Highly variable species distribution models in a subarctic stream metacommunity: patterns, mechanisms and implications

Guillermo de Mendoza[1,2,*], Riikka Kaivosoja[3], Mira Grönroos[4], Jan Hjort[3], Jari Ilmonen[5], Olli-Matti Kärnä[3], Lauri Paasivirta[6], Laura Tokola[3] & Jani Heino[7]

[1]Centre for Advanced Studies of Blanes, Spanish National Research Council (CEAB-CSIC), Accés a la Cala St. Francesc 14, ES-17300 Blanes, Spain.

[2]Laboratoire GEODE, UMR 5602 CNRS, Université Toulouse-Jean Jaurès, 5 allées Antonio Machado, FR-31058 Toulouse, France.

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

[4]University of Helsinki, Department of Environmental Sciences, Section of Environmental Ecology, Niemenkatu 73, FI-15140 Lahti, Finland.

[5]Metsähallitus, Natural Heritage Services, P.O. Box 94, FI-01301 Vantaa, Finland.

[6]Ruuhikoskenkatu 17, FI-24240 Salo, Finland.

[7]Finnish Environment Institute, Natural Environment Centre, Biodiversity, Paavo Havaksen Tie 3, FI-90570 Oulu, Finland.

* Corresponding author: guillermo.de-mendoza@univ-tlse2.fr

Summary

1. Metacommunity theory focuses on assembly patterns in ecological communities, originally exemplified through four different, yet non-exclusive, perspectives: patch dynamics, species sorting, source-sink dynamics, and neutral theory. More recently, three exclusive components have been proposed to describe a different metacommunity framework: habitat heterogeneity, species equivalence, and dispersal. Here, we aim at evaluating the insect metacommunity of a subarctic stream network under these two different frameworks.
2. We first modelled the presence/absence of 47 stream insects in northernmost Finland using binomial generalised linear models (GLMs). The deviance explained by pure local environmental (E), spatial (S), and climatic variables (C) was then analysed across species using beta regression. In this comparative analysis, site occupancy, as well as taxonomic and biological trait vectors obtained from principal coordinate analysis, were used as predictor variables.
3. Single-species distributions were better explained by in-stream environmental and spatial factors than by climatic forcing, but in a highly variable fashion. This variability was difficult to relate to the taxonomic relatedness among species or their biological trait similarity. Site occupancy, however, was related to model performance of the binomial GLMs based on spatial effects: as populations are likely to be better connected for common species due to their near ubiquity, spatial factors may also explain better their distributions.
4. According to the classical four-perspective framework, the observation of both environmental and spatial effects suggests a role for either mass effects or species sorting constrained by dispersal limitation, or both. Taxonomic and biological traits, including the different dispersal capability of species, were scarcely important, which undermines

the patch dynamics perspective, based on differences in dispersal ability between species. The highly variable performance of models makes the reliance on an entirely neutral framework unrealistic as well. According to the three-component framework, our results suggest that the stream insect metacommunity is shaped by the effect of habitat heterogeneity (supporting both species-sorting and mass effects), rather than species equivalence or dispersal limitation.

5. While the relative importance of the source-sink dynamics perspective or the species-sorting paradigm cannot be deciphered with the data at our disposal, we can conclude that habitat heterogeneity is an important driver shaping species distributions and insect assemblages in subarctic stream metacommunities. These results exemplify that the use of the three-component metacommunity framework may be more useful than the classical four perspective paradigm in analysing metacommunities. Our findings also provide support for conservation strategies based on the preservation of heterogeneous habitats in a metacommunity context.

Key-words

beta regression, comparative analysis, insects, metacommunity theory, single-species distribution models, stream macroinvertebrates, subarctic streams.

Introduction

Metacommunity theory predicts the assembly of ecological communities according to different perspectives. Originally, this idea was illustrated by Leibold et al. (2004) in the form of four metacommunity perspectives: (1) patch dynamics, which is based on a resource competition-colonisation trade-off among species, thus taking into account species' dispersal potential (Hanski, 1994); (2) species-sorting along environmental gradients, which relies on differences in environmental tolerance among species (Leibold, 1995); (3) mass effects or source-sink dynamics, whereby species may survive in poor-quality habitats owing to constant immigration from the source populations in high quality habitats (Pulliam 1988); and (4) the neutral theory, where demographic stochasticity solely explains assembly patterns (Hubbell, 2001). Deciphering which of these perspectives is more suitable in the context of metacommunity analysis seems difficult and may well depend on the context of analysis (e.g. spatial extent, biogeographic region, ecosystem type and more; Heino et al., 2015).

Nevertheless, the examples of metacommunity perspectives depicted in Leibold et al. (2004) are not mutually exclusive, and represent a fraction of possibilities which can be expanded with the inclusion of species dispersal rates, connectivity, species interactions, disturbance, priority effects, rapid local adaptation, meta-ecosystem dynamics and more (Brown, Sokol, Skelton, & Tornwall, 2017; Logue, Mouquet, Peter, & Hillebrand, 2011). The more recent proposal by Logue et al. (2011) claims that the metacommunity concept is better generalised by three major exclusive components, which decompose the metacommunity framework into (1) environmental heterogeneity, whereby habitat patches differ in environmental attributes; (2) species equivalence, in terms of niche characteristics; and (3) dispersal, referred to as the rate of dispersal among patches. Here, we aim at evaluating species distributions in a subarctic stream insect metacommunity under these two different

88 frameworks (i.e., Leibold et al., 2004 versus Logue et al., 2011), specifically so as to evaluate
89 which of the two is more adequate for the interpretation of our observations.
90 Species distribution models have previously been used to predict community-level
91 properties such as biodiversity (Ferrier & Guisan, 2006). Their accuracy in predicting
92 community-level properties appears to be higher than that of community assembly models,
93 although at a high cost in terms of model complexity (Bonthoux, Baselga, & Balent, 2013,
94 Chapman & Purse, 2011). The accuracy of single-species distribution modelling, however,
95 may also be advantageous to test ecological theories about community assembly mechanisms.
96 This is because accurately modelling the distribution of single species, one at a time, provides
97 the opportunity to proceed with a subsequent comparative analysis across species. Using a
98 comparative analysis, the variation in model performance can be related, for example, to
99 species traits and potential phylogenetic constraints.
100 Stream insect species, in particular, are highly suitable to decipher community
101 assembly processes through the comparative analysis of single-species distribution models
102 (Heino & de Mendoza, 2016). This is because of the high variability among species in
103 tolerance of environmental conditions, as well as resource exploitation, dispersal capability,
104 and habit traits (Merritt & Cummins, 1996; Tachet, Richoux, Bournaud, & Usseglio-Polatera,
105 2010; Schmidt-Kloiber & Hering, 2015; Serra, Cobo, Graça, Dolédec, & Feio, 2016). This
106 variability is valuable in evaluating which community assembly mechanism dominates in
107 each particular context of analysis. Basically, such an analysis might shed light into the
108 relevance of environmental variables, spatial variables, and dispersal capability of species on
109 model performance. Subsequently, this information can be used as an indicator of the
110 preponderance of one community assembly mechanism over another (Figure 1). For example,
111 if many species show similar spatial patterns, and if these species share the same dispersal
112 potential, we can presume that the ability to disperse may be underlying the observed general

pattern for these species. This would give us hints about the adequacy to consider one particular metacommunity theory perspective over the others. Within the classical four-perspective framework (Leibold et al., 2004), patch dynamics would likely be suitable in this case, as this perspective relies on the different capability of species to both disperse and exploit resources. Within the metacommunity framework based on three exclusive components (Logue et al., 2011), dispersal would be main driver in this case. Moreover, stream insects are also a diverse group of species, which belong to different insect orders and vary widely in physiological and morphological adaptations (Merritt & Cummins, 1996). Thus, modelling the distribution of single stream insect species and subsequently proceeding with a comparative analysis across species is also a suitable indirect practice to explore possible evolutionary constraints on community assembly processes.

In this study, we analysed the distribution of common stream insect species in the metacommunity of a subarctic drainage basin. Species differ widely in their dispersal capability (e.g. passive or active dispersers, aquatic or aerial adults) and tolerance of environmental conditions such as temperature, water flow, or habitat characteristics (Grönroos et al., 2013; Heino, 2005; Heino & Grönroos, 2014). We used environmental, climatic and spatial variables as predictors of the distributions of single stream insect species. Our aim was to elucidate, first, whether or not environmental and spatial factors are relevant for explaining the distribution of stream insect species; and second, whether or not the obtained models can be related to the different dispersal capability, site occupancy (i.e. a gradient of rarity-commonness), and biological and taxonomic traits, of stream insect species. Both considerations were used to evaluate which of the two different metacommunity frameworks, either the one based on four non-exclusive perspectives (Leibold et al., 2004) or the one based on three exclusive axes (Logue et al., 2011), is more adequate to interpret our observations of single species distributions in stream networks (Figure 1).

Methods

Study area

The field work for this study (Fig. S1) was conducted in the Tenojoki drainage basin (main stem length: 361 km, basin area: 16377 km^2, altitude of sites: from 19 to 285 m a.s.l.) in northernmost Finland (70^{o}N, 27^{o}E). This subarctic drainage basin is close to a natural state, since it is characterised by very small human populations and subsequent little impact from human development. A typical feature of the area are short cool summers and long cold winters (from early November to end of May). The mean annual temperature is about -2^{o}C in the continental areas of the drainage basin, and close to 0^{o}C near the Arctic Ocean (Dankers & Christensen, 2005). Annual precipitation ranges from 310 mm to 410 mm depending on the location in the drainage basin (Mansikkaniemi, 1970). Most of the rainfall and snowmelt enters streams and rivers, as evaporation is generally of minor importance. Vegetation is dominated by mountain birch (*Betula pubescens* ssp. c*zerepanowii*) woodlands at low altitude and barren tundra at higher altitude, but also peatlands, heathlands and riparian meadows occur commonly. Coniferous pine (*Pinus sylvestris*) woodlands occur only in scattered locations, mostly in the southern parts of the drainage basin. Wadeable streams and rivers (i.e. channel width < 25 m, water depth < 50 cm) in the area are close to a pristine state, providing excellent possibilities for examining species distributions in natural environmental conditions. We sampled altogether 55 tributary streams for this study (for details, see Kärnä et al., 2015). All these 1st to 5th order tributaries drain into the mainstem of the River Tenojoki or the River Utsjoki, and no site is located in the two mainstem rivers (Fig. S1).

