Evolution of Saxitoxin Synthesis in Cyanobacteria and Dinoflagellates

Jeremiah D. Hackett,*[1] Jennifer H. Wisecaver,[1] Michael L. Brosnahan,[2] David M. Kulis,[2] Donald M. Anderson,[2] Debashish Bhattacharya,[3] F. Gerald Plumley,[4] and Deana L. Erdner[5]

[1]Department of Ecology and Evolutionary Biology, University of Arizona

[2]Biology Department Woods Hole Oceanographic Institution

[3]Department of Ecology, Evolution, and Natural Resources, Institute of Marine and Coastal Sciences, Rutgers University

[4]Bermuda Institute of Ocean Sciences, St George's GE01, Bermuda

[5]Department of Marine Science, University of Texas

***Corresponding author**: E-mail: hackettj@email.arizona.edu.

Associate editor: Jennifer Wernegreen

Abstract

Dinoflagellates produce a variety of toxic secondary metabolites that have a significant impact on marine ecosystems and fisheries. Saxitoxin (STX), the cause of paralytic shellfish poisoning, is produced by three marine dinoflagellate genera and is also made by some freshwater cyanobacteria. Genes involved in STX synthesis have been identified in cyanobacteria but are yet to be reported in the massive genomes of dinoflagellates. We have assembled comprehensive transcriptome data sets for several STX-producing dinoflagellates and a related non-toxic species and have identified 265 putative homologs of 13 cyanobacterial STX synthesis genes, including all of the genes directly involved in toxin synthesis. Putative homologs of four proteins group closely in phylogenies with cyanobacteria and are likely the functional homologs of sxtA, sxtG, and sxtB in dinoflagellates. However, the phylogenies do not support the transfer of these genes directly between toxic cyanobacteria and dinoflagellates. SxtA is split into two proteins in the dinoflagellates corresponding to the N-terminal portion containing the methyltransferase and acyl carrier protein domains and a C-terminal portion with the aminotransferase domain. Homologs of sxtB and N-terminal sxtA are present in non-toxic strains, suggesting their functions may not be limited to saxitoxin production. Only homologs of the C-terminus of sxtA and sxtG were found exclusively in toxic strains. A more thorough survey of STX+ dinoflagellates will be needed to determine if these two genes may be specific to SXT production in dinoflagellates. The *A. tamarense* transcriptome does not contain homologs for the remaining STX genes. Nevertheless, we identified candidate genes with similar predicted biochemical activities that account for the missing functions. These results suggest that the STX synthesis pathway was likely assembled independently in the distantly related cyanobacteria and dinoflagellates, although using some evolutionarily related proteins. The biological role of STX is not well understood in either cyanobacteria or dinoflagellates. However, STX production in these two ecologically distinct groups of organisms suggests that this toxin confers a benefit to producers that we do not yet fully understand.

Key words: Saxitoxin, dinoflagellate, evolution, gene transfer, secondary metabolism, toxin.

Article

Introduction

The important role of gene transfer in the evolution of genomes and novel traits in bacteria is well established (Ochman et al. 2000; Pál et al. 2005). Whereas gene duplication is often implicated in the evolution of novel traits in multicellular eukaryotes, recent genome level analyses of microbial eukaryotes have revealed an important role for gene transfer in these organisms as well (Bowler et al. 2008; Keeling and Palmer 2008; Moustafa et al. 2009a). Dinoflagellates are a group of microbial eukaryotes that have acquired a number of striking features through lateral gene transfer such as form II RuBisCo, histone-like proteins, and large numbers of plastid-related genes (Morse et al. 1995; Hackett et al. 2005; Waller et al. 2006). They also play a significant role in marine ecosystems through primary production and the generation of a wide variety of toxic secondary metabolites associated with harmful algal blooms, or HABs (Van Dolah 2000; Hackett et al. 2004).

One of the most significant of the HAB poisoning syndromes is paralytic shellfish poisoning (PSP), a potentially fatal illness that results when humans consume shellfish that have accumulated toxins. PSP is caused by saxitoxins (STXs), a family of heterocyclic guanidines that bind to sodium channels responsible for the flux of sodium in nerve and muscle cells (Cembella 1998). STX is produced by a small number of species in two groups of organisms from different domains of life: cyanobacteria and dinoflagellates.

STX production is found in a small number of cyanobacterial species from several genera including *Anabaena*, *Cylindrospermopsis*, *Lyngbya*, *Aphanizomenon*, *Planktothrix*, and *Raphidiopsis* (Wiese et al. 2010). The complete sequence (~35 kb) of a putative STX biosynthesis gene cluster encoding

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025

© The Author 2012. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

26 proteins (sxtA-sxtZ) in the toxin-producing (STX+) cyanobacterial strain *Cylindrospermopsis raciborskii* T3 was recently determined (Kellmann et al. 2008). Subsequently, the STX gene cluster from several other cyanobacteria has been characterized (Mihali et al. 2009; Moustafa et al. 2009b; Stuken et al. 2010). Eight proteins encoded by these genes (sxtA, sxtG, sxtB, sxtD, sxtS, sxtU, sxtH/T, and sxtI) appear to be directly involved in the synthesis of STX (Kellmann et al. 2008; Pearson et al. 2010). Three additional genes encode proteins (sxtL, sxtN, sxtX) proposed to further modify the STX molecule, producing STX congeners. Phylogenomic analyses of the cyanobacterial STX synthesis genes revealed that some were vertically inherited, but many were acquired through lateral gene transfer from other bacteria (Moustafa et al 2009b).

