DR. YANG JUN (Orcid ID : 0000-0003-4914-0162)

Received Date : 26-Jun-2016

Revised Date : 29-Dec-2016

Accepted Date : 04-Jan-2017

Article type : Resource Article

DNA metabarcoding reveals that 200 µm size-fractionated filtering is unable to discriminate between planktonic microbial and large eukaryotes

LEMIAN LIU[1], MIN LIU[1,2], DAVID M. WILKINSON[3,4], HUIHUANG CHEN[1], XIAOQING YU[1] AND JUN YANG[1*]

[1]*Aquatic EcoHealth Group, Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, People's Republic of China*

[2]*University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China*

[3]*School of Natural Science and Psychology, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK*

[4] *School of Life Sciences, University of Lincoln, Lincoln LN6 7TS, UK.*

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1755-0998.12652

Running head: Eukaryotic plankton community in reservoirs

Key words: size-fractionated filtering / eukaryotic planktonic community / high-throughput sequencing / subtropical reservoir / eDNA

Correspondence: Jun Yang, Aquatic Ecohealth Group, Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, People's Republic of China. E-mail: jyang@iue.ac.cn; Tel. (+86)5926190775; Fax (+86)5926190775.

Abstract

Microeukaryotic plankton (0.2–200 μm) are critical components of aquatic ecosystems and key players in global ecological processes. High-throughput sequencing is currently revolutionizing their study on an unprecedented scale. However, it is currently unclear if we can accurately, effectively, and quantitatively depict the microeukaryotic plankton communities using traditional size-fractionated filtering combined with molecular methods. To address this, we analyzed the eukaryotic plankton communities both with, and without, pre-filtering with a 200 μm pore-size sieve - by using SSU rDNA-based high-throughput sequencing on 16 samples with 3 replicates in each sample from two subtropical reservoirs sampled from January to October in 2013. We found that ~25% reads were classified as metazoan in both size groups. The species richness, alpha and beta diversity of plankton community, and relative abundance of reads in 99.2% eukaryotic OTUs showed no significant changes after pre-filtering with a 200 μm pore-size sieve. We further found that both > 0.2 μm and 0.2–200 μm eukaryotic plankton communities, especially the abundant plankton subcommunities, exhibited very similar, and synchronous, spatiotemporal patterns and processes associated with almost identical environmental drivers. The lack of an effect on community structure from pre-filtering suggests that environmental DNA from larger metazoa is introduced into the smaller size class. Therefore, size-fractionated filtering with 200 μm is insufficient to discriminate between the eukaryotic plankton size groups in

metabarcoding approaches. Our results also highlight the importance of sequencing depth, and strict quality filtering of reads, when designing studies to characterize microeukaryotic plankton communities.

Introduction

Microeukaryotic plankton (0.2–200 μm) are extremely diverse and play essential roles in the structure and function of global aquatic ecosystems (Shen *et al*. 1990; Finlay & Fenchel 2004; Adl *et al*. 2012; Cardinale *et al*. 2012; de Vargas *et al*. 2015; Pernice *et al*. 2016). This makes understanding this diversity a major ecological goal (Green & Bohannan 2006). Microbial eukaryotic diversity has been studied through light microscopy ever since the work of the early microscopists of the European Enlightenment, such as Antoni van Leeuwenhoek and Robert Hooke (Gest 2004). This history of over 350 years of microscopy had produced extensive data sets based on morphological features observable in light microscopy (Fenchel 1987; Shen *et al*. 1990; Reynolds 2006). In recent decades molecular approaches have become more important, and the SSU rRNA gene has been widely used in microbial ecology on an unprecedented scale (Bik *et al*. 2012; Thomsen & Willerslev 2015). The analysis of phylogenetic diversity using molecular methods can provide an important supplement to microscope-based approaches for accurately identifying planktonic microbial eukaryotes, in part because some extremely small eukaryotes (e.g. pico- or nano- eukaryotes) are difficult to observe or distinguish by light microscopy, and in addition, morphologically similar individuals may belong to different cryptic species or morphologically distinct types may be from the same species (Medinger *et al*. 2010; Boenigk *et al*. 2012; Gomaa *et al*. 2015; Santoferrara *et al*. 2015). Recently, several SSU rRNA gene-based studies have successfully described and characterized microeukaryotic plankton in aquatic systems (most of them from marine ecosystems, e.g. Edgcomb *et al*. 2011; Pawlowski *et al*. 2011; de Vargas *et al*. 2015; Massana *et al*. 2015; Yu *et al*. 2015; Pernice *et al*. 2016). To date, however, our limited knowledge on microeukaryotic phylogenetic diversity has hindered the development of

theories addressing the stability of structure and function of planktonic communities across space and time (Stoeck *et al*. 2014; Konopka *et al*. 2015).

Unlike molecular surveys of aquatic prokaryotic communities which are easily separated from large eukaryotic plankton by using the 16S SSU rRNA gene (Zinger *et al*. 2012), it is difficult to separate the microeukaryotic plankton community from the large eukaryotes by directly using the 18S SSU rRNA gene (Bik *et al*. 2012; Massana & Logares 2013). To overcome this problem, previous molecular studies of microeukaryotic diversity have often used size-fractionated samples to separate microeukaryotic plankton from large organisms, an approach which has long been used in traditional microscopy based plankton studies (Hardy 1956). Many of these molecular studies have used 200 μm pore-size sieves to remove large organisms (Countway *et al*. 2007; Amaral-Zettler *et al*. 2009; Cheung *et al*. 2010; Schnetzer *et al*. 2011; Kim *et al*. 2012; Jones *et al*. 2013; Santoferrara *et al*. 2014; Wang *et al*. 2015; Abad *et al*. 2016). Although the traditional size-fractionated filtering approach has been very successful, this critical step may have some limitations. First, environmental DNA (eDNA) from large metazoan such as the cell or body breakage, excreta or epidermal cells of arthropoda, annelida, mollusca, nematoda, fish, mammals and insects, etc. (Bohmann *et al*. 2014) may remain in the smaller size fraction and contaminate the microbial eukaryotic communities (Thomsen & Willerslev 2015). Second, the irregular body forms, colony forms, complex cell cycles and ontogenic processes (e.g. egg, sporocyst and larvae) could also allow large eukaryotes to pass through the sieve (Logares *et al*. 2014). Therefore, if the limitations mentioned above commonly exist, we may not be effectively exploring the diversity of microeukaryotic plankton communities by such a sieving approach. To address this question, we determine whether the pre-filtering by a 200 μm pore-size sieve can be an accurate and reliable approach to collecting data on the microeukaryotic plankton community.