Field sampling of stream insects

We took a 3-minute kick-net sample (net mesh size: 0.3 mm) at each study site (Kärnä et al., 2015) at the same time with the environmental measurements in early and middle of June 2012 (see below). The sample for each site consisted of six 30-s subsamples that were divided between main habitats at a riffle site (*ca*. 50 m^2) based on visual inspections of variation in depth, flow, moss cover and particle size. The six subsamples were pooled in the field to obtain a composite sample. Such a sampling method has been shown to be effective in northern streams, allowing to detect patterns in community structure (Heino, Ilmonen, & Paasivirta, 2014) and distributions of single species (Heino & de Mendoza, 2016). Samples were immediately preserved in ethanol in the field and were taken to the laboratory for further processing and identification. Animals were separated from detritus and moss fragments and identified to the lowest possible taxonomic level, mostly species (Kärnä et al., 2015).

Species considered and species traits

We detected 107 insect taxa, of which 87 could be taxonomically determined to species or species group (Kärnä et al., 2015). Insects determined to genus level were discarded as they were considered too likely to include a few species, which is inappropriate to model single-species distributions. Then, we focused on 48 species that occurred at more than 10% of the 55 study sites, that is, that occurred in at least six sites. This is because modelling the distribution of species present in less than six sites is likely to produce spurious results and therefore the analysis of these species was considered unreliable (e.g. Pearce & Ferrier, 2000). In practice, we could model the occupancies of only 47 species because the mayfly *Baetis rhodani* occurred at all sites, so we could not use this species to model presence/absence. The 47 stream insect species considered in this study are listed in Table S1.

Nomenclature generally follows de Jong et al. (2014) and more specific references for the Simuliidae (Adler & Crosskey, 2016 ; Ilmonen, 2014).

Body size class, dispersal potential, functional feeding groups and habit trait groups were considered as species traits (Table S2). Functional feeding groups refer to exploitation of different resources, while habit traits define modes of locomotion and attachment to substrate (Merritt & Cummins, 1996). Body size classes and female dispersal potential followed a previous study (Heino & de Mendoza, 2016), with additional information from Tachet et al. (2010), Schmidt-Kloiber & Hering (2015) and Serra et al. (2016). Female dispersal potential was characterised as being "low" or "high". In general, all species of the Simuliidae were considered to have high dispersal potential, owing to the fact that their females feed as flying adults, in most cases searching for blood of vertebrates, and hence were assumed here to generally persist much longer as active flyers than the rest of species. In this regard, Baldwin, West, and Gomery (1975) often found their marked Simuliidae females several kilometers away from their natal streams. Owing to their small size, the Simuliidae may also be distributed long distances passively by wind (Crosskey, 1990). All other species were considered as weak dispersers except for the caddisflies *Plectrocnemia conspersa* and *Potamophylax cingulatus*, according to the information available for these taxa from previous studies (Gíslason, Hannesdóttir, Munoz, & Pálsson, 2015; Hoffsten, 2004; Müller-Peddinghaus, 2011; Müller-Peddinghaus & Hering, 2013; Schmidt-Kloiber & Hering, 2015). Although such information about dispersal abilities of stream insects is rather simple, there is currently no better information available (Schmidt-Kloiber & Hering, 2015; Serra et al., 2016; Tachet et al., 2010). Functional feeding and habit trait groups generally follow Merritt & Cummins (1996).

Local environmental, climatic, and spatial variables

211 The 55 streams were surveyed during the early northern summer, between early and middle
212 of June in 2012. We measured a set of local (i.e. proximal) environmental variables that have
213 been found important for stream insects in northern drainage basins in previous studies
214 (Heino et al., 2014; Kärnä et al., 2015). These comprised physical habitat and water physico-
215 chemical variables. For physical habitat variables, we measured current velocity (m/s) and
216 depth (cm) at 30 random spots in a riffle site. We also measured mean width of the riffle site
217 based on five cross-channel measurements, evenly spaced across the surveyed riffle site.
218 Bank height and bank slope were measured at the same locations with stream width
219 measurements. Bank height was measured as the height of the lower stream bank, i.e. the
220 height from the water level to the edge of terrestrial vegetation. Bank slope was measured
221 (perpendicular to the stream) as a stream bank rise (cm) over 2 m starting from the edge of
222 terrestrial vegetation. Moss cover (%) and particle size classes (%) were visually estimated at
223 10 squares (1 m^2) at random locations in a riffle site. We used a modified Wentworth's
224 (1922) scale of particle size classes: sand (0.25–2 mm), gravel (2–16 mm), pebble (16–64
225 mm), cobble (64–256 mm) and boulder (256–1,024 mm). Based on the visual estimates for
226 each square, we calculated mean values for each particle size class and moss cover at a site
227 and used these mean values in species distribution modelling. We also visually estimated
228 shading (%) by riparian vegetation and proportion of riparian deciduous trees (%). For
229 physico-chemical properties, we measured pH, conductivity and water temperature at each
230 site in the field using a YSI device model 556 MPS (YSI Inc., Ohio, USA) and took
231 additional water samples during the field campaign for further analysis. Water samples were
232 frozen at the end of the day at the Kevo Field Station situated in the northern part of the study
233 area, and were later analysed for total nitrogen, colour, iron and manganese in the laboratory
234 of the Finnish Environment Institute in Oulu following Finnish national standards (National
235 Board of Waters, 1981).

We also included three climatic variables, including annual air temperature sum above 5^{o}C (growing degree days), mean annual air temperature and mean July air temperature for the period 1981–2010. These variables were calculated in ArcMap 10.2 for each site from a gridded (1 x 1 km) climate data provided by the Finnish Meteorological Institute (Pirinen et al., 2012). The gridded climate data were produced using meteorological station observations and Kriging interpolation (e.g. Aalto, Pirinen, Heikkinen, & Venäläinen, 2013). The selected climatic variables are likely to be important for the distributions of insects in this subarctic area, where temperature is closely associated with insect life cycles (Danks, 2007).

Spatial variables were distance-based Moran's Eigenvector Maps (db-MEM) based on geographical distances among sites (Dray, Legendre & Peres-Neto, 2006). These spatial db-MEM variables were obtained with the function "PCNM" of the R package "PCNM" (Legendre, Borcard, Blanchet, & Dray, 2013; R Core Team, 2013). We used the largest distance in the minimum spanning tree, keeping all sites connected, as the truncation threshold. Spatial db-MEM variables represent structures of autocorrelation at all spatial scales. Only those spatial db-MEM variables showing significant positive autocorrelation were included in subsequent modelling (Borcard, Gillet & Legendre, 2011), resulting in 13 spatial variables (Figure 2). Based on eigenvalues and bubble plot maps, the spatial variables can be divided into those ranging from large-scale spatial structures (e.g. V1, V2) and those showing very small scale spatial patterns (e.g. V12, V13).

Prior to modelling species distribution, we eliminated strongly correlated (i.e. Pearson $r > .7$) predictor variables from the sets of local environmental and climatic variables (see Dormann et al., 2013). Hence, we removed one variable (i.e. annual temperature sum) from the climatic variables and four variables (i.e. water iron, colour, conductivity and boulders) from the stream environmental variables. The spatial variables were already not mutually correlated (Borcard et al., 2011).

Modelling species distributions

The distribution (i.e. presence/absence) of each species was modelled using binomial generalised linear models (i.e. binomial GLMs with logit link function), using separately local environmental, climatic and spatial variables, with the R package "Rcmdr" (Fox, 2005). The deviance explained for each species was thus obtained for each binomial GLM with each of these three different subsets of variables (Figure 2). The variables selected for each species' model were based on forward selection and Bayesian Information Criterion (BIC), separately for each variable group (i.e. environmental, climate and spatial). BIC values were used because they prevented the selection of too complex models in our case, in contrast to AIC (results not shown), which is often the case under large sample sizes (Burnham & Anderson, 2004). Moreover, the target model under BIC selection does not depend on sample size, in contrast to AIC (Burnham & Anderson, 2004). Therefore, AIC may be problematic in our case as we aim at comparing model performance between species, which may differ in the number of presences and absences. Also, deviating observations were removed from some species' models if they had Cook's distance values > 1 and hence affected profoundly a few models (Cook, 1977). For environmental variables, we registered whether the effect was positive or negative on species distributions. We then used the selected variables of these three subsets (i.e. local environmental, climatic and spatial) to perform variation (deviance) partitioning by subtraction, similarly as performed in multivariate contexts (Legendre & Legendre, 2012). Specifically, the deviance accounted for subset A, subset B, and subset A and B together, was computed, so as to obtain the different fractions of variation solely explained by each subset (i.e. unshared with other subsets). We eventually obtained adjusted D^2 values (Guisan & Zimmermann, 2000; Legendre & Legendre, 2012) which could be attributed to pure local environmental (E), climatic (C) or spatial effects (S), as well as to

total effects combining the three subsets of pure effects and their joint effects (E+C+S effects; Figure 2). Modelling methods other than GLMs could have been possible, yet species probably show linear responses to the environmental predictors due to the fact that they are on the edge of their geographical and ecological distributions, making GLMs adequate. Adding quadratic terms to binomial models is unlikely to change results substantially in these situations (e.g. Pulido, Riera, Ballesteros, Chappius, & Gacia, 2015), and increase the difficulty of interpretation of the results. Also, deviance partitioning is easy to accomplish when this is based on GLMs.

