**Homogeneous selection and stochasticity overrule heterogeneous selection across biotic taxa and ecosystems**

**Abstract**

Deterministic and stochastic factors shape ecological communities. However, a quantitative synthesis of the factors underlying the balance among different assembly processes is lacking. Here, we synthesized data from 149 datasets covering major biotic groups and ecosystem types globally. We used a null model approach based on Raup-Crick dissimilarities and Bayesian meta-regression to analyze the data. We found that communities were more under homogeneous selection than heterogeneous selection across biotic taxa and ecosystems. Environment selected species homogeneously more often at small scales while heterogeneously more often at large scales. Stochasticity also showed scale-dependence as stochastic community assembly increased with study scale. Homogeneous and heterogeneous selection were strongest at high latitudes while stochastic factors were strongest in tropics. Marine systems had the highest degree of homogeneous selection and the lowest stochasticity. We provide the first analysis of community assembly across taxa and ecosystems which should be important for a better understanding of how communities respond to environmental change.

Keywords: beta diversity, community assembly, environmental filtering, latitude, stochasticity

**Introduction**

Biological communities are structured by deterministic (i.e., abiotic environmental filtering and biotic interactions) and stochastic (i.e., random dispersal and ecological drift) factors (Chase and Myers 2011), often referred to as community assembly processes (Chase and Leibold 2003). Even if some studies strongly emphasize deterministic over stochastic factors in shaping communities (Dini-Andreote et al. 2015, Aguilar and Sommaruga 2020), mounting evidence suggests that deterministic and stochastic factors are typically not mutually exclusive but rather act in concert to shape species composition and diversity (Caruso et al. 2011, Daniel et al. 2019, He et al. 2021).

Regarding deterministic factors, abiotic environment may have opposing effects on communities as environment may induce either homogeneous selection or heterogeneous selection on communities, depending on the responses of individual species to environmental gradients and the level of environmental gradients covered (Chase et al. 2011, Ning et al. 2020). A typical scenario is that such abiotic filtering results in distance decay of community similarity as environment tends to be more or less homogeneous at small spatial scales leading to similar communities at neighboring sites (i.e., pairwise communities show high similarity due to homogeneous selection), whereas environment is more heterogeneous at broader spatial scales leading to dissimilar communities between sites located at longer distances apart (i.e., pairwise communities show low similarity due to heterogeneous selection) (Soininen et al. 2007, Graco-Roza et al. 2022). Note that stochastic factors can also explain the distance decay relationship as sites at longer distances apart typically have more dissimilar communities due to dispersal limitation while among nearby sites communities may be homogenized due to mass effects (Pulliam, 2000; Soininen et al., 2007).

What is currently lacking is a quantitative synthesis of the factors underlying the relative balance between these major community assembly processes. Key question is in which circumstances we expect the deterministic factors to dominate over stochastic factors and vice versa. One of the main factors potentially affecting the relative balance of the community assembly processes is study extent. There is some consensus that deterministic factors act more locally rather than regionally since biotic interactions typically occur at fine spatial scales and any small-scale heterogeneity in environment filters species efficiently from the species pool (Soininen 2014, Xing and He 2018). However, we emphasize that biotic interactions (Araújo and Luoto 2007, Gotelli et al. 2010) and especially climatic gradients (Bellard et al. 2012) structure communities at larger extents, too. Concerning stochastic factors, one could envisage that they shape communities perhaps more strongly at larger extents where dispersal limitation is often a predominant factor (Leibold and Chase 2018). Thus, we expect the balance to shift from deterministic to stochastic factors when the study extent increases (H1). Furthermore, we expect homogeneous selection to dominate at smaller scales while heterogeneous selection becomes stronger at broader scales due to increased environmental heterogeneity.

One of the most intriguing factors potentially affecting the relative balance of assembly processes is latitude. On one hand, studies suggest that biotic interactions should be stronger in tropics than at high latitudes due to higher herbivory, insect predation and mutualism (Schemske et al. 2009, Roslin et al. 2017,). On the other hand, environmental filtering could increase with latitude since large-scale environmental heterogeneity becomes larger at high latitudes (Terborgh 1973) and cold climate represents a specific filter for biota (Kirk et al. 2022). Therefore, we hypothesize the deterministic factors to either decrease or increase with latitude depending on whether biotic or abiotic filtering dominates while stochastic factors may not show a significant latitudinal trend (H2).

