Bacterial diversity in a large, temperate, heavily modified river, as determined by pyrosequencing

Gary E. Schultz Jr.[1,*], Jeffrey J. Kovatch[1], Emily M. Anneken[1,2]

[1]Department of Biological Sciences, Marshall University, Huntington, West Virginia 25755, USA
[2]*Present address:* US Environmental Protection Agency, Cincinnati, Ohio 45268, USA

ABSTRACT: Despite the importance of microbes in lotic systems, the structure and diversity of microbial communities in rivers are understudied. Biodiversities of bacterial communities within a large, heavily modified river (Ohio River) and a major tributary (Guyandotte River) were examined using next-generation pyrosequencing and classical ecological diversity indices. Next-generation pyrosequencing was used to determine percent abundance of 16S rRNA gene sequences from 6 sites in the Ohio River and 1 site in the Guyandotte River. *Cyanobacteria* were the numerically dominant phylum (~57%) in the Ohio River, suggesting a largely autochthonous community, and the community structure was unique relative to other aquatic and riverine systems. Multiple samples (n = 9) of the Ohio River communities had >85% similarity to one another on a sequence basis at the operational taxonomic unit level across several river kilometers, indicating the bacterial community of this river within a navigational pool is well mixed. The Guyandotte River community was only ~13–20% similar to the Ohio River community and was dominated by *Proteobacteria* (~75%), indicating a distinctly different bacterial community structure suggestive of allochthonous systems. The Guyandotte was more diverse than the Ohio River (Shannon diversity index: 4.23 and 2.29 respectively). Relative to lentic systems, the distribution of unique sequences at the phylum level were very similar, indicating that bacterial communities in both lotic and lentic systems are all composed of bacteria adapted to freshwater.

KEY WORDS: Bacterial community · Richness · Pyrosequencing · Molecular diversity · Ohio River

INTRODUCTION

Rivers carry water, organisms, nutrients, and sediment from and to vast areas of the terrestrial landscape. They give humans a convenient way to transport large amounts of cargo as well as providing a great deal of the world's electric power. Rivers provide other ecosystem services to human populations such as drinking water, fisheries, irrigation, and recreation. Despite making up only ~0.2% of the freshwater on the earth, rivers play an important and significant role in the global carbon cycle (Cole et al. 2007). As biogeochemical cycles are controlled by microbes, the bacterial community carried by rivers may significantly affect global biogeochemical budgets as well as the ecosystem services provided by rivers.

Changing land use, urbanization, industrialization, and river flow modification such as dams and interbasin transfers may negatively influence river biodiversity (Vörösmarty et al. 2010). Because microbes control biogeochemical function, knowledge of the members and dynamics of the bacterial community of an ecosystem is vital to predicting the impact of such threats upon the ecosystem. Surprisingly little is known about the diversity of lotic freshwater bacterial communities. In the mid-1980s, investigators began to use molecular tools to examine the bacteria in environmental samples (Olsen et al. 1986, Pace et al. 1986). Since these early studies, several projects

have identified the major lineages of bacteria in rivers (Böckelmann et al. 2000, Battin et al. 2001, Crump & Hobbie 2005, Winter et al. 2007, Li et al. 2008, Crump et al. 2009, 2012, Lemke et al. 2009, Ghai et al. 2011, Liu et al. 2012). Most of these studies used some combination of denaturing gradient gel electrophoresis (DGGE), fluorescence *in situ* hybridization (FISH), and 16S rRNA clone libraries, coupled with Sanger sequencing, to identify the bacteria present. These tools were the best available, but quite time-consuming, and are not capable of defining the vast majority of bacteria present at the species or genus level.

Despite these limitations, several investigators have been able to determine the most abundant bacteria within various rivers. Crump et al. (2009) determined that Arctic rivers were dominated by *Bacteroidetes*, followed by *Betaproteobacteria*, then *Actinobacteria*. Lemke et al. (2009) found that the most abundant phylum of the Upper Paraná River in Brazil was *Firmicutes*, followed by *Proteobacteria*, then *Bacteroidetes*, then *Actinobacteria*. *Cytophaga*, *Flavobacterium*, and *Bacteroidetes* were most abundant in the Danube (Winter et al. 2007). *Betaproteobacteria* were the most abundant group in several freshwater rivers, including the Elbe (Böckelmann et al. 2000), the Rotmoos stream in the Alps (Battin et al. 2001), the Ipswich River (Crump & Hobbie 2005), the Songhua River (Li et al. 2008), and the Dongjiang River in China (Liu et al. 2012). These studies were limited in the number of sequences that could be acquired by the technology available.

New technologies that allow the parallelization of sequencing have become available in the last decade (Metzker 2010). For example, 454 pyrosequencing is a next-generation sequencing tool that sequences by detecting the release of pyrophosphate during nucleotide incorporation rather than Sanger sequencing-type chain termination (Ronaghi et al. 1998). With pyrosequencing, DNA from environmental samples can be sequenced without cultivation of bacteria. The data provided by 16S rRNA pyrosequencing allow investigators to utilize analytical tools from classical ecology, allowing the first attempts to accurately measure the biological diversity of riverine microbial communities. The first pyrosequencing study of a large river (the Amazon; Ghai et al. 2011) found that at a single point in the river, the dominant phylum was *Proteobacteria* (dominated by *Betaproteobacteria*), followed by *Actinobacteria*. Significantly more studies are needed to determine the bacterial diversity and community structure contained within large rivers.

The present paper adds to this increasing flow of information in the hopes that patterns in biodiversity may emerge. Presented here is the first known effort using next-generation sequencing to describe the bacterial biodiversity of a large, temperate, heavily modified and managed river, the Ohio River in North America. This paper examines in detail the alpha diversity of the Ohio River and the Guyandotte River, a large tributary, and the beta diversity between systems. Also included are comparisons with other lotic systems and lentic systems from the literature.

MATERIALS AND METHODS

Study area and sampling

The Ohio River is the largest tributary by volume of the Mississippi River in North America. The Ohio River is ~1579 km long, and its drainage basin covers 490 603 km² (Fig. 1). The river is heavily modified for navigation and flood control and has a variety of land use types along the river, ranging from rural forests and agriculture to highly industrialized urban areas. The Ohio is separated by a series of locks and dams into 20 navigation pools. These pools vary in length from 16 to 295 km. The data in this study were collected from within the Greenup Pool, river km 449 to 549 (counting up with distance downstream) near Huntington, WV, USA.

