Global biogeographical regions reveal a signal of past human impacts

Marta Rueda[1,2*], Manuela González-Suárez[3], Eloy Revilla[2]

1. Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Sevilla, Spain.

2. Departamento de Biología de la Conservación, Estación Biológica de Doñana (EBD-CSIC), Sevilla, Spain

3. Ecology and Evolutionary Biology, School of Biological Sciences, University of Reading, Reading, UK.

Corresponding author:

Marta Rueda

Tel: +34656532402

ORCID: https://orcid.org/0000-0002-9754-6258

Email: mruedag@us.es

Abstract

Ecologists have long documented that the world´s biota is spatially organized in regions with boundaries shaped by processes acting on geological and evolutionary timescales. Although growing evidence suggests that human impact has been key in how biodiversity is currently assembled, its role as a driver of the geographical organization of biodiversity remains unclear. Here, we quantify the relative importance of human land use from ~5000 years ago to predict the current assemblage of terrestrial mammals in biogeographical regions across the Earth. Results show that past anthropogenic land use has left an imprint on the taxonomic differentiation of some of the largest biogeographical realms, whereas land use at present stands out as a driver of the taxonomic differences between medium-sized subregions, i.e., within and among continents. Our findings highlight the far-reaching effect that past anthropogenic actions have had on the organization of biodiversity globally.

Introduction

How is the world's biodiversity organized, and why do large-scale patterns of taxonomic diversity change through natural geographic regions? These questions have attracted the attention of naturalists since the early 19th century (von Humboldt, 1806; Wallace, 1876; Ricklefs, 2004; Daru et al., 2017) and are now critical to forecast the future of biodiversity in the face of global change (Botkin et al., 2017). A key step in understanding the organization of biodiversity is to identify the assemblage of regions based on their shared organisms (Wiens, 2011). Alfred R. Wallace was among the first to propose that the world's fauna is organized hierarchically in broad regions shaped by geographic and climatic factors (Wallace, 1876). About 150 years later, the development of multivariate analytical techniques has led to revaluations of Wallace's proposal (Kreft and Jetz, 2010; Procheş and Ramdhani, 2012; Holt et al., 2013; Rueda et al., 2013) and the refinement of our understanding regarding the extrinsic determinants driving the main dissimilarities among bioregions (Ficetola et al., 2017). Broadly, what we know is that the interplay of multiple factors has jointly contributed to the formation of bioregions, but their prominence varies across the globe (Riddle and Hafner 2010; Ficetola et al., 2017; Mazel et al., 2017). Processes acting deeper in the past, like plate tectonics, are the most important in explaining the separation between strongly divergent biogeographical realms. Factors representing present-day ecological barriers, like climatic heterogeneity, determine the separation between less dissimilar regions, i.e., within and among continents; while mountains - the third factor in importance - operate at all levels of biogeographical differentiation (Ficetola et al., 2017). However, although there is clear evidence that anthropogenic actions are the main driver of the extinction processes acting at multiple scales (Barnosky et al., 2011; Ceballos et al., 2017) and, thereby, of the biotic homogenization of the

ecological communities worldwide (Olden, 2006; Baiser et al., 2012), we do not yet fully understand whether and how anthropogenic impact has caused changes at a biogeographical level.

Anthropogenic impact has usually not been contemplated in the biogeographical scenario likely because extant biogeographical regions have traditionally been assumed to reflect the natural organization of biodiversity that resulted from ecological, historical, and evolutionary processes acting over millions of years (Lomolino et al., 2010). However, rising evidence showing how human-mediated species introductions are already affecting extant biogeographic regions (Capinha et al., 2015; Bernardo-Madrid et al., 2019) challenges this view. Furthermore, accumulating facts indicate that Quaternary human activities induced shifts in the plant and animal communities we see today (Lyons et al., 2016), and there is little doubt that modern humans have been a major driver in the extinction of large mammals during the late Pleistocene and early Holocene (Sandom et al., 2014; Smith et al., 2018). Moreover, archaeological, palaeoecological and genetic data suggest that cumulative human transformation of ecosystems over millennia has resulted in dramatic changes in composition, community structure, richness and genetic diversity of a diverse array of organisms across taxonomic groups (Ellis et al., 2013, 2021; Boivin, 2016; Mottl et al., 2021). This body of evidence raises the question of whether historical anthropogenic pressures may have been large enough to leave an imprint detectable today on biogeographical assemblages globally.

To answer these questions, we set up a bioregionalization and analytical procedure focused on two data sets; one describing the current distribution ranges of global terrestrial mammals, and

another set inferring their present natural ranges, which represents estimates of where species would potentially live without anthropogenic pressures (Faurby and Svenning, 2015; Faurby et al., 2018). Using both types of distribution ranges we build hierarchical bioregionalizations, upscaling from the smallest detectable bioregions to the largest realms, and compute models to identify those variables (plate tectonics, climatic heterogeneity, mountains and anthropogenic land use over the Holocene, specifically 5000 and 2000 years ago as well as present time) that best predict the assemblage of species in bioregions. We hypothesized that if anthropogenic land use has had a noteworthy effect on the biogeographical assemblages and boundaries as we know them nowadays, then (1) the biogeographic configurations built using current distribution ranges will differ from those based on natural ranges; and (2) human land use should predict the biogeographic patterns obtained using current ranges, but not patterns obtained using natural ranges. To reinforce our results, we also run complementary models using only the distribution ranges of those extant terrestrial mammals that have undergone some modification in their range due to anthropogenic pressures (536 species), i.e. they have an estimated distribution range in PHYLACINE different from that of the IUCN (2016) (Faurby et al., 2018). Besides, at the largest biogeographical scale we also compute models excluding Madagascar and Australia as these bioregions are characterized by an extremely low historical land use, which could influence results.

Our analyses reveal that historic anthropogenic land use have left an imprint on how diversity is structured on Earth at a biogeographical scale and are consistent with increasingly evidence showing that the impact of past human actions on ecosystems and organisms has been greater than previously thought.

Results

Bioregionalizations

Both for the current and natural bioregionalizations we identified a hierarchical system of biogeographic regions with four levels (Appendix 1-Figure 1). For current mammalian ranges, the broadest delineation of nine large bioregions was mostly consistent with the original maps of Wallace´s realms and subregions (Figure 1a). For natural mammalian ranges, the eight bioregions obtained showed a different biogeographic arrangement with respect to the nine large current ones, especially in the Northern Hemisphere (Figure 1b) (Table 1). Differences between current and natural bioregions ('Material and Methods') extended to the subregions or medium-sized bioregions and smaller bioregions (Appendix 1-Figure 1) and were confirmed for the bioregionalizations carried out with the 536 species whose distribution ranges have been modified by anthropogenic actions (Appendix 1-Table 1).

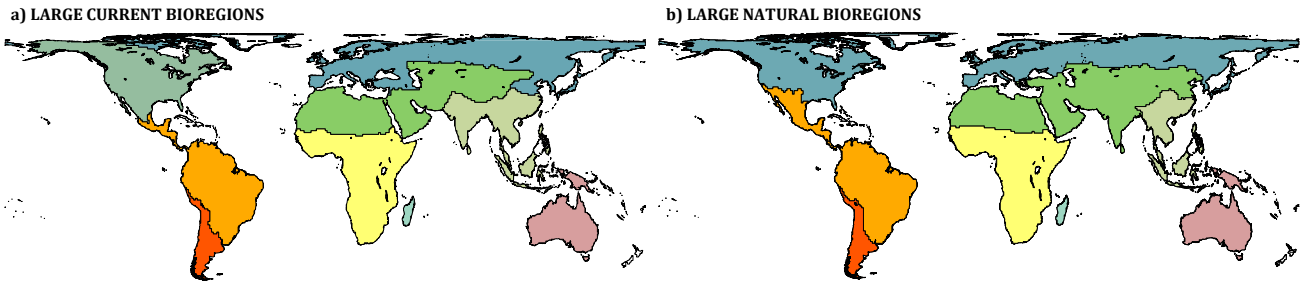

Fig. 1. Current and natural broader regions or realms. *Only results for the broader regions are shown here: 9 current bioregions (a), and 8 natural bioregions (b). The greatest differences between both biogeographical configurations are more relevant in the Northern Hemisphere, where for the natural bioregions we found that the Palearctic and the Nearctic are united forming a great extension, the typical biographical boundary between the Nearctic and the Neotropical*

realms disappears and extends towards the north, and India is no longer part of the Asian region and is included instead within the 'Saharo-Arabian' region.

