Molecular biogeography of planktonic and benthic diatoms in the Yangtze River

Citation for published version:
Wang, J, Liu, Q, Zhao, X, Borthwick, A, Liu, Y, Chen, Q & Ni, J 2019, 'Molecular biogeography of planktonic and benthic diatoms in the Yangtze River', *Microbiome*, vol. 7, 153. https://doi.org/10.1186/s40168-019-0771-x

Digital Object Identifier (DOI):
10.1186/s40168-019-0771-x

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Microbiome

Molecular Biogeography of Planktonic and Benthic Diatoms in the Yangtze River

Jiawen Wang[1†], Qingxiang Liu[1†], Xianfu Zhao[2], Alistair G. L. Borthwick[3], Yuxin Liu[1], Qian Chen[1,4], Jinren Ni[4,5*]

Author Affiliations:

[1]The Key Laboratory of Water and Sediment Sciences, Ministry of Education, College of Environmental Sciences and Engineering, Peking University, Beijing, 100871, China

[2]Institute of Hydroecology, Ministry of Water Resources, Chinese Academy of Sciences, Wuhan 430079, China

[3]Institute of Infrastructure and Environment, School of Engineering, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JL, UK

[4]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining, 810016, China

[5]Beijing Innovation Center for Engineering Science and Advanced Technology, Peking University, Beijing, 100871, P. R. China

[†]Equal contributors

[*]Correspondence: Jinren Ni, College of Environmental Sciences and Engineering,

Peking University, Beijing, 100871, China. E-mail: jinrenni@pku.edu.cn

Abstract

Background: Diatoms are of great significance to primary productivity in oceans, yet little is known about their biogeographic distribution in oligotrophic rivers.

Results: With the help of high-throughput sequencing analysis of 279 samples from the Yangtze River, we provided the first integral biogeographic pattern of planktonic and benthic diatoms over a 6030 km continuum along the world's third largest river. Our study revealed spatial dissimilarity of diatoms under varying landforms, including plateau, mountain, foothill, basin, foothill-mountain, and plain regions, from the river source to the estuary. Environmental drivers of diatom communities were interpreted in terms of photosynthetically active radiation, temperature, channel slope and nutrients, and human interference. Typical benthic diatoms, such as *Pinnularia*, *Paralia* and *Aulacoseira*, experienced considerable reduction in relative abundance downstream of the Three Gorges Dam and the Xiluodu Dam, two of the world's largest dams.

Conclusions: Our study revealed that benthic diatoms are of particular significance in characterizing motile species in the riverine environment, thus providing insight into diatom biogeography and biogeochemical cycles in large river ecosystems.

Keywords: diatoms, biogeography, water, sediment, landform, environmental driver, human interference, Yangtze River

Background

Diatoms play a particularly important role in the biogeochemical cycle [1] of primary elements such as carbon, nitrogen, phosphorus, and silica, contributing about 20-25 % of global primary production [2]. Diatoms are ubiquitous and diverse species of single-celled, eukaryotic, photosynthetic microorganisms on earth [3], and are often the dominant primary producers in marine and freshwater ecosystems [4]. Therefore, diatoms in such ecosystems may be remarkably dissimilar either in phylogenetic composition or biogeographic distribution [5,6]. Freshwater bodies typically consist of lentic (particularly lakes and wetlands) and lotic waters (including streams and rivers), which are often dominated respectively by planktonic algae and benthic species [7].

Accurate identification of diatoms depends on the reliability of the analysis methods used. Morphological analysis requires extensive taxonomic expertise, and may exhibit shortcomings in characterizing specific diatoms in rivers [8]. With the development of High-Throughput Sequencing (HTS) technology, DNA metabarcoding has become a rapid, accurate and reliable method for diatom detection [9]. Various DNA barcoding studies have been successfully conducted, based mainly on different maker genes, including COI [10], ITS [11], and 18S rDNA [9,12]. Malviya et al. [13] were the first to estimate the diversity and distribution of planktonic diatoms based on the V9 region of eukaryotic 18S rDNA. Recently, the V4 region of 18S rDNA was proposed for diatom barcoding in a study of diatoms in river and deltaic systems [9,12].

Comparing the numerous studies of diatoms and eutrophication in oceans [14,15] and lakes [16,17] to date, it is clear that the present understanding of diatoms is relatively poor for lotic and oligotrophic rivers [7]. In fact, previous reports on the dynamics of riverine diatoms have mostly focused on tributaries, small rivers, reaches, stations, and estuaries [18–20]. Many studies examined the diversity and composition of planktonic [19,20] or benthic diatoms [18,21,22] based on morphological identification. For example, Centis et al. [20] investigated planktonic diatoms dominated by physical constraints at two stations of the River Adige, Italy. Liu et al. [18] investigated the community structure of benthic diatoms in the Dong River, one of the three main tributaries of the Pearl River, China. Although Kireta et al. [23] observed that both planktonic and periphytic diatoms could be used as bio-indicators of river conditions, little is known about the distinction between planktonic and benthic diatoms regarding their spatiotemporal distributions.

Biogeography studies aim to reveal the spatial and temporal distribution of biodiversity, and provide insight into the mechanisms that generate and sustain diversity [24]. Spatial dispersal and environmental selection processes are regarded as essential drivers for the biogeographical pattern of a given bacterial community [25]. The former promotes movement of species and their establishment at a new location, whereas the latter alters the abundance and composition of species, according to the ability to survive and reproduce under local environmental conditions. A similar explanation has been proved to apply to the biogeographical pattern of planktonic or benthic diatom communities in small rivers using morphological analysis [26–28]. However, it remains unclear how the integrated spatiotemporal distributions of planktonic and benthic diatom communities are shaped by spatial dispersal and environmental selection processes in large rivers subject to complex natural and anthropogenic impacts.

To close the above gap, we implemented large-scaled synchronous monitoring of diatom communities at 62 hydrologic stations over a 6030 km continuum of the Yangtze River in China. Consequently, we successfully provided the first molecular biogeographic pattern of both planktonic and benthic diatoms in this, the largest river in Asia (Fig.1). Meanwhile, environmental drivers of diatom communities were interpreted in terms of photosynthetic active radiation, temperature, channel slope, and nutrient conditions under varying landforms.

Results

Our study generated 8,602,620 V4 18S rDNA reads, in total, from 279 samples. All sequencing reads were classified into 3947 operational taxonomic units (OTUs) at a 97% similarity threshold, with 3144 OTUs well matching 454 diatom species in our reference database. Rarefaction curves (Additional file 1: Figure S1) together with high values of Good’s coverage ranging from 0.9854 to 0.9992 illustrated that OTUs obtained by the current sequencing depth gave a reasonable representation of the diatom communities. The phylogeny tree, constructed by representative OTUs (accounting for > 90% sequence in all samples) and reference sequences (Additional file 1: Figure S2), further confirmed the accuracy of taxonomic assignment.

Molecular insight into the diatom communities

Molecular barcoding based on high-throughput sequencing (HTS) provided a detailed diatom directory for the whole Yangtze River at different taxonomy levels, i.e., 4 classes, 37 orders, 60 families, and 152 genera.

HTS is of particular use in detecting nano-sized diatoms (2-20 μm) with several HTS studies for the Yangtze River confirming the presence of *Fragilaria perminuta*, *Achnanthidium minutissimum*, *Achnanthidium saprophilum*, *Amphora pediculus*, *Fistulifera saprophila*, *Mayamaea permitis*, *Sellaphora seminulum*, *Encyonema minutum*, *Fragilaria famelica*, *Fragilaria rumpens*, *Gomphonema pumilum*, *Staurosirella pinnata*, *Planothidium frequentissimum*, *Craticula buderi*, and *Craticula molestiformis*.

