INFORMATIVE TITLE:

Bacterial diversity along a 2600 km river continuum

Authors:

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]*

Running title: River bacterioplankton diversity

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

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

email: alexander.eiler@ebc.uu.se

for submission to Environmental Microbiology

Summary

The bacterioplankton diversity in large rivers has thus far been undersampled, despite the importance of streams and rivers as components of continental landscapes. Here, we present a comprehensive dataset detailing the bacterioplankton diversity along a midstream transect of the Danube River and its tributaries. Using 16S rRNA-gene amplicon sequencing, our analysis revealed that bacterial richness and evenness gradually declined downriver in both the free-living and particle-associated bacterial communities. These shifts were also supported by the beta diversity analysis, where the effects of tributaries were negligible in regards to the overall variation. In addition, the river was largely dominated by bacteria that are commonly observed in freshwater and typical of lakes, whereas only few taxa attributed to lotic systems were detected. These freshwater taxa, which were composed of members of the acI lineage and the freshwater SAR11 group (LD12) and the Polynucleobacter, increased in proportion downriver and were accompanied by a decrease in soil and groundwater bacteria. When examining our results in a broader ecological context, we elaborate that patterns of bacterioplankton diversity in large rivers can be explained by the River Continuum Concept published in 1980, with a modification for planktonic microorganisms.

Introduction

Streams and rivers link terrestrial and lentic systems with their marine counterparts and provide numerous essential ecosystem services. They supply drinking water, are used for irrigation, industry, and hydropower, and serve as transport routes or for recreation. Of general importance is the role of lotic systems in biogeochemical nutrient cycling. Until recently, rivers and streams were mainly considered as pipes shuttling organic material and nutrients from the land to the ocean. This view has begun to change as lotic and lentic systems are now considered more akin to “leaky funnels” in regard to the cycling of elements. Indeed, they play an important role in the temporary storage (sequestration) and transformation of terrestrial organic matter (Ensign and Doyle, 2006; Cole *et al.*, 2007; Withers and Jarvie, 2008; Battin *et al.*, 2009). As a result of recognising the role of rivers and streams in the carbon cycle (see for example the report by IPCC in 2013; http://www.ipcc.ch/), the study of the diverse, ongoing processes in the water column and sediments of lotic networks has been receiving increasing interest (Kronvang *et al.*, 1999; Beaulieu *et al.*, 2010; Seitzinger *et al.*, 2010; Aufdenkampe *et al.*, 2011; Benstead and Leigh, 2012; Raymond *et al.*, 2013).

When attempting to model the mechanisms of nutrient processing in freshwater systems, bacteria are regarded as the main transformers of elemental nutrients and viewed as substantial contributors to the energy flow (Cotner and Biddanda, 2002; Battin *et al.*, 2009; Findlay, 2010; Madsen, 2011). However, in the case of open lotic systems such as rivers, there remains a major lack of knowledge concerning the diversity of bacterial communities and the link between diversity and ecosystem functioning (Battin *et al.*, 2009). There is currently no agreement on the distinctness of the river bacterioplankton from that of other freshwater systems or the variability of its diversity along entire river transects. More generally, the question of what regulates this diversity remains open.

When summarising previous studies, it can be concluded that bacteria affiliated with the phyla of *Proteobacteria (*particularly *Betaproteobacteria), Bacteroidetes, Cyanobacteria a*nd

Verrucomicrobia have dominated 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 bacteria suggest that the abundant taxa comprising the riverine bacterioplankton form a cohesive group and can thus be regarded as "typical" freshwater bacteria (Zwart *et al.*, 2002; Lozupone and Knight, 2007; Newton *et al.*, 2011). Nevertheless, the previous studies were constrained by their low sequencing depth and focus on the dominant members of the communities.

Yet, a reasonable sequencing depth is a requirement to correctly estimate the community diversity and identify fine-scale changes that occur as responses to the fluctuating environmental conditions. In one study, 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 widespread availability of second-generation sequencing technologies (Shokralla *et al.*, 2012). By targeting the short hyper-variable regions of the universal 16S rRNA gene and proceeding with ultra-deep sequencing, one not only allows for a proper investigation of the diversity and the richness of a community but also uncovers the ability to detect and investigate rare populations that may bear critical functions (Sogin *et al.*, 2006; Gilbert *et al.*, 2012; Sjöstedt *et al.*, 2012).

Regarding large rivers, microbial community studies using second-generation sequencing are scarce with only a few available concerning bacterioplankton. These publications include studies of the Amazonas River (Brazil), the Upper Mississippi River (USA), the Columbia River Estuary (USA) and the Yenisey River mainly reveal taxonomic patterns (Ghai *et al.*, 2011; Fortunato *et al.*, 2013; Staley *et al.*, 2013; Kolmakova *et al.*, 2014). Moreover, the longitudinal development of the bacterioplankton communities along the rivers could not be addressed comprehensively because only a few sites were analysed in each case. Considering

the environmental gradients that develop along such rivers (Sekiguchi *et al.*, 2002; Winter *et al.*, 2007; Velimirov *et al.*, 2011), it is expected that the bacterial communities will show a similar variation in their composition and function as one travels from the source to the mouth.

