LIFE CYCLE STUDIES OF THE RED TIDE DINOFLAGELLATE SPECIES COMPLEX *ALEXANDRIUM TAMARENSE*
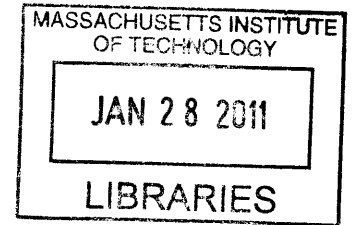
by

Michael L. Brosnahan

A.B., Dartmouth College, 2002

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

February 2011

Signature of Author _____

Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology and Woods Hole Oceanographic Institution
January 27, 2011

Certified by _____

Donald M. Anderson
Thesis Supervisor

Accepted by _____

Simon Thorrold
Chair, Joint Committee for Biological Oceanography
Woods Hole Oceanographic Institution

1

For Mom and Dad. Thank you.

TABLE OF CONTENTS

Chapter 1: Outbreeding lethality between toxic Group I and nontoxic Group III
Alexandrium tamarense spp. isolates: predominance of heterotypic encystment and
implications for mating interactions and biogeography

Chapter 2: Application of imaging flow cytometry for species-specific measurement of DNA content in dinoflagellate microplankton

Chapter 3: *de novo* assembly of four *Alexandrium tamarense* transcriptomes

ACKNOWLEDGEMENTS

Beach dinners and ice cream; to Sandy for love, support and for sticking with me through one of the hardest years of my life; to Mom for love, support and believing in me through every year of my life; and to Dad for love, support and fishing memories I'll carry for the rest of my life. I love you all.

LIFE CYCLE STUDIES OF THE RED TIDE DINOFLAGELLATE SPECIES COMPLEX *ALEXANDRIUM TAMARENSE*

by

Michael L. Brosnahan

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

ABSTRACT

Blooms of toxic species within the algal dinoflagellate species complex *Alexandrium tamarense* may cause Paralytic Shellfish Poisoning, a significant and growing environmental threat worldwide. However, blooms of closely related nontoxic *A. tamarense* also occur, sometimes in close geographical proximity to toxic blooms. This thesis explores the interactions between toxic and nontoxic blooms by examining sexual crosses between each of five ribosomal clades within the *A. tamarense* complex (termed Groups I-V). Several lines of evidence argue that these clades represent separate species. Particular emphasis was given to interactions between toxic Group I and nontoxic Group III species because they are among the most closely related *A. tamarense* clades and because they share a natural range boundary in several parts of the world.

Interspecies hybridization appeared widespread between different clades and between geographically dispersed isolates. However, subsequent germination studies of hypnozygotes produced from combinations of Group I and Group III clones failed to yield new vegetative cultures in multiple trials. The possibility that these hypnozygotes were actually inbred (i.e. the result of pairs of only Group I or only Group III gametes) was considered and rejected using a nested PCR assay that was developed to assess the parentage of individual cysts. The assay was also suitable for analyzing cysts collected from the field and was applied to individual cysts collected from Belfast Lough, an area where both Group I and Group III blooms were known to occur. Two Group I/Group III hybrids were detected in fourteen successful assays from the Belfast sample, showing that hybridization does occur in nature.

These findings have several important implications. First, the failure of Group I/Group III hybrids to produce new vegetative cultures serves as a proof that the *A. tamarense* clades represent cryptic species because they are unable to produce genetic intermediates. Second, the presence of hybrid cysts in Belfast Lough indicates ongoing displacement of a nontoxic population by a toxic one (or vice versa) in that region. Third, the inviability of toxic/nontoxic hybrids suggests a remediation strategy whereby the recurrence of toxic *A. tamarense* blooms might be combated through the introduction of nontoxic cells.

The results from these experiments also highlighted several shortfalls in our understanding of the mechanisms governing sexual compatibility between clones and also our ability to replicate these organisms' sexual cycle in the laboratory. Two initiatives were begun with the ultimate goal of better characterizing sexual processes in natural populations.

The first initiative was the application of an imaging flow cytometer to detect sexual events in natural blooms. An existing instrument, the Imaging FlowCytobot, was adapted to positively identify *A. tamarense* Group I cells in mixed species assemblages and measure cell DNA content. A collection of four samples were analyzed, three from the development and decline of a local Group I bloom and one from a Group I red tide that occurred near Portsmouth, NH and led to a major deposition of new cysts in the southern Gulf of Maine. Several unanticipated patterns were revealed including the discovery of a persistent layer of cells with 2c DNA content near the surface and disproportionately high rates of infection by an *Amoebophrya* sp. parasite in large *A. tamarense* planozygotes.

The second initiative was the application of high throughput Illumina sequencing to define the transcriptomes of toxic Group I, nontoxic Group III, and toxic Group IV clones. Additional steps were taken to isolate RNA and prepare a cDNA library from a natural sample of Group I hypnozygotes. The applications for these data are expected to be extensive and include the discovery of sexual biomarkers and further characterization of the differences between toxic and nontoxic *A. tamarense* species. Preliminary results from the sequencing of these libraries and their initial assembly are described.

Introduction

In recent decades, the apparent incidence and severity of blooms of harmful and toxic

algae has increased worldwide (Anderson 1989; Hallegraeff 1993). Many hypotheses

have been proposed to explain this phenomenon including changing patterns in our use of

coastal waters, increases in nutrient pollution, and also the spread of the scientific

expertise needed to detect these organisms. Curiously, as our scientific knowledge of

their ecology has grown, our ability to control and mitigate the impacts of these harmful

blooms has not (Anderson 1997). In this thesis, I describe new findings concerning the

mating behavior of *Alexandrium tamarense*, a globally distributed group of algal

dinoflagellates, some of which cause Paralytic Shellfish Poisoning (PSP). These findings

have implications for the global spread of toxic members within this group and also

suggest a new strategy for remediating areas contaminated by toxic *A. tamarense*. At the

same time, this work draws attention to gaps in our understanding of these species' basic

physiology and behavior, and describes ongoing work to develop methods that will

address these shortfalls.

     In the sections that follow, I will briefly describe the *A. tamarense* species

complex and its position within the genus *Alexandrium*. I will then review what is known

of the *A. tamarense* mating system and the importance of its sexual life cycle to its

persistence and dispersal to new areas. Lastly, I will give an overview of the original

work presented in the data chapters that follow.

BACKGROUND

The Alexandrium tamarense species complex

The dinoflagellate genus *Alexandrium* is comprised of more than thirty algal

species. Most are approximately spherical in their vegetative forms and range in diameter

from 20 to 60 μm. All species within the genus contain chloroplasts and can grow

photoautotrophically, but some may also supplement their nutrition through the intake of

other cells or large macromolecules. Food vacuoles containing small ciliates and

phytoplankton have been observed within *A. ostenfeldii* (Jacobson and Anderson 1996),

and *A. tamarense* cells have the ability to intake high molecular weight organic molecules

(John and Flynn 1999). As dinoflagellates, these species possess a number of unique

nuclear characteristics including exceptionally large genomes, abundant and permanently

condensed chromosomes, and a lack of typical eukaryotic nucleosomes (Spector 1984;

Hackett, Anderson et al. 2004).

Several *Alexandrium* species have also been shown to produce PSP toxins,

bioluminesce, and/or migrate vertically in the water column (Tomas 1997). Cell

locomotion is accomplished through the action of two flagella: a transverse flagellum that

lies within a furrow that encircles the cell body, and a longitudinal or trailing flagellum

that extends from the posterior portion of the cell. Their swimming behavior may enable

these cells to access high nutrient concentrations many meters from the surface at night

and return to euphotic waters during daylight hours (Anderson and Stolzenbach 1985;

MacIntyre, Cullen et al. 1997; Townsend, Bennett et al. 2005).

Though sometimes referred to as red tides, *Alexandrium* blooms only rarely cause

water discoloration and most often constitute only a small fraction of the total

phytoplankton community at any given time and location (Anderson 1997). Still, the

presence of toxic *Alexandrium* species can pose a significant health hazard. Saxitoxin and

its congeners, the causative molecules of PSP, are among the most potent natural toxins

known and concentrations of only a few hundred *Alexandrium* cells L^{-1} can rapidly

contaminate filter-feeding shellfish beyond safe limits for human consumption (Hall,

Strichartz et al. 1990). The ability to synthesize these toxins is shared with two other

dinoflagellate genera, *Pyrodinium* and *Gymnodinium*, and a handful of cyanobacteria.

The direct advantage it affords to these species is unknown, and the ability to synthesize

toxin has been gained and/or lost many times throughout the evolutionary history of these

groups. In the *Alexandrium* genus, production of saxitoxin or its congeners is restricted to

six species: *A. minutum*, *A. ostenfeldii*, *A. tamiyavanichii*, *A. tropicale* and two members

of the *A. tamarense* species complex (Cembella 1998; Lilly 2003; Hackett, Wisecaver et

al. in prep).

The *Alexandrium tamarense* species complex was first defined by the close

association of three morphotypes: *A. catenella* (Whedon et Kof.) Balech, *A. fundyense*

Balech, and *A. tamarense* (M. Lebour) (Taylor 1984). The morphotypes were long

considered to be valid species, but several phylogenetic analyses have shown that the

morphotypes are not monophyletic, invalidating the use of the morphotype names as

species descriptors (Scholin, Hallegraeff et al. 1995; Hingman, Stone et al. 2001; John,

Fensome et al. 2003; MacKenzie, de Salas et al. 2004; Lilly, Halanych et al. 2007).

Figure 2 from Lilly, E. L., K. M. Halanych, et al. (2007). "Species Boundaries and Global Biogeography of the *Alexandrium tamarense* complex (Dinophyceae)." <u>Journal of Phycology</u> **43**(6): 1329-1338.

"One of two most likely trees returned by maximum-likelihood analysis, score –ln 2756.4879. Strains are labeled with their original morphospecies designation. Strain numbers of toxic strains are indicated in bold type, while those of nontoxic strains are in nonbold type. Gray type is used where toxicity is unknown."

the
A. tamarense
complex

100
ATRU-22 *A. tamarense* Russia
ATRU-21 *A. tamarense* Russia
Alex61-2 *A. tamarense* Scotland
UW4-1 *A. tamarense* Scotland
ATBR2c *A. tamarense* Brazil
ATBR2e *A. tamarense* Brazil
GTMP *A. tamarense* USA
ATRU-29 *A. tamarense* Russia
YDC98c *A. tamarense* South Korea
BSW97 *A. tamarense* South Korea
71
PW06 *A. tamarense* USA
PI32 *A. tamarense* USA
BGt1 *A. fundyense* USA
SSW00G6-3 *A. tamarense* South Korea
AFNFA3.1 *A. fundyense* Canada
GtCA29 *A. catenella* USA
ATRU04 *A. tamarense* Russia
ATBR2g *A. tamarense* Brazil
ATUR0TA *A. tamarense* Uruguay
I72/24#1 *A. tamarense* Japan
61
OF041 *A. tamarense* Japan
OF051 *A. tamarense* Japan
YDC98a *A. tamarense* South Korea
ACC01 *A. catenella* Chile
ACC07 *A. catenella* Chile
ACQH01 *A. catenella* USA
ATBR2d *A. tamarense* Brazil
Alex61-1 *A. tamarense* Scotland
UW4-2 *A. tamarense* Scotland
K-0055 *A. tamarense* Faroe Islands
SJC00a *A. tamarense* South Korea
ULW9903.2 *A. tamarense* South Korea
UL7 *A. tamarense* South Korea
SJW97043 *A. tamarense* South Korea
HAT4 *A. tamarense* Japan
AFNFA3.2 *A. fundyense* Canada
SA2 *A. catenella* South Africa
ATSK03 *A. tamarense* South Korea
ATUR02 *A. tamarense* Uruguay
KCJ97111 *A. tamarense* South Korea
A3 *A. catenella* USA

Group I

69

100

92

100
ATRU-10 *A. tamarense* Russia
SZN08 *A. tamarense* Italy
SZN21 *A. tamarense* Italy
SZNB1 *A. tamarense* Italy
SZNB19 *A. tamarense* Italy

Group II

100

ATSW01-1 *A. tamarense* Sweden
ATFE7 *A. tamarense* England
AlexW12 *A. tamarense* England
ATIR5 *A. tamarense* Ireland
Alex35.2 *A. tamarense* Ireland
PGT183 *A. tamarense* England
AlexW2 *A. tamarense* England
ATFE6 *A. tamarense* England
Ply 173 *A. tamarense* England
Alex31.9 *A. tamarense* Ireland
ATIR3 *A. tamarense* Ireland
SP3B8-3 *A. tamarense* Spain
Alex31.6 *A. tamarense* Ireland
AlexW1 *A. tamarense* England
ATIR4 *A. tamarense* Ireland
ATIR1 *A. tamarense* Ireland
PE2V *A. tamarense* Spain
WKS-1 *A. tamarense* Japan

Group III

100

100

61

920501-C *A. cf. tamarense* China
CMC98b *A. catenella* South Korea
LAC35 *A. fundyense* Australia
YSC98a *A. catenella* South Korea
930205-1 *A. cf. tamarense* China
HK1989 *A. catenella* Hong Kong
HK1998 *A. catenella* Hong Kong
ACPP09 *A. catenella* Australia
ATCI03-1 *A. tamarense* China
ACB0PNZ *A. catenella* New Zealand
ATRU-14 *A. tamarense* Russia
ATRU-11 *A. tamarense* Russia
ATT98.11bis *A. catenella* France
DPC95b *A. catenella* South Korea
ATTL01 *A. catenella* France
ALEXSPHK *A. catenella* Hong Kong
WKS-8 *A. catenella* Japan
ACPP01 *A. catenella* Australia
G.Hope 1 *A. catenella* South Korea
TN9 *A. catenella* Japan
OF101 *A. catenella* Japan
ATCI01-1 *A. tamarense* China
940101-C *A. cf. tamarense* China
K-0270 *A. catenella* Australia
ATPR-12 *A. tamarense* Russia
ACY12 *A. catenella* Japan
ATT98.X11 *A. catenella* France
DPC7 *A. catenella* South Korea
KMC93e *A. catenella* South Korea
CMC2 *A. catenella* South Korea
YSC98I1 *A. catenella* South Korea
SJW0097-8 *A. catenella* South Korea

Group IV

95
65

100

100

ATBB01 *A. tamarense* Australia
Group V

100
100
ATMS01 *A. tamiyavanichii* Malaysia
TAMI22201 *A. tamiyavanichii* Japan
TAMI22612 *A. tamiyavanichii* Japan
TAMI22207 *A. tamiyavanichii* Japan
A. tamiyavanichii

96
CU-15 *A. tropicale* Thailand
CU-13 *A. tropicale* Thailand
CU22 *A. tropicale* Thailand
SA1 *A. tropicale* South Africa
A. tropicale

79
AJB-1 *A. affine* Hong Kong
AABCV-1 *A. affine* Mexico
CU1 *A. affine* Thailand
PA5V *A. affine* Spain
AABB01/2 *A. affine* Australia
66
100
A. affine

0.005 substitutions/site

15

Using a collection of clones from all six temperate continents, Lilly, Halanych et al. (2007) described five distinct ribosomal clades, designated Groups I-V, that strongly correlated with their clones' region of origin. Each of the clades included representatives of the *tamarense* morphotype, two included the *catenella* morphotype, and a single clade (Group I) included the *fundyense* morphotype. Another important finding from these studies was that only two of the five ribosomal clades produce PSP toxins, but every isolate tested within these toxic clades (Groups I and IV) produced saxitoxin and/or its congeners. Both the topology of the ribosomal phylogeny and their geographic separation strongly suggested that the five clades were distinct, cryptic species (Scholin, Hallegraeff et al. 1995; Lilly, Halanych et al. 2007). Experiments described in the first data chapter of this thesis provide strong support for this conclusion. Therefore, I have used the group designations suggested by Lilly, Halanych et al. as they are the only published and valid descriptors of *A. tamarense* species available.

The A. tamarense Life Cycle

A. tamarense blooms are initiated through the germination of hypnozygotes, a durable, diploid cyst stage in their life cycle. The hypnozygotes can be found throughout the water column depending on the season and the hydrographic conditions, but are most abundant on the ocean bottom and usually germinate in early spring. The germling cell that emerges from a hypnozygote cyst is called a planomeiocyte. It is a diploid swimming cell that completes meiosis with two or more divisions (Pfiester 1989). After completing meiosis, vegetative cells are produced that in turn proliferate through mitotic division and

re-form populations in surface waters where they may cause PSP. Vegetative cells remain haploid and it is these cells that are predominant within blooms. Under conditions that remain poorly understood, vegetative cells transform to gametes that then fuse in pairs to form diploid planozygotes, a swimming stage that can remain in the plankton for up to one week before returning to the hypnozygote phase (Pfiester and Anderson 1987).

The entirety of this cycle can be replicated in culture but there are significant discrepancies between the behavior of cultured cells and those collected from nature. Perhaps the most significant of these is the requirement that cultures be nutrient stressed to initiate sex (Turpin, Dobell et al. 1978; Anderson, Kulis et al. 1984). While nutrient stress does appear capable of replicating large-scale cyst deposition patterns in nature, sexual induction is also frequently observed when nutrients appear to be non-limiting (Anderson, Chisholm et al. 1983; Garces, Bravo et al. 2004; McGillicuddy, Anderson et al. 2005). Other factors may affect the yield of cysts produced in cultures, including the associated bacterial assemblage and changes in temperature, but these factors do not stimulate sexual fusion by themselves (Anderson, Kulis et al. 1984; Adachi, Kanno et al. 1999; Nagai, Matsuyama et al. 2004). Still, other stimuli likely exist because shifts toward sexual fusion and encystment appear to occur independent of nutrient depletion in the field.

A major source of the uncertainty in the conditions that cause sexual induction of *A. tamarense* complex cells in the field is the difficulty in differentiating vegetative and sexual forms, particularly gametes and vegetative cells. Different authors have described the gamete stage as hologamous (identical to one another and vegetative cells) and as

anisogamous (paired gamete cells differing in size), but in either case no characters that

differentiate free-swimming gametes from vegetative cells have been verified (Pfiester

and Anderson 1987). Instead, detection of sexuality in natural populations has been

restricted to observations of large cells that are presumed to be planozygotes (e.g.

Anderson, Chisholm et al. 1983). In culture, planozygotes are slightly larger than

vegetative cells (diameters greater than 45 μm versus 30-35 μm; Anderson and Lindquist,

1985), but cells in natural populations may be much more variable in size. This

characteristic of natural populations is perhaps due to their greater genetic heterogeneity,

or their exposure to macronutrients or other biological agents not present in culture (e.g.

bacterial assemblages, allelochemicals, or parasites).

Another very significant discrepancy between the sexual cycle in nature and in

culture is the apparent health of the hypnozygotes that the two processes produce.

Cultured hypnozygotes typically germinate at rates less than 35%, but those collected

from sediment cores can be stimulated to germinate at rates greater than 90% (Anderson

and Keafer 1987; Figueroa, Bravo et al. 2005; Matrai, Thompson et al. 2005). A caveat to

the latter observations is that Group I hypnozygotes collected from natural samples are

under the control of an endogenous circannual rhythm such that high germination rates

are only achievable within a period of 2-3 months each year (Anderson and Keafer 1987;

Matrai, Thompson et al. 2005). Appropriate experiments to demonstrate this type of

rhythmicity have only been conducted with Group I cysts collected from deep water

sediment cores in the Gulf of Maine. No similar behavior has ever been tested, or

replicated with cultured hypnozygotes.

A last source of uncertainty to the experiments presented in this thesis is the problem of assigning gender to *A. tamarense* clones. Though two vegetative clones must be cultured together to produce hypnozygotes in the laboratory, sexual compatibilities among groups of clones cannot be explained by simple male or female assignments to each (Destombe and Cembella 1990). Instead, *A. tamarense* and other dinoflagellates are frequently compared to ciliates, an allied but distantly related group in the eukaryotic protist lineage that expresses many genders (Destombe and Cembella 1990; Blackburn, Bolch et al. 2001). Mating by ciliates differs from most other sexual protists because its gender mechanism controls the initiation of sexual behavior rather than the physical attachment of two gametes (Dini and Nyberg 1993). Like dinoflagellates, the compatibility of distinct ciliate clones cannot be predicted based on simple binary gender assignments, but the probability that two ciliate clones are sexually compatible is also much higher than is observed among dinoflagellates. In this thesis, I have explored the application of a ciliate-like model to patterns in the compatibility of different *A. tamarense* clones, but the ciliate model is neither refuted nor proved. Instead, it is clear that though the *A. tamarense* gender system may have some ciliate-like characteristics, it must also have additional complexity to explain the mating patterns observed among many combinations of *A. tamarense* clones in the lab.

Mechanisms of A. tamarense Range Expansion

The toxic *A. tamarense* species (Groups I and IV) may be the most extensively studied harmful algae in the world and their spread in recent decades has been noted by

19

several authors (Anderson 1989; Smayda 1990; Hallegraeff 1993). The expansion of these species' ranges has been attributed to both natural and human-assisted mechanisms, all of which are intimately related to this group's life cycle, particularly its hypnozygote stage.

Because the hypnozygote stage is resistant to large variations in salinity and temperature, it is capable of remaining viable in areas that experience large seasonal changes in growth conditions. Moreover, once buried in anoxic sediment hypnozygotes may remain viable for years after their initial settlement in new environments. These characteristics make these organisms particularly well suited to dispersal via ship ballast water where cells are likely to experience long periods of poor growth conditions and may be deposited in a foreign habitat that is not immediately suitable for growth. Such a vector was shown to spread toxic Group IV from its native range along the coast of east Asia to Thau Lagoon, France (Lilly, Kulis et al. 2002).

Another serious concern among coastal resource managers is the potential regional scale shifts or expansions in the distribution of *A. tamarense* blooms. One mechanism for such regional shifts is the potential for expansion through coastal nutrient enrichment. Globally, coastal nutrient enrichment contributes to trends of increasing phytoplankton bloom intensities – with more nutrients available, phytoplankton blooms persist longer and spread over greater areas (Anderson, Glibert et al. 2002). A second factor that is likely to cause regional shifts in the distribution of toxic *A. tamarense* is changes in coastal ocean circulation related to climate change (Dale, Edwards et al. 2006). However regional shifts in the range of *A. tamarense* occur, they will present two

challenges. The first is the initial expansion of toxic blooms to areas that have not previously experienced them. The second, more sinister challenge may be the deposition of hypnozygote to these new areas, effectively colonizing them with populations of toxic cells that could persist for decades or longer. A prime example of such a regional shift was the appearance of persistent PSP toxicity in the western Gulf of Maine and Massachusetts Bay after a massive *Alexandrium* bloom spread south from the Bay of Fundy in 1972 (Hartwell 1974; Anderson 1997). Persistent, localized blooms from that event may have been established in tidal ponds and embayments along the eastern coast of Cape Cod, including the Nauset Marsh system (described in Chapter 2).

CHAPTER OVERVIEW

I have benefitted greatly from two unique data sets that were shared with me shortly after I arrived in Woods Hole. The first was a set of results from mating experiments that included well over fifty clones isolated from waters all over the world. Assessing the compatibility of any two clones takes time. Cultures must be prepared, mixed and incubated for three weeks or more before evidence of encystment can be observed (to say nothing about the time and energy required to isolate and maintain a large collection of clonal cultures). It's telling that the mating first experiments described here were completed in 1989, 15 years before I entered graduate school and a few months before

my 11th birthday. I'm grateful to have had the opportunity to work with this data set, especially since I could never have set out to create it by myself as a graduate student.

The second data set that I was shown was the results from a set of germination trials. These trials included hypnozygotes produced in combinations of two Group I clones, two Group III clones, and pairs of Group I and Group III clones. Uniquely and in every case, hypnozygotes from pairs of Group I and Group III clones failed to produce new viable cultures. The result suggested a barrier to genetic recombination between two of the most closely related *A. tamarense* clades and also a potential remediation strategy since Group III blooms do not synthesize saxitoxin. An overriding caveat to these results however was our uncertainty regarding how *A. tamarense* might be sexually induced and the degree to which the Group I and Group III clades had speciated from one another.

In Chapter 1 of this thesis, a set of experiments and analyses are described that verify the hybridization of Group I and Group III clones using a specially developed nested PCR assay. An immediate question that arose from this result was whether the hybridization between clades is an artifact of culturing them in the lab, or if such hybridization might also occur in nature. The question is addressed in two ways. First, the same nested PCR assay is applied to a field sample from Belfast Lough, Northern Ireland an area where both Group I and Group III blooms are known to occur. Second, the ability of a clone of one clade to choose between toxic and nontoxic clones is explored through comparisons of the rates of compatibility within ribosomal clades and between them.

The results from Chapter 1 highlight uncertainty regarding differences in the sexual behavior of *A. tamarense* species in the field versus their behavior in culture. One

overriding shortfall, noted above, has been the inability to positively identify sexual

stages in *A. tamarense* collected from the field. Such methods are needed to answer

questions about precisely when and where sexual processes occur and further describe the

likelihood of sexual interactions between different *A. tamarense* species.

Two initiatives to address the problem of identifying sexual cells are described

here. The first, presented in Chapter 2, is the modification of an imaging flow cytometer

to positively identify *A. tamarense* cells in mixed assemblages of many microplankton

taxa through both their micrographs and the detection of a species-specific ribosomal

probe. The method enables measurements of cell DNA content, and takes advantage of

established protocols and sampling procedures. Several unexpected patterns are observed

from samples collected during a bloom in Salt Pond (Eastham, MA) and a red tide that

occurred near Portsmouth, NH during July of 2009.

The second initiative is the sequencing and assembly of four dinoflagellate

transcriptomes, described in Chapter 3. To date transcriptome sequencing has been the

sole viable route to sequencing the *A. tamarense* species because their genomes are

massive (estimated to be nearly 70 times the size of the human genome; Hackett, Scheetz

et al. 2005). By sequencing expressed RNAs rather than genomic DNA, it is possible to

obtain sequences from the genes expressed by these cells. Here, I have applied a new

high throughput sequencing method called RNA-Seq to generate sequences that will be

the basis of future studies directed at biomarker discovery and the evolution of the *A.*

tamarense species. I have also applied the RNA-Seq method to an amplified set of RNAs

that were extracted from a small sample of naturally formed hypnozygotes.

A challenge associated with current high throughput sequencing methods is both

the overwhelming volume of data that they produce and their short read lengths. In

Chapter 3, I describe my efforts to assemble these many short reads into larger gene

sequences and discus ongoing work to annotate and verify my results thus far.

REFERENCES

Adachi, M., T. Kanno, et al. (1999). "Promotion of cyst formation in the toxic dinoflagellate *Alexandrium* (Dinophyceae) by natural bacterial assemblages from Hiroshima Bay, Japan." Marine Ecology Progress Series [Mar. Ecol. Prog. Ser.] **191**: 175-185.

Anderson, D. and N. Lindquist (1985). "Time-course measurements of phosphorus depletion and cyst formation in the dinoflagellate *Gonyaulax tamarensis* Lebour." Journal of Experimental Marine Biology and Ecology. **86**: 1-13.

Anderson, D. M. (1989). Toxic algal blooms and red tides: a global perspective. Red tides: biology, environmental science, and toxicology. Okaichi, D. M. Anderson and Nemoto, Elsevier Science Publishing Co., Inc.

Anderson, D. M. (1997). "Bloom dynamics of toxic *Alexandrium* species in the northeastern U.S." Limnology and Oceanography **42**(5, Pt. 2): 1009-1022.

Anderson, D. M. (1997). "Turning back the harmful red tide." Nature **388**(6642): 513.

Anderson, D. M., S. W. Chisholm, et al. (1983). "Importance of life cycle events in the population dynamics of *Gonyaulax tamarensis*." Marine Biology **76**(2): 179-189.

Anderson, D. M., P. M. Glibert, et al. (2002). "Harmful Algal Blooms and Eutrophication: Nutrient Sources, Composition, and Consequences." Estuaries **25**(4): 704-726.

Anderson, D. M. and B. A. Keafer (1987). "An endogenous annual clock in the toxic marine dinoflagellate *Gonyaulax tamarensis*." Nature **325**(6105): 616-617.

Anderson, D. M., D. M. Kulis, et al. (1984). "Sexuality and cyst formation in the dinoflagellate *Gonyaulax tamarensis* : Cyst yield in batch cultures." Journal of Phycology **20**(3): 418-425.

Anderson, D. M. and K. D. Stolzenbach (1985). "Selective retention of two dinoflagellates in a well-mixed estuarine embayment: the importance of diel vertical migration and surface avoidance." Marine Ecology Progress Series **25**: 39-50.

Blackburn, S. I., C. J. S. Bolch, et al. (2001). "Reproductive compatibility among four global populations of the toxic dinoflagellate *Gymnodinium catenatum* (Dinophyceae)." Phycologia **40**(1): 78-87.

Cembella, A. D. (1998). Ecophysiology and metabolism of paralytic shellfish toxins in marine microalgae. Physiological Ecology of Harmful Algal Blooms. D. M. Anderson, A. D. Cembella and G. M. Hallegraeff. Berlin, Springer: 381-403.

Dale, B., M. Edwards, et al. (2006). Climate Change and Harmful Algal Blooms. Ecology of Harmful Algae. E. Granéli and J. T. Turner, Springer-Verlag: 367-390.

Destombe, C. and A. Cembella (1990). "Mating-type determination, gametic recognition and reproductive success in *Alexandrium excavatum* (Gonyaulacales, Dinophyta), a toxic red-tide dinoflagellate." Phycologia **29**(3): 316-325.

Dini, F. and D. Nyberg (1993). "Sex in Ciliates." Advances in Microbial Ecology **13**: 85-153.

Figueroa, R. I., I. Bravo, et al. (2005). "Effects of nutritional factors and different parental crosses on the encystment and excystment of *Alexandrium catenella* (Dinophyceae) in culture." Phycologia **44**(6): 658-670.

Garces, E., I. Bravo, et al. (2004). "Relationship between vegetative cells and cyst production during *Alexandrium minutum* bloom in Arenys de Mar harbour (NW Mediterranean)." J. Plankton Res. **26**(6): 637-645.

Hackett, J. D., D. M. Anderson, et al. (2004). "Dinoflagellates: A Remarkable Evolutionary Experiment." American Journal of Botany **91**(10): 1523-1534.

Hackett, J. D., J. H. Wisecaver, et al. (in prep). "Independent evolution of saxitoxin synthesis in cyanobacteria and dinoflagellates."

Hall, S., G. Strichartz, et al. (1990). The saxitoxins: sources, chemistry and pharmacology. Marine Toxins: Oring, Structure ad Molecular Pharmacology. S. Hall and G. Strichartz. Washington, D. C., American Chemical Society: 29-65.

Hallegraeff, G. M. (1993). "A review of harmful algal blooms and their apparent global increase." Phycologia **32**: 79-99.

Hartwell, A. D. (1974). Hydrographic Factors Affecting the Distribution and Movement of Toxic Dinoflagellates in the Western Gulf of Maine. The First International Conference on Toxic Dinoflagellates, Boston, Massachusetts, The Massachusetts Science and Technology Foundation.

Hingman, W. A., D. M. Stone, et al. (2001). "Sequence comparisons of toxic and non-toxic *Alexandrium tamarense* (Dinophyceae) isolates from UK waters." Phycologia **40**: 256-262.

Jacobson, D. M. and D. M. Anderson (1996). "Widespread phagocytosis of ciliates and other protists by mixotrophic and heterotrophic thecate dinoflagellates." Journal of Phycology **32**: 279-285.

John, E. H. and K. J. Flynn (1999). "Amino acid uptake by the toxic dinoflagellate *Alexandrium fundyense*." Marine Biology **133**(1): 11-19.

John, U., R. A. Fensome, et al. (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.

Lilly, E. L. (2003). Phylogeny and Biogeography of the Toxic Dinoflagellate *Alexandrium*. Biological Oceanography. Woods Hole, Massachusetts Institute of Technology/ Woods Hole Oceanographic Institution. **Ph.D.:** 226.

Lilly, E. L., K. M. Halanych, et al. (2007). "Species Boundaries and Global Biogeography of the *Alexandrium tamarense* complex (Dinophyceae)." Journal of Phycology **43**(6): 1329-1338.

Lilly, E. L., D. M. Kulis, et al. (2002). "Paralytic shellfish poisoning toxins in France linked to a human-introduced strain of *Alexandrium catenella* from the western Pacific: evidence from DNA and toxin analysis." J. Plankton Res. **24**(5): 443-452.

MacIntyre, J. G., J. J. Cullen, et al. (1997). "Vertical migration, nutrition and toxicity in the dinoflagellate *Alexandrium tamarense*." Marine Ecology Progress Series (Mar. Ecol. Prog. Ser.).

MacKenzie, L., M. de Salas, et al. (2004). "The dinoflagellate genus *Alexandrium* (Halim) in New Zealand coastal waters: comparative morphology, toxicity and molecular genetics." Harmful Algae **3**(1): 71-92.

Matrai, P., B. Thompson, et al. (2005). "Circannual excystment of resting cysts of *Alexandrium* spp. from eastern Gulf of Maine populations." Deep-Sea Research II **52**(19-21): 2560.

McGillicuddy, D. J., D. M. Anderson, et al. (2005). "Mechanisms regulating large-scale seasonal fluctuations in *Alexandrium fundyense* populations in the Gulf of Maine: Results from a physical-biological model." Deep-Sea Research II **52**: 2698-2714.

Nagai, S., Y. Matsuyama, et al. (2004). "Effect of nutrients and temperature on encystment of the toxic dinoflagellate *Alexandrium tamarense* (Dinophyceae) isolated from Hiroshima Bay, Japan." Plankton Biology and Ecology **51**(2): 103-109.

Pfiester, L. A. (1989). "Dinoflagellate sexuality." International Review of Cytology **114**: 249-272.

Pfiester, L. A. and D. M. Anderson (1987). Dinoflagellate Reproduction. The Biology of Dinoflagellates. F. J. R. Taylor. Oxford, Blackwell Scientific Publications. **21:** 611-648.

Scholin, C. A., G. M. Hallegraeff, et al. (1995). "Molecular evolution of the *Alexandrium tamarense* 'species complex' (Dinophyceae): Dispersal in the North American and west Pacific regions." Phycologia **34**(6): 472-485.

Smayda, T. (1990). Novel and nuisance phytoplankton blooms in the sea: Evidence for a global epidemic. Toxic Marine Phytoplankton. E. Graneli, B. Sundstrom, L. Edler and D. M. Anderson. New York, Elsevier.

Spector, D. L. (1984). Dinoflagellate Nuclei. Dinoflagellates. D. L. Spector. Orlando, Harcourt Brace Jovanovich, Publishers: 107-147.

Taylor, F. J. R. (1984). Toxic Dinoflagellates : Taxonomic and Biogeographic Aspects with Emphasis on *Protogonyaulax*. Seafood Toxins. E. P. Ragelis.

Tomas, C. R. (1997). Identifying Marine Phytoplankton. Sand Diego, CA, Academic Press.

Townsend, D. W., S. L. Bennett, et al. (2005). "Diel vertical distributions of the red tide dinoflagellate *Alexandrium fundyense* in the Gulf of Maine." Deep-Sea Research II **52**(19-21): 2593.

Turpin, D. H., P. E. R. Dobell, et al. (1978). "Sexuality and cyst formation in Pacific strains of the toxic dinoflagellate *Gonyaulax tamarensis*." Journal of Phycology **14**(2): 235-238.

Chapter 1: Outbreeding lethality between toxic Group I and nontoxic Group III *Alexandrium tamarense* spp. isolates: predominance of heterotypic encystment and implications for mating interactions and biogeography

Michael L. Brosnahan[1], David M. Kulis[1], Andrew R. Solow[2], Deana L. Erdner[3], Linda Percy[4], Jane Lewis[4], and Donald M. Anderson[1]

1. Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, U.S.A.

2. Marine Policy Center, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, U.S.A.

3. Department of Marine Science, The University of Texas at Austin, Marine Science Institute, 750 Channel View Drive, Port Aransas, Texas 78373-5015, USA

4. School of Biosciences, University of Westminster, 115 New Cavendish St, London, W1W 6UW, UK

ABSTRACT

We report the zygotic encystment of geographically dispersed isolates in the dinoflagellate species complex *Alexandrium tamarense*, in particular, successful mating of toxic Group I and nontoxic Group III isolates. However, hypnozygotes produced in Group I/Group III co-cultures complete no more than three divisions after germinating. Previous reports have suggested a mate recognition mechanism whereby hypnozygotes

produced in co-cultures could arise from either homotypic (inbred) or heterotypic (outbred) gamete pairs. To determine the extent to which each occurs, a nested PCR assay was developed to determine parentage of individual hypnozygotes. The vast majority of hypnozygotes from pairwise Group I/Group III co-cultures were outbred, so that inviability was a result of hybridization, not inbreeding. These findings support the assertion that complete speciation underlies the phylogenetic structure of the *A. tamarense* species complex. Additionally, the ribosomal DNA (rDNA) copy numbers of both hybrid and single ribotype hypnozygotes were reduced substantially from those of haploid motile cells. The destruction of rDNA loci may be crucial for the successful mating of genetically distant conjugants and appears integral to the process of encystment.

The inviability of Group I/III hybrids is important for public health because the presence of hybrid cysts may indicate ongoing displacement of a nontoxic population by a toxic one (or vice versa). Hybrid inviability also suggests a bloom control strategy whereby persistent, toxic Group I blooms could be mitigated by introduction of nontoxic Group III cells. The potential for hybridization in nature was investigated by applying the nested PCR assay to hypnozygotes from Belfast Lough, Northern Ireland, a region where Group I and Group III populations co-occur. Two hybrid cysts were identified in 14 successful assays, demonstrating that Group I and Group III populations do interbreed in that region. However, an analysis of mating data collected over an 18-year period indicated a leaky pre-mating barrier between ribosomal species (including Group I and Group III). Whether the observed selectivity inhibits hybridization in nature is dependent on its mechanism. If the point of selectivity is the induction of gametogenesis, dissimilar ribotypes could interbreed freely, promoting displacement in cases where hybridization is lethal. If instead, selectivity occurs during the adhesion of gamete pairs, it could enable stable co-existence of *A. tamarense* species. In either case, hybrid inviability may impose a significant obstacle to range expansion. The nested PCR assay developed here is a valuable tool for investigation of interspecies hybridization and its consequences for the global biogeography of these important organisms.

1. INTRODUCTION

Mating is an integral part of the life cycle of many dinoflagellates and is associated in some species with the formation of hypnozygotic cysts. These cysts are highly resistant to environmental stresses and are a likely vector for the global expansion and persistence of many harmful dinoflagellate blooms (Anderson, 1989; Hallegraeff, 1993). The great majority of dinoflagellates proliferate as motile, haploid cells. Under certain conditions, some species cease mitotic division and undergo gametogenesis, transforming to non-dividing gametes that are competent to conjugate and fuse with one another. Importantly, the process of gametogenesis does not involve meiotic division because the progenitor motile cells are already haploid. Upon fusion, gamete pairs form planozygotes - swimming diploids that may persist for days or weeks before either returning to haploid, mitotic cell growth or metamorphosing to the hypnozygote form (Figure 1; Figueroa and Bravo, 2005; Figueroa *et al.*, 2006; Pfiester and Anderson, 1987).

Dinoflagellate species are categorized as either homothallic or heterothallic depending on whether clonal cultures can be induced to form zygotes. Heterothallic classification supposes that all pairs of gametes are heterotypic - that is, they are formed by fusion of non-sibling cells - whereas homothallic species may form hetero- or homotypic pairs. Species that have a hypnozygote stage are said to be homothallic if

31

encystment can be induced in clonal culture, or heterothallic if encystment requires

mixing of two compatible clones. However, conjugation in clonal cultures of putatively

heterothallic species, including the subject of this study, the *Alexandrium tamarense*

species-complex, is sometimes observed with low-grade or no subsequent production of

cysts (Blackburn *et al.*, 2001; Destombe and Cembella, 1990). Moreover, conjugation

and encystment are dissociated in numerous other dinoflagellates so that encystment may

not be the sole pathway for zygote maturation (Figueroa and Bravo, 2005; Figueroa *et al.*,

2006; Uchida, 2001). Still, it is the hypnozygote stage of *A. tamarense* that confers

resistance to prolonged stresses, and hypnozygotes may be the only abundant cell type

within *A. tamarense* populations for long periods (Anderson and Wall, 1978). While non-

encysting zygotes may benefit from genetic recombination, failure to encyst is likely

catastrophic in regions where sexual cycles occur in phase with extended periods of poor

growth conditions (e.g. high latitude winter). Though zygosis and encystment may be

separable in other species, these processes are strongly intertwined among the *A.*

tamarense species.

    Alexandrium tamarense species are similar to many ciliates in that their mating

system consists of more than two self-incompatible mating types (Destombe and

Cembella, 1990). Mating types in any eukaryotic microbe must affect one of two control

points in the sexual cycle: the transition to sexuality (gametogenesis); or, the physical

coupling (adhesion) of compatible gametes (Figure 1). A prime example of the latter type

of control is the unicellular chlorophyte *Chlamydomonas reinhardtii*, a species whose

gametes are either *plus* or *minus* type. These types each express specific and

complimentary flagellar agglutinins that ensure heterotypic pairing (Adair *et al.*, 1983;

Musgrave *et al.*, 1981). While conjugation by *A. tamarense* clones indicates expression of similar recognition factors, these factors may not differentiate distinct cell genders. In species where mating types act by reciprocally spurring gametogenesis, homotypic conjugation (i.e. inbreeding) may be possible. Well-elucidated examples include the hypotrich ciliate genus *Euplotes* (Dini and Nyberg, 1993) and the human fungal pathogen *Cryptococcus neoformans* (Lin *et al.*, 2005). In both *C. neoformans* and *Euplotes*, mature gametes are capable of initiating fusion with either similar or dissimilar gamete types. In the ciliate *E. crassus*, gametogenesis may also be induced non-reciprocally by co-culture with clones of its congener species *E. minuta*. In these interspecies co-cultures, only *E. crassus* clones are sexually induced so that all conjugants are homotypic pairs of *E. crassus* cells (Dini *et al.*, 1990). If interspecies mating within the *A. tamarense* species complex was similar to that observed in *Euplotes*, the heterothallic classification of *A. tamarense* species might not preclude them from homotypic encystment. Further, stimulation of inbreeding could be most profound when compatible clones are genetically distinct and stimulation of gametogenesis is non-reciprocal.

