LIMNOLOGY
and
OCEANOGRAPHY: METHODS

ASLO

Limnol. Oceanogr.: Methods 2023

doi: 10.1002/lom3.10575

Comparison of advanced methodologies for diatom identification within dynamic coastal communities

Emily Pierce [1,2] **Olivia Torano,**[1] **YuanYu Lin** [1] **Astrid Schnetzer,**[2] **Adrian Marchetti** [1]*

[1]Department of Earth, Marine and Environmental Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, North Carolina, USA

Abstract

Diatom community composition has a critical influence on global ocean health and ecological processes. Developing accurate and efficient methods for diatom identification under dynamic environmental conditions is essential to understanding the implications of diatom community changes. Two developing methods for identifying and enumerating phytoplankton, cell imaging and molecular sequencing, are experiencing rapid advancements. This study aims to compare diatom taxonomic composition results within natural assemblages derived from rapidly advancing methods, FlowCam imaging and metabarcoding of the V4 region of the 18S rRNA gene, with traditional light microscopy cell counting techniques. All three methods were implemented in tandem to analyze changes in dynamic diatom assemblages within simulated upwelling experiments conducted in the California upwelling zone. The results of this study indicate that, summed across all samples, DNA sequencing detected four times as many genera as morphology-based methods, thus supporting previous findings that DNA sequencing is the most powerful method for analyzing species richness. Results indicate that all three methods returned comparable relative abundance for the most abundant genera. However, the three methods did not return comparable absolute abundance, primarily due to barriers in deriving quantities in equal units. Overall, this study indicates that at the semi-quantitative level of relative abundance measurements, FlowCam imaging and metabarcoding of the V4 region of the 18S rRNA gene yield comparable results with light microscopy but at the qualitative and quantitative levels, enumeration metrics diverge, and thus method selection and cross-method comparison should be performed with caution.

Marine phytoplankton play key roles in aquatic ecosystems. Diatoms are an especially important group of marine phytoplankton as they account for as much as a quarter of global productivity (Nelson et al. 1995; Falkowski et al. 1998; Field et al. 1998; Sarthou et al. 2005), serve as drivers of the marine biological carbon pump (Ducklow et al. 2001; Agustí et al. 2015; Tréguer et al. 2018), and have the ability to form harmful algal blooms (Bates et al. 1989; Bates and Trainer 2006; Smith et al. 2018). Assessing diatom community composition, which describes the distribution and abundance of specific taxa within an assemblage, is a critical tool for evaluating ecosystem health, particularly in the face of rapid climate change (Smetacek and Cloern 2008; Armbrust 2009; Marinov et al. 2013).

A traditional approach to community composition analysis of large phytoplankton taxonomic groups (i.e., diatoms) has consisted of identification and enumeration based on morphological characteristics via light microscopy (Tomas 1997; Morales et al. 2001; Dunn and Everitt 2004). This method is associated with several practical drawbacks including subjective identification, extensive training requirements, and lengthy analysis time (Archibald 1984; Prygiel et al. 2002; Blanco 2020). As technology has advanced, many new methods for both high-throughput imaging and sequencing have emerged and evolved over time (Benfield 2007; Shokralla et al. 2012). These new methodologies often produce more rapid results than traditional approaches and can offer additional insights such as the detection of cryptic or rare species (Bickford et al. 2007; Jerde et al. 2011). While advancing methodologies present exciting new information and opportunities, it is critical to understand these new methods in context of one another (Malviya et al. 2016; Pierella Karlusich et al. 2020). Advancing methodologies offer different qualitative

*Correspondence: amarchetti@unc.edu

Author Contribution Statement: EP, AM, and AS conceived the study. EP, OT, YL, and AM contributed to data collection. EP, OT, YL, AS, and AM contributed to data analysis and interpretation in addition to drafting and revising the manuscript.

Additional Supporting Information may be found in the online version of this article.

and quantitative capabilities that require validation to ensure efficient and accurate phytoplankton community assessment as well as comparability and continuity with previous phytoplankton composition approaches.

The FlowCam (Yokogawa Fluid Imaging Technologies) is one example of a tool in a field of rapidly evolving imaging technology. The FlowCam combines a microscope with a camera to capture a live stream of particle images as they flow through a glass flow cell (Sieracki et al. 1998). Depending on the objective and flow cell size combination, the FlowCam can typically analyze particles between 10 and 600 μm (Karlson et al. 2010). The FlowCam outputs a collection of particle images along with metadata on each particle's shape and size. A software package can be applied to semi-automatically classify and sort particle images. The FlowCam provides several practical advantages because it can be easily transported and employed in the field and its rapid sampling capabilities allows for real-time image output (Yokogawa Fluid Imaging Technologies 2011; Spaulding et al. 2012). In addition, the FlowCam is well suited to analyzing live samples which may help to resolve several biases introduced by cell shrinkage and distortion associated with preservation for light microscopy (Zarauz and Irigoien 2008). Like the traditional light microscopy approach, the FlowCam relies primarily on cell features for phytoplankton identification, and thus is best suited to large, morphologically distinct phytoplankton groups (Lombard et al. 2019; Kenitz et al. 2023).

Amplicon sequencing is another evolving methodology for phytoplankton identification and community composition analysis (Hebert et al. 2003; Caron et al. 2012). The most commonly employed DNA marker gene for taxonomic identification and other phylogenetic studies is the small subunit (18S) ribosomal RNA gene (Medlin et al. 1988; Countway et al. 2005; Hajibabaei et al. 2007; Jahn et al. 2007). All eukaryotic phytoplankton, including diatoms, possess at least one copy of the 18S rRNA gene which contains highly conserved as well as variable regions unique to specific taxa (Woese 1987; Dunthorn et al. 2012). Thus, the 18S rRNA gene has long served as a primary target for sequencing to derive community composition data including rare taxa (Hebert et al. 2003; Zhan et al. 2013). Gene sequences are classified using similarity with known isolates included within reference databases, largely removing subjectivity used in solely morphological-based classification methods (Zimmermann et al. 2014). However, due to variation in gene copy numbers (Amend et al. 2010; Kembel et al. 2012), amplicon sequencing returns only semi-quantitative relative abundance results without the use of an additional step to achieve quantitative metrics, for which there is not yet a commonly employed field standard approach (Medinger et al. 2010; Stoeck et al. 2014; Gong and Marchetti 2019). Thus, sequencing methods are somewhat limited in producing quantitative results that other methods allow.

There has been substantial research comparing results from advancing methods for phytoplankton community composition analysis to results from traditional light microscopy techniques (Stern et al. 2018; Owen et al. 2022). Studies comparing phytoplankton cell abundance results for FlowCam and light microscopy concluded that the two methods yield comparable results for quantitative abundance (Álvarez et al. 2014; Bergkemper and Weisse 2018). However, several studies also found incongruencies in the number of genera detected between FlowCam and light microscopy methods (Le Bourg et al. 2015; Hrycik et al. 2019). The FlowCam can be ideal for rapid analysis where coarse taxonomic resolution at the genus level is acceptable (Hrycik et al. 2019). Studies comparing phytoplankton community composition derived from metabarcoding of the 18S rRNA gene with that from light microscopy have produced mixed results (Xiao et al. 2014; Kittelmann et al. 2015; Malviya et al. 2016; Mora et al. 2019; Gong et al. 2020; McKeigen et al. 2022). There is limited research comparing phytoplankton community composition across these two methods, leaving a gap in our current understanding of how these new and rapidly evolving methods compare with one another.

This study aims to compare diatom community composition within highly diverse diatom assemblages derived from three different identification approaches—FlowCam imaging, 18S rRNA gene amplicon sequencing, and traditional light microscopy. Samples that were utilized for this comparison were collected from a coastal upwelling community that had been exposed to experimental manipulation by simulating upwelling combined with iron additions and removal; thus, yielding communities that varied broadly in diatom abundances and species diversity. Details on varying aspects of the incubation study are published elsewhere (Lin et al. 2023). The results of this method comparison study will help inform future method choices and are meant to aid intercalibration of phytoplankton community composition using these approaches.

Methods

Sampling locations and experimental design

Incubation experiments were conducted onboard the *R/V Oceanus* from 27 May to 6 June 2019. Experiments were conducted at a site overlying a wide continental shelf location, 41°00.896′N 124°25.165′W, and a narrow continental shelf location, 35°55.329′N, 121°32.439′W, during non-upwelling, relaxation conditions resulting in four total experiments. In the first set of experiments (termed iron grow-out) natural phytoplankton assemblages were obtained from seawater collected at the 8–10°C isotherm (90 m at wide shelf and 80 m at narrow shelf). Cubitainers (10 L) were filled in triplicates with one of the following inoculation treatments: iron (+Fe), desferrioxamine B (+DFB), an iron chelating agent that limits the uptake of dissolved iron by phytoplankton, and an unamended control (Ctl). In the second set of experiments (termed seed populations) seawater was collected at the 10°C isotherm (90 m at wide shelf and 80 m at narrow shelf) and at the near surface (15 m at both sites). Cubitainers were filled in

triplicate with deep water, surface water, or a mixture of the two. All filled cubitainers were incubated in on-deck tanks circulated with 8–10°C water using Aqualogic water chillers and shaded to achieve 30% of the incident irradiance. Cubitainers were harvested at two time points for the Fe grow-out experiments and one time point for the seed population experiments in addition to the initial communities (T_0). Overall, samples were collected from one cubitainer triplicate at each timepoint resulting in a total of 23 samples used for method comparison between FlowCam, light microscopy, and DNA sequencing (Supporting Information Table S3). Subsampling for each of the following parameters relevant to this study were collected: size-fractionated chlorophyll *a* (Chl *a*), DNA, whole water for FlowCam measurements, and preservations for light microscopy. As the focus of this study is to compare diatom community composition assessment across methods, the effects of the different experiments and treatments on community composition are not reported here.