In this study, the eukaryotic plankton communities, with and without pre-filtering by a 200 μm pore-size sieve, were analyzed simultaneously from two subtropical reservoirs in Southeast China across four different seasons during 2013. Reservoirs have come to be especially important water sources in China, indeed Yang and Lu (2014) estimated that the total storage capacity of Chinese reservoirs is triple that of its lakes. The water quality of

these reservoirs is obviously important for their role in supplying drinking water and closely linked to the planktonic microeukaryotic communities (Yang *et al*. 2016). Over the last decade, with the rapid development of deep DNA sequencing, an increasing number of studies have started to investigate the tremendous diversity of microeukaryotic communities in a wide range of aquatic environments on an unprecedented scale (Bik *et al*. 2012; Yu *et al*. 2014; Yu *et al*. 2015; Thomsen & Willerslev 2015). However, as described above, we still lack sufficient understanding about whether traditional size-fraction filtering (e.g. pre-filtering by a 200 μm pore-size sieve) can accurately and effectively discriminate the microeukaryotic plankton communities from large eukaryotic communities in DNA-based studies. In this study, we attempt to give a comprehensive answer to this question.

Materials and methods

Study area and sampling

Two reservoirs, Shidou Reservoir (SD) and Hubian Reservoir (HB), which are located in Xiamen, southeast China, were sampled in this study; detailed descriptions of these reservoirs were provided in a previous study (Yang *et al*. 2012). Briefly, SD is a large (3.9 km^2) reservoir within a wooded catchment, while HB is a smaller (1.0 km^2) reservoir within an urban catchment. For each reservoir, three sampling stations representing three replicates were selected in the riverine zone, transitional zone, and lacustrine zone, respectively. Surface water (upper 50 cm) was collected at each station of two reservoirs in January, April/May, July and October of 2013 (See Supplementary Table S1 for details of sample information). The water from each of the 48 replicates was divided into two sub-groups after thorough mixing: one for water chemistry analysis and the other for eukaryotic plankton community analyses. All samples were transported on ice to the laboratory, where they were used for the following treatments within 4 h of sampling.

Physico-chemical analysis

Water temperature, electrical conductivity, pH, dissolved oxygen, turbidity, chlorophyll *a* (Chl *a*) of the epilimnion or surface water were measured *in situ* with a Hydrolab DS5 multi-parameter water quality analyzer (Hach, Loveland, CO, USA). The water depth at the sampling sites was determined with a Speedtech SM-5 Depthmate portable sounder (Speedtech Instruments, Great Falls, VA, USA), and water transparency was measured with a Secchi disk at the same time. Total carbon (TC), total organic carbon (TOC), total nitrogen (TN), ammonium nitrogen (NH_4-N), nitrite and nitrate nitrogen (NO_x-N), total phosphorus (TP) and phosphate phosphorus (PO_4-P) were analyzed according to standard methods (Greenberg *et al*. 1992).

DNA extraction, PCR and Illumina sequencing

For the planktonic eukaryotes in the 0.2–200 μm size fraction, ~500 mL water was pre-filtered through a 200 μm pore-size sieve and then sequentially filtered through a 0.2 μm pore-size polycarbonate membrane (47 mm diameter, Millipore, Billerica, MA, USA). For > 0.2 μm eukaryotes, ~500 mL water was filtered directly through a 0.2 μm pore-size polycarbonate filter. The 0.2–200 μm and > 0.2 μm eukaryotic communities were filtered at same time in a single sampling campaign. To reduce the likelihood of contamination, the filter bowls were rinsed successively by sterile water and sample water before each sample filtering. In total, eight 0.2–200 μm and eight > 0.2 μm eukaryotic community samples with 3 replicates in each sample were obtained from the two reservoirs. Total DNA of eukaryotic plankton community was extracted directly from the membrane using a FastDNA spin kit (Bio101, Carlsbad, CA, USA) according to the manufacturer's instructions. Before DNA extraction, the membrane that collected the 0.2-200 μm eukaryotes was cut into pieces using flame disinfected scissors, and then we scissored the membranes with > 0.2 μm eukaryotes. A set of primers with the barcode were used to amplify the hypervariable V9 region of the eukaryotic 18S rRNA gene. In this study, the forward primer was 1380F (5'-CCCTGCCHTTTGTACACAC-3'), and the reverse primer was 1510R (5'-CCTTCYGCAGGTTCACCTAC-3') (Amaral-Zettler *et al*. 2009). Each DNA sample

was individually PCR-amplified in triplicated 30 μL reactions. The reactions included an initial denaturation at 98 °C for 1 min, followed by 30 cycles of 10 s at 98 °C, 30 s at 50°C and 60 s at 72 °C. At the end of the amplification, the amplicons were subjected to final 10 min extension at 72 °C. Each reaction contained 15 μL of Phusion High-Fidelity PCR Master Mix (New England Biolabs, Beverly, MA, USA), 0.2 μM of each primer, and 10 ng of target DNA. The length of PCR products (~150 bp) were confirmed using a 1% agarose gel electrophoresis. PCR products from the triplicates were mixed in equimolar amounts. Then, mixture PCR products were purified with GeneJET Gel Extraction Kit (Thermo Scientific, Hudson, NH, USA). Sequencing libraries were generated using NEB Next Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Beverly, MA, USA) following manufacturer's recommendations and index codes were added. The library quality was assessed on the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and Agilent Bioanalyzer 2100 system (Agilent Technologies, Palo Alto, CA, USA). Finally, the library was sequenced on a single lane of Illumina MiSeq platform (Illumina, Inc., San Diego, CA, USA) using a paired-end (2 × 250 bp) sequencing strategy.

Bioinformatics

In total, 3894571 raw paired-end reads were obtained from 48 replicates, these raw paired-end reads ranged from 45027 (minimum) to 154083 (maximum) with a mean of 81137 per replicate. Pairs of reads were merged by using FLASH and the mean of contig read length was 130 ± 0.006 s.e. (Magoč & Salzberg 2011). Merged reads were then quality controlled by using the QIIME software package (Caporaso *et al*. 2010) with the following settings: maximum number of consecutive low-quality base = 3; minimum of continuous high-quality base = 75% of total read length; maximum number of ambiguous bases = 0, last quality score = 3. Chimeras were identified using UCHIME and discarded prior to further analysis (Edgar *et al*. 2011). UPARSE pipeline were used to picking operational taxonomic units (OTUs) by making an OTU table (Edgar 2013). Quality filtered reads were assigned to OTUs at 97% and 99% sequence similarity thresholds, respectively. Representative sequence from each OTU was blasted against the SILVA database (Release 115) (Quast *et al*. 2013). Unassigned

(sequence similarity to a reference sequence is < 80%), mammal (10 OTUs, 74 reads at 97% threshold), fish (2 OTUs, 12 reads at 97% threshold), higher plant (28 OTUs, 272 reads at 97% threshold), and singleton OTUs were removed before the downstream analyses. For our data analyses, we used a randomly selected subset of 36000 reads at 97% threshold and 30000 reads at 99% threshold from each replicate to standardize sequencing effort across all 48 replicates. The final total data set retained 1728000 and 1440000 reads at 97% and 99% sequence similarity levels, respectively.

Statistics

The Bray-Curtis similarity matrix is considered to be one of the most robust similarity coefficients for use in ecological studies (Kent 2012) and was applied to our community dataset of microbial eukaryotic read relative abundance at OTU level. A non-metric multidimensional scaling (NMDS) ordination was used to investigate differences in microbial eukaryotic communities between sites (Clarke & Gorley 2015). We used RELATE to calculate the Spearman's rank correlations between the similarity matrices of > 0.2 μm and 0.2–200 μm eukaryotic plankton communities. RELATE works by calculating rank correlation between two matrices, thus provides a significance test with the matching coefficient ρ, which is equivalent to Mantel's test (Clarke & Gorley 2015). All these analyses above were calculated in PRIMER 6.0.

