University of Reading

Human activity is altering the world's zoogeographical regions

Article

Accepted Version

Bernardo-Madrid, R., Calatayud, J., Gonzalez-Suarez, M. ORCID: https://orcid.org/0000-0001-5069-8900, Rosvall, M., Lucas, P. M., Rueda, M., Antonelli, A. and Revilla, E. (2019) Human activity is altering the world's zoogeographical regions. Ecology Letters, 22 (8). pp. 1297-1305. ISSN 1461-0248 doi: 10.1111/ele.13321 Available at https://centaur.reading.ac.uk/84049/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1111/ele.13321

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Title: Human activity is altering the world's zoogeographical regions
Short title: Zooregions of the Anthropocene

Authors: Rubén Bernardo-Madrid[1]*, Joaquín Calatayud[2,3,4†], Manuela González-Suárez[5†], Martin Rosvall[4], Pablo M. Lucas[1,6], Marta Rueda[1], Alexandre Antonelli[7,8,9], Eloy Revilla[1]

Author Affiliation:
[1] Department of Conservation Biology, Estación Biológica de Doñana (EBD-CSIC), Sevilla, Spain
[2] Department of Life Science, Universidad de Alcalá, Alcalá de Henares, Spain
[3] Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain
[4] Integrated Science Lab, Department of Physics, Umeå University, 901 87 Umeå, Sweden
[5] Ecology and Evolutionary Biology, School of Biological Sciences, University of Reading, Reading, UK
[6] Department of Wildlife Conservation, Institute of Nature Conservation (IOP-PAS), Kraków, Poland
[7] Gothenburg Global Biodiversity Centre, Box 461, SE-405 30 Göteborg, Sweden
[8] Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Göteborg, Sweden.
[9] Royal Botanic Gardens, Kew, Richmond, TW9 3AB, UK

[*] Corresponding author. Email: ruben.bernardo.madrid@gmail.com
[†] These two authors contributed equally

Keywords: bioregions, conservation, human impacts, global change, extinction, invasion, threatened species, species assemblages, uncertainty, robustness

Statement of authorship: R.B-M., J.C., M.G-S. and E.R. conceived the project; R.B-M., J.C., M.Ros., M.G-S., A.A. and E.R. conceived the methodological approach; R.B-M., J.C., M.Ros. and E.R. developed the methodological approach; R.B-M., M.G-S., P.M.L. and E.R prepared the data; R.B-M., J.C. and M.Ros. developed the similarity index; R.B-M. and J.C. developed the robustness and impact indexes and performed the analysis. All authors discussed and wrote the manuscript.

Data accessibility: We used open access data from IUCN and BirdLife.
Type of article: Letter
Number of words: Abstract (150), main text (4968)
Number of references: 64
Number of figures: main text (3), supporting information (14)
Number of words in the main figures: figure 1 (124), figure 2 (120), figure 3 (86)
Number of tables: main text (0), supporting information (3)

Abstract

Zoogeographical regions, or zooregions, are areas of the Earth defined by species pools that reflect ecological, historical, and evolutionary processes acting over millions of years. Consequently, researchers have assumed that zooregions are robust and unlikely to change on a human timescale. However, the increasing number of human-mediated introductions and extinctions can challenge this assumption. By delineating zooregions with a network-based algorithm, here we show that introductions and extinctions are altering the zooregions we know today. Introductions are homogenising the Eurasian and African mammal zooregions and also triggering less intuitive effects in birds and amphibians, such as dividing and redefining zooregions representing the Old and New World. Furthermore, these Old and New World amphibian zooregions are no longer detected when considering introductions plus extinctions of the most threatened species. Our findings highlight the profound and far-reaching impact of human activity and call for identifying and protecting the uniqueness of biotic assemblages.

Main text

Zooregions are areas of the Earth defined by characteristic species pools that are the outcome of long (co)evolutionary histories, local and global extinctions, and colonisation events (Slater 1858; Wallace 1876; Calatayud *et al.* 2018). Understanding changes in zooregions is essential because species assemblages on a large spatial scale influence species diversity and function and drive evolutionary processes on lower spatial scales (Elton 1958; Ricklefs 2004; Ricklefs 2008; Ricklefs *et al*. 2015; Calatayud *et al.* 2016). The role that colonisation and diversification play in shaping zooregions varies with time and space, with diverse factors such as tectonic movements, climate change, geographic barriers, and ecological interactions acting on different scales (Wallace 1876; Ficetola *et al.* 2017). Because these processes have been acting for millions of years, it has been commonly assumed that zooregions are unlikely to change on a human timescale. However, the current extent and magnitude of human activity are challenging this assumption (Capinha *et al.* 2015).

Human activity is affecting the dynamics of diversification and dispersal, triggering both local and global extinctions directly via habitat destruction and fragmentation (Tilman *et al.* 1994) and persecution (Ceballos *et al.* 2015; Burgess *et al.* 2017), and indirectly via extinction cascades (Ripple *et al.* 2015). Moreover, human trade and travel are altering dispersal dynamics by breaking down natural barriers for some species (Banks *et al.* 2015), while limiting large-scale movements for others (Tucker *et al.* 2018). These processes have already produced changes in the diversity patterns of many groups of species on both local and regional scales (Carvalheiro *et al.* 2013; Magurran *et al.* 2015). However, whether the effects of human activity scale up to the highest biotic assemblages on Earth remains largely unknown.

Theoretically, species introductions and extinctions can trigger three types of zoogeographical changes by altering the patterns of species co-occurrence (Fig. 1). Species introductions can homogenize zooregions that encompass native and introduced ranges (Winter *et al.* 2009; Capinha *et al.* 2015), but they can also differentiate invaded and non-invaded areas (Olden *et al.* 2008) by decreasing similarities within the original zooregion and fragmenting it into smaller independent zooregions, a process we call division. Complementarily, extinctions can cause homogenization, if species unique to zooregions go extinct (McKinney & Lockwood 1999; Winter *et al.* 2009; Villéger *et al.* 2011), or division, if widespread species in a zooregion go extinct due to increasing regional dissimilarity (Martin & Stuart 1995; Roberts *et al*. 2001; Fig. 1). Finally, introductions and extinctions can redefine the boundaries of zooregions without varying their number, by rearranging species distributions. For instance, an area may be reassigned to an adjacent zooregion if species associated with that region invade it or if the area loses some of its associated species (Fig. 1). In these ways, individually or in combination, species introductions and extinctions have the theoretical potential to homogenise, divide, and redefine zooregions.

To answer the question of whether human-mediated introductions are affecting zooregions as we know them today, we can compare zooregions delineated with native species distributions with those delineated with both native and introduced species distributions. In addition, information about the conservation status of species can be used to explore the potential effects of species losses by simulating global extinctions of the most threatened species. Here we followed this approach in order to assess how observed introductions and the potential extinction of the most threatened species could affect zooregion delineations of amphibians, non-marine mammals, and non-pelagic birds. We delineated zooregions under two actual scenarios: (i) *native* – the native distribution range of species, and (ii) *introduced* – both the native and the introduced distribution ranges of species; and two hypothetical scenarios: (iii) *extinct* – the native scenario excluding the most threatened species, and (iv) *introduced plus extinct* – the combination of the introduced and extinct scenarios. We delineated zooregions using recently developed classification methods based on network theory (Carstensen *et al.* 2012; Sidor *et al.* 2013; Vilhena & Antonelli 2015; Bloomfield *et al.* 2017; Edler *et al.* 2017). Network algorithms not only detect groups of grid cells representing zooregions but also identify groups of highly associated grid cells and species, representing zooregions and their highly associated species, respectively (Carstensen *et al.* 2013), and offering interesting opportunities to evaluate biogeographical patterns beyond delineation. We took advantage of this potential by exploring several new metrics which we anticipate would be useful in future studies to better understand species assemblages. Altogether, we provide new insights into how human activity is affecting biodiversity patterns on a global scale.

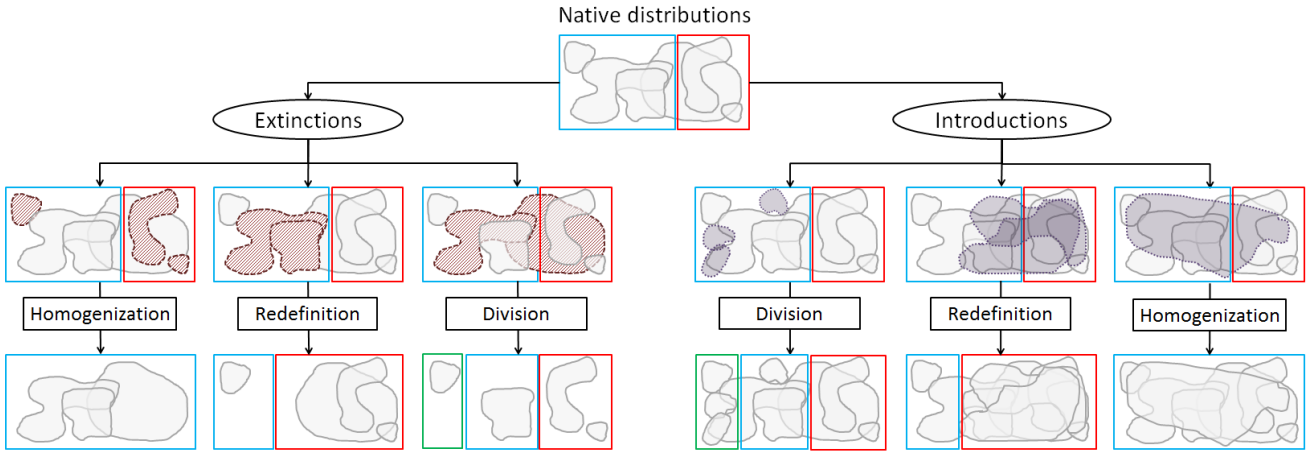


Figure 1: Theoretical changes in zooregion delineation as a consequence of human-mediated species extinctions and introductions. The blue, red, and green rectangles represent three hypothetical zooregions with inside polygons reflecting hypothetical species distributions. The diagram has three levels: (i) at the top of the figure there is a natural (pre-human impact) scenario with two zooregions (blue and red); (ii) this natural state is disrupted by either extinctions (left panels; disappearing species are marked by red polygons with dashed lines) or introductions (right panels; introduced species are marked by grey polygons with solid lines); (iii) the third level (human-affected) represents the outcome of these extinctions and introductions, and their changes to the original zooregions. The actual observed changes depend on how the composition of species is affected.

MATERIAL AND METHODS

Data

To reliably compare our four scenarios and vertebrate groups, we require an extensive dataset with broad taxonomic coverage that is as spatially consistent as possible. Therefore, we used species range maps in the form of polygons from the IUCN (2015) and BirdLife International (2015). For our analyses, we selected polygons with a presence classified as “Extant” or “Probably Extant”. We used the polygons classified as “Native” or “Reintroduced” to define the native ranges and the polygons classified as “Introduced” to

define the introduced ranges. We converted ranges into presence/absence maps in a regular global grid with cells of 111km x 111km, with a cylindrical equal-area projection (Holt *et al.* 2013). Following previous studies, we limited the analyses to the emerged land surface of the world and included only cells with ≥50% of their area classified as land (Rueda *et al.* 2013). After removing subspecies and species that did not overlap with this grid, our dataset included 6,103 amphibian species, 5,001 mammal species, and 9,834 bird species.

To explore the effects of introductions and extinctions, we generated two actual scenarios and two hypothetical scenarios. The two actual scenarios were: *native,* based on the current native distribution range of extant species, and *introduced,* based on both the native and introduced ranges of extant species, to describe the current species distributions. IUCN and BirdLife provided information about introduced ranges for 71 amphibian species, 114 mammal species, and 97 bird species. The two hypothetical scenarios were: *extinct,* derived from the native scenario but simulating the global extinction of the most threatened species, those listed as Critically Endangered or Endangered on the IUCN Red List (2015), and *introduced plus extinct,* which considers both introductions and extinctions simultaneously. Based on the IUCN Red List status, we simulated the extinction of 1,283 amphibian species, 656 mammal species, and 406 bird species.

Zoogeographical delineations

We used a network approach to delineate zooregions in our four scenarios. For this purpose, we first transformed the presence/absence matrices in networks, where nodes represent grid cells and species and links between nodes represent occurrences of species into grid cells (Appendix A; Fig. S1). Under this approach, a zooregion is a group of grid cells strongly linked to the same species pool, which we call its characteristic species. We identified groups composed by highly associated grid cells and species, called modules, by using the community-detection algorithm Infomap (Rosvall & Bergstrom 2008; Fig. S1).

Infomap exploits a central concept of information theory known as the minimum description length (MDL) principle for model selection. The MDL principle states that the best hypothesis or model to explain regularities in a given set of data is the one that can describes the data with the least amount of information (Rissanen 1978). In our biogeographical network, regularities correspond to modules of highly interconnected grid cells and species, while minimising the description length corresponds to describing the network with an optimal set of modules (see details in Appendix B). To identify the zooregion delineation that best explains regularities in the links between grid cells and species, Infomap uses an efficient search algorithm that starts from an initial delineation with a long description length of the network, hereafter codelength, and iteratively finds better delineations. At the start, each node is assigned to its own module (Fig. S2F). Infomap then iteratively selects a random node and assigns it to the module that compresses the codelength the most. If moving a node does not further compress the codelength, the node remains in the original module. Infomap continues to move nodes until the codelength cannot be further compressed (Fig. S2). Infomap can also iteratively and recursively rearrange groups of nodes to improve the delineation further. See Appendix B and Rosvall & Bergstrom (2008), Rosvall et al. (2009), for a detailed explanation.

We used a community-detection algorithm because it delineates large-scale zooregions more accurately than traditional approaches (Vilhena & Antonelli 2015; Bloomfield *et al*. 2018). We used Infomap rather than other potential community-detection algorithms because of its high overall performance (Lancichinetti & Fortunato 2009; Puxeddu *et al*. 2017) and because previous studies have shown that Infomap performs well when delineating bioregions of multiple taxonomic groups at different spatial scales (Sidor *et al.* 2013; Vilhena & Antonelli 2015; Bloomfield *et al.* 2017; Edler *et al.* 2017; Costello *et al*. 2017; Droissart *et al.* 2018; Hazzi *et al*. 2018). We also used Infomap because it suited our goals. First, it can describe the uncertainty that is inherent when assigning geographical areas to zooregions, as occurs in transitional areas (see below; Fig. S3). As Infomap investigates nodes in a stochastic order, repeated runs of Infomap will generate the same zooregion delineation only if that delineation stands out among the

alternatives. However, if there is any uncertainty, alternative zooregion delineations may explain species occurrence patterns similarly, and Infomap can detect them in repeated analyses. Second, we can use the codelength of the zooregion delineation to assess its statistical support (see below). Third, Infomap directly provides the number of hierarchical levels and zooregions within each level that minimises the codelength, removing the subjectivity inherent to other approaches where the researcher must determine those numbers to make results more comparable (see Appendix B for further details; Bloomfield *et al.* 2018; Zhao & Schranz 2019). As most bioregionalizations at a global scale primarily focus on zoogeographical realms and regions (Duellman 1999; Cox 2001; Holt *et al.* 2013; Rueda *et al.* 2013; Newbold *et al.* 2015; Vilhena & Antonelli 2015), here we only used the hierarchical levels that better matched those spatial resolutions. In particular, we only used the zooregions at the highest and lowest hierarchical levels, which we refer to as major and minor zooregions, respectively. Fourth, Infomap identifies modules as groups of grid cells and highly associated species. Grid cells grouped in the same module represent a zooregion, while species highly associated with them are the characteristic species of that zooregion. Thus, species present in a zooregion but not clustered in the corresponding module can be described as the non-characteristic species (Appendix A; Fig. S1). Zooregions are defined by the balance in composition between characteristic and non-characteristic species. Therefore, we can use species composition to go beyond zooregion delineation and quantify the *a priori* resistance by an area, i.e. grid cell, to change its zooregion assignation due to introductions and extinctions, which we call biogeographical robustness. Similarly, we can quantify how strongly introductions and extinctions are affecting that robustness, which we call biogeographical impact (see below).

Biogeographical uncertainty: Alternative zooregion delineations

Biogeographical uncertainty reflects the complexity and overlap of species distributions and the processes that shape zooregions (e.g., Wallace's Line). To better understand the effects of introductions and extinctions, we study all alternative delineations to account for uncertainty (see Appendix C for other benefits). Introductions and extinctions can trigger new zooregion delineations, but may also affect the probability of detecting delineations that already occur in the native scenario. We estimated the detection probability of each alternative zooregion delineation for each scenario, hereafter alternative solutions and their detection probabilities. Because the number of alternative solutions detected and their detection probabilities can be inaccurate if the number of analyses is low, we developed an approach that calculates how many runs we needed in order to describe each scenario and vertebrate group accurately. Please see figure S4 for a description of the workflow followed in the study.

