

Running title: Metacommunity ecology and biogeography of aquatic organisms

Metacommunity ecology meets biogeography: effects of geographical region, spatial dynamics and environmental filtering on community structure in aquatic organisms

Jani Heino¹, Janne Soininen², Janne Alahuhta^{3,4}, Jyrki Lappalainen⁵ and Risto Virtanen^{6,7,8}

¹Finnish Environment Institute, Natural Environment Centre, Biodiversity, P.O. Box 413, FI-90014 Oulu, Finland.

²University of Helsinki, Department of Geosciences and Geography, P.O. Box 64, FI-00014 University of Helsinki, Finland.

³University of Oulu, Geography Research Unit, P.O. Box 3000, FI-90014 University of Oulu, Finland.

⁴Finnish Environment Institute, Freshwater Centre, State of Surface Waters, P.O. Box 413, FI-90014 Oulu, Finland

⁵University of Helsinki, Department of Environmental Sciences, P.O. Box 65, FI-00014, University of Helsinki, Finland.

⁶ Department of Physiological Diversity, Helmholtz Center for Environmental Research – UFZ, Permoserstr. 15, Leipzig 04318, Germany.

⁷German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig D-04103, Germany.

⁸Department of Ecology, University of Oulu, PO Box 3000, FI-90014 University of Oulu, Finland.

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

Author contribution statement. All authors devised the ideas and provided data. JH ran the statistical analyses and led the writing. All authors commented on the manuscript drafts.

Abstract. We examined variation in the composition of six freshwater organismal groups across **various** drainage basins in Finland. We first modelled spatial structures within each drainage basin using Moran eigenvector maps. Second, we partitioned variation in community structure among three groups of predictors using constrained ordination: (1) local environmental variables, (2) spatial variables, and (3) dummy variable drainage basin identity. Third, we examined turnover and nestedness components of multiple-site beta diversity, and tested the best fit patterns of our datasets using the “elements of metacommunity structure” analysis. Our results showed that basin identity and local environmental variables were significant predictors of community structure, whereas within-basin spatial effects were typically negligible. In half of the organismal groups (diatoms, bryophytes, zooplankton), basin identity was a slightly better predictor of community structure than local environmental variables, whereas the opposite was true for the remaining three organismal groups (insects, macrophytes, fish). Both pure basin and local environmental fractions were, however, significant after accounting for the effects of the other predictor variable sets. All organismal groups exhibited high levels of beta diversity, which was mostly attributable to the turnover component. Our results showed consistent Clementsian type metacommunity structures, suggesting that subgroups of species responded similarly to environmental factors or drainage basin limits. We conclude that aquatic communities across large scales are mostly determined by environmental and basin effects, which leads to high beta diversity and prevalence of Clementsian community types.

Keywords. Bryophytes, diatoms, fish, invertebrates, lakes, macrophytes, metacommunities, streams.

Introduction

Biogeography and community ecology are two disciplines that combine history, dispersal, biotic interactions and environmental filtering as determinants of the structure of biotic assemblages. However, a better understanding of the determinants of biotic assemblages might benefit from a closer conceptual unification of these disciplines (Jenkins and Ricklefs 2011; Ricklefs and Jenkins 2011). Biotic assemblages can be understood to comprise either regional biotas or local communities, depending on the grain size under investigation (Beck et al. 2012; Pinel-Alloul et al. 2013). One means to investigate biogeographic and ecological influences is to compare the effects of regional, spatial and environmental drivers of local communities over large spatial extents (Shurin et al. 2009; Bini et al. 2014; Gonçalves-Souza et al. 2014).

A major aim of biogeography is to consider evolutionary, historical and climatic influences on regional biotas. One typically finds a strong relationship between present-day climate and species richness (e.g. Hawkins and Porter 2003) or composition (e.g. Heino and Alahuhta 2015) of regional biotas. The same is true for historical effects on regional biotas, which can be investigated as phylogenetic patterns (e.g. Wiens 2012) or using various statistical approaches as proxies of historical effects (e.g. Hortal et al. 2011). The degree to which the influences of these broad-scale factors remain important when the focus is on local communities is still elusive. Some studies have suggested that regional and historical influences remain significant even when the focus is on local communities (Hoeinghaus et al. 2007; Vyverman et al. 2007), but others have shown that local environmental factors account for significant variation in local community structure even over broad spatial extents (Van der Gucht et al. 2007; Gonçalves-Souza et al. 2014; Souffreau et al. 2015). The influence of

regional characteristics on local communities can be studied as an overall “region effect” on local communities (Fig. 1), and it indirectly relates to historical effects and climatic forcing on local community structure (Declerck et al. 2011; Viana et al. 2015).

Metacommunity ecology is a recently emerged subdiscipline of ecology, where dispersal among sites is considered a key to understand biotic assemblages (Leibold et al. 2004; Jocque et al. 2010). Metacommunity ecology emphasises the idea that dispersal among sites within a region is an important process affecting the structure of local communities (Leibold et al. 2004). Dispersal may be limiting or homogenizing local communities, the effects of which may not be easily distinguishable because both may induce spatial structuring in the biological data (e.g. Ng et al. 2009). However, those effects can be at least partly separated by focusing on nested spatial scales (Declerck et al. 2011; Silva and Hernández 2015), where differences among regions may denote dispersal limitation and spatial structures within a region mainly relate to homogenising effects of dispersal that can happen via mass effects (e.g. Mouquet and Loreau 2003). For freshwater organisms, regions can be individual drainage basins (e.g. Heino 2011), whereby dispersal is more likely to take place within such regions than between regions (e.g. Heino et al. 2015a). Regions could also be delineated based on drainage basin boundaries, major landscape configurations or ecoregions in terrestrial studies.

Metacommunity theory also predicts that species sorting, i.e. filtering of species by local abiotic and biotic factors, is most pronounced when dispersal rates are intermediate (Leibold et al. 2004). Such intermediate dispersal allows species to track variation in environmental conditions among sites within a region (e.g. a drainage basin), resulting in a relatively good match between environmental conditions and community structure (Leibold et al. 2004). True species sorting may be disrupted by both limiting and high dispersal rates (Ng et al. 2009; Winegardner et al. 2012), although understanding their relative importance

may be masked by spatially-structured environmental variation (Pinel-Alloul et al. 1995; Heino et al. 2015a). Spatially-autocorrelated environmental variables are a typical phenomenon in observational studies, often making it difficult to infer the relative roles of species sorting and dispersal effects on community structure (Pinel-Alloul et al. 1995; Bonada et al. 2012; Heino et al. 2015b). Hence, heuristic approaches across different scales should be used to disentangle those effects on community structure.

In addition to explaining variation in local community structure, a major aim of both biogeography and metacommunity ecology is to quantify the degree of variation (e.g. beta diversity, Baselga 2010) or describe predominant patterns (e.g. elements of metacommunity structure, Leibold and Mikkelsen 2002) in biological survey data. While beta diversity has been quantified at various spatial grains and extents (Soininen et al. 2007, Anderson et al. 2011), the elements of metacommunity structure have mostly been tested using data from local communities within relatively small regions (Leibold and Mikkelsen 2002; Heino et al. 2015c; but see Presley and Willig 2010; Meynard et al. 2013). Recently, Heino and Alahuhta (2015) applied the elements of metacommunity structure approach to encompass large spatial grain and geographical extent. They found, as opposed to sets of local communities within small regions where various patterns are typically detected (Heino et al. 2015c; Tonkin et al. 2015a), that regional beetle faunas across a broad geographical gradient showed consistent Clementsian type variation (Heino and Alahuhta 2015). **Clementsian type variation emphasises discrete ‘community types’ along ecological gradients, such that subgroups of species replace other subgroups in space (Clements 1936). Such variation also suggests that subgroups of species either responded similarly to environmental variation or are affected by similar historical effects (Heino and Alahuhta 2015; Tonkin et al. 2015b).** We here expanded this approach from single drainage basins to encompass local communities of six aquatic organismal groups surveyed across three drainage basins.

We first expected the predominance of environmental factors in affecting metacommunity organization (Cottenie 2005; Van der Gucht et al. 2007). Such patterns have been found in many stream (Göthe et al. 2013; Grönroos et al. 2013; Tonkin et al. 2015b) and lake studies (Soininen et al. 2011; Alahuhta and Heino 2013; Heino 2013). Second, we expected that basin identity and its associated biogeographical and climatic aspects, would show the strongest effect on variation in community structure. This is because biogeographical factors, including regional variation in climate, should be most important at large spatial extents (Gonçalves-Souza et al. 2014; Viana et al. 2015). We had data for three groups of organisms surveyed in streams (i.e. diatoms, bryophytes, insects) and three groups of organisms surveyed in lakes (i.e. macrophytes, zooplankton, fish). These organismal groups show wide variation in ecological and biological characteristics, including life form, body size and dispersal mode (Heino et al. 2015c). Hence, we examined whether different organismal groups would show different patterns, with (a) lake organisms being more dispersal limited than stream organisms, the communities of the latter which are better connected by dispersal than those of the former, (b) passively dispersing organisms with small propagules (i.e. diatoms, bryophytes, macrophytes, zooplankton) showing less biogeographical variation than more actively dispersing large organisms (i.e. insects, fish), and (c) passively dispersing organisms should show stronger environmental control than actively dispersing organisms across the biogeographical scales of the three drainage basins. This is because small passively dispersing organisms (e.g. diatoms), which can disperse passively via air and animal vectors (Kristiansen 1996), may overcome drainage basin boundaries more easily than actively dispersing organisms restricted to dispersal via watercourses (De Bie et al. 2012). We also examined whether Clementsian metacommunity structures and high beta diversity would be evident for all organismal groups because our

surveys comprised relatively large geographical and environmental gradients (Heino and Alahuhta 2015; Tonkin et al. 2015b).

