TITLE:

Bacterial diversity along a 2 600 km river continuum

Contributors:

Savio Domenico[1,2], Sinclair Lucas[3], Ijaz Umer Z.[4], Blaschke Alfred P.[1,5], Reischer Georg H.[2,7], Blöschl Guenter[1,5], Mach Robert L.[2], Kirschner Alexander K.T.[6,7], Farnleitner Andreas H.[1,2,7], Eiler Alexander[3]*

* Correspondence: A Eiler, Department of Ecology and Genetics, Limnology, Uppsala University, Norbyv. 18D, Uppsala, SE-75236, Sweden

email: alexander.eiler@ebc.uu.se

1 Centre for Water Resource Systems (CWRS), Vienna University of Technology, Vienna, Austria

2 Research Group Environmental Microbiology and Molecular Ecology, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria

3 Department of Ecology and Genetics, Limnology, and Science for Life Laboratory, Uppsala University, Uppsala, Sweden

4 School of Engineering, University of Glasgow, Glasgow, UK

5 Institute of Hydraulic Engineering and Water Resource Management, Vienna University of Technology, Vienna, Austria

6 Institute for Hygiene and Applied Immunology, Water Hygiene, Medical University Vienna, Vienna, Austria

7 Interuniversity Cooperation Centre Water and Health, www.waterandhealth.at

for submission to ISME J

28 **Abstract**

29 Understanding the biogeochemistry of large rivers is of high relevance as they play a major
30 role in the global carbon cycle and provide diverse ecosystem services. Since these ecosystem
31 functions are mainly mediated by bacteria, knowledge about their diversity represents a
32 centrepiece in determining the role of rivers in the landscape. Here, we present a
33 comprehensive dataset on the bacterioplankton diversity along a 2 600 km midstream transect
34 of the second largest river in Europe, the Danube River, and its tributaries using an Illumina®-
35 based sequencing approach (16S rRNA-gene amplicon sequencing). Our analysis revealed
36 that bacterial richness and evenness gradually declined downwards the river in both the 0.2-
37 3.0 µm and the > 3.0 µm size fractions. These shifts of the bacterioplankton community
38 along the river was also underpinned by beta diversity analysis where effects of tributaries
39 were negligible in regards to the overall variation. In addition, we found very few taxa typical
40 for lotic systems and that the relative contribution of so-called typical freshwater bacteria
41 often observed in lakes was increasing towards the river mouth. This supports the hypothesis
42 of a succession from soil and groundwater bacteria towards lake bacteria in lotic systems.
43 Putting our findings into a broad ecological context, we suggest that patterns of
44 bacterioplankton diversity in large rivers follow predictions of the River Continuum Concept
45 published in 1980, with a modification for planktonic microorganisms.

46

47 **Keywords**

48 bacterial diversity/gradual development/next-generation sequencing/river continuum
49 concept/riverine bacterioplankton

Introduction

Streams and rivers are important compartments of landscapes. Besides linking lakes and terrestrial with marine systems, lotic systems provide essential ecosystem services as they supply drinking water, water for irrigation and industry, hydropower, transport routes and areas for recreation. Of general importance is their role in biogeochemical nutrient cycling. Until recently, rivers and streams were mainly considered as pipes shuttling organic material and nutrients from the land to the oceans. This view has changed as lotic together with lentic systems are now considered as "leaky funnels" in the cycling of elements as they play an important role in the temporary storage (sequestration) and in the transformation of terrestrial organic matter (Ensign & Doyle, 2006; Cole *et al.*, 2007; Withers & Jarvie, 2008; Battin *et al.*, 2009). As a result of recognizing the role of rivers and streams in the carbon cycle (see for example the report by IPCC in 2013; http://www.ipcc.ch/), lotic networks and the diverse processes in their water column and sediments have received increasing interest (Kronvang *et al.*, 1999; Beaulieu *et al.*, 2010; Seitzinger *et al.*, 2010; Aufdenkampe *et al.*, 2011; Benstead & Leigh, 2012; Raymond *et al.*, 2013). Looking into the "black box" of nutrient processing in freshwater systems, bacteria are regarded as the main transformers of elemental nutrients and as substantial contributors to energy flow and nutrient cycling on a global scale (Cotner & Biddanda, 2002; Battin *et al.*, 2009; Findlay, 2010; Madsen, 2011). For open lotic systems such as rivers, however, there is still a big lack of knowledge about the diversity of bacterial communities and how this diversity is linked with ecosystem functioning (Battin *et al.*, 2009). This includes that there is still no agreement on how distinct river bacterioplankton is from that of other freshwater systems, how variable diversity is along whole river transects and what is regulating this diversity.

Summarizing previous studies, it can be concluded that bacteria affiliated to the phyla of *Proteobacteria* (especially *Betaproteobacteria*), *Bacteroidetes, Cyanobacteria* and *Verrucomicrobia* are dominating the bacterial communities in rivers (Crump *et al.*, 1999;

