due to the structural novelty afforded by HAB toxins, their representative biosynthetic pathways have been the source of intense interest by chemists and geneticists for decades, with the ultimate goal being the ability to connect molecules back to the genes encoding the biosynthetic pathway. However, this feat has been accomplished for only a select few HAB toxins, with the majority being of unknown biosynthetic origin.[74,75] The limited information available on HAB toxin biosynthetic genes is due in part to the difficulties inherent in microalgal genomics.

Saxitoxin is one such HAB toxin with a known genetic basis for biosynthesis.[76] The biosynthetic genes encoding toxin production were identified by Dr. Brett Neilan and co-workers studying the saxitoxin producing cyanobacteria *Cylindrospermopsis* using adaptor mediated PCR

to explore the region around a previously identified gene thought to be involved in biosynthesis.[77] Their efforts identified a discrete gene cluster present in the genome of *Cylindrospermopsis* that included 26 genes of putative biosynthetic function.[78] Clustering of biosynthetic genes is a common phenomenon in bacterial genomes.[79] More recent biochemical analysis has assigned discrete biochemical transformations to several of the genes in the cluster.[80–82] Additionally, several homologs to the *Cylindrospermopsis* genes can be found in dinoflagellates, suggesting a common biosynthetic pathway shared between the two divergent clades.[83,84] This genetic information has been implemented in qPCR-based environmental genetics approaches currently being developed for the monitoring of saxitoxin-producing HABs.[72,85]

Much of the progress on saxitoxin biosynthesis was made possible by the virtue of its presence in cyanobacterial genomes. Unlike bacterial genomes, dinoflagellate genomes are massive, containing up to 245,000,000,000 nucleotide base pairs (245 Gbp). Dinoflagellate genomes also contain no obvious gene clustering and instead feature highly repetitive regions that are difficult to sequence using modern genome sequencing approaches.[86,87] No biosynthetic genes have been identified for strictly dinoflagellate HAB toxins, and structurally complex molecules like palytoxin have evaded meaningful biosynthetic interrogation.

Meanwhile, the field of diatom genomics has exploded over the past 15 years. Sequenced diatom genomes range in size from around 25 Mbp to nearly 100 Mbp and contain considerably less repetitive elements than dinoflagellate genomes. Many diverse diatoms have had their genomes sequenced, including model organisms, open ocean diatoms, Antarctic diatoms and oleaginous diatoms with biofuels applications.[88–93] Two publicly available *Pseudo-nitzschia* genomes also exist for *P. multiseries* and *P. multistriata*, both of which have the

capacity to produce DA.[94] Beyond strictly genomic descriptions of diatoms, a variety of diatom transcriptomes have been sequenced to understand the active transcriptional activity of diatom genomes.[95] These transcriptomics studies have unraveled a wealth of information surrounding diatom physiological adaptations to nutrient availability in culture, a key factor for understanding the influence of nutrient inputs to diatom growth and bloom formation.[96–101] Taken together, these genomic and transcriptomic resources putatively enable the identification of the genetic basis for DA biosynthesis.

## 1.4 In This Dissertation

### 1.4.1 Chapter 2: Biosynthesis of the neurotoxin domoic acid in a bloom forming diatom

In Chapter 2 of this dissertation, I describe the discovery of the DA biosynthetic (*dab*) genes in the diatom *P. multiseries*. This study utilizes a comparative transcriptomics approach to identify genes that are "turned on" under culturing conditions that increase DA production. Our comparative transcriptomics experiments focused on a combination of phosphorous (P) limitation and increased carbon dioxide ($CO_2$) bubbling previously shown to have a nuanced effect on DA production in culture.[102] We were able to identify a suite of approximately 40 genes upregulated under both P limitation and $CO_2$ enrichment, including three genes that had predicted functions consistent with biochemical transformations hypothesized to build the DA molecular scaffold. These genes also co-localized in a compact gene cluster in the *P. multiseries* genome, suggesting a shared involvement in a common biological process. Heterologous expression of the three *dab* genes in *Escherichia coli* and *Saccharomyces cerevisiae* enabled the purification of the encoded enzymes and validation of their direct involvement in DA biosynthesis *in vitro*, establishing a genetic basis for DA production.

### 1.4.2 Chapter 3: Domoic acid biosynthesis in the red alga *Chondria armata* suggests a complex evolutionary history for toxin production

In Chapter 3 of this dissertation, I describe the identification of DA biosynthetic genes in the red macroalgae, *Chondria armata*, the originally described DA producer. Draft genome sequencing of *C. armata* enabled identification of a DA biosynthesis gene cluster featuring a highly similar genomic organization to the original *dab* gene cluster described in Chapter 2. Biosynthetic function of two red algal DA (*rad*) biosynthesis genes was confirmed *in vitro* via heterologous expression and purification of enzymes from *E. coli*. While slight differences in enzyme activity encoded by the *rad* genes revealed new wrinkles to DA biosynthesis, the reactions were overall consistent with what we observed initially in *P. multiseries*. Phylogenetic analysis of the *dab* and *rad* genes suggests an interesting evolutionary history for toxin production, whereby two of the three genes may have been acquired via separate horizontal gene transfer (HGT) events. The third gene appears to have a clear red algal or diatom ancestry, perhaps being neofunctionalized from native enzymology by the *C. armata* or *P. multiseries* host, respectively.

### 1.4.3 Chapter 4: Metatranscriptomic sequencing and barcoding of a major harmful algal bloom event in Monterey Bay, California

In Chapter 4 of this dissertation, we apply the *dab* genes to the broader marine environment in an attempt to establish a paradigm for molecular monitoring and profiling of DA-producing *Pseudo-nitzschia* HABs. Our study focuses on the progression of the 2015 North American West Coast *P. australis* bloom in Monterey Bay, California. Samples were collected

weekly throughout 2015 from Monterey Municipal Wharf II as part of routine monitoring efforts. We generated RNA barcoding and metatranscriptomics datasets from phytoplankton net tow samples to complement monitoring data also collected from the wharf including *Pseudo-nitzschia* cell counts, particulate DA concentration, and chlorophyll concentration. Sequencing the 18SV4 region and chloroplast 16S sequencing allowed us to visualize the *Pseudo-nitzschia* bloom that dominated the bay from early spring to late summer. We also implemented novel *Pseudo-nitzschia*-targeted ITS2 sequencing to describe the various species present throughout the year and highlight the nearly monospecific *P. australis* bloom. Meanwhile, metatranscriptomic sequencing enabled detection of *dab* transcripts coincident with DA production throughout the bloom event. Notably, *dab* transcripts could be detected for *P. australis*, *P. multiseries* and *P. seriata*, three species that could also be identified in the ITS2 sequencing data. Finally, gene networks analysis was implemented to analyze *dab* gene expression in the larger context of *P. australis* transcriptomics, revealing functional clustering of related transcripts that paints a picture of how bloom physiology shifts from inception to demise. These environmental sequencing efforts establish a paradigm for using *dab* transcripts in genetic monitoring of HABs.

## 1.5 Works Cited


1. B. E. Lapointe, R. A. Brewton, L. W. Herren, M. Wang, C. Hu, D. J. McGillicuddy, S. Lindell, F. J. Hernandez, P. L. Morton PL. Nutrient content and stoichiometry of pelagic *Sargassum* reflects increasing nitrogen availability in the Atlantic Basin. *Nat. Commun.* **12(1)**, 3060 (2021)

2. E. Granéli, J. T. Turner JT, Eds. Ecology of harmful algae (Springer, 2006)

3. S. A. Rasmussen, A. J. C. Andersen, N. G. Andersen, K. F. Nielsen, P. J. Hansen, T. O. Larsen. Chemical diversity, origin, and analysis of phycotoxins. *J. Nat. Prod.* **79(3)**, 662–673 (2016)

4. W. W. Carmichael. Health effects of toxin-producing cyanobacteria: "The CyanoHABs." *Hum. Ecol. Risk Assess.* **7(5)**, 1393–1407 (2001)

5. J. Heisler, P. M. Glibert, J. M. Burkholder, D. M. Anderson, W. Cochlan, W. C. Dennison, Q. Dortch, C. J. Gobler, C. A. Heil, E. Humphries, A. Lewitus, R. Magnien, H. G. Marshall, K. Sellner, D. A. Stockwell, D. K. Stoecker, M. Suddleson. Eutrophication and harmful algal blooms: a scientific consensus. *Harmful Algae* **8(1)**, 3–13 (2008)

6. F. X. Fu, A. O. Tatters, D. A. Hutchins. Global change and the future of harmful algal blooms in the ocean. *Mar. Ecol. Prog. Ser.* **470**, 207–233 (2012)

7. M. L. Wells, V. L. Trainer, T. J. Smayda, B. S. O. Karlson, C. G. Trick, R. M. Kudela, A. Ishikawa, S. Bernard, A. Wulff, D. M. Anderson, W. P. Cochlan, Harmful algal blooms and climate change: Learning from the past and present to forecast the future. Harmful Algae **49**, 68–93 (2015)

8. M. L. Wells, B. Karlson, A. Wulff, R. Kudela, C. Trick, V. Asnaghi, E. Berdalet, W. Cochlan, K. Davidson, M. De Rijcke, S. Dutkiewicz, G. Hallegraeff, K. J. Flynn, C. Legrand, H. Paerl, J. Silke, S. Suikkanen, P. Thompson, V. L. Trainer. Future HAB science: directions and challenges in a changing climate. *Harmful Algae*. **91**, 101632 (2020)

9. S. M. McKibben, W. Peterson, A. M. Wood, V. L. Trainer, M. Hunter, A. E. White, Climatic regulation of the neurotoxin domoic acid. *Proc. Natl. Acad. Sci.* **114**, 239–244 (2017)

10. L. Cavole, A. Demko, R. Diner, A. Giddings, I. Koester, C. Pagniello, M-L. Paulsen, A. Ramirez-Valdez, S. Schwenck, N. Yen, M. Zill, P. Franks. Biological impacts of the 2013–2015 warm-water anomaly in the Northeast Pacific: winners, losers, and the future. *Oceanography* **29(2)**, 273–285 (2016)

11. R. M. McCabe, B. M. Hickey, R. M. Kudela, K. A. Lefebvre, N. G. Adams, B. D. Bill, F. M. D. Gulland, R. E. Thomson, W. P. Cochlan, V. L. Trainer. An unprecedented

coastwide toxic algal bloom linked to anomalous ocean conditions. *Geophys. Res. Lett.* **43**, 10366– 10376 (2016)

12. G. P. Rossini, P. Hess. "Phycotoxins: chemistry, mechanisms of action and shellfish poisoning." In *Molecular, Clinical and Environmental Toxicology* Andreas Luch, Ed. (Springer, 2010). pp. 65–122.

13. V. Valdiglesias, M. Prego-Faraldo, E. Pásaro, J. Méndez, B. Laffon. Okadaic acid: more than a diarrheic toxin. *Mar. Drugs* **11(11)**, 4328–4349 (2013)

14. A. Contractor, C. Mulle, G. T. Swanson. Kainate receptors coming of age: milestones of two decades of research. *Trends Neurosci.* **34(3)**, 154–163 (2011)

15. M. de Lera Ruiz, R. L. Kraus. Voltage-gated sodium channels: structure, function, pharmacology, and clinical indications. *J. Med. Chem.* **58(18)**, 7093–7118 (2015)

16. D. W. Hilgemann. From a pump to a pore: how palytoxin opens the gates. *Proc. Natl. Acad. Sci.* **100(2)**, 386–388 (2003)

17. K. F. Meyer, H. Sommer, P. Schoenholz. Mussel Poisoning. *J. Prev. Med.* **2**, 365–394 (1928)

18. G. Vancouver. A voyage of discovery to the North Pacific Ocean and round the world. (1798)

19. E. J. Schantz, J. D. Mold, D. W. Stanger, J. Shavel, F. J. Riel, J. P. Bowden, J. M. Lynch, R. S. Wyler, B. Riegel, H. Sommer. Paralytic shellfish poison. VI. a procedure for the isolation and purification of the poison from toxic clam and mussel tissues. *J. Am. Chem. Soc.* **79(19)**, 5230–5235 (1957)

20. J. Bordner, W. E. Thiessen, H. A. Bates, H. Rapoport. Structure of a crystalline derivative of saxitoxin. *J. Am. Chem. Soc.* **97(21)**, 6008–6012 (1975)

21. E. J. Schantz, V. E. Ghazarossian, H. K. Schnoes, F. M. Strong, J. P. Springer, J. O. Pezzanite, J. Clardy. Structure of saxitoxin. *J. Am. Chem. Soc.* **97(5)**, 1238–1239 (1975)

22. L. E. Llewellyn. Saxitoxin, a toxic marine natural product that targets a multitude of receptors. *Nat. Prod. Rep.* **23(2)**, 200 (2006)

23. K. D. Cusick, G. S. Sayler. An overview on the marine neurotoxin, saxitoxin: genetics, molecular targets, methods of detection and ecological functions. *Mar. Drugs* **11(4)**, 991–1018 (2013)

24. P. Ciminiello, C. Dell'Aversano, E. Fattorusso, M. Forino. Palytoxins: a still haunting Hawaiian curse. *Phytochem. Rev.* **9(4)**, 491–500 (2010)

25. R. E. Moore, P. J. Scheuer. Palytoxin: a new marine toxin from a coelenterate. *Science* **172(3982)**, 495–498 (1971)

26. R. E. Moore, G. Bartolini. Structure of palytoxin. *J. Am. Chem. Soc.* **103(9)**, 2491–2494 (1981)

27. D. Uemura, K. Ueda, Y. Hirata, H. Naoki, T. Iwashita. Further studies on palytoxin. I. *Tetrahedron Lett.* **22(20)**, 1909–1912 (1981)

28. V. Ramos, V. Vasconcelos. Palytoxin and analogs: biological and ecological effects. *Mar. Drugs* **8(7)**, 2021–2037 (2010)

29. M. Usami, M. Satake, S. Ishida, A. Inoue, Y. Kan, T. Yasumoto. Palytoxin analogs from the dinoflagellate *Ostreopsis siamensis*. *J. Am. Chem. Soc.* **117(19)**, 5389–5390 (1995)

30. T. M. Perl, L. Bédard, T. Kosatsky, J. C. Hockin, E. C. D. Todd, R. S. Remis. An outbreak of toxic encephalopathy caused by eating mussels contaminated with domoic acid. *N. Engl. J. Med.* **322(25)**, 1775–1780 (1990)

31. J. L. C. Wright, R. K. Boyd, A. D. Freitas, M. Falk, R. A. Foxall, W. D. Jamieson, M. V. Laycock, A. W. McCulloch, A. G. McInnes, P. Odense, V. P. Pathak, M. A. Quilliam, M. A. Ragan, P. G. Sim, P. Thibault, J. A. Walter, M. Gilgan, D. J. A. Richard, D. Dewar. Identification of domoic acid, a neuroexcitatory amino acid, in toxic mussels from eastern Prince Edward Island. *Can. J. Chem.* **67**, 481–490 (1989)

32. T. Takemoto, K. Daigo. Constituents of *Chondria armata*. *Yakugaku Zasshi*. **6(5)**, 578 – 580 (1958)

33. K. Daigo. Studies on the constituents of *Chondria armata*. II: Isolation of an anthelmintical constituent. *Yakugaku Zasshi* **79**, 353–356 (1959).

34. S. S. Bates, C. J. Bird, A. D. Freitas, R. Foxall, M. Gilgan, L. A. Hanic LA, G. R. Johnson, A. W. McCulloch, P. Odense, R. Pocklington, M. A. Quilliam. Pennate diatom *Nitzschia pungens* as the primary source of domoic acid, a toxin in shellfish from eastern Prince Edward Island, Canada. *Can. J. Fish. Aquat. Sci.* **46**, 1203–1215 (1989)

35. A. Lelong, H. Hégaret, P. Soudant, S. S. Bates, Pseudo-nitzschia (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: Revisiting previous paradigms. *Phycologia* **51**, 168–216 (2012)

36. S. S. Bates, K. A. Hubbard, N. Lundholm, M. Montresor, C. P. Leaw, *Pseudo-nitzschia*, *Nitzschia*, and domoic acid: new research since 2011. *Harmful Algae* **79**, 3–43 (2018)

37. W. H. Gerwick, Plant sources of drugs and chemicals. *Encyclopedia of Biodiversity* **2**, 129-139 (2013)

38. I. Nitta, H. Watase, Y. Tomiie. Structure of kainic acid and its isomer, allokainic acid. *Nature*. **181(4611)**, 761–762 (1958)

39. Y. Komiya, A. Kobayashi, Techniques applied in Japan for the control of *Ascaris* and hookworm infections -- a review. *Jpn. J. Med. Sci. Biol.* **18**, 1–17 (1965)

40. S. F. Traynelis, L. P. Wollmuth, C. J. McBain, F. S. Menniti, K. M. Vance, K. K. Ogden, K. B. Hansen, H. Yuan, S. J. Myers, R. Dingledine. Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62(3)**, 405–496 (2010)

41. J. A. Larm, P. M. Beart, N. S. Cheung, Neurotoxin domoic acid produces cytotoxicity via kainate- and AMPA-sensitive receptors in cultured cortical neurones. Neurochem. Int. 31, 677–682 (1997)

42. J. Zhu, X. Y. Zheng, H. L. Zhang, Q. Luo. Kainic acid-induced neurodegenerative model: potentials and limitations. J Biomed Biotechnol. **2011**, 457079 (2011)

43. K. A. Lefebvre, P. S. Kendrick, W. Ladiges, E. M. Hiolski, B. E. Ferriss, D. R. Smith, D. J. Marcinek, Chronic low-level exposure to the common seafood toxin domoic acid causes cognitive deficits in mice. *Harmful Algae* **64**, 20–29 (2017)

44. L. M. Grattan, C. J. Boushey, Y. Liang, K. A. Lefebvre, L. J. Castellon, K. A. Roberts, A. C. Toben, J. G. Morris, Repeated dietary exposure to low levels of domoic acid and problems with everyday memory: Research to public health outreach. *Toxins* **10**, 103 (2018)

45. J. A. Funk, M. G. Janech, J. C. Dillon, J. J. Bissler, B. J. Siroky, P. D. Bell, Characterization of renal toxicity in mice administered the marine biotoxin domoic acid. *J. Am. Soc. Nephrol.* **25**, 1187–1197 (2014)

46. J. S. Ramsdell, T. S. Zabka, *In utero* domoic acid toxicity: a fetal basis to adult disease in the California sea lion (*Zalophus californianus*). *Mar. Drugs* **6**, 262–290 (2008)

47. C. A. Scholin, F. Gulland, G. J. Doucette, S. Benson, M. Busman, F. P. Chavez, J. Cordaro, R. DeLong, A. De Vogelaere, J. Harvey, M. Haulena, K. Lefebvre, T. Lipscomb, S. Loscutoff, L. J. Lowenstine, R. Marin III, P. E. Miller, W. A. McLellan, P. D. R. Moeller, C. L. Powell, T. Rowles, P. Silvagni, M. Silver, T. Spraker, V. Trainer, F. M. Van Dolah. Mortality of sea lions along the central California coast linked to a toxic diatom bloom. *Nature* **403(6765)**, 80–84 (2000)

48. S. E. Fire, Z. Wang, M. Berman, G. W. Langlois, S. L. Morton, E. Sekula-Wood, C. R. Benitez-Nelson. Trophic transfer of the harmful algal toxin domoic acid as a cause of death in a minke whale (*Balaenoptera acutorostrata*) stranding in Southern California. *Aquat. Mamm.* **36(4)**, 342–350 (2010)

49. T. M. Work, B. Barr, A. M. Beale, L. Fritz, M. A. Quilliam, J. L. C. Wright.

Epidemiology of domoic acid poisoning in brown pelicans (*Pelecanus occidentalis*) and Brandt's cormorants (*Phalacrocorax penicillatus*) in California. *J. Zoo Wildl. Med.* **24(1)**, 54–62 (1993)

50. M. Hildebrand, S. J. L. Lerch, R. P. Shrestha. Understanding diatom cell wall silicification—moving forward. *Front. Mar. Sci.* **5**, 125 (2018)

51. E. V. Armbrust. The life of diatoms in the world's oceans. *Nature* **459(7244)**, 185–192 (2009)

52. A. Mann. The dependence of the fishes on the diatoms. *Ecology* **2(2)**, 79–83 (1921)

53. J. T. Allen, L. Brown, R. Sanders, C. M. Moore, A. Mustard, S. Fielding, M. Lucas, M. Rixen, G. Savidge, S. Henson, D. Mayor. Diatom carbon export enhanced by silicate upwelling in the northeast Atlantic. *Nature* **437(7059)**, 728–732 (2005)

54. A. Yool, T. Tyrrell. Role of diatoms in regulating the ocean's silicon cycle. *Global Biogeochem. Cycles* **17(4)**, 1103 (2003)

55. S. Malviya, E. Scalco, S. Audic, F. Vincent, A. Veluchamy, J. Poulain, P. Wincker, D. Iudicone, C. De Vargas, L. Bittner, A. Zingone, C. Bowler. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci*. **113(11)**, E1516–E1525 (2016)

56. C. X. Huang, H. C. Dong, N. Lundholm, S. T. Teng, G. C. Zheng, Z. J. Tan, P. T. Lim, Y. Li. Species composition and toxicity of the genus *Pseudo-nitzschia* in Taiwan Strait, including *P. chiniana sp. nov.* and *P. qiana sp. nov. Harmful Algae* **84**, 195–209 (2019)

57. N. Lundholm, Ø. Jvind Moestrup. Morphology of the marine diatom *Nitzschia navis-varingica, sp. nov.* (Bacillariophyceae), another producer of the neurotoxin domoic acid. *J. Phycol.* **36(6)**, 1162–1174 (2000)

58. D. B. Smida, N. Lundholm, W. H. C. F. Kooistra, I. Sahraoui, M. V. Ruggiero, Y. Kotaki, M. Ellegaard, C. Lambert, H. H. Mabrouk, A. S. Hlaili. Morphology and molecular phylogeny of *Nitzschia bizertensis sp. nov.*—a new domoic acid-producer. *Harmful Algae* **32,** 49–63 (2014)

59. D. G. Mann, R. Trobajo, S. Sato, C. Li, A. Witkowski, F. Rimet, M. P. Ashworth, R. M. Hollands, E. C. Theriot. Ripe for reassessment: a synthesis of available molecular data for the speciose diatom family Bacillariaceae. *Mol. Phylogenet. Evol.* **158**, 106985 (2021)

60. V. L. Trainer, S. S. Bates, N. Lundholm, A. E. Thessen, W. P. Cochlan, N. G. Adams, C. Trick. *Pseudo-nitzschia* physiological ecology, phylogeny, toxicity, monitoring and impacts on ecosystem health. *Harmful Algae* **14**, 271–300 (2012)

61. R. H. Lampe, E. L. Mann, N. R. Cohen, C. P. Till, K. Thamatrakoln, M. A. Brzezinski, K.

W. Bruland, B. S. Twining, A. Marchetti. Different iron storage strategies among bloom-forming diatoms. *Proc. Natl. Acad. Sci.* **115(52)**, E12275–E12284 (2018)

62. P. W. Boyd, T. Jickells, C. S. Law, S. Blain, E. A. Boyle, K. O. Buesseler, K. H. Coale, J. J. Cullen, H. J. W. de Baar, M. Follows, M. Harvey, C. Lancelot, M. Levasseur, N. P. J. Owens, R. Pollard, R. B. Rivkin, J. Sarmiento, V. Schoemann, V. Smetacek, S. Takeda, A. Tsuda, S. Turner, A. J. Watson. Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. *Science* **315(5812)**, 612–617 (2007)

63. C. G. Trick, B. D. Bill, W. P. Cochlan, M. L. Wells, V. L. Trainer, L. D. Pickell. Iron enrichment stimulates toxic diatom production in high-nitrate, low-chlorophyll areas. *Proc. Natl. Acad. Sci.* **107(13)**, 5887–5892 (2010)

64. J. Ritzman, A. Brodbeck, S. Brostrom, S. McGrew, S. Dreyer, T. Klinger, S. K. Moore. Economic and sociocultural impacts of fisheries closures in two fishing-dependent communities following the massive 2015 U.S. West Coast harmful algal bloom. *Harmful Algae* **80**, 35–45 (2018)

65. X. Du, W. Peterson, J. Fisher, M. Hunter, J. Peterson J. Initiation and development of a toxic and persistent *Pseudo-nitzschia* bloom off the Oregon coast in Spring/Summer 2015. *PLoS One* **11(10)**, e0163977 (2016)

66. J. P. Ryan, R. M. Kudela, J. M. Birch, M. Blum, H. A. Bowers, F. P. Chavez, G. J. Doucette, K. Hayashi, R. Marin, C. M. Mikulski, J. T. Pennington, C. A. Scholin, G. J. Smith, A. Woods, Y. Zhang. Causality of an extreme harmful algal bloom in Monterey Bay, California, during the 2014–2016 northeast Pacific warm anomaly. *Geophys. Res. Lett.* **44(11)**, 5571–5579 (2017)

67. R. M. Kudela, A. Bickel, M. L. Carter, M. D. A. Howard, L. Rosenfeld. The Monitoring of harmful algal blooms through ocean observing: the development of the California Harmful Algal Bloom Monitoring and Alert Program. *Coastal Ocean Observing Systems* 58-75 (2015)

68. K. S. Casey, T. B. Brandon, P. Cornillon, R. Evans. "The past, present, and future of the AVHRR Pathfinder SST Program." In *Oceanography from Space* V. Barale, J.F.R. Gower, L. Alberotanza, Eds. (Springer, 2010) pp. 273–287

69. M. Wang, S. Son. VIIRS-derived chlorophyll-a using the ocean color index method. *Remote Sens. Environ.* **182**, 141–149 (2016)

70. C. R. Anderson, D. A. Siegel, R. M. Kudela, M. A. Brzezinski. Empirical models of toxigenic *Pseudo-nitzschia* blooms: potential use as a remote detection tool in the Santa Barbara Channel. *Harmful Algae* **8(3)**, 478–492 (2009)

71. C. R. Anderson, R. M. Kudela, M. Kahru, Y. Chao, L. K. Rosenfeld, F. L. Bahr, D. M. Anderson, T. A. Norris. Initial skill assessment of the California Harmful Algae Risk

Mapping (C-HARM) system. *Harmful Algae* **59,** 1–18 (2016)

72. L. A. Pearson, P. M. D'Agostino, B. A. Neilan. Recent developments in quantitative PCR for monitoring harmful marine microalgae. *Harmful Algae* **108**, 102096 (2021)

73. H. A. Bowers, R. Marin, J. M. Birch, C. A. Scholin. Sandwich hybridization probes for the detection of *Pseudo-nitzschia* (Bacillariophyceae) species: an update to existing probes and a description of new probes. *Harmful Algae* **70**, 37–51 (2017)

74. D. Tillett, E. Dittmann, M. Erhard, H. von Döhren, T. Börner, B. A. Neilan. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem. Biol.* **7(10)**, 753–764 (2000)

75. J. R. Chekan, T. R. Fallon, B. S. Moore. Biosynthesis of marine toxins. *Curr. Opin. Chem. Biol.* **59**, 119–129 (2020)

76. L. A. Pearson, T. Mihali, M. Moffitt, R. Kellmann, B. Neilan. On the chemistry, toxicology and genetics of the cyanobacterial toxins, microcystin, nodularin, saxitoxin and cylindrospermopsin. *Marine Drugs* **8(5)**, 1650–1680 (2010)

77. R. Kellmann, T. K. Michali, B. A. Neilan. Identification of a saxitoxin biosynthesis gene with a history of frequent horizontal gene transfers. *J. Mol. Evol.* **67(5)**, 526–538 (2008)

78. R. Kellmann, T. K. Mihali, Y. J. Jeon, R. Pickford, F. Pomati, B. A. Neilan. Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. *Appl. Environ. Microbiol.* **74(13)**, 4044–4053 (2008)

79. M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. de Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates, P. Cruz-Morales, S. Duddela, S. Düsterhus, D. J. Edwards, D. P. Fewer, N. Garg, C. Geiger, J. P. Gomez-Escribano, A. Greule, M. Hadjithomas, A. S. Haines, E. J. N. Helfrich, M. L. Hillwig, K. Ishida, A. C. Jones, C. S. Jones, K. Jungmann, C. Kegler, H. U. Kim, P. Kötter, D. Krug, J. Masschelein, A. V. Melnik, S. M. Mantovani, E. A. Monroe, M. Moore, N. Moss, H.-W. Nützmann, G. Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. Tian, N. J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gómez, A. Bechthold, H. B. Bode, R. Borriss, S. F. Brady, A. A. Brakhage, P. Caffrey, Y.-Q. Cheng, J. Clardy, R. J. Cox, R. De Mot, S. Donadio, M. S. Donia, W. A. van der Donk, P. C. Dorrestein, S. Doyle, A. J. M. Driessen, M. Ehling-Schulz, K.-D. Entian, M. A. Fischbach, L. Gerwick, W. H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Höfte, S. E. Jensen, J. Ju, L. Katz, L. Kaysser, J. L. Klassen, N. P. Keller, J. Kormanec, O. P. Kuipers, T. Kuzuyama, N. C. Kyrpides, H.-J. Kwon, S. Lautru, R. Lavigne, C. Y. Lee, B. Linquan, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Méndez, M. Metsä-Ketelä, J. Micklefield, D. A. Mitchell, B. S. Moore, L. M. Moreira, R. Müller, B. A. Neilan, M. Nett, J. Nielsen, F. O'Gara, H. Oikawa, A. Osbourn, M. S. Osburne, B. Ostash, S. M. Payne, J.-L. Pernodet, M. Petricek, J. Piel, O. Ploux, J. M. Raaijmakers, J. A. Salas, E. K. Schmitt, B. Scott, R. F. Seipke, B. Shen, D. H.

Sherman, K. Sivonen, M. J. Smanski, M. Sosio, E. Stegmann, R. D. Süssmuth, K. Tahlan, C. M. Thomas, Y. Tang, A. W. Truman, M. Viaud, J. D. Walton, C. T. Walsh, T. Weber, G. P. van Wezel, B. Wilkinson, J. M. Willey, W. Wohlleben, G. D. Wright, N. Ziemert, C. Zhang, S. B. Zotchev, R. Breitling, E. Takano, F. O. Glöckner, Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015)

80. S. W. Chun, M. E. Hinze, M. A. Skiba, A. R. H. Narayan. Chemistry of a unique polyketide-like synthase. *J. Am. Chem. Soc.* **140(7)**, 2430–2433 (2018)

81. A. L. Lukowski, L. Mallik, M. E. Hinze, B. M. Carlson, D. C. Ellinwood, J. B. Pyser, M. Koutmos, A. R. H. Narayan. Substrate promiscuity of a paralytic shellfish toxin amidinotransferase. *ACS Chem. Biol.* **15(3)**, 626–631 (2020)

82. A. L. Lukowski, D. C. Ellinwood, M. E. Hinze, R. J. DeLuca, J. Du Bois, S. Hall, A. R. H. Narayan. C–H hydroxylation in paralytic shellfish toxin biosynthesis. *J. Am. Chem. Soc.* **140(37)**, 11863–11869 (2018)

83. A. Stüken, R. J. S. Orr, R. Kellmann, S. A. Murray, B. A. Neilan, K. S. Jakobsen. Discovery of nuclear-encoded genes for the neurotoxin saxitoxin in dinoflagellates. *PLoS One* **6(5)**, e20096 (2011)

84. M. A. Akbar, N. Y. Mohd Yusof, N. I. Tahir, A. Ahmad, G. Usup, F. K. Sahrani, H. Bunawan. Biosynthesis of saxitoxin in marine dinoflagellates: an omics perspective. *Mar. Drugs* **18(2)**, 103 (2020)

85. K. M. McKindles, P. V. Zimba, A. S. Chiu, S. B. Watson, D. B. Gutierrez, J. Westrick, H. Kling, T. W. Davis. A multiplex analysis of potentially toxic cyanobacteria in Lake Winnipeg during the 2013 bloom season. *Toxins* **11(10)**, 587 (2019)

86. S. A. Murray, D. J. Suggett, M. A. Doblin, G. S. Kohli, J. R. Seymour, M. Fabris, P. J. Ralph. Unravelling the functional genetics of dinoflagellates: a review of approaches and opportunities. *Perspect. Phycol.* **3(1)**, 37–52 (2016)

87. S. Lin. Genomic understanding of dinoflagellates. *Res. Microbiol.* **162(6)**, 551–569 (2011)

88. E. V. Armbrust, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, D. S. Rokhsar. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004)

89. C. Bowler, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C.

Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J-P Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin–Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M-P. O. Secq, C. Napoli, M. Obornik, M. S. Parker, J-L Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, I. V. Grigoriev. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456(7219)**, 239–244 (2008)

90. M. Lommer, M. Specht, A-S. Roy, L. Kraemer, R. Andreson, M. A. Gutowska, J. Wolf, S. V. Bergner, M. B. Schilhabel, U. C. Klostermeier, R. G. Beiko, P. Rosenstiel, M. Hippler, J. LaRoche. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol.* **13(7)**, R66 (2012)

91. T. Mock, R. P. Otillar, J. Strauss, M. McMullan, P. Paajanen, J. Schmutz, A. Salamov, R. Sanges, A. Toseland, B. J. Ward, A. E. Allen, C. L. Dupont, S. Frickenhaus, F. Maumus, A. Veluchamy, T. Wu, K. W. Barry, A. Falciatore, M. I. Ferrante, A. E. Fortunato, G. Glöckner, A. Gruber, R. Hipkin, M. G. Janech, P. G. Kroth, F. Leese, E. A. Lindquist, B. R. Lyon, J. Martin, C. Mayer, M. Parker, H. Quesneville, J. A. Raymond, C. Uhlig, R. E. Valas, K. U. Valentin, A. Z. Worden, E. V. Armbrust, M. D. Clark, C. Bowler, B. R. Green, V. Moulton, C. van Oosterhout, I. V. Grigoriev. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541(7638)**, 536–540 (2017)

92. J. C. Traller, S. J. Cokus, D. A. Lopez, O. Gaidarenko, S. R. Smith, J. P. McCrow, S. D. Gallaher, S. Podell, M. Thompson, O. Cook, M. Morselli, A. Jaroszewicz, E. E. Allen, A. E. Allen, S. S. Merchant, M. Pellegrini, M. Hildebrand. Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnol. Biofuels* **9(1)**, 258 (2016)

93. A. Oliver, S. Podell, A. Pinowska, J. C. Traller, S. R. Smith, R. McClure, A. Beliaev, P. Bohutskyi, E. A. Hill, A. Rabines, H. Zheng, L. Z. Allen, A. Kuo, I. V. Grigoriev, A. E. Allen, D. Hazlebeck, E. E. Allen. Diploid genomic architecture of *Nitzschia inconspicua*, an elite biomass production diatom. *Sci. Rep.* **11(1)**, 15592 (2021)

94. S. Basu, S. Patil, D. Mapleson, M. T. Russo, L. Vitale, C. Fevola, F. Maumus, R. Casotti, T. Mock, M. Caccamo, M. Montresor, R. Sanges, M. I. Ferrante. Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol.* **215(1)**, 140–156 (2017)

95. P. J. Keeling, F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy,

P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. B. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone, A. Z. Worden, The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol.* **12**, e1001889 (2014)

96. S. R. Smith, C. Glé, R. M. Abbriano, J. C. Traller, A. Davis, E. Trentacoste, M. Vernet, A. E. Allen, M. Hildebrand. Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. *New Phytol.* **210(3)**, 890–904 (2016)

97. K. A. Lema, G. Metegnier, J. Quéré, M. Latimier, A. Youenou, C. Lambert, J. Fauchot, M. Le Gac. Inter and intra-specific transcriptional and phenotypic responses of *Pseudo-nitzschia* under different nutrient conditions. *Genome Biol. Evol.* 11(3):731–747 (2019)

98. S. T. Dyhrman, B. D. Jenkins, T. A. Rynearson, M. A. Saito, M. L. Mercier, H. Alexander, L. P. Whitney, A. Drzewianowski, V. V. Bulygin, E. M. Bertrand, Z. Wu, C. Benitez-Nelson, A. Heithoff. The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLOS One* **7(3)**, e33768 (2012)

99. S. R. Smith, C. L. Dupont, J. K. McCarthy, J. T. Broddrick, M. Oborník, A. Horák, Z. Füssy, J. Cihlář, S. Kleessen, H. Zheng, J. P. McCrow, K. K. Hixson, W. L. Araújo, A. Nunes-Nesi, A. Fernie, Z. Nikoloski, B. O. Palsson, A. E. Allen. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* **10**, 4552 (2019)

100. S. R. Smith, J. T. F. Gillard, A. B. Kustka, J. P. McCrow, J. H. Badger, H. Zheng, A. M. New, C. L. Dupont, T. Obata, A. R. Fernie, A. E. Allen. Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. *PLOS Genet.* **12(12)**, e1006490 (2016)

101. A. E. Allen, J. LaRoche, U. Maheswari, M. Lommer, N. Schauer, P. J. Lopez, G. Finazzi, A. R. Fernie, C. Bowler. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci.* **105(30)**, 10438–10443 (2008)

102. J. Sun, D. A. Hutchins, Y. Feng, E. L. Seubert, D. A. Caron, F-X Fu. Effects of changing $pCO_2$ and phosphate availability on domoic acid production and physiology of the marine harmful bloom diatom *Pseudo-nitzschia multiseries*. *Limnol. Oceanogr.* **56(3)**, 829–840 (2011)

# Chapter 2: Biosynthesis of the Neurotoxin Domoic Acid in a Bloom Forming Diatom

## 2.1 Introduction to Chapter 2

### 2.1.1 Chemical Studies on Domoic Acid Biosynthesis and Proposed Enzymology

Various efforts have been made over the past several decades to understand the biosynthesis of domoic acid (DA) in the diatom *Pseudo-nitzschia*. However, the discrete genes and corresponding enzymes responsible for DA production had remained elusive prior to our pioneering work described in this chapter. In a biosynthetic pathway to a given molecule of interest, various enzymes perform key chemical transformations on common metabolic precursors to build the "natural product," a name commonly applied to the unique specialized molecules produced by different organisms across the tree of life. Discrete biosynthetic enzymes can be identified either by activity-guided biochemical fractionation, wherein the enzyme(s) of interest are separated from the thousands of other endogenous proteins present in the producing organism based on their activity, or via heterologous expression of the candidate genes potentially encoding relevant biosynthetic enzymes. Isolation of a single enzyme via biochemical fractionation is typically a highly arduous undertaking, and identification of biosynthetic gene candidates was similarly difficult prior to the introduction of full-genome sequencing over the past several decades.[1,2]

Despite these challenges, significant progress towards understanding natural product biosynthesis can be made in the absence of genetic and biochemical information by using chemistry-guided approaches. Incorporation of various atomic isotopes into a metabolite of interest is a chemical approach that can give powerful insight into the biochemical pathways feeding directly into biosynthesis. A natural products chemist can readily generate or purchase isotopically labeled precursor molecules, such as glucose or acetate, with selected carbon or hydrogen atoms replaced with the respective heavier $^{13}C$ or $^{2}H$ nuclei. When these labeled

precursors are fed to the producing organism, the natural products chemist can then use analytical approaches such as mass spectrometry (MS) or nuclear magnetic resonance (NMR) to ascertain whether the fed isotopic label is actively incorporated into the natural product.

The three-dimensional structure for DA contains several defining characteristics but may be most recognizable for its cyclic pyrrolidine core.[3,4] Pyrrolidines are defined as fully-saturated 5-membered rings containing one nitrogen and four carbons. The pyrrolidine core of DA is further decorated with a carboxylate at C2, a methyl-carboxylate at C3, and an alkyl sidechain at C4 containing two double bonds in conjugation and an additional terminal carboxylic acid at the C7' position. This pyrrolidine-dicarboxylic acid motif is commonly referred to as a "kainoid ring" in the chemical literature, a feature named after the structurally similar algal metabolite kainic acid (KA)(Fig 1).

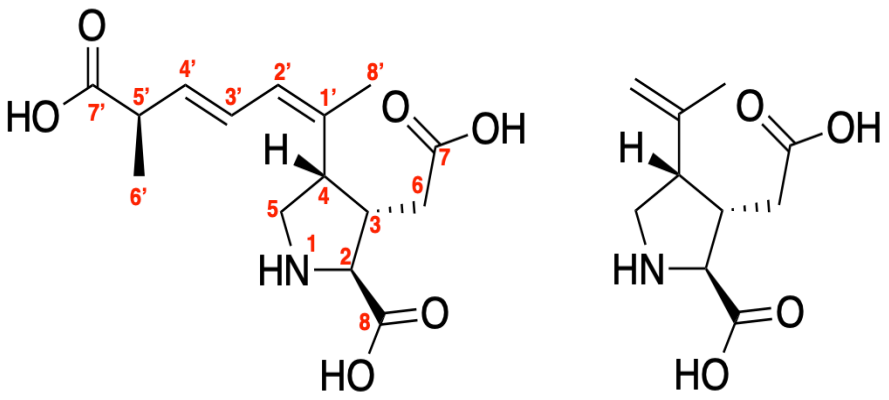



**Figure 2.1.** The Red Algal Kainoids. Domoic acid (DA, left) and Kainic acid (KA, right). Unified numbering scheme shown for DA.[3]

By simply looking at the DA structure, chemists have long hypothesized that the biosynthetic precursors to DA include L-glutamate and a monoterpene. To test this hypothesis, J.C. Wright and colleagues fed labeled [1,2-$^{13}C_2$]acetate to cultures of *Pseudo-nitzschia*

*multiseries* and extracted DA to determine the enrichment of the $^{13}C$ isotope, measured by $^{13}C$ nuclear magnetic resonance (NMR), at various positions along the molecule.[5,6] Feeding of $^{13}C$-labeled acetate is an especially powerful approach due to the wide-ranging incorporation of acetic acid in diverse metabolic pathways such as the citric acid cycle and fatty acid biosynthesis. Steroids and other terpenoids derived from the classic mevalonate (MEV) isoprenoid biosynthetic pathway readily incorporate $^{13}C$ from isotopically-labeled acetate, an expected phenomenon considering the MEV pathway begins with the condensation of two molecules of acetyl-coenzyme A (CoA) as its first committed step.[7]

As a result of these [1,2-$^{13}C_2$]acetate feeding studies, all carbons on the DA scaffold exhibited increased incorporation of $^{13}C$, although at differing levels of enrichment based on position [5,6]. Positions C6 and C7 were the most highly enriched for $^{13}C$ (30.6% and 29.2%, respectively), with a very high (~95%) probability of both positions being occupied by the heavy isotope as evidenced by the elevated $^{13}C$-$^{13}C$ coupling between the two positions revealed in the $^{13}C$-NMR spectra. This is consistent with incorporation of an intact [1,2 $^{13}C_2$]acetate unit. Positions C3 (15.6%), C2 (17.4%), and C8 (14.5%) were all similarly enriched and at a lower level than C6 and C7, with stronger $^{13}C$-$^{13}C$ coupling between C2 and C8 than what was observed between C3 and its adjacent carbons, C2 and C6. Taken together, these trends in $^{13}C$ enrichment and differences in $^{13}C$-$^{13}C$ coupling between positions suggest a tricarboxylic acid cycle (TCA)-derived metabolite. Further feeding studies with [1,2-$^{13}C_2$ $^{2}H_3$]acetate showed deuterium enrichment at C6 but not at C2, suggesting that the TCA cycle-derived DA precursor is likely derived from alpha-ketoglutarate. In this biosynthetic model, the TCA-derived C2 would have been previously occupied by a carbonyl, erasing the $^{2}H$ isotopic signature at the position from the fed labeled acetate.[6] Indeed, alpha-ketoglutarate can be converted directly to L-

glutamate, the predicted direct precursor to DA, through the transamination reactions commonly found in amino acid catabolism.

With respect to the pyrrolidine alkyl sidechain, presumably derived from the monoterpene geranyl pyrophosphate (GPP), Wright and colleagues noticed an [1,2 $^{13}C_2$]acetate labeling trend unique from the expected labeling enrichment witnessed in other isoprene-derived natural products. Every remaining carbon was enriched for $^{13}C$, consistent with the MEV isoprenoid biosynthetic pathway, but at substantially lower levels than expected for a MEV-derived terpenoid, ranging from 2.7 to 4.3-percent incorporation. The authors suggested that low $^{13}C$ incorporation may be due to a subcellular compartmentalized isoprenoid biosynthetic pathway, perhaps hampering full uptake of the labeled acetate.[6] Because the MEV pathway is generally cytoplasmic, the authors refer directly to the then-recently discovered methylerythritol phosphate (MEP) pathway acting in plant chloroplasts as a compatible explanation for the low levels of isotopic incorporation.[8]

However, labeling of every isoprenoid-derived carbon using [1,2-$^{13}C_2$]acetate is not consistent with a strict MEP origin, either. The MEP pathway was originally identified by using [1-$^{13}C$]glucose and utilizes the glycolytic precursors glyceraldehyde-3-phosphate (GAP) and pyruvate.[8] Acetate cannot be converted to either pyruvate or GAP in normal eukaryotic metabolism, although there are rare exceptions in extremophilic prokaryotes employing the enzyme pyruvate synthase.[9,10] Therefore, a MEP-derived isoprene should not have incorporated any $^{13}C$ from the acetate feeding experiment described here, creating quite the conundrum. Despite these apparent inconsistencies, metabolic trans-compartmental crosstalk between the MEV and MEP pathways has been described in plants and subsequently proposed for algae.[11,12] Such a phenomenon could explain the low rates of $^{13}C$ incorporation for the DA alkyl sidechain,

and would therefore be consistent with a primarily chloroplastic, MEP origin for the DA monoterpene precursor with only minimal input from MEV-derived isoprenes labeled by the [1,2-$^{13}C_2$]acetate.

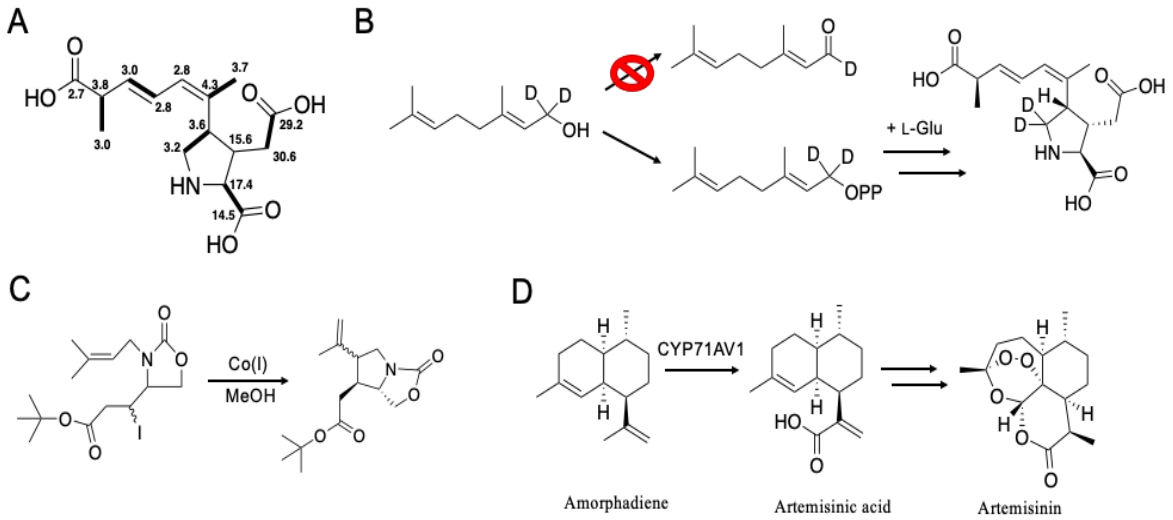



**Figure 2.2.** Insight into DA biosynthesis from isotope labeling, chemical synthesis, and homologous biosynthetic reactions. A) Labeling of 13C nuclei from [1,2-$^{13}C_2$]acetate feeding experiments of *P. multiseries*.[5,6] Numeric labels indicate percent incorporation of $^{13}C$, bold bonds indicate $^{13}C$-$^{13}C$ coupling suggestive of intact acetate incorporation. B) Feeding study of *P. australis* using [1-$^{2}H_2$]-geraniol to suggest an *N*-prenyltransferase mechanism in DA biosynthesis.[13] C) Biomemetic synthesis of kainoids using cobalt catalyst and iodinated synthetic kainoid precursors (cite). D) Conversion of a methyl (-$CH_3$) to a carboxylic acid (-COOH) in artemisinin biosynthesis by a single CYP450 enzyme.[27]

Besides these key efforts, very few studies have made progress in identifying or suggesting the possible enzymatic activities involved in biosynthesizing DA from glutamate and a monoterpene. In 2012, Savage et al provided deuterium-labeling evidence supporting GPP as the likely monoterpene precursor to DA.[13] By feeding cultures of *Pseudo-nitzschia australis* with [1-$^{2}H_2$]-geraniol, assumed to be phosphorylated to [1-$^{2}H_2$]-GPP *in vivo*, the researchers were able to observe enrichment of the DA [M+2] mass by gas chromatography mass spectrometry (GCMS). The isotopic enrichment, although marginal, was reproducible and statistically

significant compared to the controls. Despite the low enrichment of DA [M+2], a similar trend was not observed with the DA [M+1] mass. This observation suggests that GPP is the direct DA precursor and that an alternative pathway via oxidation of geraniol to geranial to enable L-glutamate *N*-prenylation is unlikely. This conclusion is consistent with the lack of precedence for such an aldehyde condensation in *N*-prenylations. On the other hand, there are known examples of enzymes capable of performing *N*-prenylation using isoprene diphosphates as substrates, although it appears that *N*-prenylation is exceedingly uncommon in nature.[14,15] These enzymes can be generally referred to as *N*-prenyltransferases, and the formation of the C-N bond in the final DA structure might be catalyzed by such an enzyme.

Complete biosynthesis of the characteristic pyrrolidine ring of DA also requires formation of an additional C-C bond between C4, derived from GPP, and C3, derived from the beta-position carbon of glutamate. This second transformation would require some sort of chemical activation at C3 for the reaction to proceed. While several routes to activation and ring formation could be envisioned, a radical cyclization mechanism as a potential route to kainoid ring formation was proposed by J.E. Baldwin, whose massive contribution to the field of chemistry includes "Baldwin's rules" of molecular cyclization and numerous investigations into the biosynthesis of penicillin and cephalosporin antibiotics.[16–19] In their work on kainoids, Baldwin and co-workers were able to construct synthetic analogues of the proposed uncyclized, *N*-prenylated precursors to kainic acid and domoic acid.[20,21] Incorporation of iodine at the position analogous to the precursor L-glutamate beta-position carbon allowed for the subsequent cobalt ($Co^{1+}$)-catalyzed abstraction of atomic iodine. Removal of iodine from this synthetic analogue leaves behind a radical to react with the nearby olefin on the prenyl chain, forming the new C3-C4 bond and completing kainoid ring formation.

Conceivably, a single enzyme could remove the C3 hydrogen to enable this radical-driven 5-exo-trig cyclization. Enzymes capable of this sort of hydrogen atom abstraction include diverse transition metal-cofactor enzymes. Iron-dependent enzymes are especially notable for their ability to catalyze this exact sort of oxidative hydrogen atom transfer (HAT) reaction, with common examples including alpha-ketoglutarate and $Fe^{2+}$-dependent dioxygenases, heme-containing oxidases, and in *S*-adenosyl-L-methionine (SAM)-dependent catalysis through involvement in iron-sulfur [Fe-S] clusters.[22–25] Although not technically a biosynthetic study, the synthetic studies by Baldwin demonstrate the feasibility of a radical-driven oxidative cyclization reaction in kainoid ring formation and make the above classes of enzymes tantalizing candidates for a potential role in DA biosynthesis. Oxidative cyclization of natural products has been well described in the chemical literature.[26]

The final defining structural characteristic of DA, the carboxylic acid at the C7' position of the pyrrolidine alkyl sidechain, has not received comparable study into its putative biosynthetic installation. However, direct generation of carboxylic acids from an extant methyl group is a well-studied reaction in natural products chemistry, especially in terpene biosynthesis. In the biosynthetic pathway to the antimalarial artemisinin, a plant-derived sesquiterpene, such a transformation is required to make artemisinic acid from the cyclic precursor amorphadiene. Only one enzyme is required to perform the three hydroxylations to install a carboxylic acid: the cytochrome P450 (CYP450) enzyme CYP71AV1.[27] Similar triple-oxidation reactions have been observed in the biosyntheses for the plant-derived diterpene hormone gibberellin, the bacterial terpene alkaloid xiamycin, and mammalian steroid-derived bile acid.[28–30] All of these reactions are catalyzed by single CYP450s, a large family of heme-containing metalloenzymes.[31]

### 2.1.2 Inducible Production of DA in Culture is a Roadmap to Biosynthetic Genes

Beyond these efforts offering chemical insight into DA biosynthesis, many of the studies on DA production in *Pseudo-nitzchia* have focused on various conditions that can increase or decrease production in culture. These conditions include physical factors, such as salinity, temperature and irradiance, biological factors such as age of culture, growth phase, and presence of associated bacteria, and also a wide range of nutritional limitation and supplementation regimes.[32] Nutrient addition and limitation conditions have received an especially high amount of attention to understand the influence of nutrient availability during bloom events on the severity of *Pseudo-nitzschia* bloom toxicity. For example, silica limitation is linked to increased DA production in culture and anomalously low silica concentrations were observed in Monterey Bay, California during the 2015 North American West Coast *Pseudo-nitzschia* HAB event coincident with historically high levels of DA detected in the water column.[33,34] Other factors including trace metal availability and dissolved $CO_2/HCO_3^-$ have also been demonstrated to augment DA production in culture.[32] The sheer number of different conditions shown to impact DA production makes drawing a discrete link between diatom physiological states, oceanographic conditions, and levels of toxin production difficult, although increased DA production does seem to be a hallmark of cellular stress.

Nevertheless, the inducibility of DA production provides a potential avenue to identifying the enzymatic machinery responsible for toxin production. Although other means of biochemical regulation could be envisioned, such as increased metabolic flux through primary metabolic pathways leading to biosynthetic precursors, inducible production of natural products commonly occurs at the level of biosynthetic gene transcription. With the advent of modern high-throughput sequencing approaches, global transcript abundance can be readily measured using quantitative

RNA sequencing. By generating transcriptomic datasets across different experimental conditions, such as those conditions that may increase or decrease metabolite production, it is possible to identify transcripts with higher or lower abundance between the experiments and identify candidate genes that have a role in metabolite biosynthesis. This "comparative transcriptomics" approach to identify biosynthetic genes has been leveraged to great success in recent years, particularly in plant systems by the Elizabeth Sattely lab. By combining transcriptomics experiments with untargeted metabolomics, Sattely and co-workers have identified novel biosynthetic pathways to various plant natural products.[35–38] One such natural product, 4-hydroxyindole-3-carbonyl nitrile (4-OH-ICN), was a previously undiscovered molecule from *Arabidopsis thaliana* with an apparent role in pathogen defense. Treatment of *A. thaliana* with different plant pathogens directly induced production of 4-OH-ICN and upregulation of discrete transcripts that were subsequently demonstrated to encode enzymes directly implicated in 4-OH-ICN biosynthesis.[38]

Comparative transcriptomics has also been implemented in diatoms to understand physiological responses to nutrient stress. Iron starvation, phosphate limitation, silica limitation, and nitrogen source effects have all been explored using comparative transcriptomic approaches using either microarrays or, more recently, modern RNA sequencing.[39–43] These efforts have yielded tremendous insight into the unique biochemistry found in diatoms, ranging from a dynamically regulated ornithine-urea cycle to non-reductive iron uptake dependent on phytotransferrin, a functional analogue to the well-studied mammalian transferrin present in blood cells.[44,45] The various transcriptomic and functional studies in diatoms have been enabled in large part thanks to substantial progress in diatom genomic sequencing.[46–48] Currently, there are two publicly available *Pseudo-nitzschia* genomes: *P. multistriata* and *P. multiseries*, with

only the *P. multiseries* genome available at the outset of our study into DA biosynthesis.[49] Various transcriptomic efforts had also been made in diverse *Pseudo-nitzschia* species, including a study on the differential expression of transcripts at different phases of logarithmic growth, a factor known to have an effect on DA production in culture.[50-52]

In our present work, we implemented transcriptomic sequencing of *P. multiseries* grown under a combination of phosphate (P) limitation and $CO_2$ enrichment regimes, two conditions previously shown to have a combinatorial effect on the induction of DA biosynthesis in culture.[53] Our differential expression analysis revealed a suite of approximately 40 transcripts consistently upregulated under the different DA-inducing conditions. Four of the upregulated genes co-localized to a common genomic location in a "biosynthetic gene cluster", a phenomenon commonly associated with genes involved in a shared biosynthetic pathway. Of these four genes, three had predicted functions consistent with the biosynthetic transformations required to build DA from glutamate and monoterpene precursors. These three genes included a putatively chloroplast-localized terpene cyclase (*dabA*), an alpha-ketoglutarate-dependent iron dioxygenase (*dabC*), and a cytochrome P450 (*dabD*). *In vitro* characterization of the enzymes encoded by each gene revealed the key chemical transformations of *N*-prenylation of L-glutamate by GPP (DabA), installation of the carboxylic acid functionality at the GPP-derived C7' position (DabD), and 5-exo-trig cyclization to form the kainoid ring structure (DabC). These unique enzymatic steps were sufficient to make isodomoic acid A, a major isomer of domoic acid, providing a core biosynthetic pathway to the DA series of molecules in *P. multiseries*.

## 2.2 Works cited for Chapter 2 Introduction


1. M. H. Medema, T. de Rond, B. S. Moore. Mining genomes to illuminate the specialized chemistry of life. *Nat Rev Genet*. **22(9)**, 553–571 (2021)

2. M. G. Chevrette, K. Gutiérrez-García, N. Selem-Mojica, C. Aguilar-Martínez, A. Yañez-Olvera, H. E. Ramos-Aboites, P. A. Hoskisson, F. Barona-Gómez. Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat Prod Rep.* **37(4)**, 566–599 (2020)

3. J. A. Walter, M. Falk, J. L. C. Wright. Chemistry of the shellfish toxin domoic acid: characterization of related compounds. *Can. J. Chem.* **72(2)**, 430–436 (1994)

4. J. Clayden, B. Read, K. R. Hebditch. Chemistry of domoic acid, isodomoic acids, and their analogues. *Tetrahedron* **61(24)**, 5713–5724 (2005)

5. D. J. Douglas, U. P. Ramsey, J. A. Walter, J. L. C. Wright. Biosynthesis of the neurotoxin domoic acid by the marine diatom *Nitzschia pungens* forma *mulsiseries*, determined with [$^{13}$C]-labelled precursors and nuclear magnetic resonance. *J. Chem. Soc. Chem. Comm.* **9**, 714–716 (1992)

6. U. P. Ramsey, D. J. Douglas, J. A. Walter, J. L. C. Wright. Biosynthesis of domoic acid by the diatom *Pseudo-nitzschia multiseries*. *Nat. Toxins.* **6(3–4)**, 137–46 (1998)

7. H. M. Miziorko. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys.* **505(2)**, 131–143 (2011)

8. H. K. Lichtenthaler, J. Schwender, A. Disch, M. Rohmer. Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. *FEBS Lett.* **400(3)**, 271–274 (1997)

9. T. Nunoura, Y. Chikaraishi, R. Izaki, T. Suwa, T. Sato, T. Harada, K. Mori, Y. Kato, M. Miyazaki, S. Shimamura, K. Yanagawa, A. Shuto, N. Ohkouchi, N. Fujita, Y. Takaki, H. Atomi, K. Takai. A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* **359(6375)**, 559–563 (2018)

10. A. Mall, J. Sobotta, C. Huber, C. Tschirner, S. Kowarschik, K. Bačnik, M. Mergelsberg, M. Boll, M. Hügler, W. Eisenreich, I. A. Berg. Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science* **359(6375)**, 563–567 (2018)

11. J. A. Bick, B. M. Lange. Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Arch. Biochem. Biophys.* **415(2)**, 146–154 (2003)

12. M. Lohr, J. Schwender, J. E. W. Polle. Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Sci*. **185–186**, 9–22 (2012)

13. T. J. Savage, G. J. Smith, A. T. Clark, P. N. Saucedo. Condensation of the isoprenoid and amino precursors in the biosynthesis of domoic acid. *Toxicon* **59(1)**, 25–33 (2012)

14. H. Sugawara, N. Ueda, M. Kojima, N. Makita, T. Yamaya, H. Sakakibara. Structural insight into the reaction mechanism and evolution of cytokinin biosynthesis. *Proc. Natl. Acad. Sci.* **105(7)**, 2734–2739 (2008)

15. A. Mattila, R-M. Andsten, M. Jumppanen, M. Assante, J. Jokela, M. Wahlsten, K. M. Mikula, C. Sigindere, D. H. Kwak, M. Gugger, H. Koskela, K. Sivonen, X. Liu, J. Yli-Kauhaluoma, H. Iwaï, D. P. Fewer. Biosynthesis of the bis-prenylated alkaloids muscoride A and B. *ACS Chem. Biol.* **14(12)**, 2683–2690 (2019)

16. J. E. Baldwin. Rules for ring closure. *J. Chem. Soc. Chem. Comm.* **18**, 734 (1976)

17. J. E. Baldwin, S. E. Abraham. The biosynthesis of penicillins and cephalosporins. *Nat. Prod. Rep.* **5(2)**, 129-145 (1988)

18. P. L. Roach, I. J. Clifton, V. Fülöp, K. Harlos, G. J. Barton, J. Hajdu, I. Andersson, C. J. Schofield, J. E. Baldwin. Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature* **375(6533)**, 700–704 (1995)

19. M. D. Lloyd, H-J Lee, K. Harlos, Z-H Zhang, J. E. Baldwin, C. J. Schofield, J. M. Charnock, C. D. Garner, T. Hara, A. C. Terwisscha van Scheltinga, K. Valegård, J. A. Viklund, J. Hajdu, I. Andersson, Å Danielsson, R. Bhikhabhai. Studies on the active site of deacetoxycephalosporin C synthase. *J. Mol. Biol.* **287(5)**, 943-960 (1999)

20. J. E. Baldwin, M. G. Moloney, A. F. Parsons. Enantioselective kainoid synthesis by cobalt-mediated cyclisation of an amino acid derivative. *Tetrahedron* **46(20)**, 7263–7282 (1990)

21. J. E. Baldwin, M. G. Moloney, A. F. Parsons. Enantioselective synthesis of kainoid analogues by cobalt-mediated cyclisations. *Tetrahedron* **47(1)**, 155–172 (1991)

22. Y. Nakano, K. F. Biegasiewicz, T. K. Hyster. Biocatalytic hydrogen atom transfer: an invigorating approach to free-radical reactions. *Curr. Opin. Chem. Biol.* **49**, 16–24 (2019)

23. Y. Wei, E. L. Ang, H. Zhao. Recent developments in the application of P450 based biocatalysts. *Curr. Opin. Chem. Biol.* **43**, 1–7 (2018)

24. J. B. Broderick, B. R. Duffus, K. S. Duschene, E. M. Shepard. Radical S-adenosylmethionine enzymes. *Chem. Rev.* **114(8)**, 4229–4317 (2014)

25. C. Krebs, D. Galonić Fujimori, C. T. Walsh, J. M. Bollinger. Non-heme Fe(IV)–oxo Iitermediates. *Acc. Chem. Res.* **40(7)**, 484–492 (2007)

26. C-M Tang, Y. Zou, K. Watanabe, C. T. Walsh, Y. Tang. Oxidative cyclization in natural

product biosynthesis. *Chem. Rev.* **117(8)**, 5226–5333 (2017)

27. R. Yu, W. Wen. Artemisinin biosynthesis and its regulatory enzymes: progress and perspective. *Pharmacogn. Rev.* **5(10)**, 189–194 (2011)

28. C. A. Helliwell, J. A. Sullivan, R. M. Mould, J. C. Gray, W. J. Peacock, E. S. Dennis. A plastid envelope location of *Arabidopsis* ent-kaurene oxidase links the plastid and endoplasmic reticulum steps of the gibberellin biosynthesis pathway. *Plant J.* **28(2)**, 201–208 (2001)

29. Q. Zhang, H. Li, S. Li, Y. Zhu, G. Zhang, H. Zhang, W. Zhang, R. Shi, C. Zhang. Carboxyl formation from methyl via triple hydroxylations by XiaM in xiamycin A biosynthesis. *Org. Lett.* **14(24)**, 6142–6145 (2012)

30. I. Holmberg-Betsholtzs, E. Lundq, I. Bjorkhemq, K. Wikvalll. Sterol 27-hydroxylase in bile acid biosynthesis. *J. Biol. Chem.* **268(15)**, 11079-11085 (1993)

31. I. G. Denisov, T. M. Makris, S. G. Sligar, I. Schlichting. Structure and chemistry of cytochrome P450. *Chem. Rev.* **105(6)**, 2253–2277 (2005)

32. A. Lelong, H. Hégaret, P. Soudant, S. S. Bates. *Pseudo-nitzschia* (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. *Phycologia* **51(2)**, 168–216 (2012)

33. Y. Pan, D. V. Subba Rao, K. H. Mann, R. G. Brown, R. Pocklington. Effects of silicate limitation on production of domoic acid, a neurotoxin, by the diatom *Pseudo-nitzschia multiseries*. I. Batch culture studies. *Mar. Ecol. Prog. Ser.* **131**, 225–233 (1996)

34. J. P. Ryan, R. M. Kudela, J. M. Birch, M. Blum, H. A. Bowers, F. P. Chavez, G. J. Doucette, K. Hayashi, R. Marin, C. M. Mikulski, J. T. Pennington, C. A. Scholin, G. J. Smith, A. Woods, Y. Zhang. Causality of an extreme harmful algal bloom in Monterey Bay, California, during the 2014–2016 northeast Pacific warm anomaly. *Geophys. Res. Lett*. **44(11)**, 5571–5579 (2017)

35. R. S. Nett, Y. Dho, Y-Y Low, E. S. Sattely. A metabolic regulon reveals early and late acting enzymes in neuroactive Lycopodium alkaloid biosynthesis. *Proc. Natl. Acad. Sci.* **118(24)**, e2102949118 (2021)

36. R. S. Nett, W. Lau, E. S. Sattely. Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* **584(7819)**, 148–153 (2020)

37. A. P. Klein, E. S. Sattely. Two cytochromes P450 catalyze S-heterocyclizations in cabbage phytoalexin biosynthesis. *Nat. Chem. Biol.* **11**, 837–839 (2015)

38. J. Rajniak, B. Barco, N. K. Clay, E. S. Sattely. A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defence. *Nature* **525**, 376–379 (2015)

39. A. E. Allen, J. LaRoche, U. Maheswari, M. Lommer, N. Schauer, P. J. Lopez, G. Finazzi, A. R. Fernie, C. Bowler. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci.* **105(30)**, 10438–10443 (2008)

40. S. R. Smith, J. T. F. Gillard, A. B. Kustka, J. P. McCrow, J. H. Badger, H. Zheng, A. M. New, C. L. Dupont, T. Obata, A. R. Fernie, A. E. Allen. Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. *PLOS Genet.* **12(12)**, e1006490 (2016)

41. S. T. Dyhrman, B. D. Jenkins, T. A. Rynearson, M. A. Saito, M. L. Mercier, H. Alexander, L. P. Whitney, A. Drzewianowski, V. V. Bulygin, E. M. Bertrand, Z. Wu, C. Benitez-Nelson, A. Heithoff. The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLOS One* **7(3)**, e33768 (2012)

42. S. R. Smith, C. Glé, R. M. Abbriano, J. C. Traller, A. Davis, E. Trentacoste, M. Vernet, A. E. Allen, M. Hildebrand. Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. *New Phytol.* **210(3)**, 890–904 (2016)

43. S. R. Smith, C. L. Dupont, J. K. McCarthy, J. T. Broddrick, M. Oborník, A. Horák, Z. Füssy, J. Cihlář, S. Kleessen, H. Zheng, J. P. McCrow, K. K. Hixson, W. L. Araújo, A. Nunes-Nesi, A. Fernie, Z. Nikoloski, B. O. Palsson, A. E. Allen. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* **10**, 4552 (2019)

44. A. E. Allen, C. L. Dupont, M. Oborník, A. Horák, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie, C. Bowler. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**, 203–207 (2011)

45. J. B. McQuaid, A. B. Kustka, M. Oborník, A. Horák, J. P. McCrow, B. J. Karas, H. Zheng, T. Kindeberg, A. J. Andersson, K. A. Barbeau, A. E. Allen. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature* **555(7697)**, 534–537 (2018)

46. E. V. Armbrust, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, D. S. Rokhsar. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306(5693)**, 79–86 (2004)

47. C. Bowler, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C.

Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J-P Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin–Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M-P. O. Secq, C. Napoli, M. Obornik, M. S. Parker, J-L Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, I. V. Grigoriev. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456(7219)**, 239–244 (2008)

48. T. Mock, R. P. Otillar, J. Strauss, M. McMullan, P. Paajanen, J. Schmutz, A. Salamov, R. Sanges, A. Toseland, B. J. Ward, A. E. Allen, C. L. Dupont, S. Frickenhaus, F. Maumus, A. Veluchamy, T. Wu, K. W. Barry, A. Falciatore, M. I. Ferrante, A. E. Fortunato, G. Glöckner, A. Gruber, R. Hipkin, M. G. Janech, P. G. Kroth, F. Leese, E. A. Lindquist, B. R. Lyon, J. Martin, C. Mayer, M. Parker, H. Quesneville, J. A. Raymond, C. Uhlig, R. E. Valas, K. U. Valentin, A. Z. Worden, E. V. Armbrust, M. D. Clark, C. Bowler, B. R. Green, V. Moulton, C. van Oosterhout, I. V. Grigoriev. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541(7638)**, 536–540 (2017)

49. S. Basu, S. Patil, D. Mapleson, M. T. Russo, L. Vitale, C. Fevola, F. Maumus, R. Casotti, T. Mock, M. Caccamo, M. Montresor, R. Sanges, M. I. Ferrante. Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol.* **215(1)**, 140–156 (2017)

50. K. R. Boissonneault, B. M. Henningsen, S. S. Bates, D. L. Robertson, S. Milton, J. Pelletier, D. A. Hogan, D. E. Housman. Gene expression studies for the analysis of domoic acid production in the marine diatom *Pseudo-nitzschia multiseries*. *BMC Mol. Biol.* **14**, 25 (2013)

51. V. Di Dato, F. Musacchia, G. Petrosino, S. Patil, M. Montresor, R. Sanges, M. I. Ferrante. Transcriptome sequencing of three *Pseudo-nitzschia* species reveals comparable gene sets and the presence of nitric oxide synthase genes in diatoms. *Sci. Rep.* **5**, 12329 (2015)

52. P. J. Keeling, F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. B.

Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone, A. Z. Worden. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol.* **12**, e1001889 (2014)

53. J. Sun, D. A. Hutchins, Y. Feng, E. L. Seubert, D. A. Caron, F-X Fu. Effects of changing $pCO_2$ and phosphate availability on domoic acid production and physiology of the marine harmful bloom diatom *Pseudo-nitzschia multiseries*. *Limnol. Oceanogr.* **56(3)**, 829–840 (2011)

## 2.3 Reprint of: "Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom"

### 2.3.1 Main text





Published in final edited form as:

*Science*. 2018 September 28; 361(6409): 1356–1358. doi:10.1126/science.aau0382.

# Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom

**John K. Brunson**[#1,2], **Shaun M. K. McKinnie**[#1], **Jonathan R. Chekan**[1], **John P. McCrow**[2], **Zachary D. Miles**[1], **Erin M. Bertrand**[2,3], **Vincent A. Bielinski**[4], **Hanna Luhavaya**[1], **Miroslav Oborník**[5], **G. Jason Smith**[6], **David A. Hutchins**[7], **Andrew E. Allen**[2,8,†], and **Bradley S. Moore**[1,9,†]

[1]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA.

[2]Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA 92037, USA.

[3]Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada.

[4]Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, La Jolla, CA 92037, USA.

[5]Institute of Parasitology, University of South Bohemia and Biology Center CAS, Branišovská 31, 370 05 České Budějovice, Czech Republic.

[6]Moss Landing Marine Laboratories, 8272 Moss Landing Road, Moss Landing, CA 95039, USA.

[7]Marine and Environmental Biology, Department of Biological Sciences, University of Southern California,Los Angeles, CA 90089, USA.

[8]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92037, USA.

[9]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

[#] These authors contributed equally to this work.

## Abstract



[†]Corresponding author. bsmoore@ucsd.edu (B.S.M.); aallen@jcvi.org (A.E.A.).

**Author contributions:** J.K.B., S.M.K.M., J.R.C., A.E.A., and B.S.M. conceived the project, designed the experiments, analyzed the data, and wrote the paper, with input from all authors. J.K.B., S.M.K.M., J.R.C., and Z.D.M. performed protein expression and did the enzymology experiments. S.M.K.M and J.K.B. performed chemical synthesis and structural characterization of enzymatic products. J.P.M., E.M.B., J.K.B., and A.E.A. performed the transcriptome analyses. V.A.B. designed vectors and expression constructs and assisted with sequencing. H.L. designed the yeast expression experiments. M.O. constructed the phylogeny. G.J.S. provided the *P. multiseries* isolate 15091C3. D.A.H. provided *P. multiseries* biomass for RNA sequencing from the DA induction experiments (17) and helped to design RNA sequencing experiments.

**Competing interests:** The authors declare no competing interests.

**Data and materials availability:** Transcriptome sequences are deposited in NCBI's Sequence Read Archive, BioProjectIDs SAMN08773590-SAMN08773622. The sequence of the *dab* cluster as amplified from *P. multiseries* 15901C3 is deposited in GenBank, accession no. MH202990. Data used to construct fig. S1 are courtesy of the NASA Ocean Biology Processing Group (OBPG) and are available from https://oceancolor.gsfc.nasa.gov. All other data needed to evaluate the paper are present in the supplementary materials.

Oceanic harmful algal blooms of *Pseudo-nitzschia* diatoms produce the potent mammalian neurotoxin domoic acid (DA). Despite decades of research, the molecular basis for its biosynthesis is not known. By using growth conditions known to induce DA production in *Pseudo-nitzschia multiseries*, we implemented transcriptome sequencing in order to identify DA biosynthesis genes that colocalize in a genomic four-gene cluster. We biochemically investigated the recombinant DA biosynthetic enzymes and linked their mechanisms to the construction of DA's diagnostic pyrrolidine skeleton, establishing a model for DA biosynthesis. Knowledge of the genetic basis for toxin production provides an orthogonal approach to bloom monitoring and enables study of environmental factors that drive oceanic DA production.

Oceanic harmful algal blooms (HABs) are intensifying in frequency and severity in association with climate change (1–3). This phenomenon was exemplified in the summer of 2015 when the North American Pacific coast experienced the largest HAB ever recorded, spanning the Aleutian Islands of Alaska to the Baja peninsula of Mexico (fig. S1) (4). This bloom caused widespread ecological and economic devastation, resulting in the deaths of marine mammals and the closure of beaches and fisheries. The dominant algal species in the 2015 HAB were pennate diatoms of the globally distributed genus *Pseudo-nitzschia*, which are often associated with high production of the excitatory glutamate receptor agonist domoic acid (DA; **1**). Mammalian consumption of DA-contaminated shellfish exerts its toxicity at the AMPA and kainate ionotropic glutamate receptors of the central nervous system (5, 6). In humans, a high, single-exposure dose of DA can cause amnesic shellfish poisoning (ASP), which involves symptoms of amnesia, seizures, coma, and in extreme cases, death. Even chronic, low-level consumption of DA may lead to kidney damage, cognitive deficit, and impairment of fetal development, making DA outbreaks an important human health problem (7–10). Similar neurotoxic symptoms have been observed in birds and marine mammals such as sea lions, which suffer spatial memory impairment linked to DA consumption, likely leading to increases in sea lion strandings (11).

Although abiotic (12) and biotic (13) factors have been shown to affect toxicity in culture, the biological and physicochemical mechanisms underlying DA production in *Pseudo-nitzschia* are unclear. Moreover, not all *Pseudo-nitzschia* blooms produce DA (12). An understanding of the genetic basis of DA biosynthesis in diatoms would facilitate determination of the cellular pathway that controls oceanic DA production and thereby diatom toxicity.

Stable isotope experiments (14, 15) suggest that DA is composed of glutamic acid (Glu) and geranyl pyrophosphate (GPP) building blocks. No pathway that involves these starting materials to construct the characteristic pyrrolidine ring of DA has been previously characterized, and thus, our ability to predict candidate bio-synthetic genes using traditional bioinformatic methods was limited. However, we postulated that a redox enzyme, such as a cytochrome P450 (CYP450), generates the 7′-carboxylic acid of DA through three successive oxidations, a reaction common in all branches of life (Fig. 1A) (16).

To identify putative DA biosynthetic genes, we examined patterns of transcriptional activity under previously established conditions—phosphate limitation and elevated $CO_2$ (17)—that stimulate DA production. We created a differential expression dataset for nearly 20,000

*Pseudo-nitzschia multiseries* RNA transcripts (table S1). A small fraction (~500; 2.5%) showed consistent up-regulation under phosphate limitation (fig. S2 and table S2), and further refinement of this subset to include genes that were also up-regulated with increasing partial pressure of $CO_2$ ($PCO_2$) highlighted only 43 (0.22%) transcripts (table S3). A CYP450 gene showed the highest fold change of all analyzed transcripts under the high $PCO_2$ and low-phosphate DA-inducing condition (Fig. 1B). This transcript was the only annotated CYP450 gene out of 20 total within the *P. multiseries* genome (220 Mbp; Joint Genome Institute accession no. PRJNA32659) that showed increased transcription (fig. S3 and table S4). When mapped to the public *P. multiseries* genome, this CYP450 was localized to a compact genomic island spanning ~8 kb that possesses three other similarly up-regulated genes that are indicative of canonical gene clustering more typically observed in bacterial- and fungal-specialized metabolism (Fig. 1C) (18). Genomic organization in diatoms is not generally typified by clustering of metabolic genes (19, 20). However, clusters of transcriptionally coregulated genes, sensitive to Fe and Si limitation, have been reported (21, 22). The annotated gene functions suggest involvement in terpenoid and redox biochemistry, which are two predicted hallmarks for DA biosynthesis. These gene candidates for DA biosynthetic (Dab) enzymes were annotated as *dabA* (terpene cyclase), *dabB* (hypothetical protein), *dabC* [α-ketoglutarate (αKG)–dependent dioxygenase], and *dabD* (CYP450) (Fig. 1C). We independently sequenced this gene cluster from an environmental isolate of *P. multiseries* in order to validate its conservation (GenBank accession no. MH202990).

