

ARTICLE

Received 15 Nov 2012 | Accepted 12 Apr 2013 | Published 21 May 2013

DOI: 10.1038/ncomms2881

Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels

Andrés Baselga¹, Tomochika Fujisawa^{2,3,†}, Alexandra Crampton-Platt^{2,3}, Johannes Bergsten^{2,3,†}, Peter G. Foster², Michael T. Monaghan^{2,3,4} & Alfried P. Vogler^{2,3}

A correlation of species and genetic diversity has been proposed but not uniformly supported. Large-scale DNA barcoding provides qualitatively novel data to test for correlations across hierarchical levels (genes, genealogies and species), and may help to unveil the underlying processes. Here we analyse sequence variation in communities of aquatic beetles across Europe (>5,000 individuals) to test for self-similarity of beta diversity patterns at multiple hierarchical levels. We show that community similarity at all levels decreases exponentially with geographic distance, and initial similarity is correlated with the lineage age, consistent with a molecular clock. Log-log correlations between lineage age, number of lineages, and range sizes, reveal a fractal geometry in time and space, indicating a spatio-temporal continuum of biodiversity across scales. Simulations show that these findings mirror dispersal-constrained models of haplotype distributions. These novel macroecological patterns may be explained by neutral evolutionary processes, acting continuously over time to produce multi-scale regularities of biodiversity.

¹Departamento de Zoología, Facultad de Biología, Universidad de Santiago de Compostela, c/ Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain. ²Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK. ³Division of Biology, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK. ⁴Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Mueggelseedamm 301, 12587 Berlin, Germany. † Present addresses: Department of Zoology, Kyoto University, Sakyo, Kyoto, 606-8502, Japan (T.F.); Department of Entomology, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden (J.B.). Correspondence and requests for materials should be addressed to A.B. (email: andres.baselga@usc.es).

The task of preserving global biodiversity requires both theoretical and empirical understanding of the processes that underpin patterns of species distribution at multiple spatio-temporal scales^{1–3}. Biodiversity is a hierarchical concept that embraces multiple levels of biological organization including genes, species and communities⁴. Ideally, models that explain the distribution of diversity should thus invoke processes that apply equally well across various temporal and spatial scales. Early attempts to examine patterns across the biological hierarchy have focused on the link between genetic and species diversity, and have described a correlation between local species richness and intra-specific genetic diversity of species (the species-genetic diversity correlation)⁵. Additional work has extended the species-genetic diversity correlation to other facets of diversity (that is, beta diversity), revealing a correlated distance decay of community similarity on either level⁶. These findings potentially link patterns of biodiversity at different scales via stochastic processes of migration, speciation/mutation and extinction^{7,8}.

An integrated spatio-temporal framework of biodiversity applicable at multiple levels of the biological hierarchy also is emerging from neutral theories of diversity that assume stochastic but spatially limited dispersal as the primary driver of local diversity^{9,10}. In neutral models, individuals are close to their parents at birth and follow a trajectory of limited dispersal, giving rise to spatial patterns of clumping that may explain macroecological properties at larger scales¹¹. Consequently, neutral models have been developed assuming that clumping arises at all spatial scales and hence the distribution of species is self-similar, that is, the proportion of species in an area that are found in a subsection of the area is independent of the area size itself¹².

One of the major challenges to empirical studies across multiple levels of the biological hierarchy is the absence of empirical data. Large-scale DNA sequencing of entire species assemblages provides the means to quantify species and genetic diversity collectively for the members of a given community^{6,13}, which can be used to test for similarity of patterns across biological levels and broad spatial scales.

Here we study genotypic diversity in mitochondrial DNA (mtDNA) in assemblages of aquatic beetles across Europe, spanning c. 30 degrees of latitude. Natural assemblages of aquatic beetles are highly suitable for testing the predictions from dispersal-constrained theoretical models. They form discrete communities of a few dozen species in lakes, ponds and small streams, comprising several families, of which Dytiscidae (predaceous diving beetles) is dominant. Being dependent on open water bodies, dispersal requires discrete migration steps across unsuitable terrestrial habitat. Therefore, using mtDNA sequences for entire local water beetle communities, we investigate if patterns of spatial turnover show self-similarity at multiple hierarchical levels that is predicted under a neutral community ecology paradigm. Our results confirm this expectation, suggesting that a uniform process determines biodiversity patterns across various spatial and temporal scales.

Results

Distance decay at species and genetic levels. Surveys of 23 regional assemblages (Fig. 1, Supplementary Table S1) yielded a total of 5,066 sequences and 2,183 unique haplotypes. The Generalized Mixed Yule-Coalescent (GMYC) procedure establishes independently evolving entities¹³ that broadly represent the species level¹⁴. The method grouped total haplotype variation into 274 putative species. The number of species in an assemblage ranged from 13–98 (mean = 38, s.d. = 20). We first used the species composition of assemblages to test for the decay of assemblage similarity with geographic distance that is expected

under a dispersal-constrained model, given continuous dispersion since the origin of lineages¹⁵. The decline of assemblage similarity with spatial distance was characterized by a simple negative exponential function with halving-distance $D_{1/2} = 1,039$ km and initial similarity (intercept with the y-axis) of $y = 0.64$ (Fig. 1a, Supplementary Table S2). We then tested for self-similarity of this pattern at the genotype level, as predicted from a common process that is uniform over time. The haplotype level also showed exponential distance decay but at a much lower initial similarity ($y = 0.16$) and the decay was slightly faster ($D_{1/2} = 854$ km) than at the species level (Fig. 1a).