This study builds on our previous research on metacommunities in northern streams (e.g. Grönroos et al. 2013) and lakes (e.g. Soininen et al. 2011). Our present study provides new comparative information about the responses of different aquatic organismal groups to region identity, within-region spatial structuring and local environmental factors. We also show that the metacommunity structures are largely invariable at a biogeographic scale regardless of organismal and ecosystem-specific differences, which adds to research conducted within smaller geographic regions (e.g. Heino et al. 2015c).

Materials and Methods

Dataset characteristics and environmental variables

We re-analysed some of our recently-collected data (Soininen et al. 2009; 2011; Grönroos et al. 2013; Alahuhta et al. 2012; 2015; Heino et al. 2015c) for three groups of stream organisms (i.e. diatoms, bryophytes, insects) and three groups of lake organisms (i.e. macrophytes, zooplankton, fish). However, in this study, we combined data such that they comprised 45 to 60 sites across three drainage basins to facilitate comparative purposes. In all cases, datasets were carefully taxonomically harmonised to guarantee that they were comparable. The three drainage basins and sites surveyed generally differed between the organismal groups (Fig. 2), but stream diatoms and bryophytes were surveyed at the same sites. Each drainage basin drains into the sea (i.e. Arctic Ocean, White Sea or Baltic Sea). Due to limited resources, sites

were sometimes sampled in different years to avoid large seasonal variation. Moreover, our experience on northern freshwaters has shown that between-year variation is likely to be less pronounced than between-season variation.

Stream diatoms

We sampled 45 stream sites across three drainage basins in Finland. The drainage systems were: (1) Koutajoki (centered on 66°N, 29°E), (2) Kemijoki (67°N, 28°E), and (3) Muonionjoki (68°N, 24°E). The spatial extent (a rectangle encompassing all study sites) comprising the three study areas was 63,609 km². Generally, 15 sites per region appeared to be a sufficient sample size, detecting the majority of the diatom species present in the regional species pool of a drainage basin (Soininen et al. 2009). Algal sampling was conducted once for each site during summer low-flows in August 2001 or 2004. Sampling was confined to near-pristine streams. All sampling was conducted by the same field crew using a strictly standardized field protocol. Each study site with a length of 10 m was divided into 5 or 10 cross-stream transects, depending on stream width. One or two stones were selected randomly in each transect, and diatoms were scraped off the stones from a predefined area (3.1 cm²), using a rubber template. Subsamples, 10 in total, were then pooled into a composite sample for each site. In the laboratory, fresh samples were carefully checked to guarantee that most diatom frustules were alive before acid combustion. We used acid combustion (HNO₃: H₂SO₄, 2:1) to clean frustules of organic material. Cleaned diatoms were mounted in Naphrax and a total of 500 frustules per sample were identified and counted, using phase-contrast light microscopy (magnification 1000×) (for details, see Soininen et al. 2009). **In total, 96% of diatoms were identified to species.** We also measured current velocity,

shading, particle size, moss cover, conductivity and pH in the field, and analyzed water samples for water colour and total phosphorus in the laboratory.

Stream bryophytes

Stream bryophytes were sampled at the same 45 sites as diatoms using a systematic sampling protocol. At each stream site, 10 50 × 50 cm plots were studied. The plots were placed at regular intervals of 2 m along the approximately 20 m long riffle section. The order of sampling proceeded from the stream margin to the center of the stream to cover habitat variation in the riffle section. Bryophytes were either identified in the field or samples of difficult-to-identify bryophytes were taken to laboratory for microscopic identification.

Bryophytes were identified to species except for thalloid liverworts of genus *Pellia*.

Stream insects

We sampled 60 near-pristine to pristine streams across three drainage basins in Finland (Grönroos *et al.*, 2013). Each drainage basin had 20 sampled streams. The drainage basins were Iijoki (centered on 65°N, 27°E), Koutajoki (centered on 66°N, 29°E) and Tenojoki (centered on 70°N, 27°E). The spatial extent comprising the three study areas was 87,101 km². Stream macroinvertebrates were sampled in the Iijoki drainage basin in the last week of May in 2009, in the Koutajoki drainage basin in the last week of May in 2008, and in the Tenojoki drainage basin in the second week of June in 2010. At each site, the field crew took a collective two-minute kicknet (net mesh size 0.3 mm) sample covering most microhabitats present in a riffle site (for details, see Grönroos *et al.* 2013). **In total, 79% of insects were identified to species in the laboratory.** Several riparian, in-stream habitat and water chemistry variables were also measured at each site (Grönroos *et al.*, 2013). Cover (%) of deciduous

trees was assessed in a 50-meter section on both banks upstream of each sampling site. Shading was estimated visually as percent canopy cover at the whole study section. Current velocity (at $0.6 \times \text{depth}$) and depth were measured at 30 random locations along cross-stream transects, the number of which depended on stream width. More transects were sampled in narrow than wide streams. Mean wetted width of each stream reach was measured based on five cross-stream transects. Macrophyte cover (%) and substratum particle class cover (%) were assessed at 10 random randomly-spaced $50 \text{ cm} \times 50 \text{ cm}$ plots. In addition, in each of the 10 plots, visual estimates of the percentage cover of five particle size classes were made based on a modified Wentworth scale (see Grönroos et al. 2013). Water samples were collected simultaneously with the field sampling and were measured for pH and conductivity.

Lake macrophytes

Macrophytes were sampled in 57 lakes with variable environmental conditions in three drainage basins in Finland (Alahuhta et al. 2015). In each of the Kymijoki (62°N , 26°E) and Vuoksi (63°N , 29°N) drainage basins, 20 lakes were surveyed, whereas 17 lakes were investigated in the Kokemäki drainage basin (62°N , 24°N). The spatial extent comprising the three study areas was $132,060 \text{ km}^2$. Lake macrophyte surveys were carried out during growing seasons between 2002 and 2008 in the Kymijoki and Vuoksi drainage basins, and between 2000 and 2011 in the Kokemäki drainage basin. Vascular plants, including helophytes and hydrophytes, were sampled using a main belt transect method, in which a varying number of 5-meter wide transects, depending on lake size, from the upper eulittoral to the outer limit of vegetation were examined. All macrophytes were identified to species. Ten hydro-morphological and water quality variables known to be important for aquatic plants were measured in each lake (Alahuhta et al. 2015). These variables consisted of lake

altitude, lake area, lake perimeter, alkalinity, turbidity, colour, Secchi depth, total phosphorus, total nitrogen and conductivity. Water quality variables represented median values for surface samples during the growing season over the period between 2000 and 2007 for the Kymijoki and Vuoksi drainage basins, and between 2000 and 2011 for the Kokemäenjoki drainage basin. Water quality data was obtained from the Hertta database maintained by the Finnish Environment Institute (www.environment.fi).

Lake zooplankton

Zooplankton were sampled from 60 small lakes in Finland during July in 2008 or 2009. The sites were sampled in three drainage basins. We sampled 20 lakes both in the Kokemäenjoki (61°N, 24°E) and the Kymijoki (63°N, 25°E) drainage basins in 2008, and 20 lakes in the Koutajoki (66°N and 29°E) drainage basin in 2009. These drainage basins were chosen because they cover a relatively large geographical extent and because the nutrient concentrations of lakes vary from ultraoligotrophic to eutrophic (Soininen et al. 2011). The spatial extent comprising the three study areas was 125,190 km². We sampled only small lakes to ensure that plankton sampling covers the site as adequately as possible. Most of the lakes within the drainage basins were not readily inter-connected to each other via water routes. For more information on the environmental characteristics of the lakes within the drainage basins, see Soininen et al. (2011). Plankton samples were collected with a tube sampler (V = 2.3 L) from three locations in the middle of the lake and pooled. We collected the samples in the middle of the lakes in order to avoid benthic taxa from the littoral entering the samples. The samples were collected at 0.5 m below the surface of the water. Zooplankton samples (6.15 L) were filtered through a 50 µm net and preserved with formaldehyde in the field. The maximum depth of the lakes as well as surface water

temperature was measured. The surface area of each lake was measured using Geographic Information System (MapInfo Version 6.5, MapInfo, Troy, NY, USA). Conductivity was measured in the field using a conductivity meter (Philips PW 9529). Samples for water chemistry analyses were collected simultaneously with the plankton sampling and analyzed in the laboratory for chlorophyll *a*, water colour, total nitrogen, and total phosphorus. In the laboratory, all zooplankton individuals were counted at magnification of 125-400× using an inverted microscope. Both crustacean zooplankton and rotifers were included in the counting. A total of 71% of zooplankton were identified to species.

Lake fishes

The lake fish data were based on postal inquiries sent to persons employed as chairmen or active members in regional fishing associations (Lappalainen and Malinen 2002). All fish were identified to species. The data were from three drainage basins: Vuoksi (centered on 63°N, 28°E), Kymijoki (centered on 62°N, 26°E) and Kokemäenjoki (centered on 61°N, 24°E). From each of the three drainage basins, 20 lakes were randomly selected for this study. The spatial extent comprising the three study areas was 150,869 km². The environmental data of the lakes were based on the Hertta database (www.environment.fi). The environmental variables available were lake area, length of the shoreline, altitude, maximum lake depth, conductivity, pH, colour and total phosphorus. Average values of water chemistry incorporating the whole water column for a period between June and September were calculated.

