

1 **New distribution data and phylogenetic approach reveal bioregionalization of the western**
2 **Palaearctic ants**

3
4 **Running title**

5 Bioregionalization of European ants

6
7 **Authors and affiliations:**

8 Runxi Wang^{1, *}, Jamie M. Kass², Christophe Galkowski³, Federico Garcia⁴, Matthew T. Hamer¹,
9 Alexander Radchenko⁵, Sebastian Salata⁶, Enrico Schifani⁷, Zalimkhan M. Yusupov⁸, Evan P.
10 Economo^{2,9} and Benoit Guénard¹

11
12 1 School of Biological Sciences, The University of Hong Kong, Kadoorie Biological Sciences
13 Building, Pok Fu Lam Road, Hong Kong SAR, China.

14 2 Biodiversity and Biocomplexity Unit, Okinawa Institution of Science and Technology
15 Graduate University, Onna, Okinawa, Japan.

16 3 Société Linnéenne de Bordeaux, 1 place Bardineau, 33000, Bordeaux, France.

17 4 Iberian Myrmecological Association, Barcelona, Spain.

18 5 Schmalhausen Institute of Zoology of the National Academy of Sciences of Ukraine, 15
19 Bogdana Khmel'nitskogo str., 01030, Kiev, Ukraine.

20 6 Department of Biodiversity and Evolutionary Taxonomy, University of Wrocław,
21 Przybyszewskiego 65, PL-51148 Wrocław, Poland.

22 7 Department of Chemistry, Life Sciences and Environmental Sustainability, University of
23 Parma, Parco Area delle Scienze 11/a, 43124 Parma, Italy.

24 8 Tembotov Institute of Ecology of Mountain Territories of the Russian Academy of Science,
25 I. Armand Street 37-A, Nalchik city, 360051, Kabardino-Balkarian Republic, Russia.

26 9 Radcliffe Institute of Advanced Study, Harvard University, Cambridge, MA, USA

27 * Corresponding: runxiwg@connect.hku.hk

28
29 **Acknowledgements**

30 RW is supported by an Early Career Scheme Grant from the Research Grants Council (ECS-
31 27106417) of the Hong Kong Government. JMK is supported by the Japan Society for the
32 Promotion of Science (JSPS) Postdoctoral Fellowships for Foreign Researchers Program; AR
33 is supported by the grant NRFU (Ukraine) No. 2020/02/0369. This work can not be done
34 without the close and extensive collaborations across borders, we want to thank the peace we
35 had in Eurasia and hope we will still have it in the future!

36

37 **Abstract and keywords**

38 **ABSTRACT**

39 **Aim:** Biogeographic regionalization has fascinated biogeographers and ecologists for centuries
40 and is endued with new vitality by evolutionary perspectives. However, progress is scant for
41 most insect groups due to shortfalls in distribution and phylogenetic information, namely
42 Wallacean and Darwinian shortfalls respectively. Here, we used the western Palearctic ants as
43 the case to tackle these shortfalls and test their biogeographic structure through novel
44 distribution data and phylogenetic approaches.

45 **Location:** Western Palearctic realm.

46 **Taxon:** Ants (Formicidae).

47 **Methods:** Firstly, we developed a refined database integrating the occurrences of 747 ant
48 species across 207 regions of the western Palearctic realm, based on newly expert-validated
49 records derived from the existing global ant biodiversity informatics. Using range estimates for
50 these species derived from polygons and species distribution modelling, we produced species
51 assemblages in 50×50 km grid cells. We calculated taxonomic and phylogenetic turnover of
52 ant assemblages, performing hierarchical clustering analysis using the Simpson dissimilarity
53 index to delineate biogeographic structure.

54 **Results:** At both the regional list- and grid assemblage-levels, the Mediterranean has higher
55 turnover and more biogeographic regions than northern Europe, both taxonomically and
56 phylogenetically. Delineations based on grid assemblages detected more detailed
57 biogeographic transitions, while those based on regional lists showed stronger insularity in
58 biogeographic structure. The phylogenetic regionalization suggested closer but varied affinities
59 between assemblages in comparison to the taxonomic approach.

60 **Main conclusions:** Here, we integrated expert-validated regional lists, species distribution
61 modelling, and a recent phylogeny to tackle Wallacean and Darwinian shortfalls for an
62 important insect group by developing a next-generation map of biogeographic regionalization
63 for the western Palearctic ants. The results of this study suggest strong constraints from
64 geographic barriers and potential effects of climatic history on ant distributions and
65 evolutionary history, and also provide baseline spatial information for future investigations of
66 regional insect distributions.

67

68 **KEYWORDS**

69 biogeography, regionalization, beta diversity, Europe, Formicidae, insect, species distribution
70 modelling, beta diversity, phylogenetic turnover

71

72 **INTRODUCTION**

73 Organisms are not uniformly distributed on Earth—most are restricted to particular areas due
74 to ecological and evolutionary processes, ultimately forming distinct biogeographic regions
75 (Wallace, 1876; Lomolino et al., 2016). The concept of bioregionalization is thus fundamental
76 to the classification of species distributions (Wallace, 1894; Kreft & Jetz, 2010; Morrone, 2018)
77 and provides a powerful framework for informing empirical studies in ecology and evolution,
78 as well as applied studies in conservation and management (Wallace, 1876; Kreft & Jetz, 2010;
79 Holt et al., 2013). For centuries, biogeographers have proposed various regionalization systems
80 mainly based on plants and terrestrial vertebrates (e.g., Wallace, 1876; Smith, 1983; Cox, 2001;
81 Kreft & Jetz, 2010; Holt et al., 2013), but similar classification efforts based on insects, which
82 represent an overwhelming component of biodiversity (Stork, 2018), have been scant.

83

84 This taxonomic bias in regionalization may raise potential issues for the general representativity
85 of described biogeographic patterns. First, insects have evolved complex and unique life
86 histories and strategies (e.g., parasitism) (Gullan & Cranston, 2014). Moreover, mounting
87 evidence has shown that since the late Quaternary, the distributions of many vertebrate groups
88 have been shaped substantially by targeted anthropogenic selection pressures like hunting and
89 domestication (Faurby & Svenning, 2015; Santos et al., 2020), which for the most part have
90 not historically had direct effects on insects. Instead, likely direct drivers of insect distributions
91 over this time period have been climate, geology, and land cover change (Elias, 1991; Rueda,
92 Rodríguez & Hawkins, 2010; Ballesteros-Mejia et al., 2017). Due to these differences,
93 expanding our knowledge of biogeographic patterns to include insects would enhance, and
94 potentially validate, our current understanding of biogeographical regionalization.

95

96 The lack of detailed distributional information for insects (i.e., Wallacean Shortfall, Lomolino,
97 2004; Guénard et al., 2017), however, represents a major challenge to studying insect
98 regionalization. Most of the existing insect regionalization systems are based on regional lists,
99 which represent the most widely used geographic units and the most comprehensive
100 distributional information for insect taxa (e.g., Dennis, Williams & Shreeve, 1991; Heiser &
101 Schmitt, 2010; Vitali & Schmitt, 2017). These regional lists provide useful biogeographical
102 information, but their geographic units are normally taxon-specific, delineated as expert range
103 polygons with varying detail and accuracy that are limited to sampled areas. Therefore, their
104 boundaries and ranges may lack quantitative validation and be limited by sampling bias,
105 resulting in regionalization patterns with potential limitations for cross-taxa comparisons
106 (Rueda et al., 2010). Species distribution models (SDMs), which estimate relationships between
107 species' occurrence localities and environmental variables to make predictions of range extents,
108 provide a potential solution for tackling the Wallacean Shortfall for taxa with few distributional

109 data such as insects (Diniz-Filho, De Marco Jr, & Hawkins 2010; Ballesteros-Mejia et al., 2017,
110 Kass et al. 2020). Species distribution modeling represents a data-driven approach to generating
111 reproducible range estimates and making predictions for unsampled areas (Peterson et al. 2018).
112 Although SDMs for low-data species in particular can be susceptible to sampling bias,
113 statistical overfitting, and other methodological issues (Galante et al., 2017), recently proposed
114 methods for tuning model complexity (Radosavljevic & Anderson 2014) and accounting for
115 sampling bias (Phillips et al. 2009) can help remedy many of these issues.

116

117 Along with a lack of distributional information for insects, there is also a dearth of phylogenetic
118 information (i.e., the Darwinian Shortfall, Lomolino, 2004), representing another big challenge
119 for understanding insect biogeography. Historically, the delineation of biogeographic regions
120 has been based solely on taxonomic information, but in recent years, progress in phylogenetic
121 methods (i.e., phylogenetic regionalization) has provided new opportunities to explore
122 evolutionary relationships for entire assemblages and enhance the objectivity and repeatability
123 of delineations (Holt et al., 2013; Daru et al., 2017; Ye et al., 2019). Especially for
124 taxonomically challenging and complex groups like insects, the phylogenetic approach is
125 attractive because it provides an inclusive measurement for taxonomic classification based on
126 phylogenetic distance. However, the role that evolutionary history has played in regionalization
127 remains rarely explored for insects (Diniz-Filho et al., 2010).

128

129 Ants (Formicidae) represent good model organisms to investigate the biogeography of insects
130 due to their wide distributions across climates and biomes, but also because of their varied and
131 dominant roles in ecosystems as keystone species (Hölldobler & Wilson, 1990; Guénard et al.,
132 2017). Modern biodiversity informatics and recent research on ant macroevolution have
133 resulted in new data on their distributions and phylogeny (Economo et al., 2018; Guénard et al.,
134 2017; Kass et al., submitted). Although new species are continuously being described across
135 the globe and significant effort directed towards sampling and identification remains needed,
136 ants within the West Palearctic realm are relatively well-documented and are arguably among
137 the best-known ant faunas on the globe.

138

139 The biogeography of the western Palearctic realm has been studied for centuries and its insect
140 distributions are thought to have been shaped by climate change since the Last Glacial
141 Maximum. Climate refugia in the Mediterranean region harbor high species endemism and
142 show high heterogeneity in species composition, while species distributions in the more
143 northern regions have a more homogenous structure (Dennis et al., 1991; Fattorini & Ulrich,
144 2012; Vitali & Schmitt, 2017) and are suggested to be driven by postglacial dispersal processes
145 (Hewitt, 1999; Schmitt, 2007; Calatayud et al., 2019). Previous western Palearctic

146 regionalization systems have mainly focused on a few taxa including butterflies (e.g., Dennis
147 et al., 1991; Rueda, Rodríguez & Hawkins, 2010), dragonflies (e.g., Heiser & Schmitt, 2010;
148 Heiser, Dapporto & Schmitt, 2014) and some beetle taxa (e.g., Fattorini & Ulrich, 2012; Vitali
149 & Schmitt, 2017). All of them, however, lack fine-scale distributional data or a phylogenetic
150 understanding. Here, we developed a novel dataset including fine-scale regional lists, SDMs,
151 and a large-scale phylogeny of the western Palearctic ants to tackle these shortfalls and to
152 delineate the most comprehensive regionalization of insects in the western Palearctic realm to
153 date (Figure S2.1).

154

155 First, we evaluated the biogeographic structure of western Palearctic ants based on both
156 regional lists and grid assemblages generated from SDMs. We expected that both sets of results
157 would have similar spatial patterns, but that the grid assemblage delineation would better
158 identify transitions between biogeographic regions and the uniqueness of ant assemblages. We
159 also expected the SDM results would help fill in gaps by making predictions for areas with
160 limited sampling, thus reflecting more objective spatial extents of ant distributions. Secondly,
161 we tested whether the phylogenetic regionalization results in different patterns from those
162 obtained with the taxonomic approach. A historical connection between regions based on
163 shared evolutionary history may fail to be detected under a taxonomic delineation (Ye et al.,
164 2019). For example, allopatric speciation events driven by geographical isolation can increase
165 the dissimilarity of species composition between two regions yet a phylogenetic affinity
166 between them would remain (Daru et al., 2017). We thus expect the relationships between
167 regions resulting from phylogenetic regionalization to differ from those of the taxonomic
168 delineation, especially for geographically isolated species, and that these differences will help
169 to reveal a more multilayered regional evolutionary history for the western Palearctic ants.