In this study, the extent and implications of homotypic and heterotypic encystment among genetically distinct groups of the *A. tamarense* species complex were assessed both in culture and in nature. Members of the *A. tamarense* complex are globally distributed and may produce saxitoxin and congener compounds, which, in turn, cause paralytic shellfish poisoning (PSP) in human and animal shellfish consumers. Most toxic clones used here are Group I and most nontoxic clones are Group III, where group designations are derived from phylogenetic relationships among the clones' large subunit ribosomal DNA sequences (LSU rDNA). These two groups occur in close geographic

33

proximity along the northern coasts of Ireland and Great Britain (Lilly *et al.* 2007). The

group designations (I and III) refer to two of five distinct clades that are largely isolated

in their global distribution. Each of the groups is consistent with respect to the toxicity of

its members: all tested Group I and IV clones produce PSP toxins, and all Group II, III

and V clones do not. Sequence divergence is less than 2% within the clades but up to

11% between them, comparable to that between the closely related species *A. tropicale*

and *A. affine*. Groups I and III, the focal groups in this study, have approximately 6%

divergence and are among the most closely related of the clades. Morphotype

designations *A. catenella*, *A. fundyense* and *A. tamarense* are commonly used to

differentiate isolates within the *A. tamarense* complex, but we do not regard these as

valid species because they are inconsistent with the complex's ribosomal phylogeny. All

three morphotypes occur but do not cluster within Groups I and IV. Only the *A.*

tamarense morphotype is known within Groups II, III and V (Lilly *et al.*, 2007; Scholin

et al., 1995).

Mating intercompatibility among the *A. tamarense* ribosomal clades was assessed

through an analysis of encystment and germination data accumulated over 18 years.

Importantly, germination studies demonstrated complete post-zygotic lethality among

hypnozygotes resulting from pairwise co-cultures of toxic Group I and nontoxic Group

III isolates. Most of these hypnozygotes failed to germinate, and those that did germinate

completed no more than three cell divisions. Because Group I/Group III progeny were

inviable, a method was needed to identify hypnozygotes as Group I, Group III or, if

present, Group I/Group III hybrids. Therefore, we developed a quantitative nested PCR

assay for genotyping single hypnozygotes, and used the assay to confirm outbreeding in

pairwise Group I/Group III co-cultures. The assay was also used to analyze hypnozygotes in sediment samples from Belfast Lough, a region in Northern Ireland where both Group I and Group III blooms are known to occur.

2. MATERIALS AND METHODS

2.1 *Clonal cultures*. One hundred nineteen clones were evaluated for hypnozygote formation in clonal culture and in pairwise co-culture with other clones. Clones were isolated from throughout the world, represented all major clades (62 Group I, 7 Group II, 39 Group III, 7 Group IV, and one Group V isolate), and included *A. tamarense, A. fundyense*, and *A. catenella* morphotypes. Additionally, two *A. affine* clones and one *A. tropicale* clone were evaluated (see Supplementary Table S1 and Figures S1 and S2). In many cases, cultures are no longer maintained and LSU rDNA sequence is not available for direct determination of a given clone's group affiliation. In these cases, group membership was assigned using LSU rDNA sequence data from clones that were isolated from the same field expeditions, or inferences were made based on available toxicity data and the known biogeographical distributions of the *A. tamarense* clades. Most clones were isolated between 1980 and 1995. The oldest clone, PGT183, was isolated in 1957. The youngest, a set of 15 Group I and 8 Group III clones were isolated from slurry germinations of Belfast Lough sediments in 2006.

2.2 *Encystment evaluation and hypnozygote storage*. A screen for encystment in pairwise co-cultures was initiated in 1989. The last experiments compiled here were completed in

Figure 1. Life cycle diagram of the *A. tamarense* species complex. Though the

hypnozygote stage may be bypassed by other encysting dinoflagellates, such an alternate

pathway of zygote maturation has not been described in *A. tamarense*.

Pellicle cyst

G2 motile cell

G1 motile cells

Gametogenesis

G2 motile cell

Gametes

Haploid

Meiosis

Adhesion

Diploid

Fusing gametes

Planomeiocyte

Planozygote

Excystment

Encystment

Hypnozygote

30 μm
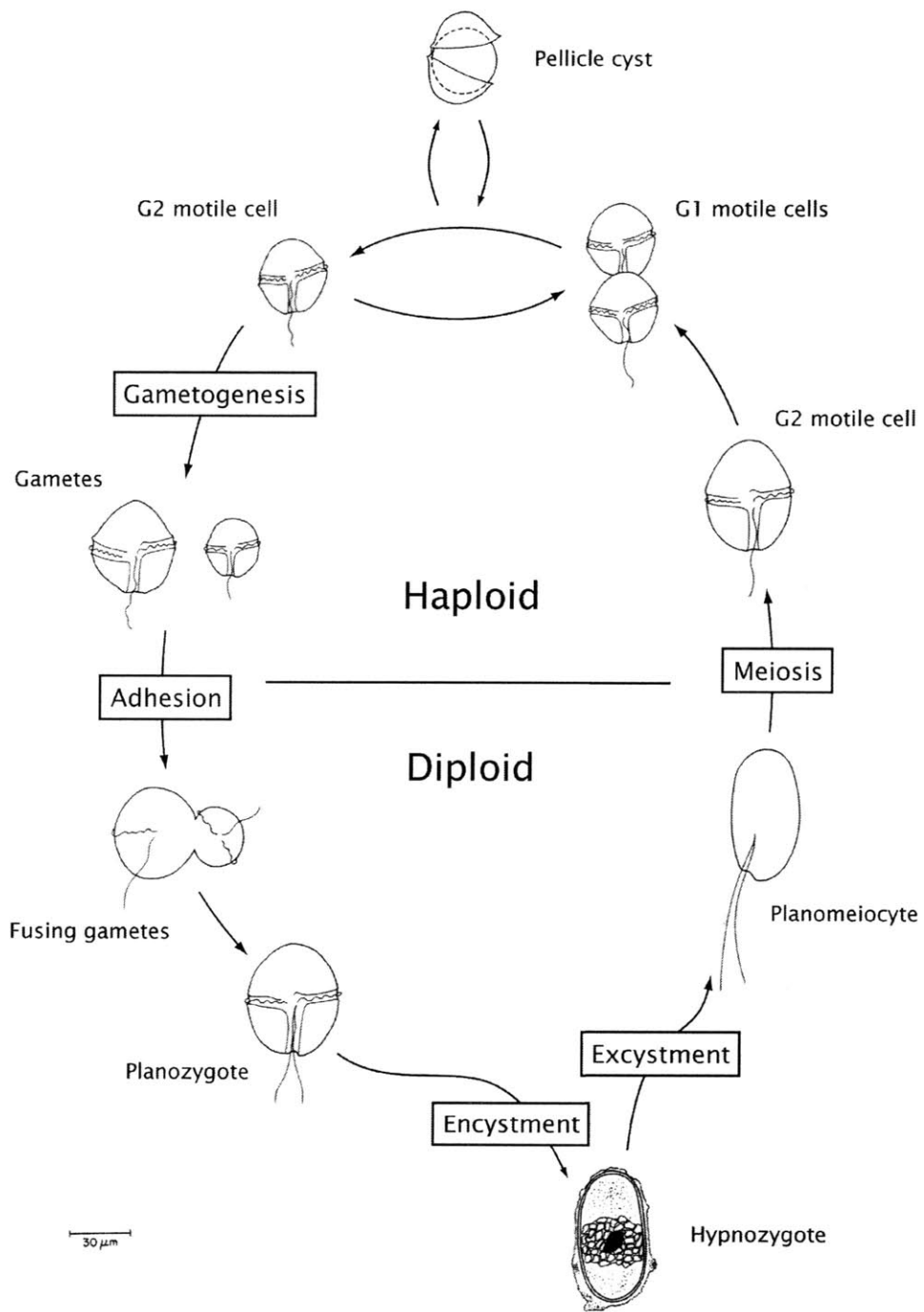
2007. All 119 isolates were tested for clonal encystment in N-limited culture, but all possible pairwise combinations were not. Four subsets of 12 to 15 clones were tested in all possible pairwise combinations (Figure 2 A-D). Two of these subsets were derived from a set of Belfast Lough isolates, and the others from American and European ('cosmopolitan') collections of isolates. One Belfast and one cosmopolitan subset were biased by selecting isolates that would lead to a higher proportion of successful encystment, while the other two subsets (one Belfast and one cosmopolitan) were unbiased. Experiments for both Belfast subsets were completed within one year of the isolation of clones (2007); the experiments for the unbiased and biased cosmopolitan subsets were completed in 1990 and 2005 respectively.

Cultures were grown in 25 mL volumes of modified f/2 (replete) or h/2 (N-limited) medium (Guillard and Ryther, 1962) in 50 mL borosilicate culture tubes at 20°C, on a 14:10-h L:D cycle with 150-200 μmol m^{-2} s^{-1} photon flux density (Anderson *et al.*, 1984). Medium modifications included: (1) elimination of NaSiO$_3$, (2) the reduction of CuSO$_4$ to 10^{-8} M, and (3) addition of H$_2$SeO$_3$ to the trace metal mixture (10^{-8} M final concentration). In the N-limited (encystment) medium, NH$_4$ was substituted for NaNO$_3$ as the nitrogen source at a final concentration of 25 μM. Additionally, 16 Belfast co-cultures were repeated in phosphate-limited medium that was prepared as described by Anderson and Lindquist (1985). Inoculum cultures maintained in replete medium were grown to mid/late-exponential phase and transferred into N-limited or phosphate-limited medium at an approximate cell density of 100-500 cells·mL^{-1} for each clone. After inoculation, the cultures were incubated for 30-40 days to maximize the opportunity for complete hypnozygote formation. The cell/cyst deposits at the bottom of the culture

tubes were sampled by Pasteur pipette and loaded into a Palmer-Maloney chamber. The contents of the chamber were scanned under 100x magnification to determine if hypnozygotes were present. Cyst yields from N-limited cultures were scored semi-quantitatively as 0, 1, 2 or 3. The ratings indicate that 0, 1-50, 50-100, or greater than 100 cysts were found per 25 mL N-limited culture. Hereafter, we refer to co-cultures scored 0 as 'negatives' and those scored 1, 2, or 3 as 'positives'. Some positives with higher yields (scored 2 and 3) were repeated to produce cysts for germination and genotyping experiments. Samples for germination and genotyping experiments were harvested into cryovials and stored in anoxic sediment as described by Anderson et al. (2003). We have previously found that exposure to anoxic sediment substantially prolongs germinability of stored culture samples compared to storage within the original culture tubes and incubation chambers.

2.3 *Hypnozygote viability.* For germination studies, single vials of hypnozygotes were removed from storage 3 to 30 months after collection. Vials were sorted on ice and under red-light to reduce stimulation and premature germination (Binder and Anderson, 1986). Hypnozygote samples were disaggregated by brief sonication and 1 to 237 cysts isolated by micropipette to separate wells of 96-well tissue culture plates containing 130-200 μL replete medium. In some cases, more than one cyst was placed in a well because the sonication step sometimes failed to eliminate clumping. Plates were incubated at 4°C, 15 °C or 20°C on a 14:10-h L:D cycle with approximately 150 μmol m^{-2} s^{-1} photon flux density. In all, hypnozygotes from 4 separate Group I only co-cultures, 7 Group III only co-cultures, and 10 Group I/III co-cultures were examined. Of these, germinations of

seven co-cultures (including four Group I/III co-cultures) were attempted at multiple

maturation periods (1 to 31 months) and temperatures (4°C, 15°C or 20°C). Isolated

hypnozygotes were monitored for excystment weekly or biweekly for at least one month.

Because hypnozygotes from Group I/Group III co-cultures never yielded viable cultures,

the total number of cells arising from these cysts was carefully observed under a

stereomicroscope so that all cells including those trapped in the air/water interface could

be counted.

2.4 *Real-time quantitative PCR assay.* Two real-time quantitative PCR (qPCR) assays

were developed for a genotyping assay and also for estimation of the LSU rDNA copy

number within Group I and Group III cysts. Consensus alignments of Group I and Group

III LSU rDNA D1-D2 hypervariable regions were constructed and primers were designed

to anneal to two regions of dissimilarity between the respective sequences (Figure 3;

Group I forward primer: 5' - GTG TTG CAC TTG CTT GAC AAG AGC - 3'; Group I

reverse primer: 5'-CAT CCC CAA GCA CAG GAA CAC AC-3'; 200 nucleotide

amplicon; Group III forward primer: 5' - GGT GAG ATT GTA GTG CTT GCT TGA

CAA TAG - 3'; Group III reverse primer: 5' - AAG GAA GGA AGC AAC CTC AAA

CAC ATG - 3'; 220 nucleotide amplicon). Specific amplification of targets was verified

by gel electrophoresis of PCRs with the respective primer pairs and genomic DNA

templates.

All qPCR assays were performed in 25 µL reactions by detection of SYBR Green

dye binding of product (Stratagene FullVelocity real-time qPCR chemistry, 400 nM

forward and reverse primers, 5 µl template, and 1 µl 1:10^3 dilution BioRad 1mM

fluorescein standard in DMSO). A fast two-step thermocycler program essentially as described in the FullVelocity manual was used (40 cycles; 10s 95°C, 30s 63°C; BioRad iCycler IQ real-time PCR detection system). Specific amplification of qPCR product was confirmed by melt curve analysis on each assay plate. Fluorescence thresholds were generally set according to default preferences of BioRad iCycler analytical software but were sometimes modified to exclude reaction failures (as marked by grossly aberrant slopes in log RFU/cycle plots).

D1R-D2C primer PCR products (Scholin *et al.*, 1994) of Group I and Group III templates were used as standards. Primers and other reactants from D1R-D2C amplifications were removed by product concentration on Qiagen QIAquick PCR Purification microcentrifuge columns. D1R-D2C products were then eluted and quantified using a NanoDrop ND-1000 spectrophotometer. Serial dilutions of these standards were made across 6 orders of magnitude, and all diluted standards were amplified in triplicate on each assay plate. Only technical replicates within 0.5 cycles of standard means were used for standard curve estimates. Standard curves were calculated from linear regression of log amplicons and threshold cycle values (Ct). Amplification efficiencies were calculated as the mean fraction of template molecules copied in each successive round of amplification.

2.5 *Single hypnozygote genotyping by nested PCR*. The parentage of individual hypnozygotes produced in Group I/Group III co-cultures was assayed by nested PCR. In brief, individual hypnozygotes were isolated directly to a D1R-D2C PCR reactant

Figure 2. Matrices of cyst yields from all pairwise co-cultures of clone subsets: (A)

unbiased cosmopolitan (1990 results only), (B) biased cosmopolitan (all yields observed,

1990-2006), (C) unbiased Belfast Lough, and (D) biased Belfast Lough. GTM253 (biased

cosmopolitan subset, B) is clonal isolate GTM253-17. Yields from each pairwise co-

culture were scored 0-3 on the basis of presence and abundance of hypnozygotes after >1

month incubation in nitrogen limited media (see text). Three co-cultures (combinations of

GT7, GTCA28, and GTCN16) from the unbiased cosmopolitan subset were repeated in

2005 and are boxed in both A (1990) and B (2005 result only).

A Unbiased Cosmopolitan

Group I		Group I											
		AFNFA3	AFNFA4	GT2	GT7	GTCA04	GTCA08	GTCA28	GTCA29	GTCN16	GTME20	GTMR01	GTPP03
Group I	AFNFA3	0	0	0	3	3	3	3	0	3	0	0	0
12 Clones	AFNFA4		0	0	3	3	3	3	0	0	0	0	0
	GT2			0	0	1	0	1	0	3	0	0	0
	GT7				0	0	0	[1]	0	[0]	0	0	0
	GTCA04					0	0	1	0	0	2	1	0
	GTCA08						0	0	1	0	1	0	0
	GTCA28							0	0	[1]	2	0	1
	GTCA29								0	0	0	1	0
	GTCN16									0	3	3	0
	GTME20										0	0	0
	GTMR01											0	0
	GTPP03												0

B Biased Cosmopolitan

Group I		Group I							Group III							
		38-3	ATRU03	GT7	GTCA28	GTCN16	GTLI21	GTM253	ATSP1-A	ATSP1-B	ATSP1-C	ATSP1-D	ATSW01-1	SP4B5-2	SP4B6-2	SP4E3-1
Group I	38-3	0	1	0	0	1	1	3	0	2	0	2	0	0	1	0
7 Clones	ATRU03		0	1	1, 0	1	1	1	0	2	0	0	0	0	1	1
	GT7			0	1,[0]	[0]	0	1	0	0	0	2	0	0	1	0
	GTCA28				0, 2, 1	1,[0]	0	2	0	1	1	0	0	0	1	0
	GTCN16					0	0	3	0	2	0	1	2, 1	0, 1	0, 1	0
	GTLI21						0	2	0	1	0	0	0	0	1	0
	GTM253							0	0	0	0	2	2	3	0	0
Group III	ATSP1-A								0	0	0	0	0, 3	0	0	0
8 Clones	ATSP1-B									0	0	0	2, 3	0	0	0
	ATSP1-C										0	0	0	0	1	0
	ATSP1-D											0	0	0, 2	2	0
	ATSW01-1												0	0	2, 3	0
	SP4B5-2													0	0, 3, 2	2
	SP4B6-2														0	3, 0
	SP4E3-1															0

C Unbiased Belfast

Group I		Group I					Group III						
		UW450	UW459	UW460	UW474	UW485	UW451	UW452	UW453	UW475	UW480	UW484	UW486
Group I	UW450	0	0	0	0	0	0	0	0	0	0	0	0
5 Clones	UW459		0	0	0	0	0	0	0	0	0	0	0
	UW460			0	0	0	0	0	0	0	0	0	0
	UW474				0	0	0	0	0	0	0	0	0
	UW485					0	0	0	0	0	0	0	0
Group III	UW451						0	2	0	2	0	2	0
7 Clones	UW452							0	1	0	3	0	2
	UW453								0	1	0	2	0
	UW475									0	3	0	2
	UW480										0	3	0
	UW484											0	3
	UW486												0

D Biased Belfast

Group I		Group I											Group III			
		UW455	UW456	UW482	UW464	UW465	UW466	UW468	UW473	UW485	UW499	UW500	UW475	UW480	UW484	UW504
Group I	UW455	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
11 Clones	UW456		0	0	0	0	0	0	0	0	0	0	2	0	0	0
	UW462			0	0	2	2	0	1	0	0	0	1	0	0	0
	UW464				0	0	0	0	0	0	0	0	0	0	0	0
	UW465					0	2	0	1	0	0	1	1	0	3	0
	UW466						0	1	0	0	0	0	0	0	0	1
	UW468							0	1	0	0	0	2	0	1	0
	UW473								0	0	0	0	0	0	0	0
	UW485									0	0	0	0	0	0	0
	UW499										0	0	0	0	0	0
	UW500											0	0	0	0	0
Group III	UW475												0	3	0	3
4 Clones	UW480													0	3	0
	UW484														0	0
	UW504															0

43

mixture. Hypnozygotes were lysed during isolation and denaturing steps of PCR

amplification. In the last step of the method, the D1R-D2C amplicons were interrogated

using the Group I and Group III qPCR assays to estimate the numbers of Group I and

Group III 'primary' amplicons (i.e. the approximately 700 bp fragments amplified during

D1R-D2C PCR). This approach increased sensitivity compared to a single amplification,

multiplex qPCR assay and also enabled qPCR replicates from single cysts.

In developing the nested PCR assay, micropipette isolation of hypnozygotes was

found to yield inconsistent results with low detectable numbers of primary PCR

amplicons. The performance of the assay improved dramatically when cysts were isolated

using a laser pressure catapult system. Sample preparation for laser catapulting was as

follows: first, (1) Director laser microdissection slides (Expression Pathology,

Gaithersburg, MD) were conditioned by an overnight wash in 1N HCl and then treated

with 0.01% poly-L-lysine; (2) for hypnozygotes from sediment samples, sediment

slurries were disaggregated by sonication and sieving, then enriched for hypnozygotes by

density centrifugation over a modified Nalco-sucrose step gradient (Schwinghamer *et al.*,

1991); (3) hypnozygote suspensions, either from sediment samples or stored pellets, were

diluted in 5–6 volumes deionized water, then sonicated using a Microsonix Sonicator

3000 (output level 2.5, 1 min) to disaggregate clumps; (4) suspensions were then dried to

the Director slides in a 55°C oven; lastly, (5) dried samples were rinsed successively in

coplin jars of deionized water, 70% ethanol, and 100% ethanol for >5 minutes each wash.

All slides were stored in 100% ethanol until hypnozygote isolation using a Zeiss PALM

inverted microscope system.

Hypnozygote cysts retained normal morphology when wet with ethanol but could also be identified by a characteristic 'raisin-like' appearance when dried. Individual cysts were laser pressure catapulted to PCR tube caps that had been pre-loaded with ~25 μL D1R-D2C (primary PCR) reactant mixture (600 nM D1R primer, 600 nM D2C primer, 1X NEB ThermoPol buffer, 2.5 U NEB Taq polymerase, 200 μM dNTPs). The caps were positioned <1 mm above the slide surface, so that the great majority of cysts were captured effectively using the AutoLPC function of the Zeiss PALM system. After launching a cyst, caps were re-positioned >2 cm above the slide then replaced into their PCR tube and stored on ice until thermocycling. The primary PCR itself was limited to 20 cycles to prevent product skew associated with plateau phase of amplification ($T_{M\,init}$ = 96°C, 5 min; $T_{M\,cycle}$ = 94°C, 40 s; T_A = 50°C, 1 min; T_E = 72°C, 1.5 min). Primers, reactants and other debris were removed from the primary PCR products using Qiagen QIAquick microcentrifuge columns and the products were eluted in 50 μL Qiagen EB buffer (10 mM Tris-Cl, pH 8.5). The purified products were then suitable for use as template in the Group I and Group III qPCR assays.

Genotyping was performed on hypnozygotes from 10 co-culture samples (Group I only co-cultures: 38-3 + GTM253-17, GTLI21 + GTM253-17; Group III only co-cultures: ATSW01-1 + SP4B6-2, SP4B5-2 + SP4B6-2; and mixed Group I/Group III co-cultures: 38-3 + ATSP1-B, 38-3 + SP4B6-2, ATSW01-1 + GTM253-17, ATSP1-D + GTM253-17, GTM253-17 + SP4B5-2, and a polyclonal co-culture of four Group I and five Group III isolates). Cell debris was also isolated from each of the Group I/Group III co-culture samples in order to assess contamination by adsorption of free Group I and Group III templates to cysts. Lastly, hypnozygotes from two sediment samples collected

```
                10        20        30        40        50        60        70        80        90       100       110       120
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Group I   ACCCGCTGAATTTAAGCATATAAGTAAGTGGTGGaAATTAAACCAAATGGGATATCTTTAGTAATTGCGAATGAAcAAGGATATGCTTAGCTTGACAAATGGAGCTATTGGCTTTGAATT
Group III ACCCGCTGAATTTAAGCATATAAGTAAGTGGTGGAAATTAAACCAAATGGGATTtCTTTAGTAATTGCGAATGAACAAGGATATGCTTAGCTTGACAAATGGAGCTTTTGGCTTTGAGTT


                370       380       390       400       410       420       430       440       450       460       470       480
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Group I   GAGAATTAAATGAGTTTGTATTTGCTGAACACAAAGTAAaCAGACTTGATTTGCTTGGTGGGAGTGTTGCACTTGCTTGACAAGAGCTTTGGGCTGTGGGTGTAATGATTCTTTCTTTGC
Group III GAGAATTAAATGAGTTTGTATTTGCTGAACACAAAGCAAACAGACTTGATTTGCTTGGTGAGATTGTAGTGCTTGCTTGACAATAGGTTTTGGCTGTGGGTGAAATGATTCTTTCTTTGC


                610       620       630       640       650       660       670       680       690       700       710
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Group I   GTGTGTGTTCCTGTGCTTGGGGATGC-TTCCTTCCTTGGACTTACAAGCCCTGACACACACATGCTGGCAAAATGCTTCTGCTTGACCCGTCTTGAAACACGGTCCAAGG
Group III GTTTTTGTTCCATGTGTTTGAGGTTGGTTCCTTCCTTGGGCTTACAAGCCCTGACACACACAATCTGGCAAAATGTTTCTGCTTGACCCGTCTTGAAACACGGTCCAAGG
```

Figure 3. Sequence alignment of Group I and Group III D1-D2 rDNA consensus sequences (segments 121–360 and 481-600 omitted). Conserved differences are highlighted in gray. Overline and underline arrows indicate forward/reverse positions of D1R/D2C (solid, bold), Group I (short dashed), and Group III (long dashed) primer pairs. Consensus sequences were constructed from 18 Group I and 6 Group III clones. Polymorphisms in the respective consensus sequences are denoted by lower case letters. Divergence between the consensus sequences is 6%.

from within Belfast Lough, Northern Ireland were also examined after analysis of isolates from slurry germinations revealed that both Group I and Group III hypnozygotes were present in the embayment.

2.6 *Hypnozygote rDNA copy number estimation.* We estimated the genomic copy number of LSU rDNA loci within Group I, Group III and hybrid hypnozygotes. Experiments completed by our labs have established that the copy number in motile cells is quite high and can vary substantially: Group III LSU rDNA copy numbers range from approximately $7x10^4 - 1.5x10^5$ and Group I rDNA copy numbers from approximately $5x10^5 - 1x10^6$ (experiments not described further in the present work, see also Galluzzi *et al.*, 2009). The Zeiss PALM system was used to collect samples of 484 – 1518 hypnozygotes from which DNA was extracted by a modification of the Invitrogen ChargeSwitch gDNA Micro Tissue kit.

Slides of co-culture or sediment material were prepared as described for single hypnozygote genotyping by nested PCR. After the final ethanol wash, slides were allowed to dry while set on the inverted microscope of the Zeiss PALM system. Batches of 100 to 300 cysts were rapidly mapped, then serially catapulted to a 500 μL PCR tube cap with ~70 μL Invitrogen proteinase K/Lysis Buffer L15. The number of cysts captured could then be assessed by raising the focal plane of the microscope to the lysis buffer surface. The lysis buffer (containing captured cysts) was transferred to PCR tubes by brief centrifugation and the process repeated 4 – 6 times for a single co-culture or sediment sample. Capture efficiency was only evaluated for 1 – 2 batches (PCR tube caps) each sample because the efficiency was consistently high (>85%) and because

counting of cysts could be time consuming, causing evaporative loss of lysis buffer. Following capture, samples were incubated overnight in a 55°C temperature block and vortexed in a suspension of 500 µM zirconium beads 3 x 1 min. PCR tubes were then punctured and lysate collected by centrifugation at 2000 g for 10 min. RNAse A treatment, wash and purification steps were followed according to the manufacturer's instructions. Extracted DNA was eluted from the ChargeSwitch magnetic beads in 150 µL Elution Buffer E5 (Tris-EDTA, pH 8.5) after brief warming in the 55°C temp block.

In all, eight samples were isolated for copy number estimation. Seven co-culture samples were examined (Group I only: 38-3 + GTM253-17, GTLI21 + GTM253-17; Group III only: SP4B5-2 + SP4B6-2; Group I/III: 38-3 + ATSP1-B, GTM253-17 + ATSP1-D, GTM253-17 + ATSW01-1, and GTM253-17 + SP4B5-2). A Group I sample was also taken from naturally produced hypnozygote cysts collected from Casco Bay, ME aboard the R/V Oceanus, Fall 2006.

2.7 *Statistcal analysis of genotyping data.* The significance of differences in the amplicon ratio of hypnozygotes from pairwise same-group and out-group co-cultures of Group I and Group III isolates was tested using a simple randomization procedure (Solow, 1990). Optimal threshold values of the Group I:Group III amplicon ratio for separating the three groups of hypnozygotes were found by minimizing the cross-validated misclassification rate. The significance of this rate was assessed by randomizing group designations (i.e. Group I only, Group III only, or Group I/Group III co-culture) and repeating the classification procedure (including threshold optimization) a total of 10,000 times. The same approach was used in separate assessments of the significance of Group I/hybrid,

Group III/hybrid, and Group I/Group III differences.

The randomization procedure demonstrated that significant differences exist in the amplicon ratio between hypnozygotes from pairwise same-group and out-group co-cultures (P < 0.0001, see Results), so that hypnozygotes are predominantly formed from heterotypic gamete pairs. The misclassification rate was substantially higher for cysts with low total numbers of amplicons, likely reflecting poor lysis efficiency. The number of amplicons detected is also related to the amplification efficiency during the primary PCR reaction, which was estimated to be 80% by qPCR of amplified and unamplified Group I and Group III genomic DNA standards (data not shown, $1.8^{20} \cong 130{,}000$-fold total amplification). It is assumed that the primary PCR efficiency was similar in the cyst genotyping experiments. Using nominal estimates of the Group I and Group III rDNA copy numbers and the estimated primary PCR efficiency, the expected amplicon sums were estimated at thresholds of 0.01, 0.1, 1, 10 and 100% lysis efficiency. The sums were evaluated for all mixtures of Group I and Group III cyst copy numbers under the constraint that the sum be the product of the cyst lysis efficiency and one whole cyst equivalent. In no case did a single cyst genotyping experiment yield more than the number of amplicons expected from complete lysis of a Group I cyst (i.e. the maximum amplicon sum expected based on results from hypnozygote copy number estimation; 100% lysis efficiency x 45,000 Group I rDNA copies x 1.8^{20} primary PCR amplification = 5.7 x 10^9 amplicons in the primary PCR product). Only 1 of 75 controls from laser catapult isolation of debris exceeded the number of copies expected from 0.01% lysis of a hypnozygote (data not shown). Accordingly, individual cyst genotyping experiments were considered successful if the sum of amplicons exceeded 0.01% efficiency (though

ratios derived from larger numbers of amplicons are reported with greater confidence).

To refine and extend the analysis, the results of cyst genotyping experiments with less than 0.01% efficiency were omitted and probability density functions (pdfs) of the log amplicon ratio:

$$x = \log \left[\frac{Group\,I\,amplicons + 1}{Group\,III\,amplicons + 1} \right]$$

were estimated by kernel density estimation. Kernel density estimation is a standard nonparametric approach to fitting a probability density function to data (Silverman, 1986). To apply kernel density estimation, it is necessary to select a bandwidth that controls the smoothness of the estimated density. Here, a common bandwidth was selected to minimize the cross-validated misclassification rate, where each cyst was classified to the group for which the estimated probability density of its log amplicon ratio was greatest. The classification thresholds for this rule correspond to the values of x at which the estimated probability density functions intersect.

The estimated pdfs can be used to assess the group membership probabilities for a cyst from an unknown group. Let $P(x \mid g)$ be the estimated probability density function for group g and assume equal prior group membership probabilities. By Bayes's Theorem, the posterior group membership probabilities are given by:

$$P(g \mid x) = \frac{P(x \mid g)}{\sum_g P(x \mid g)}$$

50

for g = Group I, Group III, and hybrid. The classification rule described above assigns a cyst to the group for which this probability is greatest. The posterior group membership probabilities for the cysts of known parentage are shown in Figure 5C. These probabilities were also calculated for cysts of unknown parentage that were collected from the polyclonal Group I/Group III co-culture and the two Belfast Lough sediment samples.

3. RESULTS

3.1 *Mating*. Encystment results from 896 co-cultures were compiled (12.8% of the possible pairwise combinations from the 119 clones; see Supplementary Table S1 and Figure S2). Of these, 224 co-cultures produced hypnozygotes (25% positives). Only five isolates produced hypnozygotes when grown clonally: two Group I clones and three Group III clones (Supplementary Figure S2). Co-cultures of clones from the same ribosomal sequence group ('same-group co-cultures') were more likely to be positives than out-group co-cultures (37.2% of same-group pairs versus 17.8% of out-group pairs). No crosses between *A. catenella* and *A. fundyense* morphotype isolates were positive but Group IV *A. catenella* did produce hypnozygotes in co-culture with Group I, III and V isolates of *A. tamarense* morphotype. Of the 21 possible out-group permutations among *A. tamarense* Group I-V, *A. affine* and *A. tropicale*, 13 were tested and 9 produced hypnozygotes. A significant caveat to these results is that our effort is dominated by co-cultures of Group I and Group III isolates; 41.0% of the co-culture experiments consist of one Group I and one Group III isolate, 23.8% are Group I only, and 12.7% are Group III

Table 1. Pairwise co-culture effort and number of positives (hypnozygote production). Three of seven possible same-group combinations (shaded) and 13 of 21 possible out-group combinations (unshaded) were tested. Positive ratios are reported in parentheses for combinations tested in more than 5 combinations.

Groups	Isolate Combinations Tested	Positives
I, I	214	68 (31.8%)
III, III	114	55 (48.2%)
IV, IV	5	1
I, II	35	1 (2.9%)
I, III	367	86 (23.4%)
I, IV	22	2
I, V	28	2
I, *A. affine*	28	1
I, *A. tropicale*	13	2
II, IV	14	0
III, IV	50	5 (10.0%)
III, V	1	0
III, *A. affine*	1	0
IV, V	2	1
IV, *A. affine*	1	0
IV, *A. tropicale*	1	0

only (see Table 1; also, Supplementary Figure S2). Among the 12 other out-group co-cultures, 7 have been attempted in fewer than 15 isolate combinations, including only 9 out-group combinations that did not include one Group I or Group III isolate.

Because the co-culture dataset is not exhaustive and experiments were pursued non-randomly, a randomization test was used to compare the ratios of positives to total combinations from same-group and out-group experiments (hereafter termed 'positive ratios'). Group assignments of the 119 isolates were permuted 1000 times to construct probability density estimates of differences between same-group and out-group positive ratios. In approximately half of these trials, the positive ratio of out-group crosses exceeded that of same-group co-cultures and in no case was the difference greater than 12% (substantially less than the 19.4% difference found in our dataset). In a similar randomization analysis restricted to results from Group I same-group, Group III same-group, and Group I/Group III co-cultures, the difference between same-group and out-group ratios also never exceeded the same-group/out-group differences in the true dataset (31.8% and 48.2% for Group I same-group and Group III same-group co-cultures respectively and 23.4% for Group I/Group III co-cultures; Table 1).

Encystment yields among positives varied substantially, including among replicates of some co-cultures, but were generally low. In all, 93 combinations (41.5%) were scored '1', 53 (23.7%) were '2', 41 (18.3%) were '3', where ratings represent 0, 1-50, 50-100, or greater than 100 cysts per 25 mL culture. The remaining positives (37 total) had variable yields (Supplementary Figure S2). The highest yields were observed in Group III same-group combinations followed by Group I same-group combinations (mean ratings of 0.47 and 0.39 respectively; N-limited co-cultures only and excluding variable yield positives). The highest average yield among out-group co-cultures was

Table 2. Comparison of positive ratios and average overall and positive yield ratings from four clone subsets that were tested in all pairwise combinations. Positive ratio is the proportion of co-cultures that were found to produce hypnozygotes. Average yield rating is the mean of yield ratings (0, 1, 2 or 3) from all subset co-cultures. Average positive yield rating is the mean of all subset co-cultures rated 1, 2 or 3. Biased subsets are shaded (see text for further description).

	Cosmopolitan		Belfast Lough	
Positive Ratio	0.250	0.429	0.159	0.181
Average Yield Rating	0.694	0.616	0.394	0.305
Average Positive Yield Rating	2.78	1.44	2.17	1.68

Group I/Group III (0.33). Yields in phosphate-limited trials were sometimes lower than yields from N-limited trials but no differences were observed in encystment compatibility.

Among the four clone subsets, the two unbiased subsets had the highest yield ratings among positive combinations (2.78 and 2.17, Figure 2 A and C), indicating more robust mating interactions versus biased subsets that had higher positive ratios. Yields in two of three co-cultures were reduced in the 2005 cosmopolitan subset from yields in the 1990 cosmopolitan subset (GT7/GTCA28 and GTCA28/GTCN16 co-cultures; Figure 2 A and B). Also noteworthy, was the appearance of clonal encystment (auto-encystment) activity by GTCA28 after the completion of the 1990 experiments. The suggestion that yields decrease in time from the cosmopolitan subsets is contradicted by the Belfast subset data (Figure 2 C and D; experiments completed within one year of clone isolations). While the yields from positive Group III combinations were higher than same-group yields in the complete dataset (2.23 vs. 1.85), yields from Belfast Group I combinations were quite low (1.38; or less than the average yield from all positive *out-group* combinations, 1.57).

Assignment of mating types within the four subsets was attempted by two different methods. In the first, we assumed a dioecious mating system in which the mating type of each isolate is fixed. If the assumptions of this model hold, all isolates should be assigned one of two mutually incompatible mating types, + or –. Of the four subsets, only one, the unbiased Belfast subset, can be divided in this manner. However, in this case, the only positives were Group III combinations. Similarly, all Group III isolates from the biased Belfast subset (excluding Group I/Group III co-culture results) can be

assigned mating types without ambiguity, but this is unsurprising given that three of the four Group III isolates are shared with the unbiased subset (Figure 2 C and D). More noteworthy from the Belfast results is that all but one combination of + and – Group III clones were positive for encystment. Neither of the cosmopolitan subsets can be sorted into only + and – mating types, either as a complete set or when results from Group I and Group III isolates are considered separately (Supplementary Figure S3 A-D).

The failure of the dioecious/fixed mating-type model contradicts one or both of its assumptions: that the number of mating types is limited to two, and that mating type is fixed within clonal cultures of each isolate. We addressed the former criterion by also considering only negative co-culture results. Clones of the respective subsets were divided into a minimal number of groups whose members are mutually incompatible. As occurs when only considering positives, some clones are poorly constrained and may be segregated to more than one group. Using this approach, the unbiased cosmopolitan and biased Belfast subsets can be solved using a three mating-type system, whereas the biased cosmopolitan subset requires four. When the biased cosmopolitan subset is broken into matrices of only Group I or Group III same-group co-cultures, the Group I clones still require four mating types but the Group III clones may be segregated into three (see Supplementary Figures S3 and S4).

3.2 *Germination.* Six Group I isolates and thirteen Group III isolates were co-cultured in 21 different pairwise combinations and in three poly-clonal combinations and the resulting cysts used for germination trials (Table 3; Supplementary Table S1). Cysts were formed, stored, and incubated numerous times over several years. The longest storage

56

time was 31 months for cysts from Group I/Group III co-cultures, 17 months for Group I co-cultures, and 10 months for Group III co-cultures. However, hypnozygotes from all three combinations remained ungerminated for more than three years when stored in cryovials as described.

Overall, 590 hypnozygotes resulting from Group I same-group combinations, 582 from Group III same-group combinations, and 866 hypnozygotes from Group I/Group III combinations (including 197 hypnozygotes from a polyclonal co-culture) were monitored for germination (Table 3). Among co-cultures from which 20 or more cysts were isolated, excystment success (defined as the proportion of hypnozygotes that gave rise to a germling cell) ranged from 0 to 76.2%. However, some excystment was observed in all ribosomal group combinations and incubation temperatures examined. Approximately half of the germling cells from hypnozygotes produced in Group I same-group and Group III same-group co-cultures were viable (i.e., completed >3 divisions after excystment), even though excystment success was low (5.1% and 34.5% overall average for Group I same-group and Group III same-group combinations respectively; Table 5). A precise count of viable germlings from same-group hypnozygotes was impossible because the cysts were sometimes isolated to tissue culture wells in clumps. When more than two hypnozygotes in a single well excysted, the well was frequently overgrown with motile cells so that it was not possible to determine the success of the respective germlings. Germination trials of hypnozygotes from pairwise Group I/Group III out-group co-cultures did not present similar difficulty because the germlings never completed more than three divisions. Overall excystment success of these hypnozygotes was 29.5%. For comparison, a much greater proportion of hypnozygotes from the polyclonal Group

57

I/Group III co-culture excysted (58.4%), and slightly less than a third of these divided more than three times.

3.3 *Hypnozygote rDNA copy number estimation*. DNA extractions from bulk samples of hypnozygotes were analyzed using Group I amplicon and Group III amplicon qPCR assays (8 samples, ranging from 484 – 1518 cysts; Table 4). The mean efficiency of the respective assays was 75.4% and 84.3%. Cross reactivities between the Group I primers and Group III templates, and the Group III primers and Group I templates were both negligible, though the former was higher (0.9% versus 0.009% of the true number of opposite standard templates 'detected').

The rDNA copy number estimates for Group I hypnozygotes were 44,454, 44,588, and 119,207 (Table 4). The latter estimate was from hypnozygotes formed during co-culture of clones GTLI21 and GTM253-17, and is considered aberrant because no cysts from this co-culture germinated (Table 5). The nominal estimated copy number was 45,000 in samples from another Group I co-culture (38-3/GTM253-17) and from Group I hypnozygotes collected from a Gulf of Maine sediment sample. Only one sample of Group III hypnozygotes was collected (SP4B5-2/SP4B6-2) but no germination trials were completed. The estimated Group III amplicon copy number was nominally 30,000 (Table 4).

Four hypnozygote samples from Group I/Group III co-cultures were analyzed for copy number and two were used in germination trials. The Group I rDNA copy number estimates were 5-10 fold greater than those for Group III in all four samples. This difference is consistent with that found during the genotyping experiments (described

below). In three of the four samples, the sum of the Group I and Group III copy estimates was intermediate between the nominal copy numbers of Group I and Group III hypnozygotes. In the one case where rDNA copy number was less than 30,000 (GTM253-17/SP4B5-2, estimated copy sum = 8,453), hypnozygotes did not excyst (n = 76). In the one other Group I/Group III sample tested for both germination and copy number (GTM253-17/ATSW01-1), 38,703 ribosomal copies were detected per cyst and 5 of 98 hypnozygotes excysted (Tables 3, 4 and 5).

3.4 *Single cyst genotyping*. The nested PCR assay effectively differentiated hypnozygotes formed by Group I and Group III parents. In an initial analysis, only hypnozygotes from pairwise co-cultures were considered because such cysts would have known parentage if *A. tamarense* clones formed only heterotypic pairs (total of 282 experiments: 42 from Group I co-culture, 24 from Group III co-culture, and 216 from Group I/Group III co-culture). Ratios between the number of Group I and Group III amplicons were used because both types were frequently detected (likely due to limited cross-reactivity between the respective templates and primers). If *Alexandrium* isolates were to behave as true heterothallic cells (no homotypic conjugation), the ratio of Group I to Group III amplicons detected should be heavily skewed toward Group I for Group I hypnozygotes, heavily skewed toward Group III for Group III hypnozygotes, and intermediate between Group I and Group III for hybrids.