Community assembly processes may vary also among major realms. We hypothesize that as marine ecosystems are large and relatively stable and individuals may disperse relatively easily among sites (Clarke 1992), marine communities show more deterministic community assembly than freshwater or terrestrial systems while they show less stochasticity (H3). Note that freshwaters especially are small and prone to frequent physical disturbances (Allan 1995) and therefore represents perhaps the most stochastic ecosystems in which dispersal is also often weaker especially in less-connected lake ecosystems (Soininen 2014).

Here, we test these hypotheses using 149 datasets collected from recent ecological literature. With a null model approach (Chase et al. 2011) and Bayesian meta-regression, we study if communities show deterministic or stochastic community assembly and how deterministic factors including homogeneous selection (i.e. communities show low between-site dissimilarities) and heterogeneous selection (i.e. communities show high between-site dissimilarities) vary with underlying factors introduced above.

**Material and methods**

We obtained 149 datasets, each representing a metacommunity, including raw data of species occurrence and spatial coordinates of the study sites. These datasets were compiled from a publicly available database designed to test hypotheses related to spatial patterns (Graco-Roza et al. 2022) and encompass 17 major biotic groups from terrestrial (n = 87), freshwater (n = 43), and marine (n = 21) environments, covering over 17,000 study sites worldwide (Figure 1). Taxa in most datasets were identified to species or morphospecies level, while in some cases, genus-level data were used if taxonomic knowledge did not permit distinguishing individual species.

*Data analyses*

We used Raup-Crick values (Chase et al. 2011) to measure species turnover in each metacommunity by comparing all pairs of communities. Raup-Crick values range from -1 to 1. A value of 1 indicates that two communities are disproportionately dissimilar, while a value of -1 indicates that they are disproportionately similar. Values approaching zero indicate that the similarity in community composition is not different than expected by chance (Chase et al. 2011). Higher positive values of Raup-Crick are often associated with deterministic niche differentiation resulting from either biotic interactions or environmental differences. Lower and negative values of Raup-Crick may suggest that the environment is filtering species and selecting for similar communities. In our datasets, which typically covered intermediate or broad spatial scales, the probability of interaction between species from two communities is relatively small. Therefore, our results are more likely related to deterministic environmental effects rather than biotic interactions.

We also drew a few predictors to study the effects of spatial distances, geographic, biological and ecosystem features both within and across datasets. At within-dataset level, we estimated the (i) spatial distances and (ii) mean latitudes between all pairs of sites to study the effect of the geographic scales over which communities are similar or dissimilar. At cross-dataset level, we included (iii) the realm in which the communities were sampled (i.e., freshwater, terrestrial or marine), (iv) the total species pool in the data, (v) the dispersal mode (i.e., active or passive), and (vi) the average body size of the organisms (drawn from the literature as log10 body mass).

We note that while dispersal mode and body size are not included in our hypotheses, these have been useful explaining variables in some macroecological meta-analyses (Soininen et al. 2007, Soininen and Luoto 2014). For example, organism body size could also influence the relative balance as smaller-sized species are expected to show more stochastic species distributions than larger species (Farjalla et al. 2012, Soininen et al. 2013). This is because smaller species show random passive dispersal and exhibit stochastic births and deaths following disturbances.

We used a linear mixed model following a Bayesian framework as a meta-regression approach to synthesize the effects of our predictors and test our hypothesis. Bayesian models are a more flexible alternative to traditional frequentist models, allowing us to specify prior distributions for our model parameters, incorporate previous knowledge, and obtain posterior distributions for our parameters of interest (Gelman et al. 2013).

Conventional statistical methods would result in substantial pseudo-replication when analyzing compositional similarity for all pairwise comparisons of sites. Specifically, with sites, there are possible comparisons, but only comparisons are independent, corresponding to the diagonal and off-diagonal elements of the matrix of all possible comparisons (Longacre et al. 2005). To avoid pseudo-replication, we focused on the first off-diagonal elements of a randomized site-by-site matrix for each study, as the diagonal elements represent self-comparisons (Newbold et al. 2016). By selecting data from the first off-diagonal of the randomized matrix, which represents unique pairwise comparisons, we ensure that no site is compared to others more than once. We then applied this selection to a set of 100 bootstrapped site-by-site matrices, yielding over 1 million observations clustered by bootstrap ID. However, running a model with this many comparisons in Bayesian analysis would require high computational effort and memory consumption. Therefore, we included a random subset of 1% of the observations in our model (N = 9100), achieving a balance between accuracy and computational efficiency.

