

DETECTION OF SAXITOXIN-PRODUCING DINOFLAGELLATES

Priority Claim

This application claims priority from United States of America Provisional Patent
5 Application No. 61/486,633 filed on 16 May 2011.

Technical Field

The invention generally relates to the field of saxitoxins and the identification of
microorganisms capable of producing them. More specifically, the invention relates to the
10 identification of genes encoding saxitoxin in dinoflagellates, and methods for the specific
detection of dinoflagellates that are producers of saxitoxins.

Background

Saxitoxin (STX) is a potent neurotoxin that occurs in aquatic environments
15 worldwide and has significant economic, environmental and human health impacts.
Ingestion of vector species can lead to paralytic shellfish poisoning, a severe human
illness that may lead to paralysis and death. An estimated 2000 cases of human paralytic
shellfish poisoning, with a mortality rate of 15%, occur globally each year. The costs of
monitoring and mitigation of STX have led to an annual economic loss from harmful
20 plankton blooms calculated at US \$895 million. In freshwater environments, STX is
predominantly produced prokaryotic cyanobacteria. However, in marine environments
eukaryotic dinoflagellates have been associated with the presence of STX. Despite the
association of a number of dinoflagellate species with STX production, the genetic basis
for the production of STX in these microorganisms remains elusive.

25 There is a need for methods to detect the presence (or absence) of STX-producing
dinoflagellates in marine samples.

Summary

A number of studies have unsuccessfully attempted to identify STX pathway genes
30 and/or enzymes in dinoflagellates by enzymatic characterisation, PCR approaches, *in*
silico analyses of expressed sequence tag (EST) libraries, or the use of other publicly
available nucleotide sequences. In addition, several previous studies have suggested that
STX may not in fact be produced by dinoflagellates but instead by co-cultured bacteria.

The present inventors have determined that dinoflagellates are in fact producers of
35 STX and have identified genes responsible for STX production in these microorganisms.
The identification of genes responsible for STX production provides a basis for numerous

molecular tests to detect STX-producing dinoflagellates in both freshwater and marine environments.

In one aspect, there is provided a method for detecting a saxitoxin-producing dinoflagellate in a sample, the method comprising:

- 5 obtaining a sample for use in the method; and
- analyzing the sample for the presence of one or more of a dinoflagellate saxitoxin A catalytic domain polynucleotide or a polypeptide encoded by said polynucleotide;
- wherein:
- 10 the saxitoxin catalytic domain is a saxitoxin A4 catalytic domain; and
- the presence of said polynucleotide or polypeptide indicates the presence of a saxitoxin-producing dinoflagellate in the sample.

In one example, the disclosure provides a method for detecting a saxitoxin-producing dinoflagellate in a sample, the method comprising:

- 15 obtaining a sample for use in the method, and
- analyzing the sample for the presence of one or more of a dinoflagellate saxitoxin A polynucleotide or a polypeptide encoded by said polynucleotide,
- wherein the presence of said polynucleotide or polypeptide indicates the presence of a saxitoxin-producing dinoflagellate in the sample.

20 In another aspect, there is provided a method for determining the absence of a saxitoxin-producing dinoflagellate in a sample, the method comprising:

- obtaining a sample for use in the method; and
- analyzing the sample for the presence of one or more of a dinoflagellate saxitoxin A catalytic domain polynucleotide or a polypeptide encoded by said
- 25 polynucleotide;
- wherein:
- the saxitoxin catalytic domain is a saxitoxin A4 catalytic domain; and
- the absence of said polynucleotide or polypeptide indicates the absence of a saxitoxin-producing dinoflagellate in the sample.

30 In one example, the disclosure provides a method for determining the absence of a saxitoxin-producing dinoflagellate in a sample, the method comprising:

- obtaining a sample for use in the method, and

analyzing the sample for the presence of one or more of a dinoflagellate saxitoxin A polynucleotide or a polypeptide encoded by said polynucleotide,

wherein the absence of said polynucleotide or polypeptide indicates the absence of a saxitoxin-producing dinoflagellate in the sample.

5 In one example, the polynucleotide comprises a saxitoxin A nucleotide sequence selected from those set forth in any one of SEQ ID NOS: 5-197, 224-227, 230-242 and 247 or a fragment or variant of any one of those sequences.

In another example, the polynucleotide comprises a saxitoxin A1 catalytic domain sequence, a saxitoxin A4 catalytic domain sequence, or a fragment thereof.

10 In a further example, the saxitoxin A4 catalytic domain sequence or fragment thereof comprises nucleotides 3115-4121 of the polynucleotide sequence set forth in SEQ ID NO: 3, nucleotides 3597-3721 of the polynucleotide sequence set forth in SEQ ID NO: 3, or a fragment or variant of either sequence.

In yet a further example, the saxitoxin A4 catalytic domain sequence or fragment
15 thereof consists of nucleotides 3115-4121 of the polynucleotide sequence set forth in SEQ ID NO: 3, nucleotides 3597-3721 of the polynucleotide sequence set forth in SEQ ID NO: 3, or a fragment or variant of either sequence.

In one example, the saxitoxin A1 catalytic domain sequence comprises nucleotides 160-1821 of the polynucleotide sequence set forth in SEQ ID NO: 1, nucleotides 277-
20 2022 of the polynucleotide sequence set forth in SEQ ID NO: 3, or a fragment or variant of either sequence.

In another example, the saxitoxin A1 catalytic domain sequence consists of nucleotides 160-1821 of the polynucleotide sequence set forth in SEQ ID NO: 1, nucleotides 277-2022 of the polynucleotide sequence set forth in SEQ ID NO: 3, or a
25 fragment or variant of either sequence.

In yet another example, the analyzing comprises amplification of polynucleotides from the sample by polymerase chain reaction.

In one example, the polymerase chain reaction utilises one or more primers comprising a sequence set forth in SEQ ID NO: 198 or SEQ ID NO: 199, or a fragment
5 or variant of either sequence.

In another example, the polymerase chain reaction utilises one or more primers consisting of a sequence set forth in SEQ ID NO: 198 or SEQ ID NO: 199, or a fragment or variant of either sequence.

In a further example, the polymerase chain reaction utilises one or more primers
10 comprising or consisting of a sequence set forth in any one of SEQ ID NOs: 198-199, 200-211, 220-223, 228-229, and 243-244, or a fragment or variant of any one of those sequences.

In yet a further example, the polypeptide comprises a saxitoxin A amino acid sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 4, or a fragment or variant of either
15 sequence.

In one example, the polypeptide consists of a saxitoxin A amino acid sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 4, or a fragment or variant of either sequence.

In another example, the saxitoxin-producing dinoflagellate is from the *Alexandrium*, *Pyrodinium* or *Gymnodinium* genus.

In a further example, the saxitoxin-producing dinoflagellate is selected from the
20 group consisting of *A. catenella*, *A. fundyense*, *A. lusitanicum*, *A. minutum*, *A. ostenfeldii*, *A. tamarense*, *G. catenatum*, and *P. bahamense var compressum*.

In yet a further example, said analysing is performed using a primer pair of the eleventh or twelfth aspect.

In another aspect, there is provided a kit for the detection of a saxitoxin-producing
25 dinoflagellate in a sample, the kit comprising at least one agent for detecting the presence of a dinoflagellate saxitoxin A catalytic domain polynucleotide or a polypeptide encoded by said polynucleotide, wherein the saxitoxin catalytic domain is a saxitoxin A4 catalytic domain.

In one example, the disclosure provides a kit for the detection of a saxitoxin-producing dinoflagellate in a sample, the kit comprising at least one agent for detecting the presence of a dinoflagellate saxitoxin A polynucleotide or a polypeptide encoded by said polynucleotide.

5 In yet another aspect, there is provided a kit for the determining the absence of a saxitoxin-producing dinoflagellate in a sample, the kit comprising at least one agent for detecting the presence of a dinoflagellate saxitoxin A catalytic domain polynucleotide or a polypeptide encoded by said polynucleotide, wherein the saxitoxin catalytic domain is a saxitoxin A4 catalytic domain.

10 In one example, the disclosure provides a kit for determining the absence of a saxitoxin-producing dinoflagellate in a sample, the kit comprising at least one agent for detecting the presence of a dinoflagellate saxitoxin A polynucleotide or a polypeptide encoded by said polynucleotide.

In one example, the agent binds specifically to a polynucleotide comprising a
15 saxitoxin A nucleotide sequence selected from those set forth in any one of SEQ ID NOS: 5-197, 224-227, 230-242 and 247 or a fragment or variant of any one of those sequences.

In one example, the agent binds specifically to a polynucleotide comprising a saxitoxin A1 catalytic domain sequence, a saxitoxin A4 catalytic domain sequence, or a fragment thereof.

5 In another example, the saxitoxin A4 catalytic domain sequence or fragment thereof comprises nucleotides 3115-4121 of the polynucleotide sequence set forth in SEQ ID NO: 3, nucleotides 3597-3721 of the polynucleotide sequence set forth in SEQ ID NO: 3, or a fragment or variant of either sequence.

10 In a further example, the saxitoxin A4 catalytic domain sequence or fragment thereof consists of nucleotides 3115-4121 of the polynucleotide sequence set forth in SEQ ID NO: 3, nucleotides 3597-3721 of the polynucleotide sequence set forth in SEQ ID NO: 3, or a fragment or variant of either sequence.

15 In yet a further example, the saxitoxin A1 catalytic domain sequence comprises nucleotides 160-1821 of the polynucleotide sequence set forth in SEQ ID NO: 1, nucleotides 277-2022 of the polynucleotide sequence set forth in SEQ ID NO: 3, or a fragment or variant of either sequence.

In one example, the saxitoxin A1 catalytic domain sequence consists of nucleotides 160-1821 of the polynucleotide sequence set forth in SEQ ID NO: 1, nucleotides 277-2022 of the polynucleotide sequence set forth in SEQ ID NO: 3, or a fragment or variant of either sequence.

20 In another example, the agent is a primer, probe or antibody.

In a further example, the agent is a primer comprising a sequence set forth in SEQ ID NO: 198 or SEQ ID NO: 199, or a fragment or variant of either sequence.

In yet a further example, the agent is a primer consisting of a sequence set forth in SEQ ID NO: 198 or SEQ ID NO: 199, or a fragment or variant of either sequence.

25 In one example, the agent is a primer comprising or consisting of a sequence set forth in any one of SEQ ID NOs: 198-199, 200-211, 220-223, 228-229, and 243-244, or a fragment or variant of any one of those sequences.

30 In another example, the agent binds specifically to a saxitoxin A amino acid sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 4, or a fragment or variant of either sequence.

In yet another example, the kit comprises two agents, wherein the two agents are a primer pair of the eleventh or twelfth aspect.

In a further example, the sample is an environmental sample.

In one example, the sample is a saltwater sample.

In another example, the sample is a freshwater sample.

In yet another example, the sample is a marine sample.

5 In one example, the disclosure provides an isolated polynucleotide comprising the sequence set forth in SEQ ID NO: 1, or a variant or fragment thereof.

In one example, the isolated polynucleotide consists of the sequence set forth in SEQ ID NO: 1, or a variant or fragment thereof.

In one example, the disclosure provides an isolated polynucleotide comprising the sequence set forth in SEQ ID NO: 3, or a variant or fragment thereof.

10 In one example, the isolated polynucleotide consists of the sequence set forth in SEQ ID NO: 3, or a variant or fragment thereof.

In one example, the disclosure provides an isolated polynucleotide comprising the sequence set forth in any one of SEQ ID NOS: 5-197, 224-227, 230-242 and 247 or a variant or fragment of any one of those sequences.

15 In one example, the isolated polynucleotide consists of the sequence set forth in any one of SEQ ID NOS: 5-197, 224-227, 230-242 and 247 or a variant or fragment of any one of those sequences.

In one example, the disclosure provides an isolated polypeptide encoded by any one of the polynucleotides disclosed herein.

20 In one example, the disclosure provides a primer or probe that binds specifically to a polynucleotide disclosed herein.

In another example, the disclosure provides an antibody that binds specifically to a polypeptide disclosed herein.

25 In one aspect, there is provided a primer pair for detecting a saxitoxin-producing dinoflagellate in a sample, wherein said primer pair consists of:

a first primer consisting of the polynucleotide sequence of SEQ ID NO: 198 and a second primer consisting of the polynucleotide sequence of SEQ ID NO: 199.

In one example, the disclosure provides an primer pair for detecting a saxitoxin-producing dinoflagellate in a sample, wherein said primer pair comprises a first primer comprising the polynucleotide sequence of SEQ ID NO: 198, or a fragment or variant thereof, and a second primer comprising the polynucleotide sequence of SEQ ID NO: 199, or a fragment or variant thereof.

In one example, the primer pair comprises a first primer consisting of the polynucleotide sequence of SEQ ID NO: 198, or a fragment or variant thereof, and a second primer consisting of the polynucleotide sequence of SEQ ID NO: 199, or a fragment or variant thereof.

In one example, the the disclosure provides a primer pair for detecting a saxitoxin-producing dinoflagellate in a sample, wherein said primer pair comprises:

a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 198 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 199; or

5 a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 200 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 201; or

a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 202 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 203; or

10 a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 204 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 205; or

a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 206 and a second primer comprising or consisting of the polynucleotide sequence of
15 SEQ ID NO: 207; or

a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 208 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 209; or

a first primer comprising or consisting of the polynucleotide sequence of SEQ ID
20 NO: 210 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 211; or

a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 220 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 221; or

25 a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 222 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 223; or

a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 228 and a second primer comprising or consisting of the polynucleotide sequence of
30 SEQ ID NO: 229; or

a first primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 243 and a second primer comprising or consisting of the polynucleotide sequence of SEQ ID NO: 244.

35

Brief Description of the Drawings

Preferred embodiments of the present invention will now be described, by way of an example only, with reference to the accompanying drawings wherein:

Figure 1 is a diagram showing the structure of *sxtA* in dinoflagellates and cyanobacteria. A. Transcript structure of *sxtA* transcripts in *A. fundyense* CCMP1719. B. Genomic *sxtA* structure of *C. raciborskii* T3. C. Structure of STX with bonds and molecules introduced by *sxtA* marked in bold.

Figure 2 is a graph showing GC content of *A. fundyense* *sxtA* transcripts and of cyanobacterial *sxtA* genes. GC content was calculated every 10 bp with a window size of 1000 bp.

Figure 3 shows an *SxtA1* phylogenetic tree. Schematic representation, drawn to scale (for full tree see Figure 5). Maximum likelihood topology is shown. Numbers on nodes represent bootstrap values of maximum likelihood and Bayesian analyses, respectively.

Figure 4A shows an *SxtA4* phylogenetic tree. Schematic representation of phylogenetic tree, drawn to scale (for full tree see Figure 6). Maximum likelihood topology is shown. Numbers on nodes represent bootstrap values of maximum likelihood and Bayesian analyses, respectively.

Figure 4B is a graph showing the genomic copy number of *sxtA4* in *A. catenella* ACSH02 at three different time-points during the growth cycle.

Figure 5 shows an *SxtA1* phylogenetic tree. Maximum likelihood topology is shown. Numbers on nodes represent bootstrap values of maximum likelihood and Bayesian analyses, respectively. Sequences in bold are transcript-derived sequences; either generated using RACE or are contigs from 454 read assembly.

Figure 6 shows an *SxtA4* phylogenetic tree. Maximum likelihood topology is shown. Numbers on nodes represent bootstrap values of maximum likelihood and Bayesian analyses, respectively. Sequences in bold are transcript-derived sequences; either generated using RACE or are contigs from 454 read assembly.

Figure 7 is a map showing Phytoplankton and *S. glomerata* sampling sites referred to Example 2. A. New South Wales, Australia, and the three estuaries in which blooms were sampled: B. Brisbane Water, C. the Georges River and D. Wagonga Inlet. Scale bar in D represents 1km in inset maps B,C and D.

Figure 8 is a graph showing a standard curve of the *sxtA4* primer pair based on dilutions of DNA from known numbers of cells of three exponentially growing *Alexandrium catenella* strains. The assay was tested using DNA concentrations representing 30-2600 cells in the different strains.

Figure 9 provides two graphs showing the abundance of *sxtA4* gene copies (primary y axis) and estimates of *Alexandrium catenella* cells (secondary y axis) based on microscopic cell identifications counts and qPCR using an LSU rDNA primer pair at A. the Georges River and B. Wagonga Inlet sampling sites, during November 2010.

Figure 10 provides a graph showing the number or copies of *sxtA4* L⁻¹ and cells of *Alexandrium catenella*, as determined using a manual count under the light microscopes, each week at the sampling site.

Figure 11 provides a graph showing a regression equation between *Alexandrium catenella* cell abundance and *sxtA4* copies.

Figure 12 shows SEM images of *Alexandrium tamarense* ATNWB01. A) Ventral view, with cell membrane intact, showing general cell size and shape, scale bar = 5 μ m, B) Dorsal view, with cell membrane intact, showing cell shape, scale bar = 10 μ m, C) Epicone in apical view, showing APC and pore on 1' plate, scale bar = 5 μ m, D) Hypocone in antapical view, showing plate patterns, scale bar = 10 μ m, E) Apical pore complex, showing shape of comma, F) Posterior sulcal plate, showing pore, scale bar = 2 μ m.

Figure 13 shows SEM images of strains of *Alexandrium tamarense* ATCJ33, ATEB01, and ATBB01 (for comparison, taken from Hallegraeff et al 1991). A-D, ATCJ33, A) ATCJ33 showing a chain of 2 cells, and general cell size and shape, scale bar = 10 μ m, B) First apical plate, showing ventral pore, C) APC, showing shape of comma, D) First apical plate showing ventral pore, E, F, H ATEB01. E) ATEB01 showing general size and shape of cells, scale bar = 10 μ m, F) ATEB01, first apical plate showing the ventral pore, G) ATBB01 showing the ventral pore and the APC, H) ATEB01 showing the APC.

Figure 14 shows 18S+ITS1-5.8S-ITS2+28S rDNA concatenated phylogeny of the *A. tamarense* complex (2821 characters). The tree is reconstructed with Bayesian inference (Phylobayes). Numbers on the internal nodes represent posterior probability and bootstrap values (>50%) for Phylobayes and RAXML (ordered; Phylobayes/RaxML). Black circles indicate a posterior probability value of 1.00 and bootstrap >90%. Group 1-4 clades have been collapsed, for an expanded version of the phylogeny see

Figure 15 is a toxin profile of ATNWB01 using HPLC, peaks as indicated were determined against PSP toxin standards.

Figure 16 shows 18S+ITS1-5.8S-ITS2+28S rDNA concatenated phylogeny of the *A. tamarense* complex (2821 characters). The tree is reconstructed with Bayesian inference (Phylobayes). Numbers on the internal nodes represent posterior probability and bootstrap values (>50%) for Phylobayes and RAXML (ordered; Phylobayes/RaxML). Black circles indicate a posterior probability value of 1.00 and bootstrap >90%.

Definitions

As used in this application, the singular form “a”, “an” and “the” include plural references unless the context clearly dictates otherwise. For example, the term “a dinoflagellate” also includes a plurality of dinoflagellates.

As used herein, the term “comprising” means “including”. Variations of the word “comprising”, such as “comprise” and “comprises,” have correspondingly varied meanings. Thus, for example, a polynucleotide “comprising” a sequence encoding a protein may consist exclusively of that sequence or may include one or more additional sequences.

As used herein, the term “saxitoxin” encompasses pure saxitoxin and analogs of thereof, examples of which include neosaxitoxin (neoSTX), gonyautoxins (GTX), decarbamoylsaxitoxin (dcSTX), non-sulfated analogs, mono-sulfated analogs, di-sulfated analogs, decarbamoylated analogs and hydrophobic analogs.

Any description of prior art documents herein, or statements herein derived from or based on those documents, is not an admission that the documents or derived statements are part of the common general knowledge of the relevant art.

Detailed Description

The present inventors have determined that dinoflagellates are producers of saxitoxin (STX) and have identified genes responsible for STX production in these microorganisms. Based on this discovery, the inventors have developed an assay capable of discerning between STX-producing dinoflagellate species and dinoflagellate species which do not produce STX.

Accordingly, certain aspects of the invention relate to the provision of STX polynucleotide and polypeptide sequences present in dinoflagellates.

Also provided are methods for the detection of STX-producing dinoflagellates in a given sample based on detecting the presence (or absence) of one or more sequences of the invention in the sample.

Also provided are kits for the detection of STX-producing dinoflagellates in a given sample comprising agent(s) for detecting the presence (or absence) of one or more sequences of the invention in the sample.

Polynucleotides and polypeptides

Disclosed herein are dinoflagellate saxitoxin polynucleotide and polypeptide sequences (“polynucleotides of the invention” and “polypeptides of the invention”,

respectively). Polynucleotides of the invention may be deoxyribonucleic acids (DNA), ribonucleic acids (RNA) or complementary deoxyribonucleic acids (cDNA).

In certain embodiments, the sequences are saxitoxin A gene (*sxtA*) polynucleotide sequences or saxitoxin A polypeptide (STXA) sequences.

5 The *sxtA* polynucleotide sequences may comprise any one or more *sxtA* gene catalytic domain(s) (i.e. the *sxtA1*, *sxtA2*, *sxtA3*, or *sxtA4* catalytic domain(s)), or fragment(s) thereof. The *sxtA1* sequence may be defined by nucleotides 160-1821 of the polynucleotide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 1) or nucleotides 277-2022 of the polynucleotide sequence set forth in GenBank
10 accession number JF343239 (SEQ ID NO: 3). The *sxtA2* sequence may be defined by nucleotides 1837-2415 of the polynucleotide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 1) or nucleotides 2038-2604 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3). The *sxtA3* sequence may be defined by nucleotides 2479-2694 of the polynucleotide sequence set
15 forth in GenBank accession number JF343238 (SEQ ID NO: 1) or nucleotides 2722-2949 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3). The *sxtA4* sequence may be defined by nucleotides 3115-4121 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3) or nucleotides 3597-3721 of the polynucleotide sequence set forth in GenBank
20 accession number JF343239 (SEQ ID NO: 3).

Other examples of *sxtA* gene polynucleotide sequences of the invention include those provided in GenBank accession numbers JF343238 and JF343239 (SEQ ID NO: 1 and SEQ ID NO: 3).

25 The STXA polypeptide sequences may comprise any one or more STX protein catalytic domain(s) (i.e. the STXA1, STXA2, STXA3, or STXA4 catalytic domain(s)) or fragment(s) thereof. The STXA1 sequence may be defined by amino acid residues 1-554 of the polypeptide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 2) or amino acid residues 1-582 of the polypeptide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 4). The STXA2 sequence may be defined by
30 amino acid residues 560-752 of the polypeptide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 2) or amino acid residues 588-776 of the polypeptide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 4). The STXA3 sequence may be defined by amino acid residues 774-845 of the polypeptide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 2) or amino acid residues
35 816-891 of the polypeptide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 4). The STXA4 sequence

may be defined by amino acid residues 947-1281 of the polypeptide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 4).

Other examples of STX polypeptide sequences of the invention include those provided in GenBank accession numbers JF343238 and JF343239 (SEQ ID NO: 2 and
5 SEQ ID NO: 4).

Preferably, the polynucleotide and polypeptide sequences are from saxitoxin-producing dinoflagellates. For example, the polynucleotide and polypeptide sequences may be from dinoflagellates of the order *Gonyaulacales* or *Gymnodiniales*. Preferably, the dinoflagellates are of the genus *Alexandrium* (formerly *Gonyaulax*), *Pyrodinium* or
10 *Gymnodinium*.