Comparative analysis across species

We performed a comparative analysis across species using beta regression (Ferrari & Cribari-Neto, 2004), where the adjusted D^2 values obtained with previous binomial GLMs were used as the dependent variable to be explained by site occupancy, taxonomic vectors or species trait vectors (Figure 2). These vectors were obtained separately from Principal Coordinate Analysis (PCO). Using the taxonomic relatedness of species, a taxonomic relatedness matrix was built using the function "taxa2dist" in the R package "vegan" (Oksanen et al., 2013), and taxonomic vectors were handled as continuous PCO vectors with the function "pco" in the R package "ecodist" (Goslee & Urban, 2007). The first four taxonomic eigenvectors were selected as these had much higher eigenvalues than the rest (Fig. S2). Similarly, species trait vectors were also computed using body size class, dispersal potential, functional feeding groups, and habit trait groups (Table S2). Species traits were considered as regular factors, except body size class which was considered as an ordered factor, to obtain a distance matrix based on Gower's metric with the function "daisy" of the R package "cluster" (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2013), and eventually trait PCO vectors with the function "cmdscale". The four trait eigenvectors obtained were considered for further

311 statistical analyses. The variation in adjusted D^2 values across species that could be attributed
312 to pure E, pure C, pure S, or E+C+S effects was fitted on site occupancy, the four taxonomic
313 and four species trait vectors selected, using beta regression with the function "betareg" of
314 the R package "betareg" (Cribari-Neto & Zeileis, 2010). Beta regression is adequate when the
315 response variable (in this case, the adjusted D^2 values) is constrained between 0 and 1.
316 We compared the explained variation by pure E, C, and S effects with a Kruskal-
317 Wallis test, with additional Mann-Whitney tests for subsequent pair-wise comparisons
318 between groups. Non-parametric tests were chosen since adjusted D^2 values data departed
319 from normality following the Shapiro-Wilk test (Zar, 1984). We also analysed the univariate
320 relationships between site occupancy, body size, dispersal potential, broad taxonomic insect
321 groups, functional feeding groups, habit trait groups, and taxonomic and trait vectors.
322 Depending on the continuous (e.g. site occupancy) or categorical (e.g. habit trait group)
323 nature of the variables involved, we followed Kruskal-Wallis tests, Mann-Whitney tests,
324 Fisher's exact test or Spearman correlations, as these variables were generally not normally
325 distributed (Zar, 1984).
326
327 **Results**
328 *Single species models*
329 Local environmental and spatial effects accounted for a higher variation in species
330 distributions (16.1% and 12.6% in average, respectively) than did climatic effects (5.4%; $p <$
331 .001, Kruskal-Wallis test), whereas the average deviance explained did not differ
332 significantly between local environmental and spatial effects ($p = .125$, Mann-Whitney test)
333 (Table S3). The local environmental factors most frequently selected in explaining species
334 distributions were water temperature, shading, and to a lesser extent, stream width, cobbles
335 and moss (Figure 3). The spatial variables most often selected were better represented by

large-scale spatial variables within the Tenojoki drainage basin (e.g. V1, V2) than by small-spatial scale variables (e.g. V12, V13), as also shown in Figure 3. Amongst the climate variables, mean annual temperature was significant in explaining the distribution of 32 species, and July air temperature of 20 species (not shown in Fig. 3).

The adjusted deviance explained by binomial GLMs was highly variable across species and difficult to relate to particular taxonomic groups (Table S3). For example, local environmental effects were particularly relevant for the stonefly *Siphonoperla burmeisteri* (i.e. accounting for 66.7% of adjusted D^2 values), the mayfly *Heptagenia dalecarlica* (50.2%), and the blackfly *Prosimulium hirtipes* (37.7%), whereas spatial effects were most relevant for the caddisfly *Rhyacophila nubila* (41.2%), the stonefly *Brachyptera risi* (31.5%), and the chironomid midge *Cardiocladius capucinus* (28.6%). Climate effects were also highly variable. They were generally low (see above), and accounted for more than 20% of adjusted D^2 values in only three cases: the stoneflies *Diura nanseni* and *Siphonoperla burmeisteri* (34.4% and 20.1%, respectively), and the chironomid midge *Orthocladius rivicola* (28.9%). Combining all effects, binomial GLMs explained on average 37.8% of the null deviance (Table S3).

Comparative analysis across species models

The highly variable species-local environment and species-climate relationships in binomial GLMs were not accounted for by site occupancy, or by taxonomic and trait vectors, in the beta regression analysis (Table 1). The deviance explained by spatial variables was, however, significantly (i.e. $p < .05$) accounted for by site occupancy (Table 1). The influence of TAX-PCO4 and TRA-PCO2 on the adjusted D^2 values predicted by spatial effects in binomial GLMs was significant as well. Also, the influence of TAX-PCO3 was marginally significant (i.e. $p < .10$), remaining like this in the binomial GLMs based on all variables combined

(Table 1). However, when repeating the beta regression analysis by using only the significant variables selected (i.e. site occupancy, TAX-PCO3, TAX-PCO4, and TRA-PCO2), only site occupancy was statistically significant (p = .017), but not TAX-PCO3, TAX-PCO4 or TRA-PCO2 (p = .943, p = .175, and p = .449, respectively, results not shown in Table 1). Analysing through beta regression the univariate relationship of these variables with the adjusted D^2 values of binomial GLMs based on spatial effects produced a similar result (site occupancy, p = .036, Fig. S3; TAX-PCO3, TAX-PCO4, and TRA-PCO2, p = .760, p = .660, and p = .524, respectively, results not shown). This univariate relationship between the adjusted D^2 values and site occupancy was not observed when the adjusted D^2 values of binomial GLMs were referred to environmental or climate effects (Fig. S3). No statistical significance was observed either for univariate relationships between separate species traits and the adjusted D^2 values in binomial GLMs, with the sole exception of body size (Fig. S3).

The TAX-PCO3 vector showed the highest species scores for blackflies (Simuliidae) and the lowest for mayflies (Ephemeroptera), and was strongly correlated (p < .001) to dispersal potential (Table S5, Figure 4a). In contrast to this taxonomic vector, TAX-PCO4 showed the highest species scores for both blackflies and mayflies (Figure 4b), and was strongly correlated to site occupancy (p = .007, Table S5). Finally, TRA-PCO2 reflects the influence of functional feeding groups and body size on model performance (Figure 4c), as indicated by the strong correlation of both variables (i.e. p < .001) with this trait vector (Table S5).

Discussion

Single species models

Our results indicated that single species distributions of stream insects are highly variable in terms of predictability, as well as the significant environmental and spatial predictors

386 underlying such distributions. There was no evident association between model accuracy and
387 particular taxonomic groups (Table S3). Nevertheless, a few generalisations can be
388 highlighted with regard to the results obtained. For example, water temperature and shading,
389 and to a lesser extent, stream width, cobbles and moss, were more relevant as environmental
390 predictors of species distributions than stream flow or water chemistry variables (Table S3,
391 Figure 3). This is in line with the well-known influence of temperature and resource
392 availability on insect life cycles at high latitudes (Danks, 2007) and indicates the influence of
393 species sorting processes along these environmental gradients. Resource availability is
394 represented in our case by shading, which indicates the proximity of terrestrial vegetation and
395 hence is a surrogate of availability of allochthonous resources from terrestrial origin for
396 aquatic insect larvae. This typically corresponds with a situation of a low-order stream which,
397 as in our case, is influenced strongly by terrestrial material from riparian vegetation which is
398 then taken as food resource by shredders, hence promoting their dominance (Vannote,
399 Minshall, Cummins, Sedell, & Cushing, 1980). Shading may also be inversely related to
400 primary productivity, but in this study, we found that the relationship of species distribution
401 with shading was always positive (Figure 3), suggesting that rather than biofilm production, it
402 is the external input of terrestrial material from riparian birch tree abundance what is likely
403 driving species distributions. In our case, shading was selected as a significant variable in
404 binomial models for some predators (*Isoperla difformis* and *Plectrocnemia conspersa*) and
405 shredders (*Leuctra* spp.), for some collector-gatherers (*Corynoneura lobata*-type,
406 *Eukiefferiella devonica*-group, *Orthocladius rhyacobius*-group and *Tvetenia discoloripes*),
407 and for some collector-filterers (*Philopotamus montanus* and *Prosimulium hirtipes*) (Table
408 S3). These latter groups perhaps benefit indirectly from the increase in potential resources
409 that the variable “shading” represents for shredders, for example, through the enhancement of

410 nutrient re-cycling by shredding coarse plant litter (Covich, Palmer, & Crowl, 1999; Wallace
411 & Webster, 1996).
412 Spatial variables were also relevant for the distributions of some species. Specifically,
413 large-scale spatial variables were more important than small-scale variables in explaining
414 species distributions in our study (Table S3, Figure 3). At a larger spatial extent (ca. 500 km
415 latitudinal gradient), previous findings indicate a stronger relevance of environmental factors,
416 compared to spatial restrictions, on single-species distributions (Heino & de Mendoza, 2016).
417 This is perhaps not surprising because increasing the spatial extent may have a strong positive
418 effect on the relevance of niche processes through larger environmental gradients (Chase,
419 2014). However, increasing the spatial extent may also preclude species to reach
420 environmentally suitable locations owing to dispersal limitation, and thus the relative
421 contribution of both environmental and spatial constraints on species distributions does not
422 always vary predictably with spatial scale (Alahuhta & Heino, 2013).
423
424 *Comparative analysis across species*
425 Comparative analysis across the species models showed a clear relationship between model
426 performance and site occupancy. Specifically, the binomial GLMs that we built upon spatial
427 variables could be related to site occupancy, and to a lesser extent, to taxonomic and trait
428 vectors, whereas none of these variables was significantly related to model performance
429 when models were based on local environmental or climate variables (Table 1). At first
430 glance, our results also suggested both a slight influence of female dispersal potential (related
431 to the taxonomic vector TAX-PCO3), and a potential influence of functional feeding groups
432 and body size (related to the trait vector TRA-PCO2), on the performance of models based on
433 spatial variables. The taxonomic vector TAX-PCO3 perhaps relates to female dispersal
434 potential, as species scores along this vector were much higher for the blackflies than for the

rest of species, and lowest for the mayflies (Figure 4). Blackflies are possibly the best active dispersers among all the insects we considered, because females feed as flying adults and in most species they must actively search for blood meals, often several kilometers away from their natal streams (Baldwin et al., 1975). However, adult mayflies, do not feed and often have extremely short life spans (Brittain, 1990). Therefore, it seems reasonable to assume that blackflies may actively disperse better than mayflies. Site occupancy and dispersal potential were not correlated (Table S5), and both taxa were the ones with highest number of sites occupied (Fig. S4). In contrast, mayflies differed in site occupancy from non-biting midges (Chironomidae) (Fig. S4), despite species in both groups can be considered weak active dispersers, as chironomid adults are also short-lived and generally weak active fliers (Armitage, 1995). On the other hand, the trait vector TRA-PCO2 suggests an influence of feeding behaviour and body size (Figure 4, Table S5) on model performance. This is because the exploitation of food resource from terrestrial origin (i.e. shredders) would facilitate the development of more complex trophic food webs with the inclusion of predators (Figure 4). This would also contribute to the positive association of body size to TRA-PCO2 (Fig. 4), as the largest insects we found are either predators or shredders (Table S2).