While Raup-crick values are continuous, defining a threshold to which homogeneous or heterogeneous selection processes prevails over stochastic assembly can help the interpretation of patterns and the generalization of results (Teittinen et al. 2022). We here assumed that homogeneous selection prevailed when Raup-crick values < -0.9, heterogeneous selection was considered prevalent if Raup-crick > 0.9, while stochastic assembly prevailed if otherwise. By doing it so, we constructed a categorical response variable with three levels which was used to build a model following a categorical distribution. Our model included spatial distances and the mean latitude between sites, the realm, organism body size, dispersal mode, species pool size of the metacommunity as predictors.

In this model, is the assembly process observed for a pair of observations , which follows a categorical distribution with probabilities . The logit of is modeled as a linear combination of fixed effects () and random effects (), where and are design matrices for the fixed and random effects, respectively. The matrix included the mean latitude between sites, the realm, organism body size, dispersal mode and species pool size of the metacommunity as predictors. The matrix include the varying slope at the spatial distance by dataset nested by bootstrap ID within dataset, and a varying intercept for each bootstrap ID within each dataset. We incorporated a random slope term to allow the relationship between spatial distances and the probability of observing a given assembly process to vary across datasets, while constraining the variation to the ID of the bootstrap procedure. Specifically, the random effect captures the deviation of study from the overall average effect, modeled as a linear combination of the fixed effects (). The random effects are assumed to follow a multivariate normal distribution () with mean 0 and covariance matrix , which is modeled as the product of the lower triangular matrix and its transpose. The matrix is obtained from the QR decomposition of X, with the argument "qrcp = TRUE" specifying that it should be column-pivoted for numerical stability. The prior distribution for the fixed effects () is multivariate normal with mean 0 and covariance matrix , which is a diagonal matrix with diagonal elements representing the variances of each coefficient. The hyperparameters are modeled as a half Student-t distribution with 3 degrees of freedom, a location parameter of 0, and a scale parameter of 10.

All models were conducted in the programming language Stan using the function brm within ‘brms’ version 2.17.0 (Bürkner and Brms 2017). Stan implements Hamiltonian Monte Carlo and its extension, the No-U-Turn Sampler. These algorithms converge much faster than other Markov chain Monte Carlo algorithms, especially for high-dimensional models (Kruschke 2015). These algorithms converge much faster than other Markov chain Monte Carlo algorithms, especially for high-dimensional models. Our models ran with four Markov chains of 2500 iterations each, discharging the first 1000 iterations per chain and resulting in 7500 posterior samples default priors. We checked model convergence by visually inspecting parameter traces, ensuring a scale reduction factor (R) below ≤ 1 and an effective sampling size (Neff) of minimally 10% of iterations. The posterior samples were summarized based on Bayesian point estimate (median), standard error (median absolute error), and the posterior certainty of the high-density interval (HDI). The HDI is a special type of credible interval that contains the required mass such that all points within the interval have a higher probability density than points outside the interval (Kelter 2020). Bayesian models offer the advantage of estimating different degrees of certainty around parameters (Kelter 2020) which can be interpreted as the degree of evidence for the relationship. We inferred strong evidence for estimates larger or smaller than zero within at least 95% of the HDI. Likewise, we inferred moderate or weak evidence for estimates that differed from zero within 89 or 80% of HDI, respectively, and interpreted no evidence for relationships outside the 80% HDI.

**Results**

Stochastic factors increased with spatial distances as we hypothesized (H1). Homogeneous selection was strongest at small spatial distances between study sites while heterogeneous selection became evident only at very long spatial distances (Figure 2, Table 1), agreeing thus with H1. Overall, heterogeneous selection was not a common inferred process in the datasets.

Homogeneous selection strongly increased with latitude while heterogeneous selection increased only very gently with latitude agreeing partly with H2 (Figure 2). In contrast, stochastic factors decreased with latitude being strongest in the tropics, disagreeing thus with the hypothesis of no latitudinal gradient in stochasticity.

