Distribution patterns of dinoflagellate communities along the Songhua River

Yangchun Gao[1,2], Yiyong Chen[1,2], Wei Xiong[1], Shiguo Li[1,2] and Aibin Zhan[1,2]

[1] Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China

ABSTRACT

Background. Dinoflagellates have the potential to pose severe ecological and economic damages to aquatic ecosystems. It is therefore largely needed to understand the causes and consequences of distribution patterns of dinoflagellate communities in order to manage potential environmental problems. However, a majority of studies have focused on marine ecosystems, while the geographical distribution patterns of dinoflagellate communities and associated determinants in freshwater ecosystems remain unexplored, particularly in running water ecosystems such as rivers and streams.

Methods. Here we utilized multiple linear regression analysis and combined information on species composition recovered by high-throughput sequencing and spatial and environmental variables to analyze the distribution patterns of dinoflagellate communities along the Songhua River.

Results. After high-throughput sequencing, a total of 490 operational taxonomic units (OTUs) were assigned to dinoflagellates, covering seven orders, 13 families and 22 genera. Although the sample sites were grouped into three distinctive clusters with significant difference ($p < 0.05$) in environmental variables, OTUs-based dinoflagellate communities among the three clusters showed no significant difference ($p > 0.05$). Among all 24 environmental factors, two environmental variables, including NO_3-N and total dissolved solids (TDS), were selected as the significantly influential factors ($p < 0.05$) on the distribution patterns of dinoflagellate communities based on forward selection. The redundancy analysis (RDA) model showed that only a small proportion of community variation (6.1%) could be explained by both environmental (NO_3-N and TDS) and dispersal predictors (watercourse distance) along the River. Variance partitioning revealed a larger contribution of local environmental factors (5.85%) than dispersal (0.50%) to the total variation of dinoflagellate communities.

Discussion. Our findings indicated that in addition to the two quantifiable processes in this study (species sorting and dispersal), more unquantifiable stochastic processes such as temporal extinction and colonization events due to rainfall may be responsible for the observed geographical distribution of the dinoflagellate community along the Songhua River. Results obtained in this study suggested that deeper investigations covering different seasons are needed to understand the causes and consequences of geographical distribution patterns of dinoflagellate biodiversity in river ecosystems.

Subjects Freshwater Biology, Environmental Impacts
Keywords Metacommunity metabarcoding, Dinoflagellate, RDA, Species sorting, Dispersal

INTRODUCTION

Aquatic ecosystems, such as rivers and lakes, support various habitats for diverse biological communities. Among those diverse communities, dinoflagellates (division Pyrrhophyta, class Dinophyceae) are ecologically important members of phytoplankton, as they play important roles in primary production, carbon cycling and oxygen release in aquatic ecosystems (*Aufdenkampe et al., 2011*; *Gaines & Elbrächter, 1987*). However, they can also pose serious economic and ecological damages to aquatic ecosystems, since some dinoflagellate species propagate quickly and form algal blooms under suitable environmental conditions. Dinoflagellate blooms largely threaten aquatic ecosystems as they cause significantly negative effects such as water fouling, oxygen deficiency, and large-scale mortality of species (*Gao et al., 2017*; *Granéli & Turner, 2006*). Hence, a deeper understanding of dinoflagellate assembly and their geographical distribution patterns is not only beneficial to the protection and management of aquatic ecosystems and associated industries such as fisheries and aquaculture, but also provides baseline information to effectively manage and conserve aquatic biodiversity (*Altermatt, 2013*; *Heino, 2013*).

Increasing evidence suggests that multiple factors can influence dinoflagellate community assembly and geographical distribution (*Aydin et al., 2015*; *Gao et al., 2018*; *Liu et al., 2012*). Major factors include environmental filtering (e.g., species sorting: species only occur at favorable environments), dispersal (e.g., mass effect: the strong mobility of individuals to reach close geographic sites; dispersal limitation: the limited ability of individuals to reach distant geographic sites) and stochastic processes (e.g., colonization and extinction). At different geographical scales, two major competing factors, dispersal and species sorting, have been commonly considered as fundamental processes in structuring biological communities in aquatic ecosystems (*Beisner et al., 2006*; *Devercelli et al., 2016*; *Heino et al., 2015*; *Isabwe et al., 2018*; *Xiong et al., 2016*; *Xiong et al., 2017*; *Wu et al., 2017*). In addition to the two major competing factors, other factors such as contrasting hydrographic conditions (e.g., wet season versus dry season) may also affect the spatial structure of aquatic communities (*Isabwe et al., 2018*). During wet seasons, frequent rainfall is expected to result in the change of connectivity among sample sites (*Larned et al., 2010*). Thus, temporal colonization and extinction events may easily occur due to the change of dispersal patterns of aquatic communities, especially for plankton communities (*Heino et al., 2015*; *Larned et al., 2010*). Such stochastic processes may largely contribute to the variation of plankton communities (*Grönros et al., 2013*; *Heino et al., 2015*). Given the complexity of various influential factors, the geographical distribution patterns of communities may largely vary among aquatic ecosystems. Although geographical variation of dinoflagellate communities and associated mechanisms have been well investigated in marine ecosystems (*Ignatiades, 2012*; *Le Bescot et al., 2016*; *Gao et al., 2018*; *Granéli & Turner, 2006*; *Jeong et al., 2010*), relevant reports in freshwater ecosystems such as rivers and streams are relatively rare.

In this study, dinoflagellate communities were collected along the Songhua River, which is located in northeast China. The Songhua River is the fifth longest river in China (1,927 km), and currently suffers from severe environmental stresses, mainly owing to increasing

disturbance derived from human activities. Mountain regions were mainly polluted by heavy metals from gold mining and metal smelting (*Zou et al., 2010*). Urban regions were polluted by organic pollutions mainly discharged by domestic sewage (*Wang et al., 2018*), and these regions were often characterized by high concentrations of chemical oxygen demand (*Lin et al., 2014*). Severe nitrogenous and phosphorus pollution derived from non-point pollution of agriculture was severely problematic in rural regions (*Yu et al., 2003*). Here we collected dinoflagellate communities from different regions of the Songhua River, and all collected communities were profiled by metabarcoding, a powerful tool for characterizing microscopic communities (*Abad et al., 2016*; *Gao et al., 2018*; *Xiong et al., 2017*; *Zhan et al., 2013*). We combined information on species composition, as well as spatial and environmental variables, to analyze the community structure and geographical distribution of dinoflagellates during the wet season of 2017. We aimed to (I) characterize the spatial distribution patterns of dinoflagellate communities and (II) disentangle the main processes that largely influence geographical distribution of dinoflagellate biodiversity.

