ECOGRAPHY

Research

Joint analysis of species and genetic variation to quantify the role of dispersal and environmental constraints in community turnover

Andrés Baselga, Carola Gómez-Rodríguez, Miguel B. Araújo, Adrián Castro-Insua, Miguel Arenas, David Posada and Alfried P. Vogler

A. Baselga (https://orcid.org/0000-0001-7914-7109) ✉ (andres.baselga@usc.es) and A. Castro-Insua, CRETUS, Dept of Zoology, Genetics and Physical Anthropology, Univ. de Santiago de Compostela, Santiago de Compostela, Spain. – C. Gómez-Rodríguez (https://orcid.org/0000-0002-2019-7176), CRETUS, Dept of Functional Biology (Area of Ecology), Univ. de Santiago de Compostela, Santiago de Compostela, Spain. – M. B. Araújo, Depto de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain and Cátedra de Biodiversidade 'Rui Nabeiro', Inst. MED, Univ. de Évora, Évora, Portugal. – M. Arenas and D. Posada, CINBIO and Dept of Biochemistry, Genetics and Immunology, Univ. of Vigo, Vigo, Spain. – A. P. Vogler, Dept of Life Sciences, Natural History Museum, London, UK and Dept of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, UK.

Ecography
2022: e05808
doi: 10.1111/ecog.05808
Subject Editor: Timothy Keitt
Editor-in-Chief:
Jens-Christian C Svenning
Accepted 29 January 2022

Spatial turnover of biological communities is determined by both dispersal and environmental constraints. However, we lack quantitative predictions about how these factors interact and influence turnover across genealogical scales. In this study, we have implemented a predictive framework based on approximate Bayesian computation (ABC) to quantify the signature of dispersal and environmental constraints in community turnover. First, we simulated the distribution of haplotypes, intra-specific lineages and species in biological communities under different strengths of dispersal and environmental constraints. Our simulations show that spatial turnover rate is invariant across genealogical scales when dispersal limitation determines the species ranges. However, when environmental constraint limits species ranges, spatial turnover rates vary across genealogical scales. These simulations were used in an ABC framework to quantify the role of dispersal and environmental constraints in 16 empirical biological communities sampled from local to continental scales, including several groups of insects (both aquatic and terrestrial), molluscs and bats. In seven datasets, the observed genealogical invariance of spatial turnover, assessed with distance–decay curves, suggests a dispersal-limited scenario. In the remaining datasets, the variance in distance–decay curves across genealogical scales was best explained by various combinations of dispersal and environmental constraints. Our study illustrates how modelling spatial turnover at multiple genealogical scales (species and intraspecific lineages) provides relevant insights into the relative role of dispersal and environmental constraints in community turnover.

Keywords: beta diversity, community turnover, dispersal limitation, ecological niche, genetic diversity, phylogenetic scaling

NORDIC SOCIETY OIKOS

www.ecography.org

Introduction

Understanding the drivers of variation of biological communities is of major importance both for fundamental and applied research. Species distributional ranges and their collective expression translated in spatial turnover across biological communities are mostly controlled by extrinsic environmental factors and the intrinsic ability of species to expand their distribution ranges. There is a large body of knowledge attempting to statistically discriminate the role of these two major processes (Borcard et al. 1992, Tuomisto et al. 2003, Gilbert and Lechowicz 2004, Peres-Neto et al. 2006, König et al. 2017). However, given that dispersal limitation is an inherently spatially-structured process and environmental factors are also spatially auto-correlated, most studies have concluded that there is a large amount of variance in community composition that cannot be uniquely attributed to either environmental constraint or dispersal limitation (Smith and Lundholm 2010, Tuomisto et al. 2012). To move forward, one needs to go beyond correlative approaches and tackle the causal nature of the spatial structure of community variation. This is particularly important in a context of climate change because, if species are strongly limited by their rate of dispersal (Svenning and Skov 2007), they might not be able to adjust their ranges by swiftly tracking suitable environments as they change (Araújo and Pearson 2005, Lenoir et al. 2020, Taheri et al. 2021). We thus need to develop a predictive framework to unequivocally link different combinations of dispersal and environmental constraints to the biodiversity patterns we observe in the real world.

Towards building such predictive framework, we examine the independent signatures of dispersal and environmental constraints on community turnover at both species and haplotype levels (genetic variants) (Baselga et al. 2015, Gómez-Rodríguez et al. 2019). Uniform processes, such as dispersal limitation, may govern the distribution of both species and haplotypes and this was the tenet of the species-genetic diversity correlation (SGDC) (Vellend 2003, Vellend and Geber 2005). These seminal studies have helped improve understanding of how different causal factors can influence the correlation between species and genetic diversity (Vellend et al. 2014, Vellend 2016). Here, we focus on the spatial turnover of community composition at both the species and haplotype levels. Community turnover between two sites is usually measured at the species level, as the proportion of species that are present at both sites (increasing community similarity) or only one of them (decreasing similarity). The same principle applies to haplotypes: community turnover at the haplotype level is driven by the proportion of haplotypes found at both sites or in just one of them. It follows that understanding the process of haplotype and species range expansion is critical for understanding the underlying processes driving community turnover.

The spatial distribution of haplotypes in neutrally evolving loci are uniquely controlled by dispersal limitation because they are the result of birth-and-death events and stochastic dispersal within the range of the focal species (Slatkin 1985). In contrast, species ranges can be controlled by both 1) dispersal limitation via the intrinsic ability of organisms to expand their ranges, and 2) environmental constraint via ecological niche filtering, with varying relative strengths. We can thus predict that (P1) if dispersal limitation is the dominant driver, the rate of spatial turnover of biological communities will be the same at both haplotype and species levels. In contrast, (P2) if environmental constraint is the dominant driver, community turnover at the species and haplotype levels are expected to differ, because species ranges would be constrained by the limits of their ecological niche, whereas neutral haplotype ranges do not experience such a constraint. In this second scenario, there is a point after which the species would not be able to further expand the range while the haplotypes would continue to do so (within the limits of the species range), given that their ranges are subject to dispersal limitation alone (Fig. 1). In fact, these predictions do not need to be restricted to the dichotomy between haplotype and species levels, since they are just the lower and upper limits of multiple scales in the genealogy of a species. Genealogical scales can be viewed as slices of increasing depth in a coalescent tree, from the most recent (haplotypes, or unique genetic variants), to intermediate (intraspecific lineages composed of several related haplotypes), and to the deepest scale (species). Therefore, assessing how community turnover varies across different genealogical scales is analogous to assessing how macroecological patterns vary across phylogenetic scales (Graham et al. 2018), but phylogenetic scales are nodes above the species level in a phylogenetic tree, while genealogical scales are nodes below the species level in a coalescent tree.

In this study, we provide a predictive framework to quantify the signatures of dispersal and environmental constraints based on the relationship between community similarity and spatial distance (i.e. distance–decay) across genealogical scales. Distance–decay models allow the quantification of community turnover in two parameters, 1) the slope of the decay function, which measures the rate at which community similarity decreases with spatial distance, and 2) the intercept with the y-axis, which measures the initial community similarity (Soininen et al. 2007). These parameters can be compared across regions (Qian et al. 2005, Fitzpatrick et al. 2013, König et al. 2017), biological groups (Soininen et al. 2007, Saito et al. 2015, Gómez-Rodríguez and Baselga 2018) and genealogical scales (Baselga et al. 2013, Gómez-Rodríguez et al. 2019, Arribas et al. 2020). Previous studies reported a striking regularity of distance–decay of community similarity across genealogical scales in various groups of invertebrates (Baselga et al. 2013, 2015, Gómez-Rodríguez et al. 2019). In these studies, 1) community similarity at short distances (intercept) regularly increases with genealogical scale, which we will refer to as *genealogical scaling* of similarity, and 2) community similarity decays with spatial distance at the same rate (slope) at all genealogical scales, which we will refer to as *genealogical invariance* of turnover rate. Genealogical invariance of turnover rate is only expected if the expansion of the spatial ranges of haplotypes, lineages and species occurs

