Bacterial metacommunity organization in a highly connected aquatic system

Langenheder, Silke

2017-04


http://hdl.handle.net/10138/214437
https://doi.org/10.1093/femsec/fiw225

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
Bacterial metacommunity organization in a highly-connected aquatic system

Silke Langenheder\(^1\), Jianjun Wang\(^2\), Satu-Maaria Karjalainen\(^4\), Tiina M. Laamanen\(^4\), Kimmo Tolonen\(^5\), Annika Vilmi\(^5\) and Jani Heino\(^5\)

\(^1\) Department of Ecology and Genetics/Limnology, Uppsala University, Sweden
\(^2\) Department of Geosciences and Geography, University of Helsinki, Helsinki FIN-00014, Finland
\(^3\) State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academic of Sciences, Nanjing 210008, China
\(^4\) Finnish Environment Institute, Freshwater Centre, FI-90570 Oulu, Finland
\(^5\) Finnish Environment Institute, Natural Environment Centre, Biodiversity, FI-90570 Oulu, Finland

for correspondence: silke.langenheder@ebc.uu.se

Note: This is final submitted version.
Abstract

The spatial structure and underlying assembly mechanisms of bacterial communities have been widely studied across aquatic systems, focusing primarily on isolated sites, such as different lakes, ponds and streams. In contrast, biodiversity patterns within large aquatic systems have received less attention. Here, our main aim was therefore to determine the underlying mechanisms for biofilm bacterial assemblages within a large, highly-connected lake system in Northern Finland using associative approaches based on taxonomic and phylogenetic alpha- and beta-diversity and a large number of abiotic and biotic variables. Furthermore, null model approaches were used to quantify the relative importance of different community assembly processes. We found that the spatial variations in bacterial communities within the lake were structured by a combination of different assembly processes, including stochasticity, species sorting and potentially even dispersal limitation. Species sorting by abiotic environmental conditions explained more of the taxonomic and particularly phylogenetic turnover in community composition compared to that by biotic variables. Finally, we observed clear differences in alpha diversity (species richness and phylogenetic diversity), which were to a stronger extent determined by abiotic compared to biotic factors, but also by dispersal effects. In summary, our study shows that the biodiversity of bacterial biofilm communities in a highly-connected lake ecosystem is driven by within-habitat gradients in abiotic conditions as well as by stochastic and deterministic dispersal processes.
Introduction

In recent decades, several related conceptual frameworks have been developed to explain why the composition of ecological communities varies across space (e.g. Leibold et al., 2004; Vellend, 2010). Firstly, spatial variations in community can be due to species sorting (i.e. environmental filtering, habitat filtering or environmental selection) when species are selected by different abiotic and biotic conditions that prevail at different locations.

Deviations from pure species sorting can occur if dispersal is limiting, so that species cannot reach locations with suitable conditions (Leibold et al., 2004; Martiny et al., 2006; Nemergut et al., 2013), rendering communities less similar than expected (Stegen et al., 2013). Alternatively, if dispersal rates are so high that they result in mass effects, species are maintained in local communities simply because dispersal rates outpace species sorting processes (Mouquet and Loreau, 2003; Winegardner et al., 2012), leading to at least partly homogenised communities. Finally, local communities can also be stochastically assembled by neutral or drift processes (Hubbell, 2001; Vellend, 2010). This means that a local community is shaped by random differences in birth, death, immigration and emigration among taxa, and hence is simply a random subsample of the regional species pool. Recent literature reviews on bacterial communities have shown that species sorting is the process that often regulates beta-diversity (i.e. differences in community composition between sites) (Hanson et al., 2012; Lindström and Langenheder, 2012; Nemergut et al., 2013). It is nevertheless clear that the other assembly processes can also be important (Hanson et al., 2012; Lindström and Langenheder, 2012; Nemergut et al., 2013; Stegen et al., 2013; Wang et al., 2013). The relative importance of species sorting, compared to other processes, depends on spatial scale at which spatial processes are linked to differences in environmental heterogeneity and dispersal rates within a metacommunity (Martiny et al., 2011; Östman et al., 2012; Wang et al., 2013; Zinger et al., 2014; Comte et al., 2016), which is defined as a set
of local communities that are connected to each other through dispersal (Leibold et al., 2004).

