SPECIAL PAPER

Identifying global zoogeographical regions: lessons from Wallace

Marta Rueda[1]*, Miguel Á. Rodríguez[2] and Bradford A. Hawkins[1]

[1]*Department of Ecology & Evolutionary Biology, University of California, Irvine, CA, USA,* [2]*Forest Ecology & Restoration Group, Department of Life Sciences, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain*

ABSTRACT

Aim When dividing the world into zoogeographical regions, Alfred Russel Wallace stipulated a set of criteria by which regions should be determined, foremost the use of generic rather than species distributions. Yet, recent updates of Wallace's scheme have not followed his reasoning, probably explaining in part the discrepancies found. Using a recently developed quantitative method, we evaluated the world's zoogeographical regions following his criteria as closely as possible.

Location Global.

Methods We subjected presence–absence data from range maps of birds, mammals and amphibians to an innovative clustering algorithm, affinity propagation. We used genera as our taxonomic rank, although species and familial ranks were also assessed, to evaluate how divergence from Wallace's criteria influences the results. We also accepted Wallace's argument that bats and migratory birds should be excluded (although he was contradictory about the birds) and devised a procedure to determine the optimal number of regions to eliminate subjectivity in delimiting the number of regions.

Results Regions attained using genera (eight for mammals and birds and six for amphibians) strongly coincided with the regions proposed by Wallace. The regions for amphibians were nearly identical to Wallace's scheme, whereas we obtained two new 'regions' for mammals and two for birds that largely coincide with Wallace's subregions. As argued by Wallace, there are strong reasons not to consider these as being equivalent to the six main regions. Species distributions generated many small regions related to contemporary climate and vegetation patterns, whereas at the familial rank regions were very broad. The differences between our generic maps and Wallace's all involve areas which he identified as being uncertain in his regionalization.

Main conclusions Despite more than 135 years of additional knowledge of distributions, the shuffling of generic concepts, and the development of computers and complex analytical techniques, Wallace's zoogeographical regions appear to be no less valid than they were when he proposed them. Recent studies re-evaluating Wallace's scheme should not be considered updates as such because they have not followed Wallace's reasoning, and all computer-based analyses, including this one, are subject to the vagaries of the particular methods used.

*Correspondence: Marta Rueda, Department of Ecology & Evolutionary Biology, University of California, Irvine, CA 92697, USA.
E-mail: mruedaga@uci.edu*

Journal of Biogeography

INTRODUCTION

It will be evident in the first place that nothing like a perfect zoological division of the earth is possible. The causes that have led to the present distribution of animal life are so varied, their action and reaction have been so complex, that anomalies and irregularities are sure to exist which will mar the symmetry of any rigid system.

A. R. Wallace (1876, Vol. 1, p. 53)

The partitioning of the world into 'natural' zoological regions, defined as areas with a consistency and distinctiveness of the fauna and flora, has attracted the attention of zoogeographers since the early 19th century. The first enduring global classification of regional faunas was developed by Sclater (1858), who established six primary zoological regions based on passerine birds. Sclater's scheme formed the basis of the arrangement employed later by Wallace (1876), in which he used a simple quantitative method to delimit global faunal regions based on the distribution of vertebrates in general, with a major focus on non-volant mammals, and further subdivided each of the six large regions into four subregions (see Figs 1a & 2a). Since then, many studies have attempted to modify Wallace's regions with limited success (e.g. Schmidt, 1954, and references therein), and although increasing knowledge about species distributions has led authors to periodically review and redraw Wallace's scheme (e.g. Smith, 1983; Cox, 2001), his remains the model largely accepted to date.

The development of computerized clustering methods, together with the availability of global species distribution databases, has renewed interest in assessing global biogeographical boundaries using more complex quantitative methods, ostensibly to evaluate the reliability of Wallace's scheme. These studies have identified a variable number of 'new' zoological regions that show some similarities but also large differences with Wallace's geographical divisions of the world's fauna (see Kreft & Jetz, 2010; Procheş & Ramdhani, 2012; Holt *et al.*, 2013a). Mismatches between these results and those of Wallace are probably caused in part by the specific methodological procedures employed. Yet, there may be additional, less apparent causes. In establishing the number, extent and boundaries of zoogeographical regions, Wallace (1876) proposed a set of principles to which regions should conform. However, none of the recent studies have followed his tenets, which make comparisons between Wallace and modern regionalizations highly tentative. But before revisiting Wallace's results using modern data and methods, it is necessary to understand the logic he used.