Water samples from ~0.5 m depth were collected August 8, 2010 by boat from 6 sites within the Greenup Pool of the Ohio River and from 1 site within the Guyandotte River (Fig. 1). These sites were located upstream and downstream of the confluence of the Guyandotte into the Ohio River (38° 26′ 01.66″ N, 82° 23′ 20.70″ W). The historical discharge of the Ohio and Guyandotte Rivers in August is ~1147 m³ s⁻¹ and ~15.5 m³ s⁻¹, respectively (US Geological Survey, National Water Information System, http://waterdata.usgs.gov/wv/nwis/rt). Ohio River samples ~2.9 km upstream of the Guyandotte were taken near the bank on the north and right descending bank (RB1), center-channel (CC1), and near the bank on the south and left descending bank (LB1); duplicate samples were taken from the Guyandotte River (GUY) several hundred meters upstream of its confluence with the Ohio River, as well as near the right descending bank across from the Guyandotte confluence (RB2), center-channel in the Ohio River at the confluence (CC2), and ~1.6 km downstream of the confluence near the left descending bank (LB2) (Fig. 1). All samples were collected imme-
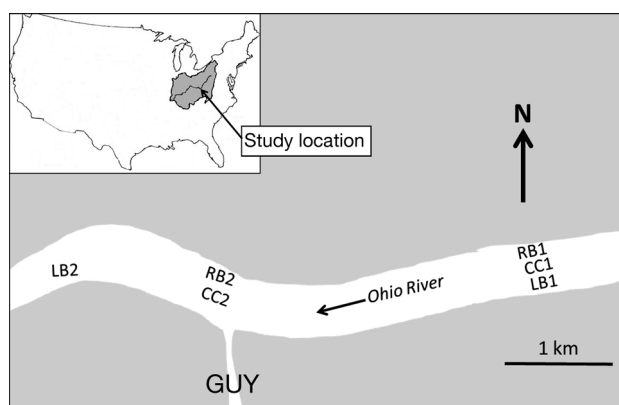
Fig. 1. Subsurface water samples were taken from 7 sites in the Greenup Pool of Ohio River and Guyandotte River (GUY), a tributary. Ohio River site designations are right (descending) bank (RB1 and RB2), left (descending) bank (LB1 and LB2), and center channel (CC1 and CC2). Arrow indicates westerly flow direction at the study site. Inset: study site on the Ohio River main stem and drainage basin (shaded) within the USA

diately below the surface using clean (autoclaved) 1 l polypropylene bottles. Following collection, samples were placed in a cooler in the dark and promptly taken back to the lab for processing.

DNA extraction

Samples were filtered through a 0.22 µm Whatman Nuclear Track-Etch membrane. After filtration, DNA was extracted using the Mo Bio PowerWater® DNA Isolation Kit (Mo Bio Laboratories). After confirmation by gel electrophoresis, extracted DNA from each sample was sent to the Research and Testing Laboratory (RTL; Lubbock, TX) on ice for pyrosequencing.

Pyrosequencing

Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) was performed as described previously (Dowd et al. 2008) at RTL. The bTEFAP was analyzed on the Genome Sequencer FLX instrument using Titanium protocols and reagents (Roche). The bTE-FAP procedures were performed at RTL based upon RTL protocols (www.researchandtesting.com/docs/ Data_Analysis_Methodology.pdf). One-step PCR with a mixture of Hot Start and HotStar high-fidelity *Taq* polymerases was utilized. The PCR primers for FLX amplicon pyrosequencing were 104F (5'-GGA CGG GTG AGT AAC ACG TG-3') and 530R (5'-GTA TTA CCG CGG CTG CTG-3'). These primers are located in the V2-V3 hypervariable regions of 16S.

Data analysis

Data analysis was performed per the RTL operating procedures (www.researchandtesting.com/docs/ Data_Analysis_Methodology.pdf) as follows. During the read quality checking and denoising stage, denoising and chimera checking was performed on all the reads for each region of data. Each remaining read was then quality-scanned to remove poor reads from each sample. Potential poor-quality ends of each read were trimmed using the scores provided by the sequencer and according to the internal RTL algorithm (see operating procedures).

Clustering was performed to classify reads into clusters. This post quality-trimming stage outputs the clustered sequences along with the information required to determine how each read joined the cluster. RTL uses USEARCH (Edgar 2010) to perform the clustering stage. Reads are first sorted from longest to shortest. USEARCH is then used to de-replicate the reads so they are clustered together into groups in which each sequence is an exact match to a portion of the seed sequence for each cluster. The sequences are then sorted from longest to shortest again, and USEARCH then clusters the sequences at a 4% divergence to determine similar clusters. These consensus sequences are then again re-sorted based on length, and any cluster not containing at least 2 member sequences is removed from consideration.

The RTL analysis pipeline uses UCHIME in de novo mode to perform chimera detection and removal on the clustered data. All possible chimeras were removed from analysis.

After removal of chimeric sequences, denoising was performed according to RTL protocols. The reads were then placed into the taxonomic analysis pipeline. Of the 152 405 raw sequence reads, 33 255 sequences were discarded through quality-filtering, chimera removal, and denoising.

The denoised and chimera-checked reads generated during sequencing were condensed into a single FASTA-formatted file. The analysis pipeline then removed all failed sequence reads, sequences with low-quality tags, and sequences less than half the expected amplicon length (or 250 bp, whichever was shorter). Remaining sequences were condensed into a single FASTA-formatted sequence and quality file. This FASTA reads archive was then used for taxonomic identification.

Remaining sequences were identified by sorting the sequences so that the FASTA-formatted file contained reads from longest to shortest. USEARCH (Edgar 2010) was then again used and sequences were clustered into operational taxonomic unit (OTU) clusters with 100% identity. The seed sequence for each cluster was put into a FASTA-formatted sequence file and queried against a database of high-quality sequences derived from the National Center for Biotechnology Information (NCBI) using a distributed .NET algorithm that utilizes BLASTn+ (KrakenBLAST; www.krakenblast.com). The resulting BLASTn+ outputs were compiled and data-reduction analysis performed using a .NET and C# analysis pipeline (Dowd et al. 2005, 2008, Sen et al. 2009, Suchodolski et al. 2009, Callaway et al. 2010, Andreotti et al. 2011, Handl et al. 2011, Ishak et al. 2011). This BLASTn+ derived sequence identity percentage was used to classify each sequence at the appropriate taxonomic level. Sequences with identity scores >97% (<3% divergence) were resolved at the species level, between 95 and 97% at the genus level, between 90 and 95% at the family level, between 85 and 90% at the order level, between 80 and 85% at the class level, and between 77 and 80% at the phylum level. All matches below a 77% identity score were discarded. In addition, if the high-score pair was not at least 75% of the query sequence, it was discarded, regardless of identity. The percent abundance of each organism, resolved upon the above parameters, was then presented at each taxonomic level, including a percentage compilation representing all sequences resolved to their primary identification (Dowd et al. 2008, Sen et al. 2009, Suchodolski et al. 2009, Callaway et al. 2010, Andreotti et al. 2011, Handl et al. 2011, Ishak et al. 2011).

