Genomic evidence of speciation and adaptation in diatoms

Julie A. Koester

A dissertation
submitted in partial fulfillment of the
requirements of

Doctor of Philosophy

University of Washington
2012

E. Virginia Armbrust, Chair
Willie Swanson
Robert Morris

School of Oceanography

University of Washington

Abstract

Diatoms are one of the most ecologically important groups of organisms in the ocean. They are the youngest and most species-rich group of phytoplankton, having colonized both marine and terrestrial ecosystems. The ocean is constantly changing, and understanding the mechanisms of species-diversification and adaptation in diatoms is important to assessing their resilience to future environmental changes. Speciation and adaptation were investigated in three diatom genera using genomic and genetic signals. One mechanism of speciation was tested by bringing isolates of two populations of the planktonic diatom *Ditylum brightwellii* into the lab to measure genome size differences indicated by cell size differences in the field. Genome sizes differed by two-fold between individuals of each population, suggesting that the populations are in fact cryptic species, thus corroborating previous research indicating reproductive isolation between the populations. Growth rates of *Ditylum* isolates from within a species differed significantly depending on where they were collected, southern or northern Pacific Ocean, suggesting that they were differentially adapted to their local environments. Natural selection acts directly on phenotypes; positively selected genes control those phenotypes and their sequences vary among populations and between species. Positively selected genes were investigated in *Pseudo-nitzschia*, *Ditylum* and *Thalassiosira*, but the greatest number of selected genes was found within a single species, *Thalassiosira pseudonana.* All of the protein coding genes from seven strains of *T. pseudonana* were analyzed and 809 (7%) were found to be positively selected. These genes encode protein-binding proteins, transcriptional regulators, and proteins associated with cell signaling and the cell wall. One quarter of the positively selected genes was novel to *T. pseudonana*, thus differentiating this species while conferring a selective advantage to individuals. Genome duplications, such as occurred in *D. brightwellii*, provide an opportunity for increased genetic variability upon which selection may act under changing environments. In the absence of polyploidization, genetic variability is maintained through mutations accrued during each round of cell division. The positively selected genes presented here for *T. pseudonana* provide future opportunities to test hypotheses concentrated on linking genotypes of positively selected genes with the phenotypes that they control and their associated selective pressures.

Genomic evidence of speciation and adaptation in diatoms

Julie A. Koester

Chair of supervisory committee:
Professor E. Virginia Armbrust
School of Oceanography

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my committee for many insightful discussions about this research. Special thanks to Ginger Armbrust, my supervisor, for the many additional opportunities provided to me in her lab and on the ocean. To Bob Morris, Gabrielle Rocap, LuAnne Thompson, thank you.

This research was conducted in the lab and on a computer.
For laboratory assistance and trouble-shooting I would like to thank: Irina Oleinikov, Rachelle Lambert, Ellen Lin, Rhonda Morales, Claire Ellis and Chris Frasier
For computational support, *many, many* thanks go to: Chris Berthiaume and Dave Schruth

To my co-authors Jarred Swalwell (math advisor) and Peter von Dassow (inspirational cheer-leader), I thank you.

The working group is large in the Environmental Center for Genomics (CEG) and the critical conversations many. For each of the CEG members listed here (past and present), and our scientific colleagues, I remember specific conversations impacting this research: Francois Ribalet, Sara Bender, Colleen Durkin, Tom Connolly, Micaela Parker, Vaughn Iverson, Adrian Marchetti, Robin Kodner, Megan Kogut, Larry Ruzzo.

Rita Peterson, our laboratory administrator, deserves my deepest gratitude for her unwavering patience and good mood. Thank you Rita!

Thank you Mark Webber for a shared love of *Ditylum* and those New Zealand clones! Mark's micrographs and conversations are always an inspiration.

Finally, I would like to thank four scientists who have seen me through from the very beginning, before UW. I continue to share support, encouragement and camaraderie in science with Nic Blouin, Susan Brawley, Lee Karp-Boss, and Jessie Muhlin.

INTRODUCTION

Diatoms are one of the most ecologically important groups of organisms in the ocean. Historically in biological oceanography, diatoms were grouped with all of the other photosynthetic organisms in the sea and measured as a single unit of phytoplankton. The measurements included bulk chlorophyll *a* concentrations as a proxy for abundance and carbon-14 uptake for total primary production. Gradually, discoveries that individual groups of phytoplankton, eukaryotic and prokaryotic, provide functionally different contributions to the ocean and atmosphere have increased the breadth of research within biological oceanography (Chisholm *et al.*, 1988, Nelson *et al.*, 1995). Today the field is rich with research focusing on community processes and the dynamics of single species using diverse tools that include genomics and bio-informatics.

Diatoms are single-celled, eukaryotic algae with cell walls made of silica that achieved ecological success rapidly. They likely originated ca. 250 mya in shallow seas during a time of mass extinction, the boundary of the Permian and Triassic periods, (Raup, 1979, Gersonde & Harwood, 1990, Sorhannus, 2007). Diatoms are the youngest members of the phytoplankton community. It is hypothesized that a change in redox state of the ocean at this time selected for groups of eukaryotic phytoplankton, including diatoms, which had different requirements for trace metals than the green algae previously dominating the community (Quigg *et al.*, 2003). Diatoms were minor contributors to ocean processes until the Cretaceous-Paleogene extinctions 65 mya, after which they radiated both geographically and in species number. Today, diatoms are the most species-rich group of phytoplankton; in the oceans there are ca. 10,000 described species, whereas the are ca. 2000 species of dinoflagellates and 500 of coccolithophores (Mann & Droop, 1996, Kooistra *et al.*, 2007). Diatoms are adapted to many marine and terrestrial ecosystems that have water and sunlight, including sea ice and sediments. The lifestyles of diatoms are varied and include benthic and planktonic habitats, and they live as colonies or single cells with or without motility. In the ocean, diatoms control the silicon cycle. Diatoms are largely responsible for the 700-fold drawn-down of silicate concentrations that occurred in the ocean over the last 65 million years (Siever, 1991). In much of today's ocean, silicic acid occurs at growth-limiting concentrations, but concentrations vary regionally and seasonally and are greatest along the coasts. Diatoms are able to take advantage of high nutrient situations, and form dense blooms through rapid asexual reproduction. Their share of primary production is 20 % of that on Earth (Nelson *et al*., 1995, Field *et al*., 1998).

The silicon and carbon cycles in the ocean are coupled through diatoms, which provide an interactive link to climate changes. Silicon is transported to the ocean by terrestrial weathering and hydrothermal activity. Weathering, and thus silicon transport, increases during periods of glaciations, and decreases during warmer times. As the ocean continues its warming trend, the availability of silicon to diatoms is likely to decrease as the water becomes more stratified, and ocean mixing decreases. Stratification is predicted to select for smaller sized diatoms (Litchman *et al*., 2009), and in extreme cases lead to a community shift from one including diatoms to one dominated by cyanobacteria (Karl *et al*., 2001). It is therefore important to understand the mechanisms of speciation and adaptation in the diatoms to begin to link environmental selection pressures to their selected phenotypes and associated genes.

Species are an arbitrary, but necessary, unit of biological classification used to aid scientists in describing the organisms they study. Species do have biological merit in that organisms are grouped by similar traits, whether they are sexual, asexual, eukaryotic or prokaryotic. Originally (since the early 1700s), diatom species were based on their distinct cell wall morphologies. The predominate phase of the diatom life history is asexual, but over time, the importance of sexual reproduction within the life history has been recognized and applied to taxonomy, and the biological species concept is generally accepted for this group (Mann, 1999). Species, as defined by the biological species concept, are groups of organisms that can interbreed; speciation is indicated when reproductive isolation develops between populations with a species (Mayr, 1969). The relationship of diatom species to other protists (Adl *et al.*, 2005), and their relationships to one another phylogenetically (Medlin & Kaczmarska, 2004, Sims *et al*., 2006, Theriot *et al*., 2010, Alverson *et al*., 2011), are well studied, but the details are still debated (Medlin, 2010, Williams & Kociolek, 2010). Diatoms have large census population sizes, and numerous means of dispersal, the most notable of which are wind and water currents (Kristiansen 1996). Widespread distribution of morphotypes is sometimes used as evidence to suggest that the majority of diatom species are cosmopolitan (Finlay *et al*., 2002). Many diatom genera and some species *are* circum-global in distribution with latitudinal boundaries, but most species have restricted distributions (Vanormelingen *et al*., 2008, Vyverman *et al*., 2010). Small differences in morphology previously characterized as phenotypic plasticity within a species are now being correlated to genetic divergence and reproductive isolation. New descriptions of cryptic species include: the multi-species complexes related to *Skeletonema costatum* (Medlin *et al.*, 1991, Sarno *et al.*, 2005, Sarno *et al.*, 2007) and *Cyclotella meneghiniana (*Beszteri *et al*., 2005, Beszteri *et al.*, 2007); *Pseudo-nitzschia mannii* within the *P. pseudodelicatissima* complex (Amato & Montresor, 2008), and *P. arenyensis* within the *P. delicatissima* complex (Quijano-Scheggia *et al.*, 2009). Hydrogeographic and ecological mechanism likely counter-act the potential of continuous dispersal (Patarnello *et al.*, 2007, Cermeno & Falkowski, 2009). However, if mating information is unavailable, Fenchel (2005) cautions that differentiating genetic and phenotypic data should be coupled with distinct geographic distributions or functions when describing a new species.

Species specific differences are noted frequently in physiology experiments on diatoms and occasionally attributed to different adaptive strategies among the species. However, fewer studies systematically test for differences with respect to species distributions and their associated environmental parameters. Estuarine, coastal, and open ocean species have tolerances to varying salinities and fluctuations in light consistent with where they are most frequently collected (Brand, 1984, Lavaud *et al.*, 2007). Salinity effects how much silica is deposited into the diatom cell walls. Freshwater species are more heavily silicified than marine species, which has an interesting evolutionary consequence; diatoms that sink in a lake are more likely to be returned to the photic zone through seasonal mixing than are diatoms that sink in the ocean (Conley *et al.*, 1989). Lightly silicified cell walls are advantageous in the ocean, but the largest diatoms have buoyancy regulation mechanisms that allow them to descend to nitrogen sources in deep water and return to the surface (Moore & Villareal, 1996, Villareal *et al.*, 1999, Boyd & Gradmann, 2002). The next example is unique because it tests the adaptive response of a single protein, ribulose-1,5-bisphophate carboxylase oxygenase (RuBisCO), with respect to temperature. The enzyme activity and substrate affinity of RuBisCO that was extracted from Antarctic and temperate species reflected the temperatures in which the diatoms grew naturally (Descolas-gros & Debilly, 1987). The differential function of RUBISCO between the species links the mechanisms of cold-adaptation in Antarctic diatoms to a selective pressure, phenotype and protein.

Adaptation is achieved when phenotypic traits of individuals are selected for by the environment, have high rates of survival, and are passed along to offspring. Natural selection requires individuals within populations to vary genetically and thus phenotypically. This variation in functional, protein coding genes can be associated with the adaptation of geographically distinct populations to local environmental conditions using genome scans of numerous individuals from each population (Namroud *et al.*, 2008, Bigham *et al.,* Fischer *et al.*, 2011). Population genetic studies in diatoms usually use microsatellites, which are neutral genetic markers, to assess genetic diversity within and between populations. Clonal diversity of these neutral markers is high, 87 %, within an asexually dividing bloom population of the planktonic diatom *Ditylum brightwellii* (Rynearson & Armbrust, 2005), suggesting that there may also be a high level of functional genetic variation. In physiology experiments, adaptive differences are found among diatoms with respect to pollutants. *Skeletonema costatum* isolated from a highly polluted fjord had higher tolerance to zinc than the strain from a less polluted fjord (Jensen *et al.*, 1974). Interestingly, estuarine strains of several species with coastal to offshore distributions tolerated novel and toxic chemicals better than strains isolated from the open ocean; the estuarine strains could not have adapted to the chemical, but the adaptations they acquired living in a fluctuating environment allowed them to survive the chemical, whereas the oceanic strains from a stable environment died (Fisher, 1977).

Although there is a large amount of literature on the taxonomy, ecology and physiology of diatoms, the genetic mechanisms of adaptation and speciation are understudied. In the following chapters, I present data on a genomic mechanism of speciation, polyploidization, in *Ditylum brightwellii* and genetics of natural selection in three genera, *Ditylum*, *Pseudo-nitzschia* and *Thalassiosira*. Polyploidization increases the size of the genome by at least one haploid complement of chromosomes, and can lead to immediate reproductive isolation. Speciation by whole genome duplication is hypothesized for two co-occurring populations of *D. brightwellii* in Puget Sound, WA because they have a two-fold difference in genome size. For the duplicated genome, the emergence of novel gene functions can occur via mutation in duplicated genes, among many other fates after duplication (Ohno, 1970, Zhang, 2003). The rate at which new polyploids can adapt, relative to diploid progenitors, is dependent upon the amount of inherited heterozygosity from the original diploid, and the amount of masking of new beneficial mutations by the original genes (Otto & Whitton, 2000, Otto, 2007).

Differences in growth rates are indicative of genetic variation. Growth rates of *D. brightwellii* differed between the two putative species, but they also differed within a species between strains from New Zealand and Puget Sound, suggesting that those populations were adapted to their local conditions. The findings in *D. brightwellii* led to investigating genes that promote adaptation in diatoms. Positively selected genes have sequences that vary among populations and between species, and they are the blueprints behind the phenotypes upon which selection is acting. Positive selection is identified by analyzing the DNA sequence and its translation to the amino acids of the protein and calculating the ratio of amino acid changing (non-synonymous) substitutions to silent (synonymous) substitutions, d_N:d_S . If d_N:d_S is > 1.0, positive selection is indicated for the gene. Ratios equal to and less than one indicate neutral and purifying selection, respectively. Over 11,000 genes shared by seven strains of *Thalassiosira pseudonana* were tested for positive selection because a preliminary analysis revealed that comparisons within a species revealed the strongest signature of positive selection.

This is the first genome-wide scan for positive selection in any group of phytoplankton. Previously, only three single gene tests for positive selection have been done on diatoms (Sorhannus, 2003, Sorhannus & Pond, 2006, Alverson, 2007), due to constraints in available DNA sequence. Advances in high-throughput sequencing technologies are creating more opportunities to work with multiple genomes simultaneously, such as was done for this study for *T. pseudonana*. *Thalassiosira pseudonana* is a good research model in many ways: it grows well in the laboratory, has a lengthy history in the literature, it has a small genome, and as the first diatom to have its genome sequenced, it is well-studied genetically (Armbrust *et al.*, 2004). *Ditylum brightwellii* may be considered a “non-model” organism, but it too has a rich literature history including physiology and population genetics studies. *Ditylum brightwellii* is easily brought into culture from natural populations. Although *Pseudo-nitzschia* plays a minor role in this research, certain species have great impacts on ocean ecosystems because they produce a neurotoxin that can severely affect humans. The identification of positively selected genes introduces a powerful analysis tool to the biological oceanography community to understand how phytoplankton are adapting to their environment.

Chapter I: Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta)

Published: Koester, J.A., Swalwell, J.E., von Dassow, P., and Armbrust, V.E. 2010. Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta). *BMC Evolutionary Biology*. **10:**1 doi:10.1186/1471-2148-10-1

Abstract

Diatoms are one of the most species-rich groups of eukaryotic microbes known. Diatoms are also the only group of eukaryotic micro-algae with a diplontic life history, suggesting that the ancestral diatom switched to a life history dominated by a duplicated genome. A key mechanism of speciation among diatoms could be a propensity for additional stable genome duplications. Across eukaryotic taxa, genome size is directly correlated to cell size and inversely correlated to physiological rates. Differences in relative genome size, cell size, and acclimated growth rates were analyzed in isolates of the diatom *Ditylum brightwellii*. *Ditylum brightwellii* consists of two main populations with identical 18s rDNA sequences; one population is distributed globally at temperate latitudes and the second appears to be localized to the Pacific Northwest coast of the USA. These two populations co-occur within the Puget Sound estuary of WA, USA, although their peak abundances differ depending on local conditions. All isolates from the more regionally-localized population (population 2) possessed 1.94 ± 0.74 times the amount of DNA, grew more slowly, and were generally larger than isolates from the more globally distributed population (population 1). The ITS1 sequences, cell sizes, and genome sizes of isolates from New Zealand were the same as population 1 isolates from Puget Sound, but their growth rates were within the range of the slower-growing population 2 isolates. Importantly, the observed genome size difference between isolates from the two populations was stable regardless of time in culture or the changes in cell size that accompany the diatom life history. The observed two-fold difference in genome size between the *D. brightwellii* populations suggests that whole genome duplication occurred within cells of population 1 ultimately giving rise to population 2 cells. The apparent regional localization of population 2 is consistent with a recent divergence between the populations, which are likely cryptic species. Genome size variation is known to occur in other diatom genera; we hypothesize that genome duplication may be an active and important mechanism of genetic and physiological diversification and speciation in diatoms.

Introduction

Genotypic and physiological variation is frequently disguised by an apparent morphological constancy traditionally assumed to be stable enough for the assignment and identification of species. Cryptic species that display subtle variations in morphology associated with reproductive isolation have been described in all major phylogenetic lineages of eukaryotic marine phytoplankton (Medlin *et al.*, 1991, Montresor *et al.*, 2003, Sáez *et al.*, 2003, Rodríguez *et al.*, 2005), despite the fact that large population sizes and ocean mixing were expected to facilitate gene flow and homogenize species distinctions. Diatoms are the youngest (Falkowski *et al.*, 2004) and the most species-rich group of phytoplankton (Mann & Droop, 1996, Kooistra et al., 2007); they have risen quickly to become important contributors to oceanic ecosystems as primary producers and intermediates in the global biogeochemical cycles of carbon and silicon (Harrison, 2000, Nelson et al., 1995, Tréguer & Pondaven, 2000). The mechanisms of speciation in diatoms remain under investigation.

Abrupt changes in an organism's genome size through polyploidy can lead to reproductive isolation and eventual speciation (Husband & Sabara, 2003, Soltis *et al.*, 2007). Diatoms are the only major group of eukaryotic phytoplankton with a diplontic life history, in which all vegetative cells are diploid and meiosis produces short-lived, haploid gametes, suggesting an ancestral selection for a life history dominated by a duplicated (diploid) genome. Polyploidization accounts for 2-4% of speciation events in flowering plants and up to 7% of speciation events in ferns (Otto & Whitton, 2000). In addition, stable polyploids were observed among laboratory populations of the diatom species *Thalassiosira weissflogii* (Grunow) Fryxell and Hasle (von Dassow *et al.*, 2008). Polyploidization may underlie the variation in chromosome number observed between and within diatom species (Kociolek & Stoermer, 1989, Giri *et al.*, 1990, Giri, 1991, Giri, 1992).

A change in genome size precipitates a cascade of cellular responses leading to nearly universal relationships among genome size, cell size and metabolic rates (Gregory, 2001, Cavalier-Smith, 2005). In accord with other divergent taxa, genome size and cell size in phytoplankton are positively correlated (von Dassow et al., 2008, Holm-Hansen, 1969, Veldhuis *et al.*, 1997, Beaton & Cavalier-Smith, 1999). Growth rates are inversely correlated with genome and cell sizes such that large-celled species with more DNA, including diatoms, grow more slowly than small-celled species with less DNA (Williams, 1964, Shuter *et al.*, 1983, Chisholm, 1992).

The relationship between cell size and genome size is of additional interest in diatoms. Asexual mitotic division produces two daughter cells, one of which is smaller than the mother cell due to the constraints of the rigid cell wall. Over time, the mean cell size within a clonal culture decreases with each successive round of division, whereas the variance in size increases (MacDonald, 1869, Pfitzer, 1871). Large cell sizes are restored through sexual reproduction, or, less frequently, through asexual enlargement (Chepurnov *et al.*, 2004). In a clonal lineage, the original sexual offspring can have 100-fold larger volumes than the smallest cells produced asexually. The smallest size that may be attained by a species is likely influenced by genome size, but the sizes of the largest cells are likely the result of genetic and environmental interactions during zygotic development.