MATERIAL AND METHOD

Sampling and DNA extraction

Dinoflagellate communities were sampled from 23 locations in July (wet season) of 2017. The sampling locations were relatively evenly distributed along the Songhua River (Fig. 1). For each sampling location, 30 l surface water (0–0.5 m) was collected and then filtered through a 25 μm mesh, and all residuals including dinoflagellates on the mesh were transferred into a 100 ml bottle and fixed with 100% alcohol for downstream analyses (*Xiong et al., 2017*). Meanwhile, 500 ml surface water was also collected and stored at 4 °C for measurement of environmental factors. Total genomic DNA of each dinoflagellate community was extracted using the DNeasy Blood and Tissue Kit (Qiagen Canada Inc., Toronto, ON, Canada). The quality of DNA was measured using NanoDrop 2000 (Thermo Scientific, Waltham, MA, USA).

Metabarcoding

All dinoflagellate communities were characterized using a metabarcoding approach. The primer pair (Uni18S: AGGGCAAKYCTGGTGCCAGC; Uni18SR: GRCGGTATCTRATC-GYCTT) (*Zhan et al., 2013*) was selected to characterize dinoflagellate communities. The pair of primers was designed based on V4 region of the 18S rDNA, and the length of amplicons varied from 400–600 bp among different species (*Zhan et al., 2013*). Although this pair of primers was originally designed for zooplankton, PCR amplification demonstrated that the pair was largely universal and could cover a wide range of aquatic taxa including dinoflagellates (*Zhan et al., 2013*; *Zhan et al., 2014*; *Zhan & MacIsaac, 2015*). In addition, the high universality of this primer pair for dinoflagellates was also confirmed by aligning representative dinoflagellate 18S with the primer pair (Fig. S1). Three PCR replicates for each sample were performed to avoid biased amplification (*Zhan et al., 2013*; *Zhan et al., 2014*; *Zhan & MacIsaac, 2015*). The PCR mixture (25 μl) consisted of 1 × *Taq* Buffer (with 20 mM Mg^{2+}; Takara, Dalian, China), 5.0 mM of each dNTP, 10 pmol of forward primer with sample-specific tags and reverse primer, 0.5 U of TaKaRa *Taq*
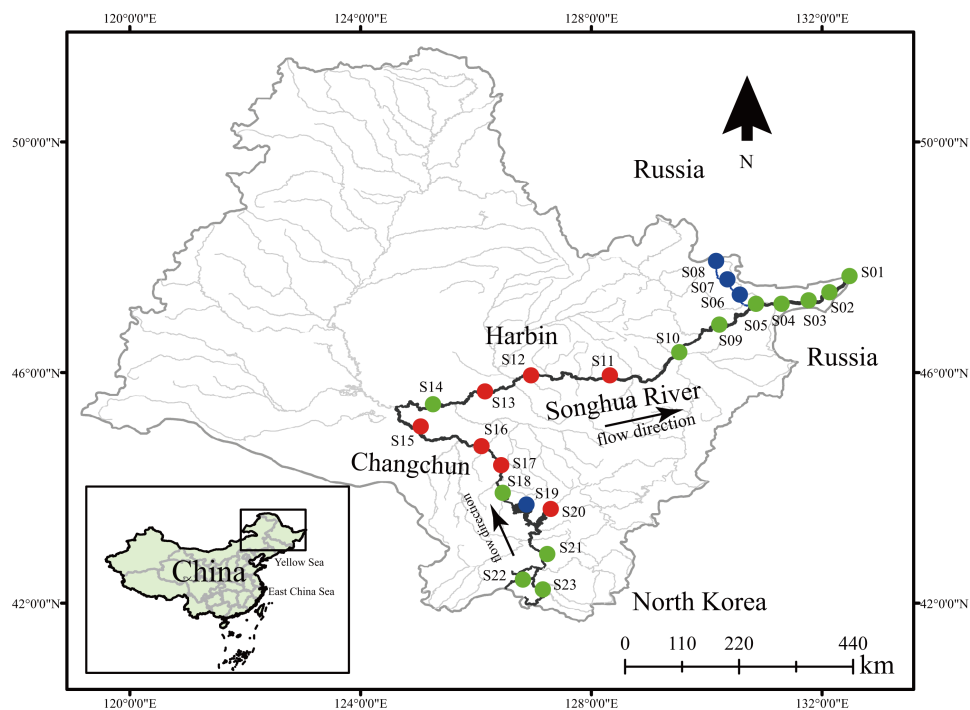
Figure 1 Sample locations of dinoflagellate communities along the Songhua River located in Northeast China. Dots in blue, red, and green represent sampling sites in Groups I, II, and III, respectively. All maps are made by ArcGIS version 10.0 (ESRI Company).

Full-size 🖼 DOI: 10.7717/peerj.6733/fig-1

(Takara) and 100 ng of genomic DNA. PCRs were performed on a Mastercycler nexus (Eppendorf, Hamburg, Germany) with the following cycle conditions: 95 °C for 5 min; then 25 cycles at 95 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s; and the final extension at 72 °C for 5 min. PCR products of the three replicates for each sample were pooled and purified using the SanPrep Spin PCR Products purification kit (Shanghai, China). Finally, a constructed sequencing library derived from the purified PCR products was sequenced using the Illumina Miseq PE300 sequencing Platform (Illumina, San Diego, CA, USA).