To ensure that chloroplast sequences were not being misidentified as cyanobacteria, an extra step in which all sequences were compared to a database containing >3800 sequences identified as chloroplast sequences was added to the analysis pipeline.

Based on the percent identity of the query sequence to the reference sequence, the taxonomic level was resolved to the best characterized OTU, whether species, genus, family, order, class, or phylum. Every sequence kept had an identity assigned to at least phylum level. Every sequence resolved to at least phylum level was considered a distinct OTU. To obtain species information for sequences identified below the 97% cut-off, sequences were reported as the nearest well-described species for each sequence

(Table S1 in the Supplement at www.int-res.com/articles/suppl/a070p169_supp.pdf). Throughout the present paper, however, species are only used in a descriptive manner. When calculations for diversity and evenness were made, the OTU data were used to ensure the most accurate results.

Diversity and community similarity

Samples were not pooled to determine beta diversity. The Bray-Curtis similarity coefficient (Bray & Curtis 1957) has been suggested to accurately reflect true community similarity (Bloom 1981, Faith et al. 1987). Bray-Curtis similarity and multidimensional scaling were done using the vegan library (Oksanen et al. 2007) in the statistical package R (R Development Core Team 2007). Similarity between the distribution of unique sequences by phylum in the Ohio River, the Guyandotte River, and in freshwater lakes was calculated using Sørensen's presence/absence measure of diversity (Sørensen 1948).

Where noted, a weighted-averages approach was used to pool replicate samples for some sites to obtain an average percent abundance for each river. Weighted averages were determined by first multiplying the number of reads per sample (N) by the OTU percentages (p_i) to determine sample size for each OTU (n_i), $n_i = Np_i$. Then OTU percent abundances for pooled samples (p_{ij}) were the sum of OTUs across samples (n_{ij}) divided by the across-sample sum of reads (N_j), $p_{ij} = \Sigma n_{ij} (\Sigma N_j)^{-1}$, where p_{ij} represented the weighted-average abundance for an OTU.

Diversity calculations were performed at the best taxonomic level available or OTU. Individual samples (n = 11) were rarified to 6300 reads using step-sizes of 300 reads, random selection without replacement, and 10 iterations per sample using SAS IML software. Rarefactions were similarly calculated for pooled samples for the Ohio (n = 9) and Guyandotte (n = 2) rivers to 18 900 reads. Richness (S_{Rare}), Shannon's diversity index (H'; Shannon & Weaver 1963), Shannon's evenness (E; Pielou 1969, 1975), and Simpson's diversity index (D; Simpson 1949) were calculated with rarified data sets. Although several studies have pointed out the disadvantages of the Shannon diversity index, it is still often used (reviewed in Magurran 2004). Comparisons of diversity estimates should be performed on samples with similarly sized sequencing depths (Gihring et al. 2012), and estimates of alpha diversity at a sequencing depth of 5000 have been shown to agree with deeper sequencing (Lundin et al. 2012).

RESULTS

Taxonomy of the Ohio River

Twenty unique phyla were represented among the 11 samples taken from both rivers. Seventeen different phyla were found in 9 Ohio River samples, and 17 were found in 2 Guyandotte samples (Table S1 in the Supplement). In the Ohio River samples, at the phylum level, *Cyanobacteria* made up the largest proportion of bacteria in all samples, averaging ~57% of the total cells sequenced in each sample (Fig. 2). The top 12 represented phyla in terms of proportional abundance (*Cyanobacteria*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, *Chloroflexi*, *Verrucomicrobia*, *Fibrobacteres*, *Firmicutes*, *Gemmatimonadetes*, *Acidobacteria*, *Planctomycetes*, and *Lentisphaerae*) were found in all Ohio River samples and—except for *Lentisphaerae*—also in the Guyandotte. TM7 (a candidate bacterial phylum) and *Synergistetes* were only seen in the Guyandotte River.

Cyanobacteria, *Actinobacteria*, and *Bacteroidetes* were the most commonly observed phyla in the Ohio River, making up >90% of the total (57, 20, and 14% respectively; Fig. 2). The dominant *Cyanobacteria* genus was *Synechococcus*, the dominant *Actinobacteria* was *Clavibacter*, and the dominant *Bacteroidetes* was *Sphingobacterium* (Table S1). In the Guyandotte, the most commonly observed phyla were *Proteobacteria*, *Actinobacteria*, and *Bacteroidetes*, making up >90% of the total, at 75, 12, and 9% respectively. The dominant *Proteobacteria* genus was *Curvibacter,* the dominant *Actinobacteria* genus was *Planktophila*, and the dominant *Bacteroidetes* genus was *Sphingobacterium* (Table S1).

All Ohio River sites showed a similar distribution of bacteria, with *Cyanobacteria* predominating (Fig. 2). Taking the weighted average of all the Ohio River samples, *Cyanobacteria* species made up ~57.0% of the bacteria, *Actinobacteria* made up 19.9%, *Bacteroidetes* made up ~14.5%, and *Proteobacteria* made up 6.4%. The remaining 2.2% were made up of some 60 different species belonging to 16 different phyla. In the Guyandotte, *Proteobacteria* species made up ~74.9% of the bacteria present, *Actinobacteria* made up ~11.8%, *Bacteroidetes* made up ~8.7%, and the remaining 4.6% were made up of 53 different bacterial species from 17 different phyla.
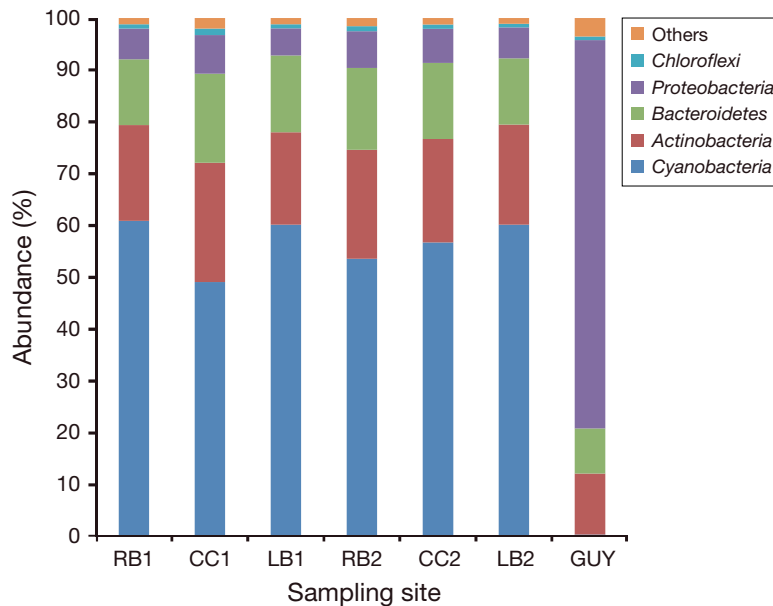

