bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pleistocene climate change and the formation of regional species pools

Joaquín Calatayud[1,2,3], Miguel Ángel Rodríguez[1], Rafael Molina-Vengas[4], María Leo[5], José Luís Hórreo[6] and Joaquín Hortal[2].

1 Departamento de Ciencias de la Vida, Universidad de Alcalá, Edificio de Ciencias, Ctra. Madrid-Barcelona km. 33,6, 28871 Alcalá de Henares, Madrid, Spain.

2 Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (MNCN-CSIC), C/José Gutiérrez Abascal 2, 28006 Madrid, Spain.

3. Integrated Science Lab, Department of Physics, Umeå University, Naturvetarhuset, byggnad G, NA plan 3, IceLab Umeå universitet, 901 87 Umeå, Sweden.

4. Institute of Plant Sciences, University of Bern, Altenbergrain 21, Bern 3013, Switzerland

5. Departamento de Biodiversidad y Conservación. Real Jardín Botánico de Madrid (CSIC). 28014 Madrid, España

6. Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (MNCN-CSIC), C/José Gutiérrez Abascal 2, 28006 Madrid, Spain.

STRAPLINE: Pleistocene effects on species pools

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Despite the description of bioregions dates back from the origin of biogeography, the processes originating their associated species pools have been seldom studied. Ancient historical events are thought to play a fundamental role in configuring bioregions, but the effects of more recent events on these regional biotas are largely unknown. We use a network approach to identify regional and sub-regional faunas of European *Carabus* beetles, and analyse the effects of dispersal barriers, niche similarities and phylogenetic history on their configuration. We identify a transition zone matching the limit of the ice sheets at Last Glacial Maximum. While southern species pools are mostly separated by dispersal barriers, in the north species are mainly sorted by their environmental niches. Strikingly, most phylogenetic structuration of *Carabus* faunas occurred since the beginning of the Pleistocene. Our results show how extreme recent historical events – such as Pleistocene climate cooling, rather than just deep-time evolutionary processes, can profoundly modify the composition and structure of geographic species pools.

Keywords: Pleistocene glaciations, regional community, species pool, regionalization, occurrence networks, dispersal, niche tracking, *Carabus* ground beetles, Europe.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Naturalists have long been captivated by the geographic distribution of world biotas. Rooted in the seminal ideas by Alexandre von Humbolt, this fascination has promoted a long-term research agenda aiming to delineate biogeographic regions according to their integrating faunas and floras (e.g. Wallace 1876, Holt et al. 2013, Rueda et al. 2013). Besides this, the large-scale eco-evolutionary processes that shape regional biotas are known to influence ecological and evolutionary dynamics at finer scales (Ricklefs 2008, 2015). For instance, regional species pools can modulate local diversity patterns (Ricklefs 2011, Medina et al. 2014, Ricklefs and He 2016), the structure and functioning of ecosystems (Naeslund and Norberg 2006), or co-evolutionary processes (Calatayud et al. 2016a). However, despite their fundamental importance, the processes that have configured regional biotas have been seldom studied (and particularly the historical ones), and most explanations on their origin and dynamics remain largely narrative (Crisp et al. 2011).

Perhaps the earliest speculations about the formation of regional species pools took place during the flourishment of bioregionalizations in the mid-19th century (reviewed by Ebach 2015). During that time, and beyond referring to geophysical factors (climate, soils, and physical barriers), some authors already started to emphasize historical influences as key elements determining the configuration of plant and animal regions. For instance, when Wallace (1876) proposed his ground-breaking zoogeographic regions, he argued that while the distribution of ancient linages such as genera and families would likely reflect major geological and climatic changes spanning the early and mid-Cenozoic, species distributions would be more influenced by recent events such as Pleistocene glaciations (see Rueda et al. 2013). These recent events could have fostered many additions and subtractions of species to regional faunas through dispersal and diversification processes. Indeed, increasing evidence suggests that Pleistocene glacial-interglacial dynamics may have driven population extinctions (e.g. Barnes et al. 2002), allopatric speciation in glacial refugia (e.g. Johnson et al. 2004) and post-glacial recolonization events (e.g. Hewitt 1999; Theissinger et al. 2013). Besides shaping phylogeographic patterns (Avise et al. 1998; Ursenbacher et al. 2006; Sommer and Nadachowski 2006; Provan and Bennett 2008), all these processes are likely underpinning diversity patterns for many taxa, particularly in the Holarctic (e.g. Svenning and Skov 2007; Hortal et al. 2011; Calatayud et al. 2016b). However, whether

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the signature of Pleistocene glaciations scales up to the configuration of regional biotas remains largely unknown.

Historical contingencies should act over the intricate interplay between ecological (i.e. environmental tolerances and dispersal) and evolutionary (i.e. diversification and adaptation to new habitats) processes underpinning the composition and structure of regional species pools. On the one hand, niche-based processes may determine the composition of regional species pools (Mittelbach and Schemske 2015), mainly throughout their effects on species distribution ranges (Soberon 2007, Hortal et al. 2010, 2012). These processes integrate responses to abiotic conditions along geographical gradients and to local and regional biotic environments (Colwell et al. 2009), which may ultimately lead to the appearance of distinct regional communities in areas of contrasted environmental conditions (Ricklefs 2015). Although species with similar environmental tolerances/preferences can coexist in regions of similar climate, their dispersal may be constrained by geographical barriers, which may lead to divergent species pools under similar environmental conditions. Finally, evolutionary processes also constrain all these mechanisms. For instance, environmental-driven regions may be expected if occupancy of new areas is constrained by niche conservatism (Hortal et al. 2011), which should also lead to species pools integrated of evolutionary related species (i.e. niche conservatism generating phylogenetically clustered species pools, Fig.1a). This, however, can be in turn filtered by biogeographical processes (Gouveia et al. 2014). As such, diversification of lineages within regions separated by strong dispersal barriers (e.g. mountain ranges) may also lead to phylogenetically clustered pools of locally adapted species (i.e. geographically driven niche conservatism; Fig.1a; Warren et al. 2014, Calatayud et al. 2016a). Historical contingencies may contribute to the configuration of regional pools by modifying the balance between these processes. For example, differential diversification rates may be the predominant driver of regional species pools during climatically stable periods (Cardillo 2011). Yet, regions with a greater influence of climatic fluctuations such as Pleistocene glaciations may harbour pools of species mostly shaped by the joint effects of current climate and post-glacial colonization dynamics (Svenning et al. 2015), thus eroding the signature of geographically-structured diversification processes.

In this study we aim to disentangle the relative importance of the processes that may contribute to the formation of regional species pools, using European *Carabus* (Coleoptera: Carabidae) as a model lineage. *Carabus* is a species-rich ground beetle

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genus of great popularity due to the beautiful jewel-like appearance of some species (Turin et al. 2003). In general, *Carabus* species are flightless nocturnal predators of snails, earthworms and caterpillars. They hold hydrophilic adaptations and are typically associated to deciduous forests (Deuve et al. 2012). Previous evidence suggests that the richness of species from this genus in Europe is determined to a large extent by both current environmental conditions (i.e. climate and habitat) and glacial-interglacial dynamics (Calatayud et al. 2016b). This makes European *Carabus* an ideal case study to evaluate the joint effects of evolutionary, ecological and historical contingency processes as drivers of regional species pools.

Specifically, we use data on the distribution and evolutionary relationships of *Carabus* species, along with network and phylogenetic analyses, to evaluate six hypotheses: First, given the presumed low dispersal capacity of the species from this genus (Turin et al. 2003), we hypothesize that (H1) European *Carabus* species pools are mainly shaped by the main orographic barriers of the continent, but also, that (H2) glacial-interglacial dynamics have led to strong differentiation between northern and southern regional species pools. If this differentiation is true, northern European *Carabus* faunas will be comprised of species that colonized newly vacant habitats after the withdrawal of the ice sheet, and hence (H3) their regional distribution will be mostly determined by current climate. In contrast, (H4) southern faunas will be mainly shaped by the joint influence of diversification events and dispersal limitations, due to the combined effect of higher climatic stability (e.g. climatic refugia) and a more complex orography (Alps, Pyrenees, Carpathians). Therefore, (H5) species forming northern regional pools will exhibit comparatively lower levels of regional endemicity, whereas those forming southern regional pools will show comparatively higher levels of regional affinity. Finally, according to Wallace (1876), the advance and retreat of the ice sheets during the Pleistocene should have determined the spatial distribution of lineages, eroding the effects of the former configuration of the distribution of the main *Carabus* lineages. Therefore, (H6) we expect a temporal signal coincident with the Pleistocene in the phylogenetic structure of *Carabus* faunas, and no effect of deep-time events on the current geographical distribution of these lineages.