Heterotypic encystment was tested by a randomization analysis of the genotyping results from pairwise co-cultures of Group I and Group III clones (described in Methods). The misclassification rate in the true dataset was 11.7%, but was never lower than 22.0%

Table 3. Germination results of hypnozygotes collected from co-culture of Group I only, Group III only and Group I/Group III isolates. Excystment Success is the proportion of hypnozygotes germinated (as demonstrated by the presence of germling cells or an archeopyle/empty hypnozygote test). Trials in which germlings did not complete more than 3 divisions are marked by an asterisk (*). Co-cultures that did not yield viable hypnozygote cysts are

shaded.

Group(s)	Co-culture	Total trials	Overall excystment success	Germination trial conditions		
				Temperature (°C)	Age at isolation (Mos.)	Excystment success
I	38-3, GTM253-17	1	2.7%	15	3	3/110
	GTCN16, GTMR01	14	10.0% (22/219)	4	4	7/15
					6	1/33
					7	2/16, 1/15, 3/11, 0/8
					8	0/18*
					9	3/20
					17	4/19
				15	4	0/21*
					7	0/1, 0/21*
					8	1/4
	GTLI21, GTM253-17	1	0.0%	15	3	0/116*
	38-3, GTCN16, GTLI21, GTM253-17	1	3.0%	15	3	5/162*
III	ATSP1-B, ATSW01-1	1	2.0%	15	3	2/100
	ATSW01-1, SP3B8-3	1	76.2%	15	1	16/21
	ATSW01-1, SP4B6-2	1	46.3%	15	3	76/162
	ATSW01-1, SP4E6-2	3	14.3% (4/28)	4	9	1/5
					10	3/16
				15	7	0/7*
	SP3B8-3, SP4E3-1	1	3.8%	15	4	1/26
	SP4E5-4, SP4E6-2	3	37.5% (3/8)	4	9	2/4
				15	4	1/3
					7	0/1*
	ATSP1-B, ATSP1-D, ATSW01-1, SP4B5-2, SP4B6-2	1	42.2%	15	3	100/237
I, III	2TOW1, PGT183	1	0.0%	20	5	0/7*
	38-3, SP4B6-2	1	61.4%	15	3	51/83*
	38-3, GT5-6	1	0.0%	20	5	0/19*
	3TOW1, PGT183	1	0.0%	20	5	0/9*
	GTCN16, SP4C7-3	1	100.0%	4	6	1/1*
	GTCN16, SP5B2-3	6	10.9% (11/101)	4	5	1/11*
					6	0/13*, 2/18*, 5/23*
					9	2/11*
				15	4	1/25*
	GTM253-17, ATSW01-1	1	5.1%	15	3	5/98*
	GTM253-17, SP4B5-2	1	0.0%	15	3	0/76*
	GTMR01, GT5-6	23	60.3% (117/194)	4	6	1/8*, 0/10*
					7	5/10*, 16/17*, 1/2*
					8	3/3*
					10	8/9*
					12	6/18*, 1/5*
					14	2/2*
					16	0/1*
					18	9/11*
					20	1/2*
					22	2/5*
					23	1/4*
					30	1/1*
					31	3/3*
					UR	7/13*
				15	6	0/1*
				20	8	19/32*
					9	14/19*
					10	14/15*
					15	3/3*
	GTMR01, SP4E5-4	7	23.1% (12/52)	4	3	0/5*
					5	2/4*
					6	1/19*, 1/2*
					9	7/14*
				15	5	0/2*
					7	1/6*
	GTMR01, SP5B3-1	3	0.0% (0/14)	4	5	0/3*
					6	0/5*
				15	4	0/6*
	MR16G6, GT5-6	1	6.7%	20	5	1/15*
	38-3, GTCN16, GTLI21, GTM253-17, ATSP1-B, ATSP1-D, ATSW01-1, SP4B5-2, SP4B6-2	1	58.4%	15	3	115/197

in several 10,000 randomization trials so that the low error rate in the true dataset is highly significant (P < 0.0001). Similarly, when the pairwise data were themselves considered in pairs (e.g. Group I and hybrid data only, not Group III), the misclassification rates were significantly lower than expected under a null hypothesis of no difference between groups (P < 0.0001 in each case). Thus, genotyping data from pairwise cultures strongly support a mating mechanism whereby homotypic conjugant pairs are prevented from encysting. Following from this finding, hypnozygotes produced in pairwise co-cultures are considered to be of known parentage (Group I, Group III or hybrid). Accordingly, hypnozygotes from pairwise co-culture are hereafter referred to as 'knowns', and those isolated from polyclonal co-cultures and sediment samples are referred to as 'unknowns'.

In estimation of the Group I, Group III and hybrid probability density functions of x, only assays that exceeded the 0.01% lysis contour were used because misclassifications were rare beyond this threshold (Figure 4 A and B). Overall, 31.5% of assays exceeded the 0.01% threshold (137/435), but most were of knowns (113 versus 24). Intersections between the pdfs were $x = -0.96$ (Group III/ hybrid boundary) and $x = 2.61$ (hybrid/Group I boundary; Figure 5B), so that the range classified as hybrids is approximately 1:9 to 407:1, Group I to Group III amplicons. This range reflects incomplete cyst lysis during isolation and consequent sub-sampling of individual cysts' rDNAs. The hybrid probability density is greatest where Group I:Group III is approximately 2:1, a skew that is comparable to that observed in copy number estimates from unamplified hypnozygote genomic DNA (approximately 8:1; Table 4). The bias toward Group III in the genotyping

data suggests a difference in D1R-D2C amplification efficiency between Group I and Group III templates that is similar to the variance observed between the efficiencies in the two qPCR assays. However, as the misclassification rate was quite low under the thresholds defined by the pdf intersections (1.8%, 2/113), such a bias during D1R-D2C amplification did not substantially affect the classification skill of the assay.

Of the 24 successful assays of unknowns, 4 hypnozygotes were classified hybrid, 9 Group I, and 11 Group III. Only 2 of 10 polyclonal unknowns were classified hybrid, but this was expected because multiple compatible clones of both types were inoculated in these co-cultures. Seven of the eight other successful polyclonal assays were Group III, a result that is consistent with higher growth rates and stationary densities observed in the Group III isolates. Only one unknown was classified with probability less than 95%, a hybrid that was isolated from Belfast Lough sediment core A7B ($x = 2.46$, estimated probability 76.5%). A second hybrid was isolated from a Belfast sediment core station A4A and was classified with probability greater than 99%. Lysis efficiency of the A4A hybrid exceeded 10% (estimated >3000 template molecules), further supporting its classification. Group I and Group III single-type cysts were also detected in the Belfast sediment samples, though more Group I were found than Group III (8 versus 4).

4. DISCUSSION

4.1 *Speciation within the A. tamarense complex.* An authoritative delineation of species within the *A. tamarense* complex has been lacking since the discovery that the group's morphotypes are uncorrelated with its ribosomal phylogeny (Scholin and Anderson, 1994). Subsequent investigations using various molecular and morphological characters have repeatedly confirmed the divergence of the five *A. tamarense* clades, the clades'

global geographic separation, and also their close kinship to sister species, *A. tropicale,*
A. tamiyavanichi, and *A. affine* (John *et al.*, 2003; Leaw *et al.*, 2005; Lilly *et al.*, 2007;
MacKenzie *et al.*, 2004). In the present study, limited mating compatibility was found
across most of these groups, a finding that is similar to reports of broad mating
compatibility among globally dispersed isolates of another PSP dinoflagellate,
Gymnodinium catenatum (Blackburn *et al.*, 2001; Oshima *et al.*, 1993). However, unlike
G. catenatum, hybrids of two closely related *A. tamarense* clades failed to yield any
viable progeny. The latter supports the assertion that complete speciation underlies the
phylogenetic topology of the *A. tamarense* complex. Such has been proposed previously
by Lilly *et al.* (2007) on the basis of ribosomal divergence between the complex's clades
and its uncontested sister species. However, the mating and germination data presented
here are only conclusive with respect to two of the five clades (Group I and Group III). It
remains to be shown whether the progeny of other mating-compatible clades are viable.

Even so, if species are defined as populations that can recombine genetically in
nature, the *A. tamarense* clades are largely speciated solely on the basis of their present-
day biogeography. Expansive intermingling of clades is limited to Groups I and III along
the northern coasts of Ireland and Great Britain and to Groups I and IV (both toxic)
within the East China Sea (Lilly *et al.*, 2007). Our demonstration of hybrid cysts in
Belfast Lough, Northern Ireland, confirms that mating is possible between cells of
different clades within a natural setting. However, because these hybrids do not yield
viable progeny, gene flow between Group I and Group III populations is not possible.
There are also small, isolated populations that commingle with the larger, interconnected
populations of other ribosomal species. These isolated populations most likely arose as a

Table 4. Copy number estimates of Group I and Group III type LSU rDNA loci in Group I, Group III and hybrid hypnozygotes. N is the estimated number of hypnozygotes collected by laser pressure catapult. Approximations of Group I and Group III copy numbers used for Figure 4 boundaries are italicized (see text). Hypnozygotes that did not excyst are marked by an asterisk (*). Hypnozygotes for which no germination trials were attempted are marked by two asterisks (**).

Groups/Clones	N	GI copy number	GIII copy number
Group I only		*(45,000)*	-
38-3, GTM253-17	744	44,588	N/T
GTLI21, GTM253-17*	1518	119,207	N/T
Gulf of Maine Sediment	608	44,454	N/T
Group III only		-	*(30,000)*
SP4B5-2, SP4B6-2**	484	N/T	30,136
Group I, Group III			
38-3, ATSP1-B**	948	28,513	2,718
GTM253-17, ATSP1-D**	664	35,059	4,462
GTM253-17, ATSW01-1	642	32,169	6,533
GTM253-17, SP4B5-2*	769	7,692	760

Table 5. Summary of findings. All cysts produced in culture had substantially lower excystment rates than those recovered from natural sediments (Anderson and Keafer, 1987; Matrai *et al.*, 2005). Both hybrid cysts and those with abnormal rDNA abundance failed to resume mitotic reproduction when stimulated to germinate, but only those with abnormal rDNA failed to excyst. (* = excystment rate commonly observed in naturally occurring Group I hypnozygotes during peak phase of the circannual germination cycle).

	Gulf of Maine	Group I	Group III	Group I/III hybrids	Abnormal rDNA/cyst
Encystment	Yes	Yes	Yes	Yes	Yes
rDNA/cyst	45,000	45,000	30,000	30,000-40,000	120,000 and 8,500
Excystment rate	90%*	5.1%	34.5%	29.5%	0% (0/192)
Return to haploid mitosis	Yes	Yes	Yes	No	No

result of global shipping (e.g. a population of Group IV, an Asian clade, in Thau Lagoon, France; Lilly *et al.*, 2002). Such human-assisted dispersal provides a mechanism to renew interaction between clades, but is unlikely to have affected the evolution of the *A. tamarense* complex because large-scale shipping did not begin until the last century.

The present-day biogeography of the *A. tamarense* complex and its genetic structure both indicate a sustained period of allopatry (geographic isolation) beginning 23-45 MYA (John *et al.*, 2003). With establishment of allopatry, genetic divergence of the clades occurred according to selection and drift in each of the their respective ranges. Such divergence is correlated with the development of incompatibilities that limit sexual interaction or that cause inviability upon inter-group mating. Groups I and III represent the latter case because encystment by Group I/III conjugants does not yield viable progeny. The timing of lethality is highly significant because contact between populations of Groups I and III may cause local elimination of the less abundant population, particularly in the absence of pre-mating barriers between these species. Elimination would occur because the probability that an individual Group I or Group III gamete will form a hybrid is proportional to the probability that it encounters gametes of the other clade. The more abundant population will have greater mating success than the less abundant one, so that, in subsequent mating cycles, numerical differences tend to be compounded and cells of the less abundant species become increasingly likely to form fatal pairs. It follows that pre-mating barriers are necessary for stable co-existence because they limit (or abolish) gametic contact between species (Paterson, 1993).

For sympatric dinoflagellates, pre-mating barriers may have a variety of mechanisms including: (1) different receptivities to biological signals, (2) variation in

Figure 4. Single hypnozygote genotyping by nested PCR. (A) Results of hypnozygotes from pairwise co-culture of Group I and Group III isolates ('knowns'). Lysis efficiency contours (0.01%, 0.1%, 1%, 10% and 100%) are expected numbers of amplicons after primary D1R-D2C PCR. Group I/hybrid and hybrid/Group III thresholds are indicated by dashed diagonal lines; gray diagonal lines indicate regions about the thresholds where classification probability is less than 95% (see text, Figure 5). (B) Results of hypnozygotes from a polyclonal Group I/Group III co-culture and two sediment samples taken within Belfast Lough (survey stations A4A and A7B). Lysis efficiency contours and classification boundaries are drawn as in (A).

A

Group I only
♦ 38-3 x GTM253-17
■ GTLI21 x GTM253-17

Group III only
△ ATSW01-1 x SP4B6-2
○ SP4B5-2 x SP4B6-2

Group I / Group III
▨ 38-3 x ATSP1-B
◉ 38-3 x SP4B6-2
● ATSW01-1 x GTM253-17
◇ ATSP1-D x GTM253-17
▲ GTM253-17 x SP4B5-2

B

◇ Group I / Group III
  polyclonal co-culture

▨ Station A4A

▲ Station A7B

69
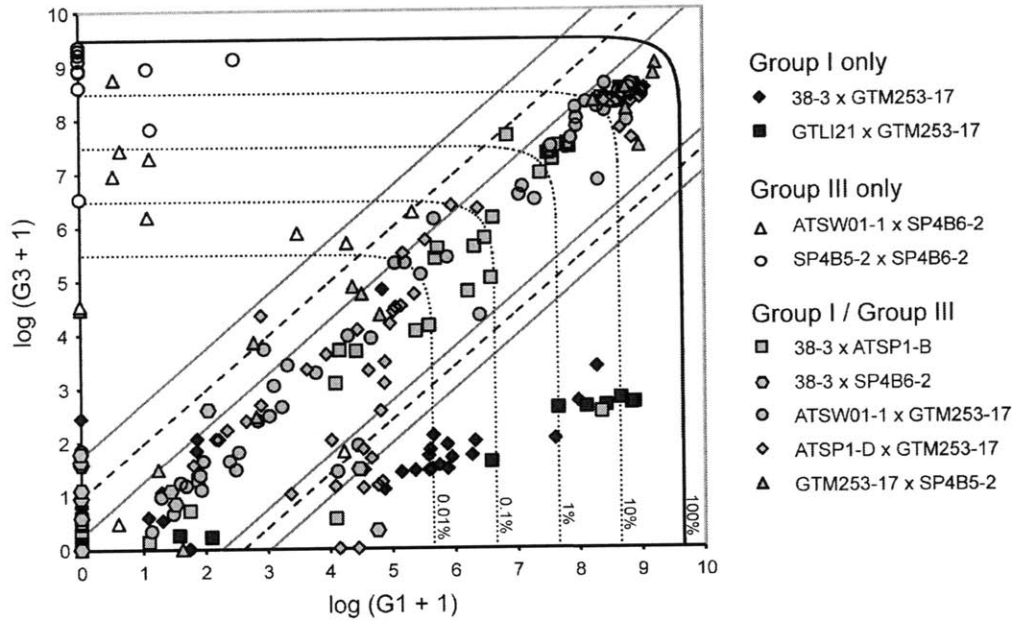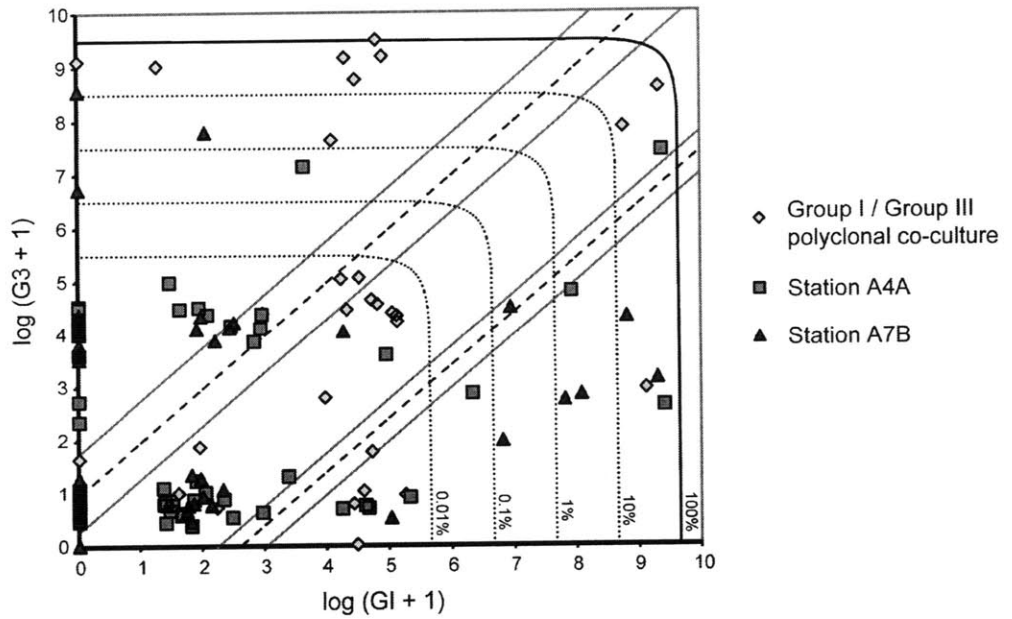
gamete aggregation patterns (e.g. depths of thin layer formations), and (3) variation in the

timing of sexual induction. The reduced positive encystment ratios in out-group

combinations of *Alexandrium* (as compared to same-group combinations; Supplementary

Figure S2) suggest that the biological signals for mating are only partially shared. This

limited reactivity between clades may be caused by non-transitive relationships among

mating factors. As an example, a 'male' signal from one group could interact positively

with a 'female' signal from another. However, this positive interaction does not ensure

similar reactivity between a 'female' signal of the first group and a 'male' signal of the

second. It is also possible that the lower positive ratio among out-group co-cultures is due

to incompatibilities that occur after mating but before encystment (i.e. during the

planozygote stage; Figure 1). Such a post-conjugation mechanism is considered less

likely because many out-group combinations do encyst successfully. It is therefore

assumed that non-transitive relationships between pre-mating factors are more likely than

between post-conjugal ones. If this assumption were violated (i.e. the lower positive ratio

were due to zygotic rather than mating incompatibilities), spatial or temporal pre-mating

barriers that limit contact between *A. tamarense* clades could still exist in areas where

their ranges overlap. However, such ecological barriers have not been shown. Further, the

presence of hybrids in Belfast Lough indicates that such barriers, if they do exist, are

incomplete (at least between Groups I and III).

The sum of these observations is that while broad mating intercompatibility has

been retained within the *A. tamarense* complex, it does not ensure gene flow between

commingling populations of divergent clades. This is because post-zygotic

incompatibility like that shown in Group I/III hybrids may prevent the proliferation of

genetic intermediates. Further, in the absence of pre-mating barriers, post-zygotic

lethality can provide a mechanism for elimination of less abundant clades (*sensu* species)

through fatal hybridization. These findings create an incongruity within our study in that

the two focal clades have been distinguished as separate species but their hybrids are used

to establish normal, intraspecies mating behavior. The approach has been borne out by its

results as virtually no inbred cysts were detected in Group I/III pairwise co-cultures.

Obligate hybridization demonstrates an inability of clonal siblings to conjugate with one

another and encyst.

4.2 *Post-zygotic lethality of Group I/Group III hybrids.* Initially, the assessment of post-

zygotic lethality was complicated because of the uncertainty in cyst parentage and the

low germinability of hypnozygotes produced in culture. Although germination success

was improved by storage of hypnozygotes in anoxic sediment, excystment rates remained

low (26.5% across all trials, Table 3). Three possible causes of low excystment were

considered: (1) that hypnozygotes were not adequately stimulated to germinate; (2) that

hypnozygotes were commonly subject to inbreeding depression; or (3) that maturation of

hypnozygotes was impaired under the conditions used in this study. The first cause was

rejected because hypnozygotes from natural sediments may be stimulated to excyst at

rates >90% under similar conditions (Anderson and Keafer, 1987). Likewise, our finding

that cysts are predominantly outbred in pairwise Group I/III co-cultures negated the

possibility of inbreeding depression. With respect to impaired maturation, the biotic and

abiotic conditions in culture were unavoidably different from those experienced in nature.

Therefore we conclude that the chosen conditions, though they prolong germinability,

71

Figure 5. Probability density estimation of log ratios from successful 'known'

genotyping experiments. (A) linear scatter of successful 'known' experiments: Group I

only co-culture – black, solid diamonds; Group III only co-culture – open circles; Group

I/Group III co-culture – black 'x' symbols. (B) Gaussian kernel estimation of probability

densities. The intersection of the estimated Group I (solid line) and the hybrid (dash-dot

line) densities is the Group I/hybrid classification threshold; the intersection of the

estimated hybrid density and the estimated Group III density (dotted line) is the

hybrid/Group III classification threshold (See Figure 4). (C) Estimated probabilities of

Group I, hybrid and Group III cyst parentage by log ratio given naïve prior probability of

hypnozygote parentage. Probabilities are calculated by Bayes's theorem.

A

B

Normalized Density

0.75

◆ ——— Group I only
○ - - - - Group III only
✕ — · — Group I/Group III

-10   -5   0   5   10

C

Probability

1

-10   -5   0   5   10

$\log (GI + 1)/(GIII + 1)$
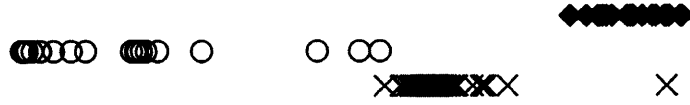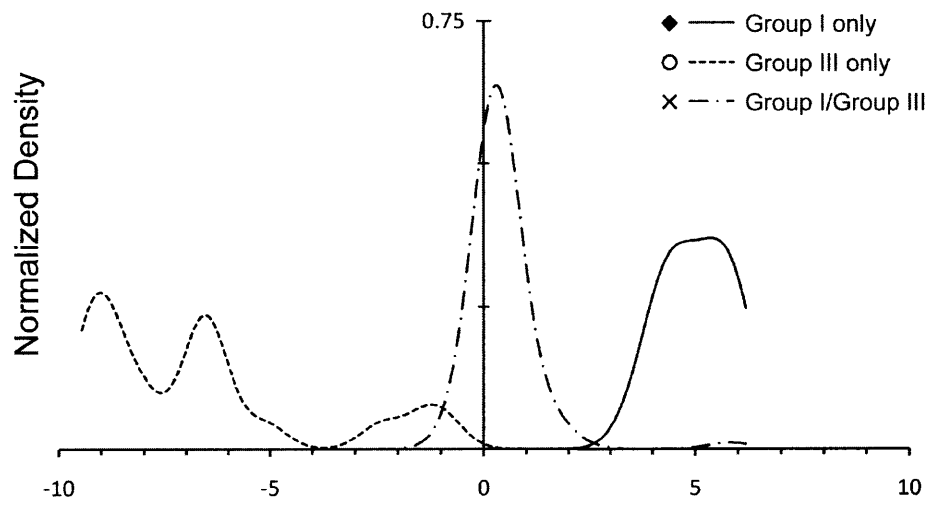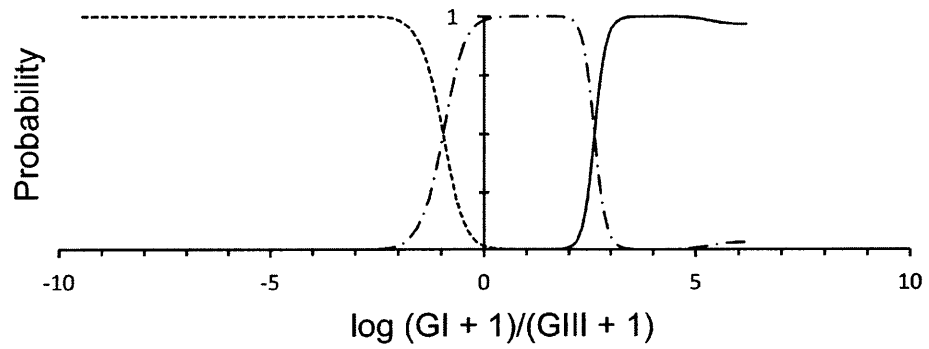
also cause germination rates to be low. We have no ready solution for improving the germinability of cultured cysts other than to follow Figueroa *et al.* (2005) who demonstrated improved germinability after isolation to un-enriched medium. A similar protocol might improve excystment success for Group I hypnozygotes but it remains unclear why natural Group I cysts will germinate efficiently in enriched medium.

In any case, low overall germinability of cultured cysts does not diminish the observation of inviability among hybrids. A much greater proportion of hybrid hypnozygotes than Group I hypnozygotes excysted in the 669 hybrid trials and, most importantly, no germlings from 152 hybrid excystments completed more than one mitotic division. The lethality of Group I/III hybrids must result from genetic recombination because it does not occur until after the completion of zygotic stages. The viability of the zygotic stages is surprising given that the estimated rDNA copy number of Group I motile cells is 5-fold greater than that for Group III motile cells (0.5-1 x 10^6 and 1-2 x 10^5 copies/cell respectively). If these copy numbers are fixed in the respective haploid genomes, as many as 2 x 10^6 copies per Group I zygote and 4 x 10^5 copies per Group III zygote should be recovered upon cyst lysis. Instead, the apparent rDNA copy number of the hypnozygotes is as much as 90-fold lower than expected and nearly equivalent between Group I and Group III cysts. Erdner *et al.* (this issue) have estimated similarly low copy numbers for Group I hypnozygotes, corroborating our results. The reductions in rDNA copy number by both Group I and Group III zygotes to nearly equivalent values may be crucial for the successful maturation observed in hybrids.

In the two cases where hypnozygote copy numbers deviated from the majority of other measurements (GTLI21/GTM253-17 and GTM253-17/SP4B5-2), hypnozygotes

uniquely failed to excyst. In the former case, the apparent copy number was substantially greater than that found in other Group I cysts so that excystment failure was associated with incomplete rDNA copy reduction. Conversely, the estimated copy number from GTM253-17/ SP4B5-2 hybrid hypnozygotes was 4- to 5- fold lower than that of hybrids that did excyst successfully. In this case, an *over*-reduction of rDNA loci was associated with excystment failure. Combined, the copy estimates from viable and inviable hypnozygotes suggest that the low rDNA copy number of hypnozygotes results from destruction of loci during the process of encystment rather than recalcitrance of DNA during extraction. Such origin regulation of rDNA loci is prominent in the sexual cycle of ciliates (Tower, 2004). However, the scale of regulation we report for *A. tamarense* is much less (compare to several thousand-fold in *Tetrahymena thermophila*).

The destruction of rDNA loci (and presumably other genes) during encystment may enable the progression of hybrids through the sexual cycle by masking disequilibria between distinct motile cell genomes. In the absence of origin regulation, differences in homolog structure or regulation might be catastrophic to any of the steps involved in the diploid phase of the sexual cycle. Tolerance of over-abundant loci in the early diploid stages may provide resilience to disequilibria upon hybrid conjugation. In turn, reductions to near equivalent copy numbers during subsequent zygote stages could be critical for successful completion of meiosis. Hybrid lethality is likely caused by genetic disequilibrium because its onset occurs after the completion of meiosis (and within the first germling mitotic cell cycle). Cell death is unavoidable once motile cells begin to re-establish higher gene copy numbers.

4.3 *Mating system models and interactions between ribosomal species.* A curious aspect

of sexual processes in nature is the ubiquity of dioecious (two mating-type) systems.

Such binary cell recognition is easily evolved from models of primitive homothallism,

and is also a sufficient mechanism to reduce inbreeding depression (Hoekstra, 1987).

Gametic dimorphism (or anisogamy) is also ubiquitous among eukaryotes and has been

reported in some *A. tamarense* cultures (Anderson and Lindquist, 1985 ; Turpin *et al.*,

1978). If sexual conjugation is limited to unlike gametes, such dimorphism confirms a

system that both prevents homotypic pairing and that also consists of only two mating

types. That anisogamous gamete pairs form within *A. tamarense* cultures suggests that

these species are dioecious, but has been contradicted by attempts to classify clones as

one of two mating types (Supplementary Figures S3 and S4; Destombe and Cembella,

1990). A critical assumption underlying these assessments was that mating types are

fixed within clones. The occurrence of auto-encystment by some clonal cultures partly

undermines this assumption since it must have occurred via a mating-type switch.

However, auto-encystment was limited to 4 of 119 isolates and cyst yields in each of

these cases were low (Supplementary Figure S2). Together, the rarity of auto-encystment

and its low yields suggest that mating-type switches by individual cells are uncommon,

even within stocks of auto-encysting clones. Further, if mating-type switching were

frequent, a much greater proportion of all pairwise combinations would have produced

cysts than was observed (Table 1).

Given the evidence of a multiple mating-type system, it was also surprising that

the positive encystment ratios remained low even among same-group combinations taken

from restricted geographical areas (e.g. consider the unbiased Belfast Lough subset where

less than 40% of same-group combinations were positive). Most multiple mating-type organisms have mating reactivity rates that approach 100% with increasing numbers of mating types (Bull and Pease, 1989). From the attempts to classify clone genders in the four isolate subsets, a minimum of four mating types must exist so that the positive ratio of same-group combinations should approach 75%. Instead, the maximum positive ratio observed for any clade was 48.4% (Group III; Table 1), much less than would be expected for a freely interbreeding species having four mating types, but very near the 50% expected under a dioecious model.

The lower positive encystment ratios among out-group combinations in *A. tamarense* is similar to broad but diminished inter-reactivity that has been described between populations of numerous ciliate genera (Dini and Nyberg, 1993). For this reason and because ciliates are closely allied to dinoflagellates within the Alveolate protist lineage, previous authors have speculated that the multiple mating-type systems of heterothallic dinoflagellates are like those of ciliates (Blackburn *et al.*, 2001; Destombe and Cembella, 1990). A comparison to the ciliate genus *Euplotes* is therefore instructive because its mating system has been described in considerable molecular detail, and because its species are also known to have more than two mating types. As mentioned previously, the mating-type system of *Euplotes* stimulates transition to the sexual cycle, but in contrast to *A. tamarense*, does not prevent homotypic conjugation. Individual *Euplotes* cells express their mating type through one of numerous pheromones that inhibit gametic differentiation when bound in homologous pairs (as would occur in clonal culture). When unlike mating types interact in polyclonal cultures, the autocrine inhibition loop is interrupted, freeing cells to transform to their competent sexual forms

(Luporini *et al.*, 2005; Vallesi *et al.*, 1995). This system enables *Euplotes* species to limit

sexual differentiation in highly clonal populations where genetic recombination is

unlikely to be beneficial, and also allows greater mating reactivity than is observed in

dioecious species. Neither the predominance of heterotypic encystment nor the low

mating reactivity of *A. tamarense* isolates refute the presence of a *Euplotes*-like mating

barrier, but its mating system must have additional complexity to explain these

characteristics.

A simple modification to the *Euplotes* model is the addition of a second barrier

that prevents adhesion of homotypic pairs (for instance, + and – types like those of *C.

reinhardtii*). This second adhesion barrier is contrasted to the first *Euplotes*-like barrier

that acts as a self-recognition mechanism. Given a system with four self-recognition

types and two adhesion types, the probability that a pairwise combination is positive

would converge toward 37.5% where each clone is characterized by one each of the self-

recognition and adhesion types. This target compares favorably to the 31.8% positive

encystment ratio found among Group I combinations but is much less than the 48.4%

observed among Group III (Table 1). The higher ratio in the latter case could be

explained by an absence of the self-recognition barrier or by a larger number of self-

recognition types among Group III clones. As an alternative to the proposed two-step

model, a one-step model would require a multiple mating type adhesion system whereby

each isolate expresses a factor that is recognizable as non-self by other isolates and in

turn promotes adhesion between them. Such a one-step system was also hypothesized for

Euplotes before the molecular characterization of its self-recognition system (Heckmann

and Kuhlmann, 1986). A weakness of such a model for *A. tamarense* species is that low

positive encystment ratios could only be explained by non-transitive interactions between mating factors, both between ribosomal groups and also within ribosomal groups where gene flow has been less restricted. Given the low positive ratios within the *A. tamarense* clades, it is unclear what advantage a multiple mating type adhesion system would lend versus a less complex dioecious system. In contrast, a two-step system would both limit gametic differentiation when sexual recombination is unprofitable and also ensures heterotypic conjugation.

The precise system that mediates mating among *Alexandrium* cells has substantial implications for the interaction of unlike, natural populations, such as the mixing of toxic Group I and nontoxic Group III cells. If one accepts that reduced encystment among out-group co-cultures results from impaired biological signaling, the pre-mating barriers must occur either during gametogenesis or adhesion. If the barrier affects adhesion, conjugation between unlike ribosomal clades would be inhibited under both the one- and two-step models. If instead the barrier affects gametogenesis, greater sexual interaction between unlike clades might be expected under certain natural conditions. This is because differentiated gametes would freely conjugate with those of other clades, particularly where populations commingle and are mutually induced to enter the sexual cycle. It would also follow that the finding of selectivity for isolates of the same ribosomal species should not be taken to represent the behavior of natural polyclonal populations. Instead, the finding of selectivity would be due to the contrivance of co-culturing only two clones. In natural, polyclonal populations such as those occurring in Belfast Lough, both may be self-stimulating so that an individual Group I cell would undergo gametogenesis upon stimulation by its conspecifics, and a Group III cell similarly by its own population. The

resulting Group I and Group III gametes could then intermingle, conjugate, and encyst unimpeded.

5. CONCLUSIONS

Co-culture experiments were first undertaken as an approach to characterize relationships among globally distributed populations, and many of these experiments pre-date our understanding of *A. tamarense* ribosomal phylogeny. Persistent questions about the nature of sexual interactions among dinoflagellates led to development of the nested PCR method for assessing parentage of hypnozygotes produced in co-cultures. Using this method, the predominance of heterotypic encystment was demonstrated and normal heterothallic sexuality (heterotypic conjugation) between two *A. tamarense* clades was verified. Substantial copy reduction of rDNA loci was also discovered upon encystment.

The discovery that clones from widely dispersed populations will encyst is similar to reports of gametic compatibility among geographically dispersed isolates of *G. catenatum*, but hybrids of genetically isolated *A. tamarense* Group I and Group III species failed to yield viable progeny. This discovery has immediate practical applications because Group I cells produce PSP toxins but Group III cells do not (Scholin *et al.*, 1994; Lilly *et al.*, 2007). Further, the ranges of the these two species only overlap at their shared boundary. Taken together, post-zygotic lethality and the species' minimal biogeographic overlap implicate interbreeding between them as a mechanism hindering expansion of one type into the range of the other, and vice versa. It also follows that deliberate introduction of the nontoxic Group III species should be considered a promising approach for mitigation of recurrent toxic Group I blooms.

6. ACKNOWLEDGEMENTS

7. REFERENCES

Adair, W.S., Hwang, C., Goodenough, U.W., 1983. Identification and Visualization of the Sexual Agglutinin from the Mating-Type Plus Flagellar Membrane of *Chlamydomonas*. Cell 33, 183-193.

Anderson, D., Lindquist, N., 1985. Time-course measurements of phosphorus depletion and cyst formation in the dinoflagellate *Gonyaulax tamarensis* Lebour. Journal of Experimental Marine Biology and Ecology. 86, 1-13.

Anderson, D.M., 1989. Toxic algal blooms and red tides: a global perspective. In:

Okaichi, Anderson, D.M., Nemoto (Eds.), Red tides: biology, environmental science, and toxicology. Elsevier Science Publishing Co., Inc.

Anderson, D.M., Fukuyo, Y., Matsuoka, K., 2003. Cyst methodologies. In: Hallegraeff, G.M., Anderson, D.M., Cembella, A.D. (Eds.), Manual on Harmful Marine Microalgae. United Nations Educational, Paris, Fr, pp. 165-190.

Anderson, D.M., Keafer, B.A., 1987. An endogenous annual clock in the toxic marine dinoflagellate *Gonyaulax tamarensis*. Nature 325 (6105), 616-617.

Anderson, D.M., Kulis, D.M., Binder, B.J., 1984. Sexuality and cyst formation in the dinoflagellate *Gonyaulax tamarensis* : Cyst yield in batch cultures. Journal of Phycology 20 (3), 418-425.

Anderson, D.M., Wall, D., 1978. Potential importance of benthic cysts of *Gonyaulax tamarensis* and G. excavata in initiating toxin dinoflagellate blooms. Journal of Phycology 14 (2), 224-234.

Binder, B.J., Anderson, D.M., 1986. Green light-mediated photomorphogenesis in a dinoflagellate resting cyst. Nature 322 (6080), 659-661.

Blackburn, S.I., Bolch, C.J.S., Haskard, K.A., Hallegraeff, G.M., 2001. Reproductive compatibility among four global populations of the toxic dinoflagellate *Gymnodinium catenatum* (Dinophyceae). Phycologia 40 (1), 78-87.

Bull, J.J., Pease, C.M., 1989. Combinatorics and Variety of Mating-Type Systems. Evolution 43 (3), 667-671.

Destombe, C., Cembella, A., 1990. Mating-type determination, gametic recognition and reproductive success in *Alexandrium excavatum* (Gonyaulacales, Dinophyta), a toxic red-tide dinoflagellate. Phycologia 29 (3), 316-325.

Dini, F., Bleyman, L.K., Giubbilini, P., 1990. Non-Mendelian Inheritance of Early Maturity in *Euplotes crassus*. Journal of Eurkaryotic Microbiology 37 (6), 475-478.

Dini, F., Nyberg, D., 1993. Sex in Ciliates. Advances in Microbial Ecology 13, 85-153.

Figueroa, R.I., Bravo, I., 2005. Sexual Reproduction and Two Different Encystment Strategies of *Lingulodinium polyedrum* (Dinophyceae) in Culture. Journal of Phycology 41 (2), 370-379.

Figueroa, R.I., Bravo, I., Garces, E., 2006. Multiple Routes of Sexuality in *Alexandrium taylori* (Dinophyceae) in culture. J. Phycol. 42, 1028-1039.

Galluzzi, L., Bertozzini, E., Penna, A., Perini, F., Garcés, E., Magnani, M., 2009. Analysis of rRNA gene content in the Mediterranean dinoflagellate *Alexandrium catenella* and *Alexandrium taylori*: implications for the quantitative real-time PCR-based monitoring methods. Journal of Applied Phycology *in press*.

Guillard, R.R.L., Ryther, J.H., 1962. Studies of marine plankton diatoms I. *Cyclotella nana* Hustedt and *Detonula confervacea* (Cleve) Gran. Canadian Journal of Microbiology 8, 229-239.

Hallegraeff, G.M., 1993. A review of harmful algal blooms and their apparent global increase. Phycologia 32, 79-99.

Heckmann, K., Kuhlmann, H.W., 1986. Mating types and mating inducing substances in *Euplotes octocarinatus*. J. Exp. Zool. 237, 87-96.

Hoekstra, R.F., 1987. The evolution of sexes. In: Stearns, S.C. (Ed.), The Evolution of Sex and its Consequences. Birkhäuser Verlag, Boston, pp. 59-91.

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.

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. Phycologia 44 (5), 550-565.

Lilly, E.L., Halanych, K.M., Anderson, D.M., 2007. Species Boundaries and Global Biogeography of the *Alexandrium tamarense* complex (Dinophyceae). Journal of Phycology 43 (6), 1329-1338.

Lilly, E.L., Kulis, D.M., Gentien, P., Anderson, D.M., 2002. Paralytic shellfish poisoning toxins in France linked to a human-introduced strain of *Alexandrium catenella* from the western Pacific: evidence from DNA and toxin analysis. J. Plankton Res. 24 (5), 443-452.

Lin, X., Hull, C.M., Heitman, J., 2005. Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature 434 (7036), 1017-1021.

Luporini, P., Alimenti, C., Ortenzi, C., Vallesi, A., 2005. Ciliate Mating Types and Their Specific Protein Pheremones. ACTA Protozoologica 44, 89-101.

MacKenzie, L., de Salas, M., Adamson, J., Beuzenberg, V., 2004. The dinoflagellate genus *Alexandrium* (Halim) in New Zealand coastal waters: comparative morphology, toxicity and molecular genetics. Harmful Algae 3 (1), 71-92.

Matrai, P., Thompson, B., Keller, M., 2005. Circannual excystment of resting cysts of *Alexandrium* spp. from eastern Gulf of Maine populations. The Ecology and Oceanography of Toxic *Alexandrium fundyense* Blooms in the Gulf of Maine 52 (19-21), 2560.

Musgrave, A., Eijk, E., Welscher, R., Broekman, R., Lens, P., 1981. Sexual agglutination factor from *Chlamydomonas eugametos*. Planta 153 (4), 362-369.

Oshima, Y., Itakura, H., Lee, K.C., Yasumoto, T., Blackburn, S., Hallegraeff, G., 1993. Toxin production by the dinoflagellate *Gymnodinium catenatum*. In: Smayda, T.J., Shimizu, Y. (Eds.), Toxic Phytoplankton Blooms in the Sea. Proceedings of the Fifth international Conference on Toxic Marine Phytoplankton, Newport, RI (USA), 28 Oct 1991. Elsevier, Amsterdam (Netherlands), pp. 907-912.

Paterson, H.E.H., 1993. Evolution and the Recognition Concept of Species. The Johns Hopkins University Press, Baltimore, MD.

Pfiester, L.A., Anderson, D.M., 1987. Dinoflagellate Reproduction. In: Taylor, F.J.R. (Ed.), The Biology of Dinoflagellates. Blackwell Scientific Publications, Oxford, pp. 611-648.

Scholin, C.A., Anderson, D.M., 1994. Identification of Group- and Strain-specific Genetic Markers for Globally Distributed *Alexandrium* (Dinophyceae). I. RFLP Analysis of SSU rRNA Genes. Journal of Phycology 30 (4), 744-754.

Scholin, C.A., Hallegraeff, G.M., Anderson, D.M., 1995. Molecular evolution of the *Alexandrium tamarense* 'species complex' (Dinophyceae): Dispersal in the North American and west Pacific regions. Phycologia 34 (6), 472-485.

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 Journal of Phycology 30 (6), 999-1011.

Schwinghamer, P., Anderson, D.M., Kulis, D.M., 1991. Separation and Concentration of Living Dinoflagellate Resting Cysts from Marine Sediments Via Density-Gradient Centrifugation. Limnology and Oceanography 36 (3), 588-592.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, New York, NY.

