2014

BIOGEOGRAPHY AND NESTED PATTERNS OF GENETIC DIVERSITY IN THE DIATOM *THALASSIOSIRA ROTULA*

Kerry A. Whittaker
University of Rhode Island, kawhitta@gmail.com

BIOGEOGRAPHY AND NESTED PATTERNS OF GENETIC DIVERSITY IN THE DIATOM *THALASSIOSIRA ROTULA*

BY

KERRY A. WHITTAKER

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

OCEANOGRAPHY

UNIVERSITY OF RHODE ISLAND

2014

DOCTOR OF PHILOSOPHY DISSERTATION

OF

KERRY A. WHITTAKER

APPROVED:

Thesis Committee:

Major Professor Tatiana Rynearson

Susanne Menden-Deuer

Jason Kolbe

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2014

ABSTRACT

Diatoms are some of the most biogeochemically important organisms on our planet, responsible for up to 40% of primary production in the ocean. Diatoms are also the most diverse group of algae, exhibiting high levels of cryptic speciation, and remarkable clonal diversity nested within species and populations. As single cells within a dynamic fluid environment, diatoms possess the potential to disperse throughout the globe. Yet, despite the global dispersal potential of these ecologically important organisms, little is understood about the structure and distribution of diatom diversity throughout the globe, or its ecological significance. The goal of this work was to explore nested patterns of diversity in the diatom *Thalasiosira rotula* by mapping its distribution over global geographic space, capturing the variations in genetic composition over time, and inferring its ecological significance.

A series of molecular markers was used to target different scales of genetic subdivision from the cryptic species to the population to the individual. Ribosomal genes (rDNA) were used to define lineages nested within the species, and clarify the relationship between *T. rotula* and its cryptic sister species, *Thalassiosira gravida.* Microsatellite markers were developed for *T. rotula*, and offered a high-resolution glimpse into the population-level subdivisions within the species. Microsatellite analysis was robust enough to identify individuals within these populations.

Ribosomal DNA analysis confirmed that *T. rotula* and *T. gravida* are, indeed, separate species. These genes also demonstrated that *T. rotula* is subdivided into three lineages diverged by 0.6 ± 0.3% at the internal transcribed spacer gene. Each lineage exhibits a unique geographic distribution, and harbors high levels of physiological

diversity. Over 400 global samples of *T. rotula* collected from 8 locations in 2010 offered a 'snapshot' of diversity nested within the species. Microsatellites revealed high levels of population divergence among these global samples. Variation over time at a single site was as great as variation between samples collected thousands of kilometers apart. Within a single site, Narragansett Bay, changes in population structure over the course of three years demonstrated that populations are comprised of many clonal lineages, harbor high adaptive potential, and may respond quickly to changes in the marine environment and local ecology.

The structure of diversity among global samples was used to infer the importance of dispersal versus environment in isolating populations and maintaining diversity. At all levels of molecular diversity, geographic distance demonstrated only a weak relationship to genetic distance. This suggests that global connectivity among populations is not limited by geographic distance. Nonetheless, *T. rotula* does not represent a single panmictic population. Significant genetic structure was detected on all levels of molecular diversity examined. A series of Mantel tests revealed that temperature and chlorophyll *a* were most highly correlated with genetic relatedness among global populations, despite our exploration of other variables including nutrient concentration, temperature, irradiance, salinity, and cell abundance. This suggests that certain populations may be better adapted to thrive in competitive high-chorophyll phytoplankton 'bloom' periods, whereas others may only thrive during non-bloom periods. Taken together, these data suggest that environmental and ecological selection may heavily influence the genetic structure of diatom populations, and help to explain the extraordinary diversity harbored within and among diatom species.

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Dr. Tatiana Rynearson, who has shared with me her wisdom, insight, and passion for knowledge throughout this process. She has provided continual support for my growth as a student, a scientist, and a person. She has set an example through her integrity, and by holding high expectations of herself and others. I have been privileged to have such a dedicated advisor, who has helped shape me as a scientist, and as a person humbly seeking a deeper understanding of our world.

Thank you to my committee members Tatiana Rynearson, Susanne Menden-Deuer, and Jason Kolbe, and Lucie Maranda. Your guidance has helped to improve this thesis, encouraging me to sharpen and distill the interpretation of data, and to tell the best story possible.

I am grateful to many funding sources. A National Science Foundation grant awarded to Dr. Rynearson has funded a large portion of this work. I also thank EPSCoR for awarding me a Graduate Student Fellowship, and scholarships from the University of Rhode Island Alumni Fund and Graduate Student Association.

Thank you to current and past members of the Rynearson Lab including Melissa Mercier, Chris Piecuch, Richard Kingsley, Alison Cleary, Sarah Flickinger, and Kelly Canesi for your friendship and your insight. To many undergraduates with whom I have worked along the way, thank you for your dedication and willingness to learn. You have taught me how to be a better mentor, a better teacher, and a better student.

I have relied upon the help of scientists around the world, who have generously offered their time, collecting and sending samples from Tasmania, Russia, Martha's

Vineyard, South Africa and beyond. Thank you for your spirit of academic service and collaboration. I look forward to passing it on. Thank you to the many scientists, colleagues, professors, fellow students, and crewmembers that I have met along the way—far too many to mention here. You inspire me, and you make me look forward to future collaborations, conferences, cruises, and fieldwork, wherever my career leads me from here.

It is with deep appreciation that I thank my immediate family, Mom, Dad, and Nate, for always believing in me, and for showing me unconditional love, understanding, patience, and support through all of my academic endeavors. Thank you to the rest of my family who, with compassion, always show an interest in what I do. A special remembrance to my Grandma Whit, who approached life with a contagious sense of humor and adventure. Thank you to Max; you share with me your love, and have kept me grounded in the moments of stress and excitement that come with this process. Thank you to my wonderful friends who bring endless music, laughter, and joy, sparking the creative part of myself that has allowed me to complete this work. You all remind me of the important things in life, and inspire me to keep learning every day.

PREFACE

This dissertation is presented in manuscript format, and is subdivided into three chapters. Chapter one is an introduction, providing context for this work and how it fits into our broader understanding of diversity and evolution in the ocean. Chapter two focuses on the biogeography of the diatom *Thalassiosira rotula*, with insight into molecular, physiological, and genome size variations among global isolates. This is a published manuscript. Chapter three explores a global snapshot of population structure in *T. rotula*, and its ecological correlates. This manuscript is in preparation for submission to Limnology and Oceanography. Chapter four addresses changes in genetic composition over time, and the succession of populations over a three year period using isolates of *T. rotula* collected from Narragansett Bay, RI and off the coast of Martha's Vineyard, MA. This manuscript is in preparation for submission to the Journal of Phycology. Chapter five is a speculative discussion on the dissertation as a whole, reflecting on the outcome of all three chapters. The appendex following these chapters includes additional data upon which figures throughout the dissertation are based.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

By

Kerry A. Whittaker

Graduate School of Oceanography, South Ferry Road, University of Rhode Island, Narragansett, RI, 02882

INTRODCUTION

With an estimated 100,000 species hypothesized to exist, diatoms are the most diverse group of algae on the planet (Mann and Vanormelingen 2013) . Molecular clock estimates place the origin of diatoms at 250 mya (Sorhannus 2007). As free-floating, single-celled organisms, it is argued that diatoms experience high dispersal throughout the global ocean, and therefore high levels of genetic connectivity, which may limit their potential for genetic isolation (Fenchel and Bland 2004; Finlay 2002). Despite this paradigm, amassing evidence demonstrates that individual diatom species exhibit genetic structure at local (Gallagher 1980; Rynearson 2004; Rynearson and Armbrust 2005), regional (Evans 2005), and global scales (Casteleyn 2010). The mechanisms driving genetic isolation and structuring genetic diversity within diatom species are little understood. By exploring the extent of genetic diversity and connectivity among populations of a single globally distributed diatom species, this study aims to investigate the ecological factors correlated with diatom diversity structure and genetic connectivity among diatom populations. Doing so broadens our understanding of the evolutionary processes have led to the vast diversity of diatoms on our planet.

We can contemplate the uniqueness of diversity structure in the ocean by comparing distribution and life history patterns of organisms between terrestrial and marine environments. Primary production is equally distributed between the land and the ocean (Field 1998), yet the patterns of evolution and diversity structure over space may be driven by very different mechanisms. For one, terrestrial organisms are dominated by small population sizes and limited dispersal

(Carr 2003). Populations are distributed over two spatial dimensions, and organisms are associated with longer generation times (Steele 1985). In contrast, marine primary producers are phytoplankton, generally smaller than 200μm, and constantly drifting with the tides and currents. This leads to a high potential for dispersal among these organisms (Finlay 2002; Kinlan 2003). Phytoplankton are also associated with massive effective population sizes (Fenchel and Bland 2004; Hochberg 1988), and can form blooms that cover vast areas of the ocean (Sverdrup 1953). Their distributions occupy three spatial dimensions, and they can rapidly evolve through fast generation times, dividing on timescales of once per day. Because of these characteristics, the evolution and diversity structure of phytoplankton may be driven by very different mechanisms than primary producers on land.

For the most part, our understanding of diatom diversity extends only to local and regional scales. The species studied to date demonstrate some common characteristics. First, it has been shown that 'super species,' or those previously thought to occur throughout the globe, can be comprised of many different species that are genetically distinct and reproductively isolated, despite the fact that they appear identical under a microscope (Amato 2007; Fenchel 2005; Koester et al. 2010; Whittaker et al. 2012). Next, all diatom species studied to date exhibit high levels of intraspecific diversity and large population sizes, with clonal diversity greater than 90% (Collins 2013). For example, a single bloom of the diatom *Ditylum brightwellii* was comprised of thousands of clonal lineages (Rynearson and Armbrust 2005). Next, populations demonstrate stability over time. For

instance, in fjords of Sweden, genetic signals from resting spores of the diatom *Skeletonema marinoi* demonstrate that a single population can persist in a local environment for over one hundred years (Härnström et al. 2011). Finally, diatom populations can be highly diverged among environments that exhibit high physical connectivity, and close local proximity. For instance, the diatom *D. brightwellii* exhibits significantly diverged populations in Puget Sound, and the neighboring Strait of Juan de Fuca (Rynearson 2004).

Our understanding of diatom diversity on global spatial scales is very limited. In fact, only one study has explored diversity within a diatom species on a global scale—this study focused on the species *Pseudo-nitzschia pungens* (Casteleyn 2010). This species is comprised of three separate lineages that are associated with unique geographic distributions. The authors used high-resolution genetic markers to explore diversity within one of these lineages, and found that genetic distance exhibited a strong relationship to geographic distance. They concluded that, for this species, dispersal is limited, and geographic distance is a strong barrier to genetic connectivity among populations. While the link between geographic distance and genetic isolation was apparent in this species, the underlying ecological mechanisms driving the isolation-by-distance (IBD) pattern remain unclear. Also, whether or not this pattern of IBD is a species-specific phenomenon, or common across diatom species is not understood.

In the global study of *Pseudo-nitzschia pungens* population structure, samples were collected over ten years and pooled over seasons and space to represent regions (Casteleyn 2010). Sampling of diatoms is often a challenge,

especially when addressing global spatial scales. Pooling samples over seasons and years is often unavoidable, but prevents an understanding of the intersection between genetic diversity and environmental diversity of the dynamic ocean, which varies widely over space and time. Because of this, we have little understanding of the relative contribution of global dispersal versus environmental selection in structuring populations and influencing the evolution of diatoms.

This dissertation broadens our understanding of diatom diversity structure on global spatial scales through the lens of the ecologically important diatom species, *Thalassiosira rotula*. *T.rotula* thrives in ecosystems throughout the globe, found in every ocean basin, and across hemispheres (Krawiec 1982). Throughout this dissertation, I explore nested patterns of diversity within samples of *T. rotula* collected at global spatial scales. I employ a sampling scheme that avoids pooling samples over time and space. Analyzing each discrete sample, each with its own environmental metadata, allows me to explore the ecological correlates of global diversity structure in this high-dispersal and globally distributed species. Doing so allows me to tease apart the key environmental correlates of diversity structure, expanding our understanding of the ecological mechanisms that influence the isolation and connectivity among diatom populations.

This dissertation explores nested patterns of diversity and their ecological correlates from several different angles. Chapter 1 addresses the following questions: 1) Does the species *Thalassiosira rotula* exhibit genetic structure on the lineage level? 2) What is the relationship between differences in genetic diversity, physiological diversity, and genome size within the species? Chapter 2 uses high-

resolution molecular markers to measure the extent of diversity nested within these lineages. Chapter two explores the following questions: 1) How are populations structured and related among global samples collected in 2010 throughout the winter-spring bloom periods in the Northern and Southern hemispheres? 2) What are the environmental correlates of diversity structure among samples? Chapter 3 addresses changes in population structure over time. The structure of diatom populations and its variation over time may play an important role in determining diatom bloom formation and a species' adaptive potential. Chapter 3 addresses the following questions: 1) How do *T. rotula* populations within Narragansett Bay, RI and off the coast of Martha's Vineyard, MA vary over a three year sampling period? 2) What are the potential environmental triggers of population shifts within the species over time? These three chapters work synergistically to approach a deeper understanding of the relative roles of high dispersal and environmental variation in structuring the diversity of single-celled organisms in the ocean. For the first time, global patterns of diversity within a diatom species are coupled with patterns of environmental variation, allowing me to investigate the relative importance of dispersal versus environmental selection in structuring populations of marine diatoms throughout the globe and over time.

Amato, A. 2007. Reproductive isolation among sympatric cryptic species in marine diatoms. Protist **158:** 193.

Carr, M. H. 2003. Comparing marine and terrestiral ecosystems: implications for the design of coastal marine reserves Ecological applications **13:** 90-107.

Casteleyn, G. 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. PNAS : Proceedings of the National Academy of Sciences **107:** 12952-12957.

Collins, S., Rost, B., Rynearson, T.A. 2013. Evolutionary potential of marine phytoplankton under ocean acidification. Evolutionary Applications **7:** 140-155.

Evans, K. M. 2005. High levels of genetic diversity and low levels of genetic differentiation in North Sea Pseudo-nitzchia pungens (Bacillariophyceae) populations. . Journal of phycology **41:** 506-514.

Fenchel, T. 2005. Cosmopolitan microbes and their'cryptic'species. Aquatic microbial ecology : international journal **41:** 49.

Fenchel, T., and J. F. Bland. 2004. The Ubiquity of Small Species: Patterns of Local and Global Diversity. BioScience **54:** 777-784.

Field, C. B. 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science (New York, N.Y.) **281:** 237-240.

Finlay, B. J. 2002. Global Dispersal of Free-Living Microbial Eukaryote Species, p. 1061-1063.

Gallagher, J. C. 1980. Population genetics of *Skeletonema costatum* (Bacillariophyceae) in Narragansett Bay. Journal of phycology **16:** 464-474.

Härnström, K., M. Ellegaard, T. R. J. Andersen, and A. Godhe. 2011. Hundred years of genetic structure in a sediment revived diatom population. Proceedings of the National Academy of Sciences **108:** 4252-4257.

Hochberg, Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika **75:** 800-802.

Kinlan, B. P. 2003. Propagule dispersal in marine and terrestrial environments: a community perspective Ecology (Durham) **84:** 2007-2020.

Koester, J. A., J. E. Swalwell, P. Von Dassow, and E. V. Armbrust. 2010. Genome size differentiates co-occurring populations of the planktonic diatom Ditylum brightwellii (Bacillariophyta). BMC Evol Biol **10:** 1.

Krawiec, R. W. 1982. Autecology and clonal variability of the marine centric diatom Thalassiosira rotula; (Bacillariophyceae) in response to light, temperature and salinity. Marine Biology **69:** 79-89.

Mann, D. G., and P. Vanormelingen. 2013. An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species. Journal of Eukaryotic Microbiology **60:** 414-420.

Rynearson, T. A. 2004. Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). Journal of phycology **40:** 34-43.

Rynearson, T. A., and E. V. Armbrust. 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom Ditylum brightwellii. Mol Ecol **14:** 1631-1640.

Sorhannus, U. 2007. A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. Marine Micropaleontology **65:** 1-12.

Steele, J. H. 1985. A comparison of terrestiral and marine ecological systems. Nature **313:** 355-358.

Sverdrup, H. U. 1953. On Conditions for the Vernal Blooming of Phytoplankton. Journal du Conseil **18:** 287-295.

Whittaker, K., D. Rignanese, R. Olson, and T. Rynearson. 2012. Molecular subdivision of the marine diatom Thalassiosira rotula in relation to geographic distribution, genome size, and physiology. BMC Evolutionary Biology **12:** 209.

CHAPTER 2

MOLECULAR SUBDIVISION OF THE MARINE DIATOM *THALASSIOSIRA ROTULA* IN RELATION TO DIFFERENCES IN GEOGRAPHIC DISTRIBUTION, GENOME SIZE, AND PHYSIOLOGY

By

KA Whittaker[1], DR Rignanese[1], RJ Olson[2], TA Rynearson[1*]

1. Graduate School of Oceanography, South Ferry Road, University of Rhode Island, Narragansett, RI, 02882
2. Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543

* Correspondence: rynearson@gso.uri.edu

Is published.
Proper citation:

Whittaker K, Rignanese D, Olson R, Rynearson T: **Molecular subdivision of the marine diatom Thalassiosira rotula in relation to geographic distribution, genome size, and physiology**. *BMC Evolutionary Biology* 2012, **12**(1):209.

ABSTRACT

Background

Marine phytoplankton drift passively with currents, have high dispersal potentials and can be comprised of morphologically cryptic species. To examine molecular subdivision in the marine diatom *Thalassiosira rotula,* variations in rDNA sequence, genome size, and growth rate were examined among isolates collected from Atlantic and Pacific Ocean basins. Analyses of rDNA included *T. gravida* because morphological studies have argued that *T. rotula* and *T. gravida* are conspecific.

Results

Culture collection isolates of *T. gravida* and *T. rotula* diverged by 7.0 ± 0.3% at the ITS1 and by 0.8 ± 0.03% at the 28S. Within *T. rotula*, field and culture collection isolates were subdivided into three lineages that diverged by 0.6 ± 0.3% at the ITS1 and 0% at the 28S. The predicted ITS1 secondary structure revealed no compensatory base pair changes among lineages. Differences in genome size were observed among isolates, but were not correlated with ITS1 lineages. Maximum acclimated growth rates of isolates revealed genotype-by-environment effects, but these were also not correlated with ITS1 lineages. In contrast, intra-individual variation in the multi-copy ITS1 revealed no evidence of recombination amongst lineages, and molecular clock estimates indicated that lineages diverged 0.68 Mya. The three lineages exhibited different geographic distributions and, with one exception, each field sample was dominated by a single lineage.

Conclusions

The degree of inter- and intra-specific divergence between *T. gravida* and *T. rotula* suggest they should continue to be treated as separate species. The phylogenetic distinction of the three closely-related *T. rotula* lineages was unclear. On the one hand, the lineages showed no physiological differences, no consistent genome size differences and no significant changes in the ITS1 secondary structure, suggesting there are no barriers to interbreeding among lineages. In contrast, analysis of intra-individual variation in the multicopy ITS1 as well as molecular clock estimates of divergence suggest these lineages have not interbred for significant periods of time. Given the current data, these lineages should be considered a single species. Furthermore, these *T. rotula* lineages may be ecologically relevant, given their differential abundance over large spatial scales.

INTRODUCTION

Photosynthetic organisms in terrestrial and marine habitats contribute to global primary production in roughly equal proportions (Field 1998; Laisk et al. 2009). For most terrestrial photoautotrophs, species distributions occupy two spatial dimensions and vary over relatively long time scales (Bolker 1999; Burger 1981; Sanmartan 2004; Simpson 1974). In marine habitats, most primary producers are unicellular phytoplankton, smaller than 200μm that drift with tides and currents. Marine phytoplankton differ from their terrestrial counterparts in that these tiny organisms have higher dispersal potentials, the ability to occupy three spatial dimensions, and species distributions that can vary over time scales of days to weeks (Carr et al. 2003; Kinlan 2003; Simon 2009).

Within the phytoplankton, diatoms are a particularly important class of algae.

These commonly-occurring organisms generate over 20% of global primary production, and thus play a key role in driving global biogeochemical cycles (Nelson et al. 1995). They are found in almost all aquatic habitats, are comprised of an estimated 200,000 species (Mann 1996) and yet only arose in the early Mesozoic (~ 250 mya) (Sorhannus 2007). Diatom species have long been identified based on their ornate siliceous frustule, or cell covering, and recently, based on DNA sequence variation. Morphological differences in the frustule are often subtle and with the advent of scanning electron microscopy, additional morphological features were visible and many new species were described (Round 1990). More recently, ribosomal DNA sequence variation has been used to identify morphologically identical, or cryptic, species (Aloisie Poulickova 2010; Alverson 2008; Amato et al. 2007; Ellegaard et al. 2008; Sarno 2005; Sarno et al. 2007; Zingone et al. 2005). Thus far, no clear barcoding gene, such as Cox1, has been identified for classifying diatom species (Evans et al. 2007; Moniz and Kaczmarska 2009).

Testing the biological species concept using diatoms poses a significant challenge because it has proven difficult to consistently control sexual reproduction in the laboratory. For those species where sexual reproduction can be controlled, there appears to be a relationship between reproductive incompatibility and compensating base changes in the stem regions of ITS2 secondary structures (Amato et al. 2007; Behnke et al. 2004; Coleman 2000; Coleman 2003). Although this same relationship has not been explicitly observed for the ITS1, it has been argued that the two genes do not evolve independently of one another, and thus their secondary structures are similarly conserved (Coleman and Vacquier 2002). In laboratory experiments with

the diatom genus *Sellaphora*, interbreeding was successful between individuals with 7.3% divergence at the ITS1 and ITS2, but not between those with 10% divergence (Behnke et al. 2004). Similarly, reproductive isolation has been observed between species of the genus *Pseudo-nitzschia* that diverged by 2.4 % at the ITS1 and ITS2. Morphologically cryptic diatom species have also been identified by 0.5% sequence divergence at the 28S rDNA (Ellegaard 2008; Sarno 2005; Sarno et al. 2007; Zingone et al. 2005). Furthermore, diatom lineages have been shown to exhibit differences in genome size (Koester et al. 2010; Von Dassow 2008), suggesting that polyploidization may play a role in driving cryptic diatom speciation, as is commonly observed in plants (Wood 2009)

The identification of morphologically cryptic species has led to the question of whether cosmopolitan species are truly globally distributed or whether these morphospecies are instead divided into multiple species with distinct biogeographic ranges. For example, the diatom *Skeletonema costatum* was once thought to be a "super" species based on its ability to thrive and even dominate phytoplankton communities in an exceptionally broad range of environments (Smayda 1958). It was recently shown to consist of several different species (Ellegaard et al. 2008; Kooistra 2008; Sarno 2005; Sarno et al. 2007; Zingone et al. 2005) that may each have unique geographic distributions (Kooistra 2008). Similarly, geographic differentiation has been shown for the harmful algal bloom-forming genus *Pseudo-nitzschia* (Casteleyn 2008; Hubbard 2008). Aside from these few examples, it is unknown whether most ecologically important diatoms with cosmopolitan distributions are true species or whether they are comprised of multiple cryptic species. Furthermore, understanding

the genetic relationships within ecologically important species or species complexes from geographically disparate regions has become an important issue in light of climate change and the novel selection pressures that may result (Behrenfeld et al. 2006; Cermeno et al. 2010 ; Hallegraeff 2010).

It has been challenging to describe general patterns of species division in diatoms because previous studies used different methods, focused on different species, and often sampled few isolates from a restricted spatial scale. We focused on identifying genetic subdivision in the diatom morphospecies *Thalassiosira rotula* by simultaneously examining variation in rDNA sequences, physiology, and genome size from isolates collected from around the globe. *Thalassiosira rotula* is a commonly-occurring diatom that can dominate phytoplankton assemblages across diverse marine habitats and hydrographic environments (eg. (Brockmann et al. 1977; Bursa 1961; Cassie 1960; Gran 1930; Krawiec 1982; Matsudaira 1964; Pratt 1959; Rytter Hasle 1976; Smayda 1957; Smayda 1958)). Here, cells were collected along a transect in the Eastern North Pacific and their rDNA (18S, ITS1, 28S) compared with isolates collected from the Pacific, the Atlantic, and Mediterranean Sea to determine the geographic distribution of rDNA sequence variants. Growth rates among isolates were used to determine the relationship between molecular and physiological diversity. Variation in genome size among isolates was measured as recent work indicated that differences in DNA content may identify cryptic species (Koester et al. 2010; Von Dassow 2008). Morphological studies suggested that *T. rotula* (Meunier 1910) and *T. gravida* (Cleve 1896) are likely a single species (Sar et al. 2011; Syvertsen 1977). To test this hypothesis, we determined rDNA sequence variation among culture collection

isolates of *T. rotula* and *T. gravida*. By examining isolates collected from around the world, we were able to identify sufficient genetic differences between *T. rotula* and *T. gravida* to warrant their continued description as distinct species and to identify genetic subdivision within *T. rotula* and its correspondence to geographic location, physiological variation and differences in genome size.

METHODS

Isolates

Cells of the diatom morphospecies *Thalassiosira rotula/ gravida* were collected from 8 locations in the Eastern and Western Pacific and Western Atlantic between 2007 and 2009 (Table 1). Culture collection isolates of *T. rotula* and *T gravida* from an additional 7 locations were also obtained (Table 1). For all field samples (sites 1-4, 7, 8, and 13), surface water was passed through a 20μm mesh net. At least 17 single cells or short chains were isolated from the >20μm size fraction using a stereomicroscope (Olympus SZ61) rinsed in sterile seawater three times, and transferred to 1 mL sterile Sargasso seawater amended with f/20 nutrients (Guillard 1975). Cells were incubated at 8ºC (site 13) or 14ºC (all other sites) and on a 12:12-h light:dark cycle at 90 μmol photons m^{-2} s^{-1} for approximately two weeks. Live cells were then examined using microscopy to record growth and detect contamination. When cell density reached ~1000 cells/ml, 20μl of each isolate were transferred to fresh f/20 media to maintain growth. Remaining cells were harvested onto a 1.2 μm filter and frozen at -80ºC until further analysis. Culture collection isolates from NCMA (National Center for Marine Algae and Microbiota, formerly CCMP), CCAP (Culture Collection of Algae and Protozoa), and Japan were grown at 4ºC (sites 14,

15) or 14ºC (all others) and maintained in exponential growth; cells were filtered and frozen as described above. Genomic DNA of both field and culture collection isolates was extracted using the DNeasy Plant Mini Kit or the DNeasy 96 Plant Kit (Qiagen, Valencia, CA, USA).

Ribosomal DNA sequencing and analysis

To quantify genetic variation among isolates, three regions of the ribosomal DNA (rDNA) were sequenced: the small subunit (18S), the D1 hypervariable region of the large subunit (28S), and the internal transcribed spacer region I (ITS1). To amplify each rDNA region, a reaction mixture containing ~5ng DNA, 0.1 mmol L^{-1} dNTPs (Bioline), 0.05 U μL^{-1} Accuzyme DNA polymerase (Bioline), 1X buffer (Bioline), and 0.5 μmol L^{-1} each of forward and reverse primers was used. Using polymerase chain reaction (PCR), the ITS1 was amplified using a newly-designed primer specific to *T. rotula* and *T. gravida*: RotIIR GTCACAGTCCAGCTCGCCACCAG and primer 1645F (Rynearson 2006). The PCR consisted of a 2 min denaturation step at 95ºC, 36 cycles of 94ºC for 30 s, 62ºC for 30 s, and 72ºC for 1 min followed by a 10 min extension at 72ºC. The ITS1 was amplified from 106 isolates, including 10 from each field sample. The 18S was amplified from 16 of those isolates using universal 18SA and 18SB primers (Medlin et al. 1988). The PCR consisted of a 2 min denaturation step at 95ºC, 33 cycles of 94ºC for 20 s, 55ºC for 60 s, and 72ºC for 2 min and 10 min at 72ºC. The D1 region of the 28S was also amplified from 16 isolates using the forward primer, 28SF: ACCCGCTGAATTTAAGCATA, and reverse primer, 28SR: ACGAACGATTTGCACGTCAG (Auwera and Wachter 1998), at the same

thermocycling conditions as the 18S PCR. All sequencing was performed on an ABI 3130xl (Applied Biosystems). Both strands of the ITS1 were sequenced to completion using primers 1645F and RotIIR. For 18S rDNA genes, both strands were sequenced to completion using primers 18SA and 18SB (Medlin et al. 1988), 18SC2, 18SE2, and 18SF3 (Rynearson 2004), and 18SD (Armbrust and Galindo 2001a). 28S was sequenced using 28SR and 28SF, respectively. All sequences are available in Genbank (accession numbers JX069320-JX069349 and JX074825-074930)

Sequences were assembled using SeqMan II 3.61 (DNASTAR inc), and aligned using Clustal W (Thompson 1994) in Mega4 (Tamura et al. 2007); boundaries of the ITS1 were determined through alignment with Genbank accession EF208798. For ITS1, 18S, and 28S, sequence divergence was calculated using a two-parameter model of nucleotide substitution using Mega4 (Kimura 1980; Tamura et al. 2007). Significant differences in rDNA sequence variation between *T. rotula* and *T. gravida* were determined using analysis of molecular variance (AMOVA) with 1000 permutations. To examine relationships among *T. rotula* ITS1 variants, a network was generated using the median joining algorithm in Network 4.5.1.6 (Fluxus Technology Ltd.) and tested for statistical significance using AMOVA with 1000 permutations. Significant differences in within-lineage diversity were determined using the population differentiation algorithm (Raymond 1995) run with 1000 steps of the Markov Chain, 1000 dememorization steps, and an alpha of 0.05. All AMOVA and population differentiation analyses were conducted using Arlequin v. 3.11 (Excoffier 2005). The RNAalifold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi) was used to construct the predicted secondary structure of the consensus ITS1

sequence from the Clustal W alignment of all isolates, using the minimum free energy and partition function, avoiding isolated base pairs (Mathews 1999).

A relaxed molecular clock with an uncorrelated lognormal model was used to generate divergence times of *Thalassiosira* species (Table 2) and *T. rotula* lineages in BEAST v1.4.6. The most frequently occurring sequence type in each *T. rotula* lineage was chosen for analysis. Trees were generated using the last 5,000 generations of a Markov chain-Monte Carlo analysis run for 1 x 10^6 generations. To obtain molecular clock estimates, we used a fossil-calibrated, 18S divergence time of 11Mya for *T. aestevalis* and *T. anguste-lineata* (Sorhannus 2007). This divergence time was forced on the *T. aestevalis* and *T. anguste-lineata* ITS1 divergence to obtain an estimated ITS1 mutation rate for the remaining Thalassiosirids compared here.

To examine intra-individual variation in the multi copy ITS1, PCR amplicons from three individuals were ligated into the pCR2.1 vector, and chemically transformed into TOP10 one-shot competent *Escherichia coli* cells using the TOPO TA Cloning Kit (Invitrogen). ITS1 inserts were amplified from cloned *E. coli* using the Illustra Templiphi Kit (GE) and 20 inserts from each isolate were sequenced as described above. Intra-individual ITS1 variants were added to the network analysis generated by Network 4.5.1.6 (Fluxus Technology Ltd.). Difference of Sums of Squares and probabilistic divergence measures in Topali v.2 were used to measure putative recombination break points among intra-individual ITS1 sequence types using a step size of 5, window size of 100, and 500 threshold runs. Additional tests for recombination were conducted in RDP3, using a UPGMA-based pairwise scanning approach (Martin et al. 2010).

Physiological Variation

Isolates of *T. rotula* collected from the Mediterranean Sea (CCMP1647, site 11), from the coast of Vancouver Island (VIA, site 2), and from the Seto Inland Sea (SIS, site 6) were made axenic to remove contaminating bacteria. Sterile glass tubes were prepared with 4.5 mL sterile f/2, 10 µL sterile bacterial test medium (Bacto TM) (5 g L^{-1} Bacto-peptone and 5 g/L malt extract), ~ 1000 cells of each isolate and different volumes of sterile antibiotic mix (50 - 400 µL) containing 0.1 g L^{-1} Penicillin G (Potassium salt), 0.0025 g L^{-1} dihydrostreptomycain sulfate, and 0.005 g L^{-1} gentamyacin. At 72, 96 and 120 hours, 5-50 µL of culture were subcultured into 4 mL of sterile f/2 media at 14ºC with a 12:12 light:dark cycle and routinely tested for bacterial contamination using Bacto TM.

The resulting three axenic isolates (CCMP1647, VIA, and SIS) as well as three additional xenic isolates (CCMP3264, VIB, and CCAP1085_21, sites 12, 3, and 10 respectively) were incubated at three temperatures (4, 10, and 17.5ºC) on a 12:12 h light:dark cycle at two light conditions (50 µmol and 112 µmol photons m^{-2} s^{-1}). Cultures were allowed to acclimate until no differences in maximum growth rate were observed between transfers. Maximum acclimated growth rates were determined following Rynearson and Armbrust (Rynearson and Armbrust 2000) and Brand (Brand et al. 1981). Briefly, the in vivo fluorescence of semi-continuous batch cultures was measured daily using a 10AU Field Fluorometer equipped with the in vivo chlorophyll optical kit (Turner). The maximum acclimated growth rate for each isolate was determined by regressing the change in the log of fluorescence over time and testing the equality of slopes from at least three serial cultures ($\alpha = 0.05$). If

slopes of serial growth curves were homogenous, the average regression coefficient was used to estimate the common slope, which represented the average acclimated growth rate. To test for differences in growth rate between axenic and xenic strains, triplicate cultures of both axenic and xenic VIA were grown at 10°C, 100 µmol photons m^{-2} s^{-1}. Analysis of variance (ANOVA), nested ANOVA, Sheffe's Test (Enderlein 1961a), and the Tukey multiple comparison test (Tukey 1949) were used to determine the significance of differences among isolates at different temperatures and light levels, and between axenic and xenic growth rates. For each isolate, the diameter of 30 cells was measured using an E800 microscope at 20X (Nikon). One-way ANOVA was used to determine the significance of differences between cell size across all temperature and light conditions. Alpha was set to 0.05 for all statistical tests.

Relative genome size

Five isolates were chosen to examine relative differences in genome size using flow cytometry: VIA (site 2), VIB (site 3), SIS (site 6), CCAP1085_21 (site 10), and CCMP3264 (site 12). To ensure that cells in both G1 and G2 phases were present in each sample, cells were grown in continuous light (20°C), 60 µmol photons m^{-2} s^{-1}. For each isolate, triplicate cultures were harvested at mid-exponential phase by centrifugation (15 min at 112 rcf). Cells were injected into 10 ml of 100% methanol at 0°C using a 1 in, 22 gauge syringe needle, incubated at 4°C for ≥ 1 h to extract chlorophyll *a*, centrifuged three times for 15 min at 112 rcf, washed each time with 5 ml PBS (137 mM NaCl, 2.7 mM KCl, 10.4 mM Na2HPO4·H2O, 1.8 mM KH2PO4, pH = 7.4) and resuspended in 3 ml PBS at >1000 cells/ml. Cellular DNA was stained

with DAPI for ≥ 20 min. A 50 μL suspension of fluorescent latex beads (UV-excited, 6 μm diameter; Invitrogen) were added to each sample as an internal standard.

Flow cytometry was conducted using a benchtop version of the Imaging FlowCytobot (IFCB) (Olson and Sosik 2007) with a 375nm, 16 mW UV laser (CUBE™375-16C, Coherent, Inc.) and a beam spot approximately 10 μm high. Dichroic and bandpass filters directed blue light (DAPI fluorescence, 425-475 nm) to the fluorescence PMT (HC 120-05M, 10 MHz preamp, Hamamatsu used for triggering) and UV light (<400 nm) to the light scattering PMT. Sheath flow (0.2 μm filtered distilled water) was gravity-driven through the laser beam at a velocity of ~ 2 ml min^{-1}. A digitizer (model AD2100-14, Chase Scientific), operating at 6 samples per microsec, recorded complete traces of the PMT response for each cell. Cell debris and clumps were removed from the dataset using images captured with each event and custom Matlab scripts, leaving only beads or well-defined *T. rotula* cells. Next, distributions of integrated fluorescence of triplicate *T. rotula* cultures for each isolate were combined and the cell data normalized to internal standard bead fluorescence. Mode values for the G1 peaks were determined from the bead-normalized distributions after smoothing the data (fastsmooth.m, width 10; T. C. O'Haver, pers. comm).

To investigate relationships between cell size and genome content, cell volume of each isolate analyzed using flow cytometry was determined by measuring cell diameter and length for 30 cells $isolate^{-1}$ using an E800 microscope at 20X (Nikon). The significance of differences in average cell volume among strains was determined using ANOVA.

RESULTS

rDNA variation

Culture collection isolates of *T. gravida* and *T. rotula* were not significantly different ($p>0.05$) at the 18S (0.1% divergence), but diverged significantly ($p <0.05$) from each other at the ITS1 (7 ± 0.3%) and 28S (0.8 ± 0.03%). Of the 97 field isolates analyzed, 10 had 28S and ITS1 sequences that were identical (100%) to *T. gravida* culture collection isolates identified by taxonomists (Table 1b) and thus were designated as *T. gravida*. rDNA sequences of the remaining field isolates were identical (100%) at the 28S and 99-100% similar at the ITS1 to *T. rotula* culture collection isolates identified by taxonomists (Table 1a) and thus were designated as *T. rotula.*

Of 92 global *T. rotula* isolates, 23 unique ITS1 sequences were detected with an average sequence divergence of 0.6 ± 0.3 %. Twenty sequences were relatively rare, identified in fewer than 4 isolates. Of these, sixteen sequences were identified just once, three sequences were identified twice, and one sequence was identified in four isolates. Three sequences (Table 3, sequences 1-3) were relatively abundant (identified in >10 isolates). The median joining network of all sequences was significant (AMOVA $p<0.001$), and was used to define three distinct lineages corresponding to the three most abundant sequences, and closely branching but rare sequences (Table 3, Figure 1). Lineage 1 was comprised of six sequences and was dominated by sequence 1, identified in 17 isolates (Table 3). All lineage 1 isolates originated from the coastal North Pacific (Figure 2, sites 1-3 and 5). Lineage 2 was comprised of three sequences and was dominated by sequence 2, identified in 11

isolates (Table 3). Lineage 2 isolates were sampled from Puget Sound (Figure 2, site 4) with the exception of six isolates sampled from coastal North Pacific waters (Figure 2, site 1). Lineage 3 was sampled from waters throughout the global ocean (Figure 2, sites 6-12) and consisted of 14 sequence types. Lineage 3 was dominated by sequence 3, identified in 38 isolates (Table 3).

