

# Oligotrophic waters of the Northwest Atlantic support taxonomically diverse diatom communities that are distinct from coastal waters

Samantha P. Setta<sup>1</sup>  | Sarah Lerch<sup>2</sup>  | Bethany D. Jenkins<sup>1,2</sup> |  
 Sonya T. Dyrhman<sup>3,4</sup>  | Tatiana A. Rynearson<sup>1</sup> 

<sup>1</sup>Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island, USA

<sup>2</sup>College of the Environment and Life Sciences, University of Rhode Island, Kingston, Rhode Island, USA

<sup>3</sup>Department of Earth and Environmental Sciences, Columbia University, Palisades, New York, USA

<sup>4</sup>Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA

## Correspondence

Tatiana A. Rynearson, Graduate School of Oceanography, University of Rhode Island, 215 S. Ferry Rd, Narragansett, RI 02882, USA.

Email: [rynearson@uri.edu](mailto:rynearson@uri.edu)

## Present address

Sarah Lerch, Institute of Marine Research, Nye Flødevigveien, Norway

## Funding information

National Science Foundation, Grant/Award Number: 1558490, 1558506, 1655686, 1756816, OIA-1004057 and OIA-1655221; Paul M. Angell Family Foundation

Editor: N. Smucker

## Abstract

Diatoms are important components of the marine food web and one of the most species-rich groups of phytoplankton. The diversity and composition of diatoms in eutrophic nearshore habitats have been well documented due to the outsized influence of diatoms on coastal ecosystem functioning. In contrast, patterns of both diatom diversity and community composition in offshore oligotrophic regions where diatom biomass is low have been poorly resolved. To compare the diatom diversity and community composition in oligotrophic and eutrophic waters, diatom communities were sampled along a 1,250 km transect from the oligotrophic Sargasso Sea to the coastal waters of the northeast US shelf. Diatom community composition was determined by amplifying and sequencing the 18S rDNA V4 region. Of the 301 amplicon sequence variants (ASVs) identified along the transect, the majority (70%) were sampled exclusively from oligotrophic waters of the Gulf Stream and Sargasso Sea and included the genera *Bacteriastrum*, *Haslea*, *Hemiaulus*, *Pseudo-nitzschia*, and *Nitzschia*. Diatom ASV richness did not vary along the transect, indicating that the oligotrophic Sargasso Sea and Gulf Stream are occupied by a diverse diatom community. Although ASV richness was similar between oligotrophic and coastal waters, diatom community composition in these regions differed significantly and was correlated with temperature and phosphate, two environmental variables known to influence diatom metabolism and geographic distribution. In sum, oligotrophic waters of the western North Atlantic harbor diverse diatom assemblages that are distinct from coastal regions, and these open ocean diatoms warrant additional study, as they may play critical roles in oligotrophic ecosystems.

## KEY WORDS

diatom, diversity, oligotrophic, richness, Sargasso Sea

## INTRODUCTION

Diatoms are a species-rich class of eukaryotic phytoplankton that generates ~20% of global primary production (Falkowski, 1998; Field, 1998; Tréguer

et al., 2018). The impact of diatom primary production on biogeochemical cycling and food web dynamics is strongly influenced by the taxonomic composition of diatom assemblages (Nelson & Brzezinski, 1997), which can vary over both space and time (Borkman

**Abbreviations:** ASV, amplicon sequencing variant; CW, coastal waters; DIN, dissolved inorganic nitrogen; GS, Gulf Stream; SS, Sargasso Sea; SRP, soluble reactive phosphorus; tb-RDA, transformation-based redundancy analysis.

& Smayda, 2009; Karentz & Smayda, 1998). In eutrophic coastal waters, taxonomically diverse diatom assemblages sustain high levels of biomass that fuel coastal ecosystems (Bracher et al., 2009; Smetacek, 2012). In persistently oligotrophic regions of the global ocean, diatom biomass is generally low (Nelson et al., 1995; Tréguer et al., 2018) except when they form intermittent blooms, which can contribute significantly to organic matter export in the open ocean (Brzezinski & Nelson, 1995; Karl et al., 2012; Nelson & Brzezinski, 1997).

Diatoms that survive in oligotrophic habitats have a variety of adaptations that allow them to grow despite low levels of macronutrients. These adaptations include the abilities to respond quickly to episodic nutrient pulses (Alexander et al., 2015; Krause et al., 2010; McGillicuddy et al., 2007), to modulate a diverse set of metabolic pathways (Armbrust, 2009), to reduce nutrient quotas (Van Mooy et al., 2009; Whitney et al., 2011), to utilize organic nutrient sources (Dyhrman et al., 2012), and to establish associations with diazotrophs (Carpenter et al., 1999; Foster & Zehr, 2019; Villareal & Lipschultz, 1995). Small diatoms have an advantage in low-nutrient waters due to their high surface-area-to-volume ratios (Finkel et al., 2010), and some larger diatoms vertically migrate through the water column to access nutrients at depth (Villareal & Lipschultz, 1995). These types of metabolic and physiological adaptations allow some diatom species to persist as sparse seed populations in oligotrophic regions and to achieve high growth rates and biomass in response to sporadic nutrient inputs (Goldman, 1993; Krause et al., 2009, 2010).

In contrast to the well-documented metabolic strategies that diatoms use to survive in oligotrophic regions, the taxonomic composition of oligotrophic diatom communities and the environmental correlates of their distributions have been less described, particularly in the oligotrophic western North Atlantic. The two quantitative records of diatom species composition that exist for the Sargasso Sea noted a diverse set of species belonging to both centric (*Bacteriastrum*, *Chaetoceros*, *Leptocylindrus*, *Rhizosolenia*, *Skeletonema*, and *Thalassiosira*) and pennate (*Nitzschia*) genera (Hulbert et al., 1960; Hulbert & Rodman, 1963). Those results have been supported by other studies that (a) used light microscopy, either applied opportunistically (e.g., personal communication notes in Bidigare et al., 1990; Malone et al., 1993) or with a focus on genus-level designation (Benitez-Nelson et al., 2007; Krause et al., 2010; McGillicuddy et al., 2007) and (b) used rRNA approaches to obtain genus-level designations (Lampe et al., 2019). Interestingly, Hulbert et al. (1960) remarked on the large number of small centric diatoms that occurred during spring but which could not be identified with light microscopy. However, the composition of those diatoms

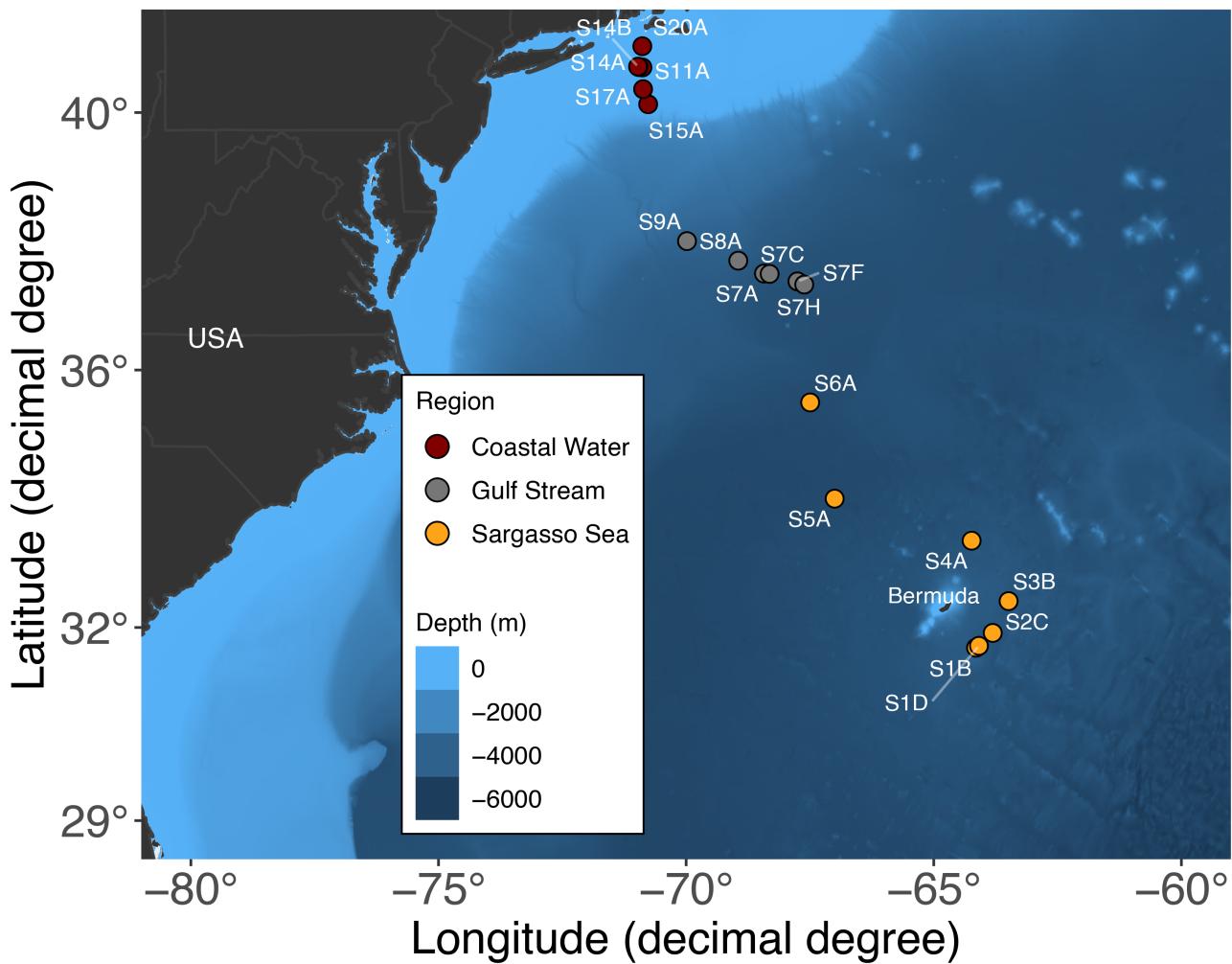
has not been further explored. Models have generated contrasting predictions that climate change will either increase (Busseni et al., 2020) or decrease (Henson et al., 2021) diatom diversity in oligotrophic open ocean regions of the North Atlantic. These predictions are difficult to evaluate given our limited understanding of North Atlantic diatom taxonomic composition and how their distributions are correlated with environmental conditions.