Molecular characterization of riverine diatoms revealed that specific diatoms had a statistically significant response to concentration of dissolved carbon dioxide ($p\text{CO}_2$, ppm) in the Yangtze River (Additional file 1: Figure S3). Planktonic diatoms *Asterionella formosa*, *Diatoma vulgare*, *Lindavia viaradiosa*, *Gomphonema pumilum*, and *Thalassiosira nordenskioeldii* were strongly associated with dissolved carbon dioxide (Spearman $r > 0.3$, $p < 0.05$), whereas benthic diatoms *Asterionella formosa*, *Encyonema prostratum*, *Eucocconeis laevis*, *Fistulifera saprophila*, and *Nitzschia sigmoidea* were highly correlated with dissolved carbon dioxide.

Moreover, fifteen genera (*Cyclotella*, *Achnanthes*, *Asterionella*, *Bacillaria*, *Cocconeis*, *Cymbella*, *Diatoma*, *Fragilaria*, *Gomphonema*, *Gyrosigma*, *Melosira*, *Navicula*, *Nitzschia*, *Rhizosolenia*, *Synedra*) were simultaneously detected in water-autumn samples by HTS and traditional morphological studies. Relative abundances of *Melosira* (19.90%), *Achnanthes* (18.55%), *Synedra* (11.84%), and *Navicula* (4.27%) identified by the morphological study were higher than those by molecular analysis (i.e., 15.12%, < 0.0001%, 0.01%, and 0.65%, respectively). *Cyclotella* dominated the riverine diatoms, as confirmed by both light microscopy (33.90% of total diatoms) and HTS (30.23% of total reads). Additional genera of considerable abundance were discovered in the Yangtze River solely by means of the HTS study, including the genus *Skeletonema* (relative abundance of 10.30%), *Paralia* (8.24%), and *Pinnularia* (4.98%).

Alpha and beta diversity of diatom communities

Six types of environmental samples were taken along the Yangtze River, including water and sediment samples from the river source region (i.e. water-plateau (12 samples) and sediment-plateau (12 samples)) and those from the mainstream in the non-plateau area (i.e. water-spring (38 samples), water-autumn (46 samples), sediment-spring (87 samples), and sediment-autumn (84 samples)). Planktonic diatoms exhibited the highest alpha-diversity (Chao1 and Shannon indices) and benthic diatoms the lowest richness (Chao1) in the plateau (Additional file 1: Figure S4). In the non-plateau area, no significant differences were observed in the alpha richness and diversity of diatom communities in the four sample types.

Non-metric multidimensional scaling (NMDS) analysis of the compositional dissimilarities between diatom communities of all samples demonstrated not only a clear spatial differentiation in diatoms between the plateau and the main body of the Yangtze, but also a division between planktonic and benthic groups (Additional file 1: Figure S5). Seasonal difference in planktonic diatoms is found much more significant than in benthic diatoms, as further confirmed by an Analysis of Similarity (ANOSIM) test (Additional file 1: Figure S6). Moreover, one-way Analysis of Variance (one-way ANOVA) indicated that more planktonic diatoms (42.75±13.98 % relative abundance, primarily belonging to *Cyclotella*, *Stephanodiscus* and *Skeletonema*) than benthic diatoms (16.58±5.06 % relative abundance, primarily belonging to *Pinnularia* and *Stephanodiscus*) exhibited significant seasonal sensitivity (Additional file 1: Figure S7).

Biogeographic patterns of diatom communities

Indicator species analysis revealed differences in diatom taxonomic composition among the six types of environmental samples. The number of indicator diatom species in the river ranged from 6 (sediment-spring) to 41 (water-plateau) (Additional file 1: Table S1). Diatom communities in the plateau region were quite different from those in the main body of the Yangtze River, as evidenced by the higher percentage of top indicator species in water-plateau and sediment-plateau samples (Additional file 1: Figure S8). The average relative abundance of indicator species in the source area exceeded 40%, and planktonic indicator species contributed more reads than benthic indicator species. Furthermore, a number of indicator species belonging to *Tabellariales* and *Hemiaulales* occurred in water-plateau and sediment-plateau samples, respectively.

Species composition and spatial distribution of planktonic and benthic diatoms at different taxonomy levels along the whole Yangtze River were then studied. At class level, *Coscinodiscophyceae* dominated planktonic diatoms whereas *Bacillariophyceae* prevailed in the mainstream sediment. In the plateau region, diatom communities displayed similar structures dominated by *Bacillariophyceae* in both water and sediment. In the lower reach, the greater abundance of planktonic *Bacillariophyceae* demonstrated a seasonal difference (Additional file 1: Figure S9).

At genus level, the relative abundance contributed by dominant genera (top 20) ranged from 55.6 to 83.6 % (Fig. 2**a**). Dominant benthic genera, *Pinnularia*, *Cyclotella*, and *Navicula*, represented 14.2, 13.4, and 13.1 % of total sequences in sediment-autumn samples. Benthic *Navicula* was adapted to wide ranges of temperature and elevation, as evident by its dominance in sediment-plateau (17.1 %) and sediment-spring (13.2 %) samples. *Cymbella* was prevalent in water-plateau (17.7 %) and sediment-plateau (9.27 %) samples.

At species level, diatom composition in terms of ecological guilds showed spatial dissimilarity in water and sediment (Additional file 1: Figure S10). Diatoms were divided into four ecological guilds according to their biological traits, including low-profile, high-profile, motile and planktic guilds in terms of different responses to nutrients and dynamic disturbances [29–31] (Methods: Ecological guilds classification). Benthic diatoms in the motile guild prevailed at most stations along the whole river, whereas those in high-profile and planktic guilds dominated upstream and downstream reaches, respectively. In addition, planktonic diatoms in the planktic guild were predominant at most stations along the Yangtze River.

Over the 6030 km continuum from river source to mouth, landform type plays a significant role in spatial differentiation of both planktonic and benthic diatom communities. Referring to a previous study on landform types in the Yangtze [32], planktonic diatoms represented by *Cymbella*, *Asterionella, Stephanodiscus*, *Melosira*, *Cyclotella* and *Conticribra* and benthic diatoms represented by *Cymbella*, *Navicula*, *Melosira*, *Conticribra, Cyclotella* and *Surirella* were respectively abundant in plateau, mountain, foothill, basin, foothill-mountain, and plain regions (Fig. 2**b**).

Natural and anthropogenic effects on diatom biogeography

Natural effects were estimated, including spatial dispersal and environmental selection processes for the biogeographical pattern of diatom communities. Significant distance-decay in diatom similarity was observed according to geographical distance (Additional file 1: Figure S11), with a greater gradient for water (slope = −0.042) than for sediment (slope = −0.038) using least squares linear regression. The Partial Mantel test demonstrated that both geographical and environmental distances played important roles in constraining diatom composition and distribution (Additional file 1: Table S2). Variation partitioning of diatom composition showed that a greater percentage (14.6-21.2%) could be explained by a purely environmental component than that (3.4-6.0%) of the total variation by a purely spatial component (Additional file 1: Figure S12), and a minor portion (0.4-5.4%) explained by spatial structured environmental heterogeneity, leaving the majority of the total variation (68.7-79.0%) inexplicable. Although environmental differentiation seems more important than spatial dispersion in shaping a diatom community, neither can fully explain the total variation in diatom composition. Common environmental parameters, including photosynthetically active radiation, temperature, channel slope, and nutrients, affect diatom communities undergoing spatial dispersal.