Nevertheless, it is important to note that taxonomic and trait vectors had a comparatively much weaker effect on predictability by spatial variables than that of site occupancy. In fact, not only did site occupancy attain a higher statistical significance (Table 1), but it could also be partly related to the capability of the taxonomic vector TAX-PCO4 to account for the adjusted D^2 values of binomial GLMs because these two predictor variables were significantly correlated (Table S5). Moreover, when repeating the beta regression analysis by using only the significant variables selected (i.e. site occupancy, TAX-PCO3, TAX-PCO4, and TRA-PCO2), only site occupancy was statistically significant, indicating that the influence of taxonomic and trait vectors on model performance is rather weak.

460 Analysing through beta regression the univariate relationship of these variables with the
461 adjusted D^2 values of binomial GLMs based on spatial effects again resulted in site
462 occupancy as the only significant variable (see Results above). Therefore, we must conclude
463 that any potential effect of taxonomic and trait vectors on model performance, including the
464 effect of female dispersal potential and body size, and that of functional feeding groups, must
465 be considered with caution: their statistical significance only appears after controlling for site
466 occupancy and the other variables considered in the full model of beta regression. In this
467 regard, the fact that *Baetis rhodani* is a widespread mayfly, which could not be modelled
468 because it was present at all sites, also gives support to the idea that dispersal abilities are not
469 so important in structuring invertebrate assemblages in high-latitude drainage basins. This is
470 because it demonstrates that mayfly species can be widespread, despite being rather weak
471 active dispersers. We also acknowledge that the rarest species (i.e. present in less than six
472 sites) were not modelled because models based on such small number of presences were
473 considered unreliable (e.g. Pierce & Ferrier, 2000). However, excluding these species does
474 not undermine the conclusion that the distributions of most common species are better
475 accounted for by models based on spatial variables than that of not-so-common species. In
476 fact, we effectively modelled 47 out of the 86 taxa available at the species (most cases) or
477 species-group (few cases) taxonomic resolution, comprising 55% of cases, which is a
478 representative subset of species in the entire metacommunity.
479
480 *Approaching the suitability of metacommunity analysis frameworks*
481 With the information above about single-species distribution models and subsequent
482 comparative analysis across species, it is possible to proceed with the evaluation of the
483 suitability of the two different frameworks of metacommunity analysis (Figure 1) considered
484 here: (1) the classical approach exemplified by the four different non-exclusive perspectives

485 described by Leibold et al. (2004) or (2) the three exclusive components as proposed by
486 Logue et al. (2011).

487 Among the four different metacommunity perspectives of the Leibold et al. (2004)
488 framework, neutral theory and patch dynamics do not rely on the effect of environmental
489 variables, in contrast to species sorting and source-sink dynamics, the latter of which also
490 incorporating a strong influence of spatial effects (Figure 1a). In our study, single-species
491 models often relied on the effect of environmental variables, particularly temperature and
492 shading, while being also dependent on large-scale spatial variables (Figure 3). As
493 environmental and spatial factors are both relevant for the distribution of species, this result
494 suggests that either species sorting along spatially structured environmental gradients, or
495 source-sink dynamics between populations of high-quality and low-quality habitats, are both
496 likely as important processes driving metacommunities. Then, the comparative analysis
497 across species showed that site occupancy is responsible for the observed differences in the
498 relevance of spatial variables on species distributions (Table 1). This suggests that common
499 species would be better able than rare species to maintain populations in low-quality habitats
500 through constant immigration, favouring the source-sink dynamics perspective over species
501 sorting.

502 Although species-sorting processes cannot be completely discarded because of the
503 demonstrated influence of environmental variables in many cases, deviance partitioning
504 suggests that the pure effects of environmental and spatial factors on species distributions are
505 stronger than their joint effects (Table S3). Also, the effect of spatial variables was better
506 explained than that of environmental factors by our explanatory variables, particularly site
507 occupancy, in the comparative analysis. These results slightly undermine the idea of species-
508 sorting across spatially structured environmental gradients as the most important process
509 shaping metacommunities. In any case, the neutral theory, which relies entirely on spatial

dynamics, is unlikely. As the dispersal potential of species has a rather weak effect on model accuracy, patch dynamics can be discarded as well as a suitable perspective of metacommunity analysis in our case. It should be acknowledged, however, that the difficulty to explain model performance with dispersal ability can also be a consequence of the coarseness of the dispersal measures currently available for freshwater invertebrates (Schmidt-Kloiber & Hering, 2015; Serra et al., 2016; Tachet et al., 2010). Moreover, the different metacommunity paradigms from Leibold et al. (2004) may always act simultaneously to a certain extent along a continuum (Figure 1a) rather than being distinct and mutually exclusive options (Brown, Sokol, et al., 2017; Gravel, Canham, Beaudet, & Messier, 2006; Logue et al., 2011).

Spatial autocorrelation may appear not only as a consequence of mass effects or species sorting along spatially structured environmental gradients when the spatial scale is not very large, but also as a consequence of dispersal limitation at very large spatial scales (Heino et al., 2015). Nevertheless, some insect species found in this study exemplify well the potential importance of the source-sink dynamics for metacommunities in subarctic streams, independently of their dispersal capability. For example, six blackfly species were examined (Table S1), of which five were present in more than 50% of sites, three of them in 75% of sites or more (Table S2). Thus, blackfly species in subarctic streams have successfully spread widely, which is advantageous to maintain metapopulations through source-sink dynamics. On the other hand, the mayflies are as widespread as the blackflies (Fig. S4), but far less capable of active dispersal. This suggests that the dispersal capability of species does not determine the metapopulation dynamics, whereas site occupancy probably does so. Spatial autocorrelation patterns have been described for the blackflies at small spatial scales, driven by strong effects of inter-specific competition for oviposition sites, and subsequent priority effects at the community level (McCreadie & Adler, 2012). The importance of priority effects

for the blackflies reinforces the idea of the relevance of site occupancy for community dynamics, where rare species are in clear disadvantage for habitat recolonisation.

Alternative to the framework of Leibold et al. (2004), we can interpret our results under the framework of Logue et al. (2011), whereby three different and mutually exclusive components can be used to analyse metacommunities: species equivalence, habitat heterogeneity and dispersal (Figure 1b). In our case, this alternative framework makes interpretation of the results much easier. At the very least, we can conclude that species equivalence is unlikely to play any role in metacommunity dynamics, similarly to discarding neutral theory under the Leibold *et al*. (2004) framework. Dispersal can also be discarded, yet again with caution due to the current lack of high resolution dispersal measures for freshwater invertebrates (Schmidt-Kloiber & Hering, 2015; Serra et al., 2016; Tachet et al., 2010). Thus, the main difference in the interpretation of the results with this alternative framework is that we can now be certain about the role of habitat heterogeneity, while under the Leibold et al. (2004) framework it is more difficult to discern whether species sorting or source-sink dynamics is the dominant process. Habitat heterogeneity is indeed related to both mechanisms. In fact, using habitat heterogeneity in space and time as the templet for ecological strategies (Southwood, 1977) could be the framework of choice in situations where it is difficult to discern species sorting processes from source-sink dynamics.

Alternative approaches, caveats and conclusions

Emergent properties at the community level are difficult to discern from field observational data alone. In this regard, population genetics can be very useful in order to gain confidence about the distinction between, e.g., source-sink dynamics and species sorting processes. This is because population genetic studies could be used to estimate the relative contribution of immigrants from nearby populations to the genetic variability of the population under study

(Bunn & Hughes, 1997; Hughes, Huey, & Schmidt, 2013; Hughes, Schmidt, & Finn, 2009). Genetic analyses would probably provide the opportunity for a more robust interpretation of our results. Genetic studies, however, are difficult to accomplish with stream insects in the field when the idea is to compare many species at a time, and they are far more expensive than the comparative approach of single species distributions we considered here. Therefore, the comparative approach presented here can be used as a first step to explore the relative contribution of environmental and spatial factors on species distributions, without using expensive and time-consuming genetic analyses. In fact, by using the comparative approach we can certainly conclude that the dispersal capability of species and neutral theory play little role in shaping subarctic stream insect metacommunities. Rather, it is habitat heterogeneity, which influences mass effects and/or species sorting processes, that matters. Subsequently, the results of our study strongly recommend the preservation of habitat heterogeneity as the conservation strategy to maintain biodiversity in these ecosystems.

Nevertheless, it should be acknowledged that one shortcoming of single-species distribution modelling is that it does not consider the influence of species interactions in structuring ecological communities. Stream ecology has considered that severe environmental conditions may weaken the potential effects of biotic interactions in structuring communities (Peckarsky, 1983). However, more recent findings pose doubts as to whether this is actually true (Cadotte & Tucker, 2017; Thomson et al., 2002). In fact, biotic interactions can reproduce patterns of community structure essentially identical to what it could be expected from environmental filtering alone. This is because environmental changes may affect population growth rates of competing species in opposite ways, and this may cause the exclusion of some species that would otherwise be able to coexist (Cadotte & Tucker, 2017). There exists also evidence indicating that biotic interactions limit the geographical range expansion of species facing environmental changes (Pigot & Tobias, 2013; Sexton et al.,

585 2009). Overall, this suggests that inter-specific interactions may also play a role in our case,
586 although the abundances of insect larvae in subarctic streams are typically low (see also
587 Heino & Grönroos, 2017) and may thus result in weak density-dependent interactions among
588 species (see also Morin, 2011).

