Journal of Biogeography (J. Biogeogr.) (2012) 39, 1391–1400

ORIGINAL ARTICLE

Atlas versus range maps: robustness of chorological relationships to distribution data types in European mammals

A. Márcia Barbosa[1,2*], Alba Estrada[3], Ana L. Márquez[4], Andy Purvis[2] and C. David L. Orme[2]

[1]Cátedra "Rui Nabeiro" – Biodiversidade, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO) – Universidade de Évora, 7004-516 Évora, Portugal, [2]Division of Biology, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK, [3]Instituto de Investigación en Recursos Cinegéticos, IREC (CSIC-UCLM-JCCM), Ronda de Toledo s/n, 13071 Ciudad Real, Spain, [4]Grupo de Biogeografía, Diversidad y Conservación, Departamento de Biología Animal, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain

*Correspondence: A. Márcia Barbosa, Cátedra "Rui Nabeiro" – Biodiversidade, CIBIO – Universidade de Évora, 7004-516 Évora, Portugal.
E-mail: barbosa@uevora.pt

ABSTRACT

Aim Chorological relationships describe the patterns of distributional overlap among species. In addition to revealing biogeographical structure, the resulting clusters of species with similar geographical distributions can serve as natural units in conservation planning. Here, we assess the extent to which temporal, methodological and taxonomical differences in the source of species' distribution data can affect the relationships that are found.

Location Western Europe.

Methods We used two data sets – the *Atlas of European mammals* and polygon range maps from the IUCN Global Mammal Assessment – both as presence–absence data for UTM 50 km × 50 km squares. We performed pairwise comparisons among 156 species for each data set to build matrices of the similarity in distribution across species, using both Jaccard's and Baroni-Urbani & Buser's indices. We then compared these similarity matrices (chorological relationships), as well as the species richness and occurrence patterns from the two data sets.

Results As expected, range maps increased both the mean prevalence per species and mean species richness per grid cell in comparison to atlas data, reflecting the general view that these data types respectively over- and underestimate species occurrence. However, species richness and occurrence patterns in atlas and range map data were positively associated and, most importantly, the chorological relationships underlying the two data sets were highly similar.

Main conclusions Despite many methodological, temporal and taxonomical differences between atlas data and range maps, the chorological relationships encountered between species were similar for both data sets. Chorological analyses can thus be robust to the data source used and provide a solid basis for analytical biogeographical studies, even over broad spatial scales.

Keywords
Chorology, conservation biogeography, data mismatch, data quality, distributional relationships, Europe, fuzzy similarity, resolution, scale.

INTRODUCTION

Comparative analyses of the distributions of species can generate profound insights into the processes and environmental factors driving spatial patterns of diversity. Such comparisons demand that an analytical approach be used in order to summarize distributions objectively (Ball, 1975; Báez *et al.*, 2005). Chorological clustering provides such a framework by detecting statistically robust clusters of species distribution types (chorotypes; Baroni-Urbani *et al.*, 1978). These can aid analyses by identifying common regions across a set of species where a common set of factors may determine the distributions shared by those species, simplifying their biogeographical interpretation (Márquez *et al.*, 1997; Real *et al.*, 2008). Chorotypes are also a good alternative to the analysis of areas of endemism for indicating the occurrence of vicariance events in a given region (Hausdorf & Hennig, 2003).

doi:10.1111/j.1365-2699.2012.02762.x

Journal of Biogeography

As with patterns of species richness (Hurlbert & White, 2005; Hurlbert & Jetz, 2007), the first step in understanding chorological relationships is to accurately describe species' distributions. However, species distribution data can come in various forms that present important differences. To date, chorological relationships have been assessed from a wide range of different distribution data types, including point occurrences (Carmona *et al.*, 1999), species checklists (Báez *et al.*, 2005), survey data gridded at different resolutions (Real *et al.*, 1997; Liébanas *et al.*, 2002), national distribution atlases (Carvalho *et al.*, 2011), and combinations of atlas and range map data gridded to river basins (Real *et al.*, 2008). However, we lack an assessment of how the use of different types of distribution data may affect the chorological relationships inferred for a given species pool.

Over wide spatial extents, species' distributions are most commonly represented in distribution atlases or as range maps. Both data types compile information gathered from multiple sources with uneven surveying effort, including literature information, museum data, records provided by volunteer naturalists and, for some species and regions, specially designed field surveys (Mitchell-Jones *et al.*, 1999; IUCN, 2010). However, there are important differences between these data types (Gaston & Fuller, 2009), which have already been shown to strongly influence the analysis of species richness patterns, the identification of diversity hotspots, and studies on the representativeness and complementarity of biodiversity in protected areas, at least up to certain resolutions (Hurlbert & White, 2005; Hurlbert & Jetz, 2007). The concept of spatial scale includes both resolution (grain) and extent: extent is the overall area encompassed by a study, while resolution refers to the size of the individual units of observation (Wiens, 1989).

Distribution atlases represent species' distributions as observed presences and absences on a regular spatial grid, providing a rough estimate of species' areas of occupancy within the study area (depending on the resolution of the grid and on each species' home range size or dispersal capacity; Gaston & Fuller, 2009). No assumptions are made about the occurrence of a species in any particular grid cell, so there are blank cells even for common species within well-recorded regions, and sometimes large areas where a species almost certainly occurs but is not documented (Mitchell-Jones *et al.*, 1999). As a failure to detect species within a grid cell is recorded as an absence, atlases often underestimate species distributions; this can be compounded if non-surveyed localities are also depicted as absences. Survey effort can strongly affect observed patterns in species' occurrence and richness (Perring & Walters, 1962; Prendergast *et al.*, 1993; Ribas *et al.*, 2007; Barbosa *et al.*, 2010; Kéry *et al.*, 2010), and it might be desirable to exclude undersampled localities from any analysis (Hurlbert & Jetz, 2007; Fontaneto *et al.*, 2012). However, information on survey effort is rarely included in large-scale distribution data sets.