Despite the rarity of the pyrrolidine kainoid skeleton in nature, the putative *dab* gene functions enabled us to initially map our genes onto the suggested biosynthetic pathway (14, 15). We hypothesized that DabA catalyzes the first committed step of DA biosynthesis: *N*-prenylation of L-Glu with GPP to form *N*-geranyl-L-glutamic acid (L-NGG; **2**). DabC and DabD would perform subsequent oxidative reactions (Fig. 2).

We began in vitro validation of the *dab* genes with *dabA*, which contains a chloroplast transit peptide sequence and an intron but low similarity to any characterized protein in the National Center for Biotechnology Information (NCBI) database. Structural prediction by using Phyre2 (23) suggested that DabA may possess a terpene cyclase–like fold, and phylogenetic analysis intimated an evolutionary history related to red algal and bacterial genes (fig. S4). Among diatoms, *dabA* appears restricted to *Pseudo-nitzschia* spp. We expressed recombinant $His_6$-DabA without the N-terminal transit peptide in *Escherichia coli* and purified the enzyme using $Ni^{2+}$ affinity and size-exclusion chromatography (fig. S5). In vitro, DabA catalyzes the *N*-geranylation of L-Glu to form L-NGG (2) in a $Mg^{+2}$-dependent manner (Fig. 3A). DabA interrogation with structurally similar substrates shows modest promiscuity toward prenyl pyrophosphates but high specificity for Glu (fig. S6). Recently, L-NGG was isolated in low abundance from the DA-containing red alga *Chondria armata* and shown through labeling experiments in *P. multiseries* to be a precursor to DA (24). These observations further implicate *dabA* as a gene that encodes a major step in DA biosynthesis.

To investigate the subsequent transformations of L-NGG, we individually interrogated the activity of the DabC and DabD oxygenases toward this substrate. In the presence of $Fe^{2+}$, L-ascorbic acid, and cosubstrate aKG, recombinant DabC purified from *E. coli* cyclized L-

NGG to form three pyrrolidine ring–containing molecules: 7′-methylisodomoic acids A, B, and C, termed dainic acids A, B, and C (**5a** to **5c**), respectively (Fig. 3A and figs. S7 and S8), of which **5a** and **5b** had been recently isolated from *C. armata* (24). DabC, however, showed minimal cyclization of other similarly *N*-prenylated glutamic acids (fig. S7). Enzymatic synthesis of the dainic acids was made more amenable through a one-pot coupled assay using L-Glu, GPP, DabA, DabC, and their requisite cofactors and cosubstrates (fig. S9). Despite DabC generating much of the structural diversity observed within the kainoid metabolites (25), the slow rate of L-NGG consumption and failure to go to completion led us to suspect that this enzyme may instead act on an oxidized substrate, placing DabD oxidation ahead of DabC cyclization in the DA biosynthetic pathway. Incubation of L-NGG with *Saccharomyces cerevisiae* micro-somes that possess coexpressed transmembrane proteins DabD (fig. S10) and *P. multiseries* CYP450 reductase (*Pm*CPR1) generated small but reproducible quantities of 7′-hydroxy-L-NGG (6) and 7′-carboxy-L-NGG (3) (Fig. 3A and fig. S11), which was validated through comparison with synthetic standards. Conversely, the dainic acid isomers were not substrates of the DabD P450 (fig. S11). 7′-carboxy-L-NGG incubation with DabC showed rapid cyclization to the *P. multiseries* natural product isodomoic acid A (4) (Fig. 3A), which did not occur in the presence of metal chelator ethylenediaminetetraacetic acid (EDTA). When comparing both DabC in vitro substrates for physiological relevance, 7′-carboxy-L-NGG was consumed at a much faster rate than L-NGG, further suggesting that 7′-carboxy-L-NGG is an on-pathway intermediate (Fig. 3B and fig. S12). Our cumulative in vitro biochemical results imply a DA biosynthetic pathway that begins with the DabA-catalyzed geranylation of L-Glu to yield L-NGG, likely in the chloroplast. DabD then performs three successive oxidation reactions at the 7′-methyl of L-NGG to produce 7′-carboxy-L-NGG, which is then cyclized by DabC to generate the naturally occurring isodomoic acid A (Fig. 2). A putative isomerase likely converts isodomoic acid A to DA. We biochemically interrogated the coclustered *dabB* gene product but did not observe isomerase activity (fig. S13). Further examination of additionally up-regulated transcripts did not suggest an obvious candidate gene (fig. S14). We are actively investigating this final isomerization reaction to complete the pathway to DA.

In addition to *Pseudo-nitzschia* spp., the kainoid structure has been observed in other diatom, red macroalgal, and fungal compounds (fig. S15) (12, 26, 27). In silico application of the *dab* genes as a kainoid ring biosynthetic query identified *dabA* and *dabC* homologs in several red algal transcriptomes (28, 29), including the known kainic acid producer *Palmaria palmata* (fig. S16). Like DA, kainic acid (26) is a structurally related glutamate agonist constructed from a shorter terpene substrate without the need of the DabD P450 oxidation.

With the establishment of the DabACD-dependent biosynthetic pathway to isodomoic acid A in *P. multiseries*, we next linked this discovery to the marine environment. We identified *dab* genes with the same genetic organization and high sequence identity in the genome of the known DA-producing *Pseudo-nitzschia multistriata* (fig. S17) (30). Moreover, of the eight publicly available *Pseudo-nitzschia* transcriptomes, only the highly toxic DA-producing species *Pseudo-nitzschia australis* expressed the *dab* genes (figs. S18 and S19) (31). No *dab* homologs were found in any other sequenced microalgal genera. By virtue of its limited distribution, *dabA* thus presents an opportunity for genetic monitoring of the DA-

producing capabilities of *Pseudo-nitzschia* blooms orthogonal to currently established mass spectrometry–based and enzyme-linked immunosorbent assay–based identification approaches. We anticipate that knowledge of the *dab* genes will allow for greater understanding of the basis and diversity of *Pseudo-nitzschia* toxicity, the physiological function of DA, and the environmental conditions that promote HAB formation so that we may better anticipate risk of exposure to this toxic marine natural product.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We acknowledge B. Duggan, A. Mrse, and Y. Su (all University of California, San Diego) for assistance and maintenance of nuclear magnetic resonance and high-resolution mass spectrometry machinery; W. Fenical (Scripps Institution of Oceanography) for helpful discussion and feedback; and M. Kahru (Scripps Institution of Oceanography) for providing processed satellite images.

Funding:

This work was supported by grants from the National Science Foundation (NSF OCE-1313747 to B.S.M., NSF-ANT-1043671 to A.E.A., and OCE 1538525 and OCE 1638804 to D.A.H.), the National Institute of Environmental Health Sciences (NIEHS P01-ES021921 to B.S.M.), the U.S. Department of Energy Genomics Science program (DE-SC0008593 and DE-SC0018344 to A.E.A.), the Gordon and Betty Moore Foundation (GBMF3828 to A.E.A. and GBMF4960 to G.J.S.), and the Czech Science Foundation (18–13458S to M.O.) and funding from the National Institutes of Health (Chemical Biology Interfaces–University of California, San Diego, training grant 5T32GM112584 to J.K.B.), the Dickinson-McCrink Fellowship (J.K.B.), the Natural Sciences and Engineering Research Council of Canada (NSERC-PDF to S.M.K.M.), and the Simons Foundation Fellowship of the Life Sciences Research Foundation (J.R.C.).

## REFERENCES AND NOTES

1. Wells ML et al., Harmful Algae 49, 68–93 (2015). [PubMed: 27011761]
2. McKibben SM et al., Proc. Natl. Acad. Sci. U.S.A. 114, 239–244 (2017). [PubMed: 28069959]
3. Gobler CJ et al., Proc. Natl. Acad. Sci. U.S.A. 114, 4975–4980 (2017). [PubMed: 28439007]
4. McCabe RM et al., Geophys. Res. Lett 43, 10366–10376 (2016). [PubMed: 27917011]
5. Stewart GR, Zorumski CF, Price MT, Olney JW, Exp. Neurol 110, 127–138 (1990). [PubMed: 2170163]
6. Larm JA, Beart PM, Cheung NS, Neurochem. Int 31, 677–682 (1997). [PubMed: 9364453]
7. Funk JA et al., J. Am. Soc. Nephrol 25, 1187–1197 (2014). [PubMed: 24511141]
8. Lefebvre KA et al., Harmful Algae 64, 20–29 (2017). [PubMed: 28427569]
9. Ramsdell JS, Zabka TS, Mar. Drugs 6, 262–290 (2008). [PubMed: 18728728]
10. Grattan LM et al., Toxins (Basel) 10, 103 (2018).
11. Cook PF et al., Science 350, 1545–1547 (2015). [PubMed: 26668068]
12. Lelong A, Hégaret H, Soudant P, Bates SS, Phycologia 51, 168–216 (2012).
13. Sison-Mangus MP, Jiang S, Tran KN, Kudela RM, ISME J 8, 63–76 (2014).
14. Ramsey UP, Douglas DJ, Walter JA, Wright JL, Nat. Toxins 6, 137–146 (1998). [PubMed: 10223629]
15. Savage TJ, Smith GJ, Clark AT, Saucedo PN, Toxicon 59, 25–33 (2012). [PubMed: 22041653]
16. Zhang X, Li S, Nat. Prod. Rep 34, 1061–1089 (2017). [PubMed: 28770915]
17. Sun J et al., Limnol. Oceanogr 56, 829–840 (2011).
18. Medema MH et al., Nat. Chem. Biol 11, 625–631 (2015). [PubMed: 26284661]
19. Armbrust EV et al., Science 306, 79–86 (2004). [PubMed: 15459382]

20. Bowler C et al., Nature 456, 239–244 (2008). [PubMed: 18923393]
21. Allen AE et al., Proc. Natl. Acad. Sci. U.S.A. 105, 10438–10443 (2008). [PubMed: 18653757]
22. Sapriel G et al., PLOS ONE 4, e7458 (2009). [PubMed: 19829693]
23. Kelley LA, Mezulis S, Yates CM, Wass MN,Sternberg MJE, Nat. Protoc 10, 845–858 (2015). [PubMed: 25950237]
24. Maeno Y et al., Sci. Rep 8, 356 (2018). [PubMed: 29321590]
25. Clayden J, Read B, Hebditch KR, Tetrahedron 61, 5713–5724 (2005).
26. Nitta I, Watase H, Tomiie Y, Nature 181, 761–762 (1958).
27. Konno K, Shirahama H, Matsumoto T, Tetrahedron Lett. 24, 939–942 (1983).
28. Saunders GW, Jackson C, Salomaki ED, Mol. Phylogenet. Evol 119, 151–159 (2018). [PubMed: 29137957]
29. Matasci N et al., Gigascience 3, 17 (2014). [PubMed: 25625010]
30. Basu S et al., New Phytol. 215, 140–156 (2017). [PubMed: 28429538]
31. Keeling PJ et al., PLOS Biol. 12, e1001889 (2014). [PubMed: 24959919]

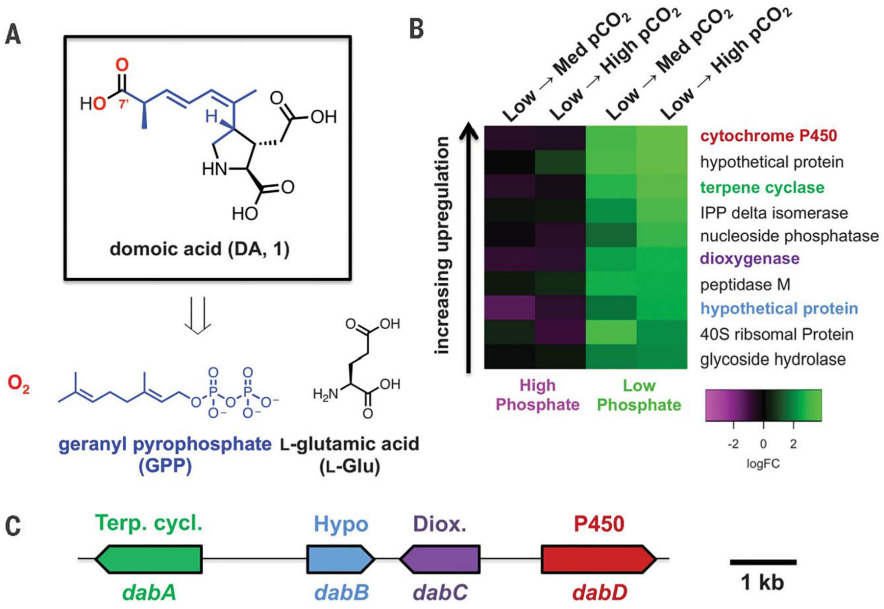



**Fig. 1. Identification of domoic acid biosynthetic genes from transcriptomic and genomic data.**
(**A**) Structure of DA (**1**) and its proposed biosynthetic building blocks (14, 15). (**B**) Ten most up-regulated *P. multiseries* transcripts under previously reported DA-induction conditions (17). Differential expression changes are each shown with reference to the $Pco_2$ conditions listed across the top of the heatmap, with phosphate concentrations held constant. (**C**) DA biosynthesis (*dab*) gene cluster in the *P. multiseries* genome.

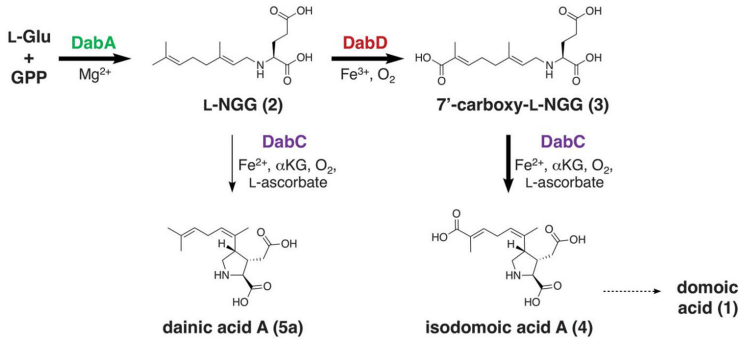



**Fig. 2.**

Domoic acid (1) biosynthetic pathway based on *dab* gene annotations and in vitro enzyme activities.

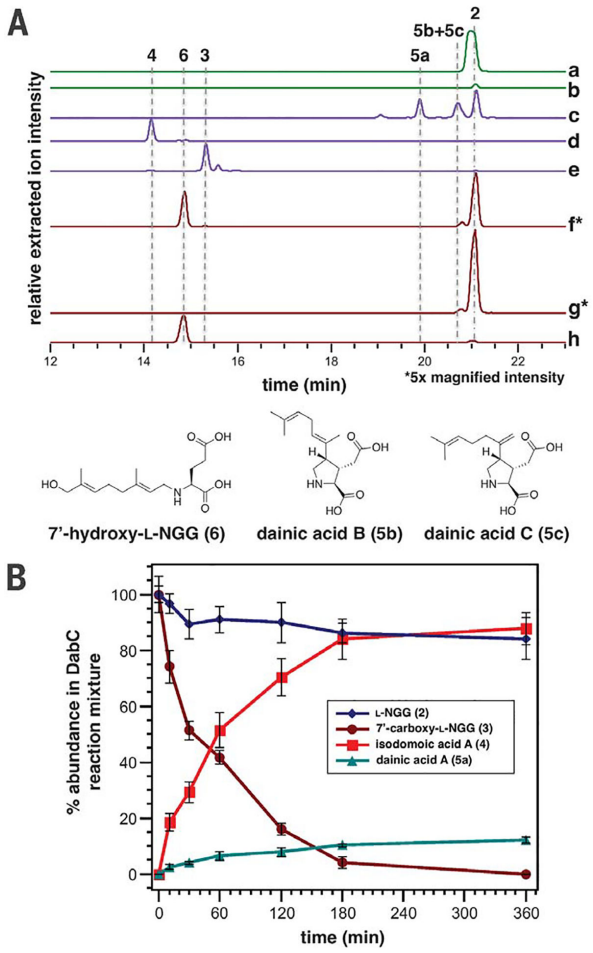



**Fig. 3. In vitro characterization of Dab enzymes.**

(**A**) Relative intensities of negative ionization–extracted ion chromatogram liquid chromatography–mass spectrometry traces [(282.1711, 298.1660, 312.1453, 280.1554, 310.1296) ± 0.01 mass/charge ratio] for in vitro Dab reactions. a, DabA, L-Glu (500 $\mu$M), GPP (500 $\mu$M), $MgCl_2$ (10 mM); b, DabA, L-Glu(500 $\mu$M), GPP (500 $\mu$M), without $MgCl_2$; c, DabC, L-NGG (**2**; 500 $\mu$M), $FeSO_4$, αKG, L-ascorbicacid; d, DabC, 7′-carboxy-L-NGG (**3**; 500 $\mu$M), $FeSO_4$, αKG, L-ascorbic acid; e, DabC, 7′-carboxy-L-NGG (**3**; 500 $\mu$M), $FeSO_4$, αKG, L-ascorbic acid, EDTA (1.0 mM); f, DabD and *Pm*CPR1 containing *S. cerevisiae* microsomes, L-NGG (**2**; 500 $\mu$M), NADPH; g, empty vector *S. cerevisiae* microsomes, L-NGG (**2**; 500 $\mu$M), NADPH; andh, synthetic 7′-hydroxy-L-NGG (**6**) standard. Additional enzyme and cosubstrate concentrations can be found in the supplementary

materials. Asterisks indicate that relative extracted ion intensities for traces f and g have been 5× magnified. (**B**) Comparison of in vitro DabC substrate consumption of L-NGG (**2**; 1.0 mM) and 7′-carboxy-L-NGG (**3**; 1.0 mM) and relative formation of respective dainic acid A (**5a**) and isodomoic acid A (**4**) products ($n = 3$ replicate DabC experiments).

## 2.3.2 Supplementary Materials including Materials and Methods

www.sciencemag.org/content/361/6409/1356/suppl/DC1

# Supplementary Material for

## Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom

John K. Brunson, Shaun M. K. McKinnie, Jonathan R. Chekan John P. McCrow, Zachary D. Miles, Erin M. Bertrand, Vincent A. Bielinski, Hanna Luhavaya, Miroslav Oborník, G. Jason Smith, David A. Hutchins, Andrew E. Allen*, Bradley S. Moore*

*Corresponding author. Email: bsmoore@ucsd.edu (B.S.M.); aallen@jcvi.org (A.E.A.)

Published 28 September 2018, *Science* **361**, 1356 (2017)
DOI: 10.1126/science.aau0382

**This PDF file includes:**



**Other Supplementary Material for this manuscript includes the following:**
(available at www.sciencemag.org/content/361/6409/1356/suppl/DC1)

Tables S1 to S4 as separate Excel files

## Table of Contents

**General Materials and Methods**

All chemicals and solvents were used as received from the commercial supplier (Sigma-Aldrich or Fisher). PCR was carried out using standard thermocycling protocols using either Phusion High-Fidelity DNA Polymerase (New England Biolabs), Prime Star MAX (TaKaRa) or OneTaq 2X MM (New England Biolabs) using the templates and buffers as specified for each reaction (detailed in experimental sections). All protein purification from *E. coli* was performed on an ÄKTApurifier instrument (GE Healthcare) with the modules Box-900, UPC-900, R-900 and Frac-900 with all solvents filtered through a nylon membrane 0.2 μm GDWP (Merck) prior to use. FPLC data was analyzed with UNICORN 5.31 (Built 743) software. Ultracentrifugation was performed using a SW 55 Ti rotor (Beckman Coulter) on an Optima L-100 XP (Beckman Coulter) using ultracentrifuge tubes from the manufacturer. All protein quantification was done by method of Bradford using the Protein Assay Dye Reagent Concentrate (Bio-Rad) on protein sample dilutions in MilliQ water.

General LCMS measurements were carried out on an Agilent Technologies 1200 Series system with a diode-array detector coupled to a Bruker Amazon SL ESI-Ion Trap mass spectrometer in negative ionization mode. The LCMS data was processed by Bruker Compass Data Analysis 4.2 SR2. Higher resolution liquid chromatography mass spectrometry (HRMS) measurements were carried out on an Agilent Technologies 1200 Series system with a diode-array detector coupled to an Agilent Technologies 6530 accurate-mass Q-TOF LCMS. In both cases, compounds were separated by reversed-phase chromatography on a Phenomenex Luna C18(2), 5 μm-100 x 4.6 mm column with water + 0.1% formic acid (solvent A) and acetonitrile + 0.1% formic acid (solvent B) as eluents. We used two specific LC methods for separation of compounds for both general LCMS as well as HRMS, both run at a flow rate of 0.75 mL/min. LC Method A uses the following gradient: hold at 5% B for 1 minute, 5% to 100% B over 20 minutes, hold at 100% B for 1.5 minutes, 100% to 5% B over 2.5 minutes, hold at 5% for 2 minutes. LC Method B uses the following gradient: hold at 5% B for 5 minutes, 5% to 35% B over 15 minutes, 35% to 100% B over 1 minute, hold at 100% B for 1.5 minutes, 100% to 5% B over 2.5 minutes, hold at 5% B for 2 minutes.

For chemical synthesis, all anhydrous reactions were performed under argon atmosphere in flame-dried glassware. Reactions were monitored by LCMS using previously described conditions or TLC (Merck, silica gel 60 F254). Products were visualized by UV at 254 nm or with $KMnO_4$ staining solution (1.5 g $KMnO_4$, 10 g $K_2CO_3$, 1.25 mL 10% NaOH in 200 mL water). TLC $R_f$ values were rounded to the nearest 0.05. Products were purified by flash column chromatography using silica gel (Alfa Aesar, 60 Å pore size). Concentration under reduced pressure was carried out by rotary evaporation. NMR spectra were recorded on a Bruker Avance III spectrometer (600 MHz) using either a 1.7 mm inverse detection triple resonance (H-C/N/D) cryoprobe or a 5 mm inverse detection triple resonance (H-C/N/D) cryoprobe, or a JEOL spectrometer (500 MHz) using $D_2O$ or $CDCl_3$ as solvents. Chemical shifts (δ) are reported in ppm and are referenced to internal sodium formate for $D_2O$ (δ = 8.44 ppm for $^1H$, δ = 171.7 ppm for $^{13}C$) or the solvent signal for $CDCl_3$ (δ = 7.26 ppm for $^1H$, δ = 77.2 ppm for $^{13}C$). Data for NMR spectra was reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, *J* = coupling constant in Hz. Infrared spectra (IR) were recorded on a Nicolet IR 100 FT-IR (Thermo) spectrometer. Absorption maxima are reported as wavenumbers in $cm^{-1}$.

**RNAseq and Differential Expression Analysis Methods**

Biomass for *Pseudo-nitzschia multiseries*, cultured across a matrix of $pCO_2$ conditions and phosphate concentrations, was obtained from experiments conducted by (*17*). Total RNA was extracted using Trizol reagent (Thermo Fisher Scientific), followed by genomic DNA removal using TURBO DNA-free™ Kit (Thermo Fisher Scientific), and Agencourt RNA Clean XP kit (Beckman Coulter) clean up the RNA. Suitable RNA quality was verified using an Agilent 2100 Bioanalyzer.

In order to construct RNAseq libraries from total RNA, ribosomal RNA was removed from 1.5 μg of total using Ribo-Zero Magnetic kits (Illumina). The composition of Removal Solutions constituted a mixture of plant, bacterial, and human/mouse/rat Removal Solution at a ratio of 2:1:1. An Agilent 2100 Bioanalyzer was used to evaluate the quality of rRNA removal. 5 ng of rRNA subtracted RNA was used for library construction using a Scriptseq v2 RNA-seq Library Preparation Kit (Illumina). An Agilent 2100 Bioanalyzer was used to verify the mean size of libraries was within a range of 360-370 bp.

PolyA mRNA transcriptomes were constructed using 300 ng of total RNA as input into the TruSeq RNA Sample Preparation Kit (Illumina), followed by the manufacture's Low-Throughput protocol. The mean size of the final libraries was inspected using an Agilent Bioanalyzer 2100 and confirmed to be within a range of 350-420 bp.

Transcriptomes were sequenced on the Illumina HiSeq platform (Illumina, San Diego, CA, USA). Reads were trimmed of primer sequences and assembled de novo using CLC Genomics workbench. Putative open ready frames on the assembled contigs were called using FragGeneScan (*32*). Putative protein sequences were annotated using hidden markov models and blastp against PhyloDB 1.02 (*33*). *Pseudo-nitzschia multiseries* RNAseq data have been deposited at the NCBI sequence read archive under accession numbers SAMN08773590-SAMN08773622.

Reads were trimmed for quality and filtered to remove primers, adaptors and rRNA sequences using Ribopicker v.0.4.3 (*34*). CLC Assembly Cell (CLCbio) was used to assemble contigs; open reading frames (ORFs) were predicted from the assembled contigs using FragGeneScan (*32*) and additional rRNA sequences were identified and removed. The remaining ORFs were annotated *de-novo* for function and taxonomy via KEGG, KO, KOG, Pfam, TigrFam and the reference dataset PhyloDB 1.02 (*33*).

Reads were aligned to the *Pseudo-nitzschia multiseries* CLN-47 reference assembly Psemu1 (https://genome.jgi.doe.gov/Psemu1) using BWA MEM (*35*). Read counts per gene were aggregated using featureCounts (*36*). Fragments per kilobase of transcript per million mapped reads (FPKM) was calculated as $10^9$ * reads / (total_mapped_reads * gene_length) and was used for assessment of relative expression. Differential expression analysis was performed using edgeR (*37*) with the exactTest function, with multiple testing correction by Benjamini-Hochberg to give FDR adjusted p-values.

Next, the model CDS sequences with reads mapped to them were filtered for only those sequences significantly upregulated under limited phosphate ($log2FC > 0$, $FDR < 0.05$), with reference only to the rRNA-depleted libraries. Specifically, we looked for those transcripts which were significantly upregulated under two phosphate and $pCO_2$ conditions: phosphate limited (0.5 μM) *P. multiseries* grown at 400 ppm $pCO_2$ and 730 ppm $pCO_2$, as compared to cultures grown with replete phosphorous (20 μM) at comparable $pCO_2$ levels. After requiring a cutoff of 50 FPKM for each CDS sequence, with FPKMs summed across all rRNA-depleted libraries (18

libraries, an average of 2.78 FPKM per CDS per library), we were left with a set of 482 CDS which correspond with upregulated transcripts under phosphate limitation.

Next, we looked for genes upregulated under both phosphate limitation and $pCO_2$ stress. We took the set of 482 CDS upregulated under phosphate limitation and filtered for those CDS which correspond to transcripts significantly upregulated (log2FC > 0, FDR < 0.05) under increasing $pCO_2$ at low phosphate concentrations. Specifically, we looked for transcripts which were upregulated when *P. multiseries* was grown under 400 ppm $pCO_2$ or 730 ppm $pCO_2$, as compared to *P. multiseries* grown under 220 ppm $pCO_2$, with all cultures being grown under limited phosphate conditions (0.5 μM). This yielded a total of 43 transcripts upregulated under both $pCO_2$ stress and strictly phosphate limitation.

**Molecular Biology/Biochemical Methods**

Diatom Strains, DNA extraction, cDNA generation, and *dab* gene amplification

For amplification of *dab* genes, *Pseudo-nitzschia multiseries* isolate 15091C3 was collected from a net tow at Monterey Wharf II in Monterey, CA on April 1st 2015. Once purified to unialgal culture, it was maintained in f/2 enriched sterilized Monterey Bay SW at 15-16 $^oC$ and 75 μmole photon $m^{-2}$ $s^{-1}$ PAR (*38*). The organism was also cultured in Natural Sea Water Media, which is 0.22 μM filtered sea water from Scripps Pier supplemented with 880 μM $NaNO_3$, 55 μM $NaH_2PO_4$, 100 μM $Na_2SiO_3$, and AQUIL trace metals and vitamins (*38*). These cultures were also maintained at 15-16 $^oC$ and 75 μmole photon $m^{-2}$ $s^{-1}$ PAR. RNA isolation was performed using the Direct-zol™ RNA MiniPrep kit (ZymoResearch) on *P. multiseries* harvested during late exponential/early stationary phase. cDNA was created using SuperScript™ III Reverse Transcriptase (ThermoFisher). Initial amplification of full-length *dabA* and *PmCPR1* was performed on cDNA samples using Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in Phusion® HF Buffer. Primer set DabA_001-2 was used for *dabA* amplification while primer set PmCPR1_001-2 was used for *PmCPR1* amplification. Initial amplification of *dabC* and *dabD* was performed from genomic DNA extracted from stationary phase *P. multiseries* using Plant DNAzol™ Reagent (Invitrogen), again using Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in Phusion® HF Buffer. Primer set DabC_001-2 was used to amplify *dabC* while primer set DabD_001-2 was used to amplify *dabD*.

Verification of *dab* cluster in an environmental isolate of *Pseudo-nitzschia multiseries*

PrimeStar MAX polymerase (TaKaRa) and primer sets ampDabC-D_001-002 and ampDabA-C_001-002 were used to amplify the intergenic regions between DabC and DabD and between DabA and DabC, respectively, from gDNA extracted from *P. multiseries* isolate 15091C3. Overhangs for Gibson assembly were added using the primer sets OH_DabA-C_001-002 and OH_DabC-D_001-002. These fragments were then individually assembled into PtGG, a promoter-less vector used here for sequencing purposes (*39*) using Gibson assembly (*40*) and transformed into TOP10 chemically competent *E. coli*. Following colony PCR verification using primer set PtGG_col001-002, we used primer sets DabA-C_seq001-010 and DabC-D_seq001-004 for Sanger sequencing (Eurofins) to achieve full coverage. These sequencing reads, along with reads from sequencing performed on *dab* gene expression vectors, were used to assemble a contig which spans the 3' end of *dabA* to the 3' end of *dabD*. Full cluster sequencing has been deposited to GenBank, accession number MH202990.

Cloning of *dabA* into pET28 backbone

The pET28a vector was linearized using PrimeStar MAX polymerase (TaKaRa) and primer set pET28_BB_001-2. Truncated Ser26 – C terminus *dabA* gene was obtained by PCR with PrimeStar polymerase and primer set DabA_p28_001-002 to incorporate the appropriate Gibson assembly overhangs and remove the putative chloroplast transit peptide predicted by SignalP 3.0 and ASAFind (*41, 42)*. NEBuilder HiFi DNA Assembly mix was used to generate the *dabA* pET28a construct. Assembly of *dabA* into pET28a also incorporated the N-terminal $His_6$ tag for affinity purification. After confirming the *dabA* sequence by Sanger sequencing, the plasmid was transformed into *E. coli* BL21 cells for subsequent expression.

Expression and purification of DabA

An overnight culture of DabA expressing BL21 *E. coli* cells was used to inoculate 1 L of Terrific Broth in a 2.8 L flask. The flask was shaken at ~200 rpm at 37 $^{o}$C until the cells reached an $OD_{600}$ of ~0.6. The incubator was then shifted to 18 $^{o}$C and the flasks were allowed to shake for an additional hour. The cells were then induced with 0.5 mM IPTG and were further incubated for ~16 h at 18 $^{o}$C. After harvesting by centrifugation, the cells were resuspended in 25 mL of buffer containing 500 mM NaCl, 20 mM Tris-HCl pH 8.0, and 10% glycerol. The cells were either frozen or immediately used for protein purification.

The cells were lysed by sonication using a Qsonica 6 mm tip at 40% amplitude for 12 cycles of 15 sec on and 45 seconds off. The lysate was centrifuged at 15,000 xg for 30 min to remove cellular debris. The cleared lysate was loaded onto a 5 mL HisTrap FF column (GE Healthcare Life Sciences) that was pre-equilibrated with buffer A1 (20 mM Tris-HCl pH 8.0, 1 M NaCl, and 30 mM imidazole). After loading, the column was washed with 40 mL of buffer A1 to remove any weakly bound protein. DabA was eluted using a linear gradient of 100% buffer A1 to 100% buffer B (20 mM Tris-HCl pH 8.0, 1 M NaCl, and 250 mM imidazole) over 40 mL while collecting 5 mL fractions, keeping flow rate at 2 mL/min. Fractions were assessed for purity using SDS-PAGE. The fractions that were at least 90% pure were combined. The protein was concentrated to 2 mL using an Amicon Ultra-15 30 kDa cutoff concentrator and further purified at a flow rate of 1 mL/min with a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare Life Sciences) pre-equilibrated with 20 mM HEPES pH 8.0, 300 mM KCl, and 10% glycerol. The protein was concentrated and either immediately used for assays or aliquoted, frozen in dry ice, and stored at -80 $^{o}$C.

DabA functional assays

DabA enzyme assays were conducted in 50 mM HEPES (pH 8.0), with 100 mM KCl, 10% glycerol, 10 mM $MgCl_2$, 500 μM geranyl pyrophosphate (or other isoprenoid), 500 μM L-glutamic acid, and 10 μM of purified DabA enzyme. Total reaction volume for small scale assays was brought to 100 μL with MilliQ water. Small scale assays were run at room temperature for ~6 hours. Reactions were quenched with 100 μL of ice cold HPLC-grade MeOH and centrifuged in a table top centrifuge at 15000 xg and 4 $^{o}$C for 20 minutes. Supernatant was then removed and subjected to subsequent HPLC and/or LCMS analysis.

Scaled up DabA reactions

Scaled up reactions for NMR characterization were run in 10 mL volume with 10 mg of GPP and 10 mg of L -Glu added as dry weight, also in 50 mM HEPES (pH 8.0) with 100 mM KCl, 10% glycerol and 10 mM $MgCl_2$. Scaled reactions were allowed to run for ~8 hours at room temperature with stirring and were quenched with 10 mL ice cold HPLC-grade MeOH, followed by centrifugation at 10000 xg and 4 $^{o}$C for 20 minutes. Supernatant was concentrated *in vacuo* to remove methanol and the remaining aqueous reaction solution was frozen and lyophilized overnight.

HPLC purification of DabA product *N*-geranyl-L-glutamic acid (**2**)

The lyophilized aqueous layer was resuspended in 0.1% aqueous formic acid and purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 30% B (5 min), 30 – 45% B (10 min), 45 – 95% B (1 min), 95% B (4 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Fractions containing **2** (retention time 15.2 minutes) were pooled, concentrated *in vacuo* and lyophilized, affording **2** as a white solid. $^{1}$H NMR (600 MHz, $D_2O$): δ 5.21 (t, *J* = 7.3 Hz, 1H), 5.12 (t, *J* = 6.7 Hz, 1H), 3.64 (t, *J* = 6.9 Hz, 2H), 3.57 (dd, *J* = 6.6, 5.7 Hz, 1H), 2.49 – 2.39 (m, 2H), 2.12 – 2.05 (m, 5H), 2.02 (ddt, *J* = 14.5, 6.8, 6.8 Hz, 1H), 1.66 (s, 3H), 1.64 (s, 3H), 1.57 (s, 3H); $^{13}$C NMR (151 MHz, $D_2O$): δ 176.6, 171.9, 146.4, 132.5, 122.4, 111.5, 58.5, 42.5, 37.6, 29.7, 24.0, 23.9, 23.5, 15.7, 14.3; HRMS (ESI) Calculated for $C_{15}H_{24}NO_4$ 282.1711, found 282.1711 $(M-H)^-$.

Marfey's Analysis of DabA assays

DabA reactions were set up as previously described except using 5 mM GPP and 5 mM L -glutamic acid (Glu), D-Glu, or racemic DL-Glu (2.5 mM each enantiomer) and incubated at room temperature for 24 h. Aliquots (10 μL) of the DabA reaction mixture were diluted with MilliQ-water to a final concentration of 1 mM. Aqueous 1 mM standard solutions of L-Glu, D-Glu, L-NGG, D-NGG, and 2 mM racemic Glu and NGG (1 mM each enantiomer) were also prepared. Saturated sodium bicarbonate (20 μL) was added to aliquots (50 μL) of each DabA reaction mixture or standard, followed by 100 μL of a freshly-prepared 1% w/v solution of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (L-FDAA) in acetone. Reactions were incubated at 37 °C for 90 minutes, then quenched by the addition of 25 μL of 1 N HCl. Reactions were centrifuged (13000 x *g*, 2 min) and the clarified supernatant (10 μL for all assays except 20 μL for DabA reaction with DL-Glu) was analyzed by analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm) at a flow rate of 1 mL/min using the following method: 5% B (5 min), 5 – 95% B (20 min), 95 – 5% B (1 min), 5% B (4 min), where A = 0.1% aqueous trifluoroacetic acid, and B = 0.1% trifluoroacetic acid in acetonitrile.

Construction of vector PtpBAD-CTHF

PtpBAD-YFP was first built by amplifying the araC-pBAD region (70048-71265) and the rrnB terminator (4171760-4171962) with primer sets PtpBAD-CTHF_001-2 and PtpBAD-CTHF_003-4 from the K12 genome and assembled into PtpBAD-YFP by ligating the three fragments with an opened portion of the PtPBR1 backbone (*39*). This PtPBR-1 fragment was generated by amplifying the backbone with primer set PtpBAD-CTHF_005-6 in order to eliminate the Amp resistance marker for replacement with the expression cassette. These four fragments were Gibson assembled with a PCR-generated fragment encoding YFP (primers PtpBAD-CTHF_007-8) and containing overhangs to the promoter and terminator fragments.

These four fragments were assembled and after colony PCR and restriction mapping, the vector was validated by Sanger sequencing. In order to generate PtpBAD-CTHF, a synthetic ultramer (uCTHF) was designed based on pNIC-CTHF vector previously described (*43*) but replacing the TEV cleavage site with thrombin. In order to open the backbone and eliminate the YFP gene, primers PtpBAD-CTHF_002 and PtpBAD-CTHF_005 are used to amplify the entire backbone as a single piece. Primer PtpBAD-CTHF_002 amplifies from the 3'end of the pBAD promoter and primer PtpBAD-CTHF_005 amplifies from rrnB terminator and the reaction generates a product of 8,239 bp. The two ultramers encoding the affinity tag region were duplexed according to the manufacturer's protocol (IDT) and the dsDNA fragment was then ligated with the backbone fragment obtained with primers PtpBAD-CTHF_002 and PtpBAD-CTHF_005. This vector was then validated via Sanger sequencing (Eurofins) and designated PtpBAD-CTHF. This vector contains a XhoI site in the uCTHF region for linearization, and proteins assembled into the XhoI site with appropriate overhangs are fused to a carboxy-terminal tag of 21 residues (ALVPRS*GHHHHHHDYKDDDDK) including a thrombin cleavage site (marked with a *), hexahistidine ($His_6$) and FLAG.

*dabC* cloning into PtpBAD-CTHF

PtpBAD-CTHF was digested with XhoI (New England Biolabs). Next, *dabC* was amplified with Gibson overhangs with homology to PtPBAD-CTHF using primer set DabC-CTHF_001-2 with Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in Phusion® HF Buffer. This amplicon was then assembled into the PtpBAD backbone in order to introduce the C-terminal $His_6$-FLAG (CTHF) affinity tag and place the CDS under the control of the arabinose-inducible promoter. The assembled vector was then transformed into TOP10 (ThermoFisher) chemically competent *E. coli*, which were then grown on LB plates with tetracycline (10 μg/mL) Following colony PCR using primer set PtpBAD-CTHF_col001-2 and OneTaq 2X MM with standard buffer (New England Biolabs), plasmid isolation using the ZR Plasmid Miniprep Classic Kit (Zymo Research) and Sanger sequencing (Eurofins) of PtpBAD-DabC-CTHF, we transformed the vector into BL21 chemically competent *E. coli* (New England Biolabs), plated on LB with tetracycline (10 μg/mL) and used these transformants for subsequent DabC expression.

DabC expression in BL21 *E. coli*

We inoculated 1 L of Terrific Broth media containing tetracyclin antibiotic (10 μg/mL) with 10 mL of PtpBAD-DabC-CTHF in BL21 grown overnight at 37 $^{o}$C, 200 rpm to stationary. Each liter of culture was grown in a shaking refrigerated incubator at 37 $^{o}$C, 200 rpm until $OD_{600}$ reached ~0.6. At this time, the temperature on the refrigerated shaker was changed to 18 $^{o}$C and flasks were allowed to adjust to temperature for approx. 30 minutes, after which cultures were induced by addition of 5g L-arabinose to a final concentration of 0.5% (w/v). Growth at 18 $^{o}$C was continued overnight (~12-18 hr) until cells were harvested by centrifugation at 8000 xg for 5 minutes.

DabC purification

Harvested DabC expression cells were resuspended and lysed using the same procedure as DabA expression cells. Due to weak binding of this DabC construct, the cleared lysate was loaded onto a 5 mL HisTrap FF column (GE Healthcare Life Sciences) that was instead pre-equilibrated with resuspension buffer A2, which lacks imidazole (1 M NaCl and 20 mM Tris-

HCl pH 8.0). After loading, the column was washed with 40 mL of buffer A2. The column was further washed in a stepwise manner with 10 mL of increasing buffer B concentrations (2, 4, 6, 8, 10, 12, and 14% buffer B). DabC was then eluted using a linear gradient to 100% buffer B over 40 mL while collecting 5 mL fractions. Fractions were assessed for purity using SDS-PAGE. The fractions that were at least 90% pure were combined and 2 mM EDTA pH 8.0 was added to remove any co-purified metals. DabC was concentrated, further purified by size exclusion chromatography, and stored in an identical manner to DabA.

DabC assay with *N*-prenylated-L-glutamate substrates

DabC enzyme assays were conducted in 100 mM HEPES (pH 8.0) with 100 mM KCl, 10% glycerol, 1 mM L-ascorbate, and 6.25 mM α-ketoglutarate. *N*-geranyl-L-glutamate (**2**) was added to 1 mM concentration, followed by 25 μM DabC purified as above and 25 μM $FeSO_4$ and allowed to run overnight (~18 hours). Total volume for the reaction is approximately 100 μL. After 18 hours, additional L-ascorbate (approx. 1 mM), DabC (approx. 25 μM) and $FeSO_4$ (approx. 25 μM) was added and allowed to continue for an additional 18-24 hours. The reaction was then quenched with 1 eq. of ice cold MeOH and centrifuged at 15000 x g and 4 $^{o}$C for 20 minutes. Reactions were set up in an analogous manner for 7'-carboxy-*N*-geranyl-L-glutamate (**3**), but showed complete substrate consumption after 3 hours and didn't require additional DabC or $FeSO_4$.

DabA and DabC coupled assays to isolate **5a-c**

DabA and DabC coupled assays were carried out at both the 100 μL and 1 mL scale. Reaction contents for both were as follows: 100 mM HEPES pH 8.0, 100 mM KCl, 10% Glycerol, 10 mM $MgCl_2$, 1 mM ascorbate, 6.25 mM μ-KG, 500 μM GPP, 500 μM L-Glu, 10 μM DabA, 25 μM DabC, and 25 μM $FeSO_4$. After 18 hours, additional ascorbate (approx. 1 mM), DabC (approx. 25 μM) and $FeSO_4$ (approx. 25 μM) was added and allowed to continue for an additional 18-24 hours. Both 100 μL and 1mL reactions were quenched with 1 eq. of ice cold MeOH and centrifuged at 15000 x g and 4 $^{o}$C for 20 minutes. The clarified supernatant was concentrated *in vacuo* to remove methanol and the remaining aqueous reaction solution was frozen and lyophilized overnight. For scaled up reactions, we set up 10 x 1 mL reactions, as larger volume incubations with light stirring (e.g. 1 x 10 mL as in DabA scale up reactions), were less efficient as these "split up" scaled reactions.

Purification of **5a-c** products from DabA and DabC coupled assays

The lyophilized aqueous layer was resuspended in 0.1% aqueous formic acid and purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 25% B (5 min), 25 – 40% B (10 min), 40 – 95% B (1 min), 95% B (4 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Fractions containing **5a** (retention time 19.4 minutes), **5b** (retention time 20.6 minutes), and **5c** (retention time 20.4 minutes) were individually pooled, concentrated *in vacuo* and lyophilized. For improved purity, additional separation of regioisomers was performed with analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm) over a gradient of 25 – 40% B (15 min) in the same solvent system. All products **5a-c** were obtained as white solids following lyophilization.

Dainic acid A (**5a**): $^{1}$H NMR (600 MHz, $D_2O$): δ 5.43 (t, *J* = 7.2 Hz, 1H), 5.16 – 5.13 (m, 1H), 3.93 (d, *J* = 6.5 Hz, 1H), 3.63 (dd, *J* = 11.8, 8.1 Hz, 1H), 3.55 (dd, *J* = 7.8, 7.8 Hz, 1H), 3.42

(dd, *J* = 11.8, 8.1 Hz, 1H), 2.95 (ddd, *J* = 9.0, 6.7, 6.5 Hz, 1H), 2.76 (ddd, *J* = 16.2, 8.0, 7.2 Hz, 1H), 2.62 – 2.58 (m, 1H), 2.40 (dd, *J* = 15.6, 6.7 Hz, 1H), 2.30 (dd, *J* = 15.6, 8.5 Hz, 1H), 1.69 (s, 3H), 1.68 (s, 3H), 1.62 (s, 3H); $^{13}$C NMR (151 MHz, $D_2O$): δ 179.4, 173.5, 134.0, 130.2, 129.4, 122.2, 65.1, 46.8, 43.7, 40.3, 36.2, 26.3, 24.7, 21.4, 16.9; HRMS (ESI) Calculated for $C_{15}H_{22}NO_4$ 280.1554, found 280.1554 $(M-H)^-$

Dainic acid B (**5b**): $^1$H NMR (600 MHz, $D_2O$): δ 5.19 (t, *J* = 7.2 Hz, 1H), 5.15 – 5.13 (m, 1H), 4.00 (d, *J* = 3.2 Hz, 1H), 3.55 (dd, *J* = 11.9, 7.5 Hz, 1H), 3.40 (dd, *J* = 11.4, 11.4 Hz, 1H), 3.01 (ddd, *J* = 8.8, 7.3, 2.5 Hz, 1H), 2.94 (dd, *J* = 7.9, 7.6 Hz, 1H), 2.78 – 2.75 (m, 2H), 2.20 (dd, *J* = 15.4, 6.5 Hz, 1H), 2.11 – 2.07 (m, 1H), 1.69 (s, 3H), 1.68 (s, 3H), 1.64 (s, 3H); $^{13}$C NMR (151 MHz, $D_2O$): δ 179.8, 173.4, 133.9, 129.8, 126.1, 122.3, 65.7, 47.0, 46.0, 42.2, 36.1, 26.4, 24.7, 16.9, 15.9; HRMS (ESI) Calculated for $C_{15}H_{22}NO_4$ 280.1554, found 280.1555 $(M-H)^-$

Dainic acid C (**5c**): $^1$H NMR (600 MHz, $D_2O$): δ 5.19 (t, *J* = 8.6 Hz, 1H), 5.04 (s, 1H), 4.77 (s, 1H), 4.06 (d, *J* = 2.4 Hz, 1H), 3.58 (dd, *J* = 11.9, 7.6 Hz, 1H), 3.42 (dd, *J* = 11.8, 11.8 Hz, 1H), 3.04 – 2.95 (m, 2H), 2.25 (dd, *J* = 15.5, 6.0 Hz, 1H), 2.22 – 2.09 (m, 4H), 2.09 – 2.06 (m, 1H), 1.68 (s, 3H), 1.61 (s, 3H); $^{13}$C NMR (151 MHz, $D_2O$): δ 179.8, 173.7, 144.1, 134.0, 123.4, 111.6, 65.7, 45.9, 43.9, 41.5, 35.9, 35.4, 25.4, 24.5, 16.6; HRMS (ESI) Calculated for $C_{15}H_{22}NO_4$ 280.1554, found 280.1556 $(M-H)^-$

Scaled up DabC assay with 7'-carboxy-*N*-geranyl-L-glutamic acid (**3**) substrate

DabC enzyme assays were conducted as previously described in a 5 mL volume using 1.8 mg of 7'-carboxy-*N*-geranyl-L-glutamic acid (**3**) as a substrate with light stirring. The reaction was quenched as previously described after 5 hours, concentrated *in vacuo* and immediately subjected to RP-HPLC purification.

HPLC purification of DabC product isodomoic acid A (**4**)

The lyophilized aqueous layer was resuspended in 0.1% aqueous formic acid and purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 30% B (15 min), 30 – 95% B (3 min), 95% B (2 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Fractions containing **4** (retention time 12.3 minutes) were pooled, concentrated *in vacuo* and lyophilized, affording **4** as a white solid (1.5 mg, 84%).

Isodomoic acid A (**4**): $^1$H NMR (600 MHz, $D_2O$): δ 6.67 (t, *J* = 7.2 Hz, 1H), 5.54 (t, *J* = 7.5 Hz, 1H), 3.97 (d, *J* = 7.3 Hz, 1H), 3.69 (dd, *J* = 11.7, 8.1 Hz, 1H), 3.63 (ddd, *J* = 7.8, 7.6, 7.6 Hz, 1H), 3.46 (dd, *J* = 11.7, 7.7 Hz, 1H), 3.01 (ddd, *J* = 12.7, 10.1, 5.4 Hz, 1H), 2.96 (ddd, 15.6, 7.7, 7.2 Hz, 1H), 2.87 (ddd, 15.8, 7.2, 7.0, 1H), 2.67 (dd, 16.5, 6.5 Hz, 1H), 2.50 (dd, 16.5, 8.6 Hz, 1H), 1.83 (s, 3H), 1.76 (s, 3H); $^{13}$C NMR (151 MHz, $D_2O$): δ 176.5, 172.9, 172.9, 141.0, 131.0, 128.1, 128.0, 65.1, 46.8, 42.4, 40.2, 34.1, 27.1, 21.3, 11.8; HRMS (ESI) Calculated for $C_{15}H_{20}NO_6$ 310.1296, found 310.1294 $(M-H)^-$

Construction of vector pBEVY-GL-dabD-CTHF-PmCPR1

DabD and PmCPR1 were heterologously expressed in *S. cerevisiae* using the vector pBEVY-GL, an expression vector featuring a bidirectional GAL1/10 promoter (*44*). pBEVY-GL was a gift from Charles Miller (Addgene plasmid #51225). The gene *dabD* was first cloned into PtpBAD-CTHF in order to introduce the C terminal $His_6$-FLAG affinity tag. Following colony PCR with OneTaq 2X MM polymerase with standard buffer (New England Biolabs) and primer set PtpBAD_col001-2, as well as Sanger sequencing (Eurofins), we then amplified *dabD*-

*CTHF* from the assembled PtpBAD-dabD-CTHF vector using primer set pBEVY-DabD_001-2 to incorporate Gibson overhangs specific to the pBEVY-GL vector. pBEVY-GL was then linearized with BamHI (New England Biolabs) and the dabD-CTHF amplicon was assembled into the linearized backbone using Gibson assembly in order to insert dabD-CTHF under the GAL1 promoter. Assembled vector was then transformed into Top10 chemically competent *E. coli* cells and plated on ampicillin selection (100 μg/mL). Colony PCR was performed using primer set pBEVY-DabD_col001-2 and OneTaq 2X MM polymerase with standard buffer. Positive clones were sent for Sanger sequencing (Eurofins) to confirm assembly of pBEVY-GL-dabD-CTHF.

Following sequence validation of pBEVY-GL-dabD-CTHF, we moved to clone *PmCPR1* under the GAL10 promoter of pBEVY-GL. *PmCPR1* was PCR amplified with primer set pBEVY-PmCPR1_001-2 to introduce Gibson overhangs specific to the vector pBEVY-GL. Following successful PCR amplification of *PmCPR1*, we used Gibson assembly to clone *PmCPR1* into pBEVY-GL-dabD-CTHF linearized with *SmaI* (New England Biolabs). This assembled vector was then transformed into TOP10 chemically competent *E. coli* cells and plated on ampicillin selection (100 μg/mL). Colony PCR was performed using primers pBEVY-PmCPR1_col001-2 and positive clones were sent for subsequent Sanger sequencing to confirm successful assembly of pBEVY-GL-dabD-CTHF-PmCPR1.

Heterologous expression of DabD and PmCPR1 in *Saccharomyces cerevisiae*

The vector pBEVY-GL-dabD-CTHF-PmCPR1 was then transformed into the protease-deficient *S. cerevisiae* strain BJ5464 (*45*) using standard lithium acetate transformation protocols (*46*). Transformants were then plated onto synthetic drop-out media (Sigma-Aldrich) without amino acid leucine (Leu) to allow selection for transformed yeast colonies. The empty vector pBEVY-GL was also similarly transformed into BJ5464 cells. Screening for DabD expression was performed by growing 50 mL cultures of individual transformants to ~1.0 $OD_{600}$ in -Leu synthetic drop-out media with galactose as the only carbon source (*47*). Cultures were then spun down (8000 xg, 10 minutes, 4 $^{o}$C and protein was extracted using a mild alkali treatment followed by resuspension in 200 μL of SDS sample buffer (*48*). Gels for western blots were run using 10 μL of resuspended sample along with the MagicMark XP Western Protein Standard (ThermoFisher) on NuPAGE 4-12% Bis-Tris protein gels (Invitrogen). Transfer to PVDF membrane (Invitrogen) was conducted in an XCell II Blot Module (Invitrogen). Membranes were then blocked and blotted with FLAG mouse monoclonal antibody (Invitrogen, 1:20000 dilution) using standard protocols, diluents and mouse secondary antibodies from the manufacturer (Invitrogen).

For large scale expression, cells were grown according to established protocols for yeast P450 expression (*47*). In short, a single colony was grown in 50 mL of synthetic dropout media (-leu) to stationary phase. A 1:100 dilution of overnight culture was made into 500 mL of YPGE media and cells were allowed to grow for 24-36 hours or until $OD_{600}$ = 1.0. When cells reached the appropriate density, cultures were induced through the addition of 20% galactose (w/v) to a final concentration of 2%. Cells were then induced for at least 24 hours prior to harvesting by centrifugation (8000 xg, 5 minutes, RT).

Isolation of microsomes from *S. cerevisiae*

Following centrifugation of *S. cerevisiae* transformed with either pBEVY-GL-dabD-CTHF-PmCPR1 or empty pBEVY-GL, cells were lysed using the mechanical procedure from

(*47*). In short, cells were incubated in TEK buffer for 5-10 minutes at room temperature and then resuspended in a minimal volume of TES-B. An equal volume of 0.5 mm glass beads were added to the resuspended yeast and were subjected to vortexer-driven bead-beating for a total of 15 minutes of active lysis, with 15 seconds of vortexing followed by 15 seconds of incubation on ice, all performed in 50 mL conical tubes. Following confirmation of cell lysis by light microscopy, the supernatant was decanted and beads were washed with 2-3 volumes of TES-B. The supernatant fraction and bead washes were then combined and spun at 12000 xg for 30 minutes at 4 $^{o}$C. The supernatant from the low speed spin was then subjected to ultracentrifugation to collect microsomal fractions (130000 xg, 90 minutes, 4 $^{o}$C). Microsomal fractions were then resuspended in the smallest volume of TES-B possible to obtain concentrated microsomal samples.

Microsomal assays

Enzyme assays using ~550 μg of microsomal protein isolated from *S. cerevisiae* strain BJ5464 transformed with either pBEVY-GL-dabD-CTHF-PmCPR1 or empty pBEVY-GL were performed in 100 mM HEPES buffer (pH ~8.0) with 100 mM KCl, 10% glycerol, 1 mM NADPH, 10 μM FAD, 10 μM FMN and 500 μM *N*-geranyl-L-glutamate as substrate. Total reaction volume was brought to 100 μL through addition of MilliQ water. Assays ran for ~8 hours at room temperature, after which the reactions were quenched through addition of 100 μL of ice-cold HPLC grade methanol and centrifuged at max speed (~17000 xg) for 20 minutes at 4 $^{o}$C. The supernatant was then removed for subsequent analysis.

DabC relative substrate consumption assay

DabC reactions were set up as previously described utilizing either 1 mM *N*-geranyl-L-glutamic acid (**2**) or 7'-carboxy *N*-geranyl-L-glutamic acid (**3**) as a substrate. Assays were set up in triplicate for each substrate in 500 μL volumes. At regular time intervals (0, 0.17, 0.5, 1, 2, 3, and 6 h), an aliquot (25 μL) of each assay was removed and quenched with a 25 μL solution of 1.0 mM phenol red in methanol. Quenched aliquots were mixed thoroughly, centrifuged at 13000 x *g* for 2 min, and the clarified supernatant (10 μL) was analyzed by analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm) at a flow rate of 1 mL/min using the following method: 10% B (5 min), 10 – 25% B (5 min), 25 – 40% B (10 min), 40 – 95% B (1 min), 95% B (4 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous trifluoroacetic acid, and B = 0.1% trifluoroacetic acid in acetonitrile. The ratio of the peak integrations at 210 nm corresponding to the substrates (**2** or **3**) or major products (**5a** or **4**) to the phenol red internal standard were compared at various time points to the t = 0 h substrate:phenol red ratio to relatively compare substrate consumption and product accumulation over time.

*dabB* Cloning

The pET28-MBP (Maltose Binding Protein)-TEV vector was linearized using PrimeStar MAX polymerase (TaKaRa) and primer set pET28-MBP_BB_001-2. The *dabB* gene was obtained by PCR with PrimeStar polymerase and primer set DabB_MBP_001-002 to incorporate the appropriate Gibson assembly overhangs. NEBuilder HiFi DNA Assembly mix was used to generate the *dabB* pET28-MBP construct yielding an N-terminal $His_6$-MBP-$His_6$-DabB fusion. After confirming the *dabB* sequence by Sanger sequencing, the plasmid was transformed into *E. coli* BL21 cells for subsequent expression.

DabB Expession and Purification

DabB was expressed and purified by a 5 mL HisTrap FF column (GE Healthcare Life Sciences) column using the same procedure as described for DabA. The DabB containing fractions were collected and concentrated to 2 mL and purified by using a HiLoad 16/60 Superdex 200 prep grade gel filtration column (GE Healthcare Life Sciences) pre-equilibrated with 20 mM HEPES pH 8.0, 300 mM KCl, and 10% glycerol at a flow rate of 1 mL/min. The DabB containing fractions were collected and the MBP tag was removed by the addition of TEV protease and incubated at 4 $^{o}$C overnight. The cut DabB was separated from MBP using a HiLoad 16/60 Superdex 75 prep grade gel filtration column (GE Healthcare Life Sciences) pre-equilibrated with 20 mM HEPES pH 8.0, 300 mM KCl, and 10% glycerol at a flow rate of 1 mL/min. The DabB containing fractions were collected, concentrated, and aliquoted before storage at -80 $^{o}$C (Fig. S5b).

DabB Activity Assays

DabB activity assays were completed using 10 μM DabB and 500 μM substrate with a buffer containing 100 mM HEPES pH 8.0, 100 mM KCl, and 10% glycerol in 100 μL of total volume. The assays were incubated for 18 h at 23 $^{o}$C and quenched with 1 equivalent of methanol. After centifuged to remove precipitated protein, they were analyzed by LCMS method B on the Agilent Technologies 1200 Series system with a diode-array detector coupled to an Agilent Technologies 6530 accurate-mass Q-TOF LCMS. Assays containing DabD were completed using the microsomal assay protocol with the addition of 10 μM DabB.

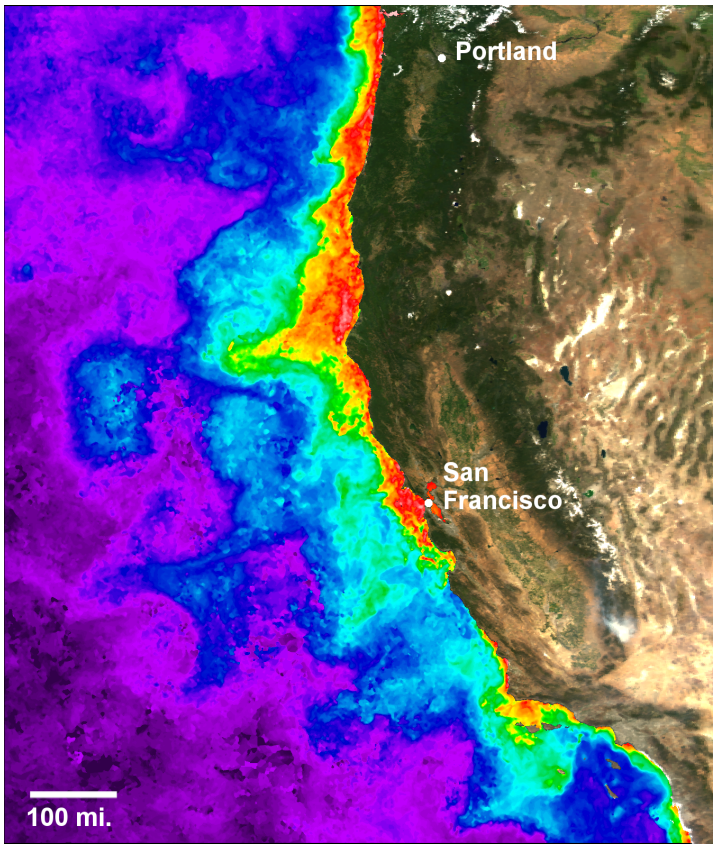



**Fig. S1.**
Extent of the Pacific Northwest *Pseudo-nitzschia* bloom in late-spring 2015. Gradation of colors purple-blue-green-yellow-red-pink in ocean areas shows increasing chlorophyll concentration towards the coast. Land is filled with quasi-true color data from MODIS-Aqua. Ocean chlorophyll is merged with data from MODIS-Aqua, MODIS-Terra and VIIRS-NPP and then composited over 15 days (May 1st – May 15th 2015). White areas are clouds. Image was processed using methods described previously (*49*). Satellite data courtesy of NASA Ocean Biology Processing Group (OBPG) and LAADS-DAAC. Image courtesy of Dr. Mati Kahru, Scripps Institution of Oceanography.


**Fig. S2.**

Heat map showing differential expression profiles of the 483 transcripts upregulated under phosphate limitation. Log2FC values come from DE analysis of these transcripts with respect to a variety of increasing $pCO_2$ conditions (low = 220 ppm, med = 400 ppm, high = 730 ppm) with phosphate held constant (high phosphate = 20 mM , low phosphate = 0.5 mM). Ordering of the rows on the heatmap is consistent with **Table S2**.

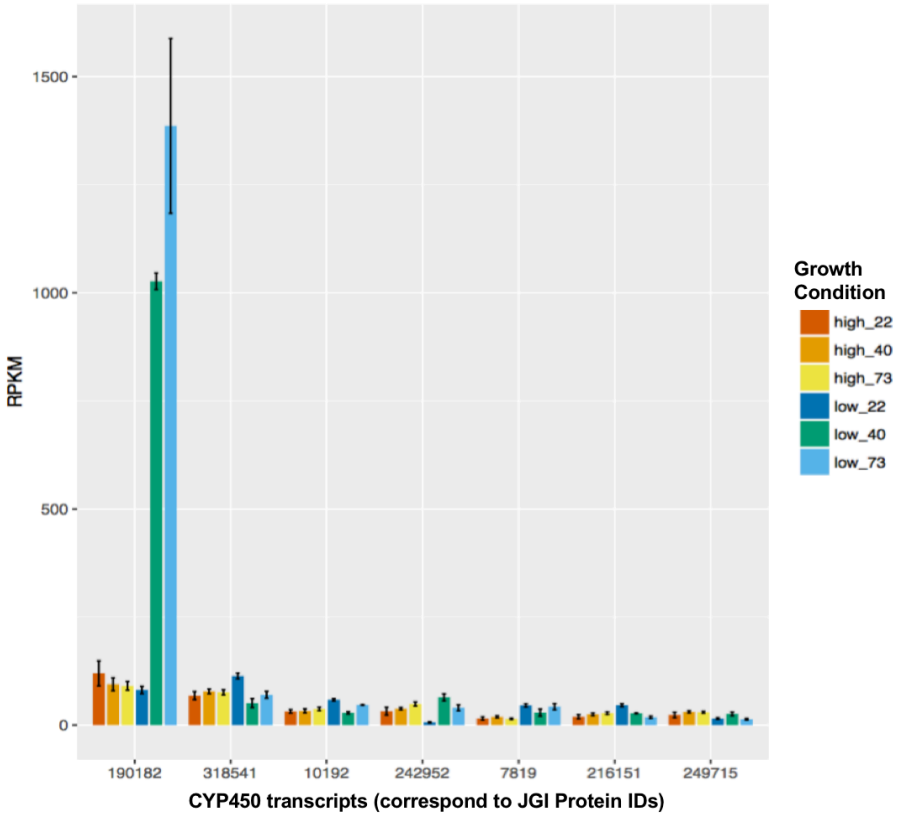



**Fig. S3.**

Histogram for average expression (RPKM) of the top 7 most highly expressed CYP450 genes in the *P. multiseries* transcriptome under the culturing conditions of high (20 μM) or low (0.5 μM) phosphate as well as 220 ppm, 400 ppm, or 730 ppm $pCO_2$. RPKM for each culturing condition was averaged across n=3 rRNA-depleted libraries, with error bars representing standard error. Of the other 13 CYP450 transcripts, trends in RPKM expression were either highly similar to those transcripts represented above (with PID: 190182 *dabD* as an exception) or were not expressed at a high enough level to be visualized on this chart. Full RPKM and differential expression analysis is available in **Table S4**.

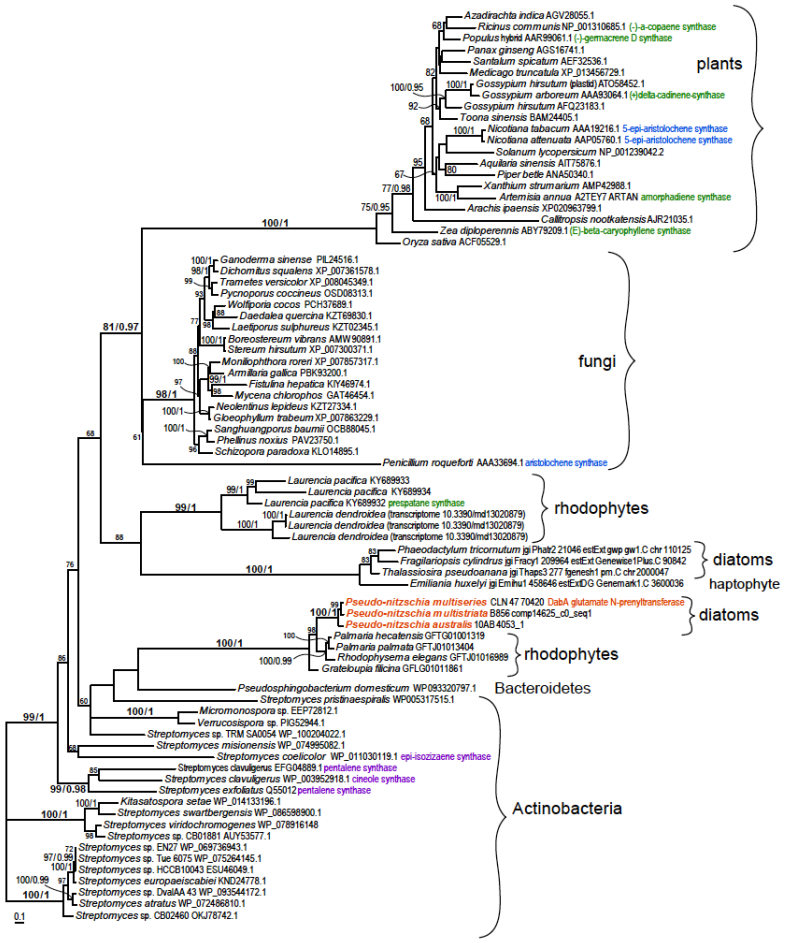



**Fig. S4.**

Phylogeny for *dabA*, red algal homologues and several putative and characterized sesquiterpene synthases. Due to the divergent amino acid sequence of DabA, candidate proteins were selected to represent a variety of diverse putative sesquiterpene synthases and were not chosen based on sequence similarity. Sequences were aligned using the CLUSTAL algorithm in frame of SeaView program (*50*). Alignment was manually edited by exclusion of gaps and ambiguously aligned regions. Maximum likelihood phylogenetic analysis was performed using IQTree v1.5.3 (*51*) under gamma-corrected LG4M matrix, determined as the best-fit model by the -TEST mode of IQTree (Akaike information criterion). The best topology and branch support values were inferred from 1,000 ultrafast bootstrap replicates (*52*). In parallel, Bayesian inference under the gamma-corrected LG matrix (4 gamma categories) was performed in PhyloBayes v3.3b (*53*).

The best topology was inferred after two independent Markov chain Monte Carlo (MCMC) chains reached 0.2-level convergence and the minimum effective size of the model parameters exceeded 100.

a) b)

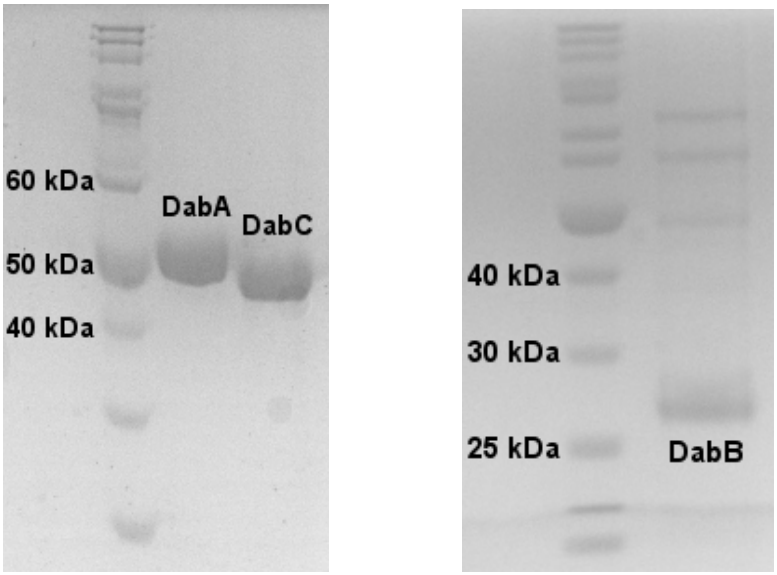
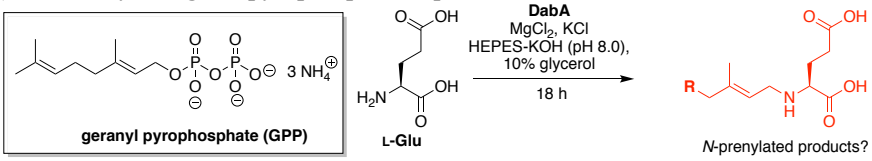
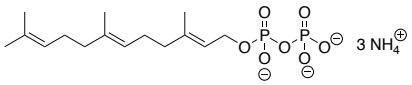



**Fig. S5.**
12% SDS-PAGE loaded with **a)** EZ-Run *Rec* Protein Ladder (Fisher Bioreagents), DabA, and DabC and **b)** EZ-Run *Rec* Protein Ladder and DabB.

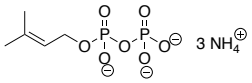
**a) Summary of organic pyrophosphate experiments**




**alternative pyrophosphate substrates tested:**

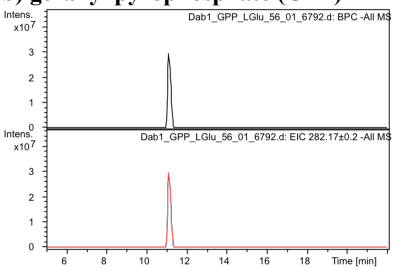



**farnesyl pyrophosphate (FPP)**

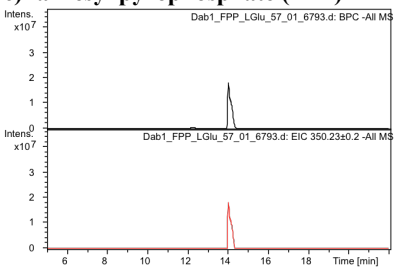



**dimethylallyl pyrophosphate (DMAPP)**

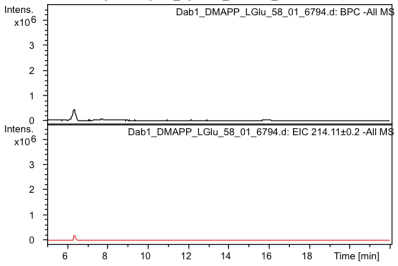
**b) geranyl pyrophosphate (GPP)**




**c) farnesyl pyrophosphate (FPP)**

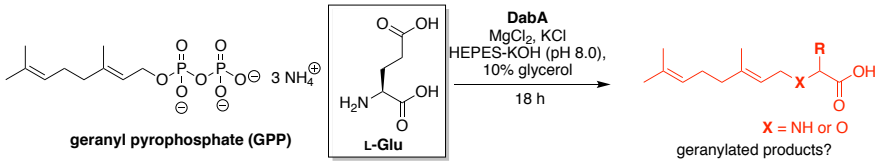



**d) dimethylallyl pyrophosphate (DMAPP)***



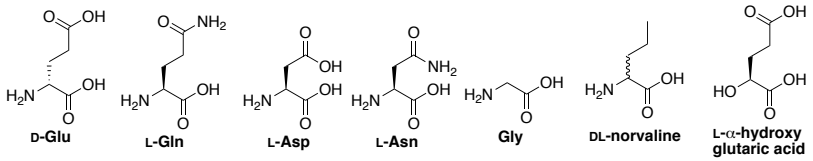

* traces shown at 10x magnification

## e) Summary of amino acid experiments

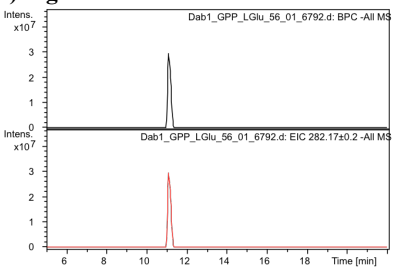



geranyl pyrophosphate (GPP)

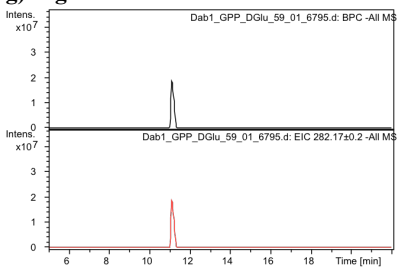
**alternative amino acid substrates tested:**



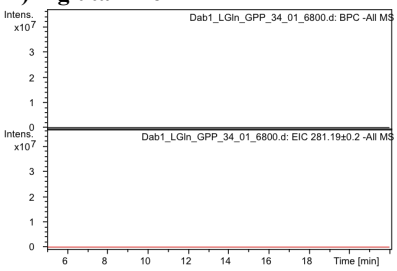

## f) L-glutamic acid



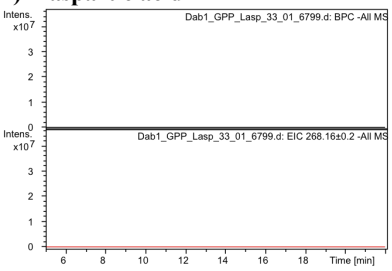

## g) D-glutamic acid



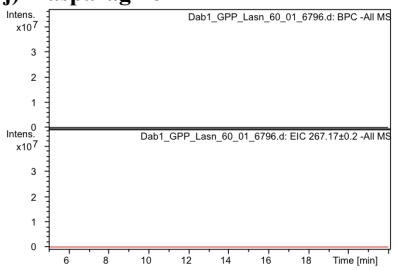

## h) L-glutamine



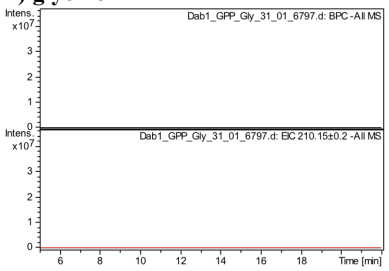

## i) L-aspartic acid



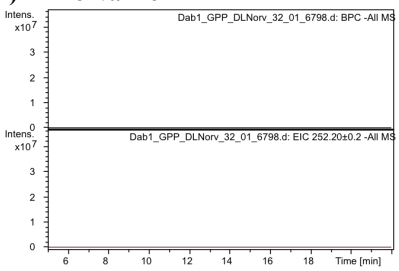
**j) L-asparagine**



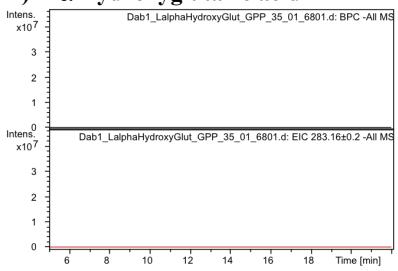

**k) glycine**




**l) DL-norvaline**




**m) L-α-hydroxyglutaric acid**




**n) Marfey's analysis of L-Glu and D-Glu substrate specificity**

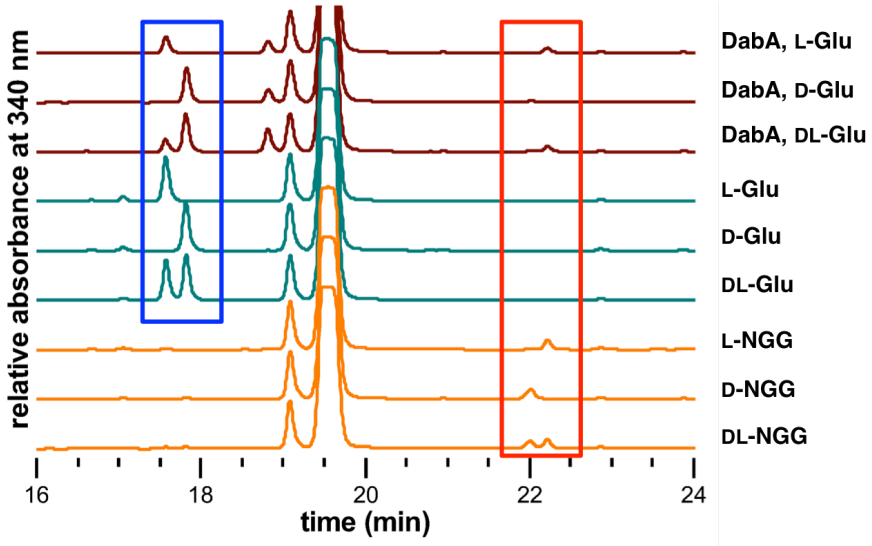
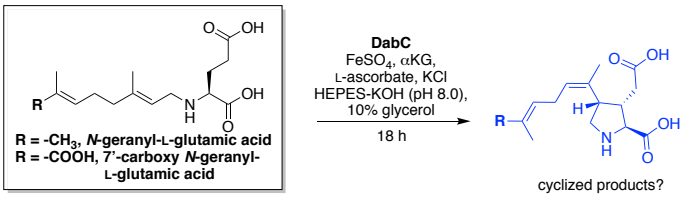
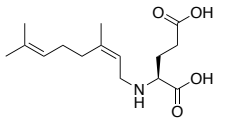


**Fig. S6.**
**a-d)** DabA reactions were set up as previously described using various organic pyrophosphates (0.5 mM) and L-glutamic acid (0.5 mM) and incubated for 18 hours at room temperature. Negative mode LCMS chromatograms of DabA pyrophosphate substrate specificities with L-Glu using LC Method A, showing the reaction base peak chromatograms (BPC, black) and extracted ion chromatograms for anticipated products (EIC ± 0.2 *m/z*, red). **e-m)** DabA reactions were set up as previously described using geranyl pyrophosphate (0.5 mM) and various amino acids/amino acid analogues (0.5 mM) and incubated for 18 hours at room temperature. Negative mode LCMS chromatograms of DabA amino acid/analogue substrate specificities with GPP using LC Method A, showing the reaction base peak chromatograms (BPC, black) and extracted ion chromatograms for anticipated products (EIC ± 0.2 *m/z*, red). **n)** RP-HPLC ($\lambda$ = 340 nm) analyses of L-FDAA derivatized DabA reactions with L-, D-, and racemic Glu, and comparison to similarly derivatized Glu (blue box) and NGG (red box) standards. DabA has a ~4:1 preference for L-Glu over D-Glu when incubated with a racemic mixture.