589 Our study considered tributary streams draining into two linear sub-elements of a
590 larger river network (Fig. S1). However, there exists growing concern about the potential role
591 of the entire dendritic river networks in shaping biodiversity patterns, community structure
592 and species distributions (Altermatt, 2013; Altermatt & Fronhofer, 2017; Brown, Wahl, &
593 Swan, 2017; Jamoneau, Passy, Soininen, Lebuocher, & Tison-Rosebery, 2017; Schmera et
594 al., 2017). For example, the consideration of whole river networks may unveil a more
595 preeminent role for spatial factors in community assembly, undermining the role of
596 environmental filtering. Therefore, studies conducted across whole dendritic networks could
597 be more in line with neutral theory, as shown by Muneepeerakul et al. (2008) for fish
598 communities, yet no environmental variable was truly considered in that study. Although we
599 focused on tributary streams draining into the main river, the consideration of whole dendritic
600 networks may help us to perceive more accurately the real connectivity pathways between
601 isolated patches. This connectivity may have consequences for metacommunity stability with
602 respect to a situation where only a linear component of this network is acting (Fagan, 2002).
603 Also, dispersal along dendritic networks implies more variability in local richness with strong
604 consequences also for community differentiation among patches (Carrara, Altermatt,
605 Rodriguez-Iturbe, & Rinaldo, 2012; Seymour, Fronhofer, & Altermatt, 2015). In any case,
606 dispersal processes in stream networks may depend on the organism group considered
607 (Schmera et al. 2017).

608 In the case of stream insects, the taxa considered and the taxonomic resolution
609 achieved prior to species-distribution modelling, may also have important consequences on

our perception of the influence of dendritic riverine networks on biodiversity patterns (Kaelin & Altermatt, 2016). Here, some taxa were discarded as it was not possible to determine the species. Provided that the influences of dendritic landscapes and biotic interactions (discussed above) may strongly affect how we understand the reality of community assemblages, it is essential to use the best taxonomic resolution possible to make accurate inferences about the mechanisms truly governing the observed patterns. In fact, the criterion of 'best taxonomic resolution possible' used in our modelling endeavours is a fundamental requirement to draw robust conclusions to be applied in biodiversity conservation.

Finally, for biodiversity conservation, it is essential to focus on maintaining habitat heterogeneity because it appears to determine metacommunity organization (Kärnä et al. 2015) and species distributions (Heino & de Mendoza, 2016) in streams at high latitudes. Unless habitat heterogeneity is not considered (along with potentially important effects of dendritic network structure and biotic interactions), conservation plans may fall short and not result in desired outcomes. .

Acknowledgements

We thank Sirkku Lehtinen and Marja Lindholm for help in the field or the laboratory. We also thank two anonymous reviewers and the Guest Editor Florian Altermatt for their very helpful and constructive comments on an earlier version of this manuscript. Kevo Subarctic Research Station provided facilities during the field work. This study is part of the project "Spatial scaling, metacommunity structure and patterns in stream communities" that was supported financially by a grant from the Academy of Finland. Further support was provided by grants (no: 273557, no: 267995 and no: 285040) from the Academy of Finland. The authors declare no conflict of interest.

References

Aalto, J., Pirinen, P., Heikkinen, J., & Venäläinen, A. (2013) Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. *Theoretical and Applied Climatology*, **112**, 99–111.

Adler, P.H., & Crosskey, R.W. (2016) *World blackflies (Diptera: Simuliidae): A comprehensive revision of the taxonomic and geographical inventory*. Retrieved from http://www.clemson.edu/cafls/biomia/pdfs/blackflyinventory.pdf.

Alahuhta, J., & Heino, J. (2013) Spatial extent, regional specificity and metacommunity structuring in lake macrophytes. *Journal of Biogeography*, **40**, 1572–1582.

Altermatt, F. (2013) Diversity in riverine metacommunities: a network perspective. *Aquatic Ecology*, **47**, 365–377.

Altermatt, F., & Fronhofer, E. A. (2017) Dispersal in dendritic networks: Ecological consequences on the spatial distribution of population densities. *Freshwater Biology*, **63**, 22–32. https://doi.org/10.1111/fwb.12951

Armitage, P. D. (1995) Behaviour and ecology of adults. In P. D. Armitage, P. S. Cranston, & L. C. V. Pinder (Eds.), *The Chironomidae: Biology and Ecology of Non-Biting Midges* (pp. 194–224). London: Chapman & Hall.

Baldwin, W. F., West, A. S., & Gomery, J. (1975) Dispersal pattern of black flies (Diptera: Simuliidae) tagged with ^{32}P. *The Canadian Entomologist*, **107**, 113–118.

Bonthoux, S., Baselga, A., & Balent, G. (2013) Assessing community-level and single-species models predictions of species distributions and assemblage composition after 25 years of land cover change. *PLoS ONE*, **8**, e54179.

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

659 Brittain, J. E. (1990) Life history strategies in Ephemeroptera and Plecoptera. In I. C.
660 Campbell (Ed.), *Mayflies and Stoneflies* (pp. 1–12). Dordrecht: Kluwer Academic
661 Publishers.
662 Brown, B. L., Sokol, E. R., Skelton, J., & Tornwall, B. (2017) Making sense of
663 metacommunities: dispelling the mythology of a metacommunity typology.
664 *Oecologia*, **183**, 643–652.
665 Brown, B. L., Wahl, C., & Swan, C. M. (2017). Experimentally disentangling the influence
666 of dispersal and habitat filtering on benthic invertebrate community structure.
667 *Freshwater Biology*, **63**, 48–61. https://doi.org/10.1111/fwb.12995
668 Bunn, S. E., & Hughes, J. M. (1997) Dispersal and recruitment in streams: evidence from
669 genetic studies. *Journal of the North American Benthological Society*, **16**, 338–346.
670 Burnham, K. P., & Anderson, D. R. (2004) Multimodel inference: understanding AIC and
671 BIC in model selection. *Sociological Methods & Research*, **33**, 261–304.
672 Cadotte, M. W., & Tucker, C. M. (2017) Should environmental filtering be abandoned?
673 *Trends in Ecology & Evolution*, **32,** 429–437.
674 Carrara, F., Altermatt, F., Rodriguez-Iturbe, I., & Rinaldo, A. (2012) Dendritic connectivity
675 controls biodiversity patterns in experimental metacommunities. *Proceedings of the*
676 *National Academy of Sciences of the United States of America*, **109**, 5761–5766.
677 Chapman, D. S., & Purse, B. V. (2011) Community versus single-species distribution models
678 for British plants. *Journal of Biogeography*, **38**, 1524–1535.
679 Chase, J. M. (2014) Spatial scale resolves the niche versus neutral theory debate. *Journal of*
680 *Vegetation Science*, **25**, 319–322.
681 Cook, R. D. (1977) Detection of influential observation in linear regression. *Technometrics*,
682 **19**, 15–18.

683 Covich, A. P., Palmer, M. A., & Crowl, T. A. (1999) The role of benthic invertebrate species
684 in freshwater ecosystems: Zoobenthic species influence energy flows and nutrient
685 cycling. *BioScience*, **49**, 119-127.
686 Cribari-Neto, F., & Zeileis, A. (2010) Beta regression in R. *Journal of Statistical Software*,
687 **34**, 1–24.
688 Crosskey, R. W. (1990) *The Natural History of Blackflies*. John Wiley & Sons, Chichester.
689 Dankers, R., & Christensen, O.B. (2005) Climate change impacts on snow coverage,
690 evaporation and river discharge in the sub-arctic Tana basin, Northern Fennoscandia.
691 *Climatic Change*, **69**, 367–392.
692 Danks, H. V. (2007) How aquatic insects live in cold climates. *The Canadian Entomologist*,
693 **139**, 443–471.
694 de Jong, Y., Verbeek, M., Michelsen, V., Bjørn, P. P., Los, W., Steeman, F., Bailly, N.,
695 Basire, C., Chylarecki, P., Stloukal, E., Hagedorn, G., Wetzel, F. T., Glöckler, F.,
696 Kroupa, A., Korb, G., Hoffmann, A., Häuser, C., Kohlbecker, A., Müller, A.,
697 Güntsch, A., Stoev, P., & Penev, L. (2014) Fauna Europaea: All animal species on the
698 web. *Biodiversity Data Journal*, **2**, e4034.
699 Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J.
700 R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne,
701 P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S.
702 (2013) Collinearity: A review of methods to deal with it and a simulation study
703 evaluating their performance. *Ecography*, **36**, 27–46.
704 Dray, S., Legendre, P., & Peres-Neto P. R. (2006) Spatial modelling: A comprehensive
705 framework for principal coordinate analysis of neighbour matrices (PCNM).
706 *Ecological Modelling*, **196**, 483–493.

Fagan, W. F. (2002) Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology*, **83**, .3243–3249.

Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31**, 799–815.

Ferrier, S., & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.

Fox, J. (2005) The R Commander: A Basic Statistics Graphical User Interface to R. *Journal of Statistical Software*, **14(9)**, 1–42.

Gíslason, G. M., Hannesdóttir, E. R., Munoz, S. S., & Pálsson, S. (2015) Origin and dispersal of *Potamophylax cingulatus* (Trichoptera: Limnephilidae) in Iceland. *Freshwater Biology*, **60**, 387–394.

Goslee, S.C. & Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22(7)**, 1–19.

Gravel, D., Canham, C. D., Beaudet, M., & Messier, C. (2006) Reconciling niche and neutrality: The continuum hypothesis. *Ecology Letters*, **9**, 399–409.

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

Guisan, A., & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hanski, I. (1994) A practical model of metapopulation dynamics. *Journal of Animal Ecology*, **63**, 151–162.

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

Heino, J., & de Mendoza, G. (2016) Predictability of stream insect distributions is dependent on niche position, but not on biological traits or taxonomic relatedness of species. *Ecography*, **39**, 1216–1226.

Heino, J., & Grönroos, M. (2014) Untangling the relationships among regional occupancy, species traits and niche characteristics in stream invertebrates. *Ecology and Evolution*, **4**, 1931–1942.

Heino, J., & Grönroos, M. (2017) Exploring species and site contributions to beta diversity in stream insect assemblages. *Oecologia*, **183**, 151–160.