In contrast, range maps tend to overestimate species' distributions. Such maps consist of continuous areas encompassing the species' known presence sites, generally estimating the species' extent of occurrence – that is, the overall geographical spread of the species' presence localities (Gaston & Fuller, 2009). Species do not occur everywhere within their geographical ranges, and internal discontinuities are generally ignored by range maps (Rapoport, 1982; Hurlbert & White, 2005; Gaston & Fuller, 2009). When range maps are converted to presence–absence data on a grid, as is necessary for most analyses, this range porosity can lead to a number of false presences that is proportional to the spatial resolution of the grid. Consequently, while at range boundaries range maps may be at a finer scale than atlas data (depending on the resolution and precision of the map), within the ranges they generally represent species' distributions at a coarser scale. Range maps often lead to local overestimates of species richness (Diniz-Filho *et al.*, 2003; Hurlbert & White, 2005; Hurlbert & Jetz, 2007) and increased spatial autocorrelation in both species' occurrences and species richness (Diniz-Filho *et al.*, 2003; Hurlbert & White, 2005).

Species distribution atlases with wide (e.g. continental) geographical coverage are available for some taxonomic groups and geographical regions. In Europe, atlases are available for vascular plants (Jalas & Suominen, 1972-94), mammals (Mitchell-Jones *et al.*, 1999), breeding birds (Hagemeijer & Blair, 1997), amphibians and reptiles (Gasc *et al.*, 1997), and some invertebrates, such as Lepidoptera, Hymenoptera and Nematoda (Heath & Leclercq, 1981). So far, atlases have been published primarily in a physical (paper) form and have not been frequently updated. Distribution range maps, on the other hand, are becoming widely available for several taxonomic groups and over wide geographical areas, following the global assessments performed by the International Union for Conservation of Nature (IUCN) and BirdLife International. Digitized range maps are currently available for mammals, amphibians, some reptiles, threatened birds, reef-building corals, groupers, wrasses, angelfish, butterflyfish, parrotfish, sea snakes, seagrasses and mangroves (IUCN, 2010). As they are published on the Internet, range maps are updated more frequently than distribution atlases.

Apart from differences in resolution scale and in the effect of survey effort, time lags also cause disparities to arise between data from different sources. These disparities can result from changes in the knowledge of species' distributions, changes in taxonomy (with the reassignment of populations to different species or genera; see Appendix S1 in Supporting Information), and expansions or reductions in species' occurrence areas. For example, the most recent mammal distribution atlas with Europe-wide coverage (Mitchell-Jones *et al.*, 1999) was published over a decade before the latest mammal range maps (IUCN, 2010). Apart from the discrepancies brought about by the false absences and false presences that each method of recording species' distribution produces (e.g. *Glis glis*; Fig. 1a), changes in species' distributions produced noticeable differences between both data sets (e.g. Iberian lynx *Lynx pardinus* and Eurasian otter *Lutra lutra*; Fig. 1b,c). Different criteria for considering particular populations as either wild or domesticated (such as the caribou/reindeer *Rangifer tarandus* in Iceland; Fig. 1d) also generated perceptible differences between the two data sets.

