1   **Pleistocene climate change and the formation of regional species pools**

2   Calatayud, J.[1,2], Rodríguez, M. Á.[1], Molina-Vengas, R.[3], Leo, M.[4], and Hortal, J[2].

3

4   1 Departamento de Ciencias de la Vida, Universidad de Alcalá, Edificio de Ciencias, Ctra. Madrid-

5   Barcelona km. 33,6, 28871 Alcalá de Henares, Madrid, Spain.

6   2 Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales

7   (MNCN-CSIC), C/José Gutiérrez Abascal 2, 28006 Madrid, Spain.

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

9   4. Departamento de Biodiversidad y Conservación. Real Jardín Botánico de Madrid (CSIC).

10   28014 Madrid, España

11

12   **Abstract**

13   Despite the description of bioregions dates back from the origin of biogeography, the

14   processes originating their associated species pools have been seldom studied. Ancient

15   historical events are thought to play a fundamental role in configuring bioregions, but

16   the effects of more recent events on these regional biotas are largely unknown. We

17   use a network approach to identify regional and sub-regional faunas of European

18   *Carabus* beetles, and analyse the effects of dispersal barriers, niche similarities and

19   phylogenetic history on their configuration. We identify a transition zone matching the

20   limit of the ice sheets at Last Glacial Maximum. While southern species pools are

21   mostly separated by dispersal barriers, in the north species are mainly sorted by their

22   environmental niches. Strikingly, most phylogenetic structuration of *Carabus* faunas

23   occurred since the beginning of the Pleistocene. Our results show how extreme recent

24   historical events –such as Pleistocene climate cooling, rather than just deep-time

25   evolutionary processes, can profoundly modify the composition and structure of

26   geographic species pools.

27

28   **Introduction**

29   Naturalists have long been captivated by the geographic distribution of world biotas.

30   Rooted in the seminal ideas by Alexandre von Humbolt, this fascination has promoted

31   a long-term research agenda aiming to delineate biogeographic regions according to

32   their integrating faunas and floras (e.g. Wallace 1876, Holt et al. 2013, Rueda et al.

33   2013). Regional biotas are known to determine ecological and evolutionary dynamics

34   taking place at finer scales (Ricklefs 2008, 2015). For instance, regional species pools

35   can modulate local diversity patterns (Ricklefs 2011, Medina et al. 2014, Ricklefs and

36   He 2016), the structure and functioning of ecosystems (Naeslund and Norberg 2006),

37   or co-evolutionary processes (Calatayud et al. 2016a). However, despite their

38   fundamental importance, the processes that have configured regional biotas have

39   been seldom studied (and particularly the historical ones), and most explanations on

40   their origin and dynamics remain largely narrative (Crisp et al. 2011).

41   Perhaps the earliest speculations about the formation of regional species pools took

42   place during the flourishment of bioregionalizations in the mid-19th century (reviewed

43   by Ebach 2015). During that time, and beyond referring to physical factors (climate,

44   soils, and physical barriers), some authors already started to emphasize historical

45   influences as key elements determining the configuration of plant and animal regions.

46   For instance, when Wallace (1876) proposed his famous zoogeographic regions, he

47   argued that while the distribution of ancient linages such as genera and families would

48   more likely reflect major geological and climatic changes spanning the early and mid-

49   Cenozoic, that of species would be more influenced by recent events such as

50   Pleistocene glaciations (see Rueda et al. 2013). Indeed, these recent events could have

51   fostered many additions and subtractions of species to regional faunas through

52   dispersal and diversification processes. Increasing evidence suggests that Pleistocene

53   glacial-interglacial dynamics may have driven population extinction (e.g. Barnes et al.

54   2002), allopatric speciation in glacial refugia (e.g. Johnson et al. 2004) and post-glacial

55   recolonization events (e.g. Hewitt 1999; Theissinger et al. 2013). As a consequence,

56   Pleistocene glaciations are known to influence current diversity patterns for many

57   taxa, particularly in the Holarctic (e.g. Svenning and Skov 2007; Hortal et al. 2011;

58   Calatayud et al. 2016b). However, the role of Pleistocene glaciations in shaping

59   regional species pools remains largely unknown.

60

61   Besides historical contingencies, niche-based processes may also determine the

62   composition of regional species pools (Mittelbach and Schemske 2015), mainly

63   throughout their effects on species distribution ranges (Soberon 2007, Hortal et al.

64   2010, Hortal et al. 2012). These processes integrate responses to abiotic conditions

65   along geographical gradients and to local and regional biotic environment (Colwell et

66   al. 2009). Furthermore, they involve a complex interplay between evolutionary,

67   ecological and biogeographical factors. For instance, species with similar climatic

68   tolerances can coexist in regions of similar climate, but their arrival (and *in situ*

69   evolution) may be constrained by geographical barriers, which may also lead to

70   climatic-tolerance divergent species pools (Fig.1a). Further, if species' resemblance in

71   climatic tolerance is phylogenetically constrained, climate-driven regional species

72   pools will be also composed of evolutionary related species (i.e. phylogenetically

73   clustered species pools), although this effect is again filtered by biogeographical

74   processes (Gouveia et al. 2014). Indeed, diversification of lineages within regions

75   separated by strong dispersal barriers (e.g. mountain ranges) may also lead to

76   phylogenetically clustered species pools (Warren et al. 2014; Fig.1a). Over the mid-

77   term, historical contingencies may contribute to erase the signature of such

78   geographically-linked diversification. For example, differential diversification rates may

79   be the predominant driver of regional species pools in climatically stable regions

80   (Cardillo 2011), yet regions where Pleistocene glaciations exerted a greater influence

81   may harbour pools of species shaped by the joint effects of current climate and post-

82   glacial dispersion (Svenning et al. 2015).

83

84   In this study we aim to disentangle the relative importance of the processes that may

85   contribute to the formation of regional species pools, using European *Carabus*

86   (Coleoptera: Carabidae) as a model lineage. *Carabus* is a species-rich ground beetle

87   genus of great popularity due to the beautiful jewel-like appearance of some species

88   (Turin et al. 2003). In general, *Carabus* species are flightless nocturnal predators of

89   snails, earthworms and caterpillars. They hold hydrophilic adaptations and are typically

90   associated to deciduous forests (Deuve et al. 2012). Previous evidence suggests that

91   the richness of species from this genus in Europe is determined to a large extent by

92   both current environmental conditions (i.e. climate and habitat) and glacial-interglacial

93   dynamics (Calatayud et al. 2016b). This makes European *Carabus* an ideal case study to

94   evaluate the joint effects of evolutionary, ecological and historical contingency

95   processes as drivers of regional species pools.

96

97 Specifically, we use data on the distribution and evolutionary relationships of *Carabus*

98 species, and network and phylogenetic analyses, to evaluate six hypotheses: First,

99 given the presumed low dispersal capacity of the species from this genus, we

100 hypothesize that (H1) European *Carabus* species pools are mainly shaped by the main

101 orographic barriers of the continent, but also, that (H2) glacial-interglacial dynamics

102 have led to strong differentiation between northern and southern regional species

103 pools. If this differentiation is true, (H3) northern European *Carabus* faunas will be

104 composed of species adapted to cold climatic conditions that colonized newly vacant

105 habitats after the withdrawal of the ice sheet, and hence their regional distribution will

106 be mostly determined by current climate. In contrast, (H4) southern faunas will be

107 mainly shaped by the joint influence of diversification events and dispersal limitations,

108 due to the combined effect of higher climatic stability towards southern latitudes (e.g.

109 climatic refugia) and a more complex orography (Alps, Pyrenees, Carpathians).

110 Therefore, (H5) species forming northern regional pools will exhibit comparatively

111 lower levels of regional endemicity, whereas those forming southern regional pools

112 will show comparatively higher levels of regional fidelity. Finally, according to Wallace

113 (1876), the advance and retreat of the ice sheets during the Pleistocene should have

114 determined the spatial distribution of lineages, eroding the effects of the former

115 configuration of the distribution of the main *Carabus* lineages. Therefore, (H6) we

116 expect a temporal signal coincident with the Pleistocene in the phylogenetic structure

117 of *Carabus* faunas, and no effect of deep-time events on the current geographical

118 distribution these lineages.