170

171 **METHODS**

172 **Ant distributions and phylogeny**

173 *European Ant Distribution (EUAD) database*

174 We developed the EUropean Ant Distribution (EUAD) database, a new collection of ant species'
175 occurrence data for the western Palearctic realm. This database is derived from the Global Ant
176 Biodiversity Informatics (GABI, Guénard et al., 2017) but features a higher spatial resolution
177 for the region. Ultimately, we compiled native ant taxa occurrence information for each of the
178 207 geographic divisions (i.e., regional lists) for the western Palearctic realm, which was
179 previously divided to 57 regions in GABI. Our definition of the western Palearctic realm does
180 not include North Africa and the Arabian Peninsula because of historically poor sampling and
181 the lack of recent taxonomic revisions for species in those regions. Geographical divisions used
182 in the database were delimited based on either administrative region (GADM, version 2.8,

183 accessed 1st Sep. 2020) or modified areas based on the physical geographic area (e.g., islands
184 and mountains), depending on data availability (details can be found in Figure S2.2).

185

186 *Validation of regional lists*

187 Preliminary versions of the database were validated by ant experts (co-authors of this study)
188 who identified dubious records and provided additional information (e.g., unpublished or
189 missing records) to complete and provide more accurate ant range maps. Occurrence records
190 were deemed dubious for reasons including nomenclatural changes in recent taxonomic
191 revisions, outdated taxonomy, and misidentifications (which can be numerous in older literature
192 or databases). Ultimately, the validation process showed that 8% of species incidence records
193 in the preliminary versions were dubious (and either corrected or excluded for later analysis)
194 and also contributed 16% new incidence records to the final database (Figure S2.3). For all ant
195 taxa in our database, we also verified nomenclature based on AntCat, an online, global catalog
196 of ants (Bolton, 2021), with validation and inclusion of taxa up to July 1st 2021. Here, we treated
197 valid subspecies as species in our analysis, which resulted in a total of 747 valid native species
198 (including 40 subspecies) for regional lists (see Table S1.1).

199

200 *Grid assemblages*

201 We first made grid-based estimates of the western Palearctic ant assemblages using range
202 estimates from SDMs developed for a global analysis on ant diversity (Kass et al., submitted).
203 All analyses were performed using the statistical computing language R 4.0.2 (R Core Team,
204 2020). Ranges were estimated for low-data species (<5 occurrence records) with univalue
205 polygons (either buffered [30 km] points or convex/alpha hull, depending on data availability),
206 and for species with sufficient data (≥ 5 occurrence records) using SDMs. We used the
207 presence-background machine-learning algorithm Maxent to train models over a study extent
208 defined by their polygon range estimate (buffered alpha hull) using 19 bioclimatic predictor
209 variables at 10 arcminute resolution (~20 km at the equator) from Worldclim 2.0 (Fick &
210 Hijmans 2017). We tuned models for optimal complexity (i.e., combinations of feature classes
211 and regularization multipliers) using sequential criteria of cross-validation results (based on the
212 10 percentile omission rate and validation AUC; Radosavljevic & Anderson 2014) with the R
213 package *ENMeval* 2.0.0 (Kass et al. 2021). We used these tuned models to make predictions of
214 suitability over the species' study extents, effectively constraining range estimates to the limits
215 of the occurrence data, and made them binary (presence/absence predictions) by thresholding
216 with the 10 percentile omission value. Range estimates represented by polygons for low-data
217 species were converted to 10 arcminute grid cells to align with the modeled range estimates.
218 We then projected the range estimates for all species to a 50 × 50 km resolution equal-area
219 extent (Albers Equal Area Conic Projection) covering our study area in the western Palearctic

220 realm. Binary model predictions were projected to this coarser resolution using bilinear
221 interpolation that selected the maximum neighborhood value, resulting in predicted presence
222 for 50 km × 50 km cells when any underlying 10 arcminute (~20 km) cells had predictions of
223 presence. Grid-based analyses used the R packages *raster* (Hijmans, 2021) and *sf* (Pebesma,
224 2018). We excluded grid cells with fewer than 5 species to control for the negative influence of
225 low species richness in the analysis (He et al., 2020). Ultimately, we obtained 4527 grid
226 assemblages representing 711 species (36 species were not included due to insufficient
227 occurrence data in our study area, Table S1.1 and Figure S2.4).

228

229 *Phylogeny*

230 We derived phylogenetic information from a recently reconstructed, large-scale ant phylogeny
231 that represents most of currently recognized genera and provides valid relationships between
232 genera and their associated uncertainties (Economio et al., 2018). We updated the nomenclature
233 of taxa in the 100 posterior phylogenetic trees and pruned those trees based on the list of native
234 taxa in the western Palearctic realm. This process used the R packages *geiger* (Slater et al.,
235 2012) and *picante* (Kembel et al., 2010). As a result, estimated phylogenetic information for
236 641 species (86% of the total) was available for further analysis (see Figure S2.4).

237

238 **Measurement of turnover**

239 We determined the uniqueness of ant assemblages by measuring turnover of taxonomic and
240 phylogenetic compositions using the Simpson's index dissimilarity metric on assemblage pairs
241 (Baselga, 2010; Kreft & Jetz, 2010). We calculated two pairwise distance metrics: taxonomic
242 dissimilarity (β_{sim}) and phylogenetic dissimilarity ($P\beta_{sim}$) with the following formula:

243

$$244 \quad 1 - (a / (\min(b, c) + a))$$

245

246 For β_{sim} , a is the total number of species shared between two assemblages, and b and c are the
247 numbers of species unique to each assemblage. For $P\beta_{sim}$, a is the total branch length shared
248 between assemblages, and b and c are the lengths of unique branches to each assemblage (Daru
249 et al., 2017; Holt et al., 2013). $P\beta_{sim}$ was based on the median value across 100 posterior trees
250 to control the uncertainty of phylogeny. Higher or lower values of β_{sim} or $P\beta_{sim}$ indicate the
251 assemblages are less or more similar to each other, respectively. The mean pairwise β_{sim} or
252 $P\beta_{sim}$ of each assemblage was calculated to show spatial patterns, where higher β_{sim} or $P\beta_{sim}$
253 denotes higher uniqueness of species composition? or evolutionary history for an assemblage.

254 We also used null models to test if the observed assemblages were more or less similar
255 (taxonomically and phylogenetically) than expected by chance. The standardized effect size
256 (SES) of turnover was calculated as:

257

258 $SES = (\text{Mean}_{\text{obs}} - \text{Mean}_{\text{null}}) / \text{s.d.}_{\text{null}}$

259

260 where Mean_{obs} is the mean of the observed βsim or $P\beta\text{sim}$, and $\text{Mean}_{\text{null}}$ and $\text{s.d.}_{\text{null}}$ are the mean
261 and standard deviation of the null distribution for each randomized assemblage. The null model
262 used the Independent Swap algorithm to control for the effects of both the species pool and
263 different richness of assemblages (Gotelli, 2000) and was run for 1000 iterations. Higher or
264 lower values of SES here indicate that the assemblages have higher or lower turnover than
265 expected, respectively. The calculations for turnover and the null model used the R package
266 *betapart* (Baselga et al., 2021). The difference of pairwise βsim and $P\beta\text{sim}$ distance matrices
267 were assessed using the Mantel correlation test in the R package *vegan* (Oksanen et al., 2020).
268 The spatial congruence between the average βsim and $P\beta\text{sim}$ of assemblages was measured
269 using a modified t-test that can control for spatial autocorrelation by correcting the degree of
270 freedom (Clifford et al., 1989), using the R package *SpatialPack* (Vallejos, Osorio &
271 Bevilacqua, 2020).

272

273 **Delineation of biogeographic structure**

274 We performed clustering analysis based on pairwise taxonomic and phylogenetic turnover for
275 the biogeographic structure classification. To choose the most appropriate clustering algorithm
276 for our data, we examined eight candidates: 1) unweighted pair-group method using arithmetic
277 averages (UPGMA); 2) unweighted pair-group method using centroids (UPGMC); 3) Ward's
278 method (WARD); 4) single lineage (SL); 5) complete lineage (CL); 6) weighted pair-group
279 method using arithmetic averages (WPGMA); 7) weighted pair-group method using centroids
280 (WPGMC); 8) divisive hierarchical clustering (DIANA). We measured algorithm performance
281 by calculating the cophenetic Pearson correlation and the Gower distance to test the degree of
282 data distortion in models (Gower, 1983; Holt et al., 2013; Legendre & Legendre, 2012), leading
283 us to select UPGMA for our analysis (Table S2.2).

284

285 Based on the dendrograms from the UPGMA analysis, we used three different metrics to choose
286 the optimal classification of biogeographic regions and subregions: 1) average silhouette width
287 (ASW) with the R package *cluster* (Maechler et al., 2021); 2) Kelly-Gardner-Sutcliffe penalty
288 (KGS) with the R package *maptree* (White & Gramacy, 2012); and 3) Bootstrap mean
289 instability (Kelley et al., 1996; Kreft & Jetz, 2010; Legendre & Legendre, 2012) with the R
290 package *fpc* (Hennig, 2020). High reliability of clusters is indicated by higher ASW and lower
291 KGS, and mean instability. We defined the biogeographic regions as regions characterized by
292 distinct and coherent ant assemblages that can be delineated clearly in space (Kreft & Jetz, 2010;
293 Morrone, 2018; He et al., 2021). We also used the R package *phyloregion* (Daru, Karunarathne

294 & Schliep, 2020) to visualize regionalization patterns.

295

296 **RESULTS**

297 **Spatial turnover of West Palearctic ants**

298 Taxonomic and phylogenetic turnover (β_{sim} and $P\beta_{\text{sim}}$) showed high correlation and spatial
299 congruence in both regional lists and grid assemblages (Pearson's correlation $R = 0.87\text{--}0.96$, P
300 < 0.001 , Table S2.3). In regional lists, the highest turnover values for both metrics were
301 observed in the Mediterranean peninsulas and islands, and southern Anatolia, while central (e.g.,
302 Alps and Carpathians) and northern Europe and the British islands showed lower β_{sim} (Figures
303 1a and e). The observed β_{sim} and $P\beta_{\text{sim}}$ was higher than expected in the southern Iberian,
304 southern Anatolia and some Mediterranean islands such as Sicily, Crete and Cyprus (Figures
305 1b and f). The spatial turnover of grid assemblages presented a similar pattern: both turnover
306 metrics were highest in the south of the western Palearctic realm and decreased poleward
307 (Figures 1c and g). The standardized effect size (SES) also suggested the observed β_{sim} and
308 $P\beta_{\text{sim}}$ of grid assemblages were significantly higher than the random patterns in the
309 southwestern Palearctic realm, similar to the results from regional lists (Figures 1d and h).

310

311 **Biogeographic structure of West Palearctic ants**

312 *Delineation of regional lists*

313 The hierarchical clustering based on pairwise β_{sim} of regional lists suggested six biogeographic
314 regions and four subregions (Figures 2a and b; Figure S2.5): (1) Sicily and Maltese islands (SM)
315 were closely grouped and distinct from the rest of the western Palearctic realm; (2) Cyprus
316 (CY), (3) Southeastern Anatolian (SEA) and (4) Eastern Mediterranean (EM) including Turkey
317 (EM1), Aegean islands and Balkans (EM2), were grouped together; (5) Western Mediterranean
318 (WM) included the Iberian Peninsula and the Balearic Islands; (6) European region (EU)
319 including Apennine Peninsula, Adriatic Balkans, Corsica and Sardinia islands (EU1) and
320 European mainland and the British Isles (EU2), were grouped in the same cluster. The western
321 Mediterranean and European regions were grouped as the sister cluster of the Eastern
322 Mediterranean region, Southern Anatolian and Cyprus.