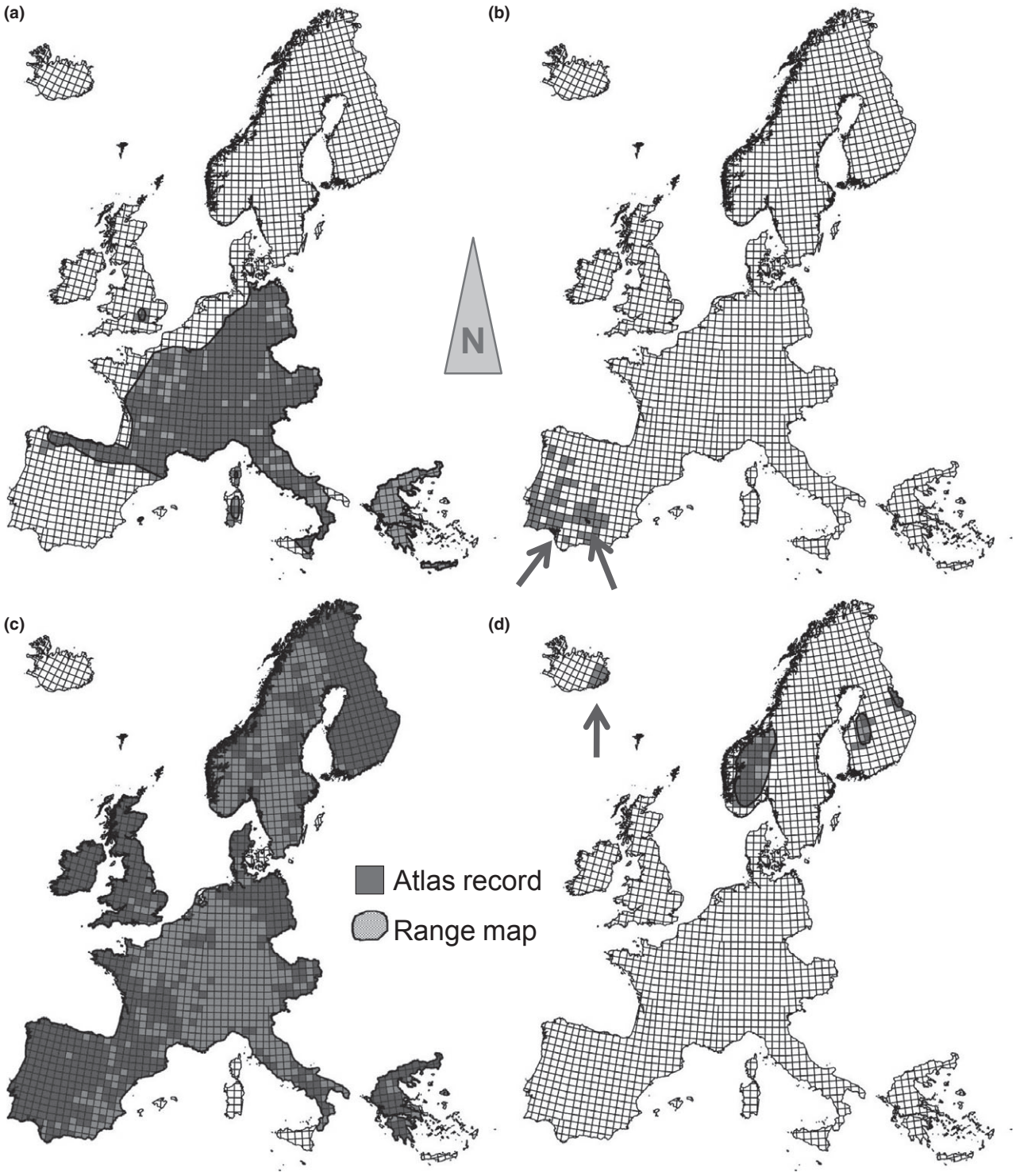

Figure 1 Western Europe under a UTM 50 km × 50 km grid (maps in Lambert Equal Area projection, datum ETRS), with four species' distributions according to the *Atlas of European mammals* (Mitchell-Jones *et al.*, 1999) and recent distribution range maps from the Global Mammal Assessment (IUCN, 2010). (a) *Glis glis*: range overestimation versus atlas underestimation of occurrence areas; (b) *Lynx pardinus*: distribution visibly contracted in recent years, and thus current range is narrower than atlas records; (c) *Lutra lutra*: distribution visibly expanded in recent years, and thus current range is wider than was recorded in the atlas; (d) *Rangifer tarandus*: Iceland population considered semi-domesticated by IUCN (2010), and thus not included in the range map albeit being included in the atlas.

In addition to affecting the biogeographical interpretation of diversity, differences between distribution data sets also have practical implications for the use of large-scale spatial analyses in systematic conservation planning. By allowing the identification of robust and coherent units of species distributions, the assessment of chorological relationships may be an important component of such studies, provided that the type of data used for analysis does not bias the results. In this study, we compared chorological relationships, along with presence–absence and species richness patterns, among the terrestrial mammals of Western Europe recorded in two different data sets: a distribution atlas published 13 years ago (Mitchell-Jones *et al.*, 1999) with species presences and absences on Universal Transverse Mercator (UTM) 50 km × 50 km squares, and the most recent range maps (IUCN, 2010) gridded at the same resolution. We examined disparities in the patterns depicted by the different data sets and compared the distributional similarity matrices obtained from them, to analyse the robustness of chorological relationships to the source of distribution data.

MATERIALS AND METHODS

Species distribution data

Data from the *Atlas of European mammals* (Mitchell-Jones *et al.*, 1999; maps available at http://www.european-mammals.org/php/mapmaker.php) were obtained in tabular format, as a list of species recorded in each UTM 50 km × 50 km square. The range maps of terrestrial mammals from the Global Mammal Assessment were downloaded from IUCN (2010) in polygon shapefile format.

The range maps depict species' global distributions. The distribution atlas refers only to Europe, and the data are particularly incomplete in Eastern Europe, where surveying effort was lower and less uniform (A.J. Mitchell-Jones, Societas Europaea Mammalogica, pers. comm.). The study area was thus set to the countries of Western Europe (following UNESCO, 2009) that were included in the mammal atlas (Mitchell-Jones *et al.*, 1999). To make the age of the data more comparable between the two data sets, we selected the atlas presences recorded after 1970 and where later extinctions were not documented (Mitchell-Jones *et al.*, 1999) and those range polygons where species are unequivocally considered to be extant (IUCN, 2010).

To make the species pools comparable, we resolved taxonomical incongruences between the two data sets that resulted from changes in species names, genus reassignments, species splits, and new species descriptions from between the publication of the distribution atlas (Mitchell-Jones *et al.*, 1999) and that of the range maps (IUCN, 2010). The taxonomic conversions performed followed IUCN (2010), which was the general nomenclatural source for this study. We used modern species names wherever possible (see Appendix S1).

Spatial data processing

We downloaded vector maps of European political boundaries and the 50 km × 50 km UTM grids covering Europe from the EDIT Geoplatform (Sastre *et al.*, 2009). We then used GRASS 6 (GRASS Development Team, 2009) through the graphical interface of QUANTUM GIS 1.7 (Quantum GIS Development Team, 2009) to select the 50 km × 50 km UTM grids covering the terrestrial area of Western Europe. Although there is some size variation in these near-equal-area grid cells, namely along the coastline and at the unions between UTM zones, these differences have shown to have a minor effect on broad-scale analyses of species richness (Nogués-Bravo & Araújo, 2006; Hurlbert & Jetz, 2007).

We imported the gridded study area to a POSTGIS spatial database under POSTGRESQL 8.4 (PostgreSQL Global Development Group, 2010), together with the shapefile of the terrestrial mammal range maps (IUCN, 2010). We intersected them to obtain a list of the species with any range within each UTM cell. All further data management and analyses were performed in R 2.10.1 (R Development Core Team, 2009) except where otherwise stated.

Biogeographical comparisons

For each data set, we converted the list of species per UTM cell into a table showing the presence or absence of each species in each cell. We then compared the species richness patterns, species occurrence (presence–absence) patterns and chorological relationships between species in the atlas and range map data.

We compared species richness in three different ways. First, we tested for systematic differences in species richness from atlas and range map data with a Wilcoxon signed rank test. Second, we checked if species richness varied concomitantly in the two data sets using Spearman's rank correlation with Dutilleul's (1993) sample size adjustment for spatial autocorrelation, implemented in the software SAM (Rangel *et al.*, 2010). Third, we calculated a measure of overall resemblance between the two species richness maps using the MAP COMPARISON KIT 3.2.2 (Geonamica/RIKS, Maastricht, The Netherlands; Visser & de Nijs, 2006). We used fuzzy numerical comparison, which considers fuzziness of locations (the notion that the representation of a cell depends on the cell itself and, to a lesser extent, also the cells in its neighbourhood). The following formula is employed to find the fuzzy resemblance (FR) of two values a and b (Hagen-Zanker *et al.*, 2006):

$$FR(a,b) = 1 - \frac{|a-b|}{\max(|a|,|b|)}. \quad (1)$$

In this case, a and b correspond to species richness values in atlas and range map data, respectively. The algorithm compares a specific grid cell in one map with the grid cells in the other map lying within the neighbourhood of that cell, thus performing pattern recognition considering local and

global similarities. We used the default values for neighbourhood radius (4) and decay (exponential, halving distance = 2), but confirmed that the results were robust to different values (for more details, see Visser & de Nijs, 2006).

