Phytoplankton composition in a eutrophic estuary: Comparison of multiple taxonomic approaches and influence of environmental factors

Weida Gong 1, Nathan Hall 2, Hans Paerl 2 and Adrian Marchetti 1,*

1 Department of Marine Sciences, University of North Carolina at Chapel Hill, Murray Hall, 123 South Rd. Chapel Hill, NC 27514.
2 Institute of Marine Sciences, University of North Carolina at Chapel Hill, 3431 Arendell Street, Morehead City, NC 28557.

Summary

To assess the comparability between taxonomic identification methods for phytoplankton, multiple approaches were used to characterize phytoplankton community composition within the Neuse River Estuary (NRE), North Carolina. Small subunit 18S rRNA gene sequencing and accessory pigment analysis displayed similar trends, indicating chlorophytes were the dominant microalgal group during most of the year, whereas results from microscopic cell counts, biovolume analysis and metatranscriptomics suggested diatom and dinoflagellate-dominated communities. Spatial environmental gradients drove variation in taxonomic composition due to preferences for specific environmental conditions among different microalgal groups. Cryptophytes were a greater proportion of the phytoplankton community within high nutrient, fresher environments whereas diatoms and dinoflagellates dominated higher salinity sections of the estuary. This study provides a detailed examination of phytoplankton communities associated with environmental gradients present in the NRE. The high level of taxonomic resolution offered by DNA sequencing (i.e., species to subspecies level) provides a better understanding of population dynamics at the base of estuarine food webs.

Introduction

Phytoplankton account for about 50% of global primary production (PPR) and serve a fundamental role at the base of most aquatic food webs (Antoine et al., 1996; Falkowski et al., 1998; Chavez et al., 2011). They possess a wide range of distinct physiological traits and nutritional qualities, where shifts in phytoplankton communities may have significant ecological and biogeochemical consequences (Guidi et al., 2016). For example, under most global warming scenarios, phytoplankton that have high optimal growth temperatures and/or motility (e.g., nanoflagellates and dinoflagellates) are predicted to be favoured with increased warming-induced stratification (Paerl et al., 2014b).

The importance of phytoplankton community composition has long been recognized and numerous methods for identification are routinely employed, of which microscopic analysis of water samples, accessory pigment analysis using high-performance liquid chromatography (HPLC), and small subunit 18S rRNA gene (18S) amplicon sequencing are the most commonly practiced (Johnson and Martiny, 2015). Conventional microscopic observation is the most direct way to characterize phytoplankton taxonomic composition and diversity and has provided many of the early descriptions of community structure (Johnson and Martiny, 2015). Despite its often limited taxonomic resolution, it has been shown to be a fundamentally important technique for phytoplankton enumeration, especially for large cells with distinguishable morphologies. However, the accuracy can be compromised for smaller cells (pico- or nano-phytoplankton) which lack discernible features. In addition, microscopic analysis is relatively labor intensive and requires extensive expertise in phytoplankton identification (Mackey et al., 1996). Chromatography-based pigment analysis by HPLC methods obtain taxonomic information by quantifying the relative abundance of taxon-specific accessory pigments to the universal chlorophyll a (chl a) content (Mackey et al., 1996). Pigment analysis is in general more efficient than conventional light microscopy; and it is suitable for analysis of a large number of samples and ground truthing of remote sensing data products.
However, an accurate classification of diagnostic accessory pigments can be challenging due to ambiguous marker pigments (pigments shared by multiple algal groups) and variabilities in cellular pigment content caused by changes in physiological states (Eker-Develi et al., 2012). Similarly, taxonomic resolution is only to the class level, at best.

Rapid development of high-throughput nucleic acid sequencing techniques has resulted in a variety of powerful tools to infer biological information from molecular data, among which, sequencing of mRNA (metatranscriptomics) has been applied to study metabolic activity and inferred physiology through gene expression analysis, and 18S amplicon sequencing has been widely used to study diversity and composition of microeukaryote communities (Sogin et al., 2006). Due to its high specificity and sequence conservation, 18S has become one of the most commonly used markers for eukaryotic plankton phylogenetic studies (Countway et al., 2005; de Vargas et al., 2015), the application of which has enabled the discovery of many previously unidentified marine protists (Caron et al., 2012). However, caution needs to be practiced when using 18S relative abundance to represent protist cell abundance. Sequencing artifacts that are inherent with genetic approaches (e.g., primer mismatches, various extraction efficiencies and sequencing errors) are proven obstacles for interpretation (Caron et al., 2012). In addition, variable 18S gene copy numbers across microorganisms can lead to biased compositional profiles if not properly corrected, further complicating the interpretation of 18S relative abundance (Countway et al., 2005; Caron et al., 2012; Gong and Marchetti, 2019).

Taxonomic composition based on pigment analysis can be in agreement with cell counts (Eker-Develi et al., 2008), but inconsistencies frequently occur (e.g. Goela et al., 2014). With the increasing application of molecular sequencing approaches in microbial ecology, it is important to directly compare such techniques with the more conventionally used methods to determine their potential biases.

