1 **Highly variable species distribution models in a subarctic stream**

2 **metacommunity: patterns, mechanisms and implications**

3

4 Guillermo de Mendoza[1,2,*], Riikka Kaivosoja[3], Mira Grönroos[4], Jan Hjort[3], Jari Ilmonen[5],

5 Olli-Matti Kärnä[3], Lauri Paasivirta[6], Laura Tokola[3] & Jani Heino[7]

6

7 [1]Centre for Advanced Studies of Blanes, Spanish National Research Council (CEAB-CSIC),

8 Accés a la Cala St. Francesc 14, ES-17300 Blanes, Spain.

9 [2]Laboratoire GEODE, UMR 5602 CNRS, Université Toulouse-Jean Jaurès, 5 allées Antonio

10 Machado, FR-31058 Toulouse, France.

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

12 [4]University of Helsinki, Department of Environmental Sciences, Section of Environmental

13 Ecology, Niemenkatu 73, FI-15140 Lahti, Finland.

14 [5]Metsähallitus, Natural Heritage Services, P.O. Box 94, FI-01301 Vantaa, Finland.

15 [6]Ruuhikoskenkatu 17, FI-24240 Salo, Finland.

16 [7]Finnish Environment Institute, Natural Environment Centre, Biodiversity, Paavo Havaksen

17 Tie 3, FI-90570 Oulu, Finland.

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

19

20   **Summary**

21   1. Metacommunity theory focuses on assembly patterns in ecological communities,

22   originally exemplified through four different, yet non-exclusive, perspectives: patch

23   dynamics, species sorting, source-sink dynamics, and neutral theory. More recently, three

24   exclusive components have been proposed to describe a different metacommunity

25   framework: habitat heterogeneity, species equivalence, and dispersal. Here, we aim at

26   evaluating the insect metacommunity of a subarctic stream network under these two

27   different frameworks.

28   2. We first modelled the presence/absence of 47 stream insects in northernmost Finland

29   using binomial generalised linear models (GLMs). The deviance explained by pure local

30   environmental (E), spatial (S), and climatic variables (C) was then analysed across

31   species using beta regression. In this comparative analysis, site occupancy, as well as

32   taxonomic and biological trait vectors obtained from principal coordinate analysis, were

33   used as predictor variables.

34   3. Single-species distributions were better explained by in-stream environmental and spatial

35   factors than by climatic forcing, but in a highly variable fashion. This variability was

36   difficult to relate to the taxonomic relatedness among species or their biological trait

37   similarity. Site occupancy, however, was related to model performance of the binomial

38   GLMs based on spatial effects: as populations are likely to be better connected for

39   common species due to their near ubiquity, spatial factors may also explain better their

40   distributions.

41   4. According to the classical four-perspective framework, the observation of both

42   environmental and spatial effects suggests a role for either mass effects or species sorting

43   constrained by dispersal limitation, or both. Taxonomic and biological traits, including

44   the different dispersal capability of species, were scarcely important, which undermines

45 the patch dynamics perspective, based on differences in dispersal ability between species.

46 The highly variable performance of models makes the reliance on an entirely neutral

47 framework unrealistic as well. According to the three-component framework, our results

48 suggest that the stream insect metacommunity is shaped by the effect of habitat

49 heterogeneity (supporting both species-sorting and mass effects), rather than species

50 equivalence or dispersal limitation.

51 5. While the relative importance of the source-sink dynamics perspective or the species-

52 sorting paradigm cannot be deciphered with the data at our disposal, we can conclude that

53 habitat heterogeneity is an important driver shaping species distributions and insect

54 assemblages in subarctic stream metacommunities. These results exemplify that the use of

55 the three-component metacommunity framework may be more useful than the classical

56 four perspective paradigm in analysing metacommunities. Our findings also provide

57 support for conservation strategies based on the preservation of heterogeneous habitats in

58 a metacommunity context.

59

60 **Key-words**

61 Metacommunity theory, single-species distribution models, comparative analysis, beta

62 regression, subarctic streams, stream macroinvertebrates, insects.

63

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

88 vs Logue *et al*., 2011), specifically so as to evaluate which of the two is more adequate for

89 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 (Chapman & Purse, 2011; Bonthoux,

94 Baselga & Balent, 2013). 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 *et al*., 2010; Schmidt-Kloiber & Hering,

105 2015; Serra *et al*., 2016). This variability is valuable in evaluating which community

106 assembly mechanism dominates in each particular context of analysis. Basically, such an

107 analysis might shed light into the relevance of environmental variables, spatial variables, and

108 dispersal capability of species on model performance. Subsequently, this information can be

109 used as an indicator of the preponderance of one community assembly mechanism over

110 another (Fig. 1). For example, if many species show similar spatial patterns, and if these

111 species share the same dispersal potential, we can presume that the ability to disperse may be

112 underlying the observed general pattern for these species. This would give us hints about the

113   adequacy to consider one particular metacommunity theory perspective over the others.

114   Within the classical four-perspective framework (Leibold *et al*., 2004), patch dynamics would

115   likely be suitable in this case, as this perspective relies on the different capability of species

116   to both disperse and exploit resources. Within the metacommunity framework based on three

117   exclusive components (Logue *et al*., 2011), dispersal would be main driver in this case.

118   Moreover, stream insects are also a diverse group of species, which belong to different insect

119   orders and vary widely in physiological and morphological adaptations (Merritt & Cummins,

120   1996). Thus, modelling the distribution of single stream insect species and subsequently

121   proceeding with a comparative analysis across species is also a suitable indirect practice to

122   explore possible evolutionary constraints on community assembly processes.

123   In this study, we analysed the distribution of common stream insect species in the

124   metacommunity of a subarctic drainage basin. Species differ widely in their dispersal

125   capability (e.g. passive or active dispersers, aquatic or aerial adults) and tolerance of

126   environmental conditions such as temperature, water flow, or habitat characteristics (Heino,

127   2005; Grönroos *et al*., 2013; Heino & Grönroos, 2014). We used environmental, climatic and

128   spatial variables as predictors of the distributions of single stream insect species. Our aim was

129   to elucidate, first, whether or not environmental and spatial factors are relevant for explaining

130   the distribution of stream insect species; and second, whether or not the obtained models can

131   be related to the different dispersal capability, site occupancy (i.e. a gradient of rarity-

132   commonness), and biological and taxonomic traits, of stream insect species. Both

133   considerations were used to evaluate which of the two different metacommunity frameworks,

134   either the one based on four non-exclusive perspectives (Leibold *et al*., 2004) or the one

135   based on three exclusive axes (Logue *et al*., 2011), is more adequate to interpret our

136   observations of single species distributions in stream networks (Fig. 1).

137

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 (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 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

163   2012 (see below). The sample for each site consisted of six 30-second subsamples that were

164   divided between main habitats at a riffle site (ca. 50 m^2) based on visual inspections of

165   variation in depth, flow, moss cover and particle size. The six subsamples were pooled in the

166   field to obtain a composite sample. Such a sampling method has been shown to be effective

167   in northern streams, allowing to detect patterns in community structure (Heino, Ilmonen &

168   Paasivirta, 2014) and distributions of single species (Heino & de Mendoza, 2016). The

169   pooled samples were immediately preserved in ethanol in the field and were taken to the

170   laboratory for further processing and identification. Animals were separated from detritus and

171   moss fragments and identified to the lowest possible taxonomic level, mostly species (Kärnä

172   *et al*., 2015).