Table 1. Values of Jaccard index between current and natural hierarchical bioregions. *Jaccard index assesses the degree of similarity between clusters (bioregions) and ranges between 0 (no similarity) and 1 (perfect match). The number of bioregions obtained during each bioregionalization processes is also provided.*

CURRENT	Jaccard coefficient	NATURAL
9 broader realms	0.57	8 broader realms
27 subregions	0.58	25 subregions
141 small bioregions	0.42	121 small bioregions
1128 smaller bioregions	0.28	1051 smaller bioregions

Global importance of predictors of the biogeographical patterns

For the broader bioregions, past human land use (5000 and 2000 years ago) was the third most important predictor of the biogeographic structure of current realms behind plate tectonics and seasonality in temperature and ahead of mountains (Figure 2c, 2e), with a moderate but non-negligible predictive value (~22% and 21% of cells classified correctly when the model was run just with human land use 2000 and 5000 years ago, respectively, see Figure 3). Human land use at present had less relevance in the classification models (Figure 2a). In accordance with our prediction, past and present human land use were poorer predictors of natural bioregions (Figure 2b, d, f). Results were consistent also if we excluded Australia and Madagascar, two biogeographic regions with minimal human land use before 2000 years ago (Appendix 1-Figure 3). Models run with the unique 536 extant terrestrial mammals whose range of distribution has been affected by human impact revealed present and past human land use as the second

161 predictor but with negligible importance with respect to plate tectonics, which was by far the
162 most important driver (Appendix 1-Figure 4). The misclassification rate (Out-of-bag, OBB
163 hereafter) for these set of models was quite low, never exceeding 3%.
164
165 For medium-sized bioregions or subregions, past and present human land use were the second
166 predictors in importance for the current bioregions (Figure 4 a, c, d; OBB values in general <
167 7.50%). As hypothesized, the importance of land use was smaller for the natural bioregions,
168 although its past land use remained a relatively strong driver (Figure 4 b, d, e). Results were
169 consistent for the models run with the unique 536 extant terrestrial mammals influenced by
170 human actions (Appendix 1-Figure 5; OBB < 8.5%).
171
172 As bioregions became smaller, many more variables were necessary to predict the classification
173 of cells into bioregions, and results for the current and natural bioregions became quite similar
174 (Appendix 1-Figure 6 and 7; OBB < 15.1% and < 35.6% for small and very small bioregions,
175 respectively). Overall, human land use at present has minor relevance in the classification
176 models, whereas past human land use remained relatively high, being third or fourth in
177 importance (Appendix 1-Figure 6 and 7).
178

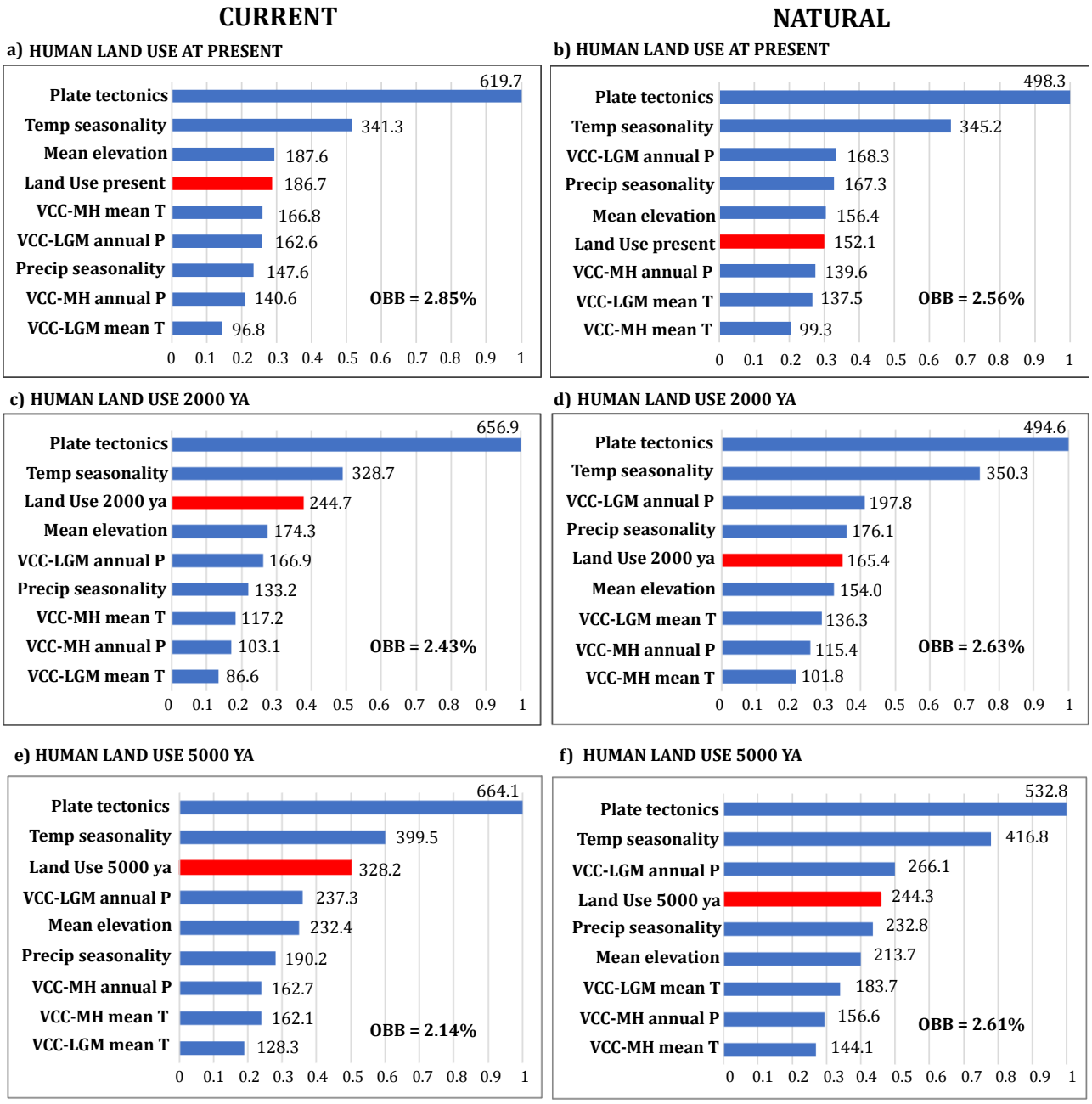

Fig. 2. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural broader bioregions*. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2000 years ago for (c) current and (d) natural bioregions, and human land use 5000 years ago for (e) current and (f) natural bioregions. Importance was measured by the drop-in classification accuracy after predictor randomization in*

random forests of 5000 trees. Higher values of mean decreased in accuracy indicate variables that are more important to the classification. OBB (Out-of-bag) represents the percentage of cells misclassified in each model. VCC = velocity of climate change; MH = Mid-Holocene; LGM = Last Glacial Maximum; annual P = annual precipitation; mean T = mean annual temperature; ya = years ago.

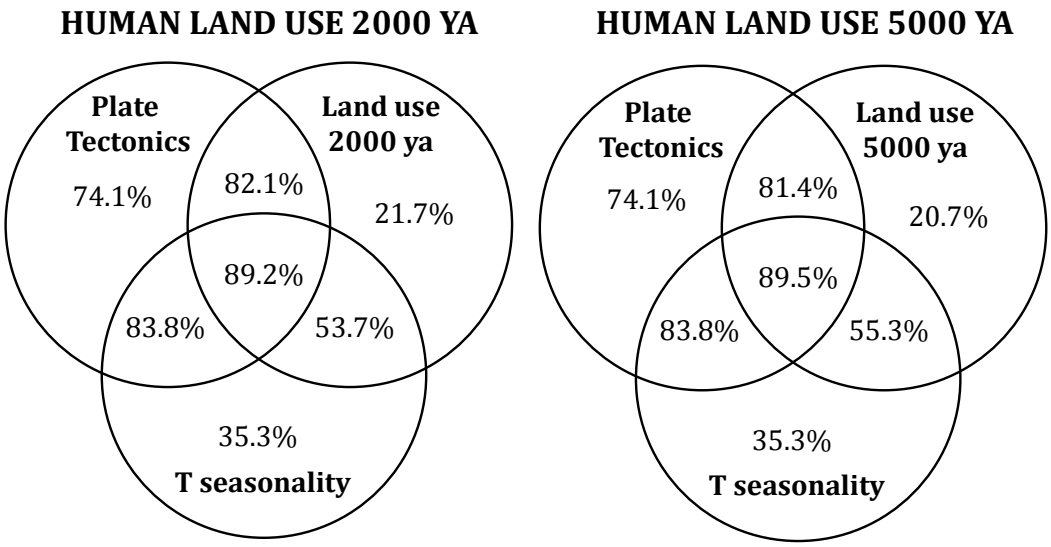

Fig. 3. Diagrams describing the accuracy in classification of the three most important drivers of the current broader bioregions for models including human land use 2000 and 5000 years ago. *Values indicate the percentage of cells correctly classified (1-OBB error) in their bioregion, and are obtained from running random forest classification models for each individual variable, by pairs and for the three most important factors together.*

