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

Guillermo de Mendoza¹²,⁰, Riikka Kaivosoja³, Mira Grönroos⁴, Jan Hjort³, Jari Ilmonen⁵,
Olli-Matti Kärnä³, Lauri Paasivirta⁶, Laura Tokola³ & Jani Heino⁷

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

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

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

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

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

⁶Ruuhikoskenkatu 17, FI-24240 Salo, Finland.

⁷Finnish 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

Metacommunity theory, single-species distribution models, comparative analysis, beta
regression, subarctic streams, stream macroinvertebrates, insects.
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 (Logue et al., 2011; Brown et al., 2017). 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 frameworks (i.e., Leibold et al., 2004).
Species distribution models have previously been used to predict community-level properties such as biodiversity (Ferrier & Guisan, 2006). Their accuracy in predicting community-level properties appears to be higher than that of community assembly models, although at a high cost in terms of model complexity (Chapman & Purse, 2011; Bonthoux, Baselga & Balent, 2013). The accuracy of single-species distribution modelling, however, may also be advantageous to test ecological theories about community assembly mechanisms. This is because accurately modelling the distribution of single species, one at a time, provides the opportunity to proceed with a subsequent comparative analysis across species. Using a comparative analysis, the variation in model performance can be related, for example, to species traits and potential phylogenetic constraints.

Stream insect species, in particular, are highly suitable to decipher community assembly processes through the comparative analysis of single-species distribution models (Heino & de Mendoza, 2016). This is because of the high variability among species in tolerance of environmental conditions, as well as resource exploitation, dispersal capability, and habit traits (Merritt & Cummins, 1996; Tachet et al., 2010; Schmidt-Kloiber & Hering, 2015; Serra et al., 2016). This variability is valuable in evaluating which community assembly mechanism dominates in each particular context of analysis. Basically, such an analysis might shed light into the relevance of environmental variables, spatial variables, and dispersal capability of species on model performance. Subsequently, this information can be used as an indicator of the preponderance of one community assembly mechanism over another (Fig. 1). For example, if many species show similar spatial patterns, and if these species share the same dispersal 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 (Heino, 2005; Grönroos et al., 2013; 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 (Fig. 1).
Methods

Study area

The field work for this study (Fig. S1, Supporting Information) was conducted in the Tenojoki drainage basin (main stem length: 361 km, basin area: 16377 km$^2$) in northernmost Finland (70°N, 27°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°C in the continental areas of the drainage basin, and close to 0°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 (Mansikkanemi, 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. czerepanowii) woodlands at low altitude and barren tundra at high 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-second subsamples that were divided between main habitats at a riffle site (ca. 50 m²) 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). The pooled 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 (Supporting Information). Nomenclature generally follows de Jong et al. (2014) and more specific references for the Simuliidae (Ilmonen, 2014; Adler & Crosskey, 2016).
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 et al. (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 (Hoffsten, 2004; Müller-Peddinghaus, 2011; Müller-Peddinghaus & Hering, 2013; Gíslason et al., 2015; Schmidt-Kloiber & Hering, 2015). Although such information about dispersal abilities of stream insects is rather simple, there is currently no better information available (Tachet et al., 2010; Schmidt-Kloiber & Hering, 2015; Serra et al., 2016). Functional feeding and habit trait groups generally follow Merritt & Cummins (1996).

