

Comprehensive Primer Sets and Cost Efficient Multiplex PCR-based eDNA Sequencing for Community Dynamics of Cyanobacteria, Eukaryotic Phytoplankton and Zooplankton in Lake

Shanmei Zou (✉ zousm912@njau.edu.cn)

Nanjing Agricultural University <https://orcid.org/0000-0002-5253-3314>

Lydia Smith

UC Berkeley: University of California Berkeley

Research

Keywords: eDNA, zooplankton in lake, cyanobacteria, eukaryotic phytoplankton, primer sets

Posted Date: October 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-97365/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Comprehensive and accurate monitoring of aquatic microbial diversity will greatly improve the understanding of marine and freshwater ecology, to which cost efficient high-throughput environmental DNA (eDNA) sequencing with effective primers will contribute. Yet our understanding about the aquatic microorganism diversity remains in its infancy due to the diversity complexity and limited gene markers. The algae blooms in lake often cause serious environmental and social problems.

Methods

We developed simple and cost efficient multiplex PCR-based eDNA Illumina sequencing and new gene primers for community diversity, spatio-temporal dynamics and interaction of cyanobacteria, eukaryotic phytoplankton and zooplankton in Lake Tai by large samples. A total of 9 gene barcodes (three regions of 18S, 28S, 23S, two regions of 16S and two regions of *rbcL*) are used, out of which primers of 18S-v4-5, 28S, 16S-new and two *rbcL* markers are newly designed.

Results

The taxonomic assignments of all gene barcodes at each of species, genus, family and phylum level differed from each other. The 23S, 16S and 16S-new are all high effective for detecting cyanobacteria. The 28S could be as useful universal marker for detecting eukaryotic phytoplankton and zooplankton. The 18S-v4-5 shows more variation than 18S-v1-3 and 18S-v9. Two new *rbcL* markers specifically assign Chlorophyta and Cryptophyta successfully. The cyanobacteria abundance accounts for the highest proportion among all taxa detected for both spring and summer, followed by zooplankton, Bacillariophyta and Chlorophyta. The phytoplankton and zooplankton diversity varies significantly from spring to summer, but shows no much geographic difference. Most interactions among phytoplankton and zooplankton are positive in the molecular ecological network. The temperature is shown as the remarkable environmental factor affecting the diversity of phytoplankton and zooplankton seasonally.

Conclusions

It is critical to employ multiple effective gene markers for accurate microbial diversity detection in lake. Our simple and cost efficient multiplex PCR-based eDNA sequencing with effective primer sets would contribute to aquatic phytoplankton and zooplankton monitoring. The diversity of cyanobacteria, eukaryotic phytoplankton and zooplankton in Lake Tai varies significantly from spring to summer, but shows no much geographic variation. The temperature is shown as the remarkable environmental factor affecting the diversity.

Introduction

It is a big challenge for biologists to quantify environmental biodiversity just relying on morphological and behavioral characters since the traditional identification is often biased, time consuming, and dependent on

a declining pool of taxonomic experts for identifying various kinds of organisms, especially for microorganisms¹⁻⁶. Next-generation sequencing (NGS) provides an alternative for enhancing biodiversity monitoring in a wide range of environmental samples by overcoming some of the challenges of labor-intensive and time consuming morphological identification^{1,7-14}. The advantages of NGS make the eDNA metabarcoding approach fast, effective, and cost efficient for monitoring biodiversity without significant damage to the target species or its habitats, where a short DNA region or multiple short DNA regions are amplified and sequenced from environmental DNA (eDNA)^{11,14-23}. The eDNA metabarcoding also has high detection probabilities for rare, cryptic and elusive species^{6,24-26}. One remarkable opportunity provided by eDNA metabarcoding for biodiversity understanding is that it can monitor the dynamics of species, populations and communities over long time periods and across large spatial scales²⁷⁻²⁸.

However, challenges and limitations also exist for amplicon-based eDNA metabarcoding, e.g. reference database and PCR primer bias. First of all, assigning abundant NGS reads at the species level is important for biodiversity monitoring^{23,28-29}, which mainly relies on assigning the reads to the available reference sequences in public databases like GenBank or EMBL. But most species just have one or a few genes, or even no genes in reference databases. In this case, some reads can only be assigned to higher taxonomic levels, which makes it difficult to associate eDNA data with existing biological and ecological knowledge⁶. The ways to build reference database also affect the identification accuracy, like blasting against the NCBI database directly, downloading a local copy of EMBL, or building a local reference²⁸. The ecoPCR is able to obtain relatively complete and accurate reference sequences for one pair primers by silico PCR^{28,30}. Secondly, successful amplification of molecular markers for eDNA depends highly on primer specificity, sensitivity, and efficiency³¹. However, it is almost impossible to amplify all taxa expected successfully for one gene marker, even for the conserved markers like 16S, ITS and 18S. Then special primers need to be designed for specific groups to get comprehensive taxonomic identification. Additionally, one short single gene usually provides insufficient information for identifying all species. Also, conserved gene primers have to be designed to get compatible short fragments with sequencing platforms within a limited length^{10,32}, which often brings difficulty to metabarcoding. Researchers suggested that multiple genetic markers for eDNA metabarcoding should be considered for accurate molecular species identification, especially for eukaryotes³³⁻³⁴.

The method for amplicon-based NGS library preparation is also critical in reducing PCR bias, which mainly differ in whether the sequence adaptors are added by ligation³⁵⁻³⁸, single-step PCR³⁹⁻⁴³ or multi-step PCR⁴⁴⁻⁴⁶. A two-step PCR library preparation is recommended since it avoids potential bias due to the use of different indices in each primary amplification, and it offers increased versatility by minimizing the number of primers that must be synthesized¹. While multiple gene barcodes are employed for amplicon-based metabarcoding the cost of library preparation should also be considered due to multifold cost in PCR, cleanup and DNA quantification, especially for large samples. Multiplex PCR is a widespread molecular biology technique for amplification of multiple gene targets in a single PCR experiment, which could be as a cost-effective method for metabarcoding if we address the challenges of suitable primers used, effective PCR conditions, and good library preparation⁴⁷.

Biodiversity monitoring by eDNA is increasingly being conducted for freshwater and marine ecosystems^{29,48-50}. However, most of these studies focused on fishes and amphibians. Although cyanobacteria, eukaryotic microalgae, and zooplankton all play key roles in aquatic ecosystems (like bloom) studies on these taxa are scarce. The water ecosystem has gradually been altered by climate change, increasing nutrient pollution, and activities of humans⁵¹⁻⁵² which often lead to eutrophication, algal growth and biomass accumulation in the upper photic zone and thus produce bloom among the globe⁵³⁻⁵⁶. Cyanobacterial blooms are currently increasing globally⁵⁵, causing the death of other aquatic animal and destroying the aquatic ecosystem. Actually, not only the cyanobacteria bloom but also many other photosynthetic algae bloom, such as green algae and diatoms, are also serious^{27,57}. While phytoplankton metabarcoding in marine environment has been triggered with 16S or 18S NGS^{27,57} the phytoplankton metabarcoding in inland lakes with bloom is lagging behind. Blooms could also cause a cascade of changes in planktonic microbial communities. There is increasing evidence that there is co-evolutionary arms race between bloom-forming cyanobacteria and their grazers⁵⁵.

Lake Tai (meaning 'Great lake' in Chinese) is China's third-largest freshwater lake and one of the most important water resources for drinking around population cities and development of agriculture and industry. However, the rapid economic and population growth has led to numerous eutrophications. The blooms in Lake Tai have occurred frequently since 1997, which brought ecosystems to the brink and led to serious environmental problems despite major government efforts to clean up the lake^{55,58-60}. Blooms of Lake Tai originally occurred just in one position: Meiliang Bay with cyanobacteria^{54,61-64}. Recent years the blooms often changed with seasons and gradually occurred among the whole lake^{34,56,64-68}. The plankton diversity monitoring among the whole lake is important to understand the lake ecosystems and control the bloom. Currently, the identification of phytoplankton and zooplankton is mainly based on morphological characters by microscope^{64,67-68} which has bias in taxonomic identification as mentioned, especially for cryptic species⁶⁹⁻⁷⁰.

Take Lake Tai as an example, here we developed simple and effective-cost multiplex PCR-based eDNA sequencing and multiple new gene barcodes, aiming to: 1) compare the detection efficiency of 9 gene loci of 18S, 28S, 16S, 23S and *rbcL* for cyanobacteria, eukaryotic phytoplankton and zooplankton; 2) validate the efficiency of multiplex PCR-based eDNA sequencing; 3) reveal the diversity patterns and seasonal community dynamics of cyanobacteria, eukaryotic phytoplankton (each phylum separately) and zooplankton from spring and summer; 4) construct the molecular associate network of phytoplankton and zooplankton; 5) uncover the impact of various environmental factors to the diversity variation of phytoplankton and zooplankton; 6) finally infer the reason for lake algae blooms. Three multiplex PCR systems were carried out to amplify 9 gene loci, including three regions of 18S (18S-v1-3, 18S-V4-5, 18S-v9), two regions of 16S rDNA, on region of 23S rDNA, one region of 28S rDNA, and two regions of *rbcL*. Five gene barcodes, 18S-v4-5, 28S, 16S-new, *rbcL*-Chlo and *rbcL*-Cry, were newly designed in this study. 48 samples from two seasons with microscope identification were used as mock communities to compare the taxonomic assignments of multiple gene barcodes.

Results

Assignments and efficiency of available and new gene barcodes

The proportion of taxonomic assignments for each marker was shown in Fig. 1. Among the gene barcodes, only 18S and 28S obtained zooplankton assignments. Particularly for 18S-v1-3 and 18S-v9, almost half of taxa were assigned to zooplankton. The 18S-v4-5 marker assigned even proportion for Chlorophyta, Bacillariophyta and zooplankton, together with a small proportion for Cryptophyta, Chrysophyta, Pyrrophyta, Ochrophyta and fungi. 28S produced similar taxonomic resolution as 18S-v4-5. Both 23S and 16S obtained a higher proportion of cyanobacteria assignments, with a small proportion of Bacillariophyta, Cryptophyta, Chrysophyta, Euglenida and Pyrrophyta. The 16S-new marker also got cyanobacteria assignments, but also got other assignments of Bacillariophyta, Chlorophyta, Chrysophyta, Cryptophyta, Pyrrophyta and bacteria. We also designed specific primers for amplifying Cryptophyta (*rbcL-Cry*) and Chlorophyta (*rbcL-Chlo*), which assigned a high proportion of Cryptophyta and Chlorophyta respectively. However, there was a proportion of reads that could not be assigned to any taxonomic level for 18S, 28S and 16S, being considered as unassigned.

