Toxic Phytoplankton Blooms in the Sea
T.J. Smayda and Y. Shimizu, editors

© 1993 Elsevier Science Publishers B.V. All rights reserved.

POPULATION ANALYSIS OF TOXIC AND NONTOXIC *Alexandrium* SPECIES USING RIBOSOMAL RNA SIGNATURE SEQUENCES

CHRISTOPHER A. SCHOLIN AND DONALD M. ANDERSON
Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA. 02543 U.S.A.

ABSTRACT

Sequence analysis was used to determine if small-subunit (Ss) and large-subunit (Ls) ribosomal RNA (rRNA) genes (rDNA) can provide distinct molecular markers for *Alexandrium* species. Analysis of the entire SsrDNA for one North American *A. fundyense* culture unexpectedly revealed the presence of two distinct genes, one of which we believe to be a pseudogene. An assay developed to detect these two sequences in PCR-amplified material was used to screen 27 globally-distributed, toxic and nontoxic *Alexandrium* species. Results indicate that the putative pseudogene occurs only in toxic North American isolates, irrespective of their species designation, but does not occur in other toxic and nontoxic forms from other regions of the world. Analysis of a portion of the LsrDNA has revealed hypervariable domains which are likely to provide highly specific signature sequences useful in delineating genetically similar populations.

INTRODUCTION

Detailed examination of cell morphology, enzyme electrophoretic studies, and comparisons of toxin composition have all been used to assess the relatedness of *Alexandrium* isolates collected from regional and globally-distributed populations [1-4]. These markers represent complex character states which are dependent on the coordinated expression of multiple genes. Consequently, the resolution of population boundaries has remained relatively coarse.

An alternative method of assessing the relatedness of isolates relies on sequence analysis of ribosomal RNA genes (rDNA) and their gene products (rRNA). rDNA sequences have been used extensively as phylogenetic and taxonomic indicators for a variety of organisms [5-7]. Ribosomal RNA genes are composed of "domains" which are both highly conserved and highly variable among all organisms [8,9]. The conserved and variable elements are valuable for both broad- and fine-scale taxonomic and phylogenetic comparisons, respectively. Analysis of rDNA sequences is not dependent on the physiological state of the organism nor the concomitant expression of other genes. Some of the fastest evolving rDNA domains may also be useful as population- or even strain-specific markers [10,11; Herzog, pers. com.]. In an

attempt to refine our concept of toxic *Alexandrium* population boundaries and provide distinct molecular markers for a variety of *Alexandrium* species, we have undertaken the sequence analysis of both small-subunit (Ss) and large-subunit (Ls) rDNA. This report summarizes our progress to date.

MATERIALS AND METHODS

Cultures used in this study are listed in Table 1. All were maintained in f/2 medium as modified and described by Anderson et.al. [12]. For nucleic acid extractions, 10 - 15 mL of a mid-log culture were harvested by gentle centrifugation, and the pellet resuspended in 217 µL of autoclaved milli-Q water at room temperature. The cell slurry was transferred to a 1.5 mL sterile centrifuge tube and adjusted to contain: 1% SDS, 10mM EDTA (8.0), 10mM Tris HCl (7.5) and 10mM NaCl. Nucleic acids in this solution were extracted, precipitated, resuspended in 10-50 µL of TE pH 7.5 and quantified by standard methods [13].

All PCR amplifications were carried out using Perkin Elmer Cetus DNA Thermal Cycler and GeneAmp PCR Core Reagents as recommended by the maufacturers. The entire SsrDNA was amplified with universal eukaryotic primers [14]. A portion of the LsDNA was amplified using primers targeted towards conserved elements at positions 25-45 ("forward," or "D1R") and 733-714 ("reverse," or "D2C"), relative to the *Prorocentrum micans* LsrRNA [15]. The sequences of these primers are: D1R- 5'ACCCGCTGAATTTAAGCATA3'; and, D2C- 5'CCTTGGTCCGTGTTTCAAGA3' [Herzog, unpublished]. Amplifications were typically carried out in triplicate and subsequently pooled prior to purification [13] and cloning of the products. 50-100 ng of SsrDNAs were digested with the following enzymes (New England Biolabs): BsaAI, BsrI or XbaI in 10-25 µL reactions. Products of the digestions were resolved on a 1.5% agarose gel.