Presence–absence patterns from atlas and range map data for each species were compared using two measures: the overall agreement (or correct classification rate: the proportion of cells with matching values in both data sets) and Cohen's kappa, which accounts for differences in prevalence between the two maps by correcting the expected percentage of agreement for the fraction of agreement expected at random. We also directly compared patterns in prevalence across species using both a Wilcoxon signed rank test and Spearman's correlation coefficient.

Chorological relationships were established between species for each data set (atlas and range maps) by creating matrices of pairwise similarities in species' geographical distributions based on two of the similarity indices most commonly employed in chorological analyses. Jaccard's (1901) index is one of the most widely used similarity indices in ecology (e.g. Real & Vargas, 1996; Chao *et al.*, 2005; Anderson *et al.*, 2009; Sillero *et al.*, 2009; Pilehvar *et al.*, 2010; Engen *et al.*, 2011). It quantifies the shared range of each pair of species as a proportion of their combined range. Jaccard's index (J) can be written as follows:

$$J = \frac{C}{A + B - C} \quad (2)$$

where A and B are the numbers of localities where each of two species is present and C is the number of localities shared by both.

Baroni-Urbani & Buser's (1976) index (BUB) is also extensively used (e.g. Márquez *et al.*, 1997; Real *et al.*, 1997; Flores *et al.*, 2004; Báez *et al.*, 2005; Real *et al.*, 2008; Olivero *et al.*, 2011), and can be written as:

$$BUB = \frac{\sqrt{CD} + C}{\sqrt{CD} + A + B - C}$$

where A, B and C are the same as in Jaccard's index and D is the number of localities from which both species are absent. An index that accounts for both shared presences and shared absences gives a more complete picture of how similar two species' distributions are. Note that if two species do not share any presence or any absence localities, $BUB = J$ (as a corollary, if they share only absence localities, their distributional similarity is still zero); but if species share both a presence and an absence area, these are both taken into account. Both J and BUB indices may vary between 0 (no distributional overlap) and 1 (identical distributions).

We used two approaches to assess the agreement between the distributional similarity matrices among data sets for both similarity indices. First, we used Mantel tests to analyse Spearman's rank correlation between the matrices, to give an overall assessment of agreement between the chorological relationships in the two data sets. Second, we quantified the degree of difference arising between pairs of closely associated species. For each species, we found the most similarly distributed species from one data set and calculated the rank of the similarity of that species pair in the other data set. Plotting the sorted ranks against cumulative numbers of species shows the degree of conservation of similar species pairs between data sets (Fig. 2).

RESULTS

We matched and analysed a total of 156 mammal species (Appendix S2). The study area included 2118 UTM 50 km × 50 km grid cells. The mammal atlas showed the presence of at least one species in 1985 (93.7%) of these cells, and the range maps intersected with 2104 cells (99.3%).

Four species (2.6%: *Apodemus uralensis*, *Eptesicus bottae*, *Lepus castroviejoi* and *Sciurus anomalus*) had the same number of presence records in the atlas and the gridded range maps, five (3.2%: *Capra ibex*, *Cricetulus migratorius*, *Lynx pardinus*, *Mustela eversmanii* and *Rupicapra pyrenaica*) had more records in the atlas than in the range maps, but the vast majority of species (the remaining 147, i.e. 94.2%) had more occurrence cells according to their range map than recorded in the atlas. Our data set was, therefore, similar to those previously analysed (e.g. Hurlbert & White, 2005; Hurlbert & Jetz, 2007; Gaston & Fuller, 2009) in that range maps provided larger estimates of species' occurrence areas than atlas or survey data.

As expected, mean species richness was significantly different in atlas and range map data (Wilcoxon paired test, $V = 35736.5$, $P < 0.001$, $n = 2218$; Fig. 2a), and it was consistently lower in the atlas, except at low diversity (Fig. 2d). However, species richness values in atlas and range data were significantly correlated (Spearman's correlation with Dutilleul's correction, $\rho = 0.76$, $P < 0.001$, $n = 2118$, corrected d.f. = 13). Likewise, the species richness maps resulting from the two data sets showed largely similar geographical trends, despite some local differences (Fig. 3). The mean fuzzy resemblance between both maps (i.e. visual similarity in spatial species richness patterns) was 0.69.

The number of presences per species on UTM grid cells ranged between 1 (for *Macaca sylvanus* and *Meriones tristrami*) and 1653 (for *Vulpes vulpes*) in the atlas and between 2 (*Macaca sylvanus* and *Meriones tristrami*) and 2103 (*Mus musculus*) according to the range maps. Across species, prevalence was significantly higher from range maps than from atlas data (Wilcoxon paired test, $V = 86.5$, $P < 0.001$, $n = 156$; Fig. 2b), as expected. However, species prevalence was highly correlated between the two data sets ($\rho = 0.969$, $P < 0.001$, $n = 156$). The overall agreement rate between species' presence–absence patterns was generally high (mean = 0.90; Fig. 2c). Incorporating the expected chance agreement by taking into account differences in prevalence between the two data sets (Cohen's kappa) showed lower agreement in presence–absence patterns (mean = 0.61; Fig. 2c). However, a measure with chance correction may not be advantageous under all circumstances (Visser & de Nijs, 2006).