For example, Östman et al. (2012) found that bacterioplankton composition can be more strongly associated with stochastic processes in homogeneous compared to heterogeneous environments, because homogeneous environmental conditions do not allow species sorting to occur. Similar conclusions were also made by Wang et al. (2013) based on a comparative survey of several within- and across-habitat studies (including soil, sediments, stream biofilm and lake water). Wang et al. (2013) also presented a conceptual model according to which bacterial communities within habitats should be either assembled by species sorting or stochastic processes, depending on the strength of environmental gradients in the system. In addition, dispersal limitation should be more important in isolated systems, whereas increased connectivity and proximity should lead to increased importance of mass effects in metacommunity organization (Wang et al., 2013; Heino et al., 2015a).

Most studies on community assembly processes in bacterial communities have used statistical approaches, such as distance-decay or variation partitioning methods, where spatial variations in taxonomic community are associated to differences in local environmental conditions and spatial distances between sites, the latter being indicative of dispersal processes (e.g., Martiny et al., 2006; Ramette and Tiedje, 2007). Using such ‘associative methods’, many studies have shown that both species sorting and spatial processes explain spatial turnover in bacterial communities; however, they often explain only a low fraction of the differences in community composition among sites (Lindström and Langenheder, 2012). It is currently not clear whether this ‘unexplained variation’ indicates that (a) bacterial communities are to a large extent stochastically assembled, (b) important environmental factors have been missed (Vellend et al., 2014; Heino et al., 2015a), or (c) because they neglect important processes connected to biotic variables, such as the diversity and community composition of other organism groups feeding on bacteria or modifying habitat conditions, which are rarely
measured in the field. In parallel, new statistical frameworks based on null model approaches incorporating taxonomic beta-diversity (the total number of species and their relative abundance) and phylogenetic beta-diversity (species relatedness in a community) have been used to quantify the relative importance of different assembly processes (e.g., Wang et al. 2013, Chase and Myers 2011). Stegen et al. (2013; 2015) developed a null model-based analysis framework to disentangle the quantitative importance of species sorting, drift, dispersal limitation and mass effects (Stegen et al. 2013). Briefly, the first step is based on the assumption that communities that are phylogenetically clustered have similar traits that have between selected in response to variation in abiotic factors over time (Webb et al., 2002). Hence, if the phylogenetic beta-diversity between a pair of communities deviates significantly from a random null model distribution, it suggests that the communities are assembled by species sorting (that is, ‘environmental selection’ by Stegen et al. (2013)). Consequently, pairs of communities that do not deviate must be assembled by other processes (Stegen et al., 2013). These are then, in the second step, dismantled by determining null model deviations of taxonomic beta-diversity instead. Communities that are more similar than expected by chance are assembled by mass effects (that is, ‘homogenizing dispersal’ by Stegen et al. (2013)), those that are less similar than expected by chance by dispersal limitation, and those that do not deviate from null model prediction by drift (Stegen et al., 2013). There are currently few studies that have used both ‘associative’ and ‘null’ approaches, and it is therefore difficult to say whether they provide supportive or conflicting results. More generally, it has also been pointed out that there is a need for studies that compare different analytical methods and biodiversity metrics, such as taxonomic and phylogenetic beta-diversity, which provide complementary information on community assembly processes (Jin et al., 2015).
Even though bacterial metacommunity organization has been studied widely in aquatic systems in recent years, most studies have focused on patterns of bacterioplankton across isolated water bodies, such as lakes, ponds and rock-pools, or have studied biofilms in stream networks (Wang et al., 2012b; Besemer, 2015; Battin et al., 2016). On the contrary, there are only very few studies on benthic biofilm communities within lakes, where different sampling sites are highly connected (but see Bartrons et al., 2012; Vilmi et al., 2016b). Generally, the biodiversity of biofilms is influenced by local abiotic (i.e. physical and chemical) and biotic conditions, and by immigration of cells from the overlying water column (Besemer, 2015; Battin et al., 2016). The latter process is influenced by passive dispersal processes that transport microorganisms to suitable locations and interactions with the local biofilm community that ultimately determine the colonization success of bacteria from the source community (Besemer, 2015; Battin et al., 2016). Stream biofilm communities are in many circumstances assembled by species sorting processes (Besemer, 2015; Peipoch et al., 2015; Battin et al., 2016), but there are also examples showing that hydrological connectivity and directional flow patterns are important (Liu et al., 2013; Freimann et al., 2015) and may potentially lead to dispersal limitation at larger spatial scales (e.g. Lear et al., 2013). On the contrary, mass effects are of limited importance and cannot ‘overrule’ species sorting processes even at the level of highly interconnected sites at small spatial scales where the flux of cells between sites is high (Besemer et al., 2009). The degree to which dispersal processes influence the structure of benthic biofilm communities in lakes is currently unclear. In addition, local environmental conditions and dispersal also influence local biodiversity, which has also primarily been studied for bacterial biofilm communities in streams (Besemer, 2015; Battin et al., 2016; Wang et al., 2016). Here, our main aim was therefore to determine the mechanisms that determine alpha- and beta-diversity of biofilm bacterial communities within a large, highly-connected lake ecosystem. We first used associative approaches based on
taxonomic and phylogenetic alpha- and beta-diversity and a large number of abiotic and biotic
variables. Second, we apply the null model approach to further quantify the relative
importance of different community assembly processes (Stegen et al. (2013)). We
hypothesized that within-habitat environmental gradients (both biotic and abiotic) are strong
enough to cause differences in alpha- as well as in beta-diversity of biofilm bacterial
communities in the lake ecosystem. We further hypothesized that the dispersal of bacteria
among sites is high enough to diminish the effects of dispersal limitation on biofilm assembly,
but may potentially cause mass effects in some cases.

