

THE
UNIVERSITY
OF RHODE ISLAND

University of Rhode Island
DigitalCommons@URI

Graduate School of Oceanography Faculty
Publications

Graduate School of Oceanography

2016

Temporal variation of *Skeletonema* community composition from a long-term time series in Narragansett Bay identified using high-throughput DNA sequencing

Kelly L. Canesi
University of Rhode Island

Tatiana A. Rynearson
University of Rhode Island, rynearson@uri.edu

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

Follow this and additional works at: <https://digitalcommons.uri.edu/gsofacpubs>

Citation/Publisher Attribution

Canesi KL, Rynearson TA (2016) Temporal variation of *Skeletonema* community composition from a long-term time series in Narragansett Bay identified using high-throughput DNA sequencing. *Mar Ecol Prog Ser* 556:1-16. <https://doi.org/10.3354/meps11843>

Available at: <https://doi.org/10.3354/meps11843>

This Article is brought to you for free and open access by the Graduate School of Oceanography at DigitalCommons@URI. It has been accepted for inclusion in Graduate School of Oceanography Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

FEATURE ARTICLE



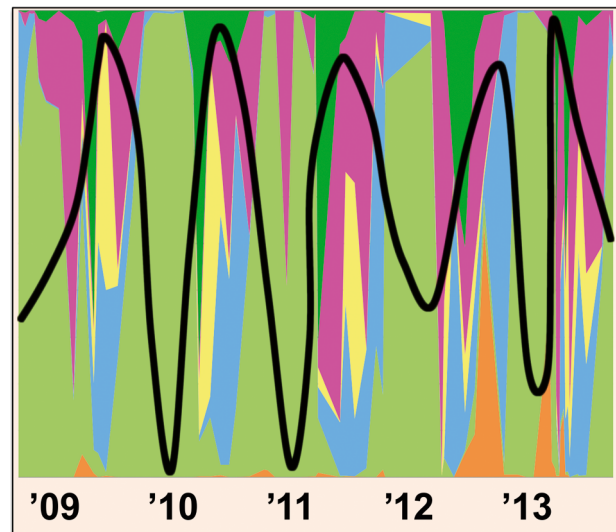
Temporal variation of *Skeletonema* community composition from a long-term time series in Narragansett Bay identified using high-throughput DNA sequencing

Kelly L. Canesi^{1,2}, Tatiana A. Rynearson^{1,*}

¹Graduate School of Oceanography, University of Rhode Island, Narragansett, RI 02882, USA

²Present Address: Gordon and Betty Moore Foundation, Palo Alto, CA 94304, USA

ABSTRACT: Phytoplankton species cannot always be identified by their morphology using light microscopy, which makes inferring the ecological and biogeochemical importance of individual species a difficult task. Here, a combination of microscopy and high-throughput DNA sequencing was used to examine morphologically cryptic and pseudo-cryptic species in the diatom genus *Skeletonema* from the Long-Term Plankton Time Series in Narragansett Bay (NBay), where *Skeletonema* is ecologically important, comprising up to 99% of microplankton cells in surface waters. The 28S rDNA from mock phytoplankton communities comprising known species was amplified and sequenced using newly developed *Skeletonema*-specific primers. The relative abundances of species in the sequence data did not match expected abundances, suggesting that 28S copy number can vary greatly, even among closely related diatom species. The 28S rDNA was also amplified from 75 field samples collected from 2008 to 2013. A total of 7 *Skeletonema* species were identified, including 5 newly detected species from NBay. *Skeletonema* species composition was highly seasonal and significantly correlated with water temperature. Winter–spring and summer–autumn communities were significantly different and characterized by low and high species richness, respectively. Species abundance during winter–spring was quantified by combining sequence data with light microscopy counts, revealing *Skeletonema marinoi* as the numerically dominant species during the winter–spring bloom. Seasonal variation in *Skeletonema* composition suggests that, although morphologically similar, species in this genus are likely adapted to different environmental conditions, raising the possibility that species composition of this important bloom-forming genus may shift as water temperatures in NBay increase due to anthropogenic influences.



High-throughput DNA sequencing revealed seasonal variation (2008–2013) in morphologically cryptic species diversity (line, 4-point moving average) in the ecologically important diatom genus *Skeletonema* (percent composition among species shown in different colors).

Image: T. A. Rynearson

KEY WORDS: Phytoplankton · Diatom · Community composition · Diversity · Cryptic species · Time series

INTRODUCTION

Diatoms are responsible for generating approximately 20% of global primary productivity (Nelson et al. 1995, Field et al. 1998, Mann 1999); yet, estimates of the total number of species are rough and range be-

*Corresponding author: ryneerson@uri.edu

tween 12 000 and 200 000 (Guiry 2012, Mann & Vanormelingen 2013). Recent taxonomic efforts have described many diatom species that are morphologically cryptic (i.e. identical) or pseudo-cryptic (i.e. small structural differences), including those in several ecologically important diatom genera such as *Cyclotella* (Beszteri et al. 2005), *Pseudo-nitzschia* (Amato et al. 2007, Lundholm et al. 2012), *Skeletonema* (e.g. Kooistra et al. 2008), *Chaetoceros* (e.g. Beszteri et al. 2005, Amato et al. 2007, Kooistra et al. 2010, Lundholm et al. 2012), *Asterionellopsis* (Kaczmarska et al. 2014) and *Leptocylindrus* (Nanjappa et al. 2013).

Because morphologically cryptic and pseudo-cryptic species are difficult or impossible to identify using routine methodology such as light microscopy, they represent a major challenge in both documenting the taxonomic diversity of marine waters (Sims et al. 2006) and understanding the impact of species composition on marine ecosystem function and biogeochemical cycles. Molecular characterizations have provided insights into the ecology of cryptic and pseudo-cryptic species (denoted hereafter as cryptic species), though studies of natural samples are limited. For example, targeted molecular analyses revealed that cryptic species in the genus *Pseudo-nitzschia* sampled from Puget Sound, Washington, varied in their relative abundances over space and time, and shifts in species composition were correlated with different environmental conditions (Hubbard et al. 2014). Similarly, cryptic species of *Pseudo-nitzschia* in the Gulf of Naples showed distinct patterns of species occurrence over time (Ruggiero et al. 2015). In the diatom family Leptocylindraceae, a targeted analysis of small subunit rDNA revealed unique temporal and geographical distributions of individual cryptic species sampled from the north-eastern Atlantic and Mediterranean coasts (Nanjappa et al. 2014). On a broader geographic scale, a global biogeography of species in the genus *Skeletonema*, assessed using the large subunit of the rDNA, revealed large-scale geographic distributions for some species and, for others, ranges that appeared restricted by temperature (Kooistra et al. 2008).

Field observations of species distributions have been supported by physiological studies showing different growth rates among cryptic species in response to environmental conditions such as temperature and light (Kaeriyama et al. 2011), salinity (Jackson et al. 1992) and nutrients (Maldonado et al. 2002). Studies of *Pseudo-nitzschia* have also described how cryptic species respond to unique environmental cues that induce sexual reproduction (Quijano-Scheggia et al. 2009). The physiological

variation observed among cryptic species in the laboratory, coupled with the few field studies of spatial and temporal variation described above, suggest that variation in cryptic species composition has important ecological and biogeochemical implications.

The diatom genus *Skeletonema* contains both cryptic and pseudo-cryptic species, and has a large influence on coastal marine ecosystems. *Skeletonema* is cosmopolitan and is known to form large blooms in coastal regions across the globe (e.g. Cleve 1900, Hasle 1973, Karentz & Smayda 1984, Cloern et al. 1985, Castillo et al. 1995, Huang et al. 2007, Borkman & Smayda 2009a), supporting food webs and contributing significantly to carbon flux (Deason 1980, Biddanda & Benner 1997). Historically, a single species, *Skeletonema costatum*, was thought to be the most ecologically important, but the genus has since been subdivided into at least 10 marine species (Guillard et al. 1974, Medlin et al. 1991, Sarno et al. 2005, 2007, Zingone et al. 2005), leaving open the question of how variation among species influences the ecology of coastal regions around the world.

Here, we examined *Skeletonema* species composition from Narragansett Bay (NBay), RI, USA, where *Skeletonema* has an enormous impact on phytoplankton productivity. It is found in the bay throughout the year and can dominate phytoplankton blooms in every season (Borkman & Smayda 2009a). On average, *Skeletonema* represents 49% of the phytoplankton community (>~10 μm) in NBay and can comprise up to 99% during blooms (Windecker 2010). Previous studies have identified a high degree of both genetic and physiological variability in *Skeletonema* sampled from NBay. For example, protein variation among isolates revealed seasonal variation in the genetic composition of what was then considered to be *S. costatum* (Gallagher 1980). In light of the newly identified species in *Skeletonema*, we hypothesized that the genetic variation observed in NBay previously may have instead reflected the differing compositions and activities of multiple *Skeletonema* species. To test this hypothesis we addressed 3 questions in this study: (1) How many species of *Skeletonema* exist in NBay? (2) Does *Skeletonema* community composition vary seasonally? (3) Do shifts in *Skeletonema* species composition correspond to seasonal fluctuations in environmental variables? We predicted that multiple cryptic species would be identified in NBay, given that 2 had previously been identified (Kooistra et al. 2008), and that there would be a high correlation between environmental variables and community composition over a multi-year time series.

