Freshwater Biology

*Freshwater Biology* (2015) **60,** 973–988 doi:10.1111/fwb.12556

# Elements of metacommunity structure and community-environment relationships in stream organisms

JANI HEINO*, TIINA NOKELA*, JANNE SOININEN†, MIKKO TOLKKINEN*, LAURA VIRTANEN‡ AND RISTO VIRTANEN§

*Finnish Environment Institute, University of Oulu, Oulu, Finland

†Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

‡Department of Environmental Sciences, University of Helsinki, Helsinki, Finland

§Department of Biology, University of Oulu, Oulu, Finland

## SUMMARY

1. Most metacommunity studies aim to explain variation in community structure using environmental and spatial variables. An alternative is to examine patterns emerging at the level of an entire metacommunity, whereby six models of metacommunity structure (i.e. random, chequerboards, nestedness, evenly spaced, Gleasonian gradients and Clementsian gradients) can be examined.
2. We aimed to test the fit of six competing models of metacommunity structure to extensive survey data on diatoms, bacteria, bryophytes and invertebrates from three drainage basins in Finland, along a latitudinal gradient from 66°N to 70°N.
3. Species were mainly distributed independently of one another (following the Gleasonian model) in the southernmost drainage basin (66°N), whereas there were discrete community types, with sets of species responding similarly along environmental gradients (following the Clementsian model), in the northernmost drainage basin (70°N). The patterns found were not directly related to an expected relationships between environmental heterogeneity and metacommunity structures, but rather to the geographical location of the drainage basin.
4. There is evidently among-region variation in the best-fit models of metacommunity structure of stream organisms. These metacommunity patterns may show some similarities among biologically disparate organismal groups sampled at the set of the same sites, although the underlying environmental drivers of those patterns may vary between the groups.

*Keywords*: bacteria, biological communities, bryophytes, diatoms, fresh waters, invertebrates

## Introduction

Metacommunity ecology is a recently emerged subdiscipline of ecology that aims to understand the structuring of local communities in their regional context. A metacommunity can thus be defined as a set of local communities potentially connected by the dispersal of species (Leibold *et al.*, 2004). While many previous metacommunity studies have taken a mechanistic perspective, trying to disentangle the roles of dispersal, spatial factors and local environmental conditions in determining local community structure (Cottenie, 2005; Meynard *et al.*, 2013; Heino *et al.*, 2015), an alternative, pattern-based approach is to examine the so-called elements of metacommunity structure (Leibold & Mikkelson, 2002). This approach tries to associate empirical patterns of variation in communities with randomness or to one of five main idealised metacommunity structures (Fig. 1): chequerboards, nestedness, evenly spaced gradients, Gleasonian gradients and Clementsian gradients (Leibold & Mikkelson, 2002; Presley, Higgins & Willig, 2010). Relatively few studies have examined the best fit of these idealised models with empirical data (Presley *et al.*, 2009; Keith *et al.*, 2011; Henriques-Silva, Lindo & Peres-Neto, 2013; Dallas & Presley, 2014). In contrast, a large number of studies have examined different patterns in

Correspondence: Jani Heino, Finnish Environment Institute, Paavo Havaksen Tie 3, FI-90014 Oulu, Finland. E-mail: jani.heino@environment.fi

 

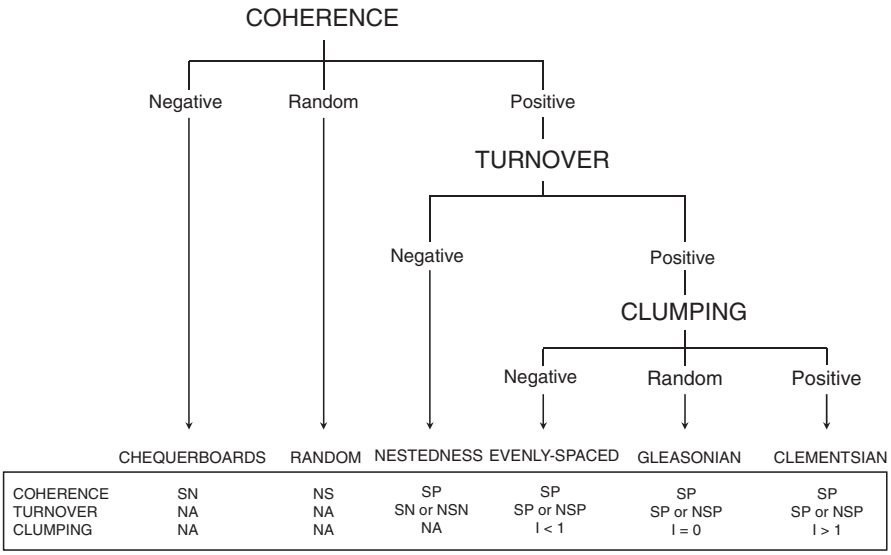


**Fig. 1** A chart describing Leibold & Mikkelson's (2002) elements of metacommunity structure (i.e. coherence, turnover and boundary clumping), resulting in six main metacommunity structures (i.e. chequerboard, random, nested, evenly spaced, Gleasonian and Clementsian). Abbreviations: SN, significantly negative; NS, non-significant; NA, not assessed; SP, significantly positive; NSN, non-significant negative; NSP, non-significant positive; I, Morisita's index value. See test for details. Figure modified from (Presley & Willig, 2010).

isolation, including chequerboards (e.g. Gotelli & McCabe, 2002), nestedness (e.g. Patterson & Atmar, 1986) and different gradient models (e.g. Hoagland & Collins, 1997).

Chequerboards have received considerable interest and spurred much controversy ever since the original suggestion by Diamond (1975) that pairs of species co-occur less than expected by chance (e.g. Stone & Roberts, 1990). Comprehensive meta-analyses and various case studies have indeed shown that some species often occur together less than expected by chance, both in macro- and microorganisms (e.g. Horner-Devine *et al.*, 2007). This pattern may be attributable to various mechanisms that often act in concert, such as interspecific competition, different habitat preferences between species and the roles of historical factors (Gotelli & McCabe, 2002; Heino, 2013). As various mechanisms may affect species distributions simultaneously, the fit of empirical data with the chequerboard pattern is often far from perfect. Thus, other idealised patterns may also fit to the same empirical dataset (e.g. Heino & Soininen, 2005).

Another idealised pattern, which has also received considerable interest, is nestedness (Patterson & Atmar, 1986). Perfect nestedness occurs when species poor communities consist of proper subsets of species in richer communities and when the ranges of widely distributed species encompass those of species with progressively less wide distributions across a set of sites (Patterson & Atmar, 1986). In general, habitat area (as a proxy of the probability of extinction of species), isolation (as a proxy for colonisation) and environmental suitability (as a proxy for species specialisation) are typical correlates of nestedness (Wright *et al.*, 1998). These ecological factors may be potential correlates of nestedness, especially if they vary in a nested fashion and species show nested responses to these factors (Hylander *et al.*, 2005). Nested patterns could be expected to prevail in regions with low environmental heterogeneity because, in such regions, turnover in species composition among sites is lower and some widespread species are able to occur in most, if not all, sites (Heino, 2011). Similarly to chequerboards, the fit of empirical data with nestedness is also typically far from perfect, leaving room for other idealised models to fit empirical data.

Different gradient models have also intrigued ecologists for a long time, originally in the context of terrestrial vegetation. Clements (1916) proposed that plant communities consist of tight associations between species, behaving like a 'superorganism', leading to discrete community boundaries (i.e. 'Clementsian gradients'). In contrast, plant communities could be mere collections of species, with individualistic responses to underlying environmental gradients (i.e. 'Gleasonian gradients'), and whose ranges happen to overlap (Gleason, 1926). Evenly spaced gradients are yet another alternative idealised model, where species range boundaries are hyperdispersed (i.e. far apart) along the underlying environmental gradient (Tilman, 1982). Evenly spaced gradients indicate maximal differences among species in environmental tolerances (Presley & Willig, 2010). Fits with gradient models should be evident in regions with high environmental heterogeneity, because environmental heterogeneity is likely to increase turnover in species composition among sites (Heino, 2011). While the debate over community gradients has been ongoing for almost 100 years, most attention has been devoted to terrestrial

plants (McIntosh, 1995), and there is thus an open field for testing the fit of these ideas in other organismal groups and ecological systems (Presley *et al.*, 2010). In river ecology, there has been a similar and long-standing debate over discrete (i.e. river zonation; Illies & Botosaneanu, 1963) and continuous (i.e. the river continuum concept; Vannote *et al.*, 1980) boundaries between biological communities along the gradient from headwaters to the lower reaches, although this has not directly addressed the six metacommunity structures examined in this study.