173

174   *Species considered and species traits*

175   We detected 107 insect taxa, of which 87 could be taxonomically determined to species or

176   species group (Kärnä *et al*., 2015). Insects determined to genus level were discarded as they

177   were considered too likely to include a few species, which is inappropriate to model single-

178   species distributions. Then, we focused on 48 species that occurred at more than 10% of the

179   55 study sites, that is, that occurred in at least six sites. This is because modelling the

180   distribution of species present in less than six sites is likely to produce spurious results and

181   therefore the analysis of these species was considered unreliable (e.g. Pearce & Ferrier,

182   2000). In practice, we could model the occupancies of only 47 species because the mayfly

183   *Baetis rhodani* occurred at all sites, so we could not use this species to model

184   presence/absence. The 47 stream insect species considered in this study are listed in Table S1

185   (Supporting Information). Nomenclature generally follows de Jong *et al*. (2014) and more

186   specific references for the Simuliidae (Ilmonen, 2014; Adler & Crosskey, 2016).

187   Body size class, dispersal potential, functional feeding groups and habit trait groups

188   were considered as species traits (Table S2). Functional feeding groups refer to exploitation

189   of different resources, while habit traits define modes of locomotion and attachment to

190   substrate (Merritt & Cummins, 1996). Body size classes and female dispersal potential

191   followed a previous study (Heino & de Mendoza, 2016), with additional information from

192   Tachet *et al.* (2010), Schmidt-Kloiber & Hering (2015) and Serra *et al.* (2016). Female

193   dispersal potential was characterised as being "low" or "high". In general, all species of the

194   Simuliidae were considered to have high dispersal potential, owing to the fact that their

195   females feed as flying adults, in most cases searching for blood of vertebrates, and hence

196   were assumed here to generally persist much longer as active flyers than the rest of species.

197   In this regard, Baldwin *et al.* (1975) often found their marked Simuliidae females several

198   kilometers away from their natal streams. Owing to their small size, the Simuliidae may also

199   be distributed long distances passively by wind (Crosskey, 1990). All other species were

200   considered as weak dispersers except for the caddisflies *Plectrocnemia conspersa* and

201   *Potamophylax cingulatus*, according to the information available for these taxa from previous

202   studies (Hoffsten, 2004; Müller-Peddinghaus, 2011; Müller-Peddinghaus & Hering, 2013;

203   Gíslason *et al.*, 2015; Schmidt-Kloiber & Hering, 2015). Although such information about

204   dispersal abilities of stream insects is rather simple, there is currently no better information

205   available (Tachet *et al.*, 2010; Schmidt-Kloiber & Hering, 2015; Serra *et al.*, 2016).

206   Functional feeding and habit trait groups generally follow Merritt & Cummins (1996).

207

208   *Local environmental, climatic, and spatial variables*

209   The 55 streams were surveyed during the early northern summer, between early and middle

210   of June in 2012. We measured a set of local (i.e. proximal) environmental variables that have

211   been found important for stream insects in northern drainage basins in previous studies

212 (Heino *et al.*, 2014; Kärnä *et al.*, 2015). These comprised physical habitat and water physico-

213 chemical variables. For physical habitat variables, we measured current velocity (m/s) and

214 depth (cm) at 30 random spots in a riffle site. We also measured mean width of the riffle site

215 based on five cross-channel measurements, evenly spaced across the surveyed riffle site.

216 Bank height and bank slope were measured at the same locations with stream width

217 measurements. Bank height was measured as the height of the lower stream bank, i.e. the

218 height from the water level to the edge of terrestrial vegetation. Bank slope was measured

219 (perpendicular to the stream) as a stream bank rise (cm) over 2 m starting from the edge of

220 terrestrial vegetation. Moss cover (%) and particle size classes (%) were visually estimated at

221 10 squares (1 m^2) at random locations in a riffle site. We used a modified Wentworth's

222 (1922) scale of particle size classes: sand (0.25-2 mm), gravel (2-16 mm), pebble (16-64

223 mm), cobble (64-256 mm) and boulder (256-1024 mm). Based on the visual estimates for

224 each square, we calculated mean values for each particle size class and moss cover at a site

225 and used these mean values in species distribution modelling. We also visually estimated

226 shading (%) by riparian vegetation and proportion of riparian deciduous trees (%). For

227 physico-chemical properties, we measured pH, conductivity and water temperature at each

228 site in the field using a YSI device model 556 MPS (YSI Inc., Ohio, USA) and took

229 additional water samples during the field campaign for further analysis. Water samples were

230 frozen at the end of the day at the Kevo Field Station situated in the northern part of the study

231 area, and were later analysed for total nitrogen, colour, iron and manganese in the laboratory

232 of the Finnish Environment Institute in Oulu following Finnish national standards (National

233 Board of Waters, 1981).

234 We also included three climatic variables, including annual air temperature sum above

235 5°C (growing degree days), mean annual air temperature and mean July air temperature for

236 the period 1981-2010. These variables were calculated in ArcMap 10.2 for each site from a

237 gridded (1 x 1 km) climate data provided by the Finnish Meteorological Institute (Pirinen *et*

238 *al*., 2012). The gridded climate data were produced using meteorological station observations

239 and Kriging interpolation (e.g. Aalto *et al*., 2013). The selected climatic variables are likely

240 to be important for the distributions of insects in this subarctic area, where temperature is

241 closely associated with insect life cycles (Danks, 2007).

242   Spatial variables were distance-based Moran's Eigenvector Maps (db-MEM) based on

243 geographical distances among sites (Dray, Legendre & Peres-Neto, 2006). These spatial db-

244 MEM variables were obtained with the function "PCNM" of the R package "PCNM"

245 (Legendre *et al*., 2013; R Core Team, 2013). We used the largest distance in the minimum

246 spanning tree, keeping all sites connected, as the truncation threshold. Spatial db-MEM

247 variables represent structures of autocorrelation at all spatial scales. Only those spatial db-

248 MEM variables showing significant positive autocorrelation were included in subsequent

249 modelling (Borcard, Gillet & Legendre, 2011), resulting in 13 spatial variables (Fig. 2).

250 Based on eigenvalues and bubble plot maps, the spatial variables can be divided into those

251 ranging from large-scale spatial structures (e.g. V1, V2) and those showing very small scale

252 spatial patterns (e.g. V12, V13).

253   Prior to modelling species distribution, we eliminated strongly correlated (i.e. Pearson

254 $r > 0.7$) predictor variables from the sets of local environment and climatic variables (see

255 Dormann *et al*., 2013). Hence, we removed one variable (i.e. annual temperature sum) from

256 the climatic variables and four variables (i.e. water iron, colour, conductivity and boulders)

257 from the stream environmental variables. The spatial variables were already not mutually

258 correlated (Borcard *et al*., 2011).

259

260 *Modelling species distributions*

261   The distribution (i.e. presence/absence) of each species was modelled using binomial

262   generalised linear models (i.e. binomial GLMs with logit link function), using separately

263   local environmental, climatic and spatial variables, with the package "*Rcmdr*" (Fox, 2005).