Spatial analysis

291 We used Moran's eigenvector maps to model spatial structures among the sites *within*
292 drainage basins and to provide spatial variables for community modelling (Borcard and
293 Legendre 2002; Legendre and Legendre 2012). We thus obtained multiple spatial variables
294 derived from geographical coordinates using Moran's eigenvector maps (MEM). These
295 variables describe spatial patterns in communities (Dray et al. 2012). In practice, this spatial
296 analysis generated orthogonal spatial variables based on information about geographic
297 coordinates, number of basins (i.e. blocks) and sites within each basin (Borcard et al. 2011;
298 Declerk et al. 2011). **Hence, as input data, we had site coordinates and indicated, in the R**
299 **script, which sites to belong to which basin. Otherwise, the MEM analysis resembles that of**
300 **the original MEM analysis without blocks.** These multiple spatial variables describe within-
301 basin spatial structures in the data, such that the sites in the other two basins get zero values
302 when the spatial structures within a focal basin are considered (Declerck et al. 2011; Silva
303 and Hernández 2015). This analysis results in a staggered matrix of MEM eigenvectors, i.e.,
304 within-region spatial variables. These variables are efficient in modelling spatial structures of
305 community composition data at multiple scales within each basin. Large-scale spatial
306 structures among drainage basins were modelled by a dummy variable "basin identity"
307 because Moran eigenvector maps do not work well when there are large gaps between
308 regions, such as those between our drainage basins. The Moran's eigenvector maps analysis
309 was run using the function "create.MEM.model" (see Declerk et al. 2011).

310 Given the facts that we had three regions in the analysis and that not all lakes were
311 connected by streams, we could not use more sophisticated methods taking into account
312 hydrological connections among sites (Blanchet et al. 2008; Borcard et al. 2011; Liu et al.
313 2013). Also, it has been previously shown for stream organisms that MEM eigenvectors
314 based on either overland or watercourse distances between sites provide similar information
315 about spatial effects on community structure (Landeiro et al. 2011; Grönroos et al. 2013).

Finally, it should be noted that the spatial component in variation partitioning analyses should be considered with certain caution (Gilbert and Bennet 2010; Smith and Lundholm 2010), and that large gradients in community composition may be challenging to model because of multiple changes in community composition (Tuomisto et al. 2012).

Forward selection of explanatory variables

The final sets of environmental and spatial variables were separately selected for the redundancy analysis (RDA; Rao 1964) models using a forward selection procedure with two stopping rules (Blanchet et al. 2008). Forward selection proceeds only if the global model, which is tested first, is significant. The first stopping rule entails exceeding the critical p-value ($\alpha=0.05$), and the second is related to the reduced model adjusted R^2 value exceeding that of the global model.

Variation partitioning

We used the raw data approach (i.e. site-by-species matrix as response) to examine variation in community structure among sites (Legendre et al. 2005; Anderson et al. 2011) in each of the six datasets. Each dataset comprised all sites in all three drainage basins. We used redundancy analysis (RDA, Rao 1964) to analyse variation in presence-absence data, as comparable abundance data were not available for all organismal groups. RDA examines variation in species composition (**Y**) in relation to sets of predictor variables that were in our present study environmental variables (**E**), spatial variables (**S**) derived from Moran's

eigenvector map analysis (see above) and dummy variable basin (**B**). Prior to the RDA, site-by-species presence-absence data were Hellinger-transformed to make the data better analysable using linear methods (Legendre and Gallagher 2001) and because Hellinger transformation/Hellinger distance was deemed highly suitable for community composition data in a comparative analysis (Legendre and de Caceres 2013). The Hellinger transformation consists in transforming the site-by-species data into relative values per site, by dividing each value by the site sum, then taking the square root of the resulting values. Hellinger transformation can be used for both presence-absence and abundance data (Legendre et al. 2005). We selected significant variables in the final RDA models of each set of variables (**E** or **S**) using the forward selection method with two stopping rules (Blanchet et al. 2008) with the function “ordiR2step” in the R package vegan (Oksanen et al. 2013). We used redundancy analysis (RDA) to partition variation in species composition (**Y**) between **E**, **S** and **B** following the widely-used variation partitioning approach (Borcard et al. 1992; Legendre and Legendre 2012). Variation partitioning of species composition (**Y**) between three sets of predictor variables results in pure environment ($\mathbf{E} \mid \mathbf{S}+\mathbf{B}$), pure spatial ($\mathbf{S} \mid \mathbf{E}+\mathbf{B}$) and pure basin ($\mathbf{B} \mid \mathbf{E}+\mathbf{S}$) fractions, as well as their shared effects and unexplained variance (**U**). In many cases, spatial variables were not significant, and we thus ran the variation partitioning between **E** and **B** only. Variation partitioning was run using the function “varpart” in the R package vegan. We reported adjusted R^2 values in all analyses because they are unbiased estimates of explained variation (Peres-Neto et al. 2006). We also tested for the significance of the total **E**, **S** and **B** fractions, and pure fractions $\mathbf{E} \mid \mathbf{S}+\mathbf{B}$, $\mathbf{S} \mid \mathbf{E}+\mathbf{B}$ and $\mathbf{B} \mid \mathbf{E}+\mathbf{S}$ using the function “anova” in the package vegan. Ecologically, we expected that **E** would be related to local environmental control, **S** to within-basin spatial dynamics and **B** to biogeographic effects.

Visual inspection of breakpoints in metacommunity structure

We also ran Principal Components Analysis (PCA) based on Hellinger-transformed presence-absence data for each organism group to visually examine breakpoints in community structure. If those breakpoints are related to among-region differences, they should be easily detectable by plotting regions using different symbols. In contrast, if the breakpoints are related to gradients in local environmental factors, they should not be related to regions. PCA based on Hellinger-transformed data was chosen among the various unconstrained ordination methods to retain comparability with the variation partitioning in RDA (see above).

Testing for different metacommunity structures

Elements of metacommunity structure (EMS) analyses were based on instructions given in Leibold and Mikkelsen (2002) and Presley et al. (2010). We followed the “range perspective” in our EMS analyses (Leibold and Mikkelsen 2002). Below, we describe the flow of analyses following previous studies (Leibold and Mikkelsen 2002; Heino et al. 2015b, 2015c).

The EMS analysis is based on three metrics: coherence, turnover and boundary clumping. In this analysis, prior to calculating those metrics, a raw data site-by-species presence-absence matrix for each organismal group was ordinated using reciprocal averaging (Leibold and Mikkelsen, 2002). Using this ordination method, the sites having similar species composition occur close to each other and the species that have similar occurrence among the sites are located close to each other along an axis (Gauch 1982). Reciprocal averaging

analyses may be sensitive to very rare species, and we thus ran the EMS analysis for each organismal group with either all species or without singletons (i.e. species occurring at a single site only were removed prior to the analysis).

Coherence is based on calculating the number of embedded absences (EAbs) in the ordinated matrix and then comparing the observed value to a null distribution of embedded absences (i.e. gaps in a species range) from simulated matrices (Leibold and Mikkelsen 2002; Presley et al. 2010). A small number of embedded absences (i.e. EAbs is significantly lower than expected by chance) mean positive coherence, whereas a large number of embedded absences (i.e. EAbs is significantly larger than expected by chance) mean negative coherence. Significantly negative coherence thus suggests a checkerboard distribution of species, non-significant coherence refers to randomness, and significantly positive coherence refers to nestedness, evenly-spaced gradients, Gleasonian structure or Clementsian structure (Leibold and Mikkelsen 2002). Turnover is evaluated only if coherence is positive. Turnover is measured as the number of times one species replaces (Rep) another between two sites in an ordinated matrix (Presley et al. 2010). Significant negative turnover (i.e. Rep is significantly lower than expected by chance) refers to nestedness, whereas significantly positive turnover (i.e. Rep is significantly larger than expected by chance) refers to evenly-spaced, Gleasonian or Clementsian structures (Leibold and Mikkelsen 2002). Furthermore, the cases of significant positive coherence and non-significant turnover can be regarded as quasi-structures (Presley et al. 2010). The evenly-spaced, Gleasonian and Clementsian metacommunity structures can be separated based on boundary clumping (Leibold and Mikkelsen 2002). This metric is assessed using Morisita's dispersion index and a subsequent chi-square test comparing observed and expected distributions of range boundary locations. Index values significantly less than 1 indicate hyperdispersed range boundaries (i.e. evenly-spaced metacommunity structure), values that are not different from 1 indicate randomly

distributed range boundaries (i.e. Gleasonian metacommunity structure), values significantly greater than 1 indicate clumped range boundaries (i.e. Clementsian metacommunity structure). Similarly, Quasi-evenly-spaced, Quasi-Gleasonian and Quasi-Clementsian metacommunity structures can be separated by boundary clumping (Presley et al. 2010).

We assessed the significance of the index values for coherence (EAbs) and turnover (Rep) using the fixed-proportional null model, where row sums are fixed (i.e. the species richness of each site was maintained), but column marginal frequencies (i.e. species frequencies of occurrence) were considered probabilities. Random matrices were produced using the “r1” method for the fixed-proportional null model as implemented in the R package *vegan* (Oksanen et al. 2013). This method is the default in the R package we used (Dallas 2013), and it has been previously used in several other studies (e.g. Heino et al. 2015c). Although a stricter fixed-fixed null model might provide slightly different results from those of fixed-proportional null model, such fixed-fixed null model was not used because it is overly conservative and because we could not have then compared our results with those from single drainage basins (e.g. Heino et al. 2015e). We used 999 simulations to provide simulated matrices. Statistical significance of EAbs or Rep was subsequently estimated by comparing the observed index value from the original matrix to the distribution of values derived from the 999 simulated matrices. Metacommunity structure was examined for each organismal group based on axis 1 of reciprocal averaging because we were interested in the most important species compositional gradient (Gauch 1982). All EMS analyses were run using the R package *metacom* (Dallas 2013), with the “r1” method borrowed from the R package *vegan* (Oksanen et al. 2013).