First, Wallace (1876, Vol. 1, pp. 54 & 63) stated that regions should be of a moderate number, approximately equal in size, and with easily defined – and remembered – boundaries corresponding as closely as possible with the major geographical divisions of the Earth's surface. As Wallace pointed out, the availability of a fairly extensive area is likely to have been an essential requirement for the development of most faunas. There are many smaller areas (e.g.

southern Europe, the Brazilian lowland rain forest or South Africa) characterized by distinct assemblages of animals that undoubtedly are the result of local climates, vegetation types, geographical barriers or past geological history, but he argued that such areas are too small, numerous and idiosyncratic to encompass the broader features of animal distributions, and their idiosyncrasies are minor when compared with the radical differences found, for example, between tropical Africa and tropical Asia (Wallace, 1876, Vol. 1, p. 68). Second, regions should also be taxonomically distinctive, whether by the possession of numerous unique genera or families or by the absence of genera or families that are abundant and widespread in adjacent regions (Wallace, 1876, Vol. 1, p. 54). Third, the delimitation of regions should not be based solely on the isolation of areas and the endemism of their fauna, but also on their richness and diversity of groups (Wallace, 1876, Vol. 1, p. 64). For example, if only isolation and levels of endemism are considered, Madagascar and New Zealand could be proposed as primary regions as a result of the presence of unique families and genera of mammals and birds. But he felt this was not appropriate for defining major biogeographical regions, although he did use endemism to define subregions (Wallace, 1876, Vol. 1, pp. 61–62 & 74–75).