The predicted folding structure of the *T. rotula* ITS1 revealed no compensatory base changes among isolates either within or between lineages (Figure 3a). Most mutations occurred in loop regions. Only two mutations (G/A, G/U) occurred in adjacent positions along a stem region, but displayed no corresponding compensatory changes. *T. gravida* and *T. rotula* exhibited different predicted ITS1 folding structures (Figures 3a and 3b).

Using a dated phylogenetic analysis, divergence times were estimated for *T. gravida,* the three *T. rotula* lineages, and several other *Thalassiosira* species (Figure 4). Divergence between *T. gravida* and *T.rotula* was approximately 3.28 Mya. Lineage 3 was calculated to be the oldest lineage, from which lineages 1 and 2 diverged 0.68 Mya. Lineage 2 diverged most recently from lineage 1 at 0.22 Mya.

The ITS1 is present in multiple copies within the genome and thus each isolate may contain intra-genomic variants. Intra-genomic variation in the ITS1 was examined in one isolate from each lineage: VIA (ST1, site 2), PS (ST2, site 4), and SIS (ST3, site 6). From isolate VIA (ST1), 17 of 20 amplicons had sequences identical to ST1, and three were novel STs that differed by 1bp from ST1. From isolate PS (ST2), 13 of 20 amplicons were identical to ST2, and the remaining 7 were novel singleton sequences that differed by 1-7 bps. From isolate SIS (ST3), 16 of 20

isolates were identical to ST3, and 4 represented novel sequences that each differed by 1bp. Average within-lineage sequence divergence was 0.18 ± 0.10%, 1.90 ± 0.50%, and 0.25 ± 0.18% for lineages 1, 2, and 3 respectively. Neither probabilistic divergence measures nor difference of sums of squares revealed recombination among sequence types ($p>0.05$). In addition, RDP3 analysis revealed no evidence for recombination among sequence types ($p>0.05$); however, with such low levels of divergence in this dataset, recombination may be below the limits of detection for RDP3 (Martin et al. 2010). When added to the network analysis, intra-individual variants did not alter the initial network structure (data not shown) and lineage groupings 1-3 remained significant ($p<0.001$).

Physiological variation

Triplicate maximum acclimated growth rates of axenic and xenic cultures of VIA were not significantly different ($p>0.05$) indicating that bacterial presence in cultures did not alter growth rate. Subsequent experiments were conducted using both xenic (CCAP1085_21, CCMP3264, VIB) and axenic (VIA, CCMP1647, SIS) isolates. Isolates VIA and VIB represented lineage 1 and were collected from the Eastern North Pacific (sites 2 and 3). Four isolates represented lineage 3, and were collected from the Western North Pacific (SIS, site 6), and the Mediterranean Sea (CCAP1085_21, CCMP1647, CCMP3264, sites 11, 12, and 15). No live isolates of lineage 2 were available.

There was no significant clustering of growth rate with ITS1 lineage at any of the six light and temperature conditions ($p>0.05$). Instead, relative growth rate among isolates differed with treatment, illustrated by extensive crossing of growth curves as

environmental conditions changed (Figures 5a and b). At high light intensity (112 μmol photons m^{-2} s^{-1}), specific growth rates ranged from no growth to 0.92 ± 0.04 day^{-1} (Figure 5a). At low light intensity (50 μmol photons m^{-2} s^{-1}), specific growth rates ranged from 0.22 ± 0.01 - 0.66 ± 0.01 day^{-1} (Figure 5b). At both light intensities, the CV was significantly larger at 4°C than at 10 or 17.5°C ($p<0.05$). Growth rate varied significantly among isolates at each temperature ($p<0.05$). There were no significant relationships between growth rate and cell diameter at any culturing condition ($p>0.05$).

The physiological response of isolates to light intensity was dependent on temperature. For example, at 4°C, there were no significant differences ($p>0.05$) in growth rates between high and low light intensities for all isolates except SIS (Figures 5a and 5b). In contrast, at 17.5°C growth rates differed significantly ($p<0.05$) between high and low light intensities for all isolates except VIB (Figures 5a and 5b).

Variation in relative genome size

Genome size was measured in five isolates. Two of these isolates represented lineage 1: VIA (ST1) and VIB (ST1). Three isolates represented lineage 3: SIS, CCMP3264, and CCAP1085_21 (ST3). No live isolates were available for lineage 2. The samples from all isolates contained cells in both G1 and G2 phases of the cell cycle, as indicated by bimodal distributions of integrated DNA fluorescence, with peaks separated by a factor of 2 in fluorescence intensity (Figure 6). G1 and G2 distributions provided an internal standard of our ability to detect changes in genome size. The relative frequency of G1 cells in VIB was significantly higher than in other isolates ($p<0.05$), likely due to different physiological responses to light and

temperature conditions used to culture cells for flow cytometry. Among isolates from lineage 3, SIS exhibited 30% less G1 integrated fluorescence than CCMP3264 and CCAP1085_21 ($p<0.05$). The mode of G1 for CCMP3264 and CCAP1085_21 was 0.15 rfu. Within ST1, the mode of G1 fluorescence signal for VIB was 0.27 rfu for G1, approximately twice as high as all other isolates. Relative genome size of VIB was significantly different from all other isolates, with a nearly two-fold increase in fluorescence intensity at both G1 and G2 peaks ($p<0.05$).

There was no relationship between genome size and cell size measurements. Average cell volume of all isolates was 21.52 ± 5.56 μm^3. There were no significant differences in cell volume between different isolates ($p>0.05$). However, VIB exhibited the largest average cell volume (28.65 ± 13.22 $mm^{3.}$), and exhibited the largest range in cell size.

DISCUSSION

To determine subdivision within the *T. rotula* morphospecies, it was first necessary to examine genetic divergence between *T. rotula* and *T. gravida*. Previous studies identified morphological plasticity in the characteristics used to define each species and argued for a single species designation (Sar et al. 2011; Syvertsen 1977). Here, we found that culture collection isolates of the two species differed by at least 7% at the ITS1 and 0.8% at the 28S rDNA. This level of divergence is comparable to that observed between different species of the diatom *Skeletonema* (0.5% 28S divergence) (Sarno 2005) and between *Pseudo-nitzschia* species (7.2% ITS1 divergence) that were confirmed using mating experiments (Amato et al. 2007). ITS1 sequence variation indicated that *T. rotula* and *T. gravida* diverged approximately 3.28

Mya. Furthermore, the predicted ITS1 secondary structures of the two species differed considerably, a characteristic related to reproductive incompatibility in protists (Behnke et al. 2004; Coleman 2000; Coleman 2003; Coleman 2007; Coleman and Vacquier 2002). Although the links between reproductive isolation and ITS1 folding structure are not as well understood as the ITS2, it has been suggested that the ITS1 and ITS2 molecules co-evolve to maintain important biochemical interactions necessary for processing the mature ribosome (Armbrust and Galindo 2001b; Coleman 2000). Here, differences in the predicted *T. gravida* and *T. rotula* folding structures indicate that significant evolution has occurred at the ITS1 (Armbrust and Galindo 2001b; Coleman 2000). Differences in rDNA sequences and predicted ITS1 secondary structures between culture collection isolates of *T. rotula* and *T. gravida* suggest that the original species designations are correct.

The majority of field isolates were not significantly different from *T. rotula* culture collection isolates at the 18S and 28S rDNA. This pool of isolates could be divided into three distinct ITS1lineages which diverged from *T. rotula* culture collection isolates by 0- 0.6%, an order of magnitude less than their divergence with *T. gravida*. This level of variation is comparable to that identified within other diatom species; for example, *Pseudo-nitzschia pungens* clades diverged by 0.5% at the ITS1 (Enderlein 1961b).

Several lines of evidence suggest that the three lineages may be able to interbreed. First, their predicted ITS1 secondary structures were identical to each other, suggesting that mutations in the ITS1 have not resulted in significant structural changes to this important molecule (Armbrust and Galindo 2001b). In addition, there

were no compensatory base changes (CBCs) in the ITS1 stem regions of the three lineages, suggesting that this gene is conserved. Overall, the lack of CBCs and conservation of secondary structure among lineages suggest that they may retain the ability to interbreed. Second, there were no consistent differences in genome size among lineages, another indication that they may be able to interbreed. Importantly, differences in genome size were observed within and not between lineages. Within lineage 1, genome size differed by roughly two fold and within lineage 3, by 30%. Genome duplication, or polyploidization, is common in plants, and has been shown to result in rapid reproductive isolation (Hegarty and Hiscock 2007; Hegarty and Hiscock 2008; Otto 2003; Otto 2007; Wood 2009). In diatoms, few studies have examined changes in DNA content. Genome size differences of two-fold have been observed among ITS1 lineages of the diatom *D. brightwellii*, diverged by only 0.8% at the ITS1, suggesting that each lineage may instead represent a distinct species (Koester et al. 2010). Furthermore, DNA content has been shown to vary among species within the genus *Thalassiosira,* suggesting that polyploidization may play a role in the evolution of diatoms as well as plants (Von Dassow 2008). Because variations in genome size did not correlate with ITS1 lineage in *T. rotula*, this metric did not provide evidence for consistent barriers to interbreeding. Instead, the observation that two out of five strains differed in genome size suggests that *T. rotula,* and perhaps diatoms in general, may have a relatively plastic genome complement (Von Dassow 2008).

Although several lines of evidence suggest that interbreeding could occur, additional data suggest that these lineages may not be actively interbreeding. To look for signatures of recombination, multiple copies of the ITS1 were sequenced from

individuals representing each lineage. If lineages were interbreeding, one might expect to find, for example, some copies or recombinants of a lineage 1 sequence in a lineage 2 individual (Butlin 1987). No signature of recombination could be detected among lineages suggesting that interbreeding, if it occurs, is infrequent and below the threshold of detection. Here, we used a single genetic marker to examine recombination; future analysis of recombination would be improved by surveying a greater number of genes. It is also worth noting that diatoms divide primarily asexually, and that sexual cycles have been examined for only a handful of the large number of described species (Levins 1969; Templeton 1981). Sexual recombination in the field has been observed (Servedio 2003; Townsend et al. 1994), but rarely, and estimates of the incidence sexual recombination in diatoms varies widely, from once per year to once every 40 yrs (Sverdrup 1953). In addition to an inability to detect recombination, it appears that gene flow between lineages has been reduced for significant time periods. A dated phylogenetic analysis indicated that lineage 3 diverged from *T. gravida* 3.28 Mya. Lineages 1 and 2 diverged later, at 0.68 Mya. Because divergence calculations can vary depending on outgroups, genes, or calibration points used in analysis (Casteleyn et al. 2010; Sorhannus 2007), divergence times should be interpreted cautiously. Even if these estimations are off by orders of magnitude, the estimated time since last interbreeding is significant.

In the marine environment, interbreeding could cease to occur through such mechanisms as isolation by distance, physical barriers to gene flow, competitive exclusion, environmental adaptation, or genetic and phenological characteristics that prevent gametes from fusing in the field (Palumbi 1994). Here, it appears that

isolation by distance was an unlikely mechanism promoting differentiation among *T. rotula* lineages. For example, genetic distance among lineages was not related to geographic distance. In fact, *T. rotula* lineage 3 had a cosmopolitan distribution ranging from the Mediterranean Sea, the N. Atlantic to the N. Pacific. This distribution is comparable to that observed in lineages of the pennate diatom *P. pungens* (Casteleyn et al. 2008) and contrasts with many terrestrial plant species, where genetic distance among lineages often correlates with geographic distance (eg.(Chiang and Schaal 1999; Sork et al. 1999)). The observation that lineages can be broadly distributed in both centric and pennate diatoms suggests that dispersal likely plays a significant role in regulating gene flow. Lack of isolation by distance observed here suggests that there are no physical barriers impeding broad dispersal.

Instead of acting as barriers, physical features, such as water recirculation, may act to reduce gene flow, allowing different lineages to arise and be maintained. For example, hydrographic features have been hypothesized to drive genetic divergence in diatoms in coastal fjords of the NE Pacific where recirculating water may retain cells inside the fjord, allowing them to remain in and adapt to a particular location (Rynearson 2004). Interestingly, lineage 2 was observed within a recirculating coastal fjord in the NE Pacific, and exhibited significant divergence from lineage 1 sampled outside of the fjord, suggesting that water recirculation may influence genetic subdivision in multiple species of phytoplankton.

Competitive exclusion and environmental adaptation may be additional mechanisms initiating and supporting lineage divergence. Here, all but one location was dominated by just a single lineage. In those locations, the probability that other

lineages were present but not detected was low. For example, a lineage representing 10% of the population would have a 99% probability of being detected in our 40 isolates collected from the N. Atlantic (Narragansett Bay and Martha's Vineyard) (Rynearson et al. 2009), suggesting that these sites were likely dominated by single lineage. This may be due to competitive exclusion of lineages not adapted to the environment in Narragansett Bay and coastal N. Atlantic. There may, however, be more complex dynamics in play. For example, all three lineages were sampled from the Queen Charlotte Islands in the NE Pacific. This location may act as a hub of intermixing between water masses and may provide an environment heterogeneous enough to support the ecological niches of all three lineages.

To explore potential signatures of environmental adaptation, we compared the physiological response of each lineage to a range of light intensities and temperatures, two important environmental variables known to affect phytoplankton growth (Bonin et al. 2007; Fleishman et al. 2009; Leal and Fleishman 2002). As established by Brand (Brand 1981), differences in acclimated growth rates can be used to identify underlying genetic variation among isolates grown in a single environmental condition. Analyzing the growth rates of isolates under different conditions allows for comparisons of genotypic versus environmental effects (Falconer 1952). If environment alone were driving differences in growth, isolates would exhibit the same relative difference in growth rate, regardless of environment. Here, there was no correlation between ITS1 lineage and growth characteristics examined. Instead, isolates exhibited significant genotype-by-environment (Falconer 1952) responses to light and temperature. The genotype-by-environment experiments conducted here

suggest that there is additional clonal diversity within each *T. rotula* lineage, similar to that identified in other phytoplankton species (Evans 2005; Iglesias-Rodriguez et al. 2006; Rynearson et al. 2009). Lack of clear differentiation between lineages in their physiological response to light and temperature does not mean that environmental adaptation has not occurred among lineages since other factors such as predation and nutrient availability may be more important drivers of environmental adaptation among lineages.

The persistence of single lineages within individual locations suggests that there may be yet another type of environmental adaptation that allows diatoms to diverge into distinct lineages. For example, all isolates collected from Narragansett Bay represented a single lineage, regardless of the month of sampling (January, February, June or October). Furthermore, isolates collected from the Gulf of Naples between 1993 and 2008 represented the same lineage, suggesting that it persisted over many years in a single habitat. Similarly, identical genotypes of *D. brightwellii* were detected in Puget Sound over a seven-year period (Rynearson et al. 2009). The persistent occurrence of lineages at individual locations may relate to the ability of many diatoms to create resting spores, which lie dormant in the sediment and may remain viable for decades (Härnström et al. 2011). *T. rotula* can form resting spores, although it is not a required part of its life cycle and the frequency of resting spore formation in the field and the rate of germination success are unknown (Garrison 1981; Karentz and Smayda 1984; Mcquoid and Hobson 1995; Mcquoid and Hobson 1996). An intriguing hypothesis is that environmental adaptation to different

phenological triggers for spore formation and germination could foster the initiation and maintenance of distinct lineages.

Finally, interbreeding could cease to occur among lineages if prezygotic barriers to gene flow no longer allow for sexual recombination among lineages. Prezygotic barriers to gene flow include changes to gametes that prevent them from recognizing each other (Palumbi 1994). In diatoms, the *Sig1* gene has been hypothesized to play an important role in gamete recognition (Armbrust and Galindo 2001a). This gene has undergone rapid evolution among strains of the diatom *T. weissflogii*, including distinct protein changes that may alter the interaction between gametes in the field (Armbrust and Galindo 2001a). This type of prezygotic barrier to gene flow may lead to speciation via reinforcement of postzygotic differentiation (Butlin 1987; Servedio 2003). Previous attempts to amplify the *Sig1* gene in *T. rotula* have failed (Armbrust and Galindo 2001a), but future transcriptional or genome-wide sequencing analyses may shed light on the potential for a prezygotic barrier to gene flow between lineages observed here. A prezygotic barrier to gene flow could well explain the conflicting data we obtained regarding interbreeding among lineages, where genome size and RNA secondary structure indicated no barriers to interbreeding but active recombination may no longer occur. In this scenario, prezygotic barriers to gene flow would prevent sexual recombination among lineages but the rDNA would not have yet diverged to levels that compare with more distantly-related species.

CONCLUSIONS

Our data suggest that genetic divergence between *T. rotula* and *T. gravida* is significant and that they should continue to be described as distinct species, with future investigations to more fully describe differences between them. From a taxonomic standpoint, the divergence among *T. rotula* lineages is less clear. On one hand, it appears as if interbreeding between lineages has not occurred for long periods of time, suggesting that they could represent recently diverged cryptic species. On the other hand, lineages exhibited no differences in ITS1 secondary structure or genome size and exhibited no clear physiological partitioning, suggesting that they are conspecific. Prezygotic barriers to gene flow and differential adaptation to environmental factors involved in vegetative growth and/or spore formation are possible mechanisms that explain these conflicting results. Given the current data, the *T. rotula* lineages should be considered a single species. Future studies to tease apart their relationships will benefit from analyses of genetic variation beyond the rDNA.

The high dispersal marine environment is clearly not a barrier to speciation in diatoms, a group of organisms with an estimated 100- 200,000 species, second only to angiosperms as the most diverse primary producers on the planet (Mann 1996; Round 1990). Confirming species identity in closely-related diatom species is particularly difficult because testing for reproductive compatibility in diatoms is often not feasible in the lab and ultimately, may not reflect recombination in the field (Mann 1999). Differences in the abundance of *T. rotula* lineages in the field suggest that past evolutionary events promoting their subdivision have led to lineage designations that

are likely ecologically relevant, highlighting that in diatoms, a close interplay of ecology and evolution may regulate their impact on global biogeochemical cycles.

COMPETING INTERESTS

This work is associated with no competing interests, financial or otherwise.

AUTHORS' CONTRIBUTIONS

TR and KW conceived and designed the present study; KW, RO, and DR performed the experiments; KW and TR analyzed molecular data, RO and KW analyzed genome size data, DR, KW and TR analyzed physiological data; KW and TR wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This research was supported by National Science Foundation grants, NSF 0727227 (to TAR) and NSF SBE0245039 (to URI, TAR). Part of the research was conducted using instrumentation supported by NSF-EPSCoR grants 0554548 and 1004057. We thank Dr. O. Kawaguchi for collecting isolates from Seto Inland Sea, Japan.

Aloisie Poulickova, J. V., Jiri Neustupa, and Pavel Skaloud. 2010. Pseudocryptic Diversity versus Cosmopolitanism in Diatoms: a Case Study on *Navicula cryptocephala* Kutz. (Bacillariophyceae) and Morphologically Similar Taxa. Protist **161:** 353-369.

Alverson, A. J. 2008. Molecular systematics and the diatom species. Protist **159:** 339-353.

Amato, A., W. H. C. F. Kooistra, J. H. Levialdi Ghiron, D. G. Mann, T. Proschold, and M. Montresor. 2007. Reproductive Isolation among Sympatric Cryptic Species in Marine Diatoms. Protist **158:** 193-207.

Armbrust, E. V., and H. M. Galindo. 2001a. Rapid evolution of a sexual reproduction gene in centric diatoms of the genus Thalassiosira. Appl Environ Microbiol **67:** 3501-3513.

Armbrust, V. E., and H. M. Galindo. 2001b. Rapid Evolution of a Sexual Reproduction Gene in Centric Diatoms of the Genus *Thalassiosira*. Applied Environmental Microbiology **67:** 3501-3513.

Auwera, G. V. D., and R. D. Wachter. 1998. Structure of the Large Subunit rDNA from a Diatom, and Comparison Between Small and Large Subunit Ribosomal RNA for Studying Stramenopile Evolution. Journal of Eukaryotic Microbiology **45:** 521-527.

Behnke, A., T. Friedl, V. A. Chepurnov, and D. G. Mann. 2004. Reproductive compatibility and rDNA sequence analysis in the Sellaphora pupula species complex (Bacillariophyta). Journal of phycology **40:** 193-208.

Behrenfeld, M. J. and others 2006. Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics. Nature **442:** 1025-1028.

Bolker, B. M. 1999. Spatial Moment Equations for Plant Competition: Understanding Spatial Strategies and the Advantages of Short Dispersal. The American naturalist **153:** 575-602.

Bonin, A., D. Ehrich, and S. Manel. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Molecular Ecology **16:** 3737-3758.

Brand, L. E. 1981. Genetic variability in reprodution rates in marine phytoplankton populations. Evolution **35:** 1117-1127.

Brand, L. E., R. R. L. Guillard, and L. S. Murphy. 1981. A method for the rapid and precise determination of acclimated phytoplankton reproduction rates. Journal of Plankton Research **3:** 193-201.

Brockmann, U. and others 1977. Parallel plastic tank experiments with cultures of marine diatoms. Helgoland Marine Research **30:** 201-216.

Burger, W. C. 1981. Why are there so many kinds of flowering plants? Bioscience **31:** 572.

Bursa, A. 1961. Phytoplankton of the *Calanus* expedition in Hudson Bay, 1953 and 1954. . J. Fish. Res. Bd Can. **18:** 51-83.

Butlin, R. 1987. Speciation by reinforcement. Trends in Ecology & Evolution **2:** 8-13.

Carr, M. H., J. E. Neigel, J. A. Estes, S. Andelman, R. R. Warner, and J. L. Largier. 2003. Comparing Marine and Terrestrial Ecosystems: Implications for the Design of Coastal Marine Reserves. Ecological Applications **13:** S90-S107.

Cassie, V. 1960. Seasonal changes in diatoms and dinoflagellates of the east coast of New Zealand during 1957 and 1958. . N.Z.J.Sci. **3:** 137-172.
Casteleyn, G. 2008. Pseudo-nitzschia pungens (Bacillariophyceae): a cosmopolitan diatom species? Harmful algae **7:** 241.
Casteleyn, G. and others 2008. Pseudo-nitzschia pungens (Bacillariophyceae): A cosmopolitan diatom species? Harmful Algae **7:** 241-257.
Casteleyn, G. and others 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. Proceedings of the National Academy of Sciences **107:** 12952-12957.
Cermeno, P., C. De Vargas, F. Abrantes, and P. G. Falkowski. Phytoplankton biogeography and community stability in the ocean. PLoS One **5:** e10037.
Chiang, T. Y., and B. A. Schaal. 1999. Phylogeography of North American populations of the moss species Hylocomium splendens based on the nucleotide sequence of internal transcribed spacer 2 of nuclear ribosomal DNA. Molecular Ecology **8:** 1037-1042.
Cleve, P. T. 1896. Redogörelse för de svenska hydrografiska undersökningarne februari 1896 V. Planktonundersökningar: Vegetabilskt plankton. Bihang till Kungliga Svenska Vetenskapsakademiens Handlingar **22:** 1-33.
Coleman, A. W. 2000. The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. Protist **151:** 1-9.
Coleman, A. W. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends in genetics **19:** 370.
---. 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucleic acids research **35:** 3322-3329.
Coleman, A. W., and V. D. Vacquier. 2002. Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (Haliotis). J Mol Evol **54:** 246-257.
Ellegaard, M. 2008. The species concept in a marine diatom: LSU rDNAÃ¢Â€Â“based phylogenetic differentiation in Skeletonema marinoi/dohrnii (Bacillariophyceae) is not reflected in morphology. Phycologia (Oxford) **47:** 156-167.
Ellegaard, M., A. Godhe, K. Härnström, and M. Mcquoid. 2008. The species concept in a marine diatom: LSU rDNA- based phylogenetic differentiation in Skeletonema marinoi/dohrnii (Bacillariophyceae) is not reflected in morphology. Phycologia (Oxford) **47:** 156-167.
Enderlein, G. 1961a. Scheffé, H.: The Analysis of Variance. Wiley, New York 1959, 477 Seiten. Biometrische Zeitschrift **3:** 143-144.
Enderlein, G. 1961b. Scheffé, H.: The Analysis of Variance. Wiley, New York 1959, 477 Seiten, $ 14,00. Biometrische Zeitschrift **3:** 143-144.
Evans, K. M. 2005. High levels of genetic diversity and low levels of genetic differentiation in North Sea Pseudo-nitzchia pungens (Bacillariophyceae) populations. . Journal of phycology **41:** 506-514.
Evans, K. M., A. H. Wortley, and D. G. Mann. 2007. An Assessment of Potential Diatom Barcode Genes (cox1, rbcL, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in Sellaphora (Bacillariophyta). Protist **158:** 349-364.

Excoffier, L. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary bioinformatics online **1:** 47.
Falconer, D. S. 1952. The Problem of Environment and Selection. The American Naturalist **86:** 293-298.
Field, C. B. 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science (New York, N.Y.) **281:** 237-240.
Fleishman, L., M. Leal, and M. Persons. 2009. Habitat light and dewlap color diversity in four species of Puerto Rican anoline lizards. Journal of Comparative Physiology A **195:** 1043-1060.
Garrison, D. L. 1981. Monterey Bay Phytoplankton. II. Resting Spore Cycles in Coastal Diatom Populations. Journal of Plankton Research **3:** 137-156.
Gran, H., T. Thompson. 1930. The diatoms and the physical and chemial conditions of the seawater of the San Juan Archipelago. Puget Sound Biol. **7:** 169-204.
Guillard, R. R. 1975. Culture of phytoplankton for feeding marine invertebrates. Plenum Press.
Hallegraeff, G. M., 2010. Ocean climate change, phytoplankton community responses, and harmful algal blooms: A formidable predictive challenge. Journal of Phycology **46:** 220-235.
Härnström, K., M. Ellegaard, T. J. Andersen, and A. Godhe. 2011. Hundred years of genetic structure in a sediment revived diatom population. Proceedings of the National Academy of Sciences.
Hegarty, M., and S. Hiscock. 2007. Polyploidy: Doubling up for Evolutionary Success. Current biology : CB **17:** R927-R929.
Hegarty, M. J., and S. J. Hiscock. 2008. Genomic Clues to the Evolutionary Success of Polyploid Plants. Current Biology **18:** R435-R444.
Hubbard, K. A. 2008. Inter and intraspecific community structure within the diatom genus Pseudo-nitzschia (Bacillariophyceae) Journal of phycology **44:** 637-649.
Iglesias-Rodriguez, M. D., O. M. Schofield, J. Batley, L. K. Medlin, and P. K. Hayes. 2006. Intraspecific genetic diversity in the marine coccolithophore *Emiliania huxleyi* (Prymnesiophyceae): the use of microsatellite analysis in marein phytoplankton population studies. Journal of Phycology **42:** 525-536.
Karentz, D., and T. J. Smayda. 1984. Tempearture and seasonal occurence patterns of 30 dominatn phytoplankton species in Narragansett Bay over a 22-year period (1969-1980. . Marine Ecology Progress Series **18:** 277-293.
Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution **16:** 111-120.
Kinlan, B. P. 2003. Propagule dispersal in marine and terrestrial environments: a community perspective Ecology (Durham) **84:** 2007-2020.
Koester, J. A., J. E. Swalwell, P. Von Dassow, and E. V. Armbrust. 2010. Genome size differentiates co-occurring populations of the planktonic diatom Ditylum brightwellii (Bacillariophyta). BMC Evol Biol **10:** 1.
Kooistra, W. 2008. Global diversity and biogeography of Skeletonema species (Bacillariophyta). Protist **159:** 177.

Krawiec, R. W. 1982. Autecology and clonal variability of the marine centric diatom Thalassiosira rotula; (Bacillariophyceae) in response to light, temperature and salinity. Marine Biology **69:** 79-89.
Laisk, A. and others 2009. Photosynthesis in Global-Scale Models, p. 465-497. *In* Govindjee [ed.], Photosynthesis <i>in silico</i>. Advances in Photosynthesis and Respiration. Springer Netherlands.
Leal, M., and L. J. Fleishman. 2002. Evidence for habitat partitioning based on adaptation to environmental light in a pair of sympatric lizard species. Proceedings of the Royal Society of London. Series B: Biological Sciences **269:** 351-359.
Levins, R. 1969. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. Bulletin of the ESA **15:** 237-240.
Mann, D. G. 1996. 3. Biodiversity, biogeography and conservation of diatoms. Hydrobiologia **336:** 19.
Mann, D. G. 1999. The species concept in diatoms. Phycologia **38:** 437-495.
Martin, D. P., P. Lemey, M. Lott, V. Moulton, D. Posada, and P. Lefeuvre. 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics **26:** 2462-2463.
Mathews, D. H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure1. Journal of molecular biology **288:** 911.
Matsudaira, Y. 1964. Cooperative studies on priamry production in coastal waters of Japan 1962-1963. Inform. Bull. Planktol. Japan **11:** 24-73.
Mcquoid, M. R., and L. A. Hobson. 1995. Importance of resting stages in diatom seasonal succession Journal of Phycology **31:** 44-50.
Mcquoid, M. R., and L. A. Hobson. 1996. Diatom resting stages. Journal of Phycology **32:** 889-902.
Medlin, L., H. J. Elwood, S. Stickel, and M. L. Sogin. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene **71:** 491-499.
Meunier, A. 1910. Microplancton des Mers de Berents et de Kara. *Duc d'Orleans, Champagne Artique de 1907***:** 1-355.
Moniz, M. B., and I. Kaczmarska. 2009. Barcoding micro- and meso-fauna. Barcoding diatoms: Is there a good marker? Molecular Ecology Resources **9:** 65-74.
Nelson, D. M., P. Treguer, M. A. Brzezinski, A. Leynaert, and B. Queguiner. 1995. Production and dissolution of biogenic silica in teh ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Global Biogeochemical Cycle **9:** 359-372.
Olson, R. J., and H. M. Sosik. 2007. A submersible imaging-in-flow instrument to analyze nano and mciroplankton: Imaging FlowCytobot. Limnology and Oceanography: Methods **5:** 195-203.
Otto, S. P. 2003. In polyploids, one plus one does not equal two. Trends in ecology & evolution (Amsterdam) **18:** 431.
Otto, S. P. 2007. The evolutionary consequences of polyploidy. Cell (Cambridge) **131:** 452.

Palumbi, S. R. 1994. Genetic Divergence, Reproductive Isolation, and Marine Speciation. Annual review of ecology and systematics **25:** 547-572.
Pratt, D. M. 1959. The phytoplankton of Narragansett Bay. Limnology and oceanography **4:** 425.
Raymond, M. 1995. An exact test for population differentiation. Evolution **49:** 1280.
Round, F. E., R.M. Crawford, D.G. Mann. 1990. The Diatoms: biology and morphology of the genera. Cambridge University Press.
Rynearson, T. A. 2004. Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). Journal of phycology **40:** 34-43.
Rynearson, T. A. 2006. Spring bloom development, genetic variation, and population succession in the planktonic diatom Ditylum brightwellii. Limnology and oceanography **51:** 1249.
Rynearson, T. A., and E. V. Armbrust. 2000. DNA Fingerprinting Reveals Extensive Genetic Diversity in a Field Population of the Centric Diatom Ditylum brightwellii. Limnology and Oceanography **45:** 1329-1340.
Rynearson, T. A., E. O. Lin, and E. V. Armbrust. 2009. Metapopulation Structure in the Planktonic Diatom Ditylum brightwellii (Bacillariophyceae). Protist **160:** 111-121.
Rytter Hasle, G. 1976. The biogeography of some marine planktonic diatoms. Deep Sea Research and Oceanographic Abstracts **23:** 319-338, IN311-IN316.
Sanmartan, I. 2004. Southern Hemisphere Biogeography Inferred by Event-Based Models: Plant versus Animal Patterns. Systematic biology **53:** 216-243.
Sar, E. A., I. S. Sunesen, A. S. Lavigne, and S. Lofeudo. 2011. Thalassiosira rotula, a heterotypic synonym of Thalassiosira gravida: morphological evidence. Diatom Research **26:** 109-119.
Sarno, D. 2005. Diversity in the genus Skeletonema (Bacillariophyceae). An assessment of the taxonomy of S. costatum-like species with the description of four new species. . Journal of phycology **41:** 151-176.
Sarno, D., W. H. C. F. Kooistra, S. Balzano, P. E. Hargraves, and A. Zingone. 2007. Diversity in the genus *Skeletonema* (Bacillariophyceae) III. Phylogenetic position and mophological variability of *Skeletonema grevillei*, with the description of *Skeletonema ardens* sp. Journal of Phycology **43:** 156-170.
Servedio, M. R., and G.-P. Srepe. 2003. Speciation as a positive feedback loop between postzygotic and prezygotic barriers to gene flow. Proceedings of the Royal Society of London. Series B: Biological Sciences **270:** 1473-1479.
Simon, N. 2009. Diversity and evolution of marine phytoplankton. Comptes rendus. Biologies **332:** 159.
Simpson, B. B. 1974. Glacial migrations of plants: island biogeographical evidence. Science (New York, N.Y.) **185:** 698-700.
Smayda, T. J. 1957. Phytoplankton studies in lower Narragansett Bay. Limnology and oceanography **2:** 342.
Smayda, T. J. 1958. Biogeographical studies of marine phytoplankton. Oikos **9:** 158.
Sorhannus, U. 2007. A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. Marine Micropaleontology **65:** 1-12.

Sork, V. L., J. Nason, D. R. Campbell, and J. F. Fernandez. 1999. Landscape approaches to historical and contemporary gene flow in plants. Trends in Ecology & Evolution **14:** 219-224.

Sverdrup, H. U. 1953. On Conditions for the Vernal Blooming of Phytoplankton. Journal du Conseil **18:** 287-295.

Syvertsen, E. E. 1977. *Thalassiosira rotula* and *T. gravida*: ecology and morphology. . Nova Hedwigia Beihefte **57:** 99-112.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Molecular Biology and Evolution **24:** 1596-1599.

Templeton, A. R. 1981. Mechanisms of Speciation--A Population Genetic Approach. Annual Review of Ecology and Systematics **12:** 23-48.

Thompson, J. D. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research **22:** 4673-4680.

Townsend, D. W., L. M. Cammen, P. M. Holligan, D. E. Campbell, and N. R. Pettigrew. 1994. Causes and consequences of variability in the timing of spring phytoplankton blooms. Deep Sea Research Part I: Oceanographic Research Papers **41:** 747-765.

Tukey, J. W. 1949. One Degree of Freedowm for Non-Additivity. Biometrics **5:** 232-242.

Von Dassow, P. 2008. Inter and intrspecific relationships between nuclear DNA content and cell size in selected members of the centric diatom genus Thalassiosira (Bacillariophyceae). Journal of phycology **44:** 335-349.

Wood, T. E. 2009. The frequency of polyploid speciation in vascular plants. PNAS : Proceedings of the National Academy of Sciences **106:** 13875-13879.

Zingone, A., I. Percopo, P. A. Sims, and D. Sarno. 2005. Diversity in the genus *Skeletonema* (Bacillariophyceae). I. A reexamination of the type material of *S. costatum* with the description of *S. grevillei* sp. . Journal of Phycology **41:** 140-150.

TABLES

Table 1. **Sites and Isolates**
Description of site and isolates collected, including isolation success and genes sequenced from each site. Site numbers refer to map featured in Figure 1. Isolation number or isolation success number was not available for culture collection strains, nor for those isolated in Japan. Culture collection accession numbers are indicated in the "Origin" column. a) All *T. rotula* isolates, including their classification as lineages 1, 2, or 3, and whether or not they were analyzed for physiological diversity. b) Description of *T. gravida* isolates

a.)

	Site	Origin	Date Collected	No. cells isolated (survived)	ITS1 sequences (#)	18S and 28S Sequences (#)	Lineage 1	Lineage 2	Lineage 3	Phys. Experiment
North Pacific	1	Queen Charlotte Islands, Canada 51.75N, 131W	5/17/2007	21(18)	10		3	6	1	
	2	Vancouver Island (A), Canada 49.65N, 127.44W	5/13/2007	48(41)	10	1	10			1
	3	Vancouver Island (B). Canada 48.87N, 125.89W	5/13/2007	32(30)	10	1	10			1
	4	Puget Sound, USA 47.74 N, 122.42 W	5/12/2007	56(56)	10	1		10		2
	5	La Jolla, USA 32.850 N, 117.25 W (CCMP1018)	1968	NA	1	1	1			1
	6	Seto Inland Sea, Japan (SIS) 34.166 N, 133.33 E	2/2007	NA	5	1			5	1
North Atlantic	7	Martha's Vineyard Coastal Observatory (MVCO), USA 41.455N, 70.5667 W	11/3/2008	17 (15)	10				10	
	8	Narragansett Bay, URI 41.53 N, 71.383W	10/10/2008	24(24)	10				10	
			1/8/2009	24(21)	10	1			10	
			2/9/2009	48(45)	10				10	
			6/26/2009	2(4)	2				2	
	9	Oban, Scotland 56.572 N, 5.434 W (CCAP1085_20)	2008	NA	1	1			1	
Mediterranean Sea	10	Gulf of Naples, Italy 40.956 N, 14.25 E (CCAP1085_21)	2008	NA	1	1			1	1
	11	Gulf of Naples, Italy 40.4900 N, 14.1500 E (CCMP1647)	11/12/2008	NA	1	1			1	
	12	Gulf of Naples, Italy 40.7500 N, 14.3300 E (CCMP3264)	1993	NA	1	1			1	1

b.)

Site	Origin	Date Collected	No. cells isolated (survived)	ITS1 sequences (#)	18S and 28S Sequences (#)
13	Iceland 60.92472N, 27.005833	5/22/2008	48(29)	10	3
14	Tromso, Norway 69.6667N, 18.966E (CCMP986, 987)	8/1/1978	NA	2	1
15	McMurdo SounR77.8333S 163.000E (CCMP1463, CCMP1462)	2/1/1991	NA	2	2

Table 2. **Genbank Accession Numbers** *Thalassiosira sp.* and Genbank accession numbers used in phylogenetic analysis.