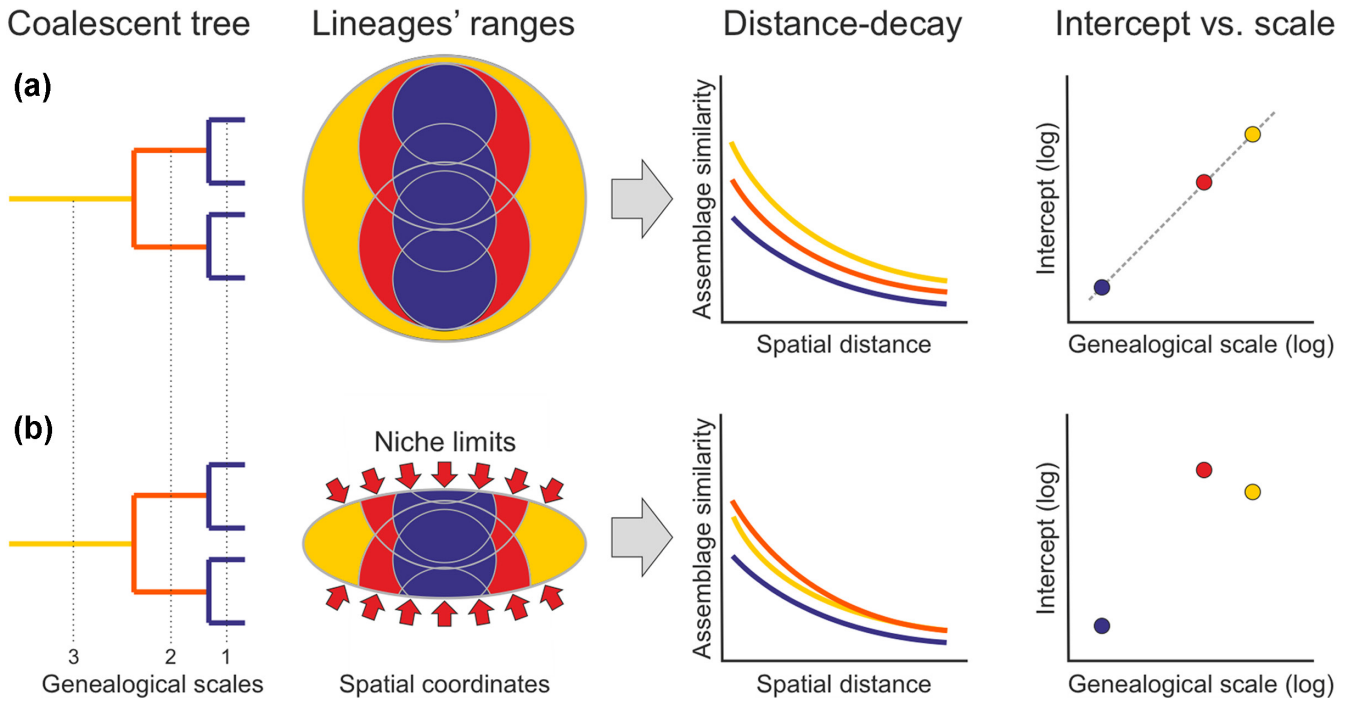

Figure 1. Schematic representation of how distance–decay curves at multiple genealogical scales account for the geometry of lower-level lineage ranges (i.e. haplotype ranges) within higher-level lineage ranges (i.e. species). Note that community similarity measures the proportion of species ranges (or lineage ranges) that overlap between two sites. Only one species and three genealogical scales (1: haplotypes, 2: intraspecific lineages, 3: species) are represented for simplicity. (a) If the species ranges (yellow) are only constrained by dispersal limitation, lineage ranges of all ages (1, 2 and 3, including species) expand their ranges at the same rate as a function of time. Therefore, similarity decays with spatial distance at the same rate at all genealogical scales and the intercepts of the curves shift regularly from level to level. (b) However, if the environment constrains the species range, range expansion will reach a limit at the species level while it will continue expanding in lineages below the species level. Consequently, community similarity decays at different rates at different genealogical scales and intercepts do not shift regularly from level to level.

at the same pace at all genealogical scales (our prediction P1, Fig. 1a). Under P1, the range size of lineages and thus community turnover depend on genealogical scale, also leading to the genealogical scaling of initial similarity. In contrast, according to our prediction P2, distance–decay rates vary across genealogical scales (*genealogical scale-dependence*) when the environment controls the species distributions (Fig. 1b). As a consequence, when the environmental conditions (e.g. climate) constrain species ranges, community similarity is predicted to decay at different rates across genealogical scales.

Our predictive framework is built upon probabilistic models that simulate the spatial distribution of lineages at multiple genealogical scales (from haplotypes to species) under different strengths of dispersal and environmental constraint. For each parameterization of the model, i.e. the specific combinations of dispersal and environmental constraints, we obtain a simulated community in which to measure how community similarity decayed with spatial distance at multiple genealogical scales. Our model thus makes specific predictions of how the shape of the distance–decay curves remains constant or varies across genealogical scales under different combinations of environmental constraint and dispersal limitation. Finally, we used our predictive framework to evaluate the strength of dispersal and environmental constraints in 16 empirical systems. Using DNA sequence data for entire local communities sampled along regional transects, we assessed how distance–decay of similarity varied across genealogical scales across these empirical communities. The observed patterns were compared against those predicted by the simulation model, implementing an approximate Bayesian computation framework (ABC, Beaumont 2010), to assess the relative relevance of dispersal limitation and environmental constraint in the spatial turnover of real-world communities.

Material and methods

Approximate Bayesian computation

We used an approximate Bayesian computation (ABC) framework (Beaumont 2010) to study how dispersal limitation and environmental constraint determine the geometry of spatial ranges of haplotypes while accounting for their genealogical relationships. ABC is a statistical approach that uses computer simulations to estimate the posterior distribution of model parameters without the need for an explicit likelihood function. To be effective, it requires a realistic simulation model and a set of summary statistics that capture

the most relevant information present in the data. In short, we simulated the expansion of haplotypes, intraspecific lineages and species as a function of dispersal ability (*tau*) and ecological niche width (*nsd*), see details below. For each combination of dispersal ability and niche width, we computed the summary statistics that characterise the distance–decay curves and their variation across genealogical scales, and used the distribution of these summary statistics to infer the dispersal ability and the ecological niche width in 16 empirical datasets, using the rejection algorithm implemented in the R package 'abc' (Csilléry et al. 2012). The rejection algorithm compares the observed summary statistics to the simulated summary statistics, and when the difference is sufficiently small, the dispersal ability and niche width values used in that specific simulation are accepted as samples of the posterior distribution of dispersal ability and niche width in the empirical dataset. On the contrary, simulations with summary statistics too far from the observed values are left out of the samples used to estimate the dispersal ability and niche width of the empirical dataset.

Simulation model

In our model, lineage evolution takes place on an environmental landscape consisting of a 300 × 300-cell grid characterised by two climatic variables, temperature and precipitation. We defined three main climatic gradients: *linear* (characterizing latitudinal gradients with tight correlation between climate and geographic distance), *bidirectional* (as produced by a mountain ridge dissecting a landscape) and *concentric* (around a landscape feature such as a singular mountain summit). Temperature and precipitation are mostly independent in the linear climatic gradient (temperature covarying with latitude and precipitation with longitude) but are highly correlated in the bidirectional and concentric climatic gradients.