Solow, A.R., 1990. A randomization test for misclassification probability in discriminant analysis. Ecology 71 (6), 2379-2382.

Tower, J., 2004. Developmental Gene Amplification and Origin Regulation. Annual Review of Genetics 38 (1), 273-304.

Turpin, D.H., Dobell, P.E.R., Taylor, F.J.R., 1978. Sexuality and cyst formation in Pacific strains of the toxic dinoflagellate *Gonyaulax tamarensis*. Journal of Phycology 14 (2), 235-238.

Uchida, T., 2001. The role of cell contact in the life cycle of some dinoflagellate species. J. Plankton Res. 23 (8), 889-891.

Vallesi, A., Giuli, G., Bradshaw, R.A., Luporini, P., 1995. Autocrine mitogenic activity of pheromones produced by the protozoan ciliate *Euplotes raikovi*. Nature 376 (6540), 522.

Supplementary Figure S1. Worldwide distribution of clonal isolates used in mating

compatibility screens. Different symbols denote different groups within the *A. tamarense*

species complex. See also Supplementary Table S1.

Legend

A. tamarense species complex

Group I ●
Group II □
Group III △
Group IV ◆
Group V ⬡

A. affine ≠
A. tropicale +

Belfast Lough
(● x 15)
(△ x 8)
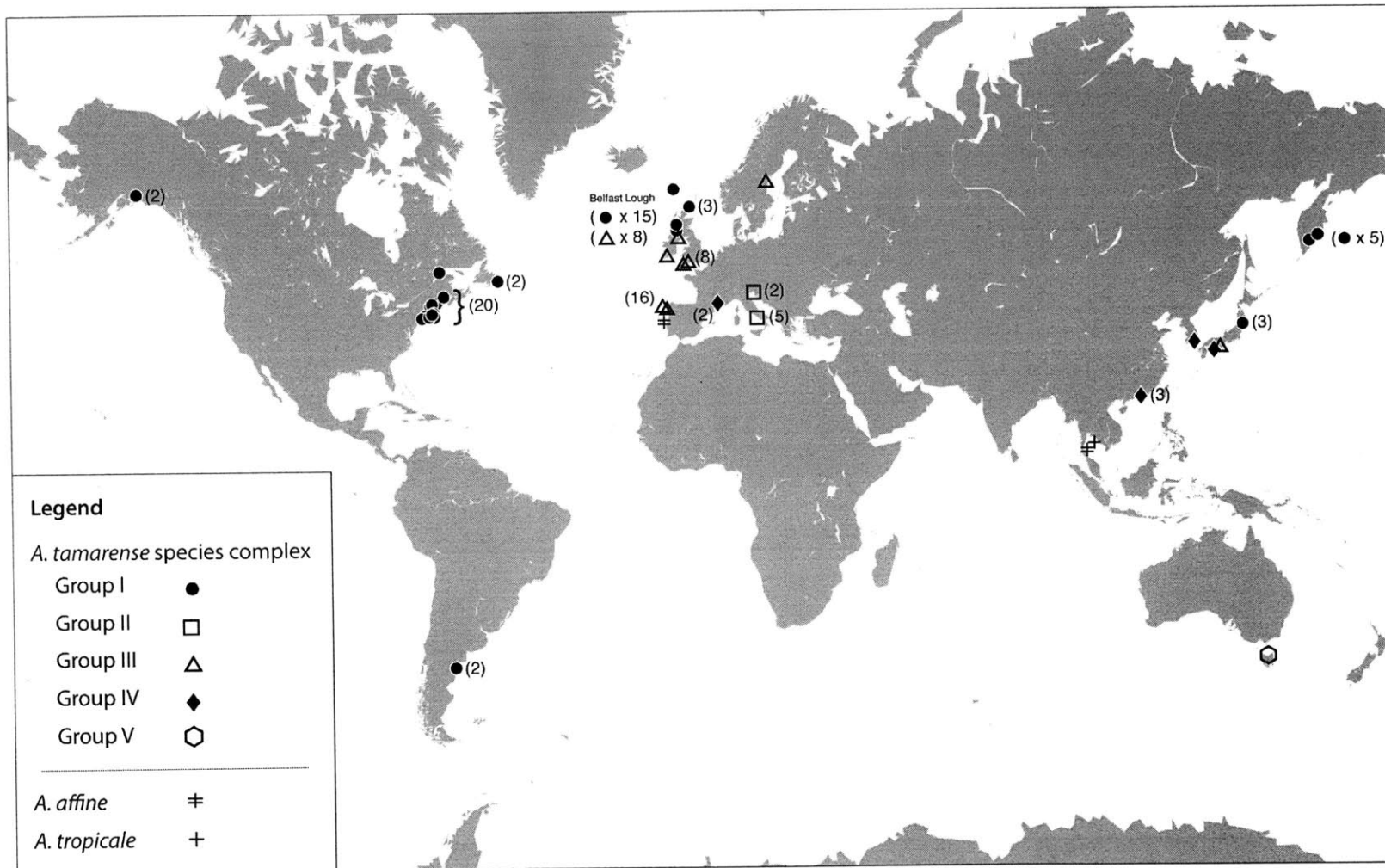
Supplementary Table S1. Clones and apparent encystment compatibilities. Parenthetical group assignments denote those for which no LSU rDNA sequence is available (see Materials and Methods). Co-cultures producing cysts that were subsequently tested for viability (germination) are in bold; the viability of progeny is denoted by superscripts V = viable and NV = inviable.

Clone	Group	Toxic	Extant	Source	Compatibilities	
2TOW1	(I)	Y	No	Portsmouth, NH, USA; plankton tow 1989		Group III: PGT183
38-3	(I)	Y	Yes	Cape Elizabeth, ME, USA; pump sample 1993	Group I: ATRU03, GTCN16, GTLI21, **GTM253-17**[V]	Group III:, ATSP1-B, ATSP1-D, GT5-6, **SP4B6-2**[NV]
3TOW1	(I)	Y	No	Portsmouth, NH, USA; plankton tow 1989		Group III: PGT183, SP4C7-3, SP5B2-3
AFNFA3	I	Y	No	Harbor Grace, New Foundland, Canada; Isolate from cyst germination 1989; Low grade hypnozygote formation in clonal culture	Group I: GT7, GTCA04, GTCA08, GTCA28, GTCN16	
AFNFA4	I	Y	No	Harbor Grace, New Foundland, Canada; Isolate from cyst germination 1989; Low grade hypnozygote formation in clonal culture	Group I: GT-7, GTCA04, GTCA08, GTCA28	Group III: GT5-6, PGT183; *A. affine*: CU-1
AFNS88	(I)	Y	No	Nantucket Shoals, MA, USA; plankton tow 1991		Group III: (PE1V); *A. tropicale*: CU-22
ATAG01	(I)	Y	No	Patagonia, Argentina; Isolate from cyst germination 1993		Group III:, GT5-6, PGT183, SP3B8-3, SP4B5-2
ATAG02	(I)	Y	No	Patagonia, Argentina; Isolate from cyst germination 1993		Group III: PGT183; Group V: ATBB01
ATKLL04	(I)	Y	No	KLL Bay, Orkney Islands, Scotland; Isolate from sediment germination 1999		Group II: SZNB8
ATRU03	I	Y	No	Mohovaya Inlet, Avachinskava Guba, Russia; Isolate from cyst germination 1992	Group I: 38-3, GT-7, GTCA28, GTCN16, GTILI21, GTM253-17	Group III: ATSP1-B, GT5-6, SP4B6-2, SP4E3-1
ATRU04	I	Y	No	Mohovaya Inlet, Avachinskava Guba, Russia; Isolate from cyst germination 1992		Group III: GT5-6
ATRU-29	I	Y	No	Avachinskaya Guba Inlet, Mys Seroglazka, Kamchatka, Russia; 2001		
ATRU3/1	(I)	-	No	Bukhta Tyulen'ye Ozero, Kamchatka, Russia; 2000		
ATRU5/1	(I)	-	No	Bukhta Tyulen'ye Ozero, Kamchatka, Russia; 2000		
AFSL12	(I)	Y	No	Baie Comeau, St. Lawrence Estuary, Canada; Isolate from cyst germination 1991		
BAH ME 182	(I)	-	No	Orkney Islands, Scotland; Single cell isolate 1997	Group I: UW4 Group IV: ATCI01-1	Group III: ATFE6, ATIR-2, ATIR-3, ATSP1-C, ATSP1-D, ATSW01-1, SP4B5-2, SP4E5-4
BAH ME 184	(I)	-	No	Orkney Islands, Scotland; Single cell isolate 1997	Group I: UW4	Group III: ATFE6, ATIR-2, ATIR-3, ATIR-5, ATSP1-D, ATSW01-1, SP4B5-2
GT1	(I)	Y	No	Campobello Island, Bay of Fundy; 1976	Group I: GTME20	
GT2	(I)	Y	No	Campobello Island, Bay of Fundy; 1976	Group I: GTCA04, GTCA28, GTCN16	Group III: GT5-6

GT7	(I)	Y	No	Campobello Island, Bay of Fundy; 1976	Group I: AFNFA3, AFNFA4, ATRU03, GTCA28, GTLI22, GTLI23, GTM253-17, GTSP1	Group III: ATSP1-D, SP4B6-2
GTCA01	(I)	Y	No	Portsmouth, NH, USA; Isolate from cyst germination 1984	Group I: GTCA04, GTCA08, GTCA28, GTCA29, GTLI21, GTLI23, GTSP1	
GTCA04	(I)	Y	No	Portsmouth, NH, USA; Isolate from cyst germination 1984	Group I: AFNFA3, AFNFA4, GT2, GTCA01, GTCA28, GTLI22, GTLI23, GTME20, GTMP 01, GTMR01,	
GTCA08	(I)	Y	No	Portsmouth, NH, USA; Isolate from cyst germination 1984	Group I: AFNFA3, AFNFA4, GTCA01, GTCA29, GTLI22, GTLI23, GTME20, GTSP1	Group III: PGT183
GTCA28	(I)	Y	Yes	Portsmouth, NH, USA; Isolate from cyst germination 1984; Low grade hypnozygote formation in clonal culture	Group I: AFNFA3, AFNFA4, GT2, GT7, GTCN16, GTLI23, GTM253-17, GTME20, GTPP03	Group III: ATSP1-B, ATSP1-C, GT5-6, PE1V, PGT183, SP4B6-2, WKS-1; Group V: ATBB01
GTCA29	(I)	Y	No	Portsmouth, NH, USA; Isolate from cyst germination 1984	Group I: GTCA01, GTCA08, GTMP 01, GTMR01, GTSP1	
GTCN16	(I)	Y	Yes	Mumford Cove, Groton, CT, USA; Isolate from cyst germination 1985; Pellicle cyst formation in clonal culture	Group I: 38-3, AFNFA3, ATRU03, GT-2, GTCA01, GTCA04, GTCA28, GTM253-17, GTME20, GTMP 01, **GTMR01**[V]	Group III: ATSP1-B, ATSP1-D, ATSW01-1, GT5-6, SP3B5-2, SP4B5-2, SP4B6-2, **SP4C7-3**[NV], SP4D7-1, SP4E6-2, **SP5B2-3**[NV]; A. tropicale: CU-22
GTLI21	(I)	Y	Yes	Moriches Bay, Long Island, NY, USA; Isolate from cyst germination 1981	Group I: 38-3, ATRU03, GTCA01, **GTM253-17**[NV]	Group III: ATSP1-B, SP4B6-2
GTLI22	(I)	Y	No	Moriches Bay, Long Island, NY, USA; Isolate from cyst germination 1981	Group I: GT7, GTCA04, GTCA08	
GTLI23	(I)	Y	No	Moriches Bay, Long Island, NY, USA; Isolate from cyst germination 1981	Group I: GT7, GTCA01, GTCA04, GTCA08, GTCA28	
GTM253-17	(I)	Y	Yes	Mitchell River, Orleans, MA, USA; Isolate of GTM253 cyst germination 2003; Pellicle cyst formation in clonal culture	Group I: **38-3**[V], ATRU03, GT7, GTCA28, GTCN16, **GTLI21**[NV]	Group III: ATSP1-D, **ATSW01-1**[NV], **SP4B5-2**[NV]
GTME20	(I)	Y	No	Mohegan Island, ME, USA; Isolate from cyst germination 1982	Group I: GT1, GTCA04, GTCA08, GTCA28, GTCN16	Group III: GT-1
GTMP 01	I	Y	No	Mill Pond, Orleans, MA, USA; Isolate from cyst germination 1978	Group I: GTCA04, GTCA29, GTCN16	Group III: GT5-6
GTMR01	(I)	Y	No	Mitchell River, Orleans, MA, USA; Isolate of GTM253 cyst germination 1986; Pellicle cyst formation in clonal culture	Group I: GTCA04, GTCA29, **GTCN16**[V]	Group III: ATSW01-1, **GT5-6**[NV], SP3B5-2, SP4D7-1, **SP4E5-4**[NV], **SP5B3-1**[NV]
GTPP03	(I)	Y	No	Perch Pond, Falmouth, MA, USA; Single cell isolate from a water sample, 1984	Group I: GTCA28,	
GTSP1	(I)	Y	No	Salt Pond, Eastham, MA, USA; 1980	Group I: GT7, GTCA01, GTCA08, GTCA29	

K-0056	(I)	Y	No	Tjaldavik, Suduroy, Faroe (Faeroe) Islands; Isolate from cyst germination 1996		
MR16B2	(I)	-	No	Isolate from germination of GTCN16 + GTMR01 hypnozygote 1993		Group III: GT5-6
MR16G6	(I)	-	No	Isolate from germination of GTCN16 + GTMR01 hypnozygote 1993		Group III: GT5-6, SP3B5-2, SP4B5-2, SP4E3-1
MR16G9	(I)	-	No	Isolate from germination of GTCN16 + GTMR01 hypnozygote 1993		
OF051	I	Y	Yes	Ofunato Bay, Iwate Prefecture, Japan; 1984		
OF84423-D3	(I)	Y	No	Ofunato Bay, Japan; 1984		Group III: GT5-6
OF87517	(I)	-	No	Ofunato Bay, Japan; 1987		
PW05	(I)	-	No	Port Benny, Prince William Sound, AK, USA		Group III: GT5-6, PGT183
PW06	I	Y	Yes	Port Benny, Prince William Sound, AK, USA		
UW4	I	Y	Yes	Ardtoe, Scottland; Isolated from sediment slurry germination 1992	Group I: BAH ME 182, BAH ME 184; Group IV: ATTL02	Group II: SZNB8; Group III: ATSP1-A, SP4B6-2
UW450	I	Y	No	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		
UW455	I	Y	No	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		Group III: UW475
UW456	I	Y	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		Group III: UW475
UW459	I	Y	No	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		
UW460	I	Y	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		
UW462	I	Y	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW465, UW466, UW473	Group III: UW475
UW464	I	Y	No	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		
UW465	I	Y	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW462, UW466, UW473, UW500	Group III: UW475, UW484
UW466	I	Y	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW462, UW465, UW468	Group III: UW504
UW468	I	-	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW466, UW473	Group III: UW475, UW484
UW473	I	Y	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW462, UW465, UW468	
UW474	I	Y	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		

91

UW485	I	Y	No	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		
UW499	I	Y	No	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		
UW500	I	Y	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW465	
A4T	(II)	-	Yes	Gulf of Trieste, Adriatic Sea; 1997		
A5T	(II)	-	Yes	Gulf of Trieste, Adriatic Sea; 1997		
SZN12	(II)	-	Yes	Gulf of Naples, Itlay; 1999		
SZNB1	II	N	Yes	Gulf of Naples, Itlay; 1999		
SZNB19	II	-	Yes	Gulf of Naples, Itlay; 1999		
SZNB21	II	-	Yes	Gulf of Naples, Itlay; 1999		
SZNB8	II	-	No	Gulf of Naples, Italy; 1999	Group I: ATKLL04, UW4	
ATFE6	III	N	Yes	Fal Estuary, England; 1995	Group I: BAH ME 182, BAH ME 184	
ATFE7	III	N	Yes	Fal Estuary, England; 1995	Group IV: ATTL01	Group III: ATFE10
ATFE10	(III)	-	Yes	Fal Estuary, England; 1995		Group III: ATFE7
ATIR-1	III	N	Yes	Cork Harbor, Ireland; 1997		
ATIR-2	III	-	Yes	Cork Harbor, Ireland; 1997	Group I: BAH ME 182, BAH ME 184	
ATIR-3	III	-	Yes	Cork Harbor, Ireland; 1997	Group I: BAH ME 182, BAH ME 184	
ATIR-4	III	-	Yes	Cork Harbor, Ireland; 1997		
ATIR-5	III	N	Yes	Cork Harbor, Ireland; 1997	Group I: BAH ME 184	
ATSP1-A	(III)	-	Yes	Tetrad daughter isolate from germination of a single ATSW01-1 + SP3B8-3 hypnozygote 1995	Group I: UW4	Group III: ATSW01-1
ATSP1-B	(III)	-	Yes	Tetrad daughter isolate from germination of a single ATSW01-1 + SP3B8-3 hypnozygote 1995	Group I: 38-3, ATRU03, GTCA28, GTCN16, GTLI21	Group III: **ATSW01-1**^{V}
ATSP1-C	(III)	-	Yes	Tetrad daughter isolate from germination of a single ATSW01-1 + SP3B8-3 hypnozygote 1995	Group I: BAH ME 182, GTCA28 Group IV: ATTL01	Group III: SP4B6-2
ATSP1-D	(III)	-	Yes	Tetrad daughter isolate from germination of a single ATSW01-1 + SP3B8-3 hypnozygote 1995	Group I: 38-3, BAH ME 182, BAH ME 184, GT7, GTCN16, GTM253-17	Group III: SP4B5-2, SP4B6-2
ATSW01-1	III	N	No	Gullmar Fjord, Essvik, Sweden; Isolate from cyst germination 1991	Group I: BAH ME 182, BAH ME 184, GTCN16, **GTM253-17**^{NV}, GTMR01 Group IV: ATTL01	Group III: ATSP1-A, **ATSP1-B**^{V}, GT5-6, SP2C9-3, SP3B5-2, **SP3B8-3**^{V}, **SP4B6-2**^{V}, **SP4E6-2**^{V}, SP5B2-3
GT-1	(III)	N	No	La Coruña Bay, Spain; Isolate from cyst germination 1993	Group I: GTME20	
GT5-6	(III)	N	No	La Coruña Bay, Spain; Isolate from cyst germination 1993; Low grade hypnozygote formation in clonal culture	Group I: 38-3, AFNFA4, ATAG01, ATRU03, ATRU04, GT2, GTCA28, GTCN16, GTMP 01, **GTMR01**^{NV}, MR16B2, MR16G6, OF84423-D3, PW05	Group III: ATSW01-1, SP3B5-2, SP4B6-2, SP4C7-3, SP5B2-3

PE1V	(III)	N	No	Ria de Vigo, Galicia, Spain; Isolate from cyst germination 1984	Group I: (AFNS88), GTCA28	
PGT183	III	N	No	Tamar Estuary, Plymouth, England; 1957	Group I: 2TOW1, 3TOW1, AFNFA4, ATAG01, ATAG02, GTCA08, GTCA28, PW05	Group III: SP3B5-2
SP2C9-3	(III)	N	No	La Coruña Bay, Spain; Isolate from cyst germination 1994		Group III: ATSW01-1, SP3B5-2, SP4B5-2, SP4D7-1, SP4E3-1, SP4E5-4, SP5B3-1
SP3B5-2	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994; Low grade hypnozygote formation in clonal culture	Group I: GTCN16, GTMR01, MR16G6	Group III: ATSW01-1, GT5-6, PGT183, SP2C9-3, SP3B8-3, SP4B6-2, SP4E6-2
SP3B8-3	(III)	N	No	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: ATAG01	Group III: **ATSW01-1**^{V}, SP3B5-2, SP4B5-2, SP4D7-1, **SP4E3-1**^{V}, SP5B3-1
SP4B5-2	(III)	-	Yes	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: ATAG01, BAH ME 182, BAH ME 184, GTCN16, **GTM253-17**^{NV}, MR16G6 Group IV: ATCI01-1	Group III: ATSP1-D, SP2C9-3, SP3B8-3, SP4B6-2, SP4E3-1, SP4E6-2
SP4B6-2	(III)	-	Yes	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: **38-3**^{NV}, ATRU03, GT7, GTCA28, GTCN16, GTLI21, GTCN16, UW4	Group III: ATSP1-C, ATSP1-D, **ATSW01-1**^{V}, GT5-6, SP3B5-2, SP4B5-2, SP4D7-1, SP4E3-1, SP4E5-4, SP5B3-1
SP4C7-3	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: 3TOW1, **GTCN16**^{NV}	Group III: GT5-6
SP4D7-1	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: GTCN16, GTMR01	Group III: SP2C9-3, SP3B8-3, SP4B6-2
SP4D10-3	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994		
SP4E3-1	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: ATRU03, MR16G6 Group IV: ATCI01-1	Group III: SP2C9-3, **SP3B8-3**^{V}, SP4B5-2, SP4B6-2, SP4E6-2
SP4E5-4	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: BAH ME 182, **GTMR01**^{NV}	Group III: SP2C9-3, SP4B6-2, **SP4E6-2**^{V}, SP5B2-3
SP4E6-2	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: GTCN16	Group III: **ATSW01-1**^{V}, SP3B5-2, SP4B5-2, SP4E3-1, **SP4E5-4**^{V}, SP5B3-1
SP5B2-3	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994	Group I: 3TOW1, **GTCN16**^{NV}	Group III: ATSW01-1, GT5-6, SP4E5-4
SP5B3-1	(III)	-	No	La Coruña Bay, Spain; Isolate from cyst germination 1994; Low grade hypnozygote yield in clonal culture	Group I: **GTMR01**^{NV}	Group III: SP2C9-3, SP3B8-3, SP4B6-2, SP4E6-2
UW451	III	-	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		Group III: UW452, UW475, UW484
UW452	III	N	No	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		Group III: UW451, UW453, UW480, UW486
UW453	III	N	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		Group III: UW452, UW475, UW484

93

UW475	III	N	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW455, UW456, UW462, UW465, UW468	Group III: UW451, UW453, UW480, UW486, UW504
UW480	III	N	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		Group III: UW452, UW475, UW484
UW484	III	N	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW465, UW468	Group III: UW451, UW453, UW480, UW486
UW486	III	N	Yes	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006		Group III: UW452, UW475, UW484
UW504	III	N	No	Belfast Lough, Northern Ireland; Isolate from germination of sediment slurry 2006	Group I: UW466	Group III: UW475
WKS-1	III	-	No	Tanabe Bay, Japan; 1987	Group I: GTCA28	
WKS-3	(IV)	-	No	Tanabe Bay, Japan; 1989		Group V: ATBB01
ATCI01-1	IV	Y	Yes	Da Yia Bay, Guan-dong (Canton) Province, China; Isolate from cyst germination 1993	Group I: ATRU03	Group III: SP4B5-2, SP4E3-1
ATCI02-1	(IV)	-	Yes	Da Yia Bay, Guan-dong (Canton) Province, China; Isolate from cyst germination 1993		
ATCI03-1	IV	Y	Yes	Da Yia Bay, Guan-dong (Canton) Province, China; Isolate from cyst germination 1993		
ATGHOPE	IV	Y	Yes	Sanchompo, South Korea; Isolated from ballast water 1991		
ATTL01	IV	Y	Yes	Thau Lagoon, France; Single cell isolate 1998	Group IV: ATTL02	Group III: ATFE7, ATSP1-C, ATSW01-1
ATTL02	(IV)	-	Yes	Thau Lagoon, France; Single cell isolate 1998	Group I: UW4 Group IV: ATTL01	
WKS-3	(IV)	-	No	Tanabe Bay, Japan; 1989		Group V: ATBB01
ATBB01	V	N	No	Bell Bay, Tasmania; 1987	Group I: ATAG02, GTCA28 Group IV: WKS-3	
CU-1	A. affine	-	No	Gulf of Thailand; 1984	Group I: AFNFA4	
PA5V	A. affine	-	No	Ria de Vigo, Galicia, Spain; 1985		
CU-22	A. tropicale	-	No	Gulf of Thailand; 1987	Group I: AFNS88, GTCN16	

94

Supplementary Figure S2. Complete matrix of hypnozygote yields from clonal and

pairwise co-cultures. Groups I-V are *A. tamarense* species complex isolates (Lilly et al.

2007). Yields are rated 0 (no encystment), 1 (1-50 cysts per 25 mL co-culture), 2 (50-100

cysts per 25 mL co-culture), or 3 (>100 cysts per 25 mL co-culture). Multiple yield

ratings reflect variable yields in experiments repeated one or more years apart in time.

Group I

	2TOW1	38-3	3TOW1	5-6	7'PUMP3	AFNFA3	AFNFA4	AFNS88	ATAG01	ATAG02	ATKLL04	ATRU03	ATRU04	ATRU-29	ATRU3/1	ATRU5/1	ATSL12/AFSL12	BAH ME 182	BAH ME 184
Group I **62 isolates** 2TOW1	0																		
38-3		0										1							
3TOW1			0																
5-6				0															
7'PUMP3					0														
AFNFA3						0													
AFNFA4							2, 0												
AFNS88								0											
ATAG01									0										
ATAG02										0									
ATKLL04											0							0	0
ATRU03												0							
ATRU04													0						
ATRU-29														0					
ATRU3/1															0				
ATRU5/1																2			
ATSL12/AFSL12																	0		
BAH ME 182																		0	0
BAH ME 184																			0
GT1																			
GT2																			
GT7																			
GTCA01																			
GTCA04																			
GTCA08																			
GTCA28																			
GTCA29																			
GTCN16																			
GTU21																			
GTU22																			
GTU23																			
GTM253-17																			
GTME20																			
GTMP 01																			
GTMR01																			
GTPP03																			
GTSP1																			
K-0056																			
MR16B2																			
MR16G6																			
MR16G9																			
OF051																			
OF84423-D3																			
OF87517																			
PW05																			
PW06																			
UW4																			
UW450																			
UW455																			
UW456																			
UW459																			
UW460																			
UW462																			
UW464																			
UW465																			
UW466																			
UW468																			
UW473																			
UW474																			
UW485																			
UW499																			
UW500																			
Group II **7 isolates** A4T																			
A5T																			
SZN12																			
SZNB1																			
SZNB19																			
SZNB21																			
SZNB8																			
Group III **39 isolates** ATFE10																			
ATFE6																			
ATFE7																			
ATIR-1																			
ATIR-2																			
ATIR-3																			
ATIR-4																			
ATIR-5																			
ATSP1-A																			
ATSP1-B																			
ATSP1-C																			
ATSP1-D																			
ATSW01-1																			
GT5-6																			
PE1V																			
PE2V																			
PGT183																			
SP2C9-3																			
SP3B5-2																			
SP3B8-3																			
SP4B5-2																			
SP4B6-2																			
SP4C7-3																			
SP4D10-3																			
SP4D7-1																			
SP4E5-1																			
SP4E6-4																			
SP4E6-2																			
SP5B2-3																			
SP5B3-1																			
UW451																			
UW452																			
UW453																			
UW475																			
UW480																			
UW484																			
UW486																			
UW504																			
WKS-1																			
Group IV **7 isolates** ATCI01-1																			
ATCI02-1																			
ATCI03-1																			
ATGHOIPE1																			
ATTL01																			
ATTL02																			
WKS-3																			
Group V ATBB01																			
A. affine **2 isolates** CU-1																			
PA5V																			
A. tropicale CU-22																			

Table columns: GT1, GT2, GT7, GTCA01, GTCA04, GTCA08, GTCA28, GTCA29, GTCN16, GTLI21, GTLI22, GTLI23, GTM253-17, GTME20, GTMP 01, GTMR01, GTPP03, GTSP1, K-0056

Group	Isolate	GT1	GT2	GT7	GTCA01	GTCA04	GTCA08	GTCA28	GTCA29	GTCN16	GTLI21	GTLI22	GTLI23	GTM253-17	GTME20	GTMP 01	GTMR01	GTPP03	GTSP1	K-0056
Group I	2TOW1																			
62 isolates	38-3			0				0			1	1		3						
	3TOW1																			
	5-6																			
	7PUMP3																			
	AFNFA3	0	0	3, 0		3	3	3	0	3						0	0	0	0	
	AFNFA4	0	0	2, 3		3	3	3	0							0	0	0	0	
	AFNS88	0																		
	ATAG01																			
	ATAG02																			
	ATKLL04																			0
	ATRU03			1				1, 0		1										
	ATRU04																			
	ATRU-29																			
	ATRU3/1																			
	ATRU5/1																			
	ATSL12/AFSL12	0																	0	
	BAH ME 182																			0
	BAH ME 184																			0
	GT1	0	0	0		0		0		0	0				1			0		
	GT2		0	0		1	1	1	0	3					0	0	0	0	0	
	GT7			0	0	0	2	1	0	0		2			0	0	0	0	3	
	GTCA01					2	2	1	0	0	2		3	1	2	1	0	0	0	
	GTCA04						0	1	1	0	0	3	3		1	0	0	0	1	
	GTCA08							0, 2, 1	0	0	0	0	3		2	0	0	0	0	
	GTCA28								0	1, 0	0	0	0	2	2	0	0	1	0	
	GTCA29									0	0				0	3	1	0	1	
	GTCN16										0		3	2	3	3	3	0		
	GTLI21											0			0					
	GTLI22												0		0					
	GTLI23													0	0					
	GTM253-17														0					
	GTME20															0	0	0	0	
	GTMP 01																0	0		
	GTMR01																	0		
	GTPP03																	0		
	GTSP1																		0	
	K-0056																			0
	MR16B2																			
	MR16G6																			
	MR16G9																			
	OF051																			
	OF84423-D3																			
	OF87517																			
	PW05																			
	PW06																			
	UW4																			
	UW450																			
	UW455																			
	UW456																			
	UW459																			
	UW460																			
	UW462																			
	UW464																			
	UW465																			
	UW466																			
	UW468																			
	UW473																			
	UW474																			
	UW485																			
	UW499																			
	UW500																			
Group II	A4T																			
7 isolates	A5T																			
	SZN12																			
	SZNB1																			
	SZNB19																			
	SZNB21																			
	SZNB8																			
Group III	ATFE10																			
39 isolates	ATFE6																			
	ATFE7																			
	ATIR-1																			
	ATIR-2																			
	ATIR-3																			
	ATIR-4																			
	ATIR-5																			
	ATSP1-A																			
	ATSP1-B																			
	ATSP1-C																			
	ATSP1-D																			
	ATSW01-1																			
	GT5-6																			
	PE1V																			
	PE2V																			
	PGT183																			
	SP2C9-3																			
	SP3B5-2																			
	SP3B8-3																			
	SP4B5-2																			
	SP4B6-2																			
	SP4C7-3																			
	SP4D10-3																			
	SP4D7-1																			
	SP4E3-1																			
	SP4E5-4																			
	SP4E6-2																			
	SP5B2-3																			
	SP5B3-1																			
	UW451																			
	UW452																			
	UW453																			
	UW475																			
	UW480																			
	UW484																			
	UW486																			
	UW504																			
	WKS-1																			
Group IV	ATCI01-1																			
7 isolates	ATCI02-1																			
	ATCI03-1																			
	ATGHOPE1																			
	ATTL01																			
	ATTL02																			
	WKS-3																			
Group V	ATBB01																			
A. affine 2 isolates	CU-1																			
	PA5V																			
A. tropicale	CU-22																			

98

Group	Isolate	MR16B2	MR16G6	MR16G9	OF051	OF84423-D	OF87517	PW05	PW06	UW4	UW450	UW455	UW456	UW459	UW460	UW462	UW464	UW465	UW466	UW468
Group I / 62 isolates	2TOW1																			
	38-3																			
	3TOW1																			
	5-6																			
	7PUMP3																			
	AFNFA3																			
	AFNFA4																			
	AFNS88																			
	ATAG01																			
	ATAG02																			
	ATKLL04									0										
	ATRU03																			
	ATRU04																			
	ATRU-29																			
	ATRU3/1																			
	ATRU5/1																			
	ATSL12/AFSL12																			
	BAH ME 182									2										
	BAH ME 184									2										
	GT1						0													
	GT2																			
	GT7																			
	GTCA01																			
	GTCA04																			
	GTCA08																			
	GTCA28																			
	GTCA29																			
	GTCN16																			
	GTLI21																			
	GTLI22																			
	GTLI23																			
	GTM253-17																			
	GTME20																			
	GTMP 01																			
	GTMR01																			
	GTPP03																			
	GTSP1																			
	K-0056									0										
	MR16B2	0																		
	MR16G6		0																	
	MR16G9			0																
	OF051				0															
	OF84423-D3					0														
	OF87517						0													
	PW05							0												
	PW06								0											
	UW4									0										
	UW450										0									
	UW455											0					0	0	0	0	0
	UW456												0	0			0	0	0	0	0
	UW459													0	0						
	UW460															0					
	UW462																0	0	2	2	0
	UW464																	0	0	0	0
	UW465																		0	2	0
	UW466																			0	1
	UW468																				0
	UW473																				
	UW474																				
	UW485																				
	UW499																				
	UW500																				
Group II / 7 isolates	A4T																			
	A5T																			
	SZN12																			
	SZNB1																			
	SZNB19																			
	SZNB21																			
	SZNB8																			
Group III / 39 isolates	ATFE10																			
	ATFE6																			
	ATFE7																			
	ATIR-1																			
	ATIR-2																			
	ATIR-3																			
	ATIR-4																			
	ATIR-5																			
	ATSP1-A																			
	ATSP1-B																			
	ATSP1-C																			
	ATSP1-D																			
	ATSW01-1																			
	GT5-6																			
	PE1V																			
	PE2V																			
	PGT183																			
	SP2C9-3																			
	SP3B5-2																			
	SP3B8-3																			
	SP4B5-2																			
	SP4B6-2																			
	SP4C7-3																			
	SP4D10-3																			
	SP4D7-1																			
	SP4E3-1																			
	SP4E5-4																			
	SP4E6-2																			
	SP5B2-3																			
	SP5B3-1																			
	UW451																			
	UW452																			
	UW453																			
	UW475																			
	UW480																			
	UW484																			
	UW486																			
	UW504																			
	WKS-1																			
Group IV / 7 isolates	ATCI01-1																			
	ATCI02-1																			
	ATCI03-1																			
	ATGHOPE1																			
	ATTL01																			
	ATTL02																			
	WKS-3																			
Group V	ATBB01																			
A. affine / 2 isolates	CU-1																			
	PASV																			
A. tropicale	CU-22																			

Group	Isolate	UW473	UW474	UW485	UW499	UW500	A4T	A5T	SZN12	SZNB1	SZNB19	SZNB21	SZNB8	ATFE10	ATFE6	ATFE7	ATIR-1	ATIR-2	ATIR-3	ATIR-4
							Group II							Group III						
Group I 62 isolates	2TOW1																			
	38-3																			
	3TOW1																			
	5-6																			
	7'PUMP3																			
	AFNFA3																			
	AFNFA4																			
	AFNS88																			
	ATAG01																			
	ATAG02																			
	ATKLL04						0	0	0	0	0	0	1	0	0	0	0	0	0	0
	ATRU03																			
	ATRU04																			
	ATRU-29																			
	ATRU3/1																			
	ATRU5/1																			
	ATSL12/AFSL12																			
	BAH ME 182						0	0	0	0	0	0	0	0	2	0	0	3	1	0
	BAH ME 184						0	0	0	0	0	0	0	0	1	0	0	1	1	0
	GT1																			
	GT2																			
	GT7																			
	GTCA01																			
	GTCA04																			
	GTCA08																			
	GTCA28																			
	GTCA29																			
	GTCN16																			
	GTLI21																			
	GTLI22																			
	GTLI23																			
	GTM253-17																			
	GTMTME20																			
	GTMP 01																			
	GTMR01																			
	GTPP03																			
	GTSP1																			
	K-0056						0	0	0	0	0	0	0	0	0	0	0	0	0	0
	MR16B2																			
	MR16G6																			
	MR16G9																			
	OF051																			
	OF84423-D3																			
	OF87517																			
	PW05																			
	PW06																			
	UW4						0	0	0	0	0	0	0	0	0	0	0	0	0	0
	UW450		0	0																
	UW455	0		0	0	0														
	UW456	0		0																
	UW459		0	0																
	UW460		0	0																
	UW462	1		0	0	0														
	UW464	0		0	0	0														
	UW465	1		0	0	0														
	UW466	0		0	0	0														
	UW468	1		0	0	0														
	UW473	0		0	0	0														
	UW474		0	0																
	UW485			0	0	0														
	UW499				0	0														
	UW500					0														
Group II 7 isolates	A4T						0													
	A5T							0												
	SZN12								0											
	SZNB1									0										
	SZNB19										0									
	SZNB21											0								
	SZNB8												0							
Group III 39 isolates	ATFE10													0		2				
	ATFE6														0	0				
	ATFE7															0				
	ATIR-1																0			0
	ATIR-2																	0		
	ATIR-3																		0	
	ATIR-4																			0
	ATIR-5																			
	ATSP1-A																			
	ATSP1-B																			
	ATSP1-C																			
	ATSP1-D																			
	ATSW01-1																			
	GT5-6																			
	PE1V																			
	PE2V																			
	PGT183																			
	SP2C9-3																			
	SP3B5-2																			
	SP3B8-3																			
	SP4B5-2																			
	SP4B6-2																			
	SP4C7-3																			
	SP4D10-3																			
	SP4D7-1																			
	SP4E3-1																			
	SP4E5-4																			
	SP4E6-2																			
	SP5B2-3																			
	SP5B3-1																			
	UW451																			
	UW452																			
	UW453																			
	UW475																			
	UW480																			
	UW484																			
	UW486																			
	UW504																			
	WKS-1																			
Group IV 7 isolates	ATCI01-1																			
	ATCI02-1																			
	ATCI03-1																			
	ATGHOPE1																			
	ATTL01																			
	ATTL02																			
	WKS-3																			
Group V	ATBB01																			
A. affine 2 isolates	CU-1																			
	PASV																			
A. tropicale	CU-22																			

	ATIR-5	ATSP1-A	ATSP1-B	ATSP1-C	ATSP1-D	ATSW01-1	GT5-6	PE1V	PE2V	PGT183	SP2C9-3	SP3B5-2	SP3B8-3	SP4B5-2	SP4B6-2	SP4C7-3	SP4D10-3	SP4D7-1	SP4E3-1
Group I 2TOW1							0			3									
62 isolates 38-3		0	2	0	2	0	3			0				0	1				0
3TOW1							0			3						1	0		
5-6							0			0									
7PUMP3							0			0									
AFNFA3							0	0	0	0									
AFNFA4							1, 2	0	0	0, 1, 2									
AFNS88							0	0		0									
ATAG01							3			3	0	0	1	1					0
ATAG02							0			2									
ATKLL04	0	0	0	0	0	0							0	0	0				
ATRU03		0	2	0		0	3			0				0	1	0	0		1
ATRU04							1			0									
ATRU29																			
ATRIU-29																			
ATRU3/1																			
ATRU5'/1																			
ATSL12/AFSL12							0	0	0	0									
BAH ME 182	0	0	0	1	3	3								0	1	0			
BAH ME 184	1	0	0	0	2	2								0	1	0			
GT1																			
GT2							1	0	0	0									
GT7		0	0	0	2	0	0	0	0	0				0	1				0
GTCA01							0	0	0	0									
GTCA04																			
GTCA08										1									
GTCA28		0	1	1	0	0	2, 1	1	0	2, 1				0	1				0
GTCA29										0									
GTCN16		0	2	0	1	2, 1	3, 1, 0	0	0	0	0	2	0	0, 1	0, 1	1	0	0	0
GTLI21		0	1	0	0	0	0	0	0	0				0	1				0
GTLI22																			
GTLI23																			
GTM253-17		0	0	0	2	2								3	0				0
GTME20							0	0	0	0									
GTMP01							0, 1			0									
GTMR01						0, 1	3	0	0	0	0	0, 3	0	0	0	0	0	1	0
GTPP03										0									
GTSP1																			
K-0056	0	0	0	0	0	0							0	0	0				
MR16B2							1			0									
MR16G6							3			0	0	1	0	2	0			0	2
MR16G9							0			0									
OF051																			
OF84423-D3										0									
OF87517							0	0	0	0									
PW05							1			1									
PW06							0			0									
UW4	0	1	0	0	0	0							0	0	1				
UW450																			
UW455																			
UW456																			
UW459																			
UW460																			
UW462																			
UW464																			
UW465																			
UW466																			
UW468																			
UW473																			
UW474																			
UW485																			
UW499																			
UW500																			
Group II A4T																			
7 isolates A5T																			
SZN12																			
SZNB1																			
SZNB19																			
SZNB21																			
SZNB8																			
Group III ATFE10																			
39 isolates ATFE6																			
ATFE7																			
ATIR-1																			
ATIR-2																			
ATIR-3																			
ATIR-4																			
ATIR-5	0																		
ATSP1-A		0	0	0	0	0, 3								0	0				0
ATSP1-B			0	0	0	2, 3								0	0				0
ATSP1-C				0	0	0								0	1				0
ATSP1-D					0	0								0, 2	2				0
ATSW01-1						0	3			0	3, 2	2	3	0	2, 3		0	0	0
GT5-6							1, 0					2			1	1	0		
PE1V								0											
PE2V									0										
PGT183										0	0	3	0	0	0		0	0	0
SP2C9-3											0	3		2, 3	2	0	0	2	1
SP3B5-2												1, 0	2	2					
SP3B8-3													0	2	2	0	0	2	3, 2
SP4B5-2														0	0, 3, 2	0	0	2	2
SP4B6-2															0	0	0	1	3, 0
SP4C7-3																0	0		
SP4D10-3																	0		
SP4D7-1																		0	
SP4E3-1																			0
SP4E5-4																			
SP4E6-2																			
SP5B2-3																			
SP5B3-1																			
UW451																			
UW452																			
UW453																			
UW475																			
UW480																			
UW484																			
UW486																			
UW504																			
WKS-1																			
Group IV ATCI01-1																			
7 isolates ATCI02-1																			
ATCI03-1																			
ATGHOPE1																			
ATTL01																			
ATTL02																			
WKS-3																			
Group V ATBB01																			
A. affine CU-1																			
2 isolates PA5V																			
A. tropicale CU-22																			

Table with 20 data columns. Row group labels on left.