We used a forward selection procedure to select the physico-chemical variables which were significantly correlated with spatiotemporal dynamics of the eukaryotic communities using the 'ordiR2step' function from vegan of R software (Blanchet *et al*. 2008). Prior to the analysis, the physico-chemical variables were log(x+1) transformed, with the exception of pH, to improve normality and homoscedasticity.

Rarefaction curves, Chao 1, abundance-based coverage estimator (ACE), Shannon-Wiener index, Simpson index and Pielou index were calculated in vegan with R software (R Core Team, 2015). Good's coverage was performed in Mothur software (Schloss *et al*. 2009). The nonparametric Mann-Whitney test was used to compare the difference in eukaryotic diversity and relative abundance of reads between the > 0.2 μm and 0.2–200 μm communities.

Biological communities are normally composed of a few abundant and many rare species, and such pattern is particularly prominent in natural microbial communities as most taxa are usually extremely rare (Logares *et al*. 2014). Abundant and rare microbial subcommunities may have fundamentally different characteristics and ecological roles (Liu *et al*. 2015). The abundant and rare taxa were defined in accordance with previous studies (Logares *et al*. 2014; Liu *et al*. 2015). Briefly, the OTUs that had a relative abundance of > 1% within a replicate and had a mean relative abundance of > 0.1% within all replicates were defined as abundant OTUs. The OTUs that had a relative abundance of < 0.01% within a replicate and had a mean relative abundance of < 0.001% within all replicates were defined as rare OTUs.

Results

Alpha-diversity

The wide range of eukaryotic plankton taxa were recovered from the studied reservoirs by this sampling. A total of 10965 OTUs were detected from 1728000 high-quality reads for all samples (36000 reads per replicate sample) at 97% sequence identity level. In all, 9520 and 9594 OTUs were identified from > 0.2 μm and 0.2–200 μm size-fractionated plankton communities, respectively. These data were the result of examining the reads of more than 1.7 million V9 rDNA reads from the pooled data set after sequence cleanup and quality control (Supplementary Fig. S1). Individual-based rarefaction curves, nonparametric asymptotic estimators (Chao 1 and ACE) and Good's coverage confirmed that the relatively abundant fraction of plankton communities was well sampled, thus allowing extraction of general patterns of plankton communities from our data set (Supplementary Table S2).

The alpha-diversity (including OTU number, Chao 1, ACE, Shannon-Wiener diversity index, Simpson diversity index, and Pielou evenness index) of all, abundant, and rare eukaryotic communities did not show significant differences between the four seasons of two reservoirs after pre-filtering of 200 μm pore-size sieve (nonparametric Mann-Whitney test, Table 1, Supplementary Tables S3, S4 and S5).

Shared and unique OTUs

Our results demonstrated a higher level stability in the abundant OTUs than in the low-abundance OTUs under different read filtering conditions. About 85% of OTUs (8149/9520 for > 0.2 μm size fraction, and 8149/9594 for 0.2–200 μm size fraction) were shared between the > 0.2 μm and 0.2–200 μm size fractions when singleton OTUs were removed, yet these shared OTUs contributed disproportionately to about 99.7% of reads in each size fraction. However, about 99% and 100% OTUs were shared between the > 0.2 μm and 0.2–200 μm size fractions when all OTUs with < 10 reads and OTUs with < 50 reads were removed, respectively (Table 2). Similarly, although only 70.5–76.6% of OTUs were shared between the two size fractions in different seasons for each reservoir when singleton OTUs were removed, the proportion of shared OTUs increased to 82.3–86.2% and 89.0–94.0% when all OTUs with < 10 reads and OTUs with < 50 reads were removed, respectively (Supplementary Table S6). Further, we randomly selected subsets of 100, 500, 1000, 5000, 10000, 20000, 30000 and 36000 cleaned reads from each replicate to test the effect of sequence depth on the shared OTUs. We found that the greater the number of sequence reads obtained, the higher the proportion of shared OTUs (Table 3), demonstrating that increasing sequencing depth per sample can lead to a more prominent increase in the number of shared OTUs than the number of unique OTUs.

Beta-diversity

We found the spatiotemporal patterns of > 0.2 μm eukaryotic plankton metacommunities in the reservoirs exhibited a remarkable similarity with the patterns for the 0.2–200 μm metacommunities ($\rho = 0.982$, $P = 0.001$, Fig. 1). Likewise, striking similarities were also found between the abundant > 0.2 μm and abundant 0.2–200 μm subcommunities ($\rho = 0.963$, $P = 0.001$). However, the patterns of rare > 0.2 μm subcommunities showed some limited differences to the rare 0.2–200 μm subcommunities ($\rho = 0.894$, $P = 0.001$) (Supplementary Fig. S2).

Relative abundance of reads at phylum or OTU levels

The > 0.2 μm plankton showed a similar relative abundance of reads with the 0.2–200 μm eukaryotic communities at the phylum or higher taxonomic rank in both reservoirs (Mann Whitney test, $P > 0.05$, Fig. 2). In total, ~22% reads were affiliated with the arthropoda and ~18% reads were affiliated with unclassified eukaryotes in both size fractions. Furthermore, 99.2% (10876/10965), 99.92% (7154/7160) and 99.95% (8642/8646) OTUs had relative abundances that were not significantly different between the > 0.2 μm and 0.2–200 μm communities in both reservoirs, in Shidou Reservoir and in Hubian Reservoir, respectively (Fig. 3). All abundant OTUs (134/134) and 99.0% rare OTUs (7498/7575) had relative abundances that were not significantly different after pre-filtering by a 200 μm pore-size sieve (Mann Whitney test, $P > 0.05$).

Influence of environmental factors on community dynamics

The environmental factors that significantly impacted the spatiotemporal variation of > 0.2 μm and 0.2–200 μm eukaryotic plankton communities were almost identical (Table 4). Water temperature, dissolved oxygen, NO_x-N, total phosphorus, PO_4-P and chlorophyll *a* were closely and significantly related to the spatiotemporal distributions of both fractionated plankton metacommunities. Further, these environmental variables explained 54.6% and 50.6% community variation of > 0.2 μm and 0.2–200 μm eukaryotic plankton, respectively. This result indicates a similar interplay between > 0.2 μm or 0.2–200 μm communities and environment, although 45.4–49.4% of community variation remained unexplained

Community analysis and comparison using 97% and 99% sequence similarity thresholds

The 99% similarity threshold produced 17657 OTUs with 1440000 reads (30000 reads per replicate sample). 15239 and 15362 OTUs were identified at a 99% similarity threshold from > 0.2 μm and 0.2–200 μm size-fractionated plankton communities, respectively. About 84% of OTUs (12944) were shared between the > 0.2 μm and 0.2–200 μm size fractions when singleton OTUs were removed, yet these shared OTUs contributed disproportionately to about 99% of reads in each size fraction. Further, about 98%, and 99.9% OTUs were shared

between the > 0.2 μm and 0.2–200 μm size fractions when all OTUs with < 10 reads and OTUs with < 50 reads were removed, respectively (Supplementary Table S7).