119

120

121 **Material and methods**

122

123 **Data origin**

124 Data on the European distribution of *Carabus* species comes from a recent analysis on

125 the determinants of diversity of this genus in the continent (Calatayud et al. 2016b).

126 Briefly, expert-based range maps of all *Carabus* species inhabiting Europe (n = 131)

127 delineated by Turing et al. (2003) were digitized and overlaid into a 100-km equal-area

4

128   grid using ARCGIS (grid available at http://dataservice.eea.europa.eu). The resulting

129   gridded distribution maps were reviewed by several experts, eventually correcting cells

130   wrongly classified as either presences or absences (see Calatayud et al. 2016b for

131   further details). The resulting presence-absence matrix was used to represent the

132   distribution of *Carabus* species in all analyses.

133

134   We used environmental and physical geography GIS data from several commonly used

135   public domain digital repositories. Bioclimatic variables were extracted from Worldclim

136   (v1.4 Hijmans et al. 2005; available at http://www.worldclim.org/). Altitudinal data

137   were derived from the 30-arcsecond digital elevation model GTOPO30 provided by the

138   U. S. Geological Survey (available at https://lta.cr.usgs.gov/GTOPO30), and the location

139   of water bodies was extracted from vector information coming from Natural Earth

140   (available at http://www.naturalearthdata.com/). Finally, data for the geographical

141   distribution of forest habitats come from MODIS Land Cover (Channan et al. 2014,

142   available at http://glcf.umd.edu/data/lc/).

143

144   **Statistical analyses**

145   **Rationale and structure of the analyses**

146   Exploring the determinants of regional faunas requires analysing ecological,

147   evolutionary and historical factors jointly. We did so through three consecutive steps

148   (Fig.1b). First, we identified distinct regional species pools within Europe by using a

149   network community detection algorithm. From this analysis we derive a species

150   pairwise similarity matrix of occurrence into different modules that represent different

151   regions. Second, we assessed the relative importance of the environmental, spatial

152   and evolutionary determinants of such similarity. We constructed four pairwise

153   matrices to describe the relationships among species; namely, *i*) a matrix of climatic

154   niche similarity, *ii*) a matrix of habitat similarity, *iii*) a matrix of spatial connectivity

155   among their ranges, and *iv*) a phylogenetic distance matrix. Then, we used generalized

156   partial matrix regression to model the similarity in species occurrences as a function of

157   these four matrices (Fig.1b). We used this workflow to explore the factors involved in

158   the configuration of *Carabus* faunas either at the European scale (i.e. co-occurrence

159   patterns of European *Carabus* species across the delimited regions, hereafter "regional

160 scale") and independently (i.e. within regions co-occurrence patterns of *Carabus*

161 species within each delimited region, "hereafter sub-regional scale"). Finally, we also

162 applied ancestral range estimation analysis in order to identify the time period from

163 which ancestral areas are estimated with less uncertainty. By doing so, we aimed to

164 detect important historical periods contributing to the regional organization of

165 *Carabus* linages.

166

167 The interpretation of the joint and independent effects of explanatory matrices can

168 shed light on the different processes configuring regional faunas (see Fig.1a). In that

169 way, if niche similarities (i.e. represented by the climatic and habitat similarity

170 matrices) and phylogenetic distances altogether explained the regional co-occurrence

171 of species, then this could be interpreted as indicative of constrained niche evolution

172 (or a tendency to resemble ancestral niches) in shaping regional faunas (Fig.1a.i).

173 However, if spatial connectivity also accounted for part of this co-occurrence, this

174 would indicate that this niche conservatism pattern can be caused by geographical

175 constrains (Fig.1a.ii). Further, the independent effects of niche similarities together

176 with spatial connectivity can be more likely the consequence of a convergence of

177 climatic niches due to geographical isolation (Fig.1a.iii). Whereas the effects of

178 connectivity and phylogeny would be indicative of a primacy of intra-regional

179 speciation driven by geographical barriers. Niche similarities alone would point to an

180 unconstrained niche evolution shaping regional faunas, while phylogeny alone would

181 indicate a primacy of geographically unconstrained intra-regional speciation events.

182 Finally, either a *cul-de-sac* effect (i.e. the accumulation of species in past climatic

183 refugia) or a primacy of vicariant speciation events could lead to the existence of

184 independent effects of connectivity and regional co-occurrence (Fig.1a.iv).

185

186 **Identification of regional species pools**

187 We took advantage of community detection analysis —borrowed from network

188 theory— to identify *Carabus* regional species pools in Europe. To do so, we derived a

189 bipartite network from the presence-absence matrix, where species and grid cells

190 constitute two disjoint sets of nodes that are mutually connected according to the

191 presence of species in grid cells (e.g. Calatayud et al. 2016a). Then, we conducted a

192   modularity analysis using the index proposed by Barber (2007) and the Louvain

193   algorithm (Blondel et al. 2008, as implemented in the Matlab function "Gen Louvain",

194   available at http://netwiki.amath.unc.edu; Mucha et al. 2010) to identify the optimal

195   modular structure of the bipartite network. That is, the optimal groups of grid cells

196   that show similar *Carabus* species composition together with their associated species

197   (i.e. network modules). We performed a heuristic search for an optimal solution,

198   where this analysis was repeated iteratively to obtain 100 different modular solutions,

199   retaining the one that showed the highest modularity value. This optimal solution was

200   used to conduct all subsequent analyses, although all the solutions were qualitatively

201   similar. We evaluated the statistical significance of the modules by comparing its

202   associated modularity value to a null distribution of values (n = 100) where the original

203   presence absence matrix was randomized using the independent swap algorithm (a

204   fixed-fixed null model implemented in the R package "picante", Kembel et al. 2010). In

205   order to detect potential sub-modules nested within modules (i.e. sub-regional species

206   pools within regional species pools), we derived a new bipartite network from each of

207   the modules previously identified in the optimal solution, and applied the procedure

208   described above in each case.

209

210   It is important to note that despite species and grid cells were assigned to just one

211   module, they could participate with different degrees in other modules. For example,

212   despite most species in a grid cell will belong to the same module the cell does, this

213   cell could also hold species of other modules. Similarly, although a species will mostly

214   be present in cells assigned to its module, it may also occur in cells from other

215   modules. Thus, we calculated the degree of module affinity for each node of the

216   bipartite network as the proportion of links shown by a given node within its module

217   divided by the total number of its links (note that this index is related to the inter-

218   modular participation index of Guimera and Amaral 2005). Species with low module

219   affinity will tend to be widespread throughout Europe, belonging to different regional

220   species pools. While species with high module affinity will be mainly distributed within

221   their corresponding modules (highly endemic species). On the other hand, grid cells

222   with low module affinity may represent transition zones between regions.

223

224

Assessing the determinants of regional species pools

226 In order to disentangle the determinants of the current configuration of *Carabus*

227 faunas in Europe, we first generated a species-per-module matrix, where each inentry

228 of the matrix represents the percentage of the distributional range of a certain species

229 that lies in a given module. Then we derived a co-occurrence pairwise similarity matrix

230 from the former matrix using the Schoener index (Schoener 1970). This index describes

231 the degree of overlap between species pairs according to their distributions

232 throughout the modules (see Krasnov et al. 2012) for a previous application to similar

233 purposes). It ranges from 0 (i.e. complete lack of overlap) to 1 (i.e. identical pattern of

234 module distribution). Note that this similarity matrix reflects the co-occurrence

235 similarities at regional scale, ignoring lower-scaled distributional patterns (i.e. two

236 species may have identical regional distribution but differ in the grid cell they are

237 present). The resultant co-occurrence pairwise similarity matrix was used as

238 dependent variable. We generated four different pairwise similarity matrices to be

239 used as explanatory variables. Two of them to account for environmental factors,

240 namely (i) a climatic-niche similarity matrix, and (ii) a habitat similarity matrix; and the

241 other two for geographical and evolutionary factors, respectively: (iii) a spatial-

242 connectivity matrix, and (iv) a phylogenetic distance matrix.

243

244 *i) Climatic-niche similarity matrix*. We characterized the climatic niche of each *Carabus*

245 species in the dataset following the approach proposed by Broennimann et al. (2012).