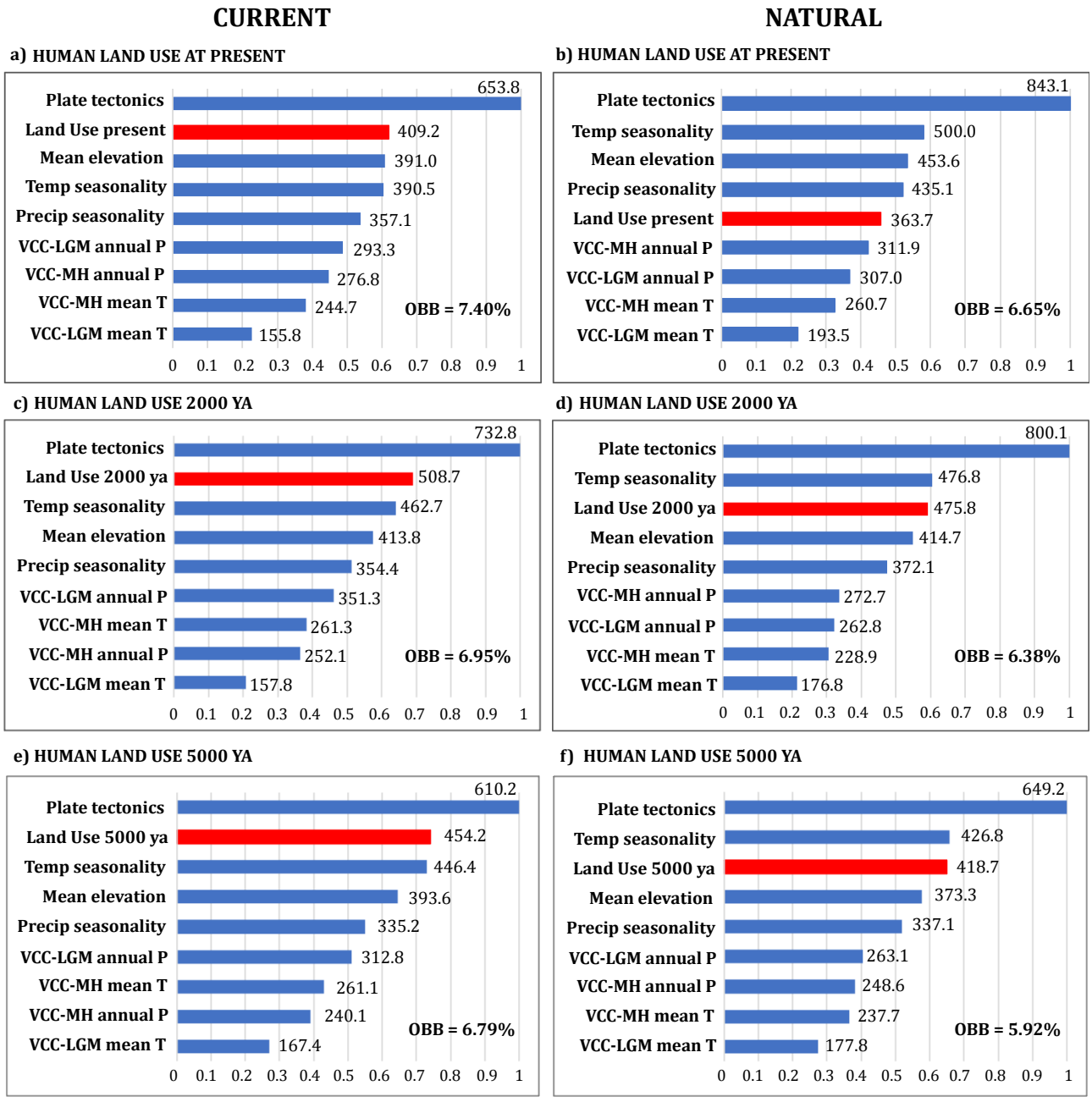

Fig. 4. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural medium-sized bioregions or subregions*. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2000 years ago for (c) current and (d) natural bioregions, and human land use 5000 years ago for (e) current and (f) natural bioregions. Specifications as in Fig. 2.*

209 **Local importance of predictors of the biogeographical patterns**

210 Maps showing local importance scores ('Material and Methods') revealed large differences in
211 the location of grid cells where past and present human land use have been analytically
212 important for the classification of the broader current bioregions (Figure 5). The cells where
213 human land use at present was most relevant are located in southern Africa, the easternmost
214 boundary between the Palearctic and the Asian regions, part of the northern boundary between
215 the Palearctic and 'Saharo-Arabian' regions, and much of the arctic zone of the Palearctic
216 (Figure 5a), although note that grid cells of the two latter areas are important because of their
217 lack of human land use (Figure 5b) ('Material and Methods'). The cells where human land use
218 2000 years ago was more important are located mainly throughout Central America and the
219 Andes mountain range, Europe except for the northernmost part, most of the Asian region, the
220 boundary between the Palearctic and the Asian regions (Figure 5c), as well as Madagascar and
221 much of the arctic zone of the Palearctic but due to lack of human land use (Figure 5d). The cells
222 where human land use 5000 years ago was more important mostly coincided with those of
223 2000 years ago, but with minor values of importance (Figure 5 e, f). All these areas mostly
224 coincided for the broader natural bioregions (Figure 6).

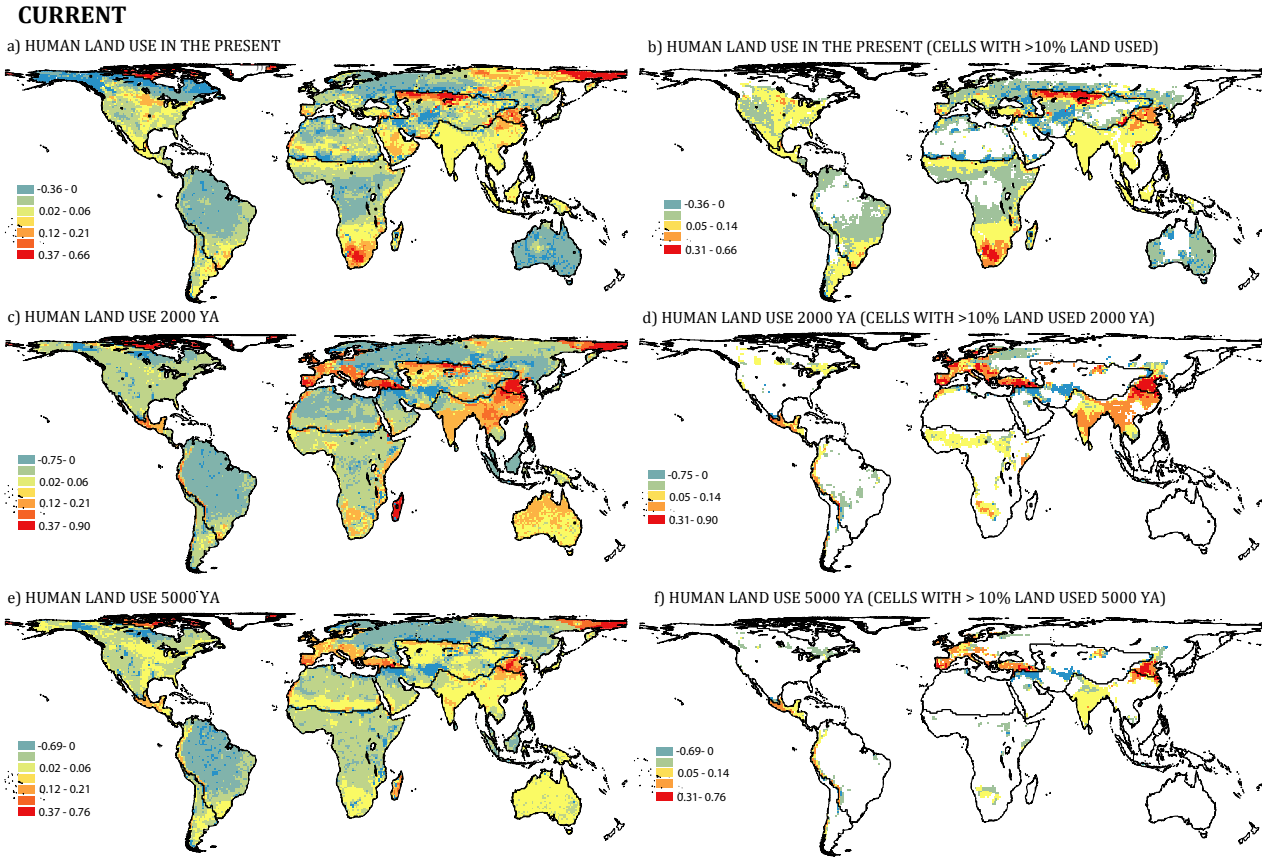

Fig. 5. Maps of the local importance of present and past human land use for the current broader bioregions. *The left panels represent the local importance values for human land use in the present (a), 2000 years ago (c) and 5000 years ago (e). The right panels represent local importance values when land use at present (b), 2000 (d), and 5000 years ago (f) was >10% in each grid cell. The local importance shows the importance of the variable (human land use here) for the classification of a single sample (grid cells). The score of the legends shows the impact on correct classification of grid cells, with blue colors indicating from negative to 0 values (incorrect to neutral classification), and yellow, orange and red indicating positive values (correct classification). Since the values of local importance do not distinguish cells where the human land use was significant and areas where it was negligible, i.e., both can be important for the classification process ('Material and Methods'), we have also mapped the importance values for those grid cells where the percentage of land use at present, as well as 2000 and 5000 years ago was >10% for a better interpretation. In these maps, white cells imply human land use values <10%.*