It is important to contrast the six different idealised models simultaneously, because such a combined approach may tell us more about the structuring of a metacommunity than if randomness, chequerboards, nestedness and the different gradient models are studied in isolation (Leibold & Mikkelson, 2002; Presley *et al.*, 2011; Meynard *et al.*, 2013; Dallas & Presley, 2014). However, the idealised models of metacommunity structure have been studied only rarely in freshwater systems (Heino & Soininen, 2005; Heino, 2009) and, to our knowledge, only three freshwater studies (Henriques-Silva *et al.*, 2013; Erős *et al.*, 2014; Fernandes *et al.*, 2014) have used the methods of 'the elements of metacommunity structure' approach (Leibold & Mikkelson, 2002). There is thus scope to examine freshwater metacommunities using the same methods as adopted for various terrestrial taxa (Leibold & Mikkelson, 2002), including rodents, bats and birds (Presley *et al.*, 2012), snails (Presley *et al.*, 2011) and plants (Meynard *et al.*, 2013). Comparing the findings of metacommunity structuring among fresh water, marine and terrestrial taxa would obviously lead to a greater understanding and generalisation in community ecology. Freshwater systems differ in many respects from terrestrial and marine systems by being more isolated and fragmented than most systems in the other two realms of life. The typically isolated and fragmented nature of freshwater systems may also mean that metacommunity patterns and underlying mechanisms are unusual, differing from those detected in most terrestrial or marine systems.

We define a metacommunity for stream organisms as a set of stream sites within a drainage basin (Peres-Neto, 2004; Heino, 2013) and examined the elements of metacommunity structure for bacteria, diatoms, bryophytes and invertebrates. We asked whether such disparate organismal groups show similar metacommunity structures along environmental gradients. Further, we examined these patterns in three drainage basins to see whether metacommunity structures in the four organismal groups depend on geography. We expected differences among the three drainage basins in metacommunity patterns and underlying processes, because they differ in environmental heterogeneity and harshness (Grönroos *et al.*, 2013; Heino, 2013).

## Methods

The test datasets come from three northern Finnish drainage basins (Figure S1), where detailed surveys of stream organisms have been conducted in recent years (Heino *et al.*, 2012; Schmera, Erős & Heino, 2013). Largely the same dataset was used in examining community–environment and community–spatial location relationships of stream diatoms, bryophytes and invertebrates (Heino *et al.*, 2012; Grönroos *et al.*, 2013). These previous studies found clear patterns in community composition along environmental gradients and virtually no spatial autocorrelation of community composition within each drainage basin, suggesting that environmental factors override any effects of spatial distance between sites (Heino *et al.*, 2012) or of connectivity (Grönroos *et al.*, 2013). Of these drainage basins, the easternmost Koutajoki basin shows the highest environmental heterogeneity, followed by the southernmost Iijoki basins and the northernmost Tenojoki basin (Grönroos *et al.*, 2013; Heino, 2013). Environmental heterogeneity was quantified and tested previously between the basins using the multivariate method called Permutational Analysis of Multivariate Dispersions (Anderson, Ellingsen & McArdle, 2006).

### *Study areas and sampling periods*

The study areas and field methods have been described previously (Heino, 2013; Schmera, Erős & Heino, 2013) and are outlined here only to aid understanding of the geographical and ecological context of the three drainage basins. Diatoms, bacteria, bryophytes and invertebrates were sampled at the same sites in each drainage basin. Diatoms and invertebrates were sampled in all three basins, while bryophytes were sampled in the Iijoki and Koutajoki basins and bacteria only in the Tenojoki basin.

The first study area was in the Iijoki drainage basin (centred on 65°N, 27°E), characterised by boreal coniferous forests and peatlands. The streams are generally slightly acidic, and nutrients range from low to moderate (Heino *et al.*, 2012; Heino, 2013). A total of 20 first- to third-order stream sites were surveyed in late May 2009, except for the bryophytes, which were sampled in July–August 2006 and 2007. The selected streams were close to a natural state. Hydrological connectivity between

sites is high due to their permanent nature and the high density of streams in the area.

The second study area is in the Koutajoki drainage basin in north-eastern Finland (centred on 66°N, 29°E). Headwater streams in the drainage basin are characterised by circumneutral to alkaline water, low-to-high concentrations of humic substances and low-to-moderate nutrient concentrations (Heino *et al*., 2012; Heino, 2013). A total of 20 first- to third-order stream sites were sampled in the Koutajoki drainage basin in late May 2008. The streams sampled were nearly pristine because they were located either in protected areas or close to those areas (Heino *et al*., 2009). Hydrological connectivity between stream sites is again high owing to their permanent nature and the high density of streams in the area.

The third study area is in the Tenojoki drainage basin (centred on 70°N, 27°E). This largely undisturbed area is characterised by an arctic–alpine vegetation, comprising mountain birch woodlands at low altitude and barren tundra at higher altitude. Streams are close to pristine and are ultraoligotrophic with circumneutral water (Heino, 2013; Schmera *et al*., 2013). A total of 30 stream sites were sampled in this drainage basin in early June 2010.

Stream organisms were sampled in different drainage basins in different years since we could not sample all sites within a short period of time in a single year. We considered it much more important to sample the sites in the same season (i.e. soon after the snowmelt in the spring) than in the same year. If the sites are not sampled within a short period of time in the same season, the results are likely to mirror seasonal rather than true, among-site, differences in stream communities in a given drainage basin (Heino, 2014). This reasoning is particularly true for organisms, such as diatoms and invertebrates, showing clear seasonal succession (Korhonen, Köngäs & Soininen, 2013). Surveys of several groups of organisms at such a large spatial extent require more resources both in the field and the laboratory and, hence, repeating sampling on several occasions at each sites would have been impossible in practice.

### *Environmental variables*

Several riparian, in-stream and water chemistry variables were measured at each stream site (Heino *et al*., 2012; Heino, 2013). For this study, three physical habitat variables and three water chemistry variables were chosen *a priori* because they often account for significant variation in the community structure of stream organisms (Malmqvist & Mäki, 1994; Heino *et al*., 2012). In each stream, (i) current velocity (at 0.6× depth) and (ii) depth were measured at 30 random locations along cross-stream transects, the number of which depended on stream width (i.e. a larger number of transects in small than large streams). The wetted width (iii) of each stream was measured based on five cross-stream transects, and this variable was a proxy for stream size. Mean values of the physical habitat measurements were used in the statistical analyses. Water samples were collected simultaneously with the field sampling, and they were analysed for (iv) pH, (v) conductivity and (vi) total phosphorus using Finnish national standards (National Board of Waters & the Environment, 1981). Because the concentrations of total phosphorus (TP) are typically below detection limits (TP < 5 μg $L^{-1}$) in the streams of the Tenojoki drainage basin, total phosphorus was not measured there (Schmera *et al*., 2013).

### *Sampling and processing of diatoms*

At each site, 10 randomly picked fist-sized stones were collected randomly from depths varying from 10 to 30 cm, biofilm containing diatoms was brushed from a pre-defined surface area of each stone (5 cm × 5 cm) and a pooled sample was thus obtained for each stream site (Heino *et al*., 2012). Diatoms were cleaned from organic material in the laboratory using wet combustion with acid ($HNO_3$:$H_2SO_4$; 2:1) and mounted in Naphrax. Approximately 500 frustules per sample were counted and identified to species using phase contrast light microscopy. In the Tenojoki drainage basin, a few samples had few frustules and such samples were omitted from the analysis.

### *Sampling and processing of bacteria*

At each stream site, 10 fist-sized stones (separately from the stones from which diatoms were sampled) were selected randomly from depths varying from 10 to 30 cm. For bacterial samples, biofilm was brushed from a predefined area (5 cm × 5 cm) of the stones surface using a piece of sterile plastic foam and disposable gloves. The resulting material and associated water were stirred in a small jar from each particular site (i.e. material from the 10 stones was pooled) and were frozen soon after sampling until the analyses in the laboratory. The samples were taken in the field with great care using disposable plastic gloves, to avoid cross-site contamination, and all analyses were undertaken by the same laboratory technician in a certified laboratory. Because bacteria are rarely included in this type of study, we provide the details of the laboratory analyses below.

DNA was first extracted from 0.25 g freeze-dried sample material using a PowerSoil DNA Isolation Kit (MOBIO Laboratories), which yielded the greatest amount of DNA in preliminary comparative trials with other kits. The 16S rDNA region of bacteria was amplified with primers 519F 5′-CAGCMGCCGCGGTAANWC-3′ and 926trP1 5′-CCTCTCTATGGGCAGTCGGTGATCCGT CAATTCCTTTRAGTTT-3′. We tested our primers using the SILVA database (Quast *et al.*, 2013). The maximum number of allowed mismatches was one, and the length of the 0-mismatch zone at the 3′-end was two bases. The PCR program we used was as follows: Stage 1, performed once (step 1: 98 °C 3:00 min); Stage 2, repeated 25 times: (Step 1: 98 °C 0, 10 min, Step 2: 64 °C 0:30 min, Step 3: 72 °C 0:20 min); and Stage 3, performed once (Step 1: 72 °C 5:00 min, Step 2: 4 °C ∞). We used different a-primer barcodes for each sample. Three separate PCRs were performed for each bacterial sample. Total volumes of one PCR were 20 μL containing 4 μL GC-buffer, 0.4 μl dNTP, 0.2 μL Phusion Hot Start II High-Fidelity DNA Polymerase, 12.4 μL sterile water (DNA Grade, DNASE, Protease free), 0.5 μL A-primer, 0.5 μL B-primer and 2 μL bacterial sample (for the concentrations used in PCRs, see Table S1 in Supporting Information). After the PCR triplicates were combined and PCR products had been cleaned with SPRIselect (Beckman Coulter), multiNA (Microchip Electrophoresis System for DNA/RNA)-analyses were performed for each cleaned PCR product. Bacterial operational taxonomic units (OTUs) were determined using ion torrent semiconductor sequencing. We used an Ion Torrent™ Personal Genome Machine® (PGM) System with 400-base read length chemistry. Silva database TestPrime results showed that our primers had 94% coverage of all bacterial taxa (for information about the primers used, see Table S1). PCRs were performed by Veriti® Thermal Cycler (Life technologies). The amplicons were sequenced using the ion torrent semiconductor system (Life Technologies).