Group	Isolate	SP4E5-4	SP4E6-2	SP5B2-3	SP5B3-1	UW451	UW452	UW453	UW475	UW480	UW484	UW486	UW504	WKS-1	ATCI01-1	ATCI02-1	ATCI03-1	ATGHOPE1	ATTL01	ATTL02
Group I 62 isolates	2TOW1																			
	38-3														0					
	3TOW1			1																
	5-6																			
	7PUMP3																			
	AFNFA3													0						
	AFNFA4													0						
	AFNS88													0						
	ATAG01	0	0		0															
	ATAG02																			
	ATKLL04	0																	0	0
	ATRU03			0											1					
	ATRU04																			
	ATRU-29														0					
	ATRU3/1														0					
	ATRU5/1																			
	ATSL12/AFSL12													0						
	BAH ME 182	1																	0	0
	BAH ME 184	0																		
	GT1																			
	GT2													0						
	GT7													0						
	GTCA01													0						
	GTCA04																			
	GTCA08																			
	GTCA28													1	0					
	GTCA29																			
	GTCN16	0	0, 1	0, 3	0									0	0					
	GTLI21													0						
	GTLI22																			
	GTLI23														0					
	GTM253-17																			
	GTME20													0						
	GTMP 01																			
	GTMR01	2, 3	0	0	2									0						
	GTPP03																			
	GTSP1																			
	K-0056	0																	0	0
	MR16B2																			
	MR16G6	0	0		0															
	MR16G9																			
	OF051														0					
	OF84423-D3																			
	OF87517													0	0					
	PW05																			
	PW06																			
	UW4	0																	0	1
	UW450					0	0	0		0	0	0								
	UW455									1										
	UW456									2	0	0	0							
	UW459					0	0	0		0	0	0								
	UW460					0	0			0	0	0								
	UW462									1			0							
	UW464									0			0							
	UW465									1	3		1							
	UW466									0			0							
	UW468									2	0	1								
	UW473									0	0		0							
	UW474					0	0	0		0	0	0								
	UW485					0	0	0		0	0	0	0							
	UW499									0			0							
	UW500									0			0							
Group II 7 isolates	A4T																		0	0
	A5T																		0	0
	SZN12																		0	0
	SZNB1																		0	0
	SZNB19																		0	0
	SZNB21																		0	0
	SZNB8																		0	0
Group III 39 isolates	ATFE10																		0	0
	ATFE6																		2	2
	ATFE7																		2	2
	ATIR-1																		0	0
	ATIR-2																		0	0
	ATIR-3																		0	0
	ATIR-4																		0	0
	ATIR-5																		0	0
	ATSP1-A														0				0	0
	ATSP1-B														0				3	1
	ATSP1-C														1				0	0
	ATSP1-D														0				0	0
	ATSW01-1	0	3, 1	2	0														1	0
	GT5-6			0, 1											0					
	PE1V																			
	PE2V																			
	PGT183	0	0	0	0										0					
	SP2C9-3	1	2																	
	SP3B5-2				2															
	SP3B8-3	0		0	2, 0															
	SP4B5-2		1												1				0	0
	SP4B6-2	1	1	0	3, 2														0	0
	SP4C7-3		0																	
	SP4D10-3		0																	
	SP4D7-1		0																	
	SP4E3-1		3, 1												1					
	SP4E5-4	0	3, 2	2															0	0
	SP4E6-2		0	0	2															
	SP5B2-3			0											0					
	SP5B3-1				0, 1															
	UW451					0	2	0	2	0	2	0								
	UW452						0	1	0	3	0	0								
	UW453							0	1	0	2	0								
	UW475								0	2	0	0	3							
	UW480									0	3	0	0							
	UW484										0	3	0							
	UW486											0	0							
	UW504												0							
	WKS-1													0						
Group IV 7 isolates	ATCI01-1														0	0	0	0		
	ATCI02-1															0	1			
	ATCI03-1																0			
	ATGHOPE1																	0		
	ATTL01																		0	1
	ATTL02																			0
	WKS-3																			
Group V	ATBB01																			
A. affine 2 isolates	CU-1																			
	PASV																			
A. tropicale	CU-22																			

		WKS-3	Group V ATBB01	A. affine CU-1	PA5V	A. tropicale CU-22
Group I	2TOW1		0			
62 isolates	38-3		0			
	3TOW1		0			
	5-6		0			
	7'PUMP3		0			
	AFNFA3		0	0	0	0
	AFNFA4		0	1, 2	0	0
	AFNS88		0	0		1, 2
	ATAG01		0			
	ATAG02		1			
	ATKLL04					
	ATRU03		0			
	ATRU04		0			
	ATRU-29					
	ATRU3/1					
	ATRU5/1					
	ATSL12/AFSL12		0	0	0	0
	BAH ME 182					
	BAH ME 184					
	GT1	0				
	GT2		0	0	0	0
	GT7		0	0	0	0
	GTCA01					
	GTCA04		0	0	0	0
	GTCA08				0	
	GTCA28		3, 1	0	0	0
	GTCA29				0	
	GTCN16		0	0	0	1
	GTLI21		0	0		0
	GTLI22					
	GTLI23					
	GTM253-17					
	GTME20		0	0	0	0
	GTMP 01				0	
	GTMR01		0	0	0	0
	GTPP03			0	0	
	GTSP1					
	K-0056					
	MR16B2		0			
	MR16G6		0			
	MR16G9		0			
	OF051					
	OF84423-D3		0			
	OF87517		0	0		0
	PW05		0			
	PW06		0			
	UW4					
	UW450					
	UW455					
	UW456					
	UW459					
	UW460					
	UW462					
	UW464					
	UW465					
	UW466					
	UW468					
	UW473					
	UW474					
	UW485					
	UW499					
	UW500					
Group II	A4T					
7 isolates	A5T					
	SZN12					
	SZNB1					
	SZNB19					
	SZNB21					
	SZNB8					
Group III	ATFE10					
39 isolates	ATFE6					
	ATFE7					
	ATIR-1					
	ATIR-2					
	ATIR-3					
	ATIR-4					
	ATIR-5					
	ATSP1-A					
	ATSP1-B					
	ATSP1-C					
	ATSP1-D					
	ATSW01-1		0			
	GT5-6	0				
	PE1V	0				
	PE2V	0				
	PGT183	0			0	
	SP2C9-3					
	SP3B5-2					
	SP3B8-3					
	SP4B5-2					
	SP4B6-2					
	SP4C7-3					
	SP4D10-3					
	SP4D7-1					
	SP4E3-1					
	SP4E5-4					
	SP4E6-2					
	SP5B2-3					
	SP5B3-1					
	UW451					
	UW452					
	UW453					
	UW475					
	UW480					
	UW484					
	UW486					
	UW504					
	WKS-1	0				
Group IV	ATCI01-1		0			
7 isolates	ATCI02-1					
	ATCI03-1					
	ATGHOPE1					
	ATTL01					
	ATTL02					
	WKS-3	0	1, 0	0		0
Group V	ATBB01		0			
A. affine	CU-1			0		
2 isolates	PA5V				0	
A. tropicale	CU-22					0

103

Supplementary Figure S3. Example solutions from application of a fixed, dioecious

mating system model to encystment data presented in Figure 1A-D. Positive co-cultures

of each isolate from the (A) unbiased cosmopolitan, (B) biased cosmopolitan, (C)

unbiased Belfast Lough, and (D) biased Belfast Lough subsets are listed. The clone

names of isolates that have been assigned a single mating type are boxed in white (mating

type +) or black (mating type −) on the basis of less constrained, positive co-cultures.

Isolates that could not be classified as a single type (i.e. have + and − type co-cultures)

are classified as ± and are boxed in gray.

A Unbiased Cosmopolitan

Isolate	Positive co-cultures
GTPP03	GTCA28
GTCA29	GTCA08, GTMR01
GT2	GTCA04, GTCA28, GTCN16
GT7	AFNFA3, AFNFA4, GTCA28
GTME20	GTCA04, GTCA08, GTCN16
GTMR01	GTCA04, GTCA29, GTCN16
AFNFA4	GT7, GTCA04, GTCA08, GTCA28
GTCA08	AFNFA3, AFNFA4, GTCA29, GTME20
AFNFA3	GT7, GTCA04, GTCA08, GTCA28, GTCN16
GTCN16	AFNFA3, GT2, GTCA28, GTME20, GTMR01
GTCA04	AFNFA3, AFNFA4, GT2, GTCA28, GTME20, GTMR01
GTCA28	AFNFA3, AFNFA4, GT2, GT7, GTCA04, GTCN16, GTME20, GTPP03

B Biased Cosmopolitan

Isolate	Positive co-cultures
ATSP1-A	ATSW01-1
ATSP1-C	GTCA28, SP4B6-2
SP4E3-1	ATRU03, SP4B5-2, SP4B6-2
ATSW01-1	GTCN16, GTM253-17, ATSP1-A, ATSP1-B, SP4B6-2
GT7	ATRU03, GTCA28, GTM253-17, ATSP1-D, SP4B6-2
GTLI21	38-3, ATRU03, GTM253-17, ATSP1-B, SP4B6-2
SP4B5-2	GTCN16, GTM253-17, ATSP1-D, SP4B6-2, SP4E3-1
ATSP1-B	38-3, ATRU03, GTCA28, GTCN16, GTM253-17, GTLI21, ATSW01-1
ATSP1-D	38-3, GT7, GTCN16, GTM253-17, SP4B5-2, SP4B6-2
38-3	ATRU03, GTLI21, GTM253-17, ATSP1-B, ATSP1-D, SP4B6-2
GTCA28	ATRU03, GT7, GTCA28, GTCN16, ATSP1-B, ATSP1-C, SP4B5-2
ATRU03	ATRU03, GT7, GTCA28, GTCN16, GTM253-17, ATSP1-B, ATSP1-D, SP4E3-1
GTCN16	38-3, GT7, GTCA28, GTM253-17, ATSP1-D, ATSW01-1, SP4B5-2, SP4B6-2
GTM253-17	38-3, ATRU03, GTCA28, GTCN16, GTLI21, ATSP1-D, ATSW01-1, SP4B5-2
SP4B6-2	38-3, GT7, GTCA28, GTCN16, GTLI21, ATSP1-C, ATSP1-D, ATSW01-1

C Unbiased Belfast

Isolate	Positive co-cultures
UW450	
UW459	
UW460	
UW474	
UW485	
UW451	UW452, UW475, UW484
UW453	UW452, UW475, UW484
UW480	UW452, UW475, UW484
UW486	UW452, UW475, UW484
UW452	UW451, UW453, UW480, UW486
UW475	UW451, UW453, UW480, UW486
UW484	UW451, UW453, UW480, UW486

D Biased Belfast

Isolate	Positive co-cultures
UW464	
UW485	
UW499	
UW455	UW475
UW456	UW475
UW500	UW465
UW480	UW475, UW484
UW504	UW466, UW475
UW473	UW462, UW465, UW468
UW484	UW465, UW466, UW468, UW480
UW466	UW462, UW465, UW468, UW473, UW504
UW462	UW465, UW466, UW473, UW475
UW468	UW466, UW473, UW475, UW484
UW465	UW462, UW466, UW473, UW500, UW475, UW484
UW475	UW455, UW456, UW462, UW465, UW468, UW480, UW504

Supplementary Figure S4. (A-D) Mating-type solution of the biased cosmopolitan subset by segregation according to negative encystment results from pairwise subset co-cultures (i.e. incompatibilities of each clone, listed in A). (B) GTCA28 could not be classified because it was found to be auto-compatible. The remaining isolates were segregated into a minimum of four mating types of mutually incompatible isolates. Isolates that are segregated to a single group are in bold; those that could be assigned to more than one group are in regular font. (C) Group I isolates are segregated to four groups of mutually incompatible clones when considered alone (GTCA again omitted). (D) Group III isolates are segregated to three groups of mutually incompatible clones when considered alone. (E and F) Mating-type solution of the biased Belfast Lough Group I isolates by segregation according to negative encystment results. (F) Isolates UW455, UW456, UW464, UW485 and UW499 were not constrained (no positive encystment findings) and were not segregated. The remaining isolates are segregated into three groups of mutually incompatible clones.

A

Biased Cosmopolitan incompatibilities

		Group I								Group III				
38-3	ATRU03	GT7	GTCA28	GTCN16	GTLI21	GTM253-17	ATSP1-A	ATSP1-B	ATSP1-C	ATSP1-D	ATSW01-1	SP4B5-2	SP4B6-2	SP4E3-1
GT7	ATSP1-A	38-3	38-3	GT7	GT7	ATSP1-A	38-3	GT7	38-3	ATRU03	38-3	38-3	GTM253-17	38-3
GTCA28	ATSP1-C	GTCN16	GTLI21	GTLI21	GTCA28	ATSP1-B	ATRU03	GTM253-17	ATRU03	GTCA28	ATRU03	ATRU03	ATSP1-A	GT7
ATSP1-A	ATSP1-D	GTLI21	ATSP1-A	ATSP1-A	GTCN16	ATSP1-C	GT7	ATSP1-A	GT7	GTLI21	GT7	GT7	ATSP1-B	GTCA28
ATSP1-C	ATSW01-1	ATSP1-A	ATSP1-D	ATSP1-A	ATSP1-C	SP4B6-2	GTCN16	ATSP1-C	GTCN16	ATSP1-A	GTLI21	GTLI21		GTCN16
ATSW01-1	SP4B5-2	ATSP1-B	ATSW01-1	SP4E3-1	ATSP1-C	SP4E3-1	GTLI21	ATSP1-D	GTLI21	ATSP1-B	ATSP1-C	ATSP1-A		GTLI21
SP4B5-2		ATSP1-C	SP4B5-2		ATSP1-D		GTM253-17	SP4B5-2	GTM253-17	ATSP1-C	ATSP1-D	ATSP1-B		GTM253-17
SP4E3-1		ATSW01-1	SP4E3-1		ATSW01-1		ATSP1-B	SP4B6-2	ATSP1-A	ATSW01-1	ATSP1-C			ATSP1-A
		SP4B5-2			SP4B5-2		ATSP1-C	SP4E3-1	ATSP1-B	SP4B5-2	ATSP1-D			ATSP1-B
		SP4E3-1			SP4E3-1		ATSP1-D		ATSP1-D	SP4E3-1	ATSW01-1			ATSP1-C
							SP4B5-2		ATSW01-1					ATSP1-D
							SP4B6-2		SP4B5-2					ATSW01-1
							SP4E3-1		SP4E3-1					

B

Biased Cosmopolitan segregation by all incompatibilities

A	B	C	D	E
38-3	ATRU03	GTCN16	GTM253-17	GTCA28
SP4B5-2	ATSP1-D	GTLI21	ATSP1-B	
		SP4E3-1	SP4B6-2	
ATSP1-A		ATSP1-A	ATSP1-A	
ATSP1-C	ATSP1-C	ATSP1-C		
ATSW01-1	ATSW01-1			
GT7		GT7		

C

Biased Cosmopolitan Group I only segregation

A	B	C	D	E
38-3	ATRU03	GTCN16	GTM253-17	GTCA28
		GTLI21		
GT7		GT7		

D

Biased Cosmopolitan Group III only segregation

A	B	C
SP4B5-2	ATSP1-D	SP4B6-2
	SP4E3-1	
ATSP1-A	ATSP1-A	ATSP1-A
ATSP1-B	ATSP1-B	ATSP1-B
ATSP1-C	ATSP1-C	
ATSW01-1	ATSW01-1	

E

Biased Belfast Group I only incompatibilities

UW462	UW465	UW466	UW468	UW473	UW500
UW468	UW468	UW473	UW462	UW466	UW462
UW500		UW500	UW465	UW500	UW466
			UW500		UW468
					UW473

F

Biased Belfast Group I only segregation

A	B	C	D
UW462	UW465	UW466	UW465
		UW473	UW466
			UW464
			UW485
UW468	UW468	UW500	UW499
	UW500		

Chapter 2: Application of imaging flow cytometry for species-specific measurement of DNA content in dinoflagellate microplankton

Michael L. Brosnahan[1], Shahla Farzan[2], Bruce A. Keafer[1], Robert J. Olson[1], Heidi M. Sosik[1], and Donald M. Anderson[1]

1. Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543

2. Department of Biology, Mount Holyoke College, South Hadley, MA 01075

ABSTRACT

Measurements of the distribution of DNA contents in protist populations can shed light on a variety of processes, including cell division, sex, the ingestion of prey, and parasite invasion. Here, we have modified Imaging Flow Cytobot (IFCB), a custom built flow cytometer that records images of microplankton particles, to measure DNA content of large dinoflagellates and other microplankton species with exceptionally high intracellular DNA mass. Additionally, the IFCB was configured to positively identify *Alexandrium tamarense* Group I (*A. t.* Group I), a photosynthetic dinoflagellate that causes Paralytic Shellfish Poisoning. The modified IFCB was used to analyze samples from the growth, peak and termination phases of an *A. t.* Group I bloom in Salt Pond (Eastham, MA, USA) during the spring of 2009 and from an offshore *A. t.* Group I 'red tide' that discolored waters near Portsmouth Harbor, ME, USA in July of 2009. The impact of infection by the dinoflagellate parasite *Amoebophrya* spp. on *A. t.* Group I cells was also investigated using cultures and field samples that were heavily impacted by these parasites. Persistent enrichment of 2c *A. t.* Group I cells at the surface in samples from Salt Pond and in the Portsmouth red tide population suggested an important role for aggregation near the air-sea interface during sexual events. However, evidence of postzygotic replication, a prerequisite for meiosis, only occurred near the bottom of Salt Pond in a population that was also heavily impacted by the *Amoebophrya* parasites. The observations suggest that gametogenesis and fusion likely occur near the surface but that zygotic maturation is completed after migration to depth. High rates of *Amoebophrya*

infection in planozygotes relative to vegetative cells suggest that planozygotes may be especially susceptible to *Amoebophyra* infection, or that they persist longer than vegetative cells after being infected. The modified IFCB system also detected and measured the DNA content of several other dinoflagellates including *Dinophysis acuminata* and also a tintinnid ciliate.

INTRODUCTION

The dinoflagellates are a morphologically and metabolically diverse protist group that sometimes endanger human health through the production of potent neurological and gastrointestinal toxins. They are among the most abundant eukaryotic microbes globally and sometimes dominate planktonic ecosystems as 'red tides', dense monospecific blooms that dramatically alter the color of surface water. Comprehensive knowledge of the dinoflagellates' ecology is needed to understand plankton dynamics and also to inform mitigation strategies against harmful species. About half of the dinoflagellates are photosynthetic (autotrophic or mixotrophic) and include planktonic and benthic species as well as coral and animal endosymbionts. Heterotrophic forms include free-swimming grazers, phagotrophs, and parasites. Cells range in size from <3 μm to 1 mm in diameter (Lin, Zhang et al. 2006). The majority are motile and many planktonic species migrate vertically in the water column (e.g. Hasle, 1950). Most important for this study, dinoflagellates share a number of unique nuclear characteristics and have among the largest genomes of any organisms: DNA mass per cell is comparable to the multinucleated ciliates and two to three orders of magnitude greater than other planktonic algae (Prescott 1994; Hou and Lin 2009).

Here, we have adapted the Imaging FlowCytobot (IFCB) to record genus- and sometimes species-specific DNA content of dinoflagellate cells. IFCB is a custom-built flow cytometer that records particle images in addition to standard measurements of laser scatter or fluorescence (Olson and Sosik 2007; Sosik and Olson 2007). It was configured to measure DNA content via fluorescence of the membrane permeable dye Hoechst 33342 and also to detect fluorescence from species-specific oligonucleotide probes. Application of a stringent trigger threshold for Hoechst fluorescence maximized sample throughput while targeting dinoflagellates. The system was especially powerful for analysis of complex species assemblages in which different species' scatter and fluorescence properties overlap. An associated image-classification system facilitated positive identification of both rare and abundant dinoflagellates so that their fluorescence and image metrics could be compared between samples.

Changes in DNA content within a species may be caused by any of several ecologically important processes: population growth (mitosis), sexual recombination (zygosis/meiosis), ingestion of other cells as prey, or infection by parasites. Of these, changes due to mitosis or sex are most recognizable because they are large and quantized (approximately 2-fold with each replication, division or fusion) causing a population undergoing mitosis or sex to have two or more distinct modes in the distribution of its DNA mass. However the incorporation of prey cells or parasites might obscure multimodal patterns. Exceptions can occur when the DNA mass of an incorporated cell is comparable to the host dinoflagellate or when an infective parasite proliferates to a narrowly defined population size range (e.g. Janse et al., 1987).

Figure 1. Life cycle diagram of the *Alexandrium tamarense* complex species. The life cycle is made up of a mitotic, reproductive cycle and a slower, sexual cycle. Each stage in the reproductive cycle is haploid (1n) and contains either a single or double complement of chromatids (1c or 2c DNA content). Sexual cycle stages may contain 1c, 2c, or 4c DNA content and its zygotic stages are diploid (2n; N.B. the DNA content of the hypnozygote stage has not been shown experimentally). The G2 phase of the life cycle is uniquely bipotent because it may give rise to two mitotic daughter cells or two gametes. Dashed lines show alternate hypothesized routes for re-entry to the reproductive cycle following germination of the hypnozygote.

REPRODUCTION   RECOMBINATION

DNA CONTENT

1c

G1 cells
1n

Gametes
1n

Meiosis II   S Phase   M Phase

2c

G2 cell
1n

Planozygote
2n

Meiosis I   'Dinoflagellate Meiosis'
             Beam and Himes 1980

Premeiotic Replication

4c

Hypnozygote
2n
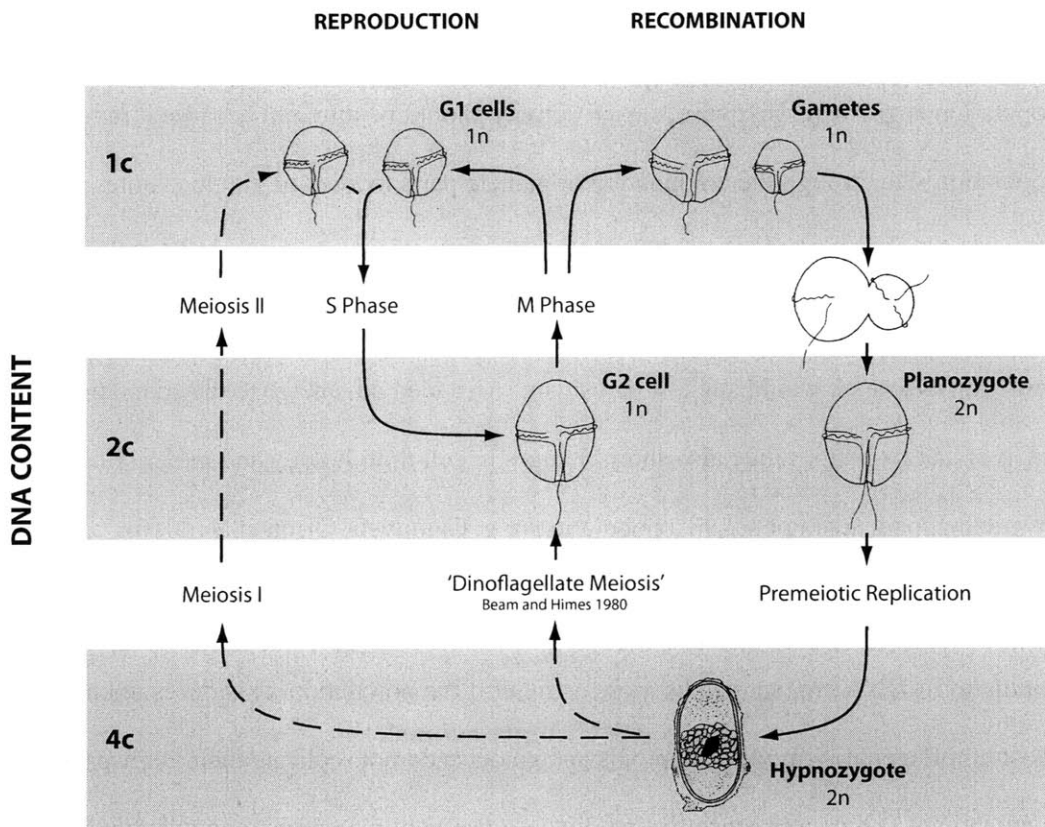
The relative likelihood that mitosis, sex or incorporation of other cells affects a

population's DNA content distribution depends on its metabolism (i.e. does the species

ingest other cells?), and also on its recent history. In the case of a growing autotrophic

population, two modes in DNA content corresponding to the G1 and G2/M cell cycle

phases are expected (1c and 2c respectively, Figure 1). Further, if a population's growth

rate is near maximal, any impact from parasites should be negligible because the

population is growing. Likewise, sexual activity should be minimal because it reduces

population size through the conjugation of gamete pairs to become single zygotes.

Conversely, if a population is declining the appearance of new modes could indicate

sexual processes (e.g. Armbrust, Chisholm et al. 1990). On the other hand, less ordered

shifts in DNA mass would suggest parasitism. An added advantage to the adaptation of

IFCB is that its images may also show changes in cell morphology and integrity caused

by grazing, parasitism or sex in a population (e.g. Campbell, Olson et al. 2010).

The discrimination of sex and mitosis from the distribution of a dinoflagellate

population's DNA content can be more difficult if the population's history is unknown.

Most dinoflagellates divide as haploids and gametes do not replicate their DNA prior to

conjugation. As a consequence the two modes observed in a growing population coincide

with modes also found during sexual events. Gametes are equivalent in DNA content to

vegetative cells in the G0/G1 phase of the cell cycle (1c), and newly formed zygotes

(planozygotes) are equivalent to G2/M phase cells (2c, see Figure 1; Cembella and Taylor

1985; Bhaud, Soyer-Gobillard et al. 1988; Cetta and Anderson 1990). Positive

identification of sexually active populations would be unambiguous if an additional 4c

114

mode were observed. A 4c mode can occur due to premeiotic DNA replication, a prerequisite for meiotic division in sexual eukaryotes. Surprisingly however, premeiotic DNA replication has been shown directly in only one dinoflagellate species, the autotroph *Prorocentrum micans* (Bhaud, Soyer-Gobillard et al. 1988). It is also possible that premeiotic replication occurs after a species has settled to the bottom in a non-swimming hypnozygote stage (Pfiester 1989).

Still, a planktonic population's DNA distribution can be used to distinguish sex and mitosis in the absence of a detectable 4c mode by considering time of sampling. Many autotrophic species have been shown to synchronize their cell cycle with the circadian light-dark cycle; and cell and sexual phases may vary substantially in duration (Van Dolah, Leighfield et al. 1995; Reguera, Bravo et al. 1996; Taroncher-Oldenburg, Kulis et al. 1997; Van Dolah and Leighfield 1999; Van Dolah, Leighfield et al. 2008). Knowledge of a species' cell cycle timing can be used to interpret patterns in its DNA distribution. If G2 phase cells are abundant only during restricted periods during the day, sampling times may be chosen to maximize the ability to detect sexual phase cells at other times.

In this study, we focused on the harmful algal species, *Alexandrium tamarense* Group I (*A. t.* Group I) because differences in the timing and duration of cell and sexual cycle phases are well studied, and also because *A. t.* Group I is a causative organism of Paralytic Shellfish Poisoning (PSP). Consumption of shellfish that are contaminated by PSP toxins can cause paralysis or death, making PSP one of the most dangerous seafood-

poisoning syndromes known. Consequently, *A. t.* Group I is the focal species of ongoing field studies in the Gulf of Maine (GoM) and within the Nauset Marsh system (NMS; Cape Cod, MA). Samples analyzed in this study were taken in conjunction with these efforts and with the foremost goal of describing their *A. t.* Group I populations.

Both the initiation and termination of *A. t.* Group I blooms are highly dependent on exit from and entry into the sexual cycle (Figure 1). Under certain conditions, *A. t.* Group I vegetative cells cease mitotic division and transform into gametes that are morphologically similar to their vegetative progenitors. Gametes conjugate in pairs and eventually form benthic hypnozygote cysts that are highly resistant to stress. The durability of its hypnozygote stage allows *A. t.* Group I to persist in areas that experience long periods of poor growth conditions, such as cold winters (Anderson and Wall 1978; Anderson, Chisholm et al. 1983). *A. t.* Group I exits the sexual cycle through meiotic division of a planomeiocyte, a swimming cell that emerges from its cyst stage (Pfiester and Anderson 1987). Vegetative cells arise from the planomeiocyte and resume mitosis in no more than three divisions (Brosnahan, Kulis et al. 2010). These mitotic cells may migrate vertically to optimize nutrient acquisition and irradiance exposure within the euphotic zone (Anderson and Stolzenbach 1985; Passow 1991; MacIntyre, Cullen et al. 1997), but it is not known whether sexual stages of *A. t.* Group I also migrate vertically to aggregate in sexually active layers. Uncertainty regarding the behavior of sexual stages is due to the difficulty in distinguishing vegetative, gamete and planozygote stages. The modified IFCB instrument described here enables a high throughput assessment of the life cycle transition from population growth to zygote formation as it occurs in the ocean.

Such methods are needed to describe the timing, location, and extent of zygote formation because it is an important process for predicting bloom termination and the recruitment of new cysts (Anderson, Stock et al. 2005; McGillicuddy, Anderson et al. 2005).

Prior to this study it had not been shown whether premeiotic replication occurred while *A. t.* Group I was in the plankton or after its zygotes had settled on the bottom as resting cysts. However, it was expected that *A. t.* Group I planozygotes and vegetative cells could be distinguished by DNA content without detection of a 4c mode because its G1 phase is longer than its other cell cycle phases across a wide range of growth conditions (Taroncher-Oldenburg, Kulis et al. 1999). Moreover, cell division by *A. t.* Group I is slow relative to many other algae (< 0.4 day^{-1}) and is synchronized to darkness via a light-dependent transition point early in the G1 phase of its cell cycle (Taroncher-Oldenburg, Kulis et al. 1997). The net effect of these behaviors is that - in a growing population of *A. t.* Group I - G2+M cells are only abundant during early morning (Tomas 1974; Taroncher-Oldenburg, Kulis et al. 1997). In contrast, *A. t.* Group I planozygotes (also 2c in DNA content) remain in the plankton for approximately one week before transforming to hypnozygote cysts (Anderson, Chisholm et al. 1983; Anderson and Lindquist 1985). Thus, if *A. t.* Group I cells with 2c DNA content are abundant during daylight hours, they are most likely planozygotes and may have accumulated over a period of several days.

A complication to this line of investigation is parasitic infection by the dinoflagellate genus *Amoebophrya* on *A. t.* Group I cells. *Amoebophrya* are intracellular

parasites that chronically infect many dinoflagellate species and may be a dominant

factor in the termination of inshore blooms (Taylor 1968; Coats 1999; Chambouvet,

Morin et al. 2008). Free-living *Amoebophrya* are small, 3 to 5 μm in diameter, and infect

their hosts as single cell dinospores. Once inside, a dinospore undergoes several rounds

of replication to become a large, multi-nucleated trophont that distends its host, filling

much of the host cytoplasm or nucleus as it matures. Ultimately, the host cell bursts and

hundreds of new infective dinospores emerge, suggesting that mature infections could

have a large impact on the DNA content and size of their hosts. We addressed this

question by examining the effects of *Amoebophrya* infection on the DNA content and

size of *A. t.* Group I cells in culture. We also examined natural plankton samples for the

presence of *Amoebophrya* spp. trophonts in *A. t.* Group I cells collected from our GoM

and NMS study sites.

MATERIALS AND METHODS

IFCB Modification and Sample Analysis

The Imaging FlowCytobot was developed to address the need for sustained, *in*

situ observation of nano- and microplanktonic algae (~10 to >100 μm in length). The

original instrument was enclosed in a watertight aluminum housing and was configured

to interrogate raw seawater withdrawn directly from its environment (Olson and Sosik

2007). For this study, the IFCB was used as a laboratory instrument instead. This allowed

samples to be concentrated and stained prior to analysis and also made it possible to

CCD camera

PMT
Cy3 emission

E

PMT
Hoechst 33342
emission

C **D**

B

A

375 nm laser

Flow cell

532 nm laser

Condenser and objective
(not pictured)

Xenon flashlamp

B
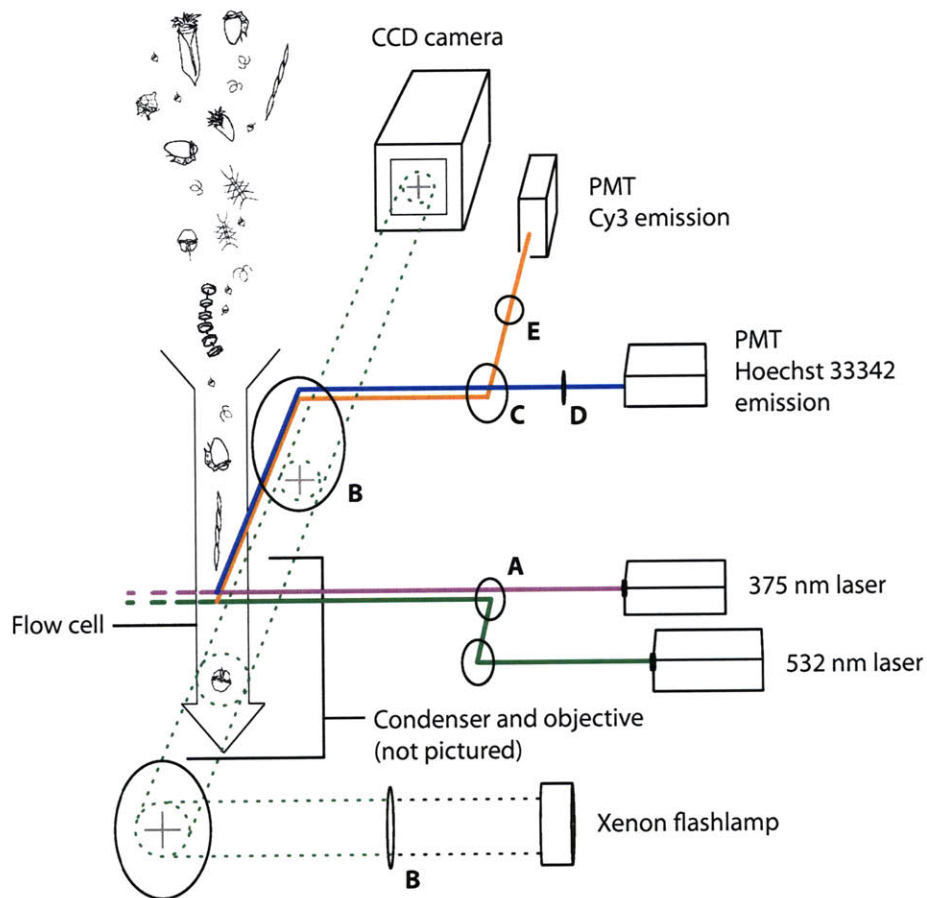
Color filters

A: 400 nm shortpass (uv / Hoechst 3342 excitation)
B: 500-550 nm bandpass (green / imager flash)
C: 560 nm shortpass (blue / Hoechst 33342 emission)
D: 400-500 nm bandpass (blue / Hoechst 33342 emission)
E: 555 nm longpass (yellow-orange / Cy3 emission)

Figure 2. Schema of the revised optical layout of the Imaging FlowCytobot.

install new lasers that do not fit within the original system's housing. The IFCB's red diode laser was replaced with a UV one (CUBE™ 375-16C, 375 nm, 16 mW by Coherent, Inc.) that excites the DNA-specific dye Hoechst 33342. A second, green laser (COMPASS™ 215M-10, DPSS-CW, 532 nm, 10 mW by Coherent, Inc.) was installed for the excitation of Cy3, a cyanine chromofluor that was attached to *A. t.* Group I and *Amoebophrya* spp. specific oligonucleotide probes. Other changes related to the installation of these lasers included the replacement of dichroic filters so that blue light (Hoechst fluorescence, 400-500 nm wavelengths) was passed to the system's trigger PMT (Hamamatsu HC120-05) and yellow-orange light (Cy3 fluorescence, >555 nm wavelengths) was passed to its second PMT. Both beams were focused through a single pair of cylindrical lenses to produce horizontally elongated elliptical beam spots across the sample channel. Because of their color difference, the two beams had different natural focal points relative to the system's sample core. The difference in focal point positions was diminished by adjusting the collimation of each beam independently with separate beam expanders. The beams were separated in space so that cells passed through first the UV beam and then the green beam (Figures 2 and 3).

Other changes to the original system included an updated computing stack (EmETX-i1701, an ETX module that was installed with a PBE-1100 adaptor, ARBOR Technology Corp.) and the addition of a high-speed trace digitizer/data acquisition card (model AD2100-14, Chase Scientific Company). The high-speed digitizer complemented the original system's signal integrators by recording complete traces of PMT response as each cell passed through the laser beams. This trace data was used to quantify
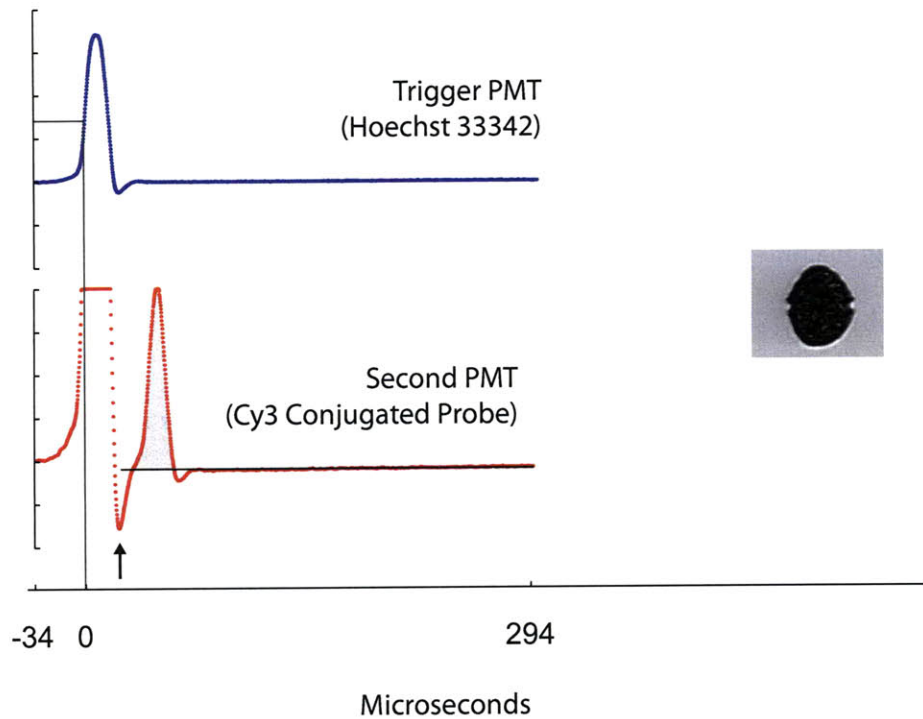
Figure 3. PMT responses as recorded by the AD-2100-4 digitizer card (Chase Scientific Company) and image acquisition. Time zero for each particle 'event' was set by the trigger threshold for blue fluorescence. The orange-yellow fluorescence minimum used to demarcate PMT response due to the UV (left) and green (right) beams is marked by an arrow, and the integral used for identification of positive probe-Cy3 staining is shaded gray. Xenon flash and CCD image acquisition timing were set by a potentiometer and are approximated.

fluorescence from Cy3-conjugated probes (green beam), which could not easily be obtained from integrated signals due to UV-induced orange fluorescence (Figure 3). Trace data from the blue light detector. Trace data from the blue light detector provided measurements of peak height width and helped to ensure that integrated signal values were not affected by detector saturation. Coefficients of variation were estimated by measuring the full width at half maximum (FWHM) of G1 phase populations and dividing by 2.36 (Shapiro 2003).

All plankton and culture samples were concentrated over 20 μm Nitex mesh. Field samples from Salt Pond were also pre-screened with 100 μm mesh to remove large plankton and detritus that could clog IFCB's fluidics system. Fixation and pigment extraction were completed as described by Anderson et al. 2005. This procedure included formalin fixation, which is known to diminish Hoechst fluorescence over time (Cetta and Anderson 1990). We attempted to minimize the effect of prolonged formalin fixation by keeping concentrated samples on ice in the field. In the lab, samples were exchanged with methanol and stored up to 4 months at -20°C before staining with oligonucleotide probes and Hoescht 33342 immediately prior to IFCB analysis.

Two Cy3-conjugated oligonucleotide probes were applied in separate experiments: the NA1 *A. t.* Group I specific ribosomal probe (5'-AGTGCAACACTCCCACCA-3', Anderson et al. 1999) and a genus-specific *Amoebophrya* ribosomal probe that was provided by Mario Sengco

(5'-CCTGCCGTGAACACTCTAAT-3'). Hybridization was completed as described by

Anderson et al. 1999, and samples were re-suspended in 5x SET/Hoechst solution (750

mM NaCl, 5 mM EDTA, 100 mM Tris-HCl, 5 μg mL^{-1} Hoechst 33342, pH 7.8) to an

estimated density of 500-1000 triggers mL^{-1}. Stained samples were stored at 4° C no

longer than six hours then stirred in an ice bath up to 30 minutes prior to analysis.

Measurements of Hoechst 33342 fluorescence were normalized to 6 μm AlignFlow Plus

beads for UV excitation (A-7305, Invitrogen). Unfortunately, the beads could not be used

as an internal standard because their brightness was below the fluorescence threshold

necessary for efficient analysis of field samples. Instead, AlignFlow Plus beads were

analyzed before and after each field sample using a higher electronic gain setting.

Similarly, stability of the green beam's excitation was tested once a day using PE

Calibrite beads (BD Biosciences). Estimated C.V. values from the AlignFlow Plus and

Calibrite beads were ~5% and ~9% respectively.

The IFCB's maximum acquisition rate is limited (to ~12Hz) by the time required

for image acquisition and transfer at each trigger event. This means that if cells are

presented at too high a rate, many of them are ignored, which presented a significant

problem because some field samples contained fewer than 2000 *A. t.* Group I cells. Loss

of a large proportion of cells would have jeopardized our ability to characterize *A. t.*

Group I populations in these samples. Two strategies were used to increase efficiency.

First, a high trigger threshold was applied so that only particles with intense Hoechst

staining were recorded (dinoflagellates and ciliates with exceptionally high DNA content

as well as blue fluorescent detrital particles). This effectively minimized the number of

non-dinoflagellate data recorded by the instrument. The second strategy used was dilution

with additional 5X SET/Hoechst 33342 solution to achieve acquisition rates of 5 Hz or

less. Samples were diluted further when minimization of sample loss was particularly

critical (i.e. when the number of A. t. Group I cells in a stained sample was less than 10^4).

