Global Ecology and Biogeography

A Journal of Macroecology

RESEARCH PAPERS

Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes

Andrés Baselga[1]*, Carola Gómez-Rodríguez[1,2] and Alfried P. Vogler[2,3]

[1]*Departamento de Zoología, Facultad de Biología, Universidad de Santiago de Compostela, c/ Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain,* [2]*Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK,* [3]*Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK*

ABSTRACT

Aim Large-scale DNA barcoding allows the simultaneous assessment of variation in community composition at species level and below. We here propose that the patterns emerging across multiple hierarchical levels can be used to discern the effects of neutral and non-neutral macroecological processes, which otherwise have proven difficult to separate.

Location Iberian Peninsula.

Methods We performed *cox1* barcoding on 20 complete assemblages of leaf beetles for 4533 individuals of 203 species. The neutrality of *cox1* sequence evolution was tested using Tajima's test. Haplotypes (n = 2020) were grouped into nested n-step networks of up to five intraspecific hierarchical levels. We then assessed whether the spatial variation in assemblage composition at all hierarchical levels from haplotype to species was self-similar (fractal) and predictable from level to level.

Results Tajima's test on a subset of widely sampled species (n = 136) was consistent with neutral evolution in 83% of the species, but only 3% of cases exhibited balancing selection. Multiple hierarchical levels representing haplotype genealogies of various ages showed a similar rate of distance decay of assemblage similarity. In addition, we found strong log-log correlations between hierarchical level (lineage age) and number of lineages, lineage range size and assemblage similarity. Similarity at the species level was strongly correlated to similarity at the haplotype level for the whole assemblage ($r^2 = 0.75$) or for within-species haplotype similarity (mean $r^2 = 0.17$, SE = 0.03).

Main conclusions These findings suggest great regularities in the pattern of assemblage variation at all lineage ages that are best explained by the enduring action of stochastic (neutral) processes of mutation and dispersal. The multi-hierarchical analysis therefore bridges predictions of the neutral theory of molecular evolution and the neutral theory of biodiversity. Neutral processes thus emerge as a unifying principle of ecology and evolution, which has deep implications in biodiversity assessment and conservation.

Keywords

Beta diversity, distance-decay of similarity, DNA barcoding, fractals, nested clade analysis, neutral theory, niche theory, species-genetic diversity correlation.

*Correspondence: Andrés Baselga, Departamento de Zoología, Facultad de Biología, Universidad de Santiago de Compostela, c/ Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain. E-mail: andres.baselga@usc.es

INTRODUCTION

Large-scale, systematic DNA sequencing of entire communities opens new opportunities for unveiling the processes that govern the distribution of biodiversity. Understanding the causes of macroecological patterns has proven difficult, despite major advances derived from empirical studies of large-scale statistical patterns of species richness, composition and abundance (e.g., Brown, 1995; Nekola & White, 1999; Willig *et al.*, 2003). Several 'unified' theories have been developed that explain regularities

in these macroecological distributions, although they emphasize different underlying processes regarding the role of niche-derived and neutral factors (McGill, 2010). Despite their fundamental differences, various models – including the Neutral Theory of Biodiversity (Hubbell, 2001; Rosindell *et al.*, 2011) and the Ecological Niche Theory (Hutchinson, 1957; Leibold, 1995; Peterson *et al.*, 2011) – are consistent with widely observed macroecological patterns such as the decline of community similarity with spatial distance (i.e. distance-decay of similarity, Nekola & White, 1999), the species abundance–range size correlations (McGill *et al.*, 2007), or the species–area curves (Drakare *et al.*, 2006). Because coinciding patterns are predicted from competing mechanisms, it has proven difficult to discern the effects of neutral and non-neutral ecological processes on biodiversity.

The integrated assessment of macroecological patterns at various levels of genotypes, genealogies and species (i.e. multi-hierarchical macroecology, see Baselga *et al.*, 2013) provides a novel avenue to disentangle the competing potential drivers of biodiversity patterns. This approach has antecedents in studies assessing the species–genetic diversity correlation (SGDC, Vellend, 2003), suggesting that species richness of a local assemblage is correlated with the genetic diversity within each of the local species because both patterns may be driven by a single process (Vellend *et al.*, 2014). The SGDC was also observed in analyses of beta diversity that show a regular decay of assemblage similarity with geographic distance at both species and genetic levels (Sei *et al.*, 2009; Odat *et al.*, 2010; Papadopoulou *et al.*, 2011). Although opening promising ways to investigate the processes underlying biodiversity, it has been shown that the SGDC alone does not suffice to separate neutral from non-neutral drivers, as both niche and neutral processes may result in the SGDC (Vellend & Geber, 2005; Vellend *et al.*, 2014), and even neutral processes alone can generate correlations of opposite sign between species and genetic diversity (Laroche *et al.*, 2015).

Multi-hierarchical macroecology is related to the SGDC, but investigates the spatial patterns of variation of entire assemblages at multiple hierarchical levels between haplotypes and species, rather than the correlation between single species genetic diversity with species diversity. The levels below species can be studied by grouping genetic variants into increasingly more inclusive genealogical entities based on the relationships of haplotypes. In the case of non-recombining mitochondrial genomes, these haplotype groups reflect different depths of time of the gene genealogy extending back to an ancestor near the species origin. As gene copies acquire mutations during lineage history, they also are subject to dispersal, which extends their spatial ranges and builds up spatial patterns of assemblage variation. Macroecological analyses can therefore be extended to lower hierarchical levels in analogy to the analysis of variation in species richness and species composition in classical macroecology. This extension to multiple hierarchical levels takes advantage of novel patterns emerging across levels, which might help to discern between neutral and non-neutral controls of assemblage variation (Baselga *et al.*, 2013).

