# Exploring species and site contributions to beta diversity in stream insect assemblages

Heino, Jani

Springer Science and Business Media Deutschland GmbH
2017-01

Heino, J & Grönroos, M 2017, 'Exploring species and site contributions to beta diversity in stream insect assemblages', Oecologia, vol. 183, no. 1, pp. 151-160. https://doi.org/10.1007/s00442-016-3754-7

http://hdl.handle.net/10138/307947
10.1007/s00442-016-3754-7

unspecified
acceptedVersion

A revised manuscript for review in Oecologia

# Exploring species and site contributions to beta diversity in stream insect assemblages

Jani Heino[1] and Mira Grönroos[2]

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

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

Correspondence: jani.heino@environment.fi

Key words: community ecology, ecological uniqueness, niche breadth, occupancy, species characteristics.

Author Contributions: JH devised the study idea, ran the analyses and led the writing. MG collected trait data, commented on all phases of the study and contributed to the writing.

## Abstract

It was recently suggested that beta diversity can be partitioned into contributions of single sites to overall beta diversity (LCBD) or into contributions of individual species to overall beta diversity (SCBD). We explored the relationships of LCBD and SCBD to site and species characteristics, respectively, in stream insect assemblages. We found that LCBD was mostly explained by variation in species richness, with a negative relationship being detected. SCBD was strongly related to various species characteristics, such as occupancy, abundance, niche position and niche breadth, but was only weakly related to biological traits of species. In particular, occupancy and its quadratic terms showed a very strong unimodal relationship with SCBD, suggesting that intermediate species in terms of site occupancy contribute most to beta diversity. Our findings of unravelling the contributions of sites or species to overall beta diversity are of high importance to community ecology, conservation and bioassessment using stream insect assemblages, and may bear some overall generalities to be found in other organism groups.

## Introduction

A major aim of ecology is to understand factors affecting spatial variation of species diversity. Species diversity can be divided into gamma, alpha and beta components (Whittaker 1960), of which the latter has received particular interest from ecologists in the last few years (Anderson et al. 2011). Recently, Legendre and De Cáceres (2013) suggested that beta diversity can be partitioned into contributions of single sites to overall beta diversity or into contributions of individual species to overall beta diversity. Hence, local contributions to beta diversity (LCBD) describe the ecological uniqueness of a site, whereas species contributions to beta diversity (SCBD) can be considered the relative importance of each species in affecting beta diversity patterns. A few studies have focused on LCBD to answer various questions of beta diversity (Lopes et al. 2014; Silva and Hernandes 2014; Tonkin et al. 2016), whereas correlates of SCBD have received little attention to date.

Patterns in LCBD and SCBD can be examined using site-based and species-based approaches, respectively. First, LCBD can be correlated with various environmental and spatial variables to see whether the ecological uniqueness of sites is based on environmental conditions (Lopes et al. 2014; Silva and Hernandes 2014; Tonkin et al. 2016), or result from community richness and abundance, which are, in turn, often determined by site environmental characteristics (Grönroos and Heino 2012; Tonkin et al. 2016). Second, SCBD can be associated with general species characteristics, such as degree of occupancy, abundance, niche position, niche breadth and species traits. These species characteristics have been shown to be inter-correlated at least to a certain degree (Tales et al. 2004; Heino 2005; Siqueira et al. 2009; Heino and Grönroos 2014). Also, as niche characteristics can be used to describe species-environment associations, they could be expected to be correlated with

SCBD. It can be assumed that generalist species with broad niches contribute less to SCBD than species with small or intermediate-sized niches. This is because species with small niche breadth may occur in environmentally restricted conditions (Brown 1984; Slatyer et al. 2013) and thus contribute more to beta diversity. Similarly, niche position can be assumed to be associated with SCBD because species occurring in marginal habitats should occur in environmentally more restricted conditions than non-marginal species (Dolédec et al. 2000; Heino and Grönroos 2014). Irrespective of niche characteristics, species biological traits, such as feeding mode, body size and dispersal capacity, might affect species contributions to beta diversity. This is because fundamental biological characteristics of species may affect species occupancy and abundance (Verberk et al. 2010; Heino and Grönroos 2014). The degree of occupancy should also itself affect SCBD, with species occurring at intermediate proportion of sites being most variable and contributing most to beta diversity. However, the effect of abundance may be different. Species having large total abundances in the data should also show most abundance variation among sites and thus show high contributions to beta diversity. These ideas are not completely new, as the mathematical relationship between mean, variance and occurrence is well known in ecological research (Gaston et al. 2006). However, these hypotheses remain to be tested rigorously, and they are important for understanding the formation of beta diversity patterns and revealing which kind of species, in terms of their ecological and biological traits, are particularly important.

We examined the contributions of single sites (LCBD) or single species (SCBD) to overall beta diversity in stream insect assemblages. Furthermore, we used community metrics (i.e. community richness and community abundance) and ecological variables (i.e. environmental features of the sites) as correlates of LCBD. Similarly, we used species metrics (i.e. species abundance and occupancy) and species characteristics (i.e. biological traits, niche position and niche breadth) as correlates of SCBD. We ran separate regression models for community metrics and ecological features as predictors for variation in LCBD or SCBD. We used test data from three northern drainage basins showing wide variation in natural environmental conditions (Grönroos et al. 2013; Heino and Grönroos 2014). Previous studies from these and other drainage basins have shown that stream insects are highly amenable model organisms for testing the relationships between species and their environment (Heino 2005; Siqueira et al. 2009) and formation of biodiversity patterns (Heino et al. 2003; Hawkins et al. 2015). However, none of these studies has examined the correlates of LCBD or SCBD.