Examples of preferred *Alexandrium* species include *A. catenella* (e.g. strains ACCC01, ACSH02, ACTRA02 and CCMP1493), *A. fundyense* (e.g. strains CCMP1719 and CCMP1979), *A. lusitanicum*, *A. minutum* (e.g. strains CCMP1888, CCMP113, ALSP01, ALSP02 and AMD16/AMAD16), *A. ostenfeldii*, and *A. tamarense* (e.g. strains
15 CCMP1771, ATBB01, ATEB01, ATCJ33 and ATNWB01). Examples of preferred *Gymnodinium* species include *G. catenatum* (e.g. strains GCTRA01 and CS-395). Examples of preferred *Pyrodinium* species include *P. bahamense* var *compressum*.

Fragments of both polynucleotides of the invention and polypeptides of the invention are also provided herein.

A polynucleotide "fragment" as contemplated herein is a polynucleotide molecule that is a constituent of a polynucleotide of the invention or variant thereof. Fragments of a polynucleotide do not necessarily need to encode polypeptides which retain biological activity although this is not excluded from being the case. In certain embodiments the fragment may be useful as a hybridization probe or PCR primer. The fragment may be
25 derived by cleaving a polynucleotide of the invention or alternatively may be synthesized by some other means, for example by chemical synthesis. A polynucleotide fragment as contemplated herein may be less than about 5000 nucleotides in length, less than about 4500 nucleotides in length, or less than about 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500, 400, 300, 250, 200, 150, 100, 75, 50, 25 or 15 nucleotides in length.
30 Additionally or alternatively, a polynucleotide fragment as contemplated herein may be more than about 15 nucleotides in length, more than about 25 nucleotides in length, or more than about 50, 75, 100, 150, 200, 250, 300, 400, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500 or 4000 nucleotides in length. Additionally or alternatively, a polynucleotide fragment as contemplated herein may be between about 25 and about 50 nucleotides in
35 length, between about 25 and about 75 nucleotides in length, or between about 25 and about 100, 100 and 250, 100 and 500, 250 and 500, 100 and 1000, 500 and 2000, 1000 and 2000 nucleotides in length.

Polynucleotide fragments of the invention comprise fragments of the *sxtA* gene. For example, polynucleotide fragments of the invention may comprise any one or more of the *sxtA1*, *sxtA2*, *sxtA3*, or *sxtA4* catalytic domain(s), or fragment(s) thereof. Polypeptide fragments of the invention comprise fragments of the STX protein. For example,
 5 polynucleotide fragments of the invention may comprise any one or more of the STXA1, STXA2, STXA3, or STXA4 catalytic domain(s), or fragment(s) thereof.

Specific examples of *sxtA* gene polynucleotide sequence fragments of the invention include those provided in GenBank accession numbers JF343240 - JF343432 (SEQ ID NO: 5 - SEQ ID NO: 197), and those set forth in SEQ ID NOs: 224-227, 230-242 and
 10 247.

Additional specific examples of *sxtA* gene polynucleotide sequence fragments of the invention include those defined by nucleotides 3115-4121 and nucleotides 3597-3721 of the polynucleotide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 3), and fragments and variants thereof.

Specific examples of *sxtA1* catalytic domain sequences include those set out in SEQ ID NOs: 224-227 and 247.
 15

Specific examples of *sxtA4* catalytic domain sequences include those set out in SEQ ID NOs: 230-242.

A polypeptide "fragment" as contemplated herein is a polypeptide molecule is a
 20 constituent of a polypeptide of the invention or variant thereof. Typically the fragment possesses qualitative biological activity in common with the polypeptide of which it is a constituent or though this is not necessarily required. A polypeptide fragment as contemplated herein may be less than about 1500 amino acid residues in length, less than about 1400 amino acid residues in length, or less than about 1300, 1200, 1100, 1000, 900,
 25 800, 700, 600, 500, 400, 300, 250, 200, 150, 100, 75, 50, 25 or 15 amino acid residues in length. Additionally or alternatively, a polypeptide fragment as contemplated herein may be more than about 15 amino acid residues in length, more than about 25 amino acid residues in length, or more than about 50, 75, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acid residues in length. Additionally or
 30 alternatively, a polypeptide fragment as contemplated herein may be between about 15 and about 25 amino acid residues in length, between about 15 and about 50 amino acid residues in length, or between about 15 and 75, 15 and 100, 15 and 150, 25 and 50, 25 and 100, 50 and 100, 50 and 150, 100 and 200, 100 and 250, 100 and 300, 100 and 500, 500 and 750, 500 and 1000, or 1000 and 1300 amino acid residues in length.

Specific examples of STXA polypeptide sequence fragments of the invention include those defined by amino acid residues X-Y of the polypeptide sequence set forth in GenBank accession number JF343248 (SEQ ID NO: 3), those defined by amino acid residues 947-1281 of the polypeptide sequence set forth in
 35

GenBank accession number JF343239 (SEQ ID NO: 4), and fragments and variants thereof.

Variants of polynucleotides of the invention and polypeptides of the invention, and fragments thereof, are also provided herein.

5 A "variant" as contemplated herein refers to a substantially similar sequence. In general, two sequences are "substantially similar" if the two sequences have a specified percentage of amino acid residues or nucleotides that are the same (percentage of "sequence identity"), over a specified region, or, when not specified, over the entire sequence. Accordingly, a "variant" of a polynucleotide and polypeptide sequence
10 disclosed herein may share at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 83%, 85%, 88%, 90%, 93%, 95%, 96%, 97%, 98% or 99% sequence identity with the reference sequence.

In general, polypeptide sequence variants possess qualitative biological activity in common. Polynucleotide sequence variants generally encode polypeptides which
15 generally possess qualitative biological activity in common. Also included within the meaning of the term "variant" are homologues of polynucleotides of the invention and polypeptides of the invention. A polynucleotide homologue is typically from a different dinoflagellate species but sharing substantially the same biological function or activity as the corresponding polynucleotide disclosed herein. A polypeptide homologue is typically
20 from a different dinoflagellate species but sharing substantially the same biological function or activity as the corresponding polypeptide disclosed herein. The term "variant" also includes analogues of the polypeptides of the invention. A polypeptide "analogue" is a polypeptide which is a derivative of a polypeptide of the invention, which derivative comprises addition, deletion, substitution of one or more amino acids, such that the
25 polypeptide retains substantially the same function. The term "conservative amino acid substitution" refers to a substitution or replacement of one amino acid for another amino acid with similar properties within a polypeptide chain (primary sequence of a protein). For example, the substitution of the charged amino acid glutamic acid (Glu) for the similarly charged amino acid aspartic acid (Asp) would be a conservative amino acid
30 substitution.

Typically, polynucleotides of the invention and polypeptides of the invention are "isolated". It will be understood that the term "isolated" in this context means that the polynucleotide or polypeptide has been removed from or is not associated with some or all of the other components with which it would be found in its natural state. For example,
35 an "isolated" polynucleotide may be removed from other polynucleotides of a larger polynucleotide sequence, or may be removed from natural components such as unrelated polynucleotides. Likewise, an "isolated" polypeptide may be removed from other polypeptides of a larger polypeptide sequence, or may be removed from natural

components such as unrelated polypeptides. For the sake of clarity, an "isolated" polynucleotide of polypeptide also includes a polynucleotide or polypeptide which has not been taken from nature but rather has been prepared *de novo*, such as chemically synthesised and/or prepared by recombinant methods. As described herein an isolated polypeptide of the invention may be included as a component part of a longer polypeptide or fusion protein.

In certain embodiments, polynucleotides of the invention may be cloned into a vector. The vector may comprise, for example, a DNA, RNA or complementary DNA (cDNA) sequence. The vector may be a plasmid vector, a viral vector, or any other suitable vehicle adapted for the insertion of foreign sequences, their introduction into cells and the expression of the introduced sequences. Typically the vector is an expression vector and may include expression control and processing sequences such as a promoter, an enhancer, ribosome binding sites, polyadenylation signals and transcription termination sequences. The invention also contemplates host cells transformed by such vectors. For example, the polynucleotides of the invention may be cloned into a vector which is transformed into a bacterial host cell, for example *E. coli*. Methods for the construction of vectors and their transformation into host cells are generally known in the art, and described in standard texts such as, for example, Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, New York; and Ausubel et al. (Eds) *Current Protocols in Molecular Biology* (2007), John Wiley and Sons, Inc..

Probes, primers and antibodies

Polynucleotides of the invention include derivatives and fragments thereof for use as primers and probes.

The derivatives and fragments may be in the form of oligonucleotides. Oligonucleotides are short stretches of nucleotide residues suitable for use in nucleic acid amplification reactions such as PCR, typically being at least about 5 nucleotides to about 80 nucleotides in length, more typically about 10 nucleotides in length to about 50 nucleotides in length, and even more typically about 15 nucleotides in length to about 30 nucleotides in length.

In one embodiment, the probe comprises or consists of a sequence as set forth in SEQ ID NO: 245 or SEQ ID NO: 246.

Probes are nucleotide sequences of variable length, for example between about 10 nucleotides and several thousand nucleotides, for use in detection of homologous sequences, typically by hybridization. Hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides.

Methods for the design and/or production of nucleotide probes and/or primers are known in the art, and described in standard texts such as Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, New York; and publications such as Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Innis et al. (Eds) (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, New York; Innis and Gelfand, (Eds) (1995) *PCR Strategies*, Academic Press, New York; and Innis and Gelfand, (Eds) (1999) *PCR Methods Manual*, Academic Press, New York..

Polynucleotide primers and probes may be prepared, for example, by chemical synthesis techniques such as the phosphodiester and phosphotriester methods (see for example Narang et al. (1979) *Meth. Enzymol.* 68:90; Brown et al. (1979) *Meth. Enzymol.* 68:109; and U.S. Patent No. 4356270), and the diethylphosphoramidite method (see Beaucage et al. (1981) *Tetrahedron Letters*, 22:1859-1862).

Polynucleotides of the invention, including the aforementioned probes and primers, may be labelled by incorporation of a marker to facilitate their detection. Techniques for labelling and detecting nucleic acids are described, for example, in standard texts such as Ausubel et al. (Eds) *Current Protocols in Molecular Biology* (2007), John Wiley and Sons, Inc.. Examples of suitable markers include fluorescent molecules (e.g. acetylaminofluorene, 5-bromodeoxyuridine, digoxigenin, and fluorescein) and radioactive isotopes (e.g. ³²P, ³⁵S, ³H, ³³P). Detection of the marker may be achieved, for example, by chemical, photochemical, immunochemical, biochemical, or spectroscopic techniques.

The probes and primers may be used, for example, to detect or isolate dinoflagellates in a sample of interest. In certain embodiments, the probes and primers may be used to detect STX-producing dinoflagellates in a sample of interest. Additionally or alternatively, the probes or primers may be used to isolate corresponding sequences in other organisms including, for example, other dinoflagellate species. Methods such as the polymerase chain reaction (PCR), hybridization, and the like can be used to identify such sequences based on their sequence homology to the sequences set forth herein. Sequences that are selected based on their sequence identity to the entire sequences set forth herein or to fragments thereof are encompassed by the embodiments. Such sequences include sequences that are orthologs of the disclosed sequences. The term "orthologs" refers to genes derived from a common ancestral gene and which are found in different species as a result of speciation. Genes found in different species are considered orthologs when their nucleotide sequences and/or their encoded protein sequences share substantial identity as defined elsewhere herein. Functions of orthologs are often highly conserved among species.

In hybridization techniques, all or part of a known nucleotide sequence is used to generate a probe that selectively hybridizes to other corresponding nucleic acid sequences

present in a given sample. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labelled with a detectable marker. Thus, for example, probes for hybridization can be made by labelling synthetic oligonucleotides based on the sequences of the invention. The level of
5 homology (sequence identity) between probe and the target sequence will largely be determined by the stringency of hybridization conditions. In particular the nucleotide sequence used as a probe may hybridize to a homologue or other variant of a polynucleotide disclosed herein under conditions of low stringency, medium stringency or high stringency. There are numerous conditions and factors, well known to those skilled
10 in the art, which may be employed to alter the stringency of hybridization such as, for example, the length and nature (DNA, RNA, base composition) of the nucleic acid to be hybridized to a specified nucleic acid; concentration of salts and other components, such as the presence or absence of formamide, dextran sulfate, polyethylene glycol etc; and altering the temperature of the hybridization and/or washing steps.

15 Under a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any organism of interest. Methods for designing PCR primers and PCR cloning are generally known in the art. Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers,
20 degenerate primers, gene-specific primers, vector-specific primers, partially-mismatched primers, and the like.

The skilled addressee will recognise that the primers described herein for use in PCR or RT-PCR may also be used as probes for the detection of dinoflagellate *stx* gene sequences.

25 Also contemplated by the invention are antibodies which are capable of binding specifically to polypeptides of the invention. The antibodies may be used to qualitatively or quantitatively detect and analyse one or more STX polypeptides in a given sample. By "binding specifically" it will be understood that the antibody is capable of binding to the target polypeptide or fragment thereof with a higher affinity than it binds to an unrelated
30 protein. For example, the antibody may bind to the polypeptide or fragment thereof with a binding constant in the range of at least about 10^{-4} M to about 10^{-10} M. Preferably the binding constant is at least about 10^{-5} M, or at least about 10^{-6} M. More preferably the binding constant is at least about 10^{-7} M, at least about 10^{-8} M, or at least about 10^{-9} M or more. In the context of the present invention, reference to an antibody specific to an STX
35 polypeptide of the invention includes an antibody that is specific to a fragment of the polypeptide.

Antibodies of the invention may exist in a variety of forms including, for example, as a whole antibody, or as an antibody fragment, or other immunologically active

fragment thereof, such as complementarity determining regions. Similarly, the antibody may exist as an antibody fragment having functional antigen-binding domains, that is, heavy and light chain variable domains. Also, the antibody fragment may exist in a form selected from the group consisting of, but not limited to: Fv, F_{ab}, F(ab)₂, scFv (single chain Fv), dAb (single domain antibody), chimeric antibodies, bi-specific antibodies, diabodies and triabodies.

An antibody "fragment" may be produced by modification of a whole antibody or by synthesis of the desired antibody fragment. Methods of generating antibodies, including antibody fragments, are known in the art and include, for example, synthesis by recombinant DNA technology. The skilled addressee will be aware of methods of synthesising antibodies, such as those described in, for example, US Patent No. 5296348 and standard texts such as Ausubel et al. (Eds) *Current Protocols in Molecular Biology* (2007), John Wiley and Sons, Inc..

Preferably, antibodies are prepared from discrete regions or fragments of the STX polypeptide of interest. An antigenic portion of a polypeptide of interest may be of any appropriate length, such as from about 5 to about 15 amino acids. Preferably, an antigenic portion contains at least about 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 amino acid residues.

Antibodies that specifically bind to a polypeptide of the invention can be prepared, for example, using purified STX polypeptides of the invention, or purified *sxt* polynucleotide sequences of the invention that encode STX polypeptides of the invention using any suitable methods known in the art. For example, a monoclonal antibody, typically containing Fab portions, may be prepared using hybridoma technology described in Harlow and Lane (Eds) *Antibodies - A Laboratory Manual*, (1988), Cold Spring Harbor Laboratory, N.Y; Coligan, *Current Protocols in Immunology* (1991); Goding, *Monoclonal Antibodies: Principles and Practice* (1986) 2nd ed; and Kohler & Milstein, (1975) *Nature* 256: 495-497. Such techniques include, but are not limited to, antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (see, for example, Huse et al. (1989) *Science* 246: 1275-1281; Ward et al. (1989) *Nature* 341: 544-546).

It will also be understood that antibodies of the invention include humanised antibodies, chimeric antibodies and fully human antibodies. An antibody of the invention may be a bi-specific antibody, having binding specificity to more than one antigen or epitope. For example, the antibody may have specificity for one or more STX polypeptides or fragments thereof, and additionally have binding specificity for another antigen. Methods for the preparation of humanised antibodies, chimeric antibodies, fully

human antibodies, and bispecific antibodies are known in the art and include, for example, those described in United States Patent No. 6995243.

Generally, a sample potentially comprising an STX polypeptide of the invention can be contacted with an antibody that specifically binds the STX polypeptide or fragment thereof. Optionally, the antibody can be fixed to a solid support to facilitate washing and subsequent isolation of the complex, prior to contacting the antibody with a sample. Examples of solid supports include, for example, microtitre plates, beads, ticks, or microbeads. Antibodies can also be attached to a ProteinChip array or a probe substrate as described above.

Detectable labels for the identification of antibodies bound to polypeptides of the invention include, but are not limited to, fluorochromes, fluorescent dyes, radiolabels, enzymes such as horse radish peroxidase, alkaline phosphatase and others commonly used in the art, and colorimetric labels including colloidal gold or coloured glass or plastic beads. Alternatively, the antibody can be detected using an indirect assay, wherein, for example, a second, labelled antibody is used to detect bound polypeptide-specific antibody.

Methods for detecting the presence of, or measuring the amount of, an antibody-marker complex include, for example, detection of fluorescence, chemiluminescence, luminescence, absorbance, birefringence, transmittance, reflectance, or refractive index such as surface plasmon resonance, ellipsometry, a resonant mirror method, a grating coupler wave guide method, or interferometry. Radio frequency methods include multipolar resonance spectroscopy. Electrochemical methods include amperometry and voltammetry methods. Optical methods include imaging methods and non-imaging methods and microscopy.

Useful assays for detecting the presence of, or measuring the amount of, an antibody-marker complex include, include, for example, enzyme-linked immunosorbent assay (ELISA), a radioimmune assay (RIA), or a Western blot assay. Such methods are described in, for example, Stites & Terr, (Eds) (1991) *Clinical Immunology*, 7th ed; and Asai, (Ed) (1993) *Methods in Cell Biology: Antibodies in Cell Biology*, volume 37.

Methods for detecting dinoflagellates

The invention provides methods for the detection and/or isolation of polynucleotides of the invention and/or polypeptides of the invention ("methods of the invention").

In one embodiment the invention provides a method for detecting a dinoflagellate in a sample. The method comprises obtaining a sample for use in the method, and detecting the presence of a polynucleotide of the invention and/or a polypeptide of the invention, or a fragment or variant thereof in the sample. The presence of the polynucleotide,

polypeptide, or variant or fragment thereof in the sample is indicative of dinoflagellates in the sample.

The present inventors have determined that the *sxtA* gene is present in saxitoxin-producing dinoflagellates but absent in dinoflagellates that do not produce saxitoxin. In particular, it has been identified that the detection of *sxtA1* and/or *sxtA4* catalytic domain(s) of the *sxtA* gene are indicative of saxitoxin-producing dinoflagellates.

Accordingly, in another embodiment the invention provides a method for detecting a saxitoxin-producing dinoflagellate in a sample. The method comprises obtaining a sample for use in the method, and detecting the presence of a polynucleotide of the invention and/or a polypeptide of the invention, or a fragment or variant thereof in the sample. The presence of the polynucleotide, polypeptide, or variant or fragment thereof in the sample is indicative of a saxitoxin-producing dinoflagellate in the sample.

In another embodiment the invention provides a method for determining an absence of saxitoxin-producing dinoflagellates in a sample. The method comprises obtaining a sample for use in the method, and determining an absence of a polynucleotide of the invention and/or a polypeptide of the invention, or a fragment or variant thereof in the sample. The absence of the polynucleotide, polypeptide, or variant or fragment thereof in the sample is indicative that saxitoxin-producing dinoflagellates are not present in the sample.

In the context of the methods of the invention (including those referred to in the paragraphs immediately above), the polynucleotide sequence may be a saxitoxin A gene (*sxtA*) sequence. The *sxtA* polynucleotide sequence may comprise any one or more *sxtA* gene catalytic domain(s) (i.e. the *sxtA1*, *sxtA2*, *sxtA3*, or *sxtA4* catalytic domain(s)), or fragment(s) thereof. Preferably, the *sxtA* polynucleotide sequence comprises an *sxtA1* and/or a *sxtA4* domain, or fragment(s) thereof. More preferably, the *sxtA* polynucleotide sequence comprises an *sxtA4* domain, or fragment(s) thereof. The polypeptide sequence may be saxitoxin A polypeptide (STXA) sequence.

In some embodiments, the polynucleotide sequence corresponds to a sequence provided in any one of GenBank accession numbers JF343238 and JF343239 (SEQ ID NO: 1 and SEQ ID NO: 3), or a fragment or a variant of either sequence.

In other embodiments, the polynucleotide sequence corresponds to a sequence provided in any one of GenBank accession numbers JF343240 - JF343432 (SEQ ID NO: 5 - SEQ ID NO: 197), or a fragment or a variant of any one of the sequences.

In other embodiments, the polynucleotide sequence comprises an *sxtA1* catalytic domain sequence provided in any one of SEQ ID NOs: 224-227 and 247, or a fragment or a variant of any one of the sequences.

In other embodiments, the polynucleotide sequence comprises an sxtA4 catalytic domain sequence provided in any one of SEQ ID NOs: 230-242, or a fragment or a variant of any one of the sequences.

In other embodiments, the polynucleotide sequence may comprise nucleotides 160-1821 of the polynucleotide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 1); nucleotides 277-2022 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3); nucleotides 1837-2415 of the polynucleotide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 1); nucleotides 2038-2604 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3); nucleotides 2479-2694 of the polynucleotide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 1); nucleotides 2722-2949 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3); nucleotides 3115-4121 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3); nucleotides 3597-3721 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3); or a fragment or a variant of any one of the sequences.

In the context of the methods of the invention (including those referred to in the paragraphs above), the polypeptide sequence may be a saxitoxin A protein (STXA) sequence. The STXA polypeptide sequence may comprise any one or more STXA protein catalytic domain(s) (i.e. the STXA1, STXA2, STXA3, or STXA4 catalytic domain(s)), or fragment(s) thereof. Preferably, the STXA polypeptide sequence comprises an STXA1 or an STXA4 domain, or fragment(s) thereof. More preferably, the STXA polypeptide sequence comprises an STXA4 domain, or fragment(s) thereof.

In some embodiments, the polypeptide sequence corresponds to a sequence provided in GenBank accession numbers JF343238 or JF343239 (SEQ ID NO: 2 and SEQ ID NO: 4), or a fragment or a variant of either sequence.

In other embodiments, the polypeptide sequence may comprise amino acid residues 1-554 of the polypeptide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 2); amino acid residues 1-582 of the polypeptide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 4); amino acid residues 560-752 of the polypeptide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 2); amino acid residues 588-776 of the polypeptide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 4); amino acid residues 774-845 of the polypeptide sequence set forth in GenBank accession number JF343238 (SEQ ID NO: 2); amino acid residues 816-891 of the polypeptide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 4); amino acid residues 947-1281 of the polypeptide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 4); or a fragment or a variant of any one of the sequences.

Dinoflagellates detected in a sample or determined to be absent from a sample using the methods of the invention may be saxitoxin-producing dinoflagellates. Without imposing any particular limitation, the dinoflagellates may be from the order *Gonyaulacales* or *Gymnodiniales*. The dinoflagellates may be from the genus *Alexandrium* (formerly *Gonyaulax*), *Pyrodinium* or *Gymnodinium*. Suitable examples of *Alexandrium* species include *A. catenella* (e.g. strains ACCC01, ACSH02, ACTRA02 and CCMP1493), *A. fundyense* (e.g. strains CCMP1719 and CCMP1979), *A. lusitanicum*, *A. minutum* (e.g. strains CCMP1888, CCMP113, ALSP01, ALSP02 and AMD16/AMAD16), *A. ostenfeldii* and *A. tamarense* (e.g. strains CCMP1771, ATBB01, ATEB01, ATCJ33 and ATNWB01). Suitable examples of *Gymnodinium* species include *G. catenatum* (e.g. strains GCTRA01 and CS-395). Suitable examples of *Pyrodinium* species include *P. bahamense* var *compressum*.

A sample for use in the methods of the invention may be "obtained" by any means. For example, the sample may be obtained by removing it from a naturally-occurring state (e.g. a sample from a lake, ocean or river), or, by removing it from a "non-natural" state (e.g. a culture in a laboratory setting, dam, reservoir, tank etc.).

