

## Characterization of Natural and Affected Environments

**New Insights into Sediment Transport in Interconnected  
River-Lake Systems Through Tracing Microorganisms**

Wenlong Zhang, Jinfei Gu, Yi Li, Li Lin, Peifang Wang, Chao Wang, Bao Qian, Haolan Wang, Lihua Niu, Longfei Wang, Huanjun Zhang, Yu Gao, Mengjie Zhu, and Siqi Fang

*Environ. Sci. Technol.*, **Just Accepted Manuscript** • DOI: 10.1021/acs.est.8b07334 • Publication Date (Web): 13 Mar 2019

Downloaded from <http://pubs.acs.org> on March 17, 2019

**Just Accepted**

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

1     **New Insights into Sediment Transport in Interconnected River-Lake Systems**  
2                                   **Through Tracing Microorganisms**

3  
4     Wenlong Zhang <sup>a</sup>, Jinfei Gu <sup>a</sup>, Yi Li <sup>a, \*</sup>, Li Lin <sup>b</sup>, Peifang Wang <sup>a</sup>, Chao Wang <sup>a</sup>, Bao  
5     Qian <sup>c</sup>, Haolan Wang <sup>a</sup>, Lihua Niu <sup>a</sup>, Longfei Wang <sup>a</sup>, Huanjun Zhang <sup>a</sup>, Yu Gao <sup>a</sup>,  
6                                   Mengjie Zhu <sup>a</sup>, Siqi Fang <sup>a</sup>

7  
8     <sup>a</sup> Key Laboratory of Integrated Regulation and Resource Development on Shallow  
9     Lakes, Ministry of Education, College of Environment, Hohai University, Nanjing  
10    210098, P.R. China

11    <sup>b</sup> Basin Water Environmental Research Department, Changjiang River Scientific  
12    Research Institute, Wuhan 430010, P.R. China

13    <sup>c</sup>Hydrology Bureau of Changjiang Water Resources Commission, Wuhan 430010, P.R.  
14    China.

15  
16    \* Corresponding author: Dr. Yi Li

17    Address: College of Environment, Hohai University, Nanjing 210098, P.R. China

18    Phone: 86-25-83786251

19    Fax: 86-25-83786090

20    E-mail: envly@hhu.edu.cn

## Abstract

A growing awareness of the wider environmental significance of diffuse sediment pollution in interconnected river-lake systems has generated the need for reliable provenance information. Owing to their insufficient ability to distinguish between multiple sources, common sediment source apportionment methods would rarely be a practical solution. Based on the inseparable relationships between sediment and adsorbed microorganisms, community-based microbial source tracking may be a novel method of identifying dominant sediment sources in the era of high-throughput sequencing. Dongting Lake was selected as a study area as it receives considerable sediment import from its inflowing rivers during the flood season. This study was conducted to characterize the bacterial community composition of sediment samples from the inflow-river estuaries and quantify their sediment microbe contributions to the central lake. Metagenomic analysis revealed that the community compositions of source sediment samples were significantly different, allowing specific sources to be identified with the machine learning classification program SourceTracker. Modified analysis using SourceTracker found that the major contributors to three major lake districts were the Songzi, Zishui, and Xinqiang Rivers. The impacts of hydrodynamic conditions on source apportionment were further verified, and suggested the practicability of this method to offer a systematic and comprehensive understanding of sediment sources, pathways, and transport dynamics. Finally, a novel framework for sediment source-tracking was established to develop effective sediment management and control strategies in river-lake systems.

- 43    **KEYWORDS:** River-lake interactions; Sediment transport; Microbial community;
- 44    SourceTracker; Dongting Lake

## 1. INTRODUCTION

Inseparable river-lake interactions are generated by the mutual exchange of water, sediment, pollutants, and organisms in river-lake systems<sup>1,2</sup>. Current research suggests that this relationship could be significantly impacted by natural and anthropogenic factors, including climate change, rapid urbanization and industrialization, as well as the operation of water conservancy projects<sup>3,4</sup>. The spatio-temporal variation of sediment transport and fate caused by the evolution of river-lake interactions should receive more attention<sup>1,5</sup>.

Sediment, an essential exchange matter and carrier in river-lake systems, simultaneously promotes the migration of adsorbed pollutants and microorganisms, which exert a fundamental control over the ecological status and ecosystem services of lakes<sup>6,7</sup>. Considerable sediment import from inflow-rivers has become the main cause of severe lake contamination and eutrophication<sup>8</sup>. Evidence-based identification of the major sediment source is required to appropriately assess the health risks and implement effective controls for lake pollutants<sup>9</sup>. Common sediment source apportionment using physical, chemical, biological, and composite fingerprinting<sup>10-13</sup> has been successfully applied as a tool to gain insights into sediment dynamics at a river-basin scale in catchments worldwide. However, there are still areas of uncertainty in traditional sediment source-tracking methods for application to large catchments (such as interconnected river-lake systems) with a complex arrangement of sources<sup>14</sup>. The sediment load imported by the inflow-river commonly represents a mixture of sediment derived from different locations and source types in upstream regions, which could limit

67 sediment source discrimination as selected physicochemical tracers are not entirely  
68 source-specific, and multiple sources within a system have similar tracer concentrations  
69 <sup>15, 16</sup>. Until now, employing existing approaches to determine dominant dispersal  
70 sources in river-lake systems has been a formidable challenge <sup>5</sup>.

71 Sediments and their adsorbed microorganisms are considered to disperse together  
72 through inlet water flows to lakes; hence, sediment microorganisms may be suitable as  
73 informative indicators for sediment sources <sup>17</sup>. Furthermore, the introduction of high-  
74 throughput sequencing (HTS) technology provides a larger amount of data about the  
75 microbial community, which could improve discrimination between potential sources  
76 and facilitate the formation of source-specific fingerprints <sup>18, 19</sup>. The source  
77 contributions of lake microbial community from specific river environmental systems  
78 identified using microbial source tracking (MST) <sup>20</sup> may shed light on the above issues  
79 existing in common sediment-tracing methods. The autochthonous microbial  
80 community of lake sediments has been reported to be shaped by the deposition of  
81 particle-associated dispersed bacteria, particularly those capable of immediate growth  
82 in the recipient habitat and competing for nutrients <sup>21, 22</sup>, which contributes to the shifts  
83 in the physicochemical state and ecological environment of aquatic ecosystems. The  
84 transport of fecal bacteria and antibiotic resistance genes is also the leading cause of  
85 microbial contamination, exposing water-users to significant health risks and  
86 undermining the use of these systems for recreational and commercial activities <sup>23-25</sup>.  
87 Therefore, tracing sediment microorganisms in interconnected river-lake systems may  
88 not only offer a systematic and comprehensive understanding of sediment sources,

pathways, and transport dynamics, but also aid in the recognition and management of the ecological environmental impact of dispersed sediment microorganisms.

In this study, we applied SourceTracker, a representative MST method, to understand the dispersal dynamics of sediment microorganisms in interconnected river-lake systems. The method utilizes a Bayesian approach to estimate the proportion of contaminating sequences in a designated sink community that originates from possible source environments <sup>26</sup>. The accuracy of the overall source assignments is improved in this method by allowing unknown sources <sup>27</sup>. SourceTracker was originally designed to detect contamination in high-throughput sequence datasets <sup>26</sup>, but has interesting applications for identifying the sources of bacterial contamination within river-lake catchments <sup>21, 22</sup>. The seasonal variation in the importance of dispersal from different sources for the lake bacterioplankton community and population dynamics was determined using SourceTracker <sup>28</sup>. The practicability and feasibility of the results derived from this approach were evaluated and proved by comparing them with the results of a well-calibrated, three-dimensional hydrodynamic tracer model <sup>29</sup>. Compared with the water samples, sediment samples containing a higher biomass and microbe taxon richness <sup>30</sup> should provide more microbial information for SourceTracker. However, knowledge regarding the complete mechanism and significance of sediment MST is lacking. Owing to the existence of substantial indigenous microbial communities in sediments <sup>31, 32</sup>, sediment source apportionment estimates obtained using SourceTracker need to be modified to increase the pertinence and practicability of the predicted proportions for identifying major external sources.

Therefore, the aims of this study were to 1) determine the relationships between the fate of sediments, nutrient spatial distribution, and lake sediment biodiversity in river-lake systems; 2) evaluate the source contributions of the microbial community in lake sediment from inflow-rivers, employing the SourceTracker model with high-throughput DNA sequencing data, and 3) provide new insights into sediment source apportionment regarding the tracing of sediment microorganisms. To achieve these goals, Dongting Lake, the second largest freshwater lake in China, was selected as a study area as it exhibits intricate river-lake interactions with considerable sediment exchange<sup>33</sup>. The characteristics and interactions of the bacterial communities in sediments collected from the inflow-river estuaries and three major lake districts were determined by multivariate statistical analysis. The analysis using SourceTracker was modified by removing the indigenous microbial community determined via the linear discriminant analysis of effect sizes, and then used for tracing sediment sources. Our findings may provide improved scientific bases to estimate sediment sources and develop effective river-lake interaction regulation strategies.