Sequences were analysed using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline (Caporaso *et al.*, 2010), where the total number of raw sequences was 558 305. The sequence library was split by samples and quality filtered based on the quality scores for each sequence. Sequences with quality scores below 25 were removed. Sequences shorter than 200 bp or longer than 1000 bp, with more than two mismatches in the primer, ambiguity or maximum homopolymer run exceeding six were also removed. After quality control, a total of 238 438 sequences were retained, with an average sequence length of 343 bp. The sequences were clustered as OTUs using the Uclust algorithm (Edgar, 2010) with 97% sequence similarities. Possible chimeras were identified using Chimera Slayer (Haas *et al.*, 2011) and removed from downstream analysis. The taxonomic assignment of each representative sequence was conducted using the Uclust consensus taxonomy assigner against the Greengene (version gg13_5) reference database. Because sequence numbers varied among samples, the OTU datasets were rarefied to the lowest uniform resample size (i.e. 1744 sequences). The bacteria we found belong to phyla typically found in freshwater ecosystems (see Figure S2 in Supporting Information; see also Harrop, Marks & Watwood, 2009; Besemer *et al.*, 2012; Wang *et al.*, 2012), although the same bacterial phyla can also be found in other ecosystems (Horner-Devine *et al.*, 2004). Thus, in our data, Proteobacteria, Bacteroidetes and Cyanobacteria dominated the samples. The most common taxa within Proteobacteria belonged to Gammaproteobacteria. The bacterial raw sequence data are available online in the European Nucleotide Archive (http://www.ebi.ac.uk/ena) and ResearchGate (http://www.researchgate.net/).

### *Sampling of bryophytes*

Bryophytes, including mosses and liverworts, were sampled using 10, 0.5 × 0.5 m, quadrats that were randomly allocated in a riffle site. All bryophyte species growing on stones and woody substrata were recorded and identified in each quadrat (Heino *et al.*, 2012). We also did a 'sweep-up' search at each site to find scarce species, hence guaranteeing that most species were detected. Bryophytes were identified in the field to species, but any difficult specimens were collected for verification in the laboratory.

### *Sampling and processing of macroinvertebrates*

To collect macroinvertebrates, we took a two-minute kick sample (net mesh size 0.3 m) covering most microhabitats present in a riffle of approximately 100 $m^2$ at each site (Heino *et al.*, 2012). Such samples contain hundreds to thousands of individuals in these boreal and subarctic streams and typically contain most of the species occurring at a site in a given season, mainly missing very rare, occasional or 'tourist' species that occur infrequently in headwater streams (Mykrä, Ruokonen & Muotka, 2006). Macroinvertebrates and associated material were immediately preserved in 70% ethanol in the field, and samples were taken to the laboratory for further processing and identification. Macroinvertebrates

were identified to the lowest possible taxonomic level, in most cases to species (> 70% of taxa), whereas the remaining taxa were determined to species group or genus (Heino, Ilmonen & Paasivirta, 2014a). We included blackflies (Diptera: Simuliidae) and non-biting midges (Diptera: Chironomidae), which are important components of stream invertebrate communities (Heino & Paasivirta, 2008), but they are often ignored in ecological and bioassessment studies (Raunio, Heino & Paasivirta, 2011). We excluded a few non-insect taxa that were identified to family level only.

### *Potential issues of different taxonomic resolutions among the four organismal groups*

Comparing results for different organismal groups determined to different taxonomic levels may potentially affect community patterns. However, based on our previous experience with diatoms (Heino & Soininen, 2007) and invertebrates (Heino, 2008), community patterns remain largely the same at the species and genus levels, and they may even be reproduced using family level data. Thus, we believe that the main patterns emerged in this study represent real natural phenomena. A uniform means of identification, perhaps based on molecular methods, would facilitate more direct comparisons of the community patterns among bacteria, diatoms, bryophytes and invertebrates.

### *Species accumulation curves*

Species accumulation curves were drawn to show how well our surveys sampled regional species pools. Species accumulation curves were based on the method 'exact' in the R package *vegan* (Oksanen *et al.*, 2013), calculating mean and standard errors for each number of sites sampled (Ugland, Gray & Ellingsen, 2003). Sampling can be considered adequate at the respective drainage basin scale if the curve levels off before the inclusion of all sites sampled.

### *Elements of metacommunity structure*

Elements of metacommunity structure (EMS) were evaluated based on the original ideas of Leibold & Mikkelson (2002), as further refined by Presley & Willig (2010), Presley *et al.* (2010) and Henriques-Silva *et al.* (2013). We took the 'range perspective', based on species distributions, in our analyses (Leibold & Mikkelson, 2002; Presley *et al.*, 2009). The analysis of the elements of metacommunity structure is based on three metrics: coherence, turnover and boundary clumping (Fig. 1). Before calculating these metrics, a site-by-species presence–absence matrix is ordinated through reciprocal averaging (i.e. correspondence analysis). Reciprocal averaging ordinates the sites so that those having similar species composition are close to each other and ordinates the species such that those having similar occurrence among the sites are close to each other (Gauch, 1982). Reciprocal averaging thereby defines a 'latent environmental gradient', and a metacommunity can be ordered along such a gradient that incorporates multiple environmental factors presumably of importance to the distributions of species (Leibold & Mikkelson, 2002; Presley & Willig, 2010). It is important to emphasise that the metacommunity structure that fits an empirical dataset best in the EMS analysis is the combination of the three elements of metacommunity structure, that is coherence, turnover and boundary clumping (Erős *et al.*, 2014). Hence, looking at the significance of boundary clumping before taking into account the significance of coherence and turnover, for example, should not be carried out.

The first metric that we evaluated is coherence. This metric is based on calculating the number of 'embedded absences' (*Abs*) in the ordinated matrix and, then, comparing the observed value to a null distribution of embedded absences from randomisations. An 'embedded absence' means that there is a gap in a species range (Leibold & Mikkelson, 2002; Presley & Willig, 2010). A small number of embedded absences (i.e. *Abs* is significantly lower than expected by chance) lead to positive coherence. In contrast, a large number of embedded absences (i.e. *Abs* is significantly larger than expected by chance) mean negative coherence. Significantly negative coherence thus points to a chequerboard distribution of species, non-significant coherence to randomness, and significantly positive coherence to any of nestedness, evenly spaced gradients, Gleasonian gradients or Clementsian gradients (Leibold & Mikkelson, 2002). The second metric is turnover, which is evaluated if coherence is positive, and it is used for deciding which gradient model best fits the data. Species turnover is measured as the number of times one species replaces (*Rep*) another between two sites in an ordinated matrix (Presley & Willig, 2010). Significantly negative turnover (i.e. *Rep* is significantly lower than expected by chance) is related to nestedness, whereas non-significant (i.e. moderately positive) or significantly positive turnover (i.e. *Rep* is significantly larger than expected by chance) points to evenly spaced, Gleasonian or Clementsian gradients (Leibold & Mikkelson, 2002; Henriques-Silva

*et al.*, 2013). These three types of gradients can be separated based on evaluation of boundary clumping (Leibold & Mikkelson, 2002; Presley *et al.*, 2010). Boundary clumping is analysed using Morisita's index and a $\chi^2$-test comparing observed and expected distributions of range boundary locations. Values of Morisita's index that are not different from 1 indicate randomly distributed range boundaries (i.e. Gleasonian gradients), values significantly larger than 1 indicate clumped range boundaries (i.e. Clementsian gradients), and values significantly less than 1 indicate hyperdispersed range boundaries (i.e. evenly spaced gradients).

Presley *et al.* (2010) further proposed that the combination of cases of significant positive coherence and non-significant turnover could be interpreted as 'quasi-structures', with non-significant negative turnover suggesting quasi-nestedness, and non-significant positive turnover suggesting quasi-Gleasonian, quasi-Clementsian or quasi-evenly spaced gradients (for details, see Presley *et al.*, 2010).