MATERIALS AND METHODS

Sample collection

NBay is a temperate estuary located on the east coast of the USA and has a mean depth of 9 m. Between 2008 and 2013, surface water samples were collected weekly from NBay (41° 34.2' N, 71° 23.4' W) and 200 to 250 ml seawater were filtered in triplicate onto 25 mm diameter, 0.22 µm pore size filters (EMD Millipore), flash-frozen in liquid nitrogen and stored at -80°C. These filtered field samples were collected as part of the NBay Long-Term Plankton Time Series (Graff & Rynearson 2011). When samples were not available from the Long-Term Plankton Time Series site, samples were collected, filtered and frozen as described above from the pier at the Graduate School of Oceanography (41° 29.5' N, 71° 25.2' W). From this archive of frozen filters, 75 dates were chosen for sequencing (see Table S1 in the Supplement at www.int-res.com/articles/suppl/m556p001_supp.pdf): (1) for each month of the time series, the sample with the highest percentage of *Skeletonema* cells in the phytoplankton community (>~10 µm) was selected for sequencing; in addition, (2) a series of high-resolution weekly samples were selected from 3 representative bloom periods (winter bloom: December 2008 to January 2009; spring bloom: March to April 2010; summer bloom: May to July 2013) (Table S1 in the Supplement).

Method validation

In order to evaluate the utility of high-throughput sequencing to identify closely related *Skeletonema* species, mock phytoplankton communities were created and sequenced. A total of 3 mock communities were generated by mixing exponentially growing cultures of 5 *Skeletonema* species in equal abundances and adding this mixture in varying percentages (80, 50 and 20%) to exponentially growing cultures of 4 other phytoplankton species that were also mixed in equal abundances. The cell concentration of each cultured species was determined by light microscope (Zeiss) and a Sedgewick-Rafter slide (Cole-Parmer). Counts were conducted in triplicate both before and after cells were mixed into the mock communities to confirm actual community composition. The species used were *Skeletonema dohrnii* (CCMP3373), *S. grethae* (CCMP1804), *S. japonicum* (CCMP2506), *S. marinoi* (isolated from NBay on 1 January 2012), *S. menzelii* (CCMP793), *Ditylum*

brightwellii (isolated from Puget Sound on 12 May 2007), *Heterocapsa triquetra* (CCMP448), *Thalassiosira pseudonana* (CCMP1335) and *T. rotula* (CCMP-3096). Each mock community comprised 4×10^6 cells and was filtered and frozen as described in 'Sample collection'.

DNA extraction and amplification

For both mock community and field samples, 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 step during cell lysis using approximately 50 triple-pure high-impact 0.5 mm zirconium beads (KSE Scientific) per sample, before vortexing and incubating in lysis buffer. A 325 bp fragment within the D1 to D4 region of the large subunit (LSU) rDNA gene was amplified using newly developed *Skeletonema* genus-specific primers (Table S2 in the Supplement) that distinguished among *Skeletonema* species. The primers were identified using all available *Skeletonema* sequences in GenBank. The specificity of the primers to amplify all *Skeletonema* species but no other protists was confirmed using both an *in silico* analysis of a phylogenetically broad protist database and test amplifications in the laboratory using an array of protists including diatoms, dinoflagellates, ciliates and raphidophytes. The *Skeletonema*-specific primers did not amplify a fragment with sufficient resolution to distinguish between *S. grethae* and *S. tropicum*. Illumina adapters of varying length were added to the 5' ends of the primers to increase amplicon length diversity in the sequencing pool and thus reduce sequencing error (Table S2). DNA was amplified in 50 µl reactions containing a final concentration of 1× BIO-X-ACT Short Mix (Bio-line USA), 0.5 µM forward primer (0.125 µM each of the 4 Illumina adapted primers), 0.5 µM reverse primer (0.125 µM each of the 4 Illumina adapted primers) and 1.25 to 5 µl template DNA. A 2-step PCR consisted of a 1 min denaturing step at 94°C; 15 cycles of 94°C for 30 s, 61°C for 30 s and 72°C for 30 s; 15 cycles of 94°C for 30 s, 66°C for 30 s and 72°C for 30 s; and a final step of 72°C for 10 min. The annealing temperature (61°C) in the first step of the PCR allowed the *Skeletonema*-specific component of the primers to anneal to and amplify template DNA. The first step contained just 15 cycles to avoid saturation and better retain the species diversity present in the template DNA. In the higher annealing temperature (66°C) second step, the 5' adapter component of

the primers annealed to and amplified only existing amplicons.

Technical replication of DNA extraction, amplification and sequencing was determined using mock community samples and randomly chosen field samples. Replicate DNA extractions were performed using triplicate filters from the 3 mock communities described above. DNA from each filter was extracted, amplified and sequenced separately. DNA amplification replicates were generated using a single DNA extract from the 15 June 2010 field sample. Each triplicate PCR was sequenced separately. DNA sequencing replicates were generated using the 4 September 2009 field sample. DNA was extracted from a single filter and amplified in a single reaction tube. The amplicons were divided into 3 aliquots and sequenced separately. By measuring technical replication at these 3 steps, we evaluated not only the precision of the method but also the propagation of error from extraction to sequencing.

High-throughput sequencing and analysis

PCR amplicons were purified using AMPure XP beads (Beckman Coulter), quantified with the Qubit high sensitivity DNA kit (Thermo Fisher Scientific) and then used in a second round of PCR to attach Nextera indices and adapters following the manufacturer's instructions (Illumina). The 9 mock communities and 79 field samples (73 separate dates plus triplicates for 2 additional dates) were pooled and quantified using the KAPA qPCR kit (Kapa Biosystems) and then sequenced on a single flow cell on the MiSeq platform (Illumina) using the MiSeq 500-cycle (2×250 bp reads) kit at the University of Rhode Island Genomics and Sequencing Center.

A total of 16.35×10^6 reads representing 88 amplicon library samples were generated, of which 9.15×10^5 quality controlled reads are available at the National Center for Biotechnology Information (NCBI) (GenBank SRA accession SRP069748). A series of tests were performed to select the optimal pipeline for sequence analysis of closely related species. First, quality filters and paired-end (PE) joining were conducted using the USEARCH (Edgar 2010) fastq_mergepairs algorithm. Raw reads were truncated from each end to the first bp with a quality score of at least Q30, either before or after PE reads were merged. At this step, PE reads were allowed from 0 to 3 bp mismatches in the overlapping sequence. Second, reads were either truncated to 315 bp (10 bp short of the full length fragment) or left at their full length using

the USEARCH fastq_filter algorithm. Third, QIIME (Caporaso et al. 2010) was used to assign operational taxonomic units (OTUs) using converted .fasta files and the pick_open_reference_otus.py script. A minimum match of 99% was used. A reference database representing a suite of *Skeletonema* species was created using *Skeletonema* 28S sequences obtained from NCBI (Table S3 in the Supplement).

Data analysis and statistical methods

For each sample, the percent sequence reads corresponding to each species was determined. Species represented by $<0.1\%$ of the sequence reads in a sample were considered potential sequencing errors (due to the high degree of similarity among *Skeletonema* 28S sequences) and removed from the sample. Community analysis was performed in PRIMER v.6.1.6 (Clarke et al. 2014). A Bray-Curtis similarity matrix was produced using the percent composition data and the CLUSTER analysis tool was used to visualize similarity among seasonal samples obtained during winter (December to February), spring (March to May), summer (June to August) and autumn (September to November). An analysis of similarities (ANOSIM) was used to test for differences in relative species composition with season.

Biological and environmental data were obtained from a variety of sources. Cell counts of *Skeletonema* spp. (Table S1 in the Supplement), surface temperature, surface salinity, and concentrations of dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP) and silicic acid (Si) were obtained from the NBay 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 NBay station ($41^\circ 38.22' N$, $71^\circ 20.34' W$) (<http://cdmo.baruch.sc.edu/>). Surface salinity and DIN concentrations 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 in PRIMER v.6.1.6 (Clarke et al. 2014).

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, the sequence data from mock communities was utilized. Each *Skeletonema* species was expected to comprise 20% of sequence reads, because the 5 species were added in equal proportions. A transformation coefficient for each species was obtained by dividing the number of expected sequence reads by the observed number of sequence reads. The transformation coefficients for each species were averaged across all 9 samples to obtain a single coefficient for each of the 5 species used in the mock communities.

Because transformation coefficients could only be determined for the 5 species used in the mock communities, we focused on winter and spring samples (generally December to April) that contained only those species and no others. For each sample, the number of sequence reads for each species was multiplied by its corresponding transformation coefficient. These transformed sequence read numbers were converted to percent composition values. Finally, percent composition was applied to absolute *Skeletonema* cell counts from each field sample to obtain abundance (cells l⁻¹) of individual species.

