Global biogeographical regions reveal a signal of past human impacts

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

1. Department of Plant Biology and Ecology, University of Seville, Seville, Spain.

2. Department of Conservation Biology, Estación Biológica de Doñana (EBD-CSIC), Seville, 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 during the Late Holocene to predict the current assemblage of terrestrial mammals in biogeographical regions across the Earth. Results show that human land use 2,000 years ago 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 and warn us about future impacts.

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[1-5] and are now critical to forecast the future of biodiversity in the face of global change[6]. A key step in understanding the organization of biodiversity is to identify the assemblage of regions based on their shared elements[4]. 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[1]. About 150 years later, the development of multivariate analytical techniques has led to revaluations of Wallace's proposal[7-10] and the refinement of our understanding regarding the extrinsic determinants driving the main dissimilarities among bioregions[11]. 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[11-13]. 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[11]. However, although there is clear evidence that anthropogenic actions are the main driver of the extinction processes acting at multiple scales[14,15], and, thereby, of the biotic homogenization of the ecological communities worldwide[16,17], we do not yet fully understand whether and how anthropogenic impact has caused changes at a biogeographical level.

71 Anthropogenic impact has usually not been contemplated in the biogeographical scenario likely
72 because extant biogeographical regions have traditionally been assumed to reflect the natural
73 organization of biodiversity that resulted from ecological, historical, and evolutionary
74 processes acting over millions of years[12]. However, rising evidence showing how human-
75 mediated species introductions are already affecting extant biogeographic regions[18,19]
76 challenges this view. Furthermore, accumulating facts indicate that Quaternary human
77 activities induced shifts in the plant and animal communities we see today[20-22], and there is
78 little doubt that modern humans have been a major driver in the extinction of large mammals
79 during the late Pleistocene and early Holocene[21,22]. Moreover, archaeological, palaeoecological
80 and genetic data suggest that cumulative human transformation of ecosystems over millennia
81 has resulted in dramatic changes in composition, community structure, richness and genetic
82 diversity of a diverse array of organisms across taxonomic groups[23,24]. This body of evidence
83 raises the question of whether anthropogenic activities may have been large enough to leave an
84 imprint detectable today on biogeographical assemblages globally.
85
86 To answer this question, we set up a bioregionalization and analytical procedure focused on
87 two data sets; one describing the current distributions of global terrestrial mammals and
88 another inferring their natural distributions (distribution of terrestrial mammals from the
89 Pleistocene to the present in the absence of human impact[25]). We then build hierarchical
90 bioregionalizations, upscaling from the smallest detectable bioregions to the largest realms,
91 applying an affinity propagation clustering algorithm[26] to the corresponding co-occurring
92 species matrix of global terrestrial mammals. Finally, we compute random forest classification
93 models to identify those variables that best predict bioregion assignation for all cells across the

94 globe. In the models, we include predictors previously identified as important determinants of
95 biogeographical boundaries[11], and predictors associated with anthropogenic activities over the
96 Holocene, specifically, the human land use at present, as well as 2,000 years ago.
97 Bioregionalizations and random forest models are also carried out using only those extant
98 species whose ranges of distribution are known to have undergone shifts due to anthropogenic
99 actions[25], as we expect a biogeographical imprint to be more apparent in this group. We
100 hypothesized that if anthropogenic actions have had a noteworthy effect on the biogeographical
101 assemblages and boundaries as we know them nowadays, then (1) the biogeographic
102 configurations built using current mammal's distributions will differ from those based on
103 natural distributions; and (2) human land use, present and past, should predict the
104 biogeographic patterns obtained using current distributions of mammals, but not patterns
105 obtained using natural distributions. Our analyses reveal that human actions have left an
106 imprint on how diversity is structured on Earth at a biogeographical scale. Our findings are
107 consistent with recent evidence showing that the impact of past human actions on ecosystems
108 and organisms has been greater than previously thought, and warn us about the effect of future
109 impacts.
110

111 **Results**

112 **Bioregionalizations**

113 Both for the current and natural bioregionalizations we identified a hierarchical system of
114 biogeographic regions with four levels showing that global biodiversity patterns can be
115 cohesively shaped from local to regional and realm scales (Fig. S1). For current mammalian
116 distributions, the broadest delineation of nine large bioregions was mostly consistent with the

117 original maps of Wallace´s realms and subregions[1,10] with some particular differences including
118 the limits of the Palearctic[7,9,10], and an extension towards the arid steppes of Mongolia of what
119 we have named the 'Saharo-Arabian' region[7,10] (Fig. 1a). For natural mammalian distributions,
120 the eight bioregions obtained showed a different biogeographic arrangement with respect to
121 the nine large current ones, especially in the Northern Hemisphere (Fig. 1b) (Table 1).
122 Differences between current and natural bioregions extended to the subregions or medium-
123 sized bioregions and smaller bioregions (Fig. S1), and were confirmed for the
124 bioregionalizations carried out with the 536 species whose range of distribution have been
125 modified by anthropogenic actions[25] (Fig. S2) (Table S1).
126

127 **Global importance of predictors of the biogeographical patterns**

128 For the broader bioregions, human land use 2,000 years ago was the third most important
129 driver of the biogeographic structure of current realms behind plate tectonics and seasonality
130 in temperature and ahead of mountains (Fig. 2c), with a moderate but non-negligible predictive
131 value (~22% of cells classified correctly when the model was run just with this predictor, see
132 Fig. 3a). Human land use at present had less relevance in the classification models (Fig. 2a). In
133 accordance with our prediction, past and present human land use were poorer predictors of
134 natural bioregions (Fig. 2b, d). Results were consistent also if we excluded Australia and
135 Madagascar, two biogeographic regions with minimal human land use before 2,000 years ago
136 (Fig. S3). Models run with the unique 536 extant terrestrial mammals whose range of
137 distribution has been affected by human impact revealed present and past human land use as
138 the second predictor after plate tectonics but with little relevance (Fig. S4). The
139 misclassification rate (OBB) for these set of models was quite low, never exceeding 3%.

For medium-sized bioregions or subregions, plate tectonics continued to be the most important determinant of the biogeographic organization of both current and natural bioregions (Fig. 4; OBB values in general < 7.50%), with human land use at present and 2,000 years ago as the second predictors in importance for the current bioregions (Fig. 4a, c). As hypothesized, the importance of land use was smaller for the natural bioregions, although its past land use remained a relatively strong predictor (Fig. 4 b, d). Results were consistent for the models run with the unique 536 extant terrestrial mammals influenced by human actions (Fig. S5; OBB < 8.5%).

As bioregions became smaller, many more variables were necessary to predict the classification of cells into bioregions, and results for the current and natural bioregions became similar (Fig. S6 and S7; OBB < 15.1% and < 35.6% for small and very small bioregions, respectively). Overall, human land use at present has minor relevance in the classification models, whereas human land use 2,000 years ago remained relatively high, being third or fourth in importance (Fig. S6c, d and S7c, d).

Local importance of predictors of the biogeographical patterns

Maps showing local importance scores revealed large differences in the location of grid cells where human land use at present and 2,000 years ago have been analytically important for the classification of the broader current bioregions (Fig. 5). The cells where human land use at present was most relevant are located in southern Africa, the easternmost boundary between the Palearctic and the Oriental regions, part of the northern boundary between the Palearctic

163 and 'Saharo-Arabian' regions, and much of the arctic zone of the Palearctic (Fig. 5a), although
164 note that grid cells of the two latter areas are important because of their lack of human land use
165 (Fig. 5c). The cells where human land use 2,000 years ago was more important are located
166 mainly throughout Central America and the Andes mountain range, Europe except for the
167 northernmost part, most of the Oriental region, the boundary between the Palearctic and the
168 Oriental regions, as well as Madagascar and much of the arctic zone of the Palearctic due to lack
169 of human land use (Fig. 5b). These areas mostly coincided for the broader natural bioregions
170 (Fig. 5e, f).
171
172 For the medium current bioregions, the cells where the human land use at present was most
173 relevant for the classification of bioregions included those highlighted for the broader
174 bioregions plus most part of Europe, southern India, eastern Australia, a large area of the
175 Nearctic, and the Sahara Desert (Fig. 6a, c). When focusing on the human land use 2,000 years
176 ago, the areas affected by human land use were similar, with some exceptions (e.g., Sahara
177 Desert), although their importance scores were much higher (Fig. 6, b, d). The location of cells
178 affected by human land use for the medium current bioregions quite coincided with that for
179 natural bioregions (Fig. 6e, f).
180
181 **Discussion**
182 Our study shows that human activity during the last few millennia has left its mark on the
183 spatial organization of the Earth´s biodiversity. First, we show that biogeographical
184 configurations based on current and natural mammal distributions differ notably, particularly
185 in the Northern Hemisphere. These differences were not solely the result of including extinct

186 megafauna - with huge distributions ranges – in the natural biogeographic regions, as
187 biogeographical patterns for the 536 extant terrestrial mammals whose ranges of distribution
188 have been affected by anthropogenic impacts confirmed our findings. Second, we show that
189 human land use is an important predictor of current bioregions but does not largely contribute
190 to predict natural bioregions. For example, for the broader regions or realms, human land use
191 2,000 years ago was the third most important predictor behind plate tectonics and temperature
192 seasonality and ahead of orography, but has minor importance on natural regions. Past
193 anthropogenic actions have seemingly left a perceptible biogeographical imprint on the
194 assembly of the global realms we recognize today. Moreover, for the medium-sized subregions,
195 human land use at present was the second most important predictor suggesting that recent
196 human activity is already causing changes in the biogeographical assemblages at the
197 subregional scale, i.e., within and among continents.
198
199 The mammalian distributions that we know today are not only a reflection of the most
200 recent human actions but also those exerted during the last few millennia[27,28]. Recent studies
201 had already shown that human actions in the present can be evident at biogeographical
202 scales[18,19], but here we also show evidence of an effect of past human land use. What did
203 happen 2,000 years ago that left such as noticeable footprint in the Earth's biogeographical
204 regions? This period coincides with the settlement of densely populated areas after the
205 Neolithic agricultural revolution[29]. Long-term impacts to enhance agricultural productivity,
206 including forest clearing, increased fire frequencies, megafauna extinctions, species invasions,
207 and soil erosion, were already apparent in some regions at this time[20,23,24]. In fact,
208 reconstructions suggest that >20% of Earth's temperate woodlands had been already impacted

209 by humans by the late-Holocene[23]. Indeed, our results on local importance show the match
210 between some of the biogeographical boundaries and regions of the New World and the
211 distribution – and presumed impact – of ancient Mesoamerican and South American pre-
212 European civilizations[30,31]. In the Old World, we note the prominence of past human land use
213 covering most of the Oriental realm territory, likely associated to the rise of rice cultivation in
214 eastern, southern and southeast Asia over the Holocene[32]. Indeed, cumulative archaeological
215 data show that the scale of agriculture and land use in these regions was significant; expansion
216 of land area used for livestock and rice (*Oryza sativa*) paddy agriculture in southern Asia was
217 sufficient to increase atmospheric methane emissions between 4000 and 1000 years ago[33].
218 Moreover, the atmospheric CO_2 decline registered between A.D. 1570 and 1620 has been
219 attributed to the uptake by vegetation following the agriculture abandonment caused by the
220 population crash after the arrival of Europeans to America[34]. Human land use 2,000 years ago
221 can also partly explain differences between the southern Palearctic boundary and the
222 Afrotropic below the Sahara. Both can be related to the transformation of European and Near
223 Eastern landscapes during the Roman period[35], and the expansion of sorghum (*Sorghum*
224 *bicolor*) cultivation 3000 to 2000 years ago in the sub-Saharan Africa[34].
225
226 While we found differences in importance of human land use 2,000 years ago between current
227 and natural medium-sized bioregions, we were surprised that past human land use did partly
228 predict natural bioregions, which were based on estimated distributions of mammals in the
229 absence of human impact. It is possible that the natural distribution maps did not fully exclude
230 human impacts as for 3,424 of the extant mammals analyzed, natural and current distributions
231 were identical. In addition, these results could also reflect an overlap between human land use