Zwart *et al.*, 2002; Cottrell *et al.*, 2005; Winter *et al.*, 2007; Lemke *et al.*, 2008; Mueller-Spitz *et al.*, 2009; Newton *et al.*, 2011; Liu *et al.*, 2012). These explorative studies on freshwater bacterial communities suggest that riverine bacterioplankton form a cohesive group with the abundant taxa in lakes and is therefore comprised of so-called "typical" freshwater bacteria (Zwart *et al.*, 2002; Lozupone & Knight, 2007; Newton *et al.*, 2011). Nevertheless, all these studies were constrained by their low sequencing depth and focus on the dominant members of the communities. A reasonable sequencing depth, however, is a prerequisite for estimating the communities' diversity and identifying fine-scale changes as a response to changing environmental conditions. To do so, a minimum sequencing depth of 1 000 and 5 000 16S rRNA gene reads per sample was suggested for a proper analysis of beta and alpha diversity, respectively (Lundin *et al.*, 2012). These methodological constraints have been overcome with the recent rise of Next-Generation Sequencing (NGS) (Shokralla *et al.*, 2012). The ultra-deep sequencing of bacterial communities by targeting short hyper-variable regions of the 16S rRNA gene not only allows a proper investigation of diversity, in particular richness, but also the detection and investigation of rare populations with critical functions for distinct ecosystems (Sogin *et al.*, 2006; Gilbert *et al.*, 2012; Sjöstedt *et al.*, 2012). In large rivers, NGS-studies on the microbial community composition are very rare with only a few available studies on the bacterioplankton communities of the Amazonas River (Brazil), Mississippi River (USA) and Columbia River Estuary (USA), mainly revealing taxonomic patterns (Fortunato *et al.*, 2013; Staley *et al.*, 2013). In these studies, the longitudinal development of the bacterioplankton communities along the rivers could not be addressed comprehensively since only a few sites were analysed. Considering the changing environmental gradients along rivers (Sekiguchi *et al.*, 2002; Winter *et al.*, 2007; Velimirov *et al.*, 2011), it can be expected that the bacterial communities will be variable as well as their function. This variability in community composition and function has been hypothesized to originate from the import of illuviated terrestrial bacteria and distinct freshwater communities from merging tributaries,

anthropogenic pollution from point sources like wastewater treatment plants or from diffuse natural or anthropogenic sources from soil erosion or agriculture (Zampella *et al.*, 2007; Tu, 2011; Besemer *et al.*, 2012). Other sources for observed variability can be impoundments or river regulation that alter hydrology, retention time or nutrient concentrations, consequently changing the structure and function of the autochthonous bacterial community (Bukaveckas *et al.*, 2002; Ruiz-González *et al.*, 2013). For macroorganisms, the large scale patterns seen in diversity from headwater streams to large rivers has been summarized in the River Continuum Concept (RCC) where diversity is expected to increase until midstream reaches, then to drop towards the river mouth. The RCC states that this pattern is due to the gradient of physical factors formed by the drainage network, as well as dynamics in chemical properties and resulting biological activity (Vannote *et al.*, 1980).

Here, we extend the RCC to river bacterioplankton by showing results from a NGS study on the bacterial community composition and how variability in bacterioplankton diversity is related to environmental conditions along a continuous river transect of 2 600 km from medium-sized reaches to the mouth of a large river. As study object the Danube River was chosen which is the second largest river of Europe by discharge and length. The Danube drains a basin of approximately 801,000 km² and borders 19 countries with 83 million inhabitants (Sommerwerk *et al.*, 2010).

Material and methods

Study sites and sample collection

Samples were taken within the frame of the second Joint Danube Survey (JDS 2) in the year 2007. The overall purpose of the Joint Danube Surveys is the comprehensive evaluation of the chemical and ecological status of the whole Danube River on the basis of the European Union Water Framework Directive (WFD) (Liska *et al.*, 2008). During sampling from Aug 15 to Sept 26 2007, 75 sites were sampled along the mainstream of the Danube River as well as 21

of the Danube's major tributaries and branches along its shippable way from river kilometre (rkm) 2 600 to rkm 0 (=river mouth) (Kirschner *et al.*, 2009). The 43 days represent the average retention time for a water mass in this part of the Danube river (for discussion about this see (Velimirov *et al.*, 2011).

For more detailed information on the sampling sites see (Kirschner *et al.*, 2009) and (Velimirov *et al.*, 2011). Samples were taken with sterile 1 L glass flasks from a water depth of approximately 30 cm. For DNA-extraction from attached bacterioplankton, depending on the biomass-concentration, 120-300 mL river water was filtered through 3.0 µm pore-sized polycarbonate filters (Cyclopore, Whatman, Germany) by vacuum filtration. The filtrate, representing the bacterioplankton fraction smaller than 3.0 µm (later referred to as "free-living" bacterioplankton) was collected in a sterile glass bottle and subsequently filtered through 0.2 µm pore-sized polycarbonate filters (Cyclopore, Whatman, Germany). The filters were stored at -80 °C until DNA extraction.

Supporting data

Within the frame of the Joint Danube Survey 2, a wide range of chemical and biological parameters was collected (Liska *et al.*, 2008). All data, sampling methods as well as analytical methods are made publicly available via the official website of the International Commission for the Protection of the Danube River (ICPDR;http://www.icpdr.org/wq-db/). Selected data from JDS 1 & 2 were published previously in several studies (Kirschner *et al.*, 2009; Janauer *et al.*, 2010; Velimirov *et al.*, 2011; von der Ohe *et al.*, 2011).

DNA extraction and quantification of bacterial DNA using quantitative PCR (qPCR)

Genomic DNA was extracted using a slightly modified protocol of a previously published phenol-chloroform bead-beating procedure (Griffiths *et al.*, 2000) using isopropanol instead of polyethylene glycol for DNA precipitation. Total DNA concentration was assessed

154 applying the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies Corporation,
155 USA) and 16S rRNA genes were quantified using domain-specific quantitative PCR.
156 Quantitative PCR reactions contained 2.5 µL of 1:4 and 1:16 diluted DNA extract as template,
157 0.2 µM of primers targeting the V1-V2 region of most bacterial 16S rRNA genes (8F, 5'-
158 AGAGTTTGATCCTGGCTCAG-3' and 338R, 5'-TGCTGCCTCCCGTAGGAGT-3') (MWG
159 Biotech AG, Ebersberg, Germany) and iQ™ SYBR® Green Supermix (Bio-Rad Laboratories,
160 Hercules, USA) (Frank *et al.*, 2007; Fierer *et al.*, 2008). Ratios of measured 16S rRNA gene
161 copy numbers in the different sample dilutions deviating markedly from 1 after multiplication
162 with the respective dilution factor were interpreted as indication for PCR-inhibition.
163
164 **Preparation of 16S rRNA gene amplicon libraries**
165 For the preparation of amplicon libraries, 16S rRNA genes were amplified and barcoded in a
166 two-step procedure in order to reduce PCR-bias introduced by long primers with barcode- and
167 sequencing adaptor-overhangs (Berry *et al.*, 2011). We followed the protocol as described by
168 Sinclair *et al.* (submitted). In short, 16S rRNA gene fragments of most bacteria were
169 amplified applying primers targeting the V3-V4 variable regions (341F, 5'-
170 CCTACGGGNGGCWGCAG-3' and 805R, 5'-GACTACHVGGGTATCTAATCC-3'
171 (Herlemann *et al.*, 2011). In 25 µL reactions containing 0.5 µM primer 341F and 805R
172 (MWG), 200 µM dNTPs (Invitrogen), 0.5 U Q5 HF DNA polymerase and provided buffer
173 (New England Biolabs, USA), genomic DNA was amplified in duplicates in 20 cycles. In
174 order to use equal amounts of bacterial template DNA for increased comparability and
175 reduction of PCR-bias, the final volume of environmental DNA extract used for each sample
176 was calculated based on 16S rRNA gene copy concentration in the respective sample
177 determined earlier by quantitative PCR (see above). For 105 samples, the self-defined
178 optimum volume of environmental DNA-extract corresponding to 6.4×10^5 16S rRNA genes
179 was spiked into the first step PCR reactions whereas for 27 samples, lower concentrations