Photosynthetically active radiation (PAR, 400-700 nm), important as solar energy, is utilized by diatoms to synthesize biomass through photosynthesis [33]. Regardless of weak changes in PAR between spring and autumn in the Yangtze River basin [34], the spatial composition and distribution of both planktonic and benthic diatom communities varied in four PAR regions (Fig. 3**a-b**, Methods: Photosynthetically Active Radiation (PAR) divisions). LefSe analysis revealed that both planktonic and benthic diatom species preferred living under different conditions of light intensity. For example, the *Caloneis*, *Cymbella*, *Fistulifera* and *Fragilaria* genera preferred very-high PAR zones, the *Papiliocellulus* genus favored medium PAR regions, and *Conticribra* and *Cyclotella* lived as biomarkers in low PAR habitats. Planktonic

Cymatopleura and *Navicula,* and benthic *Asterionella*, *Biddulphia*, *Diatoma* and *Encyonema* genera were adapted to high PAR conditions. Moreover, water temperature is a key environmental factor in structuring diatom community assemblages through its influence on diatom size and growth rate [35] in the Yangtze River (Additional file 1: Table S3). Although the richness of planktonic diatoms seems to fluctuate with PAR, the richness of benthic diatoms tends to rise with increasing temperature (Fig. 3**c**).

Stream power, often simply characterized by the river channel slope or the product of channel slope and flow discharge (except in plateau regions) [36], is another important factor altering the spatial distribution of diatoms (data sourced from Chen et al. 2001 [37]). During the wet season (autumn), the higher flow discharge weakens the correlation between planktonic community similarity and channel slope, although a stronger correlation between benthic community similarity and channel slope is maintained due to higher mobility of the streambed (Fig. 4**a-b**). In general, the varying channel slope along the Yangtze River could be simplified into three stages: steep slope in mountainous reaches (station 1~2); moderate slope in upper reaches (station 3~14); and mild slope in middle-lower reaches (station 15~24). In view of their relative abundance, planktonic diatoms are characterised by *Psammothidium*, *Nitzschia* and *Cymbella* for steep-slope environments, *Papiliocellulus* for moderate slope, and *Mayamaea*; *Pinnularia* and *Surirella* for mild slope environments. Benthic diatoms are represented by *Cocconeis*, *Entomoneis* and *Melosira* for steep slope environments, *Fallacia*, *Psammothidium* and *Skeletonema* for moderate slopes, and *Actinocyclus*, *Aulacoseira* and *Conticribra* for mild slopes (Additional file 1: Figure S13). Furthermore, slope effects on diatoms may be identified in terms of ecological guilds. Regardless of the diatoms in an unspecified ecological guild, planktonic diatoms are dominant in the planktic guild. Interestingly, species in motile guilds constitute the main component of benthic diatoms in the whole lotic river (Fig. 4**c**).

Nutrient conditions, represented by the ratio of total nitrogen to total phosphorus (TN:TP), are of great significance to diatom growth. Noting that Redfield [38] proposed that N:P = 16:1 provides an optimal criterion for algal production in the oceans, we then investigated the alpha-diversity of diatom community response to TN:TP in the Yangtze River. The alpha-diversity of planktonic diatoms could be partly interpreted in terms of annual-averaged dissolved TN:TP (data range from 2005 to 2014) in water-spring (Adj $R^2 = 0.54$) and water-autumn (Adj $R^2 = 0.41$) samples (Additional file 1: Figure S14**a-b**); however, benthic diatoms demonstrated an even weaker response to TN:TP (monitored data) (Additional file 1: Figure S14**c-d**). For relatively oligotrophic rivers like the Yangtze River, it appears that TN is more important than TP as limiting nutrients to planktonic and benthic community variability (Additional file 1: Table S3). Furthermore, the measured TN:TP provided a better explanation for community variations in benthic rather than planktonic diatoms, which could be attributed to the difference in utilization of phosphorus forms between benthic and planktonic diatoms (Additional file 1: Table S3).

Nevertheless, the aforementioned environmental factors are subject to change due to human interference. For the Yangtze River, one of the most prominent impacts on its ecology arises from the construction and operation of large dams. In the present study, a one-way ANOVA analysis revealed that certain OTUs exhibited significant variations in distribution immediately upstream and downstream of the Three Gorges Dam ($p < 0.01$) and Xiluodu Dam ($p < 0.05$). Moreover, sudden drops in relative abundance of OTUs belonging to specific benthic species (such as *Pinnularia*, *Paralia* and *Aulacoseira*) occurred downstream of the dams (Additional file 1: Figure S15).

Discussion

The molecular method provided a comprehensive explanation of the molecular biogeographical pattern of diatom communities in a large river, the Yangtze. This is for three reasons. First, plentiful information on species composition of diatom communities is provided by the HTS method (which benefits from a continuously updated NCBI database). Second, HTS characterizes diatoms based on genetic information taken from the V4 region of 18S rDNA, enabling evolutionary relationships to be distinguished, and then elucidated with confidence. Third, the HTS method greatly facilitates the identification of small-sized genera that otherwise might be either regarded as unclassified species or overlooked due to their minute size and difficulty of detection. Species-specific diatoms of different cell size make different contributions to primary productivity and bloom biomass [39], whereas nanoplanktonic diatoms may play a vital role in carbon export [40].

This study offers insights into specific planktonic and benthic diatoms, which are of great importance to the carbon cycle. Here, the diatoms exhibited a significant positive correlation with dissolved carbon dioxide [41]. In the river, diatoms take up dissolved inorganic carbon and convert it into dissolved or particulate organic matter (primary production), and then reduce carbon dioxide in the surface layer. Conversely, re-mineralization of organic matter by bacteria and zooplankton can increase carbon dioxide. Therefore, diatoms could maintain the equilibrium between carbon dioxide in the surface layer with that in the atmosphere for the river system [42].

The HTS method has revealed the existence of certain special species in the Yangtze River. A poorly known freshwater species *Skeletonema potamos*, which has high growth potential when silica is not limiting [43], was found to be a dominant species of genus *Skeletonema* in the Yangtze. Although *Paralia* has previously been reported to be abundant in marine, coastal, and river estuary areas [44,45], it has been found in considerable abundance at several stations in the mid-lower Yangtze; this may be attributed to saltwater intrusion aggravated by operation of the Three Gorges Dam during the autumn [46]. *Pinnularia* has been shown to adapt to different habitats with varying pH and nutrients, including moist soils, sediments, springs, and oceans [47]. In the river system, a large number of *Pinnularia* species (such as *Pinnularia viridiformis*, *P. subanglica*, *P. brebissonii*, *P. biceps*, etc.) were also detected, living in weakly alkaline water and sediment environments.

Taxonomic compositions of riverine diatoms in the Yangtze River are markedly different from those in lakes [48] and oceans [13] because of the distinct differences in salinity and hydraulic conditions. For example, *Actinocyclus spp.*, *Aulacoseira spp.*, *Cyclotella spp.*, *Fragilaria spp.*, and *Synedra spp.* predominated in diatom communities in Lake Kasumigaura, Japan [48]. Moreover, *Chaetocero*s, followed by *Fragilariopsis*, *Thalassiosira*, and *Corethron* were reported to be the most abundant genera in the global oceans [13]. In marine ecosystems, diatoms are abundant in nutrient-rich coastal zones particularly at high latitudes. In riverine ecosystems, the spatial variation of most species can be explained in terms of landform type. For example, diatoms of glacier origin from the Qinghai-Tibetan plateau were significantly different to those in the main body of the Yangtze River (Fig. 2, Additional file 1: Figure S8). In order of abundance, the most common riverine diatoms in water and sediment of the Yangtze River were *Cyclotella*, *Navicula*, *Pinnularia*, *Stephanodiscus,* and *Cymbella* genera.