246 We selected six bioclimatic variables to account for the main water and energy aspects

247 of climate –namely mean annual temperature, temperature of the warmest quarter,

248 temperature of the driest quarter, total annual precipitation, total precipitation of the

249 warmest quarter and total precipitation of the driest quarter– and altitudinal range to

250 account for the effects of mesoclimatic gradients within each grid cell. These variables

251 may be among the main determinants of the distribution of *Carabus* species diversity

252 within Europe (see Calatayud et al. 2016b). We conducted a principal components

253 analysis (PCA) on these variables to obtain a bidimensional climatic space defined by

254 the two main axes (81.4% of the variance, Fig.S2). Finally, we divided this climatic

255 space into 100 grid cells, and applied kernel smoothers to the species occurrence

8

256   densities in the gridded climatic space to calculate niche overlap between species

257   using the Schoener index (see above). Note that the kernel density function requires at

258   least five occurrences of the species, hence species occurring in less than five grid cells

259   were excluded from subsequent analysis (n = 12).

260

261   *ii) Habitat similarity matrix.* The distribution of *Carabus* species may also be shaped by

262   their forest preferences (Turin et al. 2003). Thus, we made use of information relative

263   to the type of vegetation where each species tends to occur in order to characterize

264   their environmental niche. We calculated the fraction of each species' distributional

265   range laying within each vegetation category, according to 10 vegetation categories

266   derived from MODIS dataset (Evergreen broadleaf forest, deciduous needle leaf forest,

267   deciduous broadleaf forest, mixed forest, closed shrub lands, open shrub lands, woody

268   savannas, savannas and grasslands). Then, we computed pairwise similarities in the

269   preference to different vegetation types using the Schoener index (see above).

270

271   *iii) Spatial connectivity matrix*. To evaluate the potential influence of dispersal barriers

272   on the current distribution of *Carabus* species we first created a dispersal-cost surface

273   by weighting each pixel in the study area according to both its topography (in this case,

274   slope) and the presence of water bodies. Slope values ranging from 0 to 100% at each

275   pixel and were determined from GTOPO30 altitudinal data using the GRASS tool

276   r.slope (GRASS Development Team 2017). Water bodies from Nature Earth were assigned

277   arbitrary values of friction to the dispersal of *Carabus* species, namely 30% for pixels

278   containing rivers and lakes and 99% for pixels that lay on sea water masses (note that

279   *Carabus* species show hydrophilic adaptations). Then, the connectivity between all

280   pairs of cells was calculated as least-cost path over the dispersal-cost surface that

281   connect both cells, using the "gdistance" R package (van Etten 2015). Finally, the spatial

282   connectivity between two species' distributional ranges was estimated as the average

283   distance among all grid cells within the range of each species. Average distances were

284   preferred over absolute least-cost distances to avoid disproportionate differences in

285   spatial connectivity between overlapping and non-overlapping distributional ranges.

286

287   *iv) Phylogenetic distance matrix*. To unravel the evolutionary history of *Carabus* lineage

288   and the importance of evolutionary processes in determining the formation of *Carabus*

289   species pool, we reconstructed a species-level time-calibrated molecular phylogeny

290   including 89 species. We used a maximum-likelihood inference based on ten DNA

291   markers (see Appendix S3 for full details). We used taxonomic information and

292   phylogenetic uncertainty methods (Rangel et al. 2015) to place species lacking

293   molecular information into the phylogeny (see Appendix S3). Thus, we derived 100

294   different phylogenetic hypotheses from the maximum-likelihood phylogeny, by

295   randomly inserting missing species within their most derived consensus clade based on

296   taxonomic knowledge.

297

298   We used multiple regression on distance matrices and variance partitioning to

299   disentangle the relative importance of climatic niche, habitat preferences, dispersal

300   barriers and evolutionary history in determining *Carabus* species pools in Europe. First,

301   we conducted single regressions between the co-occurrence pairwise similarity matrix

302   and each of the four explanatory matrices described above to seek for significant

303   associations between the variables. Given that the distribution of co-occurrence

304   pairwise values was rather bimodal with modes at 0 and 1, respectively, we set a

305   binomial family for error distribution and "logit" as the link function (see Ferrier et al.

306   2007 and Calatayud et al. 2016a for a similar approach). To assess for significance, we

307   randomized the original species per module matrix using the independent swap

308   algorithm (see above) to derive 999 null occurrence pairwise similarity matrices. Then,

309   we fitted single regressions between the null occurrence pairwise similarity matrices

310   and each of the explanatory similarity matrices, and considered a variable to have a

311   significant effect when the regression based on observed data explained a higher

312   deviance than 99% of the null regressions. In the case of phylogenetic pairwise

313   distances we repeated this procedure for each phylogenetic hypothesis to take into

314   account phylogenetic uncertainties, applying the same criterion for significance.

315   Finally, we retained those variables that showed significant associations with the co-

316   occurrence pairwise similarity matrix, and conducted partial multiple regression

317   (Legendre and Legendre 2012) on distance matrices to explore patterns of covariation

318   among niche similarities (i.e. climatic and habitat similarity matrices) dispersal barriers

10

319   and phylogenetic history. We repeated this whole procedure for each module

320   independently, to further explore the determinants of *Carabus* species co-occurrence

321   into submodules.

322

323   **Ancestral range estimation**

324   Probabilistic models of geographic range evolution have proven to be of great utility to

325   deal with historical biogeograhical questions (see Ronquist and Sanmartín 2011 and

326   references therein). Here we used these models to try to unravel whether, as

327   predicted by Wallace (1876), deep historical signals were eroded by Pleistocene

328   glaciations in the configuration of European *Carabus* faunas. We used a maximum

329   likelihood approach for inferring geographical range evolution implemented in the R

330   package BioGeoBears (Matzke 2014), using independently each one of the 100

331   phylogenetic hypotheses created before. BioGeoBears fits a variety of models that

332   differ in their rates and/or types of dispersal, extinction, sympatric speciation,

333   vicariance and founder event speciation. We fitted the six available models (DEC,

334   DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, BAYAREALIKE+J, see Matzke 2013)

335   comparing their informative capacity through the Akaike information criterion (AIC).

336   Species ranges were here coded as present/absent in each region detected in the

337   former network analyses.

338

339   The estimation of ancestral ranges usually tends to be more ambiguous in deeper

340   nodes of the phylogeny, as the potential lability of geographical ranges would tend to

341   blur deep time signals (Losos and Glor 2003). Yet, if the Pleistocene glacial periods had

342   important effects in the distribution of species it could be expected that ancestral

343   range estimations will begin to be more accurate around the Pleistocene (that is, pre-

344   Pleistocene signals on distributional range evolution will be eroded). To explore this we

345   attended to potential changes in the relationship between node age and the marginal

346   probability of the single most-probable ancestral state at each internal node. We

347   obtained these probabilities from the best model among the six geographical range

348   evolution models in BioGeoBears (see above) and in each of the phylogenetic

349   hypotheses used. Then, general additive mixed models (GAMMs) were fitted to the

350   node marginal probability as a function of node age, including the phylogenetic

11

351 hypothesis as a random factor. Further, in case we found different best models in

352 different phylogenetic hypotheses we also included a fixed variable with the model

353 type. Finally, we also carried out generalized linear mixed models (GLMMs) using a

354 piecewise regression procedure to assess for a shift in the relationship between

355 marginal probability and node age. The breakpoint in the piecewise regression was

356 assessed by including the breakpoint as a new parameter in the GLMM and then

357 minimizing the deviance of the fitted model using the function "optimize" in the R

358 package lme4 (Bates et al. 2014). Finally, we used AICc to compare the models with

359 and without breakpoint. In both cases we used a binomial family and a loglink

360 function, since the marginal probabilities of the nodes vary between zero and one. All

361 analyses were carried out in R (R core team 2015),: GAMM with the package gamm4

362 (Wood and Scheipl 2014), GLMM with Lme4 (Bates et al. 2014) and AICc with

363 AICcmodavg (Mazerolle 2011).

364

365 **Results**

366 **Identification of regional faunas**

367 The *Carabus* occurrence network was significantly modular (M=0.385, p=0.01), dividing

368 Europe in seven modules that group zoogeographically distinct regions with their

369 associated faunas; that is, different regional species pools (Figs. 2 and S1). Further, all

370 modules but module 2 showed a significant sub-modular structure, presenting a

371 decrease in modularity with latitude (mean M=0.316, ranging from 0.154 to 0.468; all

372 p-values < 0.05, see Table S1). Module 1 holds 21 species mainly living in Southwestern

373 Palearctic (Iberian Peninsula, North of Africa, Balearic Islands, Corsica, Sardinia and the

374 western half of Sicilia). This module was subdivided into four submodules. Module 2

375 only groups two species which are endemic of Crete. Module 3 identifies an East

376 Mediterranean region including the Italic Peninsula, part of Greece and Turkey. This

377 module holds 18 species and was subdivided into five submodules. Module 4 depicts a

378 Central European region embracing the Alps and Carpathian Mountains, as well as

379 Central European plains. This module is the richest one, with forty-nine *Carabus*

380 species, and was split into 4 submodules. Module 5 and module 6 correspond with

381 northern regions and are the poorest ones, holding ten species each. The former

382 expands from Iceland and the British Isles almost to the Ural Mountains. The latter

12

383   includes The Ural Mountains and expands to the easternmost zone of the study area.