232 and biogeography. If the distribution of humans 2,000 years ago responded to the same drivers
233 as the distributions of terrestrial mammals, human land use could appear as a predictor even if
234 it had no causal effect. In fact, the importance of this variable appears to be a consequence of
235 how human land use was distributed globally 2,000 years ago, or more accurately where
236 human activity was largely absent. While current human land use is distributed over wide areas
237 of the planet, that of 2,000 years ago was centralized in specific areas coinciding with the
238 distribution of ancient civilizations. This means that on a subregional scale, many bioregions
239 can be predicted by the absence of human land use. Indeed, an additional analysis that excluded
240 subregions with no reported human land use 2,000 years ago showed very little importance of
241 past human land use (Fig. S8).
242
243 In summary, our results show the value of considering impacts of past human actions to
244 understand the current organization of biodiversity globally. Previous studies have
245 documented lasting effects of human land use changes during the last millennia on current
246 biodiversity patterns[36-38], but this is the first time the signal has been recognized on the
247 taxonomic differentiation of the largest realms. This would mean that some of the
248 biogeographical boundaries proposed by Wallace 140 years ago[1] were already showing the
249 influenced of anthropogenic actions. The human transformation of ecosystems that occurred
250 in large terrestrial extensions coinciding with the distribution of ancient civilizations, makes
251 past human land use important along with orographic processes and temperature
252 seasonality to discern taxonomic differences among the large biogeographical zones of the
253 world. It is certainly difficult to disentangle direct causal effects from correlative
254 relationships in non-experimental large-scale studies such as this. We know the same

255 geographic and environmental characteristics that influenced the distribution across space
256 and time of mammals likely also influenced the distribution of our ancestors[39]. However, we
257 considered several of those predictors and still found a global imprint of past human actions
258 on the patterns of distribution of global diversity. Our results force us to reflect on how the
259 much more widespread and severe changes that have occurred since the beginning of the
260 industrial revolution will affect the organization of biodiversity in the future. We show here
261 the effects of recent human actions are already detectable at the subregional scale, and
262 possible at the broader biogeographical scale. Although it may seem like an unlikely or
263 distant future, the possibility that human activities become the main driver, above historical
264 geomorphological or climatic processes, that shapes Earth´s biodiversity cannot be ignored.
265

266 **Methods**

267 **Generating presence-absence matrices using range distributions.** Current and natural
268 distribution ranges of non-volant terrestrial mammals were obtained from PHYLACINE 1.2
269 database[25]. This database contains range maps for all 5,831 known mammal species that lived
270 since the last interglacial (~130,000 years ago until present). Briefly, ranges for current extant
271 species contained in PHYLACINE are IUCN distribution maps[40] projected to Behrmann
272 cylindrical equal area rasters with a cell size of 96.5 km by 96.5 km at 30° North and 30° South.
273 Only current and natural ranges are considered (i.e. introduced, extinct or probably extinct
274 ranges have been excluded). Moreover, PHYLACINE also contains present natural ranges, which
275 represent estimates of where species would live if they had never experience anthropogenic
276 pressures, and – in the case of extinct species – had not gone extinct. We generated two
277 presence-absences matrices, one for the current and one for the natural ranges, in which every

278 row represents a grid cell and every column a species. From theses matrices we then excluded
279 (1) *Homo* species; (2) bats and marine mammals; (3) cells containing less than 50% of land area
280 to approximate equal-size samples, as a result small islands and coastal cells were not included
281 in the analyses, and, (4) cells containing fewer than five species to reduce potential distortions
282 caused by having few taxa[9,10]. These exclusion criteria rendered a total of 14,087 cells
283 representing 3,960 terrestrial mammal species for the current matrix, and 14,151 cells
284 representing 4,306 for the natural matrix. With the aim of conducting complementary analyses
285 to support our results we also generated two additional matrices using distributions of those
286 extant species that have undergone some modification in their range of distribution due to
287 anthropogenic pressure, i.e., they have an estimated distribution range in PHYLACINE different
288 from that of the IUCN, representing 536 species of non-volant terrestrial mammals.
289
290 **Building biogeographical regionalizations.** To each of the presence-absence matrices we
291 applied a machine-learning algorithm referred to as affinity propagation[26] to build
292 biogeographical regionalizations at different biogeographical resolution. Affinity propagation
293 (AP hereafter) is a powerful clustering algorithm extensively used in bioinformatics and
294 astrophysics and is making inroads in ecology[41] and biogeography[10]. One of its main advantages
295 is that it can compress massive data sets very efficiently (i.e. with lower error). The algorithm
296 detects special data points called exemplars, and by a message-passing procedure it iteratively
297 connects every data point to the exemplar that best represents it until an optimal set of
298 exemplars and clusters emerges. Contrary to algorithms in which exemplars are found by
299 randomly choosing an initial subset of data points, AP takes as input measures of 'similarities'
300 between pairs of data points (grid cells here) and simultaneously considers all the points as

301 potential exemplars. The optimal set of exemplars is the one for which the sum of similarities of
302 each point to its exemplar is maximized. Hence, detecting exemplars goes beyond simple
303 clustering because the exemplars themselves store compressed information[42].
304
305 We used a protocol based on a successive application of AP[10] to obtain a biogeographical
306 upscaling from the smallest possible bioregions (i.e. the highest biogeographical resolution) to
307 the largest ones. We first used the presence-absence matrix to calculate pairwise similarities
308 between pairs of cells. As with clustering algorithms, there are many similarities indices or
309 taxonomic/phylogenetic turnover metrics to choose from, and none can be said to be perfect
310 from a biogeographical perspective. Here we selected Hellinger distance[43], which is calculated
311 by first modifying the species-presence data and then computing the Euclidean distance among
312 pairs of cells based on the modified data[44]. The Hellinger distance is used to avoid the 'double-
313 zero' problem, i.e. when two sites or grid cells that have no species in common are assigned the
314 same distance as two sites that share the same species; and the 'species-abundance paradox',
315 which frequently occurs when two sites share only a small fraction of all the species in the same
316 regional pool. This is expected to be a particular problem at the margins of biogeographical
317 regions where sites may be quite different from one another rather than in the centre of a
318 region where sites are likely to be very similar in their species assemblages[45,46]. We performed
319 an initial AP analysis involving all grid cells of the similarity matrix. This first AP run generates
320 the optimal solution of the highest resolution bioregions, while also identifying its exemplars.
321 We obtained 1,128 clusters and exemplars for the current distributions and 1,053 for the
322 natural distributions. Then, using the exemplars as the new units of analysis we conducted
323 again an AP, i.e., we calculated a new similarity matrix and re-run a new AP. This process was

repeated until a small and coherent number of large clusters emerged. Finally, to obtain maps of each clustering result, we classified each grid cell (row) of every presence-absence matrix according to the cluster to which they were assigned in its corresponding AP analysis. AP analyses were performed using the '*APCluster*' package[47] in R Version 3.6.1[48].

Assessing differences in biogeographical configurations. The degree of similarity between clusters (i.e. regions) of current and natural bioregions was estimated using the Jaccard index. For that, we used the *cluster_similarity* function of the '*clusteval*' package[49] in R Version 3.6.1[48]. This function computes the similarity between two clusterings of the same data. Jaccard index ranges between 0 (no similarity) to 1 (perfect match).

Predictors. To each of the cells of the presence-absence matrices we assigned a mean value of several predictors. We considered predictors previously tested as determinants of biogeographical assemblages and boundaries[11] related to climatic heterogeneity, orographic barriers, tectonic movements, and instability of past climate, plus predictors related to anthropogenic land use over the Holocene. In particular, we tested four climatic variables: annual total precipitation, mean annual temperature, seasonality in temperature and seasonality in precipitation. All climatic variables were extracted from the WorldClim dataset[50] up-scaled at a 96.5 km resolution. These variables represent both average conditions and variability within years and have been shown to be determinants of vertebrate distributions[51]. Climatic conditions have strongly shifted over the Quaternary period, and have been shown to play a role in the present-day species distributions, endemism, and assemblages[52-55]. So, to test for the potential effect of past climate change or stability, we calculated for each cell the

347 average velocity of climate change since the Last Glacial Maximum (LGM; ~22,000 years ago)
348 and since the Mid Holocene (MH; ~ 6000 years ago)[54]. To carry out these calculations, we used
349 the mean annual temperature and annual total precipitation for the MH and the LGM calculated
350 by means of the Model for Interdisciplinary Research on Climate (MIROC-ESM)[56] and available
351 in the WorldClim dataset[50]. Mountain ranges represent major barriers to dispersal for most
352 mammals, whereas plate tectonics are responsible of the long-term isolation of the biotas[11,12].
353 We use the GTOPO30 to calculate the mean elevation per grid cell. Plate tectonics were
354 obtained from Bird's dataset[57]. Each grid cell was assigned the tectonic plate to which it
355 belongs. When a cell was represented by more than one tectonic plate, it was assigned the one
356 that occupied a greater percentage of the cell. Finally, past anthropogenic land use was
357 obtained from Ellis et al.' dataset[23]. We chose the more realistic KK_{10} model, which assumes that
358 humans use land more intensively when population density is high and land scarce[58]. In
359 counterpoint, the most popular HYDE model omits land-use intensification and predicts that
360 except for the developed regions of Europe, human use of land was insignificant in every biome
361 and region before A.D. 1750. We calculated human pressures at four different time spans:
362 8,000, 5,000, and 2,000 years ago, as well as at the present time, although in the final analyses
363 only human land use at present and 2,000 years ago were used (see below).
364
365 **Evaluating potential drivers of the biogeographical configurations.** We used random
366 forest[59,60] models of 5,000 classification trees to assess the factors that may predict the
367 classification of cells in bioregions at different biogeographical resolutions and to estimate the
368 relative importance of the predictors. Random forest is a machine learning method based on a
369 combination of a large set of decision trees. Each tree is trained by selecting a random set of