With the help of synchronous water and sediment sampling along the Yangtze River, it is possible to compare the biogeography of planktonic and benthic diatoms. On the one hand, although the diversity of planktonic and benthic diatoms was very similar in non-plateau areas, their community composition was remarkably different. Planktonic diatoms were dominated by *Cyclotella* (average relative abundance: 17.66%), *Stephanodiscus* (12.81%), *Pinnularia* (7.53%), *Paralia* (7.34%), and

Skeletonema (4.90%); whereas benthic diatoms were dominated by *Navicula* (13.12%), *Cyclotella* (10.33%), *Pinnularia* (10.12%), *Surirella* (7.10%) and *Stephanodiscus* (6.53%). On the other hand, seasonal differences were more evident in planktonic diatoms than benthic diatoms. A one-way ANOVA analysis confirmed that the water temperature of the Yangtze River was significantly different between spring and autumn ($p< 0.01$), with the water temperature (average 21 °C) in autumn being more conducive than that (average 11 °C) in spring for planktonic diatom growth [49]. Moreover, seasonal fluctuations in water discharge appear to affect the community structure of planktonic diatoms, owing to the introduction of diatom species from the upstream freshwater source and to different hydrologic processes in spring and autumn [50]. Meanwhile, the weak seasonal difference of diatom communities in sediment may be ascribed to the preponderance of benthic diatoms (Additional file 1: Figure S6) that weakly responded to seasonal changes and reached a state of relative equilibrium through long-term sediment erosion and deposition processes [32].

To study the spatial distribution of planktonic and benthic diatoms, the Yangtze River can be divided into six reaches according to the surrounding landform type [32]. In both water and sediment samples, differentially abundant diatoms were detected for the six landform types across the sample site. In the Yangtze basin, landform changes are associated with variations along the river in local soil type, soil nutrients, light, altitude and temperature. The plateau reach is located at the Qinghai-Tibet Plateau which has the lowest temperature (<11 degree centigrade) and the highest altitude (>3500m) and light resource (PAR >32 $mol^{-2}m^{-1}d$). The basin reach is in Sichuan, which receives the lowest light resource. The plain reach near the river estuary is located in the Middle-Lower Yangtze Plain at the lowest altitude (< 10m) and the highest temperature and nutrient levels (nitrogen and phosphorus). Therefore, landform type, which determines the essential light, temperature, nutrients, and other environmental factors for diatom growth, had a significant impact on the spatial distribution of diatom community.

The environmental-based selection process driving the biogeographic pattern of diatoms is influenced by PAR, temperature, channel slope, and nutrient level. Benthic diatoms are typical consequences of riverine environmental selection in the Yangtze River. We now consider the influence factors in turn. First, it has been reported that sufficient PAR drives the growth and production of diatoms [51], but excess PAR can affect various cellular processes and reduce the growth or viability of diatoms [52]. In the present study, specific diatoms were identified as having adapted to different levels of light intensity (Fig. 3), indicating that diatoms possess diverse light-regulatory mechanisms and adaptive responses [53]. Although light and temperature are the essential resource for diatom growth, planktonic and benthic diatoms exhibit different preferences for PAR and water temperature. In the surface oceans, planktonic diatoms have been shown to be replaced by small phytoplankton, causing decreased primary production and carbon export, due to global warming [54]. In the

Yangtze River however, benthic diatoms are better biological indicators of temperature change. Second, as a primary driver of stream power that shapes the spatial distribution of diatoms, the channel slope not only affects competitive and succession processes among species but also alters nutritional utilization strategies and hence the production and growth of diatoms [55], leading to diatoms adapting to different flow conditions [56]. The different spatial distribution of ecological guilds suggests that benthic diatoms are appropriate markers of environmental heterogeneity along the river; however, the nutrient effect is not always significant on the dominant species in the motile guild under nutrient-oligotrophic conditions (Fig. 4). Third, the nutrient level characterized by annual-averaged dissolved TN:TP only partly explains the alpha-diversity of planktonic diatoms, whereas the monitored TN:TP provides a slightly better explanation for community variations of benthic diatoms. One reason could be related to species-specific responses to nitrogen and phosphorus in the production and growth of diatoms (Additional file 1: Figure S14 and Table S3). For example, *Nitzschia palea* [57] is phosphorus limited, but *Chaetoceros calcitrans* [58] is efficient in nitrogen assimilation. Another reason might be that algae utilize nitrogen and phosphorus in different forms; for example, ammonia is preferred to nitrate [59] and inorganic phosphate is preferred to organic phosphorus [60]. In the Yangtze River, the weak relationship between alpha-diversity of benthic diatoms and TN:TP reflects the lower influence of nutrients in this oligotrophic river.

Damming in rivers is a typical anthropogenic perturbation which could profoundly modify material fluxes and biogeochemical cycles of downstream [61,62]. Kunz et al [61] reported that sediment, carbon, nitrogen, and phosphorus were trapped by the reservoir immediately upstream of the Itezhi-Tezhi Dam, increasing the N:P ratio downstream of the dam. High flow downstream of dams has caused severe erosion of the riverbed and led to coarsening of bed materials [62]. Changes in water level also affect the light intensity and temperature to benthic diatoms. In short, a large dam disrupts the hydraulic gradient, nutrient conditions, light availability, and temperature in rivers, resulting in local changes to the environment in which benthic diatoms thrive.

Interaction between planktonic and benthic diatoms is a major concern for large river ecosystems. In the Yangtze river, such interactions have several consequences. First, the community composition of planktonic community was significantly correlated with that of benthic diatoms in paired water and sediment samples (in spring: Spearman $r = 0.3556$, $p = 0.001$; in autumn: Spearman $r = 0.1902$, $p = 0.006$). Typical benthic diatoms (e.g., *Nitzschia* and *Navicula*) were found in high abundance in the water column. The local interactions could cause benthic and planktonic habitats to become coupled through migration of algal cells, meaning that phytoplankton can be derived from benthic diatoms, and sinking planktonic algae can become benthic algae [63]. Second, the richness of planktonic diatoms appears to fluctuate with PAR, whereas the richness of benthic diatoms tends to change with temperature. Nutrient level (TN:TP) had different effects on the diversity and variation of planktonic and benthic diatoms. These phenomena further explain the local interactions in terms of light, temperature and nutrients competition [64]. Third, given the “River continuum concept” [65] and “Continuous discontinuity concept” [66], the dominance of benthic or planktonic algae changes with the natural riverine gradient (e.g. channel slope), and interactions are invariably interrupted by anthropogenic disturbances such as dams, reservoirs, and nutrient-rich discharges. Other factors affecting interactions, such as river velocity, turbulent diffusion, algal sinking and grazing remain to be further investigated [67].