384   Both modules were divided into 3 submodules. Finally, module 7 includes 21 species

385   and embraces a south-eastern central European region expanding from the Carpathian

386   Mountains to the south Ural Mountains. This module was split into 3 submodules.

387

388   The transition zones between regions appear to be associated with geographical

389   barriers such as the Pyrenees, Alps, Carpathian and Ural Mountains, as well as the

390   Bosphorus strait (Fig. 2c). Further, we also found a transition zone running through

391   Central Europe in a west-to-east belt, separating southern and northern regions.

392   Interestingly, this transition zone closely follows the southern limits of the ice sheet at

393   the Last Glacial Maximum (LGM), suggesting a link between the configuration of

394   *Carabus* regional faunas and Pleistocene glacial conditions.

395

396

397   **Correlates of regional co-occurrence**

398   Matrix regressions showed that both environmental niche similarity (i.e. climate and

399   habitat) and spatial connectivity were significantly related to species co-occurrence in

400   regions and sub-regions (p<0.01). In contrast, evolutionary relatedness was not related

401   with regional co-occurrence, as phylogenetic distances did not show a significant

402   relationship at p>0.01 for any single phylogenetic hypothesis. Environmental niche

403   similarity was the best predictor of co-occurrence in both regions and sub-regions,

404   although its effects were stronger when predicting the co-occurrence into sub-regions

405   (Fig. 3). Spatial connectivity had stronger effects when predicting the co-occurrence

406   into regions relative to sub-regions, where its effect mostly overlapped with that of

407   environmental niche. The relative importance of these two sets of factors showed a

408   clear latitudinal trend. The effects of connectivity were stronger in southern regions

409   (i.e. modules 1 and 3), which is not surprising given the more complex orography of

410   these regions. But importantly, although niche similarity was the best predictor of co-

411   occurrence in all cases, its effects became more evident towards northern regions.

412   Indeed, niche similarities and spatial connectivity were significant predictors of the co-

413   occurrence into sub-regions in all regional pools except the two northern ones (Table

414   S2). Only niche similarities (mainly climate) showed significant effects in these two

415   modules (i.e. 5 and 6), as expected if the species in these regional pools colonized the

416   northern regions only recently, and were geographically sorted according to current

417   climate.

418

419   **Ancestral range estimation**

420   BAYAREALIKE+J model was the best supported by data in 59% of the phylogenetic

421   hypotheses, followed by DEC+J and DEC, which were the best in 23 and 18% of the

422   hypotheses, respectively (Table 1). Interestingly, the BAYAREALIKE+J model does not

423   account for vicariant cladogenetic events. Rather, it considers founder effect

424   speciation, which could be seen as a signal of Pleistocene glaciations promoting

425   speciation in glacial refugia (i.e. acting as islands) and eroding the ancestors' ranges.

426   Indeed, according to this model the range contraction parameter had the highest

427   estimate (Table 1). This parameter was however of little importance in the other two

428   models, which accounted for vicariant events in both cases. Regardless the

429   biogeographical model used, GAMM results showed that node marginal probability of

430   the most probable state increased as expected towards younger nodes (P<0.01, R^2

431   =0.34). However, this increase became abrupt coinciding with the beginning of the

432   Pleistocene – which started around a 2.59 Mya.– (Fig.4). Indeed, we found that the

433   GLMM including the breakpoint was better in terms of AICc (AICc weight for the model

434   including the breakpoint = 1), and that the relationship between state probability and

435   node age changed at 2.47 Mya. (confidence interval at 95% = 2.23, 2.73). This points to

436   the signal left by the Pleistocene on the configuration of European *Carabus* faunas.

437   These results were confirmed using different approaches (see Appendix S4).

438

439

440   **Discussion**

441   More than 140 years ago, Wallace (1876) foresaw that the influence of Pleistocene

442   glaciations on the distribution of diversity had been strong enough so as to erode the

443   imprint of previous events. Our results confirm Wallace's thoughts, showing a

444   remarkable coincidence between the distribution of the ice sheet at the Last Glacial

445   Maximum and the current configuration and evolutionary structure of European

446   *Carabus* Faunas. There is a close spatial relationship between the southern limits of the

447   area glaciated at LGM, and the transition zone separating the southern and northern

448   regions identified by our analyses. Indeed, this border also coincides with the line

449   identified by Calatayud et al. (2016b), where the relationship between *Carabus* species

450   richness and current climate changes in shape and strength (Fig. 2). Thus, it seems that

451   the climate changes underwent during the Pleistocene not only shape local diversity

452   patterns (e.g. Svenning and Skov 2007, Araujo et al. 2008, Hortal et al. 2011), but they

453   have also left a strong imprint on the geographical structure of species composition at

454   a regional scale. Accordingly, the species from the northernmost region (module 5)

455   show the lowest level of endemism (Fig. S2, as expected for regional faunas composed

456   of species that have recently colonized the north of Europe from southern glacial

457   refugia; Araujo et al. 2008, Calatayud et al. 2016b). In fact, although these species are

458   widely distributed across southern Europe, all of their ranges only overlap near to the

459   northern Carpathian Mountains, an area that has been shown to be a glacial refugia

460   for some *Carabus* species (Homburg et al. 2013, Fig. S3). Additionally, the decrease in

461   modularity values with latitude also points to a lesser geographical structure of

462   northern assemblages, which can be interpreted as the result of a post-glacial

463   colonization.

464

465   Besides the Pleistocene effects in the definition and geographical structure of regional

466   pools of species, we also found evidence of the imprint of this geological period on the

467   processes configuring the distribution of *Carabus* faunas. The generally strong

468   relationship between regional patterns of co-occurrence and both niche similarities

469   and spatial connectivity shows that co-occurring species tend to have similar realized

470   environmental niches and that also tend to be geographically constrained by the same

471   dispersal barriers. This latter result was expected, given the low dispersal capacity of

472   *Carabus* species (see Turin et al. 2003), which is likely to be behind the spatial

473   coincidence of the zones of transition between regional faunas and geographical

474   barriers. Perhaps more unexpected is the relationship of regional co-occurrence and

475   niche similarities without any significant phylogenetic effect. This implies that the

476   geographical configuration of barriers to dispersal has restricted species within regions

477   of similar climate, rather than climatic-niche conservatism constraining their co-

478   occurrences. These results point to that *Carabus* niche evolution is, to some extent,

15

479   evolutionary unconstrained, which is congruent with the high adaptation capacity of

480   insects in general (e.g. Overgaard and Sørensen 2008). Going further, this questions to

481   what extent the observed species' regional occurrence is the consequence of their

482   environmental niche and not the other way around (see Hortal et al. 2012, Wüest et al.

483   2015).

484

485   Whatever the origin of the relationship between species occurrence and

486   environmental conditions, which is certainly true is that its strength changes among

487   regions. These changes follow a latitudinal gradient in the importance of

488   environmental niche similarities. In northern regions, the similarity in the realized

489   niche seems to be stronger related to the occurrence into sub-regions than in the

490   south. This might be a direct consequence of the processes that determine regional

491   pools being dependent on the particular history of each region (Ricklefs 2015). These

492   findings are also congruent with the effects of post-glacial colonization, where

493   formerly glaciated areas show a clear sorting of species northwards due to low time for

494   dispersal (Svenning and Skov 2007). This contrasts with southern regions, where

495   climate has been milder and more stable and the species have had more time to

496   diversify. This suggests that species' geographical arrangement in southern latitudes is

497   more likely the result of historical contingency rather than environmental preferences

498   (Hortal et al. 2011, Calatayud et al. 2016b).