## Materials and Methods

### *Study area and sampling*

Stream insect samples were collected and environmental variables were measured in three drainage basins in Finland: Iijoki (65ºN, 26ºE; basin total area: 14 191 km$^2$), Koutajoki (66ºN, 29ºE; basin total area: 26 100 km$^2$) and Tenojoki (70ºN, 27ºE; basin total area: 16 374 km$^2$). Largely the same dataset has been used previously for examining community assembly (Heino 2013), metacommunity patterns (Grönroos et al. 2013), occupancy-abundance relationships (Heino and Grönroos 2014) and occupancy frequency patterns (Heino 2015) in stream invertebrates. Samples from altogether 60 near-pristine to pristine stream sites were included in the present study, with a total of 20 streams sampled in each of the three basins during the spring period. Thus, to ensure that the samples in each basin were collected at the same ‘ecological season’, the Iijoki samples were collected in late May in 2009, the Koutajoki samples were collected in late May 2008, and the Tenojoki samples were collected in early to mid-June in 2010. We refer to the above mentioned papers for further information about the study system.

At each site, approximately 100 $m^2$ stream riffle section was selected for sampling. Macroinvertebrates were sampled using a kick-net (net mesh size 0.3 mm). Four 30-seconds per one-meter subsamples were divided among the different microhabitats (based on variation in velocity, depth, moss cover, and particle size) and pooled in the field. This sampling effort should catch more than 70% of species occurring at a site, mainly missing species that are only occasional in northern streams (Mykrä et al. 2006). Samples were preserved in 70% ethanol in the field and later processed in the laboratory. Samples were identified to species, species group, or genus level, including also non-biting midges (Diptera: Chironomidae) and blackflies (Diptera: Simuliidae).

Altogether 15 environmental variables were measured at each site. Percentage cover of deciduous trees (1) and canopy shading (2) were visually estimated along a 50 m stretch of the riparian zone at each site on both banks. Mean riffle width (3) was measured at each sampling site, based on five across-stream transects. Current velocity and depth were measured at 30 random locations in cross-channel transects, the number of which depended on riffle width. Moss cover (%) was visually estimated in ten randomly spaced 50 × 50 cm plots. Mean current velocity (4), standard deviation of current velocity (5), mean depth (6), standard deviation of depth (7), mean moss cover (8) and standard deviation of moss cover (9) were then calculated and used in the analyses. Substratum particle size was visually estimated from the same plots with moss cover estimates. In the analyses, we used the mean percentage cover of the size classes sand (10), gravel (11), cobble (12) and bolder (13), which correspond to diameters 0.25–2 mm, 2–16 mm, 64–256 mm and 256–1024 mm, respectively. Conductivity (14) and pH (15) were measured either in laboratory using Finnish national standards (National Board of Water and the Environment 1981) or using YSI device in the field.

*Species traits, trait vectors and statistical analyses*

We gathered information for four biological traits/trait groups: functional feeding groups, habit trait groups, body size and female dispersal strength. For functional feeding groups, we mainly used Moog (2002) where one to 10 points are given for a species depending on its association with each feeding group. Here, a species with ≥5 points for a given group was assigned to belong to the respective group. For habit trait groups and female dispersal, we mainly used US EPA (2012) trait database.  Body size information was provided by S. Dolédec (most groups), J. Ilmonen (Simuliidae) and L. Paasivirta (other dipterans). In cases where the information was not found in the primary source, we also used Merritt and Cummins (1996), Vieira et al. (2006), Tachet et al. (2010), taxonomic handbooks (e.g. Nilsson 1996) and our expert opinion. There were five functional feeding groups (i.e. predators = PRE, scrapers = SCR, filterers = FIL, gatherers = GAT and shredders = SHR), five habit trait groups (i.e. clingers = CLIN, burrowers = BUR, sprawlers = SPR, climbers = CLIM and swimmers = SWI; see Merritt and Cummins 1996), five body size classes (i.e. maximum larval length coded as an ordered factor: 0-2.5 mm, 2.5-5 mm, 5-10 mm, 10-20 mm and 20-40 mm) and two female dispersal strength classes (i.e. low and high; see Poff et al. 2006). We used this information on traits for calculating “trait distances” between species based on Gower distance coefficient using the function “gowdis” from the R package *FD* (Laliberté et al. 2014). We then ran a principal coordinates analysis (PCoA; based on the Gower distances between species) using the function ‘cmdscale’ from the R package stats (R core team 2015) to provide trait vectors for subsequent beta regression analyses. These trait vectors are, contrary to the original trait variables, continuous variables that can be used to describe trait differences between species (for a similar approach, see Heino and de Mendoza 2016). In practice, Euclidean distances between species based on the first four PCoA axes and original Gower distances between species were very strongly correlated (Mantel $r = 0.909$, $p < 0.001$), indicating that the first four PCoA axes retained most of the information about original trait distances between species.

We calculated niche position and niche breadth for each of the 203 species using the outlying mean index (OMI) analysis (Dolédec et al. 2000). This method uses the relationships between species abundances and environmental variables to produce three indices: ‘OMI’ or the niche position of a species (NicPos), ‘tolerance’ or the niche breadth of a species (NicBre), and ‘residual tolerance’ that helps one to infer how well environmental variables considered have accounted for variation in species distributions. The OMI metric measures the distance of each species to the average environmental conditions in the study area, whereas the tolerance metric measures the amplitude in the species distributions across the studied environmental gradients (Dolédec et al. 2000). Following Heino and Grönroos (2014), we used all the 15 environmental variables to calculate realised environmental niche position and environmental niche breadth for each species (analysis results shown in the above mentioned paper). We calculated OMI and tolerance values for all species, including very rare ones, as our community data analysis (see below) also incorporated all species. It has to be noted, however, that species occurring at a single site only will get a value of 0 for tolerance (NicBre). Although such zero values are unrealistic, as no species may have ‘null’ niche breadth, they are useful in comparative purposes to indicate extreme species that have very small niche breadth in a given study system.