Conclusions

This study provided the first molecular biogeographic patterns of both planktonic and benthic diatoms over a continuum of 6030 km in the Yangtze River. Significant seasonal differences in planktonic diatom communities were observed over the whole length of the mainstream. Diatoms in water and sediment exhibited differential abundance according to landform type, such as plateau, mountain, foothill, basin, foothill-mountain, and plain regions, along the river. Environmental selection overcame spatial dispersal in controlling the community structure of planktonic and benthic diatoms. Typical environmental drivers were therefore interpreted in terms of photosynthetically active radiation, hydraulic slope, nutrients, and human activities (i.e. dams). Our study revealed that benthic diatoms represented by motile species in ecological guilds are typical consequences of environmental selection in a lotic-oligotrophic river; this work enables better understanding of the specific contributions of benthic diatoms in biogeochemical cycles in world's large river ecosystems.

Methods

Sample collection

The Yangtze River is the longest river situated wholly in Asia and the third longest in the world, with a drainage basin of 1.8 million km^2. The river is over 6300 km long, has its source in the Qinghai-Tibet Plateau, and flows eastwards into the East China Sea near Shanghai. Over its length, the Yangtze River experiences great changes in landform type and hydrological regime, and supports more than 588 million people [68]. To investigate the seasonal and spatial distributions of diatom communities along the whole river, water and sediment samples were synchronously (i.e. within one week) collected for planktonic and benthic diatom identification at 50 national monitoring stations along the mainstream and six major tributaries of the Yangtze in March (spring) and October (autumn) 2014. In July 2017, we collected water and sediment samples at 12 sites in the river source. In total, 96 water samples and 183 sediment samples were obtained. Except for a very few samples missed due to restrictions of steep terrain and rapid flow as described in a previous study [32], up to four parallel samples were collected in most cases. Further details of the sampling sites are listed in Table S1. At each sampling site, 10L of well-mixed water was collected and then immediately filtered onto 0.22 μm polycarbonate membranes (Millipore, USA) within 24 h. Filtered membranes and sediment samples were stored in the laboratory at -80°C until further analysis took place.

DNA extraction, PCR amplification and sequencing

DNA was extracted in triplicate using the FastDNA® SPIN Kit for Soil (MP Biomedicals, USA) following the manufacturer’s instructions. The triplicate DNA extracts were mixed together for later PCR amplification. Amplification of the V4 region of the 18S rDNA was performed by polymerase chain reaction (PCR) (initial denaturation at 94 °C for 2 min, then 32 cycles of denaturation at 94 °C for 45 s, annealing at 50°C for 45 s, elongation at 72 °C for 60 s, and final elongation at 72 °C for 10 min, 10°C until halted by user) using DIV4for (5’-GCGGTAATTCCAGCTCCAATAG-3’) and DIV4rev3 primers (5’-CTCTGACAATGGAATACGAATA-3’) [12]. PCR mixtures (20 μL volume) prepared in triplicate contained 2 μl of 10×Buffer, 2 μl of 2.5 mM dNTPs, 0.8 μl of each primer (5 μM), 0.2 μl of rTaq Polymerase, 0.2 μl of BSA, and 1 μl of 10 ng DNA sample. Amplicons were purified using the AxyPrep DNA Gel Extraction Kit (Axygen Bioscience, Union City, CA, U.S.) according to the manufacturer’s instructions and quantified using QuantiFluorTM -ST (Promega, U.S.). Afterwards, purified amplicons were pooled in equimolar amounts and sequenced on Illumina MiSeq 2 × 250 PE platform (Majorbio Company, Shanghai, China).

Three negative control samples were used to monitor any contamination during the molecular workflow, negative filtration, DNA extraction, and PCR controls, however, no quantifiable DNA was detected for further analysis.

Bioinformatics analysis

Sequences of diatom 18S rDNA were quality-filtered using QIIME [69] as follows: (i) minimum sequence length of 300 bp, and minimum threshold quality score of Q20; (ii) maximum mismatches of 2 for matching the primer; any reads with ambiguous bases were removed; and (iii) merged pair-ended sequences that overlapped longer than 10 bp into a single sequence. UCHIME was used to remove chimeric sequences and UPARSE was used to cluster Operational Taxonomic Units (OTUs) with 97% similarity cutoff [70].

We built a reference database of 18S rRNA reads composed of 4573 unique diatom sequences. First, we extracted all diatom sequences of 18S rRNA reads from Genbank (http://www.ncbi.nlm.nih.gov/). Second, short reads (less than 100 nucleotides) were refused access to the reference database, and redundant reads were eliminated by cd-hit to increase the taxonomy identification accuracy. Third, sequence alignment was performed by Mafft (ver 7.310) [71], then the sequences were analyzed to construct an approximately-maximum-likelihood phylogenetic tree using FastTree (ver 2.1.10) [72], and any incorrect reads discarded. Finally, a total of 4573 unique sequences were retained in our reference database.

To identify taxonomically the OTUs obtained in this study against known diatom species, the BLASTN [73] program was applied to align clean 18S rRNA reads to the corrected diatom database. Those OTUs with the best BLAST hit scores, not only an e-value $\leq 10^{-5}$ but also identity $\geq 80\%$ with respect to the reference sequence were firstly selected. Then the selected OTUs were checked by means of the phylogenetic tree, and only OTUs with correct taxonomical assignment were retained for further analysis. Clean reads were further assigned to known diatom species based on our reference database.

To estimate the community structure for each site, the Mothur program [74] was used to normalize all data sets with respect to the least-well-represented data set (11049 sequences). Alpha diversity indices (chao1, Shannon and Goods coverage) were calculated using QIIME.

Statistical analysis

Diatom species that characterize each sample group were identified with Indicator Species Analysis using labdsv and indval packages in R software [75]. Indicator values were calculated based on the relative frequency and relative average abundance of a given species in six types of environmental samples. Species with indicator value $\geqslant 0.3$ and p-value $\leqslant 0.01$ were defined as indicator species at Class, Order, Family, and Genus levels. Nonmetric multidimensional scaling (NMDS) was performed to visualize the dissimilarity of different samples based on Bray–Curtis similarity matrices. Analysis of Similarity (ANOSIM) was conducted to test the significance of differences among a priori sampling groups based on environmental parameters. NMDS and ANOSIM statistics were carried out using the vegan package in R. The linear discriminant analysis effect size (LEfSe) [76] was used to discover high-dimensional biomarker and explain taxa difference at different environment conditions. One-way analysis of variance (one-way ANOVA) was carried out to test significance of group differences.

Distance-decay patterns of diatom community similarity were described by considering geographical distance from the site location to river mouth among sample sites. Mantel tests were used to examine the Spearman's rank correlation between geographical distance and diatom community similarity using Bray-Curtis distance matrices with 999 permutations in R. The distance of each sampling site was calculated using ArcGIS V10.3 software. The rate of distance-decay of diatom communities was calculated as the slope of ordinary least-squares regression line fitted to the relationship between geographic distance and community similarity.

A set of spatial variables was generated through the use of principal coordinates of neighbor matrices (PCNM) analysis based on the longitude and latitude coordinates of each sampling site [77]. The function 'envfit' was run with 999 permutations to select significant variables ($P < 0.05$). Significance testing was then assessed using the 'permutest' function based on 999 permutations in R. Partial canonical correspondence analysis (pCCA) was performed to decompose the total variation in diatom community into a pure environmental component, a pure spatial component, a spatially structured environmental component, and residual variation.

Ecological guilds classification

Based on their ecological characteristics, diatom species are classified into four ecological guilds (low profile, high profile, motile and planktic guilds) [29–31], which are expected to respond in different ways to nutrient conditions and physical disturbances. A low-profile guild is defined as having high reproduction rate, low nutrient and light availability, and slow-moving diatoms. A high-profile guild possesses characteristics of high resource availability and low disturbance. A motile ecological guild has the ability to move fast and choose the best microhabitat in a given circumstance. A planktic guild adapts to lentic environments and resists sedimentation. We extended these guilds by adding supplemented classifications used in other studies [78,79].

