University of Zurich[UZH]

Zurich Open Repository and Archive
University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2018

Application of Environmental DNA Metabarcoding for Predicting Anthropogenic Pollution in Rivers

Li, Feilong ; Peng, Ying ; Fang, Wendi ; Altermatt, Florian ; Xie, Yuwei ; Yang, Jianghua ; Zhang, Xiaowei

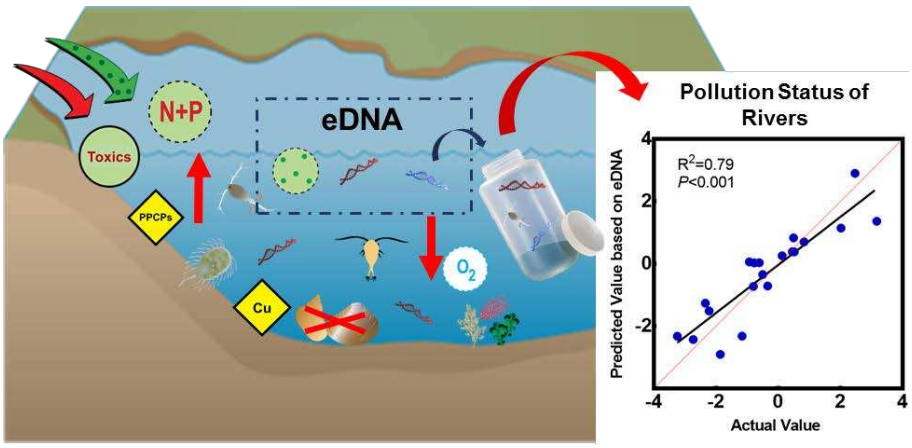
Abstract: Rivers are among the most threatened freshwater ecosystems, and anthropogenic activities are affecting both river structures and water quality. While assessing the organisms can provide a comprehensive measure of a river's ecological status, it is limited by the traditional morphotaxonomy-based biomonitoring. Recent advances in environmental DNA (eDNA) metabarcoding allow to identify prokaryotes and eukaryotes in one sequencing run, and could thus allow unprecedented resolution. Whether such eDNA-based data can be used directly to predict the pollution status of rivers as a complementation of environmental data remains unknown. Here we used eDNA metabarcoding to explore the main stressors of rivers along which community structure changes, and to identify the method's potential for predicting pollution status based on eDNA data. We showed that a broad range of taxa in bacterial, protistan, and metazoan communities could be profiled with eDNA. Nutrients were the main driving stressor affecting communities' structure, alpha diversity, and the ecological network. We specifically observed that the relative abundance of indicative OTUs was significantly correlated with nutrient levels. These OTUs data could be used to predict the nutrient status up to 79% accuracy on testing data sets. Thus, our study gives a novel approach to predicting the pollution status of rivers by eDNA data.

DOI: https://doi.org/10.1021/acs.est.8b03869

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-159261
Journal Article
Accepted Version

Originally published at:
Li, Feilong; Peng, Ying; Fang, Wendi; Altermatt, Florian; Xie, Yuwei; Yang, Jianghua; Zhang, Xiaowei (2018). Application of Environmental DNA Metabarcoding for Predicting Anthropogenic Pollution in Rivers. Environmental Science Technology, 52(20):11708-11719.
DOI: https://doi.org/10.1021/acs.est.8b03869

TOC (Table of Contents Art)



Application of environmental DNA metabarcoding for predicting anthropogenic pollution in rivers

Feilong Li[1], Ying Peng[1], Wendi Fang[1], Florian Altermatt[2,3], Yuwei Xie[4], Jianghua Yang[1], Xiaowei Zhang[1*]

[1] State Key Laboratory of Pollution Control & Resource Reuse, School of the Environment, Nanjing University, Nanjing, P. R. China, 210023

[2] Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland

[3] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[4] Toxicology Centre, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5B3, Canada

***Correspondence:**

Xiaowei Zhang, PhD, Prof

School of the Environment, Nanjing University, Nanjing, 210089, China;

E-mail: zhangxw@nju.edu.cn; howard50003250@yahoo.com

ABSTRACT

Rivers are among the most threatened freshwater ecosystems and anthropogenic activities are affecting both river structures and water quality. While assessing the organisms can provide a comprehensive measure of a river's ecological status, it is limited by the traditional morphotaxonomy-based biomonitoring. Recent advances in environmental DNA (eDNA) metabarcoding allow to identify prokaryotes and eukaryotes in one sequencing run, and could thus allow unprecedented resolution. Whether such eDNA-based data can be used directly to predict the pollution status of rivers as a complementation of environmental data remains unknown. Here we used eDNA metabarcoding to explore the main stressors of rivers along which community structure changes, and to identify the method's potential for predicting pollution status based on eDNA data. We showed that a broad range of taxa in bacterial, protistan and metazoan communities could be profiled with eDNA. Nutrient were the main driving stressor effecting communities' structure, alpha diversity and the ecological network. We specifically observed that the relative abundance of indicative OTUs significantly correlated with nutrient levels. These OTUs data could be used to predict the nutrient status up to 79% accuracy on testing data sets. Thus, our study gives novel approaches into predicting the pollution status of rivers by eDNA data.

INTRODUCTION

Rivers are exposed to multiple stressors, particularly those derived from anthropogenic pollutants, such as excess nutrient, heavy metals, pesticide, or pharmaceuticals.[1,2] Severe pollution reduces the rivers' provisioning of goods and ecosystem services.[3] To alleviate rivers' degradation, and to finally achieve 'non-toxic environment' and 'good health status' goals, governments implement laws and regulations to manage and improve the water environment.[4] For example, the European Water Framework Directive (WFD, adopted in 2000) explicitly requires the vast majority of water bodies in member states to reach a "good status" by 2015.[5] While attempts to monitor chemical contents in waters can directly evaluate the pollution status of rivers, the potential biotoxicity and ecological effects of pollutants can rarely be sufficiently assessed.[6] Alternatively, biological communities give a comprehensive indication of the physical and chemical properties of rivers, and are both the focus of river protection but can also be used as monitoring targets. Consequently, they are monitored in the context of applied environmental protection strategies in several countries.[7] This shift from a focus on chemicals to the focal community to measure quality of waters is widely recognized.[5,8] However, due to the limitations of the traditional morphology-based species identification approach, river monitoring is extremely time consuming, labor-intensive and taxonomic expertise demanding.[8,9]

Environmental DNA (eDNA) metabarcoding provides a fast and efficient way to uncover biodiversity information, which by now has routinely been used to detect individual species[10,11] or biological communities in aquatic ecosystem.[12,13] The eDNA approach has a highly sensitive detection capability and is non-invasive to the organisms themselves,[14] which gives an

unprecedented opportunity to overcome bottlenecks of traditional morphology-based biomonitoring.[15] Although environmental conditions have been speculated to influence eDNA persistence in aquatic ecosystems,[16,17] a recent study shows that the decay rate of eDNA can be modelled using first-order constant,[18] and may thus be a rather robust tool. By incorporating eDNA shedding and decay rates, the transport of eDNA can be effectively modeled to estimate species richness in a natural river ecosystem.[19] Comparisons of biodiversity information from eDNA metabarcoding and morphological datasets obtain similar results for freshwater communities.[9,20] In addition to detecting a set of target taxa, eDNA metabarcoding can also provide access to the broadest set of biodiversity present in the environment.[12,14] For example, a tree of life metabarcoding or meta-systematics approach has been applied to get a holistic biodiversity perspective at the ecosystem level.[14] This biodiversity revealed by eDNA metabarcoding carries rich information on the local community, but it is still largely underexplored what advantages it carries beyond identifying richness information only.

Recently, taxonomy-free molecular indices suggest new measurements for ecosystem assessment using supervised machine learning models.[21] This method provides a new way to monitor water pollution by using high-throughput sequence data. However, the calculation of these taxonomy-free molecular indices still needs information such as taxon-specific ecological weights or categories of tolerance to disturbance.[21,22] Especially for bacteria or foraminifera that play an important role in ecological process, these communities with a large proportion of eDNA reads could not be used to calculate indices due to the lack of relevant ecological weights information.[23] Compared with biotic indices, the composition and trophic structure of species in a community may better reflect and capture interactions between the pollution of an

ecosystem and the subsequent changes in the ecological network. For example, the abundance of arthropods decreases when pyrethroid is discharged into water, causing algal blooms due to the lack of herbivores.[24] Such changes in a river's status can only be understood and predicted by changes in species composition data. Given that eDNA has the advantage of monitoring multiple communities in one sequencing run,[14] it offers a promising tool to assess the species composition of rivers. However, whether such eDNA data can also directly predict the river's pollution status is still insufficiently known.

Here, we used eDNA metabarcoding to profile the species assemblages in rivers from the Yangtze River Delta (YRD), in order to evaluate the method's ability to associate community data with pollution levels. The YRD area is one of the most developed region in China, and serves as an indispensable water resources for agriculture and industry of 150 million people in eastern China. Large amounts of pollutants discharged into the YRD in recent decades make the study of these rivers a high priority for human welfare. Hence, the main purposes of our study are three-fold: 1) to profile species assemblages in rivers using eDNA metabarcoding; 2) to explore the main stressors of rivers along which community structure changes, and to rebuild a known stressor gradient based on environmental variables such that it can reveal multiple communities' response under this known stressor gradient; 3) to predict the pollution status of rivers using eDNA data, and to identify the method's accuracy by comparing testing and training data sets.