Raw sequence reads were denoised, trimmed and filtered using both USEARCH version 8.1 (*Edgar, 2013*) and RDP pipeline (https://pyro.cme.msu.edu/). Non-biological sequences (e.g., tags, primers and adapters) were removed using RDP. Subsequently, the expected error threshold of 0.5 was used to filter sequences with possible sequencing errors. Filtered sequences were de-replicated to obtain unique sequences, which were clustered into Operational Taxonomic Units (OTUs) at the 100% similarity (*Le Bescot et al., 2016*; *Janouškovec et al., 2017*). The obtained OTUs were annotated by searching against the Protist Ribosomal Reference database (PR2) (*Guillou et al., 2013*) using SEED version 1.46 (*Větrovský & Baldrian, 2013*) with the parameters of e value $<10^{-80}$, minimum query coverage >95% and similarity >95%. OTUs assigned to non-dinoflagellate taxa were removed from our datasets. The relative proportion of each OTU was calculated in each sample and used as the proxy of relative OTU abundance for subsequent analyses (*Hirai et al., 2015*).

Analyses of environmental factors

Environmental factors were analyzed according to the procedures described by *Xiong et al. (2017)*. Briefly, water temperature (T), pH, oxidation–reduction potential (ORP), electrical conductivity (EC) and total dissolved solid (TDS) were measured *in situ* with a multiparameter sensor (MYRON company, USA). The chlorophyll-a (Chl_a) and dissolved oxygen (DO) were determined *in situ* with a Handheld Fluorometer (Turner Designs, San Jose, CA, USA) and a portable dissolved oxygen meter (HACH company, Loveland, CO, USA), respectively. Total nitrogen (TN), nitrate nitrogen (NO_3-N) and ammonia nitrogen (NH_4-N) were measured using the alkaline potassium persulfate digestion UV spectrophotometric method, ultraviolet spectrophotometry and Nessler's reagent spectrophotometry, respectively. Total phosphorus (TP) and soluble reactive phosphorus (SRP) were determined based on the ammonium molybdate spectrophotometric method. Chemical oxygen demand (COD) and metals (K, Ca, Na, Mg, Cd, Cr, Cu, Ni, Zn, Pb, and As) were measured with HACH COD digestion vials (HACH Company, Loveland, CO, USA) and inductively coupled plasma-mass spectrometry (ICP-MS, 7500A, Plasma Quad 3, USA), respectively. A total of 24 environmental factors were collected in this study.

Spatial variables

Dispersal, such as the mass effect in running rivers, often plays a key role in structuring plankton communities. To test the role of dispersal in shaping dinoflagellate communities, we measured the actual watercourse distance and used it for dispersal proxy between sampling sites (*Beisner et al., 2006*).

Statistical analyses

Before statistical analyses, the relative abundance of OTUs and all measured environmental factors, except for pH, were $\log_{10}(x+1)$ transformed to improve homoscedasticity. We first clustered sampling sites into distinctive groups based on the Euclidean distance of environmental variables using the CLUSTER program. To characterize distribution patterns of dinoflagellate communities, we performed further tests including an analysis of similarity (ANOSIM), an analysis of similarity percentages (SIMPER), and a nonmetric multidimensional scaling (NMDS) using non-parametric multivariate methods. The relative abundance of dinoflagellates between groups was compared using ANOSIM, which was based on Bray-Curtis distance and rank dissimilarity. The SIMPER analysis was used to identify the major OTUs responsible for the contribution of community variation at both the intra- and inter-group levels. An NMDS analysis on environmental variables was performed to profile the inter-group relativeness among sampling sites. Based on the NMDS results on environmental variables, we performed another NMDS analysis on dinoflagellate communities to primarily assess the potential influence of environmental factors on dinoflagellate communities. All CLUSTER, ANOSIM, SIMPER and NMDS analyses were performed in PRMIER 6.0 (*Clarke & Gorley, 2001*).

To disentangle the main processes that largely influence geographical distribution of dinoflagellate biodiversity, we performed a linear ordination method, redundancy analysis (RDA). Since collinearity among explanatory variables can lead to the inflation of type
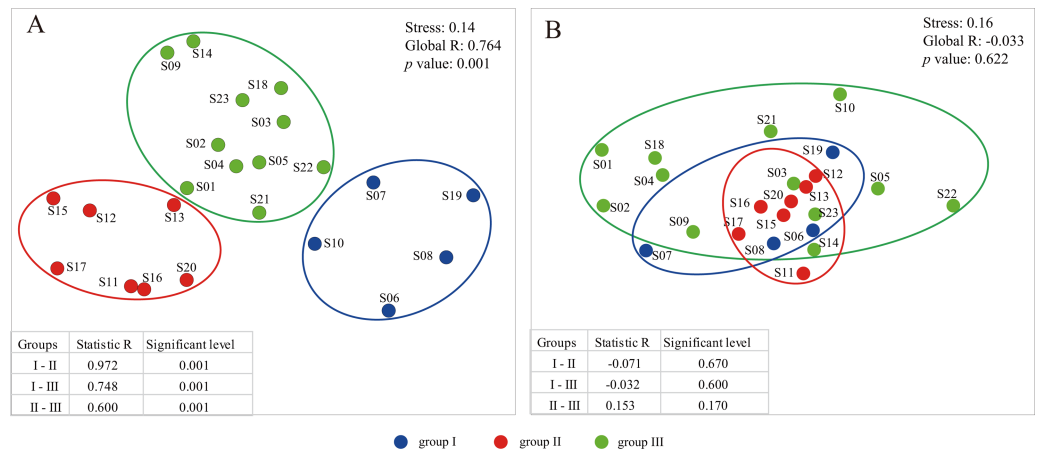
Figure 2 The plots of nonmetric multidimensional scaling ordination (NMDS) based on environmental variables (A) and dinoflagellate communities (B).

Full-size ⬜ DOI: 10.7717/peerj.6733/fig-2

I error and overestimation of the amount of explained variation (*Blanchet, Legendre & Borcard, 2008*), we conducted forward selection to select significant environmental variables using the *forward.sel* function (ANOVAS; 1,000 permutations) implemented in *packfor* package in R (*R Core Team, 2015*). To demonstrate the relative contribution of species sorting and dispersal to the community structure variation, variance partitioning and partial redundancy analyses (pRDA) were performed to estimate the proportion of dinoflagellate community variation purely explained by environmental predictors (significant environmental factors) and dispersal (watercourse distance). Variance analyses (ANOVAS; 1,000 permutations) were performed to test the significance of RDA and pRDA. The analyses including RDA, variance partitioning and pRDA were computed using *vegan* package in R (*R Core Team, 2015*).