RESULTS

Selection of an optimal sequence analysis method

Owing to the close genetic relationship among *Skeletonema* species (42 bp changes along the 325 bp fragment, with a maximum sequence similarity of 99%), computational steps were tested to identify the most accurate method of processing sequence reads from very closely related species. These tests were performed on a single mock community containing equal

concentrations of 5 species: *Skeletonema dohrnii*, *S. grethae*, *S. japonicum*, *S. marinoi* and *S. menzeli*. First, PE reads were joined and then truncated using a quality filter of Q30. This approach identified 2 species that were not part of the mock community—*S. potamos* and *S. pseudocostatum* (1 and 8% of reads, respectively; Fig. 1)—and included reads that could not be identified to species (*Skeletonema* sp.) or had no BLAST hit with species in our custom *Skeletonema* reference database (0.3 and 0.2% of reads, respectively). Based on these results, we applied the Q30 filtering before joining PE reads. Quality filtering of reads before joining PE reads yielded matches only to species that were part of the mock communities. Merged reads were truncated to 315 bp (10 bp short of the full fragment length), which 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. An additional quality filter, varying the number of mismatches allowed in the overlapping region, did not have a noticeable effect (data not shown).

Based on the results described above, the analysis pipeline applied to all field and mock community samples included (1) 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; (2) truncation of contigs to 315 bp using the USEARCH fastq_filter script; and (3) a 99% sequence match to a set of reference sequences and 2 mismatches using the QIIME pick_open_reference_otus.py script. Sequences were assigned to their best match in the database, allowing the pipeline to distinguish between sequences with as little as 1 bp difference between them (e.g. *S. dohrnii* and *S. marinoi*).

Fig. 1. Composition of *Skeletonema* sequences recovered from 3 mock phytoplankton communities. The leftmost bar represents the expected composition for all samples. The test analysis was run on Replicate (Rep) 1 of the 80% *Skeletonema* mock community and included Q30 filtering after paired end (PE) reads were joined. The remaining bars illustrate triplicate mock communities comprising 80, 50 and 20% *Skeletonema* using the following analysis parameters: Q30 filtering before joining of PE reads and truncation of merged reads to 315 bp

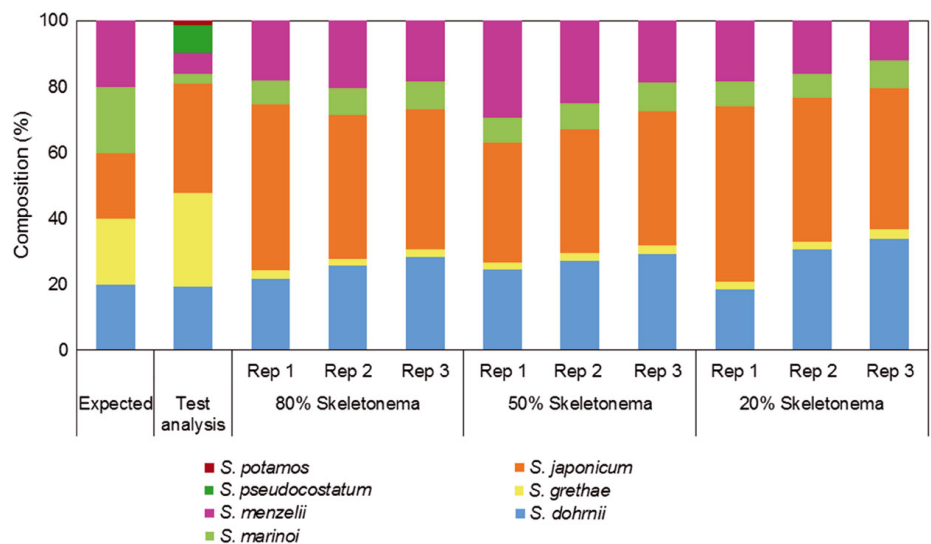


Table 1. Technical replication at 3 steps in the protocol: DNA extraction, amplification and sequencing. Values (%) represent the average \pm SD of triplicates. Mock communities were used for DNA extraction triplicates and field samples were used for amplification and sequencing triplicates. The 3 mock communities comprised 20, 50 and 80% *Skeletonema* cells and did not include *Skeletonema costatum* or *S. pseudocostatum* cells. The expected composition for each species in the mock communities was 20%. **Bold**: >15% CV among triplicate measurements

Species	Extraction triplicates (20% <i>Skeletonema</i>)	Extraction triplicates (50% <i>Skeletonema</i>)	Extraction triplicates (80% <i>Skeletonema</i>)	Amplification triplicates (Jun 2010)	Sequencing triplicates (Sep 2009)
<i>S. costatum</i>	–	–	–	0.00 \pm 0.00	0.00 \pm 0.00
<i>S. dohrnii</i>	25.25 \pm 3.40	27.05 \pm 2.39	27.70 \pm 8.02	1.32 \pm 0.29	38.95 \pm 0.53
<i>S. grethae/tropicum</i> ^a	2.23 \pm 0.29	2.26 \pm 0.29	2.48 \pm 0.34	12.09 \pm 0.05	57.67 \pm 0.52
<i>S. japonicum</i>	45.61 \pm 4.16	38.31 \pm 2.10	46.69 \pm 5.83	0.10 \pm 0.10	0.29 \pm 0.04
<i>S. marinoi</i>	7.83 \pm 0.50	8.06 \pm 0.70	7.65 \pm 0.66	7.54 \pm 0.25	0.76 \pm 0.11
<i>S. menzelii</i>	19.08 \pm 1.27	24.33 \pm 5.41	15.48 \pm 3.15	3.71 \pm 0.49	0.39 \pm 0.08
<i>S. pseudocostatum</i>	–	–	–	75.24 \pm 0.92	1.93 \pm 0.28

^a*S. grethae* was added to mock communities

Using this method, an average of 10 274 quality controlled reads per sample were recovered from each sequenced field sample or mock community (range of 112 to 33 181 reads per sample). All quality controlled reads from both mock and field communities were assigned to species in the reference database and thus de novo clustering was not implemented, although the QIIME script to pick OTUs would have done this automatically had there been unassigned reads following OTU assignment to the reference database. For each species in the reference database, no more than 2 OTUs per species were identified.

Mock communities and propagation of technical errors

A total of 3 mock communities were created in triplicate by mixing equal numbers of 5 *Skeletonema* species in various percentages (20, 50 and 80%) with 4 other phytoplankton species. Triplicate cell counts conducted immediately prior to harvesting and DNA extraction confirmed that the targeted compositions were obtained in the mock communities. In the sequencing data, all 5 *Skeletonema* species and none of the other phytoplankton species were detected, confirming that the primers when used in conjunction with high-throughput sequencing were both specific to the genus *Skeletonema* and amplified a region that could be used to distinguish among species (Fig. 1). *Skeletonema* species composition did not vary significantly among the 3 mock communities (ANOVA, $p < 0.01$) and was not affected by the absolute concentration of *Skeletonema* cells in each sample (Fig. 1). Although each species was expected

to comprise 20% of the sequence reads regardless of the total contribution of *Skeletonema* to the phytoplankton community, the observed percent composition for each *Skeletonema* species actually ranged from 2.2 to 46.7% (Table 1) and differed significantly from the expected community composition (chi-squared goodness of fit test, 4 df, $\chi^2 > 43\%$ and $p < 0.001$ for all 3 communities).

Technical replication was evaluated at 3 points: extraction, amplification and DNA sequencing. Extraction replicates were generally associated with low levels of variance, defined here as a coefficient of variation (CV) <15% (Table 1). Amplification and sequencing triplicates, conducted using field samples, also had low levels of variance with a CV <15%, except for species that represented a small fraction (<1.5%) of the sequence reads. These low abundance species included *S. dohrnii* and *S. japonicum* in the amplification replicates and *S. menzelii* in the sequencing replicates.

