Extraordinary Conservation, Gene Loss, and Positive Selection in the Evolution of an Ancient Neurotoxin

Shauna A. Murray,†[1], Troco Kaan Mihali,†[1] and Brett A. Neilan*[1]

[1]School of Biotechnology and Biomolecular Sciences, University of New South Wales, New South Wales, Australia

†These authors contributed equally to this work.

***Corresponding author**: E-mail: b.neilan@unsw.edu.au

Abstract

The recent determination of the genetic basis for the biosynthesis of the neurotoxin, saxitoxin, produced by cyanobacteria, has revealed a highly complex sequence of reactions, involving over 30 biosynthetic steps encoded by up to 26 genes clustered at one genomic locus, *sxt*. Insights into evolutionary–ecological processes have been found through the study of such secondary metabolites because they consist of a measurable phenotype with clear ecological consequences, synthesized by known genes in a small number of species. However, the processes involved in and timing of the divergence of prokaryotic secondary metabolites have been difficult to determine due to their antiquity and the possible frequency of horizontal gene transfer and homologous recombination. Through analyses of gene synteny, phylogenies of individual genes, and analyses of recombination and selection, we identified the evolutionary processes of this cluster in five species of cyanobacteria. Here, we provide evidence that the *sxt* cluster appears to have been largely vertically inherited and was therefore likely present early in the divergence of the Nostocales, at least 2,100 Ma, the earliest reliably dated appearance of a secondary metabolite. The *sxt* cluster has been extraordinarily conserved through stabilizing selection. Genes have been lost and rearranged, have undergone intra- and interspecific recombination, and have been subject to duplication followed by positive selection along the duplicated lineage, with likely consequences for the toxin analogues produced. Several hypotheses exist as to the ecophysiological role of saxitoxin: as a method of chemical defense, cellular nitrogen storage, DNA metabolism, or chemical signaling. The antiquity of this gene cluster indicates that potassium channels, not sodium channels, may have been the original targets of this compound. The extraordinary conservation of the machinery for saxitoxin synthesis, under radically changing environmental conditions, shows that it has continued to play an important adaptive role in some cyanobacteria.

Key words: cyanobacteria, polyketide synthase, paralytic shellfish toxins, saxitoxin biosynthesis, secondary metabolite.

Research article

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

Introduction

The evolutionary processes that shaped a specific phenotype are generally difficult to determine, as the link between genotype and phenotype is usually complex and partly known, consisting of multiple interacting factors (Mackay et al. 2009). Secondary metabolites differ, as they produce a measurable phenotype with clear ecological impacts, produced by discrete gene clusters in a small number of species (Fischbach et al. 2008). For this reason, they are important models for the study of evolutionary processes. However, the processes involved in the evolution of cyanobacterial secondary metabolites may be difficult to determine due to horizontal gene transfer and homologous recombination that can obscure the phylogenetic signal (Zhaxybayeva et al. 2006; Christiansen et al. 2008; Shi and Falkowski 2008).

One such secondary metabolite produced by cyanobacteria is the toxin saxitoxin. Saxitoxin and its derivatives (STXs) are carbamate alkaloid compounds that are produced via chemical reactions that are rare in metabolic pathways. These include a Claisen condensation onto an amino acid and a unique polyketide synthase (PKS)–like enzyme. Curiously, STXs are apparently synthesized via the same biosynthetic route in a group of microbial eukaryotes, the dinoflagellates (Shimizu 1993). STXs selectively block voltage-gated Na^+ and Ca^{2+} channels and are K^+ channel gating modifiers in excitable cells, affecting neural impulse generation (Llewellyn 2006).

An estimated 2,000 cases of human paralytic shellfish poisoning, mainly due to STX accumulation in shellfish, occur globally each year, with a mortality rate of 15% (Hallegraeff 1995). Other ecological impacts include changes to marine trophic structures, the deaths of marine mammals, fish, and seabirds, and, in freshwater environments, deaths of livestock (reviewed by Shumway 1990).

The gene cluster that is putatively responsible for the synthesis of STXs, *sxt*, has now been identified from five species of cyanobacteria, and a putative STX biosynthetic pathway has been proposed (fig. 1; Kellmann et al. 2008; Mihali et al. 2009; Moustafa et al. 2009; Stucken et al. 2010). Biosynthesis is initiated with *sxtA*, which contains four catalytic domains: a methyltransferase domain (*sxtA1*, MTP), a GNAT domain (*sxtA2*, ACTF) (loading of acyl carrier protein), an acyl carrier protein domain (*sxtA3*, ACP), and an AONS domain (*sxtA4*, AONS) (condensation domain) (fig. 1).

Few cyanobacterial STX-producing strains are known: strains of *Anabaena circinalis*, *Aphanizomenon* sp. NH-5, *Ap. issatschenkoi*, *Raphidiopsis brookii* and *Cylindrospermopsis*