The decay of similarity at both levels did not stem from the correlation between climatic differences and geographic distance. Given that the steepest decline occurs at shortest geographic distances, we assessed pairs of localities separated by < 1,500 km. Over these shorter distances, assemblage similarity was still significantly correlated with geographic distance, but correlation with climatic difference was negligible (Fig. 1c,d). Although other aspects of habitat type or climate not considered here may differ over distances of 1,500 km, our results indicate that the divergence in community composition at this scale can be explained by spatial parameters alone.

Distance decay at multiple hierarchical levels. We examined the temporal component of self-similarity using the evolutionary information contained in the sequencing data. By reconstructing genealogies within species¹⁶ we can identify the intermediate hierarchical levels between those of species (highest) and haplotypes (lowest). We first established a hierarchical grouping scheme from the mtDNA variation using Templeton's (1987) algorithm to generate nested clades of n steps according to the number of mutational steps between increasingly more distant haplotypes incorporated into a network¹⁷. Based on the 2,183 unique haplotypes, the method defined 268 networks, closely matching the number and extent of GMYC groups, as seen elsewhere¹⁸. Each network contained genealogies of up to five hierarchical levels. The five levels represent a convenient subdivision of lineage history, with the number of groups at each level representing the lineage size at past time intervals. Based on the criteria of the nesting algorithm¹⁷, the 268 networks (representing the five-step level, hereafter referred to as NC5; Supplementary Table S2) were further subdivided into 316, 479, 838 and 1,404 nested groups (NC4, NC3, NC2 and NC1), according to the decreasing number of mutational divergences within each level. When plotting the distance decay of community similarity at each nesting level (as a proxy for clade age), similarity decreased significantly with geographic distance and the shift from one level to the next higher level was largely uniform (Fig. 2, see intercepts a in Supplementary Table S2). The regularity of the shift between hierarchical levels suggests a uniform underlying evolutionary process.

These analyses of distance decay of similarity were performed with incidence-based methods, but a newly developed index¹⁹ allowed us to assess similarity based on absolute abundance of each entity (from haplotypes to species) among pairs of sites. The abundance-based analysis yielded almost identical results (Supplementary Fig. S1–S2). This finding demonstrates that self-similarity of distance decay at different hierarchical levels also applies to an additional level of complexity concerning species' abundances in local communities, in addition to incidences of species. From a practical point of view, this suggests that our sampling is a good representation of the studied assemblages because, for abundance data to reveal this pattern, sampling needs to be more detailed to capture the relative proportion of species than to simply record the presence of a species.