264   The deviance explained for each species was thus obtained for each binomial GLM with each

265   of these three different subsets of variables (Fig. 2). The variables selected for each species'

266   model were based on forward selection and Bayesian Information Criterion (BIC), separately

267   for each variable group (i.e. environmental, climate and spatial). BIC values were used

268   because they prevented the selection of too complex models in our case, in contrast to AIC

269   (results not shown), which is often the case under large sample sizes (Burnham & Anderson,

270   2004). Moreover, the target model under BIC selection does not depend on sample size, in

271   contrast to AIC (Burnham & Anderson, 2004). Therefore, AIC may be problematic in our

272   case as we aim at comparing model performance between species, which may differ in the

273   number of presences and absences. Also, deviating observations were removed from some

274   species' models if they had Cook's distance values > 1 and hence affected profoundly a few

275   models (Cook, 1977). For environmental variables, we registered whether the effect was

276   positive or negative on species distributions. We then used the selected variables of these

277   three subsets (i.e. local environmental, climatic and spatial) to perform variation (deviance)

278   partitioning by subtraction, similarly as performed in multivariate contexts (Legendre &

279   Legendre, 2012). Specifically, the deviance accounted for subset A, subset B, and subset A

280   and B together, was computed, so as to obtain the different fractions of variation solely

281   explained by each subset (i.e. unshared with other subsets). We eventually obtained adjusted

282   D^2 values (Guisan & Zimmermann, 2000; Legendre & Legendre, 2012) which could be

283   attributed to pure local environment (E), climatic (C) or spatial effects (S), as well as to total

284   effects combining the three subsets of pure effects and their joint effects (E+C+S effects)

285   (Fig. 2). Modelling methods other than GLMs could have been possible, yet species probably

286   show linear responses to the environmental predictors due to the fact that they are on the edge

287   of their geographical and ecological distributions, making GLMs adequate. Adding quadratic

288   terms to binomial models is unlikely to change results substantially in these situations (e.g.

289   Pulido *et al.*, 2015), and increase the difficulty of interpretation of the results. Also, deviance

290   partitioning is easy to accomplish when this is based on GLMs.

291

292   *Comparative analysis across species*

293   We performed a comparative analysis across species using beta regression (Ferrari & Cribari-

294   Neto, 2004), where the adjusted D^2 values obtained with previous binomial GLMs were used

295   as the dependent variable to be explained by site occupancy, taxonomic vectors or species

296   trait vectors (Fig. 2). These vectors were obtained separately from Principal Coordinate

297   Analysis (PCO). Using the taxonomic relatedness of species, a taxonomic relatedness matrix

298   was built using the function "taxa2dist" in the R package "*vegan*" (Oksanen *et al.*, 2013), and

299   taxonomic vectors were handled as continuous PCO vectors with the function "pco" in the R

300   package "*ecodist*" (Goslee & Urban, 2007). The first four taxonomic eigenvectors were

301   selected as these had much higher eigenvalues than the rest (Fig. S2). Similarly, species trait

302   vectors were also computed using body size class, dispersal potential, functional feeding

303   groups, and habit trait groups (Table S2). Species traits were considered as regular factors,

304   except body size class which was considered as an ordered factor, to obtain a distance matrix

305   based on Gower's metric with the function "daisy" of the R package "*cluster*" (Maechler *et*

306   *al.*, 2013), and eventually trait PCO vectors with the function "cmdscale". The four trait

307   eigenvectors obtained were considered for further statistical analyses. The variation in

308   adjusted D^2 values across species that could be attributed to pure E, pure C, pure S, or E+C+S

309   effects was fitted on site occupancy, the four taxonomic and four species trait vectors

310   selected, using beta regression with the function "betareg" of the R package "*betareg*"

311 (Cribari-Neto & Zeileis, 2010). Beta regression is adequate when the response variable (in

312 this case, the adjusted D^2 values) is constrained between 0 and 1.

313   We compared the explained variation by pure E, C, and S effects with a Kruskal-

314 Wallis test, with additional Mann-Whitney tests for subsequent pair-wise comparisons

315 between groups. Non-parametric tests were chosen since adjusted D^2 values data departed

316 from normality following the Shapiro-Wilk test (Zar, 1984). We also analysed the univariate

317 relationships between site occupancy, body size, dispersal potential, broad taxonomic insect

318 groups, functional feeding groups, habit trait groups, and taxonomic and trait vectors.

319 Depending on the continuous (e.g. site occupancy) or categorical (e.g. habit trait group)

320 nature of the variables involved, we followed Kruskal-Wallis tests, Mann-Whitney tests,

321 Fisher's exact test or Spearman correlations, as these variables were generally not normally

322 distributed (Zar, 1984).

323

324 **Results**

325 *Single species models*

326 Local environmental and spatial effects accounted for a higher variation in species

327 distributions (16.1% and 12.6% in average, respectively) than did climatic effects (5.4%) (P

328 < 0.001, Kruskal-Wallis test), whereas the average deviance explained did not differ

329 significantly between local environment and spatial effects ($P = 0.125$, Mann-Whitney test)

330 (Table S3). The local environmental factors most frequently selected in explaining species

331 distributions were water temperature, shading, and to a lesser extent, stream width, cobbles

332 and moss (Fig. 3). The spatial variables most often selected were better represented by large-

333 scale spatial variables within the Tenojoki drainage basin (e.g. V1, V2) than by small-spatial

334 scale variables (e.g. V12, V13), as was also shown in Fig. 3. Amongst the climate variables,

335   mean annual temperature was significant in explaining the distribution of 32 species, and July

336   air temperature of 20 species (not shown in Fig. 3).

337   The adjusted deviance explained by binomial GLMs was highly variable across

338   species and difficult to relate to particular taxonomic groups (Table S3). For example, local

339   environmental effects were particularly relevant for the stonefly *Siphonoperla burmeisteri*

340   (i.e. accounting for 66.7% of adjusted D^2 values), the mayfly *Heptagenia dalecarlica*

341   (50.2%), and the blackfly *Prosimulium hirtipes* (37.7%), whereas spatial effects were most

342   relevant for the caddisfly *Rhyacophila nubila* (41.2%), the stonefly *Brachyptera risi* (31.5%),

343   and the chironomid midge *Cardiocladius capucinus* (28.6%). Climate effects were also

344   highly variable. They were generally low (see above), and accounted for more than 20% of

345   adjusted D^2 values in only three cases: the stoneflies *Diura nanseni* and *Siphonoperla*

346   *burmeisteri* (34.4% and 20.1%, respectively), and the chironomid midge *Orthocladius*

347   *rivicola* (28.9%). Combining all effects, binomial GLMs explained on average 37.8% of the

348   null deviance (Table S3).

349

350   *Comparative analysis across species models*

351   The highly variable species-local environment and species-climate relationships in binomial

352   GLMs were not accounted for by site occupancy, or by taxonomic and trait vectors, in the

353   beta regression analysis (Table 1). The deviance explained by spatial variables was, however,

354   significantly (i.e. $P < 0.05$) accounted for by site occupancy (Table 1). The influence of

355   TAX-PCO4 and TRA-PCO2 on the adjusted D^2 values predicted by spatial effects in

356   binomial GLMs was significant as well. Also, the influence of TAX-PCO3 was marginally

357   significant (i.e. $P < 0.10$), remaining like this in the binomial GLMs based on all variables

358   combined (Table 1). However, when repeating the beta regression analysis by using only the

359   significant variables selected (i.e. site occupancy, TAX-PCO3, TAX-PCO4, and TRA-

360 PCO2), only site occupancy was statistically significant ($P = 0.017$), but not TAX-PCO3,

361 TAX-PCO4 or TRA-PCO2 ($P = 0.943$, $P = 0.175$, and $P = 0.449$, respectively, results not

362 shown in Table 1). Analysing through beta regression the univariate relationship of these

363 variables with the adjusted D^2 values of binomial GLMs based on spatial effects produced a

364 similar result (site occupancy, $P = 0.036$, Fig. S3; TAX-PCO3, TAX-PCO4, and TRA-PCO2,

365 $P = 0.760$, $P = 0.660$, and $P = 0.524$, respectively, results not shown). This univariate

366 relationship between the adjusted D^2 values and site occupancy was not observed when the

367 adjusted D^2 values of binomial GLMs were referred to environment or climate effects (Fig.

