Dispersal–niche continuum index: a new quantitative metric for assessing the relative importance of dispersal versus niche processes in community assembly

Annika Vilmi, Corentin Gibert Bret, Gilles Escarguel, Konsta Happonen, Jani Heino, Aurélien Jamoneau, Sophia I Passy, Félix Picazo, Janne Soininen, Juliette Tison-Rosebery, et al.

ECOGRAPHY

Research

Dispersal–niche continuum index: a new quantitative metric for assessing the relative importance of dispersal versus niche processes in community assembly

Annika Vilmi*, Corentin Gibert*, Gilles Escarguel, Konsta Happonen, Jani Heino, Aurélien Jamoneau, Sophia I. Passy, Félix Picazo, Janne Soininen, Juliette Tison-Rosebery and Jianjun Wang

A. Vilmi (https://orcid.org/0000-0001-5694-5802), F. Picazo (https://orcid.org/0000-0003-4586-6775) and J. Wang (https://orcid.org/0000-0001-7039-7136) ✉ (jjwang@niglas.ac.cn), State Key Laboratory of Lake Science and Environment, Nanjing Inst. of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China. – AV and J. Heino (https://orcid.org/0000-0003-1235-6613), Freshwater Centre, Finnish Environment Inst., Oulu, Finland. – C. Gibert, Laboratoire Paléontologie Evolution Paléoécosystèmes Paléoprimatologie (PALEVOPRIM), UMR-CNRS 7262, Univ. de Poitiers, Poitiers, France. – G. Escarguel, Univ. Lyon, Univ. Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, Villeurbanne, France. – K. Happonen and J. Soininen (https://orcid.org/0000-0002-8583-3137), Dept of Geosciences and Geography, Univ. of Helsinki, Helsinki, Finland. – A. Jamoneau and J. Tison-Rosebery, INRAE, EABX, Cestas, France. – S. I. Passy, Dept of Biology, Univ. of Texas at Arlington, Arlington, USA.

Patterns in community composition are scale-dependent and generally difficult to distinguish. Therefore, quantifying the main assembly processes in various systems and across different datasets has remained challenging. Building on the PER-SIMPER method, we propose a new metric, the dispersal–niche continuum index (DNCI), which estimates whether dispersal or niche processes dominate community assembly and facilitates the comparisons of processes among datasets. The DNCI was tested for robustness using simulations and applied to observational datasets comprising organismal groups with different trophic level and dispersal potential. Based on the robustness tests, the DNCI discriminated the respective contribution of niche and dispersal processes in pairwise comparisons of site groups with less than 40% and 30% differences in their taxa and site numbers, respectively. In the observational datasets, the DNCI suggested that dispersal rather than niche assembly was the dominant assembly process which, however, varied in intensity among organismal groups and study contexts, including spatial scale and ecosystem types. The proposed DNCI measures the relative strength of community assembly processes in a way that is simple, easily quantifiable and comparable across datasets. We discuss the strengths and weaknesses of the DNCI and provide perspectives for future research.

Keywords: assembly processes, determinism, dispersal, environmental filtering, niche processes, null models, PER-SIMPER, stochasticity

www.ecography.org

Introduction

Understanding how biological communities are assembled is one of the main goals of community ecology, macroecology and biogeography. The importance of correctly identifying the key mechanisms of community assembly increases as global change proceeds and relevant ecological information is urgently needed to help designing appropriate management and conservation programs (Mittelbach and Schemske 2015). Ecological determinism and stochasticity are considered two main forces underlying community assembly (Leibold et al. 2004, Chase 2007, Vellend et al. 2014). The deterministic paradigm postulates that biological communities are structured by niche properties, stemming from species' physiological responses, resource use and biotic interactions, such as competition (Vellend 2010). The stochastic paradigm, by contrast, considers biological communities as products of random processes, such as dispersal (Hubbell 2001). However, it is likely that niche and dispersal processes are not mutually exclusive but constitute the end points of a continuum, similar to the continuum of niche and neutral models proposed by Gravel et al. (2006). Within the dispersal–niche continuum, compositional variation of biological communities can be driven to a greater or lesser extent by each process (Holyoak et al. 2005, Mouillot 2007, Vellend 2010, Gibert and Escarguel 2019).

As patterns in community composition are scale-dependent (Chase et al. 2018) and underlying drivers are generally difficult to distinguish (Gotelli and Colwell 2001), quantifying the main assembly processes in various systems and across different datasets has remained challenging (Viana and Chase 2019). Null model approaches are commonly used for studying whether communities differ from null-expectations and to what extent deterministic (i.e. niche-related) processes affect this deviation (Chase et al. 2011). Variation partitioning in constrained ordination (Borcard et al. 1992) is another widely implemented method. It assesses the effects of environmental conditions (a proxy for niche processes) and space (a proxy for dispersal) on community assembly, and it relies on detailed data on community composition, environmental factors and geographic locations. However, traditional variation partitioning has been criticized for statistical and ecological reasons in the context of community assembly research, emphasizing the need for further methodological innovations. Specifically, in addition to capturing the relative importance of environmental conditions and space, the results of variation partitioning can also reflect the impact of spatial arrangement of environmental conditions, making interpretations of underlying processes difficult (Smith and Lundholm 2010, Clappe et al. 2018). Simulations have suggested that variation partitioning fails to correctly represent environmental and spatial components of community variation, sometimes resulting in poor environmental models, poor spatial models or both (Gilbert and Bennett 2010, Clappe et al. 2018).

To address some of the data and methodological limitations of previous methods, Gibert and Escarguel (2019) recently developed a method called PER-SIMPER. PER-SIMPER allows the qualitative identification of the dominant assembly mechanism based solely on a matrix of species occurrences across sites. Building on SIMPER (Clarke 1993), this permutation-based (hence, 'PER-') null model approach identifies the dominant assembly process accounting for compositional similarity percentages between site groups consisting of local communities (hence, 'SIMPER') within the same regional species pool (Gibert and Escarguel 2019). PER-SIMPER utilizes sites-by-taxa community matrices, generating three distinct null models, constraining rows ('niche assembly'), columns ('dispersal assembly') or both during permutations. It uses the SIMPER (Clarke 1993) method to model the original community matrix compositional similarity pattern and to detect the empirical model profile, which is then compared to the three null distribution profiles (hereafter called 'null PER-SIMPER profiles', Fig. 1). PER-SIMPER makes a qualitative assessment by identifying which null profile matches best the empirical profile. However, the sensitivity of a qualitative approach to assembly mechanisms can be limited by the fact that most communities are structured by both niche and dispersal processes (Mouillot 2007). In addition, the qualitative nature of the PER-SIMPER decision-making process prevents precise comparisons of the assembly process in different metacommunities.