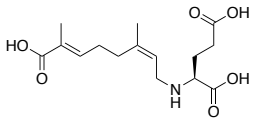
## a) DabC stereochemical selectivity experiments



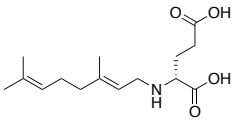

**alternative *N*-prenylated glutamic acid substrates tested:**



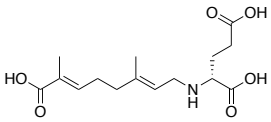
***N*-neryl-L-glutamic acid**



**7'-carboxy *N*-neryl-L-glutamic acid**

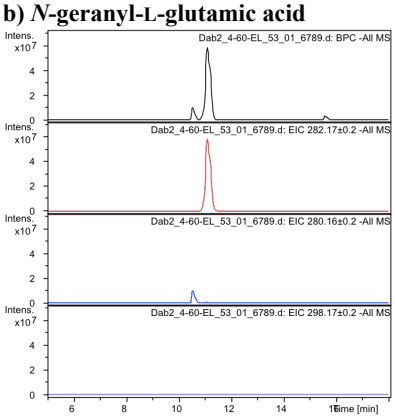


***N*-geranyl-D-glutamic acid**

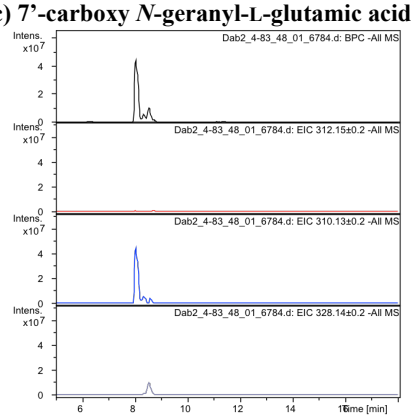


**7'-carboxy *N*-geranyl-D-glutamic acid**

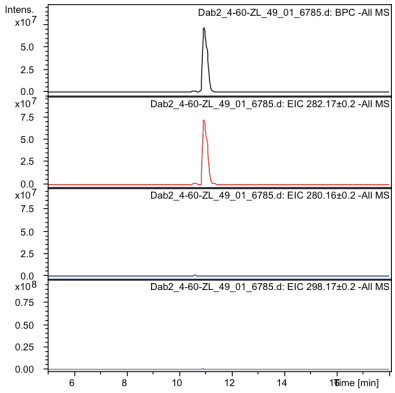
## b) *N*-geranyl-L-glutamic acid



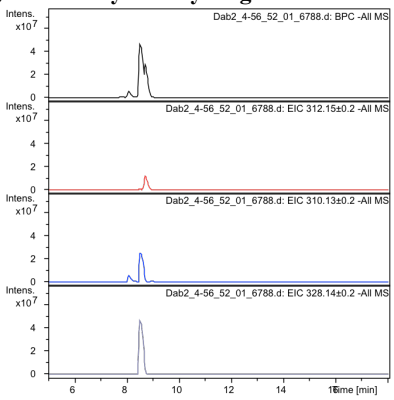

## c) 7'-carboxy *N*-geranyl-L-glutamic acid



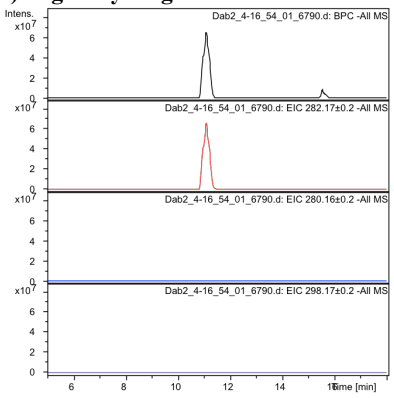
**d) *N*-neryl-L-glutamic acid**




**e) 7'-carboxy *N*-neryl-L-glutamic acid**




**f) *N*-geranyl-D-glutamic acid**




**g) 7'-carboxy *N*-geranyl-D-glutamic acid**

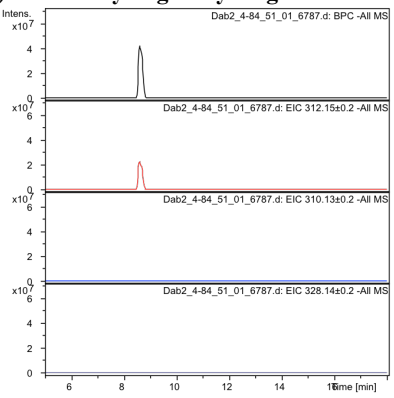



**Fig. S7.**

DabC reactions were set up as previously described using *N*-prenylated glutamic acid analogues (0.5 mM) and incubated for 18 hours at room temperature. Negative mode LCMS chromatograms of DabC substrate specificities with various *N*-prenylated glutamic acids using LC Method A, showing the reaction base peak chromatograms (BPC, black) and extracted ion chromatograms (EIC ± 0.2 *m/z*) for starting materials (red), cyclized (blue), or hydroxylated (grey) products.

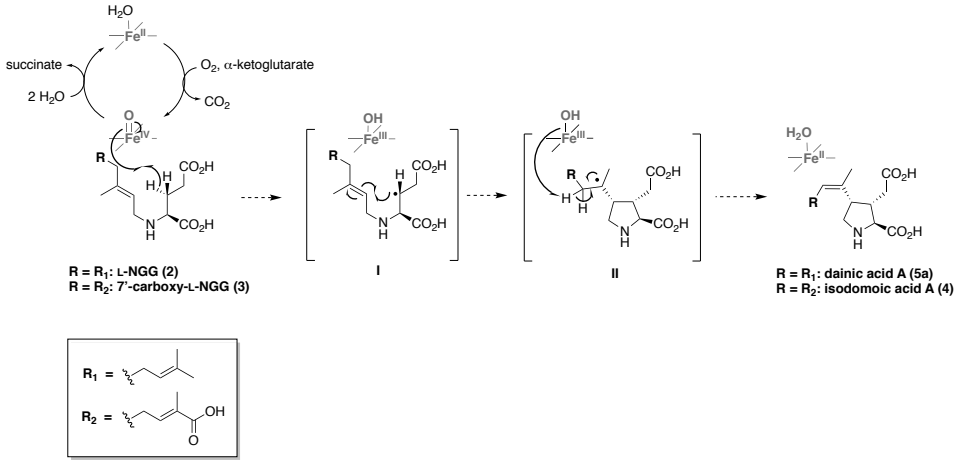
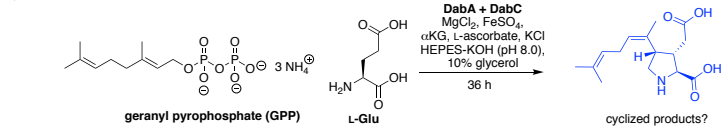


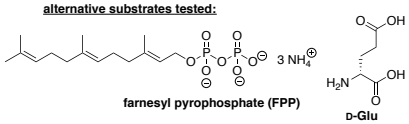

**Fig S8**.

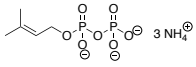
Proposed mechanism for DabC cyclization of *N*-prenylated glutamic acids **2** and **3** to major products dainic acid A (**5a**) and isodomoic acid A (**4**) respectively. Following radical abstraction of a glutamic acid β-hydrogen (intermediate I), a 5-*exo*-*trig* cyclization generates the pyrrolidine ring and tertiary radical (intermediate II). This can be resolved via a hydrogen atom transfer mechanism from the intermediate Fe(III)-OH species (*54*, *55*) to form the *cis*-double bond of the major products.

**a)**

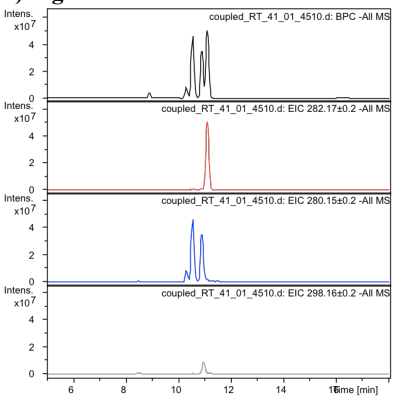



**alternative substrates tested:**

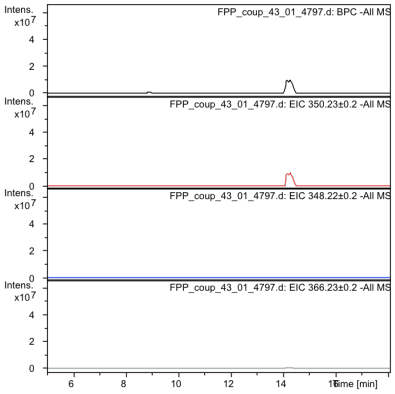


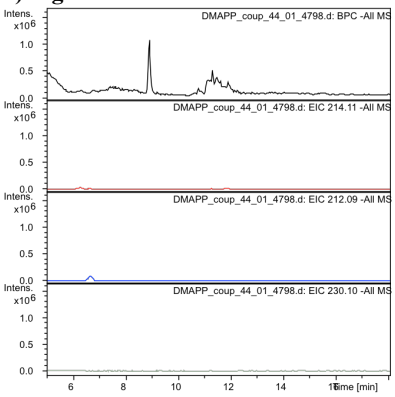

**b) L-glutamic acid + GPP**



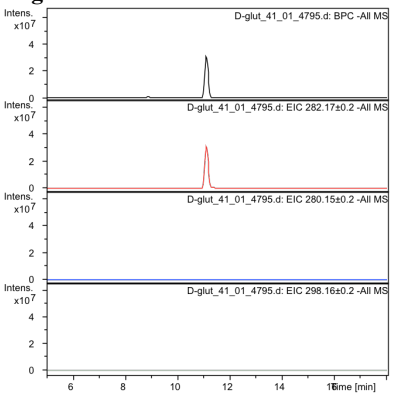

**c) L-glutamic acid + FPP**




**d) L-glutamic acid + DMAPP***




**e) D-glutamic acid + GPP**




*traces shown at 40x magnification

**Fig. S9.**
DabA and DabC coupled reactions were set up as previously described using various organic pyrophosphates (0.5 mM) or amino acids (0.5 mM) that showed individual activity in DabA assays. Negative mode LCMS chromatograms of DabA and DabC coupled reactions after 36 h using LC Method A, showing the reaction base peak chromatograms (BPC, black) and extracted ion chromatograms (EIC ± 0.2 *m/z*) for *N*-prenylated glutamic acid (red), cyclized (blue), or hydroxylated (grey) products.

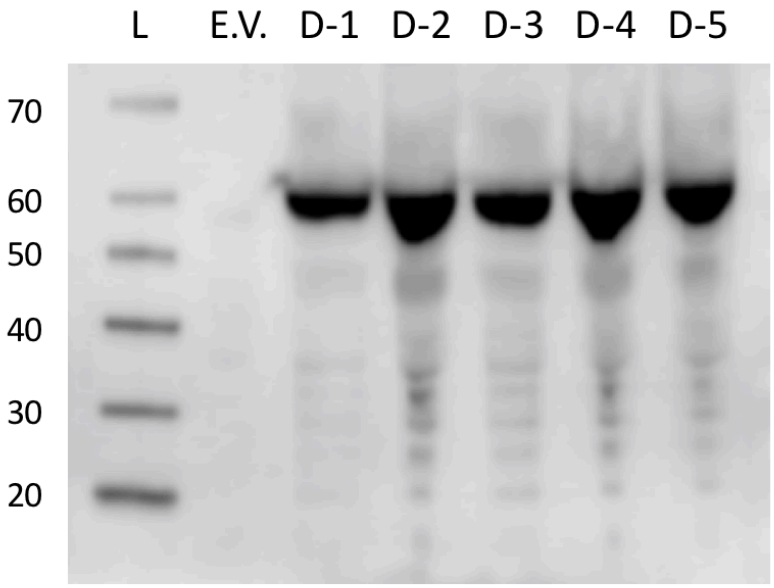



**Fig. S10.**
Western Blot for DabD using anti-FLAG mouse antibody (Sigma Aldrich) on whole cell lysates of *S. cerevisiae* BJ5464 encoding either pBEVY-GL-dabD-CTHF-PmCPR1 (D-1 through D-5) or an empty pBEVY-GL control (E.V). Ladder (L) is the Magic Mark XP Protein Standard (Thermo Fisher)

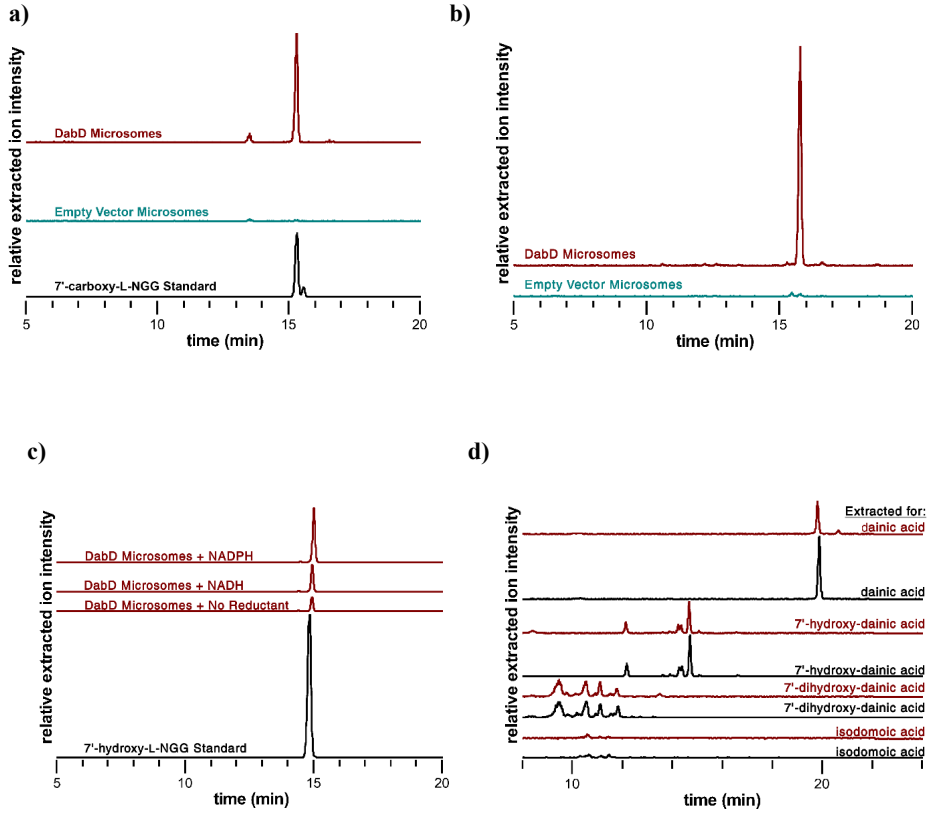



**Fig. S11.**
Negative mode LCMS chromatograms of DabD microsome assays using LC Method B, showing the extracted ion chromatograms (EIC ± 0.01 *m/z*) for **a)** 7'-carboxy-L-NGG and **b)** 7'-dihydroxy-L-NGG. **c)** Negative mode LCMS chromatograms of DabD microsome assays in the presence of different reducing agents. Traces represent the extracted ion chromatograms (EIC ± 0.01 *m/z*) for 7'-hydroxy-L-NGG. **d)** Negative mode LCMS chromatograms of DabD microsome assays with dainic acid substrate. Traces represent the extracted ion chromatograms (EIC ± 0.01 *m/z*) in the presence or absence of DabD microsomes (red and black traces, respectively). *The intensity of the 7'-carboxy-L-NGG standard trace was divided by 100 to better scale it with the experimental results.

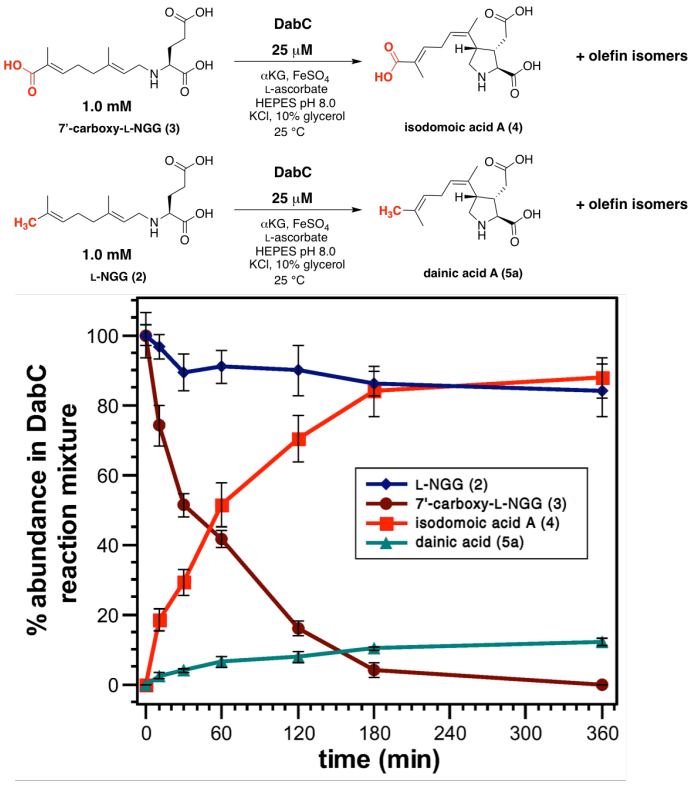



**Fig. S12.**
Relative DabC substrate consumption of *N*-prenylated substrates **2** or **3** and their respective accumulation of major products **5a** or **4** over time.

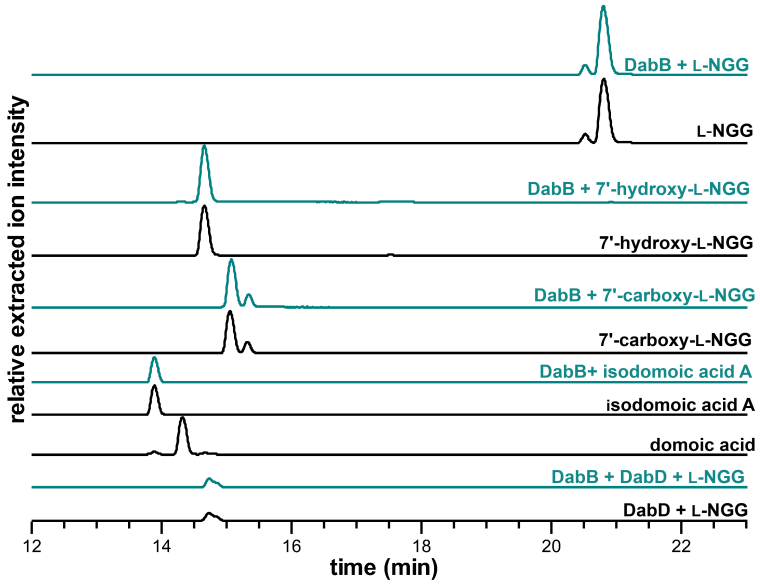



**Fig. S13.**
Negative mode LCMS chromatograms of DabB assays using LC Method B, showing the extracted ion chromatograms (EIC ± 0.1 *m/z*). The DabD containing assays were extracted for the 7'-hydroxy-L-NGG product. In each case, there did not appear to be change in retention time of the substrate, suggesting that no catalysis had occurred.

| JGI Protein ID | InterPro ID | InterPro description | HMMER Pfam description | Phyre2 prediction | Homologs in other non-*Pseudo-nitzschia* diatoms? |
|---|---|---|---|---|---|
| 190182 | IPR001128 | Cytochrome P450 | cytochrome p450 | | No |
| 311210 | | | No hit | Modeled on 3TAI: DNA double-strand break repair nuclease NurA, manganese binding (45.1% confidence, 35% coverage) | No |
| 70420 | IPR008949 | Terpenoid synthase | No hit | Modeled on 1PS1: pentalenene (terpeneoid) synthase (99.3% confidence, 59% coverage) | No |
| 323819 | | | No hit | Terpene Synthase (same ORF as 70420) | No |
| 300256 | IPR011876; IPR015797 | Isopentenyl-diphosphate delta-isomerase, type 1; NUDIX | NUDIX domain | Modeled on 2PNY: isopentenyl-diphosphate delta-isomerase 2 (100% confidence, 70% coverage) | Yes |
| 317971 | IPR000407 | Nucleoside phosphatase GDA1/CD39 | GDA1/CD39 Nucleoside phosphatase | | Yes |
| 65021 | IPR005123 | 2OG-Fe(II) oxygenase | 2OG-Fe(II) oxygenase superfamily | | No |
| 324484 | IPR006025 | Peptidase M, neutral zinc metallopeptidases, zinc-binding site | Peptidase M11 | | Yes |
| 323820 | | | No hit | Modeled on 6BLH: rsv g central conserved region bound to fab cb017.5 (56.7% confidence, 5% coverage) | No |
| 300811 | IPR000911 | Ribosomal protein L11 | Ribosomal L11 | | Yes |
| 252742 | IPR017853 | Glycoside hydrolase, catalytic core | No hit | Modeled on 1H4P: exo-1,3-beta glucanase (99.3% confidence, 51% coverage) | Yes |
| 305406 | | | No hit | Modeled on 5B5Y: metal-binding protein with carbonic anhydrase activity, low CO2 inducible protein PtLCIB | Yes |
| 211191 | IPR006021 | Staphylococcal nuclease (SNase-like) | Staphylococcal nuclease homolog | | Yes |
| 25327 | | | No hit | Modeled on 4QS4: Minor Pilin CofB (25.3% confidence, 20% coverage) | No |
| 312725 | | | No hit | Modeled on 4IW3: prolyl-4-hydroxylase (P4H) in complex with elongation factor Tu (EF-Tu) (98.7% confidence, 46% coverage) | Yes |
| 308281 | IPR003959 | AAA ATPase, core | AAA+-type ATPase | | Yes |
| 300319 | IPR011046 | WD40 repeat-like | G protein beta subunit-like protein | | Yes |
| 180541 | IPR013116 | Acetohydroxy acid | Acetohydroxy acid isomeroreductase, | | Yes |

| | | | | | |
|---|---|---|---|---|---|
| | | isomeroreductase, catalytic | catalytic domain | | |
| 193460 | IPR002303 | Valyl-tRNA synthetase, class Ia | tRNA synthetase class I | | Yes |
| 49425 | IPR002013 | Synaptojanin, N-terminal | SacI homology domain | | Yes |
| 261779 | IPR011701; IPR004737 | Major facilitator superfamily MFS-1; Nitrate transporter | Major facillitator | | Yes |
| 236889 | IPR013105 | Tetratricopeptide TPR2 | No hit | Modeled on 2HR2: tetraticopeptide repeat-like protein (98.2% confidence, 34% coverage) | Yes |
| 327114 | IPR001709; IPR001433 | Flavoprotein pyridine nucleotide cytochrome reductase; Oxidoreductase FAD/NAD(P)-binding | Flavodoxin; Oxidoreductase NAD-binding domain | | Yes |
| 242377 | IPR006271 | Phosphoserine aminotransferase, Methanosarcina type | No hit | Modeled on 3M5U: phosphoserine aminotransferase (100% confidence, 82% coverage) | Yes |
| 249540 | IPR004769 | Adenylosuccinate lyase | Adenylosuccinate lyase C-terminus | | Yes |
| 301408 | IPR001650 | DNA/RNA helicase, C-terminal | Helicase C | | Yes |
| 294204 | IPR005484 | Ribosomal protein L18/L5 | Ribosomal L18 | | Yes |
| 66239 | IPR001344 | Chlorophyll A-B binding protein | Chlorophyll A-B binding protein | | Yes |
| 198571 | IPR001827 | Homeobox protein, antennapedia type | No hit | Modeled on 5UOH: M. tuberculosis serine protease, hydrolase activity (100% confidence, 71% coverage) | Yes |
| 229110 | IPR000086 | NUDIX hydrolase, core | NUDIX hydrolase | Modeled on 5BON: human protein with 8-oxo-dGTPase activity (100% confidence, 84% coverage) | Yes |
| 327386 | IPR001841; IPR014021 | Zinc finger, RING-type; Helicase, superfamily 1 and 2, ATP-binding | Ring finger domain; Helicase conserved C-terminal domain | | No |
| 247811 | IPR006906 | Timeless protein | TIMELESS | | Yes |
| 251103 | IPR001611 | Leucine-rich repeat | Leucine-rich repeat | | No |
| 5808 | IPR010918 | AIR synthase related protein, C-terminal | AIR synthase-related protein, C-terminal domain | | Yes |
| 242642 | IPR001816; IPR002052 | Translation elongation factor EFTs/EF1B; N-6 adenine-specific DNA methylase, | Elongation factor TS | | Yes |

| | | | | | |
|---|---|---|---|---|---|
| | | conserved site | | | |
| 253335 | IPR002035 | von Willebrand factor, type A | von Willebrand factor type A domain | | Yes |
| 40563 | IPR001623 | Heat shock protein DnaJ, N-terminal | DnaJ domain | | |
| 308790 | | | No hit | Modeled on 1PD7: mad1, transcription related (42% confidence, 12% coverage) | No |
| 189252 | IPR001451; IPR003307 | Bacterial transferase hexapeptide repeat\|\|eIF4-gamma/eIF5/eIF 2-epsilon | bacterial transferase hexapeptide; eIF4-gamma/eIF5/eIF2-epsilon | | Yes |
| 255556 | IPR002048 | Calcium-binding EF-hand | No hit | | Yes |
| 182348 | | | No hit | Modeled on 1CKQ: EcoRI endonuclease (43.1% confidence, 5% coverage) | Yes |
| 163723 | | | No hit | Modeled on 5VO5: Lgd-Shrub complex (transport protein) (65.5% confidence, 16% coverage) | Yes |
| 182666 | IPR001164 | Arf GTPase activating protein | Putative GTPase activating protein for Arf | | Yes |

**Fig S14.**
Further analysis of the 43 transcripts that display significant upregulation under low phosphate and high $pCO_2$ in an attempt to identify a putative DA isomerase. The analysis platforms InterPro (*56*), pHMMER (*57*), and Phyre2 (*23*) were used to further bioinformatically characterize the translated peptide sequences. In addition, HMMER searches were able to identify putative homologs in other diatom species for select sequences.

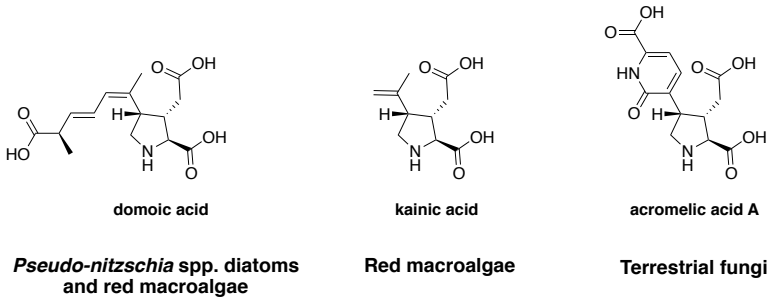



**Fig. S15.**
Kainoid-ring containing metabolites from diverse biological sources (*12*, *24*, *26*, *27*).

| Organism | *dabA* homolog | *dabC* homolog |
|---|---|---|
| *Grateloupia filicina* | GFLG01011861 (incomplete) | GFLG01012884 |
| *Palmaria palmata* | GFTH01004681 | GFTH01007334 |
| *Rhodophysema elegans* | GFTJ01016989 | GFTJ01013404 |
| *Palmaria hecatensis* | GFTG01026770, GFTG01025549, GFTG01001319 (fragments) | GFTG01010004 |

**Fig. S16.**
Table of red algal RNAseq transcript NCBI accession numbers with homology to *dabA* and *dabC*.

>DabA_PsnmuV1.4_scaffold_83-size_174963 PsnmuV1.4_scaffold_83-size_174963:169768..169842,169986..171344 (+ strand) class=CDS length=1434
ATGAAATTTGCAACGTCCATTGTCGCCGCCCTCGCCACTACAGGTGCCGCATTCACCGCGATTCCAAAC
AAGAAGGTTTCTACCGCAGGATTTGGCKTGAAGCACCAGAATCCGCTCTATGCATCGAGTACCTCAGA
CTCCATCAGCAGCCTTACGGCGGAGAGCCCATCTGAAGTCCTTTCTCGAGTCAAAAATACTGGGCTCA
CGTTGACCAACCCCAACGACTTGTACTGGATGGTCGACTTCTGCAAGGAAAAATACTACAACAAAGGC
GACTACTACTACCCCATCAAGACCGTGTGTGATGGTGAGAGTATCGACGTSAAGTTTTATTGCCCATTC
GAGCCCARCCTCAGCCCTCATTATCTCGAGCTCTACGGCTCCCGCGATCAGCGTGCYTCGATGTATGAG
ACAACCATGGAAAAATAYAACCGCATCAACTGCGAAAAAACCTCTGCAATTTGCACMCCTTACTCCA
GTTATGGCGACACCCAGATTGTCGCATACTTTTACTCTATGATGTATTACATCAATGATCAAACCGCCC
ATCTCAAACTACCCGAGAATCAGATTGAAGCGGAKCTCGTCGACATTTTGAACGACGACATTTTGATT
TACTTGAACGAGTTTCTTTCTATCTTCGAACCYAAGAATGACGAAGACTTGGAACGTATTTGGGACTTC
TTGGATTTCTATCAGCCATATTACAACAAGGTTGACGGCAAGATTGTCTTGGACGAGAAATACCAART
CAGAACCCCATCTCAGATGCCCCTAATCAAGACAATCTGYGAGTACATCTCGGAAAAYTTTGCMARCG
AAAAGAATATCACCCAAGTGATTTGGGAGGTTATCCGGTACATCAAGGGAGTRAAGGACGAAATTCA
CATCAGAGGCGACAAGAGTTTCACTCTGTCCCTTCAAGAGTACGAYGACTTCCGTGACAAGGTTACTG
CRAGTCCAATGGCGCAYGCTGTGTCCGATTTGACGCACGAGCGCTTTACCTACAAGGCATACACCAAC
CCACTTTTYATGGAGTTGGAGRACAGGTGTTCCGAGATCATCACATACTTCAACGATGTGTGCACAAG
CGAYAGAGAGCGATTGGATGAAGATCCCTTCAACTCGGTCTTCATTCTCATGGATCTTGACCCCAGTTT
GAATTTCGCAAAATCCTGCGATCTCGTGGTTCARCATGCCYAYGACAAGATGCAGGCATTCTTGAAGC
TCAAGGATGAAATTCTCGAATCGGCCTCGGACGAAGAAGAACGTTTGGCACTTGCTCAGATGATCAAG
ACYAGAGAAGATTCCTTGATTGGATATGTGTTGCATGAAGTCTGCTGTGTCGAAGAGGGSTATGCTCG
TGATCACAAGCCTTTGATGAAAGCGTTTTTGGAGGAGGAATTGGCCAAGTCGCTCTCCGAAAAGGTAT
AA

>DabB_>LN865247.1 Pseudo-nitzschia multistriata strain B856 genome assembly, scaffold: PsnmuV1.4_scaffold_83-size_174963:168137..168859
ATGTTGGTTTGCTTTGTTCCAGGTATTGAAGAAGGAAAATCCTATNCTGCAAAGGATCATTCTGGGCAC
TCAGGGAGCAGCTATCTGAAGAATTTTCTGGCTTTCTTCGCTGGCGCCGCATTCATGGCTGCGGCAATG
TATCTCAATAGTGGCAACTCCATTTCCGTTACCACTTCGGATTTAGTTGGTTCCTTCACCTCATCTGACT
CGCCTGCTATCGATATTGGAATCAATCCAACCCGTCCAAGTGTGGATGCCCGGCAAGACTTTACCAGT
GAAGGTTGGGGTAACCAAACGAGAAATGATGGGTGGCCGAGTTTTCATTTCGAAGTGTTTCACTTTTG
CCCTCACTTGATTAATTTTATGTACGAGAATCAGGTCCATCCACAGTCCCGTGTCGACCAAAGTAACGT
TCTTTATGATGCACAGCATCCTGGTGCTCTAGCGGAAGATGGTTTCGCCGAAGTCTTGCAACAAACNCT
CCCAAAATTCTATGATTACCCCGTAGGATACAGACTCAATGACCGACCCGCAGAAATCTCTGCATATT
GGACGCTCCACATGCGAATTGCACGCGGCAGCGAAGACTCGCATTGGCATGACTTTGCATCGTCTCTA
CCNGAAGGAATTAACAGAATTTATGTGAAAAAGCCCCTACGATGGAGTGGGGATCAGGTCACGCTCTT
CAGCAGCCAAACCGGCTATAATATTCTCGTGTACTAG

>DabC_PsnmuV1.4_scaffold_83-size_174963 PsnmuV1.4_scaffold_83-size_174963:166585..167709 (+ strand) class=CDS length=1125
ATGACAATGAAAGTWGTTAACAACGAAACTGCCRCTTTGACCGCAAACGAGCACGATGTGCAGGTCA
ACAAGGGCAAAACTCTCGAGCAGAGATTTCCACCTCTCGATGGTGAAAAGTTGAAATGGTTTCCGAAG
TCATCTTTGCCTGATGAGATYCCTGCCATTGATATCAGCAAAGTCAACACGAAAGAAGAGCTGGAACA
ATTTTTGGTAGACATCCGCAAGTCAGGACTTTTCTACATCGTCAACCACGGTGTTCCCGAAGACGTCTC
AATTAACGTTTACAACGCCTTCAGAGAATTTCTCTCGTCCACCACCGAAGAAGAGAGAATGAAGTATT
ACACGGACACTCATTTCCAAAATGGTGGATACGTCCCCTTCCAAGGCTCTTCCATTCGCGGAGGAAAT
TTAGGCAAGCCGCAGAAGGACCACGTCGTGAAATACTTTTGGAGAGGGCCTCAAGTTGTTAATAGGAC
CCCMAGCGAGAAATTCACCAAGGCACACGATGYCCATCATACTGAGACCTTCAACGTAGCGGAGAAG
GTGATTAGAACCATTTTCAAGGCCTTAAAGCTTCGCTTCCCTGACTTCGATCCTATGGAATTCGAAGAC
ACTATTAATTCAAAGAAGATGTTCTTCACCAATCGTGTTTACCCTCAGGCTGAAAAGAGTGACGAAGA
AACTATYACTCATCGTCTCGTTCCCCACTTAGATACCAGCTTCATCACGTTAGCGAATCAAGTTCCTGC
CGACAATGGCTTCCAAGGTCTGTTCGTTGAGACTGGAGATGGGAAAAAGGTGAAGGTCCCTGGCATCC
GTAACAGCTATTTGGTCTTCATCGGTCAAAGTCTGTCTTTCCTTACGAAGAACTACCTTCCATCGGCTC
TCCATGGTGTCGACAAGCCTCCCAAGGAAATGTTCGAAGGAAGCGAAAGATCCTCGTTGATCACTTTT
TATGAACCTGCCGAAATCATTATTCCATCCAAGAACATCAATCCCAACCCAGATGAGACCACTGATTC

GTGTCCGTTTTATGATTCTATTGGATTGGATGTAAATGATCCCAAGGGCACTACCTGGGATTTTGTGAA
GAACAAATTCATTACCGGATATTACGCGGATTAA

>DabD_PsnmuV1.4_scaffold_83-size_174963 PsnmuV1.4_scaffold_83-size_174963:164018..165697 (- strand)
class=CDS length=1680
ATGAACTTTGCGACAGCTACTTTATGTGCCGGACTMGCCGTGTCTGGGAAGTTCTTATCGAACTACTAC
CAATTCCTCTCTTCTTCACAGAGCGATGAAGTCGGCTTTCTCGAAACAGTTGACAAGCAGATGGCGAG
CAATGCTGCAGTCGCAGGCGTCGGTCTTGGTCTTATCTTCTACTGTGTCCTCAGCTTTGCAAGAAGCTA
CTACAAGTTTCGCTTCTCTCCCCTTCGTGACGCACCAGGATTTGGTCCGAAATCTTTTGTGTATGGAAT
GTTCTATGAATTCCTCGAGGCGCCATTCATGGAACCTCCAATCGAAGCCCTGAAGAAGCTACGTAAAG
GAGGCAAAGARGTCCCATTTCTTGCCTACACAACTCTTTTTGGAAGCCAAAGGCTTTTGTTATTGGACT
GCGACTTGGTGAAGCACGTCTTTACTTCGCCATCAGGTAGAGATCCAATGAGATACCCCAAGCACTAC
GTTTACTTGAGAGAGGTTGTGGGAGATGGTCTCGTCGTCGTTGAGGGCACTGAATGGAGTCGCCATCG
TCGAATCATCCAGCCKGCGTTTCAATCTGTTTTTCTGAAAGATGCTATRAGTATGGTCGTCCCCGCCTT
GGTAGAGAATCTTGTAGAGGTATGGAAGAAGACTGCTGGGGCAACGATCAATCTGAATGCCCATCTTT
CGCTAATCACTCTTGACGTGATTGGAAAAGTTGCTTTTTCGCACGAGTTCAATGCTTCCAAACTGCTCA
ACGAATGGGCCGAATGTCCRGATAAGGAARTGGGGGAGGTTGACGATCCGTTGATTTCATCCATCGGT
GATGCCTTTTCCTCYAGCCCTCTCAAATTAATGCTTACCGTGTTGAAGCAGCCATGGCTCGAGAAATAC
TTGAGTCCTGCCTTCAGAACTTCACGCAATTTACTGAACAAGGCAGCAGATGACATCGTACAGAATGC
CAAGAACATCGATGATCCTAAACGACGAAGCGTATTGAACCTGATGATGGAAGCTAAGGACGATGAA
TCCAGCAAGGCTCGCAATAAGTTGACAGACACCGAGCTCCGCGATGAAGTRAAAACATTTCTTGTTGC
TGGCCACGAAACAACYTCGACTTGGGTTCATTGGGCTTTATATGTACTTGCGATTCGCCCAGATCTCCA
AGATAAGGTACACGCTGACATCATGAAGCACGCGTCCGACGACAAGACTATTTTACTGGAACAAGCC
GATCAAATGGAATATCTGTGGGCCTTCATGAACGAAACTCTTCGTTTGTACTCACCTCTCGGATTGATC
AGCCGTGTCACTCACAAGGAAGAAAACTTCAAGGGCTACAKGATCCCAGTGGGAACGAATCTGAGAA
TTCCTATTCACTTGATTCATAGACATCCTGATCATTGGAAGGACCCCGAAGAGTTTCTACCCGAACGTT
GGTTCGACAAAGAAGAAACGAGCAAGAGACACAAGTTCGCTTTCGTCCCCTTTGCAGCAGGGGGAAG
AAACTGCATCGGTCAGCGTTTCGCTACGATGGAGGCCAAAATTATACTCGCAAATGTTGCCAAGAATT
TCGAAATCCATTTGGCTGATTCCATGAAGGGCAAGAAAATTACATTCAGTAATTTCATTTCTTTGAAGT
GCAAACCCGAGATTGAAATTCGGGTGACACCTCGAGAATAG

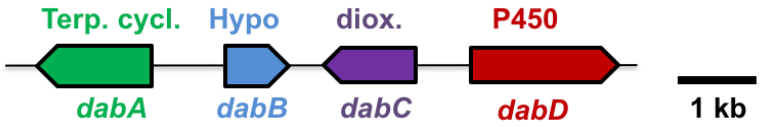



**Fig. S17.**
Sequences from the publicly available *P. multistriata* genome (*30*).

>DabA_ Pseudo_nitzschia-australis-10249_10_AB-20140214|4053_1
CTACCCTTGCCACCACCGGTGCAGCATTCACTGCAAATCCCCAAAAGGCGTTCACCGCAGGAGCCAGC
GCTAGGCAACAGGTTTCGCTCAATGCATTGAGCACCCTTAGCTCCATCGGCAGCCTTACTGCAGAAAG
CAAATCGGAAATCCTTTCTCGTGTTAAGGAGAACGGACTGACCTTGACGAATCCTAAGGATTTGTACT
GGATGGTTGACCATTTTAAGGAGAATTACTACGACGAAGGTGACTACTACTATCCCATCAAGACGGTG
TGTGATGGTGAAAGCATCGATGTCAAGTTCTACTGTCCTTTCGAGCCCAGCCTCAGCCCGCATTACCTT
CAGTTGTACGGAAGTCGTGACGAGCGTGCTTCGATTTACGAAACGACCATGGCGAAATACAACAAGAT
CAACAGCGAGAAGACTTCCGCGATTTGCACTCCATACTCCAGTTATGGCGACACCCAGATTATCGCGT
ACTTTTACTCCATGATGTATTACATCAACGATCAAACCGCCCATCTCAAGCTTCCCGAAGGAGAGATC
GAAGCTGAACTCATTGACGTCTTGAACGACGATATTTTGATTTACCTCAACGAATTCTTGAGTGTCTTT
GAACCTGAGGACGACGCTGACTTTGAACGCATTTGGGATTTCTTGGAGTTTTACCAGCCATACTTTAAC
AAGGTTGACGGAAAGATTGTGTTGGACGAGAAATACCAGAAGGGAACCCCATCGCAGATGCCCCTGA
TCCAGACCATCTGTTCGTACATCGCCGAACAATTCGCCCCCAACAAGAACATCACTCAAGTAATTTGG
GAGGTCATCAGATACATCAAGGGTGTTAAGAATGAGATCCAGATCAGAGGCGATAAAGGTTTCACTCT
ATCCCTCCAAGAGTACGACGACTTCCGTGACAAGGTTACCGCCAGTCCAATGGCTCATGCGGTTTCTG
ACTTGACCCACAAACGCTTCACCTACGAAGCATACACCGACCCACTTTTCATGGAGTTGGAGAACCGT
TGCTCCGAAATCATCACCTACTTTAACGACGTGTGCACTAGTGATAGAGAGCGATTGGACGACGATCC
GTTCAACTCGGTCTTCATTCTCATGGATCTTGATCCCACTTTGAACTTTGCCGCTTCGTGCGACCTTGTC
GTTGGTCACGCTTACGAAAAGATGGAGAGATTCTTGGAGCTTAAGGAAGAAATTCTCGTATCCGCAAG
AAGCGAGAAAGAAAAGCTTGCCTTTGCCCAGATGATCAAGACCCGAGAAGATTCGTTGATTGGATATG
TTTTGCACGAAGTTTGCTGTGTCGAGGATGGGTATGCTCGTGATCACAAACCTTTGATGAAAGCATTTT
TGGAGGAGGAGTTGACCGAGGCCCTTGAGATGGCTTAA

>DabB_Pseudo_nitzschia-australis-10249_10_AB-20140214|210_1
ATGAGTGCTAGTAAAAATAGTTATCAGTCAGCTCATGTTGAGTCCTATCGCGTTGATGGAGAATACGT
TCCTACTGAAAGGCTTCGTGATGTGTCCTCGA
AAAGCCCTCTGCTCTTGAATTTCGTGACTTTTTTCATTGGCGCTGCATTCATGGCTATAATCAGTTCAAG
AGGGGAATCAGCTGAAGCTGCTATTGATTT
TGGAATCAATCCAACCCGTCCAGGAGTGGATGCCCGGCAAGGATTTAATACCAATGAAGGATGGGGG
AACCAAACTCGAGATGATGGTTGGCCAAGTTTT
CATTTCGAGGTATTTCACTTTTGCCCTCACATGATTAACTTCTTGTACGAGAACCAGGTTCATCCACAA
TCCAAGATCGACGAAAACAATGTTATTTATG
ACGCGCAGCACAATGGTGCTATAGCGGAAGACGGCTTTGCCGAGGTTTTGCAACAAACCCTCCCCAAG
TTTTATGATTATCCCGTAGGATACAGACTTAA
CGACCGCCCCGCAGAAATTTCTGCGTTTTGGACCCTCCACATTCGAATCGCACGTGGTAGTGCGGACTC
CAATTGGCACGACTTTGCATCCTC
TCTTCCC
GAAGGTATTACCAGAATTTACGTCAAAAGGCCCCTGCGATGGGATGGTGATGGTGATGATGTTGCX

>DabC_ Pseudo_nitzschia-australis-10249_10_AB-20140214|55196_1
ATGACTGTCCAAGTAATTAACAACGAGACCAACGTGTTGACCCCCAACGAGGACGACGTCCAAATCA
ACAAGGGCAAAACCTTGGAGACGAGGTTCCCCCCCCTCAAAGCCGACGACCTCAAGTGGTTTCCCCAT
GATTCTCTGCCCGCAGAGATCCCTGCTATTGATATCAGCAAAGTAAACACGAAGGAGGAATTGGAACA
GTTTTTGGTTGACATCCGCAAATCGGGACTTTTCTACATTGTGAACCACGGTGTCCCCGAAGAGGTCTC
CATTAATGTTTACAACGCATTCAGAGAGTTTCTCTCCACCACCACCGAGGAAGAAAGGATGAAGTACT
ACACCGATACGCATTTCCAAAACGGTGGATACGTCCCCTTCCAGGGCTCCTCTATTCGCGGAGGAAAC
TTGGGAAAGCCGCAAAAGGACCACGTCATCAAGTATTTTTGGAGAGGACCACAGGTCGTCAACAGAA
CTCCTAGCGAGAGTTTCACCAAGGCCCATGACGATCACCACACCGAGACCTTCAACGTAGCTGAGAAG
GTTATTAGAACGATCTTCAAGGCTCTCAAACTTCGTTTCCCAGACTTCGACCCGATGGAGTTCGAGAAC
ACGATCAATTCCAAGAAGATGTTCTTCACCAACCGCATCTACCCGCAAGCCGAAAAGTCCGACGAGGA
AGAGCTCACCCATCGTCTTGTTCCCCACCTGGACACCAGTTTCATTACCTTGGCAAACCAGGTTCCCGC
CGATAATCGTTTCCAGGGTCTGTTCGTTGAGACCGGCGATGGGAAAAAGGTCCCCGTCCCGGGCATCC
GCAACAGTTACTTGGTCTTCATTGGTCAGAGTTTGTCGTTCCTCACCAAAAATTACCTTCCATCGGCTC
TTCACGGTGTGGACAAGCCACCCAAGGAAATCTTTGAGGGAAGCGAGAGATCCTCGTTGATCACTTTT
TACGAACCCGCGGAGATCATTATTCCATCTAAGAACATTAATCCCAACCCAGACGAGGTCACCGTGTC

GTGTCCATTCTACGATTCCATCGGCTTGAATGTTAACGACCCAGAGGGTACTACGTGGGATTTTGTCAA
GAACAAATTCATTAAGGGATACTATGCGGATTAA

>DabD_ Pseudo_nitzschia-australis-10249_10_AB-20140214|13962_1
ATGAGCAGCATTAATGTTGTAACGGCTCTTGTATCCATCGGGCTGGCCGTATCCGCTAAATTCGCATCG
AGTTACCACGATTTCCTTTCTTCTCCCAAGAGCGACGAAGTTGGATTTTTGGAAACCCTGGACATCACG
AACACGACGAGCATCGATGCGGTTGCTCACGTCGTTCTCGGTCTTGTCTCGTACTTTGTCCTTAGCTTT
GCAAGGAGCTACTACAAGTTTCGCTTTTCTCCCCTGCGTGATGCTCCGGGCTTTGGTCCGAGATCCTTC
GTGTACGGAATGTTCTATGAATTTCTTGAGGCGCCATTCATGGAACCCCCCATCTTGGCCCTCAAGAAG
CTTCGAGAAGGAGGCAAAGAAATTCCGTTTCTGGCCTACACGACACTGTTTGGAAGCCAACGGCTCTT
GTTGTTGGACTGCGACCTGATCAAGCACGTCTTTACGGCACCGTCCGGGAAGGATCCCATGAGATACC
CCAAGCACTACGTTTACTTGAGGGAGGTAGTTGGAGACGGTCTCGTAGTCGTTGAAGGTCAGGAATGG
AGTCGCCATCGTCGTATCATTCAGCCGGCTTTCCAATCCATGTTTCTCAAAGATGCCATCGGTATGGTC
GTCCCCGCCTTGGTAGAAAATCTCGTGAACGTATGGAAGAAGACCGCTGGGACGACAATCAACATGA
ACGCCCACCTTTCGCTCATCACCCTCGACGTGATTGGAAAGGTAGCGTTTTCGCACGAATTCAACGCTA
GCAAGCTACTCAACCAATGGGCCGAGTCTCCCGACAAAGAATTGGGAGAGGTCGACGATCCCTTGATT
TCGTCCATTGGCGGTTCGTTTTCCTCCAGCCCCCTCAAATTGATGCTCACCGTACTGAAACTTCCGTGG
CTCGAAAAGCACCTCAGTCCCTCGTTCCGAACCACGCGCAACCTACTCAACAAAGCAGCTGACGATAT
TGTGCAAAATGCCAGGAATATCAAAGATCCCTCGAGGCGCAGTGTGTTGAACCTAATGATGGAAGCA
AAGGACGGCGAATCCAGCAAGGCCCGCAATCAGTTGACGGACACGGAGCTTCGAGACGAGGTCAAGA
CATTCCTGGTTGCAGGCCACGAAACAACCTCCACCTGGTCCCACTGGGCGCTCTACGTGCTCGCGATCC
GCCCGGATCTCCAAGAAAAGGTGTACGCCGATGTCATGAAGCACGCACCCCCTAACGATGAGACTATT
GTGCTGGAACAAGCCGACCAAATGGAATACATGTGGGCGTTCATGAACGAAACGCTGCGTTTGTACTC
GCCCCTCGGCCTGATCAGCCGCGTTACCCACCAGGAAGAAAACTTCAAGGGATACACGATTCCGAAAG
GCACCAACCTACGAATTCCGATCCATTTGATTCACAGACATCCCGATCACTGGAAGGACCCCGAGGTT
TTCCGACCCGAGCGTTGGTTCGACAAGGAAGAAACCAGCAAGAGACACAAGTTCGCGTTCATCCCATT
TGCTGCCGGTGGAAGAAACTGCATCGGTCAGCGTTTTGCTACCATGGAGGCCAAGATCATCGTCGCGA
ACGTTGCCAAGAATTTCAAGATTCACTTGGCGGATTCGATGAAGGGCAAGGAAATCACCTTTAGCAAC
TTCATTTCTTTGAAGTGCAACCCCGAGGTTGAAATTCGTGTCGAGGCCCGAAAATAA

**Fig S18.**
Sequences from the publicly available *P. australis* RNAseq dataset (*31*).

| ***Pseudo-nitzschia*** **species/strain/variety** | ***dab*** **gene expression** | **MMETSP Library IDs** |
|---|---|---|
| *Pseudo-nitzschia arenysensis* | No | MMETSP0329 |
| *Pseudo-nitzschia australis, Strain 10249 10 AB* | Yes, *dabA-D* all expressed | MMETSP0139, MMETSP0140, MMETSP0141, MMETSP0142 |
| *Pseudo-nitzschia delicatissima B596* | No | MMETSP0327 |
| *Pseudo-nitzschia delicatissima, Strain UNC1205* | No | MMETSP1432 |
| *Pseudo-nitzschia fraudulenta, Strain WWA7* | No | MMETSP0850, MMETSP0851, MMETSP0852, MMETSP0853 |
| *Pseudo-nitzschia heimii, Strain UNC1101* | No | MMETSP1423 |
| *Pseudo-nitzschia pungens* | No | MMETSP1061 |
| *Pseudo-nitzschia pungens cf. cingulata* | No | MMETSP1060 |

**Fig. S19**.
Table of *Pseudo-nitzschia* transcriptomes analyzed for *dab* gene expression. All transcriptomes were generated as a part of the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) and are available through https://www.imicrobe.us/ using the library IDs in the table above (*31*).

**Primer Table**

| | |
|---|---|
| DabA_001 | ATG AAG TTT GCA ACA TCC ATC GTC |
| DabA_002 | TCA ATT GAG GCG AAC GGA CT |
| DabC_001 | ATGACTGTGGCAATAAATAACGA |
| DabC_002 | CTA ATC AGC GTA GTA TCC GG |
| DabD_001 | ATG AAC ATT GCA ACA GCT AGT TT |
| DabD_002 | CTA TCC ACG GGG TGT AAC ATA |
| DabA_p28_001 | GGTGCCGCGCGGCAGCCATATG ATG TCG CAC CCA AGC CAG CTC AAT GCC |
| DabA_p28_002 | GCTCGAGTGCGGCCGCAAGCTT TCA ATT GAG GCG AAC GGA CTC AGA CTC AAC CGG |
| DabB_MBP_001 | CTGTACTTCCAATCCGGATCCATGATGGCTGGTAGAAAAGGA TACCAATCAGTTATG |
| DabB_MBP_002 | GGTGGTGGTGGTGCTCGAG TCAGTAGACAAGAATGTTGTAGCCAGTTTGG |
| DabA_seq_001 | CTATATCAATGATCAAACTGCTCATCTCAAA |
| pET28_BB_001 | CATATGGCTGCCGCGCGGCACCAGG |
| pET28_BB_002 | AAGCTTGGATCCGAATTCCCCAGATCTAAAGTTTTGTCGTCT TTCC |
| pET28-MBP_BB_001 | GGATCCGGATTGGAAGTACAGGTTCTCAGATCC |
| pET28-MBP_BB_002 | TGAGATCCGGCTGCTAACAAAGCCCGAAAGG |
| PtpBAD-CTHF_001 | CTACGGGGTCTGACGCTCAGTTATGACAACTTGACGGCTACA TCATTCAC |
| PtpBAD-CTHF_002 | CCAAAAAAACGGGTATGGAGAAACAGTAGAGA |
| PtpBAD-CTHF_003 | GGAACTGCCAGGCATCAAATAAAACGAAAG |
| PtpBAD-CTHF_004 | ATTCTTGAAGACGAAAGGGCCTCGTGATACGCAAAAAGGCC ATCCGTCAGGATG |
| PtpBAD-CTHF_005 | AACAGAATTTGCCTGGCGGCGGAACTGCCAGGCATCAAATA AAACGAAAG |
| PtpBAD-CTHF_006 | TAGCCGTCAAGTTGTCATAACTGAGCGTCAGACCCCGTAGAA AAGAT |
| PtpBAD-CTHF_007 | ACCCGTTTTTTTGGATGGAGTGAAAATGGTGAGCAAGGGCG AGGAGC |
| PtpBAD-CTHF_008 | ATTTGATGCCTGGCAGTTCCTTACTTGTACAGCTCGTCCATGC CGAG |
| PtpBAD-CTHF_ultramer001 | CTCCATACCCGTTTTTTTGGGCTAACAGGAGGAATTAACCTC GAGCGCTGGTCCCTCGCGGTAGCCACCACCACCATCACCACG ACTACAAGGATGATGACGATAAGTGAAACAGAATTTGCCTG GCGGC |
| PtpBAD-CTHF_ultramer002 | GCCGCCAGGCAAATTCTGTTTCACTTATCGTCATCATCCTTGT AGTCGTGGTGATGGTGGTGGTGGCTACCGCGAGGGACCAGC GCTCGAGGTTAATTCCTCCTGTTAGCCCAAAAAAACGGGTAT GGAG |
| PtpBAD- | AAACCAATTGTCCATATTGCATCAGACATTGCC |

| | |
|---|---|
| CTHF_col001 | |
| PtpBAD-CTHF_col002 | CTTTCGTTTTATTTGATGCCTGGCAGTTCC |
| DabC-CTHF_001 | GCTAACAGGAGGAATTAACCATGACTGTGGCAATAAATAAC GAAACCGTTGT |
| DabC-CTHF_002 | CTACCGCGAGGGACCAGCGCATCAGCGTAGTATCCGGTGAT GAACTTGT |
| DabD-CTHF_001 | GCTAACAGGAGGAATTAACCATGAACATTGCAACAGCTAGT TTATGTGCCG |
| DabD-CTHF_002 | CTACCGCGAGGGACCAGCGCTCCACGGGGTGTAACATAAAT TTCAATCTCGG |
| DabCseq_001 | AGAGAATGAAGTATTATACGGACACTCATT |
| DabCseq_002 | CTCCGGTCTCAACGAAAAGAC |
| DabDseq_001 | CGTTTCTCAGGGATGCTATCAC |
| DabDseq_002 | AACAAGGAACGTTTTCACTTCGT |
| ampDabC-D_001 | GAAGTATTATACGGACACTCATTTCC |
| ampDabC-D_002 | CATAGAAGCGAAACGCTGAC |
| ampDabA-C_001 | TTCTTCATGGTCGTCTCGTAA |
| ampDabA-C_002 | CAACCCAGAAGAGACCTCTG |
| OH_DabC-D_001 | TTTGTACAAAAAAGCAGGCTGAAGTATTATACGGACACTCA TTTCC |
| OH_DabC-D_002 | TTTGTACAAGAAAGCTGGGTCATAGAAGCGAAACGCTGAC |
| OH_DabA-C_001 | TTTGTACAAAAAAGCAGGCTTTCTTCATGGTCGTCTCGTAA |
| OH_DabA-C_002 | TTTGTACAAGAAAGCTGGGTCAACCCAGAAGAGACCTCTG |
| DabA-C_seq001 | CGCGAATTCGCATTAAGAATGACTCAC |
| DabA-C_seq002 | ATGTGAAGGGTCCAATATGCAGAA |
| DabA-C_seq003 | CGTTACAACTCAACTGGTAGGCGAT |
| DabA-C_seq004 | TACAGTAGATATGCACCAGCGAACG |
| DabA-C_seq005 | AATCAATCCAACCCGTCC |
| DabA-C_seq006 | GGAGTCTCGAGTGTTAATCGTC |
| DabA-C_seq007 | CGTTCGCTGGTGCATATCTACT |
| DabA-C_seq008 | GTGAGTCATTCTTAATGCGAATTCGC |
| DabA-C_seq009 | CCATTATTTCCGTAGACACTGCCAC |
| DabA-C_seq010 | CGCTAATACTCCGATGTGCT |
| DabC-D_seq001 | ATTCTCTCTTCCTCGGTGGT |
| DabC-D_seq002 | CGGCACATAAACTAGCTGTTGC |
| DabC-D_seq003 | CGAATCGGACTACGTGTACTAGATAATCTTAC |
| DabC-D_seq004 | CTTAAAGTTCCGACTTCAGCAGTAGGC |
| PtGG_col001 | GCGCAACGTTGTTGCCATTGTCCCAGTCACGACGTTGTAAAA CGAC |
| PtGG_col002 | AATGTTGCAGCACTGACCCTTCACTATAGGGGATATCAGCTG GATGG |
| pBEVY-DabD_001 | AAAAAAAAGTAAGAATTTTTGAAAGCCACCATGAACATTGC AACAGCTAGTTTA |
| pBEVY-DabD_002 | GCCTGCAGGTCGACTCTAGATCACTTATCGTCATCATCCTTG |

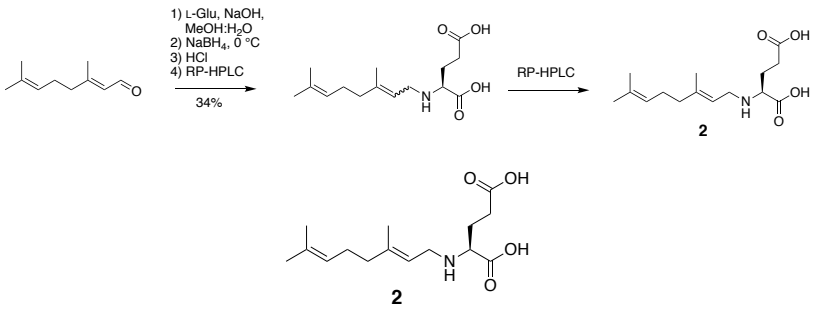
| | |
|---|---|
| pBEVY-PmCPR1_001 | TAACGTCAAGGAGAAAAAACGCCACCATGGTTAACGAAACCACCG |
| pBEVY-PmCPR1_002 | AGTGTCGAATTCGAGCTCGGCTACGACCAAAGCTCTTGAACA |
| pBEVY-DabD_col001 | TTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAAC |
| pBEVY-DabD_col002 | GAGCCAAGGCTGTTTCAATACCGTAAGCATTAATT |
| pBEVY-PmCPR1_col001 | GGGGTAATTAATCAGCGAAGCGATGAT |
| pBEVY-PmCPR1_col002 | AGCGGCTTCTGTAGTTCTCC |
| pBEVY-seq001 | GAAACTTCGAACACTGTCATCTAAAGATGC |
| pBEVY-seq002 | GAAGTGTCAACAACGTATCTACCAACG |
| PmultCPR1_seq001 | GATCAGGAATCACCATCGTTGTCG |
| PmultCPR1_seq002 | TGACGGTGTACTATGCCACC |
| PmultCPR1_seq003 | ACACGGACCCGTTGCTTTCT |
| PmultCPR_seq004 | TTGGTGACGGCTACCGTCA |

## Chemical synthesis

### Synthesis of geranyl pyrophosphate

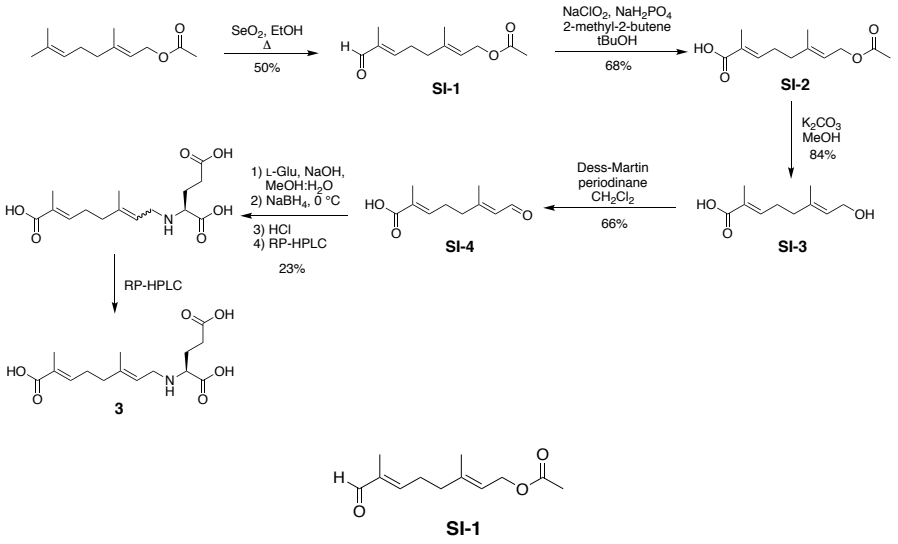
Geranyl pyrophosphate was prepared as a trisammonium salt using established protocols (*58*) and matched literature characterization.

### Synthesis of *N*-geranyl-L-glutamic acid (2)




*N*-geranyl-L-glutamic acid (**2**)
This molecule was synthesized via a modified synthetic procedure (*59*). A solution of citral (0.100 g, 0.66 mmol) in methanol (3 mL) was added to an aqueous solution (3 mL) of L-glutamic acid (0.193 g, 1.31 mmol) and sodium hydroxide (0.079 mg, 1.97 mmol) and stirred for 3 h at room temperature. The reaction mixture was cooled to 0 °C, sodium borohydride (0.032 g, 0.85 mmol) was added at once and stirred for 1 h. 1 N HCl was added until pH 3, and the reaction mixture was concentrated *in vacuo*. Allylic alcohols were removed by extraction with EtOAc (3 x 5 mL), and the aqueous layer was purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 30% B (5 min), 30 – 45% B (10 min), 45 – 95% B (1 min), 95% B (4 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Major regioisomer **2** eluted at 15.2 minutes, whereas minor *cis*-regioisomer *N*-neryl-L-glutamic acid eluted at 14.7 minutes. Fractions were pooled, concentrated *in vacuo* and lyophilized, affording a white solid (0.0664 g, 0.23 mmol, 36%, both regioisomers). Additional separation of regioisomers was performed with analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm). $^{1}$H NMR (600 MHz, $D_2O$): δ 5.21 (t, *J* = 7.8 Hz, 1H), 5.12 (t, *J* = 6.8 Hz, 1H), 3.64 (dd, *J* = 7.8, 5.7 Hz, 2H), 3.57 (dd, *J* = 6.9, 5.6 Hz, 1H), 2.47 – 2.42 (m, 2H), 2.13 – 2.07 (m, 5H), 2.01 (ddt, *J* = 14.4, 7.0, 7.0 Hz, 1H), 1.66 (s, 3H), 1.64 (s, 3H), 1.57 (s, 3H); $^{13}$C NMR (151 MHz, $D_2O$): δ 176.3, 171.8, 146.4, 132.7, 122.5, 111.5, 58.5, 42.5, 37.5, 29.3, 23.9, 23.8, 23.5, 15.7, 14.4; HRMS (ESI) Calculated for $C_{15}H_{24}NO_4$ 282.1711, found 282.1713 $(M-H)^-$
.

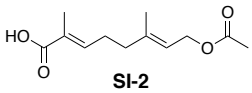
**Synthesis of 7'-carboxy-*N*-geranyl-L-glutamic acid (3)**





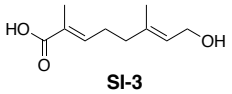


(2*E*,6*E*)-3,7-dimethyl-8-oxoocta-2,6-dien-1-yl acetate (**SI-1**)
This molecule was synthesized following a literature procedure (*60*). To a solution of selenium dioxide (3.88 g, 35.0 mmol) refluxing in 95% ethanol (125 mL) at 89 °C was added a solution of geranyl acetate (5.00 mL, 23.3 mmol) in ethanol (25 mL) over 75 minutes. The reaction mixture was stirred under reflux for 16 h, filtered through a pad of silica, and concentrated *in vacuo*. The crude reaction mixture was purified by silica flash chromatography (4:1 hexanes:EtOAc), isolating aldehyde **SI-1** as a light orange oil (2.48 g, 11.8 mmol, 50%). $R_f$ 0.7 (hexanes/EtOAc 2:1); $^1$H NMR (500 MHz, $CDCl_3$): δ 9.38 (s, 1H), 6.44 (td, $J$ = 7.3, 1.5 Hz, 1H), 5.37 (tq, $J$ = 7.0, 1.4 Hz, 1H), 4.58 (d, $J$ = 7.1 Hz, 2H), 2.49 (dt, $J$ = 7.4, 7.4 Hz, 2H), 2.23 (t, $J$ = 7.6 Hz, 2H), 2.04 (s, 3H), 1.74 (s, 3H), 1.73 (s, 3H); HRMS (ESI) Calculated for $C_{12}H_{18}NaO_3$ 233.1148, found 233.1147 $(M+Na)^+$.




(2*E*,6*E*)-8-acetoxy-2,6-dimethylocta-2,6-dienoic acid (**SI-2**)
This molecule was synthesized following a literature procedure (*61*). To a solution of **SI-1** (0.981 g, 4.67 mmol) and 2-methyl-2-butene (5.49 mL, 90%, 46.7 mmol) in *tert*-butanol (20 mL) at 0°C was added an aqueous (20 mL) solution of sodium dihydrogen phosphate (2.24 g, 18.7 mmol) and sodium chlorite (2.11 g, 80%, 18.7 mmol) over 30 minutes. Following addition, the reaction mixture was stirred at 0 °C for 1 h, then warmed to room temperature and stirred for 18 h. Organic compounds were extracted using EtOAc (3 x 25 mL), pooled organic layers were washed with water (25 mL) and brine (25 mL), dried over $MgSO_4$, filtered and concentrated *in*

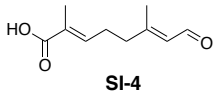
*vacuo*. The crude reaction mixture was purified by silica flash chromatography (4:1 to 2:1 hexanes:EtOAc + 0.1% AcOH), yielding **SI-2** was a clear colorless oil (0.877 g, 3.88 mmol, 68%). $R_f$ 0.24 (hexanes/EtOAc 2:1); $^1$H NMR (500 MHz, $CDCl_3$): δ 6.86 (tq, *J* = 7.3, 1.5 Hz, 1H), 5.37 (tq, *J* = 7.2, 1.4 Hz, 1H), 4.59 (d, *J* = 7.1 Hz, 2H), 2.34 (dt, *J* = 7.4, 7.4 Hz, 2H), 2.18 (t, *J* = 7.6 Hz, 2H), 2.06 (s, 3H), 1.84 (s, 3H), 1.72 (s, 3H); HRMS (ESI) Calculated for $C_{12}H_{18}NaO_4$ 249.1097, found 249.1094 $(M+Na)^+$.




**SI-3**

(2*E*,6*E*)-8-hydroxy-2,6-dimethylocta-2,6-dienoic acid (**SI-3**)

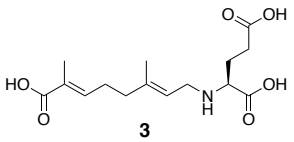
This compound was synthesized following a literature procedure (*62*). To a solution of **SI-2** (0.198 g, 0.88 mmol) in methanol (10 mL) at 0 °C, an aqueous (2 mL) solution of potassium carbonate (0.242 g, 1.75 mmol) was added over 2 minutes. Following addition, the reaction mixture was stirred at 0 °C for 5 minutes, then warmed to room temperature and stirred for 16 h. $CH_2Cl_2$ (10 mL) was added, followed by 1 N HCl until pH < 3. The layers were separated, and organic compounds were extracted with subsequent $CH_2Cl_2$ washes (2 x 10 mL). Pooled organic layers were washed with water (10 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*, using toluene co-evaporations to remove trace water. Compound **SI-3** was obtained as a light yellow oil (0.136 g, 0.74 mmol, 84%) and was used without further purification. $R_f$ 0.45 (EtOAc + 0.1% AcOH); $^1$H NMR (500 MHz, $CDCl_3$): δ 6.87 (tq, *J* = 7.3, 1.6 Hz, 1H), 5.49 – 5.39 (m, 1H), 4.17 (d, *J* = 6.8 Hz, 2H), 2.34 (dt, *J* = 7.5, 7.4 Hz, 2H), 2.16 (t, *J* = 7.6 Hz, 2H), 1.84 (d, *J* = 1.2 Hz, 3H), 1.69 (s, 3H). HRMS (ESI) Calculated for $C_{10}H_{16}NaO_3$ 207.0992, found 207.0991 $(M+Na)^+$.




**SI-4**

(2*E*,6*E*)-2,6-dimethyl-8-oxoocta-2,6-dienoic acid (**SI-4**)

Dess Martin periodinane (0.465 g, 1.10 mmol) was added to a solution of **SI-3** (0.101 g, 0.55 mmol) in dry $CH_2Cl_2$ (10 mL) at once and stirred at room temperature for 1 h. The reaction mixture was quenched with an aqueous 10% sodium thiosulfate solution (10 mL), and washed with water (10 mL). Organic compounds were extracted from pooled aqueous layers with $CH_2Cl_2$ (3 x 20 mL). Pooled organic layers were washed with brine (20 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude reaction mixture was purified by silica flash chromatography (1:1 hexanes:EtOAc + 0.1% AcOH), yielding **SI-4** as a white waxy solid (0.066 g, 0.36 mmol, 66%). $R_f$ 0.62 (EtOAc + 0.1% AcOH); IR ($CH_2Cl_2$ cast) 2934 (br), 2656, 1687, 1648, 1422, 1382, 1284, 1196, 1126 $cm^{-1}$; $^1$H NMR (600 MHz, $CDCl_3$): δ 10.00 (d, *J* = 7.9 Hz, 1H), 6.83 (tq, *J* = 7.1, 1.5 Hz, 1H), 5.90 (dq, *J* = 7.9, 1.3 Hz, 1H), 2.45 – 2.41 (m, 2H), 2.40 – 2.34 (m, 2H), 2.19 (d, *J* = 1.3 Hz, 3H), 1.85 (d, *J* = 1.3 Hz, 3H); $^{13}$C NMR (151 MHz, $CDCl_3$): δ 191.3, 172.8, 162.2, 142.5, 128.4, 127.7, 39.0, 26.5, 17.8, 12.3; HRMS (ESI) Calculated for $C_{10}H_{13}O_3$ 181.0870, found 181.0870 $(M-H)^-$.

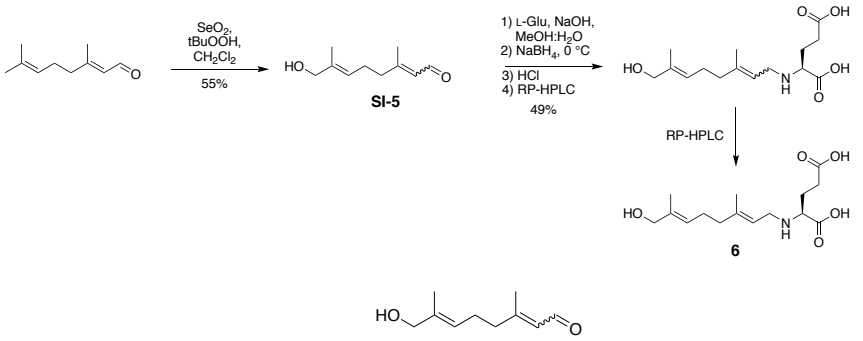



**3**

7'-carboxy-*N*-geranyl-L-glutamic acid (**3**)

A solution of **SI-4** (0.0309 g, 0.17 mmol) in methanol (5 mL) was added to an aqueous solution (5 mL) of L-glutamic acid (0.050 g, 0.34 mmol) and sodium hydroxide (0.021 g, 0.52 mmol) and stirred for 2 h at room temperature. The reaction mixture was cooled to 0 °C, sodium borohydride (0.0083 g, 0.22 mmol) was added at once, and the reaction mixture was stirred for 1 h. 1 N HCl was added until pH 3, and the reaction mixture was concentrated *in vacuo*. Alcohol **SI-3** was recovered by extraction with EtOAc (3 x 5 mL), and the aqueous layer was purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 30% B (15 min), 30 – 95% B (3 min), 95% B (2 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Major regioisomer **3** eluted at 13.7 minutes, whereas minor *cis*-regioisomer 7'-carboxy-*N*-neryl-L-glutamic acid eluted at 14.2 minutes. Fractions were pooled, concentrated *in vacuo* and lyophilized, affording a white solid (0.0123 g, 0.039 mmol, 23%, both regioisomers). Additional separation of regioisomers was performed with analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm). $^{1}$H NMR (600 MHz, $D_2O$): δ 6.71 (t, *J* = 6.8 Hz, 1H), 5.27 (t, *J* = 7.6 Hz, 1H), 3.67 (t, *J* = 6.9 Hz, 2H), 3.57 (dd, *J* = 6.2, 6.2 Hz, 1H), 2.41 (dd, *J* = 7.6, 7.4 Hz, 2H), 2.39 (td, *J* = 7.8, 7.1 Hz, 2H), 2.25 (t, *J* = 7.2 Hz, 2H), 2.10 (ddt, *J* = 14.2, 7.0, 7.0 Hz, 1H), 2.03 (ddt, *J* = 14.5, 7.0, 7.0 Hz, 1H), 1.79 (s, 3H), 1.71 (s, 3H); $^{13}$C NMR (151 MHz, $D_2O$): δ 178.6, 173.3, 173.1, 146.8, 143.1, 128.3, 113.6, 60.1, 43.9, 37.4, 31.3, 26.1, 25.4, 15.5, 11.9; HRMS (ESI) Calculated for $C_{15}H_{22}NO_6$ 312.1453, found 312.1452 $(M-H)^-$.