High genetic divergence characterizes two co-occurring populations of the common coastal diatom *Ditylum brightwellii* (T. West) Grunow in van Heurk, which has a wide-spread coastal and estuarine distribution. In Puget Sound, WA, *D. brightwellii* is composed of two metapopulations that are defined by DNA sequence differences in the ribosomal internal transcribed spacers (ITS) (Rynearson *et al.*, 2006, Rynearson *et al.*, 2009). Both metapopulations consist of two or more populations (defined by differences in microsatellite allele frequencies) that can co-occur in the water column (Rynearson & Armbrust, 2000, Rynearson & Armbrust, 2004, Rynearson & Armbrust, 2005, Rynearson et al., 2006, Rynearson et al., 2009). For simplicity, the ITS-defined metapopulations will be referred to as populations throughout this study. Based on ITS sequencing of clones from the eastern and western margins of the Pacific and Atlantic Oceans including the Yellow Sea (Genbank: EU364892) and Gulf of Maine (pers. obs.), population 1 appears to have a circum-global distribution in temperate waters, while, to date, population 2 has been found only in Puget Sound (Rynearson et al., 2009). By current taxonomic definitions, individuals from both populations are members of a single species: their 18s rDNA gene sequences are identical, and there is no variation in the patterns of the silica cell wall, which are used traditionally to delineate species (Rynearson & Armbrust, 2004, Rynearson et al., 2006, Rynearson et al., 2009). There are, however, differences between individuals from the two populations. Field isolates from population 1 are smaller than those from population 2 and the peaks of their blooms are temporally separated; this separation is differentially correlated to *in situ* silicic acid concentration and daily light exposure (Rynearson et al., 2006, Rynearson et al., 2009). Even though both populations can be found in the water column at the same time, reproduction between them is limited, as evidenced by high F_{ST} values (0.286) (Rynearson et al., 2006), which are consistent with the presence of cryptic species (Caputi *et al.*, 2007).

The observed differences in cell size and the limited gene flow between populations 1 and 2 of *Ditylum brightwellii* in the field led us to test the hypothesis that a difference in DNA content is associated with differences in acclimated growth rates and cell size separating the populations. Laboratory tests were performed within the context of two geographic scales, locally within Puget Sound, and globally, including *D. brightwellii* from New Zealand.

Materials and Methods

Cell isolation and culturing

Single cells of *Ditylum brightwellii* were isolated from Puget Sound, Washington, USA and from the mouth of Akaroa Harbour, New Zealand during the spring and summers of 2006 and 2007 (Fig. I. 1). Thirty-two clonal, non-axenic cell lines were obtained by micropipetting individual cells through three aliquots of sterilized seawater into 0.5 ml f/10 medium (Guillard & Ryther, 1962) in a 48-well plate (Costar, Corning, NY). After one week, each clone was transferred to 25 ml f/2 medium and maintained as stock cultures at 13°C, with an irradiance of approximately 40 μmol photons m^{-2} s^{-1} and a photoperiod of 16:8 h light:dark.

DNA sequencing

Fifty to 100 ml of clonal culture were filtered onto 25 mm, 5 μm pore size, polycarbonate membrane filters (Millipore) for DNA extraction using either the DNeasy Plant Mini Kit (Qiagen) or the Easy-DNA Kit (Invitrogen), following manufacturer instructions. The internal transcribed spacer sequence 1 (ITS1) was polymerase chain reaction (PCR)-amplified with primers 1645F and Dit5.8sR as described in (Rynearson et al., 2006). Products from six amplification reactions were pooled and purified in one of two ways. The pooled PCR product was either directly purified using the High Pure PCR Product Purification Kit (Roche Applied Science) or electrophoresed in 1% agarose gels and bands of the appropriate size were excised and extracted from the agarose with the QIAquick Gel Extraction Kit (Qiagen). The resulting fragments were sequenced using primers 1645F and Dit5.8sR with the DYEnamic ET Terminator Cycle Sequencing Kit (GE Healthcare Bio-sciences Corp., New Jersey) and analyzed on a MegaBACE 1000 automated sequencer (GE Healthcare Bio-sciences Corp., New Jersey). Sequences were assigned to a population by aligning them to two type-sequences [Genbank:DQ329268] (population 1; ITS1-1) and [Genbank:DQ329270] (population 2; ITS1-2). Genbank accession numbers for ITS1 sequences from our study are [Genbank: GQ370472-GQ370503].

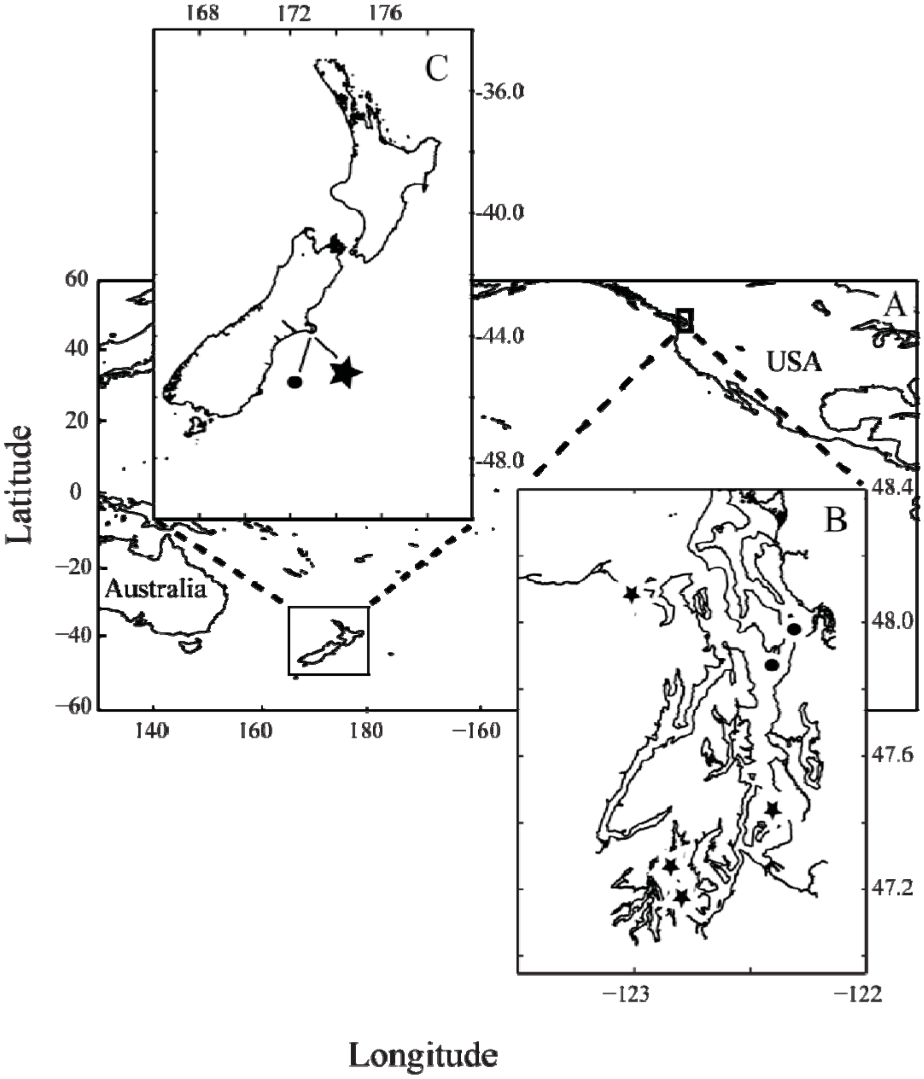


Figure I. 1. Sampling locations from which clones were isolated. A) Plan view of the Pacific Ocean with insets for: B) Puget Sound and C) New Zealand's Akaroa Harbour. Stars and circles represent sampling sites for clones isolated in 2006 and 2007, respectively.

Growth rate and size

Acclimated growth rates were determined using semi-continuous batch cultures (Brand *et al.*, 1981) of the clones isolated in 2006 plus one clone from population 1 isolated in 1997 and one from population 2 isolated in 1998 from Puget Sound by (Rynearson & Armbrust, 2000, Rynearson & Armbrust, 2004). The ITS1 sequences of the latter two clones were confirmed in this study. In total, six isolates from New Zealand, seven from Puget Sound population 1, and nine from Puget Sound population 2 were grown at 13°C under three different light conditions: continuous light of 60 and 115 µmol photons m^{-2} s^{-1}, and a 16:8 h light:dark cycle of 110 µmol photons m^{-2} s^{-1}. The 60 µmol photons m^{-2} s^{-1} continuous light experiment was completed prior to the other two, which were run simultaneously in separate incubators. Growth rates were determined by measuring chlorophyll *a* fluorescence daily with a Turner 10-AU Fluorometer (Sunnyvale, CA) and verified for a subset of clones by performing daily cell counts (data not shown). Acclimated growth rates of each clone were defined as the specific growth rates of cultures that were not significantly different over three consecutive transfers (ANCOVA; (Zar, 1996)); the common slope and associated standard error are reported as the acclimated growth rate.

Cell size was measured at two discrete times, once at the conclusion of the growth rate experiments (from the light:dark treatment) and once in conjunction with the DNA content experiments. Cells were preserved in a 1% final concentration each of formaldehyde and glutaraldehyde buffered with sterile f/2 medium. Cell width is the dimension of size reduction in *Ditylum brightwellii*; therefore, widths were measured in girdle view, perpendicular to the pervalvar (long) axis at the widest point, at 400× magnification using a Nikon Eclipse TS100 inverted microscope equipped with an ocular micrometer. Differences in size and growth rates among populations were analyzed using the statistics package SPSS (SPSS, Inc., Chicago, IL).

Relative genome size

Relative DNA content (diploid genome size) was measured using flow cytometry for 12 clones isolated in 2006 and maintained in culture for 18 months and 10 clones isolated in 2007 approximately six months and six weeks, respectively, prior to measurement. One hundred ml of each clone were grown in a 16:8 h light:dark cycle with 110 µmol photons m^{-2} s^{-1} at 13°C. Clones were harvested in mid-exponential phase, half way through the light cycle, and concentrated by centrifugation (15 min at 1700 – 2000 × g). The pellet was suspended in 15 ml of 100% methanol at 4°C for 48 h to extract chlorophyll *a* (Vaulot *et al.*, 1986). Samples were centrifuged and washed twice with 4 ml phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10.4 mM $Na_2HPO_4 \cdot H_2O$, 1.8 mM KH_2PO_4, pH = 7.4) before being resuspended in 3 ml PBS and treated with 30 µl RNase A (ca. 30 mg ml^{-1}; R4642; Sigma-Aldrich, St. Louis, MO) at 37°C for 60 min. The DNA was stained with SYBR GREEN I (Invitrogen), at 1× final concentration, for at least 20 min. Fluorescent, 1 µm latex beads (Polysciences, Warrington, PA) were added as standards to each sample. Stained samples were kept on ice in the dark until run on the flow cytometer. Each clone of a sampling set (2006 or 2007) was analyzed on a given day, and replicate clonal samples were analyzed on separate days. Clones were grown and processed independently for replicate measurements; duplicate and triplicate measurements of relative genome size were made for 2006 and 2007 isolates, respectively. Mean sample sizes ranged from 8400 – 24,000 cells per clone per replicate.

An Influx Cell Sorter (BD Biosciences, San Jose, CA) was modified to include a 500 µm sample line, and a 200 µm nozzle producing a sample stream intercepting a 300 mW, 457 nm laser focused to 20 µm. A 10× objective lens and position sensitive detectors (Swalwell *et al.*, 2009) allowed for the detection of the SYBR signal (530/20 nm bandpass filter), which was processed through an electronic integrator that produced a 16-bit data value (BD Biosciences, San Jose, CA).

Integrated SYBR signals for each cell of a single clone were usually unique; therefore, a central tendency, henceforth referred to as the mode, of the integrated SYBR signal was determined for each clone using a custom MATLAB script (MathWorks, Natick, MA). Non-uniform quantization was performed on the integrated SYBR signal for each replicate sample such that bins were variable in width. The number of cells in each bin was constant (160 cells bin^{-1}) and optimized by choosing a cell number that minimized the mean square error of the integrated SYBR values within each bin and the number of equally narrow bins. The mean value of the integrated SYBR signal from the narrowest bin, which represented the greatest concentration of cells within the smallest range of integrated SYBR signals, was taken as the mode. In the minority of cases in which two nearby bins were equally narrow, the mean of the range of the two bins was calculated as the mode. Bimodality would be indicated by equally narrow bins occurring at a distance apart.

The integrator was calibrated for linearity by collecting the integrated fluorescent signals of beads (1µm) and their doublets as the gain was increased by 2× intervals across the dynamic range of the integrator. The relationship between the linear input values of the beads and their integrated SYBR signal was determined to be of the form $f(x) = A^{bx}$, where $f(x)$ = linear value, $A = 0.6445$, $b = 5.99 \times 10^{-5}$, and x = integrated SYBR signal yielded the best fit ($SSE = 1.397$; $R^2 = 0.9972$; CurveFit Toolbox, MATLAB). Modal values of the integrated SYBR signal were applied to this equation with the linear output reported here.

Results

Thirty-two isolates from Puget Sound, WA, USA and Akaroa Harbour, New Zealand (Fig. I. 1) were assigned to one of two *Ditylum brightwellii* populations based on the seven informative sites that distinguish the two ITS1 type-sequences (Rynearson et al., 2006). The ITS1 sequences from all 10 New Zealand clones and 11 Puget Sound clones were identical to the type-sequence for population 1 (ITS1-1) from Puget Sound, including the seven informative sites, and were assigned to population 1. The ITS1 sequences for nine Puget Sound clones were identical to the ITS1-2 type sequence, including all of the informative sites, and were assigned to population 2. The ITS1 sequences for two Puget Sound clones were identical to the ITS1-2 type sequence at six of the seven informative sites. The ITS1 sequence of these two clones was polymorphic (C/T) at the fourth informative site, which was previously determined to have low-frequency variation between C and T and is therefore informative only if sequence at the other positions is known (Rynearson et al., 2009). These two clones were also assigned to population 2.

Growth rates varied among clones within each of the three groups (New Zealand and Puget Sound populations 1 and 2), but among-group differences in growth rates were greater (Fig. I. 2A-C). When clones were maintained on a light:dark cycle, the mean growth rate of clones from Puget Sound population 1 (1.24 ± 0.11 d^{-1}) was significantly faster than the mean growth rates of clones from either New Zealand (1.09 ± 0.04 d^{-1}) or Puget Sound population 2 (0.99 ± 0.13 d^{-1}) (ANOVA: $F = 11.18$, $p = 0.001$; Fig. I. 2A). Population 2 clones had the greatest range of growth rates, which included the slowest growing clones of either population, but the average growth rates of population 2 and New Zealand clones were not significantly different (Fig. I. 2A). Under conditions of continuous light, a majority of the population 2 clones failed to acclimate to the conditions within the eight week duration of each experiment

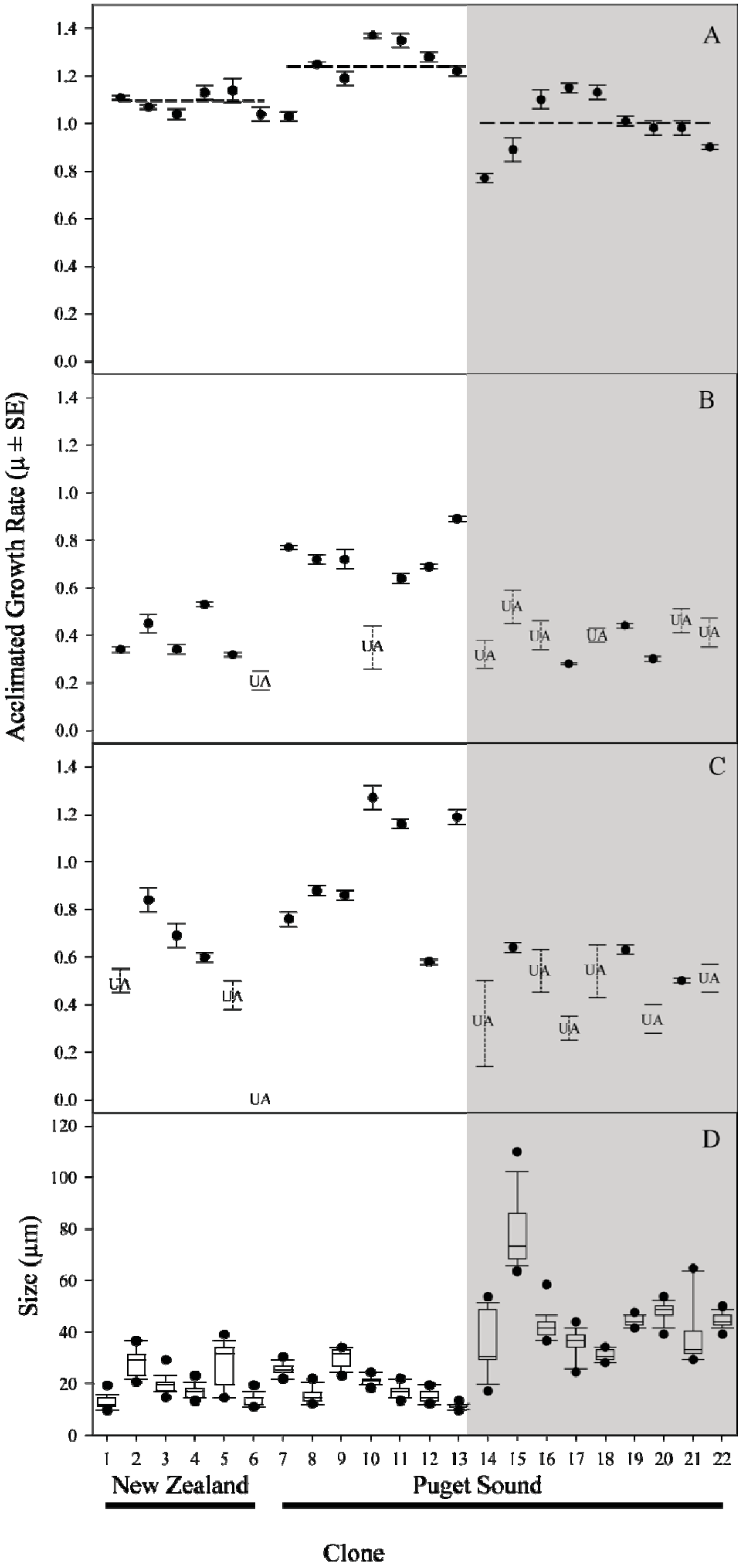

A
B
C
D
Acclimated Growth Rate (μ ± SE)
Size (μm)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
20
40
60
80
100
120
UA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
New Zealand
Puget Sound
Clone

Figure II. 2. Acclimated growth rates (A-C) and size distributions (D) of *Ditylum brightwellii*. Growth conditions: A) 110 µmol photons m^{-2} s^{-1}; 16:8 L:D; B) 115 µmol photons m^{-2} s^{-1}; 24 hour light; C) 60 µmol photons m^{-2} s^{-1}; 24 hour light. For acclimated clones, the mean growth rate ± standard error is provided. UA denotes clones unable to acclimate and the associated error bars indicate the range in growth rates. D) Boxplot parameters for the distributions of cell width (µm): bar = median; box = 1st and 3rd quartiles; whiskers = 10th and 90th percentiles; filled circles = outliers. N = 120 cells per clone. In all panels, white regions represent population 1 clones and grey regions represent population 2 clones.

(Fig. I. 2B, C). These un-acclimated clones grew, but no set of three consecutive transfer cultures grew at the same rate. In contrast, a majority of the population 1 clones were able to acclimate to growth under continuous light.

Cell widths were measured for clones included in the light:dark treatment of the growth rate experiments. Population 1 clones from New Zealand and Puget Sound had the same modal cell width of 14.6 µm (Fig. I. 2D). Population 2 clones from Puget Sound had a modal cell width of 43.9 µm, and were significantly larger than population 1 clones (Mann-Whitney U test; U = 29,672; *p* = 0.000), although individual clones from both populations did have overlapping size ranges (Fig. I. 2D).

The relative difference in DNA content between the populations of *Ditylum brightwellii* was determined using flow cytometry to measure the integrated SYBR GREEN I fluorescent signal of 12 clones from the growth rate experiments and 10 fresh isolates. The resulting distributions of DNA content were unimodal regardless of which clone was analyzed, suggesting that when grown exponentially on a 16:8 light:dark cycle all clones had progressed similarly through the cell cycle and were sampled while the majority of cells were in G1. The DNA distributions of single clones from population 1 and 2 were significantly different from each other (Fig 3; Mann-Whitney U test, U = 3 x 10^{7}, *p* = 0.000). To diminish the potential occurrence of multi-modality around the peaks of the distributions, non-uniform quantization of the signal, rather than traditional histograms, was used for subsequent analyses. The non-quantized value is termed the mode. Both methods of analysis resulted in distributions centered on single values (the modes) interpreted as the diploid DNA content of G1 cells. There was no indication of a second mode representing G2+M cells (*e.g.* Fig. I. 3). The DNA content per cell was not significantly different among individual clones within a population; the mean integrated SYBR signals of population 1 clones were 2.72 ± 0.78 and 3.11 ± 0.50, in relative units, for New

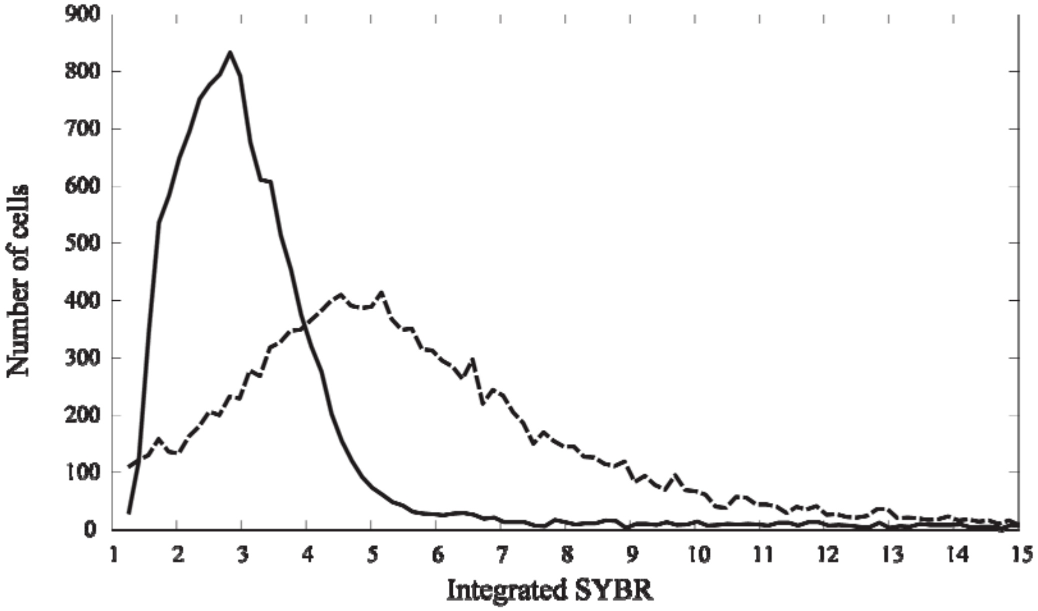


Figure I. 3. Linearly calibrated integrated SYBR signals of two clones of *Ditylum brightwellii*. Population 1 (solid line) and population 2 (dashed line) are represented by histograms each consisting of 200 bins of equal size.