368 S3). No statistical significance was observed either for univariate relationships between

369 separate species traits and the adjusted D^2 values in binomial GLMs, with the sole exception

370 of body size (Fig. S3).

371   The TAX-PCO3 vector showed the highest species scores for blackflies (Simuliidae)

372 and the lowest for mayflies (Ephemeroptera), and was strongly correlated ($P < 0.001$) to

373 dispersal potential (Table S5, Fig. 4a). In contrast to this taxonomic vector, TAX-PCO4

374 showed the highest species scores for both blackflies and mayflies (Fig. 4b), and was strongly

375 correlated to site occupancy ($P = 0.007$, Table S5). Finally, TRA-PCO2 reflects the influence

376 of functional feeding groups and body size on model performance (Fig. 4c), as indicated by

377 the strong correlation of both variables (i.e. $P < 0.001$) with this trait vector (Table S5).

378

379 **Discussion**

380 *Single species models*

381 Our results indicated that single species distributions of stream insects are highly variable in

382 terms of predictability, as well as the significant environmental and spatial predictors

383 underlying such distributions. There was no evident association between model accuracy and

384 particular taxonomic groups (Table S3). Nevertheless, a few generalisations can be

385   highlighted with regard to the results obtained. For example, water temperature and shading,

386   and to a lesser extent, stream width, cobbles and moss, were more relevant as environmental

387   predictors of species distributions than stream flow or water chemistry variables (Table S3,

388   Fig. 3). This is in line with the well-known influence of temperature and resource availability

389   on insect life cycles at high latitudes (Danks, 2007) and indicates the influence of species

390   sorting processes along these environmental gradients. Resource availability is represented in

391   our case by shading, which indicates the proximity of terrestrial vegetation and hence is a

392   surrogate of availability of allochthonous resources from terrestrial origin for aquatic insect

393   larvae. This typically corresponds with a situation of a low-order stream which, as in our

394   case, is influenced strongly by terrestrial material from riparian vegetation which is then

395   taken as food resource by shredders, hence promoting their dominance (Vannote *et al*., 1980).

396   Shading may also be inversely related to primary productivity, but in this study, we found

397   that the relationship of species distribution with shading was always positive (Fig. 3),

398   suggesting that rather than biofilm production, it is the external input of terrestrial material

399   from riparian birch tree abundance what is likely driving species distributions. In our case,

400   shading was selected as a significant variable in binomial models for some predators

401   (*Isoperla difformis* and *Plectrocnemia conspersa*) and shredders (*Leuctra* spp.), for some

402   collector-gatherers (*Corynoneura lobata*-type, *Eukiefferiella devonica*-group, *Orthocladius*

403   *rhyacobius*-group and *Tvetenia discoloripes*), and for some collector-filterers (*Philopotamus*

404   *montanus* and *Prosimulium hirtipes*) (Table S3). These latter groups perhaps benefit

405   indirectly from the increase in potential resources that the variable "shading" represents for

406   shredders, for example, through the enhancement of nutrient re-cycling by shredding coarse

407   plant litter (Wallace & Webster, 1996; Covich, Palmer & Crowl, 1999).

408   Spatial variables were also relevant for the distributions of some species. Specifically,

409   large-scale spatial variables were more important than small-scale variables in explaining

410 species distributions in our study (Table S3, Fig. 3). At a larger spatial extent (ca. 500 km

411 latitudinal gradient), previous findings indicate a stronger relevance of environmental factors,

412 compared to spatial restrictions, on single-species distributions (Heino & de Mendoza, 2016).

413 This is perhaps not surprising because increasing the spatial extent may have a strong positive

414 effect on the relevance of niche processes through larger environmental gradients (Chase,

415 2014). However, increasing the spatial extent may also preclude species to reach

416 environmentally suitable locations owing to dispersal limitation, and thus the relative

417 contribution of both environmental and spatial constraints on species distributions does not

418 always vary predictably with spatial scale (Alahuhta & Heino, 2013).

419

420 *Comparative analysis across species*

421 Comparative analysis across the species models showed a clear relationship between model

422 performance and site occupancy. Specifically, the binomial GLMs that we built upon spatial

423 variables could be related to site occupancy, and to a lesser extent, to taxonomic and trait

424 vectors, whereas none of these variables was significantly related to model performance

425 when models were based on local environmental or climate variables (Table 1). At first

426 glance, our results also suggested both a slight influence of female dispersal potential (related

427 to the taxonomic vector TAX-PCO3), and a potential influence of functional feeding groups

428 and body size (related to the trait vector TRA-PCO2), on the performance of models based on

429 spatial variables. The taxonomic vector TAX-PCO3 perhaps relates to female dispersal

430 potential, as species scores along this vector were much higher for the blackflies than for the

431 rest of species, and lowest for the mayflies (Fig. 4). Blackflies are possibly the best active

432 dispersers among all the insects we considered, because females feed as flying adults and in

433 most species they must actively search for blood meals, often several kilometers away from

434 their natal streams (Baldwin *et al*., 1975). However, adult mayflies, do not feed and often

435   have extremely short life spans (Brittain, 1990). Therefore, it seems reasonable to assume that

436   blackflies may actively disperse better than mayflies. Site occupancy and dispersal potential

437   were not correlated (Table S5), and both taxa were the ones with highest number of sites

438   occupied (Fig. S4). In contrast, mayflies differed in site occupancy from non-biting midges

439   (Chironomidae) (Fig. S4), despite species in both groups can be considered weak active

440   dispersers, as chironomid adults are also short-lived and generally weak active fliers

441   (Armitage, 1995). On the other hand, the trait vector TRA-PCO2 suggests an influence of

442   feeding behaviour and body size (Fig. 4, Table S5) on model performance. This is because

443   the exploitation of food resource from terrestrial origin (i.e. shredders) would facilitate the

444   development of more complex trophic food webs with the inclusion of predators (Fig. 4).

445   This would also contribute to the positive association of body size to TRA-PCO2 (Fig. 4), as

446   the largest insects we found are either predators or shredders (Table S2).

447   Nevertheless, it is important to note that taxonomic and trait vectors had a

448   comparatively much weaker effect on predictability by spatial variables than that of site

449   occupancy. In fact, not only did site occupancy attain a higher statistical significance (Table

450   1), but it could also be partly related to the capability of the taxonomic vector TAX-PCO4 to

451   account for the adjusted D^2 values of binomial GLMs because these two predictor variables

452   were significantly correlated (Table S5). Moreover, when repeating the beta regression

453   analysis by using only the significant variables selected (i.e. site occupancy, TAX-PCO3,

454   TAX-PCO4, and TRA-PCO2), only site occupancy was statistically significant, indicating

455   that the influence of taxonomic and trait vectors on model performance is rather weak.