Many dinoflagellate species from the genus *Alexandrium*, as well as the more distantly related *Gymnodinium catenatum* and *Pyrodinium bahamense* var. *compressum*, produce STX (Oshima et al. 1993; Usup et al. 1994; Sako et al. 2001; Lilly et al. 2005). Toxic species are paraphyletic within the genus *Alexandrium* and there are toxic and non-toxic strains of the same species, although these species designations have been questioned (Scholin et al. 1994; Lilly et al. 2007; Brosnahan et al. 2010). Phylogenetic analyses of LSU rDNA from members of the *Alexandrium tamarense* species complex (including *A. tamarense*, *A. fundyense*, and *A. catenella* morphotypes) reveal five ribosomal species, with these groupings being consistent with toxicity and geographic distribution, rather than with the subtle morphological characteristics by which the species were described (Lilly et al. 2007). Two of the species, Groups I and IV, are uniformly toxic, whereas Groups II, III, and V are uniformly non-toxic. One explanation for a disjunctive distribution of STX is that the ability to produce the toxins may stem from symbiotic bacteria and not the dinoflagellate (Kodama et al. 1988, Gallacher et al. 1997, Vasquez et al. 2001). In contrast, other studies have shown that toxin production remains in the absence of symbiotic bacteria (Hold et al. 2001). In support of the latter hypothesis, genetic analysis of *A. tamarense* Group IV has shown Mendelian inheritance of STX congener profiles, consistent with the idea that the genes encoding this function are encoded on the nuclear genome of *Alexandrium* species and not associated with prokaryotic symbionts (Sako et al. 1992).

STX synthesis has been studied in both dinoflagellates and cyanobacteria using radiotracer experiments that suggest that STX is synthesized from the same precursors (three arginines, one methionine *via* S-adenosylmethionine, and one acetate) using presumably similar biochemical reactions (Shimizu 1993). Given the unique structure of this compound and its narrow distribution in different domains, one hypothesis for the origin of STX synthesis is that toxin synthesis evolved in either cyanobacteria or dinoflagellates and was then acquired by the other through lateral gene transfer. Alternatively, toxin synthesis arose independently in each lineage, converging on a similar product. Convergent or repeated evolution of secondary metabolic pathways has been described in land plants (e.g., Pichersky and Gang 2000; Reimann et al. 2004). This process involves the evolution of secondary metabolite synthesis genes independently in separate lineages, often recruiting the new genes from the same vertically inherited precursor genes.

To test these hypotheses, we assembled a comprehensive transcriptome database for the STX+ dinoflagellate *A. tamarense* CCMP1598 (Group IV), STX+ *A. tamarense* strains SPE10-03, 38-3 and GTM-253-17 (Group I), the STX- *A. tamarense* ATSP1-B (Group III), and lower coverage transcriptomes for the two distantly related STX+ dinoflagellates (*P. bahamense* and *G. catenatum*). Using these data, we identified putative homologs of the STX synthesis genes in dinoflagellates and determined their evolutionary history.

Materials and Methods

454 Transcriptome Sequencing and Assembly

Inoculum cultures of the dinoflagellates *A. tamarense* CCMP1598 (Group IV), *A. tamarense* SPE10-03 (Group I), *Gymnodinium catenatum* GC744, and *Pyrodinium bahamense* CCFW293-B5, were maintained in modified f/2-Si medium (Anderson et al. 1994). All cultures were grown at 15°C on a 14:10 h light:dark cycle (ca. 200 μmol photon s·m-2·sec-1 irradiance provided by cool white fluorescent bulbs). Cell pellets were resuspended in Trizol (Invitrogen) and flash-frozen in liquid nitrogen. Total RNA was extracted according to the Trizol protocol and treated with DNase. First-strand cDNA synthesis was done with the Superscript First Strand (Invitrogen) synthesis kit according to the manufacturer's protocol with the addition of a modified oligo-dT primer (5′ AAG CAG TGG TAT CAA CGC AGA GTT TGT TTT TTT TTC TTT TTT TTT TVN 3′). Second-strand synthesis was done with a primer containing the Clontech primer 2A sequence 5′ and the dinoflagellate-specific 5′-trans-spliced leader sequence (5′ AAG CAG TGG TAT CAA CGC AGA GTT CCG TAG CCA TTT TGG CTC AAG 3′). This primer takes advantage of the 5′ trans-spliced leader sequence on mature dinoflagellate transcripts to produce full-length cDNAs (Lidie and Van Dolah 2007; Zhang et al. 2007). Second strand cDNA was purified using the Qiagen PCR purification kit and amplified with the Advantage 2 PCR kit (Clontech). The 454 sequencing libraries were prepared with 5 μg of cDNA and one full picotitre plate for each species was sequenced on a 454 FLX pyrosequencer with Titanium chemistry at the University of Arizona Genetics Core.

The 454 data were assembled using a custom cleaning, clustering, and assembling pipeline developed for 454 transcriptome sequences. Raw reads were first cleaned using SnoWhite v1.1.3 (http://www.evopipes.net, last accessed October 8, 2012) with the TagDust cleaning step implemented (Lassmann et al. 2009). The wcd system clustered the cleaned reads using an error threshold value of 5 (Hazelhurst et al. 2008). Based on the wcd output, a Perl script split the data set into multiple fasta and associated quality files based on cluster sizes (i.e., read depth). The multiple files were assembled separately using Mira v3.0.0 (Chevreux et al. 1999). The Mira outputs were assembled using CAP3 with an overlap percent identity cutoff of 94 (Huang and Madan 1999).

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025

Illumina RNA-seq Sequencing and Assembly

Illumina RNA-seq data was produced for *A. tamarense* CCMP1598 (Group IV), using cultures grown under 4 different conditions. Cells were grown in batch culture in nutrient-replete f/2-Si media and harvested in late-exponential phase. Cells were also grown under semi-continuous conditions in f/2-Si and in f/2-Si substituting ammonium for nitrate. These cultures were maintained in a semi-continuous mode by replacing a volume of culture with fresh medium each day such that a growth rate of 0.4 divisions·day^{-1} was achieved. Cells were harvested when the cell densities were found to be in steady state for several days. Finally, cells grown in f/2-Si were synchronized and harvested in the G1 cell cycle stage as previously described (Taroncher-Oldenburg et al. 1997). All cultures were harvested by centrifugation and the cell pellets were resuspended in 1 ml of Trizol and then flash-frozen in liquid nitrogen. *Alexandrium tamarense* 38-3 and GTM253-17 (Group I) and *A. tamarense* ATSP1-B (Group III) cell pellets were taken from late-log phase cultures grown in f/2 -Si and encystment medium (Brosnahan et al. 2010).