We used a combination of multivariate methods and beta regression analysis to examine patterns in local contribution to beta diversity (LCBD) or species contributions to beta diversity (SCBD). Following Legendre and De Cáceres (2013), we first Hellinger-transformed site-by-species abundance or presence-absence community matrix, and subsequently calculated the total beta diversity (BD total), LCBD value for each site and SCBD value for each species. In brief, this approach is based on the total variance in the community matrix, which is the total sum of squares (the sum, over all species and all sites, of the squared deviations from the species means) divided by n-1. This measure, in turn, can be decomposed into the contribution of the sites or the species to total beta diversity (Legendre and De Cáceres 2013). For further details of calculating BD total, LCBD and SCBD values, and the R function 'beta.div' for conducting those calculations, see Legendre and De Cáceres (2013). LCBD represents the ecological uniqueness of a site, i.e. it provides a measure of the relative contribution of the given sampling unit to beta diversity. SCBD indicates how large a contribution a species has to overall beta diversity in the data set. Large LCBD or large SCBD values indicate high importance of the given site or species, respectively, to the overall beta diversity (i.e. variation in species composition across sites; Anderson et al. 2011). We ran the analysis using both abundance and presence-absence data to show if there were differences between quantitative and qualitative data.

We used Kruskal-Wallis test (R function 'kruskal.test') and associated multiple comparisons (R function 'kruskalmc') to test for pairwise differences in both LCBD measures as well as community richness and community abundance among the three drainage basins. Pairwise comparisons were based on the method developed by Siegel and Castellan (1988, pp. 213-214). Kruskal-Wallis test was selected for these comparisons because our data did not meet all assumptions of parametric tests (i.e. ANOVA).

Because our response data (LCBD or SCBD) varied between 0 and 1, we used beta regression as our modelling tool (Cribari-Neto and Zeileis 2010). Beta regression is typically used to model variables that show values in the standard unit interval (i.e. vary between 0 and 1). Beta regression is based on the assumption that the dependent variable is beta-distributed and that its mean is related to a set of regressors through a linear predictor with unknown coefficients and a link function (Cribari-Neto and Zeileis 2010). The beta regression approach naturally incorporates features such as heteroskedasticity or skewness which are typically observed in response data taking values from 0 to 1. We used beta regression with logit link function for four separate models. First, we related LCBD to community metrics, namely species richness and community abundance, as well as their second order terms to account for nonlinear responses (Supporting information, Fig. S1). Second, we ran beta regression of LCBD using eight main environmental variables as predictors (i.e. pH, conductivity, shading, riparian deciduous trees, stream width, depth, velocity and moss cover). These variables have formerly been found to be influential in affecting variation in macroinvertebrate communities in northern streams (Grönroos and Heino 2012; Heino et al. 2014). Third, we used beta regression to relate SCBD to species metrics, including number of sites occupied (NumSit), total abundance of a species in the dataset (TotAbu), as well as their quadratic terms (Supporting information, Fig. S1). Finally, in the fourth model, we used beta regression to analyse variation in SCBD using niche position (NicPos), niche breadth (NicBre), and the first four PCoA trait vectors as predictor variables (see above). We also examined reduced models based on pseudo coefficients of determination (pseudo-$R^2$).

Statistical analyses, except calculations of local or species contributions to beta diversity, were run using the R packages stats (R Core Team 2015), ade4 (Dray and Dufour 2007), pgirmess (Giraudoux 2016) and betareg (Cribari-Neto and Zeileis 2010).

## Results

Total beta diversity numbers were 38.67 for SS total and 0.655 for BD total based on presence-absence data. Corresponding figures for abundance data were 37.57 for SS total and 0.636 for BD total.

LCBD based on presence-absence data was significantly related to community richness and its quadratic term, but not with community abundance (Table 1). Abundance-based LCBD was significantly related to community abundance. Also, none of the environmental variables was strongly related to LCDB, although the variable "deciduous trees" was significant in the model of abundance-based LCBD (Table 2). Also, pseudo $R^2$ values of these models were low. However, there were significant differences in abundance-based LCBD, community richness and community abundance among the three drainage basins (Kruskal-Wallis test, $p < 0.05$), but no significant differences were found for presence-absence LCBD. Based on multiple comparisons in Kruskal-Wallis test ($p < 0.05$), lower community richness was found in streams of the Tenojoki basin than in the Iijoki and Koutajoki basins. Also, community abundance was higher in the Iijoki basin than in the other two basins, and abundance-based LCBD was higher in the Koutajoki basin than in the Tenojoki basin (Fig. 1).

SCBD based on presence-absence data was significantly related to number of sites occupied and its quadratic term, but not with total abundance and its quadratic term (Table 3). The model including all four explanatory variables (i.e. number of sites occupied and its quadratic term, total abundance and its quadratic term) accounted for 85% of variation in SCBD. In addition, the model that included only the variable 'number of sites occupied' and its quadratic term accounted for 85% of variation in SCBD, and a hump-shaped relationship between the number of sites occupied and SCBD was observed (Fig. 2, Supporting information Fig. S1). SCBD based on abundance data showed rather similar patterns to those of presence-absence data. However, in this model, total abundance and its quadratic term were also significant (Table 3). The model including all explanatory variables accounted for 72% of variation. The model that only included number of sites occupied as well as its quadratic term, however, explained 63% of variation in abundance-based SCBD (Fig. 2). Beta regression also showed that SCBD-trait relationships were rather weak (Table 4), and only niche position explained significant variation in both SCBD indices. However, niche breadth was also significant predictor of SCBD based on presence-absence data, and pco3 was also significantly related to variation in SCBD based on abundance data (Table 4). This trait vector was related to variation in functional feeding mode (scrapers had high values and predators had low values), habit traits (high values for sprawlers and low values for burrowers), body size (small size classes had high values and large size classes had low values) and female dispersal potential (low dispersal had higher values than high dispersal) (Fig. 3). Inspection of the relationships between SCBD indices and original trait variables showed that female dispersal capacity was most clearly associated with species contribution to beta diversity (Supporting Information Fig. S3).

One species, the mayfly *Baetis rhodani*, had a very high Cook's distance in SCBD-related beta regression analyses. This species was the most abundant and was among the three most widely distributed species in the dataset. Its removal did not appreciably affect the interpretation of the results and it was thus included in the model. Also, when this species was excluded, a few other species appeared to have high Cook's distance.