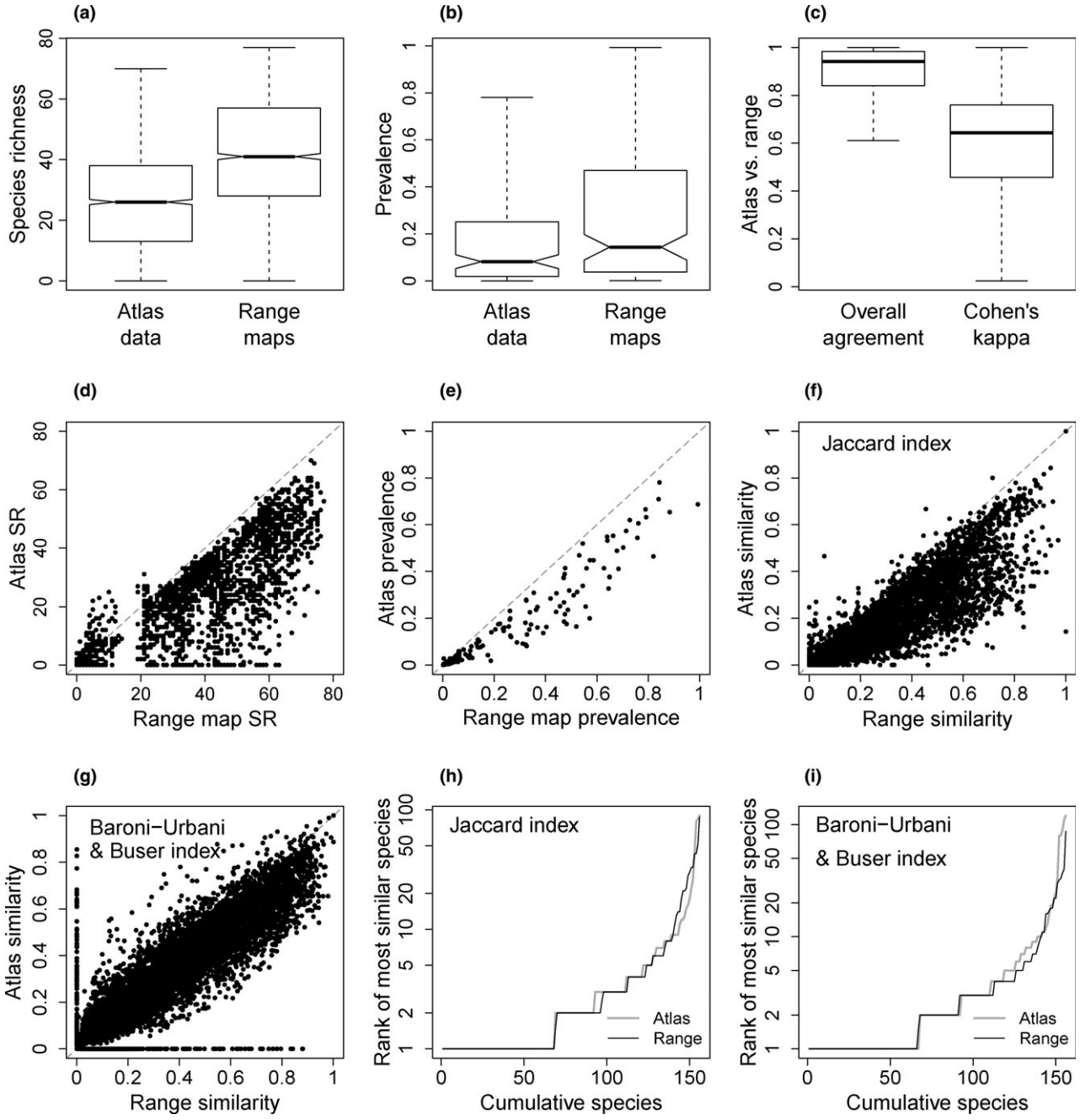

Figure 2 Comparison of distribution atlas and range map data for Western European terrestrial mammals. (a–c) Box plots showing median, upper and lower quartiles, and extreme values for (a) species richness by grid cell (n = 2118), (b) prevalence by species (n = 156), and (c) agreement between species presence–absence patterns (n = 156). (d–g) Scatter plots comparing the values of (d) species richness (SR) by grid cell (n = 2118), (e) prevalence by species (n = 156), and (f–g) pairwise distributional similarity between species (n = 156 × 156) with two different similarity indices. (h–i) Sorted ranks of the most similar species from each data set on the similarity matrix of the other data set (see text for more details).

The similarity between pairs of species' distributions was consistent between the two data types under both similarity indices (Fig. 2f,g). With Jaccard's index (J), mean pairwise similarity between species was 0.084 for atlas data and 0.14 for range map data. With *BUB*, mean similarity was 0.22 for atlas data and 0.25 for range data. For both indices, the similarity matrices obtained from atlas and range data were highly correlated (Jaccard, $\rho = 0.958$, Fig. 2f; *BUB*, $\rho = 0.943$, Fig. 2g; $P < 0.001$ in both cases, based on 9999 Mantel permutations). For 68 species (44%) with the J index and 67 (43%) with *BUB*, the species with the most similar distribution was the same regardless of the data type used (Fig. 2h,i).

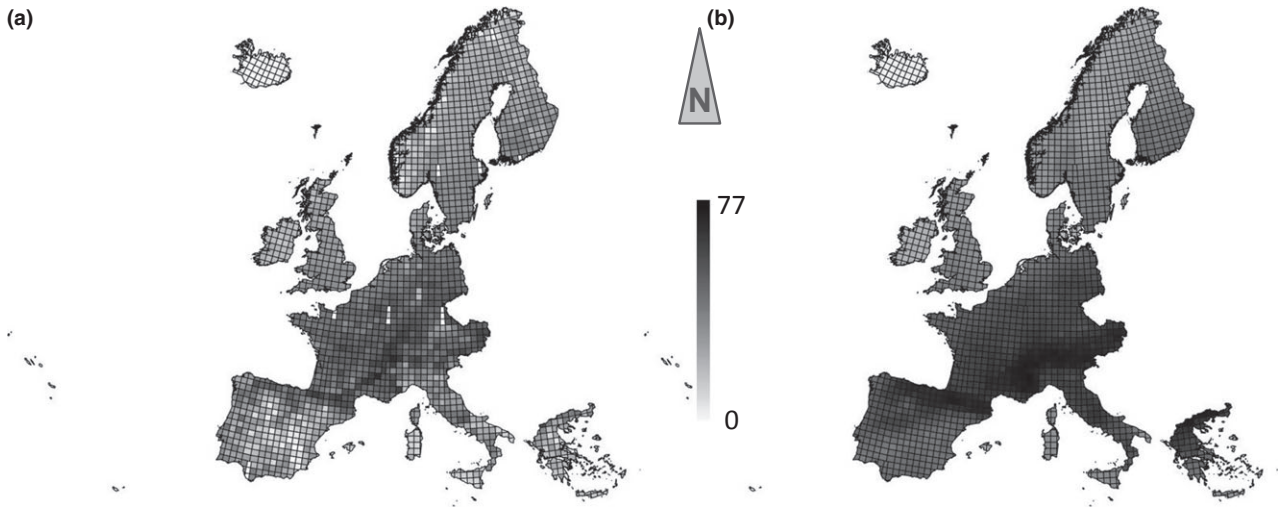

Figure 3 Terrestrial mammal species richness on Western European 50 km × 50 km UTM grid cells according to (a) the *Atlas of European mammals* (Mitchell-Jones *et al.*, 1999) and (b) the mammal distribution range maps from the Global Mammal Assessment (IUCN, 2010). Maps are in Lambert Equal Area projection (datum ETRS).