Quantifying beta diversity

We quantified beta diversity following the ideas presented by Baselga (2010). We thus partitioned total beta diversity (i.e. multiple-site beta diversity based on Sørensen coefficient) in each data set (i.e. across three drainage basins) into turnover (i.e. multiple-site beta diversity based on Simpson coefficient) and nestedness components (i.e. that resulting from nestedness-related species richness differences among sites) using the function “nestedbetasor” in the R package vegan (Oksanen et al. 2013).

Results

Overall, our results showed that basin identity and local environmental variables were significant predictors of variation in community structure, whereas within-basin spatial effects were typically negligible (Table 1). In half of the cases (diatoms, bryophytes, zooplankton), basin identity was a slightly better predictor of community structure compared with local environmental variables, whereas the opposite was true for the remaining three organismal groups (insects, macrophytes, fish). Both pure basin and local environmental fractions were, however, significant after accounting for the effects of the other predictor variable set ($p < 0.05$). Only for lake macrophytes were pure within-region spatial effects significant, but their pure effects were slightly smaller than those for local environmental and basin variables. All three pure components (i.e., pure environmental, pure within-region spatial and pure region identity components) were significant for macrophytes ($p < 0.05$). Much of the explained variation was shared between the two or three predictor variable sets. Also, our overall RDA models explained only a small fraction of variation in community structure, varying between 10 and 20 %. Total variation explained or pure environmental

fraction were not significantly related to spatial extent of the entire study regions (Spearman rank correlation, $p > 0.200$). The environmental variables selected in the RDA models were those which are often influential in aquatic metacommunity studies (Table 1). For streams, the most common environmental variable, occurring in all models, was water pH. For lakes, lake area was selected in the models of all organismal groups.

Of the organismal groups, diatoms, insects and zooplankton showed clear regional differences in two-dimensional PCA ordination plots (Fig. 3). This suggest that breakpoints in community composition are mainly related to among-region differences. In contrast, bryophytes, macrophytes and fish showed less clear regional separation of community composition, suggesting that potential breakpoints were related to variations in local environmental factors (Fig. 3).

All organismal groups showed high levels of beta diversity irrespective of the levels of gamma diversity and mean alpha diversity (Table 2). High beta diversity was largely attributable to the turnover component, whereas the nestedness component was rather high only for fishes. Such high levels of beta diversity were also reflected in coherent metacommunity structures, higher turnover than expected by chance and clear boundary clumping (Table 3). Hence, in the majority of the cases, the datasets fitted best with Clementsian metacommunity structures, with Quasi-Clementsian structures being found only for stream bryophytes and lake zooplankton (Table 3). It was notable that the beta diversity measures or the EMS analysis were not sensitive to the exclusion of rare taxa (i.e. when singletons were removed from the analyses) (Tables 2 and 3).

Discussion

We examined three sources of variation in community structure, namely drainage basin effect, spatial effect and environmental effect, which can be translated into mechanisms related to biogeography, spatial dynamics and environmental filtering, respectively. We found that (1) mainly basin identity and local environmental factors were significant determinants of community structure in all organismal groups, whereas spatial relationships between sites were influential only for lake macrophytes. We also observed (2) that all organismal groups showed high beta diversity, turnover component in particular, **across the basins (this study) and within each basin (Heino et al. 2015c)**, and (3) fitted best with Clementsian structures.

Determinants of community structure of aquatic organisms

Environmental control often dominates over all spatial effects on metacommunity organization (Cottenie 2005; Soininen 2014). We found support for this expectation in three of the six organismal groups (i.e. insects, macrophytes, fish), which corroborates many findings from streams (Landeiro et al. 2012; Grönroos et al. 2013; Alahuhta et al. 2015) and lakes (Cottenie et al. 2003; Alahuhta and Heino 2013; Heino 2013). These findings suggest that environmental filtering is the main mechanism structuring metacommunities (Cottenie 2005; Van der Gucht et al. 2007), at least if the spatial extent of a region under study is not very broad (Mykrä et al. 2007; Heino et al. 2015a). Although the maximum spatial extent in our datasets was more than 150,000 km², we did not find that basin identity (i.e. the biogeographical effect) would overcome the effect of local environmental factors on the community structure of insects, macrophytes or fish. This finding may be due to two main reasons. First, environmental ranges typically increase with increasing spatial extent, thus

providing more scope for environmental filtering provided that dispersal remains adequate (e.g. Soininen 2014). Second, in lowland regions, such as our present study area, different drainage basins may harbour rather similar biotas. This result is in contrast with findings from more topographically separated drainage basins (e.g. Hoeinghaus et al. 2007). Such small differentiation in regional faunas or floras between our drainage basins leads to apparent patterns that mainly environmental filtering drives variation in local community structure of insects, macrophytes and fish. This also suggests that biogeographic effects, such as historical influences and climatic forcing, have rather minor effects on local aquatic communities in lowland regions.

We expected that environmental conditions would overcome the effects of within-region spatial structuring. We found clear support for this expectation for the three stream organismal groups, but lake macrophytes showed significant spatial structuring along with significant environmental effects. This might result from stronger dispersal limitation in lake organisms compared to stream organisms. However, despite being significant, spatial effects on lake macrophytes were minor at best, supporting the role of environmental filtering in driving variation in community structure. Similar studies conducted across multiple regions have found corresponding results, whereby within-region spatial effects are less important than environmental control (Declerck et al. 2011; De Bie et al. 2012; Viana et al. 2015). It is interesting to note that our study regions were of intermediate size in comparison to Declerck et al.'s (2011) wetland pond study and Viana et al.'s (2015) lake study that extended over large regions in most of western Europe, and that our findings were rather similar to those studies. This suggests some similarities across broad spatial scales when there are multiple separate regions under study.

We also predicted that basin identity would overcome the effects of local environmental factors and spatial relations within drainage basins. This prediction proved to

be partly correct. Although basin effects were significant for all organismal groups, the amount of explained variation of pure basin effect was higher than that of local environmental variables for diatoms, bryophytes and zooplankton. It is possible that such region effects become even more important with increasing spatial extents, and previous findings at a large spatial extent have found similar effects on lake macrophytes and zooplankton (Viana et al. 2015). Our result that diatoms, bryophytes and zooplankton showed stronger basin effects than environmental effects is surprising, however, because small passively dispersing organisms or their tiny propagules should be able to follow variation in local environmental variables and cross drainage basin boundaries easily (Kristiansen 1996; De Bie et al. 2012). It is hence likely that some unmeasured, yet potentially influential environmental variables (e.g. temperature or geology) vary between the drainage basins, which translated into basin effect on community structure for diatoms, bryophytes and zooplankton. Moreover, Alahuhta et al. (2016) found that melting of ice sheet after the last ice age created variable local environmental conditions along even modest altitudinal gradient, further affecting present-day community composition. However, it would be very difficult to examine those effects further with the present data, as climatic, geological and historical (e.g. time since glaciation) conditions vary clearly among the basins, but are clearly less variable or not measurable within each basin. This means that those effects are hardly discernible from the effects of basin identity on aquatic communities.

We expected that lake organisms should be more dispersal limited than stream organisms, and thus the former should show more spatial structuring than the latter. This finding was partly supported, as none of the stream organismal groups exhibited significant within-region spatial structuring, whereas lake macrophytes and fish showed significant spatial structuring. This finding largely corroborates previous findings, where spatial structuring within small drainage basins is often negligible for headwater stream organisms

(Heino et al. 2012; Landeiro et al. 2012), whereas statistically significant spatial structuring has been found for some groups of lake organisms (Beisner et al. 2006; Heino 2013). This pattern may be due to the fact that stream systems are more connected than lake systems, which results in differences in the likelihood of dispersal limitation between lotic and lentic systems. However, some studies have found that species sorting through environmental heterogeneity among sites drives variation in the community structure of both riverine and lake macrophytes, whereas spatial effects are negligible (Alahuhta et al. 2015). These discrepancies in findings may be related to differences in spatial extent and the connectivity between the sites actually used in the analyses.

Dispersal may also potentially account for biogeographical variation in community structure. We **hypothesised** that small passively dispersed organisms would surpass all geographical barriers and would thus show no evidence of basin identity, whereas the opposite should be true for large actively dispersing organisms. As a related **hypothesis**, we expected small passive dispersers to show stronger environmental control (De Bie et al. 2012). We found at best little support for these conjectures, as all organismal groups showed a significant pure region effect, and pure environmental effects did not vary consistently between the passive and active dispersers. While such region effects might potentially be related to limited dispersal between the three regions (Viana et al. 2015), they may equally likely arise from climatic forcing on species distributions. However, as already indicated, it is almost impossible to disentangle overall basin effects and present-day climate or historical dispersal on our results because climate varies clearly *among* the drainage basins, but is largely invariable among sites *within* our small and predominantly lowland drainage basins.

Low explanatory power was common for the environmental, spatial and basin models. This is a typical finding in most freshwater bacterial (e.g. Souffreau et al. 2015), phytoplankton (e.g. Nabout et al. 2009), insect (e.g. Heino et al. 2015d), macrophyte (e.g.