More importantly, despite differences in the absolute numbers of OTUs obtained from the 97% and 99% sequence similarity thresholds, the diversity patterns and relative abundance of reads at the phylum or OTU levels obtained from the 99% threshold were almost identical with the results obtained from the 97% threshold (Supplementary Figs. S3, S4 and S5; Table S8). As the choice of threshold had no apparent effect on our overall results and conclusions, we focused our following discussion about the results obtained from the 97% similarity threshold.

Discussion

Our results showed that the observed species richness (number of OTUs), the estimated species richness (e.g. Chao 1 and ACE), the species diversity (e.g. Shannon-Wiener, Simpson and Pielou indices) and the relative abundance of reads at the phylum level for the > 0.2 μm eukaryotic plankton communities were very similar to the 0.2–200 μm fraction communities in all the four seasons for both reservoirs. The spatiotemporal patterns (beta-diversity) of > 0.2 μm plankton communities, especially the abundant plankton subcommunities, were also strikingly similar to the 0.2–200 μm communities. In addition, our results suggest that the environmental factors that impact on > 0.2 μm and 0.2–200 μm eukaryotic plankton communities are effectively identical – although this may, in part, be because the sieving fails to reliably separate these two fractions. Nutrients, dissolved oxygen, water temperature and chlorophyll *a* were significantly related to the spatiotemporal variation of both > 0.2 μm and 0.2–200 μm eukaryotic plankton communities. These relationships between eukaryotic plankton communities and environmental factors have also been reported in many previous studies of aquatic eukaryotic plankton communities (Lepère *et al*. 2006; Chen *et al*. 2010; Jones *et al*. 2013; Liu *et al*. 2013). However, it should be noted that about 50% of the variation in the eukaryotic community composition cannot be explained at present. This unexplained variation could be related to other variables not measured here, such as cellular lysis by virus, fish grazing and other abiotic factors (Christoffersen *et al*. 1993; Brussaard

2004). An alternative explanation is that random processes play a much larger role in the ecology and evolution of microorganism than they do for larger better studied taxa (Bonner 2013).

We found some extremely rare OTUs showed difference between the two size fractions. For example, about 15% OTUs were unique between the > 0.2 μm and 0.2–200 μm plankton communities, but these unique OTUs only represented about 0.3% of the total reads in both > 0.2 μm and 0.2–200 μm plankton communities (Table 2). Similarly, the mean relative abundance of reads in 0.8% OTUs (89/10965) had a significant difference between the > 0.2 μm and 0.2–200 μm plankton communities, but all these OTUs only represented less than 0.05% of the total reads (Fig. 3). The likely reason for the differences in OTU composition and abundance between the two size-fractions is that our sequencing depth made detection by sequencing of rare eukaryotes far less predictable (Kuczynski *et al*. 2010). If we removed the low abundance OTUs with < 50 reads, all the OTUs were fully shared between the > 0.2 μm and 0.2–200 μm plankton communities (Table 2). Further, when we investigated the effects of sequencing depth on the results of shared OTUs, again we found that the more sequence obtained, the higher the proportion of shared OTUs were (Table 3). It is very difficult to detect all eukaryotic species based on the current 2nd generation sequencing method. Recently, de Vargas *et al*. (2015) showed that sequencing of about 1.7 million V9 rDNA reads from each of the 334 plankton samples was sufficient to approach saturation of eukaryotic richness, but there were still 1.2 novel metabarcodes in every new one hundred thousand rDNA reads sequenced (1.2/100000) at the global scale. Therefore, there were always some extremely rare OTUs which cannot be detected and this resulted in the differences in low-abundance OTU composition and relative abundance between the > 0.2 μm and 0.2–200 μm plankton communities based on the current 2nd generation sequencing method. This near impossibility of sampling the entire rare biosphere in a lake is an insight that predates modern molecular methods (Hutchinson 1964), and indeed was an even larger problem for microscopy based approaches. However, our results can reflect the dynamics of natural communities which were pre-filtered by a 200 μm pore-size sieve. At least, 99.2% OTUs which account for 99.95% abundance of reads in communities changed

non-significantly after pre-filtering by a 200 pore-size sieve (Fig. 3). Clearly, our results suggest that the pre-filtering with a 200 μm pore-size sieve has no significant impact on both alpha and beta diversities of the eukaryotic plankton in reservoirs, as well as community structure (including OTU number and OTU relative abundance).

Given that about 25% reads and 11% OTUs were classified as metazoa in both size fraction groups (most of these were affiliated with arthropoda), it is interesting that the eukaryotic diversity was not affected by a 200 μm pre-filtering. One obvious explanation for this is that our high-throughput sequencing (HTS) is mainly detecting larger metazoa (arthropods) as eDNA rather than as DNA in intact organisms (Thomsen & Willerslev 2015). Recently there has been a growing interest in using eDNA from water samples as a method of surveying for rare organisms - including vertebrates such as fish or amphibians (Ficetola *et al*. 2008; Thomsen *et al*. 2012; Rees *et al*. 2014). Although we only sampled small volumes of water, our sequencing identified a Chinese longsnout catfish (*Leiocassis longirostris*) from both sized fractions. If, as we suspect, eDNA is an important part of the explanation it also raises questions about just how much of the microbial DNA detected was eDNA rather than DNA in intact organisms. As eDNA ultimately comes from live organisms there will be correlations between actual (or recent) microbial population sizes and the amount of eDNA (Dejean *et al*. 2011; Pawlowski *et al*. 2011; Thomsen *et al*. 2012). However, in ecological studies we will often want to quantify the *living* organisms and eDNA may be a complication, or contamination, in doing this via DNA sequence-based studies (Shokralla *et al*. 2012). Also, species-specific amplification behaviors and water parameters strongly impact PCR efficiencies and therefore prohibit quantitative measurement by eDNA (Engelbrektson *et al*. 2010). In addition, alternative explanations for the lack of significant impact of pre-filtering are that the eukaryotic irregular body forms, complex cell shapes, colony forms, complex cell cycles, and ontogenic processes (egg, sporocyst and larvae) allow them pass through the 200 μm mesh sieve (Logares *et al*. 2014). Several previous studies also support this point of view. For example, de Vargas *et al*. (2015) investigated global marine eukaryotic plankton diversity and showed that a large number of metazoa were detected in the 180–2000 μm, as well as in the 20–180 μm and 5–20 μm eukaryotic subcommunities. Another marine study by Massana

et al. (2015) also found metazoa contributed 50% of sequences in their filtering micro/meso (20–2000 μm) eukaryotic samples and even contributed 1–10% of sequences in the pico- (0.8–3 μm) and nano-eukaryotic (3–20 μm) samples. Clearly, additional work is needed to estimate the filtering effect in other aquatic ecosystems.