For the *BUB* index, incongruence in chorological relationships between data sets was concentrated within species with small distribution areas: the maximum difference between distributional similarity values shown by each species in atlas and range data was negatively correlated with its prevalence ($\rho = -0.56$ and -0.52 for atlas and range prevalence, respectively; $P < 0.001$). Thirty-six species (23%) showed differences in *BUB* greater than 0.5 between the two data sets; all of them had low prevalence values (≤ 0.05 in atlas data, ≤ 0.10 in the gridded range map data). With Jaccard's index, there were 24 (15%) species with a difference higher than 0.5 between their similarity values in atlas and range data and their prevalences ranged more widely, going up to 0.65 in the atlas and 0.88 in the range map data.

DISCUSSION

Biogeographical studies often depend on coarse-scale compilations of species distributions. Normally, at broad scales, the best information available is distribution atlas data or range maps. Researchers acknowledge that each of these types of information has drawbacks; for instance, they can be biased depending on the survey effort (Dennis & Thomas, 2000; Estrada *et al.*, 2008; Barbosa *et al.*, 2010), and they can represent species' distributions on a coarser scale than would be desirable (Hurlbert & White, 2005; Hurlbert & Jetz, 2007). Nevertheless, this does not preclude their use as baseline information to conduct a range of spatial analyses. When a choice has to be made between different sources or types of data to conduct a biogeographical analysis, or even when only one data type is available for a particular region or taxonomic group, it is important to rule out (or at least acknowledge) a strong influence of the data type on the results and conclusions of the study.

Distribution atlas and range map data may bias biodiversity analyses in opposite directions by respectively over- and underestimating species' areas of occurrence. These opposite biases were clearly reflected in the differences between grid-cell species richness from the two data sets analysed here (Fig. 2a,d). Local mammal species richness was generally higher in gridded range maps than in atlas data, with atlas richness varying widely for a given level of range-map richness (Fig. 2d). These results match those of previous wide-scale studies comparing bird range maps to survey (Hurlbert & White, 2005) or atlas data (Hurlbert & Jetz, 2007), and extents of occurrence to areas of occupancy (Gaston & Fuller, 2009). The characteristics of our data sets were thus analogous to those of previously analysed data. Nevertheless, atlas and range map richness were strongly correlated, even when accounting for spatial autocorrelation, indicating similar relative spatial patterns. Moreover, the species richness maps showed a relatively high fuzzy resemblance, i.e. visually similar patterns.

Mean species prevalence was also lower in atlas than in range map data, as most species had more presence grid cells according to the latter (Fig. 2b,e). This corroborates the tendency for an opposite bias, i.e. under- versus overestimation of species' occurrence areas by atlas and range map data, respectively. Nevertheless, the overall agreement between both data sets was still high (Fig. 2c), and rank correlation analysis showed that more prevalent species in range map data were also more prevalent in the atlas.

Despite the opposite bias in their estimates of species' areas of occurrence and the additional disparities caused by the temporal lag (and some differences in criteria) between the two distributional data sets, chorological relationships among species in atlas and range map data were remarkably similar (Fig. 2f–i). Clusters of chorologically related species – i.e. species with similar geographical distributions – can serve a

range of useful purposes in biogeography and conservation planning (Márquez *et al.*, 1997; Hausdorf & Hennig, 2003; Real *et al.*, 2008; Olivero *et al.*, 2011). Our results showed that the chorological relationships on which these clusters are based can be particularly robust to differences in distribution data type. This occurred both for a distributional similarity index that takes only shared presences into account (Jaccard, 1901; Fig. 2f,h) and for an index that also accounts for shared absences (Baroni-Urbani & Buser, 1976; Fig. 2g,i).

Similarity indices are typically based on the number of shared attributes (in this case, presence localities) between species. However, such indices are not otherwise spatially explicit and, in particular, they do not account for proximity between species' distributions. Consequently, the distribution areas of species living at adjacent survey units are considered just as different as those of species occurring at opposite ends of the study area. This may increase the scale-dependence of chorological relationships, as well as the effect of slight spatial errors in the georeferencing of species records. This may have particularly strong effects on the similarity values between the distributions of small-range species: the coincidence (or lack thereof) of their occurrence in just a couple of localities may mean a difference between a zero or a high similarity value, as a few localities may represent a considerable proportion of their range. This effect was especially evident in the comparison of the *BUB* similarity matrices, where a number of small-range species had zero similarity in one data set and up to 0.88 similarity in the other (Fig. 2g).

Similarity indices that account for fuzziness of location, such as the fuzzy resemblance used here to compare the species richness maps, may improve future chorological analyses, as they introduce tolerance for small spatial differences (Visser & de Nijs, 2006; Barbosa & Real, 2012). However, much development is still needed, namely in optimizing the computation of fuzzy resemblance for multiple map pairs and in determining its levels of significance. Although attempts have been made to apply fuzzy logic to the definition of chorotypes (Olivero *et al.*, 2011), these are still based on similarity indices that account for coincidence but not for proximity between presence localities, with potentially large effects on the relationships involving small-range species.

CONCLUSIONS

Despite the substantial differences between the distribution data from the two sources we have used, there was a high general agreement between species' distribution patterns. More importantly, despite these data differences and the lack of provision for spatial structure in the similarity indices, the chorological relationships (i.e. distributional similarity) between the analysed species were remarkably congruent, indicating that the type of distribution data may not significantly affect the results of such analyses, at least at this scale. An exception should be made for small-range species, for which slight differences between data sets may mean their coincidence or not in a sizeable part of their distribution areas, and hence more variable relationships. This, however, may be improved in the future by incorporating fuzzy logic and therefore tolerance for small spatial discrepancies in species' occurrence patterns.