MATERIALS AND METHODS

Study area and eDNA sampling

Twenty-two sites were sampled from the YRD area during April and May 2016 (Figure

S1). These sites are located in tributaries of the lower reach of Yangtze River (5 sites, TYR), the Qinhuai River (7 sites, QHR) and the tributaries of Tai Lake (10 sites, TTL), respectively. Qinhuai River is a tributary of Yangtze River flowing through Nanjing City. Tai Lake is the third largest freshwater lake in China as an indispensable water resources for agriculture and industry products.[25] These rivers are exposed to various sources of anthropogenic pollutants.[26,27] At each site, ten liters of surface water were sampled using sterile bottles (Thermo Fisher Scientific™, USA), and immediately transferred on ice. One liter per site was used for eDNA metabarcoding analysis (which has been shown to be sufficient in many settings),[14, 28,29] the remaining seven liters for chemical analyses (see next section). For the eDNA analysis, filtration was done within less than 6 hours after sampling. Four independent extractions of 200–250 mL were made from each one-liter water sample by filtering across a Millipore 0.22 μm hydrophilic nylon membrane (Merck Millipore, USA). The total volume of water filtered for each membrane disc depended on the turbidity of water. The membrane discs containing captured eDNA were placed in 5.0 mL centrifugal tubes, were immediately frozen and stored at −20 °C until DNA extraction.

Analysis of environmental variables

Twenty-two environmental variables were measured for each sampling site. Water temperature (WT), pH and dissolved oxygen (DO) were measured using YSI water quality analyzer *in situ* (YSI Incorporated, USA). For each site, the seven one-liter surface water samples were used to measure basic water quality variables, including permanganate index (COD), total phosphorus (TP), total nitrogen (TN), nitrate (NO_3^-), nitrite (NO_2^-), ammonia nitrogen (NH_4^+) and biochemical oxygen demand (BOD) following standard methods (NEPB,

2002), respectively. For heavy metals, one liter surface water was diluted with 2% HNO_3 and filtered through a 2.5 μm membrane filter (Whatman, UK). We then determined the concentration of Cr, Mn, Ni, Cu, Zn, As, Cd and Pb using inductively Coupled Plasma Mass Spectrometry (ICP-MS). For organic chemicals, one liter surface water was analyzed by a Thermo Ultimate 3000 high performance liquid chromatograph (Thermo Fisher, USA) coupled to a quadrupole-orbitrap instrument (Thermo QExactive Plus) equipped with a heated electrospray ionisation (ESI) source (details shown in supporting Information, SI). These organic chemicals were classified into four major classes, including pesticide, medical drug, industrial processing aid (IPA) and personal care (PerC) components. Detailed information on these chemicals is based on Peng et al.[25]

DNA extraction, PCR amplification, and next generation sequencing

eDNA was extracted directly from the filter membrane discs and blank controls (autoclaved tap water) with a DNeasy PowerWater Kit (Qiagen Canada Inc., ON, Canada) following the manufacturer's protocol. Multiple PCR assays were carried out for the target gene following the previously published protocol.[13] Briefly, a universal eukaryotic primer pair (1380F: TCCCTGCCHTTTGTACACAC; 1510R: CCTTCYGCAGGTTCACCTAC) was used to amplify the 130 bp fragment of the hypervariable region of 18S rRNA genes.[30] In analogy, the 180 bp fragment of bacterial 16s rRNA genes was amplified using the modified V3 primer pair (341F: ACCTACGGGRSGCWGCAG; 518R: GGTDTTACCGCGGCKGCTG).[31] To pool and sequence all samples in one sequencing run, unique 12-nt nucleotide codes (also known as tags) were added to the 5'-ends of the forward or reverse primers. All primers were synthesized by Shanghai Generay Biotech Co., Ltd. Each

153 eDNA sample was amplified in three PCR replicates to minimize potential PCR bias, and the
154 products were subsequently combined. PCR negative controls (nuclease-free water as DNA
155 template) were included for all assays. PCRs were carried out in 50 μL reaction mixture,
156 including 31 μL of ddH_2O, 10 μL of 5X Phusion Green HF Buffer, 1 μL of 10 mM dNTPs, 2.5
157 μL of each primer (10 μM), 2.5 μL of DNA template and 0.5 μL of Phusion Green Hot Start II
158 High-Fidelity DNA Polymerase (Thermo Fisher Scientific™, USA). The amplification
159 protocol was as follows: initial denaturation at 98 °C for 30 s followed by 30 cycles at 98 °C
160 for 5 s, 62 °C for 30 s and 72 °C for 15 s, with a final extension at 72 °C for 7 min, and the
161 PCR was cooled to 4 °C until removed. PCR products were visualized on a 2% agarose gel to
162 check the expected size of PCRs yielded amplicons. Thereafter, the PCR products were purified
163 using the E-Z 96® Cycle Pure Kit (Omega, USA). All purified PCR products were quantified
164 using Qubit™ dsDNA HS Assay Kits (Invitrogen, USA), and were pooled equally for
165 subsequent sequencing. Sequencing adaptors were linked to purified DNA fragments with the
166 Ion Xpress™ Plus Fragment Library Kit (Thermo Fisher Scientific™, USA) following the
167 manufacturer's protocol. Finally, all samples were diluted to a final concentration of 100 pM.
168 Sequencing templates were prepared with Ion OneTouch 2™ and sequenced in the Ion Proton
169 sequencer (Life Technologies, USA).

170 Low quality raw sequence (mean quality < 20, scanning window = 50, sequences
171 contained ambiguous 'N', homopolymer and sequence length: < 100 bp) were discarded using
172 *split_libraries.py* script in QIIME toolkit.[32] The cleaned reads were sorted and distinguished
173 by unique sample tags, all sequences were clustered into OTUs following the UPARSE pipeline
174 at cutoff value of 97% nucleotide similarity. The taxonomy annotation for each OTUs in

bacterial, protistan and metazoan community was assigned against the Greengenes database[33] or the Protist Ribosomal Reference database[34] using *align_seqs.py* script, and OTUs number and Shannon entropy index of each community were calculated using *alpha_diversity.py* script in QIIME toolkit.[32]

Statistical analyses

First, eDNA metabarcoding datasets were summarized in separate OTUs table, the taxonomic phylogenetic tree was built using the interactive tree of life (iTOL) online tool.[35] Then, to meet the prerequisite of parametric tests, all environmental variables except pH were log(x+1) transformed and normalized. To extract the main components explaining the variance of the environmental variables, a principle component analysis (PCA) was performed with the Kaiser-Meyer-Olkin (KMO) and Bartlett's sphericity test, eigenvalues >1 and absolute r>0.50 were taken as criterion for the extraction of the principal components (PC) and the strongly correlation between PC and environmental variables, respectively.[36] After, to rebuild a known stressor gradient, all samples were split into three levels (named low, medium and high level) using the one third of the PC1 distribution as boundaries.[37] To detect the difference of environmental variables between each level, non-parametric Kruskal-Wallis (K-W) tests were conducted, followed by post hoc Mann-Whitney-U tests.

To select the significant environmental variables in explaining the variation of bacterial, protistan and metazoan community structure, forward selection distance-based linear models (distLM), based on AIC selection criteria, were used. The significance levels of the variables were assessed by Monte Carlo permutation tests (999 permutations).[13] To illustrate the variation of communities' structure among three levels, non-metric multidimensional scaling

(nMDS) ordination based on Bray-Curtis (bacteria and protist) and Jaccard (metazoa) dissimilarity matrices were used and the significant differences were assessed by permutational multivariate analyses of variance test (PERMANOVA).[38] To identify major OTUs that were responsible for the difference in community structure between each level, a SIMPER analysis was conducted. Using multiple non-linear regression, we tested the relations between nutrient (surrogated by PC1) and Shannon index of each community. Finally, network analysis was used to explore co-occurrence ecological patterns between OTUs in complex communities, which might be more difficult to detect using either the traditional α- or β-diversity index. Network visualization of the co-occurrence relationships were generated by SparCC with 100 bootstraps to assign *P*-values.[39] Only robust and significant correlations ($|\rho| > 0.7$ and 'two tailed' $P < 0.01$) between nodes were retained in the network.

Indicative OTUs of each level were identified using *multipatt* function in the R package *Indispecies*, the Indictor Values (IndVal) were measured to reflect the conditional probability of the OTUs as an indicator, the significance was tested using a permutation test (nperm=999). To predict the pollution status of rivers based on these indicative OTUs data, predictive models were fitted for three of the four independent subsamples at each site (training data sets) using multivariate linear regression models (MLR) implemented in SPSS 22 software. The reliability and significance of each formula was tested by bootstrap resampling (n=1000). Finally, the remaining sub-sample at each site were used as testing data sets, to examine the accuracy of predicted values derived from predictive models compared with actual measured values.

All of the above statistical analyses were performed in the R statistical language (http://www.r-project.org), GraphPad Prism 6.01 software, SPSS 22 software and PRIMER7

with PERMANOVA+ add-on software (PRIMER-E Ltd, Ivybridge, UK). The network was analyzed and visualized using Cytoscape V3.[40]

RESULTS AND DISCUSSION

eDNA metabarcoding provided a wide spectrum of taxonomic diversity

We detected a total of 1,640,832 bacterial reads, 3,079,304 protistan reads and 362,672 metazoan reads across all samples after stringent quality filtering (Table S1). These eDNA data were assigned to 5,850 bacterial OTUs, 3,475 protistan OTUs and 274 metazoan OTUs (Table S1), annotating (to the highest phylogenetic level resolved) 51 phyla (98.8% of the total OTUs), 188 classes (96.8% OTUs), 347 orders (79.9% OTUs), 714 families (70.9% OTUs), 623 genera (46.9% OTUs) and 355 species (25.9% OTUs) (Figure 1a). The majority of taxonomic lineages at family level belonged to Chlorophyta, Ciliophora, and Proteobacteria (Figure 1b), the relative abundance of these taxa were also disproportionally high (Figure S2). By using different PCR assays, a wide taxonomic lineage including Proteobacteria, Actinobacteria, Chloroflexi, Ciliophora, Chlorophyta, Ochrophyta, Arthropoda, and Mollusca were recovered from the samples. Recent studies have also demonstrated that eDNA methodologies can be used to assess a broad range of prokaryotes and eukaryotes from a variety of environments (e.g. freshwater, seawater and soil).[12,14,41] However, there are still some issues (e.g. primer bias, sequencing artefacts and/or contamination) of the eDNA methodologies to be improvded,[15] and the taxonomic resolution largely depends on the choice of primer sets and corresponding reference database.[42] For example, the chloroplastic *rbcL* and nuclear ribosomal 18S genes (e.g. V4 and V9 region) have been used in algae studies,[42-44] however, there have been controversial in the taxonomic resolution of algae by these primers. These studies suggested that the *rbcL*

and V4 region of 18S are more suitable for diatom,[42,44] yet V9 region of 18S could detect a wide range of taxonomic groups.[43] The same issue of primer bias may be found in this study, for example, metazoans have fewer sequences than eukaryotic protists using single V9 region of 18S. Hence, multiple PCR assays using different gene regions are strongly advocated to assess biota in monitoring biodiversity.