## 2. MATERIALS AND METHODS

### 2.1. Study Area

Dongting Lake, the second largest freshwater lake in China, is located in the middle reaches of the Yangtze River (28°30'-30°20'N, 111°40'-113°10'E). Based on its geographical location and topography, Dongting Lake can be divided into three regions: East, South, and West Dongting Lake<sup>34</sup>. Dongting Lake receives discharge from the Yangtze River through three inlets located in the northwest catchment, including the



Songzi (SZ) , Hudu (HD) , and Ouchi (OC) Rivers, and the following tributary rivers: the Lishui (LS) and Yuanjiang (YJ) Rivers in West Dongting Lake, the Zishui (ZS), and Xiangjiang (XJ) Rivers in South Dongting Lake, and the Miluo (ML) and Xinqiang (XQ) Rivers in East Dongting Lake, and then discharges into the Yangtze River at the Chenglingji Outlet (OTL). This is a typical intricate river-lake interaction. Over 80% of the pollutant loads in Dongting Lake originate from considerable water and sediment import from inflow-rivers, which cause severe contamination and aggravated eutrophication <sup>35</sup>.

The sampling positions are shown in Figure 1. Locations L1, L2, L3, L4, and L5 are located in the central open water area of Dongting Lake, which are empirical sedimentation areas and remote from anthropogenic pollution, and thus were selected as the sink sites at which the sediment source contributions were quantified. The estuaries of the inflow-rivers, including locations LS, SZ, YJ, OC, ZS, XJ, ML, and XQ, were selected as the source sites as the abrupt slow flow in estuarine catchments causes the deposition of a large amount of sediments that can provide a record of the natural and anthropogenic inputs of contaminants into upstream rivers and form a concentrated influential expression of inflow-rivers on the lake <sup>36</sup>.

## 2.2. Sample Collection and Processing

Sampling was conducted in August 2017 (Tables S1), when the sediment discharge from inflow-rivers peaked and sediment transport processes began gradually stabilized. Water and sediment samples were collected at the same site. At each sampling site, water profiles were determined *in situ* by measuring the water temperature (T),

dissolved oxygen (DO), electric conductivity (Cond), and pH using a HACH HQ30d portable meter (HACH Company, Loveland, Co, USA). Water samples were collected in triplicate from just below the surface using a plexiglass water sampler (800, Jvchang, China), and then homogenized, filtered, and stored in 1-L polyethylene bottles. All water samples were transported on ice to the laboratory and stored at 4 °C to further characterize the water chemistry, including the total phosphorus (TP), nitrogen (TN), organic carbon (TOC), dissolved total phosphorus (DTP), particulate phosphorus (PP), chemical oxygen demand (COD<sub>Mn</sub>), and chlorophyll-a concentration (Chl-a), as previously described<sup>37-39</sup>.

Three sediment samples were collected at each site using a grab sampler (XDB0201, Pusen, China), from which the first 5 cm was sliced, mixed for homogenization, and stored in sterile 50-mL screw cap tubes. The remaining sediment (approximately 1 L) was combined and stored in a plastic bag, and then immediately transported to the laboratory in an ice box. The tubes were stored at -80 °C until the DNA could be extracted, and the sediment samples for physicochemical analysis were stored at -20 °C and freeze-dried. The organic matter (OM) content was calculated according to the loss on ignition to constant mass (4 h) at 550 °C. The total nitrogen (TN) content was determined using a continuous-flow automated analyzer (San++, Skalar, the Netherlands) after digestion with H<sub>2</sub>SO<sub>4</sub>-H<sub>2</sub>O<sub>2</sub><sup>37</sup>. The total phosphorus (TP), inorganic phosphorus (IP), organic phosphorus (OP), non-apatite inorganic phosphorus (NAIP; forms associated with oxides and hydroxides of Al, Fe, and Mn), and apatite phosphorus (AP; forms associated with Ca) contents were determined following the

Standards, Measurements, and Testing (SMT) harmonized procedure for phosphorus fractionation in freshwater sediments (SMTprotocol), which is a modified version of the Williams method. The phosphorus content of each fraction was spectrophotometrically determined following the vanadomolybdophosphate method.

### **2.3. DNA Extraction, PCR Amplification, and Sequencing Analysis**

DNA from each homogenized surface sediment sample (0.3 g) was extracted using the FastDNA spin kit for soil (Q-BIOgene, Carlsbad, CA), following the manufacturer's instructions. The quality of the extracted DNA was examined by agarose gel electrophoresis, and the DNA was stored at -20 °C until further analysis. PCR amplification and Illumina Miseq sequencing were conducted at Shanghai Biozeron Bio-Pharm Technology Co., Ltd. (Shanghai, China) using the 341F and 806R primers, targeting the V3 and V4 regions of the bacterial 16S rRNA gene. Amplicons were paired-end sequenced on the Illumina MiSeq PE250 platforms (Illumina, San Diego, CA).

After checking the quality, the raw Illumina paired-end reads that were over 75% of the amplicon length were trimmed off the barcodes and primers, and combined using the Flash software v1.2.7. The obtained raw sequence data in Fastq format were then demultiplexed and quality-filtered using QIIME v1.9.0. Operational taxonomic units (OTUs) were assigned using UCLUST based on 97% similarity identity and compared to the Silva v128 reference database using the PyNAST alignment algorithm. Taxonomic assignments were conducted against the RDP Classifier v2.2 with an 80% threshold. Data for all samples is available on the Sequence Read Archive

(<http://www.ncbi.nlm.nih.gov/sra/>) project reference SRP165943.

## **2.4. Statistical Analysis**

Statistical analyses were conducted using QIIME v1.9.0<sup>40</sup>, SPSS v20.0, RStudio v1.1.383, R v3.4.3. The trophic level index (TLI) method was followed to evaluate the eutrophication level in Dongting Lake. The alpha diversity of each sample was determined using the observed species, sample coverage, abundance-based coverage estimate (ACE), Chao 1 indices, Shannon diversity index, and Simpson diversity index based on the OTU analysis. The interactions between the alpha diversity and environmental factors were evaluated with Spearman correlation analysis (significance  $P < 0.05$ ). For comparison, a final sequence depth of 28,956 was selected for all subsequent statistical analysis. Analysis of Similarity (ANOSIM; Clarke, 1993) was conducted to identify significant differences between the bacterial communities (Beta diversity). Non-metric multidimensional scaling (NMDS) analysis, which was calculated based on the Bray-Curtis dissimilarity matrix, was used to visualize the dynamics in the bacterial community structures at the OTU-level taxa. Linear discriminant analysis of effect sizes (LEfSe) was conducted to identify highly differential taxonomy (from phylum to order level) between the samples from the central lake and inflow-rivers. Significant taxa were used to generate taxonomic cladograms and illustrate the differences between the source and sink sites.

## **2.5. SourceTracker Analysis**

The contributions of different sediment sources imported from the inflow-rivers to the bacterial community composition in the central lake sediments were predicted with

the Bayesian classifier software program SourceTracker (Knights et al., 2011) using R package SourceTracker v1.0.1. Based on its geographical location and topography, the SourceTracker analysis for Dongting Lake can be divided into three regions: West, South, and East Dongting Lake. Detailed arrangements of the designated sink and source sites in the three models are shown in Table S2. It should be noted that L2 was regarded as a sink site in West Dongting Lake and a source site in South Dongting Lake while L4 was regarded as a sink site in South Dongting Lake and a source site in East Dongting Lake.

The OTU tables derived from quality filtering and OTU picking were used as input files. SourceTracker analysis was conducted using default settings with a rarefaction depth of 1000, burn-in 100, restart 10, alpha (0.001), and beta (0.01), which has been evaluated and demonstrated to provide high sensitivity, specificity, accuracy, and precision in general <sup>41</sup>. For each source, the mean contribution proportion and standard deviation were calculated from the five independent SourceTracker runs to identify and prevent any potential false positive predictions <sup>41</sup>. The ratio of the averaged predicted proportion and standard deviation was then used to calculate the relative standard deviation (RSD) to estimate the confidence in the mean predicted source proportions and variation between different models <sup>27</sup>.

### **3. RESULTS AND DISCUSSION**

#### **3.1. Environmental Conditions**

The physiochemical characteristics of surface water and sediment samples retrieved from the fourteen investigated sites (Figure 1) are summarized in Table S3

and S4, respectively. Dongting Lake endured excessive nutrient loads (TN concentrations ranging from 1.18 to 1.76 mg/L and TP concentrations ranging from 0.06 to 0.21 mg/L in water samples) and accelerated eutrophication (TLI ranging from 45.67 to 57.59).