Material and methods

Rationale and structure of the analyses

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Exploring the determinants of regional faunas requires jointly analysing ecological, evolutionary and historical factors. We did so through three consecutive steps (Fig.1b). First, we identified distinct regional species pools within Europe by using a network community detection algorithm. From this analysis we derived a species pairwise similarity matrix of occurrence into different modules, each one representing different regions. Second, we assessed the relative importance of environmental, spatial and evolutionary determinants of such similarity. To do so, we constructed four pairwise matrices to describe ecological, topographical and evolutionary relationships among species; namely, *i*) a matrix of climatic niche similarity, *ii*) a matrix of habitat similarity, *iii*) a matrix of spatial connectivity among distributional ranges, and *iv*) a phylogenetic distance matrix. Then, we used generalized partial matrix regressions to model the similarity in species occurrences as a function of these four matrices (Fig.1b). We used this workflow to explore the factors involved in the configuration of *Carabus* faunas both at regional (i.e. through analysing *Carabus* species co-occurrence across regions) and sub-regional scale (i.e. focusing on co-occurrence patterns within sub-regions). Finally, we also applied ancestral range estimation analysis in order to identify the time period from which ancestral areas are estimated with less uncertainty. By doing so, we aimed to detect important historical periods contributing to the regional organization of *Carabus* lineages.

The interpretation of the joint and independent effects of explanatory matrices can shed light on the different processes configuring regional faunas (see Fig.1a). Thus, if niche similarities (i.e. represented by the climatic and habitat similarity matrices) and phylogenetic distances altogether explained the regional co-occurrence of species, then this could be interpreted as indicative of constrained niche evolution (or a tendency to resemble ancestral niches) in shaping regional faunas (Fig.1a.i). However, if spatial connectivity also accounted for part of this co-occurrence, this would indicate that this niche conservatism pattern can be caused by geographical constrains (Fig.1a.ii). Further, the effects of niche similarities and spatial connectivity alone (i.e. without phylogenetic signal) can be most likely the consequence of a convergence of climatic niches due to geographic isolation (Fig.1a.iii), whereas the effects of connectivity and phylogeny would be indicative of a primacy of intra-regional speciation driven by geographical barriers. Niche similarities alone would point to an unconstrained niche evolution shaping regional faunas, while phylogeny alone would indicate a primacy of geographically unconstrained intra-regional speciation events. Finally, either a *cul-de-*

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sac effect (i.e. the accumulation of species in past climatic refugia) or a primacy of vicariant speciation events could lead to the existence of independent effects of connectivity and regional co-occurrence (Fig.1a.iv).

Identification of regional species pools

We took advantage of community detection analysis —borrowed from network theory— to identify *Carabus* regional species pools in Europe. We first generated a bipartite network where species and grid cells constitute two disjoint sets of nodes that are connected according to the presence of species in grid cells (e.g. Calatayud et al. 2016a). The species presence data comes from expert-based range maps of all *Carabus* species inhabiting Europe (n = 131, Turin et al. 2003) overlaid into a 100-km equal-area grid based on the LAEA pan-European grid system (available at http://dataservice.eea.europa.euhttps://inspire.ec.europa.eu/, see Calatayud et al. 2016b for details). Then, we conducted a modularity analysis using the index proposed by Barber (2007) and the Louvain algorithm (Blondel et al. 2008) as implemented in the Matlab function "Gen Louvain", (available at http://netwiki.amath.unc.edu; Mucha et al. 2010). This analysis is intended to find groups of nodes (i.e. species and grid cells) that are more densely connected. Hence, in our case, the analysis identified groups of grid cells, each group sharing *Carabus* species mainly distributed within its cells (i.e. regions and their associated faunas). The Louvain algorithm was run 100 times, and the network partition showing highest modularity value was retained. This optimal solution was used to conduct all subsequent analyses, although all the solutions were qualitatively similar. We evaluated the statistical significance of the modules by comparing their associated modularity value to a null distribution of values (n = 100) where the original presence-absence matrix was randomized using the independent swap algorithm (a fixed-fixed null model implemented in the R package "picante", Kembel et al. 2010). Finally, to detect potential sub-modules (i.e. sub-regions) nested within modules (i.e. sub-regional species pools within regional species pools), we derived a new bipartite network from each of the previously identified modules, and applied the procedure described above in each case.

It is important to note that despite species and grid cells were assigned to just one module, they could also occur in other modules with different degrees of affinity. For example, despite most species in a grid cell will belong to the same module the cell does, this cell could also hold species that are primarily associated to other modules.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Similarly, although a species will mostly be present in cells assigned to its module, it may also occur in cells from other modules. Thus, we calculated the degree of module affinity for each node (i.e. species and grid cells) as its number of links with nodes of its module divided by its total number links (see Guimera and Amaral's 2005 inter-modular participation index for a similar metric). Thus, higher module affinities would correspond to species mainly distributed within its module (highly endemic species), as well as to cells pertaining to well-defined regions; whereas lower module affinities would reflect widespread species and cells located in transition zones.

Assessing the determinants of regional species pools

In order to disentangle the determinants of the current configuration of *Carabus* faunas in Europe, we first generated a species-per-module matrix, where each entry of the matrix represents the percentage of the distributional range of a certain species that lies in a given module. Then, we derived a co-occurrence pairwise similarity matrix from the former matrix using the Schoener's index (Schoener 1970). This metric quantifies the overlap between species pairs throughout the modules (see Krasnov et al. 2012 for a previous application) and it ranges from 0 (no overlap) to 1 (identical distribution across modules). Note that this similarity matrix reflects the co-occurrence similarities at regional scale, thus ignoring distributional patterns at lower spatial scales (i.e. two species may have identical regional distribution but appear as segregated at the local scale). The resultant co-occurrence pairwise similarity matrix was used as dependent variable. We generated four different pairwise dis/similarity matrices to be used as explanatory variables. Two of them were used to account for environmental factors: (i) a climatic-niche similarity matrix and (ii) a habitat similarity matrix. The remaining two considered geographical and evolutionary factors: (iii) a spatial-connectivity matrix and (iv) a phylogenetic distance matrix.

i) Climatic-niche similarity matrix. We characterized the climatic niche of each *Carabus* species in the dataset following a similar approach as proposed by Broennimann et al. (2012). We selected six bioclimatic variables to account for the main water and energy aspects of climate –namely mean annual temperature, temperature of the warmest quarter, temperature of the driest quarter, total annual precipitation, total precipitation of the warmest quarter and total precipitation of the driest quarter– and altitudinal range to account for the effects of mesoclimatic gradients within each grid cell. These variables

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

may be among the main determinants of the distribution of *Carabus* species diversity within Europe (see Calatayud et al. 2016b). Bioclimatic variables were extracted from Worldclim (v1.4 Hijmans et al. 2005; available at http://www.worldclim.org/), whereas altitudinal data were derived from the 30-arcsecond digital elevation model GTOPO30 (available at https://lta.cr.usgs.gov/GTOPO30). We conducted a principal component analysis on these variables to obtain a bidimensional climatic space defined by the two main axes, that explained 81.4% of the variance (Fig.S3.1). Finally, we divided this climatic space into a 100X100 grid and calculated species overlap in the gridded space using Schoener's index (see above).

ii) Habitat similarity matrix. The distribution of *Carabus* species may also be shaped by forest preferences (Turin et al. 2003). Accordingly, we used ten vegetation categories derived from MODIS Land Cover at 5-minute resolution (Evergreen broadleaf forest, deciduous needle-leaf forest, deciduous broadleaf forest, mixed forest, closed shrub lands, open shrub lands, woody savannas, savannas and grasslands; Channan et al. 2014, available at http://glcf.umd.edu/data/lc/). For each species we computed the proportion of each category overlaying its range. With this, we computed pairwise similarities in the preference for different vegetation types using Schoener's index (see above).

iii) Spatial connectivity matrix. To evaluate the potential influence of geophysical barriers to dispersal on the current distribution of *Carabus* species, we first created a dispersal-cost surface by dividing the study area in 1 Km^2 grid cells and weighting each cell according to its topography (in this case, slope) and the presence of water bodies. Slope values ranged from 0 to 100 at each pixel, being 0 the lowest dispersal cost and 100 the highest one, and were determined from GTOPO30 altitudinal data using the GRASS tool r.slope (GRASS Development Team 2017). Grid cells including water bodies as rasterized layers in the Nature Earth database (available at http://www.naturalearthdata.com/) were further weighted by assigning arbitrary values of friction to the dispersal of *Carabus* species, namely 30% for cells containing rivers and lakes and 99% for cells that lay on sea water masses (note that *Carabus* species show hydrophilic adaptations). Then, the connectivity between all pairs of cells was calculated as the least-cost path over the dispersal-cost surface that connects both cells, using the "gdistance" R package (van Etten 2015). Finally, the spatial connectivity