Material and methods

Study area

The study area, Lake Kitkajärvi, is a large (305 km²) lake system located in north-eastern
Finland where some changes in the water quality and land use have recently been reported
(Vilmi et al., 2015). In September 2013, we sampled 36 stony littoral sites for bacteria, algal
biomass, diatoms, macroinvertebrates and water (Fig. 1). The 36 sites were as evenly
distributed as possible across the perimeter of the whole lake system. The spatial
characteristics of the lake system (i.e. the areal extent and high connectivity) enable the
organisms to disperse freely among sites.

Bacterial sampling and laboratory procedures

In the field, 10 cobble-sized stones were randomly collected from the water depth of 40 cm at
each site. To collect biofilm samples, the surface of each stone was brushed for 20s with a
piece of wet foam plastic (4 cm × 4 cm × 4 cm) after which the sample was squeezed into a
sampling jar. The samples were immediately stored cold and frozen within the same day.
In the laboratory, DNA was extracted from freeze-dried sample material using a PowerSoil DNA Isolation Kit (MOBIO, Carlsbad, USA) and the 16S rRNA gene amplified with the bacterial primers 519f and 926trP1 as described in Vilmi et al (2016b) and sequences on an Ion Torrent PGM™ sequencer (Life Technologies, Gaithersburg, USA).

We processed a total number of 404,030 total reads with a mean length of 187 bp mainly using the QIIME pipeline (v1.8) (Caporaso et al., 2010) following previous studies (e.g. Wang et al., 2013). Briefly, the sequences were clustered into OTUs at 97% pairwise identity with the seed-based uclust algorithm (Edgar, 2010). After chimeras were removed via Uchime, representative sequences from each OTU were aligned to the Greengenes imputed core reference alignment V.201308 (DeSantis et al., 2006) using PyNAST (Caporaso et al., 2010). The alignments were then used to construct an approximate maximum-likelihood phylogenetic tree with Jukes-Cantor distance using FastTree (Price et al., 2010) after removing gaps and hypervariable regions using a Lane mask. Taxonomic identity of each representative sequence was determined using the RDP Classifier (Wang et al., 2007) and chloroplast or archaeal sequences were separated out. The lowest sequence depth was 704 and all samples were rarefied to 600 reads for the preparation of the final OTU tables that was used in the alpha- and beta-diversity analyses described below.

**Biotic and abiotic environmental variables**

**Biotic variables.** At each site, algal biomass was estimated as epilithic phytobenthos chlorophyll $a$, which was measured from the surfaces of 10 stones (collected randomly from 40 cm depth) using a BenthoTorch fluorometer (bbe Moldaenke, Cincinnati, USA). Further, diatoms and macroinvertebrates were sampled or surveyed as described in Vilmi et al. (2016b). Diatom samples were brushed from the surfaces of 10 cobble-sized stones from 40 cm depth at each site. In the laboratory, permanent slides were made and approximately 500
Diatom valves were identified to the lowest possible taxonomic level. Macroinvertebrates were sampled using a kick-net with a total kicking effort of 3 min and 6 m at each site and animals were preserved in ethanol. In the laboratory, the animals were sorted and identified to the lowest possible taxonomic level. Further, macroinvertebrates were assigned into different groups based on their feeding habits to separate out biofilm-eating scrapers and their abundance. The following biotic variables were used as predictor variables in the statistical analyses described below: 1) site-specific richness, Shannon’s diversity and Pielou’s evenness for diatom and macroinvertebrate communities, and biofilm-eating scrapers; 2) the first and second axes of separate non-metric multidimensional scaling analysis (NMDS) for diatom and macroinvertebrate communities, and biofilm-eating scrapers. Finally, we also used the relative abundance of the dominant primary producer *Achnanthidium minutissimum* s.l., as well as the abundance of biofilm-eating scrapers, as biotic predictor variables (see Table S1 for a summary).