The results of the Mann-Whitney U test revealed that samples collected from the estuaries of inflow-rivers had a significantly higher concentration of TN (water samples) and a lower concentration of NAIP (sediment samples) than those from the central lake ( $P < 0.05$ ), which could be explained by the sediment transport processes in interconnected river-lake systems. Due to the high velocity and strong turbulence caused by complicated hydrodynamic conditions in estuaries, substantial sediment will be transported into the lake as SS while the adsorbed nutrients may be released into the water<sup>8</sup>. SS can be deposited as a result of a drop in stream velocity and turbulence that both keep sediment suspended in the central lake and the adsorbed nutrients may also accumulate at the bed surface<sup>5</sup>. Therefore, these results suggested that sediment transport, migration, and deposition might have significant effects on the spatial distribution of nutrients in interconnected river-lake systems.

### 3.2. Bacterial Community Composition and Diversity

Bacterial community profiles were generated for the 14 collected samples. After quality filtering and trimming, a total 609,565 sequences were obtained, which were assigned to 14,312 OTUs with a 97% sequence identity threshold. The bacterial community richness and diversity indices are shown in Table 1. The average sequencing coverage for the collected samples was 94.1%, and ranged from 90.5% to

96.9%. The results of the T-test revealed that sediment samples collected from the estuaries of inflow-rivers had a significantly lower bacterial community richness (i.e., ACE and Chao1 indices) than those from the central lake ( $P < 0.05$ ), which could be explained by the substantial amount of sediment deposition from multiple inflow-rivers in the central lake.

The samples were rarefied by random subsampling to 28,956 reads per sample for comparison. The bacterial community composition of the different samples at the phylum level is displayed in Figure 2. Across all samples, Proteobacteria (43.53% - 83.26%, mean 62.03%) were the most abundant phylum, which have been well observed in freshwater sediment environments. The other ubiquitous bacterial phyla of all samples were Bacteroidetes (9.93%), Acidobacteria (8.04%), Chloroflexi (5.21%), Firmicutes (4.04%), Actinobacteria (3.46%), and Nitrospirae (1.50%), which is consistent with the results reported in previous studies<sup>34</sup>.

The beta diversity was calculated using the Bray-Curtis distance. The NMDS plots showed that the bacterial community compositions exhibited a relatively clear separation between the samples from the estuaries of inflow-rivers (Figure 3). The samples from the central lake were clustered, which could be related to biotic homogenization due to the combined contributions of sediment migration and deposition. Although there were differences observed in the NMDS plots, the ANOSIM further identified that the bacterial community structures were not significantly different between the central lake and inflow-river estuary sediments ( $r = -0.04605$ ,  $P = 0.5988$ ), suggesting similarities and relationships between our defined sink and source

sites. It is important to note that higher variability in the bacterial communities within source sites than within sink sites, and between the source and sink sites, may indicate that adequate differentiation between these sources was possible by SourceTracker.

The interactions between the bacterial diversity, richness, and environmental factors in Dongting Lake were determined through Spearman rank correlation (Figure 4). Interestingly, the alpha diversity (OTU, Chao1, and Shannon index) of sediment samples also exhibited a significant negative correlation with particulate phosphorus in the corresponding water samples (Spearman rank correlative analysis,  $P < 0.05$ ). Previous research showed that particulate phosphorus is the dominant exchange form, exhibiting a significant linear relationship with suspended sediment in Dongting Lake ( $r^2 = 0.906$ )<sup>35</sup>. Therefore, this result may be explained by the enrichment of microbial diversity in the lake sediment due to the substantial deposition of suspended sediment and adsorbed microorganisms. Moreover, these results suggested that sediment microorganisms might serve as informative indicators for sediment sources due to the inseparable transport fate of sediments and adsorbed microorganisms.

### 3.3. Optimized Analysis Using SourceTracker

The taxa in an environmental sink sample that could not be assigned to a source were designated to an unknown category in SourceTracker analysis. These unknown taxa could be explained by dispersal from other potential sources that were not included in our study, or the existence of an indigenous microbial community in the sink sites. Compared with the water samples, sediment samples contain a higher biomass of autochthonous bacteria. Thus, to reduce the predicted unknown proportions caused by



309 autochthonous bacteria in sediment samples and increase the pertinence and  
310 practicability of the predicted proportions for identifying major external sources, the  
311 source apportionment estimates obtained using SourceTracker were modified.

312 LEfSe was applied to aid the identification of the indigenous microbial community  
313 in sink sites by determining highly differential taxonomy (from phylum to order level)  
314 between samples from the central lake and inflow-rivers. The Kruskal-Wallis test was  
315 conducted to identify taxa with significantly different abundances between the source  
316 and sink sites, followed by LDA to estimate the effect size of each differentially  
317 abundant feature. A cladogram showing the phylogenetic distribution of microbial  
318 lineages significantly associated (LDA values  $> 2$ ) with sediments from the sink (red)  
319 and source sites (green) is presented in Figure 5 (a). The corresponding LDA value for  
320 each lineage is shown in Figure 5 (b). LEfSe analysis revealed that the bacterial lineages  
321 that were mainly enriched in sink sites included Nitrospirales, Nitrospira, Nitrospirae,  
322 Epsilonproteobacteria, and Campylobacteriales, while Alphaproteobacteria,  
323 Caulobacteriales, and YS2 exhibited a significant presence in source sites, as shown in  
324 Figure 5 (b). The OTUs assigned to the highly differential taxa were identified as the  
325 indigenous microbial community in sink sites and then removed in SourceTracker  
326 analysis to increase the pertinence and practicability of the predicted proportions for  
327 identifying major external sources.

328 The respective contributions of different sediment sources imported from the  
329 inflow-rivers to the bacterial community composition of central lake sediments were  
330 evaluated using the Bayesian classifier software program SourceTracker v1.0.1.

Original and optimized analyses using SourceTracker were conducted using default settings with five independent runs. For each source, the mean contribution proportion and the RSD were calculated from the five independent SourceTracker runs using the default settings. Highly relevant results between two analyses could be observed by conducted Spearman rank correlative analysis ( $r_{\text{spearman}} = 0.985$ ,  $p < 0.01$ ). The difference of the source proportional contributions derived from the original and optimized analysis could be seen in Table 2. The results of the optimized analysis suggested that the predicted unknown proportions significantly decreased (paired-samples t-test,  $P < 0.01$ ). The shifts in the ranking of the different source contributions were also observed. For instance, the ranking of the different source contributions in L1 was  $SZ > LS = YJ$  in the original analysis while  $SZ > LS > YJ$  in the optimized analysis.

#### **3.4. Sediment Source Apportionment Estimates Using SourceTracker**

The quality-filtered OTU tables (i.e., OTUs simultaneously identified at the source and sink sites) were utilized as input files. According to the source apportionments by the optimized microbial community MST, the contributions to the bacterial community of L1 in descending order were SZ (24.2%), LS (20.6%), and YJ (17.4%), with a RSD ranging from 8.5% to 11.9%, while the contributions to L2 were YJ (30.2%), SZ (27.8%), and LS (6.6%), with a RSD ranging from 4.7% to 31.4% in West Dongting Lake. The microorganisms observed at L2 were the major contributor to the sediment bacteria community of South Dongting Lake (33.0% and 17.4% for L3 and L4, respectively). The source site ML (28.4%, RSD 12.8%) also significantly seeded the

sediment bacterial community of L4. The major contributor to the bacterial community of East Dongting Lake was that of XQ (18.4%, RSD 8.2%). The remaining percentage contributions, standard deviations, and RSD in South and East Dongting Lake are provided in Figure 6 and Table 2. A considerable portion of the SourceTracker-estimated contributions to the sink sites remained unknown, with unknown contributions ranging from 23.8% to 61.2%. The results of Spearman rank correlative analysis demonstrated a significant negative correlation between the RSD values and the size of the predicted SourceTracker proportions ( $r_{\text{spearman}} = -0.884$ ,  $p < 0.01$ ), indicating that larger SourceTracker proportions have higher confidence. This trend is in accordance with previous findings <sup>41</sup>.

Although the technical replicates through the multiple SourceTracker model runs and low RSD (2.4% - 40.8%, mean 14.9%) guaranteed the stability of the prediction results, it is difficult to validate the reliability and feasibility of the estimated source proportions using independent evidence. Traditional monitoring and measurement techniques face their own limitations in terms of the practicalities and costs of deployment, both spatially and temporally. An unremarkable between-source group variability in the physicochemical tracer concentrations limited the applicability of common sediment source fingerprinting in this study. However, the reasonability and validity of this method are demonstrated in the following discussion.