© The Author 2010. 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

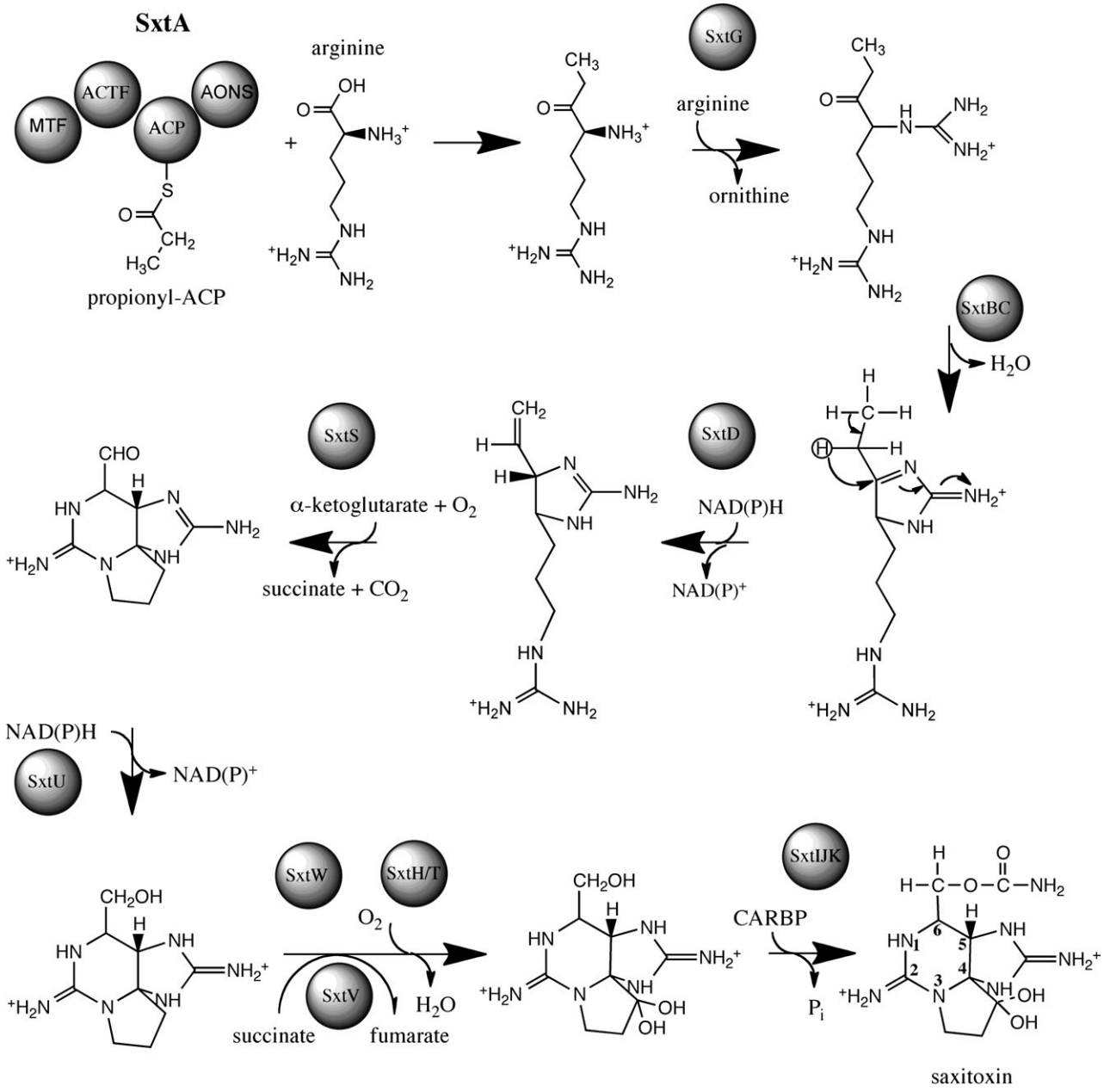

Fig. 1 The putative saxitoxin biosynthesis pathway. For more information regarding enzymes, cofactors and substrates, see Kellmann et al. (2008) and Mihali et al. (2009).

raciborskii of the order Nostocales, and *Lyngbya wollei* and *Planktothrix* sp. FP1 of the order Oscillatoriales (Mahmood and Carmichael 1986; Humpage et al. 1994; Onodera et al. 1997; Lagos et al. 1999; Pomati et al. 2000; Llewellyn et al. 2001; Dias et al. 2002; Stucken et al. 2009). These genera also contain species that are nontoxic, and different isolates of the same species, in particular, of *An. circinalis* and *C. raciborskii* (Beltran and Neilan 2000), may produce different toxin profiles or may be nontoxic. The strain-specific toxin profile may be due to the presence or absence of genes that code for enzymes responsible for tailoring of the carbon skeleton of STX. The five cyanobacterial species for which the *sxt* cluster has been characterized have varied toxin profiles (Mahmood and Carmichael 1986; Onodera et al. 1997; Lagos et al. 1999; Llewellyn et al. 2001; Stucken et al. 2009). Diversity in the STX analogues produced by the five strains may indicate that the cluster has undergone significant evolutionary changes. In this work, we determined the evolution processes by which this complex gene cluster was formed and is maintained.

Materials and Methods

Synteny Comparison and Phylogenetic Analyses

Genes putatively involved in the STX pathway in *C. raciborskii* T3, *R. brookii* D9, *An. circinalis* AWQC131C, *Aphanizomenon* sp. NH-5, and *L. wollei* have been identified and sequenced (Kellmann et al. 2008; Mihali et al. 2009; Stucken et al. 2010; Mihali et al., unpublished data). For the

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

species phylogeny, 16S rRNA genes were used as they are representative of the phylogeny based on 682 proteins (Zhaxybayeva et al. 2006; Shi and Falkowski 2008; Swingley et al. 2008). We found homologous sequences using searches in appropriate databases, inferred alignments, and determined the optimal substitution models (Abascal et al. 2005). Alignments were analyzed using maximum likelihood (ML), based on optimal model and parameters (Modeltest, Posada and Crandall 1998; Prottest, Abascal et al. 2005), and 100 bootstrap replicates (Phyml v2.4.4, Guindon and Gascuel 2003). Bayesian analysis was performed using MrBayes (Huelsenbeck and Ronquist 2001), with the appropriate substitution model until chains converged, approximately 100,000 generations, and trees were sampled every 100 generations, discarding a burn-in period of the first 250 sampling points.

Test of Recombination

If there has been recombination among divergent gene sequences, different regions of a gene will have different evolutionary histories and an alignment cannot be well represented by a single phylogenetic tree (Posada and Crandall 2002). Recombination may also impact on later analyses based on an inferred phylogeny, such as tests of selection (Anisimova and Kosiol 2009). We tested for recombination using the program GARD (Kosakovsky Pond et al. 2006), which detects evidence of segment-specific phylogenies and determines the breakpoints. GARD compares a model in which the sequence data are fitted to a single phylogeny to models in which the sequence data are partitioned into two or more regions having varying phylogenies. Support for recombination is reflected by changes in the goodness of fit between nonrecombinant and recombinant models, as assessed by the Akaike Information Criterion (Akaike 1981, Posada and Buckley 2004). Potential breakpoints were then tested with Kishino–Hasegawa tests (Kishino and Hasegawa 1989) to determine whether adjacent sequence fragments yield statistically different tree topologies, as local rate variation or heterotachy may cause regions in an alignment to appear to evolved with differing branch lengths. In simulation studies, 30 sites of recombination in a ten taxon alignment was found to increase the false-positive rate in ML-based positive selection detection methods to ~20%, whereas three breakpoints had no impact (Anisimova et al. 2003). Analyses were implemented on the GARD cluster (http://www.datamonkey.org/GARD/, Kosakovsky Pond and Frost 2005) using the multiple breakpoint detection method. The nucleotide substitution matrix was optimized for each data set via Modeltest (Posada and Crandall 1998), as implemented via the HyPhy package (Kosakovsky Pond et al. 2005), and among-site rate variation was modeled by a discretized gamma distribution with four rate classes.

Tests of Selection

Tests for positive and purifying selection were conducted using several different approaches. For each alignment, we used a distance-based approach, the codon-based *z* test (Nei and Gojobori 1986), which compares the relative abundance of synonymous and nonsynonymous substitutions within the gene sequences, as implemented in the program MEGA v3.1 (Kumar et al. 2004). For each pair of sequences, this is done by estimating the number of synonymous substitutions per synonymous site (dS), the number of nonsynonymous substitutions per nonsynonymous site (dN), and their variances, respectively, to test the null hypothesis that dN equals dS.

Distance-based methods that search for the signal of positive selection throughout the whole alignment may lack statistical power if only few sites have been affected. Therefore, an ML-based method which detects selection on a site-by-site basis was conducted using the program PARRIS (Scheffler et al. 2006). This method involves fitting a distribution of substitution rates across sites and then inferring the rate at which individual sites evolve. This method allows synonymous substitution rates to vary across sites. It also allows parts of the alignment that have experienced recombination to be analyzed based on separate trees, with nonrecombinant fragments assigned their own tree topologies and branch lengths.