Zealand and Puget Sound isolates, respectively. In contrast, the per cell DNA content of clones from population 2, 5.65 ± 0.82 relative units, was significantly greater than that of clones from population 1 (Fig. I. 4A; ANOVA: F = 73.29, *p* = 0.000). The ratio of the relative DNA content between populations 1 and 2 was 1.94 ± 0.74.

Clones associated with the DNA content experiments were isolated 18 months apart and cultured without controlling for size increases or decreases inherent in the diatom life history. Distributions of cell width were pooled for clones within each population, and the resulting two distributions were significantly different (Mann-Whitney U test, U = 63,932, *p* = 0.000). Population 2 cells, which have a two-fold larger DNA content, tended to have greater minimum and maximum widths than cells from population 1 (Fig. I. 4B). A clear correlation between genome size and cell size is obscured by the overlap of cell width distributions of individual clones from each population (Fig. I. 4B). For example, clones 14-17 from Puget Sound population 1 and clones 18-20 from population 2 are of similar size, yet they have distinct genome sizes representative of their population of origin (Fig. I. 4A and B).

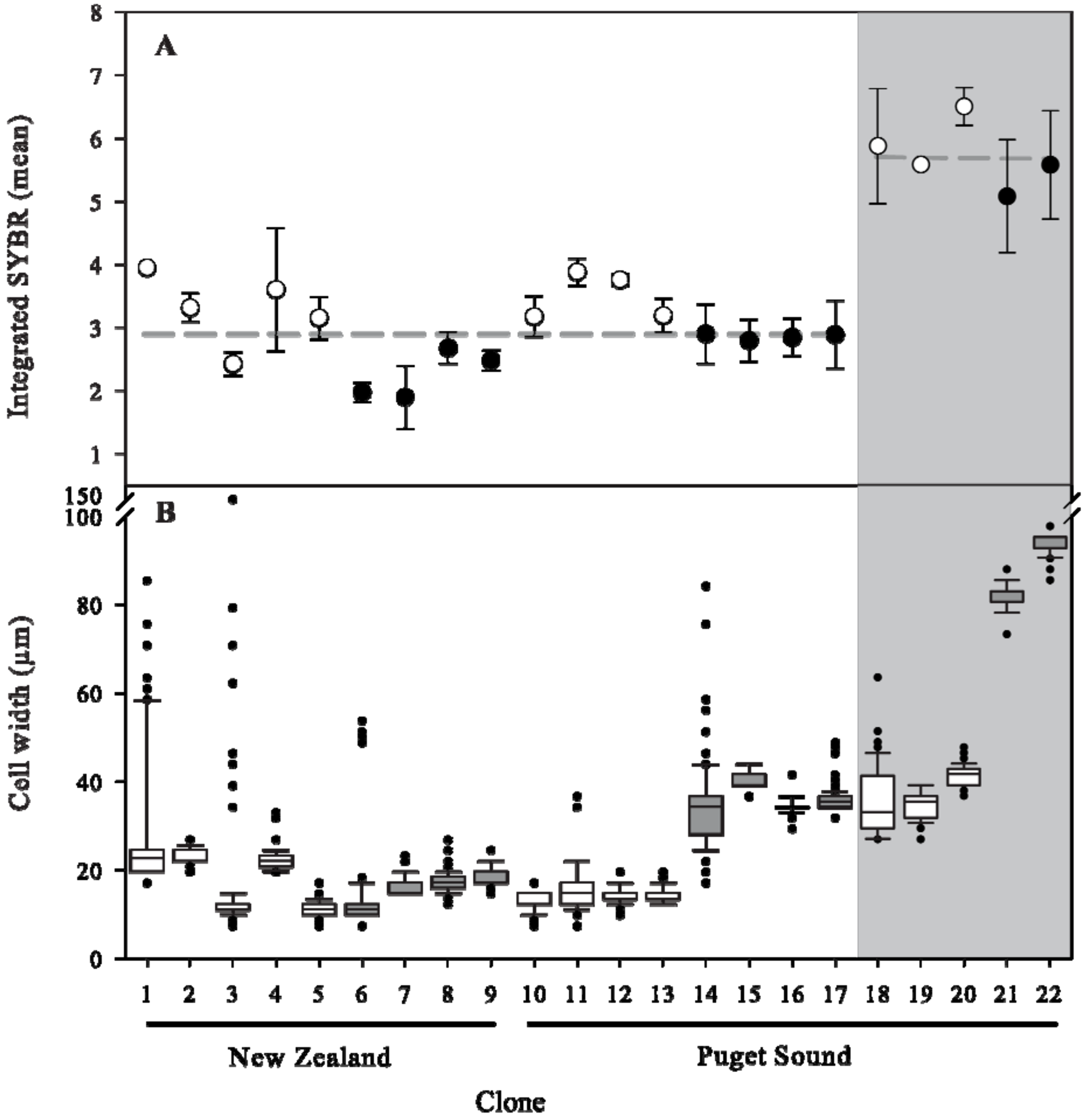


Figure I. 4. Relative genome sizes (A) and cell size distributions (B) of *Ditylum brightwellii* from two populations. A) The mean mode of the linearly calibrated integrated SYBR signal is given for clones collected in 2006 (white circles) where the whiskers represent the actual values of each duplicate, and clones collected in 2007 (black circles) where the whiskers represent the standard deviation of the triplicate samples. B) Size distributions of the cell width (µm) in each clone at the time of the flow cytometry measurements. N = 100 cells per clone. Boxplot parameters: bar = median; box = 1st and 3rd quartiles; whiskers = 10th and 90th percentiles; filled circles = outliers. White boxes represent 2006 clones and grey boxes represent 2007 clones. White panel-regions represent population 1 clones and grey panel-regions represent population 2 clones.

Discussion

Diatoms are a relatively young, but diverse, group of eukaryotic micro-organisms that arose approximately 250 million years ago (Sorhannus, 2007). The genomes of diatoms appear to be highly flexible and evolve rapidly with respect to their size and gene content, which allows for ecological differentiation (von Dassow et al., 2008, Créach *et al.*, 2006, Oliver *et al.*, 2007, Bowler *et al.*, 2008, Armbrust, 2009). Here we present evidence that genomic flexibility underlies the recent divergence of two closely related populations of *Ditylum brightwellii*, distinguished from each other by a two-fold difference in DNA content. This difference in genome size appears to be stable regardless of the amount of time isolates from the two populations have been maintained as laboratory cultures.

The DNA content of *D. brightwellii* was previously estimated to be 12.9 pg per cell (Holm-Hansen, 1969), which is the equivalent of 12.6 GB of DNA (assuming 980 MB pg^{-1}; (Cavalier-Smith, 1985)) distributed amongst anywhere from 12-50 chromosomes in each diploid cell (Gross, 1937, Eppley *et al.*, 1967). At least part of the explanation for the wide range in chromosome number is because karyotyping of diatoms is complicated by the presence of a rigid silica cell wall that prevents the consistent flattening of cells required to spread the condensed chromosomes apart for accurate enumeration (Sarma, 1983). We instead modified a flow cytometer to analyze thousands of the *D. brightwellii* cells that are up to 150 μm in width to gain an accurate estimate of the range of relative genome sizes for cultured isolates from Puget Sound, WA and from New Zealand. All distributions of DNA content were unimodal with a skew (right hand tail) towards higher DNA content. Actively dividing cells appear to spend the greatest proportion of their cell cycle in phases G1, forming the peak (or mode) of the distribution, and S, represented by cells under the tail. *Thalassiosira weissflogii*, a diatom with a six to twelve-fold smaller genome relative to *D. brightwellii* (Koester et al. unpublished data), spends a third of its cell cycle in S phase when maintained in continuous light and otherwise optimal growth conditions (Vaulot et al., 1986), suggesting that the S phase of the much larger genome of *D. brightwellii* represents an even greater proportion of the cell cycle. The level of experimental replication and consistency of the time of day that the cultures were collected and preserved argues strongly that cell cycle dynamics were comparable for clones of both populations, with few cells residing in the G2+M phase.

The most parsimonious explanation for the observed two-fold difference in DNA content between the two populations of *D. brightwellii* is a whole genome duplication event occurring within cells of population 1 creating a polyploid lineage that ultimately gave rise to population 2 cells. Mitotic and meiotic failures are potentially important and immediate mechanisms contributing to DNA content differences among diatom lineages, assuming that cells survive the initial mutation and are able to propagate asexually. Triploid and tetraploid zygotes have been observed in five genera of pennate diatoms (Mann & Stickle, 1991, Mann, 1994, Chepurnov & Roshchin, 1995, Chepurnov & Mann, 2003, Chepurnov *et al.*, 2002), and non-disjunction of spermatocytes during meiosis was observed in the same species as stable polyploids (von Dassow *et al*., 2008). Alternatively, genome size may increase via aneuploidy, gene duplication, or the rapid proliferation of transposable elements. Transposable elements were previously proposed as a primary mechanism of genome size expansion in diatoms (Bowler *et al*., 2008) because there was no evidence of whole genome or segmental duplications in either of the two diatoms with completed genome sequences, yet both contained long-terminal-repeat retrotransposons (Bowler *et al*., 2008, Armbrust *et al*., 2004).

The relationship between genome size and cell size spans five orders of magnitude for both factors and is a nearly universal trend in eukaryotes (Cavalier-Smith, 2005). Diatoms, however, are unique in that cell-wall structure causes size reduction during asexual reproduction and different cells within a clone may have vastly different sizes; therefore, a clear correlation between cell size and genome size among closely related species with genomes of similar size is unlikely to be found. Nevertheless, there are indications that genome size influences the minimum size of *Ditylum brightwellii* clones. The appearance of large cells in clones dominated by small sizes suggests that the majority of population 1 was within the sexually inducible size range for *D. brightwellii* (Koester *et al.*, 2007), and that minimum cell sizes were likely being approached by all of the clones isolated in 2006 and maintained in culture for 18 months. It is notable that the smallest cell sizes found in clones of population 2, which has the larger DNA content, tended to be larger than the smallest cell sizes found in clones from population 1. Most importantly, with regard to genome size and cell size, the diploid genome size is consistent among clones within a population, regardless of the variation in cell size caused by life history constraints.

Similar to traditional common garden experiments, our growth rate experiments tested for genetic differences among clones of *Ditylum brightwellii* from two populations, differentiated by genome size, and two groups of clones from population 1 represented by different geographic origins. The differences in growth rates between groups, populations 1 and 2 and New Zealand versus Puget Sound population 1 isolates, were greater than any within group variability. Growth rate variability may occur at the extremes of cell size within a clone (Paasche, 1973, Amato *et al.*, 2005), but those effects would be masked in our study by cell size variability within each clone and diminished by the large difference in growth rates between the groups. Population 2 clones grew more slowly than population 1 clones from the same region, consistent with the expectation that cells with larger cell sizes and larger genomes will have slower growth rates. However, population 1 clones from New Zealand grew more slowly than population 1 clones from Puget Sound, suggesting that other genetic factors are also responsible for setting rates of growth. New Zealand clones are likely adapted to oceanic conditions that are distinct from the estuarine waters of Puget Sound.

Rapid selection for cell and/or genome size among diatoms is indicated in the fossil record and appears to be associated with climate variability (Finkel *et al.*, 2005). Large cells (> 100 μm diameter) of the morpho-species *Azpeitia nodulifera* (A. Schmidt) Fryxell and Sims (previously described as *Coscinodiscus nodulifer* Schmidt) intermittently enter and exit the sedimentary record, normally dominated by cells 40 μm in diameter, over the course of thousands of years (Arrhenius, 1952, Burckle & McLaughlin, 1977). The lack of variability in morphology and the 18S rDNA gene between the two populations of *Ditylum brightwellii* suggests that the genome duplication event was recent and rapid. Genomic plasticity has likely contributed to speciation among diatoms and may be an important factor in the adaptation of diatoms to future ocean conditions.

Conclusion

The majority of phytoplankton have haplontic life histories (Graham & Wilcox, 2000); therefore, a transition to a stable duplicated genome and a diplontic life history are likely at the root of the diatoms, the only phytoplankton group whose membership is diplontic. The propensity for further duplications may be a key mechanism of speciation among diatoms. Speciation is best identified by using a suite of divergent traits including reproductive isolation, morphology, ecological and physiological differences facilitated by genetic divergence. A high F_{ST} value between the two populations of *Ditylum brightwellii* already indicated that that the process of reproductive isolation was underway, and that these two populations could represent separate species (Rynearson et al., 2006). Duplicated genes, arising from the hypothesized whole genome duplication between the populations, may be released from selective pressures allowing for mutations that may be masked until environmental regimes change and they provide an adaptive advantage to the cell (Ohno, 1970). The apparent localization of population 2 to Pacific Northwest waters appears to reflect a recent divergence of the populations initiated by a whole genome duplication event. Population 1 and 2 most likely represent cryptic species in which interbreeding is greatly reduced and phenotypic differentiation is enhanced. In conjunction with previous work on genome size variation in diatoms, these results suggest that polyploidization is an active mechanism contributing to the diversification and speciation of marine diatoms.

Authors' contributions

JAK participated in and implemented all levels of this study including experimental design, statistical analysis and drafting of the manuscript. JS executed the design, development and operation of the flow cytometer. PvD participated in preliminary experiments, final experimental design and drafting of the manuscript. EVA participated in experimental design and drafting of the manuscript. All authors have read and approve of the final manuscript.

Acknowledgements

We would like to thank Mark Webber and Raphael Hoekstra for New Zealand samples, Karie Holtermann for selected Puget Sound samples and Tatiana Rynearson for two early clonal isolates. We also thank Rachelle Lambert for laboratory assistance and Ger van den Engh and Timothy Petersen for consultation on flow cytometry. A special thanks to Vaughn Iverson for discussions on signal processing and to Kate Hubbard, Colleen Durkin and Adrian Marchetti for insightful discussions throughout the research. This research was funded by the Gordon and Betty Moore Foundation Marine Microbiology Investigator Award to E.V.A. PvD was supported by a Marie Curie International Incoming Post-doctoral Fellowship (European Commission) and an NSF Microbiology Post-doctoral Fellowship.

Chapter II: Positive selection in a diatom acts on conserved and lineage specific genes affecting transcriptional regulation and protein interactions

In preparation for publication
Authors: Julie A Koester, Willie J. Swanson, E. Virginia Armbrust

Abstract

Diatoms are the youngest and most species-rich group of phytoplankton, having adapted to both marine and terrestrial ecosystems. They are one of the most ecologically important groups of organisms in the ocean, contributing to primary production by forming dense blooms through asexual reproduction. Mutations acquired during each cell division provide genetic, and thus phenotypic, variability upon which natural selection may act. The genes promoting adaptation are positively selected and detected by determining the ratio of amino acid changing substitutions (d_N) to silent substitutions (d_S) within homologous genes of related organisms. Genome and transcriptome-wide pair-wise comparisons within three diatom genera, *Pseudo-nitzschia, Ditylum,* and *Thalassiosira* were made, allowing detection of positive selection (d_N:d_S > 1.0) at decreasing phylogenetic distances. The signal of positive selection was greatest between two strains of *T. pseudonana*, whereas purifying selection (d_N:d_S << 1.0) dominated the genes tested between species of *Pseudo-nitzschia.* Further testing among seven strains of *T. pseudonana* yielded 809 candidate genes of positive selection, representing 7 % of those coding for proteins. Orphan genes and genes encoding protein binding domains and transcriptional regulators were enriched within the set of positively selected genes relative to the genome as a whole. Positively selected genes may be enhancing survival during suboptimal growth conditions. For example, the differential co-expression of positively selected genes encoding putative cell wall proteins and a putative anti-pathogenic protein suggests that there is an adaptive response to growth in nutrient limited conditions in one strain of *T. pseudonana*. In addition, strains isolated from waters with relatively stable temperatures had the greatest number of positively selected genes among different subsets of strains tested. This was the first genome scan performed within a group of phytoplankton to detect positively selected genes. These results present an opportunity to test new hypotheses that integrate positively selected genotypes

in *T. pseudonana* with their associated phenotypes and selective forces in both natural populations and the laboratory.

Introduction

Diatoms are evolutionarily the youngest members of the phytoplankton, originating near the boundary of the Permian and Triassic periods ca. 250 mya, a time of mass extinction in the world ocean (Raup, 1979, Sorhannus, 2007). The number of diatom species has increased within the last 65 my while the global silicic acid concentration has decreased in response to its precipitation into the silica of diatom cell walls (Siever, 1991, Falkowski et al., 2004). In today's oceans, diatoms are the most species-rich group of phytoplankton (Kooistra et al., 2007), they control the biological portion of the silicon cycle, and are important primary producers (Maliva *et al.*, 1989, Tréguer & Pondaven, 2000). Until recently, marine populations were expected to be genetically homogenous and distributed widely by ocean currents. Instead, even marine species with circum-global distributions have structured populations (Casteleyn *et al.*, 2010, Rynearson & Armbrust, 2004). The radiation of diatoms into marine, freshwater, soil and ice ecosystems is evidence of genetic specialization and an exemplary ability to adapt.

Environmental fluctuations affecting diatoms, especially marine species, which are the focus of this research, can occur with short periodicity. Estuarine and coastal species must be adapted to daily fluctuations in light that may range from stressfully intense levels to very dim due to turbidity, in contrast to the consistently clear water in which open ocean species live (Lavaud et al., 2007). Coastal species may experience transitions from cool, high nutrient, upwelling conditions to relaxed, warmer, low nutrient, downwelling states that change within a week (Austin & Barth, 2002). The abundance of diatoms is normally low in the open ocean because nutrient concentrations are low, but the formation of cold-core mesoscale eddies mixes nutrients to the surface and allows diatoms to grow quickly into "blooms"; eddies can form within days and persist for up to a month (Benitez-Nelson *et al.*, 2007).

Diatoms have a high capacity to accrue mutations upon which selection may act. They are diploid, and mutations may accumulate because they are masked by redundant gene copies until environmental conditions change, rendering them useful. The life history of diatoms is dominated by asexual reproduction. Division rates are relatively quick, and cells can divide at least once per day during bloom conditions (Furnas, 1990), presenting an opportunity to acquire new mutations. Sexual generation times are ca. 2 years in marine diatoms (D'Alelio *et al.*, 2010, Holtermann *et al.*, 2010); thus, genetic variation is increased further and surviving mutations are passed along relatively quickly. The mutational load carried by diatoms is evident in the sequenced genomes of *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. *Thalassiosira pseudonana* and *P. tricornutum* have accrued as much genetic diversity in 90 my as mammals and fish have in 550 my (Bowler *et al.*, 2008); however, when normalized to generation time, the number of neutral substitutions per generation is equivalent between the pairs. In addition, the genome of *T. pseudonana* has, on average, one polymorphism per every 150 bases (Armbrust *et al.*, 2004), which is an order of magnitude less than the purple sea urchin (Sodergren *et al.*, 2006), but an order of magnitude more than humans (Altshuler *et al.*, 2001).