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between each pair of species' distributional ranges in the dataset was estimated as the average distance among all grid cells within the range of each species. Average distances were preferred over absolute least-cost distances to avoid disproportionate differences in spatial connectivity between overlapping and non-overlapping distributional ranges.

iv) Phylogenetic distance matrix. To unravel the evolutionary history of the *Carabus* lineage and assess the potential importance of evolutionary processes in determining the formation of *Carabus* species pool, we reconstructed a species-level time-calibrated molecular phylogeny including the 89 species for which we found available DNA information on ten markers (eight mitochondrial regions: 12S rDNA,16S rDNA, 28S rDNA, ND4, ND5, COI, Cytb and PEPCK; plus two nuclear ones: anonymous locus and wingless; see Tables S4.1 and S4.2). We aligned each marker independently using different algorithms: MAFFT (Katoh and Standley 2013), Clustal X (Thompson et al. 1994; Larkin et al. 2007), MUSCLE (Edgar 2004) and Kalign (Lassmann and Sonnhammer 2005), and selected the most reliable alignment for each marker using the multiple overlap score (MOS) provided by MUMSA (Lassmann and Sonnhammer 2006). We removed ambiguous or poorly aligned positions from the alignments with trimAl (Capella-Gutiérrez et al. 2009). The final dataset included 5603 characters, and was used to conduct Bayesian phylogenetic inference with BEAST v.2.4.6 software (Bouckaert et al. 2014). We used a GTR model for sequence evolution, a Random Local Clock, a birth-death model prior, and 100 million of MCMC chain length searching for convergence. There is ongoing debate on the divergence time of the genus *Carabus* (Andujar et al. 2012, Deuve et al. 2012), hence molecular dating was conducted under two different scenarios. Firstly, according to Deuve et al. (2012), the crown age of *Carabus* was set at 17.3 Mya. Secondly, and according to Andujar et al. (2012), the origin of the group was set at 25.16 Mya. We used the software Tracer v 1.6 (available at http://tree.bio.ed.ac.uk/software/tracer/) to check for MCMC chains convergence, and FigTree v.1.4.2 (available at http://tree.bio.ed.ac.uk/software/figtree/) for viewing and editing the phylogenetic trees (see Figures S3.2 and S3.3). In order to account for topological and time-calibration uncertainties, we used 100 phylogenies sampled from the posterior distribution for each molecular dating scenario. In addition, we used taxonomic information and phylogenetic uncertainty methods (Rangel et al. 2015) to place species lacking molecular information into the phylogeny (see Appendix S1).

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Thus, we derived 100 different phylogenetic hypotheses from each Bayesian posterior phylogeny by randomly inserting missing species within their most derived consensus clade based on taxonomic knowledge. In total, we generated 20,000 phylogenetic hypotheses (100 phylogenies per two molecular dating scenarios per 100 phylogenies accounting for uncertainties associated with lack of molecular data) that were used in subsequent analyses. Pairwise phylogenetic distances were calculated for each calibrated phylogeny using the function *cophenetic* implemented in the APE R package (Paradis et al. 2004).

We used generalized multiple regression on distance matrices and deviance partitioning to disentangle the relative importance of climatic niche, habitat preferences, dispersal barriers and evolutionary history in determining *Carabus* species pools in Europe. First, we conducted single regressions between the co-occurrence pairwise similarity matrix and each of the four explanatory matrices described above to seek for significant associations between the variables. We set a binomial family for error distribution and "logit" as the link function (see Ferrier et al. 2007 and Calatayud et al. 2016a for a similar approach). To assess for significance, we randomized the observed species per module matrix using the independent swap algorithm (see above) to derive 999 null occurrence similarity matrices. Then, we used simple regressions to relate each null similarity matrix with each one of the explanatory matrices. The relationship between an explanatory matrix and the observed species per module matrix was considered to be significant when it explained a higher proportion of the deviance than 99% of the regressions performed on the null matrices. In the case of phylogenetic pairwise distances we repeated this procedure for each phylogenetic hypothesis to consider phylogenetic uncertainties, applying the same criterion for significance. Finally, we retained those variables that showed significant relationships, and conducted variance partitioning among explanatory matrices (Legendre and Legendre 2012) to explore patterns of covariation among niche similarities (i.e. climatic and habitat similarity matrices), dispersal barriers and phylogenetic history. We conducted the analyses for the co-occurrence into modules (i.e. regions) and sub-modules (i.e. sub-regions) both at a European and regional (i.e. co-occurrence into sub-modules of each module) scales.

Ancestral range estimation

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Probabilistic models of geographic range evolution have proven to be of great utility to deal with historical biogeograhical questions (see Ronquist and Sanmartín 2011 and references therein). Here we used these models to try to unravel whether, according to Wallace's ideas (1876), deep historical signals were eroded by Pleistocene glaciations in the configuration of European *Carabus* faunas. We used Maximum Likelihood approach for inferring geographical range evolution implemented in the R package BioGeoBears (Matzke 2014). BioGeoBears fits a variety of models that differ in their types of dispersal, extinction, sympatric speciation, vicariance and founder event speciation. We fitted the six available models (DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, BAYAREALIKE+J, see Matzke 2013) to each of the 20,000 phylogenetic hypotheses described above, comparing their informative capacity through the Akaike Information Criterion weighted by sample size (AICc). Species ranges were coded as present/absent in each module detected in the former network clustering.

The estimation of ancestral ranges usually tends to be more ambiguous in deeper nodes of the phylogeny, as the lability of geographical ranges would tend to blur deep-time signals (Losos and Glor 2003). Similarly, if the Pleistocene glacial periods had important effects in the distribution of species it could be expected that ancestral range estimations will increase in accuracy around the Pleistocene. That is, pre-Pleistocene signals on the evolution of distributional range will be blurred. To explore this, we seek to detect changes in the relationship between node age and the marginal probability of the single most-probable ancestral state at each internal node. We obtained these probabilities from the best model among the six geographical range evolution models in BioGeoBears (see above) and for each of the phylogenetic hypotheses described above. Then, general additive mixed models (GAMMs) were fitted to the node marginal probability as a function of node age, including the phylogenetic hypothesis as a random factor. Furthermore, since we found different best models for different phylogenetic hypotheses we included model type as a fixed variable. We also used generalized linear mixed models (GLMMs) combined with piecewise regression to detect potential major breakpoints (i.e. temporal shifts) in the relationship between marginal probability and node age for each of the two dating scenarios. Given the large amount of observations (130 nodes x 10,000 phylogenies) we firstly estimated the breakpoint independently for each phylogenetic hypothesis. To do so, we included the breakpoint as a new parameter in a generalized linear model of

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

marginal probabilities as a function of node age, minimizing the deviance of the fitted model using the function "optimize" in the R package lme4 (Bates et al. 2014). Finally, to assess significance we used the averaged breakpoint value in a GLMM with the same fixed formulation but including the phylogenetic hypothesis as a random factor. Because node marginal probabilities ranged between 0 and 1, we used a binomial family and a loglink function to fit all models.

It is worth mentioning that our dataset does not fulfil the assumption of ancestral range estimation models that phylogenies should include all extant lineages, since some *Carabus* linages have representatives from outside Europe (mainly Asiatic species) that were not included in the analyses. However, the potential effect of such incomplete sampling should be similar as that of unrecorded clade extinctions, which have been estimated to have little effect on ancestral range reconstruction analyses (Matzke 2014). Nevertheless, we also conducted a simpler analysis to explore whether there is a temporal signal in the structuration of *Carabus* faunas coinciding with Pleistocene glaciations. To do so, we first generated a binomial variable based on the module affiliation of the species descending from each internal phylogenetic node. Thus, we gave a 1 to the phylogenetic nodes whose all descendant species were classified into the same region and 0 otherwise. Then, we fitted piecewise regression models of this variable as a function of node age using the approach previously explained. This analysis would serve to further elucidate if there is a breakpoint in the relationship between node age and the probability that all descendant species belong to the same region (see Appendix S2 for details and another complementary approach).

All analyses were carried out in R (R core team 2015), using the function *bam* of the package mcvg for GAMM (Wood at al. 2015) and the package Lme4 for GLMM analyses (Bates et al. 2014).