Different branches within an alignment may evolve at different rates, particularly when gene duplication events or sudden changes in the environment affecting one particular lineage may have occurred. If this has occurred, the signal of selection averaged over the entire tree may have become undetectable. Therefore, we tested for positive selection along specific branches of interest within each phylogeny using the branch-site method (Zhang et al. 2005). The results of the branch-site method based on a model allowing for positive selection in the foreground branch (FG) was compared by likelihood ratio test (LRT) with a second model, which did not allow for positive selection in the FG branch (Yang 1998). The significance of the LRTs was calculated assuming that twice the difference in the log of ML values between the two models is distributed as a χ^2 distribution, whereas the degrees of freedom were given by the difference in the number of parameters between the two nested models. In general, our alignments contained too few sequences to accurately detect positive selection operating at specific codons within the alignments. The branch-site analyses, LRTs, and PARRIS were all implemented in the HyPhy software package (Kosakovsky Pond et al. 2005).

Results and Discussion

Gene Synteny

The organization and compliment of *sxt* genes in each of the producing strains corresponds well with the species phylogeny, in that species that were most closely related had the most similar *sxt* gene content and organization, and those that were most distantly related had the most divergent *sxt* gene clusters (fig. 2). Recent cyanobacterial genome sequencing projects have shown that a core set of genes has been extraordinarily conserved and subject to strong selection pressure to maintain synteny (Shi

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

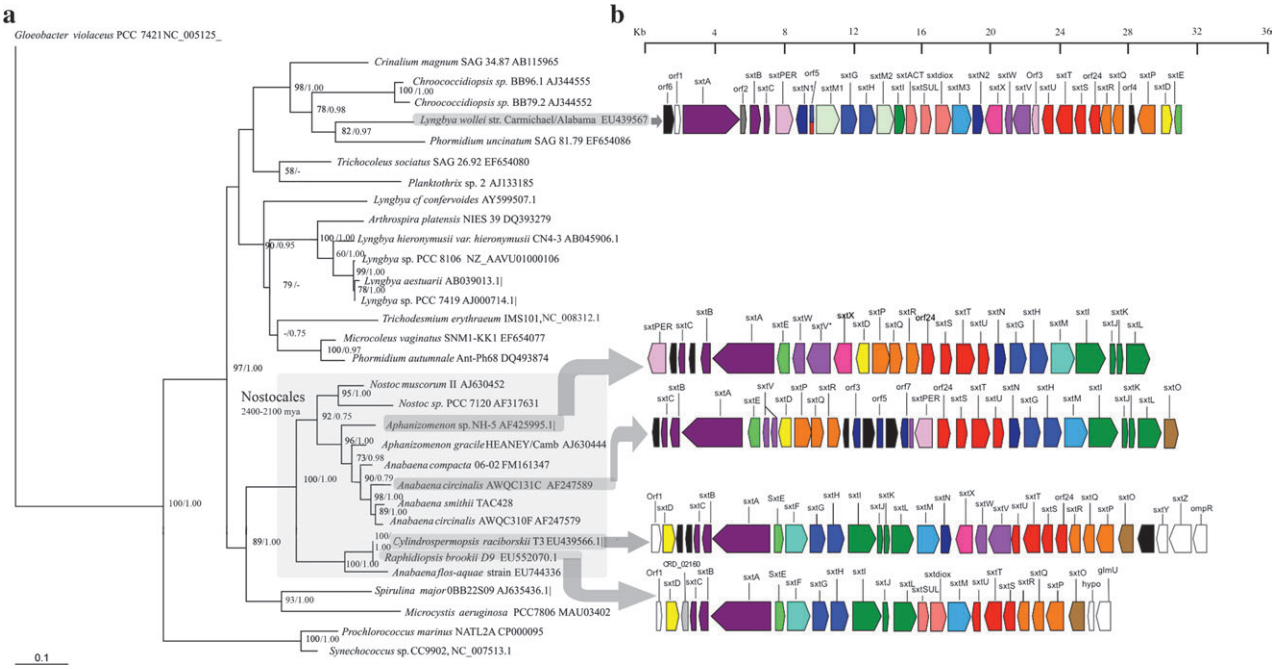

FIG. 2 Phylogeny of *sxt* containing cyanobacterial clades and the *sxt* clusters in the sequenced strains. (*a*) Phylogeny of the cyanobacterial clades relevant to this study, Nostocales, Oscillatoriales, and outgroups, based on 16s rRNA, using an ML method and a GTR + γ-substitution model, values at nodes represent BS support/Bayesian PP support, only values above 50% are shown. Likelihood = −9,053.5. (*b*) Gene synteny of the clusters in five saxitoxin producing strains: *Anabaena circinalis* 131C, *Aphanizomenon* sp. NH-5 (Mihali et al. 2009), *Cylindrospermopsis raciborskii* T3 (Kellmann et al. 2008), *Raphidiopsis brookii* D9 (Stucken et al. 2010), and *Lyngbya wollei*. Transposases are shown in black.

and Falkowski 2008; Swingley et al. 2008). We found that a complement of 14 genes or gene families were present in the *sxt* clusters of all five STX-producing strains, whereas an additional three genes were present in four of the five strains and six genes were present in three species (fig. 2*b* and table 1). Although gene order differed among clusters, groups of genes tended to be located adjacent to each other in each producing organism (fig. 2*b*). The clusters of *An. circinalis* AWQC131C and *Aphanizomenon* sp. NH-5 are similarly organized, differing mainly in the position of the group of genes *sxtP*, *Q*, and *R*, and *sxtPER*, which, in *An. circinalis*, are bordered by several transposases (fig. 2). The *C. raciborskii* T3 and *R. brookii* D9 *sxt* clusters are similarly organized, though appear to have lost a different complement of genes from their common STX-producing ancestor. *Raphidiopsis brookii* D9 lacks the largest number of *sxt* genes, in line with the reduction and streamlining of its entire genome (Stucken et al. 2010). The STX gene organization of *C. raciborskii* T3 and *R. brookii* D9 are also similar to the *An. circinalis*/*Aphanizomenon* sp. NH-5 in that the cluster of *sxtI*, *J*, *K* (missing in D9), and *sxtL* was maintained as a unit and *sxtA* was transcribed in the same direction. *Lyngbya wollei* had the most divergent cluster of those analyzed and contains the unique gene *sxtACT*, an extra *sxtN* homolog, as well as a unique direction of transcription for the group of genes *sxtA*, *B*, and *C* in relation to the entire gene cluster. Similarities between *L. wollei* and *C. raciborskii* are the common position and direction of the group of genes *sxtM*, *N*, *X*, *W*, *V*, *U*, *T*, *S*, *R*, *Q*, and *P*.