**Water chemistry.** We performed an extensive sampling campaign for water chemical measurements within two weeks of the sampling of the biotic variables. This time lag was a result of logistic constraints since various chemical variables require immediate analyses in fresh water samples. Samples were taken from a depth of 0.5-1.0 m using a LIMNOS water sampler from a boat at the littoral zone near the bacterial sampling site (i.e. a couple of meters offshore in deeper water to avoid sample contamination by disturbed bottom sediments). Water samples were analyzed within 24 hours of sampling in an accredited laboratory. A total of 35 chemical parameters were analyzed from the site-specific samples (Table S1).

**Physical characteristics.** As physical variables, bottom slope (%) and particle size distribution were measured in the field. Modified Wentworth classes were used to visually assess the coverages of different particle sizes which were mud, fine inorganic sediment (<2 mm), gravel (2-16 mm), pebbles (16-64 mm), cobbles (64-256 mm), boulders (256-1024 mm),
mm), large boulders (>1024 mm) and bedrock. Subsequently, wind fetch describing the openness of a site was calculated according to Rohweder et al. (2008). For descriptive statistics and abbreviations of the physical variables, see Table S1.

**Biodiversity estimators**

To determine beta-diversity, we calculated community dissimilarity with and without phylogenetic information. Dissimilarities based on relative abundance data were chosen because they give more weight to dominant OTUs and reduce chance effects that may be involved in the detection of rare OTUs, which may decrease the overall degree of explained variation when presence-absence data are used (Souffreau et al. 2015). Taxonomic turnover was determined using Bray-Curtis dissimilarities based on relative abundances of OTUs between a given pair of samples. To determine phylogenetic turnover, we used the mean nearest taxon distance index ($\beta$MNTD) (Fine and Kembel, 2011; Stegen et al., 2012).

$\beta$MNTD is the mean phylogenetic distance to the closest relative in a paired community for all taxa (Fine and Kembel, 2011) and is sensitive to the changes of lineages close to the phylogenetic tips. Weighted $\beta$MNTD based on relative abundance data was calculated according to Stegen et al. (2013).

Bacterial alpha diversity was quantified using species richness and Faith’s phylogenetic diversity (PD) (Faith and Baker, 2006).

**Statistical analyses**

**Beta-diversity**

To investigate the underlying mechanisms determining beta-diversity in bacterial communities, we used two different approaches. First, we used multiple regression on matrices (MRM) to tease apart the relative importance of sets of variables related to spatial (SPA), abi-
otic (ABIO, i.e. chemical and physical) and biotic variables (BIO) on bacterial community similarity. Second, we performed null model analyses according to Stegen et al. (2013) to quantify the relative importance of environmental selection (species sorting), homogenizing dispersal (mass effects), drift and dispersal limitation.

Multiple regression on matrices. To tease apart the relative importance of the three components (SPA, ABIO and BIO) as well as individual variables on taxonomic and phylogenetic community similarity, we further used the MRM approach (Legendre et al., 1994) with z-score transformed Euclidean distance matrices of the predictor variables (Euclidean distance matrices), as suggested by Martiny et al. (2011). We only included abiotic and biotic variables that were highly correlated (Pearson r > 0.8) with community similarities using the “bioenv” function of the “vegan” packages in R (Table S2, (Clarke and Ainsworth, 1993) (Oksanen et al., 2016)). To reduce the effect of spurious relationships between variables, the MRM model was run twice. After the first run, we removed non-significant variables, and we reported the results from the second run only (Martiny et al., 2011). To identify the importance of individual abiotic and biotic variables to the overall correlations, we calculated their partial regression coefficients. Partial regression coefficients provide information about the degree of change in community similarity per standardized unit of similarity for the variable of interest, while all other variables are constant, and thereby identify the variables that make the strongest independent contribution to changes in community composition. The MRM analysis was performed using the R package ecodist v1.2.9 (Goslee and Urban, 2007). Finally, we calculated distance-decay relationships for both taxonomic and phylogenetic beta-diversity to compare their degree of community turnover with increasing spatial distance across sites (Martiny et al. 2006).