The taxonomic difference of all gene barcodes was compared at each of species, genus and family level respectively for each phytoplankton phylum and zooplankton (Fig. 2–3, Figure S1-S2). The 28S marker produced the most different assignments from other markers at each of species, genus and family level for Bacillariophyta, following which the 18S-v4-5 produced the second most different species assignments from other markers (Fig. 2). For Chlorophyta, the 18S-v4-5 marker produced the most different species assignments while the 28S produced the most different genus assignments and the 16S-new produced the most different family assignments (Fig. 2). The 16S-new also produced the most different assignments at each of species, genus and family level for cyanobacteria (Fig. 3). For zooplankton, the most different assignments among all markers were from 18S-v4-5 at each of species, genus and family level (Fig. 3). For Chrysophyta and Cryptophyta, the 18S-v4-5 marker produced the most different species assignments (Figure S1). The 28S and 16S-new markers produced the most different genus assignments and most different family assignments respectively for Cryptophyta (Figure S1), and produced the most different species assignments for Euglenida and Pyrrophyta respectively (Figure S2). The species assignments of all gene barcodes were also compared with the microscope observations for each phytoplankton phylum (Figure S3). For each phylum, while some markers showed the same species assignments with the microscope observations some markers showed unique species assignments from microscope observations. For Bacillariophyta and Cyanobacteria, the microscope observations showed some unique species assignments from all gene barcodes.

Finally, all gene loci showed different assignment rate at species, genus, and family level for each phytoplankton phylum (Fig. 4). The 28S and *rbcL-Cry* markers showed higher species assignment rate for Chlorophyta, Bacillariophyta, Cryptophyta, Pyrrophyta, Heterokontophyta, Ochrophyta, and Zooplankton. The 23S marker showed the highest species assignment rate for Cyanobacteria. All gene barcodes showed generally higher genus and family assignment rate than species assignment rate.

Community diversity patterns

The community diversity analysis was performed for April and August respectively. After filtering reads with low identify score the “clean reads” for samples of April and August were 4,348,278 and 4,816,816 respectively (Fig. 5A), which were assigned to cyanobacteria, eukaryotic phytoplankton, zooplankton, fungi, few bacteria and some unassigned taxa. The proportion of various taxonomic assignments was generally consistent between April and August, where the cyanobacteria made up a large proportion followed by zooplankton, Chlorophyta, Bacillariophyta and Cryptophyta.

The top 30 species and top 30 genera from phytoplankton are shown in Fig. 5B and 5C. While the top 5 microalgae species in April were *Microcystis aeruginosa* (Cyanobacteria), *Synechococcus rubescens* (Cyanobacteria), *Cyclotella choctawhatcheeana* (Bacillariophyta), *Teleaulax acuta* (Cryptophyta), and *Cryptomonas curvata* (Cryptophyta) the top 5 microalgae species in August were *Microcystis aeruginosa* (Cyanobacteria), *Cyanobium gracile* (Cyanobacteria), *Synechococcus rubescens* (Cyanobacteria), *Aulacoseira granulate* (Bacillariophyta), and *Cyclotella choctawhatcheeana* (Bacillariophyta). It was indicated that there was a difference for the top 30 species between April and August. In April, the relative abundance of the top species in West and North was higher than that in other regions. But in August, the relative abundance of some top species was also higher in Center and East (Fig. 5B). In genus level, there was also a difference for the top genera between April and August (Fig. 5C). In April, the relative abundance of the top genera in North and West was higher than that in other regions. In August, each region of the lake showed some genera which had higher relative abundance. The top 30 phytoplankton families for the two seasons are shown in Figure S4A, and the top 30 zooplankton species for the two seasons are shown in Figure S4B. The detailed taxa identified and their relative abundance at each of species, genus and family level for phytoplankton from the two seasons are listed in Table S1, and the detailed taxonomic identification and abundance for zooplankton is listed in Table S2.

Community seasonal succession and molecular ecological network

The taxonomic assignments of all gene barcodes were combined for non-metric multidimensional scaling (NMDS) analysis, Canonical Correspondence Analysis (CCA) analysis and molecular ecological network construction as below.

The NMDS analysis was carried out for each phytoplankton phylum, zooplankton, and fungus independently. For both mock and non-mock communities, it was demonstrated that the samples in April and August were divided clearly for each of Cyanobacteria, Bacillariophyta, Chlorophyta, Chrysophyta, Cryptophyta, Euglenida, Fungi and Zooplankton (Fig. 6). The samples for Pyrrophyta from two seasons were not clearly separated despite some diversity dissimilarities. The reason was possibly that the read abundance was insufficient for clustering. However, there was not much diversity dissimilarities among samples of different regions of the whole lake in one season. For all of Cyanobacteria, Bacillariophyta, Chlorophyta, Chrysophyta, Cryptophyta, Euglenida, Fungi and Zooplankton, the samples from one season generally grouped together without obvious division.

The CCA analysis was further conducted for each phytoplankton phylum, zooplankton, and fungus respectively, to explain the impact of environmental factors to their community composition. Samples that

were amplified by different gene barcodes were analyzed separately to avoid sample bias, which means that the mock community was analyzed independently (Fig. 7, Figure S5). For all of phytoplankton, zooplankton, and fungi, it was shown that the temperature (WD) was remarkable in affecting community diversity variation among the whole lake from April to August with the longest vector for both mock and non-mock communities, where the samples in the two seasons were separately clearly (Fig. 7, Figure S5). However, the samples in one season clustered together without apparent dissimilarity for all taxa analyzed. There was close correlation for temperature and the total nitrogen (TN).

The molecular ecological network of assigned taxa from the mock community was constructed for April and August separately (Fig. 8). After filtering taxa with less abundance by standard parameters, the various taxa from Cyanobacteria, Chlorophyta, Chrysophyta, Bacillariophyta, Cryptophyta, zooplankton, fungi, few bacteria, and some un-assignments were included for molecular network construction. It was indicated that the most interactions among all taxa included were positive for samples in both April and August. Particularly for samples in August, there was only one negative correlation between two unassigned taxa. For samples in April, interactions among some taxa in Chlorophyta was negative.

Comparison of singleplex and multiplex PCR-based eDNA sequencing

The assignment consistency of singleplex and multiplex PCR was verified by observing the rate of their shared assignments to their total assignments. It was shown that there was assignment dissimilarity at each of species, genus and family level between singleplex PCR and multiplex PCR for most of the gene barcodes from 6 samples in the mock community (c1, c2, c3, c4, c5, c6) (Figure S6). We also found that some gene barcodes in the multiplex PCR systems amplified more taxa than in the singleplex PCR system, but some gene barcodes in the singleplex PCR system also amplified different taxa from the multiplex PCR system. Cross amplification existed between singleplex and multiplex PCR reactions.

Discussion

The aquatic environment exhibits enormous microbial diversity which often interacts with each other to affect the ecosystem. The enormous microbial diversity brings major challenges to model microbial systems and to explain patterns of community variation across space and time, especially for lakes with serious environmental problems like bloom. Yet, our understanding about the aquatic microorganism diversity remains in its infancy, especially for the eukaryotic organisms while the bacteria and fungi monitoring have been already triggered by NGS. Take Lake Tai as an example, here we developed simple and cost-efficient multiplex PCR-based eDNA sequencing strategy and multiple new gene barcodes for revealing diversity patterns and seasonal community dynamics of microbial community diversity with lake algae bloom.

Gene barcodes for phytoplankton and zooplankton eDNA sequencing

For both traditional DNA barcoding and eDNA metabarcoding, conserved primers of gene barcodes is useful for detecting a wide range of targets with broad sensitivity, such as the primers of 16S, ITS and 18S which are well-tested for obtaining a relatively wide range of microbial identification^{17,71-72}. But even for the conserved primers, they do not always have equal affinity for all possible DNA sequences since different species often vary in these primer-amplified regions, which could consequently induce bias during PCR amplification⁷¹. So multiple gene barcodes and specific primers would greatly contribute to microbe diversity monitoring.

In this study, we focused on the diversity analysis of prokaryotic and eukaryotic microalgae and zooplankton, especially for bloom microalgae. Firstly, we aimed to compare the identification efficiency of multiple gene barcodes at each of species, genus, family and phylum level for phytoplankton and zooplankton, especially for five gene loci newly designed. It was clear that the 9 gene barcodes produced different assignments from each other, especially for the new markers. The 16S-new could be as effective candidate gene for identifying cyanobacteria since it assigned many different taxa from 23S which specially identifies cyanobacteria. But the 16S-new primers also obtained sequences longer than 600 bp. We will focus on designing primers for amplifying shorter 16S-new fragments for compatibility with NGS sequence platforms. The 28S was effective in identifying phytoplankton and zooplankton, especially for Bacillariophyta, which could be universal marker for identifying aquatic eukaryotic organisms. For three regions of 18S, the new 18S-v4-5 assigned relatively even proportion for various groups, and produced different taxonomic assignments from 18S-v1-3 and 18S-v9 at each of species, genus and family level. We had tried to use available 18S-v4 primers to amplify the samples⁵⁷, but they failed. Among the three regions of 18S, the 18S-v4-5 region is more variable for identification. In future studies, we will continue to optimize 18S-v4-5 primers for obtaining fragments shorter than 500 bp. The 18S-v9 region has been usually used for identifying eukaryotic organisms by NGS³². In this study, the 18S-v9 was proved useful for identification of eukaryotic phytoplankton and zooplankton. Compared with 18S-v4-5 and 18S-v9, the 18S-v1-3 was more effective in distinguishing zooplankton. We also designed specific *rbcl* primers which specifically amplified Chlorophyta and Cryptophyta successfully and produced high-resolution in species identification. In conclusion, we prove that it is important to combine multiple gene barcodes to get comprehensive taxonomic diversity for microalgae and zooplankton.

Finally, the microscope species identification was compared with the taxonomic assignments of each gene marker for each phytoplankton phylum. The microscope observation detected some species of Bacillariophyta and Cyanobacteria which were not identified by multiple gene barcodes, which is because there is no any sequence of the gene barcodes deposited in the public databases for these species as references. We combined ecoPCR and blastn to get relatively complete reference database. But there were still some sequences which could not be assigned to any taxa due to the incomprehensive reference sequences in public database. So it would be important to supplement as many as marker sequences into the public database to form a comprehensive reference database⁷². On the other hand, the multiple gene barcodes assigned many species that the microscope observations did not discover. These species undiscovered by microscope are possibly cryptic species or species that are very difficult to be distinguished by tiny morphological characters⁶⁹⁻⁷⁰.

Community diversity patterns and succession for algae bloom

The cyanobacterial blooms in Lake Tai have led to serious environmental and societal problems, with long-term negative impacts on water quality, fisheries, aesthetics, tourism and other economic activities^{55,58,60}. Most studies about Lake Tai focused on the impact of environmental factors to cyanobacterial blooms^{64,66} where the microscope observation was often used for phytoplankton identification. The lack of conserved genes for identifying phytoplankton makes it significant to employ multiple gene loci for phytoplankton barcoding. Zooplankton identification by barcoding has already been conducted for Lake Tai with single COI gene⁶⁵. But it would be important to comprehensively understand zooplankton diversity among the whole lake from different seasons and its interaction with phytoplankton.