This variability has been hypothesised to originate from the import of bacteria through terrestrial illuviation and merging tributaries as well as from anthropogenic contributions such as wastewater treatment plant pollution. More diffuse phenomena such as soil erosion and agriculture should also be considered (Zampella *et al.*, 2007; Tu, 2011; Besemer *et al.*, 2012).

In the case of macroorganisms, an attempt to summarise the large-scale diversity patterns observed from headwater streams to large rivers has been undertaken with the publication of the River Continuum Concept (RCC). The RCC proposes that diversity increases from headwaters to medium-sized stream reaches, with a subsequent decrease towards the river mouth. It is suggested that this pattern is due to the gradient of physical factors formed by the drainage network, the dynamics in chemical properties and the resulting biological activity (Vannote *et al.*, 1980).

Here, we attempted to extend the RCC to include river bacterioplankton by utilising the results from a second-generation sequencing experiment detailing the bacterial community composition along a large river. Furthermore, we revealed how the variability in bacterioplankton diversity is related to the environmental variables along a continuous river transect spanning 2600 km from medium-sized reaches to the river mouth. We separately investigated the free-living communities and particle-associated communities by extracting two different size fractions (0.2-3.0 µm and >3.0 µm) for each sample. These two fractions have been shown to exhibit significant differences in activity and community dynamics in previous studies, justifying this distinction (Crump *et al.*, 1999; Velimirov *et al.*, 2011). The study site was the Danube River, the second largest river in Europe by discharge and length.

The Danube River drains a basin of approximately 801 000 km²; the area is populated with 83 million inhabitants and borders 19 countries (Sommerwerk *et al.*, 2010).

Results

General description of sequences

In total, DNA was extracted and sequenced from 132 filtered water samples originating from the Danube River and its tributaries. In addition, the same procedure was applied to 5 negative control samples. The sequencing yielded 2 030 029 read pairs ranging from 3451 to 24 873 per sample. After quality filtering and mate-pair joining as outlined in Sinclair *et al.* (in review; see Supporting information), 1 572 361 sequence reads (further referred to as "reads") were obtained.

The OTU clustering resulted in 8697 OTUs after the removal of all Plastid-, Mitochondrion-, *Thaumarchaetoa*-, *Crenarchaeota*- and *Euryarchaeota*-assigned OTUs. These undesirable sequences represented 19.1% of the reads and accounted for 625 OTUs. Next, for the alpha diversity analysis, any sample with less than 7000 reads was excluded, resulting in 8241 OTUs in the remaining 88 samples. By contrast, for the beta diversity analysis, all samples were randomly rarefied to the lowest number of reads in any one sample. This brought every sample down to 2347 reads, and any OTU containing less than two reads was discarded, which brought the total OTU count to 5082.

Core microbial community

The majority of bacteria-assigned OTUs (4402 out of 8697) were only represented by less than ten reads in the entire dataset. As a consequence, 3243 of 8697 OTUs (~37%) were present in only one to four samples, and an additional 2219 OTUs (~26%) were present in as few as five to nine samples. In addition to these rare OTUs, the core community of the Danube River, defined by all OTUs that appeared in at least 90% of all samples, was

comprised of 89 OTUs for the free-living bacterioplankton (0.2-3.0 µm) and 141 OTUs for the particle-associated microbes (>3.0 µm). The cumulative contribution of OTUs based on their occurrence along the entire river transect is shown in Fig. 1A. for both analysed size fractions. On average, 81% of all reads of the free-living river community and 63% of all reads of the particle-associated river community were part of their respective core community. A significant increase in the relative contribution of the core communities could be observed towards the river mouth for both fractions (see Fig. 1B.).

Variability of diversity along the river

The Chao1 richness estimator and Pielou's evenness index were calculated for both size fractions after adjusting all samples down to 7000 reads and discarding those that did not obtain enough reads (n=44). The estimated richness was persistently higher in the particle-associated fraction when compared to the free-living fraction with averages of 2025 OTUs and 1248 OTUs, respectively. We observed the highest diversity of all samples in the medium-sized stretches of the upstream parts of the Danube River. The richness and evenness gradually decreased downstream in both size fractions, as confirmed by the regression analysis (Table 1). The gradual development of the communities can be visualised by applying non-metric multidimensional scaling (NMDS) to the beta diversity distance matrix (Fig. 3.). In both size fractions, a significant relationship between community composition and river kilometre was observed (Table 2). The additional environmental variables that corresponded with the compositional dynamics are given in Table 2, excluding tributaries. As shown in the NMDS, the tributaries did not follow the general patterns and often formed outliers in the ordination plot.