Results

Identification of regional faunas

The *Carabus* occurrence network was significantly modular (M=0.385, p=0.01), dividing Europe in seven modules that group zoogeographically distinct regions with their associated faunas (i.e., different regional species pools; Figs. 2a and S3.4). Furthermore, all modules but module 2 showed significant sub-modular structure, presenting a decrease in modularity with latitude (mean M=0.316, ranging from 0.154 to 0.468; all p-values < 0.05, see Table S4.3). Module 1 holds 21 species mainly living

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in South-western Palearctic (Iberian Peninsula, North of Africa, Balearic Islands, Corsica, Sardinia and the western half of Sicilia). This module was subdivided into four submodules. Module 2 included only two species, both endemic of Crete. Module 3 identified an East Mediterranean region including the Italic Peninsula, part of Greece and Turkey. This module holds 18 species and was subdivided into five submodules. Module 4 depicted a Central European region embracing the Alps and the Carpathian Mountains, as well as Central European plains. This module showed the highest species richness, including 49 *Carabus* species, and was split into four submodules. Module 5 and module 6 comprised northern regions and showed the lowest species richness values, holding 10 species each. The former comprised Iceland and the British Isles and extended eastward up to the vicinity of the Ural Mountains. The latter included this mountain range and expanded to the easternmost zone of the study area. Both modules were divided into 3 submodules. Finally, module 7 included 21 species and embraced a south-eastern central European region expanding from the Carpathian Mountains to the south Ural Mountains. This module was split into three submodules.

Regarding transition zones between regions, and in agreement with our first hypothesis, we found that they were clearly associated with geographical barriers such as the Pyrenees, the Alps, the Carpathian and the Ural Mountains, as well as the Turkish Straits System (that connects the Black Sea to the Aegean separating the Anatolian and Greek peninsulas; Fig. 2b). Interestingly, we also identified a west-to-east transitional belt between southern and northern regions that closely followed the southern limits of the ice sheet at the Last Glacial Maximum (LGM). This transitional zone further suggested a link between the configuration of *Carabus* regional faunas and Pleistocene glacial conditions, supporting our hypothesis 2.

Correlates of regional co-occurrence

Matrix regressions showed that deviance of species co-occurrences across regions, across sub-regions and within each region was significantly explained ($p<0.01$), primarily by environmental niche similarity, and secondarily by spatial connectivity, except for northern regions (i.e. modules 5 and 6; Fig. 3 and Table S4.4). In contrast, relationships with evolutionary relatedness were no significant in all instances and regardless of the phylogenetic hypothesis used ($p>0.01$ in all cases, see Table S4.4). Comparing both types of subdivisions, environmental niche similarity explained more deviance across sub-regions than across regions, whereas spatial connectivity did the

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reverse (see Fig. 3). Comparing explained deviances between regions, the primacy of environmental niche similarity (mostly climate, see Table S4.4) in the northern ones (5 and 6) is consistent with the notion that northern regional pools are geographically sorted by current climate (our hypothesis 3), whereas the importance shown by spatial connectivity in the remaining regions is consistent with the more complex orography of central and southern Europe (consistent with hypothesis 4).

Ancestral range estimation

Because both phylogenetic datasets (i.e. alternative calibration scenarios) yielded very similar qualitative and quantitative results (see Figure S3.5 and Table S4.5), we only present here ancestral range estimations based on Deuve's et al. (2012) calibration. BAYAREALIKE+J and DEC+J were the best models for 79% and 19.7% of the phylogenetic hypotheses, respectively (Table 1). Interestingly, both models consider founder event speciation, which could be an indication of Pleistocene glaciations promoting speciation in glacial refugia (i.e. acting as islands). Moreover, according to the BAYAREALIKE+J model the range contraction parameter had the highest estimate (extinction parameter in Table 1), which could be also expected due to a generalized range erosion of glaciations. This parameter was however of little importance in the DEC+J model, which accounted for vicariant events. Regardless of the biogeographical model used, GAMM results showed that node marginal probability of the most probable state increased as expected towards younger nodes ($P<0.01$, $R^2=0.26$). However, this increase showed steep increments at the border between the Miocene and the Pliocene and, more abruptly, coinciding with the beginning of the Pleistocene c. 2.59 Mya (Fig.4). Indeed, piecewise regresion revealed that the relationship between marginal state probability and node age changed at 2.16 Mya (median value; with 45th and 55th percentiles at 1.68 and 2.94 Mya., respectively). This points to a distinct signal of the Pleistocene on the phylogenetic configuration of European *Carabus* faunas, providing support for our hypothesis 6. These results were confirmed by the complementary binomial analysis, which showed a breakpoint at 3.70 Mya. (45th and 55th percentiles at 2.47 and 4.48 Mya.; see also Appendix S2). The breakpoint associated to the Pliocene-Pleistocene transition was significant in all cases ($P < 0.01$).

Discussion

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

More than 140 years ago, Wallace (1876) foresaw that the influence of Pleistocene glaciations on the distribution of diversity had been strong enough to erode the imprint of previous events. Our results support Wallace's thoughts, showing a remarkable coincidence between the distribution of the ice sheets at the Last Glacial Maximum and the current configuration and evolutionary structure of European *Carabus* Faunas.

The first line of evidence supporting this idea comes from the close spatial relationship between the southern limits of the ice sheet at LGM and the transition zone separating the southern and northern regions. This border also coincides with the line – identified by Calatayud et al. (2016b)– where the relationship between *Carabus* species richness and current climate changes (Fig. 2). Thus, it seems that the climate changes underwent during the Pleistocene not only shaped phylogeographic (Avise et al. 1998; Hewitt 1999; Barnes et al. 2002; Johnson et al. 2004; Theissinger et al. 2013) and species richness patterns (e.g. Svenning and Skov 2007, Araújo et al. 2008, Hortal et al. 2011, Calatayud et al. 2016b), but that Ice ages have also left a strong imprint on the geographical structure of species composition at a regional scale. Accordingly, the species from the northernmost region (module 5) show the lowest level of endemism (Fig. S3.6), as expected for regional faunas composed of species that have recently colonized the north of Europe from southern glacial refugia (Araújo et al. 2008, Calatayud et al. 2016b, our hypothesis 5). In fact, although these species show large distribution ranges in different parts of southern Europe, their ranges only overlap near the northern Carpathian Mountains (Fig. S3.7). This area was a glacial refugia for a large and taxonomically diverse array of northern European species (Ursenbacher et al. 2006; Sommer and Nadachowski, 2006, Provan and Bennett 2008), including *Carabus* (Homburg et al. 2013, Fig. S3). Additionally, the decrease in modularity values with latitude also points to a lesser geographical structure of northern assemblages, which can be interpreted as the result of a post-glacial colonization, together with less geographic complexity in some areas.

Besides the Pleistocene effects in the definition and geographical structure of regional species pools, we also found evidence of the imprint of this geological period on the processes configuring the distribution of *Carabus* faunas. The general strong relationship between regional patterns of co-occurrence and both niche similarities and spatial connectivity shows that co-occurring species tend to have similar realized environmental niches and that also tend to be geographically constrained by the same dispersal barriers. This latter result was expected given the –presumed– low dispersal

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

capacity of *Carabus* species (see Turin et al. 2003), which is likely to be behind the spatial coincidence of module transition zones and geographical barriers. Perhaps more unexpected is the weak effect of phylogenetic distances despite the strong relationship between regional co-occurrence and niche similarities. This implies that geographical barriers rather than climatic-niche conservatism have restricted species distributions even within regions of similar climate. These results also point to that *Carabus* niche evolution is, to some extent, evolutionarily unconstrained, which is congruent with the general high adaptation capacity of insects (e.g. Overgaard and Sørensen 2008).

Whatever the origin of the relationship between species occurrence and environmental conditions, what is certainly true is that its strength changes between regions. These changes follow a latitudinal gradient in the importance of environmental niche similarities. The occurrence into sub-regions is more strongly related to the similarity in the realized niche in the north than in the south. This might be a direct consequence of the effects of post-glacial colonization, where formerly glaciated areas show a clear sorting of species due to its environmental preferences. On the contrary, in southern regions, species are expected to had have more time to diversify and sort geographically by other factors besides climate (Hortal et al. 2011). Our findings partially corroborated this idea since we found strong effects of dispersal barriers in these areas. Yet, a longer time for speciation should have also resulted in some phylogenetic signal in southern regions, which was not supported by our results.