were used due to limited amounts of bacterial genomic DNA or PCR-inhibition detected by quantitative PCR (see above). These 132 samples included eight biological replicates. Prior, we removed four samples because of their extremely low genomic DNA-concentrations and 16S rRNA gene copy numbers. Duplicates of PCR-products were pooled, diluted 1:100 and used as templates in the subsequent barcoding-PCR. In this PCR, diluted 16S rRNA gene amplicons were amplified using 50 primer pairs with unique barcode pairs (Sinclair *et al.*, submitted). The barcoding PCRs for most samples were conducted in triplicates analogously to first step PCR (n=100). Remaining 32 samples with weak bands in first step PCR due to low genomic template DNA-concentrations or high sample-dilution to avoid detected PCR-inhibition were amplified in 6-9 replicates in order to increase amplicon DNA yield. Barcoded PCR amplicons were pooled equimolarly after purification using the Agencourt AMPure XP purification system (Beckman Coulter, Danvers, MA, USA) and quantification of amplicon-concentration using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies Corporation, USA). Finally, in total, 137 samples including 5 negative controls resulted in four pools for sequencing.

Illumina® sequencing

Sequencing was performed on a Illumina® MiSeq at SciLife Lab Uppsala. For each pool, library preparation was performed separately following the TruSeq protocol with the exception of initial fragmentation and size selection. This involves the binding of the standard sequencing adapters in combination with separate Illumina®-specific MID barcodes that enables the combination of different pools on the same sequencing run (Sinclair *et al.*, submitted). After pooling, random PhiX DNA was added to provide calibration and help with the cluster generation on the MiSeq's flow cell.

16S rRNA gene amplicon data analysis

Sequence data was processed as outlined in Sinclair *et al.* (submitted). In short, paired-end 16S rRNA amplicon read assembly was conducted using PANDAseq (Masella *et al.*, 2012). After pooling of all reads originating from the four pools, reads were filtered based on Phred scores. Truncation of barcode sequences and insertion of barcodes into read-label was performed as recommended by Edgar (2013). Chimera removal and OTU (Operational taxonomic units) clustering at three percent sequence dissimilarity was performed based on all assembled reads occurring at least twice and performed applying UPARSE-denovo-picking (Edgar, 2013). Taxonomic annotation of OTUs was performed by matching to the quality checked SILVAmod-database (Lanzén *et al.*, 2012) using CREST. Plastid, mitochondrial and archaeal sequences were removed. Additionally, OTUs were also taxonomically annotated as "typical" freshwater bacteria using a freshwater database (Newton *et al.*, 2011). If necessary, OTU-resampling for standardization of sequence numbers per sample was performed using the 'rrarefy'-function implemented in the R-package vegan (Oksanen *et al.*, 2013). Biodiversity measure calculation, statistical analyses as well as plot-generation were conducted in R (R Core Team, 2013) and using python scripts.

Habitat index for the top 5 000 OTUs was determined using the SEQenv pipeline (http://environments.hcmr.gr/seqenv.html). SEQenv pipeline retrieves hits to highly similar sequences from public repositories (NCBI Genbank) and uses a text mining module to identify Environmental Ontology (EnvO) (Ref: http://environmentontology.org/) terms mentioned in the text records carrying environmental contextual information ("Isolation Source" field entry for genomes in Genbank or associated PubMed abstracts). At the time of running SEQenv on our dataset, in version 0.8, there were around 1200 EnvO terms organized into three main branches namely, *environmental material*, *environmental feature*, and *biome* and give a concise and controlled description of the environments. However, we have used SEQenv to retrieve a subset of these terms, i.e., those that contain "Habitat" (ENVO:00002036).

Sequences have been deposited...

Results

General description of sequences

In total, 132 DNA samples extracted from filtered water from the Danube River and its tributaries plus 5 negative controls were sequenced. Sequencing yielded 2 030 029 sequence read pairs ranging from 3 451 to 24 873 per sample. After all quality filtering and mate-pair assembly as outlined in Sinclair *et al.* (submitted), 1 572 361 assembled sequence reads (from here on only referred to as "reads") were obtained from the 132 high-quality samples. OTU-clustering resulted in 8 697 OTUs after removal of 625 Plastid-, Mitochondrion-, Thaumarchaetoa-, Crenarchaeota- and Euryarchaeota-assigned OTUs which represented 19.1 percent of the reads. Further, rarefaction to 2 347 reads per sample (= minimum number of reads in any one sample) and additional removal of OTUs with < 2 reads reduced the number of OTUs to 5 082 for beta diversity analysis. For alpha diversity analysis, all samples with less than 7 000 reads were excluded resulting in 8 241 OTUs in the remaining 88 samples after resampling.