## Discussion

Our findings showed that both local contribution to beta diversity (LCBD) and species contribution to beta diversity (SCBD) showed some highly predictable patterns. First, LCBD was significantly related to species richness, whereas SCBD showed significant relationships with various species metrics, particularly with number of sites occupied by a species. However, environmental variables were less influential for LCBD, and the same was true for species traits for SCBD. In the following, we will discuss these main findings.

We found that LCBD was negatively related to species richness, although the relationship was also slightly curvilinear (Table 1a). Legendre and De Cáceres (2013) and Silva and Hernandez (2014) also found that LCBD was negatively correlated with species richness, indicating that sites with unique species composition harboured low species richness. However, in our analyses, when LCBD was based on abundance data, the relationship between LCBD and species richness was not significant, but community abundance was significant (Table 1b). Our findings thus suggest that the variation in LCBD of stream insects across streams is governed by variation in species richness or abundance (depending on whether presence-absence or abundance data are used in the calculation of LCBD), which in turn may be typically correlated with factors such as stream size, moss cover and pH or vary between drainage basins (Heino et al. 2003; Heino and Grönroos 2012). However, we did not find among-drainage basin differences in LCBD based on presence-absence data, although abundance-based LCBD, species richness and community abundance did show significant differences. This finding further suggests that both local stream-level factors and drainage basin identity may affect LCBD, but these relationships may be complex (cf. presence-absence vs. abundance-based LCBD results). Also, in this study, the variation in LCBD values was not explained well by local environmental variables. This finding is rather similar to that of Tonkin et al. (2016), who found that LCBD was variably related to environmental factors in five different stream invertebrate datasets.

SCBD correlated significantly with the species metrics, notably with the number of sites occupied and its quadratic term. These two variables alone explained 85% and 68% of variation in presence-absence and abundance-based SCBD, respectively, but shape of the relationship was different (Fig. 2). Our findings showed, in the presence-absence case, that intermediate species in terms of site occupancy contribute most to the presence-absence-based beta diversity. This is because those intermediate species show most variation in occupancy among sites (see also Gaston et al. 2006). Our results also showed, in the abundance case, that species with high occupancy across sites and high total abundance in the data contribute most to the abundance-based beta diversity. The finding that widely-distributed species are also abundant relates to the positive occupancy-abundance relationship (Brown 1984; Gaston 2003), which has also been detected previously in stream ecosystems in general (e.g. Siqueira et al. 2009) and the present study system in particular (e.g. Heino and Grönroos 2014). Our findings also suggest that species contribution to beta diversity in a dataset is strongly predictable from general species characteristics, including its occupancy and abundance. This is somewhat circular reasoning, however, as SCBD, occupancy and abundance are calculated from the same data. This mathematical dependency may affect the relationships among occupancy, abundance and SCBD (see also Legendre and De Cáceres 2013). For example, high SCBD values mean that species show a large variation in abundance across locations. The mean-variance relation suggests that such species are expected to have relatively high local abundance as well as high occurrence. Similarly, if we measure SCBD using species presence-absence data, high variance is shown by species that have intermediate occurrence.

We found only weak effects of species traits on SCBD. While niche position was a significant determinant of SCBD in both cases, niche breath was significant only for SCBD based on presence-absence data, and only the third PCoA axis was significantly related with variation in SCBD based on abundance data. This finding suggests that biological traits are less important than niche position in affecting SCBD. This finding is not surprising because niche position is typically strongly correlated with occupancy (Heino 2005; Siqueira et al. 2009), and hence could be assumed to be related to SCBD. Also, niche position is related to the amount of variability in the occupancy of species that can be accounted for by environmental variables (Heino and de Mendoza 2016). This suggests that there may be complex, yet predictable relationships between SCBD, occupancy, niche position and environmental relationships of species. Biological traits fit less well into this equation, as they were not strongly related to either SCBD (this study) or species occupancy (Heino and de Mendoza 2016). However, one biological trait PCoA axis was significantly related to SCBD based on abundance data, which suggests that their role cannot be entirely ruled out. In this context, female dispersal capacity may be a key trait, with low dispersal capacity being probably related to high contribution to beta diversity (Supporting information Fig. S3).

Understanding the determinants of LCBD and SCBD is important from basic ecological, conservation and bioassessment viewpoints. First, LCBD indicates the ecological uniqueness of a site (Legendre and De Cáceres 2013), which goes beyond typical measures of beta diversity and provides a means to value single sites separately. Second, valuing single sites is important for conservation because limited resources do not allow us to conserve all sites. Therefore, preserving a set of ecologically unique sites is one option for biodiversity conservation. However, one has to consider that sites having high LCBD values are often rather species poor sites (Legendre and De Cáceres 2013; Silva and Hernandes 2014), which may limit their use in conservation if species-rich sites are a conservation goal. A compromise would be to conserve a combined set of both ecologically unique and species-rich sites. Third, from the bioassessment perspective, it would be feasible to focus on species having intermediate site occupancy, as these species contribute most to beta diversity.

Changes in the occupancies of these species should also result in clearly discernible changes in the variation of community structure across sites in the face of anthropogenic changes. This is particularly so if previously uncommon species ‘switch roles’ with previously widely-distributed or highly abundant species. For example, Hawkins et al. (2015) found that anthropogenic disturbances decreased the regional prevalence of most common taxa and increased the prevalence of several less common taxa in stream macroinvertebrate communities. This finding also suggests that species contributions to beta diversity change with anthropogenic disturbance, resulting in biodiversity patterns differing between sets of anthropogenically altered and more pristine ecosystems (but see Mayor et al. 2015).

In conclusion, we found that LCBD and SCBD were highly predictably related to species richness and species occupancy, respectively. We emphasise the importance of SCBD in this context, as the contributions of single species to overall beta diversity have been little studied using this approach. At the same time, understanding the determinants of LCBD and SCBD may hold a key to various general ecological, conservation and bioassessment issues. We thus urge other researchers to test the generality of our findings in ecological systems other than running waters.