Among realms, marine ecosystems had the highest level of homogeneous selection, followed by freshwater and terrestrial systems, agreeing with H3 (Figure 2). However, heterogeneous selection did not vary among the realms. Stochastic factors were strongest in terrestrial systems, followed by freshwaters while these were notably lower in marine systems as we hypothesized.

Overall, community assembly processes did not vary markedly with body size. Heterogeneous selection increased slightly with larger body size while homogeneous selection slightly decreased with size (Table 1). Stochastic factors slightly increased with body size. Among the dispersal modes, passively dispersed taxa had slightly lower homogeneous selection than actively dispersed taxa (Table 1). In contrast, pattern was opposite for stochastic factors for which passively dispersed taxa had slightly higher values. For gamma diversity, both homogeneous and heterogeneous selection increased with gamma diversity while stochastic factors somewhat surprisingly decreased with gamma diversity.

**Discussion**

To quantify the degree to which communities assemble through deterministic and stochastic factors is paramount for a better understanding of how biodiversity varies in space and time (Aguilar and Sommaruga 2020). However, to date, a quantitative synthesis over the community assembly in different ecosystems, taxa and geographical gradients have been lacking. Here, we showed that study scale, latitude and ecosystem type jointly drive the balance between deterministic and stochastic community assembly. Overall, one of the key findings was that heterogeneous selection is a rare process compared with homogeneous selection or stochastic community assembly suggesting that many species may respond relatively similarly to major environmental gradients, resembling thus more Clementsian community pattern than Gleasonian (McIntosh 1995) or simply because environmental gradients are not that strong. In general, it is probable that one of the underlying causes for such homogeneous selection is constantly increasing human land use that homogenizes biotic communities spatially in terrestrial and freshwater systems by decreasing the environmental heterogeneity (Budnick et al. 2019, Rodrigues et al. 2013). In marine ecosystems in turn, weak dispersal barriers increase community homogenization spatially. Next, we will discuss our major findings in detail.

*Scale*

We found some scale-dependence in community assembly processes. The scale-dependence agrees with Soininen (2014) who showed through a meta-analysis that environmental filtering decreases with increasing study extent. Note that in the current study, we were able to separate homogeneous and heterogeneous selection by environment and found that they varied meaningful way as we hypothesized; environment selected species homogeneously relatively more often at small scales while heterogeneously relatively more often at larger scales where environment should become more different. Compared with the study by Soininen (2014), which only included data sets at metacommunity scales (maximum up to hundreds of kilometers), the variability in scales was notably larger in the present study since we also covered distances up to 10 000 km. It is thus likely that the current data incorporated a complete mixture of environmental selection, dispersal limitation (acting probably mostly at broader scales) as well as dispersal surplus (i.e. mass effect, acting probably mostly at smaller scales).

One of our notable findings was that homogeneous selection overruled heterogeneous selection at all scales and heterogeneous selection only became evident at very long distances. The reason why heterogeneous selection was low remains speculative at present, but some factors can be put forward. First, patterns might relate to some underlying environmental variable that do not vary systematically with scale yet has importance on biota. This finding agrees with Soininen`s (2023) study about boreal stream diatoms where homogeneous selection by water chemistry dominated at all study scales ranging from few meters to hundreds of kilometers. Allen et al. (2020) also found that homogeneous selection by environment dominated the community assembly across 2000 km longitudinal transect in planktonic microbes in South Pacific. Second, in some data sets, human land use and species introductions may affect the degree to which communities are spatially homogenised. For example, intensive human land use (forestry, agriculture, urbanization) homogenises taxa in a wide range of ecosystems (Rodrigues et al. 2013, Budnick et al. 2019). Moreover, the number of species invasions often influences the degree of biotic homogenisation at least for fish, bird, and plant communities (Olden et al. 2004). Overall, we think that the patterns reflect the fact that even if environmental variation is notable at multiple scales in nature, it is typically not strong enough to differentiate communities. For example, physico-chemical variables may vary both at smaller and broader scales in different natural systems (Taka et al. 2022). Moreover, climatic factors vary not only at broad spatial scales but also at microscales revealed in number of microclimatic studies (De Frenne et al. 2021). Nonetheless, still species respond to environment in surprisingly similar fashion.