There are some potential limitations that merit further discussion. We consider the water sample volume may affect the representation of the large eukaryotic plankton (e.g. mesozooplankton) community. We just collected ~500 mL water on a 0.2 μm pore-size polycarbonate membrane for the eukaryotic plankton analysis, because the density of microorganisms and other particles is much higher in the reservoirs than in the pelagic ocean (Pesant *et al*. 2015). The average transparency during our sampling period was less than 2 m, and it took nearly one hour to filter 500 mL water on a membrane. Hence, we think most of the eukaryotic plankton has been collected in our samples. Although the V9 region of 18S rRNA gene has been widely used in high-throughput sequencing analyses of eukaryotic diversity (de Vargas *et al*. 2015), the taxonomic resolution of this V9 region marker is relatively lower than the entire 18S hypervariable regions (Harder *et al*. 2016). For example, we found about 18% of reads were affiliated with unclassified eukaryotes at 97% sequence similarity threshold. In this study, however, we did not focus on the exact taxonomy. The evaluation of reduced taxonomic resolution for the V9 region seems to go beyond the scope of this study. Nevertheless, the 18S V1-V2 and V4 regions and COI gene should be combined with 18S V9 region to obtain a high taxonomic resolution in future studies (Abad *et al*. 2016). Given the short length (100–150 bp) and relatively reduced variability of the V9 region (Harder *et al*. 2016), we used a 99% OTU similarity threshold to avoid clustering together congeneric species. We did find the trend of results (diversity) obtained from both 97% and 99% similarity thresholds was almost identical (Supplementary Figs S3, S4 and S5; Tables S7 and S8).

In general, our study gives a novel insight on the traditional size fraction-DNA sequencing method for eukaryotic plankton communities. It is difficult to directly compare our study with previous HTS studies, because most of previous studies focused on totally separated size-fractions (e.g. 0.2–20 μm vs. 20–200 μm) (e.g. Logares *et al*. 2014; de Vargas *et al*.

2015; Massana *et al*. 2015; Abad *et al*. 2016). However the two size-fractions in our studies are overlapping (i.e. > 0.2 μm vs. 0.2–200 μm). Therefore, our design provides a better way to analyze the effect of size-fraction filtering on plankton community. However, given that these previous studies also found lots of metazoa in small fractions, it can be speculated that some large protists or fungi will remain in small fractions. Our work thus provides strong evidence that the traditional size-fractionated filtering fails when used in SSU rDNA-based high-throughput sequencing studies, probably because of eDNA contamination, irregular body forms of large eukaryotes, etc. in natural water samples. Directly targeting SSU rRNA instead of its genes may present an alternative approach. This method can indicate ribosomes and thus can reflect living biomass and can separate DNA within viable organisms from eDNA (Geisen *et al*. 2015). In addition, there is also clearly still an important role for microscopy to visually confirm the presence of organisms in water samples in addition to molecular approaches (Albaina *et al*. 2016).

Conclusion

We compared the diversity and community composition of eukaryotic plankton between > 0.2 μm and 0.2–200 μm size fractions using high-throughput sequencing. We found about 25% sequence reads were classified as metazoan in both size-fractions communities from the Shidou and Hubian reservoirs. The species richness, alpha and beta diversities of plankton communities and relative abundance of 99.2% eukaryotic OTUs showed no significant changes after pre-filtering by a 200 μm pore-size sieve in the two subtropical reservoirs. Our results further demonstrated a higher level stability in the abundant OTUs than in the low-abundance OTUs under deeper sequencing conditions. This was most likely due to the presence of eDNA from metazoa in our water samples or irregular body forms of large eukaryotes. In contrast, only a tiny minority of extremely rare OTUs (0.8% of the total OTUs and less than 0.05% of the total reads) showed difference in occurrence and/or relative abundance between the two size fractions. Therefore, we conclude that SSU rDNA-based size-fractionated filtering, at least 200 μm pore-size filtering, is unable to discern microeukaryotic plankton from large organisms in the reservoirs. More effective and

complementary approaches should be considered as a supplement for future ecology studies on microeukaryotic plankton such as microscopy and SSU rRNA sequencing rather than just relying on rDNA-based molecular approaches.

Acknowledgements

This work was funded by the National Natural Science Foundation of China (31370471), the Natural Science Foundation for Distinguished Young Scholars of Fujian Province (2012J06009), and the Knowledge Innovation Program of the Chinese Academy of Sciences (IUEQN201505). DMW's involvement in this work was aided by a sabbatical visit to the Institute of Urban Environment, Chinese Academy of Sciences. We thank the four reviewers and the editor who provided useful comments to improve this paper.

References

Abad D, Albaina A, Aguirre M *et al*. (2016) Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. *Marine Biology*, **163**, 1–13.

Adl SM, Simpson AG, Lane CE *et al*. (2012) The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, **59**, 429–514.

Albaina A, Aguirre M, Abad D, Santos M, Estonba A (2016) 18S rRNA V9 metabarcoding for diet characterization: a critical evaluation with two sympatric zooplanktivorous fish species. *Ecology and Evolution*, **6**, 1809–1824.

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE*, **4**, e6372.

Bik HM, Porazinska DL, Creer S *et al*. (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, **27**, 233–243.

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. *Ecology*, **89**, 2623–2632.

Boenigk J, Ereshefsky M, Hoef-Emden K, Mallet J, Bass D (2012) Concepts in protistology: species definitions and boundaries. *European Journal of Protistology*, **48**, 96–102.

Bohmann K, Evans A, Gilbert MTP *et al*. (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, **29**, 358–367.

Bonner JT (2013) *Randomness in Evolution*. Princeton University Press, Princeton, USA.

Brussaard CP (2004) Viral control of phytoplankton populations - a review. *Journal of Eukaryotic Microbiology*, **51**, 125–138.

Caporaso JG, Kuczynski J, Stombaugh J *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Cardinale BJ, Duffy JE, Gonzalez A *et al*. (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**, 59–67.

Chen MJ, Chen FZ, Xing P, Li H, Wu QL (2010) Microbial eukaryotic community in response to *Microcystis* spp. bloom, as assessed by an enclosure experiment in Lake Taihu, China. *FEMS Microbiology Ecology*, **74**, 19–31.

Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *The ISME Journal*, **4**, 1053–1059.

Christoffersen K, Riemann B, Klysner A, Søndergaard M (1993) Potential role of fish predation and natural populations of zooplankton in structuring a plankton community in eutrophic lake water. *Limnology and Oceanography*, **38**, 561–573.

Clarke KR, Gorley RN (2015) *PRIMER v7: User Manual/Tutorial*. PRIMER-E Ltd, Plymouth, UK.

Countway PD, Gast RJ, Dennett MR *et al*. (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). *Environmental Microbiology*, **9**, 1219–1232.

de Vargas C, Audic S, Henry N *et al*. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science*, **348**, 1261605.

Dejean T, Valentini A, Duparc A *et al*. (2011) Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE*, **6**, e23398.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Edgcomb V, Orsi W, Bunge J *et al*. (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *The ISME Journal*, **5**, 1344–1356.

Engelbrektson A, Kunin V, Wrighton KC *et al*. (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *The ISME Journal*, **4**, 642–647.