To address the paucity of data on diatom community composition in the oligotrophic Sargasso Sea and Gulf Stream, we examined diatom 18S rDNA gene sequence variation from biomass collected along a transect in the western North Atlantic, from high-nutrient coastal waters to the low-nutrient Sargasso Sea region of the North Atlantic Subtropical Gyre, the oligotrophic gyre that is expanding most rapidly due to climate change (4.3% per year; Polovina et al., 2008). Using high-throughput amplicon sequencing and polymerase chain reaction primers designed to recover diatom diversity with more sensitivity than universal eukaryotic primers (Zimmermann et al., 2011), we examined the taxonomic composition of diatom communities along the transect and evaluated whether diatom community composition and richness corresponded to gradients in nutrients and temperature between the Sargasso Sea and waters on the US Northeast Shelf.

## METHODS

### Sample collection

Samples were collected from 17 locations in the western North Atlantic aboard the R/V Atlantic Explorer from May 2 to 15, 2018 (AE1812; Figure 1). At each sampling location, water was collected from the surface mixed layer at 25 m and the deep chlorophyll maximum (DCM), except when bottom depth was less than 25 m in shelf waters (Data S1 in the Supporting Information). Light and temperature data were obtained using a Sea Bird Scientific SBE 911plus CTD with rosette-mounted sensor that measured photosynthetically active radiation (CPAR). Dissolved inorganic nitrogen (DIN) and soluble reactive phosphorus (SRP) concentrations (hereafter phosphate) were obtained by filtering whole seawater through GF/F ( $\sim 0.7 \mu\text{m}$ ) filters (Whatman) and measuring flow-through water on a Seal Analytical AA3 HR nutrient autoanalyzer at the University of Hawaii School of Ocean and Earth Science and Technology Laboratory for Analytical Biogeochemistry. Samples with phosphate concentrations below 125 nM were also measured using the Magnesium Induced Coprecipitation (MAGIC) protocol (Karl & Tien, 1992). Chlorophyll a concentrations were measured in triplicate by filtering whole seawater over GF/F ( $\sim 0.7 \mu\text{m}$ ) filters



**FIGURE 1** Locations sampled May 2–15, 2018 from the Sargasso Sea (SS), Gulf Stream (GS), and coastal waters (CW). Ocean bathymetry is shown in the background.

(Whatman). Chlorophyll *a* was immediately extracted from filters using 100% denatured ethanol over 12 h (Jespersen & Christoffersen, 1987) and measured on a 10-AU Fluorometer (Turner); concentrations were then calculated using a predetermined calibration curve.

Seawater samples for community composition analyses using rDNA genes were pre-filtered over mesh netting of either 153  $\mu\text{m}$  (stations 1B–9A) or 200  $\mu\text{m}$  (stations 11A–20A) to exclude large macro-zooplankton grazers. The remaining biomass was collected on either 0.2- $\mu\text{m}$  filters (10  $\text{cm}^2$  polyethersulfone, Millipore Sterivex or 25 mm polyester, Sterlitech) or 5.0- $\mu\text{m}$  filters (25 mm PETE, Sterlitech), in volumes of 2–10 L depending on biomass (Data S1). Filter pore-sizes were chosen to pair with filtering for other environmental and molecular analysis, and we therefore tested for significant differences between size fractions before combining all samples for a more robust assessment of diatom richness (see [Statistical Analyses](#) section). By examining all diatoms in the <153–200- $\mu\text{m}$  size fraction, we captured the full size range of diatom diversity. Filters

were flash frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$  until DNA extraction.

### Analysis of community composition using DNA

DNA was extracted from filtered biomass using the Blood & Tissue Kit (Qiagen) following the manufacturer's animal tissues protocol with modifications. For samples collected using the 0.2- $\mu\text{m}$  pore size Sterivex filters, standard Sterivex extraction methods were followed by injecting lysis buffer and beads directly into the cartridge and incubating (Chestnut et al., 2014; Cruaud et al., 2017). Briefly, cell lysis was conducted by injecting 1 mL ATL buffer, 90  $\mu\text{L}$  of Proteinase K, and 200  $\mu\text{L}$  of 400- $\mu\text{m}$  sterilized zirconium beads with a 3-mL Luer lock sterile syringe before capping and vortexing the samples at the highest setting for 4 min. Samples were then incubated at  $56^\circ\text{C}$  for 1 h, vortexing every 15 min. After incubation, the lysis solution was removed from filter capsules using a 3-mL Luer lock

sterile syringe and placed into a sterile 5-mL conical tube with 1mL AL buffer, vortexed, and incubated at 56°C for 10min. Finally, 1mL of 96%–100% ethanol was added to samples, followed by vortexing for 15s and pipetting into the DNEasy spin column (Qiagen). Remaining steps followed the animal tissues (Qiagen) protocol. For samples collected on 0.2-μm and 5-μm polyester and polyethersulfone filters, cells were lysed using a Mini-Beadbeater-96 (BioSpec) at top-speed (36 oscillations · s<sup>-1</sup> or 2,100rpm) for 60s with 540μL of ATL buffer and ~40μL of 400-μm sterilized zirconium beads. Afterward, 60μL of Proteinase K was added to each sample, and samples were then vortexed for 15s and incubated at 56°C for 1 h, vortexing every 15min. Remaining steps followed the animal tissues (Qiagen) protocol.

A 420-bp fragment of the 18S rDNA V4 hypervariable region was amplified from extracted DNA using universal diatom primers (Zimmermann et al., 2011) with Illumina-specific adapters (Rynearson et al., 2020). Reaction mixtures included 0.3μM each of the primers D512for and D978rev (Zimmermann et al., 2011), 50% v/v HIFI Master Mix (KAPA), 3–130ng template DNA, and DNase free water for a total reaction volume of 25μL. Sequences were amplified in triplicate using a modification of the thermocycling protocol in Rynearson et al. (2020) and pooled for sequencing. Thermocycling conditions were aimed at reducing sequencing bias and included an initial denaturing step at 95°C for 3min, followed by 15 cycles of 30s each at 94°C, 55°C, and 72°C, followed by 15 cycles of 30s each at 94°C, 70°C, and 72°C, and 10min at 72°C. Amplicons were cleaned with Ampure XP beads (Beckman Coulter, Inc.), followed by quantification using the Qubit Broad Range DNA Assay Kit (Thermo Fisher Scientific, Inc.), then amplified for five cycles with Nextera indices and adapters (Illumina Inc.). Samples were cleaned once more with Ampure XP beads, and PCR products were quantified with the KAPA qPCR kit (Kapa Biosystems). Products were sequenced on an Illumina MiSeq platform using v3 chemistry (2×300bp Illumina Inc.) at the Genomics and Sequencing Center at the University of Rhode Island.

Paired end sequencing reads were first trimmed of primers and Illumina adapters using Cutadapt (version 2.7) with an error rate for primers of 0.10 (–e option; Martin, 2011). The R package dada2 (version 1.14.0) was used to filter reads from each sample with the following criteria: a max error of 2, minimum truncate length before quality filtering of 250bp forward and 225bp reverse, and minimum final quality trimmed read length of 210bp forward, 210bp reverse. Dada2 was used to generate an error model of each unique sequence using learnErrors(), then error rates were used to infer true biological sequences using the dada() command. Forward and reverse reads were merged using the dada2 command mergePairs() with a minimum overlap

(minOverlap) of 10bp. Trimmed and merged sequences were grouped into amplicon sequencing variants (ASVs) using the makeSequenceTable() command to implement the Divisive Amplicon Denoising Algorithm (DADA), which can resolve differences between reads down to one nucleotide (Callahan et al., 2016). Finally, chimeric sequences were removed using the removeBimeraDenovo() dada2 command with the consensus method to remove sequences that are combinations of abundant parent sequences.

Taxonomic classification of ASVs was conducted using the PR2 Database (version 4.12.0; Guillou et al., 2012) and the naïve Bayesian classifier method with the assignTaxonomy() command implemented in dada2 (Callahan et al., 2016). Amplicon sequencing variants were kept separate in the statistical analyses below even if they were assigned the same species identification because the region of the 18S rRNA gene we amplified had previously been shown to predominantly recover differences among species and not strains (Zimmermann et al., 2011). For this reason, ASV number was always appended to species names to preserve these distinctions. Finally, sequences were checked for contaminants using the R package decontam (version 1.6.0) to remove any contaminated sequence features using a statistical classification procedure (Davis et al., 2018). To ensure that each sample contained sufficient reads for further statistical analyses, rarefaction curves were used to set a lower limit of 200 diatom reads per sample, eliminating three samples from further analysis (Figure S1 in the Supporting Information). Two additional samples (SPS32 & SPS33) contained sequentially filtered biomass over a 5-μm filter and then a 0.2-μm filter, and therefore, ASV counts from those samples were merged for downstream analysis (SPS32\_m), yielding a total of 30 samples that were included in further statistical analyses.

## Statistical analyses

All statistical analyses were conducted in R (version 4.04). Because amplicon sequencing yields inherently compositional data (Gloor et al., 2017) and because rDNA gene copy number can vary by orders of magnitude among species (Santoferrara, 2019; Zhu et al., 2005), sequence reads were analyzed for ASV identity and presence/absence, not abundance. Relative abundances are presented only in the bar graph in Figure S2 in the Supporting Information. To obtain a robust measure of ASV richness, we grouped samples from all depths and size fractions for each site. Sample sites were assigned to Sargasso Sea (SS), Gulf Stream (GS), or coastal waters (CW) based on their temperature and salinity characteristics. Regional (SS, GS, CW) differences in DIN, phosphate, and chlorophyll a concentrations were examined using ANOVA

(aov function) and post-hoc Tukey's tests (tukeyHSD function) at an alpha of 0.05. Diatom ASV richness is reported as the total number of ASVs per sample. Diatom genus richness is reported as the total number of genera per sample and used to highlight divergence in taxonomy across region that might relate to trait differences. Differences in ASV number and genus richness by region were tested using an ANOVA (aov function). A linear regression of chlorophyll a concentration and ASV richness was calculated using the lm() function.