Images were classified using a support vector machine (SVM) based image-

processing pipeline that was developed specifically for analysis of IFCB data (Sosik and

Olson 2007). Seven species/image categories were defined: A. t. Group I singlet, A. t.

Group I doublet, Amylax triacantha, Dinophysis acuminata, Gonyaulax sp.,

Protoperidinium sp., and a tintinnid class. Accuracy of the SVM classifier can vary

substantially depending on the relative uniqueness of each category. In this case, many of

the dinoflagellates were round cells causing the classifier to be particularly error-prone

(e.g. compare A. t. Group I, A. triacantha, Gonyaulax sp., and Protoperidinium sp.,

Figure 5). Therefore, output from the SVM was manually corrected by reviewing all

image classifications and reassigning them as necessary. Data from cell clumps were not

categorized, nor were other ciliate and dinoflagellate cell types that were found in only

one or two water samples. After image classification of field samples, scatter plots of A. t.

Group I fluorescence were reviewed to confirm NA1-Cy3 staining. Because the images

were sometimes not sufficient to distinguish between A. t. Group I and other round cells,

we only included cells with positive probe fluorescence. Image features assessed in the

initial step of the image-processing pipeline include particle metrics such as size, shading

and orientation, but only diameter measurements were used in our analysis. Single A. t.

Group I cells were approximated as spheres to calculate their volume, which was of

interest because it was expected to change in unison with DNA content during cell division and because previous studies had identified *A. t.* Group I planozygotes as larger cells.

Amoebophrya spp. Studies

An *Amoebophrya* sp. isolate was obtained from M. Sengco in 2004 and propagated using cultures of SP-E10-03, an *A. t.* Group I clone that was isolated from Salt Pond in 2001. Cultures of SP-E10-03 were grown in f/2 media without silicate (Guillard and Ryther 1962) and infected with suspensions of actively parasitized *A. t.* Group I hosts. Host cells infected by *Amoebophrya* trophonts are readily identified by their green fluorescence under blue illumination. Transfers to new host cultures were completed when more than half of host cells in a parent culture were visibly infected (typically every 4-6 days), but no other efforts were made to control the parasite to host ratio. A sample of heavily infected host cells was chosen for analysis on IFCB after a transfer in the summer of 2009. Measurement of DNA content required extraction of all intracellular pigments with ice-cold methanol. Previously it had been shown that the *Amoebophrya* genus-specific probe was effective for identifying infected *A. t.* Group I except during the initial phase of infection when a single dinospore was inside the host (D. Kulis *pers comm.*). The limitation was ignored because mature infections were identified effectively and it was these older infections that were expected to cause the greatest increase in host DNA content and cell size. Culture samples were stained for

analysis as described but infected cells (rather than *A. t.* Group I) were identified by Cy3 probe fluorescence during image classification. Unstained cells were used as an internal control because their DNA content and size were indistinguishable from a sample of uninfected host cells.

In years prior to the 2009 field surveys in the NMS, high rates of *Amoeobophyra* infection had been observed in Salt Pond. Therefore, infection rates for *A. t.* Group I cells were determined from aliquots of Salt Pond and the GoM field samples by staining simultaneously with the Cy3 conjugated NA1 probe and a fluorescein isothiocyanate (FITC) conjugated version of the *Amoebophra* genus specific probe. After staining, *A. t.* Group I cells fluoresced yellow-orange under green illumination and green under blue illumination if substantially infected by the parasite. The proportion of cells infected was determined from observations of no fewer than 200 *A. t.* Group I cells per sample using an epifluorescence microscope.

Salt Pond Sampling

Samples for IFCB analysis were taken in conjunction with weekly surveys performed throughout the NMS and adjacent coastal waters between March and June 2009. Here, a brief overview of the NMS surveys is provided but readers looking for a detailed description are directed elsewhere (Crespo et al. *in prep*).

The NMS is a network of shallow channels that link three brackish ponds to a central marsh area. In turn, the central marsh is connected to the Atlantic Ocean via Nauset inlet (Figure 4 inset). The mean depth within the NMS is approximately 1.25 m but its northwestern and southeastern extremities (Salt Pond and Mill Pond respectively) each contain deep holes 8 and 10 m in depth. Town Cove, the pond that forms the system's southwestern extremity, is broader and shallower with a maximum depth of 3.5 m. Weather and tide permitting, 31 stations from throughout the NMS were sampled for *A. t.* Group I concentration, nutrients (NH_4^+, NO_2^- + NO_3^-, PO_4^-, and SiO_3^-), and salinity-temperature profiles by two boats. Surveys were scheduled to follow daytime high water from Nauset inlet to the system's extremities during the transition from flood to ebb. Water samples were taken by hand-activating a 5 L Niskin bottle just below the water surface at all stations. Where depth exceeded 2 m within the central marsh, additional samples were taken near the bottom, and four depths were sampled from the deep holes in Salt Pond and Mill Pond. Salinity-temperature profiles were taken with SPE-19 or SPE-19plus CTD instruments (Seabird, Inc.). Nutrients samples were filtered through 0.2 μm cartridge filters to remove particulates and stored in sample-rinsed, acid-washed vials. Samples were frozen after transport to the lab (typically less than 6 hours from filtration) and stored at -20°C until analysis by the Woods Hole Oceanographic Institution Nutrient Analytical Facility (Woods Hole, MA). *A. t.* Group I cells were enumerated from formalin-fixed concentrates after methanol extraction and staining with the Cy3-conjugated NA1 ribosomal probe as described by Anderson, Kulis et al. (2005).

127

A. t. Group I cell concentrations in the ponds were 2 – 3 orders of magnitude higher than in the central marsh. The result was expected because *A. t.* Group I and other dinoflagellates avoid high irradiance levels at the surface, causing the cells to be selectively retained in salt ponds during strong tidal flushing (Anderson and Stolzenbach 1985). Another byproduct of this behavior is that the Salt Pond population in the system's north is isolated from populations in Town Cove and Mill Pond in the south as shown by its later onset and demise (Supplementary Figure 1). We took advantage of the Salt Pond population's isolation to identify samples collected during its growth, peak and termination phases (May 11th, 18th, and 26th respectively). Samples for IFCB analysis were concentrated from 5 L Niskin bottles tripped at the surface (0-1 m), mid-depth (approximately 5 m), and 1 m above the bottom at a station over the 8 m hole in the center of the pond. All IFCB samples were taken in addition to normal sampling described for the surveys.

An important caveat for these surveys was that all were conducted at high slack water and during daylight. This restriction was necessary so that boats could navigate through areas of the central marsh that were drained during low water. It also determined the time of day when IFCB samples were taken from Salt Pond. Sampling on the 11th and 26th of May were completed by mid-afternoon when most vegetative *A. t.* Group I cells were expected to be in the G1 phase of the cell cycle. Sampling on the 18th was completed before mid morning when a growing population would still have been completing cell division.
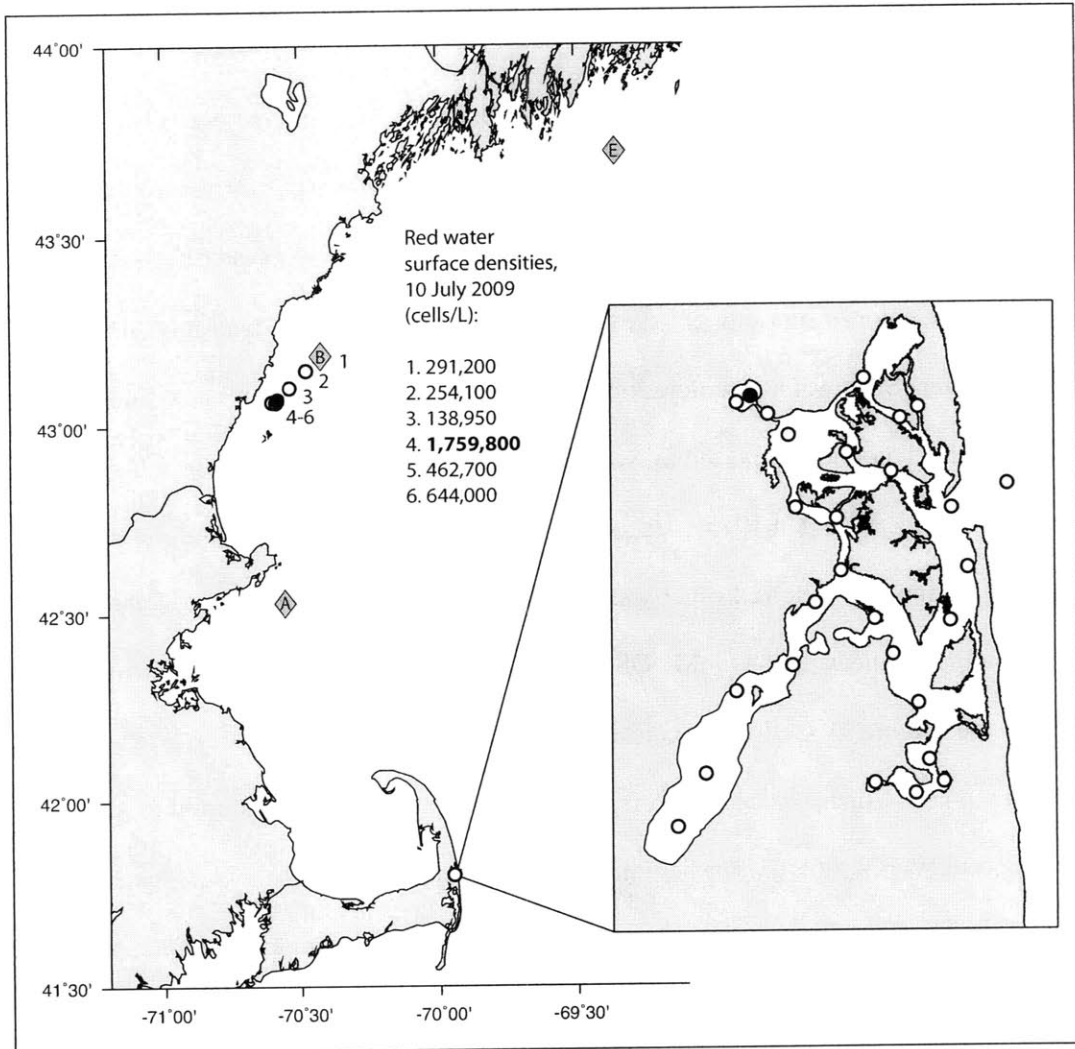
Figure 4. Gulf of Maine (GoM) map. Portsmouth red tide transect stations are denoted by numbered circles from Portsmouth Harbor, ME to GoMOOS Buoy B (diamond, also GoMOOS Buoys A and E). The cell sample from station 4 (closed circle) was analyzed by IFCB and by FACSCalibur. *Inset.* Nauset Marsh system (NMS) map. Thirty-one stations throughout the NMS were sampled weekly from late March to June 2009. Salt Pond (closed circle) is a kettle pond and forms the northwestern extremity of the NMS.

An *A. t.* Group I red tide (hereafter termed the "Portsmouth red tide") was

discovered serendipitously during a mooring deployment July 10, 2009 aboard the R/V

Gulf Challenger (University of New Hampshire, Portsmouth, NH). Water in the area

appeared the color of rust and *A. t.* Group I was the single dominant microplankter at the

water surface. Such water discoloration by *A. t.* Group I is rare in the western Gulf of

Maine and had not been reported in New Hampshire or southwestern Maine since 1972

(Hartwell 1975; Mulligan 1975). Six surface samples were taken using a bucket along a

transect beginning where red water was first encountered outside Portsmouth Harbor, ME

and heading northeast to the GoMOOS B buoy where the mooring deployment was

completed (Figure 4). Cells were concentrated over a 20 μm Nitex sieve and formalin

fixed on board. Concentrates were refrigerated immediately, then transferred to ice for

transport to Woods Hole, MA where they were extracted with methanol less than twelve

hours after sampling. Cells were stained and enumerated as described by Anderson, Kulis

et al. (2005). The maximum cell density observed was nearly 1.8 million cells L^{-1} and no

Amoebophrya parasites were detected. Hydrographic data from the area were compiled

from GoMOOS reports at Buoy A, B and E (http://www.gomoos.org). Unfortunately, the

salinity sensor at Buoy B was not functioning during the red tide, but salinity

measurements were taken at Buoy A and Buoy E, which lie downstream and upstream

respectively from Buoy B within the Western Maine Coastal Current. A fresh buoyant

mass of water was observed at Buoy A July 9 – 11, coincident with the red water

observation at Buoy B on July 10 (Supplementary Figure 2). The buoyant mass suggests

that the red tide spread from the northeast, through entrainment in the Maine river plume.

An emergency response was undertaken beginning July 19, nine days after the red

tide was discovered aboard the R/V Gulf Challenger. The response included both ship-

based water sampling and an aerial survey spanning most of the southern GoM. No water

discoloration attributable to *A. t.* Group I was observed during the aerial survey and cell

concentrations near and downstream from Buoy B fell at least 10^4-fold (B. Keafer *in*

prep). A follow-up benthic survey conducted October 15-25, 2009 found high numbers of

new *A. t.* Group I cysts in a tongue extending southeast from Portsmouth across Jeffries

Ledge (Supplementary Figures 3 and 4). The methods used to map cyst distribution

across the Gulf of Maine were described by Anderson, Stock, et al. 2005. Annual cyst

surveys have included stations near Portsmouth, NH and Cape Ann, MA in 2004, 2005,

2006, 2008, and 2009 (the southernmost stations for the 2007 survey were just north and

east of this area). Cyst abundance in both the 0-1 cm and 1-3 cm layers were greater than

100-fold higher in 2009 than in any of the previous surveys, with a maximum

concentration of 5203 cysts cm^{-3} in the 0-1 cm layer outside Portsmouth Harbor. Our

conclusion from the cyst surveys and our observations during the emergency response is

that the 2009 red tide produced an exceptionally large number of cysts and expanded the

southern extent of the mid-coast Maine *A. t.* Group I cyst bed. Because cell

concentrations were exceptionally high during the red tide and dissipated within 10 days,

and because other species and parasite loads were not significant, the population sampled

near Portsmouth July 10, 2009 was most likely undergoing high levels of sexual

encystment. Therefore the red water cells were compared to those taken in Salt Pond as a population that only reflects high levels of sexual fusion, not population growth or parasite infection.

The monospecificity and high concentrations of *A. t.* Group I in the GoM samples also enabled analysis using a conventional, commercial flow cytometer (FACSCalibur, BD Biosciences). A subsample from transect station 4 (where *A. t.* Group I cell concentration was highest) was stained using propidium iodide (PI) and treated with RNAse A as described by Taroncher-Oldenburg, Kulis et al. (1997). The FACSCalibur analysis was in addition to that described using IFCB. A similar parallel analysis was also completed using clonal *A. t.* Group I cultures grown that had significant abundance of both G1 and G2+M phase cells.

RESULTS AND DISCUSSION

Enhanced Measurement of Diverse Microplankton

Particle detection and image classification were essentially identical as described for the original IFCB instrument. The most significant change was that particle detection was based on blue fluorescence after Hoechst 33342 staining rather than red chlorophyll auto-fluorescence. The majority of particles recorded in this way in the red tide and culture samples were *A. t.* Group I cells, but the Salt Pond data sets were dominated by unclassified particles (debris) and had several other abundant species including *D.*

acuminata, a species that causes Diarrheic Shellfish Poisoning (DSP). Fluorescence measurements from *A. t.* Group I overlapped those of detrital particles, *Protoperidinium* sp., and *Dinophysis acuminata*, but these species or particles were still readily distinguished by their images (Figure 5). Similar overlap of fluorescence metrics occurred among detrital particles, *A. triacantha*, *Gonyaulax* sp., and an abundant tintinnid, but these taxa were also readily distinguished.

The overlap in yellow-orange fluorescence between *D. acuminata*, *Protoperidinium* sp. and *A. t.* Group I stained with the NA1-Cy3 probe would have interfered with analysis on a conventional flow cytometer, but did not interfere with measurements taken with the IFCB. The strong yellow-orange fluorescence of *Protoperidinium* sp. and *D. acuminata* was not due to poor specificity of the *A. t.* Group I specific NA1-Cy3 probe. Both *Protoperidinium* sp. and *D. acuminata* autofluoresce under green light even after methanol extraction. In the case of *Protoperidinium* sp., autofluorescence is broad spectrum and its source is unknown (D. Kulis *pers comm.*). On the other hand, yellow-orange autofluorescence by *D. acuminata* is attributable to poor extraction of its cryptophyte derived kleptoplastids. Separation of *D. acuminata* and *A. t.* Group I might be achieved with a conventional flow cytometer by attaching a different fluorochrome to the NA1 probe or by examining differences in the two species' forward light scatter (a proxy for the in-flow cross-section of a particle). However, in either case, it is unlikely that the relatively rare dinoflagellates would be well separated from the abundant debris particles in the Salt Pond samples. The IFCB could easily separate these species.

Assessment of DNA Content Measurements

Analysis of DNA measurements focused on *A. t.* Group I cells because this species was the target of our field sampling regimen. However, the detection of other dinoflagellates and a tintinnid ciliate suggests that the modified system will be useful for analysis of other species' DNA fluorescence within natural assemblages, both as a proxy for cell and sexual cycle shifts, and perhaps also for estimation of dinoflagellate ingestion by microplanktonic grazers.

The accuracy of DNA content measurement was first evaluated by comparing results from samples stained with Hoechst 33342 and analyzed using the modified IFCB and those stained with PI and analyzed using the FACSCalibur flow cytometer. Modes of G1 and G2+M phase cells were apparent in data from the modified IFCB, but C.V. values were large (16-28% by IFCB compared to 12-15% by FACSCalibur). The capability to distinguish G1 and G2 phase cells was confirmed by using *A. t.* Group I doublets as internal controls. Though doublets were rare in these samples, it was shown that their DNA content coincided with that of singlet 2c *A. t.* Group I cells. The comparison reassured us that measurements by IFCB were reliable for detection of 1c, 2c and potentially 4c modes in the *A. t.* Group I populations sampled. On the other hand, the large C.V. estimates produced by IFCB relative to the FACSCalibur highlight the need to further optimize the IFCB for quantitative measurement of fluorescence. Such refinement may enable species identification directly from DNA content, as was recently demonstrated for separation of populations of *Karlodinium venefecum* and various

Alexandrium species using a FACSCalibur by Figueroa et al. (2010). Without such refinement, we limited our analysis of IFCB DNA measurements to more qualitative comparisons of the presence or absence of 1c, 2c and 4c *A. t.* Group I populations.

The DNA fluorescence histograms from the peak and termination phase of the Salt Pond bloom (Figure 6, May 18th and 26th) appear shifted to lower fluorescence values, though they still show 2c modes (compare to Figure 7). Such shifts are presumed to be an artifact of formalin fixation, which has been noted to reduce Hoechst staining in other dinoflagellates (Cetta and Anderson 1990). The 'formalin effect' was observed in all field samples, where 1c unit of DNA content corresponded to approximately 70% the value observed in culture samples that were fixed and extracted in the laboratory. Especially in cases where the formalin effect obscured differences between parent and daughter mitotic cells, cell volume was a useful covariant of DNA content that helped define populations of 1c and 2c cells (Figure 7). In contrast, yellow-orange fluorescence histograms (a proxy for ribosome content in stained *A t.* Group I) from dividing populations were not bimodal and had estimated C.V. values >40% (Figure 6). The failure of NA1-Cy3 staining intensity to co-vary with cell volume and DNA content may be related to reduced precision of the Calibrite bead standards compared to the UV-excited beads or to variation in rRNA cell^{-1} with physiological condition or growth stage (Anderson, Kulis et al. 1999).

A second test of the absolute measurements of DNA content was provided by results from the Portsmouth red tide sample. The initial IFCB analysis of this sample

Figure 5. Compiled flow cytometric measurements of blue (400-500 nm) and yellow-orange (>555 nm) fluorescence from surface, mid-depth and bottom samples taken at Salt Pond May 18th 2009 and analyzed on the modified IFCB. Unclassified particles (gray) include debris, rare cell types and cell clumps, and overlapped all regions of the scatter plot.
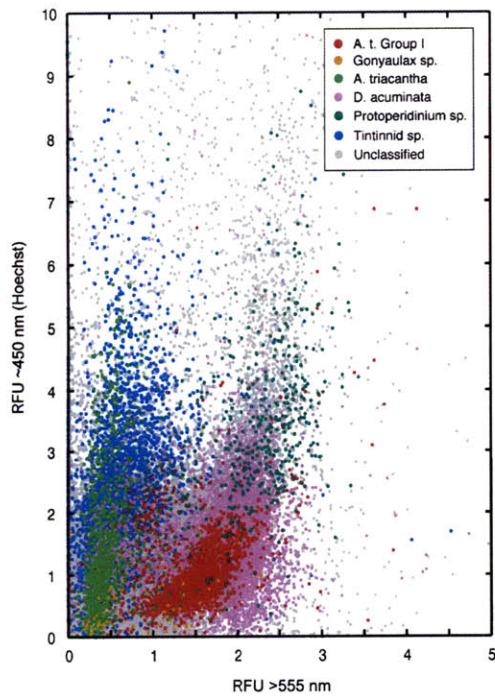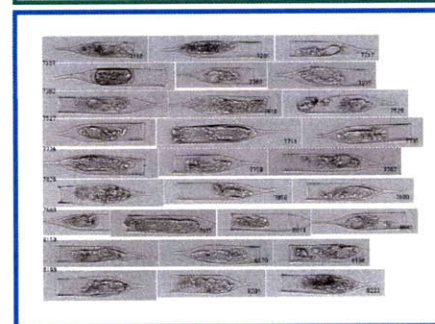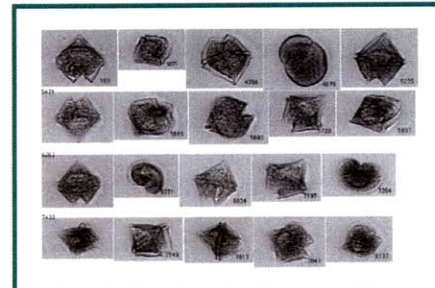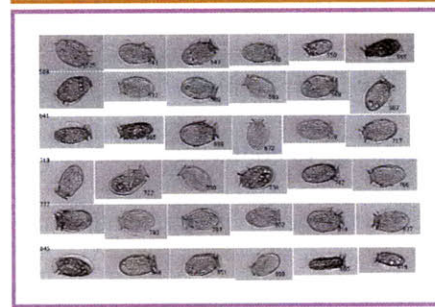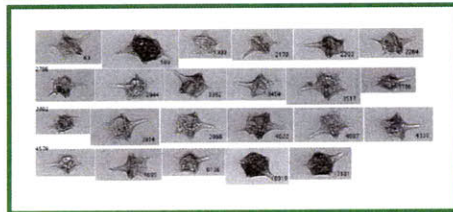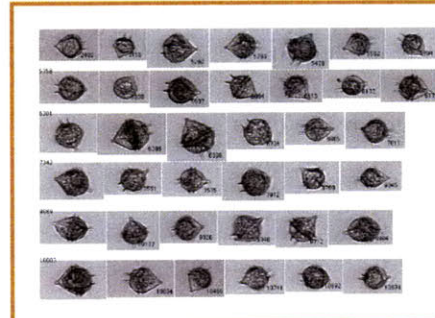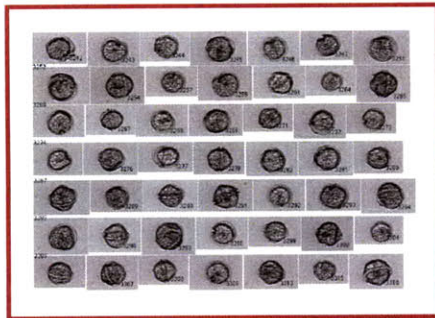
showed a dominant 2c mode (Figures 6, Red tide Sta. 4). Because the red tide sample was

essentially monospecific and contained relatively little debris, it was possible to identify

A. t. Group I cells using a conventional flow cytometer. Analysis using the FACSCalibur

instrument confirmed the result from IFCB, that the Portsmouth red tide sample was

dominated by 2c cells.

Discrimination between Sex and Parasitism by Amoebophrya spp.

The field samples chosen for analysis represented *A. t.* Group I bloom phases that

were expected to contain G1 and G2 vegetative cells, gametes, planozygotes, and cells

infected by *Amoebophrya* parasites. In particular, Salt Pond samples were expected to

show a progression from a vegetative population during the development phase of the

bloom (May 11th) towards increasing abundance of large, 2c planozygotes during bloom

termination (May 26th). Instead, the abundance of 2c cells declined between the

development and termination phases of the bloom and a broadly distributed mode of very

large 4c cells was found in the May 26th bottom sample (Figures 6 and 7). The rate of

Amoebophrya infection was also very high in that sample (Figure 8), suggesting that high

DNA content and cell size could have been due to parasite infection rather than concerted

premeiotic replication by planozygotes.

The effect of *Amoebophrya* infection on cell size and DNA content was explored

using two approaches. First, we examined cell size and infection rates in bottom and mid-

depth samples from the May 26th survey. If cell size increases were due to the presence of

Amoebophrya trophonts, we would expect all large cells to contain them, but this was not

observed. Instead, large and small *A. t.* Group I cells collected from mid-depth had

infection rates of 61.8% and 13.8% respectively, and those from the bottom sample had

infection rates of 79.0% and 42.6% respectively. This suggests that *Amoebophrya*

infection was not the direct cause of increased cell size, but that the large cells (most

likely mature planozygotes) exhibited higher rates of parasite infection than the smaller

cells. In the second approach, the effect of infection on a clonal culture of *A. t.* Group I

was examined. DNA content of infected SP-E10-03 was unchanged or increased slightly,

and cell volume remained less than a third that of the large cells collected in Salt Pond

(Figures 6 and 7). We conclude that the large, high DNA content cells in Salt Pond were

most likely mature planozygotes that had replicated their DNA.

Other flow cytometry studies of dinoflagellate sexual conjugation have not

detected premeiotic replication (Parrow, Burkholder et al. 2002; Parrow and Burkholder

2003; Figueroa, Garces et al. 2007; Figueroa, Garcés et al. 2010) and some have

suggested that dinoflagellates may undergo a unique meiosis that does not require

premeiotic replication (Beam and Himes 1980; Kita, Fukuyo et al. 1993; Elbrachter 2003;

Parrow and Burkholder 2003). The notion that dinoflagellate could carry out a unique

dinoflagellate meiosis stems from pioneering genetic studies of the saprobic

dinoflagellate, *Crypthecodinium cohnii*, which showed that these cells only rarely yield

tetratype sexual progeny (Beam, Himes et al. 1977; Beam and Himes 1980), and also by

a lack of reports of chiasmata between dinoflagellate chromosomes. The observations

from *C. cohnii* can be explained by either a normal two-division meiosis that occurs

Figure 6. Histograms of blue (400-500 nm) and yellow-orange fluorescence (>555 nm).

Left panels. Estimated DNA content distributions of *A. t*. Group I populations from Salt

Pond and the Portsmouth red tide. A 'formalin effect' of reduced Hoechst 33342

fluorescence was observed in all samples but was most severe in samples from the peak

of the Salt Pond bloom, May 18, 2009. *Top right panel*. Estimated DNA content

distribution of SP-E10-03 culture (*A. t*. Group I) infected and uninfected with an

Amoebophrya sp. Parasite. *Middle right panel*. NA1-Cy3 ribosomal probe staining

intensity from four field sample sets.

Figure showing multiple normalized counts histograms:

Left column (Relative Fluorescence Units (400-500 nm)):
- Legend: Surface (dashed), Midwater, Bottom
- 'Growing' 15:11 hrs 11 May 2009
- 'Peak' 08:02 hrs 18 May 2009
- 'Terminating' 14:46 hrs 26 May 2009
- Red tide, Sta. 4 16:32 hrs 10 July 2009

Right column:
- SP-E10-03 culture: Amoebophyra sp. +, Amoebophyra sp. − (Relative Fluorescence Units (~450 nm))
- Legend: Red tide, SP Surface, SP Midwater, SP Bottom (Relative Fluorescence Units (>555 nm))

141

without homologous recombination, or by a seemingly simple one-division haploidization that may lack premeiotic replication.

It is unfortunate that the latter is consistent with previous flow cytometry studies that detected 2c but not 4c zygotes. However, in each of the cited flow cytometry studies, zygotes were only observed up to formation, not for sufficient time to detect meiotic division, if it occurred. On a more fundamental level, it is unlikely that dinoflagellates would develop and maintain a nonstandard one-division meiosis without premeiotic replication because its evolution is improbable (Raikov 1995; Wilkins and Holliday 2009), and because a one-division system is susceptible to loss of sexual function (Haig 1993; Archetti 2004). If homologous recombination does not occur during meiosis, daughter cells from meiosis I are haploid and have nearly identical chromatids, essentially like G2 phase vegetative cells. Moreover, analysis of meiotic tetrads would yield only parental ditype or non-parental ditype progeny and not tetratypes, just as has been reported for *C. cohnii*. It remains to be shown whether the result from *C. cohnii* is generalizable to other dinoflagellates. In any case, premeiotic DNA replication most likely does occur in dinoflagellates. However, without documentation of meiotic chromosomal recombination, it is possible that a species re-enters the mitotic cell cycle after one meiotic division as 2c, G2 phase cells rather than as 1c, G0/G1 phase cells after two meiotic divisions (Figure 1).

The higher infection rate and near-bottom localization of planozygote cells suggests that they have a strong surface avoidance (or bottom-seeking) behavior and may
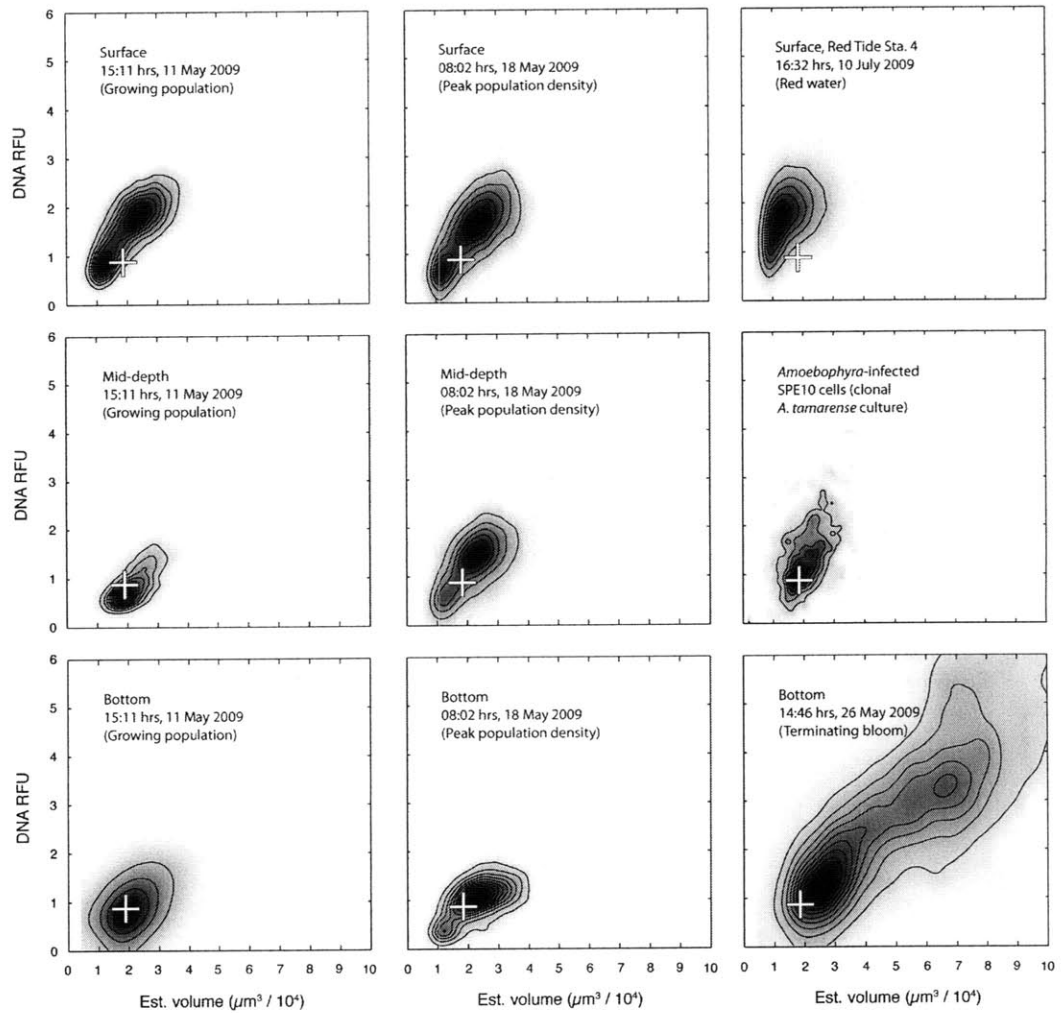
Figure 7. Contour plots of estimated cell volume versus Hoechst 33342 fluorescence (DNA RFU). The mode value of a cultured, G1 phase vegetative population (1n, 1c) is marked by a white '+' in each panel as a common reference. All DNA measurements were normalized to UV-excitable, 6 μm AlignFlow Plus beads.

143

have greater susceptibility to *Amoebophrya* infection. The latter would be surprising

because planozygotes, as diploid (2n) cells, would be expected to benefit from

heterozygous complementation of resistance factors (in our laboratory, clonal isolates

[1n, vegetative cells] vary substantially in their susceptibility to *Amoebophrya* infection).

Alternative explanations for the higher infection rate of large cells include either a delay

in the maturation of planozygotes infected by *Amoebophrya* (such that uninfected

planozygotes are 'lost' more readily through transformation to the hypnozygote stage), or

a longer duration of infection versus smaller, vegetative *A. t.* Group I cells. In the second

case, the higher infection rate of planozygotes might reflect their higher DNA content,

which is often a target for digestion by these parasites (Cachon and Cachon 1987).

Trophonts might expend the nuclear contents and lyse haploid hosts more quickly than

diploid ones, leading to a higher apparent infection rate.

Near Surface Enrichment of 2c A. t. *Group I Cells*

The collection times of field samples from growth and termination of the Salt

Pond bloom and the red tide were during mid afternoon when *A t.* Group I vegetative

populations were expected to have 1c DNA content (G1 phase of the cell cycle). The

abundance of 2c planozygotes was expected to be near zero during the growth phase of

the Salt Pond bloom (May 11th) and increase to a substantial fraction of the total

population during the bloom's termination (May 26th). Similar to the termination phase

samples, red tide samples were expected to have abundant 2c planozygotes. The latter

prediction was borne out - most red tide *A. t.* Group I contained 2c DNA content (Figures

6 and 7) - but the absolute abundance of 2c cells in Salt Pond (estimated as the product of

cell concentration and proportion of 2c and 4c cells at each of the three depths sampled)

increased less than 2-fold from May 11th to May 26th, and fell from ~25% to ~10% of the

depth integrated population (Figures 6 and 7). Given its rate of growth from May 11th to

May 18th (Figure 8), it is unlikely that a quarter of the Salt Pond population had entered

the sexual cycle by May 11th and therefore the estimate is probably inflated. Still, many

factors could have contributed to actual changes in the Salt Pond bloom's demographics,

the most likely being differential losses due to parasitism and grazing. Planozygotes were

shown to suffer higher rates of *Amoebophrya* infection and also to grow much larger than

vegetative cells, making them more attractive to size-selective grazers (Figures 7 and 8).

As in previous studies of Salt Pond, no environmental cues for sexual induction

were found (Anderson, Chisholm et al. 1983). Elevated nutrient concentrations near the

bottom of Salt Pond reflected a flux of remineralized nutrients from microbial

degradation of organic matter. Though the measured concentration of nutrients was lower

overall on May 26th than on May 11th, the flux had most likely increased due to higher

water temperature at the bottom of the pond (> 2°C change; Figure 8). Moreover, *A. t.*

Group I are able to migrate vertically to depths far greater than the maximal depth of Salt

Pond (Anderson and Stolzenbach 1985; Ralston, McGillicuddy et al. 2007), so that lower

nutrient concentrations near the surface should not have affected cell growth. Water

temperatures were below the optimal range for growth throughout the study period, but

were also below the optimal range for encystment (18-21°C; Anderson et al. 1984).

Figure 8. Salt Pond *A. t*. Group I population history from weekly surveys, March-June

2009. *Top panel*. Line graph of maximum observed cell densities recorded from all

depths sampled and bar graph of surface, mid-depth, and bottom sample rates of

Amoebophrya infection. *Middle and bottom panels*. Hydrographic profiles taken during

growth, peak, and termination phase of the Salt Pond bloom (11, 18, and 26 May 2009

respectively).
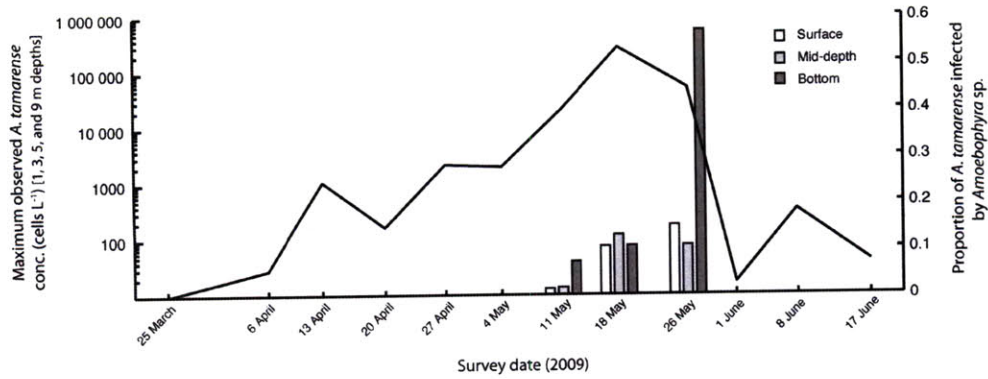
15:11 hrs, 11 May 2009
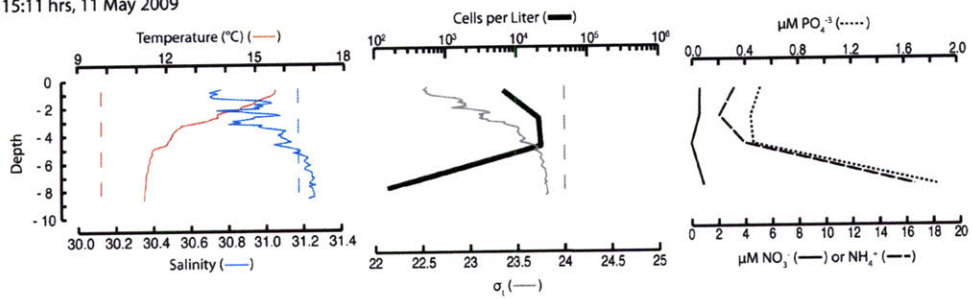
8:02 hrs, 18 May 2009
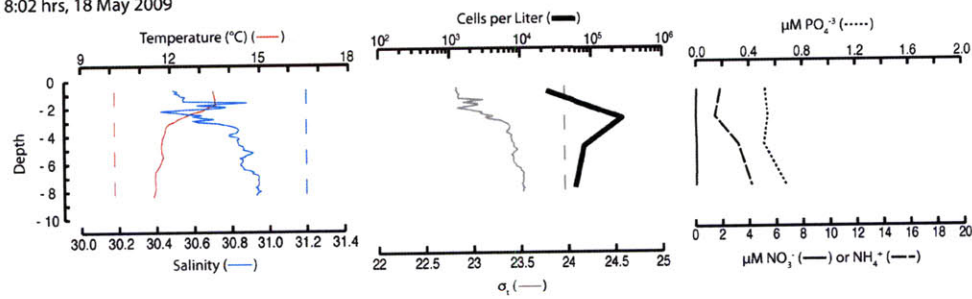
14:46 hrs, 26 May 2009
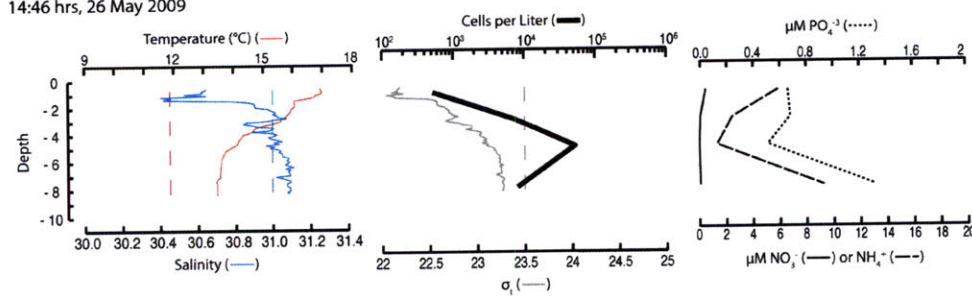
The majority of *A. t.* Group I were vegetative as shown by large concentrations of 1c cells during the May 11th and May 26th surveys, and the presence of dividing populations at all depths during the bloom's peak May 18th. Unlike the other field samples analyzed in this study, samples from May 18th were taken near daybreak when vegetative cells were expected to be completing cell division. The greater abundance of 2c cells than 1c cells at all depths was surprising given our expectation that the population would be synchronized by a light dependent transition point early in G1 (Figure 7). Most cells should have completed cell division during nighttime hours, with 1c/G1 phase daughter cells most abundant at daybreak. The later timing of division and its synchronization may be evidence of rhythmic cell cycle control similar to that observed in *Karenia brevis* (Brunelle, Hazard et al. 2007; Van Dolah, Leighfield et al. 2008).

The presence of 2c *A. t.* Group I cells in the near surface of all Salt Pond and the Portsmouth red tide samples may also reflect increased growth under higher irradiance and/or vertical migration to the surface for asynchronous cell division. Van Dolah, Leighfield et al. (2008) noted near surface enrichment of 2c/G2 *Karenia brevis* during a mesocosm experiment and attributed this pattern to unequal inheritance of daughter cells and consequent differences in phototaxis. The same may also be true of *A. t.* Group I even though the presence of 2c cells was consistent with our expectation of a large abundance of planozygotes during the red tide. If instead 2c cells were vegetative, their localization near the surface and lack of coordination with bulk division observed near daybreak could be adaptive for successful gametogenesis (rather than normal cell division).