As a solution to these limitations, we here propose a new metric, the dispersal–niche continuum index (DNCI) with broad ecological application. Specifically, the PER-SIMPER analysis returns three E-metric distributions, which correspond to the deviation between the empirical SIMPER profile and the three null PER-SIMPER profiles generated from the three permutation modes of the original matrix, where each permutation mode is linked to one of the three main assembly perspectives (i.e. dispersal assembly, niche assembly or joint dispersal–niche assembly). The new DNCI proposed here is derived from these computed E-values. Quantitative identification of the assembly process is based on the subtraction of the standard effect size (SES) of E_n (i.e. E-metric distribution from the 'niche' model) from the standard effect size of E_d (i.e. E-metric distribution from the 'dispersal' model). Thus, DNCI offers a way to both quantify and compare the strengths of the main assembly processes across datasets (Fig. 1). Positive or negative DNCI values indicate that niche or dispersal assembly is the main process structuring the studied site groups consisting of biological communities, respectively. Higher absolute values of the index represent greater potential strength of the dominant assembly process.

To test the robustness of the DNCI, we performed robustness tests with simulations. We then calculated the DNCI for two example datasets to test the applicability of the DNCI with empirical data. Using stream monitoring data from the USA, we compared whether the DNCI values differ among metacommunities of organisms varying in trophic level and dispersal capacity, including diatoms, macroinvertebrates and fish. Using a field study dataset from China, we addressed whether the DNCI values of stream bacteria and macroinvertebrates differ from each other and whether they vary with
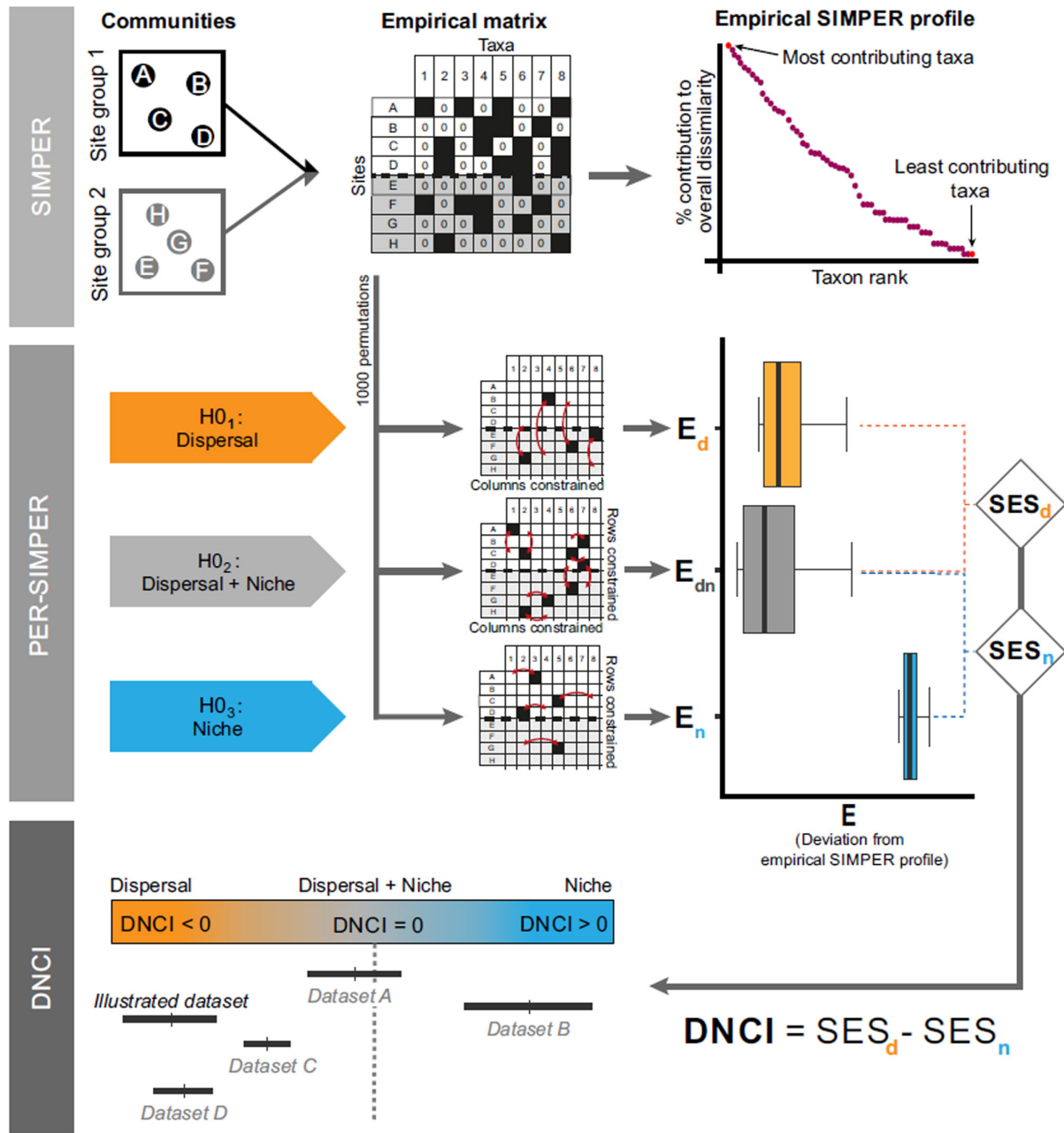
Figure 1. Flowchart of PER-SIMPER derivation from the initial SIMPER method and the development and interpretation of the DNCI. Negative DNCI values indicate that dispersal assembly is the dominating process, while positive DNCI values indicate that niche assembly is the dominating process behind community dissimilarities among site groups. With increasing absolute values, the strength of the dominating process increases. DNCI values close to zero indicate that both niche and dispersal assembly processes affect distributions. In dataset A, the DNCI does not significantly differ from zero, indicating that both dispersal and niche processes control the empirical species distributions. In dataset B, DNCI is significantly greater than zero, indicating that niche assembly is the main driver of the empirical species distribution. In datasets C and D and the illustrated dataset, DNCI is significantly less than zero, indicating that dispersal assembly is the main driver of the empirical species distributions. Dataset D and the illustrated dataset do not have significantly different DNCI values, i.e. the control of the empirical species distributions by dispersal assembly is similar. Dataset C has a DNCI significantly higher than dataset D and the illustrated dataset, i.e. dispersal assembly control in dataset C is significantly weaker than in dataset D and the illustrated dataset. SES_d = standard effect size of E_d. SES_n = standard effect size of E_n. Modified from Gibert and Escarguel's (2019) Supporting information.