Although it is very important to perform biogeographical analyses with the best data available, our results show that distribution atlas and gridded range maps produce highly concordant chorological relationships between Western European terrestrial mammals, even at a relatively fine resolution for this spatial extent (50 km × 50 km). Chorological relationships can thus be considered fairly robust to the data source, more so than patterns in species richness (Hurlbert & White, 2005; Hurlbert & Jetz, 2007; Fig. 2). This constitutes very helpful information for analytical biogeographers and conservation biogeographers, as they can assume that their results would change only slightly if they used different data sources.

ACKNOWLEDGEMENTS

This work was presented as a poster at the IBS Crete meeting, January 2011. We thank the Societas Europaea Mammalogica (particularly A.J. Mitchell-Jones) and the IUCN for providing the species distribution data. Susanne Fritz, Diego Fontaneto, Andrew Smith, Heikki Henttonen, Claudia Fontaneto, Anna Skarin and Vineet Katariya provided useful information. A.M.B. is supported by a post-doctoral fellowship (SFRH/BPD/40387/2007) from the Portuguese Foundation for Science and Technology (FCT), co-financed by the European Social Fund. A.L.M. is supported by a project (CGL2009-11316/BOS) from the Ministry of Science and Innovation (Spain) and the European Regional Development Fund. The 'Rui Nabeiro' Biodiversity Chair is financed by Delta Cafés and an FCT project (PTDC/AAC-AMB/098163/2008).

REFERENCES

Anderson, R.J., Bolton, J.J. & Stegenga, H. (2009) Using the biogeographical distribution and diversity of seaweed species to test the efficacy of marine protected areas in the warm-temperate Agulhas Marine Province, South Africa. *Diversity and Distributions*, **15,** 1017–1027.

Báez, J.C., Real, R., Vargas, J.M. & Flores-Moya, A. (2005) Chorotypes of seaweeds from the western Mediterranean Sea and the Adriatic Sea: an analysis based on the genera *Audouinella* (Rhodophyta), *Cystoseira* (Phaeophyceae) and *Cladophora* (Chlorophyta). *Phycological Research*, **53,** 255–265.

Ball, I.R. (1975) Nature and formulation of biogeographical hypothesis. *Systematic Zoology*, **24,** 407–430.

Barbosa, A.M. & Real, R. (2012) Applying fuzzy logic to comparative distribution modelling: a case study with two sympatric amphibians. *The Scientific World Journal*, article ID 428206, doi: 10.1100/2012/428206.

Barbosa, A.M., Fontaneto, D., Marini, L. & Pautasso, M. (2010) Is the human population a large-scale indicator of

the species richness of ground beetles? *Animal Conservation*, **13,** 432–441.

Baroni-Urbani, C. & Buser, M.W. (1976) Similarity of binary data. *Systematic Biology*, **25,** 251–259.

Baroni-Urbani, C., Ruffo, S. & Vigna-Taglianti, A. (1978) Materiali per una biogeografia italiana fondata su alcuni generi di Coleotteri Cicindelidi, Carabidi e Crisomelidi. *Memorie della Società Entomologica Italiana*, **56,** 35–92.

Carmona, J.A., Doadrio, I., Márquez, A.L., Real, R., Hugueny, B. & Vargas, J.M. (1999) Distribution patterns of indigenous freshwater fishes in the Tagus River basin, Spain. *Environmental Biology of Fishes*, **54,** 371–387.

Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2011) Incorporating evolutionary processes into conservation planning using species distribution data: a case study with the western Mediterranean herpetofauna. *Diversity and Distributions*, **17,** 408–421.

Chao, A., Chazdon, R.L., Colwell, R.K. & Shen, T.-J. (2005) A new statistical approach for assessing compositional similarity based on incidence and abundance data. *Ecology Letters*, **8,** 148–159.

Dennis, R.L.H. & Thomas, C.D. (2000) Bias in butterfly distribution maps: the influence of hotspots and recorder's home range. *Journal of Insect Conservation*, **4,** 73–77.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12,** 53–64.

Dutilleul, P. (1993) Modifying the *t* test for assessing the correlation between two spatial processes. *Biometrics*, **49,** 305–314.

Engen, S., Grøtan, V. & Sæther, B.-E. (2011) Estimating similarity of communities: a parametic approach to spatio-temporal analysis of species diversity. *Ecography*, **34,** 220–231.

Estrada, A., Real, R. & Vargas, J.M. (2008) Using crisp and fuzzy modelling to identify favourability hotspots useful to perform gap analysis. *Biodiversity and Conservation*, **17,** 857–871.

Flores, T., Puerto, M.A., Barbosa, A.M., Real, R. & Gosálvez, R.U. (2004) Agrupación en corotipos de los anfibios de la provincia de Ciudad Real (España). *Revista Española de Herpetología*, **18,** 41–53.

Fontaneto, D., Barbosa, A.M., Segers, H. & Pautasso, M. (2012) The 'rotiferologist' effect and other global correlates of species richness in monogonont rotifers. *Ecography*, **35,** 174–182.

Gasc, J.-P., Cabela, A., Crnobrnja-Isailovic, J., Dolmen, D., Grossenbacher, K., Haffner, P., Lescure, J., Martens, H., Martínez Rica, J.P., Maurin, H., Oliveira, M.E., Sofianidou, T.S., Veith, M. & Zuiderwijk, A. (eds) (1997) *Atlas of amphibians and reptiles in Europe*. Societas Europaea Herpetologica & Muséum National d'Histoire Naturelle (IEGB/SPN), Paris.

Gaston, K.J. & Fuller, R.A. (2009) The sizes of species' geographic ranges. *Journal of Applied Ecology*, **46,** 1–9.

GRASS Development Team (2009) *Geographic resources analysis support system*. Open Source Geospatial Foundation. Available at: http://grass.osgeo.org/ (accessed 10 January 2009).

Hagemeijer, W.J.M. & Blair, M.J. (eds) (1997) *The EBCC atlas of European breeding birds – their distribution and abundance*. European Bird Census Council, T. & A.D. Poyser, London.