RESULTS

Environmental factors

A total of 24 environmental factors were collected at the 23 sampling sites along the Songhua River (Table S1). Based on the environmental variables, all sampling sites were clustered into three distinctive groups (I, II, III) (Stress = 0.14; Figs. 1 and 2A). In addition, the ANOSIM analysis revealed significant differences in the environmental factors between the three groups (global $R = 0.764$; $p = 0.001$; Fig. 2A). The environmental factors largely varied among the three groups, particularly for NH_4-N and NO_3-N. The concentration of NH_4-N was the lowest in the group I (0.01–0.24 mg l^{-1}, mean = 0.074 mg l^{-1}), higher in the group II (0.01–0.47 mg l^{-1}, mean = 0.15 mg l^{-1}), and the highest in the group III (0.01–2.74 mg l^{-1}, mean = 0.66 mg l^{-1}; Table S1). Similarly, NO_3-N was the lowest in the group I (0.51–2.88 mg l^{-1}, mean = 1.48 mg l^{-1}) and the highest in the group III (0.42–11.61 mg l^{-1}, mean = 6.24 mg l^{-1}; Table S1).
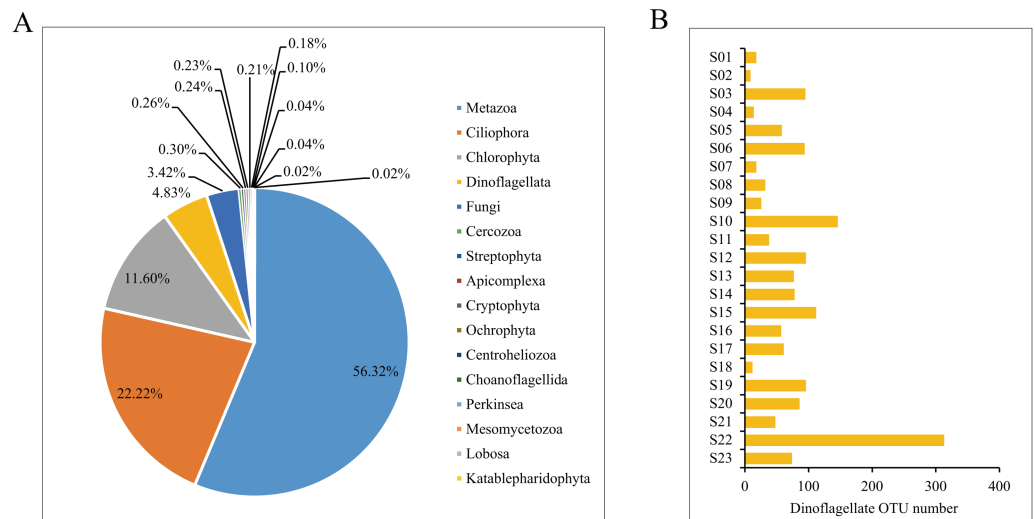
Figure 3 **The annotation of Operational Taxonomic Units (OTUs).** (A) Proportions of taxonomic composition; (B) the number of dinoflagellate OTUs per sample.

Full-size 🖼 DOI: 10.7717/peerj.6733/fig-3

Composition and distribution of dinoflagellate communities

MiSeq sequencing produced a total of 346,490 raw reads from the 23 samples (NCBI SRA accession no. SRP151183). After quality filtering and OTU clustering at 100% similarity, a total of 11,530 OTUs were obtained, and the rarefaction curves for all samples reached saturation or almost saturation (Fig. S2), suggesting that the biodiversity was well recovered. Among the 11,530 OTUs, 10,153 OTUs were successfully annotated, covering Metazoa (56.32%), followed by Ciliophora (22.22%), Chlorophyta (11.60%) and Dinoflagellata (4.83%) (Fig. 3A). All 490 dinoflagellate OTUs were used for downstream analyses. The number of OTUs per sample site largely varied from 9 at site S02 to 313 at site S22, and the average number of OTUs per sample was 72 (Fig. 3B, Table S2).

The taxonomic assignment of the dinoflagellate OTUs covered seven orders: Gonyaulacales, Peridiniales, Dinophyceae_X, Prorocentrales, Suessiales, Gymnodiniales and Dino-Group-I, of which Gonyaulacales was the most abundant taxa (307 OTUs), followed by Peridiniales (121 OTUs) and Dinophyceae_X (28 OTUs). From the seven orders, we identified 13 families, of which Ceratiaceae (307 OTUs), Kryptoperidiniaceae (94 OTUs) and Proroentraceae (20 OTUs) were identified as the first three dominant taxa. A total of 22 genera were retrieved, and the top six were *Ceratium* (307 OTUs), *Unruhdinium* (93 OTUs), *Prorocentrum* (20 OTUs), *Dinophyceae_XXX* (17 OTUs), *Thoracosphaeraceae_X* (12 OTUs) and *Woloszynskia* (7 OTUs) (Fig. 4, Table S2). Among the 490 OTUs, OTU_6 (Gouyaulacales, Ceratiaceae, *Ceratium*), OTU_7 (Peridiniales, Kryptoperidiniaceae, *Unruhdinium*) and OTU_115 (Peridiniales, Thoracosphaeraceae, *Thoracosphaeraceae_X*) were the top three taxa with highest relative abundance (Table S2).

To primarily assess the potential influence of environmental factors on distribution patterns of dinoflagellate communities, we also performed CLUSTER and NMDS analyses on dinoflagellate communities based on groups inferred from environmental variables.

Figure 4 The Operational Taxonomic Unit (OTU) number of different taxonomic groups. (A) Order level; (B) Family level; (C) Genus level.