Chl *a* concentrations

Approximately 400 mL from each sample was collected in a dark bottle and filtered onto a polycarbonate filter (5 μm pore size, 47 mm) and subsequently onto a Whatman GF/F filter (0.7 μm nominal pore size, 25 mm) using a series filter cascade and a vacuum pump. Filters were collected and stored at −20°C until onshore analysis. GF/F Filters were placed in 6 mL of 90% acetone and stored at −20°C for 24 h before analysis. Samples were analyzed using the acidification method on a Turner 10AU fluorometer blanked with a 90% acetone solution (Parsons et al. 1984).

FlowCam diatom enumeration

Approximately 25 mL of seawater was collected from each sample and processed with the FlowCam. All necessary settings for reproducing the FlowCam workflow are described below (Owen et al. 2022). The whole water samples were analyzed using a Portable Series FlowCam with a peristaltic pump set to autoimage mode using a ×10 objective with a standard 100 μm glass flow cell installed. The selected objective and flow cell combination allowed for a particle size range of 10–100 μm (Karlson et al. 2010). Particle images were captured in autofocus mode using a color camera set to 1280 × 960 pixels resolution and 10 frames per second maximum frame rate. Unpreserved samples were kept in the dark until analysis within approximately 12 h of collection. FlowCam analysis yielded a set of images associated with each sample. Autoimage mode captures images of a subarea of the flow cell at a set shutter speed and saves the particles captured in each image. The number of particles is normalized by the volume imaged to produce a sample density (particles mL^{-1}). Any images not containing particles (i.e., bubbles or smears on the flow cell caused by a disruption of the calibration setting) were deleted from the dataset and the particle density was corrected. Remaining images were automatically categorized to diatom genus using Yokogawa Fluid Imaging VisualSpreadsheet software combined with manual identification (Tomas 1997; Yokogawa Fluid Imaging Technologies 2011). Sample densities ranged between 2 and 2143 particles mL^{-1} with an average density of 673 particles mL^{-1} and median of 276 particles mL^{-1}.

18S rRNA metabarcoding

Approximately 1 L of seawater from each sample was filtered onto a Super-450 Pall filter (0.45 μm pore size, 47 mm) using a vacuum pump. Filters were flash frozen in liquid nitrogen and stored at −80°C until analysis onshore. Before DNA was extracted, 2.5×10^7 copies of the pEXPORTS plasmid were added to each sample, serving as an internal standard to allow for absolute quantification of reads (Tkacz et al. 2018; Lin et al. 2019; Gomes 2020). The pEXPORTS plasmid was constructed to resemble the natural length and structure of the 18S rRNA gene V4 region, while avoiding sequence homology (Gomes 2020). DNA was extracted using the Qiagen DNeasy Plant Mini Kit. The V4 region fragment was amplified using a set of dual index fusion primers with "heterogeneity spacers" included to improve amplicon library sequence heterogeneity (Kozich et al. 2013; Fadrosh et al. 2014). Each PCR was completed in triplicate for each sample and consisted of 0.006U TaKaRa Taq DNA polymerase, 2X PCR buffer, 2X dNTP mixture, 0.4 μL of each primer, and 2 μL of DNA template. PCR was conducted at 98°C for 5 min, followed by 30 three-step cycles consisting of 94°C for 30 s, 57°C for 45 s and 72°C for 1 min, and a last step of 72°C for 2 min. PCR triplicates were pooled in equal volume (20 μL from each triplicate) and purified using the Qiagen QIAquick DNA Purification Kit. DNA concentration for each pooled sample was measured using the Invitrogen Qubit 4 fluorometer and Qubit dsDNA HS Assay Kit. Each pooled sample was normalized to 17 nmol L^{-1} concentration and then pooled in equal volumes to produce a compiled library sample. The compiled library was sequenced by Azenta Life Sciences using an Illumina MiSeq sequencing platform. All bioinformatics processing was performed using the QIIME 2 software package (Caporaso et al. 2010; Bolyen et al. 2019). Reads (2 × 300 base pairs) were demultiplexed by custom scripts in QIIME2 and Python. Demultiplexed sequence reads were trimmed to ensure quality and merged using the cutadapt plugin (Martin 2011) and the q2-vsearch plugin, respectively (Rognes et al. 2016). Amplicon sequence variants (ASVs) were identified and clustered using a 97% similarity threshold with the DADA2 plugin (Caron et al. 2009; Callahan et al. 2016). Then ASVs were taxonomically classified using the PR2 18S v.4.13.0 gene taxonomic reference dataset (Guillou et al. 2012). While ASVs were classified to the species level, taxonomic results were aggregated by genus for comparison with the morphologically based identification approaches. A count of plasmid internal standard reads in each sample was derived by blasting the plasmid sequence to each sample sequence file. The count of plasmid reads was then used in the following equation adapted from Lin et al. 2019 (Eq. 1) to quantify the gene copies mL^{-1} in each sample (Supporting Information Table S1). Sequences were submitted to the NCBI Sequence Read Archive under project number PRJNA966115.

$$\text{Absolute abundance}\left(\text{copies mL}^{-1}\right) = \frac{\#\text{taxon reads} \times \#\text{plasmid copies added}}{\#\text{plasmid reads recovered} \times \text{volume filtered (mL)}}. \quad (1)$$

Light microscopy counts

Approximately 100 mL of seawater from each sample was collected in glass bottles and inoculated with Lugol's iodine solution to achieve a final 2% concentration. Samples were then stored in the dark at 4°C until analysis onshore. A subsample was prepared for analysis by settling 25, 50, or 100 mL of sample using an Utermöhl settling chamber (Utermöhl 1958). The volume settled for each sample was chosen based on the Chl *a* concentration of that sample. Slides were analyzed on an Olympus CKX-31 inverted microscope under ×200 magnification. Diatoms cells were enumerated and classified in 10 randomly selected fields of view per sample, with the exception of three low biomass samples (1, 13, and 11) in which the same approach was utilized for 15 fields of view. The number of cells counted for each sample ranged between 30 and 2584 cells with an average count of 776 cells per sample and a median of 365 cells per sample.

Statistical analysis

The statistical analysis for this study was completed using the PRIMERv7 (Clarke and Gorley 2015) and R (R Core Team 2020). Data visualizations were produced using the ggplot2 package (Wickham 2011). Samples 1, 2, 3, and 6 contained fewer than 100 diatom cells counted via light microscopy or fewer than 100 particles imaged via FlowCam, and were therefore removed from statistical analysis to prevent bias in the method comparison from inaccurate cell counts. Data from each sample analyzed by each method were normalized by the sum abundance of that sample to derive relative abundance datasets for each method. Absolute abundance data were $\log(x + 1)$ transformed to derive comparable absolute abundance for each genus in each sample. A Bray–Curtis similarity matrix was created for relative and absolute

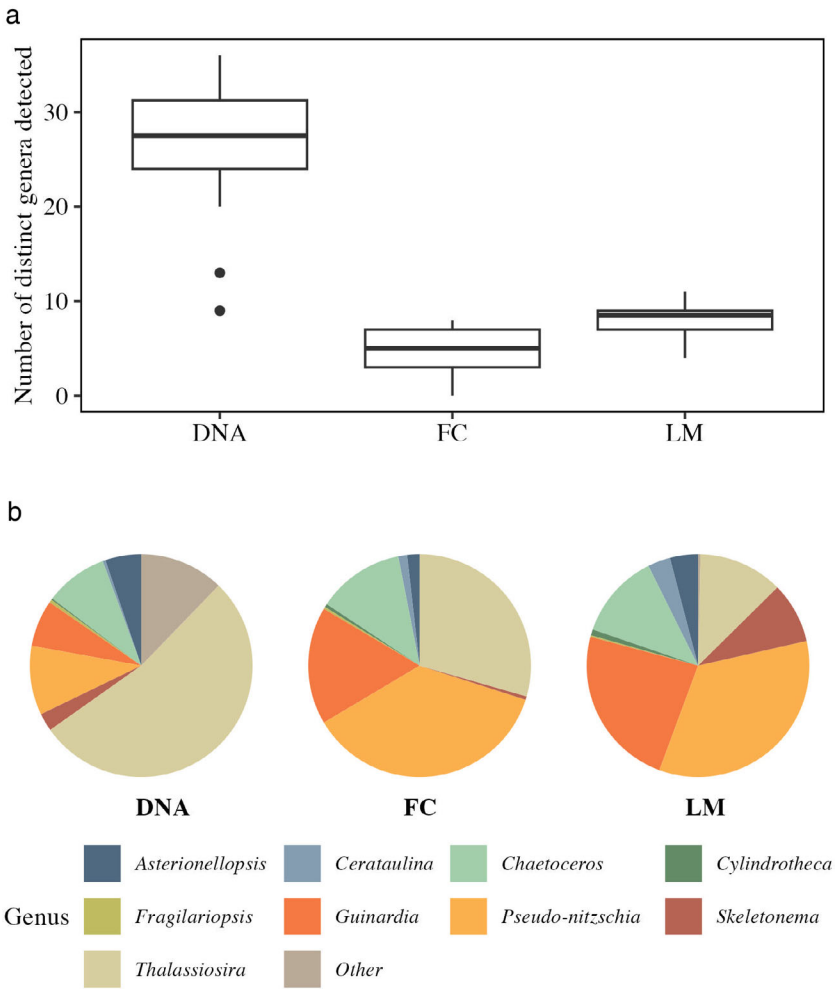

Fig. 1. A comparison across taxonomic approaches of the total number of diatom genera detected in each sample ($n = 23$) (**A**) and the relative proportion of each of the dominant diatom genera (**B**).

abundances. This matrix was then used with two-way crossed ANOSIM tests to quantify the level of similarity between methods using PRIMERv7 (Clarke 1993). The ANOSIM without sample replication test provided Spearman correlation values (p_{av}) that represent the effect size of method on diatom composition changes across samples with 0 representing no effect and 1 representing a very strong effect. The ANOSIM test also provides a p value which was considered significant using an alpha level of 0.05 Non-metric multidimensional scaling (NMDS) plots were made for relative abundance data, absolute abundance data, and presence/absence transformed data using the vegan package v2.6-2 in R (Dixon 2003). Distances between points in the NMDS plots is representative of dissimilarity between community composition of those samples. Linear relationships between datasets of each method pair were constructed for absolute and relative abundances using the lm function in the R stats package v.4.0.3.

