

[Click here to view linked References](#)

1 Running title: Metacommunity ecology and biogeography of aquatic organisms

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

4

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

6 ¹Finnish Environment Institute, Natural Environment Centre, Biodiversity, P.O. Box 413, FI-
7 90014 Oulu, Finland.

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

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

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

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

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

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

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

22

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

25

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

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

48

49 **Keywords.** Bryophytes, diatoms, fish, invertebrates, lakes, macrophytes, metacommunities,
50 streams.

51 **Introduction**

52

53 Biogeography and community ecology are two disciplines that combine history, dispersal,
54 biotic interactions and environmental filtering as determinants of the structure of biotic
55 assemblages. However, a better understanding of the determinants of biotic assemblages
56 might benefit from a closer conceptual unification of these disciplines (Jenkins and Ricklefs
57 2011; Ricklefs and Jenkins 2011). Biotic assemblages can be understood to comprise either
58 regional biotas or local communities, depending on the grain size under investigation (Beck
59 et al. 2012; Pinel-Alloul et al. 2013). One means to investigate biogeographic and ecological
60 influences is to compare the effects of regional, spatial and environmental drivers of local
61 communities over large spatial extents (Shurin et al. 2009; Bini et al. 2014; Gonçalves-Souza
62 et al. 2014).

63 A major aim of biogeography is to consider evolutionary, historical and climatic
64 influences on regional biotas. One typically finds a strong relationship between present-day
65 climate and species richness (e.g. Hawkins and Porter 2003) or composition (e.g. Heino and
66 Alahuhta 2015) of regional biotas. The same is true for historical effects on regional biotas,
67 which can be investigated as phylogenetic patterns (e.g. Wiens 2012) or using various
68 statistical approaches as proxies of historical effects (e.g. Hortal et al. 2011). The degree to
69 which the influences of these broad-scale factors remain important when the focus is on local
70 communities is still elusive. Some studies have suggested that regional and historical
71 influences remain significant even when the focus is on local communities (Hoeinghaus et al.
72 2007; Vyverman et al. 2007), but others have shown that local environmental factors account
73 for significant variation in local community structure even over broad spatial extents (Van
74 der Gucht et al. 2007; Gonçalves-Souza et al. 2014; Souffreau et al. 2015). The influence of

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

78 Metacommunity ecology is a recently emerged subdiscipline of ecology, where
79 dispersal among sites is considered a key to understand biotic assemblages (Leibold et al.
80 2004; Jocque et al. 2010). Metacommunity ecology emphasises the idea that dispersal among
81 sites within a region is an important process affecting the structure of local communities
82 (Leibold et al. 2004). Dispersal may be limiting or homogenizing local communities, the
83 effects of which may not be easily distinguishable because both may induce spatial
84 structuring in the biological data (e.g. Ng et al. 2009). However, those effects can be at least
85 partly separated by focusing on nested spatial scales (Declerck et al. 2011; Silva and
86 Hernández 2015), where differences among regions may denote dispersal limitation and
87 spatial structures within a region mainly relate to homogenising effects of dispersal that can
88 happen via mass effects (e.g. Mouquet and Loreau 2003). For freshwater organisms, regions
89 can be individual drainage basins (e.g. Heino 2011), whereby dispersal is more likely to take
90 place within such regions than between regions (e.g. Heino et al. 2015a). Regions could also
91 be delineated based on drainage basin boundaries, major landscape configurations or
92 ecoregions in terrestrial studies.