A sample for use in the methods of the invention may be suspected of comprising one or more dinoflagellates, or one or more saxitoxin-producing dinoflagellates. The sample may be a comparative or control sample, for example, a sample comprising a known concentration or density of dinoflagellates or saxitoxin-producing dinoflagellates or a sample comprising one or more known species or strains of dinoflagellates or saxitoxin-producing dinoflagellates. The sample may be derived from any source. For example, a sample may be an environmental sample. The environmental sample may be derived from, for example, saltwater, freshwater, a river, a lake, an ocean, or coastal waters. The environmental sample may be derived from a dinoflagellate bloom. Alternatively, the sample may be derived from a laboratory source, such as a culture, or a commercial source. Alternatively, the sample may be derived from a biological source such as, for example, tissue or biological fluid. The sample may be modified from its original state, for example, by purification, dilution or the addition of any other component or components.

In certain embodiments, a sample tested using the methods of the invention may provide information regarding the presence or absence of saxitoxin in animals populating the source of the sample. For example, the sample may be tested to determine the presence or absence of saxitoxin in animal seafoods such as, for example, fish (e.g. pufferfish) and in particular shellfish (e.g. mussels, clams, oysters, scallops and the like).

Polynucleotides and polypeptides for use in methods of the invention may be isolated (i.e. extracted) from microorganisms either in mixed culture or as individual species or genus isolates. Accordingly, the microorganisms of a sample may be cultured

prior to extraction or the extraction may be performed directly on a given sample. Suitable methods for the isolation (i.e. extraction) and purification of polynucleotides and polypeptides for analysis using methods of the invention are generally known in the art and are described, for example, in standard texts such as Ausubel (Eds) *Current Protocols in Molecular Biology* (2007), John Wiley and Sons, Inc; Coligan et al. (Eds) *Current*
 5 *Protocols in Protein Science* (2007), John Wiley and Sons, Inc; Walker, (Ed) (1988) *New Protein Techniques: Methods in Molecular Biology*, Humana Press, Clifton, N.J; and Scopes, R. K. (1987) *Protein Purification: Principles and Practice*, 3rd. Ed., Springer-Verlag, New York, N.Y.. Additional methods are described in Neilan (1995) Appl.
 10 Environ. Microbiol. 61:2286-2291. Suitable polypeptide purification techniques suitable for use in the methods of the invention include, reverse-phase chromatography, hydrophobic interaction chromatography, centrifugation, gel filtration, ammonium sulfate precipitation, and ion exchange.

In alternative embodiments, methods of the invention may be performed without
 15 isolating nucleic acids and/or polypeptides from the sample.

Detecting the presence (or determining the absence) of polynucleotides of the invention and/or polypeptides of the invention in a given sample may be performed using any suitable technique. Suitable techniques may typically involve the use of a primer, probe or antibody specific for any one or more polynucleotides of the invention or any
 20 one or more polypeptides of the invention. Suitable techniques include, for example, the polymerase chain reaction (PCR) and related variations of this technique (e.g. quantitative PCR), antibody based assays such as ELISA, western blotting, flow cytometry, fluorescent microscopy, and the like. These and other suitable techniques are generally known in the art and are described, for example, in standard texts such as Coligan et al.
 25 (Eds) *Current Protocols in Protein Science* (2007), John Wiley and Sons, Inc; Walker, (Ed) (1988) *New Protein Techniques: Methods in Molecular Biology*, Humana Press, Clifton, N.J; and Scopes (1987) *Protein Purification: Principles and Practice*, 3rd. Ed., Springer-Verlag, New York, N.Y..

In preferred embodiments, detecting the presence (or determining the absence) of
 30 polynucleotides of the invention in a given sample is achieved by amplification of nucleic acids extracted from a sample of interest by polymerase chain reaction using primers that hybridise specifically to the polynucleotide sequence, and detecting the amplified sequence. Under the PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify polynucleotides of the invention such as, for example, RNA
 35 (e.g. mRNA), DNA and/or cDNA polynucleotides. Suitable methods of PCR include, but are not limited to, those using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially-mismatched primers, and the like. Methods for designing PCR and RT-PCR primers are generally

known in the art and are disclosed, for example, in standard texts such as Ausubel et al. (Eds) *Current Protocols in Molecular Biology* (2007), John Wiley and Sons, Inc; Maniatis et al. *Molecular Cloning* (1982), 280-281; Innis et al. (Eds) (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis and
5 Gelfand, (Eds) (1995) *PCR Strategies* (Academic Press, New York); Innis and Gelfand, (Eds) (1999) *PCR Methods Manual* (Academic Press, New York); and Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed., Cold Spring Harbor Laboratory Press, Plainview, New York.

The skilled addressee will readily appreciate that various parameters of PCR and
10 RT-PCR procedures may be altered without affecting the ability to obtain the desired product. For example, the salt concentration may be varied or the time and/or temperature of one or more of the denaturation, annealing and extension steps may be varied. Similarly, the amount of DNA, cDNA, or RNA template may also be varied depending on the amount of nucleic acid available or the optimal amount of template required for
15 efficient amplification. The primers for use in the methods and kits of the present invention are typically oligonucleotides typically being at least about 5 nucleotides to about 80 nucleotides in length, more typically about 10 nucleotides in length to about 50 nucleotides in length, and even more typically about 15 nucleotides in length to about 30 nucleotides in length.

20 The skilled addressee will recognise that primers of the invention may be useful for a number of different applications, including but not limited to, PCR, RT-PCR, and as probes for the detection of polynucleotides of the invention. Such primers can be prepared by any suitable method, including, for example, direct chemical synthesis or cloning and restriction of appropriate sequences. Not all bases in the primer need reflect the sequence
25 of the template molecule to which the primer will hybridize. The primer need only contain sufficient complementary bases to enable the primer to hybridize to the template. A primer may also include mismatch bases at one or more positions, being bases that are not complementary to bases in the template, but rather are designed to incorporate changes into the DNA upon base extension or amplification. A primer may include
30 additional bases, for example in the form of a restriction enzyme recognition sequence at the 5' end, to facilitate cloning of the amplified DNA.

The methods of the invention involve detecting the presence (or determining the absence) of polynucleotides of the invention and/or polypeptides of the invention in a given sample. As noted above, the sequences may comprise saxitoxin A sequences
35 including any one or more of the saxitoxin A1, A2, A3 or A4 catalytic domain sequences (or fragment(s) thereof). Preferably, the sequence comprises the saxitoxin A1 and/or the saxitoxin A4 domain (or fragment(s) thereof). More preferably, the sequence comprises the saxitoxin A4 domain (or fragment(s) thereof).

The skilled addressee will recognise that any primer(s) capable of the amplifying a polynucleotide of the invention, any probe capable of detecting a polynucleotide of the invention, or any antibody capable of detecting a polypeptide of the invention, may be used when performing the methods of the invention. In preferred embodiments, the
5 primers, probes and antibodies bind specifically to any one or more of the saxitoxin A sequences referred to in the preceding paragraph (i.e. paragraph directly above).

By "binding specifically" it will be understood that the primer, probe or antibody is capable of binding to the target sequence with a higher affinity than it binds to an unrelated sequence. Accordingly, when exposed to a plurality of different but equally
10 accessible sequences as potential binding partners, the primer, probe or antibody specific for a target sequence will selectively bind to the target sequence and other alternative potential binding partners will remain substantially unbound by the primer, probe or antibody. In general, a primer, probe or antibody specific for a target sequence will preferentially bind to the target sequence at least 10-fold, preferably 50-fold, more
15 preferably 100-fold, and most preferably greater than 100-fold more frequently than to other potential sequences that are not target sequences. A primer, probe or antibody specific for a target sequence may be capable of binding to non-target sequences at a weak, yet detectable level. This is commonly known as background binding and is readily discernible from specific binding, for example, by use of an appropriate control.

20 In preferred embodiments, the primers, probes or antibodies bind specifically to a saxitoxin A1 or A4 catalytic domain polynucleotide or polypeptide sequence, or a fragment thereof. More preferably the primers, probes or antibodies bind specifically to a saxitoxin A4 catalytic domain polynucleotide or polypeptide sequence, or a fragment thereof.

25 Suitable primers and probes may bind specifically to any fragment of a saxitoxin A1 catalytic domain polynucleotide sequence. Suitable antibodies may bind specifically to a fragment of a saxitoxin A1 catalytic domain polypeptide sequence encoded by such polynucleotide sequences.

Suitable primers and probes may bind specifically to any fragment of a saxitoxin
30 A4 catalytic domain polynucleotide sequence. suitable primers and probes may bind specifically to a fragment of the saxitoxin A4 catalytic domain polynucleotide sequence defined by nucleotides 3115-4121 of the polynucleotide sequence set forth in GenBank accession number JF343239 (SEQ ID NO: 3). Suitable antibodies may bind specifically to a fragment of a saxitoxin A4 catalytic domain polypeptide sequence encoded by such
35 polynucleotide sequences.

In some embodiments, the methods of the invention may involve detecting the presence (or determining the absence) of polynucleotides of the invention in a sample using PCR amplification. Suitable oligonucleotide primer pairs for the PCR amplification

of saxitoxin A polynucleotide sequences may be capable of amplifying any one or more catalytic domain(s) of the *sxt* gene, or fragments(s) thereof. Preferably, the primers amplify a sequence comprising a saxitoxin A4 catalytic domain polynucleotide sequence, or a fragment thereof. A suitable primer pair for this purpose may comprise a first primer comprising the polynucleotide sequence of SEQ ID NO: 198, or a fragment or variant thereof, and/or a second primer comprising the polynucleotide sequence of SEQ ID NO: 199, or a fragment or variant thereof. Other examples of suitable primer pairs include those set forth in SEQ ID NOs: 200-211, 220-223, 228-229, and 243-244 (including fragments and variants of these primer pair sequences).

The skilled addressee will recognise that the exemplified primers are not intended to limit the region of the saxitoxin A gene amplified or the methods of the invention in general. The skilled addressee will also recognise that the invention is not limited to the use of the specific primers exemplified, and alternative primer sequences may also be used, provided the primers are designed appropriately so as to enable the amplification of saxitoxin polynucleotide sequences, preferably saxitoxin A polynucleotide sequences, and more preferably saxitoxin A1 and/or A4 domain polynucleotide sequences.

In other embodiments, the methods of the invention may involve detecting the presence (or determining the absence) of polynucleotides of the invention in a sample by the use of suitable probes. Probes of the invention are based on *sxt* polynucleotide sequences of the invention. Probes are nucleotide sequences of variable length, for example between about 10 nucleotides and several thousand nucleotides, for use in detection of homologous sequences, typically by hybridization. Hybridization probes of the invention may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides. Probes of the invention may be labelled by incorporation of a marker to facilitate their detection. Examples of suitable markers include fluorescent molecules (e.g. acetylaminofluorene, 5-bromodeoxyuridine, digoxigenin, fluorescein) and radioactive isotopes (e.g. ³²P, ³⁵S, ³H, ³³P). Detection of the marker may be achieved, for example, by chemical, photochemical, immunochemical, biochemical, or spectroscopic techniques. Methods for the design and/or production of nucleotide probes are generally known in the art, and are described, for example, in standard texts such as Robinson et al. (Eds) *Current Protocols in Cytometry* (2007), John Wiley and Sons, Inc; Ausubel et al. (Eds) *Current Protocols in Molecular Biology* (2007), John Wiley and Sons, Inc; Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed., Cold Spring Harbor Laboratory Press, Plainview, New York; and Maniatis et al. (1982) *Molecular Cloning*, 280-281.

In other embodiments, the methods of the invention may involve detecting the presence (or determining the absence) of polypeptides of the invention in a sample using

antibodies. The antibodies may be used to qualitatively or quantitatively detect and analyse one or more STX polypeptides of the invention in a given sample. The antibodies may be conjugated to a fluorochrome allowing detection, for example, by flow cytometry, immunohistochemistry or other means known in the art. Alternatively, the antibody may
5 be bound to a substrate allowing colorimetric or chemiluminescent detection. The invention also contemplates the use of secondary antibodies capable of binding to one or more antibodies capable of binding specifically to a polypeptide of the invention.

Kits for detecting dinoflagellates

10 The invention also provides kits for the detection and/or isolation of polynucleotides of the invention and/or polypeptides of the invention ("kits of the invention").

In certain embodiments the kits are used for detecting a dinoflagellate in a sample.

In other embodiments the kits are used for detecting a saxitoxin-producing
15 dinoflagellate in a sample.

In other embodiments the kits are used for determining an absence of saxitoxin-producing dinoflagellates in a sample.

In general, the kits of the invention comprise at least one agent for detecting the presence of one or more polynucleotides of the invention and/or one or more polypeptides
20 of the invention, and/or variants or fragments thereof (see description in the section above entitled "Polynucleotides and polypeptides"). Any agent suitable for this purpose may be included in the kits. Examples of suitable agents include primers, probes and antibodies such as those described above in the sections entitled "Probes, primers and antibodies" and "Methods for detecting dinoflagellates".

25 In preferred embodiments, the kits are for use in the methods of the invention (see description in the section above entitled "Methods for detecting dinoflagellates").

In some embodiments the invention provides a kit for the detection of a dinoflagellate in a sample, the kit comprising at least one agent for detecting in the sample the presence of one or more polynucleotides of the invention, and/or one or more
30 polypeptides of the invention, and/or a variant or fragment of either. Preferably, the dinoflagellate is a saxitoxin-producing dinoflagellate.

In other embodiments the invention provides a kit for determining the absence of a dinoflagellate in a sample, the kit comprising at least one agent for determining in the sample the absence of one or more polynucleotides of the invention, and/or one or more
35 polypeptides of the invention, and/or a variant or fragment of either. Preferably, the dinoflagellate is a saxitoxin-producing dinoflagellate.

In general, kits of the invention may comprise any number of additional components. The additional components may include

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as described in the specific embodiments without departing from the scope of the invention as broadly described.

The invention will now be described with reference to specific examples.

25 *Materials and Methods*

Saxitoxin-producing and non-producing dinoflagellate cultures were obtained from various culture collections (Table 1).

Table 1. List of dinoflagellate strains used in this study, their production of STX and whether *sxtA1* and *sxtA4* fragments were amplified from their genomic DNA.

Order	Genus	Species	Strain	STX	PCR sxtA1	PCR sxtA4
<i>Gonyaulacales</i>						
<i>Alexandrium</i>						

28

<i>affine</i>	CCMP112	n. d.	n. d.	n. d.
<i>affine</i>	AABB01/01	n. d.	n. d.	n. d.
<i>affine</i>	AABB01/02	n. d.	n. d.	n. d.
<i>andersonii</i>	CCMP1597	n. d.	n. d.	n. d.
<i>andersonii</i>	CCMP2222	n. d.	n. d.	n. d.
<i>catenella</i>	ACCC01	yes	yes	yes
<i>catenella</i>	ACSH02	yes	yes	yes
<i>catenella</i>	ACTRA02	yes	yes	yes
<i>catenella</i>	CCMP1493	yes	yes	yes
<i>fundyense</i>	CCMP1719	yes	yes	yes
<i>fundyense</i>	CCMP1979	yes	yes	yes
<i>minutum</i>	CCMP1888	yes	yes	yes
<i>minutum</i>	CCMP113	yes	yes	yes
<i>minutum</i>	ALSP01	yes	yes	yes
<i>minutum</i>	ALSP02	yes	yes	yes
<i>minutum</i>	AMD16	yes	yes	yes
<i>tamarensis</i>	CCMP1771	n. d.	yes	yes
<i>tamarensis</i>	ATBB01	n. d.	yes	yes
<i>tamarensis</i>	ATEB01	n. d.	yes	yes
<i>tamarensis</i>	ATCJ33	n. d.	yes	yes
<i>tamarensis</i>	ATNWB01	yes	yes	yes
<i>Gambierdiscus</i>				
<i>australes</i>	CAWD148	*	n. d.	n. d.
<i>Ostreopsis</i>				
<i>ovata</i>	CAWD174	*	n. d.	n. d.
<i>siamensis</i>	CAWD96	*	n. d.	n. d.
<i>Gymnodiniales</i>				
<i>Amphidinium</i>				
<i>massarti</i>	CS-259	*	n. d.	n. d.
<i>Gymnodinium</i>				
<i>catenatum</i>	GCTRA01	yes	yes	yes
<i>catenatum</i>	CS-395	yes	yes	yes
<i>Prorocentrales</i>				
<i>Prorocentrum</i>				
<i>lima</i>	CS-869	*	n. d.	n. d.

n.d. not detected, *species never reported to synthesize STX

Cultures were maintained in GSe (see method in Doblin *et al.* (1999), Growth and biomass stimulation of the toxic dinoflagellate *Gymnodinium catenatum* (Graham) by dissolved organic substances. J Exp Mar Biol Ecol 236: 33-47) or L1 media (see method in Guillard and Hargraves (1993) *Stichochrysis immobilis* is a diatom, not a chrysophyte. Phycologia 32, 234-236) at 16 - 20 °C, under a 12/12 light cycle, and a photon irradiance of ~100 micromoles of photons m⁻² s⁻¹. Toxicity of strains was determined using HPLC or LCMS. The detection limit of the HPLC method ranged from about 0.07 µg STXeq/ 100 g for C1 and C3 to 4.1 µg STXeq/ 100 g for GTX1. The detection limit for the LCMS method ranged from about 0.1 pg/ cell for NEO and STX to 0.5 pg/ cell for C1 and C2.

- RNA and DNA extraction

To isolate total RNA for the 454-library construction (see below), cultures of *Alexandrium fundyense* Balech CCMP1719 and *Alexandrium minutum* Halim CCMP113 were harvested in exponential phase through centrifugation (1 min, 1000 x g, 12°C). Cells were washed with PBS, exposed to bead-beating on dry ice with the Fast Prep bead-beater from Medinor (20 s, speed 4) using 1.4 mm beads (Medinor) and total RNA was extracted with the ChargeSwitch® Total RNA Cell kit (Invitrogen) according to the manufacturers' protocol.

For RACE analyses, polyA-enriched mRNA was isolated using the Dynabeads DIRECT kit (Invitrogen). Cells were harvested by centrifugation (2 min, 4 °C, 16000 x g), were washed twice with PBS, the lysis/ binding buffer was added, and this was homogenised using the bead-beater (20 s, step 4). After centrifugation (1 min, 4 °C, 16000 x g), the clear homogenate was transferred to the Dynabeads mix and the mRNA isolated according to protocol. Finally, mRNA was treated with TURBO™ DNase (Ambion) according to the protocol supplied.

Genomic DNA was isolated from all dinoflagellate strains listed in Table 1 by either using the Genomic DNA plant ChargeSwitch® kit (Invitrogen) according to the manufacturer's protocol, or by the CTAB method.

Quality and quantity of RNA and DNA were determined using a Nanodrop spectrophotometer (ThermoScientific), by amplifying control dinoflagellate genes (cytochrome b, actin) and/ or by visualizing them on an ethidium bromide stained agarose gel.

- cDNA library construction, 454 sequencing, assembly and analyses

Normalized polyA-enriched cDNA libraries with 454 adapters attached at each end were constructed commercially by Vertis Biotechnologie AG (<http://www.vertis-biotech.com/>). Half a plate each of *A. fundyense* CCMP 1719 and *A. minutum* CCMP113 libraries were sequenced using Roche 454 sequencing TITAN technology at the Norwegian High-Throughput Sequencing Centre (<http://www.sequencing.uio.no/>). Only 454 reads that possessed at least one cDNA adaptor were considered further. Adaptors and, where present, full and partial dinoflagellate spliced-leader (SL) sequences were removed prior to assembly using an in-house PERL script which is now integrated in the bioinformatic tool CLOTU. Reads were assembled using the software program Mira v3.0.5 with the main switches 'denovo', 'est', 'accurate' and '454'.

To identify putative *sxt* gene sequences within the two 454 libraries, custom BLAST searches were performed at the freely available online data portal 'Bioportal' (www.biportal.no). Two strategies were used: the cyanobacterial *sxt* genes were queried either against the assembled *Alexandrium* datasets or the unassembled 454 read datasets.

All hits with an e-value < 0.1 were extracted and the sequence with the lowest e-value for each gene was blasted against the non-redundant protein database at NCBI.

For *sxtA*, all retrieved sequences were re-assembled in the software program CLC Bio Main Workbench, using a minimum overlap of 10 bp and low or high alignment stringency. Resulting contig sequences were blasted against the non-redundant and EST databases at NCBI using algorithms blastn, blastx and tblastx. The structure of *sxtA* transcripts was determined by aligning their translated sequence to *sxtA* from cyanobacteria, as well as by conserved domains searches (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Catalytic and substrate-binding residues of *sxtA* from cyanobacteria were previously determined. The transcripts were searched for the presence possible signal peptides and corresponding cleavage sites using the neural networks and hidden Markov models implemented in SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) and the 3-layer approach of Signal-3L (<http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/>). Transmembrane helices were explored using TMHMM server 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) and hydrophobicity profiles with Kyte-Doolittle plots.

- RACE analyses

Primers were designed in conserved regions of the contigs with high similarity to *sxtA* using Primer3 software (<http://frodo.wi.mit.edu/primer3/>; Table 2).

Table 2. Primers used in PCR and sequencing

Name	Sequence 5' - 3'	Orientation	Description
sxt001	TGCAGCGMTGCTACTCCTACTAC (SEQ ID NO: 200)	Forward	binds within sxtA1, designed on 454 reads
sxt002	GGTCGTGGTCYAGGAAGGAG (SEQ ID NO: 201)	Reverse	binds within sxtA1, designed on 454 reads
sxt007	ATGCTCAACATGGGAGTCATCC (SEQ ID NO: 202)	Forward	binds within sxtA4, designed on 454 reads
sxt008	GGGTCCAGTAGATGTTGACGATG (SEQ ID NO: 203)	Reverse	binds within sxtA, designed on 454 reads
Additional primers used for RACE analyses and sequencing			
sxt013	GTAGTAGGAGTAGCKACGCTGCA (SEQ ID NO: 204)	Reverse	reverse complement of sxt001
sxt014	CTCCTTCCTRGACCACGACC (SEQ ID NO: 205)	Forward	reverse complement of sxt002
sxt015	GGATGACTCCCATGTTGAGCAT (SEQ ID NO: 206)	Reverse	reverse complement of sxt007
sxt016	CATCGTCAACATCTACTGGACCC (SEQ ID NO: 207)	Forward	reverse complement of sxt008
sxt019	GGCAAGTATCTCCGCAGGCTTAC (SEQ ID NO: 208)	Reverse	binds within sxtA1, upstream of sxt002

sxt020	CGTGGAGGAGCATGTTGACAGAATC (SEQ ID NO: 209)	Forward	binds within sxtA1, downstream of sxt001
sxt026	ACTCGACAGGCCGGCAGTACAGAT (SEQ ID NO: 210)	Reverse	binds with sxtA4, upstream of sxt008
sxt040	TGAGCAGGCACGCAGTCC (SEQ ID NO: 211)	Forward	binds within sxtA1 on the long transcript
Primers to amplify clones directly			
TopoF	GGCTCGTATGTTGTGTGGAATTGTG (SEQ ID NO: 212)	Forward	binds within pCR®2.1-TOPO® vector
TopoR	AGTCACGACGTTGTAAACGACGG (SEQ ID NO: 213)	Reverse	binds within pCR®2.1-TOPO® vector

First-strand cDNA was synthesized with ~95 ng polyA-enriched mRNA using the adaptor primer AP according to the manufacturer's instructions for transcripts with high GC content (3'RACE System, Invitrogen). Following RNase H treatment, the RACE product was 1:10 diluted and used as template for PCR. To amplify the 5' end of the transcript, three different protocols were used. First, the method of Zhang (Zhang *et al.* (2007) Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci USA* 104: 4618-4623) was used with slight modifications: the 3'RACE library described above was amplified with the primers AUAP (adapter primer supplied with the kit) and dinoSL to enrich for full transcripts (PCR program: 94 °C - 60 s; 30 x (94 °C - 30 s, 68 °C - 5 min); 68 °C - 10 min; 8 °C hold; PCR chemistry see below). The PCR product was 1:10 diluted and used as template in nested PCRs, which were amplified using the dinoSL primer as forward and several different internal reverse primers (Table 2). Further, these experiments used the two kits 5'RACE System (Invitrogen) and the GeneRacer kit (Invitrogen), using the provided 5' Adapter primers and several different internal reverse primers (Table 2). All products were cloned and sequenced as described below.