Results

Number of detected diatom genera

Each of the three methods used detected a different number of diatom genera (Fig. 1A). Sequencing allowed for the detection of 45 unique diatom genera among all the samples, microscopy provided the detection of 11 different diatom genera, and FlowCam analysis provided the detection of nine different diatom genera (Table 1).

The nine genera detected using FlowCam analysis were also detected using light microscopy and DNA sequencing and are subsequently referred to as the commonly shared genera. While DNA sequencing detected the highest number of genera, 89% of 18S rRNA reads were from commonly shared diatom genera (Fig. 1B).

Given that DNA sequencing was conducted using a substantially larger volume of water than FlowCam or light microscopy, it is not surprising that the method detected more genera. Of the genera detected only from DNA sequencing, all but four genera each made up less than 1% of diatom abundance on average across samples, representing relatively rare taxa in the communities. The four genera detected by DNA that were not rare across the sample communities include *Actinocyclus*, *Minidiscus*, *Minutocellus*, and *Rhizosolenia*, respectively, making up on average 2%, 1%, 2%, and 4% of diatom communities across samples. The lack of detection of these four genera in FlowCam or light microscopy samples may highlight some of the disadvantages of those methods compared to DNA sequencing. For example, *Minidiscus* and *Minutocellus* are nanoplanktonic diatoms typically below 5 μm in diameter, below the size range of detection for FlowCam or light microscopy (Vaulot et al. 2008). Their detection using DNA sequencing highlights the unique suitability of the DNA sequencing approach to capture small-sized diatoms compared to FlowCam and light microscopy methods. *Actinocyclus* is not exceptionally different from certain species of *Thalassiosira* when viewed from the valve view. Thus, a potential explanation for the lack of detection of *Actinocyclus* using the FlowCam and light microscopy may have been misidentification due to non-distinct morphological features highlighting the disadvantage of subjective, morphological based methods. Finally, many of the sequences attributed to the genera *Rhizosolenia* were also attributed to the species *delicatula*. *Rhizosolenia delicatula* is a basionym for *Guinardia delicatula* (Tomas 1997), therefore, the DNA sequencing results may have inflated the abundance of the genus *Rhizosolenia* and underestimated the presence of the genus *Guinardia*. The misidentification of *Rhizosolenia* sequence reads highlights the disadvantage for DNA sequencing relying on reference databases that may be out of date, incomplete, or inaccurate.

Table 1. List of all diatom genera detected across samples for each method.

FlowCam	Light microscopy	DNA sequencing
Pseudo-nitzschia, *Thalassiosira*, *Chaetoceros*, *Asterionellopsis*, *Cylindrotheca*, *Cerataulina*, *Guinardia*, *Skeletonema*, *Fragilariopsis*	*Pseudo-nitzschia*, *Thalassiosira*, *Chaetoceros*, *Asterionellopsis*, *Cylindrotheca*, *Cerataulina*, *Guinardia*, *Skeletonema*, *Fragilariopsis*, *Eucampia*, *Navicula*	*Actinocyclus*, *Actinoptychus*, *Arcocellulus*, *Asterionella*, *Asterionellopsis*, *Asteromphalus*, *Asteroplanus*, *Attheya*, *Bacteriastrum*, *Cerataulina*, *Chaetoceros*, *Corethron*, *Coscinodiscus*, *Cyclotella*, *Cylindrotheca*, *Dactyliosolen*, *Delphineis*, *Detonula*, *Ditylum*, *Entomoneis*, *Eucampia*, *Fragilariopsis*, *Guinardia*, *Haslea*, *Hemiaulus*, *Lauderia*, *Leptocylindrus*, *Meuniera*, *Minidiscus*, *Minutocellus*, *Navicula*, *Nitzschia*, *Odontella*, *Plagiogrammopsis*, *Pleurosigma*, *Porosira*, *Proboscia*, *Pseudo-nitzschia*, *Pseudogomphonema*, *Rhizosolenia*, *Skeletonema*, *Stephanopyxis*, *Tabularia*, *Thalassionema*, *Thalassiosira*

Relative abundance of commonly detected diatom genera

Relative abundance results (semi-quantitative) appeared to be similar, but not congruent across methods for commonly shared genera (Fig. 2).

For instance, *Skeletonema* was observed in the first 10 samples based on DNA metabarcoding and in seven of those samples using light microscopy but was only observed in two of those samples based on FlowCam analyses. Like the less abundant *Minidiscus* and *Minutocellus, Skeletonema* is a relatively small diatom (diameter range of 2–21 μm) for which the FlowCam was not well suited compared to light microscopy or metabarcoding approaches. On the other hand, *Guinardia* and *Pseudo-nitzschia* were both frequently detected at higher relative abundances in the morphology-based results compared to metabarcoding. As discussed previously, shortcomings in sequence reference databases may play a role in erroneous taxonomic assignments (e.g., *Guinardia*). The lower relative abundance of these genera in metabarcoding results could also stem from extraction inefficiency or preferential amplification of other taxa in the sample due to PCR bias, both commonly faced issued with metabarcoding approaches (Medinger et al. 2010; Gonzalez et al. 2012). Conversely, *Thalassiosira* was frequently detected in higher proportions in the metabarcoding results compared to the morphological approaches which may be related to variable gene copy number which has been observed to range by an order of magnitude for some species (Zhu et al. 2005).

The two-way crossed ANOSIM analysis indicated that the three methods produced similar patterns in relative abundance across samples (Table 2). The two-way crossed ANOSIM for the entire community profile also indicated that there was no significant effect of method on community composition changes across samples. However, when examining each of the commonly shared genera individually, it appeared that there was variability in the effect of method across samples. There was a large effect of method on the relative abundance of the genera *Skeletonema* and *Fragilariopsis* ($p_{av} > 0.5$), followed by a small effect on *Pseudo-nitzschia*, *Cerataulina*, *Cylindrothec*a, *Asterionellopsis*, and *Thalassiosira* ($p_{av} < 0.25$), and no detectable impact of method on relative abundances

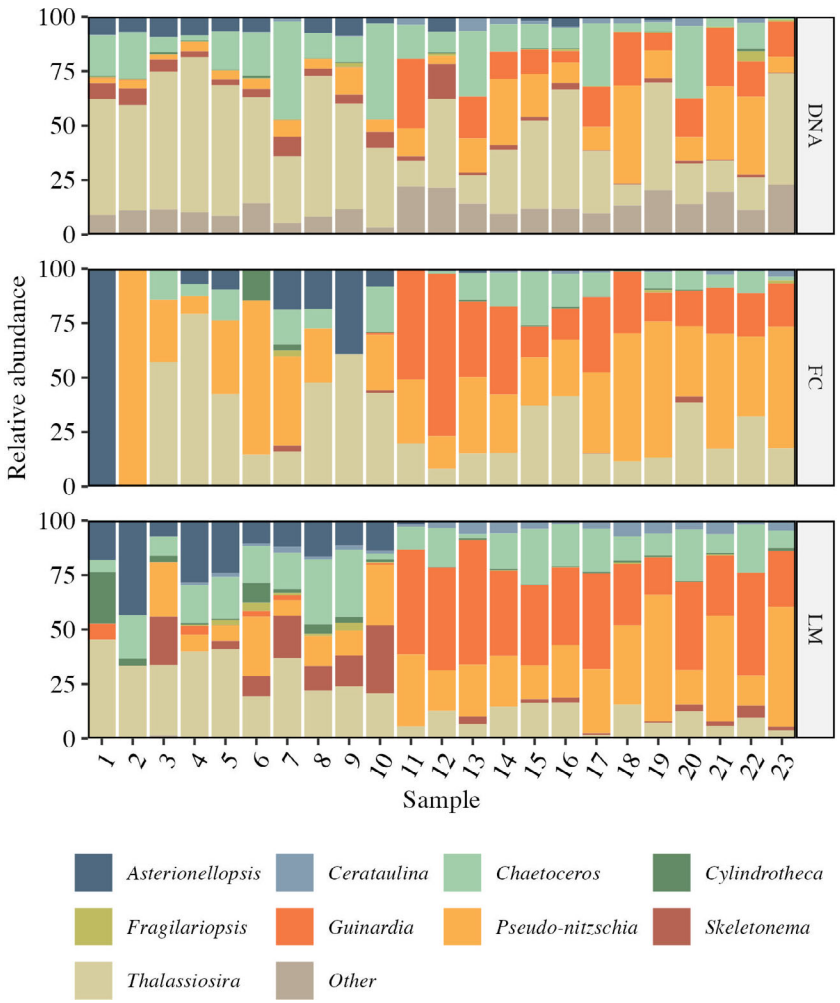

Fig. 2. Relative proportion of each commonly shared diatom genus detected in samples using each taxonomic approach. A detailed description of each treatment is provided in Supporting Information Table S3.