499

500   The lack of relationship between the phylogenetic distances among *Carabus* species

501   and their regional co-occurrence can be the outcome of either a generalized

502   preponderance of vicariant events and/or a "cul-de-sac effect" (O'Regan 2008). The

503   former will more likely imply that the generation of the dispersal barriers that shaped

504   the regions will also promote the geographical split of many lineages and subsequent

505   allopatric speciation events (Weeks et al. 2016). Yet, the formation of the geographical

506   accidents associated to the delimitation of *Carabus* regions largely predates the origin

507   of the genus (see Beccaluva et al. 1998, Deuve et al. 2012). On the other hand, a

508   generalized dispersion into climatic refugia, together with a subsequent stagnancy

509   within them (i.e. a "cul-de-sac" effect) may also produce the observed mixing of

510   unrelated linages into regions. Although it is difficult to distinguish between both

511   processes, we believe it is more plausible to think that southern regions have

512   accumulated unrelated species while acting as glacial refugia, whereas northern

513   regions were recolonized by unrelated species with similar environmental niches ---or

514   simply with higher dispersal capacity (see above).

515

516   In agreement with this idea, the model of ancestral range estimation that is best

517   supported by the data does not include vicariant events. On the contrary, this model

518   takes into account founder speciation events, a process typical from islands (Provine

519   1989). This could be interpreted on the light of southern regions acting as islands

520   during the different Pleistocene glacial maxima. Should this be true, a temporal signal

521   of this period on the spatial organization of *Carabus* lineages should be evident in their

522   phylogeny. We found such signal on the relationship between node maximum

523   probability state and node age. Indeed, the striking coincidence between the

524   breakpoint where this relationship becomes steeper and the beginning of the

525   Pleistocene argues in favour of the imprint of this epoch on the configuration of

526   *Carabus* regional faunas. We obtained similar results using several other approaches

527   (Appendix S4), supporting the notion that the current regional organization of *Carabus*

528   species and linages has its very roots at the beginning of the Pleistocene. This contrasts

529   with ancestral range estimations for clades inhabiting areas that were never glaciated,

530   where more ancient signals were found in the spatial sort of lineages (Condamine et al.

531   2015, Economo et al. 2015, Tänzler et al. 2016, Toussaint and Balke 2016). It thus

532   seems that the repeated advances and retreats of ice sheets during this geological

533   period produced the repeated cycles of retreat to southern regions and advance

534   towards northern regions of *Carabus* species, a hustle-and-bustle process that

535   ultimately led to the observed mixing of unrelated lineages.

536

537   In sum, our results provide solid arguments in favour of the importance of Pleistocene

538   glaciations along with geographical barriers in structuring the regional faunas of this

539   group. On the one hand, European *Carabus* faunas are primarily delimited by the

540   location of the southern limit of the ice sheet at LGM, which separates two large

541   regions that differ not only in species composition, but also in the processes underlying

542   the spatial organization of these species. On the other hand, the phylogenetic

543   structure of these faunas coincides with the beginning of the Pleistocene. This not only

544   implies that the geographical distribution of species and lineages is deeply shaped by

545   past climates, but also that the ecological processes (Naeslund and Norberg 2006,

546   Madrigal et al. 2016) and evolutionary mechanisms (Wüest et al. 2015, Calatayud et

547   al. 2016a) that are dependent on regional species pools may be profoundly affected by

548   the history of Earth climates. It is therefore essential to take into account past

549   historical events while trying to understand, not only current diversity patterns and the

550   processes behind them, but also processes and patterns occurring at local scales.

551

References

553 1. Araújo, M. B., D. Nogués-Bravo, J. A. F. Diniz-Filho, A. M. Haywood, P. J. Valdes,
554   and C. Rahbek. 2008. Quaternary climate changes explain diversity among reptiles
555   and amphibians. Ecography **31**:8–15.
556 2. Barber, M. J. 2007. Modularity and community detection in bipartite networks.
557   Physical Review E **76**:066102.
558 3. Barnes, I., P. Matheus, B. Shapiro, D. Jensen, and A. Cooper. 2002. Dynamics of
559   Pleistocene population extinctions in Beringian brown bears. Science **295**:2267–
560   2270.
561 4. Bates, D., M. Maechler, B. Bolker, S. Walker, et al. 2014. lme4: Linear mixed-
562   effects models using Eigen and S4. R package version **1**.
563 5. Beccaluva, L., M. Shallo, M. Coltorti, I. Premti, and F. Siena. 1998. Encyclopedia of
564   European and Asian Regional Geology. Springer.
565 6. Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding
566   of communities in large networks. Journal of Statistical Mechanics **2008**:P10008.
567 7. Broennimann, O., M. C. Fitzpatrick, P. B. Pearman, B. Petitpierre, L. Pellissier, N. G.
568   Yoccoz, W. Thuiller, M.-J. Fortin, C. Randin, N. E. Zimmermann, et al. 2012.
569   Measuring ecological niche overlap from occurrence and spatial environmental
570   data. Global Ecology and Biogeography **21**:481–497.
571 8. Calatayud, J., J. L. Hórreo, J. Madrigal-González, A. Migeon, M. Á. Rodríguez,
572   S. Magalhães, and J. Hortal. 2016*a*. Geography and major host evolutionary
573   transitions shape the resource use of plant parasites. Proceedings of the National
574   Academy of Sciences USA **113**:9840–9845.
575 9. Calatayud, J., J. Hortal, N. G. Medina, H. Turin, R. Bernard, A. Casale, V. M. Ortuño,
576   L. Penev, and M. Á. Rodríguez. 2016*b*. Glaciations, deciduous forests, water
577   availability and current geographical patterns in the diversity of European *Carabus*
578   species. Journal of Biogeography **43**:2343–2353.
579 10. Cardillo, M. 2011. Phylogenetic structure of mammal assemblages at large
580   geographical scales: linking phylogenetic community ecology with macroecology.
581   Philosophical Transactions of the Royal Society of London B **366**:2545–2553.
582 11. Channan, S., K. Collins, and W. Emanuel. 2014. Global mosaics of the standard
583   MODIS land cover type data. University of Maryland and the Pacific Northwest
584   National Laboratory, College Park, Maryland, USA .