Nutrient identified as a major stressor of rivers

The values of the twenty-two environmental variables assessed varied largely across samples. Subsequently, all of these variables were reduced to five principal components based on PCA (details were available in SI, Table S2). The first two principal components (PC1 and PC2) explained 40.40% of variances of the total variables (Figure 2). Therein, nutrient (including DO, NO_3^-, NH_4^+, TN and TP) were most strongly associated with the first principal component (PC1, Table S2, Figure 2), which explained 23.70 % of the variation in the data. We then used this structuring along the PCA axes as our main environmental descriptors based on which we wanted to study community shifts. Nutrient are in this context relevant, as they have become one of the most severe environmental problems in this region after decades of dense input of nutrient from anthropogenic activities.[45,46] We used the PC 1 as a new predictor representing nutrient gradients, and then split all samples into three levels (hereafter called "*Low nutrient*", "*Medium nutrient*", and "*High nutrient*") by using the 33rd and 66th percentile of the PC1 distribution as boundaries.[37] As expected, six nutrient-related parameters were significantly different among these three levels. Specifically, the concentrations of NH_4^+, TN and TP in *High nutrient* level were significantly higher than in the *Low nutrient* level (Figure S3), while the concentration of NO_3^-, pH and DO had an opposing distribution. The other 16

263 variables (BOD, COD and heavy metal, and other chemicals) were not significantly different
264 among the three levels.

265 **Changes of communities' structure were mainly due to nutrient**

266 Effects of nutrient on communities' structure were greater than other environment
267 variables. Based on distLM analysis, the most parsimonious models explained 60.40%, 62.24%,
268 and 35.61% of the total variation in bacterial, protistan and metazoan communities' structure,
269 respectively (Table 1). In particular, most of the variance in these communities' structure could
270 be explained by nutrient levels (approximately 40%, 30%, and 22%, respectively). These
271 results were coinciding exactly with the PCA of environmental variables indicating that
272 nutrient were the driving stressor of these rivers. Hence, the pollution status of rivers may be
273 directly revealed by such species information, and may not need to have environmental
274 variables being monitoring.[8] Importantly, the biological assessment gives an integrated
275 measure of the nutrient exposition of a community over time, while chemical measurements
276 usually only cover one specific time point. To verify this hypothesis, we then analyzed the trend
277 of communities' change under known stressor gradients (nutrient gradients), which is expect
278 to identify some taxa for characterizing the nutrient status in rivers.

279 First, we found that the dominant taxa in bacterial, protistan, and metazoan communities
280 varied across the nutrient gradients (Figure 3a). Some taxa (Myxozoa, Nitrospirae,
281 Foraminifera and Stramenopiles, Mollusca, and Arthropoda) were primarily identified in *Low*
282 and/or *Medium nutrient* level, and taxa in Opisthokonta_unknown, Choanoflagellida,
283 Centroheliozoa, Gastrotricha, and Rotifera were dominant in *High nutrient* level. These results
284 were consistent with other studies showing that eutrophication altered the composition of

communities.[47,48] For example, Mollusca and Foraminifera were sensitive to nutrient,[49,50] Rotifers increased along a gradient of increasing nutrient levels.[51] Besides, as a chemolithoautotrophic nitrite-oxidizing bacterium,[52] Nitrospirae was primarily found in *Low* and *Medium nutrient* level, which might explain why the concentration of NO_3^- in these two levels were greater.

Then, the communities' structure also largely varied following the nutrient gradient rather than across regions (Figure 3b), which were further identified as significant by the PERMANOVA tests (pseudo-$F_{bacteria}$=5.187, P<0.001; pseudo-$F_{protist}$=7.854, P<0.001; pseudo-$F_{metazoa}$=9.188, P<0.001). The SIMPER analyses revealed that *Microcystis* sp. (Cyanobacteria, OTU625), *Mycobacterium* sp. (Actinobacteri, OTU368), and ACK.M1 (Actinobacteria, OTU280) in bacterial communities were the major contributors to the dissimilarity across each level (Table S3). *Mediophyceae* sp. (Bacillariophyta, OTU726), *Cryptomonas* sp. (Cryptophyta, OTU544), and *Strobilidiidae* sp. (Ciliophora, OTU104) in protozoan communities and *Sinocalanus* sp. (Arthropoda, OTU48), *Leiosolenus* sp. (Mollusca, OTU87), and *Brachionus calyciflorus* (Rotifera, OTU103) in metazoan communities were the main contributors to the dissimilarity. As is well known, *Microcystis* became a dominant taxon during cyanobacterial bloom periods,[53] and could produce highly stable and potent polypeptides (microcystins (MCs)) that pose a serious threat to public health.[46,54] Nutrient enrichment combined with high ambient temperature was regarded as the main stressor that influenced on *Microcystis* blooms.[55] Besides, some laboratory toxicity studies found that some species in *Brachionus* could be more tolerant of ammonia than cladocerans and copepods.[56,57]

Alpha biodiversity and ecological interaction network varied with nutrient status

Alpha biodiversity (Shannon index) indicated a significantly hump-shaped response to nutrient in bacterial, protistan and metazoan communities (Figure 4a-c). Our results contribute to growing evidence of nonlinear responses of aquatic assemblages to nutrient enrichment.[58,59] Although linear responses of biodiversity to nutrient enrichment were also reported,[60] some factors might frequently obscure natural nonlinear responses of multiple taxa to stressors. For example, biotic interaction could change structure of the food web within an ecosystem,[61] so that the consumer–resource interactions in communities were often affected by other species.[62,63] The top predator species could determine how communities' conditions changed across time and space.[64] The network analysis verifies the above inference, such that the ecological interaction network between OTUs in each community revealed a distinct network topology in each nutrient level (Figure 4d). The number of nodes and edges in network were higher in *Low nutrient*, followed by *High* and *Medium nutrient* (Table S4). In addition, more complex ecological interactions in *Low* and *High nutrient* level according to the betweenness centrality and average closeness centrality parameters.[65]

Novel OTUs based indicator could rapidly predict nutrient status of rivers

Indicator analysis identified 960 OTUs in different taxa (i.e. Ochrophyta, Ciliophora, Actinobacteria, Proteobacteria and Chlorophyta) that were characteristic of each nutrient level (Figure S4). These OTUs included their trophic positioning and responsiveness to a range of nutrient gradients.[42,66] In addition, the relative abundance of indicative OTUs in Ochrophyta, Ciliophora, Arthropod, Proteobacteria, Cryptophyta and Cyanobacteria were significantly negatively correlated with nutrient, indicative OTUs in Actinobacteria, Chlorophyte and Rotifera could increase with nutrient gradient (Figure 5a). Here it is noteworthy that the

329 relationship between the relative abundance of indicative OTUs in cyanobacteria and nutrient
330 was contradictory to previous study.[55] One possible explanation is that more than half of sites
331 had reached eutrophication or severe eutrophication status (Table S2). A previous study within
332 this study area confirmed this phenomenon that high cyanobacteria cell concentrations were
333 negatively correlated with ammonia, especially in appropriate external water temperature.[67]
334 Recently, palaeo-limnological views based on subfossil DNA also supported strong
335 correlations between trophic status changes and microbial eukaryotes community succession.[68]
336 By contrast, the largest proportions of these indicator were Ochrophyta, particularly more than
337 50% of OTUs in Ochrophyta belonging to *Cyclotella* sp., *Nitzschia* sp., *Melosira* sp., or others.
338 (Bacillariophyceae). Diatoms inhabit a variety of waterbodies and different species could
339 respond differently and characteristically to environmental status.[69] In particular, the *Nitzschia*
340 sp. has been used to characterize the nutrient status in waters.[70] Besides, diatoms were
341 considered as the most sensitive group to TP, and the occurrence of OTUs declined with
342 increasing concentrations of TP.[38]

343 Using identified indicative OTUs, we could identify the nutrient status with 41%–75%
344 accuracy on training data sets, but the predictive ability of single communities was lower than
345 combined communities' data (Table 2). When comparing the nutrient (surrogated by PC1)
346 predicted value with the actual value using test data sets, there was a good consistency between
347 each value (R^2=0.51–0.79, Figure 5b). However, we also found underprediction or
348 overprediction of the nutrient status using single community data. For example, the predicted
349 value in *Low nutrient* level was higher than actual ones using bacterial or metazoan data, but
350 the lower predicted value was occurred in *High nutrient* level. Besides, almost all predictions

were higher than the actual ones in protistan data. Corresponding to single community data, we achieved up to 79% accuracy to predict the nutrient status via combining multiple communities' data. An recent study has also demonstrated that multi-trophic metabarcoding biotic index has higher predictive potential for pollutant status.[23] In another study, it has also been found that sequence data was a better predictor than environmental variables to predict cyanobacterial blooms.[28] In short, eDNA metabarcoding offers novel and promising tools to monitor and predict anthropogenic contamination of aquatic ecosystems (e.g. rivers, lakes, marine ecosystems) by DNA sequence-based data.