We started by performing 100 analyses per vertebrate group and scenario. We obtained the assignation of each grid cell to the major and minor zooregions (100 delineations for each level; see Appendix C). We identified the alternative zooregion delineations in each level and scenario by grouping the 100 Infomap outputs based on their similarity to each other (Fig. S4). To this end, we developed a similarity metric that measures which proportion of the Earth's surface of two different zooregion delineations is the same (see Appendix D for details about the similarity index). To group zooregion delineations by their similarity (similarity threshold 0.9), we used Infomap again, but this time using the 100 zooregion delineations as nodes and their similarity values as links (see Appendix C for details and assessment of the threshold). This step resulted in a set of modules representing similar zooregion delineations. We called each module an alternative solution and used the relative abundance of delineations grouped in each alternative solution as its detection probability. To evaluate whether this number of analyses was sufficient, i.e., if we detected all alternative zooregion delineations with a precise detection probability, we subsequently ran 100 new analyses per vertebrate group and scenario, i.e., grid cells and species as nodes, and extracted again the assignation of the grid cells to the major and minor zooregions. We then combined the zooregions' delineations with the first 100 and repeated the clustering

analyses, using Infomap but clustering the 200 zooregion delineations. We compared the detection probabilities of all alternative zooregion delineations based on 100 vs 200 analyses using Fisher's exact test. We continued this process of adding 100 new delineations in each step until the Fisher's *p*-value reached an asymptote near one (Appendix C). Using this approach, we found that, with 1,500 analyses per vertebrate group and scenario, we reached that asymptote in all vertebrate groups and scenarios (Fig. S5).

Finally, we tested whether the detected zooregion delineations represented trivial species' co-occurrence patterns by comparing their codelength with the codelengths of delineations obtained with a null biogeographical model (Appendix E for details). All the code developed to perform our analyses is freely available in Appendix F.

Effects of introductions and extinctions

For each hierarchical level and vertebrate group, we used the Schoener's D index (Schoener 1974) to assess how introductions and extinctions alter the detection probabilities of alternative solutions. Schoener's D index is widely used to describe the categorical overlapping of two objects (e.g., Petitpierre *et al*. 2012). In our case, we used the Schoener's D index to estimate the overlap in the detection probabilities of the alternative solutions for each pair of scenarios. Schoener's D index ranges from 0, when two scenarios share no solutions, to 1, when the detection probabilities in two scenarios are identical. While this index reflects whether and how much introductions and extinctions affect our capacity to detect specific alternative zooregion delineations, i.e., the detectability of changes, it does not capture the type of zoogeographical changes – homogenization, division, or redefinition – the location of the changes, or how much of the Earth's surface is affected. To assess the type of zoogeographical change, we first used the detection probabilities of the native scenario as a starting point and observed which alternative solutions increase and decrease their detection probabilities when considering introductions and extinctions. Second, we mapped alternative solutions to understand and locate the zoogeographical changes (Figs. 2 and S6). Finally, we quantified the extent of the Earth's surface that was altered due to introductions and extinctions. To do that, we calculated the mean similarity of all the pairs of delineations grouped into two alternative solutions.

Biogeographical robustness and impact indexes

The assignation of a given grid cell to a zooregion depends on the pool of species co-occurring within it. While the presence of some species indicates that the grid cell belongs to the assigned zooregion (i.e., characteristic species) other species dilute that zooregion assignation (i.e., non-characteristic species). Thus, the strength of the assignation of a grid cell depends on the balance between indication and dilution. However, species introductions and extinctions may alter this balance and trigger changes in the zooregion assignation of grid cells. Because of delineating zooregions considering introductions and extinctions only inform about zoogeographical changes, but does not offer clues to understand why changes occur, we propose to study the *a priori* resistance of grid cell against changes in zooregion assignation, that is biogeographical robustness. To this end, we first calculated the indicator and dilution value of each species for each grid cell. We calculated the indicator value of the characteristic species, hereafter *IndVal*, by using the formula proposed by Dufrêne & Legendre (1997), which considers the affinity and fidelity of species to their zooregion. The affinity of the species *i* is described as $A_i = R_i/Z$, where R_i depicts the distribution range of the species *i* overlapping its associated zooregion and *Z* depicts the total area of the associated zooregion. The fidelity of the species *i* is defined as $F_i = R_i/D_i$, where D_i is the total distribution area of the species *i*. Finally, *IndVal* is calculated as $IndVal_i = A_i \times F_i$. A species that occupies most of its associated zooregion (i.e., high affinity) and is not present in other zooregions (i.e., high fidelity) will have a high *IndVal*. We calculated the dilution value of the non-characteristic species, hereafter *DilVal*, as a modification of *IndVal*, substituting fidelity by infidelity (1 − *fidelity*). Thus, it was calculated as $DilVal_i = A_i \times$

(1 − F_i). A species overlapping with much of its associated region (i.e., high affinity) but mostly distributed in other zooregions (i.e., high infidelity) will have a high *DilVal*. Finally, we calculated the biogeographical robustness value of a grid cell *g* as

$$R_g = \sum_{i=1}^{C} IndVal_i - \sum_{j=1}^{N} DilVal_j \ , \quad (1)$$

where *C* represents the number of characteristic species of the zooregion in the grid cell *g*, while *N* is the number of non-characteristic species. Grid cells containing a large number of highly indicative species and a few species with low dilution values will have high biogeographical robustness and will be less vulnerable to changing their zooregion assignation due to introductions and extinctions.

Similarly, we can quantify to what degree introductions and extinctions alter the biogeographical robustness of a grid cell, that is biogeographical impact. In particular, we measured to what degree introductions and extinctions reduce *IndVal* and increase *DilVal*. Specifically, we calculated the decrease in the summation of *Indval* as

$$\Delta I = \sum_{i=1}^{C} IndVal_{native_i} - \sum_{i=1}^{C} IndVal_{focal-scenario_i} \ , \quad (2)$$

and the increase in the summation of *DilVal* as

$$\Delta D = \sum_{j=1}^{N} DilVal_{focal-scenario_j} - \sum_{j=1}^{N} DilVal_{native_j} \ , \quad (3)$$

where focal−scenario refers to the introduced or extinct scenario. The total biogeographical impact in each grid cell is the sum of both values. For example, a grid cell that loses highly indicative characteristic species and gains highly diluter non-characteristic ones will experience high biogeographical impact.

RESULTS

Comparison of the scenarios

Our native zooregions are similar to the ones first proposed by Wallace (1876). However, for amphibians and birds, we also detected an additional higher hierarchical level, delineating two main major zooregions corresponding to the Old and New World (Fig. 2, A1 and B1). The quality of observed zooregion delineations for all vertebrate groups and scenarios was significantly better than those of delineations detected under a null model (p=0.01; see Appendix E for details of the metric and significance assessment used to test it). In the major zooregions of mammals, species introductions triggered large changes in the detection probability of solutions across scenarios (D = 0.25, Fig. 2), increasing the probability of detecting a homogenised major zooregion that includes Africa, Eurasia, and the Indo-Malayan regions (Fig. 2, M3 and Tables S1 and S2). In amphibians, introductions led to a single solution that differed from the native solution (D = 0, Fig. 2, A2), with major zooregions representing the Old and New World redefined due to the relocation of the Australian region into the New World zooregion. In birds, introductions divided major

zooregions. Although this effect was less detectable than in previous cases (D = 0.83, Fig. 2), bird introductions decreased the detectability of the two major zooregions corresponding to the Old and New World (Fig. 2, B2). Regarding minor zooregions, mammalian introductions increased the similarity between the North African and Arabian minor regions (D = 0.38; Fig. S11, m2). We found no zoogeographical changes in the minor zooregions of amphibians or birds (D = 1 and D = 0.99, respectively; Fig. S6). In all cases, zoogeographical changes in major zooregions affected larger fractions of the Earth's surface than in minor zooregions (Table S3). Overall, our results revealed that introductions had a broad range of effects, triggering redefinition, homogenization, and division of the major zooregions of amphibians, mammals, and birds, respectively, as well as homogenization of the minor zooregions of mammals (Figs. 2 and S6).

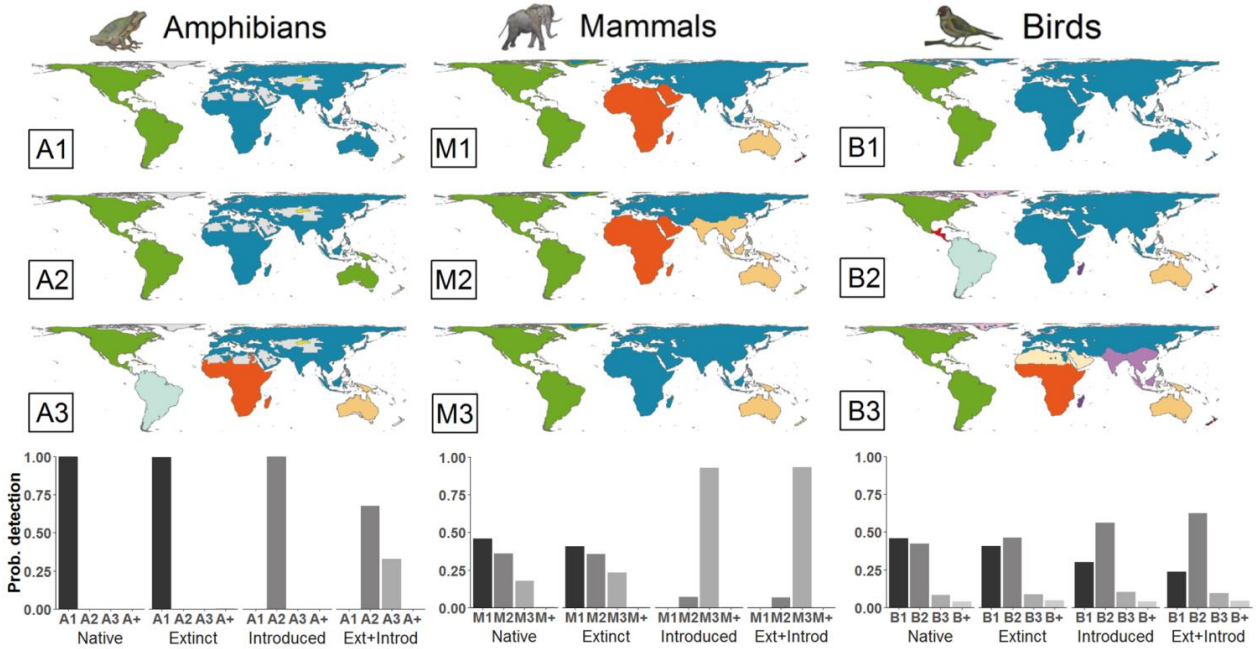


Figure 2: Observed introductions and potential extinctions alter the detection probability of the native zooregion delineations in amphibians, mammals, and birds. The maps show the delineations of the three most probable solutions for the major zooregions of amphibians, mammals, and birds (labelled as A1–A3 for amphibians, M1–M3 for mammals, and B1–B3 for birds), and the histograms show their detection probability. In the amphibian maps, areas where amphibians are not present, according to the IUCN, are in grey. The maps represent the consensus of the delineations grouped in the same solution. The histograms also show the joint probability of the detection of less frequently observed solutions, not shown in the maps (see Table S1 and Fig. S7-9 for details).

Extinction, on the other hand, had no significant effect on zooregions alone (Table S2). However, the combination of amphibian introductions and extinctions triggered new changes in which zooregions representing the Old and New World were no longer detected (Fig. 2, A3).

Robustness and impacts

Stronger biogeographical impacts often occurred in geographical areas where we detected zoogeographical changes (Figs. 2 and 3). For example, biogeographical impacts coincided with the relocation of the Australian-South American region in amphibians, the homogenization of Africa and Eurasia in mammals, and the division of the Americas in birds (Figs. 2 and 3B). Conversely, some areas with high biogeographical impacts, such as the American region in mammals or the African and Eurasian regions in birds, showed no zoogeographical changes. Biogeographical robustness across zooregions and geographical areas may explain why similar biogeographical impacts trigger changes in some circumstances but not in others (Fig. 3A and Fig. S13). For instance, the higher biogeographical robustness in birds and mammals could explain

why introductions and extinctions resulted in fewer changes in the detection probability of their alternative solutions (Fig. 3A and Table S1). Indeed, zoogeographical changes in birds and mammals encompassed a large extent of Earth's surface, but the detection probabilities of the alternative solutions were not strongly affected (Figs. 2 and S6 and Tables S1 and S3). Similarly, the higher changes in the detection probabilities of the alternative solutions observed in major zooregions compared to those of the minor zooregions could be explained by their lower biogeographical robustness (Figs. 3 and S13 and Table S2).

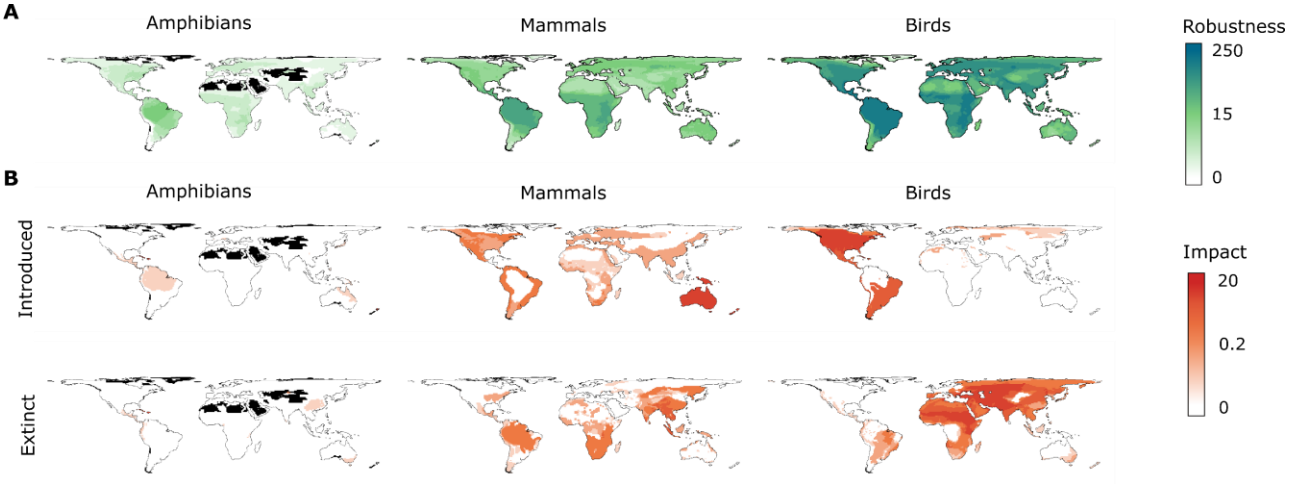


Figure 3: Maps representing biogeographical robustness and impact under different scenarios, showing how both vary across regions and vertebrate groups. We represent the (A) robustness and (B) impact values for the grid cells of the major zooregions, calculated using the most frequent solution for each vertebrate group. (A) Light and dark blue colours indicate low and high relative biogeographical robustness, respectively. (B) Light and dark red colours indicate low and high relative biogeographical impact, respectively. In amphibian delineations, black areas indicate where amphibians are not present, according to the IUCN.

DISCUSSION

Our results suggest that species introductions are already modifying the world's zooregions of amphibians, mammals, and birds at different spatial scales across the Earth. Furthermore, alterations in zooregions can be greater if most threatened species go extinct. Contrary to general assumptions, species introductions are triggering not only homogenization on a global scale but also other effects such as division and redefinition. Our results suggest that introductions lead to zoogeographical changes in a complex interplay between the impact, the species pools involved, and the intrinsic properties of the zooregions, including biogeographical robustness, which highlights the value of going beyond delineations to achieve more effective biodiversity assessments.

Human-mediated species introductions and extinctions are heterogeneously affecting the delineation of vertebrate zooregions across spatial scales. Our results suggest that introductions are homogenizing the African and Eurasian major and minor mammalian zooregions by making their already similar species assemblages even more similar (Wallace 1984; Dobson & Wright 2000; Sanmartin *et al.* 2001), redefining amphibian zooregions representing the New and Old World by reallocating Australia (Fig. 2A), and dividing bird zooregions representing the New and Old World by increasing the internal heterogeneity of the bird major zooregions due to intercontinental introductions (Dyer *et al.* 2017). Although potential extinctions appear to have little effect on zooregion delimitations when they are considered alone, the combined effect of current introductions and extinctions could induce important modifications in the future, such as the blurring of biogeographic differences between the amphibians of the Old and New World (Duellman 1999).