456   Analysing through beta regression the univariate relationship of these variables with the

457   adjusted D^2 values of binomial GLMs based on spatial effects again resulted in site

458   occupancy as the only significant variable (see Results above). Therefore, we must conclude

459   that any potential effect of taxonomic and trait vectors on model performance, including the

460   effect of female dispersal potential and body size, and that of functional feeding groups, must

461   be considered with caution: their statistical significance only appears after controlling for site

462   occupancy and the other variables considered in the full model of beta regression. In this

463   regard, the fact that *Baetis rhodani* is a widespread mayfly, which could not be modelled

464   because it was present at all sites, also gives support to the idea that dispersal abilities are not

465   so important in structuring invertebrate assemblages in high-latitude drainage basins. This is

466   because it demonstrates that mayfly species can be widespread, despite being rather weak

467   active dispersers. We also acknowledge that the rarest species (i.e. present in less than six

468   sites) were not modelled because models based on such small number of presences were

469   considered unreliable (e.g. Pierce & Ferrier, 2000). However, excluding these species does

470   not undermine the conclusion that the distributions of most common species are better

471   accounted for by models based on spatial variables than that of not-so-common species. In

472   fact, we effectively modelled 47 out of the 86 taxa available at the species (most cases) or

473   species-group (few cases) taxonomic resolution, comprising 55% of cases, which is a

474   representative subset of species in the entire metacommunity.

475

476   *Approaching the suitability of metacommunity analysis frameworks*

477   With the information above about single-species distribution models and subsequent

478   comparative analysis across species, it is possible to proceed with the evaluation of the

479   suitability of the two different frameworks of metacommunity analysis (Fig. 1) considered

480   here: 1) the classical approach exemplified by the four different non-exclusive perspectives

481   described by Leibold *et al.* (2004) or 2) the three exclusive components as proposed by

482   Logue *et al.* (2011).

483   Among the four different metacommunity perspectives of the Leibold *et al.* (2004)

484   framework, neutral theory and patch dynamics do not rely on the effect of environmental

485   variables, in contrast to species sorting and source-sink dynamics, the latter of which also

486   incorporating a strong influence of spatial effects (Fig. 1a). In our study, single-species

487   models often relied on the effect of environmental variables, particularly temperature and

488   shading, while being also dependent on large-scale spatial variables (Fig. 3). As

489   environmental and spatial factors are both relevant for the distribution of species, this result

490   suggests that either species sorting along spatially structured environmental gradients, or

491   source-sink dynamics between populations of high-quality and low-quality habitats, are both

492   likely as important processes driving metacommunities. Then, the comparative analysis

493   across species showed that site occupancy is responsible for the observed differences in the

494   relevance of spatial variables on species distributions (Table 1). This suggests that common

495   species would be better able than rare species to maintain populations in low-quality habitats

496   through constant immigration, and would favour the source-sink dynamics perspective over

497   species sorting.

498   Although species-sorting processes cannot be completely discarded because of the

499   demonstrated influence of environmental variables in many cases, deviance partitioning

500   suggests that the pure effects of environmental and spatial factors on species distributions are

501   stronger than their joint effects (Table S3). Also, the effect of spatial variables was better

502   explained than that of environmental factors by our explanatory variables, particularly site

503   occupancy, in the comparative analysis. These results slightly undermine the idea of species-

504   sorting across spatially structured environmental gradients as the most important process

505   shaping metacommunities. In any case, the neutral theory, which relies entirely on spatial

506   dynamics, is unlikely. As the dispersal potential of species has a rather weak effect on model

507   accuracy, patch dynamics can be discarded as well as a suitable perspective of

508   metacommunity analysis in our case. It should be acknowledged, however, that the difficulty

509   to explain model performance with dispersal ability can also be a consequence of the

510   coarseness of the dispersal measures currently available for freshwater invertebrates (Tachet

511   *et al*., 2010; Schmidt-Kloiber & Hering, 2015; Serra *et al*., 2016). Moreover, the different

512   metacommunity paradigms from Leibold *et al*. (2004) may always act simultaneously to a

513   certain extent along a continuum (Fig. 1a) rather than being distinct and mutually exclusive

514   options (Gravel *et al*., 2006; Logue *et al*., 2011; Brown *et al*., 2017).

515   Spatial autocorrelation may appear not only as a consequence of mass effects or

516   species sorting along spatially structured environmental gradients when the spatial scale is

517   not very large, but also as a consequence of dispersal limitation at very large spatial scales

518   (Heino *et al*., 2015). Nevertheless, some insect species found in this study exemplify well the

519   potential importance of the source-sink dynamics for metacommunities in subarctic streams,

520   independently of their dispersal capability. For example, six blackfly species were examined

521   (Table S1), of which five were present in more than 50% of sites, three of them in 75% of

522   sites or more (Table S2). Thus, blackfly species in subarctic streams have successfully spread

523   widely, which is advantageous to maintain metapopulations through source-sink dynamics.

524   On the other hand, the mayflies are as widespread as the blackflies (Fig. S4), but far less

525   capable of active dispersal. This suggests that the dispersal capability of species does not

526   determine the metapopulation dynamics, whereas site occupancy probably does so. Spatial

527   autocorrelation patterns have been described for the blackflies at small spatial scales, driven

528   by strong effects of inter-specific competition for oviposition sites, and subsequent priority

529   effects at the community level (McCreadie & Adler, 2012). The importance of priority effects

530   for the blackflies reinforces the idea of the relevance of site occupancy for community

531   dynamics, where rare species are in clear disadvantage for habitat re-colonisation.

532   Alternative to the framework of Leibold *et al*. (2004), we can interpret our results

533   under the framework of Logue *et al*. (2011), whereby three different and mutually exclusive

534   components can be used to analyse metacommunities: species equivalence, habitat

535   heterogeneity and dispersal (Fig. 1b). In our case, this alternative framework makes

536   interpretation of the results much easier. At the very least, we can conclude that species

537   equivalence is unlikely to play any role in metacommunity dynamics, similarly to discarding

538   neutral theory under the Leibold *et al*. (2004) framework. Dispersal can also be discarded, yet

539   again with caution due to the current lack of high resolution dispersal measures for freshwater

540   invertebrates (Tachet *et al*., 2010; Schmidt-Kloiber & Hering, 2015; Serra *et al*., 2016). Thus,

541   the main difference in the interpretation of the results with this alternative framework is that

542   we can now be certain about the role of habitat heterogeneity, while under the Leibold *et al*.

543   (2004) framework it is more difficult to discern whether species sorting or source-sink

544   dynamics is the dominant process. Habitat heterogeneity is indeed related to both

545   mechanisms. In fact, using habitat heterogeneity in space and time as the templet for

546   ecological strategies (Southwood, 1977) could be the framework of choice in situations

547   where it is difficult to discern species sorting processes from source-sink dynamics.

548

549   *Alternative approaches, caveats and conclusions*

550   Emergent properties at the community level are difficult to discern from field observational

551   data alone. In this regard, population genetics can be very useful in order to gain confidence

552   about the distinction between, e.g., source-sink dynamics and species sorting processes. This

553   is because population genetic studies could be used to estimate the relative contribution of

554   immigrants from nearby populations to the genetic variability of the population under study

555   (Bunn & Hughes, 1997; Hughes, Schmidt & Finn, 2009; Hughes, Huey & Schmidt, 2013).

556   Genetic analyses would probably provide the opportunity for a more robust interpretation of

557   our results. Genetic studies, however, are difficult to accomplish with stream insects in the

558   field when the idea is to compare many species at a time, and they are far more expensive

559   than the comparative approach of single species distributions we considered here. Therefore,

560   the comparative approach presented here can be used as a first step to explore the relative

561   contribution of environmental and spatial factors on species distributions, without using

562   expensive and time-consuming genetic analyses. In fact, by using the comparative approach

563   we can certainly conclude that the dispersal capability of species and neutral theory play little

564   role in shaping subarctic stream insect metacommunities. Rather, it is habitat heterogeneity,

565   which influences mass effects and/or species sorting processes, that matters. Subsequently,

566   the results of our study strongly recommend the preservation of habitat heterogeneity as the

567   conservation strategy to maintain biodiversity in these ecosystems.