Alahuhta and Heino 2013) and fish (e.g. Beisner et al. 2006) metacommunity studies based on adjusted coefficient of determination (Peres-Neto et al. 2006) and presence-absence data (Vilmi et al. 2016). There are at least **five** reasons for the low amount of explained variation: (1) there are influential missing environmental variables, (2) modelling of dispersal routes and rates is inadequate, or (3) variation in community structure just happens to be difficult to explain owing to various deterministic and stochastic factors varying simultaneously (Heino et al. 2015d). **Also, (4) very low amounts of explained variation could simply emerge by chance (T. Dallas, pers. com.). Finally, (5) the low amounts of variation explained might be related to methodological difficulties in modelling high beta diversity in a dataset, which may be due to multiple turnovers in species composition (see Tuomisto et al. 2012).** Despite these low amounts of explained variation, we could test our hypotheses about the relative roles of environmental control, spatial effects and region constraints by basing conclusions on significance testing and additional evidence on beta diversity and metacommunity structures.

Elements of metacommunity structure

Many previous studies on metacommunity structures using the EMS analysis have found highly variable patterns, varying from random through nested to Gleasonian (i.e. where species show individualistic responses to ecological gradients) and Clementsian (i.e. where subgroups of species show similar responses to ecological gradients) structures (Leibold and Mikkelsen 2002; Presley and Willig 2010; Presley et al. 2012; Dallas and Drake 2014; Heino et al. 2015c). However, many of these studies have focused on metacommunities within small regions, which might increase variability in the results (Heino et al. 2015c).

In the present study, we expected that Clementsian gradients would emerge when metacommunity structures were analysed across multiple drainage basins because this potentially means crossing multiple species pools and covering large environmental gradients. On the one hand, different species pools should result in different local communities, owing to a strong regional influence on local community structure (Heino et al. 2003; Soininen et al. 2009). On the other hand, environmental gradient lengths should increase with increasing spatial extent, resulting in stronger species composition-environment correlations among sites (Vetaas and Chaudhary 1998; Soininen 2014). Although we cannot decisively distinguish between the two main drivers of Clementsian structures because, for example, different organismal groups may show different regional vs local environmental influences on breakpoints in community composition (Fig. 3), we found strong support for such Clementsian structures. This finding is similar to those in a study of bat faunas on Caribbean islands (Presley and Willig 2010), a study of beetle faunas over northern European biogeographical provinces (Heino and Alahuhta 2015), a study on riverine invertebrates of two central German drainage basins (Tonkin et al. 2015b) and a study on wetland crustacean communities in Spanish wetlands (Gascón et al. 2016). However, this finding partly disagrees with studies conducted within small regions, including the individual drainage basins incorporated in this study. Heino et al. (2015c) found that the stream and lake metacommunities of individual drainage basins showed variable (i.e. random, nested, Gleasonian, Clementsian and quasi structures) metacommunity patterns (Table 4). This suggested some degree of scale-dependency in metacommunity structures. We hence propose that Clementsian structures are common in large-scale studies of local communities, i.e., a combination of small-grained data with broad spatial extents (Beck et al. 2012; Bini et al. 2014; Dallas and Drake 2014), whereas various patterns may be detected in small-scale studies (Heino et al. 2015e; Tonkin et al. 2015). Further indirect support for Clementsian

structures at large scales was provided by the high levels of beta diversity, turnover component in particular. This is because it is likely that high turnover results in Clementsian gradients rather than, for example, nested structures (Heino et al. 2015c).

Conclusions

The simple yet heuristic approach we used here is easily adaptable to situations where there are two spatial scales and two or more individual regions (e.g. drainage basins), providing a useful starting point for more sophisticated analyses of variation in community structure. We suggest that by analysing simultaneously three sources of variation, environmental (E), spatial (S) and basin effects (B), we can reveal interesting patterns and suggest some underlying processes for variation in metacommunity organization across broad biogeographic regions. Our findings also increase understanding of biogeographical patterns of community structure in aquatic environments by combining **beta diversity analysis with multivariate models (i.e. variation partitioning) and general ecological pattern detection (i.e. the EMS analysis)**. Indeed, our findings strongly suggest that aquatic organisms typically show high levels of beta diversity and Clementsian gradients at broad spatial extents even when the focus is on local aquatic communities.

Acknowledgements. Financial support was provided by the Academy of Finland (to JH, JS). We thank Tad Dallas, **Jamie Kneitel and an anonymous reviewer** for comments on an earlier version of this manuscript.

References

- Alahuhta J, Heino J (2013) Spatial extent, regional specificity and metacommunity structuring in lake macrophytes. *J Biogeogr* 40:1572-1582
- Alahuhta, J, Hellsten S, Kuoppala, M, Riihimäki, J (2016) Regional and local determinants of macrophyte community compositions in high-latitude lakes of Finland. *Hydrobiologia*, in press
- Alahuhta J, Rääpysjärvi J, Hellsten S, Kuoppala M, Aroviita J (2015) Species sorting drives variation of boreal lake and river macrophyte communities. *Comm Ecol* 16:76-85
- Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders N., Cornell HV, Comita LS, Davies KF, Harrison SP, Kraft NJB, Stegen JC, Swenson NG (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol Lett* 14:19–28
- Baselga A (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecol Biogeogr* 19:134–143
- Beisner BE, Peres-Neto PR, Lindström E, Barnett A, Longhi ML (2006) The role of dispersal in structuring lake communities from bacteria to fish. *Ecology* 87:2895-2991
- Beck J, Ballesteros-Mejia L, Buchmann CM, Dengler J, Fritz SA, Gruber B, Hof C, Jansen F, Knapp S, Kreft H, Schneider A-K, Winter M, Dormann CF (2012) What's on the horizon for macroecology? *Ecography* 35:673–683

675 Bini LM, Landeiro VL, Padial AA, Siqueira T, Heino J (2014) Nutrient enrichment is related
 676 to two facets of beta diversity for stream invertebrates across the United States.
 677 Ecology 95:1569-1578

678 Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables.
 679 Ecology 89:2623–2632

680 Borcard D, Gillet F, Legendre P (2011) Numerical Ecology with R. Springer, New York.

681 Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of
 682 principal coordinates of neighbour matrices. Ecol Model 153:51–68

683 Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological
 684 variation. Ecology 73:1045–1055

685 Clements FE (1936) Nature and Structure of the Climax. J Ecol 24:252-284

686 Cottenie K (2005) Integrating environmental and spatial processes in ecological community
 687 dynamics. Ecol Lett 8:1175-1182

688 Dallas T (2013) metacom: Analysis of the “elements of metacommunity structure”. R
 689 package version 1.2. <http://CRAN.R-project.org/package=metacom>

690 Dallas T, Drake JM (2014) Relative importance of environmental, geographic, and spatial
 691 variables on zooplankton metacommunities. Ecosphere 5:art104

692 da Silva PG, Hernández MIM (2015) Scale-dependence of processes structuring dung beetle
 693 metacommunities using functional diversity and community deconstruction
 694 approaches. PLoS ONE 10:e0123030

695 De Bie T, De Meester L, Brendonck L, Martens K, Goddeeris B, Ercken D, Hampel H,
 696 Denys L, Vanhecke L, Van der Gucht K, Van Wichelen J, Vyverman W,

697 Declerck SAJ (2012) Body size and dispersal mode as key traits determining
698 metacommunity structure of aquatic organisms. *Ecol Lett* 15:740–747

699 Declerck SAJ, Coronel JS, Legendre P, Brendonck L (2011) Scale dependency of processes
700 structuring metacommunities of cladocerans in temporary pools of High-Andes
701 wetlands. *Ecography* 34:296–305

702 Diniz-Filho JAF, Bini LM (2005) Modelling geographical patterns in species richness using
703 eigenvector-based spatial filters. *Glob Ecol Biogeogr* 14:177–185

704 Dray S, Péliissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, Bellier E, Bivand R,
705 Blanchet FG, De Cáceres M, Dufour AB, Heegaard E, Jombart T, Munoz F,
706 Oksanen J, Thioulouse J, Wagner HH (2012) Community ecology in the age of
707 multivariate multiscale spatial analysis. *Ecol Monogr* 82:257–275

708 Gascón, S, Cañedo-Argüelles M, Nebra A, Ruhí A, Rieradevall M, Caiola N, Sala J, Ibàñez
709 C, Quintana XD, Boix D (2016) Environmental filtering determines
710 metacommunity structure in wetland microcrustaceans. *Oecologia*, in press.