370 variables and a random sample from the training dataset (i.e., the calibration data set). The
371 number of variables randomly sampled as candidates at each split, *mtry*, is a tuning parameter
372 of random forest models. The default value for random forest classification is the squared root
373 of the number of variables (i.e., the squared root of 9 resulting in a *mtry* of 3). We also used the
374 *tuneRF* function[61] to look for the optimal *mtry* parameter. We ran the code 5 times for each
375 biogeographic scale and the results varied between 3 and 8, although the trend was always
376 higher towards 3, which consequently was the *mtry* we finally chose. To calculate the accuracy
377 of the models we used the out-of-bag (OOB), an estimate of the misclassification rate that
378 represents and unbiased error rate of the model that is calculated by counting how many cases
379 in the training set are misclassified and dividing the number by the total number of
380 observations.
381
382 Random forests are able to disentangle interacting effects and identify nonlinear and scale-
383 dependent relationships that often occur at the scale of the analysis performed here among
384 multiple correlated predictors[60]. Although random forests are generally assumed to not be
385 affected by highly correlated predictor variables, we eliminated some predictors showing a
386 moderate to high correlation ($r > 0.50$, Table S1) as some evidence from genomic studies
387 suggests that variable importance measures may show a bias towards correlated predictor
388 variables[62]. So, our final models included seasonality in temperature and precipitation as
389 representatives of climate heterogeneity; velocity of climate change since the LGM and the MH
390 to the present as representatives of past climate change; plate tectonics; mean elevation
391 representing mountains; and human land use at the present and 2,000 years ago. We decided to
392 not include human land use 8,000 years ago in the models due to its low representation at

global level (Fig. S9). We ran separate models for human land use at present and human land use 2,000 years ago. Although globally the correlation between both variables is relatively low (r = 0.29), we made the decision to perform independent models because for certain bioregions both human land use at present and 2,000 years ago can be highly correlated. We also ran additional models with the 536 extant species that have undergone some modification in their distribution range due to anthropogenic pressures. Finally, the algorithm used in the process of regionalization only uses species presence-absences, which means that humans can only have a role if present and if causing local extinctions, thereby changing the amount of shared species between grid cells. At the largest biogeographical scale two bioregions, Madagascar and Australia, are characterised by an extremely low historical human land use that could influence results. Therefore, we also ran the models excluding Australia and Madagascar. Random forest models were computed using the package '*RandomForest*'[61] in R Version 3.6.1[48].

Calculating global and local predictor importance. Random forest also calculates estimates of global variable importance for classification, which are very useful to interpret the relevance of variables for the dataset under study. We measured variable importance using the mean decrease in accuracy: variables with a large mean decrease in accuracy are more important for classification of the data. This measure is obtained by permuting randomly each variable and assessing the decrease in classification accuracy of the model[61]. Thus, the more the accuracy of the random forest decreases due to the permutation (or exclusion) of a single variable, the more important that variable is considered.

415 Moreover, the local variable importance defines the importance of a variable in the
416 classification of a single sample (grid cell here) and therefore shows a direct link between
417 variables and samples[63]. The local importance score is derived from all trees for which the
418 sample was not used to train the tree (i.e. its value is OBB). The percentage of correct votes for
419 the correct class in the permuted OBB data is subtracted from the percentage of votes for the
420 correct class in the original OOB data to assign a local importance score for the variable for
421 which the values were permuted. The score reflects the impact on correct classification of a
422 given sample: negative, 0 (the variable is neutral) and positive. Given that local importance
423 values are noisier than global importance ones we run the same classification 5 times (5 per
424 biogeographical scale) and averaged the local importance scores to obtain a robust estimation
425 of local importance values[63]. Since one value per cell was obtained, these were used to create
426 maps of the local importance of present and past human land use. However, the values of local
427 importance do not distinguish between areas where the human land use was high and areas
428 where it was negligible, i.e., both are equally valuable for the classification analysis. So, to help
429 in the interpretation we also represent in maps the local importance value for those grid cells
430 where the percentage of land used was greater than 10%.
431

432 **References**

433 1. Wallace, A. R. *The geographical distribution of animals* (Harper & Brothers, New York, 1876).
434 2. von Humboldt, A. *Essai sur la geographie des plantes; accompagné d'un tableau physique des*
435 *régions équinoxales, accompagné d'un tableau physique des régions équinoctiales* (Schoel &
436 Co., Paris, 1806).

437 3. Ricklefs, R. E. A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.* **7,** 1-
438 15 (2004).
439 4. Wiens, J. J. The niche, biogeography and species interactions. *Philos. Trans. R. Soc. Lond. B Biol.*
440 *Sci.* **366,** 2336-2350 (2011).
441 5. Daru, B. H., Elliot, T. L., Park, D. S. & Davies, T. J. Understanding the processes underpinning
442 patterns of phylogenetic regionalization. *Trends Ecol. Evolut.* **32,** 845-860 (2017).
443 6. Botkin, D. B. et al. Forecasting the effects of global warming on biodiversity. *Bioscience* **57,**
444 227-236 (2007).
445 7. Holt, B. et al. An update of Wallace´s zoogeographic regions of the world. *Science* **339,** 74–78
446 (2013).
447 8. Proches, S. & Ramdhani, S. The world´s zoogeographical regions confirmed by cross-taxon
448 analyses. *Bioscience* **62,** 260–270 (2012).
449 9. Kreft, H. & Jetz. W. A framework for delineating biogeographical regions based on species
450 distributions. *J. Biogeogr.* **37,** 2029-2053 (2010).
451 10. Rueda, M., Rodriguez, M. A. & Hawkins, B. A. Identifying global zoogeographical regions:
452 lessons from Wallace. *J. Biogeogr.* **40,** 2215–2225 (2013).
453 11. Ficetola, G. F., Mazel, F. & Thuiller, W. Global determinants of zoogeographical boundaries.
454 *Nature Ecology & Evolution* **1,** 0089 (2017).
455 12. Lomolino, M. V., Riddle, B. R., Whittaker, R. J. & Brown, J. H. *Biogeography* (Sinauer
456 Associates, 2010).
457 13. Riddle, B. R. & Hafner, D. J. Integrating pattern with process at biogeographic boundaries:
458 the legacy of Wallace. *Ecography* **33,** 321–325 (2010).

14. Barnosky et al. Has the Earth´s sixth extinction already arrived? *Nature* **471**, 51-57 (2011).

15. Ceballos, G., Ehrlich, P. R. & Dirzo, R. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl. Acad. Sci. USA* **114**, E6089-E6096 (2017).

16. Baiser, B., Olden, J. D., Record, S., Lockwood, J. L. & McKinney, M. L. Pattern and process of biotic homogenization in the new Pangea. *Proc. Royal Soc. B* **279,** 4772-4777 (2012).

17. Olden, J. D. Biotic homogenization: a new research agenda for conservation biogeography. *J. Biogeogr*. **33**, 20-27-2039 (2006).

18. Capinha, C. et al. The dispersal of alien species redefines biogeography in the Anthropocene. *Science* **348,** 1248-1251 (2015).

19. Bernardo-Madrid, R. et al. Human activity is altering the world´s zoogeographical regions. *Eco Lett,* https:/doi.org/10.1111/ele.13321 (2019).

20. Lyons, S. K. et al. Holocene shifts in the assembly of plant and animal communities implicate human impacts. *Nature* **529,** 80-85 (2016).

21. Sandom, C., Faurby, S., Sandel, B. & Svenning, J-C. Global late Quaternary megafauna extinctions linked to humans, not climate change. *Proc. Royal Soc. B* **281,** 20133254 (2014).

22. Smith, F., Elliott Smith, R. E., Lyons, S. K. & Payne, J. Body size downgrading of mammals over the late Quaternary. *Science* **360**, 310-313 (2018).

23. Ellis, E. C. et al. Used planet: a global history. *Proc. Natl. Acad. Sci. USA* **110,** 7978-7985 (2013).

24. Boivin, N. L. et al. Ecological consequences of human niche construction: Examining long-tern anthropogenic shaping of global species distributions. *Proc. Natl. Acad. Sci. USA* **113**, 6388-6296 (2016).

25. Faurby, S., Davis, M., Pedersen, R. O., Schowannek, S. D., Antonelli, A. & Svenning, J-C. PHYLACINE 1.2: The Phylogenetic atlas of mammal macroecology. *Ecology* **99**, 2626 (2018).

26. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315,** 972-976 (2007).

27. Faurby, S. & Svenning, J-C. Historic and prehistoric human-driven extinctions have reshaped global mammal diversity patterns. *Diversity Distrib.* **21**, 1155-1166 (2015).

28. Polaina, E., González-Suárez, M. & Revilla, E. The legacy of past human land use in current patterns of mammal distribution. *Ecography* **42**, 1623-1635 (2020).

29. Bocquet-Appel, J-P. When the world´s population took off: the springboard of the Neolithic demographic transition. *Science* **333**, 560-561 (2011).

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

31. Dunning, N. P., Beach, T. P. & Luzzadder-Beach, S. Kax and kol: Collapse and resilience in lowland Maya civilization. *Proc Natl Acad Sci USA* **109**, 3652-3657 (2013).

32. Silva, F. et al. Modelling the geographical origin of rice cultivation in Asia using the rice archaeological database. *PLoS ONE* **10**, e0137024 (2015).

33. Fuller, D. Q. et al. The contribution of rice agriculture and livestock pastoralism to prehistoric methane levels. An archaeological assessment. *Holocene* **21**, 743-759 (2011).

34. Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **519**, 171-180 (2015).

35. Butzer, K. W. Environmental history in the Mediterranean world: cross-disciplinary investigation of cause-and-effect for degradation and soil erosion. *J. Archaeol Sci* **32**, 1773-1800 (2005).

504 36. Dambrine, E. et al. Present forest biodiversity patterns in France related to former Roman
505 agriculture. *Ecology* **88,** 1430-1439 (2007).
506 37. van der Sande, M. et al. A 7000-year history of changing plant trait composition in an
507 Amazonian landscape; the role of humans and climate. *Ecol. Lett.* **22**, 925-935 (2019).
508 38. Haberle, S. G. Prehistoric human impact on rainforest biodiversity in highland New Guinea.
509 *Phil. Trans. R. Soc. B* **362**, 219-228 (2007).
510 39. Lomolino, M. V. On teaching "…that grand subject, …". *Front. Biogeogr.* **10**, e37812 (2018).
511 40. IUCN 2016. IUCN red list of threatened species. Version 2016-3. http://www.iucnredlist.org
512 41. Cardille, J. A. & Lambois, M. From the redwood forest to the Gulf Stream waters: human
513 signature nearly ubiquitous in representative US landscapes. *Front. Ecol. Environ*. **8,** 130–134
514 (2009).
515 42. Mézard, M. Where are the exemplars? *Science* **315,** 949-951 (2007).
516 43. Rao, C. R. A review of canonical coordinates and an alternative to correspondences analysis
517 using Hellinger distance. *Qüestiió* **19,** 23–63 (1995).
518 44. Legendre, P. & Gallagher, E. D. Ecological meaningful transformations for ordination of
519 species data. *Oecologia* **129,** 271-280 (2001).
520 45. Legendre, P. & Legendre, L. *Numerical ecology*, (Elsevier, Amsterdam, 1998).
521 46. Gagné, S. & Proulx, R. Accurate delineation of biogeographical regions depends on the use of
522 an appropriate distance measure. *J. Biogeogr.* **36,** 561–567 (2009).
523 47. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: a R package for affinity
524 propagation clustering. *Bioinformatics* **27,** 2463–2464 (2011).
525 48. R Core Team. R: *A Language and Environment for Statistical Computing, 3.6.1* (R Foundation
526 for Statistical Computing, 2019).