3

spatial distance. Finally, we discussed the strengths and weaknesses of the DNCI in inferring community assembly processes and suggested potential research avenues.

Material and methods

Description of DNCI

The initial approach of PER-SIMPER (Gibert and Escarguel 2019) compares an empirical SIMPER profile with three null PER-SIMPER profiles and qualitatively identifies which null profile and thus which assembly mode (i.e. dispersal assembly, niche assembly or both) is responsible for the composition of the analyzed communities (Fig. 1). To this end, the empirical SIMPER matrix is permuted either by maintaining the richness of localities (the row sums) in order to retain information on the potential number of ecological niches per sample locality, or by maintaining the number of localities where the taxa may have dispersed (the column sums) in order to retain information on the potential dispersal capacity of taxa. The third permutation mode maintains both row and column sums, retaining a larger part of the sampled information. While the number of localities where a species was sampled can provide indication for dispersal capacity, the taxonomic richness of a community can reflect the number of ecological niches, with richer communities having more niches (Granot and Belmaker 2020). The identification of the assembly mode relies on the E-metric, corresponding to the square of the deviations between the empirical SIMPER profile and multiple null PER-SIMPER profiles produced with the three permutation modes (by default, 1000 permutations for each assembly mode). This approach can be problematic for two main reasons. First, its sensitivity is low because most communities are structured by both niche and dispersal processes (Mouillot 2007). Second, being qualitative, this approach prevents precise comparisons (in space and time) of the assembly processes in different metacommunities. To address these weaknesses, we developed the DNCI based on the three E-metric values:

$$DNCI = SES_d - SES_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{E_{d(i)} - \overline{E_{dn}}}{\sigma E_{dn}}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{E_{n(i)} - \overline{E_{dn}}}{\sigma E_{dn}}\right)$$

where SES_d and SES_n are the standard effect sizes of E_d and E_n, respectively, E_d, E_n and E_{dn} are the E-metric values (i.e. deviation from the empirical SIMPER profile) for the three PER-SIMPER null models (i.e. 'dispersal assembly' permutation mode, 'niche assembly' permutation mode and 'dispersal and niche assembly' permutation mode, respectively), and the standalone n is the number of iterations selected for the PER-SIMPER method.

The standard deviation related to DNCI is obtained by simple error propagation from the standard deviations associated with SES_d and SES_n, such as:

$$\sigma_{DNCI} = \sqrt{\sigma_{SES_d}^2 + \sigma_{SES_n}^2}$$

If the DNCI is not significantly different from 0, the dispersal and niche processes could be assumed to contribute equally to variations in community composition. If the DNCI is significantly lower than 0, dispersal processes are the dominant drivers of community composition. Conversely, if the DNCI is significantly higher than 0, niche processes are the primary determinants of community composition (Fig. 1). It is noteworthy that negative DNCI values indicating the predominance of dispersal processes do not give information on actual dispersal rates.

DNCI robustness test with simulated communities

PER-SIMPER's original qualitative approach has shown accurate identification of the main assembly process underlying the composition of simulated communities, even when the analysed groups have different number of taxa or result from various sampling intensities. Furthermore, PER-SIMPER's qualitative identifications are particularly robust when one assembly process plays a role significantly larger than the other one. Conversely, when both processes are simulated, PER-SIMPER is more sensitive to sampling biases. In particular, the number of sampled sites (less than five sites in a site group) within the analysed dataset may bias the PER-SIMPER inference toward the niche-based expectation by underestimating the effect of dispersal processes (Gibert and Escarguel 2019).

Here, the effects of number of taxa and number of sites are tested and measured on simulated datasets prior to using the new DNCI on empirical datasets. To ensure consistency between the quantitative and qualitative robustness tests, the simulated communities are obtained from cellular automata detailed by Gibert and Escarguel (2019). The automata are designed for simulating the different displacement potential of various taxa (dispersal) as well as the influence of local environmental conditions on taxonomic richness (niche). The dispersal potential of taxa and the taxonomic richness limit inside cells can be tuned by, respectively, changing the number of individuals moving along the automaton grid (i.e. the more individuals, the more dispersal) and by changing the pattern limiting maximum richness within the cells. Using these two parameters, it is possible to simulate a complete continuum of communities from those strongly constrained by niche processes to those strongly constrained by dispersal as well as a set of intermediate types of communities.

The final occurrence matrices are composed of 30 sites per group (i.e. 60 sites in total) and 100–150 taxa. Three simulated matrices, one assembled only by niches, one assembled only by dispersal and one assembled jointly by niches and dispersal, are used as references. Their measured DNCI values are then compared to the DNCI values computed for site groups with increasing proportions of removed taxa and sites. Specifically, 10–70% of the sites are randomly removed, and 10–70% of the simulated taxa are randomly removed.

The removals were done in one site group only or in both site groups simultaneously.