The entire SsrDNA was cloned into mp18 and mp19 M13 phage [14]. LsrDNA fragments were cloned using Invitrogen's T/A cloning kit according to the recommendations of the manufacturer. Details of screening recombinant colonies and isolation of plasmid DNA will be described elsewhere. All sequencing reactions were carried out using United States Biochemical Corp. Sequenase (v. 2.0) sequencing kit and Amersham dATP [α^{35}S] label. Both strands of the SsrDNA were sequenced using 22 mp18 ("forward") and 24 mp19 ("reverse") M13 pooled clones [14; USB/sequenase protocol]. For LsrDNAs, 10 clones for each isolate were typically pooled prior to sequencing. The LsrDNA and direct SsrRNA sequencing protocols will be published elsewhere.

RESULTS AND DISCUSSION

Identification of Two SsrDNA Genes in a Clonal *A. fundyense* culture

Pilot sequencing efforts directed towards the analysis of full-length SsrDNAs from a clonal *A. fundyense* culture (CA29; Table 1) unexpectedly

TABLE I. Strain numbers, species designations, and isolation locales of *Alexandrium* cultures screened for A and B SsrDNA genes, and used in preliminary LsrDNA sequence analysis.

Geographic Block		Strain	Spp. Designation	Isolation Locale	Toxic?	Larger SsrDNA Products[1]	Restriction Sites[2] A	B	Other[3]
N. America	W. Coast	PW06	*A. tamarense*	Port Benny, Alaska	yes		+	+	
		BGt 1	*A. catenella*	Russian River, CA	yes		+	+	
	E. Coast	AFNFA3	*A. fundyense*	Newfoundland	yes	+	+	+	
		AFNFA4	*A. fundyense*	Newfoundland	yes	+	+	+	
		Gony.# 7	*A. fundyense*	Bay of Fundy	yes	+	+	+	
		Gt 429	*A. fundyense*	Ipswich Bay, MA	yes	+	+	+	
		Gt CA29	*A. fundyense*	Cape Ann, MA	yes	+	+	+	
		GtMP	*A. fundyense*	Orleans, MA	yes	+	+	+	
		GtPP01	*A. tamarense*	Falmouth, MA	yes	+	+	+	
		GtPP06	*A. tamaense*	Falmouth, MA	yes	+	+	+	
		Gt CN16	*A. tamarense*	Groton, CN	yes	+	+	+	
		Gt LI21	*A. tamarense*	Babylon, NY	yes		+	+	
		TC02	*A. andersoni*	Eastham, MA	no		+		+
W. Europe	U.K.	Pgt183	*A. tamarense*	Plymouth,U.K.	no		+		
	Spain	PE1V	*A. tamarense*	Galicia, Spain	no		+		
		PE2V	*A. tamarense*	Galicia, Spain	?		+		
		PA5V	*A. affine*	Galicia, Spain	no		+		
	Portugal	GtPort	*A. lusitanicum*	Portugal	yes		+		+
S. Pacific	Australia (mainland)	ACPP01	*A. catenella*	Port Phillip Bay, Australia	yes		+		
		ACPP02	*A. catenella*	Port Phillip Bay, Australia	yes		+		
		ACPP03	*A. catenella*	Port Phillip Bay, Australia	yes		+		
		ACPP09	*A. catenella*	Port Phillip Bay, Australia	yes		+		
		AMAD01	*A. minutum*	Port River, Australia	yes		+		+
		AMAD06	*A. minutum*	Port River, Australia	yes		+		+
	Australia (Tasmania)	ATBB01	*A. tamarense*	Bell Bay, Tasmania	no		+		
		AABB01/2	*A. affine*	Bell Bay, Tasmania	no		+		
ballast water		ACJP03	*A. catenella*	Kashima, Japan[4]	yes	+	+		
		I72/22 #2	*A. tamarense*	Muroran, Japan[4]	yes		+	+	