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

50. Fick, S. E. & Hijmans, R. J. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climat.* **37,** 4302-4315 (2017).

51. Sexton, J. P., McIntyre, P. J., Angert, A. L. & Rice, K. J. Evolution and ecology of species range limits. *Ann. Rev. Ecol. Evol. Syst.* **40,** 415-436 (2009).

52. Graham, R. W. et al. Spatial response of mammals to Late Quaternary environmental fluctuations. *Science* **272,** 1601-1606 (1996).

53. Davis, M. B. & Shaw, R. G. Range shifts and adaptive responses to Quaternary climate change. *Science* **292,** 673–679 (2001).

54. Sandel, B. et al. The influence of late Quaternary climate-change velocity on species endemism. *Science* **334,** 660-664 (2011).

55. Araujo, M. B. et al. Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography* **31,** 8-15 (2008).

56. Watanabe, S. et al. MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.* **4,** 845–872 (2011).

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

58. Kaplan, J. O. et al. Holocene carbon emissions as a result of anthropogenic land cover change. *Holocene* **21,** 775–791 (2011).

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

60. Cutler, D. R. et al. Random forests for classification in ecology. *Ecology* **88,** 2783–2792 (2007).

550 61. Liaw, A. & Wiener, M. Classification and regression by RandomForest. *R News* **2,** 18–22
551 (2002).
552 62. Nicodemus, K. K. et al. The behaviour of random forest permutation-based variable
553 importance measures under predictor correlation. *BMC Bioinformatics* **11**, 110 (2010).
554 63. Touw, W. G. et al. Data mining in the Life Sciences with Random Forest: a walk in the park or
555 lost in the jungle. *Briefings in Bioinformatics* **14**, 315-326 (2012).
556

557 **Acknowledgements**

558 This work was supported by the project 2020/125 (Programa Operativo FEDER Andalucía 2014-
559 2020) to MR, and by CGL2017-83045-R AEI/FEDER EU (Agencia Estatal de Investigación from
560 the Ministry of Economy, Industry and Competitiveness, Spain co-financed with FEDER) to ER.
561

562 **Author contributions**

563 MR, MG-S and ER designed the study; MR analyzed the data and led the writing, with input and
564 edits from all authors.
565

566 **Competing interests**

567 The authors declare no competing interests.
568

569 **Material and Correspondence**

570 Marta Rueda (mruedag@us.es)

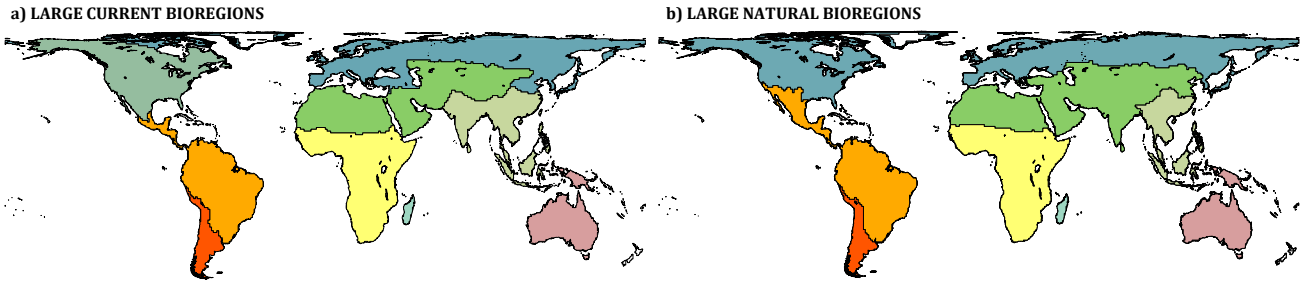

Fig. 1. Current and natural broader regions or realms. Only results for the broader regions are shown here: 9 current bioregions (a), and 8 natural bioregions (b). The greatest differences between both biogeographical configurations are more relevant in the Northern Hemisphere, where for the natural 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 Oriental region and is included instead within the 'Saharo-Arabian' region.

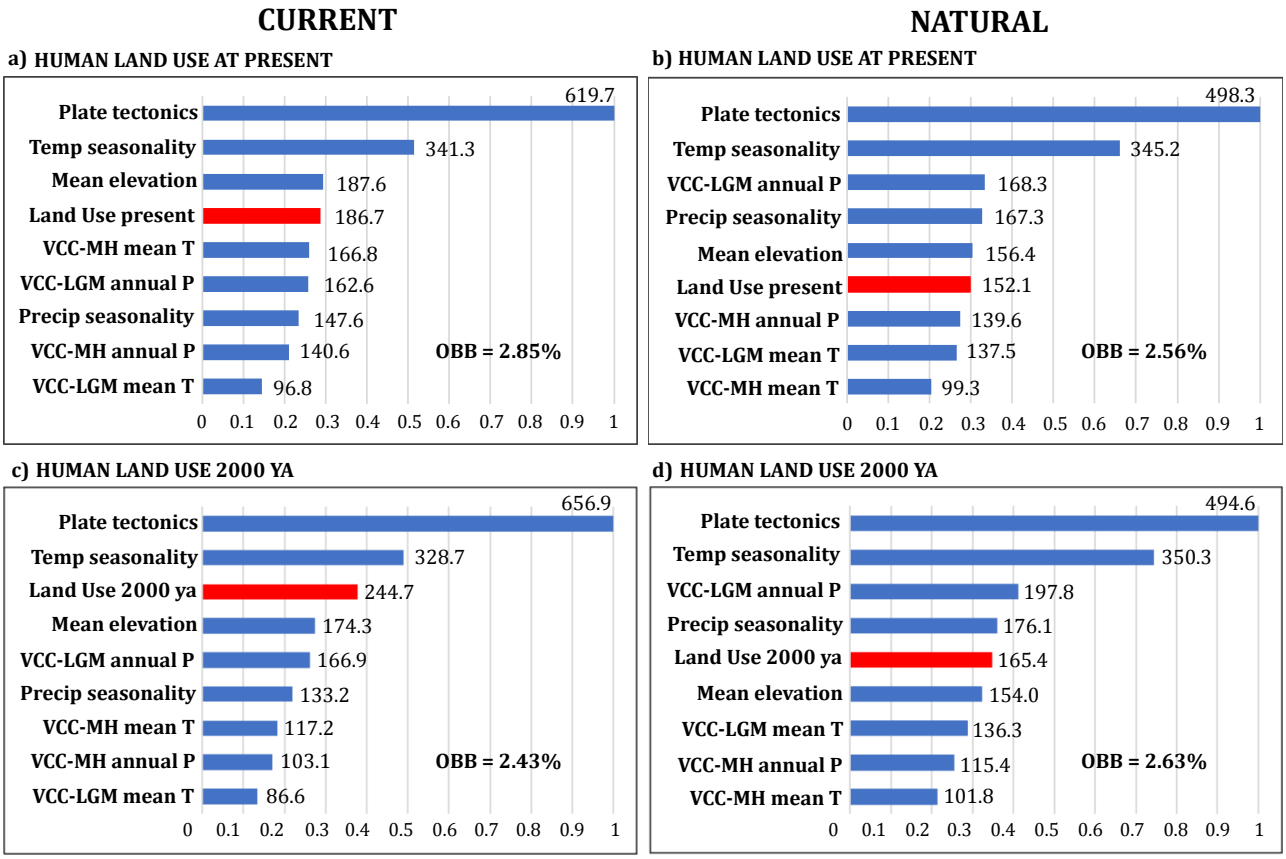

Fig. 2. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural broader bioregions. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2,000 years ago for (c) current and (d) natural bioregions. Importance was measured by the drop-in classification accuracy after predictor randomization in random forests of 5,000 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.

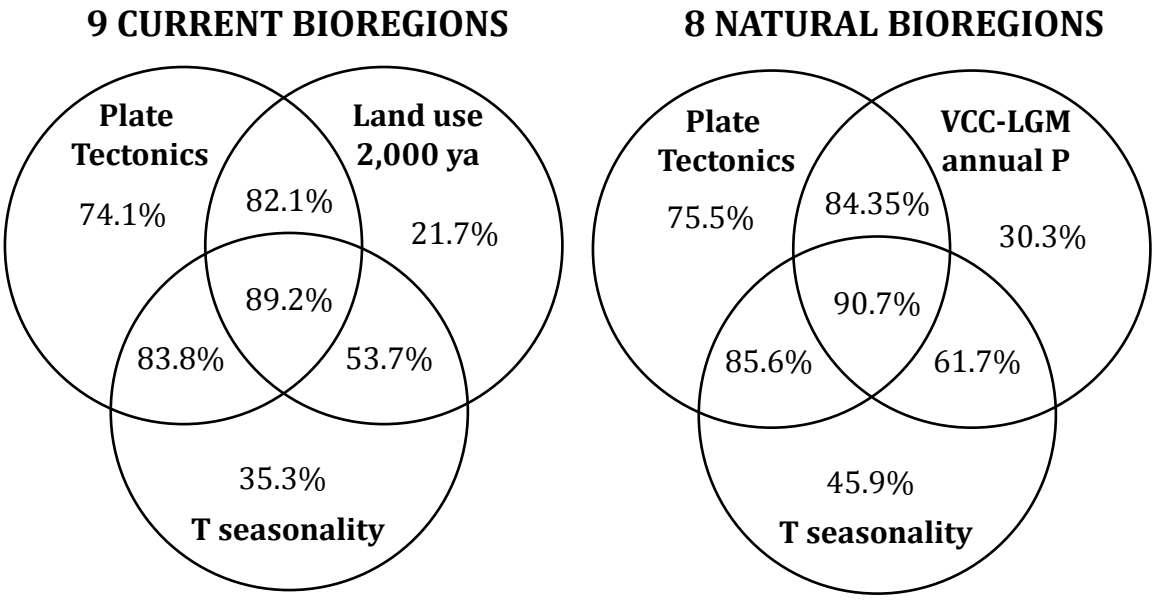

Fig. 3. Venn diagrams describing the accuracy in classification for models including the three most important drivers of the current and natural broader bioregions. Values indicate the percentage of cells correctly classified (1-OBB error) in their bioregion, and are obtained from running random forest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