Observational field survey data on diatoms, macroinvertebrates and fish from the United States

Datasets including stream diatoms, macroinvertebrates and fish were obtained from the National Water-Quality Assessment (NAWQA) Program of the U.S. Geological Survey. We selected four subregions of the Mid-Atlantic Region (hydrologic unit HUC02), situated in the northeastern part of the United States, including the Upper Hudson (HUC0202), Delaware (HUC0204), Susquehanna (HUC0205) and the Potomac (HUC0207) (map in Supporting information). In total, there were 131 sites sampled for diatoms, 128 sites sampled for macroinvertebrates and 43 sites sampled for fish across the four subregions (i.e. four site groups). Each of the subregions comprised a sufficient number of sites for the analyses (a minimum of eight sites in Delaware for fish and a maximum of 66 sites in the Potomac for macroinvertebrates), given that at least five sites per group is recommended by Gibert and Escarguel (2019). Importantly, the subregions likely shared the same regional species pools as they belong to the same hydrologic region. The streams were sampled between 1993 and 2011. We included only summer months, i.e. June, July and August, to eliminate seasonal effects on community compositions and assembly processes, which can make interpretations more complex (Li et al. 2020). We then selected taxa identified at the species level for diatoms (444 species) and fish (93 species), and at species or genus levels for macroinvertebrates (429 taxa).

Observational field survey data on bacteria and macroinvertebrates from China

In October 2014, 89 individual streams were sampled for bacteria and macroinvertebrates in the southeastern part of the Tibetan Plateau, Yunnan province, China. The sampled streams include 52 tributaries of the Salween River and 37 tributaries of the Mekong River (see map in Supporting information). The elevational ranges were pronounced, spanning from 570 to 1592 m a.s.l. in the Salween catchment and from 1270 to 2100 m a.s.l. in the Mekong catchment. Macroinvertebrate samples were identified to genus level when possible. Bacterial OTUs (i.e. operational taxonomic units) were determined based on the 16S rRNA genes using bacterial universal primers 515F and 806R that target the V4 region. The bacterial dataset was rarefied to 10 000 sequences. Further details on field sampling and laboratory procedures are presented in Wang et al. (2017) and Vilmi et al. (2020).

Statistical methods

We developed an R function ('DNCI_multigroup') for the calculation of DNCI values. The function first performs the PER-SIMPER procedure (Gibert and Escarguel 2019) and, then, as described earlier, calculates pairwise DNCI values among site groups based on the E-values from the PER-SIMPER analysis. For the US data, we used the four subregions as site groups for all three presence/absence community matrices (i.e. diatoms, macroinvertebrates and fish). Due to differing numbers of sites per group, we selected the size of the smallest group (Delaware with eight sites in the fish dataset) and randomly resampled sites from all other groups to eliminate a potential group size bias. We performed the random resampling of eight sites from each of the four groups 200 times. For the Chinese data, we used elevational bins, each containing 12 or 13 sites, as site groups in our analyses (Supporting information). In both datasets, we removed singletons (i.e. taxa occurring at one site only) prior to calculation of the DNCI values.

We employed the Kruskal–Wallis test followed by posthoc Dunn tests with p-value adjustments (method = 'bh') to study whether the DNCI differed across diatoms, macroinvertebrates and fish in the US Mid-Atlantic hydrologic subregions, and the Wilcoxon rank sum test to study whether the DNCI of bacteria and macroinvertebrates differed from each other in the Chinese catchments. Separately for bacteria and macroinvertebrates, the Kruskal–Wallis test followed by posthoc Dunn tests with p-value adjustments (method = 'bh') was used to study whether within- and across-catchment DNCIs differed from each other in the Chinese dataset. With Mantel tests, we studied whether the DNCI of Chinese bacteria and macroinvertebrates varied with distance using pairwise DNCI values and pairwise spatial distances of the site groups. As the computed DNCI values were all negative, we performed Mantel tests with absolute DNCI values (Mantel test does not process negative values). The absolute DNCI values simply indicate the distances to zero. After analyses with absolute values, the original DNCI values (i.e. negative, in our case) were plotted against the pairwise spatial Euclidean distances of the site groups. As the relationships were apparently negative, the Mantel r statistic was changed to negative. It is important to note that Mantel tests can only be used for cases where all DNCI values are either positive or negative, the latter after the transformation described above. We made network plots based on the results from China to graphically illustrate the variation of DNCI in space.

All statistical analyses and plots were done using packages DNCImper (Gibert et al. 2020), vegan (Oksanen et al. 2018), FSA (Ogle et al. 2020), ggplot2 (Wickham 2016), igraph (Csardi and Nepusz 2006) and dichromat (Lumley et al. 2013) with the R statistical software (<www.r-project.org>). The functions and codes for DNCI calculation and examples for preparing plots are available in GitHub (<https://github.com/Corentin-Gibert-Paleontology/DNCImper>) and as Supporting information.

Results

Robustness tests

If only niche effects are simulated, removal of 40% of the taxa or 30% of the sites in one site group of the dataset can lead
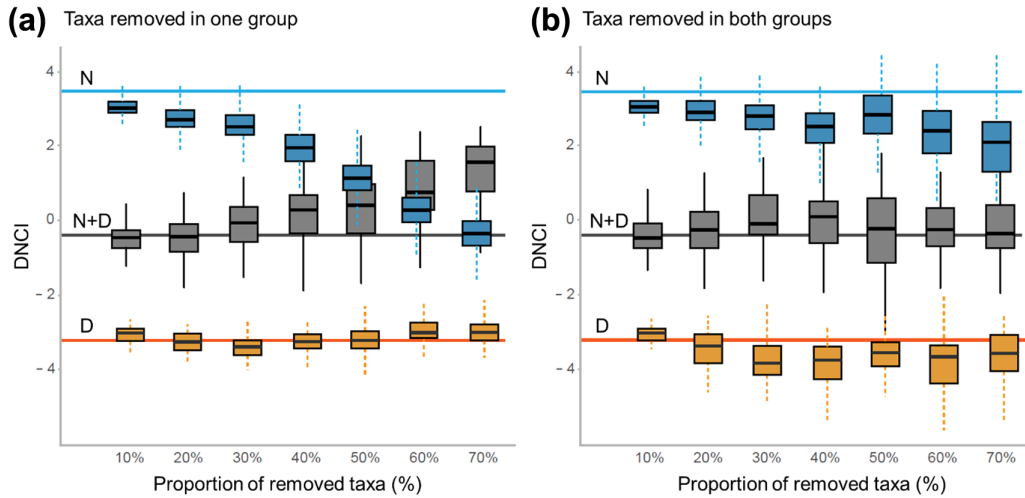
Figure 2. Distribution of DNCI (1000 iterations) after degradation of three reference datasets: niche assembled (N, blue line), niche and dispersal assembled (N + D, grey line) and dispersal assembled (D, orange line). The number of taxa is randomly decreased by 10–70% within a single group of sites of the reference dataset (a). The number of taxa is randomly decreased by 10–70% in each of the two groups of sites in the reference dataset (b).