Table 2. Results of a two-way crossed ANOSIM test of similarity across method results for relative abundances.

	p_{av}	p
Community	ns	ns
Asterionellopsis	0.161	0.018
Cerataulina	0.125	0.042
Chaetoceros	ns	ns
Cylindrotheca	0.127	0.048
Fragilariopsis	0.708	0.001
Guinardia	ns	ns
Pseudo-nitzschia	0.135	0.030
Skeletonema	0.584	0.001
Thalassiosira	0.211	0.007

Table 3. Results of a two-way crossed ANOSIM test of similarity across method results for absolute abundances.

	p_{av}	p
Community	0.901	0.001
Asterionellopsis	0.346	0.001
Cerataulina	0.176	0.018
Chaetoceros	0.143	0.028
Cylindrotheca	ns	ns
Fragilariopsis	0.462	0.002
Guinardia	ns	ns
Pseudo-nitzschia	0.208	0.008
Skeletonema	0.473	0.001
Thalassiosira	0.447	0.001

of the genera *Chaetoceros* and *Guinardia*. When splitting samples into two categories, higher biomass (Chl $a > 1\ \mu g\ L^{-1}$) and lower biomass (Chl $a < 1\ \mu g\ L^{-1}$), an ANOSIM showed that there was no significant difference in average relative abundance changes across methods between the two biomass categories ($R = 0.556$, $p = 0.1$).

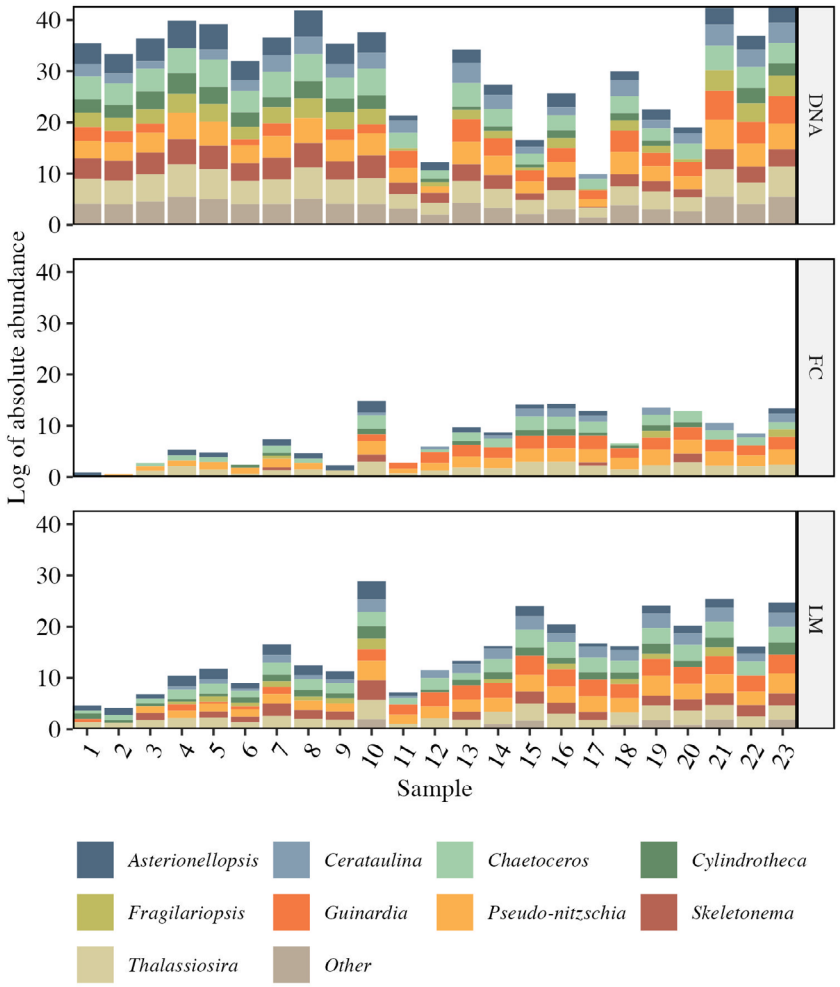

Fig. 3. The log of absolute abundance values of each commonly shared diatom genus detected in samples using each taxonomic approach. Units are as follows: LM = cells mL^{-1}, FC = particles mL^{-1}, and DNA = 18S rRNA gene copies mL^{-1}. A detailed description of each treatment is provided in Supporting Information Table S3.

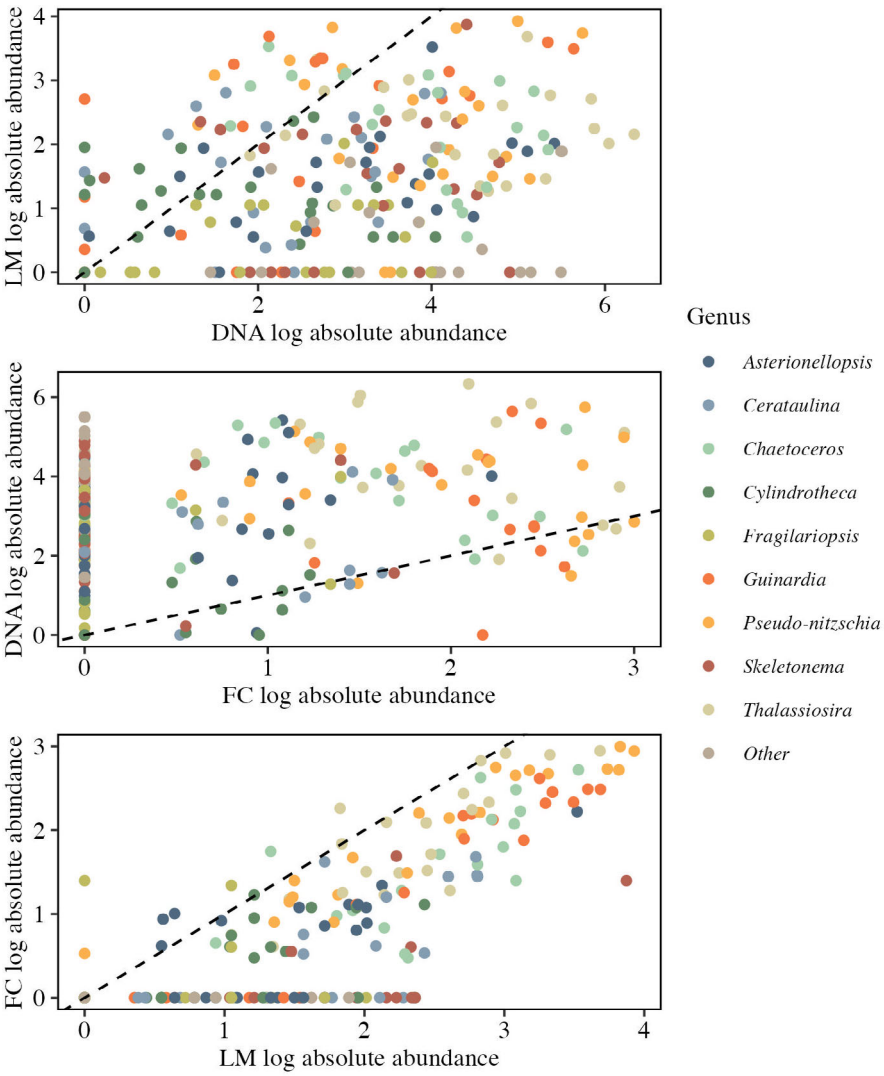

Fig. 4. Pairwise comparisons between results of logged absolute abundance for each taxonomic approach pair. The dashed line represents a 1 : 1 line of perfect agreement. Units are as follows: LM = cells mL^{-1}, FC = particles mL^{-1}, and DNA = 18S rRNA gene copies mL^{-1}.

Table 4. Linear relationships between absolute abundance results from different methods.