568   Nevertheless, it should be acknowledged that one shortcoming of single-species

569   distribution modelling is that it does not consider the influence of species interactions in

570   structuring ecological communities. Stream ecology has considered that severe environmental

571   conditions may weaken the potential effects of biotic interactions in structuring communities

572   (Peckarsky, 1983). However, more recent findings pose doubts as to whether this is actually

573   true (Thomson *et al*., 2002; Cadotte & Tucker, 2017). In fact, biotic interactions can

574   reproduce patterns of community structure essentially identical to what it could be expected

575   from environmental filtering alone. This is because environmental changes may affect

576   population growth rates of competing species in opposite ways, and this may cause the

577   exclusion of some species that would otherwise be able to coexist (Cadotte & Tucker, 2017).

578   There exists also evidence indicating that biotic interactions limit the geographical range

579   expansion of species facing environmental changes (Sexton *et al*., 2009; Pigot & Tobias,

580   2013). Overall, this suggests that inter-specific interactions may also play a role in our case,

581   although the abundances of insect larvae in subarctic streams are typically low (see also

582   Heino & Grönroos, 2017) and may thus result in weak density-dependent interactions among

583   species (see also Morin, 2011).

584   Our study considered tributary streams draining into two linear sub-elements of a

585 larger river network (Fig. S1). However, there exists growing concern about the potential role

586 of the entire dendritic river networks in shaping biodiversity patterns, community structure

587 and species distributions (Altermatt, 2013). For example, the consideration of whole river

588 networks may unveil a more preeminent role for spatial factors in community assembly,

589 undermining the role of environmental filtering. Therefore, studies conducted across whole

590 dendritic networks could be more in line with neutral theory, as shown by Muneepeerakul *et*

591 *al*. (2008) for fish communities, yet no environmental variable was truly considered in that

592 study. Although we focused on tributary streams draining into the main river, the

593 consideration of whole dendritic networks may help us to perceive more accurately the real

594 connectivity pathways between isolated patches. This connectivity may have consequences

595 for metacommunity stability with respect to a situation where only a linear component of this

596 network is acting (Fagan, 2002). Also, dispersal along dendritic networks implies more

597 variability in local richness with strong consequences also for community differentiation

598 among patches (Carrara *et al*., 2012; Seymour *et al*., 2015).

599   In the case of stream insects, the taxa considered and the taxonomic resolution

600 achieved prior to species-distribution modelling, may also have important consequences on

601 our perception of the influence of dendritic riverine networks on biodiversity patterns (Kaelin

602 & Altermatt, 2016). Here, some taxa were discarded as it was not possible to determine the

603 species. Provided that the influences of dendritic landscapes and biotic interactions (discussed

604 above) may strongly affect how we understand the reality of community assemblages, it is

605 essential to use the best taxonomic resolution possible to make accurate inferences about the

606 mechanisms truly governing the observed patterns. In fact, the criterion of 'best taxonomic

607 resolution possible' used in our modelling endeavours is a fundamental requirement to draw

608 robust conclusions to be applied in biodiversity conservation.

609   Finally, for biodiversity conservation, it is essential to focus on maintaining habitat

610   heterogeneity because it appears to determine metacommunity organization (Kärnä *et al*.

611   2015) and species distributions (Heino & de Mendoza, 2016) in streams at high latitudes.

612   Unless habitat heterogeneity is not considered (along with potentially important effects of

613   dendritic network structure), conservation plans may fall short and not result in desired

614   outcomes.   .

615

616   **Acknowledgements**

625

626   **References**

627   Aalto, J., Pirinen, P., Heikkinen, J. & Venäläinen, A. (2013) Spatial interpolation of monthly

628   climate data for Finland: comparing the performance of kriging and generalized

629   additive models. *Theoretical and Applied Climatology*, **112**, 99–111.

630   Adler, P.H. & Crosskey, R.W. (2016) *World blackflies (Diptera: Simuliidae): A*

631   *comprehensive revision of the taxonomic and geographical inventory*.

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

633   Alahuhta, J. & Heino, J. (2013) Spatial extent, regional specificity and metacommunity

634   structuring in lake macrophytes. *Journal of Biogeography*, **40**, 1572–1582.

635   Altermatt, F. (2013) Diversity in riverine metacommunities: a network perspective. *Aquatic*

636   *Ecology*, **47**, 365–377.

637   Armitage, P.D. (1995) Behaviour and ecology of adults. In: *The Chironomidae: Biology and*

638   *Ecology of Non-Biting Midges* (Eds, P.D. Armitage, P.S. Cranston & L.C.V. Pinder),

639   pp. 194–224. Chapman & Hall, London.

640   Baldwin, W.F., West, A.S. & Gomery, J. (1975) Dispersal pattern of black flies (Diptera:

641   Simuliidae) tagged with ^{32}P. *Canadian Entomologist*, **107**, 113–118.

642   Bonthoux, S., Baselga, A. & Balent, G. (2013) Assessing community-level and single-species

643   models predictions of species distributions and assemblage composition after 25 years

644   of land cover change. *PLoS ONE*, **8**, e54179.

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

646   York.

647   Brittain, J.E. (1990) Life history strategies in Ephemeroptera and Plecoptera. In: *Mayflies and*

648   *Stoneflies* (Ed I.C. Campbell), pp. 1–12. Kluwer Academic Publishers, Dordrecht.

649   Brown, B.L., Sokol, E.R., Skelton, J. & Tornwall, B. (2017) Making sense of

650   metacommunities: dispelling the mythology of a metacommunity typology.

651   *Oecologia*, **183**, 643–652.

652   Bunn, S.E. & Hughes, J.M. (1997) Dispersal and recruitment in streams: evidence from

653   genetic studies. *Journal of the North American Benthological Society*, **16**, 338–346.

654   Burnham, K.P. & Anderson, D.R. (2004) Multimodel inference: understanding AIC and BIC

655   in model selection. *Sociological Methods & Research*, **33**, 261–304.

656   Cadotte, M.W. & Tucker, C.M. (2017) Should environmental filtering be abandoned? *Trends*

657   *in Ecology & Evolution*, **32,** 429–437.

658   Carrara, F., Altermatt, F., Rodriguez-Iturbe, I. & Rinaldo, A. (2012) Dendritic connectivity

659   controls biodiversity patterns in experimental metacommunities. *Proceedings of the*

660   *National Academy of Sciences of the USA*, **109**, 5761–5766.

661   Chapman, D.S. & Purse, B.V. (2011) Community versus single-species distribution models

662   for British plants. *Journal of Biogeography*, **38**, 1524–1535.

663   Chase, J.M. (2014) Spatial scale resolves the niche versus neutral theory debate. *Journal of*

664   *Vegetation Science*, **25**, 319–322.

665   Cook, R.D. (1977) Detection of influential observation in linear regression. *Technometrics*,

666   **19**, 15–18.

667   Covich, A.P., Palmer, M.A. & Crowl, T.A. (1999) The role of benthic invertebrate species in

668   freshwater ecosystems: zoobenthic species influence energy flows and nutrient

669   cycling. *BioScience*, **49**, 119-127.

670   Cribari-Neto, F. & Zeileis, A. (2010) Beta regression in R. *Journal of Statistical Software*,

671   **34**, 1–24.