Local environmental, climatic, and spatial variables

The 55 streams were surveyed during the early northern summer, between early and middle of June in 2012. We measured a set of local (i.e. proximal) environmental variables that have been found important for stream insects in northern drainage basins in previous studies
These comprised physical habitat and water physico-chemical variables. For physical habitat variables, we measured current velocity (m/s) and depth (cm) at 30 random spots in a riffle site. We also measured mean width of the riffle site based on five cross-channel measurements, evenly spaced across the surveyed riffle site. Bank height and bank slope were measured at the same locations with stream width measurements. Bank height was measured as the height of the lower stream bank, i.e. the height from the water level to the edge of terrestrial vegetation. Bank slope was measured (perpendicular to the stream) as a stream bank rise (cm) over 2 m starting from the edge of terrestrial vegetation. Moss cover (%) and particle size classes (%) were visually estimated at 10 squares (1 m²) at random locations in a riffle site. We used a modified Wentworth’s (1922) scale of particle size classes: sand (0.25-2 mm), gravel (2-16 mm), pebble (16-64 mm), cobble (64-256 mm) and boulder (256-1024 mm). Based on the visual estimates for each square, we calculated mean values for each particle size class and moss cover at a site and used these mean values in species distribution modelling. We also visually estimated shading (%) by riparian vegetation and proportion of riparian deciduous trees (%). For physico-chemical properties, we measured pH, conductivity and water temperature at each site in the field using a YSI device model 556 MPS (YSI Inc., Ohio, USA) and took additional water samples during the field campaign for further analysis. Water samples were frozen at the end of the day at the Kevo Field Station situated in the northern part of the study area, and were later analysed for total nitrogen, colour, iron and manganese in the laboratory of the Finnish Environment Institute in Oulu following Finnish national standards (National Board of Waters, 1981).

We also included three climatic variables, including annual air temperature sum above 5°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 et al., 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 et al., 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 (Fig. 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 \( r > 0.7 \)) predictor variables from the sets of local environment 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 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 (Fig. 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 environment (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)
(Fig. 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 et al., 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 (Fig. 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 et al., 2013), and eventually trait PCO vectors with the function “cmdscale”. The four trait eigenvectors obtained were considered for further statistical analyses. The variation in adjusted $D^2$ values across species that could be attributed to pure E, pure C, pure S, or E+C+S effects was fitted on site occupancy, the four taxonomic and four species trait vectors selected, using beta regression with the function “betareg” of the R package “betareg”
Beta regression is adequate when the response variable (in this case, the adjusted $D^2$ values) is constrained between 0 and 1.

We compared the explained variation by pure E, C, and S effects with a Kruskal-Wallis test, with additional Mann-Whitney tests for subsequent pair-wise comparisons between groups. Non-parametric tests were chosen since adjusted $D^2$ values data departed from normality following the Shapiro-Wilk test (Zar, 1984). We also analysed the univariate relationships between site occupancy, body size, dispersal potential, broad taxonomic insect groups, functional feeding groups, habit trait groups, and taxonomic and trait vectors.

Depending on the continuous (e.g. site occupancy) or categorical (e.g. habit trait group) nature of the variables involved, we followed Kruskal-Wallis tests, Mann-Whitney tests, Fisher’s exact test or Spearman correlations, as these variables were generally not normally distributed (Zar, 1984).