**Synthesis of 7'-hydroxy-*N*-geranyl-L-glutamic acid (6)**

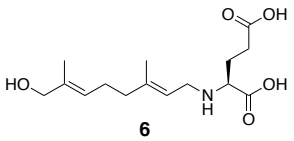





**SI-5**

(6*E*)-8-hydroxy-3,7-dimethylocta-2,6-dienal (**SI-5**)

This compound was synthesized following modification of a literature procedure (*63*). To a suspension of selenium (IV) dioxide (0.089 g, 0.80 mmol) in $CH_2Cl_2$ (5 mL), a solution of citral (0.36 mL, 95%, 2.00 mmol) in $CH_2Cl_2$ (5 mL), and aqueous *t*-butyl hydrogen peroxide (0.51 mL, 70%, 4.00 mmol) were sequentially added and stirred for 18 h at room temperature. The reaction mixture was diluted with $CH_2Cl_2$ (20 mL), filtered to remove particulates, and washed with saturated aqueous $NaHCO_3$ (25 mL) and brine (25 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude reaction mixture was purified by silica flash chromatography (1:1 hexanes:EtOAc), yielding **SI-5** as a light yellow oil (0.184 g, 1.09 mmol, 55%) as an approximately 1:1 mixture of regioisomers. $R_f$ 0.15 (hexanes/EtOAc 2:1); IR ($CHCl_3$ cast) 3408 (br), 2927, 2863, 1667, 1441, 1387, 1192, 1014 $cm^{-1}$; $^1H$ NMR (500 MHz, $CDCl_3$): δ 9.99 (d, $J$ = 8.1 Hz, 1H), 9.89 (d, $J$ = 8.1 Hz, 1H), 5.92 – 5.86 (m, 2H), 5.43 – 5.34 (m, 2H), 4.00 (s, 2H), 3.99 (s, 2H), 2.63 (t, $J$ = 7.5 Hz, 2H), 2.33 – 2.25 (m, 6H), 2.18 (d, $J$ = 1.3 Hz, 3H), 2.00 (d, $J$ = 1.4 Hz, 3H), 1.67 (d, $J$ = 1.4 Hz, 3H), 1.65 (d, $J$ = 1.2 Hz, 3H);$^{13}C$ NMR (125 MHz, $CDCl_3$): δ 191.5, 191.2, 163.9, 163.8, 137.1, 136.3, 128.7, 127.5, 123.7, 123.1, 68.4, 68.3, 40.2, 32.3, 26.6, 25.3, 25.2, 17.7, 13.8; HRMS (ESI) Calculated for $C_{10}H_{16}NaO_2$ 191.1043, found 191.1044 $(M+Na)^+$.



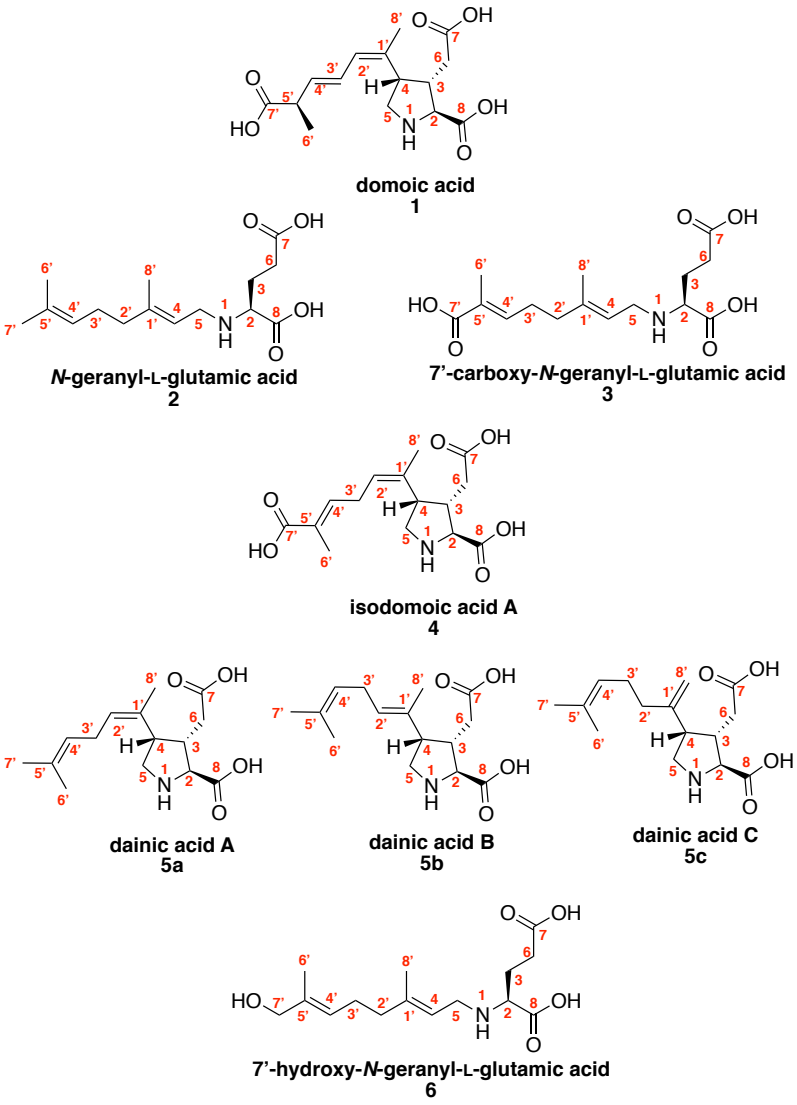
**6**

7'-hydroxy-*N*-geranyl-L-glutamic acid (**6**)

A solution of **SI-5** (0.100 g, 0.59 mmol) in methanol (2.5 mL) was added to an aqueous solution (2.5 mL) of L-glutamic acid (0.175 g, 1.19 mmol) and sodium hydroxide (0.071 mg, 1.78 mmol) and stirred for 3 h at room temperature. The reaction mixture was cooled to 0 °C, sodium borohydride (0.029 g, 0.77 mmol) was added at once and stirred for 1 h. 1 N HCl was added

until pH 3, and the reaction mixture was concentrated *in vacuo*. Allylic alcohols were removed by extraction with EtOAc (3 x 5 mL), and the aqueous layer was purified by preparative RP-HPLC (Phenomenex Luna 5u C18(2), 10.0 x 250 mm) at a flow rate of 10 mL/min using the following method: 10% B (5 min), 10 – 17.5% B (15 min), 17.5 – 95% B (3 min), 95% B (2 min), 95 – 10% B (2 min), 10% B (3 min), where A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile. Major regioisomer **6** eluted at 15.7 minutes, whereas minor *cis*-regioisomer 7'-hydroxy-*N*-neryl-L-glutamic acid eluted at 16.5 minutes. Fractions were pooled, concentrated *in vacuo* and lyophilized, affording a white solid (0.087 g, 0.29 mmol, 49%, both regioisomers). Additional separation of regioisomers was performed with analytical RP-HPLC (Phenomenex Luna 5u C18(2), 4.6 x 150 mm). $^{1}$H NMR (600 MHz, $D_2O$): δ 5.35 (tq, *J* = 5.5, 1.5 Hz, 1H), 5.21 (tq, *J* = 7.8, 1.6 Hz, 1H), 3.91 (s, 2H), 3.65 (dd, *J* = 7.8, 3.4 Hz, 2H), 3.59 (dd, *J* = 7.4, 5.4 Hz, 1H), 2.51 – 2.44 (m, 2H), 2.16 (t, *J* = 6.9 Hz, 2H), 2.14 – 2.07 (m, 3H), 2.02 (ddt, *J* = 14.0, 7.0, 7.0 Hz, 1H), 1.66 (d, *J* = 1.3 Hz, 3H), 1.59 (d, *J* = 1.3 Hz, 3H); $^{13}$C NMR (151 MHz, $D_2O$): δ 175.3, 171.2, 145.8, 133.2, 124.9, 111.6, 66.2, 58.3, 42.5, 37.1, 28.7, 23.5, 23.5, 14.2, 11.5; HRMS (ESI) Calculated for $C_{15}H_{24}NO_5$ 298.1660, found 298.1661 $(M-H)^-$.

**NMR and Compound Characterization**




Numbering scheme for synthetic and enzyme-isolated compounds **2-6** based on unified numbering scheme for domoic acid (**1**). (*64*)

**$^1$H NMR table for compounds 2 – 6**

| $^1$H | 2 (DabA) | 2 (synth) | 3 (synth) | 4 (DabC) | 5a (DabC) | 5b (DabC) | 5c (DabC) | 6 (synth) |
|---|---|---|---|---|---|---|---|---|
| **2** | 3.57, dd (6.6, 5.7) | 3.57, dd (6.9, 5.6) | 3.57, dd (6.2, 6.2) | 3.97, d (7.3) | 3.93, d (6.5) | 4.00, d (3.2) | 4.06, d (2.4) | 3.59, dd (7.4, 5.4) |
| **3** | 2.08, m 2.02, ddt (14.5, 6.8, 6.8) | 2.08, m 2.01, ddt (14.4, 7.0, 7.0) | 2.10, ddt (14.2, 7.0, 7.0) 2.03, ddt (14.5, 7.0, 7.0) | 3.01, ddd (12.7, 10.1, 5.4) | 2.95, ddd (9.0, 6.7, 6.5) | 3.01, ddd (8.8, 7.3, 2.5) | 3.03, m | 2.09, m 2.02, ddt (14.0, 7.0, 7.0) |
| **4** | 5.21, t (7.3) | 5.21, t (7.8) | 5.27, t (7.6) | 3.63, ddd (7.8, 7.6, 7.6) | 3.55, dd (7.8, 7.8) | 2.94, dd (7.9, 7.6) | 2.97, m | 5.21, tq (7.8, 1.6) |
| **5a** | 3.64, t (6.9) | 3.64, dd (7.8, 5.7) | 3.67, t (6.9) | 3.69, dd (11.7, 8.1) | 3.63, dd (11.8, 8.1) | 3.55, dd (11.9, 7.5) | 3.58, dd (11.9, 7.6) | 3.65, dd (7.8, 3.4) |
| **5b** | 3.64, t (6.9) | 3.64, dd (7.8, 5.7) | 3.67, t (6.9) | 3.46, dd (11.7, 7.7) | 3.42, dd (11.8, 8.1) | 3.40, dd (11.4, 11.4) | 3.42, dd (11.8, 11.8) | 3.65, dd (7.8, 3.4) |
| **6a** | 2.44, m | 2.45, m | 2.41, dd (7.6, 7.4) | 2.67, dd (16.5, 6.5) | 2.40, dd (15.6, 6.7) | 2.20, dd (15.4, 6.5) | 2.25, dd (15.5, 6.0) | 2.47, m |
| **6b** | 2.44, m | 2.45, m | 2.41, dd (7.6, 7.4) | 2.50, dd (16.5, 8.6) | 2.30, dd (15.6, 8.5) | 2.09, m | 2.07, m | 2.47, m |
| **2'a** | 2.10, m | 2.09, m | 2.25, t (7.2) | 5.54, t (7.5) | 5.43, t (7.2) | 5.14, m | 2.14, m | 2.12, m |
| **2'b** | 2.10, m | 2.09, m | 2.25, t (7.2) | - | - | - | 2.12, m | 2.12, m |
| **3'a** | 2.11, m | 2.10, m | 2.39, td (7.8, 7.1) | 2.96, ddd (15.6, 7.7, 7.2) | 2.76, ddd (16.2, 8.0, 7.2) | 2.76, m | 2.16, m | 2.16, t (6.9) |
| **3'b** | 2.11, m | 2.10, m | 2.39, td (7.8, 7.1) | 2.87, ddd (15.8, 7.2, 7.0) | 2.60, m | 2.76, m | 2.16, m | 2.16, t (6.9) |
| **4'** | 5.12, t (6.7) | 5.12, t (6.8) | 6.71, t (6.8) | 6.67, t (7.2) | 5.15, m | 5.19, t (7.2) | 5.19, t (8.6) | 5.35, tq (5.5, 1.5) |
| **6'** | 1.57, s | 1.57, s | 1.79, s | 1.83, s | 1.62, s | 1.64, s | 1.61, s | 1.59, d (1.3) |
| **7'** | 1.64, s | 1.64, s | - | - | 1.68, s | 1.69, s | 1.68, s | 3.91, s |
| **8'** | 1.66, s | 1.66, s | 1.71, s | 1.76, s | 1.69, s | 1.68, s | 5.04, s 4.77, s | 1.66, d (1.3) |

all spectra were collected in $D_2O$ and referenced to sodium formate (δ 8.44).

**$^{13}$C NMR table for compounds 2 – 6**

| $^{13}$C | 2 (DabA) | 2 (synth) | 3 (synth) | 4 (DabC) | 5a (DabC) | 5b (DabC) | 5c (DabC) | 6 (synth) |
|---|---|---|---|---|---|---|---|---|
| **2** | 58.5 | 58.5 | 60.1 | 65.1 | 65.1 | 65.7 | 65.7 | 58.3 |
| **3** | 23.9 | 23.8 | 25.4 | 42.4 | 43.7 | 42.2 | 41.5 | 23.5 |
| **4** | 111.5 | 111.5 | 113.6 | 40.2 | 40.3 | 47.0 | 43.9 | 111.6 |
| **5** | 42.5 | 42.5 | 43.9 | 46.8 | 46.8 | 46.0 | 45.9 | 42.5 |
| **6** | 29.7 | 29.3 | 31.3 | 34.1 | 36.2 | 36.1 | 35.9 | 28.7 |
| **7** | 176.6 | 176.3 | 178.6 | 176.5 | 179.4 | 179.8 | 179.8 | 175.3 |
| **8** | 171.9 | 171.8 | 173.3 | 172.9 | 173.5 | 173.4 | 173.7 | 171.2 |
| **1'** | 146.4 | 146.4 | 146.8 | 131.0 | 129.4 | 129.8 | 144.1 | 145.8 |
| **2'** | 37.6 | 37.5 | 37.4 | 128.1 | 130.2 | 126.1 | 35.4 | 37.1 |
| **3'** | 24.0 | 23.9 | 26.1 | 27.1 | 26.3 | 26.4 | 25.4 | 23.5 |
| **4'** | 122.4 | 122.5 | 143.1 | 141.0 | 122.2 | 122.3 | 123.4 | 124.9 |
| **5'** | 132.5 | 132.7 | 128.3 | 128.0 | 134.0 | 133.9 | 134.0 | 133.2 |
| **6'** | 15.7 | 15.7 | 11.9 | 11.8 | 16.9 | 16.9 | 16.6 | 11.5 |
| **7'** | 23.5 | 23.5 | 173.1 | 172.9 | 24.7 | 24.7 | 24.5 | 66.2 |
| **8'** | 14.3 | 14.4 | 15.5 | 21.3 | 21.4 | 15.9 | 111.6 | 14.2 |

All spectra were collected in $D_2O$ and referenced to sodium formate (δ 171.7). Carbon signals were derived from HSQC and HMBC correlations.

**HMBC correlations table for Dab enzyme-isolated compounds 2, 4, 5a-c**

| $^{13}$C | 2 (DabA) | 4 (DabC) | 5a (DabC) | 5b (DabC) | 5c (DabC) |
|---|---|---|---|---|---|
| **2** | 3, 5, 6 | 6b | 3, 4, 5a, 5b, 6a, 6b | 5a, 6a, 6b | 6a, 6b |
| **3** | 2, 6 | 2, 4, 6b | 2, 4, 5a, 5b, 6a, 6b, 8' | 4, 5a, 6a, 6b | 5a, 6a, 6b |
| **4** | 5, 2', 8' | 5a, 6a, 6b, 2', 8' | 2, 3, 5a, 5b, 6a, 6b, 2', 8' | 2, 5b, 6a, 6b, 2', 8' | 5b, 8'a, 8'b |
| **5** | 2, 4 | 4 | 2, 3, 4 | 4 | - |
| **6** | 2, 3 | 2, 4 | 2, 3, 4 | 2, 4 | - |
| **7** | 3, 6 | 6a, 6b | 3, 6a, 6b | 6a, 6b | 6a, 6b |
| **8** | 2, 3 | 2 | 2, 3 | 2 | 2 |
| **1'** | 5, 2', 3', 8' | 4, 5a, 5b, 2', 8' | 3, 5a, 5b | 4, 5b, 3', 8' | 4, 5b, 2'b |
| **2'** | 4, 3', 8' | 4, 3'a, 3'b, 8' | 4, 3'a, 3'b, 8' | 4, 3', 8' | 8'a, 8'b |
| **3'** | 2', 4' | 2', 4' | 2', 4' | 4' | - |
| **4'** | 2', 3', 6', 7' | 3'a, 3'b, 6' | 2', 3'a, 3'b, 6', 7' | 3', 6', 7' | 6', 7' |
| **5'** | 3', 6', 7' | 6' | 3'a, 3'b, 6', 7' | 6', 7' | 6', 7' |
| **6'** | 7' | 4' | 4', 7' | 4', 7' | 7' |
| **7'** | 4', 6' | 4', 6' | 4', 6' | 4', 6' | 6' |
| **8'** | 4, 2' | 4, 2' | 4, 2' | 4, 2' | 4, 2'b |

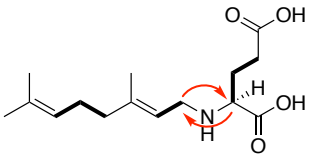
**HMBC correlations table for Dab synthetic compounds 2, 3, 6**

| $^{13}C$ | **2 (synthetic)** | **3 (synthetic)** | **6 (synthetic)** |
|---|---|---|---|
| **2** | 3, 5, 6 | 3, 5, 6 | 3, 5, 6 |
| **3** | 2, 6 | 2, 6 | 2, 6 |
| **4** | 5, 2', 8' | 5, 2', 8' | 5, 2', 8' |
| **5** | 2, 4 | 2 | 2 |
| **6** | 2, 3 | 2, 3 | 2, 3 |
| **7** | 3, 6 | 3, 6 | 3, 6 |
| **8** | 2, 3 | 2, 3 | 2, 3 |
| **1'** | 5, 2', 3', 8' | 5, 2', 3', 8' | 5, 2', 3', 8' |
| **2'** | 4, 3', 8' | 4, 3', 8' | 4, 3', 8' |
| **3'** | 2', 4' | 2' | 2' |
| **4'** | 2', 3', 6', 7' | 2', 3', 6' | 2', 3', 6', 7' |
| **5'** | 3', 6', 7' | 3', 6' | 6', 7' |
| **6'** | 4', 7' | 4' | 4', 7' |
| **7'** | 4', 6' | 4', 6' | 4', 6' |
| **8'** | 4, 2' | 4, 2' | 4, 2' |

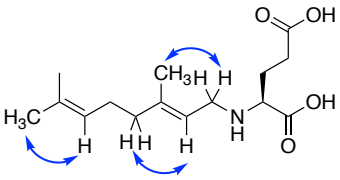
NMR correlations for compounds **2 – 6**

COSY correlations
key HMBC correlations
key NOE correlations

*N*-geranyl-L-glutamic acid – DabA product (**2**):

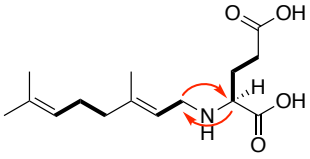


***N*-geranyl-L-glutamic acid**
**2**
(enzymatic)

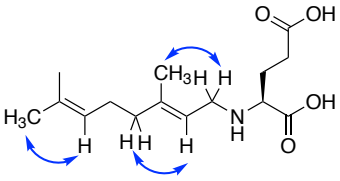


***N*-geranyl-L-glutamic acid**
**2**
(enzymatic)

*N*-geranyl-L-glutamic acid – synthetic (**2**):

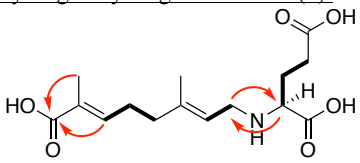


***N*-geranyl-L-glutamic acid**
**2**
(synthetic)

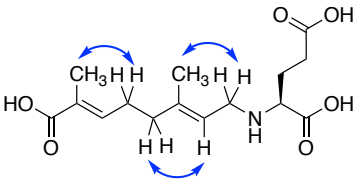


***N*-geranyl-L-glutamic acid**
**2**
(synthetic)

7'-carboxy-*N*-geranyl-L-glutamic acid (**3**):

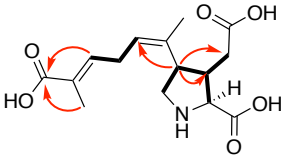


**7'-carboxy-*N*-geranyl-L-glutamic acid**
**3**

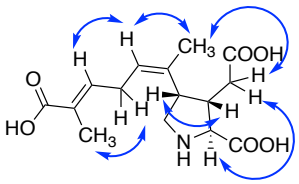


**7'-carboxy-*N*-geranyl-L-glutamic acid**
**3**

isodomoic acid A (**4**):

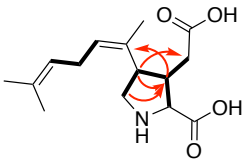


**isodomoic acid A**
**4**

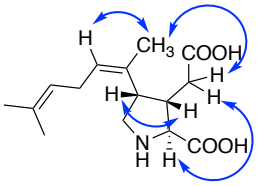


**isodomoic acid A**
**4**

dainic acid A (**5a**):

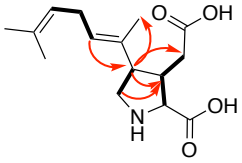


**dainic acid A**
**5a**

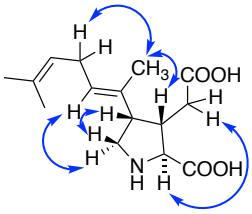


**dainic acid A**
**5a**

dainic acid B (**5b**):

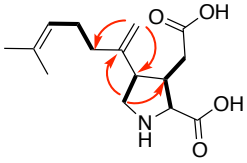


**dainic acid B**
**5b**

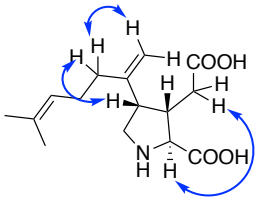


**dainic acid B**
**5b**

dainic acid C (**5c**):

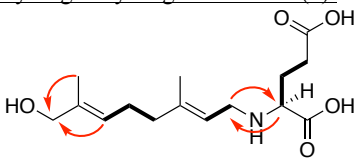


**dainic acid C**
**5c**

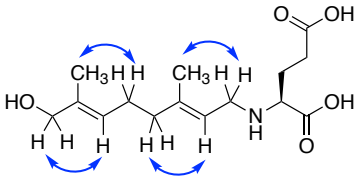


**dainic acid C**
**5c**

7'-hydroxy-*N*-geranyl-L-glutamic acid (**6**):

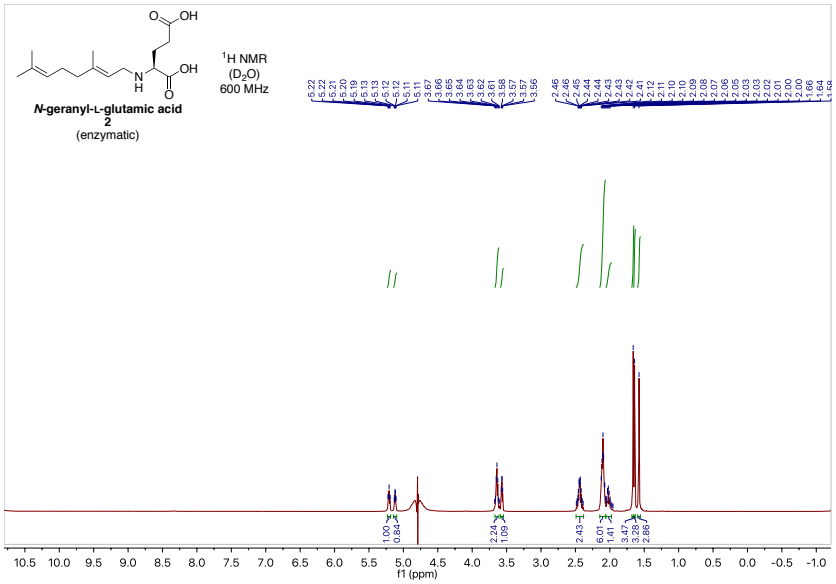


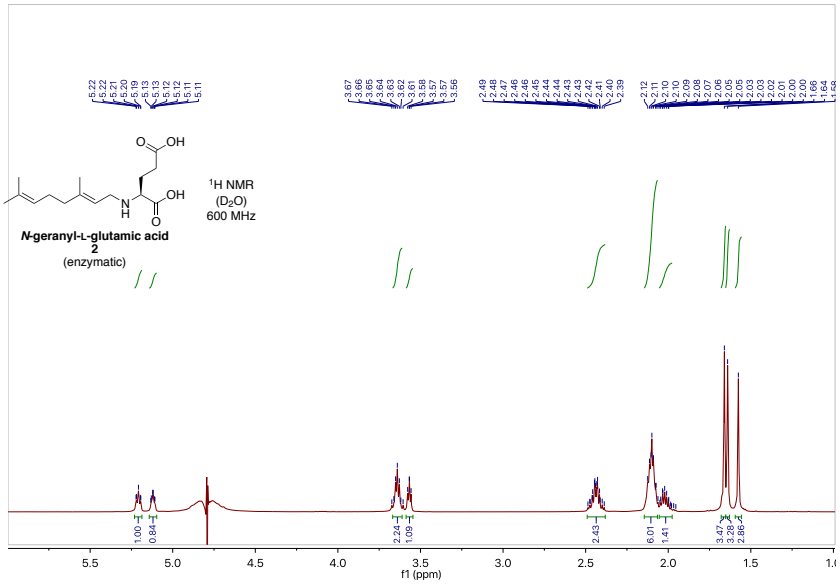
**7'-hydroxy-*N*-geranyl-L-glutamic acid**
**6**

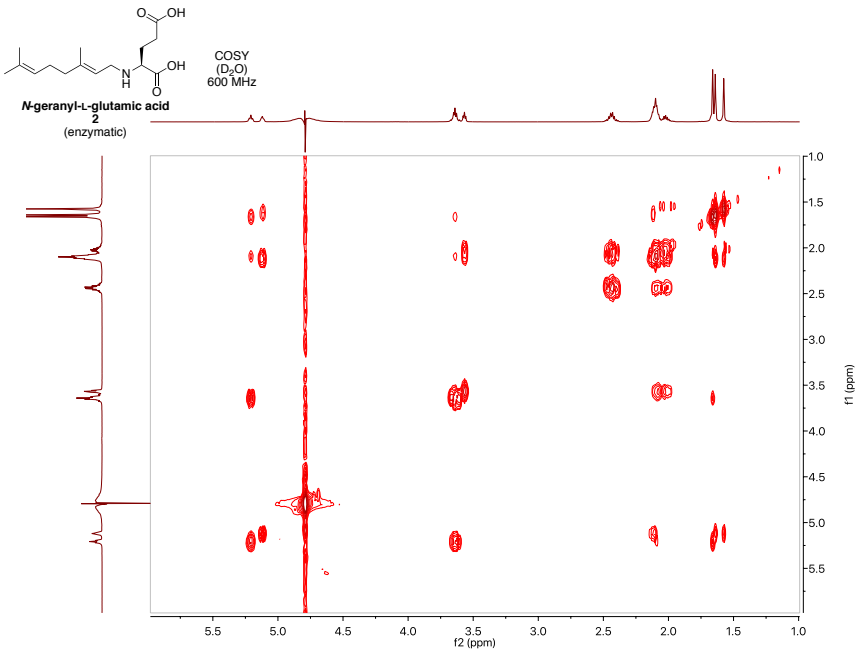


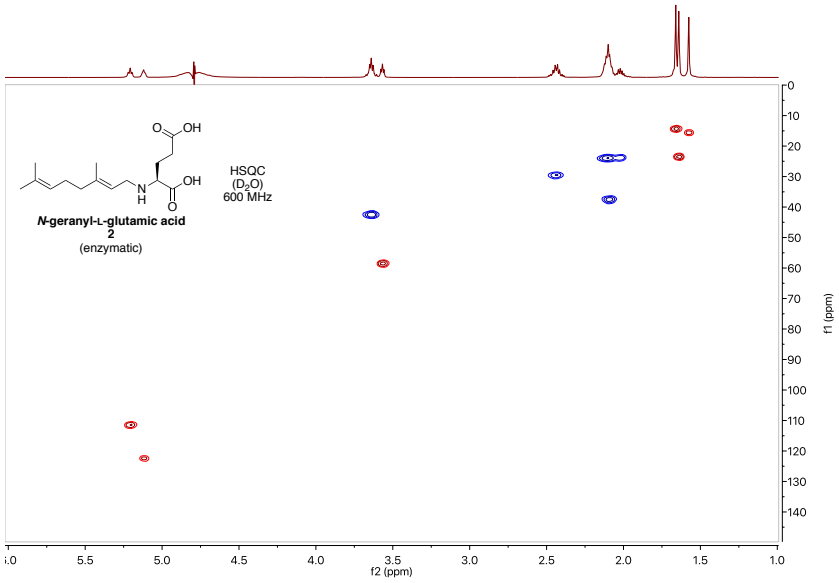
**7'-hydroxy-*N*-geranyl-L-glutamic acid**
**6**

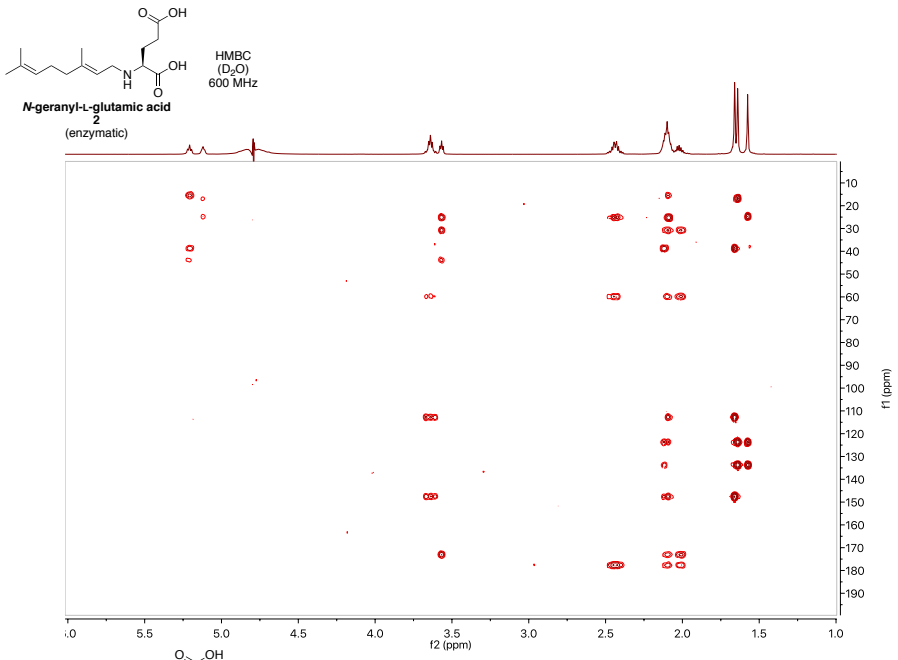
*N*-geranyl-L-glutamic acid – DabA product (**2**):

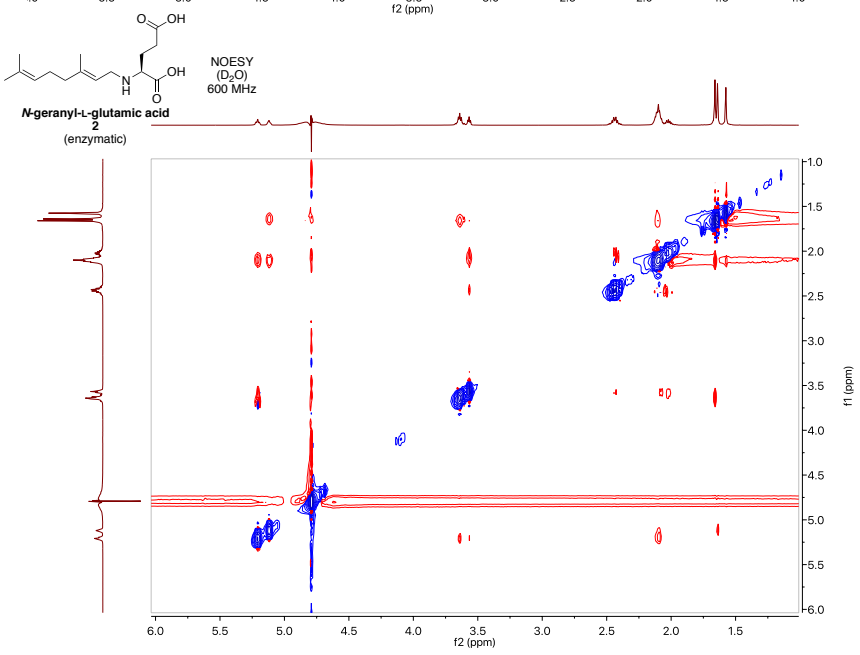


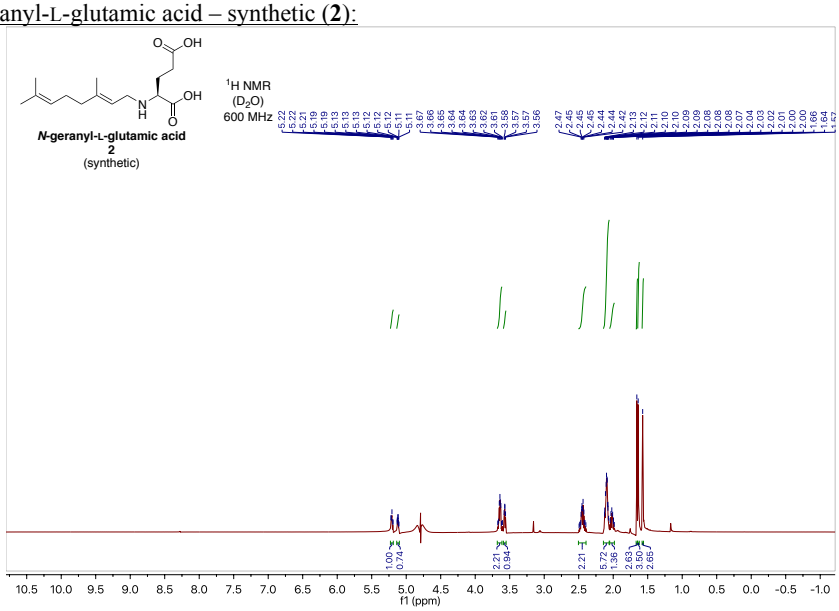

N-geranyl-L-glutamic acid
2
(enzymatic)
COSY
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

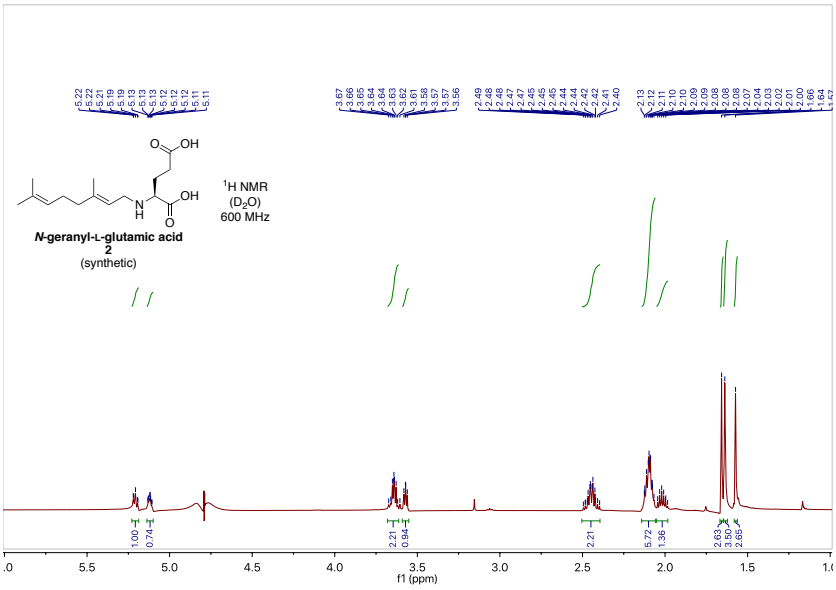

N-geranyl-L-glutamic acid
2
(enzymatic)
HSQC
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

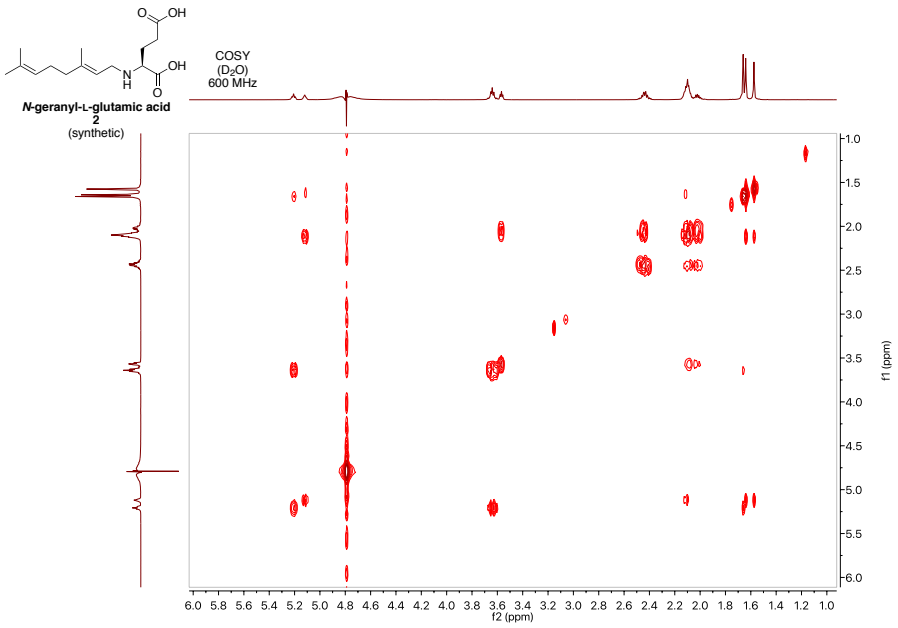

HMBC
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(enzymatic)
f2 (ppm)
f1 (ppm)

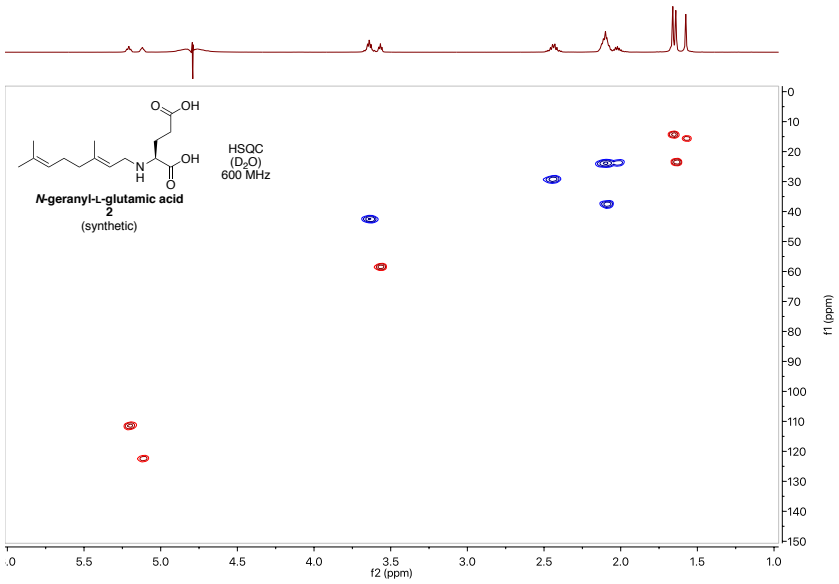

NOESY
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(enzymatic)
f2 (ppm)
f1 (ppm)

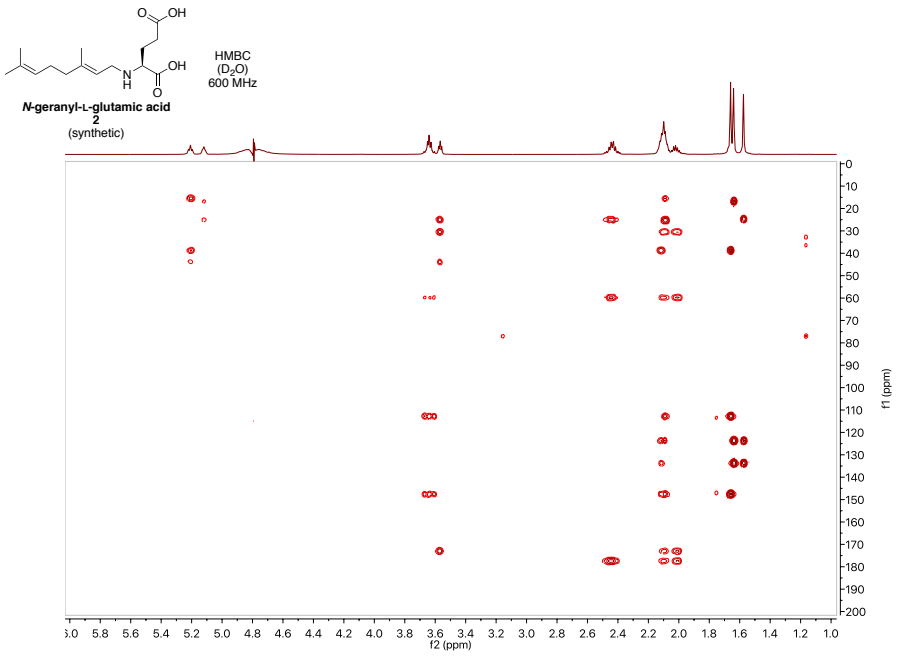
*N*-geranyl-L-glutamic acid – synthetic (**2**):

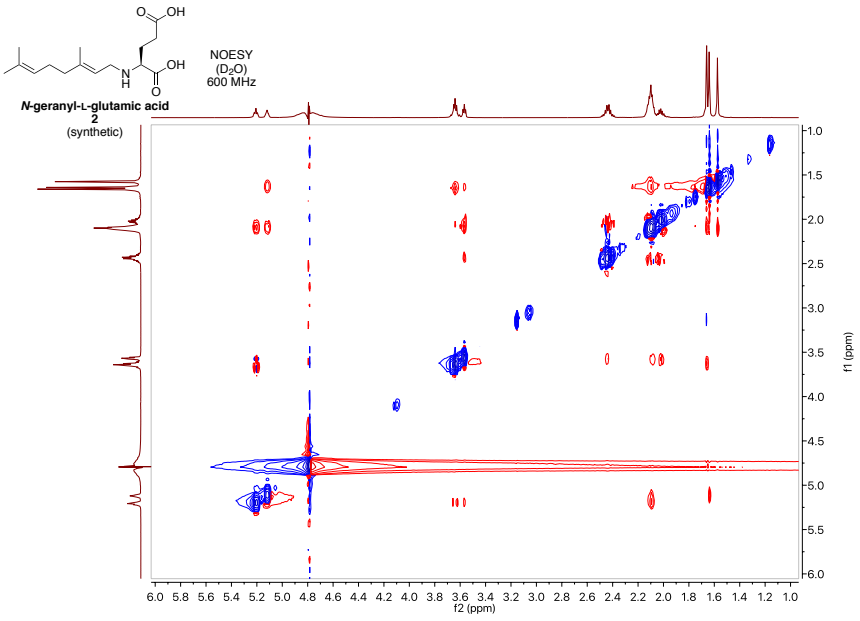


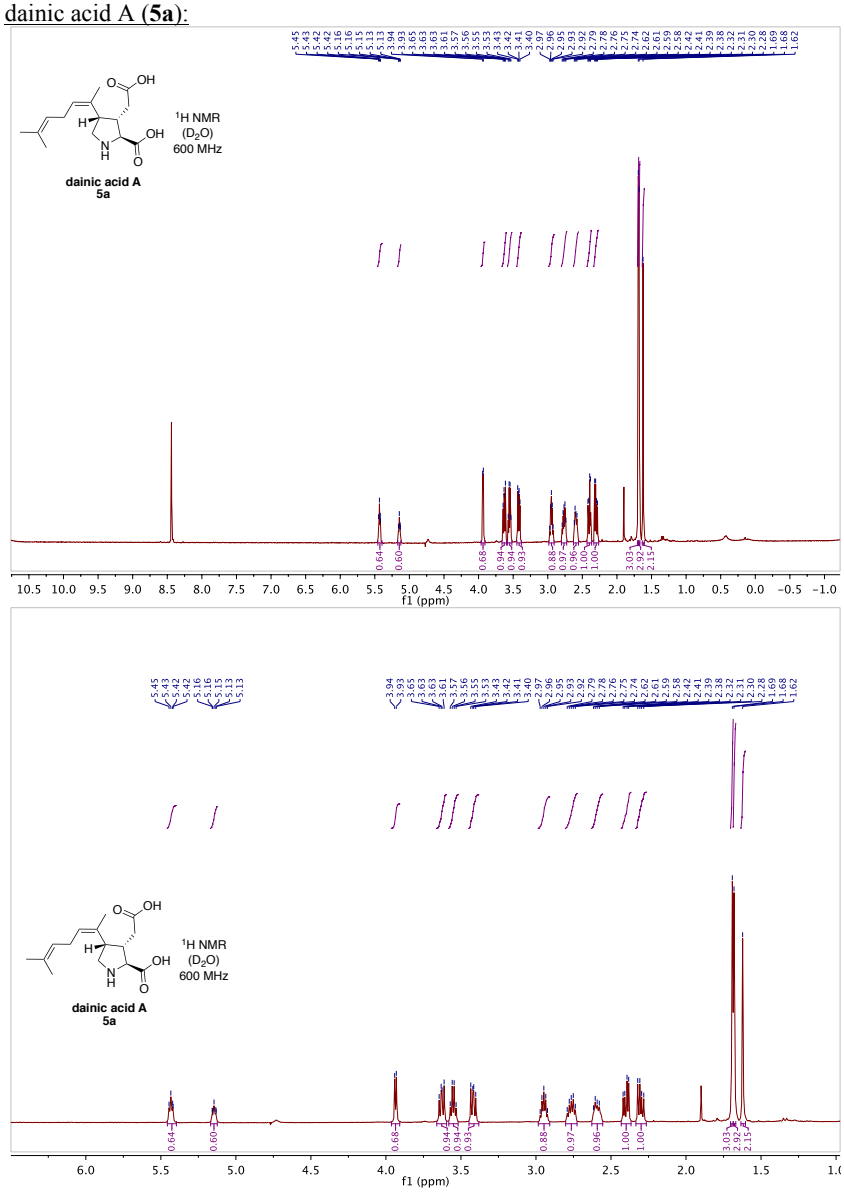

N-geranyl-L-glutamic acid
2
(synthetic)
COSY
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

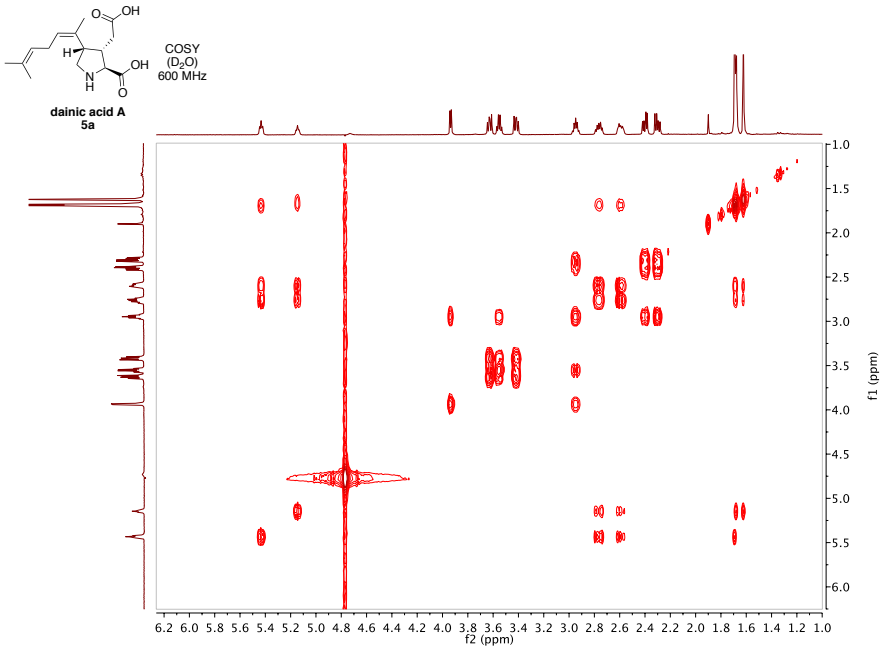

N-geranyl-L-glutamic acid
2
(synthetic)
HSQC
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

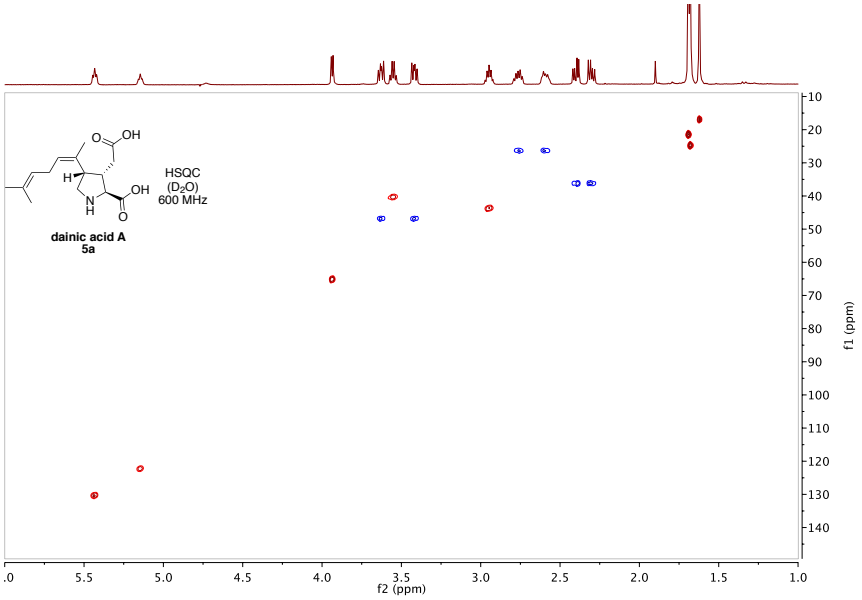

HMBC
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(synthetic)
f2 (ppm)
f1 (ppm)

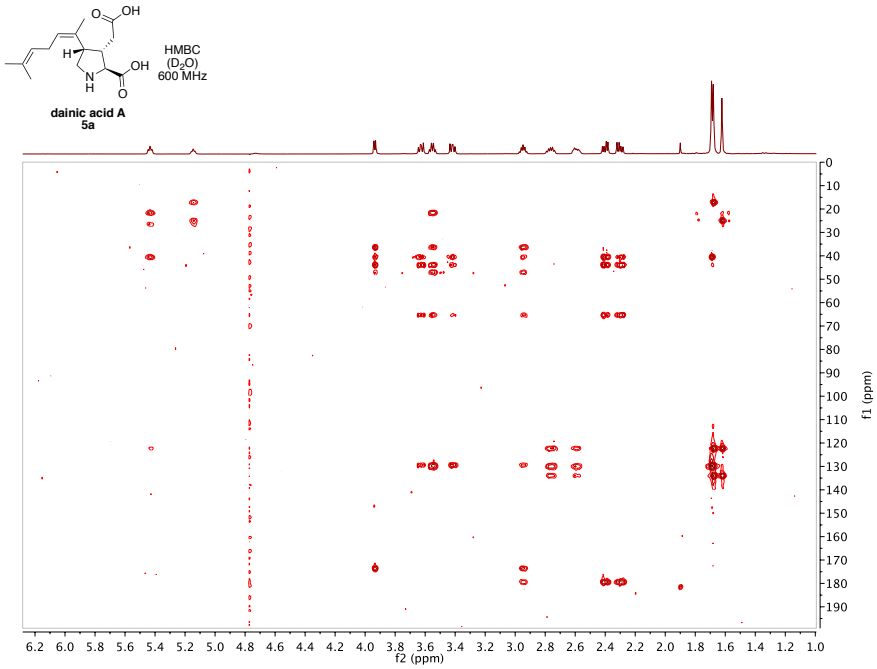

NOESY
(D2O)
600 MHz
N-geranyl-L-glutamic acid
2
(synthetic)
f2 (ppm)
f1 (ppm)

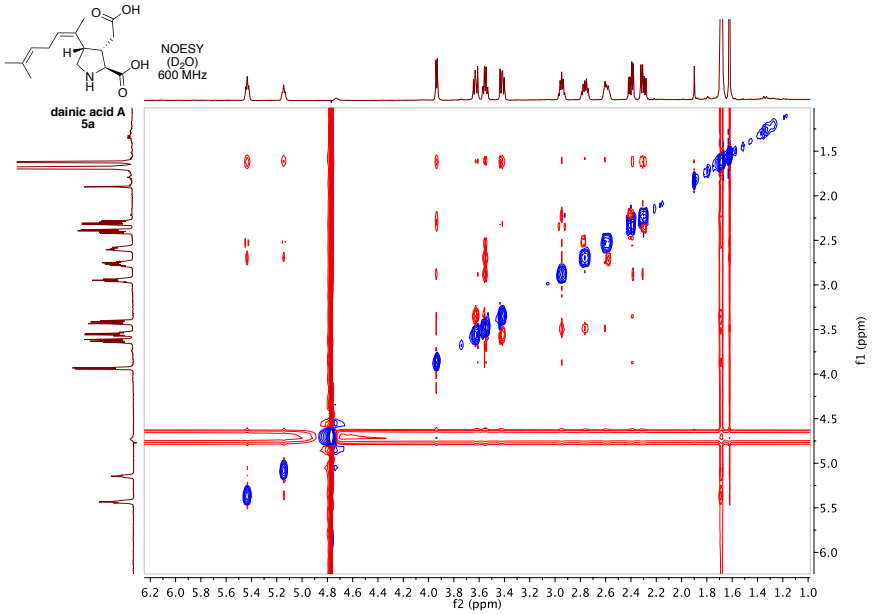
dainic acid A (**5a**):



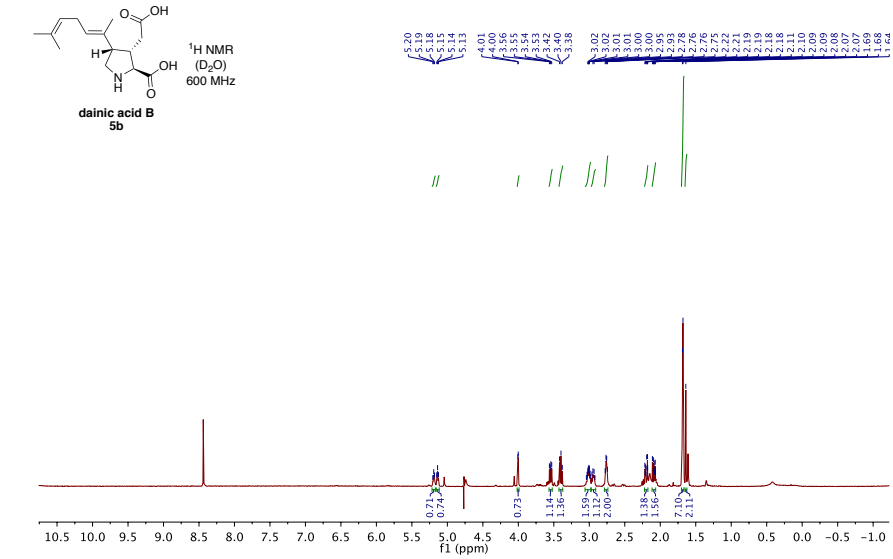

dainic acid A
5a
COSY
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

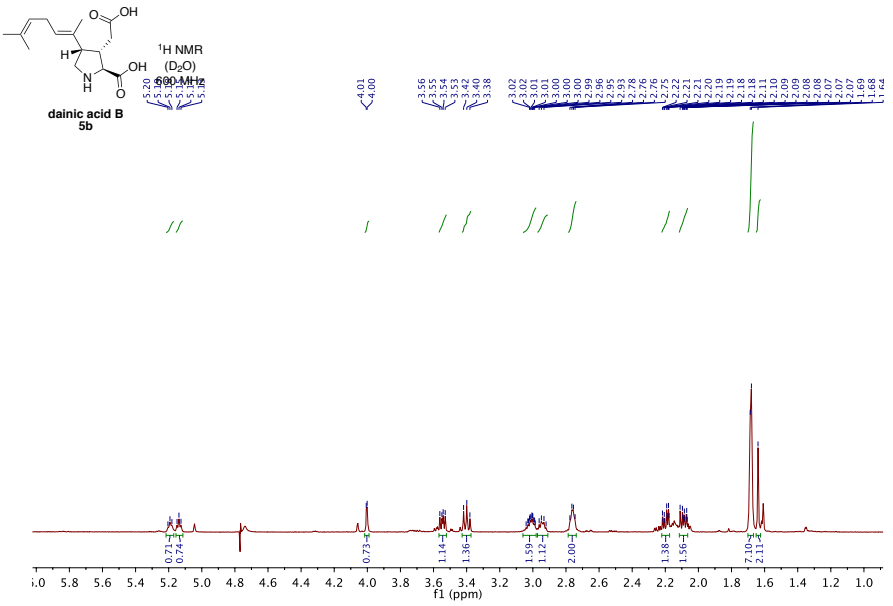

dainic acid A
5a
HSQC
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)

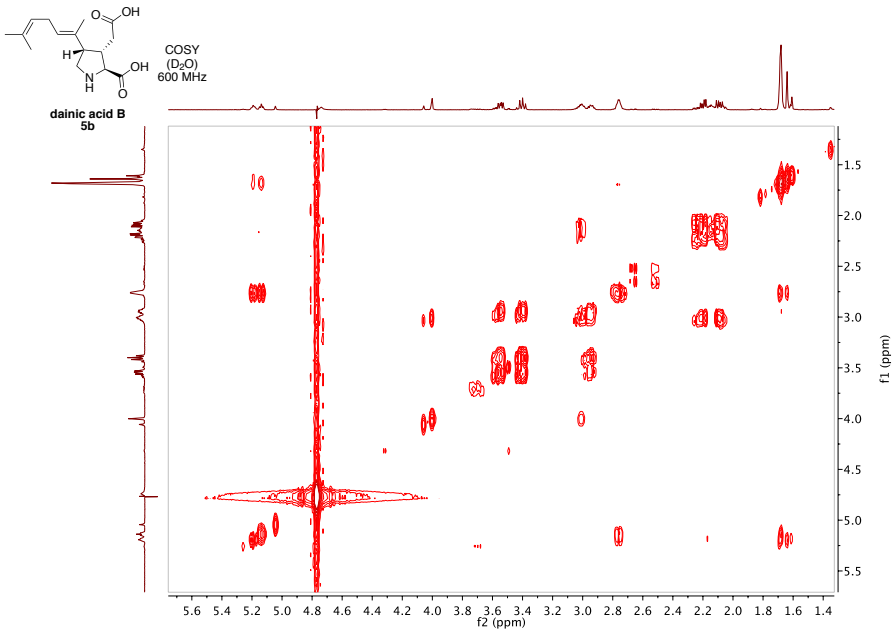

HMBC
(D2O)
600 MHz
dainic acid A
5a
f2 (ppm)
f1 (ppm)

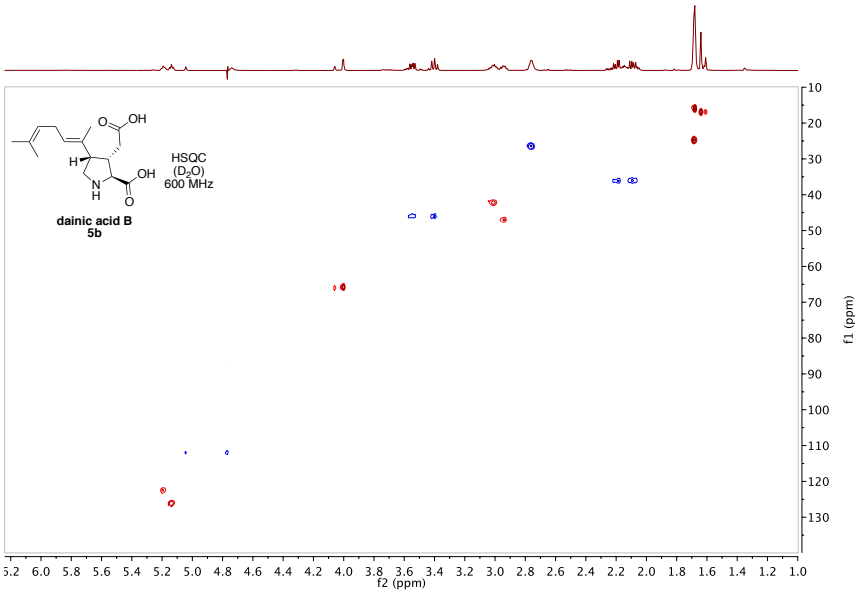

NOESY
(D2O)
600 MHz
dainic acid A
5a
f2 (ppm)
f1 (ppm)

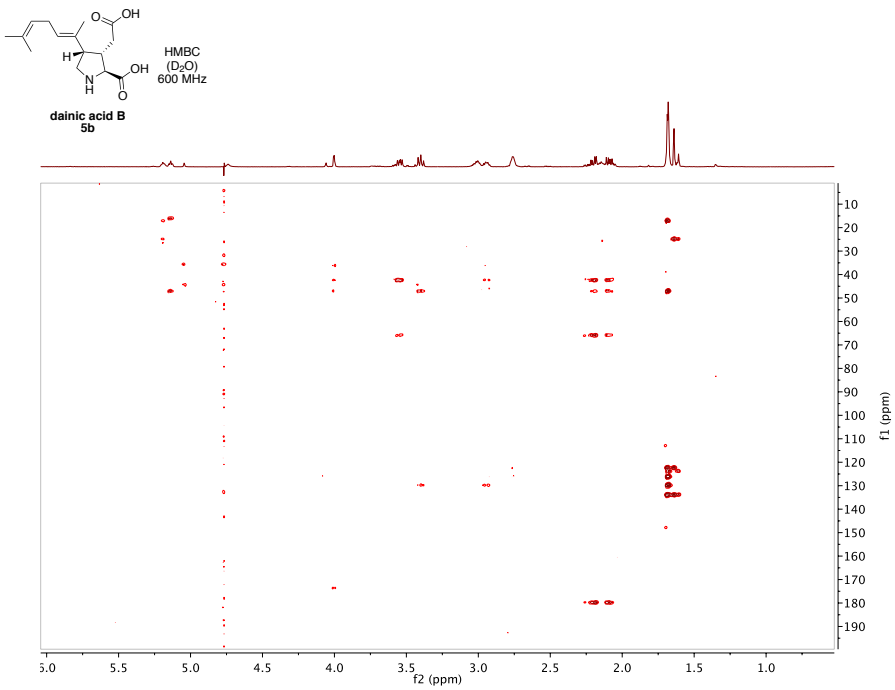
dainic acid B (**5b**):

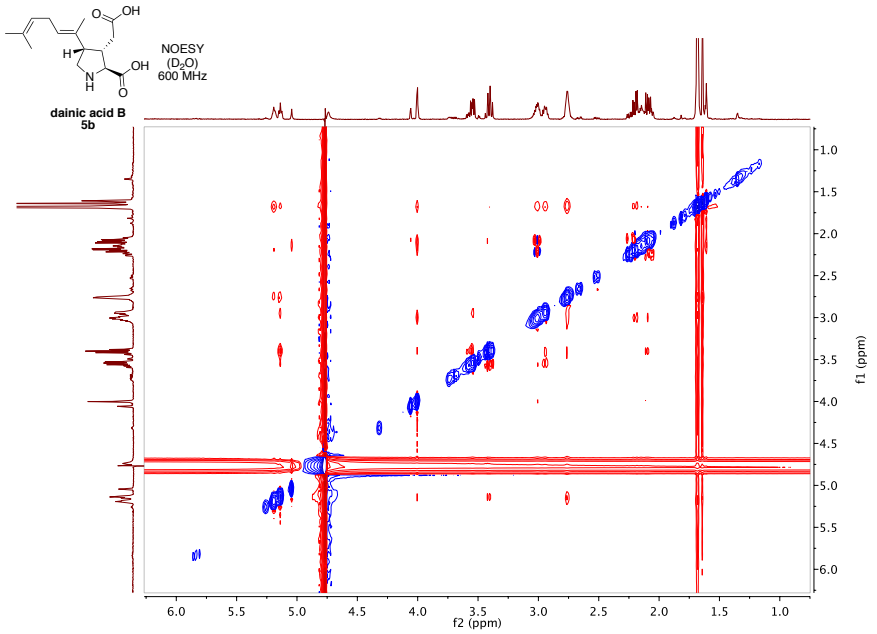


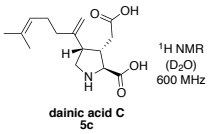

dainic acid B
5b
COSY
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)


dainic acid B
5b
HSQC
(D2O)
600 MHz
f2 (ppm)
f1 (ppm)


HMBC
(D2O)
600 MHz
dainic acid B
5b
f2 (ppm)
f1 (ppm)

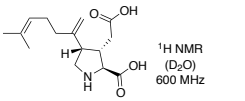

NOESY
(D2O)
600 MHz
dainic acid B
5b
f2 (ppm)
f1 (ppm)

dainic acid C (**5c**):

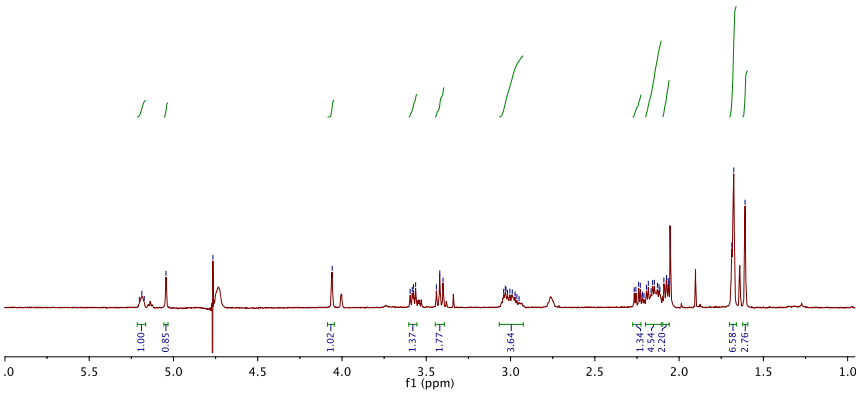


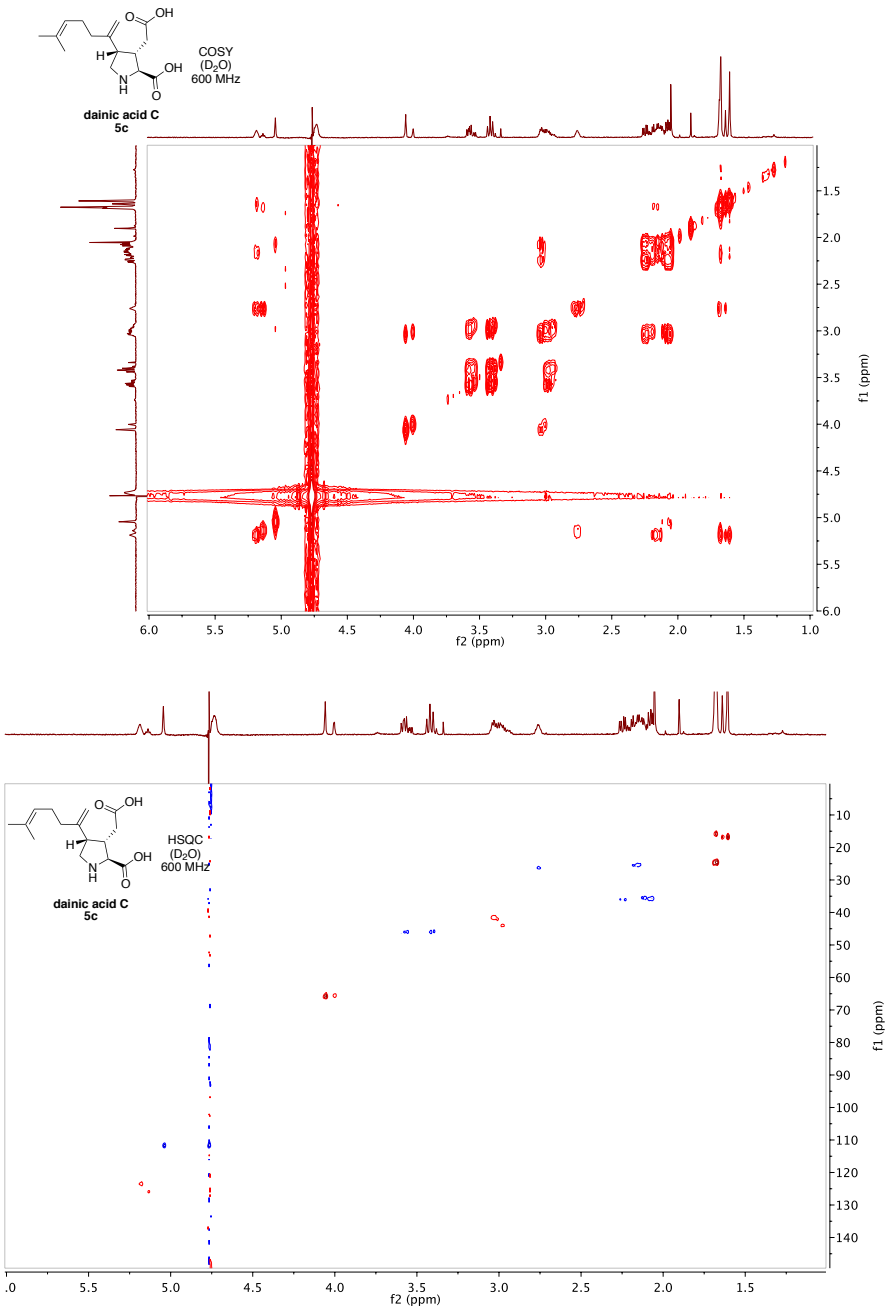

$^1$H NMR
($D_2O$)
600 MHz

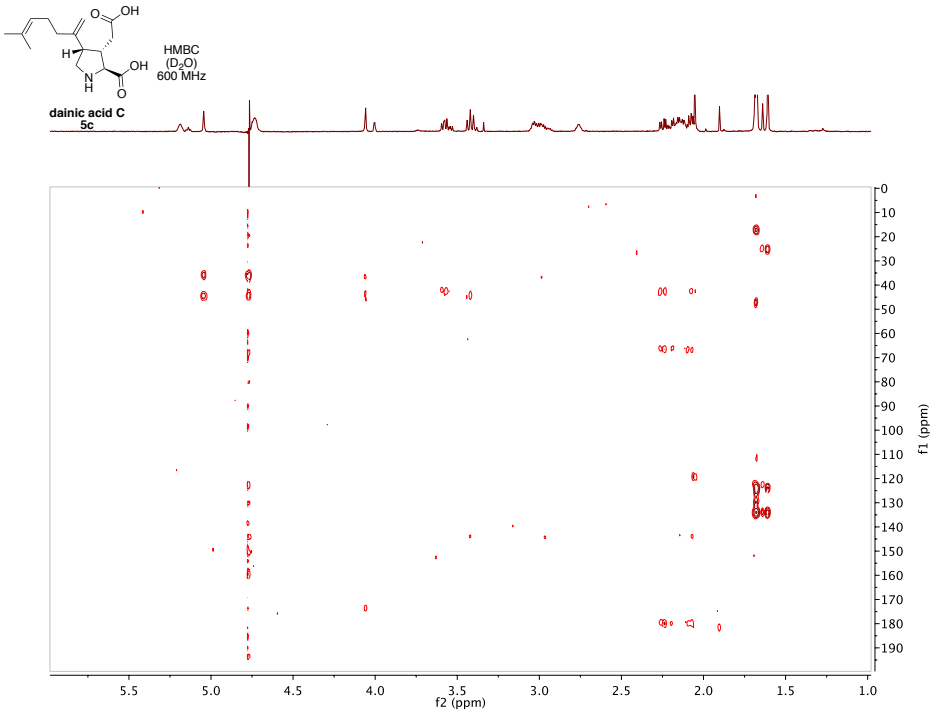


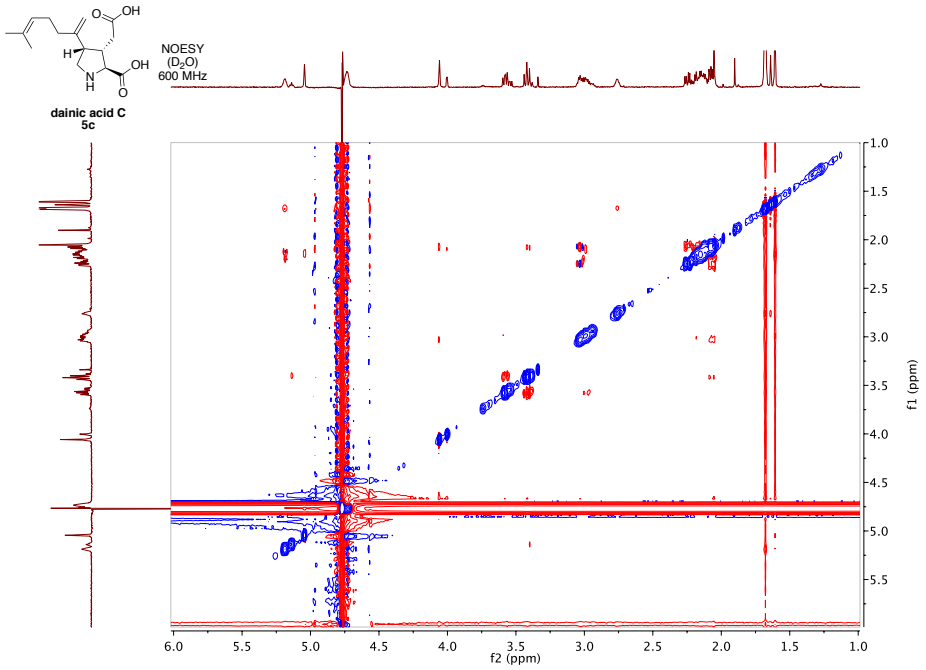




$^1$H NMR
($D_2O$)
600 MHz

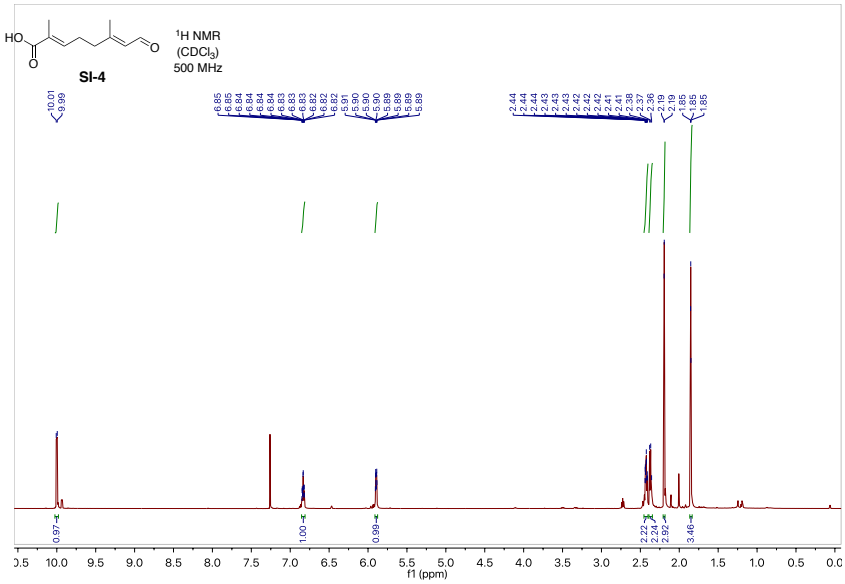


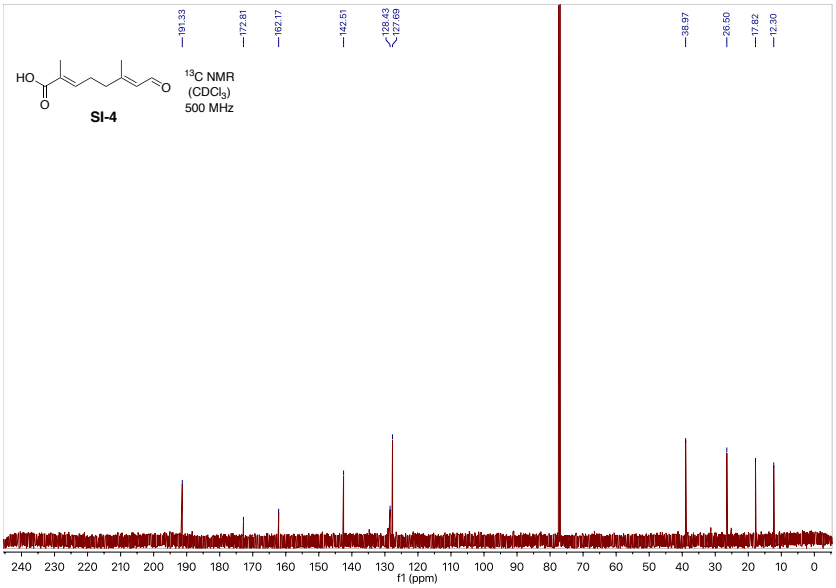

COSY
(D2O)
600 MHz
dainic acid C
5c
f2 (ppm)
f1 (ppm)
HSQC
(D2O)
600 MHz
dainic acid C
5c
f2 (ppm)
f1 (ppm)


HMBC
(D2O)
600 MHz
dainic acid C
5c
f2 (ppm)
f1 (ppm)

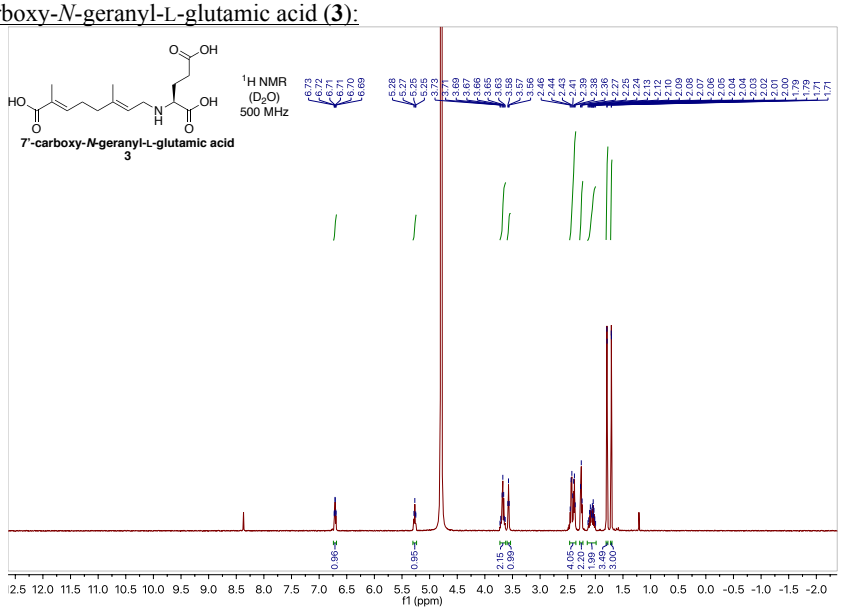

NOESY
(D2O)
600 MHz
dainic acid C
5c
f2 (ppm)
f1 (ppm)