1) indicates presence of PCR products larger than expected (Fig. 1b)
2) presence of restriction sites characteristic of A and B genes (Fig. 1a)
3) "Other:" digestion with BsaAI resulted in atypical products, but showed no evidence for multiple SsrDNAs
4) presumed origin [Hallegraeff, personal communication]

revealed both length and sequence heterogeneities in the PCR-amplified, pooled clones, and suggested substantial variation within the SsrDNAs from this isolate. Four mp18 (forward) and four mp19 (reverse) M13 clones were then sequenced individually over most of the length of the molecule. This led to the discovery of two distinct sequences, termed the "A gene" and "B gene," present in PCR-amplified material from the *A. fundyense* culture in roughly a 1:1 ratio. These two genes differ at approximately 40 positions scattered throughout the length of the molecule. When the two sequences were placed into a data base of aligned eukaryotic and prokaryotic SsrRNAs, the variance exhibited by the B gene violated many highly- to universally-conserved sequence elements [Sogin, pers. comm.]. In contrast, the A gene did not vary from any of the canonical sequence motifs. This led to our current hypothesis that the B gene is a pseudogene (ie., is mutated in such a way that it is no longer functional) that only recently diverged from its A gene counterpart given their high degree of homology.

Direct sequencing of SsrRNAs from this culture under nutrient replete and nitrogen-limited conditions, as well as over a complete circadian cycle, failed to demonstrate the presence of a stable B gene transcript. The A gene transcript, however, was clearly present at all times and under all culture conditions. The lack of a stable B gene transcript further supports the hypothesis that the B gene is a pseudogene. Additional experiments will be required to determine if the B gene is transcribed and stable in vivo.

Rapid Detection of the A/B Gene Pair

The presence of the B gene in an eastern North American *A. fundyense* and the possibility of it having only recently appeared evolutionarily suggested that it might serve as a taxonomic and/or biogeographic marker. A restriction assay capable of detecting the A and B genes in PCR-amplified material was thus developed, based upon the sequence differences between the two genes. This variation conveniently creates or eliminates restriction sites in the A and B genes (Fig. 1a). This assay, termed the "A/B restriction test," allowed rapid detection of either gene's presence in PCR-amplified material (Figs. 1b-e).

To date, this test has been used to screen an additional 27 *Alexandrium* isolates with a variety of species designations and from diverse regions of the world (Table 1). Results, which will be described in detail elsewhere, can be summarized as follows: 1) the B gene was detected in isolates of *A. tamarense*, *A. fundyense* and *A. catenella*, but not in *A. minutum*, *A. lusitanicum*, *A. affine*, or *A. andersoni;* 2) of the 11 *A. tamarense*, *A. fundyense* and *A. catenella* cultures found to harbor the B gene, 10 were from North America and one from the ballast water of a Japanese cargo vessel; 3) the A and B genes occur at an approximate 1:1 ratio in PCR-amplified material from those isolates which carry the B sequence; 4) there is no correlation between an isolate's ability to produce PSP toxins and the presence of the B gene; 5) the B gene is not universally distributed, consistent with the hypothesis that it is a pseudogene; and, finally, 6) the geographic distribution of the B gene suggests that it may be a useful marker for toxic *Alexandrium* with a North American origin, though more globally-distributed isolates will need to be examined to rigourously test

this possibility. It should also be noted that the A/B restriction test effectively samples only a small number of nucleotides within the SsrDNA, and thus other *Alexandrium* isolates could carry a similar second gene which is not detected by the enzymes used in this assay.