Recombination

Significant recombination was detected in only five of the *sxt* genes or gene families, *sxtA* (position 995), *sxtC* (position 141), *sxtH/T* (positions 132 and 465), *sxtF/M* (position 840), and *sxtL* (positions 411, 896, and 709) (supplementary table S1, Supplementary Material online). Recombination within the *sxtH/T* gene family has been intragenomic or intraspecific within *L. wollei*, *An. circinalis*, and *Aphanizomenon* sp. NH-5 (positions 133–465; fig. 3*c*), and intergeneric, from *sxtT* into *sxtH*, either from *An. circinalis* or an ancestor into a common ancestor of *An. circinalis* and *Aphanizomenon* sp. NH-5 (positions 1–133, fig. 3*b*). Within the *sxtF/M* gene family, recombination occurred in *An. circinalis*, either intragenomically, prior to the loss of one copy of this gene family, or intergenerically, between *An. circinalis* and *Aphanizomenon* sp. NH-5 or an ancestral strain (fig. 3*a* and *b*). Based on the phylogeny derived from the largest fraction of the gene, positions 1–840, an *sxtM* ortholog was retained in *An. circinalis* and one *sxtF* ortholog in *Aphanizomenon* sp. NH-5 (fig. 4*a*), whereas the fractions 841–1340 appear to constitute unique *sxtF* orthologs in *An. circinalis* and *Aphanizomenon* sp. NH-5 (fig. 4*b*). Interestingly, this could indicate that two paralogs of this gene family were originally present in *An. circinalis* and *Aphanizomenon* sp. NH-5, as in *C. raciborskii*, *R. brookii*, and *L. wollei*. The evolutionary divergence between strains may be the reason for the low level of recombination. In the NRPS/PKS gene cluster *mcy*, responsible for the hepatotoxic microcystin biosynthesis, recombination was not detected between

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

Table 1. Genes and Gene Families in the *sxt* Cluster and Their Putative Role, the Number of Species in which They Are Present, BS Support Value for the Monophyly of the Clade of that *sxt* Gene, and the Closest Sister Group to the Clade.

Gene or gene family	Putative role	No. of species/ copies[a]	BS support for monophyly of *sxt* clade	Topology within *sxt* cluster[b]	Sister group to *sxt* clade[c]
sxtA	Polyketide synthase	5	100	(Lwollei, ((T3, Rap), (NH5, 131C)))	ABM63529.1 *Candidatus endobugula sertula* BryX
sxtB	Cytidine deaminase	5	100	((Lwollei, (T3, Rap)), (NH5, 131C))	ZP 0215975 *Shewanella benthica* cytidine deaminase
sxtC	Unknown	5	n/a[d]	n/a	NP_309227.1 *Escherichia coli* DNA-binding protein
sxtD	Sterole desaturase	5	99	((Lwollei, (T3, Rap)), (NH5, 131C))	YP_722737.1 *Trichodesmium erythraeum* IMS101 sterole desaturase
sxtE	Unknown	5	99	(Lwollei, ((T3, Rap), (NH5, 131C)));	YP_001445162 *Vibrio harveyi* hypothetical protein
sxtF/M	MATE	5 + 3	100	(((((Lw_M3,(T3_M,RapM)),131C), (Lw_M2,Lw_M1)),NH_5), T3_F,RapF)	GQ979609 *Nostoc* sp trans-AT PKS gene cluster
sxtG	Amidinotransferase	5	100	((Lwollei, (T3, Rap)), (NH5, 131C))	AP009552 *Microcystis aeruginosa* NIES-843 glycine amidinotransferase
sxtH/T/diox	Phenylprop dioxygenase	5 + 8	100	(((((((131C_T,NH5_T),(T3T,Rap_T), L_d3),(Rap_diox,L_d2)), (13C_H,NH5_H)),L_H), T3_H,Rap_H)	YP_001940983 *Methylokorus infernorum* V4 Ring-hydroxylating dioxygenase
sxtI	Carbamoyltransferase	5	63	((Lwollei, (T3, Rap)), (NH5, 131C))	YP 721441 *T. erythraeum* IMS101 carbamoyltransferase
sxtJ	Unknown	4	100	(T3, Rap), (NH5, 131C)	ZP_01619107.1 *Lyngbya* sp. PCC 8106 hypothetical protein
sxtK	Unknown	3	100	(NH5, (131C, T3))	YP_721427 *T. erythraeum* hypothetical protein
sxtL	GDSL lipase	4	100	(T3, Rap), (NH5, 131C)	ZP_01619105.1 *Lyngbya* sp. PCC 8106 hypothetical protein
sxtN	Sulfotransferase	4 +1	100	(Lwollei_1, (T3, Lwollei_2), (NH5, 131C)	YP_723575.1 *T. erythraeum* aminotransferase
sxtO	ASK	3	0	n/a	n/a
sxtP	Integrins	5	100	((Lwollei, (T3, Rap)), (NH5, 131C))	ABI75126.1 *Anabaena variabilis* ATCC 29413 hypothetical protein
sxtQ	Unknown	5	92	(Lwollei, ((T3, Rap), (NH5, 131C)))	ETA_27510 *Erwinia tasmaniensis* hypothetical protein
sxtR	Transferase	5	98	((Lwollei, (T3, Rap)), (NH5, 131C))	ZP_03155608.1 *Cyanothece* sp. PCC 7822 conserved hypothetical protein
sxtPER	Permease	3	100	(NH5, (131C, Lwollei))	YP_001866958.1 *Nostoc punctiforme* PCC 73102 hypothetical protein
sxtS	Phytanoyl diox	5	100	((Lwollei, (T3, Rap)), (NH5, 131C))	YP_508892 *Jannaschia* sp phytanoyl-CoA dioxygenase
sxtU	Alcohol dehydrogenase	5	100	((T3, Rap), (Lwollei, (NH5, 131C)))	YP_001867040.1 *Nostoc punctiforme* PCC 73102 dehydrogenase
sxtV	Succinate dehydrogenase	2	100	(Lwollei, T3)	YP_001868454.1 *Nostoc punctiforme* PCC 73102 fumarate reductase
sxtW	Ferredoxin	3	78	(NH5, (T3, Lwollei))	YP_001868453.1 *Nostoc punctiforme* PCC 73102 ferredoxin
sxtX	Cephalosporin hydroxylase	3	67	(NH5, (T3, Lwollei))	ZP_01623574.1 *Lyngbya* sp PCC 8106 hypothetical protein
Orf1	Unknown	3	99	((T3, Rap), 131C)	ZP_01624160.1 *Lyngbya* sp PCC 8106 hypothetical protein

[a] Values after + indicate multiple copies of a gene family per species.
[b] In sequences with significant recombination, this represents the topology of the largest single fragment.
[c] GenBank accession number, species name, putative protein role. See Materials and Methods for further details.
[d] *sxtC* was not further analyzed due to its low homology, short length, and significant breakpoint.