	DNA vs. LM			FC vs. LM		
	Relationship	R^2	p	**Relationship**	R^2	p
Asterionellopsis		ns		$y=0.05x+3.6$	0.98	2.2×10^{-16}
Cerataulina	$y=7.9x+605.7$	0.24	0.04	$y=0.05x+1.7$	0.51	5.6×10^{-5}
Chaetoceros		ns		$y=0.15x+3.8$	0.61	8.5×10^{-6}
Cylindrotheca		ns			ns	
Fragilariopsis		ns			ns	
Guinardia	$y=35.5x-5328.2$	0.23	0.04	$y=0.07x+44.4$	0.66	2.7×10^{-5}
Pseudo-nitzschia		ns		$y=0.10x+115.1$	0.73	3.1×10^{-6}
Skeletonema		ns			ns	
Thalassiosira		ns		$y=0.2x+100.4$	0.59	1.1×10^{-4}

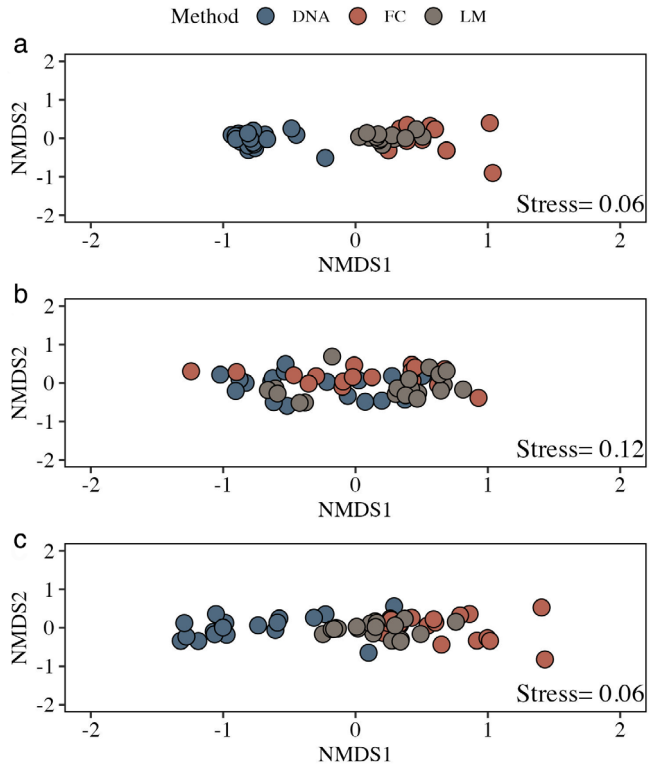

Fig. 5. Non-metric multidimensional scaling plots showing the similarity of community composition results across taxonomic approaches for three different level of quantification: qualitative presence/absence data (**A**), semi-quantitative relative abundance data (**B**), and log transformed quantitative absolute abundance data (**C**). The stress value associated with each NMDS is shown in the bottom right of each plot.

Absolute abundance of commonly detected diatom genera

Absolute abundance results differed more appreciably among methods (Fig. 3). There appeared to be similar trends across treatments between abundances obtained by FlowCam (particle mL^{-1}) and light microscopy (cell mL^{-1}), however, the results from DNA sequencing showed different patterns with much higher 18S rRNA gene abundances (gene copies mL^{-1}).

A two-way crossed ANOSIM test revealed there was a significant effect of method for the entire community profile and for all but two of the commonly shared diatom genera, *Guinardia* and *Cylindrotheca* (Table 3).

The effect size of method was small for *Cerataulina*, *Chaetoceros*, *Cylindrotheca*, and *Pseudo-nitzschia*; moderate for *Asterionellopsis* and *Thalassiosira* and high for *Skeletonema* and *Fragilariopsis* in relation to each other.

Analysis of pairwise comparisons of the untransformed absolute abundance results of each method indicated that there was more agreement between FlowCam and light microscopy results, followed by 18S rRNA metabarcoding and light microscopy, with little agreement between FlowCam and metabarcoding (Fig. 4).

Of the nine commonly shared diatom genera, seven had a significant, positive linear relationship between FlowCam and light microscopy absolute abundance results indicating that while these methods may not have derived the exact same abundances as they are in different units (particles, cells, or gene copy numbers per sample volume), they did derive similar trends. Only two of the nine commonly shared diatom genera showed significant relationships between DNA sequencing and light microscopy and none of the commonly

Table 5. Overview of practical considerations for each method discussed.

	Light microscopy	FlowCam	DNA sequencing
Unit of measurement	cells mL^{-1}	particles mL^{-1}	copies mL^{-1}—with the use of an internal standard
Ability to determine biomass	Yes	Yes	No
Amount of water utilized (mL)*	25–100	25	1000
Amount of water analyzed (mL)†	0.8–1.2	0.2–0.5	Indeterminate
Sample treatment	Preservation	None—preservation possible	Filtration
Time to sample results‡	Days	Hours	Weeks
Taxonomic resolution	Genus—potentially species	Genus—species not likely	Genus—potentially species
Objectivity of identification	Reliant on taxonomist	Reliant on taxonomist and classification algorithm	Reliant on reference database
Size range (μm)§	3–200	10–100	0.45–200

*Volumes refer to the amount of water settled, pumped through the FlowCam, or filtered, which does not directly correspond to the amount of water processed by each respective method.
†Volumes refer to the volume of concentrated preserved sample counted via light microscopy, or the summed volume captured across all imaged particles per sample from FlowCam. The volumes analyzed using light microscopy and FlowCam could potentially be increased by increasing the fields of view counted or slowing the pump speed, respectively. It is not feasible to determine the volume of water processed for DNA sequencing methods given that the volume of water filtered does not correspond to amount of DNA analyzed after amplification and sequencing.
‡The time to results shown here only accounts for time required after the sample is collected and does not account for any training time or time for building reference libraries to input into the FlowCam classification algorithm.
§Size range here is capped at upper microplankton size of 200 μm, microscopy and DNA have capabilities beyond that size range not relevant to this study.

shared diatom genera had a significant linear relationship between DNA sequencing and FlowCam, indicating less comparability between these methods (Table 4).

Method comparison across levels of quantification

NMDS plots for each of the levels of quantification that can be achieved, qualitative, semi-quantitative, and quantitative, indicate varying levels of agreement between methods (Fig. 5). At the qualitative level showing the presence/absence transformation of the data, there is significant clustering of the DNA sequencing analysis away from FlowCam and light microscopy. At the semi-quantitative level using the relative abundance data, there is significant overlap among each method indicating strong agreement. At the quantitative level, using the absolute abundance data, there is overlap especially between light microscopy and FlowCam. However, there is minimal overlap of the DNA sequencing analysis, indicating less agreement.

Discussion

Quantitative comparison of diatom enumeration methods

Results indicate varying levels of agreement between methods across modes of quantification. At the qualitative level, detecting which diatom genera are present in each sample, there is large disagreement between methods. As expected based on previous analysis of the strength of DNA sequencing in analyzing biodiversity (Caron et al. 2009; Johnson and Martiny 2015), this approach detected four-times as many diatom genera as FlowCam or light microscopy across all samples. While DNA sequencing returned a high number of genera, analysis showed that a majority of DNA reads belonged to commonly shared diatom genera, and thus each method used in this study did detect the major members of the diatom community.

When incorporating semi-quantitative information on abundance of each diatom genera through analyzing relative abundance (i.e., relative proportions), the methods derive community profiles that are not significantly different, though there is a varying level of difference for each specific genus of the shared diatoms detected. Using these methods to analyze the relative abundance of the most dominant genera yields comparable results that do not appear to shift between high and low biomass samples. The agreement in relative abundance across methods is well supported here for communities dominated by moderately-sized diatoms cosmopolitan on the U.S. West coast. Agreement may shift based on the diatom community composition of a sample, especially for communities dominated by small, morphologically indistinct, or understudied diatoms.

At the quantitative level, a comparison of absolute abundances across the different methods, these methods are not comparable and yield significantly different diatom compositions. One of the largest barriers to deriving comparable results across these methods is that each approach has different units of measure (Table 5). FlowCam has units of absolute abundance in particles mL^{-1}, light microscopy has units of absolute abundance in cells mL^{-1}, and DNA sequencing with incorporation of the internal standard has units of absolute abundance in 18S rRNA gene copies mL^{-1}. Deriving cell densities, as is provided through light microscopy, is ideal because it is most easily contextualized and used in further downstream analyses. FlowCam results in units of particles mL^{-1} may lead to underestimation of absolute abundance as there is also variability in the number of cells captured in each particle image, especially with chain-forming diatoms (Camoying and Yñiguez 2016). Some particle images contain only part of a longer chain, a chain oriented in such a way that it is impossible to count individual cells, or a chain with indistinguishable segments. The variability in the number of cells captured in a particle image presents a challenge to cell enumeration that is unique to chain-forming diatoms. Other studies have approached this issue by using biovolumes as opposed to counts (Álvarez et al. 2012; Hrycik et al. 2019) or manually counting individual cells within particle images (Camoying and Yñiguez 2016), both of which yielded more comparable units but significantly increased processing time.

In contrast, DNA sequencing results in units of V4 18S rRNA gene copies mL^{-1} may lead to overestimation of abundance as an individual organism may have many copies of the 18S rRNA gene (Countway et al. 2005; Martin et al. 2022). A real-time PCR analysis of six monoclonal diatom cultures indicated that the diatoms possessed between 61 and 36,896 18S gene copies per cell (Godhe et al. 2008). The absolute abundance copy number densities from DNA sequencing were derived using an equation reliant on the proportion of plasmid reads to total sample reads. The proof of concept for the internal standard normalization approach used here recommends that plasmid reads make up greater than 0.1% of total reads for successful detection and tested the standardization on samples with plasmid copies making up 0.7–5.7% (Lin et al. 2019). However, the proportion of plasmid reads to total reads varied widely across samples in this study (0.02–92.8%) with an average of 12.5%, outside of the range from the proof of concept (Supporting Information Table S1). The ratio of DNA derived copies mL^{-1} to light microscopy derived cells mL^{-1} provides an estimate of 18S rRNA gene copies per cell for each sample ranging from 0.07 to 8508 copies $cell^{-1}$ with an average of 1105 copies $cell^{-1}$ (Supporting Information Table S2). Six of the seven samples that are an order of magnitude below the lower limit of published copies $cell^{-1}$ estimate also contain a larger proportion of plasmid reads than the upper limit of that used for the standardization proof of concept; therefore, the plasmid standard may be overwhelming sample reads and influencing absolute abundance results for a subset of samples across a range of biomass values. While there is strong promise for the use of internal standards in amplicon sequencing (Piwosz et al. 2020) and one approach was used with some success in this analysis, the lack of standardized protocols for achieving absolute abundance results is a current limitation of the DNA sequencing approach.