ASSOCIATED CONTENT

Supporting Information

Detailed information regarding method for the detection of organic chemical and principal component analysis of environmental variables, Figures showing location of sampling sites, the relative abundance of taxa at phylum level, comparison of environmental variables, the number of indicative OTUs, Tables showing sequencing counts and shannon entropy index, summary of environmental variables and the results of PCA, SIMPER analysis, parameters of network topology. (PDF)

ACKNOWLEDGMENTS

For support, we thank the Major Science and Technology Program for Water Pollution Control and Treatment (#2017ZX07602-002) and Nature Science Foundation of Jiangsu Province (BK20160650 and BK2018043082). X.Z. was supported by the Fundamental Research Funds for the Central Universities and the Collaborative Innovation Center for

Regional Environmental Quality. F.A. was supported by the Swiss National Science Foundation Grants No PP00P3_179089 and 31003A_173074, and the University of Zurich Research Priority Programme URPP "*Global Change and Biodiversity*". Y.X. was supported by the Global Water Futures program titled "Next generation solutions to ensure healthy water resources for future generations" (#419205).

REFERENCES

(1) Woodward, G.; Gessner, M. O.; Giller, P. S.; Gulis, V.; Hladyz, S.; Lecerf, A.; Malmqvist, B.; McKie, B. G.; Tiegs, S. D.; Cariss, H.; Dobson, M.; Elosegi, A.; Ferreira, V.; Graca, M. A. S.; Fleituch, T.; Lacoursiere, J. O.; Nistorescu, M.; Pozo, J.; Risnoveanu, G.; Schindler, M.; Vadineanu, A.; Vought, L. B. M.; Chauvet, E., Continental-Scale Effects of Nutrient Pollution on Stream Ecosystem Functioning. *Science* **2012,** *336*, (6087), 1438-1440.

(2) Islam, M. S.; Ahmed, M. K.; Raknuzzaman, M.; Habibullah-Al-Mamun, M.; Islam, M. K., Heavy metal pollution in surface water and sediment: A preliminary assessment of an urban river in a developing country. *Ecol Indic* **2015,** *48*, 282-291.

(3) Arthington, A. H.; Naiman, R. J.; McClain, M. E.; Nilsson, C., Preserving the biodiversity and ecological services of rivers: new challenges and research opportunities. *Freshwater Biol* **2010,** *55*, (1), 1-16.

(4) Leese, F.; Bouchez, A.; Abarenkov, K.; Altermatt, F.; Borja, Á.; Bruce, K.; Ekrem, T.; Čiampor, F.; Čiamporová-Zaťovičová, Z.; Costa, F. O.; Duarte, S.; Elbrecht, V.; Fontaneto, D.; Franc, A.; Geiger, M. F.; Hering, D.; Kahlert, M.; Kalamujić Stroil, B.; Kelly, M.; Keskin, E.; Liska, I.; Mergen, P.; Meissner, K.; Pawlowski, J.; Penev, L.; Reyjol, Y.; Rotter, A.; Steinke, D.; van der Wal, B.; Vitecek, S.; Zimmermann, J.; Weigand, A. M., Why We Need Sustainable Networks Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0: A Perspective Derived From the DNAqua-Net COST Action. In *Advances in Ecological Research*, Bohan, D. A.; Dumbrell, A. J.; Woodward, G.; Jackson, M., Eds. Academic Press: 2018; Vol. 58, pp 63-99.

(5) Hering, D.; Borja, A.; Carstensen, J.; Carvalho, L.; Elliott, M.; Feld, C. K.; Heiskanen, A. S.; Johnson, R. K.; Moe, J.; Pont, D.; Solheim, A. L.; Van De Bund, W., The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. *Sci Total Environ* **2010,** *408*, (19), 4007-4019.

(6) Burton, G. A., Jr., The focus on chemicals alone in human-dominated ecosystems is inappropriate. *Integr Environ Assess Manag* **2017,** *13*, (4), 568-572.

(7) Birk, S.; Bonne, W.; Borja, A.; Brucet, S.; Courrat, A.; Poikane, S.; Solimini, A.; van de Bund, W. V.; Zampoukas, N.; Hering, D., Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecol Indic* **2012,** *18*, 31-41.

(8) Pawlowski, J.; Kelly-Quinn, M.; Altermatt, F.; Apothéloz-Perret-Gentil, L.; Beja, P.; Boggero, A.; Borja, A.; Bouchez, A.; Cordier, T.; Domaizon, I.; Feio, M. J.; Filipe, A. F.; Fornaroli, R.; Graf, W.; Herder, J.; van der Hoorn, B.; Iwan Jones, J.; Sagova-Mareckova, M.; Moritz, C.; Barquín, J.; Piggott, J. J.; Pinna, M.; Rimet, F.; Rinkevich, B.; Sousa-Santos, C.; Specchia, V.; Trobajo, R.; Vasselon, V.; Vitecek, S.; Zimmerman, J.; Weigand, A.; Leese, F.; Kahlert, M., The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological

assessment of aquatic ecosystems. *Sci Total Environ* **2018,** *637-638*, 1295-1310.
(9) Bista, I.; Carvalho, G. R.; Walsh, K.; Seymour, M.; Hajibabaei, M.; Lallias, D.; Christmas, M.; Creer, S., Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nat Commun* **2017,** *8*, 14087.
(10) Dougherty, M. M.; Larson, E. R.; Renshaw, M. A.; Gantz, C. A.; Egan, S. P.; Erickson, D. M.; Lodge, D. M., Environmental DNA (eDNA) detects the invasive rusty crayfish Orconectes rusticus at low abundances. *J Appl Ecol* **2016,** *53*, (3), 722-732.
(11) Sansom, B. J.; Sassoubre, L. M., Environmental DNA (eDNA) Shedding and Decay Rates to Model Freshwater Mussel eDNA Transport in a River. *Environ Sci Technol* **2017,** *51*, (24), 14244-14253.
(12) Deiner, K.; Fronhofer, E. A.; Machler, E.; Walser, J. C.; Altermatt, F., Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nat Commun* **2016,** *7*, 12544.
(13) Xie, Y.; Hong, S.; Kim, S.; Zhang, X.; Yang, J.; Giesy, J. P.; Wang, T.; Lu, Y.; Yu, H.; Khim, J. S., Ecogenomic responses of benthic communities under multiple stressors along the marine and adjacent riverine areas of northern Bohai Sea, China. *Chemosphere* **2017,** *172*, 166-174.
(14) Stat, M.; Huggett, M. J.; Bernasconi, R.; DiBattista, J. D.; Berry, T. E.; Newman, S. J.; Harvey, E. S.; Bunce, M., Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Sci Rep* **2017,** *7*, (1), 12240.
(15) Deiner, K.; Bik, H. M.; Machler, E.; Seymour, M.; Lacoursiere-Roussel, A.; Altermatt, F.; Creer, S.; Bista, I.; Lodge, D. M.; de Vere, N.; Pfrender, M. E.; Bernatchez, L., Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol Ecol* **2017,** *26*, (21), 5872-5895.
(16) Barnes, M. A.; Turner, C. R.; Jerde, C. L.; Renshaw, M. A.; Chadderton, W. L.; Lodge, D. M., Environmental Conditions Influence eDNA Persistence in Aquatic Systems. *Environ Sci Technol* **2014,** *48*, (3), 1819-1827.
(17) Jerde, C. L.; Olds, B. P.; Shogren, A. J.; Andruszkiewicz, E. A.; Mahon, A. R.; Bolster, D.; Tank, J. L., Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. *Environ Sci Technol* **2016,** *50*, (16), 8770-9.
(18) Sassoubre, L. M.; Yamahara, K. M.; Gardner, L. D.; Block, B. A.; Boehm, A. B., Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environ Sci Technol* **2016,** *50*, (19), 10456-10464.
(19) Deiner, K.; Altermatt, F., Transport distance of invertebrate environmental DNA in a natural river. *Plos One* **2014,** *9*, (2), e88786.
(20) Stoeckle, M. Y.; Soboleva, L.; Charlop-Powers, Z., Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *Plos One* **2017,** *12*, (4), e0175186.
(21) Cordier, T.; Esling, P.; Lejzerowicz, F.; Visco, J.; Ouadahi, A.; Martins, C.; Cedhagen, T.; Pawlowski, J., Predicting the Ecological Quality Status of Marine Environments from eDNA Metabarcoding Data Using Supervised Machine Learning. *Environ Sci Technol* **2017,** *51*, (16), 9118-9126.
(22) Apotheloz-Perret-Gentil, L.; Cordonier, A.; Straub, F.; Iseli, J.; Esling, P.; Pawlowski, J., Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Mol Ecol Resour* **2017,** *17*, (6), 1231-1242.
(23) Keeley, N.; Wood, S. A.; Pochon, X., Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. *Ecol Indic* **2018,** *85*, 1044-1057.
(24) De Laender, F.; Rohr, J. R.; Ashauer, R.; Baird, D. J.; Berger, U.; Eisenhauer, N.; Grimm, V.; Hommen, U.; Maltby, L.; Meliàn, C. J., Reintroducing environmental change drivers in biodiversity–ecosystem functioning research. *Trends Ecol Evol* **2016,** *31*, (12), 905-915.
(25) Peng, Y.; Fang, W.; Krauss, M.; Brack, W.; Wang, Z.; Li, F.; Zhang, X., Screening hundreds of emerging organic pollutants (EOPs) in surface water from the Yangtze River Delta (YRD): Occurrence, distribution,