Natural selection acts directly on phenotypes, and those that remain within a population are generally adaptive. The gene variants associated with those adaptive phenotypes are said to be under positive selection. Positive selection is detected in protein coding genes by determining the ratio of the number of non-synonymous substitutions at possible non-synonymous sites to the number of synonymous substitutions at synonymous sites (d_N:d_S) in homologs from two or more organisms. The conventional definition of positive selection requires d_N:d_S to be greater than one, such that amino acid changes, resulting from non-synonymous mutations, in the encoded protein are beneficial. In genes under purifying selection, amino acid changes are not tolerated within the encoded protein, therefore d_N:d_S is much less than one. When genes are evolving neutrally, the rate of synonymous and non-synonymous mutations are equal. Maximum likelihood estimation allows DNA alignments of protein coding genes and their phylogenies to be tested against different parameter-rich models of evolution (*e.g.* neutral vs. selection) (Yang, 2007). Sites models estimate a distribution of d_N:d_S values across the codons of a gene for all branches within a tree. Branch models estimate one d_N:d_S per set of branches of a tree, and branch-site models estimate different d_N:d_S distributions for specified branches (Yang, 2007). The latter models can be used to test hypotheses related to the ecological histories of certain lineages.

For a gene to become positively selected it must first pass through a period during which purifying selection is relaxed, meaning that non-synonymous mutations are allowed into the gene and homologous protein sequences diverge. Rates of protein divergence tend to be higher in lineage-specific, or “young” proteins, than in proteins with deep evolutionary histories (Alba &

Castresana, 2005, Wolf *et al.*, 2009). Concomitantly, mRNA expression of highly diverged and positively selected genes is frequently restricted to specific tissues (Kosiol *et al.*, 2008, Oliver *et al.*, 2010) and is lower than that of conserved genes which are responsible for basic cellular functions such as protein production, maintenance and division (Pál *et al.*, 2001, Subramanian & Kumar, 2004). Purifying selection is also relaxed in proteins found at the periphery of networks, and on the extracellular surface of cells interfacing with the environment (Julenius & Pedersen, 2006, Kim *et al.*, 2007). Proteins encoded by positively selected genes are frequently mediators of signal transduction, functioning in cell-cell recognition, immune response, and gamete recognition; other reproductive proteins, membrane and intracellular transporters are also selected (Bustamante *et al.*, 2005, Castillo-Davis *et al.*, 2004, Li *et al.*, 2009, Nielsen *et al.*, 2005, Namroud *et al.*, 2008, Voolstra *et al.*, 2011).

Studies of positive selection in single-celled eukaryotes and phytoplankton, specifically, remain limited (Li *et al.*, 2009, Voolstra *et al.*, 2011), primarily because little sequence data is available. Selected sites within the sexually induced gene (*SIG1*) appear to be under positive selection among four species of *Thalassiosira* (Sorhannus & Pond, 2006), but not within different strains of *Thalassiosira weissflogii* (Grunow) Fyxell and Hasle (Sorhannus, 2003, Suzuki & Nei, 2004). The ecologically important silicon transporter (*SIT*) gene family experiences strong purifying selection among 45 marine and freshwater species within the Thalassiosirales separated by 75 my of divergence (Alverson, 2007). It is possible that the detection of positive selection within the *SIT* gene family was hindered by saturating rates of synonymous mutation, which is evident among the three *SIT* genes of *P. tricornutum* (Sapriel *et al.*, 2009).

The extent of positive selection was quantified within three genera of diatoms, *Pseudo-nitzschia, Ditylum brightwellii* and *Thalassiosira pseudonana* using transcriptomic and genomic sequences. These diatoms were chosen because they represent a gradient of phylogenetic relatedness from the well established species of *Pseudo-nitzschia* that diverged 5 – 10 mya to strains within the cosmopolitan species *T. pseudonana* that diverged from *Detonula confervacea* ca. 2 mya (Sorhannus, 2007). The two cryptic species of *Ditylum brightwellii* are differentiated by genome size and likely diverged recently because the larger genome-sized species is thus far found only in the northeastern Pacific Ocean (Koester *et al.*, 2010). By rigorously testing genes within seven strains of *T. pseudonana* using a phylogenetic framework we were able to identify a large suite of positively selected genes, and to explore links between the genes, potential phenotypes and environmental selective forces using gene expression profiles and branch-site models of selection among different strains.

Materials and Methods

Sequence data and identification of homologous genes

Transcriptomic and genomic data were used to detect positive selection in three diatom genera, *Pseudo-nitzschia, Ditylum* and *Thalassiosira.* The *Pseudo-nitzschia* and *Thalassiosira* data sets were compiled from archives of E.V. Armbrust, University of Washington, Seattle, WA, USA and collaborators. cDNA was sequenced by the Sanger method for three species of *Pseudo-nitzschia , P. australis*, *P. multiseries* and *P. multistriata.* The *Pseudo-nitzschia multistriata* sequence was provided by U. John at the Alfred Wegener Institute, Bremerhaven, Germany and W.H.C.F. Kooistra at the Stazione Zoologica Anton Dohrn, Naples Italy. cDNA from two clones of *Ditylum brightwellii*, one from each of the two cryptic sister species population 1 and population 2 (Koester et al., 2010), was sequenced using pyrosequencing technology (Schuster Laboratory, University Park, PA.) and is available (CAMERA 2.0 Portal). The sequenced reads of *D. brightwellii* were assembled with CABOG default settings (Miller *et al.*, 2008).

Putatively homologous pairs of genes between three species of *Pseudo-nitzschia* and two cryptic species of *Ditylum brightwellii* were identified and prepared for analysis using an automated pipeline that integrated custom scripts with extant bioinformatic tools. In brief, the open reading frame (ORF) for each gene contig was determined by retrieving all possible ORFs at least 25 amino acids in length using *getorf* (EMBOSS) and comparing them to protein databases including those of several phytoplankton and the non-redundant protein database of NCBI using *blastp* at an *e*-value cut-off of 10^{-3} (Altschul *et al.*, 1997). The contig with the lowest *e*-value was selected for further analysis. If a contig did not have a protein match, the longest ORF was chosen. Putative homologs were paired between diatom species using a best reciprocal *blastp* at an *e*-value cut-off of 10^{-10}. Alignments of protein pairs were made using CLUSTALW2 (Larkin *et al.*, 2007), and converted to the original DNA sequence with *revtrans.py* (Wernersson & Pedersen, 2003). Aligned homologs were trimmed to blunt ends with at least three identical amino acids at each end.

Seven strains of *Thalassiosira* were sequenced with an ABI SOLiD and mapped with BWA 0.5.9 (parameters: bwa aln -k 2 -l 18 -n .001) to the reference strain (CCMP 1335) previously sequenced using the Sanger method (Armbrust et al., 2004). SOLiD sequenced reads were trimmed based on quality (Iverson *et al.*, 2012); reads less than 24 bp that did not meet the quality threshold were discarded. The majority consensus sequence of each strain was used to make gene alignments with start, stop and intron boundaries established from the *Thalassiosira pseudonana* v. 3.0 gene models (http://genome.jgi.doe.gov/Thaps3/Thaps3.home.html). Introns were removed.

Pair-wise tests for detecting positive selection within and among species

Gene alignments were converted to the PHYLIP format for input into the software package Phylogenetic Analysis by Maximum Likelihood (PAML version 4.4 (Yang, 2007)). Each set of paired genes was analyzed using the PAML program *codeml* in runmode = -2 (pairwise), model = 0, NSsites = 0 (one d_N:d_S value), and fix_omega = 0 (estimates omega) and cleandata = 1 such that sites with ambiguity codes were removed from analysis. The proportion of amino acid substitution per non-synonymous (d_N) sites and the proportion of silent amino acid substitutions per synonymous sites (d_S) were assessed individually for each gene pair. The ratio of d_N:d_S was used to evaluate at what phylogenetic distance positive selection could be detected (d_N:$d_S > 1.0$); genes saturated for synonymous substitutions ($d_S > 1.0$) were removed from this analysis.

Testing for positive selection among seven strains of Thalassiosira pseudonana

The pair-wise analysis was treated as an initial screen for genes of interest that could be tested further. Approximately 80 % of genes with a d_N:$d_S \geq 0.5$ are under positive selection (Swanson *et al.*, 2004); therefore, the 3565 genes with d_N:$d_S \geq 0.5$ in the pair-wise analysis of *T. pseudonana* were tested for positive selection among seven genetically distinct strains. A coalescent tree was generated from 3565 individual gene trees (Fig. II. 1). Gene trees and the coalescent tree were constructed with RaXML 7.2.5 (GTRPROTGAMMAWAG; (Stamatakis, 2006) and PhyloNet (unrooted minimum coalescence; (Than & Nakhleh, 2009)), respectively. Selection experiments were run using *codeml* within PAML. Individual genes were tested for positive selection by comparing a null model (M8a, nearly neutral) and a selection model (M8).

Both models allow omega (d_N:d_S for the tree of seven strains) to vary at different sites along the gene alignment. In the neutral model M8a, omega is distributed between zero and one, whereas the selection model M8 includes an additional bin allowing values of omega greater than one. The null and selection models were run three and four times, respectively, to evaluate convergence of the likelihood estimates. Any genes that did not converge within the selection model were removed from further analysis. Likelihood ratio tests (lrt) were calculated for each gene using the formula: lrt = -2.0*(Ln(likelihood)$_{\mathrm{NULL}}$ – Ln(likelihood)$_{\mathrm{SELECTION}}$). The test statistic generally follows a chi-squared distribution, therefore nominal significance (*p*-value) was determined by integrating the right-hand side of the chi-squared distribution for one degree of freedom (number of parameters: M8a = 16, M8 = 17) by the lrt statistic. Statistical significance was adjusted for multiple tests using a Bonferroni correction and a false discovery rate (FDR) of 0.01 with Q-value (parameters: lamba = range 0.0 to 0.9 by 0.05; π_0 method = bootstrapped; (Storey & Tibshirani, 2003).

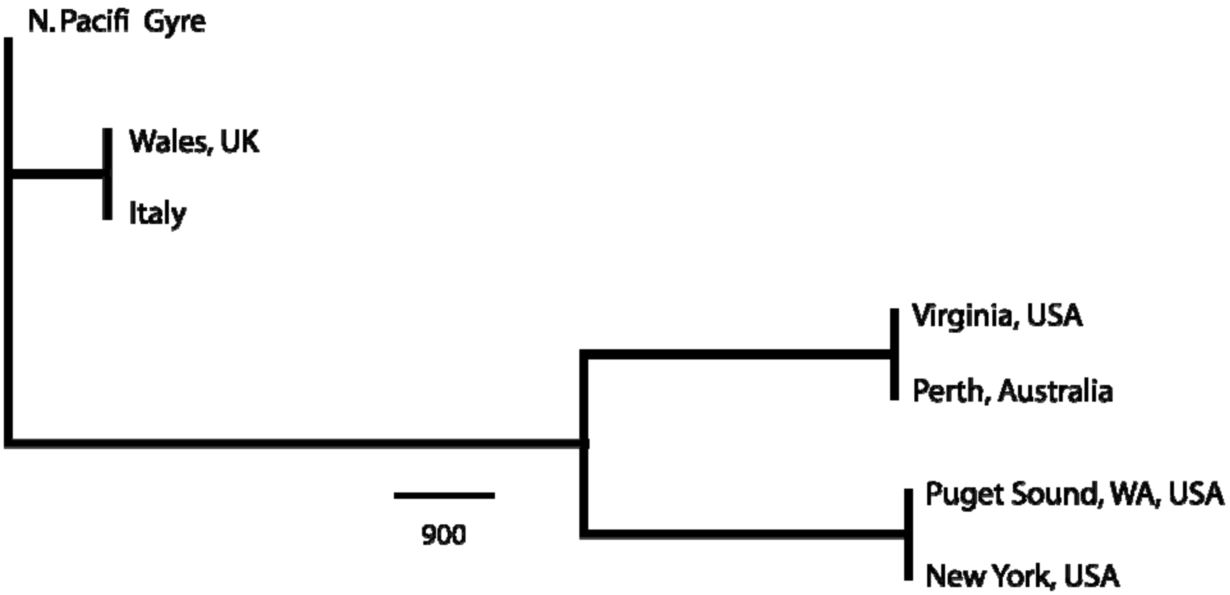


Fig. II. 1. Coalescent tree of seven strains of *Thalassiosira pseudonana*. Branch lengths are the estimated number of lineages that are required to converge on the last common ancestor.

Testing for functional enrichment of genes under positive selection.

Thalassiosira pseudonana version 3.0 gene models were submitted to InterProScan (Zdobnov & Apweiler, 2001) and *blastp* was performed against the non-redundant (nr) database at NCBI to assess their functional annotations. Orphans are operationally defined as those genes whose translated proteins either have no matches or matches with an e-value of 10^{-5} or greater from the *blast*-searched databases. Search databases included genomic and transcriptomic sequence from five additional diatoms, two oomycetes, two cryptophytes, three prasinophytes, one haptophyte, one amoeba, one ciliate, one green alga and the nr. Gene ontology (GO) terms were extracted from InterProScan results to test for significant associations of GO terms within the set of positively selected genes using GOSTATS (Falcon & Gentleman, 2007). Conditional, hypergeometric tests for over-representation of GO terms was performed separately for each of the GO ontologies, molecular function (MF), cell component (CC) and biological function (BP). A Bonferroni correction and a FDR of 0.05 were used to adjust *p*-values for multiple tests.

Expression of positively selected genes in Thalassiosira pseudonana (strain CCMP 1335)

Thalassiosira pseudonana transcription data from Mock *et al*. (2008) was used to test the hypothesis that positively selected genes have lower levels of expression than neutral or purified genes and to identify growth conditions under which positively selected genes are co-expressed. The relative level of expression for each gene was determined by taking the median fluorescence of the aggregated (previously quantile-normalized) probes and replicates of each experimental condition. Differences in the distributions of gene expression between positively selected genes and genes evolving neutrally and under purifying selection were tested for each experimental condition using one-sided Mann-Whitney tests (*wilcox.text* R v2.12.1). Two-way hierarchical clustering (Cluster 3.0 (Eisen *et al*., 1998)), using the city-blocks distance algorithm, was performed to group both genes and experimental conditions by similarity of expression patterns to identify genes that were co-expressed under specific conditions, and thus potentially co-evolving.

Testing positive selection along specified lineages of Thalassiosira pseudonana

Experiments were designed such that plausible selective forces could be associated with environmental conditions at the site each strain was collected. Specifically, the 3565 genes with

d_N:$d_S \geq 0.5$ were tested for positive selection in different subsets of strains. Two hypotheses addressed seasonal temperature variability. Strains were grouped into two groups based on the variability in seasonal sea surface temperature where they were collected and the positively selected genes associated with each group were identified. Differences in seasonal sea surface temperature (SST) were plotted and calculated using the Smith and Reynolds climatology from the NCEP NOMADS Meteorological Data Server for SST between January and July 1970 – 2000 at the location each strain was collected: (http://www.emc.ncep.noaa.gov/research/cmb/sst_analysis/#_cch2_1007146782). Hypotheses that ecosystem type and geographic isolation promote positive selection were tested using the open ocean strain (CCMP 1014) and the Adriatic strain (RcTP), respectively. These tests were performed using branch-site models. The null branch-site model A1 fixes omega to 1.0 on the branch(es) of interest, while the selection model A estimates a distribution of omega values across the gene on the chosen branch(es).

Time in culture was also tested as a potential selective force using branch models. Omega values were estimated for strains based on the decade in which they were collected. Branch models estimate one omega value per designated set of lineages for the entire gene. The null model M0 fixes omega to 1.0 and the selection model (model = 2) estimates omega along each set of lineages. Positive selection was detected for sets of strains for which the likelihood ratio test was significant and omega > 1.0 for the specified set of branches.

The branch-site and branch models used the same tree topology and statistics as described above with the exception that individual model parameters differed. Branches of the tree were annotated to support each of the different hypotheses being tested. The background omega was applied to internal branches unless both tips originating from a node were being tested with the same omega, then that internal branch would be assigned to the alternative, estimated omega.

Results

Pair-wise tests for positive selection

Pair-wise comparisons of homologs within and between species were used to determine the best phylogenetic distance at which to detect positive selection in diatoms. These tests were conservative, estimating a single d_N:d_S for each gene pair. The proportion of gene pairs saturated for synonymous substitutions was greatest in the inter-species comparisons, accounting for 89 % of those tested between *Pseudo-nitzschia multistriata* and *P. multiseries* (Table II. 1). Gene pairs saturated for synonymous substitutions could not be used to detect positive selection because the synonymous substitution rate is uncertain. Rates of synonymous and non-synonymous substitutions decreased with decreasing phylogenetic distance. The rates of synonymous substitution between homologs of the sister-species of *Ditylum* and strains of *Thalassiosira* were one and two orders of magnitude less, respectively, than the inter-species comparisons of *Pseudo-nitzschia* (Table II. 1). The length of gene alignments does not appear to affect the estimates of the rates of substitution.

Table II. 1. Rates and saturation of synonymous (d_S) and non-synonymous (d_N) mutations in homologs of three diatom genera.

				Saturated pairs $d_S \geq 1.0$		**Unsaturated Pairs d_S < 1.0**			
	Raw Seq. (#)	**Paired Genes (#)**	**Mean Align. Length (bp)**	**#**	**%**	**#**	**%**	**Mean d_N**	**Mean d_S**
***T. pseudonana* (NY)**	11390	11355	1488	9	0.08	11346	99.92	0.002	0.006
***T. pseudonana* (Wales)**	11390								
***D. brightwellii* pop. 1**	3910	113	204	4	3	110	96	0.007	0.048
***D. brightwellii* pop. 2**	477								
P. australis	920	277	324	176	64	101	36	0.090	0.726
P. multistriata	16512								
P. australis	920	404	336	266	66	138	34	0.103	0.718
P. multiseries	16535								
P. multistriata	16512	2653	483	2359	89	294	11	0.086	0.795
P. multiseries	16535								

The majority of homologous pairs that are not saturated for synonymous substitutions are under strong purifying selection with d_N:d_S ≤ 0.1 in all comparisons except the *Thalassiosira* strains (Fig. II. 2). Few *Pseudo-nitzschia* homologs had d_N:d_S > 1.0 indicating that they were under positive selection, and for most d_N:d_S was < 0.4, consistent with purifying or somewhat relaxed purifying selection. Intermediate values of d_N:d_S for both the homologs of *Ditylum* and

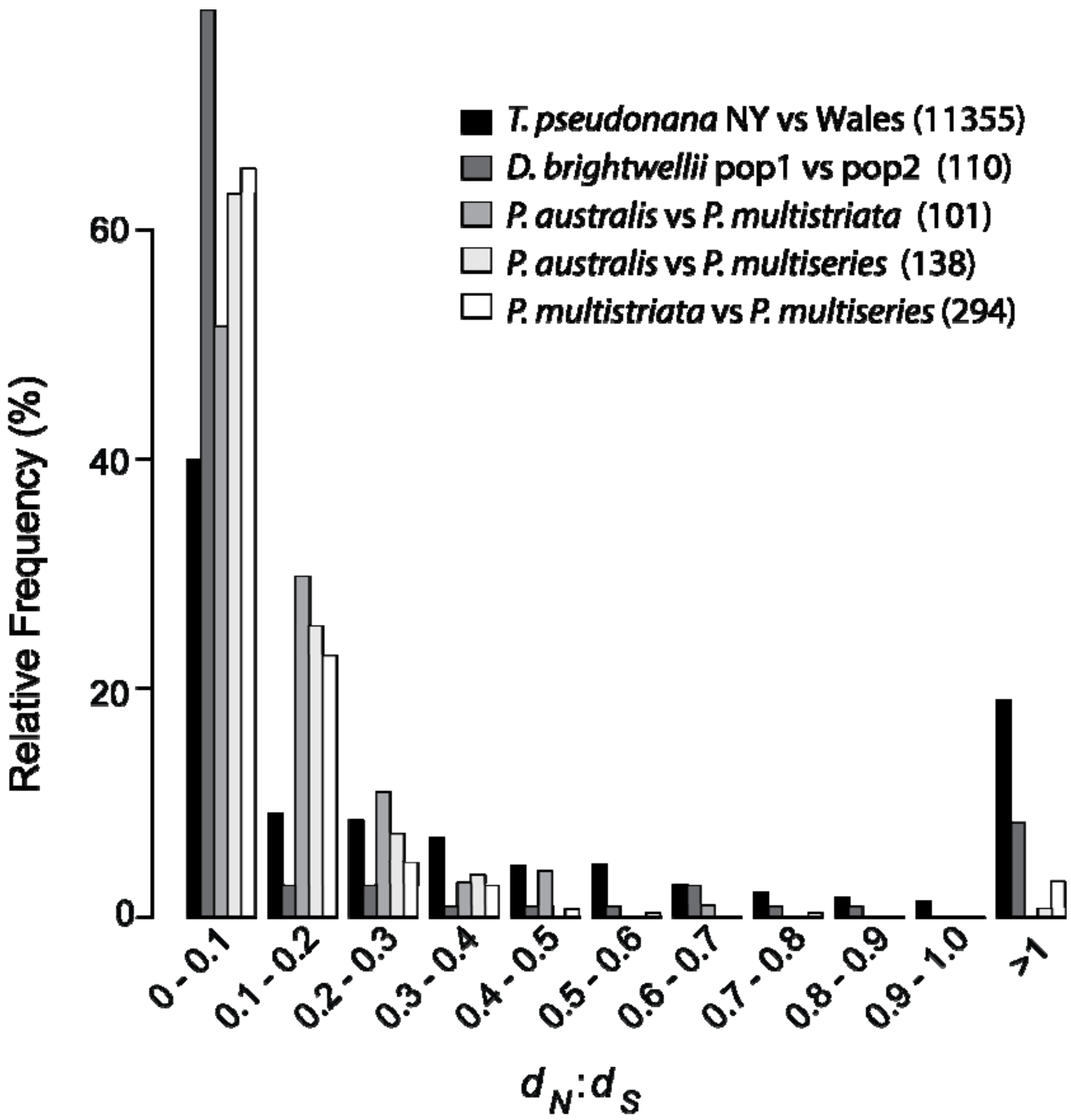


Fig. II. 2. Frequency distributions of the d_N:d_S of homologs between pairs of diatoms. Genes saturated for silent mutations (i.e. for genes where d_N:d_S < 1.0) are not included. Number of genes analyzed is in parentheses.