Fragments from polyadenylated RNAs were isolated and prepared for paired-end sequencing using a mRNA-seq reagent kit from Illumina (RS-100-0801) according to the manufacturer's instructions except that ~300 bp fragment-primer products were isolated during size selection in order to avoid overlapping paired-end reads. Sequencing of one lane for each library was completed on an Illumina Genome Analyzer 2.0 by the Biomicro Center at MIT (Cambridge, MA). Read lengths up to 108 bp were attempted but no base calls after position 72 were used in the assembly because of poor estimated read quality. The RNA-seq data from *A. tamarense* CCMP1598 were quality trimmed and assembled with the 454 data (see above) and previously sequenced expressed sequence tags (Moustafa et al. 2010) using the CLC Genomics Workbench (CLC Bio) to produce a reference transcriptome. Illumina data for *A. tamarense* Group 1 strains (38-3 and GTM-253-17) was assembled with 454 data from *A. tamarense* SPE10-3 using CLC Genomics Workbench to produce a composite Group 1 transcriptome.

Illumina data for *A. tamarense* Group III strain were assembled using ABySS v. 1.2.2 (Simpson et al. 2009). Assembly parameters were set according to Birol et al. (2009). A single value of k (k-mer length) was selected by maximizing the frequency of identical matches to MPSS tag libraries (Erdner and Anderson 2006; Moustafa et al. 2010).

Identification and Phylogenetic Analysis of STX Proteins

Protein sequences were predicted from the nucleotide assemblies of data produced in this study using FrameDP (Gouzy et al. 2009). The protein sequences from the Group IV *A. tamarense* assembly were searched against the CDD database using rpsblast (Marchler-Bauer et al. 2009). These results were queried for proteins containing the CDD domains found in the cyanobacterial STX genes. Proteins with a CDD hit *e*-value $< 1e^{-5}$ were used in phylogenetic analyses. Additional STX-related genes were identified with BLAST sequence similarity searches (tblastn, *e*-value $< 1e^{-5}$) using the cyanobacterial proteins as queries against the dinoflagellate transcriptomes. A total of 265 putative homologs of 13 cyanobacterial toxin genes were identified using these methods (Table 1). Nucleotide sequences of transcripts used in phylogenetic analyses presented in the figures have been deposited in Genbank (accession numbers JV310009–JV310320).

The evolutionary history of all putative homologs was determined using a custom phylogenomics pipeline (modified from Moustafa et al. 2008). The predicted amino acid sequences from the 265 putative homologs were queried using BLAST against a local protein database composed of representative sequences from Joint Genome Institute, NCBI's refseq and dbest, as well as the FrameDP predicted proteins from all dinoflagellate assemblies in this study (see supplementary Table S1, Supplementary Material online). For every BLAST result, full-length sequences were extracted from the database and aligned with the query with MAFFT using the L-INS-i strategy (Katoh et al. 2005). Poorly aligned positions were removed from the alignments with Gblocks set to the least stringent trimming options (Castresana 2000). Phylogenetic trees were inferred using RAxML rapid bootstrapping (100 replications) and maximum likelihood search assuming a WAG amino acid model of substitution and Γ site heterogeneity model (Stamatakis 2006). The most taxonomically specific sister group for each protein (ML bootstrap support > 75) was determined for each protein (supplementary Table S2, Supplementary Material online). Supplementary file 1 contains the tree of highest likelihood with bootstrap results for all proteins not presented in figs. 2 and 3 or in supplementary figures, Supplementary Material online.

Results and Discussion

The transcriptome assembly for *A. tamarense* Group IV included a total 10,645 Mb of sequence data comprising 112,383 contigs with a cumulative length of 98.32 Mb. To determine if this *A. tamarense* Group IV transcriptome assembly represented a comprehensive transcriptome data set, we assembled the RNA-seq libraries in all possible combinations of two (six assemblies), three (four assemblies), and four (one assembly) data sets using Abyss as described above. We compared the cumulative length of the raw data to that of the assembled contigs and showed that as the assemblies are constructed with increasing amounts of sequence data, the cumulative length of the assembly approaches an asymptote at 107 Mb (fig. 1). These results suggest that our reference assembly of the transcriptome of *A. tamarense* Group IV, with a length of 98.32 Mb, represents ~92% of the total transcriptome. Previous analyses of the *A. tamarense* transcriptome using MPSS revealed only a handful of condition-specific transcripts in this organism (Erdner and Anderson 2006; Moustafa et al. 2010), suggesting that the small number of conditions under which the cultures for this study were grown did not preclude the discovery of a large numbers of genes.

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025

Table 1. Number of Candidate STX Genes Identified in *A. tamarense* Group IV by CDD and BLAST Search.

Cyanobacterial gene	Conserved domains (CDD)	*A. tamarense* contigs
sxtA—N-terminus	No conserved domains	0[4]
sxtA—C-terminus	cd06454, Aspartate aminotransferase	12
sxtB	PRK09027, Cytidine deaminase	2
sxtC	No conserved domains	0
sxtD	COG3000, ERG3 Sterol desaturase	18
sxtE	No conserved domains	0
sxtF/M	PRK01766, multidrug efflux protein	28
sxtG	pfam02274, Amidinotransferase	2[2]
sxtH/T	COG4638, Phenylpropionate dioxygenase	18
sxtI	COG2192, NodU family carbamoyl transferase	2
sxtJ	No conserved domains	0
sxtK	No conserved domains	0
sxtL	cd00229, SGNH hydrolase	3
sxtN	pfam00685, Sulfotransferase	26
sxtP	No conserved domains	0[1]
sxtQ	pfam08892, YqcI/YcgG family	0
sxtR	No conserved domains	0
sxtS	COG5285, Phytanoyl-CoA dioxygenase	14
sxtU	COG4221, Short-chain alcohol dehydrogenase	128
sxtW	No conserved domains	0
sxtX	pfam04989, Cephalosporin hydroxylase	5
sxtY	TIGR02135, phoU	0

Note.—Numbers in brackets indicate contigs identified exclusively through BLAST search.

