Domoic acid biosynthesis in the red alga *Chondria armata* suggests a complex evolutionary history for toxin production

Taylor S. Steele[a,b,1], John K. Brunson[a,c,1], Yukari Maeno[d], Ryuta Terada[e], Andrew E. Allen[c,f], Mari Yotsu-Yamashita[d], Jonathan R. Chekan[g,2], and Bradley S. Moore[a,h,2]

[a]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093; [b]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093; [c]Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA 92037; [d]Graduate School of Agricultural Science, Tohoku University, Sendai 980-8572, Japan; [e]United Graduate School of Agricultural Sciences, Kagoshima University, Kagoshima 890-0065, Japan; [f]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92037; [g]Department of Chemistry and Biochemistry, University of North Carolina Greensboro, Greensboro, NC 27412; and [h]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093

Edited by Rodney Croteau, Institute of Biological Chemistry, Washington State University, Pullman, WA; received September 22, 2021; accepted December 10, 2021

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 *Pseudonitzschia multiseries*, establishing the genetic basis for the global production of this potent neurotoxin. Herein, we sequenced the 507-Mb genome of *Chondria armata*, the red macroalgal seaweed from which DA was first isolated in the 1950s, identifying 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. The 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 like the homologous diatom enzyme despite higher amino acid similarity to red algal kainic acid synthesis enzymes. A 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 the horizontal gene transfer of *N*-prenyltransferases and kainoid synthases to establish DA production within the algal lineages.

natural products | neurotoxin | genomics | seaweed | biosynthetic gene cluster

Harmful algal blooms of the diatom genus *Pseudonitzschia* produce high levels of domoic acid (DA), a neurotoxic glutamate receptor agonist with far-reaching food web implications because of its bioaccumulation in shellfish (1). The 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 *Pseudonitzschia*, the compound was originally discovered in the 1950s by Daigo and coauthors 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. The production of KA is limited to red algae (Rhodophyta), and while multiple orders within Rhodophyta contain species capable of making KA, the distribution of known kainoid-producing genera is sparse (*SI Appendix*, Table S1 and Fig. S1). 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. 1*A*). Among diatoms, only

Significance

Originally isolated from the red alga *Chondria armata*, domoic acid (DA) is best known as a potent marine neurotoxin produced by oceanic harmful algal blooms of planktonic diatoms. Sequencing efforts to date of kainoid-producing red algae have focused exclusively on a closely related molecule, kainic acid, leaving a gap in the understanding of DA biosynthesis in red algae and its evolutionary linkage to diatoms. Here, we present the phylogenetic and biochemical investigation of DA biosynthesis in *C. armata*. This work demonstrates the high synteny of DA biosynthetic genes between relatively distant taxonomic groups of algae and suggests a complex evolutionary history for DA biosynthesis involving gene transfer and neofunctionalization.

Author contributions: T.S.S., J.K.B., J.R.C., and B.S.M. designed research; T.S.S., J.K.B., Y.M., M.Y.-Y., and J.R.C. performed research; R.T. contributed new reagents/analytic tools; T.S.S., J.K.B., Y.M., M.Y.-Y., J.R.C., and B.S.M. analyzed data; T.S.S., J.K.B., J.R.C., and B.S.M. wrote the paper; and A.E.A. provided helpful discussions.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]T.S.S. and J.K.B. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: jrchekan@uncg.edu or bsmoore@ucsd.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2117407119/-/DCSupplemental.

Published February 2, 2022.