323

324 The delineation based on $P\beta_{\text{sim}}$ also recognized six biogeographic regions and presented
325 similar spatial patterns compared to the results of β_{sim} (Figures 2c and d; Figure S2.5).
326 However, biogeographic affinities between some regional assemblages were different in the
327 phylogenetic delineation. The Apennine Peninsula, Corsica and Sardinia islands (WM1) were
328 identified as the sister group of the Iberian Peninsula and the Balearic Islands (WM2), and a
329 part of the Western Mediterranean (WM). The western Anatolian (EM1), Aegean islands and
330 Mediterranean Turkey (EM2) were distinct from the rest of the Eastern Mediterranean (EM).

331 The Adriatic part of the Balkans was assigned to be a part of the European region (EU). The
332 Southeastern Anatolia (SEA), Sicily and Maltese (SM) and Cyprus (CY) islands were
333 suggested as the outgroup of the rest of the western Palearctic.

334

335 *Delineation of grid assemblages*

336 Grid assemblage clustering consistently recognized three biogeographic regions of western
337 Palearctic ants: Western Mediterranean (WM), Eastern Mediterranean (EM) and European
338 region (EU) (Figure 3; Figure S2.6). There were six and seven subregions detected based on
339 β sim and P β sim, respectively. The Western Mediterranean region included Corsica, Sardinia,
340 Sicily and Maltese islands (WM1, Figure 3d) and also included the Mediterranean coast of
341 France (WM2, Figure 3d) in the phylogenetic but not in the taxonomic delineation (Figure 3b).
342 The Eastern Mediterranean region extended from the Eastern Caucasus (ECA, Figures 3b and
343 d) to the Balkan Peninsula and even reached the Apennine Peninsula and Corsica island (EM3,
344 Figure 3b) in the taxonomic delineation. Crete island (EM2, Figure 3b) was suggested to be
345 distinct from the rest of the Eastern Mediterranean region based on β sim while the southern
346 Balkans, Aegean regions, southern Anatolia and Cyprus were grouped together based on P β sim
347 (EM1, Figure 3d). The Western and Eastern Mediterranean regions were grouped as the sister
348 cluster of the European region. Some lowland areas in the south of the European continent were
349 identified as a subregion (EU1) of the European region (EU, Figure 3): lowland in the north of
350 Caucasus, steppe in the north of Black Sea, plains in the south of Carpathian and Alps
351 mountains and in the north of Balkan, Dinaric Alps and Apennine mountains. This subregion
352 also included southern France and northern Iberia, except the Pyrenees, in the taxonomic
353 delineation (Figure 3b) while the Apennine Peninsula and Adriatic Balkans were included in
354 the phylogenetic delineation, (Figure 3d).

355

356 *Biogeographic boundaries*

357 Most of the boundaries of biogeographic regions and subregions of western Palearctic ants were
358 along the mountain chains or other geographic barriers like seas (Figure 4). The Pyrenees,
359 Apennines, Dinaric Alps, Balkan mountains, Black Sea, and the Caucasus mountains separated
360 the southwestern (i.e., Mediterranean regions) and northwestern (i.e., European region)
361 Palearctic realm of ants, from west to east. Mediterranean regions were mainly divided by the
362 Mediterranean, Tyrrhenian and Ionian Seas while the Mediterranean and Aegean Seas, Taurus
363 and Lesser Caucasus mountains were the major boundaries of subregions. The boundary of
364 subregions in the European region was also located along mountain ranges including the Alps,
365 Carpathians and Great Caucasus. The phylogenetic approach detected more boundaries in the
366 Eastern Mediterranean region compared to taxonomic delineation (Figures 4b and c).
367 Biogeographic boundaries identified based on grid assemblages were more consistent across

368 dissimilarity metrics and better matched geographic barriers compared to the boundaries
369 estimated using regional lists. And boundaries delineated using both geographic units were
370 well-matched when the regional polygons are delimited along mountains, for example, on the
371 Balkans, Taurus and Caucasus mountains (Figures 4b and c).

372

373 **DISCUSSION**

374 In this study, we proposed a new and comprehensive regionalization system for the western
375 Palearctic ants. To the best of our knowledge, this novel system represents the first
376 comprehensive delineation for Palearctic insects at a fine geographic scale informed by a large-
377 scale phylogeny. The regionalization of ants proposed in this study largely follows the typical
378 biogeographic structure of western Palearctic fauna, but at the same time reveals differences
379 across geographic units and dissimilarity metrics.

380

381 **Biogeographic regionalization of ants in the western Palearctic realm**

382 Overall, our results show that the western Palearctic ant fauna has a clear biogeographic
383 structure: it can be separated into southwestern (i.e., Mediterranean regions) and northwestern
384 parts (i.e., European region), with the southwestern part having stronger regionalization (i.e.,
385 higher turnover and more biogeographic regions) compared to the northwest (Figures 1-3). This
386 biogeographic divergence of the western Palearctic realm is consistent with the first
387 zoogeographic regionalization proposed by Wallace (1876) and other systems proposed
388 afterwards based on vertebrates (e.g., Holt et al., 2013; Ficetola et al., 2018) and insects (e.g.,
389 Heiser et al., 2014; Vitali & Schmitt, 2017).

390

391 This biogeographic divergence of ants reveals the strong impact of historical climate changes
392 on species distributions in the western Palearctic. The comparison of fossil and modern ant
393 assemblages has shown that many lineages went extinct in the western Palearctic realm during
394 the cooling period following the Miocene (Guénard, Perrichot & Economo, 2015). As more
395 diverse and unique ant assemblages currently exist in Mediterranean regions, perhaps due to
396 higher climate stability (Schmitt, 2007), this suggests that the historical signal of western
397 Palearctic ant fauna may persist in these areas (Figure 1). In contrast, the homogenous structure
398 of ants in the northwestern Palearctic realm may indicate that the fauna in the north was
399 assembled by poleward dispersals of species from their southern glacial refugia after the Last
400 Glacial Maximum (Baroni Urbani & Collingwood, 1977; Pusch et al., 2006; Leppänen et al.,
401 2013), a pattern which is also found in many other taxa (Dennis et al., 1991; Fattorini & Ulrich,
402 2012; Vitali & Schmitt, 2017). The Pyrenees, Balkans and Caucasus are likely to be important
403 postglacial recolonization centres for ants due to the strong biogeographic affinity between
404 these regions and the northwestern Palearctic realm (Figures 3 and S2.7).

405

406 The strong biogeographic regionalization in Mediterranean regions also suggests the
407 importance of dispersal limitation in shaping ant biogeographic structure. Compared to the
408 northern part of the western Palearctic realm, Mediterranean regions have more islands and
409 varied topography, which may be responsible for restricting ant distributions (Figure 4).
410 Especially for the East Mediterranean region, its remarkable uniqueness in species composition
411 and biogeographic isolation could also be the consequence of complex topography and
412 geological history (Vitali & Schmitt, 2017; Ficetola et al., 2018; Ahmadi et al., 2021; Kiran &
413 Karaman, 2021). Thus, the dispersal limitation may explain why geographic barriers (e.g.,
414 mountains and seas) represent major boundaries separating biogeographic regions and
415 subregions of the western Palearctic ants (Ficetola, Mazel & Thuiller, 2017).

416

417 Moreover, the strong biogeographic affinities between some isolated regions also reveal the
418 legacy of historical land connections. The high similarity in the fauna of the Apennine Peninsula
419 and Adriatic Balkans, the Corsica and Sardinia islands, and the Sicilian and Maltese islands
420 (Figures 2 and 3) can be explained by the biotic exchanges through land bridges during the
421 period of partial disappearance of the Mediterranean Sea in late Miocene (i.e., Messinian
422 salinity crisis) and Quaternary glacial periods (Randi, 2007; Schlick-Steiner et al., 2007;
423 Schmitt, 2007; Dapporto et al., 2014; Schmitt et al., 2021). The biogeographic affinity of ant
424 assemblages between the Tyrrhenian islands and the Iberian Peninsula suggests the
425 evolutionary history shared by their fauna (Figures 2c-d and 3c-d). This may be due to the
426 historical expansion of old lineages through land bridges (Senczuk et al., 2017; Dapporto et al.,
427 2019; Schmitt et al., 2021), or to the legacy of more ancient geographic connections. Corsica
428 and Sardinia islands were parts of the Tyrrhenian microplate which was separated from the
429 Iberian Peninsula, and thus some lineages may persist with the same origins as the Iberian fauna
430 (Ketmaier, Caccone & Silva-Opps, 2013; Schmitt et al., 2021).

431

432 **Effects of geographic units and dissimilarity metrics on regionalization**

433 *Regional lists versus grid cells*

434 Both regional lists and grid assemblages result in a very similar biogeographic structure,
435 although several species are not included in grid-level analysis (Figures 2, 3 and S2.5). The
436 regionalization based on grid assemblages shows more detail regarding biogeographic
437 transition and seemingly detects stronger effects of geographic barriers in comparison to that
438 of the regional lists. The grid-level delineations reveal not only the close relationship between
439 the Pyrenees, Balkans, Caucasus and the northwestern Palearctic ant fauna, but also suggest
440 differences between the southern subregion and the European region (EU in Figure 3). The
441 southern subregion (i.e., EU1 in Figure 3) includes several glacial refugia (Hewitt, 1999;

442 Schmitt, 2007) and may harbor species extinct in the north during the Last Glacial Maximum
443 and unable to disperse since across physical barriers like the Alps and the Carpathian mountains
444 (Figures 4 and S2.8).

445

446 The accuracy of regional polygons would influence the regionalization results. Due to their
447 relatively coarse scale, regional polygons likely do not reflect species' range boundaries as
448 accurately as grid cells (Figure 4). Some regional polygons based on administrative divisions
449 may extend across multiple biogeographic regions and thus regional lists may include distinct
450 fauna. For example, the regional list of southeastern Anatolia has a relatively large extent that
451 may include some species from the Arabian Peninsula, making its fauna very different from the
452 rest of the Eastern Mediterranean region. Notably, delineations of regional lists highlight the
453 biogeographic uniqueness of Sicily, Maltese and Cyprus islands (Figure 2). The SDMs we used
454 to make range predictions do not correct for dispersal limitations, so the long-term isolation
455 and complex geologic history of these islands (Poulakakis et al., 2013; Schmitt et al., 2021) and
456 other regions with similar histories were not considered for the grid assemblage regionalization.

457

458 *Taxonomic versus phylogenetic metrics*

459 The spatial patterns of taxonomic and phylogenetic turnover were highly congruent (Figure 1
460 and Table S2.3). This suggests that the spatial turnover of ant phylogeny may be caused by
461 species replacement in the western Palearctic realm, while the null model analysis suggests
462 many coastal and insular Mediterranean regions present strong phylogenetic turnover which is
463 independent of species richness and compositions. The phylogenetic regionalization further
464 reveals the affinity between the Tyrrhenian islands and the Iberian Peninsula and the divergence
465 between northern and southern Anatolia (Figures 2c-d and 3c-d); the latter is perhaps even
466 under-estimated due to the limitation of phylogenetic data (Figure S2.4). Thus, the phylogenetic
467 approach provides evolutionary support and some new perspectives into the biogeographic
468 history of the West Palearctic realm.

469

470 **Tackling the Wallacean and Darwinian shortfalls for diverse insect groups**

471 Answering some of the oldest and most important questions in biology relies on basic
472 information about species distributions and their phylogenetic relationships, which represent
473 important shortfalls for most insect groups. These knowledge gaps have long prevented
474 biogeographic classifications based on taxa other than plants and vertebrates. In particular,
475 classifications made based on vertebrates are particularly problematic, as current data may hide
476 profound assemblage changes that occurred following ancient human settlement and resulting
477 environmental modifications (Faurby & Svenning, 2015; Santos et al., 2020). As a result,
478 datasets that expand the taxonomic breadth of biogeographic classifications are needed to

479 confirm previous results if any broad patterns across taxa are to be confirmed. Although rare
480 for insects, this study was able to produce a comprehensive regionalization of the western
481 Palearctic ants due to the development of accurate and exhaustive distributional information,
482 as well as a large-scale phylogeny.