Based on the multiple gene barcodes, we aim to reveal the diversity patterns, spatiotemporal dynamics and molecular network of phytoplankton and zooplankton communities in Lake Tai, in combination with the impact of environment factors. Firstly, the diversity analysis at each of species, genus, family and phylum level was performed. The proportion of each group assigned between April and August was generally consistent, but the total reads of samples in August was higher than that in April for the same sample size, which indicates that the species abundance in summer is larger than spring. The top phytoplankton taxa at each of species, genus and family level between April and August were also different between April and August. These diversity variations suggest that the growth of phytoplankton changes with the time. However, there was no apparent difference about the diversity composition from spring to summer. The heatmap analysis showed that the diversity of various top taxa was slightly different among the five regions of the lake in one season.

The NMDS results indicated that the samples between April and August were separated clearly for each phytoplankton phylum, zooplankton and fungi, but the samples within one season were not separated clearly. These suggest that the diversity composition of phytoplankton and zooplankton in Lake Tai varies from spring to summer. Then what could cause the diversity variance of phytoplankton, zooplankton and fungi from different seasons? The CCA analysis was conducted for each phytoplankton phylum, zooplankton and fungi, which indicated that the samples from April and August were also separated clearly for all of phytoplankton, zooplankton and fungi. The temperature was shown as the remarkable environmental factor affecting the diversity variance among temperature, TP, TN, NH₄⁺-N, and NO₃⁻-N. So the CCA analysis could give an explanation that the diversity of phytoplankton and zooplankton in April is significantly different from August since the temperature in summer is much higher than spring. The satellite data of 11 years from the lake showed that high temperatures and nutrient concentrations in spring promoted cyanobacterial growth⁵⁶. Although the detailed bloom mechanism is not determined it has been pointed that the temperature stimulates cyanobacteria bloom in several ways like nitrogenase activity and growth rates^{56,73-80}. Additionally, from our results, we point that except cyanobacteria the temperature could also affect the diversity variance of eukaryotic phytoplankton and zooplankton. Finally, the molecular association network showed that the interactions among most taxa of phytoplankton, zooplankton and fungi were positive for samples in both spring and August. This is possibly because the taxa among the

different taxonomic groups have positive correlations in the aquatic food chain⁵⁵. While microalgae grow quickly their grazer can also have high biomass. In future studies, we will associate the bacteria diversity and perform biodiversity monitor for long time of years with more environmental factors involved to better understand the bloom.

Strategy of multiplex PCR-based eDNA sequencing for microbial diversity

While multiple gene barcodes are used in high-throughput amplicon sequencing the cost of amplicon library preparation is also multifold in target PCR, DNA cleanup, PCR enrichments, DNA quantity and index and adaptor ligation, especially for large samples. Except the cost, the laboratory work is also very time-consuming for multiple gene loci. Multiplex PCR is similar to the singleplex PCR except that each sample is designed to amplify and detect multiple target sequences rather than only a single target, which makes it possible to prepare the multiple amplicon NGS libraries once while the multiple targets are amplified together. However, multiplex methods must deal with interactions between multiple sets of primers that may have different annealing temperatures and may cross-react and generate primer dimers. Henrik et al. (2018)⁴⁷ employed multiplex PCR in arthropod phylogenetic analysis to show the efficiency of multiplex PCR-based amplicon Illumina sequencing.

Here we developed multiplex PCR-based amplicon Illumina sequencing in environmental microbial diversity. Firstly, we selected and tested abundant primers of various gene loci to ensure the PCR efficiency of single locus. Then we optimized the annealing temperatures under which the multiple gene loci could be amplified successfully in a reaction mixture, without cross-reaction targets and apparent primer dimers. Finally, three multiplex PCR systems were constructed for the 9 gene loci since it would be impossible to combine all the gene loci in one reaction mixture because some of them had different annealing temperatures. So based on the two-step library preparation, we used a total of three locus-specific PCRs, three cleanup reactions and one indexing PCR for 9 gene loci of one specimen, saving at least 6 cleanup reactions and 6 DNA quantities for each specimen, at a total cost of < 5 USD due to reagent difference. By Illumina Miseq sequencing, about 100000 raw reads were generated for each specimen in this study, costing about 6–8 USD to recover all loci. The sequence cost depends on the choice of sequence yield. Thus, our multiplex PCR-based amplicon NGS sequencing make it possible to generate nine locus dataset for a total cost of < 4 USD with sequence yield of 50000 raw reads per specimen.

However, the taxonomic difference existed between singleplex and multiplex PCR, which means that their taxonomic assignments were not completely consistent. We found that both multiplex and singleplex PCR reactions generated some unique assignments. There was cross-identification between singleplex and multiplex PCR strategies. PCR bias affecting amplicon accuracy is inevitable, like DNA Polymerase, PCR cycle number, template concentration and annealing temperature, even for single amplicon PCR¹. Even the singleplex and multiplex PCR had almost same PCR conditions their locus primers were significantly different and thus could affect the PCR reaction. Whatever, the multiplex PCR provides cost-efficient, quick and accurate amplicon NGS sequencing despite some bias. We will further focus on optimizing the multiplex PCR condition for assignment accuracy.

Material And Methods

Sampling, microscope identification and DNA extraction

The Lake Tai was divided into 5 regions based on the reported plankton distribution experience (Figure S7). 20–25 samples were collected from each region on April 1 to 3 and August 1 to 3 in 2017. The number of total samples collected in April and August was the same (125 samples for each season). The samples were collected by use of a plankton net (0.2 μm) through which 30 L of water were passed. Mixed samples were collected at each sampling site by combining an equal volume of water samples collected from 0.5 m depth, 1.0 m depth and 2.0 m depth. The phytoplankton of 20 samples from each season were identified by microscope as mock community for comparing metabarcoding assignments, where the 20 samples were generally uniformly distributed among the 5 regions. DNA was extracted by OMEGA E.Z.N.ATM Mag-Bind Soil DNA Kit for each sample. The water temperature (WD), total phosphorous (TP), total nitrogen (TN), $\text{NH}_4^+\text{-N}$ and $\text{NO}_3^-\text{-N}$ were detected and analyzed based on a standard method common used⁶⁴.

Primer selected and new primer design

First, multiple gene barcodes that have previously proved useful were selected from published papers. We aimed to select multiple gene barcodes which were useful to identify a wide range of general assignments for both phytoplankton and zooplankton. The different regions of each gene marker should also be considered due to their different variance to ensure sufficient diversity for taxonomic identification. One pair of locus-specific primers should only amplify one single clear target. The length of the target should also be below 600 bp for compatibility with the maximum read length of the Illumina MiSeq system. Based on these criteria, about 50 published pairs of locus-specific primers were selected and tested initially. Finally, 18S-v1-3, 18S-v9, 23S and 16S which already have available primers and could be amplified successfully were selected (Table 1). Due to PCR failure of available 18S-v4 primers, new 18S-v4 primers were designed. New primers of 28S were also designed. We also designed primers for *rbcL* which could be a specific marker to obtain Chlorophyta and Cryptophyta. After downloading related reference sequences from EMBL the program DegePrime was used to design potential primers⁸¹.

Table 1
Multiplex PCR systems and primers used and designed in this study.

PCR system	Fragment	Primer Name	Sequence (5' - 3')	Reference	Fragment size (bp)
A	16S	CYA359F	GGGGAATYTTCCGCAATGGG	Ulrich et al., 1997	450
		CYA781R	GACTACWGGGGTATCTAATCCCWTT	Ulrich et al., 1997	
	18S-v1-3	Euklf	CTGGTTGATCCTGCCAG	Diez et al., 2001	560
		Euk516r	ACCAGACTTGCCCTCC	Diez et al., 2001	
	18S-v9	V8f	ATAACAGGTCTGTGATGCCCT	Bradley et al., 2016	375
		1510R	CCTTCYGCAGGTTACCTAC	Amaral-Zettler, et al., 2009	
	23S	p23SrV_f1	GGACAGAAAGACCCTATGAA	Sherwood et al., 2007	414
		p23SrV_r1	TCAGCCTGTTATCCCTAGAG	Sherwood et al., 2007	
B	28S	1FC1RF272F	GAGACCGATAGCRMACAAGT	This study	310
		1FC1RF593R	CTYGGTCCGTGTTTCWAGAC	This study	
	18S-v4-5	V4-384F	GYTGCAGTTAAAAMGCTCGT	This study	500–520
		V4-879R	TTCAGNCTTGCGACCATACT	This study	
C	16S-new	16SF174F	CGGTAAGACRGRGGATGCAA	This study	550–700
		16SR457R	CGACACGAGCTGACGACAGC	This study	
	<i>rbcL</i> -Chlo	<i>rbcL</i> ChloF24F	CGTATGACTCCWCAACWGG	This study	400
		<i>rbcL</i> Chlo411R	GGTTTAATWGTACAWCCTAA	This study	
	<i>rbcL</i> -Cry	<i>rbcL</i> CryF84F	ATGTTCCGTAYGACWCCTCA	This study	350

PCR system	Fragment	Primer Name	Sequence (5' - 3')	Reference	Fragment size (bp)
		<i>rbclCry405R</i>	GCWGGACCTTGGWAAGTCTT	This study	

Multiplex PCR and Library preparation

Firstly, we amplified each locus individually using a Qiagen Hot-Start PCR kit to confirm that they will produce only one single clear fragment, as visualized after agarose gel electrophoresis. PCR reactions were carried out in a total volume of 10 μ L, using 5 μ L Master Mix, 1 μ L DNA template (2–10 ng/ μ L), and 0.1 μ M of forward and reverse primer. Then multiple pairs of locus-specific primers which amplified single clear fragment were mixed together for gradient PCR (annealing temperatures from 45 °C to 68 °C) to select optimal multiplex PCR conditions under which multiple gene loci could be amplified together with same annealing temperatures and without cross-reactions. After optimization, four gene targets (18S-v1-3, 18S-v9, 23S and 16S) could be amplified in one PCR reaction (A) with annealing temperature 63 °C, two gene targets (*rbcl-Chlo* and *rbcl-Cry*) could be amplified in one PCR reaction (B) with annealing temperature 55 °C, and two gene targets (28S and 18S-v4) could be amplified in one PCR reaction (C) with annealing temperature 55 °C (Table 1). Multiplex PCR reactions were carried out in a total volume of 25 μ L with Qiagen Multiple PCR Plus kit, using 12.5 μ L Master Mix, 1 μ L DNA template (2–10 ng/ μ L), 0.1 μ M of forward and reverse primer, and Q-Solution. PCR conditions for the three multiplex PCR systems were: 95 °C for 5 min, 30 cycles of 95 °C for 30 s, primer-specific annealing temperatures for 1 min 30 s, 72 °C for 1 min 30 s, with a final extension of 72 °C for 1 min. Based on the optimized multiplex PCR conditions, two-step NGS library preparation was performed subsequently. Primers were synthesized with the locus-specific sequence on the 3' end and a 5' tail containing sequence matching the TruSeq sequencing primer binding site (The forward primer was synthesized with a 5' tail of ACACTCTTCCCTACACGACGCTCTTCCGATCT and the reverse primer with a 5' tail of GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT).