If the STX biosynthetic pathway evolved independently in cyanobacteria and dinoflagellates, BLAST searches may not identify all potential functional homologs of the cyanobacterial genes. To expand our search to proteins that may have similar biochemical activity, but little sequence similarity, we annotated our *A. tamarense* Group IV transcriptome using the CDD database and identified proteins with domains found in the respective cyanobacterial proteins. Using both approaches, we identified 265 putative homologs of 13 cyanobacterial STX genes in the *A. tamarense* Group IV transcriptome (Table 1). These genes comprise all of the genes directly implicated in toxin synthesis (sxtA, sxtG, sxtB, sxtD, sxtS, sxtU, sxtH/T, and sxtI), the three proteins proposed to be involved in the modification of saxitoxin into related congeners (sxtL, sxtN, and sxtX), potential toxin transporters (sxtF/M) and a STX-binding protein (sxtP). Eight cyanobacterial proteins (sxtC, sxtE, sxtJ, sxtK, sxtQ, sxtR, sxtW, and sxtY) did not contain conserved domains and did not have any significant BLAST hits against the *A. tamarense* Group IV transcriptome. Our search strategy likely identified many genes that are not involved in toxin synthesis. However, because of the comprehensive nature of the transcriptome data set (fig. 1), we suggest that it is unlikely that additional, undiscovered homologs of the cyanobacterial genes are expressed. We analyzed the predicted protein sequences using

Fig. 1. Graph depicting the relationship between the amount of Illumina input data and the cumulative assembly length for the *de novo* dinoflagellate transcriptome assemblies.

our phylogenomic pipeline, which included the other dinoflagellate transcriptomes in this study, to determine the evolutionary history of each protein. If the STX synthesis pathway evolved once in either dinoflagellates or cyanobacteria and was then transferred to the other lineage, we expect cyanobacterial and dinoflagellate homologs to group closely in phylogenetic analyses. If the pathways were assembled independently, we expect these proteins to show divergent evolutionary origins.

The cyanobacterial gene sxtA encodes a polyketide synthase proposed to perform the first step in toxin synthesis, loading the acyl carrier protein (ACP) domain with acetate from acetyl-CoA and then methylating acetyl-ACP to propionyl-ACP (Kellmann et al. 2008). The aminotransferase domain of sxtA then performs a Claisen condensation of propionyl-ACP with arginine. Previous analyses have shown that cyanobacterial sxtA resulted from the fusion of two proteins acquired from different Eubacterial sources (Moustafa et al. 2009b). The N-terminus, which contains the acetyltransferase and ACP domain with a phosphopantetheinyl-attachment site (PP-binding), originated from a proteobacterium, whereas the C-terminus contains a Class II aminotransferase domain and was likely acquired from an actinobacterium.

Our analysis of dinoflagellate transcriptomes shows that the two sxtA domains are not fused in dinoflagellates and are encoded on different proteins. We identified contigs using BLAST in *A. tamarense* Group IV, *A. tamarense* Group I, *P. bahamense*, and STX- *A. tamarense* Group III, with sequence similarity to the N-terminal portion of sxtA containing an acyltransferase and PP-binding domain. Our search also included all available dinoflagellate ESTs deposited in Genbank. In these data, we indentified ESTs from *A. catenella* with similarity to this region of sxtA. The dinoflagellate proteins form a strongly supported monophyletic group within a