Photosynthetically Active Radiation (PAR) divisions

Solar radiation is the most important source of energy required for plant growth. Solar radiation with wavelengths (400-700 nm), called Photosynthetically Active Radiation (PAR) is able to convert light energy into biomass [33] through photosynthesis by plants and algae. Monteith reported the linear correlation between net primary production (NPP) and PAR absorbed by green foliage [80]. Xudong et al. [34] also suggested that the spatial distribution of annually-averaged PAR is complex and inhomogeneous across China, using data for the period 1961−2007. Thus, we define four zones of PAR intensity in different regions across the Yangtze River basin as follows:

I. Very high, PAR > 32 mol m^{-2} d^{-1};

II. High, 26 < PAR ≤ 32 mol m^{-2} d^{-1};

III. Medium, 23 < PAR ≤ 26 mol m^{-2} d^{-1};

IV. Low, PAR ≤ 23 mol m^{-2} d^{-1}.

Xudong et al. found slowly changing rates of spring-averaged and autumn-averaged PAR occurred in the Yangtze River basin.

Additional files

Additional file 1:

Figure S1. Rarefaction curves of diatom richness per sample at cutoff levels of 3%.

Figure S2. Phylogenetic distribution of reference sequence and abundance of OTUs. The color range displays class level taxonomy information on the node branch. The outmost heatmap indicates relative abundance of OTUs in six sample types: WS, water-spring; SS, sediment-spring; WA, water-autumn; SA, sediment-autumn; WP, water-plateau; SP, sediment-plateau.

Figure S3. Spearman relationships for relative abundance of diatoms and dissolved CO_2 (pCO_2, ppm). Strong correlations between dissolved CO_2 and the planktonic and benthic diatoms are marked in red..

Figure S4. Alpha diversity index per diatom community obtained for each type of sample.

Figure S5. Nonmetric multidimensional scaling (NMDS) diagram of compositional (Bray-Curtis) dissimilarities between diatom communities for all samples in the Yangtze River..

Figure S6. ANOSIM statistics concerning differences in diatom communities within and between sample types.

Figure S7. Diatom genera exhibiting significant seasonal differences in water (**a**) and sediment samples (**b**).

Figure S8. Proportion of indicator diatoms in each sample type obtained using indicator taxa analysis at Class (**a**), Order (**b**), Family (**c**), and Genus (**d**) levels.

Figure S9. Biogeographical distribution of diatom species at class level throughout the mainstream of the Yangtze River for: (**a**) water-spring, (**b**) water-autumn, (**c**) sediment-spring, and (**d**) sediment-autumn samples. For comparison, water-plateau samples are displayed in (**a**) and (**b**); sediment-plateau samples are displayed in (**c**) and (**d**).

Figure S10. Biogeographical distribution of ecological guilds at species level throughout the mainstream of the Yangtze River for all samples. For comparison, water-plateau samples are displayed in (**a**) and (**b**)**;** sediment-plateau samples are displayed in (**c**) and (**d**).

Figure S11. Relationships between diatom community (Bray–Curtis) similarity and geographic distance for water (**a**) and sediment (**b**) samples. Values of Mantel

Spearman correlation (r) and correlation significance (P) are provided. Red lines indicate the ordinary least squares linear regression across all samples.

Figure S12. Variation in community composition explained by environmental, spatial, and spatially structured environmental component.

Figure S13. LEfSe cladogram of planktonic (**a**) and benthic (**b**) diatom communities for the three channel slope regions. Diatom taxa with a mean relative abundance of ⩾ 0.1% in all samples, assigned to kingdom (innermost), phylum, class, order, family, and genus (outermost), are used to determine taxa or clades most likely to explain differences between channel slope regions. Differentially abundant taxa (biomarkers) are colored according to the slope regions in which they are most abundant; i.e. red, green and blue circles stand for biomarkers in steep, moderate and mild slope regions.

Figure S14. Relationships between Shannon diversity and TN:TP for water-spring (**a**), water-autumn (**b**), sediment-spring (**c**), and sediment-autumn (**d**) samples. Annually-averaged TN:TP data during 2005−2014 are used for water samples, whereas monitored data in spring and autumn 2014 are used for sediment samples. Distance relationship of TN:TP for sampling sites along the mainstream is shown in (**e**).

Figure S15. Significant differences in abundance of benthic diatoms upstream and downstream of Xiluodu Dam (**a**) and Three Gorges Dam (**b**).

Table S1. Numbers of Indicator species and Top Indicator species across sample sites.

Table S2. Partial Mantel test for Spearman correlations between community similarity and geographic and environmental distances.

Table S3. Effects of selected environmental and spatial factors on diatom community composition.

Abbreviations

OTUs: operational taxonomic units; HTS: high-throughput sequencing; NMDS: Nonmetric multidimensional scaling; ANOSIM: Analysis of Similarity; one-way ANOVA: One-way Analysis of Variance; LefSe: linear discriminant analysis effect size; PAR: Photosynthetically Active Radiation; TN:TP: ratio of total nitrogen to total phosphate

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Availability of data and material

Complete datasets supporting the findings of this article are in the NCBI Sequence Read Archive (SRA) database (Accession Number: SRP153344).

Competing interests

The authors declare no competing financial interests.

Funding

Financial support from National Natural Science Foundation of China (No. 91647211 and 51721006) is appreciated.

Author’s contributions

J.R.N. designed the research. J.W.W. and Q.X.L performed research. J.W.W., Q.X.L., J.R.N. and A.G.L.B. wrote the paper. X.Z. and Y.L. contributed new ideas and information. All of the authors contributed to interpretation of the findings.

Acknowledgments

Support from BIC-ESAT and Majorbio Company (Shanghai, China) is gratefully acknowledged.

References

1. Armbrust EV. The life of diatoms in the world’s oceans. Nature. 2009;459:185–92.
2. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science. 1998/07/10. 1998;281:237–40.
3. Sims PA, Mann DG, Medlin LK. Evolution of the diatoms: insights from

fossil, biological and molecular data. Phycologia. 2006;45:361–402.

4. Mann DG, Droop SJM. Biodiversity, biogeography and conservation of diatoms. Hydrobiologia. 1996;336:19–32.

5. Oeding S, Taffs KH. Developing a regional diatom index for assessment and monitoring of freshwater streams in sub-tropical Australia. Ecol. Indic. 2017;80:135–46.

6. Cermeño P, Falkowski PG. Controls on Diatom Biogeography in the ocean. Science. 2009;325:1539–41.

7. Orlando Necchi JR. River Algae. Necchi JR O, editor. Cham: Springer International Publishing; 2016.

8. Kelly MG. Use of the trophic diatom index to monitor eutrophication in rivers. Water Res. 1998;32:236–42.

9. Zimmermann J, Glöckner G, Jahn R, Enke N, Gemeinholzer B. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Mol. Ecol. Resour. 2015;15:526–42.

10. Evans KM, Wortley AH, Mann DG. An Assessment of Potential Diatom "Barcode" Genes (cox1, rbcL, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in Sellaphora (Bacillariophyta). Protist. 2007;158:349–64.

11. Moniz MBJ, Kaczmarska I. Barcoding of Diatoms: Nuclear Encoded ITS Revisited. Protist. 2010;161:7–34.

12. Visco JA, 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;49:7597–605.

13. Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, et al. Insights into global diatom distribution and diversity in the world's ocean. Proc. Natl. Acad. Sci. U. S. A. 2016;113:E1516–25.