Fenchel T (1987) *Ecology of Protozoa: The Biology of Free-living Phagotropic Protists*. Springer-Verlag, Berlin, Germany.

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**, 423–425.

Finlay BJ, Fenchel T (2004) Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist*, **155**, 237–244.

Geisen S, Laros I, Vizcaíno A, Bonkowski M, Groot G (2015) Not all are free-living: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Molecular Ecology*, **24**, 4556–4569.

Gest H (2004) The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of the Royal Society. *Notes and Records of the Royal Society*, **58**, 187–201.

Gomaa F, Yang J, Mitchell EAD *et al*. (2015) Morphological and molecular diversification of Asian endemic *Difflugia tuberspinifera* (Amoebozoa, Arcellinida): a case of fast morphological evolution in protists? *Protist*, **166**, 122–130.

Green J, Bohannan BJ (2006) Spatial scaling of microbial biodiversity. *Trends in Ecology & Evolution*, **21**, 501–507.

Greenberg A, Clesceri L, Eaton A (1992) *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association, Washington DC, USA.

Harder CB, Rønn R, Brejnrod A *et al*. (2016) Local diversity of heathland Cercozoa explored by in-depth sequencing. *The ISME Journal*, **10**, 2488–2497.

Hardy A (1956) *The Open Sea: The World of Plankton*. Collins, London, UK.

Hutchinson GE (1964) The lacustrine microcosm reconsidered. *American Scientist*, **52**, 334–341.

Jones AC, Liao T, Najar FZ *et al*. (2013) Seasonality and disturbance: annual pattern and response of the bacterial and microbial eukaryotic assemblages in a freshwater ecosystem. *Environmental Microbiology*, **15**, 2557–2572.

Kent M (2012) *Vegetation Description and Data Analysis*. Wiley-Blackwell, Chichester, UK.

Kim DY, Countway PD, Yamashita W, Caron DA (2012) A combined sequence-based and fragment-based characterization of microbial eukaryote assemblages provides taxonomic context for the terminal restriction fragment length polymorphism (T-RFLP) method. *Journal of Microbiological Methods*, **91**, 527–536.

Konopka A, Lindemann S, Fredrickson J (2015) Dynamics in microbial communities: unraveling mechanisms to identify principles. *The ISME Journal*, **9**, 1488–1495.

Kuczynski J, Liu ZZ, Lozupone C *et al*. (2010) Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nature Methods*, **7**, 813–819.

Lepère C, Boucher D, Jardillier L, Domaizon I, Debroas D (2006) Succession and regulation factors of small eukaryote community composition in a lacustrine ecosystem (Lake Pavin). *Applied and Environmental Microbiology*, **72**, 2971–2981.

Liu LM, Yang J, Yu XQ, Chen GJ, Yu Z (2013) Patterns in the composition of microbial communities from a subtropical river: effects of environmental, spatial and temporal Factors. *PLoS ONE*, **8**, e81232.

Liu LM, Yang J, Yu Z, Wilkinson DM (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *The ISME Journal*, **9**, 2068–2077.

Logares R, Audic S, Bass D *et al*. (2014) Patterns of rare and abundant marine microbial eukaryotes. *Current Biology*, **24**, 813–821.

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

Massana R, Gobet A, Audic S *et al*. (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environmental Microbiology*, **17**, 4035–4049.

Massana R, Logares R (2013) Eukaryotic versus prokaryotic marine picoplankton ecology. *Environmental Microbiology*, **15**, 1254–1261.

Medinger R, Nolte V, Pandey RV *et al*. (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular Ecology*, **19**, 32–40.

Pawlowski J, Christen R, Lecroq B *et al*. (2011) Eukaryotic richness in the abyss: insights from pyrotag sequencing. *PLoS ONE*, **6**, e18169.

Pernice MC, Giner CR, Logares R *et al*. (2016) Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *The ISME Journal*, **10**, 945–958.

Pesant S, Not F, Picheral M *et al*. (2015) Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, **2**, 150023.

Quast C, Pruesse E, Yilmaz P *et al*. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, **41**, D590–D596.

R Core Team. (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. http://www.R-project.org.

Rees HC, Maddison BC, Middleditch DJ, Patmore JR, Gough KC (2014) REVIEW: The detection of aquatic animal species using environmental DNA-a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, **51**, 1450–1459.

Reynolds CS (2006) *The Ecology of Phytoplankton*. Cambridge University Press, Cambridge, UK.

Santoferrara LF, Grattepanche JD, Katz LA, McManus GB (2014) Pyrosequencing for assessing diversity of eukaryotic microbes: analysis of data on marine planktonic ciliates and comparison with traditional methods. *Environmental Microbiology*, **16**, 2752–2763.

Santoferrara LF, Tian M, Alder VA, McManus GB (2015) Discrimination of closely related species in tintinnid ciliates: new insights on crypticity and polymorphism in the genus *Helicostomella*. *Protist*, **166**, 78–92.

Schloss PD, Westcott SL, Ryabin T *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Schnetzer A, Moorthi SD, Countway PD *et al*. (2011) Depth matters: microbial eukaryote diversity and community structure in the eastern North Pacific revealed through environmental gene libraries. *Deep Sea Research Part I: Oceanographic Research Papers*, **58**, 16–26.

Shen YF, Zhang ZS, Gong XJ *et al*. (1990) *Modern Biomonitoring Techniques Using Freshwater Microbiota*. China Architecture & Building Press (in Chinese), Beijing, China.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, **21**, 1794–1805.

Stoeck T, Breiner HW, Filker S *et al*. (2014). A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. *Environmental Microbiology*, **16**, 430–444.

Thomsen P, Kielgast J, Iversen LL *et al*. (2012) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, **21**, 2565–2573.

Thomsen PF, Willerslev E (2015) Environmental DNA - an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, **183**, 4–18.

Wang YM, Liu LM, Chen HH, Yang J (2015) Spatiotemporal dynamics and determinants of planktonic bacterial and microeukaryotic communities in a Chinese subtropical river. *Applied Microbiology and Biotechnology*, **99**, 9255–9266.

Yang J, Lv H, Yang J *et al*.(2016) Decline in water level boosts cyanobacteria dominance in subtropical reservoirs. *Science of the Total Environment*, **557/558**, 445–452.

Yang J, Yu XQ, Liu LM, Zhang WJ, Guo PY (2012) Algae community and trophic state of subtropical reservoirs in southeast Fujian, China. *Environmental Science and Pollution Research*, **19**, 1432–1442.

Yang XK, Lu XX (2014) Drastic change in China's lakes and reservoirs over the past decades. *Scientific Reports*, **4**, 6041.

Yu LY, Zhang WJ, Liu LM, Yang J (2015) Determining microeukaryotic plankton community around Xiamen Island, southeast China, using Illumina MiSeq and PCR-DGGE techniques. *PLoS ONE*, **10**, e0127721.

Yu Z, Yang J, Zhou J, *et al*. (2014) Water stratification affects the microeukaryotic community in a subtropical deep reservoir. *Journal of Eukaryotic Microbiology*, **61**,126–133.

