2015

COMMUNITY COMPOSITION OF THE MORPHOLOGICALLY CRYPTIC DIATOM GENUS SKELETONEMA IN NARRAGANSETT BAY

Kelly Canesi
University of Rhode Island, kelly_canesi@my.uri.edu

COMMUNITY COMPOSITION OF THE MORPHOLOGICALLY CRYPTIC DIATOM GENUS *SKELETONEMA* IN NARRAGANSETT BAY

BY

KELLY CANESI

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

MASTER OF SCIENCE

IN

OCEANOGRAPHY

UNIVERSITY OF RHODE ISLAND

2015

MASTER OF SCIENCE THESIS

OF

KELLY CANESI

APPROVED:

Thesis Committee:

Major Professor: Tatiana Rynearson

Candace Oviatt

Christopher Lane

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2015

ABSTRACT

It is well known that morphologically cryptic species are routinely present in planktonic communities but their role in important ecological and biogeochemical processes is poorly understood. I investigated the presence of cryptic species in the genus *Skeletonema*, an important bloom-forming diatom, using high-throughput genetic sequencing and examined the ecological dynamics of communities relative to environmental conditions. Samples were obtained from the Narragansett Bay Long-Term Plankton Time Series, where *Skeletonema* spp. can comprise >95% of microplankton cells during blooms. DNA was extracted and sequenced from monthly samples between December 2008 and December 2013, and *Skeletonema* specific primers were used to exclusively target and differentiate known species. Seven species of *Skeletonema* were found in Narragansett Bay over the sampling period, including five species that were previously undetected in this location. *Skeletonema* community composition was highly seasonal, and temperature had the greatest effect on composition changes over time. Winter and spring communities demonstrated less species richness than did summer and autumn communities. These data suggest that a *Skeletonema* community can be comprised of different species, even in one estuary, and that seasonal changes appear to have a substantial effect on community structure, and perhaps important ecological processes, over time.

ACKNOWLEDGMENTS

Many people were involved in the formation of this thesis. First and foremost, I must thank Dr. Tatiana Rynearson for her knowledge, advice, patience, and good humor throughout my time in graduate school. She helped me grow as a scientist and student by encouraging me to overcome all obstacles in my way and to seek out answers to even the most difficult of questions, both in science and in life. She demonstrated to me the importance of leading with passion, sincerity, and respect. I am honored to have had Tatiana as a mentor and will carry her guidance with me throughout my career and life.

Many thanks to my thesis committee consisting of Dr. Candace Oviatt and Dr. Chris Lane for their thoughtful feedback and advice, as well as Dr. Bethany Jenkins for serving as defense chair and providing her own useful insight. During my time at GSO, I also worked closely with the Metcalf Institute for Marine and Environmental Reporting, and for that I must thank Dr. Sunshine Menezes, Katharine McDuffie, and Karen Southern for opening my eyes to new possibilities and allowing me to expand my horizons beyond the lab bench. GSO will always have a special place in my heart because of the many wonderful faculty, staff, and students I met there. I would like to particularly thank Dr. Susanne Menden-Deuer, Dr. David Smith, and Meredith Clark for their hard work at GSO and their personal guidance, as well as the many faculty members who contributed to my studies in oceanography.

The research described in this thesis was supported by Rhode Island Sea Grant. Additional academic support was provided by the Metcalf Institute for Marine and Environmental Reporting, the Graduate School of Oceanography Alumni Fellowship, and the Lance A. Ricci Fellowship.

I truly appreciate the time and expertise shared by Dan Roche, Maria Casas, and Andrea Drzewianowski to learn the necessary techniques to carry out my research. Thank you to Malcolm McFarland and Ryan Morse for providing technical assistance as well as samples and data for certain dates within the time series. Extra special thanks to my wonderful lab mates for being a source of inspiration and friendship: Kerry Whittaker, Gang Chen, Amanda Montalbano, Sarah Flickinger, Olivia Ahern, Kristen Hunter-Cereva, Shu Cheng, Chris O'Brien and Maddie Willert, as well as the many wonderful undergraduates with whom I worked. Additionally, many thanks to those who provided their valuable advice and moral support over the past few years: Mary Kane, Francoise Morison, Alison Cleary, Sam DeCuollo, Becca Trueman, Kevin Aviles, Jhoset Burgos, Melanie Garate, Megan Gahm, Kelsey James, and Ashley Marranzino. As an honorary member of the Rynearson lab, thank you to Elena May Menden for being a source of joy and laughter with her special appearances in the office.

Finally, I would like to thank my friends and family for supporting me through an exciting, and sometimes tumultuous, period in my life. I must particularly thank four people without whom I would not be where I am today: my sister Andrea, who has always been my role model and reminds me every day the importance of being true to be myself; Meg, for our unwavering and unforgettable friendship throughout the past 15 years (and many more to come); and my parents, Bill and Maureen, who are my everything and taught me to never lose faith in the world and, more importantly, in myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Diatoms are single-celled planktonic algae that account for 40-45% of oceanic carbon production, meaning they are capable of providing nearly half of the energy required by ocean life (Field, 1998). Their incredible contribution to marine ecosystems necessitates a strong understanding of the dynamics controlling diatom populations and communities, which is tied to the understanding and long-term monitoring of ecosystem health. An estimated 12,000 species have been described, but some researchers estimate that greater than 200,000 diatom species exist, within which a vast majority are undescribed (Guiry, 2012). The discovery of undescribed diatom species can lead to improved understanding of the dynamics and interactions within phytoplankton communities, particularly when considering the important ecological role diatoms play in marine ecosystems.

A large portion of these undescribed species belong to a category called 'cryptic species', loosely defined as being morphologically indistinguishable (at least superficially) but genetically distinct (Bickford et al., 2007). The concept of cryptic speciation is not new to microbial communities, but marine systems are so diverse that they remain key targets for the study of cryptic species (Mann, 1999). Marine habitats are known for their vastly unknown diversity among not only diatoms but also many other groups and the myriad specialized interactions between unique organisms. Diatoms in particular play an important functional role in marine habitats, but how cryptic species may differ in their ecological roles is unclear.

One example of a cryptic diatom genus is *Skeletonema,* once considered to be the single species *Skeletonema costatum* with a high degree of genetic variability but is in

fact a complex of at least 10 marine species. Species identified include *S. menzelii* (Guillard et al., 1974), *S. pseudocostatum* (Medlin et al., 1991), and *S. subsalum* (Hasle and Evensen, 1975). Sarno et al. (2005) discovered four additional species of *Skeletonema* using light microscopy, electron microscopy and genetic sequencing. These species included: *S. dohrnii*, *S. marinoi*, *S. japonicum*, and *S. grethae*. Other species have been described, including *S. ardens* and *S. grevelli* (Sarno et al., 2007). Few studies have addressed species composition *in situ* and only two species have been identified in Narragansett Bay (Kooistra et al., 2008).

Skeletonema is one of the most abundant phytoplankton genera found in Narragansett Bay. It often dominates phytoplankton blooms and is found in the bay throughout the year (Windecker, 2010). It is a cosmopolitan genus, also found in marine ecosystems globally (Kooistra et al., 2008). *Skeletonema* represent an average of 49% of the phytoplankton community in Narragansett Bay and can comprise up to 99% of the total phytoplankton cells during peak bloom times (Windecker, 2010).

Over time, *Skeletonema* abundance in Narragansett Bay has decreased from an average of 2137 cells ml^{-1} between the years 1959 and 1980 to 1128 cells ml^{-1} between 1980 and 1997, with average cell abundance remaining at approximately 1220 cells ml^{-1} until 2005 (Borkman and Smayda, 2009). *Skeletonema* abundance has changed in other locations as well. A study by Biswas et al. (2009) demonstrated that average cell abundance phytoplankton changed over a period of nearly two decades from 1990 to 2007 in the Bay of Bengal. *Skeletonema* increased from a relative cell abundance of 1.42% (out of the total phytoplankton population) in 1990 to 12.66% in 2000, and even as high as 18.95% in 2007. This differs from the study in Narragansett Bay not only because

the relative cell abundance appears to be much lower than Narragansett Bay's average *Skeletonema* cell abundance of 49%, but also because the relative abundance has actually increased in recent years. While it is apparent that these changes are occurring both in Narragansett Bay and globally, a mechanism has not yet been described to explain these shifts. Furthermore, the presence of cryptic species of *Skeletonema* in Narragansett Bay is unknown and may account for changes in community composition based on potential physiological differences among species.

Studying changes in *Skeletonema* community composition over time is a first step towards understanding differences among morphologically cryptic species and the roles they play in marine ecosystems. These distinctions may also inform projections of future change in phytoplankton communities as the environment and climate undergo rapid changes over time, not only in Narragansett Bay but also in coastal ecosystems around the globe. We address three questions in this study: 1) Do multiple species of *Skeletonema* exist in Narragansett Bay? 2) Does *Skeletonema* community composition vary seasonally, or does community composition remain the same throughout the year? 3) Do *Skeletonema* community shifts correspond to seasonal fluctuations in environmental variables, such as temperature, irradiance, and nutrient concentration? We predict that multiple cryptic species will be detected in Narragansett Bay using novel molecular techniques, and that environmental conditions will cause community composition to change on seasonal time scales due to genetic, and possibly physiologic, differences among species.

METHODS

Sample collection

Filtered field samples, cell counts, and environmental data were provided by the Narragansett Bay Long-Term Plankton Time Series (www.gso.uri.edu/phytoplankton). Surface water samples were collected weekly from Narragansett Bay (41 34.2N, 71 23.4W) and 200-250 ml seawater were filtered in triplicate onto 25 mm diameter 0.22 μm pore-size filters (EMD Millipore), flash-frozen in LN_2, and stored at -80°C. From this archive of frozen filters, a series of monthly samples was selected between December 2008 and December 2013, based on the sampling date with the highest percentage of *Skeletonema* cells in the phytoplankton community. In addition, a series of weekly samples were selected from three representative bloom periods (winter bloom: December 2008-January 2009; spring bloom: March 2010-April 2010; summer bloom: May 2013-July 2013) (Table 1).

Method validation

In order to evaluate the utility of high-throughput sequencing to identify closely-related *Skeletonema* species, simulated communities of mixed phytoplankton species were created and included along with analysis of field samples. Three simulated communities were generated by mixing exponentially growing cultures of five *Skeletonema* species in equal abundances and adding this to exponentially growing cultures of four other phytoplankton species that were also mixed in equal abundances. The three simulated communities were comprised of an 80%, 50% and 20% mixture of *Skeletonema* cells. The remainder of each community was comprised of the mixture of other phytoplankton. The species used were *S. dohrnii* (CCMP3373), *S. grethae*

(CCMP1804), *S. japonicum* (CCMP2506), *S. marinoi* (isolated from Narragansett Bay on January 1, 2012), *S. menzelii* (CCMP793), *Ditylum brightwellii* (isolated from Puget Sound on May 12, 2007), *Heterocapsa triquetra* (CCMP448), *Thalassiosira pseudonana* (CCMP1335), and *Thalassiosira rotula* (CCMP3096). To generate the simulated communities, cell abundance in each culture was determined microscopically and used to determine the volume needed to create communities with a total of 4×10^6 cells and the appropriate *Skeletonema* concentration (20, 50 or 80%). Three simulated communities were produced by combining appropriate volumes of all cultures in an acid-washed bottle and inverting 40 times. Communities were immediately harvested in triplicate onto 25 mm diameter 0.22 μm pore-size filters (EMD Millipore), flash-frozen in LN_2, and stored at -80°C, resulting in a total of nine samples with three filters per community.

Technical replication was measured at three steps of the method: DNA extraction, DNA amplification, and DNA sequencing. Technical replicates were selected from simulated community samples and randomly from the available field samples. At the DNA extraction level, three simulated communities of varying *Skeletonema* abundance were filtered in triplicate and DNA was extracted from each filter in separate reactions. Each extraction was then treated as an individual sample in the following steps, resulting in nine total samples that were amplified and sequenced to represent extraction replication. At the DNA amplification level, one sampling date was selected: June 15, 2010. DNA was extracted from a single filter collected on this date, and three separate PCRs were run using DNA from the single extraction. Each PCR was then treated as an individual sample during sequencing, resulting in three total samples to represent amplification replication. At the DNA sequencing level, one sampling date was selected:

September 4, 2009. DNA was extracted and amplified from a single filter collected on this date, and the PCR product was then split into three aliquots of equal volume to be sequenced separately. By measuring technical replication at these three steps, we evaluated not only the precision of the method but also the propagation of error through a series of steps involved in characterizing *Skeletonema* diversity.