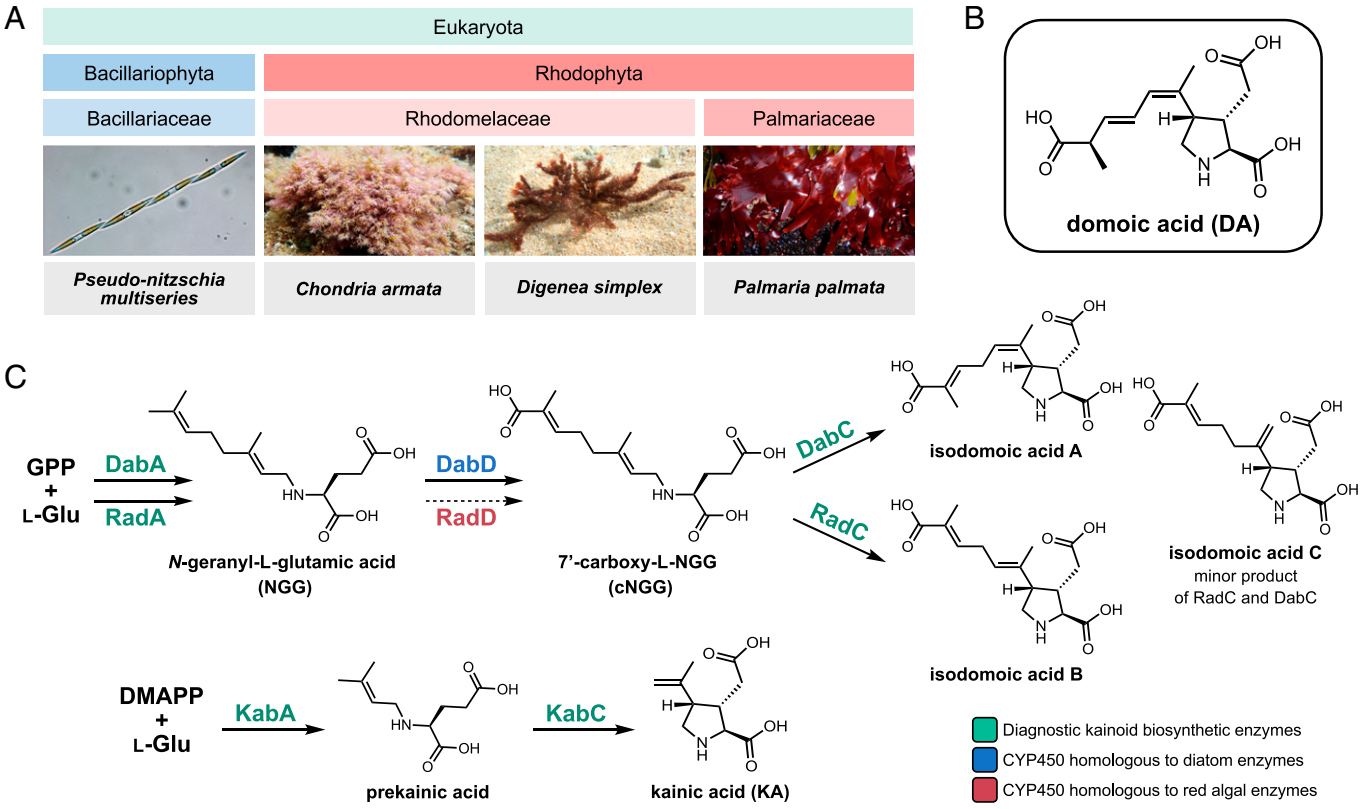

Fig. 1. Kainoid biosynthetic genes and producing organisms. (*A*) DA-producing diatom *P. multiseries* (Bacillariophyta) and KA-producing red algae *D. simplex* and *P. palmata* with DA-producing red alga *C. armata* (Rhodophyta), two significantly divergent phyla of kainoid-producing algae. (*B*) Structure of 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 (17, 18) shown with solid arrows.

the genera *Pseudonitzschia* 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. 1*B*) biosynthesis (*dab*) and KA biosynthesis (*kab*) were recently elucidated (Fig. 1*C*) (17, 18). Both pathways begin with the *N*-prenylation of L-glutamate by either geranyl diphosphate (GPP) or dimethylallyl diphosphate (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. 1*C*). 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. The discovery of the various kainoid pathways was facilitated by transcriptomic and genomic sequencing of the DA-producing diatom *Pseudonitzschia 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 cocluster 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 insights 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.

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 (20), repeat content was estimated to be 45.49%, and kmer distributions were consistent with tetraploidy (*SI Appendix*, Fig. S2). Hybrid assembly using MaSuRCA v3.4.2 (21) yielded a genome with a total size, N50, and GC-content of 507 Mb, 643 kb, and 45.34%, respectively (*SI Appendix*, Table S3). 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). Because of a lack of rhodophyta-specific datasets, we assessed the 10 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 (*SI Appendix*, Table S4 and Fig. S3). To further assess overall genome completeness, we mapped ONT reads to the assembled genome using minimap2 (27), in which 88% of ONT reads were present in the assembled genome, a clearer indicator of genome completeness. Genome sequences and assembly are deposited in the National Center for Biotechnology Information (NCBI), BioProject identification PRJNA762367.