While our results suggest substantial changes worldwide, there are some limitations in the available data on species distribution. IUCN and BirdLife distribution maps are the most comprehensive global sources of range data for the three vertebrate groups we analysed, but these maps do not reflect all human-mediated introductions and extinctions. For example, native range distributions may not reflect recent range contractions due to habitat loss and climate change (Rondinini et al. 2011) or past human-mediated extinctions such as the Holocene megafaunal extinctions (Lorenzen et al. 2011; Faurby & Svenning 2015). Introduced ranges also likely underestimate the actual extent of introduced species because of the difficulties of tracking ongoing range expansion. Indeed, the available IUCN data do not yet include updated introduced ranges such as those of the hog deer (*Axis porcinus*) in Australia (Bentley 1998). Thus, despite the substantial changes reported, our findings are in fact conservative, even when simulating potential extinctions. We are also aware that the distribution ranges of amphibians, mammals, and birds have not been recorded with the same sampling effort across the world, which can potentially bias our results. As updated information on a global scale becomes available (Rondinini *et al.* 2011; Faurby & Svenning 2015; Dawson *et al.* 2017; Dyer *et al.* 2017), we may be able to refine our understanding of human impacts on the distribution of life on Earth. In the meantime, our results strongly indicate that the delineation of biogeographical regions is already being affected by human impacts.

In our study, we only used species composition to detect zooregion changes. These changes are directly applicable to conservation policies, as species (and sometimes populations) are currently the main management and conservation targets. However, there are proposals to consider additional biodiversity metrics as conservation targets (Vézquez & Gittleman 1998; Rosenfeld 2002; Olden *et al*. 2004), such as phylogenetic clades and functional traits (Olden *et al*. 2004; Devictor *et al.* 2010). We expect that bioregionalization analyses based on multiple dimensions of biodiversity will lead to a holistic understanding of human impacts on biodiversity. Therefore, we encourage future studies to assess whether human activity is also altering other aspects of biodiversity, as well as the potential consequences of such alterations.

Our results show that biogeographical robustness, as well as the location and intensity of human activities, is highly associated with changes in zooregion delineation. In particular, biogeographical robustness helps to understand whether a given impact can trigger changes in zooregion assignation and to explain the differences among vertebrate groups and spatial scales. We suggest that biogeographical robustness captures different elements such as environmental and topographic conditions (Ficetola *et al.* 2017) and dispersal abilities. For instance, the relatively high dispersal capabilities of many birds (particularly as compared to amphibians) allow them to occupy a large extent of their associated zooregions, which gives them high indicator values (Figs. 3 and S13). Thus, dispersal capabilities may explain why major zooregions of birds are more robust than those of mammals, which in turn are more robust than those of amphibians, the most dispersal-limited group analysed (Figs. 3 and S13). Similarly, dispersal capabilities could explain the differences between major and minor zooregions. Because there are few species with continental distribution ranges, only a few species are highly indicative in major zooregions compared to minor zooregions, making them less robust (Figs. 3, S13, and S14).

Differences in robustness between major and minor zooregions can also explain why introductions trigger more zoogeographical changes than extinctions. Simulated extinctions corresponded to species with generally small distribution ranges. Consequently, those species, which tend to have low indicative values, would probably affect minor zooregions. This fact, together with the high robustness of minor zooregions, can make that simulated extinctions trigger few zoogeographical changes. In contrast, introductions can be widespread (e.g. *Rhinella marina*, *Oryctolagus cuniculus*, or *Passer domesticus*), affecting major zooregions. Since major zooregions tend to be less robust, widespread introductions can trigger changes in major zooregions. Moreover, the biased spatial distribution of introductions can reinforce this effect. For example, we found changes in the detection probabilities of alternative solutions in the New World by

including introduced ranges for fewer than 100 alien bird species, however, we note that 40% of alien introductions affected North America (including native North American species introduced to other areas and species from other areas introduced to North America). In summary, if introduction exchanges mostly affect the same geographical areas, relatively few introductions can trigger zoogeographical changes.

Nonetheless, we detected some changes that were not clearly explained by biogeographical robustness and impact, which may reflect more complex consequences of extinctions and introductions. For instance, it is difficult to predict changes in areas where many introduced species from several zooregions occur, such as mammal introductions from America and Eurasia to Oceania. Although we did not detect changes in some highly impacted areas, probably because of their high biogeographical robustness, current alterations of species assemblages are affecting their biogeographical robustness. Consequently, smaller future impacts could trigger changes in zooregion delineations due to cumulative effects. Altogether, these results suggest a negative association between biogeographical robustness and detectability of zoogeographical changes, i.e., variations in detection probabilities of alternative solutions. While no change occurs without impact, robustness seems to be a key metric in predicting zoogeographical changes. We found that the proposed biogeographical robustness metric is useful for identifying vulnerable biogeographic areas. While additional studies will be required to unravel the specific causes and consequences of bioregional robustness, we propose that our methodology can be used as an additional tool in wide-scale biodiversity conservation.

In summary, our findings suggest that human activity is changing zooregions, which are arguably the largest conservation targets of our biologically diverse world. Comparable reconfigurations have previously been reported only after glaciations and mass extinction events (Sidor *et al*. 2013; Calatayud *et al*. 2018). Modifications in the species composition of zooregions come with important consequences for the evolutionary (Jablonski 2005; Erwin 2014; Calatayud *et al.* 2016; Edie *et al.* 2018) and ecological (Olden *et al*. 2004; Naeslund & Norberg 2006; Ricklefs 2015) processes that characterise and depend on large-scale biotas. Our results highlight the urgency in identifying and protecting the uniqueness of large-scale species assemblages.

ACKNOWLEDGMENTS

We thank the Conservation Biology group of Estación Biológica de Doñana, CSIC, and Elena Marmesat for helpful suggestions. We also thank B. Hawkins, C. Venditti, T. Oliver, O. Gordo, A. Jiménez-Valverde, A. Perrigo, S. Faurby, and Debbie Kolp for insightful discussions and feedback. This work was supported by the following grants and projects: Agencia Estatal de Investigación from the Ministry of Economy, Industry and Competitiveness, Spain, with projects CGL2012-35931 and CGL2017-83045-R AEI/FEDER EU, co-financed with FEDER to E.R., R.B.M., M.G-S., and P.M.L., the Predoctoral Fellowship BES-2013-065753 from the same institution to R.B.M.; a Juan de la Cierva post-doctoral fellowship (JCI-2011- 09158) to M.G-S; project 707587 H2020-MSCA-IF-2015 from the EU H2020 to M.Ru. and E.R.; the Swedish Research Council grant 2016-00796 to M.Ros; the National Science Centre in Poland (2016/22/Z/NZ8/00121) through the BiodivERsA call (project BearConnect, national funders ANR/NCN/BMBF/CCCDI–UEFISCDI/RCN) to P.M.L; and the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013, ERC Grant Agreement n. 331024), the Swedish Foundation for Strategic Research (FFL 15-0196), the Swedish Research Council (B0569601), a Wallenberg Academy Fellowship (2014.0216), the Faculty of Sciences at the University of Gothenburg, and the David Rockefeller Center for Latin American Studies at Harvard University to A.A.

References

Banks N.C., Paini D.R., Bayliss K.L. & Hodda M. (2015). The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecology Letters*, 18, 188-199.

Bentley A. (1998). An Introduction to the Deer of Australia: with special reference to Victoria. 3rd edn. (Australian Deer Research Foundation Ltd: Melbourne)

BirdLife (2015). Bird species distribution maps of the world. Version 5.0. Available at: http://datazone.birdlife.org.

Bloomfield N.J., Knerr N. & Encinas-Viso F. (2018). A comparison of network and clustering methods to detect biogeographical regions. *Ecography*, 41, 1-10.

Burgess M.G., Costello C., Fredston-Hermann A., Pinsky M.L., Gaines S.D., Tilman D*., et al.* (2017). Range contraction enables harvesting to extinction. *Proceedings of the National Academy of Sciences*, 114, 3945.

Calatayud J., Hórreo J.L., Madrigal-González J., Migeon A., Rodríguez M.Á., Magalhães S*., et al.* (2016). Geography and major host evolutionary transitions shape the resource use of plant parasites. *Proceedings of the National Academy of Sciences*, 113, 9840-9845.

Calatayud J., Rodriguez M. A., Molina-Venegas R., Leo M., Horreo J. L. & Hortal J. (2018). Pleistocene climate change and the formation of regional species pools. bioRxiv, 149617. 4 peer-reviewed by Peer Community In Evolutionary Biology https://doi.org/10.1101/149617.

Capinha C., Essl F., Seebens H., Moser D. & Pereira H.M. (2015). The dispersal of alien species redefines biogeography in the Anthropocene. *Science*, 348, 1248-1251.

Carstensen D.W., Dalsgaard B., Svenning J.-C., Rahbek C., Fjeldså J., Sutherland W.J*., et al.* (2012). Biogeographical modules and island roles: a comparison of Wallacea and the West Indies. *J. Biogeogr.*, 39, 739-749.

Carstensen D.W., Lessard J.-P., Holt B.G., Borregaard M.K. & Rahbek C. (2013). Introducing the biogeographic species pool. *Ecography*, 36, 1310-1318.

Carvalheiro L.G., Kunin W.E., Keil P., Aguirre-Gutiérrez J., Ellis W.N., Fox R*., et al.* (2013). Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. *Ecology letters*, 16, 870-878.

Ceballos G., Ehrlich P.R., Barnosky A.D., García A., Pringle R.M. & Palmer T.M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, 1.

Costello M.J., Tsai P., Wong P.S., Cheung A.K.L., Basher Z. & Chaudhary C. (2017). Marine biogeographic realms and species endemicity. Nature Communications, 8, 1057.

Cox B. (2001). The biogeographic regions reconsidered. J. Biogeogr., 28, 511-523.

Dawson W., Moser D., van Kleunen M., Kreft H., Pergl J., Pyšek P., et al. (2017). Global hotspots and correlates of alien species richness across taxonomic groups. Nature Ecology &Amp; Evolution, 1, 0186.

Devictor V., Mouillot D., Meynard C., Jiguet F., Thuiller W. & Mouquet N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecology letters*, 13(8), 1030-1040.

Dobson M. & Wright A. (2000). Faunal Relationships and Zoogeographical Affinities of Mammals in North-West Africa. *J. Biogeogr.*, 27, 417-424.

Droissart V., Dauby G., Hardy O.J., Deblauwe V., Harris D.J., Janssens S*., et al.* (2018). Beyond trees: Biogeographical regionalization of tropical Africa. *J. Biogeogr.*, 45, 1153-1167.

Duellman W.E. (1999). *Patterns of distribution of amphibians: a global perspective*. JHU Press, pp.

Dufrêne M. & Legendre P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological monographs*, 67, 345-366.

Dyer E.E., Cassey P., Redding D.W., Collen B., Franks V., Gaston K.J*., et al.* (2017). The global distribution and drivers of alien bird species richness. *PLoS biology*, 15, e2000942.

Edie S.M., Jablonski D. & Valentine J.W. (2018). Contrasting responses of functional diversity to major losses in taxonomic diversity. *Proceedings of the National Academy of Sciences*, 201717636.

Edler D., Guedes T., Zizka A., Rosvall M. & Antonelli A. (2017). Infomap Bioregions: Interactive Mapping of Biogeographical Regions from Species Distributions. *Systematic Biology*, 66, 197-204.

Elton, C. S. (1958). The ecology of invasions by animals and plants. University of Chicago Press.

Erwin D.H. (2014). Temporal acuity and the rate and dynamics of mass extinctions. *Proceedings of the National Academy of Sciences*, 111, 3203-3204.

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

Ficetola G.F., Mazel F. & Thuiller W. (2017). Global determinants of zoogeographical boundaries. *Nature Ecology & Evolution*, 1, 0089.

Hazzi N. A., Moreno J. S., Ortiz-Movliav C., & Palacio R. D. (2018). Biogeographic regions and events of isolation and diversification of the endemic biota of the tropical Andes. *Proceedings of the National Academy of Sciences*, 115, 7985-7990.

Holt B.G., Lessard J.-P., Borregaard M.K., Fritz S.A., Araújo M.B., Dimitrov D*., et al.* (2013). An Update of Wallace's Zoogeographic Regions of the World. *Science*, 339, 74-78.

IUCN (2015). The IUCN Red List of Threatened Species. Version 2015.3. Available at: http://www.iucnredlist.org.

Jablonski D. (2005). Mass extinctions and macroevolution. *Paleobiology*, 31, 192-210.

Lancichinetti A. & Fortunato S. (2009). Community detection algorithms: a comparative analysis. *Physical review E*, 80, 056117.

Lorenzen E.D., Nogués-Bravo D., Orlando L., Weinstock J., Binladen J., Marske K.A*., et al.* (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, 479, 359.

Magurran A.E., Dornelas M., Moyes F., Gotelli N.J. & McGill B. (2015). Rapid biotic homogenization of marine fish assemblages. *Nature communications*, 6, 8405.

Martin P.S. & Stuart A.J. (1995). Mammoth extinction: two continents and Wrangel Island. *Radiocarbon*, 37, 7-10.

McKinney M.L. & Lockwood J.L. (1999). Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends in Ecology & Evolution*, 14, 450-453.

Naeslund B. & Norberg J. (2006). Ecosystem consequences of the regional species pool. *Oikos*, 115, 504-512.

Newbold T., Hudson L.N., Hill S.L.L., Contu S., Lysenko I., Senior R.A*., et al.* (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, 520, 45-50.

Olden J. D., Poff N. L., Douglas M. R., Douglas M. E. & Fausch K. D. (2004). Ecological and evolutionary consequences of biotic homogenization. Trends in ecology & evolution, 19, 18-24.

Olden J.D., Kennard M.J. & Pusey B.J. (2008). Species invasions and the changing biogeography of Australian freshwater fishes. *Global Ecology and Biogeography*, 17, 25-37.

Petitpierre B., Kueffer C., Broennimann O., Randin C., Daehler C. & Guisan A. (2012). Climatic niche shifts are rare among terrestrial plant invaders. Science, 335, 1344-1348.

Puxeddu M. G., Petti M., Pichiorri F., Cincotti F., Mattia D. & Astolfi L. (2017). Community detection: Comparison among clustering algorithms and application to EEG-based brain networks. *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3965-3968.

Ricklefs R. E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecology letters*, 7, 1-15.

Ricklefs R. E. (2008). Disintegration of the ecological community: American Society of Naturalists Sewall Wright award winner address. *The American Naturalist*, 172, 741-750.

Ricklefs R.E. (2015). Intrinsic dynamics of the regional community. *Ecology letters*, 18, 497-503.

Ripple W.J., Newsome T.M., Wolf C., Dirzo R., Everatt K.T., Galetti M*., et al.* (2015). Collapse of the world's largest herbivores. *Science Advances*, 1.

Rissanen J. (1978). Modeling by shortest data description. *Automatica*, 14, 465-471.

Roberts R.G., Flannery T.F., Ayliffe L.K., Yoshida H., Olley J.M., Prideaux G.J., *et al.* (2001). New ages for the last Australian megafauna: continent-wide extinction about 46,000 years ago. *Science*, 292, 1888-1892.

Rondinini C., Di Marco M., Chiozza F., Santulli G., Baisero D., Visconti P.*, et al.* (2011). Global habitat suitability models of terrestrial mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 2633-2641.

Rosenfeld J. S. (2002). Functional redundancy in ecology and conservation. *Oikos*, 98, 156-162.

Rosvall M., Axelsson D., Bergstrom C. T. (2009). The map equation. *The European Physical Journal Special Topics*, 178, 13-23.

Rosvall M. & Bergstrom C.T. (2008). Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences, 105, 1118-1123.

Rueda M., Rodriguez M.A. & Hawkins B.A. (2013). Identifying global zoogeographical regions: lessons from Wallace. *J. Biogeogr.*, 40, 2215-2225.

Sanmartin I., Enghoff H. & Ronquist F. (2001). Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society*, 73, 345-390.

Schoener T.W. (1974). Resource Partitioning in Ecological Communities. *Science*, 185, 27.

Slater P.L. (1858). On the general Ggographical distribution of the members of the class aves. Journal of the Proceedings of the Linnean Society of London. Zoology, 2, 130-136.

Sidor C.A., Vilhena D.A., Angielczyk K.D., Huttenlocker A.K., Nesbitt S.J., Peecook B.R.*, et al.* (2013). Provincialization of terrestrial faunas following the end-Permian mass extinction. *Proceedings of the National Academy of Sciences*, 110, 8129-8133.

Tilman D., May R.M., Lehman C.L. & Nowak M.A. (1994). Habitat destruction and the extinction debt. *Nature*, 371, 65.

Tucker M.A., Böhning-Gaese K., Fagan W.F., Fryxell J.M., Van Moorter B., Alberts S.C.*, et al.* (2018). Moving in the Anthropocene: Global reductions in terrestrial mammalian movements. *Science*, 359, 466.

Vézquez, D. P., & Gittleman, J. L. (1998). Biodiversity conservation: does phylogeny matter?. *Current Biology*, 8, 379-381.