483

484 This study demonstrates that a combination of expert opinion and modeling can help tackle the
485 Wallacean Shortfall for insects. The expert-validation step ensures quality control of the
486 distributional data, while models like SDMs represent powerful tools for estimating detailed
487 delimitations of species distributions. However, we also recognize some limitations in this
488 approach. For example, the SDM predictions we used here represent the responses of ant
489 species to climatic variables and do not explicitly consider species' dispersal abilities, biotic
490 interactions, and other local effects. Although techniques exist to consider dispersal limitations
491 (Monsimet et al. 2020) and biotic interactions (Kass et al. 2020, Wisz et al. 2013) in SDMs,
492 this is currently difficult for ants because the necessary data does not exist for most ant species.
493 Thus, the range estimates from the regional lists and the SDM predictions can suggest different
494 affinities and spatial extents in biogeographic structure (Figures 2, 3 and S2.7). Such important
495 ecological information could improve the modeling of species distributions and the resulting
496 delineation of biogeographic structure once available for more ant species (e.g., Jaeschke et al.,
497 2013).

498

499 The Darwinian Shortfall for insects may be particularly important for many regions around the
500 globe. Some of the diverse insect groups for which large-scale phylogenies have been
501 developed could be ideal models for understanding the macroevolution and macroecology of
502 insects (e.g., ants: Economo et al., 2018, butterflies: Earl et al., 2021). The estimates of
503 phylogeny could be useful for solving limitations due to limited molecular data available in
504 most insect groups, especially for metrics like phylogenetic turnover which is based on the
505 branches length, and are not sensitive to the uncertainty in phylogenetic topology (Economo et
506 al., 2018; Jetz & Pyron, 2018). However, more taxonomic revisions and molecular sequencing
507 are needed to complete comprehensive and robust phylogenetic constructions for insect species.

508

509 **Conclusion**

510 Our study shows that expert validation and modeling of species distributions combined with a
511 large-scale phylogeny can help us develop a comprehensive regionalization for a diverse insect
512 group, and thus directly tackle Wallacean and Darwinian shortfalls for insects. The new
513 bioregionalization of western Palearctic ants we present here supports a biogeographic
514 divergence between the more homogenous northwest European region and the more
515 regionalized southwest Mediterranean region. These biogeographic structures reveal potential

516 effects of Quaternary climate changes, and even deeper geological processes like plate tectonics,
517 on ant distributions. The western Palearctic ants represent an ideal system to investigate how
518 the insect distributions have responded to historical processes. Moreover, the remarkable
519 uniqueness of coastal and insular Mediterranean areas highlights their historical roles as glacial
520 refugia and potential significance for the future conservation of ant diversity. The estimation of
521 regionalization systems for other insect groups would provide the possibility for comparisons
522 with our delineations, which would contribute to essential knowledge of insect biogeography
523 and would provide important information about their uniqueness to guide future conservation
524 efforts.

525

526 **DATA AVAILABILITY STATEMENT**

527 The regional lists, binary range maps and metadata of species distribution modelings used for
528 this study will be available in Dryad Digital Repository once the study has been accepted.

529

530 **REFERENCES**

- 531 Ahmadi, M., Hemami, M. R., Kaboli, M., Nazarizadeh, M., Malekian, M., Behrooz, R., ... &
532 Zimmermann, N. E. (2021). The legacy of Eastern Mediterranean mountain uplifts: rapid
533 disparity of phylogenetic niche conservatism and divergence in mountain vipers. *BMC*
534 *Ecology and Evolution*, 21(1), 1-13.
- 535 Ballesteros-Mejia, L., Kitching, I. J., Jetz, W., & Beck, J. (2017). Putting insects on the map:
536 Near-global variation in sphingid moth richness along spatial and environmental gradients.
537 *Ecography*, 40(6), 698-708.
- 538 Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity.
539 *Global ecology and biogeography*, 19(1), 134-143.
- 540 Baselga, A., Orme, D., Villeger, S., De Bortoli, J., Leprieur, F. and Logez, M. (2021). betapart:
541 Partitioning Beta Diversity into Turnover and Nestedness Components. R package version
542 1.5.4. <https://CRAN.R-project.org/package=betapart>
- 543 Bolton, B. 2020. An online catalog of the ants of the world. Available from <https://antcat.org>.
544 (accessed 19th Nov. 2020)
- 545 Calatayud, J., Rodríguez, M. Á., Molina-Venegas, R., Leo, M., Horreo, J. L., & Hortal, J.
546 (2019). Pleistocene climate change and the formation of regional species pools.
547 *Proceedings of the Royal Society B*, 286(1905), 20190291.
- 548 Cox, B. (2001). The biogeographic regions reconsidered. *Journal of biogeography*, 28(4), 511-
549 523.
- 550 Clifford, P., Richardson, S., & Hémon, D. (1989). Assessing the significance of the correlation
551 between two spatial processes. *Biometrics*, 123-134.
- 552 Dapporto, L., Fattorini, S., Vodă, R., Dincă, V., & Vila, R. (2014). Biogeography of western

- 553 Mediterranean butterflies: combining turnover and nestedness components of faunal
554 dissimilarity. *Journal of Biogeography*, 41(9), 1639-1650.
- 555 Dapporto, L., Cini, A., Vodă, R., Dincă, V., Wiemers, M., Menchetti, M., ... & Vila, R. (2019).
556 Integrating three comprehensive data sets shows that mitochondrial DNA variation is
557 linked to species traits and paleogeographic events in European butterflies. *Molecular*
558 *Ecology Resources*, 19(6), 1623-1636.
- 559 Daru, B. H., Elliott, T. L., Park, D. S., & Davies, T. J. (2017). Understanding the processes
560 underpinning patterns of phylogenetic regionalization. *Trends in ecology & evolution*,
561 32(11), 845-860.
- 562 Daru, B. H., Karunarathne, P., & Schliep, K. (2020). *phyloregion*: R package for
563 biogeographical regionalization and macroecology. *Methods in Ecology and Evolution*,
564 11(11), 1483-1491.
- 565 Dennis, R. L. H., Williams, W. R., & Shreeve, T. G. (1991). A multivariate approach to the
566 determination of faunal structures among European butterfly species (Lepidoptera:
567 Rhopalocera). *Zoological Journal of the Linnean Society*, 101(1), 1-49.
- 568 Diniz-Filho, J. A. F., De Marco Jr, P. A. U. L. O., & Hawkins, B. A. (2010). Defying the curse
569 of ignorance: perspectives in insect macroecology and conservation biogeography. *Insect*
570 *Conservation and Diversity*, 3(3), 172-179.
- 571 Earl, C., Belitz, M. W., Laffan, S. W., Barve, V., Barve, N., Soltis, D. E., ... & Guralnick, R.
572 (2021). Spatial phylogenetics of butterflies in relation to environmental drivers and
573 angiosperm diversity across North America. *iScience*, 24(4), 102239.
- 574 Economo, E. P., Narula, N., Friedman, N. R., Weiser, M. D., & Guénard, B. (2018).
575 Macroecology and macroevolution of the latitudinal diversity gradient in ants. *Nature*
576 *communications*, 9(1), 1-8.
- 577 Elias, S. A. (1991). Insects and climate change. *Bioscience*, 41(8), 552-559.
- 578 Fattorini, S., & Ulrich, W. (2012). Spatial distributions of European Tenebrionidae point to
579 multiple postglacial colonization trajectories. *Biological Journal of the Linnean Society*,
580 105(2), 318-329.
- 581 Faurby, S., & Svenning, J. C. (2015). Historic and prehistoric human-driven extinctions have
582 reshaped global mammal diversity patterns. *Diversity and Distributions*, 21(10), 1155-
583 1166.
- 584 Ficetola, G. F., Mazel, F., & Thuiller, W. (2017). Global determinants of zoogeographical
585 boundaries. *Nature Ecology & Evolution*, 1(4), 1-7.
- 586 Ficetola, G. F., Falaschi, M., Bonardi, A., Padoa-Schioppa, E., & Sindaco, R. (2018).
587 Biogeographical structure and endemism pattern in reptiles of the Western Palearctic.
588 *Progress in Physical Geography: Earth and Environment*, 42(2), 220-236.
- 589 Fick, S. E., & Hijmans, R. J. (2017). *WorldClim 2*: new 1-km spatial resolution climate surfaces

- 590 for global land areas. *International journal of climatology*, 37(12), 4302-4315.
- 591 Galante, P. J., Alade, B., Muscarella, R., Jansa, S. A., Goodman, S. M., & Anderson, R. P.
592 (2018). The challenge of modeling niches and distributions for data-poor species: a
593 comprehensive approach to model complexity. *Ecography*, 41(5), 726-736.
- 594 Guénard, B., Perrichot, V., & Economo, E. P. (2015). Integration of global fossil and modern
595 biodiversity data reveals dynamism and stasis in ant macroecological patterns. *Journal of*
596 *Biogeography*, 42(12), 2302-2312.
- 597 Guénard, B., Weiser, M. D., Gomez, K., Narula, N., & Economo, E. P. (2017). The Global Ant
598 Biodiversity Informatics (GABI) database: synthesizing data on the geographic
599 distribution of ant species (Hymenoptera: Formicidae). *Myrmecological*
600 *News/Osterreichische Gesellschaft fur Entomofaunistik*, 24, 83-89.
- 601 Gower, J. C. (1983). Comparing classifications. In *Numerical taxonomy* (pp. 137-155).
602 Springer, Berlin, Heidelberg.
- 603 Gullan, P. J., & Cranston, P. S. (2014). *The insects: an outline of entomology*. John Wiley &
604 Sons.
- 605 He, J., Lin, S., Li, J., Yu, J., & Jiang, H. (2020). Evolutionary history of zoogeographical
606 regions surrounding the Tibetan Plateau. *Communications biology*, 3(1), 1-9.
- 607 Heiser, M., & Schmitt, T. (2010). Do different dispersal capacities influence the biogeography
608 of the western Palearctic dragonflies (Odonata)? *Biological Journal of the Linnean Society*,
609 99(1), 177-195.
- 610 Heiser, M., Dapporto, L., & Schmitt, T. (2014). Coupling impoverishment analysis and
611 partitioning of beta diversity allows a comprehensive description of Odonata
612 biogeography in the Western Mediterranean. *Organisms Diversity & Evolution*, 14(2),
613 203-214.
- 614 Hennig, C. (2020). *fpc: Flexible Procedures for Clustering*. R package version 2.2-9.
615 <https://CRAN.R-project.org/package=fpc>
- 616 Hewitt, G. (1999). Post-glacial re-colonization of European biota. *Biological journal of the*
617 *Linnean Society*, 68(1-2), 87-112.
- 618 Hijmans, R. J. (2021). *raster: Geographic Data Analysis and Modeling*. R package version 3.4-
619 13. <https://CRAN.R-project.org/package=raster>
- 620 Hölldobler, B., & Wilson, E. O. (1990). *The ants*. Harvard University Press.
- 621 Holt, B. G., Lessard, J. P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., ... &
622 Rahbek, C. (2013). An update of Wallace's zoogeographic regions of the world. *Science*,
623 339(6115), 74-78.
- 624 Jaeschke, A., Bittner, T., Reineking, B., & Beierkuhnlein, C. (2013). Can they keep up with
625 climate change?—Integrating specific dispersal abilities of protected Odonata in species
626 distribution modelling. *Insect Conservation and Diversity*, 6(1), 93-103.