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

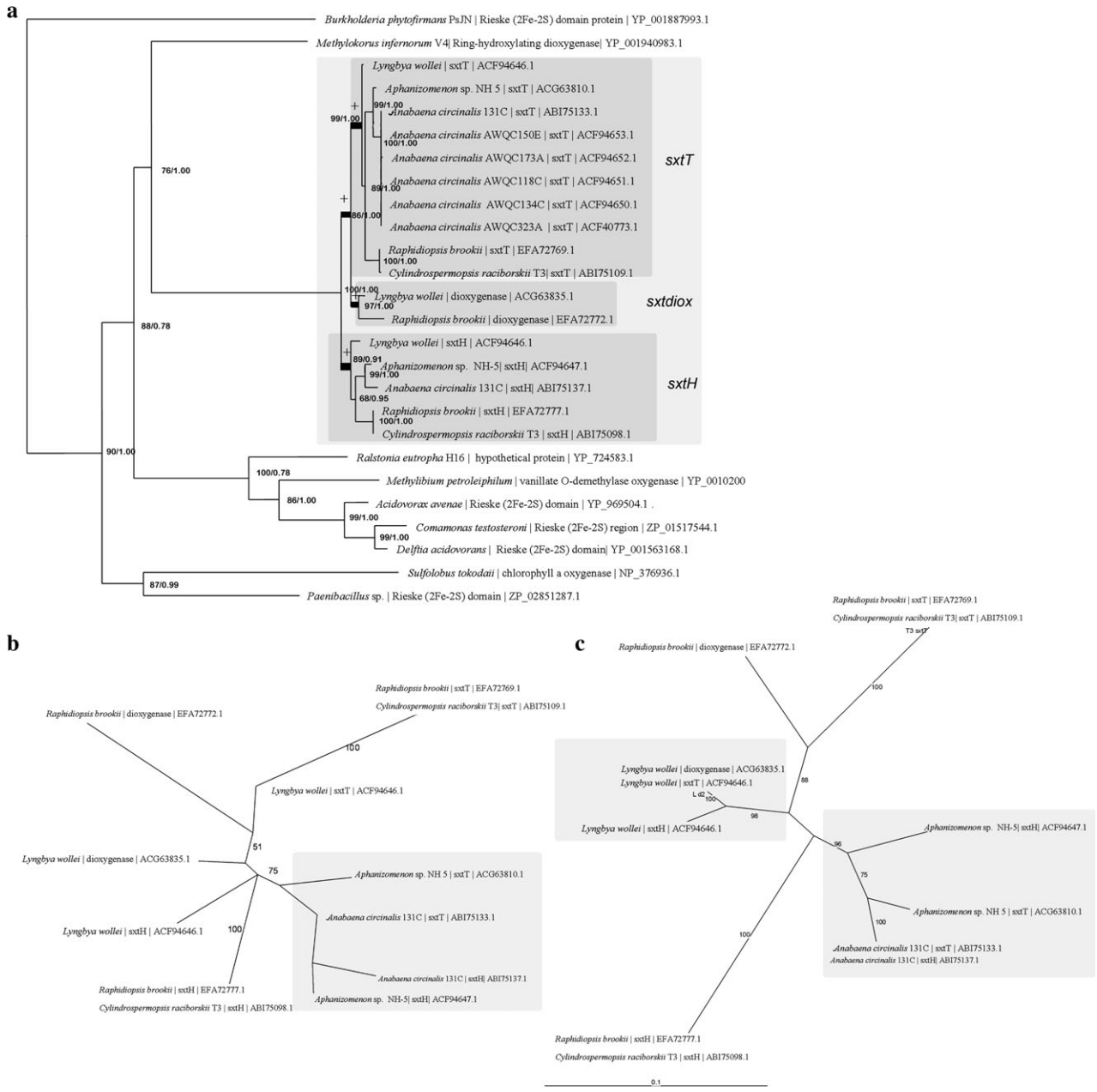

FIG. 3 Phylogeny of the *sxtH*/*T*/*diox* family and the closest paralogs. (*a*) Phylogeny showing the lineages with significant positive selection (thick black lines and +), based on proteins using an ML method and a WAG + I + G + F substitution model, values at nodes represent BS support/Bayesian PP support, only values above 50% are shown, representing the largest nonrecombinant fragment (nucleotide positions 465–1020). Likelihood = −13,101.12. (*b*) Unrooted phylogeny of the fragment positions 1–133, based on nucleotides and GTR + γ-substitution model. Box shows significant incongruence to main phylogeny indicating recombination in these organisms. (*c*) Unrooted phylogeny of the fragment positions 133–465, based on nucleotides and GTR + γ-substitution model. Boxes show significant incongruence to main phylogeny indicating recombination in these organisms.

genera, whereas it appeared to occur often between closely related strains and intragenomically (Rantala et al. 2004; Christiansen et al. 2008; Tooming-Klunderud et al. 2008).

Phylogenetic Analyses

Based on the individual phylogenetic analyses, the *sxt* gene cluster appears to have had a single origin in cyanobacteria and is only distantly related to other extant gene clusters, in line with conclusions from other studies (Moustafa et al. 2009). Genes from the five strains formed highly supported (63–100% BS, mean = 94%) monophyletic clades in alignments of all but 1 of the 23 genes or gene families analyzed (table 1 and supplementary table S1, Supplementary Material online). Branch lengths leading to the *sxt* clade were much longer than within-clade branch lengths, indicating the absence of close orthologs and the high level of sequence conservation. The most frequent topologies (12 of 14 genes common to all clusters) mirrored the relationships among gene clusters indicated by synteny and species phylogeny (fig. 2 and table 1). In analyses of genes common