Zinger L, Gobet A, Pommier T (2012) Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, **21**, 1878–1896.

Conflict of interest

All authors have no conflict of interest to declare.

Author contributions

J.Y. and L.L. designed the research; L.L., H.C. and X.Y. performed the experiments; L.L., M.L. and J.Y. analyzed the data; L.L., D.M.W., and J.Y. wrote the paper.

Data accessibility

All raw sequence data from this study have been deposited in the public NCBI database (http://www.ncbi.nlm.nih.gov/) under the accession number SRP062446. Both OTU table and environmental data were uploaded under Dryad doi: http://dx.doi.org/10.5061/dryad.4m7m1.

Supplementary information

Fig. S1 Rarefaction curves of similarity-based operational taxonomic unit (OTU) at 97% sequence similarity threshold. A - the 48 replicate samples (12 replicate samples in both size fractions for each reservoir), B - the two sample pools that were from > 0.2 μm and 0.2–200 μm communities, respectively. Colors indicate the different size fractions (red for > 0.2 μm, green for 0.2–200 μm).

Fig. S2 NMDS ordination for > 0.2 μm (left, n = 24) and 0.2–200 μm (right, n = 24) eukaryotic plankton communities from two subtropical reservoirs (Shidou and Hubian). Abundant – abundant taxa or OTUs, Rare – rare taxa or OTUs. The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold. The sample names are indicated using the format: reservoir_month (triplicate samples for each reservoir in each season). Two sampling reservoirs are shown in different shapes, while the sampling time is shown in different colors. The ρ value indicates the correlation coefficient of RELATE between the > 0.2 μm and 0.2–200 μm eukaryotic plankton communities. Correlations were calculated as pair-wise comparisons of all similarity matrix data (i.e. 276 pairwise combinations for the 24 replicate samples in each size-fraction).

Fig. S3 NMDS ordination for > 0.2 μm (A, n = 24) and 0.2–200 μm (B, n = 24) eukaryotic plankton communities from two subtropical reservoirs (Shidou and Hubian). The operational taxonomic units (OTUs) were defined at 99% sequence similarity threshold. The sample names are indicated using the format: reservoir_month (triplicate samples for each reservoir in each season). Two sampling reservoirs are shown in different shapes, while the sampling time is shown in different colors. The ρ value indicates the correlation coefficient of RELATE between the > 0.2 μm and 0.2–200 μm eukaryotic plankton communities. Correlations were calculated as pair-wise comparisons of all similarity matrix data (i.e. 276 pairwise combinations for the 24 replicate samples in each size-fraction).

Fig. S4 The comparison of relative abundance of reads at phylum or higher taxonomic rank between the > 0.2 μm (n = 24) and 0.2–200 μm (n = 24) eukaryotic plankton communities based on a 99% OTU similarity threshold. Statistical analysis is nonparametric Mann-Whitney test, and all *P* values are > 0.05 in all comparisons between the > 0.2 μm and 0.2–200 μm fractions. Values and error bars indicate mean and standard error (n = 24), respectively. Undetermined – unclassified eukaryotes (sequence similarity > 80%). The eukaryotic plankton groups were classified according to Adl *et al*. (2012), note that the Silva annotation is not entirely identical with Adl *et al*. (2012).

Fig. S5 Plot of square root-transformed mean relative abundances of reads at OTU level (99% similarity threshold) in the > 0.2 μm (n = 24) and 0.2–200 μm (n = 24) eukaryotic plankton communities. Statistical analysis is nonparametric Mann-Whitney test. Data are presented as mean of 24 replicates for 17657 OTUs from both reservoirs. Only 85 (0.48%) OTUs had significant difference between the two size fractions with *P* value < 0.05. The black line is y = x.

Table S1 Sample information of eukaryotic plankton communities from two subtropical reservoirs

Table S2 Diversity, predicted richness and Good's coverage of eukaryotic plankton communities

Table S3 Comparison of OTU number, Chao 1, ACE, Shannon-Wiener, Simpson and Pielou indices between two size fractions (> 0.2 μm *vs* 0.2–200 μm) of all eukaryotic plankton communities from Shidou and Hubian reservoirs (3 replicates in each sample)

Table S4 Comparison of OTU number, Shannon-Wiener, Simpson and Pielou indices between two size fractions of abundant eukaryotic plankton subcommunities from Shidou and Hubian reservoirs

Table S5 Comparison of OTU number, Shannon-Wiener, Simpson and Pielou indices between two size fractions of rare eukaryotic plankton subcommunities from Shidou and Hubian reservoirs

Table S6 The OTUs that are unique and shared between the > 0.2 μm and 0.2–200 μm fraction communities in different seasons and reservoirs

Table S7 The OTUs and reads that are unique and shared between the > 0.2 μm (n = 24) and 0.2–200 μm (n = 24) eukaryotic plankton communities from Shidou and Hubian reservoirs

Table S8 Comparison of OTU number, Shannon-Wiener, Simpson and Pielou indices between the > 0.2 μm (n = 3) and 0.2–200 μm (n = 3) eukaryotic plankton samples

Figure and table legends

Fig. 1 High correlation and synchrony in spatiotemporal variability of eukaryotic plankton community across different size-fractions from the two subtropical reservoirs (Shidou and Hubian). NMDS ordination for the > 0.2 μm (A, n = 24) and 0.2–200 μm (B, n = 24) eukaryotic plankton communities based on the Bray-Curtis similarity (stress: 0.04 for both ordinations). The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold. The sample names are indicated using the format: reservoir_month (triplicate samples for each reservoir in each season). Two sampling reservoirs are shown in different shapes, while the sampling time is shown in different colors. Spearman correlation of Bray-Curtis similarity (RELATE analysis) between the > 0.2 μm and 0.2–200 μm eukaryotic plankton communities (C). Correlations were calculated as pairwise comparisons of all

similarity matrix data (i.e. 276 pairwise combinations for the 24 replicate samples in each size-fraction); the ρ value indicates the correlation coefficient, and the black line is y = x.

Fig. 2 No significant difference was found in community composition of eukaryotic plankton at phylum or higher taxonomic rank with and without 200 μm filtering in both reservoirs (All), in Shidou Reservoir (SD), and in Hubian Reservoir (HB). Statistical analysis is nonparametric Mann-Whitney test, and all *P* values are > 0.05 in all comparisons between the > 0.2 μm and 0.2–200 μm fractions. Data are presented as mean of 24 replicates for both reservoirs, 12 replicates for Shidou Reservoir (SD) and 12 replicates for Hubian Reservoir (HB), respectively. Values and error bars indicate mean and standard error, respectively. Undetermined – unclassified eukaryotes (sequence similarity > 80%). The eukaryotic plankton groups were classified according to Adl *et al*. (2012), note that the Silva annotation is not entirely identical with Adl *et al*. (2012).