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

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

106 In addition to explaining variation in local community structure, a major aim of both
107 biogeography and metacommunity ecology is to quantify the degree of variation (e.g. beta
108 diversity, Baselga 2010) or describe predominant patterns (e.g. elements of metacommunity
109 structure, Leibold and Mikkelsen 2002) in biological survey data. While beta diversity has
110 been quantified at various spatial grains and extents (Soininen et al. 2007, Anderson et al.
111 2011), the elements of metacommunity structure have mostly been tested using data from
112 local communities within relatively small regions (Leibold and Mikkelsen 2002; Heino et al.
113 2015c; but see Presley and Willig 2010; Meynard et al. 2013). Recently, Heino and Alahuhta
114 (2015) applied the elements of metacommunity structure approach to encompass large spatial
115 grain and geographical extent. They found, as opposed to sets of local communities within
116 small regions where various patterns are typically detected (Heino et al. 2015c; Tonkin et al.
117 2015a), that regional beetle faunas across a broad geographical gradient showed consistent
118 Clementsian type variation (Heino and Alahuhta 2015). **Clementsian type variation**
119 **emphasises discrete ‘community types’ along ecological gradients, such that subgroups of**
120 **species replace other subgroups in space (Clements 1936). Such variation also suggests that**
121 **subgroups of species either responded similarly to environmental variation or are affected by**
122 **similar historical effects (Heino and Alahuhta 2015; Tonkin et al. 2015b). We here expanded**
123 **this approach from single drainage basins to encompass local communities of six aquatic**
124 **organismal groups surveyed across three drainage basins.**

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

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

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

158

159 **Materials and Methods**

160

161 **Dataset characteristics and environmental variables**

162

163 We re-analysed some of our recently-collected data (Soininen et al. 2009; 2011; Grönroos et
164 al. 2013; Alahuhta et al. 2012; 2015; Heino et al. 2015c) for three groups of stream organisms
165 (i.e. diatoms, bryophytes, insects) and three groups of lake organisms (i.e. macrophytes,
166 zooplankton, fish). However, in this study, we combined data such that they comprised 45 to
167 60 sites across three drainage basins to facilitate comparative purposes. In all cases, datasets
168 were carefully taxonomically harmonised to guarantee that they were comparable. The three
169 drainage basins and sites surveyed generally differed between the organismal groups (Fig. 2),
170 **but stream diatoms and bryophytes were surveyed at the same sites.** Each drainage basin
171 drains into the sea (i.e. Arctic Ocean, White Sea or Baltic Sea). Due to limited resources, sites

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

175

176 *Stream diatoms*

177 We sampled 45 stream sites across three drainage basins in Finland. The drainage systems
178 were: (1) Koutajoki (centered on 66°N, 29°E), (2) Kemijoki (67°N, 28°E), and (3)
179 Muonionjoki (68°N, 24°E). The spatial extent (a rectangle encompassing all study sites)
180 comprising the three study areas was 63,609 km². Generally, 15 sites per region appeared to
181 be a sufficient sample size, detecting the majority of the diatom species present in the
182 regional species pool of a drainage basin (Soininen et al. 2009). Algal sampling was
183 conducted once for each site during summer low-flows in August 2001 or 2004. Sampling
184 was confined to near-pristine streams. All sampling was conducted by the same field crew
185 using a strictly standardized field protocol. Each study site with a length of 10 m was divided
186 into 5 or 10 cross-stream transects, depending on stream width. One or two stones were
187 selected randomly in each transect, and diatoms were scraped off the stones from a
188 predefined area (3.1 cm²), using a rubber template. Subsamples, 10 in total, were then pooled
189 into a composite sample for each site. In the laboratory, fresh samples were carefully checked
190 to guarantee that most diatom frustules were alive before acid combustion. We used acid
191 combustion (HNO₃: H₂SO₄, 2:1) to clean frustules of organic material. Cleaned diatoms were
192 mounted in Naphrax and a total of 500 frustules per sample were identified and counted,
193 using phase-contrast light microscopy (magnification 1000×) (for details, see Soininen et al.
194 2009). **In total, 96% of diatoms were identified to species.** We also measured current velocity,