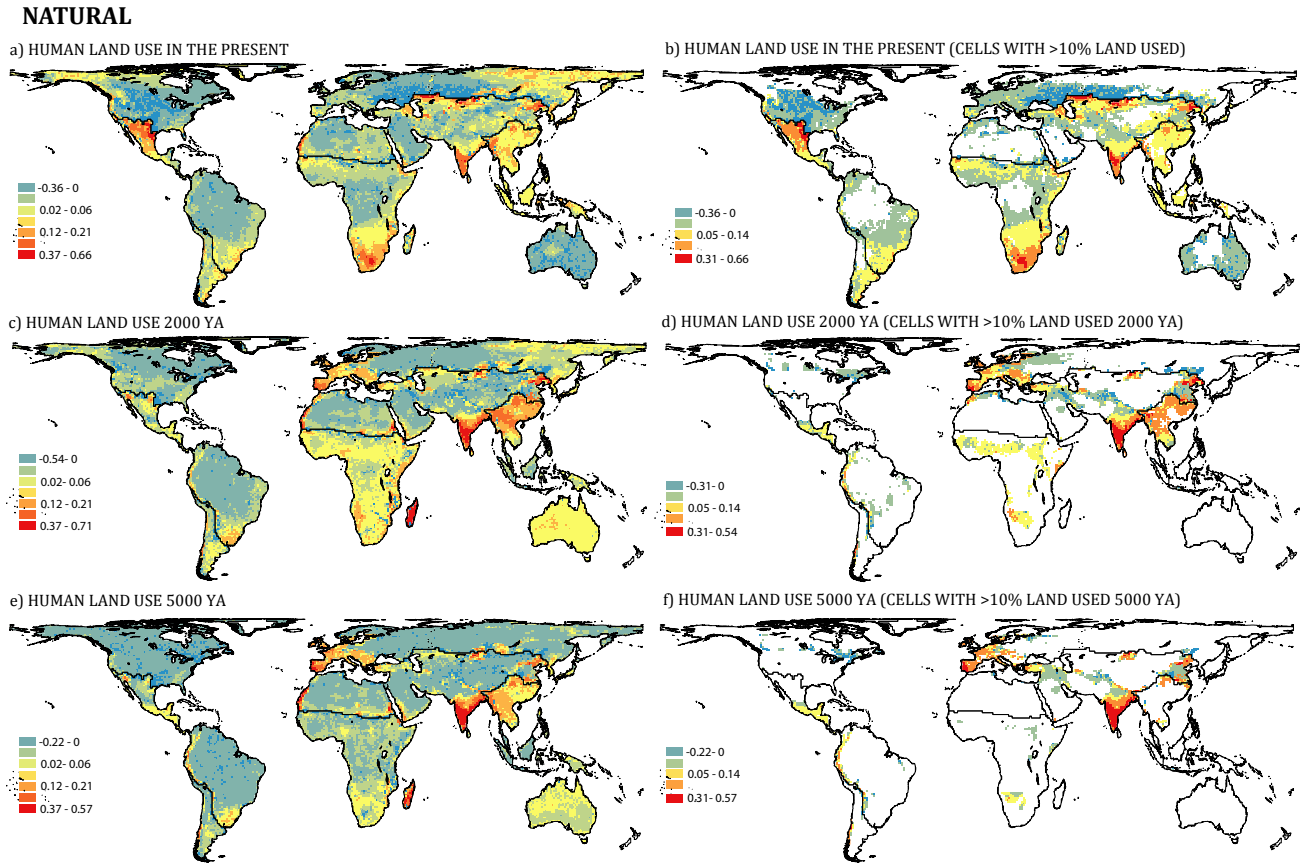

Fig. 6. Maps of the local importance of present and past human land use for the natural broader bioregions. *The left panels represent the local importance values for human land use in the present (a), 2000 years ago (c) and 5000 years ago (e). The right panels represent local importance values when land use at present (b), 2000 (d), and 5000 years ago (f) was >10% in each grid cell. The score of the legends shows the impact on correct classification of grid cells, with blue colors indicating from negative to 0 values (incorrect to neutral classification), and yellow, orange and red indicating positive values (correct classification).*

For the medium current bioregions, the grid cells where the human land use at present was most relevant for the classification of bioregions included those highlighted for the broader bioregions plus most part of Europe, southern India, eastern Australia, a large area of the Nearctic, and the Sahara Desert (Figure 7 a, b). When focusing on the human land use 2000 years ago, the areas affected by human land use were similar to those of the broader bioregions, but note that other areas also became very important (e.g. Sahara Desert, Mongolian Plateau) due to their lack of human land use (Figure 7 c, d). Again, the cells where human land use 5000 years ago was more important mostly coincided with those of 2000 years ago but with minor importance values (Figure 7 e, f). The location of cells affected by human land use for the medium current bioregions quite coincided with that for natural bioregions (Figure 8).

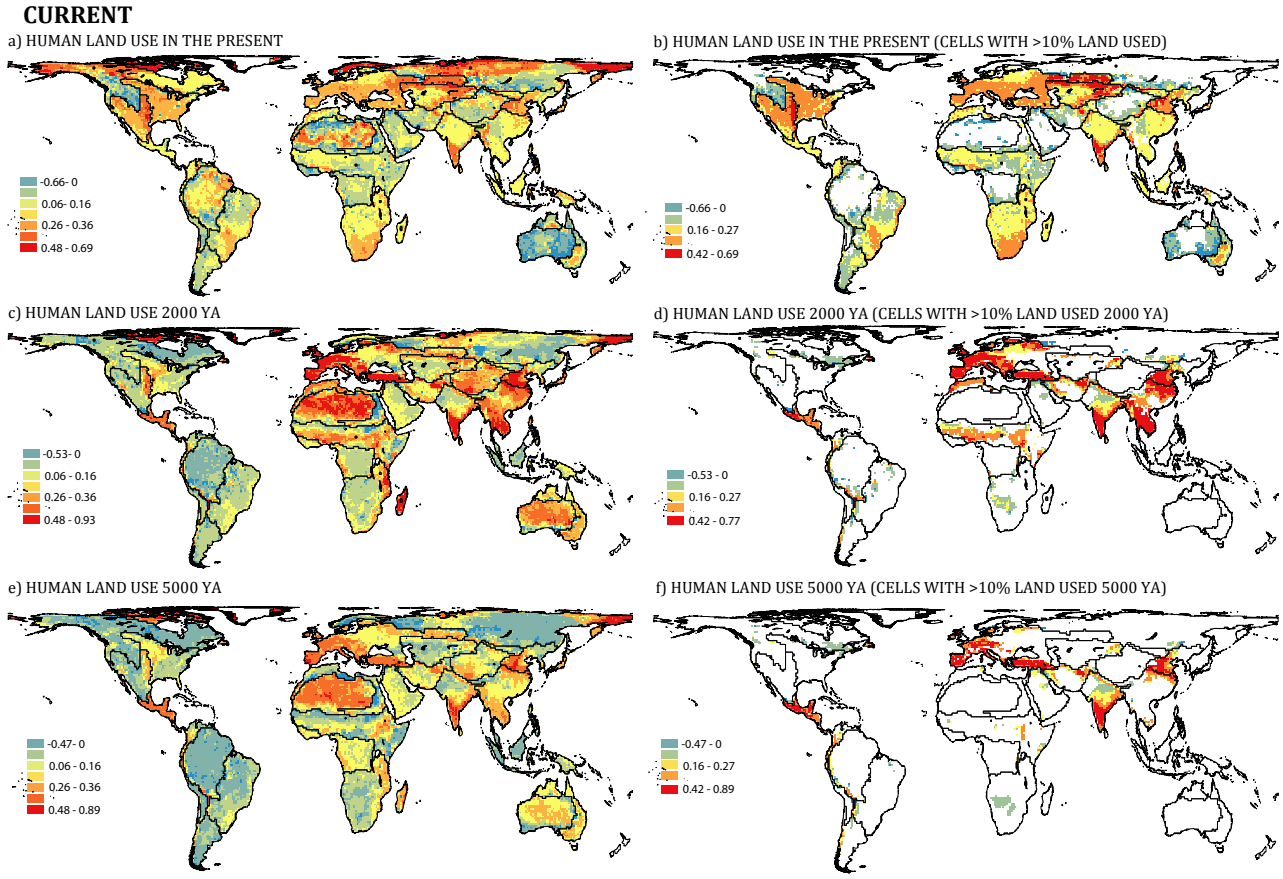

Figure 7. Maps of the local importance of present and past human land use for the medium current bioregions. *The left panels represent the local importance values for human land use in the present (a), 2000 years ago (c) and 5000 years ago (e). The right panels represent local importance values when land use at present (b), 2000 (d), and 5000 years ago (f) was >10% in each grid cell. The score of the legends shows the impact on correct classification of grid cells, with blue colors indicating from negative to 0 values (incorrect to neutral classification), and yellow, orange and red indicating positive values (correct classification).*