Null model analysis. In the first step, we calculated standardized effect size of βMNTD, which measures (in units of SDs) how much observed βMNTD deviated from the mean of
null distribution (999 null iterations) based on random shuffling of OTU labels across the tips of the phylogeny (Hardy, 2008; Fine and Kembel, 2011; Stegen et al., 2012). This randomization keeps the observed species richness, species occupancy and species turnover constant. We used a significance cut-off of < -2 or > 2, respectively to determine the proportion of community pairs that is phylogenetically more or less similar than expected by chance, respectively. Both cases indicate that environmental selection determines observed compositional differences between samples (Stegen et al., 2013; Dini-Andreote et al., 2015). For all cases where \( \beta \)MNTD did not deviate significantly from the null model distribution (i.e. communities that were not assembled by environmental selection), we calculated the Raup-Crick beta-diversity metric for each pair of local communities after a total of 1,000 iterations (Chase et al., 2011), but based on species relative abundances (RC\text{bray}) as in Stegen et al. (2013). Observed RC\text{bray} values were compared with those of a random null model distribution according to Chase et al. (2011) and then we followed the procedure described in detail in Stegen et al. (2013) to disentangle the importance of drift, dispersal limitation and mass effects: RC\text{bray} values between -0.95 and +0.95 indicate drift, RC\text{bray} values > +0.95 indicate that communities are less similar than expected by chance as a result of dispersal limitation, and RC\text{bray} values < -0.95 indicate that communities are more similar than expected by chance as a result of mass effects.

**Alpha-diversity**

Variation in alpha diversity was partitioned between the three components (SPA, ABIO and BIO) using linear models (Borcard et al., 1992; Anderson and Gribble, 1998). By generating models with the three sets of explanatory variables, we estimated the proportions of variation in bacterial diversity explained by the pure effects of SPA, ABIO and BIO, and by the intersections of these three components. For spatial variables, principal coordinates of
neighborhood matrices (PCNM; Borcard and Legendre, 2002; Borcard et al., 2004) were used to represent original spatial distance matrices as sets of orthogonal eigenvectors. The first PCNM eigenvector represents the broadest spatial gradient, while each successive eigenvector represents a finer spatial scale. A set of PCNM eigenvectors for each analysis was determined by selecting only positive eigenvectors, which were significant (alpha = 0.05) explanatory variables in a distance-based redundancy analysis (Fig. S1). The multiple regression analysis showed that abiotic factors had a relatively stronger effect on both the taxonomic (Bray-Curtis similarities) and phylogenetic (βMNTD) turnover than spatial distance (which had no significant effect) or biotic parameters, where the partial regression coefficients were lower and only marginally significant (Table 1). Moreover, when MRM was run to tease apart the relative importance of individual environmental variables, the partial regression coefficients of algal biomass were significant in the case of taxonomic turnover, whereas no single biotic variable had significant partial regression coefficients in the case of phylogenetic turnover. Among abiotic variables, significant partial regression coefficients were found for NO$_x$, alkalinity and NH$_4$ in the case of Bray-Curtis similarities and NO$_x$ in case of βMNTD. Generally, however, the explanatory power of the MRM was low ($R^2$ values < 0.25 in all cases), so that the largest fraction in differences in community composition remained unexplained.

The null model-based approach showed that the majority of pairs of communities were assembled by drift (56% of all pairwise comparisons) whereas 14% were assembled by
environmental selection, 24 % by dispersal limitation and 6 % by homogenizing dispersal (Fig. 2).

**Alpha diversity.** At the local scale, species richness and PD at the local scale were significantly correlated to each other ($r^2 = 0.544$, $p < 0.05$, Fig. 3). Generally, a lower fraction of variation in bacterial diversity across sampling sites could be explained by abiotic (7%) compared to 18%) and spatial variables (16% compared to 26%) for richness than for PD, respectively. Larger proportion of variation in local diversity were explained by spatial variables than by local environmental conditions, whereas mainly smaller spatial scale variables were significant for species richness and various spatial scale variables for PD. We also found that abiotic variables accounted for significant fractions of variation in both alpha diversity metrics, while the abiotic factors explaining variation in species richness and PD were different (Fig. S2). Richness was negatively correlated to suspended solids and nitrogen, whereas PD was positively correlated to alkalinity, but negatively correlated to aluminium concentrations and Fetch (Fig. S2). Moreover, for PD, we found a significant, albeit minor, effect of biotic variables, as well as considerable co-variation between abiotic, spatial and biotic variables (Fig. 3). Of the biotic variables, PD was positively related to algal biomass and macroinvertebrate richness (Fig. S2).