585   12. Colwell, R. K., and T. F. Rangel. 2009. Hutchinson's duality: the once and future
586   niche. Proceedings of the National Academy of Sciences USA **106**:19651–19658.
587   13. Condamine, F. L., E. F. Toussaint, A.-L. Clamens, G. Genson, F. A. Sperling, and G. J.
588   Kergoat. 2015. Deciphering the evolution of birdwing butterflies 150 years after
589   Alfred Russel Wallace. Scientific Reports **5**:11860.
590   14. Crisp, M. D., S. A. Trewick, and L. G. Cook. 2011. Hypothesis testing in
591   biogeography. Trends in Ecology and Evolution **26**:66–72.
592   15. Deuve, T., A. Cruaud, G. Genson, and J.-Y. Rasplus. 2012. Molecular systematics
593   and evolutionary history of the genus *Carabus* (Col. Carabidae). Molecular
594   Phylogenetics and Evolution **65**:259–275.
595   16. Ebach, M.C. 2015. Origins of Biogeography. Springer, New York.
596   17. Economo, E. P., E. M. Sarnat, M. Janda, R. Clouse, P. B. Klimov, G. Fischer, B. D.
597   Blanchard, L. N. Ramirez, A. N. Andersen, M. Berman, et al. 2015. Breaking out of
598   biogeographical modules: range expansion and taxon cycles in the hyperdiverse
599   ant genus *Pheidole*. Journal of Biogeography **42**:2289–2301.
600   18. Ehlers, J., and P. L. Gibbard. 2004. Quaternary glaciations-extent and chronology:
601   part I: Europe. Elsevier.
602   19. Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized
603   dissimilarity modelling to analyse and predict patterns of beta diversity in regional
604   biodiversity assessment. Diversity and Distributions **13**:252–264.
605   20. Gouveia, S. F., J. Hortal, M. Tejedo, H. Duarte, F. A. Cassemiro, C. A. Navas, and
606   J. A. F. Diniz-Filho. 2014. Climatic niche at physiological and macroecological
607   scales: the thermal tolerance–geographical range interface and niche
608   dimensionality. Global Ecology and Biogeography **23**:446–456.
609   21. GRASS Development Team, 2017. Geographic Resources Analysis Support System
610   (GRASS GIS) Software, Version 7.2. Open Source Geospatial Foundation.
611   http://grass.osgeo.org.
612   22. Guimerà, R., and L. A. N. Amaral. 2005. Functional cartography of complex
613   metabolic networks. Nature **433**:895–900.
614   23. Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. Biological
615   Journal of the Linnean Society **68**:87–112.
616   24. Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high
617   resolution interpolated climate surfaces for global land areas. International
618   Journal of Climatology **25**:1965–1978.
619   25. Holt, B. G., J.-P. Lessard, M. K. Borregaard, S. A. Fritz, M. B. Araújo, D. Dimitrov, P.-
620   H. Fabre, C. H. Graham, G. R. Graves, K. A. Jønsson, et al. 2013. An update of
621   Wallace's zoogeographic regions of the world. Science **339**:74–78.
622   26. Homburg, K., C. Drees, M. M. Gossner, L. Rakosy, A. Vrezec, and T. Assmann. 2013.
623   Multiple glacial refugia of the low-dispersal ground beetle *Carabus irregularis*:
624   molecular data support predictions of species distribution models. PloS One
625   **8**:e61185.
626   27. Hortal, J., P. De Marco Jr, A. Santos, and J. A. F. Diniz-Filho. 2012. Integrating
627   biogeographical processes and local community assembly. Journal of
628   Biogeography **39**:627–628.
629   28. Hortal, J., J. A. F. Diniz-Filho, L. M. Bini, M. Á. Rodríguez, A. Baselga, D. Nogués-
630   Bravo, T. F. Rangel, B. A. Hawkins, and J. M. Lobo. 2011. Ice age climate,

19

631   evolutionary constraints and diversity patterns of European dung beetles. Ecology
632   Letters **14**:741–748.

633 29. Hortal, J., J. M. Lobo, and A. Jiménez-Valverde. 2012. Basic questions in
634   biogeography and the (lack of) simplicity of species distributions: putting species
635   distribution models in the right place. Natureza & Conservação **10**:108–118.

636 30. Hortal, J., N. Roura-Pascual, N. Sanders, and C. Rahbek. 2010. Understanding
637   (insect) species distributions across spatial scales. Ecography **33**:51.

638 31. Johnson, N. K., C. Cicero, and K. Shaw. 2004. New mitochondrial DNA data affirm
639   the importance of Pleistocene speciation in North American birds. Evolution
640   **58**:1122–1130.

641 32. Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D.
642   Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating
643   phylogenies and ecology. Bioinformatics **26**:1463–1464.

644 33. Krasnov, B. R., M. A. Fortuna, D. Mouillot, I. S. Khokhlova, G. I. Shenbrot, and
645   R. Poulin. 2012. Phylogenetic signal in module composition and species
646   connectivity in compartmentalized host-parasite networks. The American
647   Naturalist **179**:501–511.

648 34. Legendre, P., and L. F. Legendre. 2012. Numerical ecology. Elsevier.

649 35. Losos, J. B., and R. E. Glor. 2003. Phylogenetic comparative methods and the
650   geography of speciation. Trends in Ecology and Evolution **18**:220–227.

651 36. Madrigal-González, J., P. Ruiz-Benito, S. Ratcliffe, J. Calatayud, G. Kändler,
652   A. Lehtonen, J. Dahlgren, C. Wirth, and M. A. Zavala. 2016. Complementarity
653   effects on tree growth are contingent on tree size and climatic conditions across
654   Europe. Scientific Reports **6**:32233.

655 37. Matzke, N. J., 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood)
656   Evolutionary Analysis in R Scripts. University of California, Berkeley, Berkeley, CA.
657   http://CRAN.R-project.org/package=BioGeoBEARS.

658 38. Matzke, N. J. 2014. Model selection in historical biogeography reveals that
659   founder-event speciation is a crucial process in island clades. Systematic Biology
660   **63**:51–70.

661 39. Mazerolle, M. J. 2011. AICcmodavg: model selection and multimodel inference
662   based on (Q) AIC (c). R package version **1**.

663 40. Medina, N. G., B. Albertos, F. Lara, V. Mazimpaka, R. Garilleti, D. Draper, and
664   J. Hortal. 2014. Species richness of epiphytic bryophytes: drivers across scales on
665   the edge of the Mediterranean. Ecography **37**:80–93.

666 41. Mittelbach, G. G., and D. W. Schemske. 2015. Ecological and evolutionary
667   perspectives on community assembly. Trends in Ecology and Evolution **30**:241–
668   247.

669 42. Mucha, P. J., T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. 2010.
670   Community structure in time-dependent, multiscale, and multiplex networks.
671   Science **328**:876–878.

672 43. Naeslund, B., and Norberg, J. 2006. Ecosystem consequences of the regional
673   species pool. *Oikos*, **115***:504-512.

674 44. O'Regan, H. J. 2008. The Iberian Peninsula–corridor or cul-de-sac? Mammalian
675   faunal change and possible routes of dispersal in the last 2 million years.
676   Quaternary Science Reviews **27**:2136–2144.

45. Overgaard, J., and J. G. Sørensen. 2008. Rapid thermal adaptation during field temperature variations in *Drosophila melanogaster*. Cryobiology **56**:159–162.

46. 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.

47. R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

48. 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.

49. Ricklefs, R. E. 2008. Disintegration of the ecological community. The American Naturalist **172**:741–750.

50. 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.

51. Ricklefs, R. E. 2015. Intrinsic dynamics of the regional community. Ecology Letters **18**:497–503.

52. 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.

53. Ronquist, F., and I. Sanmartín. 2011. Phylogenetic methods in biogeography. Annual Review of Ecology, Evolution and Systematics **42**:441–464.

54. Rueda, M., M. Á. Rodríguez, and B. A. Hawkins. 2013. Identifying global zoogeographical regions: lessons from Wallace. Journal of Biogeography **40**:2215–2225.

55. Schoener, T. W. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. Ecology **51**:408–418.

56. Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters **10**:1115–1123.

57. 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.

58. 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.

59. 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.

60. 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.

61. 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.

62. 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,

723   L. Penev, and A. Casale, editors. The Genus *Carabus* in Europe: A Synthesis.
724   Pensoft Publishers.

725 63. van Etten, J., 2015. gdistance: Distances and Routes on Geographical Grids.
726   http://CRAN.R-project.org/package=gdistance.

727 64. Wallace, A. R. 1876. The geographical distribution of animals: with a study of the
728   relations of living and extinct faunas as elucidating the past changes of the earth's
729   surface. Cambridge University Press.

730 65. Warren, D. L., M. Cardillo, D. F. Rosauer, and D. I. Bolnick. 2014. Mistaking
731   geography for biology: inferring processes from species distributions. Trends in
732   Ecology and Evolution **29**:572–580.

733 66. Weeks, B. C., S. Claramunt, and J. Cracraft. 2016. Integrating systematics and
734   biogeography to disentangle the roles of history and ecology in biotic assembly.
735   Journal of Biogeography **43**:1546–1559.

736 67. Wood, S., and F. Scheipl, 2014. gamm4: Generalized additive mixed models using
737   mgcv and lme4. http://CRAN.R-project.org/package=gamm4.

738 68. Wüest, R. O., A. Antonelli, N. E. Zimmermann, and H. P. Linder. 2015. Available
739   climate regimes drive niche diversification during range expansion. The American
740   Naturalist **185**:640–652.

741

742   **Table 1.** Results of the best ancestral range estimation models. d= dispersion, e=

743   extinction; j= founder speciation. Prop.best = proportion of phylogenetic hypotheses

744   where a model was the best in terms of AICc. Mean values among different

745   phylogenetic hypotheses (incl. standard deviation within brackets), are shown for each

746   parameter.