Species	**Accession no.**
T. pseudonana	EF208793
T. guillardii	EF208788
T. weissflogii	FJ432753
T. anguste-lineata	EF208800
T. aestivalis	EF208797
T. oceanica	EF208795
T. punctigera	EF208796

Table 3. **ITS1 Variation**

ITS1 variation among 76 isolates of *Thalassiosira rotula* collected from around the globe. This table includes only those sequences represented by >1 isolate. No singleton sequences are displayed. Base pairs indicated with a slash (eg. A/C) indicate ambiguous signals in sequencing, and thus the possible presence of two alleles at that position.

Lineage	Sequence ID	n	Origin	Site #	BP Position 41	44	50	82	95	108	149	200	201
1	1	17	Vancouver Island (A and B)	2,3	A	G	C	C	C	A	-	G	G
	4	2	Vancouver Island (A and B)	2,3	-	-	-	-	A/C	A/T	-	-	-
	5	2	Queen Charlotte Islands, La Jolla, CA	1,5	-	-	-	-	-	-	C	-	-
2	2	11	Puget Sound, Queen Charlotte Islands	4,1	G/A	-	-	-	-	-	-	-	A
	6	4	Puget Sound	4	G	-	-	-	-	-	-	-	G/A
3	3	38	Martha's Vineyard, Narragansett Bay, Japan, Mediterranean, Scotland	6,7,8, 9,10,12	-	-	-	T	-	-	-	T	-
	7	2	Mediterranean and Narragansett Bay	11,8	-	A	C/T	T	-	-	-	T	-

FIGURE LEGENDS

Figure 1. **Network Analysis**

Network analysis representing the most parsimonious relationship between sequence variants, separated by single base pair mutations (dotted lines). From 92 isolates, three lineages were defined (those with more than 10 isolates/sequence type, plus their most closely associated sequence types). Each color represents a lineage, each circle represents a sequence variant, and its size indicates the number of isolates comprising each sequence variant. Numbers inside each node the number of isolates comprising that sequence type. All other nodes represent 1, 2, or 4 isolates, scaled to size.

Figure 2. **Sample Map**

Map of global sample locations from where *T. rotula* isolates were collected. Numbers correspond to location and sample information given in Table 1.

Figure 3. **ITS1 rDNA Folding Structures**

a. Predicted RNA folding structure of ITS1 in *T. rotula* based on consensus sequences of all global isolates. Bars and letters represent alternate base pairs seen in sequence variants. b. Predicted RNA folding structure of ITS1 in *T. gravida* based on consensus sequences of all isolates from Iceland and culture collections.

Figure 4. **Phylogenetic Analysis**

Bayesian analysis of divergence times among *Thalassiosira sp*. based on rDNA ITS1 sequence alignment. Time estimates are derived from a relaxed molecular clock calibrated using Sorhannus [12]. Branch numbers represent time of divergence (Mya). Chronogram shows that *T. rotula* split from *T. gravida* approximately 3.28 Mya. Within *T. rotula,* Lineages 1 and 2 split from Lineage 3 approximately 0.68 Mya. Lineages 1 and 2 diverged from one another approximately 0.22 Mya. The tree topology matches that of Sorhannus [12], with the divergence of *T. pseudonana* from all other *T. sp* at 30 Mya. Placement of *T. weissflogii* and *T. guillardii* differ from Sorhannus [12], which may be due to differences in ITS1 and 18S mutation rates.

Figure 5. **Physiological Experiments**

Black lines represent lineage 1, grey lines represent lineage 3. a) Growth rates for six strains at 4, 10, and 17.5°C, high light (112 um photons m-2 s-1) b) Growth rates for six strains at 4, 10, and 17.5°C, low light (50 um photons m-2 s-1).

Figure 6. **Genome Size**

Average integrated fluorescence signal for five strains of *T. rotula* at G1 and G2 cell phases, measured using the IFCB. Fluorescence is normalized to 6um beads. Dotted line represents mode fluorescence of each strain in G1 cell phase.

Figure 1.

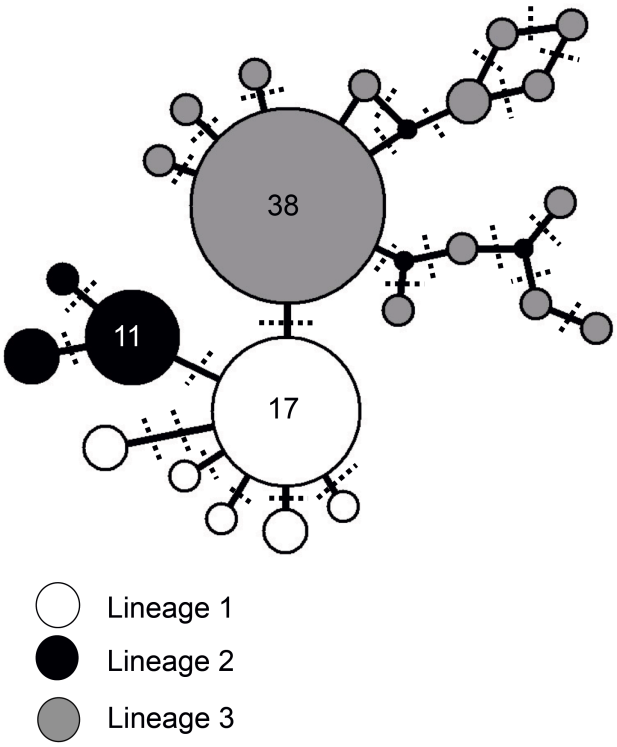


Figure 2.

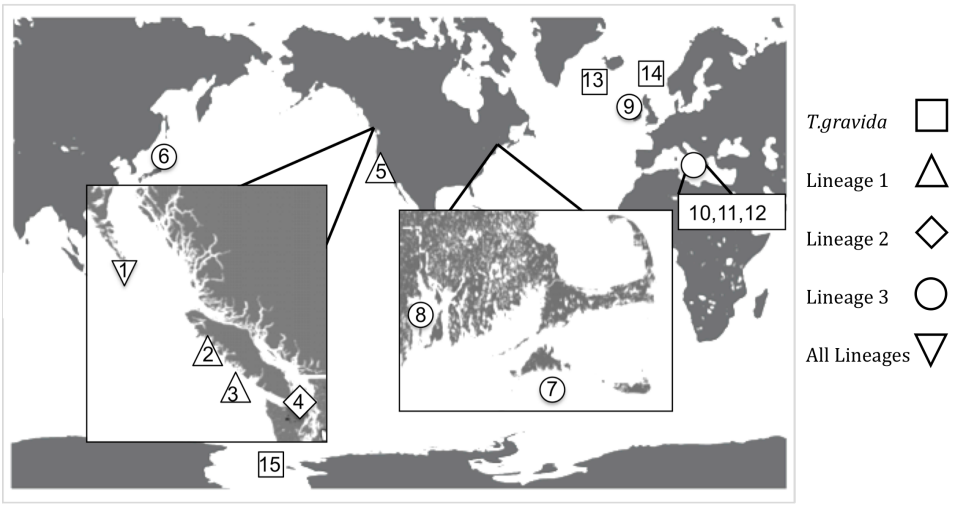

13
14
9
6
5
10,11,12
1
8
2
3
4
7
15
T.gravida
Lineage 1
Lineage 2
Lineage 3
All Lineages

Figure 3.

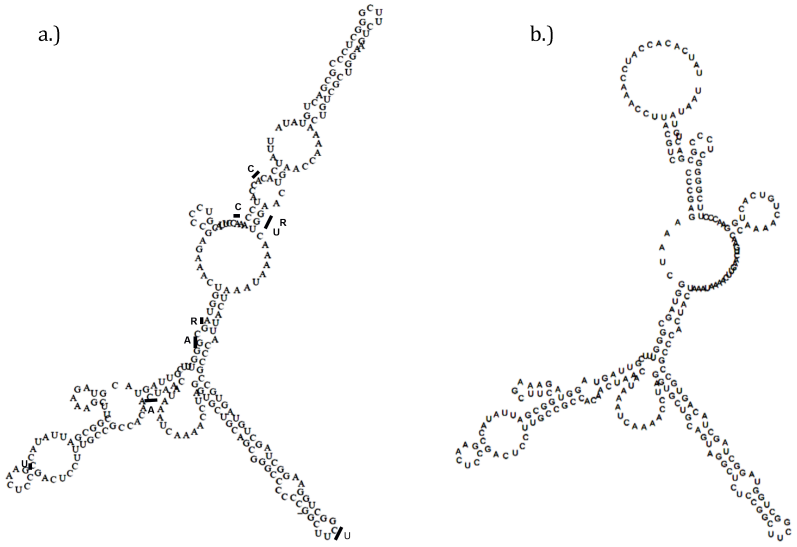


Figure 4.

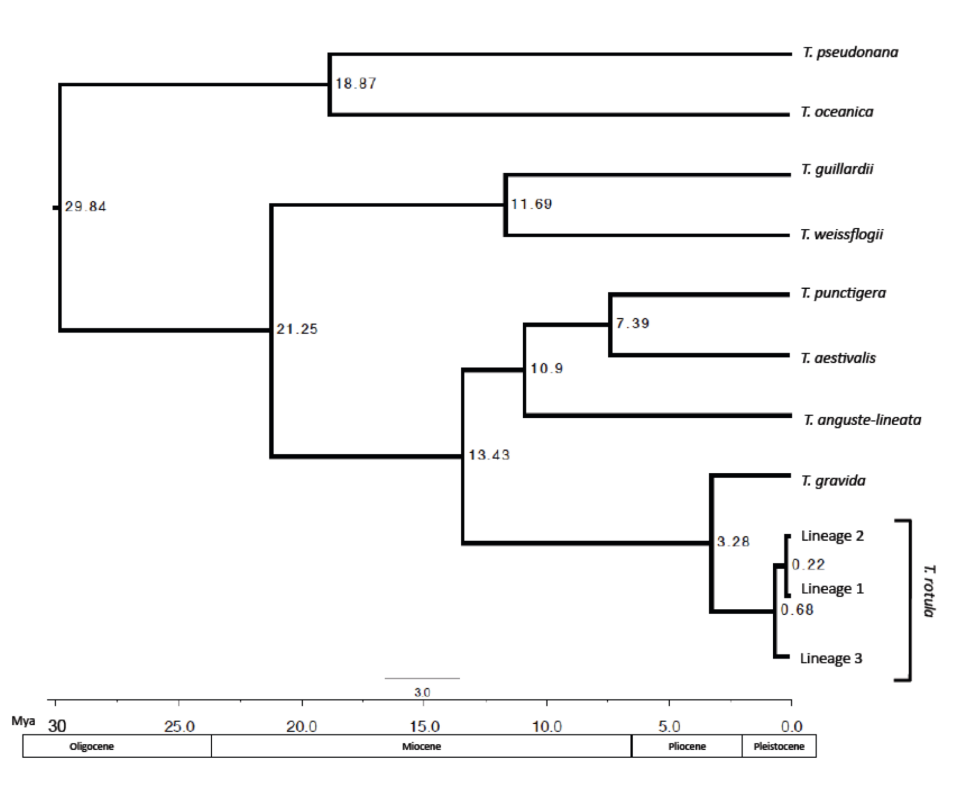

T. pseudonana
18.87
T. oceanica
T. guillardii
29.84
11.69
T. weissflogii
T. punctigera
7.39
21.25
T. aestivalis
10.9
T. anguste-lineata
13.43
T. gravida
3.28
Lineage 2
0.22
Lineage 1
0.68
Lineage 3
T. rotula
3.0
Mya
30
25.0
20.0
15.0
10.0
5.0
0.0
Oligocene
Miocene
Pliocene
Pleistocene

Figure 5.

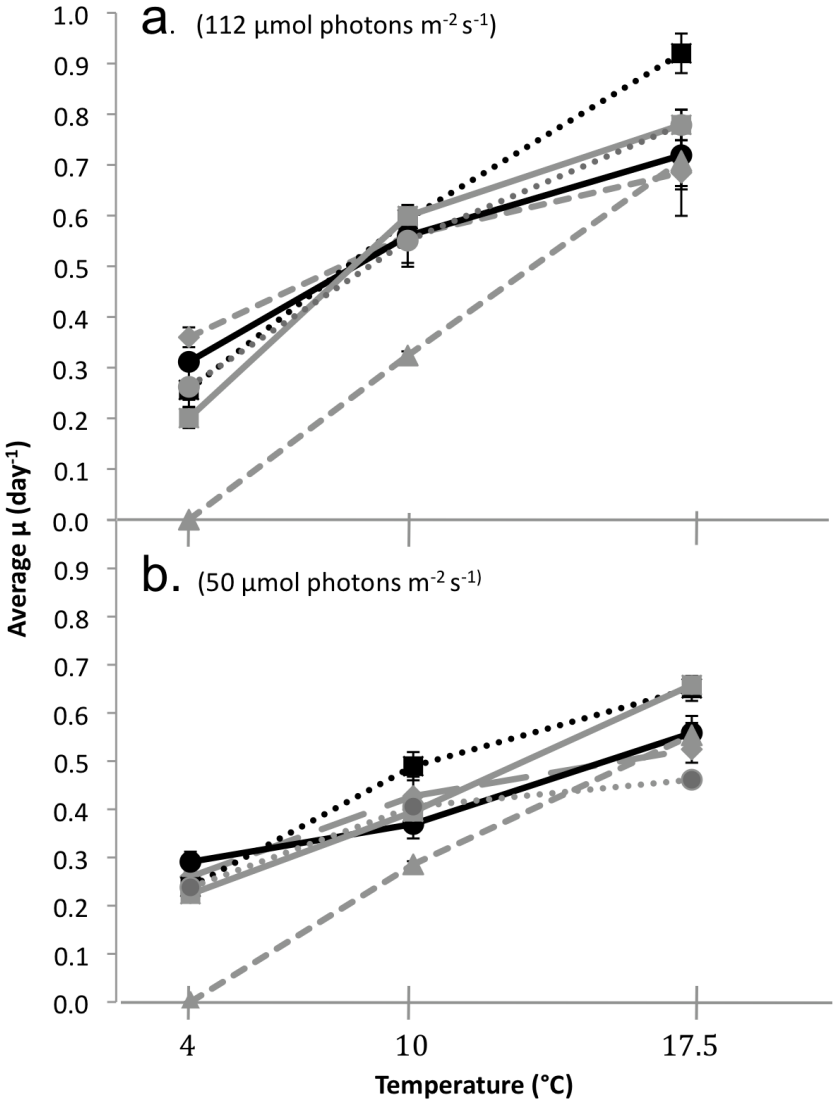
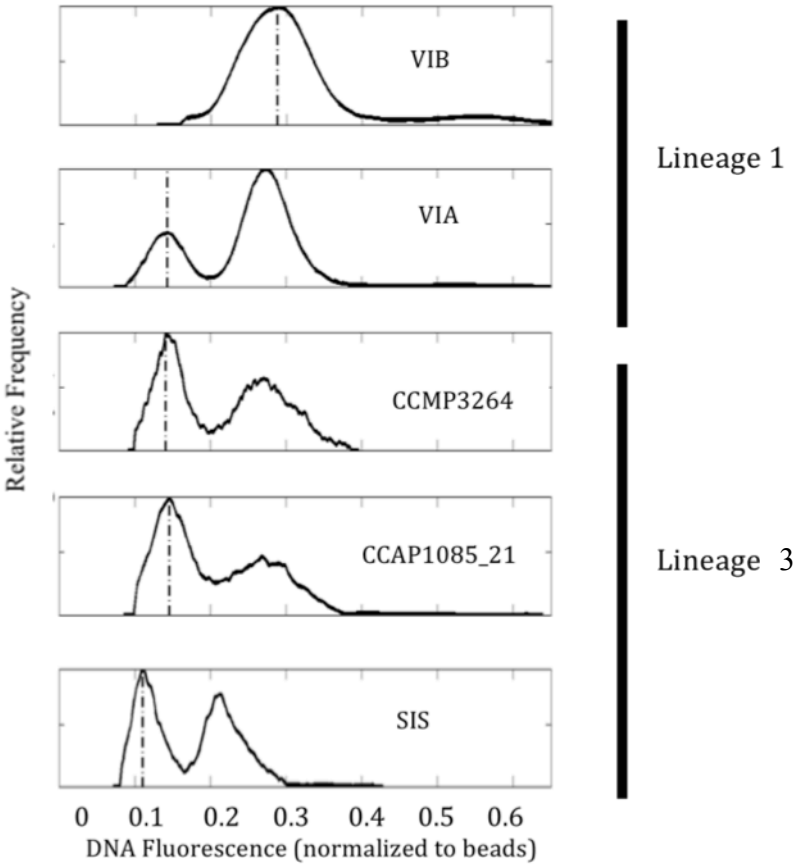

a. (112 µmol photons m-2 s-1)
b. (50 µmol photons m-2 s-1)
Average µ (day-1)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
4
10
17.5
Temperature (°C)

Figure 6.


VIB
VIA
CCMP3264
CCAP1085_21
SIS
Relative Frequency
0
0.1
0.2
0.3
0.4
0.5
0.6
DNA Fluorescence (normalized to beads)
Lineage 1
Lineage 3

CHAPTER 3

PATTERNS AND DRIVERS OF GLOBAL POPULATION STRUCTURE IN THE MARINE DIATOM *THALASSIOSIRA ROTULA*

By

Kerry A. Whittaker[1] and Tatiana A. Rynearson[1]

1. Graduate School of Oceanography, South Ferry Road, University of Rhode Island, Narragansett, RI, 02882

ABSTRACT

Diatoms exhibit astounding levels of inter and intraspecific diversity, yet the mechanisms driving their diversification are little understood. We examined global population structure among 449 isolates of the cosmopolitan, ecologically important diatom species *Thalassiosira rotula.* Isolates sampled over a single bloom period at sites distributed across the Atlantic, Pacific and Indian Ocean basins and between the northern and southern hemispheres were genotyped using six microsatellite markers. Clonal diversity within sites was greater than 99%, and divergence between sites exhibited F_{ST} values upwards of 0.14. The presence of genetically distinct populations demonstrates that significant levels of genetic isolation can occur in the ocean despite the enormous dispersal potential of these planktonic organisms. Isolation-by-distance measures demonstrated that genetic distance was weakly related to geographic distance (r^2 =0.2), suggesting that distance is not a strong barrier to connectivity between diatom populations. A series of Mantel tests was used to examine correlations between pairwise genetic differences among sites, and differences in environmental and ecological factors including temperature, salinity, *T. rotula* cell abundance, and chlorophyll *a*. The combination of chlorophyll *a* and temperature accounted for 53% of genetic variability among sites (p=0.02), suggesting that environmental variables play a role in the structuring genetic variation among diatom populations. In fact, these data suggest that environmental selection may play a more important role than dispersal potential in structuring diversity and controlling geneflow among global populations of *T.rotula.*

INTRODUCTION

The relative influence of dispersal and environmental selection on the genetic connectivity of marine diatom populations has been widely debated for decades (Fenchel and Bland 2004; Finlay 2002; Medlin 2007). On local and regional scales, diatoms possess enormous diversity, often morphologically cryptic, both within and among species (Beszteri et al. 2005; Casteleyn 2008; Casteleyn 2009b; D'alelio et al. 2009; Ellegaard 2008; Evans 2005; Gallagher 1980; Kooistra 2008; Rynearson 2004; Rynearson 2005; Rynearson 2006; Rynearson and Armbrust 2000; Rynearson et al. 2009; Sarno 2005; Whittaker et al. 2012). As single-celled marine organisms, diatoms possess a high potential to disperse across the global ocean. The effect of this global dispersal capacity on the connectivity of diatom populations remains little understood, despite its potential importance for maintaining genetic diversity within species. The extent to which diatom populations are connected over space and time may influence their ability to adapt to variations in the ocean environment. Thus, identifying the structure of diatom diversity over large spatial scales and across diverse marine environments is essential to understanding interactions between their dispersal capacity, environmental selection, and genetic isolation. To date, few studies have attained the molecular or geographic resolution to empirically examine the factors structuring diatom diversity on the global spatial scale.

Molecular studies examining the extent and structure of diatom population diversity, for the most part, have targeted local and regional spatial scales, and species studied to date demonstrate some common characteristics. First, all exhibit high levels of genetic diversity and large population sizes, with H_e values upwards of 0.9 (Godhe

and Härnström 2010; Rynearson 2004; Rynearson et al. 2009), and clonal diversity greater than 90% (Rynearson and Armbrust 2005). For example, a single bloom of the diatom *Ditylum brightwellii* was comprised of thousands of clonal lineages (Rynearson 2005). Second, populations demonstrate stability over time. For instance, in fjords of Sweden, genetic signals from resting spores of the diatom *Skeletonema marinoi* demonstrate that a single population can persist in a local environment for over one hundred years (Härnström et al. 2011). Third, diatom populations exhibit evidence of sympatric isolation, where environmental selection is suspected to be the driving force behind isolation of different populations, possibly leading to speciation, within a single location. For instance, the diatom *D. brightwellii* is subdivided into two distinct populations within Puget Sound, each present during different times of the year (Rynearson 2006). It has been suggested that these populations are, in fact, separate species (Koester, 2011). Early allozyme work shows a similar pattern in Narragansett Bay, where two populations of *Skeletonema costatum* were identified in the spring and fall, respectively (Gallagher 1980). Diatom species studied to date exhibit some differences in population structure as well. For example, in the centric diatom species *D. brightwellii* and *S. marinoi*, significantly diverged populations have been sampled on small geographic scales (e.g. < 500 km). These populations exhibited high levels of divergence indicative of limited gene flow despite the high physical connectivity between sites (Rynearson, 2004; Godhe and Härnström 2010). In contrast, little population structure and high gene flow were observed among samples of the pennate diatom species *Pseudo-nitzschia pungens* in the Southern Bight of the North Sea (Casteleyn 2009a).

The characteristics of diatom population structure over global spatial scales are virtually unknown. For marine algae, it has been widely argued that high dispersal may eliminate genetic distance as a barrier to genetic connectivity (Finlay 2002; Foissner 2006; Kinlan 2003; Medlin 2007). Only a single study has examined global population structure in a diatom species. In this study, *Pseudo-nitzschia pungens* populations exhibited evidence for allopatric isolation, where geographic distance was suspected to be a strong barrier to gene flow (Casteleyn 2010). The pattern of allopatric isolation in this pennate species is in contrast to patterns that one might predict in the centric diatoms examined thus far, which exhibit high levels of population differentiation among sites experiencing high physical connectivity and separated by small geographic distances (Godhe and Härnström2010; Härnström et al. 2011; Rynearson 2006; Rynearson et al. 2009). Differences in life history and biogeography between pennate and centric diatoms, or even among individual diatom species, may result in their different patterns of population structure. Thus, it remains unclear whether or not geographic limits to gene flow, as found in *P. pungens*, are common to diatoms.

In general, diatom studies have relied largely on samples from culture collections, or have pooled samples from regions collected over many years and across seasons. For *P. pungens*, evidence for genetic stability over time justified the pooling of samples over seasons and years (Casteleyn 2010). The pooling of samples over long periods of time and across seasons may not be appropriate for other species, such as *D. brightwellii* or *S. marinoi,* which exhibit significantly diverged populations that exist within the same local or regional environment. While often unavoidable, pooling

samples affects the interpretation of results by preventing our ability to tease apart the influence of time and the environment on popultion structure.

Molecular tools targeting population-level diversity offer a unique opportunity to explore the relative influence of environment, ecology, and dispersal in structuring diatom populations over large spatial scales. Here, we used microsatellite markers to identify genetically distinct populations within the globally distributed bloom-forming diatom species, *Thalassiosira rotula*. Global samples captured bloom periods in the northern and southern hemispheres over a single year, providing a 'snapshot' of global population structure in this species. Analyzing discrete samples across the bloom period allowed for a better understanding of both spatial and temporal variation in genetic composition. The relative importance of dispersal and environmental selection in driving genetic isolation can be inferred by correlating the occurrence of genetically distinct populations with environmental conditions from where they were sampled. The goal of this work was to determine population-genetic structure across global geographic space in order to better understand those factors of the ocean system that may influence the generation of diversity and distribution of populations throughout the globe, and over time.

METHODS

Field samples and isolates

Four-hundred forty-nine isolates of *T. rotula* were collected throughout 2010 as part of a simultaneous global sampling effort targeting twenty locations throughout the Atlantic and Pacific Ocean basins, and across hemispheres (Table1, Figure 1). The majority of sampling sites were chosen based on their alignment with established time

series and thus contained significant metadata (Table 2). For sites not associated with time series measurements, ancillary data (e.g. temperature and salinity) was collected via in-situ sensors (e.g. YSI, Xylem). Whole seawater was rush-shipped from these twenty locations to the University of Rhode Island, USA at one to five time points per site throughout the year, targeting specifically the winter-spring phytoplankton bloom periods in the northern and southern hemispheres.

Whole seawater was shipped in 1 L dark Nalgene bottles in small coolers containing ice packs and sample information. In seawater samples where *T. rotula* could be identified morphologically, single cells or chains were isolated, cultures maintained, and DNA extracted according to Whittaker *et al.* (Whittaker et al. 2012). In brief, surface seawater was concentrated over a 20 μm mesh net. For each field sample where *T. rotula* was present, up to 96 single cells or chains were isolated from this >20μm size fraction using a stereo microscope (Olympus SZ61), rinsed in sterile seawater three times, and transferred to 1 mL sterile Sargasso seawater amended with f/20 nutrients (Guillard 1975). Isolates were incubated in either at 4, 10, 14, or 20°C, according to closest surface seawater temperature (SST) at isolation, and under a 12:12-h light:dark cycle of 90-120 μmol photons m^{-2} s^{-1}. Upon reaching approximately 1000 cells/mL (1 to 3 weeks depending on growth rate), and upon confirmation that the cultures were free of algal contamination, cells were filtered and DNA extracted.

In addition to the collection of field isolates, whole seawater samples were processed for community analysis in two ways. First, concentrated samples were fixed using a 2% acid Lugol's solution, and stored in 20 ml scintillation vials. Second,

whole seawater was filtered onto three 0.2 μm polyester filters (Millipore), at 100 mls per filter, and stored at -80°C.

Because shipping time took anywhere from 1 to 4 days, a simple experiment was conducted to test whether or not genetic diversity in a 1 L bottle changed due to the shipping process. Two liters of whole seawater were collected from Narragansett Bay on January 26, 2010 (Table 2). One liter was immediately processed according to the procedure outline above, and 48 single cells isolated. The second liter was shipped to Washington, USA and back to Rhode Island, USA, at a transit time of four days (simulating the longest transit time for any sample). Upon returning to Rhode Island, 48 single cells were isolated from the shipped sample. DNA extracted from these isolates was genotyped and analyzed based on the procedures described below.

Microsatellite discovery and optimization

Six microsatellite loci (TR1, TR3, TR7, TR10, TR8, TR27) were isolated from the *T. rotula* genome using two methods. Loci TR1 and TR3 were isolated using the microsatellite enrichment method described in Hamilton *et al.* (Hamilton et al. 1999) and adapted by Spies *et al.* (Spies et al. 2005) wherein a single step is used to both digest genomic DNA and ligate a universal 'SNX' linker to resulting DNA fragments. Genomic DNA (gDNA) of the National Culture of Marine Algae (NCMA) culture strain CCMP3096 was extracted using the QIAGEN DNeasy Plant Extraction Midi kit according to the manufacturer's protocol. Approximately 1 ug of genomic DNA (gDNA) was digested using restriction enzymes *HincII* and *XmnI* (NEB, New England Biolabs), and SNX linkers added using the reaction and temperature profile described in Spies *et al*. (Spies et al. 2005). Microsatellites were enriched using two biotin-

labeled oligonicleotide probes (GACA) 4 and (GA) 4 immobilized on Dynabeads (Dynal Biotech, Inc.) From the bead hybridization products of each probe, polymerase chain reaction (PCR) was used to amplify DNA enriched for microsatellites; this product was ligated into the pCR2.1 vector, and cloned into Top10 one-shot chemically competent *E.coli* cells using the TOPO TA Cloning Kit (Invitrogen) according to the manufacturer's instructions. Hundreds of colonies were screened first using an M13 PCR reaction to confirm successful transformation followed by plasmid amplification via the Illustra TempliPhi 100 Amplification Kit (GE Healthcare Life Sciences) and sequencing on an ABI 3130*xl* (Applied Biosystems). Microsatellites were identified from sequences using SeqMan II (DNAstar, Inc.) software, and selected for downstream analysis based on quality of repeat (greater than 10 uninterrupted repeats) and adequate flanking region for primer design.

The microsatellite discovery method described in Abdelkrim et al., 2009 (Abdelkrim et al. 2009) was used to isolate the remaining four loci (TR7, TR8, TR10, TR27) from 454 sequencing of gDNA of NCMA strain CCM3096. One eighth of a plate of Roche GS-FLX Titanium sequencing was performed on 5 ug of gDNA extracted from CCMP3096 using the QIAGEN DNeasy Plant Extraction Midi kit. Msatcommander software (Faircloth 2008) was used to digitally identify repeat regions within the resulting contig and raw-read 454 datasets using the following search parameters: mono to hexa-nucleic acid repeat lengths of 10 or greater, and enough flanking region to allow for primer design at least 50 bp away from the microsatellite region. This search resulted in the discovery of 862 microsatellite loci,

over 200 of which contained enough flanking region to design primers. Thirty of these loci were chosen for downstream marker development and optimization based on 'clean' repeat quality (e.g., no repeat interruptions), and primers designed in flanking regions free of any repeat motif.

Optimization and verification of all 34 microsatellite markers (four isolated via bead hybridization enrichment, and 30 isolated via 454 sequencing) involved the following steps: To check for amplification success and allelic variability, microsatellite loci were amplified in DNA extracted from eight strains of *T. rotula* either purchased from the NCMA culture collection, formerly Culture Collection of Marine Plankton (CCMP3264, 1647, 3096), Culture Collection of Algae and Protozoa, UK (CCAP1085_21,1085_20), or maintained in the laboratory and cited in Whittaker et al., 2012 (P17F4, and NB12A2). These strains were chosen based on previous knowledge of their genetic and physiological variability (Whittaker et al. 2012). Microsatellites were amplified initially using un-labeled forward and reverse primers (Table 3), and the following reaction mixture: 2.5 mM $MgCl_2$, 0.2 mM dNTP mix (Bioline), 0.4 µM forward primer, 0.4 µM reverse primer, 0.2 mg/mL BSA (Bioline), and 0.03U/µl Mango Taq Polymerase (Bioline). The thermocycler profile was as follows, with annealing temperature (Ta) set initially at 2°C below the lowest predicted melting temperature (Tm) for each primer set: 2 min denaturation step at 94°C, 30 cycles of 94°C for 20 s, Ta for 20 s, and 72°C for 30 s followed by a 10 min extension at 72°C. PCR product was visualized on a 6% polyacrylamide gel, and examined for amplification success, size validation, and allelic variability across strains. Amplification yield and specificity were optimized by testing gradients of

annealing temperature, primer and MgCl2 concentrations, and diluting the DNA template reflecting the DNA concentration of field isolates. Once variability and amplification success was confirmed, fluorescent primers were used to amplify PCR product for downstream fragment analysis. Loci were fluorescently labeled with 6FAM, NED, or VIC (Table 3). Markers were further optimized for fragment analysis, resulting in the final PCR conditions used to genotype field isolates (Table 3). Six loci were chosen for downstream analysis of field isolates and amplicons sequenced to confirm identity.

T-RFLP Screening of Field DNA

Species within the *Thalassiosira* genus are difficult to identify microscopically. To verify the species-level identification of *T. rotula* field isolates, a terminal restriction length polymorphism (T-RFLP) screening protocol was developed, based on the polymorphisms occurring at the internal transcribed spacer 1 region (ITS1) among *Thalassiosira* species. First, the ITS1 was amplified using *Thalassiosira*-specific primers: TRscreenF-TTA TCA CAC CAT CCA AAC CTT A and TRscreenR-6FAM - TTA GGG ACA CTC AGA TAA GCA; the reverse primer was modified with a 6FAM fluorescent tag on the 5' end. The restriction enzyme *BsrI* (NEB, New England Biolabs) was used to digest amplicons using the following thermal profile: 65°C for 1.5 hours, with an 85°C denaturation for 10 min. Due to polymorphisms that exist among *Thalassiosira* species at the ITS1, a *BsrI* differentially digests amplicons from each species, resulting in fluorescently labeled fragments that differ in size among *Thalassiosira* species (Figure 2). Digested amplicons were analyzed using fragment analysis performed on an ABI 3130xl. T-

RFLP screening was performed on all isolates collected, and a subset of results were confirmed via ITS1 sequencing.

Amplification of microsatellite and ITS1 regions in field DNA

To quantify genetic variation among isolates at both the population and lineage levels, isolates were genotyped at six microsatellite loci, and sequenced at the ITS1. Six loci were amplified from the gDNA of 449 *T. rotula* isolates using the following PCR reactions: Locus TR27 was amplified using the three-primer method described in Blacket et al., 2012 (Blacket et al. 2012). This method involves the addition of a tag on the forward primer that anneals to a fluorescent probe with a complementary tag sequence, GCCTTGCCAGCCCGC with a VIC 5' label. The reaction mixture is as follows: ~3 ng gDNA, 1X Mango Buffer (Bioline), 0.25 mM dNTP mix (Bioline), 2.50 mM $MgCl_2$, 0.08 µM forward (tailed) primer, 0.25 µM reverse primer, 0.10 µM fluorescent tag, H_2O, and 0.03 U/µl Mango Taq Polymerase (Bioline). Loci TR1 and TR3 were multiplexed with the following reaction: 1X Mango Buffer (Bioline), 2.5 mM $MgCl_2$, 0.2 mM dNTP mix (Bioline), 0.2 µM TR1F primer, 0.2 µM TR1R primer, 0.3 µM TR3F primer, 0.3 µM TR3R primer, 0.2 mg mL^{-1} BSA, and 0.03U μL^{-1} Mango Taq Polymerase (Bioline). TR8 and TR10 were multiplexed using the following reaction: 1X Mango Buffer (Bioline), 2.5 mM $MgCl_2$, 0.2mM dNTP mix (Bioline), 0.3 µM TR8F primer, 0.3 µM TR8R primer, 0.3 µM TR10F primer, 0.3 µM TR10R primer, 0.2 mg mL^{-1} BSA, and 0.03 U μL^{-1} Mango Taq Polymerase (Bioline). TR7 was amplified using the following reaction: 2.5 mM $MgCl_2$, 0.2 mM dNTP mix (Bioline), 0.4 µM TR7 or TR2F primer, 0.4 µM TR7 or TR2R primer, 0.2 mg mL^{-1} BSA, and 0.03 U μl^{-1} Mango Taq Polymerase (Bioline). Thermal profiles of all

microsatellite reactions consisted of an initial 5min denaturation step at 94°C, followed by a set of 30-35 cycles with annealing conditions optimized to each primer pair with fluorescent labels (Table 3), a 10 cycle probe-annealing step for TR27 only consisting of 94°C for 20 s, 54°C/61°C for 20s and 72°C for 30s, and finally (for all loci) a 10 min extension at 72°C. Alleles were scored using an ABI 3130*xl* (TR1,TR3, TR7) or an ABI 3730*xl* (TR8, TR10, TR27), and analyzed using the software Gene Mapper 5 ® (Life Technologies).

To amplify the ITS1 rDNA region, a reaction mixture containing ~5ng DNA, 0.1 mmol L^{-1} dNTPs (Bioline), 0.05 U μL^{-1} Bio X Act DNA polymerase (Bioline), 1X buffer (Bioline), and 0.5 μmol L^{-1} each of primers 1645F (Rynearson 2006) and RotIIR (Whittaker et al. 2012). The PCR consisted of a 2 min denaturation step at 95°C, 36 cycles of 94°C for 30 s, 62°C for 30 s, and 72°C for 1 min followed by a 10 min extension at 72°C. The ITS1 was amplified from 350 isolates, targeting a subset of each field sample. All sequencing was performed on an ABI 3130 *xl* (Applied Biosystems), and sequences are available in Genbank.

Analysis of microsatellite alleles

Summary statistics for each locus and each site were generated using the Excel Microsatellite Toolkit v. 3.1.1 (Park 2001), and GenePop v4.2 (Raymond and Rousset 1995). These statistics included observed and expected heterozygosity for each locus, and globally, as well as polymorphism information content (PIC) values for each locus and globally. The number of alleles per locus was also calculated. The number of unique genotypes among all individuals (G:N) for each site was calculated by identifying matching multilocus genotypes (MLGs) in the Excel Microsatellite Toolkit

v. 3.1.1 (Park 2001). Deviations from Hardy-Weinberg equilibrium (HWE) were calculated in Genepop v4.2 (Raymond and Rousset 1995), and inbreeding coefficient F_{IS} calculated in GenAlEx 6.5(Peakall and Smouse 2012). P-values were adjusted to account for multiple tests via the Bonferroni technique (Hochberg 1988).

Statistical analysis of population structure

Using microsatellite genotypes and allele frequencies, the presence of hierarchical population structure could be tested among samples over space and time. The significance of genetic differentiation among sites was determined using the exact G test and Fisher's exact probability test (Raymond 1995) in GENEPOP v4.2 (Raymond and Rousset 1995). The extent of pairwise differentiation between sites, $F_{ST,}$ was calculated in GenAlEx 6.5 (Peakall and Smouse 2012). Pairwise tests between sample sites were conducted separately for each locus, as well as among all loci. All p-values were Bonferroni corrected. Analysis of molecular variance (AMOVA) was calculated using GenAlEx 6.5, testing for the presence of genetic structure at global and basin scales. Principal Coordinates Analysis (PCoA) was used to ordinate pairwise F_{ST} in 2D space among sites with sample size greater than 10 individuals. The Bayesian clustering program STRUCTURE v2.2 (Pritchard et al. 2003) was used to examine hierarchical resolution of population structure in the global dataset. STRUCTURE is a Bayesian analysis that inferrs the presence of distinct populations, assigns individuals to populations, and estimates population allele frequencies in situations where many individuals are migrants or admixed. In the STRUCTURE program, populations (K) from 2 to 13 were tested in 10 independent runs using the admixture model and correlated allele frequencies, with locations as

prior, a burn-in of 100,000, and 100,000 repetitions. A maximum K of 13 was tested to reflect the total number of samples. Estimates of the number of populations (K) were based on Pritchard *et al*. (Pritchard et al. 2000) and Evanno *et al.* (Evanno et al. 2005). StructureHarvester v.6.93 (Earl 2012) was used to determine the optimal K, and generate input files for downstream analyses. The Greedy algorithm in CLUMPP v.1.11 (Jensen et al. 2005) was used to evaluate agreement between independent STRUCTURE runs, and to arrange cluster labels. DISTRUCT v1.1 (Rosenberg 2004) was used to visualize results of STRUCTURE.