- PCR and sequencing

All PCR reactions were carried out in 25 µl volumes containing template, 1 unit 10X BD Advantage 2 PCR buffer (BD Biosciences), 0.2 mM dNTPs, 0.5 µM of each forward and reverse primer (Table 2), DMSO (10 % final concentration) and 0.25 units 50X BD Advantage 2 Polymerase Mix (BD Biosciences). If not stated otherwise, PCRs were amplified as follows: 94 °C - 2.5 min; 5 x (94 °C - 30 s; 68 °C - variable); 5 x (94 °C - 30 s; 66 °C - 30 s; 68 °C - variable); 25x (94 °C - 30 s; 64 °C - 30 s; 68 °C - variable); 68 °C - 10 min; 8 °C - hold. PCR products were visualized on 1 % ethidium bromide stained agarose gels, cut out and cleaned with the Wizard® SV Gel and PCR Clean-up System (Promega) and cloned with the TOPO TA® cloning kit according to the manufacturer's instructions (Invitrogen; pCR®2.1-TOPO® vector; One Shot® Mach1™ T1 Phage-Resistant Chemically Competent E. coli cells). Individual colonies were directly added to 25 µl PCR reactions containing 1 unit 10 X standard PCR buffer (Qiagen), 0.4 µM primer TopoF and TopoR (Table 2), 0.2 mM dNTPs, and 1 unit

HotStarTaq (Qiagen). Cycling conditions were 95 °C - 15 min, 30 x (94 °C - 30 s; 60 °C - 30 s; 72 °C - 90 s), 72 °C - 5 min, 8 °C - hold. PCR products were diluted and Sanger sequenced directly from both sides using the primers M13F and M13R supplied with the cloning kit.

5

- *SxtA1 and sxtA4 genomic amplification*

All dinoflagellate strains (Table 1) were tested for the presence of putative *sxtA1* and *sxtA4* genes. PCRs were run using gDNA according to the protocol described above. The *sxtA1* fragment was amplified with primers sxt001 & sxt002 (~550 bp) and the *sxtA4* fragment with the primers sxt007 & sxt008 (~750 bp) (Table 2).

10

- *Phylogenetic analyses*

Dinoflagellate nucleotide sequences were aligned manually using MacClade v4.07 (see Maddison and Maddison (1992) MacClade. 3 ed: Sinauer Associates) considering the coding sequence in the correct reading frame before being translated to the corresponding amino-acid sequence. The dinoflagellate amino acid sequences were subsequently aligned, using MAFFT v6 L-INS-I model to the orthologous *sxt* sequences for cyanobacteria, in addition to a selection of closely related NCBI nr Blastp hits, constituting the outgroup. Resulting alignments were checked manually and poorly aligned positions excluded using MacClade v4.07 (see Maddison and Maddison (1992) *supra*).

20

ProtTest v2.4 (see Abascal *et al.* (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104-2105) determined WAG as the optimal evolutionary model for all inferred alignments. Maximum Likelihood (ML) analyses were performed with RAxML-VI-HPC v7.2.6, PROTCATWAG model with 25 rate categories (see Stamatakis (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690). The most likely topology was established from 100 separate searches and bootstrap analyses were performed with 100 pseudo-replicates. Bayesian inferences were performed using Phylobayes v3.2e (see Lartillot and Philippe (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21: 1095-1109; and Lartillot and Philippe (2006) Computing Bayes factors using thermodynamic integration. Syst Biol 55: 195-207) under the same substitution model with a free number of mixing categories and a discrete across site variation under 4 categories. Trees were inferred when the largest maximum difference between the bipartitions (chains) was <0.1. All model estimation and phylogenetic analyses were done on the freely available 'Bioportal' (<http://www.bioportal.uio.no/>).

30

35

- Copy number determination

Triplicate 200 ml batch cultures of *Alexandrium catenella* strain ACSH02 were grown as previously described, and abundance was counted every three days using a Sedgewick-Rafter chamber and inverted light microscope (Leica Microsystems). Ten ml samples for gDNA extraction were taken in early exponential, late exponential and stationary phase.

Primers suitable for qPCR were designed based on conserved regions in an alignment of *A. fundyense* and *A. minutum* 454 reads covering the *sxtA4* region using Primer 3 software amplifying a 161 bp product. qPCR cycles were carried out on a Rotor Gene 3000 (Corbett Life Science) using SYBR Green PCR Master Mix (Invitrogen). qPCR assays were performed in a final volume of 25 μ l volume consisting of 12.5 μ l SYBR Green PCR master mix, 1 μ l of template DNA, 1 μ l of each primer pair, 1 μ l of BSA and 8.5 μ l of MilliQ water. qPCR assays were performed in triplicate with the following protocol: 95 °C for 10 s, and 35 cycles of 95 °C for 15 s and 60 °C for 30 s. Melting curve analysis was performed at the end of each program to confirm amplification specificity, and select PCR products were sequenced. The standard curve was constructed from a 10-fold dilution series of a known concentration of fresh PCR product, ranging from $2 - 2 \times 10^{-5}$ ng. The molecules of PCR product were determined: $(A \times 6.022 \times 10^{23}) \times (660 \times B)^{-1}$ with A: concentration of PCR product, 6.022×10^{23} ; Avogadro's number, 660: average molecular weight per base pair and B: length of PCR product. The number of molecules in the unknown samples were determined and divided by the known number of cells in the qPCR template to obtain copy number per cell. The detection limit was around 5000 copies of the gene sequence (i.e. ~20 - 30 cells per assay, each with ~200 copies of the sequence). However, the analyses were run with 10 - 100-fold this number of cells, and thus not run at or close to the detection limit.

Results

- Identification of *sxt* sequences in the transcriptome of *A. minutum* and *A. fundyense*

454 sequencing resulted in 589,410 raw reads for *A. minutum* and 701,870 raw reads for *A. fundyense* (SRA028427.1: samples SRS151150.1 and SRS151148.1, respectively). After quality control, the reads were assembled into 44,697 contigs and 539 singletons for *A. minutum* and 51,861 contigs and 163 singletons for *A. fundyense*. The contig lengths and GC contents were similar for both libraries: the mean sequence lengths ($\pm$ SD) of 669 bp ($\pm$ 360) and 678 bp ($\pm$ 361) and a GC content of 59 % and 58 % were calculated for *A. minutum* and *A. fundyense*, respectively.

Searching the unassembled 454 cDNA library reads with the cyanobacterial *sxtA* gene resulted in 94 hits for *A. fundyense* and 88 hits for *A. minutum*, respectively. The same search on the assembled datasets returned 10 contigs from the *A. fundyense* and 9

from the *A. minutum* library. After pooling of all sequences and re-assembly, two contigs showed a high similarity to *sxtA* from cyanobacteria: one to the domain *sxtA1* (contig length = 1450 bp, GC = 60.1 %, bit score = 213, e-value = $5e^{-61}$) and the other to *sxtA4* (contig length = 1059 bp, GC = 65 %, bit score = 195, e-value = $1e^{-47}$). Both contigs
5 contained sequences from both *Alexandrium* libraries, but neither contained a full ORF, a dinoflagellate spliced leader sequence or a polyA-tail. The two contigs were used to design *sxtA1* and *sxtA4* primers for genomic amplification, RACE analyses and sequencing.

The results of the *in silico* search for the remaining core *sxt* genes are summarized
10 in Table 3.

Table 3. Blast analyses of the core *sxt* genes from *C. raciborskii* T3 against the assembled *A. fundyense* and *A. minutum* 454 libraries; given are: the number of contigs with an E-value ≤ 0.1 present in each library; the top blastX hit, its accession number, taxonomy, score and E-value when the top contig is blasted against the non-redundant protein database of NCBI, as well as the closest hit to *sxt* genes from cyanobacteria from the same analysis.

454 library	Number of contigs	Top score/ E-value	Uppermost blastX hit of top contig against NCBI nr-database	Accession	Taxonomy	Uppermost blastX score/ E-value	Top <i>sxt</i> hit score/ E-value
<i>sxtA</i>							
<i>A. fundyense</i>	10	105/ 2e-51	polyketide synthase [<i>Mycrococcus xanthus</i> DK 1622]	YP_63211	Bacteria; Proteobacteria	183/ 5e-44	182/ 7e-44
<i>A. minutum</i>	9	108/ 3e-61	SxtA [<i>Lyngbya wollei</i>]	ACG63826	Bacteria; Cyanobacteria	236/ 2e-65	236/ 2e-65
<i>sxtB</i>							
<i>A. fundyense</i>	1	46/ 7e-11	cytidine deaminase [<i>Plectocystis pacifica</i> SIR-1]	ZP_01910517	Bacteria; Proteobacteria	91/ 9e-27	67/ 1e-11
<i>A. minutum</i>	1	35/ 0.094	none				
<i>sxtF/sxtM</i>							
<i>A. fundyense</i>	4	51/ 4e-06	putative efflux protein, MATE [<i>Polysphondylium pallidum</i> PN500]	EFA81712	Eukaryota; Amoebozoa	136/ 2e-30	62/ 5e-08
<i>A. minutum</i>	1	34/ 0.01	putative efflux protein, MATE [<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>]	XP_002873960	Eukaryota; Viridiplantae	78/ 8e-23	none
<i>sxtG</i>							
<i>A. fundyense</i>	9	57/ 2e-27	glycine amidinotransferase [<i>Amycolatopsis mediterranei</i> U32]	YP_003768377	Bacteria; Actinobacteria	163/ 3e-38	140/ 2e-31
<i>A. minutum</i>	7	55/ 2e-25	glycine amidinotransferase [<i>Amycolatopsis mediterranei</i> U32]	YP_003768377	Bacteria; Actinobacteria	143/ 2e-32	117/ 1e-24
<i>sxtH/sxtI</i>							
<i>A. fundyense</i>	7	43/ 2e-12	Rieske (2Fe-2S) region [<i>Anabaena variabilis</i> ATCC 29413]	YP_321575	Bacteria; Cyanobacteria	197/ 6e-86	80/ 1e-12
<i>A. minutum</i>	6	41/ 5e-06	Rieske (2Fe-2S) region [<i>Anabaena variabilis</i> ATCC 29413]	YP_321575	Bacteria; Cyanobacteria	119/ 5e-38	60/ 2e-07
<i>sxtI</i>							
<i>A. fundyense</i>	3	68/ 1e-13	Carbamoyltransferase [<i>Nocardia dactyloides</i> DSM	YP_003679504	Bacteria; Actinobacteria	131/ 9e-29	89/ 9e-16

<i>A. minutum</i>	1	67/ 1e-13	431111 carbamoyl transferase [<i>Streptomyces griseoflavus</i> Tu4000]	ZP_05536710	Bacteria; Actinobacteria	132/ 6e-29	91/ 1e-16
<i>xxIR</i>							
<i>A. fundyense</i>	3	36/ 0.063	atp-citrate synthase [<i>Ectocarpus siliculosus</i>]	CBJ30109	Eukaryota; stramenopiles	349/ 8e-96	none
<i>A. minutum</i>	1	38/ 0.015	atp-citrate synthase [<i>Ectocarpus siliculosus</i>]	CBJ30109	Eukaryota; stramenopiles	516/ 1e-144	none
<i>xxIS</i>							
<i>A. minutum</i>	1	36/ 0.05	hypothetical protein [<i>Perkinsus marinus</i> ATCC 50983]	XP_002767298	Eukaryota; Alveolata	91/ 4e-34	none
<i>xxIU</i>							
<i>A. fundyense</i>	33	83/ 2e-16	predicted protein [<i>Chlamydomonas reinhardtii</i>]	XP_001689640	Eukaryota; Viridiplantae	214/ 4e-54	107/ 8e-22
<i>A. minutum</i>	27	84/ 2e-16	hypothetical protein [<i>Schizophyllum commune</i> H4-8]	XP_003034688	Eukaryota; Fungi	116/ 1e-24	797/ 2e-13

Apart from *sxtA*, contigs with a good alignment score (bit score > 55) and a highly significant e-value ($< e^{-20}$) were recovered for the amidinotransferase gene *sxtG* in both libraries. Re-blasting the contigs with the lowest e-values against the NCBI nr protein database showed that the most similar gene was an actinobacterial glycine aminotransferase, while the similarity to *sxtG* from cyanobacteria was less but still highly significant (Table 3). For the core biosynthesis genes *sxtB*, *sxtF/M*, *sxtH/T*, *sxtI*, *sxtR* and *sxtU*, contigs with an e-value ≤ 0.1 were recovered from both *Alexandrium* libraries, while *sxtS* only had a hit in the *A. minutum* library (Table 3). No matches were recovered for *sxtC*, *sxtD* and *sxtE* in either of the libraries. *SxtC* and *sxtE* are unknown proteins and *sxtD* is a sterol desaturase-like protein. It is possible that dinoflagellate proteins with no similarity to the cyanobacterial genes carry out their function. Alternatively, these genes were not present in the dataset generated. While the dataset is comprehensive, it is not complete. For example, some regions of the *sxtA* transcripts were also not recovered in the 454 dataset, but only obtained through RACE analyses (see above). Re-blasting against NCBI nr protein database retrieved hits to proteins for *sxtB* (*A. fundyense* only), *sxtF/M*, *sxtH/T*, *sxtI*, and *sxtU* that are similar to those encoded in the corresponding cyanobacterial *sxt* genes. The actual sequence similarity was less conserved and no significant hits between the *Alexandrium* contigs and the cyanobacterial *sxt* genes were observed.

- Transcript structure of *sxtA* in *A. fundyense*

The RACE experiments resulted in two different *sxtA* - like transcript families. Both had dinoflagellate spliced-leader sequences at the 5' end and polyA-tails at the 3' end, but they differed in sequence, length, and in the number of *sxt* domains they encode. The shorter transcripts encode the domains *sxtA1*, *sxtA2* and *sxtA3*, while the longer transcripts encodes all four *sxtA* domains, which are also encoded by the cyanobacterial *sxtA* gene (Fig. 1).

The consensus sequence of the shorter transcripts was 3136 bp excluding polyA-tail. Eight clones with SL-leader were sequenced, and three different 5'UTRs were uncovered. The sequences were almost identical; however, one clone had a 15 bp and another had a 19 bp insert exactly following the SL-sequence. The two sequence inserts were, apart from the length, identical. The nine 3'UTR that were sequenced were almost identical and the polyA-tail started at the same position in each clone. The domain structure of this shorter *sxtA* transcript was as follows: Amino acid residues 1-27 encode a signal peptide. Residues 28-531 correspond to *sxtA1*, which contains three conserved motifs (I: VDTGCGDGSL (SEQ ID NO: 214), II: VDASRTLHVR (SEQ ID NO: 215), III: LEVSFGLCVL (SEQ ID NO: 216)). Residues 535-729 correspond to *sxtA2* with the catalytic domains 557-W, 648-T, 663-H, 711-R; while *sxtA3*, the final domain of the short

transcript, corresponds to residues 750-822 with the phosphopantetheinyl attachment site 783-DSL-785.

The consensus sequence of the longer *sxtA* transcript was 4613 bp (majority rule, longest 3'UTR, without polyA-tail, Fig. 1). Five clones with SL-sequences were characterized. One of those had a slightly divergent SL-sequence with an A at position 15 instead of the usual G. All 5'UTRs were 97 bp long (excluding SL sequence) and almost identical in sequence. Each of the four 3' clones sequenced had a different length (342, 407, 446 and 492 bp). The domain structure of the longer *sxtA* transcript was as follows: Amino acid residues 1-25 encode a signal peptide. Residues 26-530 correspond to domain *sxtA1* with the three conserved motifs: I: VVDTGCGDG (SEQ ID NO: 217), II: VDPSRSLHV (SEQ ID NO: 218) and III: LQGSFGLCML (SEQ ID NO: 219); residues 535-724 correspond to domain *sxtA2*, with the catalytic residues 556-W, 661-T, 693-H, 708-R; *sxtA3* corresponds to the residues 763-539 where 799-DSL-801 is the phosphopantetheinyl attachment site; finally, domain *sxtA4* corresponds to residues 894-1272.

The GC content of the two *Alexandrium sxtA* transcripts was consistently higher than the cyanobacteria *sxtA* genes (Fig. 2). The GC contents were 69 % (long transcript), 62 % (short transcript) and 43 % (all cyanobacteria *sxtA* genes).

All algorithms predicted the presence of signal peptides (SP) and corresponding cleavage sites for both transcripts. However, transmembrane helices that may indicate class I transit peptides in dinoflagellates were not predicted. Neither of the transcripts matched the criteria for class II and class III transit peptides.

The Genbank accession numbers are JF343238 for the short and JF343239 for the long *sxtA* transcripts (majority rule consensus sequences), and JF343357 - JF343432 for the remaining cloned RACE sequences of *A. fundyense* CCMP 1719.

- Phylogeny of dinoflagellate *sxtA1* and *sxtA4* sequences

The *sxtA1* and *sxtA4* primers designed in this study (Table 2) amplified single bands of ~550 bp (*sxtA1*) and ~750 bp (*sxtA4*) length in 18 *Alexandrium* strains comprising five species and two *Gymnodinium catenatum* strains, which had a range of toxicities (Table 1). No *sxtA1* or *sxtA4* PCR products were amplified for five non-STX-producing *Alexandrium affine* and *Alexandrium andersonii* strains, nor for non-STX-producing dinoflagellate strains of the genera *Gambierdicus*, *Ostreopsis*, *Prorocentrum*, *Amphidinium* (Table 1). These PCR-based results are generally in agreement with the toxin measurements. However, *sxtA1* and *sxtA4* fragments were amplified from the genomic DNA of four *A. tamarense* strains (ATCJ33, ATEB01, CCMP1771, ATBB01) in which no STX were detected (Table 1).

The phylogenetic analyses of *sxtA1* (Fig. 3; Fig. 5) show that all *sxtA1* sequences formed one fully supported cluster, divided into two sub-clusters. Some clones of the same strain were identical, however, slightly different clones were observed for most strains. These different clones were distributed throughout the phylogeny, generally without species- or strain-related patterns. Only sequences from *G. catenatum* formed a tight branch within one of the sub-clusters. The closest relatives to the dinoflagellate cluster were the cyanobacterial *sxtA* genes and proteobacterial polyketide synthases (Fig. 3; Fig. 5).

All *sxtA4* sequences formed one well-supported cluster, with clones from the same strain distributed throughout (Fig. 4A; Fig. 6). The cyanobacterial *sxtA* genes and actinobacterial aminotransferases formed the closest sister clades.

The Genbank accession numbers for the genomic *sxtA1* and *sxtA4* fragments are JF343240 - JF343356.

15 - Copy number and polymorphisms of *sxtA4*

Between 100 - 240 genomic copies of *sxtA4* in *A. catenella* were found in triplicate batch cultures of ACSH02 collected at three time points with different growth rates, based on the qPCR assay (Fig. 4B).

Analysis of a 987 bp contig, which covered the *sxtA4* domain and was based on *A. fundyense* 454 reads revealed at least 20 single nucleotide polymorphisms (SNPs), 15 of which were silent. SNPs were defined as a base pair change that occurred in at least two of the reads. Homopolymer stretches and indels were ignored.

Discussion

25 - *Sxt* genes are encoded in dinoflagellate genomes

The dinoflagellate genome is unusually large [1.5 – 200 pg DNA cell⁻¹; 52] and highly divergent. Recent estimates predict that dinoflagellate genomes contain between 38,000 and almost 90,000 protein-encoding genes, which correspond to 1.5 - 4.5 the number of genes encoded in the human genome. The results of sequencing > 1.2 million ESTs in this study demonstrate that close homologues of the genes involved in STX biosynthesis in cyanobacteria are also present in STX-producing dinoflagellates (Table 3). To further confirm their dinoflagellate origins *sxtA* was investigated being the unique starting gene of the biosynthesis pathway. The transcriptome of *A. fundyense* CCMP 1719 contained two different transcript families that had the same domain architecture as *sxtA* in cyanobacteria. The two transcript families varied in length, sequence, and the number of catalytic domains they encode. The longer transcripts contained all four domains present in the known cyanobacterial *sxtA* genes, however, the shorter transcripts lacked the terminal aminotransferase domain (Fig. 1). In contrast to bacterial transcripts, both

transcript families possessed eukaryotic polyA-tails at the 3'end and dinoflagellate spliced-leader sequences at the 5'end. Hence, these results clearly show that at least *sxtA*, and possibly other *sxt* genes, are encoded in the nuclear genome of dinoflagellates and that STX-synthesis in dinoflagellates does not originate from co-cultured bacteria. These bacteria may still, however, play an important role in modulating STX biosynthesis in dinoflagellates.

The signal peptides identified in both transcripts indicate a specific targeting of both Sxt products. Many genes in the nuclear genomes of dinoflagellates are plastid-derived and their products targeted to the plastid. These proteins are translated in the cytosol and then transported to the plastid through the plastid membranes. In peridinin-containing dinoflagellates like *Alexandrium*, this process requires the presence of signal and transfer peptide motifs. Both *sxtA* transcripts are predicted to contain signal peptides, but transfer-peptide structures were not identified. Thus, it seems that both *sxtA* proteins are targeted out of the cytosol, but the region of target need to be experimentally investigated.

The dinoflagellate *sxtA* transcripts did not only differ from the cyanobacterial counterparts by the presence of signal peptides, SL sequences and polyA-tails, but also in their GC content. The *A. fundyense* ESTs had a considerably higher GC content (Fig. 2). Transcribed genes from *Alexandrium* species have been reported to have an average GC content > 56%, while filamentous cyanobacteria, such as the STX-producing genera *Cylindrospermopsis*, *Anabaena*, *Aphanizomenon* and *Lyngbya*, have a genomic GC content around 40 %. This indicates that the GC content of *sxtA* has diverged significantly from the progenitor *sxtA* possessing ancestor, in line with the rest of the genome in these microorganisms.

The involvement of the two different *sxtA* transcripts and their role in STX-synthesis is presently unclear, but the differences in GC content (Fig. 2) indicate that they are under different selection pressures.

- The non-identical copies of *sxtA*: variation at the genome and transcriptome level

One typical feature of dinoflagellate genomes is that genes may occur in multiple copies, which may or may not be identical. This is possibly related to highly unusual genetic mechanisms such as the recycling of processed cDNAs. It appears that *sxtA* also occurs in multiple copies within dinoflagellate genomes. It was estimated that 100 - 240 copies of the *sxtA4* domain were present in the genomic DNA of *A. catenella* ACSH02 (temperate Asian ribotype). The copy number differences detected throughout the cell cycle are likely related to the growth rate of the batch culture and the proportion of cells in various cell cycle phases. All genomic *sxtA4* sequences from 15 different *Alexandrium* and one *G. catenatum* strains formed one well-supported phylogenetic cluster, with

several slightly different clone sequences of the same strain distributed throughout the tree. *SxtA1* was also found to occur in multiple, non-identical copies in all strains analysed (Fig. 5). Further, the separation of the dinoflagellate *sxtA1* cluster into two sub-clades indicates that *sxtA1* may be encoded by two separate gene classes, at least in some strains.