Regarding stochastic factors, we note that dispersal comprises both dispersal limitation and dispersal surplus (i.e., mass effect), the latter of which refers to a situation where populations can be temporally present at sub-optimal sites due to the constant flux of immigrants from nearby sites (Pulliam 2000). Given that notable part of our data cover long distances and dispersal limitation is more likely to take place at broader spatial scales, it is not surprising that stochastic factors increased with spatial distance. Cottenie (2005) also found in his comprehensive synthesis about metacommunity types that neutrality in communities increased with study scale due to stronger dispersal limitation at broader scales.

*Latitude*

Community assembly processes did show some latitudinal trends as we expected. Notably, homogeneous and heterogeneous selection representing deterministic factors were strongest at high latitudes while stochastic factors were strongest in the tropics. We think that overall stronger environmental selection at higher latitudes mirrors the fact that environmental heterogeneity tends to become stronger towards the poles (Terborgh 1973) and harsh environment selects species efficiently from the regional pool. In contrast, as was seen in present results, communities in low latitudes are relatively more stochastic perhaps due to more benign environmental conditions, higher complexity, and diversity. This finding concurs partly with Graco-Roza et al. (2022) global analysis in which distance decay along environment was weakest at mid-latitudes at temperate region and stronger in the tropics or near the poles suggesting that environmental distance decay did not show simple latitudinal clines across wide range of taxa. In contrast, Nishizawa et al. (2022) found the opposite exclusively for plants since they documented the strongest environmental filtering at mid latitudes (20-30˚) in their meta-analysis. The present result is in line with Qiao et al. (2015) who found that environmental filtering increased toward higher latitudes due to harsh environment in Chinese forests and with Kirk et al. (2022) who observed that stress related to lower temperatures filter species efficiently from the species pool in freshwater fish. In sum, latitudinal gradients in community assembly rules need still further studying yet there seems some evidence showing that environmental filtering and especially homogeneous selection increases with latitude.

*Realm*

In general, marine ecosystems are environmentally more homogeneous than terrestrial or freshwater ecosystems, at least in the open ocean (Clarke 1992), and typically show weaker dispersal barriers than terrestrial or freshwater ecosystems (Bierne et al. 2003, Cornell and Harrison 2014). Therefore, it was not surprising that marine systems had the highest degree of homogeneous selection and the lowest stochasticity indicating low dispersal limitation. These findings agree with Soininen et al. (2007) who found that marine systems have overall significantly lower beta diversity than other systems. We also emphasize that as marine systems are large, they probably face less physical disturbances (e.g. hydrological disturbance, land use change, fire) than other systems, thus making communities less stochastic and more stable (Korhonen et al. 2010).

*Species traits*

Community assembly processes did not vary notably with body size, thus disagreeing with studies from freshwater systems (Farjalla et al. 2012, Soininen et al. 2013) and more generally across ecosystems and taxa (Soininen and Luoto 2014). The present finding also disagrees with Luan et al. (2020) who found that stochastic processes were relatively more common in the smallest soil microorganisms while deterministic processes were more common in larger soil eukaryotes such as nematodes or fungi. However, Zinger et al. (2019) concluded the opposite and showed that soil mesofauna have more stochastic community assembly than microbes. We did not find clear differences in community assembly processes between actively and passively dispersing taxa either. In this sense, our results disagree from a study by Cottenie (2005) who found that dispersal mode was an important determinant of metacommunity type across taxa. In his analysis, passively dispersed taxa such as freshwater plankton and terrestrial plants tracked well environmental heterogeneity and active dispersers were strongly related to environmental variation in lakes and marine systems while having a low degree of stochasticity. One major difference between the present study and Cottenie (2005) is that our data cover much larger spatial extents than the former analysis, which was concentrated on regional metacommunity datasets; such differences in the ranges of scales probably has important influence on the outcome too.

*Conclusions*

To conclude, we found that overall, communities showed more homogeneous rather than heterogeneous selection across biotic taxa and ecosystems. Nonetheless, results had an interesting structure since environment selected species homogeneously relatively more often at small scales while heterogeneously relatively more often at larger scales. Stochasticity also showed scale-dependence as stochastic assembly increased with study scale as we hypothesized. We further found that latitude and ecosystem type affected the balance between assembly processes as deterministic factors were strongest at high latitudes while stochastic factors were strongest in the tropics. Our results reinforce the idea that marine systems are somewhat unique systems in their biogeographical patterns as they had the highest degree of homogeneous selection and the lowest stochasticity indicating lower environmental heterogeneity and less dispersal limitation. Our analysis provides the first comprehensive study of community assembly processes across taxa and ecosystems and should be important for a better understanding of how communities respond to ongoing environmental change.