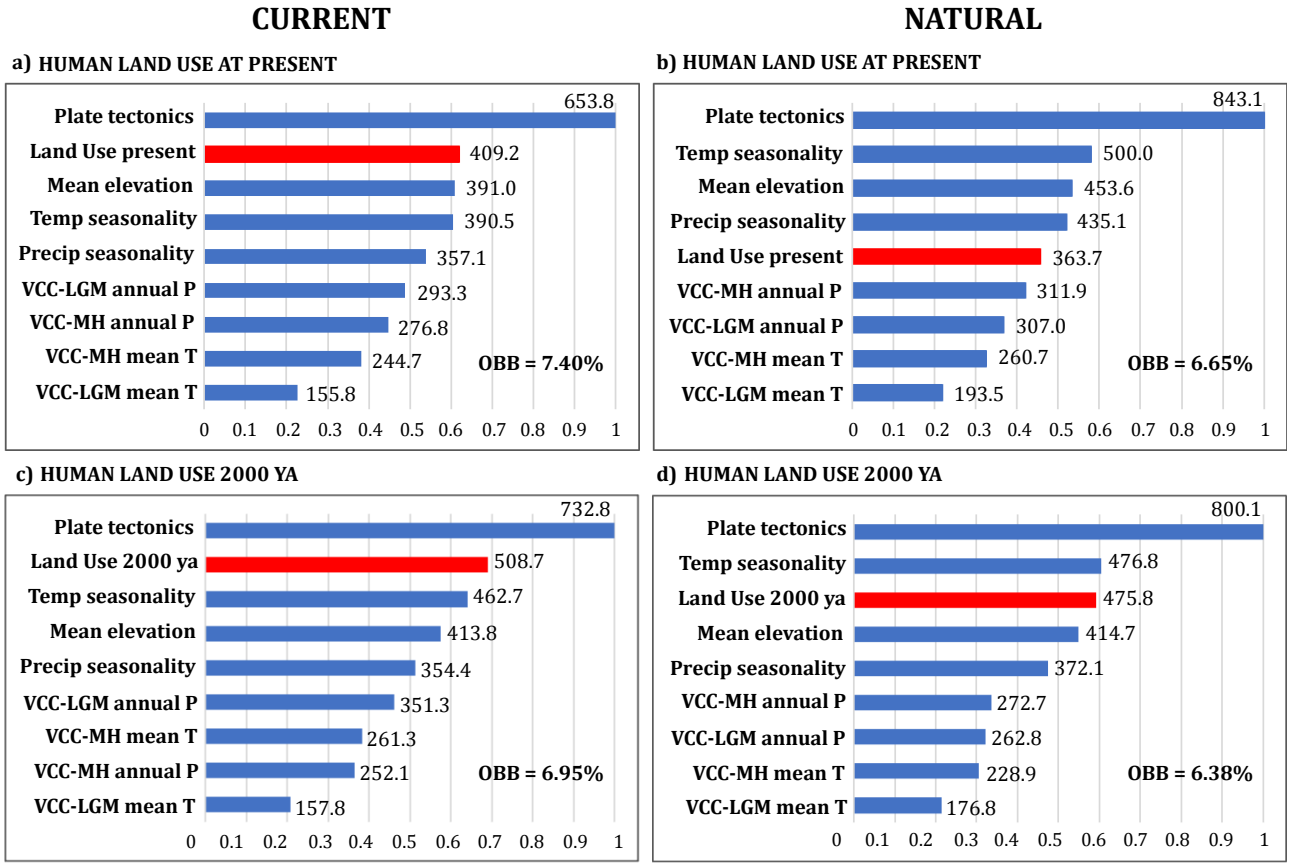

Fig. 4. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural medium-sized bioregions or subregions. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2,000 years ago for (c) current and (d) natural bioregions. Specifications as in Fig. 2.

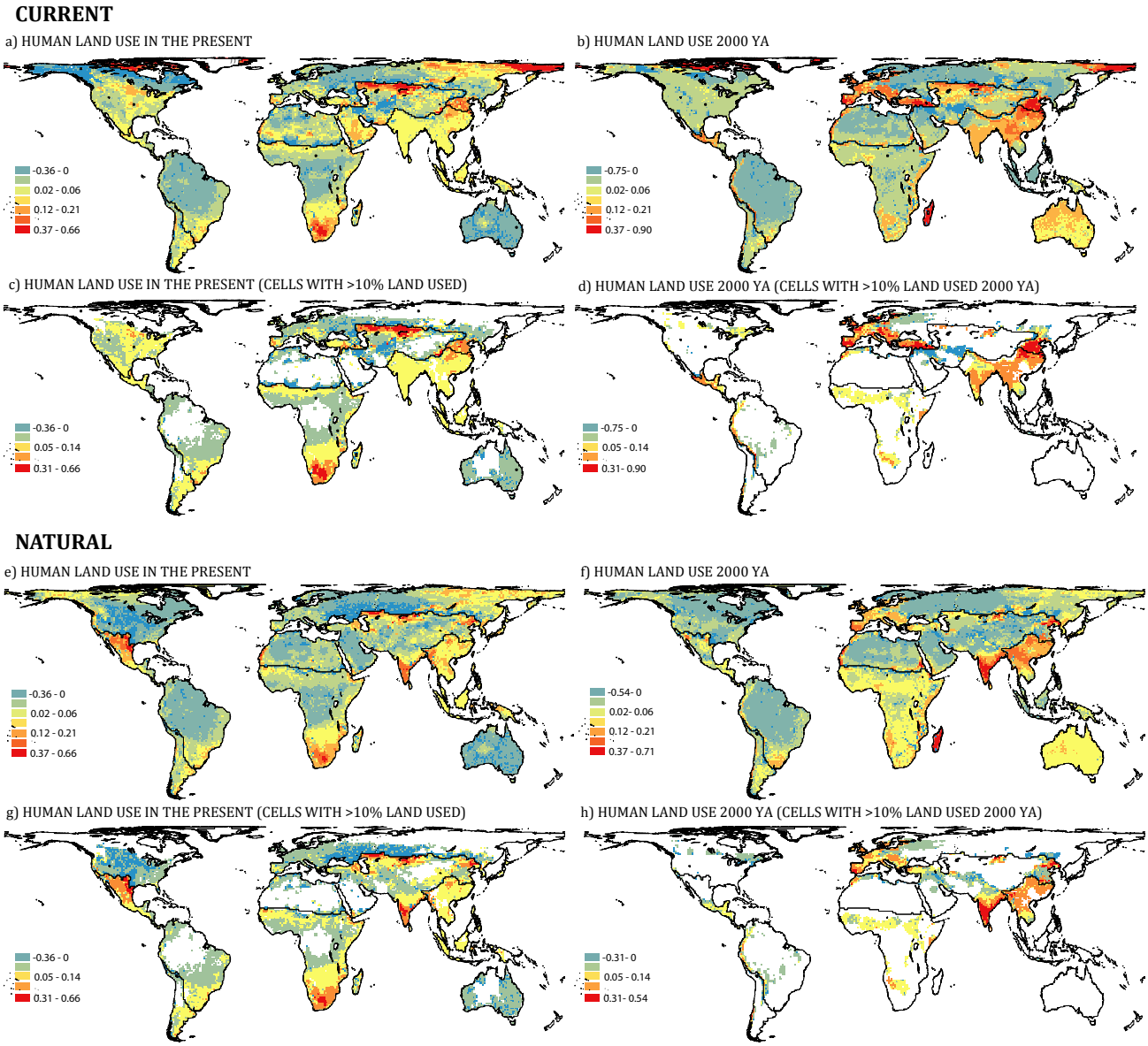

Fig. 5. Maps of the local importance of present (left panels) and past (right panels) human land use for current (the four maps above) and natural (the four maps below) broader bioregions. The local importance shows the importance of the variable for the classification of a single sample (grid cells here). 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, we have also mapped the importance values for those grid cells where the percentage of land use at present and 2,000 years ago was >10% (c, d, g, h). In these maps, white cells imply human land use values <10%.

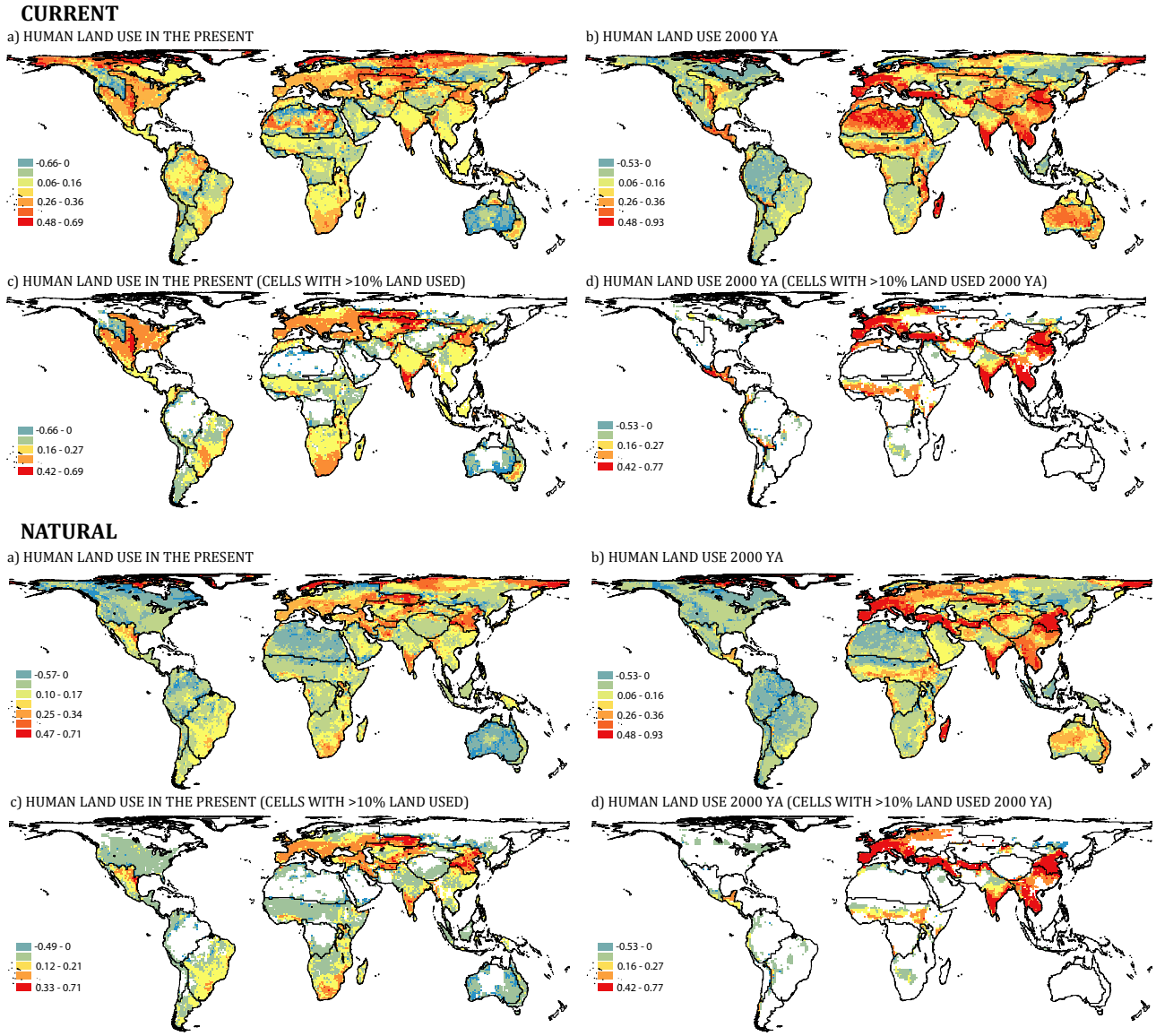

Fig. 6. Maps of the local importance of present (left panels) and past (right panels) human land use for current and natural medium-sized subregions. Specifications as in Fig. 5.

625 **Table 1**. **Values of Jaccard index between current and natural hierarchical bioregions**.

626 Jaccard index assesses the degree of similarity between clusters (bioregions) and ranges

627 between 0 (no similarity) and 1 (perfect match). The number of bioregions obtained during

628 each bioregionalization processes is also provided.