Phylogenetic Analysis

To determine genetic structure and the extent of genetic diversity on the lineage level, 257 ITS1 sequences were sequenced and analyzed following the methods described in Whittaker *et al.* (Whittaker et al. 2012). SeqMan II 3.61 (DNASTAR, Inc.) was used to curate raw sequencing data, and sequences were aligned using Clustal W (Thompson 1994) in Mega4 (Tamura et al. 2007). The boundaries of the ITS1 were determined through alignment with Genbank accession EF208798, and *T. rotula* sequence types compared with previously identified ITS1 lineages within the species *T. rotula* (Whittaker et al. 2012). To examine relationships among *T. rotula* ITS1 variants, a network was generated using the median joining algorithm in Network 4.5.1.6 (Fluxus Technology Ltd.).

Seascape Genetics

Correlations between pair-wise genetic distance and pair-wise Euclidean distance of environmental variables among sites were examined. These calculations relied upon extensive metadata associated with each sample. Metadata were

generously provided by the institutions responsible for the time series associated with each site, or by individuals who collected the samples (Table 2). Parameters examined included the following: location, sea surface temperature (SST), salinity, chlorophyll *a*, and abundance of *T. rotula* in each sample. *Thalassiosira rotula* cell abundance was calculated from Lugol's fixed samples using an E800 microscope at 20X (Nikon), wherein triplicate counts of *T. rotula* were performed from each sample using a 1mL Sedgewick-Rafter microscope slide.

A series of Mantel tests was performed. Isolation by distance (Mcrae 2006) (IBD) was examined using the software IBDWS v3.16 (Jensen et al. 2005). The IBD analysis assessed correlations between pairwise F_{ST} and geographic distance between sites. Pairwise geographic distances between sites were measured as the 'shortest ocean path,' or shortest continuous connection across the sea, as calculated in Google Earth. In the software Primer-E, BIOENV (Clarke and Ainsworth 1993) was performed to identify those factors, or combination of factors, of the ocean environment most correlated with a pairwise F_{ST} resemblance matrix. BIOENV relies upon matrices calculated as the Euclidean distance between sites based on every combination of environmental variables, in this case, temperature, salinity, chlorophyll *a*, and cell abundance. It tests the correlation between these environment-based matrices with genetic distance matrices (F_{ST}) among sites to find the combination of variables that best correlates with pairwise genetic distances, providing a Rho value to demonstrate the strength of the correlation.

RESULTS

Six microsatellite markers were developed and optimized for *T. rotula*. Markers TR1 and TR3 were developed via bead hybridization. The remaining 4 loci were developed via 454 sequencing which generated more than 128k reads in approximately 45 Mb of sequencing data. The median length read for the run was 386 bp. Twenty-six percent of the reads were fully assembled into contigs greater than 500 bp. The largest contigs were over 28,000 bp long. A search for repeat motifs over 10 repeats long recovered over 800 loci from the dataset; 231 contained flanking sequences long enough for primer design. Through optimization tests on 30 of these loci, six demonstrated the amplification success and variability required for downstream genotyping (Table 3). Polymorphism information content (PIC) was highest for locus TR10, at 0.94, as averaged across all global isolates genotyped. PIC values for the remaining five loci across all global isolates were 0.87, 0.58, 0.69, 0.47, and 0.79 for TR8, TR1, TR3, TR7, and TR27, respectively.

Population diversity of *T. rotula*

To test effect of long-distance shipping on *T. rotul*a genetic diversity and population structure, two samples were collected simultaneously from Narragansett Bay. One sample was processed immediately and the other was shipped in a cooler for 4 days (samples NBa and NBb, respectively). NBa and NBb did not differ in their isolation success rates (both at 100%), but differences in sample size reflect differences in the number of positively identified *T. rotula* during isolation. This suggests that cells in shipped sample were healthy, as the probability of successful isolation (growth) remained the same in NBa and BBb. In those isolates that survived,

and were positively ID as *T.rotula,* diversity statistics did not differ significantly between shipped and immediately processed samples. This suggests that selection for certain genotypes was not occurring in the sample during shipping, and those cells that survived represented the diversity of the sample as if it had been immediately processed. In fact, diversity statistics for NBa and NBb were consistent for expected and observed heterozygosity, and F_{ST} values desmonstrated that NBa and NBb were not diverged. For example, expected heterozygosity was 0.683 and 0.687 for NBa and NBb, respectively. Observed heterozygosity was 0.521 and 0.556 for NBa and NBb respectively. On average, NBa and NBb showed no significant divergence (F_{ST}) based on genotypic richness and allele frequency distribution, demonstrating that shipping samples at a transit time of four days resulted in no effect on population diversity, and did not introduce any bias in genetic diversity or population structure.

In total, 449 *T. rotula* isolates were obtained from across the globe at eight locations, and thirteen discrete time points. The species identity of all isolates was confirmed via T-RFLP screening (Figure 2), 257 of which were further confirmed via ITS1 sequencing (Table 2, Figure 1). Isolation was unsuccessful for one sample from Iceland, wherein 100% of cells died either due to incompatible isolation conditions, or viral lysis. For all other samples, isolation success rate was high; of all cells isolated, on average 96% survived.

In total, 439 alleles were identified in 449 global isolates of *T. rotula*, or 73 alleles per locus on average. The average number of alleles per locus per site was 15.85. *Thalassiosira rotula* exhibited extremely high levels of clonal diversity; only four identical six-locus genotypes were observed more than once in 449 individuals

genotyped. Samples also exhibited high levels of genetic diversity, as measured by expected and observed heterozygosity. Across all loci, expected heterozygosity was as high as 0.83, and even higher for individual loci (Figure 3, Tables 3, 4 and 5). Observed heterozygosity was lower than expected for all sites, at all loci (Table 4, table 5). In fact, significant departures from Hardy-Weinberg equilibrium in the form of heterozygote deficiencies were observed for most site and locus combinations (Table 4). The lowest observed heterozygosity was detected in California (0.44), and the lowest expected heterozygosity was detected in New Zealand (0.55). The highest observed heterozygosity was detected in Puget Sound, WA (0.61), and the highest expected heterozygosity was detected on the coast of Washington, USA (0.83).

Diversity statsistics were compared between ocean basins. Measures of allelic richness and heterozygosity did not differ between Atlantic and Pacific Ocean basins, as determined by a one-way ANOVA tests ($p= 0.703$, and $p=0.985$, respectively).

Ribosomal ITS1 diversity of *T.rotula*

Analysis of 257 ITS1 sequences revealed two distinct lineages, corresponding to lineages 1 and 3 previously detected in global *T. rotula* isolates (Whittaker et al. 2012) (Figure 4). Lineage 1 represented isolates sampled predominantly from the Eastern Pacific, whereas lineage 3 represented isolates sampled across ocean basins, including the Western Pacific, Atlantic, Indian Ocean, and Eastern Pacific. Lineage 1 was not restricted to the Pacific, and observed in the Atlantic in seven individuals from Narragansett Bay and France. Lineage 3 was not restricted to the Atlantic, as it was observed in 10 individuals distributed across all Pacific locations studied. Lineage 2,

detected previously in *T. rotula* isolates from Puget Sound, was not identified in these 257 isolates.

Global population structure

Throught this paper, a 'population' is defined as any group of samples that exhibit significant levels of divergence (F_{ST}) based on Fisher's exact test, and measures of allele frequency and heterozygosity. This general definition differs from populations identified using the software STRUCTURE (K), which were determined using a Bayesian clustering algorithm. STRUCTURE 'populations,' or clusters, will be referenced in the discussion of STRUCTURE analysis output.

Tests of allelic and genotypic differentiation identified five significantly diverged populations within *T. rotula*. Measurements of F_{ST} ranged from 0.009 to 0.139 (Table 6). We tested whether microsatellites showed the same deliniation as ITS1 sequences using Principal coordinates analysis (PCoA); this test showed no differentiation between lineages when individuals from lineages 1 and 3 were analyzed together (Figure 5). Because of this, we chose not to analyze population structure separately for lineages 1 and 3. PCoA axis 1 and 2 explained 44% and 20% of variation, respectively (Figure 5). AMOVA demonstrated that populations were not significantly structured across ocean basins (Table 7). Genetic variability was significantly structured among samples, among individuals, and within individuals (p=0.001) (Table 7).

Not all isolates from each water sample were equally or significantly diverged (Table 6). For example, Puget Sound, WA and South Africa, and NBa/b were three samples (with sample sizes >10) that exhibited statistically significant divergence

from all other sites analyzed. New Zealand was also significantly diverged from all other samples, but this result should be treated with caution due to low number of isolates (n=8). Population structure over geographic space was not spatially intuitive. For instance, a significant population included FRa, and NBc from the Atlantic Basin, and WAa and Cali from the Eastern Pacific excluding other isolates collected from France and Narragansett Bay at different times throughout the sampling period. This also demonstrated that samples collected at different time points within France, Narragansett Bay, and Washington were, at times, more closely related to individuals collected from other global sites than to individuals collected at a later sampling time at the same location. In fact, F_{ST} values among samples collected at the same site were 60% as great as those among sites separated by distances of 15,000 km.

STRUCTURE analysis was able to assign *T. rotula* individuals to different clusters (Figure 6). The optimal K was estimated based on the Evanno method, which uses ΔK, or the rate of change of the log probability of data between successive *K* values (Evanno et al. 2005) (Figure 7). The highest ΔK was observed at K=11 (Figure 7a. and b.). However, due to a marked increase in the standard deviation of L(K) after K=4, and because it was associated with the second highest ΔK, (K=4) was also chosen as the most likely number of populations in this study (Figure 7a. and b.). Choosing K=4 also allowed for clearer interpretation of STRUCTURE results, as the membership coefficient of individuals into a single STRUCTURE group was generally higher when K=4 than when K=11 was used, allowing for better understanding of the hierarchical clustering of individuals. In general, STRUCTURE clustering of K=4 agreed with PCoA groupings. NBa and NBb showed high

membership in the 'red' population. SA formed a unique group, with high membership into a 'blue' population, exhibiting little admixture with other STRUCTURE populations. Puget Sound also showed a unique pattern, with almost exclusive membership into a 'green' population. NBd, FRb, WAb, and Helg showed similar patterns of mixed membership among green, blue, and red populations. Finally, NBc, FRa, Waa, and Cali showed similar patterns of admixture between the red, green, and yellow STRUCTURE clusters. Isolation by distance measures identified a weak relationship between genetic distance and shortest ocean distance, or shortest continuous ocean distance between sites ($p=0.007$, $r^2=0.20$); the result was significant, but the r^2 value was so low as to be uninformative (Figure 8). This relationship was not observed when samples from either ocean basin were analyzed separately (data not shown).

Environmental variation and genetic structure

Environmental and ecological variables (salinity, temperature, chlorophyll *a*, and cell abundance) differed considerably across sample sites and over time. For instance, *T. rotula* cell abundance ranged from 100 cells L^{-1} to 55,000 cells L^{-1}. At a single site, Narragansett Bay, cell abundance increased from ~2,000 cells/L to over 55,000 cells/L over the course of three sampling periods distributed over two months (Figure 9). Environmental conditions varied widely among samples. Temperature of sample collection ranged from 1°C to 14.14°C. Chlorophyll *a* concentration ranged from 0.1 to 22.64 $\mu g\ L^{-1}$. Salinity ranged from 27.44 to 32.02 ppt. A series of Mantel tests revealed significant correlations between environmental variables (salinity, temperature, chlorophyll *a*, and *T. rotula* cell abundance) and pairwise genetic

distance (F_{ST}) among isolates (Table 6, Figure 10). The combination of temperature and chlorophyll *a* concentration explained 53% of the variance in global population structure ($p = 0.02$). Temperature alone explained 27%, and chlorophyll *a* concentration alone explained 36% of the variance in global population structure (Table 8).

DISCUSSION

This work, the first to empirically measure population structure in a diatom species across global space and over time, revealed several unexpected results: First, the scale of genetic divergence over time in a single location for *T. rotula* was 60% as large as the genetic divergence observed over the largest distances sampled (15,000 km). Second, geographic distance showed only a weak relationship to genetic connectivity. This suggests that geographic distance in *T. rotula* is not a strong barrier to gene flow given the time scale of global circulation and potential for dispersal. Third, environmental selection was demonstrated to be a force structuring populations across space and time.

Variability over time

Samples were collected at multiple time points at three sites: Narragansett Bay, RI, USA, Roscoff, France, and the coast of Washington, USA. Over the course of two months, we observed three significantly diverged populations in Narragansett Bay, and two significantly diverged populations in both Roscoff, France and off of the coasts of Washington, USA. The detection of significantly diverged populations in samples collected over short periods of time contrasts with the population stability over 100 years observed in *S. marinoi* in a Danish fjord (Härnström et al. 2011). This

contrast may be due to differences in environmental variability and physical dynamics of the study sites. For example, the Mariager fjord in Denmark is associated with a relatively lower flushing rate than observed in Narragansett Bay, RI coastal Roscoff and Washington. The residence time of Mariager fjord is eight months (Härnström et al. 2011). This is in contrast to Narragansett Bay, which is associated with a very short residence time between 10 and 40 days (Pilson 1985a; Pilson 1985b). Roscoff is a coastal site within the Western English Channel, experiences high connectivity to both the North Sea and Atlantic water masses, and is dominated by persistent tidal flushing (Robinson et al. 1986; Southward et al. 2004). The Washington coast is a region that experiences dynamic upwelling, and our sample site was associated with a highly exposed coast versus the more protected fjord site. Narragansett Bay, France, and Washington experience high levels of physical connectivity to surrounding waters, characteristics that may support the population differentiation over time observed here. Different populations may be introduced from connected waters, and be specifically adapted and tightly coupled with environmental conditions that can vary widely over short periods of time.

The observation of significantly diverged populations over short periods of time in *T. rotula* may also relate to characteristics of the life history of the species. For instance, *T. rotula* and other diatoms like *S. marinoi* form resting spores, or dormant cells that remain viable for many decades in the sediment (Mcquoid and Hobson 1995; Mcquoid and Hobson 1996). The triggers of resting spore formation and germination are little understood for many species, but relatively well understood for *S. marinoi*; it has been hypothesized that the accumulation of cysts in the sediment

may support the persistence of a single fjord population, preventing the introduction of new populations to the system (Godhe and Härnström2010; Härnström et al. 2011). Although *T. rotula* is known to create resting spores, the success of *T. rotula* resting spore germination, triggers of spore formation, and level of accumulation in different sediments is virtually unknown. Resting spore dynamics, and the diversity of spores in sediments throughout the globe, may influence the variability of population structure over time. Hypothetically, sediments may harbor high levels of resting spores, and act as a 'seed bank' for diverse populations whose accumulation in the water column is triggered by favorable environmental conditions (Garrison 1981; Godhe and Härnström 2010; Mcquoid and Hobson 1995).

The scale of divergence over time observed in *T. rotula* is similar to that observed in other diatom species. For the diatom *D. brightwellii* in the Puget Sound fjord of the Eastern Pacific, populations significantly diverged by F_{ST} values of 0.19 have been consistently observed during different times of the year. Within Narragansett Bay alone, samples separated over three months were associated with three populations diverged by a maximum F_{ST} of 0.064. Within two samples from France collected over the course of one month, two populations were observed, significantly diverged by an F_{ST} of 0.023. Washington samples were separated by a significant F_{ST} of 0.028. This is first time that temporal divergence among diatom populations has been compared with geographic divergence over a global spatial scale. However, evidence suggests that other diatom species may also exhibit population divergence and variability over time at the scale observed over global geographic space.

Due to the difficulty of obtaining diatom samples, individuals from culture collections or random sampling events have traditionally been pooled to represent locations of interest, despite sampling times being separated by decades, or distributed across seasons. The act of pooling samples over time or regional space relies on the assumption that spatial variability is significantly greater than variability over time. Evidence from *T. rotula* strongly refutes this assumption. Pooling samples over many decades may ignore the large variations in genetic composition embedded in time, and lead to false conclusions about the structure of populations, and factors driving that structure. The methods used here provide a viable option for obtaining simultaneous (non-pooled) samples over large spatial scales, because shipping time (4 days) was shown to have no effect on *T. rotula* population structure. By analyzing discrete samples separately, I was able to examine changes in genetic composition over time. Pooling samples over space and time may result in erroneous clustering of individuals into *a priori* groups, leading to assumptions about genetic divergence that mask underlying genetic substructure; this may result in false measurements of heterozygosity, such as those described by the Wahlund effect (Wahlund 1928). Because so little is known in regards to the spatial and temporal variability of diatoms, pooling of samples should be avoided wherever possible.

Dispersal and Geography

Phylogeographic inference from the ribosomal ITS1 gene has revealed that *T. rotula* can be subdivided into three lineages that exhibit differences in their geographic distributions (Whittaker et al. 2012). Lineage 1 was observed predominantly in the Eastern Pacific, Lineage 2 within Puget Sound, and Lineage 3 in both both the

Atlantic and Pacific Ocean basins. A molecular clock estimated the divergence time between Lineage 1 and Lineage 3 at around 0.7 Mya. This period is similar to the split between *P. pungens* lineages, which also exhibit differences in geographic distribution, one being cosmopolitan and the other restricted to the NE Pacific (Casteleyn, 2008). It has been hypothesized that punctuated glaciation, and fluctuations in sea level during the Pleistocene contributed to this pattern of lineage divergence, and the associated geographic structure in *P. pungens* (Casteleyn 2010).

Ribosomal ITS1 sequencing of some individuals sampled in this study revealed that *T. rotula* lineages do not adhere strictly to the geographic ranges previously identified. Whereas previous research showed clear differences in the geographic distribution of lineages, the increase in sample size in this study (257 ITS1 sequences, vs. 96 previously) allowed for detection of lower frequency haplotypes that violate these geographic ranges. For instance, Lineage 2, previously detected only in the Eastern Pacific, was observed in seven samples from the Atlantic. Lineage 3, previously observed in only a single site in the Eastern Pacific, was detected in all Eastern Pacific sites sampled here. Detection of Lineage 2 in Atlantic samples suggests a possible hybridization of lineages, present despite the concerted evolution acting on the ITS1 marker, as well as the multi-copy nature of the gene that may mask intra-individual variability (Baldwin 1992). In fact, these same characteristics of the ITS1 may mask lineage hybrids among individuals studied here. High frequency of reticulate evolution, or hybridization, is common in plants (Arnold and Fogarty 2009; Linder and Rieseberg 2004; Rieseberg and Wendel 1993). Reticulate evolution is also known to influence the genetic and even morphological variation of corals (Arnold

and Fogarty 2009). The observation of geographic mixing among *T. rotula* lineages suggests that hybridization may be facilitated by current ocean conditions. Thus, *T. rotula* lineages may reflect historical subdivisions within the species rather than contemporary patterns of gene flow among them. In fact, global population structure demonstrated by *T. rotula* shows no correlation with lineage-level subdivisions of the species, pointing to contemporary gene flow among them. These data confirm that *T. rotula* lineages retain the ability to interbreed.

Interestingly, distance was only a weak barrier to dispersal over global space, explaining only 20% of genetic structure. No isolation by distance was observed on basin scales. In addition, contemporary population structure revealed by microsatellites demonstrate that sites separated by large distances, and across ocean basins such as the Eastern Pacific and Western Atlantic may be more related than individuals from within the same basin over the winter/spring bloom period. These data suggest that the time scale of global circulation (~1,000 years) is sufficient to allow for the gene flow between individuals from distant sites, despite their vast spatial extent. Finally, these results suggest that the dispersal capacity of *T. rotula* is sufficient to permit global connectivity.

Three samples were significantly diverged from all other samples: Puget Sound, South Africa, and one sample from Narrgansett Bay. These three samples were collected under unique environmental or hydrographic conditions. For Narragansett Bay, the high-chlorophyll environment under which the samples were collected may have led to conditions of elevated competitive exclusion for which that single population was uniquely adapted. As the only fjord site sampled, Puget Sound

represents a location of higher water retention and physical isolation than the other coastal and estuarine sites examined. Two other species, *D. brightwellii* and *S. marinoi,* have also demonstrated that isolates collected from fjord environments exhibit significant genetic divergence from nearby coastal waters (Rynearson, 2004; Härnström, 2011). South Africa was the only Indian Ocean site sampled, and there exists no comparison for phytoplankton population structure in this region. STRUCTURE analysis demonstrated that membership of South African isolates was high for a single population with relatively little admixture. The unique oceanographic locations of Puget Sound (a fjord) and the Indian Ocean suggest that differences in circulation patterns may be responsible for driving the genetic isolation of populations detected there. The weak correlation between genetic distance and geographic distance observed in this study suggests that other physical properties of the ocean environment (e.g., currents, residence time) may better define 'distance' in the ocean. The coupling between ocean hydrodynamics and genetic structure of diatoms is an appropriate direction for future study.

The weak correlation between geographic and genetic distance in *T. rotula* conflicts with the observation that geographic distance limits gene flow among global *P. pungens* populations. For instance, isolation by distance (IBD) r^2 was 0.75 for *P. pungens*, but only 0.20 for *T. rotula*. This contrast may relate to distinct life history differences between the species. For one, *P. pungens,* and many other pennate diatoms, are heterothallic, meaning that male and female mating types must be present for successful sexual reproduction. Centric diatoms are more typically homothallic, although a diversity of mating systems have evolved across diatom species. Overall,

the timing and triggers of sexual reproduction in diatoms are little understood (Von Dassow and Montresor 2011). However, differences in environmental or temporal triggers of sex between the two species may greatly impact patterns of population structure across the wide ecological gradients sampled. The timing and triggers of programmed cell death (PCD) may be an additional source of variation between the two species, dictating their dispersal capacity over broad geographic space. PCD can occur as a response to adverse environmental conditions, but also occur as a normal step in the cell life cycle (Bidle and Bender 2008; Von Dassow and Montresor 2011). Although triggers of PCD in either species have not been examined, a greater tendency towards PCD in *P. pungens* under adverse conditions could prevent their long-distance dispersal, making distance a greater barrier to gene flow.

A more likely explanation of differences between the species' population structure may be differences in their ability to form resting spores, or dormant cells, that may provide a stepping-stone for dispersal in those species that can form them. Life stages of cell quiescence and resting spore formation may affect the ability for cells to traverse large distances of the ocean. Resting spores may remain viable in environmentally unfavorable conditions, and the ability to form viable resting spores may be essential in facilitating dispersal in these organisms (Mcquoid and Hobson 1995; Mcquoid and Hobson 1996). *Pseudo-nitzschia pungens* does not form resting spores, but *T. rotula* does (Garrison 1981); this may lead to differences in their dispersal potential, explaining why geographic distance is a greater barrier to geneflow in *P. pungens* than *T. rotula*. Starkly different patterns of population structure in the

two species points to the need to expand our understanding of diatom population diversity over larger spatial scales and across more species.

Ecological drivers of population structure

Thalassiosira rotula, as a species, does not represent a single panmictic population. In fact, despite evidence of connectivity over broad spatial scales, significant population structure within *T. rotula* over time and space demonstrates that 'everything is *not* everywhere.' But does the environment select? Distinct populations within global samples of *T. rotula* demonstrate that mechanisms exist to support their genetic isolation, and reduce gene flow. So, although geographic distance is not a strong barrier to gene flow amongst populations, other mechanisms for selection must exist to support the patterns of population structure observed. One mechanism supporting this pattern may be environmental selection.

On the global scale, in *T. rotula*, environment was strongly correlated with patterns of genetic relatedness among samples. Specifically, the combination of temperature and chlorophyll *a* was significantly correlated with pairwise genetic distances, explaining 53% of genetic variability ($p = 0.02$). Two possible hypotheses explaining the correlation between environment and genetic relatedness are 1) sexual selection based on environmental triggers of reproduction or 2) isolation of populations due to environmental selection.

To maintain the population structure observed here, sexual reproduction amongst isolates within a population must occur more frequently than between isolates from different populations. Diatoms exhibit both clonal and sexual life cycles, and the frequency of sex in the field is little understood for the majority of species;

observations of sex in the field are rare, and laboratory breeding experiments often unsuccessful. Environmental conditions can trigger sexual reproduction in diatoms (Chepurnov et al. 2004). In addition, the genes responsible for gamete recognition (*Sig*1 in particular) have been shown to evolve rapidly, exhibiting high levels of inter and intra-specific divergence within the *Thalassiosra* genus (Armbrust 1999; Armbrust and Galindo 2001). Rapid evolution of gamete recognition genes may regulate the breeding success of individuals from unique populations, thus promoting their isolation.

Here, we observed the genetic signature of sexual reproduction in *T. rotula.* Isolates across sites and loci exhibited high levels of heterozygote deficiencies, as has been observed in other diatom species (Casteleyn 2009a; Evans 2009; Rynearson 2004; Rynearson 2006). The consistency of this pattern across loci and samples suggests that this pattern is a real feature, and not solely a signature of null alleles. Exclusive clonality and asexual reproduction would result in an excess of heterozygotes (Balloux et al. 2003). Thus, heterozygote deficiency may be an imprint of frequent sexual reproduction in the species. Heterozygote deficiency has been observed across diatom species, and yet this common phenomenon is not fully understood. *Thalassiosira rotula* exhibited extremely high levels of clonal diversity; only four identical six-locus genotypes were observed more than once in 449 individuals genotyped. 99% clonal diversity suggests that the environment is heterogonous enough to prevent any one clone from becoming dominant (Rynearson 2005). In addition, sexual reproduction must be frequent in the field, and a force responsible for generating and maintaining diversity.

Environmental selection is perhaps the more parsimonious explanation of correlations between environmental variables (temp and chlorophyll *a*), and pairwise relateness between *T. rotula* samples. In general, the coexistence of populations at different times within a single location suggests that the environment supports their presence only during favorable conditions. These data demonstrate that population variability may be tightly coupled with environmental variability. For instance, populations most adapted to conditions of high chlorophyll (or bloom conditions) and high temperatures are more closely related, despite when or where those conditions occur. Whether or not environmental selection of these populations has lead to subsequent mechanisms of reproductive isolation between them cannot be determined here.

CONCLUSIONS

The structuring of *T. rotula* populations across time and global geographic space demonstrates that "everything is *not* everywhere"; this species is not genetically homogeneous, nor fully admixed. In *T. rotula* genetic composition can vary as much over time as it does over large spatial scales; this poses a caution to pooling samples over seasons or years, or grouping samples over space when exploring population-level diversity in diatoms. This work advances a decades-long debate on the relative importance of dispersal and the environment in structuring single-celled species at constant flux in the ocean environment. In *T. rotula*, geographic distance plays only a minor role in limiting gene flow and structuring populations. Instead, populations that vary as widely over time as they do in space are structured and maintained by the

environment. These data strongly suggest that it is the environment, rather than dispersal, that isolates and poses barriers to gene flow within *T. rotula*.

ACKNOLWEDGEMENTS

This research was supported by National Science Foundation grants, NSF 0727227 (to TAR) and NSF SBE0245039 (to URI, TAR). Part of the research was conducted using instrumentation supported by NSF-EPSCoR grants 0554548 and 1004057. We thank the following people who took part in seawater collection in 2010: Caitlin Lawrence, Heidi M. Sosik, Taylor Crockford, Barney Balch, Karen Wiltshire, Nicole Alberle-Malzahn, Erla Bjork , Hugh MacIntyre, Justin Liefer, Lucie Novoveska, Fabienne Jalabert, David A. Caron, Erica Seubert, Mya Keyzers, Anthony Odell, Bruce L. Wing, Tatiana Orlova, Ольга Шевченко, Martina Doblin, Katherine Baer-Jones, Sophie Mormede, Peter Thompson, Cathy Johnston, Beatriz Yancinnelli, and Johan Van der Molen. Environmental data was generously provided by several monitoring institutions including: Helgoland Roads, Southern California Coastal Observing System (SCCOS), Narragansett Bay Phytoplankton Time Series, Martha's Vineyard Coastal Observatory, Washington State Department of Ecology, and Service d'Observation en Milieu LITtoral (SOMLIT).

Abdelkrim, J., B. Robertson, J.-A. Stanton, and N. Gemmell. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. Biotechniques **46:** 185-192.

Armbrust, E. V. 1999. Identification of a New Gene Family Expressed during the Onset of Sexual Reproduction in the Centric DiatomThalassiosira weissflogii. Applied and environmental microbiology **65:** 3121-3128.

Armbrust, E. V., and H. M. Galindo. 2001. Rapid evolution of a sexual reproduction gene in centric diatoms of the genus Thalassiosira. Appl Environ Microbiol **67:** 3501-3513.

Arnold, M. L., and N. D. Fogarty. 2009. Reticulate evolution and marine organisms: the final frontier? International journal of molecular sciences **10:** 3836-3860.

Baldwin, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. Molecular phylogenetics and evolution **1:** 3-16.

Balloux, F. O., L. Lehmann, and T. De Mee√ªs. 2003. The Population Genetics of Clonal and Partially Clonal Diploids. Genetics **164:** 1635-1644.

Beszteri, B., E. Acs, and L. K. Medlin. 2005. Ribosomal DNA sequence variation among sympatric strains of the Cyclotella meneghiniana complex (Bacillariophyceae) reveals cryptic diversity. Protist **156:** 317-333.

Bidle, K. D., and S. J. Bender. 2008. Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom Thalassiosira pseudonana. Eukaryotic cell **7:** 223-236.

Blacket, M., C. Robin, R. Good, S. Lee, and A. Miller. 2012. Universal primers for fluorescent labelling of PCR fragments‚Äîan efficient and cost‚Äêeffective approach to genotyping by fluorescence. Molecular ecology resources **12:** 456-463.

Casteleyn, G. 2008. Pseudo-nitzschia pungens (Bacillariophyceae): a cosmopolitan diatom species? Harmful algae **7:** 241.

Casteleyn, G. 2009a. Lack of population genetic structuring in the marine planktonic diatom Pseudo-nitzschia pungens (Bacillariophyceae) in a heterogeneous area in the Southern Bight of the North Sea. Marine biology **156:** 1149-1158.

Casteleyn, G. 2009b. Natural hybrids in the marine diatom Pseudo-nitzschia pungens (Bacillariophyceae): genetic and morphological evidence. Protist **160:** 343.

Casteleyn, G. 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. PNAS : Proceedings of the National Academy of Sciences **107:** 12952-12957.

Chepurnov, V. A., D. G. Mann, K. Sabbe, and W. Vyverman. 2004. Experimental Studies on Sexual Reproduction in Diatoms, p. 91-154. International Review of Cytology. Academic Press.

Clarke, K., and M. Ainsworth. 1993. A method of linking multivariate community structure to environmental variables. Marine Ecology-Progress Series **92:** 205-205.

D'alelio, D., A. Amato, W. H. Kooistra, G. Procaccini, R. Casotti, and M. Montresor. 2009. Internal transcribed spacer polymorphism in Pseudo-nitzschia

multistriata (Bacillariophyceae) in the Gulf of Naples: recent divergence or intraspecific hybridization? Protist **160:** 9-20.
Earl, D. A. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources **4:** 359-361.
Ellegaard, M. 2008. The species concept in a marine diatom: LSU rDNA-based phylogenetic differentiation in Skeletonema marinoi/dohrnii (Bacillariophyceae) is not reflected in morphology. Phycologia (Oxford) **47:** 156-167.
Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology **14:** 2611-2620.
Evans, K. M. 2005. High levels of genetic diversity and low levels of genetic differentiation in North Sea Pseudo-nitzchia pungens (Bacillariophyceae) populations. . Journal of phycology **41:** 506-514.
Evans, K. M. 2009. Highly differentiated populations of the freshwater diatom Sellaphora capitata suggest limited dispersal and opportunities for allopatric speciation. Protist **160:** 386.
Faircloth, B. C. 2008. msatcommander: detection of microsatellite repeat arrays and automated, locus,Äêspecific primer design. Molecular Ecology Resources **8:** 92-94.
Fenchel, T., and J. F. Bland. 2004. The Ubiquity of Small Species: Patterns of Local and Global Diversity. BioScience **54:** 777-784.
Finlay, B. J. 2002. Global Dispersal of Free-Living Microbial Eukaryote Species, p. 1061-1063.
Foissner, W. 2006. Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta protozoologica **45:** 111.
Gallagher, J. C. 1980. Population genetics of *Skeletonema costatum* (Bacillariophyceae) in Narragansett Bay. Journal of phycology **16:** 464-474.
Garrison, D. L. 1981. Monterey Bay Phytoplankton. II. Resting Spore Cycles in Coastal Diatom Populations. Journal of Plankton Research **3:** 137-156.
Godhe, A., and K. Härnström. 2010. Linking the planktonic and benthic habitat: genetic structure of the marine diatom Skeletonema marinoi. Molecular Ecology **19:** 4478-4490.
Guillard, R. R. 1975. Culture of phytoplankton for feeding marine invertebrates. Plenum Press.
Hamilton, M. B., E. L. Pincus, A. Di Fiore, and R. C. Fleischer. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. BioTechniques **27:** 500-502, 504-507.
Härnström, K., M. Ellegaard, T. R. J. Andersen, and A. Godhe. 2011. Hundred years of genetic structure in a sediment revived diatom population. Proceedings of the National Academy of Sciences **108:** 4252-4257.
Hochberg, Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika **75:** 800-802.
Jensen, J. L., A. J. Bohonak, and S. T. Kelley. 2005. Isolation by distance, web service. BMC Genet **6:** 13.

Kinlan, B. P. 2003. Propagule dispersal in marine and terrestrial environments: a community perspective Ecology (Durham) **84:** 2007-2020.
Kooistra, W. 2008. Global diversity and biogeography of Skeletonema species (Bacillariophyta). Protist **159:** 177.
Linder, C. R., and L. H. Rieseberg. 2004. Reconstructing patterns of reticulate evolution in plants, p. 1700-1708.
Mcquoid, M. R., and L. A. Hobson. 1995. Importance of resting stages in diatom seasonal succession Journal of Phycology **31:** 44-50.
Mcquoid, M. R., and L. A. Hobson. 1996. Diatom resting stages. Journal of Phycology **32:** 889-902.
Mcrae, B. H. 2006. Isolation by resistance. Evolution **60:** 1551-1561.
Medlin, L. K. 2007. If everything is everywhere, do they share a common gene pool? Gene **406:** 180-183.
Park, S. 2001. Trypanotolerance in West African Cattle and the Population Genetics Effects of Selection.
Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research,Äîan update. Bioinformatics **28:** 2537-2539.
Pilson, M. E. Q. 1985a. Annual cycles of nutrients and chlorophyll in Narragansett Bay, Rhode Island. Journal of Marine Research **43:** 849-873.
Pilson, M. Q. 1985b. On the residence time of water in Narragansett Bay. Estuaries **8:** 2-14.
Pritchard, J., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155:** 945 - 959.
Pritchard, J. K., W. Wen, and D. Falush. 2003. Documentation for STRUCTURE software: version 2.
Raymond, M. 1995. An exact test for population differentiation. Evolution **49:** 1280.
Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. Journal of Heredity **86:** 248-249.
Rieseberg, L. H., and J. F. Wendel. 1993. Introgression and its consequences in plants. Hybrid zones and the evolutionary process**:** 70-109.
Robinson, G., J. Aiken, and H. Hunt. 1986. Synoptic surveys of the western English Channel. The relationships between plankton and hydrography. Journal of the Marine Biological Association of the United Kingdom **66:** 201-218.
Rosenberg, N. 2004. DISTRUCT: a program for the graphical display of structure results. Molecular Ecology Notes **4:** 137 - 138.
Rynearson, T. A. 2004. Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). Journal of phycology **40:** 34-43.
Rynearson, T. A. 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom Ditylum brightwellii. Molecular ecology **14:** 1631-1640.
Rynearson, T. A. 2006. Spring bloom development, genetic variation, and population succession in the planktonic diatom Ditylum brightwellii. Limnology and oceanography **51:** 1249.

Rynearson, T. A., and E. V. Armbrust. 2000. DNA Fingerprinting Reveals Extensive Genetic Diversity in a Field Population of the Centric Diatom Ditylum brightwellii. Limnology and Oceanography **45:** 1329-1340.

Rynearson, T. A., and E. V. Armbrust. 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom Ditylum brightwellii. Mol Ecol **14:** 1631-1640.

Rynearson, T. A., E. O. Lin, and E. V. Armbrust. 2009. Metapopulation Structure in the Planktonic Diatom Ditylum brightwellii (Bacillariophyceae). Protist **160:** 111-121.

Sarno, D. 2005. Diversity in the genus Skeletonema (Bacillariophyceae). An assessment of the taxonomy of S. costatum-like species with the description of four new species. . Journal of phycology **41:** 151-176.

Southward, A. J. and others 2004. Long-term oceanographic and ecological research in the western English Channel. Advances in Marine Biology **47:** 1-105.

Spies, I. B., S. Lowe, Y. Hong, and M. F. Canino. 2005. Development and characterization of seven novel di-, tri-, and tetranucleotide microsatellite markers in Atka mackerel (Pleurogrammus monopterygius). Molecular Ecology Notes **5:** 469-471.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Molecular Biology and Evolution **24:** 1596-1599.

Thompson, J. D. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research **22:** 4673-4680.

Von Dassow, P., and M. Montresor. 2011. Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. Journal of plankton research **33:** 3-12.

Wahlund, S. 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas **11:** 65-106.

Whittaker, K., D. Rignanese, R. Olson, and T. Rynearson. 2012. Molecular subdivision of the marine diatom Thalassiosira rotula in relation to geographic distribution, genome size, and physiology. BMC Evolutionary Biology **12:** 209.