The first PCR was conducted with these tailed primers using annealing temperatures 55 °C for PCR system A and B, and 49 °C for PCR system C. Triple PCR products for each sample were pooled, which was then cleaned by Solid Phase Reversible Immobilization (SPRI) Beads. After purification, the cleaned PCR products were quantified. Then equal amounts of PCR products from the three multiplex PCR systems for each sample were pooled for each sample. Indexing PCRs were performed for each pooled amplicon with indexing primers annealing to 5'-tails of the locus-specific PCR primers. The index PCR conditions was: 95 °C for 5 min, 12 cycles of 95 °C for 30 s, 60 °C for 1 min 30 s, 72 °C for 1 min, with a final extension of 72 °C for 1 min. The indexing primers include dual indexes and Illumina sequencing adapters provided by Vincent Coates DNA Sequencing Laboratory at UC. The indexed products were cleaned up and quantified by Qubit. Then all the amplicons were pooled in equal amounts for sequencing on an Illumina MiSeq 600 cycles. The whole workflow was shown in Figure S8.

40 out all samples from two seasons were set as mock community to compare the taxonomic difference of microscopic observations and eDNA metabarcoding, where the various phytoplankton phyla had been

previously identified by microscope. The five new gene barcodes designed in this study were amplified in the mock community for comparing identification efficiency of all gene loci and microscopic observations. For comparing the assignment consistency between singleplex and multiplex PCR, we selected 6 samples to perform the single PCR for each locus per specimen, where the PCR conditions were the same as multiplex PCR.

Bioinformatics analysis

The ecoPCR⁸² was firstly run to build the original database for each gene marker on the EMBL standard nucleotide database from both std and wgs⁸³. For each marker, 3 mismatches was set for silico PCR. As the CRUX module in Emily (2019)³⁰ reported, blastn⁸⁴ was then used to query the seed databases against the NCBI non-redundant nucleotide database to increase the breadth of reference sequences and capture sequences without barcode primers in EMBL⁸⁵. Taxonomy files were retrieved using Entrez-qiime⁸⁶. We also added our previous published phytoplankton sequences from traditional DNA barcoding into the databases for 16S and *rbcL*. Finally, if there were still reads which could not be assigned to any taxa we blasted them against the whole NCBI nucleotide database again to get the unique perfect-matched species assignment.

Sequence analysis for all gene barcodes was performed using QIIME2⁸⁷. Raw sequence fastq files were firstly demultiplexed to separate the samples based on the indexes. PCR primers were removed by cutadapt. Dada2 was applied for denoising, dereplicate-sequence, filtering chimera and merging the paired reads. For 18S-v1-3, 18S-v4-5 and 16S-new, paired reads merged with good quality score were used or only forward reads were used if the reads could not be perfectly jointed. After quality filter, taxonomic classification was performed by feature-classifier module. For reads that were identified as unsigned we blasted them against the whole NCBI (nt/nr) database to get the unique perfect-matched species assignment with the highest identity score 1. Finally, the taxonomic assignments and reads abundance for all gene loci were obtained for biodiversity analysis.

The mock community was used to compare the identification efficiency of all gene loci. The proportion of different taxa for each marker was counted at the phylum level. Then at each of species, genus and family level, the taxonomic dissimilarity of all markers for phytoplankton and zooplankton was compared by Venn package in R, where the comparison was performed for each phytoplankton phylum, including Cyanobacteria, Chlorophyta, Bacillariophyta, Cryptophyta, Pyrrhophyta, Euglenophyta and Chrysophyta. The assignment rate of each gene marker at each of species, genus and family level was also performed respectively for each phytoplankton phylum, which meant the proportion of reads assigned to each taxonomic level to total reads from assignments of each marker for each phytoplankton phylum.

Taxonomic assignments from all gene loci were combined for biodiversity analysis. The community diversity analysis of two seasons was conducted respectively from the total ASV (amplicon sequence variant) of all samples including the mock community, which was performed by heatmap implemented in R package. For all samples, pairwise community dissimilarity was calculated using Bray-Curtis as implemented in the vegan package⁸⁸. The non-metric multidimensional scaling (NMDS) plots based on Bray-Curtis distance was generated to reveal the diversity dynamics of samples among the whole lake in

April and August respectively by *vegan*⁸⁸, which was performed for each phytoplankton phylum and all of zooplankton. The Canonical Correspondence Analysis (CCA) analysis was performed to evaluate the correlations of environmental factors and diversity dynamics of samples among the whole lake in April and August, in the correspondence analysis by calling the “*cca*” function from *vegan* package⁸⁸. The CCA analysis was also conducted for each phytoplankton phylum and all of zooplankton, where the samples amplified by different gene loci were analyzed independently to avoid sample bias. Permutation tests were performed to evaluate the significance of overall models. The *vegan* package was employed in R 3.6.1. For understanding the interactions among various taxa, the ecological association networks were constructed by molecular ecological networks (MENs) models in Deng (2012)⁸⁹. The cytoscape was used to visualize the networks (<https://cytoscape.org/>). For obtaining comprehensive associations, the samples from the mock community were used for the molecular network construction.

Conclusions

Comprehensive and accurate identification of aquatic microbial diversity is important to understand marine and freshwater ecology, especially for algae blooms. The comprehensive primer sets of 9 gene markers among 18S, 28S, 23S, 16S and *rbcL*, including our newly designed ones, assign cyanobacteria, eukaryotic phytoplankton and zooplankton differently at each of species, genus, family and phylum level, which suggests that it is critical to employ multiple effective gene markers for aquatic accurate microbial diversity detection. Our newly developed simple and cost efficient multiplex PCR-based eDNA sequencing with effective primer sets would contribute to lake phytoplankton and zooplankton monitoring. The cyanobacteria is dominant in the phytoplankton, followed by Bacillariophyta and Chlorophyta in both spring and summer in Lake Tai. The phytoplankton and zooplankton diversity varies significantly from spring to summer, but shows no much geographic difference among the lake. Most interactions among phytoplankton and zooplankton are positive in the molecular ecological network. The temperature is shown as the remarkable factor affecting the community dynamics. These results suggest that the temperature possibly plays key role in algae bloom. In future studies, we will associate the bacteria diversity and perform microbial biodiversity monitor for long time of years to better understand the lake microbial diversity and algae bloom.

Declarations

Ethics approval and consent to participate

This study is not involved in human participate, human data and human tissue.

Consent for publication

This study does not contain any individual person's data in any form.

Availability of data and materials

The NGS sequence data of all samples obtained in this study are deposited in NCBI. The SRA accession for the NGS sequences is PRJNA655206. The submission information for each sample mainly includes the reads sequence in fastaq form, the sampling location, the sequence platform. These data will be public after the paper is published.

Competing interests

The authors have no competing interests.

Funding

The financial support from the China Postdoctoral Science Foundation (2014M561661, 2015T80558), National Natural Science Foundation of China (NSFC) (31600294) and the Fundamental Research Funds for the Central Universities (KJQN201742, Y0201600141) was gratefully acknowledged.

Author contributions

Shanmei Zou designed the project, conducted the experiment, data analysis and writing. Lydia Smith contributed the experiment and writing.

Acknowledgement

The project was supported by the Bioinformatics Center of Nanjing Agricultural University. Thanks very much the experimental platform support from the Laboratory of ecosystem ecology of Nanjing Agricultural University. We also thank the suggestions from Emily E. Curd (University of California - Los Angeles) about the reference library building, some coding and figure help from Meng Wang and Yuxiao Hua (Nanjing Agricultural University).

References

1. Gohl, D. M. et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol.* **34**, 942–949 (2016).
2. Burivalova, Z., Game, E.T. & Butler, R.A. The sound of a tropical forest. *Science* **363**, 28–29 (2019).
3. Khelifa, R. Sensitivity of biodiversity indices to life history stage, habitat type and landscape in Odonata community. *Biol Conserv.* **237**, 63–69 (2019).
4. Rajan, S.C., Athira, K., Jaishanker, R., Sooraj, N.P. & Sarojkumar, V. Rapid assessment of biodiversity using acoustic indices. *Biodivers Conserv.* **28**, 2371–2383 (2019).
5. Outhwaite, C.L. et al. Complex long-term biodiversity change among invertebrates, bryophytes and lichens. *Nat Ecol Evol* **4**, 384–392 (2020).
6. Beng, K.C. & Corlett, R.T. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. *Biodivers Conservn.* **29**, 2089–2121 (2020).
7. Pace, N.R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–40 (1997).

8. Caron, D.A. New accomplishments and approaches for assessing protistan diversity and ecology in natural ecosystems. *Bioscience* **59**, 287–99 (2009).
9. Fuhrman, J.A., Cram, J.A. & Needham, D.M. Marine microbial dynamics and their ecological interpretation. *Nat Rev Microbiol.* **13**, 133–46(2015).
10. Bahram, M., Anslan, S., Hildebrand, F., Bork, P. & Tedersoo, L. Newly designed 16S rDNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. *Environ Microbiol Rep.* 9999(9999) (2018).
11. Alexander, J.B. et al. Development of a multiassay approach for monitoring coral diversity using eDNA metabarcoding. *Coral Reefs.* **39**, 159–171 (2020).
12. Cowart, D.A., Matabos, M., Brandt, M.I., Marticorena, J., Sarrazin, J. Exploring environmental DNA (eDNA) to assess biodiversity of hard substratum faunal communities on the lucky strike vent field (Mid-Atlantic Ridge) and investigate recolonization dynamics after an induced disturbance. *Front Mar Sci.* **6**, 783(2020).
13. Sales, N.G. et al. Space-time dynamics in monitoring neotropical fish communities using eDNA metabarcoding. *BioRxiv.* <https://doi.org/10.1101/2020.02.04.933366> (2020).
14. Abad, D. et al. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. *Mar Biol.* **163**, 149 (2016).
15. Eric, J.M. et al. A novel ultra high-throughput 16S rDNA gene amplicon sequencing library preparation method for the Illumina HiSeq platform. *Microbiome* **5**, 68(2017).
16. Powell, J.E. et al. Modulation of the honey bee queen microbiota: Effects of early social contac. *Plos One*, **13**(7), e0200527 (2018).
17. Eisenstein, M., Microbiology: making the best of PCR bias. *Nat Methods.* **15**(5), 317 (2018).
18. Alberdi, A., Aizpurua, O., Gilbert, MTP. & Bohmann, K. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods Ecol Evol.* **9**, 134–147(2018).
19. Leempoel, K., Hebert, T. & Hadly, E.A. A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. *Proc Royal Soc B Biol Sci.* **287**, 20192353 (2020).
20. Takahara, T., Iwai, N., Yasumiba, K. & Igawa, T. Comparison of the detection of 3 endangered frog species by eDNA and acoustic surveys across 3 seasons. *Freshw Sci.* **39**, 18–27 (2020).
21. Ficetola, G.F., Manenti, R. & Taberlet, P. Environmental DNA and metabarcoding for the study of amphibians and reptiles: species distribution, the microbiome, and much more. *Amphibia-Reptilia.* **40**, 129–148 (2019).
22. Sutter, M. & Kinziger, A.P. Rangewide tidewater goby occupancy survey using environmental DNA. *Conserv Genet.* **20**, 597–613 (2019).
23. Thomsen, P.F & Sigsgaard, E.E. Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecol Evol.* **9**, 1665–1679 (2019).
24. Carvalho, S. et al. Beyond the visual: using metabarcoding to characterize the hidden reef cryptobiome. *P Roy Soc B-Biol Sci.* **286**, 20182697 (2019).