Fig. 2. Percent abundances of the 5 most abundant bacterial phyla for 6 sampling sites in Ohio River and 1 site in Guyandotte River. See Fig. 1 for site abbreviations

Similarity

The similarity of the bacterial community of the Ohio River relative to the Guyandotte was determined at the OTU level using Bray-Curtis similarity. Similarity was >85% between all Ohio River samples (Table 1). Replicate similarity, when taken, was >91% for the Ohio River samples, and 78% for the Guyandotte samples. The similarity between the bacterial communities in the Ohio River and the Guyandotte was much lower, and ranged from 13 to 20% (Table 1). Multidimensional scaling analysis supported these findings: Samples within the Ohio River formed a cluster distinct from the Guyandotte River samples (Fig. 3).

Diversity and richness measures

Rank abundance plots of the pooled Ohio River samples and the pooled Guyandotte River samples both show communities dominated by a few genera, but with many rare species also present (Fig. 4). Observed OTU richness for the Guyandotte River was higher than that of the Ohio River, with $S_{obs} = 583$ and 467, respectively, despite the number of reads in the Ohio River being 5-fold greater (Table 2). There were more rare species, singletons, and doubletons in the Guyandotte River (Table 2), consistent with the patterns in the Whittaker plots (Fig. 4).
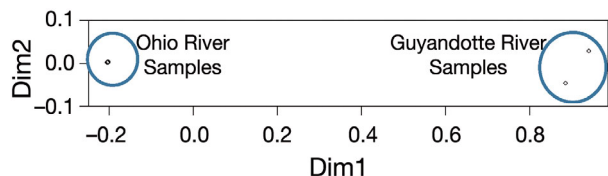
Fig. 3. Multidimensional scaling diagram showing the degree of similarity among samples between bacterial communities of Ohio River (n = 9; points overlap on graph) and Guyandotte River (n = 2). Community similarity (Bray-Curtis) was calculated with relative abundance of operational taxonomic units (OTUs). Dim: dimension
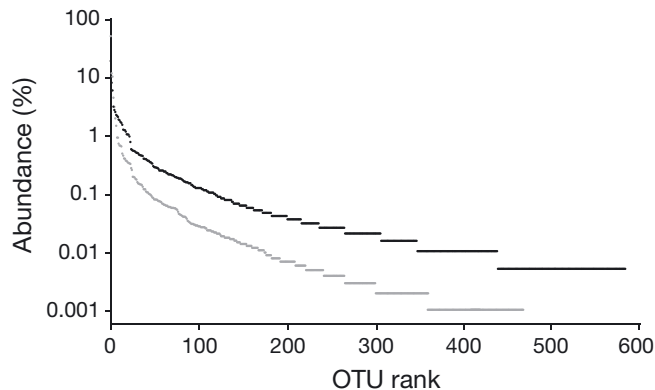

Fig. 4. Bacteria rank-abundances from Ohio (●) and Guyandotte (●) rivers. Weighted-average percent abundance of operational taxonomic units (OTUs) for pooled samples in Ohio (n = 9) and Guyandotte rivers (n = 2)

Species richness increased when within-river samples were pooled and total sequence depth increased. Pooled data for each river were rarified to 18 900 reads, resulting in decreased Ohio diversity (S_{Rare} = 291.1) relative to the Guyandotte River (S_{Rare} = 582.3). Both the Shannon and Simpson diversity indices were higher in the Guyandotte River relative to the Ohio River, with H'_{Guy} = 4.23 and H'_{OH} = 2.29, and D_{Guy} = 18.95 and D_{OH} = 3.56 (Table 2). The Shannon derived evenness for the entire suite of Ohio River samples was 0.373, and ranged from 0.386 to 0.450. For the pooled Guyandotte samples, evenness was 0.665.

DISCUSSION

Proteobacteria are typically the most abundant bacterial phylum reported in rivers (Böckelmann et al. 2000, Battin et al. 2001, Sekiguchi et al. 2002, Crump & Hobbie 2005, Li et al. 2008, Liu et al. 2012), but occasionally the most abundant phyla have been *Actinobacteria* (Ghai et al. 2011), *Bacteroidetes* (Crump et al. 2009), or *Firmicutes* (Lemke et al.

2009). It is interesting, then, to see that *Cyanobacteria* were the numerically dominant bacteria within the Greenup Pool of the Ohio River (Fig. 2). This is, however, generally consistent with the predictions of the river continuum concept that pelagic primary productivity increases in large rivers (Vannote et al. 1980). The Ohio River was also quite different from other rivers described in that *Betaproteobacteria* made up <5% of the total community. The Guyandotte River, however, had characteristics more similar to other rivers, with a small fraction of the total sequences found belonging to *Cyanobacteria* (<0.5%), but showing a large population of *Betaproteobacteria* (~65%), mostly *Curvibacter* (~21% of total number of bacteria).

There has only been one other set of published data from a large river using 454 pyrosequencing that describes the taxonomy of the bacterial community (Ghai et al. 2011). The free-flowing upper course of the Amazon River described by Ghai et al. (2011) is very different than the heavily managed Ohio River. Ghai et al. (2011) used direct 454 pyrosequencing and classified the 16S rRNA gene fragments recovered rather than pyrosequencing a section of 16S rRNA genes amplified by PCR. Despite the differences in study methods, latitude, land use, and flow between the 2 rivers, it is still informative to compare the taxonomy of the dominant bacteria of these 2 large rivers. The dominant phylum of the relatively pristine Amazon River was *Proteobacteria* (~50%), followed by *Actinobacteria* (~23%), with *Bacteroidetes*, *Acidobacteria*, *Gemma-*

Table 1. Bray-Curtis similarity coefficients for bacterial diversity for all Ohio River and Guyandotte River sample sites. Identical communities have a value of 1.0. See Fig. 1 for site abbreviations; 'a' indicates duplicate samples