**Discussion**

This study shows that within-habitat environmental gradients in one large, highly-connected lake ecosystem were strong enough to cause differences in alpha- and beta-diversity of biofilm bacterial communities. Further, we show that abiotic conditions explained more of the taxonomic and phylogenetic turnover in community composition compared to biotic
variables, and that drift, species sorting and dispersal processes contribute to differences in the composition of bacterial biofilm communities between sites.

Even though it has been shown that within-lake beta-diversity of bacterial communities is lower compared to that between lakes (Yannarell and Triplett, 2004; Wang et al., 2013), it has become clear that there are significant spatial signals for the bacterial communities in sediments (Wang et al., 2013), biofilm (Vilmi et al., 2016b) and lake water (Jones et al., 2012; Lear et al., 2014) within lakes. Data from global surveys in marine systems has also shown that there is stronger community turnover in habitats with stronger environmental gradients (e.g., sediment vs. plankton and coastal vs. open ocean habitats) (Zinger et al., 2011; Zinger et al., 2014). For inland waters, previous studies attributed these spatial patterns in communities to both species sorting and dispersal effects (Wang et al., 2013; Lear et al., 2014; Vilmi et al., 2016b), which was also supported by the results of our study for bacterial biofilms within a lake system. The null model approach showed, moreover, that drift was the predominant assembly process, which fits with conceptual ideas that stochastic assembly should prevail within lakes where environmental gradients should be relatively weak (Wang et al. (2013) and because they are relatively homogenous, well-mixed system where biofilm bacteria are recruited from the water (Besemer, 2015; Battin et al., 2016). At the same time, we might have even underestimated stochastic processes, since they might be masked by indirect species sorting processes by conditions in the water, i.e., species sorting that acted on planktonic bacteria, which then randomly colonized the biofilms. Hence, the variations in biofilm composition reflect the differences in the composition of plankton source communities, which have shown to vary in composition within systems at relatively small spatial scales (Lear et al., 2014). However, as biofilms form at the interface between the substrate (in our case stones) and the water, and are dependent on inorganic nutrients and
organic matter from the water to support their growth, it seems unlikely that species sorting
effects were indirect. Hence, it has been shown that biofilm assembly is a selective process
and not just the results of random dispersal from the surrounding water (Besemer et al., 2012).

This finding that almost 25% of all pairwise assemblages are assembled by dispersal
limitation is puzzling, in particular because the MRM showed that spatial variables were not
significantly related to community similarity. One possible explanation for these deviating
results is that the null model analysis overestimates the effect of dispersal limitation. Stegen et
al. (2013) define that pairs of communities are assembled by dispersal limitation if $R_{C_{bray}}$ is
close to 1, indicating that communities are less similar than expected by chance. This can,
however, also be the result of strong biotic forces that create very different communities at
adjacent sites (Chase et al., 2011). Here, we found that biotic factors had stronger effects on
beta-diversity in the case of taxonomic compared to phylogenetic beta-diversity, and since
Stegen et al.’s (2013) approach only used the latter in the identification of species sorting
processes, it seems possible that their definition of dispersal limitation to some extent masks
the effects of biotic sorting. Another possibility is that the spatial distance matrix that we used
in the MRM analyses does not depict the hydrodynamics of our study lake, therefore resulting
in non-significant spatial effects. Collectively, the results from different statistical analyses are
contradictory, but we cannot currently rule out that dispersal limitation can produce spatial
differences in biofilm composition even within a highly-connected lake ecosystem. To fully
understand how dispersal influences biofilm community assembly, future studies should
therefore utilize new statistical approaches that disentangle effects of directional dispersal
through water masses and non-directional processes (e.g. aggregation) at various spatial
scales (Bertolo et al., 2012) and integrate direct measurements of water flow rates and
directions. In addition to deviating results regarding the importance of dispersal limitation by
the MRM and null model approaches, we also found that the results of MRM model differed
between taxonomic and phylogenetic beta-diversity. Generally, both biotic and abiotic factors
explained less of the similarities in phylogenetic compared to taxonomic beta-diversity. This
is, for example, in contrast to a study that compared taxonomic and phylogenetic beta-
diversity in a vertical soil gradient, where species sorting was more important in case of the
latter (Hu et al., 2015). These deviating results might indicate that the environmental gradients
in the highly-connected lake system were not strong enough to select for traits that are
phylogenetically conserved, and that community turnover along environmental gradients is
therefore better captured by taxonomic diversity metrics.