25. Franklin, T.W. et al. Using environmental DNA methods to improve winter surveys for rare carnivores: DNA from snow and improved noninvasive techniques. *Biol Cons.* **229**, 50–58 (2019).
26. Shelton, A.O. et al. Environmental DNA provides quantitative estimates of a threatened salmon species. *Biol Conserv.* **237**, 383–391 (2019).
27. Needham, D.M. & Fuhrman, J.A. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat Microbiology*, **1(4)**, 16005 (2016).
28. Pierre, T., Aurélie, B., Lucie, Z. & Eric, C. EnvironMantel DNA For Biodiversity Research and Monitoring. OXFORD (2018).
29. Fujii, K. et al. Environmental DNA metabarcoding for fish community analysis in backwater lakes: A comparison of capture methods. *PLoS One* **14**, e0210357–e0210357 (2019).
30. Emily, E.C. et al. Anacapa Toolkit: An environMantel DNA toolkit for processing multilocus metabarcode datasets. *Methods Ecol Evol*, **10**, 1469-1475 (2019).
31. Nichols, R.V. et al. Minimizing polymerase biases in metabarcoding. *Mol Ecol Resour.* **18**, 927–939 (2018).
32. Bradley, I.M., Pinto, A.J. & Guest, J.S. Design and Evaluation of Illumina MiSeq-Compatible, 18S rDNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. *Appl Environ Microbiol.* **82(19)**, 5878 (2016).
33. Bucklin, A., Lindeque, P. K., Rodriguez-Ezpeleta, N., Albaina, A., & Lehtiniemi, M. Metabarcoding of marine zooplankton: Prospects, progress and pitfalls. *J Plankton Res.* **38**, 393–400 (2016).
34. Zhang, M. et al. Long-term dynamics and drivers of phytoplankton biomass in eutrophic Lake Tai. *Sci Total Environ.* **645**, 876–886 (2018).
35. Zhou, H.-W. et al. BIPES, a cost-effective high-throughput method for assessing microbial diversity. *ISME J.* **5**, 741–749 (2011).
36. Degnan, P.H. & Ochman, H. Illumina-based analysis of microbial community diversity. *ISME J.* **6**, 183–194 (2012).
37. Gloor, G.B. et al. Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS One* **5**, e15406 (2010).
38. Claesson, M.J. et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rDNA gene regions. *Nucleic Acids Res.* **38**, e200 (2010).
39. Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. & Schloss, P.D. 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 (2013).
40. Caporaso, J.G. et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
41. Fadrosch, D.W. et al. An improved dual-indexing approach for multiplexed 16S rDNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**, 6 (2014).

42. Bartram, A.K., Lynch, M.D.J., Stearns, J.C., Moreno-Hagelsieb, G. & Neufeld, J.D. Generation of multimillion-sequence 16S rDNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol.* **77**, 3846–3852 (2011).
43. Salipante, S.J. et al. Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rDNA-based bacterial community profiling. *Appl Environ Microbiol.* **80**, 7583–7591 (2014).
44. Illumina 16S metagenomic sequencing library preparation (Illumina Technical Note 15044223 Rev. A). Illumina http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf (2013).
45. Faith, J.J. et al. The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013).
46. Lundberg, D.S., Yourstone, S., Mieczkowski, P., Jones, C.D. & Dangl, J.L. Practical innovations for high-throughput amplicon sequencing. *Nat Method* **10**, 999–1002 (2013).
47. Henrik, K., Susan, R.K, Alexandra, Rueda., Athena, Lam. & Rosemary, G.G. Scaling up DNA barcoding – Primer sets for simple and cost efficient arthropod systematics by multiplex PCR and Illumina amplicon sequencing. *Methods Ecol Evol.* (2018).
48. Civade, R. et al. Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PLoS One* **11**, e0157366–e0157366 (2016).
49. Gillet, B. et al. Direct fishing and eDNA metabarcoding for biomonitoring during a 3-year survey significantly improves number of fish detected around a South East Asian reservoir. *PLoS One* **13**, e0208592–e0208592 (2018).
50. Rivera, S.F., Vasselon, V. & Bouchez, A. Diatom metabarcoding applied to large scale monitoring networks: Optimization of bioinformatics strategies using Mothur software. *Ecol Indic.* DOI: 10.1016/j.ecolind.2019.105775 (2019).
51. Drake, B. G. Rising sea level, temperature, and precipitation impact plant and ecosystem responses to elevated CO₂ on a Chesapeake Bay wetland: review of a 28-year study. *Global Change Biol.* **20**, 3329–3343 (2014).
52. Baldwin, A. H., Jensen, K. & Schoenfeldt, M. Warming increases plant biomass and reduces diversity across continents, latitudes, and species migration scenarios in experimental wetland communities. *Global Change Biol.* **20**, 835–850 (2014).
53. Scheffer, M., Hosper, S. H., Meijer, M. L., Moss, B. & Jeppesen, E. Alternative equilibria in shallow lakes. *Trends Ecol Evol.* **8**, 275–279 (1993).
54. Paerl, H.W. et al. Controlling harmful blooms of cyanobacteria in a hyper-eutrophic lake (Lake Tai, China): the need for a dual nutrient (N & P) management strategy. *Water Res.* **45**, 1973–1983 (2011).
55. Jef, Huisman. et al. Cyanobacterial blooms. *Nat Rev Microbiol.* **16**, 471-483 (2018).
56. Shi, K. et al. Long-term MODIS observations of cyanobacterial dynamics in Lake Tai: responses to nutrient enrichment and meteorological factors. *Sci Rep* **7**, 40326 (2017).

57. Berdjeb, L. et al. Short-term dynamics and interactions of marine protist communities during the spring–summer transition. *Isme J*, 2018.
58. Duan, H. et al. Two-decade reconstruction of algae blooms in China's Lake Tai. *Environ Sci Technol.* **43**, 3522–3528 (2009).
59. Duan, H. et al. Distribution and incidence of algae blooms in Lake Tai. *Aquat Sci.* **77**, 9–16 (2015).
60. Qin, B. et al. A drinking water crisis in Lake Tai, China: linkage to climatic variability and lake management. *Environ Manag.* **45**, 105–112 (2010).
61. Qin, B., Xu, P., Wu, Q., Luo, L. & Zhang, Y. Environmental issues of Lake Taihu, China. *Hydrobiologia* **581**, 3–14 (2007).
62. Deng, J. et al. Earlier and warmer springs increase cyanobacterial (*Microcystis* spp.) blooms in subtropical Lake Tai, China. *Freshw Biol.* **59**, 1076–1085 (2014).
63. Xu, H. et al. Determining critical nutrient thresholds needed to control harmful blooms of cyanobacteria in eutrophic Lake Tai, China. *Environ Sci Technol.* **49**, 1051–1059 (2015).
64. Li, D. et al. Factors associated with blooms of cyanobacteria in a large shallow lake, China. *Environ Sci Eur.* **30(1)**, 27 (2018).
65. Yang, J. et al. Indigenous species barcode database improves the identification of zooplankton. *PLoS One* **12(10)**, e0185697 (2017).
66. Chao, J.Y. et al. Long-term moderate wind induced sediment resuspension meeting phosphorus demand of phytoplankton in the large shallow eutrophic Lake Tai. *Plos One*, **12(3)**, e0173477 (2017).
67. Chen, Y. et al. Phytoplankton Community Structure and Its relationship with Environmental Factors in Different Regions of Taihu Lake. *J Hydroecology.* **38**, 39-44 (2017) (in Chinese).
68. Zhang, Y., Li, W. & Chen, Q. Spatial-temporal variance of the intensity of algae bloom and related environmental and ecological factors in Lake Tai. *Acta Ecologica Sinica.* **36**, 4338-4345 (2016).
69. Zou, S. et al. How DNA barcoding can be more effective in microalgae identification: a case of cryptic diversity revelation in *Scenedesmus* (Chlorophyceae). *Sci Rep*, **6**, 36822 (2016a).
70. Zou, S. et al. Combining and Comparing Coalescent, Distance and Character - Based Approaches for Barcoding Microalgae : A Test with *Chlorella*- Like Species (Chlorophyta). *PLoS One*, **11(4)**, e0153833 (2016b).
71. Knight, R. et al. Best practices for analysing microbiomes. *Nat Rev Microbiol.* **16**, 410–422 (2018).
72. Karst, S.M. et al. Retrieval of a million high-quality, full-length microbial 16S and 18S rDNA gene sequences without primer bias. *Nat Biotechnol.* **36(2)** (2018).
73. Brauer, V. S. et al. Low temperature delays timing and enhances the cost of nitrogen fixation in the unicellular cyanobacterium *Cyanothece*. *ISME J.* **7**, 2105–2115 (2013).
74. Breitbarth, E., Oschlies, A. & La Roche, J. Physiological constraints on the global distribution of *Trichodesmium*: effects of temperature on diazotrophy. *Biogeosciences* **4**, 53–61 (2007).
75. Paerl, H. W. & Huisman, J. Blooms like it hot. *Science* **320**, 57–58 (2008).
76. Jöhnk, K. D. et al. summer heatwaves promote blooms of harmful cyanobacteria. *Glob Chang Biol.* **14**, 495–512 (2008).

77. Kosten, S. et al. Warmer climates boost cyanobacterial dominance in shallow lakes. *Glob Chang Biol.* **18**, 118–126 (2012).
78. Su, C., Lei, L., Duan, Y., Zhang, K.Q. & Yang, J. Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Appl Microbiol Biotechnol.* **93**, 993–1003 (2012).
79. Zhang, Y., Li, W. & Chen, Q. Spatial-temporal variance of the intensity of algae bloom and related environmental and ecological factors in Lake Tai. *Acta Ecologica Sinica.* **36**, 4338-4345 (2016).
80. Paerl, H. W., Gardner, W. S., Mccarthy, M. J., Peierls, B. L. & Wilhelm, S. W. Algae blooms: noteworthy nitrogen. *Science* **346**, 175–175 (2014).
81. Hugerth, L.W. et al. DegePrime, a Program for Degenerate Primer Design for Broad-Taxonomic-Range PCR in Microbial Ecology Studies. *Appl Environ Microbiol.* **80**(16), 5116-23 (2014).
82. Ficetola, G.F. et al. An in silico approach for the evaluation of DNA barcodes. *BMC Genom.* **11**, 434 (2010).
83. Stoesser, G. et al. The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* **30**(1), 21-26 (2002).
84. Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinformatics.* **10**, 421 (2009).
85. Pruitt, K.D., Tatusova, T., & Maglott, D.R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
86. Baker, C. bakerccm/entrez_qiime: entrez_qiime v2.0. <https://doi.org/10.5281/zenodo.159607> (2016).
87. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* **37**, 852–857 (2019).
88. Oksanen, J. et al. vegan: Community Ecology Package R package version 2.5-5 (2019)
89. Deng, Y., Jiang, Y., Yang, Y. et al. Molecular ecological network analyses. *BMC Bioinformatics* **13**, 113 (2012). <https://doi.org/10.1186/1471-2105-13-113>.