DNA extraction and amplification

DNA was extracted from frozen filters using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions with the addition of a disruption during the lysis step using triple-pure high-impact 0.5 mm zirconium beads (KSE Scientific). A small scoop was used to add approximately 50 beads to each sample before vortexing and incubating. A 325 bp fragment within the D1-D4 region of the LSU rDNA gene was amplified via polymerase chain reaction (PCR) using *Skeletonema* genus-specific primers (forward: 5'-GTTAAAGAGTACCTGAAATTGTTAA-3'; reverse: 5'-TGTTACTTTCATTACGCATATCA-3'). Primers amplified a region of the DNA that can be used to distinguish among *Skeletonema* species, with the exception of *S. grethae*, which could not be distinguished from *S. tropicum* (Table 2). DNA sequences from cultured cells and available reference sequences indicate higher than 99% similarity between these two species in the LSU D1-D4 region. Illumina adapters of varying length were added to the 5' ends of the primers to increase diversity in the sequencing pool and thus reduce sequencing error (Table 3). DNA was amplified in 50 µl reactions containing 25 µl 2x Bio-x-act Short Mix (Bioline USA, Inc.), 0.5 µM forward primer (0.125 µM each of the four Illumina adapted primers) 0.5 µM reverse primer (0.125 µM each of the four Illumina adapted primers), 1.25-5 µl template DNA, and H_2O. A two-step PCR

consisted of a 10 minute denaturing step at 94°C; 15 cycles of 94°C for 30 seconds, 61°C for 30 seconds, and 72°C for 30 seconds; 15 cycles of 94°C for 30 seconds, 66°C for 30 seconds, and 72°C for 30 seconds; and a final step of 72°C for 10 minutes. The first step of the PCR allowed primers to bind to the DNA and amplify the fragment of interest, and the second step of the PCR extended those fragments to include the Illumina adapters. Amplicons were visualized and confirmed using agarose gel electrophoresis and cleaned via ethanol precipitation.

High-throughput sequencing and analysis

PCR amplicons were submitted to the URI Genomics and Sequencing Center and were sequenced using the MiSeq platform (Illumina, Inc.). PCR amplicons were cleaned using Ampure XP beads (Beckman Coulter, Inc.) then quantified with the Qubit high sensitivity DNA kit (Thermo Fisher Scientific, Inc.) to verify DNA concentration. Next, cleaned amplicons underwent a second round of PCR with five cycles to attach Nextera indices and adapters (Illumina, Inc.) followed by a second round of Ampure XP clean up. Amplicons were visualized by agarose gel electrophoresis and analyzed using the BioAnalyzer DNA1000 chip (Agilent Technologies, Inc.) to confirm length. DNA concentrations were quantified again using the Qubit high sensitivity DNA kit. Samples were pooled and quantified using the KAPA qPCR kit (Kapa Biosystems). The pooled DNA contained all 96 simulated communities and field samples, which were multiplexed on a single flow cell and run on the MiSeq platform using the MiSeq 500-cycle (2x250 bp reads) kit. Barcoded sequences were then uploaded to BaseSpace (www.basespace.com) by the URI GSC and downloaded for analysis.

A total of 16.35 x 10^6 reads representing 96 amplicon library samples were analyzed. Each sample produced two .fastq files, one containing forward reads and one containing reverse reads. A series of tests were performed to select the optimal methods for merging paired-end (PE) reads. First, we compared the QIIME (Caporaso et al., 2010) join_paired_ends.py algorithm and the USEARCH (Edgar, 2010) –fastq_mergepairs algorithm. Ultimately, PE reads were merged using the USEARCH algorithm. Within the USEARCH algorithm, a quality filter was tested. Raw reads were truncated to the first base pair with a quality score of at least Q30, and this was done separately both before and after PE reads were merged. Based on the accuracy of the resulting data, we opted to perform Q30 trimming before merging PE reads. Also within the –fastq_mergepairs algorithm, PE reads were removed if one or more mismatches occurred in the overlapping sequence. Additional quality filtering and conversion to .fasta format was performed using the USEARCH fastq_filter algorithm, which further truncated sequences to a length of 315 bp to generate reads of equal length for downstream processing. Converted .fasta files were imported to the Center for Computation and Visualization (Brown University) where QIIME was used to combine .fasta files into a single .fna file using the add_qiime_labels.py script. Sequences were assigned OTUs using QIIME's pick_open_reference_otus.py script. A minimum percent ID of 60% (which is the default value used in QIIME) was compared to a minimum percent ID of 99%, and the 99% ID filter produced more reliable results. A reference database representing a suite of *Skeletonema* species was created by searching NCBI for *Skeletonema* 28S sequences (Table 4). All sequences were clustered into OTUs based on the reference database and *de novo* clustering was not necessary.

Data analysis and statistical methods

For each sample, the percent sequence reads corresponding to each species was determined by dividing the number of sequence reads per species by the total number of sequence reads. Species represented by < 0.1% of the sequence reads in a sample were considered as potential sequencing errors (due to the high degree of similarity among *Skeletonema* 28S sequences) and removed from the data set.

Community analysis was performed in PRIMER v6.1.6 (PRIMER-E Ltd.). Percent composition data were used to evaluate similarities among samples. A Bray-Curtis similarity matrix was produced and the CLUSTER analysis tool was used to visualize similarity among seasonal samples obtained during winter (December-February), spring (March- May), summer (June- August) and autumn (September-November). Significant differences between seasons was determined by performing an Analysis of Similarity (ANOSIM).

Environmental data were obtained from a variety of sources. Surface temperature, surface salinity, DIN, DIP, and Si data were obtained from the Narragansett Bay Long-Term Plankton Time Series and from the Marine Ecosystems Research Laboratory (http://www.gso.uri.edu/merl/merl.html) when data were not available from the time series. Photosynthetically active radiation (PAR) data were obtained from the National Estuarine Research Reserve System's Centralized Data Management Office for the Narragansett Bay station (41.637N, 71.339W) (http://cdmo.baruch.sc.edu/). Surface salinity and DIN values were log transformed and all data were normalized. An environmental PCA was constructed and correlated with the community similarity matrix using the BIO-ENV analysis.

Calculation of absolute species abundance for winter samples

For a subset of the samples, absolute species abundance was calculated. To generate the conversions from relative to absolute species abundance, we used the sequence data from the simulated communities. For each of three simulated communities (analyzed in triplicate), the expected number of sequence reads for each *Skeletonema* species was 20% of the total sequence reads since the five species were added in equal proportions. To obtain a transformation coefficient for each species, the number of expected sequence reads was divided by the observed number of sequence reads. The transformation coefficients were averaged across all 9 samples to obtain a single coefficient for each of the five species used in the simulated communities (Table 5).

Because transformation coefficients could only be determined for the five species used in the simulated communities, two species identified in Narragansett Bay (*S. costatum* and *S. pseudocostatum)* were not included in the analysis. Since we did not have an estimate of over- or underrepresentation for either of these two species, we did not consider field samples with > 1% of either or both species present. As a result, only field samples containing the five species with coefficients (and possibly minute amounts of two additional species) were considered. Transformation coefficients were first multiplied by the number of sequence reads. Next, the transformed sequence reads were converted to percent composition values. Finally, percent composition values were applied to absolute *Skeletonema* cell counts from each field sample to obtain individual species cell counts. Since spring, summer, and autumn samples tended to have high levels of *S. costatum* and *S. pseudocostatum* and could not be considered in our analysis, we focused on winter samples that consistently lacked these two species.

RESULTS

Propagation of error in DNA processing

Technical replication was evaluated at three steps of the method: extraction, amplification (PCR), and sequencing of DNA. Extraction replicates were derived from simulated communities that were filtered and extracted in triplicate. Extraction replicates were generally associated with low levels of variance, defined as a coefficient of variation (CV) less than 15%, although variance exceeded 15% for some species. Variance exceeding 15% was associated with *S. menzelii* in the 50% *Skeletonema* community and with *S. dohrnii* and *S. menzelii* in the 20% *Skeleonema* community (Table 6). Among extraction triplicates, the overall contribution of a single species to a community was not associated with standard deviation, as species with both higher and lower contributions to the community experienced lower levels of variance than those that exceeded 15% variance.

Amplification and sequencing triplicates were associated with less error than extraction triplicates (Table 6). Species with low sequence read abundance were associated with high levels of variance but were rare among the samples. In contrast, species that dominated samples with high sequence read abundance were associated with low levels of variance. For example, *S. pseudocostatum* in the June 2010 amplification triplicates comprised 75.42% of sequence reads in that sample and demonstrated only 0.92% variance among triplicates (Table 6). Overall, technical errors associated with amplification and sequencing were low and tended to affect low abundance species much more than high abundance species.

Selection of optimal sequence analysis method

Due to the complexity of amplicon MiSeq data and the close genetic relationship among *Skeletonema* species, computational steps were extensively tested to identify the most accurate method of processing high-throughput sequencing data. These tests were performed on a single simulated community containing equal concentrations of five species: *S. dohrnii, S. grethae, S. japonicum, S. marinoi,* and *S. menzelii*. The final workflow consisted of quality filtering of raw sequence reads, length truncation of merged reads, and prefiltering of sequences against a set of reference sequences (Figure 1). This section describes options that were compared at each step and how the final workflow was selected.

First, two algorithms commonly used to merge paired-end (PE) amplicon reads were compared: QIIME's join_paired_ends.py algorithm and USEARCH's –fastq_mergepairs algorithm. These two methods were similar when comparing the total number of merged sequences with a minimum base pair overlap filter implemented for a single sample (Figure 2). The USEARCH algorithm resulted in a maximum yield of 95449 PE reads compared to a maximum yield of only 88619 PE reads in QIIME. The USEARCH –fastq_mergepairs algorithm was selected to merge forward and reverse reads.

Within the USEARCH –fastq_mergepairs algorithm, several parameters were tested to further optimize data processing. Reads were truncated inwards from each end of a read to the first base pair with a quality score of Q30; this was performed separately both before and after PE reads were merged. When Q30 truncation occurred after PE reads were joined, two species were identified in the sequence data that were known to be

absent from the simulated community: *S. potamos* and *S. pseudocostatum*. A small number of reads were identified as *Skeletonema sp.* (0.3%) or had no BLAST hit (0.2%) when they failed to cluster with OTUs in our custom *Skeletonema* reference database. Also, up to eight OTUs were classified for a single taxon using this method. When quality filtering occurred before PE reads were joined, no unexpected OTUs appeared in the data and only a single OTU was assigned to each species (Figure 3).

After merging PE reads, two additional parameters were tested. First, a maximum length of 315 bp (10 bp short of the full fragment length) was found to improve the accuracy of results. It removed only highly conserved bps with no phylogenetic information and ensured that the length of all reads did not exceed the length of any reference sequences (Figure 3). Second, a 99% identity filter in the OTU picking script eliminated OTUs that did not match sequences of species that were known to be present in the sample. When OTUs were required to match reference sequences at either 60% (the default value) or 99%, the 99% sequence match resulted in loss of about 80% of the sequence data compared to the default value but eliminated OTUs that were identified as *Skeletonema sp.* or had no BLAST hit (Figure 3). This only had a noticeable effect when Q30 trimming was performed after merging PE reads. An additional quality filter, varying the number of mismatches allowed in the overlapping region, did not have a noticeable effect on OTU picking results (Figure 3).

Based on the results described above, a combination of USEARCH –fastq_mergepairs and –fastq_filter algorithms and QIIME pick_open_reference_otus.py algorithm was used to process sequences, and a suite of quality filters and script parameters were compared for a single sample with known concentrations of five

Skeletonema species (Figure 1). The results of these tests revealed that the ideal method of processing sequence reads included a) Q30 trimming before merging PE reads and elimination of sequences containing one or more mismatches in the overlapping region using the USEARCH –fastq_mergepairs algorithm; b) contig length truncation to 315 base pairs using the USEARCH –fastq_filter script; and c) a 99% prefilter ID using the QIIME pick_open_reference_otus.py script (Figure 1).

Simulated communities

Three simulated field communities were created in triplicate by mixing known numbers of five *Skeletonema* species and four other phytoplankton species in culture. The *Skeletonema* species were included in equal proportions and combined to comprise 20, 50 or 80% of the total cells in the community. All five *Skeletonema* species were detected by high-throughput sequencing and none of the other phytoplankton species were detected, confirming that the primers were both specific to the genus *Skeletonema* and amplified a region that could be used to distinguish among species (Figure 4). The average coefficient of variation (CV) within species was 13.77% and did not exceed 15% for any species with the exception of *S. dohrnii* in the 20% *Skeletonema* community (CV=31.38%) and *S. menzelii* in the 80% *Skeletonema* community (CV=24.07%) and the 20% *Skeletonema* community (CV=22.02%).

Both within and among mixed communities, *Skeletonema* species composition did not vary significantly and was not affected by the absolute concentration of *Skeletonema* cells in each sample as evidenced by the consistent representation of species across communities (ANOVA, $p<0.01$) (Figure 5). Although the five species were expected to contribute equally to the total *Skeletonema* community in all three experimental

conditions (i.e. each species was expected to comprise 1/5 of the sequence data regardless of the total contribution of *Skeletonema* to the phytoplankton community), the observed percent composition for each *Skeletonema* species actually ranged from 1.97% to 53.41% and differed significantly from the expected community compositions (Chi-square goodness of fit test with 4 degrees of freedom, χ^2>43% and p<0.001 for all three communities).