672   Crosskey, R.W. (1990) *The Natural History of Blackflies*. John Wiley & Sons, Chichester.

673   Dankers, R. & Christensen, O.B. (2005) Climate change impacts on snow coverage,

674   evaporation and river discharge in the sub-arctic Tana basin, Northern Fennoscandia.

675   *Climatic Change*, **69**, 367–392.

676   Danks, H.V. (2007) How aquatic insects live in cold climates. *Canadian Entomologist*, **139**,

677   443–471.

678   de Jong, Y., Verbeek, M., Michelsen, V., Bjørn, P.P., Los, W., Steeman, F., Bailly, N.,

679   Basire, C., Chylarecki, P., Stloukal, E., Hagedorn, G., Wetzel, F.T., Glöckler, F.,

680   Kroupa, A., Korb, G., Hoffmann, A., Häuser, C., Kohlbecker, A., Müller, A.,

681   Güntsch, A., Stoev, P. & Penev, L. (2014) Fauna Europaea: all animal species on the

682   web. *Biodiversity Data Journal*, **2**, e4034.

683   Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz,

684   J.R., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C.,

685   Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D. & Lautenbach,

686   S. (2013) Collinearity: a review of methods to deal with it and a simulation study

687   evaluating their performance. *Ecography*, **36**, 27–46.

688   Dray, S., Legendre, P. & Peres-Neto P.R. (2006) Spatial modelling: a comprehensive

689   framework for principal coordinate analysis of neighbour matrices (PCNM).

690   *Ecological Modelling*, **196**, 483–493.

691   Fagan, W.F. (2002) Connectivity, fragmentation, and extinction risk in dendritic

692   metapopulations. *Ecology*, **83**, .3243–3249.

693   Ferrari, S.L.P. & Cribari-Neto, F. (2004). Beta regression for modelling rates and

694   proportions. *Journal of Applied Statistics*, **31**, 799–815.

695   Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level.

696   *Journal of Applied Ecology*, **43**, 393–404.

697   Fox, J. (2005) The R Commander: A Basic Statistics Graphical User Interface to R. *Journal*

698   *of Statistical Software*, **14**, 1–42.

699   Gíslason, G.M., Hannesdóttir, E.R., Munoz, S.S. & Pálsson, S. (2015) Origin and dispersal of

700   *Potamophylax cingulatus* (Trichoptera: Limnephilidae) in Iceland. *Freshwater*

701   *Biology*, **60**, 387–394.

702   Goslee, S.C. & Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of

703   ecological data. *Journal of Statistical Software*, **22**, 1–19.

704   Gravel, D., Canham, C.D., Beaudet, M. & Messier, C. (2006) Reconciling niche and

705   neutrality: the continuum hypothesis. *Ecology Letters*, **9**, 399–409.

706   Grönroos, M., Heino, J., Siqueira, T., Landeiro, V.L., Kotanen, J. & Bini, L.M. (2013)

707   Metacommunity structuring in stream networks: roles of dispersal mode, distance

708   type, and regional environmnental context. *Ecology and Evolution*, **3**, 4473–4487.

709   Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology.

710   *Ecological Modelling*, **135**, 147–186.

711   Hanski, I. (1994) A practical model of metapopulation dynamics. *Journal of Animal Ecology*,

712   **63**, 151–162.

713   Heino, J. (2005) Positive relationship between regional distribution and local abundance in

714   stream insects: a consequence of niche breadth or niche position? *Ecography*, **28**,

715   345–354.

716   Heino, J. & de Mendoza, G. (2016) Predictability of stream insect distributions is dependent

717   on niche position, but not on biological traits or taxonomic relatedness of species.

718   *Ecography*, **39**, 1216–1226.

719   Heino, J. & Grönroos, M. (2014) Untangling the relationships among regional occupancy,

720   species traits and niche characteristics in stream invertebrates. *Ecology and Evolution*,

721   **4**, 1931–1942.

722   Heino, J. & Grönroos, M. (2017) Exploring species and site contributions to beta diversity in

723   stream insect assemblages. *Oecologia*, **183**, 151–160.

724   Heino, J., Ilmonen, J. & Paasivirta, L. (2014) Continuous variation of macroinvertebrate

725   communities along environmental gradients in northern streams. *Boreal*

726   *Environment Research*, **19**, 21–38.

727   Heino, J., Melo, A.S., Siqueira, T., Soininen, J., Valanko, S. & Bini, L.M. (2015)

728   Metacommunity organisation, spatial extent and dispersal in aquatic systems:

729   patterns, processes and prospects. *Freshwater Biology*, **60**, 845–869.

730   Hoffsten, P.-O. (2004) Site-occupancy in relation to flight-morphology in caddisflies.

731   *Freshwater Biology*, **49**, 810–817.

732   Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*.

733   Princeton University Press, Princeton.

734   Hughes, J.M., Schmidt, D.J. & Finn, D.S. (2009) Genes in streams: Using DNA to

735   understand the movement of freshwater fauna and their riverine habitat. *BioScience*,

736   **59**, 573–583.

737   Hughes, J.M., Huey, J.A. & Schmidt, D.J. (2013) Is realised connectivity among populations

738   of aquatic fauna predictable from potential connectivity? *Freshwater Biology*, **58**,

739   951–966.

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

741   91–95.

742   Kaelin, K. & Altermatt, F. (2016) Landscape-level predictions of diversity in river networks

743   reveal opposing patterns for different groups of macroinvertebrates. *Aquatic*

744   *Ecology*, **50**, 283–295.

745   Kärnä, O.-M., Grönroos, M., Antikainen, H., Hjort, J., Ilmonen, J., Paasivirta, L. & Heino, J.

746   (2015) Inferring the effects of potential dispersal routes on the metacommunity

747   structure of stream insects: as the crow flies, as the fish swims or as the fox runs?

748   *Journal of Animal Ecology*, **84**, 1342–1353.

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

750   Amsterdam.

751   Legendre, P., Borcard, D., Blanchet, F.G. & Dray, S. (2013). *PCNM: MEM spatial*

752   *eigenfunction and principal coordinate analyses*. R package version 2.1-2/r109.

753   http://R-Forge.R-project.org/projects/sedar/

754 Leibold, M.A. (1995) The niche concept revisited: mechanistic models and community

755   context. *Ecology*, **76**, 1371–1382.

756 Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F.,

757   Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. & Gonzalez, A. (2004) The

758   metacommunity concept: a framework for multi-scale community ecology. *Ecology*

759   *Letters*, **7**, 601–613.

760 Logue, J.B., Mouquet, N., Peter, H. & Hillebrand, H. (2011) Empirical approaches to

761   metacommunities: a review and comparison with theory. *Trends in Ecology and*

762   *Evolution*, **26**, 482–491.

763 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2013) *cluster: Cluster*

764   *analysis basics and extensions*. R package version 1.14.4. http://CRAN.R-

765   project.org/package=cluster.

766 Mansikkaniemi, H. (1970) Deposits of sorted material in the Inarijoki-Tana river valley in

767   Lapland. *Reports of Kevo Subarctic Research Station*, **6**, 1–63.

768 McCreadie, J.W. & Adler, P.H. (2012) The roles of abiotic factors, dispersal, and species

769   interactions in structuring stream assemblages of black flies (Diptera: Simuliidae).

770   *Aquatic Biosystems*, **8**, 14.

771 Merritt, R.W. & Cummins, K.W. (1996) *An Introduction to the Aquatic Insects of North*

772   *America (3rd edition)*. Kendall/Hunt Publishing, Dubuque.