Model	N° param.	d	e	j	Prop. Best
BAYAREALIKE+J	3	0.015(0.001)	0.063(0.017)	0.015(0.005)	0.59
DEC+J	3	0.025(0.001)	<0.001(<0.001)	0.008(0.003)	0.23
DEC	2	0.026(0.001)	0.003(0.001)	0	0.18
BAYAREALIKE	2	0.019(0.001)	0.111(0.006)	0	0
DIVALIKE	2	0.031(0.001)	0.002(0.001)	0	0
DIVALIKE+J	3	0.03(0.001)	0(0)	0.006(0.003)	0

747

748

749   **Figure 1.** Hypothetical examples of the factors configuring regional faunas and work

750   flow. a) Figure showing four hypothetical processes configuring regional faunas.

751   Dotted lines depict different regions while colours correspond with different climates.

752   In each case, the tips of the phylogeny point to regional distribution of the species. b)

753   Workflow and potential results: 1) Hypothetical results of modularity analysis over the

754   occurrence network; 2) similarity matrix of occurrence into modules; 3) pairwise

755   matrix of environmental niche similarities; phylogenetic distances and topographical

756   connectivity; and 4) hypothetical results and interpretations of a partial matrix

757   regression on species occurrence similarities as a function of niche similarities,

758   phylogenetic distances and connectivity.

759

760   **Figure 2.** European *Carabus* regions found by the network community detection

761   analysis. a) Geographical location of modules (i.e. regions) and submodules. b) Module

762   simplification of the occurrence network. Circles represent a module, being their size

763   proportional to the species group within them. Links depict the species shared among

764   regions, being its width proportional to the number of species. c) Values of module

765   affinity per grid cell; green colours (i.e. cells with low affinity) identify transition zones.

766   The dotted black line corresponds with the southern limit of the ice sheet at LGM

767   (extracted from Ehlers and Gibbard 2004). The blue line depicts the breakpoint where

768   the temperature--*Carabus* richness relationship changes, as found in Calatayud et al.

769   (2016b).

770

771   **Figure 3.** Results of the partial generalized matrix regression of similarity in regional

772   co-occurrence, as a function of environmental niche similarity (climate and habitat),

773   topographical connectivity and phylogenetic distances. The first and second bars

774   correspond with the models including all regions and subregions, respectively. The

775   remaining bars correspond with the models where the similarities in subregional

776   occurrence were analysed independently for the species of each region. Con.= spatial

777   connectivity; Envir. = environmental niche similarities; Overl.= Overlap.

778

779   **Figure 4.** Results of the ancestral range estimation. a) Marginal probability of the most

780   probable state at each internal node for a phylogenetic hypothesis where the

781   BAREALIKE+J model was the best supported by the data. The size of the circles depicts

782   the value of probability, while the colours represent the combinations of regions

783   forming the state. The phylogeny is labelled at subgenera level. b) GAMM predictions

784   of the marginal probability as a function of node age for the BAREALIKE+J model. The

785   shaded area corresponds with the interval confidence at 95%. The dotted black line

786   represent the breakpoint found by picewise GLMM regression. The interval confidence

787   at 95% for the breakpoint is also provided. See Fig. S5 for the predictions of the other

788   models.

789

790   **Figure 1.**
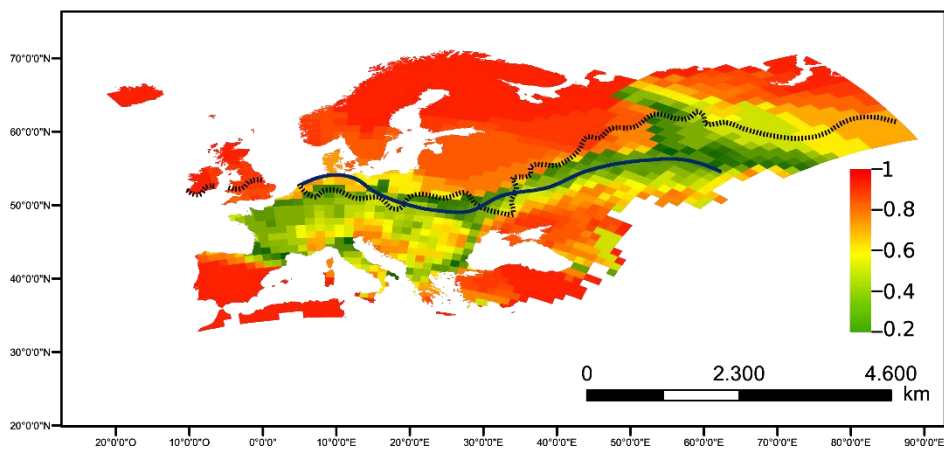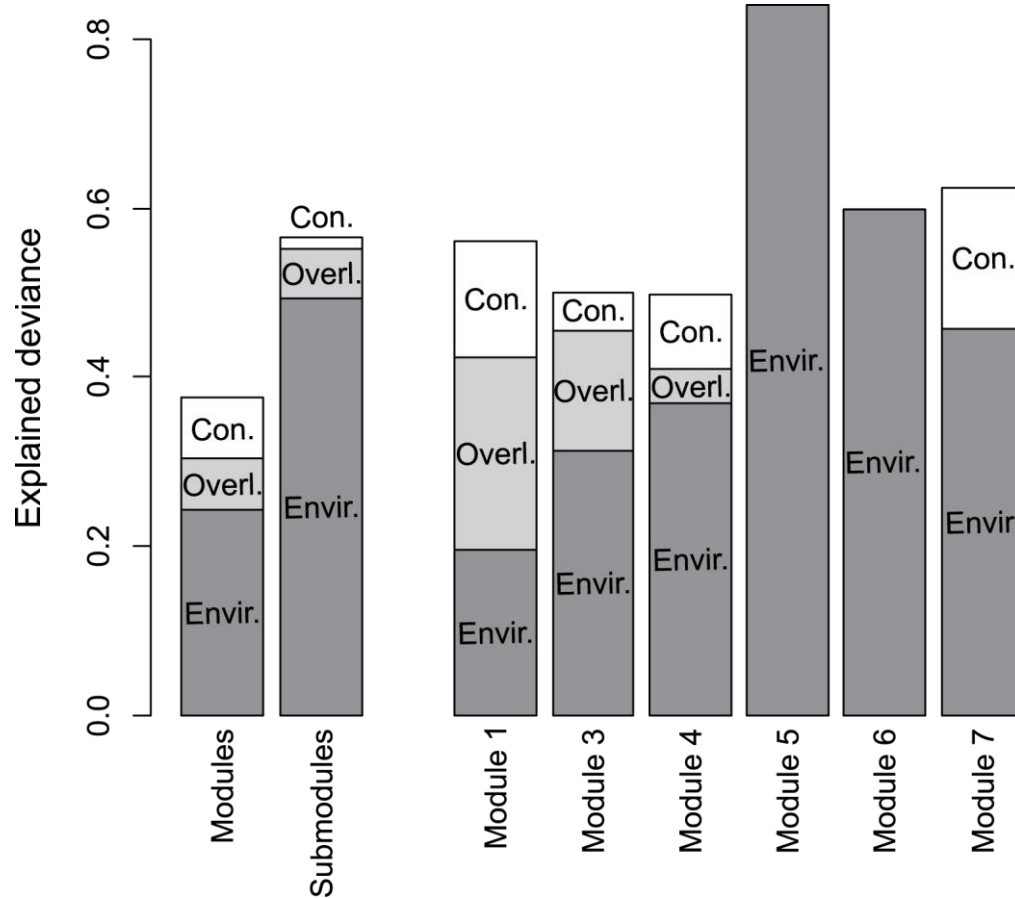
791