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

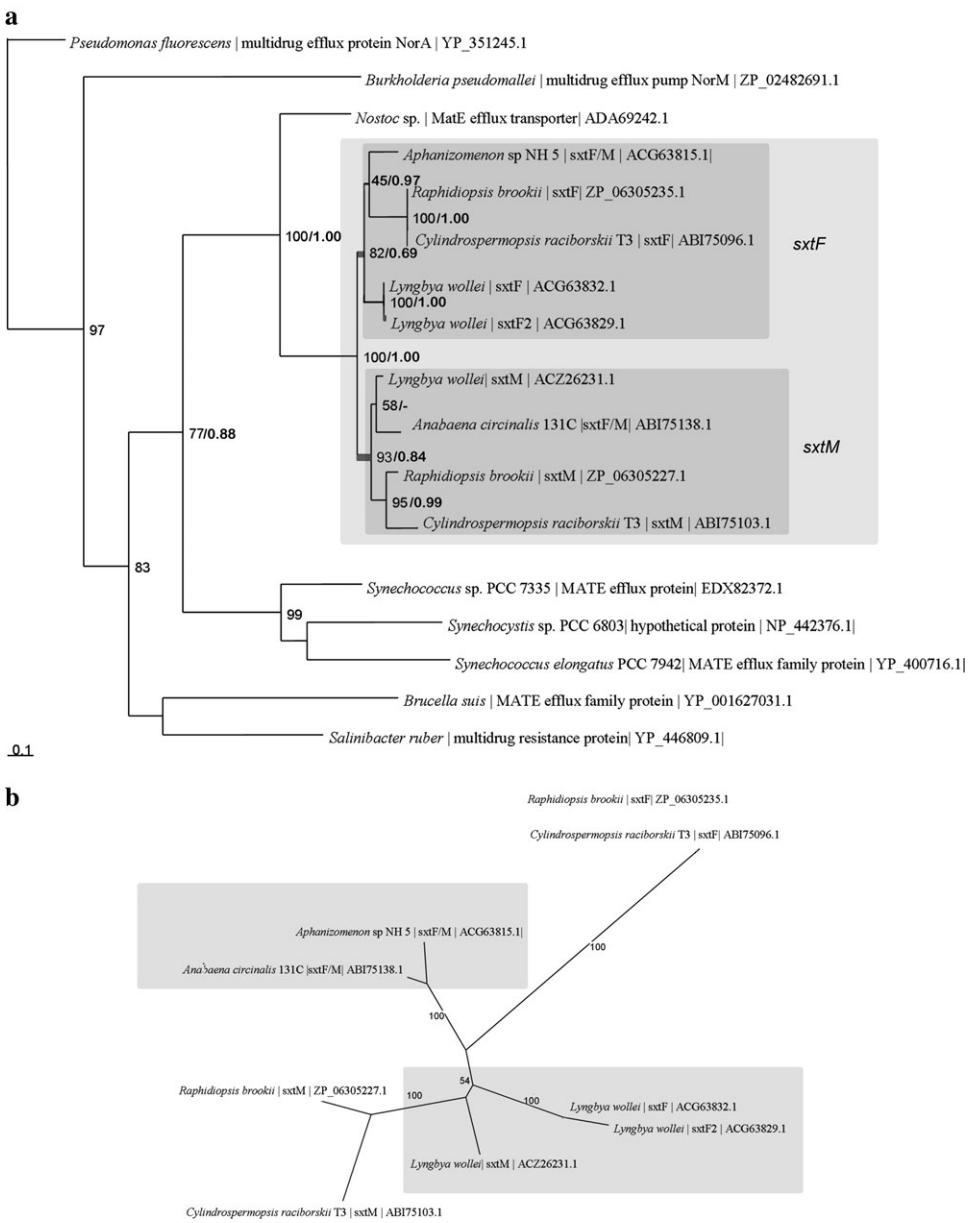

Fig. 4 Phylogeny of the *sxtF/M* family and the closest paralogs. (*a*) Phylogeny showing the lineages with significantly different selective regimes to the rest of the tree (gray thick lines), based on proteins using an ML method and a CpREV + I + G + F substitution model, values at nodes represent BS support/Bayesian PP support, only values above 50% are shown, representing nucleotide positions 1–840 of the alignment. Likelihood = −11,082.45. (*b*) Unrooted phylogeny of the fragment positions 840–1340, using an ML method, based on nucleotides and GTR + γ-substitution model. Box shows significant incongruence to main phylogeny indicating recombination in these organisms.

to all strains ($n = 14$), orthologs from *L. wollei* (*sxtA*, *E*, *F*, *H*, *Q*, and *T*) formed the most basal lineage or *L. wollei* was the basal lineage in a clade with C. *raciborskii* T3 and *R. brookii*, with *An. circinalis* and *Aphanizomenon* sp. NH-5 forming a separate clade (*sxtB*, *D*, *I*, *G*, *P*, and *S*; supplementary figs. S1–S20, Supplementary Material online). Infrequently (*sxtU* and *F*), *L. wollei* was the most basal lineage in a clade with *An. circinalis* AWQC131C and *Aphanizomenon* sp. NH-5, with C. *raciborskii* and *R. brookii* forming a separate clade, although this topology was not well supported (BS = 52, 58%, respectively). Orthologs from *R. brookii* and *C. raciborskii* always formed a highly supported clade and were 99–100% identical for all but three of the genes present in both strains. *Anabaena circinalis* AWQC131C and *Aphanizomenon* sp. NH-5 formed a highly supported clade in analyses of 17 of the 19 genes present in both strains (fig. 3 and supplementary figs. S1–S20, Supplementary Material online).

In the progenitor STX-producing strain, the gene cluster appears to have had a mosaic origin, as while many (71%, 17 of 24) genes in the cluster have distant paralogs in other cyanobacteria: *Trichodesmium erythraeum*, *Synechococcus elongatus*, *Microcystis aeruginosa*, *Nostoc punctiforme*, *An. variabilis*, and *Lyngbya* sp. PCC 8106 (table 1 and supplementary figs. S1–S20, Supplementary Material online), some (29%, 7 of 24) have no extant cyanobacterial homologs and may have been laterally acquired from other

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

bacteria, mainly gammaproteobacteria: *Escherichia coli*, *Shewanella benthica*, *Candidatus endobugula*, *Methylokorus infernorum*, and *Erwinia tasmaniensis*.

Selection Analyses

Alignment-wide stabilizing selection was overwhelmingly found in analyses using the distance-based and ML-based methods, with only two genes (*sxtV* and *W*) showing alignment-wide neutral selection (supplementary table S1, Supplementary Material online). Alignment-wide positive selection was not found in any alignment, using either the *z* test or the ML-based approach (supplementary table S1, Supplementary Material online), although in eight of the genes analyzed, at least one lineage was not significantly affected by stabilizing selection. High levels of sequence conservation (99–100% similar) were found in analyses of *sxtG*, *H*, *T*, and *I* in eight strains of *An. circinalis* that differed from one another in 16S rRNA gene sequences (Beltran and Neilan 2000; Moustafa et al. 2009; fig. 3 and supplementary figs. S1–S20, Supplementary Material online). Similarly, *An. circinalis* ACBU02 appeared to differ in only 4 of the 26 genes (*sxtF*, *J*, *K*, and *N*) as compared with *An. circinalis* AWQC131C, despite differences in 16S rRNA gene sequence (Moustafa et al. 2009).