629

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
1,128 smaller bioregions	0.28	1,051 smaller bioregions

Supplementary Materials for

Global biogeographical regions reveal a signal of past human impacts

Marta Rueda*, Manuela González-Suárez, Eloy Revilla

**Corresponding author. Email: mruedag@us.es*

This PDF files includes:

Figure S1. Nested global bioregionalizations for terrestrial mammals.
Figure S2. Broader regions and subregions for the 536 extant mammals whose distributions rangers have shifted due to human impact.
Figure S3. Variable importance for the drivers of the current and natural broader bioregions without Australia and Madagascar.
Figure S4. Variable importance for the drivers of the current and natural broader bioregions built with the 536 extant mammals whose distributions rangers have shifted dur to human impact.
Figure S5. Variable importance for the drivers of the current and natural medium-sized bioregions built with the 536 extant mammals whose distributions rangers have shifted dur to human impact.
Figure S6. Variable importance for the drivers of the current and natural small bioregions.
Figure S7. Variable importance for the drivers of the current and natural smaller bioregions.
Figure S8. Variable importance for the drivers of the current and natural medium-sized subregions without subregions not affected by human land use 2,000 years ago.
Figure S9. Percentage of human land use at present, 2,000 and 8,000 years ago.
Table S1. Jaccard index for current and natural hierarchical bioregions build with the 536 extant mammals whose distributions rangers have shifted due to human impact.
Table S2. Pearson's correlations between continuous factors.

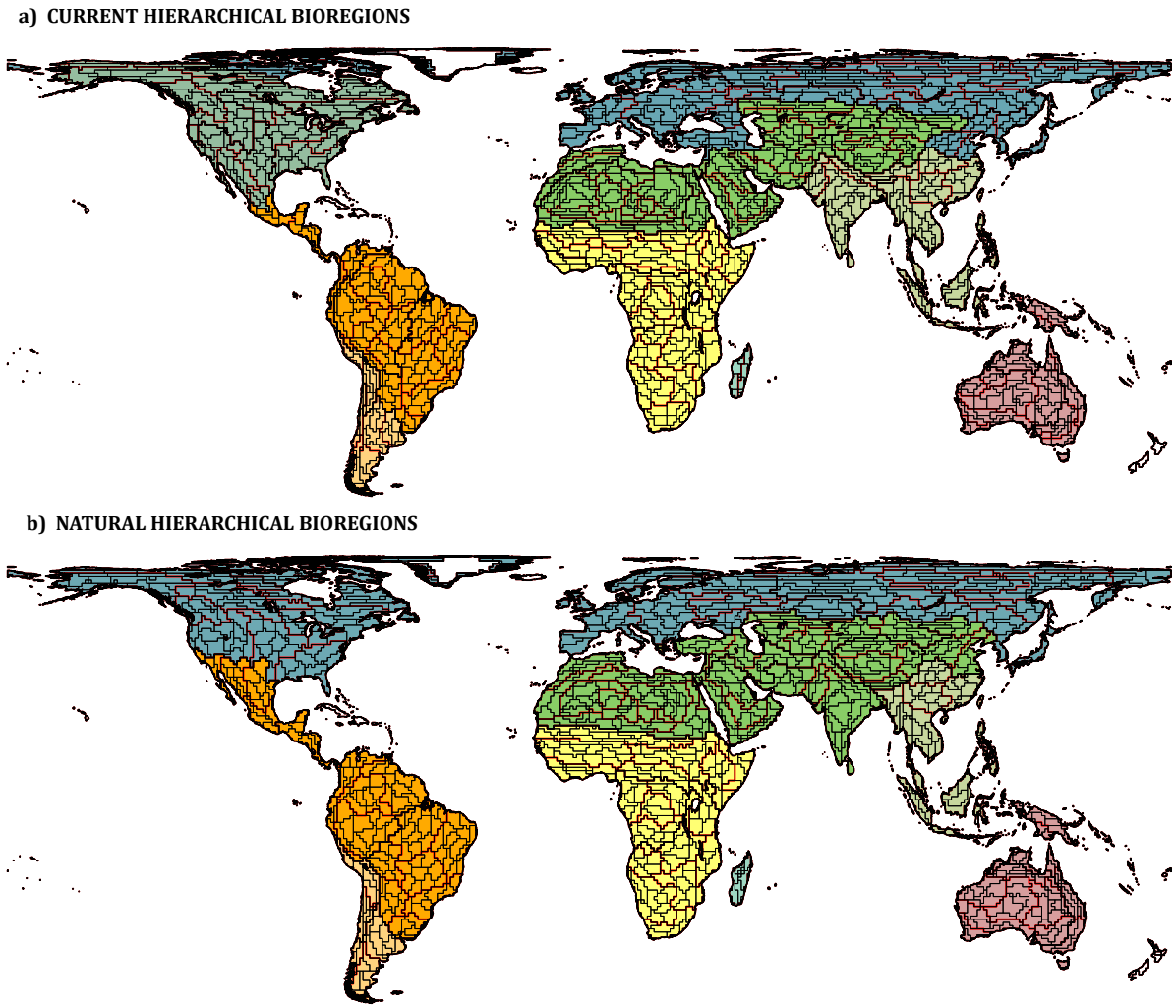

Fig. S1. Nested global bioregionalizations for non-volant terrestrial mammals. The hierarchical application of the affinity propagation algorithm resulted in a nested global bioregionalization containing for the current hierarchical bioregions (a): 9 (filled colours), 27 (red boundaries), 141 (black boundaries) and 1128 (grey boundaries) bioregions from the largest extant realms to the smallest obtained bioregions; and for the natural hierarchical bioregions (b): 8 (filled colours), 25 (red boundaries), 121 (black boundaries), and 1,051 (grey boundaries) bioregions.

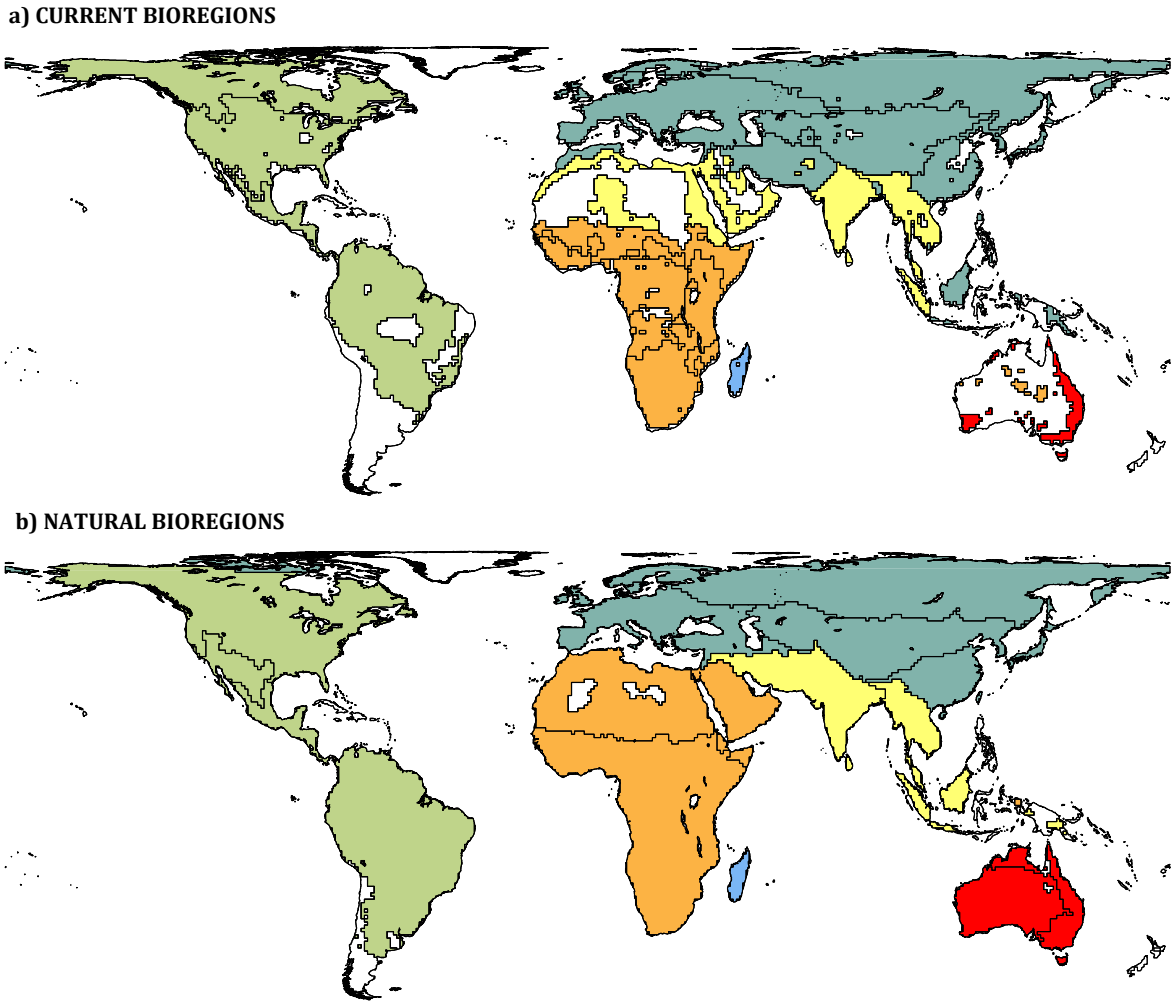

Fig. S2. Broader regions and medium-sized bioregions for the 536 extant non-volant terrestrial mammals whose distribution ranges have shifted due to human actions. The affinity propagation clustering algorithm resulted in 6 broader bioregions (filled colours), and 18 subregions (black boundaries) for (a) Current bioregions; and 6 broader bioregions (filled colours) and 17 subregions (black boundaries) for (b) Natural bioregions. Although not represented in the map, the clustering algorithm resulted in 93 and 1,333 smaller regions for the current, and 98 and 997 smaller regions for the natural bioregions.