## Acknowledgements

This research was supported by grants from the Academy of Finland, Maj and Tor Nessling Foundation, and Kone Foundation. We thank Sylvain Dolédec, Jari Ilmonen and Lauri Paasivirta for insect body size information, and Tommi Karhu, Maija Niva and Heikki Mykrä for help with the field and laboratory work. We also thank Jamie Kneitel and two anonymous reviewers for excellent comments on earlier drafts of the manuscript.

## References

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

Brown JH (1984) On the relationship between abundance and distribution of species. Am Nat 124: 255-279.

Cribari-Neto F, Zeileis A (2010) Beta Regression in R. J Statistical Softw 34:1-24

Dolédec S, Chessel D, Gimaret-Carpentier C 2000. Niche separation in community analysis: a new method. Ecology 81:2914–2927

Dray S, Dufour AB (2007): The ade4 package: implementing the duality diagram for ecologists. J Statistical Softw 22: 1-20

Gaston KJ (2003) The structure and dynamics of geographic ranges. Oxford University Press, Oxford

Gaston KJ, Borges PAV, He F, Gaspar C (2006) Abundance, spatial variance and occupancy: arthropod species distribution in the Azores. J Anim Ecol 75:646–656

Giraudoux P (2016) pgirmess: Data Analysis in Ecology. R package version 1.6.4. http://CRAN.R-project.org/package=pgirmess

Grönroos M, Heino J (2012) Species richness at the guild level: effects of species pool and local environmental conditions on stream macroinvertebrate communities. J Anim Ecol 81:679-691

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

Hawkins CP, Mykrä H, Oksanen J, Vander Laan JJ (2015) Environmental disturbance can increase beta diversity of stream macroinvertebrate assemblages. Glob Ecol Biogeogr 24: 483–494

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

Heino J (2013) Environmental heterogeneity, dispersal mode, and co-occurrence in stream macroinvertebrates. Ecol Evol 3:344–355

Heino J (2015) Deconstructing occupancy frequency distributions in stream insects: effects of body size and niche characteristics in different geographical regions. Ecol Entomol 40:491-499

Heino J, Grönroos M (2014) Untangling the relationships among regional occupancy, species traits and niche characteristics in stream invertebrates. Ecol Evol 4:1931-1942

Heino J, Ilmonen J, Paasivirta L (2014) Continuous variation of macroinvertebrate communities along environmental gradients in northern streams. Boreal Env Res 19:21-38

Heino J, Muotka T, Paavola R (2003) Determinants of macroinvertebrate diversity in headwater streams: regional and local influences. J Anim Ecol 72:425-434

Laliberté E, Legendre P, Shipley B (2014) FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-12

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

Lopes PM, Bini LM, Declerck SAJ, Farjalla V F, Vieira LCG, Bonecker CC, Lansac-Toha FA, Esteves FA, Bozelli RL (2014) Correlates of zooplankton beta diversity in tropical lake systems. PLoS ONE 9:e109581

Mayor SJ, Boutin S, He F, Cahill JF Jr (2015) Limited impacts of extensive human land use on dominance, specialization, and biotic homogenization in boreal plant communities. BMC Ecology 15:5

Merritt RW, Cummins KW (1996) An introduction to the aquatic insects of North America, third edition. Kendall/Hunt

Mykrä H, Ruokonen T, Muotka T (2006) The effect of sample duration on the efficiency of kick-sampling in two streams with contrasting substratum heterogeneity. Ver Int Ver Theor Ang Limnol 29:1351–1355

National Board of Waters and the Environment 1981. Vesihallinnon analyysimenetelmät. – Publ. Nat. Board Wat. Finland. Report 213:1–136

Nilsson A (ed.) (1997) Aquatic insects of North Europe, Volume 2. Apollo Books, Stenstrup

Poff NL, Olden JD, Vieira NKM, Finn DS, Simmons MP, Kondratieff BC (2006) Functional trait niches of North American lotic insects: traits-based ecological applications in light of phylogenetic relationships. J N Amer Bentholl Soc 25:730-755

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Siegel S, Castellan NJ (1988) Nonparametric statistics for the behavioural sciences. MacGraw Hill International, New York

Silva PG da, Hernández MIM (2014) Local and Regional Effects on Community Structure of Dung Beetles in a Mainland-Island Scenario. PLoS ONE 9:e111883

Siqueira T, Bini LM, Cianciaruso MV, Roque FO, Trivinho-Strixino S (2009) The role of niche measures in explaining the abundance–distribution relationship in tropical lotic chironomids. Hydrobiologia 636:163-172

Slatyer, R.A., Hirst, M. & Sexton, J.P. (2013) Niche breadth predicts geographical range size: a general ecological pattern. Ecol Lett 16: 1104–1114

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

Tales E, Keith P, Oberdorff T (2004) Density-range size relationships in French riverine fishes. Oecologia 138:360–370

Tonkin JD, Heino J, Sundermann A, Haase P, Jähnig S (2016) Context dependency in biodiversity patterns of central German stream metacommunities. Freshwater Biol 61: 607-620

U.S. EPA. Freshwater Biological Traits Database (Final Report) (2012) U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-11/038F, 2012

Verberk WCEP, van der Velde G, Esselink H (2010) Explaining abundance-occupancy relationships in specialists and generalists: a case study on aquatic macroinvertebrates in standing waters. J Anim Ecol 79: 589-601

Vieira NKM, Poff N, Carlisle DM, Moulton SR, Koski ML, Kondratieff BC, (2006) A database of lotic invertebrate traits for North America: U.S. Geological Survey Data Series 18

Whittaker RH (1960) Vegetation of the Siskiyou Mountains, Oregon and California. Ecol Monogr 30:279–338

## Tables

Table 1. Results of beta regression analyses when the response variable, local contributions to beta diversity (LCBD), was explained by community metrics (i.e. community richness and abundance, and their quadratic terms). Both (a) presence-absence and (b) abundance data were used for calculating LCBD.