Moreover, based on their bacterial composition, a clear separation was formed between the two filter fractions, as confirmed by PERMANOVA analysis (R^2=0.156, p-value<0.01). The apparent synchrony in the gradual development of the two size fractions along the river's

course could also be demonstrated using a Procrustes test (R=0.96, p<0.001). Nevertheless, the application of a permutation test to the beta dispersion values of each size fraction revealed a significantly higher variability in the >3.0 µm fraction when compared to the 0.2-3.0 µm fraction (p-value=0.002) (see Fig. S2.).

Typical river bacterioplankton

We used the 9322 total OTUs to perform a similarity search against the database of freshwater bacteria 16S rRNA sequences developed by Newton and colleagues (2011). The analysis revealed that a large proportion of the bacterial population inhabiting the Danube could be assigned to previously described freshwater taxa (Fig. 4.). In particular, these included representatives of the LD12-tribe belonging to the subphylum of *Alphaproteobacteria*, as well as the acI-B1-, acI-A7- and acI-C2-tribes belonging to the phylum *Actinobacteria*.

Interestingly, in the free-living size fraction, an increase in the relative abundance of the four previously mentioned tribes was clearly observed towards the river mouth (Fig. 4A.), contributing up to 35% of the community. Correspondingly, it is possible to observe a clear decrease in the proportion of atypical freshwater taxa in the free-living fraction (labelled “Everything else”) with an increasing number of OTUs assigned to the tribe-level as one goes down the river (Fig. 4B.). In the particle-associated fraction, these typical freshwater taxa are much less abundant (Fig. 4B.). Nevertheless, the OTUs that could be assigned to typical freshwater taxa increased downriver.

In a similar manner, the 8697 bacterial OTUs were BLASTed against the NCBI-NT database; next, any environmental descriptive terms occurring in the search results were retrieved and classified according to the Environmental Ontology (EnvO) terminology. By running a PERMANOVA analysis, we confirmed that the bacterial communities of the different size fractions have distinct habitat preferences (PERMANOVA; R^2=0.42, p<0.0001). Restricting the analysis to only 'groundwater' and 'soil' terms indicated that the proportion of bacteria

199 potentially originating from these two sources decreased towards the river mouth (Fig. 5A.
200 and B.). By using only the contribution of 'river' and 'sediment' terms, contrary to our
201 expectations, we could not demonstrate any trend along the river transect. It is worth noting
202 that by applying this procedure, most OTUs were not dominated by the 'river' environmental
203 term and only a total of four OTUs received an ontology comprising 50% or more of the
204 'river' keyword.
205
206 **Discussion**
207 The tremendous diversity within the microbial communities inhabiting all types of aquatic
208 environments is being revealed by a rapidly increasing number of studies applying high-
209 throughput sequencing technologies to environmental samples (e.g. Sogin *et al*., 2006;
210 Andersson *et al*., 2009; Galand *et al*., 2009; Eiler *et al*., 2012; Peura *et al*., 2012). At the same
211 time, the factors modulating this diversity are also being described (Besemer *et al*., 2012;
212 Hanson *et al*., 2012; Lindström and Langenheder, 2012; Szekely *et al*., 2013). However, very
213 few studies investigating river bacterioplankton are available, and all the studies are based on
214 either relatively small sample sets (Ghai *et al*., 2011; Fortunato *et al*., 2013; Staley *et al*.,
215 2013; Kolmakova *et al*., 2014) or are of low resolution (Sekiguchi *et al*., 2002; Winter *et al*.,
216 2007; Liu *et al*., 2011, 2012). Here, we describe the diversity of lotic bacterioplankton along a
217 2600 kilometre transect using high spatial and taxonomic resolution and explain the observed
218 patterns in the context of the River Continuum Concept (RCC; Vannote *et al*., 1980).
219
220 *Towards a typical freshwater bacteria community along the river*
221 In addition to an obvious gradual change in beta diversity, we recorded a significant decrease
222 in bacterial richness and evenness in the free-living and particle-associated communities along
223 the river. The gradual change in beta diversity not only significantly correlated with river
224 kilometre but also correlated with alkalinity, dissolved silicates, and nitrate. In addition, the

particle-associated community composition correlated significantly with phytoplankton biomass, total suspended solids and total bacterial production. As particles derived from autochthonous matter increased downriver, these correlations, together with the accompanied change in the particle-associated communities, point towards a distinction between communities inhabiting autochthonous and allochthonous particles.

Another distinction was the remarkably higher richness found in particle-attached communities when compared to the free-living bacterioplankton fraction. We ascribe this phenomenon to the higher availability of distinct ecological niches inside and directly on the particles. The suspended particles not only included detritus derived from terrestrial and aquatic sources or mobilised sediments but also included living organisms such as planktonic algae or zooplankton. Therefore, the high diversity of particles in combination with the resulting spectrum of microenvironments (including anoxic habitats) provides an explanation for the higher richness observed in the particle-associated fraction. Furthermore, differences in diversity between the two size fractions were apparent in the results of the EnvO term analysis, indicating the distinct habitat preferences of bacteria. Taking a closer look at these results, we found that the proportion of bacteria originating from soils and groundwater sources was constantly higher in the particle-associated communities, which is likely due to the quantity of suspended particles from soils.