- 627 Jetz, W., & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation
628 with present imperilment across the amphibian tree of life. *Nature ecology & evolution*,
629 2(5), 850-858.
- 630 Kass, J. M., Anderson, R. P., Espinosa-Lucas, A., Juárez-Jaimes, V., Martínez-Salas, E.,
631 Botello, F., ... & Sánchez-Cordero, V. (2020). Biotic predictors with phenological
632 information improve range estimates for migrating monarch butterflies in Mexico.
633 *Ecography*, 43(3), 341-352.
- 634 Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C. L., Pinilla-Buitrago, G. E., Boria, R. A., ...
635 & Anderson, R. P. (2021). ENMeval 2.0: redesigned for customizable and reproducible
636 modeling of species' niches and distributions. *Methods in Ecology and Evolution*, 12:
637 1602– 1608.
- 638 Kelley, L. A., Gardner, S. P., & Sutcliffe, M. J. (1996). An automated approach for clustering
639 an ensemble of NMR-derived protein structures into conformationally related subfamilies.
640 *Protein Engineering, Design and Selection*, 9(11), 1063-1065.
- 641 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D.,
642 Blomberg, S.P. and Webb, C.O. 2010. Picante: R tools for integrating phylogenies and
643 ecology. *Bioinformatics* 26:1463-1464.
- 644 Ketmaier, V., Caccone, A., & Silva-Opps, M. (2013). Twenty years of molecular biogeography
645 in the West Mediterranean islands of Corsica and Sardinia: lessons learnt and future
646 prospects. *Current progress in biological research*, 71-93.
- 647 Kiran, K., & Karaman, C. (2021). Ant fauna (Hymenoptera: Formicidae) of Central Anatolian
648 Region of Turkey. *Turkish Journal of Zoology*, 45(3), 161-196.
- 649 Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on
650 species distributions. *Journal of Biogeography*, 37(11), 2029-2053.
- 651 Legendre, P., & Legendre, L. (2012). *Numerical ecology*. Elsevier.
- 652 Leppänen, J., Vepsäläinen, K., Anthoni, H., & Savolainen, R. (2013). Comparative
653 phylogeography of the ants *Myrmica ruginodis* and *Myrmica rubra*. *Journal of*
654 *biogeography*, 40(3), 479-491.
- 655 Lomolino, M. V., Riddle, B. R., & Whittaker, R. J. (2016). *Biogeography Fifth Edit.*
- 656 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2021). *cluster: Cluster*
657 *Analysis Basics and Extensions*. R package version 2.1.2.
- 658 Monsimet, J., Devineau, O., Petillon, J., & Lafage, D. (2020). Explicit integration of dispersal-
659 related metrics improves predictions of SDM in predatory arthropods. *Scientific Reports*,
660 10(1), 1-12.
- 661 Morrone, J. J. (2018). The spectre of biogeographical regionalization. *Journal of Biogeography*,
662 45(2), 282-288.
- 663 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.

- 664 R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. and Wagner,
665 H. (2020). *vegan: Community Ecology Package*. R package version 2.5-7.
666 <https://CRAN.R-project.org/package=vegan>
- 667 Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The*
668 *R Journal* 10 (1), 439-446, <https://doi.org/10.32614/RJ-2018-009>
- 669 Peterson, A. T., Navarro-Sigüenza, A. G., & Gordillo, A. (2018). Assumption-versus data-
670 based approaches to summarizing species' ranges. *Conservation Biology*, 32(3), 568-575.
- 671 Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S.
672 (2009). Sample selection bias and presence-only distribution models: implications for
673 background and pseudo-absence data. *Ecological Applications*, 19(1), 181-197.
- 674 Poulakakis, N., Kapli, P., Kardamaki, A., Skourtanioti, E., Göcmen, B., Ilgaz, Ç., ... &
675 Lymberakis, P. (2013). Comparative phylogeography of six herpetofauna species in
676 Cyprus: late Miocene to Pleistocene colonization routes. *Biological Journal of the Linnean*
677 *Society*, 108(3), 619-635.
- 678 Pusch, K., Seifert, B., Foitzik, S., & Heinze, J. (2006). Distribution and genetic divergence of
679 two parapatric sibling ant species in Central Europe. *Biological Journal of the Linnean*
680 *Society*, 88(2), 223-234.
- 681 R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation
682 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 683 Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species
684 distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, 41(4), 629-
685 643.
- 686 Randi, E. (2007). Phylogeography of south European mammals. In *Phylogeography of southern*
687 *European refugia* (pp. 101-126). Springer, Dordrecht.
- 688 Rueda, M., Rodríguez, M. Á., & Hawkins, B. A. (2010). Towards a biogeographic
689 regionalization of the European biota. *Journal of Biogeography*, 37(11), 2067-2076.
- 690 Santos, A. M., Cianciaruso, M. V., Barbosa, A. M., Bini, L. M., Diniz-Filho, J. A. F., Faleiro,
691 F. V., ... & Hortal, J. (2020). Current climate, but also long-term climate changes and
692 human impacts, determine the geographic distribution of European mammal diversity.
693 *Global Ecology and Biogeography*, 29(10), 1758-1769.
- 694 Senczuk, G., Colangelo, P., De Simone, E., Aloise, G., & Castiglia, R. (2017). A combination
695 of long term fragmentation and glacial persistence drove the evolutionary history of the
696 Italian wall lizard *Podarcis siculus*. *BMC evolutionary biology*, 17(1), 1-15.
- 697 Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial
698 trends. *Frontiers in zoology*, 4(1), 1-13.
- 699 Schmitt, T., Fritz, U., Delfino, M., Ulrich, W., & Habel, J. C. (2021). Biogeography of Italy
700 revisited: genetic lineages confirm major phylogeographic patterns and a pre-Pleistocene

- 701 origin of its biota. *Frontiers in zoology*, 18(1), 1-13.
- 702 Slater, G. J., Harmon, L. J., Wegmann, D., Joyce, P., Revell, L. J., & Alfaro, M. E. (2012).
703 Fitting models of continuous trait evolution to incompletely sampled comparative data
704 using approximate Bayesian computation. *Evolution: International Journal of Organic*
705 *Evolution*, 66(3), 752-762.
- 706 Smith, C. H. (1983). A system of world mammal faunal regions. I. Logical and statistical
707 derivation of the regions. *Journal of Biogeography*, 455-466.
- 708 Stork, N. E. (2018). How many species of insects and other terrestrial arthropods are there on
709 Earth? *Annual Review of Entomology*, 63, 31-45.
- 710 Vallejos, R., Osorio, F., Bevilacqua, M. (2020). *Spatial Relationships Between Two*
711 *Georeferenced Variables: with Applications in R*. Springer, New York. ISBN 978-3-030-
712 56680-7
- 713 Vitali, F., & Schmitt, T. (2017). Ecological patterns strongly impact the biogeography of
714 western Palaearctic longhorn beetles (Coleoptera: Cerambycoidea). *Organisms Diversity*
715 *& Evolution*, 17(1), 163-180.
- 716 Wallace, A. R. (1876) *The geographical distribution of animals*. Harper & Brothers, New York.
- 717 Wallace, A. R. (1894). What are zoological regions. *Nature*, 49, 610-613.
- 718 Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., ... &
719 Svenning, J. C. (2013). The role of biotic interactions in shaping distributions and realised
720 assemblages of species: implications for species distribution modelling. *Biological*
721 *Reviews*, 88(1), 15-30.
- 722 Ye, J., Lu, L., Liu, B., Yang, T., Zhang, J., Hu, H., ... & Chen, Z. (2019). Phylogenetic
723 delineation of regional biota: A case study of the Chinese flora. *Molecular Phylogenetics*
724 *and Evolution*, 135, 222-229.
- 725
- 726

727 **TABLE & FIGURES**

728 **Figure 1.** Spatial patterns of turnover of the western Palearctic ant assemblages.

729 **Figure 2.** Regionalization of the western Palearctic ant fauna based on regional lists.

730 **Figure 3.** Regionalization of the western Palearctic ant fauna based on grid assemblages.

731 **Figure 4.** Comparison of biogeographic structure transitions.

732

733 **SUPPLEMENTARY**

734 **Appendix 1.**

735 **Table S1.1.** Information of species included in the study.

736 **Appendix 2.**

737 **Table S2.2.** Evaluation of clustering algorithms.

738 **Table S2.3.** Correlation of taxonomic and phylogenetic pairwise dissimilarity matrix and
739 spatial turnover.

740 **Figure S2.1.** Diagram of the workflow of this study.

741 **Figure S2.2.** Map of geographical units used in the European Ant Distribution (EUAD)
742 database.

743 **Figure S2.3.** Contribution of experts to the new database.

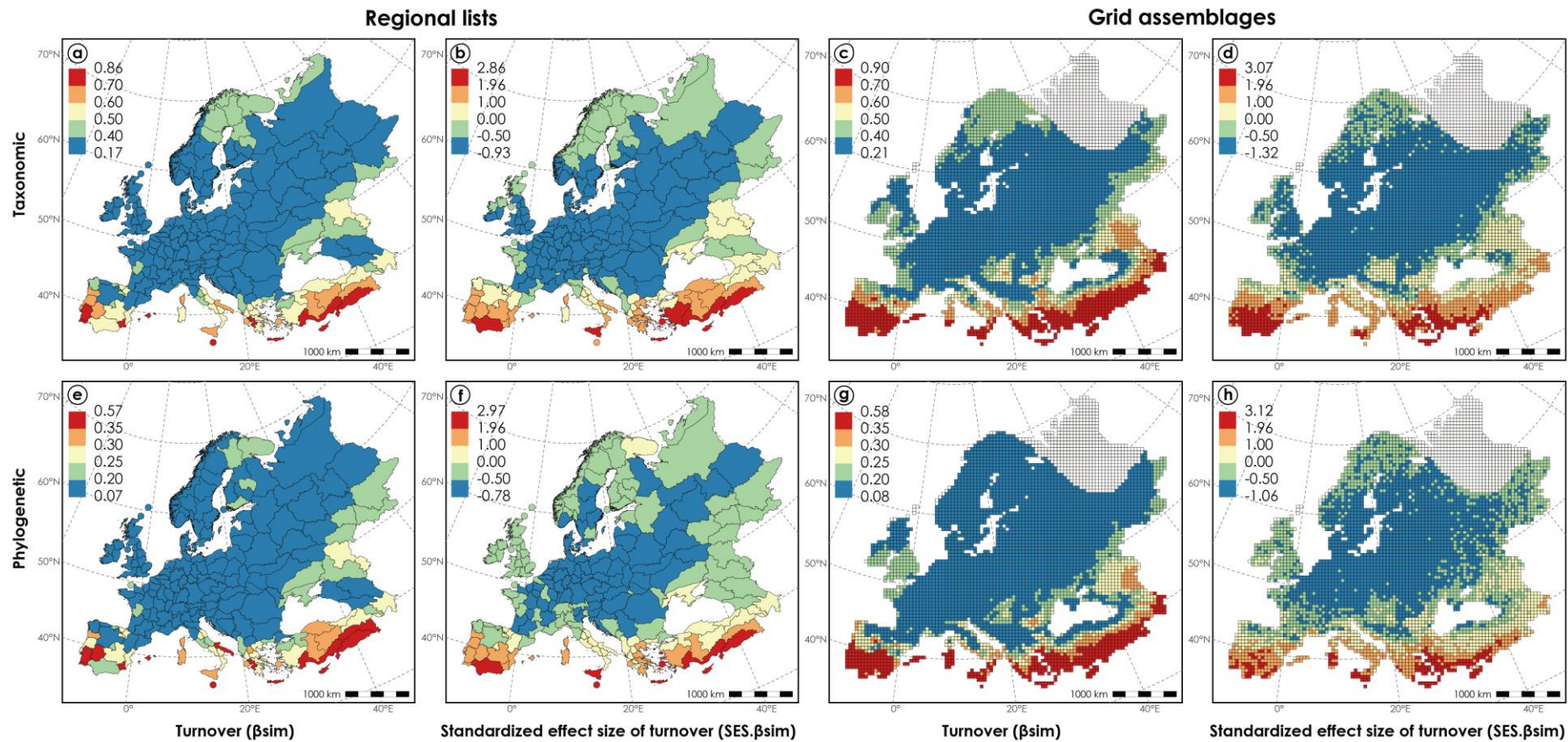
744 **Figure S2.4.** Spatial pattern of species without species distribution modellings or phylogenetic
745 data.

746 **Figure S2.5.** Evaluation of UPGMA hierarchical clustering of ant regional lists.

747 **Figure S2.6.** Evaluation of UPGMA hierarchical clustering of ant grid assemblages.

748 **Figure S2.7.** Different subregions identified within the European region of ants.