| a) LCBD p-a | Estimate | SE | z | p | | Model Pseudo-$R^2$ |
|---|---|---|---|---|---|---|
| (Intercept) | -3.394 | 0.130 | -26.035 | 0.000 | *** | |
| Richness | -0.041 | 0.010 | -3.895 | 0.001 | *** | |
| Richness$^2$ | 0.001 | 0.000 | 3.380 | 0.001 | *** | |
| Abundance | -0.000 | 0.000 | -1.487 | 0.137 | | |
| Abundance$^2$ | 0.000 | 0.000 | 1.393 | 0.164 | | 0.368 |

| b) LCBD abund. | Estimate | SE | z | p | | Model Pseudo-$R^2$ |
|---|---|---|---|---|---|---|
| (Intercept) | -3.748 | 0.210 | -17.860 | 0.000 | *** | |
| Richness | -0.020 | 0.016 | -1.249 | 0.212 | | |
| Richness$^2$ | 0.001 | 0.000 | 1.758 | 0.079 | | |
| Abundance | -0.001 | 0.000 | -12.127 | 0.003 | * | |
| Abundance$^2$ | 0.000 | 0.000 | 1.654 | 0.098 | | 0.166 |

Table 2. Results of beta regression analyses when the response variable, local contributions to beta diversity (LCBD), was explained by local environmental variables. Both (a) presence-absence and (b) abundance data were used for calculating LCBD.

| a) LCBD p-a | Estimate | SE | z | p | | Model Pseudo-$R^2$ |
|---|---|---|---|---|---|---|
| (Intercept) | -3.459 | 0.631 | -5.479 | 0.000 | *** | |
| pH | -0.054 | 0.093 | -0.577 | 0.564 | | |
| Conductivity | 0.005 | 0.014 | 0.367 | 0.714 | | |
| Shading | 0.000 | 0.001 | 0.059 | 0.953 | | |
| Deciduous trees | -0.000 | 0.000 | -0.977 | 0.329 | | |
| Stream width | -0.000 | 0.000 | -0.557 | 0.577 | | |
| Depth | -0.002 | 0.003 | -0.762 | 0.446 | | |
| Velocity | -0.270 | 0.151 | -1.790 | 0.073 | | |
| Macrophytes | -0.002 | 0.001 | -1.546 | 0.122 | | 0.113 |

| b) LCBD abund. | Estimate | SE | z | p | | Model Pseudo-$R^2$ |
|---|---|---|---|---|---|---|
| (Intercept) | -3.654 | 0.772 | -4.732 | 0.000 | *** | |
| pH | -0.037 | 0.113 | -0.326 | 0.745 | | |
| Conductivity | 0.023 | 0.016 | 1.402 | 0.161 | | |
| Shading | -0.000 | 0.001 | -0.571 | 0.568 | | |
| Deciduous trees | -0.003 | 0.001 | -2.536 | 0.011 | * | |
| Stream width | -0.000 | 0.000 | -0.141 | 0.888 | | |
| Depth | 0.002 | 0.004 | 0.494 | 0.621 | | |
| Velocity | -0.195 | 0.186 | -1.049 | 0.294 | | |
| Macrophytes | -0.001 | 0.002 | -0.634 | 0.526 | | 0.217 |

Table 3. Results of beta regression analyses when species contributions to beta diversity (SCBD) was explained by species metrics. Both (a) presence-absence and (b) abundance data were used to calculate SCBD. NumSit = number of sites occupied, TotAbu = a species total abundance in the dataset.

| a) SCBD p-a | Estimate | SE | z | p | | Model Pseudo-$R^2$ |
|---|---|---|---|---|---|---|
| (Intercept) | -6.598 | 0.049 | -134.450 | 0.000 | *** | |
| NumSit | 0.171 | 0.006 | 29.422 | 0.000 | *** | |
| $\text{NumSit}^2$ | -0.003 | 0.000 | -18.757 | 0.000 | *** | |
| TotAbu | 0.000 | 0.000 | -0.024 | 0.981 | | |
| $\text{TotAbu}^2$ | 0.000 | 0.000 | 1.153 | 0.249 | | 0.851 |

| b) SCBD abund. | Estimate | SE | z | p | | Model Pseudo-$R^2$ |
|---|---|---|---|---|---|---|
| (Intercept) | -7.186 | 0.107 | -66.911 | 0.000 | *** | |
| NumSit | 0.141 | 0.009 | 15.563 | 0.000 | *** | |
| $\text{NumSit}^2$ | -0.002 | 0.000 | -11.513 | 0.000 | *** | |
| TotAbu | 0.002 | 0.000 | 11.248 | 0.000 | *** | |
| $\text{TotAbu}^2$ | -0.000 | 0.000 | -9.521 | 0.000 | *** | 0.717 |

Table 4. Results of beta regression analyses when species contributions to beta diversity (SCBD) was explained by species ecological and biological traits. Both (a) presence-absence and (b) abundance data were used to calculate SCBD. NicPos = Niche position, NicBre = Niche breadth. Niche characteristics were derived from the OMI analysis, and the first four PCoA axes (pco1-pco4) were used as synthetic trait vectors.

| a) SCBD p-a | Estimate | SE | z | p | | Model Pseudo-$R^2$ |
|---|---|---|---|---|---|---|
| (Intercept) | -4.698 | 0.056 | -83.801 | 0.000 | *** | |
| Niche position | -0.125 | 0.006 | -21.616 | 0.000 | *** | |
| Niche breadth | 0.084 | 0.018 | 4.612 | 0.001 | *** | |
| pco1 | -0.060 | 0.109 | -0.553 | 0.580 | | |
| pco2 | 0.098 | 0.174 | 0.567 | 0.571 | | |
| pco3 | 0.299 | 0.179 | 1.666 | 0.096 | | |
| pco4 | 0.164 | 0.205 | 0.802 | 0.423 | | 0.395 |

| a) SCBD abund. | Estimate | SE | z | p | | Model Pseudo-$R^2$ |
|---|---|---|---|---|---|---|
| (Intercept) | -5.085 | 0.125 | -40.791 | 0.000 | *** | |
| Niche position | -0.042 | 0.007 | -6.359 | 0.000 | *** | |
| Niche breadth | 0.053 | 0.034 | 1.583 | 0.113 | | |
| pco1 | -0.283 | 0.206 | -1.370 | 0.171 | | |
| pco2 | 0.106 | 0.322 | 0.328 | 0.743 | | |
| pco3 | 0.773 | 0.340 | 2.275 | 0.023 | * | |
| pco4 | 0.098 | 0.392 | 0.250 | 0.802 | | 0.288 |