Thalassiosira suggested that many were experiencing relaxed purifying selection. Approximately 10 and 20 % of the *Ditylum* and *Thalassiosira* homologs, respectively, had signatures of positive selection.

Positive selection among seven strains of Thalassiosira pseudonana

The 3565 genes in *T. pseudonana* with $d_N{:}d_S \geq 0.5$ were tested more rigorously to determine the statistical support for positive selection of these genes. The maximum likelihood approach incorporated a phylogenetic model that included seven genetically distinct strains, and allowed omega ($d_N{:}d_S$ for a gene tree) to vary across the alignment of each gene. Maximum likelihood analysis identified 2035 genes with a nominal *p*-value ≤ 0.05 (Table II. S1). After correcting for multiple tests, 809 (Bonferroni, *p*-value $< 1.5\times10^{-5}$) and 1784 (FDR = 0.01) genes emerged as strong candidates of positive selection representing 7 and 16 % of the protein coding genome, respectively (Table II. S1). The set of 809 genes is a subset of 1784 genes, and will be referred to as the positively selected set.

Functions of proteins encoded by the 809 positively selected genes were assessed using the Gene Ontology (GO) data structure and InterPro for greater resolution of information concerning protein domains. GO terms were assigned to 329 (41 %) of the 809 positively selected genes and 5875 (52 %) of the 11390 total coding genes. Enrichment analysis was performed for those genes with GO terms. Six GO terms, that included 146 unique genes, were over-represented in the positively selected set of genes relative to the distribution of GO terms for all protein coding genes. The majority (112) of over-represented genes encoded proteins involved in protein-protein interactions (Table II. 2). One third (34 %) of the 329 positively selected genes with GO terms were represented by GO:0005515, the term for protein binding, in contrast to 18 % of all genes with GO terms. Genes encoding biosynthetic and metabolic regulatory proteins were also over-represented within the positively selected set (Table II. 2). Transcription factors dominated the regulatory genes with the majority (25 of 28) represented by GO:0009889; 8.5 % of positively selected genes versus 4.3 % of all genes with GO terms were represented by this term (Table II. S2). Protein domains with the greatest representation are WD40, zinc fingers, tetratricopeptides (TPR), PDZ and domains associated with heat shock (Table II. S2). Orphan genes, for which no homologs were found in other organisms, represent 24 % (191 genes) of the positively selected genes and 15 % (1718 genes) of all coding genes.

Table II. 2. Functional enrichment in 809[a] positively selected genes of *Thalassiosira pseudonana.*

Ontology	*p*-cut[c]	GO ID	*p*-value	Odds-Ratio	Exp[d]	Obs[e]	Total[f]	Term	
MF[b]	18	GO:0005515	6.05E-12	2.43	62	112	1098	protein binding	*†
		GO:0009982	0.001353	4.34	2.23	8	39	pseudouridine synthase activity	
CC	6	GO:0005634	0.007638	1.89	16	26	313	nucleus	
BP	23	GO:0009889	2.69E-05	2.63	13	29	282	regulation of biosynthetic process	*†
		GO:0060255	3.53E-05	2.54	14	30	301	regulation of macromolecule metabolic process	*†
		GO:0090304	4.20E-05	2.19	24	43	530	nucleic acid metabolic process	*†
		GO:0051171	7.57E-05	2.46	14	29	298	regulation of nitrogen compound metabolic process	*†
		GO:0031323	0.000126	2.34	15	30	322	regulation of cellular metabolic process	*†
		GO:0080090	0.000194	2.31	15	29	314	regulation of primary metabolic process	†
		GO:0001522	0.00022	5.90	2	8	37	pseudouridine synthesis	†
		GO:0006355	0.00039	2.48	10	22	219	regulation of transcription, DNA-dependent	†
		GO:0032774	0.000819	2.33	11	22	231	RNA biosynthetic process	†
		GO:0050789	0.000878	1.89	24	39	513	regulation of biological process	†
		GO:0034641	0.001622	1.64	47	65	1017	cellular nitrogen compound metabolic process	†
		GO:0016567	0.001646	10.51	1	4	12	protein ubiquitination	†

[a] 329 of 809 genes were annotated by GO terms

[b] Ontologies: MF = molecular function, BP = Biological Process, CC = cellular component.

[c] *p*-cut: the number of GO IDs with p-values less than 0.05 in the hypergeometric tests for each GO ID.

[d] Exp: expected number of genes for the GO term in positively selected set if the function is not enriched

[e] Obs: Observed number of genes in the positively selected set of genes

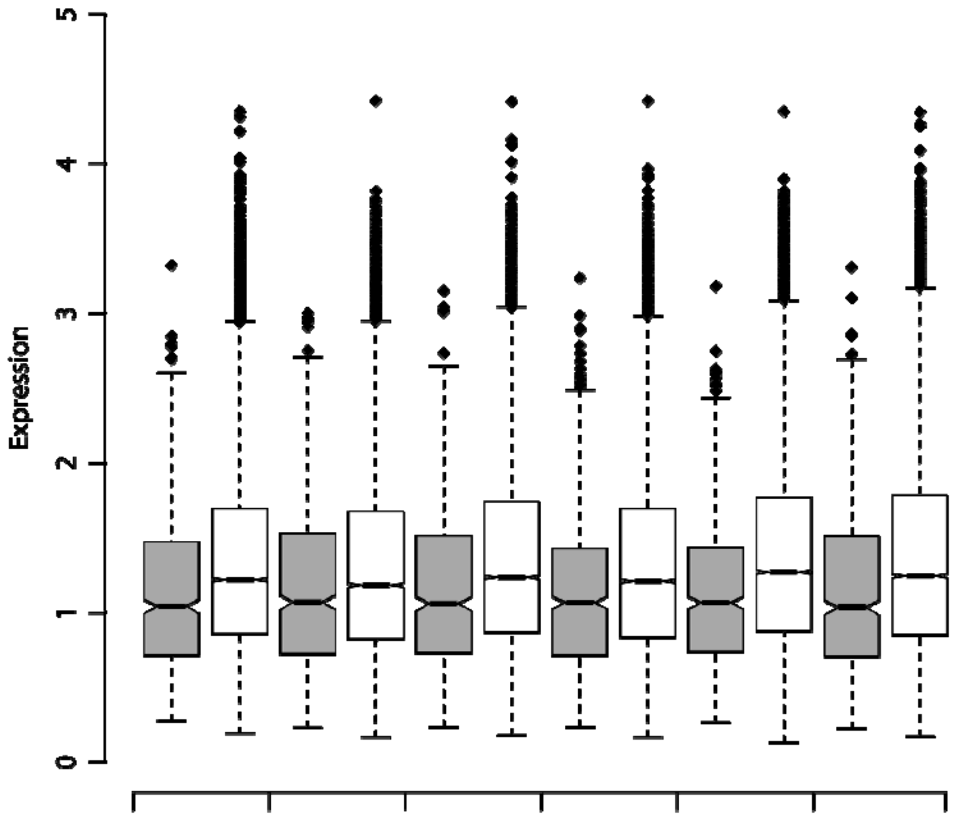
[f] Total: The total number of genes in the *T. pseudonana* genome annotated for a specific GO term

* Significant with Bonferroni corrections per Ontology (MF =1.4E-4; CC = 3.7E-4; BP = 9.6E-5)

† Significant with false discovery rate of 0.05 using q-value correction

Expression of positively selected genes Thalassiosira pseudonana (strain CCMP 1335)

Relative mRNA expression of *T. pseudonana* CCMP 1335 from a publicly available data set was analyzed with respect to the positively selected genes. *Thalassiosira pseudonana* had been limited for growth by silicic acid, iron, nitrate, carbon dioxide, or low temperature (4 °C) or maintained under nutrient replete conditions (Mock *et al.*, 2008). The dataset of expressed genes was comprised of 8996 genes of which 636 were in the positively selected set of 809. Positively selected genes are expressed at lower levels than those under neutral and purifying selection (Fig. II. 3). In each experimental treatment, the highest value of expression was an order of



	Ctrl	Si	Fe	NO_3	Temp	CO_2
W	2191502	2334484	2192834	2246692	2082194	2153474
p-value	7.704E-14*	1.542E-07*	9.026E-14*	3.778E-11*	2.2E-16*	6.963E-16*

Figure II. 3. Boxplots and statistics of the expression of positively selected (gray) and neutral and purified (white) genes of *Thalassiosira pseudonana* (CCMP 1335) grown in a nutrient replete control, nutrient limitation (Si, Fe, NO_3 and CO_2), and a 4 °C cold treatment. Boxplots: Box = 1st and 3rd quartiles, whiskers = 1.5 × IQR, notch = median of expression as $\log_{10}$(median aggregated probe intensity per gene). Mann-Whitney tests (N = 636 selected genes, N = 8357 neutral and purified genes; * significance at p << 0.01).

magnitude less for positively selected genes than genes evolving neutrally or under purifying selection (Fig. II. 3). In addition, the frequency distributions of gene expression for positively selected and neutral and purified genes were significantly different (Mann-Whitney; Fig. II. 3). Two-way hierarchical clustering of the expression of the 636 positively selected genes highlighted similarities and differences among groups of genes with respect to the different treatments (Fig. II. 4). Each gene is represented by a row in the heatmap of Fig. II. 4, and the dendrogram on the left groups genes together by how similar their expression patterns are across the treatments. The upper dendrogram groups treatments by the similarity of the patterns of expression for all genes. Two groups of genes are of special interest because they have very low and specific expression patterns, respectively. The first is a large group of genes that putatively encode proteins associated with sexual reproduction and had consistently low expression across all of the treatments (Fig. II. 4; Table II. S3). The second group contains five genes, four of which were differentially upregulated in the silicic acid and iron limited conditions relative to the control. One gene encodes a transcription factor, three appear to encode extracellular proteins including one with a chitin binding domain, and the fifth has a transmembrane helix and possibly binds lectins (Fig. II. 5).

Positive selection along specified lineages of Thalassiosira pseudonana

The different *T. pseudonana* strains were grouped according to similarities in environmental conditions at the locations the strains were collected to test hypotheses that those conditions promoted positive selection. Four strains were isolated from locations where sea surface temperature is relatively stable year-round, with a seasonal fluctuation up to 6 °C (Table II. 3). These strains have five-fold more genes under positive selection than strains isolated from regions where the sea surface temperature is more variable with seasonal fluctuations from 13 – 15 °C (Table II. 3). Forty-one of the 121 positively selected genes associated with stable sea surface temperatures had GO terms. Eighteen (44 %) of the 41 genes interacted with other proteins, GO:0005515, and were statistically enriched after Bonferroni correction (GOSTATS: *p*-value 0.00012). Gene ontology terms were assigned to nine of the 22 positively selected genes associated with variability in sea surfaces temperature; four of the encoded proteins had protein binding activity, but none were related to heat stress. The strain from the Pacific Gyre was the only representative from the open ocean, which is characterized by lower nutrient concentrations

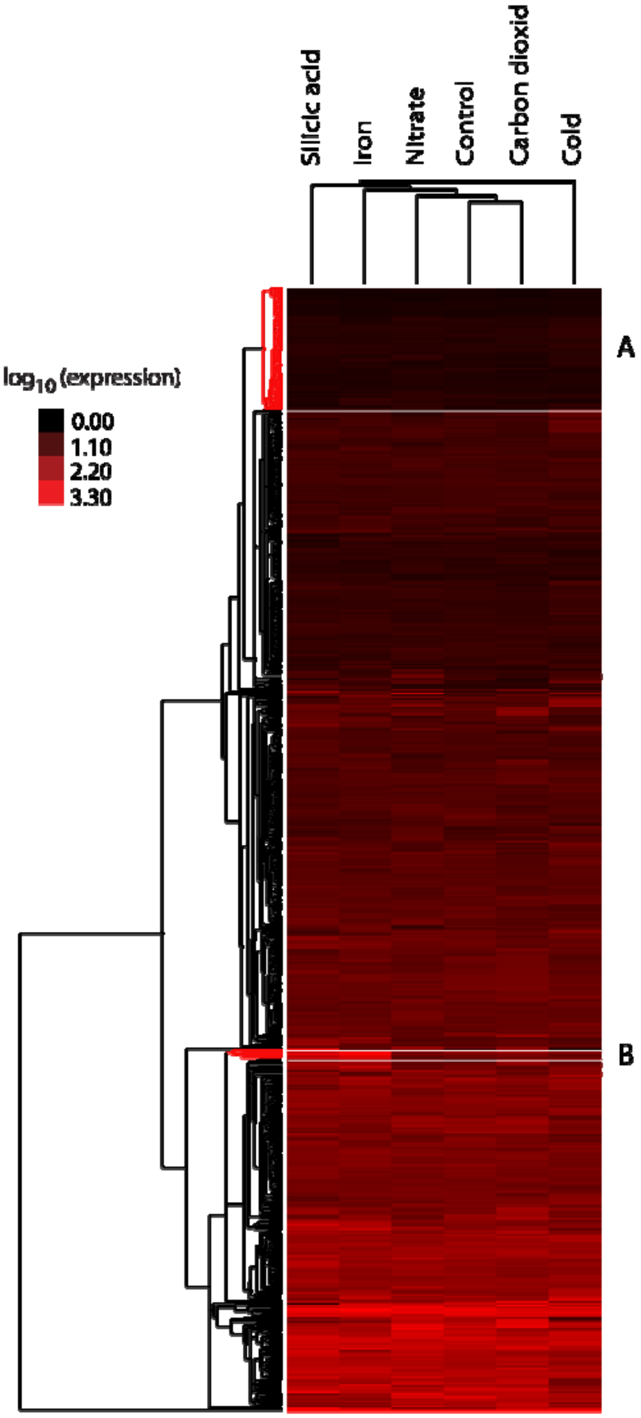
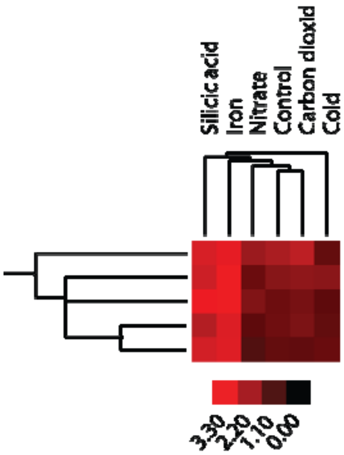


Figure II. 4. Relative expression for 636 of 809 positively selected genes in *Thalassiosira pseudonana* (CCMP 1335) grown under nutrient limitation (Si, Fe, NO_3 and CO_2), a nutrient replete control, and a 4 °C cold treatment. The similarity relationship of gene expression is represented by the dendrogram to the left. The upper dendrogram represents similarity of responses among treatments. A) Genes associated with sexual reproduction, B) Five putative cell wall-associated. Data is from Mock et al. 2008; Log$_{10}$(median aggregated probe intensity per gene) is plotted.



Prot ID	SP	Target	DE	Annotation
20959	-	-	-	TF_PP Tp_Myb1R4
12594	Y	S	Y	girdle band-associated, chitin binding, cell surface (p150) paralog
21085	-	-	Y	TM domain containing / Superfamily = lectin binding
21587	Y	S	Y	TM domain containing / conserved hypotheical protein
21795	Y	?	Y	V5 / Tpx-1 domain containing

Figure II. 5. Relative expression and annotations of five co-expressed cell wall-associated genes that are positively selected in *Thalassiosira pseudonana*. Table headers: SP = signal peptide, DE = differentially expressed relative to control in Mock et al. (2008). Within table: Y = yes, S = secretory pathway, TF = transcription factor; PP = protein-protein interacting, TM = transmembrane region. Scale is $\log_{10}$(median aggregated probe intensity per gene).

than coastal and estuarine areas. Two genes are positively selected in the Pacific Gyre strain, and neither was assigned a GO term.

The population of *T. pseudonana* from which the northern Adriatic Sea strain was collected is potentially geographically isolated from other populations because there are three hydrogeographic and genetic barriers for other species between it and the Atlantic Ocean (Patarnello *et al.*, 2007). Similar to the open ocean strain, no positively selected genes were detected within the Adriatic Sea strain (Table II. 3).

Influences of culturing were tested with a branch model to test for positive selection upon strains based on the decade in which they were isolated (Table II. 3). Time in culture does not appear to promote positive selection. There were 468 genes with a nominal *p*-value ≤ 0.05, meaning that estimated omegas of the specified branches were significantly different from one, but they could be significantly greater or less than one. Twenty-five of the 468 genes were significant at a Bonferroni corrected value ($p < 1.5\times10^{-5}$), but none had an omega ≥ 1.0.

Table II. 3. Number of genes under positive selection along specific lineages of *Thalassiosira pseudonana*.

					Tested Strains (shaded gray)				(by decade)
					Branch-site models				**Branch model**
Strain	**Isolation location**	**Lat**	**Long**	**ΔSST °C**	**Low ΔSST**	**High ΔSST**	**Ocean Environment**	**Adriatic isolation**	**Date of Isolation**
CCMP1014	North Pacific Gyre	28 N	155 W	0	Low	Low	Open Ocean	connected	1971
CCMP1015	San Juan Island, WA, USA	48.54 N	123.01 W	6	Low	Low	Coastal/Estuarine	connected	1985
CCMP1007	Virginia USA	37.95 N	79.94 W	14	Hi	Hi	Coastal/Estuarine	connected	1964
CCMP1335	New York USA	40.76 N	72.82 W	13	Hi	Hi	Coastal/Estuarine	connected	1958
CCMP1013	Wales, United Kingdom	53.28 N	3.83 W	6	Low	Low	Coastal/Estuarine	connected	1973
IT (RcTP)	Adriatic Sea, Italy	44.9 N	12.42 E	15	Hi	Hi	Coastal/Estuarine	isolated	2006
CCMP1012	Perth, Western Australia	31.99 S	115.83 E	2	Low	Low	Coastal/Estuarine	connected	1965
Results[†]									
genes with nominal p ≤ 0.05					744	353	87	70	468
significant genes (Bonferroni)					**121**	**22**	**2**	**0**	**0**
genes in positively selected set of 809					111	20	2	0	0

* ΔSST = difference in averaged sea surface temperature between January and July 1979 – 2000

[†] number of parameters for branch-site models: selection model = 16, null model = 17, df = 1
Branch model (date): selection model = 14, null model = 17, df = 3

Discussion

Phylogenetic distance at which positive selection is best detected in diatoms

The greatest number of positively selected genes was identified in the intra-species comparison of two *Thalassiosira pseudonana* strains. Therefore, an appropriate phylogenetic distance to detect the greatest signal of positive selection in diatoms appears to be intra-specific. *Thalassiosira pseudonana* diverged from *Detonula confervacea* ca. 2 mya, based on a molecular clock applied to the divergence of the DNA encoding the 18S subunit ribosomal RNA (Sorhannus, 2007). The 18S DNA sequences of *T. pseudonana* and *D. confervacea* differ by 0.4 % over 1758 base pairs (from accession numbers used in (Sorhannus, 2007)); therefore comparisons of diatoms with similar 18S sequence divergence should be tractable for identifying positively selected genes. This is not to say that positive selection cannot be detected at greater phylogenetic distances. Many sexual reproduction genes evolve rapidly, and those involved in gamete recognition are implicated in maintaining reproductive isolation (Lyon & Vacquier, 1999). The putative diatom gamete recognition gene, *SIG1*, is positively selected among four species of *Thalassiosira* (Sorhannus & Pond, 2006). *SIG1* is not positively selected within the seven strains of *T. pseudonana*, suggesting that these laboratory strains would still function within the reproductive species unit even though they were collected over distances of time and space. Selecting the appropriate phylogenetic distance to detect positively selected genes is, therefore, also contingent upon the functions of the proteins they encode.