The central hypothesis of multi-hierarchical macroecology is that the effect of neutral processes (including neutral mutation, dispersal limitation, birth and death of lineages) is uniform across hierarchical levels, while the effect of non-neutral processes differs among levels. In particular, Ecological Niche Theory (Hutchinson, 1957; Leibold, 1995; Soberón & Nakamura, 2009) proposes that species distributions are constrained by abiotic (e.g., climate) and biotic factors (e.g., interactions), via their effects on the fitness of individuals and the viability of populations. In contrast, there is no suggestion that specific mitochondrial DNA (mtDNA) haplotypes are linked to a particular niche (i.e. different from haplotype niches of conspecifics). The evolution of mtDNA has been considered to be predominantly controlled by neutral processes (Slatkin, 1985; Avise, 1994). Some recent work proposed that mtDNA is controlled by selective processes (Bazin *et al.*, 2006), but this did not overturn the view that mtDNA variation is predominantly neutral (e.g., Mulligan *et al.*, 2006; Albu *et al.*, 2008; Nei *et al.*, 2010; Karl *et al.*, 2012). In addition, evolutionary neutrality can be empirically tested in each specific study system by assessing how DNA variation is distributed across common and rare haplotypes (Tajima, 1989). This establishes the basis of our predictive framework: if neutral molecular evolution of mtDNA is not rejected, this is evidence for the absence of selective regimes that would determine the spatial distribution of genetic variation. Hence, the spatial range of mtDNA haplotypes within a species range would be controlled by neutral processes only (Slatkin, 1985; Avise, 1994). Consequently, the variation in haplotype composition of assemblages would provide a benchmark of ecological neutrality against which to compare the variation of assemblages at the species level.

This leads to the major prediction of multi-hierarchical macroecology that, under neutral dynamics, the variation of composition of entire assemblages at and below the species level is self-similar at any temporal and spatial scale (Baselga *et al.*, 2013). Self-similarity, in turn, should be eroded under the effect of niche processes at the species level as non-neutral processes affect spatial ranges of lineages differently at various levels. We thus develop a set of tests involving multiple hierarchical levels for distinguishing neutral and non-neutral scenarios of assemblage variation. The approach includes first testing the hypothesis of neutral evolution of mitochondrial haplotypes to assess whether their distributions can be used as the required neutral benchmark. The second step involves assessing self-similarity of assemblage variation across hierarchical levels, by establishing hierarchical groups from haplotype networks that represent lineages of regularly older ages (from haplotypes to species), which then can be assessed for their spatial patterns at different levels. Under neutral dynamics the spatial biotic ranges of lineages are controlled by dispersal limitation and stochastic mutation and extinction only, and if assessed at different lineage ages, it is predicted that the resulting pattern exhibit a fractal self-similarity. Hence, in a third step of the protocol we test how well variation in haplotype composition can predict the variation in species composition, and whether deviations from this relationship are related to environmental differences between sites, which would point to niche effects at the species level.

We conducted whole-assemblage DNA sequencing of the mitochondrial cytochrome oxidase I (*cox1*) 'barcode' marker to test for patterns of sequence variation, using assemblages of leaf beetles (Chrysomelidae) across the Iberian Peninsula. Their composition may plausibly be affected by niche driven processes of host plant associations and latitudinal and altitudinal environmental gradients. They are thus an excellent system to assess the prevalence of neutral and niche based processes. The study area, in the Iberian Peninsula, is comparatively little affected by Pleistocene glacial events, and this relative long-term climatic stability should produce equilibrium conditions either with environmental factors due to niche-based processes or with spatial distributions from long-term stochastic dispersal. Iberian leaf beetles therefore are suitable to test the predictability of biodiversity patterns across levels, from genetic to species diversity, expected under neutral processes, and the generality of fractal patterns in assemblage composition.

MATERIALS AND METHODS

Field sampling and taxonomic species identification

Leaf beetle assemblages were sampled in 20 localities along a South–North transect (820 km) in Spain (Table S1, Appendix S1 in Supporting Information) in April–June 2010. Each locality was separated from the closest locality by a minimum of 36 km [ALC-UBG] and a maximum of 106 km [ANC-EUM]. Sampling localities were selected to maximize the climatic variation and spanned an altitudinal range between 120 and 1270 m a.s.l. All localities were well preserved areas (most Natural Parks or areas with some degree of protection) combining forests dominated by different oak species, with shrubs and meadows. Each locality was intensively sampled by sweeping and beating all types of vegetation, including trees, shrubs and herbs, for 20 sampling periods of 30 min (18 sampling units in UBG). Collecting permits were issued by the corresponding regional governments: Junta de Andalucía (ALC, UBG, SNS), Junta de Extremadura (JCB, HOR, COR, VER, SSP, DEL), Junta de Castilla y León (FRN, ADS, ADN, SAN, OMA, TUE) and Xunta de Galicia (LAS, LAR, ANC, MAC, EUM). All specimens were preserved in 100% ethanol for DNA extraction. Specimens were identified to species level using the taxonomic monographs for the European (Warchalowski, 2003) and the Iberian (Petitpierre, 2000) leaf beetle faunas. When necessary, male and female genitalia were dissected and mounted together with specimens using dimethyl-hydantoin formaldehyde resin (DMHF).

Sequence data, DNA-based species delimitation and phylogenetic analysis

Genomic DNA was extracted from muscle tissue in the prothorax region with Wizard SV 96-well plates (Promega, UK) or with a BioSprint 96 workstation (Qiagen, Germany). A 655 base pair region from the 5′ end of mitochondrial *cox1* was amplified with primers CO1F2 (TCTACYAATCATAAAGATATTGGTAC) and

CO1R2 (ACTTCTGGATGACCAAAGAATCA) in most cases or with standard LCO/HCO primers (Folmer *et al.*, 1994) when the previous primers failed. Amplification was performed with Bioline BioTaq and the following cycling: 95 °C for 2 min, 35 cycles of 95 °C for 30 s, 40 °C for 30 s and 72 °C for 45 s, and final extension of 72 °C for 5 min. PCR products were cleaned with 96-well Millipore multiscreen plates and sequenced in both directions using ABI dye terminator sequencing. Sequence chromatograms were assembled and manually edited using Genious 5.6. DNA sequences are available under GenBank accession numbers KF134544 – KF134651 and KF652242 – KF656666.