	RB1	LB1	CC1	CC2	CC2a	RB2	RB2a	LB2	LB2a	GUY
LB1	0.95	1								
CC1	0.86	0.86	1							
CC2	0.94	0.95	0.88	1						
CC2a	0.91	0.90	0.90	0.91	1					
RB2	0.92	0.91	0.90	0.93	0.93	1				
RB2a	0.89	0.90	0.93	0.92	0.92	0.93	1			
LB2	0.94	0.92	0.85	0.94	0.91	0.91	0.89	1		
LB2a	0.93	0.91	0.85	0.93	0.91	0.91	0.89	0.95	1	
GUY	0.17	0.16	0.20	0.17	0.18	0.18	0.18	0.18	0.17	1
GUYa	0.13	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.78

Table 2. Diversity and operational taxonomic unit (OTU) richness estimates for individual and pooled samples (n) from the Ohio and Guyandotte rivers. S_{obs}: observed OTUs, singles: singletons (sequences seen only once in the system), doubles: doubletons (sequences seen only twice in the system). Sample rarefactions to 6300 reads for individual samples (*italicized*) and 18 900 reads for pooled river samples (**bold**) were used to calculate comparable richness of OTUs (S_{Rare}), Shannon's (H') and Simpson's (D) diversity indices, and Shannon's evenness (E). See Fig. 1 for site abbreviations; 'a' indicates duplicate samples; (-) not applicable

Sample	n	Reads	S_{obs}	Singles	Doubles	S_{Rare}	H'	D	E
RB1	1	11 013	226	-	-	*181.7*	*2.14*	*3.16*	*0.395*
CC1	1	10 510	218	-	-	*182.0*	*2.42*	*4.27*	*0.450*
LB1	1	12 225	232	-	-	*180.1*	*2.11*	*3.14*	*0.386*
RB2	1	9708	245	-	-	*207.5*	*2.38*	*3.83*	*0.433*
RB2a	1	10 873	245	-	-	*201.3*	*2.33*	*3.90*	*0.423*
CC2	1	14 639	265	-	-	*192.3*	*2.24*	*3.46*	*0.402*
CC2a	1	6354	187	-	-	*186.4*	*2.30*	*3.71*	*0.439*
LB2	1	12 651	254	-	-	*194.2*	*2.25*	*3.40*	*0.407*
LB2a	1	12 211	241	-	-	*192.7*	*2.28*	*3.55*	*0.416*
Ohio	9	100 184	467	107	60	**291.1**	**2.29**	**3.56**	**0.373**
GUY	1	9095	449	-	-	*395.6*	*4.25*	*21.76*	*0.695*
GUYa	1	9871	464	-	-	*399.5*	*4.11*	*16.42*	*0.670*
Guyandotte	2	18 966	583	145	91	**582.3**	**4.23**	**18.95**	**0.665**

timonadetes, and *Cyanobacteria* making up smaller fractions (<5% each) (Ghai et al. 2011). In the Ohio River, *Cyanobacteria* (~57%) was the dominant phylum, followed by *Actinobacteria* (~20%) and then *Bacteroidetes* (~14%). *Acidobacteria* and *Gemmatimonadetes* made up a very small fraction of the Ohio samples (<0.2% of the pooled Ohio samples). Of the 19 most abundant bacterial genera making up the Amazon River community, only *Streptomyces* (2%) and *Polynucleobacter* (0.7%) made up >0.5% of the bacterial community in the Ohio River (see Table S1 in the Supplement for details).

The bacterial community of the Guyandotte River was more similar in structure to that of the Amazon than the Ohio River. In both the Guyandotte and the Amazon Rivers, *Proteobacteria* was the dominant phylum, followed by *Actinobacteria*, and the most common bacteria were in the Class *Betaproteobacteria*. By contrast, very few *Cyanobacteria* were found in the Guyandotte (<1%) relative to the Ohio River (~57%). The turbidity of the Guyandotte (134 nephelometric turbidity units [NTU] at the time of sampling; data not shown) was an order of magnitude higher than in the Ohio River (15 NTU). These observations indicate that large rivers may contain very different assemblages of bacteria. The community differences observed in the present study are likely a function of the higher proportion of allochthonous material in the Guyandotte as compared to the autochthonous-dominated Ohio River. Consistent with predictions of the river continuum concept (Vannote et al. 1980), the pelagic autotrophic community was more abundant in the Ohio River, which has much less riparian cover and receives a larger amount of sunlight relative to the Guyandotte.

Distribution of bacterial sequences relative to freshwater-lake bacteria

Some bacterial groups are freshwater-specific and have a global distribution (Newton et al. 2011). Comparing the taxon distribution patterns seen in freshwater lakes with those seen in the Ohio and Guyandotte Rivers reveals that 28 different phyla were represented across the 3 systems (Fig. 5). Although *Cyanobacteria* were numerically dominant in the Ohio River, *Proteobacteria* sequences made up nearly 60% of the number of different sequences found (Fig. 5). Of the 21 phyla represented by sequences reported by Newton et al. (2011), 14 were also represented in the Ohio River and 17 were also represented in the Guyandotte River. All 3 freshwater assemblages were quite similar in terms of phylum distribution patterns, with the Sørensen measure of diversity between the 3 systems ranging from 79 to 82% (Fig. 5). These data suggest that although different freshwater systems may have community structures consisting of the same freshwater-specific taxa, these taxa may be distributed quite differently.