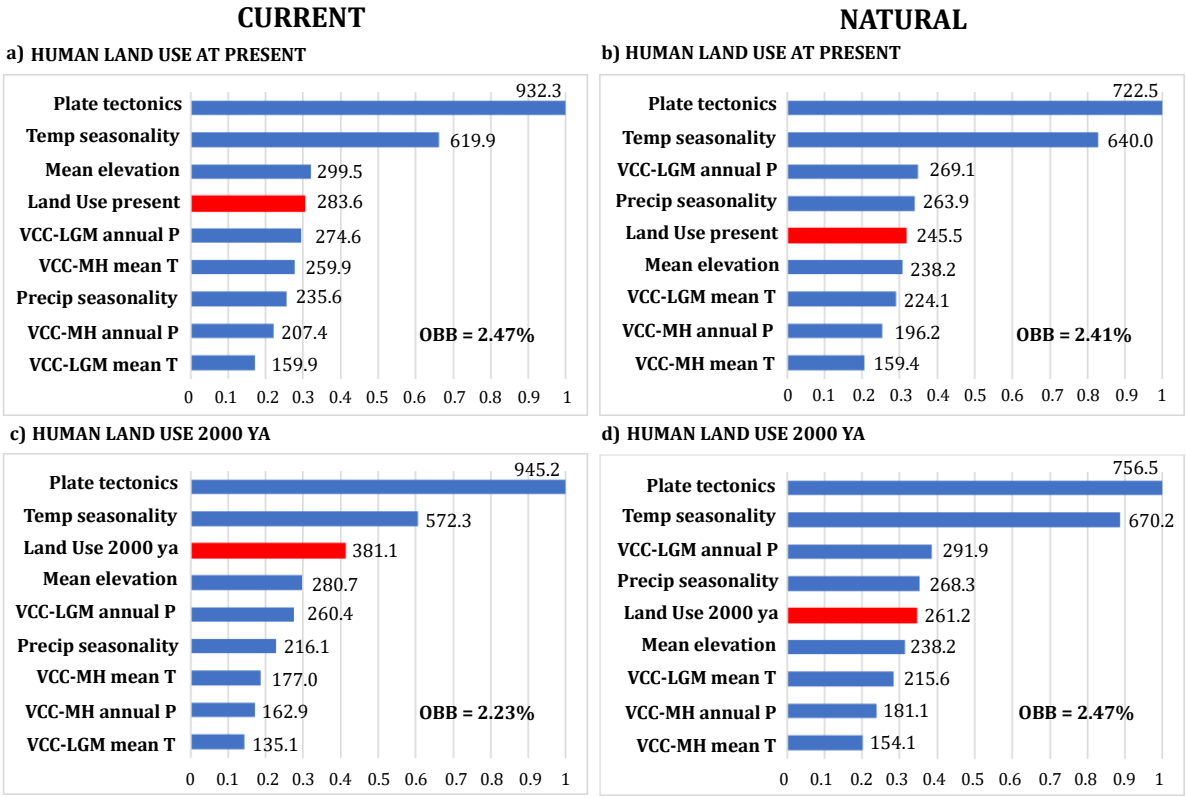

Fig. S3. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural broader bioregions without Australia and Madagascar. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2,000 years ago for (c) current and (d) natural bioregions. Importance was measured by the drop-in classification accuracy after predictor randomization in random forests of 5,000 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.

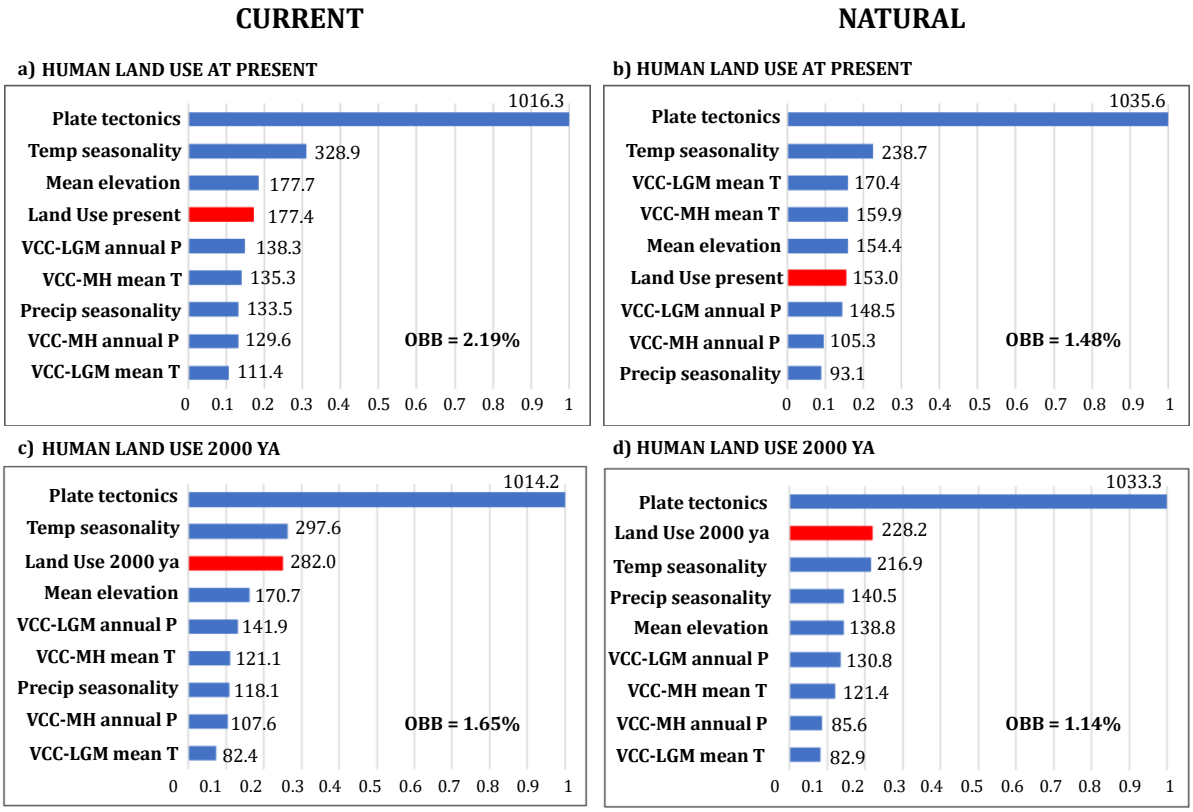

Fig. S4. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural broader bioregions for the 536 extant non-volant terrestrial mammals whose distribution ranges have shifted due to human actions. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2,000 years ago for (c) current and (d) natural bioregions. VCC = velocity of climate change; MH = Mid-Holocene; LGM = Last Glacial Maximum; annual P = annual precipitation; mean T = mean annual temperature; ya = years ago.

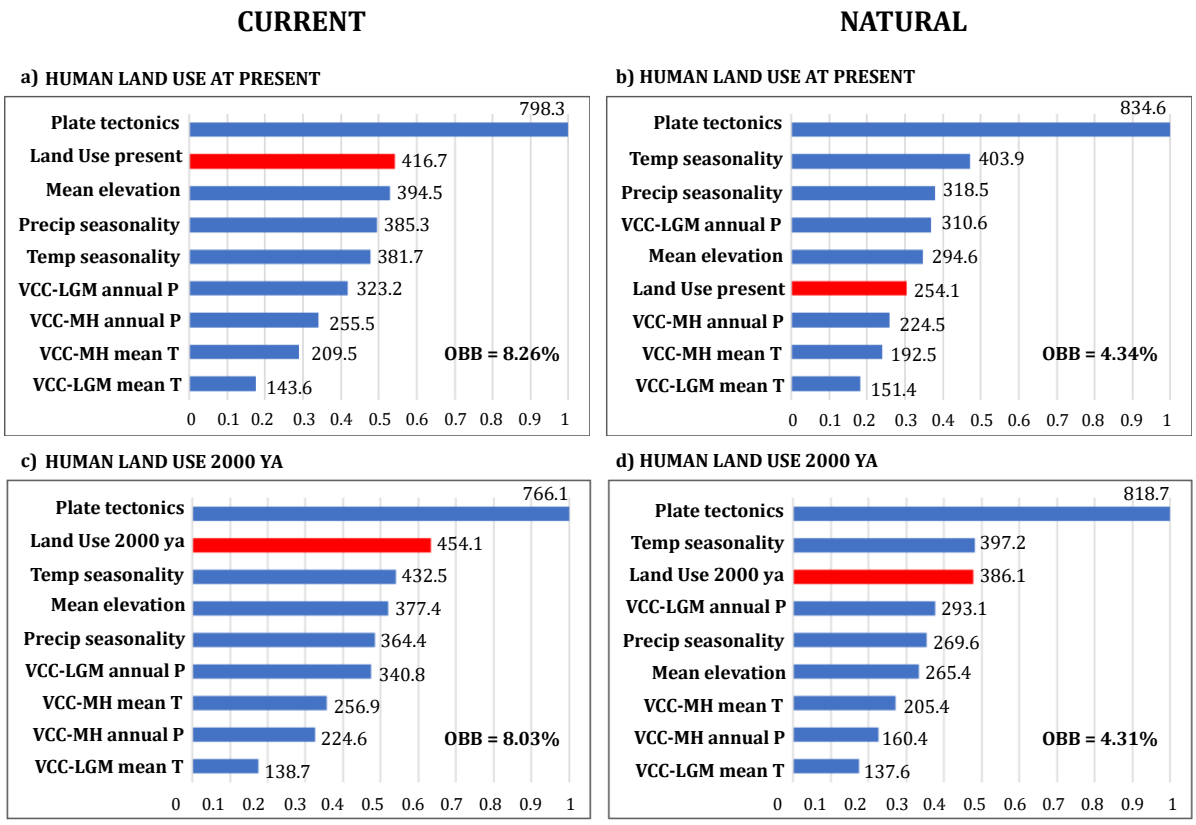

Fig. S5. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural medium-sized subregions for the 536 extant non-volant terrestrial mammals whose distribution ranges have shifted due to human actions. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2,000 years ago for (c) current and (d) natural bioregions. VCC = velocity of climate change; MH = Mid-Holocene; LGM = Last Glacial Maximum; annual P = annual precipitation; mean T = mean annual temperature; ya = years ago.

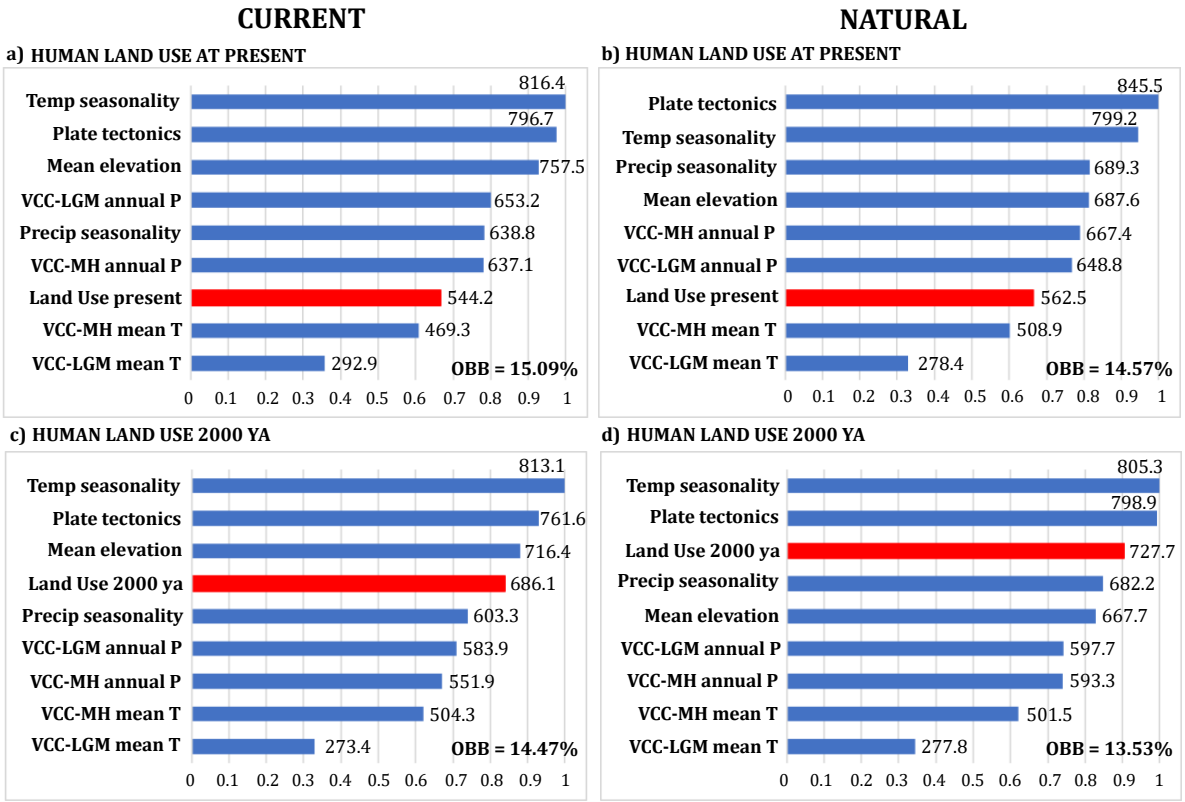

Fig. S6. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural small subregions. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2,000 years ago for (c) current and (d) natural bioregions. VCC = velocity of climate change; MH = Mid-Holocene; LGM = Last Glacial Maximum; annual P = annual precipitation; mean T = mean annual temperature; ya = years ago.