14. Boyd PW, Watson AJ, Law CS, Abraham ER, Trull T, Murdoch R, et al. A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. Nature. 2000;407:695–702.

15. Tsuda A, Takeda S, Saito H, Nishioka J, Nojiri Y, Kudo I, et al. A mesoscale iron enrichment in the western subarctic pacific induces a large centric diatom bloom. Science. 2003;300:958–61.

16. Heinsalu A, Alliksaar T, Leeben A, Nõges T. Sediment diatom assemblages and composition of pore-water dissolved organic matter reflect recent eutrophication history of Lake Peipsi (Estonia/Russia). Hydrobiologia. 2007;584:133–43.

17. Kitner M, Poulíčková A. Littoral diatoms as indicators for the eutrophication of shallow lakes. Hydrobiologia. 2003;506–509:519–24.

18. Liu J, Soininen J, Han B-P, Declerck SAJ. Effects of connectivity, dispersal directionality and functional traits on the metacommunity structure of river benthic diatoms. J. Biogeogr. 2013;40:2238–48.

19. Kheiri S, Solak CN, Edlund MB, Spaulding S, Nejadsattari T, Asri Y, et al.

Biodiversity of diatoms in the Karaj River in the Central Alborz, Iran. Diatom Res. 2018;33:355–80.

20. Centis B, Tolotti M, Salmaso N. Structure of the diatom community of the river Adige (North-Eastern Italy) along a hydrological gradient. Hydrobiologia. 2010;639:37–42.

21. Rimet F. Benthic diatom assemblages and their correspondence with ecoregional classifications: Case study of rivers in North-Eastern France. Hydrobiologia. 2009;636:137–51.

22. Dalu T, Wasserman RJ, Magoro ML, Mwedzi T, Froneman PW, Weyl OLF. Variation partitioning of benthic diatom community matrices: Effects of multiple variables on benthic diatom communities in an Austral temperate river system. Sci. Total Environ. 2017;601–602:73–82.

23. Kireta AR, Reavie ED, Sgro G V, Angradi TR, Bolgrien DW, Hill BH, et al. Planktonic and periphytic diatoms as indicators of stress on great rivers of the United States: Testing water quality and disturbance models. Ecol. Indic. 2012;13:222–31.

24. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. 2006;4:102–12.

25. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat. Rev.

Microbiol. 2012;10:1–10.

26. Keck F, Franc A, Kahlert M. Disentangling the processes driving the biogeography of freshwater diatoms: A multiscale approach. J. Biogeogr. 2018;45:1582–92.

27. Heino J, Bini LM, Karjalainen SM, Mykrä H, Soininen J, Vieira LCG, et al. Geographical patterns of micro-organismal community structure: Are diatoms ubiquitously distributed across boreal streams? Oikos. 2010;119:129–37.

28. Soininen J. Determinants of benthic diatom community structure in boreal streams: The role of environmental and spatial factors at different scales. Int. Rev. Hydrobiol. 2004;89:139–50.

29. Passy SI. Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. Aquat. Bot. 2007;86:171–8.

30. Rimet F, Bouchez A. Life-forms, cell-sizes and ecological guilds of diatoms in European rivers. Knowl. Manag. Aquat. Ecosyst. 2012;01.

31. B-Béres V, Lukács Á, Török P, Kókai Z, Novák Z, T-Krasznai E, et al. Combined eco-morphological functional groups are reliable indicators of colonisation processes of benthic diatom assemblages in a lowland stream. Ecol. Indic. 2016;64:31–8.

32. Liu T, Zhang AN, Wang J, Liu S, Jiang X, Dang C, et al. Integrated biogeography of planktonic and sedimentary bacterial communities in the Yangtze

River. Microbiome. 2018;6:1–14.

33. Jacovides CP, Timvios FS, Papaioannou G, Asimakopoulos DN, Theofilou CM. Ratio of PAR to broadband solar radiation measured in Cyprus. Agric. For. Meteorol. 2004;121:135–40.

34. Zhu X, He H, Liu M, Yu G, Sun X, Gao Y. Spatio-temporal variation of photosynthetically active radiation in China in recent 50 years. J. Geogr. Sci. 2010;20:803–17.

35. Montagnes DJS, Franklin DJ. Effect of temperature on diatom volume, growth rate, and carbon and nitrogen content: Reconsidering some paradigms. Limnol Ocean. 2001;46:2008–18.

36. Yu G-A an, Brierley G, Huang HQ, Wang Z, Blue B, Ma Y. An environmental gradient of vegetative controls upon channel planform in the source region of the Yangtze and Yellow Rivers. Catena. 2014;119:143–53.

37. Chen Z, Li J, Shen H, Zhanghua W. Yangtze River of China: Historical analysis of discharge variability and sediment flux. Geomorphology. 2001;41:77–91.

38. Redfield AC, Ketchum BH, Richards FA. The influence of organisms on the composition of seawater. Sea. New York, USA: Wiley-Interscience; 1963.

39. Assmy P, Smetacek V, Montresor M, Klaas C, Henjes J, Strass VH, et al. Thick-shelled, grazer-protected diatoms decouple ocean carbon and silicon cycles in the iron-limited Antarctic Circumpolar Current. Proc. Natl. Acad. Sci. U. S. A. 2013;110:20633–8.

40. Leblanc K, Quéguiner B, Diaz F, Cornet V, Michel-Rodriguez M, Durrieu de Madron X, et al. Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. Nat. Commun. 2018;9:953.

41. Tréguer P, Bowler C, Moriceau B, Dutkiewicz S, Gehlen M, Aumont O, et al. Influence of diatom diversity on the ocean biological carbon pump. Nat. Geosci. 2018;11:27–37.

42. Smetacek V, Ban S, Burns G, Castel J, Chaudron Y, Christou R, et al. Diatoms and the Ocean Carbon Cycle. Protist. 1999;150:25–32.

43. Duleba M, Ector L, Horváth Z, Kiss KT, Molnár LF, Pohner Z, et al. Biogeography and Phylogenetic Position of a Warm-stenotherm Centric Diatom, Skeletonema potamos (C.I. Weber) Hasle and its Long-term Dynamics in the River Danube. Protist. 2014;165:715–29.

44. Sims PA, Crawford RM. Earliest records of Ellerbeckia and Paralia from Cretaceous deposits: a description of three species, two of which are new. Diatom Res. 2017;32:1–9.

45. Zong Y. Implications of paralia sulcata abundance in scottish isolation basins. Diatom Res. 1997;12:125–50.

46. Jianjun Z, Man Z. Effect of dams on the regime of the mid-lower Yangtze River runoff and countermeasures.pdf. J. Lake Sci. 2018;30:1471–88.

47. Souffreau C, Verbruggen H, Wolfe AP, Vanormelingen P, Siver PA, Cox EJ, et al. A time-calibrated multi-gene phylogeny of the diatom genus Pinnularia.

Mol. Phylogenet. Evol. 2011;61:866–79.

48. Arai H, Fukushima T. Impacts of long-term increase in silicon concentration on diatom blooms in Lake Kasumigaura, Japan. Ann. Limnol. 2014;50:335–46.

49. Cohn SA, Farrell JF, Munro JD, Ragland RL, Weitzell RE, Wibisono BL. The effect of temperature and mixed species composition on diatom motility and adhesion. Diatom Res. 2003;18:225–43.