792
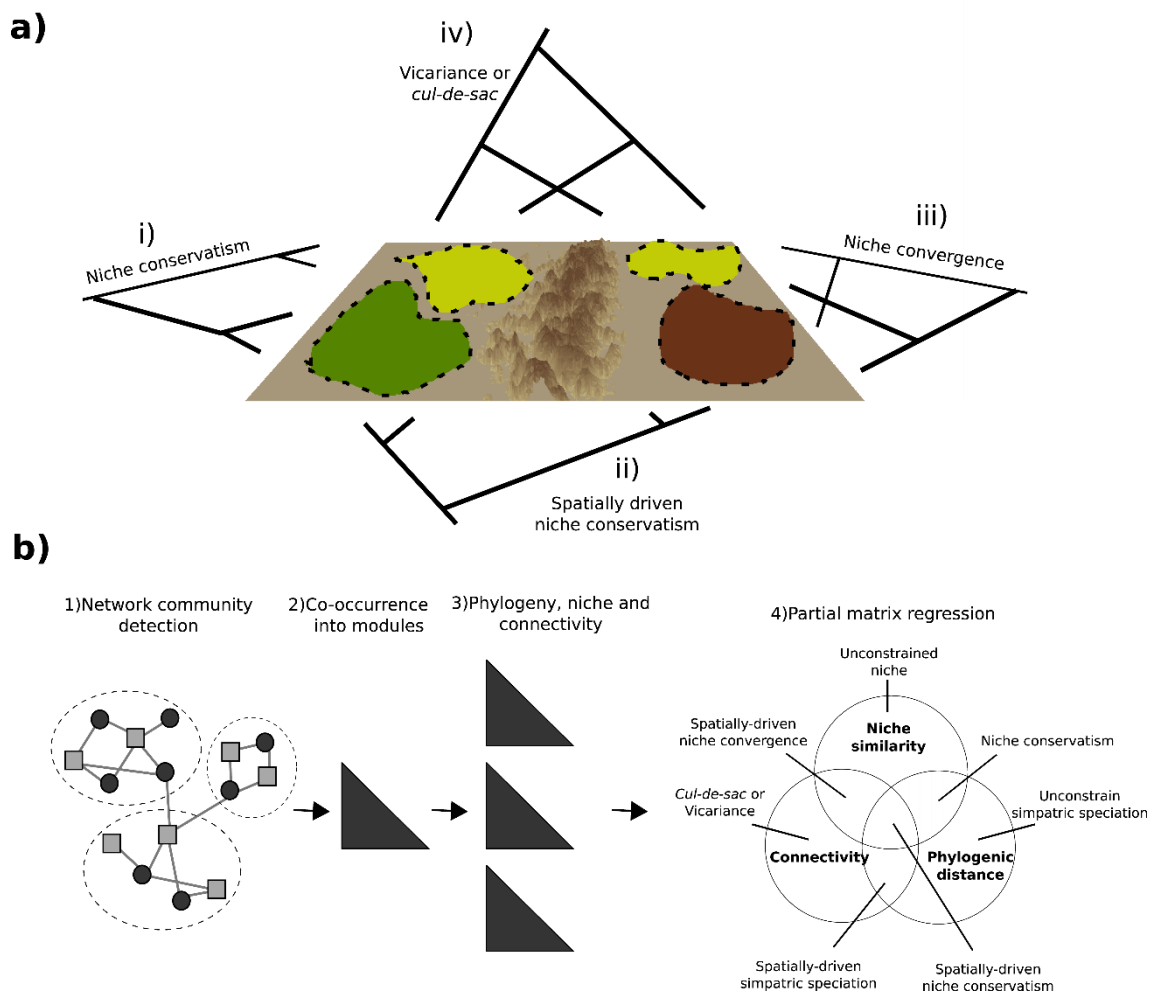
793

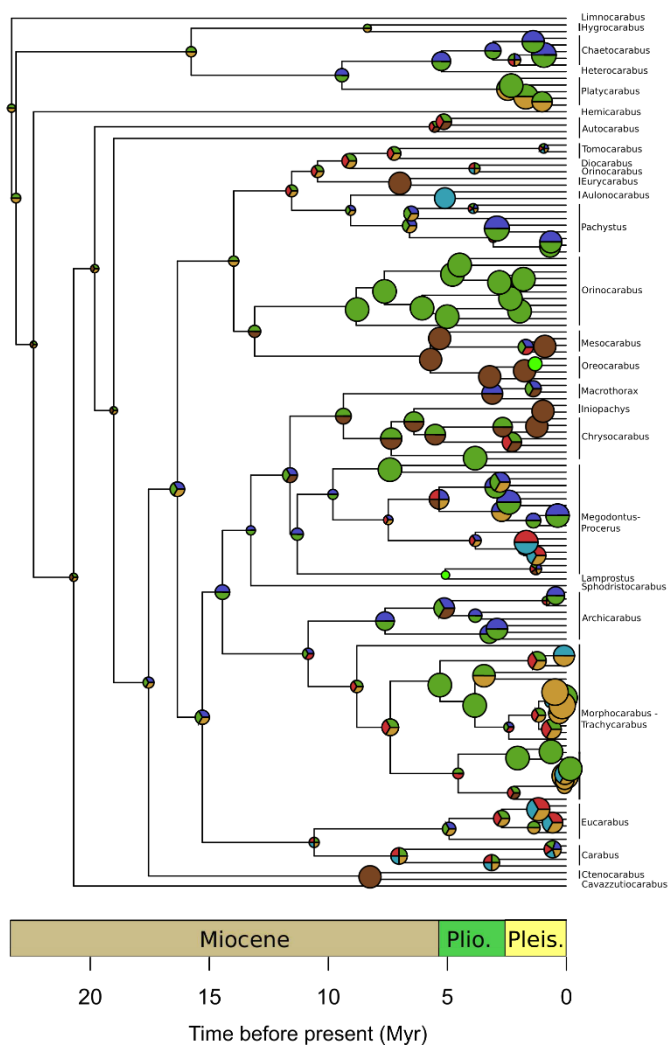794 **Figure 2.**

a)
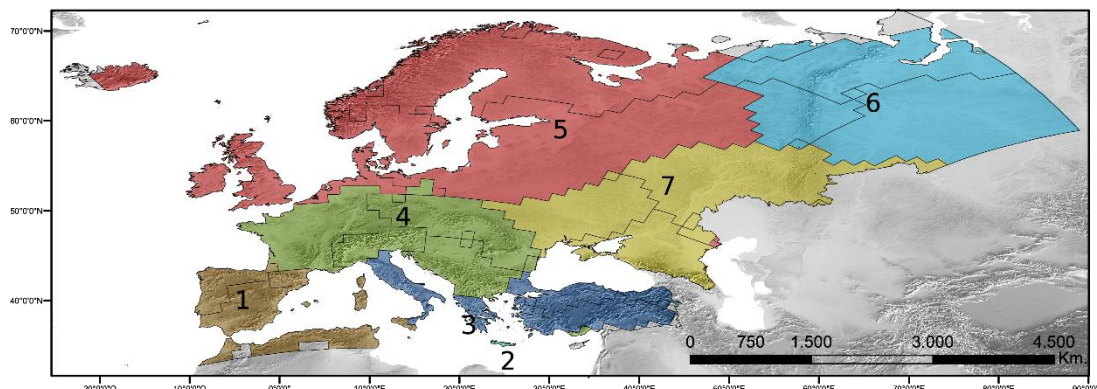
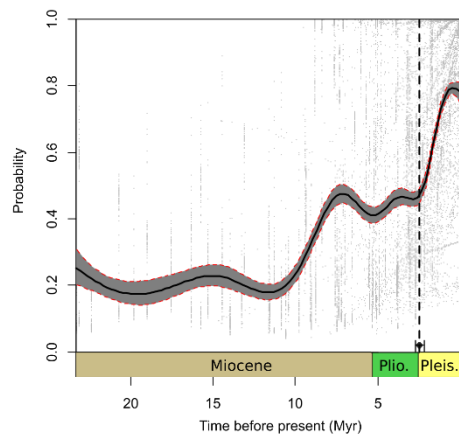
b)
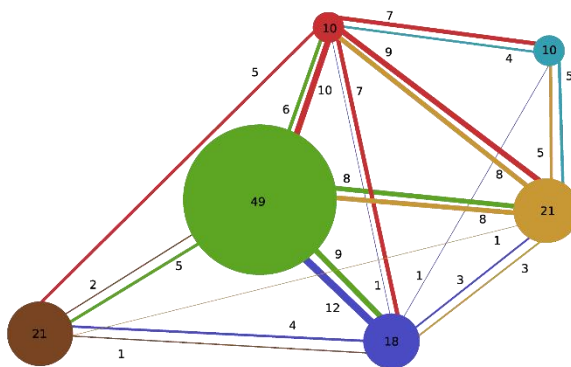
c)

795

796

797 **Figure 3**

798

799

800   **Figure** 4.

a)

b)

801