Fig. 3 Almost all (> 99%) OTUs showed no significant difference in relative abundance of reads between the > 0.2 μm and 0.2–200 μm eukaryotic plankton communities in both reservoirs (All), in Shidou Reservoir (SD), and in Hubian Reservoir (HB). Statistical analysis is nonparametric Mann-Whitney test. Data are presented as mean of 24 replicates for 10965 OTUs in both reservoirs, 12 replicates for 7160 OTUs in Shidou Reservoir (SD) and 12 replicates for 8646 OTUs in Hubian Reservoir (HB), respectively. The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold. Only 89 (0.8%), 6 (0.08%), and 4 (0.05%) OTUs were significant difference between the > 0.2 μm and 0.2–200 μm size-fractions in both reservoirs (All), in Shidou Reservoir (SD) and in Hubian Reservoir (HB), respectively. The black line is y = x.

Table 1 Comparison of OTU number, Shannon-Wiener, Simpson and Pielou indices between the > 0.2 μm (n = 3) and 0.2–200 μm (n = 3) eukaryotic plankton samples

Table 2 The OTUs and reads that are unique and shared between the > 0.2 μm (n = 24) and 0.2–200 μm (n = 24) eukaryotic plankton communities from Shidou and Hubian reservoirs

Table 3 The OTUs that are unique and shared between the > 0.2 μm (n = 24) and 0.2–200 μm (n = 24) communities after a randomly selected subsets of 100, 500, 1000, 5000, 10000, 20000, 30000, and 36000 cleaned reads from each of the 48 replicates

Table 4 The environmental factors that significantly correlated with spatiotemporal variation of > 0.2 μm and 0.2–200 μm eukaryotic plankton communities

Table 1 Comparison of OTU number, Shannon-Wiener, Simpson and Pielou indices between the > 0.2 μm (n = 3) and 0.2–200 μm (n = 3) eukaryotic plankton samples

	OTU		*P*	Shannon-Wiener		*P*
	> 0.2	0.2–200		> 0.2	0.2–200	
SD_Jan	2002±59	2018±43	> 0.05	5.42±0.09	5.55±0.15	> 0.05
SD_May	2010±71	1941±57	> 0.05	5.24±0.43	4.82±0.18	> 0.05
SD_Jul	1956±53	1935±27	> 0.05	5.23±0.25	5.21±0.09	> 0.05
SD_Oct	1856±58	1839±132	> 0.05	4.82±0.30	4.77±0.49	> 0.05
HB_Jan	1830±84	1816±77	> 0.05	5.07±0.16	5.13±0.11	> 0.05
HB_Apr	1887±31	1908±38	> 0.05	5.11±0.05	5.19±0.08	> 0.05
HB_Jul	2021±82	2039±38	> 0.05	5.44±0.07	5.43±0.05	> 0.05
HB_Oct	2026±33	1997±17	> 0.05	5.48±0.08	5.40±0.09	> 0.05
	Simpson		*P*	Pielou		*P*
	> 0.2	0.2–200		> 0.2	0.2–200	
SD_Jan	0.98±0.00	0.99±0.00	> 0.05	0.71±0.01	0.73±0.02	> 0.05
SD_May	0.96±0.03	0.94±0.01	> 0.05	0.69±0.05	0.64±0.02	> 0.05
SD_Jul	0.97±0.02	0.97±0.01	> 0.05	0.69±0.03	0.69±0.01	> 0.05
SD_Oct	0.95±0.02	0.93±0.03	> 0.05	0.64±0.04	0.63±0.06	> 0.05
HB_Jan	0.98±0.00	0.98±0.00	> 0.05	0.67±0.02	0.68±0.01	> 0.05
HB_Apr	0.97±0.00	0.98±0.00	> 0.05	0.68±0.01	0.69±0.01	> 0.05

HB_Jul	0.99±0.00	0.98±0.00	> 0.05	0.72±0.01	0.71±0.01	> 0.05
HB_Oct	0.98±0.00	0.98±0.00	> 0.05	0.72±0.01	0.71±0.01	> 0.05

SD – Shidou Reservoir, HB – Hubian Reservoir. The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold. Statistical analysis is nonparametric Mann-Whitney test, and all *P* values are > 0.05 in all comparisons. Data are expressed as mean ± s.e. (n = 3).

Table 2 The OTUs and reads that are unique and shared between the > 0.2 μm (n = 24) and 0.2–200 μm (n = 24) eukaryotic plankton communities from Shidou and Hubian reservoirs

		> 0.2 unique	> 0.2 shared	0.2–200 unique	0.2–200 shared
OTU	1[a]	1371 (14.40%)	8149 (85.60%)	1445 (15.06%)	8149 (84.94%)
	10[b]	28 (0.51%)	5479 (99.49%)	36 (0.65%)	5479 (99.35%)
	50[c]	0 (0%)	2550 (100%)	0 (0%)	2550 (100%)
Sequence	1	2587 (0.30%)	861413 (99.70%)	2651 (0.31%)	861349 (99.69%)
	10	185 (0.02%)	857287 (99.98%)	169 (0.02%)	857303 (99.98%)
	50	0 (0%)	834408 (100%)	0 (0%)	834408 (100%)

a: the singleton OTUs were removed, b: the OTUs with < 10 sequences were removed, c: the OTUs with < 50 sequences were removed. The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold. The relative contributions (percent) of OTUs richness or read number in each size fraction are given in parentheses.

Table 3 The OTUs that are unique and shared between the > 0.2 μm (n = 24) and 0.2–200 μm (n = 24) communities after a randomly selected subsets of 100, 500, 1000, 5000, 10000, 20000, 30000, and 36000 cleaned reads from each of the 48 replicates

Sequences per sample	> 0.2 unique	> 0.2 shared	0.2–200 unique	0.2–200 shared
100	392 (50.3%)	387 (49.7%)	397 (50.6%)	387 (49.4%)
500	725 (39.8%)	1098 (60.2%)	746 (40.5%)	1098 (59.5%)
1000	937 (36.8%)	1607 (63.2%))	973 (37.7%)	1607 (62.3%)
5000	1472 (28.6%)	3681 (71.4%)	1474 (28.6%)	3681 (71.4%)
10000	1624 (24.8%)	4923 (75.2%)	1702 (25.7%)	4923 (74.3%))
20000	1655 (20.2%)	6530 (79.8%)	1749 (21.1%)	6530 (78.9%)
30000	1520 (16.6%)	7624 (83.4%)	1569 (17.1%)	7624(82.9%)
36000	1371 (14.4%)	8149 (85.6%)	1445 (15.1%)	8149 (84.9%)

The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold. The relative contributions (percent) of unique or shared OTUs to total OTUs in each size fraction are given in parentheses.

Table 4 The environmental factors that significantly correlated with spatiotemporal variation of > 0.2 μm and 0.2–200 μm eukaryotic plankton communities

	> 0.2 μm (n = 24)		0.2–200 μm (n = 24)	
	Adjusted R^2	*P*	Adjusted R^2	*P*
Temperature	0.305	0.002	0.290	0.004
Dissolved oxygen	0.470	0.002	0.447	0.006
NO_x-N	0.414	0.010	0.396	0.010
Total phosphorus	0.517	0.016	0.491	0.020
PO_4-P	0.361	0.008	0.341	0.008
Chlorophyll *a*	0.239	0.002	0.222	0.002
All variables	0.546		0.506