The significance of the index values for coherence and turnover was tested using the fixed-proportional null model (Gotelli, 2000). In this null model, the species richness of each site was maintained (i.e. row sums are fixed), but species ranges (i.e. columns) are filled based on their marginal probabilities. This model is ecologically plausible because species richness typically varies along main environmental gradients (Presley *et al.*, 2009). The fixed-proportional null model is recommended for use in the EMS analyses, because it is not highly sensitive to type I or type II errors (Presley *et al.*, 2009). By contrast, an alternative fixed–fixed null model would have been very strict and conservative, might have incorporated 'too much' biology into the null model (e.g. that the pattern is incorporated in the randomised matrices) and might hence have been prone to type II error (Gotelli, 2000; Presley *et al.*, 2009). Hence, we used only a fixed-proportional null model in our analyses. Random matrices were produced by the 'r1' method for the fixed-proportional null model as implemented in the R package *vegan* (Oksanen *et al.*, 2013). We used 999 simulations to provide random matrices, with the exception of bacteria where the very high number of OTUs and resulting long computation time forced us to use 99 simulations. Statistical significance was then assessed by comparing the observed index value from the original matrix to the distribution of values derived from the randomisations (Manly, 1995). Elements of metacommunity structure were evaluated for each organismal group and for both primary (axis 1) and secondary (axis 2) correspondence analysis (CA) axis. Elements of metacommunity structure were analysed using the R package *metacom* (Dallas, 2013) in the R environment (version 3.0.1, R Core Team 2013).

The EMS approach (Leibold & Mikkelson, 2002) has been criticised recently by some authors (Ulrich & Gotelli, 2013). Their criticism mainly concentrated on the intuitive and often-detected finding that a given dataset may show multiple significant patterns if those patterns are tested using different methods (e.g. various nestedness and co-occurrence metrics) and approaches. While we agree that a given empirical dataset is likely to show a number of significant patterns, we also emphasise that such an approach based on separate metacommunity analyses does not answer to our main study question about the 'best fit' metacommunity type in an empirical dataset. Other researchers have also recently shown that the EMS approach is a good means of indicating the best fit of an empirical dataset with one of the idealised patterns (Henriques-Silva *et al.*, 2013; Meynard *et al.*, 2013; Erős *et al.*, 2014; de la Sancha *et al.*, 2014; Heino & Alahuhta, 2015). Moreover, one major difference between the EMS approach and the traditional nestedness or co-occurrence analyses is that the EMS approach examines multiple different patterns along a single major ordination axis (i.e. a latent environmental gradient) in the data, whereas the traditional nestedness and co-occurrence analyses examine a given pattern in the whole site-by-species matrix (Presley *et al.*, 2010; Dallas, 2014; Dallas & Presley, 2014).

### *Community–environment relationships*

Community–environment relationships among the organismal groups were compared based on a set of *a priori* determined environmental variables (i.e. pH, conductivity, total phosphorus, stream width, depth, current velocity). Our statistical tool was canonical correspondence analysis (CCA), which is a constrained extension of correspondence analysis (Ter Braak, 1986; Legendre & Legendre, 2012), thus proving a direct link to the EMS analysis, which is based on correspondence analysis (Leibold & Mikkelson, 2002). We used CCA for examining the community–environment relationships along the first two ordination axes based on intraset correlations that maximise the correlation between an environmental variable and linear combination site scores along each CCA axis. We used these correlations to infer which environmental factors were best related to variation in community composition based on presence–absence data and to compare whether the same variables were important for each organismal group in each drainage basin.

CCAs were run using the R package *vegan* (Oksanen *et al*., 2013). We also correlated (Pearson coefficient) species richness with the same environmental variables that were used in the CCAs. These correlations were used to show whether species richness varied strongly along the measured environmental gradients.

### *Inclusion of rare species or operational taxonomic units*

All species (diatoms, bryophytes, invertebrates) or OTUs (bacteria) were included in the EMS and CCA analyses because excluding rare species would prejudice the results, given the pronounced rarity of many species in stream communities. Our previous research has shown that a large share of species of diatoms (Soininen & Heino, 2005), bryophytes (Heino & Virtanen, 2006) and invertebrates (Heino, 2005) and a high proportion of OTUs of bacteria (Heino *et al*., 2014b) are very rare (i.e. occur at a single site only) in northern regions. Although excluding rare species may or may not have a scientific basis (Cao, Williams & Williams, 1998), we chose to include all species or OTUs in all analyses because there is no unambiguous threshold for the exclusion of species (Poos & Jackson, 2012). Furthermore, trial analyses of the community data without rare species showed that the main patterns did not change from those reported here.

## Results

We found a total of 170, 182 and 118 diatom species in the Iijoki, Koutajoki and Tenojoki drainage basins, respectively. In the three drainage basins, the corresponding numbers of invertebrate species were 149, 164 and 98. We also found 21 and 40 bryophyte species in the Iijoki and Koutajoki drainage basins, and 6070 bacterial OTUs in the Tenojoki drainage basin. Species accumulation curves showed that the expected species richness started to level off, although not completely, before the inclusion of all sites surveyed in each drainage basin (Fig. 2). This finding suggested that our surveys captured a large share of all potentially occurring species at the drainage basin scale. While this finding held true for diatoms, bryophytes and invertebrates, bacteria showed a sharply contrasting pattern with their species richness increasing almost linearly with increasing number of sites sampled. This was due to the fact that the bacterial data incorporated a large number of very rare OTUs, with those occurring at a single site only accounting for 63% of all bacterial OTUs.

Based on the fixed-proportional null model, most metacommunities showed either Gleasonian or Clementsian gradients along the first CA axis (Table 1). There were some differences among the drainage basins, with all organismal groups fitting Gleasonian gradients in the southernmost Iijoki basin, diatoms and invertebrates showed Gleasonian gradients and bryophytes fitted Clementsian gradients in the Koutajoki basin, and diatoms, bacteria and invertebrates fitted Clementsian gradients in the northernmost Tenojoki basin. Along the second CA axis, there were Gleasonian, Clementsian, nested and random distributions, and there seemed to be no clear differences among the drainage basins or the organismal groups (Table 1). These results suggested that Gleasonian and Clementsian metacommunity structures dominated in describing the distributions of stream organisms. However, the interpretation of the elements of metacommunity structure also turned out to be sensitive to the approach of interpretation used. When the results were interpreted based on the protocol presented by Presley *et al.* (2010), many Gleasonian or Clementsian metacommunity structures were actually represented by quasi-structures (Table 1).

The most important environmental correlates of variation in community composition varied among organismal groups and drainage basins (Table 2). Thus, diatom community composition was most strongly related to velocity (axis 1) and stream width (axis 2) in the Iijoki basin, to depth (axis 1) and total phosphorus (axis 2) in the Koutajoki basin, and to pH (axis 1) and velocity (axis 2) in the Tenojoki basin. Variation in bacterial community composition was most strongly related to pH (axis 1) and depth (axis 2) in the Tenojoki basin. Variation in bryophyte community composition was mostly strongly related to stream width (axis 1) and velocity (axis 2) in the Iijoki basin and to total phosphorus (axis 1) and conductivity (axis 2) in the Koutajoki basin. For invertebrates, the most important variables for community composition were stream width in the Iijoki drainage basin, velocity (axis 1) and stream width (axis 2) in the Koutajoki basin, and stream width (axis 1) and depth (axis 2) in the Tenojoki basin.

As for community composition, the most influential environmental variables related to species richness varied among organismal groups and drainage basins (Table 2). For diatoms, stream width was the most important variable in the Iijoki and Tenojoki basins and pH in the Koutajoki basin. For bacteria, no environmental variable was strongly correlated with variation in the number of OTUs. Bryophyte richness was