Community composition and seasonality of Skeletonema *in Narragansett Bay*

Total abundance of *Skeletonema* in Narragansett Bay between December 2008 and December 2013 ranged from 0 cells L^{-1} to 4.8E+7 cells L^{-1} with the maximum abundance occurring on April 13, 2010. The timing of the major *Skeletonema* bloom varied from year to year with large blooms (approximately 2.0E+7 to 4.8E+7 cells L^{-1}) occurring in winter-spring of 2010, early-mid summer of 2011 and 2013, and autumn of 2009 (Figure 6). Smaller blooms occurred during winter-spring of 2011, 2012, and 2013.

Between December 2008 and December 2013, seven species of *Skeletonema* were detected in Narragansett Bay: *S. costatum, S. dohrnii, S. grethae (tropicum), S. japonicum, S. marinoi, S. menzelii*, and *S. pseudocostatum*. Note that *S. grethae* cannot be distinguished from *S. tropicum*. One of three reference sequences for *S. tropicum* matched those of *S. grethae* completely. Two other *S. tropicum* reference sequences only had up to three bp differences from *S. grethae*, including two bp differences that were ambiguous (R and Y, respectively for purines and pyrimidines), and the percent similarity between *S. grethae* and *S. tropicum* reference sequences always exceeded 99% (Table 2). Sequences identified as *S. costatum* only appeared on August 10, 2009 and comprised 0.3% of the total sequences. Four of the species consisted of one OTU each: *S. costatum,*

S. japonicum, S. marinoi, and *S. pseudocostatum.* The remaining three species, *S. dohrnii*, *S. grethae*, and *S menzelii*, were clustered into two OTUs each, and sequence reads were combined to obtain a total number of sequences for each species.

Percent composition of sequence reads was used to analyze community composition. Since simulated communities suggested that species like *S. japonicum* are overrepresented while others like *S. marinoi* are underrepresented (Figure 4), it was not appropriate to treat percent composition of sequence reads as equivalent to percent composition of species abundance. Instead, percent composition of sequence reads was used to evaluate relative change over time in the community.

Variation in *Skeletonema* community composition was strongly associated with season (Figure 6). Field samples clustered into two distinct groups, one primarily dominated by winter and spring samples and the other primarily dominated by summer and autumn samples (Figure 7). This reflects the major shift observed between May and June (Figure 6) where low richness during winter and spring months quickly transitioned to higher richness during summer and fall months (Figure 8). The winter-spring cluster appears to have more similar community composition among field samples than the summer-autumn cluster, which also reflects the shift to a richer and unpredictable community. Pairwise comparisons among seasons further indicate that significant differences exist between all seasons ($p=0.001$) with the exception of winter-spring and summer-autumn (Table 7). Winter and spring months generally consisted of a single species (*S. marinoi*) and occasionally presented other species, especially *S. dohrnii, S. japonicum,* and *S. menzelii*. With the exception of 2011, community shifts were drastic between April and June, switching from a community primarily dominated by *S. marinoi*

to a richer community of *Skeletonema*. In 2011, *S. marinoi* continued to be the primary species during the month of June, and the shift to a richer community occurred after June.

Effect of fluctuating environmental conditions on community composition

Between December 2008 and December 2013, Narragansett Bay experienced a wide range of physical and chemical environmental conditions. Temperature during this time period ranged from 0°C to 24.6°C. Salinity ranged from 14.1 psu to 32 psu; it should be noted that the minimum salinity occurred during April 2010 and coincided with severe rains and heavy influx of freshwater to Narragansett Bay. Dissolved inorganic nitrogen (DIN) ranged from 0.14 to 18.92 μM, dissolved inorganic phosphorus (DIP) ranged from 0.04 to 1.31 μM, and Si ranged from 0.083 to 38.121 μM. Photosynthetically active radiation (PAR) ranged from 16.21 to 572.05 mmol/m^2. Seasonal shifts were primarily dependent on temperature (Figure 9).

Principal Components Analysis showed distinct separation of seasons in two directions (Figure 9). The first two axes of the PCA explained 72.2% of the variability in environmental data. The first PC axis roughly separated spring and autumn samples by nutrient availability (DIN, DIP, Si), with winter and summer samples distributed about equally along the PC axis. Low nutrients were more strongly correlated with spring samples and high nutrients were more strongly correlated with autumn samples. The second PC axis separated samples by temperature and PAR, correlating winter samples with low temperature/light and summer samples with high temperature/light. Spring samples had approximately equal distributions of high and low temperatures/light, while autumn samples tended to cluster with higher temperatures/light.

Environmental conditions, primarily temperature, were correlated with species composition. Surface temperature and DIP were significantly correlated with species composition and explained 48.0% of the variation in *Skeletonema* communities ($p<0.01$) (Table 8). Temperature alone explained 46.2% of the variability, while DIP explained less than 2% of the variability and may not have as large an effect on community composition. When temperature was removed from analysis, DIP had the greatest effect on community composition, explaining 21.9% of the variability in the data, but was not significantly correlated with species composition (Table 9).

Contribution of S. marinoi *to winter communities*

Transformation coefficients for five species (*S. dohrnii*, *S. grethae*, *S. japonicum*, *S. marinoi*, and *S. menzelii*) were used convert raw sequence reads to absolute cell counts and eliminated over- and underrepresentation of species (Table 5). We focused on winter samples (December-February), which showed that *S. marinoi* comprised 8-100% of the community during these months. Contribution of *S. marinoi* generally exceeded 80% of the community and was rarely present in low abundance.

Other species were rarely abundant in winter communities. Dates that coincided with low *S. marinoi* abundance relative to other species appeared to be influenced by unusual conditions in the bay. For example, in December 2011 community composition was 29% *S. dohrnii*, 24% *S. grethae*, 44% *S. marinoi*, and 3% *S. menzelii*, and in January 2013 community composition was 8% *S. marinoi*, 90% *S. dohrnii*, and 2% *S. menzelii*. On these dates, both *Skeletonema* abundance and overall phytoplankton abundance were low (Figure 10). December 2011 also experienced a higher than normal surface temperature of 11.5°C (January 2013 surface temperatures were within the typical range).

Chlorophyll-*a* concentration generally followed changes in overall *Skeletonema* abundance and was particularly high in January 2010 (25 ug/L). This coincided with a large *Skeletonema* bloom that was dominated by *S. marinoi*.

DISCUSSION

Skeletonema is a diatom genus comprised of up to 10 morphologically cryptic species. Previously, studies have focused on genus-level identification to measure differences among *Skeletonema* populations, but we approached the Narragansett Bay community by identifying cryptic species and their seasonal distributions over a five year time period. We found that *Skeletonema* experiences distinct shifts over seasons that may relate to environmental changes, particularly in temperature.

Use of high-throughput molecular methods to evaluate community composition

Skeletonema species are closely related, and methods of determining taxonomy must be sensitive enough to detect small differences in DNA sequence and sequence read counts among such species. In this study, high-throughput sequencing was effective in characterizing a marine diatom community consisting of closely related, morphologically cryptic *Skeletonema* species. Recent technology advances have made it easier to detect presence of unique species in microbial communities (Amaral-Zettler et al., 2009), but since we were interested in not only presence-absence but also relative and absolute change over time we included replicate measurements as well as the construction and analysis of simulated communities to evaluate accuracy. High-throughput sequencing also provides many opportunities within the experimental setup for errors to occur, either by the user or by mechanical or computational processes. When billions of bases are being sequenced rapidly and for hundreds of samples at a time, the possibility of experimental error is great (reviewed in Robasky et al., 2014).

Errors typically had a greater effect on species that revealed low sequence read abundance than on those that made up a majority of the community. These errors were

most evident during the DNA extraction step. In contrast, downstream steps of amplification and sequencing had much lower levels of variation. Despite the presence of technical error, high-throughput sequencing is still a trustworthy method to characterize microbial communities. This allows us to focus on species that are more abundant and potentially play a more significant ecological role in marine systems.

Closely-related species like *Skeletonema* are difficult to study due to experimental errors that can accumulate throughout the method and cause false identification in downstream processing steps. By applying strict quality controls that target the specific needs of the study, we mitigated some of this error and eliminated many false reads and identifications. One technique that was notably different in our study compared to others was the application of Q30 trimming. It is typically recommended to perform Q30 trimming after merging PE reads (Bokulich et al., 2013). We attempted Q30 trimming both before and after merging and found spurious reads when it was performed after merging. Trimming before merging PE reads, however, eliminated OTUs associated with *Skeletonema sp.*, *Skeletonema potamos* (a freshwater, not marine, species) (Duleba et al., 2014), and unidentified species. Since bps reflective of genetic variation mostly occurred in the middle of the sequence where forward and reverse reads overlapped, it is possible that Q30 trimming before merging reads eliminated low quality bps that were otherwise being included in the merged sequences. We found that developing a custom method of DNA sequence processing, tested on communities of known *Skeletonema* spp. abundances, provided us with more accurate and reliable data.

Accuracy of relative and absolute abundance sequence data

Simulated communities revealed that *Skeletonema* species can successfully be detected and distinguished in mixed phytoplankton communities; however sequence read abundances did not match expected abundances when cell concentrations of *Skeletonema* spp. were known (Figure 4). For example, abundances for *S. japonicum* were overestimated while *S. marinoi* and *S. grethae* were underestimated. A likely cause for these biases is variation in the LSU rDNA copy number. Theoretically, a species with a greater number of target gene copies will generate a greater number of sequence reads, producing false representation of that species, and vice versa. Unique *Skeletonema* species were only discovered within the past couple of decades, but it is possible that species have been diverged long enough to develop significantly different numbers of copies of the LSU rDNA gene. A study by Zhu et al. (2005) determined that 18S copy number can vary from 1 to 12,000 for a variety of algal strains, including dinoflagellates and picoplankton. More specifically, within the genus *Thalassiosira* copy number can vary by 10-fold among species (Zhu et al., 2005). Little work has been done to investigate the potential for variable copy number among *Skeletonema* species, although LSU copy number in *S. marinoi* has been estimated to be 61 copies (Ellegaard et al., 2008). The straightforward method for quantifying copy number per cell in these studies was quantitative PCR, and this technique could be applied to *Skeletonema* species to verify the over- and underestimation of cell abundance based on LSU copy number.

Seasonal separation of Skeletonema *communities based on environmental conditions*

Studies of *Skeletonema* community composition have only recently begun to include distinctions of morphologically cryptic species (Yamada et al., 2014). Here,

seven species of *Skeletonema* were detected in Narragansett Bay, five of which had previously never been detected in this area. Previous studies have confirmed the presence of *S. grethae* and *S. japonicum* in Narragansett Bay (Kooistra et al., 2008). In this study, *S. costatum*, *S. dohrnii*, *S. marinoi*, *S. menzelii*, and *S. pseudocostatum* were detected for the first time in Narragansett Bay. Some of these species have been found in nearby estuaries, such as *S. marinoi* in Long Island Sound and *S. menzelii* off the coast of Cape Cod (Kooistra et al., 2008). The results suggest that diversity within estuarine phytoplankton communities may be greater than previously thought.

Communities were found to be significantly different when comparing seasons, except when comparing winter and spring communities and summer and autumn communities. The most noticeable difference between these seasonal communities was the low species richness during winter and spring when *S. marinoi* was the primary species in Narragansett Bay compared to the high species richness and more unpredictable community composition during the summer. This phenomenon was also noted during several bloom periods that were monitored by sampling each week for the duration of a bloom. A winter bloom from December 2008 to January 2009 indicated no significant change in community composition over a period of eight weeks, and a spring bloom from March 2010 to April 2010 also showed consistent community composition over a period of five weeks. A summer bloom between May 2013 and July 2013, however, showed distinct changes over time in the composition of the bloom community. *S. japonicum*, *S. marinoi*, and *S. menzelii* comprised the early stages of the bloom, but composition shifted to a mix of *S. dohrnii*, *S. japonicum*, *S. menzelii*, and *S. pseudocostatum* during the third week of June. This suggests that summer conditions

initiated a transition in the *Skeletonema* community that occurred rapidly within a single bloom period, while winter and spring community composition remained consistent for even up to several months.

A similar separation of communities within a single location was found for morphologically cryptic species of *Pseudo-nitzschia* in Puget Sound. Initially, at least 16 species of *Pseudo-nitzschia* were found using another molecular method of automated ribosomal intergenic spacer analysis (ARISA) (Hubbard et al., 2008). Further investigation revealed that these species are spatially separated, even over short distances within a single estuary. While *P. pungens* and *P. delicatissima* were widely spread throughout Puget Sound, others were patchy and only found in certain locations. The community composition at each of these locations was strongly associated with environmental conditions, especially temperature, salinity, and ammonium concentration (Hubbard et al., 2014). The effect of cryptic species on community composition along an estuarine gradient is similar to our observed shift in community composition over time relative to seasonal conditions. Furthermore, these studies demonstrate that molecular methods, whether sequencing or ARISA, are useful tools for investigating cryptic species in marine field samples.