Similarity within a river navigation pool

One of our aims was to determine whether the Ohio River is homogeneous with regard to bacterial communities within a navigational pool. The study area is located within 1 of 20 large regulated navigational pools in the Ohio River. These pools are maintained at a minimum depth of 3.7 m, with areas in the Greenup Pool >9 m deep. Free surface flow in large rivers with variable depth contributes to momentum flux (Chow et al. 1988), which should ensure the river main channel is vertically and horizontally well mixed. Therefore, it was suspected that the center channel was approximately homogeneous with regard to bacterial diversity. However, it was
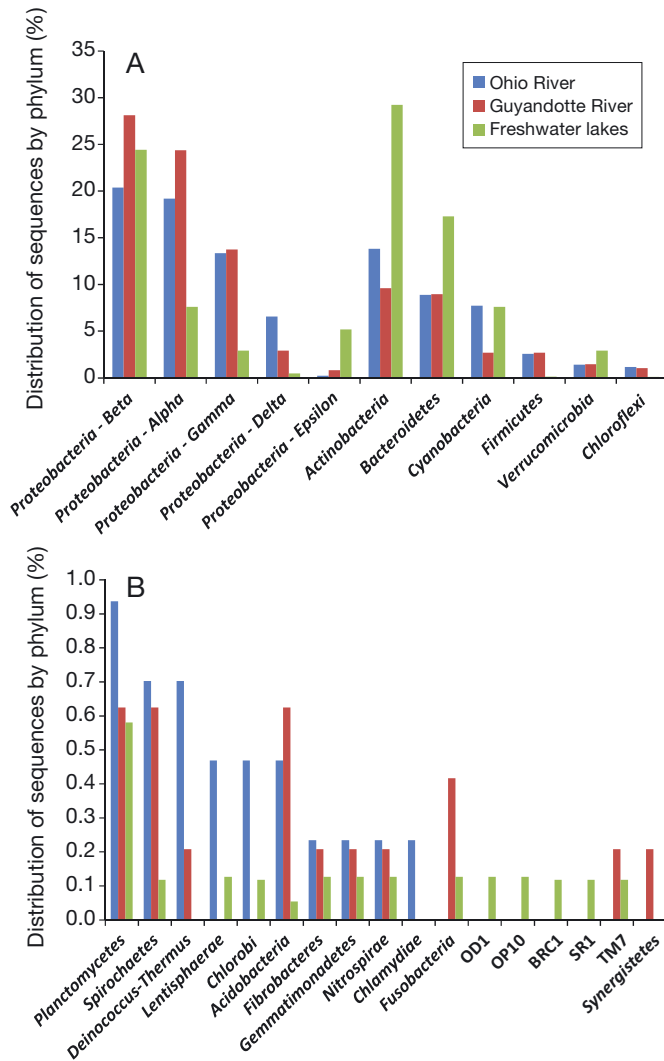
Fig. 5. Distribution of sequences by phylum in the Ohio River, the Guyandotte River, and freshwater lakes. Each bar represents the number of unique sequences found in each phylum relative to the total number of unique sequences found in each system (expressed as a percentage). *Proteobacteria* were separated by Class (e.g. Alpha) to show the large number of different *Proteobacteria* sequences present. (A) All phyla that make up >1% of the total community. (B) All phyla that make up <1% of the total community. Proportions for the rivers are from pooled sample sequences, n = 100 184 and n = 18 966, for the Ohio and Guyandotte Rivers, respectively. A database of 11 500 sequences gathered from 68 studies was used to determine the sequences in freshwater lakes (adapted from Newton et al. 2011). Sørensen's index was 0.79 between the Ohio River and freshwater lakes, 0.82 between the Ohio and Guyandotte Rivers, and 0.79 between the Guyandotte River and freshwater lakes. OD1, OP10, BRC1, SR1, and TM7 are candidate phyla

unknown whether differences in river flow, flora, fauna, or input from the banks caused significant changes in the bacterial community across the width of the channel or even within short distances downstream. The data show that the bacterial communi-

ties across the channel were all quite similar, even as similar as duplicate samples (Table 1). This homogeneity allows each pool of the Ohio River to be perceived as a single entity rather than a patchwork quilt of different environments and habitats. This similarity of the bacterial community over a few kilometers within regions that are similar in terms of flow and input is likely to be common among large deep rivers. However, an important caveat must be noted. As pointed out by Leopold et al. (1964), the river acts like a conveyor belt. New water and its associated microbial community are continually passing any given point of the river, never to return. Therefore, any samples taken are literally a record of that section of the river at that particular instance in time, with no guarantees that the bacterial community of the river will be similar at a later date.

Diversity comparisons with the Guyandotte River

The Guyandotte is 267 km long, with an average annual discharge about 100 times less than that of the Ohio River (US Geological Survey, Water Resources of the United States, http://water.usgs.gov). When our samples were collected, the Guyandotte had a distinctly different bacterial community (Fig. 2) from that of the Ohio River. At the sample site 1.6 km directly downstream from the Guyandotte (Fig. 1), however, the bacterial community of the Ohio River was still 91 to 94% similar to the center channel of the Ohio above the Guyandotte (Table 1). This indicates that the Ohio River main-stem bacterial community was stable and resilient against tributary input. This buffering capacity may exist at the microbial level as main-stem bacteria out-compete tributary bacteria, or it may simply be a dilution effect. From data collected over >50 yr (US Geological Survey, National Water Information System, http://waterdata.usgs.gov/wv/nwis/rt), the average monthly mean discharge of the Ohio River in the month of August was ~1147 m^3 s^{-1} and that of the Guyandotte was approximately 74 times lower at ~15.5 m^3 s^{-1}. If the buffering is due to dilution, the tributary bacteria may grow and contribute to the main-stem bacterial community downstream.

Diversity indices

The Ohio River seems to be spatially well mixed based on the similarity of diversity estimates across the 9 samples (Table 2). Rarefaction to the least com-

mon sequence depth has been advised for comparing diversity estimates from samples of unequal size or effort (Gihring et al. 2012, Lundin et al. 2012). Comparison of individual samples rarefied to 6300 reads (which is the lowest sequence depth; found in Sample CC2a) suggests diversity is very similar among the 9 Ohio River samples, and that the Ohio River samples are distinct from the Guyandotte River samples (Fig. 6A). Diversity analysis of these Ohio samples further suggests that diversity is higher in the Guyandotte River (H' = 2.29 and 4.23, respectively), and that across-sample diversity in the Ohio River is very similar, with H' ranging from 2.11 to 2.42 (Table 2). We are confident about these findings, as sequencing depths of 5000 to 7000 are sufficient for estimating common diversity indices, including H' (Lundin et al. 2012).

When within-river samples were pooled, estimates of diversity fell within the range of values seen by each separate sample in the respective river (Table 2),
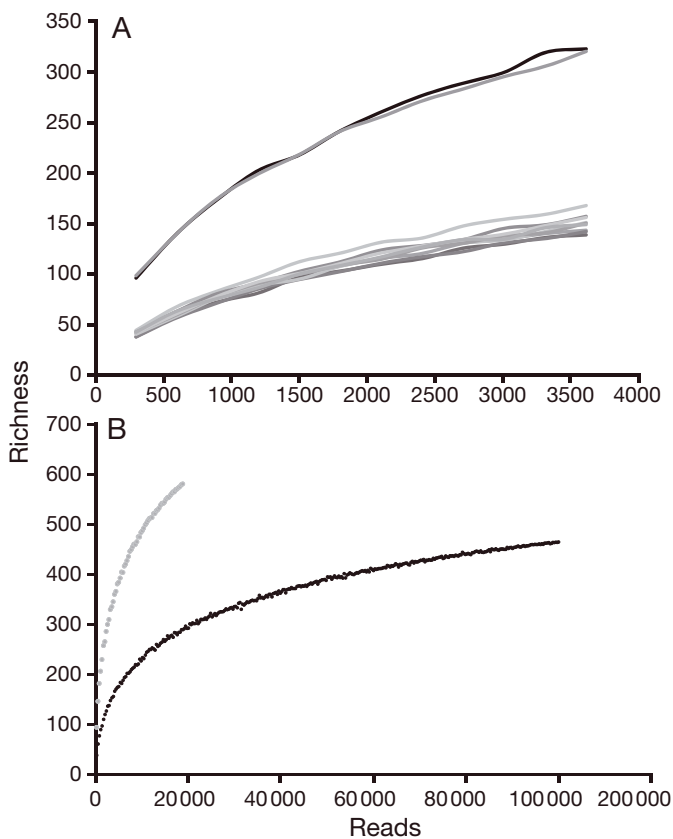