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

197

198 *Stream bryophytes*

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

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

206

207 *Stream insects*

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

219 trees was assessed in a 50-meter section on both banks upstream of each sampling site.
220 Shading was estimated visually as percent canopy cover at the whole study section. Current
221 velocity (at $0.6 \times$ depth) and depth were measured at 30 random locations along cross-stream
222 transects, the number of which depended on stream width. More transects were sampled in
223 narrow than wide streams. Mean wetted width of each stream reach was measured based on
224 five cross-stream transects. Macrophyte cover (%) and substratum particle class cover (%)
225 were assessed at 10 random randomly-spaced 50 cm \times 50 cm plots. In addition, in each of the
226 10 plots, visual estimates of the percentage cover of five particle size classes were made
227 based on a modified Wentworth scale (see Grönroos et al. 2013). Water samples were
228 collected simultaneously with the field sampling and were measured for pH and conductivity.

229

230 *Lake macrophytes*

231 Macrophytes were sampled in 57 lakes with variable environmental conditions in three
232 drainage basins in Finland (Alahuhta et al. 2015). In each of the Kymijoki (62°N, 26°E) and
233 Vuoksi (63°N, 29°N) drainage basins, 20 lakes were surveyed, whereas 17 lakes were
234 investigated in the Kokemäki drainage basin (62°N, 24°N). The spatial extent comprising the
235 three study areas was 132,060 km². Lake macrophyte surveys were carried out during
236 growing seasons between 2002 and 2008 in the Kymijoki and Vuoksi drainage basins, and
237 between 2000 and 2011 in the Kokemäki drainage basin. Vascular plants, including
238 helophytes and hydrophytes, were sampled using a main belt transect method, in which a
239 varying number of 5-meter wide transects, depending on lake size, from the upper eulittoral
240 to the outer limit of vegetation were examined. **All macrophytes were identified to species.**
241 Ten hydro-morphological and water quality variables known to be important for aquatic
242 plants were measured in each lake (Alahuhta et al. 2015). These variables consisted of lake

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

249

250 *Lake zooplankton*

251 Zooplankton were sampled from 60 small lakes in Finland during July in 2008 or 2009. The
252 sites were sampled in three drainage basins. We sampled 20 lakes both in the Kokemäenjoki
253 (61°N, 24°E) and the Kymijoki (63°N, 25°E) drainage basins in 2008, and 20 lakes in the
254 Koutajoki (66°N and 29°E) drainage basin in 2009. These drainage basins were chosen
255 because they cover a relatively large geographical extent and because the nutrient
256 concentrations of lakes vary from ultraoligotrophic to eutrophic (Soininen et al. 2011). The
257 spatial extent comprising the three study areas was 125,190 km². We sampled only small
258 lakes to ensure that plankton sampling covers the site as adequately as possible. Most of the
259 lakes within the drainage basins were not readily inter-connected to each other via water
260 routes. For more information on the environmental characteristics of the lakes within the
261 drainage basins, see Soininen et al. (2011). Plankton samples were collected with a tube
262 sampler (V = 2.3 L) from three locations in the middle of the lake and pooled. We collected
263 the samples in the middle of the lakes in order to avoid benthic taxa from the littoral entering
264 the samples. The samples were collected at 0.5 m below the surface of the water.
265 Zooplankton samples (6.15 L) were filtered through a 50 µm net and preserved with
266 formaldehyde in the field. The maximum depth of the lakes as well as surface water

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

275

276 *Lake fishes*

277 The lake fish data were based on postal inquiries sent to persons employed as chairmen or
278 active members in regional fishing associations (Lappalainen and Malinen 2002). **All fish**
279 **were identified to species.** The data were from three drainage basins: Vuoksi (centered on
280 63°N, 28°E), Kymijoki (centered on 62°N, 26°E) and Kokemäenjoki (centered on 61°N,
281 24°E). From each of the three drainage basins, 20 lakes were randomly selected for this
282 study. The spatial extent comprising the three study areas was 150,869 km². The
283 environmental data of the lakes were based on the Hertta database (www.environment.fi).
284 The environmental variables available were lake area, length of the shoreline, altitude,
285 maximum lake depth, conductivity, pH, colour and total phosphorus. Average values of water
286 chemistry incorporating the whole water column for a period between June and September
287 were calculated.