The SourceTracker results revealed that the unknown contributions in West and South Dongting Lake ranged from 23.8% to 38.0%. Compared with the limited influence of the Yangtze River, the YJ, ZS, ML, and XQ tributaries were the major

contributors to the sediment bacteria community in the three regions of Dongting Lake, according to the source apportionments by SourceTracker. The evaluated percentage contributions of SZ and OC, which were attributed to Yangtze River input, ranged from 24.2% to 27.8% in West Dongting Lake, and 8.8% to 6.4% in South Dongting Lake. These results were in accordance with the actual situation. Dongting Lake receives over 80% of its pollutant loads by water and sediment discharge from the Yangtze River through three inlets and upstream tributaries in the lake basin<sup>35</sup>. With the operation of the Three Gorges Dam (TGR) and intensified human activity in the Dongting Lake Basin, tremendous evolution of river–lake interactions has been observed, accompanied by changes in the regional water regime, hydrodynamic characteristics, sediment transportation, spatial distribution of nutrients, biodiversity, and ecological conditions<sup>42</sup>. Substantial sediments were retained in the Three Gorges Reservoir, which has led to a reduction in the sediment input from the Yangtze River through the three inlets to Dongting Lake (the first large lake downstream of the TGR)<sup>35, 43</sup>. The contributions of sediment input from the three inlets to Dongting Lake decreased sharply from 80% to 20%, according to the Changjiang Sediment Bulletin (2017).

### 3.5. Factors Determining Sediment Source Apportionment

Hydrodynamic conditions have important impacts on sediment transport, which were further verified by the SourceTracker results. With increasing sediment diffusion distance in our research area, the contributions of source sites to Dongting Lake decreased significantly ( $r_{\text{spearman}} = -0.530$ ,  $p = 0.029$ ). The contribution of LS to West Dongting Lake decreased from L1 (20.6%) to the relatively remote site of L2 (6.6%).

Correspondingly, a marked decline in the contribution of ML to L2 was observed with an increase in the dispersed distance. This could be explained by the gradual deposition of large amounts of sediment from inflow-rivers due to energy decay of water flow along the imported waterway in Dongting Lake. Water flow direction codetermined by the river-lake hydrodynamics and lake morphometry also have important influences on source contributions. For example, despite the relatively remote nature of L2 compared with L1 (17.4%), the site exhibited a higher predicted source contribution from YJ (30.2%), which may be explained by the specific directional introduction of estuarine plumes (Yuanjiang River) to West Dongting Lake.

### **3.6. Implications for Sediment Source Tracking**

Owing to the multi-source interactions and spatio-temporal variation in the transport and fate of sediments in interconnected river-lake systems, directly identifying the sediment sources of contamination and implementing targeted mitigation strategies are a persistent challenging topic in sediment control. By combining microbial community profiling with the machine learning classification program SourceTracker, we provide new insights into sediment source apportionment. Advantages include the limited amount of field data collection required (i.e. collection of source material and sediment samples as opposed to expensive long-term monitoring at a wide range of sites) and the range of complementary information that can be obtained. It is essential to consider what degree sample size and sampling time required to reduce uncertainty and deviation in the SourceTracker analyses. Following the predicted source contribution status, the risk of different inflow-rivers in diffuse sediment pollution can

be ranked, thereby allowing priorities to be established when mitigating sediment spread and designing effective control strategies. The common sediment source fingerprinting technique is a well-used tool for evaluating sediment sources in catchments at a river-basin scale. After determining the dominant inflow-rivers of sediment dissemination, the sediment source-fingerprinting technique could be applied to determine the relative importance of potential sediment sources while focusing on source types according to actual regional circumstances (i.e., land use, lithology or erosional processes). This novel framework affords a valuable and effective alternative approach to the design of effective sediment management and control strategies in interconnected river-lake systems.

Recently, new methods have been developed that amplicon sequence variants (ASVs) can be resolved exactly without imposing the arbitrary dissimilarity thresholds that define OTUs <sup>44, 45</sup>. Multiple studies suggest that the improvements in reusability, reproducibility and comprehensiveness are sufficiently great that ASVs should replace OTUs as the standard unit of marker-gene analysis and reporting <sup>46, 47</sup>. Therefore, the analyses based on ASVs were also carried out through the DADA2 method in QIIME2 and corresponding results were shown in the Supporting Information. Similar bacterial community composition (Figure S1 and S2) and diversity (Figure S3 and S4) were determined through comparing the results of analyses based on the OTU and ASV methods. The resulting ASV table was used in the SourceTracker analysis. Unexpectedly, results of SourceTracker analysis showed that 100% sink microbiome was unknown in all independent runs. After checking the resulting ASV table, it was

found that only a few overlaps of ASVs in source and sink environments were determined, which may be the reason for the limitation of the applicability of SourceTracker analysis. As is known to all, SourceTracker is a Bayesian approach to identify sources and proportions of contamination through modelling contamination as the mixture of entire source communities into a sink community which requires modest overlapping of species between sink and source samples<sup>26</sup>. The situation of the QIIME2 deblur aims at finding the exact sequence variants<sup>45</sup>. Fewer overlaps in the species caused by finer resolution might make SourceTracker analysis out of operation. Research to date has not employed the SourceTracker analysis based on ASV method. Compared with the ASV method, the practicability and feasibility of the results derived from the SourceTracker analysis based on OTU method have already been evaluated and proved by the previous research<sup>27, 29, 48, 49</sup>. Therefore, we recommend employing SourceTracker analysis based on the OTU method to estimate dominant sediment sources in interconnected river-lake systems. ASVs should replace OTUs as the standard unit of marker-gene analysis and reporting, however, few overlaps in the species should be considered especially when the ASV tables are directly used to compare bioinformatics of samples.

In conclusion, we have determined the significant effects of sediment transport, migration, and deposition on the spatial distribution of nutrients, ecological conditions, and sediment biodiversity in interconnected river-lake systems. To the best of our knowledge, the MST method was first used to trace sediment sources in interconnected river-lake systems. SourceTracker was determined to be capable of distinguishing and

appointing sediment sources. Modified analysis using SourceTracker found that sediment import from the Yangtze River has declined remarkably with the evaluation of intricate river-lake interactions in the Dongting Lake Basin. Moreover, the important contributions of hydrodynamic conditions to source apportionment were also determined in this study. However, there are several limitations for accurate identification in SourceTracker analysis, including the inadequate source and sink samples derived from snapshot sampling, the lack of multiple hydrological periods to interpret significant temporal variation in sediment sources, and the succession of sediment microbes dependent on environmental factors during sediment transport, which should be further investigated. As such, an improved understanding of sediment sources through a novel framework for sediment source-tracking in interconnected river-lake systems could assist watershed managers in developing effective sediment management strategies.

## **SUPPORTING INFORMATION**

Four tables are provided in the Supporting Information section to present information on sampling events included in this study, the detailed arrangements of designated sink and source sites in SourceTracker analysis, physiochemical characteristics of the surface water and sediment samples retrieved from the fourteen investigated sites. The results and comparisons of the analyses in terms of Amplicon Sequence Variants (ASVs) and Operational taxonomic units (OTUs) are provided in the Supporting Information.

## **ACKNOWLEDGMENTS**

This study was financially supported by the National Natural Science Foundation of



China (51879079), the Fundamental Research Funds for the Central Universities (2018B14614 and 2018B42314), the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (51421006), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJKY19\_0535), the China Postdoctoral Science Foundation (2015M580389), the Priority Academic Program Development of Jiangsu Higher Education Institutions, and the Top-Notch Academic Programs Project of Jiangsu Higher Education Institutions (TAPP).

## REFERENCES

- (1) Gao, J. H.; Jia, J.; Kettner, A. J.; Xing, F.; Wang, Y. P.; Xu, X. N.; Yang, Y.; Zou, X. Q.; Gao, S.; Qi, S., Changes in water and sediment exchange between the Changjiang River and Poyang Lake under natural and anthropogenic conditions, China. *Sci. Total Environ.* **2014**, *481* (1), 542-553.
- (2) Lesack, L. F. W.; Melack, J. M., Flooding hydrology and mixture dynamics of lake water derived from multiple sources in an Amazon floodplain lake. *Water Resour. Res.* **1995**, *31* (2), 329-345.
- (3) Milner, A. M.; Robertson, A. L.; Mcdermott, M. J.; Klaar, M. J.; Brown, L. E., Major flood disturbance alters river ecosystem evolution. *Nat. Clim. Change* **2013**, *3* (2), 137-141.
- (4) Ye, X.; Zhang, Q.; Liu, J.; Li, X.; Xu, C. Y., Distinguishing the relative impacts of climate change and human activities on variation of streamflow in the Poyang Lake catchment, China. *J. Hydrol.* **2013**, *494* (12), 83-95.