In addition to the riparian zones, merging tributaries or microbial pollution sources could be providing allochthonous particles and bacteria to the river. However, we argue that the gradual exchange of soil and groundwater bacteria with typical freshwater bacterioplankton along the midstream of the river is mostly unaffected by the merging tributaries. This can be explained by estimating the mixing behaviour of the most important tributaries, as conducted by Velimirov and colleagues (2011). In this publication, the authors proposed that the tributaries and other point sources have a negligible effect on the composition of the midstream communities due to the long mixing times of the incoming water and the restrained dilution

that this entails. This hypothesis was based on their prior observation of a gradual change in bacterial counts, cell volumes and morphotype composition, which were all significantly correlated with several physicochemical parameters. A few years later, Kolmakova and colleagues (2014) also reported that in the case of the of the Yenisei river, large incoming tributaries could conserve a parallel flow and only merge into the main stream over several kilometres.

Focusing on the taxonomic composition, our data shows that “typical” freshwater bacteria, including members of the acI lineage (c.f. Newton *et al*., 2011), the freshwater SAR11 group (LD12) and the *Polynucleobacter* genus, formed to a major part the bacterial “core community”, particularly in the free-living fraction. The close resemblance between riverine bacterial communities and those of lakes strongly corroborates the existence of a so-called “typical freshwater bacteria” group (Zwart *et al*., 2002; Lozupone and Knight, 2007; Newton *et al*., 2011).

Explaining patterns in river bacterioplankton

The observed shift towards a more typical freshwater bacteria-dominated community is most likely driven by decreasing inputs of allochthonous bacteria to the midstream from soils and groundwaters on the one hand and by competitive advantages of these taxa on the other hand. The first explanation is supported by previous observations where bacterial communities were similar to or at least heavily impacted by soil communities (Besemer *et al*., 2012; Crump *et al*., 2012). In this regard, it was stated that the inputs of allochthonous organisms outweigh the rate of local extinction (e.g. Leibold *et al*., 2004). The second argument, a downstream rise in competitiveness of downstream-specific OTUs, is suggested by the observed simultaneous decrease in evenness together with bacterial richness in both size fractions along the river transect. Such a rise of few OTUs with a competitive advantage downriver was already predicted by the RCC (Vannote *et al*., 1980).

The RCC proposes that more refractory and relatively high molecular weight compounds are exported downstream and accumulate along the river, whereas labile allochthonous organic compounds are rapidly used by heterotrophic organisms or physically absorbed in the headwaters. In this case, the highest diversity of soluble organic compounds was proposed to be due to a maximum interface with the landscape (Vannote *et al*., 1980). Our assumption that downstream-specific OTUs possess a competitive advantage in utilising nutrient-poor organic compounds is also supported by the increasing relative abundance of typical freshwater taxa such as LD12 and acI, which represent small cells with an oligotrophic lifestyle (Salcher *et al*., 2011; Garcia *et al*., 2013). A general trend towards smaller cells along the Danube River was previously described by Velimirov *et al*. (2011), which potentially highlights the decreasing availability of nutrients (larger surface-to-volume-ratio). In addition to the selection for smaller cells based on competitive advantages of oligotrophic bacteria, the starvation of copiotrophic cells originating from terrestrial sources, which are better adapted to higher quality and nutrient-rich compounds (Barcina *et al*., 1997), could contribute to the trend towards smaller cell volumes.

To demonstrate the role of organic matter sources in the apparent decline of richness towards the river mouth, an assessment of the organic matter composition and bioavailability should be included in future studies. Furthermore, loss factors such as sedimentation and (selective) top-down control such as grazing and viral lysis have been shown to vary over environmental gradients and can substantially influence microbial diversity (Ayo *et al*., 2001; Langenheder and Jürgens, 2001; Weinbauer, 2004; Pernthaler, 2005; Bouvier and Del Giorgio, 2007).

Necessary adjustments to the RCC for the application to river bacterioplankton

When combining ours and previous results (Besemer *et al*., 2012, 2013; Crump *et al*., 2012; Staley *et al*., 2013), we propose that the RCC – although developed for macroorganisms – can be transferred to river bacterioplankton. For macroorganisms, the RCC proposes the highest

diversity in medium-sized streams, which is mainly based on their model parameter of diel temperature variability. However, for bacterioplankton, we observed a continuous decrease from headwaters to river mouth, which could be interpreted to be in conflict with the RCC. Since Vannote and colleagues did not consider bacterioplankton, we argue that the RCC is open for interpretation in this respect. Nevertheless, this requires the careful illumination of the following important points: (i) The primarily passive transport of bacterioplankton contrasts the motility and sessility of macroorganisms, such as aquatic invertebrates, fish or macrophytes; (ii) the large contact zone of small rivers and the surrounding environment (soil and groundwater) is constantly contributing allochthonous microbes to the river bacterioplankton community (Besemer *et al*., 2012; Crump *et al*., 2012); (iii) soil communities harbour a much higher diversity when compared to planktonic communities (e.g., Crump *et al*., 2012); (iv) these allochthonous bacteria should be at least temporarily capable of proliferating in their new lotic environment when compared to, e.g., terrestrial insects that fall or are washed into streams or rivers.