G2 cells represent a uniquely bipotent stage of the *A. tamarense* life cycle: they can divide to produce new daughter vegetative cells or to enter the sexual cycle as gametes (Figure 1). In culture, the 'decision' of G2 phase cells is driven toward gametogenesis by starvation and toward reproduction by nutrient availability (Brosnahan, Kulis et al. 2010). However, no nutrient stress was apparent in Salt Pond even though large planozygote cells were observed during bloom termination. Brosnahan, Kulis et al. (2010) suggested that *A. tamarense* species may express a 'self-recognition mating type' that essentially acts as a quorum mechanism for sexual induction. If such a quorum mechanism were to exist, the migration of gametogenesis poised *A. t.* Group I to the surface would aid their aggregation apart from the larger mitotic population. Moreover, G2 *A. t.* Group I could quickly aggregate to densities substantially higher than their ambient concentration via simple phototaxis (Lewis 2005). Indeed, any persistence of planozygotes in the near surface ought to be maladaptive because passive aggregation at the surface also heightens exposure to motile grazer species and increases exposure to damaging levels of irradiance. The planozygote stage is the product of successful encounter by two compatible gametes and therefore would seem to have already realized the benefit of migrating up to the surface. In contrast, a G2 phase cell might assess whether gametogenesis is likely to yield successful encystment, and if not, could instead proceed toward normal mitotic division (though asynchronous with the larger population). It is noteworthy that in a culture study, the only condition that prolongs the G2 phase of the cell cycle to a length comparable to G1 in *A. t.* Group I was severe

phosphate stress, a condition that is known to stimulate mating between compatible cultures (Anderson and Lindquist 1985; Taroncher-Oldenburg, Kulis et al. 1999).

The resolution of the identity of 2c cells at the surface and their role in the ecology of *A. t.* Group I blooms awaits further analysis using tools like the IFCB and other new molecular methods in development for positively identifying sexual stages of this species' life cycle. Higher frequency sampling and IFCB analysis of the Salt Pond bloom should better delimit the behavior of 2c cells near the surface in Salt Pond and also in large-scale offshore blooms occurring within the GoM. A powerful aspect of the IFCB analysis described here is that sample preparation closely follows established methods for formalin fixation and methanol extraction of field samples. This characteristic enabled analysis of cell samples taken from the Portsmouth red tide that might otherwise have been unavailable.

ACKNOWLEDGMENTS

REFERENCES

Anderson, D. and N. Lindquist (1985). "Time-course measurements of phosphorus depletion and cyst formation in the dinoflagellate *Gonyaulax tamarensis* Lebour." Journal of Experimental Marine Biology and Ecology. **86**: 1-13.

Anderson, D. M., S. W. Chisholm, et al. (1983). "Importance of life cycle events in the population dynamics of *Gonyaulax tamarensis*." Marine Biology **76**(2): 179-189.

Anderson, D. M., D. M. Kulis, et al. (1984). "Sexuality and cyst formation in the dinoflagellate *Gonyaulax tamarensis* : Cyst yield in batch cultures." Journal of Phycology **20**(3): 418-425.

Anderson, D. M., D. M. Kulis, et al. (1999). "Detection of the toxic dinoflagellate *Alexandrium fundyense* (Dinophyceae) with oligonucleotide and antidbody probes: variability in labeling intensity with physiological condition." Journal of Phycology **35**: 870-883.

Anderson, D. M., D. M. Kulis, et al. (2005). "Identification and enumeration of *Alexandrium* spp. from the Gulf of Maine using molecular probes." Deep-Sea Research II **52**(19-21): 2467.

Anderson, D. M., C. A. Stock, et al. (2005). "*Alexandrium fundyense* cyst dynamics in the Gulf of Maine." Deep-Sea Research II **52**(19-21): 2522.

Anderson, D. M. and K. D. Stolzenbach (1985). "Selective retention of two dinoflagellates in a well-mixed estuarine embayment: the importance of diel vertical migration and surface avoidance." Marine Ecology Progress Series **25**: 39-50.

Anderson, D. M. and D. Wall (1978). "Potential importance of benthic cysts of *Gonyaulax tamarensis* and *G. excavata* in initiating toxin dinoflagellate blooms." Journal of Phycology **14**(2): 224-234.

Archetti, M. (2004). "Loss of complementation and the logic of two-step meiosis." Journal of Evolutionary Biology **17**(5): 1098-1105.

Armbrust, E. V., S. W. Chisholm, et al. (1990). "Role of Light and the Cell Cycle on the Induction of Spermatogenesis in a Centric Diatom." Journal of Phycology **26**(3): 470-478.

Beam, C. A. and M. Himes (1980). Sexuality and meiosis in dinoflagellates. Biochemistry and Physiology of Protozooa. M. Levandowsky and S. H. Hutner, Academic Press. **3:** 171-206.

Beam, C. A., M. Himes, et al. (1977). "Genetic evidence of unusual meiosis in the dinoflagellate *Crypthecodinium cohnii*." Genetics **87**(1): 19-32.

Bhaud, Y., M.-O. Soyer-Gobillard, et al. (1988). "Transmission of gametic nuclei through a fertilization tube during mating in a primitive dinoflagellate, *Prorocentrum micans* Ehr." J Cell Sci **89**(2): 197-206.

Brosnahan, M. L., D. M. Kulis, et al. (2010). "Outbreeding lethality between toxic Group I and nontoxic Group III *Alexandrium tamarense* spp. isolates: Predominance of heterotypic encystment and implications for mating interactions and biogeography." Deep Sea Research Part II **57**(3-4): 175-189.

Brunelle, S. A., E. S. Hazard, et al. (2007). "Characterization of a Dinoflagellate Cryptochrome Blue-Light Receptor with a Possible Role in Circadian Control of the Cell Cycle." Journal of Phycology **43**(3): 509-518.

Cachon, J. and M. Cachon (1987). Parasitic Dinoflagellates. The Biology of Dinoflagellates. F. J. R. Taylor. Oxford, Blackwell Scientific Publications: 571-610.

Campbell, L., R. J. Olson, et al. (2010). "First Harmful *Dinophysis* (Dinophyceae, Dinophysiales) Bloom in the U.S. is Revealed by Automated Imaging Flow Cytometry." Journal of Phycology **46**(1): 66-75.

Cembella, A. D. and F. J. R. Taylor (1985). Biochemical variability within the *Protogonyaulax tamarensis/catenella* species complex. Toxic Dinoflagellates. D. M. Anderson. New York, Elsevier: 55-60.

Cetta, C. M. and D. M. Anderson (1990). "Cell cycle studies of the dinoflagellates *Gonyaulax polyedra* Stein and *Gyrodinium uncatenum* Hulburt during asexual and sexual reproduction." Journal of Experimental Marine Biology and Ecology [J. EXP. MAR. BIOL. ECOL.] **135**(1): 69-84.

Chambouvet, A., P. Morin, et al. (2008). "Control of Toxic Marine Dinoflagellate Blooms by Serial Parasitic Killers." Science **322**(5905): 1254-1257.

Coats, D. W. (1999). "Parasitic Life Styles of Marine Dinoflagellates." Journal of Eukaryotic Microbiology **46**(4): 402-409.

Elbrachter, M. (2003). "Dinophyte Reproduction: Progress and Conflicts." Journal of Phycology **39**(4): 629-632.

Figueroa, R. I., E. Garces, et al. (2007). "Comparative Study of the Life Cycles of *Alexandrium tamutum* and *Alexandrium minutum* (Gonyaulacales, Dinophyceae) in Culture1." Journal of Phycology **43**(5): 1039-1053.

Figueroa, R. I., E. Garcés, et al. (2010). "The use of flow cytometry for species identification and life-cycle studies in dinoflagellates." Deep Sea Research II **57**(3-4): 301-307.

Guillard, R. R. L. and J. H. Ryther (1962). "Studies of marine plankton diatoms I. *Cyclotella nana* Hustedt and *Detonula confervacea* (Cleve) Gran." Canadian Journal of Microbiology **8**: 229-239.

Haig, D. (1993). "Alternatives to Meiosis: The Unusual Genetics of Red Algae, Microsporidia, and Others." J. Theor. Biol. **163**: 15-31.

Hartwell, A. D. (1975). Hydrographic factors affecting the distribution and movement of toxic dinoflagellates in the western Gulf of Maine. Toxic Dinoflagellate Blooms. V. R. LoCicero. Wakefield, MA, Massachusetts Science and Technology Foundation: 47-68.

Hasle, G. R. (1950). "Phototactic Vertical Migration in Marine Dinoflagellates." Oikos **2**(2): 162-175.

Hou, Y. and S. Lin (2009). "Distinct Gene Number-Genome Size Relationships for Eukaryotes and Non-Eukaryotes: Gene Content Estimation for Dinoflagellate Genomes." PLoS ONE **4**(9): e6978.

Janse, C. J., P. H. van Vianen, et al. (1987). "*Plasmodium* species: Flow cytometry and microfluorometry assessments of DNA content and synthesis." Experimental Parasitology **64**(1): 88-94.

Kita, T., Y. Fukuyo, et al. (1993). "Sexual reproduction of *Alexandrium hanoi*." Bull. Plankton Soc. Japan **39**: 79-85.

Lewis, D. M. (2005). "A simple model of plankton population dynamics coupled with a LES of the surface mixed layer." Journal of Theoretical Biology **234**: 565-591.

Lin, S., H. Zhang, et al. (2006). "Development of a Dinoflagellate-Oriented PCR Primer Set Leads to Detection of Picoplanktonic Dinoflagellates from Long Island Sound." Appl. Environ. Microbiol. **72**(8): 5626-5630.

MacIntyre, J. G., J. J. Cullen, et al. (1997). "Vertical migration, nutrition and toxicity in the dinoflagellate *Alexandrium tamarense*." Marine Ecology Progress Series (Mar. Ecol. Prog. Ser.).

McGillicuddy, D. J., D. M. Anderson, et al. (2005). "Mechanisms regulating large-scale seasonal fluctuations in *Alexandrium fundyense* populations in the Gulf of Maine: Results from a physical-biological model." Deep-Sea Research II **52**: 2698-2714.

Mulligan, H. F. (1975). Oceanographic factors associated with New England red-tide blooms. Toxic Dinoflagellate Blooms. V. R. LoCicero. Wakefield, MA, Massachusetts Science and Technology Foundation: 23-40.

Olson, R. J. and H. M. Sosik (2007). "A submersible imaging-in-flow instrument to analyze nano- and microplankton: Imaging FlowCytobot." Limnology and Oceanography: Methods **5**: 195-203.

Parrow, M., J. M. Burkholder, et al. (2002). "Vegetative and sexual reproduction in *Pfiesteria* spp. (Dinophyceae) cultured with algal prey, and inferences for their classification." Harmful Algae **1**(1): 5-33.

Parrow, M. W. and J. M. Burkholder (2003). "Reproduction and Sexuality in *Pfisteria shumwayae* (Dinophyceae)." Journal of Phycology **39**(4): 697-711.

Passow, U. (1991). "Vertical migration of *Gonyaulax catenata* and *Mesodinium rubrum*." Marine biology , Heidelberg **110**(3): 455-463.

Pfiester, L. A. (1989). "Dinoflagellate sexuality." International Review of Cytology **114**: 249-272.

Pfiester, L. A. and D. M. Anderson (1987). Dinoflagellate Reproduction. The Biology of Dinoflagellates. F. J. R. Taylor. Oxford, Blackwell Scientific Publications: 611-648.

Prescott, D. M. (1994). "The DNA of Ciliated Protozoa." Microbiological Reviews **58**(2): 233-267.

Raikov, I. B. (1995). "Meiosis in Protists: Recent Advances and Persisting Problems." European Journal of Protistology [Eur. J. Protistol.]. **31**: 1-7.

Ralston, D. K., D. J. McGillicuddy, Jr., et al. (2007). "Asynchronous vertical migration and bimodal distribution of motile phytoplankton." J. Plankton Res. **29**(9): 803-821.

Reguera, B., I. Bravo, et al. (1996). Phased cell division and other biological observations in field populations of *Dinophysis* spp during cell cycle studies. Paris (France), UNESCO.

Shapiro, H. M. (2003). Practical Flow Cytometry. Hoboken, NJ, John Wiley & Sons, Inc.

Sosik, H. M. and R. J. Olson (2007). "Automated taxonomic classification of phytoplankton sampled with imaging in-flow cytometry." Limnology and Oceanography: Methods [Limnol. Oceanogr. Methods]. **5**: 204-216.

Taroncher-Oldenburg, G., D. M. Kulis, et al. (1997). "Toxin variability during the cell cycle of the dinoflagellate *Alexandrium fundyense*." Limnology and Oceanography [Limnol. Oceanogr.] **42**(5, Pt. 2): 1178-1188.

Taroncher-Oldenburg, G., D. M. Kulis, et al. (1999). "Coupling of Saxitoxin Biosynthesis to the G1 Phase of the Cell Cycle in the Dinoflagellate *Alexandrium fundyense*: Temperature and Nutrient Effects." Natural Toxins **7**: 207-219.

Taylor, F. J. R. (1968). "Parasitism of the toxin-producing dinoflagellate *Gonyaulax catenella* by the endoparasitic dinoflagellate *Amoebophrya ceratii*." J. Fish. Res. Bd. Can. **25**(10): 2241-2245.

Tomas, R. (1974). "Cell Division in *Gonyaulaz catenella*, a Marine Catenate Dinoflagellate." Journal of Eukaryotic Microbiology [J. EUKARYOT. MICROBIOL.]. **21**(2): 316-321.

Van Dolah, F. M. and T. A. Leighfield (1999). "Diel Phasing of the Cell-Cycle in the Florida Red Tide Dinoflagellate, *Gymnodinium breve*." Journal of Phycology **35**(6): 1404-1411.

Van Dolah, F. M., T. A. Leighfield, et al. (2008). "Cell cycle behavior of laboratory and field populations of the Florida red tide dinoflagellate, *Karenia brevis*." Continental Shelf Research **28**(1): 11-23.

Van Dolah, F. M., T. A. Leighfield, et al. (1995). "Cell Division in the Dinoflagellate *Gambierdiscus toxicus* is Phased to the Diurnal Cycle and Accompanied by Activation of the Cell Cycle Regulatory Protein, CDC2 Kinase." Journal of Phycology **31**(3): 395-400.

Wilkins, A. S. and R. Holliday (2009). "The Evolution of Meiosis From Mitosis." Genetics **181**(1): 3-12.
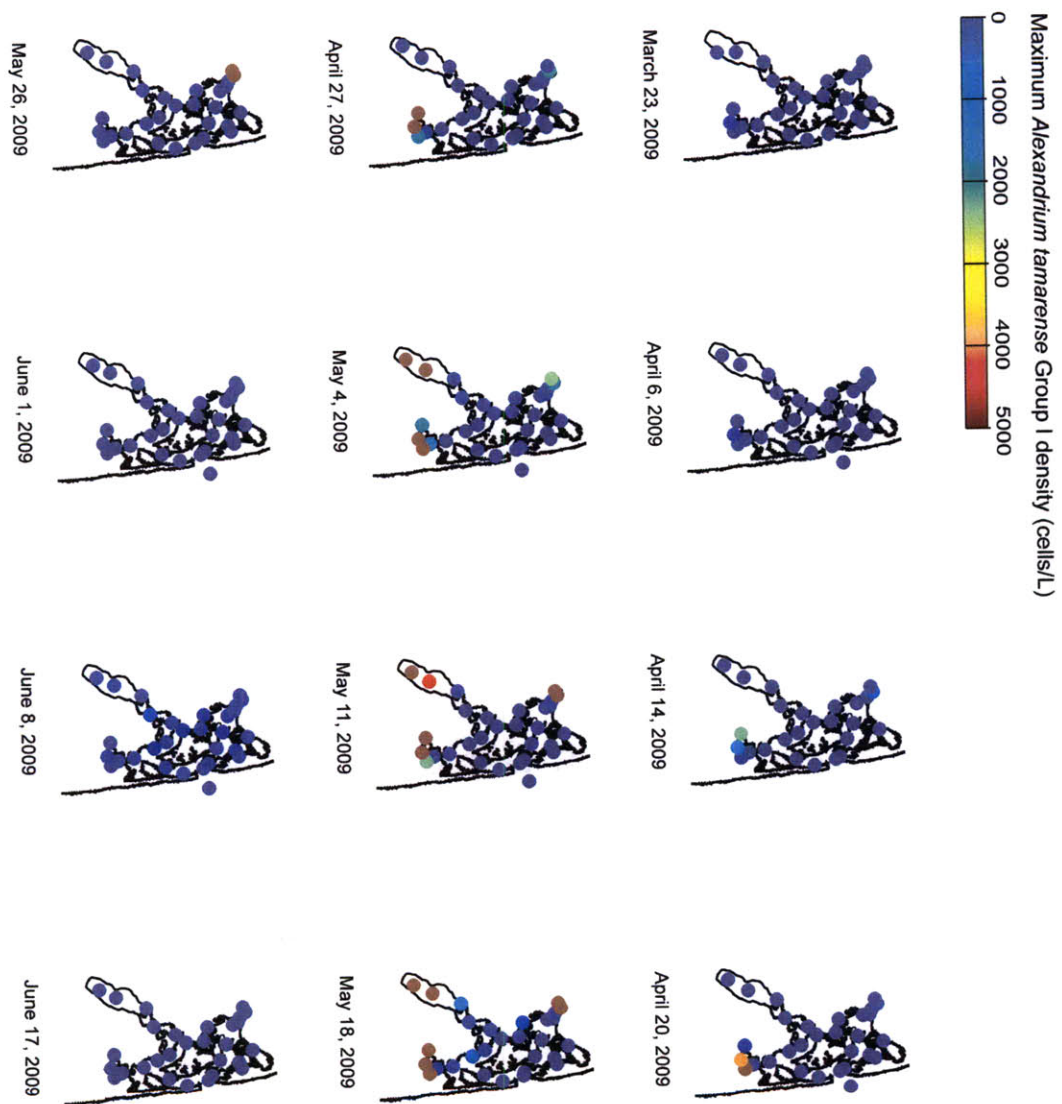
Supplementary Figure 1. Maximum observed *A. t.* Group I densities during weekly

survey of the Nauset Marsh system, March – June 2009.

Supplementary Figure 2. Hydrographic data from GoMOOS Buoys A, B, and E (red, blue and black respectively) leading up to discovery of the Portsmouth red tide July 10th and the emergency response survey conducted 19-20 July, 2009. Current, temperature, and salinity measurements were taken at a depth of 1m.

Wind (6 h bins)

Current (6 h bins)

Salinity (hourly)

Density (hourly)

Temperature (hourly)

Red water observed during buoy deployment near the GOMOOS B mooring.

No red water observed during aerial survey (0 - 5 mi. from coast; Hyannis, MA to Bar Harbor, ME).

159
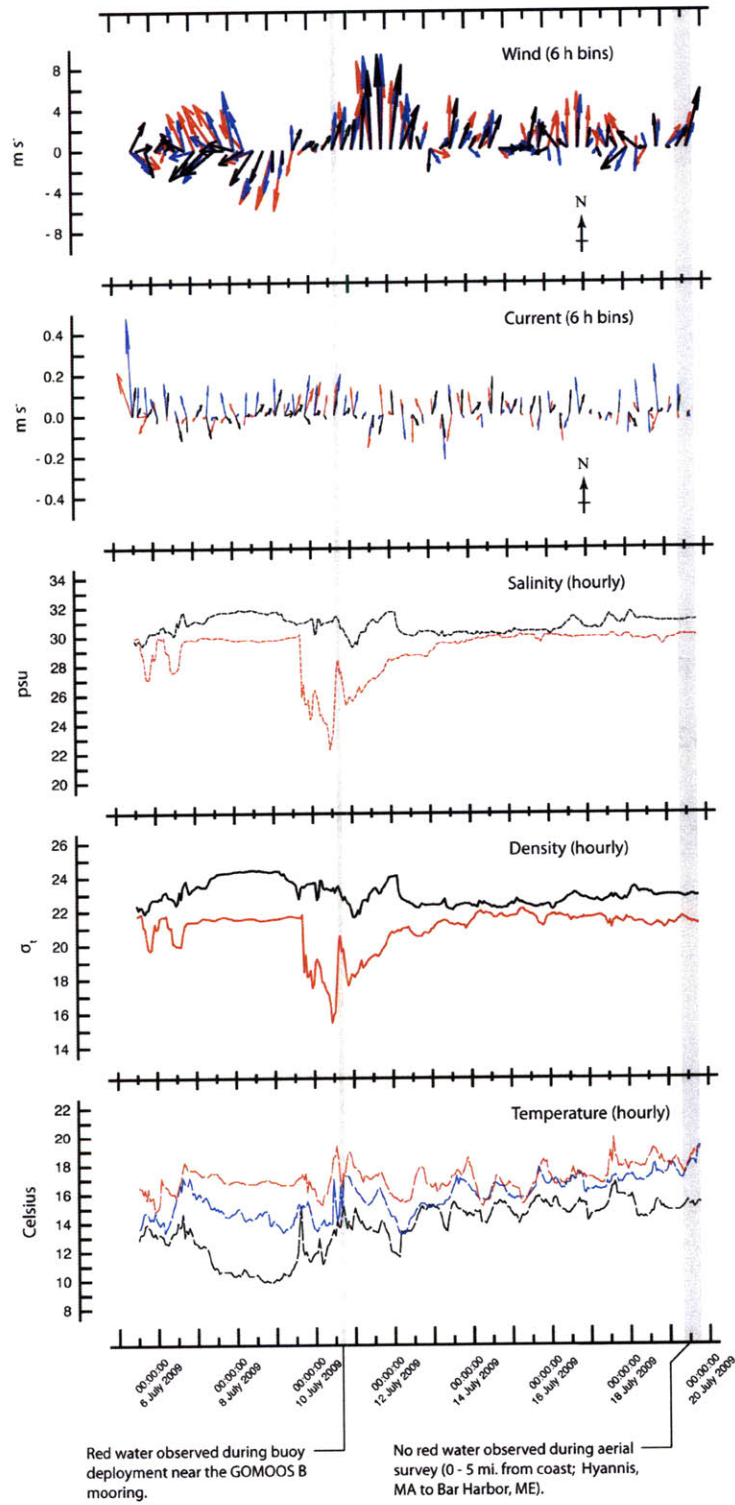
Supplementary Figure 3. Surface layer (0-1 cm) surveys of *A. t.* Group I hypnozygote
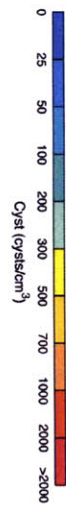
abundance 2005-2009.
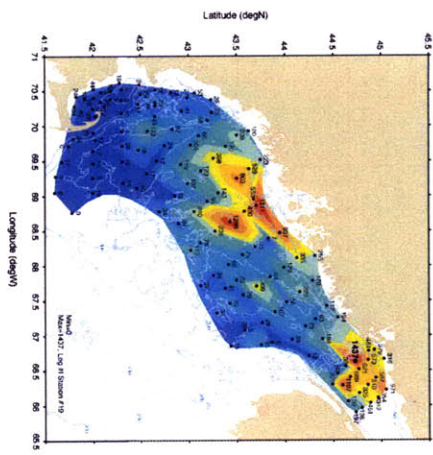
Cyst Counts, Sediment layer 1cm
R/V Oceanus OC416, September 21–29, 2005
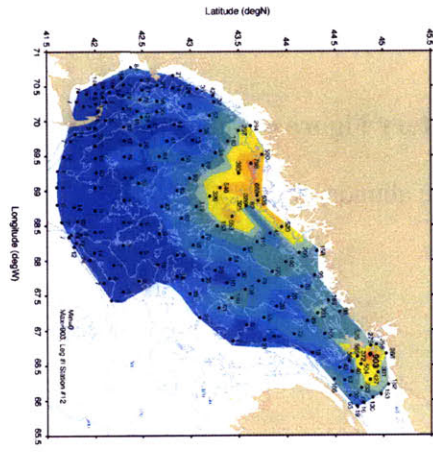
Cyst Counts, Sediment layer 1cm
R/V Oceanus OC433, October 31 – November 12, 2006
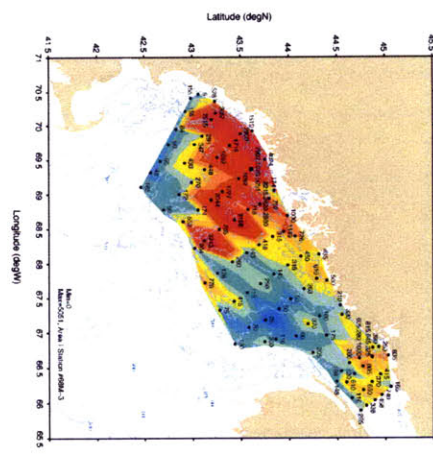
Cyst Counts, Sediment layer 1cm
R/V Endeavor EN456, October (14–24) 2008
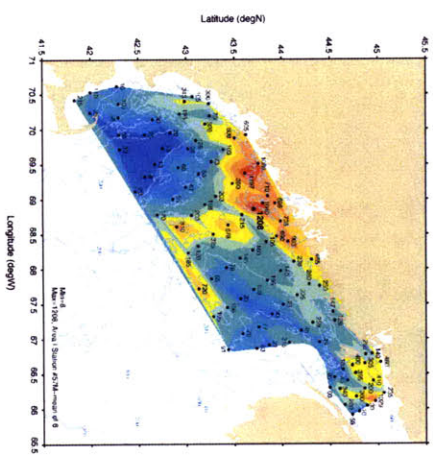
Cyst Counts, Sediment layer 1cm
R/V Oceanus OC458, October (15–25) 2009

Cyst Counts, Sediment layer 1cm
R/V Oceanus OC440, October (8–18), 2007
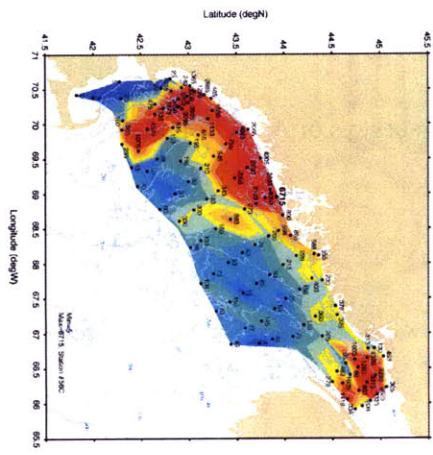
Supplementary Figure 4. Sub-surface layer (1-3 cm) surveys of *A. t.* Group I
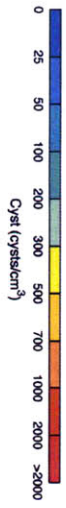
hypnozygote abundance 2005-2009.
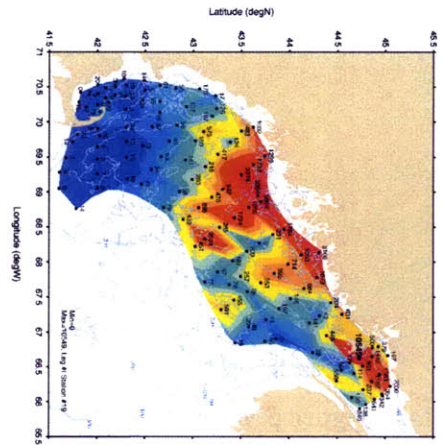
Cyst Counts, Sediment layer 1-3 cm
R/V Oceanus OC416, September 21-29, 2005
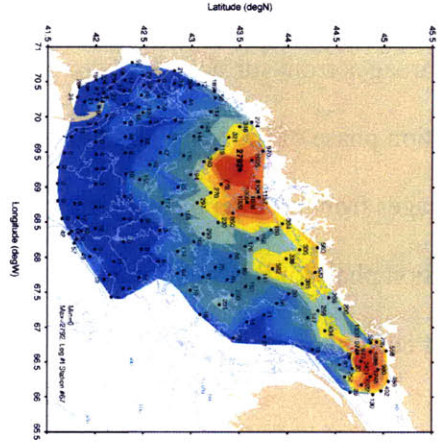
Cyst Counts, Sediment layer 1-3 cm
R/V Oceanus OC433, October 31 – November 12, 2006
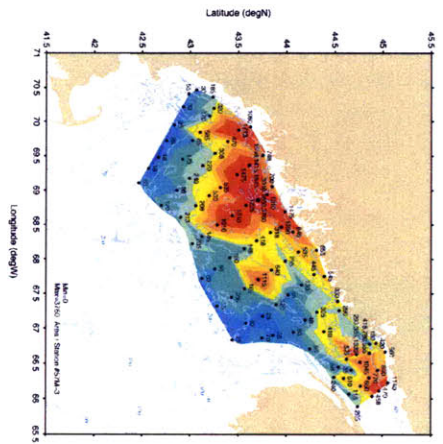
Cyst Counts, Sediment layer 1-3 cm
R/V Endeavor EN456, October (14-24) 2008

Cyst Counts, Sediment layer 1-3 cm
R/V Oceanus OC459, October (15-25) 2009

Cyst Counts, Sediment layer 1-3 cm
R/V Oceanus OC440, October (8-18) 2007

Cyst (cysts/cm^3)

163
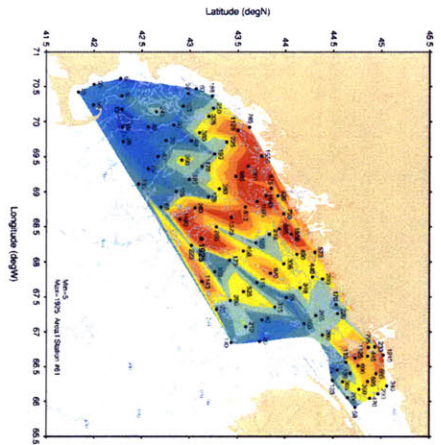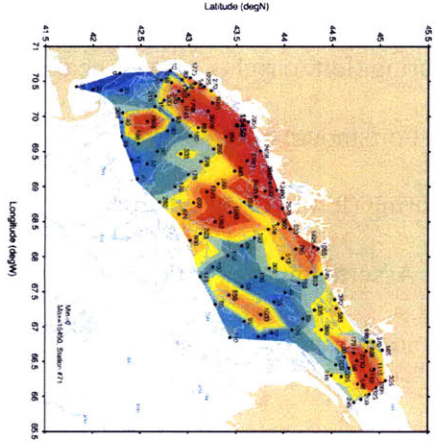
Supplementary Figure 5. *Row A*. Depth-specific histograms of NA1-Cy3 staining

intensity for field samples from Salt Pond and Station 4 of the Portsmouth red tide. *Row*

B. Assorted images from surface *A. t*. Group I taken during (left to right) growth, peak

and termination phase of the Salt Pond bloom, and the Portsmouth red tide. *Row C*.

Assorted images from mid-depth *A. t*. Group I taken during (left to right) growth, peak

and termination phase of the Salt Pond bloom. *Row D*. Assorted images from near bottom

A. t. Group I taken during (left to right) growth, peak and termination phase of the Salt

Pond bloom.

A

Normalized Counts

'Growing'
15:11 hrs
11 May 2009

Surface
Midwater
Bottom

Relative Fluorescence Units (>555 nm)

'Peak'
08:02 hrs
18 May 2009

'Terminating'
14:46 hrs
26 May 2009

Red tide Sta. 4
16:32 hrs
10 July 2009

B

C

D

165
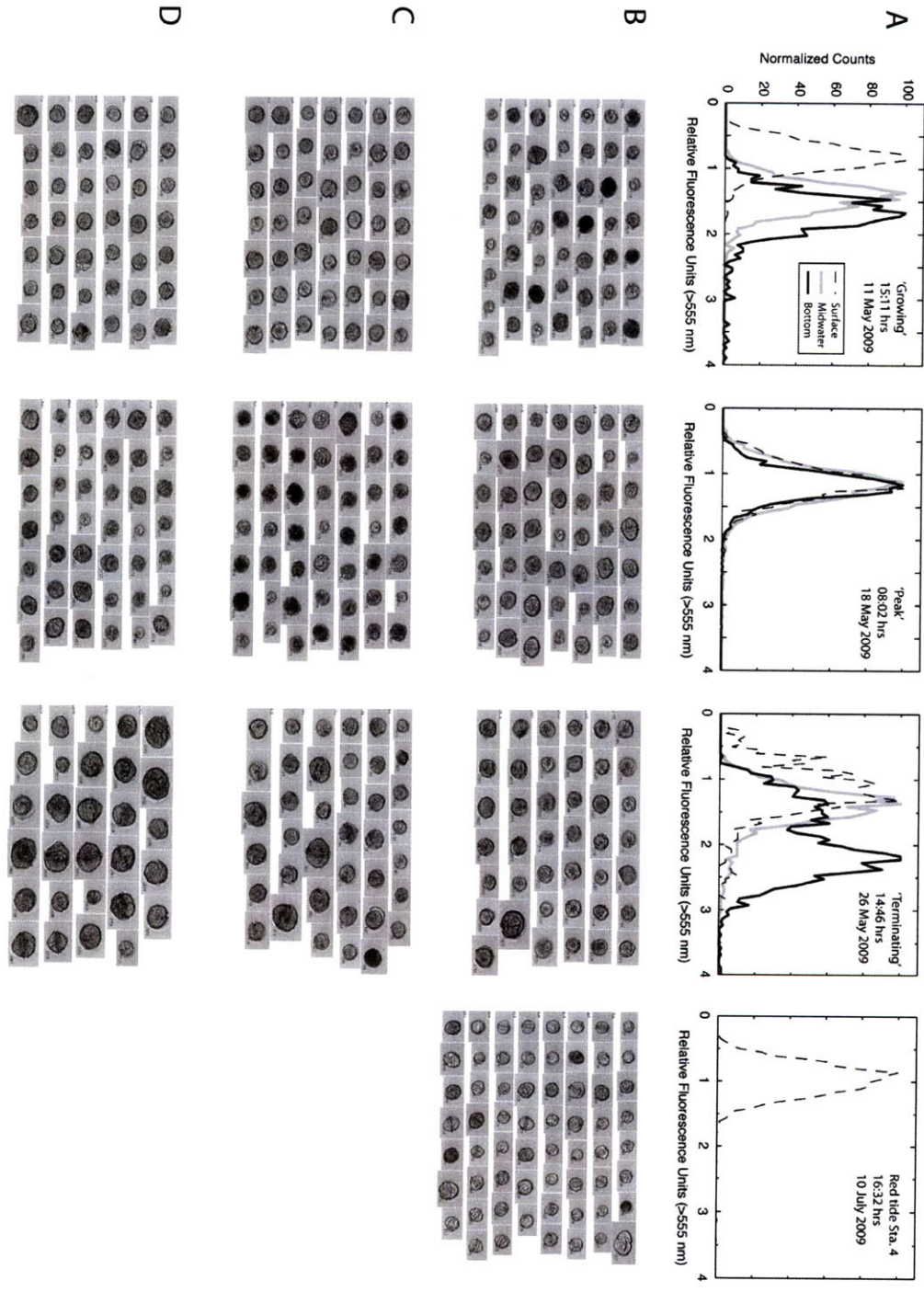
Chapter 3: *de novo* assembly of four *Alexandrium tamarense* transcriptomes

ABSTRACT

Recent advancements have dramatically increased the capacity to produce DNA sequence and opened the field of whole transcriptome quantitative gene expression to studies of non-model organisms. Here, commercially available reagent kits and protocols were used to sequence the transcriptomes of three species within the *Alexandrium tamarense* species complex, a group of dinoflagellates that sometimes cause Paralytic Shellfish Poisoning (PSP). A total of twelve cDNA libraries were sequenced using two different sequencing platforms: 454/Roche pyro-sequencing and Illumina reversible terminator sequencing. Preparation of cDNA libraries is described for ten RNA-Seq libraries from vegetative cell phase cultures and a single library prepared from naturally occurring hypnozygotes. As a first step in the analysis of these datasets, whole transcriptome assembly was completed using the parallelized assembler ABySS across a large range of the de Bruijn graph parameter k. Individual builds were assessed using various contig and assembly metrics as well as the fraction of MPSS transcript signatures with matches in the final assemblies. All of the criteria were assessed and used to justify final parameter selection according to the fraction of MPSS tags matched. Future steps in the assembly process will include mixed parameter transcriptome models and contig annotations derived from BLAST. Future applications for these datasets include evolutionary studies of transcriptome sequence divergence within the *A. tamarense* species complex, quantitative comparisons of transcript abundance, and other gene mining studies.

INTRODUCTION

In recent years, new DNA sequencing platforms have dramatically reduced the cost of genome and transcriptome sequencing projects while simultaneously enabling

huge increases in their scale. Here, new platforms developed by 454/Roche and Illumina

(Margulies, Egholm et al. 2005; Bentley, Balasubramanian et al. 2008) were used to

sequence mRNA transcripts from three algal dinoflagellates in the *Alexandrium*

tamarense species complex. The vast majority of new sequence data were derived from

the Illumina platform using a commercial RNA-Seq protocol. The advantages of this

approach are that a tremendous volume of data is produced ($>10^8$ 'reads' in a single

sequencing run) and that the relative abundance of sequencing templates is maintained.

The resulting sequencing reads can be mapped to a reference genome or transcriptome

assembly to generate quantitative estimates of expression for all genes transcribed by a

cell population. Moreover, the dynamic range of RNA-Seq expression estimates is

superior to previous technologies such as microarrays and gene signature methods (e.g.

SAGE and MPSS; Wang et al. 2009). A handicap of the RNA-Seq approach marketed

by Illumina has been its relatively short reads (initially only 30 bp from a single template

end). The tremendous volume of data and low information content of individual reads

(compared to previous generation sequencers) has necessitated the development of new

algorithms for their *de novo* assembly (Miller, Koren et al. 2010).

The assembly of non-normalized transcriptome sequences is an especially

difficult problem because different transcript species vary substantially in abundance,

requiring large datasets for sufficient sequencing coverage of rare transcripts. At the time

this project was initiated, several short-read assembly algorithms were in active open-

source development including: Velvet (Zerbino and Birney 2008), VCAKE (Jeck,

Reinhardt et al. 2007), SSAKE (Warren, Sutton et al. 2007), EULER-SR (Chaisson and

Pevzner 2008), ALLPATHS (Butler, MacCallum et al. 2008), and ABySS (Simpson, Wong et al. 2009). Each of these use a de Bruijn graph-based approach that greatly reduces the computational resources required for assembly in comparison to the previous standard approach, OLC (Overlap/ Layout/ Consensus; see Miller et al. 2010). ABySS was chosen for this project because it was the first suite applied to the problem of *de novo* transcriptome assembly (Birol, Jackman et al. 2009), and because it enables the distribution of assembly steps across multiple processing nodes, negating computational limits to the amount of data that could be used. Additional steps taken to aid transcriptome assembly directly from RNA-Seq data sets included the use of paired-end sequencing whereby templates were read from both forward and reverse primers during Illumina sequencing, and the application of additional sequencing reagents to extend read lengths up to 108 base pairs.

The necessity of assembling transcriptome sequences rather than those derived from genomes was due in part to the nature of the *A. tamarense* species. Dinoflagellates like *A. tamarense* are known for their unique chromosomal structure and exceptionally large genomes (ca. 2×10^{11} bases or approximately 70x the size of the human genome; Hackett et al. 2005). The large size of the *A. tamarense* genomes but low gene density (Erdner and Anderson, 2006; Hou et al. 2009; Moustafa et al. 2010) strongly indicated *de novo* transcriptome assembly for maximum efficiency in gene discovery. Three *A. tamarense* species were sequenced: Group I, Group III, and Group IV. Two of these species, Group I and Group IV, pose a significant human health threat because they produce saxitoxin and congener compounds that cause Paralytic Shellfish Poisoning

(PSP) in human and animal consumers of contaminated seafood. The group designations for the species were proposed by Lilly et al. to describe five ribosomal clades found among globally distributed isolates, each with distinct ranges, and also consistent with respect to their PSP toxicity (Lilly, Halanych et al. 2007). Brosnahan et al. (2010) showed that though clones of the respective clades may form sexual cysts, those formed by conjugation of two of the most closely related clades, Group I and Group III, are unable to resume vegetative cell division. Together, the results reported by Lilly et al. and Brosnahan et al. established complete speciation between the *A. tamarense* clades in spite of their close morphological resemblance and mating intercompatibility.

Previous gene discovery and whole transcriptome gene expression studies of *A. tamarense* species had been restricted to the haploid, vegetative phase of their life cycle and had yet to define the complete complement of gene sequences expressed by these cells (Hackett, Scheetz et al. 2005; Erdner and Anderson 2006; Moustafa, Evans et al. 2010). A comprehensive assembly of all *A. tamarense* genes expressed during the vegetative cell phase was sought using the newly available sequencing methods by 454/Roche and Illumina. The transcriptome of the diploid hypnozygote stage was also explored using naturally occurring cysts collected from a sediment core in the Gulf of Maine. The impetus to use naturally produced hypnozygotes rather than those from cultures was the poor germinability of cultured cysts (Brosnahan, Kulis et al. 2010) and the difficulty in separating cultured hypnozygotes from other life cycle stages because of the large mucilage aggregates they form during encystment.

MATERIALS AND METHODS

Cultures and Resting Cyst Isolation

Cultures for vegetative phase libraries were prepared with four *A. tamarense*
clones: two PSP toxic Group I clones (38-3 and GTM253-17), one nontoxic Group III
clone (ATSP1-B), and one PSP toxic Group IV clone (CCMP 1598). Group I and III cell
pellets were harvested during late log phase of cultures grown in f/2 -Si and encystment
medium (Brosnahan, Kulis et al. 2010). Pellets were lysed by resuspension in Trizol
(Invitrogen), then flash frozen and stored in liquid nitrogen until RNA extraction and
library construction. Unlike Group I and Group III cultures, the Group IV clone was
axenized prior to this study and was propagated using sterile techniques. Four Group IV
pellets were prepared: (1) a late-log phase f/2 -Si culture, (2) a semi-continuous culture
grown in f/2 -Si (0.4 division day^{-1} steady-state condition > 3 days), (3) a semi-
continuous culture grown in f/2 -Si with ammonium substitution of nitrate (0.4 divisions
day^{-1} steady-state condition > 3 days), and (4) a G1 phase culture grown in f/2 -Si and
synchronized as described by Taroncher-Oldenburg et al. (1997). An additional log phase
culture was grown in f/2 -Si medium for 454/Roche sequencing (see Hackett et al. *in prep*
for a complete description).