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025

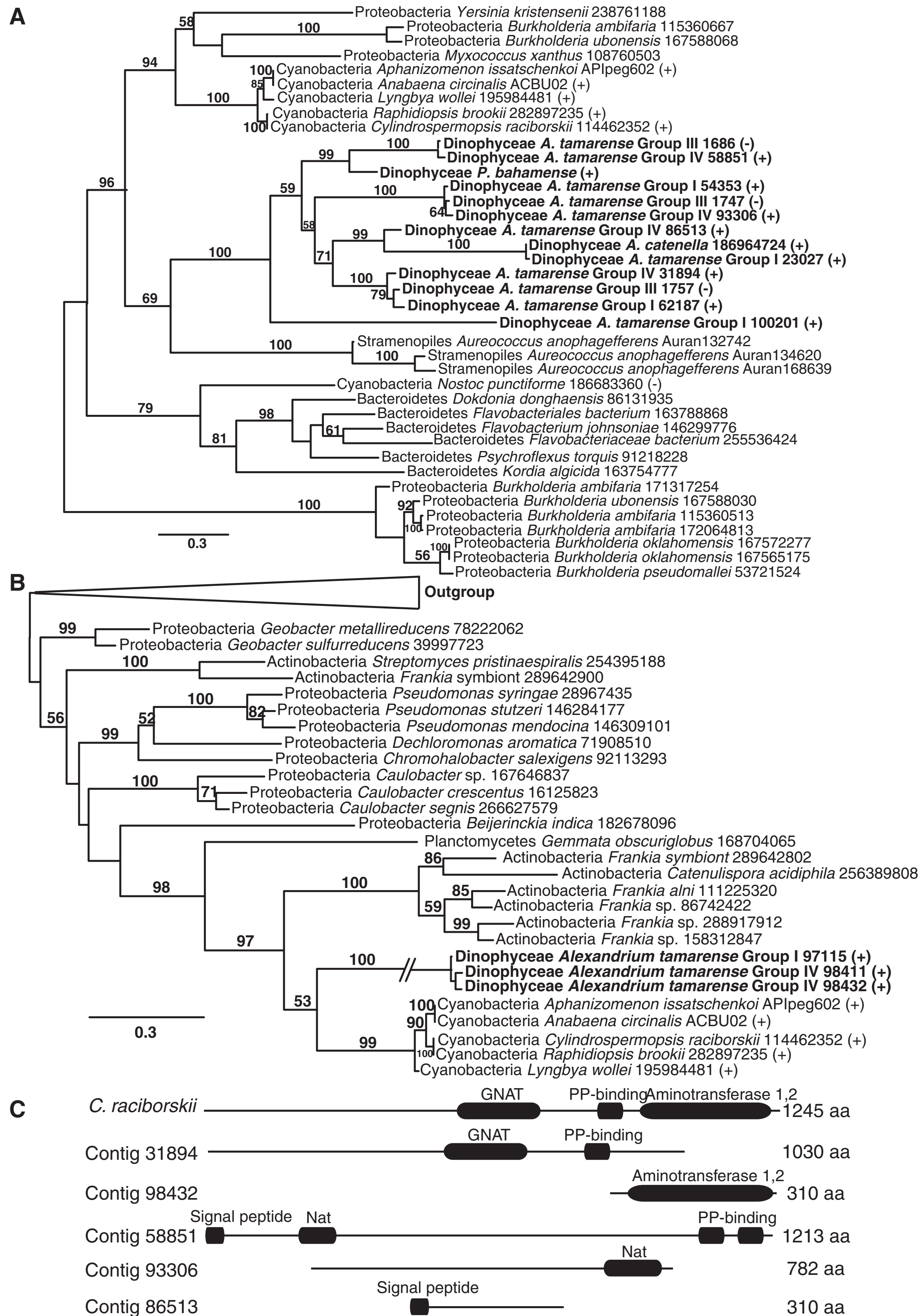

Fig. 2. (*A*) Partial maximum likelihood (ML) phylogeny of the N-terminal region of sxtA and candidate proteins containing the amidinotransferase domain. The full tree is in supplementary fig. S1, Supplementary Material online. For all phylogenies, numbers at the branches show the results of 100 bootstrap replicates (<50). The names at the tips indicate the taxonomy, species name, and GI or contig number of the sequences. STX-producing dinoflagellates are indicated in bold type. The (+) indicates a STX-producing organism. (*B*) ML phylogeny of the C-terminal region of sxtA and candidate proteins. (*C*) Domain structure of sxtA-related proteins in *A. tamarense* Group 4 compared to sxtA from *Cylindrospermopsis raciborskii* T3 (GI: 114462352). Protein domains were predicted using InterPro Scan. GNAT, GCN5-related *N*-acetyltransferase (IPR000182); PP-binding, Phosphopantetheine-binding (IPR006163); Aminotransferase 1,2, Aminotransferase class I/II (IPR004839); Nat, Acyl-CoA N-acyltransferases (IPR016181). Signal peptides were predicted by SignalP in InterproScan.

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025

clade containing Stramenopiles, Proteobacteria, and the N-terminal portion of sxtA from toxic cyanobacteria (fig. 2A).

Searches against the CDD database identified 12 contigs in *A. tamarense* Group IV that contain an aminotransferase domain similar to that in cyanobacterial sxtA (Table 1). Most of these clustered with other eukaryotes in phylogenetic analyses (supplementary Table S2, Supplementary Material online). Two contigs clustered closely with STX+ cyanobacteria, forming a monophyletic group with a protein from STX+ *A. tamarense* Group I within a well-supported clade containing actinobacteria and the C-terminal portion of sxtA from STX+ cyanobacteria (fig. 2B and supplementary fig. S1, Supplementary Material online). We could not identify any putative homologs of these proteins in the non-toxic *A. tamarense* Group III.

Only one of the *A. tamarense* proteins corresponding to the N-terminal portion of sxtA (Contig 31894, fig. 2C) contains the same predicted domain structure as cyanobacterial sxtA with a GCN5-related N-acyltransferase (GNAT) domain. The other proteins have a different predicted domain structure, including some with signal peptides. *A. tamarense* Group IV Contig 31,894 seems to be the most likely functional homolog of the N-terminus of sxtA in this organism based on this predicted structure. The other proteins, although clearly related to sxtA, may either participate in STX synthesis in some unknown fashion or have been modified for another process. Proteins with similarity to the N-terminal portion were indentified in both STX+ and STX- strains, as well in the stramenopile *Aureococcus anophagefferens*. This distribution suggests that this protein is not specific to STX synthesis and has been adapted for other uses in a wide diversity of organisms. In contrast, proteins similar to the C-terminus were found only in toxic *A. tamarense* species.

The sxtA-related proteins in dinoflagellates are not fused and are excluded from a sister relationship with cyanobacterial the N-terminal sxtA tree (fig. 2A). Together, these results suggest that these genes were not exchanged directly between cyanobacteria and dinoflagellates. The strong monophyly of sxtA-related proteins from STX+ dinoflagellates and, separately, from cyanobacteria supports the hypothesis that precursors of these *sxt*A genes were acquired independently by each lineage.

We also found contigs predicted to encode proteins containing an amidinotransferase domain, present in sxtG, the second enzyme in the STX biosynthetic pathway. This protein is proposed to incorporate an amidino group from a second arginine molecule into the STX intermediate (Kellmann et al. 2008). One of these proteins (Contig 87049) from *A. tamarense* Group IV clusters with homologs from the three other toxic dinoflagellates in a poorly supported clade containing sxtG from toxic cyanobacteria (fig. 3A). The dinoflagellate proteins are excluded from a sister group relationship with the cyanobacterial sxtG proteins by a supported relationship with protein from *Beggiatoa* sp. The dinoflagellate proteins appear to share a common ancestor with the cyanobacterial homologs, but were acquired independently by each lineage prior to the evolution of saxitoxin synthesis. Another *A. tamarense* Group IV protein (Contig 8098) falls in an unsupported position sister to an actinobacterium outside of this clade (fig. 3A). The other two amidinotransferase candidates from *A. tamarense* cluster either in a well-supported position within a clade of Eubacterial arginine deaminases (Contig 22,175) or in a group of homologs from other STX+ dinoflagellates outside the clade shown in fig. 2A (supplementary fig. S2, Supplementary Material online).