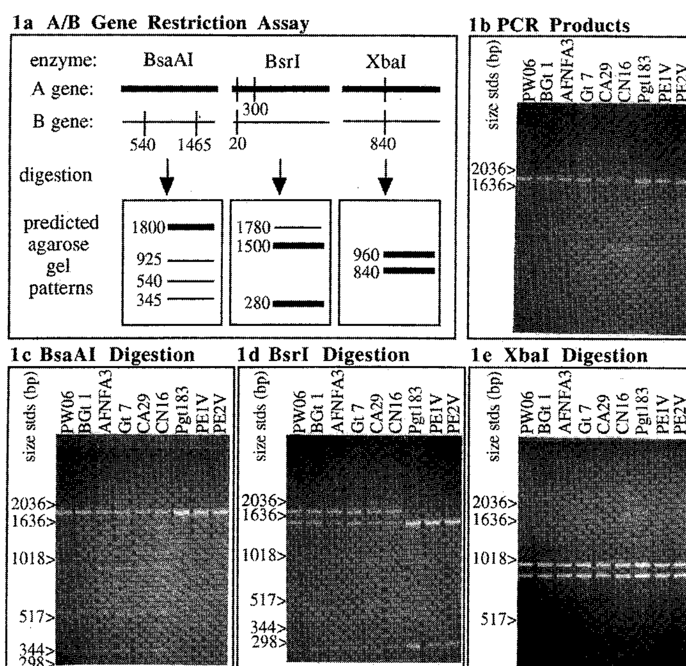

FIG. 1. A/B gene restriction test. 1a: schematic representation of the assay. Top: restriction maps of A and B genes for BsaAI (B cutter), BsrI (A cutter) and XbaI (A/B cutter; numbers refer to cut sites); bottom: schematized agarose gel resolution of corresponding digestion products [numbers refer to length of the fragments in base pairs (bp)]. 1b-e: agarose gels of PCR-amplified products from noted strains (1b) and corresponding fragments produced by BsaAI (1c), BsrI (1d) and XbaI (1e) digestion (see also Table 1).

Two peculiarities of the SsrDNA amplified products were noted and deserve special mention. First, all samples containing both A and B genes display minor BsaAI digestion products which do not conform to the size classes

predicted by a computer program analysis of the two sequences (see Fis. 1a, c). These products are the result of chimeric molecules which are either artifacts generated by PCR template strand switching [16], or true representatives of yet another class of SsrDNA that exists in vivo. Second, amplified SsrDNAs from all toxic, eastern North American *Alexandrium* contain larger PCR products in addition to the predicted major product (Fig. 1b; Table 1). It is unlikely that contaminants are the the cause of these differences, since all cultures used in this study were maintained with the same stocks of media. We believe the larger SsrDNA amplification products are of *Alexandrium* origin and probably due to insertions within, or rearrangements of, the SsrDNA tandem repeats. These insertions or rearrangements are quite possibly unique to eastern North American populations of toxic *Alexandrium* and may serve as yet another independent, fine-scale biogeographic marker. Interestingly, the *A. tamarense* isolated from the ballast water of a Japanese freighter (I72/22#2), which was found to carry the B gene, also displays the larger SsrDNA PCR products.

<u>SsrRNA and LsrDNA Sequence Comparisons</u>

In our efforts to detect B gene expression, we examined cultures from both the east and west coasts of North America. Direct sequencing of a portion of the SsrRNA isolated from these cultures, spanning 2 of the most highly variable domains within this molecule [V3 and V4, 8; Sogin, pers. comm.], revealed little to no variation between any of the isolates (data not shown). This suggests that actively-expressed, functional *Alexandrium* SsrDNA may evolve too slowly to be useful as a population-specific marker. On the other hand, the B gene, or other unexpressed/rearranged SsrDNAs, may be extremely useful in this context as they should be evolving at an accelerated rate. However, detailed sequence analysis of these probable pseudogenes is labor intensive and difficult given the high background of conserved, actively-expressed genes. As a consequence, we have focussed our attention on a portion of the LsrDNA [D1 and D2 domains] thought to contain some of the most highly varible ribosomal sequences [9; Herzog, pers. comm.].

We recently succeeded in PCR-amplifying, cloning and sequencing an approximately 700bp fragment, roughly spanning positions 25-733 relative to the *P. micans* LsrRNA [15] for the 28 isolates listed in Table 1. Analysis of these data is incomplete, yet several important trends have emerged and warrant discussion. First, the LsrDNA fragments can exhibit both length and sequence heterogeneities, suggesting substantial sequence variation exists between different copies of these genes within particular, clonal *Alexandrium* cultures. This variation, however, appears conserved among isolates from the same geographic population. As an example, we find a particular pattern of sequence ambiguities and length heterogeneities which are identical in all eastern North American *A. tamarense* and *A. fundyense*. Significantly, the *A. tamarense* ballast water isolate (I72/22#2) which harbored the B gene and gave rise to the larger SsrDNA amplification products, also carries the predominant eastern North American LsrDNA sequence morph. We are now considering the possibility that eastern North American, toxic *A. tamarense* may have been

introduced to Japan in the recent past, and subsequently transported to other regions of the world.

Distinct populations of *Alexandrium* are also evident on a larger biogeographic scale. For example, within North America, west coast isolates appear distinguishable from their east coast counterparts using LsrDNA sequences. Furthermore, all Australian *A catenella* can be easily distinguished from North American *A. catenella*, *A. tamarense* and *A. fundyense*. When comparing very different species, such as *A. catenella* and *A. minutum*, even greater differences are seen. Cross-genera comparisons yield the most dramatic demonstration of sequence variation [9,17].