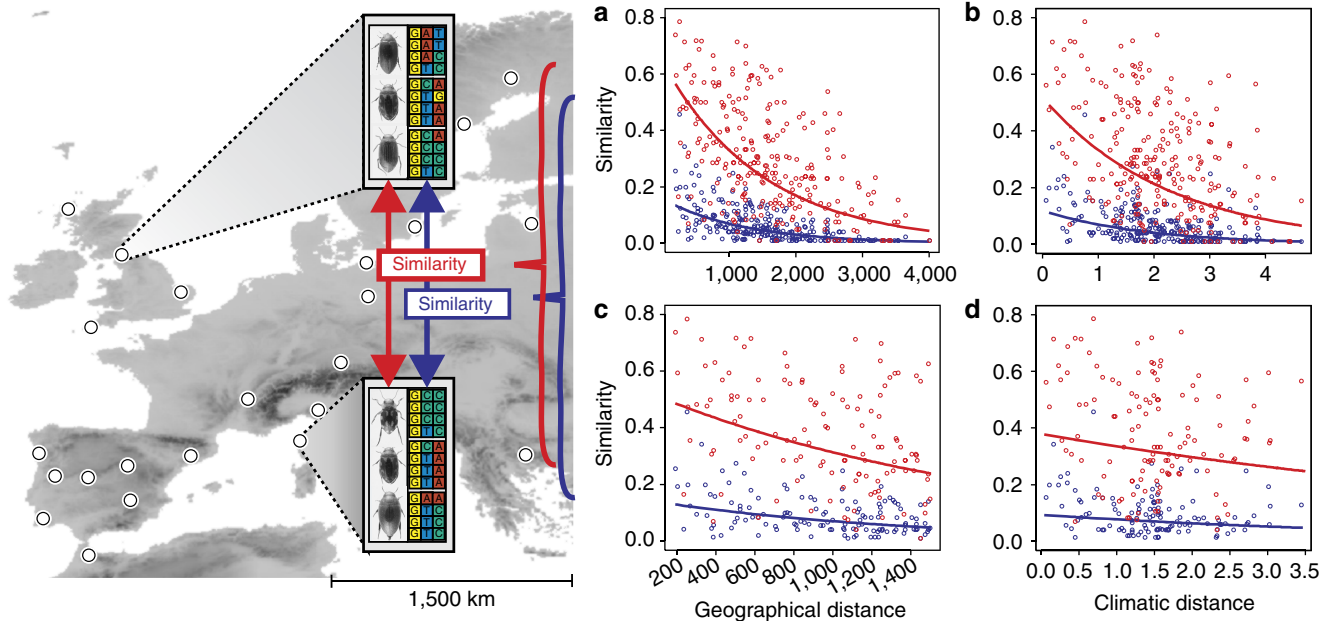


Figure 1 | Distance decay of similarity in water beetle assemblages across Europe. Fitted curves (species level, red; haplotype level, blue) are exponential decay functions. **(a)** Decay of similarity against geographic distance across Europe: exponential decay curves were fitted at the species ($r^2 = 0.32$, Mantel test $P < 0.001$, $n = 253$) and haplotype ($r^2 = 0.46$, $P < 0.001$) levels. **(b)** Distance decay of similarity against climatic distance across Europe at the species ($r^2 = 0.18$, $P < 0.001$, $n = 253$) and haplotype ($r^2 = 0.25$, $P < 0.001$) levels. **(c)** Distance decay of similarity against geographic distance for the subset of sites separated by $< 1,500$ km at the species ($r^2 = 0.09$, $P = 0.001$, $n = 127$) and haplotype ($r^2 = 0.13$, $P = 0.001$) levels. **(d)** Distance decay of similarity against climatic distance for the subset of sites separated by $< 1,500$ km at the species ($r^2 = 0.02$, $P = 0.07$, $n = 127$) and haplotype ($r^2 = 0.03$, $P = 0.01$) levels.

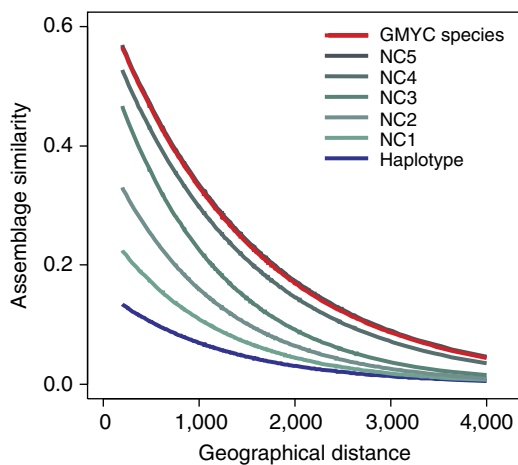


Figure 2 | Distance decay of similarity of assemblages across Europe at multiple levels. The colour of curves identifies the species (red), haplotype (blue) and intermediate nested-clade (green tones) levels. The fitted lines are exponential decay functions ($0.46 > r^2 > 0.31$, Mantel test $P < 0.001$, $n = 253$).

The fractal geometry of biotic ranges. Under neutral dynamics the time since the common ancestor is negatively correlated with the number of lineages, and this correlation is predicted to be linear in a log–log plot indicating fractality⁸. Accordingly, our analysis showed a very strict negative log–log linear correlation (Fig. 3b) with slope -1.24 . This exponential increment in the number of entities at each hierarchical level is consistent with a fractal pattern. A fractal pattern was also found in the spatial geometry of distribution ranges. The number of entities at each nesting level was tightly correlated with the entities’ range sizes

(estimated as the number of occupied sites). Range size showed a very tight correlation with the number of entities at each nesting level in a log–log plot (Fig. 3c). Moreover, a striking feature was the great regularity by which the mean similarity and the intercepts of distance decay curves (Fig. 3d) declined with the n -step level. This regularity implies that lower-level ranges are uniformly distributed within higher-level ranges throughout the whole phylogenetic history between haplotypes and species. As the n -step level provides an approximate time axis for lineage diversification due to the broadly clock-like rate of change in mtDNA²⁰, the correlation with community similarity implies a spatio-temporally continuous process that operates across all hierarchical levels to produce self-similarity at any scale (Fig. 3a). These patterns do not simply reflect a constraint of haplotype ranges upon species ranges, as alternative situations are conceivable under which the similarity is higher at lower hierarchical levels (Supplementary Fig. S3).

Simulations of neutral against niche-based processes. We used simulations to test whether our observed empirical patterns favour dispersal-constrained over niche-constrained scenarios. Communities were simulated that either followed a neutral trajectory of stochastic dispersion and lineage diversification, or that experienced constraints in the dispersion of species by the environment, reflecting niche-constrained conditions (see details in Materials and Methods). The latter disconnects species ranges constrained by niches from the unconstrained dispersion of individuals within the species ranges. Simulated distributions showed that community similarity was correlated with geographic distance at both the species and haplotype levels (Fig. 4), suggesting that migration is at least partially responsible for range size in all cases. Nonetheless, only under neutral conditions were the shapes of the haplotype and species decay curves consistently