- (5) Vercruysse, K.; Grabowski, R. C.; Rickson, R. J., Suspended sediment transport dynamics in rivers: Multi-scale drivers of temporal variation. *Earth Sci. Rev.* **2017**, *166*, 38-52.
- (6) Jamieson, R.; Joy, D. M.; Lee, H.; Kostaschuk, R.; Gordon, R., Transport and deposition of sediment-associated *Escherichia coli* in natural streams. *Water Res.* **2005**, *39* (12), 2665-2675.
- (7) Withers, P. J.; Jarvie, H. P., Delivery and cycling of phosphorus in rivers: a review. *Sci. Total Environ.* **2008**, *400* (1), 379-395.
- (8) Bilotta, G. S.; Brazier, R. E., Understanding the influence of suspended solids on water quality and aquatic biota. *Water Res.* **2008**, *42* (12), 2849-2861.
- (9) Scott, T. M.; Rose, J. B.; Jenkins, T. M.; Farrah, S. R.; Lukasik, J., Microbial source tracking: current methodology and future directions. *Appl. Environ. Microb.* **2002**, *68* (12), 5796-5803.
- (10) Collins, A. L.; Walling, D. E.; Leeks, G., Source type ascription for fluvial suspended sediment based on a quantitative composite fingerprinting technique. *Catena* **1997**, *29* (1), 1-27.
- (11) Walling, D. E.; Peart, M. R.; Oldfield, F.; Thompson, R., Suspended sediment sources identified by magnetic measurements. *Nature* **1979**, *281* (5727), 110-113.
- (12) Gibbs, M. M., Identifying source soils in contemporary estuarine sediments: a new compound-specific isotope method. *Estuar. Coast.* **2008**, *31* (2), 344-359.
- (13) Nosrati, K.; Govers, G.; Ahmadi, H.; Sharifi, F.; Amoozegar, M. A.; Merckx, R.; Vanmaercke, M., An exploratory study on the use of enzyme activities as sediment

- 529 tracers:biochemical fingerprints? *Int. J. Sediment Res.* **2011**, *26* (2), 136-151.
- 530 (14) Vale, S. S.; Fuller, I. C.; Procter, J. N.; Basher, L. R.; Smith, I. E., Characterization  
531 and quantification of suspended sediment sources to the Manawatu River, New  
532 Zealand. *Sci. Total Environ.* **2016**, *543* (Pt A), 171-186.
- 533 (15) Walling, D. E., Tracing suspended sediment sources in catchments and river  
534 systems. *Sci. Total Environ.* **2005**, *344* (1), 159-184.
- 535 (16) Pulley, S.; Foster, I.; Antunes, P., The uncertainties associated with sediment  
536 fingerprinting suspended and recently deposited fluvial sediment in the Nene river  
537 basin. *Geomorphology* **2015**, *228* (228), 303-319.
- 538 (17) Droppo, I. G., Rethinking what constitutes suspended sediment. *Hydrol. Processes*  
539 **2001**, *15* (9), 1551-1564.
- 540 (18) Kircher, M.; Kelso, J., High-throughput DNA sequencing--concepts and  
541 limitations. *Bioessays* **2010**, *32* (6), 524-536.
- 542 (19) Bamshad, M. J.; Ng, S. B.; Bigham, A. W.; Tabor, H. K.; Emond, M. J.; Nickerson,  
543 D. A.; Shendure, J., Exome sequencing as a tool for Mendelian disease gene  
544 discovery. *Nat. Rev. Genet.* **2011**, *12* (11), 745-755.
- 545 (20) Staley, C.; Reckhow, K. H.; Lukasik, J.; Harwood, V. J., Assessment of sources of  
546 human pathogens and fecal contamination in a Florida freshwater lake. *Water Res.*  
547 **2012**, *46* (17), 5799-5812.
- 548 (21) Adams, H. E.; Crump, B. C.; Kling, G. W., Metacommunity dynamics of bacteria  
549 in an arctic lake: the impact of species sorting and mass effects on bacterial  
550 production and biogeography. *Frontiers in Microbiology* **2013**, *5* (3), 82.

- 551 (22) Crump, B. C.; Amaral-Zettler, L. A.; Kling, G. W., Microbial diversity in arctic  
552 freshwaters is structured by inoculation of microbes from soils. *ISME J.* **2012**, *6* (9),  
553 1629-1639.
- 554 (23) Mallin, M. A.; Williams, K. E.; Esham, E. C.; Lowe, R. P., Effect of human  
555 development on bacteriological water quality in coastal watersheds. *Ecol. Appl.* **2000**,  
556 *10* (4), 1047-1056.
- 557 (24) Green, H. C.; Shanks, O. C.; Sivaganesan, M.; Haugland, R. A.; Field, K. G.,  
558 Differential decay of human faecal *Bacteroides* in marine and freshwater. *Environ.*  
559 *Microbiol.* **2011**, *13* (12), 3235-3249.
- 560 (25) Brown, C. M.; Staley, C.; Wang, P.; Dalzell, B.; Chun, C. L.; Sadowsky, M. J., A  
561 High-Throughput DNA-Sequencing Approach for Determining Sources of Fecal  
562 Bacteria in a Lake Superior Estuary. *Environ. Sci. Technol.* **2017**, *51* (15), 8263-8271.
- 563 (26) Knights, D.; Kuczynski, J.; Charlson, E. S.; Zaneveld, J.; Mozer, M. C.; Collman,  
564 R. G.; Bushman, F. D.; Knight, R.; Kelley, S. T., Bayesian community-wide culture-  
565 independent microbial source tracking. *Nat. Methods* **2011**, *8* (9), 761-763.
- 566 (27) Staley, C.; Kaiser, T.; Lobos, A.; Ahmed, W.; Harwood, V. J.; Brown, C. M.;  
567 Sadowsky, M. J., Application of SourceTracker for accurate identification of fecal  
568 pollution in recreational freshwater: a double-blinded study. *Environ. Sci. Technol.*  
569 **2018**, *52* (7), 4207-4217.
- 570 (28) Comte, J.; Berga, M.; Severin, I.; Logue, J. B.; Lindstrom, E. S., Contribution of  
571 different bacterial dispersal sources to lakes: Population and community effects in  
572 different seasons. *Environ. Microbiol.* **2017**, *19* (6), 2391-2404.

- (29) McCarthy, D. T.; Jovanovic, D.; Lintern, A.; Teakle, I.; Barnes, M.; Deletic, A.; Coleman, R.; Rooney, G.; Prosser, T.; Coutts, S.; Hipsey, M. R.; Bruce, L. C.; Henry, R., Source tracking using microbial community fingerprints: Method comparison with hydrodynamic modelling. *Water Res.* **2017**, *109*, 253-265.
- (30) Zinger, L.; Amaral-Zettler, L. A.; Fuhrman, J. A.; Hornerdevine, M. C.; Huse, S. M.; Welch, D. B. M.; Martiny, J. B. H.; Sogin, M.; Boetius, A.; Ramette, A., Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *Plos One* **2011**, *6* (9), e24570.
- (31) Lucie, Z.; Amaral-Zettler, L. A.; Fuhrman, J. A.; M Claire, H. D.; Huse, S. M.; Welch, D. B. M.; Martiny, J. B. H.; Mitchell, S.; Antje, B.; Alban, R., Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *Plos One* **2011**, *6* (9), e24570.
- (32) Yi, L.; Chao, W.; Wenlong, Z.; Peifang, W.; Lihua, N.; Jun, H.; Jing, W.; Linqiong, W., Modeling the Effects of Hydrodynamic Regimes on Microbial Communities within Fluvial Biofilms: Combining Deterministic and Stochastic Processes. *Environ. Sci. Technol.* **2015**, *49* (21), 12869-12878.
- (33) Feng, L.; Hu, C.; Chen, X.; Zhao, X., Dramatic inundation changes of China's two largest freshwater lakes linked to the Three Gorges Dam. *Environ. Sci. Technol.* **2013**, *47* (17), 9628-9634.
- (34) Huang, W.; Jiang, X., Profiling of sediment microbial community in Dongting Lake before and after impoundment of the Three Gorges Dam. *Int.. J. Env. Res. Pub. He.* **2016**, *13* (6), 617-631.