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

Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S., & Bini, L. M. (2015) Metacommunity organisation, spatial extent and dispersal in aquatic systems: Patterns, processes and prospects. *Freshwater Biology*, **60**, 845–869.

Hoffsten, P.-O. (2004) Site-occupancy in relation to flight-morphology in caddisflies. *Freshwater Biology*, **49**, 810–817.

Hubbell, S. P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press.

Hughes, J. M., Huey, J. A. & Schmidt, D. J. (2013) Is realised connectivity among populations of aquatic fauna predictable from potential connectivity? *Freshwater Biology*, **58**, 951–966.

Hughes, J. M., Schmidt, D. J., & Finn, D.S. (2009) Genes in streams: Using DNA to understand the movement of freshwater fauna and their riverine habitat. *BioScience*, **59**, 573–583.

Ilmonen, J. (2014) Checklist of the family Simuliidae (Diptera) of Finland. *ZooKeys*, **441**, 91–95.

Jamoneau, A., Passy, S. I., Soininen, J., Lebuocher, T., & Tison-Rosebery, J. (2017) Beta diversity of diatom species and ecological guilds: Response to environmental and spatial mechanisms along the stream watercourse. *Freshwater Biology*, **63**, 62–73. https://doi.org/10.1111/fwb.12980

Kaelin, K., & Altermatt, F. (2016) Landscape-level predictions of diversity in river networks reveal opposing patterns for different groups of macroinvertebrates. *Aquatic Ecology*, **50**, 283–295.

Kärnä, O.-M., Grönroos, M., Antikainen, H., Hjort, J., Ilmonen, J., Paasivirta, L., & Heino, J. (2015) Inferring the effects of potential dispersal routes on the metacommunity structure of stream insects: as the crow flies, as the fish swims or as the fox runs? *Journal of Animal Ecology*, **84**, 1342–1353.

Legendre, P., & Legendre L. (2012) *Numerical Ecology (3rd English edition)*. Elsevier, Amsterdam.

Legendre, P., Borcard, D., Blanchet, F.G., & Dray, S. (2013). *PCNM: MEM spatial eigenfunction and principal coordinate analyses*. R package version 2.1-2/r109. Retrieved from http://R-Forge.R-project.org/projects/sedar/

Leibold, M. A. (1995) The niche concept revisited: mechanistic models and community context. *Ecology*, **76**, 1371–1382.

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.

780 Logue, J. B., Mouquet, N., Peter, H., & Hillebrand, H. (2011) Empirical approaches to
781 metacommunities: a review and comparison with theory. *Trends in Ecology and*
782 *Evolution*, **26**, 482–491.
783 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2013) *cluster: Cluster*
784 *analysis basics and extensions*. R package version 1.14.4. Retrieved from
785 http://CRAN.R-project.org/package=cluster.
786 Mansikkaniemi, H. (1970) Deposits of sorted material in the Inarijoki-Tana river valley in
787 Lapland. *Reports of Kevo Subarctic Research Station*, **6**, 1–63.
788 McCreadie, J. W., & Adler, P. H. (2012) The roles of abiotic factors, dispersal, and species
789 interactions in structuring stream assemblages of black flies (Diptera: Simuliidae).
790 *Aquatic Biosystems*, **8**, 14.
791 Merritt, R. W., & Cummins, K. W. (1996) *An Introduction to the Aquatic Insects of North*
792 *America, 3rd ed*. Dubuque: Kendall/Hunt Publishing.
793 Morin, P. J. (2011) *Community Ecology, 2nd ed.* Oxford: Wiley-Blackwell.
794 Müller-Peddinghaus, E. (2011) *Flight-morphology of Central European caddisflies (Insecta:*
795 *Trichoptera) in relation to their ecological preferences*. PhD Thesis, University of
796 Duisburg-Essen, Essen.
797 Müller-Peddinghaus, E., & Hering, D. (2013) The wing morphology of limnephilid
798 caddisflies in relation to their habitat preferences. *Freshwater Biology*, **58**, 1138–
799 1148.
800 Muneepeerakul, R., Bertuzzo, E., Lynch, H. J., Fagan, W. F., Rinaldo, A., & Rodriguez-
801 Iturbe, I. (2008) Neutral metacommunity models predict fish diversity patterns in
802 Mississippi-Missouri basin. *Nature*, **453**, 220–222.

National Board of Waters (1981) *Vesihallinnon analyysimenetelmät*, Tiedotus 213. Helsinki: Vesihallitus.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2013) *vegan: Community Ecology Package*. R package version 2.0-9. Retrieved from http://CRAN.R-project.org/package=vegan.

Pearce, J., & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.

Peckarsky, B. L. (1983) Biotic interactions or abiotic limitations? A model of lotic community structure. In T. D. Fontaine III, & S. M. Bartell (Eds.), *Dynamics of Lotic Ecosystems* (pp. 303–323). Ann Arbor: Ann Arbor Science Publishers.

Pigot, A. L., & Tobias, J. A. (2013) Species interactions constrain geographic range expansion over evolutionary time. *Ecology Letters*, **16**, 330–338.

Pirinen, P., Simola, H., Aalto, J., Kaukoranta, J. P., Karlsson, P., & Ruuhela, R. (2012) Climatological statistics of Finland 1981–2010. *Finnish Meteorological Institute Reports*, **1,** 1–96.

Pulido, C., Riera, J. L., Ballesteros, E., Chappuis, E., & Gacia, E. (2015) Predicting aquatic macrophyte occurrence in soft-water oligotrophic lakes (Pyrenees mountain range). *Journal of Limnology*, **74,** 143–154.

Pulliam, H. R. (1988) Sources, sinks, and population regulation. *American Naturalist*, **132**, 652–661.

R Core Team (2013) *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. Retrieved from http://www.R-project.org/.

Schmera, D., Árva, D., Boda, P., Bódis, E., Bolgovics, Á., Borics, G., Csercsa, A., Deák, C., Krasznai, E. Á., Lukács, B. A., Mauchart, P., Móra, P., Sály, P., Specziár, A., Süveges

K., Szivák, I., Takács, P., Tóth, M., Várbíró, G., Vojtkó, A. E., & Erős, T. (2017). Does isolation influence the relative role of environmental and dispersal-related processes in stream networks? An empirical test of the network position hypothesis using multiple taxa. *Freshwater Biology*, **63**, 74–85. https://doi.org/10.1111/fwb.12973

Schmidt-Kloiber, A., & Hering, D. (2015) www.freshwaterecology.info – An online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences. *Ecological Indicators*, **53**, 271–282.

Serra, S. R. Q., Cobo, F., Graça, M. A. S., Dolédec, S., & Feio M. J. (2016) Synthesising the trait information of European Chironomidae (Insecta: Diptera): Towards a new database. *Ecological Indicators*, **61**, 282–292.

Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009) Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 415–436.

Seymour, M., Fronhofer, E. A., & Altermatt, F. (2015) Dendritic network structure and dispersal affect temporal dynamics of diversity and species persistence. *Oikos*, **124**, 908–916.

Southwood, T. R. E. (1977) Habitat, the templet for ecological strategies? *Journal of Animal Ecology*, **46,** 337–365.

Tachet, H., Richoux, P., Bournaud, M., & Usseglio-Polatera, P. (2010) *Invertébrés d'eau douce: systématique, biologie, écologie (nouvelle édition revue et augmentée)*. Paris: CNRS.

Thomson, J. R., Lake, P. S., & Downes, B. J. (2002) The effect of hydrological disturbance on the impact of a benthic invertertebrate predator. *Ecology*, **83**, 628–642.

852 Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C.E. (1980) The
853 river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**,
854 130–137.
855 Wallace, J. B., & Webster, J. R. (1996) The role of macroinvertebrates in stream ecosystem
856 function. *Annual Review of Entomology*, **41**, 115–139.
857 Wentworth, C. K. (1922) A scale of grade and class terms for clastic sediments. *Journal of*
858 *Geology*, **30,** 377–392.
859 Zar, J. H. (1984) *Biostatistical Analysis, Second Edition*. Prentice Hall, Englewood Cliffs.
860

Supporting Information

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

Table S1. Insect species considered.

Table S2. Species traits considered and site occupancy.

Table S3. Results of binomial GLMs.

Table S4. Taxonomic and trait vectors from Principal Coordinate Analysis (PCO), with corresponding scores for each species.

Table S5. Statistical significance of the correlations among site occupancy, species traits, trait vectors, and taxonomic vectors.

Figure S1. A map of the study area located in the Tenojoki drainage basin.

Figure S2. Eigenvalues from taxonomic Principal Coordinate Analysis (PCO) based on taxonomic distances between species.

Figure S3. Results of binomial GLMs in relation to site occupancy and species traits.

Figure S4. Comparison of site occupancy values between different insect groups.

Table 1. Results of beta regression showing the effects of site occupancy, biological trait vectors and taxonomic vectors on different fractions of variation (adjusted deviance, Adj. D^2) explained by binomial GLMs: local environment (E) effects, climate (C) effects, spatial (S) effects, and combined (E+C+S) effects. Significant values ($p < .05$) are shown in boldface; marginally significant values ($p < .10$) in italics.