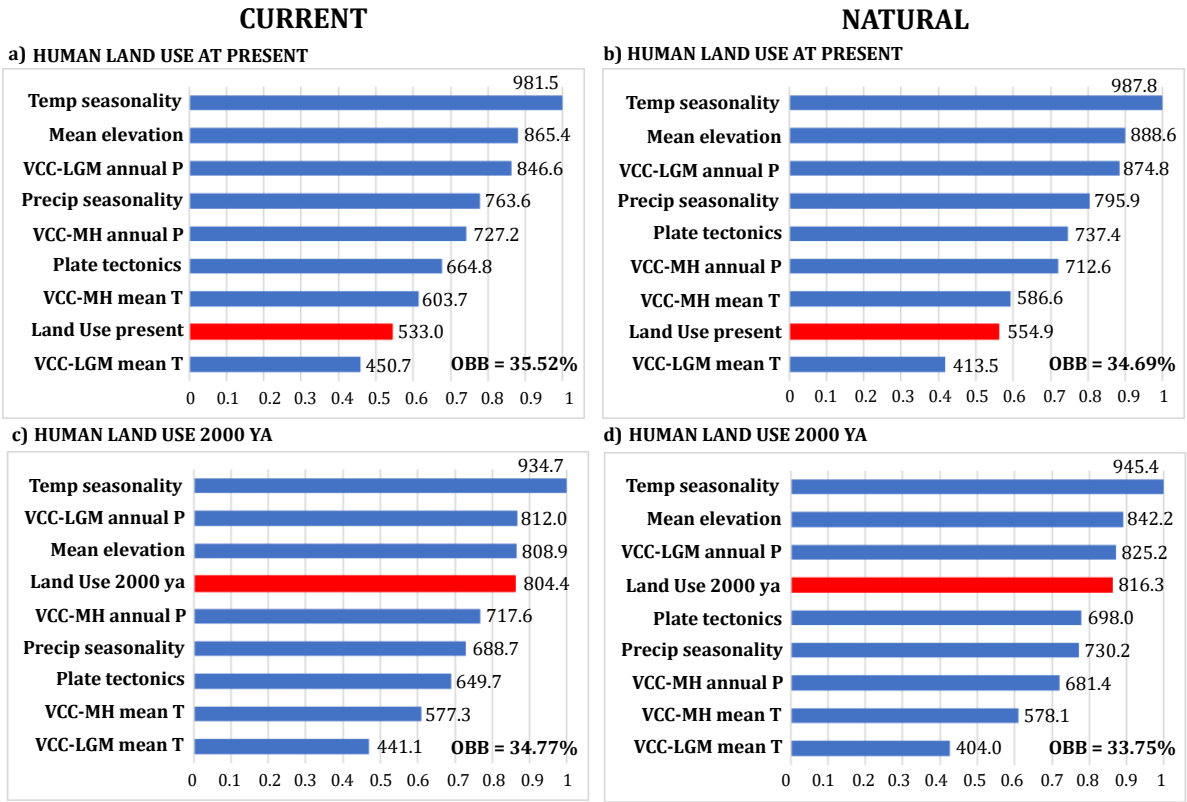

Fig. S7. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural smaller subregions. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2,000 years ago for (c) current and (d) natural bioregions. VCC = velocity of climate change; MH = Mid-Holocene; LGM = Last Glacial Maximum; annual P = annual precipitation; mean T = mean annual temperature; ya = years ago.

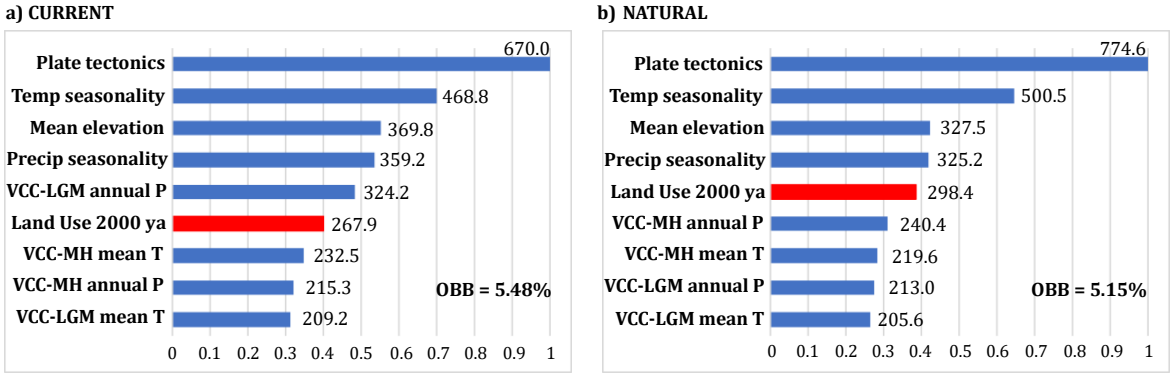

Fig. S8. Ranking of importance values for the drivers of taxonomic differentiation for the current (a) and natural (b) medium-sized subregions when subregions without human land use 2,000 years ago are removed from the analysis. Subregions removed were Madagascar, the two Australian subregions, the Sahara Desert (only for the current subregions), the Tibetan plateau, and the American Arctic. VCC = velocity of climate change; MH = Mid-Holocene; LGM = Last Glacial Maximum; annual P = annual precipitation; mean T = mean annual temperature; ya = years ago.

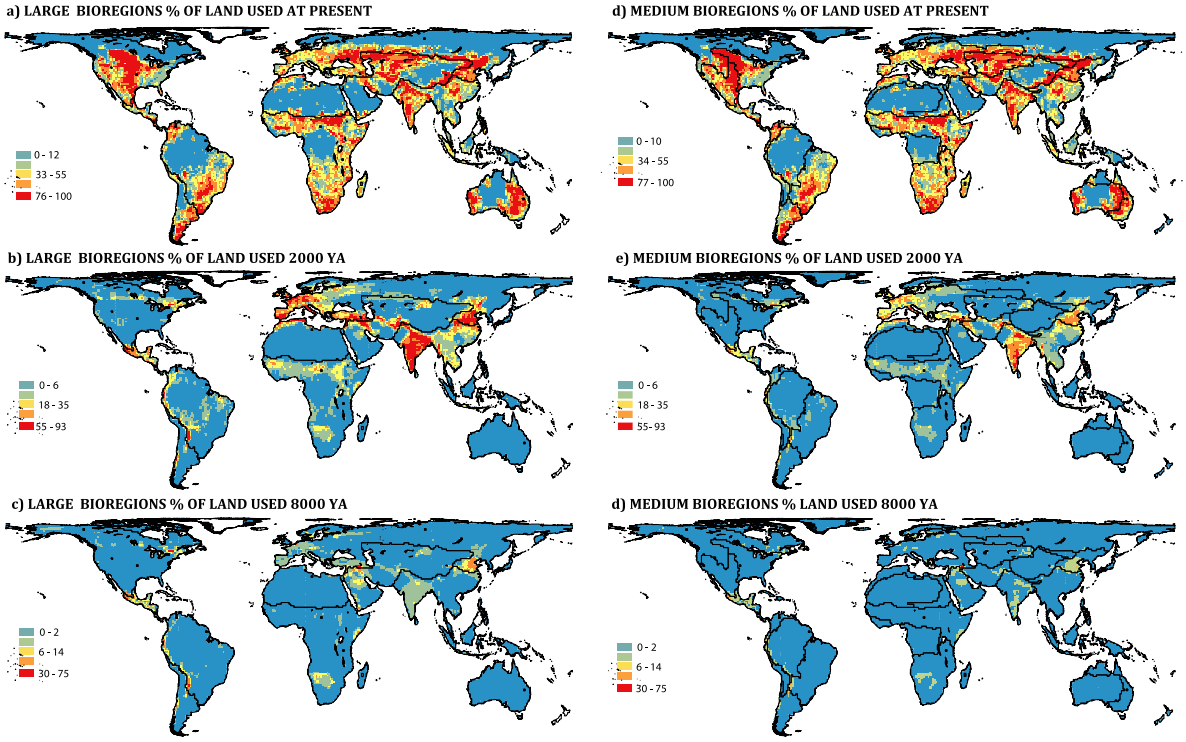

Fig. S9. Percentage of human land used at present, 2,000 and 8,000 years ago for broader regions and medium-sized bioregions.

Table S1. Jaccard index for current and natural hierarchical bioregions built with the 536 extant mammals whose distributions ranges have shifted due to human impact. 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 of the bioregionalization processes is also provided.

CURRENT	Jaccard coefficient	NATURAL
6 broader realms	0.48	6 broader realms
18 subregions	0.33	17 subregions
93 small bioregions	0.20	98 small bioregions
1,333 smaller bioregions	0.19	997 smaller bioregions

Table S2. Pearson's correlations between continuous determinants. MAT = Mean annual temperature; VCC = velocity of climate change; MH = Mid-Holocene; LGM = Last Glacial Maximum; ya = years ago.

	Elevation	MAT	Annual precip	Temp season	Precipit season	VCC-LGM Temp	VCC-MH Temp	VCC-LGM Precip	VCC-MH Precip	Land use present	Land use 2,000 ya	Land use 5,000 ya	Land use 8,000 ya
Elevation	1												
MAT	-0.21	1											
Annual precipitation	-0.15	0.35	1										
Temp seasonality	-0.007	-0.84	-0.55	1									
Precipit seasonality	0.21	0.30	-0.18	-0.18	1								
VCC-LGM Temp	-0.27	-0.41	-0.12	0.44	-0.29	1							
VCC-MH Temp	-0.21	0.13	0.03	-0.06	0.09	0.32	1						
VCC-LGM Precip	-0.17	-0.16	0.33	0.04	-0.19	0.37	0.04	1					
VCC-MH Precip	-0.09	0.36	0.24	-0.36	0.37	-0.17	0.19	0.11	1				
Land use present	0.07	0.20	-0.06	-0.10	0.13	-0.08	-0.05	-0.12	0.13	1			
Land use 2,000 ya	-0.0007	0.12	0.07	-0.15	0.13	-0.01	-0.01	-0.08	0.04	0.29	1		
Land use 5,000 ya	0.016	0.07	0.05	-0.06	0.11	-0.04	-0.03	-0.04	-0.01	0.22	0.87	1	
Land use 8,000 ya	0.006	0.06	0.008	-0.04	0.08	-0.01	-0.03	-0.04	-0.03	0.13	0.63	0.86	1