288

289 **Spatial analysis**

290

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 Declerck 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 Declerck 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).

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

320

321 **Forward selection of explanatory variables**

322

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

329

330 **Variation partitioning**

331

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

339 eigenvector map analysis (see above) and dummy variable basin (**B**). Prior to the RDA, site-
340 by-species presence-absence data were Hellinger-transformed to make the data better
341 analysable using linear methods (Legendre and Gallagher 2001) and because Hellinger
342 transformation/Hellinger distance was deemed highly suitable for community composition
343 data in a comparative analysis (Legendre and de Caceres 2013). **The Hellinger transformation**
344 **consists in transforming the site-by-species data into relative values per site, by dividing each**
345 **value by the site sum, then taking the square root of the resulting values. Hellinger**
346 **transformation can be used for both presence-absence and abundance data (Legendre et al.**
347 **2005). We selected significant variables in the final RDA models of each set of variables (**E****
348 **or **S**) using the forward selection method with two stopping rules (Blanchet et al. 2008) with**
349 **the function “ordiR2step” in the R package vegan (Oksanen et al. 2013). We used**
350 **redundancy analysis (RDA) to partition variation in species composition (**Y**) between **E**, **S****
351 **and **B** following the widely-used variation partitioning approach (Borcard et al. 1992;**
352 **Legendre and Legendre 2012). Variation partitioning of species composition (**Y**) between**
353 **three sets of predictor variables results in pure environment (**E | S+B**), pure spatial (**S | E+B**)**
354 **and pure basin (**B | E+S**) fractions, as well as their shared effects and unexplained variance**
355 **(**U**). In many cases, spatial variables were not significant, and we thus ran the variation**
356 **partitioning between **E** and **B** only. Variation partitioning was run using the function**
357 **“varpart” in the R package vegan. We reported adjusted R^2 values in all analyses because**
358 **they are unbiased estimates of explained variation (Peres-Neto et al. 2006). We also tested for**
359 **the significance of the total **E**, **S** and **B** fractions, and pure fractions **E | S+B**, **S | E+B** and**
360 ****B | E+S** using the function “anova” in the package vegan. **Ecologically, we expected that **E******
361 **would be related to local environmental control, **S** to within-basin spatial dynamics and **B** to**
362 **biogeographic effects.**

363

364 **Visual inspection of breakpoints in metacommunity structure**

365

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

373

374

375 **Testing for different metacommunity structures**

376

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

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

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

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

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

416 We assessed the significance of the index values for coherence (EAbs) and turnover
417 (Rep) using the fixed-proportional null model, where row sums are fixed (i.e. the species
418 richness of each site was maintained), but column marginal frequencies (i.e. species
419 frequencies of occurrence) were considered probabilities. Random matrices were produced
420 using the “r1” method for the fixed-proportional null model as implemented in the R package
421 vegan (Oksanen et al. 2013). This method is the default in the R package we used (Dallas
422 2013), and it has been previously used in several other studies (e.g. Heino et al. 2015c).

423 Although a stricter fixed-fixed null model might provide slightly different results from those
424 of fixed-proportional null model, such fixed-fixed null model was not used because it is
425 overly conservative and because we could not have then compared our results with those
426 from single drainage basins (e.g. Heino et al. 2015e). We used 999 simulations to provide
427 simulated matrices. Statistical significance of EAbs or Rep was subsequently estimated by
428 comparing the observed index value from the original matrix to the distribution of values
429 derived from the 999 simulated matrices. Metacommunity structure was examined for each
430 organismal group based on axis 1 of reciprocal averaging because we were interested in the
431 most important species compositional gradient (Gauch 1982). All EMS analyses were run
432 using the R package metacom (Dallas 2013), with the “r1” method borrowed from the R
433 package vegan (Oksanen et al. 2013).