Effect of physiology and environmental conditions on community composition

In addition to the phylogenetic separation of cryptic species, studies have also found that physiology can differ significantly among cryptic species, including *Skeletonema.* In a study by Gallagher (1980), seasonal field samples from Narragansett Bay exhibited significantly different genotypic frequencies. Within summer blooms community compositions were 96% similar, and within winter blooms community

compositions were 97% similar. Between summer and winter blooms, however, communities were only 27% similar. Further study revealed that isolates of *S. costatum* from Narragansett Bay exhibited significantly different growth rates and chlorophyll concentration per cell during different seasons (Gallagher, 1982). In these studies, differences were attributed to diversity among clonal lineages prior to the knowledge of cryptic species. Now that we have detected at least seven species of *Skeletonema* in Narragansett Bay, these seasonal differences could more likely be attributed to species diversity and the unique genetic and physiological characteristics associated with each species.

Different physiologies are a likely cause for shifting community composition across seasons. In a study comparing growth rates of *Skeletonema* species and temperature, cultures of *S. marinoi* grew faster than all other species at the lowest tested temperatures (10°C and 15°C) and displayed a specific growth rate of 1.0-1.6 d^{-1} between temperatures of 10°C and 30°C (Kaeriyama et al., 2011). This could explain the dominance of *S. marinoi* in winter samples. Other studies of *Skeletonema* physiology have focused on seasonally separated clonal lineages, such as those of *S. marinoi* off the coast of Sweden (Saravanan and Godhe, 2010), and significant differences in growth rate and abundance were detected. If significant differences exist among population lineages, it is likely that physiologies differ even more among species of *Skeletonema*.

Temperature had by far the greatest effect on *Skeletonema* community composition. We discovered that temperature alone is the sole environmental condition that significantly correlated with diversity. Another study of plankton community composition in a temperate lake found that species richness and diversity were highest in

summer, coinciding with higher temperatures (Graham et al., 2004). This phenomenon is apparent not only in small lakes and estuaries but also globally. A study by Thomas et al. (2012) measured growth rates of 130 species of phytoplankton under varying temperatures and found that growth rates of individual species are strongly dependent on optimum temperature niches. Rapid ocean warming could alter species ranges and diversity within a community by challenging these temperature niches, but high genetic diversity within a genus can help prevent loss of ecologically important groups of organisms like *Skeletonema.*

On a broader scale, marine planktonic organisms occupy a latitudinal gradient that relies heavily on temperature and other environmental variables. Diversity of bacterial communities decreases with increasing latitude and strongly correlates with temperature (Fuhrman et al., 2008). In a modeled global ocean, phytoplankton experienced the same phenomenon of decreasing diversity with increasing latitude that may be dependent on temperature gradients (Barton et al., 2010). Temperature is a possible driver of diversity for marine microbes, where high temperatures result in more diverse communities. Furthermore, transport from the Gulf Stream into Narragansett Bay has been found to affect bloom behavior of *Skeletonema* and could allow tropical and sub-tropical species that are more adapted to warm temperatures to enter the Bay and occupy seasonal niches (Borkman and Smayda, 2009). This could provide an opportunity for a species like *S. tropicum* to enter the bay. At this time it is unclear if *S. tropicum* is actually excluded from the Narragansett Bay *Skeletonema* community due to the lack of resolution between LSU sequences of *S. grethae* and *S. tropicum*. *S. tropicum* has been found in marine and brackish environments mostly clustered near low latitudes but extends as far south as the

coastal waters of Uruguay in South America (Kooistra et al., 2008). With an optimum growth rate occurring at 25°C (Kaeriyama et al., 2011), *S. tropicum* could thrive in Narragansett Bay during the warm summer months, but thus far does not appear to do so. Further clarification of the phylogenetic and physiologic differences between *S. grethae* and *S. tropicum* can help distinguish between species and their potential roles in the Narragansett Bay ecosystem.

Composition of winter communities and implications for future changes in community composition

We used the results from simulated communities of known *Skeletonema* spp. proportions to calculate absolute cell abundances of those species during the winter. Based on these calculations, we confirmed that *S. marinoi* is a dominant winter species and may also dominate other seasons, since simulated communities suggest that *S. marinoi* is underrepresented when field samples are sequenced. The discovery of seasonally separated communities of *Skeletonema* and particularly winter communities that are dominated by a single species is one that could have implications for future changes to phytoplankton communities. Since temperature may partially affect the physiology of *Skeletonema,* changes in ocean temperature could result in a shift in growth of cryptic species and community composition. *S. marinoi* is an abundant primary producer during the winter and plays a significant ecological role that can be affected by such changes in the environment, so further investigation of how this particular species responds to environmental change is crucial.

Conclusions

Skeletonema is an ecologically important and cosmopolitan diatom genus that consists of several morphologically cryptic species. In this study, we used molecular tools to investigate how *Skeletonema* communities vary in Narragansett Bay due to the presence of cryptic species and what environmental factors may influence We conclude the following: 1) In this study, *Skeletonema* community composition was investigated for the first time over a span of several years in an estuarine environment. Measures of experimental error suggest that careful processing of samples and data are necessary and that recognizing these errors is critical to interpreting molecular data. 2) Sequencing data tends to misrepresent actual cell abundances of species, and measurement of LSU rDNA copy number or the construction of simulated communities is necessary to convert relative changes in composition to absolute changes. 3) At least seven species of *Skeletonema* reside in Narragansett Bay, five of which were not known to exist there previously. Species inhabit seasonal niches, with *S. marinoi* occurring primarily in the winter and spring and other species occurring in the summer and fall. The seasonal change in diversity may result from varying physiologies. 4) Seasonal distributions are primarily controlled by temperature. 5) *S. marinoi* is the dominant species in the winter, and perhaps throughout the year. Its strong presence in the wintertime, compounded with a changing environment, could result in significant shifts in the *Skeletonema* community.

This study of *Skeletonema* community composition addresses the recently discovered cryptic species and how they are influenced by their environment. Future work can expand to include even longer time series and physiological comparisons of species isolated from Narragansett Bay. Our methods and results can also serve as a

guide for researchers in other parts of the globe who wish to investigate *Skeletonema* community composition. Further research into *Skeletonema* community dynamics over time on both local and global scales will ultimately help researchers understand and project future changes to marine phytoplankton communities.

TABLES

Table 1. Field samples collected from Narragansett Bay between December 2008 and December 2013. Samples were chosen based on the highest percentage *Skeletonema* in the phytoplankton community within a month. Some months are represented by multiple samples, including December 2008-January 2009, March 2010-April 2010, and May 2013-July 2013. Samples were not collected in July 2011 and between February-April 2012. Areas within the table marked by a '-' did not have associated abundance data.

Date collected	Total phytoplankton abundance (cells L^{-1})	*Skeletonema* abundance (cells L^{-1})	Percent *Skeletonema*
12/09/08	1866000	1690000	90.57%
12/17/08	841000	796000	94.65%
12/22/08	125000	85000	68.00%
12/30/08	43000	4000	9.30%
01/13/09	137000	114000	83.21%
01/20/09	173000	125000	72.25%
01/27/09	242000	164000	67.77%
02/02/09	207000	24000	11.59%
02/09/09	110000	16000	14.55%
03/03/09	921000	11000	1.19%
04/13/09	5147000	3000	0.06%
05/27/09	1153000	9000	0.78%
06/23/09	4680000	4088000	87.35%
07/27/09	10784000	9408000	87.24%
08/10/09	25849000	20074000	77.66%
09/04/09	17383000	15977000	91.91%
10/08/09	541000	20000	3.70%
11/23/09	438000	195000	44.52%
12/31/09	728000	577000	79.26%
01/26/10	19612000	19346000	98.64%
02/02/10	10633000	10308000	96.94%
03/29/10	6085000	252000	4.14%
04/05/10	4516000	3596000	79.63%
04/13/10	49346000	48221000	97.72%
04/20/10	907000	828000	91.29%
04/27/10	172000	50000	29.07%
05/28/10	4942000	4450000	90.04%
06/15/10	12709000	11637000	91.57%
07/19/10	18144000	14466000	79.73%
08/20/10	8015000	5334000	66.55%

09/15/10	599000	7000	1.17%
10/05/10	2101000	299000	14.23%
11/16/10	1474000	83000	5.63%
12/30/10	1179000	413000	35.03%
01/11/11	4991000	2434000	48.77%
02/01/11	3539000	2087000	58.97%
03/09/11	2840000	47000	1.65%
04/01/11	1253000	83000	6.62%
04/20/11	14207667	4820000	33.93%
05/31/11	4361000	1405000	32.22%
06/06/11	19456000	15670000	80.54%
06/13/11	3527000	165000	4.68%
08/19/11	1861000	169000	9.08%
09/06/11	3590000	2772000	77.21%
10/03/11	2460000	1932000	78.54%
11/07/11	454000	160000	35.24%
12/07/11	165000	93000	56.36%
12/28/11	68000	0	0.00%
01/04/12	699000	539000	77.11%
05/23/12	-	-	-
06/25/12	-	-	-
07/02/12	-	-	-
08/01/12	-	-	-
09/05/12	-	-	-
10/05/12	-	-	-
11/02/12	-	-	-
01/02/13	45200	1100	2.43%
02/12/13	8272800	7870000	95.13%
03/12/13	-	-	-
04/02/13	807800	535000	66.23%
05/21/13	192300	8300	4.32%
05/28/13	478000	161000	33.68%
06/06/13	6374300	2390000	37.49%
06/11/13	22438200	740000	3.30%
06/19/13	15745200	8110000	51.51%
06/24/13	7044600	6270000	89.00%
07/03/13	-	-	-
07/08/13	12231400	10200000	83.39%
07/15/13	48604100	47640000	98.02%
07/22/13	352400	81000	22.99%
08/16/13	2769200	1455000	52.54%

09/09/13	5512100	3375000	61.23%
10/28/13	-	-	-
11/20/13	-	-	-
12/02/13	-	-	-

Table 2. Inter- and intraspecific variation within the targeted 325 bp fragment of the 28S gene in *Skeletonema* species. A dot indicates that the base pair is identical to the reference sequence and a slash indicates a deletion. Sequence numbers can be cross-referenced with those in Table 4 for the associated accession numbers.

Species	Sequence number	35	45	49	51	57	64	65	75	76	81	113	115	117	118	119
Skeletonema dohrnii (reference)	1-2,9	G	G	A	C	A	A	G	C	T	C	C	A	T	T	C
Skeletonema dohrnii	3-4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Skeletonema marinoi	5-8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Skeletonema costatum	10-12	.	.	.	.	T	.	.	.	C	T	G	C	G	C	A
Skeletonema costatum	13	.	.	.	.	T	.	.	.	C	T	G	C	G	C	A
Skeletonema potamos	14-15	.	.	.	.	T	.	.	.	C	T	G	C	G	C	A
Skeletonema japonicum	16-20	.	.	T	.	.	C	.	T	C	.	.	.	.	.	.
Skeletonema grethae	21-23	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.
Skeletonema menzelii	24	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.
Skeletonema menzelii	25	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.
Skeletonema menzelii	26-28	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.
Skeletonema pseudocostatum	29-33	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.
Skeletonema tropicum	34	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.
Skeletonema tropicum	35-37	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.
Skeletonema tropicum	38	.	.	T	.	.	.	.	T	C	.	.	.	.	.	.
Skeletonema grevillei	39	T	T	.	A	.	C	.	.	.	.	.	C	.	G	.
Skeletonema grevillei	40	T	T	.	A	.	C	.	.	.	.	.	C	.	G	.
Skeletonema grevillei	41	T	T	.	A	.	C	.	.	.	.	.	C	.	G	.
Skeletonema grevillei	42	T	T	.	A	.	C	.	.	.	.	.	C	.	G	.
Skeletonema grevillei	43	T	T	.	A	.	C	.	.	.	.	.	C	.	G	.
Skeletonema ardens	44-48	.	T	.	A	.	.	T	.	.	.	.	C	.	G	.