Vilhena D.A. & Antonelli A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, 6, 6848.

Villéger S., Blanchet S., Beauchard O., Oberdorff T. & Brosse S. (2011). Homogenization patterns of the world's freshwater fish faunas. *Proceedings of the National Academy of Sciences*, 108, 18003-18008.

Wallace A.R. (1876). The geographical distribution of animals. With a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface. Macmillan and Co.,, London, 89-149 pp.

Wallace A.R. (1894). What are zoogeographical regions*. Nature,* 49, 610-613*.*

Winter M., Schweiger O., Klotz S., Nentwig W., Andriopoulos P., Arianoutsou M.*, et al.* (2009). Plant Extinctions and Introductions Lead to Phylogenetic and Taxonomic Homogenization of the European Flora. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 21721-21725.

Zhao T., & Schranz M. E. (2019). Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. *Proceedings of the National Academy of Sciences*, 116, 2165-2174.

Appendix A: Biogeographical network

Bioregions are geographical areas of Earth's surface that share a similar pool of species. We thus need two kinds of data to delineate zooregions: the geographical areas (defined by grid cells), and species (presence/absence in the grid cells). Networks can schematize the distribution of species by using grid cells and species as nodes, and species occurrences in grid cells as links between them (Fig. S1; Thebault 2013; Carstensen *et al*. 2013). Species are not randomly distributed across grid cells, but there are co-occurrence patterns that allow us to identify zooregions. We thus expect that a biogeographical network will have a modular structure, groups of nodes densely connected between them compared to the rest of network nodes. Each module will represent a cluster of grid cells and species highly linked, i.e. what we consider the zooregions and their characteristic species. Community detection algorithms allow us to identify those modules (see Figure S.1, for an example of a biogeographical network and its modular structure). Note that community detection algorithms not provide groups of grid cells representing zooregions, but groups of highly associated grid cells and species.

Appendix B: Infomap

Infomap consists of (i) an objective function that measures the quality of a given network clustering that assigns grid cells and species to modules, and (ii) a search algorithm that optimises the objective function over different clusterings. To explain how Infomap identifies modules of highly interconnected species and grid cells, we demonstrate how it clusters the schematic biogeographical network in Fig. S2.

The objective function of Infomap, which is called the map equation, relies on two ingredients. The first ingredient is a diffusion process on the network that captures interactions beyond nearest neighbours. The map equation models the diffusion process with a random walk that consists of a succession of random steps between grid cells and species (Rosvall & Bergstrom 2008). The random walk starts from a randomly selected grid cell, continues to one of the species present in the grid cell selected at random, and then to a grid cell within the species' range also selected at random, and so on repeatedly (Fig. S2B). If the network has modules of highly interconnected grid cells and species, the walk will persist for relatively long times within those modules. For example, in the schematic network, the illustrated walk persists for several steps among the top four nodes and the bottom four nodes, respectively (Fig. S2C). To capture the general behaviour of the random walk, which we call network flows, rather than a particular realisation, the map equation uses the stationary distribution of the random walk (Rosvall & Bergstrom 2008). It is the long-term visit rates of nodes and links averaged over sufficient time such that they are unaffected by a shift in time (see Rosvall & Bergstrom (2008) and Rosvall *et al.* (2009) for further details).

The second ingredient is based on the minimum description length principle, a central concept in information theory and statistical modelling, which states that the best hypothesis for a given set of data is the one that compresses the data the most by using regularities in the data (Rissanen 1978). The map equation takes advantage of this information-theoretic principle by measuring in bits, the basic unit of information, the smallest amount of information required, on average, to specify which node the random walk will visit next given a network clustering (Rosvall & Bergstrom 2008). The metric used to measure it is known as codelenght (Rosvall & Bergstrom 2008, Rosvall *et al.* 2009). Network clusterings with modules in which the network flows persist for relatively long times in the biogeographical network require little information, and the clustering that minimises the map equation is the one that captures most regularities (Rissanen 1978; Rosvall & Bergstrom 2008; Rosvall *et al.* 2009).

To measure the smallest amount of information required to describe the network flows per step, the map equation exploits the fact that information theory provides the compression limit of a sequence of symbols based on the frequency of the symbols (Rosvall & Bergstrom 2008). The limit measures in bits the minimal possible average codelength per symbol and is called the Shannon entropy, $H(P) = -\sum p_i \log_2 p_i$, where P is the probability distribution of symbols and p_i is the probability of symbol i. The map equation uses symbols to describe when the random walk visits nodes, enters modules and exits modules (see an example with the interactive app in www.mapequation.org/apps/MapDemo.html#applet; Rosvall & Bergstrom 2008). Their corresponding probabilities depend on the visit rates of nodes and the enter and exit rates of modules. In an undirected network, node i's visit rate is the total weight w_i of its links divided by the total link weight of all N nodes in the network, $W = \sum_{i=1}^{N} w_i$. Our biogeographical data form an undirected network because the relationship between grid cells and species is symmetric. That is, a random walk can move along links in both directions, from grid cells to species and from species to grid cells such that each undirected link weight is counted twice in the total link weight W. In addition, because we treat all species occurrences as equally important, our network is unweighted with all link weights set to 1. Therefore, node i's visit rate is its number of links divided by twice the total number of links in the network. In a similar way, the exit and enter rates of a module are the total number of links that crosses its boundary divided by W. To calculate the per-step codelength for a set of m symbols required to describe how the random walk moves within and exits a module or enters a module, respectively, their corresponding weights $w_1, w_2, \ldots, w_m$ are first normalized to probabilities by dividing by their total weight $w = \sum_{j=1}^{m} w_j$.

Because the Shannon entropy of the probabilities gives the per-symbol codelength, to express the per-step codelength it must be weighted by the symbols' total use rate, w/W. Therefore, with a compressed notation, the per-step codelength for m symbols with weights $w_1, w_2, \ldots, w_m$ is $w/W\, H(w_1, w_2, \ldots, w_m) = -w/W \sum_{i=1}^{m} w_i/w \log_2 w_i/w$.

Describing network flows in the schematic network with one module that contains all nodes such that $m = N$ only requires symbols for nodes and not for entering or exiting modules (Fig. S2D). Accordingly, the total weight of all required symbols, $w = 18$, is the same as the total weight of all nodes, $W = 18$, and the per-step codelength L_1 for the one-module solution is

$$\begin{aligned} L_1 &= \frac{w}{W}\left(-\sum_{i=1}^{n} w_i/w \log_2 w_i/w\right) \\ &= \frac{w}{W} H(w_1, w_2 .. w_n) \\ &= \frac{18}{18} H(2,2,2,3,3,2,2,2) = 2.97 \text{ bits,} \end{aligned}$$

where each weight in the last line is the total link weight of each node in the network.

To describe walks on the network with two modules (Fig. S2E), the map equation requires some extra symbols in the sequence. The exit and enter rates of a module are the total weight of all links that crosses its boundary divided by W. Because the map equation assumes that it should be possible to describe any walk exactly without ambiguity, the sequence must contain an extra symbol in each module when the walk exits it, and one symbol for each module for telling which module the walk enters next. With one module for the top four nodes and one for the bottom four nodes in the schematic network (Fig. S2E), one link with weight one crosses each boundary such that all enter and exit rates are $1/18$. To measure the total codelength required to describe network flows with two modules, the map equation adds up one codelength for moving into modules, one for movements in one module, and one for movements in the other module, all weighted by their rate of use

$$L_2 = \frac{10}{18}\underbrace{H(2,2,2,3,1)}_{2.25 \text{ bits}} + \frac{10}{18}\underbrace{H(2,2,2,3,1)}_{2.25 \text{ bits}} + \frac{2}{18}\underbrace{H(1,1)}_{1 \text{ bit}} = 2.61 \text{ bits.}$$

The two-module clustering has shorter codelength because it requires fewer bits to describe movements within the smaller modules despite the extra exit symbol and the cost of switching between them because switching happens so rarely. Compared with the one-module clustering, the two-module clustering better captures regularities in network flows on the biogeographical network (Fig. S2D and E). In fact, this two-module clustering minimises the map equation.

To optimise the map equation over different clusterings, Infomap uses an efficient search algorithm that starts from an initial clustering with long description length and iteratively finds better clusterings. At the start, each node is clustered in its own module. Infomap then iteratively selects a random node and assigns it to the module that decreases the map equation the most. If moving a node does not decrease the map equation, the node remains in the original module. Infomap moves nodes until no movement can decrease the map equation. This procedure forms small modules, see Fig. S2F-J.

Then Infomap rebuilds the network with these small modules defined as super nodes with internal nodes, and, as before, decreases the map equation by moving super nodes into new modules. Infomap repeats this network-rebuilding procedure until the map equation cannot be decreased further. This procedure happens to be sufficient to find the best clustering for the schematic network, see Fig S2H-I. However, when Infomap moves several nodes at once, which occurs when modules have been defined as super nodes, some of them may be assigned to suboptimal modules. Therefore, Infomap can also iteratively and recursively rearrange individual nodes and groups of nodes to refine the solution. See Rosvall & Bergstrom 2008 and the tutorial of Infomap http://www.mapequation.org/assets/publications/mapequationtutorial.pdf for an extensive explanation of how the algorithm works, including the hierarchical search for multiple levels of clusters in clusters.

For a network with n nodes assigned to m modules, there are n^m possible clusterings, which even for moderate-sized networks makes it impossible to test them all and guarantee that Infomap has found the best one. In practice, it is more informative to investigate good clusterings than the very best, because small changes in the network can trigger a new best clustering. Because Infomap treats nodes in random order, multiple runs of Infomap give the same clustering only if it stands out among other solutions and typically give more or less similar solutions with comparable description lengths. Consequently, if Infomap gives different solutions with comparable description lengths, it is not clear how to assign some geographical areas to particular zooregions.

For a given run, Infomap provides the hierarchical classification of modules that minimizes the codelength. In other words, the same search process conducted by Infomap for grouping nodes into modules directly provides the optimal number of hierarchical levels and modules within each level. Thus, although using Infomap or another traditional approach is already a subjective decision, researchers using traditional clustering approaches need to make chose how to define how many hierarchical levels there are. Thus, Infomap reduces that subjectivity of identifying hierarchical classifications of zooregions.

To assess the significant difference between various scenarios, we compared the detection probability of each solution in a scenario.

Appendix C: Alternative zooregion delineations

In this section we explain (i) why we prefer to use all alternative zooregion delineations instead of only the one with the lowest description length, and (ii) how we detect all alternative zooregion delineations.

Benefits of using all alternative zooregions

Exploring all alternative zooregion delineations offers more information on the system as a whole for different reasons. First, it informs about uncertainty when delineating zooregions. If we perform multiple analyses and obtain always the same zooregion delineation we could consider there is no uncertainty when identifying zooregions. However, if we detect multiple zooregion delineations we must acknowledge the uncertainty. Our results for minor zooregions provide examples of both cases. Second, we can identify the geographical areas where there is uncertainty when assigning to zooregions. To do that, we can compare two alternative zooregion delineations and search which geographical areas vary in their zooregion assignation. For example, a region that is assigned sometimes to one zooregion and sometimes to another zooregion can indicate that it is a transitional area (Fig. S3). Third, we can avoid underestimations or overestimations of the effects triggered by introductions and extinctions. Imagine a hypothetical native scenario with two alternative zooregion delineations A and B, whose detection probabilities are 51% and 49%, respectively, and an introduced scenario with the same solutions but their probabilities changed to A=49% and B=51%. If we assess the effects of introductions exclusively by using the zooregion delineation with the highest detection probability, we would say that introductions changes zooregion configuration from A to B. However, when we assess the detection probabilities of all zooregion delineations, we find that the effect of introductions is very small. Similarly, we could also overestimate the effects of introductions if we only use the zooregion delineation with the lowest code length as the representation of each scenario. Imagine that the zooregion delineation A in the native scenario has the lowest code length, while the zooregion delineation B has the lowest code length in the introduced scenario. Thus, assessing and comparing the detection probabilities of all alternative zooregion delineations informs about the effects of introductions and extinctions more accurately.

Identification of alternative zooregion delineations

It is impossible to know a priori the number of alternative zooregion delineations per scenario and their stable detection probabilities. We consider accurate those probabilities that do not vary as we increase the number of analyses. For example, we know that the probabilities of heads or tails when flipping a coin are 50%, however, if we calculate the probability with a few trials we can get wrong estimates of the probabilities. Thus, we developed an approach to calculate how many analyses we need to obtain all alternative zooregion delineations and their detection probabilities. In particular, our approach is based on performing an increasing number of analyses until the detection probability of the alternative zooregion delineations does not change significantly. The conceptual approach is similar to the cumulative species curves for detecting species richness in a study area but also considering detection probabilities.

For each scenario and vertebrate group, we started performing 100 analyses using Infomap (Fig. S4). We used the presence/absence matrix as Infomap input, where the grid cells and species were nodes and species occurrences in grid cells were their links. The output of these 100 analyses was 100 hierarchical zooregion delineations. We extracted the assignation of each grid cell to major and minor zooregions, i.e. highest and lowest hierarchical levels, obtaining 100 delineations of the major and minor zooregions (Fig. S4). Zooregion delineations from different analyses may only differ by a small number of grid cells, reflecting changes in a few boundaries grid cells which are of little interest for our goals. We therefore consider that alternative zooregion delineations are those that differed by at least 10% of the Earth's surface (see below). To detect the alternative zooregion delineations, we calculated the similarity between all pairs of zooregion delineations (see Appendix D for details about the similarity index) and grouped them

by using their similarity. To group zooregions delineations, we used Infomap for a second time. In particular, for each scenario, we conducted new analyses with Infomap considering each zooregion delineation as a node and their similarity values as links. To ensure that all zooregion delineations assigned to the same group were similar, we transformed all similarity values below 0.9 to 0. Consequently, only the zooregion delineations whose similarity was higher or equal than 90% where linked. The output of this second analysis were clusters of zooregion delineations that represent the alternative delineations and their relative detection probability. To avoid misunderstandings, we call each cluster of alternative zooregions delineations an alternative solution. Later, we performed 100 new analyses and repeated this grouping process but using the 200 analyses together, i.e. 100 new and old analyses. Finally, we assessed if the detection probabilities of the alternative solutions varied significantly when using 100 vs. 200 analyses. To do that, we performed a Fisher's exact test with p-values estimated from 10,000 Marko Carlo simulations (function fisher.test of the R package stats; R Core Team 2017). We considered that p-values close to one indicated that the solutions had similar relative occurrence, and therefore that we detected all supported solutions and their stable detection probabilities. If p-values were far from 1, we performed 100 new analyses. Overall, for each scenario and vertebrate group we needed 1,500 analyses to describe all alternative zooregion delineations and their detection probabilities (Fig. S4; R codes in Appendix F).

Although we show in the main text the results considering two zooregion delineations as alternative solutions when they differed in at least 10% of the Earth's surface, we obtained similar results by using 5% and 15% as thresholds (Table S2).

Appendix D: Similarity index

We measured the similarities between pairs of zooregion delineations using a generalised version of the Jaccard index. Specifically, we calculated the index between all pairs of zooregions C_i^A, C_j^B in two delineations, C^A and C^B, and weighted them by the fraction of all grid cells *n* in the intersection. This gives the weighted Jaccard partition index (equation 1)

$$J(C^A, C^B) = \sum_{i,j} \frac{|C_i^A \cap C_j^B|}{n} \frac{|C_i^A \cap C_j^B|}{|C_i^A \cup C_j^B|}. \quad (4)$$

We preferred this index over others due to its meaning. As we only compared the assignation of grid cells in different delineations, this index roughly represents the proportion of overlap between the zooregions in two delineations. In other words, it approximately measures the amount of Earth's surface classified under the same zooregion in two delineations, providing a similarity index which is biogeographically meaningful. This weighted Jaccard index equals 1 when two zooregion configurations are the same, and tends to 0 as they become different. We used the similarity values to group highly similar delineations, so that each group of delineations represents alternative supported solutions.

Appendix E: Null model

We checked the quality of our solutions by assessing that the complexity of the co-occurrence patterns (i.e. zooregions) could not be obtained by chance. To do that, we calculated a compressibility metric for all our delineations. The compressibility metric is simply calculated with the codelength of the network considering that all nodes are modules, the first step-search of Infomap (Fig. S2F), divided by the codelenght of Infomap output, the clustering with the lowest codelength (Fig. S2L). This division allows comparing network with different sizes; note we have two scenarios with less number of nodes because of the simulated extinctions. Note that Infomap output provides both codelengths in the output file. In particular, we compared the compressibility metric of our solutions to the ones of solutions based on randomised null scenarios. If our detected solutions were trivial, we would expect their compressibility metric to be equal or lower to the ones of the solutions based on randomised null scenarios. To make this comparison, we delineated 100 null databases for each scenario and vertebrate group (see below) and ran Infomap for each of these null sets. This resulted in 400 "null" zooregion delineations per vertebrate group with their corresponding compressibility metrics. The results showed that all real solutions had a lower compressibility metric than those of the null solutions, suggesting that our zooregion configurations are non-trivial.