Only two contigs were found in the *A. tamarense* Group IV transcriptome that encode proteins with a cytidine deaminase domain as in sxtB, which is predicted to perform the third step in toxin synthesis. These group in a strongly supported clade with proteins from *A. tamarense* Group I and STX-Group III, as well as proteins from two Eubacteria and sxtB proteins from STX+ cyanobacteria (fig. 3B). As with sxtA and sxtG, the putative sxtB proteins in dinoflagellates show similarity to the cyanobacterial proteins, but phylogenetic analyses do not support the transfer of these genes directly between cyanobacteria and dinoflagellates.

Searches for proteins with amino acid sequence or domain similarity to sxtD, sxtS, sxtU, sxtH/T, and sxtI produced many candidate proteins (Table 1), but none with a close phylogenetic relationship to proteins from toxic cyanobacteria. Supplementary figs. S3–S6, Supplementary Material online, show trees from phylogenomic analyses that contain the remaining cyanobacterial genes directly implicated in toxin synthesis (sxtS, sxtU, sxtH/T, and sxtI). We did not identify any homologs of the sterol desaturase sxtD of Eubacterial origin in any of our dinoflagellate transcriptomes. Phylogenies for two additional saxitoxin-related cyanobacterial genes, sxtF/M and sxtS, which contain dinoflagellate sequences are shown in supplementary figs. S7 and S8.

Conclusion

We assembled a high-coverage, comprehensive transcriptome for the STX+ *A. tamarense* Group IV and identified putative homologs of all cyanobacterial genes predicted to be directly involved in the synthesis of STX. Phylogenetic analyses show that only proteins involved in the first three steps of toxin synthesis in *A. tamarense* are closely related to the cyanobacterial proteins. Since proteins involved in these three steps have numerous homologs among a diverse group of organisms, it is unlikely that the biochemistry of these three steps is unique to saxitoxin synthesis. Rather it is combination of these enzymes in a particular sequence on the precursors of saxitoxin that is unique to this pathway. The sxtA, sxtG, and sxtB in cyanobacteria and dinoflagellates seem to be derived from common ancestral proteins. It is unlikely that these ancestral were involved in saxitoxin synthesis since they are also found in many organisms that are not known to make saxitoxin and are only found together in saxitoxin-producing cyanobacteria and dinoflagellates. The phylogeny of the C-terminus of sxtA is inconclusive and is the most likely candidate for direct lateral transfer from STX+ cyanobacteria to dinoflagellates. It would likely have been transferred prior to the fusion of sxtA in cyanobacteria since gene fission is much less frequent than gene fusion (Kummerfeld and Teichmann, 2005, Durrens et al. 2008). The sxtG tree

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025

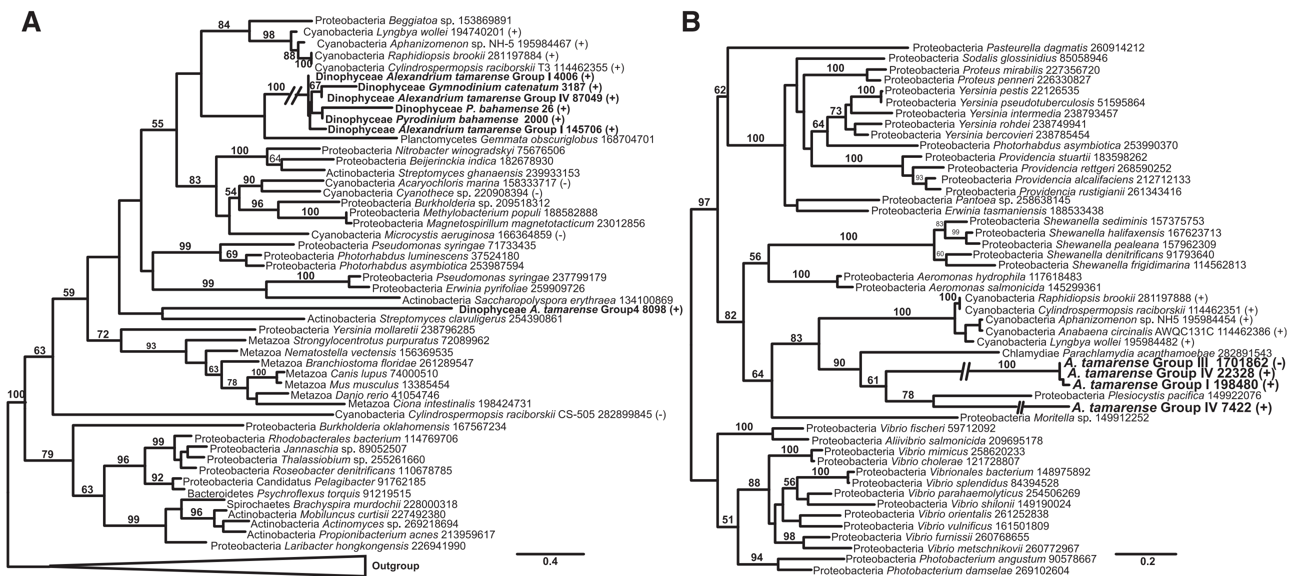

Fig. 3. Format is the same as in fig. 1. (*A*) Partial ML phylogeny of sxtG candidate proteins, the complete tree is in supplementary fig. S2, Supplementary Material online. (*B*) ML phylogeny of sxtB candidate proteins.

(fig. 3*A*) also does not conclusively exclude the possibility of direct transfer between STX+ cyanobacteria and dinoflagellates, if the *Beggiatoa* sp. protein resulted from an independent transfer from cyanobacteria to this species. These results may indicate that it is the C-terminal portion of sxtA and sxtG that are critical for saxitoxin synthesis and that the remaining steps in the synthesis pathway can be performed by other enzymes.