711 Gauch HG (1982) *Multivariate Analysis in Community Ecology*. Cambridge University
712 Press, Cambridge

713 Gilbert B, Bennett JR (2010) Partitioning variation in ecological communities: do the
714 numbers add up? *J Appl Ecol* 47:1071–1082

715 Gonçalves-Souza T, Romero GQ, Cottenie K (2014) Metacommunity versus biogeography:
716 A case study of two groups of neotropical vegetation-dwelling arthropods.
717 *PLoS ONE* 9:e115137

718 Göthe E, Angeler DG, Sandin L (2013) Metacommunity structure in a small boreal stream
719 network. *J Anim Ecol* 82:449–458

720 Grönroos M, Heino J, Siqueira T, Landeiro VL, Kotanen J, Bini LM (2013) Metacommunity
721 structuring in stream networks: roles of dispersal mode, distance type and
722 regional environmental context. *Ecol Evol* 3:4473–4487

723 Hawkins BA, Porter EE (2003) Water–energy balance and the geographic pattern of species
724 richness of western Palearctic butterflies. *Ecol Entomol* 28:678–686

725 Heino J (2013) Does dispersal ability affect the relative importance of environmental control
726 and spatial structuring of littoral macroinvertebrate communities? *Oecologia*
727 171:971-980

728 Heino J, Alahuhta J (2015) Elements of regional beetle faunas: faunal variation and
729 compositional breakpoints along climate, land cover and geographical
730 gradients. *J Anim Ecol* 84:427–441

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

733 Heino J, Melo AS, Bini LM, Altermatt F, Al-Shami SA, Angeler D, Bonada N, Brand C,
734 Callisto M, Cottenie K, Dangles O, Dudgeon D, Encalada A, Göthe E,
735 Grönroos M, Hamada N, Jacobsen D, Landeiro VL, Ligeiro R, Martins RT,
736 Miserendino ML, Md Rawi CS, Rodrigues M, Roque FO, Sandin L, Schmera
737 D, Sgarbi LF, Simaika J, Siqueira T, Thompson RM, Townsend CR (2015d) A
738 comparative analysis reveals weak relationships between ecological factors and
739 beta diversity of stream insect metacommunities at two spatial levels. *Ecol Evol*
740 5:1235-1248

741 Heino J, Melo AS, Siqueira T, Soininen J, Valanko S, Bini LM (2015a) Metacommunity
742 organisation, spatial extent and dispersal in aquatic systems: patterns, processes
743 and prospects. *Freshw Biol* 60:845-869

744 Heino J, Muotka, T, Paavola R (2003) Determinants of macroinvertebrate diversity in
745 headwater streams: regional and local influences. *Journal of Animal Ecology*
746 72:425-434

747 Heino J, Nokela T, Soininen J, Tolkkinen M, Virtanen L, Virtanen R (2015e) Elements of
748 metacommunity structure and community-environment relationships in stream
749 organisms. *Freshw Biol* 60:973-988

750 Heino J, Soininen J, Alahuhta J, Lappalainen J, Virtanen R (2015c) A comparative analysis
751 of metacommunity types in the freshwater realm. *Ecol Evol* 5:1525-1537

752 Hoenighaus DJ, Winemiller KO, Birnbaum JS (2007) Local and regional determinants of
753 stream fish assemblage structure: inferences based on taxonomic vs. functional
754 groups. *J Biogeogr* 34:324–338

755 Hortal J, Diniz-Filho JAF, Bini LM, Rodríguez MÁ, Baselga A, Nogués-Bravo D, Rangel
756 TF, Hawkins BA, Lobo JM (2011) Ice age climate, evolutionary constraints and
757 diversity patterns of European dung beetles. *Ecol Lett* 14:741–748

758 Jenkins DG, Ricklefs RE (2011) Biogeography and ecology: two views of one world. *Proc R*
759 *Soc Lond Ser B* 366:2331-2335

760 Jocque M, Field R, Brendonck L, de Meester L (2010) Climatic control of dispersal –
761 ecological specialization trade-offs: a metacommunity process at the heart of
762 the latitudinal diversity gradient? *Glob Ecol Biogeogr* 19:244–252

763 Kristiansen J (1996) Dispersal by freshwater algae – a review. *Hydrobiologia* 336:151–157

764 Lappalainen J, Malinen T (2002) Effects of area and location on pikeperch yields in Finnish
765 lakes. In: Cowx IG (ed) *Management and Ecology of Lake and River Fisheries*.
766 Blackwell. pp. 35-45

767 Landeiro VL, Bini LM, Melo AS, Pes AMO, Magnusson WE (2012) The roles of dispersal
768 limitation and environmental conditions in controlling caddisfly (Trichoptera)
769 assemblages. *Freshw Biol* 57:1554–1564

770 Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the
771 spatial variation of community composition data. *Ecol Monogr* 75:435-450

772 Legendre P, De Cáceres M (2013) Beta diversity as the variance of community data:
773 dissimilarity coefficients and partitioning. *Ecol Lett* 16:951-963

774 Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of
775 species data. *Oecologia* 129:271-280

776 Legendre P, Legendre L (2012) *Numerical Ecology*. Third Edition. Elsevier, Amsterdam

777 Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD,
778 Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A (2004) The
779 metacommunity concept: a framework for multi-scale community ecology. *Ecol*
780 *Lett* 7:601-613

781 Leibold MA, Mikkelsen GM (2002) Coherence, species turnover, and boundary clumping:
782 elements of meta-community structure. *Oikos* 97:237–250

783 Liu J, Soininen J, Han B-P, Declerck SAJ (2013) Effects of connectivity, dispersal
784 directionality and functional traits on the metacommunity structure of river
785 benthic diatoms. *J Biogeogr* 40:2238-2248

786 Logue JB, Mouquet N, Peter H, Hillebrand H, The Metacommunity Working Group (2011)
787 Empirical approaches to metacommunities: a review and comparison with
788 theory. *Trends in Ecology and Evolution* 26:482-491

789 Meynard CN, Lavergne S, Boulangeat I, Garraud L, Van Es J, Mouquet N, Thuiller W (2013)
790 Disentangling the drivers of metacommunity structure across spatial scales. *J*
791 *Biogeogr* 40:1560-1571

792 Mouquet N, Loreau M (2003) Community patterns in source-sink metacommunities. *Am Nat*
793 162:544-557

794 Mykrä H, Heino J, Muotka T (2007) Scale-related patterns in the spatial and environmental
795 components of stream macroinvertebrate assemblage variation. *Glob Ecol*
796 *Biogeogr* 16:149–159

797 Nabout JC, Siqueira T, Bini LM, Nogueira IS (2009) No evidence for environmental and
798 spatial processes in structuring phytoplankton communities. *Acta Oecologica*
799 35:720-726

800 Ng ISY, Carr C, Cottenie K (2009) Hierarchical zooplankton metacommunities:
801 distinguishing between high and limiting dispersal mechanisms. *Hydrobiologia*
802 619:133-143

803 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL,
804 Solymos P, Stevens MHH, Wagner H (2013) *vegan: Community Ecology*
805 *Package*. R package version 2.0-10. <http://CRAN.R-project.org/package=vegan>

806 Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data
807 matrices: estimation and comparison of fractions. *Ecology* 87:2614–2625

808 Pinel-Alloul B, André A, Legendre P, Cardille JA, Patalas K, Salki A (2013) Large-scale
809 geographic patterns of diversity and community structure of pelagic crustacean
810 zooplankton in Canadian lakes. *Glob Ecol Biogeogr* 22:784–795

811 Pinel-Alloul B, Niyonsenga T, Legendre P (1995) Spatial and environmental components of
812 freshwater zooplankton structure. *Écoscience* 2:1–19

813 Presley SJ, Cisneros LM, Patterson BD, Willig MR (2012) Vertebrate metacommunity
814 structure along an extensive elevational gradient in the tropics: a comparison of
815 bats, rodents and birds. *Glob Ecol Biogeogr* 21:968-976

816 Presley SJ, Higgins CL, Willig MR (2010) A comprehensive framework for the evaluation of
817 metacommunity structure. *Oikos* 119:908-917

818 Presley SJ, Willig MR (2010) Bat metacommunity structure on Caribbean islands and the
819 role of endemics. *Glob Ecol Biogeogr* 19:185-199

820 Rao CR (1964) The use and interpretation of principal component analysis in applied
821 research. *Sankhyā: The Indian J Stat Ser A* 26:329–358

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

824 Ricklefs RE, Jenkins DG (2011) Biogeography and ecology: towards the integration of two
825 disciplines. *Proc R Soc Lond Ser B* 366:2438–2448

826 Rundle SD, Bilton DT, Foggo A (2007) By wind, wings or water: body size, dispersal and
827 range size in aquatic invertebrates. In: Hildrew AG, Raffaelli DG, Edmonds-
828 Brown R (eds) *Body size: The Structure and Function of Aquatic Ecosystems*,
829 pp. 186-209. Cambridge University Press, Cambridge

830 Shurin JB, Cottenie K, Hillebrand H (2009) Spatial autocorrelation and dispersal limitation in
831 freshwater organisms. *Oecologia* 159:151-159

832 Smith TW, Lundholm JT (2010) Variation partitioning as a tool to distinguish between niche
833 and neutral processes. *Ecography* 33:648-655

834 Soininen J (2014) A quantitative analysis of species sorting across organisms and
835 ecosystems. *Ecology* 95:3284-3292

836 Soininen J, Heino J, Kokocinski M, Muotka T (2009) Local-regional diversity relationship
837 varies with spatial scale in lotic diatoms. *J Biogeogr* 36:720-727

838 Soininen J, Korhonen JJ, Karhu J, Vetterli A (2011) Disentangling the spatial patterns in
839 community composition of prokaryotic and eukaryotic lake plankton. *Limnol*
840 *Oceanogr* 56:508-520

841 Soininen J, McDonald R, Hillebrand H (2007) The distance decay of similarity in ecological
842 communities. *Ecography* 30:3-12

843 Souffreau C, Van der Gucht K, van Gremberghe I, Kosten S, Lacerot G, Lobão LM, de
844 Moraes Huszar VL, Roland F, Jeppesen E, Vyverman W, De Meester L (2015)
845 Environmental rather than spatial factors structure bacterioplankton
846 communities in shallow lakes along a > 6000 km latitudinal gradient in South
847 America. *Env Microbiol* 17:2336–2351

848 Tonkin JD, Stoll S, Jähnig SC, Haase P (2015) Contrasting metacommunity structure and
849 beta diversity in a river-floodplain system. *Oikos* in press