The Neuse River Estuary (NRE), North Carolina is a dynamic estuarine system that displays large spatiotemporal dynamics in phytoplankton composition and abundance (Peierls et al., 2012; Hall et al., 2013; Gong et al., 2018). The varied nature of the phytoplankton community in the NRE provides an excellent case-study for comparing different taxonomic identification methods.

Diatoms, dinoflagellates, chlorophytes, cryptophytes and cyanobacteria are the major phytoplankton groups in the NRE, each contributing, on average, about 20% to the total chl a (Pinckney et al., 1998). Community dynamics among these groups in the NRE demonstrate that phytoplankton with competitive advantages under certain environmental conditions increase their relative contribution of the total plankton community at the expense of other groups when environmental conditions change (Piehler et al., 2004). For example, dinoflagellates can be dominant during the late winter when freshwater discharge is high and light is limiting within surface waters due to their motility and low grazing pressure, whereas cyanobacteria are typically more abundant in summer due to their high optimal growth temperatures (Pinckney et al., 1998; Hall et al., 2013). It is therefore important to have a basic understanding of the influence of environmental factors on phytoplankton community composition to better predict future shifts due to anthropogenic influences and climate change.

This study used both ‘conventional’ methods based on microscopic cell counts and phytoplankton diagnostic pigments along with high throughput sequencing of RNA and DNA to characterize the taxonomic composition of the major phytoplankton groups under different environmental conditions in the NRE, providing a comprehensive comparison of these taxonomic identification methods. Pigment analysis was found to represent the most similar compositional profiles with 18S sequencing whereas microscopic analysis was biased towards large-sized diatoms in the NRE. We further investigated the spatiotemporal variations in phytoplankton community composition at the operational taxonomic unit (OTU) level in the NRE, using 18S sequencing data, and examined their correlations with environmental factors. Significant differences in community composition were detected among different sections of the estuary which were mainly attributed to salinity and nutrient gradients present in the NRE.

Results

Comparison of taxonomic composition based on different methods

The eukaryotic phytoplankton community in the NRE is mainly comprised of diatoms, dinoflagellates, chlorophytes and cryptophytes (Pinckney et al., 1998). To examine the relative abundance of the four major microalgal groups, cell abundance, biovolume, accessory pigment abundance, 18S gene and mRNA sequence abundances were measured and compared at four distinct stations along the estuary spanning from fresh water to brackish waters (Fig. 1). Diatoms dominated all sites with respect to cell abundance, biovolume and mRNA sequence abundance (68.6%, 48.3% and 46.0% of the total abundance of the four major phytoplankton groups, respectively). In contrast, chlorophytes were the more abundant group based on pigment analysis and 18S sequencing (43.2% and 59.3%, respectively) (Fig. 2A, Table S1). Cryptophytes were consistently the least...
represented group across all methods, comprising less than 20% at all sites.

Substantial compositional differences were observed among various taxonomic identification methods, among which, 18S sequencing and pigment analysis represented the most similar taxonomic proportions at the examined stations (Fig. 2B). A principal component analysis (PCA) also indicated that 18S and pigment analyses yielded the most similar representation of the phytoplankton community structure. Based on the PCA, elevated dinoflagellates relative abundance was observed in biovolume analysis and is attributed to the difference between biovolume and other measurements (Fig. 2C). A hierarchical clustering algorithm was then applied to have a more comprehensive overview of the differences among each taxonomic method. Similarly, accessory pigment analysis and 18S were observed to be the most consistent among methods, having the least differences (Fig. 2D). Higher proportions of diatoms and dinoflagellates were also detected in metatranscriptomic analyses, which led to the differences observed with 18S and pigment analysis. Cryptophytes were the least abundant group, exerting minimal influence on differences among methods (Fig. 2C).

Analysis of 18S sequences from the NRE yielded 266 OTUs, of which 22, 38, 71 and 34 OTUs were from the four major algal groups of diatoms, dinoflagellates, chlorophytes and cryptophytes, respectively. Trebouxiophyceae, a chlorophyte, accounted for the majority of the 18S read counts in each sample and was present in all water samples (Table S2). In contrast, RNA-based sequences from metatranscriptomics analysis indicated that the dinoflagellate Karlodinium contained the most transcripts whereas Trebouxiophyceae was ranked 42nd in mRNA transcript abundance.