Core microbial community

The majority of all bacteria-assigned OTUs (4 402 of 8 697) was only represented by less than ten reads in all samples. As a consequence, 3 243 of 8 697 OTUs (~ 37 percent) were represented in only one to four samples and additional 2 219 OTUs (~ 26 percent) in just as few as five to nine samples. Besides these rare OTUs, the core community in the Danube River was comprised of 89 OTUs for the free-living bacterioplankton community as defined by all OTUs that are represented in at least 90 percent of all samples of the respective size fraction. The corresponding number of OTUs for the particle-attached core community was 141. The cumulative contribution of OTUs based on their occurrence along the entire river

transect is shown in Figure 1A for both analysed size fractions (0.2-3.0 µm and > 3.0 µm, representing the free-living and particle-attached bacterial communities). Based on this figure the core communities of the free-living and particle-attached bacteria were defined by their presence in at least 90 percent of the samples of the respective size fraction. On average, 81 percent of all reads of the free-living river community and 63 percent of all reads of the attached river community belonged to the respective core community. For both core communities a significant increase of their relative quantitative contribution towards the river mouth could be observed (see Figure 1B).

Variability of diversity along the river

Chao1 richness estimator as well as Pielou's evenness (J) were separately calculated for both size fractions of the main river after resampling of samples with originally more than 7 000 reads (n=88). The estimated richness was persistently higher in the attached fraction with an average of 2 025 OTUs compared to the free-living fraction with 1 248 OTUs on average. We observed a gradual decrease of the bacterial richness in both size fractions along the Danube River (Figure 2A) which was confirmed by regression analysis (Table 1). A similar significant decrease was observed for J along the course of the river (Figure 2B).

The visualization of beta diversity by applying non-metric multidimensional scaling (NMDS) mirrored the gradual development (Figure 3) and a significant relationship was observed between community composition from both size fractions and river kilometre (Table 2). Additional environmental variables corresponding with compositional dynamics when excluding tributaries are given in Table 2. As shown in the NMDS, tributaries did not follow the general patterns and often formed outliers in the multidimensional space. Moreover, the clear separation of the bacterial communities into two distinct groups based on the filter fractions could be confirmed by PERMANOVA (R^2=0.156, p-value<0.01). The apparent synchrony in the gradual development in the two size fractions along the river's course could

284 also be shown by Procrustes test (R = 0.96, p < 0.001). Still, the permutation test on the
285 betadispersion of each size fraction revealed a significantly higher variability in the >3 µm
286 fraction than the 0.2-3.0 µm fraction along the river transect (p-value=0.002) (see Fig. S2).
287
288 **Typical River bacterioplankton**
289 Results from the screening procedure for so-called typical freshwater-bacteria according to
290 Newton and colleagues (2011) revealed a high proportion of particular OTUs assigned to
291 previously described freshwater taxa (Figure 4). These included the LD12-tribe belonging to
292 the subphylum of *Alphaproteobacteria* as well as the acI-B1-, acI-A7- and acI-C2-tribes
293 belonging to the order of *Actinomycetales* of the phylum *Actinobacteria*. Interestingly, in the
294 free-living size fraction a clear trend was observed where the relative abundance of the four
295 above-mentioned tribes increased clearly towards the river mouth (Figure 4A) contributing at
296 most 35 percent to the total free-living community. In correspondence, there is a clear
297 decrease of atypical freshwater reads (labelled "Everything else") for the free-living fraction
298 while tribe-level-assignable reads steadily increased along the river as shown in Figure 4B. In
299 the attached fraction, these typical freshwater taxa are much less abundant (see Figure 4B).
300 Here, reads that could not be assigned to previously described typical freshwater taxa clearly
301 dominate the community with the exception of the samples closest to the river mouth, in
302 which an increase of the above mentioned typical freshwater bacteria could also be observed.
303 Annotation of sequences based on environmental descriptive terms occurring within their
304 Genbank (NCBI) records corroborate the specific findings of our study that bacteria of the
305 different size fractions seem to have distinct preferred habitats as indicated by statistical
306 analysis using the habitat annotations (PERMANOVA R^2=0.42; p<0.0001). Restricting the
307 analysis to 'groundwater' and 'soil' terms indicated that the proportion of bacteria potentially
308 originating from these sources decreased towards the river mouth (Figure 5A and B).
309 Opposite to what we expected, we could not depict a trend along the Danube River transect

using the contribution of 'river' and 'sediment' terms. Furthermore, we found only four OTUs annotated as 'typical river bacterioplankton' as defined by river terms dominating their environmental onthology.

Discussion

A rapidly increasing number of studies applying novel high-throughput sequencing technologies have revealed a tremendous diversity within microbial communities residing in all kinds of aquatic environments (e.g., Sogin *et al.*, 2006; Andersson *et al.*, 2009; Galand *et al.*, 2009; Eiler *et al.*, 2012; Peura *et al.*, 2012). However, river bacterioplankton has so far been widely neglected with few exceptions (Ghai *et al.*, 2011; Fortunato *et al.*, 2013; Staley *et al.*, 2013). Here, we describe biogeographic patterns in lotic bacterioplankton communities over a 2 600 km longitudinal transect from mid-sized stream reaches (rkm 2 600) to the mouth (rkm 0) (corresponding to stream order 4 to 12 according to the RCC; Vannote *et al.*, 1980) by investigating their taxonomic composition and their specificity to the “river biome”.

A previous publication from this survey has already shown that the bacterial compartment follows continuous trends along the Danube River by investigating bacterial abundance, morphotype composition and bacterial production (Velimirov *et al.*, 2011). In that study, it was shown that bacterial numbers, cell volumes, morphotype composition and attached bacterial production exhibited significant correlation with river kilometre and with several other environmental variables (Velimirov *et al.*, 2011) corresponding to our findings in conjunction with bacterial community composition (Table 2). Moreover, we recorded a significant decrease in bacterial richness in both examined size fractions, which lets us propose that the development of the bacterioplankton community in large rivers is in accordance with the 1980 published RCC (Vannote *et al.*, 1980).