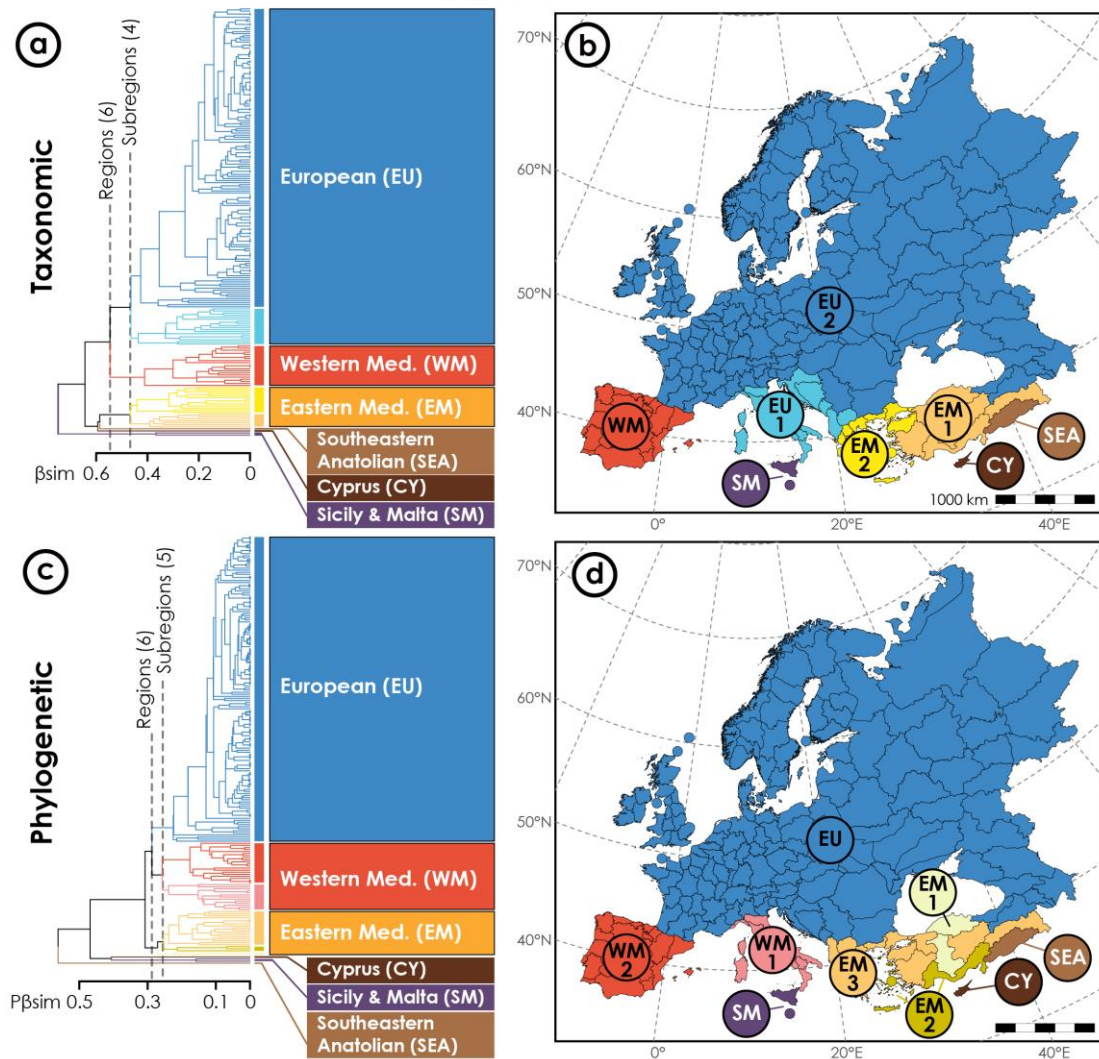
749



750

751 **Figure 1.** Spatial patterns of turnover of the western Palearctic ant assemblages. Taxonomic (a-d) and phylogenetic (e-h) turnover of regional lists (a, b, e, f)
 752 and grid assemblages (c, d, g, h). Both observed average value of taxonomic (β_{sim}) (a, c) and phylogenetic ($P\beta_{sim}$) (e, g) turnover and their standardized effect
 753 size (SES) results from randomization test (Independent Swap) were calculated (b and d for SES. β_{sim} ; f and h for SES. $P\beta_{sim}$). Values that are not available
 754 were indicated by the white colour on the map.

Regional lists



755

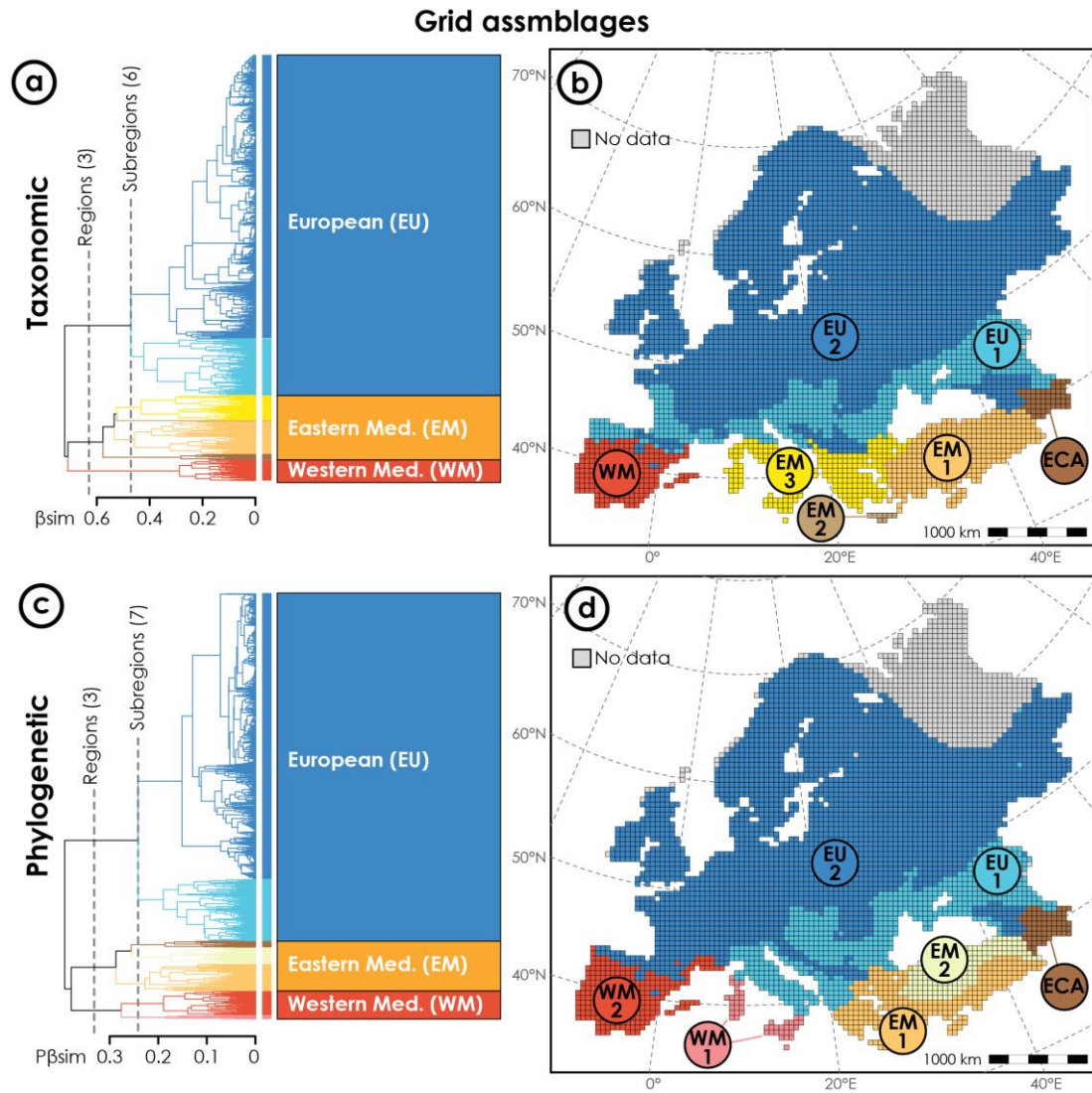
756 **Figure 2.** Regionalization of the western Palearctic ant fauna based on regional lists.

757 Dendrograms (a, c) and maps (b, d) resulting from the unweighted pair-group method using

758 arithmetic average (UPGMA) hierarchical clustering based on βsim (a-b) and $P\beta\text{sim}$ (c-d)

759 matrices. Colours used to characterize particular regions in dendrograms and maps are identical.

760



761

762 **Figure 3.** Regionalization of the western Palearctic ant fauna based on grid assemblages.

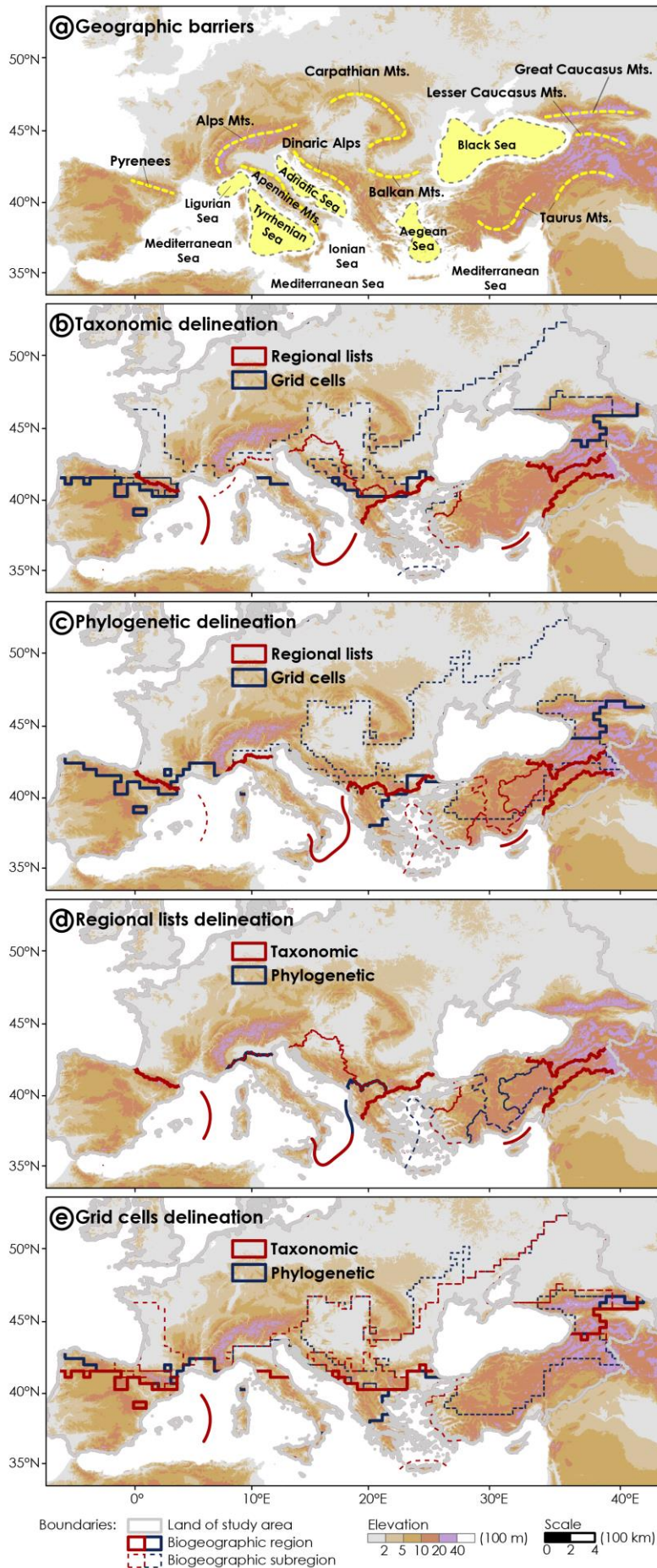
763 Dendrograms (a, d) and maps (b, e) resulting from UPGMA hierarchical clustering based on

764 β_{sim} (a-c) and $P\beta_{sim}$ (d-f) matrices. Colours used to characterize particular regions in

765 dendrograms and maps are identical. Values that are not available were indicated by the grey

766 colour on the map.