Conflict of interest statement: Authors declare no conflict of interest.

Data archiving statement: Data are not yet published but once published, data and code will be freely available in Zenodo (zenodo.org). The data DOI will be included at the end of the article once published.

**References**

Aguilar, P. and Sommaruga, R. 2020. The balance between deterministic and stochastic processes in structuring lake bacterioplankton community over time. – Mol. Ecol. 29: 3117-3130.

Allan, J.D. 1995. Stream Ecology. – Kluwer Academic Publishers.

Allen, R., Hoffmann, L.J., Larcombe, M.J., Louisson, Z. and Summerfield, T.C. 2020. Homogeneous environmental selection dominates microbial community assembly in the oligotrophic South Pacific Gyre. – Mol. Ecol. 29: 4680-4691.

Araújo, M.B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. ─ Global Ecol. Biogeogr. 16: 743-753.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. and Courchamp. F. 2012. Impacts of climate change on the future of biodiversity. ─ Ecol. Lett. 15: 365-377.

Bierne, N., Bonhomme, F. and David P. 2003. Habitat preference and the marine-speciation paradox*.* ─Proc. Royal Soc. B. 270: 1399–1406.

Budnick, W.R., Leboucher, T., Beilard, J., Soininen, J., Lavoie, I., Pound, K. et al. 2019. Local and regional drivers of taxonomic homogenisation in stream communities along a land use gradient. ─ Global Ecol. Biogeogr. 28: 1597-1609.

Bürkner, P. and Brms, C. 2017. An R package for Bayesian multilevel models using Stan. ─ J. Stat. Softw.80: 1–28.

Caruso, T., Chan, Y., Lacap, D.C., Lau, M.C.Y., McKay, C.P. and Pointing, S.B. 2011. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. ─ ISME J. 5: 1406-1413.

Chase, J.M. and Leibold, M.A. 2003. Ecological Niches. ─ The University of Chicago Press.

Chase, J.M. and Myers, J.A. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. ─ Phil. Trans. Royal Soc. B 366: 2351-2363.

Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M. and Inouye, B.D. 2011. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. ─ Ecosphere 2: 24.

Clarke, A. 1992. Is there a latitudinal diversity cline in the sea? ─ Trends Ecol. Evol. 7: 286–287.

Cornell, H.V. and Harrison, S.P. 2014. What are species pools and when are they important? ─ Annu. Rev. Ecol. Evol. Syst. 45: 45–67.

Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. ─ Ecol. Lett. 8: 1175–1182.

Daniel, J., Gleason, J.E., Cottenie, K. and Rooney, R.C. 2019. Stochastic and deterministic processes drive wetland community assembly across a gradient of environmental filtering. ─ Oikos 128: 1158-1169.

De Frenne P., J. Lenoir, M. Luoto, B. Scheffers, F. Zellweger, J. Aalto et al. 2021. Forest microclimates and climate change: importance, drivers and future research agenda. ─ Global Change Biol. 27: 2279-2297.

Dini-Andreote, F., J.C. Stegen, J.D. van Elsas, and J.F. Salles. 2015. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. ─ PNAS 112: E1326-E1332.

Farjalla, V. F., D. S. Srivastava, N. A. C. Marino, F. D. Azevedo, V. Dib, P. M. Lopes et al. 2012. Ecological determinism increases with organism size. ─ Ecology 93: 1752–1759.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B Rubin. 2013. Bayesian data analysis. ─ CRC press.

Gotelli, N.J., G.R. Graves, and C. Rahbek. 2010. Macroecological signals of species interactions in the Danish avifauna. ─ PNAS 107: 5030-5035.

Graco-Roza, C., S. Aarnio, N. Abrego, A. T. R. Acosta, J. Alahuhta, J. Altman et al. 2022. Distance decay 2.0 – A global synthesis of taxonomic and functional turnover in ecological communities. ─ Global Ecol. Biogeogr. 31: 1399– 1421.

He, X., C. Brown, and & L. Lin. 2021. Relative importance of deterministic and stochastic processes for beta diversity of bird assemblages in Yunnan, China. ─ Ecosphere 12: e03545.