On top of the virtual landscape, we simulated random coalescent haplotype genealogies for each species with the function *rcoal*() of the R package 'ape' (Paradis et al. 2004). Tree height (i.e. the distance between the root and the tips of the coalescent tree) is set to a number of arbitrary time units. The spatial range of species and haplotypes is simulated allowing lineages to spread as follows (see animation in the Supporting information). The ancestral haplotype of each species is placed at random in one of the central 200 × 200 cells of the landscape. For each species, the ancestral haplotype is allowed to expand its spatial range n times; n being the number of time units between its origin and the first node in the genealogy. Because genealogies are independently simulated for each species, n varies in each case. At the first node, a new haplotype is generated and assigned to a random location within the spatial range of the ancestral haplotype. Again, both haplotypes (ancestral and new) are allowed to expand their ranges a number of times until reaching the next node in the tree, when new haplotypes are again added to the spatial grid in the same manner, until the tips of the tree are reached.

At each step, the spatial expansion of the range of the haplotypes (and therefore of the species range) is probabilistic and takes place as follows. The probability of expanding to a new location in the landscape is modelled as a negative exponential function with mean *tau*. Thus, the probability of reaching a particular location decays exponentially with the distance to the occupied cells, and the slope of the decay depends on *tau*. The species ecological niche is modelled as a normal distribution with mean at the species optimum (defined as the climatic conditions in the species' initial location) and width defined by a single parameter, *nsd*. Hence, the probability of a species expanding to a given location depends on the species ecological niche width and on the climatic difference between the target location and the species optimum. Lineage extinction was not considered because the simulation aims to mimic the spread of spatial ranges of the tips and lineages of a coalescent tree. At the end of each simulation, cell locations are collapsed into regional squares (by combining 10 × 10 local cells), in which the presence/absence of each haplotype, intraspecific lineage and species is recorded. The simulated biological communities are thus the set of lineages present in any given regional square.

Simulation parameterization

We simulated genealogies for 20 species, each one with 10–20 sampled haplotypes. In all cases, we set tree height to 100 arbitrary time units. For each of the three climatic gradients (linear, bidirectional and concentric), we explored combinations of seven niche widths (from very narrow to unlimited) and seven dispersal rates (from very low to unlimited), resulting in a total of 48 scenarios per climatic setting (after ignoring the combination of unlimited niche and unlimited dispersal, which leads to a trivial uniform community in which all haplotypes, lineages and species are present everywhere). Each of these scenarios (144 combinations of *tau* × *nsd* × climatic gradient) was replicated 1000 times, resulting in 144 000 simulations.

ABC summary statistics

For each simulation, we calculated the pairwise community similarity between regional squares using Simpson's similarity (Simpson 1943), an index independent of richness differences (Baselga 2010), at five genealogical scales, using the presence/absence in each square of: 1) haplotypes (tips in the coalescent tree), 2) intraspecific lineages delimited at 25 time units, 3) intraspecific lineages delimited at 50 time units, 4) intraspecific lineages delimited at 75 time units and 5) species. For each of these five genealogical scales, we fitted a negative exponential function describing the decay of community similarity with spatial distance, as $y = a \times e^{-bx}$, where y is similarity at distance x, a initial similarity and b the rate of distance–decay (Gómez-Rodríguez and Baselga 2018). The spatial (Euclidean) distance between regional squares was computed from their spatial coordinates in the virtual landscape. Similarity matrices and distance–decay models (estimating the a and b parameters from the simulated data) were computed using the R package 'betapart' (Baselga and Orme 2012, Baselga et al. 2020). From these five distance–decay curves, we computed two summary statistics measuring

how distance–decay patterns vary across genealogical scales (Baselga et al. 2013). The first summary statistic was the log–log correlation between a and genealogical level (r_a), which measures how regularly similarity shifts across genealogical scales (with low values of r_a indicating genealogical scale-dependence, and high values indicating genealogical scaling of initial similarity). The second summary statistic was the coefficient of variation of b across genealogical scales (cv_b), which quantifies how different the decay rates are across genealogical scales (with low values of cv_b indicating genealogical invariance in decay rate). These two summary statistics account for the variance in distance–decay patterns across genealogical scales (Fig. 1). Additionally, we computed the average r^2 of decay curves across all genealogical scales, *mean.* r^2, which accounts for the goodness of fit in the relationship between pairwise community similarity and spatial distance.

Power of the ABC framework

We assessed the ability of our ABC framework to estimate *tau* and *nsd* with an independent set of 14 400 simulations (144 combinations of *tau* × *nsd* × climatic gradient that were replicated 100 times), under a rejection tolerance of 0.01 (i.e. only the 1% of simulations with closest summary statistics to the observed ones were used to estimate *tau* and *nsd*). We found that our ABC method was able to estimate *tau* and *nsd* with acceptable accuracy (Supporting information).

Analysis of empirical datasets

We implemented our ABC framework to infer the relative importance of dispersal and environmental constraints across 16 publicly available datasets that consisted of fully sequenced (mitochondrial *cox1-5′* or *cox1-3′*) biological communities in multiple sites (mean number of sites = 11, SD = 6). The criteria to select those datasets were a minimum number of sites (5), and a large number of sequences per site (minimum average number of sequences per site > 50). Mitochondrial *cox1-5′* or *cox1-3′* fragments were selected as these are putatively neutral markers (Avise 1994), so the haplotype ranges should not be controlled by environmental constraint. Datasets encompassed different biological groups (mostly insects), world regions and geographical extents: leaf beetles of the Iberian Peninsula (Baselga et al. 2015), water beetles of Europe (Baselga et al. 2013), water beetles of Australia (Hendrich et al. 2010), water beetles of Madagascar (Isambert et al. 2011), darkling beetles of the Aegean Islands (Papadopoulou et al. 2011), dung beetles of Costa Rica (García-López et al. 2013), bats of South East Asia (Francis et al. 2010), bats of Guyana (Clare et al. 2011), terrestrial molluscs of the Iberian Peninsula (Gómez-Rodríguez et al. 2019), dragonflies of New Brunswick (Curry et al. 2012), caddisflies of New Brunswick (Curry et al. 2012), caddisflies of Pennsylvania (Bringloe 2013), ants of Mauritius (Smith and Fisher 2009), butterflies of Romania (Dincă et al. 2011), nymphalid butterflies of Yucatan (Prado et al. 2011) and lepidopterans from Papua (Craft et al. 2010). Sequences were available in public repositories (BOLD database and GenBank, see accession numbers and references in the Supporting information), and geographical coordinates were obtained from the original sources or kindly provided by the authors. When sites were not explicitly defined in the original sources, we grouped the individual data-points into a discrete number of sites using a cluster analysis (Ward algorithm) on the geographical coordinates of individual sequences. The spatial extent of datasets varied between 40 and 4000 km (mean = 1043, SD = 1331).

For each dataset, we delimited lineages and putative species using haplotype networks built with the TCS software (Clement et al. 2000) and the nesting algorithm implemented in ANeCA v.1.2 (Panchal 2007). This algorithm creates a nested design (Templeton et al. 1992) by hierarchically clustering haplotypes into entities corresponding to intermediate lineages defined by n mutational steps between haplotypes (n-step networks). We then fitted distance–decay curves at all genealogical scales using negative exponential models and computed the same summary statistics r_a, and cv_b, as in the simulations. We also computed the average coefficient of determination of decay curves, *mean.*r^2. To estimate dispersal ability (*tau*) and ecological niche width (*nsd*) in each empirical dataset, and thus the relative importance of dispersal versus environmental constraints, we evaluated the observed summary statistics against the distribution of summary statistics across simulated communities considering variation of *tau* and *nsd* parameters (ABC framework). To do this, we first classified each of the empirical datasets into one of three climatic gradients (linear, bidirectional or concentric) based on the observed maps of mean annual temperature and annual precipitation (Supporting information) and used the simulations from the corresponding climate setting with the ABC framework for each specific empirical dataset. The values of *tau* and *nsd* estimated for each empirical dataset can be interpreted in relative terms. Low values indicate strong dispersal limitation and narrow niches, respectively, and high values indicate unlimited dispersal and wide niches, respectively.