Adj. D^2 of E effects						
	Estimate	SE	z	*p*	Log-likelihood	Pseudo R^2
(Intercept)	-1.6002	0.2220	-7.209	**<0.001**	49.06	0.2500
Site occupancy	-0.0042	0.0091	-0.469	0.646		
TAX-PCO1	-378.2072	323.7056	-1.168	0.243		
TAX-PCO2	91.4160	161.7362	0.565	0.572		
TAX-PCO3	225.8850	218.0393	1.036	0.300		
TAX-PCO4	-112.1618	101.9824	-1.100	0.271		
TRA-PCO1	-0.5557	0.8628	-0.644	0.520		
TRA-PCO2	1.3525	1.0712	1.263	0.207		
TRA-PCO3	0.3959	0.9475	0.418	0.676		
TRA-PCO4	-1.3110	1.3459	-0.974	0.330		
Adj. D^2 of C effects						
	Estimate	SE	z	*p*	Log-likelihood	Pseudo R^2
(Intercept)	-2.9806	0.2682	-11.115	**<0.001**	94.66	0.2309
Site occupancy	0.0029	0.0103	0.278	0.781		
TAX-PCO1	-579.9412	375.5543	-1.544	0.123		
TAX-PCO2	83.1028	185.9703	0.447	0.655		
TAX-PCO3	352.6843	252.8857	1.395	0.163		
TAX-PCO4	-1.9501	120.4650	-0.016	0.987		
TRA-PCO1	-1.6362	1.0045	-1.629	0.103		
TRA-PCO2	0.4327	1.2476	0.347	0.729		
TRA-PCO3	0.6454	1.1129	0.580	0.562		
TRA-PCO4	-2.1768	1.5519	-1.403	0.161		
Adj. D^2 of S effects						
	Estimate	SE	z	*p*	Log-likelihood	Pseudo R^2
(Intercept)	-2.4436	0.2222	-10.997	**<0.001**	60.19	0.2137
Site occupancy	0.0208	0.0083	2.498	**0.012**		
TAX-PCO1	309.6453	360.2543	0.860	0.390		
TAX-PCO2	-4.0452	162.3764	-0.025	0.980		
TAX-PCO3	365.5554	205.9633	1.775	*0.076*		
TAX-PCO4	217.5296	102.0912	2.131	**0.033**		
TRA-PCO1	-1.0051	0.8333	-1.206	0.228		

TRA-PCO2	2.5228	1.1999	2.102	**0.036**		
TRA-PCO3	1.4151	0.9765	1.449	0.147		
TRA-PCO4	-0.7984	1.3342	-0.598	0.550		

Adj. D^2 of E+C+S effects

	Estimate	SE	z	*p*	Log-likelihood	Pseudo R^2
(Intercept)	-0.7133	0.2331	-3.060	**0.002**	17.76	0.1791
Site occupancy	0.0111	0.0095	1.164	0.244		
TAX-PCO1	-140.2227	352.1883	-0.398	0.691		
TAX-PCO2	32.5012	178.6572	0.182	0.856		
TAX-PCO3	407.2860	228.4095	1.783	*0.075*		
TAX-PCO4	81.0016	110.2921	0.734	0.463		
TRA-PCO1	-1.3052	0.9142	-1.428	0.153		
TRA-PCO2	1.6446	1.1935	1.378	0.168		
TRA-PCO3	0.9742	1.0367	0.940	0.347		
TRA-PCO4	-1.0409	1.4575	-0.714	0.475		

881

882 **Figure legends**

883 **Figure 1** Conceptual representation of (a) the four non-exclusive classical approaches in
884 metacommunity studies (Leibold et al. 2004), and (b) the more recent framework of
885 metacommunity analysis based on three exclusive components (Logue et al. 2011); according
886 to the relative relevance of the variables used in this study (axes): spatial variables (x-axis),
887 environmental variables (y-axis) and the different dispersal capability of species (z-axis).
888 Circles represent the theoretical location where the emphasis of each approach is situated
889 across the three axes.

890 **Figure 2** Flow chart of the statistical analyses performed in this study.

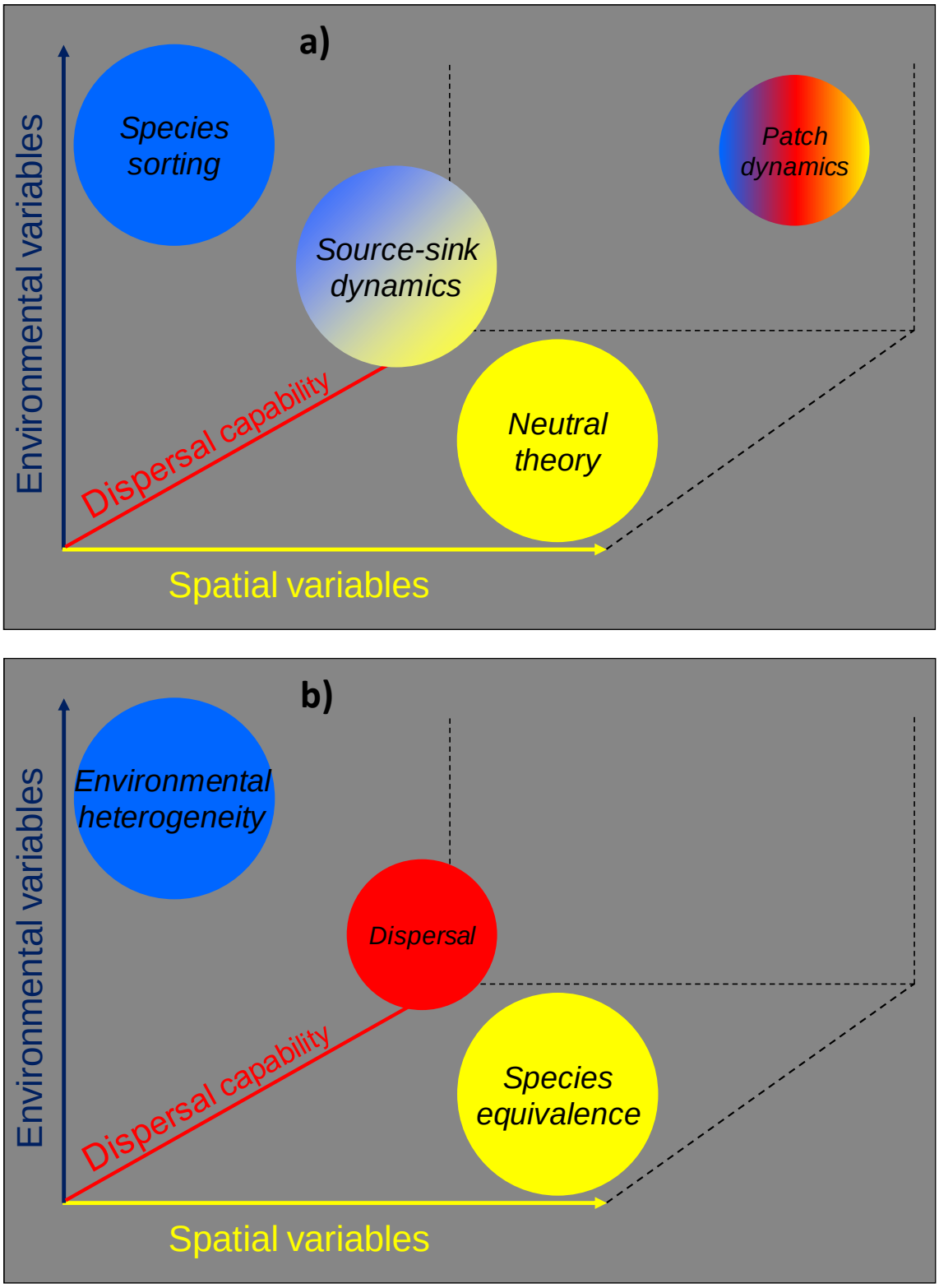
891 **Figure 3** Frequency of local environmental variables (top) and spatial variables (bottom)
892 selected as significant in explaining species distributions through binomial GLMs. Spatial
893 variables are arranged from small-scale (i.e. V13) to large-scale extent (i.e. V1), and include
894 V11 which was never selected. Climate variables are not shown (mean annual temperature
895 was selected 32 times, and July air temperature 20 times). Species-environment relationships
896 are shown in black when positive and grey when negative. Specific information for each
897 species can be found in Table S2.

898 **Figure 4** Species scores on taxonomic vectors TAX-PCO3 (a) and TAX-PCO4 (b), and on
899 trait vector TRA-PCO2 (c), arranged from lowest to highest values. For the trait vector, the
900 location of the different functional feeding (FFG) and habit trait groups (HTG), is indicated,
901 as well as the four different body size classes considered (BS, represented by columns of four
902 different sizes), and those insects considered as of high female dispersal potential (H).