Interestingly, homologs of the C-terminus of sxtA and sxtG have been found exclusively in toxic dinoflagellate species, with sxtG also found in *G. catenatum* and *P. bahamense*. The single non-toxic species examined here contained closely related homologs of N-terminal sxtA and sxtB. Since no strong candidates for the remaining saxitoxin genes were indentified in the toxic species, their distribution was not examined in the non-toxic *A. tamarense* Group III. A more thorough survey of STX+ and STX- dinoflagellates will be needed to determine if C-terminal sxtA and sxtG are specific to STX producing dinoflagellates.

It is still unclear if saxitoxin is synthesized by the same biochemical mechanisms in both dinoflagellates and cyanobacteria. It is possible that the later steps in the pathway are performed with very different biochemical reactions or enzymes, in which case, the genes involved in those steps would not have been identified by our analysis. Also, our results do not exclude the possibility that some of the steps of toxin synthesis could be performed or facilitated by associated bacteria. The biological role of STX is not well understood in either cyanobacteria or dinoflagellates. However, the evolution of STX production in these two ecologically distinct groups of organisms suggests that this toxin confers a benefit to producers that we do not yet fully understand. Functional genetic studies, such as expression of candidate genes in recombinant bacteria, will be required to definitively prove these genes play a role in STX synthesis. Currently, there are no genetic tools available for producing gene knock-outs in dinoflagellates. However, our analysis the *A. tamarense* transcriptome has identified strong candidates for many of the genes involved in toxin synthesis.

Supplementary Material

Supplementary file 1, figures S1–S8, and tables S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors would like to thank Katrina Dlugosh and Mike Barker for assistance with the SnoWhite software and 454 assembly. This work was funded by a collaborative grant from the NSF (EF-0732440) to F.G.P., D.B., J.D.H., and D.M.A., by an NSF grant (OCE-0723498) to J.D.H., D.L.E., D.B., and D.M.A., by an NIEHS grant (R01 ES013679-01A2) to D.B. and D.M.A., and by the Woods Hole Center for Oceans and Human Health through NSF (OCE-0430724) and NIEHS (1 P50 ES012742). Following the submission of this article, Stüken and colleagues (2011) published a paper describing the first discovery of dinoflagellate saxitoxin genes in *Alexandrium fundyense*, including sxtA as a single protein. Our data indicated, at the writing of this manuscript that sxtA was split into two separate transcripts. Reanalysis of our data in light of the findings of Stüken et al. show that our data are consistent with their discovery of sxtA as a single protein.

References

Anderson DM, Kulis DM, Doucette GJ, Gallager JC, Balech E. 1994. Biogeography of toxic dinoflagellates in the genus *Alexandrium* from the northeast United States and Canada as determined by morphology, bioluminescence, toxin composition, and mating compatibility. *Mar Biol.* 120:467–478.

Birol I, Jackman SD, Nielsen CB, et al. (15 co-authors). 2009. De novo transcriptome assembly with ABySS. *Bioinformatics* 25:2872–2877.

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025

Bowler C, Allen AE, Badger JH, et al. (77 co-authors). 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–244.

Brosnahan ML, Kulis DM, Solow AR, Erdner DL, Percy L, Lewis J, Anderson DM. 2010. Outbreeding lethality between toxic Group I and nontoxic Group III Alexandrium tamarense spp. isolates: predominance of heterotypic encystment and implications for mating interactions and biogeography. *Deep-Sea Res. II* 57(3–4):175–189.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17: 540–552.

Cembella AD. 1998. Ecophysiology and metabolism of paralytic shellfish toxins in marine microalgae. In: Anderson DM, Cembella AD, Hallegraeff GM, editors. Physiological ecology of harmful blooms. Berlin, Heidelberg: Springer. p. 381–403.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. *Comp Sci Biol GCB.* 99:45–56.

Durrens P, Nikolski M, Sherman D. 2008. Fusion and fission of genes define a metric between fungal genomes. *PLoS Comp Biol.* 4: e1000200.

Erdner DL, Anderson DM. 2006. Global transcriptional profiling of the toxic dinoflagellate *Alexandrium fundyense* using massively parallel signature sequencing. *BMC Genomics* 7:88.

Gallacher S, Flynn KJ, Franco JM, Brueggemann EE, Hines HB. 1997. Evidence for production of paralytic shellfish toxins by bacteria associated with *Alexandrium* spp. (Dinophyta) in culture. *App Env Microbiol.* 63:239–245.

Gouzy J, Carrere S, Shiex T. 2009. FrameDP: sensitive peptide detection on noisy matured sequences. *Bioinformatics* 25:670–671.

Hackett JD, Anderson D, Erdner D, Bhattacharya D. 2004. Dinoflagellates: a remarkable evolutionary experiment. *Am J Botany.* 91:1523–1534.

Hackett JD, Scheetz TE, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Bhattacharya D. 2005. Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics* 6:80.

Hazelhurst S, Hide W, Liptak Z, Nogueira R, Starfield R. 2008. An overview of the wcd EST clustering tool. *Bioinformatics* 24:1542–1546.

Hold G, Smith EA, Rappe MS, Maas EW, Moore ERB, Stroempl C, Stephen JR, Prosser JI, Birkbeck TH, Gallacher S. 2001. Characterization of bacterial communities associated with toxic and non-toxic dinoflagellates: *Alexandrium* spp. and *Scrippsiella trochoidea*. *FEMS Microb Ecol.* 37:161–173.

Huang XQ, Madan A. 1999. Cap3: A DNA Sequence Assembly Program. *Genome Research* 9:868–877.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33:511–518.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet.* 9:605–618.

Kellmann R, Mihali TK, Jeon YJ, Pickford R, Pomati F, Neilan BA. 2008. Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. *App Environ Microbiol.* 74: 4044–4053.

Kodama M, Ogata T, Sato S. 1988. Saxitoxin-producing bacterium isolated from Protogonyaulax tamarensis. In: Okaichi T, Anderson DM, Nemoto T, editors. Red tides: biology, environmental science and toxicology. New York: Elsevier. p. 363–366.

Kummerfeld SK, Teichmann SA. 2005. Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* 21:25–30.