Results

Our simulations revealed how the decay of community similarity at different genealogical scales changes under varying strengths of dispersal limitation and environmental constraint (Fig. 2). The proportion of variance explained by negative exponential models was relatively high (mean pseudo-r^2 across genealogical scales > 0.5) under most simulations in the linear climatic gradient, except when both dispersal limitation and environmental constraint were loose, but only under tight dispersal limitation in the bidirectional and concentric climatic gradients (Supporting information). Only simulations involving tight dispersal limitation and no or very loose environmental constraint produced patterns of distance–decay in community turnover characterised by genealogically invariant distance–decay slopes and initial similarity regularly scaling from the species to the haplotype scale (Fig. 2a, e, i). In contrast, when

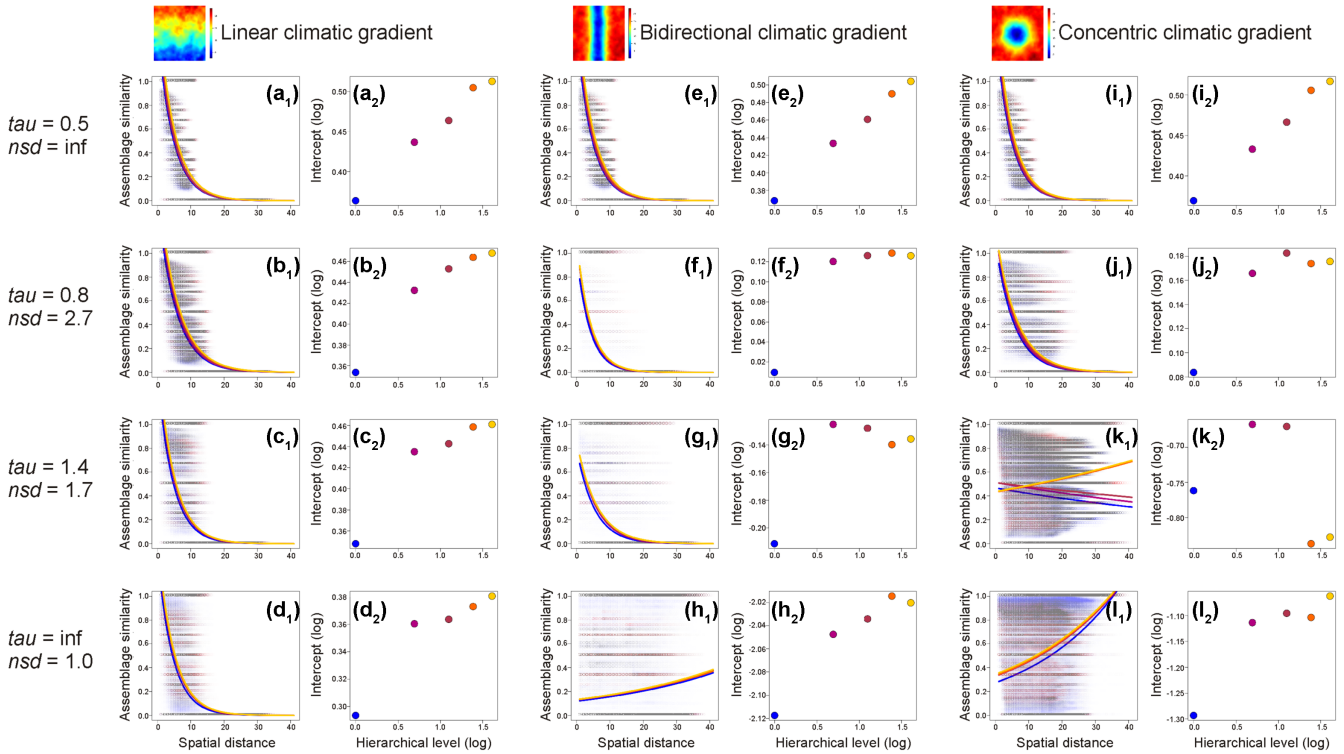

Figure 2. Simulated distance–decay patterns at multiple genealogical scales from haplotypes (blue) to species (yellow) for representative combinations of dispersal limitation (*tau*), niche width (*nsd*) and climate setting (linear, bidirectional or concentric gradient). The three major columns represent the three simulated climatic settings (heat maps): linear gradient (left column, a–d), bidirectional gradient (central column, e–h) and concentric gradient (right column, i–l), and the rows represent different combinations of dispersal limitation and environmental constraint, from pure dispersal limitation (a, e, i) to pure environmental constraint (d, h, l). For each scenario, two plots are included: predicted distance–decay patterns at multiple genealogical scales (subscript = 1), and log–log relationship between initial similarity (intercept) and genealogical scale (subscript = 2). In both columns, colours represent the genealogical scales from haplotypes (blue) to species (yellow). Genealogical invariance of distance–decay curves (a, e, i) is characterized by similar slopes at all scales and a regular shift in the intercepts from level to level. Unlimited dispersal ability or niche width is marked as infinite (inf).

environmental constraint was strong, distance–decay of similarity showed genealogical scale-dependence (Fig. 2b–d, f–h, j–l). Genealogical scale-dependence was particularly marked when climatic conditions changed rapidly over short distances (e.g. mimicking altitudinal gradients) (Fig. 2g–h, k–l). In these cases, at the shortest spatial distance, communities may even be more similar at the haplotype level than at the species level (Fig. 2k). This occurs because the proportion of haplotypes (which are more numerous than species) shared between two locations becomes high if haplotypes disperse widely within the species ranges (which only occupy a small spatial range due to environmental constraint). Altogether, among all the combinations of different strengths of dispersal limitation and environmental constraint, only when dispersal ability was tightly limited and under lack of (or loose) environmental constraint, our simulations resulted in distance–decay curves invariant across genealogical scales (Fig. 3). In turn, distance–decay curves were different across genealogical scales when environmental constraint is effective at limiting species ranges.

When these simulations were used as a benchmark to infer the processes driving empirical patterns, the ABC analyses estimated low values of dispersal ability (*tau*) and high values of niche width (*nsd*) in 11 datasets (Table 1), suggesting the predominance of dispersal limitation in these biological systems. These datasets showed regular genealogical scaling of initial similarity and genealogically invariant distance–decay slopes (Fig. 4a–e, g–l), and in seven of these datasets (leaf beetles of the Iberian Peninsula, water beetles of Europe, Australia and Madagascar, darkling beetles of the Aegean Islands, bats of South East Asia, terrestrial molluscs of the Iberian Peninsula and caddisflies of New Brunswick) the coefficient of determination was relatively high ($mean.r^2 > 0.36$). However, in four of them (water beetles of Madagascar, bats of Guyana, dragonflies of New Brunswick and caddisflies of Pennsylvania) this coefficient was lower ($mean.r^2 < 0.15$), suggesting some uncertainty in these four datasets. Finally, the ABC analysis estimated low values of ecological niche width in the five remaining datasets, pointing to a preponderance of environmental constraint. These datasets exhibited a marked genealogical scale-dependence in distance–decay curves (Fig. 4g, m–p): the dung beetles of Costa Rica, butterflies of Romania, nymphalid butterflies of Yucatan, lepidopterans from Papua, and ants of Mauritius (Table 1).