The genomic variation of *sxtA* is also present in the *Alexandrium* transcriptomes. Adding the transcriptome data to the *sxtA1* tree showed that the upper clade corresponds to the longer *sxtA* transcripts, whereas the lower clade corresponds to the shorter transcripts (Fig. 3, Fig. 5). Analyses at the nucleotide level of the *sxtA4* region in the transcriptome of *A. fundyense* revealed many of SNP sites, two-thirds of which were silent.

- Correlation between *sxtA1*, *sxtA4* and saxitoxin production

The *sxtA1* and *sxtA4* genomic sequences identified during this study were present in all STX-producing dinoflagellate strains analysed, including two *G. catenatum* and 14 *Alexandrium* strains of the species *A. catenella*, *A. minutum*, *A. fundyense* and *A. tamarensis*. Neither of the two *sxt* fragments were amplified from two *A. andersoni* and three *A. affine* strains. Homologs were also not detected in *Gambierdiscus australes*, *Amphidinium massartii*, *Prorocentrum lima*, *Ostreopsis siamensis* and *Ostreopsis ovata*, none of which are known to produce STX (Table 1).

Despite the very good correlation between the presence of *sxtA1* and *sxtA4* and STX content for most of the strains analysed, both fragments were also amplified from *A. tamarensis* strains for which no STX-production was detected (Table 1). RACE analyses of *A. tamarensis* strain CCMP1771 revealed that *sxtA1* and *sxtA4* were transcribed in this strain (data not shown). It is postulated that the amount of STX produced by *A. tamarensis* is lower than the detection limit of the HPLC/MS toxin determination methods used, since a very sensitive saxiphilin assay used to investigate *A. tamarensis* strain ATBB01 found it to be toxic. Transcript abundance has been suggested to be positively related to the number of gene copies present in a dinoflagellate genome. Hence, it is possible that strains with low levels of STX have fewer copies of the *sxt* genes compared to those with greater STX-production. If this holds true, then the presence of *sxtA1* and *sxtA4* would indicate toxicity and molecular methods could be developed to detect STX-producing cells in the environment.

- Evolution of STX-synthesis in eukaryotes and its role in the diversification of *Alexandrium*

The cyanobacterial *sxt* genes are highly conserved between cyanobacteria species and the gene cluster is thought to have arisen at least 2100 million years ago. The results

herein show that dinoflagellate *sxtA* transcripts that are phylogenetically closely related to a clade of the cyanobacteria *sxtA* sequences and other bacterial putative toxin-related genes (Fig. 3 & Fig. 4) also have the same domain structure as cyanobacterial *sxtA* genes (Fig. 1). It is proposed that this striking similarity is most likely due to a horizontal gene transfer (HGT) event between ancestral STX-producing bacteria and dinoflagellates. Within dinoflagellates, STX are produced by species of the genera *Alexandrium* and *Pyrodinium*, which belong to the family Gonyaulacaceae within the order Gonyaulales, as well as by one species of the genus *Gymnodinium*, which belongs to the family Gymnodiniaceae in the order Gymnodiniales. Thus, these toxins are produced by two genera within one family and by a single species from a distant dinoflagellate order. This distribution of STX-synthesis within the dinoflagellates as well as the close relationship between *Alexandrium* and *Gymnodinium catenatum* *sxtA* sequences (Fig. 3, Fig. 4, Fig. 5, and Fig. 6), suggests that the bacteria-to-dinoflagellate HGT likely took place prior to the origin of the genera *Alexandrium* and *Pyrodinium*, and was followed by a dinoflagellate-to-dinoflagellate transfer into *G. catenatum*. The extent of eukaryote-to-eukaryote HGTs is often underestimated due to difficulties in detecting such events, however, recent work highlights the importance and prevalence of such gene transfers.

The relationship among the dinoflagellate *sxtA* sequences in this study was not resolved in this study, as most the internal nodes were not statistically supported (Fig. 5 and Fig. 6). Therefore, it was not possible to determine with certainty whether the evolution of the *sxtA* genes mirrors that of the genus *Alexandrium*, or to determine the origins of a putative HGT from *Alexandrium* into *G. catenatum*. However, the *sxtA1* and *sxtA4* gene copies from multiple strains of *G. catenatum*, *A. minutum*, and *A. catenella* tended to be clustered by species indicating that their history reflects the evolution of these species. The non-amplification of *sxtA1* and *sxtA4* from the non-STX-producing species *A. affine* and *A. andersoni* may indicate that the *sxtA* genes have either been lost from these lineages or have mutated so much, that the primers developed here were not able to amplify them.

The two *Alexandrium* EST datasets contained transcripts, which encoded homologs to the majority of core *sxt* genes identified from cyanobacteria (Table 3). Even though the similarity to the cyanobacterial *sxt* genes was often significant, it was much less than observed for *sxtA*. The closest hits were to other bacterial or eukaryotic genes present in the database. This indicates that different genes in the *sxt* pathway may have separate origins in dinoflagellates. Further work is required to elucidate the complex origins of this gene cluster and will lead to further advances regarding the genomes and molecular biology of these ancient and important microorganisms.

Example 2: Quantitative method for detecting and quantifying STX production in dinoflagellates

Materials and Methods

5 - Culture maintenance

Dinoflagellate cultures (Table 4) were maintained in GSe (Doblin et al., 1999, *supra*) or L1 media (Guillard & Hargraves, 1993 *supra*) at 16-20°C. Light was provided by white fluorescent bulbs (Crompton Light), with photon flux of 60-100 $\mu\text{mol photon m}^{-2} \text{ sec}^{-1}$ on 12/12 hour dark/light cycle. Strains used were provided by the University of
 10 Tasmania (isolated by M. de Salas) the Australian National Culture Collection of Marine Microalgae, Provasoli-Guillard Culture Collection (CCMP) and the Cawthron Institute Culture Collection.

15 **Table 4.** Dinoflagellate strains tested, STXs content and whether the *sxtA4* qPCR primer pair resulted in a product. All samples were tested with a positive control to ensure PCR inhibitors were not present.

Dinoflagellate	Strain number	STXs detected ¹	<i>sxtA4</i> qPCR product
<i>Alexandrium</i>			
<i>affine</i>	CCMP112	-	-
<i>affine</i>	CS-312/02	-	-
<i>andersonii</i>	CCMP1597	-	-
<i>andersonii</i>	CCMP2222	-	-
<i>catenella</i>	ACCC01	+	+
<i>catenella</i>	ACSH02	+	+
<i>catenella</i>	ACTRA02	+	+
<i>fundyense</i>	CCMP1719	+	+
<i>minutum</i>	CCMP113	+	+
<i>minutum</i>	CS-324	+	+
<i>tamarense</i>	ATCJ33	-	+
<i>tamarense</i>	ATNWB01	+	+
<i>Gambierdiscus</i>			
<i>australes</i>	CAWD148	-	-
<i>Ostreopsis</i>			
<i>ovata</i>	CAWD174	-	-
<i>siamensis</i>	CAWD	-	-
<i>Amphidinium</i>			
<i>massarti</i>	CS-259	-	-
<i>Gymnodinium</i>			
<i>catenatum</i>	GCTRA01	+	+
Environmental water sample containing:		n/a	-
<i>Protoceratium reticulatum</i>			
<i>Prorocentrum micans</i> , <i>Karenia sp</i>			

Karlodinium veneficum
Polarella glacialis
Symbiodinium sp

- DNA extraction and PCR

Culture density was determined regularly using a Sedgewick Rafter cell (Proscitech) and an inverted light microscope (Leica Microsystems). Known numbers of cultured cells were harvested during exponential growth phase. DNA was extracted from the cell pellets using the CTAB method (see Doyle and Doyle, 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:1-5), with an additional overnight DNA precipitation at -20°C. Quality and quantity of DNA was determined using a Nanodrop (Thermoscientific), and by amplifying a control dinoflagellate gene (*cytb* or SSU rRNA), according to the protocols of Lin et al. (2009), using the primer pair 4f and 6r, which amplify a 440 bp fragment, or 18S r DNA primers 18SF08 (5'-TTGATCCTGCCAGTAGTCATATGCTTG-3'(SEQ ID NO: 220)) and R0ITS (5'-CCTTGTTACGACTTCTCCTTCCTC-3'(SEQ ID NO: 221)) that amplify ~1780 bp.

15

- *sxtA4* qPCR assay development and copy number determination

An alignment of *sxtA4* genomic and sequences from 9 strains of the species *Alexandrium catenella*, *A. tamarense*, *A. minutum*, *A. fundyense* and *Gymnodinium catenatum* (GenBank accession numbers JF343238 – JF343239, JF343259-JF343265) was constructed. The degree of conservation of the gene sequences was checked for a 440 bp fraction of the *sxtA4* domain and found to be 94-98% between *Alexandrium* species, and 89% between *Gymnodinium catenatum* and *Alexandrium* species. Primers specific for *sxtA4* were designed using Primer3 software and a consensus sequence. The specificity of the primer sequences was then confirmed using BLAST (Basic Local Alignment Search Tool) on NCBI (National Centre for Biotechnology Information). The sequences of the primers were *sxtA4F* 5' CTGAGCAAGGCGTTCAATTC 3' (SEQ ID NO: 198) and *sxtA4R* 5' TACAGATMGGCCCTGTGARC 3' (SEQ ID NO: 199), resulting in an 125 bp product.

25

To determine their specificity to STX-producing, or potentially STX-producing strains, the *sxtA4* primer pair was amplified from 6 species of *Alexandrium*, *Gymnodinium catenatum*, 3 other toxin producing species of Gonyaulacales: *Ostreopsis ovata*, *Gambierdiscus australes*, *Ostreopsis siamensis*, an additional dinoflagellate, *Amphidinium massartii*, and an environmental sample containing a mixed phytoplankton community, including 6 identified dinoflagellate species (Table 4). PCR amplification was performed in 20 µl reactions containing template, 0.5 µM of each primer, 3 mM MgCl₂, 1 µl BSA (NEB), and 10 µL Immomix (Bioline), containing dNTPs, Immolase

35

Taq polymerase and reaction buffer or 20 µl containing template, 0.2 µM of each primer, 3 mM MgCl₂, 1 µl BSA, 2 µl MyTaq reaction buffer (Bioline) containing dNTPs, 0.2 µl MyTaq (Bioline) hot start polymerase and H₂O. Hot start PCRs were performed with an initial denaturing step of 95°C for 5-10 min, and 35 cycles of 30 s at 95°C, 30 s at 55 or 60°C (for the *cytb* and *sxtA4* primers, respectively), 30s at 72°C followed by a final extension step of 7 min at 72°C. The 18S fragment was amplified in 25 µL reactions containing template, 1 unit 10X BD Advantage 2 PCR buffer (BD Biosciences), 5 mM dNTPs, 0.2 µM of each primer, DMSO (10 % final concentration) and 0.25 units 50X BD Advantage 2 Polymerase Mix (BD Biosciences). PCRs were amplified as follows: 94 °C - 1 min; 30 x (94 °C - 30 s; 57 °C - 30 s; 68 °C - 120 s); 68 °C - 10 min; 8 °C - hold. Products were analysed using 3% agarose gel electrophoresis, stained with ethidium bromide and visualized.

qPCR was also performed using a primer pair specific for the temperate Asian ribotype of *Alexandrium catenella*, found in Australian temperate waters, based on a region of the large subunit (LSU) ribosomal RNA region (Hosoi Tanabe and Sako, 2005), amplifying an 160 bp fragment, catF (5'-CCTCAGTGAGATTGTAGTGC-3' (SEQ ID NO: 222)) and catR (5'-GTGCAAAGGTAATCAAATGTCC-3' (SEQ ID NO: 223)). Assays were performed on environmental samples, and new standard curves of this LSU rRNA primer pair were constructed based on strains isolated from Australian waters by M. de Salas (UTAS): ACCC01, isolated from Cowan Creek, NSW, approximately 20 km from the Brisbane Water site and 50 km from Georges River site, ACSH02, isolated from Sydney Harbour, approximately 35 km north of the Georges River site, and ACTRA02, isolated from Tasmania, Australia.

qPCR cycling was carried out on a Rotor Gene 3000 (Corbett Life Science) using SSOFast Evagreen supermix (Biorad). qPCR assays were performed in a final volume of 20 µl consisting of 10 µl Evagreen master mix (containing DNA intercalating dye, buffer and *Taq* polymerase), 1 µl of template DNA, 0.5 µM of each primer, and 1 µl of BSA. qPCR assays were performed in triplicate with the following cycles: 95°C for 10 s, and 35 replicates of 95°C for 15 s and 60°C for 30 s. Melting curve analysis was performed at the end of each cycle to confirm amplification specificity, and selected PCR products were sequenced.

Standard curves for both *sxtA4* and LSU rRNA were constructed in two ways: (1) Using a dilution series of a known concentration of fresh PCR product, ranging from 5.7–5.7 x 10⁻⁵ ng (n=6). Standard curves using PCR product were used to determine the efficiency of the assay (see method in Pfaffl, (2001) A new mathematical model for relative quantification in real-time RT-PCR Nucleic Acids Res. 2001 May 1;29(9):e45), as well as to determine copy number. The molecules of PCR product were determined: (A x 6.022 x 10²³) x (660 x B)⁻¹ with A: concentration of PCR product, 6.022 x 10²³:

Avogadro's number, 660: average molecular weight per base pair and B: length of PCR product. The number of molecules in the unknown samples were determined and divided by the known number of cells in the DNA qPCR template, to obtain copy number per cell. (2) Extracting DNA from duplicate samples of known numbers of cells of strains of *Alexandrium catenella* (ACCC01, ACSH02, ACTRA02) taken during exponential growth phase, and diluting the DNA at 50% over 3 orders of magnitude (n=6).

To estimate the environmental abundance of *A. catenella* in the samples based on the LSU rRNA assay, the equations from (2) were extrapolated and applied to the CT values measured for these samples. Because variability has been found in copy numbers of the rRNA genes among strains of some *Alexandrium* species, as well as a variability of up to a factor of 2 expected due to variability in growth and cell cycle conditions of cells, the copy number of the LSU rRNA gene region in duplicate samples of each of the 3 strains was determined. The final *A. catenella* abundances in the environmental samples were determined as the mean and standard deviation of 6 independent estimates.

- Phytoplankton and oyster sample collection

The phytoplankton community was sampled daily at mid-tide during the 15–20th November, 2010, at standard monitoring sites close to Sydney rock oyster (*Saccostrea glomerata*) farms in Wagonga Inlet, Narooma, NSW, -36° 13' E 150° 6' S and the Georges River, NSW -34° 1' E 151° 8' S (Fig. 7). Samples were also taken at Brisbane Water, NSW -33° 28' E 151° 18' S on 22nd July, 2010 (Fig. 7).

Triplicate 4 L bottle samples were taken each day for molecular analysis. A further 500 ml bottle was taken and immediately fixed with Lugol's iodide for microscopic identification and counting. Samples were filtered using 3 µm Millipore filters and frozen at -20°C until DNA extraction. Ten individual *S. glomerata* samples were taken from farms in the immediate vicinity of the phytoplankton sampling site on the 17/11/10 and the 19/11/10. *S. glomerata* samples were pooled for toxin testing.

To determine the specificity of the primer pair in mixed environmental samples, a phytoplankton community in which no known STX-producing species were present was sampled. 1 L of the surface community at Jervis Bay was sampled on 19th January, 2011, and preserved, concentrated, identified and counted species present from 500 ml, as above. Dinoflagellates present were identified as *Protoceratium reticulatum*, *Karlodinium cf. veneficum*, *Karenia* sp., *Prorocentrum micans*, *Polarella glacialis*, *Pfiesteria shumwayae*, and *Symbiodinium* sp. 500 ml of the remaining sample was filtered and performed DNA extraction and PCR as described above.

- *Alexandrium* cell counts using microscopy

Phytoplankton cells in ~300 ml of the Lugol's preserved samples were concentrated by gravity assisted membrane filtration on to 5 µm cellulose ester filters (Advantec) prior

to washing into 4 ml and counting. *Alexandrium* species were identified and counted using a Sedgewick Rafter cell and a Zeiss Axiolab microscope equipped with phase-contrast optics. The number of cells counted varied among samples, depending on *Alexandrium* abundance, and standard error rates were calculated using the equation:
5 Error = $2 / \sqrt{n}$, where n is the number of cells observed in the sample.

- Toxin determination in oysters and cultures

Shellfish samples and dinoflagellate cell pellets were tested using HPLC, according to the AOAC Official Method 2005.06 for paralytic shellfish poisoning toxins in shellfish
10 at the Cawthron Institute, New Zealand. A matrix modifier as described in the original protocol was not used, instead used average spike recoveries for each separate compound were used. HPLC analysis was performed on a Waters Acquity UPLC system (Waters) coupled to a Waters Acquity FLR detector. Separation was achieved with a Waters Acquity C18 BEH 1.7 μm 2.1 x 50 mm column at 30°C, eluted at 0.2 mL min⁻¹. Mobile
15 phases were 0.1 M ammonium formate (A) and 0.1 M ammonium formate in 5% acetonitrile (B), both adjusted to pH 6. The gradient consisted of 100% A for 0.5 min, a linear gradient to 80% B over 3.5 min, then returning to initial conditions over 0.1 min and held for 1.9 min. The fluorescence detector had excitation set to 340 nm and emission to 395 nm. Analytical standards for the STX analogs were obtained from the National
20 Research Council, Canada. The detection limit of the HPLC of the cell cultures was considered to be 0.1 pg cell⁻¹ for NEO and STX, 0.2 pg cell⁻¹ for GTX1/4, GTX6 (B2) and GTX5 (B1), 0.5 pg cell⁻¹ for C1,2, and <0.3 pg cell⁻¹ for the analogs C3,4.

Results

25 - Specificity, sensitivity and efficiency of the primer pair

The primers designed in this study were found to amplify a fragment of the correct size in all tested STX-producing dinoflagellates of the species: *Alexandrium minutum*, *A. catenella*, *A. fundyense*, *A. tamarense* and *Gymnodinium catenatum* (Table 4). In addition, it amplified a fragment of the correct size from the species *A. tamarense*, strain ATCJ33,
30 Tasmanian ribotype, which was not found to produce STXs at a level above the detection limit of the HPLC method utilised. Sequencing of the products confirmed this to be a homolog of *sxtA4*.

The *sxtA4* primer pair did not amplify DNA from the non-STX producing related Gonyaulacalean species *Alexandrium andersonii*, *Alexandrium affine*, *Gambierdiscus*
35 *australes*, *Ostreopsis ovata* or *Ostreopsis siamensis* nor from the more distantly related dinoflagellate species *Amphidinium massartii*. In addition, the *sxtA4* primer pair did not amplify DNA from the phytoplankton samples, which contained a mixed planktonic community including bacteria, diatoms, picoplankton and the dinoflagellates

Protoceratium reticulatum, *Karlodinium cf veneficum*, *Karenia* sp., *Prorocentrum micans*, *Polarella glacialis*, *Pfiesteria shumwayae*, and *Symbiodinium* sp. (Table 4). In contrast, DNA from all samples was amplified using the positive control primer pair to ensure the reaction template was intact and free of inhibitors.

The efficiency of the *sxtA4* assay based on this primer pair was 97% as calculated using a dilution series of fresh PCR product over 6 orders of magnitude. The assay was 93–107% efficient as calculated using a duplicate 50% dilution series of gDNA from the three strains of *A. catenella* (Fig. 8). For standard curves based on both PCR product and based on gDNA, r^2 values of the regression equations were 0.95 or greater (Fig. 8). The assay was sensitive to DNA quantities representing ~30 to >2000 cells of the three strains of *A. catenella*. Therefore, if collection of samples was carried out following a similar protocol to that utilised, and 4 L of seawater was collected, extracted and eluted in 15 μ L, of which 1 μ L was assayed, then the assay would detect environmental concentrations of *A. catenella* with a lower limit of approximately 110 cells L⁻¹.

- Copy number of *sxtA4* genes

The copy number of *sxtA4* in the 3 cultured strains of *Alexandrium catenella* had a mean of 178–280 cell⁻¹ (Table 5). Toxicity of these strains was 3.1–6.6 pg STX equivalent cell⁻¹ (Table 5). In the environmental samples, the copy number of *sxtA4* was estimated to be 226 and 376 cell⁻¹ in the Georges River and Wagonga Inlet samples, respectively, and most variable amongst the estimates based on the Wagonga Inlet samples.

Table 5. STXs present in *Alexandrium catenella* strains, in pg cell⁻¹ and in *Saccrostrea glomerata* from the sampling sites, in μ g STX equivalents kg⁻¹ of shellfish, and mean copy number of *sxtA4* genes in the strain or in all phytoplankton samples from that sampling site. A reading of 0 indicates levels were below the detection limit of the test. The *S. glomerata* samples were taken on 17/11/10, 19/11/10 and 22/7/10 for the Georges River, Wagonga Inlet and Brisbane Water, respectively.

	Total STXs	GTX-1,4	GTX-6	Cl,2	GTX-5 (B1)	NEO/STX	C-3,4	B2	<i>sxtA4</i> cell ⁻¹ +/- sd in strain or in plankton sample
cultures									
ACSH02	5.25	1.75	0.60	2.40	0.5	<0.1	<0.3	0	178 +/- 49 (n=9)
ACCC01	6.60	1.15	0	2.55	0	0	1.00	1.9	240 +/- 97 (n=3)
ACTRA02	3.13	1.13	0	2.00	0	0	0	0	280 +/- 85 (n=3)
<i>S. glomerata</i>									
Georges River	200	160	trace	30	10	0	trace	0	226 +/- 97

Wagonga Inlet	48	32	0	16	0	0	0	0	(n=15) 376 +/- 257 (n=12)
Brisbane Water	145	53	92		0	0	0	0	275 (n=1)

Environmental samples

sxtA4 was detected in the single Brisbane Water sample, as well as in the Georges River and Wagonga Inlet sample sets (Fig. 9). Sequencing and melt-curve analysis confirmed this to be *sxtA4*, with an average identity of 99% or higher to the corresponding gene from the *Alexandrium catenella* strain. A positive relationship between cell number, as estimated from microscopy, cell number as estimated from LSU rDNA, and the *sxtA4* copy number was observed in both sets of environmental samples. The correlation between cell number as estimated from LSU rRNA gene qPCR and the estimated *sxtA4* gene copy number was very high for the Georges River samples ($r^2=0.97$, slope=0.0059, $p<0.001$), and lower for the Wagonga Inlet sample ($r^2=0.70$, slope= 0.001, $p<0.07$) (Fig. 9).

STXs in pooled *S. glomerata* samples were detected from each of these three sites, with the highest concentrations reported for the Georges River site (200 µg STX equivalent kg⁻¹ of shellfish) with lower levels recorded for both Wagonga Inlet and Brisbane Water (48 and 145 µg STX equivalent kg⁻¹ of shellfish, respectively, Table 5).

Discussion

Provided herein is a new method for detecting and quantifying the potential for STX production in marine environmental samples. The assay is based on the detection of the gene *sxtA* that encodes a unique enzyme putatively involved in the *sxt* pathway. The method described detected *sxtA* gene in all STX-producing cultures, and did not detect it in the non STX-producing cultures or the environmental sample that did not contain known STX-producing species. However, *sxtA* genes were also detected in the non-producing strain of *Alexandrium tamarense*, Tasmanian ribotype, ATCJ33. As a very closely related strain of the Tasmanian ribotype of this species has been found to produce STXs (unpublished data), it is possible that the strain ATCJ33 has the potential to produce STXs under certain circumstances. The amplification of *sxtA4* from the *Alexandrium* and *Gymnodinium catenatum* species and strains in this study are in line with findings in Example one above, in which approximately 550 bp and 750 bp fragments of *sxtA1* and *sxtA4* were amplified from the same strains tested, and no amplification of these fragments from the species *Alexandrium affine* and *A. andersonii* was detected.