## Figure legends

**Fig. 1**. Difference in local contributions to beta diversity (LCBD) (a, b), community richness (c) and community abundance (d) among the three drainage basins. Note that there was much variation within each basin in LCBD values, even though community richness and abundance also varied clearly among the basins. II=Iijoki basin, KO=Koutajoki basin, TE=Tenojoki basin. Twenty sites were surveyed in each basin

**Fig. 2.** Relationship between species contribution to beta diversity (SCBD) based on presence-absence data (a) or abundance data (b) and number of sites occupied by species. N = 203 species

**Fig. 3.** Relationships between the first four PCoA axes (pco1-pco4) and the four trait variables. FFG = functional feeding group, HTG = habit trait group, SIZE = maximum larval body length, and FDISP = female dispersal potential. See text for individual trait abbreviations

Fig. 1.

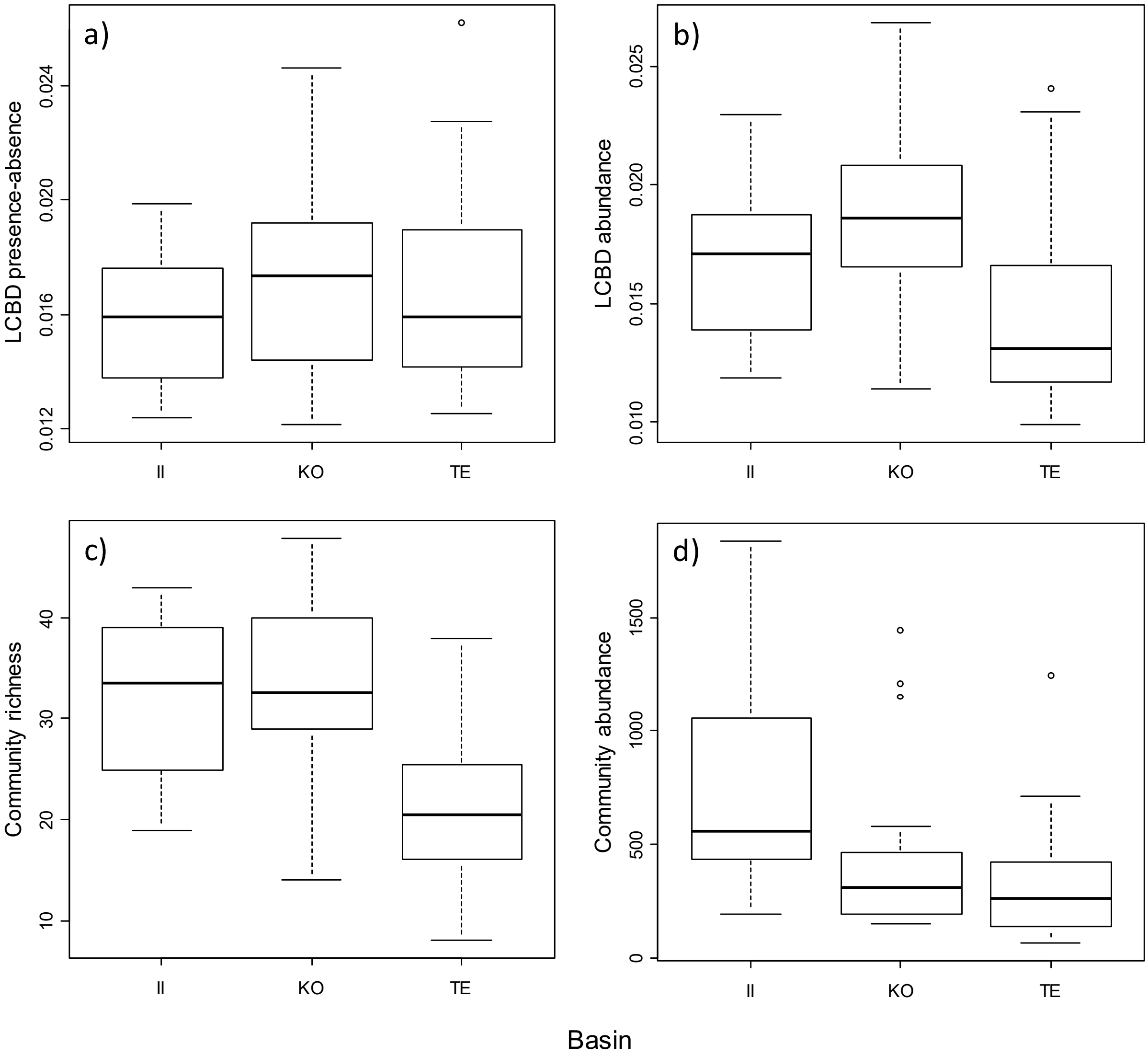

a)
LCBD presence-absence
0.012
0.016
0.020
0.024
b)
LCBD abundance
0.010
0.015
0.020
0.025
c)
Community richness
10
20
30
40
d)
Community abundance
0
500
1000
1500
II
KO
TE
Basin

Fig. 2.