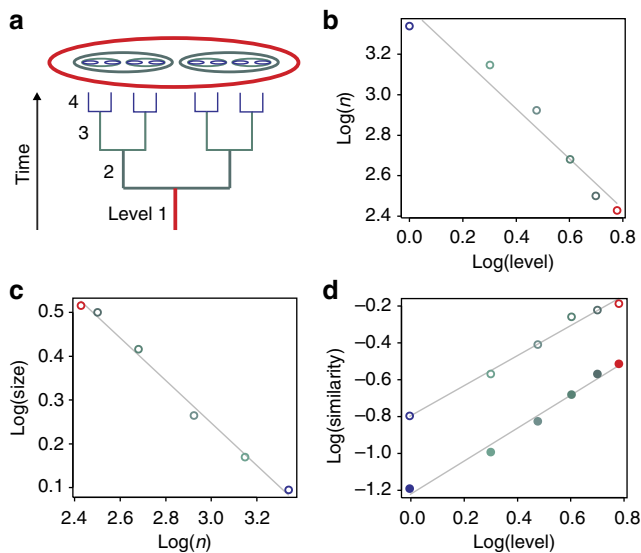


Figure 3 | Evidence for fractal patterns of biodiversity in time and space. (a) Scheme illustrating how neutral mutations cause a regular branching of lineages through time (molecular clock), and because dispersal is the parameter controlling clade ranges, younger lineages nested in older lineages will have smaller ranges that are regularly nested in older ranges. (b) Observed relationship (linear regression $r^2 = 0.96$, $F_{1,4} = 87.1$, $P < 0.001$, $n = 6$) between the number of clades and hierarchical nesting level from haplotypes (level 1) to five-step networks (level 6); (c) Relationship between range size and number of clades ($r^2 = 0.99$, $F_{1,4} = 647.4$, $P < 0.001$, $n = 6$); (d) Relationship between initial similarity (open circles) and hierarchical n -step level ($r^2 = 0.99$, $F_{1,4} = 364.7$, $P < 0.001$, $n = 6$), and between mean similarity (closed circles) and hierarchical n -step level ($r^2 = 0.99$, $F_{1,4} = 407.4$, $P < 0.001$, $n = 6$). Red and blue dots represent the species and haplotype levels, respectively, while green tones represent the intermediate nested-clade levels. Grey lines are the fitted functions.

self-similar; the niche-based model produced increasingly flatter slopes at the haplotype level with reduced community sizes compared with the species level (Fig. 4, Supplementary Table S3). The correlation of community variation with climatic distance, rather than spatial distance, was weak at both species and haplotype levels (Supplementary Fig. S4, Supplementary Table S3). Taken together, the dispersal-constrained scenario mirrored the empirical data for European water beetles, whereas the niche-constrained scenario did not.

Discussion

The regular shift of the empirical distance decay curves below the species level can be interpreted simultaneously in temporal and spatial dimensions. Spatial distributions of species appear as larger-scale extensions of ranges of lower-level clades (haplotypes and nested clades), while entities at progressively lower nesting levels comprise a set of more recently diversified subordinate lineages. The observed fractal geometry of clade ranges supports the idea that the same neutral processes are underlying the haplotype and species ranges, and the two levels differ in spatial scale only due to the difference in length of time for dispersal (clade age).

Existing theoretical models predicting fractal scaling properties of species distributions are based on purely statistical properties^{12,21} and do not incorporate a mechanistic foundation. Yet, they are consistent with the assumption of stochastic dispersal, birth and death of demographically equivalent individuals of neutral community models⁸ that have been shown to produce