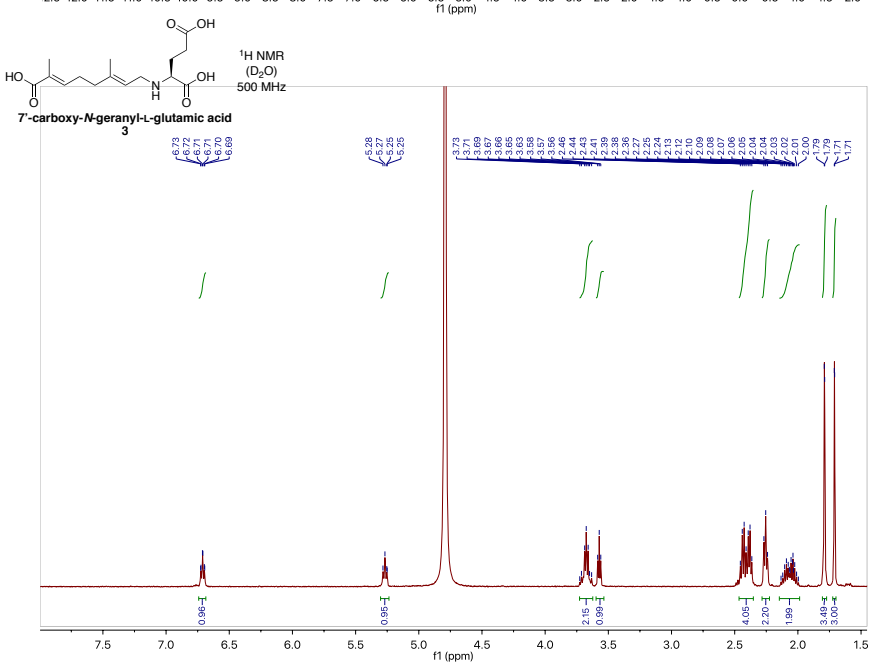
(2*E*,6*E*)-2,6-dimethyl-8-oxoocta-2,6-dienoic acid (**SI-4**):

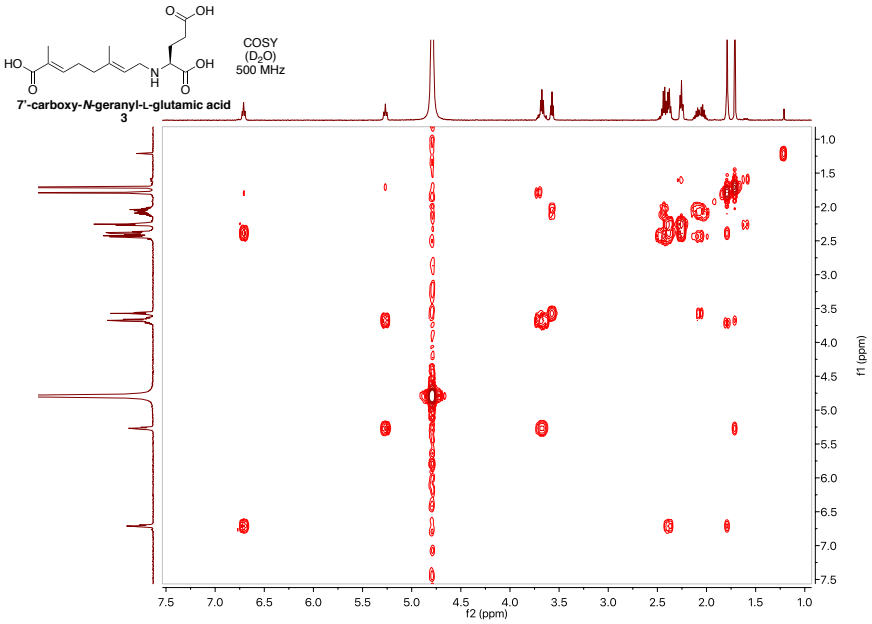


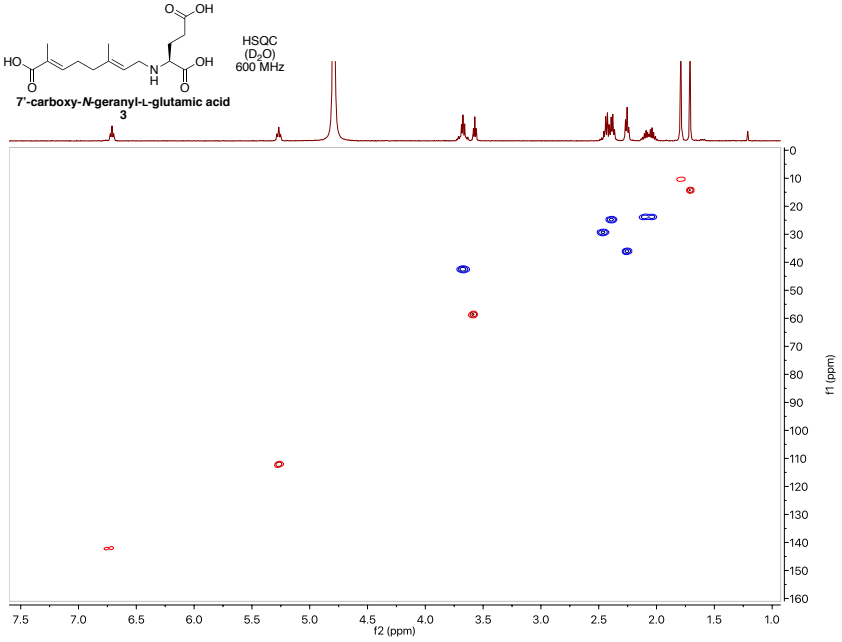



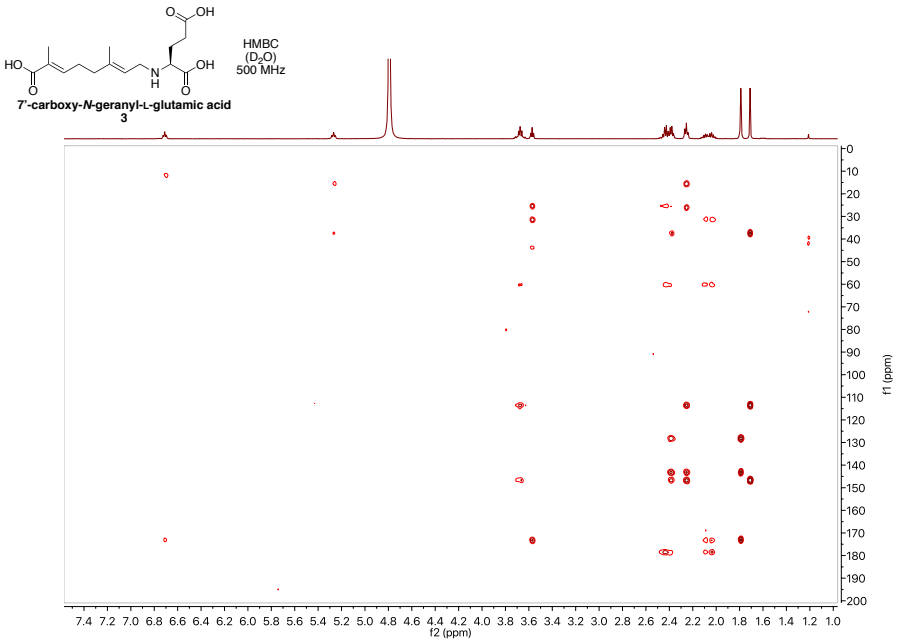
7'-carboxy-*N*-geranyl-L-glutamic acid (**3**):

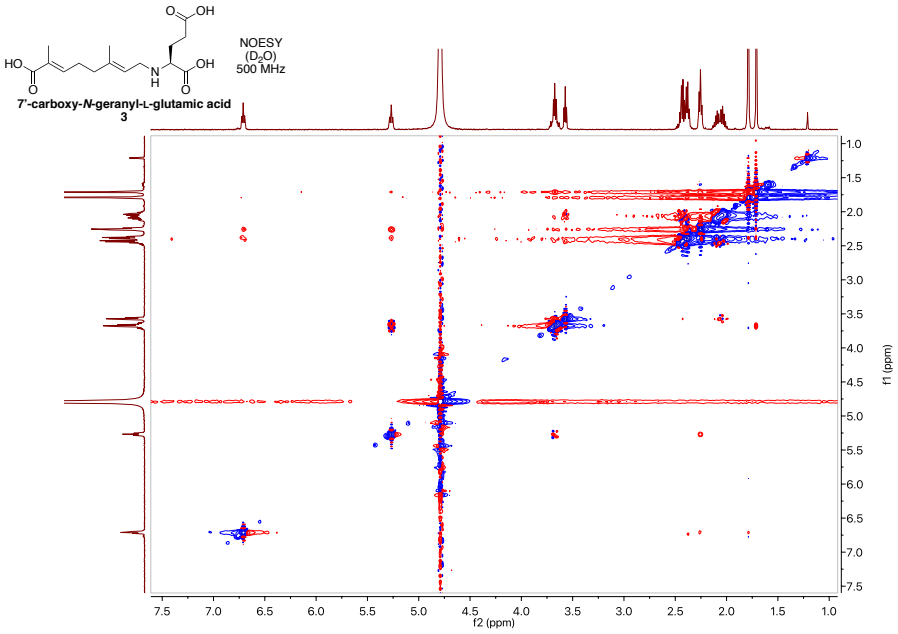


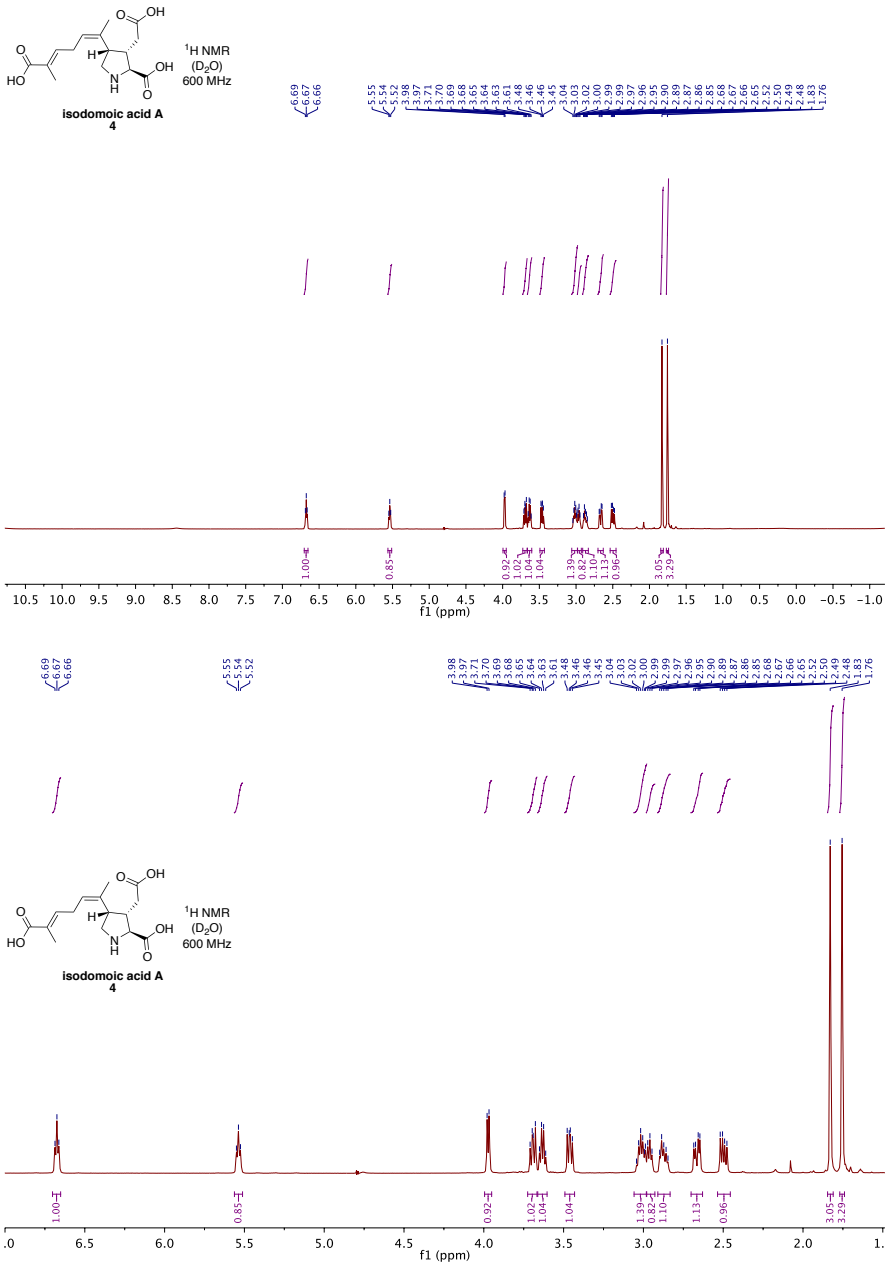



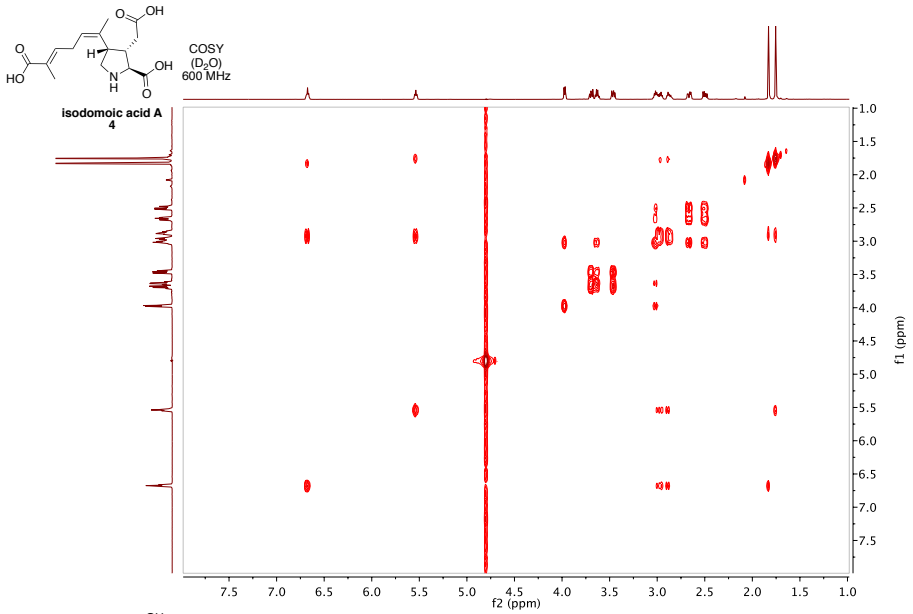

COSY
(D2O)
500 MHz
7'-carboxy-N-geranyl-L-glutamic acid
3
f2 (ppm)
f1 (ppm)

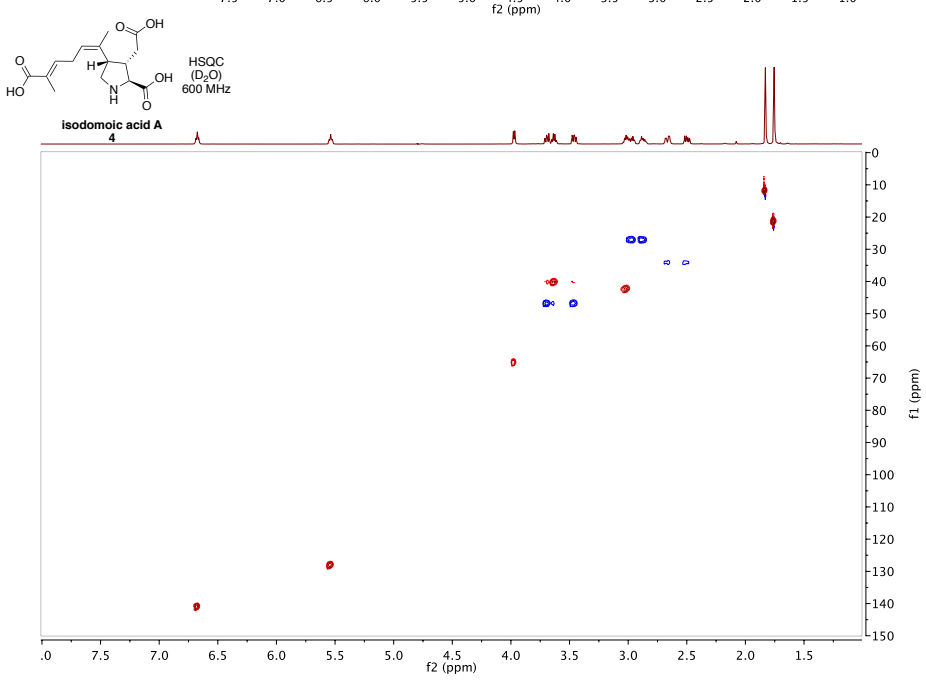

HSQC
(D2O)
600 MHz
7'-carboxy-N-geranyl-L-glutamic acid
3
f2 (ppm)
f1 (ppm)

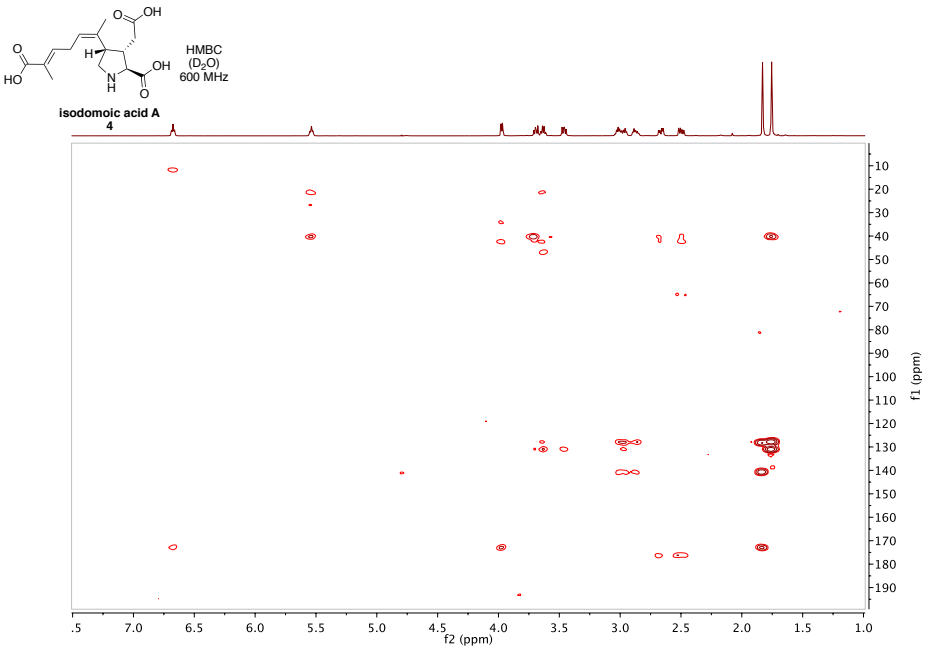

HMBC
(D2O)
500 MHz
7'-carboxy-N-geranyl-L-glutamic acid
3
f2 (ppm)
f1 (ppm)

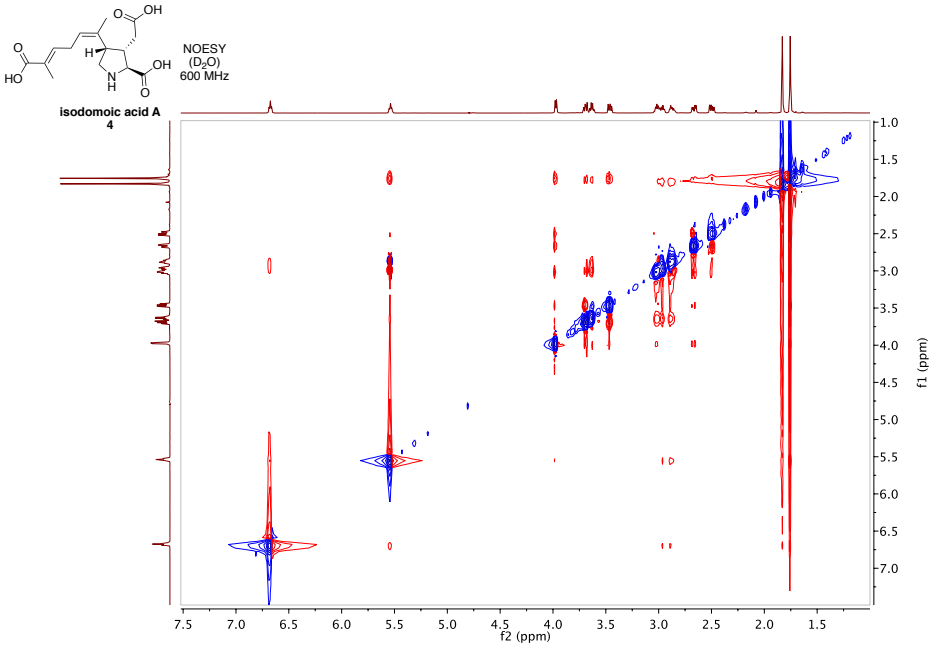

NOESY
(D2O)
500 MHz
7'-carboxy-N-geranyl-L-glutamic acid
3
f2 (ppm)
f1 (ppm)


isodomoic acid A
4
$^1$H NMR
($D_2O$)
600 MHz
f1 (ppm)
isodomoic acid A
4
$^1$H NMR
($D_2O$)
600 MHz
f1 (ppm)

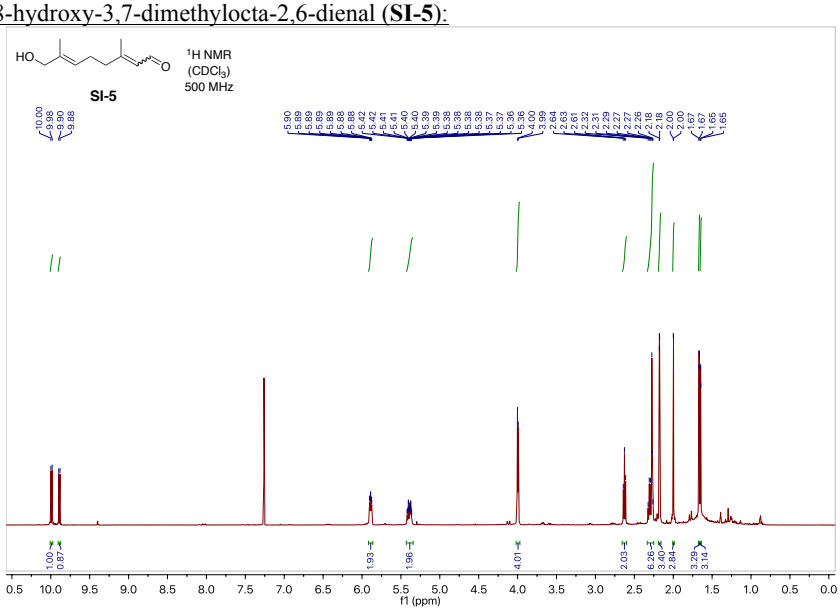

COSY
(D2O)
600 MHz
isodomoic acid A
4
f2 (ppm)
f1 (ppm)

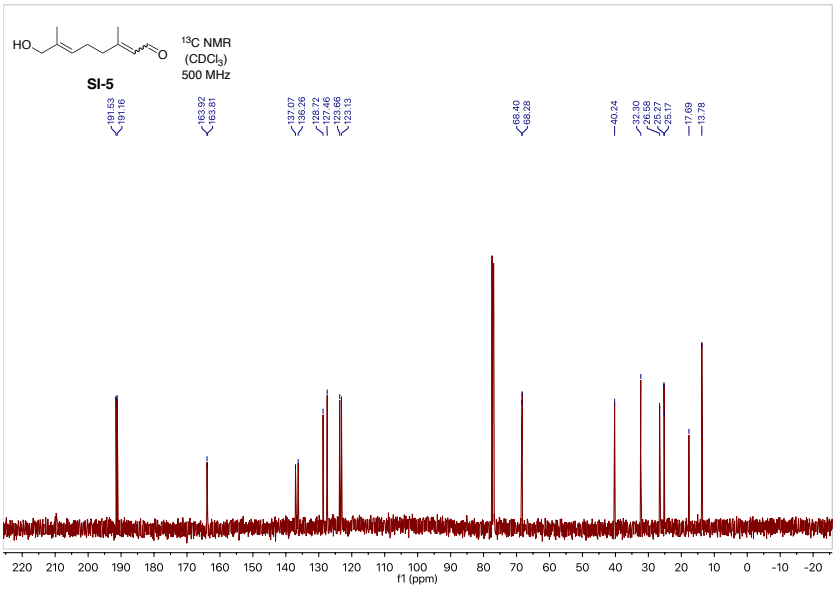

HSQC
(D2O)
600 MHz
isodomoic acid A
4
f2 (ppm)
f1 (ppm)


HMBC
(D2O)
600 MHz
isodomoic acid A
4
f2 (ppm)
f1 (ppm)

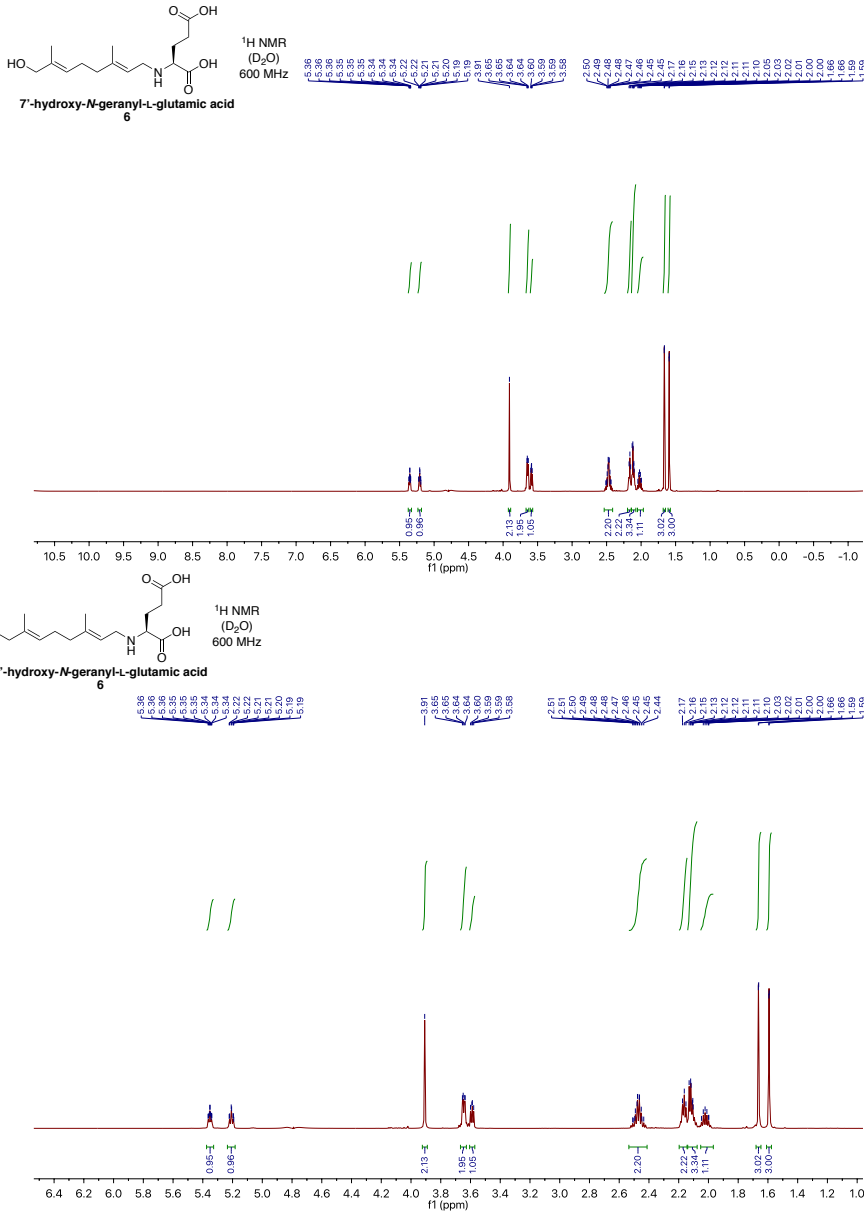

NOESY
(D2O)
600 MHz
isodomoic acid A
4
f2 (ppm)
f1 (ppm)

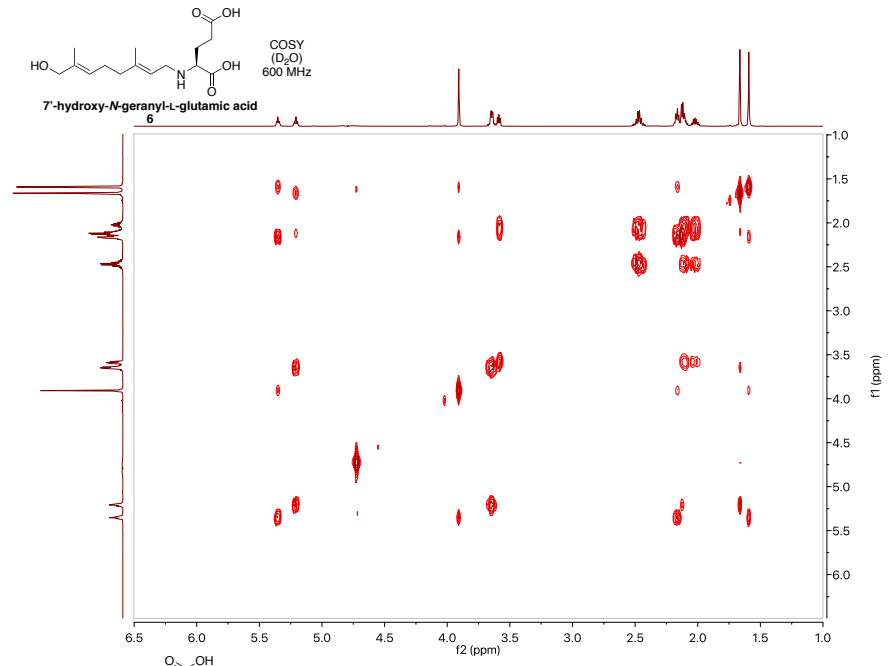
(6*E*)-8-hydroxy-3,7-dimethylocta-2,6-dienal (**SI-5**):

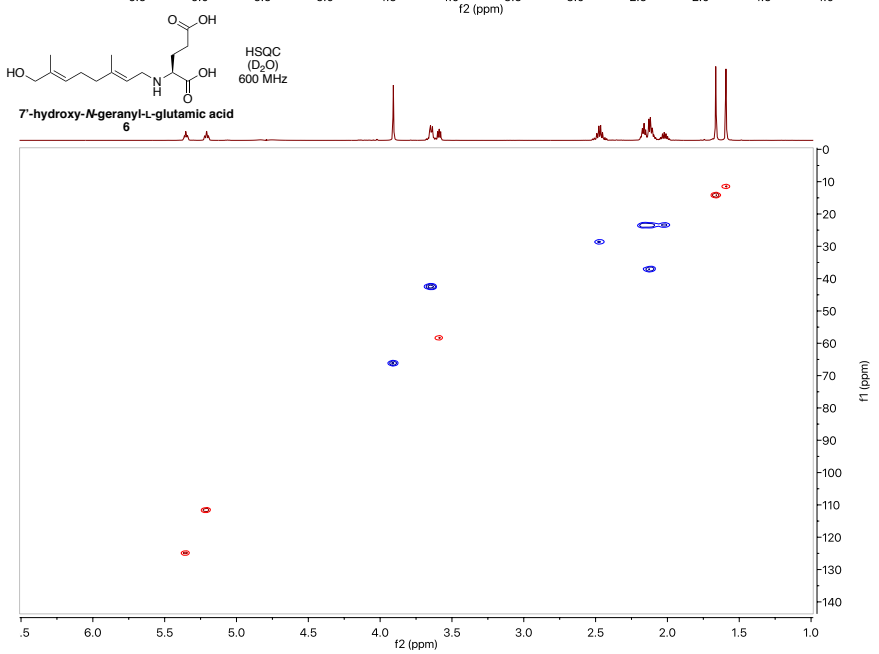


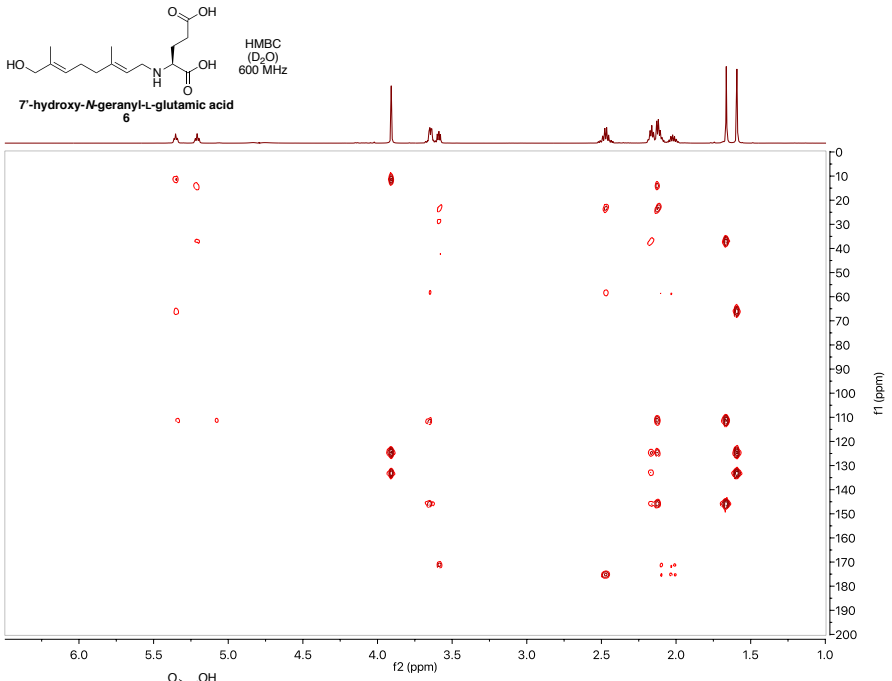



7'-hydroxy-*N*-geranyl-L-glutamic acid (**6**):




COSY
(D2O)
600 MHz
7'-hydroxy-N-geranyl-L-glutamic acid
6
f2 (ppm)
f1 (ppm)


HSQC
(D2O)
600 MHz
7'-hydroxy-N-geranyl-L-glutamic acid
6
f2 (ppm)
f1 (ppm)


HMBC
(D2O)
600 MHz
7'-hydroxy-N-geranyl-L-glutamic acid
6
f2 (ppm)
f1 (ppm)

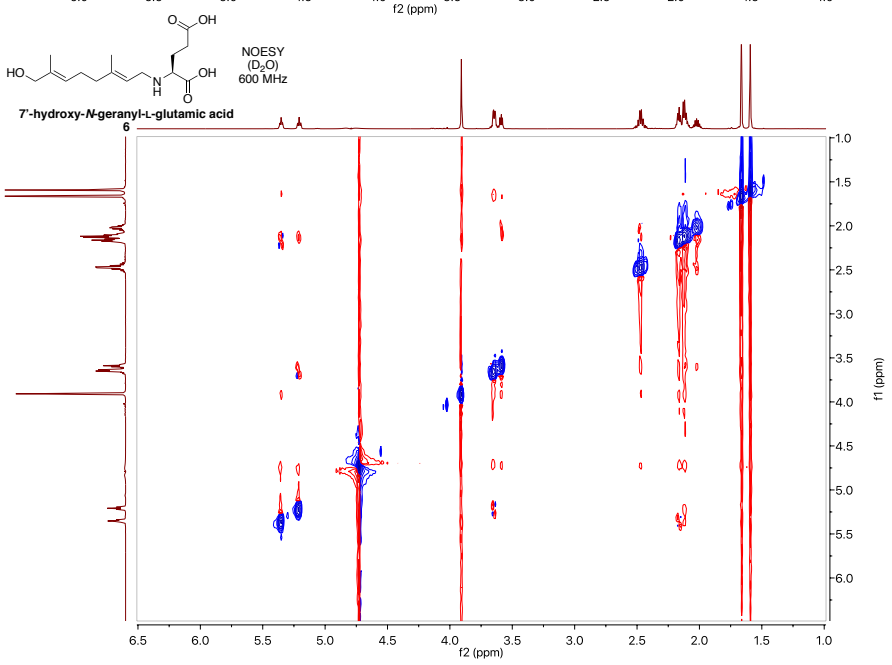

NOESY
(D2O)
600 MHz
7'-hydroxy-N-geranyl-L-glutamic acid
6
f2 (ppm)
f1 (ppm)

**Table S1. (separate file)**

Table showing RPKM, annotation, TMM values for edgeR, and edgeR differential expression (DE) analysis for all 19704 *P. multiseries* transcripts under the nutritional conditions described in this manuscript and a previous publication (*19*). Libraries prepared with either Ribo-Zero Magnetic kits (Illumina) or TruSeq RNA Sample Preparation Kit (Illumina) are labeled as "rRNA" or "truseq" respectively.

**Table S2. (separate file)**

Table showing RPKM, annotation, TMM values for edgeR, and edgeR differential expression (DE) analysis for *P. multiserieis* transcripts upregulated under phosphate starvation. Expression data (RPKM, TMM) is also shown for all of the nutritional conditions described in this manuscript and a previous publication (*19*). Libraries prepared with either Ribo-Zero Magnetic kits (Illumina) or TruSeq RNA Sample Preparation Kit (Illumina) are labeled as "rRNA" or "truseq" respectively.

**Table S3. (separate file)**

Table showing RPKM, annotation, TMM values for edgeR, and edgeR differential expression (DE) analysis for *P. multiserieis* transcripts upregulated under phosphate starvation and under increasing $pCO_2$ with constant low phosphate. Expression data (RPKM, TMM) is also shown for all of the nutritional conditions described in this manuscript and a previous publication (*19*). Libraries prepared with either Ribo-Zero Magnetic kits (Illumina) or TruSeq RNA Sample Preparation Kit (Illumina) are labeled as "rRNA" or "truseq" respectively.

**Table S4. (separate file)**

Table showing RPKM, annotation, TMM values for edgeR, and edgeR differential expression (DE) analysis for all CYP450 annotated transcripts. Expression data (RPKM, TMM) is shown for all of the nutritional conditions described in this manuscript and a previous publication (*19*). Libraries prepared with either Ribo-Zero Magnetic kits (Illumina) or TruSeq RNA Sample Preparation Kit (Illumina) are labeled as "rRNA" or "truseq" respectively.

**References and Notes**


1. M. L. Wells, V. L. Trainer, T. J. Smayda, B. S. O. Karlson, C. G. Trick, R. M. Kudela, A. Ishikawa, S. Bernard, A. Wulff, D. M. Anderson, W. P. Cochlan, Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful Algae* **49**, 68–93 (2015). doi:10.1016/j.hal.2015.07.009 Medline

2. S. M. McKibben, W. Peterson, A. M. Wood, V. L. Trainer, M. Hunter, A. E. White, Climatic regulation of the neurotoxin domoic acid. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 239–244 (2017). doi:10.1073/pnas.1606798114 Medline

3. C. J. Gobler, O. M. Doherty, T. K. Hattenrath-Lehmann, A. W. Griffith, Y. Kang, R. W. Litaker, Ocean warming since 1982 has expanded the niche of toxic algal blooms in the North Atlantic and North Pacific oceans. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4975–4980 (2017). doi:10.1073/pnas.1619575114 Medline

4. R. M. McCabe, B. M. Hickey, R. M. Kudela, K. A. Lefebvre, N. G. Adams, B. D. Bill, F. M. D. Gulland, R. E. Thomson, W. P. Cochlan, V. L. Trainer, An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions. *Geophys. Res. Lett.* **43**, 10366–10376 (2016). doi:10.1002/2016GL070023 Medline

5. G. R. Stewart, C. F. Zorumski, M. T. Price, J. W. Olney, Domoic acid: A dementia-inducing excitotoxic food poison with kainic acid receptor specificity. *Exp. Neurol.* **110**, 127–138 (1990). doi:10.1016/0014-4886(90)90057-Y Medline

6. J. A. Larm, P. M. Beart, N. S. Cheung, Neurotoxin domoic acid produces cytotoxicity via kainate- and AMPA-sensitive receptors in cultured cortical neurones. *Neurochem. Int.* **31**, 677–682 (1997). doi:10.1016/S0197-0186(97)00030-2 Medline

7. J. A. Funk, M. G. Janech, J. C. Dillon, J. J. Bissler, B. J. Siroky, P. D. Bell, Characterization of renal toxicity in mice administered the marine biotoxin domoic acid. *J. Am. Soc. Nephrol.* **25**, 1187–1197 (2014). doi:10.1681/ASN.2013080836 Medline

8. K. A. Lefebvre, P. S. Kendrick, W. Ladiges, E. M. Hiolski, B. E. Ferriss, D. R. Smith, D. J. Marcinek, Chronic low-level exposure to the common seafood toxin domoic acid causes cognitive deficits in mice. *Harmful Algae* **64**, 20–29 (2017). doi:10.1016/j.hal.2017.03.003 Medline

9. J. S. Ramsdell, T. S. Zabka, In utero domoic acid toxicity: A fetal basis to adult disease in the California sea lion (*Zalophus californianus*). *Mar. Drugs* **6**, 262–290 (2008). doi:10.3390/md6020262 Medline

10. L. M. Grattan, C. J. Boushey, Y. Liang, K. A. Lefebvre, L. J. Castellon, K. A. Roberts, A. C. Toben, J. G. Morris, Repeated dietary exposure to low levels of domoic acid and problems with everyday memory: Research to public health outreach. *Toxins (Basel)* **10**, 103 (2018). doi:10.3390/toxins10030103 Medline

11. P. F. Cook, C. Reichmuth, A. A. Rouse, L. A. Libby, S. E. Dennison, O. T. Carmichael, K. T. Kruse-Elliott, J. Bloom, B. Singh, V. A. Fravel, L. Barbosa, J. J. Stuppino, W. G. Van Bonn, F. M. D. Gulland, C. Ranganath, Algal toxin impairs sea lion memory and hippocampal connectivity, with implications for strandings. *Science* **350**, 1545–1547 (2015). doi:10.1126/science.aac5675 Medline

12. A. Lelong, H. Hégaret, P. Soudant, S. S. Bates, *Pseudo-nitzschia* (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: Revisiting previous paradigms. *Phycologia* **51**, 168–216 (2012). doi:10.2216/11-37.1

13. M. P. Sison-Mangus, S. Jiang, K. N. Tran, R. M. Kudela, Host-specific adaptation governs the interaction of the marine diatom, *Pseudo-nitzschia* and their microbiota. *ISME J.* **8**, 63–76 (2014). doi:10.1038/ismej.2013.138 Medline

14. U. P. Ramsey, D. J. Douglas, J. A. Walter, J. L. Wright, Biosynthesis of domoic acid by the diatom *Pseudo-nitzschia multiseries*. *Nat. Toxins* **6**, 137–146 (1998). doi:10.1002/(SICI)1522-7189(199805/08)6:3/4<137:AID-NT28>3.0.CO;2-L Medline

15. T. J. Savage, G. J. Smith, A. T. Clark, P. N. Saucedo, Condensation of the isoprenoid and amino precursors in the biosynthesis of domoic acid. *Toxicon* **59**, 25–33 (2012). doi:10.1016/j.toxicon.2011.10.010 Medline

16. X. Zhang, S. Li, Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* **34**, 1061–1089 (2017). doi:10.1039/C7NP00028F Medline

17. J. Sun, D. A. Hutchins, Y. Feng, E. L. Seubert, D. A. Caron, F.-X. Fu, Effects of changing $pCO_2$ and phosphate availability on domoic acid production and physiology of the marine harmful bloom diatom *Pseudo-nitzschia multiseries*. *Limnol. Oceanogr.* **56**, 829–840 (2011). doi:10.4319/lo.2011.56.3.0829

18. M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. de Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates, P. Cruz-Morales, S. Duddela, S. Düsterhus, D. J. Edwards, D. P. Fewer, N. Garg, C. Geiger, J. P. Gomez-Escribano, A. Greule, M. Hadjithomas, A. S. Haines, E. J. N. Helfrich, M. L. Hillwig, K. Ishida, A. C. Jones, C. S. Jones, K. Jungmann, C. Kegler, H. U. Kim, P. Kötter, D. Krug, J. Masschelein, A. V. Melnik, S. M. Mantovani, E. A. Monroe, M. Moore, N. Moss, H.-W. Nützmann, G. Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. Tian, N. J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gómez, A. Bechthold, H. B. Bode, R. Borriss, S. F. Brady, A. A. Brakhage, P. Caffrey, Y.-Q. Cheng, J. Clardy, R. J. Cox, R. De Mot, S. Donadio, M. S. Donia, W. A. van der Donk, P. C. Dorrestein, S. Doyle, A. J. M. Driessen, M. Ehling-Schulz, K.-D. Entian, M. A. Fischbach, L. Gerwick, W. H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Höfte, S. E. Jensen, J. Ju, L. Katz, L. Kaysser, J. L. Klassen, N. P. Keller, J. Kormanec, O. P. Kuipers, T. Kuzuyama, N. C. Kyrpides, H.-J. Kwon, S. Lautru, R. Lavigne, C. Y. Lee, B. Linquan, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Méndez, M. Metsä-Ketelä, J. Micklefield, D. A. Mitchell, B. S. Moore, L. M. Moreira, R. Müller, B. A. Neilan, M. Nett, J. Nielsen, F. O'Gara, H. Oikawa, A. Osbourn, M. S. Osburne, B. Ostash, S. M. Payne, J.-L. Pernodet, M. Petricek, J. Piel, O. Ploux, J. M. Raaijmakers, J. A. Salas, E. K. Schmitt, B. Scott, R. F. Seipke, B. Shen, D. H. Sherman, K. Sivonen, M. J. Smanski, M. Sosio, E. Stegmann, R. D. Süssmuth, K. Tahlan, C. M. Thomas, Y. Tang, A. W. Truman, M. Viaud, J. D. Walton, C. T. Walsh, T. Weber, G. P. van Wezel, B. Wilkinson, J. M. Willey, W. Wohlleben, G. D. Wright, N. Ziemert, C. Zhang, S. B. Zotchev, R. Breitling, E. Takano, F. O. Glöckner, Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015). doi:10.1038/nchembio.1890 Medline

19. E. V. Armbrust, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, D. S. Rokhsar, The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004). doi:10.1126/science.1101156 Medline

20. C. Bowler, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J.-P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M. P. Oudot-Le Secq, C. Napoli, M. Obornik, M. S. Parker, J.-L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, I. V. Grigoriev, The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008). doi:10.1038/nature07410 Medline

21. A. E. Allen, J. Laroche, U. Maheswari, M. Lommer, N. Schauer, P. J. Lopez, G. Finazzi, A. R. Fernie, C. Bowler, Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10438–10443 (2008). doi:10.1073/pnas.0711370105 Medline

22. G. Sapriel, M. Quinet, M. Heijde, L. Jourdren, V. Tanty, G. Luo, S. Le Crom, P. J. Lopez, Genome-wide transcriptome analyses of silicon metabolism in *Phaeodactylum tricornutum* reveal the multilevel regulation of silicic acid transporters. *PLOS ONE* **4**, e7458 (2009). doi:10.1371/journal.pone.0007458 Medline

23. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015). doi:10.1038/nprot.2015.053 Medline

24. Y. Maeno, Y. Kotaki, R. Terada, Y. Cho, K. Konoki, M. Yotsu-Yamashita, Six domoic acid related compounds from the red alga, *Chondria armata*, and domoic acid biosynthesis by the diatom, *Pseudo-nitzschia multiseries*. *Sci. Rep.* **8**, 356 (2018). doi:10.1038/s41598-017-18651-w Medline

25. J. Clayden, B. Read, K. R. Hebditch, Chemistry of domoic acid, isodomoic acids, and their analogues. *Tetrahedron* **61**, 5713–5724 (2005). doi:10.1016/j.tet.2005.04.003

26. I. Nitta, H. Watase, Y. Tomiie, Structure of kainic acid and its isomer, allokainic acid. *Nature* **181**, 761–762 (1958). doi:10.1038/181761b0 Medline

27. K. Konno, H. Shirahama, T. Matsumoto, Isolation and structure of acromelic acid A and B. New kainoids of *Clitocybe acromelalga*. *Tetrahedron Lett.* **24**, 939–942 (1983). doi:10.1016/S0040-4039(00)81571-3

28. G. W. Saunders, C. Jackson, E. D. Salomaki, Phylogenetic analyses of transcriptome data resolve familial assignments for genera of the red-algal Acrochaetiales-Palmariales Complex (Nemaliophycidae). *Mol. Phylogenet. Evol.* **119**, 151–159 (2018). doi:10.1016/j.ympev.2017.11.002 Medline

29. N. Matasci, L.-H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, T. Warnow, S. Ayyampalayam, M. Barker, J. G. Burleigh, M. A. Gitzendanner, E. Wafula, J. P. Der, C. W. dePamphilis, B. Roure, H. Philippe, B. R. Ruhfel, N. W. Miles, S. W. Graham, S. Mathews, B. Surek, M. Melkonian, D. E. Soltis, P. S. Soltis, C. Rothfels, L. Pokorny, J. A. Shaw, L. DeGironimo, D. W. Stevenson, J. C. Villarreal, T. Chen, T. M. Kutchan, M. Rolf, R. S. Baucom, M. K. Deyholos, R. Samudrala, Z. Tian, X. Wu, X. Sun, Y. Zhang, J. Wang, J. Leebens-Mack, G. K.-S. Wong, Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**, 17 (2014). doi:10.1186/2047-217X-3-17 Medline

30. S. Basu, S. Patil, D. Mapleson, M. T. Russo, L. Vitale, C. Fevola, F. Maumus, R. Casotti, T. Mock, M. Caccamo, M. Montresor, R. Sanges, M. I. Ferrante, Finding a partner in the ocean: Molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol.* **215**, 140–156 (2017). doi:10.1111/nph.14557 Medline

31. P. J. Keeling, F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. B. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone, A. Z. Worden, The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol.* **12**, e1001889 (2014). doi:10.1371/journal.pbio.1001889 Medline

32. M. Rho, H. Tang, Y. Ye, FragGeneScan: Predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191 (2010). doi:10.1093/nar/gkq747 Medline

33. E. M. Bertrand, J. P. McCrow, A. Moustafa, H. Zheng, J. B. McQuaid, T. O. Delmont, A. F. Post, R. E. Sipler, J. L. Spackeen, K. Xu, D. A. Bronk, D. A. Hutchins, A. E. Allen, Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 9938–9943 (2015). doi:10.1073/pnas.1501615112 Medline

34. R. Schmieder, Y. W. Lim, R. Edwards, Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2012). doi:10.1093/bioinformatics/btr669 Medline

35. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN] (2013).

36. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). doi:10.1093/bioinformatics/btt656 Medline

37. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010). doi:10.1093/bioinformatics/btp616 Medline

38. R. A. Anderson, Ed., *Algal Culturing Techniques* (Elsevier, 2005).

39. R. E. Diner, V. A. Bielinski, C. L. Dupont, A. E. Allen, P. D. Weyman, Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. *Front. Bioeng. Biotechnol.* **4**, 65 (2016). doi:10.3389/fbioe.2016.00065 Medline

40. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison 3rd, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009). doi:10.1038/nmeth.1318 Medline

41. J. D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004). doi:10.1016/j.jmb.2004.05.028 Medline

42. A. Gruber, G. Rocap, P. G. Kroth, E. V. Armbrust, T. Mock, Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **81**, 519–528 (2015). doi:10.1111/tpj.12734 Medline

43. P. Savitsky, J. Bray, C. D. O. Cooper, B. D. Marsden, P. Mahajan, N. A. Burgess-Brown, O. Gileadi, High-throughput production of human proteins for crystallization: The SGC experience. *J. Struct. Biol.* **172**, 3–13 (2010). doi:10.1016/j.jsb.2010.06.008 Medline

44. C. A. Miller 3rd, M. A. Martinat, L. E. Hyman, Assessment of aryl hydrocarbon receptor complex interactions using pBEVY plasmids: Expressionvectors with bi-directional promoters for use in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **26**, 3577–3583 (1998). doi:10.1093/nar/26.15.3577 Medline

45. E. W. Jones, Tackling the protease problem in *Saccharomyces cerevisiae*. *Methods Enzymol.* **194**, 428–453 (1991). doi:10.1016/0076-6879(91)94034-A Medline

46. D. C. Amberg, D. J. Burke, J. N. Strathern, *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press, 2005).

47. D. Pompon, B. Louerat, A. Bronine, P. Urban, Yeast expression of animal and plant P450s in optimized redox environments. *Methods Enzymol.* **272**, 51–64 (1996). doi:10.1016/S0076-6879(96)72008-6 Medline

48. V. V. Kushnirov, Rapid and reliable protein extraction from yeast. *Yeast* **16**, 857–860 (2000). doi:10.1002/1097-0061(20000630)16:9<857:AID-YEA561>3.0.CO;2-B Medline

49. M. Kahru, R. M. Kudela, M. Manzano-Sarabia, B. G. Mitchell, Trends in the surface chlorophyll of the California Current: Merging data from multiple ocean color satellites. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **77–80**, 89–98 (2012). doi:10.1016/j.dsr2.2012.04.007

50. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010). doi:10.1093/molbev/msp259 Medline

51. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). doi:10.1093/molbev/msu300 Medline

52. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013). doi:10.1093/molbev/mst024 Medline

53. N. Lartillot, T. Lepage, S. Blanquart, PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009). doi:10.1093/bioinformatics/btp368 Medline

54. N. P. Dunham, W. C. Chang, A. J. Mitchell, R. J. Martinie, B. Zhang, J. A. Bergman, L. J. Rajakovich, B. Wang, A. Silakov, C. Krebs, A. K. Boal, J. M. Bollinger Jr., Two distinct mechanisms for C–C desaturation by iron(II)- and 2-(oxo)glutarate-dependent oxygenases: Importance of α-heteroatom assistance. *J. Am. Chem. Soc.* **140**, 7116–7126 (2018). doi:10.1021/jacs.8b01933 Medline

55. M.-C. Tang, Y. Zou, K. Watanabe, C. T. Walsh, Y. Tang, Oxidative cyclization in natural product biosynthesis. *Chem. Rev.* **117**, 5226–5333 (2017). doi:10.1021/acs.chemrev.6b00478 Medline

56. R. D. Finn, T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, A. J. Bridge, H.-Y. Chang, Z. Dosztányi, S. El-Gebali, M. Fraser, J. Gough, D. Haft, G. L. Holliday, H. Huang, X. Huang, I. Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D. A. Natale, M. Necci, G. Nuka, C. A. Orengo, Y. Park, S. Pesseat, D. Piovesan, S. C. Potter, N. D. Rawlings, N. Redaschi, L. Richardson, C. Rivoire, A. Sangrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato, G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, I. Xenarios, L.-S. Yeh, S.-Y. Young, A. L. Mitchell, InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45** (D1), D190–D199 (2017). doi:10.1093/nar/gkw1107 Medline

57. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29-37 (2011). doi:10.1093/nar/gkr367 Medline

58. A. B. Woodside, Z. Huang, C. D. Poulter, Trisammonium geranyl diphosphate. *Org. Synth.* **66**, 211 (1988). doi:10.15227/orgsyn.066.0211

59. G. Verardo, P. Geatti, E. Pol, A. G. Giumanini, Sodium borohydride: A versatile reagent in the reductive N -monoalkylation of α-amino acids and α-amino methyl esters. *Can. J. Chem.* **80**, 779–788 (2002). doi:10.1139/v02-083

60. G. R. Labadie, R. Viswanathan, C. D. Poulter, Farnesyl diphosphate analogues with ω-bioorthogonal azide and alkyne functional groups for protein farnesyl transferase-catalyzed ligation reactions. *J. Org. Chem.* **72**, 9291–9297 (2007). doi:10.1021/jo7017747 Medline

61. Y.-J. Zhao, T.-P. Loh, Practical synthesis of 1,5-dimethyl substituted conjugated polyenes from geranyl acetate. *Tetrahedron* **64**, 4972–4978 (2008). doi:10.1016/j.tet.2008.03.094

62. H. L. Jackson, G. T. Nadolski, C. Braun, S. F. Lockwood, Efficient total synthesis of lycophyll (ψ,ψ-Carotene-16,16‘-diol). *Org. Process Res. Dev.* **9**, 830–836 (2005). doi:10.1021/op050137f

63. S. Xie, S. Uesato, T. Fujita, H. Inouye, Biosynthesis of iridoid glucosides in *Patrinia gibbosa*. *J. Nat. Prod.* **52**, 701–705 (1989). doi:10.1021/np50064a005

64. J. A. Walter, M. Falk, J. L. C. Wright, Chemistry of the shellfish toxin domoic acid: Characterization of related compounds. *Can. J. Chem.* **72**, 430–436 (1994). doi:10.1139/v94-064

## 2.4 Acknowledgements and Author Contributions

The main text and supplemental information sections of Chapter 2 are reprints of the material as it appears "**J. K. Brunson***, S. M. K. McKinnie*, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. Smith, D. A. Hutchins, A. E. Allen, B. S. Moore. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* 361(6409), 1356-1358 (2018)." The dissertation author was one of two equally contributing primary investigators and authors of this manuscript.

J.K.B., S.M.K.M., J.R.C., A.E.A., and B.S.M. conceived the project, designed the experiments, analyzed the data, and wrote the paper, with input from all authors. J.K.B., S.M.K.M., J.R.C., and Z.D.M. performed protein expression and did the enzymology experiments. S.M.K.M and J.K.B. performed chemical synthesis and structural characterization of enzymatic products. J.P.M., E.M.B., J.K.B., and A.E.A. performed the transcriptome analyses. V.A.B. designed vectors and expression constructs and assisted with sequencing. H.L. designed the yeast expression experiments. M.O. constructed the phylogeny. G.J.S. provided the P. multiseries isolate 15091C3. D.A.H. provided P. multiseries biomass for RNA sequencing from the DA induction experiments and helped to design RNA sequencing experiments.

We also acknowledge B. Duggan, A. Mrse, and Y. Su (all University of California, San Diego) for assistance and maintenance of nuclear magnetic resonance and high-resolution mass spectrometry machinery; W. Fenical (Scripps Institution of Oceanography) for helpful discussion and feedback; and M. Kahru (Scripps Institution of Oceanography) for providing processed satellite images. This work was supported by grants from the National Science Foundation (NSF OCE-1313747 to B.S.M., NSF-ANT-1043671 to A.E.A., and OCE 1538525 and OCE 1638804 to D.A.H.), the National Institute of Environmental Health Sciences (NIEHS P01-ES021921 to

B.S.M.), the U.S. Department of Energy Genomics Science program (DE-SC0008593 and DE-SC0018344 to A.E.A.), the Gordon and Betty Moore Foundation (GBMF3828 to A.E.A. and GBMF4960 to G.J.S.), and the Czech Science Foundation (18-13458S to M.O.) and funding from the National Institutes of Health (Chemical Biology Interfaces–University of California, San Diego, training grant 5T32GM112584 to J.K.B.), the Dickinson-McCrink Fellowship (J.K.B.), the Natural Sciences and Engineering Research Council of Canada (NSERC-PDF to S.M.K.M.), and the Simons Foundation Fellowship of the Life Sciences Research Foundation (J.R.C.)

# Chapter 3: Domoic acid biosynthesis in red alga *Chondria armata* suggests a complex evolutionary history for toxin production

## 3.1 Introduction to Chapter 3

Domoic acid (DA) and the structurally related molecule kainic acid (KA) were originally isolated from Japanese seaweeds in the 1950s.[1–3] The chemical constituents of seaweeds were of particular interest due to their historical use as de-worming, or anthelmintic, agents in Japan for hundreds of years. Indeed, both DA and KA display potent activity against roundworm infections, and both compounds were subsequently named after the common, Japanese names for the anthelmintic seaweeds they were isolated from.[1,4] The red macroalgal seaweed *Chondria armata*, known colloquially in Japan as "domoi", was the first described source of DA, and KA was originally isolated from the red algae *Digenea simplex*, referred to locally as "kaininsō" seaweed.[5]

After our initial discovery of DA biosynthesis in the diatom *Pseudo-nitzschia multiseries*, as described in Chapter 2, our lab next turned to unraveling the biosynthesis of KA in seaweeds. Fortunately, our work on DA biosynthesis revealed two unique enzymes involved in building the characteristic pyrrolidine core shared by both DA and KA: a glutamate *N*-prenyltransferase and an alpha-ketoglutarate dependent kainoid synthase.[6] Although the glutamate *N*-prenyltransferase has substantial structural homology to known bacterial terpene cyclases, the enzyme appears to be especially unique in its amino acid sequence with no clear homologous sequences found in the standard NCBI database.[7] The kainoid synthase enzyme is also unique in amino acid sequence space, displaying only limited homology to known bacterial biosynthetic enzymes involved in penicillin (isopenicillin N synthase, IPNS) and cephalosporin (deacetoxycephalosporin C synthase, DAOCS) biosynthesis.[8,9] The novelty of the gene sequences encoding glutamate *N*-prenyltransferase and kainoid synthase activities provide a useful roadmap or "bioinformatic hook" for identifying DA and KA biosynthesis in other

organisms (as described here, in Chapter 3) and in the broader marine environment (as described in Chapter 4).

Our group leveraged this genetic information to identify candidate KA biosynthetic (*kab*) genes in publicly available red algal transcriptomic datasets.[10] To confirm the presence of these candidate genes in KA-producing seaweeds, our group sequenced draft genomes for *D. simplex* (aka "kaininsō") and *Palmaria palmata* using Oxford Nanopore Technologies (ONT) sequencing. The ONT approach has gained recent popularity for its ability to generate long reads of nucleotide sequencing data at relatively low cost, an incredible boon for generating a draft genome sequence and understanding genomic organization in various organisms.[11–14] Following assembly of ONT data, it was revealed that the *kab* genes also co-cluster in the *D. simplex* and *P. palmata* genomes, a phenomena also reflected by the diatom DA biosynthetic (*dab*) gene cluster. Functional characterization of purified KabA and KabC enzymes revealed a succinct two step biosynthetic pathway to KA using L-glutamate and dimethyl-allyl pyrophosphate as precursors.[10] This pathway is highly reminiscent of DA biosynthesis in diatoms, featuring the key glutamate *N*-prenyltransferase and kainoid synthase activities as inferred by bioinformatic analysis.[6]

Description of kainoid biosynthetic genes in both seaweeds and diatoms raises quite a few questions regarding the evolution of DA and KA biosynthesis. With respect to DA production, only a handful of red macroalgae have been described to produce the potent neurotoxin, all within the family Rhodomelaceae.[15,16] In diatoms, DA biosynthesis appears to be even more constrained. The diatom genus *Pseudo-nitzschia* contains a sizable number of DA producing species, with 26 of 52 known species demonstrating toxin production in culture according to a recent review.[17] Besides *Pseudo-nitzschia*, the genus *Nitzschia* is the only other taxonomic group containing species capable of DA production. Within *Nitzschia*, only *N.*

*bizertensis* and *N. navis-vargincia* have been described to make DA.[18,19] These two species do not clade well with *Pseudo-nitzschia* or even other *Nitzschia*, and the latter genus now appears to be polyphyletic according to recent molecular phylogenetics.[20] Although KA production appears to be more widespread than DA production in seaweeds, KA production is nevertheless sparse among described red algal orders and families and is seemingly absent from the described pool of diatom metabolites.[15,16] Neither molecule has been described in any other macroalgal or microalgal lineage besides red algae and diatoms, respectively.

At present, the evolutionary history of the kainoid molecules DA and KA remains unclear. However, the similarity between the diatom and red algal enzyme activities and gene sequences suggests a potential origin via horizontal gene transfer (HGT), especially considering the large evolutionary divergence between these algal clades. While diatoms acquired many genes directly from red algae through endosymbiosis events in deep evolutionary time, one would expect a broader distribution of kainoid biosynthesis in both lineages if toxin biosynthesis was inherited in this manner.[21,22] An HGT-based hypothesis is further buoyed by the observation that no clear ancestor of the glutamate *N*-prenyltransferase (DabA/KabA) or kainoid synthase (DabC/KabC) exists in sequenced diatom and seaweed genomes. However, acquisition of kainoid biosynthetic enzymes through HGT cannot be confidently assigned in the absence of additional sequencing information. The “smoking gun” sequencing experiment would identify a discrete bacterial host or viral intermediate capable of moving kainoid biosynthetic enzymes between the disparate algal taxa. Such sequencing data has not been generated or described to date.

Despite this apparent “missing link”, HGT is an established phenomenon in both diatoms and reg algae. Approximately 3-5% of diatom protein-coding genes appear to be of relatively

recent acquisition from bacteria.[23,24] Perhaps the best illustration of diatom HGT is exemplified by the presence of an ornithine-urea cycle.[25] While this metabolic pathway is common in metazoans for the purpose of excreting nitrogen in the form of urea, diatoms have leveraged the pathway to redistribute organic nitrogen sources in response to nitrogen availability.[26] Many of the enzymes enabling the diatom ornithine-urea cycle have been acquired laterally from bacteria, creating a "Frankenstein pathway" together with native host enzymology.[25] Among red algae, HGT has been suggested for the acquisition of bacterial carbonic anhydrases to aid in carbon concentration.[27] Notably, a few terpene cyclases have been described in red algae that appear to be of bacterial origin.[28] While the glutamate *N*-prenyltransferase reaction in DA and KA biosynthesis is catalyzed by a terpene cyclase-like enzyme, these *N*-prenyltransferases appear to be unrelated to the enzymes previously described in red algae.[7]

To better understand the evolution of DA biosynthesis, we sequenced the genome of the originally-described DA-producer, *Chondria armata* ("domoi") using a mix of ONT long-read sequencing and Illumina short-read sequencing. To our surprise, we discovered a compact red algal DA ("*rad*") gene cluster encoding the three major steps DA biosynthesis: glutamate *N*-prenylation (*radA*), kainoid ring cyclization (*radC*) and presumed CYP450 sidechain hydroxylation (*radD*). While incomplete copies of this gene cluster exists elsewhere in the *C. armata* genome, the complete *rad1* cluster is nearly identical to the cluster we originally described in diatoms with respect to gene synteny, or the order that genes appear in the genome along a strand of DNA.[6] Heterologous expression and purification of RadA1 and RadC1 proteins from *Escherichia coli* yielded soluble enzymes capable of performing the key chemical transformations required to make DA. Comparison of RadC1 activity to diatom DabC and *D. simplex* KabC activity further implicates RadC1 in DA biosynthesis and begins to paint an

interesting evolutionary picture. While RadC1 is more like the KA-producing KabC in terms of amino acid sequence, the enzyme displays an activity much more like the diatom-derived DabC enzyme involved in DA biosynthesis.

Phylogenetic analysis of the kainoid synthase enzymes further supports these sequence-based relationships, and an HGT-derived origin for these enzymes is supported by monophyletic cladding of the DabC/KabC/RadC sequences in spite of the evolutionary divergence of their respective red algal and diatom hosts. Meanwhile, the RadD1 enzyme clades strictly with other red algal CYP450 enzymes, while DabD clades nicely with other diatom CYP450 sequences. This phenomenon suggests that host CYP450 enzymology was neofunctionalized and repurposed for the installation of the characteristic 7'-carboxylic acid on the DA alkyl side chain. While the true ecological role of DA remains unknown, the observation that neofunctionalization of native biochemistry may have happened two times in the evolution towards DA biosynthesis in diatoms and red algae underscores a potential advantage afforded by DA production in algae, shedding additional light on the mysterious origins of this marine neurotoxin.

## 3.2 Works Cited for the Chapter Introduction


1. K. Daigo. Studies on the constituents of *Chondria armata*. II: Isolation of an anthelmintical constituent. *Yakugaku Zasshi* **79**, 353–356 (1959).

2. S. Murakami, T. Takemoto, Z. Shimizu. Studies on the effective principles of *Digenea simplex* Aq. I. *Yakugaku Zasshi* **73(9)**, 1026–1028 (1953)

3. I. Nitta, H. Watase, Y. Tomiie. Structure of kainic acid and its isomer, allokainic acid. *Nature*. **181(4611)**, 761–762 (1958)

4. Y. Komiya, A. Kobayashi. Techniques applied in Japan for the control of *Ascaris* and hookworm infections -- a review. *Jpn. J. Med. Sci. Biol.* **18**, 1–17 (1965)

5. W. H. Gerwick, Plant sources of drugs and chemicals. *Encyclopedia of Biodiversity* **2**, 129-139 (2013)

6. J. K. Brunson, S. M. K. McKinnie, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. Smith, D. A. Hutchins, A. E. Allen, B. S. Moore. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018)

7. J. R. Chekan, S. M. K. McKinnie, J. P. Noel, B. S. Moore, Algal neurotoxin biosynthesis repurposes the terpene cyclase structural fold into an *N*-prenyltransferase. *Proc. Natl. Acad. Sci.* **117**, 12799–12805 (2020)

8. P. L. Roach, I. J. Clifton, V. Fülöp, K. Harlos, G. J. Barton, J. Hajdu, I. Andersson, C. J. Schofield, J. E. Baldwin. Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature* **375(6533)**, 700–704 (1995)

9. M. D. Lloyd, H-J Lee, K. Harlos, Z-H Zhang, J. E. Baldwin, C. J. Schofield, J. M. Charnock, C. D. Garner, T. Hara, A. C. Terwisscha van Scheltinga, K. Valegård, J. A. Viklund, J. Hajdu, I. Andersson, Å Danielsson, R. Bhikhabhai. Studies on the active site of deacetoxycephalosporin C synthase. *J. Mol. Biol.* **287(5)**, 943-960 (1999)

10. J. R. Chekan, S. M. K. McKinnie, M. L. Moore, S. G. Poplawski, T. P. Michael, B. S. Moore. Scalable biosynthesis of the seaweed neurochemical, kainic acid. *Angew. Chem. Int. Ed.* **58**, 8454–8457 (2019)

11. M. Jain, H. E. Olsen, B. Paten, M. Akeson. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. **17(1)**, 239 (2016)

12. W. De Coster, P. De Rijk, A. De Roeck, T. De Pooter, S. D'Hert, M. Strazisar, K. Sleegers, C. Van Broeckhoven. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res.* **29(7)**, 1178–1187 (2019)

13. S. Goldstein, L. Beka, J. Graf, J. L. Klassen. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genomics* **20(1)**, 23 (2019)

14. T. P. Michael, F. Jupe, F. Bemm, S. T. Motley, J. P. Sandoval, C. Lanz, O. Loudet, D. Weigel, J. R. Ecker. High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* **9(1)**, 541 (2018)

15. M. V. Laycock, A. S. W. de Freitas, J. L. C. Wright. Glutamate agonists from marine algae. *J. Appl. Phycol.* 1(2), 113–122 (1989)

16. M. Sato, T. Nakano, M. Takeuchi, N. Kanno, E. Nagahisa, Y. Sato. Distribution of neuroexcitatory amino acids in marine algae. *Phytochemistry* **42(6)**, 1595–1597 (1996)

17. S. S. Bates, K. A. Hubbard, N. Lundholm, M. Montresor, C. P. Leaw, *Pseudo-nitzschia*, *Nitzschia*, and domoic acid: new research since 2011. *Harmful Algae* **79**, 3–43 (2018).

18. D. B. Smida, N. Lundholm, W. H. C. F. Kooistra, I. Sahraoui, M. V. Ruggiero, Y. Kotaki, M. Ellegaard, C. Lambert, H. H. Mabrouk, A. S. Hlaili. Morphology and molecular phylogeny of *Nitzschia bizertensis sp. nov.*—a new domoic acid-producer. *Harmful Algae* **32,** 49–63 (2014)

19. N. Lundholm, Ø. Jvind Moestrup. Morphology of the marine diatom *Nitzschia navis-varingica, sp. nov.* (Bacillariophyceae), another producer of the neurotoxin domoic acid. *J. Phycol.* **36(6)**, 1162–1174 (2000)

20. D. G. Mann, R. Trobajo, S. Sato, C. Li, A. Witkowski, F. Rimet, M. P. Ashworth, R. M. Hollands, E. C. Theriot. Ripe for reassessment: a synthesis of available molecular data for the speciose diatom family Bacillariaceae. *Mol. Phylogenet. Evol.* **158**, 106985 (2021)

21. D. Moreira, P. Deschamps. What was the real contribution of endosymbionts to the eukaryotic nucleus? Insights from photosynthetic eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6**, a016014 (2014)

22. H. Qiu, H. S. Yoon, D. Bhattacharya, Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Front. Plant Sci.* **4**, 366 (2013)

23. E. Vancaester, T. Depuydt, C. M. Osuna-Cruz, K. Vandepoele, Comprehensive and functional analysis of horizontal gene transfer events in diatoms. *Mol. Biol. Evol.* **37**, 3243–3257 (2020)

24. C. Bowler, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J-P Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K.

Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin–Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M-P. O. Secq, C. Napoli, M. Obornik, M. S. Parker, J-L Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, I. V. Grigoriev. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456(7219)**, 239–244 (2008)

25. A. E. Allen, C. L. Dupont, M. Oborník, A. Horák, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie, C. Bowler. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**, 203–207 (2011)

26. S. R. Smith, C. L. Dupont, J. K. McCarthy, J. T. Broddrick, M. Oborník, A. Horák, Z. Füssy, J. Cihlář, S. Kleessen, H. Zheng, J. P. McCrow, K. K. Hixson, W. L. Araújo, A. Nunes-Nesi, A. Fernie, Z. Nikoloski, B. O. Palsson, A. E. Allen. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* **10**, 4552 (2019)

27. D. Wang, X. Yu, K. Xu, G. Bi, M. Cao, E. Zelzion, C. Fu, P. Sun, Y. Liu, F. Kong, G. Du, X. Tang, R. Yang, J. Wang, L. Tang, L. Wang, Y. Zhao, Y. Ge, Y. Zhuang, Z. Mo, Y. Chen, T. Gao, X. Guan, R. Chen, W. Qu, B. Sun, D. Bhattacharya, Y. Mao. *Pyropia yezoensis* genome reveals diverse mechanisms of carbon acquisition in the intertidal environment. *Nat. Commun.* **11**, 4028 (2020)

28. G. Wei, Q. Jia, X. Chen, T. G. Köllner, D. Bhattacharya, G. K-S. Wong, J. Gershenzon, F. Chen. Terpene biosynthesis in red algae is catalyzed by microbial type but not typical plant terpene synthases. *Plant Physiol.* **179**, 382–390 (2019)

## 3.3 Reproduction of manuscript submission: "Domoic acid biosynthesis in red alga *Chondria armata* suggests a complex evolutionary history for toxin production"

### 3.3.1 Main Text

#### 3.3.1.1 Abstract

Domoic acid (DA), the causative agent of amnesic shellfish poisoning, is produced by select organisms within two distantly related algal clades: planktonic diatoms and red macroalgae. The biosynthetic pathway to isodomoic acid A was recently solved in the harmful algal bloom-forming diatom *Pseudo-nitzschia multiseries*, establishing the genetic basis for the global production of this potent neurotoxin. Herein we sequenced the genome of *Chondria armata*, the red macroalgal seaweed from which DA was first isolated in the 1950s and identified several copies of the red algal DA (*rad*) biosynthetic gene cluster. The *rad* genes are organized similarly to the diatom DA biosynthesis cluster in terms of gene synteny, including a cytochrome P450 (CYP450) enzyme critical to DA production that is notably absent in red algae that produce the simpler kainoid neurochemical, kainic acid (KA). Biochemical characterization of the *N*-prenyltransferase (RadA) and kainoid synthase (RadC) enzymes support a slightly altered DA biosynthetic model in *C. armata* via the congener isodomoic acid B, with RadC behaving more similarly to the homologous diatom enzyme despite higher amino acid similarity to red algal kainic acid synthesis enzymes. Phylogenetic analysis of the *rad* genes suggests unique origins for the red macroalgal and diatom genes in their respective hosts, with native eukaryotic CYP450 neofunctionalization combining with horizontal gene transfer of *N*-prenyltransferases and kainoid synthases to establish DA production within the algal lineages.

### 3.3.1.2 Introduction

Harmful algal blooms of the diatom genus *Pseudo-nitzschia* produce high levels of domoic acid (DA), a neurotoxic glutamate receptor agonist with far-reaching food web implications due to its bioaccumulation in shellfish.[1] Consumption of seafood contaminated with DA can cause acute amnesic shellfish poisoning (ASP) in humans, a malady characterized by seizures, short-term memory loss, and even death.[2,3] In 1987, the first recorded outbreak of ASP occurred on Prince Edward Island, Canada, wherein 107 people contracted the illness from eating mussels containing high levels of DA.[4,5] While oceanic DA production is primarily linked to the harmful algal bloom-forming diatom genus *Pseudo-nitzschia*, the compound was originally discovered in the 1950s by Daigo and co-authors from the red macroalga *Chondria armata*.[6] The chemical constituents of *C. armata*, along with several other seaweeds, were of particular interest due to the historical usage of algae as anthelmintic agents in Japan, with the name 'domoic acid' deriving from the word 'domoi', the Japanese name for *C. armata*.[6,7] Indeed, both DA and the related red algal metabolite kainic acid (KA) display anthelmintic activity, with the latter being used to treat roundworm infections until the 1990s.[8,9] These two compounds are commonly referred to as kainoids, sharing a homologous glutamate and isoprenoid-derived pyrrolidine scaffold.