Lassmann T, Hayashizaki Y, Daub CO. 2009. TagDust-a program to eliminate artifacts from next generation sequencing data. *Bioinformatics* 25:2839–2840.

Lidie KL, Van Dolah FM. 2007. Spliced leader RNA-mediated trans-splicing in a dinoflagellate, Karenia brevis. *J Euk Microbiol.* 54: 427–435.

Lilly EL, Halanych KM, Anderson DM. 2005. Phylogeny, biogeography, and species boundaries within the *Alexandrium minutum* group. *Harmful Algae* 4:1004–1020.

Lilly EL, Halaynch KM, Anderson DM. 2007. Species boundaries and global biogeography of the *Alexandrium tamarense* complex (Dinophyceae). *J Phycol.* 43:1329–1338.

Marchler-Bauer A, Anderson JB, Chitsaz F, et al. (28 co-authors). 2009. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* 37:D205–D210.

Mihali TK, Kellmann R, Neilan BA. 2009. Characterisation of the paralytic shellfish toxin biosynthesis gene clusters in *Anabaena circinalis* AWQC131C and *Aphanizomenon* sp. NH-5. *BMC Biochem.* 10:8.

Morse D, Salois P, Markovic P, Hastings JW. 1995. A nuclear-encoded-form II RuBisCO in dinoflagellates. *Science* 268:1622–1624.

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009a. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324:1724–1726.

Moustafa A, Chan CX, Danforth M, Zear D, Ahmed H, Jadhav N, Savage T, Bhattacharya D. 2008. A phylogenomic approach for studying plastid endosymbiosis. *Genome Informatics* 21:165–176.

Moustafa A, Erdner DL, Kulis DM, Anderson DM, Bhattacharya D. 2010. Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence. *PLoS One* 5:e9688.

Moustafa A, Loram JE, Hackett JD, Anderson DM, Plumley FJ, Bhattacharya D. 2009b. Origin of saxitoxin biosynthetic genes in cyanobacteria. *PLoS One* 4:e5758.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.

Oshima YH, Itakura K-C, Lee T, Yasumoto S, Blackburn G, Hallegraeff G. 1993. Toxin production by the dinoflagellate *Gymnodinium catenatum*. In: Smayda TJ, Shimizu Y, editors. Toxic phytoplankton blooms in the sea. New York: Elsevier. p. 907–912.

Pál C, Papp B, Lercher MJ. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet.* 37: 1372–1375.

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

Pichersky E, Gang DR. 2000. Genetics and biochemistry ofsecondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 5:1360–1385.

Reimann A, Nurhayati N, Backenkohler A, Ober D. 2004. Repeated evolution of the pyrrolizidine alkaloid-mediated defense system in separate angiosperm lineages. *Plant Cell* 16:2772–2784.

Sako Y, Kim C-H, Ishida Y. 1992. Mendelian inheritance of paralytic shellfish poisoning toxin in the marine dinoflagellate *Alexandrium catenella*. *Biosci Biotech Biochem.* 56:692–694.

Sako Y, Yoshida T, Uchida A, Arakawa O, Noguchi T, Ishida Y. 2001. Purification and characterization of a sulfotransferase specific to N-21 of saxitoxin and gonyautoxin 2 + 3 from the toxic

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025

dinoflagellate *Gymodinium catenatum* (Dinophyceae). *J Phycol.* 37: 1044–1051.

Scholin CA, Herzog M, Sogin M, Anderson DM. 1994. Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae) species. II. Sequence analysis of a fragment of the large-subunit ribosomal RNA gene. *J Phycol.* 30: 999–1011.

Shimizu Y. 1993. Microalgal metabolites. *Chem Rev.* 93:1685–1698.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19: 1117–1123.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

Stüken A, Jakobsen KS. 2010. The cylindrospermopsin gene cluster of *Aphanizomenon* sp. strain 10E6: organization and recombination. *Microbiology* 156:2438–2451.

Stüken A, Orr RJS, Kellmann R, Murray SA, Neilan BA, Jakobsen KS. 2011. Discovery of Nuclear-Encoded Genes for the Neurotoxin Saxitoxin in Dinoflagellates. *PLoS ONE* 6:e20096.

Taroncher-Oldenburg G, Kulis DM, Anderson DM. 1997. Toxin variability during the cell cycle of the dinoflagellate *Alexandrium fundyense*. *Limnol Oceanogr.* 42:1178–1188.

Taroncher-Oldenburg G, Kulis DM, Anderson DM. 1999. Coupling of saxitoxin biosynthesis to the G1 phase of the cell cycle in the dinoflagellate Alexandrium fundyense: temperature and nutrient effects. *Nat Toxins.* 7:207–219.

Usup G, Kulis DM, Anderson DM. 1994. Growth and toxin production of the toxic dinoflagellate *Pyrodinium bahamense* var. *compressum* in laboratory cultures. *Nat Toxins.* 2:254–262.

Waller RF, Slamovitis CH, Keeling P. 2006. Lateral gene transfer of a multigene region from cyanobacteria to dinoflagellates resulting in a novel plastid-targeted fusion protein. *Mol Biol Evol.* 23: 1437–1443.

Wiese M, D'Agostino PM, Mihali TK, Moffitt MC, Neilan BA. 2010. Neurotoxic alkaloids: saxitoxin and its analogs. *Mar Drugs.* 8: 2185–2211.

Van Dolah F. 2000. Marine algal toxins: origins, health effects, and their increased occurrence. *Env Health Persp.* 108:133–141.

Vasquez M, Gruttner C, Gallacher S, Moore ERB. 2001. Detection and characterization of toxigenic bacteria associated with *Alexandrium catenella* and *Aulacomya ater* contaminated with PSP. *J Shellfish Res.* 20:1245–1249.

Zhang H, Hou YB, Miranda L, Campbell DA, Sturm NR, Gaasterland T, Lin SJ. 2007. Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci U S A.* 104:4618–4623.

Downloaded from https://academic.oup.com/mbe/article/30/1/70/1018966 by guest on 19 April 2025