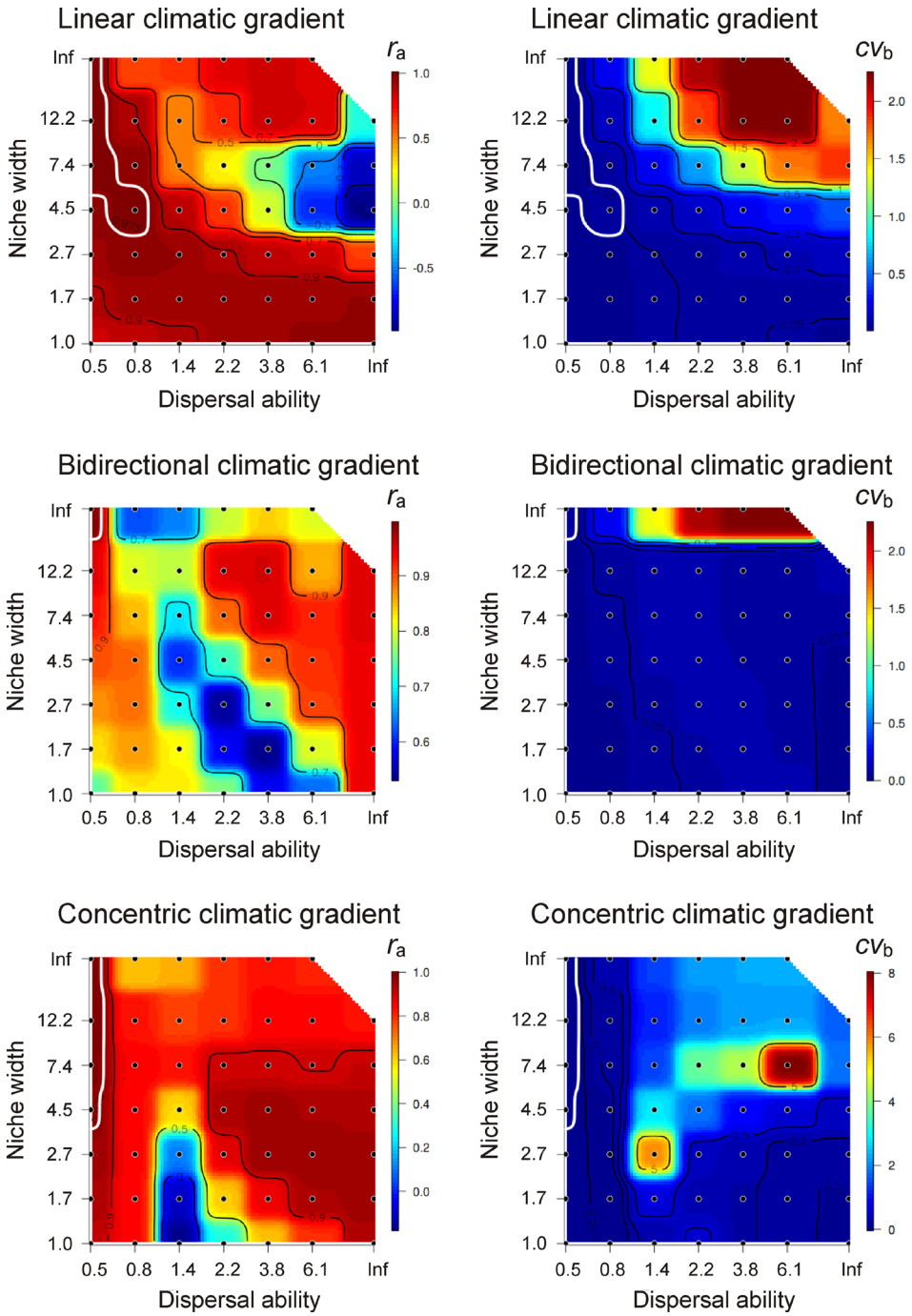

Figure 3. Interpolated parameter space of two summary statistics under different values of dispersal ability, niche width and three climatic settings. Columns show the log–log correlation between distance–decay intercepts and genealogical scale (r_a), and the coefficient of variation of distance–decay slopes (cv_b). Colour represents the average statistic across 1000 simulations, which was interpolated across the whole range of dispersal and niche width combinations. The white line depicts the region defined by $r_a > 0.975$, and $cv_b < 0.2$ (genealogical scaling of intercepts and genealogical invariance of distance–decay slopes).

Finally, to analyse how spatial scale determines the relative importance of dispersal and environmental constraints, we assessed the spatial scale at which genealogical invariance in distance–decay was detected across the empirical systems. The spatial scale of analysis ranged from a few tens of kilometres in the ants of Mauritius to several thousand kilometres in European and Australian water beetles (Supporting information). In the datasets dominated by dispersal limitation, the rate at which similarity decayed with distance showed a logarithmic relationship with the spatial scale of the dataset (Fig. 5, $r^2 = 0.79$, $F_{1,5} = 19.36$, $p = 0.007$). The remaining datasets (i.e. those affected by environmental constraint) fell

Table 1. Summary statistics for the 16 empirical datasets and estimated dispersal ability (*tau*) and niche width (*nsd*) with ABC. For each dataset, the observed summary statistics r_a and cv_b were evaluated against the summary statistics derived from the simulations under the corresponding climate setting observed in the empirical dataset (Supporting information). The median and 0.1–0.9 quantiles (in brackets) of the posterior distributions of the estimated *tau* and *nsd* parameters are shown in the right columns.

Dataset	Climate	r_a	cv_b	*mean.r*^{2}	*tau*	*nsd*
Leaf beetles-Spain	Linear	0.99	0.17	0.73	0.5 (0.5–0.8)	7.39 (2.7–12.2)
Water beetles-Europe	Linear	0.99	0.18	0.39	0.5 (0.5–0.8)	7.39 (2.7–12.2)
Water beetles-Australia	Linear	0.99	0.09	0.52	0.5 (0.5–0.8)	4.48 (1.7–12.2)
Water beetles-Madagascar	Concentrical	0.97	0.26	0.14	0.5 (0.5–3.7)	4.48 (1.6–7.4)
Darkling beetles-Aegean Islands	Linear	0.99	0.18	0.42	0.5 (0.5–0.8)	7.39 (2.7–12.2)
Dung beetles-Costa Rica	Bidirectional	0.98	0.19	0.17	20 (20–20)	1.65 (1.0–4.5)
Bats-SE Asia	Linear	0.99	0.14	0.86	0.5 (0.5–0.8)	4.48 (1.7–12.2)
Bats-Guyana	Concentrical	0.99	0.51	0.1	0.5 (0.5–0.5)	7.39 (4.5–20)
Molluscs-Spain	Linear	0.99	0.07	0.37	0.5 (0.5–0.8)	7.39 (2.7–12.2)
Dragonflies-New Brunswick	Linear	0.98	0.25	0.07	0.5 (0.5–0.8)	4.48 (1.7–7.4)
Caddisflies-New Brunswick	Linear	0.99	0.15	0.47	0.5 (0.5–0.8)	4.48 (1.7–12.2)
Caddisflies-S. Pennsylvania	Linear	0.98	0.63	0.11	0.5 (0.5–0.8)	4.48 (1–7.4)
Ants-Mauritius	Concentrical	0.31	0.31	0.11	1.359 (0.8–3.7)	1 (1–20)
Butterflies-Romania	Bidirectional	0.97	0.53	0.12	20 (20–20)	2.72 (1–4.5)
Nymphalids-Yucatan	Bidirectional	1.00	1.49	0.02	20 (20–20)	2.72 (1–4.5)
Lepidoptera-Papua	Bidirectional	0.98	1.72	0.04	20 (20–20)	2.72 (1–4.5)

in a narrow area above this logarithmic function, showing flatter slopes of the decay curves. For example, the water beetles of Madagascar showed a slope in distance–decay curves like those of the water beetles of Europe and Australia, suggesting that they have a similar dispersal ability. However, at this relative small spatial scale (Madagascar), the water beetle dispersal ability is sufficient to make their spatial range being environmentally constrained, while European and Australian water beetle communities are dispersal limited because of the relatively larger spatial scales of these datasets. A similar outcome was observed for the bat communities of SE Asia (dispersal limited) and Guyana (environmentally constrained).