The evolutionary origin of kainoid biosynthesis in marine algae remains unclear. Production of KA is limited to red algae (Rhodophyta), and while multiple orders within Rhodophyta contain species capable of making KA, distribution of known kainoid producing genera is sparse (Supplementary Information, Table S3.2 & Fig. S3.1). Meanwhile, DA production in red algae is further constrained to just one family, Rhodomelaceae. After the initial discovery of DA in *C. armata,* further chemical studies identified potential kainoid production in

other red algae, including *Alsidium corallinum*,[10,11] *Digenea simplex,*[12] *Amansia glomerata*,[12] *Chondria baileyana*,[13] and *Vidalia obtusiloba*.[12] The ability of diatoms (Bacillariophyta) to produce DA is surprising considering no other taxa has been described to produce the neurotoxin besides Rhodophyta and Bacillariophyta, two significantly divergent clades of algae (Fig. 3.1A). Among diatoms, only the genera *Pseudo-nitzschia* and *Nitzschia* contain species that have been demonstrated to produce DA, although it is possible that other understudied diatoms, such as species belonging to the genus *Amphora,* may also produce the toxin.[14-16] Understanding the distribution and evolutionary history of DA biosynthesis may help identify understudied emerging sources of toxicity in the marine environment.

The genes encoding DA (Fig. 3.1B) biosynthesis (*dab*) and KA biosynthesis (*kab*) were recently elucidated (Fig. 3.1C).[17,18] Both pathways begin with the *N*-prenylation of L-glutamate by either geranyl pyrophosphate (GPP) or dimethylallyl pyrophosphate (DMAPP) to establish the linear precursors to DA and KA, respectively.[19] Subsequent oxidative cyclization to build the characteristic pyrrolidine ring of either molecule is catalyzed by an α-ketoglutarate (αKG) dependent $Fe^{2+}$ oxidase, also known as a kainoid synthase enzyme (DabC/KabC). These two unusual enzymatic transformations are diagnostic for kainoid biosynthesis in algae, forming a core biosynthetic gene cluster and serving as bioinformatic hooks to probe sequencing data for kainoid biosynthetic pathways (Fig. 3.1C). The biosynthesis of DA, unlike that of KA, requires further oxidation to install the carboxylic acid functionality on the monoterpene-derived alkyl side chain via a cytochrome P450-catalyzed (DabD) reaction occurring prior to oxidative cyclization of the pyrrolidine. Additionally, the biosynthesis of DA also involves an uncharacterized isomerization transformation, catalyzed by a yet to be identified enzyme. Discovery of the various kainoid pathways was facilitated by transcriptomic and genomic

sequencing of the DA producing diatom *Pseudo-nitzschia multiseries,*[17] and of the KA producing red algae *Digenea simplex* and *Palmaria palmata*.[18] Notably, the kainoid biosynthetic genes from all organisms sequenced thus far appear to co-cluster within each respective genome. Follow-up sequencing efforts of red algae to date have focused exclusively on organisms producing KA, leaving a gap in our understanding of the enzymology underlying DA biosynthesis in red algae.

To address this fundamental question, we sequenced the genome of *Chondria armata* and identified unique copies of a biosynthetic gene cluster suggestive of DA biosynthesis. Subsequent *in vitro* characterization of the *C. armata* pyrrolidine-forming enzymes further implicates their role in DA biosynthesis, leading to the designation of these sequences as red algal domoic acid (*rad*) biosynthesis genes. Comparative enzyme-substrate assays, syntenic comparisons of gene cluster organization, and phylogenetic analysis provide key insight into the evolution of DA biosynthesis in diatoms and red algae. Our findings suggest a combination of horizontal gene transfer (HGT) and neofunctionalization of native enzymology served to establish neurotoxin production in the divergent algal lineages.

### 3.3.1.3 Results and Discussion

Genome Sequencing and Assembly

We collected *C. armata* from Kyushu Island, Japan and sequenced its genomic DNA with a combination of Oxford Nanopore Technologies (ONT) and Illumina platforms. Genome size was estimated to be 480 Mb using the 17-mer histogram generated by Platanus v1.2.4, repeat content was estimated to be 45.49%, and kmer distributions were consistent with tetraploidy (Supplementary Information, Fig. S3.2).[20] Hybrid assembly using MaSuRCA v3.4.2

yielded a genome with a total size, N50, and %GC of 507 Mb, 643 kb, and 45.34%, respectively (Supplementary Information, Table S3.3).[21] Notably, the longest assembled contig was 3.3 Mb in length. Genome contiguity and completeness was assessed with the eukaryota benchmarking universal single-copy orthologs (BUSCO v4.0.5) database.[22] Due to a lack of rhodophyta-specific datasets, we assessed the ten publicly available red macroalgal genomes with the eukaryota_odb10 BUSCO database.[23-26] We found the *C. armata* assembly to be relatively complete containing 69% of complete, single-copy eukaryotic gene orthologs, where current published red macroalgal genomes contain on average 65% of complete, single-copy eukaryotic gene orthologs (Supplementary Information, Table S3.4 & Fig. S3.3). To further assess overall genome completeness, we mapped ONT reads to the assembled genome using minimap2, where 88% of ONT reads were present in the assembled genome, a clearer indicator of genome completeness.[27] Genome sequences and assembly are deposited in NCBI, BioProjectID PRJNA762367.

Kainoid Cluster Identification and Comparison

To identify kainioid biosynthesis genes in the *C. armata* genome, we used the *dab* and *kab* genes as genetic hooks.[17,18] tBLASTn searches with the highly conserved *N*-prenyltransferase gene initially revealed three chromosomal copies of a kainoid biosynthetic gene cluster. Notably, unlike previously characterized red algal kainoid gene clusters, two of the *C. armata* clusters included a cytochrome P450 (CYP450), which is suggestive of DA biosynthesis. We refer to these clusters as red algal domoic acid (*rad*) biosynthetic gene clusters. The first *rad* cluster (*rad1*, GenBank accession no. OK169902) is contained within a 7.9 kb region, and complete coding regions for all three key DA biosynthetic genes were clearly

identified (Fig. 3.2). The second copy of the *rad* cluster (*rad2*, OK169903) is contained within a shorter 7.5 kb region, and complete coding regions of an *N*-prenyltransferase and kainoid synthase were identified. However, this copy of the *rad* cluster appears to contain a degraded CYP450 with premature stop codons. The third copy of the *rad* cluster (*rad3*, OK169904) is contained within a 5.8 kb region, and completely lacks a CYP450. This cluster is located 271 kb into an assembled contig, 443 kb in length. Further analysis did not reveal any DA CYP450 gene homologs or general CYP450 enzymes within this assembled sequence space (Supplementary Information, Fig. S3.4). Curiously, despite *C. armata* appearing to be tetraploid, only three copies of the *rad* cluster were initially identified. To determine if one copy of the *rad* clusters was a collapsed paralog, the coverage of each *rad* cluster copy was calculated using mapped ONT reads. The mean coverage of *rad1* was almost double (39X) that of *rad2* (21X) and *rad3* (21X) and is indicative of a collapsed paralog, supporting the presence of all four expected genomic copies of the *rad* cluster.

Publicly available DA and KA gene clusters were used to construct a comparative analysis with new *rad* gene clusters from *C. armata* (Fig. 3.2).[28] Despite core kainoid biosynthetic genes being more similar by amino acid percent identity to the KA producing red alga *D. simplex,* a relative of *C. armata* also belonging to the family Rhodomelaceae, all-versus-all global amino acid alignments revealed the overall structure and organization of the *rad* gene cluster to be more like *dab* clusters from diatoms. Aside from a lack of the hypothetical 'DabB' protein found in the *dab* clusters, the *rad* and *dab* clusters surprisingly display nearly identical gene organization and orientation (Fig. 3.2). Amino acid percent identity remains consistent across the *N*-prenyltransferase (*radA*/*dabA*) and kainoid synthase (*radC*/*dabC*) enzymes, at 54%

and 56%, respectively. However, despite conservation of gene location and orientation, the amino acid percent identity between the CYP450s, *radD* and *dabD*, decreased to 24%.

*In vitro* Validation of Biosynthetic Gene Function

Most of the known DA isomers were originally discovered in *C. armata*.[29-31] Although production of isodomoic acid A, B, and C have all been described in *Pseudo-nitzschia*, these isomers are found at substantially lower levels than DA in the diatom species and isolates studied thus far.[14,32] A similar trend has also been observed in isolates of *C. armata*.[33] However, high-resolution liquid chromatography-mass spectrometry (LC-MS) analysis of methanolic algal extracts suggests that isodomoic acid B is an especially abundant DA isomer in our independently collected *C. armata* isolate (Fig. 3.3A) (Supplementary Information, Fig. S3.5). We hypothesized that the differential abundance of DA isomers in *C. armata* extracts is linked to the activities of one or more kainoid synthase enzymes expressed by the seaweed, a hypothesis consistent with our previous observations that the kainoid synthase DabC can generate structural diversity.[17]

Confirmation of kainoid synthase activity in *C. armata* was achieved by expressing full-length RadC1 from the intact *rad1* cluster as a N-terminal hexahistidine (N-$His_6$) soluble construct in *Escherichia coli* (Supplementary Information, Fig. S3.6). After affinity chromatography purification, RadC1 was observed to perform oxidative cyclization on the linear precursor cNGG, similar to the previously described activity for DabC.[17] Overnight RadC1 and DabC reactions both proceed to near-complete substrate consumption when cNGG was added. However, RadC1 makes isodomoic acid B as its major cyclization product, whereas DabC produces isodomoic acid A (Fig. 3.3A). While isodomoic acid B production by RadC1 is

consistent with the relative isomer abundance observed in our *C. armata* extracts, the presence of isodomoic acid A and C in the algae remains to be explained biochemically, although trace quantities of isodomoic acid C are observed from the RadC1 reaction. Initial expression, purification, and enzymatic assay of RadC2, encoded in the *rad2* cluster, did not reveal substantial activity towards cNGG, although exceedingly small quantities of isodomoic acid B is produced. As the *rad2* cluster also contains a degraded copy of the co-clustered RadD CYP450 with several premature STOP codons, we suspect that this second cluster may be undergoing pseudogenization and that the encoded enzymes may not function *in vivo*. The third copy of the kainoid synthase enzyme, RadC3, displays 100% amino acid sequence identity to RadC2 and is assumed to be similarly inactive.

Both RadC1 and DabC generate a similar series of dainic acid molecules from the less-favored NGG substrate, although RadC appears to approach complete substrate turnover whereas DabC does not (Supplementary Information, Fig. S3.7).[17] Because RadC1 is more closely related in amino acid sequence to the KA-forming KabC enzyme than the diatom DabC enzyme, the substrate specificity of KabC from *D. simplex* (DsKabC) was also further investigated. Intriguingly, DsKabC does not appear to catalyze cyclization on either the cNGG or NGG precursor (Fig. 3.3A, Supplementary Information, Fig. S3.7). As demonstrated previously, DsKabC efficiently cyclizes the shorter prekainic acid substrate to KA, and this activity can also be seen in this study for RadC and DabC, although neither of the latter reactions proceed to completion (Supplementary Information, Fig. S3.8).[18] This inability of DsKabC to catalyze cyclization on the longer cNGG and NGG substrates represents a marked difference between the DA isomer-forming RadC and DabC versus the KA-forming KabC. While no obvious single residue can be attributed to these differences in substrate specificity, structural prediction via

AlphaFold reveals all known KabC enzyme models to have an extended loop that pushes further into the active site than in RadC or DabC models (Fig. 3.3B).[34,35] Quantification of active site area and volume solvent accessible surface area using CASTp identified an overall decrease in KabC modeled pocket volume (115 $Å^3$) when compared to RadC (208 $Å^3$) and DabC (262 $Å^3$) models (Fig. 3.3C, Supplementary Information, Fig. S3.9).[36] This follows an expected trend according to each enzyme's observed substrate specificities (Fig. 3.3A).

The glutamate *N*-prenyltransferase enzyme, RadA1, was also expressed as a soluble construct following a seven amino acid N-terminal truncation and fusion to maltose binding protein (MBP) containing an N-$His_6$- tag (Supplementary Information, Fig. S3.10). Initial overnight assays with purified MBP-Δ7-RadA1 suggest both GPP and DMAPP can be accepted as substrates to make NGG and prekainic acid respectively. This observation contrasts the selectivity previously shown by DabA and to a lesser extent KabA.[17,18] Ongoing exhaustive experiments on RadA kinetics and structure will provide additional information regarding potential prenyl-group promiscuity to compare directly with previously described DabA and KabA activity.[19] Nevertheless, the ability of RadA1 and RadC1 to produce DA intermediates and isomers, similar to what was observed previously in diatoms, supports their role in overall biosynthesis of DA in *C. armata*.[17]

Phylogenetic Analysis

Previous efforts have attempted to describe the evolutionary history of the glutamate *N*-prenyltransferase enzyme present in both DA and KA biosynthesis.[19] Despite structural homology to known bacterial terpene cyclases, the enzyme does not clade well with any extant terpene cyclase due to its unique amino acid sequence (Supplementary Information, Fig. S3.11).

To further explore the evolutionary history of the remaining DA and KA biosynthetic genes, maximum likelihood (ML)-phylogenetic trees were constructed for both the αKG-dependent kainoid synthases (DabC/KabC/RadC) and the CYP450s (DabD/RadD). Beginning with RadC, initial BLAST searches revealed few closely related homologs outside of known KA or DA biosynthetic genes in the National Center for Biotechnology Information (NCBI) protein database. The closest homologs were primarily from bacteria, including the well-studied αKG-dependent oxidase deacetoxycephalosporin C synthase (*cefE*, UniProtKB/Swiss-Prot ID: P18548.1) in *Streptomyces clavuligerus*, an enzyme that performs the penicillin to cephalosporin ring expansion. To ensure a broad understanding of kainoid synthase enzyme phylogeny in relation to other αKG-dependent $Fe^{2+}$-containing oxidases, representative sequences of key taxonomic groups were generated from the isopenicillin N synthase-like InterPro family (IPR027443). These representative sequences, in addition to top BLAST hits and other known kainoid synthases, were used to construct a ML-phylogenetic tree. This process of using BLAST hits and representative InterPro family sequences to construct a ML tree was repeated with RadD. Unlike RadC, BLAST searches using RadD revealed several significant alignments with other red algal CYP450s. To ensure a well-populated phylogeny, representative sequences from the cytochrome P450 InterPro family (IPR001128) were generated from the same taxonomic groups as before. These sequences, in addition to representative diatom P450 sequences from publicly available transcriptomics and top BLASTp hits for RadD and DabD, were used to construct a ML-phylogenetic tree.[37]

Based on our phylogenetic analysis, kainoid synthases form their own distinct branch independent of taxonomic origin (Fig. 3.4A), a finding that is consistent with previous phylogenetic studies on the *N*-prenyltransferase enzymes.[19] Notably, the bacterial enzyme cefE is

the only well-characterized close relative to RadC. The monophyletic cladding of both the *N*-prenyltransferase enzymes and kainoid synthase enzymes from these distantly related taxa suggests a possible horizontal gene transfer event. This observation is further supported by high levels of overall sequence homology between the two enzymatic groups. On the other hand, the CYP450 enzymes are independently nested within their respective taxonomic groups, an observation that supports convergent evolution towards installation of the carboxylic acid on the prenyl sidechain (Fig. 3.4B). Based on this taxonomic clustering, the presence of a CYP450 enzyme in DA biosynthesis may be a result of gene duplication and neofunctionalization of extant P450 enzymes, a phenomenon that has been previously described in plants.[38,39] This hypothesis is further strengthened by the presence of additional CYP450s closely related to RadD or DabD within the genome of *C. armata* or *P. multiseries*, respectively (Supplementary Information, Fig. S3.12, Fig. S3.13). The evolutionary history of CYP450 families in diatoms has been the subject of previous research, with diatoms displaying substantial P450 gene family acquisition and loss despite encoding relatively few CYP450 genes compared to land plants.[40] The diatom DabD protein bears the most homology to other CYP450 enzymes found in diatoms, including, *Thalassiosira pseudonana* and *Fragilariopsis cylindrus* (bootstrap >99%), suggestive of an ancient acquisition of this CYP450 family in the diatom lineage. The biochemical function of these related enzymes is unknown, and the specific CYP450 family classification of DabD-like enzymes remains to be determined (Supplementary Information, Fig. S3.12).

#### 3.3.1.4 Conclusions and Discussion

Our discovery of the DA biosynthetic machinery within the genome of the red alga *C. armata*, the first-described DA producer, provides crucial insight into the evolutionary history of

kainoid biosynthesis in diatoms and red algae. The overall sparse distribution of kainoid biosynthetic genes in these two distantly related taxa and conservation of gene synteny in *C. armata* and *Pseudo-nitzschia* suggests a possible HGT event. However, direct HGT of the entire gene cluster from red algae to diatoms, or vice versa, is not likely in modern ecological contexts. Alternatively, many genes in diatoms and other microalgae were indeed acquired from red algae in distant evolutionary history through endosymbiotic gene transfer (EGT). Endosymbiosis of unicellular red algae played a major role in the evolution of the diatom clade, and several red algal genes were transferred to the host nuclear genome via EGT to contribute to the diverse origins of diatom genomes.[41,42] However, an ancient EGT event moving the DA biosynthetic gene cluster from red algae to diatoms is unlikely given the limited distribution of kainoid biosynthetic enzymes in modern diatom genera, whereas genes acquired via EGT are typically widespread in the diatom lineage. Nevertheless, the lack of clear homology of the glutamate *N*-prenyltransferase (RadA, KabA, and DabA) and kainoid synthase (RadC, KabC, and DabC) enzymes to known algal proteins suggests HGT from a potential third host as a mechanism for the acquisition of these kainoid biosynthetic enzymes. Furthermore, the observation that the enzymes RadD and DabD are most like native CYP450s within the red algal and diatom genomes, respectively, supports a history invoking two modes of evolution towards DA biosynthesis: HGT combined with gene duplication and neofunctionalization of native biochemistry.

From our biochemically supported phylogenetic analysis, it appears that the glutamate *N*-prenyltransferase and kainoid synthase genes form a core biosynthetic gene cluster and were acquired via HGT from an unknown source. These specific chemical transformations are unprecedented in biochemistry outside of diatom and red algal kainoid biosynthesis and do not

seem to be repurposed enzymology from within the respective algal lineages, as evidenced by their low similarity to terpene cyclases and αKG-dependent $Fe^{2+}$ oxidases from public algal genomes and transcriptomes. These enzymes may be bacterial in origin due to the structural homology of the *N*-prenyltransferase to bacterial terpene cyclases and the modest amino acid sequence homology of the kainoid synthase to known bacterial enzymes, although no strong candidates for bacterial kainoid biosynthetic enzymes have been found to date.[19] Notably, HGT from bacteria is well-described in diatoms and red algae. Diatoms especially have been the subject of substantial genome sequencing efforts, revealing that approximately 5% of their functional genes have been acquired from bacteria in relatively recent evolutionary history.[43,44] Acquisition of HGT-derived bacterial genes, such as urea cycle enzymes and proteorhodopsins, has contributed to the unique biochemistry and subsequent ecological success of the diatom lineage.[45,46] Similarly, HGT is an important mechanism for the expansion of genes related to both primary and secondary metabolism in red algae. The extremophilic red alga *Galdieria sulphuraria* is estimated to have acquired at least 5% of protein-coding genes through HGT.[47] These HGT-derived genes typically confer an adaptive advantage to the organism, such as in the genus *Pyropia* wherein a HGT-derived carbonic anhydrase (CA) gene acts as a key component of the carbon-concentrating mechanism and aids in overcoming carbon limitations.[48] Red algae also seem to have acquired other terpene cyclases via HGT from bacteria, although these enzymes appear unrelated to the *N*-prenyltransferase enzymes in DA and KA biosynthesis.[49]

Following acquisition of the core kainoid biosynthetic genes, both *C. armata* and *Pseudo-nitzschia* spp. appear to have repurposed extant CYP450 enzymology to install the characteristic carboxylic acid on the terpene-derived DA sidechain important for toxicity.[50] Enzymes within the CYP450 family catalyze a diversity of chemical transformations, and conversion of a methyl

group to a carboxylic acid is a well-established biotransformation with precedence in diverse eukaryotic and prokaryotic organisms.[51] Although the activity of RadD1 remains to be confirmed *in vitro*, the high degree of synteny between the *dab* and *rad1* clusters suggests that the red algal RadD1 fulfills a similar role to DabD in diatoms. The observed activity towards the cNGG substrate in the RadC1 and DabC systems also supports the hypothesis that a common sequence of DA biosynthetic reactions is shared between diatoms and red algae, with CYP450-catalyzed triple oxidation to a carboxylic acid occurring upstream of oxidative cyclization in both organisms.[17] Independent neofunctionalization of CYP450 activities has been previously implicated in the evolution of the plant hormone diosgenin, a plant steroid saponin with a distinctive CYP450-installed spiroketal motif that is not limited to a specific taxon of plants but is instead broadly and sparsely distributed.[38] An additional example of CYP450-driven convergent evolution is seen in the furanocoumarin pathway, also in plants.[39] The independent selection for these enzyme activities in unrelated plants provides precedence for the neofunctionalization of CYP450s to install similar chemical features in distantly related organisms, a phenomenon that may have occurred in the evolution towards DA biosynthesis as reflected by the red algal-derived RadD and diatom-derived DabD.

Evolution towards DA production is further supported by the apparent substrate scope of the different kainoid synthases. Although RadC1 is more like the red algal KabC in its amino acid sequence, RadC1 displays an activity more similar to the diatom enzyme DabC in its ability to convert both NGG and cNGG to DA-like molecules. These substrates are not similarly cyclized by the KA-forming DsKabC. Despite the similar activities exhibited by the *C. armata* and *Pseudo-nitzschia* kainoid synthases, the difference in DA isomers generated by RadC1 (isodomoic acid B) and DabC (isodomoic acid A) raises additional biosynthetic questions.

Notably, the 1,3-olefin isomerization step required to make DA itself continues to remain elusive.[17,29] Because isodomoic acid B would require both a 1,3-isomerization and a separate *trans*-to-*cis* isomerization of the second olefin closest to the kainoid ring, it is worth reconsidering whether the key 1,3-isomerization occurs before or after kainoid synthase cyclization to create the most abundant, canonical DA isomer in diatoms and red algae. A similar 1,3-isomerization step occurs immediately following triple hydroxylation of a methyl group to a carboxylic acid in lysergic acid biosynthesis, a process thought to be carried out by one or more CYP450 enzymes in the ergot fungus *Claviceps purpurea*.[52] However, such an olefin isomerization was not revealed by previous studies on DabD expressed in *Saccharomyces cerevisiae*.[17] Meanwhile, hypothetical protein DabB identified from the diatom biosynthetic cluster is missing from the respective *C. armata* gene cluster and does not appear to be encoded within the sequenced *C. armata* genome. In any case, the responsible isomerase remains unidentified. Ongoing investigations of candidate genes present in these two distantly related species, including further study of the activities exhibited by the co-clustered CYP450 enzymes RadD and DabD, will shed additional light on this remaining piece of the DA biosynthetic puzzle.

The apparent evolution of kainoid synthase substrate selectivity in combination with evidence for CYP450 neofunctionalization and co-clustering suggests a potential advantage afforded by DA biosynthesis over KA biosynthesis. While DA is chemically distinct from KA and has enhanced bioactivity against ionotropic glutamate receptors, the true ecological function of DA has remained elusive.[53] Previous studies have demonstrated limited trace metal chelation by DA, affording a potential selective advantage for diatoms thriving in iron limited coastal regimes.[54,55] Other work has suggested a role for DA in grazer defense, particularly with respect

to copepod grazing in the diatom system.[56] However, no studies to date have provided a concrete ecological mechanism beyond reasonable doubt in both diatoms and red algae. While DA could conceivably serve different functions in the red alga *C. armata* and *Pseudo-nitzschia* diatoms, the evolutionary history described here provides new insight to the ecological importance of this marine biotoxin.

### 3.3.1.5 Main Text Figures

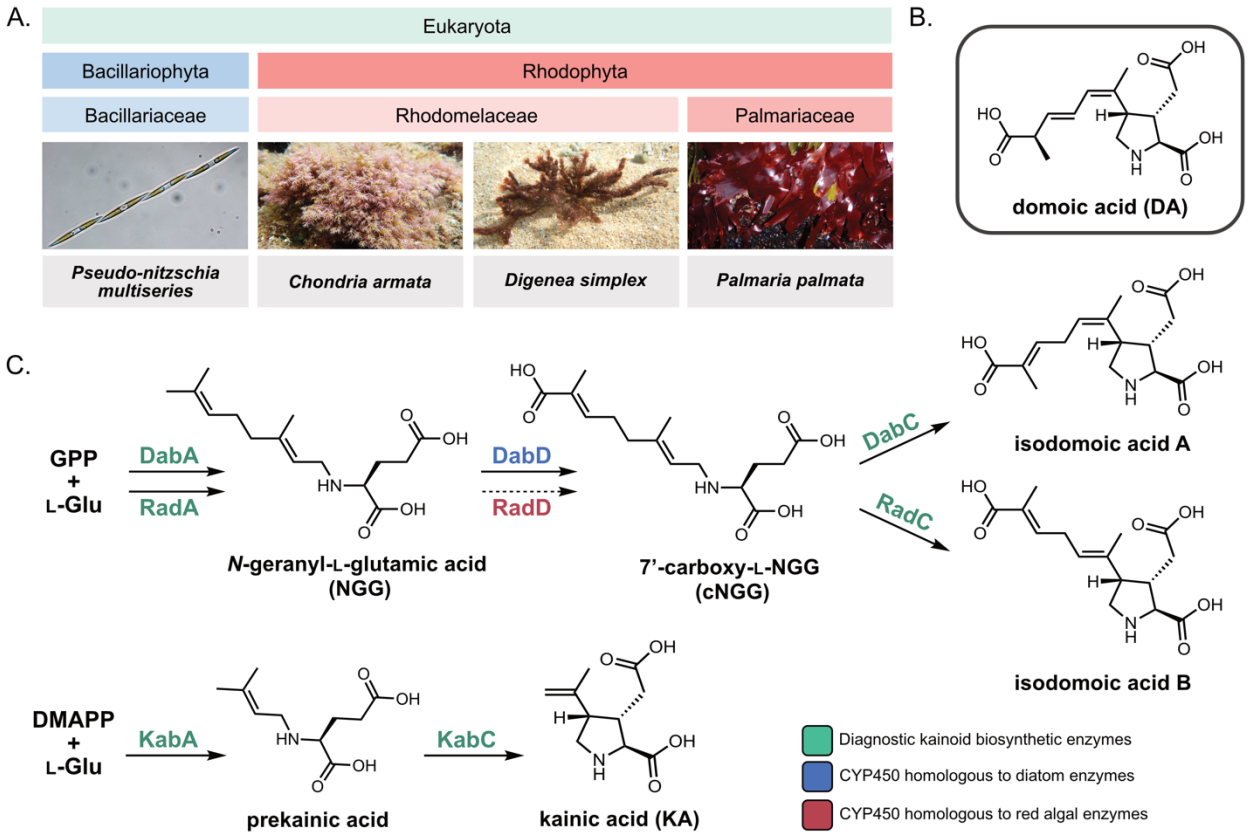



**Figure 3.1.** Kainoid biosynthetic genes and producing organisms. (*A*) DA producing diatom *P. multiseries* (Bacillariophyta) and KA producing red algae *D. simplex* and *P. palmata* with DA producing red alga *C. armata* (Rhodophyta), two significantly divergent phyla of kainoid producing algae. (*B*) Structure of domoic acid (DA). (*C*) Proposed DA and KA biosynthetic pathways showing diagnostic kainoid biosynthetic enzymes in teal performing the two key enzymatic transformations forming the core kainoid pyrrolidine scaffold. Verified activities and major products from this work and previous studies shown with solid arrows.[17,18]

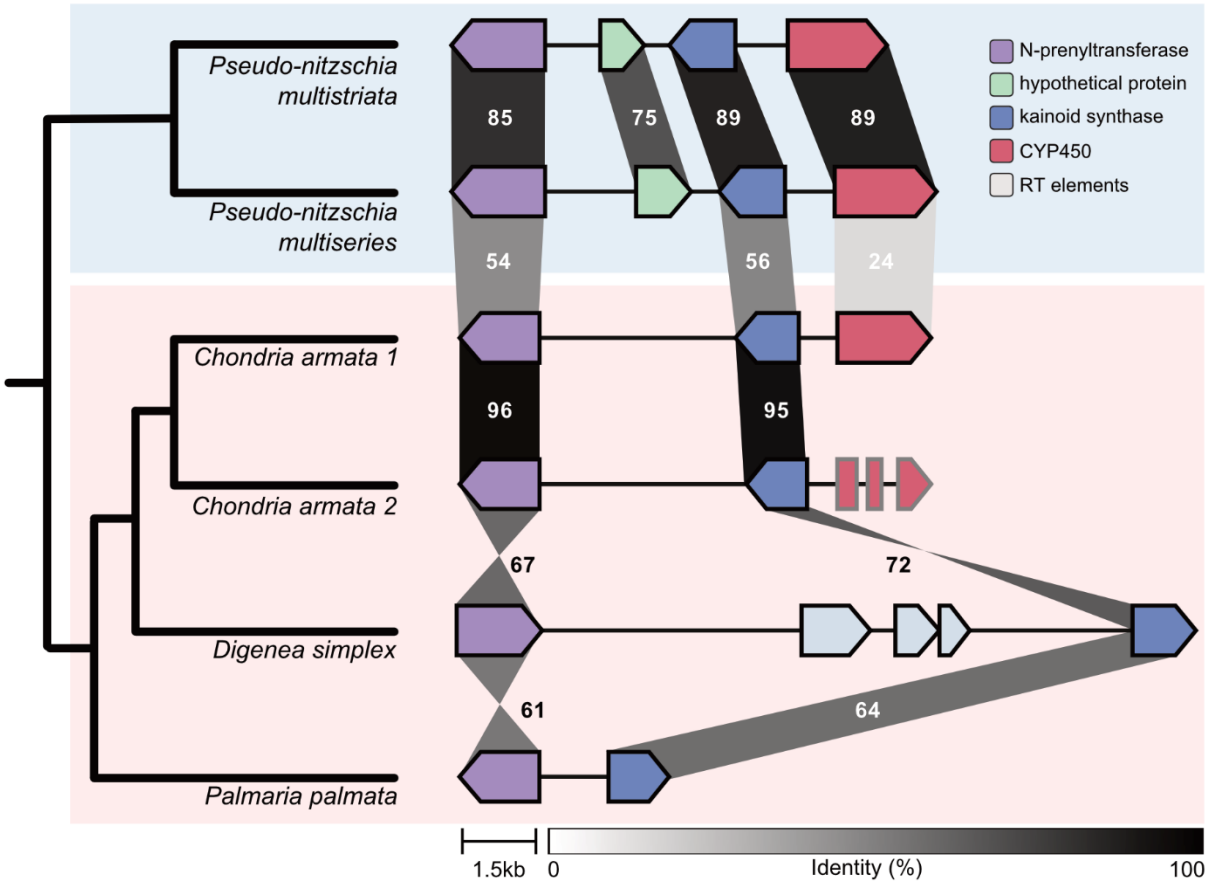



**Figure 3.2.** Visualization of syntenic comparisons between *dab*, *rad*, and *kab* gene clusters. Blue background highlights diatom sequences, and red background highlights red algal sequences. Cladogram connects clusters based on taxonomic relationships of organisms, not to scale with respect to evolutionary time. Intensity of shading between genes is relative to the similarity of gene pairs by amino acid sequence percent identity, in agreement with discrete measures of sequence percent identity as labeled. Retro transposable elements are abbreviated as RT elements.

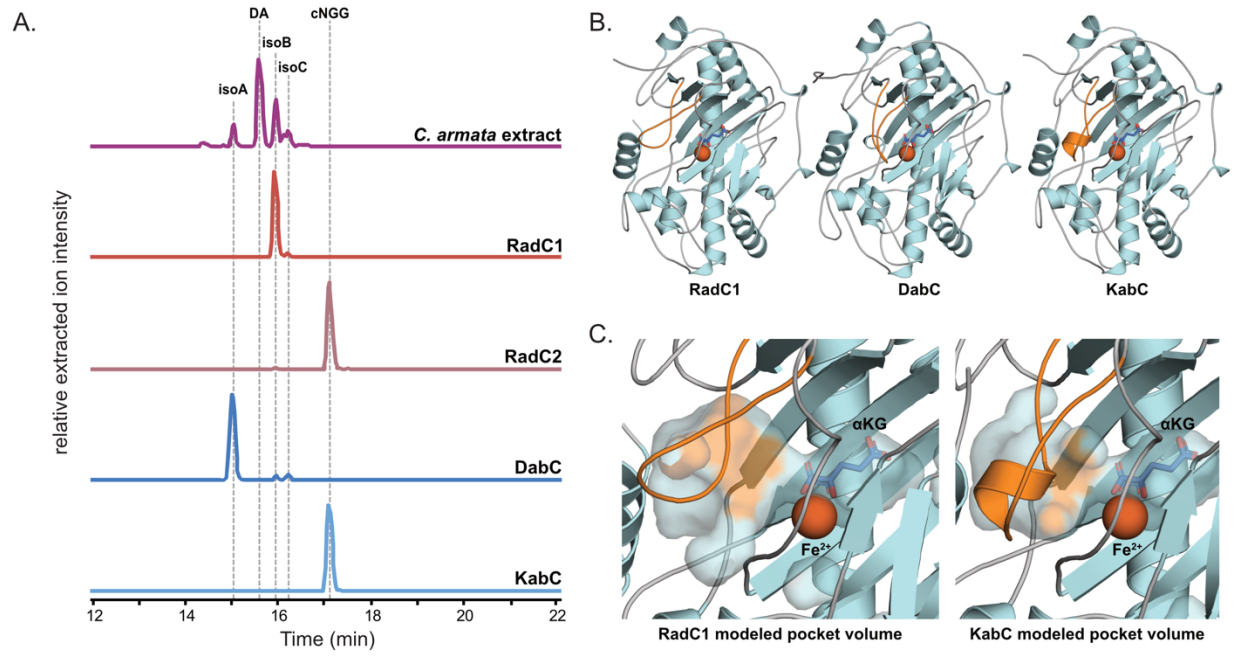



**Figure 3.3.** Activities and predicted structures of the known kainoid synthase enzymes. (*A*) Comparison of combined extracted ion chromatogram (EIC) profiles for cNGG substrate (312.1453 +/- 0.2 m/z) and isodomoic acid (310.1293 +/- 0.2 m/z) from RadC1, RadC2, DabC, and KabC assays and *C. armata* extract. Relative intensity of extracted ions is shown, and all substrates and observed DA isomers are compared to purified standards (Supplementary Information, Fig. S3.5). (*B*) AlphaFold2 predicted models for RadC1, DabC and KabC. (*C*) Modeled pocket volume for RadC1 and KabC.

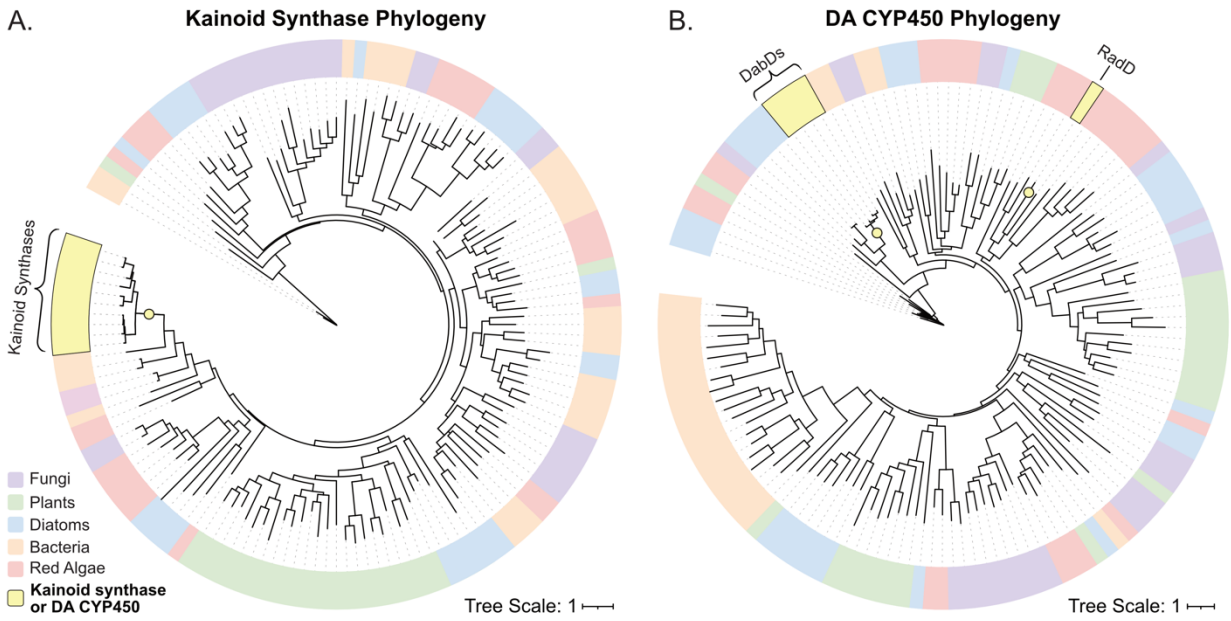



**Figure 3.4.** Phylogenetic analysis of (*A*) kainoid synthase and (*B*) co-clustered DA CYP450 enzymes (yellow highlights, yellow circle denotes key branch points). The maximum likelihood trees were built using representative sequences from key taxonomic groups (Supplementary Information, Table S3.5 and Table S3.6). Kainoid synthase enzymes form their own distinct branch, independent of taxonomic origin, while the DA CYP450s are independently nested within their respective taxonomic groups.

### 3.3.1.6 Works Cited to the Main Text


1. V. L. Trainer, S. S. Bates, N. Lundholm, A. E. Thessen, W. P. Cochlan, N. G. Adams, C. Trick. *Pseudo-nitzschia* physiological ecology, phylogeny, toxicity, monitoring and impacts on ecosystem health. *Harmful Algae* **14**, 271–300 (2012)

2. T. M. Perl, L. Bédard, T. Kosatsky, J. C. Hockin, E. C. Todd, L. A. McNutt, R. S. Remis. Amnesic shellfish poisoning: a new clinical syndrome due to domoic acid. *Can. Dis. Wkly. Rep. Rapp. Hebd. Mal. Au Can.* **16 Suppl 1E**, 7–8 (1990)

3. O. M. Pulido, Phycotoxins by Harmful Algal Blooms (HABs) and human poisoning: an overview. *Int. Clin. Pathol. J.* **2** (2016)

4. J. L. C. Wright, R. K. Boyd, A. D. Freitas, M. Falk, R. A. Foxall, W. D. Jamieson, M. V. Laycock, A. W. McCulloch, A. G. McInnes, P. Odense, V. P. Pathak, M. A. Quilliam, M. A. Ragan, P. G. Sim, P. Thibault, J. A. Walter, M. Gilgan, D. J. A. Richard, D. Dewar. Identification of domoic acid, a neuroexcitatory amino acid, in toxic mussels from eastern Prince Edward Island. *Can. J. Chem.* **67**, 481–490 (1989)

5. S. S. Bates, C. J. Bird, A. D. Freitas, R. Foxall, M. Gilgan, L. A. Hanic LA, G. R. Johnson, A. W. McCulloch, P. Odense, R. Pocklington, M. A. Quilliam. Pennate diatom *Nitzschia pungens* as the primary source of domoic acid, a toxin in shellfish from eastern Prince Edward Island, Canada. *Can. J. Fish. Aquat. Sci.* **46**, 1203–1215 (1989)

6. K. Daigo, Studies on the constituents of *Chondria armata*. II: Isolation of an anthelmintical constituent. *Yakugaku Zasshi* **79**, 353–356 (1959)

7. W. H. Gerwick, Plant sources of drugs and chemicals. *Encyclopedia of Biodiversity* **2**, 129-139 (2013)

8. Y. Komiya, A. Kobayashi, Techniques applied in Japan for the control of *Ascaris* and hookworm infections -- a review. *Jpn. J. Med. Sci. Biol.* **18**, 1–17 (1965)

9. J. F. Tremblay, Shortage of kainic acid hampers neuroscience research. *Chem. Eng. News Arch.* **78**, 14–15 (2000)

10. G. Impellizzeri, S. Mangiafico, G. Oriente, M. Piattelli, S. Sciuto, E. Fattorusso, S. Magno, C. Santacroce, D. Sica. Amino acids and low-molecular-weight carbohydrates of some marine red algae. *Phytochemistry* **14**, 1549–1557 (1975)

11. Y. Shimizu, S. Gupta, K. Masuda, L. Maranda, C. K. Walker, R. Wang. Dinoflagellate and other microalgal toxins: chemistry and biochemistry. *Pure Appl. Chem.* **61**, 513–516 (1989)

12. M. Sato, T. Nakano, M. Takeuchi, N. Kanno, E. Nagahisa, Y. Sato. Distribution of neuroexcitatory amino acids in marine algae. *Phytochemistry* **42(6)**, 1595–1597 (1996)

13. M. V. Laycock, A. S. W. de Freitas, J. L. C. Wright, Glutamate agonists from marine algae. *J. Appl. Phycol.* **1**, 113–122 (1989)

14. S. S. Bates, K. A. Hubbard, N. Lundholm, M. Montresor, C. P. Leaw, *Pseudo-nitzschia*, *Nitzschia*, and domoic acid: new research since 2011. *Harmful Algae* **79**, 3–43 (2018)

15. D. G. Mann, R. Trobajo, S. Sato, C. Li, A. Witkowski, F. Rimet, M. P. Ashworth, R. M. Hollands, E. C. Theriot. Ripe for reassessment: a synthesis of available molecular data for the speciose diatom family Bacillariaceae. *Mol. Phylogenet. Evol.* **158**, 106985 (2021)

16. B. C. Dhar, L. Cimarelli, K. S. Singh, L. Brandi, A. Brandi, C. Puccinelli, S. Marcheggiani, R. Spurio, Molecular detection of a potentially toxic diatom species. *Int. J. Environ. Res. Public. Health* **12**, 4921–4941 (2015).

17. J. K. Brunson, S. M. K. McKinnie, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. Smith, D. A. Hutchins, A. E. Allen, B. S. Moore. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018)

18. J. R. Chekan, S. M. K. McKinnie, M. L. Moore, S. G. Poplawski, T. P. Michael, B. S. Moore. Scalable biosynthesis of the seaweed neurochemical, kainic acid. *Angew. Chem. Int. Ed.* **58**, 8454–8457 (2019)

19. J. R. Chekan, S. M. K. McKinnie, J. P. Noel, B. S. Moore, Algal neurotoxin biosynthesis repurposes the terpene cyclase structural fold into an *N*-prenyltransferase. *Proc. Natl. Acad. Sci.* **117**, 12799–12805 (2020)

20. R. Kajitani, D. Yoshimura, M. Okuno, Y. Minakuchi, H. Kagoshima, A. Fujiyama, K. Kubokawa, Y. Kohara, A. Toyoda, T. Itoh. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nat. Commun.* **10** (2019).

21. A. V. Zimin, D. Puiu, M. C. Luo, T. Zhu, S. Koren, G. Marçais, J. A. Yorke, J. Dvořák, S. L. Salzberg. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res*. **5**, 787-792 (2017)

22. M. Seppey, M. Manni, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol. Clifton NJ* **1962**, 227–245 (2019)

23. J. Collén, B. Porcel, W. Carré, S. G. Ball, C. Chaparro, T. Tonon, T. Barbeyron, G. Michel, B. Noel, K. Valentin, M. Elias. Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci.* **110**, 5247–5252 (2013)

24. J. Lee, E. C. Yang, L. Graf, J. H. Yang, H. Qiu, U. Zelzion, C. X. Chan, T. G. Stephens, A. P. Weber, G. H. Boo, S. M. Boo. Analysis of the draft genome of the red seaweed *Gracilariopsis chorda* provides insights into genome size evolution in Rhodophyta. *Mol. Biol. Evol.* **35**, 1869–1886 (2018)

25. S. Jia, G. Wang, G. Liu, J. Qu, B. Zhao, X. Jin, L. Zhang, J. Yin, C. Liu, G. Shan, S. Wu. High-quality de novo genome assembly of *Kappaphycus alvarezii* based on both PacBio and HiSeq sequencing. *bioRxiv*, 2020.02.15.950402 (2020)

26. S. H. Brawley, N. A. Blouin, E. Ficko-Blean, G. L. Wheeler, M. Lohr, H. V. Goodson, J. W. Jenkins, C. E. Blaby-Haas, K. E. Helliwell, C. X. Chan, T. N. Marriage. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc. Natl. Acad. Sci.* **114**, E6361–E6370 (2017)

27. H. Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018)

28. C. L. M. Gilchrist, Y-H. Chooi, Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473-2475 (2020)

29. J. Clayden, B. Read, K. R. Hebditch, Chemistry of domoic acid, isodomoic acids, and their analogues. *Tetrahedron* **61**, 5713–5724 (2005)

30. M. Meda, T. Kodama, T. Tanaka, H. Yoshizumi, T. Takemoto, K. Nomoto, T. Fujita. Structures of isodomoic acids A, B and C, novel insecticidal amino acids from the red alga *Chondria armata*. *Chem. Pharm. Bull.* **34**, 4892–4895 (1986)

31. L. Zaman, O. Arakawa, A. Shimosu, Y. Onoue, S. Nishio, Y. Shida, T. Noguchi. Two new isomers of domoic acid from a red alga, *Chondria armata*. *Toxicon* **35**, 205–212 (1997)

32. A. Lelong, H. Hégaret, P. Soudant, S. S. Bates, *Pseudo-nitzschia* (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. *Phycologia* **51**, 168–216 (2012)

33. Y. Maeno, Y. Kotaki, R. Terada, Y. Cho, K. Konoki, M. Yotsu-Yamashita. Six domoic acid related compounds from the red alga, *Chondria armata*, and domoic acid biosynthesis by the diatom, *Pseudo-nitzschia multiseries*. *Sci. Rep.* **8**, 356 (2018)

34. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021)

35. M. Mirdita, S. Ovchinnikov, M. Steinegger, ColabFold - Making protein folding accessible to all. *bioRxiv*, 2021.08.15.456425 (2021)

36. W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018)

37. P. J. Keeling, F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. B. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone, A. Z. Worden. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol.* **12**, e1001889 (2014)

38. B. Christ, C. Xu, M. Xu, F. S. Li, N. Wada, A. J. Mitchell, X. L. Han, M. L. Wen, M. Fujita, J. K. Weng. Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat. Commun.* **10**, 3206 (2019)

39. C. Villard, R. Munakata, S. Kitajima, R. Van Velzen, M. E. Schranz, R. Larbat, A. Hehn. A new P450 involved in the furanocoumarin pathway underlies a recent case of convergent evolution. *New Phytol.* **231,** 1923-1939 (2021)

40. L. Teng, X. Fan, D. R. Nelson, W. Han, X. Zhang, D. Xu, H. Renault, G. V. Markov, N. Ye. Diversity and evolution of cytochromes P450 in stramenopiles. *Planta* **249**, 647–661 (2019)

41. H. Qiu, H. S. Yoon, D. Bhattacharya, Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Front. Plant Sci.* **4**, 366 (2013)

42. D. Moreira, P. Deschamps, What was the real contribution of endosymbionts to the eukaryotic nucleus? Insights from photosynthetic eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6**, a016014 (2014)

43. E. Vancaester, T. Depuydt, C. M. Osuna-Cruz, K. Vandepoele, Comprehensive and functional analysis of horizontal gene transfer events in diatoms. *Mol. Biol. Evol.* **37**, 3243–3257 (2020)

44. C. Bowler, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J-P Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C.

Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin–Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M-P. O. Secq, C. Napoli, M. Obornik, M. S. Parker, J-L Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, I. V. Grigoriev. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456(7219)**, 239–244 (2008)

45. A. E. Allen, C. L. Dupont, M. Oborník, A. Horák, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie, C. Bowler. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**, 203–207 (2011)

46. A. Marchetti, D. Catlett, B. M. Hopkinson, K. Ellis, N. Cassar, Marine diatom proteorhodopsins and their potential role in coping with low iron availability. *ISME J.* **9**, 2745–2748 (2015)

47. G. Schönknecht, W. H. Chen, C. M. Ternes, G. G. Barbier, R. P. Shrestha, M. Stanke, A. Bräutigam, B. J. Baker, J. F. Banfield, R. M. Garavito, K. Carr. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**, 1207–1210 (2013)

48. D. Wang, X. Yu, K. Xu, G. Bi, M. Cao, E. Zelzion, C. Fu, P. Sun, Y. Liu, F. Kong, G. Du, X. Tang, R. Yang, J. Wang, L. Tang, L. Wang, Y. Zhao, Y. Ge, Y. Zhuang, Z. Mo, Y. Chen, T. Gao, X. Guan, R. Chen, W. Qu, B. Sun, D. Bhattacharya, Y. Mao. *Pyropia yezoensis* genome reveals diverse mechanisms of carbon acquisition in the intertidal environment. *Nat. Commun.* **11**, 4028 (2020)

49. G. Wei, Q. Jia, X. Chen, T. G. Köllner, D. Bhattacharya, G. K-S. Wong, J. Gershenzon, F. Chen. Terpene biosynthesis in red algae is catalyzed by microbial type but not typical plant terpene synthases. *Plant Physiol.* **179**, 382–390 (2019)

50. Y. Maeno, Y. Kotaki, R. Terada, M. Hidaka, Y. Cho, K. Konoki, M. Yotsu-Yamashita. *Org. Biomol. Chem.* (2021)

51. X. Zhang, S. Li, Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* **34**, 1061–1089 (2017)

52. T. Haarmann, I. Ortel, P. Tudzynski, U. Keller, Identification of the cytochrome P450 monooxygenase that bridges the clavine and ergoline alkaloid pathways. *ChemBioChem* **7**, 645–652 (2006)

53. J. A. Larm, P. M. Beart, N. S. Cheung, Neurotoxin domoic acid produces cytotoxicity via kainate- and ampa-sensitive receptors in cultured cortical neurones. *Neurochem. Int.* **31**, 677–682 (1997)

54. C. G. Trick, B. D. Bill, W. P. Cochlan, M. L. Wells, V. L. Trainer, L. D. Pickell. Iron enrichment stimulates toxic diatom production in high-nitrate, low-chlorophyll areas. *Proc. Natl. Acad. Sci.* **107**, 5887–5892 (2010)

55. E. Rue, K. Bruland, Domoic acid binds iron and copper: a possible role for the toxin produced by the marine diatom Pseudo-nitzschia. *Mar. Chem.* **76**, 127–134 (2001)

56. S. Harðardóttir, S. Wohlrab, D. M. Hjort, B. Krock, T. G. Nielsen, U. John, N. Lundholm. Transcriptomic responses to grazing reveal the metabolic pathway leading to the biosynthesis of domoic acid and highlight different defense strategies in diatoms. *BMC Mol. Biol.* **20**, 7 (2019)

### 3.3.2 Supplementary Information including Materials and Methods

#### 3.3.2.1 Methods: General Methods/Chemical Methods

General

All chemicals and solvents were used as received from the commercial supplier (SigmaAldrich or Fisher). Domoic acid and kainic acid were purchased from commercial suppliers (National Research Council of Canada or Chem-Impex, respectively). All protein purification from *Escherichia coli* was performed on an ÄKTA Pure 25 L1 instrument (Cytiva) with a fraction collector F9-C and sample pump S9 with all solvents filtered through a nylon 0.2 µm GDWP membrane (Merck) prior to use. FPLC data was analyzed with UNICORN version 7 software. All protein quantification was done by method of Bradford using the Protein Assay Dye Reagent Concentrate (Bio-Rad) on protein sample dilutions in MilliQ water.

High resolution liquid chromatography mass spectrometry (HRMS) measurements were carried out on an Agilent Technologies 1200 Series system with a diode-array detector coupled to an Agilent Technologies 6530 accurate-mass Q-TOF LCMS run in negative ionization mode. Compounds were separated by reversed-phase chromatography on a Phenomenex Kinetex 5 mm C18 100 Å 150 x 4.6 mm LC column with water + 0.1% formic acid (solvent A) and acetonitrile + 0.1% formic acid (solvent B) as eluents. Two LC methods were used for separation of all compounds for HRMS. For LC Method A, the following gradient was applied at a flow rate of 0.75 mL/min: hold at 5% B for 1 minute, 5% to 35% B over 30 min, 35 to 100% B over 1 minute, hold at 100% B for 1.5 min, 100% to 5% B over 2.5 min, hold at 5% B for 2 min. For LC Method B, the following gradient was applied at a flow rate of 0.75 mL/min: hold at 5% B

for 1 minute, 5% to 35% B over 15 min, 35 to 100% B over 1 minute, hold at 100% B for 1.5 min, 100% to 5% B over 2.5 min, hold at 5% B for 2 min.

Preparation of substrates and standards

Preparation of all non-commercial substrates and standards has been described elsewhere in the chemical literature.[1-3] Briefly, the chemical syntheses of *N*-geranyl glutamate (NGG), 7'-carboxy-*N*-gernayl glutamate (cNGG), prekainic acid, dimethyl-allyl pyrophosphate (DMAPP), and geranyl pyrophosphate (GPP) were performed as previously described.[1,2] Enzyme biocatalytic synthesis of dainic acid A, dainic acid C, and isodomoic acid A was performed as previously described.[1] Purification of isodomoic acid B from *Chondria armata* was performed as previously described.[3] Isodomoic acid C was co-purified with isodomoic acid B using the previously described methodology for isodomoic acid B. The two compounds were separated using a Mightysil RP-18 GP Aqua column (4.6 × 250 mm, 5 μm, Kanto Chemical, Tokyo, Japan) with $H_2O$–MeCN–HCOOH (90 : 10 : 0.1, v/v/v). Proton nuclear magnetic resonance ($^1$H-NMR) measurement of isodomoic acid C was performed in $D_2O$ at 600 MHz (Fig. S3.14). The signal of the residual MeOD was adjusted at δ 3.30 ppm as the internal reference in $D_2O$. The identity of isodomoic acid B was similarly confirmed by $^1$H-NMR by comparison to literature spectra.[3,4]

### 3.3.2.2 Molecular Biology/Biochemical Methods

DNA Extraction

*Chondria armata* was collected from Hanazezaki (31˚11'40" N 130˚30'30" E), Ibusuki City, Kagoshima Prefecture, Kyushu Island, Japan. The sample was dried by lyophilization and

finely ground. 1 g of this *C. armata* powder was used to extract DNA according to the previously described protocol.[2] Extracted DNA was further purified using a Short Read Eliminator Kit, XSRE (Circulomics) and size selected using the BluePippin system with a High Pass Plus 15 kb cassette (Sage Science).

Oxford Nanopore Sequencing

Size-selected, high-molecular weight (HMW) DNA was sequenced using the Oxford Nanopore MinION and PromethION platforms (Oxford Nanopore Technologies (ONT), Oxford, UK). Two one-dimensional (1d) libraries were prepared with 900 ng HMW DNA using the Ligation Sequencing Kit (SQK-LSK109) (ONT, Oxford, UK). An initial library was loaded onto a R9.4 flowcell and run for 48 hr, resulting in 6.9 Gb of sequence with a read length N50 of 15 kb. A second library was generated using the same method as before and loaded onto a R9.4.1 flowcell and run using the PromethION platform 24 (P24), yielding 46 Gb of sequence with a read length N50 of 12.5 kb. The data from these two runs were pooled and used for assembly. NCBI Sequence Read Archive accession: SRR15927349 and SRR15927348.

Illumina Sequencing

NEBnext sequencing libraries were generated to polish genome assemblies (New England Biolabs, Beverly, MA, USA). NEBnext sequencing libraries were created with 100 ng of DNA and quality-controlled on a bioanalyzer. Resulting libraries were sequenced on an Illumina MiSeq 2x150 bp to check quality and quantity (Illumina, San Diego, CA). The libraries were then sequenced on an Illumina NovaSeq S4 200 (PE100, 10x) run that resulted in 74 Gb of sequence. NCBI Sequence Read Archive accession: SRR15927350.

Genome assembly

Illumina reads were trimmed with Trimmomatic v0.3 and used for genome size estimation and initial assembly by Platanus v1.2.4 using the following parameters: -k 17 -s 10 -u 0.2 -t 24 -m 500.[5,6] Using the 17-mer histogram, *C. armata* genome size was estimated to be 480 Mb (Figure S3.2).

Raw pooled ONT and illumina reads were assembled using the MaSuRCA pipeline v3.4.2.[7] Configuration file parameter FLYE_ASSEMBLY=1 was used for final assembly of corrected mega-reads. All other parameters were set to default settings.

Genome quality assessment

*C. armata* genome assembly contiguity and completeness were assessed with BUSCO (v4.0.5) using the eukaryota benchmarking universal single-copy orthologs eukaryota_odb10 dataset.[8] This method was applied to the ten publicly available red macroalgal genomes – *Agarophyton vermiculophyllum* (AgarVerm_1.0), *Asparagopsis taxiformis* (ASM1839795v1), *Chondrus crispus* (ASM35022v2), *Digenea simplex* (ASM479842v1), *Gracilariopsis chorda* (GraCho1.0), *Gracilariopsis lemaneiformis* (Glem_v01), *Kappaphycus alvarezii* (ASM220596v3), *Neoporphyra haitanensis* (ASM982973v1), Neopyropia yezoensis (ASM982973v1) and *Porphyra umbilicalis* (P_umbilicalis_v1) – to establish a relative comparison of *C. armata* eukaryotic gene orthologs. Raw ONT reads were mapped to the final assembled genome using minimap2.[9] Mapped reads were processed, sorted, and indexed using SAMtools with the sorted bam file reporting 88.32% of reads mapped to the assembled

genome.[10] The average depth of read coverage for *rad1, rad2*, and *rad3* was calculated to be 39x, 21x, and 21x, respectively (Fig. S3.4).

PCR and cloning

Polymerase chain reaction (PCR) was carried out using standard thermocycling protocols using Prime Star MAX (TaKaRa) using the templates and primers as specified for each reaction. All reactions using plasmids as template were treated with DpnI (New England Biolabs) following PCR amplification to remove contaminating template DNA. Assembly of all PCR fragments was performed using the NEBuilder HiFi DNA Assembly Mix (New England Biosciences) following manufacturer protocols. Initial amplification of *radA1*, *radC1*, and *radC2* was performed using the primer sets RadA_p28_N_001-002 and RadC_p28_N_001-002 with *C. armata* gDNA as template. The pET28a vector was linearized using the primer set p28-N_BB_001-002. Assembly of *radA* and *radC* amplicons into the linearized pET28a vector incorporated the N-terminal $His_6$ affinity tag, whereupon the assembled constructs were transformed into *E. coli* DH5a chemically competent cells (ThermoFisher) and plated on kanamycin (50 mg/mL) LB plates. Purification of plasmid DNA and subsequent Sanger sequencing (GeneWiz) confirmed the presence of *radA1*, *radC1*, and *radC2* inserts. All constructs were transformed into chemically competent *E. coli* BL21(DE3) (New England Biolabs).

Following expression testing of constructs, further truncations and tags were screened for *radA1* and *radC2* to yield soluble protein expression. A 7-amino acid N-terminal truncation was generated for *radA1* using the primers RadA_d7_MBP and p28_3prime_Amp using the construct generated above as a template. A modified pET28 vector containing an N-terminal

maltose binding protein (MBP) with N-terminal $His_6$ tag was linearized using the primer set MBP_BB_001-002.[11] Assembly of the truncated *radA1* into the linearized MBP vector yielded the construct N-$His_6$-MBP-Δ7-RadA1, which was transformed and purified, as described above, and verified by Sanger sequencing.

An 11-amino acid N-terminal truncation was generated for *radC2* using the primers RadC2_D11_p28-C and RadC2_CtermOH with the original RadC2 construct generated above as template. The pET28a vector was linearized using the primer set p28-C_BB_001-002 to allow the addition of a C-terminal $His_6$ tag to the insert. Assembly of the truncated *radC2* into the linearized pET28a vector yielded the construct Δ11-RadC2-$His_6$-C, which was transformed, purified, and sequenced as described above to verify the construct.

Extraction of domoic acid isomers from *C. armata*

To extract domoic acid isomers from *C. armata,* 50 mg of dried algae was pulverized and extracted with 500 μL of 50% methanol. This mixture was mashed and vortexed every 15 mins, for 90 mins at 60˚C. The mixture was then centrifuged for 15 mins at (21000 xg, 20 min, 20 ˚C) to pellet debris. The supernatant was gently removed and filtered prior to injection (10 μL) on LC-HRMS.

Protein Purification

Both DabC and DsKabC were purified as described previously.[1,2] For the newly generated N-$His_6$-MBP-Δ7-RadA1 (RadA1, hereafter), N-$His_6$-RadC1 (RadC1, hereafter) and Δ11-RadC2-$His_6$-C (RadC2, hereafter) protein expression constructs, culturing and protein purification was performed using standard methodology. Cultures of *E. coli* BL21 (DE3)

chemically competent cells co-transformed with RadA1, RadC1, and RadC2 protein expression plasmids were shaken at 37 ˚C in 1 L of Terrific Broth (Fisher Bioreagents) to an $OD_{600}$ of ~0.6. Then, cultures were chilled to 18 ˚C and induced with 1 mM of isopropylthio-b-galactoside (IPTG). Flasks were shaken overnight (~16 hr) and cells were harvested by centrifugation (8000 xg, 10 min). Cell pellets were resuspended in 25 mL of lysis buffer (20 mM Tris pH 8, 500 mM NaCl, 10 mM imidazole, 10% glycerol) and stored at -80 ˚C for future purification.

All frozen cell pellets were defrosted, and cells were lysed by sonication using a Qsonica 6 mm tip at 40% amplitude for 20 cycles of 15 seconds on and 45 seconds off, gently mixing after 15 cycles. Cell lysate was then centrifuged for 30 mins at 20,000xg to pellet cell debris. Initial immobilized metal-affinity chromatography (IMAC) purification of RadA1, RadC1, and RadC2 was performed similarly for all soluble protein constructs. Briefly: clarified lysate was loaded at 2mL/min onto a 5mL HisTrap FF column (Cytiva) using a pre-equilibrated with wash buffer (20 mM Tris, 500 mM NaCl, 10mM imidazole, pH 8). The loaded column was washed with 10 column volumes of 8% elution buffer (20 mM Tris, 500 mM NaCl, 250 mM imidazole, pH 8). Protein was eluted using a linear gradient of 8-100% elution buffer over 15 column volumes in 4 mL fractions. Fractions were assessed using SDS-PAGE, and target protein containing fractions were pooled and concentrated using Amicon Ultra-15 50 kDa-cutoff (RadA1) and 30 kDa-cutoff (RadC1 and RadC2) centrifugal filters (Millipore Sigma).

RadC1 and RadC2 were further purified by size exclusion chromatography. Briefly: protein was concentrated to 1 mL using an Amicon Ultra-15 30 kDa-cutoff concentrator and further purified at a flow rate of 1 mL/min with a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare Life Sciences) pre-equilibrated with 50 mM HEPES pH 8.0, 250 mM NaCl, and

10% glycerol. Purity was checked using SDS-PAGE and relevant fractions were pooled, concentrated, and stored immediately at -80 ˚C.

Following IMAC purification, concentrated RadA1 protein was desalted and buffer-exchanged using PD-10 columns (Sephadex G-25 M, Cytiva) pre-equilibrated with storage buffer (50 mM HEPES pH 8, 250 mM NaCl, 10% glycerol) also containing 5 mM $MgCl_2$ to help with RadA1 stability. The buffer-exchanged RadA1 was further concentrated, aliquoted, and stored immediately at -80 ˚C. RadA1 was not stable upon attempted size exclusion chromatography using 50 mM HEPES pH 8.0, 250 mM NaCl, 5 mM $MgCl_2$, and 10% glycerol.

Enzymatic activity assays

Enzyme assays to demonstrate kainoid synthase (RadC1, RadC2, DabC, DsKabC) function were carried out as previously described.[1] Assays were conducted in 100 mM HEPES (pH 8.0), 100 mM KCl, 10% glycerol buffer with 1 mM ascorbate and 6.25 mM alpha-ketoglutarate (aKG). Prenylated glutamate substrate (NGG, cNGG, prekainic acid) was added to 1 mM, followed by 50 mM of enzyme and 50 mM of $FeSO_4$. Total reaction volume was 100 mL. Reactions were allowed to incubate at room temperature (~25 ˚C) overnight (~18 hr) and were then quenched with 100 mL of ice-cold methanol. Quenched reactions were centrifuged (21000 xg, 20 min, 4 ˚C) in a tabletop centrifuge and subsequently filtered with a nylon 0.22 mm pore CA membrane (Costar Spin-X) prior to injection (10 mL) on LC-HRMS. LC Method A was used for all kainoid synthase assays using the geranylated substrates (NGG and cNGG) and associated standards, whereas LC Method B was used for kainoid synthase assays on pre-kainic acid and associated standards. Substrates and products were identified using retention time and mass [M-1] in comparison to injections of prepared chemical standards. A similar 10 mL injection of *C.*

*armata* DA extract, prepared as described above, was also analyzed by LC-HRMS in comparison with cNGG enzyme assays and standards to ascertain relative DA isomer abundance in our *C. armata* isolate.

Enzyme assays to demonstrate *N*-prenyltransferase function were carried out as previously described, with modifications.[1,2,12] Assays were conducted in 100 mM HEPES (pH 8.0), 100 mM KCl, 10% glycerol buffer with 5 mM $MgCl_2$. Isoprene diphosphate (DMAPP or GPP) was added to 1 mM together with 20 mM of L-glutamate. Reactions were allowed to incubate at room temperature (~25 ˚C) overnight (~18 hr) and were then quenched with 100 mL of ice-cold methanol. Quenched reactions were centrifuged, filtered, and injected (10 mL) onto LC-HRMS. LC Method B was used for all *N*-prenyltransferase assays described here. Substrates and products were identified using retention time and mass [M-1] in comparison to injections of prepared chemical standards.

Phylogenetic analysis

Kainoid synthase tree: Initial BLAST searches revealed few closely related homologs outside of known KA or DA biosynthetic genes. These top tBLASTn hits were pooled with representative sequences from the isopenicillin N synthase-like (IPNS-like) InterPro family (IPR027443). Representative sequences were selected from UniProt using the UniRef50 function to provide a clustered set of sequences. Select sequences were further clustered using the CD-HIT suite, and a representative sequence from the largest 150 clusters was selected for further analysis. A multiple sequence alignment was generated using kalign (v2.04).[13] A maximum-likelihood (ML) phylogenetic tree was built using IQ-TREE, and the best-fit substitution model, VT+R5, was automatically selected.[14] This process of using top BLAST hits and representative

InterPro family sequences to construct a ML tree was repeated with RadD using the cytochrome P450 InterPro family (IPR001128). In addition to top BLAST hits and representative sequences, diatom P450 sequences were mined from publicly available transcriptomic datasets and a subset of related P450s was seeded into the tree. The substitution model LG+R6 was automatically selected. All trees were visualized in iTOL.[15]

### 3.3.2.3 Supplemental Tables

**Table S3.1.** Primers used in this study

| | |
|---|---|
| RadA_p28_N_001 | GGTGCCGCGCGGCAGCCATATGAAGGTACTTGCAGAAGACGACCC |
| RadA_p28_N_002 | GGTGGTGGTGGTGCTCGAGTCAAGTCGCTGTACTAATTATTTTAACGAGCTCG |
| RadC_p28_N_001 | GGTGCCGCGCGGCAGCCATATGTTTACGATCAAAGGAACGGAACTGAAC |
| RadC_p28_N_002 | GGTGGTGGTGGTGCTCGAGCTAGTAGTAGCCATGAAGAAACTTGTATTTTACG |
| p28_3prime_Amp | TGGTGGTGGTGGTGCTCGAG |
| RadA_d7_MBP | CTGTACTTCCAATCCGGATCCGACCCAAATGATGCACTAGCCCGTATCAAAACC |
| RadC2_d11_p28-C | CTTTAAGAAGGAGATATACCATGGATTTTAATCCCTTGGAAGTAGAGAAGCTCAATTG |
| RadC2_CtermOH | GCTCGAGTGCGGCCGCAAGCTTGTAGTACCCATGAGTAAACTTGTATTTTACGTAAGTGA |
| MBP_BB_001 | GGATCCGGATTGGAAGTACAGGTTCTCAGATCC |
| MBP_BB_002 | CTCGAGCACCACCACCACCACCACTGAG |
| p28-N_BB_001 | CATATGGCTGCCGCGCGGCACC |
| p28-N_BB_002 | CTCGAGCACCACCACCACCACCACTGAG |
| p28-C_BB_001 | CATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTG |
| p28-C_BB_002 | AAGCTTGCGGCCGCACTCGAGC |

**Table S3.2.** Summary of known KA and DA red algal producers from the literature. Right columns indicate if the study included chemical characterization of extracts and/or validation of biosynthetic genes.[16-19] *A homolog for RadD was not found in the sequenced *D. simplex* sample.

**Kainic Acid**

| order | family | genus | species | extract | gene |
|---|---|---|---|---|---|
| Ceramiales | Rhodomelaceae | Digenea | simplex | yes | yes |
| Ceramiales | Rhodomelaceae | Laurencia | papiillosa | yes | no |
| Ceramiales | Rhodomelaceae | Vidalia | obtusiloba | yes | no |
| Ceramiales | Ceramiaceae | Centroceras | clavulatum | yes | no |
| Palmariales | Palmariaceae | Palmaria | palmata | yes | yes |
| Palmariales | Palmariaceae | Palmaria | hecatensis | yes | no |
| Palmariales | Rhodophysemataceae | Rhodophysema | elegans | yes | yes |
| Halymeniales | Halymeniaceae | Grateloupia | filicina | yes | yes |

**Domoic Acid**

| order | family | genus | species | extract | gene |
|---|---|---|---|---|---|
| Ceramiales | Rhodomelaceae | Digenea | simplex | yes | no* |
| Ceramiales | Rhodomelaceae | Chondria | armata | yes | yes |
| Ceramiales | Rhodomelaceae | Osmundaria | obtusiloba | yes | no |
| Ceramiales | Rhodomelaceae | Amansia | glomerata | yes | no |
| Ceramiales | Rhodomelaceae | Alsidium | helminthochorton | yes | no |

**Table S3.3.** Quast statistics of the final assembled *C. armata* genome.

| **Statistics without reference** | **CA_masurca_bothONT** |
|---|---|
| # contigs | 2991 |
| # contigs (>= 0 bp) | 2991 |
| # contigs (>= 1000 bp) | 2818 |
| # contigs (>= 5000 bp) | 2271 |
| # contigs (>= 10000 bp) | 2035 |
| # contigs (>= 25000 bp) | 1769 |
| # contigs (>= 50000 bp) | 1490 |
| Largest contig | 3292845 |
| Total length | 507692717 |
| Total length (>= 0 bp) | 507692717 |
| Total length (>= 1000 bp) | 507578445 |
| Total length (>= 5000 bp) | 506026847 |
| Total length (>= 10000 bp) | 504384905 |
| Total length (>= 25000 bp) | 499967534 |
| Total length (>= 50000 bp) | 489680988 |
| N50 | 643002 |
| N75 | 256360 |
| L50 | 226 |
| L75 | 532 |
| GC (%) | 45.34 |
| **Mismatches** | |
| # N's | 19500 |
| # N's per 100 kbp | 3.84 |

**Table S3.4.** BUSCO assessment of publicly available red macroalgal genomes

| Scientific name | Assembly | Complete | Single-copy | Duplicated | Fragmented | Missing | n=255 |
|---|---|---|---|---|---|---|---|
| Chondria armata | | 74.5 | 70.2 | 4.3 | 7.5 | 18.0 | 255 |
| Agarophyton vermiculophyllum | AgarVerm_1.0 | 59.6 | 58.8 | 0.8 | 6.3 | 34.1 | 255 |
| Asparagopsis taxiformis | ASM1839795v1 | 78.0 | 69.4 | 8.6 | 5.5 | 16.5 | 255 |
| Chondrus crispus | ASM35022v2 | 72.2 | 71.0 | 1.2 | 8.2 | 19.6 | 255 |
| Digenea simplex | ASM479842v1 | 71.4 | 69.8 | 1.6 | 7.5 | 21.1 | 255 |
| Gracilariopsis chorda | GraCho1.0 | 76.9 | 74.9 | 2.0 | 5.5 | 17.6 | 255 |
| Gracilariopsis lemaneiformis | Glem_v01 | 75.3 | 74.1 | 1.2 | 6.7 | 18.0 | 255 |
| Kappaphycus alvarezii | ASM220596v3 | 59.6 | 41.6 | 18.0 | 11.8 | 28.6 | 255 |
| Neoporphyra haitanensis | OUC_PyHait | 52.6 | 51.4 | 1.2 | 13.3 | 34.1 | 255 |
| Neopyropia yezoensis | ASM982973v1 | 54.9 | 54.5 | 0.4 | 13.7 | 31.4 | 255 |
| Porphyra umbilicalis | P_umbilicalis_v1 | 42.4 | 42.0 | 0.4 | 14.1 | 43.5 | 255 |

**Table S3.5.** Kainoid synthase representative sequences. UniProt accession numbers of representative sequences of key taxonomic groups from the isopenicillin N synthase-like (IPNS-like) InterPro family (IPR027443). Sequences highlighted in grey represent NCBI accession numbers of top BLAST hits.

| | | | | |
|---|---|---|---|---|
| D4E802 | A0A2V3J2T9 | A0A6S7ZS46 | A0A6P4B8E0 | A0A0W0G3P1 |
| A0A485B8P9 | A0A6V0PIX2 | A0A6V1FBN5 | A0A371EF89 | A0A094A4Z7 |
| A0A1I2FBH3 | M2WQW5 | A0A6V1VUR4 | A0A4S4F1G2 | A0A093Z4A8 |
| A0A0Q0DE95 | R7QT26 | A0A6S8Y206 | A0A5N5L5D4 | A0A7D7ZSX3 |
| A0A2X4YWX6 | A0A5J4YHC7 | A0A6U4KKR1 | B9RI27 | WP_075477798 |
| A0A6M4A7N8 | A0A6T5W5I9 | K0RCE9 | A0A2U1NEA5 | WP_112713920 |
| W6RJX4 | A0A6T6N7U3 | A0A6V3GZM2 | A0A2G3D1B6 | A0A1R2C824 |
| A0A653YTM6 | A0A5J4Z228 | A0A6U1BI44 | A0A5N5P4G6 | A0A1R2CN02 |
| A0A1H4S654 | A0A6T9YZV4 | A0A6U3NAE4 | A0A5J9TM82 | P18548 |
| A0A3N1V457 | A0A6T6NYH4 | A0A6V0NRI9 | A0A094H7E5 | |
| A0A562MRU1 | A0A6V0MET1 | A0A6U0GHJ2 | A0A135S7A4 | |
| A2WDT8 | A0A5J4Z328 | A0A6U6F6P8 | A0A139IR59 | |
| M5D0U2 | A0A6T6LKS5 | A0A6V3DN33 | A0A1S9RV91 | |
| A0A663ASE4 | A0A2V3J2G5 | A0A6U2NKB2 | A0A5N6J8W1 | |
| A0A4R3ZW85 | A0A6T6KDA2 | A0A124SC96 | A0A2G8SRB9 | |
| A0A2N3PW13 | M1VKP3 | A0A438J382 | A0A3M7LXI8 | |
| A0A0P9P600 | A0A6V0QHG1 | A0A2U1NH25 | A0A1L9TCD4 | |
| A0A387HMF6 | A0A1X6PK80 | A0A4V3WLV0 | A0A1X6MM30 | |
| A0A286HD09 | A0A5J4Z952 | A0A6A2YJD0 | A0A444RNV4 | |
| A0A021VNH7 | A0A6V1LTB9 | A0A5B6VHC0 | A0A0F2M517 | |
| Q98N28 | A0A6V1I903 | A0A3S3QTW4 | A0A4U0WWH5 | |
| I3UAA2 | A0A6S8QUQ0 | W9QYP6 | A0A395SJD4 | |
| Q3JLE9 | A0A6U3NP22 | A0A498I7M4 | A0A0G2FG99 | |
| A0A6P2CA42 | A0A6U1FFP2 | M4CSL2 | A0A2P2HFL0 | |
| C4WFH3 | A0A6T0NQR4 | A0A3Q7I749 | A0A364MXH3 | |
| A0A6T6Q0U9 | A0A6T0LLC5 | A0A540MTJ5 | A0A135SK80 | |
| A0A6T6NV76 | A0A6V1Y066 | A0A314KJM9 | A0A4P7NST2 | |
| R7QC14 | A0A6S8G773 | A0A5J5BV07 | B8MRF9 | |
| A0A6T6CML0 | A0A6T2XKK6 | A0A0B0MG73 | W3WSK3 | |
| A0A5J4Z4R2 | A0A6T0K2C4 | A0A1J6JT00 | A0A1S9D541 | |

**Table S3.6.** CYP450 representative sequences. UniProt accession numbers of representative sequences of key taxonomic groups from the cytochrome P450 InterPro family (IPR001128). Sequences highlighted in grey represent the Protein ID of sequences from the JGI PhycoCosm.

| | | | | |
|---|---|---|---|---|
| A0A6V1GDW8 | A0A4R2CXH8 | A0A094AXQ3 | A0A498I3I0 | A0A6T6AFA8 |
| A0A6V1V780 | A0A2Z4V4B8 | A0A2N1NSW2 | A0A3Q7IZP9 | M2VTT7 |
| A0A3S7L8P2 | A0A1H5YLD1 | A0A094JJ65 | A0A5N6QLC6 | M1VHJ8 |
| A0A6U5MHZ5 | A0A4R4VQ17 | A0A2G7FJ09 | A0A200RBS5 | A0A2V3IHK9 |
| A0A6V3BWI7 | A0A249PFW7 | A0A094HR63 | A0A2G3C5P5 | 247725 |
| K0RTJ9 | A0A0M4Q8A0 | A0A135TVV1 | A0A6A1UQ32 | 202814 |
| A0A6S9PJ72 | A0A290XS85 | A0A2G7FQT3 | A0A6L2M770 | 130274 |
| A0A6U3SSW4 | A0A498PLX0 | A0A0G4KS67 | A0A5J4ZT80 | 32491 |
| A0A6U0ZEH0 | X8CRZ0 | W9CUQ1 | A0A371HFR0 | |
| A0A6U1BAX8 | A0A2G6CH09 | A0A2P4QZ99 | A0A1R3I738 | |
| A0A6U2T721 | A0A4D4L8P2 | A0A165K1U7 | A0A6T5VV58 | |
| A0A6T0HIW2 | A0A3N6E6F4 | A0A0G4NKU9 | A0A1X6P1P4 | |
| A0A6U0GZL9 | A0A0P0R9V1 | A0A175W394 | A0A6V0RYK1 | |
| A0A6V0Y2P1 | A0A1I4AUB2 | A0A4Y9ZEV9 | A0A6T6MKE5 | |
| A0A6V0Q6S4 | A0A2I7WCV3 | A0A4U9EEN7 | A0A5J4YL64 | |
| A0A6S8EW11 | I7FQV2 | A0A6N2LYY4 | A0A5J4YNM0 | |
| A0A6T9B9Y7 | A0PL28 | A0A5D2RYA7 | A0A2V3IWP3 | |
| A0A6V2X6A8 | A0A6H9IY50 | A0A2N9HUK6 | A0A6T6AE48 | |
| A0A6V2E486 | A0A3G2J7I5 | A0A2N9H9F4 | A0A6T6Q6L2 | |
| A0A1E7EPM2 | A0A1Q5HD23 | A0A5N5LFD9 | A0A1X6PC05 | |
| A0A6S8LNV9 | A0A2U3E828 | A0A200Q069 | A0A5J4YWA9 | |
| A0A6V2JND1 | A0A421J420 | A0A6A2XWI3 | R7QH57 | |
| A0A6V4QEM2 | A0A1S9DKV5 | A0A498J2J3 | A0A6V0RHZ8 | |
| A0A6U1UVS9 | A0A484GA20 | A0A6N2LLD8 | A0A6T6K777 | |
| A0A6U3S297 | A0A094D2J3 | A0A166FQ11 | A0A1X6P2T0 | |
| A0A6N0WX86 | A0A2U3DT16 | A0A1Q3CNA7 | A0A6V0RF11 | |
| U5ELP1 | A0A2N1NTF4 | A0A6A3B2M2 | R7QBA0 | |
| A0A3L8C4K3 | J4KML6 | A0A5B6WX83 | A0A6T6B6C5 | |
| C2W9K6 | A0A4S9AE14 | A0A498I2M7 | A0A6T6ATK7 | |
| A0A3Q8XTT3 | A0A1S9RMP3 | A0A498KMK2 | A0A6V0MZH5 | |

### 3.3.2.4 Supplemental Figures

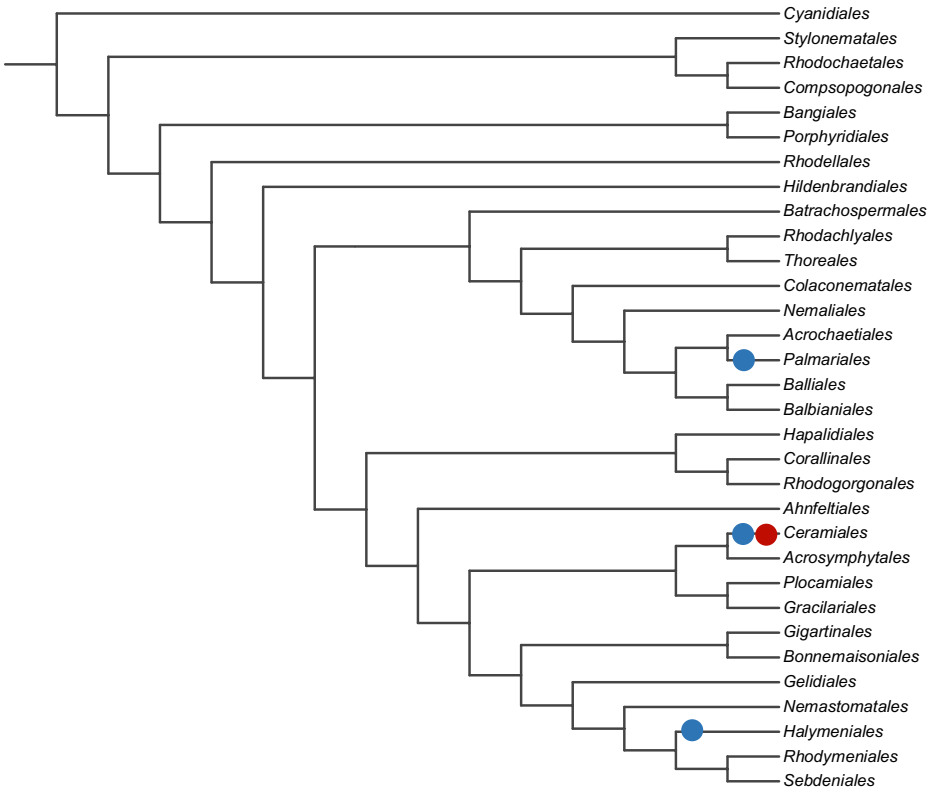



**Figure S3.1.** Cladogram of DA and KA producers by order in Rhodophyta. Blue and red dots indicate KA and DA producing orders, respectively. See Table S3.1 for a list of producing organisms.