Kainoid Cluster Identification and Comparison. To identify kainoid biosynthesis genes in the *C. armata* genome, we used *dab* and *kab* genes as genetic hooks (17, 18). Previous work demonstrated the remarkably low sequence similarity of *dabA* and *kabA* genes to any publicly available protein sequence (19). tBLASTn searches with the highly conserved *N*-prenyltransferase genes initially revealed three copies of a kainoid biosynthetic gene cluster. Notably, unlike previously characterized red algal kainoid gene clusters, two of the *C. armata* clusters included a coclustered 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: OK169902) is contained within a 7.9-kb region, and complete coding regions for all three key DA biosynthetic genes were clearly identified (Fig. 2). The second copy of the *rad* cluster (*rad2*, GenBank accession: 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*, GenBank accession: 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 (*SI Appendix*, Fig. S4). 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 (39×) that of *rad2* (21×) and *rad3* (21×) and is indicative of a collapsed paralog, supporting the presence of all four expected genomic copies of the *rad* cluster.

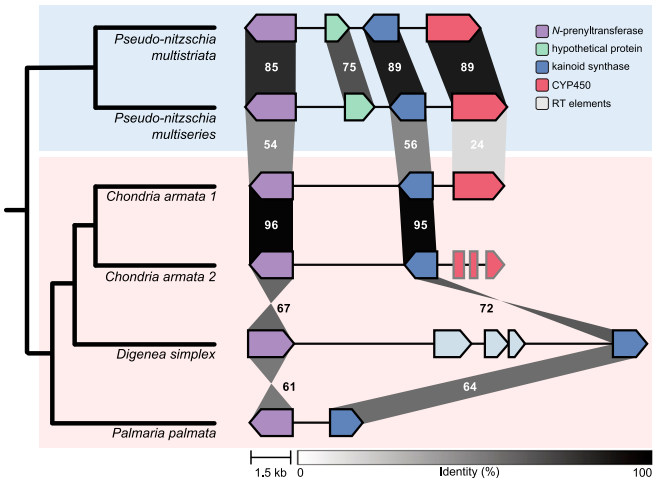

Fig. 2. Visualization of syntenic comparisons between *dab*, *rad*, and *kab* gene clusters. Blue background highlights diatom sequences, and red background highlights red algal sequences. Cladogram connects clusters based on taxonomic relationships of organisms, not to scale with respect to evolutionary 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.

Publicly available *dab* and *kab* gene clusters were used to construct a comparative analysis with new *rad* gene clusters from *C. armata* (Fig. 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 lacking the hypothetical "DabB" protein found in the *dab* clusters, the *rad* and *dab* clusters surprisingly display nearly identical gene organization and orientation (Fig. 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 the conservation of gene location and orientation, the amino acid percent identity between the coclustered CYP450s *radD* and *dabD* decreased to 24%.

In Vitro Validation of Biosynthetic Gene Function. In addition to DA, several additional isomers have also been isolated from *C. armata* and other sources (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 with *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*A*) (*SI Appendix*, Fig. S5). 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).

To determine the basis for differences in isomer production, we examined the newly identified *rad1* cluster from *C. armata*, beginning with kainoid synthase. Full-length RadC1 from the intact *rad1* cluster was expressed as an N-terminal hexahistidine (N-His_6)–soluble construct in *Escherichia coli* (*SI Appendix*, Fig. S6). 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). 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 is yet to be explained biochemically. Although trace quantities of isodomoic acid C are observed from the RadC1 reaction, it remains to be seen whether stereoisomer diversity is also enriched upstream of cyclization through incorporation of neryl diphosphate. Initial expression, purification, and enzymatic assay of RadC2, encoded in the *rad2* cluster, did not reveal substantial activity toward cNGG, although small quantities of isodomoic acid B are produced. As the *rad2* cluster also contains a degraded