Discussion

Our framework predicts how varying strengths of environmental and dispersal constraints modify the relationship between community similarity and spatial distance at multiple genealogical scales (haplotypes, intraspecific lineages and species). Spatial turnover is expected to shift regularly across genealogical scales (i.e. genealogical scaling of initial similarity) and to decay with spatial distance at the same rate across genealogical scales (i.e. genealogical invariance of distance–decay slopes), when the process is driven exclusively by dispersal. On the contrary, the regularity is disrupted if the environment constraints the distributional dynamics of species (but not the distributions at lower genealogical scales, haplotypes and intraspecific lineages). We have used simulations in which the expansion of haplotype and species ranges is driven by varying combinations of dispersal and environmental constraints. In the simulated biological communities, the relative contributions of dispersal versus environmental constraints in controlling community turnover are known. Therefore, the joint analysis of distance–decay patterns at multiple genealogical scales in these simulated communities produces unequivocal predictions of how varying combinations of dispersal and environmental constraints impact the variance in distance–decay patterns across genealogical scales (measured with two summary statistics). Finally, an approximate Bayesian framework (ABC) allows estimating the relative relevance of dispersal limitation and environmental constraint in any empirical biological systems, by comparing the summary statistics from empirical datasets against those from the full spectrum of simulations. Because these simulations represent general scenarios with simple climatic structures and simple rules for range expansion, they can be used as a general benchmark in future studies aiming to estimate the relative roles of dispersal limitation and niche constraint. They might also be used as a basis to build more specifically tuned simulations, with particular climatic settings, dispersal abilities or environmental constraints across species and regions.

Our simulations show that the genealogical invariance of distance–decay rates and the genealogical scaling of initial similarity is only expected when dispersal is the main driver of range dynamics. In such scenario, the rate of decay in community similarity with spatial distance only depends on the organisms' ability to expand their ranges. Hence the slope of distance–decay curves can be considered as a proxy for the dispersal ability of each biological group (Qian 2009, Wetzel et al. 2012, Saito et al. 2015, Chust et al. 2016, Gómez-Rodríguez and Baselga 2018). This direct link between dispersal ability and the slope of distance–decay curves allows two additional inferences. First, dispersal limitation will be effective at different spatial scales depending on the intrinsic dispersal ability of organisms. Evidently, dispersal limitation will be relevant at small spatial scales only in organisms with the poorest dispersal abilities while, at large spatial scales, dispersal limitation may also play a role in structuring the diversity patterns of organisms with better dispersal abilities. This implies that, for any given dispersal ability, there

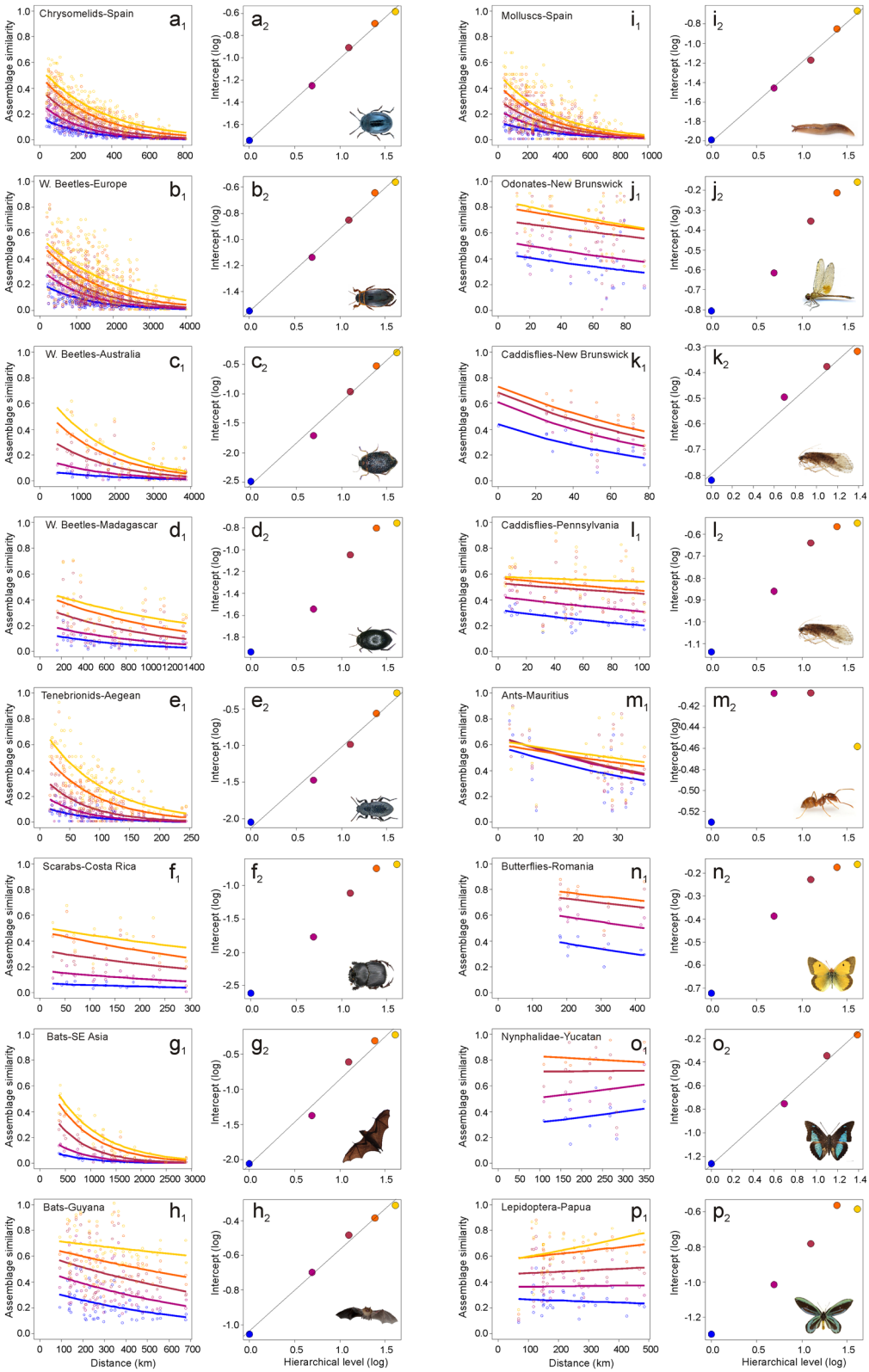

Figure 4. Observed distance–decay patterns at multiple genealogical scales from haplotypes to species for different taxonomic groups and geographic areas. For each empirical dataset, two plots are included: distance–decay patterns at multiple genealogical scales (plot with subscript = 1), and log–log relationship between initial similarity (intercept) and genealogical scale (subscript = 2). In both columns, colours represent the genealogical scales from lower (haplotypes, blue) to higher levels (putative species, yellow). The linear fit is shown when $r_a \geq 0.99$. Pictures were downloaded from Wikimedia Commons and other sources under Creative Commons license (ant by Michael Bentley; bats by Christoph F. Robiller and Hasitha Tudugalle; beetles by Udo Schmidt; butterflies by Didier Descouens, Vítězslav and Maňák Robert Nash; caddisfly by CBG Photography Group, Centre for Biodiversity Genomics; dragonfly by William Haber; slug by Africa Gómez).