This highly debated concept, originally developed for aquatic macroorganisms, states that biological diversity should be highest in medium-sized stream reaches (stream order of 4 to

336 6), corresponding to the most upstream sites of our transect (rkm 2 600), and should decline
337 from thereon towards the river mouth. Such a pattern of decreasing richness was previously
338 observed for the bacterioplankton community development along the midstream reaches of
339 the Upper Mississippi River where 10 sites along approximately 400 km were sampled
340 (Staley *et al.*, 2013).
341 The fact that bacterioplankton is primarily passively transported in contrast to fish or aquatic
342 invertebrates calls for modifications of the RCC. Vannote and colleagues emphasized the
343 consideration of riparian influence, substrate or flow as potentially important factors affecting
344 the biological diversity leading to the assumption that the riparian zone constantly provides
345 allochthonous microbes to the river bacterioplankton species pool. This assumption is
346 corroborated by our results of the SEQenv analysis were a decreasing contribution of
347 groundwater (Fig. 5A) and soil (Fig. 5B) bacteria was observed from the mid-sized stream
348 region towards the Danube River delta. Concomitant with decreasing influence from riparian
349 zone and increasing stream order, the proportion of autochthonous river bacteria increased as
350 indicated by the pronounced rise in the contribution of so-called typical freshwater bacteria
351 especially in the free-living size fractions (see Figure 4). This suggests that at intermediate
352 stream orders where both allochthonous and autochthonous bacteria can proliferate, diversity
353 should be highest.
354 An additional explanation for the observed gradual decline in richness in the free-living
355 fraction and especially in the attached fraction could be the decreasing diversity in the organic
356 matter composition along our transect, with highest expected diversity in terms of quality and
357 availability in the mid-sized stream reaches (first sites in our study). There, both
358 allochthonous and autochthonous organic matter should be equally important providing a
359 broad range of substrates for bacteria, with dominance of allochthonous organic compounds
360 towards headwaters and autochthonous organic compounds towards the river's mouth. This
361 compositional change regarding the source of organic matter along our transect can be

362 accompanied by a qualitative change along the continuum from more labile to more refractory
363 organic compounds. Such a qualitative deterioration in turn should favour specialized and
364 highly competitive bacteria capable of using low quality and homogenous organic matter
365 sources in the downstream parts of the river. This assumption is also in accordance with the
366 observed trend to smaller cells by Velimirov and colleagues (2011), e.g. based on a significant
367 and constant increase of small coccoid cells towards the river mouth. Furthermore, this
368 coincides with the observed increasing relative contribution of typical freshwater taxa such as
369 LD12 and acI, which represent small cells with an oligotrophic lifestyle (Salcher *et al.*, 2011;
370 Garcia *et al.*, 2013). The trend of a "selection" for smaller cells might additionally result from
371 starvation of copiotrophic cells originating from terrestrial sources which are adapted to high
372 quality and nutrient-rich compounds (Barcina *et al.*, 1997; Egli, 2010). However, proof for the
373 role of organic matter sources in the apparent decline of richness towards the river mouth
374 needs to be provided by assessing organic matter composition and bioavailability in future
375 studies. Besides these bottom-up mechanisms, loss factors like sedimentation and top-down
376 control (grazing and viral lysis) influence microbial diversity and have been shown to vary
377 over environmental gradients (Ayo *et al.*, 2001; Weinbauer, 2004; Pernthaler, 2005) and can
378 be selective against specific community members (Bouvier & Del Giorgio, 2007;
379 Langenheder & Jürgens, 2001).
380 Moreover, our results show a concomitant decrease in evenness with bacterial richness in both
381 size fractions along the river transect which points to a rise in competitiveness of few species
382 attributable to an increase in stability and uniformity of the system along the river continuum.
383 An alternative explanation for the decreasing evenness along the Danube could be a
384 continuously progressive eutrophication of the system due to increasing availability of
385 nutrients from allochthonous inputs such as from agricultural areas or sewage sources
386 (Withers & Jarvie, 2008). However, the latter seems unlikely as a change in diversity
387 estimates in the main river communities did not significantly correlate with chlorophyll a and

388 phosphorous as proxies for eutrophication (inferred by Spearman rank correlation). This
389 finding is reinforced by the observation that despite a remarkable increase in total
390 phosphorous and chlorophyll a concentrations as well as bacterial production (Velimirov *et*
391 *al.*, 2011) along the stretch between Budapest (rkm 1 632) to Belgrade (rkm 1 168), we could
392 not observe a marked decrease in richness and evenness. Similarly, there was also an increase
393 in microbial faecal pollution along the stretch from Vienna (rkm 1 925) to Belgrade (rkm 1
394 168), indicative for wastewater effluents or the influence from highly polluted tributaries
395 (Kirschner *et al.*, 2009), which was not reflected in richness and evenness. However, when
396 including chloroplast-assigned sequence reads in the analysis, a decrease in evenness was
397 observed in these stream reaches. Plastid-associated reads and high concentrations of
398 chlorophyll a (4.34 – 30.64, mean = 15.12 ng/µL) revealed a diatom bloom
399 (*Thalassiosiraceae* as indicted by chloroplast 16S rRNA gene) in this stretch (rkm 1 632 to 1
400 132).
401 Illuminating our observations from a theoretical point of view and focussing more on the
402 process of community assembly over time, the RCC states that „the concept of biological
403 succession is of little use for river continua, because the communities in each reach have a
404 continuous heritage rather than an isolated temporal composition within a sequence of
405 discrete successional stages“; and „the concept implies that in natural river systems total
406 absence of a population is rare, and biological subsystems are simply shifting spatially and not
407 in the temporal sense typical for plant succession” (Vannote *et al.*, 1980). Latter hypothesis
408 applied to bacterioplankton is partially supported by findings from bacterioplankton dynamics
409 along the Upper Mississippi River (Staley *et al.*, 2013) and implied by a previous low
410 resolution study on the Danube River (Winter *et al.*, 2007). The former identified a bacterial
411 assemblage that appears spatially stable over ~ 400 km changing only minimally along the
412 Upper Mississippi River. Similarly to these findings, our results show less first-time
413 occurrences of new, not yet observed OTUs along the river transect with increasing distance