For most water quality monitoring programs, microscopy has been the standard measurement of phytoplankton identification and enumeration, providing a high taxonomic resolution community profile, especially for large cells with discernible morphological features (e.g., diatoms and dinoflagellates). However, the ability to identify smaller cells (e.g., chlorophytes and cryptophytes) can be difficult due to limited optical resolution (Johnson and Martin, 2015). Taxonomic composition based on microscopy-based cell abundance was compared to 18S at the genus level, when possible. Diatoms showed the least differences, with Cyclotella being the most abundant diatom genus in both measurements (Figs 3, S1). More substantial differences were detected between the other three major microalgal groups in the NRE. For example, Oxyrrhis and Chroomonas were the most abundant dinoflagellate and cryptophyte genera, accounting for, on average, 12.7% and 58.7% of the total dinoflagellate and cryptophyte cell abundances, respectively, despite contributing <1% of the total dinoflagellate and cryptophyte 18S read counts (Figs 3 and S1); similarly, Trebouxiophyceae was the most dominant chlorophyte class by 18S sequence abundance but was not readily identified by microscopy (Figs 3 and S1).
To further investigate the inconsistencies among methods, a redundancy analysis (RDA) was conducted to identify environmental factors that are associated with discrepancies in taxonomic profiles obtained using each approach. Diatom relative cell abundance measured by microscopy was associated with chl a and salinity, which suggests that higher diatom proportions were observed at high biomass and saltier, lower estuarine stations.
Fig. 3. Taxonomic composition of total OTUs based on 18S sequences.
A. OTUs are clustered at the phylum level. Some groups are further partitioned.
B. Composition of the four major microalgal groups in the NRE are partitioned into lower taxonomic groups.
(Fig. S2A). A similar positive association was observed between biovolume analysis, salinity and chl a for dinoflagellate relative proportions, suggesting an increase in dinoflagellates in the lower estuary than at upper estuary stations (Fig. S2B). In contrast, elevated proportions of chlorophytes and cryptophytes were observed at high NO$_3^-$ locations that resembled upper estuarine conditions (Fig. S2C and D). In general, different phytoplankton groups were favoured across environmental gradients through using microscopy-based cell abundance and biovolume analyses. Noticeably, taxonomic profiling inferred through metatranscriptomic sequence analysis is less impacted by environmental variations (shown by the short arrow representing metatranscriptome loading in Fig. S2), and thus variations in community composition inferred by metatranscriptomic sequencing were less pronounced across environmental gradients in the NRE.

**Spatiotemporal variations in phytoplankton community composition in the NRE**

The NRE is a highly dynamic estuary, characterized by increasing salinity and decreasing nutrient concentrations along its axis (Gong et al., 2017). Steep nutrient gradients have been characterized as a main difference between the upper (Station 20) and lower estuary stations (Station 70, 120 and 180) and have been shown to cause dramatic shifts in phytoplankton physiology and metabolic profiles along the NRE (Gong et al., 2018). A similar distinction was observed in community composition based on 18S sequencing along the NRE. A PCA was used to characterize the changes in community composition among all the samples. There was a clear difference between plankton communities from upper and lower sections of the estuary with the exception of Station 120 in June (Fig. 4A). During this sampling period, higher than typical NO$_3^-$ concentrations were observed further down the estuary. Changes across temporal gradients throughout the year were not as evident during our sampling period (Fig. S3). Analysis of similarity (ANOSIM) was used to test whether the divergence in community composition between the upper and lower NRE was significant. Consistent with PCA results, there was a significant spatial variation in community composition based on 18S sequencing while temporal differences were not significant (Figs 4B and S3).

To explore the environmental and ecological factors that correlate to spatial variation in the phytoplankton community, we performed a mantel-test to identify parameters that had strong correlations with changes in overall community composition. Nutrient concentrations (NO$_3^-$, NH$_4^+$ and PO$_4^{3-}$, salinity, dissolved inorganic carbon (DIC) and turbidity strongly correlated with changes in taxonomic composition (Fig. 4C). Notably, strong spatial variations were observed in all of these environmental factors during our sampling period (Fig. S4).

**Correlations between plankton groups and the surrounding environment**

A key to investigating the underlying mechanisms of variations in plankton community composition is an understanding of different groups’ environmental preferences. To examine the correlations between OTUs and environmental factors, 18S relative proportions (including both autotrophic and heterotrophic protist sequence data) were subjected to a weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008). WGCNA clusters OTUs into modules (ME designated by colours) based on their similarity in changes of relative proportions to the measured environmental parameters. ME blue and ME yellow were found to be positively correlated with nitrogen concentrations (NH$_4^+$ and NO$_3^-$) and negatively correlated with salinity (Fig. 5A). These two
modules mainly consist of chlorophytes, cryptophytes and metazoan OTUs (Table S3). Noticeably, all cryptophyte OTUs were assigned to MEyellow (Fig. 5B), suggesting preference for high nutrient concentrations, low salinity conditions present in the upper estuary waters. Cryptophytes were shown to have more inner-group connections than the other three major microalgal groups in a microbial association network analysis (Fig. 6, Tables S4 and S5). MEbrown was positively correlated with PO$_4^{3-}$ concentration and primary productivity (PPR), comprised mainly of diatoms and dinoflagellates (Fig. 5B). In general, the major microalgal groups apart from chlorophytes were mostly assigned to one or two ME modules and have demonstrated consistent group-specific ecological associations. In contrast, chlorophytes had more diverse responses to differences in their environmental surroundings (Fig. 5B).