767



769 **Figure 4.** Comparison of biogeographic structure transitions. The major geographic barriers
770 (a) are indicated by the yellow dotted lines. The boundaries of biogeographic
771 regions/subregions based on regional lists and grid cells are overlapped in taxonomic (b) and
772 phylogenetic (c) delineations while the boundaries of taxonomic and phylogenetic
773 delineations are also compared at regional list- (d) and grid cell-level (e). The biogeographic
774 boundaries over seas are also indicated.
775

776 **Table S2.3.** Evaluation of clustering algorithms. Algorithm with higher cophenetic Pearson
 777 correlation (Cor.) and lower Gower distance (Gower dist.) is considered to have a better
 778 performance. Red and blue colours indicate better and worse performance in each criterion,
 779 respectively. Abbreviations: UPGMA, unweighted pair-group method using arithmetic
 780 averages; 2) UPGMC, unweighted pair-group method using centroids; 3) WARD, Ward's
 781 method; 4) SL, single lineage; 5) CL, complete lineage; 6) WPGMA, weighted pair-group
 782 method using arithmetic averages; 7) WPGMC, weighted pair-group method using centroids;
 783 8) DIANA, divisive hierarchical clustering.

	β sim distance		P β sim distance	
	Cor.	Gower dist.	Cor.	Gower dist.
<i>Regional lists</i>				
UPGMA	0.77	532	0.74	218
UPGMC	0.79	1508	0.70	487
WPGM				
A	0.56	532	0.68	341
WPGMC	0.77	1432	0.58	485
SL	0.69	2896	0.67	960
CL	0.54	5252	0.62	3158
Ward	0.62	2419	0.62	40072
DIANA	0.68	4289	0.73	2106
<i>Grid assemblages</i>				
UPGMA	0.81	272733	0.78	105631
UPGMC	0.75	974126	0.72	316449
WPGM				
A	0.57	617566	0.72	134312
WPGMC	0.75	1084549	0.57	316449
SL	0.70	2422040	0.69	695378
CL	0.47	2788312	0.71	1460044
Ward	0.71	1526776	0.66	715461437
DIANA	0.51	2792959	0.78	1366657

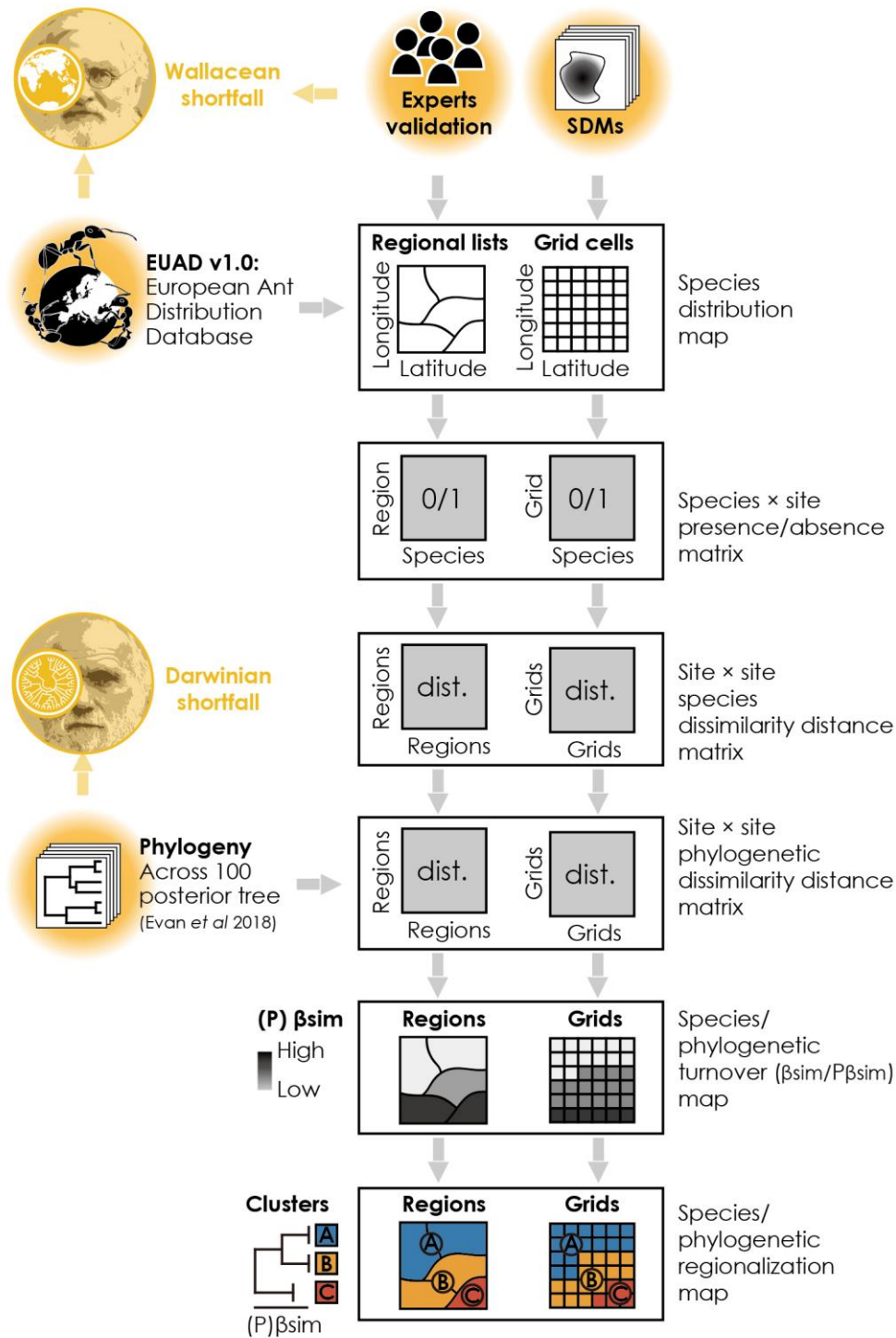
784

785

786 **Table S2.4.** Correlation of taxonomic and phylogenetic pairwise dissimilarity matrix and
787 spatial turnover. Pearson's correlation assessed by Mantel test is shown above the diagonal and
788 spatially corrected Pearson's correlation calculated by modified t test is shown below the
789 diagonal. All correlations are significant at $P < 0.001$ level.

		β sim	SES. β sim	P β sim	SES.P β sim
Regional assemblage s (N = 207)	β sim	—	0.87	0.94	0.85
	SES. β sim	0.93	—	—	0.95
	P β sim	0.96	—	—	0.82
	SES.P β sim	0.90	0.94***	0.88	—
Grid assemblage s (N = 4403)	β sim	—	0.88	0.95	0.84
	SES. β sim	0.92	—	—	0.95
	P β sim	0.96	—	—	0.82
	SES.P β sim	0.91	0.97	0.85	—

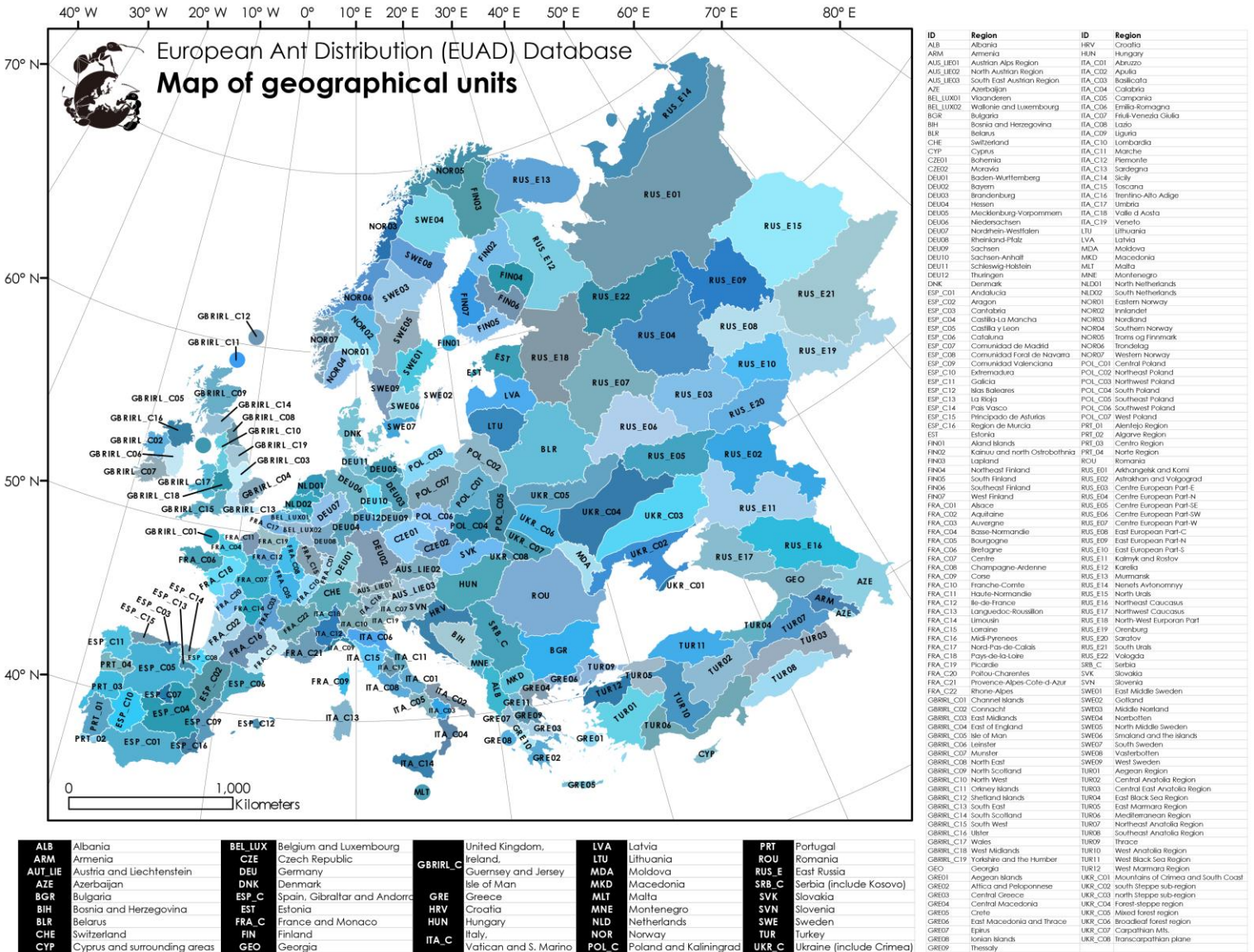
790
791



792

793 **Figure S2.1.** Conceptual diagram of the workflow of this study.

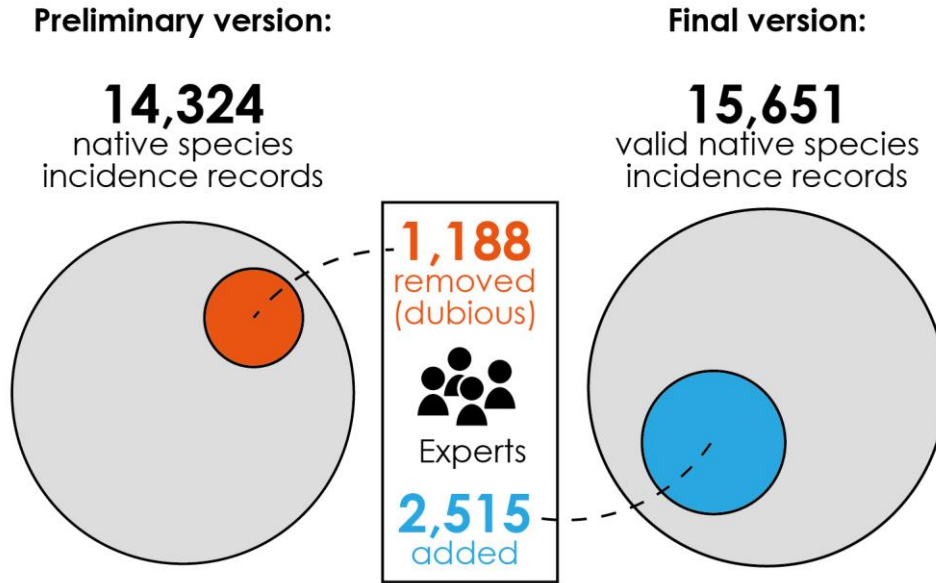
794



795 **Figure S2.2.** Map of geographical units used in the European Ant Distribution (EUAD)
 796 database. Abbreviations indicate the administrative or geographic regions delimited in the
 797 EUAD database.