This generalized lack of phylogenetic structuration of *Carabus* faunas can be the outcome of –relatively– recent speciation events due to vicariance and/or a "cul-de-sac effect" (O'Regan 2008). The former would imply the formation of dispersal barriers promoting the geographical split of many lineages and subsequent allopatric speciation (Weeks et al. 2016). Yet, the geophysical accidents that can be associated with the limits of the *Carabus* regions we found here largely predate the origin of the genus (see Beccaluva et al. 1998, Deuve et al. 2012). On the other hand, a generalized dispersion into climatic refugia, together with a subsequent stagnancy within them (i.e. a "cul-de-sac" effect) may also produce the observed mixing of unrelated linages into regions. Although it is difficult to distinguish between both processes, the latter seems more plausible, with southern regions accumulating unrelated species while acting as glacial refugia, and northern ones being recolonized by unrelated species with similar environmental niches and/or simply higher dispersal capacity (Svenning and Skov

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2007). Future studies are required to elucidate which specific mechanisms allowed some *Carabus* species to colonize northern latitudes.

In agreement with this idea, the model of ancestral range estimation that is best supported by the data does not include vicariant events. On the contrary, this model considers founder speciation events, a process typically associated with islands (Provine 1989). As such, this could be interpreted on the light of southern regions acting as islands during the different Pleistocene glacial maxima. Nevertheless, these results must be taken with caution because some assumptions of ancestral estimation models were not fully satisfied, and because of the low predictive power of these models (i.e. low marginal probabilities). Nevertheless, marginal probabilities showed a steeper increase coinciding with the Pliocene-Pleistocene transition, which is a more reliable result. Indeed, this result was consistent with the more conservative binomial analysis and across the different time calibration scenarios. This remarkable temporal coincidence supports that the current regional organization of *Carabus* species and linagese is rooted at the beginning of the Pleistocene, which also explains the lack of phylogenetic structure of regional *Carabus* faunas. Our results contrast with ancestral range estimations for clades inhabiting areas that were never glaciated, where more ancient signals were found in the spatial sorting of lineages (Condamine et al. 2015, Economo et al. 2015, Tänzler et al. 2016, Toussaint and Balke 2016). Hence, it seems that the repeated advances and retreats of ice sheets and glacial conditions that characterize the European Pleistocene produced repeated cycles of retreat to southern regions and advance towards the north of *Carabus* species, a hustle-and-bustle process that ultimately led to the observed mixing of unrelated lineages.

To summarize, our results provide solid arguments in favour of the importance of Pleistocene glaciations along with geographical barriers and niche-based processes in structuring the regional faunas of European *Carabus*. On the one hand, this group's faunas are primarily delimited by the location of the southern limit of the ice sheet at LGM, which separates two large regions that differ not only in species composition, but also in the processes underlying the spatial organization of these species. On the other hand, the phylogenetic structure of these faunas coincides with the beginning of the Pleistocene. This implies that the geographical distribution of species and lineages is profoundly shaped by past climates. Moreover, our results also suggest that local ecological processes (Naeslund and Norberg 2006, Madrigal et al. 2016) and evolutionary mechanisms (Wüest et al. 2015, Calatayud et al. 2016a) may be

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

profoundly affected by the history of Earth climatic. Hence, the study of these historical events may be essential to unravel both large and local scale diversity patterns.

Acknowledgements

We are very grateful to Achille Casale for providing data on *Carabus* habitat preferences and comments on an early version of the phylogeny, and the Scientific Computation Centre of Andalusia (CICA) for the computing services they provided. We acknowledge insightful discussions with Hortal lab members. JC was supported by a FPU-fellowship of the Spanish Ministry of Education (FPU12/00575). This work is partly supported by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) project SCARPO (CGL2011-29317) to JC and JH. MAR and RMV were supported through the grants CGL2017-86926-P and CGL2013-48768-P. ML and JLH were supported by MINECO FPI and Juan de La Cierva grants, respectively.

Author contributions

JC and JH conceived research. JC designed the study with contributions of all authors. JC, RMV, ML and JLH analysed the data. All authors discussed results. JC wrote the paper with contributions of all authors.

References

1. Andújar, C., Serrano, J. and Gómez-Zurita, J. 2012. Winding up the molecular clock in the genus Carabus (Coleoptera: Carabidae): assessment of methodological decisions on rate and node age estimation. BMC evolutionary biology, **12**: 40.
2. Araújo, M. B., D. Nogués-Bravo, J. A. F. Diniz-Filho, A. M. Haywood, P. J. Valdes, and C. Rahbek. 2008. Quaternary climate changes explain diversity among reptiles and amphibians. Ecography **31**:8–15.
3. Avise, J. C., Walker, D., and Johns, G. C. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proceedings of the Royal Society of London B: Biological Sciences **265**: 1707-1712.
4. Barber, M. J. 2007. Modularity and community detection in bipartite networks. Physical Review E **76**:066102.
5. Barnes, I., P. Matheus, B. Shapiro, D. Jensen, and A. Cooper. 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. Science **295**:2267–2270.
6. Bates, D., M. Maechler, B. Bolker, S. Walker, et al. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version **1**.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