Kirk, M. A., F. J. Rahel, and D. C. Laughlin. 2022. Environmental filters of freshwater fish community assembly along elevation and latitudinal gradients. ─ Global Ecol. Biogeogr. 31: 470– 485.

Korhonen, J.J., J. Soininen, and H. Hillebrand. 2010. A quantitative analysis of temporal turnover in aquatic species assemblages across ecosystems. ─ Ecology 91: 508-517.

Kruschke, J. K. 2015. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. ─ Academic Press.

Leibold, M.A., and J.M. Chase. 2018. Metacommunity Ecology. ─ Princeton University Press.

Luan, L., Y. Jiang, M. Cheng, F. Dini-Andreote, Y. Sui, Q. Xu et al. 2020. Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. ─ Nat. Comm. 11: 6406.

Longacre, A., L. R. Scott, and J.S. Levine. 2005. Linear independence of pairwise comparisons of DNA microarray data. ─ J. Bioinform. Comp. Biol. 3: 1243-1262.

McIntosh, R.P. 1995. H.A. Gleason`s “individualistic concept” and theory of animal communities: a continuing controversy. ─ Biol. Rev. 70: 317-357.

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Gray, C.L., Scharlemann, J.P.W., Börger, L., Phillips, H.R.P., Sheil, D., Lysenko, I. and Purvis, A. 2016. Global patterns of terrestrial assemblage turnover within and among land uses. ─ Ecography 39: 1151-1163.

Ning, D. L., M.T. Yuan, L.W. Wu, Y. Zhang, X. Guo, X. S. Zhou et al. 2020. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. ─ Nat. Comm. 11: 4717.

Nishizawa, K., N. Shinohara, M.W. Cadotte, and A.S. Mori. 2022. The latitudinal gradient in plant community assembly processes: A meta-analysis. ─ Ecol. Lett. 25: 1711– 1724.

Olden, J.D., N.L. Poff, M.R. Douglas, M.E. Douglas, and K.D. Fausch. 2004.

Ecological and evolutionary consequences of biotic homogenisation. ─ Trends Ecol. Evol. 19: 18-24.

Pulliam, H.R. 2000. On the relationship between niche and distribution. ─ Ecol. Lett. 3: 349-361.

Qiao, X., F. Jabot, Z. Tang, M. Jiang, and J. Fang. 2015. A latitudinal gradient in tree community assembly processes evidenced in Chinese forests. ─ Global Ecol. Biogeogr. 24: 314-323.

Rodrigues, J.L., V. H. Pellizari, R. Mueller, K. Baek, C. Jesus Eda, F.S. Paula et al. 2013. Conversion of the Amazon rainforest to agriculture results in biotic homogenisation of soil bacterial communities. ─ PNAS 110: 988-993.

Roslin, T. V., B. Hardwick, V. Novotny, W.K. Petry, N.R. Andrew, A. Asmus et al. 2017. Higher Predation Risk for Insect Prey at Low Latitudes and Elevations. ─ Science 356: 742–744.

Schemske, D. W., G. G. Mittelbach, H. V. Cornell, J. M. Sobel & K. Roy. 2009. Is there a latitudinal gradient in the importance of biotic interactions? ─ Annu. Rev. Ecol. Evol. Syst. 40: 245-269.

Soininen, J. 2014. A quantitative analysis of species sorting across organisms and ecosystems. ─ Ecology 95: 3284-3292.

Soininen, J. 2023. Are diatom community assembly processes scale invariant in streams? ─ Freshwater Biol. 68: 502-508.

Soininen, J., and M. Luoto. 2014. Predictability in species distributions: a global analysis across organisms and ecosystems. ─ Global Ecol. Biogeogr. 23: 1264-1274.

Soininen, J., R. McDonald, and H. Hillebrand. 2007. The distance decay of similarity in ecological communities. ─ Ecography 30: 3-12.

Soininen, J., J. Korhonen, and M. Luoto. 2013. Stochastic species distributions are driven by organism size. ─ Ecology 94: 660-670.

Taka, M., N. Sillanpää, T. Niemi, L. Warsta, T. Kokkonen, and H. Setälä. 2022. Heavy metals from heavy land use? Spatio-temporal patterns of urban runoff metal loads. ─ STOTEN 817: 152855.