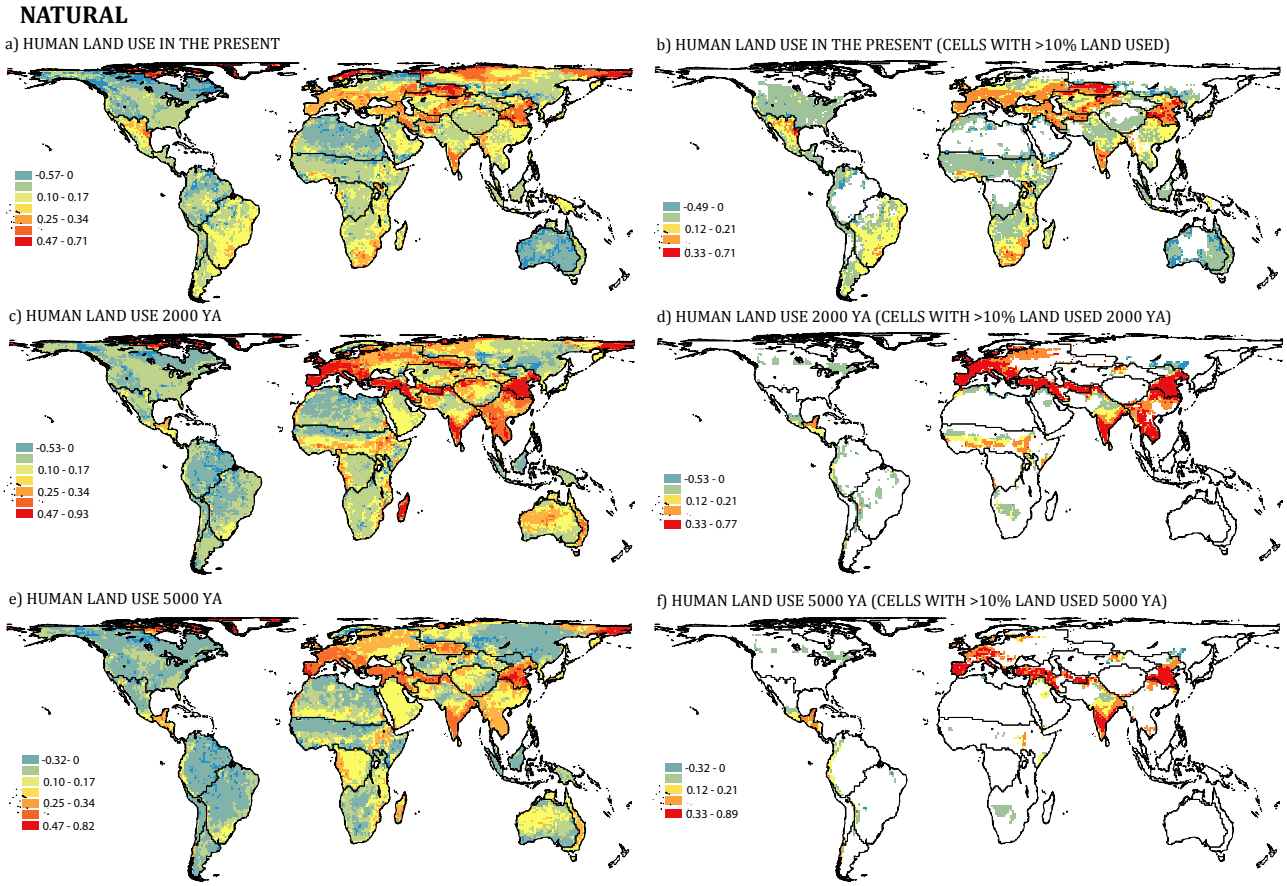

Figure 8. Maps of the local importance of present and past human land use for the medium natural bioregions. *The left panels represent the local importance values for human land use in the present (a), 2000 years ago (c) and 5000 years ago (e). The right panels represent local importance values when land use at present (b), 2000 (d), and 5000 years ago (f) was >10% in each grid cell. The score of the legends shows the impact on correct classification of grid cells, with blue colors indicating from negative to 0 values (incorrect to neutral classification), and yellow, orange and red indicating positive values (correct classification).*

286 **Discussion**

287 Our study shows that human activity during the last few millennia has left its mark on the
288 spatial organization of the Earth´s biodiversity. First, we show that biogeographical
289 configurations based on current and natural mammal distributions differ, particularly in the
290 Northern Hemisphere. These differences were not solely the result of including extinct
291 megafauna - with huge distributions ranges – in the natural biogeographic regions, as
292 biogeographical patterns for the 536 extant terrestrial mammals whose ranges of distribution
293 have been affected by anthropogenic impacts confirmed our findings. Second, we show that
294 human land use is an important predictor of current bioregions but does not largely contribute
295 to predict natural bioregions. For example, for the broader regions or realms, past human land
296 use was the third most important predictor behind plate tectonics and temperature seasonality
297 and ahead of mountains, but has minor importance on natural regions. Past anthropogenic
298 actions have seemingly left a perceptible biogeographical imprint on the assembly of the global
299 realms we recognize today. Moreover, for the medium-sized subregions, human land use at
300 present was the second most important predictor suggesting that recent human activity is
301 already causing changes in the biogeographical assemblages at the subregional scale, i.e., within
302 and among continents.

303

304 The mammalian distributions that we know today are not only a reflection of the most
305 recent human actions but also those exerted during the last few millennia (Faurby and
306 Svenning, 2015; Polaina et al., 2020). Recent studies had already shown that human actions in
307 the present, in terms of human-mediated species introductions, can be evident at
308 biogeographical scales (Capinha et al., 2015; Bernardo-Madrid et al., 2019), but here we also

309 show evidence of an effect of past human land use. What did happen 5000 years ago that left
310 such as noticeable footprint in the Earth's biogeographical regions? This period coincides with
311 the emergence of an “intense” agricultural land expansion and the settlement of densely
312 populated areas in the Mediterranean, Southwest Asia, South Asia, and eastern China (Kirck,
313 2005; Ellis et al., 2013; Stephens et al. 2019). Long-term impacts to enhance agricultural
314 productivity, including forest clearing, increased fire frequencies, species invasions, and soil
315 erosion, were already apparent in some regions at this time (Ellis et al., 2013; Boivin et al.,
316 2016), and may have transformed vegetation structure and species composition across many
317 regions, from the Mediterranean to the Tropics (Stephens et al. 2019). In fact, reconstructions
318 suggest that >20% of Earth’s temperate woodlands had been already impacted by humans by
319 the late-Holocene (Ellis et al., 2013). Indeed, our results on local importance show the match
320 between some of the biogeographical boundaries and regions of the New World and the
321 distribution – and presumed impact – of ancient Mesoamerican and South American pre-
322 European civilizations (Richard, 1997; Dunning et al., 2013). In the Old World, we note the
323 prominence of past human land use covering most of the Asian realm territory, likely
324 associated to the rise of rice cultivation in eastern, southern and southeast Asia over the
325 Holocene (Silva et al., 2015). Indeed, cumulative archaeological data show that the scale of
326 agriculture and land use in these regions was significant; expansion of land area used for
327 livestock and rice (*Oryza sativa*) paddy agriculture in southern Asia was sufficient to increase
328 atmospheric methane emissions between 4000 and 1000 years ago (Fuller et al., 2015).
329 Moreover, the atmospheric CO_2 decline registered between A.D. 1570 and 1620 has been
330 attributed to the uptake by vegetation following the agriculture abandonment caused by the
331 population crash after the arrival of Europeans to America (Lewis and Maslin, 2015). Past

human land use can also partly explain differences between the southern Palearctic boundary and the Afrotropic below the Sahara. Both can be related to the transformation of European and Near Eastern landscapes during the Roman period (Butzer, 2005), and the expansion of sorghum (*Sorghum bicolor*) cultivation 3000 to 2000 years ago in the sub-Saharan Africa (Boivin et al., 2016).

While we found differences in importance of human land use 5000 and 2000 years ago between current and natural medium-sized bioregions, we were surprised that past human land use was an important determinant of natural bioregions, which were based on distributions of mammals in the absence of human impact. These natural ranges are inferred `natural´ distributions and it is possible that they did not fully exclude human impacts, as for 3424 of the extant mammals analyzed, natural and current distributions were identical. In addition, these results could also reflect an overlap between human land use and biogeography. If the distribution of humans during the middle and late Holocene responded to the same drivers as the distributions of terrestrial mammals, human land use could appear as a predictor even if it had no causal effect. In fact, the importance of this variable appears to be a consequence of how human land use was distributed globally 5000 and 2000 years ago, or more accurately where human activity was largely absent. Note that while current human land use is distributed over wide areas of the planet, that of 5000 and 2000 years ago was centralized in specific areas coinciding with the distribution of ancient civilizations. This means that on a subregional scale, many bioregions can be well predicted by the absence of human land use. Indeed, an additional analysis that excluded subregions with no reported human land use 2000 years ago showed

very little importance of past human land use for both current and natural medium-sized bioregions (Appendix 1-Figure 8).