Full-size ⊡ DOI: 10.7717/peerj.6733/fig-4

However, no significant variation was found among groups with a global R of -0.033 ($p = 0.622$; Fig. 2B), and the two clusters based on the dendrogram were not coincide with environmental groups (Fig. 2A; Fig. S3). In addition, we observed a high level of dissimilarity at the inter-group level; the average dissimilarity was 83.04 ($R = -0.071$, $p = 0.670$) between groups I and II, 77.94 ($R = -0.032$, $p = 0.600$) between groups I and III, and 68.91 ($R = 0.153$, $p = 0.170$) between groups II and III (Table S3; Fig. 2B). However, we found a low level of similarity at the intra-group level: the average similarity values were 12.39, 21.44 and 29.15 in groups I, II, and III, respectively (Table S3). In addition, the taxa responsible for the intra-group similarity were almost similar. For example, OTU_6 (Ceratiaceae, Ceratiaceae, *Ceratium*), which had wide geographical distribution and large variation in relative abundance among sample sites, contributed 54.74%, 83.44%, and 76.00% to the similarity within the groups I, II, III, respectively (Table S4). Meanwhile, the OTU_6 (Ceratiaceae, Ceratiaceae, *Ceratium*) was also one of the top contributors to the dissimilarity between groups, contributing 28.61%, 27.22%, and 28.12% to the dissimilarity between groups I and II, groups I and III, and groups II and III, respectively (Table S3).

Influence of environmental factors on community structure

To construct a parsimonious RDA model, environmental factors with significant influence on dinoflagellate community structure were selected based on the results of forward selection. The results showed that two environmental factors (NO$_3$-N and TDS) were selected and included in the RDA framework (Fig. 5A). The RDA was globally significant
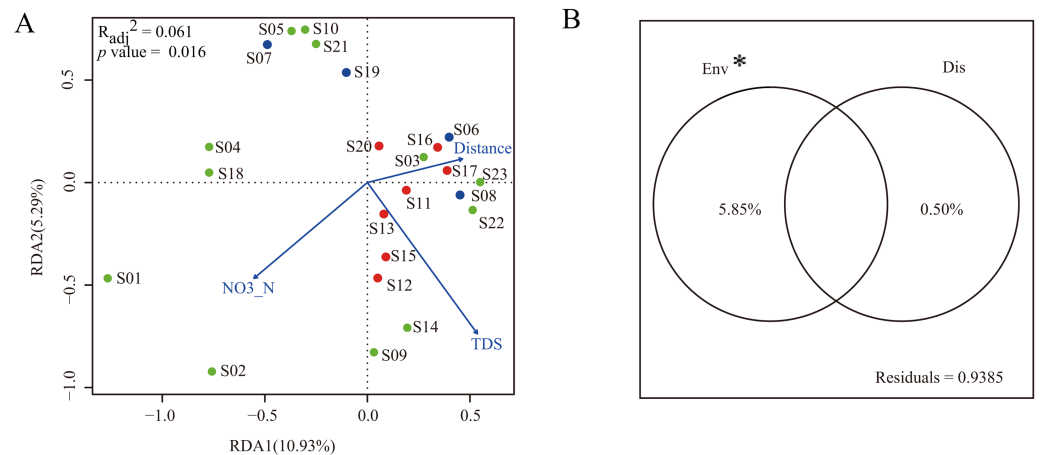
Figure 5 **Multiple linear regress analysis.** (A) The ordination plot based on redundancy analysis of dinoflagellate community with selected environmental predictor (NO_3-N and TDS) and dispersal predictor (watercourse distance). Dots in blue, red, and green represent sampling sites in Groups I, II, and III, respectively; (B) the variance partitioning of the selected environmental predictors (NO_3-N and TDS) and dispersal predictor (watercourse distance). "Env" and "Dis" represent environmental and dispersal predictors, respectively. "*" indicates significant difference ($p < 0.05$).

Full-size 🖼 DOI: 10.7717/peerj.6733/fig-5

($p = 0.016$) with an adjusted coefficient of determination (R_{adj}^2) of 0.061. The first two axes of the RDA model explained 10.93% and 5.29% of the total variation, respectively (Fig. 5A).

In order to characterize the relative contribution of environmental predictors (NO_3-N and TDS) and spatial predictor (watercourse distance) to the observed dinoflagellate community structure, a variance partitioning was performed. The results demonstrated that the environmental predictors, which explained 5.85% of the total variation, had a significant influence on dinoflagellate community structure when excluding effects of the spatial predictor ($p < 0.05$). Conversely, when excluding the influence of environmental predictors, the spatial predictor only explained 0.50% of the total variation and did not significantly influence the dinoflagellate community structure ($p > 0.05$, Fig. 5B).

DISCUSSION

Geographical distribution of dinoflagellate communities

In the present study, only two environmental predictors (NO_3-N and TDS) significantly influenced the geographical distribution of dinoflagellate communities along the Songhua River. After variance partitioning, the significant influence still remained, further highlighting the importance of these two environmental factors in shaping distribution patterns of dinoflagellate communities along the Songhua River. In general, NO_3-N is essential for the growth and proliferation of plankton and has been reported with strong correlations with dinoflagellate abundance (*Lafrancois et al., 2004*). As expected, our results highlight the importance of NO_3-N in affecting the variation of dinoflagellate abundance along the Songhua River. In contrast, TDS can have toxic effects on plankton (*Ivanova & Kazantseva, 2006*). Thus, the varied TDS concentration levels among sites can lead to

different abundances of dinoflagellates (e.g., low abundance at site S10 with high TDS; Fig. 5A), further leading to geographical variation of communities.