To test for variation in ASV richness by size fraction, an ANOVA (aov function) was used to test the influence of filter pore size on richness estimates. Because the results of the ANOVA were insignificant, we then grouped samples by region (rather than filter pore size) for further diversity analysis. To examine variation in cell size among species identified exclusively from coastal waters compared with offshore regions, cell volumes of diatom ASVs identified to species were obtained from LeBlanc et al. (2012) and then separated into two size classes of cell volumes: small ( $\leq 6,500 \mu\text{m}^3$ ) and large ( $> 6,500 \mu\text{m}^3$ ). The cell volume cut offs correspond to the volume of a cylindrical cell with both cell height and diameter of 20  $\mu\text{m}$ . The percent of ASVs in each volume range (small or large) was then calculated for each region. To test for differences among regions, depth, and filter pore size, the vegan (version 2.5–7) function ordinate() was used with the Jaccard similarity index and max iterations of 500. Differences among groups in the Jaccard similarity matrix were tested in vegan (Adonis function) following homogeneity tests (betadisper function). A Mantel test (Mantel & Valand, 1970) was used to test for differences in diatom community composition with geographic distance using the Jaccard similarity index. Finally, a transformation-based redundancy analysis (tb-RDA) was used to correlate diatom presence with environmental variables across the three regions. A subset of 27 samples out of the total 30 were included in the tb-RDA; three samples were removed because two (SPS34 & SPS35, CW station S20A) did not have associated nutrient measurements and one (SPS02, SS station S1B) had an outlier dissolved organic nitrogen value ( $> 1.89 \mu\text{mol} \cdot \text{L}^{-1}$ ) concentration of  $11.25 \mu\text{mol} \cdot \text{L}^{-1}$  DIN. Prior to ordinating using a tb-RDA, the presence of ASVs was first normalized with vegan (decostand function) using a Hellinger standardization to give less weight to rare ASVs only present in one sample that might have influenced multivariate analysis (Legendre & Gallagher, 2001). To achieve normality, phosphate was log transformed, and both phosphate and temperature were standardized with z-scores followed by redundancy analysis (rda function) and significance testing using ANOVA and a holm  $R^2$  adjustment for multiple comparisons. Silicate ( $\text{SiO}_2$ ), DIN, phosphate, temperature, CPAR and salinity were all included in exploratory analysis, but only temperature and phosphate were significant ( $p \leq 0.05$ ), and the final

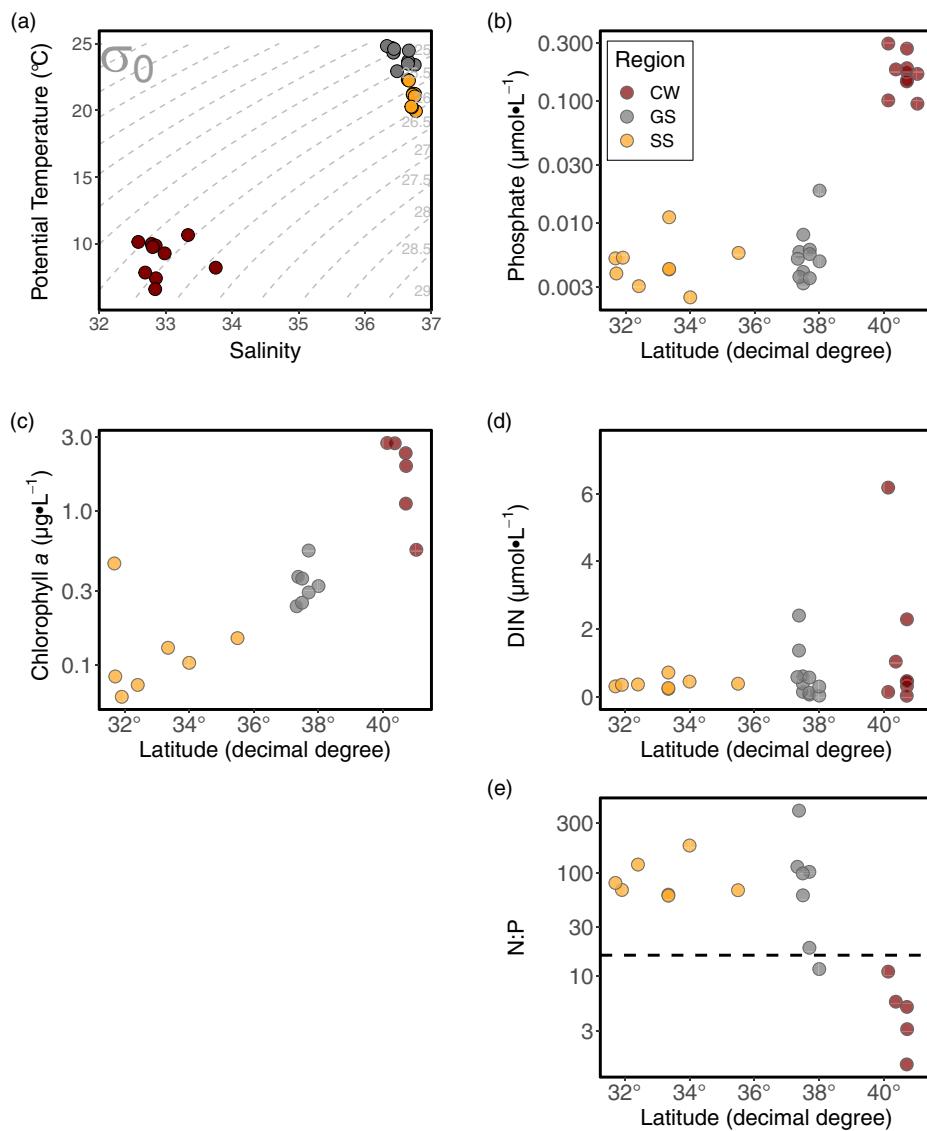
redundancy analysis only included these significant environmental variables.

## RESULTS

Samples collected in the western North Atlantic (Figure 1) during May 2018 were defined as originating from nearshore coastal water (CW, salinities of 32–33 and temperatures of 6–10°C), and offshore regions of the Sargasso Sea (SS, salinity 37 and temperatures of 20–22°C) and Gulf Stream (GS, high salinities of 36–37 and temperatures of 23–25°C; Figure 2a; Data S1). Compared with the Sargasso Sea and Gulf Stream, coastal water had significantly higher chlorophyll a (ANOVA,  $F_{2,17}=23.6$ ,  $p < 1.26\text{e}^{-5}$ ) and phosphate (ANOVA,  $F_{2,27}=70$ ,  $p < 2.03\text{e}^{-11}$ ) concentrations but not DIN concentrations (ANOVA,  $F_{2,25}=0.487$ ,  $p=0.62$ ; Figure 2b–d; Tables S1 and S2 in the Supporting Information). There were no significant differences between the Gulf Stream and Sargasso Sea (Tukey post-hoc HSD) in terms of phosphate, DIN, or chlorophyll a (Table S2). Dissolved inorganic nitrogen: phosphate (N:P) ratios were largely above 16 in the Sargasso Sea and Gulf Stream and below 16 in coastal water (Figure 2e).

Sequencing of the amplified rDNA yielded 657,709 sequences over 34 samples with an average of 19,344 reads per sample. On average,  $21.4\% \pm 21.7\%$  of sequence reads matched the class Bacillariophyceae in each sample (Figure S3 in the Supporting Information). Thirty-one samples exceeded the minimum threshold of 200 diatom reads (class Bacillariophyceae) per sample established following a rarefaction analysis to prevent under sampling (see Methods section and Figures S1 and S3). After merging two samples (SPS32 and SPS33, see Methods section), the final 30 samples were comprised of nine Sargasso Sea samples, 11 Gulf Stream samples, and 10 coastal water samples. A total of 301 ASVs were identified as Bacillariophyceae, and of those, 181 could be identified to genus level and 90 identified to species level (Data S1). Reads from the genera *Thalassiosira*, *Minidiscus*, and *Skeletonema* were abundant in coastal water, while *Pseudo-nitzschia* and *Nitzschia* reads were abundant in the offshore regions (Figure S2). In samples from both nearshore and offshore regions, a large proportion of reads (21%–87%) could not be assigned even to the genus level.

There were no regional differences in diatom richness, either at the ASV (ANOVA,  $F_{2,27}=0.45$ ,  $p=0.644$ ) or genus levels (ANOVA,  $F_{2,27}=0.42$ ,  $p=0.659$ ; Figure 3a,b). On average, each water sample along the transect had an ASV richness of  $30 \pm 17$  and genus richness of  $10 \pm 5$ . There was no significant relationship between diatom diversity (ASV richness) and total chlorophyll a (linear regression,  $R^2 < 0.01$ ,  $F_{1,26}=0.045$ ,  $p=0.834$ ; Figure 3c). There were no significant differences in ASV diversity with either size fraction (either

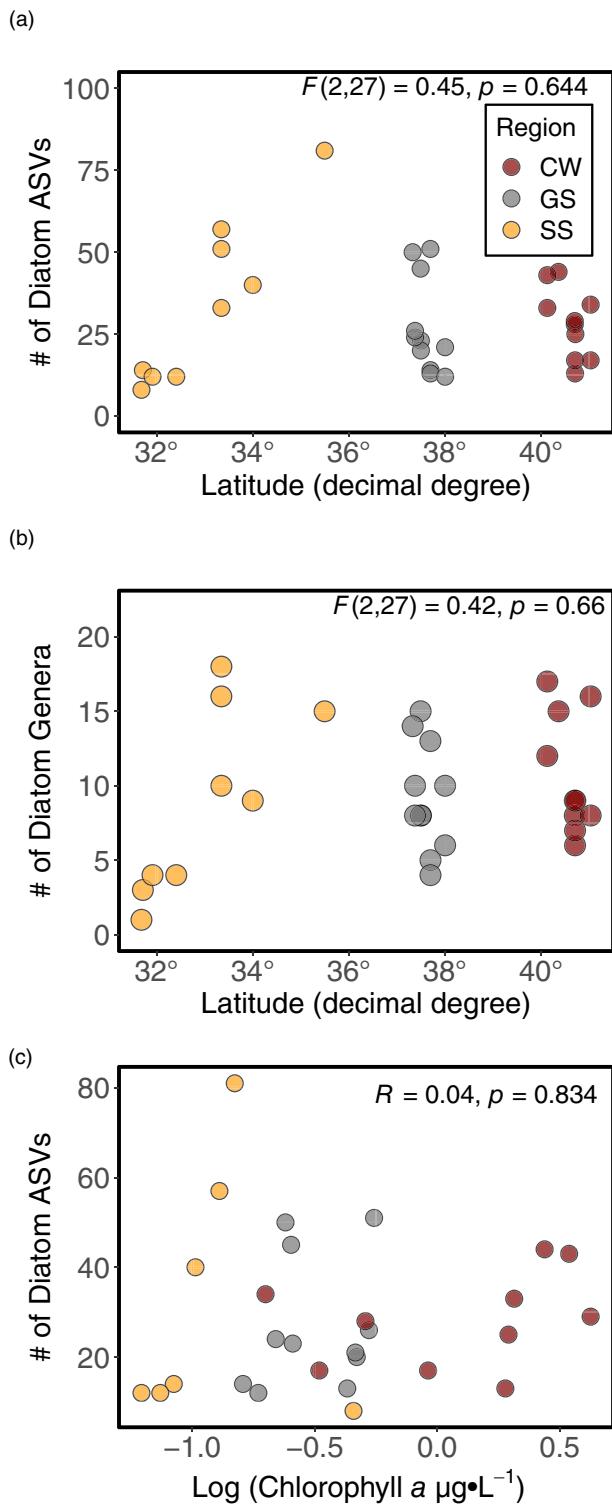


**FIGURE 2** Characterization of water samples including temperature and salinity (a); and concentrations of phosphate (b); Chlorophyll a (c); dissolved inorganic nitrogen, DIN (d); and N:P (DIN:phosphate) ratios (e). Concentrations of phosphate, Chlorophyll a, and N:P ratio are graphed on a log scale. Dissolved inorganic nitrogen in sample SPS02 was  $11.25 \mu\text{mol} \cdot \text{L}^{-1}$  and considered an outlier and removed from the DIN graph in (d). Regions include, coastal water (CW), Gulf Stream (GS), and Sargasso Sea (SS).