773 Morin, P.J. (2011) *Community Ecology, 2nd edition*. Wiley-Blackwell, Oxford.

774 Müller-Peddinghaus, E. (2011) *Flight-morphology of Central European caddisflies (Insecta:*

775   *Trichoptera) in relation to their ecological preferences*. PhD Thesis, University of

776   Duisburg-Essen, Essen.

777   Müller-Peddinghaus, E. & Hering, D. (2013) The wing morphology of limnephilid

778   caddisflies in relation to their habitat preferences. *Freshwater Biology*, **58**, 1138–

779   1148.

780   Muneepeerakul, R., Bertuzzo, E., Lynch, H.J., Fagan, W.F., Rinaldo, A. & Rodriguez-Iturbe,

781   I. (2008) Neutral metacommunity models predict fish diversity patterns in

782   Mississippi-Missouri basin. *Nature*, **453**, 220–222.

783   National Board of Waters (1981) *Vesihallinnon analyysimenetelmät*, Tiedotus 213.

784   Vesihallitus, Helsinki.

785   Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson,

786   G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2013) *vegan: Community Ecology*

787   *Package*. R package version 2.0-9. http://CRAN.R-project.org/package=vegan.

788   Pearce, J. & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species

789   distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.

790   Peckarsky, B.L. (1983) Biotic interactions or abiotic limitations? A model of lotic community

791   structure. In: *Dynamics of Lotic Ecosystems* (Eds T.D. Fontaine III & S.M. Bartell),

792   pp. 303–323. Ann Arbor Science Publishers, Ann Arbor.

793   Pigot, A.L. & Tobias, J.A. (2013) Species interactions constrain geographic range expansion

794   over evolutionary time. *Ecology Letters*, **16**, 330–338.

795   Pirinen, P., Simola, H., Aalto, J., Kaukoranta, J.P., Karlsson, P & Ruuhela, R. (2012)

796   Climatological statistics of Finland 1981-2010. *Finnish Meteorological Institute*

797   *Reports*, **1,** 1–96.

798   Pulido, C., Riera, J.L., Ballesteros, E., Chappuis, E. & Gacia, E. (2015) Predicting aquatic

799   macrophyte occurrence in soft-water oligotrophic lakes (Pyrenees mountain range).

800   *Journal of Limnology*, **74,** 143–154.

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

802   652–661.

803   R Core Team (2013) *R: A language and environment for statistical computing. R Foundation*

804   *for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org/.

805   Schmidt-Kloiber, A. & Hering, D. (2015) www.freshwaterecology.info – An online tool that

806   unifies, standardises and codifies more than 20,000 European freshwater organisms

807   and their ecological preferences. *Ecological Indicators*, **53**, 271–282.

808   Serra, S.R.Q., Cobo, F., Graça, M.A.S., Dolédec, S. & Feio M.J. (2016) Synthesising the trait

809   information of European Chironomidae (Insecta: Diptera): Towards a new database.

810   *Ecological Indicators*, **61**, 282–292.

811   Sexton, J.P., McIntyre, P.J., Angert, A.L. & Rice, K.J. (2009) Evolution and ecology of

812   species range limits. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 415–

813   436.

814   Seymour, M., Fronhofer, E.A. & Altermatt, F. (2015) Dendritic network structure and

815   dispersal affect temporal dynamics of diversity and species persistence. *Oikos*, **124**,

816   908–916.

817   Southwood, T.R.E. (1977) Habitat, the templet for ecological strategies? *Journal of Animal*

818   *Ecology*, **46,** 337–365.

819   Tachet, H., Richoux, P., Bournaud, M. & Usseglio-Polatera, P. (2010) *Invertébrés d'eau*

820   *douce: systématique, biologie, écologie (nouvelle édition revue et augmentée)*. CNRS,

821   Paris.

822   Thomson, J.R., Lake, P.S. & Downes, B.J. (2002) The effect of hydrological disturbance on

823   the impact of a benthic invertertebrate predator. *Ecology*, **83**, 628–642.

824   Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. & Cushing, C.E. (1980) The

825   river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**,

826   130–137.

827   Wallace, J.B. & Webster, J.R. (1996) The role of macroinvertebrates in stream ecosystem

828   function. *Annual Review of Entomology*, **41**, 115–139.

829   Wentworth, C.K. (1922) A scale of grade and class terms for clastic sediments. *Journal of*

830   *Geology*, **30,** 377–392.

831   Zar, J.H (1984) *Biostatistical Analysis, Second Edition*. Prentice Hall, Englewood Cliffs.

832

833   **Supporting Information**

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

835   **Table S1.** Insect species considered.

836   **Table S2.** Species traits considered and site occupancy.

837   **Table S3.** Results of binomial GLMs.

838   **Table S4.** Taxonomic and trait vectors from Principal Coordinate Analysis (PCO), with

839   corresponding scores for each species.

840   **Table S5.** Statistical significance of the correlations among site occupancy, species traits,

841   trait vectors, and taxonomic vectors.

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

843   **Figure S2.** Eigenvalues from taxonomic Principal Coordinate Analysis (PCO) based on

844   taxonomic distances between species.

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

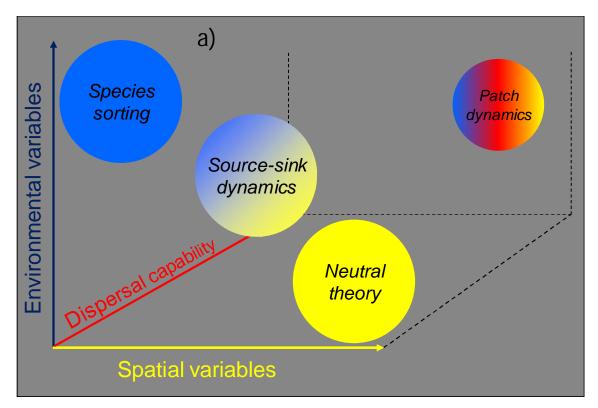846   **Figure S4.** Comparison of site occupancy values between different insect groups.

847

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

	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		

	Estimate	SE	z	P		
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		

853

Figure legends

855   **Fig. 1.** Conceptual representation of **a)** the four non-exclusive classical approaches in

856   metacommunity studies (Leibold *et al*. 2004), and **b)** the more recent framework of

857   metacommunity analysis based on three exclusive components (Logue *et al*. 2011); according

858   to the relative relevance of the variables used in this study (axes): spatial variables (x-axis),

859   environmental variables (y-axis) and the different dispersal capability of species (z-axis).

860   Circles represent the theoretical location where the emphasis of each approach is situated

861   across the three axes.

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

863   **Fig. 3.** Frequency of local environment variables (top) and spatial variables (bottom) selected

864   as significant in explaining species distributions through binomial GLMs. Spatial variables

865   are arranged from small-scale (i.e. V13) to large-scale extent (i.e. V1), and include V11

866   which was never selected. Climate variables are not shown (mean annual temperature was

867   selected 32 times, and July air temperature 20 times). Species-environment relationships are

868   shown in black when positive and grey when negative. Specific information for each species

869   can be found in Table S2.

870   **Fig. 4.** Species scores on taxonomic vectors TAX-PCO3 (a) and TAX-PCO4 (b), and on trait

871   vector TRA-PCO2 (c), arranged from lowest to highest values. For the trait vector, the

872   location of the different functional feeding (FFG) and habit trait groups (HTG), is indicated,

873   as well as the four different body size classes considered (BS, represented by columns of four

874   different sizes), and those insects considered as of high female dispersal potential (H).

875

Fig. 1.

878 **Fig. 2.**

879

Fig. 3.

Fig. 4.