- 595 (35) Tian, Z.; Zheng, B.; Wang, L.; Li, H.; Wang, X., Effects of river-lake interactions  
596 in water and sediment on phosphorus in Dongting Lake, China. *Environ. Sci. Pollut.*  
597 *Res.* **2017**, *24* (29), 23250-23260.
- 598 (36) Ip, C. C.; Li, X. D.; Zhang, G.; Farmer, J. G.; Wai, O. W.; Li, Y. S., Over one  
599 hundred years of trace metal fluxes in the sediments of the Pearl River Estuary, South  
600 China. *Environ. Pollut.* **2004**, *132* (1), 157-172.
- 601 (37) Wu, H.; Li, Y.; Zhang, J.; Niu, L.; Zhang, W.; Cai, W.; Zhu, X., Sediment bacterial  
602 communities in a eutrophic lake influenced by multiple inflow-rivers. *Environ. Sci.*  
603 *Pollut. Res.* **2017**, *24* (24), 19795-19806.
- 604 (38) Cai, W.; Li, Y.; Niu, L.; Zhang, W.; Wang, C.; Wang, P.; Meng, F., New insights  
605 into the spatial variability of biofilm communities and potentially negative bacterial  
606 groups in hydraulic concrete structures. *Water Res.* **2017**, *123*, 495-504.
- 607 (39) Li, Y.; Wu, H.; Shen, Y.; Wang, C.; Wang, P.; Zhang, W.; Gao, Y.; Niu, L.,  
608 Statistical determination of crucial taxa indicative of pollution gradients in sediments  
609 of Lake Taihu, China. *Environ. Pollut.* **2019**, *246*, 753-762.
- 610 (40) J Gregory, C.; Kyle, B.; Bushman, F. D.; Desantis, T. Z.; Andersen, G. L.; Rob,  
611 K., PyNAST: a flexible tool for aligning sequences to a template alignment.  
612 *Bioinformatics* **2010**, *26* (2), 266-267.
- 613 (41) Henry, R.; Schang, C.; Coutts, S.; Kolotelo, P.; Prosser, T.; Crosbie, N.; Grant, T.;  
614 Cottam, D.; O'Brien, P.; Deletic, A.; McCarthy, D., Into the deep: Evaluation of  
615 SourceTracker for assessment of faecal contamination of coastal waters. *Water Res.*  
616 **2016**, *93*, 242-253.

- 617 (42) Gao, B.; Yang, D.; Yang, H., Impact of the Three Gorges Dam on flow regime in  
618 the middle and lower Yangtze River. *Quat. Int.* **2013**, *304* (447), 43-50.
- 619 (43) Dai, Z.; Liu, J. T.; Wei, W.; Chen, J., Detection of the Three Gorges Dam influence  
620 on the Changjiang (Yangtze River) submerged delta. *Sci. Rep.* **2014**, *4* (1), 6600.
- 621 (44) Eren, A. M.; Maignien, L.; Sul, W. J.; Murphy, L. G.; Grim, S. L.; Morrison, H.  
622 G.; Sogin, M. L., Oligotyping: Differentiating between closely related microbial taxa  
623 using 16S rRNA gene data. *Methods in Ecology & Evolution* **2013**, *4* (12), 1111-  
624 1119.
- 625 (45) Callahan, B. J.; Mcmurdie, P. J.; Rosen, M. J.; Han, A. W.; Johnson, A. J.; Holmes,  
626 S. P., DADA2: High-resolution sample inference from Illumina amplicon data. *Nat.*  
627 *Methods* **2016**, *13* (7), 581-583.
- 628 (46) Callahan, B. J.; McMurdie, P. J.; Holmes, S. P., Exact sequence variants should  
629 replace operational taxonomic units in marker-gene data analysis. *ISME J.* **2017**, *11*  
630 (12), 2639-2643.
- 631 (47) Eren, A. M.; Morrison, H. G.; Lescault, P. J.; Reveillaud, J.; Vineis, J. H.; Sogin,  
632 M. L., Minimum entropy decomposition: Unsupervised oligotyping for sensitive  
633 partitioning of high-throughput marker gene sequences. *ISME J.* **2015**, *9* (4), 968-  
634 979.
- 635 (48) Li, L. G.; Yin, X. L.; Zhang, T., Tracking antibiotic resistance gene pollution from  
636 different sources using machine-learning classification. *Microbiome* **2018**, *6* (1), 93-  
637 104.
- 638 (49) Baral, D.; Dvorak, B. I.; Admiraal, D.; Jia, S. G.; Zhang, C.; Li, X., Tracking the

639 Sources of Antibiotic Resistance Genes in an Urban Stream during Wet Weather  
640 using Shotgun Metagenomic Analyses. *Environ. Sci. Technol.* **2018**, 52 (16), 9033-  
641 9044.  
642



Figure captions

Figure 1. Location of the sampling sites in Dongting Lake, China. Black points - sampling sites in the estuaries of inflow-rivers; red stars - sampling sites in central Dongting Lake; red triangle - Chenglingji outlet (OTL).

Figure 2. Bacterial community profiles based on the relative abundance of different phyla and classes of Proteobacteria at the different study sites. Low abundance phyla (< 1% in all samples) are not represented.

Figure 3. Non-metric multidimensional scaling (NMDS) ordination visualization of the bacterial community compositions (Bray-Curtis distances) sampled in the estuaries of inflow-rivers and central Dongting Lake.

Figure 4. Correlation between the sediment bacteria's alpha diversity (OTU, ACE, Chao1, Coverage, Simpson, and Shannon), aquatic environmental variables (COD, Chl-a, TOC, TN, TP, DTP, PP, TLI), and aquatic sediment environmental variables (OM, TN, TP) in Dongting Lake. TNS and TPS represent the TN and TP of sediment samples, respectively. This plot leaves blank on no significant coefficient ( $P > 0.05$ ).

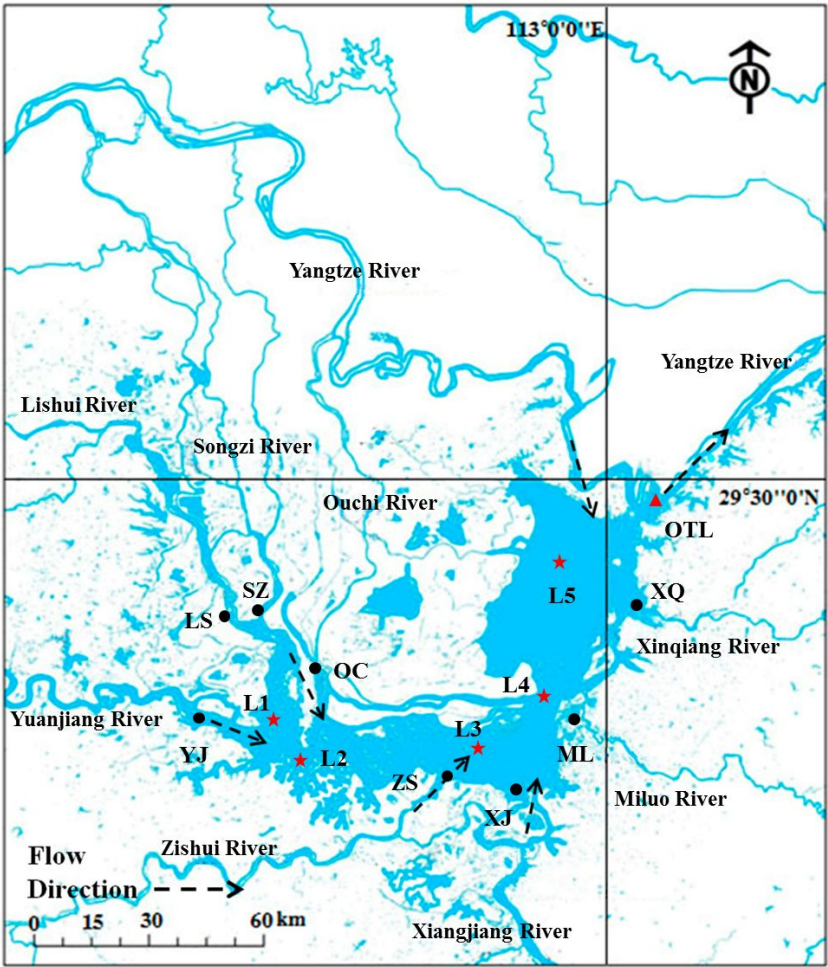
Figure 5. (a) Cladogram indicating the phylogenetic distribution of microbial lineages significantly associated with the sediments from sink (red) and source (green) sites, and (b) highly differential taxonomy with LDA values greater than 2.

Figure 6. Results of the optimized SourceTracker analysis showing the contribution of different source communities to three regions: West, South, and East Dongting Lake.