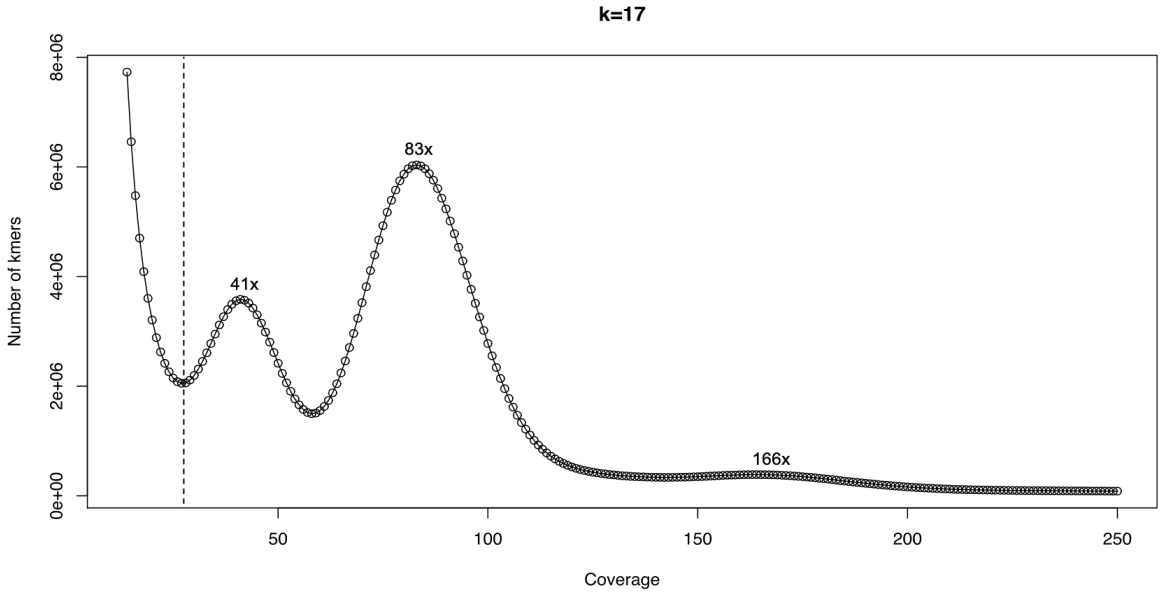



**Figure S3.2.** K-mer analysis of the *Chondria armata* genome. Tetraploid (41x), diploid (83x), and haploid (166x) peaks are labeled. Sequences to the left of the dashed line were omitted. Haploid genome size was estimated to be 480 Mb.

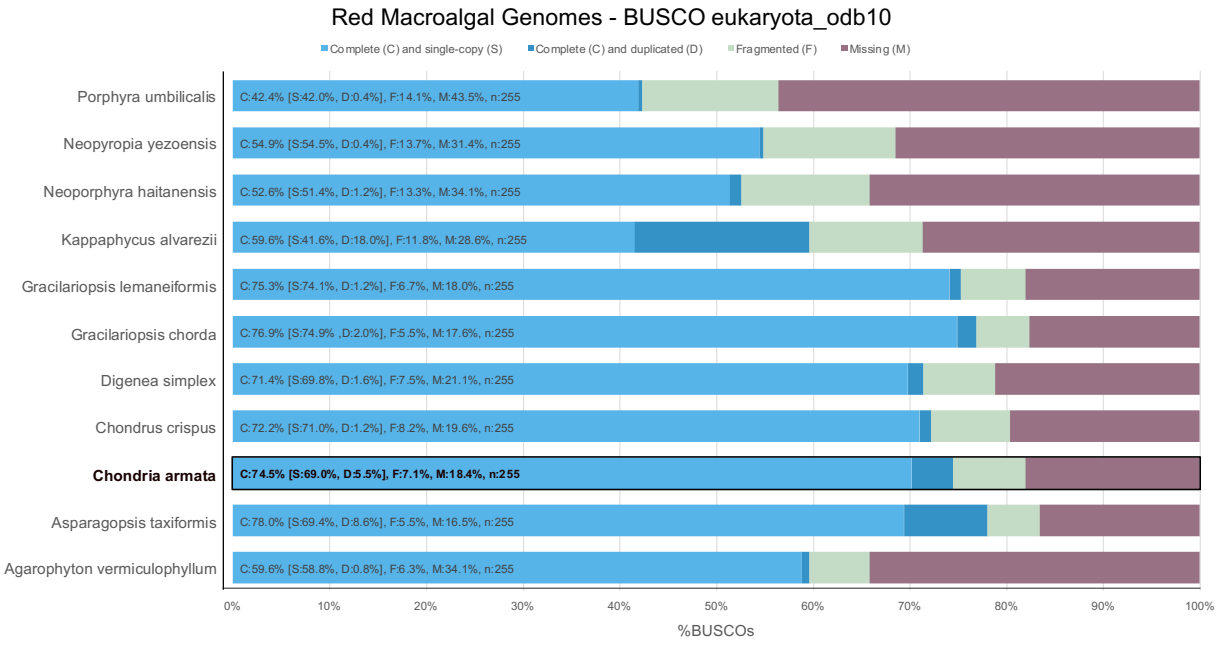



**Figure S3.3.** BUSCO assessment of red macroalgal genomes. *C. armata* BUSCO assessment highlighted in black box.

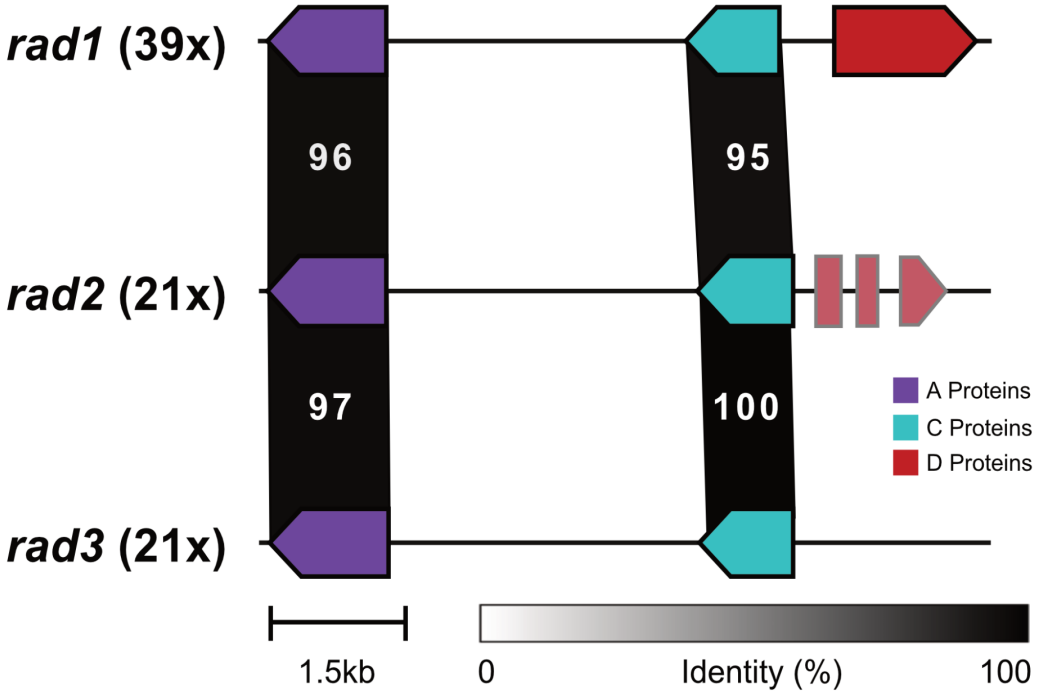



**Figure S3.4.** The three copies of the *rad* cluster within the *C. armata* genome. Mean coverage across each copy in parentheses. The sequences of *rad1*, *rad2*, and *rad3* are deposited in GenBank, accession numbers OK169902, OK169903, and OK169904, respectively.

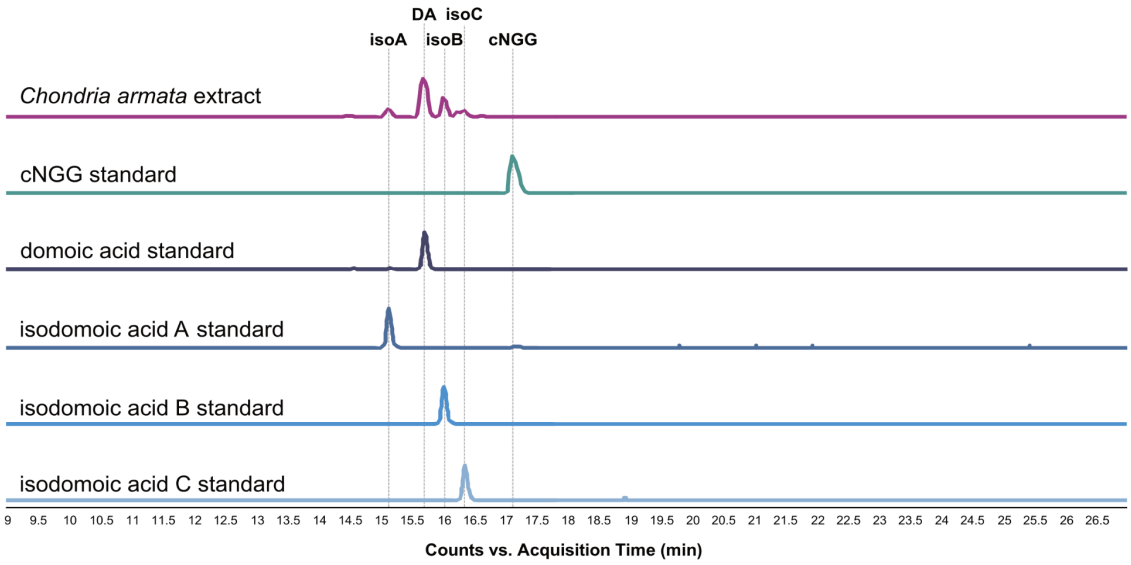



**Figure S3.5.** Domoic acid isomers in *C. armata.* Negative mode LC-HRMS chromatograms using LC Method A of domoic acid (DA), isodomoic acids (isoA, isoB, isoC), and cNGG standards compared with *C. armata* extract, showing extracted ion chromatograms for anticipated DA isomer and cNGG masses (EIC 310.1293 m/z and 312.1453 m/z, respectively, +/- 0.2 m/z).

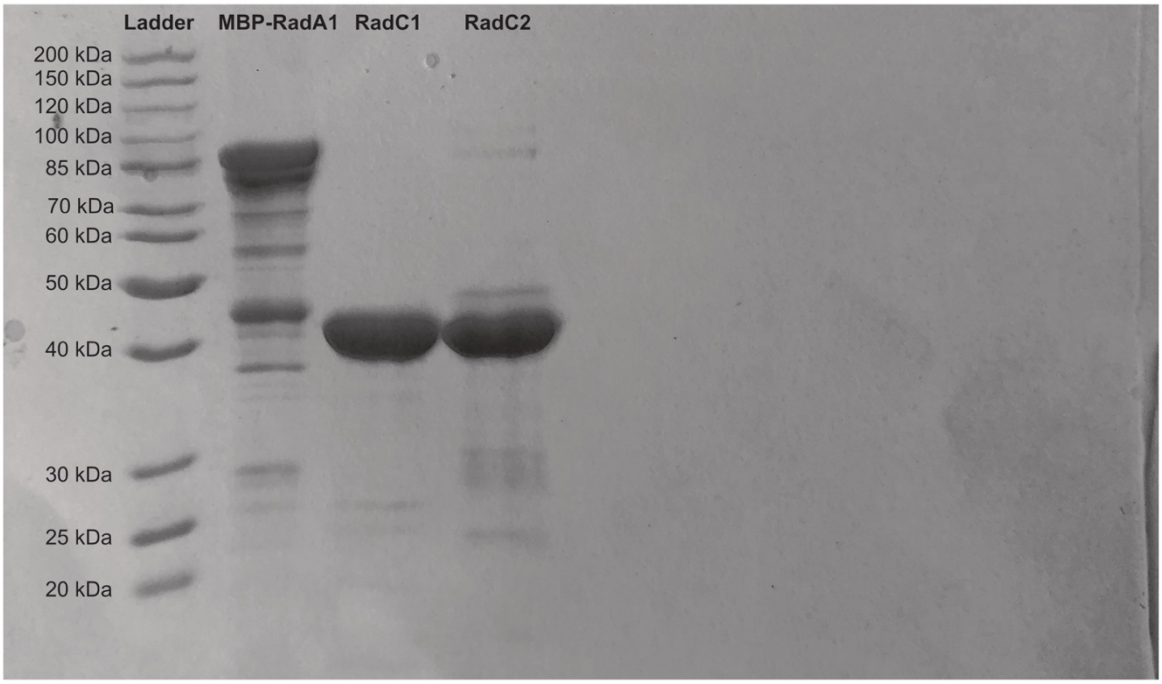



**Figure S3.6**. 10% SDS-PAGE gel. Sample order is as follows: Ladder, MBP-RadA1, RadC1, RadC2. Note the additional band in the RadA lane is the MBP tag without RadA. RadA was unable to be further purified by gel filtration due to instability of the construct.

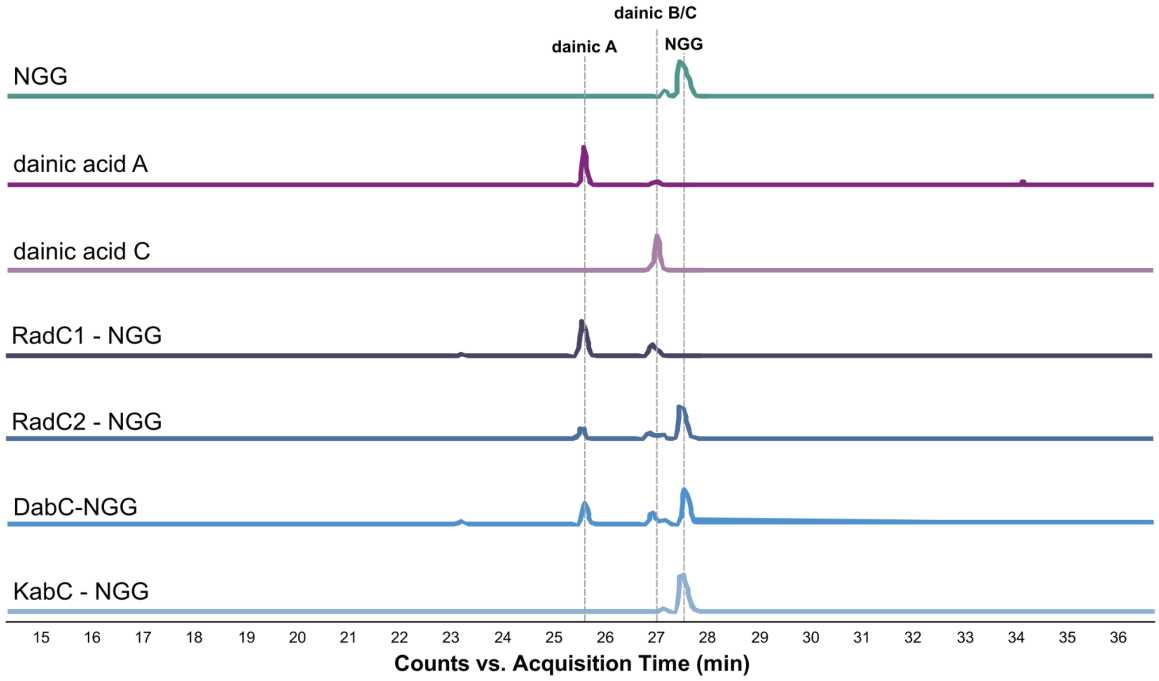



**Figure S3.7.** Kainoid synthase reactions were set up as previously described using NGG (1 mM). Negative mode LC-HRMS chromatograms using LC Method A of kainoid synthase reactions are shown together with dainic acid standards, showing combined extracted ion chromatograms for anticipated products and substrates (EIC 280.1554 m/z and 282.1711 m/z, respectively, +/- 0.2 m/z).

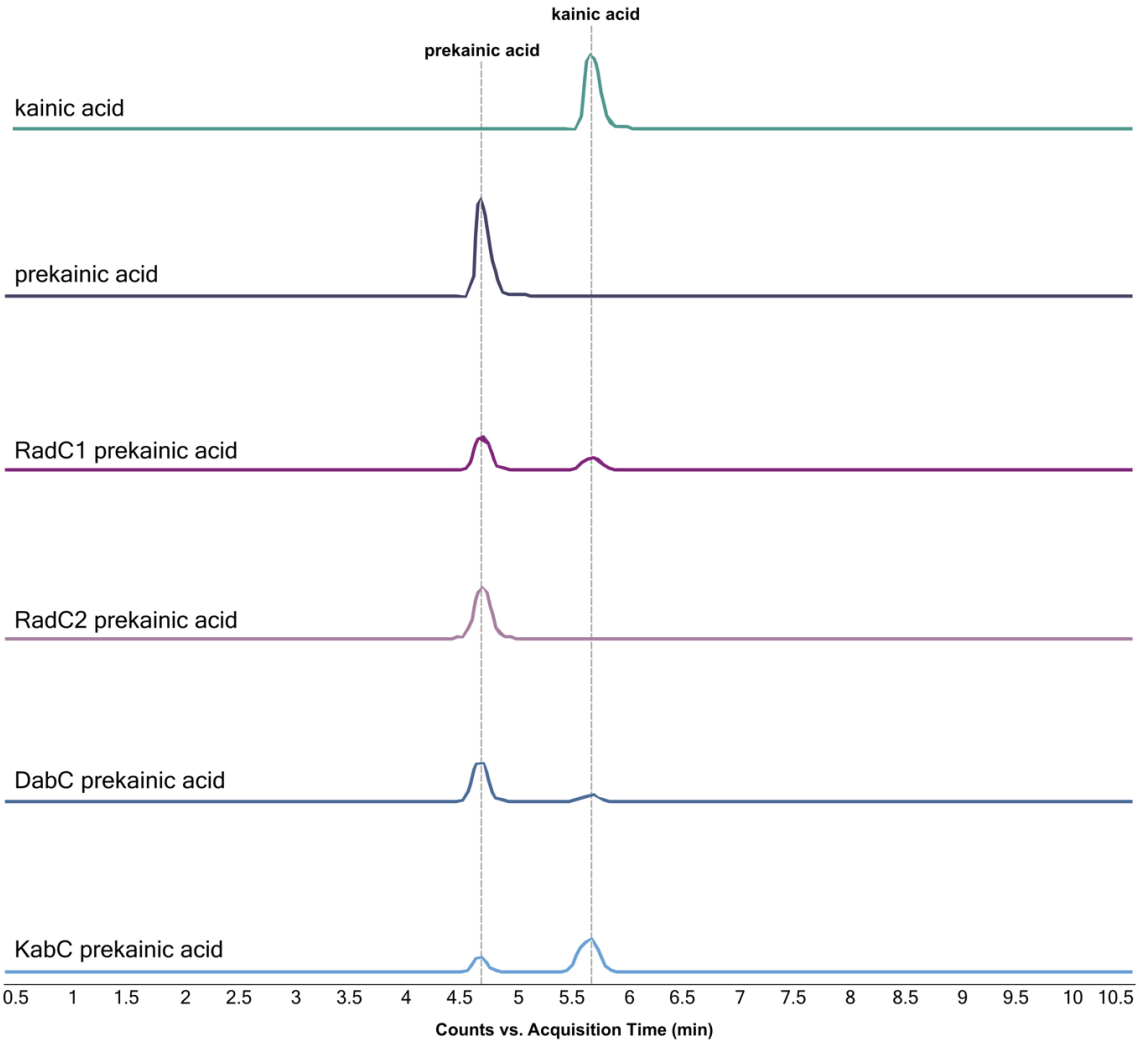



**Figure S3.8.** Prekainic acid substrate screen. Kainoid synthase reactions were set up as previously described using prekainic acid (1 mM). Negative mode LC-HRMS chromatograms using LC Method B of kainoid synthase reactions are shown together with prekainic acid and kainic acid standards, showing combined extracted ion chromatograms for anticipated products and substrates (EIC 212.0300 m/z and 214.0500 m/z, respectively, +/- 0.2 m/z).

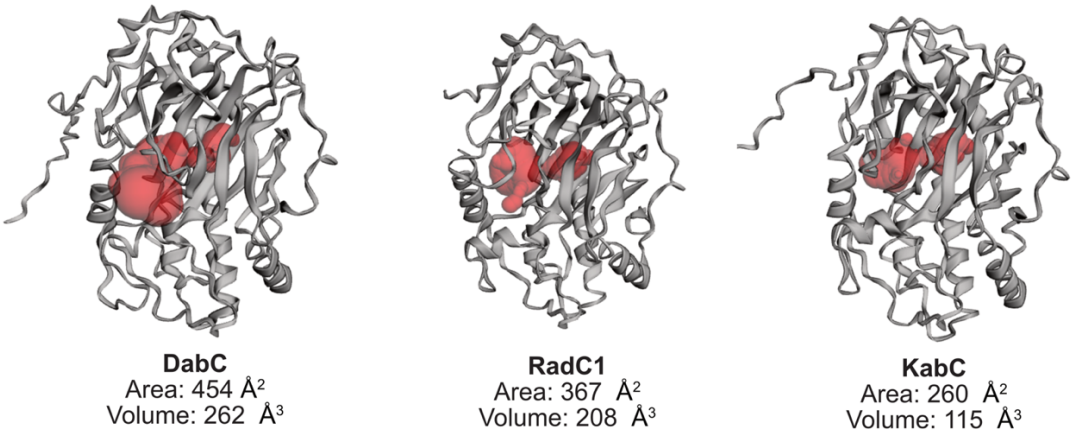



**Figure S3.9.** CASTp output of modeled pocket volumes for *Pseudo-nitzschia multiseries* DabC, *C. armata* RadC1, and *Digenea simplex* KabC.[20] Binding pockets are highlighted in red, calculated enzyme pocket areas and volumes are labeled below protein models in terms of solvent accessible surface area.

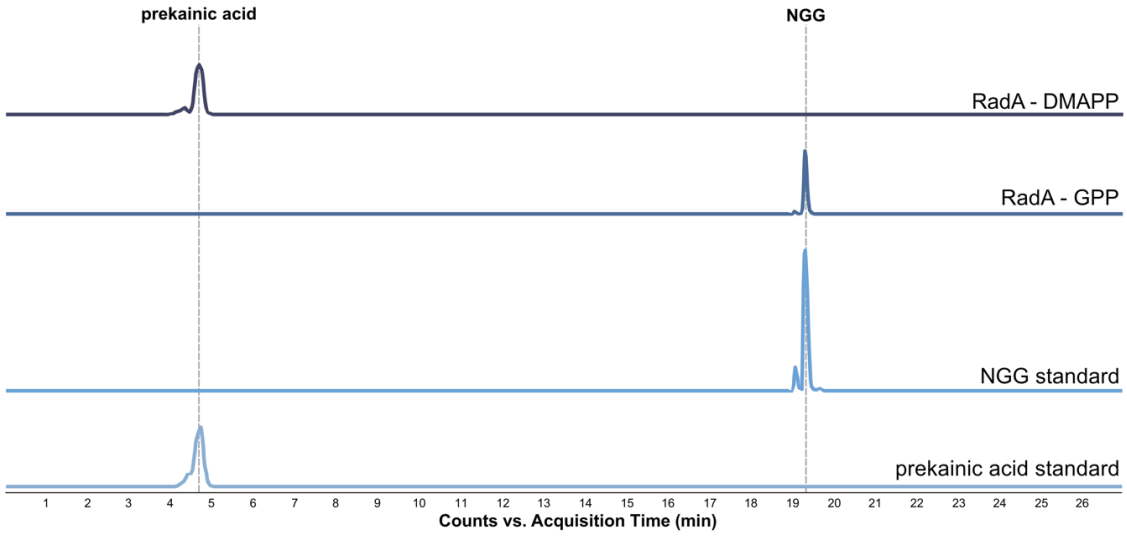



**Figure S3.10**. Overnight assays with purified MBP-Δ7-RadA1. Substrate screening suggest both GPP and DMAPP can be accepted as substrates to make NGG and prekainic acid, respectively. RadA1 *N*-prenyltransferase reactions were set up as previously described using L-glutamate acid (20 mM) and DMAPP or GPP prenyl donors (1 mM). Negative mode HRMS chromatograms using LC Method B of RadA1 reactions are shown together with prekainic acid and NGG standards, showing individual extracted ion chromatograms for anticipated NGG and prekainic acid products (EIC 282.1711 m/z and 214.0500 m/z, respectively, +/- 0.2 m/z).

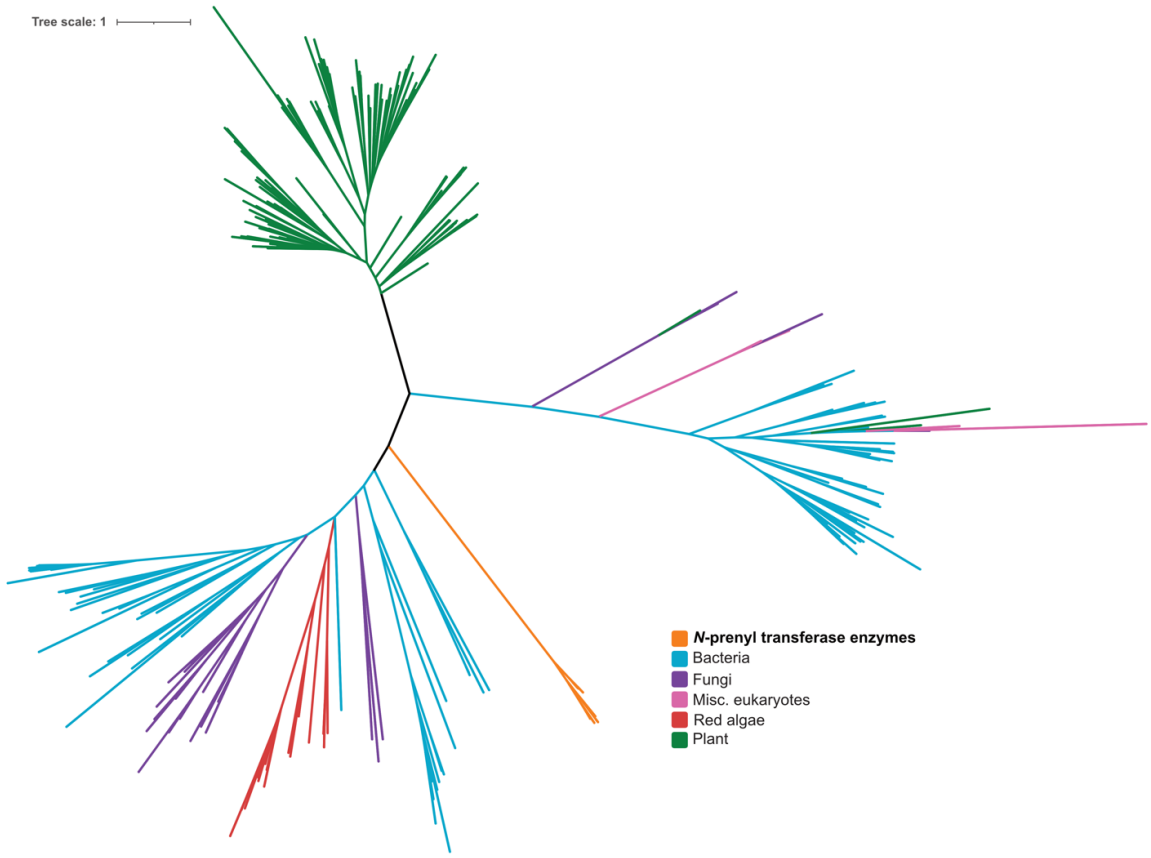



**Figure S3.11.** *N*-prenyltransferase enzyme maximum-likelihood phylogenetic tree. RadA sequences mapped to previously reported tree.[12] They clade with the other *N*-prenyltransferases and are very distantly related to other terpene cyclases.

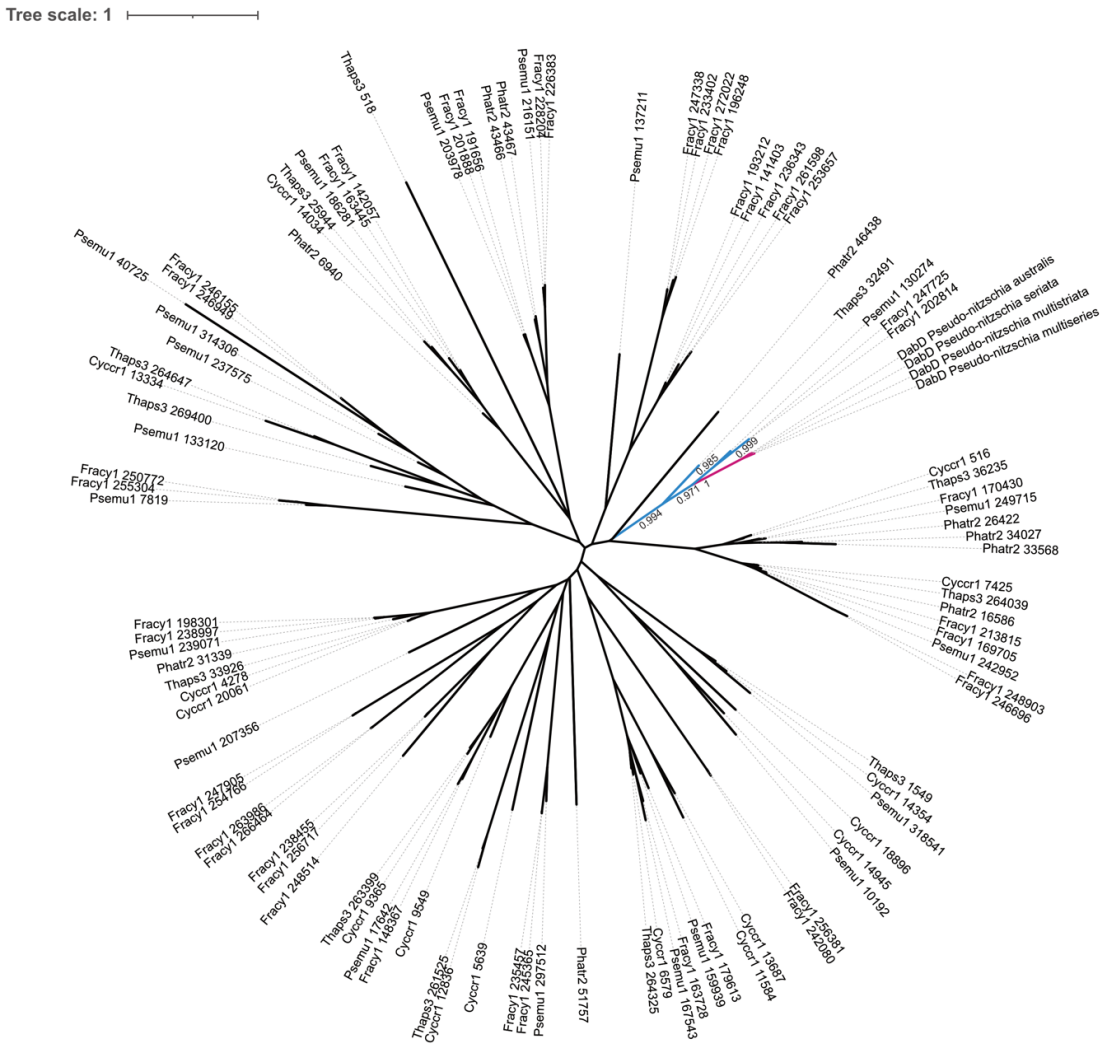



**Figure S3.12.** Expanded diatom CYP450 maximum-likelihood phylogenetic tree. Diatom sequences used in Figure 4B highlighted in blue and red. Numeric identifiers are Joint Genome Institute Protein IDs (JGI PIDs) from the respective publicly available diatom genomes through JGI.

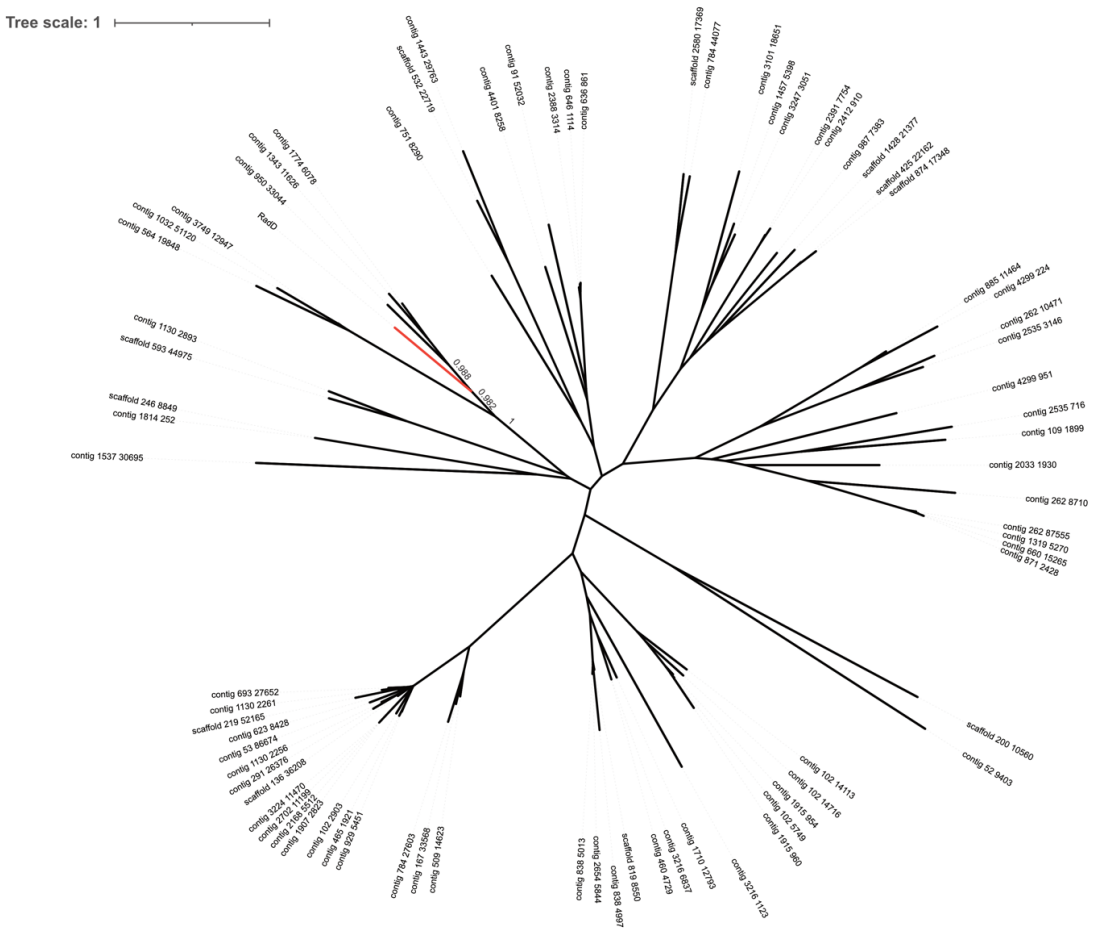



**Figure S3.13.** Expanded *C. armata* CYP450 maximum-likelihood phylogenetic tree

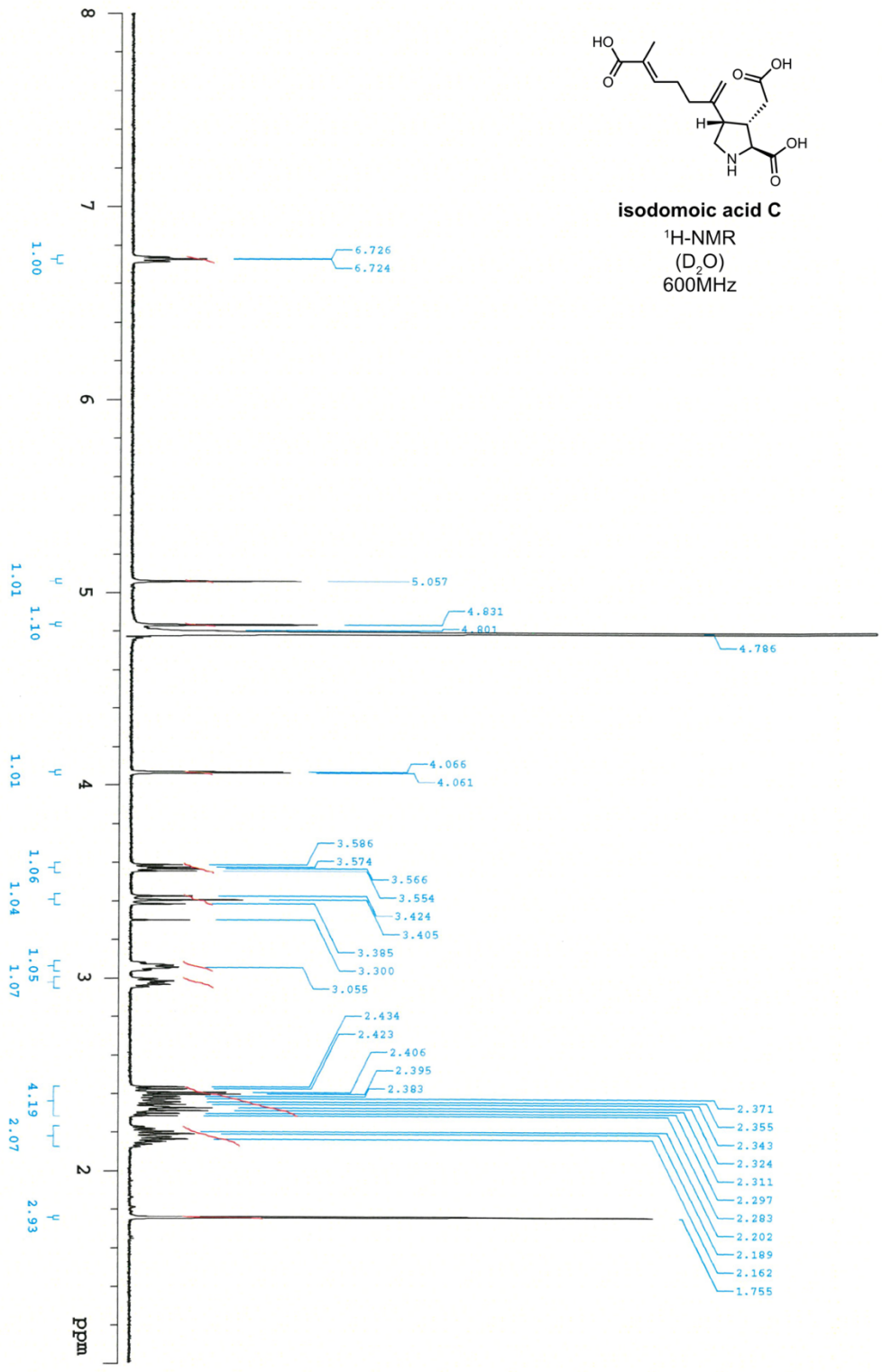



**Figure S3.14.** [1]H-NMR spectrum of isodomoic acid C. Analysis performed in $D_2O$/600 MHz. The signal of the residual MeOD was adjusted at δ 3.30 ppm as the internal reference in $D_2O$.

### 3.3.2.5 Works Cited to the Supplementary Information


1. J. K. Brunson, S. M. K. McKinnie, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. Smith, D. A. Hutchins, A. E. Allen, B. S. Moore. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018)

2. J. R. Chekan, S. M. K. McKinnie, M. L. Moore, S. G. Poplawski, T. P. Michael, B. S. Moore. Scalable biosynthesis of the seaweed neurochemical, kainic acid. *Angew. Chem. Int. Ed.* **58**, 8454–8457 (2019)

3. Y. Maeno, Y. Kotaki, R. Terada, Y. Cho, K. Konoki, M. Yotsu-Yamashita. Six domoic acid related compounds from the red alga, *Chondria armata*, and domoic acid biosynthesis by the diatom, *Pseudo-nitzschia multiseries*. *Sci. Rep.* **8**, 356 (2018)

4. M. Meda, T. Kodama, T. Tanaka, H. Yoshizumi, T. Takemoto, K. Nomoto, T. Fujita. Structures of isodomoic acids A, B and C, novel insecticidal amino acids from the red alga *Chondria armata*. *Chem. Pharm. Bull.* **34**, 4892–4895 (1986)

5. A. M. Bolger, M. Lohse, B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014)

6. R. Kajitani, D. Yoshimura, M. Okuno, Y. Minakuchi, H. Kagoshima, A. Fujiyama, K. Kubokawa, Y. Kohara, A. Toyoda, and T. Itoh. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nat. Commun.* **10**, 1702 (2019)

7. A. V. Zimin, D. Puiu, M-C. Luo, T. Zhu, S. Koren, G. Marçais, J. A. Yorke, J. Dvořák, S. L. Salzberg. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Re*s. **5**, 787-792 (2017)

8. M. Seppey, M. Manni, E. M. Zdobnov. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019)

9. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018)

10. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin. 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009)

11. S. W. Lee, D. A. Mitchell, A. L. Markley, M. E. Hensler, D. Gonzalez, A. Wohlrab, P. C. Dorrestein, V. Nizet, J. E. Dixon. Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc. Natl. Acad. Sci.* **15**, 5879-5884 (2008)

12. J. R. Chekan, S. M. K. McKinnie, J. P. Noel, B. S. Moore. Algal neurotoxin biosynthesis repurposes the terpene cyclase structural fold into an N-prenyltransferase. *Proc. Natl. Acad. Sci.* **23**, 12799-12805 (2020)

13. T. Lassmann. Kalign 3: multiple sequence alignment of large datasets. *Bioinformatics* **36**, 1928–1929 (2020)

14. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020)

15. I. Letunic, P. Bork. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. **49**, W293-W296 (2021)

16. G. Impellizzeri, S. Mangiafico, G. Oriente, M. Piattelli, S. Sciuto, E. Fattorusso, S. Magno, C. Santacroce, D. Sica. Amino acids and low-molecular-weight carbohydrates of some marine red algae. *Phytochemistry* **14**, 1549–1557 (1975)

17. Y. Shimizu, S. Gupta, K. Masuda, L. Maranda, C. K. Walker, R. Wang. Dinoflagellate and other microalgal toxins: chemistry and biochemistry. *Pure Appl. Chem.* **61**, 513–516 (1989)

18. M. Sato, T. Nakano, M. Takeuchi, N. Kanno, E. Nagahisa, Y. Sato. Distribution of neuroexcitatory amino acids in marine algae. *Phytochemistry* **42(6)**, 1595–1597 (1996)

19. M. V. Laycock, A. S. W. de Freitas, J. L. C. Wright. Glutamate agonists from marine algae. *J. Appl. Phycol.* **1**, 113–122 (1989)

20. W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018)

### 3.4 Acknowledgments and Author Contributions

The main text and supplemental information sections of Chapter 3 have been submitted for publication as "T. S. Steele*, **J. K. Brunson***, Y. Maeno, R. Teruda, A. E. Allen, M. Yotsu-Yamashita, J. R. Chekan, B. S. Moore. Domoic acid biosynthesis in red alga *Chondria armata* suggests a complex evolutionary history for toxin production (2021) Manuscript in submission." The dissertation author was one of two equally contributing primary investigators and authors of this manuscript.

T.S.S., J.K.B., J.R.C., and B.S.M., conceived the project, designed the experiments, analyzed the data, and wrote the paper, with input from all authors. T.S.S. and J.R.C. performed sequencing, genome mining, and analysis experiments. T.S.S. performed evolution and phylogenetic analysis experiments. T.S.S. and J.K.B. performed protein expression experiments. J.K.B. performed enzymology experiments. J.K.B. and J.R.C. designed vectors and expression constructs. T.S.S. and J.R.C. performed molecular modeling experiments. Y.M. and M.Y.Y isolated and characterized isodomoic acid standards. R.T. collected *C. armata* seaweed samples. A.E.A. for helpful discussions.

We thank T. Michael (Salk Institute for Biological Studies) for assistance with DNA sequencing, G. J. Smith (Moss Landing Marine Laboratories) for providing the diatom image, T. Teruya (University of Ryukyus) and G. Saunders (University of New Brunswick) for providing seaweed images, and G. Rouse and T. Fallon (Scripps Institution of Oceanography) for helpful discussions. This research was supported by the National Oceanic and Atmospheric Administration (NA19NOS4780181 to B.S.M. and A.A.), the National Science Foundation through a GRFP fellowship to T.S.S., the National Institutes of Health (F31ES030613 to J.K.B.), the Simons Foundation Fellowship of the Life Sciences Research Foundation to J.R.C., and the

University of North Carolina at Greensboro research start-up funds to J.R.C. We dedicate the manuscript described in this chapter in memory of Professor Ryan Botts, undergraduate advisor to T.S.S. (2020).

# Chapter 4: Metatranscriptomic sequencing and barcoding of a major harmful algal bloom event in Monterey Bay, California

## 4.1 Introduction to Chapter 4

The California Current is one of four major currents making up the North Pacific Subtropical Gyre, together with the Kuroshio Current off the coast of Japan and East Asia, the North Pacific Current to the north, and the North Equatorial Current to the south.[1,2] Of these four major currents, the California Current is the eastern boundary current, bounded by the North American West Coast at the eastern edge of the North Pacific. In general, eastern boundary currents are prime locations for strong, biologically relevant upwelling due to the action of prevailing winds fueling Eckman transport of surface waters away from the coastline. Coastal water transported in this manner is readily replaced by cold, high salinity deep water that is also typically rich in nutrients required for phytoplankton growth and biological productivity.[3] The four major eastern boundary upwelling systems, characterized by especially strong upwelling and high biological productivity include the Benguela Current off the southwest coast of Africa, the Humboldt Current off the western coast of South America, the Canary Current off the Northwest coast of Africa, and the California current. The effects of biologically-relevant upwelling in the California current has been well studied within the California Current Ecosystem (CCE) at a variety of trophic levels.[4–9]

Because the CCE is a major upwelling zone, the North American West Coast is especially susceptible to the initiation and development of harmful algal blooms (HABs), especially in the Spring and Summer months when seasonal, upwelling-favorable winds are strongest.[10] Many different types HAB events have been reported off the coast of North America for the past several decades, with major HAB taxa including saxitoxin-producing dinoflagellates and domoic acid (DA) producing blooms of the diatom *Pseudo-nitzschia*.[11] The first recorded DA outbreak occurred on the North American West Coast in 1991, just a few years after the

biotoxin was identified as a shellfish poison in Prince Edward Island mussels in 1987. The 1991 outbreak resulted in the deaths of seabirds in Monterey Bay, California, including pelicans and cormorants.[12,13] Contamination of DA was also detected in razor clams and Dungeness crab in Northern California and the Pacific Northwest.[14] While information suggesting DA outbreaks pre-1991 is limited, diatoms belonging to the genus *Pseudo-nitzschia* have been described off the North American West Coast as far back as 1920, although these diatoms were at the time classified as *Nitzschia seriata*.[15] A subsequent DA outbreak in 1998 resulted in the mass mortality and stranding of sea lions and other marine mammals along the central California coastline, with razor clam fisheries in the Pacific Northwest reporting high levels of accumulated DA similar to the 1991 event.[16,17]

Following these reports of toxic events in 1991 and 1998, DA-producing HABs have been detected and described off the North American West Coast nearly every year since.[11,18] Various "hotspots" for *Pseudo-nitzschia* HAB activity have emerged at discrete locations along the North American West Coast, with the Juan de Fuca Eddy and the Santa Barbara Channel serving as prime examples of hotspots off the coast of Washington and Southern California, respectively.[19–21] One of the major *Pseudo-nitzschia* and DA hotspots off the coast of Central California is Monterey Bay, a large body of water just south of San Francisco and San Jose that experiences regular periods of strong seasonal upwelling followed by dramatic relaxation of upwelling every year, thereby driving phytoplankton bloom and bust cycles and contributing to local primary productivity.[22–24] Two major sources of nutrient rich, cold-water upwelling have been identified: Point Año Nuevo just north of Monterey Bay and Point Sur, immediately to the south (see Fig. 4.1 in section 4.3).[22] Both upwelling centers lie outside of the bay itself and flow into the mouth of Monterey Bay, with upwelling from Point Año Nuevo providing an "upwelling

shadow" at the northern mouth of the bay contributing to retention of water masses.[23] Upwelling-favorable winds in Monterey Bay typically start in early spring and persist into the summertime. Relaxation of these seasonal winds and subsequent relaxation of upwelling results in the migration of low-salinity oligotrophic waters into the bay from offshore due to a "meandering" California Current that is no longer bounded by the nearshore upwelling centers at Pt. Año Nuevo and Pt. Sur.[22] These periods of upwelling and relaxation have been directly demonstrated to influence phytoplankton populations in Northern Monterey Bay.[24,25]

As one might expect, the seasonal cycles of nutrient-rich upwelling in Monterey Bay have the potential to fuel HAB events, particularly those caused by DA-producing *Pseudo-nitzschia*. However not all *Pseudo-nitzschia* blooms occurring within Monterey Bay are equally toxic. One study by Bowers et al. focused on two specific bloom events in the bay: a fall 2013 bloom of *Pseudo-nitzschia fraudulenta* and the local spring 2015 bloom of *Pseudo-nitzschia australis* that coincided with the larger North American West Coast *P. australis* event discussed previously in Chapter 1.[26] Despite both bloom events displaying similar concentrations of *Pseudo-nitzschia* cells, the 2015 event produced orders of magnitude more DA per liter. While this discrepancy in toxicity may be linked to the dominant species present, with *P. australis* contributing to the bulk of DA regularly produced off the Central California Coast, local oceanographic conditions may have also played a role. For example, the 2015 event was also characterized by historically low concentrations of silica in Monterey Bay, a key nutrient required for diatom growth that has also been demonstrated to induce DA production in culture when supply of the nutrient is limited.[25,27,28] Further study is required to fully understand oceanographic inputs into bloom toxicity to better monitor and forecast toxic *Pseudo-nitzschia* events in the future.

Despite our limited understanding of the geochemical and physical drivers leading to DA production in the marine environment, it is nevertheless apparent that some species of *Pseudo-nitzschia* are more commonly toxic than others.[29,30] As such, reliable, rapid identification of *Pseudo-nitzschia* species is an area of substantial interest for the improved monitoring of DA-producing HABs. Classical assignment of *Pseudo-nitzschia* species identity has been performed using electron microscopy as a "gold standard" approach for describing the unique features of diatom silica frustules in order to differentiate one species from another.[29] Unsurprisingly, this approach is time and resource intensive, especially in complex samples potentially containing multiple species. Molecular probes have proven to be especially useful for the routine, rapid identification of dominant *Pseudo-nitzschia* species from sample filters using a "sandwich hybridization" approach.[31] However, new molecular probes need to be developed to detect additional species, and the approach lacks specificity and precision in some cases.

Meanwhile, high-throughput sequencing of 18S ribosomal regions has yielded enormous insight into the diversity of diatoms in the world's oceans, yielding genus-level identification and relative abundance information comparable to what can be achieved by light microscopy.[32] Despite the clear advantages afforded by sequencing 18S ribosomal regions, the approach is much better suited for genus-level taxonomic assignment and lacks sufficient resolution at the level of individual species. To address these limits implicit in 18S sequencing information, the field has begun to investigate new molecular "barcoding" regions for the species-level identification of *Pseduo-ntizschia* isolates. One study by Lim et al demonstrated that the internal transcribed spacer 2 (ITS2) region can reliably identify and delineate species of *Pseudo-nitzschia* in a manner that is wholly consistent with classical species identification by electron microscopy.[33] We envision that a combination of high-resolution species identification via ITS2

sequencing together with detection of our newly discovered DA biosynthesis (*dab*) genes (described in Chapters 2 and 3) could form a robust basis for the molecular characterization of toxic *Pseudo-nitzschia* HABs.

To demonstrate the potential of novel molecular approaches to inform *Pseudo-nitzschia* HAB monitoring, we generated a robust molecular dataset to complement weekly pier-based monitoring and sampling conducted during the major 2015 West Coast *P. australis* HAB event. To better characterize the 2015 bloom in Monterey Bay, we extracted total RNA from weekly phytoplankton net tow sample filters collected from the Monterey Wharf II (MWII) pier at the southern mouth of Monterey Bay. This extracted RNA was then used to generate cDNA for 18SV4 and novel *Pseudo-nitzschia*-targeted ITS2 amplicon sequencing. High-throughput sequencing of 18SV4 revealed a major *Pseudo-nitzschia* bloom event emerging in mid-April, consistent with both microscopy-based enumeration of *Pseudo-nitzschia* at MWII and previous studies on the bloom event in the larger bay. Novel ITS2 sequencing confirmed the dominance of *P. australis* throughout the bloom event while also yielding additional insight into other toxic and non-toxic species present before the monospecific bloom, some of which had been identified previously in other parts of the bay. We also implemented polyA-enriched RNA-sequencing to characterize *dab* gene transcription at the molecular level. Following *de novo* assembly of metatranscriptomes, we were able to identify *dab* transcripts expressed by *P. australis*, *P. multiseries* and *P. seriata* based on sequence homology to known *dab* sequences from all three species.[34,35] As a further verification, all three species were also identified in the ITS2 sequencing libraries on the exact dates the respective *dab* sequences were detected in the metatranscriptomics datasets. Notably, presence of *dab* transcripts in the metatranscriptomics data neatly coincides with detection of particulate DA from pier samples, suggesting that

monitoring for *dab* transcripts may serve as a valuable proxy for detecting active DA production. Finally, we implemented weighted gene correlation network analysis (WGCNA) to place *P. australis dab* transcription in the context of larger trends in relative gene expression by the nearly monospecific bloom. This gene-networks approach has yielded valuable preliminary insight into the shifting physiology of a *Pseudo-nitzschia* HAB throughout its life history. Taken as a whole, our study represents the most robust molecular description of a major marine HAB event described to date, and our molecular dataset thoroughly demonstrates the power of future environmental molecular approaches for describing DA-producing *Pseudo-nitzschia* blooms and HABs in general.

## 4.2 Works Cited for the Chapter Introduction


1. D.M. Karl. A sea of change: biogeochemical variability in the North Pacific Subtropical Gyre. *Ecosystems*. **2(3)**, 181–214 (1999)

2. D. Karl, R. Bidigare, R. Letelier. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: the domain shift hypothesis. *Deep Sea Res. Part II Top. Stud. Oceanogr*. **48(8–9)**, 1449–1470 (2001)

3. M-E. Carr, E. J. Kearns. Production regimes in four Eastern Boundary Current systems. *Deep Sea Res. Part II Top. Stud. Oceanogr*. **50(22–26)**, 3199–3221 (2003)

4. L. Z. Allen, E. E. Allen, J. H. Badger, J. P. McCrow, I. T. Paulsen, L. D. Elbourne, M. Thiagarajan, D. B. Rusch, K. H. Nealson, S. J. Williamson, J. C. Venter, A. E. Allen. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *ISME J*. **6(7)**, 1403–1414 (2012)

5. B. M. Stephens, S. D. Wankel, J. M. Beman, A. J. Rabines, A. E. Allen, L. I. Aluwihare. Euphotic zone nitrification in the California Current Ecosystem. *Limnol. Oceanogr.* **65(4)**, 790–806 (2020)

6. S. A. Kranz, S. Wang, T. B. Kelly, M. R. Stukel, R. Goericke, M. R. Landry, N. Cassar. Lagrangian studies of marine production: a multimethod assessment of productivity relationships in the California Current Ecosystem upwelling region. *J. Geophys. Res. Ocean* **125(6)**, e2019JC015984 (2020)

7. B. A. Black, I. D. Schroeder, W. J. Sydeman, S. J. Bograd, B. K. Wells, F. B. Schwing. Winter and summer upwelling modes and their biological importance in the California Current Ecosystem. *Glob. Chang. Biol.* **17(8)**, 2536–2545 (2011)

8. A. Gutierrez-Rodriguez, M. R. Stukel, A. Lopes dos Santos, T. Biard, R. Scharek, D. Vaulot, M. R. Landry, F. Not. High contribution of Rhizaria (Radiolaria) to vertical export in the California Current Ecosystem revealed by DNA metabarcoding. *ISME J*. **13(4)**, 964–976 (2019)

9. D. G. Ainley, L. B. Spear, C. T. Tynan, J. A. Barth, S. D. Pierce, R. Glenn Ford, T. J. Cowles. Physical and biological variables affecting seabird distributions during the upwelling season of the northern California Current. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **52(1–2)**, 123–143 (2005)

10. G. Pitcher, B. Jiménez, R. Kudela, B. Reguera. Harmful algal blooms in Eastern boundary upwelling systems: a GEOHAB core research project. *Oceanography* **30(1)**, 22–35 (2017)

11. A. J. Lewitus, R. A. Horner, D. A. Caron, E. Garcia-Mendoza, B. M. Hickey, M. Hunter, D. D. Huppert, R. M. Kudela, G. W. Langlois, J. L. Largier, E. J. Lessard, R. RaLonde, J. E. Jack Rensel, P. G. Strutton, V. L. Trainer, J. F. Tweddle. Harmful algal blooms along

the North American west coast region: history, trends, causes, and impacts. *Harmful Algae* **19**, 133–159 (2012)

12. L. Fritz, M. A. Quilliam, J. L. C. Wright, A. M. Beale, T. M. Work. An outbreak of domoic acid poisoning attributed to the pennate diatom *Pseudo-nitzschia australis*. *J. Phycol.* **28(4)**, 439–442 (1992)

13. T. M. Work, B. Barr, A. M. Beale, L. Fritz, M. A. Quilliam, J. L. C. Wright. Epidemiology of domoic acid poisoning in brown pelicans (*Pelecanus occidentalis*) and Brandt's cormorants (*Phalacrocorax penicillatus*) in California. *J. Zoo Wildl. Med.* **24(1)**, 54–62 (1993)

14. J. C. Wekell, E. J. Gauglitz Jr., H. H. Barnett, C. L. Hatfield, M. Eklund. The occurrence of domoic acid in razor clams (*Siliqua patula*), Dungeness crab (*Cancer magister*), and anchovies (*Engraulis mordax*). *J. Shellfish Res.* **13**, 587–593 (1994)

15. G. A. Fryxell, M. C. Villac, L. P. Shapiro LP. The occurrence of the toxic diatom genus *Pseudo-nitzschia* (Bacillariophyceae) on the West Coast of the USA, 1920–1996: a review. *Phycologia* **36(6)**, 419–437 (1997)

16. C. A. Scholin, F. Gulland, G. J. Doucette, S. Benson, M. Busman, F. P. Chavez, J. Cordaro, R. DeLong, A. De Vogelaere, J. Harvey, M. Haulena, K. Lefebvre, T. Lipscomb, S. Loscutoff, L. J. Lowenstine, R. Marin III, P. E. Miller, W. A. McLellan, P. D. R. Moeller, C. L. Powell, T. Rowles, P. Silvagni, M. Silver, T. Spraker, V. Trainer, F. M. Van Dolah. Mortality of sea lions along the central California coast linked to a toxic diatom bloom. *Nature* **403(6765)**, 80–84 (2000)

17. N. A. Adams, A. M. Lesoing, V. L. Trainer. Environmental influences on domoic acid accumulation in razor clams on the Washington coast. *J. Shellfish Res*. **19,** 1007–1015 (2000)

18. J. Smith, P. Connell, R. H. Evans, A. G. Gellene, M. D. A. Howard, B. H. Jones, S. Kaveggia, L. Palmer, A. Schnetzer, B. N. Seegers, E. L. Seubert, A. O. Tatters, D. A. Caron. A decade and a half of *Pseudo-nitzschia spp*. and domoic acid along the coast of southern California. *Harmful Algae* **79**, 87–104 (2018)

19. C. G. Trick, V. L. Trainer, W. P. Cochlan, M. L. Wells, B. F. Beall. The successional formation and release of domoic acid in a *Pseudo-nitzschia* bloom in the Juan de Fuca Eddy: a drifter study. *Harmful Algae* **79**, 105–114 (2018)

20. C. R. Anderson, M. A. Brzezinski, L. Washburn, R. Kudela. Circulation and environmental conditions during a toxigenic *Pseudo-nitzschia australis* bloom in the Santa Barbara Channel, California. *Mar. Ecol. Prog. Ser*. **327**, 119–133 (2006)

21. C. R. Anderson, D. A. Siegel, M. A. Brzezinski, N. Guillocheau. Controls on temporal patterns in phytoplankton community structure in the Santa Barbara channel, California. *J*

*Geophys Res Ocean*. **113(4)**, 1–16 (2008)

22. L. K. Rosenfeld, F. B. Schwing, N. Garfield, D. E. Tracy. Bifurcated flow from an upwelling center: a cold water source for Monterey Bay. *Cont. Shelf Res.* **14(9)**, 931–964 (1994)

23. W. M. Graham, J. L. Largier. Upwelling shadows as nearshore retention sites: the example of northern Monterey Bay. *Cont. Shelf Res.* **17(5)**, 509–532 (1997)

24. J. P. Ryan, M. A. McManus, R. M. Kudela, M. Lara Artigas, J. G. Bellingham, F. P. Chavez, G. Doucette, D. Foley, M. Godin, J. B. J. Harvey, R. Marin, M. Messié, C. Mikulski, T. Pennington, F. Py, K. Rajan, I. Shulman, Z. Wang, Y. Zhang. Boundary influences on HAB phytoplankton ecology in a stratification-enhanced upwelling shadow. *Deep Sea. Res. Part II Top. Stud. Oceanogr.* **101**, 63–79 (2014)

25. J. P. Ryan, R. M. Kudela, J. M. Birch, M. Blum, H. A. Bowers, F. P. Chavez, G. J. Doucette, K. Hayashi, R. Marin, C. M. Mikulski, J. T. Pennington, C. A. Scholin, G. J. Smith, A. Woods, Y. Zhang. Causality of an extreme harmful algal bloom in Monterey Bay, California, during the 2014–2016 northeast Pacific warm anomaly. *Geophys. Res. Lett.* **44(11)**, 5571–5579 (2017)

26. H. A. Bowers, J. P. Ryan, K. Hayashi, A. L. Woods, R. Marin, G. J. Smith, K. A. Hubbard, G. J. Doucette, C. M. Mikulski, A. G. Gellene, Y. Zhang, R. M. Kudela, D. A. Caron, J. M. Birch, C. A. Scholin. Diversity and toxicity of *Pseudo-nitzschia* species in Monterey Bay: perspectives from targeted and adaptive sampling. *Harmful Algae* **78**, 129–141 (2018)

27. Y. Pan, D. V. Subba Rao, K. H. Mann, R. G. Brown, R. Pocklington. Effects of silicate limitation on production of domoic acid, a neurotoxin, by the diatom *Pseudo-nitzschia multiseries*. I. Batch culture studies. *Mar. Ecol. Prog. Ser.* **131**, 225–233 (1996)

28. J. Fehling, K. Davidson, C. J. Bolch, S. S. Bates. Growth and domoic acid production by *Pseudo-nitzschia seriata* (Bacillariophyceae) under phosphate and silicate limitation. *J Phycol*. **40(4)**, 674–683 (2004)

29. A. Lelong, H. Hégaret, P. Soudant, S. S. Bates, *Pseudo-nitzschia* (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. *Phycologia* **51**, 168–216 (2012)

30. S. S. Bates, K. A. Hubbard, N. Lundholm, M. Montresor, C. P. Leaw, *Pseudo-nitzschia*, *Nitzschia*, and domoic acid: new research since 2011. *Harmful Algae* **79**, 3–43 (2018)

31. H. A. Bowers, R. Marin, J. M. Birch, C. A. Scholin. Sandwich hybridization probes for the detection of *Pseudo-nitzschia* (Bacillariophyceae) species: an update to existing probes and a description of new probes. *Harmful Algae* **70**, 37–51 (2017)

32. S. Malviya, E. Scalco, S. Audic, F. Vincent, A. Veluchamy, J. Poulain, P. Wincker, D. Iudicone, C. De Vargas, L. Bittner, A. Zingone, C. Bowler. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci*. **113(11)**, E1516–E1525 (2016)

33. H. C. Lim, S. N. Tan, S. T. Teng, N. Lundholm, E. Orive, H. David, S. Quijano-Scheggia, S. C. Y. Leong, M. Wolf, S. S. Bates, P. T. Lim, C. P. Leaw. Phylogeny and species delineation in the marine diatom *Pseudo-nitzschia* (Bacillariophyta) using cox1, LSU, and ITS2 rRNA genes: a perspective in character evolution. *J. Phycol.* **54(2)**, 234–248 (2018)

34. J. K. Brunson, S. M. K. McKinnie, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. Smith, D. A. Hutchins, A. E. Allen, B. S. Moore. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018)

35. S. Harðardóttir, S. Wohlrab, D. M. Hjort, B. Krock, T. G. Nielsen, U. John, N. Lundholm. Transcriptomic responses to grazing reveal the metabolic pathway leading to the biosynthesis of domoic acid and highlight different defense strategies in diatoms. *BMC Mol. Biol.* **20**, 7 (2019)

## 4.3 Draft manuscript for "Metatranscriptomic sequencing and barcoding of a major harmful algal bloom event in Monterey Bay, California"

### 4.3.1 Main Text

#### 4.3.1.1 Abstract

In 2015, the Northeast Pacific was host to the largest harmful algal bloom (HAB) ever recorded. Dominated by the toxic diatom *Pseudo-nitzschia australis*, this bloom produced extremely high levels of the neurotoxin domoic acid (DA). Monterey Bay, California was one of the hotspots for this toxic event, with record-breaking levels of DA recorded throughout the bay. We have leveraged samples from weekly plankton net sampling and monitoring from Monterey Municipal Wharf II at the southern end of the bay to create metatranscriptomic and barcoding datasets in order to characterize the shifting microbial community and profile the transcriptional activity of the *P. australis* bloom. By implementing ITS2 sequencing, we are able to resolve *Pseudo-nitzschia* community members at the species level, revealing that the 2015 event implicated at least three toxic species, with *P. australis* as the most abundant species and *P. multiseries* and *P. seriata* as minor contributors to toxicity. We are also able to identify active transcription of domoic acid biosynthesis (*dab*) genes from all three species during 2015, with one additional *dab* contig of unknown origin. Measurement of *dab* transcription coincides with detection of particulate DA in all relevant samples. Furthermore, we implemented gene-network analysis revealing clustering of functionally related transcripts providing insight into diatom physiology during different phases of the HAB even. Taken together, these data are an important contribution to HAB monitoring programs and will hopefully inform future predictive efforts in HAB mitigation.

#### 4.3.1.2 Introduction

Harmful algal blooms (HABs) pose a significant threat to human health and the environment. Many HABs also produce small molecule toxins, such as saxitoxin or domoic acid, which bioaccumulate in seafood.[1] Because these toxins have severe human health impacts, fisheries must often close operations in response to HAB events. For these reasons, successful and robust HAB monitoring is of particular interest. HAB monitoring has classically been accomplished via microscopy-based detection of HAB species and chemical detection of toxins with satellite-based chlorophyll measurements lending additional information. These data, taken together with macronutrient data and physical oceanographic parameters, inform predictive monitoring approaches.[2,3] While these models have proven extremely useful for bloom prediction, they yield a significant false-positive results. A more complete understanding of toxin producer physiology and the genetic basis of HAB toxin biosynthesis would improve forecasting and enable rapid molecular monitoring approaches.

Domoic acid (DA) is one of the main HAB toxins threatening the North American West Coast. Blooms of the DA-producing diatom genus *Pseudo-nitzschia* develop along the coastline with near annual frequency.[4] In 2015, a bloom of *Pseudo-nitzschia australis* spanned the Gulf of Alaska to Point Conception, California, making this the largest HAB ever recorded in the Northeast Pacific Ocean. The toxic *P. australis*, a species common to the central California coast, expanded its range northward due to the warm northeast pacific anomaly.[5] In Monterey Bay, California, the bloom of *P. australis* produced extremely high levels of DA, prompting further study into the oceanographic mechanisms fueling the local event.[6,7] Monterey Bay is a well-studied body of water that experiences frequent phytoplankton blooms driven by the

seasonal upwelling activity occurring outside the bay at Point Año Nuevo to the north and Point Sur to the south (Figure 4.1).[8–10] In 2015, the especially toxic local *P. australis* HAB event coincided with historically low silicate concentrations recorded throughout Monterey Bay.[6] Silicate limitation, among other abiotic and biotic factors, is a common method for inducing DA production in the laboratory.[11–13] However, the specific physiological mechanism by which different stressors induce DA production remains unclear.

Recently, we elucidated the DA biosynthetic pathway in *Pseudo-nitzschia multiseries*.[14] By using an RNA-sequencing approach to guide comparative expression analysis, we identified key DA biosynthetic (*dab*) genes that were upregulated under phosphate limitation and increased $pCO_2$, two conditions which also increase DA production in laboratory cultures.[15] The enzymes encoded by *dabA*, *dabC* and *dabD* were demonstrated to perform the key biosynthetic steps to make DA, namely glutamate *N*-prenylation, kainoid ring cyclization, and side chain hydroxylation, respectively. In addition to *P. multiseries*, these DA biosynthetic (*dab*) genes have so far been identified in *P. multistriata*, *P. australis*, and *P. seriata*.[14,16]

Monitoring for *Pseudo-nitzschia* and DA, among other HAB species and toxins, has been accomplished along the California coast by The California Harmful Algal Bloom Monitoring and Alert Program (CalHABMAP) through weekly pier sampling.[17] Microscopy-based monitoring for harmful algae is time-intensive and making species-level identifications using light microscopy alone is challenging. Environmental genetics, on the other hand, can enable rapid identification of HAB species at significantly improved precision, and has been deployed to great success for routine monitoring of *Pseudo-nitzschia* in Monterey Bay.[18] While current molecular approaches for species-level identification of *Pseudo-nitzschia* has enabled highly informative descriptions of HAB events, next-generation sequencing of amplicon libraries has been

demonstrated to provide rich information regarding phytoplankton and microbial diversity around the world and represents a crucial next step in describing the species composition of *Pseudo-nitzschia* blooms in a rapid, high-throughput manner.[19] Meanwhile, detection and identification of HAB toxins is routinely accomplished via liquid chromatography-mass spectrometry (LC-MS) or by enzyme-linked immunosorbent assays (ELISAs), two approaches that have been implemented by CalHABMAP.

Although LC-MS is arguably the most accurate method of toxin detection and quantification, analysis of samples is also time-intensive and both LC-MS and ELISA require expensive laboratory equipment. On the other hand, the detection of toxin biosynthetic genes has been proposed as a possible route for monitoring and forecasting HAB toxicity, and the cyanobacterial HAB community is beginning to implement molecular monitoring guided by the biosynthetic genes for toxins such as microcystin and saxitoxin.[20] Detection of *dab* transcripts has not yet been developed for the monitoring of DA-producing *Pseudo-nitzschia* HABs, and at present it is unclear how *dab* transcription is linked to DA production in the marine environment.

In this study, we combine 18S and novel ITS2 amplicon sequencing with polyA-enriched RNA-sequencing to characterize toxic *Pseudo-nitzschia* species and *dab* gene transcription at the molecular level. These new metabarcoding and metatranscriptomic resources were generated from weekly phytoplankton net-tow samples collected from Monterey Municipal Wharf II (MWII) in Monterey Bay, California throughout the 2015 bloom event (Fig. 4.1). Analysis of these datasets reveals the phytoplankton succession leading up to the HAB event as community compositions transition from non-toxic diatom species to a near-monospecific bloom of DA producing *Pseudo-nitzschia australis*. Our metatranscriptomics analysis is also able to link discrete *dab* transcripts back to their producing species, combining ITS2 species identification

with literature *dab* sequences. Finally, we use weighted gene correlation network analysis (WGCNA) to identify suites of potentially co-regulated transcripts to contextualize *dab* expression and provide physiological insight into *Pseudo-nitzschia* HAB inception, progression, and demise. This study represents an important advance in the monitoring of eukaryotic HABs and demonstrates feasibility for future environmental genetic approaches to describe DA events.

#### 4.3.1.3 Results and Discussion

Amplicon Sequencing Reveals the Spring Phytoplankton Succession

Temperature and nitrate measurements taken from MWII in 2015 suggest cold, nutrient rich upwelling beginning in late March and early April (Supplemental Information, Fig. S4.1). This observation of upwelling is supported by previous geophysical and ecosystem dynamics studies on the oceanographic conditions fueling the 2015 West Coast HAB in Monterey Bay, CA.[6] Amplicon sequencing of the 18S-V4 region reveals a stark transition from a zooplankton-dominant ecosystem to a diatom-dominant (Bacillariophyta) community in early to mid-March following commencement of spring upwelling, an observation further reflected by increased chlorophyll concentrations during this time period (Fig. 4.2A & Fig. 4.2B). Diatoms remained the dominant members of the microbial community until mid-July when their relative abundance briefly decreased. The diatom-dominated phytoplankton population returned once more in August before its ultimate demise at the end of September, giving way to copepods and other grazers in the autumn months.