Another important finding from our study is that biotic variables were less important
than abiotic conditions in structuring biofilm bacterial communities. The most significant
biotic factor was algal biomass, which suggests that overall availability of organic substrates
released from periphyton influenced bacterial community composition (Wagner et al., 2015;
Battin et al., 2016). Algal biomass was significantly correlated to taxonomic but not to
phylogenetic turnover, which shows that OTUs irrespective of their phylogenetic affiliation
responded to changes in algal biomass. This might reflect the fact that the organic substrates
released by periphyton are often highly available and readily used by a wide range of different
bacteria (Wagner et al., 2014), and, hence, are not phylogenetically conserved (Martiny et al.,
2013). Effects of grazers were generally weak and not significant. For diatom communities in
stream biofilms, it has also been shown that abiotic parameters are relatively more important
than biotic variables (Göthe et al., 2013). However, the importance of grazers in structuring
diatom assemblages was different when the analyses were done for guilds that differed in
traits, such as growth form or body size (Göthe et al., 2013). This suggests that the trait
composition of the response community also affects the influence of grazers as a structuring
force of the spatial turnover of the grazed communities. Moreover, it might also be possible
that the effects of selective grazers, such as protozoa, would have shown a similar picture in
our study. Among abiotic variables, nitrite/nitrate concentrations had the strongest
independent contribution to the changes in taxonomic and phylogenetic community similarity,
confirming previous studies that have shown that nitrogen concentrations are an important
structuring factor for biofilm bacterial communities (Kelly et al., 2014; Zeglin, 2015). One
methodological caveat of our study was that there was a gap between the sampling for biotic
variables (including biofilm composition) and abiotic variables, which were sampled two
weeks later. The fact that abiotic factors nevertheless could explain differences in biofilm
composition suggests that the time of the sampling was not a major problem. However, it is
possible that we actually underestimated the contribution of abiotic factors, which would
further strengthen our conclusion because they were more important in structuring biofilm
communities compared to biotic factors.

Interestingly, alpha diversity were also affected by similar environmental factors
which were most important for determining beta-diversity. However, different factors
correlated with the variation in taxonomic richness and PD across sampling sites, and the
relationship between species richness and PD, $r^2 = 0.544$, was relatively weak compared to
other studies (e.g. Wang et al., 2012a). This is in congruence with unimodal productivity-
diversity relationships, which have also been found in bacteria (Song et al., 2016). PD, on the
contrary, was positively related to Mg concentration, algal biomass and macroinvertebrate
richness. This shows that PD is higher in more benign alkaline environments and that
interactions with other trophic groups may promote the co-existence of a phylogenetically
diverse bacterial biofilm community as well. Moreover, high algal biomass and
macroinvertebrate richness may reflect thicker and more mature biofilm that provide more
physical niches and resources. This idea is supported by the finding that fetch, reflecting effects of physical disturbances by wave action, was negatively related to PD. Moreover, Al concentrations were negatively related to PD, which many generally reflect negative effects of metal pollution on biofilm richness (Bier et al., 2015). Taken together, the results suggests that adaptation to environmental pollution and resistance towards strong physical disturbances require traits that might be phylogenetically conserved (Martiny et al., 2015) and therefore decreases PD. Opposite to what we found for beta-diversity, spatial variables had a stronger effect than environmental variables on differences in alpha diversity between sites. This may indicate that the total number of species found in a locality can be influenced by dispersal (including rates and pathways) that transports taxa to a location (Besemer et al., 2013; Zha et al., 2016), but that most taxa remain inactive and contribute primarily to seed-bank at local sites and only occasionally to the active community (Shade et al., 2014). We are, however, aware of the fact that the sequencing depth in our study might have been too low to obtain robust estimates of alpha-diversity (Lundin et al., 2012), so these results need to be interpreted with care.

To summarize, there is now an increasing number of studies that show clear turnover of bacterial communities at relatively small spatial scales within aquatic ecosystems. Such turnover can be attributed to a combination of different assembly processes, including stochasticity, species sorting and potentially even dispersal limitation. More studies are, however, still needed to integrate the importance of different community assembly processes at different spatial scales depending on connectivity patterns and dispersal pathways between and within different types of inland water bodies.