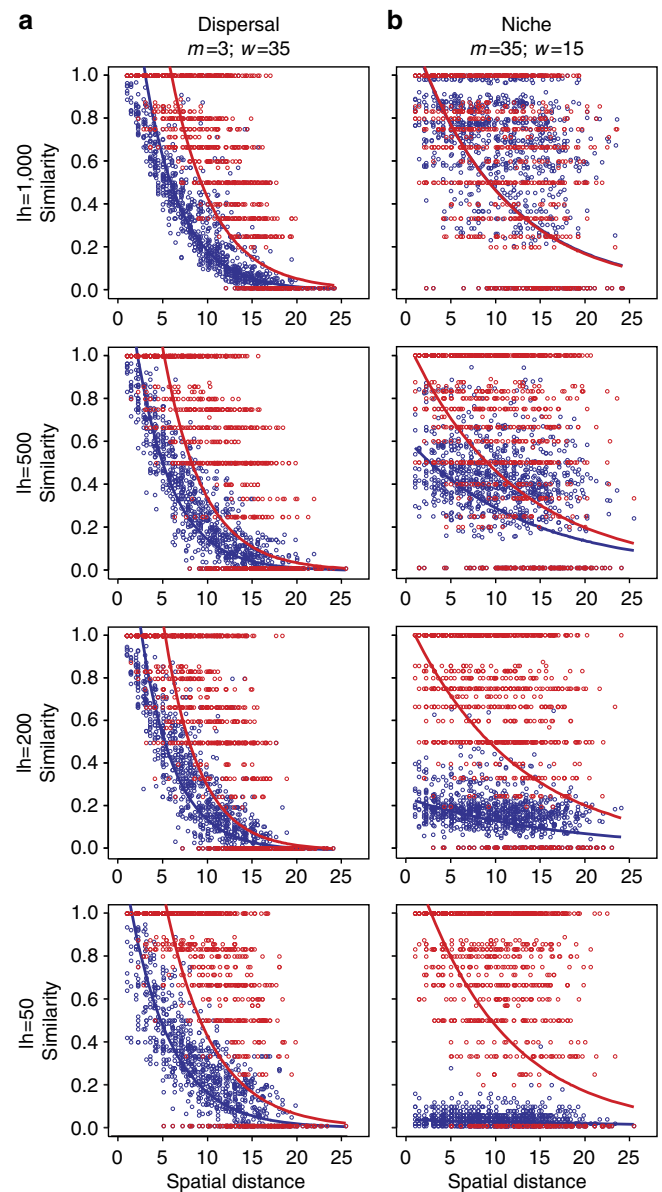


Figure 4 | Simulations under dispersal- and niche-constrained dynamics. Plots show the decay of similarity with spatial distance for each type of simulation (haplotypes: blue circles and lines; species: red circles and lines) and for different 'carrying capacities' in local assemblages (lh). (a) Under the dispersal-constrained scenario, correlation between similarity and spatial distance was always higher than for the niche-constrained scenario. Slopes for the species and haplotype levels were similar for a wide range of carrying capacities when dispersal-constrained, while intercepts were different. (b) Under the niche scenario the slope and intercepts are the same for the species and haplotype levels when the local number of haplotypes is not limited (lh = 1,000), while slopes are always flatter for the haplotype level than for the species level when the local number of haplotypes is limited.

established species-level macroecological patterns^{8,9}. The latter models also add a component of neutral evolution, whereby community level abundances determine speciation and extinction probabilities, and, therefore, the longevity of lineages. These differences among lineages determined by metacommunity parameters result in phylogenies that are self-similar at any scale of evolutionary time⁸. Our empirical results not only confirm this pattern of fractal branching in time, but also detect

self-similarity in the spatial dimension from the fractal geometry of clade ranges. These findings link the origin of new lineages over time with the spatial structure of lineage ranges across all hierarchical levels, presumably down to the individual. Finally, if the distance decay is explained by individual-based dispersal, this predicts that adjacent communities also are similar in abundance of the member species⁸. This was confirmed here (Supplementary Fig. S1–S2) and further supports the hypothesis of stochastic dispersal.

Potentially, non-neutral processes could generate these patterns if environmental factors controlling biotic ranges were spatially structured. There is no evidence for this in the current data, as over a distance of <1,500 km, for which the correlation between spatial distance and climatic difference is very low, environmental differences are negligible to explain the biotic turnover at species and genetic levels (Supplementary Table S2). Our model does not preclude that niche differentiation takes place during the process of dispersal-limited divergence, but it establishes that this is not fundamental to generating the statistical pattern. In addition, haplotype divergences in presumably neutral mtDNA are likely to reflect restricted gene flow among populations in the absence of selection. As the pattern of spatial turnover in these neutral entities is mirrored by the pattern at the species level, this correlation would either be due to complicated dependencies of the (neutral) mtDNA haplotype pattern to follow the (non-neutral) species level pattern, or species-level patterns would also be neutral. As the latter explanation requires only one simple mechanism (stochastic but limited dispersal), the neutral model is a more parsimonious explanation for the reported patterns. The use of neutral markers, such as mtDNA, is a prerequisite for articulating these processes, as the aim was to assess if species ranges, which may be neutral or non-neutral, mirror neutral haplotype ranges. Finally, whereas statistical patterns do not directly inform on the underlying process, we also used simulations to explore mechanistically whether neutral or niche-based processes are able to generate the observed patterns. These simulations suggest that only neutral processes yield the observed patterns, as niche filtering decouples the behaviour of haplotype and species decay curves (Fig. 4, Supplementary Fig. S4).

An integrated spatio-temporal framework of biodiversity can greatly benefit from analyses at the DNA level, linking stochastic population- and species-level processes under common assumptions of migration, speciation/mutation and extinction^{5,8}. In future, genetic surveys may be supplemented with abundance data to make an explicit link to mechanistic individual-based models of neutral community ecology²². A next step will be to assess the generality of the observed patterns in a broad range of taxa. MtDNA haplotype data are generated increasingly for assemblages and entire ecosystems, in particular as part of global DNA barcoding efforts²³. If evidence for neutrality is obtained generally, this provides a scaffold for biodiversity surveys and comparative analysis. One important consequence of the fractal relationships at multiple levels is that the completeness of sampling at lower genetic levels (for example, total number of haplotypes) can be assessed from regressions through the more completely sampled higher hierarchical groups (see Fig. 3b). Likewise, the slope of the relationship of range size and hierarchical level (Fig. 3c) may be used as a benchmark measure for comparing the spatial scale of community differentiation in various taxa, geographic regions or ecological settings. This slope reflects mean historical dispersal in assemblages, as it measures how neutral ranges expand over time, and hence the rate of past geographical shifts and disturbance of assemblages and ecosystems. Global change of Earth's biota is increasingly driven by changed dispersal rates due to landscape fragmentation, moving

climate zones, metapopulation dynamics after ecosystem disturbance and other factors²⁴. A dispersal-based framework of biodiversity that integrates across spatial and temporal scales will reveal global change through its effect on the dispersal dynamics of species assemblages.