7. Beccaluva, L., M. Shallo, M. Coltorti, I. Premti, and F. Siena. 1998. Encyclopedia of European and Asian Regional Geology. Springer.
8. Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics **2008**:P10008.
9. Broennimann, O., M. C. Fitzpatrick, P. B. Pearman, B. Petitpierre, L. Pellissier, N. G. Yoccoz, W. Thuiller, M.-J. Fortin, C. Randin, N. E. Zimmermann, et al. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. Global Ecology and Biogeography **21**:481–497.
10. Calatayud, J., J. L. Hórreo, J. Madrigal-González, A. Migeon, M. Á. Rodríguez, S. Magalhães, and J. Hortal. 2016*a*. Geography and major host evolutionary transitions shape the resource use of plant parasites. Proceedings of the National Academy of Sciences USA **113**:9840–9845.
11. Calatayud, J., J. Hortal, N. G. Medina, H. Turin, R. Bernard, A. Casale, V. M. Ortuño, L. Penev, and M. Á. Rodríguez. 2016*b*. Glaciations, deciduous forests, water availability and current geographical patterns in the diversity of European *Carabus* species. Journal of Biogeography **43**:2343–2353.
12. Capella-Gutiérrez, S., J.M. Silla-Martínez, and T. Gabaldón. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972–1973.
13. Cardillo, M. 2011. Phylogenetic structure of mammal assemblages at large geographical scales: linking phylogenetic community ecology with macroecology. Philosophical Transactions of the Royal Society of London B **366**:2545–2553.
14. Channan, S., K. Collins, and W. Emanuel. 2014. Global mosaics of the standard MODIS land cover type data. University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA.
15. Colwell, R. K., and T. F. Rangel. 2009. Hutchinson's duality: the once and future niche. Proceedings of the National Academy of Sciences USA **106**:19651–19658.
16. Condamine, F. L., E. F. Toussaint, A.-L. Clamens, G. Genson, F. A. Sperling, and G. J. Kergoat. 2015. Deciphering the evolution of birdwing butterflies 150 years after Alfred Russel Wallace. Scientific Reports **5**:11860.
17. Crisp, M. D., S. A. Trewick, and L. G. Cook. 2011. Hypothesis testing in biogeography. Trends in Ecology and Evolution **26**:66–72.
18. Deuve, T., A. Cruaud, G. Genson, and J.-Y. Rasplus. 2012. Molecular systematics and evolutionary history of the genus *Carabus* (Col. Carabidae). Molecular Phylogenetics and Evolution **65**:259–275.
19. Ebach, M.C. 2015. Origins of Biogeography. Springer, New York.
20. Economo, E. P., E. M. Sarnat, M. Janda, R. Clouse, P. B. Klimov, G. Fischer, B. D. Blanchard, L. N. Ramirez, A. N. Andersen, M. Berman, et al. 2015. Breaking out of biogeographical modules: range expansion and taxon cycles in the hyperdiverse ant genus Pheidole. Journal of Biogeography 42:2289–2301.
21. Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**:1792–1797.
22. Ehlers, J., and P. L. Gibbard. 2004. Quaternary glaciations-extent and chronology: part I: Europe. Elsevier.
23. Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Diversity and Distributions **13**:252–264.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24. Gouveia, S. F., J. Hortal, M. Tejedo, H. Duarte, F. A. Cassemiro, C. A. Navas, and J. A. F. Diniz-Filho. 2014. Climatic niche at physiological and macroecological scales: the thermal tolerance–geographical range interface and niche dimensionality. Global Ecology and Biogeography **23**:446–456.
25. GRASS Development Team, 2017. Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2. Open Source Geospatial Foundation. http://grass.osgeo.org.
26. Guimerà, R., and L. A. N. Amaral. 2005. Functional cartography of complex metabolic networks. Nature **433**:895–900.
27. Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society **68**:87–112.
28. Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965–1978.
29. Holt, B. G., J.-P. Lessard, M. K. Borregaard, S. A. Fritz, M. B. Araújo, D. Dimitrov, P.-H. Fabre, C. H. Graham, G. R. Graves, K. A. Jønsson, et al. 2013. An update of Wallace's zoogeographic regions of the world. Science **339**:74–78.
30. Homburg, K., C. Drees, M. M. Gossner, L. Rakosy, A. Vrezec, and T. Assmann. 2013. Multiple glacial refugia of the low-dispersal ground beetle *Carabus irregularis*: molecular data support predictions of species distribution models. PloS One **8**:e61185.
31. Hortal, J., P. De Marco Jr, A. Santos, and J. A. F. Diniz-Filho. 2012. Integrating biogeographical processes and local community assembly. Journal of Biogeography **39**:627–628.
32. Hortal, J., J. A. F. Diniz-Filho, L. M. Bini, M. Á. Rodríguez, A. Baselga, D. Nogués-Bravo, T. F. Rangel, B. A. Hawkins, and J. M. Lobo. 2011. Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. Ecology Letters **14**:741–748.
33. Hortal, J., J. M. Lobo, and A. Jiménez-Valverde. 2012. Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. Natureza & Conservação **10**:108–118.
34. Hortal, J., N. Roura-Pascual, N. Sanders, and C. Rahbek. 2010. Understanding (insect) species distributions across spatial scales. Ecography **33**:51.
35. Johnson, N. K., C. Cicero, and K. Shaw. 2004. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. Evolution **58**:1122–1130.
36. Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution **30**:772–780.
37. Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics **26**:1463–1464.
38. Krasnov, B. R., M. A. Fortuna, D. Mouillot, I. S. Khokhlova, G. I. Shenbrot, and R. Poulin. 2012. Phylogenetic signal in module composition and species connectivity in compartmentalized host-parasite networks. The American Naturalist **179**:501–511.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39. Lassmann, T. and E. L. Sonnhammer. 2005. Kalign - an accurate and fast multiple sequence alignment algorithm. BMC bioinformatics **6**:1–9.
40. Lassmann, T. and E. L. Sonnhammer. 2006. Kalign, Kalignvu and Mumsa: web servers for multiple sequence alignment. Nucleic Acids Research **34**: W596-W599
41. Legendre, P., and L. F. Legendre. 2012. Numerical ecology. Elsevier.
42. Losos, J. B., and R. E. Glor. 2003. Phylogenetic comparative methods and the geography of speciation. Trends in Ecology and Evolution **18**:220–227.
43. Madrigal-González, J., P. Ruiz-Benito, S. Ratcliffe, J. Calatayud, G. Kändler, A. Lehtonen, J. Dahlgren, C. Wirth, and M. A. Zavala. 2016. Complementarity effects on tree growth are contingent on tree size and climatic conditions across Europe. Scientific Reports **6**:32233.
44. Matzke, N. J., 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. University of California, Berkeley, Berkeley, CA. http://CRAN.R-project.org/package=BioGeoBEARS.
45. Matzke, N. J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Systematic Biology **63**:51–70.
46. Mazerolle, M. J. 2011. AICcmodavg: model selection and multimodel inference based on (Q) AIC (c). R package version **1**.
47. Medina, N. G., B. Albertos, F. Lara, V. Mazimpaka, R. Garilleti, D. Draper, and J. Hortal. 2014. Species richness of epiphytic bryophytes: drivers across scales on the edge of the Mediterranean. Ecography **37**:80–93.
48. Mittelbach, G. G., and D. W. Schemske. 2015. Ecological and evolutionary perspectives on community assembly. Trends in Ecology and Evolution **30**:241–247.
49. Mucha, P. J., T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. 2010. Community structure in time-dependent, multiscale, and multiplex networks. Science **328**:876–878.
50. Naeslund, B., and Norberg, J. 2006. Ecosystem consequences of the regional species pool. *Oikos*, **115***:*504-512.
51. O'Regan, H. J. 2008. The Iberian Peninsula–corridor or cul-de-sac? Mammalian faunal change and possible routes of dispersal in the last 2 million years. Quaternary Science Reviews **27**:2136–2144.
52. Overgaard, J., and J. G. Sørensen. 2008. Rapid thermal adaptation during field temperature variations in *Drosophila melanogaster*. Cryobiology **56**:159–162.
53. Paradis, E., J. Claude and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics **20**:289-290.
54. Provan, J., and Bennett, K. D. 2008. Phylogeographic insights into cryptic glacial refugia. Trends in ecology & evolution **23**: 564-571.
55. Provine, W. B., 1989. Founder effects and genetic revolutions in microevolution and speciation: an historical perspective. Pp. 43–76 *in* L. Val Giddings and K. Y. Kaneshiro, editors. Genetics, speciation, and the founder principle. Oxford University Press.
56. R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

57. Rangel, T. F., R. K. Colwell, G. R. Graves, K. Fučíková, C. Rahbek, and J. A. F. Diniz-Filho. 2015. Phylogenetic uncertainty revisited: Implications for ecological analyses. Evolution **69**:1301–1312.
58. Ricklefs, R. E. 2008. Disintegration of the ecological community. The American Naturalist **172**:741–750.
59. Ricklefs, R. E. 2011. Applying a regional community concept to forest birds of eastern North America. Proceedings of the National Academy of Sciences USA **108**:2300–2305.
60. Ricklefs, R. E. 2015. Intrinsic dynamics of the regional community. Ecology Letters **18**:497–503.
61. Ricklefs, R. E., and F. He. 2016. Region effects influence local tree species diversity. Proceedings of the National Academy of Sciences USA **113**:674–679.
62. Ronquist, F., and I. Sanmartín. 2011. Phylogenetic methods in biogeography. Annual Review of Ecology, Evolution and Systematics **42**:441–464.
63. Rueda, M., M. Á. Rodríguez, and B. A. Hawkins. 2013. Identifying global zoogeographical regions: lessons from Wallace. Journal of Biogeography **40**:2215–2225.
64. Schoener, T. W. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. Ecology **51**:408–418.
65. Sommer, R. S., and Nadachowski, A. 2006. Glacial refugia of mammals in Europe: evidence from fossil records. Mammal Review **36**: 251-265.
66. Svenning, J.-C., W. L. Eiserhardt, S. Normand, A. Ordonez, and B. Sandel. 2015. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. Annual Review of Ecology, Evolution and Systematics **46**:551–572.
67. Svenning, J.-C., and F. Skov. 2007. Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? Ecology Letters **10**:453–460.
68. Tänzler, R., M. H. Van Dam, E. F. Toussaint, Y. R. Suhardjono, M. Balke, and A. Riedel. 2016. Macroevolution of hyperdiverse flightless beetles reflects the complex geological history of the Sunda Arc. Scientific Reports **6**:18793.
69. Theissinger, K., M. Bálint, K. A. Feldheim, P. Haase, J. Johannesen, I. Laube, and S. U. Pauls. 2013. Glacial survival and post-glacial recolonization of an arctic–alpine freshwater insect (*Arcynopteryx dichroa, Plecoptera, Perlodidae*) in Europe. Journal of Biogeography **40**:236–248.
70. Thompson, J. D., D. G. Diggins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22**:4673–4680.
71. Toussaint, E. F., and M. Balke. 2016. Historical biogeography of Polyura butterflies in the oriental Palaeotropics: trans-archipelagic routes and South Pacific island hopping. Journal of Biogeography **43**:1560–1572.
72. Turin, H., L. Penev, A. Casale, E. Arndt, T. Assmann, K. Makarov, D. Mossakowski, G. Szél, F. Weber, H. Turin, et al., 2003. Species accounts. Pp. 151–284 *in* H. Turin, L. Penev, and A. Casale, editors. The Genus *Carabus* in Europe: A Synthesis. Pensoft Publishers.
73. Ursenbacher, S., Carlsson, M., Helfer, V., Tegelström, H., and Fumagalli, L. 2006. Phylogeography and Pleistocene refugia of the adder (Vipera berus) as inferred from mitochondrial DNA sequence data. Molecular Ecology **15**: 3425-3437.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

74. van Etten, J., 2015. gdistance: Distances and Routes on Geographical Grids. http://CRAN.R-project.org/package=gdistance.
75. Wallace, A. R. 1876. The geographical distribution of animals: with a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface. Cambridge University Press.
76. Warren, D. L., M. Cardillo, D. F. Rosauer, and D. I. Bolnick. 2014. Mistaking geography for biology: inferring processes from species distributions. Trends in Ecology and Evolution **29**:572–580.
77. Weeks, B. C., S. Claramunt, and J. Cracraft. 2016. Integrating systematics and biogeography to disentangle the roles of history and ecology in biotic assembly. Journal of Biogeography **43**:1546–1559.
78. Wood, S.N., Goude, Y. and Shaw S. 2015. Generalized additive models for large datasets. Journal of the Royal Statistical Society, Series C **64**: 139–155.
79. Wüest, R. O., A. Antonelli, N. E. Zimmermann, and H. P. Linder. 2015. Available climate regimes drive niche diversification during range expansion. The American Naturalist **185**:640–652.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Results of ancestral range estimation models. LnL= model likelihood; Nº param.= number of model parameters; d= dispersion, e= extinction; j= founder speciation. W. AICc= Weighted AICc. Prop. best= proportion of phylogenetic hypotheses where a model was the best in terms of AICc. Mean values among different phylogenetic hypotheses (incl. confidence interval at $\alpha = 0.05$) are shown for each parameter.