Fig. 5. Weighted Gene Co-expression Network Analysis (WGCNA) demonstrating the correlations between modules and environmental parameters. A. WGCNA. The numbers in each cell are Pearson’s correlation coefficients and p values of the correlation test (in brackets). Colour of the cell indicates the correlation between modules and corresponding parameters, with red and blue indicating a positive or negative correlation, respectively. The MEgrey module is comprised of OTUs that could not be assigned to other WGCNA modules. B. Proportions of each major microalgal group present in particular WGCNA modules.

**Discussion**

Comparison of phytoplankton composition based on different taxonomic identification measurements

We examined phytoplankton community composition in the NRE using five different taxonomic approaches, including conventional methods as well as high-throughput sequencing techniques. Microscopy based cell abundance detected high proportions of diatom species, potentially due to their large size rendering them more easily discernable using microscopy, whereas the abundance of relatively small chlorophytes and cryptophytes may have been underestimated. Dinoflagellates contributed less to the total cell abundance compared with the other methods, particularly as determined by 18S sequencing, likely due to their larger genome sizes and high 18S gene copy numbers (Fig. 2) (Lin, 2011; Gong and Marchetti, 2019). Similar to microscopy counts, diatoms also contributed the largest fraction of the phytoplankton community based on biovolume analysis. Microscopy-based approaches (i.e., cell abundance and biovolume) were shown to have a strong bias towards specific taxonomic groups under various environmental conditions (Fig. S2). Along with chlorophytes, diatoms and dinoflagellates that were over-represented at different sections of the NRE had substantial variability.
across different samples. Taxonomic profiles based on microscopy typically contained a single phytoplankton group dominating the community (Fig. 2). In comparing microscopy to 18S sequence analysis at the genus level, when possible, substantial differences were observed among the four major microalgal groups, especially for the smaller-sized chlorophytes and cryptophytes (Fig. 3, Fig. S1). Optical resolution provided by light microscopy often limits the ability to obtain higher resolution taxonomic information from the small phytoplankton groups that lack discernible features, resulting in underestimation and/or misclassification of dominant genera. Underrepresentation of certain microalgal groups may also be attributed to cell preservation issues.

Pigment analysis determines taxonomic composition by binning diagnostic pigments into various taxonomic groups. However, phytoplankton pigment contents can vary as a function of physiological status, primarily as a function of light levels and nutrient status (Eker-Develi et al., 2012). Different phytoplankton groups have been shown to employ distinct metabolic strategies across the spatiotemporal gradients in the NRE. For example, photosynthetic activities are generally elevated in the upper estuary, while increased nutrient acquisition mechanisms are commonly observed in the lower estuary (Gong et al., 2018). Thus, a station- and time-specific phytoplankton pigment content matrix that incorporates such physiological variability would aid in a more accurate characterization of the phytoplankton community.

Metatranscriptomic sequencing is now often being used to infer phytoplankton physiological status by examining gene expression profiles. Although not primarily used for determining taxonomic composition due to variations in RNA expression, such information can be obtained by measuring the transcript abundance of each taxonomically resolved group. Highly conserved genes that display constitutive gene expression may be particularly useful in elucidating microbial community composition as opposed to those genes that are influenced in expression by varying environmental conditions. Due to the limited number of available phytoplankton genomes and transcriptomes within publicly available reference databases, metatranscriptomics does not currently provide the same level of taxonomic resolution as with 18S sequencing (i.e., to the genus or species level). For example, Gong and colleagues (2017) identified transcripts from Lavanderina sp., a bloom-forming dinoflagellate species in the NRE, that were routinely identified as other dinoflagellate species due to the absence of a Lavanderina transcriptome database. Trebouxiophyceae, the most abundant algal class based on 18S sequencing, was ranked 42nd based on mRNA abundance. In fact, only about half of the genera identified by 18S sequencing were also detected in the metatranscriptome sequences, suggesting inaccurate taxonomic binning of transcripts at lower taxonomic levels. In addition, taxonomic proportions of metatranscriptomic sequences can vary as a function of their metabolic rates and processes which may have divergent responses to changes in environmental conditions. Therefore, assessing taxonomy through metatranscriptomics should be used with caution unless the expression patterns of specific genes can be identified to track cell abundance fairly well or when combined with other taxonomic approaches. We attempted an examination of taxonomic proportions of specific gene transcripts to identify expressed genes that resembled our 18S sequence dataset. However, the results lacked statistical significance due to the limited number of metatranscriptomic samples (data not shown). Further comparisons are warranted to identify functional genes that can potentially be used to infer phytoplankton taxonomic composition through metatranscriptomic sequencing.