Species	Sequence number	121	124	128	129	130	135	136	138	144	177	178	180	186	188
Skeletonema dohrnii (reference)	1-2,9	G	C	A	T	C	G	C	A	T	A	C	T	A	C
Skeletonema dohrnii	3-4	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Skeletonema marinoi	5-8	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Skeletonema costatum	10-12	T	.	.	.	.	.	G	.	.	.	.	C	.	.
Skeletonema costatum	13	T	.	G	.	.	.	G	.	.	.	.	C	.	.
Skeletonema potamos	14-15	T	.	.	C	T	.	.	T	C	.	.	C	.	.
Skeletonema japonicum	16-20	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Skeletonema grethae	21-23	.	.	.	C	.	.	.	.	C	.	T	.	.	.
Skeletonema menzelii	24	.	.	.	C	.	.	.	.	.	.	.	C	.	.
Skeletonema menzelii	25	.	.	.	C	T	.	.	.	.	.	.	.	G	T
Skeletonema menzelii	26-28	.	.	.	C	.	A	.	.	.	G	.	C	G	T
Skeletonema pseudocostatum	29-33	.	.	.	C	.	.	.	T	C	G	T	.	.	.
Skeletonema tropicum	34	.	.	.	C	.	.	.	.	.	R	T	.	.	.
Skeletonema tropicum	35-37	.	.	.	C	.	.	.	.	.	.	T	.	.	.
Skeletonema tropicum	38	.	.	.	C	.	.	.	.	C	.	T	.	.	.
Skeletonema grevillei	39	.	T	.	C	.	.	.	.	.	.	T	A	G	T
Skeletonema grevillei	40	.	T	.	C	.	.	.	.	.	.	T	A	G	T
Skeletonema grevillei	41	.	T	.	C	.	.	.	.	.	.	T	A	G	T
Skeletonema grevillei	42	.	T	.	C	Y	.	.	.	.	.	T	A	G	T
Skeletonema grevillei	43	.	T	.	Y	.	.	.	.	.	.	T	A	G	T
Skeletonema ardens	44-48	.	.	.	C	T	.	.	.	.	G	.	.	.	.

Species	Sequence number	193	195	198	228	229	230	231	232	234	237	239	241	245
Skeletonema dohrnii (reference)	1-2,9	G	C	G	A	A	T	A	A	A	A	G	T	C
Skeletonema dohrnii	3-4	.	.	.	.	.	.	.	.	.	.	.	.	.
Skeletonema marinoi	5-8	.	.	.	.	.	.	.	.	.	.	.	.	.
Skeletonema costatum	10-12	A	.	.	C	T	G	.	C	.	/	.	.	.
Skeletonema costatum	13	A	.	.	C	T	G	.	C	.	/	.	.	.
Skeletonema potamos	14-15	.	T	.	C	T	G	.	c	.	/	.	.	.
Skeletonema japonicum	16-20	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema grethae	21-23	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema menzelii	24	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema menzelii	25	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema menzelii	26-28	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema pseudocostatum	29-33	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema tropicum	34	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema tropicum	35-37	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema tropicum	38	.	.	.	.	.	.	.	.	G	.	.	A	.
Skeletonema grevillei	39	.	.	A	C	.	.	G	.	G	.	T	.	A
Skeletonema grevillei	40	.	.	A	C	.	.	G	.	G	.	T	.	A
Skeletonema grevillei	41	.	.	A	C	.	.	G	.	G	.	T	.	A
Skeletonema grevillei	42	.	.	A	C	.	.	G	.	G	.	T	.	A
Skeletonema grevillei	43	.	.	A	C	.	.	G	.	G	.	T	.	A
Skeletonema ardens	44-48	.	.	.	C	.	.	G	.	G	.	T	.	A

Table 3. *Skeletonema* genus-specific primers and Illumina adapters used in PCR reactions to target all *Skeletonema* species in the mixed community and field samples. Adapters for both the forward and reverse primers were added onto the 5' end of the primer.

Forward primer	5'-GTTAAAGAGTACCTGAAATTGTTAA-3'
Adapter 1	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'
Adapter 2	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGT-3'
Adapter 3	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTC-3'
Adapter 4	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGA-3'
Reverse primer	5'-TGTTACTTTCATTACGCATATCA-3'
Adapter 1	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'
Adapter 2	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGC-3'
Adapter 3	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCA-3'
Adapter 4	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACT-3'

Table 4. Complete list of sequences included in the *Skeletonema* 28S LSU reference database, obtained from the NCBI Nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore). Sequences are numbered for cross-referencing with Table 1. Search terms in NCBI included "*Skeletonema* 28S" and "*Skeletonema* LSU". Sequences that were previously published (with the exception of *Skeletonema grevillei*) were accepted as good candidates for the database, downloaded, and aligned in CLC Genomics Workbench.

Species	Sequence number	Accession number	Published by
Skeletonema dohrnii	1	AB572711	Yamada et al., 2010
	2	AB572710	Yamada et al., 2010
	3	AB572708	Yamada et al., 2010
	4	AJ633538	Sarno et al., 2005
	5	AJ633537	Sarno et al., 2005
Skeletonema marinoi	6	FR823449	Kremp et al., 2012
	7	FR823448	Kremp et al., 2012
	8	AJ633535	Sarno et al., 2005
	9	AJ633533	Sarno et al., 2005
Skeletonema costatum	10	AB728776	Yamada et al., 2013
	11	AB728769	Yamada et al., 2013
	12	AB728755	Yamada et al., 2013
	13	AB728767	Yamada et al., 2013
Skeletonema potamos	14	KF621300	Duleba et al., 2014
	15	KF621299	Duleba et al., 2014
Skeletonema japonicum	16	AJ633524	Sarno et al., 2005
	17	AB572714	Yamada et al., 2010
	18	AB572736	Yamada et al., 2010
	19	AB572757	Yamada et al., 2010
	20	AB572793	Yamada et al., 2010
Skeletonema grethae	21	AJ633523	Sarno et al., 2005
	22	AJ633522	Sarno et al., 2005
	23	AJ633521	Sarno et al., 2005
Skeletonema menzelii	24	AJ633528	Sarno et al., 2005
	25	AJ633527	Sarno et al., 2005
	26	AJ633526	Sarno et al., 2005
	27	AB573023	Yamada et al., 2010
	28	AB555561	Yamada et al., 2010

Skeletonema pseudocostatum	29	AJ633507	Sarno et al., 2005
	30	AJ633510	Sarno et al., 2005
	31	AJ633512	Sarno et al., 2005
	32	AJ633514	Sarno et al., 2005
	33	AB572831	Yamada et al., 2010
Skeletonema tropicum	34	AJ633519	Sarno et al., 2005
	35	AJ633516	Sarno et al., 2005
	36	AB572812	Yamada et al., 2010
	37	AB572808	Yamada et al., 2010
	38	AB572825	Yamada et al., 2010
Skeletonema ardens	39	AB572616	Yamada et al., 2010
	40	AB572613	Yamada et al., 2010
	41	AB572612	Yamada et al., 2010
	42	AB572610	Yamada et al., 2010
	43	AB572609	Yamada et al., 2010
Skeletonema grevillei	44	AB728798	Unpublished
	45	AB728795	Unpublished
	46	AB728792	Unpublished
	47	AB728787	Unpublished
	48	AB728783	Unpublished

Table 5. Using data from the simulated field communities, transformation coefficients were determined to convert relative percent composition data into more accurate absolute abundance data. A single sample is represented below to demonstrate how coefficients produce more reliable sequence read data. Transformation coefficients were multiplied by the number of sequence reads (before transformation) to obtain a more accurate number of sequence reads (after transformation). The target for expected number of sequence reads per species for this sample was 3665 and percent composition was 20%.

Species name	Transformation coefficient	Number of sequence reads (before transformation)	Percent composition (before transformation)	Number of sequence reads (after transformation)	Percent composition (after transformation)
Skeletonema japonicum	0.465	7834	43%	3647	20%
Skeletonema marinoi	2.56	1496	8%	3833	21%
Skeletonema dohrnii	0.773	5201	28%	4019	22%
Skeletonema grethae	8.71	399	2%	3478	19%
Skeletonema menzelii	1.08	3395	19%	3660	20%
Total number of sequences		18325		18637	

Table 6. Percent composition and standard deviation of species for five filtered field samples. Technical replication of the method was tested at three different steps: DNA extraction, amplification, and sequencing. Three sets of technical triplicates were run for DNA extraction by extracting from triplicate filters of simulated communities, where "Simulated community 1" represents 20% *Skeletonema*, "Simulated community 2" represents 50% *Skeletonema*, and "Simulated community 3" represents 80% *Skeletonema*. One set of technical triplicates was run for DNA amplification (June 2010) and sequencing (September 2009) by setting up three separate reactions for a single extraction. Values in bold represent greater than 15% coefficient of variation among triplicate measurements.

Species	Extraction triplicates (Simulated community 1)	Extraction triplicates (Simulated community 2)	Extraction triplicates (Simulated community 3)	Amplification triplicates (June 2010)	Sequencing triplicates (September 2009)
Skeletonema costatum	-	-	-	0.00%±0.00%	0.00%±0.00%
Skeletonema dohrnii	25.25%±3.40%	27.05%±2.39%	**27.70%±8.02%**	**1.32%±0.29%**	38.95%±0.53%
Skeletonema grethae	2.23%±0.29%	2.26%±0.29%	2.48%±0.34%	12.09%±0.05%	57.67%±0.52%
Skeletonema japonicum	45.61%±4.16%	38.31%±2.10%	46.69%±5.83%	**0.10%±0.10%**	0.29%±0.04%
Skeletonema marinoi	7.83%±0.50%	8.06%±0.70%	7.65%±0.66%	7.54%±0.25%	0.76%±0.11%
Skeletonema menzelii	19.08%±1.27%	**24.33%±5.41%**	**15.48%±3.15%**	3.71%±0.49%	**0.39%±0.08%**
Skeletonema pseudocostatum	-	-	-	75.24%±0.92%	1.93%±0.28%

Table 7. Pairwise similarity among *Skeletonema* species composition between seasons (ANOSIM). Global R was equal to 0.386 and significant comparisons are denoted in bold. Community compositions were significantly different (p=0.001) except between winter-spring and summer-autumn.

Groups	R statistic	Significance Level (%)
Winter, spring	0.009	26.4
Winter, summer	**0.581**	**0.1**
Winter, autumn	**0.578**	**0.1**
Spring, summer	**0.591**	**0.1**
Spring, autumn	**0.533**	**0.1**
Summer, autumn	0.023	24.7

Table 8. Correlation of Narragansett Bay environmental conditions to *Skeletonema* species composition. Six environmental variables were included in the BIOENV analysis: surface temperature (°C), surface salinity (psu), average PAR (mmol/m^2), DIN (μM), DIP (μM), and Si (μM). Surface temperature and DIP combined explained 48% of the variability in community structure, although temperature alone explained 46% of that variability. The global BEST match permutation test has a ρ=0.480 and p=0.01.

Number of variables	Correlation	Variables
2	0.480	Surface temperature, DIP
1	0.462	Surface temperature
3	0.422	Surface temperature, average PAR, DIP
2	0.412	Surface temperature, Si
3	0.405	Surface temperature, DIP, Si
3	0.393	Surface temperature, surface salinity, DIP

Table 9. When temperature was removed from the BIOENV analysis, DIP is most highly correlated with community structure. Alone, DIP explains 21.9% of variability in the data. Combined with average daily PAR, it explains 25.0%. The global BEST match permutation test has a ρ=0.25 and p=0.01.

Number of variables	Correlation	Variables
2	0.250	Average PAR, DIP
3	0.236	Average PAR, DIP, Si
1	0.219	DIP
4	0.201	Average PAR, DIN, DIP, Si
3	0.199	Surface salinity, average PAR, DIP
2	0.194	DIP, Si

FIGURES

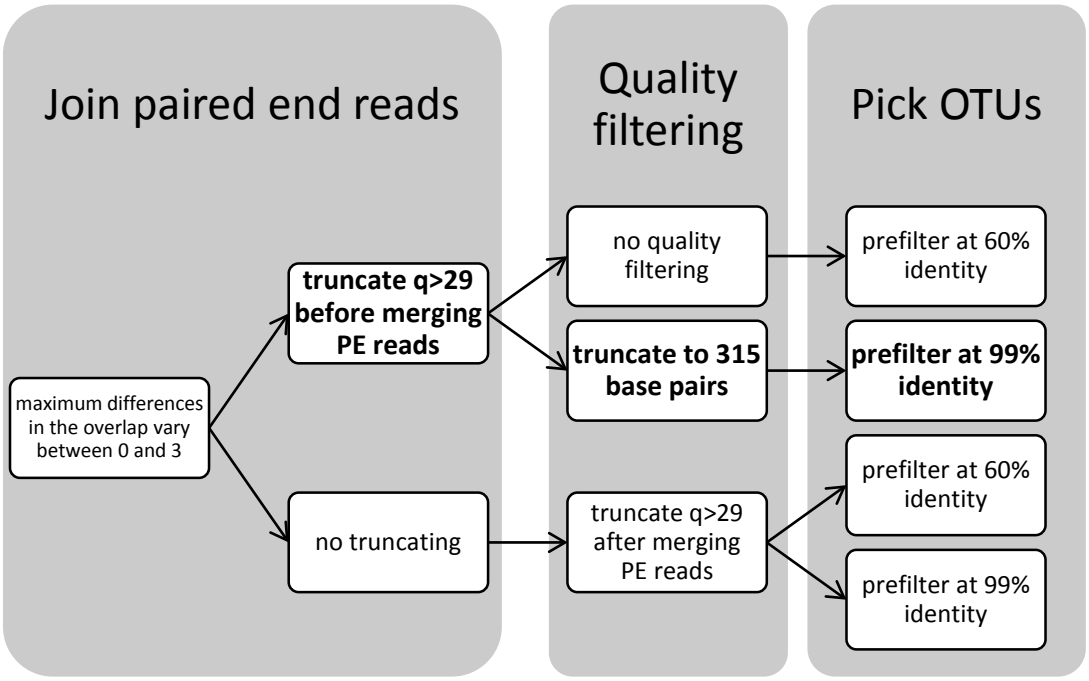

Figure 1. Data processing consisted of three main steps—merging paired end reads, quality filtering, and picking OTUs. Due to the close relationship among *Skeletonema* species and the possibility for error in downstream sequence processing, we tested various options within these three steps to ensure optimal sequence identification using simulated communities of known *Skeletonema* species concentrations. Steps highlighted in bold represent the final workflow selected to analyze the full data set. In the first step, we determined that a maximum difference of one bp in the overlapping region was sufficient.