TABLES

Table 1. **Sites and Locations**
Names and locations of sites from which whole seawater was received in 2010. This includes locations from which samples were received, but *T. rotula* was not observed or isolated.

	Site	Latitude	Longitude
1.	Alaska, USA	58.38	-134.64
2.	Puget Sound, Washington, USA	47.88	-122.48
3.	Olympic Peninsula, Washington, USA	47.04	-125.09
4.	Newport Beach, California, USA	33.61	-117.93
5.	Boothbay Beach, Maine, USA	43.72	-68.02
6.	Martha's Vineyard Coastal Observatory, Massachusetts, USA	41.29	74.55
7.	Narragnasett Bay, Rhode Island, USA	41.57	-71.38
8.	Manasquoa, New Jersey, USA	40.11	-74.03
9.	Dauphin Island, Alabamable, USA	30.23	-87.77
10.	Chile	-71.41	-29.91
11.	Iceland	65.38	22.92
12.	Helgoland, Germany	51.18	7.90
13.	Roscoff, France	49.00	3.93
14.	Durban, South Africa	-29.87	31.05
15.	Vladivastok, Russia	43.18	131.90
16.	Sydney, Australia	-33.86	151.26
17.	Port Hacking, NSW	-34.07	151.12
18.	Tasmania	-42.89	147.34
19.	Greta Point, New Zealand	-41.30	174.81
20.	Southland Current, New Zealand	-45.79	170.83

Table 2. **Samples**
List of samples containing *T. rotula*, sample location, site abbreviation (used throughout the study, and in subsequent tables and figures), longitude and latitude, date of isolation, and number of cells isolated. Metadata source refers to the site-related time series, or established oceanographic sampling institutions. Websites for monitoring institutions and data hubs are as follows: 1. http://www.gso.uri.edu/phytoplankton/ 2. http://www.ecy.wa.gov/database.html 3. http://somlit.epoc.u-bordeaux1.fr/fr/
4.http://www.awi.de/en/research/research_divisions/biosciences/shelf_sea_ecology/long_term_studies/helgoland_roads_long_term_data_series/
5. http://www.sccoos.org/interactive-map/

Season	Sample Site	Site Name	Long.	Lat.	Date Collected	# Cells Isolated	# Cells Survived	# Confirmed T. rotula	Long-term Monitoring/ Metadata Source
Northern Hemisphere Winter/Spring	Narragansett Bay, RI, USA (incubated)	NBa	41.56	-71.38	1/26/10	48	48	44	Phytoplankton of Narragansett Bay [1]
	Narragansett Bay, RI, USA (shipped)	NBb	41.56	-71.38	1/26/10	48	48	36	Phytoplankton of Narragansett Bay [1]
	Narragansett Bay, RI, USA	NBc	41.56	-71.38	2/2/10	22	20	19	Phytoplankton of Narragansett Bay [1]
	Narragansett Bay, RI, USA	NBd	41.56	-71.38	3/22/10	12	12	11	Phytoplankton of Narragansett Bay [1]
	Olympic Peninsula, WA, USA	WAa	47.04	-125.09	2/16/10	48	39	37	Washington State Department of Ecology [2]
	Roscoff, France	FRa	49	3.93	3/9/10	48	45	33	SOMLIT[3]
	Helgoland, Germany	Ger	51.18	7.9	3/18/10	20	19	19	Helgoland Roads[4]
	Newport Beach, CA, USA	Cali	33.61	-117.93	3/22/10	56	55	51	Southern California Coastal Observing System[5]
	Roscoff, France	FRb	49	3.93	3/24/10	48	45	26	SOMLIT[3]
	Olympic Peninsula, WA, USA	WAb	47.04	-125.09	3/29/10	41	40	40	Washington State Department of Ecology [2]
	Puget Sound, WA	Puget	47.88	-122.48	4/14/10	35	35	29	Washington State Department of Ecology [2]
	Iceland	Ice	65.37	22.92	4/20/10	48	0	0	NA
Austral Winter/Spring	Greta Point, NZ	NZ	41.3	174.8	10/15/10	8	8	8	Sampler
	Usha Pier, SA	SA	-29.02	-31.05	11/19/10	96	96	96	Sampler

Table 3. **Microsatellite Loci**

List of microsatellite loci, repeat motifs, and their optimal conditions for amplification. Italicized primer segment of locus TR27 refers to the forward fluorescent tag used in the three-primer amplification method. Touchdown cycles were used to improve amplification of loci in low-concentration field DNA, and may not be necessary for high-concentration DNA samples. Optimal temperature refers to the temperature of annealing that balances the need for high specificity, and suitable amplification. Modification to the 5' end of the forward primer refers to a fluorescent probe, allowing for detection of amplicons in fragment analysis.

Locus	Repeat	Size range (bp)	Forward Primer	Reverse Primer	Touchdown start temp. /Increment/Cycle #	Optimal Temp./Cycle #	5' Modification
TR1	$(GTT)^5GCAAGC(TGG)^9$	120-166	GGCAAAGCCCACTTCCAATTT	ATCCCCTCCGCCTCTTCC	60ºC/-1ºC/7	55ºC/28	Vic
TR3	$(GA)^{16}$	102-276	CCAGAGGGTATGGCTGTAAG	ATGCTGCCGGTAGTGAGTTG			Ned
TR7	$(AC)^{13}$	126-354	AGCGTAGGTGTCCCAGAAC	ATGCCATGTACACCGTTGC	NA	60.1ºC/34	6Fam
TR8	$(AG)^{12}$	139-211	TGACCTAGGCAGCCAGAATG	ATGGCGATGTGATTGGTGC	67ºC/-0.2ºC/37	55ºC/12	Vic
TR10	$(ACT)^{11}$	121-400	CATCGAGTGCCCGTTGTTG	GTTGATTGGGTCCAAGGGG			6Fam
TR27	$(GTT)^{10}$	125-328	*GCCTCCCTCGCGCCA-*CGTTCCTTCTTCGGAGGGG	CAGCTCCCTCACCACCAG	68ºC/-1ºC/10	57ºC/35	Vic

Table 4. **Summary Statistics per Site and Locus**

Summary of diversity statistics and allelic richness for each site, at each locus and each sample. Number of individuals typed for each locus at each site (N), number of alleles (N_a), expected heterozygosity (H_e), observed heterozygosity (H_o), and inbreeding coefficient (F_{is}). F_{is} values in bold represent samples that exhibit significant heterozygote deficiency.

	Locus	Site Name												
		Nba	**NBb**	**NBc**	**NBd**	**Fra**	**FRb**	**WAa**	**WAb**	**Cali**	**Puget**	**NZ**	**SA**	**Helg**
TR8	N	44	36	17	7	25	19	30	40	36	24	8	94	16
	N_A	27	22	20	8	24	16	19	34	18	16	7	44	22
	Size range (bp)	139-255	152-270	133-306	139-203	129-278	139-141	130-195	136-272	133-218	136-220	149-189	136-246	127-247
	H_e	0.89	0.89	0.94	0.90	0.96	0.93	0.89	0.95	0.84	0.91	0.84	0.93	0.96
	H_o	0.39	0.50	0.71	0.57	0.52	0.63	0.23	0.53	0.19	0.54	0.63	0.53	0.88
	F_{IS}	**0.57**	**0.44**	0.25	0.38	**0.46**	**0.33**	**0.74**	**0.45**	**0.77**	**0.41**	0.27	**0.43**	0.09
TR1	N	42	36	19	10	33	23	37	37	49	29	7	92	19
	N_A	3	4	4	5	5	9	7	4	7	9	1	11	6
	Size range (bp)	123-129	120-129	120-129	120-129	122-129	120-122	115-166	123-139	110-129	114-151	123	110-148	110-139
	H_e	0.65	0.62	0.66	0.76	0.68	0.71	0.72	0.63	0.66	0.65	0.00	0.42	0.64
	H_o	0.64	0.50	0.63	0.40	0.61	0.39	0.54	0.40	0.31	0.52	0.00	0.37	0.63
	F_{IS}	0.01	**0.19**	0.04	**0.49**	0.1	**0.46**	**0.25**	**0.37**	**0.54**	**0.21**	NA	0.11	0.02
TR3	N	44	33	19	11	33	22	37	40	51	27	8	93	19
	N_A	5	12	11	6	8	14	14	17	12	11	6	15	9
	Size range (bp)	108-137	106-137	103-141	108-134	108-256	102-205	106-147	103-187	108-139	104-135	108-167	106-276	105-125
	H_e	0.62	0.77	0.86	0.80	0.71	0.83	0.79	0.82	0.81	0.77	0.62	0.61	0.62
	H_o	0.73	0.76	0.74	1.00	0.58	0.59	0.81	0.75	0.63	0.85	0.63	0.71	0.42
	F_{IS}	-0.17	0.01	**0.14**	-0.27	**0.18**	**0.29**	**-0.03**	**0.09**	**0.22**	**-0.1**	-0.01	**-0.17**	**0.33**
TR7	N	42	31	19	10	33	20	36	36	48	29	6	90	18
	N_A	10	7	3	4	5	11	8	16	11	13	1	15	9
	Size range (bp)	193-227	192-212	201-215	175-213	201-215	199-242	169-215	162-258	201-355	166-328	212	126-354	165-345
	H_e	0.43	0.32	0.28	0.51	0.46	0.61	0.47	0.69	0.42	0.69	0.00	0.45	0.70
	H_o	0.17	0.23	0.32	0.00	0.12	0.30	0.19	0.25	0.21	0.59	0.00	0.28	0.33
	F_{IS}	**0.62**	0.3	-0.13	**1**	**0.74**	**0.51**	**0.59**	**0.64**	**0.51**	0.15	NA	**0.39**	**0.53**
TR27	N	41	36	17	10	32	25	36	35	48	21	8	95	12
	N_A	8	9	9	13	17	14	17	33	40	23	8	22	13
	Size range (bp)	157-274	157-253	188-221	125-257	182-274	184-222	185-222	125-307	140-321	156-328	108-207	138-247	164-256
	H_e	0.54	0.56	0.80	0.95	0.86	0.81	0.84	0.94	0.96	0.96	0.88	0.80	0.94
	H_o	0.49	0.64	0.53	0.50	0.56	0.60	0.53	0.66	0.67	0.57	0.75	0.76	0.58
	F_{IS}	0.1	-0.15	**0.35**	**0.49**	**0.35**	**0.26**	**0.38**	**0.3**	**0.31**	**0.41**	0.15	**0.06**	**0.39**
TR10	N	42	35	17	10	26	24	29	38	37	26	8	83	17
	N_A	38	38	19	14	31	28	31	36	40	29	14	68	19
	Size range (bp)	143-291	142-295	139-263	142-267	145-343	143-353	137-331	147-311	121-400	125-400	158-400	147-353	130-310
	H_e	0.96	0.97	0.96	0.97	0.98	0.97	0.97	0.95	0.98	0.97	0.98	0.98	0.96
	H_o	0.71	0.71	0.53	0.50	0.42	0.75	0.48	0.53	0.65	0.58	0.75	0.51	0.47
	F_{IS}	**0.26**	**0.26**	**0.45**	**0.5**	**0.57**	**0.23**	**0.51**	**0.45**	**0.34**	**0.41**	**0.25**	**0.48**	**0.51**

Table 5. **Diversity Statistics**
Summary of diversity statistics and allelic richness for each site, across all loci.
Included are diversity statistics averaged across all loci for each site, and the respective standard deviation (SD) including unbiased expected heterozygosity (H_e), observed heterozygosity (H_o), and number of alleles (no. alleles) averaged across loci for each site. Total sample size is given, along with average (global) diversity statistics across the entire sample set.

Population	Sample size	H_e	H_e SD	H_o	H_o SD	# Alleles	No Alleles SD
NBa	44	0.683	0.083	0.521	0.031	15.167	14.077
NBb	36	0.687	0.097	0.556	0.035	15.333	12.707
NBc	19	0.749	0.103	0.575	0.048	11.000	7.239
NBd	11	0.814	0.070	0.495	0.066	8.333	4.227
FRa	33	0.773	0.081	0.468	0.037	15.000	10.863
FRb	26	0.810	0.055	0.544	0.043	15.333	6.683
WAa	37	0.780	0.071	0.465	0.035	16.000	8.764
WAb	40	0.831	0.058	0.518	0.033	23.333	12.925
Cali	51	0.778	0.086	0.442	0.030	21.333	14.882
Puget	29	0.824	0.057	0.608	0.039	16.833	7.705
Nz	8	0.553	0.181	0.458	0.077	6.167	4.875
SA	96	0.699	0.098	0.525	0.021	29.167	22.409
Helg	19	0.804	0.068	0.552	0.049	13.000	6.293
Total/Average	**449**	**0.752**	**0.085**	**0.518**	**0.042**	**15.846**	**10.281**

Table 6. **Pairwise F_{ST}**

Multilocus F_{ST} values among pairs of populations (below diagonal). Significant results (bold) at each locus (after Bonferroni correction) are shown above diagonal.

	Nba	Nbb	Nbc	Nbd	Fra	Frb	Waa	Wab	Calia	Puget	Nzb	SA	Helg
Nba	0.000	0.014	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
Nbb	0.009	0.000	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
Nbc	0.044	0.044	0.000	0.003	0.038	**0.001**	0.466	**0.001**	**0.002**	**0.001**	**0.001**	**0.001**	**0.001**
Nbd	0.064	0.051	0.029	0.000	0.171	0.439	0.009	0.062	0.012	**0.001**	**0.001**	**0.001**	0.184
Fra	0.044	0.045	0.009	0.006	0.000	**0.001**	0.251	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
Frb	0.064	0.046	0.046	0.000	0.023	0.000	**0.001**	**0.001**	**0.001**	**0.001**	0.003	**0.001**	0.017
Waa	0.031	0.035	0.000	0.020	0.002	0.036	0.000	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
Wab	0.052	0.044	0.031	0.009	0.018	0.018	0.028	0.000	**0.001**	**0.001**	**0.001**	**0.001**	0.018
Calia	0.049	0.047	0.017	0.015	0.018	0.038	0.011	0.034	0.000	**0.001**	**0.001**	**0.001**	**0.001**
Puget	0.094	0.094	0.072	0.045	0.055	0.048	0.065	0.024	0.058	0.000	**0.001**	**0.001**	**0.001**
Nzb	0.135	0.118	0.139	0.065	0.093	0.035	0.110	0.053	0.127	0.104	0.000	**0.001**	**0.001**
SA	0.092	0.083	0.076	0.037	0.041	0.038	0.063	0.028	0.072	0.079	0.060	0.000	**0.001**
Helg	0.075	0.069	0.046	0.005	0.027	0.012	0.042	0.010	0.044	0.022	0.059	0.049	0.000

Table 7. **AMOVA Results**

Analysis of molecular variance (AMOVA) results, describing estimated variance (Est.Var.) distributed among regions, populations, individuals, and within individuals. Regions refers to Atlantic and Pacific Ocean basins. Pops refers to thirteen discrete samples. Significant variation was observed among pops, individuals, and within individuals, but not among regions. [a]degrees of freedom [b]sum of squares [c]mean squared deviations

Source	**df**[a]	**SS**[b]	**MS**[c]	**Est. Var.**	**%**
Among Regions	1	13.499	13.499	0.000	0%
Among Pops	11	127.034	11.549	0.127	5%
Among Indiv	436	1445.574	3.316	0.884	35%
Within Indiv	449	695.000	1.548	1.548	60%
Total	897	2281.107		2.559	100%

Table 8. **BIOENV Results**
Shown are the best BIOENV results for correlation of environmental differences among global samples with pairwise F_{ST} values. Variables include Chlorophyll *a* (chl*a*), salinity (sal), temperature (temp), cell abundance (CA). Table displays number of variables tested (# Vars), their correlations, and the variables associated with each test. For the global test, Rho = 0.529, and p =0.02.

# Variables	Correlation	Variables
2	0.529	chl*a*, temp
1	0.364	chl*a*
3	0.331	chl*a*,sal, temp
2	0.268	temp
2	0.208	chl*a*, sal
2	-0.095	sal, temp
1	-0.254	ca
2	-0.254	sal, ca

FIGURE LEGENDS

Figure 1. **Global Map**

Sites in red indicate locations where samples were collected, but *T. rotula* was not observed. Sites in blue indicate locations where samples were collected, and *T. rotula* successfully isolated. Numbers refer to locations listed in Table 1.

Figure 2. **T-RFLP**

Transcription fragment length polymorphism (T-RFLP) data showing the differences in fragment length between *T. rotula* and commonly misidentified *Thalassiosira* species, allowing for efficient species-level identification of field isolates. Species name and fragment length are given. X axis of each graph represents base pair (bp) length of fragment. Y axis indicates the signal intensity of the fragment in one particular run.

Figure 3. **Expected vs. Observed Heterozygosity**

Comparison of expected vs. observed heterozygosity for each site. Expected heterozygosity (H_e) is in black, and observed heterozygosity (H_o) is in grey. Error bars represent one SD across all 6 loci.

Figure 4. **Network Analysis**

Network analysis displaying the relationship among ITS1 sequence types of individuals from each site. Circles represent unique sequence types. Size of each circle represents the relative number of isolates associated with that sequence type.

Hash marks represent single nucleotide differences between sequence types. Pie charts represent the proportion of isolates from each region containing each sequence type: blue = Eastern Pacific, red= Atlantic, yellow= Western Pacific, green = Indian Ocean. Outlines indicate Lineages 1 and 3.

Figure 5. **Principal Coordinates Analysis**

Principal coordinates analysis (PCoA) translating pair-wise F_{ST} between global sites into Euclidean distance over a 2D space. Lines grouping sites represent different levels of pairwise F_{ST} genetic distance. Small dotted lines represent F_{ST} values of 0.024. Dashed lines represent F_{ST} values of 0.050. Thin solid line represents F_{ST} values of 0.073. Thick solid line represents F_{ST} values of 0.098. Pie charts associated with each site represent proportion of individuals representing lineages 1 (blue) or 2 (red).

Figure 6. **STRUCTURE Results**

STRUCTURE shows the Bayesian clustering of individuals into K=4 groups. Number of K groups were chosen based on the graphical analysis of STRUCTURE results, following the Evanno *et al.* method [43]. Individuals are organized by sampling site. Each individual is represented by a single bar. Each K group is represented by a different color. For each individual, color bar represents membership probability of that individual into any number of K groups (y axis from 0-1). Admixture is apparent when membership coefficient for individuals is < 1 into two or more groups.

Figure 7. **Detection of K Groups**

Graphical method allowing for the detection of K including a.) the mean log of K (L(K)) ± SD over 10 runs for each value of K (2-13) b. Delta K, as described in Evanno *et al.*[43]. K =11 was chosen as the most likely number of populations. However, K=4 exhibited the second highest delta K, and a smaller standard deviation of L(K) than K=11. Both were chosen for visualization of STRUCTURE results to improve understanding of hierarchical structure within the data.

Figure 8. **Isolation-by-Distance**

Isolation-by-distance (IBD) analysis displays the regression of pair-wise genetic ($F_{ST}/(1-F_{ST})$) with pairwise geographic distance (measured in shortest continuous ocean path between sites) among 13 global samples. $R^2= 0.20$, $p=0.007$.

Figure 9. **Cell Abundance**

Cells/L of *T. rotula* for each water sample from which *T. rotula* was isolated. Error bars represent variation from triplicate cell counts of each sample. Sample from Helgoland was not well preserved, and thus cells could not be enumerated.

Figure 10. **Isolation-by-Environment**

Isolation-by-environment graph shows regression of pair-wise genetic (F_{ST}) with pair-wise environmental distance (measured as Euclidean distance of environmental factors

among sites. In this isolation-by-environment regression, 'environmental distance' relates the Euclidean distance between sites when two variables, chlorophyll *a* and temperature, are considered. These are the two variables determined by BIOENV to correlate best with genetic distance, with a significant Rho of 0.529 and p value of 0.002.

FIGURES

Figure 1.

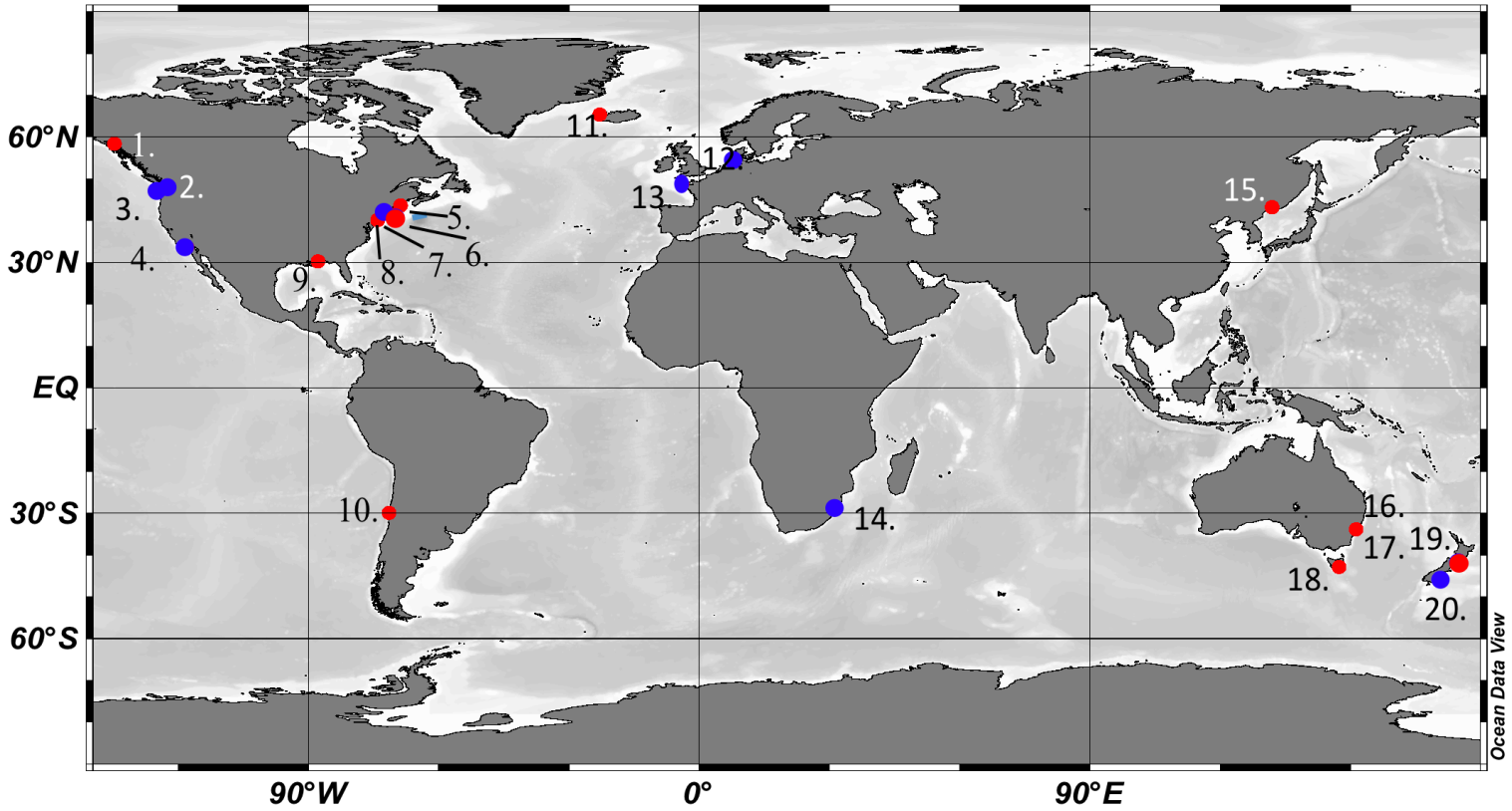


Figure 2.

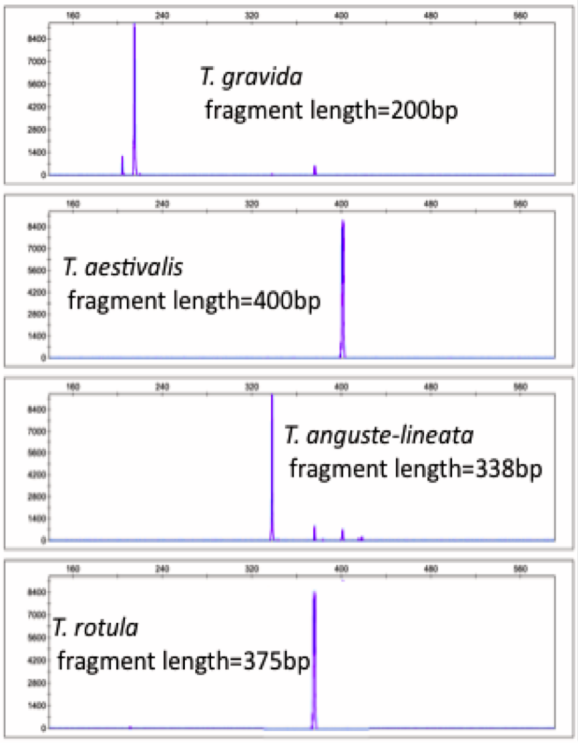

T. gravida
fragment length=200bp
T. aestivalis
fragment length=400bp
T. anguste-lineata
fragment length=338bp
T. rotula
fragment length=375bp

Figure 3.

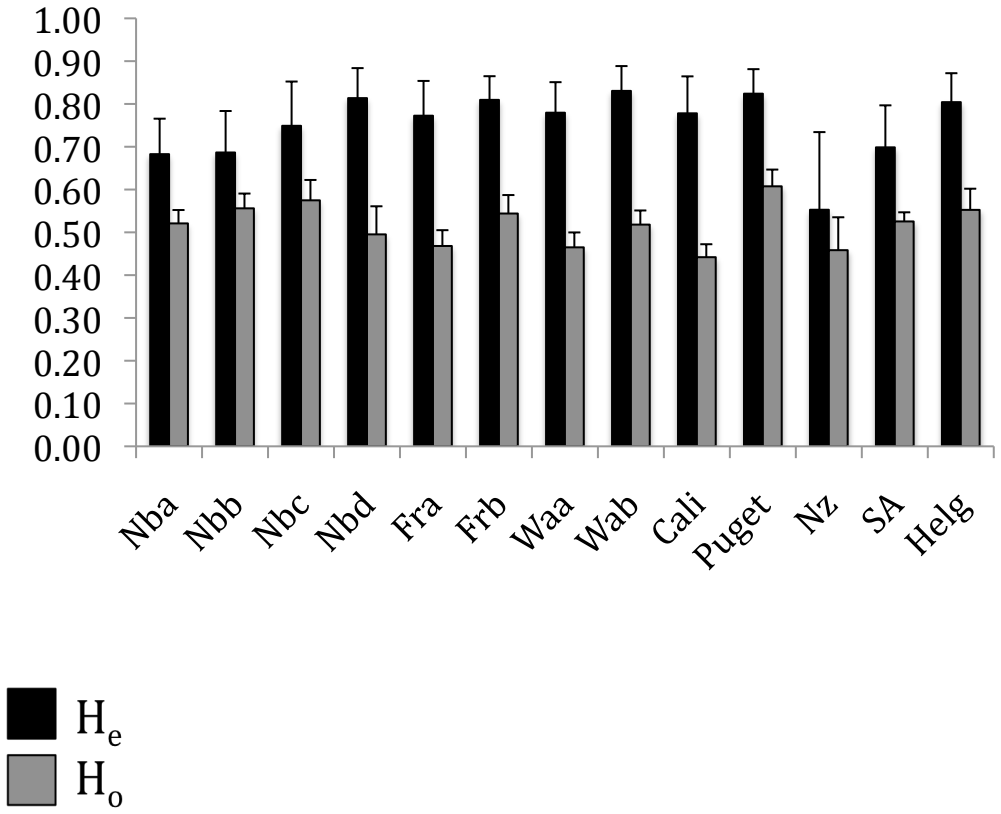

1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Nba
Nbb
Nbc
Nbd
Fra
Frb
Waa
Wab
Cali
Puget
Nz
SA
Helg
H_e
H_o

Figure 4.

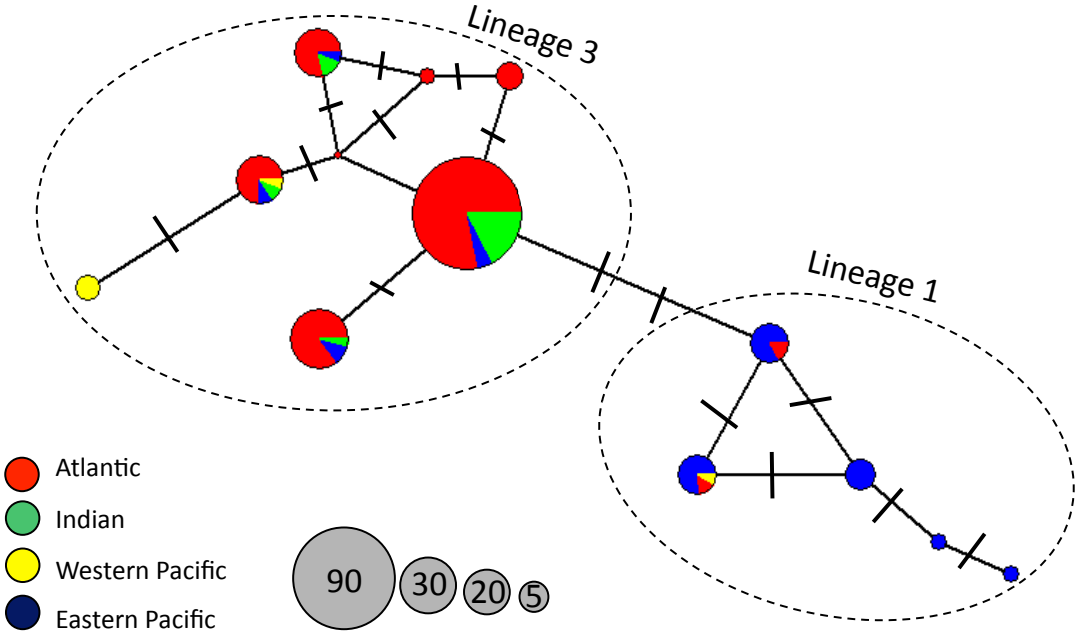

Lineage 3
Lineage 1
Atlantic
Indian
Western Pacific
Eastern Pacific
90
30
20
5

Figure 5.

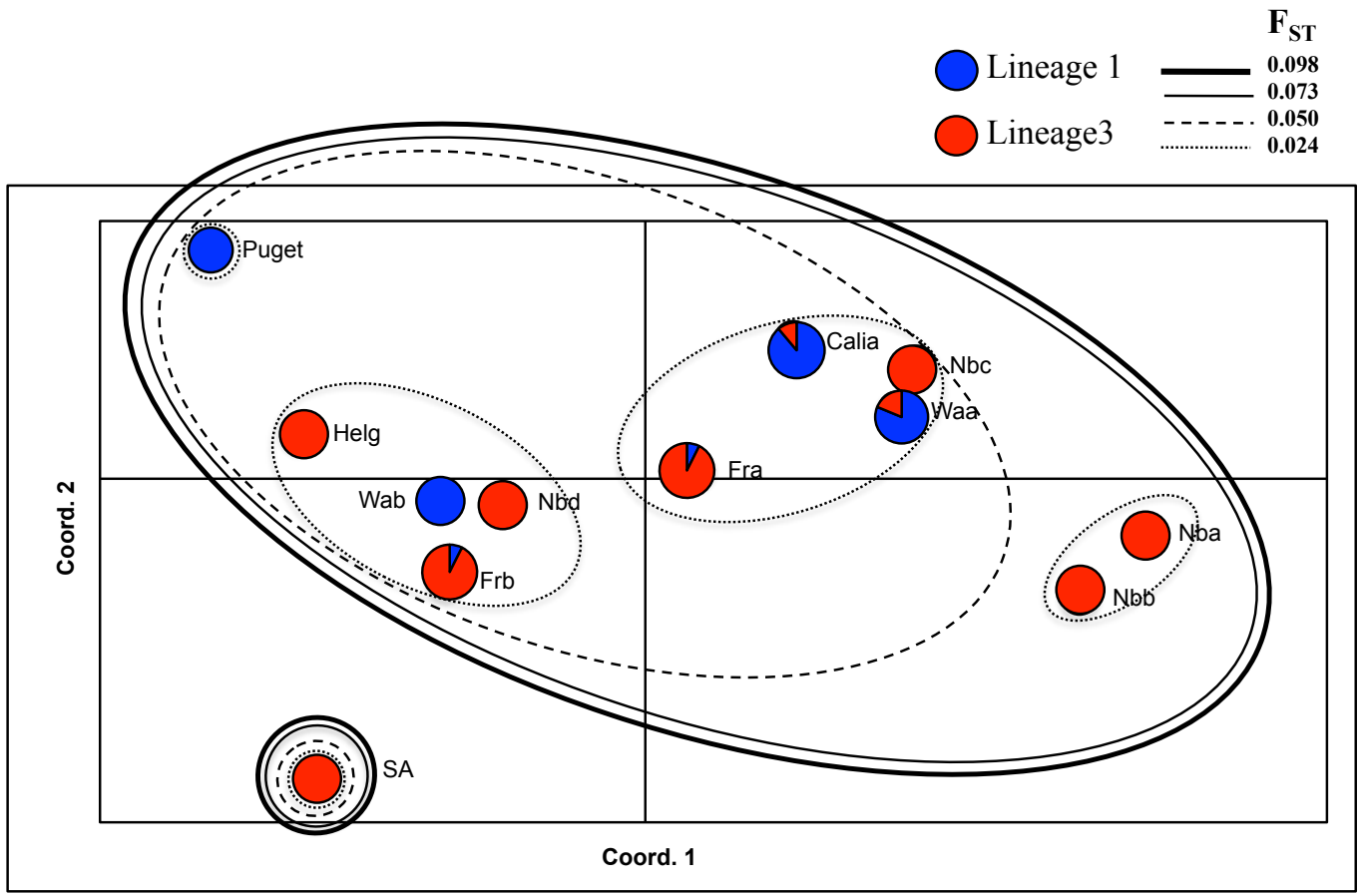

Lineage 1
Lineage3
F_{ST}
0.098
0.073
0.050
0.024
Puget
Helg
Wab
Nbd
Frb
Calia
Nbc
Waa
Fra
Nba
Nbb
SA
Coord. 2
Coord. 1

Figure 6.

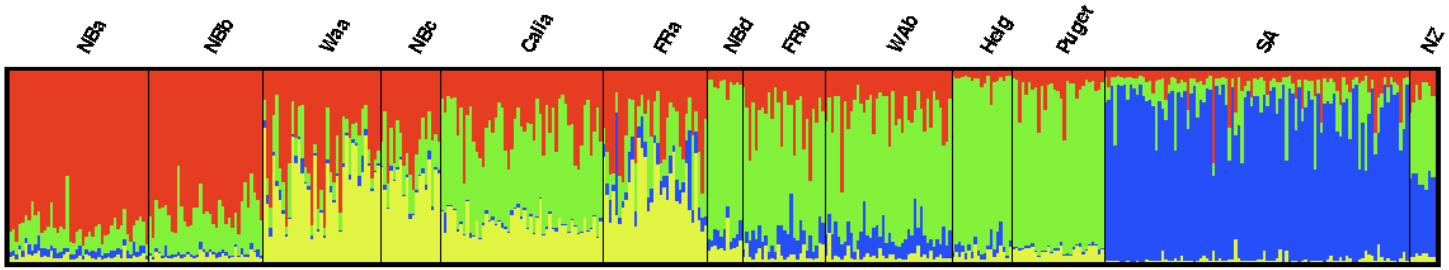
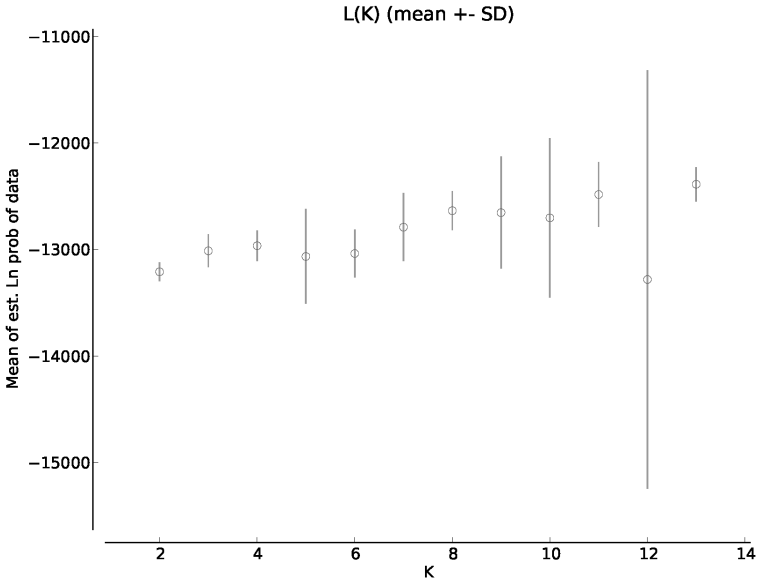


Figure 7. a. and b.



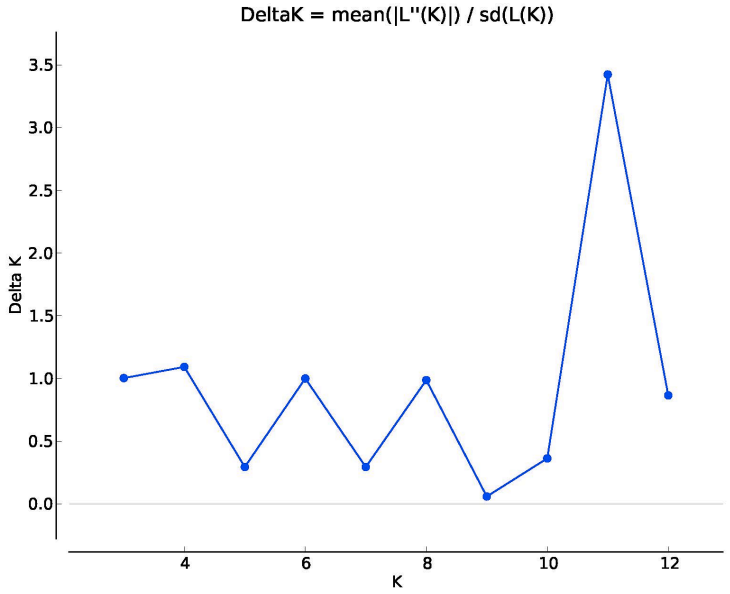


Figure 8.