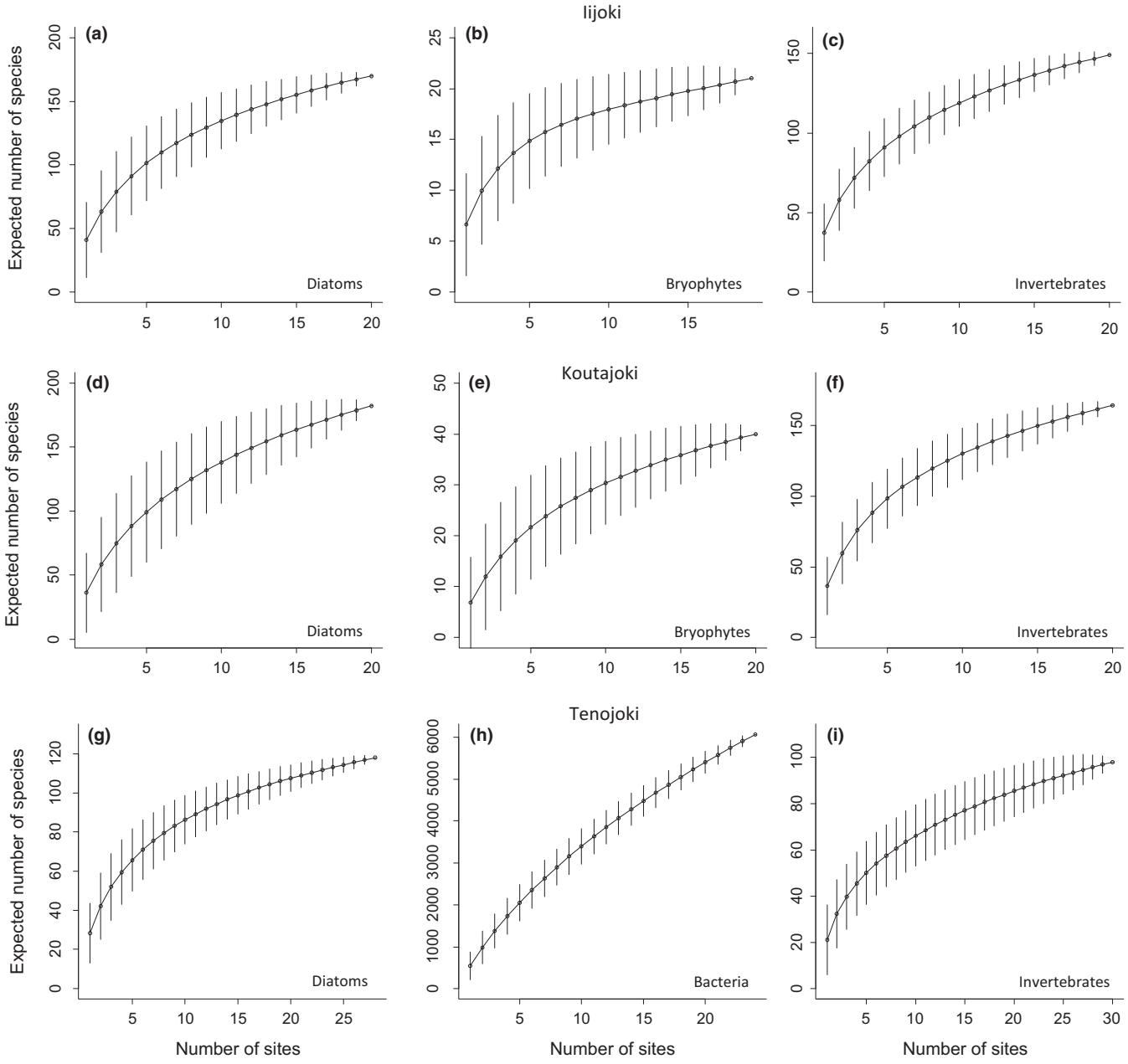


**Fig. 2** Species accumulation curves for each taxonomic group in each drainage basin. Note that the *y*-axis scale varies among the subfigures : The data from the three basins are as follows: Iijoki (a - c), Koutajoki (d - f) and Tenojoki (g - i).

most strongly correlated with conductivity in the Iijoki basin and with total phosphorus in the Koutajoki basin. Invertebrate species richness showed the strongest correlations with depth in the Iijoki basin and with total phosphorus in the Koutajoki basin. No variable explained invertebrate species richness adequately in the Tenojoki basin. In all drainage basins, species richness was only weakly related to the variables measured.

## Discussion

The elements of metacommunity structure (EMS) approach (Leibold & Mikkelson, 2002) provides a novel way to understand metacommunity structuring, complementing more conventional 'mechanistic' approaches in community ecology (Meynard *et al.*, 2013; Dallas & Presley, 2014). Most previous studies on stream metacommunities have also relied on mechanistic approaches,

**Table 1** Results of the EMS analysis for stream organisms. These results were based on the fixed-proportional null model. Interpretations followed Leibold & Mikkelson (2002; L&M) and Presley *et al.* (2010)

| | Coherence | | | | | Turnover | | | | | Boundary clumping | | | Interpretation | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Abs | Z-value | *P*-value | Sim mean | Sim SD | Rep | Z-value | *P*-value | Sim mean | Sim SD | Index | *P*-value | df | L&M | Presley *et al.* |
| CA axis 1 | | | | | | | | | | | | | | | |
| Iijoki | | | | | | | | | | | | | | | |
| Diatoms | 1361 | −7.098 | **0.001** | 1979 | 87 | 77330 | 3.465 | **0.001** | 35621 | 12035 | 0.716 | 0.291 | 167 | Gleasonian | Gleasonian |
| Bryophytes | 108 | −2.437 | **0.014** | 148 | 17 | 1678 | 0.170 | 0.864 | 1590 | 513 | 0.995 | 0.533 | 18 | Gleasonian | Quasi-Gleasonian |
| Invertebrates | 1416 | −3.780 | **0.001** | 1683 | 71 | 42006 | 1.490 | 0.136 | 28723 | 8913 | 1.463 | 0.157 | 146 | Gleasonian | Quasi-Gleasonian |
| Koutajoki | | | | | | | | | | | | | | | |
| Diatoms | 1620 | −5.716 | **0.001** | 2221 | 105 | 67199 | 1.222 | 0.221 | 47667 | 15972 | 0.256 | 0.066 | 179 | Gleasonian | Quasi-Gleasonian |
| Bryophytes | 259 | −2.055 | **0.039** | 320 | 29 | 8906 | 1.056 | 0.291 | 6593 | 2188 | 2.00 | **0.001** | 37 | Clementsian | Quasi-Clementsian |
| Invertebrates | 1411 | −6.466 | **0.001** | 1946 | 83 | 100784 | 5.988 | **0.001** | 35454 | 10909 | 1.613 | 0.106 | 161 | Gleasonian | Gleasonian |
| Tenojoki | | | | | | | | | | | | | | | |
| Diatoms | 1509 | −4.408 | **0.001** | 1872 | 83 | 69452 | 2.837 | **0.004** | 35317 | 12027 | 2.350 | **0.001** | 115 | Clementsian | Clementsian |
| Bacteria | 89452 | −7.533 | **0.001** | 117930 | 3780 | 75769955 | 6.607 | **0.001** | 14761506 | 9232680 | 5.862 | **0.024** | 6067 | Clementsian | Clementsian |
| Invertebrates | 1396 | −3.110 | **0.002** | 1655 | 83 | 60240 | 2.181 | **0.029** | 34073 | 11998 | 3.600 | **0.001** | 95 | Clementsian | Clementsian |
| CA axis 2 | | | | | | | | | | | | | | | |
| Iijoki | | | | | | | | | | | | | | | |
| Diatoms | 1526 | −4.933 | **0.001** | 1978 | 91 | 25872 | −0.741 | 0.458 | 34969 | 12280 | 1.193 | 0.335 | 167 | Nested | Quasi-nested |
| Bryophytes | 118 | −1.989 | **0.046** | 149 | 15 | 425 | −2.201 | **0.027** | 1570 | 520 | 1.357 | **0.032** | 18 | Nested | Nested clumped species loss |
| Invertebrates | 1515 | −2.454 | **0.014** | 1684 | 69 | 37944 | 1.021 | 0.307 | 28772 | 8975 | 1.254 | 0.281 | 146 | Gleasonian | Quasi-Gleasonian |
| Koutajoki | | | | | | | | | | | | | | | |
| Diatoms | 1745 | −4.641 | **0.001** | 2222 | 102 | 59520 | 0.838 | 0.401 | 46684 | 15305 | 2.560 | **0.002** | 179 | Clementsian | Quasi-Clementsian |
| Bryophytes | 264 | −1.857 | 0.063 | 319 | 30 | 5890 | −0.371 | 0.710 | 6719 | 2232 | 2.221 | 0.001 | 37 | Random | Random |
| Invertebrates | 1514 | −5.390 | **0.001** | 1947 | 80 | 97128 | 5.213 | **0.001** | 35791 | 11764 | 1.152 | 0.364 | 161 | Gleasonian | Gleasonian |
| Tenojoki | | | | | | | | | | | | | | | |
| Diatoms | 1554 | −3.998 | **0.001** | 1868 | 79 | 47734 | 1.117 | 0.264 | 35234 | 11191 | 1.186 | **0.003** | 115 | Clementsian | Quasi-Clementsian |
| Bacteria | 108862 | −3.193 | **0.001** | 118962 | 3161 | 33461280 | 2.693 | **0.007** | 13280859 | 7492241 | 11.725 | **0.001** | 6067 | Clementsian | Clementsian |
| Invertebrates | 1748 | 1.086 | 0.277 | 1654 | 86 | 17505 | −1.358 | 0.174 | 33924 | 12087 | 6.620 | 0.001 | 95 | Random | Random |

Significant p-values are indicated by bold font

 

**Table 2** Community–environment relationships along the first two CCA axes as evidenced by intraset correlations between each environmental variable and the ordination axes. Also, shown are species richness–environmental variables correlations. The strongest correlations between environmental variables and ordination axes or species richness are shown by bold font