Abundance, composition and seasonality of *Skeletonema* in NBay

Abundance of *Skeletonema* in NBay between December 2008 and December 2013 ranged from 0 to 4.8×10^7 cells l^{-1} (Fig. 2). The timing of the major *Skeletonema* bloom varied from year to year, with large blooms (approximately 1.5 to 4.8×10^7 cells l^{-1}) occurring in late summer–autumn of 2009, winter–spring of 2010 and 2012, and early–mid summer of 2011 and 2013. Smaller blooms occurred during late summer 2010, winter–spring of 2011 and 2013, and early summer 2012. During blooms, *Skeletonema*

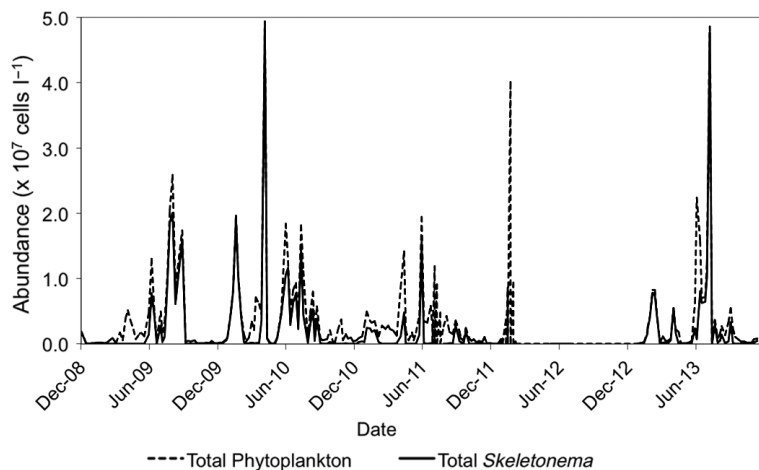


Fig. 2. Abundance of *Skeletonema* (solid line) and all phytoplankton (>10 μm ; dashed line) cells in Narragansett Bay (NBay) from December 2008 to December 2013 (no data available March to April 2012 and August to December 2012). *Skeletonema* is an important primary producer in the Bay, at some points comprising up to 99% of the total microplankton cells

often dominated the entire phytoplankton community (>10 μm), comprising up to 99% of all cells (see Table S1 in the Supplement).

Analysis of sequence data from 75 field samples revealed that 7 species of *Skeletonema* were present in NBay between December 2008 and December 2013: *S. costatum*, *S. dohrnii*, *S. grethae/tropicum*, *S. japonicum*, *S. marinoi*, *S. menzelii* and *S. pseudocostatum*. Note that *S. grethae* could not be distinguished from *S. tropicum* using the primer set utilized here. Sequences identified as *S. costatum* only appeared on 10 August 2009 and comprised 0.3% of the sequence reads.

There was little intraspecific variation in LSU sequence reads. Four species were represented by just a single LSU sequence or ribotype: *S. costatum*, *S. japonicum*, *S. marinoi* and *S. pseudocostatum*. The remaining 3 species, *S. dohrnii*, *S. grethae/tropicum* and *S. menzelii*, were each represented by 2 LSU ribotypes differing by <1.23%, matching previously identified variation within species. In *S. dohrnii*, there is low variability across the entire published 28S, and the primers used here target a region that encompasses the 2 known ribotypes. In *S. menzelii*, there are 4 ribotypes represented in the published 28S. The primer target region can distinguish among 3 ribotypes, 2 of which were detected in this study. In the *S. grethae/tropicum* group, there are 3 ribotypes in the target region; one is shared between the 2 species and it was the most frequently identified ribotype in the dataset. An additional 2 ribotypes differ by 1 bp and are unique to *S. tropicum*. One of those

was identified in the dataset but only rarely (never >0.1% of reads in a given sample).

Percent composition of sequence reads was used to analyze community composition. As data from mock communities revealed that species like *S. japonicum* were overrepresented while others, like *S. marinoi*, were underrepresented in sequence data (Fig. 1), 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 community composition.

There were large temporal shifts in *Skeletonema* community composition (Fig. 3). Samples from winter and spring months generally comprised a single species (*S. marinoi*) and occasionally other species,

especially *S. dohrnii*, *S. japonicum* and *S. menzelii*. Samples from summer and autumn months comprised a more diverse community of *Skeletonema*. Species composition differed significantly between seasons ($p = 0.001$) with the exception of winter vs. spring and summer vs. autumn ($p > 0.05$) (Table 2). Field samples were clustered into 2 distinct groups, one primarily comprising winter and spring samples, and the other summer and autumn samples (Fig. 4). Winter and spring communities clustered together more tightly than did summer and autumn communities, reflecting the relatively stable community composition in winter–spring and the more diverse and temporally dynamic species composition in summer–autumn. Only a few samples did not cluster as expected. For example, 2 samples from December 2011 (12/7/11 and 12/28/11, dates given as mm/dd/yy) clustered with summer and autumn samples. They were collected when water temperatures were nearly double the typical temperature for that time of year (11.5°C compared to typical early December temperatures of ~5°C). Similarly, a sample collected in October 2013 (10/28/13) clustered with winter–spring samples when water temperature was lower than normal. Other exceptions to the seasonal clustering pattern occurred in months bordering the transition from winter–spring to summer–autumn (5/27/09, 6/23/09, 6/6/11, 6/6/13 and 11/23/09, 11/16/10, 11/20/13). The only exception that could not be explained by water temperature was a sample collected in January 2013 (1/2/13) when water temperatures were normal for that time of year (1.8°C).

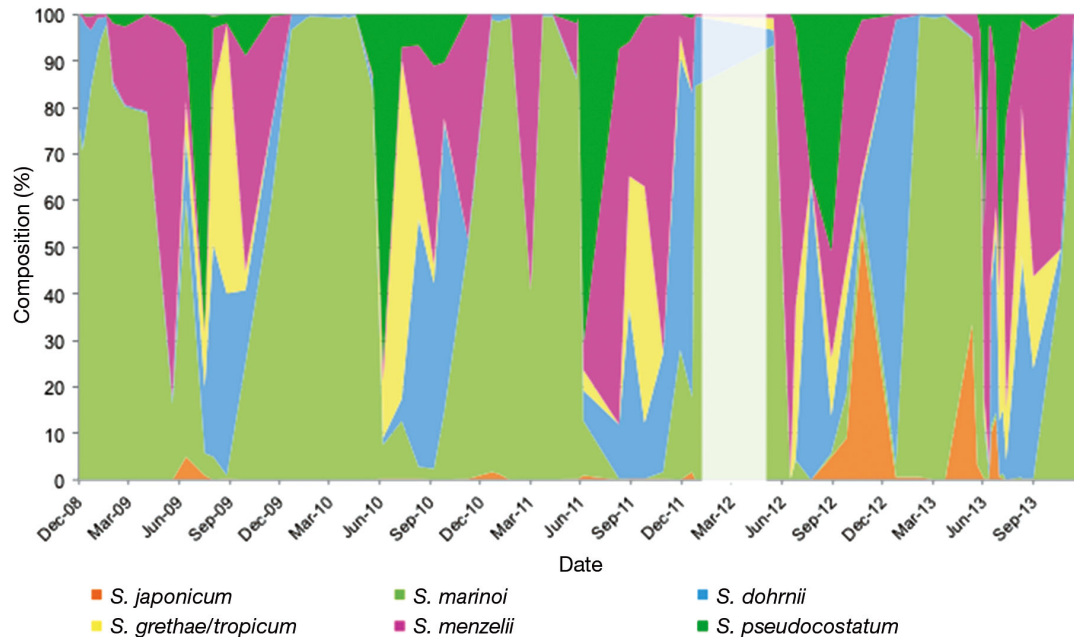


Fig. 3. Composition of *Skeletonema* sequence reads from December 2008 to December 2013 in NBay. Data include monthly samples as well as high-resolution weekly samples, which occurred during December 2008 to January 2009, March 2010 to April 2010 and May 2013 to July 2013. Samples were not collected in July 2011, February to April 2012 (white bar) and December 2012

Like community composition, species richness also shifted with season. There was a general shift in species richness between May and June where low species richness (~2 species) during winter–spring months transitioned to higher species richness (5 to 6 species) during summer–autumn months (Fig. 3). Although there were exceptions, this general shift in species richness was consistent from year to year.

High temporal resolution of species composition during 3 bloom periods

Three sets of high-resolution (weekly) samples collected during *Skeletonema* blooms revealed that

Table 2. Pairwise similarity between *Skeletonema* communities collected during different seasons (ANOSIM). The top triangle of cells represents R values (global R = 0.386) and the lower triangle represents p-values (**bold**: significantly different at $p \leq 0.05$). Community composition was significantly different among all seasons except between winter and spring, and between summer and autumn

	Winter	Spring	Summer	Autumn
Winter	–	0.009	0.581	0.578
Spring	0.264	–	0.591	0.533
Summer	0.001	0.001	–	0.023
Autumn	0.001	0.001	0.247	–

species composition varied dramatically with season. During the 2009 winter bloom, 9 weekly samples were analyzed between December 2008 and February 2009. Chl *a* concentrations reached $6.9 \mu\text{g l}^{-1}$ and *Skeletonema* comprised up to 95% of the $>10 \mu\text{m}$ phytoplankton community (Table S1 in the Supplement). During that time period, sequence reads were dominated by *S. marinoi* (see results regarding quantitative assessment of *S. marinoi* below) (Fig. 5A). A spring bloom in March to April 2010 had similar chl *a* concentrations and dominance by *Skeletonema* in the phytoplankton. Sequence reads from these samples were also dominated by *S. marinoi* over a period of 5 wk (Fig. 5B). The May to July 2013 summer bloom contrasted starkly from winter blooms. Chl *a* reached a peak of $23.5 \mu\text{g l}^{-1}$ and *Skeletonema* comprised up to 98% of phytoplankton cells. In terms of species composition, the summer bloom was more diverse (6 species present) and changed distinctly over time, with sequences dominated by *S. marinoi* then *S. menzelii*, followed by a mixture of 4 species (Fig. 5C).