Hagen-Zanker, A., Engelen, G., Hurkens, J., Vanhout, R. & Uljee, I. (2006) *Map Comparison Kit 3 user manual*. RIKS BV, Maastricht.

Hausdorf, B. & Hennig, C. (2003) Biotic element analysis in biogeography. *Systematic Biology*, **52,** 717–723.

Heath, J. & Leclercq, J. (eds) (1981) *Provisional atlas of the invertebrates of Europe. Maps 1–27*. Institute of Terrestrial Ecology, Cambridge, UK.

Hurlbert, A.H. & Jetz, W. (2007) Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences USA*, **104,** 13384–13389.

Hurlbert, A.H. & White, E.P. (2005) Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecology Letters*, **8,** 319–327.

IUCN (2010) *IUCN red list of threatened species. Version 2010.4*. Available at: http://www.iucnredlist.org/technical-documents/spatial-data (accessed 2 March 2011).

Jaccard, P. (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Societé Vaudoise des Sciences Naturelles*, **37,** 547–579.

Jalas, J. & Suominen, J. (1972–94) *Atlas Florae Europaeae*, Vols 1–10. Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.

Kéry, M., Gardner, B. & Monnerat, C. (2010) Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, **37,** 1851–1862.

Liébanas, G., Peña-Santiago, R., Real, R. & Márquez, A.L. (2002) Spatial distribution of dorylaimid and mononchid nematodes from southeast Iberian Peninsula: chorological relationships among species. *Journal of Nematology*, **34,** 390–395.

Márquez, A.L., Real, R., Vargas, J.M. & Salvo, A.E. (1997) On identifying common distribution patterns and their causal factors: a probabilistic method applied to pteridophytes in the Iberian Peninsula. *Journal of Biogeography*, **24,** 613–631.

Mitchell-Jones, A.J., Amori, G., Bogdanowicz, W., Kryštufek, B., Reijnders, P.J.H., Spitzenberger, F., Stubbe, M., Thissen, J.B.M., Vohralík, V. & Zima, J. (1999) *The atlas of European mammals*. Academic Press, London.

Nogués-Bravo, D. & Araújo, M.B. (2006) Species richness, area and climate correlates. *Global Ecology and Biogeography*, **15,** 452–460.

Olivero, J., Real, R. & Márquez, A.L. (2011) Fuzzy chorotypes as a conceptual tool to improve insight into biogeographic patterns. *Systematic Biology*, **60,** 645–660.

Perring, F.H. & Walters, S.M. (1962) *Atlas of the British flora*. Botanical Society of the British Isles, London.

Pilehvar, B., Veiskarami, G., Abkenar, K.T. & Soosani, J. (2010) Relative contribution of vegetation types to regional biodiversity in Central Zagross forests of Iran. *Biodiversity and Conservation*, **19,** 3361–3374.

PostgreSQL Global Development Group (2010) *PostgreSQL: the world's most advanced open-source database*. Available at: http://www.postgresql.org (accessed 9 April 2010).

Prendergast, J.R., Wood, S.N., Lawton, J.H. & Eversham, B.C. (1993) Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiversity Letters*, **1,** 39–53.

Quantum GIS Development Team (2009) *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project. Available at: http://qgis.osgeo.org/ (accessed 10 January 2009).

R Development Core Team (2009) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Available at: http://www.r-project.org/ (accessed 10 January 2009).

Rangel, T.F., Diniz-Filho, J.A.F. & Bini, L.M. (2010) SAM: a comprehensive application for Spatial Analysis in Macroecology. *Ecography*, **33,** 46–50.

Rapoport, E. (1982) *Areography: geographical strategies of species*. Pergamon, New York.

Real, R. & Vargas, J.M. (1996) The probabilistic basis of Jaccard's index of similarity. *Systematic Biology*, **45,** 380–385.

Real, R., Pleguezuelos, J.M. & Fahd, S. (1997) The distribution patterns of reptiles in the Riff region, northern Morocco. *African Journal of Ecology*, **35,** 312–325.

Real, R., Olivero, J. & Vargas, J.M. (2008) Using chorotypes to deconstruct biogeographical and biodiversity patterns: the case of breeding waterbirds in Europe. *Global Ecology and Biogeography*, **17,** 735–746.

Ribas, A., Barbosa, A.M., Casanova, J.C., Real, R., Feliu, C. & Vargas, J.M. (2007) Geographical patterns of the species richness of helminth parasites of moles (*Talpa* spp.) in Spain: separating the effect of sampling effort from those of other conditioning factors. *Vie et Milieu*, **57,** 45–52.

Sastre, P., Roca, P., Lobo, J.M. & EDIT co-workers (2009) A Geoplatform for improving accessibility to environmental cartography. *Journal of Biogeography*, **36,** 568.

Sillero, N., Brito, J.C., Skidmore, A.K. & Toxopeus, A.G. (2009) Biogeographical patterns derived from remote sensing variables: the amphibians and reptiles of the Iberian Peninsula. *Amphibia–Reptilia*, **30,** 185–206.

UNESCO (2009) *EFA Global Monitoring Report: North America and Western Europe*. Available at: http://www.unesco.org/en/efareport/regions/north-america-and-western-europe (accessed 23 September 2009).

Visser, H. & de Nijs, T. (2006) The map comparison kit. *Environmental Modelling and Software*, **21,** 346–358.

Wiens, J.A. (1989) Spatial scaling in ecology. *Functional Ecology*, **3,** 385–397.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Conversions applied to the species that did not match between data sets.

Appendix S2 List of species used in the analysis.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCH

A. Márcia Barbosa is a post-doctoral fellow jointly hosted by Imperial College London (UK) and the Research Centre in Biodiversity and Genetic Resources (CIBIO) at the University of Évora (Portugal). Her research interests include biogeography, macroecology, distribution modelling, comparative phylogeography, biodiversity patterns and conservation. The team have a common interest in the analysis of biogeographical patterns.

Author contributions: A.M.B. conceived and designed the study, gathered and analysed the data, and led the writing. A.E., A.L.M., A.P. and C.D.L.O. provided ideas for additional analyses and improved writing, interpretation and presentation.

Editor: Malte Ebach