| | Diatoms | | | Bryophytes | | | Invertebrates | | |
|---|---|---|---|---|---|---|---|---|---|
| | CCA1 | CCA2 | Richness | CCA1 | CCA2 | Richness | CCA1 | CCA2 | Richness |
| Iijoki | | | | | | | | | |
| Eigenvalue | 0.22 | 0.21 | | 0.21 | 0.20 | | 0.28 | 0.22 | |
| Conductivity | −0.14 | 0.13 | 0.16 | −0.21 | −0.09 | **0.34** | 0.43 | 0.16 | 0.13 |
| pH | −0.64 | 0.12 | 0.13 | −0.54 | 0.41 | 0.31 | −0.27 | 0.49 | 0.25 |
| Total phosphorus | 0.32 | 0.01 | −0.13 | 0.26 | −0.32 | 0.14 | 0.49 | 0.06 | −0.11 |
| Stream width | −0.13 | **−0.87** | **0.48** | **−0.81** | −0.17 | 0.32 | **0.59** | **0.64** | 0.05 |
| Depth | −0.46 | −0.36 | 0.38 | 0.13 | 0.43 | −0.09 | 0.00 | 0.03 | **−0.43** |
| Velocity | **−0.83** | 0.06 | 0.06 | −0.13 | **0.83** | −0.11 | −0.42 | 0.63 | −0.38 |
| Koutajoki | | | | | | | | | |
| Eigenvalue | 0.29 | 0.25 | | 0.40 | 0.38 | | 0.35 | 0.30 | |
| Conductivity | −0.18 | 0.16 | −0.37 | 0.70 | **−0.56** | 0.18 | 0.36 | −0.25 | 0.19 |
| pH | −0.33 | −0.04 | **−0.43** | 0.45 | −0.42 | 0.25 | 0.54 | −0.14 | 0.04 |
| Total phosphorus | 0.11 | **0.90** | 0.19 | **−0.83** | −0.46 | **0.32** | −0.53 | −0.15 | **0.53** |
| Stream width | 0.40 | −0.48 | 0.10 | −0.16 | −0.51 | −0.14 | 0.21 | **0.84** | 0.11 |
| Depth | **0.94** | 0.01 | 0.28 | 0.05 | −0.57 | −0.17 | −0.07 | 0.76 | 0.53 |
| Velocity | −0.30 | −0.43 | −0.37 | 0.46 | −0.32 | 0.17 | **0.95** | 0.03 | −0.28 |
| Tenojoki | | | | | | | | | |
| Eigenvalue | 0.15 | 0.13 | | 0.47 | 0.45 | | 0.22 | 0.14 | |
| Conductivity | 0.22 | −0.63 | 0.01 | −0.48 | −0.48 | −0.12 | 0.26 | 0.17 | −0.11 |
| pH | **0.84** | 0.09 | −0.04 | **−0.62** | −0.60 | −0.12 | 0.41 | 0.57 | 0.05 |
| Stream width | −0.21 | −0.15 | −0.31 | 0.06 | −0.53 | 0.05 | **0.79** | 0.18 | 0.13 |
| Depth | −0.34 | −0.05 | 0.03 | 0.57 | **−0.73** | 0.13 | 0.22 | **0.61** | 0.06 |
| Velocity | 0.18 | **−0.77** | −0.15 | 0.13 | −0.41 | −0.04 | 0.72 | −0.13 | 0.10 |

aiming to understand the relative roles of dispersal and environmental factors for community structure (see review by Heino *et al*., 2015). Hence, our aim was to find out if key organismal groups in streams show similar patterns of the elements of metacommunity structure that result in the same metacommunity structures. We found mostly Gleasonian and Clementsian metacommunities along the first two ordination axes, when following the original interpretations of Leibold & Mikkelson (2002). There were no notable differences among the organismal groups, but Gleasonian structures dominated in the southernmost Iijoki basin, with intermediate environmental heterogeneity, and Clementsian structures did so in the northernmost Tenojoki basin, with the lowest environmental heterogeneity. Therefore, these findings did not agree with our expectation that high environmental heterogeneity would promote the fit of metacommunity data to the gradient models (i.e. clearer nestedness in the low-heterogeneity drainage basin, and increasing turnover and a clearer fit with the gradient models in the high-heterogeneity drainage basin). This may be due to the low power available to test this hypothesis (i.e. three basins), the high heterogeneity of all natural stream systems and the fact that different organisms perceive environment (and therefore its heterogeneity) at different scales (Resh *et al*., 1994). Rather, the geographical gradient was obvious, with a change from Gleasonian to Clementsian structures from south to north. Similarly, in a large-scale study of lake fish, Henriques-Silva *et al.* (2013) also found clear geographical variation in metacommunity structuring, with Clementsian metacommunities dominating in the southern drainage basins, while nested metacommunities occurred across the whole Canadian province of Ontario. It is thus highly likely that there are differences in metacommunity structures and underlying processes among geographical regions, ecological settings and the biology of the organisms in question (Presley *et al*., 2012; Dallas & Presley, 2014).

An interesting finding was that organismal groups as different as bacteria, diatoms and invertebrates fitted best with Clementsian metacommunity structures in the northernmost drainage basin (70°N). Not only are these organisms very different ecologically, but their diversity varied to a considerable extent in this drainage basin, which was obviously partly related to the means of

identification (DNA-based versus morphological identification). This difference in identification might have been expected to bring about different patterns for the three organismal groups. While invertebrates (total diversity: 98 species) and diatoms (total diversity: 118 species) showed low diversity, typical of very high-latitude regions (Heino, 2011; Mittelbach, 2012), bacteria appeared to be a highly diverse group in terms of the number of operational taxonomic units (total diversity: 6070 OTUs). It is also worth noting that the regional pool of bacteria was clearly undersampled compared to those of diatoms, bryophytes and invertebrates (Fig. 2), suggesting that much more cryptic bacterial biodiversity exists in the study area (see also Esteban & Finlay, 2010). Based on both differences in biology and diversity, we could have expected these taxa to show different metacommunity structures. Rather, all these key organismal groups in northern streams fitted best with Clementsian metacommunity structures, suggesting some generalities among disparate organisms, which nevertheless had different relationships to underlying environmental gradients. It is possible that there are two or more groups of species in each organismal group that respond similarly to harsh environmental conditions in the northernmost drainage basin, leading to Clementsian gradients. It is also possible that, in these environmentally harsh systems, positive interactions (e.g. conditioning of the environment by bacteria) among species are necessary and lead to the occurrence of tightly associated sets of species (see also Boon *et al.*, 2014).

In our study system, environmental harshness comprises short ice-free periods (i.e. approximately five months in a year), challenging icy conditions in streams during winter (e.g. ice scour and freezing to the bed), low water temperature and very low potential productivity (e.g. total phosphorus values are typically $<5$ μg $L^{-1}$), which all are characteristics typical of high-latitude freshwater ecosystems (Wrona *et al.*, 2013). Such harshness may mean that species are more dependent on each other than in less harsh systems. However, also evidence opposite to that reasoning has been obtained, and some studied have shown increasing importance of biotic interactions in less harsh conditions than those in high-latitude streams (Schemske *et al.*, 2009; Defossez *et al.*, 2011). However, based on the present descriptive analyses, any comprehensive answer to the question of underlying mechanisms of the prevalence of Clementsian structures in the northernmost drainage basin would be premature.

The methods used to infer the best fit of empirical data with idealised metacommunity models also has an important role in our understanding of metacommunity patterns (Leibold & Mikkelson, 2002; Presley *et al.*, 2010; Ulrich & Gotelli, 2013; Dallas & Presley, 2014). Using different methods, Heino & Soininen (2005) suggested that Gleasonian gradients usually fitted best the distribution patterns of stream diatoms, although they also found instances of chequerboard and nested subset patterns in their dataset. That case study clearly showed the necessity for a rigorous approach in identifying the best fit of metacommunity structures to empirical data. Various studies of invertebrate and diatom metacommunities in streams have also shown that both nestedness (Heino, 2009; Soininen & Köngäs, 2012) and chequerboards (Heino & Soininen, 2005; Larsen & Ormerod, 2014) are common in these systems, although their prevalence and strength may be contingent on the environmental context (Heino, Mykrä & Rintala, 2010; Heino, 2013). For example, it is interesting to speculate that low environmental heterogeneity leads to nestedness, whereas high environmental heterogeneity leads to chequerboards (Heino, 2013; McCreadie & Bedwell, 2013). Most of the above-mentioned stream studies did not, however, consider simultaneously Clementsian, Gleasonian or evenly spaced gradients, emphasising the difficulty in associating different idealised patterns with empirical data based only on nestedness or co-occurrence analyses. It is also possible that a site-by-species matrix as a whole shows either nestedness or chequerboards (Heino & Soininen, 2005; Heino, 2009), but each particular ordination axis shows nestedness, Gleasonian or Clementsian gradients (Presley & Willig, 2010; Presley *et al.*, 2010).

The environmental correlates of the first two ordination axes differed among organismal groups and drainage basins. For example, diatom communities correlated most strongly with either physical (i.e. velocity, stream width) or chemical (i.e. total phosphorus, pH) variables, depending on the drainage basin. Similarly, bryophyte communities correlated with different chemical (i.e. total phosphorus, conductivity) and physical (i.e. velocity, depth) variables in different drainage basins. Invertebrate communities varied most clearly along physical habitat gradients (i.e. stream width, velocity, depth) although the relative importance of these variables also differed between the three drainage basins. Although such context-dependent community–environment relationships have often been observed in stream systems (Malmqvist & Mäki, 1994; Soininen, 2004; Heino *et al.*, 2012; Grönroos *et al.*, 2013), it was particularly evident that different organismal groups responded to different limiting or constraining environmental factors.

One might argue that large environmental gradients should lead to different gradient structures, whereas small environmental gradients should lead to nestedness (Heino, 2011). However, to our knowledge, no study has shown the effects of increasing environmental gradient lengths on metacommunity structures, although some have shown the effects of increasing environmental heterogeneity on increasingly more segregated distributions of species across a set of sites (Heino, 2013; McCreadie & Bedwell, 2013). We detected a nested pattern only once, which was perhaps not surprising because species richness did not show particularly strong correlations with any of the environmental variables measured. A strong richness gradient is a pre-requisite for a nested subset pattern (Patterson & Atmar, 1986), where species richness typically varies along a single major ecological gradient and there is an ordered gain or loss of species (Hylander *et al*., 2005). Instead, Gleasonian and Clementsian gradients occurred frequently irrespective of the underlying environmental heterogeneity in a drainage basin. This finding may be related to the fact that natural stream systems are inherently highly heterogeneous (Ward, Malard & Tockner, 2002) and, hence, responses of species to environmental gradients are typically more complex than a simple gain or loss of species along ecological gradients (Heino *et al*., 2010; Heino, 2013).