Qualitative comparison of diatom enumeration methods

Aside from the quantitative comparison of method performance, there are practical caveats to consider (Table 5). Despite sequences being aggregated at the genus level for this analysis, the sequencing approach did allow for identification at the species level which may be an advantage for species-specific research questions. Species level identification of some taxa may be possible for FlowCam or microscopy methods based on clear morphological distinctions, but this depth of analysis would require more labor time for the taxonomist and exemplary particle image quality from the FlowCam which is difficult to achieve. In addition, as revealed by the number of diatom genera detected by each of the approaches, DNA sequencing allows for deeper analysis that is preferred when trying to capture rarer taxa or obtain estimates of diatom species richness and diversity (Dawson and Hagen 2009). Closer inspection of the genera detected by metabarcoding, but not FlowCam or light microscopy approaches highlight the ability of DNA based methods to analyze nanoplanktonic diatoms and morphologically indistinct genera. Deeper taxonomic analysis may not be achieved with metabarcoding for environments with a lack of representative sequences in reference databases (Abad et al. 2016; Apothéloz-Perret-Gentil et al. 2021). While DNA sequencing allowed for deeper analysis in this study, it does have practical downsides such as not yielding easily interpreted quantitative estimates of community composition. As there is further progress in the field in obtaining the number of 18S rRNA gene copies per cell of dominant diatoms (Martin et al. 2022), the results from DNA analysis may become more quantitative. In addition, metabarcoding approaches are multi-step and there may be variability in the time required to troubleshoot successful amplification or potentially in the wait times for samples to be sequenced. While the variability and multi-step process of the metabarcoding approach may lengthen the total time to results, samples are run in parallel, and thus all results are received at once and the process is less labor intensive overall when working with a larger number of samples.

While FlowCam and traditional light microscopy provide more easily interpretable quantitative units, they can be more labor intensive and rely on the professional expertise in morphological identification from the user (Kenitz et al. 2023). FlowCam results require some processing time to fully derive community composition, the imager does provide real-time images which may be a large advantage in the field for rapid bloom or harmful algae detection for example. Because the FlowCam uses an algorithm to pre-sort images based on a training set, there may be less user knowledge and training required compared to morphological identification by light microscopy; however, the user must still create the training set and correct any misidentifications from the algorithm requiring both time and expertise. One benefit of the FlowCam is that an image of each particle captured is automatically saved to a dataset and can be revisited later. While image capture is possible with traditional light microscopy, the process is more laborious and time consuming than automatic FlowCam imaging. As high-throughput imaging technologies like the FlowCam continue to advance, so do technical tools for image classification that may improve identification accuracy (e.g., EcoTaxa) (Picheral et al. 2017). In addition to the potential for increased identification accuracy, FlowCam image libraries have been used as a quick and accessible training tool for new taxonomists (Clayton et al. 2022). The image clarity of particle images captured by the FlowCam can be variable, and thus limits taxonomic classification of some diatoms to the genus level, confirming previously published results that the FlowCam is not ideal for high-resolution taxonomic analysis (e.g., species level) (Álvarez et al. 2014). The size range limitation of the FlowCam as used in this study is also a disadvantage of the FlowCam approach for diatom community composition analysis since many genera may fall outside of the 10–100 μm range as evidenced by the low detection of *Skeletonema* in FlowCam results of this study.

Conclusions

Each approach used in this study has its own sets of advantages and disadvantages that should be considered in the context of a particular research objective when choosing which method to utilize. Our findings indicate that relative abundance results from these methods are comparable for the commonly shared diatom genera making up a majority of the community composition in a sample. Whereas, because absolute abundance results are in different units, this limits the extent of intercomparability. Thus, comparing community composition results between studies that have employed different diatom identification and enumeration methods should be performed with care and note the limitations of each method.

Automated cell imaging techniques will provide the most rapid results which may be ideal for field monitoring programs; however, these techniques do have several drawbacks to consider. Automated cell imaging techniques are limited in taxonomic depth and less likely to allow for the detection of rarer taxa given the volumes that are typically analyzed. In addition, automated cell imaging techniques may skew abundance data for chain-forming diatoms, presenting hurdles for researchers interested in detecting specific ecologically important species (e.g., harmful algal bloom forming species). FlowCam methods also have a more limited size range capability and are thus not the most suitable for analyzing especially small or large phytoplankton. Light microscopy requires more time and expertise, but it appears to detect the greatest number of taxa for those that can be morphologically differentiated, potentially making it the best option for researchers aiming to detect specific ecologically important species. DNA sequencing removes the inherent subjectivity in FlowCam and light microscopy, but there are significant hurdles to

converting metabarcoding results to cell abundances. In addition, taxonomic information derived from sequencing approaches is reliant on the strength of the taxonomic reference database, making the approach less reliable for understudied regions or organisms. Several studies have proposed a combination of morphological and nucleic acid-based methods to derive the most informative analysis of plankton community composition (Groendahl et al. 2017; Banerji et al. 2018; Malashenkov et al. 2021; Santi et al. 2021; Hammond et al. 2022).

Future work on comparing advancing methodologies for identification of diatoms and other phytoplankton groups should consider evaluating methods using a defined mixture of phytoplankton cells and nucleic acid molecules to simulate the composition of a natural community (i.e., mock community). Other studies have used mock communities to validate DNA barcoding workflows (Kermarrec et al. 2013; Smith et al. 2017; Catlett et al. 2020), but there has been little research to date utilizing mock communities to assess multiple advancing methodologies such as FlowCam and DNA sequencing. This study gives insights into how the results of these field-based methods compare within dynamic diatom communities, providing valuable information for researchers making methodological choices or comparing across study results; however, there is no way to determine which results are the most accurate. The use of mock communities with known diatom composition and cell abundance would allow for an additional analysis of not only how the methods in question compare to each other, but also how well they accurately characterize known community composition.

References

Abad, D., A. Albaina, M. Aguirre, A. Laza-Martínez, I. Uriarte, A. Iriarte, F. Villate, and A. Estonba. 2016. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Mar. Biol. **163**: 149. doi:10.1007/s00227-016-2920-0

Agustí, S., J. I. González-Gordillo, D. Vaqué, M. Estrada, M. I. Cerezo, G. Salazar, J. M. Gasol, and C. M. Duarte. 2015. Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. Nat. Commun. **6**: 7608. doi:10.1038/ncomms8608

Álvarez, E., Á. López-Urrutia, and E. Nogueira. 2012. Improvement of plankton biovolume estimates derived from image-based automatic sampling devices: Application to FlowCAM. J. Plankton Res. **34**: 454–469. doi:10.1093/plankt/fbs017

Álvarez, E., M. Moyano, Á. López-Urrutia, E. Nogueira, and R. Scharek. 2014. Routine determination of plankton community composition and size structure: A comparison between FlowCAM and light microscopy. J. Plankton Res. **36**: 170–184. doi:10.1093/plankt/fbt069

Amend, A. S., K. A. Seifert, and T. D. Bruns. 2010. Quantifying microbial communities with 454 pyrosequencing: Does read abundance count? Mol. Ecol. **19**: 5555–5565. doi:10.1111/j.1365-294X.2010.04898.x

Apothéloz-Perret-Gentil, L., A. Bouchez, T. Cordier, A. Cordonier, J. Guéguen, F. Rimet, V. Vasselon, and J. Pawlowski. 2021. Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. Mol. Ecol. **30**: 2959–2968. doi:10.1111/mec.15646

Archibald, R. M. 1984. Diatom illustrations——An appeal. Bacillaria **7**: 173–178.

Armbrust, E. 2009. The life of diatoms in the world's oceans. Nature **459**: 185–192. doi:10.1038/nature08057

Banerji, A., M. Bagley, M. Elk, E. Pilgrim, J. Martinson, and J. Santo Domingo. 2018. Spatial and temporal dynamics of a freshwater eukaryotic plankton community revealed via 18S rRNA gene metabarcoding. Hydrobiologia **818**: 71–86. doi:10.1007/s10750-018-3593-0

Bates, S. S., and V. L. Trainer. 2006. The ecology of harmful diatoms, p. 81–93. *In* E. Granéli and J. T. Turner [eds.], Ecology of harmful algae. Springer. doi:10.1007/978-3-540-32210-8_7

Bates, S., and others. 1989. Pennate diatom *Nitzschia pungens* as the primary source of domoic acid, a toxin in shellfish from eastern Prince Edward Island, Canada. Can. J. Fish. Aquat. Sci. **46**: 1203–1215. doi:10.1139/f89-156

Benfield, M. C., and others. 2007. RAPID: Research on automated plankton identification. Oceanography **20**: 172–187. doi:10.5670/oceanog.2007.63

Bergkemper, V., and T. Weisse. 2018. Do current European lake monitoring programmes reliably estimate phytoplankton community changes? Hydrobiologia **824**: 143–162. doi:10.1007/s10750-017-3426-6

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. **22**: 148–155. doi:10.1016/j.tree.2006.11.004

Blanco, S. 2020. Diatom taxonomy and identification keys, p. 25–38. *In* G. Cristóbal, S. Blanco, and G. Bueno [eds.], Modern trends in diatom identification: Fundamentals and applications. Springer International Publishing. doi:10.1007/978-3-030-39212-3_3

Bolyen, E., and others. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. **37**: 852–857. doi:10.1038/s41587-019-0209-9

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods **13**: 581–583. doi:10.1038/nmeth.3869

Camoying, M. G., and A. T. Yñiguez. 2016. FlowCAM optimization: Attaining good quality images for higher taxonomic classification resolution of natural phytoplankton samples.