The homologs that were detected between *Pseudo-nitzschia* species are currently subject to strong purifying selection, suggesting that their functions are conserved in other groups of organisms as well. Interestingly, the distribution of d_N:d_S in *Pseudo-nitzschia* homologs is of the same shape, and the rates of non-synonymous mutations per codon are in the same range as genes shared by mice and humans that also have homologs in plants and yeast (Alba & Castresana, 2005). These genes have deep evolutionary histories of hundreds of millions of years (Alba & Castresana, 2005). The apparent difference in the time required to accrue similar amounts of genetic variation, 10 versus 100s of million years, is likely due to the high frequency of asexual reproduction in diatoms providing an opportunity for mutation on a daily basis.

The cryptic sister species of *Ditylum* have a two-fold difference in genome size that is hypothesized to be the result of whole genome duplication (Koester et al., 2010). Positively selected genes would potentially be associated with speciation and niche differentiation reflecting the two species different adaptive strategies. The proportion of genes indicated to be under positive selection in *Ditylum* was intermediate to the *Pseudo-nitzschia* species and *Thalassiosira* strain pairs. This may be because the *Ditylum* species diverged only recently, and too little time has passed to identify selective substitutions. Increasing the length of alignments and the number of homologs compared between the *Ditylum* species may lead to the identification of more genes likely to be under positive selection. In addition to the divergence of alleles between the two species, positive selection may act on paralogs within the putative tetraploid (Han *et al.*, 2009), but this hypothesis remains to be tested.

Candidate genes under positive selection among seven strains of Thalassiosira pseudonana

Seven percent of the 11390 known genes in *Thalassiosira pseudonana* are strong candidates for positive selection. This estimate is statistically conservative, but it may include genes whose sequence variation is the product of unidentified gene duplications collapsed onto a single locus when sequenced reads are mapped to the reference genome. This is a challenging complication to consider: the signal of positive selection is accurate, but instead of applying to a single-locus alignment including sequence from seven strains, the alignment contains the consensus sequence of multiple loci for one or more strains. Therefore, the indication of positive selection may be based both on intra- and inter-strain divergence, instead of just inter-strain differences.

Positively selected genes are accrued at different rates across organisms, highlighting differences in selective pressures and mechanisms retaining beneficial mutations within populations. *Thalassiosira pseudonana* and humans both have ca. 7% of their loci under positive selection (Biswas & Akey, 2006), but the length of time that humans have been diverging from their last common ancestor is two to three times longer than that between *T. pseudonana* and *D. confervacea*. Two coral species *Acropora millepora* and *A. palmata* diverged from one another 10 – 12 mya (van Oppen *et al.*, 2001), and have a similar proportion of positively selected genes in their genomes (Voolstra *et al.*, 2011). Approximating sexual generation times for diatoms, acroporid corals, and humans as 2, 4, 20 yr respectively, these groups have accrued the same relative number of positively selected genes in 1.0×10^6 , 2.8×10^6, and 3.0×10^5 generations. Even though positively selected genes are the most rapidly evolving genes within a genome, many factors affect the fate of beneficial mutations within a population. The moderate proportion of positively selected genes within *T. pseudonana* might be due to large population sizes, which increase the time required for beneficial alleles to become fixed within a population. Cell counts for single diatom species can range as high as a million cells per liter during blooms (Lassiter *et al.*, 2006), although the effective population size is likely much smaller than that. Other factors affecting the fate of beneficial mutations include the amount of allelic variability within a population, gene flow with other populations and the interaction of these factors with the strength and periodicity of environmentally selective forces (Lynch *et al.*, 1991).

Versatility within the regulatory networks of gene expression provides a mechanism for cells to react quickly to environmental fluctuations (Li & Chen, 2010). Transcription factors are nodes within a network; therefore, positive selection in a single transcription factor potentially affects the expression of many proteins at once. Transcription factors comprise ca. two percent of the protein coding genes in *T. pseudonana* (Rayko *et al.*, 2010) and were the most specific group of over-represented proteins encoded by the positively selected genes (Table II. 2; Table II. S2). Ten percent of the 258 transcription factors are positively selected in *T. pseudonana* suggesting that differential gene regulation is an important component of adaptation for these strains. Positively selected genes were identified within the families of basic leucine zipper (bZip), heat shock, Myb, and zinc finger transcription factors. Transcription factors are also over-represented within the positively selected genes among human populations (Bustamante *et al.*, 2005), and between two species of the nematode *Caenorhabditis* (Castillo-Davis *et al.*, 2004), but not between two closely related corals (Voolstra *et al*., 2011). In mammals, families of zinc finger transcription factors are extensive, and positively selected genes have been identified within this family (Emerson & Thomas, 2009). In *T. pseudonana* only the TAZ group of zinc finger transcription factors is under positive selection; two of the six members of this family are selected. In mammals, TAZ factors co-regulate the response to hypoxic stress and are integrated into protein complexes of embryos that sense cell density and mediate embryo development (Freedman *et al.*, 2003, Varelas *et al.*, 2010).

Responding to the molecular signals of biotic and abiotic stressors is within the purview of both bZip and heat shock transcription factors; genes from both families also function under normal homeostatic physiologies. These two families are of special interest in diatoms because their relative sizes differ from those of other organisms. There are 10-fold fewer bZip transcription factors in diatoms and their phylogenetic group, the stramenopiles, than in plants and animals (Montsant *et al.*, 2007, Rayko *et al.*, 2010). Yet, 20 % of the genes in the bZip transcription family are under positive selection in *T. pseudonana*. Diatoms have an expanded family of heat shock transcription factors (HSFs), currently estimated to contain 94 members in contrast to the one to four HSFs found in yeast and metazoans (Montsant *et al.*, 2007, Rayko *et al.*, 2010, Fujimoto & Nakai, 2010). Six of the 39 (ca. 15 %) heat shock factors in a diatom-only clade, Group 2 HSFs, are positively selected (Fig. 1 of Rayko *et al.* 2010). The Group 2 clade of transcription factors highlights the importance of gene duplication in evolution and the potential for mutations in one of the duplicates to eventually provide a selective advantage (Zhang *et al.*, 1998, Briscoe *et al.*, 2010).

In *T. pseudonana*, 77 % of the proteins over-represented within any functional category are engaged in protein-protein interactions. The specific GO term over-represented in *T. pseudonana* is not among those over-represented among positively selected genes in other organisms, but genes encoding proteins with related functions including signal transducers, and receptors for various proteins are over-represented in animals (Bustamante et al., 2005, Voolstra et al., 2011). Adaptive amino acid substitutions in interacting proteins can alter the binding efficiency and function of the protein complex, thus providing the potential to optimize the interaction through co-evolution of binding partners (Presgraves & Stephan, 2007, Clark *et al.*, 2009). Interacting proteins are co-expressed; therefore the expression of the encoding mRNA may follow a similar pattern. No obvious examples of interacting proteins, such as heat shock transcription factors and heat shock proteins, were found within groups of co-expressed genes in *T. pseudonana* strain CCMP 1335. Genes encoding protein binding domains are expressed simultaneously but additional testing is required to determine if any of these proteins are actually binding partners, and consequently co-evolving.

Lineage-specific genes are found throughout the evolutionary tree of life (Tautz & Domazet-Lošo, 2011). These fast evolving genes experience relaxed levels of purifying selection; the increased numbers of non-synonymous mutations may lead to novel protein functions and adaptive advantages (Domazet-Lošo & Tautz, 2003, Voolstra *et al.*, 2011). Orphan genes of *T. pseudonana* were defined as those genes lacking any sequence-homology to the genes of other organisms at an *e*-value cut-off of 10^{-5}. There are proportionally more orphan genes in the positively selected set (24 %) than in the genome as a whole (15 %). These genes, encoding proteins of unknown function, are important to conferring an adaptive advantage to individuals of *T. pseudonana.* For example, the gene families of cyclins and heat shock factors have been greatly expanded in diatoms, and both families have genes found only in diatoms (Huysman *et al.*, 2010, Rayko *et al.*, 2010). A small number of genes are positively selected within the diatom-only cyclins and heat-shock factors. Two major mechanisms are hypothesized to create orphan or lineage-specific genes. First, one member of a duplicated pair may mutate faster than the other rendering it unrecognizable as a homolog (Domazet-Lošo & Tautz, 2003); alternatively, homologs along specific evolutionary branches may mutate quickly without duplication (Elhaik *et al.*, 2006). The second hypothesis promotes non-coding DNA to protein coding genes through spurious transcription of RNA (Cai *et al.*, 2009). The creation of orphan genes provides a long-term source of genetic variability upon which selection may act.

Expression of positively selected genes

Levels of gene expression in yeast and vertebrates are inversely correlated to protein divergence, and are hypothesized to exercise indirect control over mutation rates, such that highly expressed genes are the most conserved (Pál et al., 2001, Subramanian & Kumar, 2004). In addition to being expressed at lower levels than genes subject to neutral or purifying selection, positively selected genes tend to be expressed in restricted conditions, or specific tissues in multi-cellular organisms (Kosiol et al., 2008). Positively selected genes in *T. pseudonana* CCMP 1335 had significantly lower expression than more conserved genes across six different experimental treatments (Fig. II. 4). Similar to other organisms, the adaptive genes in *T. pseudonana* are likely functioning in specialized capacities and to enhance survival during sub-optimal growth conditions.

Sexual reproduction genes are a classic example of genes with restricted expression and rapid rates of evolution in diverse organisms (Clark *et al.*, 2006, Oliver *et al.*, 2010). *Thalassiosira pseudonana* belongs to the multipolar diatoms that produce flagellated sperm, as do the centric diatoms. Genes encoding flagella-associated and putative sperm activating proteins are under positive selection in *T. pseudonana*, and are among the genes with the lowest levels of expression among six experimental growth conditions (Fig. II. 4; Mock *et al.*, 2008). This pattern of expression for genes involved in sexual reproduction is consistent with observations that there were no cells differentiating into gametangia in these experiments (Mock et al. 2008). Sexual reproduction is not yet documented in *T. pseudonana*, but it is an obligate phase of the life history of many diatoms (Chepurnov *et al*., 2004). The signature of positive selection suggests that these genes are functional; otherwise, they would be expected to degrade into non-functional pseudogenes. Sexual reproduction in diatoms tends to be episodic, thus the genes would only be expressed periodically. For example, the putative gamete recognition gene *SIG 1* is positively selected among related species of *Thalassiosira* (Sorhannus & Pond, 2006), but it is expressed above background levels only after the initiation of gametogenesis in *T. weissfloggi* (Armbrust, 1999).

Co-expression of multiple genes suggests that their protein products are functioning in the same metabolic pathway or that they are affecting the same phenotype. The environmental conditions in which the genes are expressed provide additional information when the function of their encoded proteins is unknown. One of the most obvious patterns of co-expression in *T. pseudonana* CCMP 1335 was associated with five positively selected genes that were up-regulated when the cells were limited for silicon and iron (Fig. II. 5, Mock *et al*, 2008). The first gene in this group is a *MYB* transcription factor that also is highly expressed under other conditions, similar to expression patterns of *MYB* genes in plants where they respond to environmental stress and regulate growth and development (Chen *et al.*, 2006). Two genes encode proteins that appear to interact with chitin: protein 12594 possesses a chitin binding domain and a signal peptide suggesting that it is secreted; protein 21085 does not have a signal peptide, but is does have a transmembrane domain and it is a member of a superfamily of proteins that bind lectins, suggesting that it may also interact with chitin. The fourth member of this cluster encodes a protein (21587) with a transmembrane domain, but possesses no other functional annotation. The fifth gene contains a V5/Tpx-1 domain, commonly found in extracellular sensory proteins that recognize signals and proteins from other, frequently antagonistic and pathogenic organisms (Cantacessi *et al.*, 2009). Gene expression of this cluster is associated with a distinctive morphology induced by both silicon and iron limitation – an elongated, bent-cell phenotype in which chitin is deposited in the girdle region of the wall (Durkin *et al.*, 2009). Healthy cells have tight connections between silica cell wall components; therefore, secretion of chitin-binding proteins is hypothesized to shore up the silica cell wall when those connections are weakened (Davis *et al.*, 2005, Durkin *et al.*, 2009). The pattern of expression and putative localization of these five proteins provides a possible link between selective agents of nutrient limitation, functional genes and a cell wall phenotype.

Selection of specified strains with respect to environment

Branch-site models of positive selection were used to explore the hypothesized selective forces of geographic isolation, ocean environment and temperature, on subsets of the strains collected from sites associated with these environmental variables (Table II. 3). Three hydrogeographic barriers separate the Adriatic Sea from the Atlantic Ocean, and population structure is differentiated by these boundaries in other planktonic organisms living in the Mediterranean Sea (Patarnello *et al*., 2007, Yebra *et al.*, 2011). Although the Adriatic strain is potentially geographically isolated, there do not appear to be any differentiating environmental pressures. Instead, the genes of the Adriatic strain are evolving under purifying and neutral selection. Genes evolving neutrally have non-synonymous mutations that contribute to protein divergence, but their rates do not outpace those of silent mutations. Interestingly, the Adriatic strain and the strain from Wales share a branch on the coalescent tree, and do not have intermediate lineages separating them, suggesting that they are from the same population. However, manual review of numerous gene alignments indicates that the relationship between the two strains is an artifact of long-branch attraction grouping them together because they consistently had the most different protein sequences. Together, these results suggest that the Adriatic strain may be from a population subject to random genetic drift. Further population level experiments are required to test this hypothesis.

The open ocean strain from the Pacific Gyre was collected from an environment very different than the other six estuarine strains, including the one from the Adriatic. The gyre has stable light, salinity and low nutrients compared to coastal regions and estuaries, which are characterized by dynamic fluctuations of those same parameters (Karl, 1999). Therefore, it was hypothesized that the open ocean strain would have a strong genetic signal of differential adaptation to its unique environment. Surprisingly, only two genes were found to be positively selected in the open ocean strain. Both encode predicted proteins with unknown functions. Increasing the number of strains from other oceanic regions would increase the power to detect genes that are being selected for by this environment.

Temperature can be a strong selective agent (Husby *et al.*, 2011), and two hypothesis tests analyzed genes for positive selection associated with seasonal variation in sea-surface temperatures. The four strains collected from locations where summer-winter sea surface temperatures varied 6 °C, or less, shared 121 positively selected genes. The presence of three heat shock factors, one heat shock protein, and over-represented protein binding proteins within this set of positively selected genes suggests adaptation to stress conditions. Fewer positively selected genes are shared by the three strains collected from regions where there is a strong seasonal fluctuation in temperature, but one gene is notable because it encodes a UV radiation resistance protein. This protein may not be directly affected by temperature, but it would be affected by available sunlight altering the temperature of water. These data provide circumstantial evidence that seasonal temperature variation may act as selective agents; however, it is more likely that absolute temperature, the timing of seasonal changes and interactions between available light and temperature are more effective in selecting fit phenotypes (*e.g.* Namroud *et al.,* 2008).

Branch-site experiments are usually performed to identify positive selection associated with speciation events in macroscopic, non-planktonic organisms, where differentiating environmental niches are readily inferred (Briscoe et al., 2010, Shen *et al.*, 2010). Although the results from our branch-sites experiments are difficult to interpret without ambiguity, these data are robust in that the majority of genes found to be positively selected among subsets of strains were identified in the original set of 809 positively selected genes. Therefore, the positively selected genes identified by the branch-site models remain strong candidates for further physiological experiments in the laboratory to test the fitness of the alleles carried by each strain (Yang & dos Reis, 2011).

Conclusion

These were the first transcriptome and genome scans performed to identify positively selected genes within a phytoplankton group. Intra-species comparisons are appropriate for detecting the greatest number of genes evolving under positive selection, but inter-species analyses are still important for genes of certain functions. As more genomes become available for diatom species with high ecological and human impact, analyzing multiple strains of those species will be important to understanding their survival strategies and the conditions under which they bloom. Seven percent of the protein coding genes in *Thalassiosira pseudonana* are strong candidates for positive selection. Selection for transcription factors suggests that the adaptive response cascades and affects many more genes than just those that are selected by the fluctuating conditions in which diatoms live. Five positively selected genes may be linked to a cell wall phenotype facilitating survival in stressful conditions, specifically shortages of the obligate nutrients of silicon and iron, both of which are in growth limiting concentrations in large regions of the world's oceans. Specific alleles of each of these genes are functioning in individual diatoms to confer a selective advantage; however, to understand which genetic variant is connected to which phenotype and selective pressure more directed research is required. These results present an opportunity to test hypotheses that integrate positively selected genes in *T. pseudonana* with their associated phenotypes and selective forces through manipulative laboratory experiments and in the field by taking a population genetics approach to determine allele distributions for populations living in different regions.

Author contributions

JAK, WJS and EVA designed the experiments. JAK cultured *Ditylum* and constructed and curated the data set. JAK performed in *silico* experiments and analyzed the data. JAK, WJS and EVA wrote the paper.

Acknowledgments

We would like to thank Renee George, Megan Kogut, Bob Morris and Larry Ruzzo for insightful discussion and Robin Kodner for assistance with the *T. pseudonana* consensus tree. Special thanks are given to those who shared data sets. The *Thalassiosira pseudonana* data set was constructed and curated by Rhonda Morales, Ellen Lin, Vaughn Iverson, Dave Schruth and Chris Berthiaume. The *Pseudo-nitzschia multiseries* and *P. australis* data sets are courtesy of Micaela S. Parker, and Uwe John and Wiebe Kooistra provided the *Pseudo-nitzschia multistriata* data set.

CONCLUSION

The rapid diversification and success of diatoms were facilitated by genetic diversity within populations that resulted in colonization and adaption to new environments. Partial and whole-genome duplications are just two of many ways to introduce genetic diversification into populations. The ability to form stable polyploids occurs in multiple diatom genera and appears to be the mechanism of speciation in *Ditylum brightwellii* found in the northeastern Pacific Ocean. Mutations accrued in one of the copies of the duplicate genes may lead to functional innovations in the encoded proteins, thus promoting adaptation to different environmental conditions. The putative tetraploid *D. brightwellii* population 2 is localized to the northeastern Pacific Ocean and blooms during late spring and early summer when silicate concentrations are less than during the peak bloom period of population 1 (Rynearson *et al*. 2006), suggesting that it has adapted to those lower concentrations. This is a testable hypothesis that may be explored through physiological experiments and genetic analysis on several strains from each species. The hypothesis that silicate concentrations drive selection could be tested with kinetic uptake experiments of silicate and competition experiments between the species at varying silicate concentrations. In addition, increasing the sample size of the number of genes compared may point to genes affected by positive selection and environmental pressures other than silicon depending on encoded protein functions.

Diatoms have the potential to be a model system to study polyploidization. Several genera are suspected of harboring polyploid species, but have yet to be tested (Rines, pers. comm.). Two genera with putative polyploid species, *Thalassiosira* and *Ditylum*, have species that occur frequently in environmental samples and are easily brought into culture. Interactions between the environment and polyploids as well as the cellular dynamics of whole-genome duplication may be tested. For example, biogeographic mapping may be used to determine if there is a correlation between ploidy and range; the susceptibility of a species to form polyploids repeatedly and/or successively could be tested with laboratory manipulations. The genes and genomes themselves present countless hypotheses regarding the divergence of protein function between the species and its association with the environment.

In the absence of polyploidization, genetic diversity is maintained in diatom populations through mutations likely accrued during asexual reproduction. The physiological difference in the growth rates of *Ditylum brightwellii* population 1 strains from New Zealand and Puget Sound is likely due to the selection for different alleles that promote their local adaptation. The collection and characterization of *D. brightwellii* using a neutral genetic marker, genome size analysis, and physiology broadened our detailed understanding of this cosmopolitan species.

Genes linked directly to adaptation were intensively investigated in *Thalassiosira pseudonana* by analyzing every protein coding gene in the genome for positive selection using seven different strains. Seven percent of *T. pseudonana's* genes are statistically strong candidates for positive selection. The results from this study present many new hypotheses that may be tested in *T. pseudonana*. Population genetics studies can be used to assess the distribution of alleles within populations from different regions to understand what environmental variables are associated with them. Manipulative laboratory experiments can be used to pinpoint selective pressures, and *T. pseudonana* can be engineered to fluorescently tag the most interesting proteins whose functions are unknown to better link genotypes to phenotypes. A putative link between selective pressures, a phenotype, and positively selected genes was identified by combining the results from the analysis for positive selection with those from a previously published experiment measuring gene expression of one *T. pseudonana* strain grown under six environmentally relevant conditions. Combined analyses such as this one are powerful, but will be more influential once similar data sets become available for more species with high ecological and human impacts.

References

Adl, S. M., Simpson, A. G. B., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., James, T. Y., Karpov, S., Kugrens, P., Krug, J., Lane, C. E., Lewis, L. A., Lodge, J., Lynn, D. H., Mann, D. G., McCourt, R. M., Mendoza, L., Moestrup, O., Mozley-Standridge, S. E., Nerad, T. A., Shearer, C. A., Smirnov, A. V., Spiegel, F. W. & Taylor, M. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **52**:399-451.