The elevated contribution of allochthonous bacteria to the upstream river bacterioplankton is corroborated by our results of the SEQenv analysis (Fig. 5A and B), where an increased impact from soil and groundwater bacteria to the communities was detected. The importance of the impact from the riparian zone on suspended microbial communities was also reported in previous studies on headwater stream networks and the runoff-process from hill slopes via headwaters to a lake, suggesting terrestrial environments as critical reservoirs of microbial diversity for downstream surface waters (Besemer *et al*., 2012, 2013; Crump *et al*., 2012). Crump and colleagues found that the dominant bacteria in an arctic lake were all first observed in soil waters and other upslope environments draining into the system. Additional support for a steady decrease in diversity from headwaters to river mouths was provided by a similar decreasing trend in microbial diversity of benthic microbial biofilms from headwaters

to mid-sized streams, which are proposed to be settled by the suspended bacterial community (Besemer *et al.*, 2012, 2013).

Taking these features into account, we propose that patterns in bacterioplankton diversity can indeed be incorporated into the RCC. By highlighting the riparian zone, substrate availability and flow as important determinants of community structure, Vannote and colleagues already provided a conceptual framework to explain the patterns of bacterioplankton diversity in both size fractions along the Danube River. In addition, an increase in the competitiveness of several freshwater taxa attributable to an increase in stability and uniformity of the system along the river continuum is in accordance with the RCC. Furthermore, our study shows that the influence of dispersal from soil, groundwaters and other allochthonous sources in determining the patterns of diversity decreased downriver, whereas internal processes, such as the impact of environmental conditions in rivers, increased in importance. Although we were able to show that the contribution of dispersal and environmental conditions in determining community composition was linked to hydrology, the link between the patterns of diversity and ecosystem function remains to be determined.

Experimental Procedures

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).

Study sites and sample collection

Samples were collected within the frame of the second Joint Danube Survey project (JDS 2) in 2007. The overall purpose of the Joint Danube Surveys is to produce a comprehensive evaluation of the chemical and ecological status of the entire Danube River on the basis of the European Union Water Framework Directive (WFD) (Liska *et al.*, 2008). During sampling from Aug 15th to Sept 26th 2007, 75 sites were sampled along the mainstream of the Danube River along its shippable way from river kilometre (rkm) 2600 to the river mouth at rkm 0 (Kirschner *et al.*, 2009; Fig. S1.). In addition, 21 samples from the Danube's major tributaries and branches were included. At the most upstream sites, the Danube River is representative of a typical stream of the rithron and characterised by its tributaries Iller, Lech and Isar (Kavka and Poetsch, 2002). The trip took 43 days and is equivalent to the average retention time of a water body in this part of the Danube River (for discussion of this issue, see Velimirov *et al.*, 2011). Samples were collected with sterile 1 L glass flasks from a water depth of approximately 30 cm. Glass flasks were sterilised by rinsing with 0.5% HNO_3 and autoclaving them. For DNA extraction of the particle-associated 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, which represented 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.

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

applying the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies Corporation, USA), and 16S rRNA genes were quantified using domain-specific quantitative PCR. Quantitative PCR reactions contained 2.5 µL of 1:4 and 1:16 diluted DNA extract as the template, 0.2 µM of primers 8F and 338 (Frank *et al*., 2007; Fierer *et al*., 2008) targeting the V1-V2 region of most bacterial 16S rRNA genes and iQ™ SYBR® Green Supermix (Bio-Rad Laboratories, Hercules, USA). All primer information is available in Table S1. The ratios of measured 16S rRNA gene copy numbers in the different sample dilutions that deviated markedly from 1 after multiplication with the respective dilution factor were interpreted as an indicator for PCR-inhibition.