In summary, our results show the value of considering impacts of past human actions to understand the current organization of biodiversity globally. Previous studies have documented lasting effects of human land use changes during the last millennia on current biodiversity patterns (Dambrine et al, 2007; Haberle, 2007; van der Sande et al., 2019), but this is the first time the signal has been recognized on the taxonomic differentiation of the largest realms. This would mean that some of the biogeographical boundaries proposed by Wallace 140 years ago (Wallace, 1876) were already showing the influence of anthropogenic actions. The human transformation of ecosystems that occurred in large terrestrial extensions coinciding with the distribution of ancient civilizations, makes past human land use important along with orographic processes and temperature seasonality to discern taxonomic differences among the large biogeographical zones of the world. It is certainly difficult to disentangle direct causal effects from correlative relationships in non-experimental large-scale studies such as this. We know the same geographic and environmental characteristics that influenced the distribution across space and time of mammals likely also influenced the distribution of our ancestors (Lomolino, 2018). However, we considered several of those predictors and still found a global imprint of past human actions on the patterns of distribution of global diversity. Our results force us to reflect on how the much more widespread and severe changes that have occurred since the beginning of the industrial revolution will affect the organization of biodiversity in the future. We show here the effects of recent human actions are already detectable at the subregional scale, and

377 possible at the broader biogeographical scale. Although it may seem like an unlikely or
378 distant future, the possibility that human activities become the main driver, above historical
379 geomorphological or climatic processes, that shapes Earth´s biodiversity cannot be ignored.
380

381 **Material and Methods**

382 **Generating presence-absence matrices using range distributions**

383 Current and natural distribution ranges of terrestrial mammals were obtained from
384 PHYLACINE 1.2 (Faurby et al., 2018), which contains range maps for all 5831 known mammal
385 species that lived since the last interglacial (~130,000 years ago until present). Ranges for
386 current extant species contained in PHYLACINE are IUCN (2016) distribution maps, whereas
387 present natural ranges represent estimates of where species would potentially live if they had
388 never experience anthropogenic pressures, and – in the case of extinct species – had not gone
389 extinct (see Faurby and Svenning, 2015 for details). Both current and natural ranges in
390 PHYLACINE are projected to Behrmann cylindrical equal area rasters with a cell size of 96.5 km
391 by 96.5 km at 30° North and 30° South. We generated two presence-absences matrices, one for
392 the current and one for the natural ranges, in which every row represents a grid cell and every
393 column a species. From theses matrices we excluded (1) *Homo* species; (2) bats and marine
394 mammals; (3) cells containing less than 50% of land area to approximate equal-size samples,
395 and (4) cells containing fewer than five species to reduce potential distortions caused by having
396 few taxa (Kreft and Jetz, 2010; Rueda et al., 2013). These exclusion criteria rendered a total of
397 14,087 cells representing 3960 terrestrial mammal species for the current matrix, and 14,151
398 cells representing 4306 for the natural matrix. With the aim of conducting complementary
399 analyses to support our results we also generated two additional matrices using only those

400 extant mammals that have undergone some modification in their distribution ranges due to
401 anthropogenic pressure, i.e., they have an estimated distribution range in PHYLACINE different
402 from that of the IUCN, representing 536 species.
403
404 **Building biogeographical regionalizations**
405 To each of the presence-absence matrices we applied a machine-learning algorithm referred to
406 as affinity propagation (AP hereafter) (Frey and Dueck, 2007) to build biogeographical
407 regionalizations at different biogeographical resolution. We chose AP because it can compress
408 massive data sets very efficiently (i.e. with lower error), and its good performance in
409 hierarchical bioregionalization procedures (from few grid cells to large realms) has already
410 been demonstrated in a previous work (Rueda et al., 2013). The AP algorithm works detecting
411 special data points called exemplars, and by a message-passing procedure it iteratively
412 connects every data point to the exemplar that best represents it until an optimal set of
413 exemplars and clusters emerges. Contrary to algorithms in which exemplars are found by
414 randomly choosing an initial subset of data points, AP takes as input measures of 'similarities'
415 between pairs of data points (grid cells here) and simultaneously considers all the points as
416 potential exemplars. The optimal set of exemplars is the one for which the sum of similarities of
417 each point to its exemplar is maximized. Hence, detecting exemplars goes beyond simple
418 clustering because the exemplars themselves store compressed information (Mézard, 2007).
419
420 We first used the presence-absence matrices to calculate pairwise similarities between pairs of
421 cells, and selected Hellinger distance as a similarity index. The Hellinger distance is a
422 modification of the Euclidean distance (Legendre and Gallagher, 2001) and is used to avoid the

423 ‘double-zero’ problem, i.e. when two sites that have no species in common are assigned the
424 same distance as two sites that share the same species; and the ‘species-abundance paradox’,
425 which frequently occurs when two sites share only a small fraction of all the species in the same
426 regional pool. This is expected to be a particular problem at the margins of biogeographical
427 regions where sites may be quite different from one another rather than in the centre of a
428 region where sites are likely to be very similar in their species assemblages (Legendre and
429 Legendre, 1998; Gagné and Proulx, 2009). We then used a protocol based on a successive
430 application of AP (Rueda et al., 2013) to obtain a biogeographical upscaling from the smallest
431 possible bioregions (i.e. the highest biogeographical resolution) to the largest ones. We
432 performed an initial AP analysis involving all grid cells of the similarity matrix. This first AP run
433 generates the optimal solution of the highest resolution bioregions, while also identifying its
434 exemplars. We obtained 1128 clusters and exemplars for the current distributions and 1053 for
435 the natural distributions. Then, using the exemplars as the new units of analysis we conducted
436 again an AP, i.e., we calculated a new similarity matrix and re-run a new AP. This process was
437 repeated until a small and coherent number of large clusters emerged. Finally, to obtain maps
438 of each clustering result, we classified each grid cell (row) of every presence-absence matrix
439 according to the cluster to which they were assigned in its corresponding AP analysis. AP
440 analyses were performed using the '*APCluster*' R package (Bodenhofer et al., 2001).
441

442 **Assessing differences in biogeographical configurations**

443 The degree of similarity between clusters (i.e. regions) of current and natural bioregions was
444 estimated using the Jaccard index. For that, we used the *cluster_similarity* function of the
445 ‘*clusteval*’ R package (Ramey, 2012). This function computes the similarity between two

446 clusterings of the same data. Jaccard index ranges between 0 (no similarity) to 1 (perfect
447 match).
448
449 **Predictors**
450 To each of the cells of the presence-absence matrices we assigned a mean value of several
451 predictors. We considered variables previously tested as determinants of biogeographical
452 boundaries related to climatic heterogeneity, orographic barriers, tectonic movements, and
453 instability of past climate (e.g. Ficetola et al., 2017) plus variables related to anthropogenic land
454 use over the Holocene. In particular, we tested four climatic variables: annual total
455 precipitation, mean annual temperature, seasonality in temperature and seasonality in
456 precipitation. All climatic variables were extracted from the WorldClim dataset (Fick and
457 Hijmans, 2017) up-scaled at a 96.5 km resolution. These variables represent both average
458 conditions and variability within years and have been shown to be determinants of vertebrate
459 distributions (Sexton et al., 2009). Climatic conditions have strongly shifted over the
460 Quaternary period, and have been shown to play a role in the present-day species distributions,
461 endemism, and assemblages (Graham, 1996; Davis and Shaw, 2001; Araujo et al., 2008; Sandel
462 et al., 2011). To test for the potential effect of past climate change or stability, we calculated the
463 average velocity of climate change since the Last Glacial Maximum (LGM; ~22,000 years ago)
464 and since the Mid Holocene (MH; ~ 6000 years ago) (see Sandel et al., 2011). For that, we used
465 the mean annual temperature and annual total precipitation for the MH and the LGM calculated
466 by means of the Model for Interdisciplinary Research on Climate (MIROC-ESM) (Watanabe et
467 al., 2011). Mountain ranges represent major barriers to dispersal for most mammals, whereas
468 plate tectonics are responsible of the long-term isolation of the biotas (Ficetola et al., 2017,