Differences in dispersal abilities among animals also led Wallace (1876, Vol. 1, pp. 56–57) to argue for the impossibility of setting boundaries which best define the distribution of one taxonomic group that would be equally true for all other groups (also see opening quote), and he chose to construct the 'standard' zoogeographical regions based on the 'most appropriate group' – mammals for him – using the distributions of other groups only to clarify doubtful points. Moreover, Wallace was somewhat conflicted about how to deal with highly dispersive animals. On the one hand he excluded bats because they are almost universally distributed, and he thought they contained no information about the natural divisions of the Earth (Wallace, 1876, Vol. 1, p. 64). Migratory birds were more of a problem, and he was unclear about how to deal with them. Referring to Palaearctic birds, he thought that most of the genera that migrate southwards had been year-round residents in past epochs when the temperature of the Earth was warmer but are only able to maintain their presence in the north by migration in winter. Under this reasoning, he concluded, they should be considered as belonging to their breeding range, i.e. the temperate zones (Wallace, 1876, Vol. 1, p. 185). However, when discussing migration, Wallace did not accept that the breeding place of a migratory bird represents its true home instead of where it resides in winter, arguing that the likely reason why migratory birds leave the warm regions to breed in temperate latitudes is because they find abundant food there in summer. Consequently, because favourable breeding areas would be susceptible to changes of climate, he argued that the true range of a species should comprise the area that it occupies regularly for any part of the year (Wallace, 1876, Vol. 1, pp. 27–28), a criterion difficult to deal with in
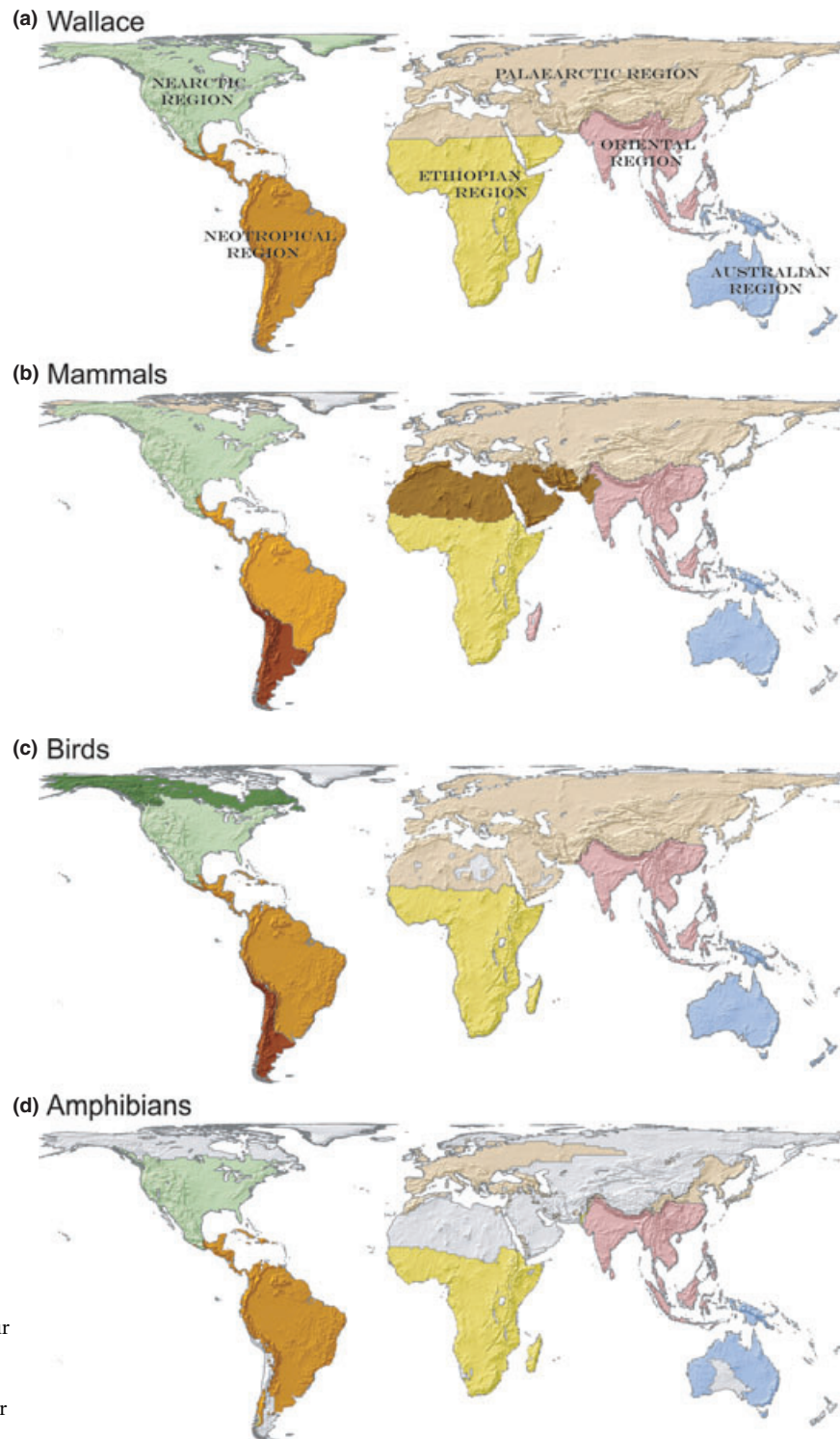
Figure 1 (a) The six zoogeographical regions as proposed by Wallace (1876) compared with the regions generated by our affinity-propagation protocol for (b) mammal, (c) bird and (d) amphibian genera. Grey areas represent cells with fewer than five genera.

practice and likely to generate inconsistent patterns among bird genera.