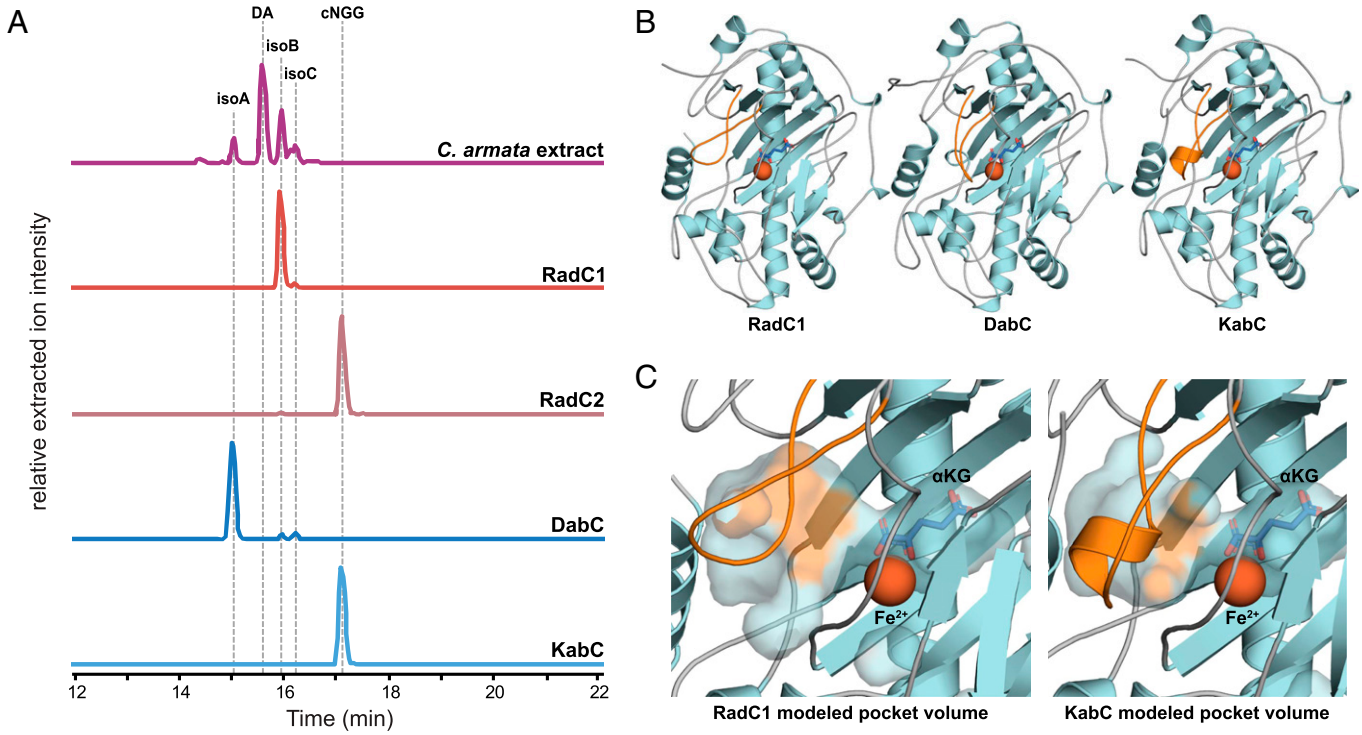

Fig. 3. Activities and predicted structures of the known kainoid synthase enzymes. (*A*) Comparison of combined extracted ion chromatogram profiles for cNGG substrate (*m/z* 312.1 ± 0.2) and isodomoic acid (*m/z* 310.1 ± 0.2) from RadC1, RadC2, DabC, and KabC assays and *C. armata* extract. Relative intensity of extracted ions is shown, and all substrates and observed DA isomers are compared to purified standards (*SI Appendix*, Fig. S5). (*B*) AlphaFold2 predicted models for RadC1, DabC, and KabC. (*C*) Modeled pocket volume for RadC1 and KabC.

copy of the coclustered 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, a set of higher isoprene homologs of kainic acid, from the less-favored NGG substrate. RadC1 appears to approach complete substrate turnover in our preliminary, qualitative in vitro analysis, whereas DabC does not (*SI Appendix*, Fig. S7) (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, GenBank accession: QCC62383.1) was also further investigated. Intriguingly, KabC does not catalyze cyclization on either the cNGG or NGG precursor (Fig. 3 and *SI Appendix*, Fig. S7). As demonstrated previously, KabC 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 (*SI Appendix*, Fig. S8) (18). This inability of KabC 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 AlphaFold2 (34, 35) reveals KabC enzyme models to have an extended loop that pushes further into the active site than in RadC or DabC models (Fig. 3*B*). The quantification of active site area and volume solvent–accessible surface area using CASTp (36) 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*C* and *SI Appendix*, Fig. S9). This follows an expected trend according to each enzyme's observed substrate specificities (Fig. 3*A*).