414 from upstream regions, pointing towards a core community along the Danube River. In our
415 study, the proposed ubiquitous “core bacterial community“ was composed of approximately a
416 hundred OTUs but contributed more than fifty percent to the total amplicon pool. Such OTUs
417 belonging to the core community can be argued to represent superior competitors, especially
418 when considering that they accumulated with increasing stream order. A likely consequence of
419 their accumulation could be the decrease in alpha diversity along the river transect.
420 Focussing on the taxonomic composition along the river, our data shows that ”typical“
421 freshwater bacteria including members of the acI lineage, freshwater SAR11 (LD12) and the
422 Polynucleobacter genus formed to a major part the „core bacterial community“. This finding
423 that riverine bacterial communities can resemble those of lakes, clearly corroborates the
424 existence of typical freshwater bacteria (Zwart *et al.*, 2002; Lozupone & Knight, 2007;
425 Newton *et al.*, 2011). However, a modification of the 'typical freshwater bacteria concept' for
426 rivers could be the consideration of the “succession” from soil and groundwater bacteria to
427 “lake bacteria“ in the flowing wave with increasing stream order.
428 Summing up these findings, we propose that riverine bacterioplankton in a large river shifts
429 towards a composition very similar to that of lakes with increasing stability of the system and
430 reduced influence from the riparian zone. The resulting accumulation of typical freshwater
431 bacteria is accompanied by a decrease in alpha diversity which is in accordance with the
432 RCC.
433

434 **Acknowledgement**

435 This study was supported by the Austrian Science Fund (FWF) as part of the DKplus “Vienna
436 Doctoral Programme on Water Resource Systems” (W1219-N22) and the FWF project
437 P25817-B22 as well as the research project “Groundwater Resource Systems Vienna” in
438 cooperation with Vienna Water (MA31). AE and LS are funded by the Swedish Foundation
439 for Strategic Research (ICA10-0015). Infrastructure (cruise ships and floating laboratory) and

logistics for taking, storing and transporting samples were provided by the International Commission for the Protection of the Danube River (ICPDR). The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project "b2011035".

Conflict of Interest Statement

The authors declare no conflict of interest

References

Andersson AF, Riemann L, Bertilsson S. (2009). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* **4**:171–181.

Aufdenkampe AK, Mayorga E, Raymond PA, Melack JM, Doney SC, Alin SR, *et al.* (2011). Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. *Front Ecol Environ* **9**:53–60.

Ayo B, Santamaría E, Latatu A, Artolozaga I, Azúa I, Iriberri J. (2001). Grazing rates of diverse morphotypes of bacterivorous ciliates feeding on four allochthonous bacteria. *Lett Appl Microbiol* **33**:455–460.

Barcina I, Lebaron P, Vives-Rego J. (1997). Survival of allochthonous bacteria in aquatic systems: a biological approach. *FEMS Microbiol Ecol* **23**:1–9.

Battin TJ, Luyssaert S, Kaplan LA, Aufdenkampe AK, Richter A, Tranvik LJ. (2009). The boundless carbon cycle. *Nature Geosci* **2**:598–600.

Beaulieu JJ, Tank JL, Hamilton SK, Wollheim WM, Hall RO, Mulholland PJ, *et al.* (2010). Nitrous oxide emission from denitrification in stream and river networks. *Proc Natl Acad Sci U S A* **108**:214–219.

Benstead JP, Leigh DS. (2012). An expanded role for river networks. *Nature Geosci* **5**:678–679.

Berry D, Ben Mahfoudh K, Wagner M, Loy A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl Environ Microbiol* **77**:7846–7849.

Besemer K, Peter H, Logue JB, Langenheder S, Lindström ES, Tranvik LJ, *et al.* (2012). Unraveling assembly of stream biofilm communities. *The ISME journal* **6**:1459–1468.

Bouvier T, Del Giorgio PA. (2007). Key role of selective viral-induced mortality in determining marine bacterial community composition. *Environ Microbiol* **9**:287–297.

Bukaveckas P, Williams J, Hendricks S. (2002). Factors regulating autotrophy and heterotrophy in the main channel and an embayment of a large river impoundment. *Aquat Ecol* **36**:355–369.

Cole JJ, Prairie YT, Caraco NF, McDowell WH, Tranvik LJ, Striegl RG, *et al.* (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems* **10**:172–185.

Cotner JB, Biddanda BA. (2002). Small players, large role: Microbial influence on biogeochemical processes in pelagic aquatic ecosystems. *Ecosystems* **5**:105–121.

Cottrell MT, Waidner LA, Yu L, Kirchman DL. (2005). Bacterial diversity of metagenomic and PCR libraries from the Delaware River: Metagenomic analysis of freshwater bacteria. *Environ Microbiol* **7**:1883–1895.

Crump BC, Armbrust EV, Baross JA. (1999). Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. *Appl Environ Microbiol* **65**:3192–3204.

Edgar RC. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**:996–998.

Egli T. (2010). How to live at very low substrate concentration. *Water Res* **44**:4826–4837.

Eiler A, Heinrich F, Bertilsson S. (2012). Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J* **6**:330–342.

Ensign SH, Doyle MW. (2006). Nutrient spiraling in streams and river networks. *J Geophys Res* **111**. http://doi.wiley.com/10.1029/2005JG000114 (Accessed December 3, 2013).

Fierer N, Hamady M, Lauber CL, Knight R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci U S A* **105**:17994–17999.

Findlay S. (2010). Stream microbial ecology. *J North Am Benthol Soc* **29**:170–181.

Fortunato CS, Eiler A, Herfort L, Needoba JA, Peterson TD, Crump BC. (2013). Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. *ISME J* **7**:1899–1911.

Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* **104**:13780–13785.

Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A* **106**:22427–22432.

Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Sczyrba A, Stepanauskas R, *et al.* (2013). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *ISME J* **7**:137–147.

Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D, Rinke R, Cristina Souza de Oliveira T, *et al.* (2011). Metagenomics of the Water Column in the Pristine Upper Course of the Amazon River. *PLoS One* **6**:e23785.

Gilbert JA, Steele JA, Caporaso JG, Steinbruck L, Reeder J, Temperton B, *et al.* (2012). Defining seasonal marine microbial community dynamics. *ISME J* **6**:298–308.

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. (2000). Rapid Method for Coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA- and rRNA-Based Microbial Community Composition. *Appl Environ Microbiol* **66**:5488–5491.

Herlemann DP, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF. (2011). Transitions in bacterial communities along the 2000[thinsp]km salinity gradient of the Baltic Sea. *ISME J* **5**:1571–1579.

Janauer GA, Schmidt-Mumm U, Schmidt B. (2010). Aquatic macrophytes and water current velocity in the Danube River. *Ecol Eng* **36**:1138–1145.

Kirschner AKT, Kavka GG, Velimirov B, Mach RL, Sommer R, Farnleitner AH. (2009). Microbiological water quality along the Danube River: Integrating data from two whole-river surveys and a transnational monitoring network. *Water Res* **43**:3673–3684.

Kronvang B, Hoffmann C, Svendsen L, Windolf J, Jensen J, Dørge J. (1999). Retention of nutrients in river basins. *Aquat Ecol* **33**:29–40.

Langenheder S, Jürgens K. (2001). Regulation of bacterial biomass and community structure by metazoan and protozoan predation. *Limnol Oceanogr* **46**:121–134.

Lanzén A, Jørgensen SL, Huson DH, Gorfer M, Grindhaug SH, Jonassen I, *et al.* (2012). CREST – Classification Resources for Environmental Sequence Tags. *PLoS One* **7**:e49334.

Lemke MJ, Lienau EK, Rothe J, Pagioro TA, Rosenfeld J, DeSalle R. (2008). Description of freshwater bacterial assemblages from the Upper Paraná River floodpulse system, Brazil. *Microb Ecol* **57**:94–103.

Liska I, Slobodnik J, Wagner F. (2008). Joint Danube Survey 2, Final Scientific Report. *International Commission for the Protection of the Danube River* 242.

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.

Lozupone CA, Knight R. (2007). Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* **104**:11436–11440.

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

Madsen EL. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Curr Opin Biotechnol* **22**:456–464.

Masella A, Bartram A, Truszkowski J, Brown D, Neufeld J. (2012). PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**:1–7.

Mueller-Spitz SR, Goetz GW, McLellan SL. (2009). Temporal and spatial variability in nearshore bacterioplankton communities of Lake Michigan. *FEMS Microbiol Ecol* **67**:511–522.

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.

Von der Ohe PC, Dulio V, Slobodnik J, De Deckere E, Kühne R, Ebert R-U, *et al.* (2011). A new risk assessment approach for the prioritization of 500 classical and emerging organic microcontaminants as potential river basin specific pollutants under the European Water Framework Directive. *Sci Total Environ* **409**:2064–2077.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, *et al.* (2013). vegan: Community Ecology Package. R package version 2.0-10. http://CRAN.R-project.org/package=vegan.

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

Peura S, Eiler A, Bertilsson S, Nykanen H, Tiirola M, Jones RI. (2012). Distinct and diverse anaerobic bacterial communities in boreal lakes dominated by candidate division OD1. *ISME J* **6**:1640–1652.

Raymond PA, Hartmann J, Lauerwald R, Sobek S, McDonald C, Hoover M, *et al.* (2013). Global carbon dioxide emissions from inland waters. *Nature* **503**:355–359.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ruiz-González C, Proia L, Ferrera I, Gasol JM, Sabater S. (2013). Effects of large river dam regulation on bacterioplankton community structure. *FEMS Microbiol Ecol* **84**:316–331.

Salcher MM, Pernthaler J, Posch T. (2011). Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria 'that rule the waves' (LD12). *ISME J* **5**:1242–1252.

Seitzinger SP, Mayorga E, Bouwman AF, Kroeze C, Beusen AHW, Billen G, *et al.* (2010). Global river nutrient export: A scenario analysis of past and future trends: GLOBAL RIVER EXPORT SCENARIOS. *Global Biogeochem Cycles* **24**:n/a–n/a.

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.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M. (2012). Next-generation sequencing technologies for environmental DNA research. *Mol Ecol* **21**:1794–1805.

Sinclair L, Osman OA, Bertilsson S, Eiler A. (submitted). Microbial community composition and diversity: evaluating the Illumina platform.

Sjöstedt J, Koch-Schmidt P, Pontarp M, Canbäck B, Tunlid A, Lundberg P, *et al.* (2012). Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. *Appl Environ Microbiol* **78**:1361–1369.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, *et al.* (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci U S A* **103**:12115–12120.

Sommerwerk N, Bloesch J, Paunović M, Baumgartner C, Venohr M, Schneider-Jacoby M, *et al.* (2010). Managing the world's most international river: the Danube River Basin. *Mar Freshw Res* **61**:736–748.

Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, *et al.* (2013). Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *J Appl Microbiol* **115**:1147–1158.

Tu J. (2011). Spatial and temporal relationships between water quality and land use in northern Georgia, USA. *J Integr Env Sci* **8**:151–170.

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. (1980). The River Continuum Concept. *Can J Fish Aquat Sci* **37**:130–137.

Velimirov B, Milosevic N, Kavka G, Farnleitner A, Kirschner AT. (2011). Development of the bacterial compartment along the Danube River: a continuum despite local influences. *Microb Ecol* **61**:955–967.

Weinbauer MG. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**:127–181.

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.

Withers PJA, Jarvie HP. (2008). Delivery and cycling of phosphorus in rivers: A review. *Sci Total Environ* **400**:379–395.

Zampella RA, Procopio NA, Lathrop RG, Dow CL. (2007). Relationship of land-use/land-cover patterns and surface-water quality in the Mullica River basin. *J Am Water Resour Assoc* **43**:594–604.