the DNCI analysis to fail in identifying the correct assembly mechanism (Fig. 2a, 3a, blue boxplots). For assemblages simulated solely via dispersal effects, the removal of 50% of the sites in one site group can also lead to the identification of the wrong assembly mechanisms (Fig. 3a, orange boxplot). Consequently, the DNCI is more robust to variation in number of taxa than in number of sites but can also be highly sensitive to specific variation in number of sites (Fig. 2, 3). These results are consistent with preliminary robustness tests conducted on the qualitative PER-SIMPER method (Gibert and Escarguel 2019). However, the computation of robustness tests on a quantitative index provides a more precise understanding of the consequences of sampling effort variation on

the method's ability to detect assembly mechanisms and the intensity of their respective contributions. First, the computation of a single, robust DNCI value requires groups with a limited difference in number of taxa and sites (Fig. 2a, 3b). Specifically, comparing groups with deviation in number of taxa exceeding 40% is not recommended as it leads to underestimation of the niche processes (Fig. 2a). In addition, to discriminate the respective contribution of niche and dispersal processes, the DNCI must be computed for site groups with less than 30% differences in site numbers, which is especially important for communities assembled by both niche and dispersal processes (Fig. 3a). The DNCI is, therefore, to some extent sensitive to the symmetry of the
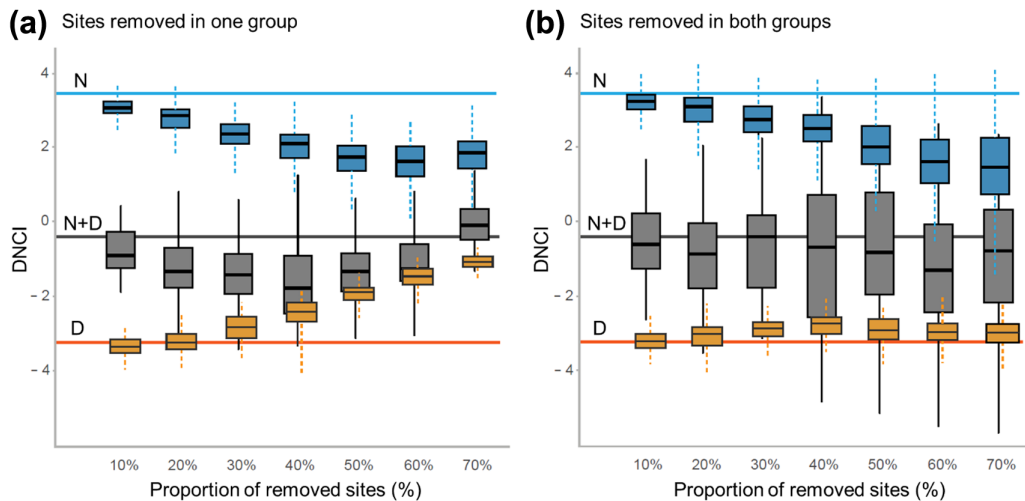

Figure 3. Distribution of DNCI (1000 iterations) after degradation of three reference datasets: niche assembled (N, blue line), niche and dispersal assembled (N + D, grey line) and dispersal assembled (D, orange line). The number of sites is randomly decreased by 10–70% within a single group of sites in the reference dataset (a). The number of sites is randomly decreased by 10–70% in each of the two groups of sites in the reference dataset (b).

occurrence matrix of interest. Second, DNCI values can be compared accurately across datasets of different size both in terms of taxa and site numbers, as DNCI values are insensitive to symmetric removal of sites or taxa (at least up to 70% differences; Fig. 2b, 3b). Consequently, random re-sampling (Sanders 1968, Raup 1972, Gotelli and Colwell 2001) may be necessary when computing DNCI between site groups of dissimilar size within a dataset, but not when comparing DNCI values between distinct datasets of dissimilar size. This is a significant improvement in terms of precision compared to PER-SIMPER, which confounds the effect of matrix symmetry with the effect of unequal sample size.

Specific applications of DNCI

Identifying the main community assembly process in major stream organismal groups

The DNCI values of diatoms, macroinvertebrates and fish from the Mid-Atlantic hydrological subregions of the US were negative, suggesting that dispersal was the dominant assembly process for all three organismal groups (Fig. 4). The DNCI values of diatoms significantly differed from the DNCI values of macroinvertebrates and fishes ($p < 0.001$). Dispersal assembly was, in a relative sense, most intense
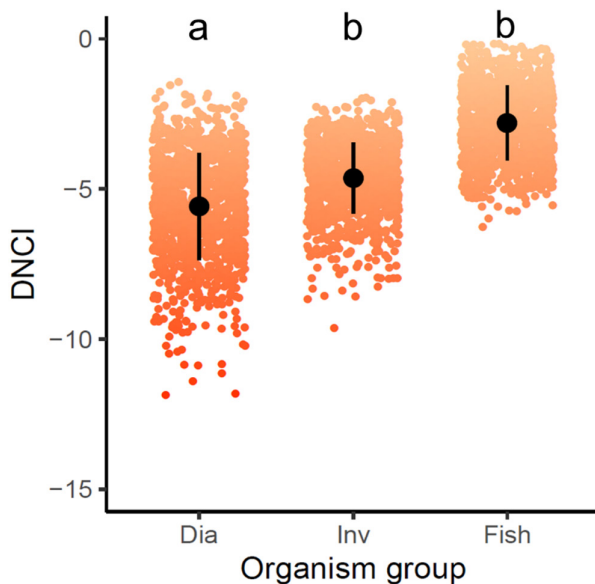

Figure 4. DNCI values of diatoms (Dia), macroinvertebrates (Inv) and fish (Fish) from four Mid-Atlantic hydrological subregions in the US. Each dot represents the DNCI value between two groups computed in one iteration, i.e. the total number of DNCI values computed for each organismal group is 1200 (six pairwise comparisons across site groups, 200 iterations). Negative DNCI values indicate that dispersal is the dominant assembly process. Higher absolute values (darker colour) indicate greater strength of the dominant assembly process. Different letters represent statistically different DNCI values following the Kruskal–Wallis and post-hoc Dunn tests ($p < 0.001$). The black dots illustrate the mean values and the lines represent the standard deviation.

for diatoms and less intense for macroinvertebrates and fishes (Fig. 4).