**Results**

*Single species models*

Local environmental and spatial effects accounted for a higher variation in species distributions (16.1% and 12.6% in average, respectively) than did climatic effects (5.4%) ($P < 0.001$, Kruskal-Wallis test), whereas the average deviance explained did not differ significantly between local environment and spatial effects ($P = 0.125$, Mann-Whitney test) (Table S3). The local environmental factors most frequently selected in explaining species distributions were water temperature, shading, and to a lesser extent, stream width, cobbles and moss (Fig. 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 was also shown in Fig. 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. $P < 0.05$) 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. $P < 0.10$), 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 ($P = 0.017$), but not TAX-PCO3, TAX-PCO4 or TRA-PCO2 ($P = 0.943$, $P = 0.175$, and $P = 0.449$, 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, $P = 0.036$, Fig. S3; TAX-PCO3, TAX-PCO4, and TRA-PCO2, $P = 0.760$, $P = 0.660$, and $P = 0.524$, 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 environment 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 ($P < 0.001$) to dispersal potential (Table S5, Fig. 4a). In contrast to this taxonomic vector, TAX-PCO4 showed the highest species scores for both blackflies and mayflies (Fig. 4b), and was strongly correlated to site occupancy ($P = 0.007$, Table S5). Finally, TRA-PCO2 reflects the influence of functional feeding groups and body size on model performance (Fig. 4c), as indicated by the strong correlation of both variables (i.e. $P < 0.001$) 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 underlying such distributions. There was no evident association between model accuracy and particular taxonomic groups (Table S3). Nevertheless, a few generalisations can be
highlighted with regard to the results obtained. For example, water temperature and shading, and to a lesser extent, stream width, cobbles and moss, were more relevant as environmental predictors of species distributions than stream flow or water chemistry variables (Table S3, Fig. 3). This is in line with the well-known influence of temperature and resource availability on insect life cycles at high latitudes (Danks, 2007) and indicates the influence of species sorting processes along these environmental gradients. Resource availability is represented in our case by shading, which indicates the proximity of terrestrial vegetation and hence is a surrogate of availability of allochthonous resources from terrestrial origin for aquatic insect larvae. This typically corresponds with a situation of a low-order stream which, as in our case, is influenced strongly by terrestrial material from riparian vegetation which is then taken as food resource by shredders, hence promoting their dominance (Vannote et al., 1980). Shading may also be inversely related to primary productivity, but in this study, we found that the relationship of species distribution with shading was always positive (Fig. 3), suggesting that rather than biofilm production, it is the external input of terrestrial material from riparian birch tree abundance what is likely driving species distributions. In our case, shading was selected as a significant variable in binomial models for some predators (Isoperla difformis and Plectrocnemia conspersa) and shredders (Leuctra spp.), for some collector-gatherers (Corynoneura lobata-type, Eukiefferiella devonica-group, Orthocladius rhyacobius-group and Tvetenia discoloripes), and for some collector-filterers (Philopotamus montanus and Prosimulium hirtipes) (Table S3). These latter groups perhaps benefit indirectly from the increase in potential resources that the variable “shading” represents for shredders, for example, through the enhancement of nutrient re-cycling by shredding coarse plant litter (Wallace & Webster, 1996; Covich, Palmer & Crowl, 1999).

Spatial variables were also relevant for the distributions of some species. Specifically, large-scale spatial variables were more important than small-scale variables in explaining
species distributions in our study (Table S3, Fig. 3). At a larger spatial extent (ca. 500 km latitudinal gradient), previous findings indicate a stronger relevance of environmental factors, compared to spatial restrictions, on single-species distributions (Heino & de Mendoza, 2016).

This is perhaps not surprising because increasing the spatial extent may have a strong positive effect on the relevance of niche processes through larger environmental gradients (Chase, 2014). However, increasing the spatial extent may also preclude species to reach environmentally suitable locations owing to dispersal limitation, and thus the relative contribution of both environmental and spatial constraints on species distributions does not always vary predictably with spatial scale (Alahuhta & Heino, 2013).