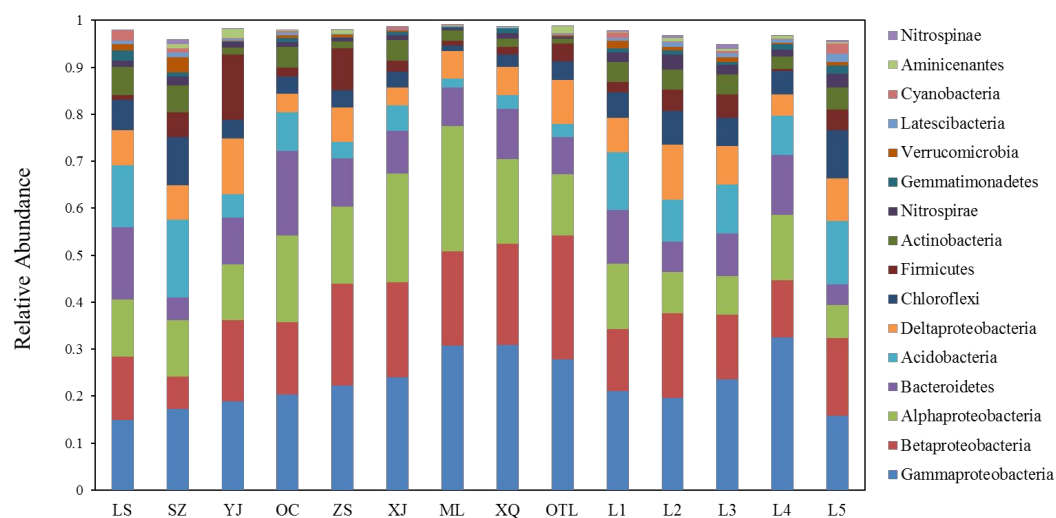
Table captions

Table 1. Alpha diversity estimates of bacterial communities.

Table 2. Mean percentage contributions and relative standard deviations (RSD) of the potential sources for sediment bacteria in Dongting Lake calculated by different SourceTracker models.



677 Figure 1. Location of the sampling sites in Dongting Lake, China. Black points - sampling sites in  
678 the estuaries of inflow-rivers; red stars - sampling sites in central Dongting Lake; red triangle -  
679 Chenglingji outlet (OTL).



680 Figure 2. Bacterial community profiles based on the relative abundance of different phyla and  
 681 classes of Proteobacteria at the different study sites. Low abundance phyla (< 1% in all samples)  
 682 are not represented.

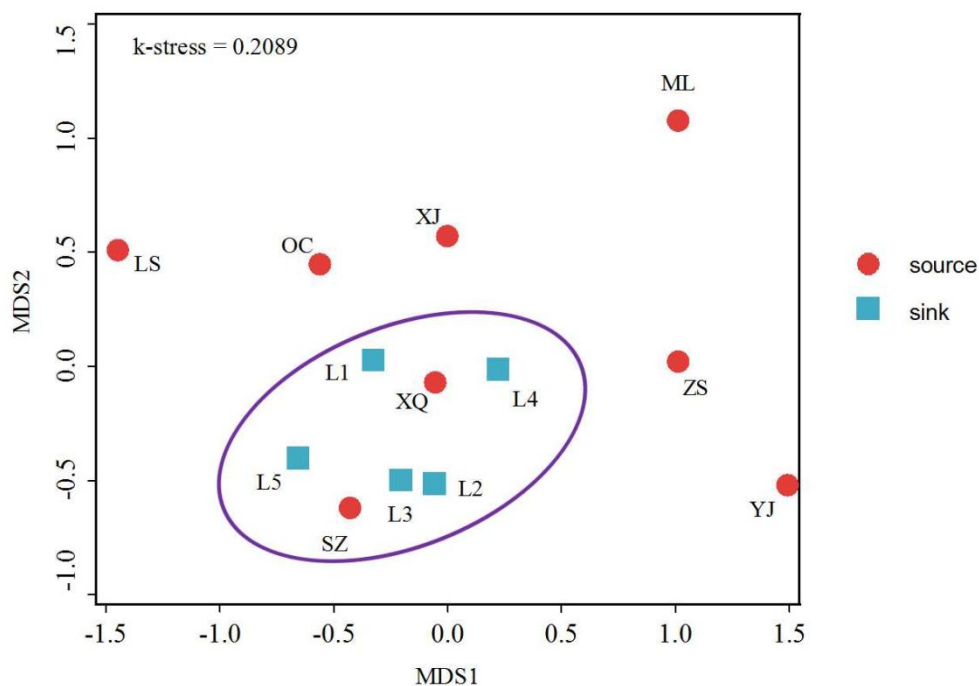
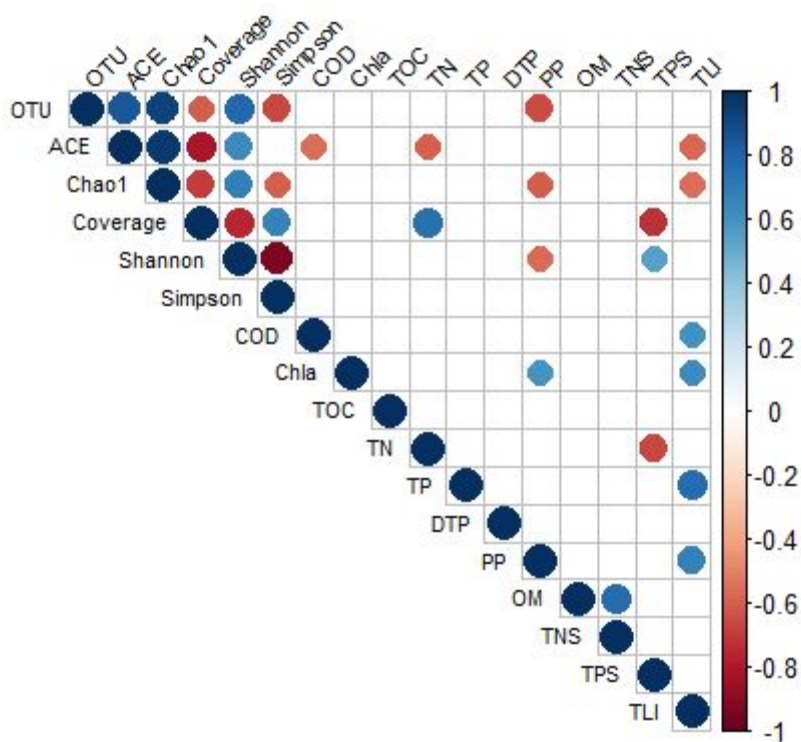
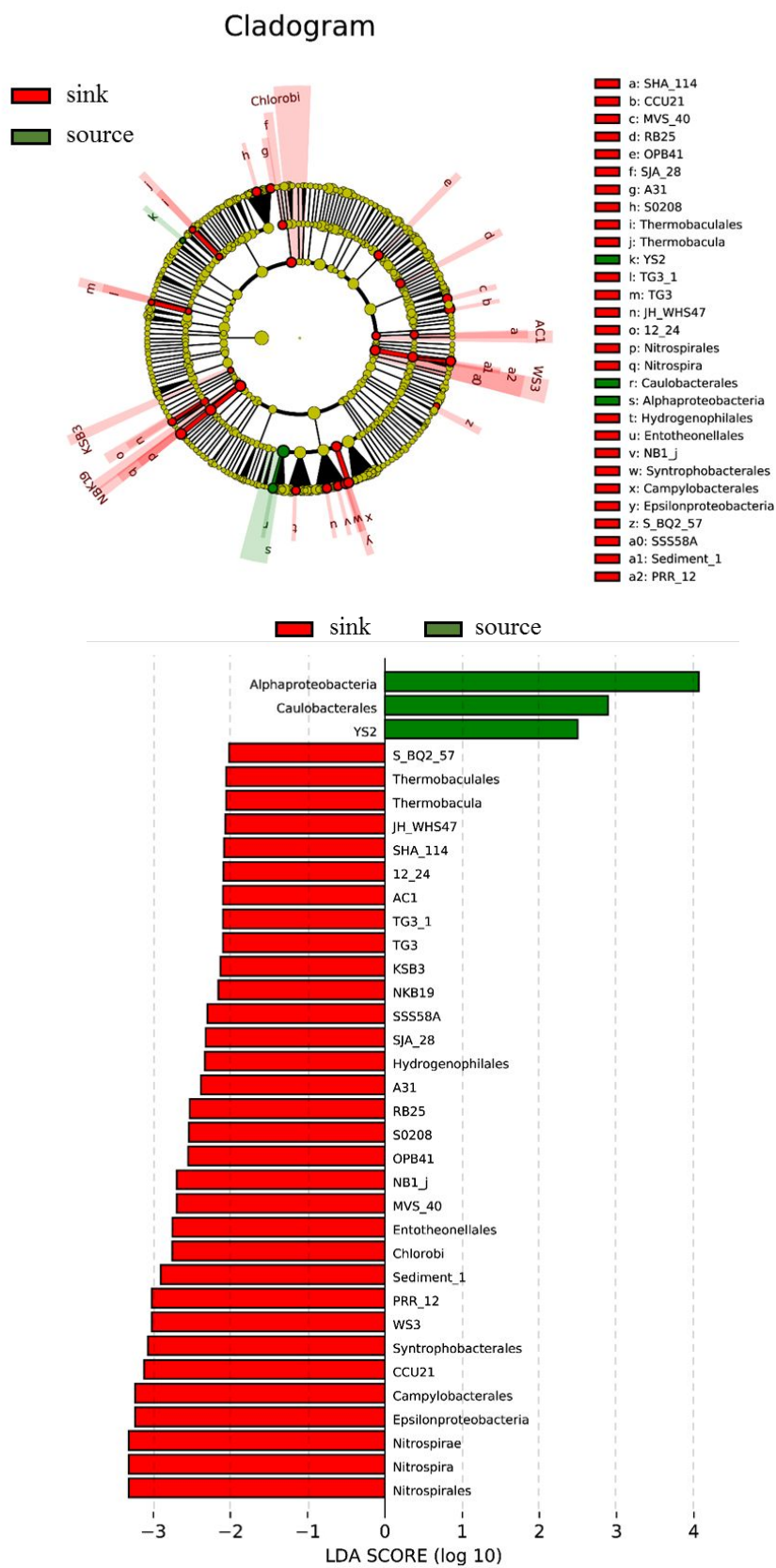