However, both environmental and spatial predictors only explain 6.1% of the total variation, indicating that some unquantifiable variables play stronger roles in structuring the spatial distribution of dinoflagellate communities. These observations suggest that stochastic processes, such as temporal colonization and extinction events, should be additional factors in structuring dinoflagellate communities along the Songhua River (*Grönros et al., 2013*; *Heino et al., 2015*). Temporal colonization and extinction events have been frequently observed in ponds, especially temporary ponds where "strong small island effects" and "elements of chance" are expected to play key roles in structuring communities (*Heino et al., 2015*; *Oertli et al., 2010*). However, rapid extinction and colonization events are rarely observed in running water systems in general, as excessive dispersal can homogenize biological community structure in rivers and streams (*Heino et al., 2015*). Since we collected samples along the Songhua River during the wet season when it rained frequently and parts of waters may derive from basin ponds, the sampled dinoflagellate communities, or parts of the sampled communities, may derive from ponds. Consequently, only a small proportion of community variation could be explained by both environmental and spatial factors. Indeed, the small proportion of community variation explained by environmental and spatial factors have been commonly observed in aquatic ecosystems (*Beisner et al., 2006*; *Devercelli et al., 2016*). For example, one study in a lake ecosystem showed that only 8% of the variation of phytoplankton communities was explained by spatial factors and measured environmental parameters, and the authors argued that one-time sampling of phytoplankton communities and the absence of some important environmental variables (e.g., disturbance frequency) were the major reasons for this finding (*Beisner et al., 2006*). A study at fine geographical scale of a river floodplain showed that less than 17.8% of phytoplankton community variation was explained by environmental and spatial variables, and the authors suggested that random dispersion, ecological drift, and priority effects were important ecological processes responsible for phytoplankton meta-communities (*Devercelli et al., 2016*). Collectively, the low explanation power of environmental and spatial factors in this study may be attributed to four factors: (I) temporal colonization and extinction due to rainfall; (II) unmeasured environmental factors, such as relative light intensity and disturbance frequency, where the former is essential for the reproduction and growth of dinoflagellates and the latter has been shown to structure phytoplankton communities (*Beisner, 2001*); (III) higher trophic level predators, such as zooplankton, were not considered in this study; the abundance and taxa of predators can make significant influence on the prey composition and abundance variation (*Kozak, Goldyn & Dondajewska, 2015*); (IV) although the metabarcoding method shows robust power for diversity assessments of various communities (*Zhan et al., 2014*; *Xiong et al., 2017*; *Yang et al., 2018*), this method still cannot completely quantify the abundance of community composition (*Sun et al., 2015*), especially for taxonomic groups with large size variation. The poor relationship between sequence abundance based on molecular methods and real species abundance (*Godhe et al., 2008*) may disturb statistical

analyses considerably for the exploration of complex interactions between organisms and environments.

Two genera, including *Ceratium* and *Unruhdinium*, were identified with high relative abundance along the Songhua River. Some species of *Ceratium*, such as *Ceratium hirundinella* and *Ceratium furcoides*, can form algae blooms (*Van Ginkel, Hohls & Vermaak, 2001*; *Cassol et al., 2014*). As such algae blooms may kill fish and lead to mass economic losses, as seen in Japan (*Taylor, Fukuyo & Larson, 1995*), harmful algae should be closely monitored for a better understanding of their population dynamics to manage economic and ecological issues along the Songhua River. Blooms of *Ceratium* spp. have been reported as the consequences of environmental changes, such as fluctuations of nutrient levels, temperature, and dissolved oxygen concentration (*Periotoo et al., 2007*; *da Silva et al., 2012*). Specifically, *Ceratium* spp. blooms were reported with relation to the increase of eutrophic levels in many rivers and lakes such as Furnas Reservoirs, Brazil (*da Silva et al., 2012*); however, related studies reported that the density of *Ceratium* spp. tended to become high under mesotrophic conditions (*Periotoo et al., 2007*). Thus, the relationship between trophic level and *Ceratium* spp. abundance was still unclear. Our results demonstrated that the relative abundance of *Ceratium* spp. varied largely among sites along the Songhua River. Although the RDA showed that the environmental predictors (NO_3-N and TDS) should be responsible for the variation, the lower explanatory extent (<6.1%) made it difficult to infer a conclusion on the determinants of *Ceratium* spp. Hence, a deeper investigation covering different seasons is largely required to further recover the population fluctuation dynamics of *Ceratium*.

The relative roles of species sorting and dispersal

The selected environmental factors (NO_3-N and TDS) significantly explained only 5.85% of the total variation of dinoflagellate communities when the influence of watercourse distance was excluded. Interestingly, the dispersal predictor (watercourse distance) did not significantly influence dinoflagellate communities when excluding the influence of the environmental factors. The results obtained here are consistent with those in several other river systems such as Parana River floodplain (*Devercelli et al., 2016*) and Jiulong River (*Isabwe et al., 2018*), whereas opposite findings have been frequently observed (*Datry et al., 2016*; *Heino et al., 2015*). Usually, the inconsistent results may derive from whether there are significant environmental gradients in river systems. The Songhua River flows through three types of regions with different levels of pollutants such as NO_3-N. Similar environmental gradients were also observed in Jiulong River (*Isabwe et al., 2018*) and Chaobai River (*Xiong et al., 2017*). The strong environmental gradients provide preconditions for species sorting, where species only occur at favorable environments. However, we cannot rule out a possible reason of discrepancy caused by the difference of communities (e.g., dinoflagellate versus phytoplankton). Since the relative contribution of species sorting and dispersal also depends on the taxonomic groups with diverse dispersal abilities and life histories (*Beisner et al., 2006*; *Devercelli et al., 2016*; *Isabwe et al., 2018*; *Lin et al., 2014*; *Xiong et al., 2017*), further studies on different taxonomic groups should be performed to verify whether the relative contribution of species sorting and dispersal differ among taxonomic groups.

CONCLUSIONS

In summary, we used a metabarcoding-based approach to analyze the geographical distribution of dinoflagellate communities and identified factors responsible for the observed patterns. The minor proportion (6.1%) of community variation explained by environmental and spatial predictors indicates that additional stochastic processes, such as temporal extinction and colonization events, may play crucial roles in structuring dinoflagellate communities along the Songhua River during the wet season. The dissimilarity of dinoflagellate communities at the intra-group level was significantly greater than that at the inter-group level, suggesting in addition to two quantifiable processes (species sorting and dispersal), more complex processes should be involved in determining the community structure. Our study suggests that deeper investigations covering different seasons are required to understand the causes and consequences of geographical distribution of dinoflagellate communities and causative factors for the observed patterns on river ecosystems. Such information is crucial for both ecological surveys and conservation/management of biodiversity in different habitats.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Yangchun Gao conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Yiyong Chen performed the experiments, analyzed the data, prepared figures and/or tables, approved the final draft.

- Wei Xiong performed the experiments, prepared figures and/or tables, approved the final draft.
- Shiguo Li performed the experiments, contributed reagents/materials/analysis tools, approved the final draft.
- Aibin Zhan conceived and designed the experiments, contributed reagents/materials/-analysis tools, approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The sequencing data is available at NCBI (SRA accession no. SRP151183 and Bioproject PRJNA477641).

Data Availability

The following information was supplied regarding data availability:

The raw measurements are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.6733#supplemental-information.

REFERENCES

Abad D, Albaina A, Aguirre M, Laza-Martínez A, Uriarte I, Iriarte A, Villate F, Estonba A. 2016. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. *Marine Biology* **163**:Article 149 DOI 10.1007/s00227-016-2920-0.