Comparative analysis across the species

Comparative analysis across the species models showed a clear relationship between model performance and site occupancy. Specifically, the binomial GLMs that we built upon spatial variables could be related to site occupancy, and to a lesser extent, to taxonomic and trait vectors, whereas none of these variables was significantly related to model performance when models were based on local environmental or climate variables (Table 1). At first glance, our results also suggested both a slight influence of female dispersal potential (related to the taxonomic vector TAX-PCO3), and a potential influence of functional feeding groups and body size (related to the trait vector TRA-PCO2), on the performance of models based on spatial variables. The taxonomic vector TAX-PCO3 perhaps relates to female dispersal potential, as species scores along this vector were much higher for the blackflies than for the rest of species, and lowest for the mayflies (Fig. 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 (Fig. 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 (Fig. 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.
Analysing through beta regression the univariate relationship of these variables with the
adjusted $D^2$ values of binomial GLMs based on spatial effects again resulted in site
occupancy as the only significant variable (see Results above). Therefore, we must conclude
that any potential effect of taxonomic and trait vectors on model performance, including the
effect of female dispersal potential and body size, and that of functional feeding groups, must
be considered with caution: their statistical significance only appears after controlling for site
occupancy and the other variables considered in the full model of beta regression. In this
regard, the fact that *Baetis rhodani* is a widespread mayfly, which could not be modelled
because it was present at all sites, also gives support to the idea that dispersal abilities are not
so important in structuring invertebrate assemblages in high-latitude drainage basins. This is
because it demonstrates that mayfly species can be widespread, despite being rather weak
active dispersers. We also acknowledge that the rarest species (i.e. present in less than six
sites) were not modelled because models based on such small number of presences were
considered unreliable (e.g. Pierce & Ferrier, 2000). However, excluding these species does
not undermine the conclusion that the distributions of most common species are better
accounted for by models based on spatial variables than that of not-so-common species. In
fact, we effectively modelled 47 out of the 86 taxa available at the species (most cases) or
species-group (few cases) taxonomic resolution, comprising 55% of cases, which is a
representative subset of species in the entire metacommunity.

Approaching the suitability of metacommunity analysis frameworks

With the information above about single-species distribution models and subsequent
comparative analysis across species, it is possible to proceed with the evaluation of the
suitability of the two different frameworks of metacommunity analysis (Fig. 1) considered
here: 1) the classical approach exemplified by the four different non-exclusive perspectives
described by Leibold *et al.* (2004) or 2) the three exclusive components as proposed by

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

Although species-sorting processes cannot be completely discarded because of the demonstrated influence of environmental variables in many cases, deviance partitioning suggests that the pure effects of environmental and spatial factors on species distributions are stronger than their joint effects (Table S3). Also, the effect of spatial variables was better explained than that of environmental factors by our explanatory variables, particularly site occupancy, in the comparative analysis. These results slightly undermine the idea of species-sorting across spatially structured environmental gradients as the most important process 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 (Tachet et al., 2010; Schmidt-Kloiber & Hering, 2015; Serra et al., 2016). Moreover, the different metacommunity paradigms from Leibold et al. (2004) may always act simultaneously to a certain extent along a continuum (Fig. 1a) rather than being distinct and mutually exclusive options (Gravel et al., 2006; Logue et al., 2011; Brown et al., 2017).

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 re-colonisation.

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 (Fig. 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 (Tachet et al., 2010; Schmidt-Kloiber & Hering, 2015; Serra et al., 2016). 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, Schmidt & Finn, 2009; Hughes, Huey & Schmidt, 2013). 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 (Thomson et al., 2002; Cadotte & Tucker, 2017). 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 (Sexton et al., 2009; Pigot & Tobias, 2013). Overall, this suggests that inter-specific interactions may also play a role in our case, although the abundances of insect larvae in subarctic streams are typically low (see also Heino & Grönroos, 2017) and may thus result in weak density-dependent interactions among species (see also Morin, 2011).
Our study considered tributary streams draining into two linear sub-elements of a larger river network (Fig. S1). However, there exists growing concern about the potential role of the entire dendritic river networks in shaping biodiversity patterns, community structure and species distributions (Altermatt, 2013). For example, the consideration of whole river networks may unveil a more preeminent role for spatial factors in community assembly, undermining the role of environmental filtering. Therefore, studies conducted across whole dendritic networks could be more in line with neutral theory, as shown by Muneepeerakul et al. (2008) for fish communities, yet no environmental variable was truly considered in that study. Although we focused on tributary streams draining into the main river, the consideration of whole dendritic networks may help us to perceive more accurately the real connectivity pathways between isolated patches. This connectivity may have consequences for metacommunity stability with respect to a situation where only a linear component of this network is acting (Fagan, 2002). Also, dispersal along dendritic networks implies more variability in local richness with strong consequences also for community differentiation among patches (Carrara et al., 2012; Seymour et al., 2015).