Alba, M. M. & Castresana, J. 2005. Inverse Relationship Between Evolutionary Rate and Age of Mammalian Genes. *Mol. Biol. Evol.* **22**:598-606.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389-402.

Altshuler, D., Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z. M., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N. S., Zody, M. C., Linton, L., Lander, E. S. & Int, S. N. P. M. W. G. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**:928-33.

Alverson, A. J. 2007. Strong purifying selection in the silicon transporters of marine and freshwater diatoms. *Limnol. Oceanogr.* **52**:1420-29.

Alverson, A. J., Beszteri, B., Julius, M. L. & Theriot, E. C. 2011. The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*. *BMC Evol. Biol.* **11**.

Amato, A. & Montresor, M. 2008. Morphology, phylogeny, and sexual cycle of *Pseudo-nitzschia mannii* sp. nov. (Bacillariophyceae): a pseudo-cryptic species within the *P. pseudodelicatissima* complex. *Phycologia* **47**:487-97.

Amato, A., Orsini, L., D'Alelio, D. & Montresor, M. 2005. Life cycle, size reduction patterns, and ultrastructure of the pennate planktonic diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae). *J. Phycol.* **41**:542-56.

Armbrust, E. V. 1999. Identification of a new gene family expressed during the onset of sexual reproduction in the centric diatom *Thalassiosira weissflogii*. *Appl. Environ. Microbiol.* **65**:3121-28.

Armbrust, E. V. 2009. The life of diatoms in the world's oceans. *Nature* **459**:185-92.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V., Kroeger, N., Lau, W. W. Y., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Rynearson, T. A., Saito, M. A., Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P. & Rokhsar, D. S. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**:79-86.

Arrhenius, G. 1952. Sediment cores of the East Pacific. *Reports of the Swedish Deep Sea Expeditions 1947-1948* Stockholm : Swedish Natural Science Research Council, pp. 1-91.

Austin, J. A. & Barth, J. A. 2002. Variation in the position of the upwelling front on the Oregon shelf. *Journal of Geophysical Research-Oceans* **107**.

Beaton, M. J. & Cavalier-Smith, T. 1999. Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes. *Proc. R. Soc. Lond. B. Biol. Sci.* **266**:2053-59.

Benitez-Nelson, C. R., Bidigare, R. R., Dickey, T. D., Landry, M. R., Leonard, C. L., Brown, S. L., Nencioli, F., Rii, Y. M., Maiti, K., Becker, J. W., Bibby, T. S., Black, W., Cai, W. J., Carlson, C. A., Chen, F. Z., Kuwahara, V. S., Mahaffey, C., McAndrew, P. M., Quay, P. D., Rappe, M. S., Selph, K. E., Simmons, M. P. & Yang, E. J. 2007. Mesoscale eddies drive increased silica export in the subtropical Pacific Ocean. *Science* **316**:1017-21.

Beszteri, B., Acs, E. & Medlin, L. K. 2005. Ribosomal DNA sequence variation among sympatric strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) reveals cryptic diversity. *Protist* **156**:317-33.

Beszteri, B., John, U. & Medlin, L., K. 2007. An assessment of cryptic genetic diversity within the *Cyclotella meneghiniana* species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphisms. *Eur. J. Phycol.* **42**:47-60.

Bigham, A., Bauchet, M., Pinto, D., Mao, X. Y., Akey, J. M., Mei, R., Scherer, S. W., Julian, C. G., Wilson, M. J., Herraez, D. L., Brutsaert, T., Parra, E. J., Moore, L. G. & Shriver, M. D. Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. *PLoS Genet.* **6**.

Biswas, S. & Akey, J. M. 2006. Genomic insights into positive selection. *Trends Genet.* **22**:437-46.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J. P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kroger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jezequel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M. P., Napoli, C., Obornik, M., Parker, M. S., Petit, J. L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van De Peer, Y. & Grigoriev, I. V. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**:239-44.

Boyd, C. M. & Gradmann, D. 2002. Impact of osmolytes on buoyancy of marine phytoplankton. *Mar. Biol.* **141**:605-18.

Brand, L. E. 1984. The salinity tolerance of 46 marine phytoplankton isolates. *Estuarine Coastal and Shelf Science* **18**:543-56.

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

Briscoe, A. D., Bybee, S. M., Bernard, G. D., Yuan, F., Sison-Mangus, M. P., Reed, R. D., Warren, A. D., Llorente-Bousquets, J. & Chiao, C.-C. 2010. Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. *Proc. Natl. Acad. Sci. U.S.A.* **107**:3628-33.

Burckle, L. H. & McLaughlin, R. B. 1977. Size changes in the marine diatom *Coscinodiscus nodulifer* A. Schmidt in the equatorial Pacific. *Micropaleontology (New York)* **23**:216-22.

Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Todd Hubisz, M., Glanowski, S., Tanenbaum, D. M., White, T. J., Sninsky, J. J., Hernandez, R. D., Civello, D., Adams, M. D., Cargill, M. & Clark, A. G. 2005. Natural selection on protein-coding genes in the human genome. *Nature* **437**:1153-57.

Cai, J. J., Macpherson, J. M., Sella, G. & Petrov, D. A. 2009. Pervasive hitchhiking at coding and regulatory sites in humans. *PLoS Genet.* **5**.

Cantacessi, C., Campbell, B. E., Visser, A., Geldhof, P., Nolan, M. J., Nisbet, A. J., Matthews, J. B., Loukas, A., Hofmann, A., Otranto, D., Sternberg, P. W. & Gasser, R. B. 2009. A portrait of the "SCP/TAPS" proteins of eukaryotes - Developing a framework for fundamental research and biotechnological outcomes. *Biotechnol. Adv.* **27**:376-88.

Caputi, L., Andreakis, N., Mastrototaro, F., Cirino, P., Vassillo, M. & Sordino, P. 2007. Cryptic speciation in a model invertebrate chordate. *Proc. Natl. Acad. Sci. U.S.A.* **104**:9364-69.

Casteleyn, G., Leliaert, F., Backeljau, T., Debeer, A.-E., Kotaki, Y., Rhodes, L., Lundholm, N., Sabbe, K. & Vyverman, W. 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proceedings of the National Academy of Sciences* **107**:12952-57.

Castillo-Davis, C. I., Kondrashov, F. A., Hartl, D. L. & Kulathinal, R. J. 2004. The functional genomic distribution of protein divergence in two animal phyla: Coevolution, genomic conflict, and constraint. *Genome Res.* **14**:802-11.

Cavalier-Smith, T. 1985. *The Evolution of genome size.* J. Wiley, Chichester [West Sussex], New York,

Cavalier-Smith, T. 2005. Economy, speed and size matter: Evolutionary forces driving nuclear genome miniaturization and expansion. *Ann. Bot.* **95**.

Cermeno, P. & Falkowski, P. G. 2009. Controls on diatom biogeography in the ocean. *Science* **325**:1539-41.

Chen, Y. H., Yang, X. Y., He, K., Liu, M. H., Li, J. G., Gao, Z. F., Lin, Z. Q., Zhang, Y. F., Wang, X. X., Qiu, X. M., Shen, Y. P., Zhang, L., Deng, X. H., Luo, J. C., Deng, X. W., Chen, Z. L., Gu, H. Y. & Qu, L. J. 2006. The MYB transcription factor superfamily of *Arabidopsis*: Expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol. Biol.* **60**:107-24.

Chepurnov, V. A. & Mann, D. G. 2003. Auxosporulation of *Licmophora communis* (Bacillariophyta) and a review of mating systems and sexual reproduction in araphid pennate diatoms. *Phycol. Res.* **52**:1-12.

Chepurnov, V. A., Mann, D. G., Sabbe, K. & Vyverman, W. 2004. Experimental studies on sexual reproduction in diatoms. *Int. Rev. Cytol.* **237**:91-154.

Chepurnov, V. A., Mann, D. G., Vyverman, W., Sabbe, K. & Danielidis, D. B. 2002. Sexual reproduction, mating system, and protoplast dynamics of *Seminavis* (Bacillariophyceae). *J. Phycol.* **38**:1004-19.

Chepurnov, V. A. & Roshchin, A. M. 1995. Inbreeding influence on sexual reproduction of *Achnanthes longipes* Ag. (Bacillariophyta). *Diatom Res.* **10**:21-25.

Chisholm, S. W. 1992. Phytoplankton size. *In*: Falkowski, P. G. & Woodhead, A. D. [Eds.] *Primary Productivity and biogeochemical cycles in the sea: Environmental science research.* Plenum Press, New York, pp. 213-37.

Chisholm, S. W., Olson, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B. & Welschmeyer, N. A. 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**:340-43.

Clark, N. L., Aagaard, J. E. & Swanson, W. J. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* **131**:11-22.

Clark, N. L., Gasper, J., Sekino, M., Springer, S. A., Aquadro, C. F. & Swanson, W. J. 2009. Coevolution of Interacting Fertilization Proteins. *PLoS Genet.* **5**.

Conley, D. J., Kilham, S. S. & Theriot, E. 1989. Differences in silica content between marine and fresh-water diatoms. *Limnol. Oceanogr.* **34**:205-13.

Créach, V., Ernst, A., Sabbe, K., Vanelslander, B., Vyverman, W. & Stal, L. J. 2006. Using quantitative PCR to determine the distribution of a semicryptic benthic diatom, *Navicula phyllepta* (Bacillariophyceae). *J. Phycol.* **42**:1142-54.

D'Alelio, D., d'Alcala, M. R., Dubroca, L., Sarno, D., Zingone, A. & Montresor, M. 2010. The time for sex: A biennial life cycle in a marine planktonic diatom. *Limnol. Oceanogr.* **55**:106-14.

Davis, A. K., Hildebrand, M. & Palenik, B. 2005. A stress-induced protein associated with the girdle band region of the diatom *Thalassiosira pseudonana* (Bacillariophyta). *J. Phycol.* **41**:577-89.

Descolas-gros, C. & Debilly, G. 1987. Temperature adaptation of RuBP carboxylase: kinetic properties in marine Antarctic diatoms. *J. Exp. Mar. Biol. Ecol.* **108**:147-58.

Domazet-Lošo, T. & Tautz, D. 2003. An evolutionary analysis of orphan genes in Drosophila. *Genome Res.* **13**:2213-19.

Durkin, C. A., Mock, T. & Armbrust, E. V. 2009. Chitin in diatoms and its association with the cell wall. *Eukaryot. Cell* **8**:1038-50.

Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences* **95**:14863-68.

Elhaik, E., Sabath, N. & Graur, D. 2006. The "inverse relationship between evolutionary rate and age of mammalian genes" is an artifact of increased genetic distance with rate of evolution and time of divergence. *Mol. Biol. Evol.* **23**:1-3.

Emerson, R. O. & Thomas, J. H. 2009. Adaptive evolution in zinc finger transcription factors. *PLoS Genet.* **5**.

Eppley, R. W., Holmes, R. W. & Paasche, E. 1967. Periodicity in cell division and physiological behavior of *Ditylum brightwellii*, a marine planktonic diatom, during growth in light-dark cycles. *Arch Mikrobiol* **56**:305-23.

Falcon, S. & Gentleman, R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**:257-58.

Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O. & Taylor, F. J. R. 2004. The evolution of modern eukaryotic phytoplankton. *Science* **305**:354-60.

Fenchel, T. 2005. Cosmopolitan microbes and their 'cryptic' species. *Aquat. Microb. Ecol.* **41**:49-54.

Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**:237-40.

Finkel, Z. V., Katz, M. E., Wright, J. D., Schofield, O. M. E. & Falkowski, P. G. 2005. Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic. *Proceedings of the National Academy of Sciences* **102**:8927-32.

Finlay, Bland J., Monaghan, Elaine B. & Maberly, Stephen C. 2002. Hypothesis: The rate and scale of dispersal of freshwater diatom species is a function of their global abundance. *Protist* **153**:261-73.

Fischer, M. C., Foll, M., Excoffier, L. & Heckel, G. 2011. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Mol. Ecol.* **20**:1450-62.

Fisher, N. S. 1977. On the differential sensitivity of estuarine and open-ocean diatoms to exotic chemical stress. *Am. Nat.* **111**:871-95.

Freedman, S. J., Sun, Z. Y. J., Kung, A. L., France, D. S., Wagner, G. & Eck, M. J. 2003. Structural basis for negative regulation of hypoxia-inducible factor-1 alpha by CITED2. *Nat. Struct. Biol.* **10**:504-12.

Fujimoto, M. & Nakai, A. 2010. The heat shock factor family and adaptation to proteotoxic stress. *FEBS J.* **277**:4112-25.

Furnas, M. J. 1990. In situ growth rates of marine phytoplankton: approaches to measurement, community and species growth rates. *J. Plankton Res.* **12**:1117-51.

Gersonde, R. & Harwood, D. M. 1990. Lower Cretaceous diatoms from ODP Leg 113 Site 693 (Weddell Sea), Part 1: Vegetative cells. *In*: P.F., B., J.P., K., S., O. C., S., B., W.R., B., L.H., B., P.K., E., D.K., F., R.E., G., X., G., N., H., L., L., D.B., L., M., L., B., M., T., N., C.P.Q., P., C.J., P., C.M., R., E., S., V., S., L.D., S., E., T., K.F.M., T. & S.W.Jr., W. [Eds.] *Proceedings of the Ocean Drilling Program, Scientific Results.* Ocean Drilling Program, College Station, TX, pp. 365-402.

Giri, B. S. 1991. Karyology of the genus *Cyclotella* Kütz (Bacillariophyceae). *Cytologia* **56**:494-94.

Giri, B. S. 1992. Nuclear cytology of naviculoid diatoms. *Cytologia* **57**:173-79.

Giri, B. S., Chowdary, Y. B. K. & Sarma, Y. S. R. K. 1990. Cytological studies on some pennate diatoms. *The Nucleus* **33**:141-44.

Graham, L. E. & Wilcox, L. W. 2000. *Algae.* Prentice Hall, Upper Saddler River, NJ,

Gregory, T. R. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews* **76**:65-101.

Gross, F. 1937. The life history of some marine plankton diatoms. *Philos Trans R Soc Lond B Biol Sci* **228**:1-47.

Guillard, R. R. L. & Ryther, J. H. 1962. Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Can. J. Microbiol.* **8**:229-39.

Han, M. V., Demuth, J. P., McGrath, C. L., Casola, C. & Hahn, M. W. 2009. Adaptive evolution of young gene duplicates in mammals. *Genome Res.* **19**:859-67.

Harrison, K. G. 2000. Role of increased marine silica input on paleo-$p\mathrm{CO}_2$ levels. *Paleoceanography* **15**:292-98.

Holm-Hansen, O. 1969. Algae: amounts of DNA and organic carbon in single cells. *Science* **163**:87-88.

Holtermann, K. E., Bates, S. S., Trainer, V. L., Odell, A. & Virginia Armbrust, E. 2010. Mass sexual reproduction in the toxigenic diatoms *Pseudo-nitzschia australis* and *P. pungens* (Bacillariophyceae) on the Washington Coast, USA. *J. Phycol.* **46**:41-52.

Husband, B. C. & Sabara, H. A. 2003. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytol.* **161**:703-13.

Husby, A., Visser, M. E. & Kruuk, L. E. B. 2011. Speeding up microevolution: The effects of increasing temperature on selection and genetic variance in a wild bird population. *PLoS Biol.* **9**.

Huysman, M. J. J., Martens, C., Vandepoele, K., Gillard, J., Rayko, E., Heijde, M., Bowler, C., Inze, D., Van de Peer, Y., De Veylder, L. & Vyverman, W. 2010. Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biology* **11**.

Iverson, V., Morris, R. M., Frazar, C. D., Berthiaume, C. T., Morales, R. L. & Armbrust, E. V. 2012. Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine Euryarchaeota. *Science* **335**:587-90.

Jensen, A., Rystad, B. & Melsom, S. 1974. Heavy metal tolerance of marine phytoplankton. I. Tolerance of 3 algal species to zinc in coastal seawater. *J. Exp. Mar. Biol. Ecol.* **15**:145-57.

Julenius, K. & Pedersen, A. G. 2006. Protein evolution is faster outside the cell. *Mol. Biol. Evol.* **23**:2039-48.

Karl, D. M. 1999. A sea of change: Biogeochemical variability in the North Pacific Subtropical Gyre. *Ecosystems* **2**:181-214.

Karl, D. M., Bidigare, R. R. & Letelier, R. M. 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**:1449-70.

Kim, P. M., Korbel, J. O. & Gerstein, M. B. 2007. Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. *Proceedings of the National Academy of Sciences* **104**:20274-79.

Kociolek, J. P. & Stoermer, E. F. 1989. Chromosome numbers in diatoms: A review. *Diatom Res.* **4**:47-54.

Koester, J. A., Brawley, S. H., Karp-Boss, L. & Mann, D. G. 2007. Sexual reproduction in the marine centric diatom *Ditylum brightwellii* (Bacillariophyta). *Eur. J. Phycol.* **42**:351-66.

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

Kooistra, W. H. C. F., Gersonde, R., Medlin, L. K. & Mann, D. G. 2007. The origin and evolution of the diatoms: Their adaptation to a planktonic existence. *In*: Falkowski, P. G. & Knoll, A. H. [Eds.] *Evolution of primary producers in the sea.* Elsevier Academic Press, Amsterdam, Boston, pp. 210-50.

Kosiol, C., Vinar, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R. & Siepel, A. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet.* **4**.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947-48.

Lassiter, A. M., Wilkerson, F. P., Dugdale, R. C. & Hogue, V. E. 2006. Phytoplankton assemblages in the CoOP-WEST coastal upwelling area. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **53**:3063-77.

Lavaud, J., Strzepek, R. F. & Kroth, P. G. 2007. Photoprotection capacity differs among diatoms: Possible consequences on the spatial distribution of diatoms related to fluctuations in the underwater light climate. *Limnol. Oceanogr.* **52**:1188-94.

Li, C. W. & Chen, B. S. 2010. Identifying functional mechanisms of gene and protein regulatory networks in response to a broader range of environmental stresses. *Comp. Funct. Genomics*.

Li, Y. D., Liang, H., Gu, Z., Lin, Z., Guan, W., Zhou, L., Li, Y. Q. & Li, W. H. 2009. Detecting positive selection in the budding yeast genome. *J. Evol. Biol.* **22**:2430-37.

Litchman, E., Klausmeier, C. A. & Yoshiyama, K. 2009. Contrasting size evolution in marine and freshwater diatoms. *Proc. Natl. Acad. Sci. U.S.A.* **106**:2665-70.

Lynch, M., Gabriel, W. & Wood, A. M. 1991. Adaptive and demographic responses of phytoplankton populations to environmental change. *Limnol. Oceanogr.* **36**:1301-12.

Lyon, J. D. & Vacquier, V. D. 1999. Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. *Dev. Biol.* **214**:151-59.

MacDonald, J. D. 1869. On the structure of the diatomaceous frustule, and its genetic cycle. *Mag. Nat. Hist.* **4**:1-8.

Maliva, R. G., Knoll, A. H. & Siever, R. 1989. Secular change in chert distribution: A reflection of evolving biological participation in the silica cycle. *Palaios* **4**:519-32.

Mann, David G. 1999. The species concept in diatoms. *Phycologia* **38**:437-95.

Mann, D. G. 1994. Auxospore formation, reproductive plasticity and cell structure in *Navicula ulvacea* and the resurrection of the genus *Dickieia* (Bacillariophyta). *Eur. J. Phycol.* **29**:141-57.

Mann, D. G. & Droop, S. J. M. 1996. Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* **336**:19-32.

Mann, D. G. & Stickle, A. J. 1991. The genus *Craticula*. *Diatom Res.* **6**:79-107.

Mayr, E. 1969. The biological meaning of species. *Biol. J. Linn. Soc.* **1**:311-20.

Medlin, L. K. 2010. Pursuit of a natural classification of diatoms: An incorrect comparison of published data. *Eur. J. Phycol.* **45**:155-66.

Medlin, L. K., Elwood, H. J., Stickel, S. & Sogin, M. L. 1991. Morphological and genetic variation within the diatom *Skeletonema costatum* (Bacillariophyta): Evidence for a new species, *Skeletonema pseudocostatum*. *J. Phycol.* **27**:514-24.

Medlin, L. K. & Kaczmarska, I. 2004. Evolution of the diatoms: V. Morphological and cytological support for the major clades and taxonomic revision. *Phycologia* **43**:245-70.

Miller, J. R., Delcher, A. L., Koren, S., Venter, E., Walenz, B. P., Brownley, A., Johnson, J., Li, K., Mobarry, C. & Sutton, G. 2008. Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* **24**:2818-24.

Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., Durkin, C., BonDurant, S., Splinter, Richmond, K., Rodesch, M., Kallas, T., Huttlin, E., L., Cerrina, F., Sussman, M., R. & Armbrust, E., Virginia 2008. Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of Sciences*:Early Edition.

Montresor, M., Sgrosso, S., Procaccini, G. & Kooistra, W. H. C. F. 2003. Intraspecific diversity in *Scrippsiella trochoidea* (Dinophyceae): evidence for cryptic species. *Phycologia* **42**:56-70.