Altermatt F. 2013. Diversity in riverine metacommunities: a network perspective. *Aquatic Ecology* **47(3)**:365–377 DOI 10.1007/s10452-013-9450-3.

Aufdenkampe AK, Mayorga E, Raymond PA, Melack JM, Doney SC, Alin SR, Aalto RE, Yoo K. 2011. Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. *Frontiers in Ecology & the Environment* **9(1)**:53–60 DOI 10.1890/100014.

Aydin H, Yürür EE, Uzar S, Küçüksezgin F. 2015. Impact of industrial pollution on recent dinoflagellate cysts in Izmir Bay (Eastern Aegean). *Marine Pollution Bulletin* **94(1–2)**:144–152 DOI 10.1016/j.marpolbul.2015.02.038.

Beisner BE. 2001. Plankton community structure in fluctuating environments and the role of productivity. *Oikos* **95(3)**:496–510 DOI 10.1034/j.1600-0706.2001.950315.x.

Beisner BE, Peres-Neto PR, Lindström ES, Barnett A, Longhi ML. 2006. The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology* **87(12)**:2985–2991 DOI 10.1890/0012-9658(2006)87[2985:TROEAS]2.0.CO;2.

Blanchet FG, Legendre P, Borcard D. 2008. Forward selection of explanatory variables. *Ecology* **89(9)**:2623–2632 DOI 10.1890/07-0986.1.

Cassol APV, Pereira Filho W, Oliveira MA, Domingues AL, Correa FS, Buriol GA. 2014. First record of a bloom of the invasive species *Ceratium furcoides* (levander) langhans 1925 in rio grande do sul state, brazil. *Brazilian Journal of Biology* **74(2)**:515–517 DOI 10.1590/1519-6984.05413.

Clarke KR, Gorley RN. 2001. *PRIMER v5: user manual/tutorial.* Plymouth: PRIMER-E Ltd.

da Silva LC, Leone IC, dos Santos-Wisniewski MJ, Peret AC, Rocha O. 2012. Invasion of the dinoflagellate Ceratium furcoides (levander) langhans 1925 at tropical reservoir and its relation to environmental variables. *Biota Neotropica* **12(2)**:93–100 DOI 10.1590/S1676-06032012000200010.

Datry T, Melo AS, Moya N, Zubieta J, De la Barra E, Oberdorff T. 2016. Metacommunity patterns across three Neotropical catchments with varying environmental harshness. *Freshwater Biology* **61(3)**:277–292 DOI 10.1111/fwb.12702.

Devercelli M, Scarabotti P, Mayora G, Schneider B, Giri F. 2016. Unravelling the role of determinism and stochasticity in structuring the phytoplanktonic metacommunity of the Parana River floodplain. *Hydrobiologia* **764(1)**:139–156 DOI 10.1007/s10750-015-2363-5.

Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**:996–998 DOI 10.1038/nmeth.2604.

Gaines G, Elbrächter M. 1987. Heterotrophic nutrition. In: Taylor FJR, ed. *The biology of dinoflagellates. Botanical monographs.* Oxford: Blackwell Science Publications.

Gao Y, Dong Y, Li H, Zhan A. 2017. Harmful non-indigenous dinoflagellate cysts in China: a review. *Aquatic Ecosystem Health & Management* **20(4)**:413–422 DOI 10.1080/14634988.2017.1403269.

Gao Y, Dong Y, Li H, Zhan A. 2018. Influence of environmental factors on spatial–temporal distribution patterns of dinoflagellate cyst communities in the South China Sea. *Marine Biodiversity* Epub ahead of print Feb 01 2018 DOI 10.1007/s12526-018-0850-4.

Godhe A, Asplund ME, Harnstrom K, Saravanan V, Tyagi A, Karunasagar I. 2008. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Applied and Environmental Microbiology* **74**:7174–7182 DOI 10.1128/AEM.01298-08.

Granéli E, Turner JT. 2006. *Ecology of Harmful Algae. Ecological studies series 189*, Heidelberg: Springer-Verlag.

Grönros M, Heino J, Siqueira T, Landeiro VL, Kotanen J, Bini LM. 2013. Metacommunity structuring in stream networks: roles of dispersal mode, distance type, and regional environmental context. *Ecology & Evolution* **3(13)**:4473–4487 DOI 10.1002/ece3.834.

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 JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahe F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet AL, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R. 2013. The Protist Ribosomal

Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* **41(D1)**:D597–D604 DOI 10.1093/nar/gks1160.

Heino J. 2013. The importance of metacommunity ecology for environmental assessment research in the freshwater realm. *Biological Reviews* **88(1)**:166–178 DOI 10.1111/j.1469-185X.2012.00244.x.

Heino J, Melo AS, Siqueira T, Soininen J, Valanko S, Bini LM. 2015. Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects. *Freshwater Biology* **60(55)**:845–869 DOI 10.1111/fwb.12533.

Hirai J, Yasuike M, Fujiwara A, Nakamura Y, Hamaoka S, Katakura S, Takano Y, Nagai S. 2015. Effects of plankton net characteristics on metagenetic community analysis of metazoan zooplankton in a coastal marine ecosystem. *Journal of Experimental Marine Biology and Ecology* **469**:36–43 DOI 10.1016/j.jembe.2015.04.011.

Ignatiades L. 2012. Mixotrophic and heterotrophic dinoflagellates in eutrophic coastal waters of the Aegean Sea (eastern Mediterranean Sea). *Botanica Marina* **55(1)**:39–48 DOI 10.1515/bot-2012-0096.

Isabwe A, Yang JR, Wang Y, Liu L, Chen H, Yang J. 2018. Community assembly processes underlying phytoplankton and bacterioplankton across a hydrologic change in a human-impacted river. *The Science of the Total Environment* **630**:658–667 DOI 10.1016/j.scitotenv.2018.02.210.

Ivanova MB, Kazantseva TI. 2006. Effect of water pH and total dissolved solids on the species diversity of pelagic zooplankton in lakes: a statistical analysis. *Russian Journal of Ecology* **37(4)**:264–270 DOI 10.1134/s1067413606040084.