Methods

Sequence data and phylogenetic analysis. Each regional assemblage of aquatic beetles along a 4,500-km transect from northern Sweden to Morocco (see Fig. 1 and Supplementary Table S1) was composed of several localities within a <50-km radius from both lotic and lentic habitats. Samples were collected in 100% alcohol and identified in the field, to include a total of 204 recognizable species. At every unique locality within a region, all specimens up to a maximum of five individuals per field-identified species were used for DNA extraction. Field identifications were used to maximize the diversity and to obtain a good representation of the total variation, but these identifications were not further utilized. Hence the data reflect the evolutionary history of the target taxa as inferred from a single locus, but any inconsistencies that may result from gene tree–species tree incongruence presumably are avoided due to the very large number of lineages sampled and the relatively rare occurrence of incongruence²⁵. Note that with only the mtDNA sequenced, gene tree incongruence could only be assessed indirectly against the morphological species limits; where such incongruence with the mtDNA groups was evident, this usually affected only some portions of the species ranges²⁶, that is, any resulting error would affect the details of geographic turnover but not the discovery of species-level groups *per se*.

Genomic DNA was extracted from muscle tissue in the prothorax region with Wizard SV 96-well plates (Promega, UK). An 825 base pair region from the 3' end of mitochondrial cytochrome oxidase I was amplified with primers Jerry²⁷, Ron Inosine, Ron Dyt, Patty, and Pat Dyt²⁸. Amplification conditions used with Bioline Taq were 94 °C for 2 min, 35–40 cycles of 94 °C for 30 s, 53 °C for 60 s and 70 °C for 120 s, and a final extension of 70 °C for 10 min. PCR products were cleaned with a 96-well Millipore multiscreen plate, and sequenced in both directions using ABI dye terminator sequencing. Approximately 8,000 individuals were processed. Trace files of forward and reverse strands were analysed with the base calling software Phred²⁹ under default parameters. Only sequences with >600 bps of Phrap score >20 were retained, resulting in >5,000 high-quality sequences. A small proportion (1.2%) of sequences with indels and in-frame stop codons were removed as putative pseudogenes. A subset of sequences for the subfamily Agabinae was already used in a previous study after manual editing²⁶, but the raw data were recompiled here using the automated editing, to make them compatible with all other data.

Before tree building, identical haplotypes were collapsed into a single sequence. Gene trees were constructed with RAXML 7.0.3 (ref. 30) under the GTR + G + I model selected by jModeltest³¹ on a reduced data set (one sequence from one taxonomic species; ~10% of the haplotype-collapsed data set). The maximum likelihood tree was made ultrametric with Pathd8 (ref. 32), which implements a mean path length method for establishing a molecular clock. The ultrametric trees were used for quantitative species delimitation with the GMYC method that establishes the point of transition from slow to faster branching rates of the gene tree expected at the species boundary¹⁴ using the R package *splits* (SPecies LImits by Threshold Statistics)³³. The GMYC analysis uses the mode of branching to separate intra-specific (coalescent) from inter-specific (Yule) variation¹⁴ and has been widely used to demarcate the species boundary. All genotype information was included in the geographic analysis after application of the GMYC procedure, by restoring ('uncollapse') the haplotype data removed before tree building using a custom R script. Minor uncertainty in the haplotype identity due to partially missing sequence data was resolved by randomly assigning the haplotypes to one of the fully sequenced genotypes.

Haplotype networks were created from the data set using TCS software³⁴ implemented in ANeCA v.1.2 (ref. 35). 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 analysis³⁶. 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. The procedure implements a stopping rule for the maximum level by which haplotypes are included into a single network based on the probability of encountering homoplastic changes in the connections of two haplotype groups. The large size of the data set required that the TCS software was applied separately to smaller subclades. We, therefore, performed the nested-clade analysis individually on 18 clades defined in a higher-level phylogenetic analysis of Dytiscidae³⁷. Custom R and Perl scripts were used for generating relevant input and output files for the various steps.

Statistical analysis for empirical patterns. Similarity of assemblages among regions (hereafter 'sites') was assessed at haplotype, nested clades (*n*-step levels) and GMYC (species) levels by means of the Simpson index of similarity, that is,

$1-\beta_{\text{sim}}$ (refs 38,39), using the R package *betapart*⁴⁰. To assess the relationship of assemblage similarity with spatial distance and environmental (climatic) differences, nonlinear regression was used to fit exponential decay curves expressed as $y = a \times e^{-bx}$, where y is similarity at distance x , a initial similarity and $-b$ the rate of distance decay. Curves were not forced through the intercept with the y-axis (geographic distance = 0). Spatial distance between sites was computed in km as the Euclidian distance between the centroids of regions. Climatic distance was computed as the Euclidian distance in a multidimensional space consisting of three topo-climatic variables, that is, mean altitude, annual mean temperature and annual precipitation⁴¹ that were standardized (mean = 0, s.d. = 1). These climate variables have proven to be highly correlated to diversity patterns across various taxonomic groups⁴², and were selected here because they yielded the highest correlations with the first three axes in a Principal Coordinate Analysis, that also included other variables as maximum temperature of the warmest month, minimum temperature of the coldest month, precipitation of driest quarter or precipitation of wettest quarter. The degree of association between biotic similarity and spatial or climatic distance was measured as Pearson correlation (r) after linearizing the decay curves through log-transformation. An arbitrary small quantity (0.01) was summed to values of similarity = 0 to avoid undefined values. Because of possible correlation at large spatial scales, calculations were performed for the full transect, as well as for a reduced data set of pairs of sites separated < 1,500 km in which climatic and geographic distance were not closely correlated. The significance of r values was assessed with Mantel tests using the R package *vegan*⁴³. Finally, to assess the robustness of results to variation in abundance, we repeated the aforementioned analyses using a measure of similarity accounting for balanced variation in abundance between localities but independent from abundance gradients: the balanced variation component of Bray–Curtis similarity¹⁹. This is analogous to Simpson similarity but for abundance data.