Montsant, A., Allen, A. E., Coesel, S., De Martino, A., Falciatore, A., Mangogna, M., Siaut, M., Heijde, M., Jabbari, K., Maheswari, U., Rayko, E., Vardi, A., Apt, K. E., Berges, J. A., Chiovitti, A., Davis, A. K., Thamatrakoln, K., Hadi, M. Z., Lane, T. W., Lippmeier, J. C., Martinez, D., Parker, M. S., Pazour, G. J., Saito, M. A., Rokhsar, D. S., Armbrust, E. V. & Bowler, C. 2007. Identification and comparative genomic analysis of signaling and regulatory components in the diatom *Thalassiosira pseudonana*. *J. Phycol.* **43**:585-604.

Moore, J. K. & Villareal, T. A. 1996. Size-ascent rate relationships in positively buoyant marine diatoms. *Limnol. Oceanogr.* **41**:1514-20.

Namroud, M.-C., Beaulieu, J., Juge, N., Laroche, J. & Bousquet, J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* **17**:3599-613.

Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A. & Quéguiner, B. 1995. Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochem Cycles* **9**:359-72.

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., White, T. J., Sninsky, J. J., Adams, M. D. & Cargill, M. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**:976-85.

Ohno, S. 1970. Evolution by Gene Duplication. Springer-Verlag, Berlin, New York, pp. 160.

Oliver, M. J., Petrov, D., Ackerly, D., Falkowski, P. & Schofield, O. M. 2007. The mode and tempo of genome size evolution in eukaryotes. *Genome Res.* **17**:594-601.

Oliver, T. A., Garfield, D. A., Manier, M. K., Haygood, R., Wray, G. A. & Palumbi, S. R. 2010. Whole-genome positive selection and habitat-driven evolution in a shallow and a deep-Sea urchin. *Genome Biology and Evolution* **2**:800-14.

Otto, S. P. 2007. The evolutionary consequences of polyploidy. *Cell* **131**:452-62.

Otto, S. P. & Whitton, J. 2000. Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**:401-37.

Paasche, E. 1973. The influence of cell size on growth rate, silica content and some other properties of four marine diatoms species. *Norwegian Journal of Botany* **20**:197-204.

Pál, C., Papp, B. & Hurst, L. D. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* **158**:927-31.

Patarnello, T., Volckaert, F. A. M. J. & Castilho, R. 2007. Pillars of Hercules: is the Atlantic–Mediterranean transition a phylogeographical break? *Mol. Ecol.* **16**:4426-44.

Pfitzer, E. 1871. Untersuchungen über Bau und Entwickelung der Bacillariaceen (Diatomaceen). *Bot. Abhandl. (ed. Hanstein)* **1**:1-189.

Presgraves, D. C. & Stephan, W. 2007. Pervasive adaptive evolution among interactors of the *Drosophila* hybrid inviability gene, Nup96. *Mol. Biol. Evol.* **24**:306-14.

Quigg, A., Finkel, Z. V., Irwin, A. J., Rosenthal, Y., Ho, T. Y., Reinfelder, J. R., Schofield, O., Morel, F. M. M. & Falkowski, P. G. 2003. The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. *Nature* **425**:291-94.

Quijano-Scheggia, S. I., Garces, E., Lundholm, N., Moestrup, O., Andree, K. & Campi, J. 2009. Morphology, physiology, molecular phylogeny and sexual compatibility of the cryptic

Pseudo-nitzschia delicatissima complex (Bacillariophyta), including the description of *P. arenysensis* sp. nov. *Phycologia* **48**:492-509.

Raup, D. M. 1979. Size of the Permo-Triassic Bottleneck and Its Evolutionary Implications. *Science* **206**:217-18.

Rayko, E., Maumus, F., Maheswari, U., Jabbari, K. & Bowler, C. 2010. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol.* **188**:52-66.

Rodríguez, F., Derelle, E., Guillou, L., Le Gall, F., Vaulot, D. & Moreau, H. 2005. Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ. Microbiol.* **7**:853-59.

Rynearson, T. A. & Armbrust, E. V. 2000. DNA fingerprinting reveals extensive genetic diversity in a field population of the centric diatom *Ditylum brightwellii*. *Limnol. Oceanogr.* **45**:1329-40.

Rynearson, T. A. & Armbrust, E. V. 2004. Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *J. Phycol.* **40**:34-43.

Rynearson, T. A. & Armbrust, E. V. 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom *Ditylum brightwellii*. *Mol. Ecol.* **14**:1631-40.

Rynearson, T. A., Lin, E. O. & Armbrust, E. V. 2009. Metapopulation structure in the planktonic diatom *Ditylum brightwellii* (Bacillariophyceae). *Protist* **160**:111-21.

Rynearson, T. A., Newton, J. A. & Armbrust, E. V. 2006. Spring bloom development, genetic variation, and population succession in the planktonic diatom *Ditylum brightwellii*. *Limnol. Oceanogr.* **51**:1249-61.

Sáez, A. G., Probert, I., Geisen, M., Quinn, P., Young, J. R. & Medlin, L. K. 2003. Pseudo-cryptic speciation in coccolithophores. *Proc. Natl. Acad. Sci. U.S.A.* **100**:7163-68.

Sapriel, G., Quinet, M., Heijde, M., Jourdren, L., Tanty, V. r., Luo, G., Le Crom, S. p. & Lopez, P. J. 2009. Genome-wide transcriptome analyses of silicon metabolism in *Phaeodactylum tricornutum* reveal the multilevel regulation of silicic acid transporters. *Plos One* **4**:e7458.

Sarma, Y. S. R. K. 1983. Algal karyology and evolutionary trends *In*: Sharma, A. K. & Sharma, A. [Eds.] *Chromosomes in evolution of eukaryotic groups.* CRC Press, Boca Raton, Florida, pp. 177-224.

Sarno, D., Kooistra, W. H. C. F., Balzano, S., Hargraves, P. E. & Zingone, A. 2007. Diversity in the genus *Skeletonema* (Bacillariophyceae): III. Phylogenetic position and morphological variability of *Skeletonema costatum* and *Skeletonema grevillei*, with the description of *Skeletonema ardens* sp. nov. *J. Phycol.* **43**:156-70.

Sarno, D., Kooistra, W. H. C. F., Medlin, L. K., Percopo, I. & Zingone, A. 2005. Diversity in the genus *Skeletonema* (Bacillariphyceae). II. An assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *J. Phycol.* **41**:151-76.

Shen, Y. Y., Liang, L., Zhu, Z. H., Zhou, W. P., Irwin, D. M. & Zhang, Y. P. 2010. Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proc. Natl. Acad. Sci. U.S.A.* **107**:8666-71.

Shuter, B. J., Thomas, J. E., Taylor, W. D. & Zimmerman, A. M. 1983. Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells. *Am. Nat.* **122**:26-44.

Siever, R. 1991. Silica in the oceans: biological-geochemical interplay. *In*: Shieder, S. H. & Boston, P. J. [Eds.] *Scientists on Gaia.* The MIT Press, Boston, MA, pp. 287-95.

Sims, P., A., Mann, D., G. & Medlin, L., K. 2006. Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* **45**:361-402.

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D., Coffman, J. A., Dean, M., Elphick, M. R., Ettensohn, C. A., Foltz, K. R., Hamdoun, A., Hynes, R. O., Klein, W. H., Marzluff, W., McClay, D. R., Morris, R. L., Mushegian, A., Rast, J. P., Smith, L. C., Thorndyke, M. C., Vacquier, V. D., Wessel, G. M., Wray, G., Zhang, L., Elsik, C. G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M. J., Mackey, A. J., Maglott, D., Panopoulou, G., Poustka, A. J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C. A., Worley, K., Durbin, K. J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M. L., Jackson, A. R., Milosavljevic, A., Tong, M., Killian, C. E., Livingston, B. T., Wilt, F. H., Adams, N., BellÃ©, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., GeneviÃ¨re, A.-M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A. J., Goldstone, J. V., Cole, B., Epel, D., Gold, B., Hahn, M. E., Howard-Ashby, M., Scally, M., Stegeman, J. J., Allgood, E. L., Cool, J., Judkins, K. M., McCafferty, S. S., Musante, A. M., Obar, R. A., Rawson, A. P., Rossetti, B. J., Gibbons, I. R., Hoffman, M. P., Leone, A., Istrail, S., Materna, S. C., Samanta, M. P., Stolc, V., Tongprasit, W., et al. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**:941-52.

Soltis, D. E., Soltis, P. S., Schemske, D. W., Hancock, J. F., Thompson, J. N., Husband, B. C. & Judd, W. S. 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* **56**:13-30.

Sorhannus, U. 2003. The effect of positive selection on a sexual reproduction gene in *Thalassiosira weissfloggii* (Bacillariophyta): Results obtained from maximum-likelihood and parsimony-based methods. *Mol. Biol. Evol.* **20**:1326-28.

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

Sorhannus, U. & Pond, S. L. K. 2006. Evidence for positive selection on a sexual reproduction gene in the diatom genus *Thalassiosira* (Bacillariophyta). *J. Mol. Evol.* **63**:231-39.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688-90.

Storey, J. D. & Tibshirani, R. 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100**:9440-45.

Subramanian, S. & Kumar, S. 2004. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics* **168**:373-81.

Suzuki, Y. & Nei, M. 2004. False-positive selection identified by ML-based methods: Examples from the Sig1 gene of the diatom *Thalassiosira weissflogii* and the tax gene of a guman T-cell lymphotropic virus. *Mol. Biol. Evol.* **21**:914-21.

Swalwell, J. E., Petersen, T. W. & van den Engh, G. 2009. Virtual-core flow cytometry. *Cytometry* **75A**:960-65.

Swanson, W. J., Wong, A., Wolfner, M. F. & Aquadro, C. F. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* **168**:1457-65.

Tautz, D. & Domazet-Lošo, T. 2011. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**:692-702.

Than, C. & Nakhleh, L. 2009. Species Tree Inference by Minimizing Deep Coalescences. *PLoS Comput Biol* **5**:e1000501.

Theriot, E. C., Ashworth, M., Ruck, E., Nakov, T. & Jansen, R. K. 2010. A preliminary multigene phylogeny of the diatoms (Bacillariophyta): challenges for future research. *Plant Ecology and Evolution* **143**:278-96.

Tréguer, P. & Pondaven, P. 2000. Silica control of carbon dioxide. *Nature* **406**:358-59.

van Oppen, M. J. H., McDonald, B. J., Willis, B. & Miller, D. J. 2001. The evolutionary history of the coral genus *Acropora* (Scleractinia, Cnidaria) based on a mitochondrial and a nuclear marker: reticulation, incomplete lineage sorting, or morphological convergence? *Mol. Biol. Evol.* **18**:1315-29.

Vanormelingen, P., Verleyen, E. & Vyverman, W. 2008. The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodivers. Conserv.* **17**:393-405.

Varelas, X., Samavarchi-Tehrani, P., Narimatsu, M., Weiss, A., Cockburn, K., Larsen, B. G., Rossant, J. & Wrana, J. L. 2010. The Crumbs complex couples cell density sensing to Hippo-dependent control of the TGF-beta-SMAD pathway. *Dev. Cell* **19**:831-44.

Vaulot, D., Olson, R. J. & Chisholm, S. W. 1986. Light and dark control of the cell cycle in two marine phytoplankton species. *Exp. Cell Res.* **167**:38-52.

Veldhuis, M. J. W., Cucci, T. L. & Sieracki, M. E. 1997. Cellular DNA content of marine phytoplankton using two new fluorochromes: Taxonomic and ecological implications. *J. Phycol.* **33**:527-41.

Villareal, T. A., Pilskaln, C., Brzezinski, M., Lipschultz, F., Dennett, M. & Gardner, G. B. 1999. Upward transport of oceanic nitrate by migrating diatom mats. *Nature* **397**:423-25.

von Dassow, P., Petersen, T. W., Chepurnov, V. A. & Armbrust, E. V. 2008. Inter- and intraspecific relationships between nuclear DNA content and cell size in selected members of the centric diatom genus *Thalassiosira* (Bacillariophyceae). *J. Phycol.* **44**:335-49.

Voolstra, C. R., Sunagawa, S., Matz, M. V., Bayer, T., Aranda, M., Buschiazzo, E., DeSalvo, M. K., Lindquist, E., Szmant, A. M., Coffroth, M. A. & Medina, M. n. 2011. Rapid evolution of coral proteins responsible for interaction with the environment. *Plos One* **6**:e20392.

Vyverman, W., Verleyen, E., Wilmotte, A., Hodgson, D. A., Willems, A., Peeters, K., Van de Vijver, B., De Wever, A., Leliaert, F. & Sabbe, K. 2010. Evidence for widespread endemism among Antarctic micro-organisms. *Polar Science* **4**:103-13.

Wernersson, R. & Pedersen, A. G. 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* **31**:3537-39.

Williams, D. W. & Kociolek, J. P. 2010. Towards a comprehensive diatom classification and phylogeny (Bacillariophyta). *Plant Ecology and Evolution* **143**:265-70.

Williams, R. B. 1964. Division rates of salt marsh diatoms in relation to salinity and cell size. *Ecology* **45**:877-80.

Wolf, Y. I., Novichkov, P. S., Karev, G. P., Koonin, E. V. & Lipman, D. J. 2009. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. *Proc. Natl. Acad. Sci. U.S.A.* **106**:7273-80.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**:1586-91.

Yang, Z. & dos Reis, M. 2011. Statistical Properties of the Branch-Site Test of Positive Selection. *Mol. Biol. Evol.* **28**:1217-28.

Yebra, L., Bonnet, D., Harris, R. P., Lindeque, P. K. & Peijnenburg, K. T. C. A. 2011. Barriers in the pelagic: population structuring of *Calanus helgolandicus* and *C. euxinus* in European waters. *Marine Ecology-Progress Series* **428**:135-49.

Zar, J. H. 1996. *Biostatistical analysis.* Prentice-Hall, Upper Saddle River, New Jersey,

Zdobnov, E. M. & Apweiler, R. 2001. InterProScan: an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**:847-48.

Zhang, J. Z. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**:292-98.

Zhang, J. Z., Rosenberg, H. F. & Nei, M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. U.S.A.* **95**:3708-13.

Julie Anna Koester

EDUCATION

2012	**Ph.D. Biological Oceanography** University of Washington
April 2008	Advanced Phycology Course Stazione Zoologica Anton Dohrn
2005	**M.Sc.** Botany and Plant Pathology University of Maine Title: Sexual Reproduction in the marine centric diatom *Ditylum brightwellii*
1991 –1992	East/West Marine Biology Program Northeastern University
1991	**Honors B.Sc**. Marine Biology University of British Columbia Title: Resource allocation in the rockweed *Fucus gardnerii*

OTHER WORK EXPERIENCE

1992 –1995	Fisheries biologist and cartographer, Willamette National Forest; Educator, Catalina Island Marine Institute; Observer, Commercial Fishing Industry; Licensed massage therapist, self-employed

HONORS AND AWARDS

2008, 2009	Association for the Science of Oceanography and Limnology Student presentation awards for talks
2004	University of Maine Graduate Summer Research Grant ($4000)
2004	Phycological Society of America Grant in Aid of Research ($1000)
2004	University of Maine Graduate Student's Travel Grant ($525)

PEER REVIEWD PUBLICATIONS

Koester, J.A., Swalwell, J.E., von Dassow, P., and Armbrust, V.E. 2010. Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta). *BMC Evolutionary Biology*. **10:**1 doi:10.1186/1471-2148-10-1

Koester, J.A., Brawley, S.H.B., Karp-Boss, L. and Mann, D. 2007. Sexual reproduction in the marine centric diatom *Ditylum brightwellii* (Bacillariophyta). *European Journal of Phycology*. 42(4):351-366

Sharpe, S.C., **Koester, J.A.**, Loebl, M., Cockshutt, A. M., Irwin, A.J., and Finkel, Z.V. (submitted for publication 2012). Metabolic size scaling within and across cryptic species of *Ditylum brightwellii* (Bacillariophyceae).

Koester, J.A, Swanson, W.J., and Armbrust, E.V. (in prep). Positive selection in a diatom acts on conserved and lineage specific genes affecting transcriptional regulation and protein interactions.

INVITED PRESENTATION

2011 University of Rhode Island's *Integrative and Evolutionary Biology seminar*
Koester, J.A. *Genomic evidence of speciation and adaption in diatoms*

CONTRIBUTED CONFERENCE PRESENTATIONS

2011 **Koester, J.A,** Swanson, W.J., and Armbrust, E.V. Positive selection is detected within a diatom species and affects regulatory genes. *Phycological Society of America / International Society of Protistologists* joint meeting, Seattle, WA.

2011 **Koester, J.A.** and Armbrust, E.V. Darwinian selection is most evident in closely related diatom species. *Association for the Science of Oceanography and Limnology* meeting, San Juan, Puerto Rico.

2011 Sharpe, S., **Koester, J.A,** Cockshutt, A., Loebl, M., and Finkel, Z**.** Testing the ¾ rule of metabolic scaling within and across two populations of the diatom *Ditylum brightwellii*. *Association for the Science of Oceanography and Limnology* meeting, San Juan, Puerto Rico.

2011 Parker M.S., Iverson, V.I., Berthiaume, C., Lin, E.O, Morales, R., Oleinikov, I.,Knight, M.W., Schruth, D., **Koester, J.A.**, and Armbrust, E.V. Intra-species genetic diversity of the centric diatom Thalassiosira pseudonana: Whole genome comparisons of seven strains. *1st International Conference: "The Molecular Life of Diatoms*, Atlanta, GA.

2010 **Koester, J.A,** Berthiaume, C., Schruth, D and Armbrust, E.V. Seeking evidence of positive selection in diatoms. *Ocean Sciences* meeting, Portland, OR.

2009 **Koester, J.A,** Swalwell, J., and Armbrust, E.V. Changes in genome size as a mechanism of population divergence in the marine planktonic diatom, *Ditylum brightwellii*. *Association for the Science of Oceanography and Limnology* meeting, Nice, France.

2008 **Koester, J.A,** Swalwell, J., van den Engh, G., and Armbrust, E.V. Relationships between DNA content, cell size, and growth rate in populations of the planktonic diatom Ditylum brightwellii. *Association for the Science of Oceanography and Limnology* meeting, St. John's, New Foundland, Canada

2007 **Koester, J.A.** and Armbrust, E.V. Physiological variation in the marine diatom, *Ditylum brightwellii*. *Association for the Science of Oceanography and Limnology* meeting, Santa Fe, NM. (poster)

2005 **Koester, J.A.,** Karp-Boss, L, Brawley, S.H. Induction of sexual reproduction and the mating system of the planktonic diatom *Ditylum brightwellii*. *Association for the Science of Oceanography and Limnology* meeting, Salt Lake, UT. (poster)

2005 **Koester, J.A.,** Brawley, S.H. Sexual reproduction in *Ditylum brightwellii*. *45th Northeast Algal Society Symposium*, Rockland, ME.

TEACHING EXPERIENCE

Invited lectures and demonstrations in Marine Phytoplanton (OCEAN 531):

2008 "Dinoflagellates: Diversity and Ecology"

2008, 2009 "Use of microscopy and flow-cytometry to identify phytoplankton diversity"

Laboratory and field instructor, University of Maine: Biology 100 (Fall 2002), Biology 200 (Spring 2005), Human Anatomy and Physiology (Spring 2003, 2004), Marine Ecology Field Course (Fall 2003, 2004).

MENTORSHIP

Mentor: Advised and supervised laboratory work of undergraduate researchers Rachelle Lambert (2006 – 2009) and Kevin Tran (2009 – 2010) from the University Washington. Advised experimental progress of Susan Sharpe* from Mount Allison University (2008 – 2011)

- * has gone on to pursue a higher degree in science

Outreach Mentor (2010 – 2011): Girls in Engineering, Math and Science, a program through Association for Women in Science, Seattle Chapter. Mentored monthly meetings of 30 middle school girls in scientific activities.

SYNERGISTIC ACTIVITIES

Student Member, Phycological Society of America, Program Committee (2010 to present). Duties include organization and planning of symposia, sessions, and workshops for annual meetings. Planned and particpateded in a field trip for international scientists to Seattle's municipal watershed in 2011.

Co-organizer (2009): Cascadian Interdisciplinary Seminar Series, University of Washington, School of Oceanography. Four speakers from the Pacific Northwest addressed current topics in chemical, physical, biological oceanography, and astrobiology.

Researcher for scientific cruises

2010 SeaFlow flow-through cytometry and concurrent seawater filtration for nucleic acids (Seattle, Washington to Honolulu, Hawaii)

2009 Oceans and Human Health education and research cruise (Puget Sound, Washington)

2006 Puget Sound Regional Synthesis (PRISM) biannual cruise

PROFESSIONAL MEMBERSHIP

Phycological Society of America (2003 to present)
Association for the Science of Oceanography and Limnology (2007 to present)