Janouškovec J, Gavelis GS, Burki F, Dinh D, Bachvaroff TR, Gornik SG, Bright KJ, Imanian B, Strom SL, Delwiche CF, Waller RF, Fensome BA, Leander BS, Rohwer FL, Saldarriaga JF. 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences of the United States of America* **114(2)**:E171–E180 DOI 10.1073/pnas.1614842114.

Jeong HJ, Yoo YD, Kim JS, Seong KA, Kang NS, Kim TH. 2010. Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. *Ocean Science Journal* **45(2)**:65–91 DOI 10.1007/s12601-010-0007-2.

Kozak A, Goldyn R, Dondajewska R. 2015. Phytoplankton composition and abundance in restored maltanski reservoir under the influence of physico-chemical variables and zooplankton grazing pressure. *PLOS ONE* **10(4)**:e0124738 DOI 10.1371/journal.pone.0124738.

Lafrancois BM, Nydick KR, Johnson BM, Baron JS. 2004. Cumulative effects of nutrients and pH on the plankton of two mountain lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **61(7)**:1153–1165 DOI 10.1139/f04-053.

Larned ST, Datry T, Arscott DB, Tockner K. 2010. Emerging concepts in temporary-river ecology. *Freshwater Biology* **55(4)**:717–738 DOI 10.1111/j.1365-2427.2009.02322.x.

Le Bescot N, Mahé F, Audic S, Dimier C, Garet M-J, Poulain J, Wincker P, De Vargas C, Siano R. 2016. Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environmental Microbiology* **18(2)**:609–626 DOI 10.1111/1462-2920.13039.

Lin S, Wang Y, Lin J, Quan C. 2014. Response of planktonic and benthic microbial community to urban pollution from sewage discharge in Jilin Reach of the second Songhua River, China. *Clean-Soil Air Water* **42(10)**:1376–1383 DOI 10.1002/clen.201200328.

Liu D, Shi Y, Di B, Sun Q, Wang Y, Dong Z, Shao H. 2012. The impact of different pollution sources on modern dinoflagellate cysts in Sishili Bay, Yellow Sea, China. *Marine Micropaleontology* **84–85**:1–13 DOI 10.1016/j.marmicro.2011.11.001.

Oertli B, Biggs J, Céréghino R, Grillas P, Joly P, Lachavanne JB. 2010. Conservation and monitoring of pond biodiversity: introduction. *Aquatic Conservation Marine & Freshwater Ecosystems* **15(6)**:535–540 DOI 10.1002/aqc.752.

Periotoo M, Prósperi C, Daga C, Rincón A. 2007. Estudio de Ceratium hirundinella en E1 Embalse La Quebrada, Cordoba, Argentina. In: *International Congress on Development Environment and Natural Resources: Multi-level and Multi-scale Sustainability. Cochabamba, Bolívia.*

R Core Team. 2015. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/*.

Sun C, Zhao Y, Li H, Dong Y, Macisaac HJ, Zhan A. 2015. Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities. *Aquatic Biology* **24(1)**:9–15 DOI 10.3354/ab00629.

Taylor F, Fukuyo Y, Larson J. 1995. Taxonomy of harmful dinoflagellate. In: Hallengraeff GM, Aderson DM, Cembella AD, eds. *Mannual of harmful marine microalgae.* Paris: Intergovernmental Oceanographic Commission, UNESCO, 551.

Van Ginkel CE, Hohls BC, Vermaak E. 2001. A Ceratium hirundinella (OF Müller) bloom in Hartbeespoort Dam, South Africa. *Water SA* **27(2)**:269–276 DOI 10.4314/wsa.v27i2.5000.

Větrovský T, Baldrian P. 2013. Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED. *Biological Fertily of Soils* **49(10)**:1027–1037 DOI 10.1002/hbm.22096.

Wang W, Zhang W, Liang H, Gao D. 2018. Seasonal distribution characteristics and health risk assessment of typical antibiotics in the Harbin section of the Songhua River basin. *Environmental Technology* **14**:1–12 DOI 10.1080/09593330.2018.1449902.

Wu W, Lu HP, Sastri A, Yeh YC, Gong GC, Chou WC, Hsieh CH. 2017. Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *The ISME Journal* **12**:485–494 DOI 10.1038/ismej.2017.183.

Xiong W, Li J, Chen Y, Shan B, Wang W, Zhan A. 2016. Determinants of community structure of zooplankton in heavily polluted river ecosystems. *Scientific Reports* **6**:22043 DOI 10.1038/srep22043.

Xiong W, Ni P, Chen Y, Gao Y, Shan B, Zhan A. 2017. Zooplankton community structure along a pollution gradient at fine geographical scales in river ecosystems: the importance of species sorting over dispersal. *Molecular Ecology* **26**(16):4351–4360 DOI 10.1111/mec.14199.

Yang Y, Ni P, Gao Y, Xiong W, Zhao Y, Zhan A. 2018. Geographical distribution of zooplankton biodiversity in highly polluted running water ecosystems: Validation of fine-scale species sorting hypothesis. *Ecology and Evolution* **8**(10):4830–4840 DOI 10.1002/ece3.4037.

Yu SX, Shang JC, Zhao JS, Guo HX. 2003. Factor analysis and dynamics of water quality of the Songhua River, Northeast China. *Water Air Soil Pollution* **144**(1–4):159–169 DOI 10.1023/a:1022960300693.

Zhan A, Bailey SA, Heath DD, MacIsaac HJ. 2014. Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Molecular Ecology Resources* **14**(5):1049–1059 DOI 10.1111/1755-0998.12254.

Zhan A, Hulak M, Sylvester F, Huang X, Adebayo AA, Abbott CL, Adamowicz SJ, Heath DD, Cristescu ME, MacIsaac HJ. 2013. High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution* **4**(6):558–565 DOI 10.1111/2041-210x.12037.

Zhan A, MacIsaac HJ. 2015. Rare biosphere exploration using high-throughput sequencing: research progress and perspectives. *Conservation Genetics* **16**(3):513–522 DOI 10.1007/s10592-014-0678-9.

Zou TT, Wang N, Zhang G, Zhao DD. 2010. Environment spatial distribution of mercury pollution in Songhua River upstream gold mining areas. *Environmental Science* **31**(9):2228–2233.