850 Tonkin JD, Sundermann A, Jähnig SC, Haase P (2015) Environmental controls on river
851 assemblages at the regional scale: an application of the Elements of
852 Metacommunity Structure framework. *PLOS ONE* 10:e0135450

853 Tuomisto H, Ruokolainen L, Ruokolainen K (2012) Modelling niche and neutral dynamics:
854 on the ecological interpretation of variation partitioning results. *Ecography*
855 35:961-971

856 Van der Gucht K, Cottenie K, Muylaert K, Vloemans N, Cousin S, Declerck S, Jeppesen E,
857 Conde-Porcuna J-M, Degans H, Vyverman W, De Meester L (2007) The power
858 of species sorting: local factors drive bacterial community composition over a
859 wide range of spatial scales. *Proc Natl Acad Sci USA* 104:20404-20409

860 Vetaas OR, Chaudhary R. (1998) Scale and species-environment relationships in a central
861 Himalayan oak forest, Nepal. *Plant Ecol* 134:67-76

862 Viana DS, Figuerola J, Schwenk K, Manca M, Hobæk A, Mjelde M, Preston CD, Gornall RJ,
863 Croft JM, King RA, Green AJ, Santamaría L (2015) Assembly mechanisms
864 determining high species turnover in aquatic communities over regional and
865 continental scales. *Ecography* in press

866 Vilmi, A, Karjalainen, SM, Hellsten, S, Heino, J (2016) Bioassessment in a metacommunity
867 context: are diatom communities structured solely by species sorting? *Ecol Ind*
868 62:86-94

869 Vyverman W, Verleyen E, Sabbe K, Vanhoutte K, Sterken M, Hodgson DA, Mann DG,
870 Juggins S, Van de Vijver B, Jones V, Flower R, Roberts D, Chepurnov VA,

871 Kilroy C, Vanormelingen P, De Wever A (2007) Historical processes constrain
872 patterns in global diatom diversity. Ecology 88:1924-1931

873 Wiens JJ (2012) Why biogeography matters: historical biogeography versus phylogeography
874 and community phylogenetics for inferring evolutionary and ecological
875 processes. Front Biogeogr 4:128-135

876

Tables and Figures

Table 1. Results of variation partitioning for each organismal group. For each organismal group, the response data were Hellinger-transformed (presence-absence) site-by-species matrix. E = environmental effects, B = basin effect and S = within-region spatial effect. Significance of shared effects (\cap) cannot be tested. Global models of spatial effects were never significant for these organismal groups except for lake macrophytes and lake fishes.

Stream diatoms				
Fraction	Df	Adj. R2	p	Variables in the model
E	3	0.039	0.001	Moss cover, pH, total phosphorus
B	2	0.069	0.001	Dummy variable basin
E+B	5	0.085	0.001	All variables above
$E \mid B$	3	0.016	0.002	
$E \cap B$	0	0.023		
$B \mid E$	2	0.046	0.001	
U		0.915		
Stream insects				
Fraction	Df	Adj. R2	p	Variables in the model
E	8	0.155	0.001	pH, shading, deciduous trees, stream width, depth, velocity, macrophytes, sand
B	2	0.132	0.001	Dummy variable basin
E+B	10	0.187	0.001	All variables above
$E \mid B$	8	0.055	0.001	
$E \cap B$	0	0.100		
$B \mid E$	2	0.032	0.001	
U		0.813		
Stream bryophytes				
Fraction	Df	Adj. R2	p	Variables in the model
E	2	0.049	0.001	pH, total phosphorus
B	2	0.057	0.001	Dummy variable basin
E+B	4	0.097	0.001	All variables above
$E \mid B$	2	0.039	0.002	
$E \cap B$	0	0.010		
$B \mid E$	2	0.048	0.001	
U		0.903		

Table 1. Continues on the next page...

Lake zooplankton				
Fraction	Df	Adj. R2	p	Variables in the model
E	2	0.075	0.010	Water temperature, lake area
B	2	0.078	0.001	Dummy variable basin
E+B	4	0.089	0.001	All variables above
$E \mid B$	2	0.010	0.073	
$E \cap B$	0	0.065		
$B \mid E$	2	0.014	0.020	
U		0.911		
Lake macrophytes				
Fraction	Df	Adj. R2	p	Variables in the model
E	4	0.173	0.001	Conductivity, secchi depth, altitude, lake area
S	3	0.061	0.001	MEM.15, MEM.8, MEM.2
B	2	0.027	0.008	Dummy variable basin
E+S	7	0.189	0.001	Conductivity, secchi depth, altitude, lake area, MEM.15, MEM.8, MEM.2
E+B	6	0.190	0.001	Conductivity, secchi depth, altitude, lake area, dummy variable basin
S+B	5	0.092	0.001	MEM.15, MEM.8, MEM.2, dummy variable basin
E+S+B	9	0.207	0.001	All variables above
$E \mid S+B$	4	0.116	0.001	
$S \mid E+B$	3	0.017	0.013	
$B \mid S+E$	2	0.018	0.004	
$E \cap S$	0	0.048		
$S \cap B$	0	-0.001		
$E \cap B$	0	0.012		
$E \cap S \cap B$	0	-0.003		
U		0.793		
Lake fish				
Fraction	Df	Adj. R2	p	Variables in the model
E	4	0.101	0.001	Lake area, altitude, colour, conductivity
S	2	0.036	0.001	MEM.1, MEM.12
B	2	0.028	0.011	Dummy variable basin
E+S	6	0.110	0.001	Lake area, altitude, colour, conductivity, MEM.1, MEM.12
E+B	6	0.118	0.001	Lake area, altitude, colour, conductivity, dummy variable basin
S+B	4	0.066	0.001	MEM.1, MEM.12, dummy variable basin
E+S+B	8	0.128	0.001	All variables above
$E \mid S+B$	4	0.063	0.001	
$S \mid E+B$	2	0.011	0.121	
$B \mid S+E$	2	0.018	0.037	
$E \cap S$	0	0.028		
$S \cap B$	0	-0.001		
$E \cap B$	0	0.012		
$E \cap S \cap B$	0	-0.001		
U		0.872		

887

888

889 Table 2. Multiple site beta diversity for each organismal group. Singletons were either included (yes) in or omitted (no) from the calculations of
890 gamma, alpha and beta diversity. Total beta diversity (Sorensen) was also decomposed into turnover (Simpson) and nestedness components
891 (Nested). S.D. = standard deviation of alpha diversity. The numbers of sites surveyed varied from 45 to 60 among the datasets.

892

Organismal group	Singletons	Gamma diversity	Alpha diversity		Beta diversity		
			Mean	S.D.	Sorensen	Simpson	Nested
Stream diatoms	Yes	305	50.8	15.5	0.939	0.916	0.023
Stream diatoms	No	225	49.0	14.2	0.935	0.912	0.023
Stream bryophytes	Yes	49	6.13	3.3	0.949	0.914	0.036
Stream bryophytes	No	31	5.73	3.0	0.944	0.905	0.039
Stream insects	Yes	203	28.6	9.86	0.956	0.938	0.018
Stream insects	No	144	27.6	9.59	0.954	0.935	0.019
Lake macrophytes	Yes	101	26.6	9.36	0.934	0.894	0.041
Lake macrophytes	No	88	26.3	9.14	0.933	0.892	0.041
Lake zooplankton	Yes	55	8.3	3.04	0.952	0.929	0.023
Lake zooplankton	No	37	8.0	2.91	0.949	0.924	0.025
Lake fish	Yes	25	12.7	2.88	0.879	0.770	0.109
Lake fish	No	24	12.6	2.87	0.878	0.768	0.110

893

Table 3. Results of the elements of metacommunity structure analysis. EAbs = embedded absences, Rep = replacements, I = Morisita's index, Mean Sim = Mean index value from 999 randomisations. Q-Clementsian = Quasi-Clementsian. The numbers of sites surveyed varied from 45 to 60 among the datasets.

Organismal group	Singletons	Coherence		Turnover			Clumping				Interpretation
		EAbs	p	Mean Sim	Rep	p	Mean Sim	I	p	df	
Stream diatoms	Yes	7432	< 0.001	9407	767451	< 0.001	272374	1.98	< 0.001	302	Clementsian
Stream diatoms	No	5945	< 0.001	6606	287325	0.003	144143	2.14	< 0.001	222	Clementsian
Stream bryophytes	Yes	709	< 0.001	1051	51751	0.669	45242	4.48	< 0.001	46	Q-Clementsian
Stream bryophytes	No	501	< 0.001	656	35036	0.025	22932	1.93	< 0.001	28	Clementsian
Stream insects	Yes	5988	< 0.001	8217	992307	< 0.001	284808	3.12	< 0.001	200	Clementsian
Stream Insects	No	4461	< 0.001	5562	422762	< 0.001	151359	2.38	< 0.001	141	Clementsian
Lake macrophytes	Yes	2325	< 0.001	3303	125361	0.001	63488	7.08	< 0.001	98	Clementsian
Lake macrophytes	No	2037	< 0.001	2812	88138	< 0.001	44729	6.74	< 0.001	85	Clementsian
Lake zooplankton	Yes	1427	0.003	1778	75124	0.662	64798	2.02	< 0.001	52	Q-Clementsian
Lake zooplankton	No	1085	0.427	1136	40136	0.383	31481	1.69	< 0.001	34	Random
Lake fish	Yes	411	0.003	518	7318	0.020	3722	2.22	< 0.001	22	Clementsian
Lake fish	No	369	< 0.001	494	6846	0.006	3297	2.12	< 0.001	21	Clementsian

Table 4. A comparison of elements of metacommunity structures (EMS) at the within-basin (Heino et al. 2015c; two to five different basins per organismal group) and across-basins (this study) spatial extents. Q = Quasi. The results suggest a clear shift from various different structures to Clementsian structures at large spatial extents.