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 (*SI Appendix*, Fig. S10). 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). Overall, 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 because of its unique amino acid sequence (*SI Appendix*, Fig. S11). 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 NCBI protein database. The closest homologs were primarily from bacteria, including the well-studied αKG-dependent oxidase deacetoxycephalosporin C synthase (*cefE*, UniProtKB/Swiss-Prot identification: 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

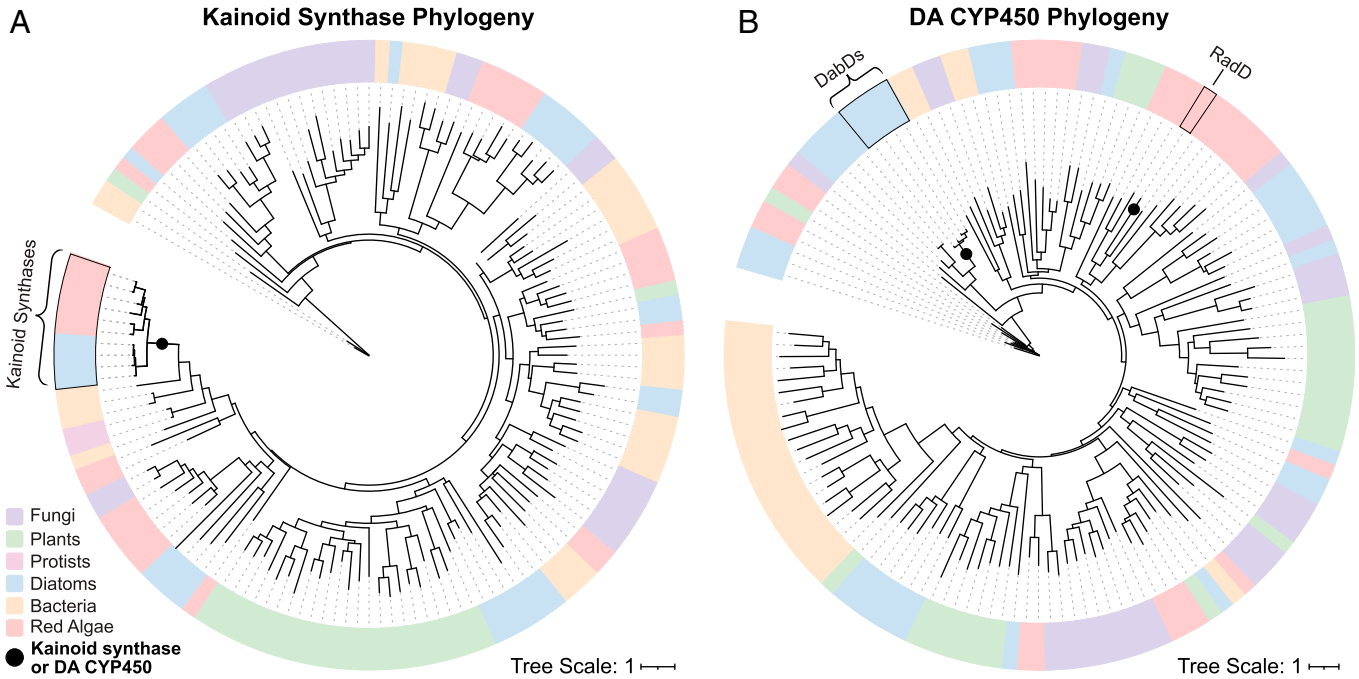

Fig. 4. Phylogenetic analysis of (*A*) kainoid synthase and (*B*) coclustered DA CYP450 enzymes (yellow highlights, yellow circle denotes key branch points). The maximum likelihood trees were built using representative sequences from key taxonomic groups (*SI Appendix*, Tables S5 and S6). Kainoid synthase enzymes form their own distinct branch, independent of taxonomic origin, while the DA CYP450s are nested within their respective taxonomic groups.