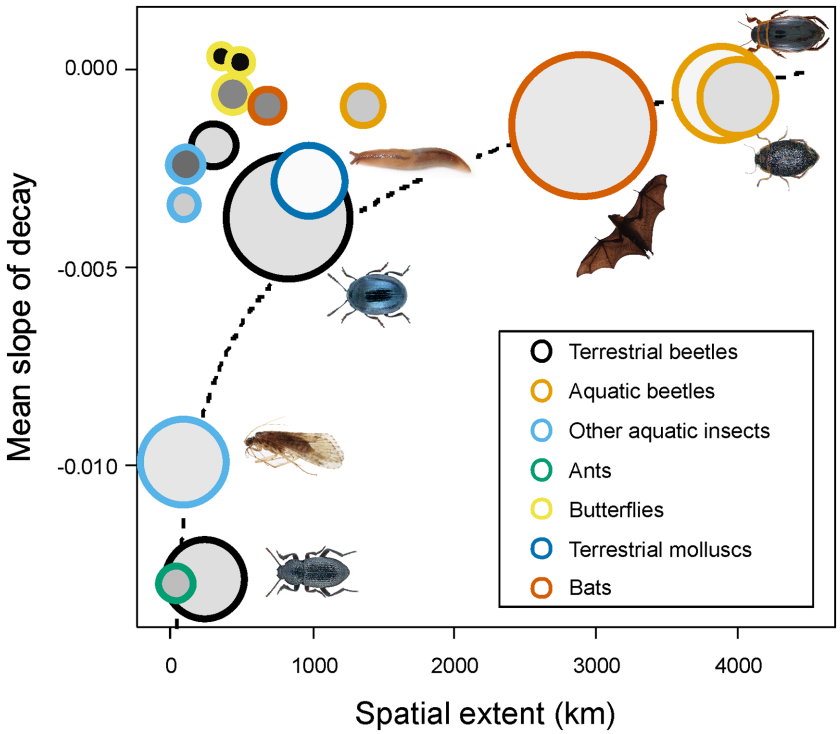

Figure 5. Relationship between distance–decay slope (mean value across genealogical scales) of the distance–decay pattern and spatial extent (maximum distance between sites in each empirical dataset). Summary statistics are indicated by the circle filling shade (cv_b, variability of slopes across genealogical scales, from low variability = light grey to high variability = dark grey) and circle size ($mean.r^2$, fit of decay curves across scales), showing that datasets presenting a genealogical invariance of distance–decay patterns occupy a specific region in the space defined by dispersal rate (slope of decay) and geographical extent. The dashed line is the logarithmic relationship ($r^2 = 0.79$, $F_{1,5} = 19.36$, $p = 0.007$) between slope and extent for those datasets in which genealogical invariance was observed (marked with pictures).

exists a minimum spatial extent below which environmental constraint becomes an effective limit to expansion of species ranges and, therefore, the distance–decay curves show genealogical scale-dependence. Below that minimum spatial extent, dispersal limitation would not effectively constraint species ranges and, instead, the environment may become the most relevant factor limiting species distributions. Second, in dispersal-limited systems, where range expansion at all genealogical scales is simply time-dependent, the relationship between the lineages' spatial range sizes and genealogical depth will depend on the dispersal ability of each biological group. Given that dispersal ability is accounted for by the slope of the distance–decay curves, the spatial and genealogical scales of community turnover become equivalent, i.e. they are different expressions of the same dispersal process over time and space.

Our predictive framework is built upon simulating a mechanistic process of range expansion of haplotypes while accounting for the genealogical relationships among them. We assume a macroecological model of lineage birth and range expansion, whereby lineage ranges gradually spread from their origin at a uniform rate, which causes similarity to decrease with distance at the same rate at all genealogical scales, unless ranges at the species level are under environmental constraint (see animation in the Supporting information). This simulation represents a simplified model of community formation that only considers lineage birth, age, lineage dispersal and species ecological niche, without considering local species interactions in the communities (García-Callejas et al. 2019). We do not further explore species interactions or alternative ecological niche processes because we are focused on the environmental constraints affecting species assembly processes, which is most likely the dominating process across terrestrial biological systems (McGill 2010, Mendoza and Araújo 2019). As a further simplification, we do not consider extinction, in line several studies of species ranges and lineage ages across macroecological settings (Barraclough and Vogler 2000). This model is powerful for the analysis of large-scale patterns of ranges without the detail of individual-based models of neutral metacommunities and thus allows the analysis of thousands of entities (and can be applied at multiple genealogical scales).

Among the empirical datasets used, a large fraction (11 of 16 cases) showed genealogical invariance in the distance–decay curves although the variance explained by the distance–decay curves was low in four cases. Therefore, the conservative estimate is that dispersal limitation is the dominant process in seven case studies. That is more than the cases for which a preponderance of environmental constraint was inferred. It is worth noting that these dispersal-limited biological systems include organisms with low mobility, like darkling beetles, leaf beetles, caddisflies and terrestrial molluscs, as well as

highly mobile organisms like water beetles and bats sampled at very large, continental scales. In contrast, the systems where climatic constraints were found to be preponderant included organisms with high dispersal ability (dung beetle, lepidopterans and ants) that were analysed at relatively small spatial scales. For example, the ants of Mauritius Island showed a large variation in distance–decay slopes across genealogical scales and a complex pattern of increase–decrease–increase in the distance–decay intercepts from lower to higher genealogical scales. Our ABC predictive framework shows that similar distance–decay curves and summary statistics (r_a and cv_b) are produced when simulating a strong environmental constraint and relatively low dispersal ability in a mountain climatic gradient (Fig. 2k). Likewise, the Papuan butterflies showed large variation in distance–decay slopes (most of them positive), low distance–decay intercept at the species level, and very poor fit of exponential decay models at all genealogical scales. Our ABC predictive framework shows that similar summary statistics were obtained in simulations under narrow environmental tolerance and extremely high dispersal ability in a mountain climatic gradient (Fig. 2h), a scenario fitting the proposed biological scenarios of butterfly communities in Papua (Craft et al. 2010). It should be noted that, as with any other method, the accuracy of our inferences depends on the quality of the data. Sampling effort (number of sampled sites and number of sequenced specimens) differed across empirical datasets, so the robustness of the inference might also vary.

In conclusion, the proposed predictive framework introduces two major features that, taken together, can help improve our ability to discern the underlying causal factors for the spatial turnover among biological communities. First, by directly manipulating the haplotype and species range expansion, we can produce specific predictions of how community turnover will remain scale-invariant or vary across genealogical scales depending on the relative strengths of dispersal limitation and niche constraint. Second, these predictions can be used to estimate the relative strength of dispersal limitation and environmental constraint in real biological communities. Whether dispersal limitation or environmental constraint control the distribution of biological diversity at varying spatial scales is crucial for understanding and predicting global change effects on the composition and functioning of biological communities (Dornelas et al. 2014). Our results are consistent with the view that biodiversity models should take into account the taxon-specific traits affecting dispersal limitation and environmental constraint. Such traits would include body size, flight ability, degree of ecological specialisation, trophic preferences and others, in line with the observed relationship between the strength of the latitudinal richness gradient and body mass (Weiser et al. 2018). Importantly, we demonstrate that assessing biodiversity patterns at multiple genealogical scales allows estimating the relative importance of dispersal and environmentally driven processes as constraints of community turnover. Further research should focus on gathering empirical evidence on whether dispersal limitation or environmental constraint become critical range determinants for a wider spectrum of biological groups and geographical regions.

Acknowledgements – We are grateful for the thorough comments provided by Fabian Laroche, Victor Saito, David Storch and Timothy Keitt to previous versions of this manuscript. Computer simulations were run in the High Performance Computing Service of Galicia (<www.cesga.es/>), Spain.
Funding – This research was supported by the Spanish Ministerio de Cienci e Innovación and the European Regional Development Fund (ERDF) through grants CGL2016-76637-P and PID2020-112935GB-I000. DP and MA receive support from the Xunta de Galicia. MBA and MA acknowledge support from Ministerio de Ciencia, Innovación y Universidades through the grants PGC2018-099363-B-I00 and RYC-2015-18241, respectively.

Author contributions

Andrés Baselga: Conceptualization (lead); Data curation (equal); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Software (equal); Writing – original draft (lead); Writing – review and editing (equal). **Carola Gómez-Rodríguez**: Conceptualization (lead); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Writing – original draft (lead); Writing – review and editing (equal). **Miguel B. Araújo**: Conceptualization (equal); Writing – review and editing (equal). **Adrián Castro-Insua**: Conceptualization (equal); Formal analysis (equal); Software (equal); Writing – review and editing (equal). **Miguel Arenas**: Conceptualization (equal); Formal analysis (equal); Methodology (equal); Writing – review and editing (equal). **David Posada**: Conceptualization (equal); Methodology (equal); Writing – review and editing (equal). **Alfried P. Vogler**: Conceptualization (lead); Investigation (equal); Writing – review and editing (equal).

Transparent Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.05808>.

Data availability statement

All the datasets analysed during the current study are already available in GenBank, see accession numbers in the Supporting information (Table S1). No new data is used in this paper.