Fig. 6. Rarefaction curves for bacterial richness from river water samples. (A) Individual samples were rarified to 6300. The 2 upper diversity functions are from Guyandotte River. The lower 9 functions are from Ohio River. (B) Pooled samples from Guyandotte (grey) and Ohio (black) Rivers. All rarefaction curves were generated with step-size of 300 reads, no replacement, and 10 iterations

consistent with Lundin et al.'s (2012) findings. Richness values, however, increased with increased sample size in both cases. In the case of the Guyandotte River, pooling the 2 samples effectively doubled the number of reads for diversity analysis from ~9400 to 18 900 (Table 2), resulting in a 45% increase in species richness (Fig. 6B). For the Ohio River, pooling the 9 samples increased the number of reads to 100 184. This pooling increased the estimated richness, S_{Rare}, by 40 to 60% (Table 2).

To make appropriate diversity comparisons to the Guyandotte River, we rarefied the 100 184 reads in the pooled Ohio River sample to 18 900 (Fig. 6B). Richness estimates at 18 900 reads were 50% lower in the Ohio River than the Guyandotte River. The richness function for the Guyandotte River suggests that the actual richness or asymptotic richness is much higher than observed (Fig. 6B). Estimated richness using the Chao1 estimator (Chao 1987) for the Guyandotte River is 20% higher than observed, with S_{obs} = 583 and S_{Chao1} = 698.5.

One of the goals of the present paper was to estimate the bacterial richness of the Ohio River. With such a large trimming of data (~81%), much information is likely lost, resulting in lower species richness estimations for the Ohio River, with S_{Rare} = 291.1 instead of the S_{obs} = 467 (Table 2). The rarefaction curve suggests that 100 000 reads is not enough to reach the asymptotic richness of the river (Fig. 6B). Estimation of species richness using all 100 184 reads suggests that richness of the Ohio River is S_{Chao1} = 562.4. Although not directly comparable with the Guyandotte River due to the difference in number of reads, the S_{Chao1} for the Ohio River confirms that the Ohio River is less diverse and less rich than the Guyandotte River (Table 2, Fig. 4).

The lower bacterial diversity seen in the Ohio River relative to the Guyandotte River may be due to several potential loss mechanisms. Predation by small heterotrophic flagellates and viral lysis has been shown to have potentially significant impacts on bacteria community structure in aquatic systems (e. g. Pernthaler 2005, Simek et al. 2007, Zhang et al. 2007, Berdjeb et al. 2011). Alternatively, the Guyandotte River was almost 9 times more turbid than the Ohio River, likely from recent rain events in the Guyandotte watershed that contributed to the hypothesized increased allochthonous microbial community. Bacteria washed in from surrounding soils (Crump et al. 2012) not adapted to living in flowing freshwater may die quickly in a flowing aquatic system and be removed from the river community as the water travels downstream. The prevalent bacterial

species in the Ohio River samples may then be autochthonous and represent the functional river community. Future investigations on the spatial and temporal stability of this functional river community are needed to validate this point.

The bacterial community evenness of the Ohio River is low ($E = 0.373$; Table 2). A system with low evenness will have a lower metabolic potential at the genetic level and may be less able to react to changes in the environment. Therefore, the Ohio River bacterial community may be inefficient at utilizing the resources available to its members and may be strongly influenced by environmental changes.

Low-abundance species

Pedrós-Alió (2006, 2007) suggested that the long 'tail' in rank-abundance plots of low abundance species (Fig. 4) might constitute a 'seed bank' of species that could become more numerous if conditions changed. This scenario is quite plausible for marine systems but less likely to occur in rivers. In rivers in general, rare bacteria at point 'A' may become more numerous by the time they reach point 'B', but they are not capable of returning and impacting the community at point 'A' and thus could not constitute a seed bank for the pelagic river. These bacteria, however, could 'seed' other specific ecological niches to be occupied, such as embayments, biofilms, etc., that occur along the river. For large rivers, our hypothesis for the occurrence of these rare species is that low-abundance species are those that are washed into the river and are either not capable of thriving in the freshwater lotic environment or are washed into the river in such small numbers that they are unable to reach significant abundances prior to exiting the river.

Lotic 'conveyor belts'

Bacterioplankton, by definition, go where the river goes. Bacterial growth rates are not fast enough to maintain a planktonic population against a flowing river. Without a mechanism for bacteria to return upstream, what investigators of a river see is a conveyor belt (Leopold et al. 1964) of bacterial communities flowing past. This conveyor belt, however, is not random, and patterns are likely to exist that govern the content of the belt.

Bacteria are constantly carried into the river with the water that makes up the river. Those bacteria that are capable of thriving in freshwater will grow and utilize the resources found in the river. These bacteria account for the significant role that Cole et al. (2007) described for rivers in the carbon cycle. In the Ohio River, the numerically dominant bacteria, and thus presumably the active bacteria, are different from those seen thus far in other large river systems. The sources of these bacteria may be different, or differences in the river itself may allow the growth of different communities of bacteria as the river flows downstream.

LITERATURE CITED

Andreotti R, Pérez de León A, Dowd S, Guerrero F, Bendele K, Scoles G (2011) Assessment of bacterial diversity in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus* through tag-encoded pyrosequencing. BMC Microbiol 11:6–16

Battin TJ, Wille A, Sattler B, Psenner R (2001) Phylogenetic and functional heterogeneity of sediment biofilms along environmental gradients in a glacial stream. Appl Environ Microbiol 67:799–807

Berdjeb L, Pollet T, Domaizon I, Jacquet S (2011) Effect of grazers and viruses on bacterial community structure and production in two contrasting trophic lakes. BMC Microbiol 11:88, doi:10.1186/1471-218U-11-88

Bloom SA (1981) Similarity indices in community studies: potential pitfalls. Mar Ecol Prog Ser 5:125–128

Böckelmann U, Manz W, Neu TR, Szewzyk U (2000) Characterization of the microbial community of lotic organic aggregates ('river snow') in the Elbe River of Germany by cultivation and molecular methods. FEMS Microbiol Ecol 33:157–170

Bray JR, Curtis JT (1957) An ordination of upland forest communities of southern Wisconsin. Ecol Monograph 27:325–349