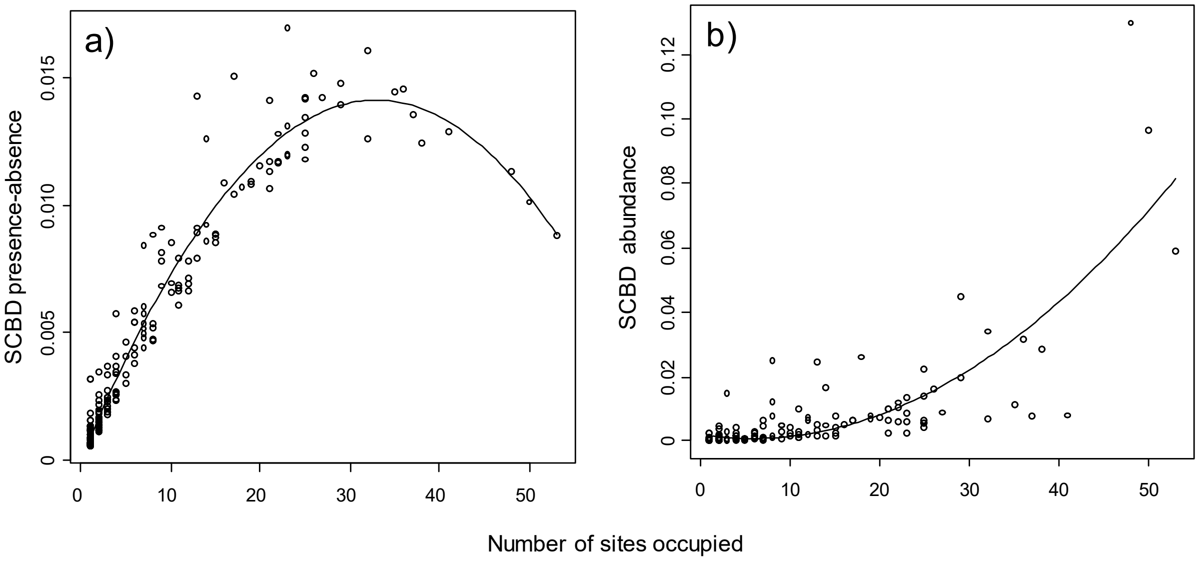
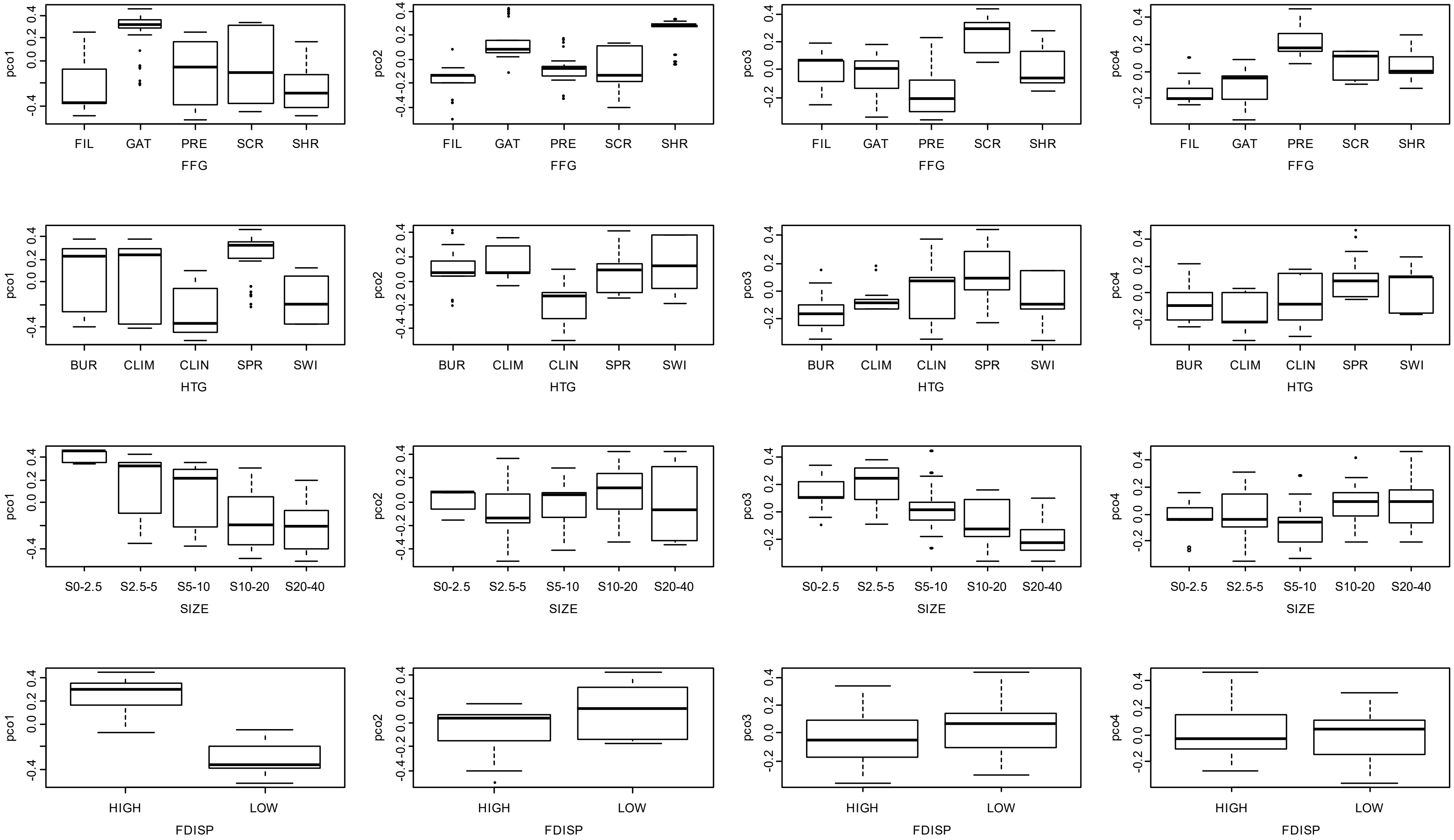

a)
SCBD presence-absence
0
0.005
0.010
0.015
b)
SCBD abundance
0
0.02
0.04
0.06
0.08
0.10
0.12
0
10
20
30
40
50
Number of sites occupied

Fig. 3.


pco1
pco2
pco3
pco4
-0.4
-0.2
0.0
0.2
0.4
FIL
GAT
PRE
SCR
SHR
FFG
BUR
CLIM
CLIN
SPR
SWI
HTG
S0-2.5
S2.5-5
S5-10
S10-20
S20-40
SIZE
HIGH
LOW
FDISP