ecological risk. *Environ Pollut* **2018,** *241*, 484-493.
(26) Shen, J. H.; Gutendorf, B.; Vahl, H. H.; Shen, L.; Westendorf, J., Toxicological profile of pollutants in surface water from an area in Taihu Lake, Yangtze Delta. *Toxicology* **2001,** *166*, (1-2), 71-78.
(27) Wu, B.; Zhao, D. Y.; Jia, H. Y.; Zhang, Y.; Zhang, X. X.; Cheng, S. P., Preliminary Risk Assessment of Trace Metal Pollution in Surface Water from Yangtze River in Nanjing Section, China. *B Environ Contam Tox* **2009,** *82*, (4), 405-409.
(28) Tromas, N.; Fortin, N.; Bedrani, L.; Terrat, Y.; Cardoso, P.; Bird, D.; Greer, C. W.; Shapiro, B. J., Characterising and predicting cyanobacterial blooms in an 8-year amplicon sequencing time course. *Isme J* **2017,** *11*, (8), 1746-1763.
(29) Machler, E.; Deiner, K.; Spahn, F.; Altermatt, F., Fishing in the Water: Effect of Sampled Water Volume on Environmental DNA-Based Detection of Macroinvertebrates. *Environ Sci Technol* **2016,** *50*, (1), 305-312.
(30) Amaral-Zettler, L. A.; McCliment, E. A.; Ducklow, H. W.; Huse, S. M., A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *Plos One* **2009,** *4*, (7), e6372.
(31) Klindworth, A.; Pruesse, E.; Schweer, T.; Peplies, J.; Quast, C.; Horn, M.; Glockner, F. O., Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **2013,** *41*, (1), e1.
(32) Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Tumbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R., QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **2010,** *7*, (5), 335-336.
(33) DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L., Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microb* **2006,** *72*, (7), 5069-5072.
(34) Guillou, L.; Bachar, D.; Audic, S.; Bass, D.; Berney, C.; Bittner, L.; Boutte, C.; Burgaud, G.; de Vargas, C.; Decelle, J.; del Campo, J.; Dolan, J. R.; Dunthorn, M.; Edvardsen, B.; Holzmann, M.; Kooistra, W. H. C. F.; Lara, E.; Le Bescot, N.; Logares, R.; Mahe, F.; Massana, R.; Montresor, M.; Morard, R.; Not, F.; Pawlowski, J.; Probert, I.; Sauvadet, A. L.; Siano, R.; Stoeck, T.; Vaulot, D.; Zimmermann, P.; Christen, R., The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* **2013,** *41*, (D1), D597-D604.
(35) Letunic, I.; Bork, P., Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **2016,** *44*, (W1), W242-5.
(36) Kuppusamy, M. R.; Giridhar, V. V., Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore. *Environ Int* **2006,** *32*, (2), 174-179.
(37) Gutierrez-Canovas, C.; Millan, A.; Velasco, J.; Vaughan, I. P.; Ormerod, S. J., Contrasting effects of natural and anthropogenic stressors on beta diversity in river organisms. *Global Ecol Biogeogr* **2013,** *22*, (7), 796-805.
(38) Chariton, A. A.; Stephenson, S.; Morgan, M. J.; Steven, A. D. L.; Colloff, M. J.; Court, L. N.; Hardy, C. M., Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environ Pollut* **2015,** *203*, 165-174.
(39) Friedman, J.; Alm, E. J., Inferring correlation networks from genomic survey data. *Plos Comput Biol* **2012,** *8*, (9), e1002687.
(40) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: A software environment for integrated models of biomolecular interaction networks.

Genome Res **2003,** *13*, (11), 2498-2504.
(41) Drummond, A. J.; Newcomb, R. D.; Buckley, T. R.; Xie, D.; Dopheide, A.; Potter, B. C.; Heled, J.; Ross, H. A.; Tooman, L.; Grosser, S.; Park, D.; Demetras, N. J.; Stevens, M. I.; Russell, J. C.; Anderson, S. H.; Carter, A.; Nelson, N., Evaluating a multigene environmental DNA approach for biodiversity assessment. *Gigascience* **2015,** *4*, 46.
(42) Visco, J. A.; Apotheloz-Perret-Gentil, L.; Cordonier, A.; Esling, P.; Pillet, L.; Pawlowski, J., Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. *Environ Sci Technol* **2015,** *49*, (13), 7597-7605.
(43) De Vargas, C.; Audic, S.; Henry, N.; Decelle, J.; Mahe, F.; Logares, R.; Lara, E.; Berney, C.; Le Bescot, N.; Probert, I.; Carmichael, M.; Poulain, J.; Romac, S.; Colin, S.; Aury, J. M.; Bittner, L.; Chaffron, S.; Dunthorn, M.; Engelen, S.; Flegontova, O.; Guidi, L.; Horak, A.; Jaillon, O.; Lima-Mendez, G.; Lukes, J.; Malviya, S.; Morard, R.; Mulot, M.; Scalco, E.; Siano, R.; Vincent, F.; Zingone, A.; Dimier, C.; Picheral, M.; Searson, S.; Kandels-Lewis, S.; Acinas, S. G.; Bork, P.; Bowler, C.; Gorsky, G.; Grimsley, N.; Hingamp, P.; Iudicone, D.; Not, F.; Ogata, H.; Pesant, S.; Raes, J.; Sieracki, M. E.; Speich, S.; Stemmann, L.; Sunagawa, S.; Weissenbach, J.; Wincker, P.; Karsenti, E.; Coordinators, T. O., Eukaryotic plankton diversity in the sunlit ocean. *Science* **2015,** *348*, (6237).
(44) Kermarrec, L.; Franc, A.; Rimet, F.; Chaumeil, P.; Humbert, J. F.; Bouchez, A., Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Mol Ecol Resour* **2013,** *13*, (4), 607-619.
(45) Duan, H. T.; Ma, R. H.; Xu, X. F.; Kong, F. X.; Zhang, S. X.; Kong, W. J.; Hao, J. Y.; Shang, L. L., Two-Decade Reconstruction of Algal Blooms in China's Lake Taihu. *Environ Sci Technol* **2009,** *43*, (10), 3522-3528.
(46) Shi, K.; Zhang, Y. L.; Xu, H.; Zhu, G. W.; Qin, B. Q.; Huang, C. C.; Liu, X. H.; Zhou, Y. Q.; Lv, H., Long-Term Satellite Observations of Microcystin Concentrations in Lake Taihu during Cyanobacterial Bloom Periods. *Environ Sci Technol* **2015,** *49*, (11), 6448-6456.
(47) Eiler, A.; Bertilsson, S., Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. *Environ Microbiol* **2004,** *6*, (12), 1228-1243.
(48) Suikkanen, S.; Pulina, S.; Engstrom-Ost, J.; Lehtiniemi, M.; Lehtinen, S.; Brutemark, A., Climate change and eutrophication induced shifts in northern summer plankton communities. *Plos One* **2013,** *8*, (6), e66475.
(49) Ferguson, C. A., Nutrient pollution and the molluscan death record: Use of mollusc shells to diagnose environmental change. *J Coastal Res* **2008,** *24*, (1a), 250-259.
(50) Bouchet, V. M. P.; Alve, E.; Rygg, B.; Telford, R. J., Benthic foraminifera provide a promising tool for ecological quality assessment of marine waters. *Ecol Indic* **2012,** *23*, 66-75.
(51) Yang, J.; Zhang, X.; Xie, Y.; Song, C.; Sun, J.; Zhang, Y.; Giesy, J. P.; Yu, H., Ecogenomics of Zooplankton Community Reveals Ecological Threshold of Ammonia Nitrogen. *Environ Sci Technol* **2017,** *51*, (5), 3057-3064.
(52) Daims, H.; Nielsen, J. L.; Nielsen, P. H.; Schleifer, K. H.; Wagner, M., In situ characterization of Nitrospira-like nitrite oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microb* **2001,** *67*, (11), 5273-5284.
(53) Qin, B. Q.; Xu, P. Z.; Wu, Q. L.; Luo, L. C.; Zhang, Y. L., Environmental issues of Lake Taihu, China. *Hydrobiologia* **2007,** *581*, 3-14.
(54) Otten, T. G.; Xu, H.; Qin, B.; Zhu, G.; Paerl, H. W., Spatiotemporal Patterns and Ecophysiology of Toxigenic Microcystis Blooms in Lake Taihu, China: Implications for Water Quality Management. *Environ Sci Technol* **2012,** *46*, (6), 3480-3488.
(55) Liu, X.; Lu, X. H.; Chen, Y. W., The effects of temperature and nutrient ratios on Microcystis blooms in Lake Taihu, China: An 11-year investigation. *Harmful Algae* **2011,** *10*, (3), 337-343.
(56) De Araujo, A. B.; Hagiwara, A.; Snell, T. W., Effect of unionized ammonia, viscosity and protozoan