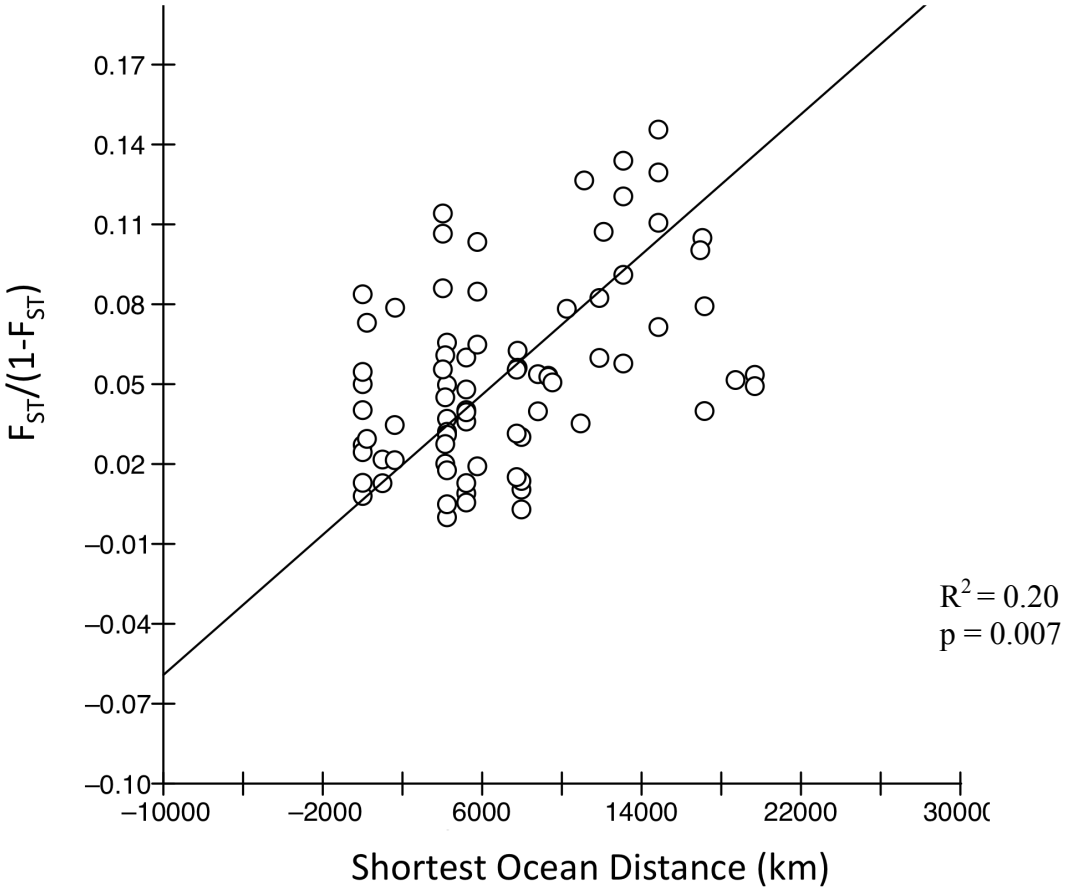

0.17
0.14
0.11
0.08
0.05
0.02
–0.01
–0.04
–0.07
–0.10
$F_{ST}/(1-F_{ST})$
–10000
–2000
6000
14000
22000
30000
Shortest Ocean Distance (km)
$R^2 = 0.20$
p = 0.007

Figure 9.

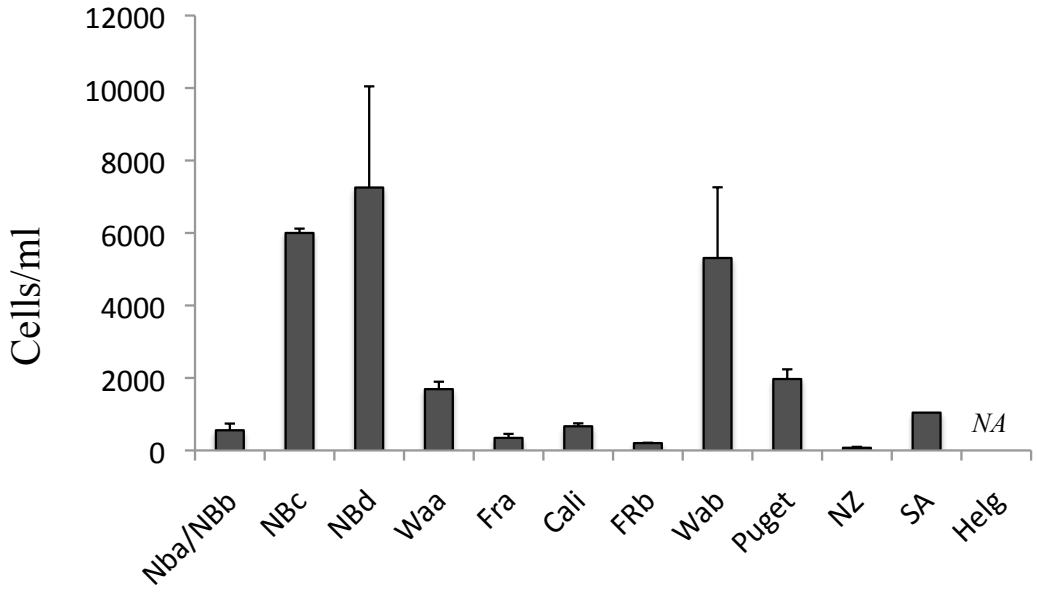
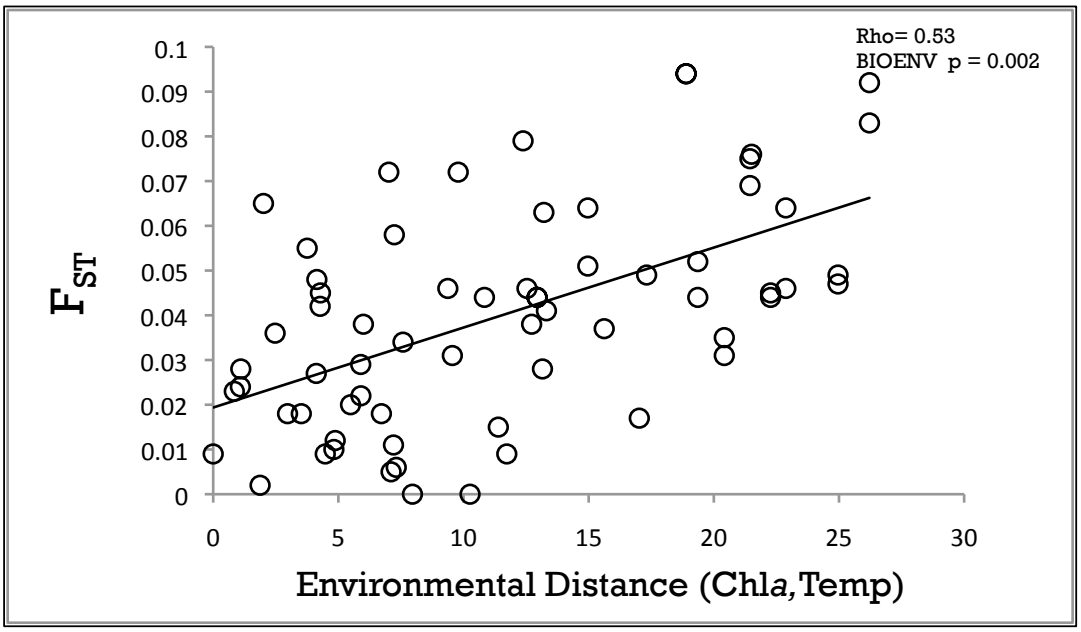

12000
10000
8000
6000
4000
2000
0
Cells/ml
NA
Nba/NBb
NBc
NBd
Waa
Fra
Cali
FRb
Wab
Puget
NZ
SA
Helg

Figure 10.


Rho= 0.53
BIOENV p = 0.002
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
F_{ST}
0
5
10
15
20
25
30
Environmental Distance (Chl*a*,Temp)

CHAPTER 4

POPULATION SUCCESSION OF *THALASSIOSIRA ROTULA* OVER THREE YEARS IN NARRAGANSETT BAY AND COASTAL MARTHA'S VINEYARD IN RELATION TO SEASONAL, ENVIRONMENTAL, AND ECOLOGICAL VARIABILITY

By

Kerry A. Whittaker[1] and Tatiana A. Rynearson[1]

1. Graduate School of Oceanography, South Ferry Road, University of Rhode Island, Narragansett, RI, 02882

ABSTRACT

The extent to which diatom population-genetic structure varies over time may play an important role in determining diatom bloom formation and potential to adapt to environmental change. Here, we examined changes in genetic diversity of the diatom *Thalassiosira rotula* throughout a 3-year period within Narragansett Bay and off the coast of Martha's Vineyard. Samples were comprised of over 99% unique clones; growth experiments demonstrated that individual clones were physiologically diverse, suggesting that each population harbors a high potential to adapt to environmental change. Eight significantly diverged populations were identified during this period, with F_{ST} values approaching 0.2. Samples were not consistently diverged between Narragansett Bay and Martha's Vineyard, suggesting that the physical connectivity between sites and high dispersal potential of *T. rotula* permit their genetic connectivity. Changes in intraspecific diversity of *T. rotula* over time exhibited a significant positive relationship to community-level interspecific diversity in Narragansett Bay, pointing to parallel mechanisms that structure gnetic diversity and drive isolation on these different genetic levels. Population succession exhibited no correlation with environmental changes on the seasonal scale. In fact, population structure was not correlated significantly with abiotic factors such as temperature, salinity, irradiance, or nutrient concentrations. Instead, genetic relationships were most correlated with chlorophyll *a* concentration. For example, those isolates collected at high concentrations of chlorophyll *a* were more closely related with eachother than with those collected at low concentrations of chlorophyll *a*, and vice versa. Here, we demonstrate that populations of *T. rotula* can vary over short time

scales, harbor enormous adaptive potential, and may respond rapidly to changes in the marine environment and local ecology.

INTRODUCTION

Diatoms contribute an estimated 40% of global carbon production (Nelson, Treguer et al., 1995). As biogeochemical powerhouses, the factors affecting their production and abundance over space and time are of particular interest. The genetic diversity of diatom communities, its seasonal variability, and sensitivity to long-term climate trends, is thought to influence the health and productivity of marine ecosystems (Beaugrand et al. 2010 ; Diaz et al. 2006; Duffy 2002; Folke et al. 2004; Stachowicz et al. 2002; Worm et al. 2006). The diversity and composition of diatom species in the field has traditionally been monitored using microscopy. More recent use of molecular tools has provided a glimpse into the fine-scale genetic underpinnings of diatom diversity, on the population and even clonal levels (Casteleyn 2009; D'alelio et al. 2009; Härnström et al. 2011; Rynearson 2004; Rynearson 2006; Rynearson et al. 2009). These finer-scale genetic subdivisions may be tightly coupled with environmental variability over time (Doney et al. 2004; Schmidt et al. 2008). Thus, clarifying the scale and triggers of change in genetic composition over time may enhance our understanding of the interactions between diatom diversity, evolution, and environmental variability.

Changes in diatom population-genetic structure over time may influence the timing of their bloom formation. Diatom blooms are some of the most prominent features of biogeochemical variability in open-ocean and coastal marine environments. Blooms occur when cell growth rates outpace loss rates; growth rates can vary with

changes in nutrient, temperature, light regimes, physical forcing, grazing pressure, or viral infection (Behrenfeld 2010; Cloern 1996; Sverdrup 1953; Townsend et al. 1994). The timing of these physical and environmental triggers of diatom growth rate relates to the temporal variation of cell abundance in the marine environment (Cloern 1996). Differences in growth rate among individual strains have clear genetic underpinnings as shown in genotype-by-environment experiments (Brand and Guillard 1981; Brand et al. 1981). This classical physiological work is supported by ever-increasing evidence from diatom transcriptomic data, demonstrating the diversity of molecular and metabolic responses of diatoms to stress factors such as nutrient limitation across species and individuals (Bowler et al. 2008; Dyhrman et al. 2012; Hwang et al. 2008; Mock et al. 2008; Sapriel et al. 2009). Whereas these genetic underpinnings of growth rate may be apparent in the laboratory, the extent to which they influence patterns of population-genetic diversity in the field is little understood. Exploring the temporal variations in diatom population-genetic structure provides a link to understanding the interactions between selection pressures of the ocean system and the physiological underpinnings of diatom diversity.

Although the structure and variability of diatom populations over time are little understood, a few key studies have used molecular tools to provide important insight. The first molecular study of diatoms used allozymes to demonstrate that the diatom *Skeletonema costatum* was subdivided into two populations differentially present during spring and fall in Narragansett Bay, RI, USA (Gallagher 1980). Whereas these allozyme 'populations' are now thought to be reproductively isolated species, this work was the first to demonstrate two characteristics of diatom species that are gaining

support. First, diatoms exhibit a high potential for cryptic diversity within 'super species' previously assumed to be cosmopolitan in their distribution (Sarno, 2005). Second, diatoms exhibit evidence for sympatric isolation, where genetic isolation can occur within the same location, driven by selective pressures of the environment. For example, populations of the diatom *Ditylum brightwellii* co-occurring in space have been detected at different times of the year (Rynearson 2004; Rynearson 2006), although data on genome size differences between these populations suggest that they may be different species (Koester et al. 2010). If both *D. brightwelli* and *S.costatum* seasonal populations are, in fact, different species, this suggests that reproductive isolation of diatoms may occur in sympatry, although alternative paths of speciation cannot be ruled out.

Understanding the extent and distribution of diatom diversity over time, and in relation to variations in the marine environment, is particularly significant in the context of a changing climate. Evidence suggests that a rapidly changing climate is shifting primary productivity throughout the global ocean; these changes in ocean productivity are most likely due to shifts in the distribution of phytoplankton populations over space and time (Beaugrand et al. 2010 ; Hays et al. 2005; Karl et al. 2001). The resiliency of phytoplankton populations to climate change ultimately depends on the extent and distribution of genetic diversity (Duffy 2002; Folke et al. 2004; Ptacnik et al. 2008; Stachowicz et al. 2002; Worm et al. 2006), which dictates both their metabolic and evolutionary potential to respond to environmental stress (Bell and Collins 2008). Unfortunately, little is known about the ways in which the molecular diversity of phytoplankton interacts with changes in climate, either through

short-term metabolic acclimation, or through adaptation over evolutionary time scales. This study explores the evolutionary and adaptive potential of the species *Thalassiosira rotula*, using isolates collected over time in Narragansett Bay, RI. We compare clonal diversity calculated from multilocus genotypes with physiological variability amongst isolates to infer the flexibility of the species in response to environmental variability. The extent and distribution of diversity, as well as the strength of genetic connectivity between populations within a single diatom species, may play a significant role in determining its ability to adapt to environmental change.

The goal of this work was to explore changes in population-genetic diversity over time in *Thalassiosira rotula,* a bloom-forming member of the phytoplankton community within the highly productive, well-studied estuary, Narragansett Bay, Rhode Island. Isolates from within Narragansett were compared with isolates collected from an adjacent coastal site (off of Martha's Vineyard, MA) over three years in order to better understand changes in genetic composition over time and in relation to local spatial connectivity and environmental variability.

METHODS

Field samples and isolates

To examine changes in population-genetic diversity over time, 442 isolates of *T. rotula* were collected from Narragansett Bay, RI (41.56 long. -71.38 lat.) at fourteen time points between 2009 and 2011, and 147 were collected from Martha's Vineyard Coastal Observatory (MVCO) (41.29 long. -74.55 lat.) at five time points between 2009 to 2011 (Table 1, Figure 1). Whole seawater samples were collected in 1 L dark Nalgene bottles from a single station in Narragansett Bay, RI ('Station 2') and a single

station off the coast of Martha's Vineyard (MVCO) (Figure 1). Both sites are part of time series that collect data on phytoplankton diversity, chlorophyll *a,* and physical parameters. Because samples aligned with time series measurements, isolates were associated with metadata including species-specific phytoplankton abundance, and temperature, salinity, and nutrient concentrations. In seawater samples where *T. rotula* could be identified morphologically, single cells or chains were isolated, cultures maintained, and DNA extracted according to Whittaker *et al.* (Whittaker et al. 2012). In brief, surface seawater was concentrated over a 20 μm mesh net. For each field sample where *T. rotula* was present, up to 96 single cells or chains were isolated from this >20 μm size fraction using a stereo microscope (Olympus SZ61), rinsed in sterile seawater three times, and transferred to 1 mL sterile Sargasso seawater amended with f/20 nutrients (Guillard 1975). Isolates were incubated at either at 4, 10, 14, or 20°C according to closest surface seawater temperature (SST) at isolation, on a 12:12-h light:dark cycle of 90-120 μmol photons m^{-2} s^{-1}. Upon reaching approximately 1000 cells ml^{-1} (1 to 3 weeks depending on growth rate), and upon confirmation that the cultures were free of algal contamination, cells were filtered and DNA extracted and stored at -80°C until further analysis.

In addition to the collection of field isolates, whole seawater samples were processed for whole community analysis in two ways: First, concentrated samples were fixed using a 2% acid Lugol's solution, and stored in 20 mL scintillation vials. Second, whole seawater was filtered onto three 0.2 μm polyester filters (Millipore), at 100 mLs per filter, and stored at -80°C.

An experiment was conducted to test whether or not genetic diversity in a 1L

bottle changed due to sample shipping, and is explained in depth in Chapter 3 (Whittaker, in prep). In short, two leters of water were collected from Narragansett Bay on January 26, 2010. One liter was immediately processed (as described above), and the second liter was shipped to Washington, USA and back to Rhode Island, USA at a transit time of four days. Upon returning to Rhode Island, 48 single cells were isolated from the shipped sample. These two samples are named NBf and NBg in this study, to keep all sample names temporally consistent. The same samples are named NBa and NBb in Chapter 2, again, for internal consistency with time.

Strain physiology and culture conditions

Twelve strains of *T. rotula* were collected in August, 2013 when sea surface temperature (SST) was 22.3°C, and their growth rates were determined at five temperatures (4, 10, 17.5, 22.3, and 30°C), and under 24 hr light at120 μmol photons m^{-2} s^{-1}. Cultures were allowed to acclimate until no differences in maximum growth rate were observed between transfers (about 10 generations). Maximum acclimated growth rates were determined following Rynearson and Armbrust (Rynearson and Armbrust 2000) and Brand (Brand et al. 1981). Briefly, the *in vivo* fluorescence of semi-continuous batch cultures was measured daily using a 10AU Field Fluorometer equipped with the *in vivo* chlorophyll optical kit (Turner Designs). The maximum acclimated growth rate for each isolate was determined by regressing the change in the log of fluorescence over time and testing the equality of slopes from at least three serial cultures (Analysis of Variance $\alpha = 0.05$). If slopes of serial growth curves were homogeneous, the average regression coefficient was used to estimate the common slope, which represented the average acclimated growth rate. Analysis of variance

(ANOVA), Sheffe's Test (Enderlein 1961) and the Tukey multiple comparison test (Tukey 1949) were used to determine the significance of differences among all isolates, and among isolates at different temperatures. Alpha was set to 0.05 for all statistical tests.

Species identification and genotyping

To confirm species-level identity of Narragansett Bay and Martha's Vineyard isolates, the T-RFLP procedure was performed, as described in Whittaker and Rynearson, 2014 (in prep). To quantify genetic variation among isolates at both the population and lineage levels, isolates were genotyped at six microsatellite loci (TR1, TR3, TR7, TR8, TR10, TR27). All six loci were amplified in the gDNA of *T. rotula* isolates using the PCR conditions described in Whittaker and Rynearson (in prep). Alleles were scored using an ABI 3130*xl* (TR1, TR3, TR7) or an ABI 3730*xl* (TR8, TR10, TR27), and analyzed using the software Gene Mapper 5 ® (Life Technologies).

Analysis of microsatellite alleles

Summary statistics for each locus at each site and time point were generated using the Excel Microsatellite Toolkit v. 3.1.1 (Park 2001), and GenePop v4.2 (Raymond and Rousset 1995). These statistics included observed and expected heterozygosity for each locus, and across all loci, as well as polymorphism information content (PIC) values for each locus, and across all loci. Total number of alleles per locus and per site was also calculated. The number of unique genotypes in the 589-isolate sample set (also known as the G:N ratio) for each time point, and each site, was calculated by identifying matching multilocus genotypes (MGLs). Deviations from Hardy-Weinberg equilibrium (HWE) were calculated in Genepop v4.2

(Raymond and Rousset 1995). Inbreeding coeficients F_{IS} were generated in GenAlEx version 6.5 (Peakall and Smouse 2012). Problems with multiple tests were alleviated by calculating corrected p-values via the Bonferroni technique.

Statistical analysis of population structure

The significance of differentiation among time points and between sites was determined using the exact G test and Fisher's exact probability test (Raymond 1995)in GENEPOP v4.2(Raymond and Rousset 1995). Pairwise tests between samples were conducted separately for each locus, and across all loci. GENEPOP and GenAlEx were used to calculate pairwise F_{ST} among samples. Further estimations of population structure including G'_{ST}, and G''_{ST} were calculated in GenAlEx 6.5 (Peakall and Smouse 2006, 2012). All p-values were corrected using the Bonferroni method. Analysis of molecular variance (AMOVA) was calculated using GenAlEx 6.5, testing for the presence of genetic structure at hierarchical space and time scales within samples, between samples, and between sites. The software Bayesian clustering program STRUCTURE v2.2 (Pritchard et al. 2003) was performed on the full dataset including all time points and both sites. Populations (K) from 2 to 19 were tested in 10 independent runs using the admixture model, correlated allele frequencies, with no prior population identification. The program was run with a burn-in of 100,000, and 100,000 repetitions. Nineteen populations were chosen as the maximum K, based on the total number of samples analyzed. Estimates of the number of populations (K) were calculated based on Pritchard *et al*. (Pritchard et al. 2000; Pritchard et al. 2003) and Evanno *et al*. (Evanno et al. 2005). The Greedy algorithm in CLUMPP v.1.11 (Jakobsson and Rosenberg 2007) was used to evaluate agreement between

independent STRUCTURE runs, and used to arrange cluster labels. DISTRUCT v1.1 (Rosenberg 2004) was used to visualize results of STRUCTURE. Population structure over time, between sites, and between sites over time was further visualized using principal coordinates analysis (PCoA) in GenAlEx 6.5. These analyses were used to assess changes over time within Narragansett Bay, but also used to determine the extent and scale of population connectivity between Narragansett Bay and its close geographic neighbor MVCO.

Environmental and Ecological Correlations

Correlations between pairwise genetic distance and environmental distance among samples were tested. Metadata were provided by the Phytoplankton of Narragansett Bay time series and Martha's Vineyard Coastal Observatory (MVCO). Parameters examined for all samples included the following: sea surface temperature (SST), salinity, chlorophyll *a*, temperature, and irradiance. Nutrient data were available for ten out of fourteen Narragansett Bay samples, and were included in an additional analysis. Principal components analysis (PCA) was used to ordinate environmental distances among samples. In addition, cell abundance of 245 phytoplankton species, measured by the Phytoplankton of Narragansett Bay time series for each sample, was used to compare community composition with intraspecific genetic composition of *T. rotula.* Community diversity metrics such as species richness, Shannon Index, and Chao1, and Simpson Index were calculated in EstimateS (Colwell 2010).

A series of nonparametric Mantel tests was performed to test correlations between genetic distance and variables of environmental distance among samples.

The statistical test known as BIOENV (Clarke and Ainsworth 1993) was performed to identify those factors, or combination of factors, of the ocean environment most correlated with a pairwise F_{ST} matrix over time both within and between sites. BIOENV uses the Euclidean distance among sites for all combinations of environmental factors, including temperature, solar irradiance, salinity, chlorophyll *a*, and cell abundance, and compares those matrices with the F_{ST} matrix to determine which combination of factors best correlates with genetic differentiation among sites. An additional BIOENV was performed for ten Narragansett Bay samples for which nutrient data were available. For these samples, the combination of the variables temperature, salinity, solar irradiance, chlorophyll *a,* cell abundance, nitrate, phosphate, and ammonia concentrations were tested against a pairwise F_{ST} matrix.

RESULTS

Isolate collection and strain physiology

Twelve isolates, collected in August, 2013 at 22.3ºC, were used to examine the extent of physiological adaptation to their temperature of isolation. Growth rates of strains incubated at five temperatures, 4ºC, 10ºC, 17.5ºC, 22.3ºC, and 30ºC exhibited significant differences (Figure 2). No strain grew at 30ºC. Lowest growth rates occurred at 4ºC, with an average μ of 0.22 ± 0.11, differing significantly from all other temperature tested ($p<0.01$). No significant differences in growth rate were observed between strains grown at 10ºC and 17.5ºC (AMOVA, $p=0.43$). Average growth rates at 10ºC and 17.4ºC were 0.40 ± 0.063 and 0.43 ± 0.10, respectively. Significant differences in growth rate were observed between isolates grown at 10ºC and 22.3ºC, and 17.5ºC and 22.3ºC ($p<0.01$). Strains exhibited the highest growth rates at 22.3ºC,

with average μ of 0.70 ± 0.14, approximately 60% times faster than when grown at 10ºC or 17.4ºC, and 350% times faster than when grown at 4ºC. Highest coefficient of variation (CV) among growth rates was observed at 4ºC, with a CV of 48%. Lowest CV was observed at 10ºC (15%), and second lowest CV was observed at 22.3ºC (20%). Every isolate exhibited a unique physiological signature across all temperatures (e.g., no isolate grew the same as any other across all temperatures) (Figure 3).

In the collection of 589 isolates of *T. rotula* from Narragansett Bay and Martha's Vineyard across 19 time points between 2009 and 2011 (Table 1), isolation success was high, 99% on average (Table 1).

Population diversity

Five hundred eighty-nine isolates of *T. rotula* were genotyped at 6 microsatellite loci (TR1, TR 3, TR 7, TR 8, TR10, and TR27). Polymorphism information content (PIC) values were highest for TR10 (0.92), and were 0.86, 0.51, 0.57, 0.65, and 0.81 for loci TR8, TR1, TR3, TR7, and TR27, respectively. In total, 544 alleles were identified in 589 isolates, at an average of 90 per locus. Average number of alleles per locus per sample was 15.4 (Table 2). Across all loci, expected heterozygosity was as high as 0.86 (Table 3). Observed heterozygosity was consistently lower than expected for all sites, and across all loci (Table 2, Table 3, Figure 4). In fact, significant heterozygote deficiencies were observed across most loci and samples. Highest expected heterozygosity was observed in Narragansett Bay isolates collected in February, 2009 (NBe). This did not correspond to the highest observed heterozygosity, which occurred in February, 2010 (NBh). Lowest expected

heterozygosity was observed in isolates collected in January, 2011 from Martha's Vineyard (MVd). Lowest observed heterozygosity was demonstrated by isolates collected from Martha's Vineyard in October, 2009 (MVa). High levels of clonal diversity were observed—out of 589 individuals sampled only two had matching multi-locus genotypes.

Population Structure

Throught this paper, a 'population' is defined as any group of samples that exhibit significant levels of divergence (F_{ST}), as determined using analysis of molecular variance, based on measures of allele frequency and heterozygosity. This general definition differs from STRUCTURE populations (K), which were determined using a Bayesian clusterina algorithm. STRUCTURE 'populations,' or clusters, will only be reference in the discussion of STRUCTURE analysis output.

Allelic and genotypic tests detected significant population structure among individuals collected from Narragansett Bay and Martha's Vineyard over time, and between sites. Measurements of F_{ST}, an index of pairwise genetic divergence between samples, were as high as 0.194 (Table 4). Not all time points were significantly different. Isolates from the two sites, Narragansett Bay and Martha's Vineyard, were not clearly subdivided by location. Analysis of molecular variance (AMOVA) demonstrated that only 1% of genetic variance was distributed among Narragansett and Martha's Vineyard sites (Table 5). PCoA was used to ordinate samples in a 2D space based on genetic distance, or F_{ST} (Figure 5). Axis 1 and 2 explained 52% and 23% of genetic variability, respectively. Principal coordinates analysis (PCoA) further demonstrated that Narragansett Bay and Martha's Vineyard samples were not

significantly structured based on sample location. Samples were also not clustered by season. For example, Narragansett Bay samples collected in fall/winter of 2010 grouped separately from Narragansett Bay those collected in fall/winter of 2009. Isolates collected in Narragansett Bay in January, 2009 (NBA) were statistically different (Fisher's exact test, $p<0.001$) from isolates collected in Narragansett Bay in January, 2011 (NBN). Samples collected only one day apart in January, 2011 from Martha's Vineyard and Narragansett Bay (NBN and MVE) were not significantly different. There was significant divergence amongst individuals to assign them to three different groups as defined by STRUCTURE, the Bayesian clustering program (Figure 6). Individuals were clustered into K=3 groups, as determined using the Evanno *et al* method [41]. The structure results suggest considerable levels of admixture between the three groups (Figure 6). Individuals from NBf, NBg, NBh, and NBi have high membership in cluster 1 (yellow). Samples NBa, NBb, NBc, NBd, NBe, and MVa, MVb, and MVc show high membership in cluster 2 (blue). NBm, NBl, and NBm show considerable admixture among cluster 3 (green), yellow, and blue clusters. All other samples including NBj, NBk, and MVd, showed the highest membership in the green cluster. These unbiased groups agree with population structure outlined in the PCoA (Figure 5).

Environmental and Ecological Correlations

Significant correlations were found between environmental and genetic variability over time. Environmental and ecological conditions varied considerably at both sites throughout the sampling period. Over the course of 3 years, samples collected at temperatures ranging from 1°C to 19.5°C in Narragansett Bay, and from

2°C to 14.1°C in Martha's Vineyard. Chlorophyll *a* was also quite variable, ranging from 0.6 μgL^{-1} to 22.1 μgL^{-1} in Narragansett Bay, and from 3.98 μgL^{-1} to 0.1 μgL^{-1} in Martha's Vineyard. The Euclidean ordination of differences in salinity, chlorophyll *a,* irradiance, and temperature are shown using PCA (Figure 7). BIOENV statistical analysis determined that 59% of variability in pairwise genetic distances between sites can be explained by the combination of irradiance and chlorophyll *a* differences ($p = 0.01$) (Table 6). However, chlorophyll *a* alone accounts for 58% of variability, and is thus the most important correlate. Ten Narragansett Bay samples were associated with nutrient data. Pairwise F_{ST} between these samples were included in a second BIOENV analysis, testing the most correlated combination of variables including: temperature, salinity, irradiance, chlorophyll *a,* nitrate, phosphate, nitrite, nitrate + nitrite, and ammonia. The highest correlation was observed using a combination of variables including ammonia, phosphate, chlorophyll *a,* and salinity, explaining 69% of genetic variability ($p = 0.04$) (Table 7). But, again, chlorophyll *a* demonstrated the highest correlation with genetic distance, explaining 67% of genetic variability, far exceeding the correlation with any other single variable. Cell counts of 245 species for each sample in Narragansett Bay was attained from the Phytoplankton of Narragansett Bay time series. Species richness was calculated for each sample. To compare whole-community diversity with intraspecific diversity of *T. rotula*, I ran a regression between *T. rotula* heterozygosity and whole-community species richness; this correlation was significant with an R^2 of 0.36 ($p = 0.04$) (Figure 8).

DISCUSSION

Observation of population structure over time in *T. rotula* provides another example of the extensive diversity nested within diatom species. This fine-scale analysis of the *T. rotula* population succession in Narragansett Bay demonstrates the following: a) High physical connectivity coupled with the vast dispersal potential of *T. rotula* supports genetic connectivity between Narragansett Bay and waters off the coast of Martha's Vineyard. b) Each population is made up of many clonal lineages which are physiologically diverse, contributing to the adaptive potential of populations in response to changes in the marine environment c) There was no correlation between changes in population structure and seasonal changes. Instead, competitive exclusion and community interactions may play an important role in structuring populations over time at the local spatial scale.

Population Succession

The variability in diversity over time in Narragansett Bay and Martha's Vineyard is best described in terms of the succession of populations. Eight distinct populations were observed throughout the sampling period. At the start of the period, in January 2009, *T. rotula* cells were isolated from Narragansett Bay at three time points (NBa-d). They represented a unique population that remained in the bay for the entire month. In February, a second population was detected (NBe) closely related to samples collected in January. On October 15, 2009, cells of *T. rotula* were isolated from Martha's Vineyard (MVa). They represented another unique population, closely related to those samples in Narragansett Bay collected in January and February 2009. This population was detected in Martha's Vineyard again on the 22nd of October (MVb). Martha's Vineyard isolates sampled on November 7th, 2009 (MVc) were not

significantly different from samples collected in Narragansett Bay throughout the month of January 2009. The STRUCTURE Bayesian clustering program identified these three '2009' populations as a single group. The next winter in 2010, two entirely different populations were detected in Narragansett Bay, identified as a single cluster in STRUCTURE; the first population was detected in January 2010 in Narragansett Bay (NBf, NBg), and the subsequent population persisted in the Bay from February to March (NBh, NBi). Later that year, beginning in September, three additional populations were detected in Narragansett Bay and Martha's Vineyard, more closely related to one another than any previous population detected; this was confirmed by their STRUCTURE clustering, which demonstrated very little admixture with the other two clusters. The first population was detected in Narragansett Bay in September. (NBj). A slightly divergent population was detected in Martha's Vineyard on October 14th, 2010 (MVd), and two days later Narragansett Bay on October 18th, 2010 (NBk). These populations were followed by a final shift in late October, which was detected in Narragansett Bay (NBl). In Narragansett Bay, samples from early November 2010 (NBm) grouped with populations collected in January, 2009, but had shifted back to the population containing samples NBj and MVe by January (NBn). Samples from January 10, 2011 from Martha's Vineyard and January 11, 2011 from Narragansett Bay clustered in the same population.

What we can infer from this pattern of succession is a) Waters off the coast of Martha's Vineyard and Narragansett Bay appear to be highly connected and undifferentiated b) Shifts in populations can occur rapidly over time within these sites

c) Population structure is unrelated to seasonal variability, and instead varies with environmental or ecological changes over shorter periods of time.

Potential sources of populations

While inferring population sources is a major challenge for many high dispersal marine organisms (Roman and Darling 2007), there are several feasible scenarios that would allow for different populations to become established over short periods of time. First, cells of any given population may be below detection, or rare in the water column until environmental or ecological conditions trigger their rapid growth, and our detection of them (Caron and Countway 2009). Second, dormant cells in the form of resting spores or cysts may remain viable in the sediment; these may be a source for populations that arise only when cues for their growth are triggered, or when conditions of the surface water are favorable (Mcquoid and Hobson 1995). Third, high connectivity with coastal environments, and high dispersal potential, may allow for colonization and establishment of new populations in each site when conditions are favorable such that they can outcompete and replace incumbent populations. Reduced diversity may be a signature of colonization events, where recently introduced populations exhibit lower diversity than source populations (Davies et al. 1999). Heterozygote deficiency was a general characteristic of all samples, and observed across loci. Diversity was lower in samples collected in October 2009 off the coast of Martha's Vineyard (MVa), and in January 2011 in Narragansett Bay (NBn). For instance, for MVa, observed heterozygosity was as low as 0.36. Nonetheless, expected heterozygosity was generally quite high for all samples. Because there was no consistent time lag between the appearance of populations in Narragansett Bay or

Martha's Vineyard, and diversity did not differ consistently between samples from each site, it seems that neither is a consistent source of populations for the other. Instead, favorable conditions may select for the growth of individuals from different populations existing in the rare biosphere or sediment at each site. In fact, it has been observed in the diatom *Skeletonema marinoi* that genetic diversity in the water column is tightly coupled with that of the sediment (Godhe and Härnström 2010).

Historical *T. rotula* dynamics in Narragansett Bay

Population structure in *T. rotula* observed over three years of sampling was insightful when placed in the context of long-term patterns of abundance in Narragansett Bay, thanks to a historical record of phytoplankton species abundance dating back to 1950s. Since this time, *T. rotula* has demonstrated a bimodal distribution of abundance with temperature, consistent over many decades (Krawiec 1982). Whereas the timing of *T. rotula* presence in the bay is somewhat sporadic from year to year, and low to medium abundances of the species occur throughout the year, the highest abundances of the species occur at 4ºC and 17ºC, respectively, with a dip in abundance at 10ºC (Krawiec, 1978). Isolates in this study were collected across a broad range of temperatures in the Bay, from 1ºC to 19.5ºC. *T. rotula* populations were not subdivided according to temperature in Narragansett Bay, suggesting that the bimodal peak in abundance with temperature does not reflect the differential growth of two separate populations, as has been demonstrated by *Skeletonema costatum* in the bay (Gallagher 1980). In addition, it has been shown from *T. rotula* autecology experiments that growth rates are consistently higher at 10ºC than at 4ºC (Krawiec 1982); this demonstrates that a dip in abundance at 10ºC is not necessarily related to a

reduction in cell-specific growth rate at this temperature. Taken together, these data suggest that the pattern of bimodal abundance of *T. rotula* in Narragansett Bay does not relate to genetic or physiological characteristics of the species. Instead, factors such as grazing or physical mixing may impact the low accumulation cells in the bay at 10ºC, and high abundance of cells at 4ºC and 17.5ºC. In fact, grazing by protists has been shown to remove up to 200% of daily phytoplankton production in Narragansett Bay (Lawrence and Menden-Deuer 2012). These data also suggest that cell abundance of *T. rotula* in the field may not relate directly to differences in growth rate, or the physiological differences of populations observed in this study. In fact, no relationship between cell abundance and population structure was observed here. Instead, cell abundance in the bay may relate more strongly to predation or stratification in the bay, which can dictate predator-prey encounter rates (Gerritsen and Strickler 1977). In addition, cell abundance in the field at different temperatures may not directly relate to the physiological adaptation of individuals or populations to those environmental conditions.