Two of the alignments showed evidence of significantly different selective regimes on branches following gene duplication events, representing the first evidence that positive selection has acted in the evolution of genes in the *sxt* cluster (figs. 3 and 4 and supplementary table S1, Supplementary Material online). Although gene duplications are a major source of evolutionary innovation (Jordan et al. 2004; He and Zhang 2005; Brunet et al. 2006; Byrne and Wolfe 2006; Johnston et al. 2007; Chapman et al. 2008), higher dN/dS ratios on branches following duplications have relatively rarely been found as a proportion of duplicated genes (Raes and Van de Peer 2003). Higher dN/dS ratios may also be due to a relaxation of purifying selection rather than positive selection. In the analysis of *sxtH/T* family, the largest nonrecombinant fragment (positions 465–1020) was found to be under significantly different selective regimes in different parts of the tree (fig. 3*a*). The FG branches were affected by positive selection (mean dN/dS = 7.5, $P < 0.0001$; fig. 3 and supplementary table S1, Supplementary Material online). The largest region of this gene appears to have duplicated in *L. wollei*, or an ancestral strain, to give rise to the clade of *sxtT* in *L. wollei*, *C. raciborskii* T3, *An. circinalis* AWQC131C, and *Aphanizomenon* sp. NH-5 and again to give rise to the *sxtdiox* clade in *R. brookii* and *L. wollei*. *sxtH/T* are putatively involved in the formation of the hydrated ketone function at C-12, a so-called geminal hydroxyl (Kellmann et al. 2008). The double hydroxylation of C-12 has been shown to be an important residue in sodium channel binding and therefore influences toxicity. The *sxtdiox* gene is found in only *L. wollei* and *R. brookii* and may be related to the C-12 carbinol STX derivatives produced by *L. wollei*.

In the analysis of *sxtF* and *sxtM* orthologs, the null model that the two FGs (fig. 4) were under the same selective regime as the rest of the tree could be rejected (fig. 4 and supplementary table S1, Supplementary Material online, $P < 0.03$) for the largest nonrecombinant fraction (positions 0–840). The mean dN/dS for these three branches was 4.6 compared with the significant stabilizing selection on the alignment (supplementary table S1, Supplementary Material online). However, the null hypothesis that these branches were under positive selection could not be rejected ($P = 0.24$), indicating possible neutral selection in one or both lineages or a significantly different selective regime elsewhere in the tree. The gene family *sxtF/M* is putatively involved in the export of STXs and has high sequence similarity to sodium-driven multidrug and toxic compound extrusion (MATE) proteins of the *NorM* family (Brown et al. 1999). The duplication and diversification of the original MATE gene may be related to the synthesis of novel and diverse STX derivatives, which required specific transporters or by pressure to increase and/or specifically regulate extracellular transport of STXs.

Cyanobacteria are an ancient group of organisms, with a presence in the fossil record of greater than 2,500 Ma (Tomitani et al. 2006, Knoll 2008). The Nostocales, the clade of cyanobacteria capable of cell differentiation and nitrogen fixation, is monophyletic and diverged once between 2,450 and 2,100 Ma (Tomitani et al. 2006, Swingley et al. 2008) (fig. 2), based on analyses of multiple genes, the fossil record, and geochemical evidence for the first appearance of nitrogen fixation. Our study indicates that it is likely that a common ancestor of the *C. raciborskii*/*R. brookii* and the *Anabaena*/*Aphanizomenon* clade, an early diverging Nostocales, possessed the *sxt* cluster (fig. 2). It may have been present in an earlier common ancestor of the Nostocales and *L. wollei*, a member of the Oscillatoriales, as the phylogenetic evidence was equivocal on whether the genes from *L. wollei* were the sister group to those of all other species or sister to those of *C. raciborskii* and *R. brookii* (table 1). Evidence of a putative excision site in the strain of *An. circinalis* 310F that does not produce STXs (Mihali et al. 2009) indicates that widespread secondary loss may account for the patchy distribution of the *sxt* cluster in strains of *Anabaena*. The *sxt* gene clusters contain transposases, which are associated with gene deletion and duplication events (fig. 2). This model of evolution parallels what has been proposed for the *mcy* cluster (Rantala et al. 2004), in which transposable elements were responsible for widespread *mcy* gene deletion, with remnants of the *mcy* gene cluster remaining in nontoxic strains (Christiansen et al. 2008). Phylogenetic analyses of STX-producing and nonproducing strains of *An. circinalis*, *Aphanizomenon* sp., and *C. raciborskii* have found that species were monophyletic and producing and nonproducing strains were closely related (Beltran and Neilan 2000, Moustafa et al. 2009), suggesting that any putative gene loss events have occurred relatively recently in their evolution.

Concluding Remarks

Several hypotheses exist as to the ecophysiological role of saxitoxin: as a method of chemical defense, cellular

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

nitrogen storage, DNA metabolism, or chemical signaling (Pomati et al. 2004; Cembella and John 2006; Zimmer and Ferrer 2007). Our analysis based on gene synteny, complement, phylogenetic analyses of primary sequences, the strong stabilizing selection, and the relatively low recombination rate indicates that this cluster has been largely vertically inherited and therefore may have emerged at least 2,100 Ma. At that time, the environment and biosphere were substantially different to those of today, containing organisms that had not yet evolved voltage-gated sodium channels, the most widely studied STX target. Extant bacteria generally do not contain sodium channels. Potassium channels are evolutionary predecessors of voltage-gated sodium channels (Goldin 2002) and were the likely targets of STXs before the emergence of sodium channels. The ecophysiological role of potassium channels in bacteria is largely unknown, and the effect of STXs on these channels has not been studied. Research on the effects of this ancient toxin on potassium channels will be key to understanding its past and present function. The strong sequence conservation in the *sxt* cluster, despite radical changes in organism diversity, levels of oxygen, ultraviolet irradiation, pH, and salinity, shows that it has maintained a vital adaptive role in cyanobacteria since its initial divergence.

Supplementary Material

Supplementary table S1 and figures S1–20 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

S.M. and B.N. are fellows of the Australian Research Council. We thank Art Poon and Sergei Kosakovsky Pond for advice with and modifications to HyPhy. This work was supported by the Australian Research Council.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 21:2104–2105.

Akaike H. 1981. Likelihood of a model and information criteria. *J Econom*. 16:3–14.

Anisimova M, Kosiol C. 2009. Investigating protein-coding sequence evolution with probabilistic codon substitution models. *Mol Biol Evol*. 26:255–271.

Anisimova M, Nielsen R, Yang Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164:1229–1236.

Beltran EC, Neilan BA. 2000. Geographical segregation of the neurotoxin-producing cyanobacterium *Anabaena circinalis*. *Appl Environ Microbiol*. 66:4468–4474.

Brown MH, Paulsen IT, Skurray RA. 1999. The multidrug efflux protein NorM is a prototype of a new family of transporters. *Mol Microbiol*. 31:394–395.

Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol Biol Evol*. 23:1808–1816.

Byrne KP, Wolfe KH. 2006. Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. *Genetics*. 175:1341–1350.

Cembella AD, John U. 2006. Molecular physiology of toxin production and growth regulation in harmful algae. In: Granéli E, Turner JT, editors. Ecology of harmful algae. Heidelberg (Germany): Springer Verlag. p. 339.

Chapman MA, Leebens-Mack JH, Burke JM. 2008. Positive selection and expression divergence following gene duplication in the sunflower Cycloidea gene family. *Mol Biol Evol*. 25:1260–1273.