Organismal group	Within basins (Heino et al. 2015c)	Across basins (this study)
Stream diatoms	Gleasonian, Q-Gleasonian, Clementsian	Clementsian
Stream bryophytes	Q-Gleasonian, Q-Clementsian, Clementsian	Q-Clementsian
Stream insects	Q-Gleasonian, Gleasonian, Clementsian	Clementsian
Lake macrophytes	Clementsian, Q-Clementsian	Clementsian
Lake zooplankton	Random, Q-nested, Q-Gleasonian	Q-Clementsian
Lake fish	Q-Nested, Q-Clementsian, Clementsian	Clementsian

Fig. 1. Our model systems encompass three metacommunities, each with several local communities indicated by black dots (e.g. a stream site). Black arrows connecting the metacommunities denote among-region dispersal and other region effects (a). Statistical approach includes Moran's eigenvector maps, redundancy analysis (RDA), calculation of multiple site beta diversity and definition of metacommunity structures (b). B = basin effect, E = environmental effect and S = spatial effect.

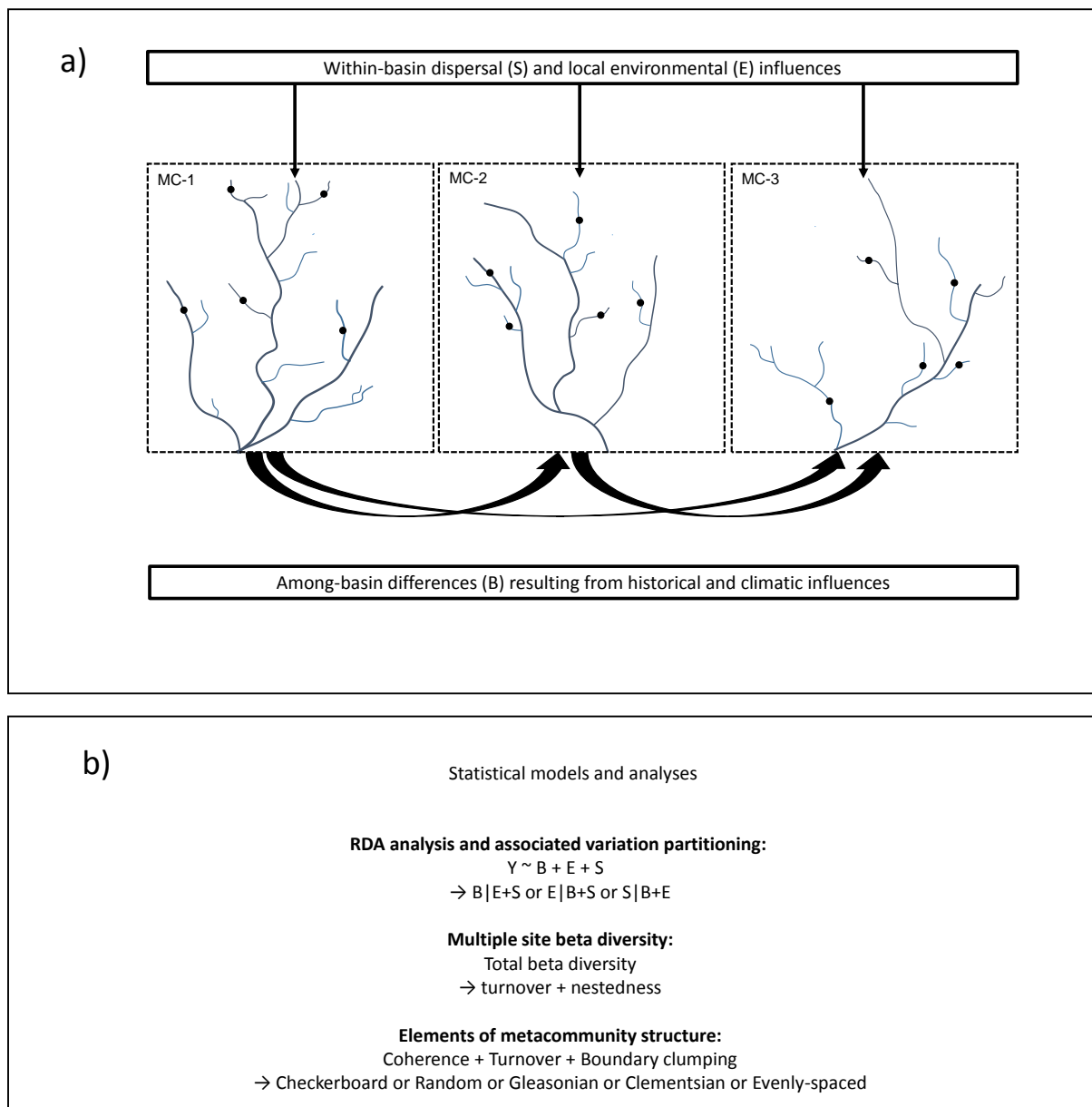


Fig. 2. Drainage basin boundaries in the datasets studied. Shown are the drainage basins sampled for diatoms and bryophytes (A), insects (B), macrophytes and fish (C) and zooplankton (D). Drainage basins are delineated to include only areas within the Finnish borders, because all surveys were done in Finland despite some drainage basins exceed national borders.

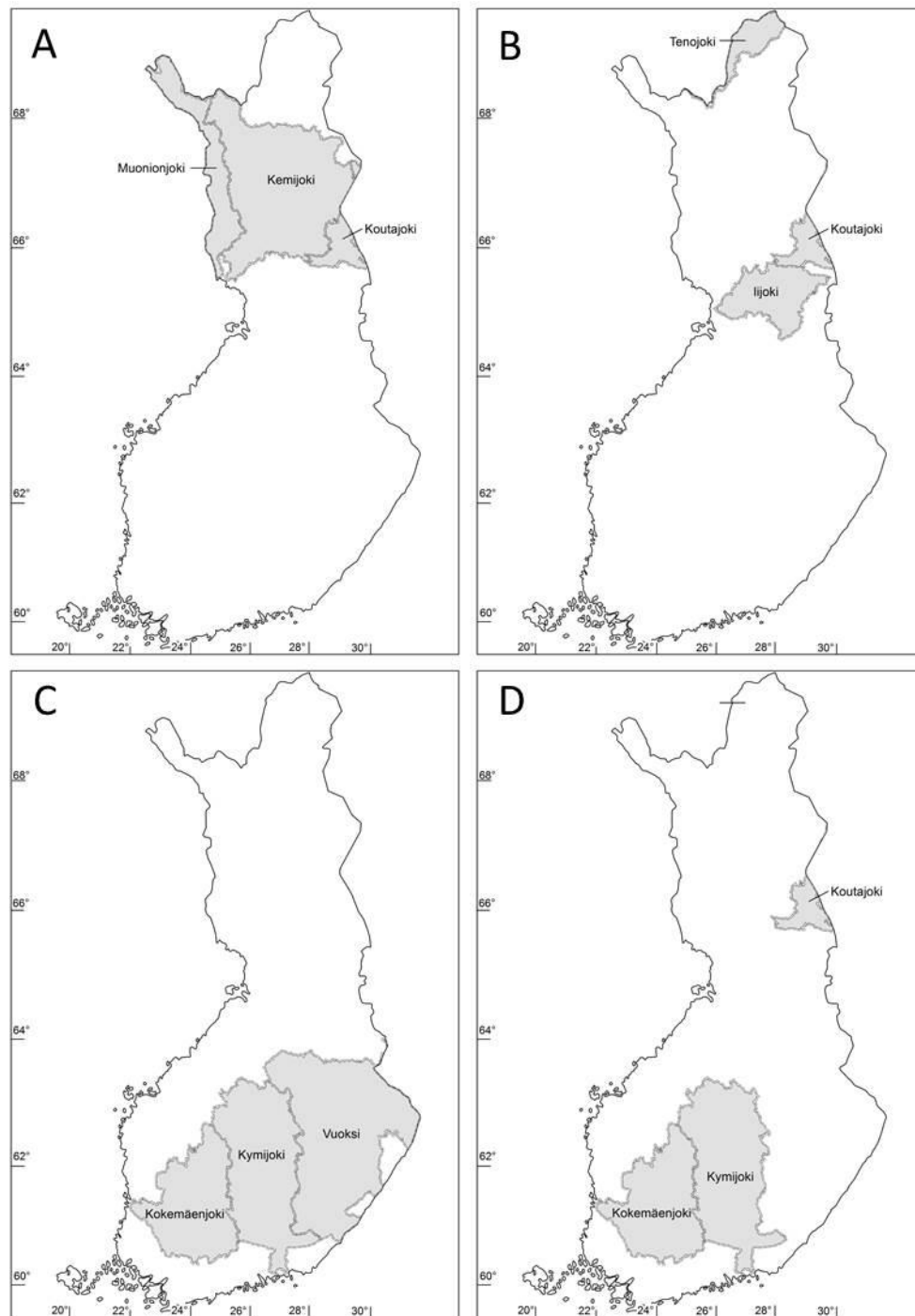


Fig. 3. PCA ordination plots for each organismal group. Different drainage basins are denoted by different symbols.