Preparation of 16S rRNA gene amplicon libraries

For the preparation of amplicon libraries, 16S rRNA genes were amplified and barcoded in a two-step procedure to reduce PCR bias that is introduced by long primers and sequencing adaptor-overhangs (Berry *et al*., 2011). We followed the protocol as described by Sinclair *et al*. (unpublished, see Supporting information). In short, 16S rRNA gene fragments of most bacteria were amplified by applying primers Bakt_341F and Bakt_805R (Herlemann *et al*., 2011; Table S1) targeting the V3-V4 variable regions. In 25 µL reactions containing 0.5 µM primer Bakt_341F and Bakt_805R, 0.2 µM dNTPs (Invitrogen), 0.5 U Q5 HF DNA polymerase and the provided buffer (New England Biolabs, USA), genomic DNA was amplified in duplicate in 20 cycles. To use equal amounts of bacterial template DNA to increase the comparability and reduction of PCR bias, the final volume of environmental DNA extract used for each sample was calculated based on 16S rRNA gene copy concentration in the respective sample determined earlier by quantitative PCR (see above). For 105 samples, the self-defined optimum volume of environmental DNA extract corresponding to 6.4×10^5 16S rRNA genes was spiked into the first step PCR reactions; however, for 27 samples, lower concentrations were used due to limited amounts of bacterial

406 genomic DNA or PCR inhibition detected by quantitative PCR (see above). These 132
407 samples included eight biological replicates. Prior to the analysis, we removed four samples
408 due to their extremely low genomic DNA concentrations and 16S rRNA gene copy numbers.
409 Duplicates of PCR products were pooled, diluted to 1:100 and used as templates in the
410 subsequent barcoding PCR. In this PCR, diluted 16S rRNA gene amplicons were amplified
411 using 50 primer pairs with unique barcode pairs (Sinclair *et al.*, in review; Table S1). The
412 barcoding PCRs for most samples were conducted in triplicates analogous to the first PCR
413 (n=100). The remaining 32 samples that had weak bands in first step PCR due to low genomic
414 template DNA concentrations or high sample dilution were amplified in 6-9 replicates to
415 increase amplicon DNA yield. Barcoded PCR amplicons were pooled in an equimolar fashion
416 after purification using the Agencourt AMPure XP purification system (Beckman Coulter,
417 Danvers, MA, USA) and quantification of amplicon-concentration using the Quant-iT™
418 PicoGreen® dsDNA Assay Kit (Life Technologies Corporation, USA). Finally, a total of 137
419 samples including 5 negative controls resulted in four pools for sequencing.
420
421 *Illumina® sequencing*
422 The sequencing was performed on an Illumina® MiSeq at the SciLife Lab Uppsala. For each
423 pool, the library preparation was performed separately following the TruSeq protocol with the
424 exception of the initial fragmentation and size selection procedures. This involves the binding
425 of the standard sequencing adapters in combination with separate Illumina®-specific MID
426 barcodes that enables the combination of different pools on the same sequencing run (Sinclair
427 *et al.*, unpublished). After pooling, random PhiX DNA was added to provide calibration and
428 help with the cluster generation on the MiSeq's flow cell.
429
430 *16S rRNA gene amplicon data analysis*

The sequence data were processed as outlined in Sinclair *et al.* (in review). In short, after sequencing the libraries of 16S rRNA amplicons, the read pairs were demultiplexed and joined using the PANDAseq software (Masella *et al.*, 2012). Next, reads that did not bear the correct primer sequences at the start and end of their sequences were discarded. Reads were then filtered based on their PHRED scores. Chimera removal and OTU (operational taxonomic unit) clustering at 3% sequence dissimilarity was performed by pooling all reads from all samples together and applying the UPARSE algorithm (Edgar, 2013). Here, any OTU containing less than two reads was discarded. Each OTU was subsequently taxonomically classified by operating a similarity search against the SILVAmod database and employing the CREST assignment algorithm (Lanzén *et al*., 2012). Plastid, mitochondrial and archaeal OTUs were removed. In addition, OTUs were also taxonomically annotated against the freshwater database (Newton *et al*., 2011) using the same method. If necessary, OTU rarefying for the purpose of standardising sequence numbers was performed using the 'rrarefy'-function implemented in the R-package vegan (Oksanen *et al*., 2013). Biodiversity measure calculation, statistical analyses and plot-generation were conducted in R (R Core Team, 2013) using python scripts. The habitat index for the top 5000 OTUs was determined using the SEQenv pipeline (http://environments.hcmr.gr/seqenv.html). The 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 associated contextual information records ("Isolation Source" field entry for genomes in Genbank or associated PubMed abstracts). At the time of running SEQenv on our dataset (version 0.8), there were approximately 1200 EnvO terms organised into three main branches (namely, *environmental material*, *environmental feature*, and *biome*). However, we used SEQenv to retrieve a subset of these terms, i.e., those that contain "Habitat" (ENVO:00002036). Raw sequence data were submitted to the NCBI Sequence Read Archive (SRA) under accession number SRP045083.

Acknowledgements

This study was supported by the Austrian Science Fund (FWF) as part of the DKplus "Vienna Doctoral Program on Water Resource Systems" (W1219-N22) and the FWF project P25817-B22, as well as the research project "Groundwater Resource Systems Vienna" in cooperation with Vienna Water (MA31). AE and LS are funded by the Swedish Foundation for Strategic Research (ICA10-0015). Infrastructure (cruise ships, floating laboratory) and logistics for collecting, storing and transporting samples were provided by the International Commission for the Protection of the Danube River (ICPDR). The analyses were performed using resources provided by the SNIC through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project "b2011035".