434

435 **Quantifying beta diversity**

436

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

443

444 **Results**

445

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

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

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

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

481

482 **Discussion**

483

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

493

494 **Determinants of community structure of aquatic organisms**

495

496 Environmental control often dominates over all spatial effects on metacommunity
497 organization (Cottenie 2005; Soininen 2014). We found support for this expectation in three
498 of the six organismal groups (i.e. insects, macrophytes, fish), which corroborates many
499 findings from streams (Landeiro et al. 2012; Grönroos et al. 2013; Alahuhta et al. 2015) and
500 lakes (Cottenie et al. 2003; Alahuhta and Heino 2013; Heino 2013). These findings suggest
501 that environmental filtering is the main mechanism structuring metacommunities (Cottenie
502 2005; Van der Gucht et al. 2007), at least if the spatial extent of a region under study is not
503 very broad (Mykrä et al. 2007; Heino et al. 2015a). Although the maximum spatial extent in
504 our datasets was more than 150,000 km², we did not find that basin identity (i.e. the
505 biogeographical effect) would overcome the effect of local environmental factors on the
506 community structure of insects, macrophytes or fish. This finding may be due to two main
507 reasons. First, environmental ranges typically increase with increasing spatial extent, thus

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

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

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

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

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

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

567 Dispersal may also potentially account for biogeographical variation in community
568 structure. We **hypothesised** that small passively dispersed organisms would surpass all
569 geographical barriers and would thus show no evidence of basin identity, whereas the
570 opposite should be true for large actively dispersing organisms. As a related **hypothesis**, we
571 expected small passive dispersers to show stronger environmental control (De Bie et al.
572 2012). We found at best little support for these conjectures, as all organismal groups showed
573 a significant pure region effect, and pure environmental effects did not vary consistently
574 between the passive and active dispersers. While such region effects might potentially be
575 related to limited dispersal between the three regions (Viana et al. 2015), they may equally
576 likely arise from climatic forcing on species distributions. However, as already indicated, it is
577 almost impossible to disentangle overall basin effects and present-day climate or historical
578 dispersal on our results because climate varies clearly *among* the drainage basins, but is
579 largely invariable among sites *within* our small and predominantly lowland drainage basins.

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

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

596

597 **Elements of metacommunity structure**

598

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

606 In the present study, we expected that Clementsian gradients would emerge when
607 metacommunity structures were analysed across multiple drainage basins because this
608 potentially means crossing multiple species pools and covering large environmental
609 gradients. On the one hand, different species pools should result in different local
610 communities, owing to a strong regional influence on local community structure (Heino et al.
611 2003; Soininen et al. 2009). On the other hand, environmental gradient lengths should
612 increase with increasing spatial extent, resulting in stronger species composition-environment
613 correlations among sites (Vetaas and Chaudhary 1998; Soininen 2014). Although we cannot
614 decisively distinguish between the two main drivers of Clementsian structures because, for
615 example, different organismal groups may show different regional vs local environmental
616 influences on breakpoints in community composition (Fig. 3), we found strong support for
617 such Clementsian structures. This finding is similar to those in a study of bat faunas on
618 Caribbean islands (Presley and Willig 2010), a study of beetle faunas over northern European
619 biogeographical provinces (Heino and Alahuhta 2015), a study on riverine invertebrates of
620 two central German drainage basins (Tonkin et al. 2015b) and a study on wetland crustacean
621 communities in Spanish wetlands (Gascón et al. 2016). However, this finding partly disagrees
622 with studies conducted within small regions, including the individual drainage basins
623 incorporated in this study. Heino et al. (2015c) found that the stream and lake
624 metacommunities of individual drainage basins showed variable (i.e. random, nested,
625 Gleasonian, Clementsian and quasi structures) metacommunity patterns (Table 4). This
626 suggested some degree of scale-dependency in metacommunity structures. We hence propose
627 that Clementsian structures are common in large-scale studies of local communities, i.e., a
628 combination of small-grained data with broad spatial extents (Beck et al. 2012; Bini et al.
629 2014; Dallas and Drake 2014), whereas various patterns may be detected in small-scale
630 studies (Heino et al. 2015e; Tonkin et al. 2015). Further indirect support for Clementsian