Finally, and most importantly, Wallace (1876, Vol. 1, pp. vii–viii; 1880, p. vii) did not consider species distributions to be appropriate for defining zoological regions. Species were intentionally ignored in his evaluations of the geographical distributions of taxa, because he argued that their more recent origins would not be as clearly related to long-term geographical change as would genera. He felt that genera represent the more permanent distribution of an 'organic type', being more clearly influenced by biotic changes and modifications of the physical environment operating over deep time. Wallace was very much fixed on defining regions based on major geological and climatic
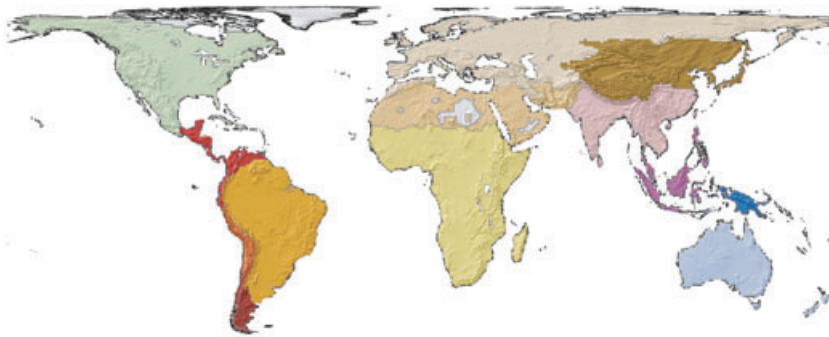
(a) Wallace
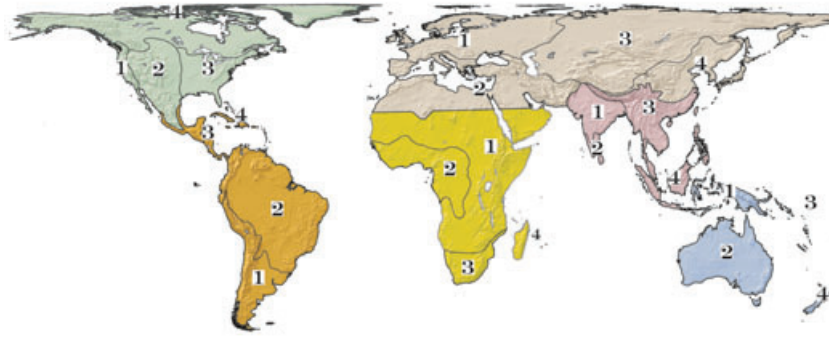
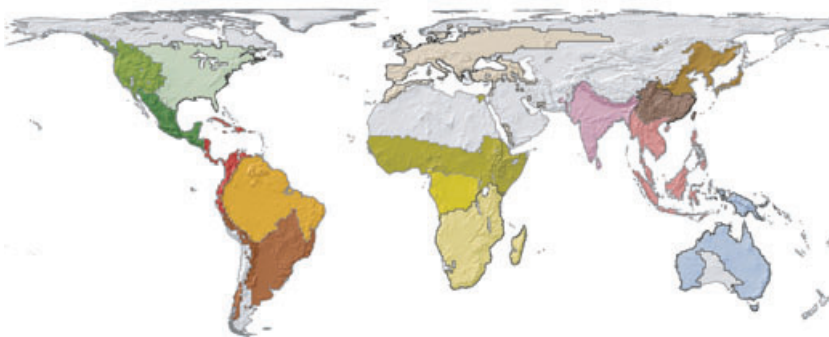**(b) Mammals**
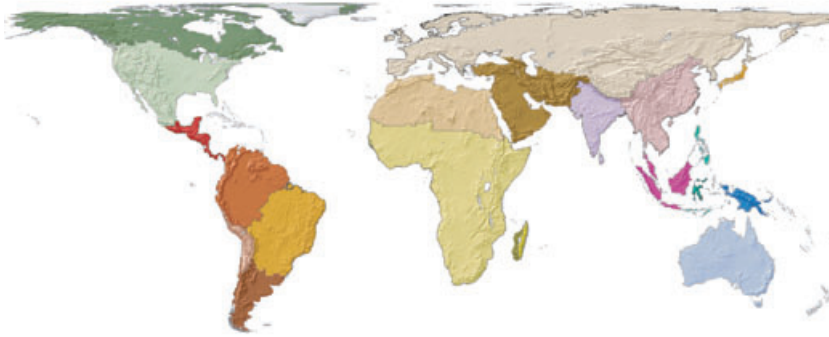
(c) Birds