The null hypothesis of neutral evolution of mtDNA was tested using Tajima's *D* test (Tajima, 1989) on the *cox1* sequences of those species with > 3 available sequences (4376 sequences of 136 species in total). Tajima's *D* tests were conducted using the R package *pegas* (Paradis, 2010). Gene trees were constructed with RAxML 7.0.3 under the GTRGAMMA model, appropriate for large, complex data sets. Prior to tree building, identical haplotypes were collapsed into a single sequence. The maximum likelihood tree was made ultrametric with Pathd8 (Britton *et al.*, 2007), which uses a mean path length method for establishing a molecular clock. The ultrametric trees were used for DNA-based species delimitation with the Generalized Mixed Yule-Coalescent (GMYC) method that establishes the point of transition from slow to faster branching rates of the gene tree expected at the species boundary (Pons *et al.*, 2006) using the R package *splits* (SPecies LImits by Threshold Statistics) (Ezard *et al.*, 2009). All genotype information was included in the geographic analysis after application of the GMYC procedure, by restoring ('uncollapse') the haplotype data removed prior to tree building using a custom R script. Minor uncertainty in the haplotype identity in 28 sequences due to partially missing sequence data was resolved by assigning the haplotypes to one of the fully sequenced genotypes in the same locality.

Haplotype networks were created using the TCS software (Clement *et al.*, 2000) implemented in ANeCA v.1.2 (Panchal, 2007). TCS uses statistical parsimony to estimate haplotype networks of closely related individuals from DNA sequence data, which is defined by the 95% confidence interval for connections between haplotypes to be non-homoplastic in the network (Templeton *et al.*, 1992). The nesting algorithm generates *n*-step hierarchical nested clades of increasingly more inclusive groups of haplotypes, following rules for initially connecting all haplotypes that can be linked by one mutational step ('1-step networks'), which then are incorporated into groups requiring a connection of maximally two steps ('2-step networks'), and so on, until connections between haplotypes exceed the probability of encountering homoplastic changes. The large size of the dataset required that the TCS software was applied separately to smaller subclades. We therefore performed the nested clade analysis individually on five clades defined in a higher-level phylogenetic analysis of Chrysomelidae (Gómez-Zurita *et al.*, 2008): (1) donaciines + criocerines, (2) cryptocephalines + cassidines + lamprosomatines + eumolpines, (3) chrysomelines, (4) galerucines and (5) alticines.

Statistical analysis of assemblage similarity patterns

Pair-wise similarity of assemblages among localities was assessed at the levels of haplotypes, nested clades (*n*-step levels), GMYC groups and morphologically delimited species. At each level, patterns of similarity were analysed excluding the effects of spatial nestedness by means of the Simpson index of similarity, i.e. $1-\beta_{sim}$ (Lennon *et al.*, 2001; Baselga, 2010), using the R package *betapart* (Baselga & Orme, 2012). To assess the relationship of assemblage similarity with spatial distance and environmental (topo-climatic) differences, non-linear regression on similarity matrices was used to fit exponential decay curves expressed as $y = a{\ast}e^{-bx}$, where y is similarity at distance x, a initial similarity and $-b$ the rate of distance decay. Spatial distance was computed in km as the Euclidian distance between the centroids of localities. Topo-climatic distance was computed as the Euclidian distance in a multidimensional space consisting of seven standardized topo-climatic variables (Hijmans *et al.*, 2005), i.e. mean altitude, annual mean temperature (Bio1 in WorldClim data), maximum temperature of warmest month (Bio5), minimum temperature of coldest month (Bio6), annual precipitation (Bio12), precipitation of wettest quarter (Bio16) and precipitation of driest quarter (Bio17). These topo-climate variables have proven to be highly correlated to diversity patterns across various taxonomic groups (Hawkins *et al.*, 2003). The degree of association between biotic similarity and spatial or climatic distance was measured with the coefficient of determination (r^2) after linearizing the decay curves through log-transformation of similarity values. An arbitrary small quantity (0.01) was summed to all values to avoid undefined log-transformed values. The significance of all relationships was assessed with Mantel tests using the R package *vegan* (Oksanen *et al.*, 2011). Variance partitioning (Borcard *et al.*, 1992) was used to quantify the unique contributions of spatial and climatic distance to explaining assemblage similarity. To assess the predictability of biodiversity patterns across hierarchical levels, the relationship between patterns among hierarchical levels was computed by log-log regressions on similarity matrices (e.g., species vs. haplotype). Deviations from a perfect relationship can be used to assess the influence of non-neutral processes on assemblages. The rationale is that if niche processes control species distributions, similarity at the species level would be higher than predicted from the haplotype level in pairs of sites with similar climate, and lower than predicted from the haplotype level in sites with different climate

Similarity matrices were used to depict a network of assemblages for each hierarchical level. The networks were plotted with the R package *network* (Butts *et al.*, 2013). In a first stage, we built fully connected networks in which nodes were the local assemblages and edges were the similarities between them. Thereafter, networks at all levels were pruned by first considering all edges totally connected (link = 1) and then sequentially removing (link = 0) the first (i.e. 25% lowest similarities), second and third quartiles of each similarity matrix. This produced networks with decreasing number of edges. The networks were compared across hierarchical levels (haplotype,

NC1-NC4 and GMYC) by measuring its modularity and the degree of nodes at each stage of edge removal (first, second and third quartile). These parameters were computed using the *optimal.community* and *degree* functions, respectively, of the R package *igraph* (Csardi & Nepusz, 2006).

Finally, we assessed whether the correlation between species-level assemblage dissimilarity and haplotype-level similarity was also detectable when computing haplotype-level similarity from single species data (the classic SGDC test at the beta diversity level). To do so, we computed pairwise dissimilarity between sites using presence-absence tables of haplotypes within a single species (i.e. among all sites where any given species is present) and correlated it with the assemblage dissimilarity computed from species presence-absence tables for the same sites. Such correlation was computed for all the species that were present in more than 5 sites and had more than 5 individuals per site (n = 30).