Identifying the main community assembly process for stream bacteria and macroinvertebrates along an elevational gradient

The DNCI values of bacteria and macroinvertebrates were negative, suggesting dominance of dispersal assembly in the two Chinese catchments. The DNCI differed significantly between the two organismal groups ($p = 0.016$), with bacteria being more strongly assembled by dispersal than macroinvertebrates. The DNCI of bacteria showed a decreasing trend with increasing spatial distances (Mantel $r = -0.486$, $p = 0.043$; Fig. 5a). For macroinvertebrates, a relationship between the DNCI and distance was not evident (Mantel $r = -0.153$, $p = 0.251$; Fig. 5b). The DNCI of bacteria was constant within and across the studied catchments (Fig. 5c; $p = 0.326$). The difference in DNCI of macroinvertebrates within the Mekong and Salween catchments was close to the $p = 0.05$ level of statistical significance (overall Kruskal–Wallis test $p = 0.056$, post-hoc Dunn test for comparisons of within-catchment DNCIs $p = 0.050$; Fig. 5d), with DNCI values being more negative within the Salween than the Mekong catchment. Network plots (Fig. 5e–f) illustrated the spatial setting and relationships between site groups (here, elevational bins), and complement the information presented in Fig. 5a–d. For bacteria, dispersal assembly appeared to be strongest between the lower mid elevations (S-MID2) and high elevations (S-HIGH) within the Salween catchment, as well as between the lower mid elevations of the Salween catchment (S-MID2) and high elevations of the Mekong catchment (M-HIGH) (Fig. 5e). For macroinvertebrates, the strongest dispersal assembly signal was observed across the catchments, i.e. between the low elevations in the Salween catchment and mid elevations of the Mekong catchment (Fig. 5f). Interestingly, for both bacteria and macroinvertebrates, dispersal assembly was weaker in the highest elevations across the two catchments (S-HIGH and M-HIGH). For macroinvertebrates, community dissimilarities were more weakly driven by dispersal assembly across the highest elevations in the Mekong catchment and all other elevations within and across catchments (M-HIGH versus all other elevational bins).

Discussion

Here, we developed a novel quantitative index, DNCI, to disentangle the importance of niche and dispersal processes in community assembly, with distinct advantages over existing methods, including earlier null model approaches (Chase et al. 2011), variance partitioning in constrained ordination (Borcard et al. 1992) and PER-SIMPER's original qualitative inference procedure (Gibert and Escarguel 2019). The advantages of the DNCI include, 1) reliance on taxa distributions rather than summary statistics, with the potential to better assess assembly processes (cf. Chase et al. 2011), 2) relative simplicity, given that it does not require abundance,
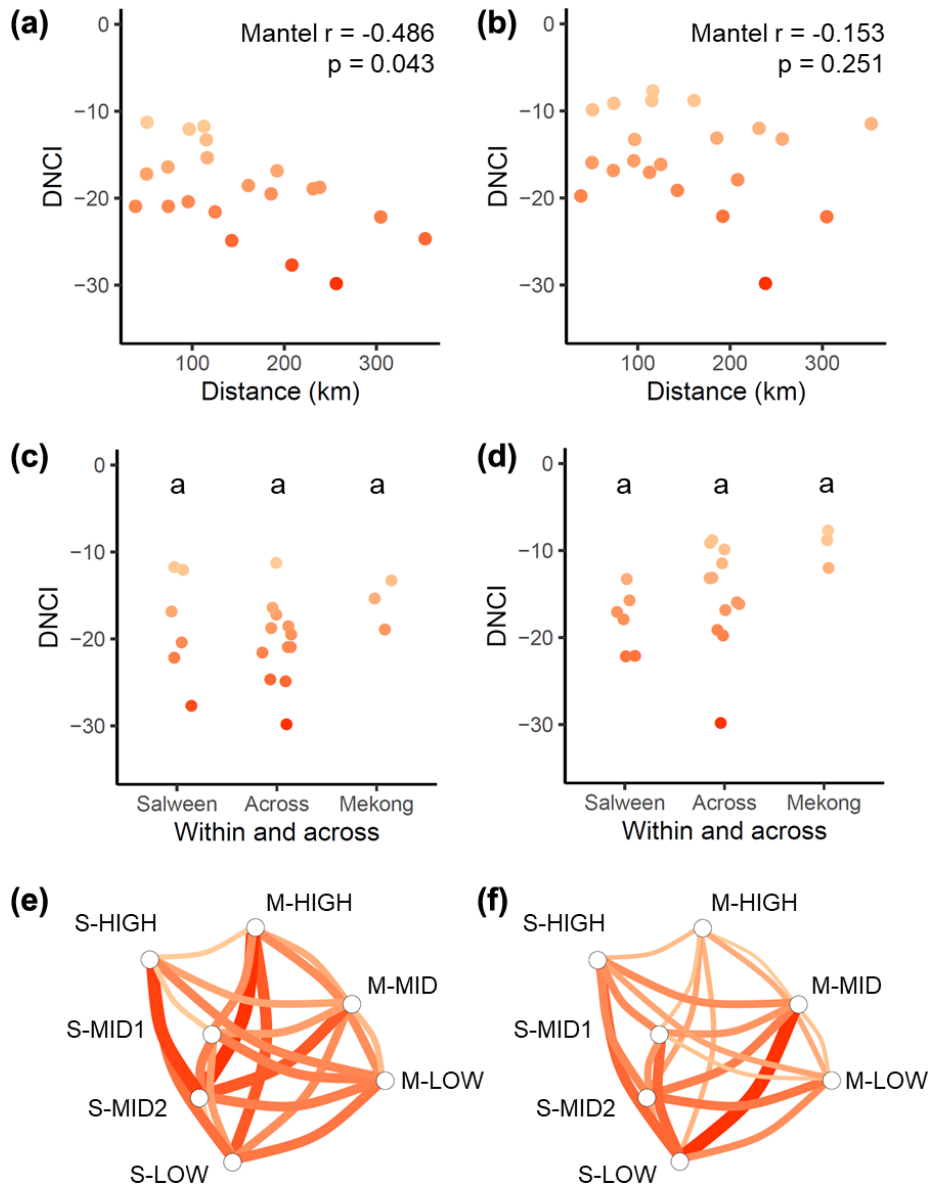
Figure 5. The DNCI values for bacteria (left column) and macroinvertebrates (right column) in the Chinese dataset. There was a significant negative relationship between DNCI of bacterial communities and distance (a), suggesting that intensity of dispersal assembly grew with increasing distances. There was no clear relationship between the DNCI of macroinvertebrate communities and distance (b). Based on the Kruskal–Wallis test, the DNCI of bacteria within and across catchments did not differ from each other (c; p = 0.326). Based on the Kruskal–Wallis and post-hoc Dunn tests, the DNCI of macroinvertebrates was almost statistically significantly different within the Mekong and the Salween catchments (d; overall Kruskal–Wallis p = 0.056, post-hoc Dunn p = 0.050). Network plots illustrate the DNCI, i.e. the strength of the dominating assembly process, within and across catchments for bacteria (e) and macroinvertebrates (f). The nodes in the network plots illustrate site groups, each containing 12 or 13 sites, along elevational gradients in the Salween (S) and Mekong (M) catchments. The edges connecting the nodes in the network plots illustrate the intensity of the dominant assembly process; the darker the colour and wider the line, the more intense the process (i.e. the bigger the absolute value of the DNCI).