(d) Amphibians

Figure 2 (a) The 24 subregions identified by Wallace (1876) compared with regions found for (b) mammal, (c) bird and (d) amphibian species. The numerical codes in (a) are those used by Wallace in his global map (Wallace, 1876): Nearctic (1) Californian, (2) Rocky Mountain, (3) Alleghanian, (4) Canadian; Neotropical (1) Chilean, (2) Brazilian, (3) Mexican, (4) Antillean; Palaearctic (1) European, (2) Mediterranean, (3) Siberian, (4) Manchurian; Ethiopean (1) East African, (2) West African, (3) South African, (4) Malagasy; Oriental (1) Indian, (2) Ceylon and South Indian, (3) Himalayan or Indo-Chinese, (4) Indo-Malayan or Malayan; Australian (1) Austro-Malayan, (2) Australian, (3) Polynesian, (4) New Zealand.

events spanning the early and mid-Cenozoic, which he felt that generic distributions were more likely to reflect, whereas species distributions were largely influenced by more recent events (e.g. the Pleistocene ice ages) and so were not useful for defining major biogeographical regions. He also pointed out that species can show greater geographical anomalies in

their distribution than genera. For example, islands may contain species that are relicts that have survived as a result of long, continued isolation and protection from competition (Wallace, 1876, Vol. 1, p. 361). In such cases, the use of species would lead us to define most tropical and subtropical islands as zoological regions (which would also violate his principle that they should be roughly comparable in size). Wallace (1876, Vol. 1, pp. 77–78 & 451–452) illustrated this problem using New Zealand bird species: if we based our classification on bird species distributions, we might consider New Zealand as a zoological region itself. However, when looking at its birds as a whole, most belong to Australian genera.

Wallace's advice about the utility of genera for defining biogeographical regions could be problematical, because all taxonomic ranks above species are generally considered to be arbitrary decisions by taxonomists. This is especially relevant to evaluating Wallace's results in the 21st century. There have been many taxonomic revisions of vertebrates over the past 135 years, the outcome of which is very difficult to predict with respect to global regionalizations. For example, non-volant mammals comprised 404 genera in Wallace (1876), but the classification used in the IUCN mammal range maps (see below) comprises 947. Although some of these genera may comprise species unknown in Wallace's time, presumably genera were also taxonomically broader than they are now, and this difference could affect the regionalization scheme depending on how species have been shuffled among the many new genera. Despite this potential issue, Wallace's reasons for not using species distributions are clearly stated, and if we are going to test his scheme here we must do so following his criteria.

In this paper we evaluate whether the world's zoogeographical regions proposed by Wallace (1876) can be confirmed if his general classification criteria are followed as closely as possible. To compare against other recent regionalizations we had to select a quantitative clustering method, of which there are several to choose from: none having logical priority from a biogeographical perspective. Based on recent advances in methodology we selected an innovative clustering analysis – affinity propagation (Frey & Dueck, 2007) – and assessed the regions for the vertebrate groups for which global distribution maps are available (birds, mammals and amphibians). We base our results and discussion on the generic rank, although species and familial rank regionalizations were also generated to illustrate how using them results in schemes substantially different from what Wallace intended even when using a standardized analytical method. We also excluded both bats and long-distance migratory birds; the former because he did, and the latter because of his contradictory point of view about where they belong geographically. Also, small islands are excluded to meet his criterion that regions should not be too small and driven by a few endemic genera. Finally, we utilize a clustering process referred to as adaptive affinity propagation to compute the optimal number of clusters to define regional

solutions that are not arbitrarily selected from a wide range of solutions.

MATERIALS AND METHODS

Species distribution data and distance metric

Species range maps were obtained from BirdLife International for birds (http://www.birdlife.org/datazone/info/spcdownload, accessed June, 2011) and from the IUCN for mammals (http://www.iucnredlist.org/initiatives/mammals/description/download-gis-data, accessed January, 2010) and amphibians http://www.iucnredlist.org/initiatives/amphibians/description/download-gis-data, accessed January, 2010). The maps were processed in ArcGIS 9.3 (ESRI, Redlands, CA, USA) to extract species occurrences into a global grid with a 9309 km^2 grain size – based on Behrmann's projection – to generate presence–absence matrices in which every row represents a grid cell and every column a species. We then excluded: (1) cells containing less than 50% of land area, so that small islands and coastal cells are not included in further analyses; (2) cells containing fewer than five species to reduce potential distortions caused by having few taxa and to make the results more directly comparable to at least some pre-existing analyses (e.g. Kreft & Jetz, 2010); (3) bats; and (4) pelagic and transcontinental migratory birds. These exclusion criteria rendered a total of 13,483 cells and 7875 species for birds, 14,097 cells and 3901 species for mammals, and 8537 cells and 5829 species for amphibians. Additionally, using the taxonomies provided in the downloaded files, we combined distributions of all species in each genus (1861 for birds, 947 for mammals and 504 for amphibians) and family (224 for birds, 119 for mammals and 55 for amphibians), and presence–absence matrices were constructed for these data.