RESULTS

A total of 5102 individuals from 20 local assemblages across the Iberian Peninsula were processed for DNA sequencing of the barcode marker (*cox1*). This yielded 4533 high-quality sequences and 2020 unique haplotypes from 203 species (as currently defined based on morphological delimitations). The number of species in an assemblage ranged from 26–63 (mean = 39, SD = 8). Sequence-based species delimitation with the GMYC procedure grouped total haplotype variation into 269 groups, ranging from 30–68 putative species (mean = 43, SD = 9) in an assemblage. Nested clade analysis yielded 1311, 811, 504 and 398 lineages at the first to fourth hierarchical levels (Nested Clade 1 to Nested Clade 4, NC1-NC4). These numbers of lineages were tightly log-log correlated with the hierarchical level ($r^2 = 0.96$, $F_{1,5} = 120.9$, $P < 0.0001$).

We first tested the hypothesis of neutral evolution of mtDNA, using Tajima's D for the 136 species with > 3 *cox1* sequences. In 113 species (83%) the null hypothesis of neutral evolution was not rejected ($-2 < D < 2$, $P > 0.05$). When neutrality was rejected, Tajima's D was significantly lower than expected ($D < -2$, $P < 0.05$, suggesting either purifying selection or population size expansion) in 19 species (14%), while it was higher than expected ($D > 2$, $P < 0.05$, suggesting either balancing selection or decrease in population size) in 4 species (3%). The distribution of Tajima's D statistic across species (mean = -0.67, SD = 1.23) was not significantly different from a normal distribution with the same mean and SD (Kolmogorov-Smirnov test $D = 0.12$, $P = 0.24$). The fact that neutral evolution was not rejected in the majority of species was not likely related to the small sample size in some species (e.g., 36 species had less than 10 specimens) because the relationship between Tajima's D statistic and the number of specimens was negative (i.e. species with more individuals were more likely to have more negative D) but negligible ($r^2 = 0.07$, $F_{1,134} = 9.6$, $P = 0.002$).

Assemblage similarity at species and haplotype levels decreased exponentially with spatial (Table 1, Fig. 1a) and, to a lesser degree, with climatic distance (Fig. 1b). Intermediate levels (NC1 to NC4) also showed similar patterns. Variance par-

Table 1 Relationship between biotic similarity and spatial or climatic distance at multiple hierarchical levels: haplotype, nested clade levels NC1 to NC4, GMYC (i.e. Generalized Mixed Yule-Coalescent) groups and species.

		Spatial distance				Climatic distance			
	N	r^2	P	a	b	r^2	P	a	b
haplotype	2020	0.74	< 0.0001	0.15	−0.0038	0.48	< 0.0001	0.17	−0.36
NC1	1311	0.75	< 0.0001	0.26	−0.0043	0.48	< 0.0001	0.28	−0.40
NC2	811	0.76	< 0.0001	0.41	−0.0043	0.47	< 0.0001	0.44	−0.40
NC3	504	0.75	< 0.0001	0.55	−0.0039	0.47	< 0.0001	0.58	−0.36
NC4	398	0.74	< 0.0001	0.62	−0.0035	0.42	< 0.0001	0.62	−0.31
GMYC	269	0.70	< 0.0001	0.60	−0.0026	0.52	< 0.0001	0.66	−0.26
species	203	0.67	< 0.0001	0.65	−0.0024	0.48	< 0.0001	0.71	−0.24

The analysis is based on pairwise Simpson similarity in assemblage composition. N is the number of lineages at each hierarchical level. Regression parameters corresponding to the formula $y = a*e^{-bx}$, where y similarity and x is spatial or climatic distance, and coefficients of determination (r^2), and p values obtained by Mantel tests are shown.

Figure 1 Assemblage similarity patterns of Iberian leaf beetle assemblages at multiple hierarchical levels. (a) Decay of similarity against geographic distance at the haplotype (blue), GMYC (Generalized Mixed Yule-Coalescent) groups (red), and species (pink) levels. See parameters in Table 1. (b) Decay of similarity against climatic distance at the haplotype (blue), GMYC (red), and species (pink) levels. (c) Log-transformed decay of similarity with spatial distance at multiple hierarchical levels: blue = haplotype, green tones = NC1-NC4 (i.e. nested clades 1 to 4), red = GMYC, pink = species. (d) Density plots showing the distribution of similarity at multiple hierarchical levels (colours as before).
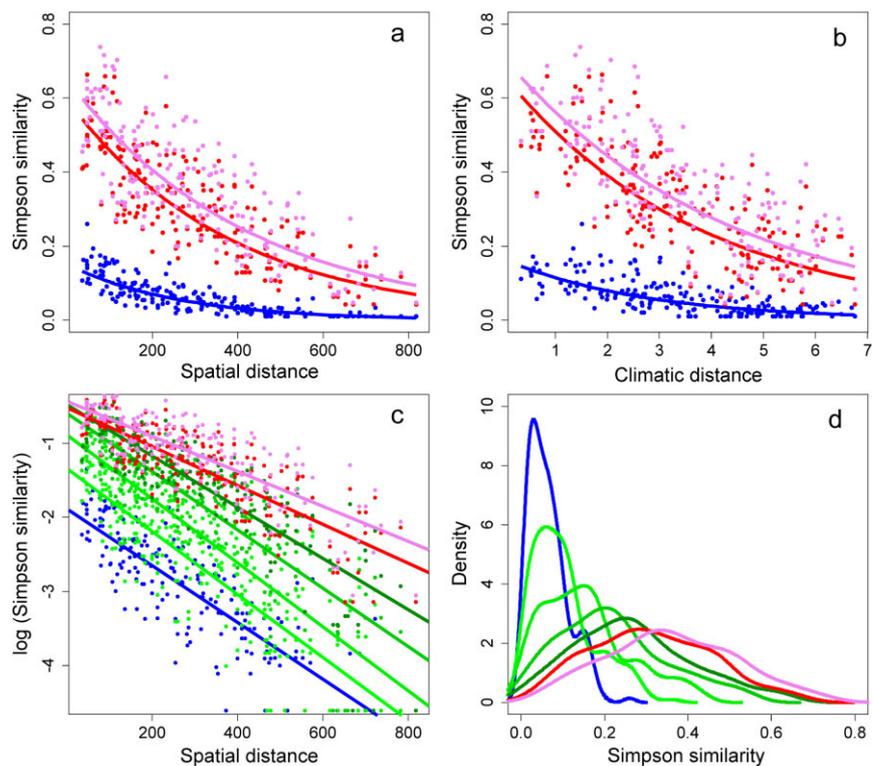
titioning showed that when both spatial and climatic distances are included in the model, the unique contribution to explained variance is still relevant for spatial distance (r^2 between 0.22 and 0.33, see Table S2) but negligible for climatic distance (r^2 between 0.01 and 0.04). The exponential decay of assemblage similarity with spatial distance had a very similar slope at each level, but assemblage similarity uniformly increased from level to level (Fig. 1c, Table 1). The latter is seen also from density plots that demonstrate the shift of similarity distributions to higher values with each level (Fig. 1d). The relationship of initial similarity and hierarchical level was log-log correlated (r^2 = 0.96, $F_{1,5}$ = 130.0, P < 0.0001). Equally, mean similarity of assemblage composition at each hierarchical level was log-log correlated