Funding
This study was funded by the European Regional Development Fund through the Kitka-MuHa project. Annika Vilmi was supported by the Maj and Tor Nessling Foundation, Jianjun Wang was supported by the Emil Aaltonen Foundation and Silke Langenheder by funding from the Swedish Research Council.

Acknowledgements

We thank Yrjö Kivinen, Susan Kunnas, Minna Kuoppala, Sari Latvala, Marja Lindholm, Osmo Murtovaara and Juha Riihimäki for sampling and data collection as well as Marko Suokas from Biocenter Oulu Sequencing Center.

References


*Bioinformatics* **26**: 2460-2461.


Table 1. Multiple regression analysis on distance matrices for taxonomic (Bray-Curtis) and phylogenetic beta diversity metrics ($\beta$MNTD). Overall $R^2$ values as well as partial regression coefficients for sets of abiotic (ABIO), biotic (BIO) and spatial (SPA) variables selected by the bioenv analysis, as well as individual ABIO and BIO variables are shown. Note that only variables with significant partial regression coefficient of at least one of the beta-diversity metrics ($p < 0.05$) are listed. Partial regression coefficients and p-values for all variables are shown in Table S4.

<table>
<thead>
<tr>
<th>Sets of variables</th>
<th>Bray- Curtis</th>
<th>p-value</th>
<th>$\beta$MNTD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.217</td>
<td>0.000</td>
<td>0.189</td>
<td>0.011</td>
</tr>
<tr>
<td>ABIO</td>
<td>0.025</td>
<td>0.001</td>
<td>0.012</td>
<td>0.021</td>
</tr>
<tr>
<td>BIO</td>
<td>0.011</td>
<td>0.044</td>
<td>0.005</td>
<td>0.047</td>
</tr>
<tr>
<td>SPA</td>
<td>0.002</td>
<td>0.699</td>
<td>-0.001</td>
<td>0.643</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual variables</th>
<th>Bray- Curtis</th>
<th>p-value</th>
<th>$\beta$MNTD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.241</td>
<td>0.000</td>
<td>0.217</td>
<td>0.039</td>
</tr>
<tr>
<td>Alkalinity</td>
<td>0.026</td>
<td>0.000</td>
<td>0.012</td>
<td>0.028</td>
</tr>
<tr>
<td>$\text{NH}_4^+$</td>
<td>-0.028</td>
<td>0.010</td>
<td>-0.001</td>
<td>0.643</td>
</tr>
<tr>
<td>$\text{NO}_x$-N</td>
<td>0.043</td>
<td>0.000</td>
<td>0.012</td>
<td>0.028</td>
</tr>
<tr>
<td>Particle mean size</td>
<td>0.10</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algal biomass</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure legends

Figure 1. The study area with 36 littoral sampling sites.

Figure 2. Proportions of community pairs assembled by drift, species sorting (Selection), dispersal limitation (Disp. Lim) and mass effects or homogenizing dispersal (Hom. Disp.).

Figure 3. (A) Relationship between species richness and phylogenetic diversity. (B) and (C) Variation in bacterial species richness (B) and phylogenetic diversity (C) related to spatial distance (SPA), abiotic (ABIO), and biological variables (BIO). Inorganic SS: inorganic suspended solids (µg L⁻¹), NOx: NO₂+NO₃-N (µg L⁻¹), Mg: Mg concentration (mg L⁻¹), Al: Aluminum concentration (mg L⁻¹), Macro.comm.richness: Species richness of macroinvertebrates. PCNM eigenvectors were determined as described in the material and methods.
Figure 1
Figure 2

Process Estimates

Proportion (%)

Selection: 14.44
Disp. Lim.: 24.13
Hom. Disp.: 5.71
Drift: 55.71
Figure 3

A. Relationship between Richness and PD

B. Variation partitioning: Richness

C. Variation partitioning: PD

- **ABIO**: Inorganic SS, NO₃, p = 0.049
- **BIO**: ns
- **SPA**: PCNM 16, 14, 4; p = 0.018
- **ABIO**: Mg, Fetch, Al; p = 0.004
- **BIO**: Algal Biomass, macro comm. richness; p = 0.025
- **SPA**: PCNM 4, 16, 14, 3, 1, 2; p = 0.004