- Copy number of *sxtA4* and STXs content

The abundance of *sxtA4* was found to be relatively similar among the strains and environmental samples tested, with a range of 178-376 copies cell⁻¹ (Table 5). This

supports results in Example one above in which 100-240 copies cell⁻¹ were found throughout the growth of *Alexandrium catenella* strain ACSH02 using qPCR. The total STX equivalent toxicity of the three strains of *Alexandrium catenella* was 3.1–6.6 pg STX equivalents cell⁻¹ (Table 5). This is within the of other STX-producing species, such as strains of *A. minutum*, *A. catenella* and *A. tamarense* (0.66–9.8 pg STX equivalents cell⁻¹), depending on nutrient supply and culture growth, but lower than the most toxic strains such as *Gymnodinium catenatum* (26–28 pg STX equivalent cell⁻¹), and *Alexandrium ostenfeldii* (up to 217 pg STX equivalent cell⁻¹)

10 - *sxtA4*, *A. catenella* and STXs in south-eastern Australia

Alexandrium catenella was sampled on three occasions in southeastern Australian estuaries throughout this study period and, in each case, *sxtA4* was detected (Table 4). For the Georges River sample set, the correlation between *sxtA4* copies L⁻¹ and cell abundance L⁻¹, as determined by LSU rDNA, was highly significant (Fig. 9). At the Georges River sampling site, mean abundances of 3150-26450 cells L⁻¹ were recorded throughout the 5 day sampling period based on the estimate of the LSU rDNA assay, and 7900-38000, based on microscope cell counts from selected days. On the final day of sampling, variability in cell counts was found amongst triplicate samples taken at the site, reflecting patchiness in the distribution of *A. catenella*. Despite this, the correlation between *sxtA4* copies L⁻¹ and *Alexandrium catenella* cell number based on rRNA qPCR was very strong ($r^2=0.97$, slope=0.0059, $p<0.001$), and total STX loads in oyster samples taken during this week were 200 µg STX equivalents kg⁻¹ of shellfish, below the regulatory level for public health monitoring (800 µg STX equivalents kg⁻¹ of shellfish) but the highest of the three samples taken during this study.

At Wagonga Inlet, mean abundances of 30-288 cells L⁻¹ were found based on the estimate of the LSU rDNA qPCR assay and 80–540 cells L⁻¹ based on microscope cell counts throughout the sampling period (Fig. 9). The correlation between *sxtA4* copies L⁻¹ and cell number, as calculated from the LSU rDNA qPCR assay, was lower than that of the Georges River samples ($r^2=0.70$, slope= 0.001, $p=0.07$). This may reflect the fact that two of these samples contained fewer than 110 cells L⁻¹, and were thus at the lower limit of reliable detection of this assay. Alternatively, the lower correlation coefficient of this sampling set may be attributed to the presence of different strains of *Alexandrium catenella* which differed in copy number of *sxtA*.

35 - Detection methods for STXs and *Alexandrium* species

Generally, the enumeration of HAB-forming phytoplankton and their toxins for industry and for biological oceanographic research relies on microscope-based counting of species and direct toxin detection methods. Quantification of STXs is generally

conducted by mouse bioassay, instrumental HPLC, LC-MS or antibody-based immunoassays, such as enzyme linked immunosorbent assays (ELISA). HPLC is a time-consuming and expensive process, requiring a well-equipped analytical laboratory and pure standards of STX and its numerous analogs. While newly developed ELISA methods have overcome some of these drawbacks, they are not available for several common STX derivatives, and have problems of cross-reactivity, as toxin profiles are often quite complex.

Molecular genetic and antibody-based methods for marine phytoplankton species identification and enumeration have many advantages when compared to microscope-based counts and direct toxin detection methods: their simplicity, with a much lower requirement for training and experience compared to microscope-based taxonomic identification, cost effectiveness (qPCR reagents generally cost less than ~US \$1 per sample), speed and potential for automation (up to 30 samples may be run in triplicate in under 2 hours on a standard qPCR machine using 96 well plates). Real time qPCR machines have substantially decreased in cost in recent years, and it is possible to operate them with a basic training in molecular biology techniques.

A reliable detection limit of ~ 110 cells L^{-1} of *Alexandrium catenella* is achievable using the qPCR method reported here. Assays based on qPCR for the detection of *Alexandrium* may be more sensitive than microscopy-based methods at low cell abundances and where the species of interest may be a minor component of the phytoplankton. The Sedgewick-Rafter counting chamber method, as it is applied in the majority phytoplankton monitoring programs, is considered to have a reliable detection limit of 1000 cells L^{-1} . However, this is dependent on the volume of sample observed. In the present study, levels of detection down to < 20 cells L^{-1} were achieved using the Sedgewick-Rafter counting chamber method, by filtering the sample such that larger volumes of sample were observed. The standard error of microscope-based counts is dependent on the number of cells observed, and increases with decreasing cell number. For molecular genetic based methods, the standard error associated with cell counts is independent of the abundance of cells, for cell abundances greater than the detection threshold of the assay. Molecular genetic identification and enumeration methods have reported detection thresholds in the order of 10-100 cells L^{-1} of *Alexandrium* species using qPCR and FISH probes, depending on the volume of water (typically 1-8 L) sampled using these methods. As concentrations as low as 200 cells L^{-1} of *Alexandrium* species have been associated with STX uptake in shellfish, the reliable detection of species at low cell abundances is an important advantage of qPCR based enumeration methods over microscope counts as currently practiced in the majority of phytoplankton monitoring programs.

While many advantages have been noted in molecular genetic based monitoring methods, current methods have several drawbacks. qPCR for species enumeration using marker genes requires the use of multiple probes in habitats where several species of *Alexandrium*, *Pyrodinium bahamense* and *Gymnodinium catenatum* occur and produce STXs. Species not previously documented in a particular habitat are occasionally identified, and they may not be noticed if a suitable probe is not available for their identification. In addition, the enumeration of certain target species requires research to culture and determine the toxicity of local STX-producing species, as this may vary between regions. High abundances of *Alexandrium catenella* in regions in which this species generally produces STXs have not always been correlated with STXs in shellfish, suggesting that population level differences in production of STXs may occur.

A final drawback of most qPCR-based counting methods is that ribosomal RNA genes, which are commonly used for detection, as their relatively fast divergence rates allow for the design of species-specific markers, can vary significantly in copy number among strains of some species of *Alexandrium*, and in some cases, during species growth. This may be due to the presence of unstable rDNA pseudogenes in some *Alexandrium* species, and possibly the presence of extra-chromosomal rDNA molecules. The effect of this variation is to cause disparities between the estimated gene number and cell number, and consequently inconsistencies between abundance estimates based on microscopy and those based on qPCR. For this reason, in the present study, the copy number of LSU rDNA genes was determined in replicates of three strains of *A. catenella* isolated from local waters, in order to obtain a reliable estimate of cell number based on LSU rDNA copy number.

The novel method presented here relies on the direct detection of a gene (*sxtA*) involved in the synthesis of STXs. Therefore, it may be more closely correlated with STX production than the abundance of any particular species. In addition, it is considerably faster and cheaper to detect *sxtA* than the actual toxins using analytical instruments. Using the disclosed primer pair, *sxtA4* was not amplified from two non-STX producing *Alexandrium* species tested but was amplified from the relatively distantly related STX-producing species *Gymnodinium catenatum*. This allows for the use of a single assay to simultaneously detect several different STX-producing genera, including potential STX producing species not previously known from a particular site. However, the assay also amplified a product from a strain of *Alexandrium tamarense* ATCJ33 that has not been found to produce STXs, mirroring findings in Example one above that this strain possessed *sxtA1* and *sxtA4*, and showing that this assay may not, in a small percentage of cases, be indicative of the presence of STXs.

Transcription level regulation may play a relatively minor role in the expression of many dinoflagellate genes, compared to regulation in other organisms, as genes that are

up-regulated have been reported to increase in transcript abundance by no more than ca. 5-fold, compared to levels during standard growth. This has led to the theory that the duplication of genomic copies of highly expressed genes in dinoflagellates may function as a means of increasing their transcription. If this were true, there may be a relationship
 5 between the copy number of *sxtA* cell⁻¹ and the quantity of STXs produced by a particular strain. In this study, the species tested had a relatively similar STXs cell quotas and no significant difference in *sxtA* cell⁻¹.

10 Example 3: Investigation of *sxtA* sequences in saxitoxin producing and non-producing dinoflagellate species

Aim

To amplify and sequence genes involved in the synthesis of saxitoxin (STX) in dinoflagellates which are known to produce saxitoxin and those which are not known to
 15 produce STX.

Materials and Methods

Dinoflagellate strains were grown in GSe media and maintained in a culture cabinet at 18 degrees, and a 12/12 light cycle. DNA was extracted from 20 ml of exponentially
 20 growing culture by harvesting by centrifuging at 3000 g for 5 minutes, then using the CTAB method.

DNA template was PCR amplified using Advantage GC rich PCR polymerase (Clontech), which contains 10% BSA, in a Thermo Cycler with the following PCR conditions: an initial 5 minute 95°C denaturing before 35 cycles of (1) 30 sec 94°C
 25 denaturing, (2) 30 sec annealing (variable temperature), and (3) 1-2 minute 72°C extension, with a final 10 minute extension at the same temperature. PCR products were gel excised using Promega Wizard SV Gel and PCR Clean-Up System (Promega), before direct sequencing with an ABI3730 DNA analyzer (Applied Biosystems) using primers as below.

Sxt001	F	TGCAGCGMTGCTACTCCTACTAC (SEQ ID NO: 200)	57.1
Sxt002	R	GGTCGTGGTCYAGGAAGGAG (SEQ ID NO: 201)	55.9

30 Species newly investigated for this study were the STX producing species: *Alexandrium tamarense*, strain CAWD121, and the non-producing species *Alexandrium* sp, strain AAKT01, *Amphidinium massartii* CS-259, *Amphidinium mootonorum* CAWD161, *Coolia monotis* CAWD98, *Gambierdiscus australes* CAWD148,

Prorocentrum lima CAWD, *Protoceratium reticulatum* CAWD, *Gonyaulax spinifera* CAWD.

Results

5 An *sxtA1* domain gene of the correct size was identified from the saxitoxin producing species *Alexandrium tamarense* strain CAWD121 (SEQ ID NO: 224), and the presumed saxitoxin producing species *Alexandrium catenella*, strain ACNC50 (SEQ ID NOs: 225-226 show *sxtA1* domain sequences from two clones of strain ACNC50). This was found to be 99% similar to the corresponding gene from *Alexandrium catenella* strain
10 ACSH02.

A gene for *sxtA1* (SEQ ID NO: 227) was also amplified from the *Alexandrium* strain AAKT01, which has not previously been reported to produce saxitoxin.

No genes could be amplified from any of the remaining non-producing species of dinoflagellates.

15

Conclusion

The presence of genes putatively involved in the synthesis of saxitoxin in dinoflagellates have been confirmed in all species which produce saxitoxin. In this study, they were found in a further producing strain which has caused harmful algal blooms containing
20 saxitoxin in New Zealand, isolated as CAWD121. In addition, one species of *Alexandrium*, strain AAKT01, closely related to species that produce saxitoxin, was found to possess the gene *sxtA1*. This species has not previously been reported to produce saxitoxin but is now suggested to have the potential to produce it under certain circumstances. Toxin analyses using more sensitive detection techniques may be required
25 in order to verify this hypothesis. For example, Negri et al (2003) reported that the strain *A. tamarense*, ATBB01 showed STX activity when tested with the saxiphilin assay, but had no detectable toxins when tested with HPLC methods (see Negri, et al. (2003). "Paralytic shellfish toxins are restricted to few species among Australia's taxonomic diversity of cultured microalgae", J. Phycol. 39(4), 663-667.

30 None of the species of other dinoflagellate orders investigated in this study, not known to produce saxitoxin and not closely related to saxitoxin producing species, were found to possess the gene *sxtA1*.

Example 4: *Alexandrium catenella* from Opua Bay, New Zealand, as monitored by a qPCR assay based on *sxtA*

Aim

- 5 To determine whether the number of copies of the gene *sxtA* involved in the synthesis of saxitoxin in marine dinoflagellates, is correlated with a manual count of the number of cells of *Alexandrium catenella* in marine environmental samples.

Methods

- 10 Samples were collected at Opua Bay, at the south-eastern region of Onapua Bay in Queen Charlotte Sound, South Island, New Zealand. Once a week over 4 weeks (22/2/12 – 13/3/12), triplicate 500 ml samples of an integrated sample from the 0-15m depth from the water column were taken and stored at -80 degrees C. After defrosting, 200 ml of each sample was filtered through a 0.45um Millipore filter. The filter was rinsed in ~50 ul
15 of seawater into an eppendorf tube, vortexed and centrifuged. The pellet was extracted using the Mobio DNA soil kit, according to the manufacturer's instructions. DNA was eluted twice in 100 ul (same eluate was rinsed through filter twice). The DNA was concentrated into 20 ul of sample using a high salt/ethanol precipitation.

- The sequences of the primers were *sxtA*4F 5'-CTGAGCAAGGCGTTCAATTC-3' (SEQ ID NO: 198) and *sxtA*4R 5'-TACAGATMGGCCCTGTGARC-3' (SEQ ID NO: 20 199), resulting in a 125 bp product.

- Standard curves for *sxtA4* was constructed using a dilution series of a known concentration of fresh PCR product, ranging from $5.7-5.7 \times 10^{-5}$ ng (n=6). Standard curves using PCR product were used to determine the efficiency of the assay, as well as to
25 determine the number of copies. The molecules of PCR product were determined: $(A \times 6.022 \times 10^{23}) \times (660 \times B)^{-1}$ with A: concentration of PCR product, 6.022×10^{23} : Avogadro's number, 660: average molecular weight per base pair and B: length of PCR product.

- qPCR cycling was carried out on a Rotor Gene 3000 (Corbett Life Science) using
30 SSO Fast Evagreen supermix (Biorad). qPCR assays were performed in a final volume of 20 µl consisting of 10 µl Evagreen master mix (containing DNA intercalating dye, buffer and Taq polymerase), 1 µl of template DNA, 0.5 µM of each primer, and 1 µl of BSA. qPCR assays were performed in triplicate with the following cycles: 95°C for 10 s, and 35 replicates of 95°C for 15 s and 60°C for 30 s. Melting curve analysis was performed at the

end of each cycle to confirm amplification specificity, and selected PCR products were sequenced.

Results

5 The number of copies of *sxtA4* L⁻¹ detected was found to reach a peak in the third sampling week and then drop (Figure 10), similar to the change in abundance of *Alexandrium catenella* cells. The number of copies of *sxtA4* L⁻¹ in the triplicate water samples was found to be significantly correlated with the mean abundance of *Alexandrium catenella* cells L⁻¹ (Figure 11, R² = 0.86).

Conclusion

10 The quantitative qPCR assay based on the gene *sxtA* was found to be a reliable method of determining the potential for saxitoxin presence in marine environmental samples containing *Alexandrium catenella*. The abundance of copies of the gene *sxtA* were found to be significantly correlated with the abundance of the species *Alexandrium catenella*.

20 **Example 5: Amplification of *sxtA1* and *sxtA4* sequences and investigation of saxitoxin production and *sxtA* in the 'non-toxic' *Alexandrium tamarense* Group V clade**

Summary

25 The three *Alexandrium* species *A. tamarense*, *A. fundyense* and *A. catenella* include strains that can be potent producers of the neurotoxin saxitoxin (STX) and its analogues, the causative agents of Paralytic Shellfish Poisoning (PSP). These three species are morphologically highly similar, differing from each other only in the possession of a ventral pore, or in the ability to form chains. The appropriateness of these morphological characters for species delimitation has been extensively debated. A distinctive clade of this species complex, Group V, Tasmanian clade, is found in southern Australia, and occasionally occurs in bloom proportions. This clade has been considered non-toxic, and 30 no PSP toxins have been found in shellfish following blooms of this species. In the present study, a Tasmanian strain of *Alexandrium tamarense*, Group V, was identified that produces STX and possesses the gene, *sxtA* that is putatively involved in STX production. The toxin profile was determined and is unusual, including a high proportion of GTX5 and a small amount of STX, and differs from that of co-occurring *A. catenella* (Group IV).

A putative bloom of *A. tamarense* that occurred in October 2010, and the subsequent finding of STX in Sydney Rock Oysters (*Saccostrea glomerata*), may suggest that some naturally occurring strains of this species could produce STX.

5 Introduction

Three common and widespread species of the dinoflagellate genus *Alexandrium*, *A. catenella*, *A. tamarense* and *A. fundyense*, possess highly similar, sometimes overlapping morphological features (Balech, 1995; Fukuyo, 1985; Steidinger, 1990). This clade is considered to comprise a 'species complex', as it consists of five genetically distinct groups (John et al., 2003; Orr et al., 2011; Scholin et al., 1994; Lilly et al., 2007). The characteristics that are used for the identification of these species include the cell shape, shape of the apical pore complex (APC), presence (*A. tamarense*) or absence (*A. catenella*/*A. fundyense*) of a ventral pore on the first apical plate, and whether the cells show a tendency for chain formation (*A. catenella*) or not (*A. tamarense*/*A. fundyense*) (Balech, 1995). Some forms with morphologies intermediate between these three species have also been observed (Cembella et al., 1988; Gayoso and Fulco, 2006; Orlova et al., 2007; Sako et al., 1990; Orr et al., 2011).

In contrast to the information based on morphology, the many phylogenetic studies of *Alexandrium* species, based on regions of the rRNA operon, including the SSU, ITS/5.8s, and LSU genes, have clearly distinguished clades (Groups I-V) from one another (John et al., 2003; Scholin et al., 1994); (Kim and Sako, 2005; Leaw et al., 2005; Lilly et al., 2007; Montresor et al., 2004; Rogers et al., 2006; Orr et al., 2011). Based on a survey of dinoflagellate diversity and its relationship to rDNA sequences, Litaker et al., (2007) suggested that a conservative "species level" marker in dinoflagellates could be considered a difference of 4% (=uncorrected genetic distance of 0.04) in aligned regions of ITS1/5.8S/ITS2 rDNA. These clades of *Alexandrium tamarense/catenella/fundyense* differ from one another by 13-18 % in aligned sequences of ITS1/5.8S/ITS2 (Orr et al., 2011), therefore, at a level 3-4 times that in some other dinoflagellate species.

The identification of *Alexandrium tamarense/catenella/fundyense* strains to a particular genetic clade (Groups I-V) has been considered more predictive of their propensity for STX production than species identifications based on morphology (Scholin et al., 1994; John et al., 2003; Kim and Sako, 2005; Leaw et al., 2005; Montresor et al., 2004; Lilly et al., 2007; Rogers et al., 2006). All strains of Groups I and IV analysed to date produce varying quantities of STX, with diverse toxin profiles (Table 7, Anderson et

al., 1994), while no strains of Group II have been reported to produce STX (John et al., 2003). The toxicity of strains of Groups III and V is unclear. They have generally been considered non-toxic (Lilly et al., 2007; Scholin et al., 1994; Genovesi et al., 2011; Bolch and de Salas, 2007; Hallegraeff et al., 1991). A single strain with a genetic sequence placing it within Group III, CCMP116, has been reported to be toxic (Penna and Magnani, 1999). While Group V strains have generally been considered non-toxic (Hallegraeff et al., 1991; Bolch and de Salas, 2007), it has been suggested that very low levels of STXs may be produced by the strain ATBB01/CS298 from Bell Bay, Tasmania (Scholin et al., 1994; Negri et al., 2003). The toxin profile was not determined.

10 In Australian marine waters, the *Alexandrium* species *A. catenella* and *A. minutum* produce STX, and have occurred in bloom proportions, resulting in STX uptake in shellfish (Hallegraeff et al., 1988; Hallegraeff et al., 1991; Bolch and de Salas, 2007). Of the species of the *A. tamarensis* species complex, two groups have been consistently found in the region: Group V *A. tamarensis* and Group IV *A. catenella* (Bolch and de Salas, 2007). No other groups of this species complex have been found, during investigations over the past 20 years (Hallegraeff et al., 1988; Hallegraeff et al., 1991; Bolch and de Salas, 2007). Several theories have been put forward as to the origins of these *A. tamarensis* 'species complex' strains in Australian marine waters, including their introduction by ballast water (Group IV), or their long term presence in the region (Group V) (Bolch and de Salas, 2007).

20 Blooms of *A. catenella* (Group IV), *A. minutum* and the species *Gymnodinium catenatum*, have been associated with uptake of STX in shellfish vectors on multiple occasions at sites in New South Wales, South Australia, Victoria and Tasmania, Australia (reviewed in Bolch and de Salas, 2007). Potential shellfish vectors that have been investigated for the presence of STX, either experimentally or in the course of monitoring, in Australian waters are Sydney Rock Oysters (*Saccostrea glomerata*), Pacific Oysters (*Crassostrea gigas*), and Pearl Oysters *Pinctada imbricata* (Murray et al., 2009). Blooms of *A. tamarensis* Group V have occurred intermittently throughout the region, but have not been reported to cause STX uptake in shellfish (Hallegraeff et al., 1991).

30 In the course of investigating the genetic basis of STX production, genes for the putative *sxtA* domains *sxtA1* and *sxtA4* were discovered in three strains of *A. tamarensis* Group V (Stüken et al., 2011). These strains were reinvestigated to determine their genetic affinities and their potential for STX production.

This study describes the toxin profile, morphology and molecular phylogeny of a strain of *A. tamarense* that was found to produce STX. Furthermore, a finding of STX presence in samples of *S. glomerata* from New South Wales in 2010, following a putative bloom of this species is reported.

5

Materials and Methods

Culture maintenance

Dinoflagellate cultures were maintained in GSe media at 18°C. Light was provided by white fluorescent bulbs (Crompton Light), with photon flux of 60-100 μmol photon m^{-2} sec^{-1} on 12/12 hour dark/light cycle. Strains used were ATCJ33, isolated from Cape Jaffa, South Australia, Australia (-36.94,139.70); ATNWB01, isolated from North West Bay, Tasmania, Australia (-43.08,147.31); and ATEB01, isolated from Emu Bay, Burnie, Tasmania, Australia (-41.05,145.91), by M de Salas. Bulk cultures for toxin determination were inoculated on the same day in 2 L Erlenmeyer flasks and were harvested together, during late logarithmic or early stationary phase, for extraction of toxins. Cell abundance was determined by counting 1 ml subsamples using a Sedgewick Rafter counting chamber under a Leica DMIL Inverted Light microscope. Cultures were centrifuged and immediately frozen at -20 °C prior to HPLC analysis or DNA extraction. Cell pellets for HPLC analysis contained 1.25 – 2.25 x 10⁶ cells.

15
20

LM and SEM

Cell size and shape was determined using a Leica DMIL Inverted Light Microscope with 40 or 100 x magnification. For scanning electron microscopy, two methods were used, in order to either keep the cell membrane intact or to expose it. Cultures were fixed in 2% osmium tetroxide for 10 minutes, or in 4% glutaraldehyde for 1-2 hr. They were placed on a polylysine-coated cover slips or on 5 μm Millipore filters, rinsed in distilled water, and dehydrated in a series of increasing ethanol concentrations (30, 50, 70, 90, 100 %), followed by critical point drying (Baltec). When completely dry, they were mounted on stubs and sputter coated with gold. They were observed using a Zeiss Ultra Plus Field Emission Scanning Electron Microscope (FESEM) at the University of Sydney (Australian Centre for Microscopy and Microanalysis) at 5-15 kV.

25
30

DNA extraction and PCR

DNA was extracted from the cell pellets using the CTAB method, with an additional overnight DNA precipitation at -20°C. Quality and quantity of DNA was determined using a Nanodrop (Thermoscientific), and by amplifying a control dinoflagellate gene (*cytb*), according to the protocols of (Lin et al., 2009), using the primer pair 4f and 6r, which amplify a 440 bp fragment.