Physiological and genetic coupling

To test the relative physiological adaptation of populations or individuals to different environments, we measured the growth rate of individuals at five different temperatures, reflecting the temperature range of Narragansett Bay. Twelve isolates collected at 22.3ºC were grown at 4ºC, 10ºC, 17.5ºC, 22.3ºC, and 30ºC. These data confirm previous autoecology work by Krawiec (Krawiec 1982) showing that *T. rotula* grows more slowly at 4ºC than at all other temperatures measured. No difference in growth rate was observed between isolates grown at 10ºC and 17.5ºC,

suggesting that the difference in these temperature conditions may not trigger rapid increases in growth rate in these isolates, which were collected at higher temperatures (22.3ºC). In fact, individuals grew on average 60% faster at their temperature of isolation (22.3ºC) than at other temperatures, and none grew at 30ºC. This suggests that these isolates may be particularly adapted to the temperature from which they were collected, and that this condition may be an environmental trigger to the increased growth rates allowing for their detection in the field.

Isolates exhibited dramatically different patterns in growth among individuals (Figure 3). In fact, every isolate exhibited a unique physiological footprint across temperatures tested. This physiological diversity reflects the genetic and clonal diversity observed within *T. rotula* isolates. For instance, we observed 99.6% unique clones amongst samples collected here; out of 589 isolates genotyped, only 2 were sampled with matching multi-locus genotypes. Thus, *T. rotula* clonal lineages harbored within each population are associated with high physiological diversity. It has been hypothesized that the rapid and frequent environmental change in the ocean contributes to high levels of clonal diversity, preventing the dominance of any single clonal lineage (Rynearson 2005); this holds true only if clonal diversity is related to physiological diversity, as observed here. The magnitude of variation in growth rate amongst *T. rotula* isolates, related to clonal diversity, may explain the rapid changes in the composition of populations in the field; this has also been demonstrated for the diatom *D. brightwellii* (Rynearson 2004). Narragansett Bay isolates show a clear coupling between the scales of genetic and physiological diversity; this suggests that

populations harbor enormous adaptive potential, offering ecological flexibility in the highly variable marine environment.

Inter vs. intraspecific diversity

Species abundance data for all Narragansett Bay samples allowed for comparisons between species-level community composition of the Bay and the genetic composition of *T. rotula* over time. It has been hypothesized that the species-level diversity of field communities may correlate with the genetic diversity within species; this relationship is thought to exist due to the similar selective pressures structuring diversity and driving evolution on both the inter and intra-specific levels (Vellend 2005; Vellend and Geber 2005). In addition, processes such as mutation, drift, and migration may be acting in parallel on the intraspecific and species levels. Here, we observed a significant correlation between species richness and intraspecific heterozygosity (genetic diversity). This correlation suggests that parallel factors are responsible for structuring and maintaining the diversity of species in the field community, as well as the genotypic diversity within species. The heterogeneity of the marine environment, especially in coastal temperate ecosystems like Narragansett Bay, most likely provides ecological niches highly variable over space and time that are filled on different scales by diverse species and genotypes. The fact we observed this correlation in a planktonic organism, constantly impacted by a dynamic fluid environment, suggests that the mechanisms maintaining diversity and driving evolution may be consistent both within species, and on the whole-community level.

Ecological correlates of genetic structure

Although changes in genetic composition over time showed no correlation with seasons, years, or space, a strong relationship was observed between chlorophyll *a* concentration and population structure. In fact, chlorophyll concentration explained 53% of the variance in divergence among samples ($p=0.01$). Chlorophyll is related to productivity (Behrenfeld et al. 2005; Cloern et al. 1995), and changes in its concentration in the field can reflect a generalized response of phytoplankton to an influx in nutrients, temperature, stratification, as well as loss factors such as grazing and mixing (Marra et al. 1990; Riemann et al. 1989). The relationship between chlorophyll and genetic structure suggests that biological interactions such as competitive exclusion and resource competition during bloom periods may dictate the relative success of different populations over time. Different populations may be better adapted to different competitive regimes, such as those occurring during bloom (high chlorophyll) or non-bloom periods (low chlorophyll). The role of resource competition in structuring the diversity of microbial communities is widely discussed (Follows and Dutkiewicz 2011; Fuhrman 2009); it makes sense that these dynamics may also impact the intraspecific genetic structure of these organisms. Because population succession was less correlated with single abiotic variables such as temperature and salinity, this suggests that a more complex interaction of environmental and ecological factors may dictate the favorable conditions under which different populations can thrive. It is this complex interaction of ecological and environmental variables, fluctuating with high frequency over time that may play a significant role in structuring populations over time, and dictate the frequency of interbreeding among individuals from those populations.

CONCLUSIONS

Isolates from Narragansett Bay and off the coast of Martha's Vineyard exhibit enormous adaptive potential, suggesting that populations may be evolutionarily flexible to environmental change over time. The heterogeneity of the environment undoubtedly supports and maintains the high levels of clonal and population-level diversity observed here. In a changing climate, reduction in environmental variability due to loss of seasonal extremes may reduce the strength of those evolutionary forces supporting diversity within communities and species. Yet, we know that phytoplankton can evolve rapidly, and quickly fill new ecological niches (Falkowski 2004; Fussmann et al. 2003; Litchman and Klausmeier 2008; Lohbeck et al. 2012). Here, rapid growth rates, accompanied by high physiological variability, were most likely responsible for the rapid response and succession of populations. These populations harbor high levels of physiological diversity, which most likely relates to their relative production in the field. This is particularly significant in terms of the timing and productivity of phytoplankton blooms. The formation of phytoplankton blooms is, at times, sporadic, and triggers of bloom formation little understood. *T. rotula* demonstrates that populations can respond rapidly to changes in the local environment, whether that be interactions with other species, a release of nutrient stress, or changes in abiotic factors. Here, population structure over time was unrelated to seasonal changes. This work shows that population succession may be tightly coupled with ecological interactions in the field, dictated just much by resource competition and competitive exclusion as the fluctuation of abiotic factors in the environment.

ACKNOWELDGEMENTS

This research was supported by National Science Foundation grants, NSF 0727227 (to TAR) and NSF SBE0245039 (to URI, TAR). Part of the research was conducted using instrumentation supported by NSF-EPSCoR grants 0554548 and 1004057. We thank the following people who took part in seawater collection in 2010: Caitlin Lawrence Heidi M. Sosik, and Taylor Crockford. Environmental data was generously provided by monitoring institutions including the Narragansett Bay Phytoplankton Time Series, Martha's Vineyard Coastal Observatory.

Beaugrand, G., M. Edwards, and L. Legendre, 2010. Marine biodiversity, ecosystem functioning, and carbon cycles, p. 10120-10124.
Behrenfeld, M. J. 2010. Abandoning Sverdrup's Critical Depth Hypothesis on phytoplankton blooms. Ecology **91:** 977-989.
Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea. 2005. Carbon‚Äêbased ocean productivity and phytoplankton physiology from space. Global Biogeochemical Cycles **19**.
Bell, G., and S. Collins. 2008. Adaptation, extinction and global change. Evolutionary Applications **1:** 3-16.
Bowler, C. and others 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456:** 239-244.
Brand, L. E., and R. R. L. Guillard. 1981. The effects of continuous light and light intensity on the reproduction rates of twenty-two species of marine phytoplankton. Journal of Experimental Marine Biology and Ecology **50:** 119-132.
Brand, L. E., R. R. L. Guillard, and L. S. Murphy. 1981. A method for the rapid and precise determination of acclimated phytoplankton reproduction rates. Journal of Plankton Research **3:** 193-201.
Caron, D. A., and P. D. Countway. 2009. Hypotheses on the role of the protistan rare biosphere in a changing world. Aquatic Microbial Ecology **57:** 227.
Casteleyn, G. 2009. Lack of population genetic structuring in the marine planktonic diatom Pseudo-nitzschia pungens (Bacillariophyceae) in a heterogeneous area in the Southern Bight of the North Sea. Marine biology **156:** 1149-1158.
Clarke, K., and M. Ainsworth. 1993. A method of linking multivariate community structure to environmental variables. Marine Ecology-Progress Series **92:** 205-205.
Cloern, J. E. 1996. Phytoplankton bloom dynamics in coastal ecosystems: A review with some general lessons from sustained investigation of San Francisco Bay, California. Reviews of Geophysics **34:** 127-168.
Cloern, J. E., C. Grenz, and L. Videgar-Lucas. 1995. An empirical model of the phytoplankton chlorophyll: carbon ratio-the conservation factor between productivity and growth rate. Limnology and Oceanography **40:** 1313-1321.
Colwell, R. 2010. EstimateS: statistical estimation of species richness and shared species from samples. 2009. Version 8.2. User's guide and application.
D'alelio, D., A. Amato, W. H. Kooistra, G. Procaccini, R. Casotti, and M. Montresor. 2009. Internal transcribed spacer polymorphism in Pseudo-nitzschia multistriata (Bacillariophyceae) in the Gulf of Naples: recent divergence or intraspecific hybridization? Protist **160:** 9-20.
Diaz, S., J. Fargione, F. S. Chapin, Iii, and D. Tilman. 2006. Biodiversity Loss Threatens Human Well-Being. PLoS Biol **4:** e277.
Davies, N., F. X. Villablanca, and G. K. Roderick. 1999. Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. Trends in Ecology & Evolution **14:** 17-21.

Doney, S. C., M. R. Abbott, J. J. Cullen, D. M. Karl, and L. Rothstein. 2004. From genes to ecosystems: the ocean's new frontier. Frontiers in Ecology and the Environment **2:** 457-468.

Duffy, J. E. 2002. Biodiversity and ecosystem function: the consumer connection. Oikos **99:** 201-219.

Dyhrman, S. T. and others 2012. The transcriptome and proteome of the diatom Thalassiosira pseudonana reveal a diverse phosphorus stress response. PloS one **7:** e33768.

Enderlein, G. 1961. Scheffé, H.: The Analysis of Variance. Wiley, New York 1959, 477 Seiten, $ 14,00. Biometrische Zeitschrift **3:** 143-144.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology **14:** 2611-2620.

Falkowski, P. G. 2004. The Evolution of Modern Eukaryotic Phytoplankton. Science (New York, N.Y.) **305:** 354-360.

Folke, C. and others 2004. Regime Shifts, Reslilience, and Biodiversity in Ecosystem Management. Annual Review of Ecology, Evolution, and Systematics **35:** 557-581.

Follows, M. J., and S. Dutkiewicz. 2011. Modeling diverse communities of marine microbes. Annual review of marine science **3:** 427-451.

Fuhrman, J. A. 2009. Microbial community structure and its functional implications. Nature **459:** 193-199.

Fussmann, G. F., S. P. Ellner, and N. G. Hairston. 2003. Evolution as a critical component of plankton dynamics. Proceedings of the Royal Society of London. Series B: Biological Sciences **270:** 1015-1022.

Gallagher, J. C. 1980. Population genetics of *Skeletonema costatum* (Bacillariophyceae) in Narragansett Bay. Journal of phycology **16:** 464-474.

Gerritsen, J., and J. R. Strickler. 1977. Encounter probabilities and community structure in zooplankton: a mathematical model. Journal of the Fisheries Board of Canada **34:** 73-82.

Godhe, A., and K. Härnström. 2010. Linking the planktonic and benthic habitat: genetic structure of the marine diatom Skeletonema marinoi. Molecular ecology **19:** 4478-4490.

Guillard, R. R. 1975. Culture of phytoplankton for feeding marine invertebrates. Plenum Press.

Härnström , K., M. Ellegaard, T. R. J. Andersen, and A. Godhe. 2011. Hundred years of genetic structure in a sediment revived diatom population. Proceedings of the National Academy of Sciences **108:** 4252-4257.

Hays, G. C., A. J. Richardson, and C. Robinson. 2005. Climate change and marine plankton. Trends in Ecology & Evolution **20:** 337-344.

Hwang, Y.-S., G. Jung, and E. Jin. 2008. Transcriptome analysis of acclimatory responses to thermal stress in Antarctic algae. Biochemical and biophysical research communications **367:** 635-641.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23:** 1801-1806.

Karl, D. M., R. R. Bidigare, and R. M. Letelier. 2001. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: The domain shift hypothesis. Deep Sea Research Part II: Topical Studies in Oceanography **48:** 1449-1470.

Koester, J. A., J. E. Swalwell, P. Von Dassow, and E. V. Armbrust. 2010. Genome size differentiates co-occurring populations of the planktonic diatom Ditylum brightwellii (Bacillariophyta). BMC Evol Biol **10:** 1.

Krawiec, R. W. 1982. Autecology and clonal variability of the marine centric diatom Thalassiosira rotula; (Bacillariophyceae) in response to light, temperature and salinity. Marine Biology **69:** 79-89.

Lawrence, C., and S. Menden-Deuer. 2012. Drivers of protistan grazing pressure: seasonal signals of plankton community composition and environmental conditions. Marine Ecology Progress Series **459:** 39-52.

Litchman, E., and C. A. Klausmeier. 2008. Trait-based community ecology of phytoplankton. Annual Review of Ecology, Evolution, and Systematics **39:** 615-639.

Lohbeck, K. T., U. Riebesell, and T. B. Reusch. 2012. Adaptive evolution of a key phytoplankton species to ocean acidification. Nature Geoscience **5:** 346-351.

Marra, J., R. Bidigare, and T. Dickey. 1990. Nutrients and mixing, chlorophyll and phytoplankton growth. Deep Sea Research Part A. Oceanographic Research Papers **37:** 127-143.

Mcquoid, M. R., and L. A. Hobson. 1995. Importance of resting stages in diatom seasonal succession Journal of Phycology **31:** 44-50.

Mock, T. and others 2008. Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana identifies genes involved in silicon bioprocesses. Proceedings of the National Academy of Sciences **105:** 1579-1584.

Park, S. 2001. Trypanotolerance in West African Cattle and the Population Genetics Effects of Selection.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research,Äîan update. Bioinformatics **28:** 2537-2539.

Pritchard, J., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155:** 945 - 959.

Pritchard, J. K., W. Wen, and D. Falush. 2003. Documentation for STRUCTURE software: version 2.

Ptacnik, R. and others 2008. Diversity predicts stability and resource use efficiency in natural phytoplankton communities. Proceedings of the National Academy of Sciences **105:** 5134-5138.

Raymond, M. 1995. An exact test for population differentiation. Evolution **49:** 1280.

Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. Journal of Heredity **86:** 248-249.

Riemann, B., P. Simonsen, and L. Stensgaard. 1989. The carbon and chlorophyll content of phytoplankton from various nutrient regimes. Journal of Plankton Research **11:** 1037-1045.

Roman, J., and J. A. Darling. 2007. Paradox lost: genetic diversity and the success of aquatic invasions. Trends in Ecology & Evolution **22:** 454-464.

Rosenberg, N. 2004. DISTRUCT: a program for the graphical display of structure results. Molecular Ecology Notes **4:** 137 - 138.

Rynearson, T. A. 2004. Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). Journal of phycology **40:** 34-43.

Rynearson, T. A. 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom Ditylum brightwellii. Molecular ecology **14:** 1631-1640.

Rynearson, T. A. 2006. Spring bloom development, genetic variation, and population succession in the planktonic diatom Ditylum brightwellii. Limnology and oceanography **51:** 1249.

Rynearson, T. A., and E. V. Armbrust. 2000. DNA Fingerprinting Reveals Extensive Genetic Diversity in a Field Population of the Centric Diatom Ditylum brightwellii. Limnology and Oceanography **45:** 1329-1340.

Rynearson, T. A., E. O. Lin, and E. V. Armbrust. 2009. Metapopulation Structure in the Planktonic Diatom Ditylum brightwellii (Bacillariophyceae). Protist **160:** 111-121.

Sapriel, G. and others 2009. Genome-wide transcriptome analyses of silicon metabolism in Phaeodactylum tricornutum reveal the multilevel regulation of silicic acid transporters. PLoS One **4:** e7458.

Sarno, D. 2005. Diversity in the genus Skeletonema (Bacillariophyceae). An assessment of the taxonomy of S. costatum-like species with the description of four new species. . Journal of phycology **41:** 151-176.

Schmidt, P. S. and others 2008. Ecological genetics in the North Atlantic: Environmental gradients and adaptation at specific loci. Ecology **89:** S91-S107.

Stachowicz, J. J., H. Fried, R. W. Osman, and R. B. Whitlatch. 2002. Biodiversity, invasion resistance, and marine ecosystem function: reconciling pattern and process. Ecology **83:** 2575-2590.

Sverdrup, H. U. 1953. On Conditions for the Vernal Blooming of Phytoplankton. Journal du Conseil **18:** 287-295.

Townsend, D. W., L. M. Cammen, P. M. Holligan, D. E. Campbell, and N. R. Pettigrew. 1994. Causes and consequences of variability in the timing of spring phytoplankton blooms. Deep Sea Research Part I: Oceanographic Research Papers **41:** 747-765.

Tukey, J. W. 1949. One Degree of Freedom for Non-Additivity. Biometrics **5:** 232-242.

Vellend, M. 2005. Species diversity and genetic diversity: parallel processes and correlated patterns. The American Naturalist **166:** 199-215.

Vellend, M., and M. A. Geber. 2005. Connections between species diversity and genetic diversity. Ecology Letters **8:** 767-781.

Whittaker, K., D. Rignanese, R. Olson, and T. Rynearson. 2012. Molecular subdivision of the marine diatom Thalassiosira rotula in relation to geographic distribution, genome size, and physiology. BMC Evolutionary Biology **12:** 209.

Worm, B. and others 2006. Impacts of Biodiversity Loss on Ocean Ecosystem Services, p. 787-790.

Table 1. **Samples**
Isolation date and temperature for NB and MVCO samples, as well as number of cells isolated and survived from each sample. Number of cells survived also represents those isolates that were extracted and genotyped at six microsatellite loci. NBf/g refers to samples which were collected from Narragansett Bay and either immediately processed (NBf), or shipped to Washington, USA and back (4 days) in order to test the effect of shipping time on *T. rotula* population structure.

Sample Site	Site Abbreviation	Date Collected	Cells Isolated	Cells Survived	SST at Collection
Narragansett Bay, RI, USA	NBa	1/6/09	20	19	2.7
	NBb	1/13/09	24	20	2
	NBc	1/22/09	10	10	0.6
	NBd	1/27/09	14	14	2.1
	NBe	2/9/09	31	31	2.3
	NBf/g	1/26/10	48/48	44/36	3.8
	NBh	2/2/10	20	18	1
	NBi	3/22/10	12	11	6.5
	NBj	9/27/10	48	47	19.5
	NBk	10/18/10	48	48	14.8
	NBl	10/25/10	56	56	12.9
	NBm	11/3/10	48	47	12
	NBn	1/11/11	40	40	2.2
Martha's Vineyard Coastal Observatory, MA, USA	MVa	10/15/09	40	40	14.1
	MVb	10/22/09	30	19	13
	MVc	11/7/09	32	32	11.5
	MVd	10/14/10	48	48	16.4
	MVe	1/10/11	8	8	2.3

Table 2. **Diversity Statistics**
Summary of diversity statistics and allelic richness for each site, across all 6 loci. Included are diversity statistics averaged across all loci for each site, and the respective standard deviation (SD) including unbiased expected heterozygosity (H_e), observed heterozygosity (H_o), and number of alleles (no. alleles). Total sample size is given, along with average (global) diversity statistics across the entire sample set.

Population	Sample size	H_e	H_e SD	H_o	H_o SD	No. Alleles	No. Alleles SD
NBa	19	0.73	0.10	0.43	0.05	13.33	8.31
NBb	20	0.81	0.05	0.42	0.05	14.00	5.06
NBc	10	0.79	0.09	0.44	0.07	8.67	4.27
NBd	14	0.79	0.07	0.51	0.06	8.67	4.23
NBe	31	0.86	0.04	0.38	0.04	17.17	6.18
NBf	44	0.68	0.08	0.52	0.03	15.17	14.08
NBg	36	0.69	0.10	0.56	0.03	15.33	12.71
NBh	19	0.75	0.10	0.57	0.05	10.83	7.00
NBi	11	0.81	0.07	0.50	0.07	8.33	4.23
NBj	47	0.71	0.08	0.51	0.03	19.83	11.72
NBk	48	0.70	0.12	0.47	0.03	20.50	16.72
NBL	56	0.73	0.09	0.49	0.03	22.50	12.63
NBm	47	0.78	0.07	0.48	0.03	19.00	8.85
NBn	40	0.77	0.08	0.54	0.03	19.67	10.23
MVa	40	0.84	0.07	0.36	0.04	21.50	10.56
MVb	19	0.82	0.08	0.47	0.05	14.33	8.66
MVc	32	0.77	0.09	0.40	0.04	19.17	9.22
MVd	48	0.65	0.11	0.44	0.03	18.83	11.70
Mve	8	0.72	0.12	0.61	0.08	5.83	2.99
Total/Average:	**589**	**0.76**	**0.08**	**0.48**	**0.04**	**15.40**	**8.91**

Table 3. **Diversity Statistics per Site and Locus**

Summary of diversity statistics and allelic richness for each site, at each locus and each sample. Number of individuals typed for each locus at each site (N), number of alleles (N_a), expected heterozygosity (H_e), observed heterozygosity (H_o), and inbreeding coefficient (F_{is}). F_{is} values in bold represent samples that exhibit significant heterozygote deficiency.

		Site Name																		
		Nba	**NBb**	**NBc**	**NBd**	**Nbe**	**NBf**	**NBg**	**NBh**	**Nbi**	**NBj**	**NBk**	**NBL**	**NBm**	**NBn**	**Mva**	**MVb**	**MVc**	**MVd**	**Mve**
TR1	N	19	20	10	14	30	42	36	19	10	47	48	54	40	40	27	17	27	46	7
	N_A	7	8	3	3	6	3	4	4	5	7	3	9	11	10	9	6	7	9	3
	Size range (bp)	117-275	107-297	110-129	110-129	117-157	123-129	120-129	120-129	120-129	91-129	110-129	110-132	110-259	104-174	107-141	110-173	107-144	110-147	108-114
	H_e	0.49	0.68	0.48	0.47	0.74	0.65	0.62	0.66	0.76	0.48	0.21	0.47	0.67	0.60	0.62	0.68	0.58	0.35	0.44
	H_o	0.26	0.35	0.20	0.43	0.48	0.64	0.50	0.63	0.40	0.43	0.23	0.33	0.43	0.38	0.21	0.31	0.33	0.37	0.57
	F_{IS}	**0.47**	**0.49**	**0.6**	0.08	**0.35**	0.01	**0.19**	0.04	**0.49**	**0.1**	-0.09	**0.3**	**0.37**	**0.38**	**0.66**	**0.55**	**0.43**	-0.7	-0.33
TR3	N	19	20	10	15	30	44	33	19	11	46	48	56	42	40	39	19	30	46	6
	N_A	6	13	7	9	18	5	12	11	6	14	13	10	8	9	19	6	11	12	2
	Size range (bp)	104-118	104-163	106-202	108-202	104-237	108-137	106-137	103-141	108-134	88-273	104-204	106-273	104-208	104-210	101-210	105-269	101-172	78-135	198-215
	H_e	0.58	0.72	0.58	0.72	0.78	0.62	0.77	0.86	0.80	0.50	0.53	0.54	0.50	0.45	0.65	0.49	0.41	0.31	0.89
	H_o	0.26	0.55	0.40	0.57	0.60	0.73	0.76	0.74	1.00	0.46	0.48	0.64	0.50	0.55	0.33	0.05	0.23	0.32	0.40
	F_{IS}	**0.55**	**0.24**	0.33	0.22	**0.23**	-0.17	**0.015**	**0.14**	-0.27	0.08	0.1	-0.2	-0.01	-0.22	**0.49**	**0.9**	**0.4**	-0.02	**0.58**
TR7	N	16	18	9	9	26	42	31	19	10	45	34	51	40	38	19	9	19	45	5
	N_A	7	13	11	7	21	10	7	3	4	13	12	20	20	18	17	11	20	13	6
	Size range (bp)	149-232	168-239	144-213	178-241	122-273	193-227	192-212	201-215	175-213	200-364	183-255	191-283	166-281	186-294	161-247	179-295	130-297	195-284	181-207
	H_e	0.48	0.69	1.00	0.78	0.92	0.43	0.32	0.28	0.51	0.68	0.66	0.61	0.83	0.80	0.91	0.91	0.86	0.59	0.89
	H_o	0.13	0.28	0.44	0.44	0.50	0.17	0.23	0.32	0.00	0.53	0.44	0.31	0.38	0.29	0.42	0.63	0.47	0.35	1.00
	F_{IS}	**0.75**	**0.61**	**0.54**	**0.45**	**0.46**	**0.62**	0.3	-0.13	**1**	**0.21**	**0.33**	**0.49**	**0.55**	**0.64**	**0.54**	**0.33**	**0.46**	**0.41**	**-0.13**
TR8	N	18	19	9	13	25	44	36	17	7	47	48	53	41	40	40	18	29	47	7
	N_A	22	19	13	12	18	27	22	20	8	30	29	31	30	21	35	23	24	26	10
	Size range (bp)	139-304	135-240	130-252	135-201	130-278	139-255	152-270	133-306	139-203	125-206	138-218	126-228	127-215	136-219	135-274	135-242	135-333	137-197	110-153
	H_e	0.97	0.94	0.97	0.93	0.84	0.89	0.89	0.93	0.90	0.92	0.93	0.93	0.95	0.89	0.96	0.97	0.91	0.87	0.24
	H_o	0.50	0.42	0.44	0.31	0.21	0.39	0.50	0.71	0.57	0.45	0.46	0.49	0.46	0.50	0.45	0.61	0.38	0.45	0.25
	F_{IS}	**0.49**	**0.56**	**0.56**	**0.68**	**0.75**	**0.57**	**0.44**	**0.24**	**0.38**	**0.52**	**0.51**	**0.47**	**0.51**	**0.44**	**0.54**	**0.4**	**0.59**	**0.49**	-0.04
TR10	N	16	18	9	13	30	42	35	17	10	47	48	48	41	33	31	19	31	41	5
	N_A	25	21	13	16	22	38	38	19	14	38	50	42	28	37	36	28	32	40	6
	Size range (bp)	140-346	152-458	152-285	146-317	136-268	143-291	142-295	139-263	142-267	148-295	127-359	136-417	147-377	136-391	144-354	129-375	144-381	125-359	163-344
	H_e	0.99	0.95	0.97	0.95	0.95	0.96	0.97	0.96	0.97	0.87	0.98	0.96	0.89	0.97	0.98	0.98	0.97	0.97	0.89
	H_o	0.56	0.44	0.44	0.58	0.10	0.71	0.71	0.53	0.50	0.47	0.46	0.52	0.49	0.67	0.30	0.47	0.39	0.48	0.80
	F_{IS}	**0.44**	**0.54**	**0.56**	**0.4**	**0.9**	**0.23**	**0.26**	**0.45**	**0.5**	**0.47**	**0.54**	**0.46**	**0.45**	**0.32**	**0.7**	**0.52**	**0.6**	**0.51**	0.11
TR27	N	13	10	7	9	15	41	36	17	10	41	48	54	36	38	26	11	28	48	7
	N_A	13	10	5	8	16	8	9	9	13	17	16	23	17	23	17	12	21	14	6
	Size range (bp)	150-228	143-224	192-224	192-222	160-300	157-274	157-253	188-221	125-257	174-221	163-295	182-264	182-210	162-256	181-357	160-300	157-263	166-231	158-189
	H_e	0.86	0.91	0.81	0.88	0.96	0.54	0.56	0.80	0.95	0.84	0.87	0.87	0.88	0.92	0.92	0.90	0.89	0.83	0.94
	H_o	0.85	0.50	0.71	0.75	0.40	0.49	0.64	0.53	0.50	0.73	0.77	0.63	0.64	0.84	0.42	0.73	0.61	0.71	0.63
	F_{IS}	0.01	**0.46**	0.13	0.16	**0.59**	0.1	-0.15	**0.35**	**0.49**	**0.14**	**0.12**	**0.28**	**0.28**	**0.09**	**0.54**	**0.2**	**0.32**	**0.15**	**0.35**

Table 4 . **Pairwise F_{ST}**

Multilocus F_{ST} values among pairs of populations (below diagonal). Significant results (after Bonferroni correction) are shown in bold above diagonal.

	NBf	NBg	NBh	NBl	NBj	NBm	NBk	NBl	NBn	Nba	NBb	NBc	NBd	Nbe	Mva	MVb	MVc	MVd	Mve
NBf	0.000	**0.015**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
NBg	0.010	0.000	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
NBh	0.046	0.043	0.000	0.010	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
NBl	0.082	0.061	0.027	0.000	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	0.002	**0.001**	0.002	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
NBj	0.133	0.127	0.111	0.079	0.000	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	0.003	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
NBm	0.102	0.090	0.072	0.032	0.023	0.000	**0.001**	**0.001**	**0.001**	**0.001**	0.002	0.074	0.003	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	0.005
NBk	0.152	0.136	0.127	0.089	0.028	0.045	0.000	**0.001**	**0.001**	**0.001**	**0.001**	0.007	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	0.007
NBl	0.110	0.094	0.071	0.040	0.024	0.019	0.025	0.000	**0.001**	0.003	**0.001**	0.003	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	0.003
NBn	0.132	0.120	0.091	0.059	0.026	0.020	0.032	0.013	0.000	**0.001**	**0.001**	0.009	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	0.095
Nba	0.135	0.130	0.084	0.046	0.024	0.024	0.026	0.017	0.027	0.000	0.351	0.366	0.054	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
NBb	0.133	0.129	0.084	0.045	0.039	0.019	0.062	0.040	0.041	0.001	0.000	0.385	0.213	0.002	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
NBc	0.145	0.137	0.111	0.052	0.026	0.010	0.022	0.029	0.018	0.001	0.002	0.000	0.439	0.003	0.003	0.145	0.307	0.002	0.040
NBd	0.143	0.135	0.102	0.042	0.052	0.020	0.044	0.044	0.041	0.013	0.005	0.000	0.000	0.002	0.032	0.098	0.016	**0.001**	0.008
Nbe	0.134	0.125	0.092	0.047	0.081	0.036	0.105	0.080	0.073	0.052	0.018	0.026	0.023	0.000	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
Mva	0.171	0.159	0.143	0.083	0.086	0.040	0.076	0.086	0.068	0.052	0.044	0.024	0.012	0.046	0.000	0.118	**0.001**	**0.001**	**0.001**
MVb	0.194	0.181	0.163	0.098	0.068	0.035	0.060	0.081	0.059	0.048	0.037	0.008	0.009	0.048	0.005	0.000	0.037	**0.001**	**0.001**
MVc	0.143	0.125	0.121	0.062	0.046	0.024	0.034	0.050	0.041	0.034	0.035	0.002	0.014	0.044	0.018	0.009	0.000	**0.001**	0.007
MVd	0.147	0.137	0.127	0.094	0.022	0.043	0.018	0.023	0.028	0.031	0.067	0.038	0.076	0.124	0.103	0.091	0.055	0.000	**0.001**
Mve	0.180	0.162	0.141	0.093	0.067	0.031	0.034	0.040	0.011	0.059	0.073	0.022	0.032	0.083	0.047	0.051	0.031	0.069	0.000

Table 5. **AMOVA**
Analysis of molecular variance (AMOVA) results, describing estimated variance distributed among regions, populations, individuals, and within individuals. Regions refer to Narragansett Bay and Martha's Vineyard sites. Pops refers to sixteen discrete temporal samples. [a]degrees of freedom [b]sum of squares [c]mean squared deviations

Source	df[a]	SS[b]	MS[c]	Est. Var.	%
Among Regions	1	27.86	27.86	0.03	1%
Among Pops	17	218.80	12.87	0.16	6%
Among Indiv	570	1872.07	3.28	1.00	40%
Within Indiv	589	759.00	1.29	1.29	52%
Total	1177	2877.72		2.47	100%

Table 6. **BIOENV Results**
Shown are the best BIOENV results for correlation of environmental differences among samples with pairwise F_{ST} values. Variables include 1. chlorophyll *a,* 2. temperature 3. irradiance 4. salinity. Table displays number of variables tested (# Vars), their correlations, and the variables associated with each test. For the global test, Rho = 0.591, and p =0.01.

# Variables	Correlation	Variables
2	0.591	1,3
1	0.587	1
2	0.585	1,4
3	0.583	1,3,4
3	0.455	1,2,4
4	0.454	all
2	0.459	1,2
3	0.449	1-3

Table 7. **BIOENV Results Including Nutrient Concentrations**
Shown are the best BIOENV results for correlation of environmental differences among ten samples from Narragansett Bay or which all metadata was available, and their pairwise F_{ST} values. Variables include 1. ammonia 2. phosphate 3. nitrate + nitrite 4. nitrate 5. nitrite 6. chlorophyll *a* 7. temperature 8. irradiance 9. salinity. Table displays number of variables tested (# Vars), their correlations, and the variables associated with each test. For the global test, Rho = 0.68, and p =0.04.

# Variables	Correlation	Variables
4	0.687	1,2,6,9
2	0.686	1,6
3	0.684	1,2,6,8
4	0.682	1,6,8
3	0.682	1,6,9
5	0.681	1,2,6,8,9
3	-0.681	2,6,9
1	-0.669	6

FIGURE LEGENDS

Figure 1. **Station Map**

Map showing sampling sites in Narragansett Bay Station 2 (red) and Martha's Vineyard Coastal Observatory (yellow) field sites. Whole seawater was consistently sampled from these two locations throughout the sampling period.

Figure 2. **Average Growth Rate**

Bars represent average of triplicate maximum acclimated growth rates of 12 isolates from Narragansett Bay isolated at 22.3°C, and incubated at 5 temperatures (4, 10, 17.5, and 22.3, and 30°C. Asterisk designates significant difference in growth rate from all other temperatures ($p<0.05$), found in isolates grown at 4°C and 22.3°C. Isolates grown at 10°C and 17.5°C were not significantly different. NG signifies no growth (at 30°C)

Figure 3. **Isolate Physiology**

Triplicate maximum acclimated growth rates of 12 isolates from Narragansett Bay isolated at 22.3°C. Each line represents the growth rate of a different isolate across the temperature gradient.

Figure 4. **Expected vs. Observed Heterozygosity**

Comparison of expected (black) vs. observed heterozygosity (grey) for each sample, and SD (error bars) across all 6 loci.

Figure 5. **Principal Coordinates Analysis**

Principal coordinates analysis (PCoA) based on pairwise F_{ST} values of Narragansett Bay and MV samples. Lines grouping sites represent different levels of pairwise F_{ST} distance. Small dotted lines represent F_{ST} values of 0.022. Dashed lines represent F_{ST} values of 0.055. Thin solid line represents F_{ST} values of 0.066. Thick solid line represents F_{ST} values of 0.088. Color bar represents a time line, moving forward from 1/6/09 to the last sampling date of 1/11/11. Sites are colored based on their sampling time.

Figure 6. **STRUCTURE Results**

STRUCTURE results showing Bayesian clustering of individuals from all samples into K=3 groups. Number of K groups was chosen based on the graphical analysis of STRUCTURE results, following the Evanno *et al.*method [43]. Individuals are organized by sampling site. Each individual is represented by a single bar. Each K group is represented by a different color. For each individual, color bar represents membership probability of that individual into any number of K groups. Admixture is apparent when membership coefficient for individuals is < 1 into two or more groups.

Figure 7. **Isolation-by-Environment**

Isolation-by-environment graph shows regression of pair-wise genetic (F_{ST}) with pair-wise environmental distance (measured as Euclidean distance of environmental factors among sites. This graph shows the regression of F_{ST} with pair-wise chlorophyll *a*

differences among sites; chlorophyll *a* was the variable found by BIOENV to be the most significantly correlated with pair-wise F_{ST} (Rho = 0.062, p=0.01).

Figure 8. **Species Richness vs. Heterozygosity**

Regression of species richness (community-level) for each sample from Narragansett Bay (n=14), with observed heterozygosity of *T. rotula* in each sample. Correlation between heterozygosity and species richness was significant with a R^2 of 0.36 (p = 0.04).

Figure 1.

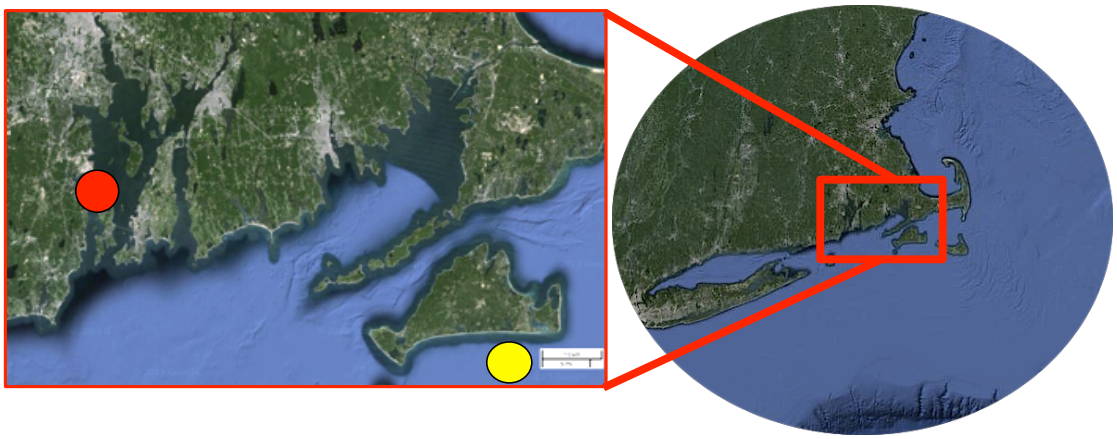


Figure 2.