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

634

635 **Conclusions**

636

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

649

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

653

654 **References**

655

656 Alahuhta J, Heino J (2013) Spatial extent, regional specificity and metacommunity
657 structuring in lake macrophytes. *J Biogeogr* 40:1572-1582

658 Alahuhta, J, Hellsten S, Kuoppala, M, Riihimäki, J (2016) Regional and local determinants of
659 macrophyte community compositions in high-latitude lakes of Finland.
660 *Hydrobiologia*, in press

661 Alahuhta J, Rääpysjärvi J, Hellsten S, Kuoppala M, Aroviita J (2015) Species sorting drives
662 variation of boreal lake and river macrophyte communities. *Comm Ecol* 16:76-
663 85

664 Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders N.,
665 Cornell HV, Comita LS, Davies KF, Harrison SP, Kraft NJB, Stegen JC,
666 Swenson NG (2011) Navigating the multiple meanings of β diversity: a
667 roadmap for the practicing ecologist. *Ecol Lett* 14:19–28

668 Baselga A (2010) Partitioning the turnover and nestedness components of beta diversity.
669 *Global Ecol Biogeogr* 19:134–143

670 Beisner BE, Peres-Neto PR, Lindström E, Barnett A, Longhi ML (2006) The role of dispersal
671 in structuring lake communities from bacteria to fish. *Ecology* 87:2895-2991

672 Beck J, Ballesteros-Mejia L, Buchmann CM, Dengler J, Fritz SA, Gruber B, Hof C, Jansen F,
673 Knapp S, Kreft H, Schneider A-K, Winter M, Dormann CF (2012) What's on
674 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, Pélissier 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 Hoeninghaus 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

877 **Tables and Figures**

878

879 Table 1. Results of variation partitioning for each organismal group. For each organismal

880 group, the response data were Hellinger-transformed (presence-absence) site-by-species

881 matrix. E = environmental effects, B = basin effect and S = within-region spatial effect.

882 Significance of shared effects (\cap) cannot be tested. Global models of spatial effects were

883 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 B	3	0.016	0.002	
E ∩ B	0	0.023		
B 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 B	8	0.055	0.001	
E ∩ B	0	0.100		
B 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 B	2	0.039	0.002	
E ∩ B	0	0.010		
B E	2	0.048	0.001	
U		0.903		

884

885 Table 1. Continues on the next page...

886

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 B	2	0.010	0.073	
E∩B	0	0.065		
B 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 S+B	4	0.116	0.001	
S E+B	3	0.017	0.013	
B S+E	2	0.018	0.004	
E∩S	0	0.048		
S∩B	0	-0.001		
E∩B	0	0.012		
E∩S∩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 S+B	4	0.063	0.001	
S E+B	2	0.011	0.121	
B S+E	2	0.018	0.037	
E∩S	0	0.028		
S∩B	0	-0.001		
E∩B	0	0.012		
E∩S∩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			df	Interpretation
		EAbs	p	Mean Sim	Rep	p	Mean Sim	I	p		
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.