with the hierarchical level (r^2 = 0.99, $F_{1,5}$ = 554.7, P < 0.0001). The log-log linear correlations suggest that the patterns of assemblage variation across hierarchical levels can be described by a fractal geometry.

The self-similarity of variation in assemblage composition across levels was also evident from network analyses. These networks are visual representations of matrices of assemblage similarity among sites at multiple levels (haplotypes, nested clades, species). The nodes of networks are thus the sites (or assemblages) and the links are binary transformations of assemblage similarity using different thresholds (first, second and third quartiles of similarities). Here we use the modularity of networks, which measures the number of different clusters, as a
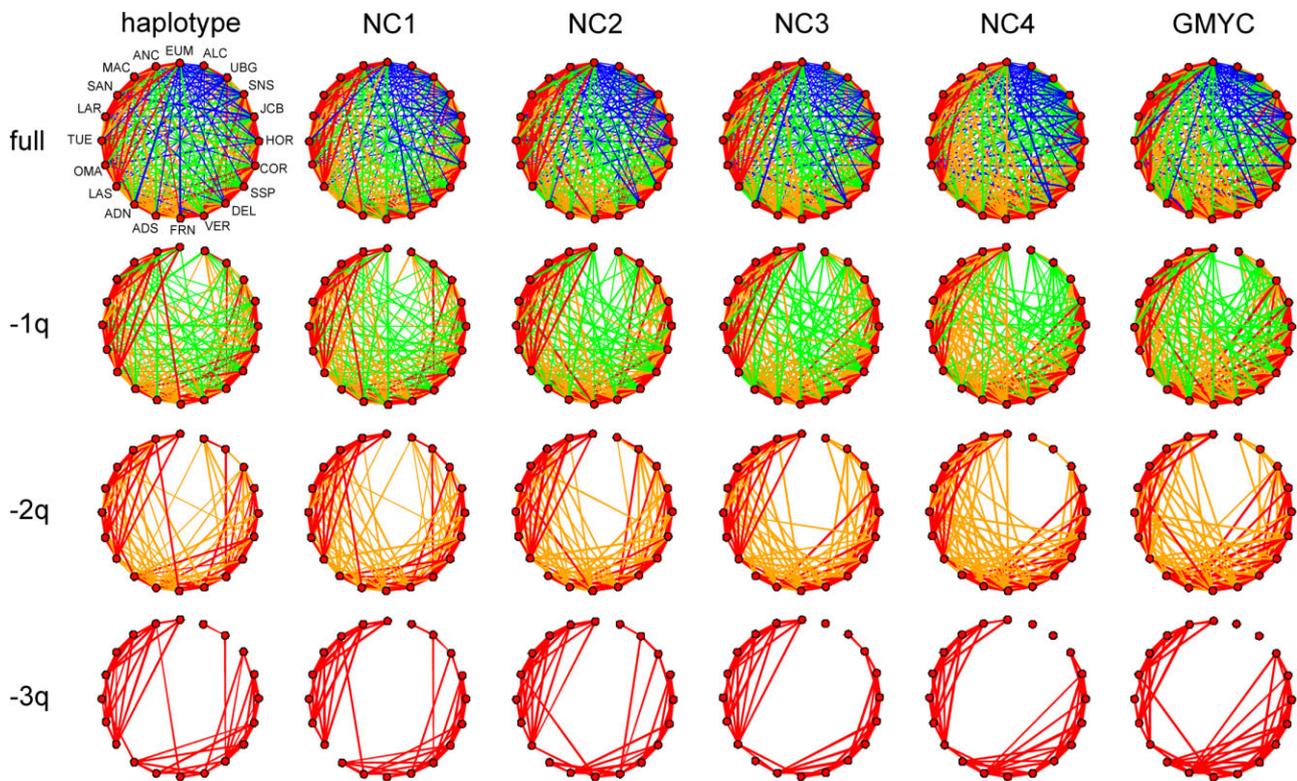
Figure 2 Assemblage networks at multiple hierarchical levels, i.e. haplotype, NC1-NC4 (nested clades 1 to 4), and GMYC (Generalized Mixed Yule-Coalescent groups). Nodes represent the 20 leaf beetle assemblages and edges represent similarity. Edge colour corresponds to the first (blue), second (green), third (orange) and fourth (red) quartiles of assemblage similarity at each hierarchical level. Fully connected networks were pruned by sequentially removing the edges corresponding to similarities in the first (-1q), second (-2q) and third (-3q) quartiles, revealing extremely similar topologies at all levels but at increasingly higher similarities from the haplotypes to the GMYC level.