Thus, the most important, yet still preliminary conclusion, from the LsrDNA analysis is that there appears to be a gradation of sequence specificity from regional populations to globally-distributed populations to species to genera. Each level of specificity carries a particular "signature." Regional populations are the most similar, while greater taxonomic separation is evidenced by more significant sequence divergence. This is precisely what one would predict if, in fact, the LsrDNA sequence is evolving at a fast rate. We believe that this sequence will therefore be useful as an independent, fine-scale taxonomic and biogeographic marker for *Alexandrium* species. Furthermore, the combined use of Ss and LsrDNA signatures may prove useful in addressing controversies surrounding the apparent spread of *Alexandrium* species throughout the world. One alternative explanation to an actual introduction of a species to a region via ballast water discharge [18] or unusual hydrographic events [1,19] is that an endogenous population's growth is enhanced, perhaps by pollution or eutrophication. A truly endogenous, sexually isolated population will likely contain sequence morphs that are unique relative to other populations of the same species found elsewhere. An introduced species, on the other hand, should carry signatures indicative of the population from which it has spread.

ACKNOWLEDGEMENTS

We greatfully acknowledge the advice and assistance of E. Delong, M. Herzog, D. Kulis, D. Prasher and M. Sogin. The following individuals kindly provided cultures: I. Bravo, S. Hall, G. Hallegraeff, and A. White. This work was supported in part by a grant from the National Science Foundation (OCE 8911226) and the Ocean Ventures Fund (WHOI). Contribution No. 8113 from the Woods Hole Oceangraphic Institution.

REFERENCES

1. B. A. Hayhome, D. M. Anderson, D. M. Kulis and D. J. Whitten, Mar. Biol. 101, 427-435 (1989).
2. L. Maranda, D. M. Anderson and Y. Shimizu, Esturine, Coastal and Shelf Sci 21, 401-410 (1985).
3. A. D. Cembella and F. J. R. Taylor, Biochem.Sys. and Ecol. 14, 311-323 (1986).
4. A. D. Cembella, J. J. Sullivan, G. Boyer, F. J. R. Taylor and R. J. Andersen, Biochem. Sys. and Ecol. 15, 171-186 (1987).

5. G. J. Olsen, D. J. Lane, S. J. Giovannoni, N. R. Pace and D. A. Stahl, Ann.Rev.Microbiol. 40, 337-365 (1986).
6. K. G. Field, G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace and R. A. Raff, Sci. 239, 748-753 (1988).
7. M. L. Sogin, H. J. Elwood and J. H. Gunderson, Proc. Nat. Acad. Sci. 83, 1383-1387 (1986).
8. M. L. Sogin and J. H. Gunderson, Ann. N.Y. Acad. Sci 503, 125-139 (1987).
9. G. Lenaers, C. A. Scholin, Y. Bhaud, D. Saint-Hilaire and M. Herzog, J. Mol. Evol. 32, 53-63 (1991).
10. U. B. Gobel, A. Geiser and E. J. Stanbridge, J. Gen. Microbiol. 133, 1969-1974 (1987).
11. R. I. Amann, L. Krumholz and D. A. Stahl, J.Bact. 172, 762-770 (1990).
12. D. M. Anderson, D. M. Kulis and B. J. Binder, J.Phycol. 20, 418-425 (1984).
13. in: Current Protocols in Molecular Biology, F. M. Ausubel et al., eds. (Greene Publishing and Wiley-Interscience, N.Y., 1989) Vols. 1 and 2.
14. L. Medlin, H. J. Elwood, S. Stickel and M. L. Sogin, Gene 71, 491-499 (1988).
15. G. Lenaers, L. Marteaux, B. Michot and M. Herzog, J. Mol. Evol. 29, 40-51 (1989).
16. H. A. Erlich, D. Gelfand, J. J. Sninsky, Sci. 252, 1643-1651 (1991).
17. G. Lenaers, H. Nielsen, J. Engberg and M. Herzog, BioSystems 21, 215-222 (1988).
18. G. Hallegraeff and C. J. Bolch, Mar. Poll. Bull. 22, 27-30 (1991).
19. D. M. Anderson in: Red Tides: Biology, Environmental Science, and Toxicology, T. Okaichi, D. M. Anderson and T. Nemoto, eds. (Elsevier, N.Y., 1989) pp. 11-16.