Regarding the creation of null delineations or models, many studies have proposed approaches to assess whether observed patterns can be explained by chance alone (Gotelli 2000; Strona *et al*. 2014). In these approaches, most of the null models randomise species occurrence by fixing the total species occurrence and the total species richness per grid cell (Gotelli 2000; Strona *et al*. 2014). However, these null models do not consider the spatial autocorrelation of species occurrences. Thus, occurrences are spread around the study system, in our case the Earth's surface. The use of these null models would always suggest that our study systems are not random; however, those may be trivial null models for our goals. A potential solution is using null models that consider the spatial autocorrelation of the occurrences of the species. Recently, Nunes & Pearson (2016) proposed a null model that translocate and rotate the distribution of species. Although this null model does consider the spatial autocorrelation of species occurrences, it is limited to specific distribution ranges because of spatial limitations. For example, a species widespread across Eurasia would have a distribution range with the shape of Eurasia. Because no other geographical area may allow allocating a distribution with this size and shape, the null model proposed by Nunes & Pearson (2016) is useful to our purposes as we could not randomise the distributions of many widespread species. To overcome this limitation, we developed a biogeographic null model that does consider the spatial autocorrelation of all species occurrences but allows altering the shape of the species distribution.

To create null datasets, we allocated the distribution ranges of species one by one. We first selected one species randomly. Second, we selected one of its occurrences and allocate it to a random grid cell (hereafter starting grid cell). To consider the spatial autocorrelation of the occurrences, we allocated the rest of the occurrences in the remainder grid cells as a function of the geographical distance between all grid cells and the starting grid cell. In particular, we provided a linear relationship between the geographical distance of the grid cells with the starting grid cell and their probability of storing a new occurrence. We defined the probability of a grid cell to store a new occurrence as: $1 - (D_{g_i} / \sum_{i=1}^{N} D_i)$, where D_{g_i} is the geographical distance between the focal grid with the starting grid cell, D_i the geographical distance between the grid cell *i*. Thus, our null model fixes the number of the species occurrences and considers their spatial autocorrelation. Like the approach proposed by Nunes & Pearson (2016), our approach does not maintain the observed species richness of the grid cells. Nevertheless, our approach approximates the observed species richness in all grid cells. To do that, we calculated for each grid cell a probability of becoming a starting cell based on its observed proportional species richness: $S_f / \sum_{i=1}^{T} Si$, where S_f was the observed species richness of the focal grid cell, S_i the observed species

richness of each grid cell, and T is the total number of grid cells of our study system. We calculated this proportional richness before allocating any species and used it to weight the probability of each cell being selected as the first starting grid cell. To consider the filling of grid cells when we introduce species one by one, after each species is allocated, we recalculated the proportional richness of each grid cell. In that way, we can approximate species richness of each grid cell in our null delineations to the real richness. To do that, we subtracted one value per each species stored in a grid cell to their original observed species richness. This allowed us to generate a model that, while limited in its ability to capture true richness, incorporated the spatial properties that are key for biogeographical zonation and thus, was most useful for our goal.

Appendix F: Used code

This appendix shows the code used in each step of our methodological approach.

(1) Create the Infomap input

Here we explain how to create an Infomap input from a presence/absence data by using R. Infomap accepts different input formats for running analyses. However, here we only show the one we used. Please see the web www.mapequation.org for other formats. We used a format with four elements: (i) a line indicating the number of nodes in the network, i.e. number of grid cells and species; (ii) a two-column table that relates species and grid cells names with numeric ids. The names should be consecutive and start in 1 (see below); (iii) a line indicating the number of links in the network, i.e. the total of species occurrences in the data; and (iv) a two-column table informing about species occurrences in grid cells but using the numeric ids.

We started reading a table whose rows represent a species occurrence in a grid cell, for example:

```
##   Species Grid_cells
## 1    Sp_1      gc_1
## 2    Sp_1      gc_2
## 3    Sp_1      gc_3
## 4    Sp_2      gc_1
## 5    Sp_2      gc_3
```

We created a list of species names and another list of grid cells names

```
list_species <- as.character(as.unique(db$Species))
list_grid_cells <- as.character(as.unique(db$Grid_cells))
```

Later, we renamed species and grid cells names with consecutive numbers starting from 1.

```
db_list_names <- data.frame(Nodes = c(list_species, list_grid_cells))
db_list_names$Numeric_names <- 1:nrow(db_list_names)
```

The number of rows of the new data.frame is the number of nodes in the network. The rows of this new table represent the relationship between the old and new names of the nodes. The data informing about the links between nodes need the numeric names. Thus, we created a table with those numeric names.

```
merge_grid_cells <- merge (db, db_list_names, by.x="Grid_cells", by.y="Nodes", all.x=T)
colnames(merge_grid_cells)[colnames(merge_grid_cells)=="Numeric_names"] <- "Grid_cells_num"
merge_species <- merge (merge_grid_cells, db_list_names, by.x="Species", by.y="Nodes", all.x=T)
colnames(merge_species)[colnames(merge_species)=="Numeric_names"] <- "Species_num"
db_links <- merge_species[,c("Species_num", "Grid_cells_num")]
```

Finally, we created a file that is the input of Infomap.

```
pre_input <- list(cbind("*Vertices",length (nrow(db_list_names))), db_links, cbind("*Edges",
nrow(db_links)), db_links)
path <- c("Directory")
file<-"Infomap_input.net" # Name with extension ".net"
write.table(pre_input[[1]], paste(path,file,sep=""), row.names=F,col.names=F,quote=F )
write.table(pre_input[[2]], paste(path,file,sep=""), append= T,row.names=F,col.names=F,quote=c(2))
write.table(pre_input[[3]], paste(path,file,sep=""), append= T,row.names=F,col.names=F,quote=F )
write.table(pre_input[[4]], paste(path,file,sep=""), append= T,row.names=F,col.names=F,quote=F)
```

(2) Run Infomap to obtain zooregion delineations

We downloaded the code for running Infomap from www.mapequation.org. We ran our analyses in Ubuntu. Here we use the previous input representing species ocurrence data to obtain zooregion delineations. To perform 1,500 analyses, we used the next code:

```
for i in {1..1500}; do ./Infomap -s$i -N100 Infomap_input.net output_folder; mv output_folder/output_file.tree output_folder/output_file_$i.tree; done
```

Now we will explain by parts. First:

```
for i in {1..1500}
```

This part of the code indicates that we want to repeat the analyses 1,500 times. Second:

```
do ./Infomap -s$i -N100 Infomap_input.net output_folder;
```

We indicate what we want to do in each iteration of the loop. In particular, we called Infomap and indicated the seed of the analyses 's$i'. This command ensures that Infomap starts searching a zooregion delineation from a different starting point. Because of we indicated the seed, we can replicate our results. For each of the 1,500 iterations, we indicated that Infomap should perform 100 runs and provides as output the best one, i.e. '-N100'. We thus performed 150,000 analyses per vertebrate group and scenario but selected only 1,500. Finally, we indicated the file that we want to use as input and the folder where we want to store the result:

```
mv output_folder/output_file.tree output_folder/output_file_$i.tree;
```

We change the name of the output. Because of Infomap call each output with the name of the input but with and extension '.tree' by default, when we create a loop we can overwrite the outputs. For that reason, in each iteration we change the output name by using a numeric count. Consequently, we can identify which solution belongs to each run, allowing a complete replication of the analyses. Note we did the same for the seed.

The computing time for generating the 1,500 delineations varied between 7 and 20 days depending on the vertebrate group.

(3) Similarity matrix

Here we indicate the code we used to calculate the similarity between all pairs of zooregion delineations. Note we performed this process twice, once for the major zooregions and another one for the minor zooregions. To calculate the similarity, we need to perform several steps. First, we need to extract the assignation of each grid cell to each hierarchical level. To do that, we provided a function that allows extracting all hierarchical levels from Infomap output. We stored the function in an R object called "Open_Ifnomap_hierarchy.RData", which can be downloaded from our supporting information. The output of our function is a table with the next columns:

1.- ID_row: This is an ID we created to get the table (see the R function)
2.- Node_flow: This is the flow of each node calculated by Infomap (see www.mapequation.org and Rosvall & Bergsrtom 2008 for further details).
3.- real_ID: The real name of the nodes used.
4.- numeric_ID: The numeric ID createdin the previous step.
5.- count_node: This is a count useful to know how many nodes are within each module
6.- Module: The assignation of that node to a module

The rest of the columns represent the nested modules. It is possible that the number of nested modules varied across modules. We can know if a module has no nested modules if it has a 0 in the name.

```
# We provided a function to do it
load("Open_Infomap_hierarchy.RData")
Table_hierarchical_classification <- open_hierarchical_Infomap_output(output_name.tree)
```

Second, we created two datasets: one for major zooregions, and another one for minor zooregions. For each one, we will have a table with two columns: (i) the nodes name, i.e. column called “real_ID” and the name of the module, i.e. column “Module” for the major zooregions, and the last column called “Nested_module_n” for the minor zooregions, where n is the number of nested modules. Third, for each hierarchical level, we create a matrix to calculate the similarity between all pairs of zooregion delineations. To do that, we created a matrix where the first row was the name of the nodes, and the rest of the columns is the assignation of the node to a module in each zooregion delineation, in our case 1,500 zooregion delineations. Because of we only use the grid cells, the matrix will have as many rows as grid cells names, and as many columns as the number of zooregion delineations + 1, since the first column indicates the name of the nodes, hereafter matrix_pre_sim. Thus, we will have a matrix for the major zooregions and another one for the minor zooregions. Finally, we calculated the similarity between all pairs of outputs by using the next code:

```
# Jaccard formula
jaccard_W <- function(p1,p2){
 p1l<-tapply(c(1:length(p1)),p1,list)
 p2l<-tapply(c(1:length(p2)),p2,list)

 cp1<-c()
 for(i in 1:length(p1l)){
  com <- unlist(lapply(p2l,function(x)sum(x%in%p1l[[i]])))
  tot <- unlist(lapply(p2l,function(x)length(unique(c(x,p1l[[i]])))))
  cp1[i]<-sum((com/tot)*(com/length(p1)))
 }

 sum(cp1)
}

# Calculate the total number of combinations
combinaciones_part<-t(combn(c(1:ncol(matrix_pre_sim)),2))
store_sim <- c() # Store similarity
for(i in 1:nrow(combinaciones_part)){
 store_sim <- c(store_sim, jaccard_W(par[,combinaciones_part[i,1]],par[,combinaciones_part[i,2]]))
 print(i)
}
```

We can create the input for clustering the zooregion delineations by their similarity

```
Infomap_input_zooregions <- cbind (combinaciones_part, store_sim)
library(dplyr)
Infomap_input_zooregions <- Infomap_input_zooregions %>%
 filter (store_sim >= 0.90)
```

You can use the code provided in point 1 of this appendix, just by changing “db” by “Infomap_input_zooregions”.

(4)Run Infomap to obtain clusers of similar zooregions

We used the next code in Linux

```
./Infomap -2N100 Infomap_input_zooregions.net output_folder
```

Now we only selected the best partition of 100 runs. Note we included a 2 before the "N". This is to do not calculate a hierarchical modular configuration. Please see details about code in www.mapequation.org

(5)Calcualte the zoogeographical robustness and impact

Please download the R function "robutsness_impact_func.RData" to calculate the zoogeographical robustness and impact from the supporting information. This function need four objects:

1.- The data about the links in the native scenario

2.- The data about the links in the perturbed scenario

3.- Infomap output in the native scenario

4.- Indicate the hierarchical level

(6)Null model

Please download the R function "biogeographical_null_model.RData" to create a null model This function needs two objects:

1.- A table where each row represent a species occurrences in a grid cell. The column of the grid cells should be called "Grid_cells".

2.- A distance matrix with the distance between each pair of grid cells.

The output is a list where each element indicates the distribution range of one species.

Supplementary figures

Figure S1: Example of the processes needed to transform species distribution in grid cells to modules defined by grid cells and species, i.e. zooregions and characteristic species. Plot A represents species distribution of grid cells called using chess terminology (columns letters and rows numbers). Plot B represents the split of grid cells and species in two kinds of information. The links represent species occurrences in those particular grid cells. Plot C represents the likely modules identified by Infomap. Nodes, i.e. grid cells and species, with the same colour belong to the same module, i.e. blue and green modules. In particular, all grid cells with the same colour represent a zooregion and species with the same colour its characteristic species. Blue and green links represent the links between the nodes of the same module. Whereas red links represent the links between nodes from different modules. These red links identify which not highly associated species occupy the zooregion, what we consider the non-characteristic species. Plot D is another version of plot C that allows identifying better zooregions and their characteristic and non-characteristic species.

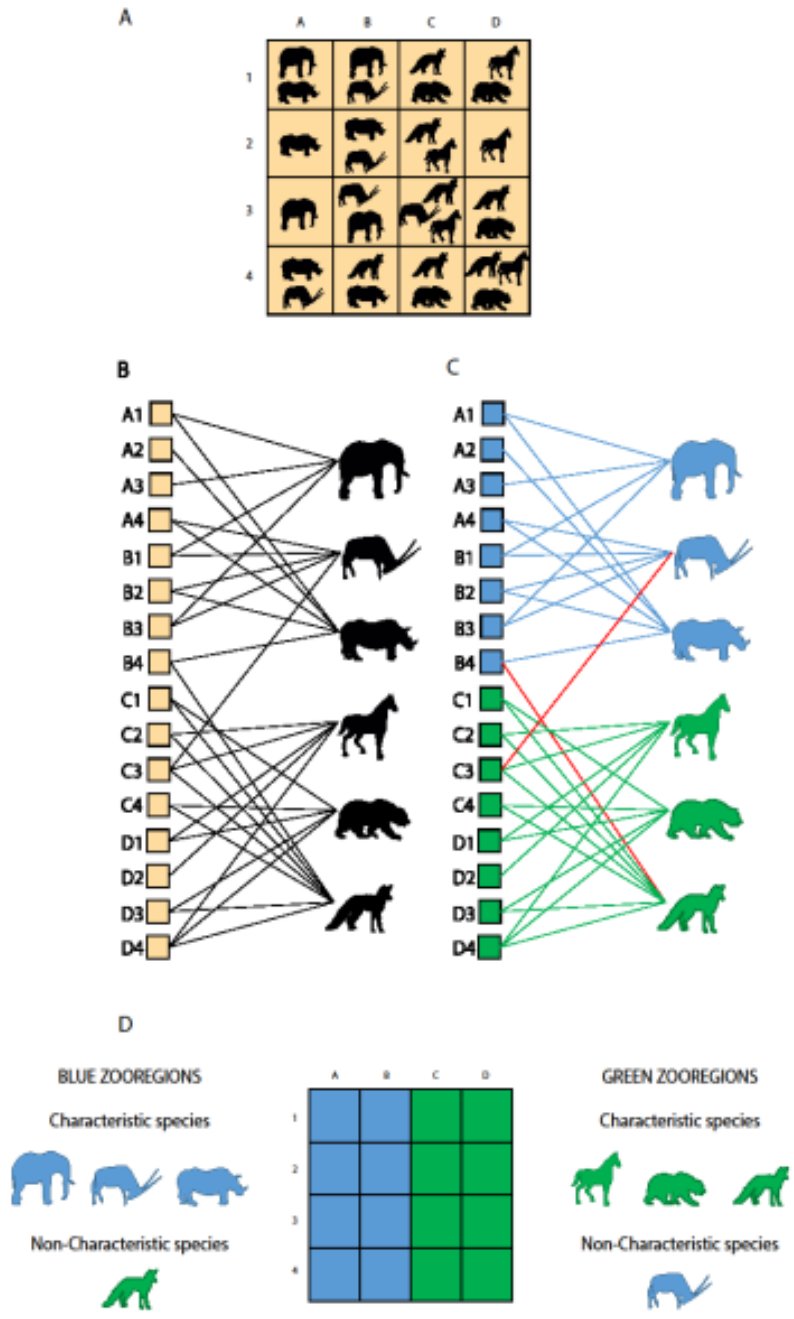


Figure S2: Illustrations are representing how Infomap identifies the modular configuration that best compresses the codelength. (A) Represent a biogeographical network composed by four grid cells and four species (see Appendix A for details about the network). (B) This illustration simulates the travel performed by a random walker along the network: from a grid cell to a random linked species and from this species to a random linked grid cell. (C) This illustration reflects two stretches of the path, i.e. purple and orange colours, where the random walker would spend a relatively long time travelling. For example, when the random walker visits a node marked with orange or purple colour, it will spend a relative long time until it goes to a node with other colour because nodes with the same colour are highly linked. Thus, colours would represent the modules that Infomap wants to detect. (D) This illustration indicates that all nodes are assumed to belong to the same module. The fractions represent the proportional number of links that a node has. The "L" represents the codelenght and indicate the amount of information in bits needed to describe the network. Please see Appendix B for details about the calculation of the codelength. (E) This illustration assumes that there are two modules, one purple and another orange. Note that in addition to the links observed in plot D, there is a new link in the middle "1/18" that represent the link between the modules. In other words, we weighted the links at a node scale, i.e. the lowest hierarchical level of the modular configuration, but also the links at higher hierarchical scales. The "L" represents the new amount of information to describe that modular structure. (F) This illustration simulates the starting modular configuration that uses Infomap for searching the best modular configuration. Note that each node belongs to a different module represented with a different colour. (G-L) These illustrations represent the iterative clustering of nodes and the new amount of information to describe that modular structure.