>0.2 or >5 µm, ANOVA,  $F_{1,28} = 1.779, p = 0.193$ ) or depth (surface or deep, ANOVA,  $F_{1,28} = 0.084, p = 0.744$ ; Data S1). The proportion of Bacillariophyceae ASVs that could be identified to species level was similar across regions: 70% and 65% of reads from the offshore region (Gulf Stream & Sargasso Sea) and coastal water, respectively.

Of Bacillariophyceae ASVs identified, just four were shared among all three regions and of those, only one could be identified to genus and species level (ASV\_69, *Guinardia striata*). The Sargasso Sea and Gulf Stream shared 71 ASVs including several species of *Pseudo-nitzschia*, *Nitzschia*, and *Leptocylindrus* (Figure 4a; Figure S4 in the Supporting Information). A total of 211 ASVs were found exclusively in the offshore region (70%; Sargasso Sea and Gulf

Stream), while 67 ASVs were found exclusively in the nearshore coastal water (25%; Figure 4a). Coastal water samples shared 10 ASVs with the Sargasso Sea samples, including those identified as *Guinardia delicatula*, *Guinardia flaccida*, and *Minutocellus polymorphus*, and shared none with the Gulf Stream (Figure 4a; Figure S4). Of the 211 ASVs found only in the offshore region, 168 ASVs (80%) could be identified to genus level, of which 129 ASVs (77%) were within the same genus as ASVs found only in the nearshore region (Figure 4). Diatom ASVs that were found only in the nearshore or offshore regions and that were identified to species level (64 of 111 ASVs identified to species level, Data S1) had different cell volumes. Amplicon sequence variants observed only in offshore regions (45 of 64 ASVs with cell volumes) were



**FIGURE 3** Trends in diatom diversity with latitude (a, b) and chl a concentration (c) from the Sargasso Sea (SS), Gulf Stream (GS), and coastal waters (CW). Richness of ASVs (a) and genera (b) did not differ significantly by region, and there was no correlation with chl a. Results of ANOVA tests (a, b) and linear regression (c) are in the upper right corner of each panel.

evenly split between the small (51%) and large (49%) size classes. In contrast, a majority of nearshore species were large (79%), falling within the  $>6,500 \mu\text{m}^3$

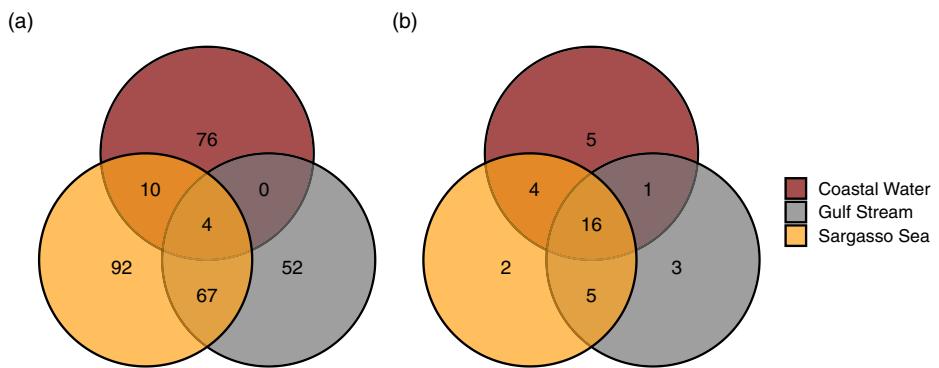
cell volume range ( $\sim 20 \mu\text{m}$  cell size; Table S3 in the Supporting Information).

More genera were observed found across all three regions (44%) compared with genera found only in coastal waters (14%) or offshore regions (28%; Figure 4b). Of the diatom genera found only in offshore regions, 18% were found only in the larger ( $>5 \mu\text{m}$ ) size fraction. While many genera were present across all regions, the number of ASVs per genus varied by region (Figure S5 in the Supporting Information). For example, *Pseudo-nitzschia*, *Nitzschia*, and *Bacteriastrum* had higher ASV diversity in the Sargasso Sea and Gulf Stream, while *Thalassiosira* and *Minidiscus* had higher diversity in coastal water.

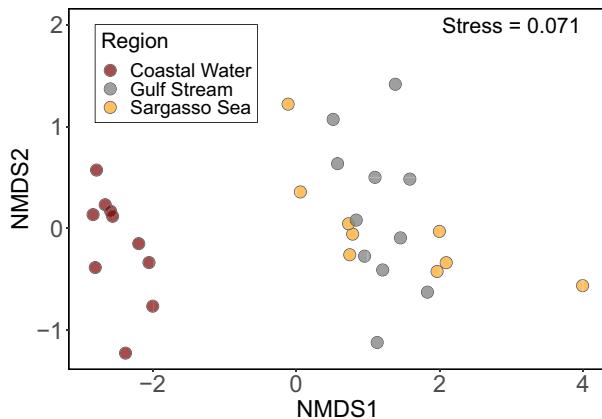
The Jaccard dissimilarity index was used to assess differences in community structure between samples. Community structure differed significantly with geographic region (ANOVA,  $F_{2,27} = 4.01, p = 0.001$ ; Figure 5), revealing one cluster formed by both Gulf Stream and Sargasso Sea samples and a second cluster of coastal water samples. There were no significant differences among samples with depth (ANOVA,  $F_{1,28} = 1.23, p = 0.161$ ) and the effect of filter pore size could not be tested, as this set of samples did not have homogenous variance (Levene's test,  $F_{1,28} = 12.80, p = 0.001$ ).

Environmental variables in the nearshore (coastal water) and offshore (Sargasso Sea and Gulf Stream) regions were correlated with ASV presence/absence in a tb-RDA. Here, phosphate concentration (ANOVA,  $F_{1,24} = 6.03, p = 0.002$ ) and temperature (ANOVA,  $F_{1,24} = 5.42, p = 0.002$ ) were significantly correlated with diatom community composition in nearshore versus offshore regions (Figure 6a). The first tb-RDA axis was significant (ANOVA-like permutation for redundancy analysis, RDA1,  $F_{1,24} = 6.05, p = 0.002$ ), explained 19.4% of the variability in the data set, and showed phosphate concentration increasing toward the coastal water and temperature increasing toward the offshore regions (Figure 6a). The second axis explained 3.5% of the variability in the dataset, was not significant (ANOVA,  $F_{1,24} = 1.09, p = 0.311$ ), and aligned primarily with differences between the Sargasso Sea and Gulf Stream regions. Mantel tests showed that Jaccard dissimilarity values of microbial communities were significantly correlated with geographic distance (Mantel test,  $r = 0.36, p = 0.0007$ ), although this correlation had weaker predictive power than the environmental variables of temperature (Mantel test,  $r = 0.73, p = 1\text{e}^{-4}$ ) and phosphate (Mantel test,  $r = 0.70, p = 1\text{e}^{-4}$ ).

Several diatom species contributed significantly to differences in community composition, which separated nearshore and offshore regions within the RDA (redundancy analysis,  $p < 0.05$ ; Figure 6b). Of the ASVs with the largest associations with the two major axes of variation (axis loadings  $> 0.14$ ), one set characterized nearshore coastal waters and included *Chaetoceros* sp., *Minidiscus trioculatus* (2 ASVs),



**FIGURE 4** Venn diagram of the number of ASVs (a) and number of genera (b) observed in the nearshore coastal waters and offshore (Gulf Stream and Sargasso Sea) regions.



**FIGURE 5** Non-metric multidimensional scaling of Jaccard dissimilarity of diatom community composition separated by region. The Jaccard index was significantly different across the three regions ( $F_{2,27}=4.01, p=0.001$ ) and met the assumption of homogeneity ( $F_{2,27}=0.13, p=0.881$ ).

*Pseudo-nitzschia delicatissima*, *Skeletonema marinii*, *Thalassiosira rotula*, and *Thalassiosira* sp. (three ASVs). The presence of these species was significantly, positively correlated with higher phosphate concentrations (ANOVA permutation for redundancy analysis,  $F_{1,24}=6.03, p=0.002$ ) and associated with lower temperatures (Figure 6b). A second set of ASVs characterized offshore regions and included *Nitzschia* sp., *Haslea* sp., *Pseudo-nitzschia* spp., and *Leptocylindrus convexus*.