798

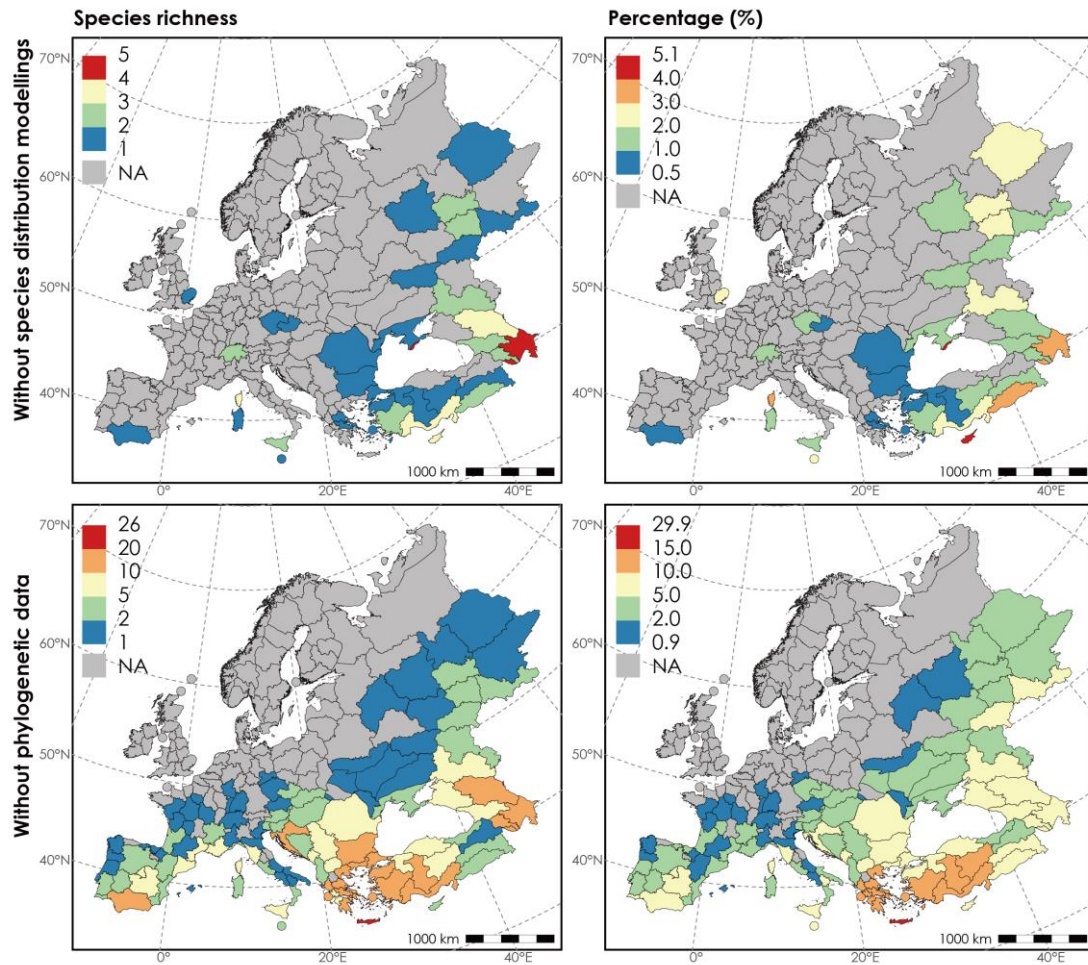
799



800 Incidence record: the unique occurrence of one species in one region

801 **Figure S2.3.** Contribution of experts to the new database. Dubious and new records are
802 removed and added based on experts' opinions, respectively.

803

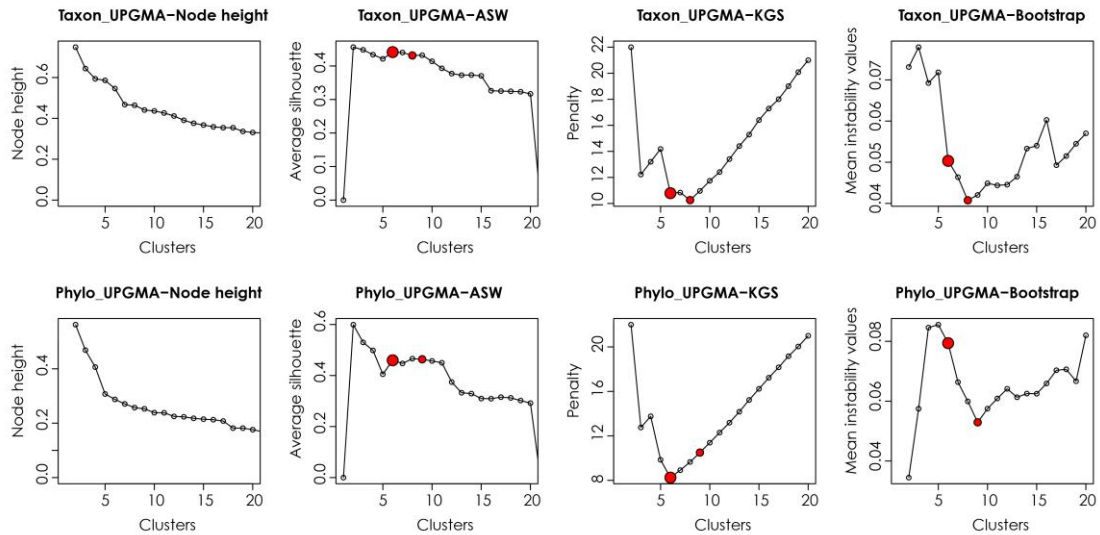


804

805 **Figure S2.4.** Spatial pattern of species without species distribution modellings or phylogenetic

806 data. Both richness and the percentage of missing species are calculated for each region.

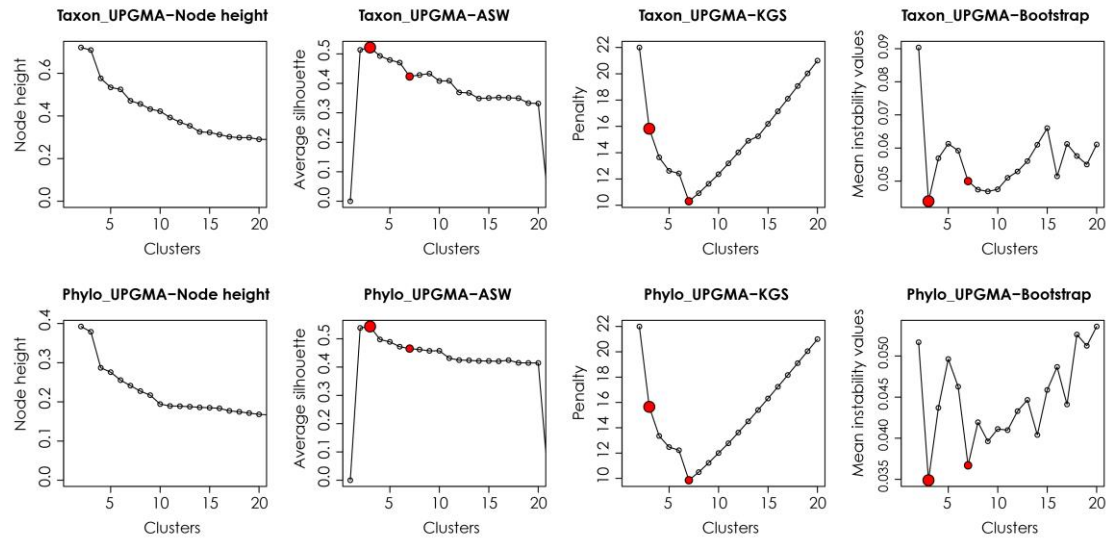
807



808

809 **Figure S2.5.** Evaluation of UPGMA hierarchical clustering of ant regional lists. The node
810 height, average silhouette (ASW), values of Kelly-Gardner-Sutcliffe penalty (KGS) and mean
811 instability values of Bootstrap test are calculated based on taxonomic (upper row) and
812 phylogenetic (lower row) dissimilarity. The red dots indicate the number of clusters chosen for
813 biogeographic regions (large dots) and subregions (small dots).

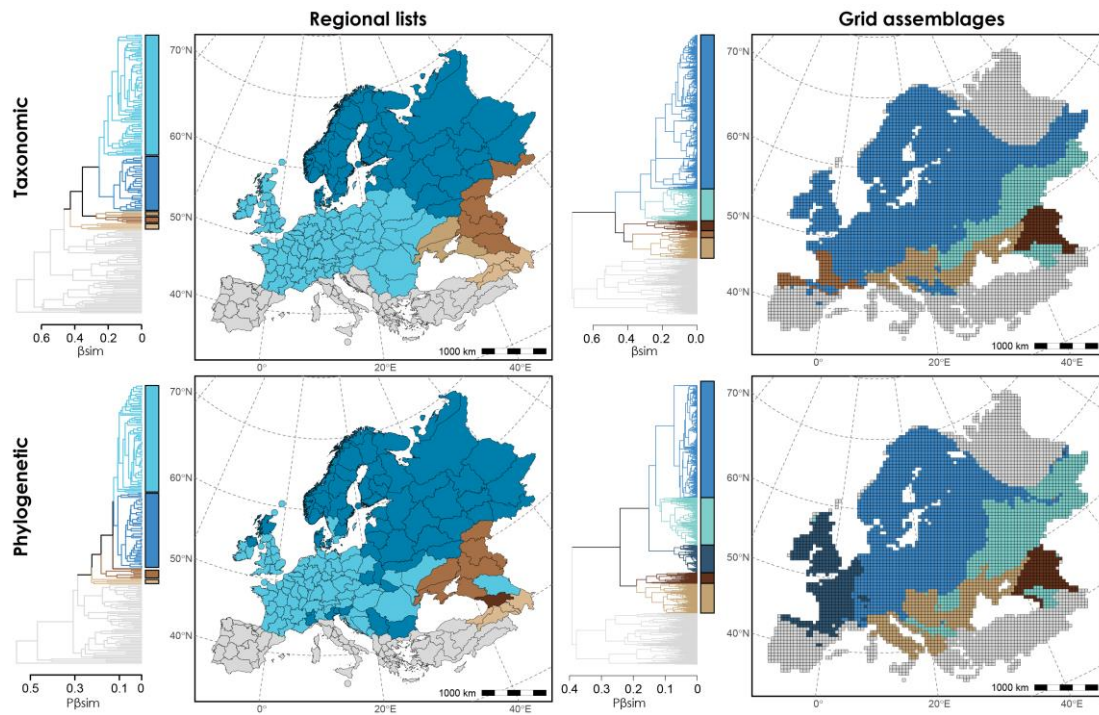
814



815

816 **Figure S2.6.** Evaluation of UPGMA hierarchical clustering of ant grid assemblages. The node
817 height, average silhouette (ASW), values of Kelly-Gardner-Sutcliffe penalty (KGS) and mean
818 instability values of Bootstrap test are calculated based on taxonomic (upper row) and
819 phylogenetic (lower row) dissimilarity. The red dots indicate the number of clusters chosen for
820 biogeographic regions (large dots) and subregions (small dots).

821



822

823 **Figure S2.7.** Different subregions identified within the European region (EU) of ants.

824 Subregions are delineated based on different geographic units (i.e., regional lists and grid

825 cells) and approaches (i.e., taxonomic and phylogenetic).