environmental or spatial data and, 3) high comparability across datasets, even those that differ greatly in numbers of sampled localities and taxa. Thus, the DNCI allows evaluating whether the dominant assembly mechanism and its strength vary across different datasets.

Based on observational data from the US and China, the DNCI-based results suggested that communities of distinct organismal groups, from microbes to fishes, were mainly assembled by dispersal processes. Differences in the DNCI were partly associated with organismal groups and geographical contexts. In the empirical datasets from the US and China, DNCI values varied among organismal groups, with dispersal assembly being slightly but significantly stronger for the metacommunities of smaller organisms. Furthermore, in the

Chinese data, our results implied that the intensity of dispersal assembly may increase at greater distances for microbes. Although the DNCI does not inform us about actual dispersal rates, it can perhaps be reasoned that the distance-related DNCI pattern observed for the Chinese microbes may suggest that the occurrences are, to some degree, limited by dispersal in the two catchments. Overall, dispersal was the dominant assembly process in the studied regional datasets, but further research is needed to assess whether the intensity of dispersal assembly changes across larger spatial distances and environmental gradients in other datasets.

For the Chinese data, some pairs of site groups were very strongly structured by dispersal both within and across catchments, while some pairs of site groups were structured by a more balanced mix of dispersal and niche processes within and across catchments (Fig. 5e–f). These two examples illustrate some of the limitations of the DNCI approach related to the use of occurrence matrices and pairwise comparisons. First, the use of a single occurrence matrix without additional environmental information makes the use of DNCI extremely straightforward, but it also implies a lack of information when it comes to the effect of niche mechanisms on the distribution of taxa. Only the assumed effect of niche mechanisms on the presence/absence of taxa and, therefore, on the richness of localities (i.e. row sums) are compared with the assumed effect of dispersal potential of taxa (i.e. column sums). This bias might be important especially for small organisms, such as bacteria, where large differences in abundance between sites rather than richness may be observed. Ecological differences between and within the two catchments were not strong enough compared to differences in dispersal potentials to explain the dissimilarity structure and thus the composition (in terms of occurrence) of assemblages. This was evidenced by the highly negative DNCI values obtained from the datasets collected from the two catchments. For macroinvertebrates, the three site groups within the Mekong catchment showed a slightly more balanced situation between dispersal and niche processes, which was indicated by higher DNCI values and, therefore, a comparatively increased role of niche conditions on the presence and absence of the sampled taxa. Second, the DNCI approach analyses pairs of assemblages. Therefore, a single analysis does not allow us to conclude on the nature of the structure of one community but on the nature of the assembly processes that differentiate groups of local communities. Here, network plots can be useful for interpreting results. By simultaneously observing the DNCI values of multiple pairs of site groups, we can identify the pairs intensely structured by dispersal (such as S-MID2–M-HIGH for bacteria and S-LOW–M-MID for macroinvertebrates) as well as those comparatively more marked by the effect of niche conditions (such as M-HIGH paired with all other site groups for macroinvertebrates). Our general results that communities from different spatial site groups were more differentiated by dispersal than niche processes can be partly seen as a consequence of the pairwise nature of the method and, in our example datasets, of a comparatively

weak environmental effect, assumingly portrayed by the local taxonomic richness, on the presence or absence of taxa.

Although the DNCI values based on our observational example datasets were all negative, indicating that dispersal assembly was dominant for all tested aquatic metacommunities from two continents, it is likely that datasets from different biological, environmental, spatial and temporal settings may produce both positive and negative DNCI values. Exploring the sensitivity of DNCI can guide future research to assess how community assembly is determined by the following factors, for instance: 1) spatial extent of studied areas (i.e. site groups in the analysis); 2) spatial distances among studied areas (site groups) and within studied areas (sites); 3) physical connectivity within and across site groups; 4) organismal groups, taxonomic levels and functional traits (e.g. dispersal ability, stress tolerance or feeding behaviour), 5) environmental heterogeneity versus homogeneity; 6) successional phase; 7) seasonality; 8) temporal environmental trends, such as progressive eutrophication, land use changes and climate change. The more the proposed DNCI approach is applied, the more information we obtain on how accurate the pairwise calculations based on presence–absence matrices of site groups are for inferring the intensities of dispersal and niche processes in community assembly in various study settings.