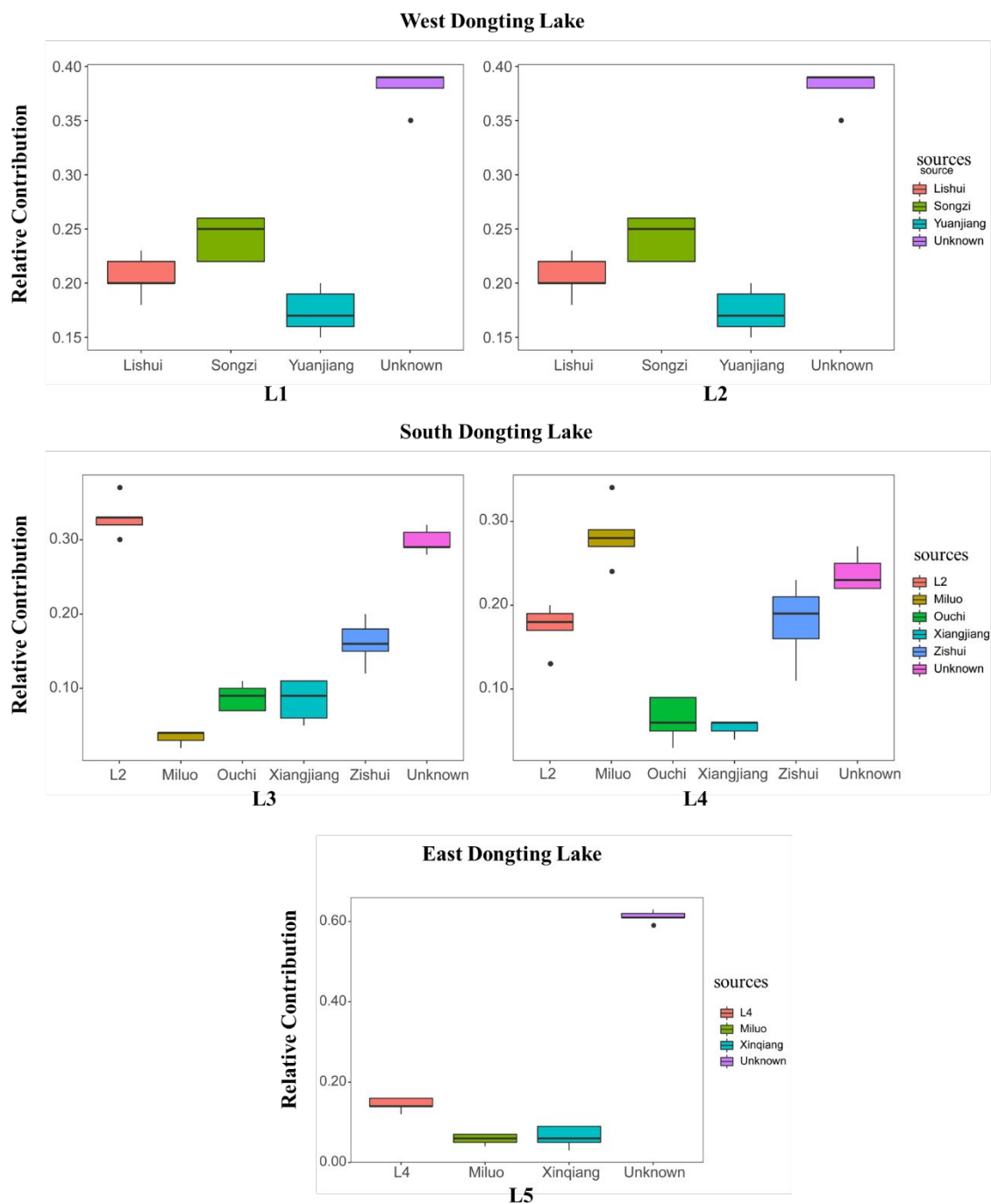
Figure 3. Non-metric multidimensional scaling (N-MDS) ordination visualization of the bacterial community compositions (Bray-Curtis distances) sampled in the estuaries of inflow-rivers and central Dongting Lake.



686 Figure 4. Correlation between the sediment bacteria's alpha diversity (OTU, ACE, Chao1,  
 687 Coverage, Simpson, and Shannon), aquatic environmental variables (COD, Chl-a, TOC, TN, TP,  
 688 DTP, PP, TLI), and aquatic sediment environmental variables (OM, TN, TP) in Dongting Lake.  
 689 TNS and TPS represent the TN and TP of sediment samples, respectively. This plot leaves blank  
 690 on no significant coefficient ( $P > 0.05$ ).



691 Figure 5. (a) Cladogram indicating the phylogenetic distribution of microbial lineages  
692 significantly associated with the sediments from sink (red) and source (green) sites, and (b) highly  
693 differential taxonomy with LDA values greater than 2.



694 Figure 6. Results of SourceTracker analysis showing the contribution of different source  
 695 communities to three regions: West, South, and East Dongting Lake.

696     Table 1. Alpha diversity estimates of bacterial communities

Sample ID	Reads	OTU	ACE	Chao1	Coverage	Shannon	Simpson
LS	53,020	6,307	11,225	9,664	0.950	7.23	0.0038
SZ	30,942	5,503	11,788	9,218	0.914	7.26	0.0035
YJ	30,881	2,887	7,206	5,015	0.952	5.58	0.0176
OC	48,052	6,162	12,741	10,373	0.939	7.13	0.0029
ZS	43,359	4,086	9,775	7,133	0.951	6.11	0.0102
XJ	52,656	5,074	9,832	7,986	0.957	6.51	0.0085
ML	57,436	3,548	8,020	6,193	0.969	5.14	0.0289
XQ	43,941	4,386	10,327	7,745	0.949	5.73	0.0251
OTL	40,188	3,039	7,321	5,497	0.961	5.39	0.0200
L1	54,350	7,318	14,677	11,841	0.938	7.20	0.0043
L2	44,226	6,147	13,157	10,367	0.933	6.87	0.0081
L3	29,268	5,161	12,943	9,687	0.905	6.87	0.0081
L4	52,290	5,474	12,047	9,390	0.949	5.90	0.0274
L5	28,956	5,271	11,622	8,885	0.911	7.35	0.0022

697



Table 2. Mean percentage contributions and relative standard deviations (RSD) of the potential sources for sediment bacteria in Dongting Lake calculated by SourceTracker

Sink	Source	Original analysis		Optimized analysis	
		Proportion (%)	RSD (%)	Proportion (%)	RSD (%)
West Dongting Lake					
L1	LS	17.8%	10.8%	20.6%	9.5%
	SZ	24.4%	6.9%	24.2%	8.5%
	YJ	17.8%	7.3%	17.4%	11.9%
	Unknown	40.0%	2.8%	37.8%	4.6%
L2	LS	6.2%	28.9%	6.6%	31.4%
	SZ	27.8%	5.9%	27.8%	4.7%
	YJ	28.0%	5.1%	30.2%	6.8%
	Unknown	38.0%	2.9%	35.4%	4.0%
South Dongting Lake					
L3	L2	36.0%	6.2%	33.0%	7.7%
	ML	4.2%	19.9%	3.4%	26.3%
	OC	8.0%	17.7%	8.8%	20.3%
	XJ	6.4%	23.7%	8.4%	33.3%
	ZS	13.2%	16.4%	16.2%	18.7%
	Unknown	32.2%	2.6%	30.2%	5.5%
L4	L2	20.0%	9.4%	17.4%	15.5%

701

702

## Abstract art