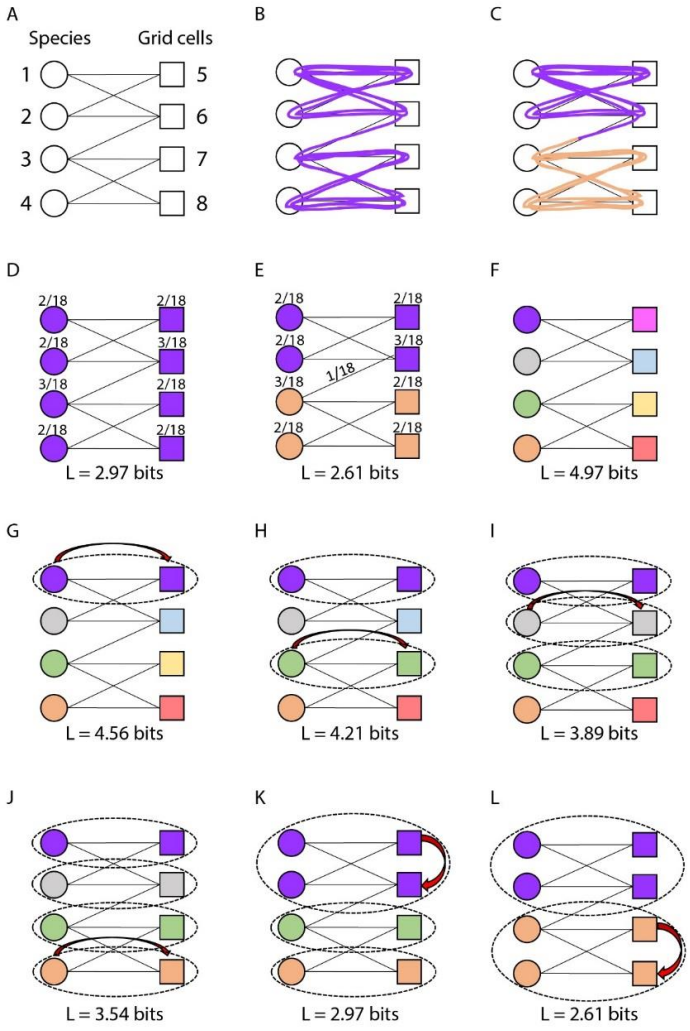


Figure S3: Hypothetical case with a biogeographical uncertainty for assigning boundaries to the zooregions. (A) Representation of the species distribution of two species. (B and C) Representation of the two potential configurations of the delineations based on the same species distribution. This figure reflects the uncertainty of assigning the transition areas of two zooregions.

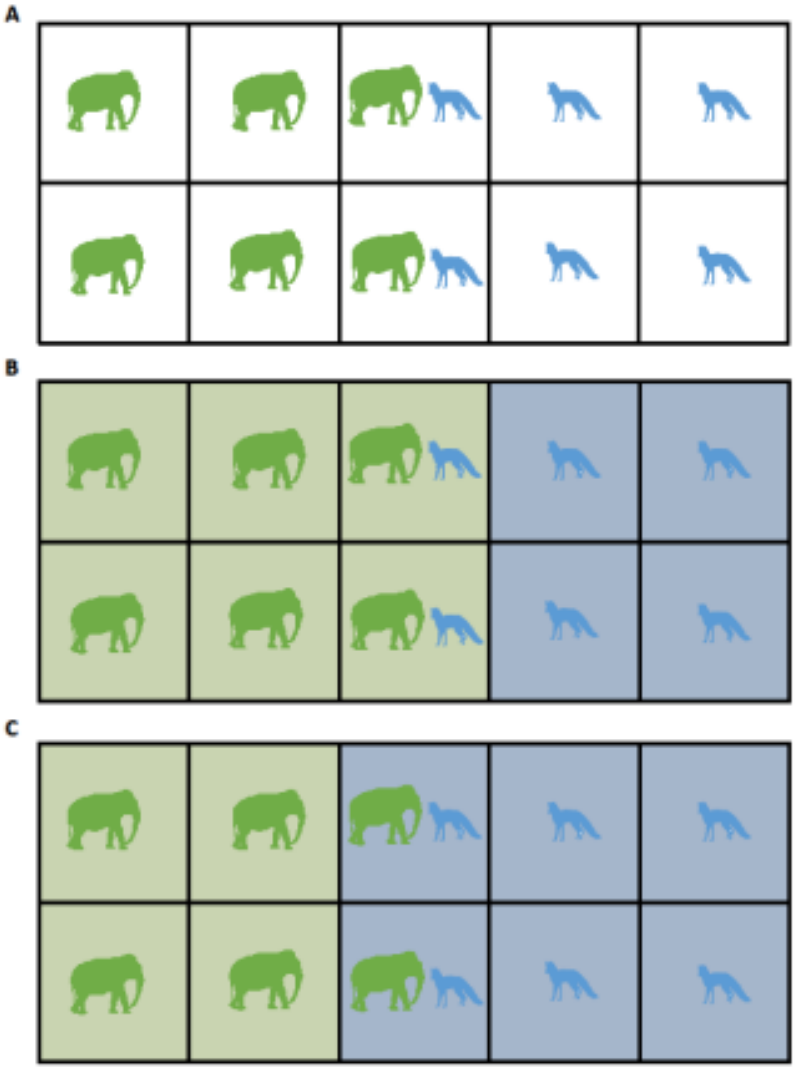


Figure S4: Workflow figure providing a general overview of our methodological steps. (A) We represent the transformation of presence/absence data into a network. On the left we have the grid cells, and on the right the species. The links represent the occurrence of one species in a grid cell. (B) We used Infomap to obtain 100 zooregion delineations of the same dataset. (C) We obtained the assignation of each grid cell to the highest and lowest hierarchical levels, what we call major and minor zooregions. (D) We calculated the similarity between all pairs of delineations by using our weighted Jaccard index (see Appendix D for details). (E) We used Infomap to obtain clusters of similar zooregion delineations by using zooregion delineations as nodes and similarity values as links. (F) We calculated the detection probability of each as alternative solutions, i.e. modules of the previous analyses. (G) We represented the trend of the Fisher's p-value, which assess how much similar are the detection probabilities of the alternative solutions when using a different number of analyses. See code for all these methodological steps in Appendix E.

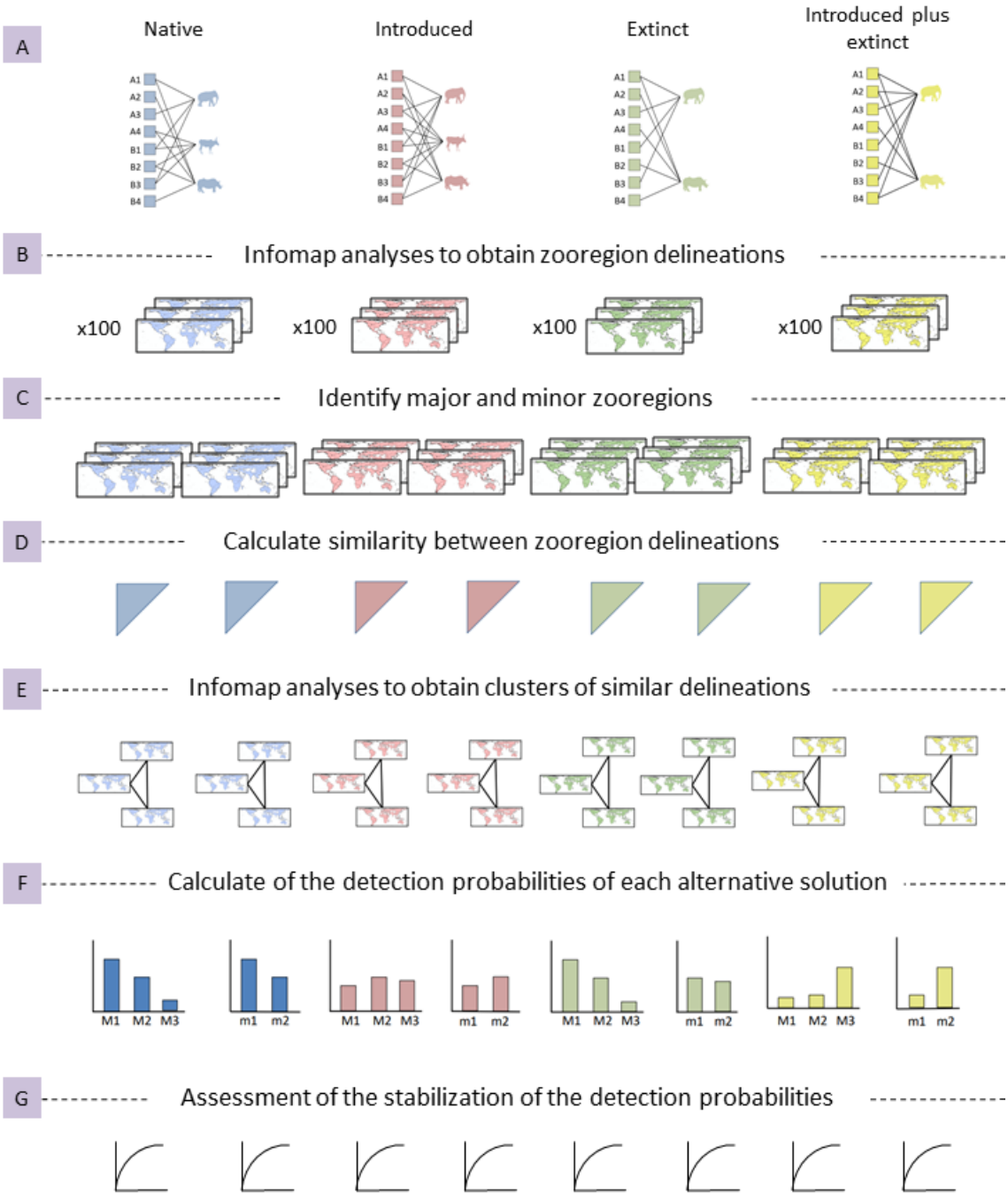


Figure S5: Number of community detection analyses needed to represent all alternative supported solutions. We represented the p-values of the Fisher's exact test (y axis) from comparing two frequencies of occurrence for the solutions of the same scenario but using a different number of delineations (x axis). Solid lines depict the Fisher's p values for the minor zooregions, where dashed depict the ones of the major zooregions. In particular, we compared the frequency of probability of the solutions from a scenario using x number of community detection analyses with the frequency of probability of the solutions from the same scenario using x+100 community detection analyses. Values close to 1 would indicate that frequency of occurrence of the solutions is similar independently of the number of analyses, and that probably new community detection analyses would not provide new supported solutions (see SI Appendix, Appendix C). For our study subjects, we show that 1,500 community detection analyses are enough to identify all alternative supported delineations of zooregions.

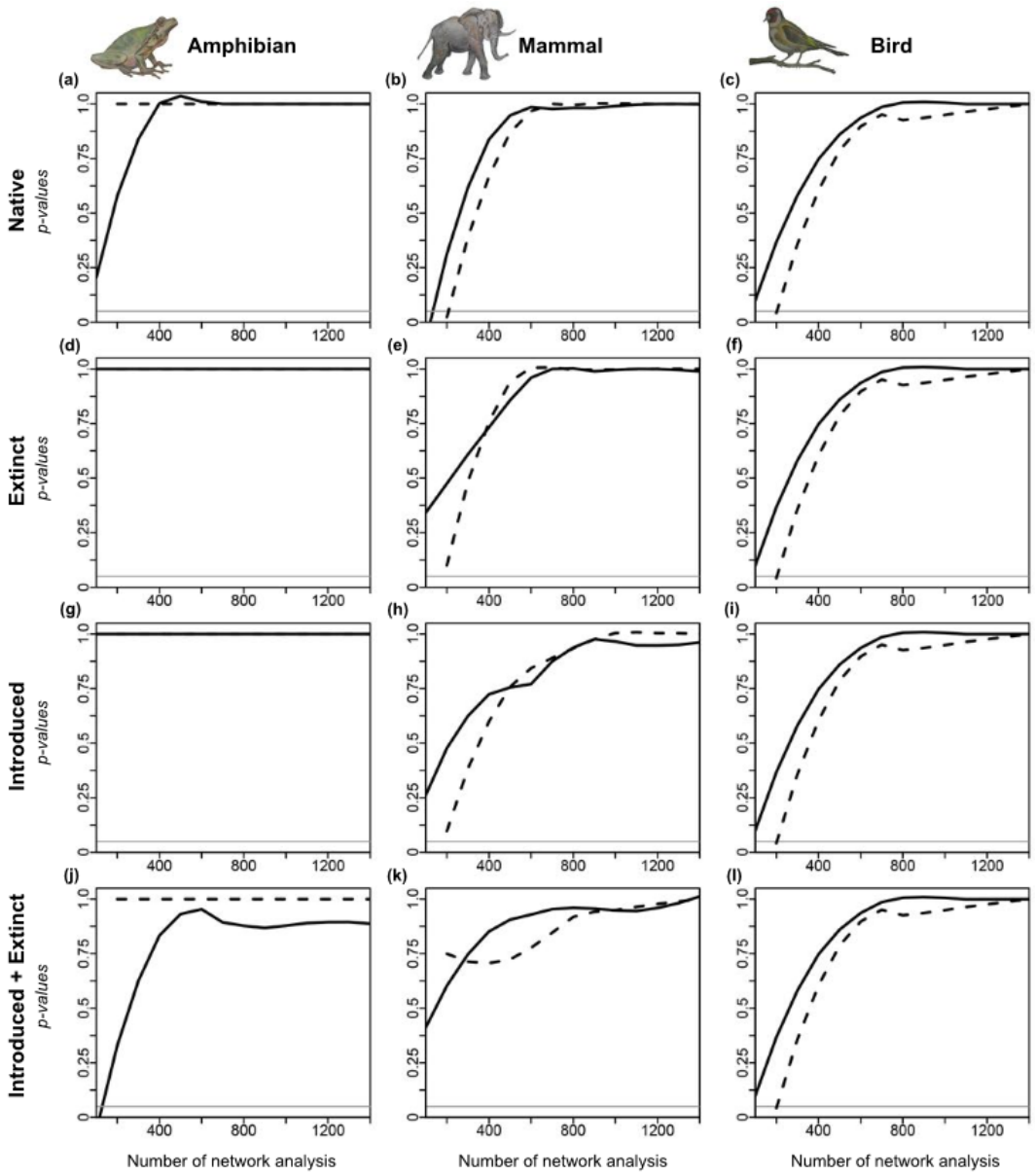


Figure S6: Representation and frequencies of the alternative solutions of the minor zooregions. Zoogeographical delineation of the unique zooregion of the minor zooregions of amphibians and zoogeographical delineation of the two most probable solutions for the minor zooregions of mammals, and birds (labelled as a1 for amphibians, m1-m2 for mammals and b1-b2 for birds), and their detectability probability across scenarios (histograms; see SI Appendix Table S1 and Figs. S10-S12 for details). The histograms also show the joint probability of occurrence of supported but non-main solutions, not shown in maps (m+ for solutions 3-11 of mammals; b+ for solution 3 of birds; see Table S1 for details). See Table S3 for details about the extension of the Earth's surface that differ between solutions.