Factors other than environmental heterogeneity may also affect metacommunity structures. For example, connectivity between sites and associated dispersal processes is very important for metacommunity organisation (Leibold *et al*., 2004; Brown *et al*., 2011; Lindström & Langenheder, 2012; Heino *et al*., 2015), although associating dispersal with the idealised metacommunity structures may be difficult (Meynard *et al*., 2013). This is because the EMS approach indirectly assumes that there is a 'latent environmental gradient' driving variation in community composition among sites (Presley & Willig, 2010). Hence, with adequate levels of connectivity and intermediate dispersal rates between sites, we could also expect that species sorting is most effective (Leibold *et al*., 2004), resulting in clear variation in community composition along environmental gradients (Heino & Peckarsky, 2014). Our previous work in the same three study areas has shown that connectivity, measured indirectly as the distance along aquatic pathways between sites, has a minor role in structuring invertebrate metacommunities (Grönroos *et al*., 2013), and that environmental factors mainly determine variation in the community composition of diatoms, bryophytes and invertebrates (Heino *et al*., 2012). We suggest, therefore, that variations in stream communities are probably driven mostly by environmental differences between sites in our high-latitude study regions (Heino *et al*., 2012; Grönroos *et al*., 2013).

To conclude, stream organisms fitted best Gleasonian and Clementsian metacommunity structures along environmental gradients. Although we found no clear relationships between underlying environmental heterogeneity and metacommunity patterns, we did detect among-region variation in metacommunity patterns, with Gleasonian gradients dominating in the southernmost drainage basins and Clementsian gradients being common in the northernmost drainage basin. We propose that further studies should examine the effects of increasing environmental gradient length and site connectivity on idealised metacommunity structures. Such approach may not only provide interesting information about possible changes in the patterns that fit the idealised models, but also offer a means of inferring the mechanisms underlying metacommunity structure.

## Acknowledgments

This study was supported by grants from the Academy of Finland, Kone Foundation and Maj and Tor Nessling Foundation. Mira Grönroos, Tommi Karhu and Heikki Mykrä provided assistance during the field work, and Marko Suokas helped us with the bacterial analyses. We thank Alan Hildrew and two anonymous reviewers for excellent comments that improved this manuscript.

## References

Anderson M.J., Ellingsen K.E. & McArdle B.H. (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, **9**, 683–693.

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

Boon E., Meehan C.J., Whidden C., Wong D.H., Langille M.G. & Beiko R.G. (2014) Interactions in the microbiome: communities of organisms and communities of genes. *FEMS Microbiology Reviews*, **38**, 90–118.

Brown B.L., Swan C.M., Auerbach D.A., Grant E.H.C., Hitt N.P., Maloney K.O. *et al.* (2011) Metacommunity theory as a multispecies, multiscale framework for studying the influence of river network structure on riverine communities and ecosystems. *Journal of the North American Benthological Society*, **30**, 310–327.

Cao Y., Williams D.D. & Williams N.E. (1998) How important are rare species in aquatic community ecology and bioassessment? *Limnology and Oceanography*, **43**, 1403–1409.

Caporaso J.G., Justin Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Clements F.E. (1916) *Plant succession: an analysis of the development of vegetation*. Carnegie Institution of Washington, Washington.

Cottenie K. (2005) Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, **8**, 1175–1182.

Dallas T. (2013) metacom: Analysis of the "elements of metacommunity structure". R package version 1.2. Available at: http://CRAN.R-project.org/package=metacom

Dallas T. (2014) metacom: an R package for the analysis of metacommunity structure. *Ecography*, **37**, 402–405.

Dallas T. & Presley S.J. (2014) Relative importance of host environment, transmission potential and host phylogeny to the structure of parasite metacommunities. *Oikos*, **7**, 866–874.

Defossez E., Courbaud B., Marcais B., Thuiller W., Granda E. & Kunstler G. (2011) Do interactions between plant and soil biota change with elevation? A study on Fagus sylvatica. *Biology Letters*, **7**, 699–701.

Diamond J.M. (1975) Assembly of species communities. In *Ecology and evolution of communities*. (Cody M.L. & Diamond J.M. eds). Harvard University Press, Cambridge, pp. 342–444.

Edgar R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Erős T., Sály P., Takács P., Higgins C.L., Bíró P. & Schmera D. (2014) Quantifying temporal variability in the metacommunity structure of stream fishes: the influence of non-native species and environmental drivers. *Hydrobiologia*, **722**, 31–43.

Esteban G.F. & Finlay B.J. (2010) Conservation work is incomplete without cryptic biodiversity. *Nature*, **463**, 293.

Fernandes I.M., Henriques-Silva R., Penha J., Zuanon J. & Peres-Neto P.R. (2014) Spatiotemporal dynamics in a seasonal metacommunity structure is predictable: the case of floodplain-fish communities. *Ecography*, **37**, 464–475.

Gauch H.G. (1982) *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge.

Gleason H.A. (1926) The individualistic concept of the plant association. *Bulletin of Torrey Botanical Club*, **53**, 7–26.

Gotelli N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606–2621.

Gotelli N.J. & McCabe D.J. (2002) Species coexistence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology*, **83**, 2091–2096.

Grönroos M., Heino J., Siqueira T., Landeiro V.L., Kotanen J. & Bini L.M. (2013) Metacommunity structuring in stream networks: roles of dispersal mode, distance type and regional environmental context. *Ecology and Evolution*, **3**, 4473–4487.

Haas B.J. *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, **21**, 494–504.

Harrop B.L., Marks J.C. & Watwood M.E. (2009) Early bacterial and fungal colonization of leaf litter in Fossil Creek, Arizona. *Journal of the North American Benthological Society*, **28**, 383–396.

Heino J. (2005) Positive relationship between regional distribution and local abundance in stream insects: a consequence of niche breadth or niche position? *Ecography*, **28**, 345–354.

Heino J. (2008) Influence of taxonomic resolution and data transformation on biotic matrix concordance and assemblage-environment relationships in stream macroinvertebrates. *Boreal Environment Research*, **13**, 359–369.

Heino J. (2009) Species co-occurrence, nestedness and guild-environment relationships in stream macroinvertebrates. *Freshwater Biology*, **54**, 1947–1959.

Heino J. (2011) A macroecological perspective of diversity patterns in the freshwater realm. *Freshwater Biology*, **56**, 1703–1722.

Heino J. (2013) Environmental heterogeneity, dispersal mode and co-occurrence in stream macroinvertebrates. *Ecology and Evolution*, **3**, 344–355.

Heino J. (2014) Taxonomic surrogacy, numerical resolution and responses of stream macroinvertebrate communities to ecological gradients: are the inferences transferable among regions? *Ecological Indicators*, **36**, 186–194.

Heino J. & Alahuhta J. (2015) Elements of regional beetle faunas: faunal variation and compositional breakpoints along climate, land cover and geographical gradients. *Journal of Animal Ecology*, **84**, 427–441.

Heino J., Grönroos M., Soininen J., Virtanen R. & Muotka T. (2012) Context dependency and metacommunity structuring in boreal headwater streams. *Oikos*, **121**, 537–544.

Heino J., Ilmonen J., Kotanen J., Mykrä H., Paasivirta L., Soininen J. *et al.* (2009) Surveying biodiversity in protected and managed areas: algae, macrophytes and macroinvertebrates in boreal forest streams. *Ecological Indicators*, **9**, 1179–1187.

Heino J., Ilmonen J. & Paasivirta L. (2014a) Continuous variation of macroinvertebrate communities along environmental gradients in northern streams. *Boreal Environment Research*, **19**, 21–38.

Heino J., Melo A.S., Siqueira T., Soininen J., Valanko S. & Bini L.M. (2015) Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects. *Freshwater Biology*, doi:10.1111/fwb.12533, in press.

Heino J., Mykrä H. & Rintala J. (2010) Assessing patterns of nestedness in stream insect assemblages along environmental gradients. *Ecoscience*, **17**, 345–355.

Heino J. & Paasivirta L. (2008) Unravelling the determinants of stream midge biodiversity in a boreal drainage basin. *Freshwater Biology*, **53**, 884–896.

Heino J. & Peckarsky B.L. (2014) Integrating behavioral, population and large-scale approaches for understanding stream insect communities. *Current Opinion in Insect Science*, **2**, 7–13.

Heino J. & Soininen J. (2005) Assembly rules and community models for unicellular organisms: patterns in diatoms of boreal streams. *Freshwater Biology*, **50**, 567–577.

Heino J. & Soininen J. (2007) Are higher taxa adequate surrogates for species-level assemblage patterns and species richness in stream organisms? *Biological Conservation*, **137**, 78–89.

Heino J., Tolkkinen M., Pirttilä A.M., Aisala H. & Mykrä H. (2014b) Microbial diversity and community–environment relationships in boreal streams. *Journal of Biogeography*, **41**, 2234–2244.