469 Lomolino et al., 2010). We used the GTOPO30 to calculate the mean elevation per grid cell.
470 Plate tectonics were obtained from Bird (2003). Each grid cell was assigned the tectonic plate
471 to which it belongs. When a cell was represented by more than one tectonic plate, it was
472 assigned the one that occupied a greater percentage of the cell. Finally, past anthropogenic land
473 use was obtained from Ellis et al. (2013). This dataset consists of two spatially explicit global
474 reconstructions of land-use from two main models of land use across the Holocene, HYDE and
475 KK_{10} models. Among them, we chose the more realistic KK_{10} model, which assumes that humans
476 use land more intensively when population density is high and land scarce (Kaplan et al., 2011).
477 In counterpoint, the HYDE model omits land-use intensification and predicts that except for the
478 developed regions of Europe, human use of land was insignificant in every biome and region
479 before A.D. 1750. This dataset comprises raster information for human land use for 10 different
480 time periods; B.C. 6000, 3000, and 1000, and A.D. 0, 1000, 1500, 1750, 1900, 1950 and 2000.
481 We calculated human land use at four different time spans: 8000, 5000, and 2000 years ago, as
482 well as at the present time (A.D. 2000). However, we decided do not include land use 8000
483 years ago in the random forest models due to its low representation at global level (Appendix
484 1-Figure 9).
485
486 **Evaluating potential drivers of the biogeographical configurations**
487 We used random forest (Breiman, 2001) models of 5000 classification trees to assess the
488 factors that may predict the classification of cells in bioregions at different biogeographical
489 resolutions and to estimate the relative importance of the predictors. Random forest is a
490 machine learning method based on a combination of a large set of decision trees. Each tree is
491 trained by selecting a random set of variables and a random sample from the training dataset

492 (i.e., the calibration data set). The accuracy of the models is given by the out-of-bag (OOB), an
493 estimate of the misclassification rate that represents and unbiased error rate of the model that
494 is calculated by counting how many cases in the training set are misclassified and dividing the
495 number by the total number of observations.
496
497 Random forests are able to disentangle interacting effects and identify nonlinear and scale-
498 dependent relationships that often occur at the scale of the analysis performed here among
499 multiple correlated predictors (Cutler et al., 2007). Although random forests are generally
500 assumed to not be affected by highly correlated predictor variables, we eliminated some
501 predictors showing a moderate to high correlation ($r > 0.50$, Appendix 1-Table 2) as some
502 evidence from genomic studies suggests that variable importance measures may show a bias
503 towards correlated predictor variables (Nicodemus et al., 2010). So, our final models included
504 seasonality in temperature and precipitation as representatives of climate heterogeneity;
505 velocity of climate change since the LGM and the MH to the present as representatives of past
506 climate change; plate tectonics; mean elevation representing mountains; and human land use at
507 the present, 5000 and 2000 years ago. Human land use 5000 and 2000 years ago showed a
508 large correlation ($r = 0.87$), so we decided to perform separate models for each time period.
509 Globally the correlation between human land use at present and 2000 years ago is relatively
510 low ($r = 0.29$), however we also decided to perform independent models for both predictors
511 because for certain bioregions both human land use at present and 2000 years ago can be
512 highly correlated. We also ran additional models with the 536 extant species that have
513 undergone some modification in their distribution range due to anthropogenic pressures.
514 Finally, the algorithm used in the process of regionalization only uses species presence-

515 absences, which means that humans can only have a role if present and if causing local
516 extinctions, thereby changing the amount of shared species between grid cells. At the largest
517 biogeographical scale two bioregions, Madagascar and Australia, are characterised by an
518 extremely low historical human land use that could influence results. Therefore, we also ran the
519 models excluding Australia and Madagascar. Random forest models were computed using the R
520 package '*RandomForest*' (Liaw and Wiener, 20002).
521
522 **Calculating global and local predictor importance**
523 We measured variable importance using the mean decrease in accuracy, which is obtained by
524 permuting randomly each variable and assessing the decrease in classification accuracy of the
525 model (Liaw and Wiener, 2002). Thus, the more the accuracy of the random forest decreases
526 due to the permutation (or exclusion) of a single variable, the more important that variable is
527 considered, i.e., variables with a large mean decrease in accuracy are more important for
528 classification of the data.
529
530 Random forest also calculates the local variable importance, which defines the importance of a
531 variable in the classification of a single sample (grid cell here) and therefore shows a direct link
532 between variables and samples (Touw et al., 2012). The local importance score is derived from
533 all trees for which the sample was not used to train the tree (i.e. its value is OBB). The
534 percentage of correct votes for the correct class in the permuted OBB data is subtracted from
535 the percentage of votes for the correct class in the original OOB data to assign a local
536 importance score for the variable for which the values were permuted. The score reflects the
537 impact on correct classification of a given sample: negative, 0 (the variable is neutral) and

538 positive. Given that local importance values are noisier than global importance ones we run the
539 same classification 5 times (5 per biogeographical scale) and averaged the local importance
540 scores to obtain a robust estimation of local importance values (Touw et al., 2012). Since one
541 value per cell was obtained, these were used to create maps of the local importance of present
542 and past human land use. However, note that the values of local importance do not distinguish
543 between areas where the human land use was high and areas where it was negligible, i.e., both
544 are equally valuable for the classification analysis. So, to help in the interpretation we also
545 represent in maps the local importance value for those grid cells where the percentage of land
546 used was greater than 10%.
547
548 All analyses were performed in the R Version 3.6.1 (R Core Team 2019).
549
550 **Data availability**
551 All Data supporting this article will be deposited in Dryad upon acceptance.
552
553 **References**
554 Araujo MB, Nogués-Bravo D, Haywood AM, Valdes PJ, Rahbek C (2008). Quaternary climate
555 changes explain diversity among reptiles and amphibians. *Ecography* 31: 8-15.
556 https://doi.org/10.1111/j.2007.0906-7590.05318.x
557 Barnosky et al. (2011). Has the Earth´s sixth extinction already arrived? *Nature* 471: 51-57.
558 https://doi.org/10.1038/nature09678

Bernardo-Madrid R, Calatayud J, González-Suárez M, Rosvall M, Lucas PM, Rueda M, Antonelli A, Revilla E (2019). Human activity is altering the world´s zoogeographical regions. *Ecology Letters* 22: 1297-1305. https://doi.org/10.1111/ele.13321

Baiser B, Olden J D, Record S, Lockwood JL, McKinney ML (2012). Pattern and process of biotic homogenization in the new Pangea. *Proceedings of the Royal Society B* 279: 4772-4777. https://doi.org/10.1098/rspb.2012.1651

Bird P (2003). An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems* 4, 1027. https://doi.org/10.1029/2001GC000252

Bodenhofer U, Kothmeier A, Hochreiter S (2011). APCluster: a R package for affinity propagation clustering. *Bioinformatics* 27, 2463–2464. https://doi.org/10.1093/bioinformatics/btr406

Boivin NL et al. (2016). Ecological consequences of human niche construction: Examining long-tern anthropogenic shaping of global species distributions. *Proceedings of the National Academy of Science USA* 113: 6388-6296. https://doi.org/10.1073/pnas.1525200113

Botkin DB et al. (2007). Forecasting the effects of global warming on biodiversity. *Bioscience* 57, 227-236. https://doi.org/10.1641/B570306

Breiman L (2001). Random forests. *Machine Learning 45:* 5–32.

Butzer KW (2005). Environmental history in the Mediterranean world: cross-disciplinary investigation of cause-and-effect for degradation and soil erosion. *Journal of Archaeological Sci*ence 32: 1773-1800. https://doi.org/10.1016/j.jas.2005.06.001

Capinha C, Essl F, Seebens H, Moser D, Pereira HM (2015). The dispersal of alien species redefines biogeography in the Anthropocene. *Science* 348: 1248-1251. https://DOI: 10.1126/science.aaa8913

Ceballos G, Ehrlich PR, Dirzo R (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Science USA* 114: E6089-E6096. https://doi.org/10.1073/pnas.1704949114

Cutler DR, Edwards Jr. TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler J (2007). Random forests for classification in ecology. *Ecology* 88: 2783–2792. https://doi.org/10.1890/07-0539.1

Dambrine E, Dupouey J-L, Laüt L, Humbert L, Thinon M, Beaufils T, Richard H (2007). Present forest biodiversity patterns in France related to former Roman agriculture. *Ecology* 88: 1430-1439. https://doi.org/10.1890/05-1314

Daru BH, Elliot TL, Park DS, Davies TJ (2017). Understanding the processes underpinning patterns of phylogenetic regionalization. *Trends in Ecology and Evolution* 32: 845-860. https://doi.org/10.1016/j.tree.2017.08.013

Davis MB, Shaw RG (2001). Range shifts and adaptive responses to Quaternary climate change. *Science* 292: 673–679. https://DOI: 10.1126/science.292.5517.673

Dunning NP, Beach TP, Luzzadder-Beach S (2013). Kax and kol: Collapse and resilience in lowland Maya civilization. *Proceedings of the National Academy of Science USA* 109: 3652-3657. https://doi.org/10.1073/pnas.1114838109

Ellis EC, Kaplan JO, Fuller DQ, Vavrus S, Goldewijk KK, Verburg PH (2013). Used planet: a global history. *Proceedings of the National Academy of Science USA* 110, 7978-7985. https://doi.org/10.1073/pnas.1217241110

Ellis EC et al. (2021). People have shaped most of terrestrial nature for at least 12,000 years. *Proceedings of the National Academy of Science USA.* 118 (17) e2023483118. https://doi.org/10.1073/pnas.2023483118