Teittinen, A., J. Soininen, and L. Virta. 2022. Studying biodiversity-ecosystem function relationships in experimental microcosms among islands. ─ Ecology 103: e3664.

Terborgh, J. 1973. On the notion of favorableness in plant ecology. ─ The Am. Nat. 107: 481–501.

Zinger, L., P. Taberlet, H. Schimann, A. Bonin, F. Boyer, M. De Barba et al. 2019. Body size determines soil community assembly in a tropical forest. ─ Mol. Ecol. 28: 528–543.

Xing, D., and F. He. 2019. Environmental ﬁltering explains a U-shape latitudinal pattern in regional β-deviation for eastern North American trees. ─ Ecol. Lett. 22: 284-291.

**Table 1:** Estimates and 95% credible intervals (CI) for predictor effects on heterogeneous and homogeneous selection. This table presents estimates and 95% credible intervals (CI) for predictor effects on selection. Predictor variables include spatial distance, mean latitude, realm (with freshwater as the reference), body size, dispersal mode (with active dispersal as the reference), and species pool size. Stochasticity, used as a reference in the model, is not included. Estimates are derived from a Bayesian regression model using a categorical distributional family. The model was estimated using 5 chains and 2500 iterations (including warm-up) through Markov Chain Monte Carlo (MCMC) sampling.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Heterogeneous selection** | | **Homogeneous selection** | |
|  | **Estimate [std.error]** | **95% CI** | **Estimate [std.error]** | **95% CI** |
| Intercept | **-9.23[0.54]** | **-10.31, -8.2** | **-4.61[0.26]** | **-5.13, -4.09** |
| Spatial distance [km] | **0.14[0.06]** | **0.03, 0.25** | -0.22[0.02] | -0.26, -0.18 |
| Mean latitude | 0.01[0.00] | 0.00, 0.02 | **0.01[0.00]** | **0.01, 0.01** |
| Realm (ref.: Freshwater) |  |  |  |  |
| *Marine* | **0.71[0.30]** | **0.13, 1.29** | **1.20[0.14]** | **0.92, 1.48** |
| *Terrestrial* | 0.06[0.19] | -0.30, 0.42 | -0.18[0.10] | -0.37, 0.00 |
| Body size | 0.16[0.02] | 0.11, 0.21 | **-0.04[0.01]** | **-0.07, -0.02** |
| Dispersal mode (ref.: Active) | |  |  |  |
| *Passive* | **0.68[0.14]** | **0.42, 0.95** | **-0.32[0.07]** | **-0.47, -0.18** |
| Species Pool | **0.94[0.07]** | **0.8, 1.09** | **0.73[0.04]** | **0.65, 0.81** |

**Figure captions**

**Figure 1 | Global distribution of datasets and key characteristics.** **(a)** Map displaying the global distribution of sites within datasets, represented by filled circles. Each dot represents a site within a dataset. The datasets comprise over 14,000 local communities from 147 different datasets. **(b)** Number of datasets per realm: Freshwaters (blue), Marine (pink), and Terrestrial (green). **(c)** Bar chart showing the distribution of datasets by dispersal mode. **(d)** Histogram of spatial extent on a logarithmic scale. **(e)** Histogram of species pool size on a logarithmic scale. The figure provides an overview of the global dataset distribution, highlighting realms, dispersal modes, spatial extent, and species pool sizes across datasets.

**Figure 2 | Probability distributions of ecological processes. (a)** Probability distribution of the effect of spatial distance on the prevalence of ecological processes, including heterogeneous selection (orange), homogeneous selection (brown), and stochastic processes (purple). The x-axis represents the spatial distance between pairs of sites within a dataset, while the y-axis represents the posterior predicted probability. **(b)** Probability distribution of the effect of mean latitude on the same ecological processes, indicated by the corresponding color scheme. The x-axis represents the mean latitude of pairs of sites within a dataset. **(c)** Probability distribution of the effect of realm on the ecological processes, represented by the color scheme. The x-axis represents the realm, which is a single value for each dataset. The shaded regions represent the 80%, 90%, and 95% credible intervals. The Bayesian regression model employed a categorical distributional family to capture the deterministic or stochastic nature of species assembly. The model was estimated using Markov Chain Monte Carlo (MCMC) sampling, with 5 chains and 2500 iterations (including warm-up). The posterior inference was based on 1000 draws.