Heino J. & Virtanen R. (2006) Relationships between distribution and abundance vary with spatial scale and ecological group in stream bryophytes. *Freshwater Biology*, **51**, 1879–1889.

Henriques-Silva R., Lindo Z. & Peres-Neto P.R. (2013) A community of metacommunities: exploring pattern in species distributions across large geographical scales. *Ecology*, **94**, 627–639.

Hoagland B.W. & Collins S.L. (1997) Gradient models, gradient analysis, and hierarchical structure in plant communities. *Oikos*, **78**, 23–30.

Horner-Devine M.C., Lage M., Hughes J. & Bohannan B.J.M. (2004) A taxa-area relationship for bacteria. *Nature*, **432**, 750–753.

Horner-Devine M.C., Silver J.M., Leibold M.A., Bohannan B.J., Colwell R.K., Fuhrman J.A. *et al.* (2007) A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology*, **88**, 1345–1353.

Hylander K., Nilsson K., Jonsson B.G. & Göthner T. (2005) Differences in habitat quality explain nestedness in a land snail metacommunity. *Oikos*, **108**, 351–361.

Illies J. & Botosaneanu L. (1963) Problémes et méthodes de la classification et de la zonation éologique des eaux courantes, considerées surtout du point de vue faunistique. Mitteilungen internationalen Vereinigung theoretische und angewandte Limnologie, **12**, 1–57.

Keith S.A., Newton A.C., Morecroft M.D., Golicher D.J. & Bullock J.M. (2011) Plant metacommunity structure remains unchanged during biodiversity loss in English woodlands. *Oikos*, **120**, 302–310.

Korhonen J.J., Köngäs P. & Soininen J. (2013) Temporal variation of diatom assemblages in oligotrophic and eutrophic streams. *European Journal of Phycology*, **48**, 141–151.

Larsen S. & Ormerod S.J. (2014) Anthropogenic modification disrupts species co-occurrence in stream invertebrates. *Global Change Biology*, **20**, 51–60.

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. *Ecology Letters*, **7**, 601–613.

Leibold M.A. & Mikkelson G.M. (2002) Coherence, species turnover, and boundary clumping: elements of metacommunity structure. *Oikos*, **97**, 237–250.

Legendre P. & Legendre L. (2012) *Numerical Ecology*. Elsevier, Amsterdam.

Lindström E.S. & Langenheder S. (2012) Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports*, **4**, 1–9.

Malmqvist B. & Mäki M. (1994) Benthic macroinvertebrate assemblages in north Swedish streams: environmental relationships. *Ecography*, **17**, 9–16.

Manly B.J.F. (1995) A note on the analysis of species co-occurrences. *Ecology*, **76**, 1109–1115.

McCreadie J.W. & Bedwell C.R. (2013) Patterns of co-occurrence of stream insects and an examination of a causal mechanism: ecological checkerboard or habitat checkerboard? *Insect Conservation and Diversity*, **6**, 105–113.

McIntosh R.P. (1995) H. A. Gleason's "individualistic community concept" and theory of animal communities: a continuing controversy. *Biological Reviews*, **70**, 317–357.

Meynard C.N., Lavergne S., Boulangeat I., Garraud L., Van Es J., Mouquet N. *et al.* (2013) Disentangling the drivers of metacommunity structure across spatial scales. *Journal of Biogeography*, **40**, 1560–1571.

Mittelbach G.G. (2012) *Community Ecology*. Sinauer, Sunderland.

Mykrä H., Ruokonen T. & Muotka T. (2006) The effect of sample duration on the efficiency of kick-sampling in two streams with contrasting substratum heterogeneity. *Verhandlungen der Internationale Vereinigung für Theoretische and Angewandte Limnologie*, **29**, 1351–1355.

National Board of Waters and the Environment (1981) Vesihallinnon analyysimenetelmät. *Publications of National Board of Waters*, **213**, 1–136.

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 2.0-6. http://CRAN.R-project.org/package=vegan.

Patterson B.D. & Atmar W. (1986) Nested subsets and the structure of insular mammal faunas and archipelagos. *Biological Journal of the Linnean Society*, **28**, 65–82.

Peres-Neto P.R. (2004) Patterns in the coexistence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*, **140**, 353–360.

Poos M.S. & Jackson D.A. (2012) Addressing the removal of rare species in multivariate bioassessments: the impact of methodological choices. *Ecological Indicators*, **18**, 82–90.

Presley S.J., Cisneros L.M., Patterson B.D. & Willig M.R. (2012) Vertebrate metacommunity structure along an extensive elevational gradient in the tropics: a compari-

son of bats, rodents and birds. *Global Ecology and Biogeography*, **21**, 968–976.
Presley S.J., Higgins C.L., López-González C. & Stevens R.D. (2009) Elements of metacommunity structure of Paraguayan bats: multiple gradients require analysis of multiple axes of variation. *Oecologia*, **160**, 781–793.
Presley S.J., Higgins C.L. & Willig M.R. (2010) A comprehensive framework for the evaluation of metacommunity structure. *Oikos*, **119**, 908–917.
Presley S.J. & Willig M.R. (2010) Bat metacommunity structure on Caribbean islands and the role of endemics. *Global Ecology and Biogeography*, **19**, 185–199.
Presley S.J., Willig M.R., Bloch C.P., Castro-Arellano I., Higgins C.L. & Klingbeil B.T. (2011) A complex metacommunity structure for gastropods along an elevational gradient. *Biotropica*, **43**, 480–488.
Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P. *et al*. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, **41**, D590–D596.
R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/
Raunio J., Heino J. & Paasivirta L. (2011) Non-biting midges in biodiversity conservation and environmental assessment: findings from boreal freshwater ecosystems. *Ecological Indicators*, **11**, 1057–1064.
Resh V., Hildrew A.G., Statzner B. & Townsend C.R. (1994) Theoretical habitat templets, species traits and species richness: a synthesis of long-term ecological research on the Upper Rhone River in the context of concurrently developed ecological theory. *Freshwater Biology*, **31**, 539–554.
de la Sancha N., Higgins C.L., Presley S.J. & Strauss R.E. (2014) Metacommunity structure in a highly fragmented forest: has deforestation in the Atlantic Forest altered historic biogeographic patterns? *Diversity and Distributions*, **9**, 1058–1070.
Schemske D.W., Mittelbach G.G., Sobel J.M., Cornell H.V. & Roy K. (2009) Is there a latitudinal gradient in the importance of biotic interactions? *Annual Review of Ecology, Evolution, and Systematics*, **40**, 245–269.
Schmera D., Erős T. & Heino J. (2013) Habitat filtering determines spatial variation of macroinvertebrate community traits in northern headwater streams. *Community Ecology*, **14**, 77–88.
Soininen J. (2004) Determinants of benthic diatom community structure in boreal streams: the role of environmental and spatial factors at different scales. *International Review of Hydrobiology*, **89**, 139–150.
Soininen J. & Heino J. (2005) Relationships between local population persistence, local abundance and regional occupancy of species: patterns in diatoms of boreal streams. *Journal of Biogeography*, **32**, 1971–1978.
Soininen J. & Köngäs P. (2012) Analysis of nestedness in freshwater communities – patterns across species and trophic levels. *Freshwater Science*, **31**, 1145–1155.
Stone L. & Roberts A. (1990) The checkerboard score and species distribution. *Oecologia*, **85**, 74–79.
Ter Braak C.J.F. (1986) Canonical Correspondence Analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167–1179.
Tilman D. (1982) *Resource Competition and Community Structure*. Princeton University Press, Princeton, New Jersey.
Ugland K.I., Gray J.S. & Ellingsen K.E. (2003) The species-accumulation curve and estimation of species richness. *Journal of Animal Ecology*, **72**, 888–897.
Ulrich W. & Gotelli N.J. (2013) Pattern detection in null model analysis. *Oikos*, **122**, 2–18.
Vannote R.L., Minshall G.W., Cummins K.W., Sedell J.R. & Cushing C.E. (1980) The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 130–137.
Wang J., Soininen J., He J. & Shen J. (2012) Phylogenetic clustering increases with elevation for microbes. *Environmental Microbiology Reports*, **4**, 217–226.
Ward J.V., Malard F. & Tockner K. (2002) Landscape ecology: a framework for integrating pattern and process in river corridors. *Landscape Ecology*, **17**, 35–45.
Wright D.H., Patterson B.D., Mikkelson G.M., Cutler A. & Atmar W. (1998) A comparative analysis of nested subset patterns of species composition. *Oecologia*, **113**, 1–20.
Wrona F.J., Reist J.D., Amundsen P.-E., Chambers P.A., Christoffersen K., Culp J.M. *et al*. (2013) Freshwater ecosystems. In: *Arctic Biodiversity Assessment. Status and Trends in Arctic Biodiversity* (Ed. H. Meltofte), pp. 335–377. Conservation of Arctic Flora and Fauna Arctic Council, Iceland.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** The concentrations used in PCRs and information about pipetting leading to final concentrations.

**Fig. S1.** Maps showing the locations of the study sites.

**Fig. S2.** The proportions of different bacterial groups of the total amount of sequence.

*(Manuscript accepted 9 February 2015)*