Conflict of Interest Statement

The authors declare no conflict of interest.

References

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

Aufdenkampe, A.K., Mayorga, E., Raymond, P.A., Melack, J.M., Doney, S.C., Alin, S.R., *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., and 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., and Vives-Rego, J. (1997) Survival of allochthonous bacteria in aquatic systems: a biological approach. *FEMS Microbiol Ecol* **23**: 1–9.

Battin, T.J., Luyssaert, S., Kaplan, L.A., Aufdenkampe, A.K., Richter, A., and Tranvik, L.J. (2009) The boundless carbon cycle. *Nat Geosci* **2**: 598–600.

Beaulieu, J.J., Tank, J.L., Hamilton, S.K., Wollheim, W.M., Hall, R.O., Mulholland, P.J., *et al.* (2010) Nitrous oxide emission from denitrification in stream and river networks. *Proc Natl Acad Sci* **108**: 214–219.

Benstead, J.P. and Leigh, D.S. (2012) An expanded role for river networks. *Nat Geosci* **5**: 678–679.

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

Besemer, K., Peter, H., Logue, J.B., Langenheder, S., Lindström, E.S., Tranvik, L.J., and Battin, T.J. (2012) Unraveling assembly of stream biofilm communities. *ISME J* **6**: 1459–1468.

Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W., and Battin, T.J. (2013) Headwaters are critical reservoirs of microbial diversity for fluvial networks. *Proc R Soc B Biol Sci* **280**:

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

Cole, J.J., Prairie, Y.T., Caraco, N.F., McDowell, W.H., Tranvik, L.J., Striegl, R.G., *et al.* (2007) Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems* **10**: 172–185.

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

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

Crump, B.C., Amaral-Zettler, L.A., and Kling, G.W. (2012) Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J* **6**: 1629–1639.

Crump, B.C., Armbrust, E.V., and Baross, J.A. (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, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**: 996–998.

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

Ensign, S.H. and Doyle, M.W. (2006) Nutrient spiraling in streams and river networks. *J Geophys Res* **111**: doi:10.1029/2005JG000114

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

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

Fortunato, C.S., Eiler, A., Herfort, L., Needoba, J.A., Peterson, T.D., and Crump, B.C. (2013) Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. *ISME J* **7**: 1899–1911.

Frank, D.N., St. Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci* **104**: 13780–13785.

Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci* **106**: 22427–22432.

Garcia, S.L., McMahon, K.D., 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, K.D., 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**:

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

Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G., and Bailey, M.J. (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.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Micro* **10**: 497–506.

Herlemann, D.P., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011) Transitions in bacterial communities along the 2000km salinity gradient of the Baltic Sea. *ISME J* **5**: 1571–1579.

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

Kavka, G.G. and Poetsch, E. (2002) Microbiology. In, Literathy,P., Koller-Kreinel,V., and Liska,I. (eds), *Joint Danube Survey: Technical report of the international commision for the protection of the Danube River*. ICPDR, Vienna, pp. 138–150.

Kirschner, A.K.T., Kavka, G.G., Velimirov, B., Mach, R.L., Sommer, R., and Farnleitner, A.H. (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.

Kolmakova, O.V., Gladyshev, M.I., Rozanov, A.S., Peltek, S.E., and Trusova, M.Y. (2014) Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation sequencing. *FEMS Microbiol Ecol* doi: 10.1111/1574-6941.12355

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

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

Lanzén, A., Jørgensen, S.L., Huson, D.H., Gorfer, M., Grindhaug, S.H., Jonassen, I., *et al.* (2012) CREST – Classification Resources for Environmental Sequence Tags. *PLoS ONE* **7**:

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* **7**: 601–613.

Lemke, M.J., Lienau, E.K., Rothe, J., Pagioro, T.A., Rosenfeld, J., and DeSalle, R. (2008) Description of freshwater bacterial assemblages from the Upper Paraná River floodpulse system, Brazil. *Microb Ecol* **57**: 94–103.

Lindström, E.S. and Langenheder, S. (2012) Local and regional factors influencing bacterial community assembly. *Environ Microbiol Rep* **4**: 1–9.

Liska, I., Slobodnik, J., and Wagner, F. (2008) Joint Danube Survey 2, Final Scientific Report. *Int. Comm. Prot. Danube River*.

Liu, L., Yang, J., and Zhang, Y. (2011) Genetic diversity patterns of microbial communities in a subtropical riverine ecosystem (Jiulong River, southeast China). *Hydrobiologia* **678**: 113–125.

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

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

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

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

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

Mueller-Spitz, S.R., Goetz, G.W., and McLellan, S.L. (2009) Temporal and spatial variability in nearshore bacterioplankton communities of Lake Michigan. *FEMS Microbiol Ecol* **67**: 511–522.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., and Bertilsson, S. (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* **75**: 14–49.