measure of how much the network is compartmentalized, and the degree of nodes, which is the number of links of a node with other nodes. The expectation under self-similarity of assemblage variation across hierarchical levels is that (i) modularity remains constant at all hierarchical levels when removing the same fraction of edges (i.e. first, second and third quartiles of similarities), and (ii) the correlation between vectors describing the nodes' degrees at different hierarchical levels is high. The sequential removal of edges from the graphs showed that network topologies were very similar across hierarchical levels (Fig. 2). Removing the edges corresponding to the first quartile of similarities (i.e. the 25% lower values) yielded modularity values for each hierarchical level ranging between 0.11 and 0.13, with a mean = 0.12 and SD = 0.01; removing also the second quartile yielded modularity values ranging from 0.22 to 0.27 (mean = 0.25, SD = 0.02); and removing the third quartile yielded modularity values ranging from 0.44 to 0.48 (mean = 0.44, SD = 0.01). Hence, the networks are equally compartmentalized at each hierarchical level when the differences in similarity across levels are controlled for. Furthermore, the node degrees (the number of links of each node) were highly correlated across hierarchical levels, as evident from the relationship between node degrees of the haplotype and GMYC level networks after removing the first ($r^2 = 0.72$, $F_{1,18} = 45.2$,

$P < 0.0001$), second ($r^2 = 0.70$, $F_{1,18} = 42.7$, $P < 0.0001$), and third quartiles ($r^2 = 0.59$, $F_{1,18} = 25.5$, $P < 0.0001$), respectively.

Assemblage similarity at the GMYC level was predicted from assemblage similarity at the haplotype level with high accuracy (Fig. 3, $r^2 = 0.75$, Mantel $P < 0.0001$). We also found a positive correlation between within-species haplotype similarity and assemblage similarity in species composition (i.e. the SGDC at the beta diversity level) in 29 out of 30 species (Pearson r ranged between -0.18 and 0.88; mean $r^2 = 0.17$, SE = 0.03). The influence of niche-based processes on species turnover across assemblages was assessed from the residuals of the regression between assemblage similarity at GMYC and haplotype level similarities (size of the circles are proportional to climatic distance in Fig. 3a). When these residuals were correlated against climatic distances this correlation was found to be significant but weak ($r^2 = 0.05$, Mantel $P = 0.0059$), indicating a negligible contribution of climatic factors to species turnover that is independent of the neutral haplotype variation.

DISCUSSION

The proposed framework has proven useful to discern the effects of neutral and non-neutral processes on biodiversity. The set of results of our analyses allows disentangling neutral and
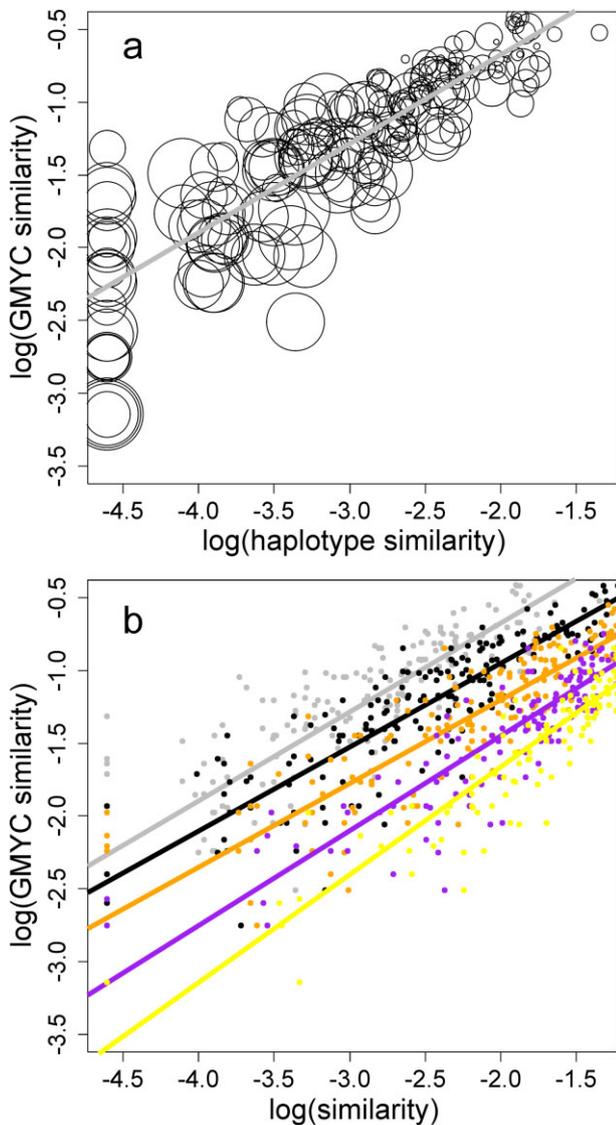
Figure 3 Predictability of assemblage similarity at GMYC level (Generalized Mixed Yule-Coalescent groups) from lower hierarchical levels (i.e. haplotype and nested clades 1 to 4, NC1-NC4). (a) Log-log relationship between pairwise similarity of two assemblages at the GMYC and haplotype levels. The size of circles is proportional to climatic distance between the two assemblages, but no marked relationship between the residuals of this regression and climatic distance ($r^2 = 0.05$, Mantel $P = 0.0059$) was observed. (b) Log-log relationship between assemblage similarity at the GMYC level and similarity at lower levels (grey = haplotype, black = NC1, orange = NC2, purple = NC3, yellow = NC4).

niche processes because the neutral distributions of mtDNA haplotypes provide a benchmark for stochastic dispersal processes. Tajima's tests showed that the evolution of mtDNA haplotypes was consistent with neutral evolution on the vast majority of species, and when it was not, the tests suggested purifying (i.e. the removal of disadvantageous mutations) rather than balancing selection that would have been indicative of

adaptations to the environment. The distribution of Tajima's D across species was not significantly different from a normal distribution centred at slightly negative values, suggesting again that selection had a weak effect on genetic variation. Therefore, considering (i) the high intra-specific and intra-populational mtDNA variation not expected under strong selection, (ii) the general consistency of mtDNA evolution with neutral dynamics, and (iii) the fact that any deviations from neutrality were biased towards purifying instead of balancing selection, we consider extremely unlikely that haplotype distributions could be linked to particular haplotype-specific environmental niches.