Partial sequences of the rRNA genes LSU and SSU and complete 5.8s/ITS genes were amplified using previously published primers: SS3, SS5 (Medlin et al., 1988), D1R, D3b (Scholin et al., 1994) and Alex5.8s (Orr et al., 2011). Typical cycling conditions for PCR consisted of an initial denaturing step of 94 °C for 2 min, followed by 35 cycles of 94 °C for 20 s, 56 °C for 30 s, and 72 °C for 1 min, followed by a final extension step of 7 min. PCR products were separated by agarose gel electrophoresis and stained with Ethidium Bromide, and visualised by UV transillumination. Fragments to be sequenced were excised from the gel, DNA was purified using a Bioline gel purification kit (Bioline, USA), eluted in 2 x 10 µl of elution buffer, and the concentration checked by use of a Nanodrop. Approximately 40 ng of PCR product was then used for direct sequencing with the same primers used for the initial amplification of the product.

Toxin determination

Dinoflagellate cell pellets were tested using HPLC, according to the AOAC Official Method 2005.06 for paralytic shellfish poisoning toxins in shellfish at the Cawthron Institute, New Zealand. A matrix modifier as described in the original protocol was not used; instead average spike recoveries were used for each separate compound. HPLC analysis was performed on a Waters Acquity UPLC system (Waters) coupled to a Waters Acquity FLR detector. Separation was achieved with a Waters Acquity C18 BEH 1.7 µm 2.1 x 50 mm column at 30°C, eluted at 0.2 mL min⁻¹. Mobile phases were 0.1 M ammonium formate (A) and 0.1 M ammonium formate in 5% acetonitrile (B), both adjusted to pH 6. The gradient consisted of 100% A for 0.5 min, a linear gradient to 80% B over 3.5 min, then returning to initial conditions over 0.1 min and held for 1.9 min. The fluorescence detector had excitation set to 340 nm and emission to 395 nm. Analytical standards for the STX analogs were obtained from the National Research Council, Canada. The detection limit of the HPLC of the cell cultures was considered to be 0.1 pg cell⁻¹ for NEO and STX, 0.2 pg cell⁻¹ for GTX1/4, GTX6 (B2) and GTX5 (B1), 0.5 pg cell⁻¹ for C1,2, and <0.3 pg cell⁻¹ for the analogs C3,4.

Phylogenetic analyses

New Group V *Alexandrium tamarense* sequences that were generated in this study; (1) 18S (Small Sub Unit) rDNA, (2) Internal Transcribed Region (ITS) 1 and 2 plus 5.8S rDNA, and (3) 28S (Large Sub Unit) rDNA, were concatenated to construct a 2,821 character region of the rDNA operon for strain ATNWB01 (GenBank accession numbers JQ991015, JQ991016, JQ991017). In addition, new LSU rDNA sequences were generated from strains ATCJ33 (874bp) and ATEB01 (678 bp length, GenBank Accession numbers JQ991018 and JQ991019). This was aligned together with all orthologues from Group V sequence data in the NCBI nr nucleotide database <http://www.ncbi.nlm.nih.gov/> (as of 10.2010), and an *A. affine* outgroup. MAFFT v6 Q-INS-I model (Hofacker et al., 2002; Katoh and Toh, 2008; Kiryu et al., 2007), considering secondary RNA structure, was used to align the dataset (default parameters used) and the resulting alignment checked manually using MacClade v4.07 (Madison and Madison, 1992). The alignment was then inferred with Gblocks v0.91b (Castresana, 2000), under the least stringent parameters, to exclude poorly aligned positions and divergent regions from the phylogenetic inference. MODELTEST (Posada and Crandall, 1998) established the optimal model of nucleotide evolution; for all alignments GTR was preferred for both the Akaike and Bayesian information Criterion (AIC and BIC). Maximum Likelihood (ML) analyses were performed with RAXML-VI-HPC v7.2.6, GTRCAT model with 25 rate categories (Stamatakis, 2006). The most likely topology was established from 100 separate searches and bootstrap analyses were performed with 100 pseudoreplicates. Bayesian inferences were performed using Phylobayes v3.2e (Lartillot et al., 2007; Lartillot and Philippe, 2004) under the GTRCAT substitution model with a free number of mixing categories and a discrete across site variation under 4 categories. Trees were inferred when the largest maximum difference between the bipartitions (chains) was <0.1. All model estimation and phylogenetic analyses were performed on the freely available Bioportal (Kumar et al., 2009) at the University of Oslo (<http://www.bioportal.uio.no/>).

Phytoplankton and shellfish sample collection, counting and identification

Phytoplankton was collected as part of the NSW Shellfish Program fortnightly monitoring at the Hastings River, New South Wales, Australia (-31.42 E, 152.87 S) over two consecutive weeks in October and November 2010, 25/10/10 and 9/11/10. In the second week, Sydney Rock Oysters (*Saccostrea glomerata*) were collected from farms and tested using a Jellett PSP test (Jellett Rapid Testing Ltd, Canada) according to the manufacturer's instructions.

For counting of bottle samples, phytoplankton cells in 500 ml of Lugol's preserved samples were concentrated by gravity assisted membrane filtration on to 5 µm cellulose ester filters (Advantec) prior to washing into 4 ml and counting. *Alexandrium* species were identified and counted using a Sedgewick Rafter cell and a Zeiss Axiolab microscope equipped with phase-contrast optics. Species were identified by Dr S. Brett, who has identified harmful phytoplankton as part of the NSW Shellfish Program since 2003. *Alexandrium* species occur commonly in NSW waters, the most commonly identified species being *A. catenella* and *A. pseudogonyaulax*. The number of cells counted varied among samples, depending on *Alexandrium* abundance, and standard error rates were calculated using the equation: $\text{Error} = 2 / \sqrt{n}$, where n is the number of cells observed in the sample.

Results

Light and Scanning Electron Microscopy

Cells of ATNWB01 were rounded, 25 – 45 µm long ($\mu = 34.2$, $n=20$), 27 - 40 µm wide ($\mu = 33.6$, $n=20$) (Fig. 12, A-B). Cells were almost always single. Chains of two cells were rarely observed. The first apical plate extended to the apical pore complex and contained a small pore, the ventral pore (Fig. 12, C). The apical pore plate (Po) had a characteristic comma shape, surrounded by small marginal pores (Fig. 12E). Occasionally cells with a pore in the posterior sulcal plate were observed (Fig. 12 F). Most cells lacked a pore in the posterior sulcal plate.

Cells of the cultures ATCJ33 (Fig. 13 A-D), ATEB01 (Fig. 13, E, F, H) and ATBB01 (Fig. 13 G) had rounded cells. Cells were mostly single, occasionally in chains of 2 cells (Fig. 13 A). Cells of ATCJ33 were 23-38 µm long ($\mu = 30.6$, $n=20$), 23 - 38 µm wide ($\mu = 29.5$, $n=20$), cells of ATEB01 were 25-40 µm long ($\mu = 33.5$, $n=20$), 23 - 35 µm wide ($\mu = 31.6$, $n=20$). The first apical plate extended to the apical pore and contained a ventral pore (Fig. 13 B, D, H). The apical pore plate showed a characteristic comma shape (Fig. 13 C, G, H).

Phylogenetic analysis

The phylogenetic analyses including the new rDNA sequences (Fig. 14, Fig. 16), the SSU rDNA, ITS/5.8s and partial LSU rDNA sequences of the toxin producing strain ATNWB01 and the LSU sequences of the two non-toxic strains ATCJ33 and ATEB01, show that all of these form a well-supported clade (1.00/100, for PP/BS support) together

with other strains of *A. tamarens* Group V. No difference could be observed in this clade based on full length rDNA, including the most variable ITS regions.

Comparison of sxtA1 and sxtA4 domains of sxtA genes in the strains

5 The nucleotide content of sequences for the two domains of *sxtA* in strains of *A. tamarens* Group 5 was compared, previously sequenced (Stülken et al., 2011). A 440 bp sequence of domain *sxtA4* showed that the strain ATNWB01 differed from the two other strains of Group 5 *A. tamarens* analysed, ATEB01 and ATCJ33, in only 2 nucleotides (99.5% sequence identity). In contrast, these three strains showed 89-98% sequence
10 identity with these genes from other species of *Alexandrium* and *Gymnodinium catenatum*. In a comparison of a 450 bp region of the domain *sxtA1*, strain ATNWB01 differed from the other 2 strains of Group 5 by 2-2.5% (97.5-98% sequence identity), as compared to a sequence identity of 70-93% for other *Alexandrium* and *Gymnodinium catenatum* sequences of *sxtA1* domain.

15

Toxins

Of the three strains of *A. tamarens*, Group V, approximately $1.5 - 2.2 \times 10^6$ cells were tested for toxicity using HPLC. Two strains, ATCJ33 and ATEB01, had negative profiles, with no detectable toxins. One strain, ATNWB01, showed positive results with a
20 profile consisting of mostly GTX5, with some STX, C1,2 and deSTX (Fig. 15, Table 7), and a cell quota of $15.3 \text{ fmol cell}^{-1}$.

Environmental samples

Alexandrium cells were detected at levels of 350 cells L⁻¹ in routine monitoring
25 samples collected from the Hastings River on 25/10/10 (Table 6). Subsequent Jellet tests on oyster samples from the same site collected on 9/11/10 were positive for PSP toxins. *Alexandrium* cells in the samples were single cells and had a generally rounded shape. Cells were 35 μm in length, and 32-35 μm in width. Examination of *Alexandrium* theca revealed a ventral pore in the 1' plate, a 1' plate of the size and shape of *A. tamarens*, and
30 the absence of a connecting pore in the APC.

Of the species of *Alexandrium* observed in Australian waters, the size and shape of cells, the shape of the 1' plate, the presence of the ventral pore on the 1' plate, and the shape of the APC, appear most consistent with *Alexandrium tamarens*. No cultures were

established from this bloom event, however, and molecular sequences nor toxin profiles could be confirmed.

Table 6: Results from phytoplankton and oyster toxicity monitoring, Hastings River.

Site	Species	Cells L ⁻¹	Date	Jellet test result on oyster samples
Hastings River	<i>Alexandrium tamarense</i>	350	25/10/2010	Negative
	<i>Alexandrium tamarense</i>	350	9/11/2010	Positive (PSP)

Table 7. Toxin profiles of *Alexandrium catenella*/fundyense/tamarensis species complex, showing molar % and total toxin content. ND=not detected

Species	Strain	Clade	Profile	C3.4	GTX 1,4	GTX 2,3	GTX (B1)	GTX6 (B2)	neoSTX	STX	deSTX	Toxin content fmol/cell	Reference
<i>Alexandrium tamarense</i>	ATBR2c	I	70	-	10	3	-	-	16	1	-	42-199*	Persich et al 2006
<i>Alexandrium tamarense</i>	ATBR2d	I	80	-	10	2	-	-	8	-	-	42-199	Persich et al 2006
<i>Alexandrium tamarense</i>	ATBR2e	I	60	-	19	2	-	-	19	-	-	42-199	Persich et al 2006
<i>Alexandrium tamarense</i>	ATBR2g	I	63	-	33	2	-	-	2	-	-	42-199	Persich et al 2006
<i>Alexandrium catenella</i>	PFB38	I	-	-	23.0	77.1	-	-	-	-	-	18.5	Aguilera-Belmonte et al 2011
<i>Alexandrium catenella</i>	PFB39	I	44.4	-	9.2	4.2	14.2	26.4	-	1.6	-	24.7	Aguilera-Belmonte et al 2011

<i>Alexandrium catenella</i>	PF336	I	26.2	-	13.7	13.7	33.7	0.1	12.5	-	-	92.0	Aguilera- Belmonte et al 2011
<i>Alexandrium catenella</i>	PF342	I	-	-	25.9	74.0	-	-	-	-	-	18.3	Aguilera- Belmonte et al 2011
<i>Alexandrium catenella</i>	PF345	I	1.1	-	13.3	27.6	15.9	15.5	20.1	2.5	-	96.9	Aguilera- Belmonte et al 2011
<i>Alexandrium catenella</i>	PF337	I	-	-	59.6	40.3	-	-	-	-	-	11.7	Aguilera- Belmonte et al 2011
<i>Alexandrium catenella</i>	PF341	I	-	-	38.1	60.2	-	-	-	-	-	8.5	Aguilera- Belmonte et al 2011
<i>Alexandrium tamarense</i>	SZN01	II	-	-	-	-	-	-	-	-	-	ND	John et al 2003
<i>Alexandrium tamarense</i>	SZN08	II	-	-	-	-	-	-	-	-	-	ND	John et al 2003

<i>Alexandrium tamarense</i>	SZN19	II															ND	John et al 2003
<i>Alexandrium tamarense</i>	SZN21	II															ND	John et al 2003
<i>Alexandrium tamarense</i>	Various strains	III															ND	Higman et al 2001
<i>Alexandrium catenella</i>	ATIL01	IV	54	2	-	-	44	-	-	-	-	-	-	-	-	-	44.3	Lilly et al 2002
<i>Alexandrium catenella</i>	ATIL02	IV	33	20	1	-	46	-	-	-	-	-	-	-	-	-	5.3	Lilly et al 2002
<i>Alexandrium catenella</i>	ACFP09	IV	22.2	13.7	30.1	0.5	1.9	30.4	1.0	-	-	-	-	-	-	-		Hallegre ff et al 1991
<i>Alexandrium catenella</i>	ACFP02	IV	11.9	4.6	21.3	0.2	4.0	57.3	0.4	-	-	-	-	-	-	-		Hallegre ff et al 1991
<i>Alexandrium catenella</i>	CAWD44	IV	53	35	8	2	-	-	2	-	-	-	-	-	-	-	150.2	MacKenzi e et al 2004*/me

an of 13 isolates)												
<i>Alexandrium tamarense</i>	CAWD12 ₁	IV	8	90	2	-	-	-	-	-	328.5	MacKenzie et al 2004
<i>Alexandrium catenella</i>	ACSH02	IV	41	-	35	-	10	13	-	-	12.1	Murray et al 2011
<i>Alexandrium catenella</i>	ACCC01	IV	36	14	16	-	-	34	-	-	14.7	Murray et al 2011
<i>Alexandrium catenella</i>	ACTRA0 ₂	IV	59	-	40	-	-	-	-	-	3.5	Murray et al 2011
<i>Alexandrium tamarense</i>	ATEB01	V									ND	Orr et al 2011
<i>Alexandrium tamarense</i>	ATEB01	V									ND with HPLC, but some response with saxiphilin assay	Negri et al 2003
<i>Alexandrium tamarense</i>	ATCJ33	V									ND	This study
<i>Alexandrium tamarense</i>	ATEB01	V									ND	This study
<i>Alexandrium tamarense</i>	ATNWB0 ₁	V	0.1	-	-	-	86	-	11	2	15.3	This study

Discussion

In general among strains of toxic species of the genus *Alexandrium*, the toxin profile appears to remain largely constant over time (Cembella and Destombe, 1996) apart from situations of extreme nutrient depletion (Boczar et al., 1988). In contrast the amount of toxin produced (cell quota) has been found to vary over time. While toxin profiles appear to be stable within a strain, the relationship between a particular toxin profile and the identification of a strain to a particular genetic group (I-V) of this species complex is unclear. Considerable variation in toxin profiles have been found amongst strains of each group examined (Table 7).

This study is the first report of the toxin profile of a strain of *Alexandrium tamarense*, Group V. The study shows that, based on morphological features such as the presence of the ventral pore, and long sequences of rDNA genes, including the variable ITS1/5.8s/ITS2 regions, the strain ATNWB01 was a member of Group V. Previously, strains of this group have been tested for toxicity by means of HPLC and found non-toxic (Hallegraeff et al., 1991; Salas et al., 2001; Bolch and De Salas, 2007). Negri et al (2003) reported that the strain *A. tamarense*, ATBB01 showed STX activity when tested with the saxiphilin assay, but had no detectable toxins when tested with HPLC methods. In the same study, a culture that had been identified as *A. tamarense* from Australian waters (strain ATTRA03) was found to produce STX (Negri et al., 2003). Strain ATTRA03 subsequently died without verification of its identity, and it is therefore unclear which species or group this strain represented. It may have been derived from an introduced cyst, as it was isolated as a cyst from a shipping port used for exports (Bolch and de Salas, 2007).

Three strains from Japan, including At304, isolated from Mikawa Bay, Japan, were found to be members of the Group V clade in the phylogenetic analysis (Fig. 14). This is the first report of a Group V strain being present in the north western Pacific region, as it was previously thought to be confined to southern Australia (Bolch and de Salas 2007). The toxicity of strains from Japan was not determined, and needs to be examined.

The strain ATNWB01 had a STX cell quota well within the detectable range using standard HPLC methods (15.3 fmol cell⁻¹ Table 7). This is similar to the toxin quota reported for many common STX-producing strains of this species complex (Table 7, 3.5-328 fmol cell⁻¹, 0.66 - 9.8 pg STX equivalents cell⁻¹). The toxin profile of this strain was relatively unusual, as GTX5 has not commonly been reported to be a major part of the toxin profile of strains of *Alexandrium tamarense* species complex (Table 7; Anderson et

al., 1994). In general, strains of *Alexandrium catenella* (Group IV), the most common source of PSTs in shellfish in New South Wales, have contained high proportions of C1/2 and GTX1/4 as major components (Murray et al, 2011; Negri et al 2003; Hallegraeff et al., 1991). GTX5 has been found to be a major component of a strain of *Alexandrium tamarens* Group I from Chile, (33.7%, Aguilera-Belmonte et al., 2011), and two strains of *Alexandrium catenella* (Group IV) from France, (44-46%, Lilly et al., 2002).

No cultures were made of the *A. tamarens* strain identified from the Hastings River sample, and no molecular sequences were determined, therefore it is not possible to definitively identify it as *A. tamarens* Group V. Fortnightly phytoplankton monitoring has been undertaken at 69 sites in 32 estuaries in New South Wales for the past 7 years. *Alexandrium catenella* has been identified on 8 sampling occasions in 2010 and 2011. On 7 of these, PSP toxins were subsequently found in sample oysters from neighbouring oyster harvest areas, using Jellet PSP tests (NSW Food Authority, unpublished data). *Alexandrium tamarens* was identified from 4 samples in 2010 and 2011. The sample from the Hastings River is the first report of it being associated with PSP toxicity in Australian oysters. The analogues produced by *Alexandrium tamarens* Group V, in particular, the high proportion of GTX5, has a low equivalent toxicity when compared to the most toxic analogues such as STX, and it is estimated to be about 10% as toxic as STX (Oshima, 1995). This low STX equivalent toxicity may explain why this is the first report of an incident of shellfish toxicity in this region that is putatively linked to a bloom of this species. Further sampling of *Alexandrium tamarens* Group V populations in NSW coastal waters is required in order to verify whether local populations can indeed produce toxins.

The three strains of *A. tamarens* have all been found to possess the gene *sxtA* (Stüken et al., 2011). It was found that *sxtA* genes were closely related to each other in all Group V strains (0.5 – 2.5% differences in aligned sequences for domains *sxtA1* and *sxtA4*). The presence or absence of these genes in these strains therefore appears to be unrelated to toxin production.

Differences in growth and toxin production in the species *Alexandrium catenella*, *Alexandrium tamarens* and *Alexandrium minutum* have been previously reported, related to the environmental conditions of the culture, such as salinity, light and nutrients, and the growth phase of the culture (Hamasaki et al., 2001; Hu et al., 2006; Lippemeier et al., 2003; Anderson et al. 1990; Grzebyk et al., 2003). In the present study, each of the three *A. tamarens* strains were cultured under identical light conditions, in the same media and

seawater, and were inoculated and then harvested on the same days. Therefore, it seems unlikely that the lack of toxicity found in the other two *Alexandrium tamarense* Group V strains is a result of differences in environmental conditions promoting the differential expression of STX production.

5 Some cultured strains of *Alexandrium* may lose toxicity over time in culture, for example, in *Alexandrium minutum* (reported as *A. lusitanicum* in Martins et al., 2004). This was initially thought to be related to antibiotic exposure, as co-cultured and symbiotic bacteria have been shown to play a role in mediating toxin production in *Alexandrium* species (Ho et al., 2006). In the present study, the cultures of Group V
10 examined most likely contain mixed bacterial communities in line with that of the seawater at the site of isolation, and none had been exposed to antibiotics.

 Mating experiments have found that subclones of a toxic clonal strain of *Alexandrium tamarense*, (Group IV) can be non-toxic (Cho et al., 2008). The non-toxic characteristics of one strain of *A. tamarense*, an axenic non-toxic subclone of a toxic
15 strain, were confirmed at the attomole per cell level. Three out of nine toxic subclones of this same strain became non-toxic over a relatively short period of time (4–6 years), while the other toxic subclones retained their toxicity and the non-toxic subclones remained non-toxic (Cho et al., 2008).

 High levels of population genetic differences have been found within the species
20 *Alexandrium catenella* Group IV (Masseret et al., 2009), *Alexandrium tamarense* Group I (Nagai et al., 2007), (Alpermann et al., 2009), *A. fundyense* Group I (Erdner et al., 2011) and *A. minutum* (McCauley et al., 2009), using microsatellite markers. This, in combination with the results of strains with differing toxin production following intraspecific mating experiments (Cho et al., 2008), suggests that significant population
25 level differences in toxin production may also exist within *A. tamarense* (Group V). Further studies using microsatellite markers on multiple clonal cultures of this strain may help to determine whether toxin production is restricted to one particular population of this species, which may allow for the design of predictive genetic tools for the identification of this population.