With the continuing development of molecular technologies and bioinformatics algorithms, 18S sequencing is rapidly replacing conventional microscopy analysis as a preferred method to examine plankton community composition (Johnson and Martiny, 2015). However, 18S sequencing is not without pitfalls, and caution needs to be taken to avoid potential bias. One of the unresolved biases in 18S rDNA analysis is the 18S gene copy number variability across different plankton species. Both genomic sequencing and quantitative polymerase chain reaction (qPCR) methods have been developed to estimate 18S gene copy number in phytoplankton isolates (Zhu et al., 2005; Gong and Marchetti, 2019); however, neither of these approaches could be applied to this study due to a lack of complete reference genomes and absolute measurement of 18S sequence abundance. Instead, we took a comparative approach that examined the ratios of 18S gene copy number between microalgal groups using proportional 18S abundances and microscopic cell counts. At a very coarse taxonomic level, dinoflagellates were found to have consistently higher 18S gene copy numbers than diatoms in all samples in the NRE, whereas the ratios of the other microalgal groups were more variable and likely to be biased (Table S6). As the ratios were inferred from normalizing 18S proportions to cell concentrations, the results are subject to biases in microscopy and is limited in allowing for more quantitative interpretations. Sequencing chloroplast 16S rRNA genes has been shown to provide concordant results with 18S gene sequencing (Needham and Fuhrman, 2016). Although there can also be multiple chloroplast 16S rRNA gene copies per cell, there is less variability compared to copy numbers of 18S genes in phytoplankton.
Spatial environmental gradients drive changes in phytoplankton composition

Lin and colleagues (2017) demonstrated that a few key species can contribute disproportionately and explain a large proportion of the variability in potential carbon export of a given region. Therefore, it is necessary to examine the phytoplankton community with high taxonomic resolution to better understand phytoplankton community dynamics and environmental interactions. Sequencing has revolutionized microbial ecological studies due to its high taxonomic resolution. Strong environmental differences in both chemical and physical properties between upper and lower sections of the NRE result in phytoplankton exhibiting distinct molecular profiles along the estuary (Peierls et al., 2012; Hall et al., 2013; Pael et al., 2014a; Gong et al., 2018). Using 18S sequencing, we observed that these environmental gradients also influence the phytoplankton taxonomic composition. Plankton composition was found to strongly correlate with nutrient concentrations (NO$_3^-$ and NH$_4^+$) (Figs. 4, 5). Nitrogen (N) and phosphorus (P) are essential nutrients for phytoplankton growth. Indeed, P can at times be a limiting nutrient in the NRE (Rudek et al., 1991), but productivity is mostly controlled by availability of N sources, particularly in the lower estuary (Paerl et al., 2010). Phytoplankton groups can regulate metabolic processes to adjust to decreasing N concentrations along the estuary, but their relative abundances are commonly altered due to intrinsic differences in nutrient requirements, nutrient uptake strategies as well as other physiological traits (Litchman, 2007; Falkowski and Knoll, 2011). For example, some dinoflagellates can perform vertical migration behaviours that offer them an opportunity to obtain nutrients deeper in the water column, providing a competitive advantage when nutrients near the surface are depleted as a result of water column stratification (Hall and Pearl, 2011).

Phytoplankton biomass also co-varies with salinity in the NRE (Pinckney et al., 1997). A significant correlation was detected between salinity and plankton community composition (Fig. 4). From our WGCNA analysis, all cryptophyte OTUs were assigned to the MElow module, which negatively correlated with salinity, whereas most diatoms and dinoflagellates were assigned to MEturquoise, a group that is positively correlated with salinity. Our results suggest that cryptophytes in the NRE are primarily freshwater species while most of the diatoms and dinoflagellates prefer marine conditions and are perhaps tolerant to brackish waters.

Group-specific correlations for certain environmental conditions

Louca and colleagues (2016) reported marine microbial responses to environmental gradients are mainly attributed to functional variations instead of changes in taxonomic composition. In our study, both compositional and functional variations were observed along the NRE, which suggests that different eukaryotic phytoplankton groups have specific metabolic niches and low functional redundancy (Gong et al., 2018). By performing a co-occurrence network analysis, cryptophytes were shown to be the most connected microalgal group despite their low relative 18S proportions in the NRE (Fig. 6). Cryptophyte inner-group co-occurrences were observed more often than either connections between cryptophytes and the other three major microalgal groups or inner-group connections within other groups (Table S4). Co-occurrence in microbial networks represent organisms having similar preferences for environmental conditions and/or committing similar or complementary ecological functions (Steele et al., 2011). The strong connection detected between cryptophyte OTUs implies a shared ecological niche in the NRE, which is further verified by the mapping of cryptophyte OTUs to a single WGCNA module. Such correlations suggest cryptophytes prefer high nutrient, relatively low salinity environments like those found in the upper estuary in the NRE. Distinct from diatoms, dinoflagellates and cryptophytes (which had the most consistent group-specific ecological associations), chlorophytes displayed a larger diversity of responses to environmental variations. The majority of chlorophytes were assigned to MEgrey, which is comprised of OTUs with distinctive environmental associations that could not be assigned to other WGCNA modules. Small-sized chlorophytes may be under strong selective grazing pressure in the NRE (Cira et al., 2016), which could obscure the correlations between chlorophytes and bottom-up, abiotic factors as these top-down controls could be a stronger influence on chlorophyte abundance and distribution.