542 contamination on reproduction and enzyme activity of the rotifer Brachionus rotundiformis. *Hydrobiologia* **2001,**
543 *446*, 363-368.
544 (57) Arauzo, M.; Valladoli, M., Short-term harmful effects of unionised ammonia on natural populations of Moina
545 micrura and Brachionus rubens in a deep waste treatment pond. *Water Res* **2003,** *37*, (11), 2547-2554.
546 (58) Taylor, J. M.; King, R. S.; Pease, A. A.; Winemiller, K. O., Nonlinear response of stream ecosystem structure
547 to low-level phosphorus enrichment. *Freshwater Biol* **2014,** *59*, (5), 969-984.
548 (59) Johnson, R. K.; Hering, D., Response of taxonomic groups in streams to gradients in resource and habitat
549 characteristics. *J Appl Ecol* **2009,** *46*, (1), 175-186.
550 (60) Wang, J.; Pan, F.; Soininen, J.; Heino, J.; Shen, J., Nutrient enrichment modifies temperature-biodiversity
551 relationships in large-scale field experiments. *Nat Commun* **2016,** *7*, 13960.
552 (61) Blottiere, L.; Jaffar-Bandjee, M.; Jacquet, S.; Millot, A.; Hulot, F. D., Effects of mixing on the pelagic food
553 web in shallow lakes. *Freshwater Biol* **2017,** *62*, (1), 161-177.
554 (62) Wisz, M. S.; Pottier, J.; Kissling, W. D.; Pellissier, L.; Lenoir, J.; Damgaard, C. F.; Dormann, C. F.;
555 Forchhammer, M. C.; Grytnes, J. A.; Guisan, A.; Heikkinen, R. K.; Hoye, T. T.; Kuhn, I.; Luoto, M.; Maiorano,
556 L.; Nilsson, M. C.; Normand, S.; Ockinger, E.; Schmidt, N. M.; Termansen, M.; Timmermann, A.; Wardle, D. A.;
557 Aastrup, P.; Svenning, J. C., The role of biotic interactions in shaping distributions and realised assemblages of
558 species: implications for species distribution modelling. *Biol Rev* **2013,** *88*, (1), 15-30.
559 (63) Tikhonov, G.; Abrego, N.; Dunson, D.; Ovaskainen, O., Using joint species distribution models for evaluating
560 how species-to-species associations depend on the environmental context. *Methods Ecol Evol* **2017,** *8*, (4), 443-
561 452.
562 (64) Van Allen, B. G.; Rasmussen, N. L.; Dibble, C. J.; Clay, P. A.; Rudolf, V. H. W., Top predators determine how
563 biodiversity is partitioned across time and space. *Ecol Lett* **2017,** *20*, (8), 1004-1013.
564 (65) Gonzalez, A. M. M.; Dalsgaard, B.; Olesen, J. M., Centrality measures and the importance of generalist
565 species in pollination networks. *Ecol Complex* **2010,** *7*, (1), 36-43.
566 (66) Jiang, Y.; Xu, H. L.; Hu, X. Z.; Zhu, M. Z.; Al-Rasheid, K. A. S.; Warren, A., An approach to analyzing spatial
567 patterns of planktonic ciliate communities for monitoring water quality in Jiaozhou Bay, northern China. *Mar*
568 *Pollut Bull* **2011,** *62*, (2), 227-235.
569 (67) Zhang, X. Q.; Recknagel, F.; Chen, Q. W.; Cao, H. Q.; Li, R. N., Spatially-explicit modelling and forecasting
570 of cyanobacteria growth in Lake Taihu by evolutionary computation. *Ecol Model* **2015,** *306*, 216-225.
571 (68) Capo, E.; Debroas, D.; Arnaud, F.; Guillemot, T.; Bichet, V.; Millet, L.; Gauthier, E.; Massa, C.; Develle, A.
572 L.; Pignol, C.; Lejzerowicz, F.; Domaizon, I., Long-term dynamics in microbial eukaryotes communities: a
573 palaeolimnological view based on sedimentary DNA. *Mol Ecol* **2016,** *25*, (23), 5925-5943.
574 (69) Pandey, L. K.; Bergey, E. A.; Lyu, J.; Park, J.; Choi, S.; Lee, H.; Depuydt, S.; Oh, Y. T.; Lee, S. M.; Han, T.,
575 The use of diatoms in ecotoxicology and bioassessment: Insights, advances and challenges. *Water Res* **2017,** *118*,
576 39-58.
577 (70) Van Dam, H.; Mertens, A.; Sinkeldam, J., A coded checklist and ecological indicator values of freshwater
578 diatoms from The Netherlands. *Netherland Journal of Aquatic Ecology* **1994,** *28*, (1), 117-133.
579

580 # TABLES

Table 1. Distance-based linear model (distLM) results of bacterial, protistan and metazoan community structures against 22 environmental variables in the full analysis (9999 permutations). Proportion of variation explained (Prop. (%)) and cumulative proportion of variation explained (Cumul. (%)) are given. Environmental variables not significantly correlated with community structure (P<0.05) are not shown.

Community	Marginal tests			Forward selection sequential tests		
	Variables	Pseudo-F	Prop. (%)	Pseudo-F	Prop. (%)	Cumul. (%)
Bacteria	pH	6.85	7.54	18.48	18.03	18.03
	NO_3^-	4.88	5.49	8.42	7.55	25.58
	NH_4^+	12.78	13.21	8.26	6.81	32.39
	Ni	3.34	3.82	8.59	6.48	38.87
	COD	4.14	4.70	7.40	5.18	44.04
	TN	4.65	5.25	6.28	4.12	48.16
	Cd	9.79	10.44	8.92	5.32	53.48
	TP	4.69	5.29	6.10	3.42	56.90
	Mn	3.67	4.18	6.71	3.50	60.40
Protist	NH_4^+	8.50	9.49	8.50	9.49	9.49
	BOD	7.65	8.63	8.92	9.08	18.58
	NO_3^-	7.06	8.02	9.90	9.06	27.64
	PerC	6.90	7.85	7.53	6.37	34.01
	TP	5.51	6.37	7.71	6.00	40.01
	Cu	5.52	6.38	6.89	4.99	45.00
	DO	7.23	8.20	8.48	5.59	50.59
	Cd	7.90	8.89	10.11	5.94	56.53
	NO_2^-	6.08	6.98	11.04	5.71	62.24
Metazoa	NH_4^+	5.07	6.49	5.07	8.49	6.49
	NO_2^-	3.12	4.10	4.48	5.48	13.97
	Cd	4.69	6.03	6.23	5.11	19.08
	TN	3.98	5.64	4.28	4.66	23.74
	Cr	2.76	3.64	3.37	3.55	27.29
	PerC	2.28	2.72	2.43	2.51	29.80
	TP	3.26	4.27	3.13	3.13	32.93
	Drug	2.40	3.18	2.74	2.67	35.61

581

582

Table 2. Based on multivariate linear regression models (MLR) to predict nutrient status (as dependent variable, y) of rivers using the relative abundance of indicative OTUs data (% indicative OTUs, as independent variable, x) in training data sets. Combine, integration of bacteria, protistan and metazoan communities; Adj-R^2, the adjust R^2.

Predictor variables	Predictor formula	Adj-R^2	F
% indicative OTUs (Bacteria)	y= 1.61-98.33*% Proteobacteria	0.41	16.41
% indicative OTUs (Protist)	y= 3.91+27.86*% Stramenopiles-30.93*% Ciliophora-28.36*% Chlorophyta-68.52*% Cryptophyta	0.60	12.94
% indicative OTUs (Metazoa)	y= 0.12-6.68*% Arthropoda	0.43	7.13
% indicative OTUs (Combine)	y= 3.45-48.46*% Proteobacteria+39.33*% Stramenopiles-41.02*% Ciliophora-7.84*% Arthropoda-358.39*% Bacteroidetes-10.24*% Ochrophyta	0.75	12.63

583
584
585

FIGURE CAPTIONS

Figure 1. Assignment of eDNA metabarcoding sequences recovered from rivers in this study. (**a**) Line graph representing the number of each taxonomic rank in different communities, resolved to the highest taxonomic resolution for each OTUs respectively (**b**) Taxonomic phylogenetic tree, built using 'tree of life' (ToL) metabarcoding, and bar graphs showing the family numbers per phylum (only families with >3 OTUs are shown).

Figure 2. The results of principle component analysis (PCA) on twenty-two environmental variables (**a**) and the linear relationship between the first principle components (PC1) and six nutrient-related parameters (**b**). The bubble size represents the scores of the PC1 in each samples, and the blue line points to the direction of the increase for a given variable, only the strongly correlation (absolute r>0.5) between PC1 and variables are shown (**a**); the dash lines are the 95% confidence interval (CI) fitting value (**b**).

Figure 3. Distribution of dominant taxonomic OTUs at phylum or class level (**a**) and non-metric multidimensional scaling (nMDS) analysis of communities' structure (**b**). The position in the triangle indicates the relative abundance of each phyla or class taxa in bacterial, protistan and metazoan community (**a**) across three nutrient levels, and the size of the circle represented the relative abundance of each taxon. Significant differences based on Bray-Curtis (bacteria and protist) and Jaccard (metazoa) dissimilarity matrices in the communities' structure are found among the three levels (**b**). TTL, QHR and TYR are the abbreviations of the tributaries of Tai Lake, the Qinhuai River and the tributaries of Yangtze River, respectively.

608

609 **Figure 4.** Responses of Shannon index to nutrient (surrogated by PC1, **a-c**), and ecological
610 interaction network between bacterial, protistan and metazoan communities in each nutrient
611 level (**d**). Non-linear polynomial regression included 95% CI (the dash lines) in bacteria
612 (quadratic), protistan (quadratic) and metazoan (cubic) communities. The correlations between
613 each OTUs were generated by SparCC with 100 bootstraps to assign *P*-values. Only when the
614 absolute r> 0.7 and a 'two tailed' *P* value < 0.01, the nodes and edges in network were remained.

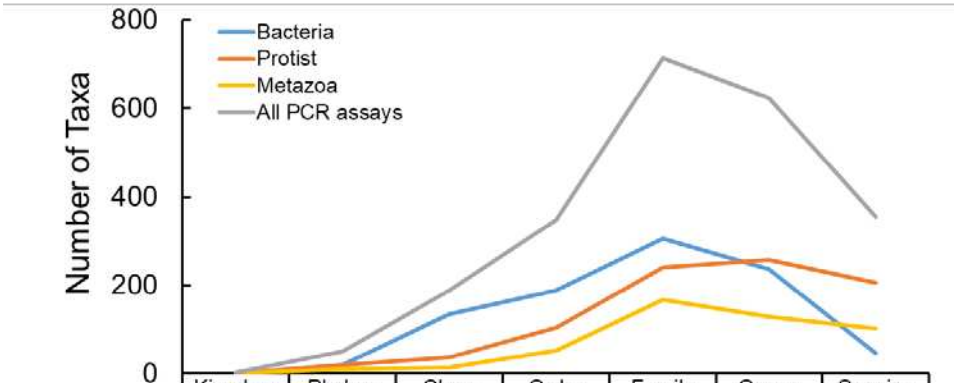
615

616 **Figure 5.** The relationship between nutrient (surrogated by PC1) and the relative abundance of
617 indicative OTUs, the dashed lines are the 95% CI fitting value, only significant correlation
618 ($P<0.05$) are shown (**a**). Comparison between the PC1 predicted value given by the indicative
619 OTUs in test samples and the PC1 actual value derived from environmental variables, the red
620 diagonal lines represent the ratio (1:1) between predicted and measured values (**b**).