Given that haplotype distributions are neutral, we can predict that if assemblage similarity at the species level were associated with niche-based factors such as climate, the fractal geometry of clade ranges would be lost, as haplotype and species ranges would be controlled by different processes. Thus, under these non-neutral circumstances the extreme regularity in the slope of similarity decay with spatial distance and the regular shift across hierarchical levels in the genealogy would not be expected. Moreover, the deviations from the perfect relationship between haplotype and species level similarities would be predicted to correlate with climatic differences due to the effects of climatic niche at the species level. However, we found that the residuals of the species-haplotype relationship showed weak correlations with climatic differences between sites, and that the lineage ranges displayed a fractal geometry, with assemblage similarity decaying at similar rates at all hierarchical levels and initial similarity regularly shifting from level to level. The only alternative explanation we see for niche-based processes generating this fractal pattern would imply that species climatic niches were further fractally subdivided into nested haplotype-specific niches, which is neither supported by our empirical test of neutrality of mtDNA evolution, nor assumed under current niche theories (Hutchinson, 1957; Leibold, 1995). In sum, the most parsimonious conclusion is that neutral processes are the major drivers of the presented macroecological patterns.

It could be argued that the self-similarity of patterns at multiple hierarchical levels is an inevitable consequence of the nestedness of lower-level entities' ranges within the higher-level entities' ranges, but this is not the case. We have already shown that the distance decay of similarity at haplotype and species level can be decoupled (Baselga *et al.*, 2013), and contrary to the findings of lower similarity at the haplotype levels, the reverse may be the case under various scenarios of spatial distribution of the haplotypes within the ranges of species. In addition, alternative distributions of haplotype variation within the species ranges do not produce the congruent patterns of distance decay. For example, simulating haplotype distributions to recreate the potential maximum and minimum overlap in haplotype ranges given the empirical species distributions, or distributing the haplotypes randomly within the empirical species ranges, produces entirely different patterns (Appendix 2).

The existence of a fractal pattern emerging across hierarchical levels provides a temporal (historical) dimension of biodiversity

variation by defining various stages of lineage history in the form of nested groups that include more genetic divergence at each level. Similar results were observed in European aquatic beetles (Baselga *et al.*, 2013). Self-similarity of patterns of assemblage variation was further supported here by the fact that network modularity remained constant and topology of networks was very similar across levels for pruned networks. This provides support for the self-similarity of assemblage variation across hierarchical levels from a spatially-explicit network-based analysis, in addition to the general statistical correlations. The fractal patterns observed both in Iberian leaf beetles and previously in European water beetles (Baselga *et al.*, 2013) suggest a general macroecological regime. Despite ecological differences and great evolutionary distances of both lineages, we find remarkable similarities in the patterns of variation in assemblage composition. First, the log-log regressions on correlations of hierarchical level and numbers of entities were similar, with a slope of −1.20 in leaf beetles vs. −1.24 in water beetles. Second, the slopes of log-log regressions on the initial and mean assemblage similarity against hierarchical level were 0.91 vs. 0.90, and 0.78 vs. 0.82, respectively. This result shows that the fractal geometry is basically the same in both unrelated groups. However, both groups differed in the slopes of the distance decay curves, which were steeper in leaf beetles (−0.0026 and −0.0038 at the haplotype and GMYC levels) than the water beetles (−0.00067 and −0.00081) with halving distances of 181 and 264 km at the haplotype and GMYC levels, versus 854 and 1039 km, respectively. Hence, in both systems the genealogies based on the *cox1* gene and their patterns of distribution are very similar leading to the same fractal geometry, but they differ greatly in spatial scales likely depending on dispersal rate. Dispersal is expected to be high in aquatic beetles because of the need for movement between fairly short-lived water bodies, according to the habitat stability hypothesis (Ribera *et al.*, 2003), while association with vegetation in leaf beetles may not require long-range movement for population persistence. These effects may be exacerbated by long term relative climatic stability of the Iberian Peninsula (Ribera & Vogler, 2004) compared to the pattern across all of Europe including northern regions highly affected by glaciations in the aquatic beetle assemblages.

To conclude, our results suggest that the multi-hierarchical correlations of community similarity may be a general outcome across taxa and geographical realms, pointing to neutral dynamics as a widely applicable principle of biodiversity and bridging between the neutral theory of evolution (Kimura, 1983) and the neutral theory of biodiversity (Hubbell, 2001). These theories attribute the biological diversity to genetic drift of allelic variants in populations and individual-based ecological drift in metacommunities, respectively. Combined, they are consistent with a framework of stochastic processes of lineage branching and distance-limited dispersal that is captured by the community-scale sequence data and multi-hierarchical analysis. However, it should be noted that the dominant effect of neutral dynamics is probably scale dependent. At smallest spatial scales, individuals are not dispersal-limited and therefore niche processes should gain rel-evance. At largest spatial scales, neutral dynamics would be constrained by major biogeographic and ecological boundaries, while within a biome or ecoregion, neutral processes seem to be the major drivers of assemblage variation. The multi-hierarchical macroecology approach can discern between neutral and non-neutral dynamics, as it provides the means to compare empirical patterns emerging across hierarchical levels with contrasting predictions derived from the neutral and niche paradigms.

REFERENCES

Albu, M., Min, X.J., Hickey, D. & Golding, B. (2008) Uncorrected nucleotide bias in mtDNA can mimic the effects of positive darwinian selection. *Molecular Biology and Evolution*, **25**, 2521–2524.

Avise, J.C. (1994) *Molecular markers, natural history and evolution*. Chapman & Hall, New York.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.