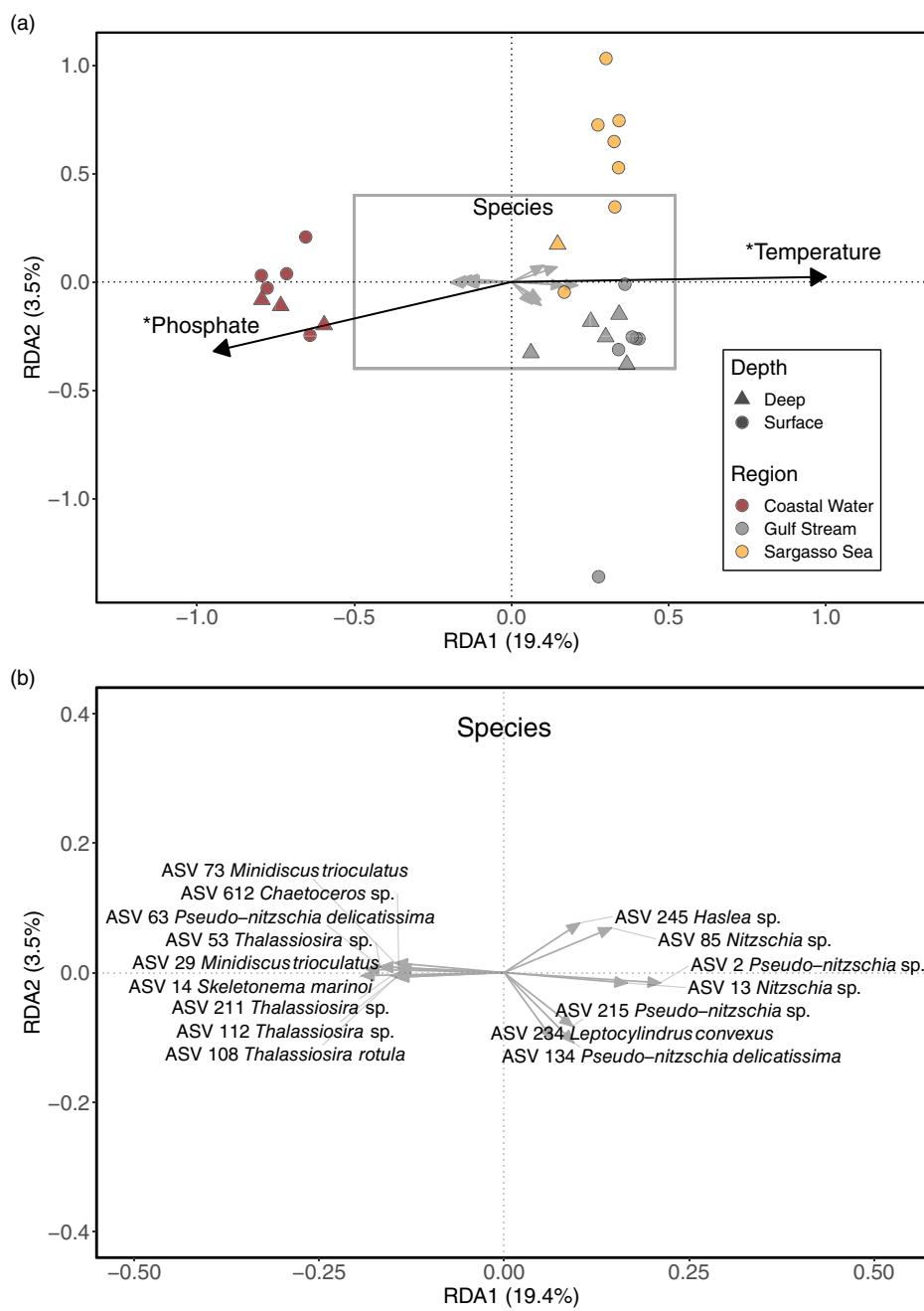
## DISCUSSION

Here we examined diatom 18S rDNA gene sequence variation along a transect in the western North Atlantic from high-nutrient coastal waters to the low-nutrient Sargasso Sea in order to expand observations of diatom diversity in oligotrophic waters and understand variations in community composition with environmental variables. Diatom ASVs were detected at every station along the transect from high-nutrient coastal waters to

oligotrophic open ocean waters. Although higher nutrient coastal waters are most often associated with flourishing diatom communities (Bracher et al., 2009; Smetacek, 2012), the highest number of both ASVs and genera were observed in the oligotrophic Sargasso Sea. Furthermore, there was no significant relationship between ASV richness and chlorophyll a concentration, supporting recent work that observed relatively high diatom richness even in low chlorophyll a waters (Fontaine & Rynearson, 2023).

Along the transect, ASV richness estimates were likely reasonable but incomplete reflections of actual diatom species diversity. On the one hand, intra-individual and intra-specific variation in the 18S rDNA gene could lead to overestimations of species richness (reviewed in Santoferara, 2019). However, these differences are not resolved well using the 18S rDNA V4 region we amplified (Zimmermann et al., 2011), and thus diatom species richness is unlikely to be strongly overestimated by this variation. On the other hand, the rDNA region we amplified distinguishes among many, but not all, diatom species (Zimmermann et al., 2011), leading to underestimates of diatom diversity. This is a known issue for species that share identical sequences at the V4 rDNA region and cannot be distinguished from each other, including species in frequently occurring and ecologically important diatom genera observed in our dataset (e.g., *Pseudo-nitzschia*, Ruggiero et al., 2022; *Skeletonema*, Luddington et al., 2012; *Thalassiosira*, Rynearson et al., 2020).

Overall, diatom richness, whether defined as the total number of ASVs or total genera per sample, did not change significantly from coastal to offshore waters, supporting previous studies that observed a comparable pattern (Caputi et al., 2019; Malviya et al., 2016). Notably, this pattern differs from two studies that observed lower diatom diversity in oligotrophic waters (Busseni et al., 2020; James et al., 2022), which may be explained by a variety of underlying causes. First, patterns of diversity may differ among locations due to oceanographic conditions. For example, the decreasing nearshore-offshore gradient of Shannon diversity



**FIGURE 6** Transformation-based redundancy analysis of diatom community composition highlighting environmental variables (a) and species (b) that best explain the relationship among water samples. Phosphate concentration ( $F_{1,24}=6.03, p=0.002$ ) and temperature ( $F_{1,24}=5.42, p=0.002$ ) were the only significant environmental variables (asterisk). Inset from gray box (a) expanded in B and showing only species with axis loadings greater than 0.14. Only axis RDA 1 was statistically significant (RDA1,  $F_{1,24}=6.05, p=0.002$ ). Surface samples corresponded to depths of 25 m (Sargasso Sea & Gulf Stream) or 5–15 m (coastal water) above the deep chlorophyll maximum (circles), and deep samples corresponded to depths taken at the deep chlorophyll maximum or below (40–80 m in SS & GS; 25 m in coastal water).

observed in the Southern California Current (James et al., 2022) was related to nitracline depth that was influenced by variations in regional upwelling. Second, different diversity metrics, such as species richness and the Shannon-Weiner index, reflect different aspects of community composition and are not directly comparable (Santini et al., 2017). Finally, patterns of diversity obtained may be related to the size range of

diatoms examined, which varies among studies, including those focused on diatoms  $>20\text{ }\mu\text{m}$  in diameter (Busseini et al., 2020). Here, 29% of the ASVs that could be identified to the species level had reported cell diameters of  $<20\text{ }\mu\text{m}$ . The prevalence of small diatoms in our study and the studies of others (James et al., 2022; Malviya et al., 2016) suggests that in the future, diatoms should be harvested on  $\leq 3\text{ }\mu\text{m}$  filters in both oligotrophic

and eutrophic waters to capture the full size spectrum of diatom species and allow for further investigation of diversity patterns across oceanographic regions.

Although diatom richness was indistinguishable across regions, there were significant differences in community composition. Dissimilarities in community composition were significantly correlated with geographic distance, although the predictive power of that relationship was only about half that of the environmental variables of phosphate concentration and temperature, supporting experimental work that has documented swift species sorting in response to nutrient availability and temperature (Anderson et al., 2022). Surface waters of the western Sargasso Sea are low in phosphate relative to other ecosystems (Cavender-Bares et al., 2001; Lomas et al., 2010; Wu, 2000), and N:P ratios in the offshore samples were well above the Redfield ratio (Redfield, 1958). Phytoplankton communities in the western Sargasso Sea frequently exhibit evidence of phosphorus stress, such as phosphorus sparing, and the utilization of organic phosphorus sources (Lomas et al., 2004, 2010; Van Mooy et al., 2009). Oligotrophic diatoms may also have novel low-nutrient adaptations, such as buoyancy control and migration to the nutricline (Arrieta et al., 2020; Falciatore & Bowler, 2002). Stress responses to phosphorus limitation can be taxon-specific (Lomas et al., 2004), and it may be that different diatom phosphorus stress adaptations (Dell'Aquila et al., 2020; Dyhrman et al., 2012) play a role in the significant correlation observed between phosphate concentrations and diatom community composition in this study. Community composition also varied significantly with temperature, and temperature gradients, such as the 18°C range observed in this study, are expected to select for taxa with differing thermal tolerances (Anderson et al., 2021; Anderson & Rynearson, 2020; Thomas et al., 2012). Diatoms surviving in relatively warmer, oligotrophic waters likely have a higher thermal maximum than coastal taxa of the colder northeast shelf, a factor known to contribute to diatom community structure and distribution (Anderson et al., 2021; Thomas et al., 2012). Several ASVs were responsible for explaining variation between offshore and nearshore community structure. For example, ASVs in the genera *Pseudo-nitzschia*, *Nitzschia*, and *Leptocylindrus* were positively correlated with both the lower phosphate concentrations and higher temperatures of the offshore region. In contrast, ASVs in the genera *Thalassiosira*, *Chaetoceros*, and *Skeletonema* were positively correlated with higher phosphate concentrations and lower temperatures of the nearshore region.

Although individual ASVs were associated with particular environments, over half of the genera in this dataset were present in both oligotrophic and eutrophic environments, suggesting that closely related species within the same genus can have distinct adaptations to

either environment. For example, some species within the cosmopolitan genera *Rhizosolenia* and *Chaetoceros* are known to form large blooms in eutrophic coastal waters (*Chaetoceros socialis*, Booth et al., 2002; and *Rhizosolenia delicatula*, Sournia et al., 1987), while other species in these same genera survive in oligotrophic waters by fulfilling their nitrogen requirements through symbioses with nitrogen-fixing cyanobacteria (*Rhizosolenia clevei*, Villareal, 1989; and *Chaetoceros compressus*, Gómez et al., 2005). Furthermore, up to 11% of all ASVs found in this study could potentially be associated with nitrogen-fixing cyanobacteria (Foster & Zehr, 2019), and many of those are known to have congeners that have no known associations with diazotrophs. A number of adaptations to both low N and P have been identified in diatoms from nutrient-rich coastal waters, including luxury nitrate uptake (Dortch, 1982; Dortch et al., 1984) and phospholipid substitution (Dyhrman et al., 2012; Martin et al., 2011; Zhang et al., 2022). However, studies of diatom nutrient physiology are rarely done with isolates from oligotrophic regions like the Sargasso Sea. Our data suggest that many of the oligotrophic species we identified are members of the same genera as well-studied eutrophic species, highlighting the value of evaluating adaptive responses in diatom isolates from warm, oligotrophic regions as they may carry different traits than isolates from colder, more nutrient-rich waters.