Trebouxiophyceae (belonging to Chlorophyta) were proportionally the most dominant phytoplankton group in the NRE, on average contributing 28% of the total 18S sequences, and were detected to have higher relative proportions in the lower estuary than in the upper estuary (Fig. 3). Trebouxiophyceae is a diverse algal class found in a variety of environments and can cope relatively well under a broad range of environmental conditions (Lemieux et al., 2014; Hodac et al., 2016). Their resilience to environmental stress may render them more competitive in the lower estuary where N concentrations are routinely below the detection limit (perhaps due to their ability to grow on regenerated sources of N) and salinity is not optimal for other microalgae that prefer freshwater environments. Despite its high relative abundance of 18S sequences, members of Trebouxiophyceae were not readily identified from our microscopic analysis, perhaps due to issues with preservation, which further
highlights how major constituents of the plankton community can be overlooked depending on the approach used to determine their composition.

Conclusions

The NRE has been a focus for targeted nutrient reduction strategies to alleviate the impacts of eutrophication and resulting harmful algal blooms. However, management efforts to reduce nutrient loading have not been successful at preventing issues of poor water quality, in part, because phytoplankton community responses resulting in blooms cannot always be traced to unique causes. Through studying the phytoplankton community with high taxonomic and functional resolution using recently developed nucleic acid sequencing techniques, this and previous studies (e.g. Gong et al., 2018) provide a comprehensive examination of the relationships between the plankton community dynamics and their environmental surroundings. This research highlights the added benefits to incorporating molecular approaches for taxonomic identification into water-quality monitoring programs. With continuing efforts addressing currently unresolved biases (Yeh et al., 2019; Catlett et al., 2020), the application of 18S gene sequencing will provide for a more informative assessment of plankton communities and aid in better predicting their responses to future environmental conditions.

Experimental procedures

Sample collection and environmental measurements

Sampling took place in 2012 in conjunction with the NRE modelling and monitoring (ModMon) program (http://paerllab.web.unc.edu/projects/modmon/) operated by the UNC-Chapel Hill Institute of Marine Sciences, and the North Carolina Department of Environmental Quality that has collected water samples in the NRE on a bimonthly to monthly basis since 1994. Water sampling at four of the 11 routinely sampled stations (Modmon stations 20, 70, 120 and 180) were performed in conjunction to the field measurements collected as part of the ModMon program (Fig. 1). Sampling occurred in the months of February, April, June, August, September and December of 2012 to assess the seasonal succession of the phytoplankton assemblages in the NRE. 18S sequences were changed every 15 min or when the flow of water decreased due to particle clogging. Individual filters were placed in Ziploc bags, wrapped in aluminium foil and immediately placed in liquid nitrogen. A minimum of six filters were collected from each station and sampling date. Onshore, filters were stored at −80°C until DNA extractions were performed. Environmental parameters including temperature, salinity, dissolved oxygen, in vivo fluorescence, photosynthetically active radiation, dissolved inorganic nutrients (including nitrite/nitrate $[\text{NO}_2^- - \text{NO}_3^-]$), phosphate $[\text{PO}_4^{3-}]$ and silicate $[\text{SiO}_4^{2-}]$, and primary productivity were measured as described in (Gong et al., 2017). Photopigment analysis via HPLC was performed as described by Pinckney and colleagues (1996) on a Shimadzu LC-20AB HPLC coupled to a Shimadzu SPD M20A in-line photodiode array spectrophotometer (Shimadzu Inc.).

Microscopic quantification

Phytoplankton cell abundances and biovolumes were determined from Lugol’s preserved samples using 5–15 ml in Utermöhl settling chambers followed by identification and enumeration using Leica DMIRB inverted microscope (Wetzlar, Germany). Phytoplankton species were identified to the lowest possible taxonomic level and enumerated at 400 times magnification using phase contrast. For each species, individuals were counted until reaching one of two-stop criteria: (i) 400 individual cells of that species were counted or (ii) 100 fields of view were counted. For some cell types, identification to the genus or even class level was not possible. For example, small, solitary centric diatoms and small flagellates (generally <5 μm maximum dimension) were grouped as small diatoms or small flagellates and not identified further. Cell dimensions were measured using the divisions of an ocular Whipple grid. Biovolumes of simply shaped phytoplankton were calculated using standard geometric formulae (e.g., sphere, ellipsoid, cylinder, parallelepiped, etc.) while biovolumes of cells with more complicated shapes (e.g., Ceratium spp.) were estimated based on cell size categories according to values provided in Olenina (2006).

HPLC pigment analysis

Aliquots of 100–300 ml of water were filtered onto 47-mm Whatman GF/F filters (nominal porosity of 0.7 μm) and immediately frozen at −20°C for diagnostic accessory pigment analysis by HPLC. Within a month of sample collection, filters were extracted in 100% acetone, sonicated and the extract stored at −20°C for approximately 24 h. Following procedures described by Pinckney and
colleagues (2001), acetone extracts (200 μl) were then analysed via HPLC on a Shimadzu LC-20AB HPLC coupled to an in-line photodiode array spectrophotometer (Shimadzu-Benelux, Antwerp, Belgium). Pigment identification and quantification were based on elution time and absorbance spectra comparisons against commercially obtained pigment standards (DHI, Denmark). For all samples measured by HPLC, Chemtax matrix factorization (Mackey et al., 1996) was used to determine the proportion of total chl a contained within five dominant phytoplankton taxonomic classes that were microscopically observed: diatoms, dinoflagellates, chlorophytes, cyanobacteria, and cryptophytes.