Model	LnL	Nº param.	d	e	j	W. AICc	Prop. best
DEC	-470.874±0.184	2	0.038±0.000	0.005±0.000	0.000±0.000	0.038±0.002	0.015
DEC+J	-466.778±0.171	3	0.035±0.000	0.000±0.000	0.013±0.000	0.184±0.006	0.196
DIVALIKE	-494.288±0.251	2	0.043±0.000	0.001±0.000	0.000±0.000	0.000±0.000	0.000
DIVALIKE+J	-491.668±0.242	3	0.041±0.000	0.000±0.000	0.008±0.000	0.000±0.000	0.000
BAYAREALIKE	-484.996±0.314	2	0.024±0.000	0.153±0.000	0.000±0.000	0.004±0.001	0.002
BAYAREALIKE+J	-460.390±0.200	3	0.021±0.000	0.066±0.001	0.019±0.000	0.775±0.007	0.790

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1.

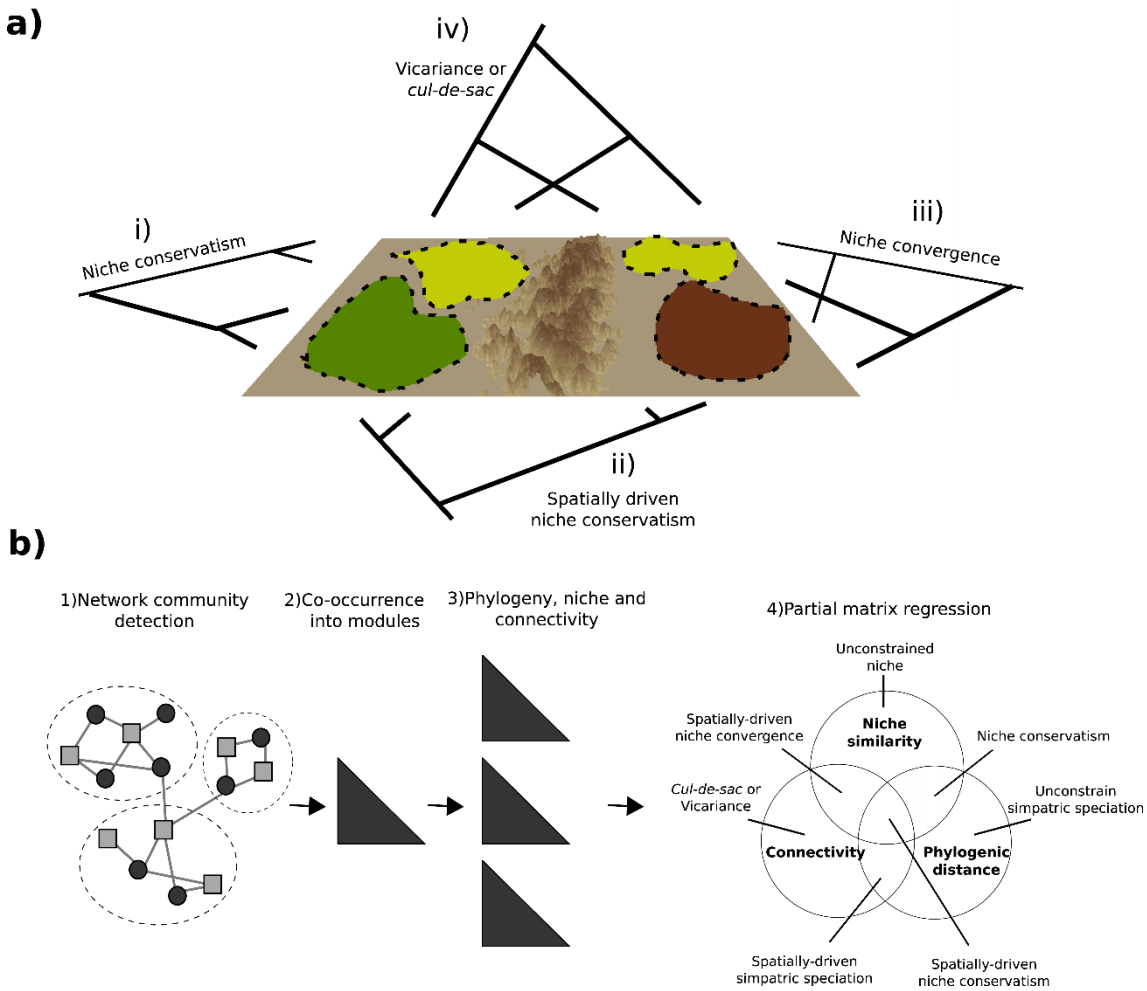

Figure 1. Identifying the factors configuring regional faunas. a) Four (out of seven, see b) hypothetical processes that may configure Regional faunas. Dotted lines depict different regions while colours correspond with different climates. In each case, the tips of the phylogeny point to regional distribution of the species. b) Workflow and potential results: 1) Hypothetical results of modularity analysis over the occurrence network; 2) similarity matrix of occurrence into modules; 3) pairwise matrix of environmental niche similarities; phylogenetic distances and topographical connectivity; and 4) hypothetical results and interpretations of a partial matrix regression on species occurrence similarities as a function of niche similarities, phylogenetic distances and connectivity.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2.