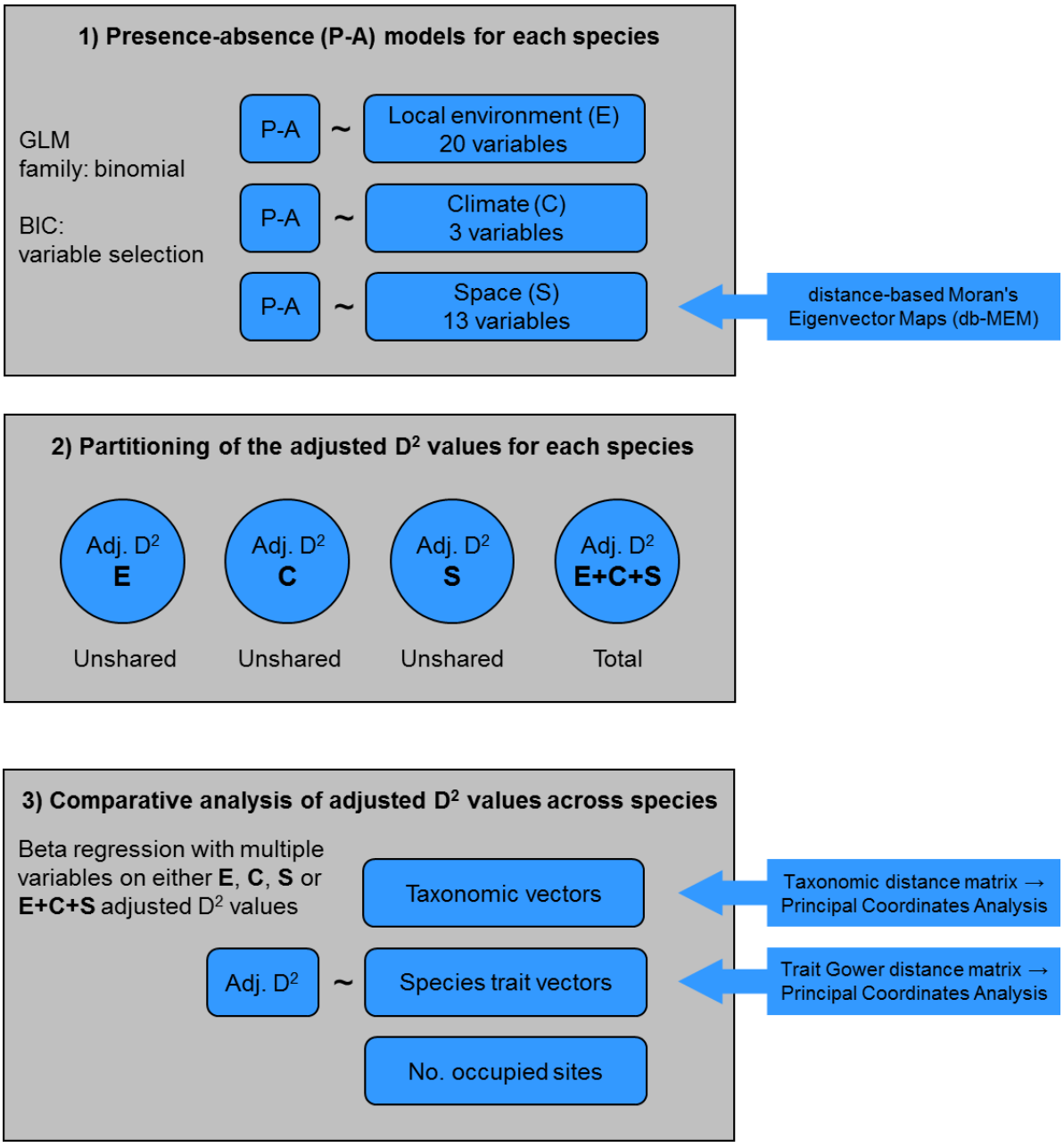
903

904 **Figure 1**

905



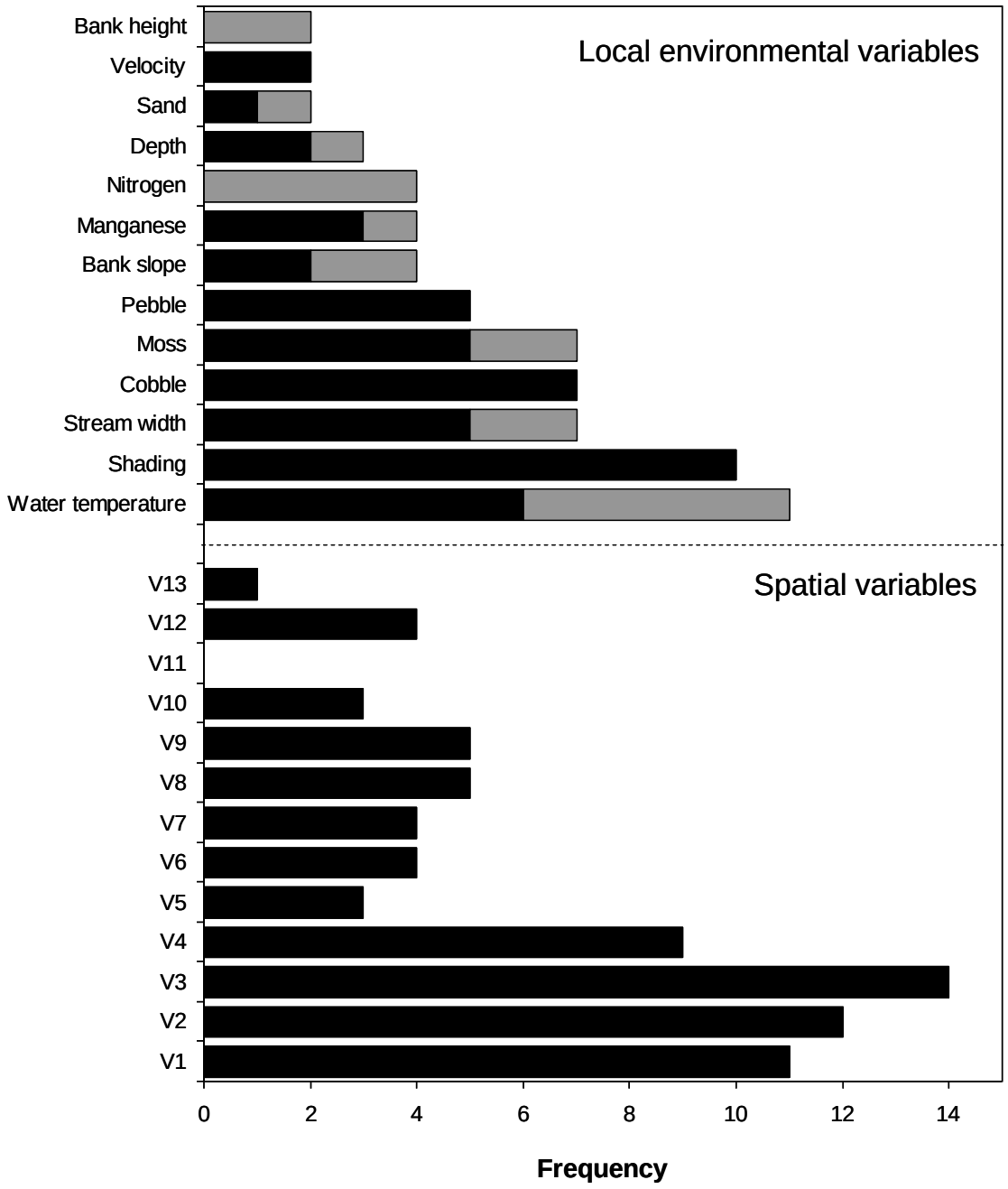
906 **Figure 2**



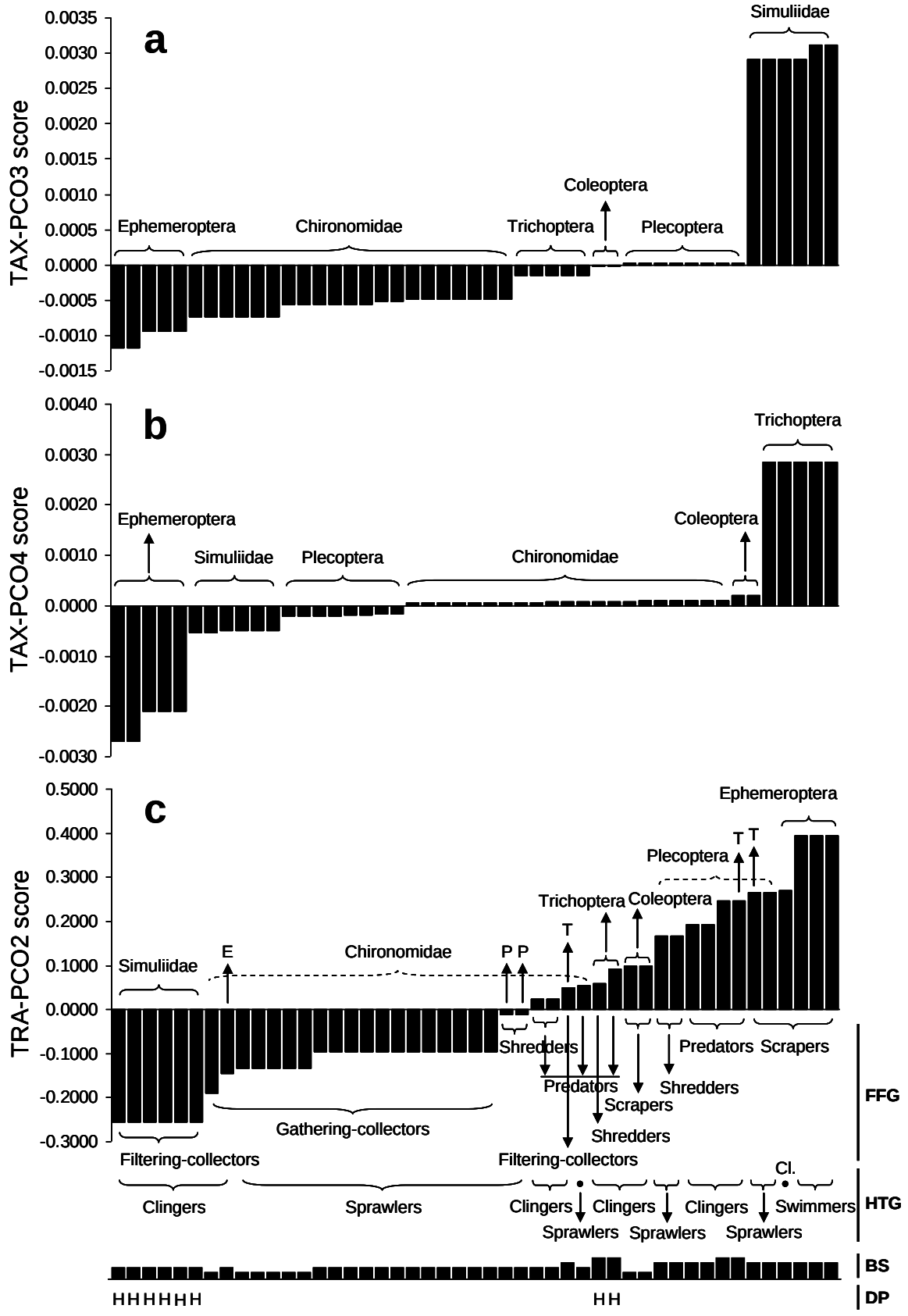
907

908 **Figure 3**

909



910 **Figure 4**


a
TAX-PCO3 score
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010
-0.0015
Ephemeroptera
Chironomidae
Trichoptera
Coleoptera
Plecoptera
Simuliidae
b
TAX-PCO4 score
0.0040
0.0030
0.0020
0.0010
0.0000
-0.0010
-0.0020
-0.0030
Ephemeroptera
Simuliidae
Plecoptera
Chironomidae
Coleoptera
Trichoptera
c
TRA-PCO2 score
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
-0.1000
-0.2000
-0.3000
Simuliidae
E
Chironomidae
P P
T
Trichoptera
Coleoptera
Plecoptera
T T
Ephemeroptera
Shredders
Predators
Scrapers
Predators
Shredders
Scrapers
Shredders
Gathering-collectors
Filtering-collectors
Filtering-collectors
Cl.
Clingers
Sprawlers
Clingers
Sprawlers
Clingers
Sprawlers
Clingers
Sprawlers
Swimmers
FFG
HTG
BS
DP
HHHHHH
HH