Figures

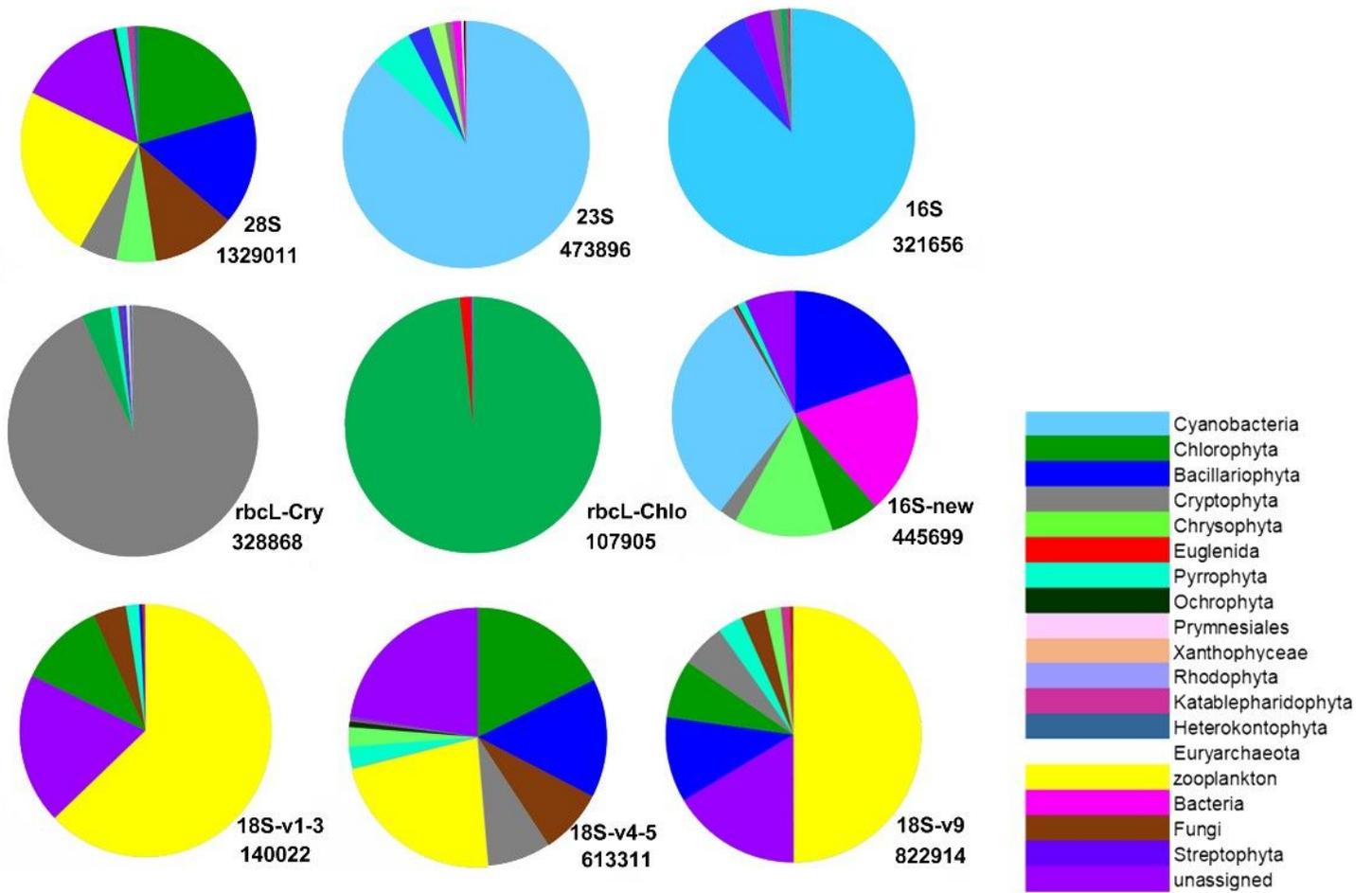
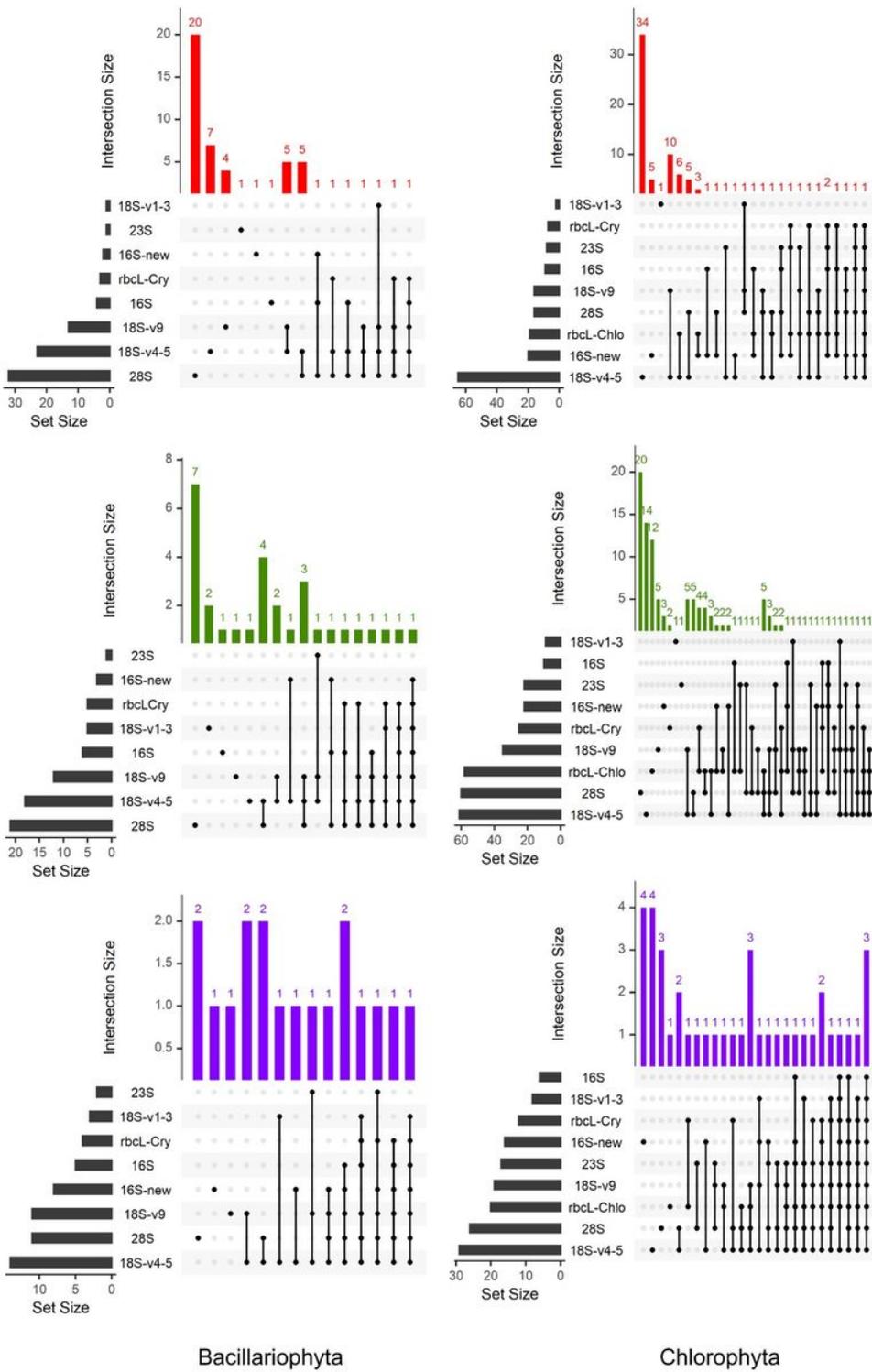


Figure 1

The proportion of taxa assigned from each gene marker.



Bacillariophyta

Chlorophyta

Figure 2

The difference of taxonomic assignments among all gene barcodes at each of species, genus and family level for Bacillariophyta and Chlorophyta. Red: species level; Cyan: genus level; Lilac color: family level.

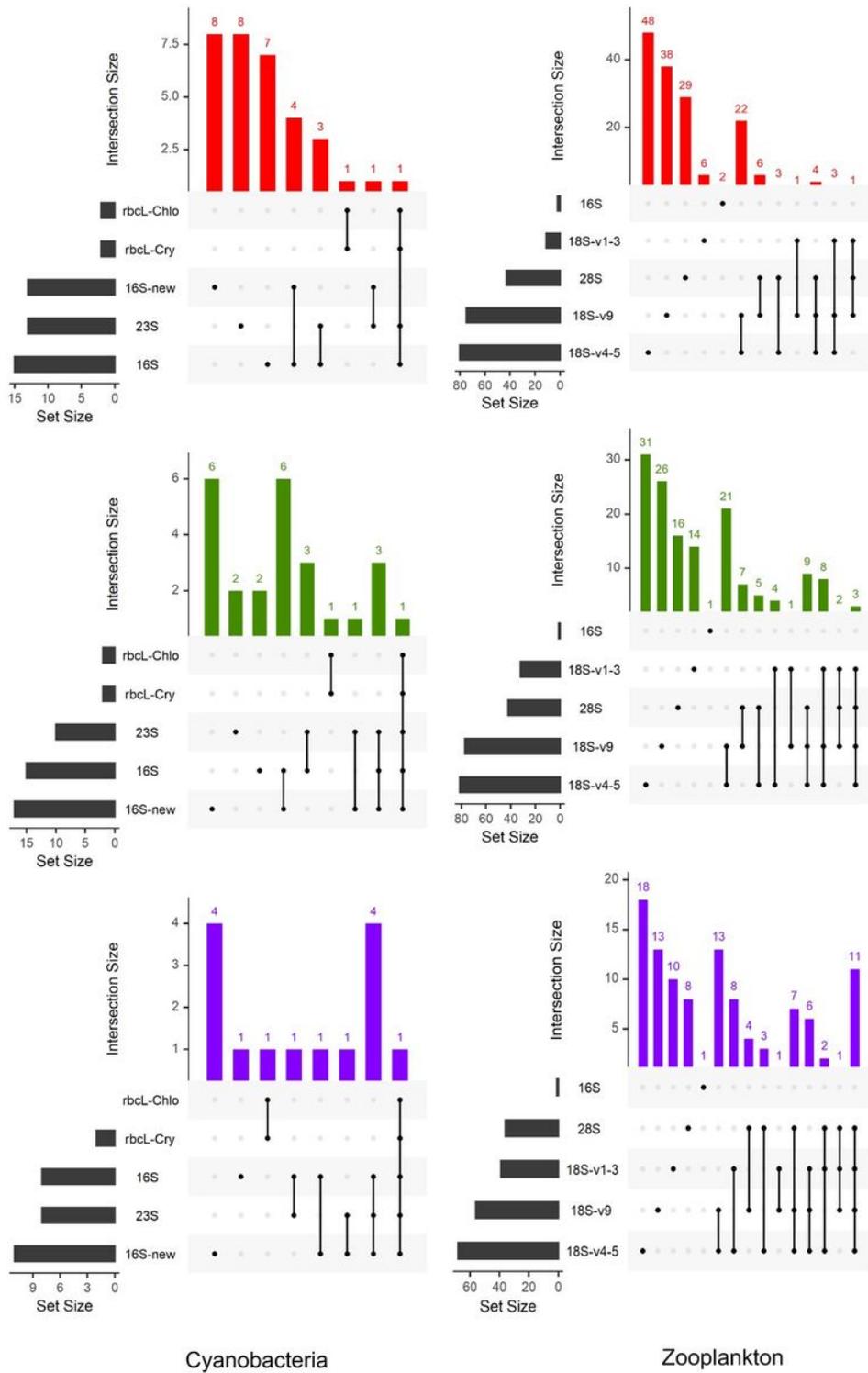


Figure 3

The difference of taxonomic assignments among all gene barcodes at species, genus and family level for Cyanobacteria and Zooplankton. Red: species level; Cyan: genus level; Lilac color: family level.