621

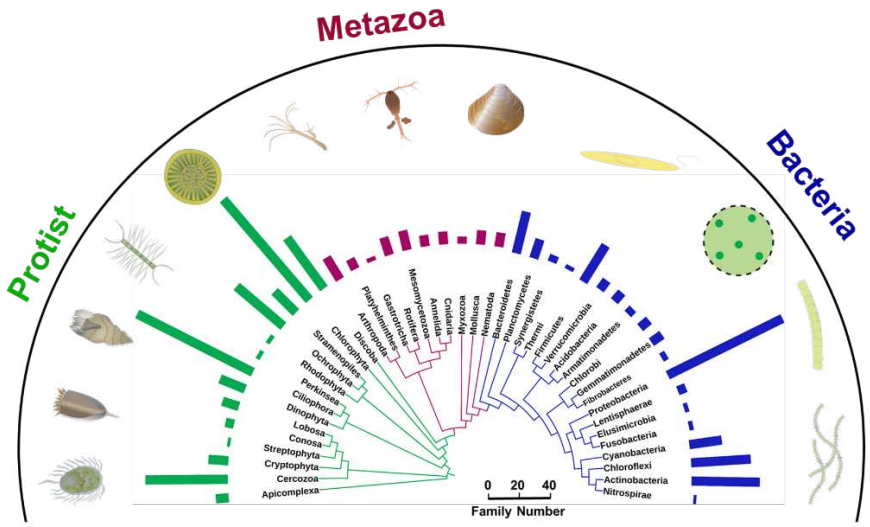
Figure 1

a



	Kingdom	Phylum	Class	Order	Family	Genus	Species
Bacteria	1	20	135	189	306	236	47
Protist	1	20	38	105	241	258	205
Metazoa	1	11	15	53	167	129	103
All PCR assays	2	51	188	347	714	623	355