**Sample preparation and sequencing**

For DNA extractions, filters were briefly thawed on ice. DNA was extracted from individual cut-up filters using the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocols with an initial bead beating step. Extracted DNA was then checked on a 1% agarose gel and quantified with a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

DNA was diluted 1:10 and amplified with PCR targeted for V4 region of the 18S gene with custom primers. The custom primers include universal linker primer sequences, which were used to link to Illumina adapter sequences and barcodes in the downstream library preparation process, as well as four degenerative bases and 18S V4 primer sequences (Table S7). The 18S V4 primer set from Lin and colleagues (2017) was used, which were developed to avoid mismatches with haplotypes. One micromolar of forward and reverse primers, 1 μM deoxynucleotide triphosphates (dNTPs), 15 μl of Milli-Q H2O, 5 μl of DNA, 1× ExTaq buffer, and 0.125 U ExTaq polymerase (Takara Biotechnology, Madison, WI) were contained in the PCR reaction. The PCR protocol included one cycle of 95°C for 5 min, 27 cycles of 95°C for 40 s, 59°C for 2 min and 72°C for 1 min, with one cycle of 72°C for 7 min. Products were cleaned with the Qiaquick PCR Purification Kit (Qiagen, Hilden, Germany), then amplified with a second PCR reaction to link to the Illumina adapters and custom barcodes (Table S7 for sequences). The PCR reaction contained 0.3 μM of forward and reverse primers, 1 μM dNTPs, 9.5 μl of MilliQ H2O, 2 μl of DNA, 1× ExTaq buffer, and 0.1 U ExTaq polymerase (Takara Biotechnology, Madison, WI), and the protocol included one cycle of 95°C for 5 min, eight cycles of 95°C for 40 s, 59°C for 2 min and 72°C for 1 min, with one cycle of 72°C for 10 min. PCR products were checked on a 2% agarose gel and pooled based on band intensity in order to have a similar amount of DNA from each sample. The pooled sample was run on a 1% SYBR Green (Thermo Fisher Scientific, Waltham, MA) stained gel, and the band was cut out and extracted with Qiaquick Gel Extraction Kit (Qiagen, Hilden, Germany). The product was submitted for sequencing at the UNC High-throughput Sequencing Facility on a single lane of the Illumina Miseq platform (San Diego, CA). Table S8 reports the number of reads for each 18S rDNA sample. Information of metatranscriptomic samples is provided by Gong and colleagues (2018).

For metatranscriptomic analysis, RNA from multiple filters collected at each station were extracted, poly-A selected and then pooled to achieve a minimum mRNA concentration of 80 ng as described by Gong and colleagues (2017). RNA was then provided to the UNC-CH High-throughput Sequencing Facility for library synthesis using the TruSeq mRNA Library Preparation kit and then sequenced using the Illumina HiSeq2000 platform in paired-end mode.

**Bioinformatic pipeline**

For 18S gene sequencing, FastQC (v 0.11.5) was used to assess read quality (Andrews, 2010), and Illumina adapters and/or custom barcodes were trimmed using fastx toolkit (v 0.0.14) (Andrews, 2010). Paired-end 2×250 reads were merged with QIME (v 1.9.0) (Caporaso et al., 2010b). OTUs calling was performed at a sequence similarity at 97% with pick_open_reference command in QIME using the uclust algorithm (Edgar, 2010). The representative sequence of each OTU was then aligned using PyNAST to SILVA 18S gene database as reference (v. 123) to obtain taxonomic information (Caporaso et al., 2010a; Quast et al., 2013).

RNA sequences were analysed as described in Gong and colleagues (2017). In brief, FastQC was used to assess read quality (Andrews, 2010). Paired-end sequence reads from each sample were individually assembled into larger contigs using ABYSS v 1.3.5 with multiple k-mer sizes (from 52 to 96 with a step of 2). The number of sequence reads that aligned to each contig were obtained with SAMtools v0.1.19. MarineRefll (http://ssharma.marsci.uga.edu/Lab/MarineRef2/), a custom-made reference organism database of marine microbial eukaryotes and prokaryotes maintained by the Moran Lab at the University of Georgia, Athens, was used for taxonomical annotation and the Kyoto Encyclopedia of Genes and Genomes database was used for functional annotation.

**Statistical analysis**

A PCA was applied to visualize the distance of taxonomic relative proportions from different measurements. For the taxonomic comparison, only diatoms, dinoflagellates,
chlorophytes and cryptophytes were included in the relative compositional matrix as these were the major eukaryotic phytoplankton groups identified in the NRE and enumerated in all the taxonomic methods being compared.