Effect of environmental conditions on community composition

Between December 2008 and December 2013, temperature in NBay ranged between 0 and 24.6°C . Salinity ranged from 14.1 to 32; it should be noted that

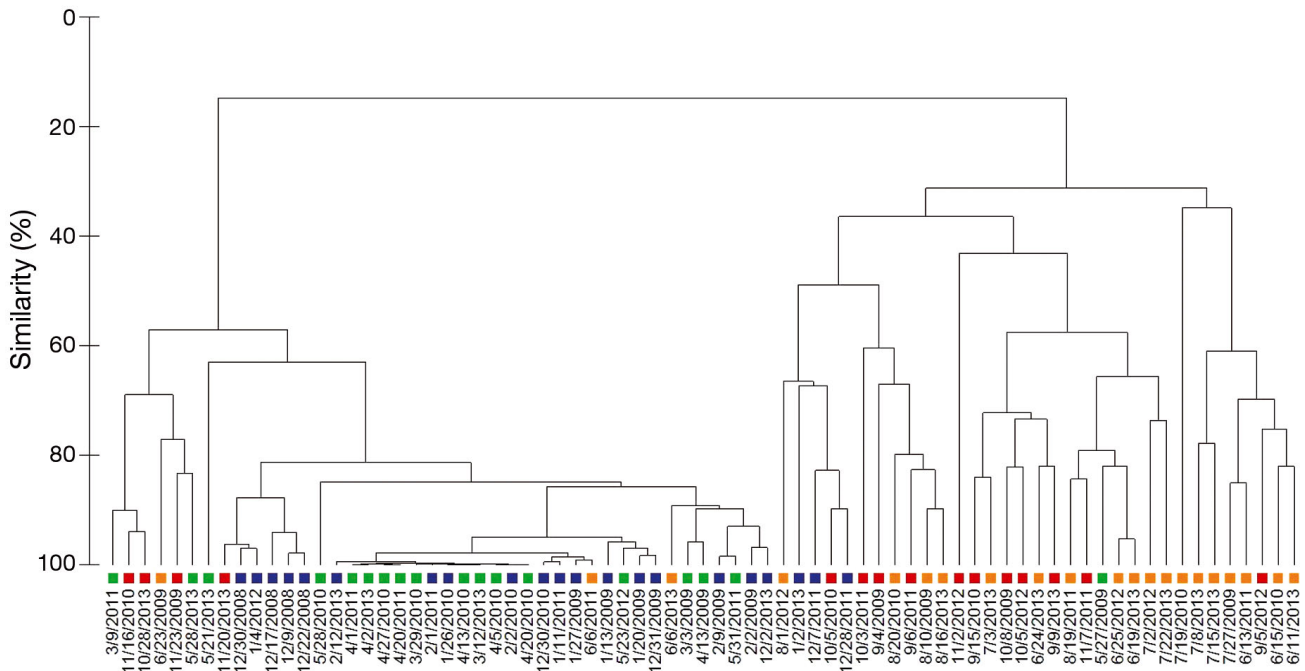


Fig. 4. Dendrogram representing the Bray-Curtis similarity of *Skeletonema* species composition. Samples are categorized by season (winter, blue; spring, green; summer, orange; autumn, red). Shorter branch lengths represent samples that are more similar and longer branch lengths represent samples that are more dissimilar. Dates given as mm/dd/yy

the minimum salinity occurred in April 2010 and coincided with severe rains and heavy influx of freshwater to NBay. Salinity most often ranges from 27 to 32. Concentrations of DIN ranged from 0.14 to 18.92 μM , DIP from 0.04 to 1.31 μM and Si from 0.08 to 38.12 μM . PAR ranged from 18 to 636 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$. BIOENV analysis showed that surface temperature and DIP were significantly correlated with species composition and explained 48.0% of the variation in *Skeletonema* communities ($p = 0.01$) (Table 3). Temperature alone explained 46.2% of the variation in species composition. Salinity, Si and DIN did not explain any significant variation in species composition.

Contribution of *Skeletonema marinoi* to winter–spring communities

Transformation coefficients were determined for the 5 species added to mock communities (*S. dohrnii*, *S. grethae/tropicum*, *S. japonicum*, *S. marinoi* and *S. menzeli*) and then used to convert relative sequence reads to absolute cell abundance for samples collected during winter–spring months that comprised only those species (Table 4). Other months were not considered as they contained 2 species not included in mock communities and for which no transformation coefficients were obtained (*S. costatum* and *S. pseudocostatum*).

The calculation of transformation coefficients yielded abundance estimates for mock communities that differed by 2%, at most, from their actual composition.

Application of transformation coefficients to field samples and combination with *Skeletonema* cell counts revealed that winter–spring months were numerically dominated by *S. marinoi*, when it generally exceeded 80% of the community composition (Fig. 6). Other *Skeletonema* species were not dominant in winter–spring communities, with the exception of November 2011. In November 2011, *S. menzeli* comprised 88% of the community (Fig. 6D). It is important to note that *S. menzeli* cell size is typically less than 10 μm and may not typically be captured in cell count observations, but that it was only detected in high abundance in the sequence data in November 2011, when both *Skeletonema* abundance and overall phytoplankton abundance were low (Fig. 6D).

DISCUSSION

Evaluating community composition of cryptic species

High-throughput DNA sequencing of a segment of the LSU rDNA was effective in characterizing a marine diatom community consisting of closely related,

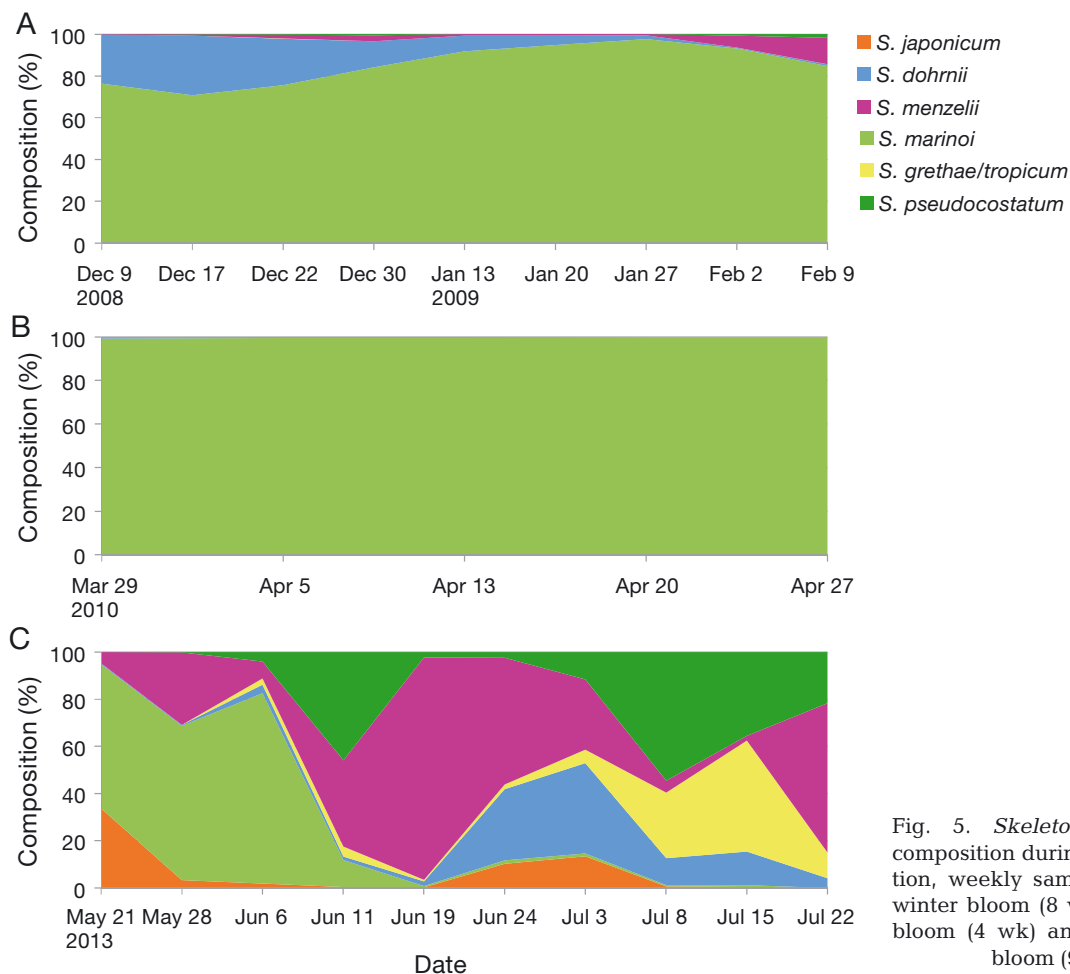


Fig. 5. *Skeletonema* species composition during high-resolution, weekly sampling of a (A) winter bloom (8 wk), (B) spring bloom (4 wk) and (C) summer bloom (9 wk)

morphologically cryptic and pseudo-cryptic *Skeletonema* species. Other regions of the ribosomal cistron, such as the commonly used small subunit (SSU) V4 and V9 regions, are very effective for identifying diversity across broad taxonomic ranges (Stoeck et al. 2010, de Vargas et al. 2015). However, the SSU would not have been variable enough to detect the different species of *Skeletonema* examined here. Conversely, the LSU species database across broad taxonomic divisions is far smaller than the SSU database. As a result, the choice of genetic marker used for high-throughput sequencing should be tailored to the taxonomic resolution of interest and the magnitude of the database available.