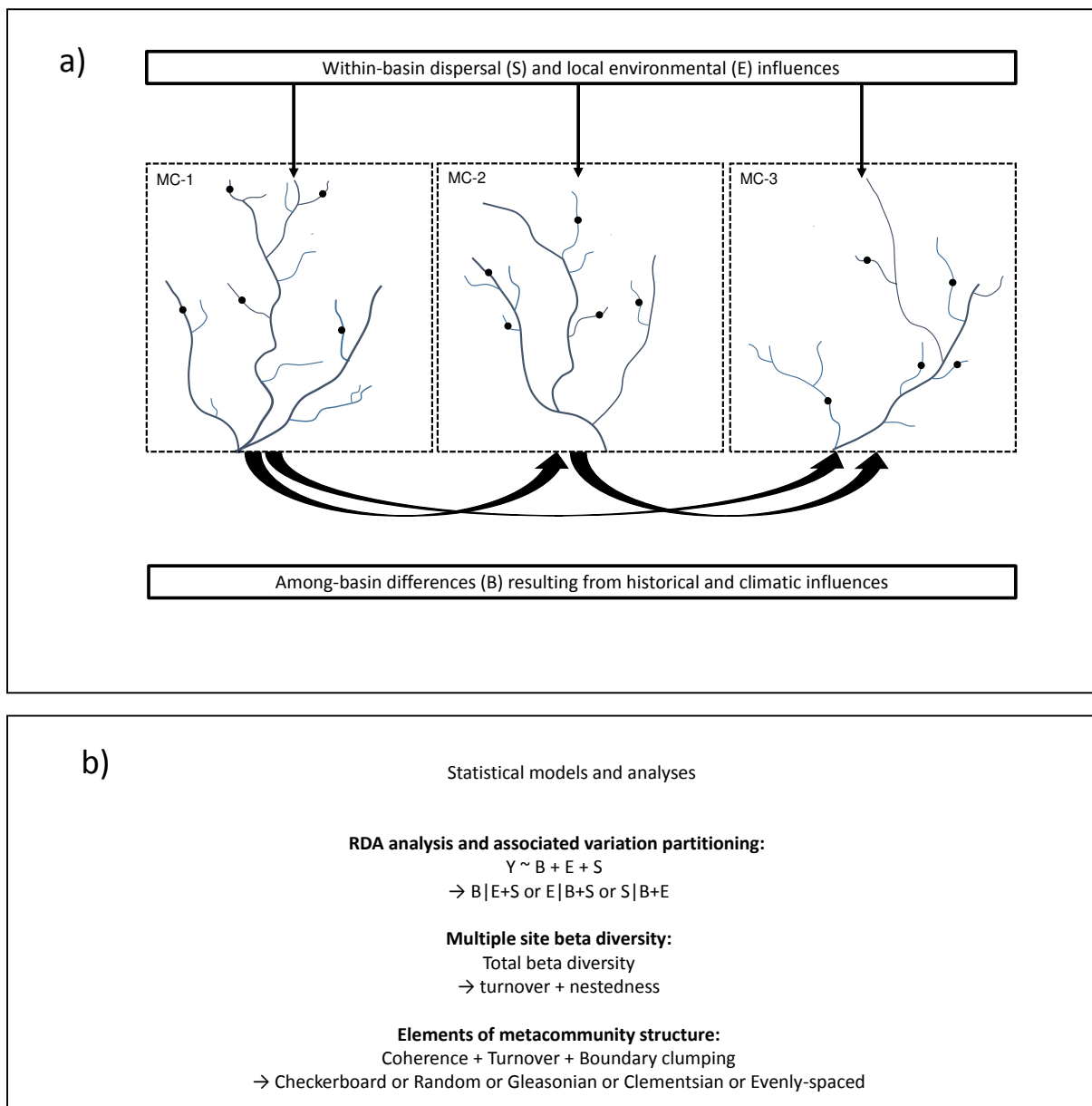


Fig. 2. Drainage basin boundaries in the datasets studied. Shown are the drainage basins sampled for diatoms and bryophytes (A), insects (B), macrophytes and fish (C) and zooplankton (D). Drainage basins are delineated to include only areas within the Finnish borders, because all surveys were done in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