Christiansen G, Molitor C, Philmus B, Kurmayer R. 2008. Nontoxic strains of cyanobacteria are the result of major gene deletion events induced by a transposable element. *Mol Biol Evol*. 25:1695–1704.

Dias E, Pereira P, Franca S. 2002. Production of paralytic shellfish toxins by *Aphanizomenon* sp. LMECYA31 (cyanobacteria). *J Phycol*. 38:1–9.

Fischbach MA, Walsh CT, Clardy J. 2008. Evolution of gene collectives: how natural selection drives chemical innovation. *Proc Natl Acad Sci U S A*. 105:4601–4608.

Guindon S, Gascuel O. 2003. PHYML—a simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.

Goldin AL. 2002. Evolution of voltage-gated Na+ channels. *J Exp Biol*. 205:575–584.

Hallegraeff GM. 1995. Harmful algal blooms: a global overview. In: Hallegraeff GM, Anderson DM, Cembella AD, editors. Manual on harmful marine microalgae. Paris: UNESCO. p. 1–22.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169:1157–1164.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Humpage AR, Rositano J, Bretag AH, Brown R, Baker PD, Nicholson BC, Steffensen DA. 1994. Paralytic shellfish poisons from Australian cyanobacterial blooms. *Aust J Mar Fresh Res*. 45:761–771.

Johnston CR, O'Dushlaine C, Fitzpatrick DA, Edwards RJ, Shields DC. 2007. Evaluation of whether accelerated protein evolution in chordates has occurred before, after, or simultaneously with gene duplication. *Mol Biol Evol*. 24:315–323.

Jordan IK, Wolf YI, Koonin EV. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol*. 4:22.

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

Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol*. 29:170–179.

Knoll AH. 2008. Cyanobacteria and earth's history. In: Herrero A, Flores E, editors. The cyanobacteria, molecular biology, genomics and evolution. Norfolk (United Kingdom): Caister academic press. p. 1–19.

Kosakovsky Pond SL, Frost SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.

Kosakovsky Pond SL, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.

Kosakovsky Pond SL, Sergei L, Posada D, Gravenor MB. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol*. 23:1891–1901.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform*. 5:150–163.

Lagos N, Onodera H, Zagatto PA, Andrinolo D, Azevedo SM, Oshima Y. 1999. The first evidence of paralytic shellfish toxins in

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011

the freshwater cyanobacterium *Cylindrospermopsis raciborskii*, isolated from Brazil. *Toxicon* 37:1359–1373.
Llewellyn LE. 2006. Saxitoxin, a toxic marine natural product that targets a multitude of receptors. *Nat Prod Rep*. 23:200–222.
Llewellyn LE, Negri AP, Doyle J, Baker PD, Beltran EC, Neilan BA. 2001. Radioreceptor assays for sensitive detection and quantitation of saxitoxin and its analogues from strains of the freshwater cyanobacterium, *Anabaena circinalis*. *Environ Sci Technol*. 35:1445–1451.
Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet*. 10:565–577.
Mahmood NA, Carmichael WW. 1986. Paralytic shellfish poisons produced by the freshwater cyanobacterium *Aphanizomenon flos-aquae* NH-5. *Toxicon* 24:175–186.
Mihali TK, Kellmann R, Neilan BA. 2009. Characterisation of the paralytic shellfish toxin biosynthesis gene cluster in *Anabaena circinalis* 131C and *Aphanizomenon*. *BMC Biochem*. 10:8.
Moustafa A, Loram JE, Hackett JD, Anderson DM, Plumley FG, Bhattacharya D. 2009. Origin of saxitoxin biosynthetic genes in cyanobacteria. *PLoS One*. 4:e5758.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 3:418–426.
Onodera H, Satake M, Oshima Y, Yasumoto T, Carmichael WW. 1997. New saxitoxin analogues from the freshwater filamentous cyanobacterium *Lyngbya wollei*. *Nat Toxins*. 5:146–151.
Pomati F, Rossetti C, Manarolla G, Burns BP, Neilan BA. 2004. Interactions between intracellular Na^+ levels and saxitoxin production in *Cylindrospermopsis raciborskii* T3. *Microbiology* 150:455–461.
Pomati F, Sacchi S, Rossetti C, Giovannardi S, Onodera H, Oshima Y, Neilan BA. 2000. The freshwater cyanobacterium *Planktothrix* sp. FP1: molecular identification and detection of paralytic shellfish poisoning toxins. *J Phycol*. 36:553–562.
Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over Likelihood Ratio Tests. *Syst Biol*. 53:793–808.
Posada D, Crandall K. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
Posada D, Crandall KA. 2002. The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol*. 54:396–402.
Raes J, Van de Peer Y. 2003. Gene duplication, the evolution of novel gene functions, and detecting functional divergence of duplicates in silico. *Appl Bioinformatics*. 2:91–101.
Rantala A, Fewer DP, Hisbergues M, Rouhiainen L, Vaitomaa J, Börner T, Sivonen K. 2004. Phylogenetic evidence for the early evolution of microcystin synthesis. *Proc Natl Acad Sci U S A*. 101:568–573.
Scheffler K, Martin DP, Seoighe C. 2006. Robust inference of positive selection from recombining coding sequences. *Bioinformatics* 22:2493–2499.
Shi T, Falkowski PG. 2008. Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc Natl Acad Sci U S A*. 105:2510–2515.
Shimizu Y. 1993. Microalgal metabolites. *Chem Rev*. 93:1685–1698.
Shumway SE. 1990. A review of the effects of algal blooms on shellfish and aquaculture. *J World Aquacult Soc*. 21:65–104.
Stucken K, John U, Cembella A, Murillo AA, Soto-Liebe K. 2010. The smallest known genomes of multicellular and toxic cyanobacteria: comparison, minimal gene sets for linked traits and the evolutionary implications. *PLoS One*. 5:e9235.
Stucken K, Murillo AA, Soto-Liebe K, Fuentes-Valdés JJ, Méndez MA, Vásquez M. 2009. Toxicity phenotype does not correlate with phylogeny of *Cylindrospermopsis raciborskii* strains. *Syst Appl Microbiol*. 32:37–48.
Swingley WD, Blankenship RE, Raymond J. 2008. Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Mol Biol Evol*. 25:643–654.
Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives. *Proc Natl Acad Sci U S A*. 103:5442–5447.
Tooming-Klunderud D, Fewer P, Rohrlack T, Jokela J, Rouhiainen L, Sivonen K, Kristensen T, Jakobsen KS. 2008. Evidence for positive selection acting on microcystin synthetase adenylation domains in three cyanobacterial genera. *BMC Evol Biol*. 8:256.
Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15:568–573.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22:2472–2479.
Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res*. 16:1099–1108.
Zimmer RK, Ferrer RP. 2007. Neuroecology, chemical defense, and the keystone species concept. *Biol Bull*. 213:208–225.

Downloaded from http://mbe.oxfordjournals.org/ at Vienna University Library on November 11, 2011