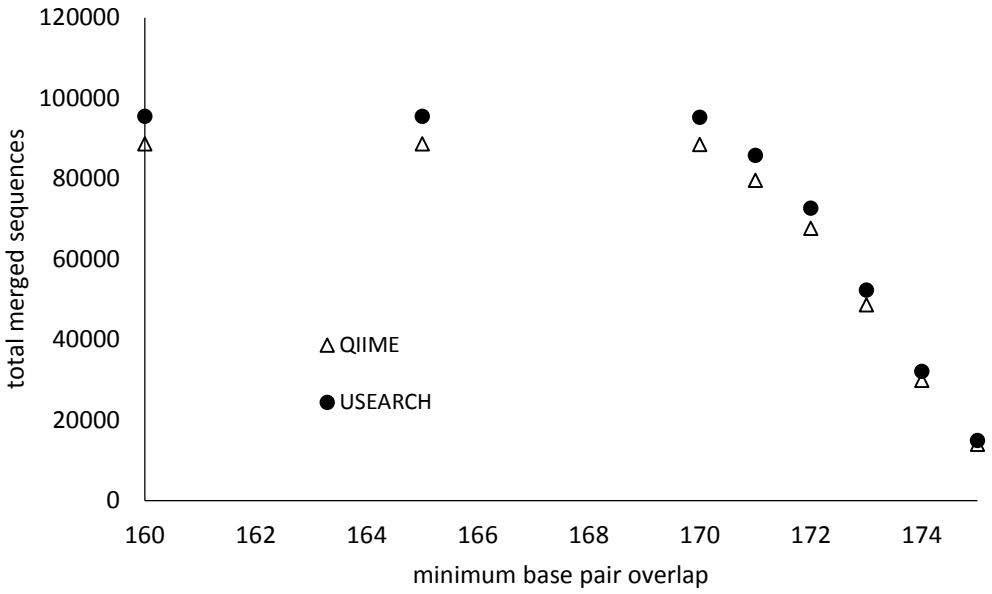

Figure 2. Two algorithms for merging PE sequence data were compared for a single sample: QIIME's join_paired_ends.py, represented by the open triangles, and USEARCH's –fastq_mergepairs, represented by the closed circles. Each algorithm was also tested with a minimum base pair overlap quality filter varying from 160 base pairs to 175 base pairs (the maximum possible overlap). The total number of merged sequences was similar between the algorithms, but the USEARCH algorithm yielded a maximum of 95449 merged reads while the QIIME algorithm yielded a maximum of only 88619 merged reads. The USEARCH algorithm was selected as the preferred method to merge PE reads.

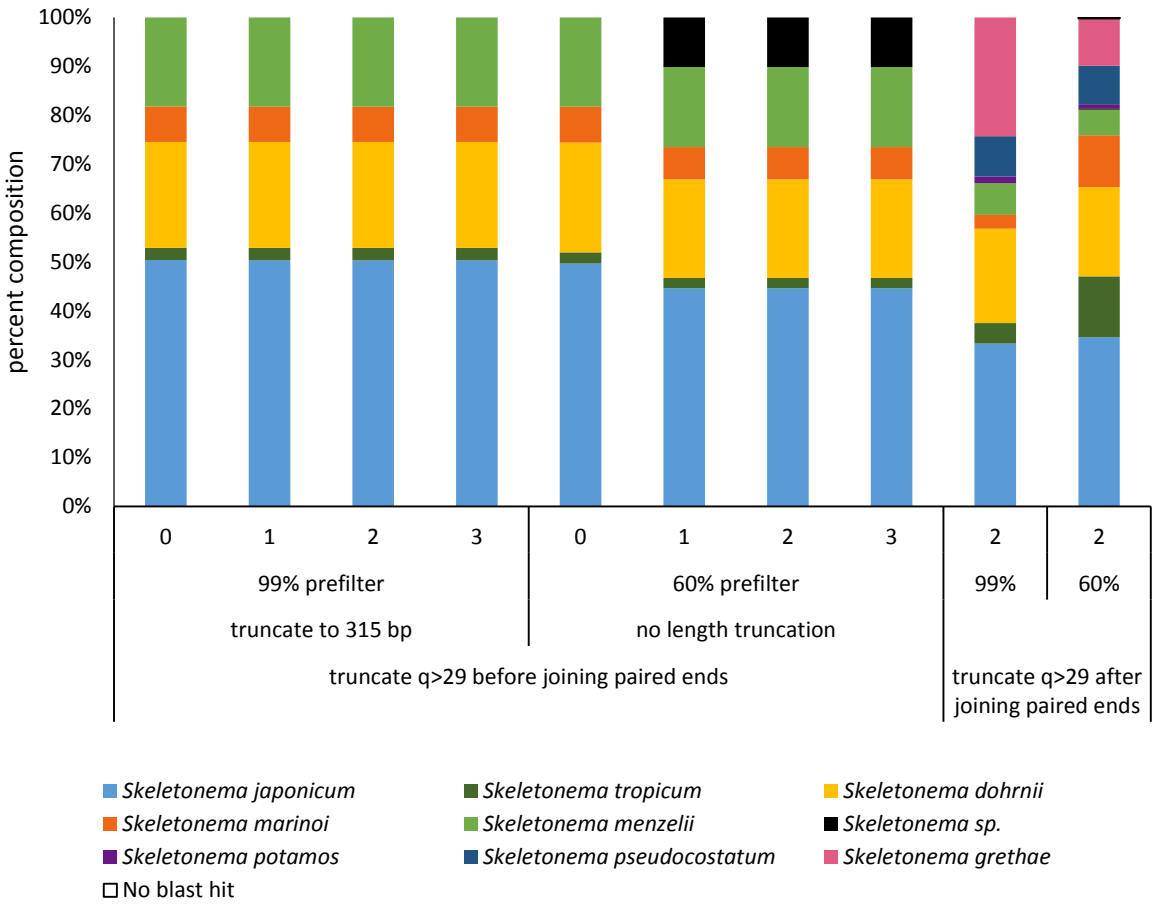

Figure 3. An optimal sequence analysis workflow was determined by testing a variety of parameters using data from a simulated *Skeletonema* community of known species composition. The first row of the x-axis denotes the number of mismatches allowed in the overlapping sequence, ranging from zero to three mismatches. The second row denotes if percent identity was pre-filtered at 60% (the default value) or 99%. The third row denotes whether or not the merged reads were truncated to 315 base pairs. The fourth row denotes whether truncation up to the first Q30 bp occurred before or after PE reads were joined. The second bar from the left denotes the combination of quality filters used in the final workflow.

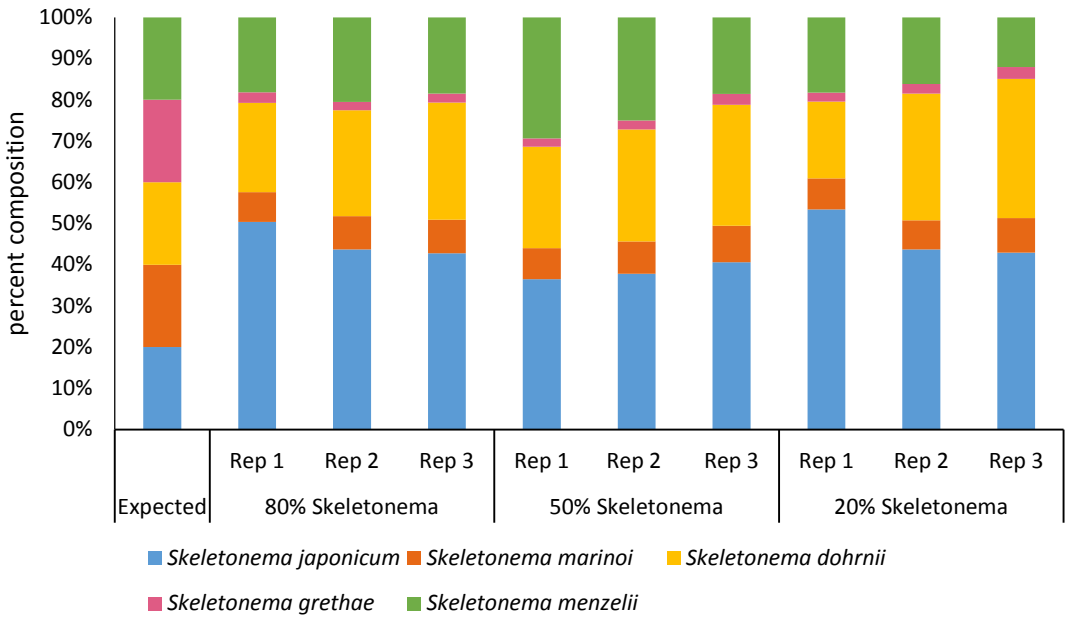

Figure 4. Percent composition of *Skeletonema* spp sequences recovered from three simulated phytoplankton communities generated to test the sensitivity of MiSeq sequencing with low-diversity sequences. The leftmost bar represents expected percent composition for all samples. Triplicate simulated communities of 80%, 50%, and 20% *Skeletonema* are represented by the remaining bars. Within the *Skeletonema* portion of the community, five different *Skeletonema* species were added in equal proportions. The remainder of each community contained other diatoms and dinoflagellates. Each experiment used the same total number of 4 x 10^6 cells.

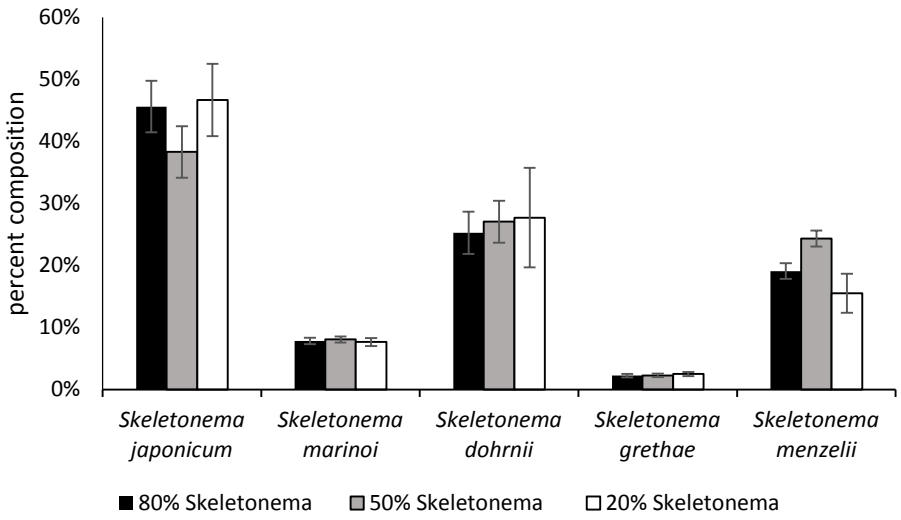

Figure 5. Simulated field communities of varying *Skeletonema* concentration produced unexpected results. Each of 5 species was added in equal proportions to make up 1/5 of the total *Skeletonema* community, but sequencing and downstream processing led to over- or underrepresentation of some species. However, triplicate measurements were precise and representation of species was consistent regardless of concentration of *Skeletonema* in the simulated community. This suggests that, no matter how abundant *Skeletonema* cells are in a mixed phytoplankton community, sequencing of environmental DNA can consistently detect present species in at least a relative manner. Error bars represent standard deviation.