Limnol. Oceanogr.: Methods **14**: 305–314. doi:10.1002/lom3.10090

Caporaso, J. G., and others. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **7**: 335–336. doi:10.1038/nmeth.f.303

Caron, D. A., and others. 2009. Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl. Environ. Microbiol. **75**: 5797–5808. doi:10.1128/AEM.00298-09

Caron, D. A., P. D. Countway, A. C. Jones, D. Y. Kim, and A. Schnetzer. 2012. Marine protistan diversity. Ann. Rev. Mar. Sci. **4**: 467–493. doi:10.1146/annurev-marine-120709-142802

Catlett, D., P. G. Matson, C. A. Carlson, E. G. Wilbanks, D. A. Siegel, and M. D. Iglesias-Rodriguez. 2020. Evaluation of accuracy and precision in an amplicon sequencing workflow for marine protist communities. Limnol. Oceanogr.: Methods **18**: 20–40. doi:10.1002/lom3.10343

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Aust. J. Ecol. **18**: 117–143. doi:10.1111/j.1442-9993.1993.tb00438.x

Clarke, K., and R. Gorley. 2015. Getting started with PRIMER v7. PRIMER-E.

Clayton, S., L. Gibala-Smith, K. Mogatas, C. Flores-Vargas, K. Marciniak, M. Wigginton, and M. R. Mulholland. 2022. Imaging technologies build capacity and accessibility in phytoplankton species identification expertise for research and monitoring: Lessons learned during the COVID-19 pandemic. Front. Microbiol. **13**: 823109. doi:10.3389/fmicb.2022.823109

Countway, P. D., R. J. Gast, P. Savai, and D. A. Caron. 2005. Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. J. Eukaryot. Microbiol. **52**: 95–106. doi:10.1111/j.1550-7408.2005.05202006.x

Dawson, S. C., and K. D. Hagen. 2009. Mapping the protistan 'rare biosphere'. J. Biol. **8**: 105. doi:10.1186/jbiol201

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. **14**: 927–930. doi:10.1111/j.1654-1103.2003.tb02228.x

Ducklow, H. W., D. K. Steinberg, and K. O. Buesseler. 2001. Upper ocean carbon export and the biological pump. Oceanography **14**: 50–58. doi:10.5670/oceanog.2001.06

Dunn, G., and B. S. Everitt. 2004. An introduction to mathematical taxonomy. Courier Corporation.

Dunthorn, M., J. Klier, J. Bunge, and T. Stoeck. 2012. Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. J. Eukaryot. Microbiol. **59**: 185–187. doi:10.1111/j.1550-7408.2011.00602.x

Fadrosh, D. W., B. Ma, P. Gajer, N. Sengamalay, S. Ott, R. M. Brotman, and J. Ravel. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome **2**: 6. doi:10.1186/2049-2618-2-6

Falkowski, P. G., R. T. Barber, and V. Smetacek. 1998. Biogeochemical controls and feedbacks on ocean primary production. Science **281**: 200–206. doi:10.1126/science.281.5374.200

Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science **281**: 237–240. doi:10.1126/science.281.5374.237

Godhe, A., M. E. Asplund, K. Härnström, V. Saravanan, A. Tyagi, and I. Karunasagar. 2008. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. Appl. Environ. Microbiol. **74**: 7174–7182. doi:10.1128/AEM.01298-08

Gomes, K. 2020. Determining impacts of iron limitation on diatom chloroplast metabolism and physiology. Dissertation. Univ. of Rhode Island.

Gong, W., and A. Marchetti. 2019. Estimation of 18S gene copy number in marine eukaryotic plankton using a next-generation sequencing approach. Front Mar Sci **6**: 219. doi:10.3389/fmars.2019.00219

Gong, W., N. Hall, H. Paerl, and A. Marchetti. 2020. Phytoplankton composition in a eutrophic estuary: Comparison of multiple taxonomic approaches and influence of environmental factors. Environ. Microbiol. **22**: 4718–4731. doi:10.1111/1462-2920.15221

Gonzalez, J. M., M. C. Portillo, P. Belda-Ferre, and A. Mira. 2012. Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. PloS One **7**: e29973. doi:10.1371/journal.pone.0029973

Groendahl, S., M. Kahlert, and P. Fink. 2017. The best of both worlds: A combined approach for analyzing microalgal diversity via metabarcoding and morphology-based methods. PloS One **12**: e0172808. doi:10.1371/journal.pone.0172808

Guillou, L., and others. 2012. The protist ribosomal reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. **41**: D597–D604. doi:10.1093/nar/gks1160

Hajibabaei, M., G. A. Singer, P. D. Hebert, and D. A. Hickey. 2007. DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. **23**: 167–172. doi:10.1016/j.tig.2007.02.001

Hammond, S. W., L. Lodolo, S. K. Hu, and A. L. Pasulka. 2022. Methodological 'lenses' influence the characterization of phytoplankton dynamics in a coastal upwelling ecosystem. Environ. Microbiol. Rep. **14**: 897–906. doi:10.1111/1758-2229.13116

Hebert, P. D., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications through DNA barcodes. Proc. Biol. Sci. **270**: 313–321. doi:10.1098/rspb.2002.2218

Hrycik, A. R., A. Shambaugh, and J. D. Stockwell. 2019. Comparison of FlowCAM and microscope biovolume measurements for a diverse freshwater phytoplankton community. J. Plankton Res. **41**: 849–864. doi:10.1093/plankt/fbz056

Jahn, R., H. Zetzsche, R. Reinhardt, and B. Gemeinholzer. 2007. Diatoms and DNA barcoding: A pilot study on an environmental sample, p. 63–68. *In* W.-H. Kusber and R. Jahn [eds.], Proceedings of the 1st central European diatom meeting. Freie Universität Berlin. doi:10.3372/cediatom.113

Jerde, C. L., A. R. Mahon, W. L. Chadderton, and D. M. Lodge. 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conserv. Lett. **4**: 150–157. doi:10.1111/j.1755-263X.2010.00158.x

Johnson, Z. I., and A. C. Martiny. 2015. Techniques for quantifying phytoplankton biodiversity. Ann. Rev. Mar. Sci. **7**: 299–324. doi:10.1146/annurev-marine-010814-015902

Karlson, B., C. Cusack, and E. Bresnan. 2010. Microscopic and molecular methods for quantitative phytoplankton analysis. UNESCO.

Kembel, S. W., M. Wu, J. A. Eisen, and J. L. Green. 2012. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput. Biol. **8**: e1002743. doi:10.1371/journal.pcbi.1002743

Kenitz, K. M., and others. 2023. Convening expert taxonomists to build image libraries for training automated classifiers. Limnol. Oceanogr.: Bull. **32**: 89–97. doi:10.1002/lob.10584

Kermarrec, L., A. Franc, F. Rimet, P. Chaumeil, J. F. Humbert, and A. Bouchez. 2013. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: A test for freshwater diatoms. Mol. Ecol. Resour. **13**: 607–619. doi:10.1111/1755-0998.12105

Kittelmann, S., S. R. Devente, M. R. Kirk, H. Seedorf, B. A. Dehority, and P. H. Janssen. 2015. Phylogeny of intestinal ciliates, including *Charonina ventriculi*, and comparison of microscopy and 18S rRNA gene pyrosequencing for rumen ciliate community structure analysis. Appl. Environ. Microbiol. **81**: 2433–2444. doi:10.1128/aem.03697-14

Kozich, J. J., S. L. Westcott, N. T. Baxter, S. K. Highlander, and P. D. Schloss. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. **79**: 5112–5120. doi:10.1128/AEM.01043-13

Le Bourg, B., V. Cornet-Barthaux, M. Pagano, and J. Blanchot. 2015. FlowCAM as a tool for studying small (80–1000 μm) metazooplankton communities. J. Plankton Res. **37**: 666–670. doi:10.1093/plankt/fbv025

Lin, Y., S. Gifford, H. Ducklow, O. Schofield, N. Cassar, and E. V. Stabb. 2019. Towards quantitative microbiome community profiling using internal standards. Appl. Environ. Microbiol. **85**: e02634-18. doi:10.1128/AEM.02634-18

Lin, Y., O. Torano, L. Whitehouse, E. Pierce, C. P. Till, M. Hurst, R. Freiberger, T. Mellett, M. T. Maldonado, J. Guo, M. Sutton, D. Zeitz and A. Marchetti. 2023. Variability in the phytoplankton response to upwelling across an iron limitation mosaic within the California Current System. BioRxiv. doi:10.1101/2023.09.22.558967

Lombard, F., and others. 2019. Globally consistent quantitative observations of planktonic ecosystems. Front. Mar. Sci. **6**: 196. doi:10.3389/fmars.2019.00196

MacKeigan, P. W., and others. 2022. Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages across hundreds of lakes. Harmful Algae **113**: 102–187. doi:10.1016/j.hal.2022.102187

Malashenkov, D. V., V. Dashkova, K. Zhakupova, I. A. Vorobjev, and N. S. Barteneva. 2021. Comparative analysis of freshwater phytoplankton communities in two lakes of Burabay National Park using morphological and molecular approaches. Sci. Rep. **11**: 16130. doi:10.1038/s41598-021-95223-z