Von der Ohe, P.C., 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, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *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., and Jones, R.I. (2012) Distinct and diverse anaerobic bacterial communities in boreal lakes dominated by candidate division OD1. *ISME J* **6**: 1640–1652.

Raymond, P.A., 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/.

Salcher, M.M., Pernthaler, J., and 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, S.P., Mayorga, E., Bouwman, A.F., Kroeze, C., Beusen, A.H.W., Billen, G., *et al.* (2010) Global river nutrient export: A scenario analysis of past and future trends: GLOBAL RIVER EXPORT SCENARIOS. *Glob Biogeochem Cycles* **24**: n/a–n/a.

Sekiguchi, H., Watanabe, M., Nakahara, T., Xu, B., and 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, J.L., Gibson, J.F., and Hajibabaei, M. (2012) Next-generation sequencing technologies for environmental DNA research. *Mol Ecol* **21**: 1794–1805.

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, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al.* (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc Natl Acad Sci* **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, T.J., Jarvis, B., Phillips, J., Cotner, J.B., and Sadowsky, M.J. (2013) Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *J Appl Microbiol* **115**: 1147–1158.

Szekely, A.J., Berga, M., and Langenheder, S. (2013) Mechanisms determining the fate of dispersed bacterial communities in new environments. *ISME J* **7**: 61–71.

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

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R., and Cushing, C.E. (1980) The River Continuum Concept. *Can J Fish Aquat Sci* **37**: 130–137.

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

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

Winter, C., Hein, T., Kavka, G., Mach, R.L., and Farnleitner, A.H. (2007) Longitudinal changes in the bacterial community composition of the Danube River: a whole-river approach. *Appl Environ Microbiol* **73**: 421–431.

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

Zampella, R.A., Procopio, N.A., Lathrop, R.G., and Dow, C.L. (2007) Relationship of land-use/land-cover patterns and surface-water quality in the Mullica River basin. *JAWRA J Am Water Resour Assoc* **43**: 594–604.

Zwart, G., Byron C. Crump, Miranda P. Kamst-van Agterveld, Ferry Hagen, and 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.

474

475

Table and Figure legends

Fig. 1A. Cumulative graph of the quantitative contribution of OTUs based on their presence in the respective sample fraction. The X-axis displays the fraction of samples, and the Y-axis shows the cumulative number of reads corresponding to the OTUs that appear in the respective sample fraction. The blue line represents the particle-associated 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 particle-associated fraction along the Danube River. Core communities were defined by including all OTUs that are present in at least 90% 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 particle-associated 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.

Fig. 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 communities of 0.2-3.0 µm and > 3.0 µm (corresponding to free-living and particle-associated bacterioplankton, respectively). Red symbols indicate samples from the free-living fraction (n=27); blue symbols samples from the particle-associated 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.

Fig. 3. The visualisation of the beta diversity analysis based on the Bray-Curtis dissimilarity index shows the compositional dissimilarity between sites along the Danube River and its tributaries. The 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 particle-associated bacterial communities (>3.0 µm). Open symbols display tributary samples, whereas full symbols indicate Danube River communities. The gradient from blue-black to light blue indicates the position of the sampling site upstream from the river mouth. The official assignment of river kilometres (rkm) for the Danube River is defined in a reverse fashion starting from the mouth at rkm 0 and progressing towards the source with our most upstream site at rkm 2600.

Fig. 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%. 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 on the left side of the panel including “F” in the sample ID; Particle-associated samples are displayed in the middle including “A” in the sample ID; both fractions of tributary samples are arranged at the right side with “F” for free-living and “A” for particle-associated samples in the sample ID.

Fig. 5. Results from the SEQenv analyses scoring sequences according to their environmental context. The 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 indicate 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.

Table 1. Summary of regression statistics (multiple R-squared and p-value) for fitted linear models between Chao1 richness (Fig. 2A), Pielou's evenness (J; Fig. 2B), and the core community proportions and river kilometre (Fig. 1B) for both size fractions (0.2-3.0 µm and >3.0 µm) in the investigated Danube River samples.

Table 2. Summary statistics of correspondence between environmental variables and the projections of bacterioplankton community samples in the NMDS ordination based on either free-living or particle-associated fractions from the Danube River. The results were obtained using the function 'envfit' included in the R-package 'vegan' (Oksanen *et al.*, 2013).

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig S1. Map showing all sampling sites along the Danube River that were sampled during the Joint Danube Survey 2 according to Liska et al. (2008)

Table S1. List of used primers and barcodes for Illumina®-Sequencing.

Sinclair et al. (in review). Manuscript under revision containing information about experimental procedures regarding Illumina®-Sequencing and bioinformatical data-analysis.

Table S2. Results of all measured environmental and chemical parameters during the Joint Danube Survey 2. Copy of JDS2-database-content.