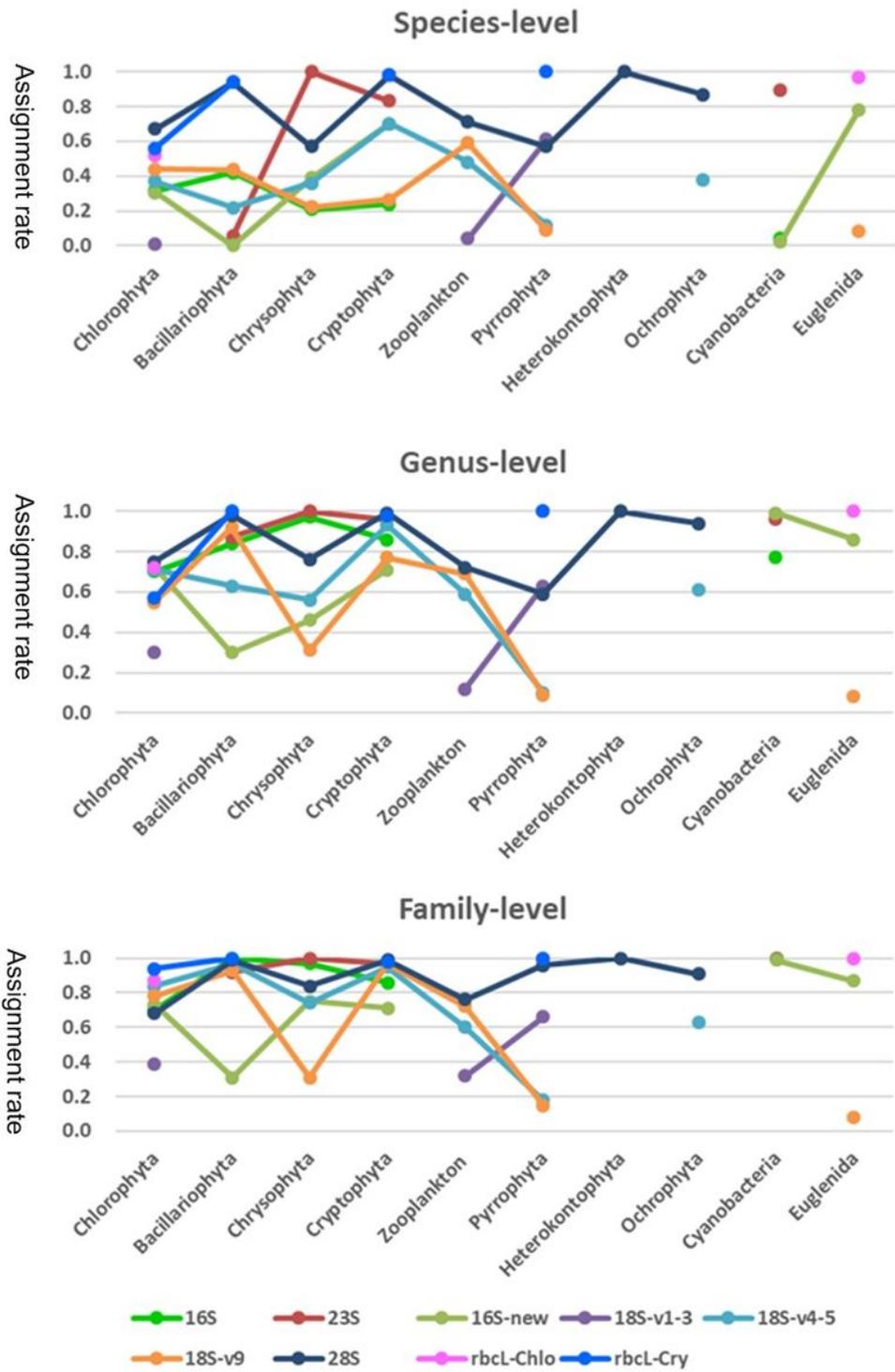


Figure 4

The assignment rate of each gene marker at each of species, genus and family level for each phytoplankton phylum, which means the rate of reads assigned at each taxonomic level to the total reads referring to final assignment after filter based on identify score.

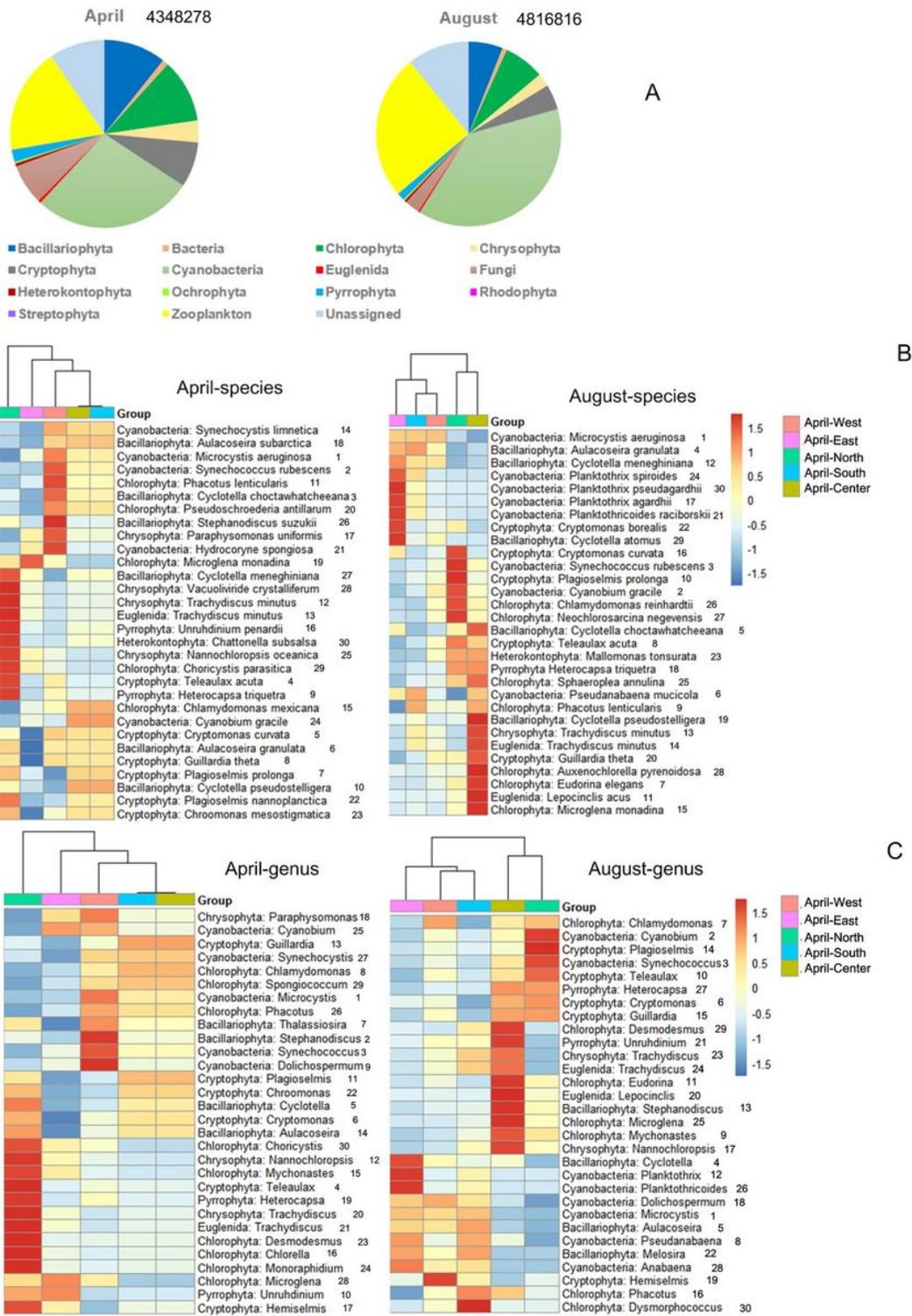


Figure 5

A: The proportion of various taxonomic groups for April (total reads number: 4348278) and August (total reads number: 5112288); B: Heatmap analysis for phytoplankton species: diversity of top 30 species for April and August. The phylum name (left) and order number for abundance (right) is besides species name. The color represents the degree of the abundance; C: Heatmap analysis for phytoplankton genus: diversity of top 30 genera for April and August. The phylum name (left) and order number for abundance (right) is besides genus name. The color represents the degree of the abundance.