To assess the extent of error in our high-throughput sequencing approach, replicate measures for DNA extraction, amplification and sequencing were conducted (reviewed in Robasky et al. 2014). Variation in the relative abundance of each species associated with replicate amplification and sequencing was low, and tended to affect only low-abundance species (<1.5% composition), whereas variation associated with replicate DNA extractions was larger and

affected species that represented a moderate fraction of sequence reads (on average ~22% composition). Although DNA extraction represented the largest source of error in estimating the relative abundance of species, the average CV at this step was only 12.7%, which is comparable to variation observed in

Table 3. Correlation of NBay environmental conditions with *Skeletonema* species composition (BIOENV). A total of 6 environmental variables were analyzed: surface temperature ($^{\circ}\text{C}$), surface salinity, average PAR ($\mu\text{mol photons m}^{-2} \text{s}^{-1}$), 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%. The global BEST match permutation test had $p = 0.480$ and $p = 0.01$

No. 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 4. Transformation of a single mock community sample from relative percent composition to absolute percent composition to demonstrate the method applied to field samples. Using data from mock field communities, transformation coefficients were calculated to convert relative percent composition data into absolute abundance data. Average transformation coefficients were calculated using sequence data from all 9 mock field samples and then multiplied by individual sequence read counts for each sample. As species were added in equal proportions to the mock field communities, the actual composition was 20% for each species

Species	Sequence reads (N)	Composition of reads (%)	Transformation coefficient	Sequence reads after transformation (N)	Composition after transformation (%)
<i>Skeletonema japonicum</i>	7834	43	0.465	3647	20
<i>S. marinoi</i>	1496	8	2.56	3833	21
<i>S. dohrnii</i>	5201	28	0.773	4019	22
<i>S. grethae</i>	399	2	8.71	3478	19
<i>S. menzelii</i>	3395	19	1.08	3660	20

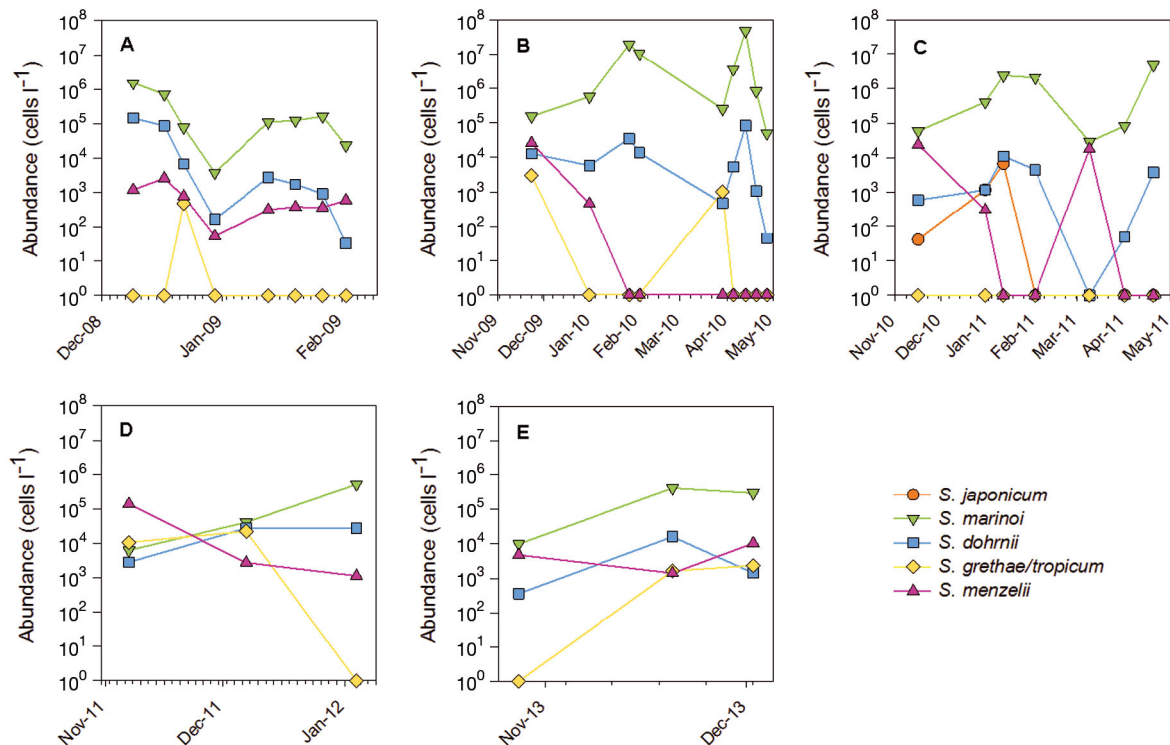


Fig. 6. Quantitative assessment of *Skeletonema* species composition during winter-spring in NBay in (A) 2008 to 2009, (B) 2009 to 2010, (C) 2010 to 2011, (D) 2011 to 2012 and (E) 2012 to 2013. Only time periods containing the 5 species for which transformation coefficients could be calculated are included

microscopy studies based on cell counts of 400 cells per sample (Lund et al. 1958).

The analysis pipeline utilized here was further validated using mock communities of known composition. This allowed for evaluation of a variety of quality controls to reduce false reads and misidentifications. One technique that was notably different in this study compared to others was the application of Q30 trimming, which is typically recommended after merging PE reads (Bokulich et al.

2013). Here, mock communities contained spurious species (i.e. *Skeletonema* sp. and the freshwater *S. potamos*) when trimming after merging PE reads. Trimming prior to merging PE reads eliminated spurious species. As bps reflective of genetic variation mostly occurred in the middle of the amplicon where forward and reverse reads overlapped, it is possible that Q30 trimming before merging reads eliminated low-quality bps that were otherwise included in the merged sequences.

Mock communities revealed that closely related *Skeletonema* species could be detected and distinguished in mixed phytoplankton communities; however, sequence read abundances did not match expected abundances. For example, abundances for *Skeletonema japonicum* were overestimated by 2.2-fold, while *S. marinoi* and *S. grethae/tropicum* were underestimated by 2.5-fold and 8.6-fold, respectively, based on the percentage of observed vs. expected sequence reads. A likely cause for the bias is variation in LSU rDNA copy number among species, which can lead to differential amplification in PCR and result in sequence read abundances that do not match species abundances. The number of rDNA cistron copies per species varies from 1 to 12 000 for a variety of algal taxa, including diatoms (Zhu et al. 2005). For example, within the diatom genus *Thalassiosira*, copy number can vary by 10-fold among species (Zhu et al. 2005). Little work has been done to investigate copy number variation among *Skeletonema* species, although LSU copy number in *S. marinoi* was estimated at 61 copies per cell (Ellegaard et al. 2008). By combining that estimate with the transformation coefficients identified here, we estimate that copy number of the 5 species used in the mock communities ranged by ~20-fold, from 11 to 208 copies per cell.

Seasonal variation in *Skeletonema* community composition

Here, 7 species of *Skeletonema* were detected in NBay, 5 of which had not been detected previously (*S. costatum*, *S. dohrnii*, *S. marinoi*, *S. menzelii* and *S. pseudocostatum*). *S. grethae* and *S. japonicum* were previously identified in NBay via single cell isolates (Kooistra et al. 2008). Some of the newly detected 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). *S. costatum*, *S. dohrnii* and *S. pseudocostatum* have all been previously identified in the Northeast Atlantic (NCBI accession EF433522.1; Kooistra et al. 2008). *S. tropicum* has been isolated primarily from subtropical waters, including the Gulf of Mexico (Sarno et al. 2005). The diversity of *Skeletonema* observed in NBay is similar to that found in coastal waters around Japan, where up to 11 species have been identified using cloning and sequencing techniques (Yamada et al. 2010, Yamada et al. 2013, Yamada et al. 2014). This is in stark contrast to previous studies of *Skeletonema* in other regions, such as the Skagerrak-

Kattegat and Baltic Sea, showing that *S. marinoi* appears to be the sole representative of the genus there (e.g. Godhe et al. 2006, Ellegaard et al. 2008, Godhe et al. 2013).

Pronounced seasonal changes in species diversity were observed within NBay. Comparison of species compositions from different seasons revealed 2 significantly different communities; a winter–spring community and a summer–autumn community. The most noticeable difference between the seasonal communities was the low species richness during winter–spring, when *S. marinoi* was the primary species in NBay, compared to higher species richness and more temporally dynamic community composition during summer–autumn. Previous research in NBay on what was thought to be *S. costatum* detected significant genotypic differences between winter–spring and summer–autumn isolates (Gallagher 1980). Based on the high-throughput sequencing data, *S. costatum* is all but absent in NBay (appearing on just 1 sampling date and comprising 0.3% of reads), suggesting that previous observations of *S. costatum* were of other species in the genus. Furthermore, the sequencing data suggest that protein variation among seasonal isolates, identified using allozymes, most likely captured inter- rather than intra-specific variation in the genus *Skeletonema*.