Baselga, A. & Orme, C.D.L. (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**, 808–812.

Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P.G., Monaghan, M.T. & Vogler, A.P. (2013) Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications*, **4**, 1892, 1–7. doi: 10.1038/ncomms2881.

Bazin, E., Glemin, S. & Galtier, N. (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570–572.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

Britton, T., Anderson, C.L., Jacquet, D., Lundqvist, S. & Bremer, K. (2007) Estimating divergence times in large phylogenetic trees. *Systematic Biology*, **56**, 741–752.

Brown, J.H. (1995) *Macroecology*. University of Chicago Press, Chicago, IL.

Butts, C.T., Handcock, M.S. & Hunter, D.R. (2013) network: Classes for Relational Data. R package version 1.7.2. Irvine, CA. Available at: http://statnet.org/.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems*, **1695**, 1–9.

Drakare, S., Lennon, J.J. & Hillebrand, H. (2006) The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecology Letters*, **9**, 215–227.

Ezard, T., Fujisawa, T. & Barraclough, T.G. (2009) splits: SPecies' LImits by Threshold Statistics. R package version 1.0-14/r31. Available at: http://R-Forge.R-project.org/projects/splits/.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Gómez-Zurita, J., Hunt, T. & Vogler, A.P. (2008) Multilocus ribosomal RNA phylogeny of the leaf beetles (Chrysomelidae). *Cladistics*, **24**, 34–50.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ.

Hutchinson, G.E. (1957) Population studies – animal ecology and demography – Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

Karl, S.A., Toonen, R.J., Grant, W.S. & Bowen, B.W. (2012) Common misconceptions in molecular ecology: echoes of the modern synthesis. *Molecular Ecology*, **21**, 4171–4189.

Kimura, M. (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.

Laroche, F., Jarne, P., Lamy, T., David, P. & Massol, F. (2015) A neutral theory for interpreting correlations between species and genetic diversity in communities. *American Naturalist*, **185**, 59–69.

Leibold, M.A. (1995) The niche concept revisited – Mechanistic models and community context. *Ecology*, **76**, 1371–1382.

Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology*, **70**, 966–979.

McGill, B.J. (2010) Towards a unification of unified theories of biodiversity. *Ecology Letters*, **13**, 627–642.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F.L., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, **10**, 995–1015.

Mulligan, C.J., Kitchen, A. & Miyamoto, M.M. (2006) Comment on 'Population size does not influence mitochondrial genetic diversity in animals. *Science*, **314**, 1390.

Nei, M., Suzuki, Y. & Nozawa, M. (2010) The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics*, **11**, 265–289.

Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**, 867–878.

Odat, N., Hellwig, F.H., Jetschke, G. & Fischer, M. (2010) On the relationship between plant species diversity and genetic diversity of *Plantago lanceolata* (Plantaginaceae) within and between grassland communities. *Journal of Plant Ecology*, **3**, 41–48.

Oksanen, J., Blanchet, G., Kindt, R., Minchin, P.R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2011) vegan: Community Ecology Package. R package version 2.0-2, Available at http://cran.r-project.org/.

Panchal, M. (2007) The automation of nested clade phylogeographic analysis. *Bioinformatics*, **23**, 509–510.

Papadopoulou, A., Anastasiou, I., Spagopoulou, F., Stalimerou, M., Terzopoulou, S., Legakis, A. & Vogler, A.P. (2011) Testing the species-genetic diversity correlation in the Aegean archipelago: toward a haplotype-based macroecology? *American Naturalist*, **178**, 560–560.

Paradis, E. (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**, 419–420.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M. (2011) *Ecological niches and geographic distributions*. Princeton University Press, Princeton, NJ.

Petitpierre, E. (2000) *Coleoptera, chrysomelidae I*. Museo Nacional de Ciencias Naturales, Madrid.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Ribera, I. & Vogler, A.P. (2004) Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*, **13**, 179–193.

Ribera, I., Foster, G.N. & Vogler, A.P. (2003) Does habitat use explain large scale species richness patterns of aquatic beetles in Europe? *Ecography*, **26**, 145–152.

Rosindell, J., Hubbell, S.P. & Etienne, R.S. (2011) The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology & Evolution*, **26**, 340–348.

Sei, M., Lang, B.K. & Berg, D.J. (2009) Genetic and community similarities are correlated in endemic-rich springs of the northern Chihuahuan Desert. *Global Ecology and Biogeography*, **18**, 192–201.

Slatkin, M. (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics*, **16**, 393–430.

Soberón, J. & Nakamura, M. (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19644–19650.

Tajima, F. (1989) Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, **123**, 585–595.

Templeton, A.R., Crandall, K.A. & Sing, C.F. (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.

Vellend, M. (2003) Island biogeography of genes and species. *American Naturalist*, **162**, 358–365.

Vellend, M. & Geber, M.A. (2005) Connections between species diversity and genetic diversity. *Ecology Letters*, **8**, 767–781.

Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S.W. & Garant, D. (2014) Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology*, **23**, 2890–2901.

Warchalowski, A. (2003) *Chrysomelidae. The leaf-beetles of Europe and the Mediterranean area*. Natura optima dux Foundation, Warszawa.

Willig, M.R., Kaufman, D.M. & Stevens, R.D. (2003) Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology Evolution and Systematics*, **34**, 273–309.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 Supplemetary tables S1 (Spatial position and mean climatic conditions of the 20 localities) and S2 (variance partitioning showing the unique contribution of spatial and climatic distances to the explained variance in assemblage similarity).

Appendix S2 Supplemetary results showing the independence of haplotype-level and species-level assemblage similarity.

BIOSKETCHES

Andrés Baselga is interested in several biodiversity-related disciplines (from basic taxonomy to phylogenetics and macroecology), with special emphasis on beetles.

Carola Gómez-Rodríguez is interested in freshwater ecology, phylogenetics and macroecology.

Alfried P. Vogler is interested in the phylogeny and evolution of beetles.

Editor: José Alexandre Diniz-Filho