There are a few caveats for applying the DNCI approach. First, an appropriate spatial scale must be defined, whereby all local communities share at least partly the same regional species pool (i.e. dispersal of organisms potentially occurs within the study context). Second, based on our simulations, the numbers of taxa and sites per group should not vary more than 40% and 30%, respectively. Third, since this approach uses occurrence data, the assembly processes identified by this approach are limited to those resulting from the presence or absence of taxa, and the driving mechanisms of abundance variation are left undetected, suggesting a further need for methodological development of the DNCI. Fourth, distances between sites and site groups and, more generally, the overall spatial setting, are not directly accounted for (but indirectly via the occurrence matrix structure), potentially complicating interpretation of results. Fifth, unlike methods which use numerous environmental variables as predictors (e.g. variation partitioning), the DNCI alone cannot show which types of environmental variables contribute to variation in community structure across sites and site groups.

Keeping in mind the requirements for shared regional species pools and limited size variations among site groups as well as the caveats described above, the proposed DNCI approach can shed light into the importance of niche and dispersal processes by measuring the strength of community assembly processes in a way that is simple, quantifiable and easily comparable across datasets. We believe that this method is a useful addition to the toolbox of biogeographers, ecologists and palaeontologists; if used for datasets from different environmental and spatial settings, it would increase understanding of community assembly in nature.

References

Borcard, D. et al. 1992. Partialling out the spatial component of ecological variation. – Ecology 73: 1045–1055.

Chase, J. M. 2007. Drought mediates the importance of stochastic community assembly. – Proc. Natl Acad. Sci. USA 104: 17430–17434.

Chase, J. M. et al. 2011. Using null models to disentangle variation in community dissimilarity from variation in a-diversity. – Ecosphere 2: 24.

Chase, J. M. et al. 2018. Embracing scale-dependence to achieve a deeper understanding of biodiversity and its change across communities. – Ecol. Lett. 21: 1737–1751.

Clappe, S. et al. 2018. Beyond neutrality: disentangling the effects of species sorting and spurious correlations in community analysis. – Ecology 99: 1737–1747.

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. – Aust. J. Ecol. 18: 117–143.

Csardi, G. and Nepusz, T. 2006. The igraph software package for complex network research. – InterJ. Complex Syst. 1695, <http://igraph.org>.

Gibert, C. and Escarguel, G. 2019. PER-SIMPER – a new tool for inferring community assembly processes from taxon occurrences. – Global Ecol. Biogeogr. 28: 374–385.

Gibert, C. et al. 2020. DNCImper: assembly process identification based on SIMPER analysis. – R package ver. 0.0.1.0000, <https://github.com/Corentin-Gibert-Paleontology/DNCImper>.

Gilbert, B. and Bennett, J. R. 2010. Partitioning variation in ecological communities: do the numbers add up? – J. Appl. Ecol. 47: 1071–1082.

Gotelli, N. J. and Colwell, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. – Ecol. Lett. 4: 379–391.

Granot, I. and Belmaker, J. 2020. Niche breadth and species richness: correlation strength, scale and mechanisms. – Global Ecol. Biogeogr. 29: 159–170.

Gravel, D. et al. 2006. Reconciling niche and neutrality: the continuum hypothesis. – Ecol. Lett. 9: 399–409.

Holyoak, M. et al. 2005. A framework for large scale community ecology. – In: Holyoak, M. et al. (eds), Metacommunities: spatial dynamics and ecological communities. Univ. of Chicago Press, pp. 1–31.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. – Princeton Univ. Press.

Leibold, M. A. et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. – Ecol. Lett. 7: 601–613.

Li, Z. et al. 2020. Seasonal changes in metacommunity assembly mechanisms of benthic macroinvertebrates in a subtropical river basin. – Sci. Total Environ. 729: 139046.

Lumley, T. et al. 2013. dichromat: color schemes for dichromats. – <https://CRAN.R-project.org/package=dichromat>.

Mittelbach, G. G. and Schemske, D. W. 2015. Ecological and evolutionary perspectives on community assembly. – Trends Ecol. Evol. 30: 241–247.

Mouillot, D. 2007. Niche-assembly vs dispersal-assembly rules in coastal fish metacommunities: implications for management of biodiversity in brackish lagoons. – J. Appl. Ecol. 44: 760–767.

Ogle, D. H. et al. 2020. FSA: fisheries stock analysis. – R package ver. 0.8.31, <https://github.com/droglenc/FSA>.

Oksanen, J. et al. 2018. Package 'vegan', ver. 2.5-2. – <http://cran.r-project.org>, <http://vegan.r-forge.r-project.org/>.

Raup, D. M. 1972. Taxonomic diversity during the Phanerozoic. – Science 177: 1065–1071.

Sanders, H. L. 1968. Marine benthic diversity: a comparative study. – Am. Nat. 102: 243–282.

Smith, T. W. and Lundholm, J. T. 2010. Variation partitioning as a tool to distinguish between niche and neutral processes. – Ecography 33: 648–655.

Vellend, M. 2010. Conceptual synthesis in community ecology. – Q. Rev. Biol. 85: 183–206.

Vellend, M. et al. 2014. Assessing the relative importance of neutral stochasticity in ecological communities. – Oikos 123: 1420–1430.

Viana, D. S. and Chase, J. M. 2019. Spatial scale modulates the inference of metacommunity assembly processes. – Ecology 100: e02576.

Vilmi, A. et al. 2020. Ecological processes underlying the assembly of aquatic bacterial and macroinvertebrate communities under contrasting climates on the Tibetan Plateau. – Sci. Total Environ. 702: 134974.

Wang, J. et al. 2017. Regional and global elevational patterns of microbial species richness and evenness. – Ecography 40: 393–402.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. – Springer, <https://ggplot2.tidyverse.org>.