Sequences were rarefied at 20,000 reads per sample. Relative compositional data from 18S sequencing results were used to assess the spatiotemporal changes in phytoplankton community in the NRE. We performed a log transformation to the relative abundances matrix and applied a low-abundance filter to remove OTUs whose relative abundances were less than 0.0001 in all the samples. To study the correlations between environmental data and taxonomic composition, we performed Mantel correlations to distance matrices of taxonomic compositional and environmental data (9999 permutations). Environmental data were z-transformed before computing for distance matrix, and distance matrices were calculated with the Euclidean method.

Relative compositional proportions were subjected to a WGCNA to assess the correlations between OTU abundances and their environmental surroundings (Langfelder and Horvath, 2008). Non-phytoplankton OTUs were also included to explore their relationships among all protists. Relative abundance was centre log-ratio (clr) transformed to ensure subcompositional coherence. OTUs with similar changes in relative proportions were clustered into a module, and correlations between each module and environmental measurements were examined by WGCNA using Pearson correlation.

To assess the associations between different species in the microbial community, we used Spearman correlation to calculate the relationships among OTUs. OTUs that occurred in <4 samples were excluded from the network analysis. Only statistically significant associations were kept as edges in the network map. All statistical analyses were carried out using vegan and igraph packages in R (Csárdi and Nepusz, 2006; Oksanen et al., 2015).

**Comparative estimates of 18S copy number ratios**

Comparative 18S copy number ratios were estimated as

\[
\text{Ratios of 18S copy number} = \frac{18S_{A/B} \times \text{microscopy}_{A/B}}{18S_{B} \times \text{microscopy}_{B}}
\]

where 18S\%_{A/B} and microscopy\%_{A/B} represent the relative proportions of 18S sequence abundance and cell abundance for functional groups A and B, respectively.

**Data deposition**

All 18S sequences have been deposited in the National Center for Biotechnology Information’s Sequence Read Archive under Bioproject PRJNA413761. All metatranscriptomics sequences are under Bioproject PRJNA304171. ModMon data are publicly available at the following website (http://paerlweb.unc.edu/projects/modmon/) or by request.

**Acknowledgements**

We thank J. Browne, J. Braddy (IMS) and the other members of the ModMon Team who assisted with sample collection and analysis during our study period. We are grateful to K. Delong (UNC) for assistance with DNA extractions and Y. Lin (Duke) for providing guidance and primers for 18S sequencing. We thank N. Cohen, C. Moreno and R. Lampe (UNC) for their assistance in the development of the bioinformatics pipeline. We also thank N. Cassar (Duke), S. Gifford and M. Alperin (UNC) for assistance with data analyses and manuscript preparation. UNC Research Computing generously provided cluster time for sequence analysis. In 2012, ModMon was funded by NC-Department of Environment and Natural Resources, Division of Water Resources, the Lower Neuse River Basin Association and Neuse River Compliance Association and National Fish and Wildlife Foundation. This research was supported by NC Sea Grant projects R/12-HCE-3-NA10OAR4170080 and R/14-HCE-2-NA14OAR4170073 to A. Marchetti and H. Paerl as well as start-up funds to A. Marchetti.

**Author contributions**

Adrian Marchetti and Hans Paerl designed research; Weida Gong contributed new analytical tools; Weida Gong, Hans Paerl, Nathan Hall and Adrian Marchetti performed research; Weida Gong and Adrian Marchetti analysed data and wrote the paper.

**Conflict of interest**

The authors declare no conflict of interest.

**References**


short- and long-term ecological change in the anthropo-
genically and climatically impacted Neuse River estuary, North Carolina, USA. *Estuar Coasts* 33: 485–497.


**Supporting Information**

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

**Fig. S1.** Taxonomic composition of the four major functional groups from microscopy-based cell counts.

**Fig. S2.** Application of RDA to relate relative proportions of four major phytoplankton group measured by different methods to environmental gradients. (A) Diatoms; (B) Dinoflagellates; (C) Chlorophytes; (D) Cryptophytes. Sample stations are indicated by level of shading.

**Fig. S3.** PCA of taxonomic composition inferred from 18S rRNA gene sequences grouped by sampling time.

**Fig. S4.** Physical and chemical parameters measured by the Modmon water quality monitoring program in 2012. Plotted are the annual averages and associated standard deviations of surface samples.

**Table S1.** Relative abundance of four the major algal groups measured by different methods.

**Table S2.** Relative abundance of OTUs measured by 18S rRNA gene sequencing.

**Table S3.** Average relative abundance of phytoplankton groups in WGCNA modules.

**Table S4.** Number of network connections between major algal groups.

**Table S5.** OTU look-up table for network analysis.

**Table S6.** Ratios of 18S rDNA gene copy number between functional groups.

**Table S7.** Primer sequences. Bases in bold are forward and reverse primers for the V4 region of the 18S rRNA gene sequencing, bases in red and green are illumina adaptor sequences, bases in light yellow and orange are universal linker sequences and bases in blue are custom barcodes.

**Table S8.** 18S rRNA gene sequencing statistics.