Callaway TR, Dowd SE, Edrington TS, Anderson RC and others (2010) Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. J Anim Sci 88:3977–3983

Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. Biometrics 43:783–791

Chow VT, Maidment DR, Mays LW (1988) Applied hydrology. McGraw Hill, New York, NY

Cole JJ, Prairie YT, Caraco NF, McDowell WH and others (2007) Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. Ecosystems 10:172–185

Crump BC, Hobbie JE (2005) Synchrony and seasonality of bacterioplankton communities in two temperate rivers. Limnol Oceanogr 50:1718–1729

Crump BC, Peterson BJ, Raymond PA, Amon RM, Rinehart A, McClelland JW, Holmes RM (2009) Circumpolar syn-

chrony in big river bacterioplankton. Proc Natl Acad Sci USA 106:21208−21212

Crump BC, Amaral-Zettler LA, Kling GW (2012) Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. ISME J 6:1629−1639

Dowd SE, Zaragoza J, Rodriguez JR, Oliver MJ, Payton PR (2005) Windows .NET Network Distributed Basic Local Alignment Search Toolkit (W.ND-BLAST). BMC Bioinformatics 6:93−106

Dowd SE, Callaway TR, Wolcott RD, Sun Y, McKeehan T, Hagevoort RG, Edrington TS (2008) FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol 8:125−132

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460−2461

Faith DP, Minchin PR, Belbin L (1987) Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69:57−68

Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D and others (2011) Metagenomics of the water column in the pristine upper course of the Amazon river. PLoS ONE 6: e23785

Gihring TM, Green SJ, Schadt CW (2012) Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. Environ Microbiol 14:285−290

Handl S, Dowd SE, Garcia-Mazcorro JF, Steiner JM, Suchodolski JS (2011) Massive parallel 16S rRNA gene pyrosequencing reveals highly diverse fecal bacterial and fungal communities in healthy dogs and cats. FEMS Microbiol Ecol 76:301−310

Ishak HD, Plowes R, Sen R, Kellner K and others (2011) Bacterial diversity in *Solenopsis invicta* and *Solenopsis geminata* ant colonies characterized by 16S amplicon 454 pyrosequencing. Microb Ecol 61:821−831

Lemke MJ, Lienau EK, Rothe J, Pagioro TA, Rosenfeld J, Desalle R (2009) Description of freshwater bacterial assemblages from the upper Parana River floodpulse system, Brazil. Microb Ecol 57:94−103

Leopold LB, Wolman MG, Miller JP (1964) Fluvial processes in geomorphology. WH Freeman, San Francisco, CA

Li D, Yang M, Li Z, Qi R, He J, Liu H (2008) Change of bacterial communities in sediments along Songhua River in Northeastern China after a nitrobenzene pollution event. FEMS Microbiol Ecol 65:494−503

Liu Z, Huang S, Sun G, Xu Z, Xu M (2012) Phylogenetic diversity, composition and distribution of bacterioplankton community in the Dongjiang River, China. FEMS Microbiol Ecol 80:30−44

Lundin D, Severin I, Logue JB, Östman O, Andersson AF, Lindström EF (2012) Which sequencing depth is sufficient to describe patterns in bacterial α- and β-diversity? Environ Microbiol Rep 4:367−372

Magurran AE (2004) Measuring biological diversity. Blackwell Publishing, Malden, MA

Metzker ML (2010) Sequencing technologies — the next generation. Nat Rev Genet 11:31−46

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev 75:14−49

Oksanen J, Kindt R, Legendre P, O'Hara RB (2007) vegan: community ecology package for R, version 1.9-25. R Foundation for Statistical Computing, Vienna. Available at http://cran.r-project.org

Olsen G, Lane D, Giovannoni S, Pace N, Stahl D (1986)

Microbial ecology and evolution: a ribosomal RNA approach. Annu Rev Microbiol 40:337−365

Pace N, Stahl D, Lane D, Olsen G (1986) The analysis of natural microbial populations by rRNA sequences. Adv Microb Ecol 9:1−55

Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? Trends Microbiol 14:257−263

Pedrós-Alió C (2007) Dipping into the rare biosphere. Science 315:192−193

Pernthaler J (2005) Predation on prokaryotes in the water column and its ecological implications. Nat Rev Microbiol 3:537−546

Pielou EC (1969) An introduction to mathematical ecology. Wiley-Interscience, New York, NY

Pielou EC (1975) Ecological diversity. Wiley, New York, NY

R Development Core Team (2007) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. www.R-project.org

Ronaghi M, Uhlén M, Nyrén P (1998) A sequencing method based on real-time pyrophosphate. Science 281:363−365

Sekiguchi H, Watanabe M, Nakahara T, Xu B, Uchiyama H (2002) Succession of bacterial community structure along the Changjiang River determined by denaturing gradient gel electrophoresis and clone library analysis. Appl Environ Microbiol 68:5142−5150

Sen R, Ishak HD, Estrada D, Dowd SE, Hong E, Mueller UG (2009) Generalized antifungal activity and 454-screening of *Pseudonocardia* and *Amycolatopsis* bacteria in nests of fungus-growing ants. Proc Natl Acad Sci USA 106: 17805−17810

Shannon CE, Weaver W (1963) The mathematical theory of communication. University of Illinois Press, Urbana, IL

Šimek K, Weinbauer MG, Hornák K, Jezbera J, Nedoma J, Dolan JR (2007) Grazer and virus-induced mortality of bacterioplankton accelerates development of *Flectobacillus* populations in a freshwater community. Environ Microbiol 9:789−800

Simpson EH (1949) Measurement of diversity. Nature 163: 688

Sørensen T (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biol Skr K Danske Vidensk Selsk 5:1−34

Suchodolski JS, Dowd SE, Westermarck E, Steiner JM, Wolcott RD, Spillmann T, Harmoinen JA (2009) The effect of the macrolide antibiotic tylosin on microbial diversity in the canine small intestine as demonstrated by massive parallel 16S rRNA gene sequencing. BMC Microbiol 9: 210−225

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE (1980) The river continuum concept. Can J Fish Aquat Sci 37:130−137

Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D and others (2010) Global threats to human water security and river biodiversity. Nature 467:555−561

Winter C, Hein T, Kavka G, Mach RL, Farnleitner AH (2007) Longitudinal changes in the bacterial community composition of the Danube River: a whole-river approach. Appl Environ Microbiol 73:421−431

Zhang R, Weinbauer MG, Qian PY (2007) Viruses and flagellates sustain apparent richness and reduce biomass accumulation of bacterioplankton in coastal marine waters. Environ Microbiol 9:3008−3018