b

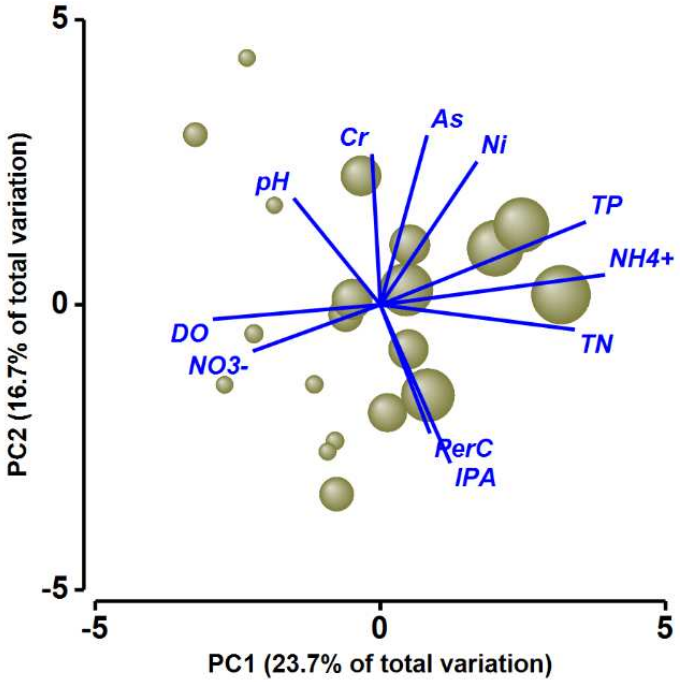


622

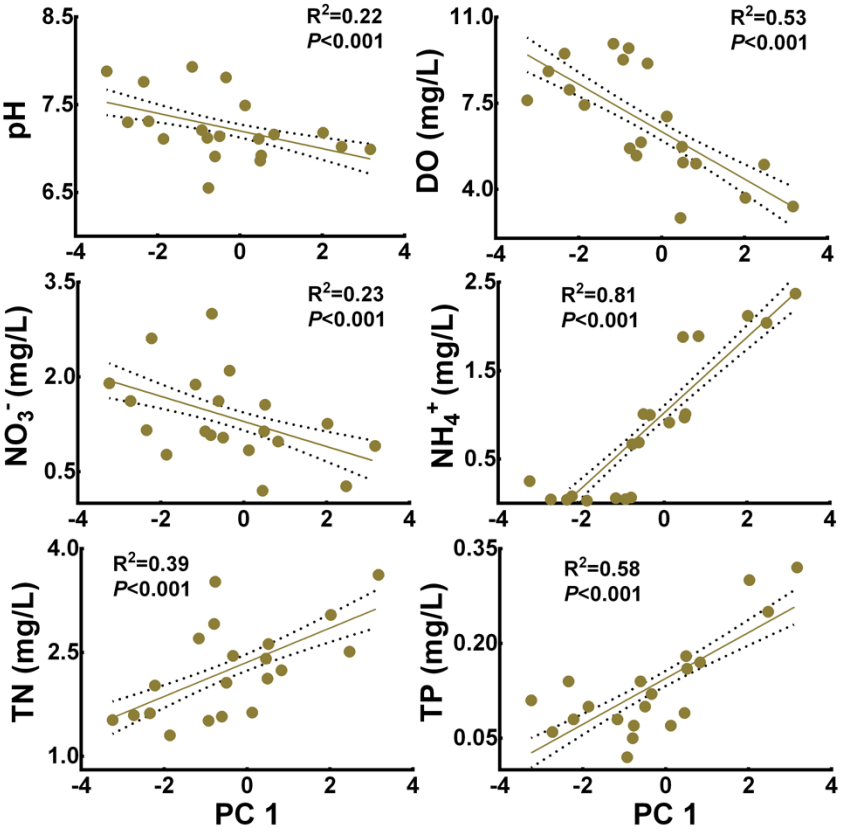
623

Figure 2

a



b

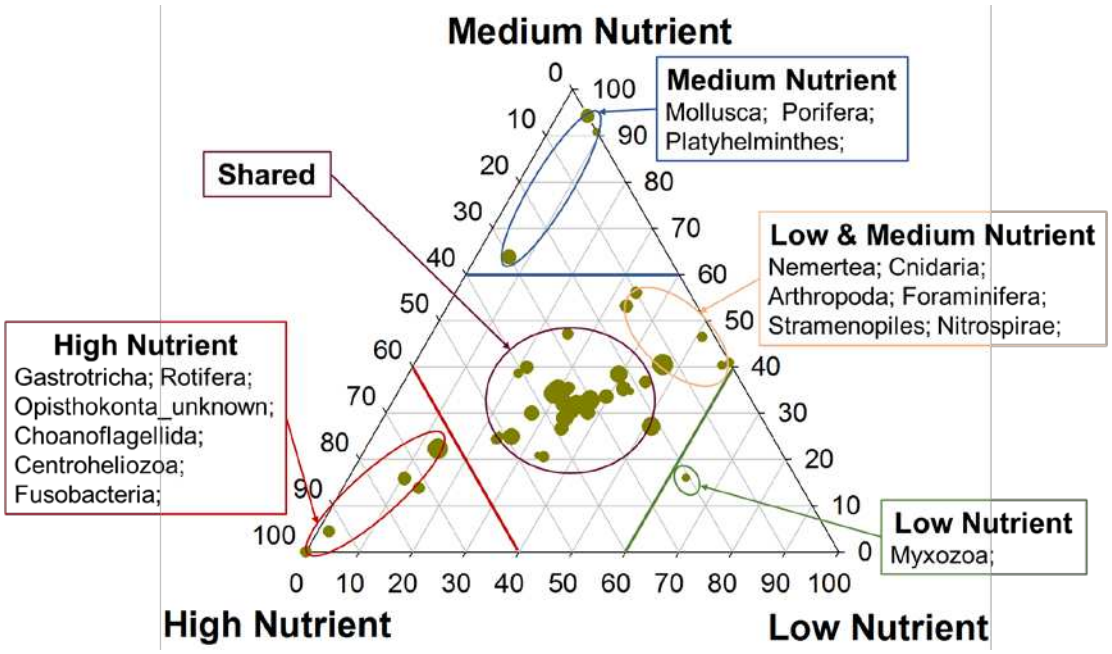


624

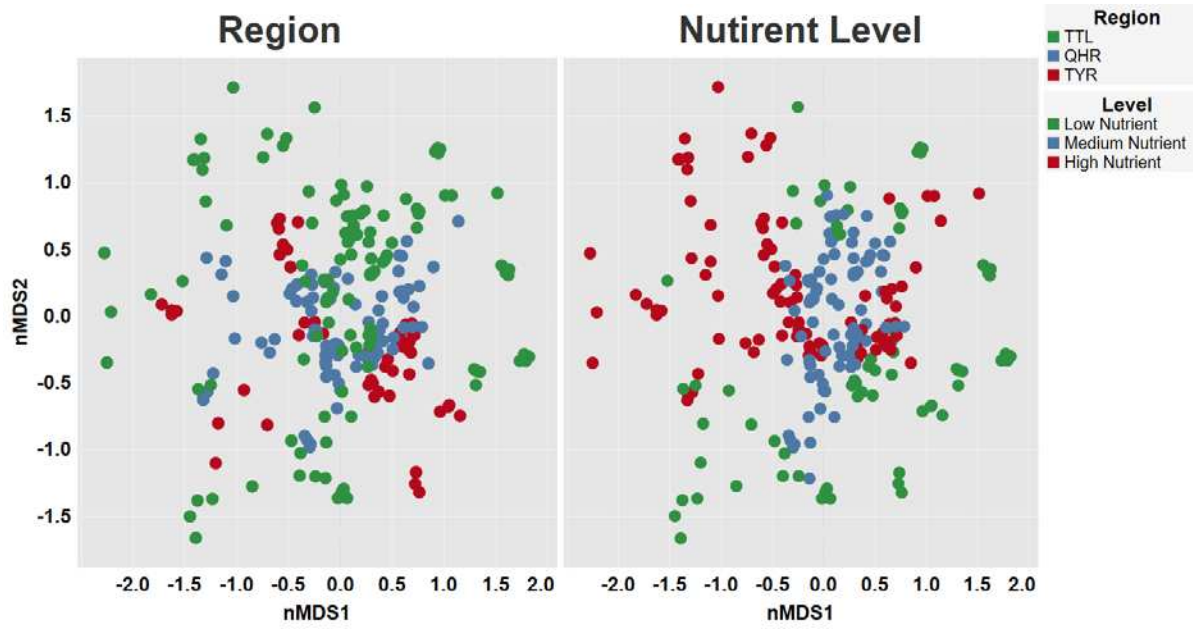
625

Figure 3

a

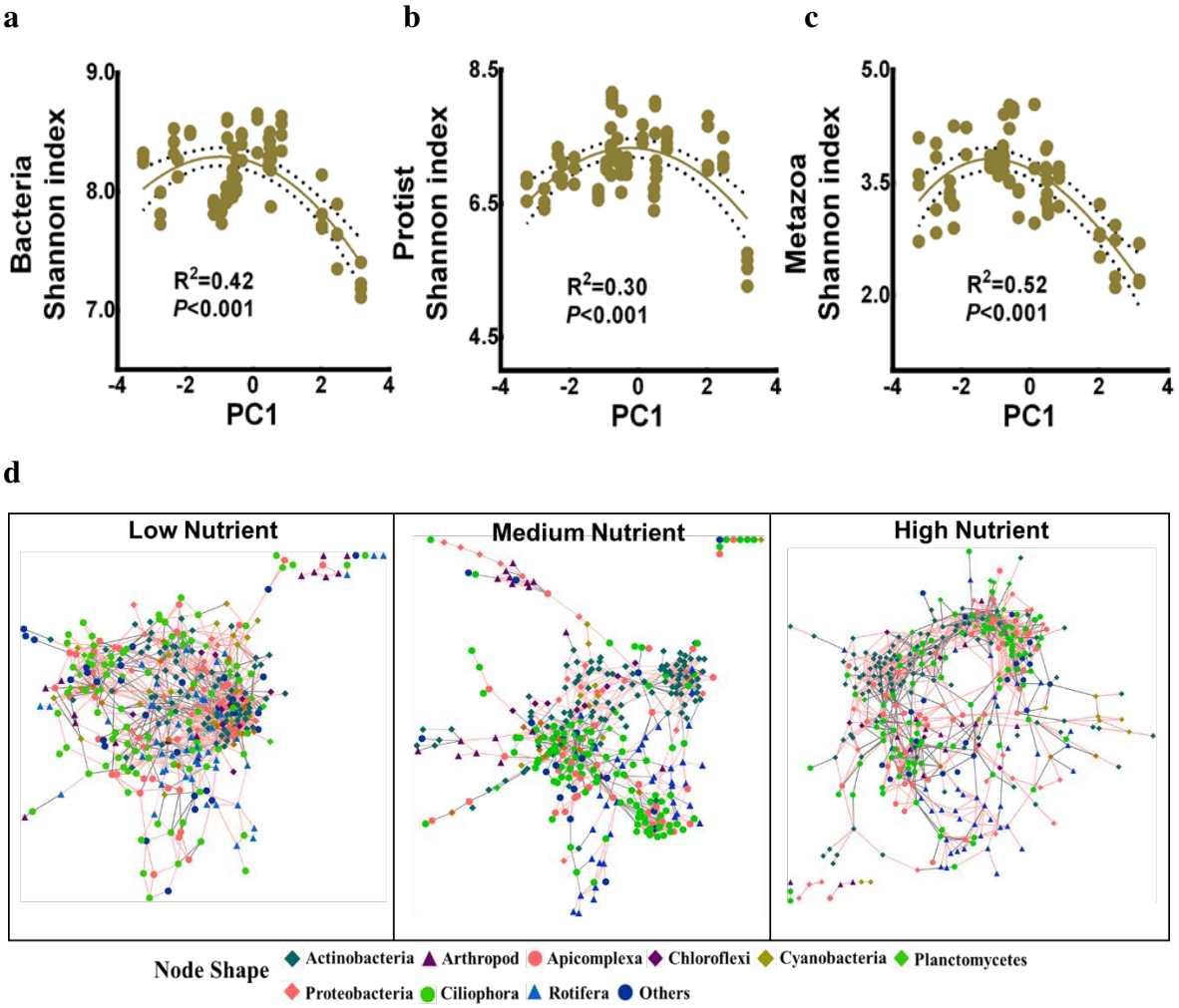


b



626

Figure 4

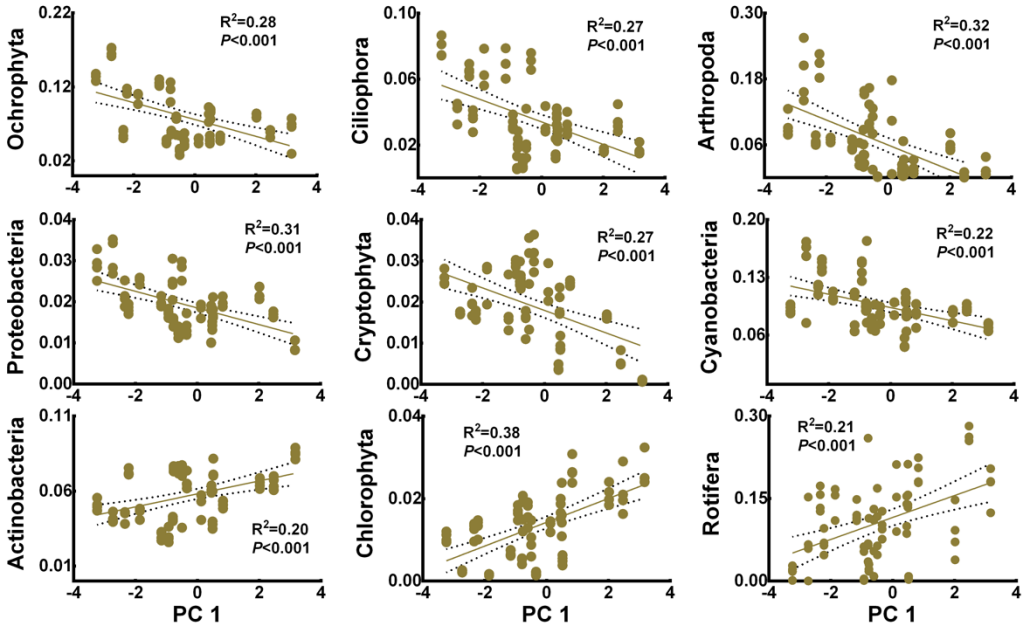


627

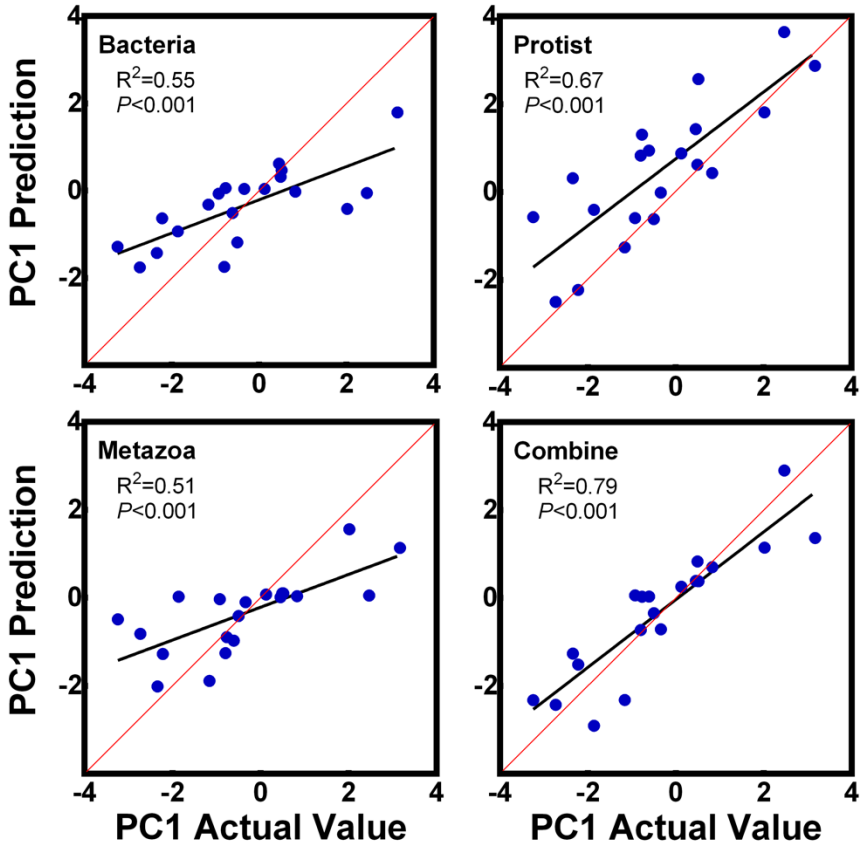
628

Figure 5

a



b



629