This dataset provided the opportunity to investigate whether shifts in community composition between oligotrophic and eutrophic waters were associated with shifts in species with different cell sizes. Cell volumes for species found in our study suggest less than a quarter of the diatom community in coastal waters was comprised of small cells compared to half of the diatom community in oligotrophic regions. Due to their high surface-area-to-volume ratios, nutrient uptake kinetics in small-celled species are advantageous in oligotrophic regions (Finkel et al., 2010). Recent work has pointed to the ecological and biogeochemical importance of small diatoms (Leblanc et al., 2018), particularly in the offshore waters of the western North Atlantic during the spring bloom (Bolaños et al., 2020). Although they are small, diatoms like *Minidiscus* can achieve rapid sinking rates primarily via cell aggregation, and thus can contribute significantly to export production (Leblanc et al., 2018; Richardson & Jackson, 2007). Notably, small diatoms are also being increasingly recognized as ecologically important inhabitants in high-nutrient waters (Arsenieff et al., 2020; Rynearson et al., 2020), waters in which we also identified numerous small-celled species, including those in the genera *Minidiscus* as well as *P. delicatissima* and *Thalassiosira minima*. Although analysis of cell volumes was limited to those ASVs identified to species level and whose cell size measurements were included in the database compiled by Leblanc et al. (2018), patterns of cell sizes

emerged between offshore and nearshore regions and provided an initial examination of cell size differences across nutrient gradients in the western North Atlantic.

Diatom abundance in the Sargasso Sea changes throughout an annual cycle (Hulbert et al., 1960; Steinberg et al., 2001), and the diatom community we identified represents a snapshot in time. Although taxonomic investigations of species composition in the Sargasso Sea are rare, the most species-rich time of year appears to be mid-April, just 2 weeks before our samples were collected, and studies have observed five species found in our study; *Bacteriastrum hyalinum*, *Chaetoceros decipiens*, *Guinardia flaccida*, *Leptocylindrus danicus*, and *Thalasionema frauenfeldii* (Hulbert et al., 1960; Hulbert & Rodman, 1963). The persistence of these diatom species over decades suggest they are long-term inhabitants of the oligotrophic Sargasso Sea. More broadly, we recovered 36 genera, including 13 genera identified in Hulbert et al. (1960) and Hulbert and Rodman (1963). Importantly, these historical studies reported that small centric diatoms were the most abundant cells at depth during the April bloom, but they were too small to be identified to species level. Our work suggests these may have included the species *Minidiscus trioculatus*, *Minutocellus polymorphus*, or *Thalassiosira minima* (for full species list see Table S4 in the Supporting Information). Following the April bloom, species diversity and abundance in surface waters were low throughout the remainder of the year (Hulbert et al., 1960), an observation supported by pigment data from the Bermuda Atlantic Time Series (Steinberg et al., 2001). Intermittent diatom-dominated blooms have been observed at other times of year in response to eddy-driven nutrient inputs (Allen et al., 2005; Krause et al., 2010; McGillicuddy et al., 2007). Our work suggests that potential seed populations for such blooms contain high diversity, with the majority of all ASVs identified here (70%) originating exclusively from oligotrophic waters.

Only a small percentage of ASVs (5%) were found in both nearshore and offshore regions and included several unidentified Bacillariophyceae, three species of *Guinardia* (*G. striata*, *G. flaccida*, and *G. delicatula*), *Minutocellus polymorphus*, *Cerataulina pelagica*, and *Chaetoceros costatus*, among others. Species in the genera *Minutocellus* and *Chaetoceros* are cosmopolitan (De Luca et al., 2019; Leblanc et al., 2018; Malviya et al., 2016) and able to survive across different ocean environments. In addition, *C. pelagica* has been found in a variety of environmental conditions, including both high-nutrient coastal and oligotrophic waters (Brun et al., 2015; Carstensen et al., 2015). The ability of certain diatom species to survive across disparate environments may relate to their metabolic plasticity and the flexibility of gene regulation reflecting a generalist strategy (Glibert, 2016; Margalef, 1978, 1979). Alternatively, there are also well-known examples of

ecologically distinct diatom species with identical 18S V4 rDNA regions (Cerino et al., 2005; Luddington et al., 2012), highlighting that some of the observed cosmopolitan ASVs may in fact represent different species along the nutrient and temperature gradients we sampled. The percentage of reads that could not be identified even to the genus level was 40%, consistent with previous work examining diatom diversity (Malviya et al., 2016) and highlighting large gaps in our ability to connect ASVs and their patterns of occurrence with species-specific or even genus-specific physiology.

## CONCLUSIONS

Understanding the diversity and distribution of diatoms is important in establishing their contribution to nutrient cycling and carbon export, considering that diatoms account for ~20% of global carbon cycling (Falkowski, 1998; Field, 1998; Tréguer et al., 2018) yet are understudied in open ocean oligotrophic regions that constitute a majority of the global ocean (Busseni et al., 2020; Malviya et al., 2016). Across this transect, just one-third of the ASVs could be identified to the species level (30%). This speaks to the need for continued work to connect described species with their rDNA gene sequences as well as for additional taxonomic and physiological research to identify and interrogate the metabolic capabilities of new species (e.g., Luddington et al., 2012). Strikingly, the percentage of reads that could be identified to species level was similar across the transect from nearshore to offshore, indicating no substantial difference in the ability to assign taxonomy to diatoms from the two habitats.

Our study determined that oligotrophic regions of the western North Atlantic supported taxonomically diverse diatom communities that were distinct from those in coastal waters. Resource partitioning through niche differentiation and competition between species likely facilitate this diversity, especially in oligotrophic environments where there is competition for scarce resources (Huisman & Weissing, 1999; Kerr et al., 2002; Menden-Deuer & Rowlett, 2014; Tilman et al., 1982). This species diversity might also allow for resilience in the face of environmental change through a diversity of physiological capacity or through functional redundancy that allows one diatom species to take the place of another without driving shifts in ecosystem functions like primary production or carbon export (Tilman et al., 2014). Given that oligotrophic regions are predicted to expand in coming years (Polovina et al., 2008), understanding patterns of diversity in low-nutrient, open ocean regions is important for projecting diatom diversity and resilience in the future ocean and recognizing the implications this has for global biogeochemical cycling.

## AUTHOR CONTRIBUTIONS

**Samantha P. Setta:** Conceptualization (equal); data curation (lead); formal analysis (equal); investigation (lead); methodology (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal).

**Sarah Lerch:** Investigation (supporting); visualization (supporting); writing – review and editing (equal). **Bethany D. Jenkins:** Conceptualization (equal); funding acquisition (equal); writing – review and editing (equal).

**Sonya T. Dyrhman:** Conceptualization (equal); funding acquisition (equal); writing – review and editing (equal).

**Tatiana A. Rynearson:** Conceptualization (equal); formal analysis (equal); funding acquisition (equal); investigation (supporting); methodology (supporting); project administration (lead); resources (lead); supervision (lead); validation (lead); visualization (equal); writing – original draft (supporting); writing – review and editing (equal).

## ACKNOWLEDGMENTS

Special thanks to B. Van Mooy and A. Ebling for determining phosphorus concentrations of water samples. Thanks to the captain and crew of the Atlantic Explorer, aboard which samples were collected. Thanks to E. McDermith, K. Gomes, E. Salomaki, I. Bishop, S. Anderson, M. Harke, M. Hernández Limón, H. Holm, and A. Ebling for help with sample collection. Thanks to J. Atoyan at the URI Genomics and Sequencing Center for her help and advice during sequencing sample preparation.

## FUNDING INFORMATION

This research was supported by National Science Foundation (NSF) awards 1558490 (T.A.R. & B.D.J.), 1655686 (T.A.R.), 1756816 (B.D.J) and 1558506 (S.T.D.), and by a grant from the Paul M. Angell Family Foundation (S.T.D.). The Genomics and Sequencing Center, University of Rhode Island and the Center for Computation and Visualization, Brown University were supported by NSF awards OIA-1004057 and OIA-1655221.

## DATA AVAILABILITY STATEMENT

Chlorophyll a (<https://www.bco-dmo.org/dataset/82094/8/data>) and phosphorus (<https://www.bco-dmo.org/dataset/739973/data>) data is accessible via BCO-DMO. Metabarcoding sequences used in this study can be found in the sequence read archive at the National Center for Biotechnology Information (PRJNA938782).

## ORCID

Samantha P. Setta  <https://orcid.org/0000-0001-9075-7573>

Sarah Lerch  <https://orcid.org/0000-0002-7905-1745>

Sonya T. Dyrhman  <https://orcid.org/0000-0003-2901-699X>

Tatiana A. Rynearson  <https://orcid.org/0000-0003-2951-0066>

## REFERENCES

- Alexander, H., Jenkins, B. D., Rynearson, T. A., & Dyrhman, S. T. (2015). Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *The Proceedings of the National Academy of Sciences of the United States of America*, 112, E2182–E2190.
- Allen, J. T., Brown, L., Sanders, R., Mark Moore, C., Mustard, A., Fielding, S., Lucas, M., Rixen, M., Savidge, G., Henson, S., & Mayor, D. (2005). Diatom carbon export enhanced by silicate upwelling in the Northeast Atlantic. *Nature*, 437(7059), 728–732.
- Anderson, S. I., Barton, A. D., Clayton, S., Dutkiewicz, S., & Rynearson, T. A. (2021). Marine phytoplankton functional types exhibit diverse responses to thermal change. *Nature Communications*, 12, 6413.
- Anderson, S. I., Franzè, G., Kling, J. D., Wilburn, P., Kremer, C. T., Menden-Deuer, S., Litchman, E., Hutchins, D. A., & Rynearson, T. A. (2022). The interactive effects of temperature and nutrients on a spring phytoplankton community. *Limnology and Oceanography*, 67(3), 634–645.
- Anderson, S. I., & Rynearson, T. A. (2020). Variability approaching the thermal limits can drive diatom community dynamics. *Limnology and Oceanography*, 65(9), 1961–1973.
- Armbrust, E. V. (2009). The life of diatoms in the world's oceans. *Nature*, 459, 185–192.
- Arrieta, J., Jeanneret, R., Roig, P., & Tuval, I. (2020). On the fate of sinking diatoms: The transport of active buoyancy-regulating cells in the ocean. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(20179), 20190529.
- Arsenieff, L., Le Gall, F., Rigaut-Jalabert, F., Mahé, F., Sarno, D., Gouhier, L., Baudoux, A. C., & Simon, N. (2020). Diversity and dynamics of relevant nanoplanktonic diatoms in the Western English Channel. *The ISME Journal*, 14, 1966–1981.
- Benitez-Nelson, C. R., Bidigare, R. R., Dickey, T. D., Landry, M. R., Leonard, C. L., Brown, S. L., Nencioli, F., Rii, Y. M., Maiti, K., Becker, J. W., Bibby, T. S., Black, W., Cai, W. J., Carlson, C. A., Chen, F., Kuwahara, V. S., Mahaffey, C., McAndrew, P. M., Quay, P. D., ... Yang, E. J. (2007). Mesoscale eddies drive increased silicate export in the subtropical Pacific Ocean. *Science*, 316(5827), 1017–1021.
- Bidigare, R., Marra, J., Dickey, T., Iturriaga, R., Baker, K., Smith, R., & Pak, H. (1990). Evidence for phytoplankton succession and chromatic adaptation in the Sargasso Sea during spring 1985. *Marine Ecological Progress Series*, 60, 113–122.
- Bolaños, L. M., Karp-Boss, L., Choi, C. J., Worden, A. Z., Graff, J. R., Haëntjens, N., Chase, A. P., Penna, A. D., Gaube, P., Françoise, M., Menden-Deuer, S., Westberry, T. K., O'Malley, R. T., Boss, E., Behrenfeld, M. J., & Giovannoni, S. J. (2020). Small phytoplankton dominate western North Atlantic biomass. *The ISME Journal*, 14, 1663–1674.
- Booth, B. C., Larouche, P., Bélanger, S., Klein, B., Amiel, D., & Mei, Z. P. (2002). Dynamics of *Chaetoceros socialis* blooms in the north water. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(22–23), 5003–5025.
- 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, 84–94.
- Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Rottgers, R., & Peeken, I. (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences*, 6, 751–764.
- Brun, P., Vogt, M., Payne, M. R., Gruber, N., O'Brien, C. J., Buitenhuis, E. T., le Quéré, C., Leblanc, K., & Luo, Y. W. (2015). Ecological niches of open ocean phytoplankton taxa: Niches of open ocean phytoplankton. *Limnology and Oceanography*, 60, 1020–1038.