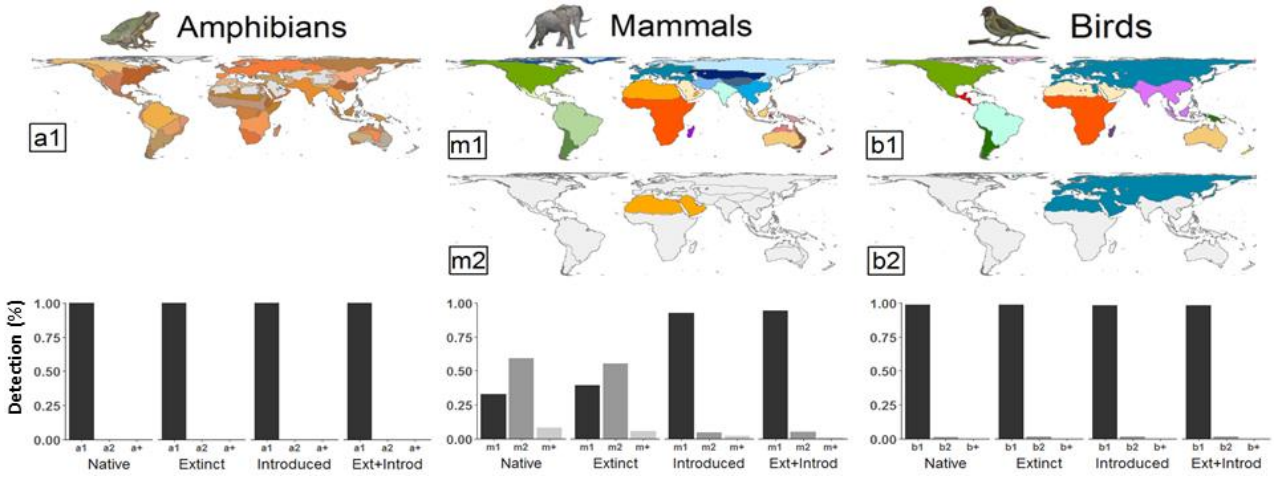


Figure S7: All six solutions for the major zooregions of amphibians. The zooregions represented are the consensus of geographical delineations grouped in each solution. See Table S1 for the detection probability of each solution under each scenario. We called the solutions of the major zooregions of amphibians with the capital letter "A" and a consecutive number. See Table S3 for details about the extension of the Earth's surface that differ between solutions.

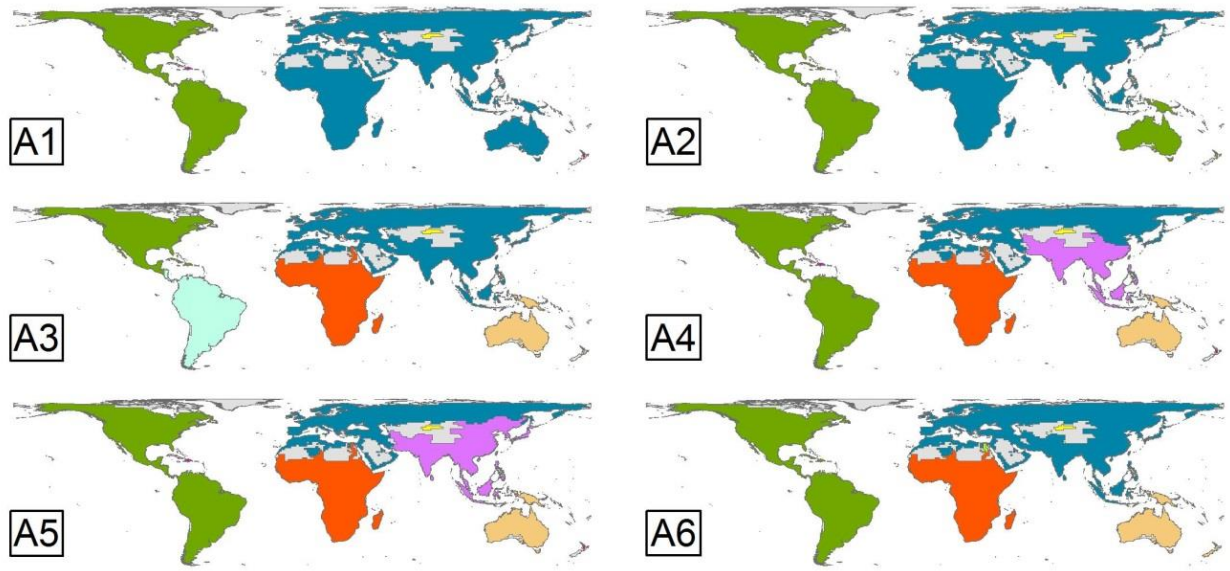


Figure S8: All five solutions for the major zooregions of mammals. The zooregions represented are the consensus of geographical delineations grouped in each solution. See Table S1 for the detection probability of the different solutions. We called the solutions of the major zooregions of mammals with the capital letter "M" and a consecutive number. See Table S3 for details about the extension of the Earth's surface that differ between solutions.

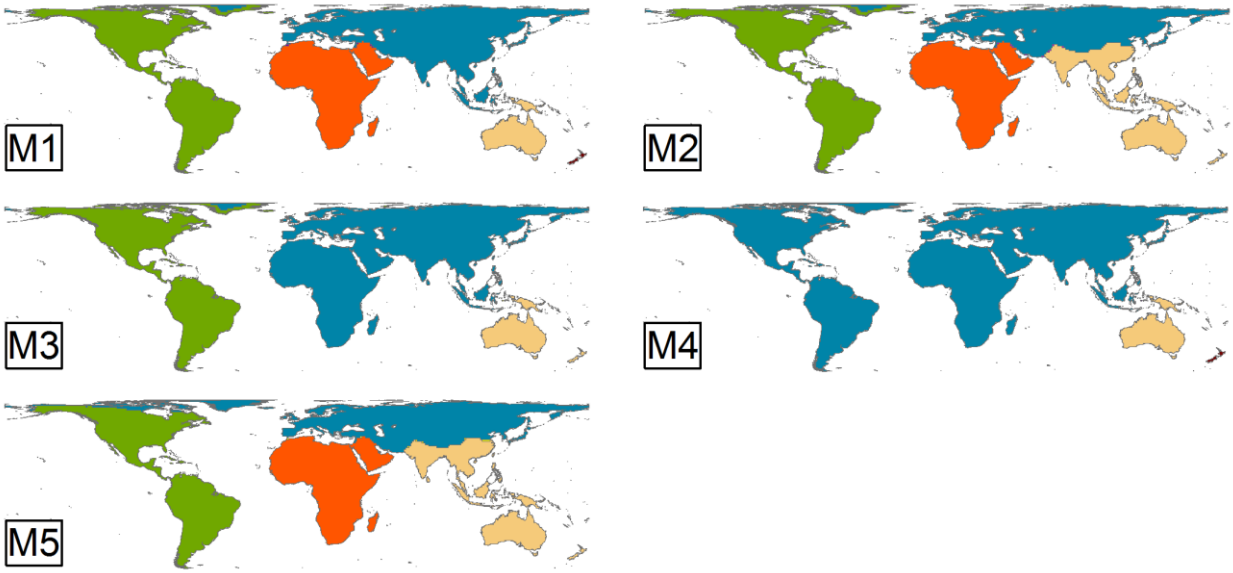


Figure S9: All eight solutions for the major zooregions of birds. The zooregions represented are the consensus of geographical delineations grouped in each solution. See Table S1 for the detection probability of the different solutions. We called the solutions of the major zooregions of birds with the capital letter "B" and a consecutive number. See Table S3 for details about the extension of the Earth's surface that differ between solutions.

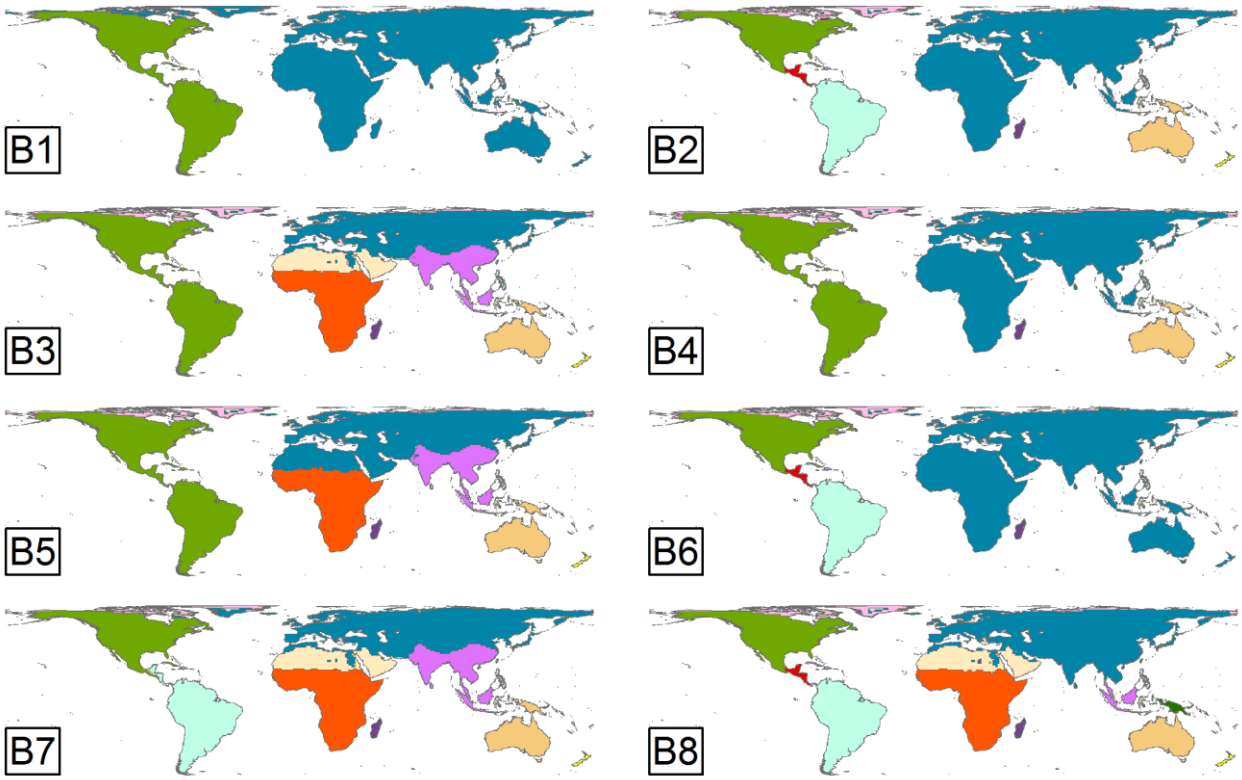


Figure S10: Unique solution for the minor zooregions of amphibians. The zooregions represented are the consensus all delineations. Grey color represents the areas without information about amphibians by the IUCN. See Table S3 for details about the extension of the Earth's surface that differ between solutions.

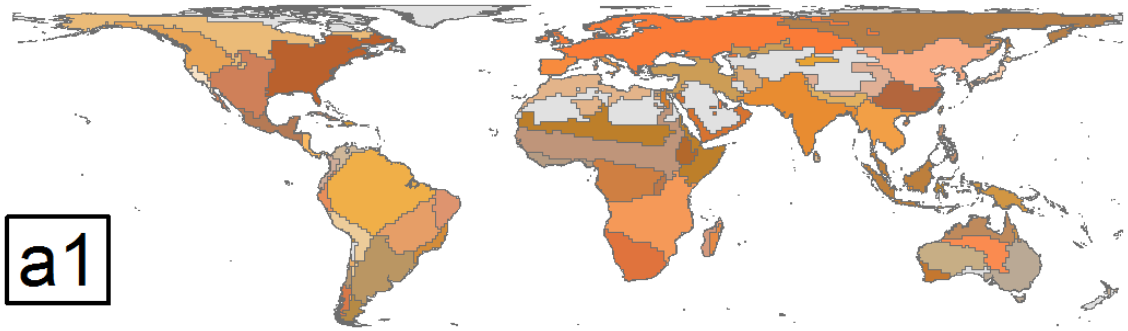


Figure S11: Eleven solutions for the minor zooregions of mammals. The zooregions represented are the consensus of the delineations grouped in each solution. See Table S1 for the detection probability of the different solutions. We called the solutions of the minor zooregions of mammals with the lower letter "m" and a consecutive number. Solution m1 is the most frequent solution in the native scenario and all regions are distinguished by colour. For solutions m2-m11 we used colours only to highlight changes in relationship to m1 (these colours are the same as the general region colour in m1), and used grey colour and line boundaries for zooregions in which no changes were detected compared to m1. See Table S3 for details about the extension of the Earth's surface that differ between solutions.

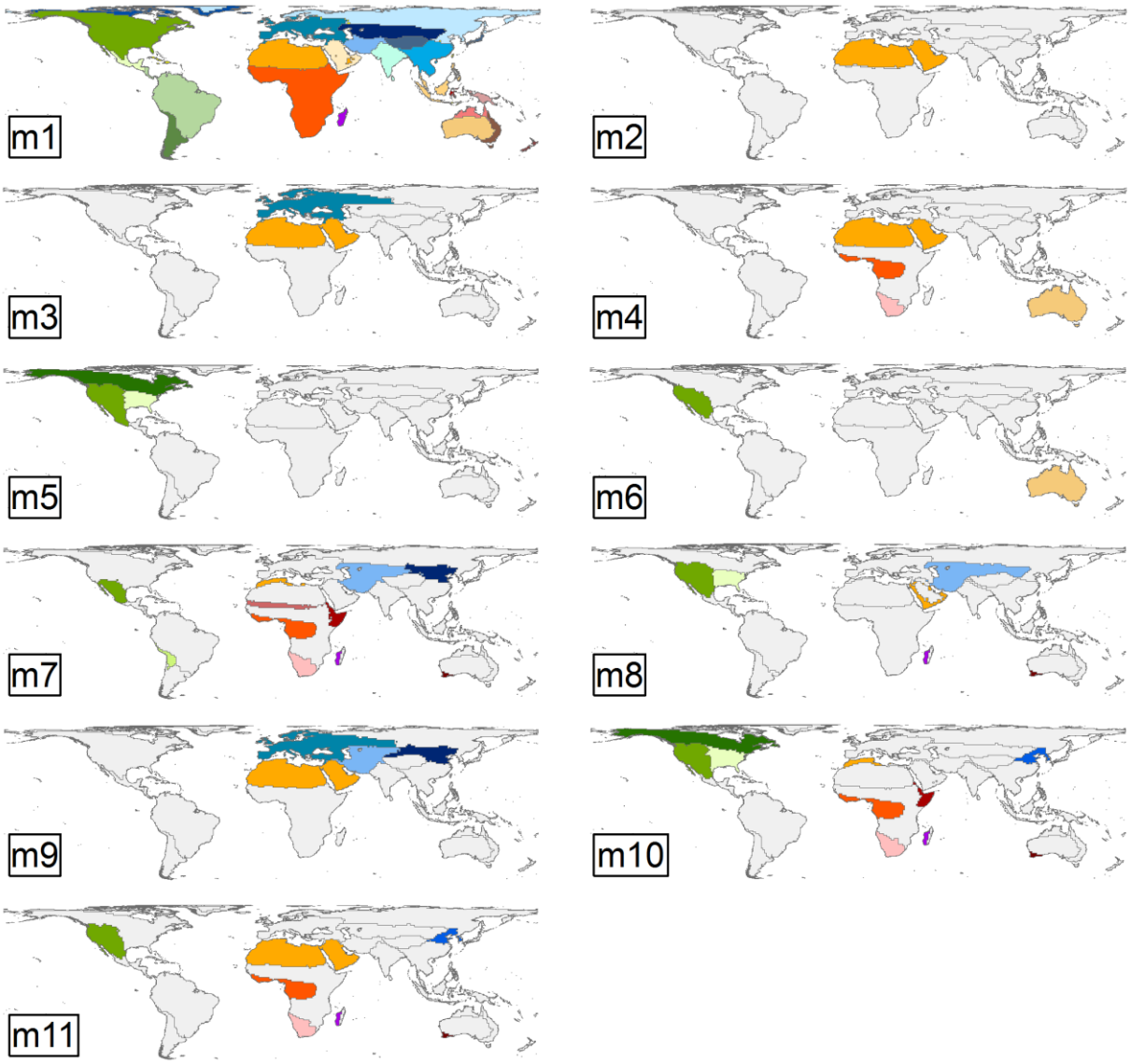


Figure S12: Three solutions for the minor zooregions of birds. The zooregions represented are the consensus of the delineations grouped in each solution. See Table S1 for the detection probability of the different solutions. We called the solutions of the minor zooregions of birds with the lower letter "b" and a consecutive number. The colours of the solutions represent the variations between the solutions with the most frequent solution in the native scenario (b1). The zooregions coloured with grey colour in the solutions b2 and b3 indicate that are like the zooregions of the solution m1 (the most frequent solution in the native scenario). In contrast, the zooregions with non-grey colour in the solutions b2 and b3 are different to the zooregions of the solution b1. See Table S3 for details about the extension of the Earth's surface that differ between solutions.