Supporting information

The supporting information associated with this article is available from the online version.

References

Araújo, M. B. and Pearson, R. G. 2005. Equilibrium of species' distributions with climate. – Ecography 28: 693–695.

Arribas, P. et al. 2020. The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding. – Mol. Ecol. 30: 48–61.
Avise, J. C. 1994. Molecular markers, natural history and evolution. – Chapman and Hall.
Barraclough, T. G. and Vogler, A. P. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. – Am. Nat. 155: 419–434.
Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. – Global Ecol. Biogeogr. 19: 134–143.
Baselga, A. and Orme, C. D. L. 2012. betapart: an R package for the study of beta diversity. – Methods Ecol. Evol. 3: 808–812.
Baselga, A. et al. 2013. Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. – Nat. Commun. 4: 1892.
Baselga, A. et al. 2015. Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. – Global Ecol. Biogeogr. 24: 873–882.
Baselga, A. et al. 2020. betapart: partitioning beta diversity into turnover and nestedness components. R package ver. 1.5.2. – <http://CRAN.R-project.org/package=betapart>.
Beaumont, M. A. 2010. Approximate Bayesian computation in evolution and ecology. – Annu. Rev. Ecol. Evol. Syst. 41: 379–406.
Borcard, D. et al. 1992. Partialling out the spatial component of ecological variation. – Ecology 73: 1045–1055.
Bringloe, T. 2013. Lotic Trichoptera Larvae of Southern Pennsylvania. – BOLD Database, Project LTLSP.
Chust, G. et al. 2016. Dispersal similarly shapes both population genetics and community patterns in the marine realm. – Sci. Rep. 6: 28730.
Clare, E. L. et al. 2011. Neotropical bats: estimating species diversity with DNA barcodes. – PLoS One 6: e22648.
Clement, M. et al. 2000. TCS: a computer program to estimate gene genealogies. – Mol. Ecol. 9: 1657–1659.
Craft, K. J. et al. 2010. Population genetics of ecological communities with DNA barcodes: an example from New Guinea Lepidoptera. – Proc. Natl Acad. Sci. USA 107: 5041–5046.
Csilléry, K. et al. 2012. abc: an R package for approximate Bayesian computation (ABC). – Methods Ecol. Evol. 3: 475–479.
Curry, J. C. et al. 2012. Congruence of biodiversity measures among larval dragonflies and caddisflies from three Canadian rivers. – Freshwater Biol. 57: 628–639.
Dincă, V. et al. 2011. Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. – Proc. R. Soc. B 278: 347–355.
Dornelas, M. et al. 2014. Assemblage time series reveal biodiversity change but not systematic loss. – Science 344: 296–299.
Fitzpatrick, M. C. et al. 2013. Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. – Proc. R. Soc. B 280: 20131201.
Francis, C. M. et al. 2010. The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia. – PLoS One 5: e12575.
García-Callejas, D. et al. 2019. Spatial trophic cascades in communities connected by dispersal and foraging. – Ecology 100: e02820.
García-López, A. et al. 2013. Beta diversity at multiple hierarchical levels: explaining the high diversity of scarab beetles in tropical montane forests. – J. Biogeogr. 40: 2134–2145.
Gilbert, B. and Lechowicz, M. J. 2004. Neutrality, niches, and dispersal in a temperate forest understory. – Proc. Natl Acad. Sci. USA 101: 7651–7656.
Gómez-Rodríguez, C. and Baselga, A. 2018. Variation among European beetle taxa in patterns of distance decay of similarity suggests a major role of dispersal processes. – Ecography 41: 1825–1834.
Gómez-Rodríguez, C. et al. 2019. Understanding dispersal limitation through the assessment of diversity patterns across phylogenetic scales below the species level. – Global Ecol. Biogeogr. 28: 353–364.
Graham, C. H. et al. 2018. Phylogenetic scale in ecology and evolution. – Global Ecol. Biogeogr. 27: 175–187.
Hendrich, L. et al. 2010. Mitochondrial *cox1* sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. – PLoS One 5: e14448.
Isambert, B. et al. 2011. Endemism and evolutionary history in conflict over Madagascar's freshwater conservation priorities. – Biol. Conserv. 144: 1902–1909.
König, C. et al. 2017. Dissecting global turnover in vascular plants. – Global Ecol. Biogeogr. 26: 228–242.
Lenoir, J. et al. 2020. Species better track climate warming in the oceans than on land. – Nat. Ecol. Evol. 4: 1044–1059.
McGill, B. J. 2010. Towards a unification of unified theories of biodiversity. – Ecol. Lett. 13: 627–642.
Mendoza, M. and Araújo, M. B. 2019. Climate shapes mammal community trophic structures and humans simplify them. – Nat. Commun. 10: 5197.
Panchal, M. 2007. The automation of nested clade phylogeographic analysis. – Bioinformatics 23: 509–510.
Papadopoulou, A. et al. 2011. Testing the species-genetic diversity correlation in the Aegean archipelago: toward a haplotype-based macroecology? – Am. Nat. 178: 560–560.
Paradis, E. et al. 2004. APE: analyses of phylogenetics and evolution in R language. – Bioinformatics 20: 289–290.
Peres-Neto, P. R. et al. 2006. Variation partitioning of species data matrices: Estimation and comparison of fractions. – Ecology 87: 2614–2625.
Prado, B. R. et al. 2011. Beyond the colours: discovering hidden diversity in the Nymphalidae of the Yucatan Peninsula in Mexico through DNA barcoding. – PLoS One 6: e27776.
Qian, H. 2009. Beta diversity in relation to dispersal ability for vascular plants in North America. – Global Ecol. Biogeogr. 18: 327–332.
Qian, H. et al. 2005. Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. – Ecol. Lett. 8: 15–22.
Saito, V. S. et al. 2015. Dispersal traits drive the phylogenetic distance decay of similarity in Neotropical stream metacommunities. – J. Biogeogr. 42: 2101–2111.
Simpson, G. G. 1943. Mammals and the nature of continents. – Am. J. Sci. 241: 1–31.
Slatkin, M. 1985. Gene flow in natural populations. – Annu. Rev. Ecol. Syst. 16: 393–430.
Smith, M. A. and Fisher, B. L. 2009. Invasion, DNA barcodes, and rapid biodiversity assessment using ants of Mauritius. – Front. Zool. 6: 31.
Smith, T. W. and Lundholm, J. T. 2010. Variation partitioning as a tool to distinguish between niche and neutral processes. – Ecography 33: 648–655.
Soininen, J. et al. 2007. The distance decay of similarity in ecological communities. – Ecography 30: 3–12.

Svenning, J. C. and Skov, F. 2007. Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? – Ecol. Lett. 10: 453–460.
Taheri, S. et al. 2021. Improvements in reports of species redistribution under climate change are required. – Sci. Adv. 7: eabe1110.
Templeton, A. R. et al. 1992. A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data. 3. Cladogram estimation. – Genetics 132: 619–633.
Tuomisto, H. et al. 2003. Dispersal, environment and floristic variation of western Amazonian forests. – Science 299: 241–244.
Tuomisto, H. et al. 2012. Modelling niche and neutral dynamics: on the ecological interpretation of variation partitioning results. – Ecography 35: 961–971.
Vellend, M. 2003. Island biogeography of genes and species. – Am. Nat. 162: 358–365.
Vellend, M. 2016. The theory of ecological communities. – Princeton Univ. Press.
Vellend, M. and Geber, M. A. 2005. Connections between species diversity and genetic diversity. – Ecol. Lett. 8: 767–781.
Vellend, M. et al. 2014. Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. – Mol. Ecol. 23: 2890–2901.
Weiser, M. D. et al. 2018. Toward a theory for diversity gradients: the abundance–adaptation hypothesis. – Ecography 41: 255–264.
Wetzel, C. E. et al. 2012. Distance decay of similarity in neotropical diatom communities. – PLoS One 7: e45071.