The pace of temporal variation in *Skeletonema* species composition in NBay depended on season. Data analyzed from high-resolution weekly samples collected during winter, spring and summer blooms suggested that winter and spring blooms tended to have consistent species composition over weeks to months, whereas summer blooms were much more dynamic, with shifts in relative species abundance occurring weekly. Rapid changes in species composition over a period of weeks could be due to stochastic effects. However, given the observation of physiological variation within the genus (Kaeriyama et al. 2011), it is more likely that rapid compositional changes may indicate that diversification within the genus *Skeletonema* has led to species with finely tuned differences in their physiologies, allowing the genus to dominate the microplankton across a range of environmental conditions.

Effect of environmental conditions on community composition

Water temperature was the single most important environmental condition that correlated with *Skeletonema* community composition. Temperature is thought

to influence phytoplankton species richness and diversity (Graham et al. 2004), in part because temperature strongly influences the growth rates of individual species (Thomas et al. 2012), including those in the genus *Skeletonema*. A comparison of growth rates under different temperatures revealed that cultures of *S. marinoi* grew faster than other *Skeletonema* species at the lowest tested temperatures (10 and 15°C) (Kaeriyama et al. 2011). This could explain the dominance of *S. marinoi* in winter–spring samples, although wintertime temperatures in NBay drop into the single digits, lower than previously examined. Earlier work in NBay on what was thought to be *S. costatum* revealed significant differences in growth rates and chl *a* concentration per cell in isolates collected from different seasons (Gallagher 1982). Now that at least 7 species of *Skeletonema* have been identified from NBay, it is likely that previously observed differences in physiological characteristics were associated with variation among species.

Water temperatures in NBay can reach up to 25°C. It has been hypothesized that transport of warm, Gulf Stream waters into NBay could allow tropical and sub-tropical species that are more adapted to warm temperatures to enter the Bay during summer months (Borkman & Smayda 2009b). Water transport could provide an opportunity for species like *S. tropicum* to enter the bay, occupy seasonal niches, and thus affect the magnitude and composition of *Skeletonema* blooms. At this time, it is unclear if *S. tropicum* is part of the NBay *Skeletonema* community, due to lack of resolution between *S. grethae* and *S. tropicum* LSU rDNA sequences analyzed here. *Skeletonema grethae* is a temperate species and has been previously isolated from NBay (Sarno et al. 2005, Kooistra et al. 2008). In contrast, *S. tropicum* has been found in marine and brackish environments, mostly clustered near low latitudes (Hulburt & Guillard 1968, Kooistra et al. 2008). With an optimum growth rate at 25°C and cessation of growth below 13°C (Hulburt & Guillard 1968, Kaeriyama et al. 2011), *S. tropicum* could thrive in NBay only during summer months. *S. grethae/tropicum* sequences were observed generally between May and October, when temperatures range from 12 to 25°C, providing the opportunity for both the temperate and tropical species to thrive. These 2 species are typically distinguishable by their chloroplast composition, as *S. tropicum* has numerous and more visible chloroplasts than *S. grethae* (Sarno et al. 2005). Future analyses should take into account occasions where microscopy is a solution for unresolved sequences.

Quantifying morphologically cryptic species in the field

It has been particularly challenging to quantify species that are morphologically cryptic from field samples, at least at the level of the light microscope. Here, we combined multiple approaches to ultimately extract quantitative abundance data on cryptic species from field samples by combining genus-level cell counts with high-throughput DNA sequence data and transformation coefficients based on mock communities. This multi-pronged approach was applied to samples collected during winter–spring months of each year, revealing that *S. marinoi* is the numerically dominant winter–spring bloom species in NBay. At times, we found that *S. marinoi* comprised up to 99.86% of *Skeletonema* cells in the bay during winter–spring blooms (e.g. 2 February 2010) and is present at cell numbers of up to 4.8×10^7 cells l^{-1} . Phytoplankton count data combined with the transformed sequence data illustrated that during the winter–spring bloom, *S. marinoi* was not only the dominant *Skeletonema* species but also the dominant microplankton species in the entire phytoplankton community. The consistent dominance of *S. marinoi* during the winter–spring bloom over 5 yr highlights the importance of a single species to the functioning of NBay.

The disconnect between species abundances obtained using microscopy vs. amplicon sequencing has been observed previously and identified as an important consideration in the improvement of studies aiming to taxonomically resolve samples using molecular methods (e.g. Wollschläger et al. 2014). For example, using mock communities we found that there could be nearly a 10-fold difference in the percent composition of a species observed in the sequencing data vs. the actual percent composition. The mismatch in species abundances between cell counts obtained using microscopy vs. amplicon sequencing has been noted rarely, but should be an important consideration in any study aiming to quantify community structure using high-throughput sequencing.

CONCLUSIONS

Although they are rarely used in high-throughput DNA sequencing surveys (Degnan & Ochman 2012), the mock communities utilized here served to enhance the ecological relevance of high-throughput sequencing data. The mock communities allowed for

validation of analysis pipelines, leading to the elimination of spurious species that would have otherwise been included in analyses, and provided insights into the extent of copy number variation in the rDNA cistron among species, even those that are very closely related. Finally, mock communities were used to obtain transformation coefficients that were then applied to field samples, yielding a quantitative assessment of abundance among cryptic species.

The 5 yr time series of field samples analyzed here revealed that different *Skeletonema* species inhabit different seasonal niches, with *Skeletonema marinoi* dominating the community in the winter and spring, and other species, such as *S. menzeli* and *S. pseudocostatum*, occurring in summer and autumn. The seasonal change in diversity was significantly correlated with water temperature and may be due to species-specific growth responses to temperature that exist among *Skeletonema* species (Kaeriyama et al. 2011). The association of community composition with water temperature indicates that on-going environmental change has the potential to shift *Skeletonema* species compositions in NBay. For example, over the past 50 yr, NBay has warmed significantly, with wintertime increases of 1.6 to 2°C (Nixon et al. 2009). Over this same time period, shifts in *Skeletonema* abundance were observed, with a nearly 2-fold drop in occurrence in the 1980s (Borkman & Smayda 2009a). Long-term changes in *Skeletonema* abundance may be related to the composition of cryptic species, potential physiological differences among species and long-term temperature changes in surface waters. Future work should aim to examine factors influencing the composition of morphologically cryptic species complexes like *Skeletonema* in order to ultimately understand and project future changes to marine phytoplankton communities.

Acknowledgements. Special thanks to D. Roche and A. Drezwianowski for their assistance in the lab. Funding was provided by Rhode Island Sea Grant under NOAA Award NA10OAR4170076, and the Rhode Island Science and Technology Advisory Council. This study was conducted using the Rhode Island Genomics and Sequencing Center, University of Rhode Island, and the Center for Computation and Visualization, Brown University, both supported in part by NSF EPSCoR 0554548 and EPS-1004057.

LITERATURE CITED

- Amato A, Kooistra WHCF, Levialdi Ghiron JH, Mann DG, Proschold T, Montresor M (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 158:193–207
- Beszteri B, Acs E, Medlin LK (2005) Ribosomal DNA sequence variation among sympatric strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) reveals cryptic diversity. *Protist* 156:317–333
- Biddanda B, Benner R (1997) Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by marine phytoplankton. *Limnol Oceanogr* 42:506–518
- Bokulich NA, Subramanian S, Faith JJ, Gevers D and others (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* 10:57–59
- Borkman DG, Smayda T (2009a) Multidecadal (1959–1997) changes in *Skeletonema* abundance and seasonal bloom patterns in Narragansett Bay, Rhode Island, USA. *J Sea Res* 61:84–94
- Borkman DG, Smayda TJ (2009b) Gulf Stream position and winter NAO as drivers of long-term variations in the bloom phenology of the diatom *Skeletonema costatum* 'species-complex' in Narragansett Bay, RI, USA. *J Plankton Res* 31:1407–1425
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K and others (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336
- Castillo J, Meave-del Castillo M, Hernandez-Becerril D (1995) Morphology and distribution of species of the diatom genus *Skeletonema* in a tropical coastal lagoon. *Eur J Phycol* 30:107–115
- Clarke KR, Gorley RN, Somerfield PJ, Warwick RM (2014) Change in marine communities: an approach to statistical analysis and interpretation, 3rd edn. PRIMER-E, Plymouth
- Cleve P (1900) The plankton of the North Sea, the English Channel and the Skagerak in 1899. *K Sven Vet Akad Handl* 34:3–31
- Cloern J, Cole B, Wong R, Alpine A (1985) Temporal dynamics of estuarine phytoplankton: a case study of San Francisco Bay. *Hydrobiologia* 129:153–176
- de Vargas C, Audic S, Henry N, Decelle J and others (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1261605
- Deason EE (1980) Grazing of *Acartia hudsonica* (*A. clausi*) on *Skeletonema costatum* in Narragansett Bay (USA): influence of food concentration and temperature. *Mar Biol* 60:101–113
- Degnan PH, Ochman H (2012) Illumina-based analysis of microbial community diversity. *ISME J* 6:183–194
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461
- 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* 47:156–167
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281:237–240
- Gallagher JC (1980) Population genetics of *Skeletonema costatum* (Bacillariophyceae) in Narragansett Bay. *J Phycol* 16:464–474
- Gallagher JC (1982) Physiological variation and electrophoretic banding patterns of genetically different seasonal populations of *Skeletonema costatum* (Bacillariophyceae). *J Phycol* 18:148–162
- Godhe A, McQuoid MR, Karunasagar I, Karunasagar I,