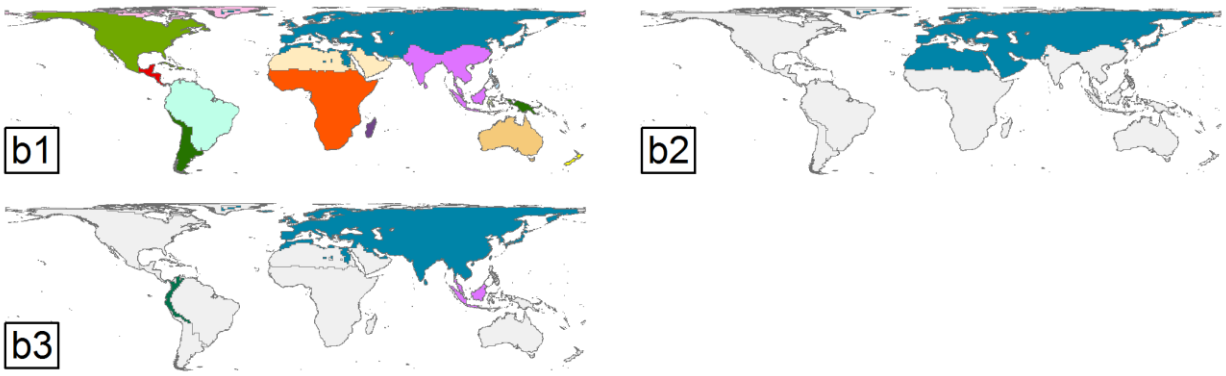


Figure S13: Maps representing robustness and impact for the different scenarios, showing how both vary across regions and vertebrate groups. We represent the (A) robustness and (B) impact values for the grid cells of the minor zooregions, calculated using the most frequent solution for each vertebrate group. (A) Light and dark blue colours indicate low and high relative robustness respectively. (B) Light and dark red colours indicate low and high relative impact respectively.

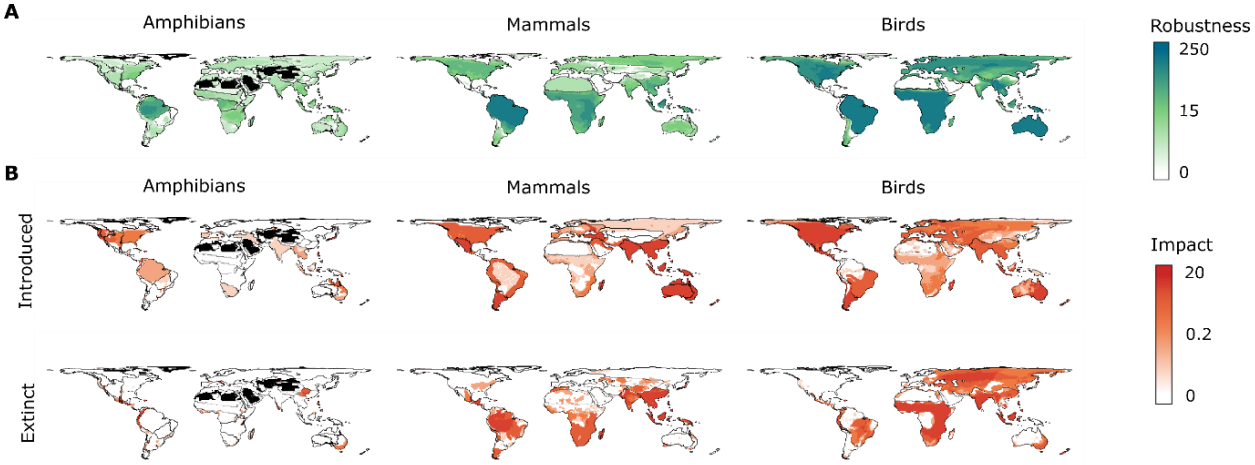
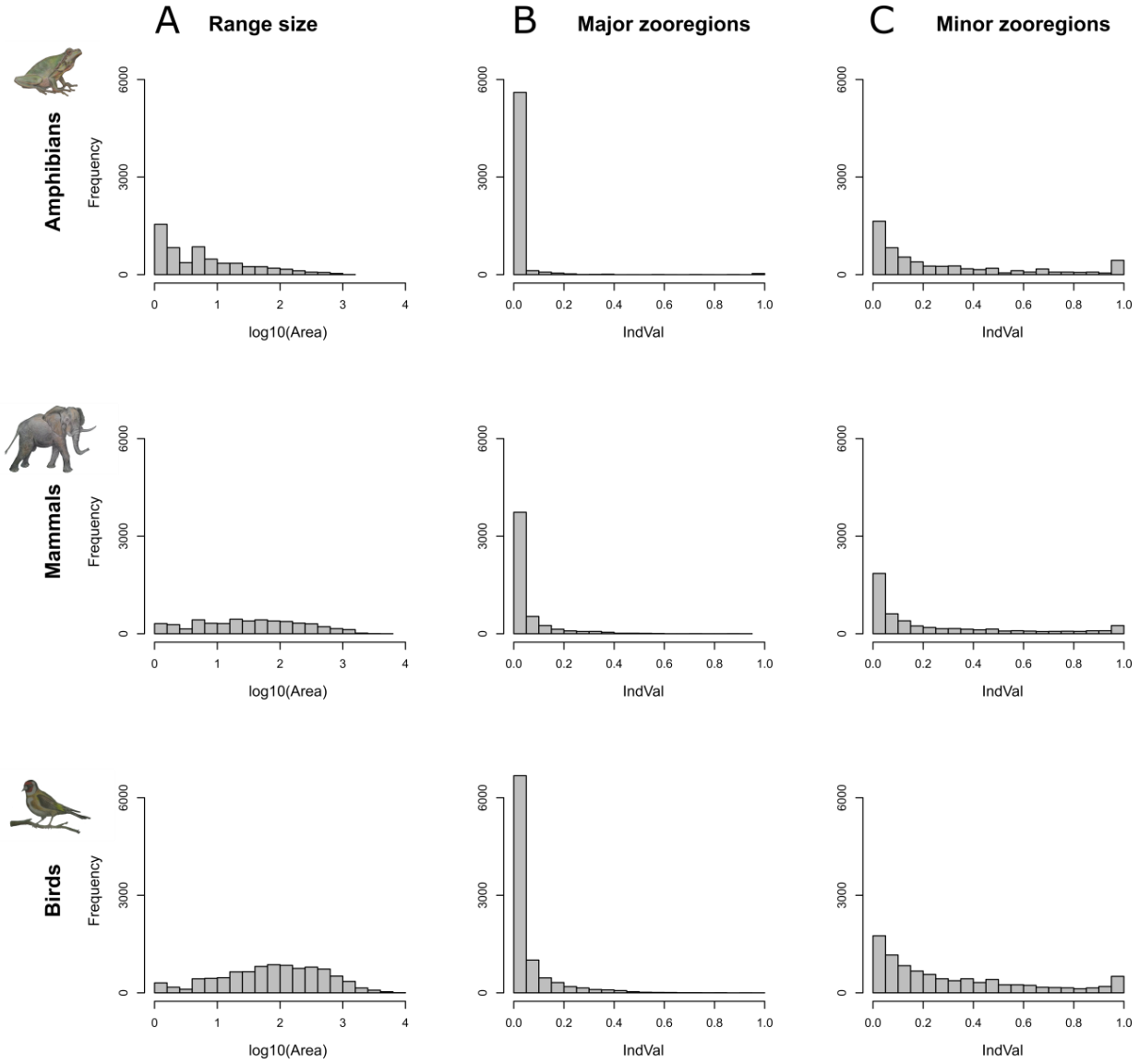


Figure S14: Histograms. (A) Histograms showing the frequency of species with different distribution areas. The axis x represents the distribution areas of the species in a logarithmic scale for amphibians, mammals, and birds. (B) Histogram showing the *IndVal* of the species in the major zooregions. (C) Histogram showing the *IndVal* of the species in the minor zooregions. To calculate the *IndVal*, we used the major and minor zooregions of the most probable solutions in the native scenario.



Supplementary tables

Table S1: Detection probability of the alternative supported solutions of the major and minor zooregions. Obtained from 1,500 independent network analyses for each vertebrate group and scenario. The rows show the probability to detect the alternative solution per scenario. We labelled the different solutions using the first letter of the vertebrate group, and identifying major regions with capitals and minor with lowercase (e.g. A1 = Solution 1 of the major zooregions of amphibians; m9 = Solution 9 of the minor zooregions of mammals). Ext+Introd= Extinction + Introduction scenario..

Amphibian major	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**
Native	0.999	0	0	0.001	0	0
Extinct	0.996	0	0	0.003	0.001	0
Introduced	0	0.999	0.001	0	0	0
Extinct plus Introduced	0	0.673	0.326	0	0	0.001

Mammal major	**M1**	**M2**	**M3**	**M4**	**M5**
Native	0.459	0.358	0.179	0.003	0.001
Extinct	0.408	0.356	0.231	0.004	0.001
Introduced	0.001	0.069	0.928	0	0.002
Extinct plus Introduced	0	0.067	0.932	0	0.001

Bird major	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**
Native	0.459	0.421	0.083	0.023	0.013	0.001	0	0
Extinct	0.405	0.463	0.087	0.027	0.015	0.003	0	0
Introduced	0.298	0.563	0.101	0.017	0.015	0.003	0.003	0
Extinct plus Introduced	0.239	0.624	0.095	0.013	0.017	0.006	0.005	0.001

Amphibian minor	**a1**
Native	1
Extinct	1
Introduced	1
Extinct plus Introduced	1

Mammal minor	**m1**	**m2**	**m3**	**m4**	**m5**	**m6**	**m7**	**m8**	**m9**	**m10**	**m11**
Native	0.592	0.326	0.043	0.021	0.013	0.003	0.001	0.001	0	0	0
Extinct	0.552	0.393	0.035	0.004	0.013	0.001	0	0	0.001	0.001	0
Introduced	0.049	0.928	0.003	0.013	0.005	0	0	0	0.001	0	0.001
Extinct plus Introduced	0.050	0.943	0.001	0.001	0.003	0	0	0	0.002	0	0

Bird minor	**b1**	**b2**	**b3**
Native	0.987	0.013	0
Extinct	0.985	0.015	0
Introduced	0.984	0.016	0
Extinct plus Introduced	0.981	0.018	0.001

Table S2: Differences between the detection probabilities of the solutions across scenarios. The upper and lower triangular matrixes of each vertebrate group show the value of the Schoener's D index of comparing the detection probability of the solutions of the major and minor zooregions respectively. The tables A, B, and C represented the Schoener's D index for a similarity threshold of 90%, 85% and 95% respectively (see Methods). We used the threshold of 90% in our analyses.

A (0.90)

AMPHIBIANS

	Native	**Extinct**	**Introduced**	**Extinct+Introduced**
Native	-	0.998	0	0
Extinct	1	-	0	0
Introduced	1	1	-	0.675
Extinct+Introduced	1	1	1	-

MAMMALS

	Native	**Extinct**	**Introduced**	**Extinct+Introduced**
Native	-	0.947	0.251	0.247
Extinct	0.930	-	0.303	0.299
Introduced	0.397	0.470	-	0.995
Extinct+Introduced	0.379	0.449	0.982	-

BIRDS

	Native	**Extinct**	**Introduced**	**Extinct+Introduced**
Native	-	0.946	0.833	0.769
Extinct	0.997	-	0.883	0.819
Introduced	0.997	0.999	-	0.930
Extinct+Introduced	0.995	0.997	0.998	-

B (0.85)

AMPHIBIANS

	Native	Extinct	Introduced	Extinct+Introduced
Native	-	0.998	0	0
Extinct	1	-	0	0
Introduced	1	1	-	0.675
Extinct+Introduced	1	1	1	-

MAMMALS

	Native	Extinct	Introduced	Extinct+Introduced
Native	-	0.947	0.239	0.247
Extinct	1	-	0.291	0.299
Introduced	1	1	-	0.985
Extinct+Introduced	1	1	1	-

BIRDS

	Native	Extinct	Introduced	Extinct+Introduced
Native	-	0.946	0.833	0.769
Extinct	0.949	-	0.883	0.819
Introduced	0.949	1	-	0.932
Extinct+Introduced	0.949	1	1	-

C (0.95)

AMPHIBIANS

	Native	Extinct	Introduced	Extinct+Introduced
Native	-	0.998	0	0
Extinct	0.209	-	0	0
Introduced	0.801	0.111	-	0.675
Extinct+Introduced	0.275	0.680	0.215	-

MAMMALS

	Native	Extinct	Introduced	Extinct+Introduced
Native	-	0.947	0.233	0.247
Extinct	0.791	-	0.285	0.299
Introduced	0.408	0.541	-	0.979
Extinct+Introduced	0.313	0.460	0.853	-

BIRDS

	Native	Extinct	Introduced	Extinct+Introduced
Native	-	0.944	0.830	0.767
Extinct	0.971	-	0.881	0.818
Introduced	0.955	0.982	-	0.927
Extinct+Introduced	0.932	0.957	0.973	-

Table S3: Proportion of the Earth's surface that is similar between solutions. The upper triangle indicates the mean similarity between all pairs of delineations of two solutions when considering a similarity threshold of 0.90. The lower triangle indicates the ratio between the similarities of delineations across solutions and the similarities of delineations within solutions. Values above one indicate that the inherent variability within solutions is lower than those across solutions; indicating that the grouping of delineations within solutions is biological relevant. We included no table for minor amphibian zooregions because there is only one solution.

Amphibian major

	1	**2**	**3**	**4**	**5**	**6**
1	-	0.81	0.47	0.53	0.52	0.63
2	1.24	-	0.49	0.50	0.48	0.61
3	2.11	2.05	-	0.67	0.65	0.84
4	1.87	1.99	1.48	-	0.90	0.83
5	1.93	2.07	1.54	1.11	-	0.80
6	1.59	1.63	1.19	1.21	1.25	-

Mammal major

	1	**2**	**3**	**4**	**5**
1	-	0.70	0.59	0.53	0.58
2	1.43	-	0.83	0.75	0.39
3	1.69	1.19	-	0.86	0.30
4	1.82	1.28	1.12	-	0.31
5	1.71	2.59	3.31	3.10	-

Bird major

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	-	0.64	0.43	0.82	0.50	0.86	0.54	0.63
2	1.53	-	0.45	0.77	0.51	0.74	0.32	0.36
3	2.23	2.16	-	0.57	0.5	0.36	0.81	0.71
4	1.19	1.28	1.70	-	0.65	0.72	0.43	0.50
5	1.92	1.92	1.13	1.52	-	0.43	0.69	0.62
6	1.12	1.32	2.66	1.36	2.26	-	0.47	0.55
7	1.77	3.04	1.19	2.27	1.39	2.03	-	0.84
8	1.54	2.73	135	1.98	1.57	1.75	1.13	-

Mammal minor

	1	2	3	4	5	6	7	8	9	10	11
1	-	0.87	0.83	0.83	0.81	0.85	0.81	0.70	0.81	0.77	0.79
2	1.06	-	0.86	0.82	0.81	0.84	0.79	0.75	0.76	0.75	0.79
3	1.11	1.07	-	0.77	0.76	0.79	0.83	0.69	0.71	0.70	0.73
4	1.09	1.09	1.17	-	0.74	0.78	0.75	0.79	0.83	0.83	0.72
5	1.12	1.11	1.19	1.20	-	0.84	0.71	0.82	0.79	0.68	0.86
6	1.08	1.08	1.16	1.16	1.06	-	0.75	0.76	0.83	0.72	0.85
7	1.10	1.11	1.07	1.16	1.21	1.18	-	0.65	0.72	0.75	0.73
8	1.32	1.22	1.32	1.12	1.08	1.20	1.32	-	0.85	0.74	0.77
9	1.15	1.21	1.29	1.07	1.12	1.10	1.19	1.18	-	0.79	0.75
10	1.20	1.22	1.31	1.08	1.29	1.27	1.14	1.35	1.27	-	0.7
11	1.17	1.16	1.25	1.24	1.02	1.08	1.17	1.30	1.33	1.44	-

Bird minor

	1	2	3
1	-	0.82	0.84
2	1.16	-	0.74
3	1.12	1.29	-

References supplementary

Carstensen D. W., Lessard J. P., Holt B. G., Krabbe Borregaard M., & Rahbek C. (2013). Introducing the biogeographic species pool. *Ecography*, 36, 1310-1318.

Gotelli N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81, 2606-2621.

Nunes L. A. & Pearson R. G. (2017). A null biogeographical test for assessing ecological niche evolution. Journal of biogeography, 44, 1331-1343.

R Core Team (2017) R: A language and environment for statistical computing.

Rissanen J. (1978). Modeling by shortest data description. Automatica, 14, 465-471.

Rosvall M., Axelsson D. & Bergstrom C.T. (2009). The map equation. *The European Physical Journal Special Topics*, 178, 13-23.

Rosvall M. & Bergstrom C.T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105, 1118-1123.

Strona G., Nappo D., Boccacci F., Fattorini S. & San-Miguel-Ayanz J. (2014). A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature communications*, 5, 4114.

Thébault E. (2013). Identifying compartments in presence–absence matrices and bipartite networks: insights into modularity measures. *Journal of Biogeography*, 40, 759-768.

Vilhena D.A. & Antonelli A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, 6, 6848.