The matrices were used to calculate pairwise distances between pairs of cells for each animal group and taxonomic rank. As with clustering algorithms, there are many distance or taxonomic/phylogenetic turnover metrics to choose from, and none can be said to be perfect from a biogeographical perspective. We selected Hellinger distances (Rao, 1995), which are calculated by first modifying the species presence–absence data and then computing the Euclidean distance among sites – or pairs of cells – based on the modified data (see details in Legendre & Gallagher, 2001). Accordingly, if $Y = [y_{ij}]$ is a presence–absence matrix of size $(n \times p)$ with grid cells (rows) $i = \{1\ldots n\}$ and species (columns) $j = \{1\ldots p\}$, the data $Y = [y_{ij}]$ are first transformed into $Y' = [y'_{ij}]$ as follows:

$$[y'_{ij}] = \sqrt{\frac{y_{ij}}{y_{i+}}}$$

where y_{i+} is the sum of the occurrences per row (grid cell richness in our case); then the Euclidean distance between the rows (grid cells) of transformed data is identical to the Hellinger distance between the original rows (grid cells) of the species occurrence data:

$$D_{\text{Hellinger}}(x_1, x_2) = \sqrt{\sum_{j=1}^{p} (y\prime_{1j} - y\prime_{2j})^2}$$

where x_1 and x_2 are grid cells described by the $j = \{1\ldots p\}$ species list, and y_1 and y_2 are occurrence values of j for the grid cells in question. The Hellinger distance is used to avoid both the 'double-zero problem', i.e. when two sites or grid cells that have no species in common are assigned the same distance as two sites that share species; and the 'species-abundance paradox', which frequently occurs when two sites share only a small fraction of all the species in the same regional pool (Legendre & Legendre, 1998; Gagné & Proulx, 2009).

Clustering method and analytical procedure

We applied a machine-learning algorithm referred to as affinity propagation (Frey & Dueck, 2007) to generate zoogeographical regions for each animal group and taxonomic rank. Affinity propagation (AP hereafter) is extensively used in bioinformatics and astrophysics and is making inroads in ecology (e.g. Cardille & Lambois, 2009) as a more powerful alternative to popular methods such as *k*-means. AP can compress a massive data set very efficiently (i.e. with lower error), while simultaneously identifying and retaining the single member that best represents each group (see Frey & Dueck, 2007, 2008).

Contrary to algorithms in which cluster centres are found by randomly choosing an initial subset of data points, AP takes as input measures of 'similarities' between pairs of data points (cells in this case) and simultaneously considers all the points as potential cluster centres (called 'exemplars' in AP terminology). AP identifies these exemplars via 'message passing' between points and searches for clusters through an iterative process until a high-quality set of exemplars and corresponding clusters emerges (see Frey & Dueck, 2007, for technical details). One important advantage is that the number of clusters need not be specified beforehand. Instead, AP takes as input a similarity value for each data point so data points with greater similarities are more likely to be chosen as exemplars. These values are referred to as 'preferences'. However, AP clustering has the limitation that it is difficult to know the value of 'preference' that yields the optimal clustering solution. To solve this, we used a variant of the AP algorithm referred to as adaptive-AP (Wang *et al.*, 2007). Adaptive-AP searches for the best solution within the range of possible preference values for the data set, from the minimum preference value (for which one or at most two clusters would be obtained) to the maximum (for which as many clusters as data points would be obtained). Thus, a series of results with different numbers of clusters is obtained. Subsequently, the adaptive-AP algorithm uses a cluster validation technique – the 'silhouette index' – to evaluate the optimal clustering solution for the data. The silhouette index reflects the compactness and sep-

aration of clusters and is applicable to both the estimation of the optimal number of clusters and evaluation of clustering quality (Rousseeuw, 1987; see also Wang *et al.*, 2007); the largest overall average silhouette value identifies the highest clustering quality and the optimal number of clusters.