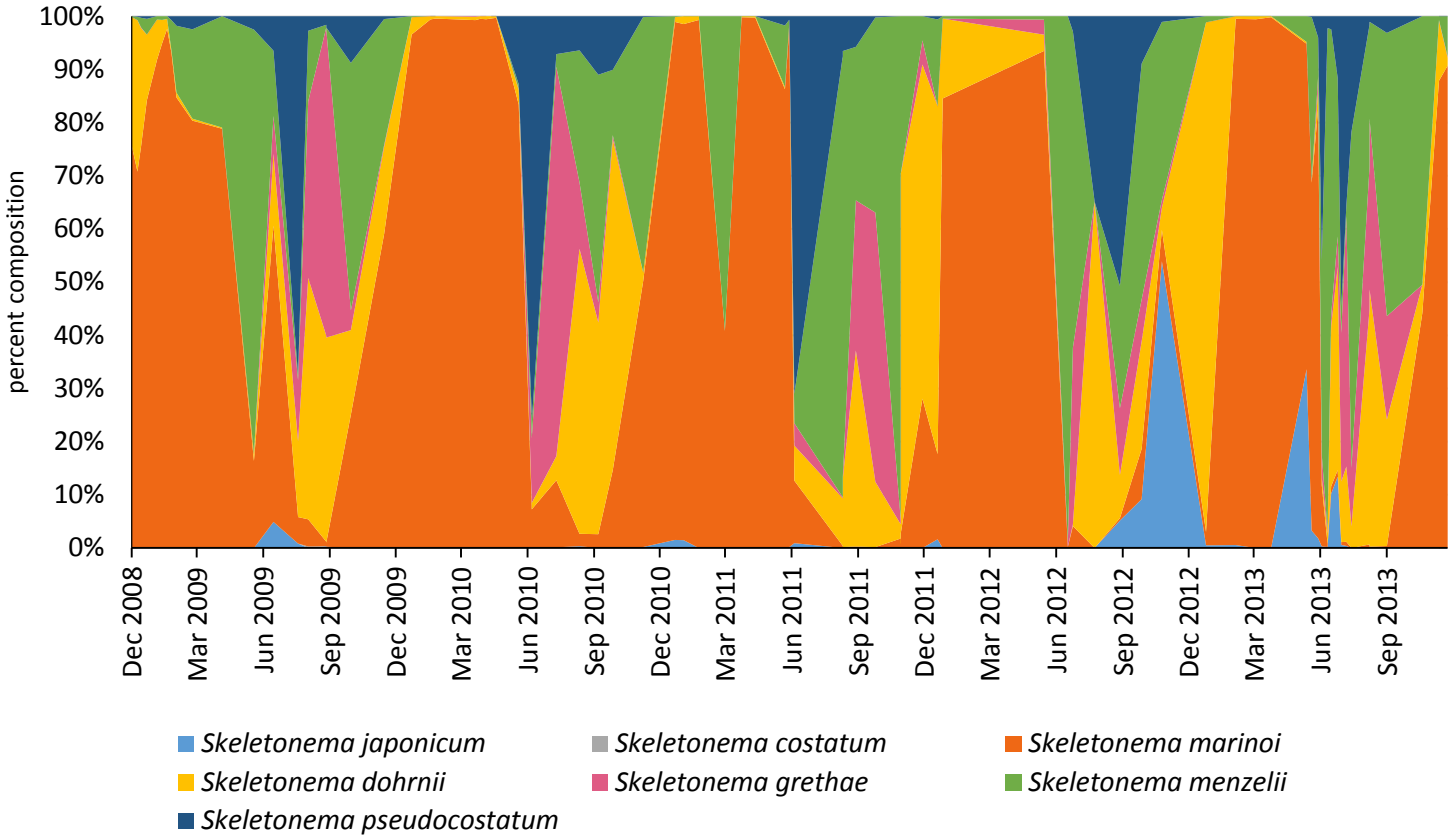

Figure 6. Percent composition of sequence reads of *Skeletonema* species from December 2008 to December 2013 in Narragansett Bay. The y-axis represents percent composition, with the individual contributions of each species accumulating to 100% for each sample. Data includes all monthly samples as well as samples taken over several weeks, which occurred during December 2008-January 2009, March 2010-April 2010, and May 2013-July 2013. Samples were not collected in July 2011 and between February-April 2012. A general pattern appears with fewer species occurring in the winter and spring and more species occurring in the summer and autumn, and the community structure transitions rapidly between seasons. *S. marinoi* appears to be the primary species in the bay during winter and spring.

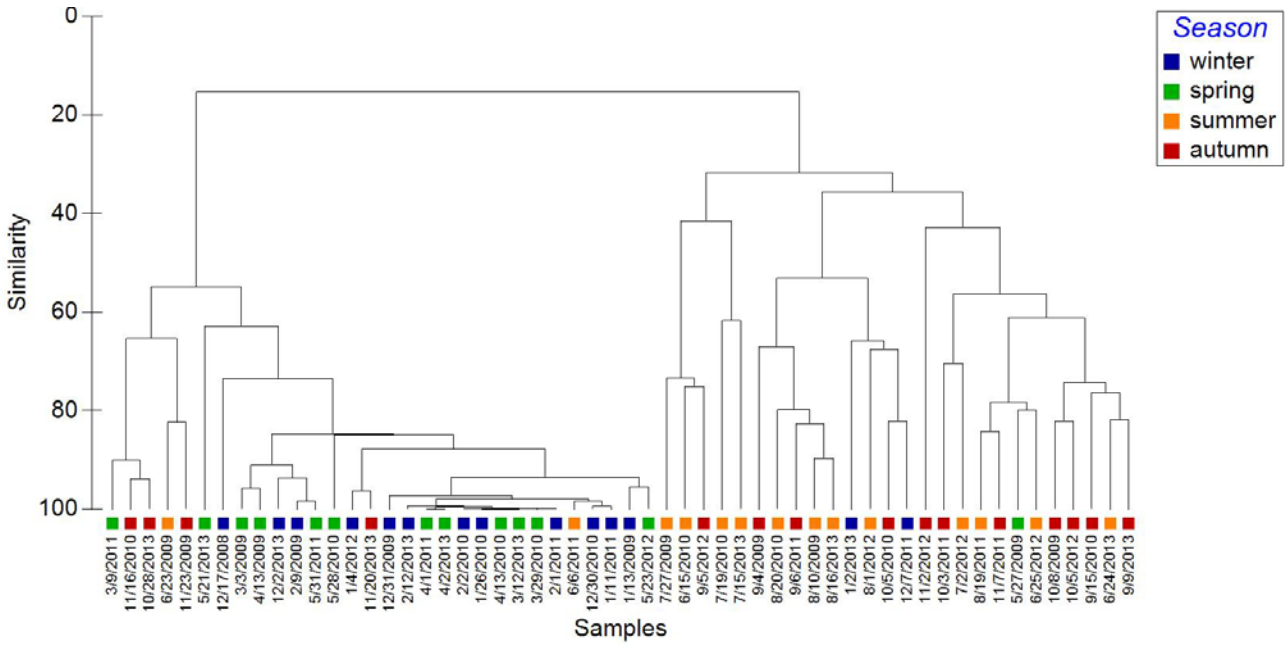

Figure 7. Field samples were grouped based on Bray-Curtis similarity and categorized by season (winter=blue, spring=green, summer=orange, autumn=red). Samples are represented by boxes on the x-axis and percent similarity is represented by the y-axis. Similarity is measured from nodes joining two samples or two groups of samples, and measurement from the node to the bottom of the dendrogram such that shorter branch lengths represent more similar samples (closer to 100% similarity) and longer branch lengths represent more dissimilar samples (further from 100% similarity).

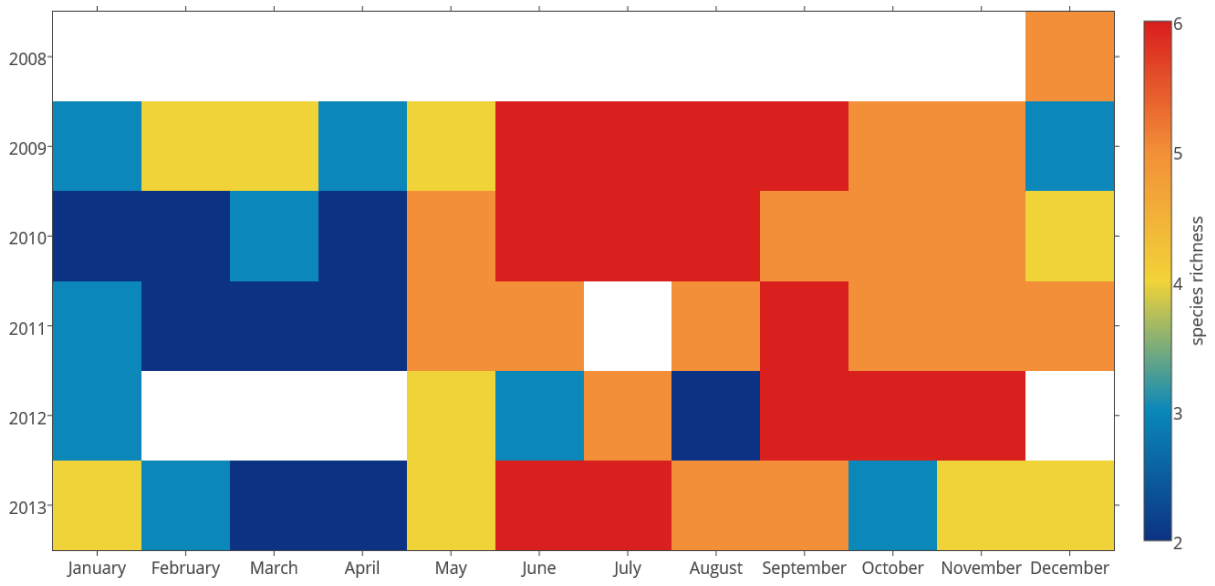

Figure 8. Monthly species richness, defined as the total number of *Skeletonema* species present in a sample, between December 2008 and December 2013. The cluster of cool colors between January and April indicates low species richness during winter and spring, while the cluster of warm colors between June and September indicate high species richness during summer. Other months tend to demonstrate moderate species richness, which often occurs in the transition from warmer seasons to cooler seasons. White squares indicate a month where no sample was available.

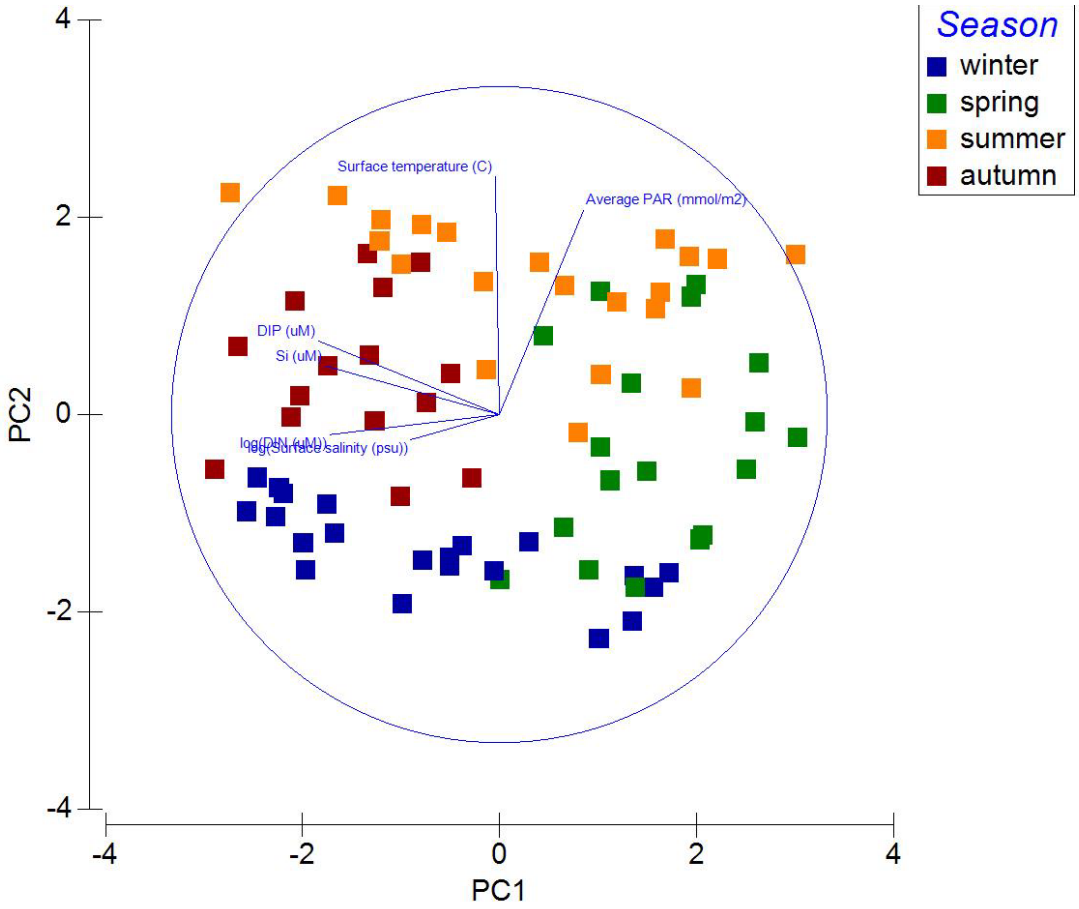

Figure 9. Principal components analysis of Narragansett Bay environmental data. Variables included surface temperature (°C), surface salinity (psu), average daily PAR (mmol/m^2), DIN (µM), DIP (µM), and Si (µM). Values of surface salinity and DIN were log transformed. Nutrient concentrations mostly contributed to PC1 and surface temperature and PAR contributed to PC2. Along PC1, high-nutrient autumn and winter samples separated from low-nutrient spring and summer samples. Along PC2, low-temperature winter and spring samples separated from high temperature summer and autumn samples. Surface salinity did not have an effect on seasonality. PC1 and PC2 explained 72.2% of the variance in environmental data.