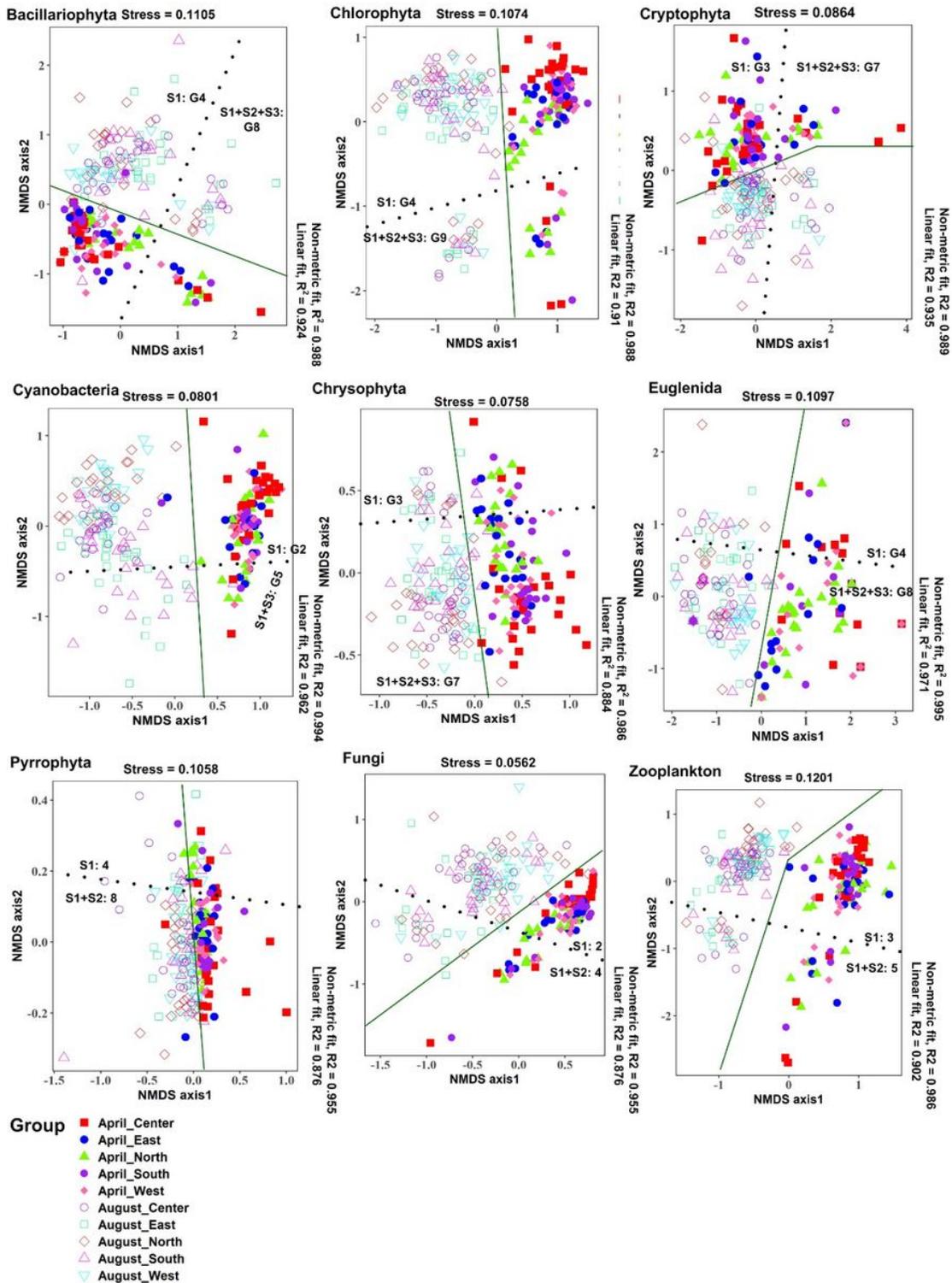


Figure 6

NMDS plots showing diversity dissimilarities among all samples of the whole lake in April and August. The NMDS analysis was performed for each phytoplankton phylum, zooplankton and fungi respectively. The R² values were shown for each plot. The solid line indicated the cut-off rule of the samples from two seasons. The dotted line indicated the cut-off rule of samples amplified by different gene loci (mock community and non-mock community). Besides the dotted line the annotation S_n represents PCR systems and the G_n represents the number of gene loci.

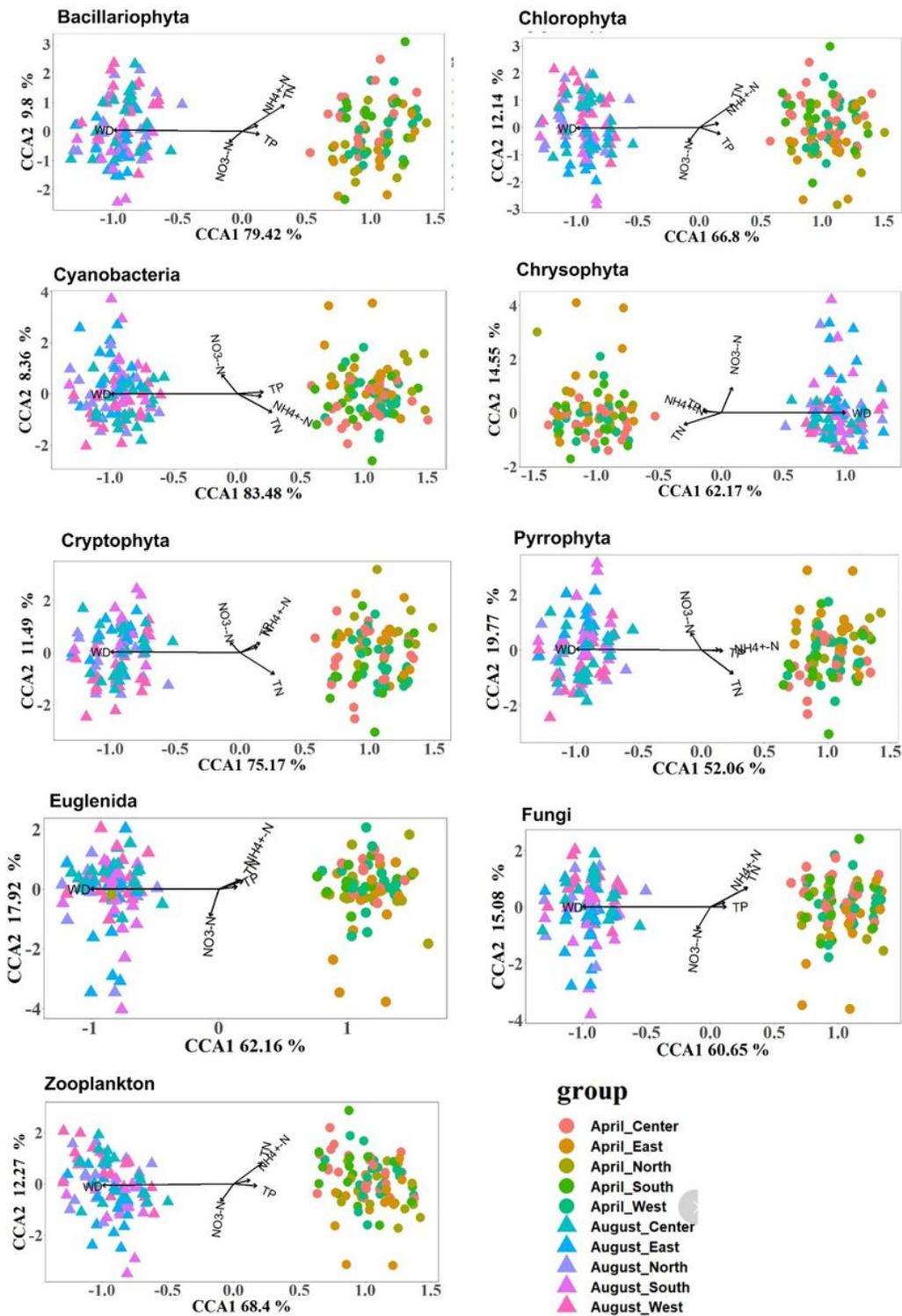


Figure 7

Canonical correspondence analysis (CCA) ordination diagram for correlations of diversity dynamics and environmental factors for each phytoplankton phylum, zooplankton, and fungi. Samples analyzed were amplified by gene barcodes from PCR system A (non-mock community). The length of the vector is proportional to its importance, and the angle between two vectors reflects the degree of correlation between variables.

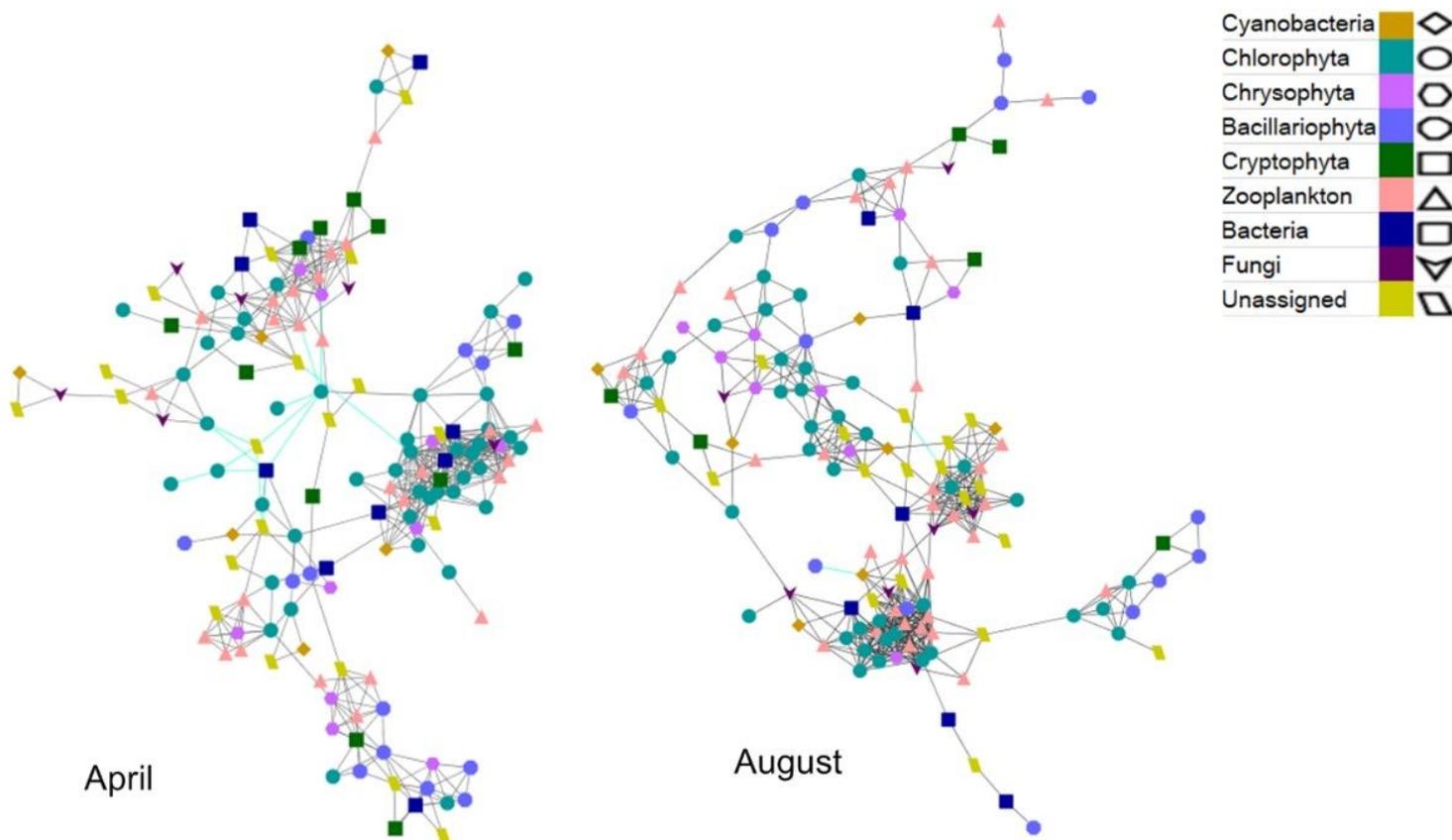


Figure 8

The molecular network among various taxonomic assignments for April and August respectively. The connection lines in gray and green indicated the positive and negative correlations respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [FigureS1.tif](#)
- [FigureS2.tif](#)
- [FigureS3.tif](#)
- [FigureS4.tif](#)
- [FigureS5.tif](#)
- [FigureS6.tif](#)
- [FigureS7.tif](#)
- [FigureS8.tif](#)