In the case of stream insects, the taxa considered and the taxonomic resolution 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), conservation plans may fall short and not result in desired outcomes.

Acknowledgements

We would like to 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 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


http://www.clemson.edu/cafls/biomia/pdfs/blackflyinventory.pdf.


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 ($P < 0.05$) are shown in boldface; marginally significant values ($P < 0.10$) in italics.

#### Adj. $D^2$ of E effects

<table>
<thead>
<tr>
<th>Estimate</th>
<th>SE</th>
<th>$z$</th>
<th>$P$</th>
<th>Log-likelihood</th>
<th>Pseudo $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.6002</td>
<td>0.2220</td>
<td>-7.209</td>
<td>$&lt;0.001$</td>
<td>0.2500</td>
</tr>
<tr>
<td>Site occupancy</td>
<td>-0.0042</td>
<td>0.0091</td>
<td>-0.469</td>
<td>0.646</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO1</td>
<td>-378.2072</td>
<td>323.7056</td>
<td>-1.168</td>
<td>0.243</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO2</td>
<td>91.4160</td>
<td>161.7362</td>
<td>0.565</td>
<td>0.572</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO3</td>
<td>225.8850</td>
<td>218.0393</td>
<td>1.036</td>
<td>0.300</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO4</td>
<td>-112.1618</td>
<td>101.9824</td>
<td>-1.100</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO1</td>
<td>-0.5557</td>
<td>0.8628</td>
<td>-0.644</td>
<td>0.520</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO2</td>
<td>1.3525</td>
<td>1.0712</td>
<td>1.263</td>
<td>0.207</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO3</td>
<td>0.3959</td>
<td>0.9475</td>
<td>0.418</td>
<td>0.676</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO4</td>
<td>-1.3110</td>
<td>1.3459</td>
<td>-0.974</td>
<td>0.330</td>
<td></td>
</tr>
</tbody>
</table>

#### Adj. $D^2$ of C effects

<table>
<thead>
<tr>
<th>Estimate</th>
<th>SE</th>
<th>$z$</th>
<th>$P$</th>
<th>Log-likelihood</th>
<th>Pseudo $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.9806</td>
<td>0.2682</td>
<td>-11.115</td>
<td>$&lt;0.001$</td>
<td>0.2309</td>
</tr>
<tr>
<td>Site occupancy</td>
<td>0.0029</td>
<td>0.0103</td>
<td>0.278</td>
<td>0.781</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO1</td>
<td>-579.9412</td>
<td>375.5543</td>
<td>-1.544</td>
<td>0.123</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO2</td>
<td>83.1028</td>
<td>185.9703</td>
<td>0.447</td>
<td>0.655</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO3</td>
<td>352.6843</td>
<td>252.8857</td>
<td>1.395</td>
<td>0.163</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO4</td>
<td>-1.9501</td>
<td>120.4650</td>
<td>-0.016</td>
<td>0.987</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO1</td>
<td>-1.6362</td>
<td>1.0045</td>
<td>-1.629</td>
<td>0.103</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO2</td>
<td>0.4327</td>
<td>1.2476</td>
<td>0.347</td>
<td>0.729</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO3</td>
<td>0.6454</td>
<td>1.1129</td>
<td>0.580</td>
<td>0.562</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO4</td>
<td>-2.1768</td>
<td>1.5519</td>
<td>-1.403</td>
<td>0.161</td>
<td></td>
</tr>
</tbody>
</table>