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, with all sequence accession numbers listed in *SI Appendix*, Table S5. 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 the hypothetically assigned 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 (*SI Appendix*, Table S6). 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. 4*A*), a finding that is consistent with previous phylogenetic studies on *N*-prenyltransferases (19). Notably, the bacterial enzyme cefE is the only well-characterized close relative to RadC. The monophyletic clading 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 identity 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 toward the installation of the carboxylic acid on the prenyl sidechain (Fig. 4*B*). Based on this taxonomic clustering, the presence of a CYP450 enzyme in DA biosynthesis may be a result of gene duplication and neofunctionalization of 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 found within the genome of *C. armata* or *P. multiseries*, respectively (*SI Appendix*, Figs. S12 and S13). 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 (*SI Appendix*, Fig. S12). The biochemical function of these related enzymes is unknown, and the specific CYP450 family classification of DabD-like enzymes remains to be determined.

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 *Pseudonitzschia* 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 toward 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 ~5% of their functional genes have been acquired from bacteria in relatively recent evolutionary history (43, 44). The 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 an HGT-derived carbonic anhydrase 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). HGT has also been suggested to play a role in the evolution of saxitoxin biosynthesis. Saxitoxin, another notable HAB neurotoxin and the causative agent of paralytic shellfish poisoning, is produced by prokaryotic freshwater cyanobacteria and eukaryotic marine dinoflagellates (50). Genes encoding saxitoxin biosynthesis were first identified in the cyanobacterium *Cylindrospermopsis raciborskii* and later in dinoflagellates belonging to the genus *Alexandrium*, suggesting a potentially ancient HGT event from cyanobacteria to dinoflagellates based on conservation of diagnostic domain architectures between key biosynthetic enzymes (51–53).

Following acquisition of the core kainoid biosynthetic genes, both *C. armata* and *Pseudonitzschia* spp. appear to have repurposed CYP450 enzymology to install the characteristic carboxylic acid on the terpene-derived DA side chain important for toxicity (54). 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 (55). 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 toward 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 oxidation to a carboxylic acid occurring prior to cyclization in both organisms (17). However, until RadD is experimentally characterized, we recognize that this reaction series may differ in red algae given the apparent activity of RadC1 toward NGG. Regardless of reaction order, we hypothesize that a similar biosynthetic reaction toward installing a carboxylic acid on the DA scaffold has evolved independently in diatoms and red algae. An analogous 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 toward DA biosynthesis as reflected by the red algal–derived RadD and diatom-derived DabD.

Evolution toward 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 *Pseudonitzschia* 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 the 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* (56). 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 coclustered CYP450 enzymes DabD and hypothetically assigned RadD, 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 coclustering 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 (57). Previous studies have demonstrated limited trace metal chelation by DA, affording a potential selective advantage for diatoms thriving in iron-limited coastal regimes (58, 59). Other work has suggested a role for DA in grazer defense, particularly with respect to copepod grazing in the diatom system (60). 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 *Pseudonitzschia* diatoms, the evolutionary history described here provides insight to the ecological importance of this marine biotoxin.

Materials and Methods

SI Appendix includes methods for expression, purification, and mutagenesis of all enzymes used in this study. Methodologies for activity assays, sequencing, and phylogenetics are also detailed. Finally, general chemical procedures and small molecule characterization are available in *SI Appendix*. The remaining data and methods are present in *SI Appendix*. The NCBI accession numbers of sequences deposited and used in this study are listed in *SI Appendix*, Table S7.

Data Availability. BioProject, BioSample, Genome, Sequence Read Archive (SRA), and GenBank data have been deposited in NCBI (PRJNA762367, SAMN21392177, JAIWHZ000000000, SRR15927350, SRR15927349, SRR15927348, OK169902, OK169903, OK169904). All other study data are included in the article and/or *SI Appendix*. Previously published data were used for this work (DOI: 10.1126/science.aau0382, https://doi.org/10.1002/anie.201902910).

ACKNOWLEDGMENTS. 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. We acknowledge R. Botts (Point Loma Nazarene University) for his valuable mentorship and guidance. This research was supported by the National Oceanic and Atmospheric Administration (NA19NOS4780181 to B.S.M. and A.E.A.), the NSF through the Graduate Fellowship Research Program to T.S.S., the NIH (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 startup funds to J.R.C.