Simulations of metacommunities. Haplotype and species distribution ranges were modelled in a hypothetical landscape to test patterns of species and haplotype composition under alternative processes of stochastic dispersal and niche-based filtering. We were interested in comparing the observed macroecological patterns (which are defined by the presence/absence of clades in a set of localities) against comparable predictions derived from simplified simulations. Given the observed agreement between empirical incidence and abundance-based patterns, there is little gain in simulating the dispersal of individuals, and, therefore, none of the existing individual-based neutral models²² are applicable here. Instead, we simulated the spread of haplotypes and species ranges with time as a function of migration capacity and the origination of new haplotypes as a function of range sizes. Time and spatial extents of the simulations were arbitrary, but it should be noted that increasing time and spatial extents would be strictly equivalent to reducing the migration parameter. Therefore, our simulations allow exploring the effects of dispersal limitation and niche width in the patterns of haplotype and species turnover.

We first defined the landscape by a lattice of 30×30 cells, and set two climatic variables (for example, temperature and precipitation), which were partly correlated with spatial dimension x and y by assigning spatial coordinates of 1 to 30 to each cell, and setting values for climate variables to the respective spatial coordinates plus a random value between -10 and 10 . This sets the correlation between spatial coordinates and climatic variables to $r^2 = 0.67$. Simulation started with 10 haplotypes, each representing a different species that were located randomly in the landscape. Each haplotype was allowed to disperse according to (i) the dispersal distance m , which was set equal for all haplotypes, and (ii) species niche width w , defined by a maximum distance from the species optimum. Optima were defined as the environmental conditions at the initial, random position of origin of each species. Niche width was a fixed attribute of species that was set equal for all species. Therefore, after origination, the range of the haplotype was allowed to spread to reach all the cells located at spatial distance $< m$ and with climatic conditions differing from the optimum $< w$. We repeated the simulation for three different dispersal capacities ($m = 3, 5, 35$) combined with three different niche widths ($w = 10, 15, 35$). A further parameter was set to control the size of local assemblages. Given that abundances were not modelled, we set a limit to the number of local haplotypes (lhs) as a surrogate. Simulations were replicated for lhs values between 50 and 1,000 lhs, representing highly saturated to completely unsaturated local assemblages. Origination of new lineages was simulated at a given time by adding new haplotypes at random points within the existing species ranges, with the mutation rate constant per cell occupied. New and old haplotypes continued to disperse as defined by m . This process of sequential origin of new haplotypes and subsequent expansion of their ranges was repeated until each species included haplotypes of four different ages (hierarchical levels) and, consequently, four different range sizes. To avoid edge effects, we did not score the outer five rows of cells at each side of the lattice, so subsequent analyses were performed within the central 20×20 cells ($n = 400$). Again, it should be noted that allowing more origination/spread cycles would be strictly equivalent to increasing the migration parameter, which was allowed to vary from strong dispersal limitation to unlimited dispersal to cope with all situations. We then computed Simpson similarity between all pairs of cells at the species and haplotype levels, fitted exponential decay curves to test the variation explained (r^2) by spatial and climatic distance, and assessed its significance with a Mantel test (see above). Note that a low dispersal rate combined with a wide niche is equivalent to a dispersal-

constrained model, whereas under high dispersal combined with narrow niche width species ranges are controlled almost exclusively by climate. All simulations were performed in R⁴⁴.