Malviya, S., and others. 2016. Insights into global diatom distribution and diversity in the world's ocean. Proc. Natl. Acad. Sci. USA **113**: E1516–E1525. doi:10.1073/pnas.1509523113

Marinov, I., S. C. Doney, I. D. Lima, K. Lindsay, J. K. Moore, and N. Mahowald. 2013. North-south asymmetry in the modeled phytoplankton community response to climate change over the 21st century. Glob. Biogeochem. Cycles **27**: 1274–1290. doi:10.1002/2013GB004599

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. **17**: 10–12. doi:10.14806/ej.17.1.200

Martin, J. L., I. Santi, P. Pitta, U. John, and N. Gypens. 2022. Towards quantitative metabarcoding of eukaryotic plankton: An approach to improve 18S rRNA gene copy number bias. Metabarcoding Metagenom. **6**: e85794. doi:10.3897/mbmg.6.85794

Medinger, R., V. Nolte, R. V. Pandey, S. Jost, B. Ottenwälder, C. Schlötterer, and J. Boenigk. 2010. Diversity in a hidden world: Potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol. Ecol. **19**: 32–40. doi:10.1111/j.1365-294X.2009.04478.x

Medlin, L., H. J. Elwood, S. Stickel, and M. L. Sogin. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene **71**: 491–499. doi:10.1016/0378-1119(88)90066-2

Mora, D., and others. 2019. Morphology and metabarcoding: A test with stream diatoms from Mexico highlights the complementarity of identification methods. Freshw. Sci. **38**: 448–464. doi:10.1086/704827

Morales, E. A., P. A. Siver, and F. R. Trainor. 2001. Identification of diatoms (Bacillariophyceae) during ecological assessments: Comparison between light microscopy and scanning electron microscopy techniques. Proc. Acad. Nat. Sci. Phila. **151**: 95–103.

Nelson, D. M., P. Tréguer, M. A. Brzezinski, A. Leynaert, and B. Quéguiner. 1995. Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Glob. Biogeochem. Cycles **9**: 359–372. doi:10.1029/95GB01070

Owen, B. M., C. S. Hallett, J. J. Cosgrove, J. R. Tweedley, and N. R. Moheimani. 2022. Reporting of methods for automated devices: A systematic review and recommendation for studies using FlowCam for phytoplankton. Limnol. Oceanogr.: Methods **20**: 400–427. doi:10.1002/lom3.10496

Parsons, T. R., Y. Maita, and C. M. Lalli. 1984. Fluorometric determination of chlorophylls, p. 107–109. *In* T. R. Parsons, Y. Maita, and C. M. Lalli [eds.], A manual of chemical & biological methods for seawater analysis. Pergamon.

Picheral, M., S. Colin, and J. Irisson. 2017. EcoTaxa, a tool for the taxonomic classification of images. URL Httpecotaxa Obs-Vlfr Fr.

Pierella Karlusich, J. J., F. M. Ibarbalz, and C. Bowler. 2020. Phytoplankton in the Tara Ocean. Annu. Rev. Mar. Sci. **12**: 233–265. doi:10.1146/annurev-marine-010419-010706

Piwosz, K., T. Shabarova, J. Pernthaler, T. Posch, K. Šimek, P. Porcal, and M. M. Salcher. 2020. Bacterial and eukaryotic small-subunit amplicon data do not provide a quantitative picture of microbial communities, but they are reliable in the context of ecological interpretations. mSphere **5**: e00052-20. doi:10.1128/mSphere.00052-20

Prygiel, J., and others. 2002. Determination of the biological diatom index (IBD NF T 90–354): Results of an intercomparison exercise. J. Appl. Phycol. **14**: 27–39. doi:10.1023/A:1015277207328

R Core Team. 2020. RStudio: Integrated development environment for R. Rstudio.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ **4**: e2584. doi:10.7717/peerj.2584

Santi, I., P. Kasapidis, I. Karakassis, and P. Pitta. 2021. A comparison of DNA metabarcoding and microscopy methodologies for the study of aquatic microbial eukaryotes. Diversity **13**: 180. doi:10.3390/d13050180

Sarthou, G., K. R. Timmermans, S. Blain, and P. Tréguer. 2005. Growth physiology and fate of diatoms in the ocean: A review. J. Sea Res. **53**: 25–42. doi:10.1016/j.seares.2004.01.007

Shokralla, S., J. L. Spall, J. F. Gibson, and M. Hajibabaei. 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. **21**: 1794–1805. doi:10.1111/j.1365-294X.2012.05538.x

Sieracki, C. K., M. Sieracki, and C. S. Yentsch. 1998. An imaging-in-flow system for automated analysis of marine microplankton. Mar. Ecol. Prog. Ser. **168**: 285–296. doi:10.3354/meps168285

Smetacek, V., and J. E. Cloern. 2008. On phytoplankton trends. Science **319**: 1346–1348. doi:10.1126/science.1151330

Smith, K. F., G. S. Kohli, S. A. Murray, and L. L. Rhodes. 2017. Assessment of the metabarcoding approach for community analysis of benthic-epiphytic dinoflagellates using mock communities. N. Z. J. Mar. Freshw. Res. **51**: 555–576. doi:10.1080/00288330.2017.1298632

Smith, J., and others. 2018. A decade and a half of *Pseudo-nitzschia* spp. and domoic acid along the coast of southern California. Harmful Algae **79**: 87–104. doi:10.1016/j.hal.2018.07.007

Spaulding, S. A., D. H. Jewson, R. J. Bixby, H. Nelson, and D. M. McKnight. 2012. Automated measurement of diatom size. Limnol. Oceanogr.: Methods **10**: 882–890. doi:10.4319/lom.2012.10.882

Stern, R., and others. 2018. Molecular analyses of protists in long-term observation programmes—Current status and future perspectives. J. Plankton Res. **40**: 519–536. doi:10.1093/plankt/fby035

Stoeck, T., H.-W. Breiner, S. Filker, V. Ostermaier, B. Kammerlander, and B. Sonntag. 2014. A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. Environ. Microbiol. **16**: 430–444. doi:10.1111/1462-2920.12194

Tkacz, A., M. Hortala, and P. S. Poole. 2018. Absolute quantitation of microbiota abundance in environmental samples. Microbiome **6**: 110. doi:10.1186/s40168-018-0491-7

Tomas, C. R. 1997. Identifying marine phytoplankton. Elsevier.

Tréguer, P., and others. 2018. Influence of diatom diversity on the ocean biological carbon pump. Nat. Geosci. **11**: 27–37. doi:10.1038/s41561-017-0028-x

Utermöhl, H. 1958. Methods of collecting plankton for various purposes are discussed. SIL Commun. **9**: 1–38. doi:10.1080/05384680.1958.11904091

Vaulot, D., W. Eikrem, M. Viprey, and H. Moreau. 2008. The diversity of small eukaryotic phytoplankton (≤3 μm) in marine ecosystems. FEMS Microbiol. Rev. **32**: 795–820. doi:10.1111/j.1574-6976.2008.00121.x

Wickham, H. 2011. ggplot2. Wiley Interdiscip. Rev. Comput. Stat. **3**: 180–185. doi:10.1002/wics.147

Woese, C. R. 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271. doi:10.1128/mr.51.2.221-271.1987

Xiao, X., H. Sogge, K. Lagesen, A. Tooming-Klunderud, K. S. Jakobsen, and T. Rohrlack. 2014. Use of high throughput sequencing and light microscopy show contrasting results in a study of phytoplankton occurrence in a freshwater environment. PloS One **9**: e106510. doi:10.1371/journal.pone.0106510

Yokogawa Fluid Imaging Technologies. 2011. FlowCam manual. Fluid Imaging Technologies.

Zarauz, L., and X. Irigoien. 2008. Effects of Lugol's fixation on the size structure of natural nano–microplankton samples, analyzed by means of an automatic counting method. J. Plankton Res. **30**: 1297–1303. doi:10.1093/plankt/fbn084

Zhan, A., and others. 2013. High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. Methods Ecol. Evol. **4**: 558–565. doi:10.1111/2041-210X.12037

Zhu, F., R. Massana, F. Not, D. Marie, and D. Vaulot. 2005. Mapping of picoeukaryotes in marine ecosystems with

quantitative PCR of the 18S rRNA gene. FEMS Microbiol. Ecol. **52**: 79–92. doi:10.1016/j.femsec.2004.10.006

Zimmermann, J., N. Abarca, N. Enk, O. Skibbe, W.-H. Kusber, and R. Jahn. 2014. Taxonomic reference libraries for environmental barcoding: A best practice example from diatom research. PloS One **9**: e108793. doi:10.1371/journal.pone.0108793

Acknowledgments

We thank the captain and crew of the *R/V Oceanus* as well as the participants of the PUPCYCLE I cruise. We thank Kristofer Gomes and Bethany Jenkins for supplying the 18S rRNA gene plasmid used in this study. We thank Yajuan Lin, Alexandria Niebergall, and Nicolas Cassar for supplying primers used in this study and bioinformatics guidance. We thank Logan Whitehouse for assistance with sample collection. We thank Miriam Sutton for producing educational content based on this research. This work was funded by the National Science Foundation Grant OCE-1751805 to A.M.

Submitted 26 January 2023
Revised 06 September 2023
Accepted 07 September 2023

Associate editor: Paul F. Kemp