- Rehnstam-Holm AS (2006) Comparison of three common molecular tools for distinguishing among geographically separated clones of the diatom *Skeletonema marinoi* Sarno et Zingone (Bacillariophyceae). *J Phycol* 42:280–291
- Godhe A, Egardt J, Kleinhans D, Sundqvist L, Hordoir R, Jonsson PR (2013) Seascape analysis reveals regional gene flow patterns among populations of a marine planktonic diatom. *Proc R Soc B* 280:20131599
- Graff J, Ryneerson T (2011) Extraction method influences the recovery of phytoplankton pigments from natural assemblages. *Limnol Oceanogr Methods* 9:129–139
- Graham JM, Kent AD, Lauster GH, Yannarell AC, Graham LE, Triplett EW (2004) Seasonal dynamics of phytoplankton and planktonic protozoan communities in a northern temperate humid lake: diversity in a dinoflagellate dominated system. *Microb Ecol* 48:528–540
- Guillard RRL, Carpenter EJ, Reimann BEF (1974) *Skeletonema menzelii* sp. nov., a new diatom from the western Atlantic Ocean. *Phycologia* 13:131–138
- Guiry MD (2012) How many species of algae are there? *J Phycol* 48:1057–1063
- Hasle G (1973) Morphology and taxonomy of *Skeletonema costatum* (Bacillariophyceae). *Nord J Bot* 20:109–137
- Huang CJ, Wang C, Dong QX, Lin XP (2007) Population dynamics of *Skeletonema costatum* in Zhelin Bay of eastern Guangdong. *Acta Ecol Sin* 27:142–151
- Hubbard KA, Olson CE, Armbrust EV (2014) Molecular characterization of *Pseudo-nitzschia* community structure and species ecology in a hydrographically complex estuarine system (Puget Sound, Washington, USA). *Mar Ecol Prog Ser* 507:39–55
- Hulburt E, Guillard RRL (1968) The relationship of the distribution of the diatom *Skeletonema tropicum* to temperature. *Ecology* 49:337–339
- Jackson AE, Ayer SW, Laycock MV (1992) The effect of salinity on growth and amino acid composition in the marine diatom *Nitzschia pungens*. *Can J Bot* 70:2198–2201
- Kaczmarek I, Mather L, Luddington IA, Muise F, Ehrman JM (2014) Cryptic diversity in a cosmopolitan diatom known as *Asterionellopsis glacialis* (Fragilariaceae): implications for ecology, biogeography, and taxonomy. *Am J Bot* 101:267–286
- Kaeriyama H, Katsuki E, Otsubo M, Yamada M, Ichimi K, Tada K, Harrison PJ (2011) Effects of temperature and irradiance on growth of strains belonging to seven *Skeletonema* species isolated from Dokai Bay, southern Japan. *Eur J Phycol* 46:113–124
- Karentz D, Smayda TJ (1984) Temperature and seasonal occurrence patterns of 30 dominant phytoplankton species in Narragansett Bay over a 22-year period (1959–1980). *Mar Ecol Prog Ser* 18:277–293
- Kooistra WHCF, Sarno D, Balzano S, Gu H, Andersen RA, Zingone A (2008) Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). *Protist* 159:177–193
- Kooistra WHCF, Sarno D, Hernandez-Becerril DU, Assmy P, Di Prisco C, Montresor M (2010) Comparative molecular and morphological phylogenetic analyses of taxa in the *Chaetocerotaceae* (Bacillariophyta). *Phycologia* 49:471–500
- Lund JWG, Kipling C, Le Cren ED (1958) The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia* 11:143–170
- Lundholm N, Bates SS, Baugh KA, Bill BD, Connell LB, Léger C, Trainer VL (2012) Cryptic and pseudo-cryptic diversity in diatoms—with descriptions of *Pseudo-nitzschia hasleana* sp. nov. and *P. fryxelliana* sp. nov. *J Phycol* 48:436–454
- Maldonado MT, Hughes MP, Rue EL (2002) The effect of Fe and Cu on growth and domoic acid production by *Pseudo-nitzschia multiseriata* and *Pseudo-nitzschia australis*. *Limnol Oceanogr* 47:515–526
- Mann DG (1999) The species concept in diatoms. *Phycologia* 38:437–495
- Mann DG, Vanormelingen P (2013) An inordinate fondness? The number, distributions, and origins of diatom species. *J Eukaryot Microbiol* 60:414–420
- Medlin LK, Elwood HJ, Stickel S, Sogin ML (1991) Morphological and genetic variation within the diatom *Skeletonema costatum* (Bacillariophyta): evidence for a new species, *Skeletonema pseudocostatum*. *J Phycol* 27:514–524
- Nanjappa D, Kooistra WHCF, Zingone A (2013) A reappraisal of the genus *Leptocylindrus* (Bacillariophyta), with the addition of three species and the erection of *Tenuicylindrus* gen. nov. *J Phycol* 49:917–936
- Nanjappa D, Audic S, Romac S, Kooistra WHCF, Zingone A (2014) Assessment of species diversity and distribution of an ancient diatom lineage using a DNA metabarcoding approach. *PLOS ONE* 9:e103810
- Nelson DM, Tréguer P, Brzezinski MA, 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* 9:359–372
- Nixon SW, Fulweiler RW, Buckley BA, Granger SL, Nowicki BL, Henry KM (2009) The impact of changing climate on phenology, productivity, and benthic–pelagic coupling in Narragansett Bay. *Estuar Coast Shelf Sci* 82:1–18
- Quijano-Scheggia SI, Garcés E, Lundholm N, Moestrup Ø, Andree K, Camp 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
- Robasky K, Lewis NE, Church GM (2014) The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* 15:56–62
- Ruggiero MV, Sarno D, Barra L, Kooistra WHCF, Montresor M, Zingone A (2015) Diversity and temporal pattern of *Pseudo-nitzschia* species (Bacillariophyceae) through the molecular lens. *Harmful Algae* 42:15–24
- Sarno D, Kooistra WHCF, Medlin LK, 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. *J Phycol* 41:151–176
- Sarno D, Kooistra WHCF, Balzano S, Hargraves PE, 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* 43:156–170
- Sims PA, Mann DG, Medlin LK (2006) Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* 45:361–402
- Stoeck T, Bass D, Nevbel M, Christen R, Jones M, Breiner H, Richards T (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex

- eukaryotic community in marine anoxic water. *Mol Ecol* 19:21–31
- Thomas MK, Kremer CT, Klausmeier CA, Litchman E (2012) A global pattern of thermal adaptation in marine phytoplankton. *Science* 338:1085–1088
 - Windecker L (2010) Ten years of phytoplankton species abundance patterns in mid-Narragansett Bay, Rhode Island: 1999–2008. University of Rhode Island, Kingston, RI
 - Wollschläger J, Nicolaus A, Wiltshire KH, Metfies K (2014) Assessment of North Sea phytoplankton via molecular sensing: a method evaluation. *J Plankton Res* 36:695–708
 - Yamada M, Katsuki E, Otsubo M, Kawaguchi M and others (2010) Species diversity of the genus *Skeletonema* (Bacillariophyceae) in the industrial harbor Dokai Bay, Japan. *J Oceanogr* 66:755–771
 - Yamada M, Otsubo M, Tsutsumi Y, Mizota C and others (2013) Species diversity of the marine diatom genus *Skeletonema* in Japanese brackish water areas. *Fish Sci* 79:923–934
 - Yamada M, Otsubo M, Kodama M, Yamamoto K and others (2014) Species composition of *Skeletonema* (Bacillariophyceae) in planktonic and resting-stage cells in Osaka and Tokyo Bays. *Plankton Benthos Res* 9:168–175
 - Zhu F, Massana R, Not F, Marie D, Vaultot D (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* 52: 79–92
 - Zingone A, Percopo I, Sims P, Sarno D (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). I. A reexamination of the type material of *S. costatum* with the description of *S. grevillei* sp. nov. *J Phycol* 41:140–150

Editorial responsibility: Katherine Richardson, Copenhagen, Denmark

*Submitted: November 25, 2015; Accepted: July 19, 2016
Proofs received from author(s): August 23, 2016*