50. Hauptmann AL, Markussen TN, Stibal M, Olsen NS, Elberling B, Bælum J, et al. Upstream Freshwater and Terrestrial Sources Are Differentially Reflected in the Bacterial Community Structure along a Small Arctic River and Its Estuary. Front. Microbiol. 2016;7:1474.

51. Domingues RB, Anselmo TP, Barbosa AB, Sommer U, Galvão HM. Light as a driver of phytoplankton growth and production in the freshwater tidal zone of a turbid estuary. Estuar. Coast. Shelf Sci. 2011;91:526–35.

52. Buma AGJ, Engelen AH, Gieskes WWC. Wavelength-dependent induction of thymine dimers and growth rate reduction in the marine diatom Cyclotella sp. exposed to ultraviolet radiation. Mar. Ecol. Prog. Ser. 1997;153:91–7.

53. Depauw FA, Rogato A, D’Alcalá MR, Falciatore A. Exploring the molecular basis of responses to light in marine diatoms. J. Exp. Bot. 2012;63:1575–91.

54. Bopp L, Aumont O, Cadule P, Alvain S, Gehlen M. Response of diatoms

distribution to global warming and potential implications : A global model study. Geophys. Res. Lett. 2005;32:1–4.

55. Peters F, Arin L, Marrasé C, Berdalet E, Sala MM. Effects of small-scale turbulence on the growth of two diatoms of different size in a phosphorus-limited medium. J. Mar. Syst. 2006;61:134–48.

56. Mitrovic SM, Chessman BC, Davie A, Avery EL, Ryan N. Development of blooms of Cyclotella meneghiniana and Nitzschia spp. (Bacillariophyceae) in a shallow river and estimation of effective suppression flows. Hydrobiologia. 2008;596:173–85.

57. Han L, Xu B, Qi F, Chen Z. Effect of nitrogen/phosphorus concentration on algal organic matter generation of the diatom Nitzschia palea: Total indicators and spectroscopic characterization. J. Environ. Sci. (China). 2016;47:130–42.

58. Tantanasarit C, Englande AJ, Babel S. Nitrogen, phosphorus and silicon uptake kinetics by marine diatom Chaetoceros calcitrans under high nutrient concentrations. J. Exp. Mar. Bio. Ecol. 2013;446:67–75.

59. Zeng Q, Qin L, Bao L, Li Y, Li X. Critical nutrient thresholds needed to control eutrophication and synergistic interactions between phosphorus and different nitrogen sources. Environ. Sci. Pollut. Res. 2016;23:21008–19.

60. Cotner, Jr JB, Wetzel RG. Uptake of dissolved inorganic and organic phosphorus compounds by phytoplankton and bacterioplankton. Limnol. Oceanogr. 1992;37:232–43.

61. Kunz MJ, Wüest A, Wehrli B, Landert J, Senn DB. Impact of a large tropical reservoir on riverine transport of sediment, carbon, and nutrients to downstream wetlands. Water Resour. Res. 2011;47:W12531.

62. Graf WL. Downstream hydrologic and geomorphic effects of large dams on American rivers. Geomorphology. 2006;79:336–60.

63. Tekwani N, Majdi N, Mialet B, Tornés E, Urrea-Clos G, Buffan-Dubau E, et al. Contribution of epilithic diatoms to benthic-pelagic coupling in a temperate river. Aquat. Microb. Ecol. 2013;69:47–57.

64. Jäger CG, Diehl S. Resource competition across habitat boundaries: asymmetric interactions between benthic and pelagic producers. Ecol. Monogr. 2014;84:287–302.

65. Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. The River Continuum Concept. Can. J. Fish. Aquat. Sci. 1980;37:130–7.

66. Ward JV, Stanford JA. The serial discontinuity concept of lotic ecoystems. In: Fontaine, T.D., Bartell SM, editor. Dynamics of Lotic Ecosystems. Ann Arbor Science Publications. Ann Arbor, Michigan; 1983.

67. Jäger CG, Borchardt D. Longitudinal patterns and response lengths of algae in riverine ecosystems: A model analysis emphasising benthic-pelagic interactions. J. Theor. Biol. 2018;442:66–78.

68. Qu S, Wang L, Lin A, Zhu H, Yuan M. What drives the vegetation restoration in Yangtze River basin, China: Climate change or anthropogenic factors?

Ecol. Indic. 2018;90:438–50.

69. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods. 2010;7:335–6.

70. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods. 2013;10:996–8.

71. Yamada KD, Tomii K, Katoh K. Application of the MAFFT sequence alignment program to large data - Reexamination of the usefulness of chained guide trees. Bioinformatics. 2016;32:3246–51.

72. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490.

73. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.

74. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing MOTHUR: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Env. Microbiol. 2009;75:7537–41.

75. Dufrtne M, Legendre P. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecol. Monogr. 1997;67:345–66.

76. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al.

Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12:R60.

77. Borcard D, Legendre P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbor matrices. Ecol Model. 2002;153:51–68.

78. Berthon V, Bouchez A, Rimet F. Using diatom life-forms and ecological guilds to assess organic pollution and trophic level in rivers: A case study of rivers in south-eastern France. Hydrobiologia. 2011;673:259–71.

79. B-Béres V, Török P, Kókai Z, Lukács Á, T-Krasznai E, Tóthmérész B, et al. Ecological background of diatom functional groups: Comparability of classification systems. Ecol. Indic. 2017;82:183–8.

80. Monteith JL. Solar-Radiation and Productivity in Tropical Ecosystems. J. Appl. Ecol. 1972. p. 747–66.

Figure legends

Figure 1. Map of sampling sites in the Yangtze River covering the actual sinuous channel reach of length 6030 km (equivalent to 1.83 times the 3290 km straight line from start to end sampling sites). The molecular method provides insight into the biogeographic pattern of diatoms along the mainstream of the Yangtze River. Photosynthetically Active Radiation (PAR), temperature, channel slope, and nutrients characterized by ratio of total nitrogen to total phosphate (TN:TP) are used to interpret the biogeographic patterns of diatom communities.

Figure 2. (**a**) Circular visualization of dominant diatoms at genus level in six sample types. Inner circular diagram shows relative abundance of different diatom genus in six sample types. Only the dominant genus with a mean relative abundance of ⩾ 1% in all samples is depicted. The width of ribbons for each diatoms is directly proportional to their relative abundance in each sample type. Similarly, different colored ribbons of different width for each sample type describe the distribution of different genera. (**b**) Representative diatoms genera in different landform types from the river source to mouth along the Yangtze River.

Figure 3. LEfSe cladogram of planktonic (**a**) and benthic (**b**) diatom communities from four PAR regions. Diatom taxa with a mean relative abundance of ⩾ 0.1% in all samples, assigned to kingdom (innermost), phylum, class, order, family, and genus (outermost), are used to determine taxa or clades most likely to explain differences between PAR regions. Differentially abundant taxa (biomarkers) are colored by their

most abundant PAR regions, i.e. red, green, blue and purple circles stand for biomarkers in regions of very high, high, medium and low abundance. Orange and blue circles display the average alpha-diversity (Chao1) of planktonic and benthic diatoms respectively, in different photosynthetically active radiation (PAR) regions, with size corresponding to the Chao1 index (**c**).

Figure 4. Relationships between community similarity and river channel slope for water-spring (**a**), water-autumn (**b**), sediment-spring (**c**)**,** and sediment-autumn (**d**) samples. Values of Mantel Spearman correlation (r) and correlation significance (P) are also provided. Gray lines denote ordinary least squares linear regression fits across all samples. Spatial distributions of ecological guilds for different channel slopes are shown in (**e**).