30

References

Aguilera-Belmonte, A., Inostroza, I., Franco, J.M., Riob , P., Gomez, P.I., 2011. The growth, toxicity and genetic characterization of seven strains of *Alexandrium catenella*

- (Whedon and Kofoid) Balech 1985 (Dinophyceae) isolated during the 2009 summer outbreak in southern Chile. Harmful Algae. In Press.
- Alpermann, T.J., Beszteri, B., John, U., Tillmann, U., Cembella, A.D., 2009. Implications of life history transitions on the population genetic structure of the toxigenic marine dinoflagellate *Alexandrium tamarens*. Mol. Ecol. 18(10), 2122-2133.
- 5 Anderson, D., Kulis, D., Doucette, G., Gallagher, J., Balech, E., 1994. Biogeography of toxic dinoflagellates in the genus *Alexandrium* from the northeastern United States and Canada. Mar. Biol. 120(3), 467-478.
- Anderson, D., Kulis, D., Sullivan, J., Hall, S., 1990. Toxin composition variations in one isolate of the dinoflagellate *Alexandrium fundyense*. Toxicon 28(8), 885-893.
- 10 Balech, E. 1995. The genus *Alexandrium* Halim (Dinoflagellata). Shenkin Island Press.
- Boczar, B.A., Beitler, M.K., Liston, J., Sullivan, J.J., Cattolico, R.A., 1988. Paralytic shellfish toxins in *Protogonyaulax tamarens* and *Protogonyaulax catenella* in axenic culture. Plant Physiol. 88(4), 1285.
- 15 Bolch, C.J.S., de Salas, M.F., 2007. A review of the molecular evidence for ballast water introduction of the toxic dinoflagellates *Gymnodinium catenatum* and the *Alexandrium*. Harmful Algae 6(4), 465-485.
- Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17(4), 540-552.
- 20 Cembella, A., Destombe, C., 1996. Genetic differentiation among *Alexandrium* populations from eastern Canada. In: Harmful and Toxic Algal Blooms. Intergovernmental Oceanographic Commission, UNESCO, Paris, 447-450.
- Cembella, A., Taylor, F., Therriault, J.C., 1988. Cladistic analysis of electrophoretic variants within the toxic dinoflagellate genus *Protogonyaulax*. Bot. Mar. 31(1), 39-52.
- 25 Cho, Y., Hiramatsu, K., Ogawa, M., Omura, T., Ishimaru, T., Oshima, Y., 2008. Non-toxic and toxic subclones obtained from a toxic clonal culture of *Alexandrium tamarens* (Dinophyceae): Toxicity and molecular biological feature. Harmful Algae 7(6), 740-751.
- de Salas, M.F., Emmerik, M.J., Hallegraeff, G., Negri, A., Vaillancourt, R., Bolch, C., 2001. Toxic Australian *Alexandrium* Dinoflagellates: Introduced or indigenous? In: Hallegraeff, G.M., Blackburn, S.I., Bolch, C.J., Lewis, R.J. (Eds.), Proceedings of the Ninth International Conference on Harmful Algal Blooms. Intergovernmental Oceanographic Commission of UNESCO, Hobart, Tasmania, Harmful Algal Blooms 2000., pp 477-480.
- 30

- Erdner, D.L., Richlen, M., McCauley, L.A. R., Anderson, D.M., 2011. Diversity and dynamics of a widespread bloom of the toxic dinoflagellate *Alexandrium fundyense*. PloS One 6(7), e22965.
- Fukuyo, Y., 1985. Morphology of *Protogonyaulax tamarens* (Lebour) Taylor and
5 *Protogonyaulax catenella* (Whedon and Kofoid) Taylor from Japanese coastal waters. Bull. Mar. Sci. 37(2), 529-537.
- Gayoso, A.M., Fulco, V.K., 2006. Occurrence patterns of *Alexandrium tamarens* (Lebour) Balech populations in the Golfo Nuevo (Patagonia, Argentina), with observations on ventral pore occurrence in natural and cultured cells. Harmful Algae 5(3),
10 233-241.
- Genovesi, B., Shin-Grzebyk, M.S., Grzebyk, D., Laabir, M., Gagnaire, P.A., Vaquer, A., Pastoureaud, A., Lasserre, B., Collos, Y., Berrebi, P., 2011. Assessment of cryptic species diversity within blooms and cyst bank of the *Alexandrium tamarens* complex (Dinophyceae) in a Mediterranean lagoon facilitated by semi-multiplex PCR. J. Plankt.
15 Res. 33(3), 405.
- Grzebyk, D., Bechemin, C., Ward, C.J., Verite, C., Codd, G.A., Maestrini, S.Y., 2003. Effects of salinity and two coastal waters on the growth and toxin content of the dinoflagellate *Alexandrium minutum*. J. Plankt. Res. 25(10), 1185.
- Hallegraeff, G., Bolch, C., Blackburn, S., Oshima, Y., 1991. Species of the toxigenic
20 dinoflagellate genus *Alexandrium* in southeastern Australian waters. Bot. Mar. 34(6), 575-588.
- Hallegraeff, G., Steffensen, D., Wetherbee, R., 1988. Three estuarine Australian dinoflagellates that can produce paralytic shellfish toxins. J. Plankt. Res. 10(3), 533.
- Hamasaki, K., Horie, M., Tokimitsu, S., Toda, T., Taguchi, S., 2001. Variability in
25 toxicity of the dinoflagellate *Alexandrium tamarens* isolated from Hiroshima Bay, western Japan, as a reflection of changing environmental conditions. J. Plankt. Res. 23(3), 271.
- Higman, W.A., Stone, D.M., Lewis, J.M., 2001. Sequence comparisons of toxic and non-toxic *Alexandrium tamarens* (Dinophyceae) isolates from UK waters. Phycologia 40 (3),
30 256-262.
- Ho, A.Y.T., Hsieh, D.P.H., Qian, P.Y., 2006. Variations in paralytic shellfish toxin and homolog production in two strains of *Alexandrium tamarens* after antibiotic treatments. Aquat. Microb. Ecol. 42(1), 41-53.

- Hofacker, I.L., Fekete, M., Stadler, P.F., 2002. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* 319(5), 1059-1066.
- Hu, H., Chen, W., Shi, Y., Cong, W., 2006. Nitrate and phosphate supplementation to increase toxin production by the marine dinoflagellate *Alexandrium tamarense*. *Mar. Poll. Bull.* 52(7), 756-760.
- John, U., Fensome, R.A., Medlin, L.K., 2003. The application of a molecular clock based on molecular sequences and the fossil record to explain biogeographic distributions within the *Alexandrium tamarense* species complex (Dinophyceae). *Mol. Biol. Evol.* 20(7), 1015-1027.
- 10 Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. *Briefings Bioinf.* 9(4), 286-298.
- Kim, C.J., Sako, Y., 2005. Molecular identification of toxic *Alexandrium tamiyavanichii* (Dinophyceae) using two DNA probes. *Harmful Algae* 4(6), 984-991.
- Kiryu, H., Tabei, Y., Kin, T., Asai, K., 2007. Murelet: a practical multiple alignment tool
15 for structural RNA sequences. *Bioinformatics* 23(13), 1588-1598.
- Kumar, S., Skjölveland, Å., Orr, R., Enger, P., Ruden, T., Mevik, B.H., Burki, F., Botnen, A., Shalchian-Tabrizi, K., 2009. AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics* 10(1), 357.
- 20 Lartillot, N., Brinkmann, H., Philippe, H., 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* 7(Suppl 1), S4.
- Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21(6), 1095-1109.
- 25 Leaw, C.P., Lim, P.T., Ng, B.K., Cheah, M.Y., Ahmad, A., Usup, G., 2005. Phylogenetic analysis of *Alexandrium* species and *Pyrodinium bahamense* (Dinophyceae) based on theca morphology and nuclear ribosomal gene sequence. *J. Phycol.* 44(5).
- Lilly, E., Kulis, D., Gentien, P., Anderson, D., 2002. Paralytic shellfish poisoning toxins in France linked to a human-introduced strain of *Alexandrium catenella* from the western
30 Pacific: evidence from DNA and toxin analysis. *J. Plankt. Res.* 24(5), 443.
- Lilly, E.L., Halanych, K.M., Anderson, D.M., 2007. Species boundaries and global biogeography of the *Alexandrium tamarense* complex (Dinophyceae) 1. *J. Phycol.* 43(6), 1329-1338.

- Lin, S., Zhang, H., Hou, Y., Zhuang, Y., Miranda, L., 2009. High-level diversity of dinoflagellates in the natural environment, revealed by assessment of mitochondrial *cox1* and *cob* genes for dinoflagellate DNA barcoding. *Appl. Environ. Microbiol.* 75(5), 1279.
- Lippemeier, S., Frampton, D.M.F., Blackburn, S.I., Geier, S.C., Negri, A.P., 2003. Influence of phosphorous limitation on toxicity and photosynthesis of *Alexandrium minutum* (Dinophyceae) monitored by online detection of variable chlorophyll fluorescence. *J. Phycol.* 39(2), 320-331.
- Litaker, R.W., Vandersea, M.W., Kibler, S.R., Reece, K.S., Stokes, N.A., Lutzoni, F.M., Yonish, B.A., West, M.A., Black, M.N.D., Tester, P.A., 2007. Recognizing dinoflagellate species using ITS rDNA sequences. *J. Phycol.* 43(2), 344-355.
- Madison, W., Madison, D., 1992. *MacClade: interactive analysis of phylogeny and character evolution*. Sunderland, Massachusetts: Sinauer Associates.
- Martins, C.A., Kulis, D., Franca, S., Anderson, D.M., 2004. The loss of PSP toxin production in a formerly toxic *Alexandrium lusitanicum* clone. *Toxicon* 43(2), 195-205.
- Masseret, E., Grzebyk, D., Nagai, S., Genovesi, B., Lasserre, B., Laabir, M., Collos, Y., Vaquer, A., Berrebi, P., 2009. Unexpected genetic diversity among and within populations of the toxic dinoflagellate *Alexandrium catenella* as revealed by nuclear microsatellite markers. *Appl. Environ. Microbiol.* 75(7), 2037-2045.
- McCauley, L.A.R., Erdner, D.L., Nagai, S., Richlen, M.L., Anderson, D.M., 2009. Biogeographic analysis of the globally distributed harmful algal bloom species *Alexandrium minutum* (Dinophyceae) based on rRNA gene sequences and microsatellite markers. *J. Phycol.* 45(2), 454-463.
- Medlin, L., Elwood, H.J., Stickel, S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* 71(2), 491-499.
- Montresor, M., John, U., Beran, A., Medlin, L.K., 2004. *Alexandrium tamatum* sp. nov. (dinophyceae): A new nontoxic species in the genus *Alexandrium*. *J. Phycol.* 40(2), 398-411.
- Murray, S.A., O'Connor, W.A., Alvin, A., Mihali, T.K., Kalaitzis, J., Neilan, B.A., 2009. Differential accumulation of paralytic shellfish toxins from *Alexandrium minutum* in the pearl oyster, *Pinctada imbricata*. *Toxicon* 54(3), 217-223.
- Murray, S.A., Wiese M., Stüken, A., Brett, S. Kellmann, R., Hallegraeff, G.M., Neilan, B.A., 2011. *sxtA*-based quantitative molecular assay to identify saxitoxin-producing harmful algal blooms in marine waters. *Appl. Environ. Microbiol.* 77:7050-7057.

- Nagai, S., Lian, C., Yamaguchi, S., Hamaguchi, M., Matsuyama, Y., Itakura, S., Shimada, H., Kaga, S., Yamauchi, H., Sonda, Y., 2007. Microsatellite markers reveal population genetic structure of the toxic dinoflagellate *Alexandrium tamarense* (dinophyceae) in Japanese coastal waters. *J. Phycol.* 43(1), 43-54.
- 5 Negri, A., Llewellyn, L., Doyle, J., Webster, N., Frampton, D., Blackburn, S., 2003. Paralytic shellfish toxins are restricted to few species among Australia's taxonomic diversity of cultured microalgae. *J. Phycol.* 39(4), 663-667.
- Orlova, T.Y., Selina, M.S., Lilly, E.L., Kulis, D.M., Anderson, D.M., 2007. Morphogenetic and toxin composition variability of *Alexandrium tamarense* (Dinophyceae) from the east coast of Russia. *J. Phycol.* 46(5).
- 10 Orr, R.J.S., Stüken, A., Rundberget, T., Eikrem, W., Jakobsen, K.S., 2011. Improved phylogenetic resolution of toxic and non-toxic *Alexandrium* strains using a concatenated rDNA approach. *Harmful Algae* 10(6), 676-688.
- Oshima, Y., 1995. Postcolumn derivatization liquid chromatographic method for paralytic shellfish toxins. *J. AOAC International* 78(2), 528-532.
- 15 Parkhill, J.P., Cembella, A.D., 1999. Effects of salinity, light and inorganic nitrogen on growth and toxigenicity of the marine dinoflagellate *Alexandrium tamarense* from northeastern Canada. *J. Plankt. Res.* 21(5), 939-955.
- Penna, A., Magnani, M., 1999. Identification of *Alexandrium* (Dinophyceae) species using PCR and rRNA targeted probes. *J. Phycol.* 35(3), 615-621.
- 20 Persich, G., Kulis, D., Lilly, E., Anderson D. M., Garcia V.M.T. 2006. Probable origin and toxin profile of *Alexandrium tamarense* (Lebour) Balech from southern Brazil. *Harmful Algae* 5, 36-44.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14(9), 817-818.
- 25 Rogers, J.E., Leblond, J.D., Moncreiff, C.A., 2006. Phylogenetic relationship of *Alexandrium monilatum* (Dinophyceae) to other *Alexandrium* species based on 18S ribosomal RNA gene sequences. *Harmful Algae* 5(3), 275-280.
- Sako, Y., Kim, C.H., Ninomiya, H., Adachi, M., Ishida, Y., 1990. Isozyme and cross analysis of mating populations in the *Alexandrium catenella/tamarense* species complex. In: *Toxic Marine Phytoplankton*. Elsevier, New York, 320-323.
- Scholin, C.A., Herzog, M., Sogin, M., Anderson, D.M., 1994. Identification of group and strain specific genetic markers for globally distributed *Alexandrium* (dinophyceae). II Sequence analysis of a fragment of the LSU rRNA gene. *J. Phycol.* 30(6), 999-1011.

- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21), 2688.
- Steidinger, K.A., 1990. Species of the *tamarensis/catenella* group of *Gonyaulax* and the fucoxanthin derivative-containing gymnodinioids. In: Graneli, E., Sundstrom, B., Edler, L., Anderson, D.M. (Eds.), *Toxic Marine Phytoplankton*. Elsevier, New York., pp. 11-16.
- Stüken, A., Orr, R.J.S., Kellmann, R., Murray, S.A., Neilan, B.A., Jakobsen, K.S., 2011. Discovery of nuclear-encoded genes for the neurotoxin saxitoxin in dinoflagellates. *PloS One* 6(5), e20096.
- Wohlrab, S., M. H. Iversen, and U. John. 2010. A molecular and co-evolutionary context for grazer induced toxin production in *Alexandrium tamarense*. *PloS One* 5:e15039.
- Yang, I., U. John, S. Beszteri, G. Glöckner, B. Krock, A. Goesmann, and A. D. Cembella. 2010. Comparative gene expression in toxic versus non-toxic strains of the marine dinoflagellate *Alexandrium minutum*. *BMC Genomics* 11:248.

15 **Example 6: qPCR reactions for the detection of sxtA4 and discerning between toxic and non-toxic strains of *Alexandrium minutum***

Materials and Methods

10 10 µl or 20 µl qPCR reactions were run on a Roche LightCycler®480 system in a white 96 well plate. Each reaction contained 5 or 10 µl LightCycler® 480 SYBR Green I Master, 125 nM of each primer (sxt072 and sxt073) and template (*Alexandrium* cDNA or gDNA). Reactions were run in duplicate. Standard curves and non-template controls were included on each plate. Standard curves were generated with 10x serial dilutions of a gel-purified PCR amplicon generated from strain CCMP113 with primers sxt007 & sxt008

25 (PCR to get amplicon as as described in Stüken et al. 2011, PlosOne). qPCR cycling parameters were: hot-start: 1x (95°C, 10min); amplification: 45x (94°C, 10s; 64°C, 20s; 72°C, 10s, single acquisition; melting curve: 1x (95°C, 5s; 65°C, 1min; up to 97°C continuous measurements); Cooling: 1x (40°C, 10s). Crossing point and meltcurve analyses were carried out using the software supplied by Roche.

30

The qPCR primers used for the detection of sxtA in these experiments were:

sxt072 CTTGCCCCGCCATATGTGCTT (SEQ ID NO: 228)

sxt073 GCCCGGCGTAGATGATGTTG (SEQ ID NO: 229)

Results

The following strains were tested by the qPCR outlined above.

	Species	Strain	SXT-synthesis?	Result
5	<i>Alexandrium minutum</i>	1022 Rance	under investigation	double peak
	<i>Alexandrium minutum</i>	771 Penzé	under investigation	double peak
	<i>Alexandrium tamarense</i>	ATNWB01	yes	single peak
	<i>Alexandrium tamarense</i>	ATCJ33	no	single peak
	<i>Alexandrium tamarense</i>	ATEB01	no	single peak
10	<i>Alexandrium catenella</i>	ACTRA	yes	single peak
	<i>Alexandrium catenella</i>	ACCC01	yes	single peak
	<i>Alexandrium andersonii</i>	CCMP2222	no	*unclear
	<i>Alexandrium insuetum</i>	CCMP2082	no	*unclear
	<i>Alexandrium affine</i>	CCMP112	no	*unclear
15	<i>Alexandrium affine</i>	PA8V	no	no amplification
	<i>Alexandrium minutum</i>	CCMP113	yes	single peak
	<i>Alexandrium minutum</i>	AL24V	yes	single peak
	<i>Alexandrium minutum</i>	Min3	yes	single peak
	<i>Alexandrium minutum</i>	VGO650	no	double peak
20	<i>Alexandrium minutum</i>	VGO651	no	double peak
	<i>Alexandrium minutum</i>	AL10C	yes	single peak
	<i>Alexandrium catenella</i>	CCMP1493	yes	single peak
	<i>Alexandrium fundyense</i>	CCMP1719	yes	single peak
	<i>Alexandrium fundyense</i>	MDQ1096	yes	single peak

25

Sequences for saxitoxin-producing strains were generated as follows:

Alexandrium minutum CCMP113_sxtA4_gDNA_consensus (SEQ ID NO: 230)

Alexandrium minutum CCMP113_sxtA4_cDNA_consensus (SEQ ID NO: 231)

Alexandrium minutum AL24V_sxtA4_gDNA_consensus (SEQ ID NO: 232)

30 *Alexandrium minutum* AL24V_sxtA4_cDNA_consensus (SEQ ID NO: 233)

Alexandrium minutum Min3_sxtA4_gDNA_consensus (SEQ ID NO: 234)

Alexandrium minutum Min3_sxtA4_cDNA_consensus (SEQ ID NO: 235)

Alexandrium minutum VGO650_sxtA4_gDNA_consensus (SEQ ID NO: 236)

Alexandrium minutum VGO651_sxtA4_gDNA_consensus (SEQ ID NO: 237)

Alexandrium minutum VGO651_sxtA4_cDNA_consensus (SEQ ID NO: 238)

Alexandrium minutum AL10C_sxtA4_cDNA_consensus (SEQ ID NO: 239)

Alexandrium catenella CCMP1493_sxtA4_cDNA_consensus (SEQ ID NO: 240)

Alexandrium fundyense CCMP1719_sxtA4_cDNA_consensus (SEQ ID NO: 241)

5 Alexandrium fundyense MDQ1096_sxtA4_cDNA_consensus (SEQ ID NO: 242)

The primers utilised detected sxtA in STX producing species and can also discriminate between toxic and non-toxic strains of *Alexandrium minutum*. Non-toxic *A. minutum* have two different sxtA copies in their genome, which results in a bimodal melting curve, when a Sybr Green assay is used (see Fig. 17: toxic *A. minutum* AL1V, 10 two curves; non-STX *A. minutum* VGO651, one curve). Other species also have characteristic melt-curves, but discrimination between toxic and non-toxic species based on the meltcurve only works for *A. minutum*.

The primers sxt072 and sxt073 were also demonstrated to work well with the Universal Probe Library probe from Roche, #142 (data not shown). This assay is very 15 specific and useful for detection of sxtA in environmental samples.

Example 7: sxtA1 and sxtA4 are absent in dinoflagellate strains that do not produce saxitoxins

20

Materials and Methods

sxtA1 and sxtA4 PCRs were performed as described in Example 1 above.

Results

25 sxtA (1/4) was not detected in any of the dinoflagellate strains listed below:

Species/Taxon	sxtA (1/4)
<i>Adenoides eludens</i> CCMP1891	n.d.
<i>Alexandrium insuetum</i> CCMP2082	n.d.
<i>Amphidinium carteri</i> UIO081	n.d.
<i>Amphidinium mootonorum</i> CAWD161	n.d.
<i>Azadinium spinosum</i> RCC2538	n.d.

<i>Ceratium longipes</i> CCMP1770	n.d.
<i>Coolia monotis</i>	n.d.
<i>Gambierdiscus australes</i> CAWD148	n.d.
<i>Gymnodinium aureolum</i> SCCAP K-1561	n.d.
<i>Heterocapsa triquetra</i> RCC2540	n.d.
<i>Karlodinium veneficum</i> RCC2539	n.d.
<i>Lepidodinium chlorophorum</i> RCC2537	n.d.
<i>Lingulodinium polyedrum</i> CCMP1931	n.d.
<i>Pentaparsodinium dalei</i> SCCAP K-1100	n.d.
<i>Polarella glacialis</i> CCMP2088	n.d.
<i>Prorocentrum micans</i> UIO292	n.d.
<i>Prorocentrum minimum</i> UIO085	n.d.
<i>Protoceratium reticulatum</i>	n.d.
<i>Pyrocystis noctiluca</i> CCMP732	n.d.
<i>Scrippsiella trochoideae</i> BS-46	n.d.
<i>Thecadinium kofoidii</i> SCCAP K-1504	n.d.
n.d. not detected	

CLAIMS:

1. A method for detecting a saxitoxin-producing dinoflagellate in a sample, the method comprising:
 - obtaining a sample for use in the method; and
 - analyzing the sample for the presence of one or more of a dinoflagellate saxitoxin A catalytic domain polynucleotide or a polypeptide encoded by said polynucleotide;wherein:
 - the saxitoxin catalytic domain is a saxitoxin A4 catalytic domain; and
 - the presence of said polynucleotide or polypeptide indicates the presence of a saxitoxin-producing dinoflagellate in the sample.
2. A method for determining the absence of a saxitoxin-producing dinoflagellate in a sample, the method comprising:
 - obtaining a sample for use in the method; and
 - analyzing the sample for the presence of one or more of a dinoflagellate saxitoxin A catalytic domain polynucleotide or a polypeptide encoded by said polynucleotide;wherein:
 - the saxitoxin catalytic domain is a saxitoxin A4 catalytic domain; and
 - the absence of said polynucleotide or polypeptide indicates the absence of a saxitoxin-producing dinoflagellate in the sample.
3. The method according to claim 1 or claim 2, wherein the saxitoxin A4 catalytic domain sequence comprises nucleotides 3115-4121 of the polynucleotide sequence set forth in SEQ ID NO: 3, nucleotides 3597-3721 of the polynucleotide sequence set forth in SEQ ID NO: 3.
4. The method according to claim 1 or claim 2, wherein said analyzing comprises amplification of polynucleotides from the sample by polymerase chain reaction.
5. The method according to claim 4, wherein said polymerase chain reaction utilises one or more primers consisting of a sequence set forth in any one of SEQ ID

NOs: 198-199.

6. The method according to claim 1 or claim 2, wherein the polypeptide comprises a saxitoxin A4 catalytic domain amino acid sequence as set forth in residues 947-1281 of SEQ ID NO: 4.

7. The method according to claim 1 or claim 2, wherein the saxitoxin-producing dinoflagellate is from the *Alexandrium*, *Pyrodinium* or *Gymnodinium* genus.

8. The method according to claim 7, wherein the saxitoxin-producing dinoflagellate is selected from the group consisting of *A. catenella*, *A. fundyense*, *A. lusitanicum*, *A. minutum*, *A. ostenfeldii*, *A. tamarense*, *G. catenatum* and *P. bahamense* var *compressum*.

9. A kit for the detection of a saxitoxin-producing dinoflagellate in a sample, the kit comprising at least one agent for detecting the presence of a dinoflagellate saxitoxin A catalytic domain polynucleotide or a polypeptide encoded by said polynucleotide, wherein the saxitoxin catalytic domain is a saxitoxin A4 catalytic domain.

10. A kit for the determining the absence of a saxitoxin-producing dinoflagellate in a sample, the kit comprising at least one agent for detecting the presence of a dinoflagellate saxitoxin A catalytic domain polynucleotide or a polypeptide encoded by said polynucleotide, wherein the saxitoxin catalytic domain is a saxitoxin A4 catalytic domain.

11. The kit according to claim 9 or claim 10, wherein the saxitoxin A4 catalytic domain sequence comprises nucleotides 3115-4121 of the polynucleotide sequence set forth in SEQ ID NO: 3, nucleotides 3597-3721 of the polynucleotide sequence set forth in SEQ ID NO: 3.

12. The kit according to claim 9 or claim 10, wherein the agent is a primer, probe or antibody.

13. The kit according to claim 9 or claim 10, wherein the agent is a primer consisting of a sequence set forth in any one of SEQ ID NOs: 198-199.

14. The kit according to claim 9 or claim 10, wherein the agent binds specifically to a saxitoxin A4 catalytic domain amino acid sequence as set forth in residues 847-1281 of SEQ ID NO: 4.

15. A primer pair for detecting a saxitoxin-producing dinoflagellate in a sample, wherein said primer pair consists of:

a first primer consisting of the polynucleotide sequence of SEQ ID NO: 198 and a second primer consisting of the polynucleotide sequence of SEQ ID NO: 199.