- Brzezinski, M. A., & Nelson, D. M. (1995). The annual silica cycle in the Sargasso Sea near Bermuda. *Deep Sea Research Part I: Oceanographic Research Papers*, 42, 1215–1237.
- Busseni, G., Caputi, L., Piredda, R., Fremont, P., Hay Mele, B., Campese, L., Scalco, E., de Vargas, C., Bowler, C., d'Ovidio, F., Zingone, A., d'Alcalà, M. R., & Iudicone, D. (2020). Large scale patterns of marine diatom richness: Drivers and trends in a changing ocean. *Global Ecology Biogeography*, 29, 1915–1928.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581–583.
- Caputi, L., Carradec, Q., Eveillard, D., Kirilovsky, A., Pelletier, E., Pierella Karlusich, J. J., Rocha Jimenez Vieira, F., Villar, E., Chaffron, S., Malviya, S., Scalco, E., Acinas, S. G., Alberti, A., Aury, J., Benoiston, A., Bertrand, A., Biard, T., Bittner, L., Boccardo, M., & Wincker, P. (2019). Community-level responses to iron availability in open ocean plankton ecosystems. *Global Biogeochemical Cycles*, 33(3), 391–419.
- Carpenter, E., Montoya, J., Burns, J., Mulholland, M., Subramaniam, A., & Capone, D. (1999). Extensive bloom of a N<sub>2</sub>-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean. *Marine Ecological Progress Series*, 185, 273–283.
- Carstensen, J., Klais, R., & Cloern, J. E. (2015). Phytoplankton blooms in estuarine and coastal waters: Seasonal patterns and key species. *Estuarine, Coastal and Shelf Science*, 162, 98–109.
- Cavender-Bares, K. K., Karl, D. M., & Chisholm, S. W. (2001). Nutrient gradients in the western North Atlantic Ocean: Relationship to microbial community structure and comparison to patterns in the Pacific Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 48, 2373–2395.
- Cerino, F., Orsini, L., Sarno, D., Dell'Aversano, C., Tartaglione, L., & Zingone, A. (2005). The alternation of different morphotypes in the seasonal cycle of the toxic diatom *pseudo-nitzschia galaxiae*. *Harmful Algae*, 4, 33–48.
- Chestnut, T., Anderson, C., Popa, R., Blaustein, A. R., Voytek, M., Olson, D. H., & Kirshtein, J. (2014). Heterogeneous occupancy and density estimates of the pathogenic fungus *Batrachochytrium dendrobatidis* in waters of North America. *PLoS ONE*, 9, e106790.
- Cruaud, P., Vigneron, A., Fradette, M. S., Charette, S. J., Rodriguez, M. J., Dorea, C. C., & Culley, A. I. (2017). Open the Sterivex™ casing: An easy and effective way to improve DNA extraction yields. *Limnology and Oceanography Methods*, 15, 1015–1020.
- Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, 6, 226.
- De Luca, D., Kooistra, W. H. C. F., Sarno, D., Gaonkar, C. C., & Piredda, R. (2019). Global distribution and diversity of *Chaetoceros* (Bacillariophyta, Mediophyceae): Integration of classical and novel strategies. *PeerJ*, 7, e7410.
- Dell'Aquila, G., Zauner, S., Heimerl, T., Kahnt, J., Samel-Gondesen, V., Runge, S., Hempel, F., & Maier, U. G. (2020). Mobilization and cellular distribution of phosphate in the diatom *Phaeodactylum tricornutum*. *Frontiers in Plant Science*, 11, 579.
- Dortch, Q. (1982). Effect of growth conditions on accumulation of internal nitrate, ammonium, amino acids, and protein in three marine diatoms. *Journal of Experimental Marine Biology and Ecology*, 61(3), 243–264.
- Dortch, Q., Clayton, J. R., Thoresen, S. S., & Ahmed, S. I. (1984). Species differences in accumulation of nitrogen pools in phytoplankton. *Marine Biology*, 81(3), 237–250.
- Dyrhman, S. T., Jenkins, B. D., Rynearson, T. A., Saito, M. A., Mercier, M. L., Alexander, H., Whitney, L. P., Drzewianowski, A., Bulygin, V. V., Bertrand, E. M., Wu, Z., Benitez-Nelson, C., & Heithoff, A. (2012). The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS One*, 7, e33768.
- Falciatore, A., & Bowler, C. (2002). Revealing the molecular secrets of marine diatoms. *Annual Review of Plant Biology*, 53(1), 109–130.
- Falkowski, P. G. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science*, 281, 200–206.
- Field, C. B. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, 281, 237–240.
- Finkel, Z. V., Beardall, J., Flynn, K. J., Quigg, A., Rees, T. A. V., & Raven, J. A. (2010). Phytoplankton in a changing world: Cell size and elemental stoichiometry. *Journal of Plankton Research*, 32, 119–137.
- Fontaine, D. N., & Rynearson, T. A. (2023). Multi-year time series reveals temporally synchronous diatom communities with annual frequency of recurrence in a temperate estuary. *Limnology and Oceanography*. <https://doi.org/10.1002/limo.12400>
- Foster, R. A., & Zehr, J. P. (2019). Diversity, genomics, and distribution of phytoplankton-cyanobacterium single-cell symbiotic associations. *The Annual Review of Microbiology*, 73, 435–456.
- Glibert, P. M. (2016). Margalef revisited: A new phytoplankton mandala incorporating twelve dimensions, including nutritional physiology. *Harmful Algae*, 55, 25–30.
- Gloer, G. B., Mackeaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, 8, 2224.
- Goldman, J. C. (1993). Potential role of large oceanic diatoms in new primary production. *Deep Sea Research Part I: Oceanographic Research Papers*, 40, 159–168.
- Gómez, F., Furuya, K., & Takeda, S. (2005). Distribution of the cyanobacterium *Richelia intracellularis* as an epiphyte of the diatom *Chaetoceros compressus* in the western Pacific Ocean. *Journal of Plankton Research*, 27(4), 323–330.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Bouette, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Besco, N., Logares, R., ... Christen, R. (2012). The Protist ribosomal reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41, D597–D604.
- Henson, S. A., Cael, B. B., Allen, S. R., & Dutkiewicz, S. (2021). Future phytoplankton diversity in a changing climate. *Nature Communications*, 12, 5372.
- Huisman, J., & Weissing, F. J. (1999). Biodiversity of plankton by species oscillations and chaos. *Nature*, 402, 407–410.
- Hulbert, E. M., & Rodman, J. (1963). Distribution of phytoplankton species with respect to salinity between the coast of southern New England and Bermuda. *Limnology and Oceanography*, 8, 263–269.
- Hulbert, E. M., Ryther, J. H., & Guillard, R. R. L. (1960). The phytoplankton of the Sargasso Sea off Bermuda. *ICES Journal of Marine Science*, 25, 115–128.
- James, C. C., Barton, A. D., Allen, L. Z., Lampe, R. H., Rabines, A., Schulberg, A., Zheng, H., Goericke, R., Goodwin, K. D., & Allen, A. E. (2022). Influence of nutrient supply on plankton microbiome biodiversity and distribution in a coastal upwelling region. *Nature Communications*, 13, 2448.
- Jespersen, A. M., & Christoffersen, K. (1987). Measurements of chlorophyll-A from phytoplankton using ethanol as extraction solvent. *Archiv für Hydrobiologie*, 109, 445–454.
- Karentz, D., & Smayda, T. J. (1998). Temporal patterns and variations in phytoplankton community organization and abundance in Narragansett Bay during 1959–1980. *Journal of Plankton Research*, 20, 145–168.

- Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M., & Mahaffey, C. (2012). Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proceedings of the National Academy of Sciences*, 109, 1842–1849.
- Karl, D. M., & Tien, G. (1992). MAGIC: A sensitive and precise method for measuring dissolved phosphorus in aquatic environments. *Limnology and Oceanography*, 37, 105–116.
- Kerr, B., Riley, M. A., Feldman, M. W., & Bohannan, B. J. M. (2002). Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature*, 418, 171–174.
- Krause, J. W., Nelson, D. M., & Lomas, M. W. (2009). Biogeochemical responses to late-winter storms in the Sargasso Sea, II: Increased rates of biogenic silica production and export. *Deep Sea Research Part I: Oceanographic Research Papers*, 56, 861–874.
- Krause, J. W., Nelson, D. M., & Lomas, M. W. (2010). Production, dissolution, accumulation, and potential export of biogenic silica in a Sargasso Sea mode-water eddy. *Limnology and Oceanography*, 55, 569–579.
- Lampe, R. H., Wang, S., Cassar, N., & Marchetti, A. (2019). Strategies among phytoplankton in response to alleviation of nutrient stress in a subtropical gyre. *The ISME Journal*, 13(12), 2984–2997.
- Leblanc, K., Arístegui, J., Armand, L., Assmy, P., Beker, B., Bode, A., Breton, E., Cornet, V., Gibson, J., Gosselin, M. P., Kopczynska, E., Marshall, H., Peloquin, J., Piontovski, S., Poult, A. J., Quéguiner, B., Schiebel, R., Shipe, R., Stefels, J., ... Yallop, M. (2012). A global diatom database – Abundance, biovolume and biomass in the world ocean. *Earth System Science Data*, 4, 149–165.
- Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu de Madron, X., Bowler, C., Malviya, S., Thyssen, M., Grégori, G., Rembauville, M., Grosso, O., Poulin, J., de Vargas, C., Pujo-Pay, M., & Conan, P. (2018). Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nature Communications*, 9, 953.
- Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271–280.
- Lomas, M. W., Burke, A. L., Lomas, D. A., Bell, D. W., Shen, C., Dyhrman, S. T., & Ammerman, J. W. (2010). Sargasso Sea phosphorus biogeochemistry: An important role for dissolved organic phosphorus (DOP). *Biogeosciences*, 7, 695–710.
- Lomas, M. W., Swain, A., Shelton, R., & Ammerman, J. W. (2004). Taxonomic variability of phosphorus stress in Sargasso Sea phytoplankton. *Limnology and Oceanography*, 49, 2303–2309.
- Luddington, I. A., Kaczmarśka, I., & Lovejoy, C. (2012). Distance and character-based evaluation of the V4 region of the 18S rRNA gene for the identification of diatoms (Bacillariophyceae). *PLoS ONE*, 7(9), e45664.
- Malone, T., Pike, S. E., & Conley, D. J. (1993). Transient variations in phytoplankton productivity at the JGOFS Bermuda time series station. *Deep Sea Research Part I: Oceanographic Research Papers*, 40, 903–924.
- Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulin, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A., & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences*, 113(11), E1516–E1525.
- Mantel, N., & Valand, R. S. (1970). A technique of nonparametric multivariate analysis. *Biometrics*, 26, 547–558.
- Margalef, R. (1978). Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanologica Acta*, 1(4), 493–509.
- Margalef, R. (1979). Functional morphology of organisms involved in red tides, as adapted to decaying turbulence. *Developments in Marine Biology*, 1, 84–89.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17, 10–12.
- Martin, P., Van Mooy, B. A., Heithoff, A., & Dyhrman, S. T. (2011). Phosphorus supply drives rapid turnover of membrane phospholipids in the diatom *Thalassiosira pseudonana*. *The ISME Journal*, 5(6), 1057–1060.
- McGillicuddy, D. J., Anderson, L. A., Bates, N. R., Bibby, T., Buesseler, K. O., Carlson, C. A., Davis, C. S., Ewart, C., Falkowski, P. G., Goldthwait, S. A., Hansell, D. A., Jenkins, W. J., Johnson, R., Kosnyrev, V. K., Ledwell, J. R., Li, Q. P., Siegel, D. A., & Steinberg, D. A. (2007). Eddy/wind interactions stimulate extraordinary mid-ocean plankton blooms. *Science*, 316, 1021–1026.
- Menden-Deuer, S., & Rowlett, J. (2014). Many ways to stay in the game: Individual variability maintains high biodiversity in planktonic microorganisms. *Journal of the Royal Society Interface*, 11, 20140031.
- Nelson, D. M., & Brzezinski, M. A. (1997). Diatom growth and productivity in an oligo-trophic midocean gyre: A 3-yr record from the Sargasso Sea near Bermuda. *Limnology and Oceanography*, 42, 473–486.
- Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A., & Quéguiner, B. (1995). Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles*, 9, 359–372.
- Polovina, J. J., Howell, E. A., & Abecassis, M. (2008). Ocean's least productive waters are expanding. *Geophysical Research Letters*, 35(3).
- Redfield, A. C. (1958). The biological control of chemical factors in the environment. *American Journal of Science*, 46(3), 205–221.
- Richardson, T. L., & Jackson, G. A. (2007). Small phytoplankton and carbon export from the surface ocean. *Science*, 315, 838–840.
- Ruggiero, M. V., Kooistra, W. H. C. F., Piredda, R., Sarno, D., Zampicinini, G., Zingone, A., & Montresor, M. (2022). Temporal changes of genetic structure and diversity in a marine diatom genus discovered via metabarcoding. *Environmental DNA*, 4, 763–775.
- Rynearson, T. A., Flickinger, S. A., & Fontaine, D. N. (2020). Metabarcoding reveals temporal patterns of community composition and realized thermal niches of *Thalassiosira* spp. (Bacillariophyceae) from the Narragansett Bay long-term plankton time series. *Biology*, 9, 19.
- Santini, L., Belmaker, J., Costello, M. J., Pereira, H. M., Rossberg, A. G., Schipper, A. M., Ceaşu, S., Dornelas, M., Hilbers, J. P., Hortal, J., Huijbregts, M. A. J., Navarro, L. M., Schifflers, K. H., Visconti, P., & Rondinini, C. (2017). Assessing the suitability of diversity metrics to detect biodiversity change. *Biological Conservation*, 213, 341–350.
- Santoferrara, L. F. (2019). Current practice in plankton metabarcoding: Optimization and error management. *Journal of Plankton Research*, 41, 571–582.
- Smetacek, V. (2012). Making sense of ocean biota: How evolution and biodiversity of land organisms differ from that of the plankton. *Journal of Biosciences*, 37, 589–607.
- Sournia, A., Birrien, J. L., Douville, J. L., Klein, B., & Viollier, M. (1987). A daily study of the diatom spring bloom at Roscoff (France) in 1985. I. the spring bloom within the annual cycle. *Estuarine, Coastal and Shelf Science*, 25(3), 355–367.
- Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., & Knap, A. H. (2001). Overview of the US JGOFS Bermuda Atlantic time-series study (BATS): A decade-scale look at ocean biology and biogeochemistry. *Deep Sea Research Part II: Topical Studies in Oceanography*, 48, 1405–1447.
- Thomas, M. K., Kremer, C. T., Klausmeier, C. A., & Litchman, E. (2012). A global pattern of thermal adaptation in marine phytoplankton. *Science*, 338, 1085–1088.

- Tilman, D., Isbell, F., & Cowles, J. M. (2014). Biodiversity and ecosystem functioning. *Annual Review of Ecology, Evolution, and Systematics*, 45, 471–493.
- Tilman, D., Kilham, S. S., & Kilham, P. (1982). Phytoplankton community ecology: The role of limiting nutrients. *Annual Review of Ecology, Evolution, and Systematics*, 13, 349–372.
- Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., Bittner, L., Dugdale, R., Finkel, Z., Iudicone, D., Jahn, O., Guidi, L., Lasbleiz, M., Leblanc, K., Levy, M., & Pondaven, P. (2018). Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience*, 11, 27–37.
- Van Mooy, B. A. S., Fredricks, H. F., Pedler, B. E., Dyhrman, S. T., Karl, D. M., Kobližek, M., Lomas, M. W., Mincer, T. J., Moore, L. R., Moutin, T., Rappé, M. S., & Webb, E. (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature*, 458, 69–72.
- Villareal, T. A. (1989). Division cycles in the nitrogen-fixing *Rhizosolenia* (Bacillariophyceae)-*Richelia* (Nostocaceae) symbiosis. *British Phycological Journal*, 24(4), 357–365.
- Villareal, T. A., & Lipschultz, F. (1995). Internal nitrate concentrations in single cells of large phytoplankton from the Sargasso Sea. *Journal of Phycology*, 31, 689–696.
- Whitney, L. P., Lins, J. J., Hughes, M. P., Wells, M. L., Chappell, P. D., & Jenkins, B. D. (2011). Characterization of putative iron responsive genes as species-specific indicators of iron stress in *Thalassiosiroid* diatoms. *Frontiers in Microbiology*, 2, 234.
- Wu, J. (2000). Phosphate depletion in the western North Atlantic Ocean. *Science*, 289, 759–762.
- Zhang, K., Li, J., Wang, J., Lin, X., Li, L., You, Y., Wu, X., Zhou, Z., & Lin, S. (2022). Functional differentiation and complementation of alkaline phosphatases and choreography of DOP scavenging in a marine diatom. *Molecular Ecology*, 31, 3389–3399.
- Zhu, F., Massana, R., Not, F., Marie, D., & Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, 52, 79–92.
- Zimmermann, J., Jahn, R., & Gemeinholzer, B. (2011). Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms, Diversity & Evolution*, 11, 173–192.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Figure S1.** Rarefaction curves of Bacillariophyceae sequences in all samples (A) and samples with less than 1,000 sequences (B). Color of boxes indicates region and stations corresponding to listed sample names are shown in Figure S2. Samples SPS05, SPS12, & SPS26 were below the 200-sequence cutoff (red line) and were removed from further analysis.

**Figure S2.** Relative abundances of center log ratio transformed reads following correction for zero counts (zCompositions package) using the pipeline described in Gloor et al. (2017) at each station (bottom x-axis) and across region (top x-axis). Sequences that could not be

identified to genus are grouped under the label “Class Bacillariophyceae” and sequences that represented <1% of reads are group under the label “Abnd < 1%.” Sequences were identified using the PR2 database.

**Figure S3.** Total number of sequencing reads in each sample for all taxa (A) and identified as the class Bacillariophyceae (B). Region of each sample and station is indicated in the top box of the figure. The red dashed line indicates the 200-sequence minimum used to prevent undersampling (see Methods section).

**Figure S4.** Percent occurrence of each ASV in A) coastal water (CW), Gulf Stream (GS), Sargasso Sea (SS); B), in CW and SS; and C), in GS and SS. Percent occurrence is the # of times an ASV was seen in each region normalized to the number of samples in each region.

**Figure S5.** Number of ASVs of each genus in coastal water, Gulf Stream, and the Sargasso Sea.

**Data S1.** Datasheet includes sample location and collection information, ctd data, and nutrient concentrations from the AE1812 Cruise associated with diatom richness data.

**Table S1.** Mean and standard deviation (parentheses) of dissolved inorganic nitrogen (DIN), phosphate, and chlorophyll a (Chl a) concentrations in the Sargasso Sea (SS), Gulf Stream (GS), and coastal waters (CW).

**Table S2.** Tukey post-hoc test results for phosphate and chlorophyll a concentrations for the Sargasso Sea (SS), Gulf Stream (GS), and coastal waters (CW).

**Table S3.** Number and percent of ASVs observed only in the nearshore region (coastal water) or offshore region (Sargasso Sea & Gulf Stream) within each cell volume ( $\mu\text{m}^3$ ) range. Only ASVs (64) with published cell volumes from LeBlanc et al. (2012) and observed only in offshore or nearshore regions are included in the table.

**Table S4.** Species list of small diatoms found in our study, the cell shape symmetry (centric vs. pennate), average cell volume ( $\mu\text{m}^3$ ; LeBlanc et al., 2012).

**How to cite this article:** Setta, S. P., Lerch, S., Jenkins, B. D., Dyhrman, S. T., & Rynearson, T. A. (2023). Oligotrophic waters of the Northwest Atlantic support taxonomically diverse diatom communities that are distinct from coastal waters. *Journal of Phycology*, 00, 1–15. <https://doi.org/10.1111/jpy.13388>