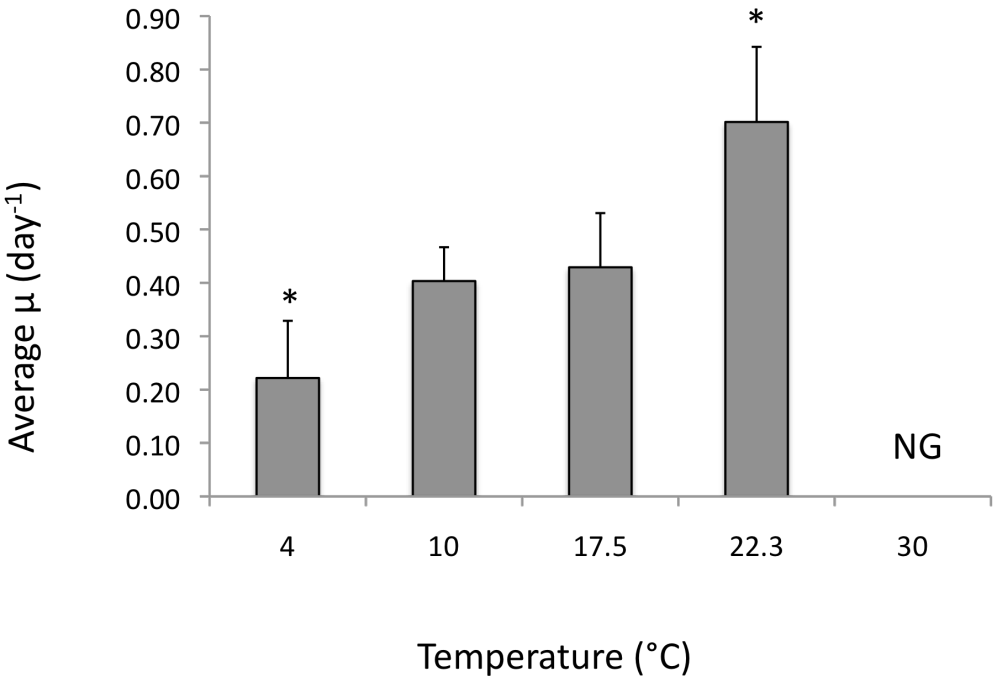

0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Average μ (day-1)
*
*
NG
4
10
17.5
22.3
30
Temperature (°C)

Figure 3.

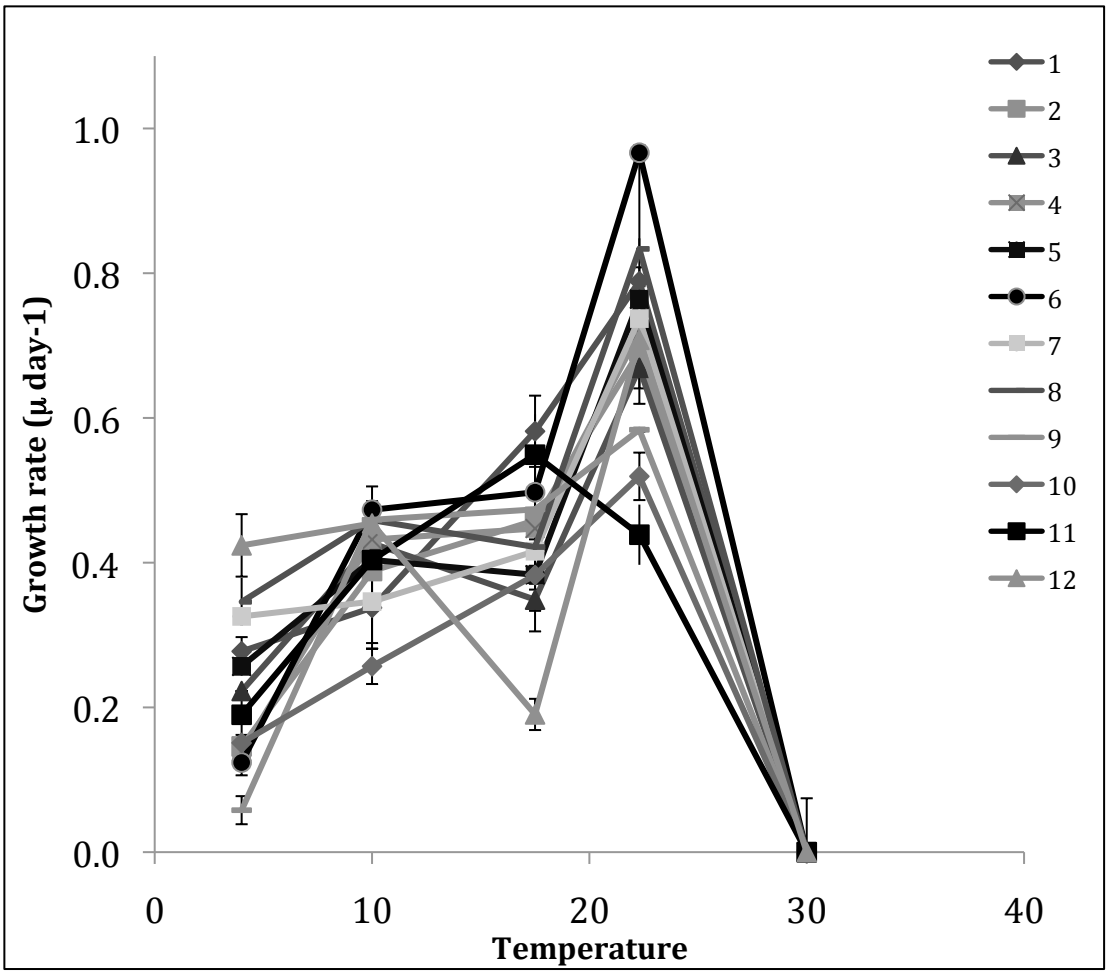

Growth rate (μ day-1)
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Temperature
1
2
3
4
5
6
7
8
9
10
11
12

Figure 4.

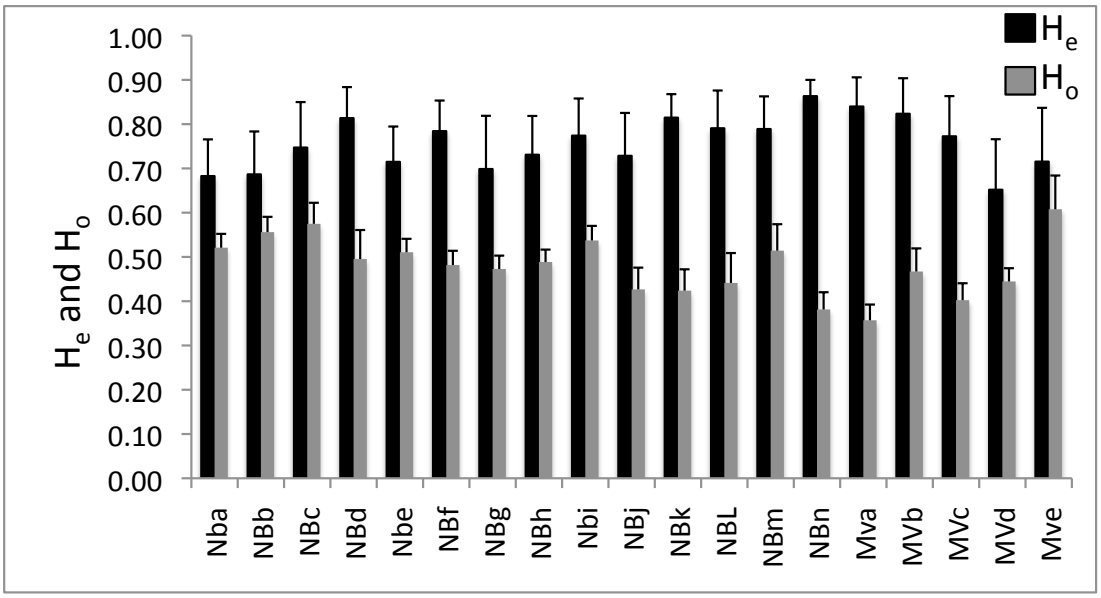

H_e
H_o
H_e and H_o
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Nba
NBb
NBc
NBd
Nbe
NBf
NBg
NBh
Nbi
NBj
NBk
NBL
NBm
NBn
Mva
MVb
MVc
MVd
Mve

Figure 5.

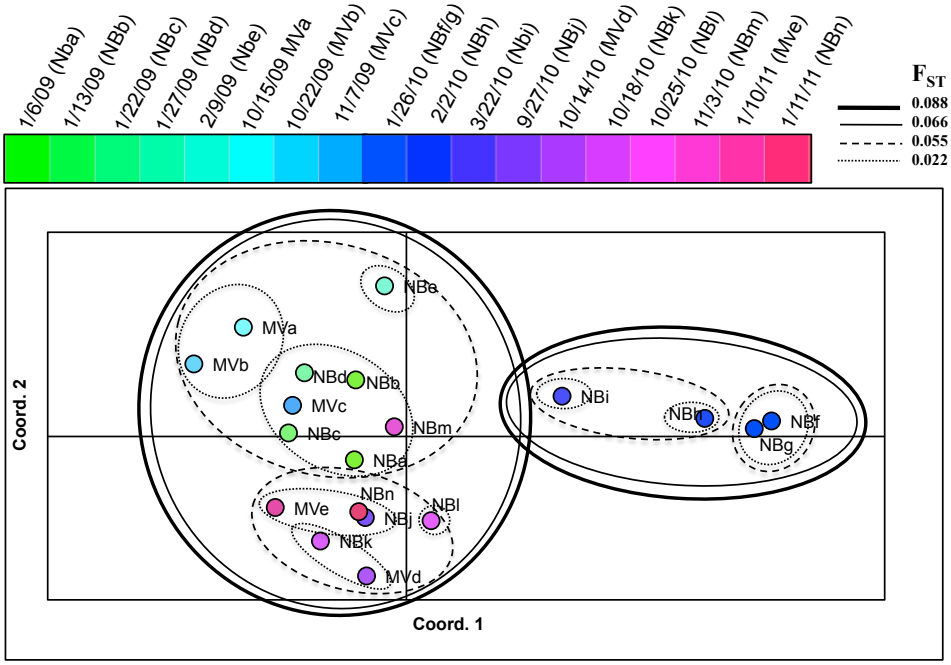

1/6/09 (Nba)
1/13/09 (NBb)
1/22/09 (NBc)
1/27/09 (NBd)
2/9/09 (Nbe)
10/15/09 MVa
10/22/09 (MVb)
11/7/09 (MVc)
1/26/10 (NBf/g)
2/2/10 (NBh)
3/22/10 (Nbi)
9/27/10 (NBj)
10/14/10 (MVd)
10/18/10 (NBk)
10/25/10 (NBl)
11/3/10 (NBm)
1/10/11 (Mve)
1/11/11 (NBn)
F_ST
0.088
0.066
0.055
0.022
NBa
MVa
MVb
NBd
NBb
MVc
NBc
NBm
NBe
NBn
MVe
NBj
NBl
NBk
MVd
NBi
NBh
NBf
NBg
Coord. 2
Coord. 1

Figure 6.

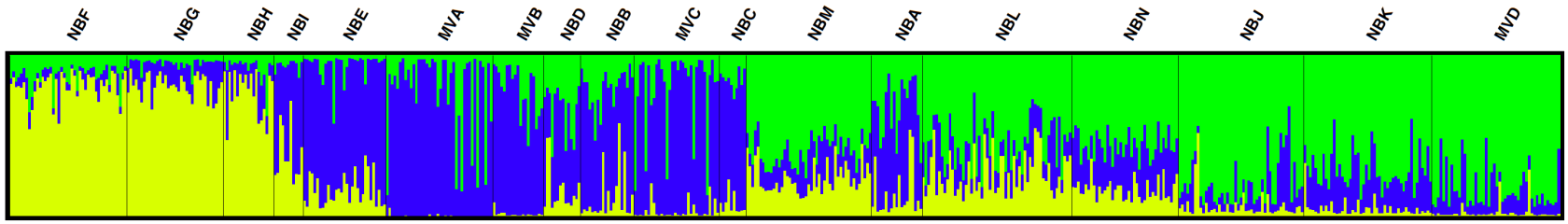

NBF
NBG
NBH
NBI
NBE
MVA
MVB
NBD
NBB
MVC
NBC
NBM
NBA
NBL
NBN
NBJ
NBK
MVD

Figure 7.

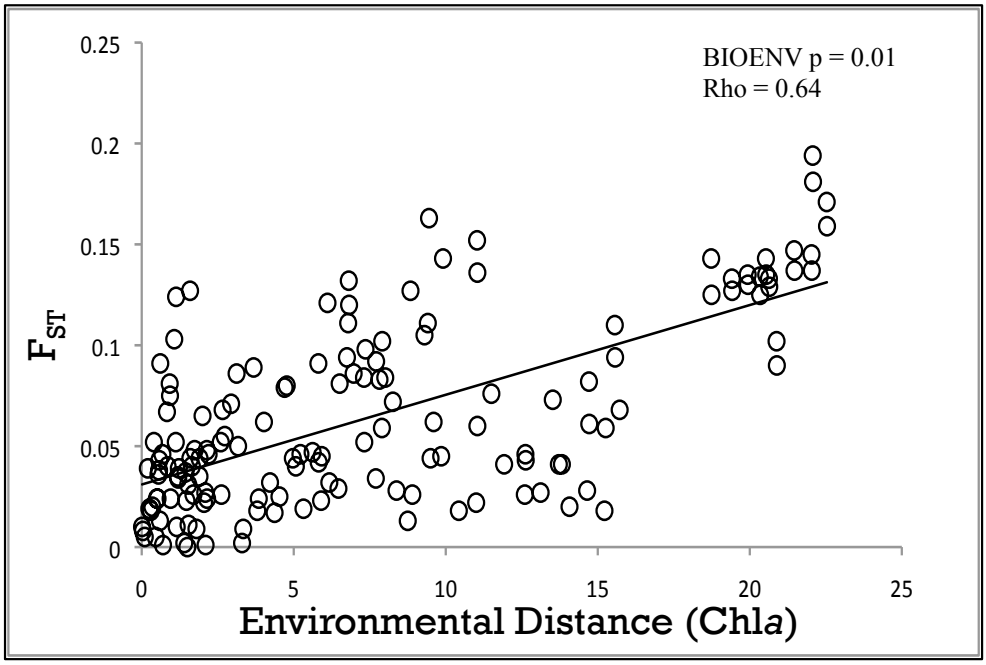
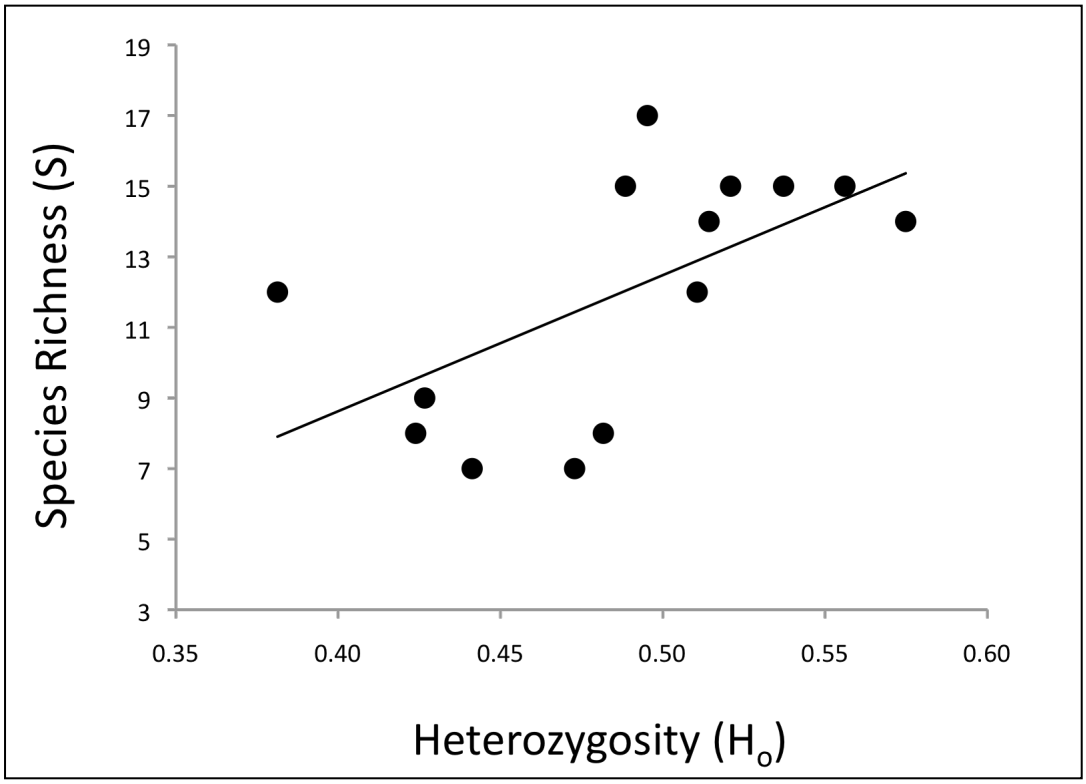

BIOENV p = 0.01
Rho = 0.64
0.25
0.2
0.15
0.1
0.05
0
F_{ST}
0
5
10
15
20
25
Environmental Distance (Chl*a*)

Figure 8.


Species Richness (S)
19
17
15
13
11
9
7
5
3
0.35
0.40
0.45
0.50
0.55
0.60
Heterozygosity (H_o)

CHAPTER 5

DISCUSSION

By

Kerry A. Whittaker

Graduate School of Oceanography, South Ferry Road, University of Rhode Island, Narragansett, RI, 02882

DISCUSSION

Diatoms are of biogeochemical importance to the earth's climate, playing an important role in the global cycling of carbon (Nelson et al. 1995). Found in virtually every marine environment, their ecological success cannot be denied. The vast extent of diversity both within and between diatom species most likely contributes to their widespread ecological success and productivity in the marine environment; for instance, this diversity has been shown to be functionally significant (Sarthou et al. 2005; Whittaker et al. 2012). We know that phytoplankton can evolve rapidly, and quickly fill new ecological niches (Falkowski 2004; Fussmann et al. 2003; Litchman and Klausmeier 2008; Lohbeck et al. 2012). In addition, as single-celled planktonic organisms, diatoms maintain the potential to disperse throughout the globe. Yet, despite the global presence and ecological importance of diatoms, virtually nothing is known about the ways in which their intraspecific diversity is distributed and structured on global spatial scales. Even less understood are the ecological factors that influence the evolution of unique populations within species, and support their genetic connectivity over space and time. This dissertation sought to address these questions by a) exploring the extent and structure of diversity nested within the diatom species *T. rotula* over global space and time b) identifying factors of the marine environment that have contributed to the evolution of this diversity.

Understanding the extent and distribution of diatom diversity is particularly important in the context of climate change. For one, diatoms are highly productive

photosynthetic organisms that play an essential role in global cycling of carbon (Nelson et al. 1995; Sarthou et al. 2005). Evidence suggests that a rapidly changing climate is shifting primary productivity throughout the global ocean; these changes in ocean productivity are most likely due to shifts in the distribution of phytoplankton populations over space and time (Beaugrand et al. ; Cermeno et al. 2010 ; Hallegraeff, 2010). The resiliency of phytoplankton populations to climate change ultimately depends on the extent and distribution of genetic diversity, which dictates both their metabolic and evolutionary potential to respond to environmental stress (Beaugrand et al., 2010).

The results presented here suggest that *T. rotula* harbors high levels of clonal diversity, as has been observed in other species (Godhe and Härnström2010; Rynearson 2005). More importantly, these high levels of clonal diversity relate to high levels of physiological variation observed among individuals collected from around the globe (Whittaker et al. 2012). In samples from Narragansett Bay and off the coast of Martha's Vineyard collected over three years, *T. rotula* populations were highly diverse, and shifted over short time scales. This suggests that the water column or sediment harbors a repository of populations that may respond differently and rapidly to changes in the environment or ecology. *T. rotula* exhibits high levels of genetic and physiological diversity within populations that shift rapidly over time; if these characteristics are common to other diatom species, it suggests that they may be well-suited for adaptation to rapidly changing environmental conditions. On the other hand, in a changing climate, reduction in environmental variability due to

loss of seasonal extremes may the reduce strength of those evolutionary forces supporting diversity within communities and species. More likely, characteristics of rapid growth rates, accompanied by high physiological variability, contribute heavily to the rapid response and succession of *T. rotula* populations; these characteristics suggest that *T. rotula*, and other diatom species, harbor high evolutionary potential that has contributed to their ecological success in the past, and may aid in their ability to adapt to future changes in the earth's climate.

The extent to which diatom diversity varies over time and space may play an important role in determining their bloom formation. Diatom blooms are some of the most prominent features of biogeochemical variability in open-ocean and coastal marine environments (Sverdrup 1953; Townsend et al. 1994). Diatom blooms are defined as a rapid increase in cell biomass, which contribute a large drawdown of carbon from the atmosphere, and rapid flux of fixed carbon that can sink to depths when the bloom declines (Sverdrup 1953). Despite their massive impact on marine ecosystems, the timing of blooms is often sporadic and unpredictable, and triggers of blooms little understood; but in general, diatom blooms occur when cell growth rates outpace loss rates (Lucas et al. 1999a; Lucas et al. 1999b; Smayda 2000). Differences in growth rate among individual strains been shown in the lab to have genetic underpinnings (Rynearson et al. 2009; Whittaker et al. 2012). In the field, exploring the temporal and spatial variations in diatom population-genetic structure provides a link to understanding the interactions between selection pressures of the ocean system and the physiological underpinnings of their diversity and bloom formation.

The data presented here suggest a strong link between the spatial and temporal variations in diatom population structure and their bloom formation. For example, I observed that population-genetic structure within *T. rotula* was significantly correlated with whole- community chlorophyll *a* content on local and global scales, and over time. This means that individuals collected under conditions of low chlorophyll (or non-bloom conditions) were more closely related than with individuals collected at high chlorophyll concentrations (or bloom conditions), and vice versa. In Narragansett Bay, *T. rotula* populations shifted rapidly over time, suggesting that they can respond quickly to changes in the local environment, whether that relates to interactions with other species, a release of nutrient stress, or changes in the environment. This work demonstrates that population succession may be tightly coupled with ecological interactions in the field, dictated just much by resource competition and competitive exclusion as the fluctuation of abiotic factors in the environment. In particular, these data suggest that certain populations may be particularly adapted to bloom and non-bloom periods; this implies that the dynamics of adaptation and diversity below the species level may play an essential role in determining the timing and triggers of phytoplankton blooms, and may explain why blooms are so difficult to predict.

This work advances a decades-long debate on the relative importance of dispersal and the environment in structuring single-celled species at constant flux in the ocean environment. In terms of dispersal, for *T. rotula*, geographic distance plays only a minor role in limiting gene flow and structuring populations. This is in contrast to a strong isolation by distance pattern observed in global samples of the

diatom *Pseudo-nitzschia pungens* (Casteleyn 2010). The two species may differ in their patterns of diversity for several reasons. For one, pennate and centric diatoms differ in their modes of sexual reproduction (Von Dassow and Montresor 2011). These differences may contribute to their genetic structure, as discussed in chapter 2. More probable may be differences in dispersal capacity between the species, namely the ability for *T. rotula* to form resting spores, but not *P. pungens* (Garrison 1981). Resting spores are dormant cysts formed by some diatom species that remain viable for up to 100 years, and often settle in the sediment (Garrison 1981; Harnstrom et al. 2011; Mcquoid and Hobson 1995). Because of this, they may provide a seed bank of diversity that can act as a stepping-stone of dispersal for these organisms (Godhe et al. 2013). The ability to form resting spores may increase the potential for *T. rotula* to disperse throughout the globe, across less favorable environmental conditions. Because *P.pungens* cannot form resting spores, this may explain why dispersal is a greater barrier to gene flow among populations. More broadly, differences in genetic structure observed between these two species stresses the need for future research on population structure in more diatom species to uncover common patterns and shared selection pressures.

This is the first study to examine global diatom population structure alongside variations in the marine environment; thus, it is unclear if patterns observed in *T. rotula* are common to other species. For *T. rotula,* these data demonstrate that the environment, rather than dispersal, isolates populations and dictates geneflow among them. The structuring of *T. rotula* populations across time and global geographic space demonstrates that this species is not genetically homogenous, nor fully admixed.

In *T. rotula* genetic composition can vary as much over time as it does over large spatial scales; this poses a caution to pooling samples over seasons or years, or grouping samples over space when exploring population-level diversity in diatoms. In addition, these results suggest that more work is needed to understand the scale of variability in diatom populations over time. If geographic space is not a good predictor of *T. rotula* population structure, then care must be taken to better understand variations in population structure over time, and across diverse ecological conditions. This work broadens our understanding of these interactions, by attempting to tease apart variations diatom diversity over large spatial scales, time, and across a wide range of environmental conditions. Overall, this work demonstrates that distance may not be a great barrier to geneflow in these organisms, and instead, environmental selection may support high levels of diversity within diatom species, and dictate the extent of geneflow among populations.

Beaugrand, G., M. Edwards, and L. Legendre, 2010. Marine biodiversity, ecosystem functioning, and carbon cycles, p. 10120-10124.
Casteleyn, G. 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. PNAS : Proceedings of the National Academy of Sciences **107:** 12952-12957.
Cermeno, P., C. De Vargas, F. Abrantes, and P. G. Falkowski, 2010. Phytoplankton biogeography and community stability in the ocean. PLoS One **5:** e10037.
Falkowski, P. G. 2004. The Evolution of Modern Eukaryotic Phytoplankton. Science (New York, N.Y.) **305:** 354-360.
Fussmann, G. F., S. P. Ellner, and N. G. Hairston. 2003. Evolution as a critical component of plankton dynamics. Proceedings of the Royal Society of London. Series B: Biological Sciences **270:** 1015-1022.
Garrison, D. L. 1981. Monterey Bay Phytoplankton. II. Resting Spore Cycles in Coastal Diatom Populations. Journal of Plankton Research **3:** 137-156.
Godhe, A., J. Egardt, D. Kleinhans, L. Sundqvist, R. Hordoir, and P. R. Jonsson. 2013. Seascape analysis reveals regional gene flow patterns among populations of a marine planktonic diatom. Proceedings of the Royal Society B: Biological Sciences **280:** 20131599.
Godhe, A., and K. Härnström. 2010. Linking the planktonic and benthic habitat: genetic structure of the marine diatom Skeletonema marinoi. Molecular ecology **19:** 4478-4490.
Hallegraeff, G. M, 2010. Ocean climate change, phytoplanktonc community responses, and harmful algal blooms: A formidable predictive challenge. Journal of Phycology **46:** 220-235.
Härnström, K., M. Ellegaard, T. R. J. Andersen, and A. Godhe. 2011. Hundred years of genetic structure in a sediment revived diatom population. Proceedings of the National Academy of Sciences **108:** 4252-4257.
Litchman, E., and C. A. Klausmeier. 2008. Trait-based community ecology of phytoplankton. Annual Review of Ecology, Evolution, and Systematics **39:** 615-639.
Lohbeck, K. T., U. Riebesell, and T. B. Reusch. 2012. Adaptive evolution of a key phytoplankton species to ocean acidification. Nature Geoscience **5:** 346-351.
Lucas, L. V., J. R. Koseff, J. E. Cloern, S. G. Monismith, and J. K. Thompson. 1999a. Processes governing phytoplankton blooms in estuaries. I: The local production-loss balance. Marine ecology. Progress series **187:** 1-15.
Lucas, L. V., J. R. Koseff, S. G. Monismith, J. E. Cloern, and J. K. Thompson. 1999b. Processes governing phytoplankton blooms in estuaries. II: The role of horizontal transport. Marine Ecology-Progress Series **187:** 17-30.
Mcquoid, M. R., and L. A. Hobson. 1995. Importance of resting stages in diatom seasonal succession Journal of Phycology **31:** 44-50.
Nelson, D. M., P. Treguer, M. A. Brzezinski, A. Leynaert, and B. Queguiner. 1995. Production and dissolution of biogenic silica in teh ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Global Biogeochemical $_{\text{Cycle}}$ **9:** 359-372.
Rynearson, T. A. 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom Ditylum brightwellii. Molecular ecology **14:** 1631-1640.

Rynearson, T. A., E. O. Lin, and E. V. Armbrust. 2009. Metapopulation Structure in the Planktonic Diatom Ditylum brightwellii (Bacillariophyceae). Protist **160:** 111-121.

Sarthou, G. R., K. R. Timmermans, S. P. Blain, and P. Treguer. 2005. Growth physiology and fate of diatoms in the ocean: a review. Journal of Sea Research **53:** 25-42.

Smayda, T. 2000. Ecological features of harmful algal blooms in coastal upwelling ecosystems. South African Journal of Marine Science **22:** 219-253.

Sverdrup, H. U. 1953. On Conditions for the Vernal Blooming of Phytoplankton. Journal du Conseil **18:** 287-295.

Townsend, D. W., L. M. Cammen, P. M. Holligan, D. E. Campbell, and N. R. Pettigrew. 1994. Causes and consequences of variability in the timing of spring phytoplankton blooms. Deep Sea Research Part I: Oceanographic Research Papers **41:** 747-765.

Von Dassow, P., and M. Montresor. 2011. Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. Journal of plankton research **33:** 3-12.

Wahlund, S. 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas **11:** 65-106.

Whittaker, K., D. Rignanese, R. Olson, and T. Rynearson. 2012. Molecular subdivision of the marine diatom Thalassiosira rotula in relation to geographic distribution, genome size, and physiology. BMC Evolutionary Biology **12:** 209.

APPENDIX

A. Growth rates of *T. rotula* strains

The table below contains raw growth rate data for strains referenced in Chapters 2 and 4. Table contains culture/strain name, average growth rate, average growth rate standard error, and temperature of incubaion. A growth rate of "0" refers to cultures that were incubated at a certain temperature, but failed to grow.

Strain/Culture Name	Average growth rate (mu*day-1)	(mu*day-1) SE	Temp.
4C Nb12a2	0.277447	0.019715	4
10C Nb12a2	0.337721	0.048799	10
17.5C Nb12a2	0.581584	0.049222	17.5
22.3C Nb12A2	0.788809	0.169322	22.3
30C nb12a2	0.000000	0.000000	30
4C Nba	0.146088	0.013383	4
10C Nba	0.388438	0.031565	10
17.5C Nba	0.458412	0.011415	17.5
22.3C Nba	0.69201	0.008761424	22.3
30C Nba	0	0	30
4C Nbb	0.222625801	0.021239502	4
10C Nbb	0.426800707	0.043867487	10
17.5C Nbb	0.348901299	0.028425068	17.5
22.3C Nbb	0.669246521	0.074429587	22.3
30C Nbb	0	0	30
4C NBc	0.139005334	0.010171578	4
10C NBc	0.431945434	0.031575926	10
17.5C NBc	0.447225027	0.037015346	17.5
22.3C NBc	0.717538452	0.044459162	22.3
30C NBc	0	0	30
4C Nbd	0.256710438	0.009467267	4
10C Nbd	0.403796531	0.026227082	10
17.5C Nbd	0.383270891	0.012450994	17.5
22.3C Nbd	0.764077106	0.04410796	22.3
30C Nbd	0	0	30
4C Nbe	0.123835579	0.017669177	4

10C Nbe	0.473105424	0.032255292	10
17.5C Nbe	0.497354353	0.043859613	17.5
22.3C Nbe	0.966439934	0.010302169	22.3
30C Nbe	0	0	30
4C Nbf	0.325739178	0.00320383	4
10C Nbf	0.346171771	0.065398596	10
17.5C Nbf	0.415858836	0.053080554	17.5
22.3C Nbf	0.737272939	0.044681703	22.3
30C Nbf	0	0	30
4C Nbg	0.345886039	0.033071875	4
10C Nbg	0.458761376	0.014711676	10
17.5C Nbg	0.421606373	0.022195664	17.5
22.3C Nbg	0.833600185	0.014582375	22.3
30C Nbg	0	0	30
4C Nbh	0.057979034	0.019399648	4
10C Nbh	0.459208912	0.020846898	10
17.5C Nbh	0.473644673	0.021153651	17.5
22.3C Nbh	0.583662343	0.002728415	22.3
30C Nbh	0	0	30
4C Nbi	0.150956273	0.011055567	4
10C Nbi	0.256864136	0.024549864	10
17.5C Nbi	0.382824097	0.049386286	17.5
22.3C Nbi	0.519282506	0.032800147	22.3
30C Nbi	0	0	30
4C Nbj	0.189828262	0.027683743	4
10C Nbj	0.403595337	0.027959872	10
17.5C Nbj	0.549215328	0.055819952	17.5
22.3C Nbj	0.438639734	0.041585811	22.3
30C Nbj	0	0	30
4C Nbl	0.423883679	0.043156064	4
10C Nbl	0.454158725	0.030875144	10
17.5C Nbl	0.190296196	0.021566123	17.5
22.3C Nbl	0.707904659	0.023335654	22.3
30C Nbl	0	0	30
4° JPNTR18	0.359921258	0.019538984	4

10° JPNTR18	0.56064458	0.053989511	10
17.5° JPNTR18	0.68487896	0.085385097	17.5
25°JPNTR18	0.540674833	0.007730284	25
30°JPNTR18	0		30
35°JPNTR18	0		35
4° GSO101	0.311433287	0.01270744	4
10° GSO101	0.56004463	0.061001022	10
17.5° GSO101	0.719025431	0.06064188	17.5
25°GSO101	0.587994809	0.004420992	25
30°GSO101	0	0	30
35°GSO101	0	0	35
4° CCMP1647	0	0	4
10° CCMP1647	0.323901288	0.008247406	10
17.5° CCMP1647	0.707816223	0.05610438	17.5
25°CCMP1647	0.62379	0.022427278	25
30°CCMP1647	0	0	30
35°CCMP1647	0	0	35
4° P17F4	0.256044574	0.019504296	4
10° P17F4	0.592622328	0.006626868	10
17.5° P17F4	0.92013566	0.039054618	17.5
4° CCMP3264	0.200500248	0.019728648	4
10° CCMP3264	0.599255024	0.017810276	10
17.5° CCMP3264	0.779039995	0.030000725	17.5
25°CCMP3264	0.648043096	0.02273065	25
30°CCMP3264	0	0	30
35°CCMP3264	0	0	35
4° CCAP1085_21	0.262153425	0.039370683	4
10° CCAP1085_21	0.550623393	0.005974966	10
17.5° CCAP1085_21	0.778676102	0.030269254	17.5
25°CCAP1085_21	0.607127493	0.015811256	25
30°CCAP1085_21	0	0	30
35°CCAP1085_21	0	0	35

B. Microsatellite Sequences

Primer	Seq ID	uSAT	Primer_left	Primer_right
TR1	TR1 (from bead hyb)	$(GTT)^{3}GCAAGC(TGG)^{9}$	ggcaaagcccacttccaattt	atcccctccgcctcttcc
TR3	TR3 (from bead hyb)	$(GA)16$	ccagagggtatggctgtaag	atgctgccggtagtgagttg
TR7	GIWPZQA01A3T0T	$(AC)^{13}$	agcgtaggtgtcccagaac	atgccatgtacaccgttgc
TR8	GIWPZQA01A3VGT	$(AG)^{12}$	tgacctaggcagccagaatg	atggcgatgtgattggtgc
TR10	GIWPZQA01A56PH	$(ACT)^{11}$	catcgagtgcccgttgttg	gttgattgggtccaagggg
TR27	GIWPZQAO1A0IO4	$(GTT)^{10}$	cgttccttcttcggagggg	cagctccctcaccaccag

C. Environmental Data, Narragansett Bay/ MVCO

Sample	Date	Chl*a*	Salinity	Temp	Cell Abundance	NH4	DIP	NO3+2	NO3	NO2	Irradiance (kWh/m2/day)
NBa	1/8/09	2.20	28.46	2.7	6000	0.959	0.38	3.731	0.163	3.567	1.7
NBb	1/13/09	1.49	28.55	2	3000	3.829	0.756	7.965	0.332	7.633	1.7
NBc	1/22/09	1.51	27.44	0.6	4000	3.878	0.732	12.019	0.443	11.575	1.7
NBd	1/27/09	2.18	30.43	2.1	8000	*NA*	*NA*	*NA*	*NA*	*NA*	1.7
NBe	2/9/09	2.63	30.44	2.3	3000	2.24	0.05	0.56	0.21	0.34	2.46
NBf	1/26/10	22.64	30.33	3.81	1991	0.16	0.04	0.22	0.03	0.185	1.7
NBg	1/26/10	22.64	30.33	3.81	2017	0.16	0.04	0.22	0.03	0.185	1.7
NBh	2/2/10	10.01	30.13	1	6000	0.16	0.04	0.02	0.01	0.00	2.46
NBi	3/22/10	7.92	26.06	6.51	55699	1.76	0.04	0.31	0.16	0.14	3.53
NBj	9/27/10	3.22	28.65	19.5	12000	*NA*	*NA*	*NA*	*NA*	*NA*	4.21
NBk	10/25/10	11.60	28.11	12.9	27000	*NA*	*NA*	*NA*	*NA*	*NA*	2.97
NBl	11/3/10	7.07	28.59	12	9000	1.27	0.33	0.68	0.28	0.40	1.8
NBm	1/11/11	15.82	28.66	1.8	15000	7.81	1.04	8.26	0.81	7.44	1.7
NBn	10/17/10	1.75	28.37	14.8	1000	*NA*	*NA*	*NA*	*NA*	*NA*	2.97
MVa	10/15/09	0.1	31.59	14.14	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	2.97
MVb	10/22/09	0.56	31.32	12.95	NA	*NA*	*NA*	*NA*	*NA*	*NA*	2.97
MVc	11/7/09	3.98	31.43	11.57	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	1.8
MVe	1/10/11	1.17	32.02	2.43	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	1.7

D. Environment Data, Global Samples

Site	Date Collected	Chl*a*	Salinity	SST	Cells/L
Nba	1/26/10	22.64	30.33	3.81	1991
NBb	1/26/10	22.64	30.33	3.81	2017
NBc	2/2/10	10.01	30.13	1.00	6000
Waa	2/16/10	2.70	17.56	8.20	6510
Fra	3/9/10	0.83	34.93	8.30	1330
Ger	3/18/10	0.30	*NA*	2.60	*NA*
NBd	3/22/10	7.92	26.06	6.51	55699
Cali	3/22/10	0.33	33.62	15.00	1667
FRb	3/24/10	0.36	34.98	9.00	*NA*
Wab	3/29/10	3.80	14.68	8.26	17520
Puget	4/14/10	4.50	29.46	9.09	8125
NZ	10/15/10	*NA*	*NA*	13.90	*NA*
SA	11/19/10	0.23	35.60	22.20	1042