Our initial intent was to generate several hypnozygote samples from multiple time
points during the transition from dormancy to germination, but only one sample was
successfully carried to cDNA synthesis. Several steps were taken to ensure that cysts

would be plentiful, and that their viability would be maximized and validated prior to library construction. The samples were collected from a set of three Craib cores taken during an October 2007 benthic survey of the Gulf of Maine aboard the R/V Oceanus and within Casco Bay, an area where cyst abundance was known to exceed 1000 cc^{-1} mud. Because cysts are sensitive to light, elevated temperature, and oxygen (Binder and Anderson 1986; Anderson, Taylor et al. 1987), the cores were refrigerated immediately after retrieval, then their 1-6 cm layers combined on board in a single 500 cc amber jar with minimal headspace. The combined sample was stored at 4°C until aliquoted to several 60 cc amber jars in May 2008 under an N$_2$ (+ 0.3% H$_2$) atmosphere and under a red light, then returned to 4°C storage until removal of individual jars at three time points in May 2008, January 2009, and April 2009. These time points corresponded to two expected maxima (May and April) and a minimum (January) of the cysts' germinability that result from an endogenous circannual rhythm (Anderson and Keafer 1987; Matrai, Thompson et al. 2005). Each sample was diluted with cold 0.2 μm filtered seawater to 600 mL and aliquoted to 30 culture flasks containing approximately 4000 cysts in a dark, refrigerated room. Flasks were then transferred to a 6°C incubator under a 14:10 light:dark cycle and maximal irradiance less than 15 μE s^{-1} m^{-2} to slow germination for observation of chloroplast reconstitution that is known to precede germination (Anderson, Stock et al. 2005).

Maximum vigor of the cysts was validated by monitoring germination of 250-500 cysts isolated to 96-well plates within twenty-four hours of dilution. Five flasks were removed from the 6°C incubator during the first light period, their slurries combined and

disaggregated using a sonicating probe, and cysts concentrated by centrifugation over a Nalco-sucrose gradient (Schwinghamer, Anderson et al. 1991). Cysts were isolated by micropipette as single cells or clumps in a cold room, then incubated under 14:10 light:dark cycle in a 15°C incubator. The plates were monitored daily for cyst germination up to ten days, and germination frequencies of 91.2%, 26.3%, and 76.5% were observed for cysts collected in May 2008, January 2009, and April 2009 respectively. The result confirmed that cysts were living and that their germination was under control of a circannual rhythm. In addition to the proportion germinating, cyst abundance and fluorescence were monitored daily for five days following initial dilution and every other day thereafter.

Finally, clutches of 300 to 500 cysts were isolated by micropipette for RNA extraction and amplification from separate slurry flasks while cyst abundance was still high enough for efficient extraction (up to 13 days after dilution). Flasks for clutch isolation were removed from the 6°C incubator within 1 hour of 'daybreak' (incubator lights turned on), sonicated using a probe, and centrifuged over a Nalco-sucrose gradient to separate cysts from heavy sediment and debris. Individual cysts were transferred by micropipette to a microcentrifuge tube with 200 μL 0.2 μm filtered cold seawater, again in a cold room and within three hours of sonication. After three hours, the microcentrifuge tube was capped and spun for 5 minutes, >12,000 rcg, then aspirated by pipette to a total volume less than 50 μL. Lastly, the cyst clutches were resuspended in Trizol and flash frozen in liquid nitrogen.

RNA Extraction, Library Preparation, and Sequencing

All samples except a single CCMP 1598 (Group IV) sample were prepared for sequencing on the Illumina Genome Analyzer II platform. However, the basic steps in preparing templates for both the 454 and Illumina platforms are much the same. First, poly-adenylated RNA is extracted from cell samples, fragmented, and used as a template for cDNA synthesis. Next, platform appropriate primers are ligated to the cDNA fragments, and products are size selected for delivery to the sequencer. Additional steps were taken to normalize transcript species and to select full-length poly-adenylated RNAs for libraries sequenced by 454 (see Hackett et al. *in prep*). Libraries prepared for Illumina sequencing were prepared as described below.

Total RNA was extracted from both vegetative cell pellet and cyst samples according to the manufacturer's protocol for Trizol. After extraction, RNA from vegetative cell pellets was precipitated and resuspended in RNAse free water, then prepared for sequencing using an RNA-Seq reagent kit from Illumina (RS-100-0801). All manufacturer instructions were followed except that slightly larger (~300 bp) fragment-primer products were isolated during size selection in order to avoid overlapping paired-end reads during sequencing.

The expected yield from RNA extraction of cyst samples was less than 30 ng, much less than the yield from vegetative cell pellets. Therefore, no precipitation was attempted from these samples and the aqueous phase of these RNA extracts was concentrated using an RNAdvance Tissue Isolation kit (Agencourt, Inc.) as described by

Quigley et al.(2010). Antisense copies of poly-adenylated transcripts were then

selectively amplified by the Eberwine method using a MessageAMP kit (Applied

Biosystems, formerly Ambion). Success of the isolation/amplification scheme was

assessed for each sample from a BioAnalyzer 2100 trace (RNA 6000 Pico Kit, Agilent

Technologies). Only one sample, which was collected within twenty-four hours of

dilution during the first active germination period (May 2008), yielded more than 50 ng

product and was the only sample prepared further for sequencing. The amplification

product was fragmented without additional poly-A selection and prepared as previously

described for vegetative cell libraries using an Illumina RNA-Seq reagent kit.

Each of the Illumina libraries was sequenced using a single lane of a flow cell on

an Illumina Genome Analyzer II by the Biomicro Center at MIT (Cambridge, MA). Read

lengths up to 108 bp were attempted, but no base calls after position 72 were used during

assembly because of poor estimated read quality.

Transcriptome Assembly

Three computer systems were used for assembly: a Macintosh workstation

(2x2.66 GHz Quad-Core Intel Xeon processors, 48 GB 1066 MHz DDR3 RAM, Mac OS

X 10.6), a Linux workstation (2x2.66 GHz Quad-Core Intel Xeon processors, 48 GB

1066 MHz DDR3 RAM, RHEL 5.3), and a Linux cluster (16 2x2 GHz Quad-Core Intel

Xeon nodes with 2 GB 1066 MHz DDR3 RAM each, CentOS 5, Sun Grid Engine

version 6.1). ABySS version 1.2.0 was installed on both the Macintosh workstation and

Linux cluster, and version 1.2.2 was installed on the Linux workstation. Only minor

changes were made in the update from 1.2.0 to 1.2.2, but both versions were substantially

altered from version 1.0 used in the publication by Birol et al. (2009). The changes from version 1.0 included automated quality-based trimming of reads and read-depth calculations that set parameters c, e, E, and n. The parameter settings were over-ridden by manually setting them to 2, 0, 0, and 2 respectively according to directions posted by Shaun Jackman (BC Cancer Agency, Vancouver, Canada) at the ABySS users webpage (http://groups.google.com/group/abyss-users). The most important remaining parameter was k, a basic variable in all de Bruijn graph based assemblers that determines the initial fragmentation of sequencing reads. In the first step of a de Bruijn based assembler, each read of length L is broken into L-k+1 short, overlapping subsequences. The longer the length of k, the longer the repeat motifs that can be faithfully reconstructed during assembly, but also the more false contigs that are generated due to sequencing errors. Due to this trade-off, builds were completed across a range in k-space (values between 20 and 64) before a single build was selected for further analysis.

Assemblies from builds were compared according to several build properties, including: mean, median, and max contig lengths; N50 value; the sum of all contigs greater than 100 bp; and the fraction of MPSS tags with identical matches (See Figures 2 and 3). The last criterion was in addition to those prescribed by Birol et al. and was justified by the rationale that all MPSS tags ought to have identical matches in the transcriptome assemblies if both data sets were complete and derived from a single source (vegetative clone). Such is the case for the assembly of the Group IV transcriptome where both the MPSS libraries and cDNA libraries were derived from CCMP 1598 (Moustafa, Evans et al. 2010). For Group I and Group III clones, assemblies

were queried by MPSS tags from GTCA28, a Group I clone (Erdner and Anderson, 2006), because it was the most closely related by ribosomal sequence (Lilly, Halanych et al. 2007). The process of matching tags to contigs was completed using Bowtie version 0.12.7 (Langmead, Trapnell et al. 2009). Upon reviewing all of the criteria, the value of k that maximized the fraction of MPSS tags matched was identical or intermediate to the optimal k values for the other criteria (Figure 2), and therefore was sole criterion used for k-selection.

Final assemblies included all the cDNA libraries sequenced from a given *A. tamarense* clone, but the comprehensiveness of each of the clonal assemblies was also assessed by repeating the assembly process with combinations of the vegetative cell libraries from CCMP 1598 (e.g. six combinations of two libraries, four combinations of three libraries, etc.). Each library represented a distinct level of sequencing 'effort', defined as the quantity of sequence data delivered to the ABySS assembler after quality-based trimming. The size of the transcriptome was estimated by plotting the effort level of each assembly combination versus the sum of its contig lengths and approximating the relationship between these quantities using the Michaelis-Menten equation (Figure 4). After the analysis suggested that the transcriptome of CCMP 1598 is more than 97% complete, its contigs were used to estimate their relative expression using CLC Genomics Workbench (CLC Bio). Relative expression was expressed as reads per kilobase (contig length) per million sequenced reads (RPKM), a standard measure of gene expression in RNA-seq experiments. Then, Group IV contigs were sorted into 20 quantiles according to

their relative expression, and the average contig length and fraction of MPSS tags

matched for each quantile was calculated using Bowtie and a perl script (Figure 4).

Table 1. Sequence library statistics before and after quality-based trimming.

Clone	Libraries	Read length (bases)		Total no. reads (forward and reverse)	Bases after trimming
		Before trimming	Average after trimming		
Group I					
38-3	f/2	108	43.2	30,364,758	1.3×10^9
	encystment	108	32.9	23,113,423	7.6×10^8
GTM253-17	f/2	108	26.1	25,735,899	6.7×10^8
	Encystment	72	67.7	42,858,761	2.9×10^9
Casco Bay	hypnozygotes	108	84.3	32,224,174	2.7×10^9
Group III					
ATSP1-B	f/2	108	36.2	29,311,726	1.1×10^9
	encystment	72	68.9	38,059,250	2.6×10^9
Group IV					
CCMP 1598	f/2	72	68.5	40,351,460	2.8×10^9
	f/2 454	--	309.3	1,206,135	3.7×10^8
	f/2 G1	72	69.3	12,341,882	8.6×10^8
	f/2 semi	72	65.1	56,350,834	3.7×10^9
	+NH$_4$$^{+1}$ semi	72	68.4	40,117,750	2.8×10^9
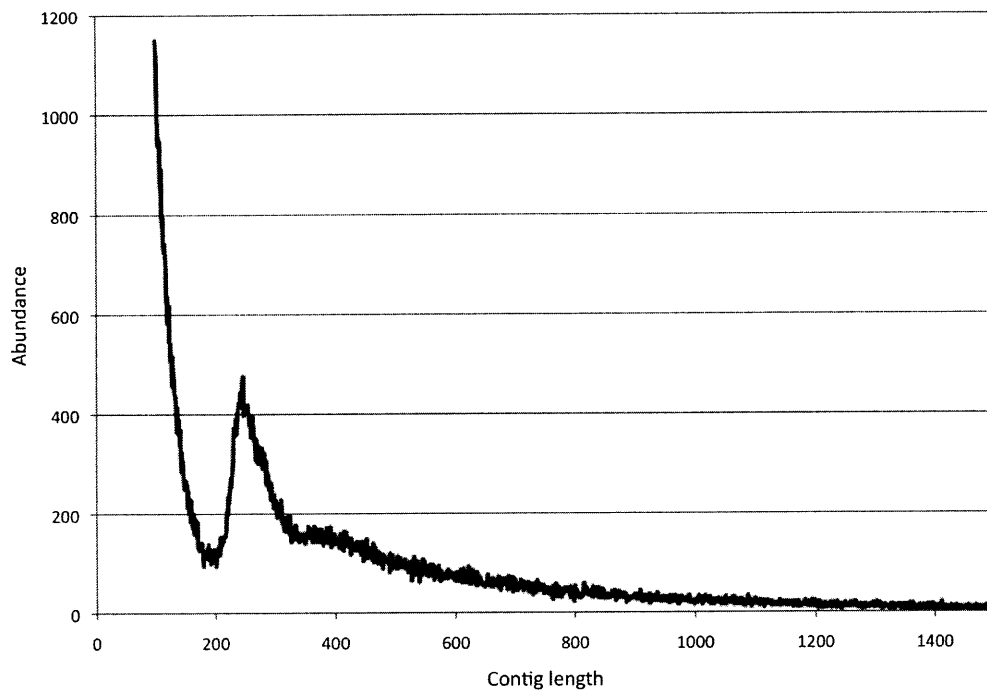
Figure 1. Contig size distribution from an assembly of RNA-Seq data (all RNA-Seq
libraries from CCMP 1598, k = 36). The majority of contigs in all assemblies were short,
and their size distribution had local maxima of length less than the assembly N50 value
(here equal to 580 bases).

RESULTS AND DISCUSSION

Sequencing and Assembly

Two Illumina sequencing flow cells were used to sequence a total of eleven
libraries. The first of these included five lanes sequenced to 108 bp from both the forward
and reverse template ends, but less than half of the base calls were useable (45 bases per

read on average). In contrast, the second flow cell returned better quality data but only 72

bases were sequenced from each template end, and the average usable read length was 68

bases (Table 1). The discrepancy in data output between the flow cells was not due to

differences in the quality of the cDNA templates because two of the template libraries

sequenced in the second run were prepared simultaneously with the five libraries

sequenced on the first flow cell. Moreover, all of the RNA-Seq libraries passed

BioAnalyzer and qPCR quality controls performed by the Biomicro Center at MIT

immediately prior to sequencing. Instead, the variation in read quality and quantity

between runs is due to variations in the sequencer itself and perhaps the maturation of

protocols in the time between the two sequencing runs (completed 6 months apart from

one another). Though no specific protocol changes were noted, sequencing chemistry and

standards for the Illumina Genome Analyzer have evolved rapidly in the three years since

its commercial release.

The distribution of contig lengths in the final assemblies of the four vegetative

clones all had local maxima between 276 and 488 bases, but the assemblies also

contained substantial numbers of longer contigs (Table 2 and Figure 1). The patterns

observed in the size distribution are characteristic of these types of assemblies and are

partly described by the statistic N50, a contig length cutoff over which 50% of the total

bases in an assembly are contained. The lengths of N50 were near 1 kb for all final

vegetative assemblies, indicating that a large proportion of contigs in each were at least

as large as typical expressed sequence tags from Sanger-based chemistry. K-mer sizes

that maximized the size of N50 were also nearly identical to those that produced the maximal fraction of MPSS tags matched.

A contrast to the results reported by Birol et al. (2009) from assembly of a human transcriptome was that the criteria used to assess different k-mer assemblies were stable through the range of values explored (Figure 2). The result suggests that there is little difference among the respective transcriptome assemblies for a given vegetative clone. However, the metrics used in this assessment do not address the actually sequence content of the contigs. One possibility is that different k-mer assemblies may exchange spurious chimera between valid contigs. The stability of the assemblies with respect to k remains to be assessed by pair-wise comparisons of the actual contig sequences.

In contrast to assemblies of the vegetative clone transcriptomes, the hypnozygote library did not yield a substantial proportion of large contigs. The N50 values of the best assemblies were less than one third the length of those from vegetative assemblies and the maximum fraction of Group I MPSS tags matched was less than 5%. In an effort to trouble shoot the hypnozygote assembly, Group I MPSS tags were queried against a database made from the raw sequencing reads of the hypnozygote library using BlastN (Altschul, Madden et al. 1997). The result was that 25.2% of the tags were matched, less than the fraction of Group I tags that matched the Group III vegetative transcriptome assembly (a different species, Table 2). For comparison, a similar query of the Group IV MPSS tags versus a database of all Group IV sequence libraries returned greater than 70% of tags having an identical, un-gapped match.

In one respect, the failure to match similarly large numbers of Group I MPSS tags

to the hypnozygote library was encouraging because the Group I MPSS tags were derived

from two vegetative phase samples, suggesting that a large number of novel transcripts

are expressed by the diploid cyst phase in these organisms' life cycle. However, given the

small size of contigs produced by the assembler and the possibility of contamination from

other cell species (because cysts were isolated from sediment cores), it was decided that

priority would be given to assessing the vegetative clone assemblies. Future evaluation of

the hypnozygote assembly will begin by prioritizing the annotation of large contigs and

those with Group I MPSS tag matches.

Table 2. Assembly metrics from final builds of each clone.

Clone(s)	k-mer	Contig length (bases)			N50 length (bases)	Sum of contigs	Fraction of MPSS tags matched
		Mean	Median	Max			
38-3 (Group I)	25	527	347	8,911	808	6.8×10^7	48.4%
GTM253 (Group I)	41	529	276	11,590	995	9.4×10^7	65.3%
ATSP1-B (Group III)	40	574	300	14,811	1,088	9.7×10^7	25.7%
CCMP 1598 (Group IV)	44	711	488	14,010	1,115	1.0×10^8	67.3%

Initial Analysis of Vegetative Clone Assemblies

The large fraction of MPSS tags matched in each of the clonal assemblies

suggested that many, but not all of the *A. tamarense* transcripts were assembled.

However, this conclusion takes for granted that the number of MPSS tags is
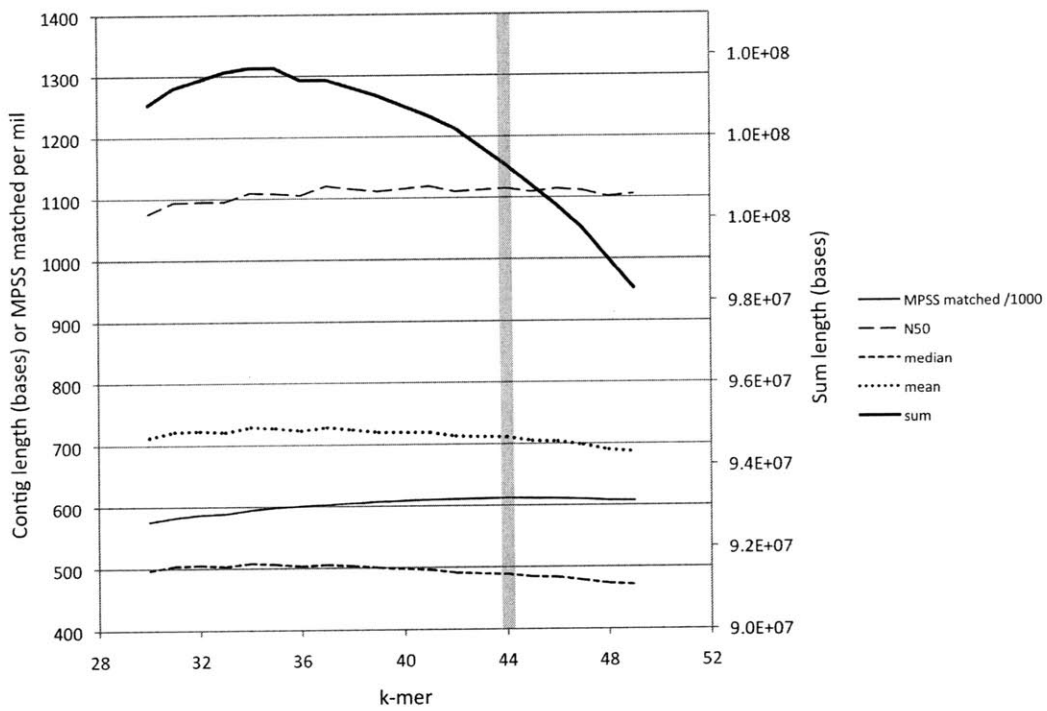
Figure 2. Comparison of various assembly metrics for CCMP 1598. The gray vertical bar

highlights the k-mer size that maximized the fraction of MPSS tags matched.

an accurate measure of the number of genes in these organisms. In fact, it is likely that some tag sequences are spurious, either because they were derived from contaminants or else from repeated sequencing errors (Meyers, Tej et al. 2004). Therefore, a second method was employed to assess the comprehensiveness of the sequencing effort: the construction of a 'pseudo-rarefaction curve'.

In this case, none of the cDNA libraries were true replicates of one another even though all were expected to include most (if not all of) the same gene sequences. Combinations of the libraries still behaved as one would expect if effort were increased toward sequencing a single cDNA library. As more of the libraries were used, gains in the size of the total assembly were diminished (Figure 3), but the largest assembly still utilized all five of the CCMP 1598 libraries. The sum of the contig lengths from the largest assembly was approximately 1.04×10^8 bases, only slightly less than the maximum 1.07×10^8 bases predicted after fitting a Michaelis-Menten equation to the Group IV curve. While the result is sensitive to the choices in regression model and the method of parameterization, it does indicate a much higher level of comprehensiveness that is suggested by MPSS tag matching (up to 97% completion versus 67%; Table 2). No similar process was completed using the Group I and Group III libraries because there were fewer of these libraries. However it is of note that Hacket et al. found orthologs for 14 cyanobacterial genes implicated in the synthesis of saxitoxin in contigs from each of these four clonal transcriptome assemblies, including the nontoxic Group III species.

As a last step to better assess the matching of MPSS tags as a metric to evaluate the transcriptome assemblies, the relative expression of the assembly contigs was

184
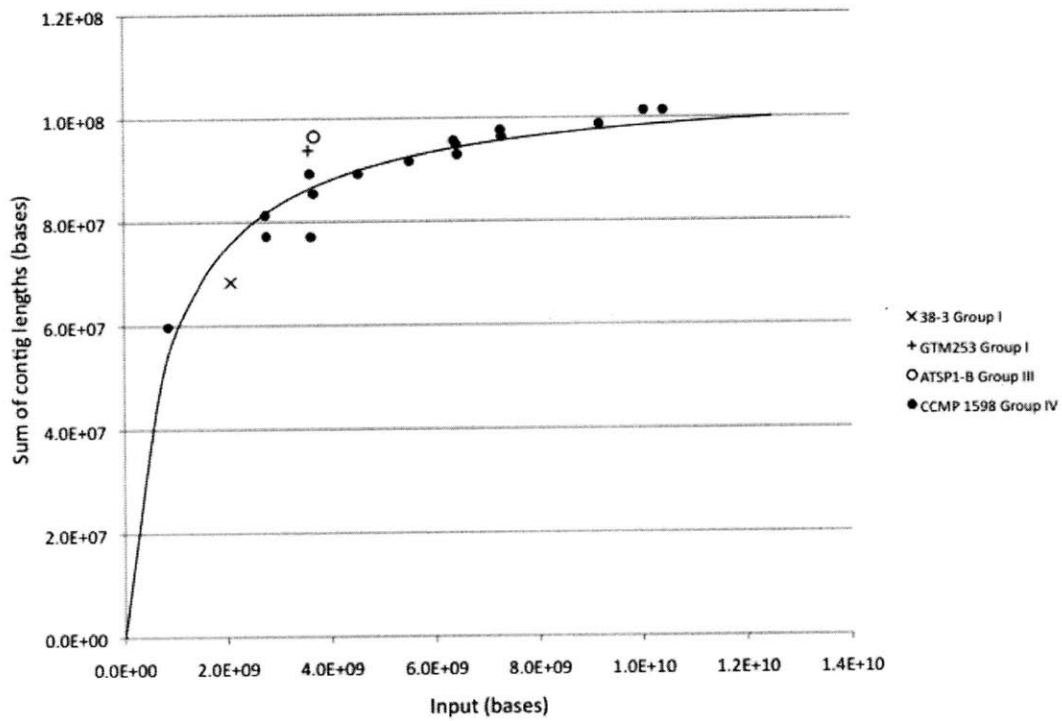
Figure 3. Pseudo-rarefaction of sequencing effort (input) versus assembly size (measured as sum of contig lengths). A Michaelis-Menten equation was fit to points from the Group IV (CCMP 1598) library combinations (solid line) in order to estimate the size of the complete *A. tamarense* Group IV transcriptome.

estimated in RPKM, sorted into 20 quantiles, and the mean length and MPSS tags matched of each quantile compared (Figure 4). The expectation was that contigs with low expression (and therefore low sequencing coverage) would be fragmented. Therefore, on average lowly expressed contigs would be smaller and match a smaller fraction of tags. Both predictions were validated, but somewhat surprisingly the most highly expressed contigs were also relatively short in length. The latter result may have been due to greater inclusion of large homolog families known to occur in dinoflagellate genomes (Hou and Lin 2009). Single nucleotide differences among multiple gene loci present the same problem to de Bruijn graph-based assemblers than sequencing errors do: they cause graph branches that result in contig fragmentation. A test of this interpretation of shorter contig lengths at high levels of gene expression will be the annotation of the contig sequences using BLAST, a step that is in progress.

Next Steps

Many assembly algorithms and protocols are in active development toward the assembly of transcriptomes from short read datasets. Recent reports have indicated that optimal assemblies may only be achieved by comparison of results across multiple k-mer values (Robertson, Schein et al. 2010; Surget-Groba and Montoya-Burgos 2010). While such composite assembly is beyond the scope of this dissertation, the first steps have been completed here. What remains is the comparison of contig sequences from the many
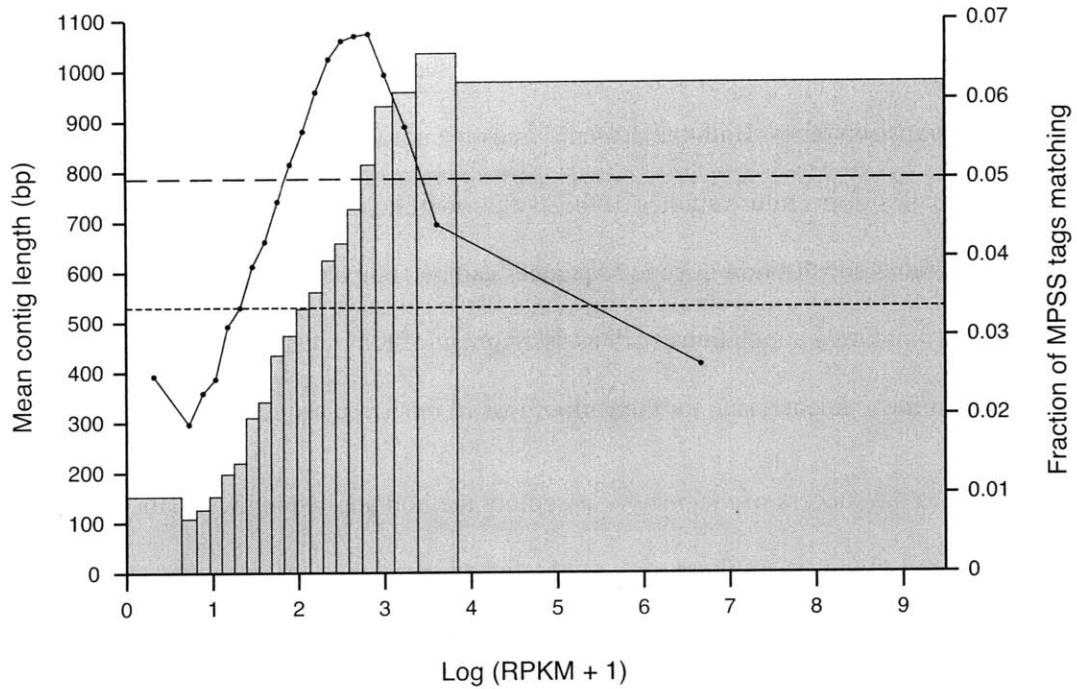
Figure 4. Cascade graph of the final Group IV assembly sorted into 20 quantiles according to RPKM expression estimates. Each gray bar represents a quantile (approximately 7,120 contigs). The width of a bar denotes the range in expression of its contigs, and bar height is the fraction of MPSS tags matched. Horizontal dashed lines are the levels of tags matched expected if all tag matches were distributed evenly among the quantiles (long dashed is the case where all tags were mapped; short dashed where only those tags that did match were considered). The solid circles/line are the mean lengths of the contigs within each quantile.

assemblies produced as part of this effort. BLAST-based annotation of the contigs in the four vegetative assemblies are in progress. Results from this effort will be used to identify contaminant sequences and also to compare the genes present in each of the *A. tamarense* transcriptomes. Both the presence and absence of some contigs between the assemblies will inform future studies of the evolution of this species complex. Rates of non-synonymous substitution in gene sequences can be compared within genes that are shared, and functional annotations such as assignment of KEGG and Gene Ontology terms may inform mechanisms for the gain or loss of unshared sequences.

Given the success of the *de novo* assembly method applied to these data sets already, more refined assembly may be facilitated by using the four transcriptomes as references to one another. Un-assembled reads like those from the hypnozygote library and short contigs from the vegetative phase transcriptomes may be linked to longer contigs in other assemblies to reduce fragmentation of some genes. As a corollary, an aspect of this project that requires more attention is the relationship between gene model quality and quantitative estimates of gene expression. For instance, what is the consequence of high levels of fragmentation in low and high expressed contigs? If these contigs were unfragmented, would their expression level estimates increase or decrease?

For a variety of reasons, there are no obvious answers to these questions. Future assessments of the *de novo* assemblies will have to compare contig sequences to known, complete gene sequences. Still, the present assemblies are a useful basis for other molecular studies as transcriptome 'drafts', and will be a starting point for finding gene

sequences from these important organisms. A very recent example, is the discovery of orthologs for 14 cyanobacterial genes implicated in the synthesis of saxitoxin in both the toxic Group I and Group IV assemblies, and also the nontoxic Group III assembly by Hackett et al (*in prep*).

ACKNOWLEDGEMENTS

REFERENCES

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucl. Acids Res. 25(17): 3389-3402.

Anderson, D. M. and B. A. Keafer (1987). "An endogenous annual clock in the toxic marine dinoflagellate *Gonyaulax tamarensis*." Nature **325**(6105): 616-617.

Anderson, D. M., C. A. Stock, et al. (2005). "Alexandrium fundyense cyst dynamics in the Gulf of Maine." Deep-Sea Research II **52**(19-21): 2522.

Anderson, D. M., C. D. Taylor, et al. (1987). "The Effects of Darkness and Anaerobiosis on Dinoflagellate Cyst Germination." Limnology and Oceanography **32**(2): 340-351.

Bentley, D. R., S. Balasubramanian, et al. (2008). "Accurate whole human genome sequencing using reversible terminator chemistry." Nature **456**(7218): 53-59.

Binder, B. J. and D. M. Anderson (1986). "Green light-mediated photomorphogenesis in a dinoflagellate resting cyst." Nature **322**(6080): 659-661.

Birol, I., S. D. Jackman, et al. (2009). "*De novo* transcriptome assembly with ABySS." Bioinformatics **25**(21): 2872-7.

Brosnahan, M. L., D. M. Kulis, et al. (2010). "Outbreeding lethality between toxic Group I and nontoxic Group III Alexandrium tamarense spp. isolates: Predominance of heterotypic encystment and implications for mating interactions and biogeography." Deep Sea Research Part II **57**(3-4): 175-189.

Butler, J., I. MacCallum, et al. (2008). "ALLPATHS: de novo assembly of whole-genome shotgun microreads." Genome Res **18**(5): 810-20.

Chaisson, M. J. and P. A. Pevzner (2008). "Short read fragment assembly of bacterial genomes." Genome Res **18**(2): 324-30.

Erdner, D. and D. Anderson (2006). "Global transcriptional profiling of the toxic dinoflagellate *Alexandrium fundyense* using Massively Parallel Signature Sequencing." BMC Genomics **7**(1): 88.

Hackett, J. D., T. E. Scheetz, et al. (2005). "Insights into a dinoflagellate genome through expressed sequence tag analysis." BMC Genomics **6**(1): 80.

Hackett, J. D., J. H. Wisecaver, et al. (in prep). "Independent evolution of saxitoxin synthesis in cyanobacteria and dinoflagellates."

Hou, Y. and S. Lin (2009). "Distinct Gene Number-Genome Size Relationships for Eukaryotes and Non-Eukaryotes: Gene Content Estimation for Dinoflagellate Genomes." PLoS ONE **4**(9): e6978.

190

Jeck, W. R., J. A. Reinhardt, et al. (2007). "Extending assembly of short DNA sequences to handle error." Bioinformatics 23(21): 2942-2944.

Langmead, B., C. Trapnell, et al. (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol 10(3): R25.

Lilly, E. L., K. M. Halanych, et al. (2007). "Species Boundaries and Global Biogeography of the *Alexandrium tamarense* complex (Dinophyceae)." Journal of Phycology 43(6): 1329-1338.

Margulies, M., M. Egholm, et al. (2005). "Genome sequencing in microfabricated high-density picolitre reactors." Nature 437(7057): 376.

Matrai, P., B. Thompson, et al. (2005). "Circannual excystment of resting cysts of *Alexandrium* spp. from eastern Gulf of Maine populations." Deep-Sea Research II 52(19-21): 2560.

Meyers, B. C., S. S. Tej, et al. (2004). "The Use of MPSS for Whole-Genome Transcriptional Analysis in *Arabidopsis*." Genome Res. 14(8): 1641-1653.

Miller, J. R., S. Koren, et al. (2010). "Assembly algorithms for next-generation sequencing data." Genomics 95(6): 315-327.

Moustafa, A., A. N. Evans, et al. (2010). "Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence." PLoS ONE 5(3): e9688.

Quigley, M., F. Pereyra, et al. (2010). "Transcriptional analysis of HIV-specific CD8+ T cells shows that PD-1 inhibits T cell function by upregulating BATF." Nat Med 16(10): 1147-1151.

Robertson, G., J. Schein, et al. (2010). "*De novo* assembly and analysis of RNA-seq data." Nat Methods 7(11): 909-12.

Schwinghamer, P., D. M. Anderson, et al. (1991). "Separation and Concentration of Living Dinoflagellate Resting Cysts from Marine Sediments Via Density-Gradient Centrifugation." Limnology and Oceanography 36(3): 588-592.

Simpson, J. T., K. Wong, et al. (2009). "ABySS: a parallel assembler for short read sequence data." Genome Res 19(6): 1117-23.

Surget-Groba, Y. and J. I. Montoya-Burgos (2010). "Optimization of *de novo* transcriptome assembly from next-generation sequencing data." Genome Res 20(10): 1432-40.

Taroncher-Oldenburg, G., D. M. Kulis, et al. (1997). "Toxin variability during the cell cycle of the dinoflagellate *Alexandrium fundyense*." Limnology and Oceanography [Limnol. Oceanogr.] **42**(5, Pt. 2): 1178-1188.

Wang, Z., M. Gerstein, et al. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.

Warren, R. L., G. G. Sutton, et al. (2007). "Assembling millions of short DNA sequences using SSAKE." Bioinformatics **23**(4): 500-501.

Zerbino, D. R. and E. Birney (2008). "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." Genome Res **18**(5): 821-9.

Conclusions

This thesis has explored several aspects of the life cycle of the *Alexandrium tamarense* species complex and opened a new approach to the remediation of harmful algae. Pairwise co-culture experiments, cyst germinations and single cyst genotyping experiments described in **Chapter 1** were used to show that two closely related *A. tamarense* clades were indeed different species, that they could still fuse with one another to create new diploids but could not yield new genetic intermediates (or toxic cells), and that such hybridizations do occur in nature.

In proposing the introduction of any organism as a biological control mechanism, one must be sensitive to any deleterious impacts the action may have and also be realistic about its prospects for success. At least in some instances, the deleterious impacts of introducing nontoxic *A. tamarense* Group III are likely minimal. Consider the introduction of nontoxic Group III cells to embayments within the Irish and British Isles that are regularly contaminated by both toxic and nontoxic *A. tamarense* blooms by natural processes. Still, the question of collateral impacts is real and hardly understood – especially in any consideration of introductions outside the natural range of Group III nontoxic cells. What role do toxic (or nontoxic) *A. tamarense* play in the community ecology of marine ecosystems? What might be the impact to higher and lower trophic levels? Are the impacts predictable? How successful would purposeful introductions actually be in reducing the quantity of toxic cells in the future?

One reason to approach these questions less guardedly is that the introduction of nontoxic cells to areas further afield might be occurring already, especially in light of the known traffic of toxic *A. tamarense* through global shipping (Hallegraeff and Bolch 1991; Lilly, Kulis et al. 2002). Moreover, in at least one case, the negative impact of a toxic *A. tamarense* species has been shown to cause selection on animals from higher trophic levels (Bricelj, Connell et al. 2005).

Bricelj, Connell et al. (2005) showed that one sodium channel variant in the soft shell clam, *Mya arenaria* is overwhelmingly present in populations affected by toxic *A. tamarense*, and that the mutation reduces the animal's susceptibility to poisoning. It remains to be shown whether primitive *M. arenaria* populations existed that were especially sensitive to PSP toxins but the prevalence of susceptible individuals in unaffected areas certainly suggests it to be the case. The result from *M. arenaria* is especially significant if one considers the affect of new toxic *A. tamarense* introductions to areas that are unaccustomed and therefore particularly sensitive to *A. tamarense* blooms. In some instances, the potential benefits of introducing nontoxic *A. tamarense* may outweigh its risks, especially since the latter are unknown.

The second cause for caution before investing heavily in the introduction of nontoxic *A. tamarense* as a remediation strategy is that it would be difficult to predict how successful a set of deliberate introductions could be. This question was partly addressed in **Chapter 1** through the detection of naturally occurring hybrids in Belfast

Lough and an analysis of the compatibility rates between the *A. tamarense* clades. The results were mixed in the sense that, though hybrids do occur in nature, there was also an apparent preference for toxic cells to mate with other toxic cells and similarly for nontoxic cells to mate with other nontoxic cells. One question that could not be addressed in the study is the degree to which the Irish and British Isles are an area of active speciation between the toxic Group I and nontoxic Group III species. It was noted that their proximity in northern Europe and their ability to form hybrid diploids could be a significant selective force driving them toward resistance to the proposed remediation strategy. In a sense, this is just a restatement of the obvious fact that hybridization between these species is a fatal choice for any two individual cells.

Given the lethality of hybridization between Group I and Group III species, the presence of hybrids and also the apparent boundary between their native ranges suggest that hybridization may act as a barrier to species invasion. More broadly regarding the entire *A. tamarense* complex, it would be interesting to know the history of areas that have recently experienced a colonization by toxic or nontoxic *A. tamarense* populations. One example already noted is Thau Lagoon, France where an invasion by toxic Group IV cells was discovered in part because the area had an extensive log of plankton samples that showed no previous *A. tamarense* cells (Lilly, Kulis et al. 2002). It may be that the absence of other *A. tamarense* species is a prerequisite for invasion success by a new population.

In considering the impacts of new interactions between *A. tamarense* species, an

195

understanding of the basic mechanisms that govern their mating behavior is needed. Long

standing discrepancies between the conditions needed to replicate the *A. tamarense* life

cycle in culture and those observed during sexual events in nature led to two separate

projects described in **Chapter 2** and **Chapter 3** that were begun with the long term goal

of describing sexual transitions in natural systems and better understanding (and perhaps

controlling) interactions between toxic and nontoxic *A. tamarense* blooms.

In **Chapter 2**, an imaging flow cytometer was adapted to positively identify *A.*

tamarense Group I cells in mixed plankton assemblages taken from field samples and

also to measure the DNA content of these cells. While DNA content is not always an

unambiguous identifier of sexual stages, confusion between G2 phase vegetative cells

and young planozygotes (which both contain '2c' or 2 haploid units of DNA) was

avoided through the timing of sample collection. One striking result was the presence of

2c cells in all four surface samples examined. In one case, a red tide that occurred near

Portsmouth, 2c cells were almost certainly planozygotes given a collection of other data

that showed a massive cyst deposition in the area surrounding the bloom. But 2c cells

were also observed at the surface in a Salt Pond bloom that was still dividing by mitosis.

The latter result suggests either that sexual fusions in nature occur during development of

A. tamarense blooms and not just during their termination phase, or that G2 phase mitotic

cells preferentially aggregate near the surface, perhaps as part of a quorum-like sensing

behavior.

Another intriguing observation made during the flow cytometry study was the

disproportionate impact of *Amoebophyra* parasite on mature planozygote cells. Several hypotheses to explain this observation remain to be explored but the implication to species invasions (or introductions) may be quite significant, especially if parasites are either highly selective toward endemic flora as has been suggested by Chambouvet, Morin et al. (2008) or are introduced with new *A. tamarense* cells.

In the final chapter of this thesis (**Chapter 3**), I have described the first steps toward building comprehensive catalogs of the genes expressed by three different *A. tamarense* species including toxic Group I and nontoxic Group III. In addition, efforts to extract and sequence mRNAs from naturally occurring Group I hypnozygotes were described. The technology used for this effort was developed by Illumina Inc. (San Diego, CA) and yields tremendous amounts of data (as much as 10 Gb per experiment). Perhaps the most exciting advantage promised by the Illumina platform and other similar "next generation" sequencing instruments is the ability to combine gene discovery and quantitative estimates of gene expression in a single experiment. A caveat to this promise is that the sequencing reads generated by the Illumina instrument are short relative to previous generation instruments (i.e. Sanger based systems), and gene discovery relies on new assembly algorithms that are in active development at this time (Birol, Jackman et al. 2009; Robertson, Schein et al. 2010). The work described in **Chapter 3** represents an initial drafting and evaluation of the transcriptomes from *A. tamarense* Group I, Group III, and Group IV species. More work is underway to compare and characterize the differences in the genes expressed by these closely related species.

In the near future, with a wealth of gene sequence information in hand, I and

others can continue a search for molecular markers of sexual stages in the *A. tamarense*

life cycle. The sequences databases will also be an obvious starting point for new studies

to examine the evolution of the *A. tamarense* clades and develop new strategies for their

control.

REFERENCES

Birol, I., S. D. Jackman, et al. (2009). "*De novo* transcriptome assembly with ABySS."
    Bioinformatics **25**(21): 2872-7.

Bricelj, V. M., L. Connell, et al. (2005). "Sodium channel mutation leading to saxitoxin
    resistance in clams increases risk of PSP." **434**(7034): 763.

Chambouvet, A., P. Morin, et al. (2008). "Control of Toxic Marine Dinoflagellate
    Blooms by Serial Parasitic Killers." Science **322**(5905): 1254-1257.

Hallegraeff, G. M. and C. J. Bolch (1991). "Transport of toxic dinoflagellate cysts via
    ships' ballast water." Marine Pollution Bulletin **22**(1): 27-30.

Lilly, E. L., D. M. Kulis, et al. (2002). "Paralytic shellfish poisoning toxins in France
    linked to a human-introduced strain of *Alexandrium catenella* from the western
    Pacific: evidence from DNAand toxin analysis." J. Plankton Res. **24**(5): 443-452.

Robertson, G., J. Schein, et al. (2010). "*De novo* assembly and analysis of RNA-seq
    data." Nat Methods **7**(11): 909-12.