References

- Brown, J. H. & Maurer, B. A. Macroecology - the division of food and space among species on continents. *Science* **243**, 1145–1150 (1989).
- Gaston, K. J. & Blackburn, T. M. *Pattern and Process in Macroecology* (Blackwell Science Ltd., 2000).
- Whittaker, R. J. *et al.* Conservation biogeography: assessment and prospect. *Divers. Distrib.* **11**, 3–23 (2005).
- Noss, R. F. Indicators for monitoring biodiversity - a hierarchical approach. *Conserv. Biol.* **4**, 355–364 (1990).
- Vellend, M. & Geber, M. A. Connections between species diversity and genetic diversity. *Ecol. Lett.* **8**, 767–781 (2005).
- Papadopoulou, A. *et al.* Testing the species-genetic diversity correlation in the Aegean archipelago: toward a haplotype-based macroecology? *Am. Nat.* **178**, 560–560 (2011).
- Vellend, M. Island biogeography of genes and species. *Am. Nat.* **162**, 358–365 (2003).
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, 2001).
- Bell, G. Neutral macroecology. *Science* **293**, 2413–2418 (2001).
- Rosindell, J., Hubbell, S. P. & Etienne, R. S. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol. Evol.* **26**, 340–348 (2011).
- McGill, B. J. Towards a unification of unified theories of biodiversity. *Ecol. Lett.* **13**, 627–642 (2010).
- Harte, J., McCarthy, S., Taylor, K., Kinzig, A. & Fischer, M. L. Estimating species-area relationships from plot to landscape scale using species spatial-turnover data. *Oikos* **86**, 45–54 (1999).
- Craft, K. J. *et al.* Population genetics of ecological communities with DNA barcodes: An example from New Guinea Lepidoptera. *Proc. Natl Acad. Sci. USA* **107**, 5041–5046 (2010).
- Pons, J. *et al.* Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* **55**, 595–609 (2006).
- Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* **26**, 867–878 (1999).
- Avise, J. C. *Molecular Markers, Natural History and Evolution* (Chapman & Hall, 1994).
- Templeton, A. R., Boerwinkle, E. & Sing, C. F. A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol-dehydrogenase activity in *Drosophila*. *Genetics* **117**, 343–351 (1987).
- Hart, M. W. & Sunday, J. Things fall apart: biological species form unconnected parsimony networks. *Biol. Lett.* **3**, 509–512 (2007).
- Baselga, A. Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. *Methods Ecol. Evol.* (in press) doi:10.1111/2041-210X.12029 (2013).
- Papadopoulou, A., Anastasiou, I. & Vogler, A. P. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol. Biol. Evol.* **27**, 1659–1672 (2010).
- Harte, J., Conlisk, E., Ostling, A., Green, J. L. & Smith, A. B. A theory of spatial structure in ecological communities at multiple spatial scales. *Ecol. Monographs* **75**, 179–197 (2005).
- Chave, J. & Leigh, E. G. A spatially explicit neutral model of beta-diversity in tropical forests. *Theor. Popul. Biol.* **62**, 153–168 (2002).
- Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* **270**, 313–321 (2003).
- Thuiller, W. *et al.* Predicting global change impacts on plant species' distributions: future challenges. *Perspect. Plant Ecol.* **9**, 137–152 (2008).
- Monaghan, M. T. *et al.* Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Syst. Biol.* **58**, 298–311 (2009).
- Bergsten, J. *et al.* The effect of geographical scale of sampling on DNA barcoding. *Syst. Biol.* **61**, 851–869 (2012).
- Simon, C. *et al.* Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. *Ann. Entomol. Soc. Am.* **87**, 651–701 (1994).
- Isambert, B. *et al.* Endemism and evolutionary history in conflict over Madagascar's freshwater conservation priorities. *Biol. Conserv.* **144**, 1902–1909 (2011).
- Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
- Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

31. Posada, D. jModelTest: Phylogenetic Model Averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008).
32. Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S. & Bremer, K. Estimating divergence times in large phylogenetic trees. *Syst. Biol.* **56**, 741–752 (2007).
33. Ezard, T., Fujisawa, T. & Barraclough, T. G. splits: SPecies' Limits by Threshold StatisticsR package version 1.0-14/r31. <http://R-Forge.R-project.org/projects/splits/> (2009).
34. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659 (2000).
35. Panchal, M. The automation of nested clade phylogeographic analysis. *Bioinformatics* **23**, 509–510 (2007).
36. Templeton, A. R., Crandall, K. A. & Sing, C. F. A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data.3. cladogram estimation. *Genetics* **132**, 619–633 (1992).
37. Ribera, I., Vogler, A. P. & Balke, M. Phylogeny and diversification of diving beetles (Coleoptera: Dytiscidae). *Cladistics* **24**, 563–590 (2008).
38. Baselga, A. Partitioning the turnover and nestedness components of beta diversity. *Global Ecol. Biogeogr.* **19**, 134–143 (2010).
39. Lennon, J. J., Koleff, P., Greenwood, J. J. D. & Gaston, K. J. The geographical structure of British bird distributions: diversity, spatial turnover and scale. *J. Anim. Ecol.* **70**, 966–979 (2001).
40. Baselga, A. & Orme, C. D. L. betapart: an R package for the study of beta diversity. *Methods Ecol. Evol.* **3**, 808–812 (2012).
41. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
42. Hawkins, B. A. *et al.* Energy, water, and broad-scale geographic patterns of species richness. *Ecology* **84**, 3105–3117 (2003).
43. Oksanen, J. *et al.* vegan: Community Ecology PackageR package version 2.0-2, available at <http://cran.r-project.org/> (2011).
44. R Development Core Team. R: A language and environment for statistical computing. Version 2.13.1. Available at <http://www.r-project.org> (2011).

Acknowledgements

We are grateful to collectors of specimens, in particular Drs G.N. Foster, D.T. Bilton, M. Balke, L. Hendrich, J. Geijer, H. Hermann, A.N. Nilsson and I. Ribera and to M. Elliott and R. Wild for technical help in the laboratory. We thank T. Barraclough, I. Ribera and J. Rosindell for invaluable discussions. Funding was provided by NERC (grant NE/C510908/1 to APV, T. Barraclough and MTM) and the Spanish Ministry of Science and Innovation (grant no. CGL2009-10111 to A.B.).

Author contributions

A.B., M.T.M. and A.P.V. conceived the study, A.B. performed statistical analyses and modelling work, J.B. and A.C.P. collected the data, T.F. and A.C.P. performed phylogeographic analysis, P.G.F. performed sequence compilation, A.B. and A.P.V. wrote a draft of the paper, and all authors contributed substantially to revisions.

Additional information

Accession codes: Sequences have been deposited in EMBL Protein Knowledgebase database under accession numbers JN840019–JN845080.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Baselga, A. *et al.* Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nat. Commun.* **4**:1892 doi: 10.1038/ncomms2881 (2013).