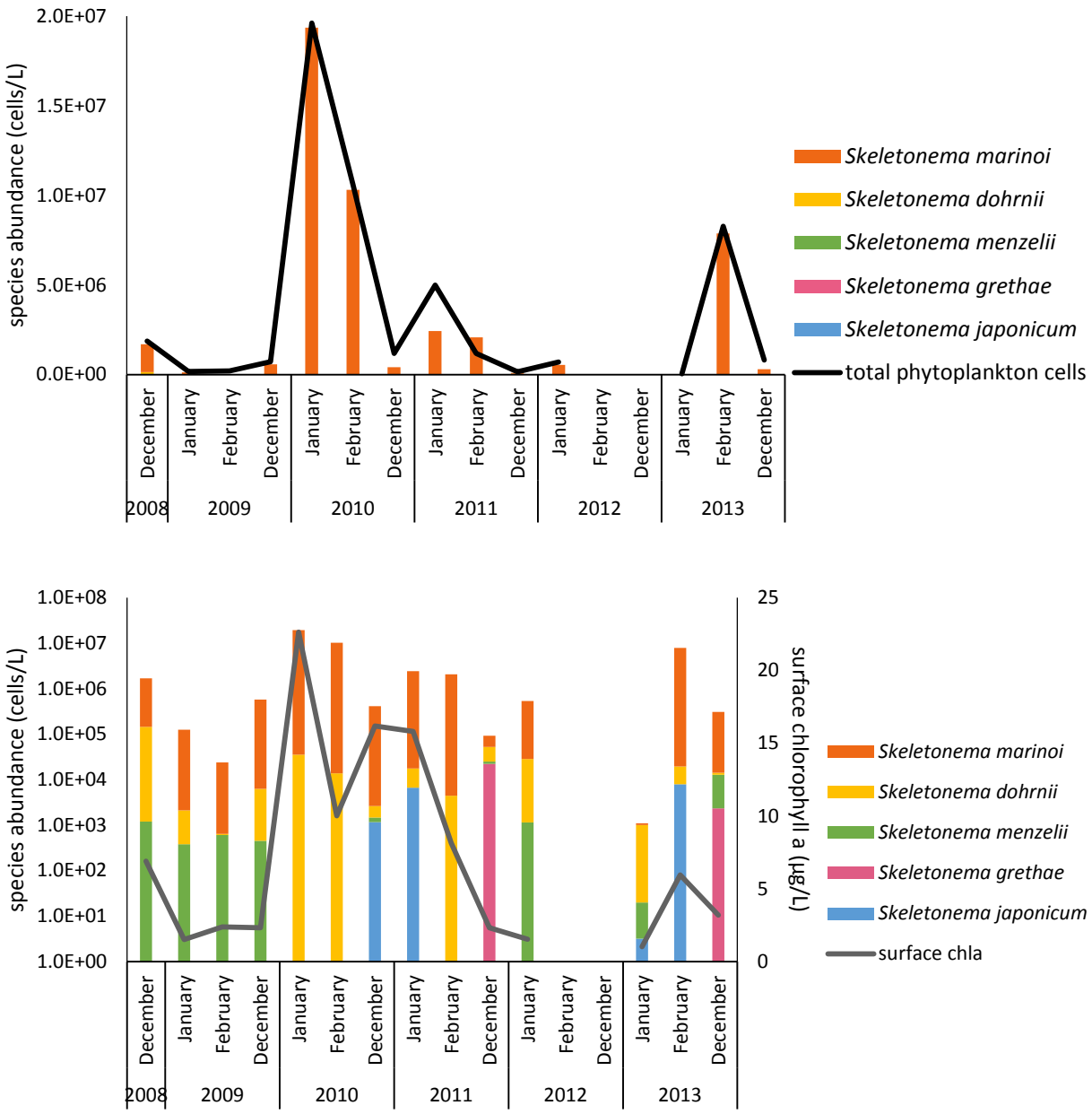

Figure 10. Percent composition data were converted to absolute abundance for five species by multiplying raw sequence reads by a transformation coefficient, which should provide more accurate representation of species. Winter samples (Dec-Feb) were analyzed because they tended to have low concentrations of the two species for which we did not calculate transformation coefficients. The left-hand y axis represents abundance in cells/L, which was obtained for each species by multiplying the transformed percent composition by total *Skeleonema* abundance. In the bottom figure, the right-hand y axis represents surface chlorophyll-*a* levels (solid line). *S. marinoi* dominated winter in both bloom and non-bloom years.

BIBLIOGRAPHY

Amaral-Zettler, L. a, McCliment, E. a, Ducklow, H. W., & Huse, S. M. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PloS One*, *4*(7), e6372. doi:10.1371/journal.pone.0006372

Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., & Follows, M. J. (2010). Patterns of diversity in marine phytoplankton. *Science (New York, N.Y.)*, *327*(5972), 1509–11. doi:10.1126/science.1184961

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, & I. Das. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, *22*(3). doi:10.1016/j.tree.2006.11.004

Biswas, H., Dey, M., Ganguly, D., De, T. K., Ghosh, S., & Jana, T. K. (2009). Comparative Analysis of Phytoplankton Composition and Abundance over a Two-Decade Period at the Land–Ocean Boundary of a Tropical Mangrove Ecosystem. *Estuaries and Coasts*, *33*(2), 384–394. doi:10.1007/s12237-009-9193-5

Bokulich, N. a, Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., … Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods*, *10*(1), 57–59. doi:10.1038/nmeth.2276

Borkman, D. G., & Smayda, T. (2009). Multidecadal (1959–1997) changes in *Skeletonema* abundance and seasonal bloom patterns in Narragansett Bay, Rhode Island, USA. *Journal of Sea Research*, *61*(1-2), 84–94. doi:10.1016/j.seares.2008.10.004

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–336. doi:10.1038/nmeth.f.303

Duleba, M., Ector, L., Horváth, Z., & Tihamér, K. (2014). Biogeography and Phylogenetic Position of a Warm-stenotherm Centric Diatom, *Skeletonema* potamos (C. I. Weber) Hasle and its Long-term Dynamics in the River. *Annals of Anatomy*, *165*(5), 715–729. doi:10.1016/j.protis.2014.08.001

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460-2461. Doi:10.1093/bioinformatics/btq461

Ellegaard, M., Godhe, A., Härnström, K., & McQuoid, M. (2008). The species concept in a marine diatom: LSU rDNA–based phylogenetic differentiation in *Skeletonema* marinoi/dohrnii (Bacillariophyceae) is not reflected in morphology. *Phycologia*, *47*(March), 156–167. doi:10.2216/07-09.1

Field, C. B., M. J. Behrenfeld, J. T. Randerson, & P. Falkowski. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, *281*(5374), 237–240. doi:10.1126/science.281.5374.237

Fuhrman, J. A., Steele, J. A., Hewson, I., Schwalbach, M. S., Brown, M. V, Green, J. L., & Brown, J. H. (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(22), 7774–7778.

Gallagher, J. C. (1980). Population genetics of *Skeletonema costatum* (Bacillariophyceae) in Narragansett Bay. *Journal of Phycology*, *16*, 464-474. doi:10.1111/j.1529-8817.1980.tb03061.x

Gallagher, J. C. (1982). Physiological variation and electrophoretic banding patterns of genetically different seasonal populations of *Skeletonema* costatum (Bacillariophyceae). *Journal of Phycology*, *18*, 148–162.

Graham, J. M., Kent, a. D., Lauster, G. H., Yannarell, a. C., Graham, L. E., & Triplett, E. W. (2004). Seasonal dynamics of phytoplankton and planktonic protozoan communities in a northern temperate humic lake: Diversity in a dinoflagellate dominated system. *Microbial Ecology*, *48*(1), 528–540. doi:10.1007/s00248-004-0223-3

Guillard, R. R. L., Carpenter, E. J., & Reimann, B. E. F. (1974). *Skeletonema* menzelii sp . nov ., a new diatom from the western. *Phycologia*, *13*(2), 131–138.

Guiry, M. D. (2012). How many species of algae are there? *Journal of Phycology*, *48*, 1057-1063. doi:10.1111/j.1529-8817.2012.01222.x

Hasle, G. R., & Evensen, D. L. (1975). Brackish-water and fresh-water species of the diatom genus *Skeletonema* Grev. I. *Skeletonema* subsalsum (A. Cleve) Bethge. *Phycologia*, *14*(4), 283–297. doi:10.2216/i0031-8884-14-4-283.1

Hubbard, K. A., G. Rocap, & E. V. Armbrust. (2008). Inter- and intraspecific community structure within the diatom genus *Pseudo-nitzschia* (Bacillariophyceae). *Journal of Phycology*, *44*(3). Doi:10.1111/j.1529-8817.2008.00518.x

Hubbard, K. A., C. E. Olson, & E. V. Armbrust. (2014). Molecular characterization of *Pseudo-nitzschia* community structure and species ecology in a hydrographically complex estuarine system (Puget Sound, Washington, USA). *Marine Ecology Progress Series*, *507*, 39-55. Doi:10.3354/meps10820

Kaeriyama, H., Katsuki, E., Otsubo, M., Yamada, M., Ichimi, K., Tada, K., & Harrison, P. J. (2011). Effects of temperature and irradiance on growth of strains belonging to seven *Skeletonema* species isolated from Dokai Bay, southern Japan. *European Journal of Phycology*, *46*(2), 113–124. doi:10.1080/09670262.2011.565128

Kooistra, W. H. C. F., Sarno, D., Balzano, S., Gu, H., Andersen, R. A., & Zingone, A. (2008). Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). *Protist*, *159*(2), 177–93. doi:10.1016/j.protis.2007.09.004

Mann, D. G. (1999). The species concept in diatoms. *Phycologia*, *38*(6), 437-495. doi:10.2216/i0031-8884-38-6-437.1

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. *Journal of Phycology*, *27*, 514–524.

Robasky, K., Lewis, N. E., & Church, G. M. (2014). The role of replicates for error mitigation in next-generation sequencing. *Nature Rev Genet.*, *15*(1), 56–62. doi:10.1016/j.biotechadv.2011.08.021.Secreted

Saravanan, V., & Godhe, A. (2010). Genetic heterogeneity and physiological variation among seasonally separated clones of *Skeletonema* marinoi (Bacillariophyceae) in the Gullmar Fjord, Sweden. *European Journal of Phycology*, *45*(2), 177–190. doi:10.1080/09670260903445146

Sarno, D., Kooistra, W. H. C. F., Medlin, L. K., Percopo, I., & Zingone, A. (2005). Diversity in the Genus *Skeletonema* (Bacillariophyceae). Ii. an Assessment of the Taxonomy of S. Costatum -Like Species With the Description of Four New Species. *Journal of Phycology*, *41*(1), 151–176. doi:10.1111/j.1529-8817.2005.04067.x

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. *Journal of Phycology*, *43*(1), 156–170. doi:10.1111/j.1529-8817.2006.00305.x

Thomas, M. K., C. T. Kremer, C. A. Klausmeier, & E. Litchman. (2012). A global pattern of thermal adaptation in marine phytoplankton. *Science, 388*, 1085. doi:10.1126/science.1224836

Windecker, L.A. (2010). Ten years of phytoplankton species abundance patterns in mid-Narragansett Bay, Rhode Island: 1999-2008. Graduate School of Oceanography, University of Rhode Island. M.S. Thesis.

Yamada, M., Otsubo, M., Kodama, M., Yamamoto, K., Nishikawa, T., Ichimi, K., Harrison, P. J. (2014). Species composition of *Skeletonema* (Bacillariophyceae) in planktonic and resting-stage cells in Osaka and Tokyo Bays. *Plankton Benthos Res*, *9*(3), 168–175.

Zhu, F., R. Massana, F. Not, D. Marie, & D. Vaulot. (2005). Mapping of picoeukaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology, 52*, 79-92. doi:10.1016/j.femsec.2004.10.006