605 Faurby S, Svenning J-C. (2015). Historic and prehistoric human-driven extinctions have
606 reshaped global mammal diversity patterns. *Diversity and Distributions* 21: 1155-1166.
607 https://doi.org/10.1111/ddi.12369
608 Faurby S, Davis M, Pedersen RO, Schowannek SD, Antonelli A, Svenning J-C (2018). PHYLACINE
609 1.2: The Phylogenetic atlas of mammal macroecology. *Ecology* 99: 2626
610 DOI: 10.1002/ecy.2443
611 Frey BJ, Dueck D (2007). Clustering by passing messages between data points. *Science* 315:
612 972-976. DOI: 10.1126/science.1136800
613 Ficetola GF, Mazel F, Thuiller W (2017). Global determinants of zoogeographical boundaries.
614 *Nature Ecology & Evolution* 1, 0089. https://doi.org/10.1038/s41559-017-0089
615 Fick SE, Hijmans RJ (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for
616 global land areas. *International Journal of Climatology* 37: 4302-4315.
617 https://doi.org/10.1002/joc.5086
618 Fuller DQ, van Etten J, Manning K, Castillo C, Kingwell-Banham E, Weiiskopf A, Qin L, Sato Y,
619 Hijmans RJ (2011). The contribution of rice agriculture and livestock pastoralism to
620 prehistoric methane levels. An archaeological assessment. *Holocene* 21: 743-759.
621 https://doi.org/10.1177/0959683611398052
622 Gagné S, Proulx R (2009). Accurate delineation of biogeographical regions depends on the use
623 of an appropriate distance measure. *Journal of Biogeography* 36, 561–567.
624 https://doi.org/10.1111/j.1365-2699.2008.01990.x
625 Graham RW et al. (1996). Spatial response of mammals to Late Quaternary environmental
626 fluctuations. *Science* 272: 1601-1606. DOI: 10.1126/science.272.5268.1601

Haberle SG (2007). Prehistoric human impact on rainforest biodiversity in highland New Guinea. *Philosophical Transactions of the Royal Society B* 362: 219-228. https://doi.org/10.1098/rstb.2006.1981

Holt B, et al. (2013). An update of Wallace´s zoogeographic regions of the world. *Science* 339: 74–78. DOI: 10.1126/science.1228282

IUCN 2016. IUCN red list of threatened species. Version 2016-3. http://www.iucnredlist.org

Kaplan JO,Krumhardt KM, Ellis EC, Ruddiman WF, Lemmen C, Goldewijk KK (2011). Holocene carbon emissions as a result of anthropogenic land cover change. *Holocene* 21: 775–791. https://doi.org/10.1177/0959683610386983

Kreft H, Jetz WA (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography* 37: 2029-2053. https://doi.org/10.1111/j.1365-2699.2010.02375.x

Kirck PV (2005). Archaeology and Global Change: The Holocene Record. *Annual Review of Environment and Resources* 30: 409-440. https://doi.org/10.1146/annurev.energy29.102403.140700

Legendre P, Legendre L (1998). *Numerical ecology*, (Elsevier, Amsterdam, 1998).

Legendre P, Gallagher ED (2011). Ecological meaningful transformations for ordination of species data. *Oecologia* 129: 271-280.

Lewis SL, Maslin MA (2015). Defining the Anthropocene. *Nature* 519: 171-180. https://doi.org/10.1038/nature14258

Liaw A, Wiener M (2002). Classification and regression by RandomForest. *R News* 2: 18–22.

Lomolino MV, Riddle BR, Whittaker RJ, Brown JH (2010). *Biogeography* (Sinauer Associates, 2010).

Lomolino MV (2018). On teaching "…that grand subject, …". *Front. Biogeogr.* 10: e37812. 10.21425/F5FBG37812

Lyons SK et al. (2016). Holocene shifts in the assembly of plant and animal communities implicate human impacts. *Nature* 529: 80-85. https://doi.org/10.1038/nature16447

Mazel F, Wüest RO, Lessard J-P, Renaud J, Ficetola GF, Lavergne S, Thuiller W (2017). Global patterns of β-diversity along the phylogenetic time-scale: The role of climate and plate tectonics. *Global Ecology and Biogeography* 26: 1211– 1221. https://doi.org/10.1111/geb.12632

Mézard M (2007). Where are the exemplars? *Science* 315: 949-951. DOI: 10.1126/science.1139678

Mottl O et al. (2021). Global acceleration in rates of vegetation change over the past 18,000 years. *Science* 372: 860-864. DOI: 10.1126/science.abg1685

Nicodemus KK, Malley, JD, Strobl C, Ziegler A (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinformatics* 11: 110.

Olden JD (2006). Biotic homogenization: a new research agenda for conservation biogeography. *Journal of Biogeogr*aphy 33: 20-27-2039. https://doi.org/10.1111/j.1365-2699.2006.01572.x

Polaina E, González-Suárez M, Revilla E (2020). The legacy of past human land use in current patterns of mammal distribution. *Ecography* 42, 1623-1635. https://doi.org/10.1111/ecog.04406

Procheş Ş, Ramdhani S (2012). The world´s zoogeographical regions confirmed by cross-taxon analyses. *Bioscience* 62, 260–270 (2012). https://doi.org/10.1525/bio.2012.62.3.7

R Core Team. R: *A Language and Environment for Statistical Computing, 3.6.1* (R Foundation for Statistical Computing, 2019).

Ramey JA (2012). Clusteval: Evaluation of Clustering Algorithms. R package version 0.1. https: //CRAN-R-project.org/package=clusteval.

Rao CR (1995) A review of canonical coordinates and an alternative to correspondences analysis using Hellinger distance. *Qüestiió* 19, 23–63.

Richard EWA (1997). *Ancient civilizations of the new world: Essays in world history* (Westview Press, Boulder, Co, 1997).

Ricklefs RE (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters* 7: 1-15. https://doi.org/10.1046/j.1461-0248.2003.00554.x

Riddle BR, Hafner DJ (2010). Integrating pattern with process at biogeographic boundaries: the legacy of Wallace. *Ecography* 33: 321–325. doi: 10.1111/j.1600-0587.2010.06544.x

Rueda M, Rodriguez MA, Hawkins BA (2013) Identifying global zoogeographical regions: lessons from Wallace. *Journal of Biogeography* 40: 2215–2225. https://doi.org/10.1111/jbi.12214

Sandel B, Dalsgaard AB, Davies RG, Gaston KJ, Sutherland WJ, Svenning J-C (2011). The influence of late Quaternary climate-change velocity on species endemism. *Science* 334: 660-664. DOI: 10.1126/science.1210173

Sandom C, Faurby S, Sandel B, Svenning J-C (2014). Global late Quaternary megafauna extinctions linked to humans, not climate change. *Proceedings of the Royal Society B* 281: 20133254. https://doi.org/10.1098/rspb.2013.3254

Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009). Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics* 40: 415-436. https://doi.org/10.1146/annurev.ecolsys.110308.120317

Silva F, Stevens CJ, Weisskopf A, Castillo C, Qin L, Bevan A, Fuller DQ (2015). Modelling the geographical origin of rice cultivation in Asia using the rice archaeological database. *PLoS ONE* 10: e0137024. https://doi.org/10.1371/journal.pone.0137024

Smith F, Elliott Smith RE, Lyons SK, Payne J (2018). Body size downgrading of mammals over the late Quaternary. *Science* 360: 310-313. DOI: 10.1126/science.aao5987

Stephens L, et al. (2019) Archaeological assessment reveals Earth's early transformation through land use. *Science* 365: 897-902. DOI: 10.1126/science.aax1192

Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M, van Hijum SAFT (2012). Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle. *Briefings in Bioinformatics* 14: 315-326. https://doi.org/10.1093/bib/bbs034

van der Sande M, Gosling W, Correa-Metrio A, Prado-Junior J, Poorter L, Oliveira RS, Mazzei L, Bush MB (2019). A 7000-year history of changing plant trait composition in an Amazonian landscape; the role of humans and climate. *Ecology Letters:* 22, 925-935. https://doi.org/10.1111/ele.13251

von Humboldt A (1806). *Essai sur la geographie des plantes; accompagné d'un tableau physique des régions équinoxales, accompagné d'un tableau physique des régions équinoctiales* (Schoel & Co., Paris, 1806).

Wallace AR (1876). *The geographical distribution of animals* (Harper & Brothers, New York, 1876).

Watanabe S et al. (2011). MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development* 4: 845–872. https://doi.org/10.5194/gmd-4-845-2011

Wiens JJ (2011). The niche, biogeography and species interactions. *Philosophical Transactions of the Royal Society B* 366: 2336-2350. https://doi.org/10.1098/rstb.2011.0059

Acknowledgements

This work was supported by the project 2020/125-US/JUNTA/FEDER EU (Programa Operativo FEDER/Junta de Andalucía 2014-2020) to MR, and by CGL2017-83045-R AEI/FEDER EU (Agencia Estatal de Investigación from the Ministry of Economy, Industry and Competitiveness, Spain co-financed with FEDER) to ER.

Competing interests

The authors declare no competing interests.