Zwart G, Byron C. Crump, Miranda P. Kamst-van Agterveld, Ferry Hagen, Suk-Kyun Han. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol* **28**:141–155.

Titles and legends to figures

Figure 1 A Cumulative graph of the quantitative contribution of OTUs based on their presence in the respective sample fraction. X-axis displays the fraction of samples, y-axis shows the cumulative number of reads corresponding to the OTUs that appear in the respective sample fraction. The blue line represents the attached bacterial fraction (>3.0 µm); the red line shows the free-living bacterial fraction (0.2-3.0 µm).
B Gradual development of the read-proportion assigned to the operationally defined "core community" of the free-living and attached fraction along the Danube River. Core communities were defined by including all OTUs that are present in at least 90 percent of all samples of the respective size fraction. Red symbols indicate samples from the free-living fraction (0.2-3.0 µm); Blue symbols indicate the attached sample fraction (>3.0 µm). Dark blue lines represent fitted linear models with confidence intervals of 0.95 in red and blue for the respective fractions. Detailed regression statistics are shown in Table 1.

Figure 2 The gradual development of the bacterial richness (chao1; **A**) and Pielou´s evenness (J; **B**) along the Danube River in the two size fractions, representing the bacterioplankton communiies 0.2-3.0 µm and > 3.0 µm (corresponding to free-living and attached bacterioplankton). Red symbols indicate samples from the free-living fraction (n=27) and blue symbols samples from the attached fraction (n=40). Dark blue lines represent fitted linear models with confidence intervals of 0.95 in red and blue for the respective fractions. Detailed regression statistics are shown in Table 1.

Figure 3 The visualization of the beta diversity analysis based on the Bray-Curtis dissimilarity index performed in order to investigate the compositional dissimilarity between sites along the Danube River and its tributaries. Stress value of the non-metric

multidimensional scaling (NMDS) was 0.17. Circles represent free-living bacterial communities (0.2-3.0 µm), triangles represent attached bacterial communities (>3.0 µm). Open symbols display tributary-samples whereas full symbols indicate Danube River communities. Gradient from blue-black to light blue indicate the position of the sampling site upstream from the river mouth. The official assignment of river kilometres (rkm) for the Danube River is defined reverse from mouth (rkm 0) towards source with our most upstream site at rkm 2 600.

Figure 4 A heatmap (**A**) revealing the dynamics of the eleven most abundant typical freshwater tribes along the Danube river transect according to (Newton *et al.*, 2011). The gradient from light blue to black-blue indicates the relative quantitative contribution to all sequences in any one sample with a maximum of 16 percent. The overall contribution of typical freshwater tribes, clades and lineages (Newton *et al.*, 2011) to the river bacterioplankton amplicon sequences is depicted in panel **B**. Free-living Danube River samples are arranged in the middle of the panel including "F" in the sample-ID; Attached samples are displayed on the right side including "A" in the sample-ID; Both fractions of tributary-samples are arranged at the left side with "F" for free-living and "A" for attached samples in the sample-ID;

Figure 5 Results from the SEQenv analyses scoring sequences according to their environmental context. Y-axis represents the proportion of groundwater (**A**) and soil (**B**) terms associated with sequence reads per sample along the Danube River transect (x-axis). Red symbols indicate samples from the 0.2-3.0 µm fraction (n=27) and blue symbols samples from the >3.0 µm fraction (n=40). Dark blue lines represent fitted linear models with confidence intervals of 0.95 in red and blue for the respective fractions. Detailed regression statistics are given in the figure.

503

504 **Table 1.** Summary of regression statistics (Multiple R-squared and p-value) for fitted linear
505 models between chao1 richness (Fig. 2A), Pielou's evenness (J; Fig. 2B) as well as core-
506 community proportions and river kilometre (Fig. 1B) for both size fractions (0.2-3.0 µm and
507 >3.0 µm) in the investigated Danube River samples.

508

509 **Table 2.** Summary statistics of correspondence between environmental variables and the
510 projections of bacterioplankton community samples in the NMDS ordination based on either
511 free-living or attached size fraction from the Danube River. Results were obtained using
512 function 'envfit' included in the R-package 'vegan' (Oksanen *et al.*, 2013).

513

514 **Figure S1.** Absolute quantitative abundance of OTUs in all samples. Displayed is the
515 decreasing trend from many OTUs with very few representative reads (10□-10²; left side) to
516 fewer OTUs with high abundance (10²-10□ reads; right side). Y-axis displays the number of
517 OTUs belonging to each circle, with the corresponding number of representative reads on the
518 x-axis.

519

520 **Figure S2.** Boxplot of variability in bacterial communities in different size fractions (0-.2-3.0
521 µm and >3.0 µm) based on betadispersion of Bray-Curtis dissimilarities. Left: Variability
522 (distance from centroid) in the free-living bacterial community; Right: Variability in the
523 attached bacterial community.

524

525

	R2	P-value
Chao1 richness free-living	-0.392	<0.001
Pielou's Evenness free-living	-0.483	<0.001
Chao1 richness attached	-0.173	<0.01
Pielou's Evenness attached	-0.501	<0.001
Total read proportion of free-living Core Community	0.455	<0.001
Total read proportion of attached Core Community	0.316	<0.001

Variables	free-living R2	attached R2
River_km	0.844***	0.795***
WaterTemperature	0.264**	0.035
DissolvedOxygen	0.05	0.002
pH	0.330***	0.214**
Conductivity	0.303*	0.093
Alkalinity	0.626***	0.333***
Ammonium	0.117	0.004
Nitrite	0.285**	0.254**
Nitrate	0.705***	0.419***
OrganicNitrogen	0.024	0.076
OrthophosphatePhosphorus	0.186*	0.099.
TotalPhosphorus	0.069	0.059
Silicates_Dissolved	0.504***	0.515***
Phytoplankton_Biomass_Chla	0.179*	0.380***
TotalSuspendedSolids	0.150.	0.345***
Production_total	0.051.	0.397***
DOC	0.042	0.183*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1