1. V. L. Trainer *et al.*, *Pseudo-nitzschia* physiological ecology, phylogeny, toxicity, monitoring and impacts on ecosystem health. *Harmful Algae* **14**, 271–300 (2012).
2. T. M. Perl *et al.*, Amnesic shellfish poisoning: A new clinical syndrome due to domoic acid. *Can. Dis. Wkly. Rep.* **16**, 7–8 (1990).
3. O. M. Pulido, Phycotoxins by harmful algal blooms (HABs) and human poisoning: An overview. *Int. Clin. Pathol. J.* **2**, 145–152 (2016).
4. J. L. C. Wright *et al.*, 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 *et al.*, 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" in *Encyclopedia of Biodiversity*, S. A. Levin, Ed. (Elsevier, vol. **2**, 2013), pp. 129–139.
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 *et al.*, Amino acids and low-molecular-weight carbohydrates of some marine red algae. *Phytochemistry* **14**, 1549–1557 (1975).
11. Y. Shimizu *et al.*, Dinoflagellate and other microalgal toxins: Chemistry and biochemistry. *Pure Appl. Chem.* **61**, 513–516 (1989).
12. M. Sato *et al.*, Distribution of neuroexcitatory amino acids in marine algae. *Phytochemistry* **42**, 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 *et al.*, 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 *et al.*, Molecular detection of a potentially toxic diatom species. *Int. J. Environ. Res. Public Health* **12**, 4921–4941 (2015).
17. J. K. Brunson *et al.*, Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018).
18. J. R. Chekan *et al.*, Scalable biosynthesis of the seaweed neurochemical, kainic acid. *Angew. Chem. Int. Ed. Engl.* **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. U.S.A.* **117**, 12799–12805 (2020).
20. R. Kajitani *et al.*, Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nat. Commun.* **10**, 1702 (2019).
21. A. V. Zimin *et al.*, 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.* **27**, 787–792 (2017).
22. M. Seppey, M. Manni, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
23. J. Collén *et al.*, Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5247–5252 (2013).
24. J. Lee *et al.*, 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 *et al.*, High-quality de novo genome assembly of *Kappaphycus alvarezii* based on both PacBio and HiSeq sequencing. bioRxiv [Preprint] (2020). https://doi.org/10.1101/2020.02.15.950402. Accessed 16 June 2021.
26. S. H. Brawley *et al.*, Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc. Natl. Acad. Sci. U.S.A.* **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 (2021).
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 *et al.*, Structures of isodomoic acids A, B and C, novel insecticidal amino acids from the red alga *Chondria armata*. *Chem. Pharm. Bull. (Tokyo)* **34**, 4892–4895 (1986).
31. L. Zaman *et al.*, 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 *et al.*, 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 *et al.*, 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 [Preprint] (2021). https://doi.org/10.21203/rs.3.rs-1032816/v1. Accessed 8 September 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 *et al.*, 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 *et al.*, Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat. Commun.* **10**, 3206 (2019).
39. C. Villard *et al.*, A new P450 involved in the furanocoumarin pathway underlies a recent case of convergent evolution. *New Phytol.* **231**, 1923–1939 (2021).
40. L. Teng *et al.*, 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 *et al.*, The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).
45. A. E. Allen *et al.*, 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 *et al.*, Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**, 1207–1210 (2013).
48. D. Wang *et al.*, *Pyropia yezoensis* genome reveals diverse mechanisms of carbon acquisition in the intertidal environment. *Nat. Commun.* **11**, 4028 (2020).
49. G. Wei *et al.*, Terpene biosynthesis in red algae is catalyzed by microbial type but not typical plant terpene synthases. *Plant Physiol.* **179**, 382–390 (2019).
50. 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**, 991–1018 (2013).
51. R. Kellmann *et al.*, Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. *Appl. Environ. Microbiol.* **74**, 4044–4053 (2008).

52. A. Stüken *et al.*, Discovery of nuclear-encoded genes for the neurotoxin saxitoxin in dinoflagellates. *PLoS One* **6**, e20096 (2011).
53. R. J. S. Orr, A. Stüken, S. A. Murray, K. S. Jakobsen, Evolution and distribution of saxitoxin biosynthesis in dinoflagellates. *Mar. Drugs* **11**, 2814–2828 (2013).
54. Y. Maeno *et al.*, Preparation of domoic acid analogues using a bioconversion system, and their toxicity in mice. *Org. Biomol. Chem.* **19**, 7894–7902 (2021).
55. X. Zhang, S. Li, Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* **34**, 1061–1089 (2017).
56. 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).
57. 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).
58. C. G. Trick *et al.*, Iron enrichment stimulates toxic diatom production in high-nitrate, low-chlorophyll areas. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5887–5892 (2010).
59. 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).
60. S. Harðardóttir *et al.*, 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).