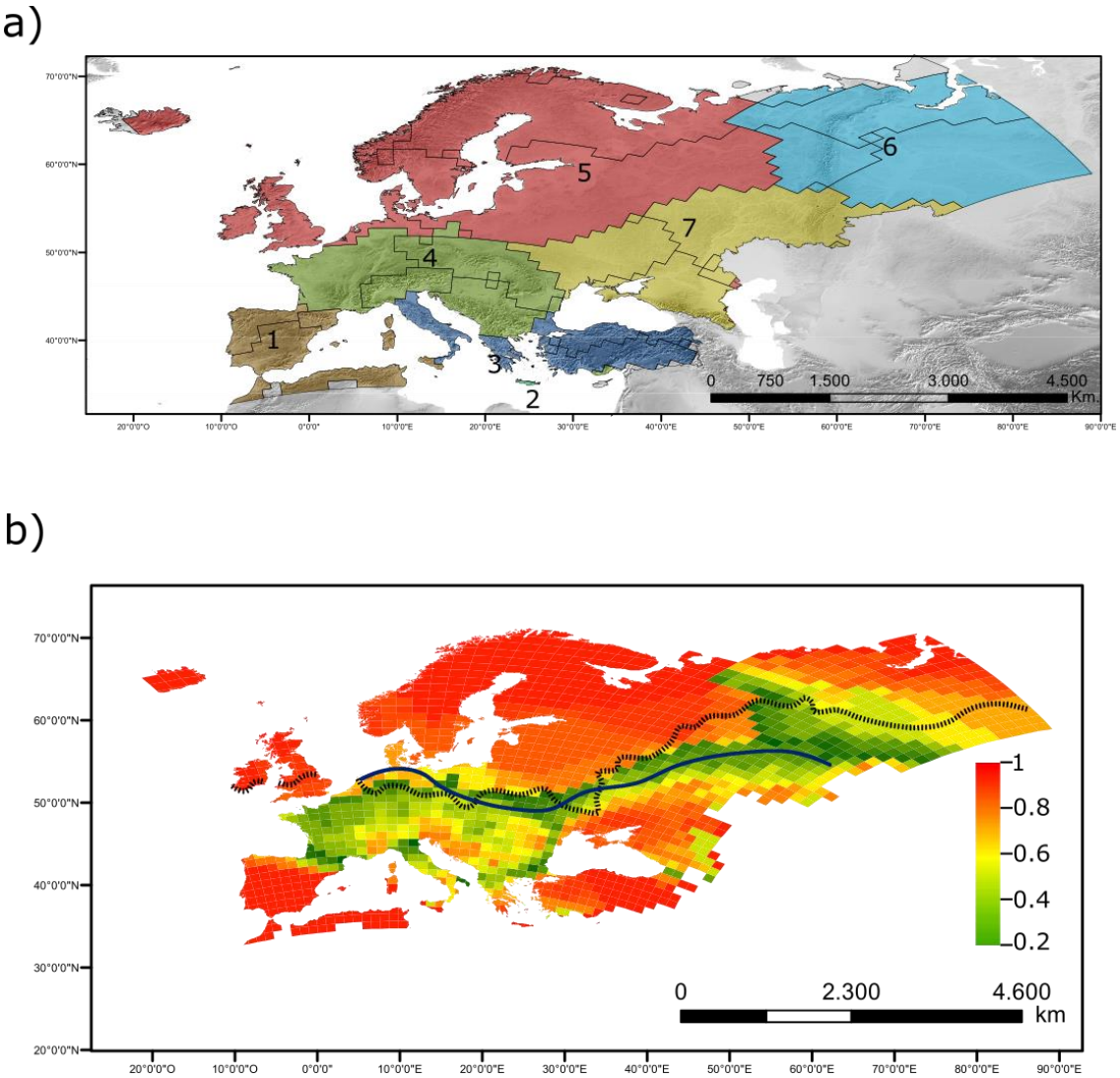

Figure 2. European *Carabus* regions found by the network community detection analysis. a) Geographical location of modules (i.e. regions) and submodules (i.e. sub-regions). b) Values of module affinity per grid cell; green colours (i.e. cells with low affinity) identify transition zones. The dotted black line corresponds with the southern limit of the ice sheet at LGM (extracted from Ehlers and Gibbard 2004). The blue line depicts the breakpoint where the temperature-*Carabus* richness relationship changes, as found in Calatayud et al. (2016b).

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3

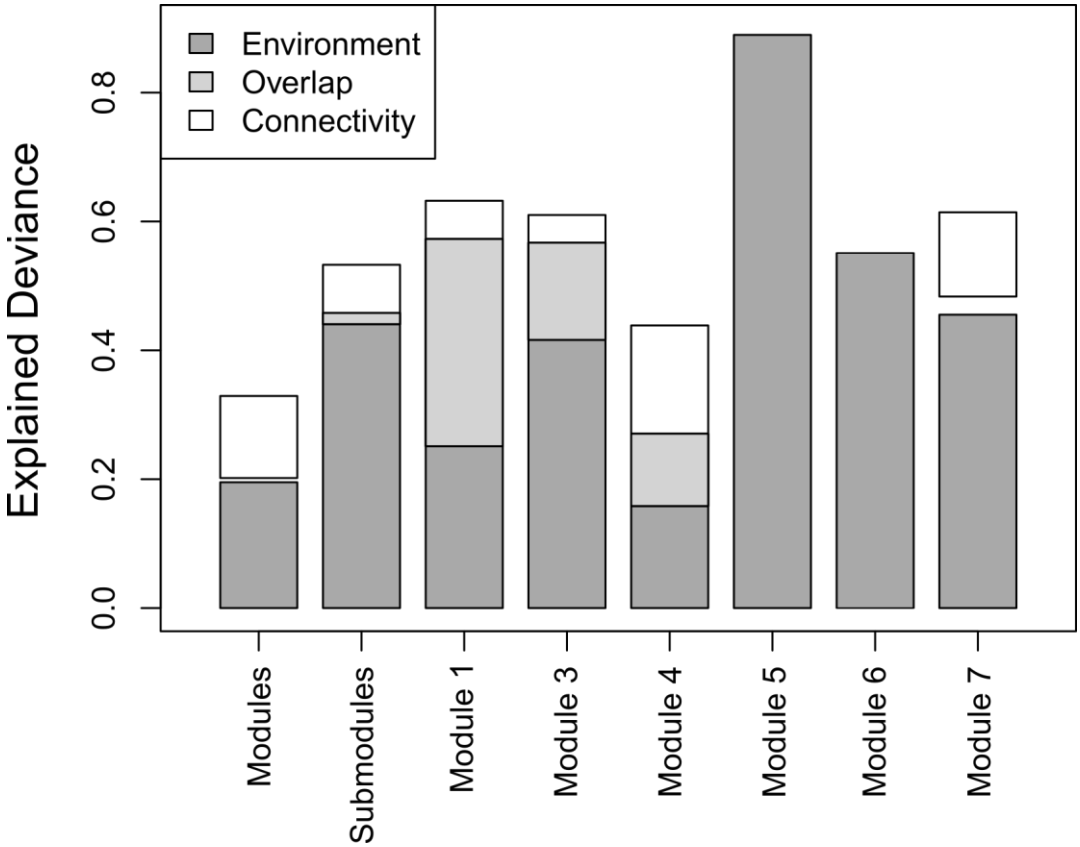

Figure 3. Results of the partial generalized matrix regression of similarity in regional co-occurrence, as a function of environmental niche similarity (climate and habitat), topographical connectivity and phylogenetic distances. The first and second bars correspond with the models including co-occurrence similarities among all modules and submodules, respectively. The remaining bars correspond with the models where the similarities in submodule occurrence were analysed independently for the species of each module.

bioRxiv preprint doi: https://doi.org/10.1101/149617; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4

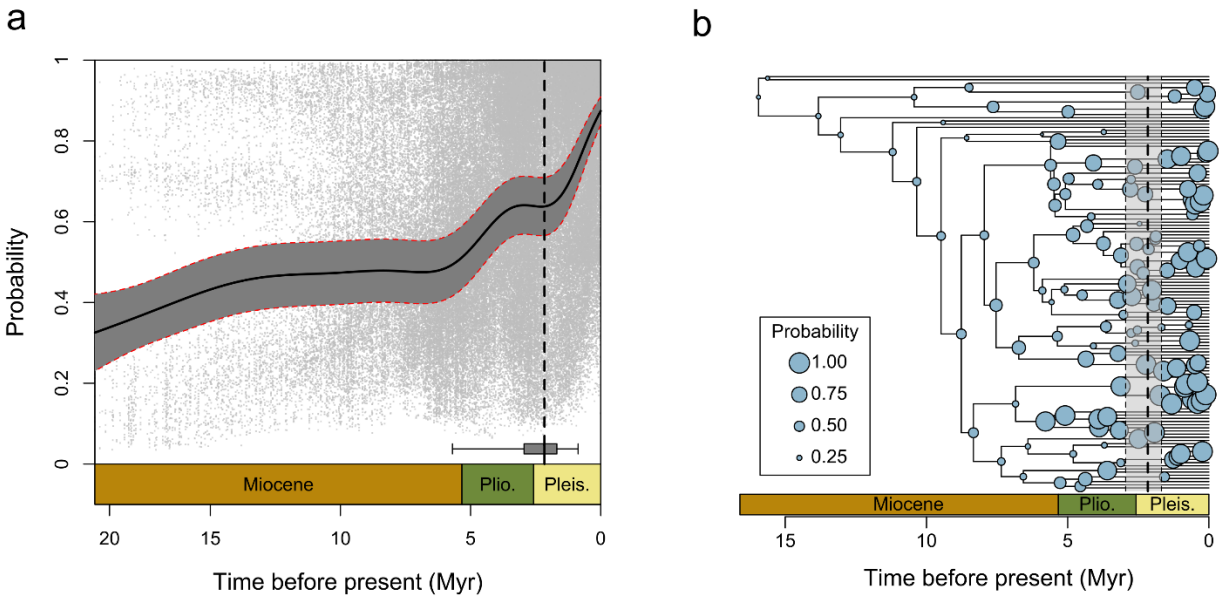

Figure 4. Results of the ancestral range estimation. a) GAMM predictions of the marginal probability as a function of node age for the BAREALIKE+J model. The shaded area corresponds with the interval confidence at 95%. The dotted black line represents the median of the breakpoint found by piecewise GLM regressions. The boxplot at the bottom represent the 45th and 55th percentile breakpoint values, whereas the whiskers depict the 25th and 75th percentiles. b) An example of a phylogenetic hypothesis where the BAREALIKE+J model was the best supported by the data. The size of the internal phylogenetic nodes is proportional to the marginal probability of the most probable state. The dotted line corresponds with the average breakpoint (median) in the relationship between node age and marginal probability. The shaded area depicts the 45th and 55th percentiles.