#### Adj. $D^2$ of S effects

<table>
<thead>
<tr>
<th>Estimate</th>
<th>SE</th>
<th>$z$</th>
<th>$P$</th>
<th>Log-likelihood</th>
<th>Pseudo $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.4436</td>
<td>0.2222</td>
<td>-10.997</td>
<td>$&lt;0.001$</td>
<td>0.2137</td>
</tr>
<tr>
<td>Site occupancy</td>
<td>0.0208</td>
<td>0.0083</td>
<td>2.498</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO1</td>
<td>309.6453</td>
<td>360.2543</td>
<td>0.860</td>
<td>0.390</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO2</td>
<td>-4.0452</td>
<td>162.3764</td>
<td>-0.025</td>
<td>0.980</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO3</td>
<td>365.5554</td>
<td>205.9633</td>
<td>1.775</td>
<td>0.076</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO4</td>
<td>217.5296</td>
<td>102.0912</td>
<td>2.131</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO1</td>
<td>-1.0051</td>
<td>0.8333</td>
<td>-1.206</td>
<td>0.228</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>z</td>
<td>P</td>
<td>Log-likelihood</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td>--------</td>
<td>----------------</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-0.7133</td>
<td>0.2331</td>
<td>-3.060</td>
<td>0.002</td>
<td>17.76</td>
</tr>
<tr>
<td>Site occupancy</td>
<td>0.0111</td>
<td>0.0095</td>
<td>1.164</td>
<td>0.244</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO1</td>
<td>-140.2227</td>
<td>352.1883</td>
<td>-0.398</td>
<td>0.691</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO2</td>
<td>32.5012</td>
<td>178.6572</td>
<td>0.182</td>
<td>0.856</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO3</td>
<td>407.2860</td>
<td>228.4095</td>
<td>1.783</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>TAX-PCO4</td>
<td>81.0016</td>
<td>110.2921</td>
<td>0.734</td>
<td>0.463</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO1</td>
<td>-1.3052</td>
<td>0.9142</td>
<td>-1.428</td>
<td>0.153</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO2</td>
<td>1.6446</td>
<td>1.1935</td>
<td>1.378</td>
<td>0.168</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO3</td>
<td>0.9742</td>
<td>1.0367</td>
<td>0.940</td>
<td>0.347</td>
<td></td>
</tr>
<tr>
<td>TRA-PCO4</td>
<td>-1.0409</td>
<td>1.4575</td>
<td>-0.714</td>
<td>0.475</td>
<td></td>
</tr>
</tbody>
</table>
Figure legends

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

Fig. 2. Flow chart of the statistical analyses performed in this study.

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

Fig. 4. Species scores on taxonomic vectors TAX-PCO3 (a) and TAX-PCO4 (b), and on trait vector TRA-PCO2 (c), arranged from lowest to highest values. For the trait vector, the location of the different functional feeding (FFG) and habit trait groups (HTG), is indicated, as well as the four different body size classes considered (BS, represented by columns of four different sizes), and those insects considered as of high female dispersal potential (H).
Fig. 1.

Spatial variables

Environmental variables

Species sorting

Source-sink dynamics

Neutral theory

Patch dynamics

Environmental heterogeneity

Dispersal capability

Dispersal

Species equivalence

Spatial variables
Fig. 2.

1) Presence-absence (P-A) models for each species

- GLM family: binomial
- BIC: variable selection
- P-A ~ Local environment (E) 20 variables
- P-A ~ Climate (C) 3 variables
- P-A ~ Space (S) 13 variables

2) Partitioning of the adjusted $D^2$ values for each species

- Adj. $D^2_E$ Unshared
- Adj. $D^2_C$ Unshared
- Adj. $D^2_S$ Unshared
- Adj. $D^2_{E+C+S}$ Total

3) Comparative analysis of adjusted $D^2$ values across species

- Beta regression with multiple variables on either E, C, S or E+C+S adjusted $D^2$ values
- Adj. $D^2$ ~ Species trait vectors
- No. occupied sites

- Taxonomic distance matrix → Principal Coordinates Analysis
- Trait Gower distance matrix → Principal Coordinates Analysis
Fig. 3. 

Local environmental variables

Spatial variables

Frequency
Fig. 4.