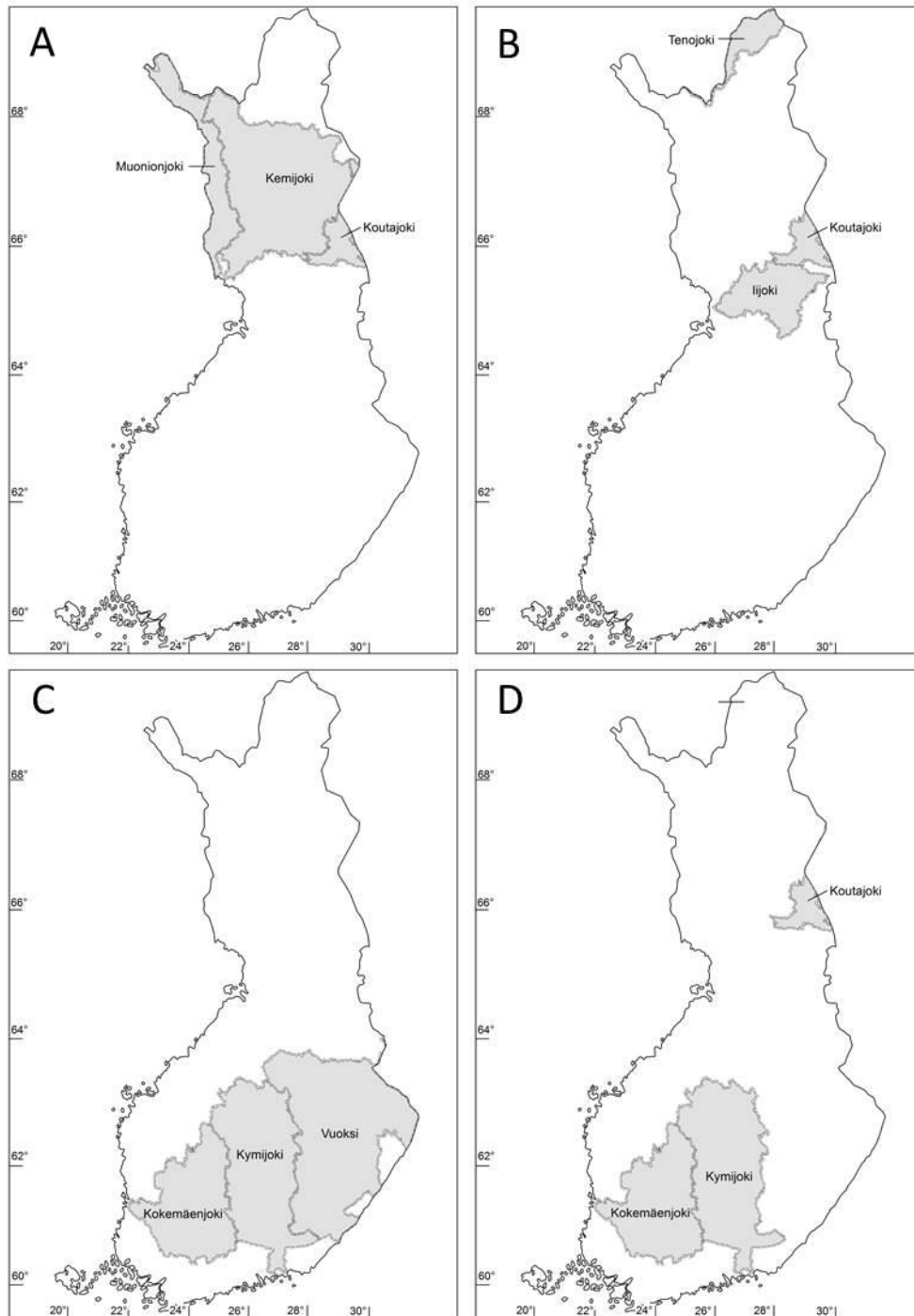


Fig. 3. PCA ordination plots for each organismal group. Different drainage basins are denoted by different symbols.