Analysis of diatom classes and genera within phylum Bacillariophyta reveals that the early spring bloom was dominated by non-toxic, centric diatoms (class Mediophyceae) prior to transitioning to a community comprised almost entirely of pennate (class Bacillariophyceae)

diatoms (Fig. 4.2B & Fig. 4.2C). Further analysis of the genera which comprise Mediophyceae and Bacillariophyceae in our samples revealed that *Chaetoceros* and *Pseudo-nitzschia* were the dominant centric and pennate diatoms, respectively (Supplemental Information, Fig. S4.2 & Fig. S4.3). On April 22nd, when the highest MWII chlorophyll levels for the year were measured, the diatom population was split between comparable relative populations of *Chaetoceros* and *Pseudo-nitzschia*. Diatoms belonging to the *Pseudo-nitzschia* genus dominated the rest of the bloom season through early autumn. These observations are further supported by sequencing of 16S chloroplast sequences (Supplemental Information, Fig. S4.4). Alpha-diversity analysis revealed that diatoms indeed experienced a reduction in diversity during the bloom event. This significant reduction in diversity was primarily driven by pennate diatoms resulting from *Pseudo-nitzschia* dominance, with centric diatom diversity appearing largely unaffected throughout the *P. australis* bloom (Fig. 4.2D & Fig. 4.2E).

ITS2 Sequencing reveals *Pseudo-nitzschia* species diversity prior to *P. australis* dominance

While we can confidently resolve the spring phytoplankton succession at the genus level to correctly identify the presence of a large *Pseudo-nitzschia* bloom, resolution at the species level was low for both 18S and 16S sequencing. To fully characterize which *Pseudo-nitzschia* species were present throughout the year, we implemented amplicon sequencing of the ITS2 region. The ITS2 region has been demonstrated to offer good species-level resolution consistent with classical methods of species delineation, such as electron microscopy.[21] Our metabarcoding experiments verified *P. australis* as the dominant species during the bloom event, as reported previously.[5,7] Much of the early spring phytoplankton succession dominated by *Chaetoceros* appears to also include low levels of various *Pseudo-nitzschia* species such as *P. fraudulenta*, *P.*

*cuspidata, P. americana*, and *P. pungens*. All these species, barring *P. americana*, have been demonstrated to produce low-to-modest amounts of DA in the lab, and their presence here coincides with very low levels of DA production (~7ng/L, March 11th, Fig. 3A-C).[22,23]

Beginning in late April, the community shifts to near-complete dominance of *P. australis*, the species which was previously identified to be the main driver of the bloom in Monterey Bay and along the entire NA West Coast.[5–7] However, we also identify a substantial subpopulation of potentially toxic *P. multiseries* present throughout the month of April in conjunction with modestly increased DA concentrations (~22 ng/L, April 15th, Fig. 3A-C). Like *P. australis*, *P. multiseries* is also notable for its high levels of DA production compared to most other species of *Pseudo-nitzschia*. Our detection of the *P. multiseries* subpopulation is further corroborated by identification of *P. multiseries* during a similar time period at the northern point of Monterey Bay at Santa Cruz Wharf by sandwich hybridization assays and cellular isolation.[7] A third subpopulation of potentially high-toxicity *Pseudo-nitzschia*, *P. seriata*, was also observed in early to mid-May, albeit at a much lower percent abundance than *P. australis*. These data suggest that several species may have contributed to DA production throughout different phases of the bloom season, and that a sizable population of *P. multiseries* may have been a significant contributor to initial DA production at MWII prior to the emergence of *P. australis* as the dominant species.

Expression of *dab* genes reflects toxin production by the *Pseudo-nitzschia* HAB

To detect and measure the expression of *dab* genes during the 2015 bloom event, we performed RNA sequencing on polyA-enriched RNA extracted from sample filters taken from weekly phytoplankton net tow monitoring. Following assembly and annotation, we were able to

detect *dab* transcripts with near identical sequence to known *dab* genes from *P. australis, P. multiseries*, and *P. seriata*, three species identified by ITS2 sequencing during the bloom months.[14,16] Specifically, we were able to identify the full suite of core DA biosynthetic genes (*dabA, dabC,* and *dabD*) as expressed by the dominant bloom species *P. australis*. These *P. australis dab* transcripts could be detected throughout the entire bloom event (April 15$^{th}$ – September 30$^{th}$). Our assemblies also yielded *P. multiseries* transcripts for *dabA* and *dabC*, both of which displayed expression profiles coinciding with the detection of the species through ITS2 amplicon sequencing (April 1$^{st}$ – mid May). Two additional *dab* genes were identified from the *de novo* assembly: one could be confidently assigned to *P. seriata* (*dabA*, April 22$^{nd}$ – June 10$^{th}$) and the other of unknown origin (*dabC*, March 11$^{th}$ - April 8$^{th}$). The unknown *dabC* transcript was present early in the spring phytoplankton succession, coincident with the presence of various low-toxicity *Pseudo-nitzschia spp.* and low-level detection of pDA. In fact, active *dabA* and *dabC* gene transcription co-occurs with detection of pDA throughout most of the spring and summer months (Table 4.1, Fig. 4.4).

Phytoplankton sampling and pDA measurements taken from MWII, combined with our molecular barcoding and *dab* transcription measurements, suggests two phases to the persistent *P. australis* bloom as observed from the wharf. The first phase involves proliferation of the bloom and steady build-up of cellular DA quotas during April and May. Cellular DA (cDA), calculated by dividing pDA by *Pseudo-nitzschia* cells per liter, steadily increases during April and May coincident with consistent *dab* gene expression by the *P. australis* and *P. multiseries* subpopulations during these months (Fig. 4A). The second phase of the bloom occurs after the July disappearance of *Pseudo-nitzschia* and subsequent decrease in pDA and cDA (Fig. 4.3A-B & Fig. 4.4B). Beginning in late July with the sudden resurgence of *Pseudo-nitzschia* cells and

DA, periodic increases in cDA are mirrored by similarly increased abundance of *P. australis* *dabA* and *dabC* transcripts, suggesting that a combination of intermittent increased *P. australis* cell density and high levels of *dab* gene expression combine to periodic elevated DA production during the late phase of the *Pseudo-nitzschia* HAB (Fig. 4.3A-B & Fig. 4.4B). In both early (highlighted in blue) and late (highlighted in orange) phases of the bloom, increases in cDA concentrations are neatly coupled with spikes in transcript abundance for *dabA* and *dabC*, the two diagnostic DA biosynthetic genes (Fig. 4.4B-D).[14]

An Upwelling Model for the Resurgence of *P. australis* in Late July

Mid-to-late summer resurgence of the bloom at MWII could be explained by resumed upwelling from outside the bay, bringing in fresh nutrients to refuel the bloom event. The initial decline in *Pseudo-nitzschia* cell counts in early to mid-July coincides with a period of rapid freshening of the near-surface water column in the middle of the bay, as reported previously.[6] This is likely due to an influx of lower-salinity water from the California Current as it meanders shoreward with the relaxation of upwelling favorable winds and therefore upwelling during this period, a commonly described seasonal phenomenon in Monterey Bay.[8,9] Indeed, relaxed upwelling can be observed from satellite-measured surface temperature plots, showing no substantially decreased sea surface temperature at the main upwelling locations of Point Año Nuevo and Point Sur on July 7$^{th}$ and July 11$^{th}$ (Fig. 4.5A & Fig. 4.5B).

The return of the bloom to MWII in late-July and early-August, an event that has not been discussed in previous analyses of the 2015 bloom in Monterey Bay, could be explained by resumed upwelling from outside the bay. In contrast to relaxation of upwelling in early July, resumed upwelling can be detected in the satellite sea surface temperature scans as shown by

masses of cold surface water located at Point Año Nuevo and Point Sur on July 24th and July 27th (Fig. 4.5C-D). On July 27th, a cold water upwelling filament can be observed extending towards the mouth of the bay from Point Año Nuevo to the north, a phenomenon that nicely coincides with a brief, sharp increase in subsurface salinity at the end of July recorded previously.[6] This cold upwelling filament is also accompanied by a stark increase in chlorophyll-a concentration on July 27th as measured by the satellite Visible Infrared Imaging Radiometer Suite (VIIRS), indicative of active phytoplankton growth (Fig. 4.5E). Substantially elevated chlorophyll-a concentrations measured at Pt. Año Nuevo, along the upwelling filament, and to a lesser extent at Pt. Sur. Additional, weaker upwelling can also be observed on August 14th at Point Año Nuevo and Point Sur (Fig. 4.5F).

Relative expression profiles of *P. australis* transcripts indicates shifting bloom physiology

In order to describe *dab* gene expression in the context of *P. australis* physiology and the overall progression of the bloom event, we employed weighted gene correlation network analysis (WGCNA) on a highly-expressed subset of the *de-novo* assembled *P. australis* HAB metatranscriptome to define modules of transcripts with similar expression profiles.[24] Out of 14873 contigs with *P. australis* species annotation, 437 contigs met statistical cut-offs for WGCNA, wherein 80% of sequencing libraries contained at least 10 reads mapping to the contigs of interest. Read counts were normalized to the sum of total reads mapping to *P. australis*-annotated contigs in each library to account for both library size and *P. australis* population size to approximate relative transcript abundance. From this analysis, we were able to identify seven modules of similarly expressed transcripts (modules 1-7), including one module of

contigs with low overall variance explained by the model (module 0) (Fig. 4.6, Supplemental Information S4.5 & S4.6).

Both *dabA* and *dabD* were found in the smallest module (module 7), together with two genes encoding key isoprenoid biosynthesis enzymes: 4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS) and isopentenyl-diphosphate delta-isomerase (IDI) (Fig. 4.7A). Both proteins are directly implicated in biosynthesis of the DA precursor geranyl pyrophosphate and are known to be co-expressed with the *dab* gene products.[16] Besides these key DA biosynthetic reactions, most of the other contigs in the "black" module appear unrelated to toxin production and are not known to indicate a specific diatom physiological state. Two additional isoprenoid biosynthesis transcripts, 1-deoxy-D-xylulose-5-phosphate synthase (DXS) and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR), are found in modules 2 and 4, respectively. All four isoprenoid biosynthesis genes included in this analysis are implicated in the chloroplastic, non-mevalonate pathway thought to feed directly into DA biosynthesis.[25]

The remaining key DA biosynthetic gene *dabC* was found in module 4, together with HDR but separate from *dabA* and *dabD* (Fig. 4.7A). While *dabC* is co-expressed with *dabA* and *dabD* in *P. multiseries* and *P. seriata* cultures, it is conceivable that different steps of DA biosynthesis might be differentially regulated.[14,16] Module 4 also displays strong functional clustering of photosynthetic machinery. Diatoms encode many copies of chlorophyll binding proteins and light-harvesting protein complexes.[26] Out of 19 contigs annotated as chlorophyll binding proteins in the WGCNA dataset, 14 were present in module 4 (Fig. 4.7B). Other photosynthesis and carbon fixation-related transcripts found in the module include the iron-sulfur subunit of the cytochrome b6f complex, photosystem I reaction center subunit VIII, photosystem

II stabilizing-protein PsbU, ferredoxin-$NADP^+$ reductase, transketolase, and fructose-1,6-bisphosphatase.

The largest WGCNA module, module 1, includes transcripts that display an increased relative abundance at the beginning of the bloom and are among the first *P. australis* transcripts detectable. Many of these transcripts are related to active cell growth and proliferation. Out of 16 ribosomal subunit transcripts in the WGCNA dataset, 8 were found in module 1 (Fig. 4.7B). Five ribosomal subunits included in the dataset are predicted to reside in the chloroplast, all of which were found in module 1, potentially indicating increased chloroplast ribosome biogenesis early in the bloom progression. Transcripts for F-type ATP synthase subunits were also primarily found in module 1, suggesting increased ATP production in conjunction with a contig encoding a mitochondrial ATP/ADP carrier protein, also found in the module. One V-ATPase subunit was found in this module along with clathrin, two proteins that play an important role in diatom cell division and cell wall silicification.[27,28] However, most of the V-type ATPase subunits in the analysis were found in module 2, perhaps implying dynamic regulation of V-type ATPase subunits for different cellular functions throughout different phases of the bloom.

Modules 2 and 5 include transcripts with higher relative abundance at the end of the bloom season, including some transcripts that may be linked to general physiological stress or cell death. Subunits of the proteasome, required for protein degradation and turnover, are enriched in these two modules. Out of 13 proteasome 20S/26S subunits in the WGCNA analysis, 8 were found in module 2 or 5 (Fig. 4.7B). Increased protein turnover versus new protein translation has been described previously in diatom systems as a side effect of decreased growth, cell division and proliferation, and may be associated with nutrient and temperature stress.[29] In general, transcripts associated with protein turnover are enriched in these two modules, with 26

of 46 transcripts of relevant KOG class annotation ("Posttranslational modification, Protein turnover, Chaperones") found in modules 2 and 5 (Fig. 4.7B). Additional transcripts from modules 2 and 5 associated with protein turnover and related stress response include ubiquitin ligases, aspartyl protease, glutaredoxin and thioredoxin, cyclophilin, and chaperones DnaJ, GroES, and GroEL.

The second-smallest module, module 6, includes select transcripts involved in diatom response to nitrogen starvation. Urea transporters and specific ammonia transporters are known to be upregulated in diatoms under nitrogen limiting conditions.[30] Both a urea transporter and the *P. australis* homolog of AMT1_1, an ammonia transporter from *P. tricornutum* induced in low-N conditions, can be found in this module. Nitrogen limitation in diatoms is also coupled to an increase in fatty acid biosynthesis and accumulation. Three fatty acid desaturases, including a Δ9-desaturase upregulated in *P. tricornutum* lacking functional nitrate reductase, were also observed in this module.[31] The remaining module, module 3, does not display obvious functional clustering at present and lacks key diatom markers of physiological states.

#### 4.3.1.4 Discussion and Conclusions

Our study provides a robust molecular framework for understanding the progression, toxicity, and physiology of the 2015 *P. australis* HAB event in Monterey Bay, CA. Using 18S, 16S and ITS2 sequencing, we can visualize the transition from a *Chaetoceros*-dominated phytoplankton community to a nearly monospecific bloom of DA-producing *P. australis*. The specific oceanographic mechanisms underlying the transition between the two genera remains unclear. *Pseudo-nitzschia* has specific molecular adaptations allowing it to outcompete *Chaetoceros* in low-iron conditions, and widespread iron-limitation is well-documented along

the California coast, particularly in subsurface chlorophyll maxima.[32,33] The role of iron limitation in the development and toxicity of the 2015 HAB event remains undescribed. The eventual dominance of *P. australis* over other toxic and non-toxic *Pseudo-nitzschia* species present before the bloom, such as *P. multiseries* and *P. fraudulenta*, is also unexplained in terms of competitive advantages and species adaptations.

Detection of *dab* transcripts in the metatranscriptomics dataset supports a molecular monitoring approach to *Pseudo-nitzschia* HAB toxicity. Active *dab* transcription appears to be inextricably linked to accumulation of pDA in monitoring samples, and consistent detection of *dab* expression in multiple sampling timepoints may help predict severe bloom events. Metatranscriptomic sequencing of additional *Pseudo-nitzschia* bloom events, including more frequent sampling on the timescale of days rather than weeks, will be necessary to assess the potential for *dab* detection to inform bloom forecasting. Additionally, the precise nucleotide sequence of *dab* transcripts may lend insight into DA producing species. In this study, high sequence similarity of *de novo* assembled contigs to literature *dab* sequences allows us to assign toxin biosynthetic transcripts to discrete producing species. Given the paucity of *dab* transcripts currently identified, further sequencing of *dab* genes and transcripts will be required to fully explore this trend among all DA-producing *Pseudo-nitzschia* species and isolates.

Our metatranscriptomic approach has also enabled preliminary insight into the changing physiology of *Pseudo-nitzschia* HABs from inception to demise. From our analysis, we were able to describe trends in the relative expression of genes involved in active growth, photosynthesis, toxin production, and environmental stress. The identification of modules containing similarly expressed genes of comparable functions from our study was encouraging, especially given the nature of our MWII sampling approach, conducted weekly from a fixed

location on the shore. We expect that a more targeted, LaGrangian sampling of surface waters or subsurface chlorophyll maxima throughout a bloom event will yield improved functional clustering of gene expression by following discrete *Pseudo-nitzschia* subpopulations in the water column. Models for *Pseudo-nitzschia* HAB physiology will be improved further if LaGrangian sampling is conducted more frequently than the weekly timepoints employed by our study.

To our knowledge, this study represents the most comprehensive molecular description of any eukaryotic marine HAB event. Our work connects routine monitoring data, such as toxin measurements and phytoplankton cell counts, to molecular HAB species identification and toxin biosynthesis gene expression. The synthesis of this data provides new avenues for *Pseudo-nitzschia* HAB monitoring at the molecular level and will inform improved descriptions of HAB physiology as it relates to nutritional and physical conditions. The continued study of molecular physiology in HABs promises to lend essential insight into the nature of toxin production and bloom formation in the changing world ocean.

#### 4.3.1.5 Main Text Figures and Table

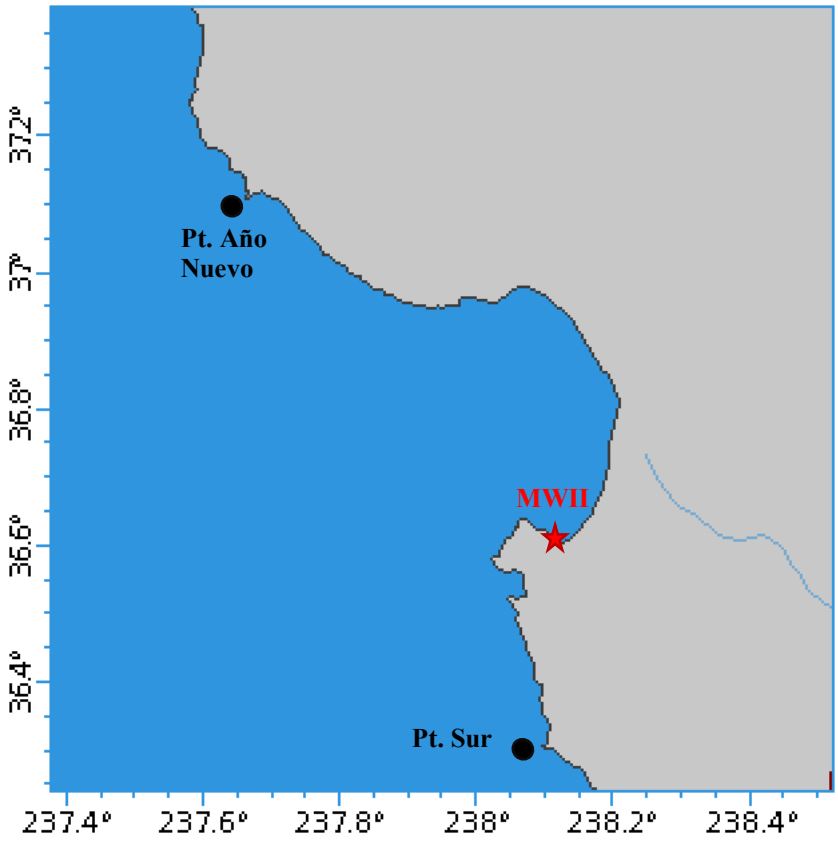



**Figure 4.1.** Monterey Bay, California. Sampling location at Monterey Wharf II (MWII) is marked together with the upwelling centers outside the bay, Pt. Año Nuevo and Pt. Sur.

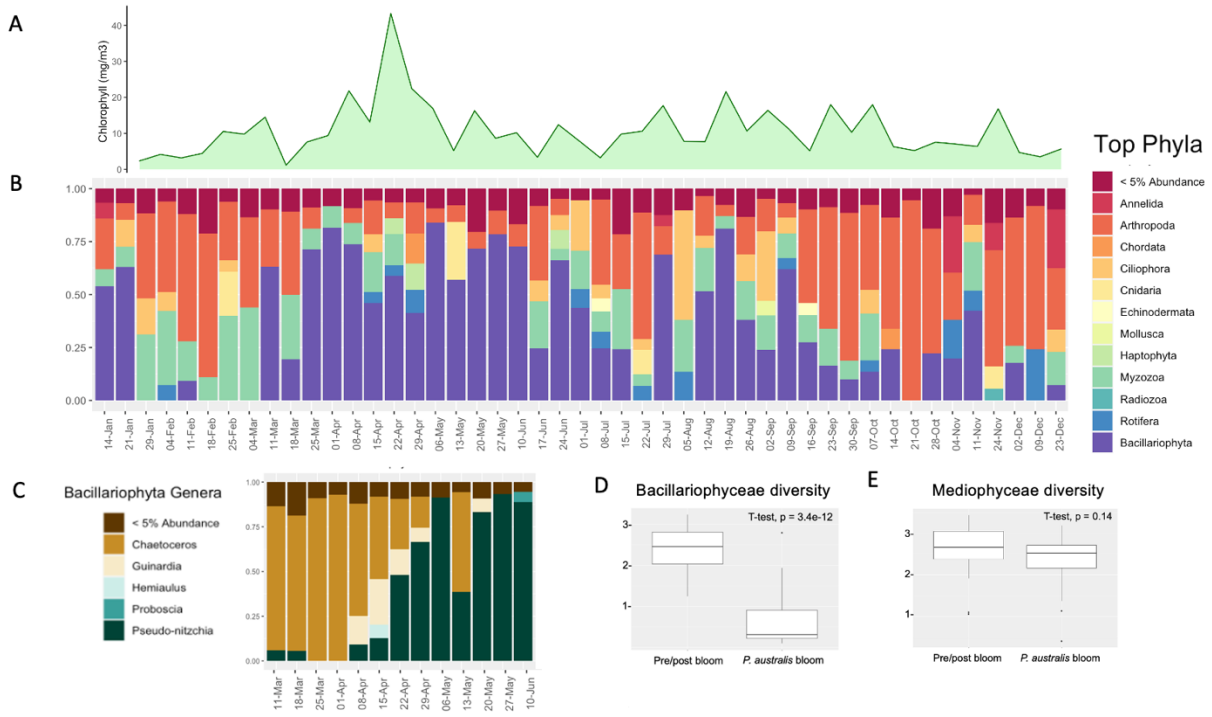



**Figure 4.2.** 18S-V4 Amplicon Sequencing of 2015 MWII Samples. (A) Chlorophyll A concentration (mg/m$^3$), (B) Relative abundance of phyla in 18S-V4 amplicon sequencing libraries, (C) Relative abundance of diatom (Bacillariophyta) genera during the spring phytoplankton succession, (D) Shannon alpha-diversity of pennate (Bacillariophyceae) diatoms comparing *P. australis* bloom samples (Apr 22$^{nd}$ – Sep 30$^{th}$) with non-bloom samples (rest of year), (E) Shannon alpha-diversity of centric (Mediophyceae) comparing bloom samples (Apr 22$^{nd}$ – Sep 30$^{th}$) with non-bloom samples (rest of year).

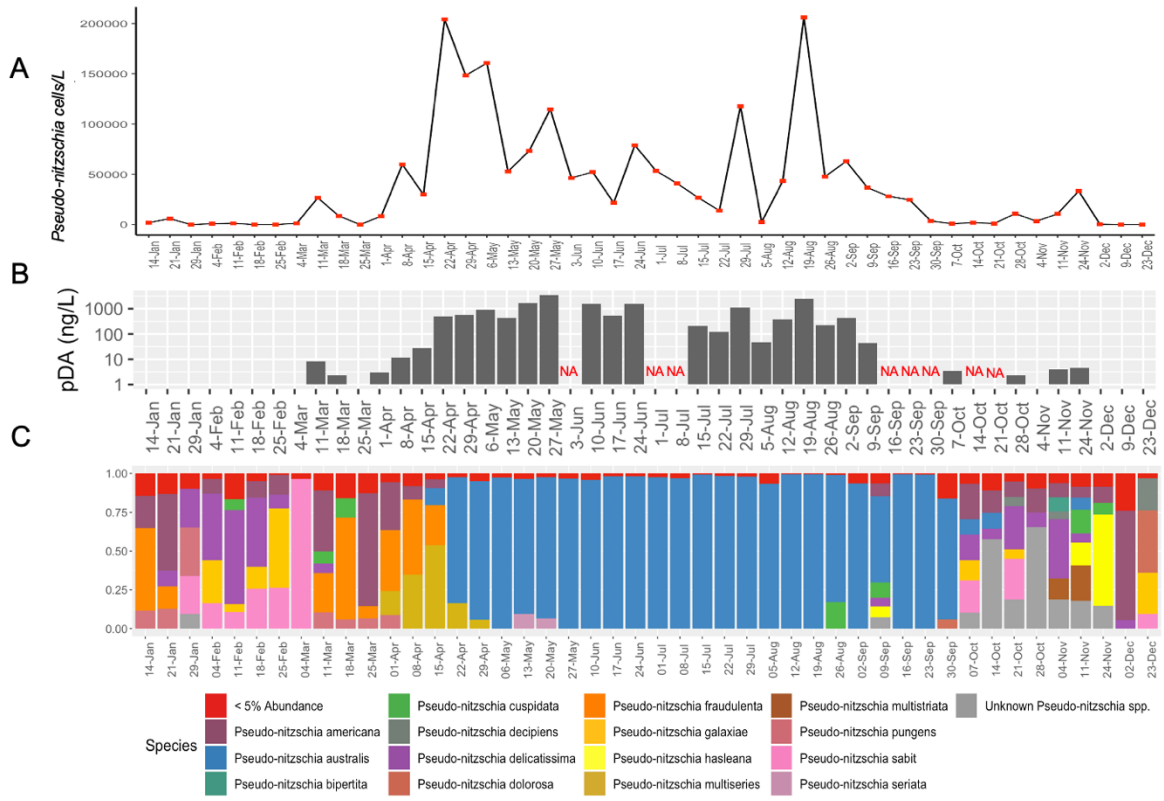



**Figure 4.3.** *Pseudo-nitzschia* targeted ITS2 Amplicon Sequencing. (A) *Pseudo-nitzschia* cell counts from weekly MWII phytoplankton net tows, as reported previously by CalHABMAP monitoring,[7] (B) Particulate DA (pDA, ng/L) measurements from filtered seawater, as reported by CalHABMAP monitoring (NA= data not available), and (C) Relative abundance of *Pseudo-nitzschia* species as determined by ITS2 amplicon sequencing. Non-ITS2 sequences were removed prior to analysis.

**Table 4.1.** Contigs encoding *dab* genes from *de novo* metatranscriptomic assembly

| **orf_id** | ***dab* gene** | **Producing species** | **Dates Detected** |
|---|---|---|---|
| contig_392124_1_1473_+ | dabA | P. multiseries | 4/1 - 5/13 |
| contig_469576_120_1547_- | dabA | P. australis | 4/15 – 9/30 |
| contig_519263_1_1377_- | dabA | P. seriata | 4/22 – 6/10 |
| contig_390219_292_1467_- | dabC | unknown | 3/11-4/08 |
| contig_392175_3_581_-<br>contig_698932_2_493_- | dabC | P. multiseries | 4/1 – 5/20 |
| contig_519989_125_1249_- | dabC | P. australis | 4/15 – 9/30 |
| contig_383882_249_1943_- | dabD | P. australis | 4/15 – 9/30 |

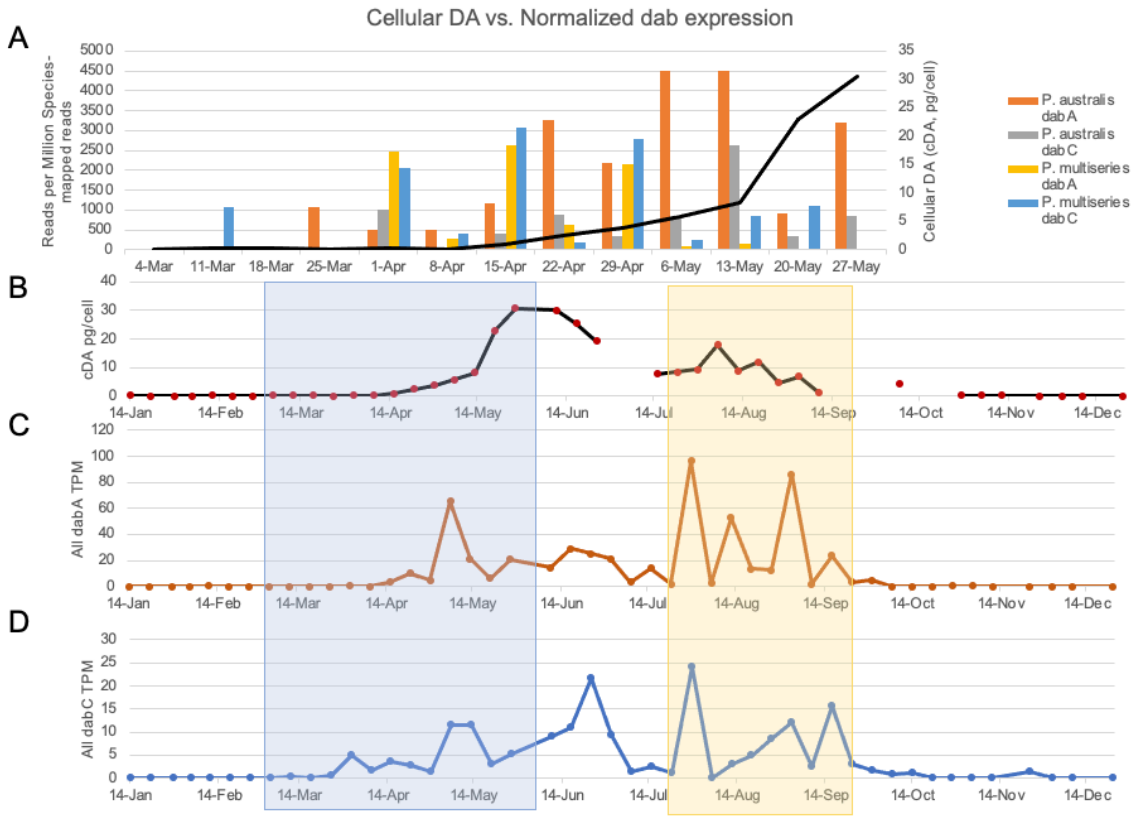



**Figure 4.4.** Transcription of *dabA* and *dabC* throughout the bloom. (A) Normalized expression of *dabA* and *dabC* transcripts plotted against cellular DA (cDA, pg/cell) during bloom initiation in the spring. Assigned reads for *P. australis* or *P. multiseries dab* transcripts were normalized by the total number of *P. australis* or *P. multiseries*-mapped reads per library to estimate relative expression within each DA-producing species, (B) cDA values for the 2015 calendar year, (C) Summed expression of all *dabA* transcripts, represented as library normalized per-million reads, (D) Summed expression of all *dabC* transcripts, represented as library normalized per-million reads.

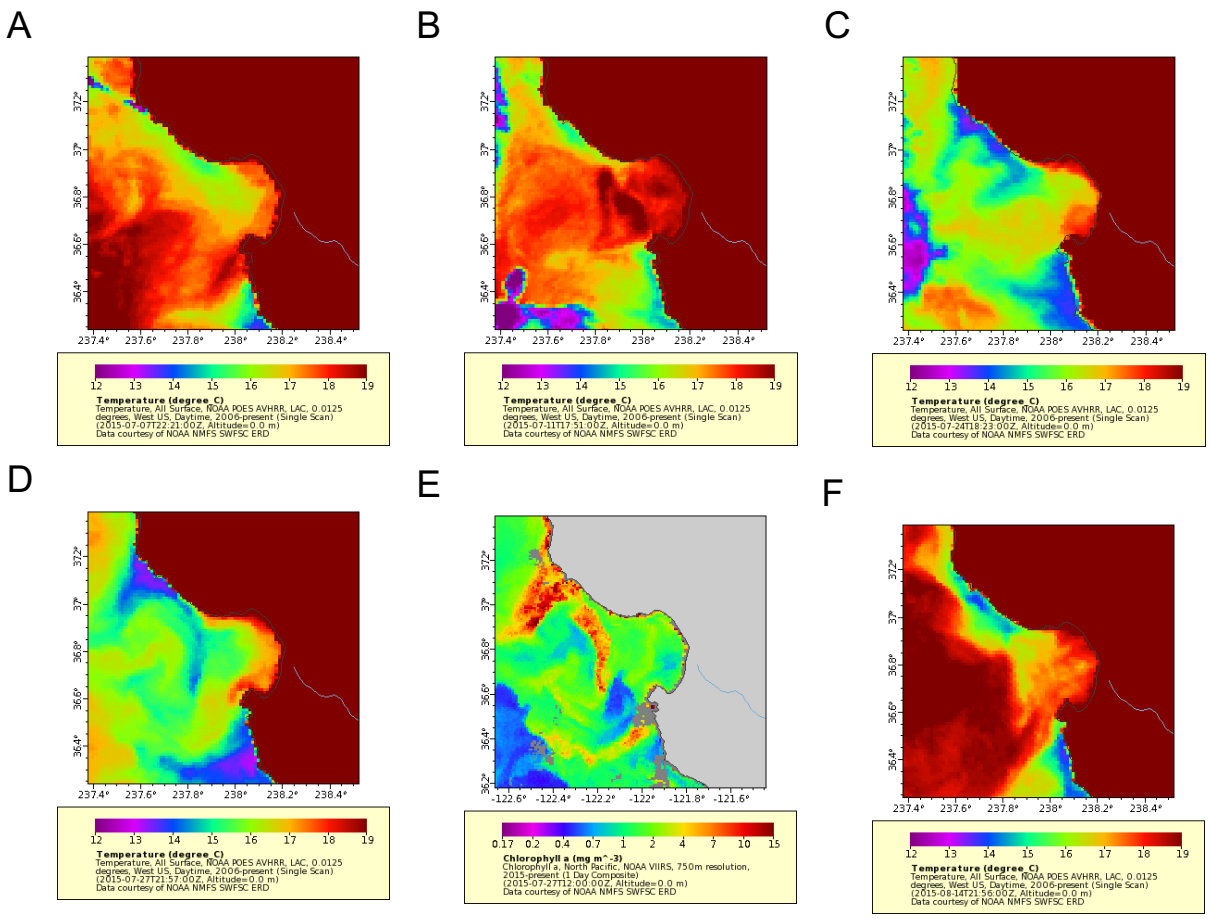



**Figure 4.5.** Satellite imaging of Monterey Bay to capture late July upwelling. (A) All surface temperature profile (°C) of Monterey Bay on July 7th, captured by NOAA AVHRR. (B) All surface temperature profile (°C) of Monterey Bay on July 11th, captured by NOAA AVHRR. (C) All surface temperature profile (°C) of Monterey Bay on July 24th, captured by NOAA AVHRR. (D) All surface temperature profile (°C) of Monterey Bay on July 27th, captured by NOAA AVHRR. (E) Chlorophyll a concentration ($mg/m^3$) in Monterey Bay on July 27th, captured by NOAA VIIRS. (F) All surface temperature profile (°C) of Monterey Bay on August 14th, captured by NOAA AVHRR.

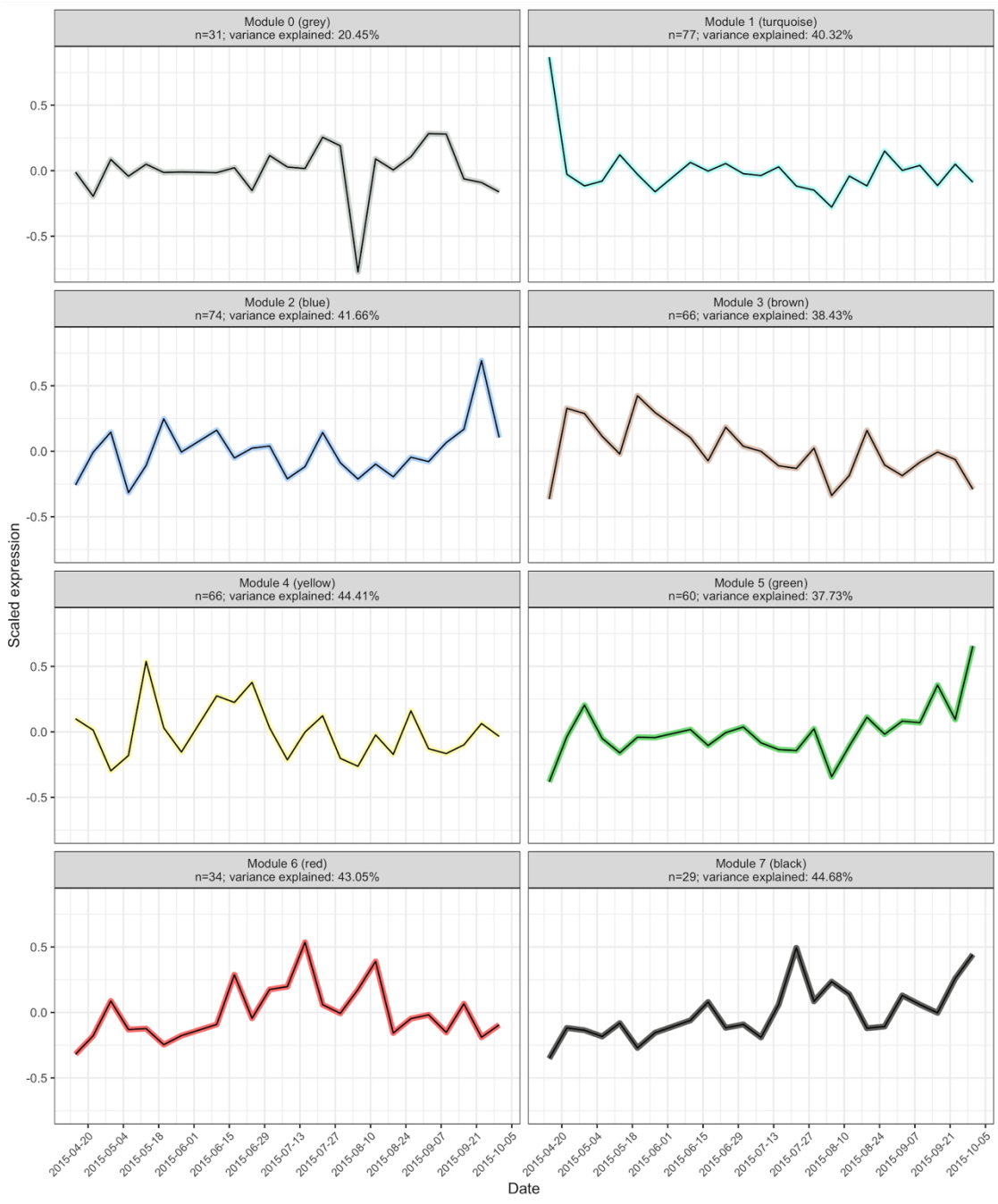



**Figure 4.6.** Relative Expression Modules for *P. australis* transcripts throughout the HAB event determined by WGCNA.

A

| DA or isoprenoid biosynthesis gene names | WGCNA Module |
|---|---|
| **dabA** | Module 7 |
| **dabC** | Module 4 |
| **dabD** | Module 7 |
| **HDS** | Module 7 |
| **IDI** | Module 7 |
| **DXS** | Module 2 |
| **HDR** | Module 4 |

B

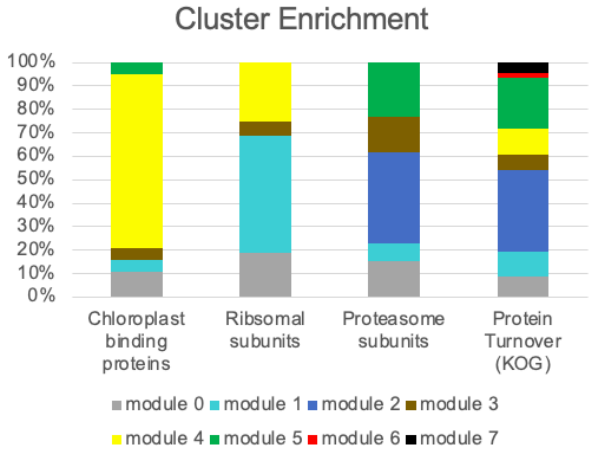



**Figure 4.7.** Annotations of interest from WGCNA clustering analysis (*A*) Functional clustering of DA biosynthesis (*dab*) and isoprenoid biosynthesis transcripts. (*B*) Enrichment of ORF annotations in WGCNA modules. Chloroplast binding proteins, ribosomal subunits and proteasome subunits were determined on the basis of PFam annotation, among other annotations. "Protein turnover (KOG)" ORFs includes all proteins in the KOG class "Posttranslational modification, Protein turnover, Chaperones"

#### 4.3.1.6 Works Cited to the Main Text


1. D. M. Anderson, E. Fensin, C. J. Gobler, A. E. Hoeglund, K. A. Hubbard, D. M. Kulis, J. H. Landsberg, K. A. Lefebvre, P. Provoost, M. L. Richlen, J. L. Smith, A. R. Solow, V. L. Trainer. Marine harmful algal blooms (HABs) in the United States: history, current status and future trends. *Harmful Algae* **102**, 101975 (2021)

2. C. R. Anderson, E. Berdalet, R. M. Kudela, C. K. Cusack, J. Silke, E. O'Rourke, D. Dugan, M. McCammon, J. A. Newton, S. K. Moore, K. Paige, S. Ruberg, J. R. Morrison, B. Kirkpatrick, K. Hubbard, J. Morell. Scaling up from regional case studies to a global harmful algal bloom observing system. *Front. Mar. Sci*. **6**, 250 (2019)

3. C. R. Anderson, R. M. Kudela, M. Kahru, Y. Chao, L. K. Rosenfeld, F. L. Bahr, D. M. Anderson, T. A. Norris. Initial skill assessment of the California Harmful Algae Risk Mapping (C-HARM) system. *Harmful Algae* **59,** 1–18 (2016)

4. J. Smith, P. Connell, R. H. Evans, A. G. Gellene, M. D. A. Howard, B. H. Jones, S. Kaveggia, L. Palmer, A. Schnetzer, B. N. Seegers, E. L. Seubert, A. O. Tatters, D. A. Caron. A decade and a half of *Pseudo-nitzschia spp*. and domoic acid along the coast of southern California. *Harmful Algae* **79**, 87–104 (2018)

5. R. M. McCabe, B. M. Hickey, R. M. Kudela, K. A. Lefebvre, N. G. Adams, B. D. Bill, F. M. D. Gulland, R. E. Thomson, W. P. Cochlan, V. L. Trainer, An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions. *Geophys. Res. Lett.* **43**, 10366– 10376 (2016)

6. J. P. Ryan, R. M. Kudela, J. M. Birch, M. Blum, H. A. Bowers, F. P. Chavez, G. J. Doucette, K. Hayashi, R. Marin, C. M. Mikulski, J. T. Pennington, C. A. Scholin, G. J. Smith, A. Woods, Y. Zhang. Causality of an extreme harmful algal bloom in Monterey Bay, California, during the 2014–2016 northeast Pacific warm anomaly. *Geophys. Res. Lett.* **44(11)**, 5571–5579 (2017)

7. H. A. Bowers, J. P. Ryan, K. Hayashi, A. L. Woods, R. Marin, G. J. Smith, K. A. Hubbard, G. J. Doucette, C. M. Mikulski, A. G. Gellene, Y. Zhang, R. M. Kudela, D. A. Caron, J. M. Birch, C. A. Scholin. Diversity and toxicity of *Pseudo-nitzschia* species in Monterey Bay: perspectives from targeted and adaptive sampling. *Harmful Algae* **78**, 129–141 (2018)

8. L. K. Rosenfeld, F. B. Schwing, N. Garfield, D. E. Tracy. Bifurcated flow from an upwelling center: a cold water source for Monterey Bay. *Cont. Shelf Res.* **14(9)**, 931–964 (1994)

9. W. M. Graham, J. L. Largier. Upwelling shadows as nearshore retention sites: the example of northern Monterey Bay. *Cont. Shelf Res.* **17(5)**, 509–532 (1997)

10. J. P. Ryan, M. A. McManus, R. M. Kudela, M. Lara Artigas, J. G. Bellingham, F. P.

Chavez, G. Doucette, D. Foley, M. Godin, J. B. J. Harvey, R. Marin, M. Messié, C. Mikulski, T. Pennington, F. Py, K. Rajan, I. Shulman, Z. Wang, Y. Zhang. Boundary influences on HAB phytoplankton ecology in a stratification-enhanced upwelling shadow. *Deep Sea. Res. Part II Top. Stud. Oceanogr.* **101**, 63–79 (2014)

11. Y. Pan, D. V. Subba Rao, K. H. Mann, R. G. Brown, R. Pocklington. Effects of silicate limitation on production of domoic acid, a neurotoxin, by the diatom *Pseudo-nitzschia multiseries*. I. Batch culture studies. *Mar. Ecol. Prog. Ser.* **131**, 225–233 (1996)

12. J. Fehling, K. Davidson, C. J. Bolch, S. S. Bates. Growth and domoic acid production by *Pseudo-nitzschia seriata* (Bacillariophyceae) under phosphate and silicate limitation. *J Phycol*. **40(4)**, 674–683 (2004)

13. K. A. Lema, M. Latimier, É. Nézan, J. Fauchot, M. Le Gac. Inter and intra-specific growth and domoic acid production in relation to nutrient ratios and concentrations in *Pseudo-nitzschia*: phosphate an important factor. *Harmful Algae* 64:11–19 (2017)

14. J. K. Brunson, S. M. K. McKinnie, J. R. Chekan, J. P. McCrow, Z. D. Miles, E. M. Bertrand, V. A. Bielinski, H. Luhavaya, M. Oborník, G. J. Smith, D. A. Hutchins, A. E. Allen, B. S. Moore. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018)

15. J. Sun, D. A. Hutchins, Y. Feng, E. L. Seubert, D. A. Caron, F.-X. Fu, Effects of changing pCO2 and phosphate availability on domoic acid production and physiology of the marine harmful bloom diatom *Pseudo-nitzschia multiseries*. *Limnol. Oceanogr.* 56, 829–840 (2011)

16. S. Harðardóttir, S. Wohlrab, D. M. Hjort, B. Krock, T. G. Nielsen, U. John, N. Lundholm. Transcriptomic responses to grazing reveal the metabolic pathway leading to the biosynthesis of domoic acid and highlight different defense strategies in diatoms. *BMC Mol. Biol.* **20**, 7 (2019)

17. R. M. Kudela, A. Bickel, M. L. Carter, M. D. A. Howard, L. Rosenfeld. The Monitoring of harmful algal blooms through ocean observing: the development of the California Harmful Algal Bloom Monitoring and Alert Program. *Coastal Ocean Observing Systems* 58-75 (2015)

18. H. A. Bowers, R. Marin, J. M. Birch, C. A. Scholin. Sandwich hybridization probes for the detection of *Pseudo-nitzschia* (Bacillariophyceae) species: an update to existing probes and a description of new probes. *Harmful Algae* **70**, 37–51 (2017)

19. S. Malviya, E. Scalco, S. Audic, F. Vincent, A. Veluchamy, J. Poulain, P. Wincker, D. Iudicone, C. De Vargas, L. Bittner, A. Zingone, C. Bowler. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci*. **113(11)**, E1516–E1525 (2016)

20. K. M. McKindles, P. V. Zimba, A. S. Chiu, S. B. Watson, D. B. Gutierrez, J. Westrick, H. Kling, T. W. Davis. A multiplex analysis of potentially toxic cyanobacteria in Lake Winnipeg during the 2013 bloom season. *Toxins* **11(10)**, 587 (2019)

21. H. C. Lim, S. N. Tan, S. T. Teng, N. Lundholm, E. Orive, H. David, S. Quijano-Scheggia, S. C. Y. Leong, M. Wolf, S. S. Bates, P. T. Lim, C. P. Leaw. Phylogeny and species delineation in the marine diatom *Pseudo-nitzschia* (Bacillariophyta) using cox1, LSU, and ITS2 rRNA genes: a perspective in character evolution. *J. Phycol.* **54(2)**, 234–248 (2018)

22. A. Lelong, H. Hégaret, P. Soudant, S. S. Bates, *Pseudo-nitzschia* (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. *Phycologia* **51**, 168–216 (2012)23.

23. S. S. Bates, K. A. Hubbard, N. Lundholm, M. Montresor, C. P. Leaw, *Pseudo-nitzschia*, *Nitzschia*, and domoic acid: new research since 2011. *Harmful Algae* **79**, 3–43 (2018)

24. P. Langfelder, S. Horvath. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**, 559 (2008)

25. M. Lohr, J. Schwender, J. E. W. Polle. Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Sci*. **185–186**, 9–22 (2012)

26. S. R. Smith, J. T. F. Gillard, A. B. Kustka, J. P. McCrow, J. H. Badger, H. Zheng, A. M. New, C. L. Dupont, T. Obata, A. R. Fernie, A. E. Allen. Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. *PLOS Genet.* **12(12)**, e1006490 (2016)

27. D. P. Yee, M. Hildebrand, M. Tresguerres. Dynamic subcellular translocation of V-type $H^+$-ATPase is essential for biomineralization of the diatom silica cell wall. *New Phytol.* **225(6)**, 2411–2422 (2020)

28. B. Xu, C-S. Luo, J-R. Liang, D-D. Chen, W-H. Zhuo, Y-H. Gao, C-P. Chen, S-S. Song. Cellular metabolic responses of the marine diatom *Pseudo-nitzschia multiseries* associated with cell wall formation. *Mar. Genomics* **16**, 29–38 (2014)

29. S. Thangaraj, M. Giordano, J. Sun. Comparative proteomic analysis reveals new insights into the common and specific metabolic regulation of the diatom *Skeletonema dohrnii* to the silicate and temperature availability. *Front. Plant Sci.* 11, 578915 (2020)

30. S. R. Smith, C. L. Dupont, J. K. McCarthy, J. T. Broddrick, M. Oborník, A. Horák, Z. Füssy, J. Cihlář, S. Kleessen, H. Zheng, J. P. McCrow, K. K. Hixson, W. L. Araújo, A. Nunes-Nesi, A. Fernie, Z. Nikoloski, B. O. Palsson, A. E. Allen. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* **10**, 4552 (2019)

31. J. K. McCarthy, S. R. Smith, J. McCrow, M. Tan, H. Zheng, K. Beeri, R. Roth, C. Lichtle, U. Goodenough, C. Bowler, C. Dupont, A. E. Allen. Nitrate reductase knockout uncouples

nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell*. **29(8)**, 2047–2070 (2017)

32. R. H. Lampe, E. L. Mann, N. R. Cohen, C. P. Till, K. Thamatrakoln, M. A. Brzezinski, K. W. Bruland, B. S. Twining, A. Marchetti. Different iron storage strategies among bloom-forming diatoms. *Proc. Natl. Acad. Sci.* **115(52)**, E12275–E12284 (2018)

33. S. L. Hogle, C. L. Dupont, B. M. Hopkinson, A. L. King, K. N. Buck, K. L. Roe, R. K. Stuart, A. E. Allen, E. L. Mann, Z. I. Johnson, K. A. Barbeau. Pervasive iron limitation at subsurface chlorophyll maxima of the California Current. *Proc. Natl. Acad. Sci.* **115(52)**, 13300–13305 (2018)

### 4.3.2 Supplementary Information including Materials and Methods

#### 4.3.2.1 Materials and Methods

Routine Monitoring Data Availability and Visualization of Satellite Data

All routine monitoring data collected weekly from Monterey Municipal Wharf II (MWII) during the study period, including *Pseudo-nitzschia* cell counts, DA measurements, local chlorophyll concentration, temperature, and nutrient data, is made publicly available through the Southern California Coastal Ocean Observing System (SCCOOS, https://sccoos.org/harmful-algal-bloom/) as part of the California Harmful Algal Bloom Monitoring and Alert Program (CalHABMAP). Satellite data was acquired and visualized via the Environmental Research Division's Data Access Program (ERDDAP) data server from the National Oceanic and Atmospheric Association (NOAA). All-surface temperature data was collected in Local Area Coverage (LAC) format by the Advanced Very High Resolution Radiometer (AVHRR) scanner onboard NOAA's Polar Orbiting Environmental Satellite (POES). Satellite-measured remote chlorophyll concentration ($mg/m^3$) was collected by the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi National Polar-Orbiting Partnership (Suomi NPP) satellite.

Phytoplankton Net Tow and Sample Filtering

Weekly net tows from MWII were obtained using a 20 cm diameter, 20 μm mesh net to concentrate surface waters to a depth of 5 meters, as is routine for monitoring at MWII. From the net tows, 100 mL of concentrated phytoplankton samples were applied to 0.22 μm polycarbonate (PC) filters, which were then stored in 1 mL of TRIzol (Invitrogen) reagent in 15 mL conical

tubes, mixed by vortexing, and kept at -80ºC for future processing. A total of 53 samples were processed and stored in this manner.

RNA extraction and cDNA generation

Samples were removed from -80ºC storage. The PC filter was removed from the TRIzol solution and samples were subsequently centrifuged for 20 minutes at 2500 xg and 4 ºC to pellet particulate matter. The supernatant was then transferred to a new tube. RNA extraction begins with the addition of 200 µL of chloroform to each sample, which were then shaken vigorously for 15 seconds and allowed to stand for 12–15 minutes at room temperature (~25 ºC). The resulting mixture was then centrifuged at 12,000 xg for 15 minutes at 4 ºC. Next, the aqueous phase was transferred to a new tube, adding an equal volume of 100% ethanol and mixing well. The mixture was then transferred to a Zymo-Spin IIICG Column (Zymo Research), after which the RNA extraction was completed using the Direct-zol RNA Miniprep Plus kit (Zymo Research) following the manufacturer protocol, including an on-column DNA digestion step using DNase I (New England Biolabs). Suitable RNA quality was verified using an Agilent 2100 Bioanalyzer.

Synthesis of cDNA to be used as template for generation of 18S-V4, 16S and ITS2 amplicons was preformed using the SuperScript III First-Strand Synthesis System (Life Technologies). 100 ng of total RNA per sample was used with random hexamer primers to make 20 µL of cDNA following kit standard protocol.

Preparation and sequencing of 18SV4, 16S, and ITS2 amplicon libraries

Amplicon libraries for 18SV4, 16S and ITS2 sequences were generated using the cDNA libraries from above as template for the one-step PCR reactions to simultaneously amplify target sequences and incorporate Illumina adaptors, linker sequences, and unique barcoded indices. All one-step PCR reactions were performed using the TruFi DNA Polymerase PCR kit (Azura). All primers used in this study, together with attached adaptors, linkers and barcoded indices are listed in Table S4.1. The 18SV4 region was amplified using the primer pair V4F and V4RB.[1] The 16S(V4-V5) region was amplified using the primer pair 515F-Y and 926R.[2] The ITS2 region was amplified using an unpublished primer pair 5.8SF and 28SR. Primer pairs containing unique combinations of barcorded indices, allowing downstream demultiplexing, are generated in 10 μM concentration and stored in 96-well plates.

PCR reactions are then set up using 1 μL of cDNA as template, 0.4 μM of primer pair, and TruFi DNA Polymerase and buffer mix per manufacturer protocol and brought to a total reaction volume of 25 μL using molecular grade water. PCR reactions for 18SV4 and 16S amplicon generation were performed with an initial denaturing step at 95°C for 1 minute followed by 30 cycles of denaturing at 95°C for 15 seconds, annealing at 56°C for 15 seconds, and extension at 72°C for 30 seconds. Initial PCR reactions for ITS2 followed a similar protocol, however improved amplification was observed when the annealing temperature was decreased to 51-53°C. To confirm amplification and correct size of amplicon, 2.5 μL of each PCR reaction was ran on a 1.8% agarose gel.

The PCR products were then cleaned up using AMPure XP beads (Agencourt) using the standard PCR purification protocol per manufacturer recommendations and eluted in 35 μl of elution buffer. Cleaned PCR products were then quantified using the PicoGreen Quant-IT dsDNA Quantitation Reagent (Life Technologies). Equal quantities of 18SV4, 16S or ITS2

amplicons (~10 ng per reaction) were pooled, cleaned, and concentrated using AMPure XP beads using standard protocol, eluting in 45 µl of elution buffer. Final library quality was assessed using a TapeStation (Agilent) and quantified on a Qubit fluorometer (ThermoFisher). Respective pools were then sequenced on the MiSeq PE300 (Illumina) generating 250 bp paired end reads for all amplicon libraries. Sequencing was performed at the UC Davis Sequencing Core.

Assembly, taxonomic classification and visualization of amplicon sequencing

The 18SV4, 16S and ITS2 amplicon sequences were imported separately into Qiime2 where the dada2 plug-in was used to merge, quality filter the reads, identify and remove chimeric reads, and generate Amplicon Sequence Variants (ASVs) and count tables.[3,4] Taxonomy was assigned to the 18S ASVs using a qiime2 naïve-bayes classifier trained on the PR2 version 4.11.1 database.[5] The 16S ASVs were annotated first using a qiime2 naïve-bayes classifier trained on the Silva-132 database in order to differentiate bacterial and mitochondrial sequences from chloroplast sequences.[6] Chloroplast ASVs were extracted in Qiime2 and further annotated using a qiime2 naïve-bayes classifier trained on the PhytoREF database.[7] Genus and species classifications for ITS2 ASVs were assigned using the "classify-consensus-blast" feature classifier, using a database of *Pseudo-nitzschia* ITS2 sequences as the reference BLAST database.[8] Conglomeration of ASV counts by taxonomic level and subsequent generation of stacked bar plots was performed in PhyloSeq v.1.34.0.[9]

Preparation and sequencing of polyA-enriched RNA

Starting with 500 ng of total RNA per sample, we used the TruSeq Stranded mRNA prep kit (Illumina) to make polyA-enriched libraries suitable for RNA sequencing, following manufacturer protocols. Following preparation of polyA-enriched libraries, samples were combined into two pools and ran on the HiSeq4000 PE150 platform (Illumina) at the UC Davis Sequencing Core.

Assembly and annotation of RNA sequencing libraries

The resulting demultiplexed HiSeq4000 read libraries were processed via the RNAseq Annotation Pipeline v0.4.[10] Reads were first trimmed for quality and filtered to remove primers, adaptors and rRNA sequences using Ribopicker v.0.4.3.[11] CLC Genomics Workbench 9.5.3 (QIAGEN) was used to assemble contigs, first by library, then globally. Open reading frames (ORFs) were predicted from the assembled contigs using FragGeneScan.[12] The resulting ORFs were annotated *de novo* for function and taxonomy via KEGG, KO, KOG, Pfam, TigrFam and the reference dataset PhyloDB 1.02.[10] Reads were mapped back to assembled contigs using BWA MEM to generate read counts.[13]

Defining expression modules for the *P. australis* bloom

The Weighted Gene Correlation Network Analysis (WGCNA) R package was implemented to identify modules of *P. australis* ORFs with similar expression patterns to define functional clusters.[14] The bloom period for WGCNA analysis purposes was defined as April 15$^{th}$ through September 30$^{th}$. *P. australis* ORFs were identified and filtered based on taxonomic annotation above, leveraging the PhyloDB 1.02 database which contains a *P. australis* annotated transcriptome.[10,15] This yielded 14873 *de novo* assembled *P. australis* ORFs. To obtain

statistically relevant results, only ORFs with 10 or more reads in 80% of the libraries comprising the bloom period were considered, yielding 437 ORFs for WGCNA analysis.

Following the filtering step, expression was normalized by dividing reads mapped to a given ORF by the total number of reads mapped to all assembled *P. australis* ORFs to account for library and *P. australis* population size prior to WGCNA analysis and construction of a correlation matrix. An adjacency matrix was built from the correlation matrix input by applying a power function (AF(s)=s^b) to the input data, where "b" is defined as the soft-thresholding parameter. A soft-thresholding parameter of b = 6 was found to be lowest value at which a scale-free topology $R^2$ value exceeding 0.8 was achieved (Supplementary Information, Fig. S5). This information was used to construct the consensus matrix and dendrogram using the function "blockwiseConsensusModules" using the parameters: b = 6, TOMType = "signed", detectCutHeight = 0.995, reassignThreshold = 0 (Supplementary Information, Fig. S5). The "moduleEigengenes" function was used with default parameters to select an "eigengene," or the ORF whose expression profile is most representative of the other profiles contained in the module. The "eigengene" for each module is featured in Figure 6. No minimum module size was set and no modules were merged to generate modules containing genes with the highest possible similar expression profiles to one another, maximizing the variance explained by each eigengene.

### 4.3.2.2 Supplemental Table and Figures

**Table S4.1.** Primers used in this study. Sequence in red and green represent Illumina sequencing adaptors, black sequence represents library specific index sequences (8 nt), blue sequence represents amplicon specific annealing sequence.

| Amplicon | Primer name | Primer Sequence |
|---|---|---|
| 18SV4 | V4F | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNXXXXXXXXCCAGCASCYGCGGTAATTCC |
| 18SV4 | V4RB | CAAGCAGAAGACGGCATACGAGATXXXXXXXXTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCTACTTTCGTTCTTGATYR |
| 16S | 515F-Y | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNXXXXXXXXGTGYCAGCMGCCGCGGTAA |
| 16S | 926R | CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCCGYCAATTYMTTTRAGTTT |
| ITS2 | 5.8SF | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNXXXXXXXXTGCTTGTCTGAGTGTCTGTGGA |
| ITS2 | 28SR | CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTTATGCTTAAATTCAGCGGGT |

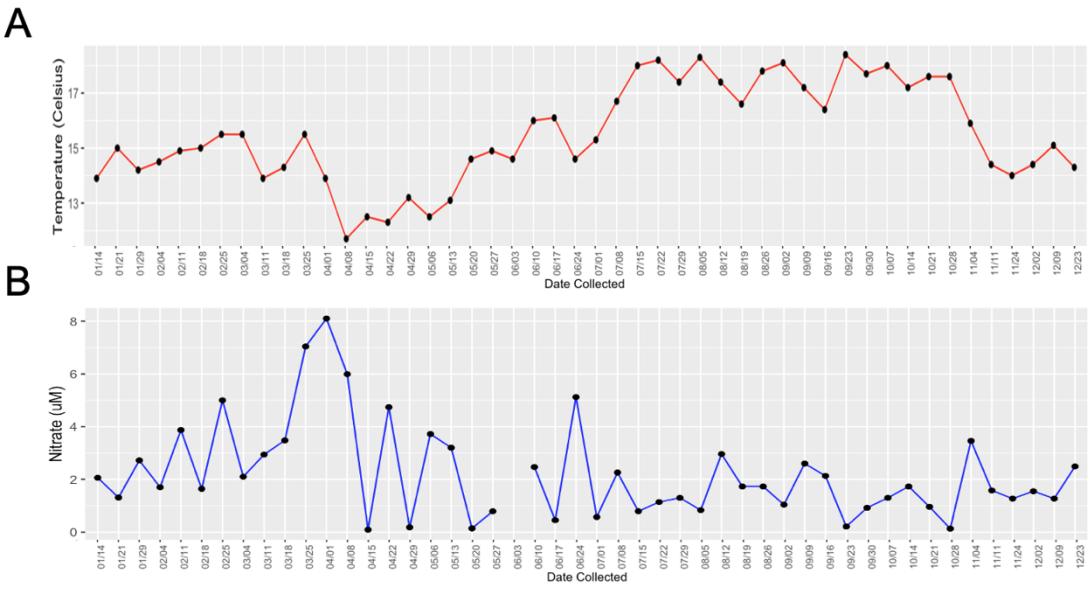



**Figure S4.1.** (*A*) Temperature (°C) and (*B*) nitrate (μM) measurements collected from MWII.

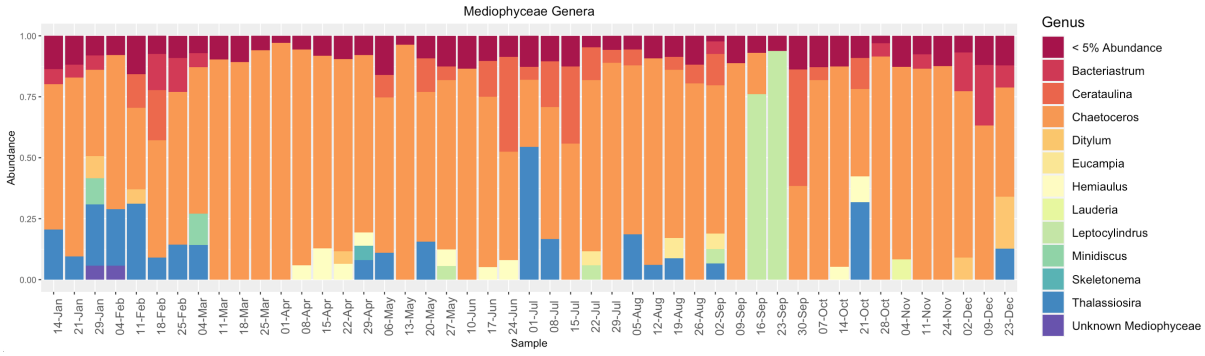



**Figure S4.2.** Genus composition of Mediophyceae (centric) diatoms determined by 18SV4 sequencing from weekly samples throughout 2015

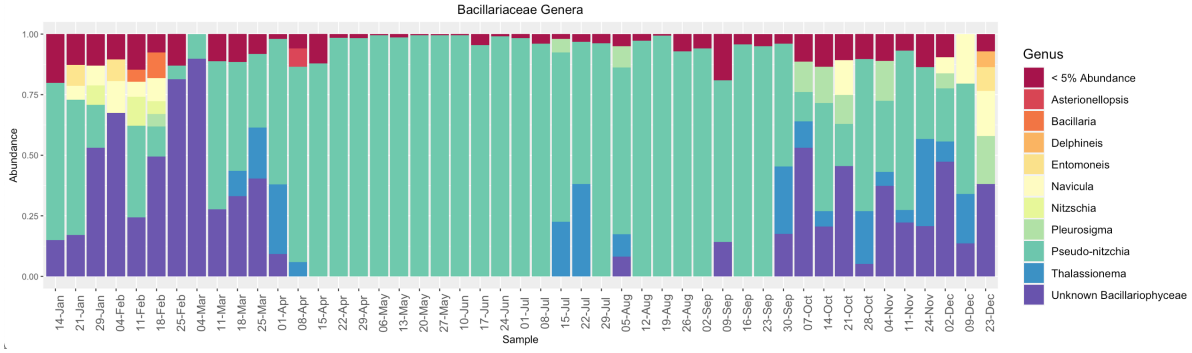



**Figure S4.3.** Genus composition of Bacillariophyceae (pennate) diatoms determined by 18SV4 sequencing from weekly samples throughout 2015

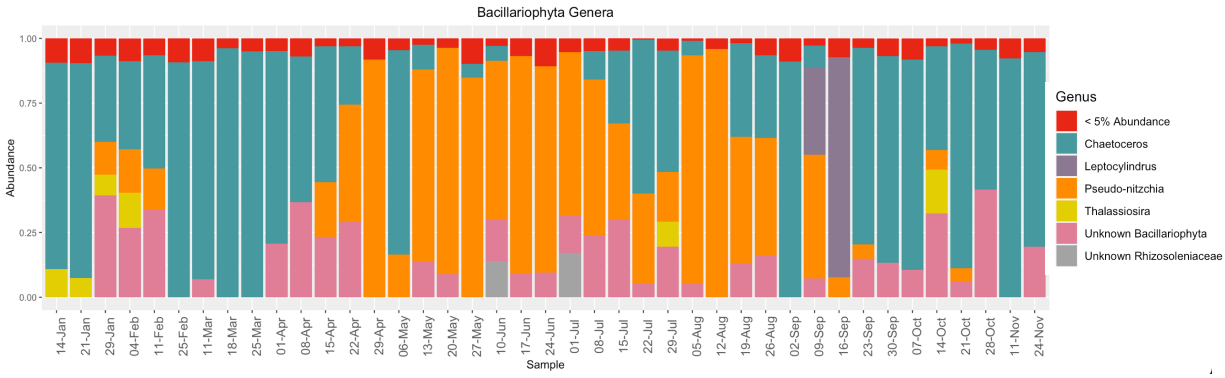



**Figure S4.4.** Genus composition of diatoms (Phylum Bacillariophyta) determined by 16S chloroplast sequencing from weekly samples throughout 2015

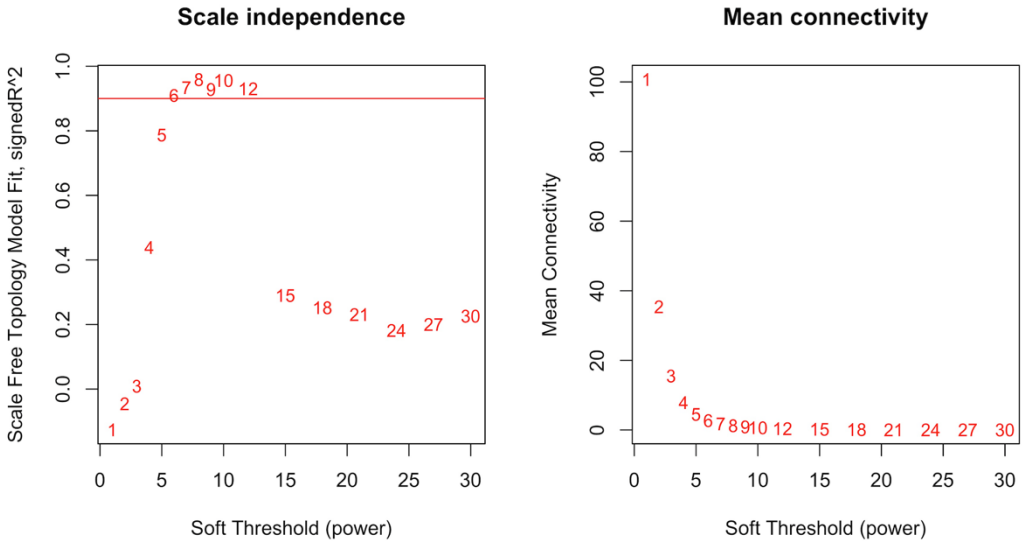



**Figure S4.5.** Determination of ideal soft-thresholding parameter "b" to test for the lowest value of "b" to exceed a scale-free topology $R^2$ value of 0.8, showing scale-free fit index and mean connectivity.

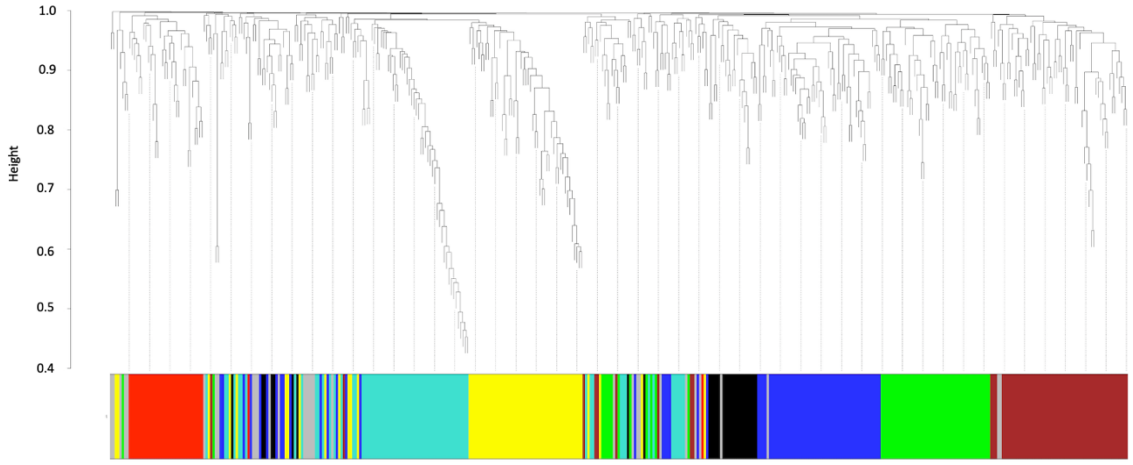



**Figure S4.6.** Clustering dendrogram of *P. australis* ORF expression profiles, together with assigned module colors.

#### 4.3.2.3 Works Cited to the Supplementary Information


1. S. Balzano, E. Abs, S. Leterme S. Protist diversity along a salinity gradient in a coastal lagoon. *Aquat. Microb. Ecol.* **74(3)**, 263–277 (2015)

2. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol [Internet]. 2016 May;18(5):1403–1414. Available from:

3. E. Bolyen, J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J. E. Bisanz, K. Bittinger, A. Brejnrod, C. J. Brislawn, C. T. Brown, B. J. Callahan, A. M. Caraballo-Rodríguez, J. Chase, E. K. Cope, R. Da Silva, C. Diener, P. C. Dorrestein, G. M. Douglas, D. M. Durall, C. Duvallet, C. F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J. M. Gauglitz, S. M. Gibbons, D. L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G. A. Huttley, S. Janssen, A. K. Jarmusch, L. Jiang, B. D. Kaehler, K-B Kang, C. R. Keefe, P. Keim, S. T. Kelley, D. Knights, I. Koester, T. Kosciolek, J. Kreps, M. G. I. Langille, J. Lee, R. Ley, Y-X. Liu, E. Loftfield, C. Lozupone, M. Maher, C. Marotz, B. D. Martin, D. McDonald, L. J. McIver, A. V. Melnik, J. L. Metcalf, S. C. Morgan, J. T. Morton, A. T. Naimey, J. A. Navas-Molina, L. F. Nothias, S. B. Orchanian, T. Pearson, S. L. Peoples, D. Petras, M. L. Preuss, E. Pruesse, L. B. Rasmussen, A. Rivers, M. S. Robeson, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S. J. Song, J. R. Spear, A. D. Swafford, L. R. Thompson, P. J. Torres, P. Trinh, A. Tripathi, P. J. Turnbaugh, S. Ul-Hasan, J. J. J. van der Hooft, F. Vargas, Y. Vázquez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, Y. Wan, M. Wang, J. Warren, K. C. Weber, C. H. D. Williamson, A. D. Willis, Z. Z. Xu, J. R. Zaneveld, Y. Zhang, Q. Zhu, R. Knight, J. G. Caporaso. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37(8)**, 852–857 (2019)

4. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13(7)**, 581–583 (2016)

5. L. Guillou, D. Bachar, S. Audic, D. Bass, C. Berney, L. Bittner, C. Boutte, G. Burgaud, C. de Vargas, J. Decelle, J. del Campo, J. R. Dolan, M. Dunthorn, B. Edvardsen, M. Holzmann, W. H. C. F. Kooistra, E. Lara, N. Le Bescot, R. Logares, F. Mahé, R. Massana, M. Montresor, R. Morard, F. Not, J. Pawlowski, I. Probert, A-L. Sauvadet, R. Siano, T. Stoeck, D. Vaulot, P. Zimmermann, R. Christen. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41(D1)**, D597–D604 (2012)

6. P. Yilmaz, L. W. Parfrey, P. Yarza, J. Gerken, E. Pruesse, C. Quast, T. Schweer, J. Peplies, W. Ludwig, F. O. Glöckner. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res*. **42(D1)**, D643–D648 (2014)

7. J. Decelle, S. Romac, R. F. Stern, E. M. Bendif, A. Zingone, S. Audic, M. D. Guiry, L.

Guillou, D. Tessier, F. Le Gall, P. Gourvil, A. L. Dos Santos, I. Probert, D. Vaulot, C. de Vargas, R. Christen. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Mol. Ecol. Resour.* **15(6)**, 1435–1445 (2015)

8. H. C. Lim, S. N. Tan, S. T. Teng, N. Lundholm, E. Orive, H. David, S. Quijano-Scheggia, S. C. Y. Leong, M. Wolf, S. S. Bates, P. T. Lim, C. P. Leaw. Phylogeny and species delineation in the marine diatom *Pseudo-nitzschia* (Bacillariophyta) using cox1, LSU, and ITS2 rRNA genes: a perspective in character evolution. *J. Phycol.* **54(2)**, 234–248 (2018)

9. P. J. McMurdie, S. Holmes. phyloseq: an R Package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8(4), e61217 (2013)

10. E. M. Bertrand, J. P. McCrow, A. Moustafa, H. Zheng, J. B. McQuaid, T. O. Delmont, A. F. Post, R. E. Sipler, J. L. Spackeen, K. Xu, D. A. Bronk, D. A. Hutchins, A. E. Allen, Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci.* **112**, 9938–9943 (2015).

11. R. Schmieder, Y. W. Lim, R. Edwards, Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2012).

12. M. Rho, H. Tang, Y. Ye, FragGeneScan: Predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191 (2010).

13. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio.GN]* (2013).

14. P. Langfelder, S. Horvath. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**, 559 (2008)

15. P. J. Keeling, F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. B. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone, A. Z. Worden, The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol.* **12**, e1001889 (2014).

## 4.4 Acknowledgments and Author Contributions

The main text and supplemental information sections of Chapter 4 are currently in preparation for manuscript submission as "**J. K. Brunson**, A. J. Rabines, H. Zheng, P. Venepally, J. P. Ryan, B. S. Moore, G. J. Smith, A. E. Allen. Metatranscriptomic sequencing and barcoding of a major harmful algal bloom event in Monterey Bay, California. (2021) Manuscript in preparation." The dissertation author was the primary investigator and author of this manuscript.

J.K.B., B.S.M., G.J.S., and A.E.A. conceived the project and designed the experiments, J.K.B. analyzed the bulk of the data and wrote the current manuscript. G.J.S. prepared samples from weekly phytoplankton net tows. H.Z. purified RNA from samples and prepared RNA sequencing libraries. J.K.B prepared cDNA libraries and performed PCR to generate amplicon sequencing libraries. A.J.R. prepared amplicon sequencing libraries for sequencing and assisted with data analysis for molecular barcoding experiments. P.V. assembled the RNA sequencing libraries and ran the annotation pipeline. J.K.B. analyzed satellite data with the guidance of J.P.R.

This research was supported by the National Oceanic and Atmospheric Administration (NA19NOS4780181 to B.S.M. and A.E.A) and the National Institutes of Health (F31ES030613 to J.K.B.).