Ideally, we would have implemented the adaptive-AP algorithm beginning with the first pass through the data. However, it was computationally unfeasible to manage the large similarity matrices generated, because the size of a similarity matrix increases with the square of the number of items; for example, our smallest matrix (amphibians) contained 72,897,444 pairwise distance values (i.e. 8538 × 8538). To solve this computational limitation we devised the following procedure. We first ran an AP analysis using the full similarity matrices, i.e. based on the complete presence–absence matrices. AP is computationally less demanding than its adaptive version because it does not scan all possible preference values.

For this first AP analysis we used the R package APClus-ter (Bodenhofer *et al.*, 2011), setting the preference value as the median of the pairwise similarities. This is standard practice when all data points are equally suitable as exemplars (Frey & Dueck, 2007) and results in a large number of clusters. For example, at the generic rank we obtained 935, 1026 and 607 clusters for birds, mammals and amphibians, respectively (see Table 1). Then, to reduce the size of the similarity matrices, we classified each grid cell (row) of the original presence–absence matrices according to the cluster to which they were assigned in the first analysis and calculated the mean number of presences per cluster for each species/genus/family (columns). For example, at the generic rank this procedure allowed us to reduce the size of the original matrices, from 13,483 grid cells (rows) for birds, 14,097 for mammals and 8538 for amphibians, to new matrices with far fewer rows (935 for birds, 1026 for mammals and 607 for amphibians; see Table 1) to construct similarity matrices with practical computational sizes. We then ran adaptive-AP clustering in Matlab 7.0 using the code provided by Wang *et al.* (2007). However, although adaptive-AP results in an optimal number of clusters, this does not indicate that this solution is also the optimal minimum number of clusters that can be obtained from the data. From a biogeographical perspective it should also be remembered that following Wallace we are looking for a moderate number of broad regions. Thus, we used the clusters obtained in this second analytical run to classify every grid cell of the original presence–absence matrices according to the cluster to which they have been assigned in the analysis, calculated the mean of the presences per cluster and generated a second new matrix (and similarity matrix) on which we re-ran adaptive-AP. We repeated this process until the silhouette index approached, or was surprisingly equal to, 1 (Table 1), indicating that we had obtained both the optimal regional configurations and the optimal minimum number of regions.

Table 1 Number of clusters and silhouette index values (*S*) resulting from each run of the affinity propagation (AP) and the adaptive-AP clustering for the three vertebrate groups (birds, mammals and amphibians) at three taxonomic ranks (species, genus and family). Sequential adaptive-AP runs were conducted on each data set until the highest silhouette index was achieved, which in all cases save one was the maximum possible value (= 1).

	Run 1 AP	Run 2 Adaptive-AP	Run 3 Adaptive-AP	Run 4 Adaptive-AP	Run 5 Adaptive-AP
BIRDS					
Species	1017	31 (S = 0.22)	13 (S = 0.34)	13 (S = 1)	
Genus	935	11 (S = 0.26)	8 (S = 0.69)	8 (S = 1)	
Family	802	7 (S = 0.23)	5 (S = 0.58)	5 (S = 1)	
MAMMALS					
Species	1123	66 (S = 0.19)	25 (S = 0.29)	21 (S = 0.80)	21 (S = 1)
Genus	1026	10 (S = 0.23)	8 (S = 0.76)	8 (S = 1)	
Family	793	8 (S = 0.26)	5 (S = 0.53)	5 (S = 1)	
AMPHIBIANS					
Species	790	70 (S = 0.18)	15 (S = 0.15)	15 (S = 1)	
Genus	607	6 (S = 0.28)	6 (S = 1)		
Family	214	2 (S = 0.42)			

RESULTS

The optimal regions using modern genera (eight for mammals and birds and six for amphibians; Table 1) largely coincide with the six zoogeographical regions and boundaries proposed by Wallace (1876) (Fig. 1). As expected, clustering at the species rank yields many more minimum regions (21 for mammals, 13 for birds and 15 for amphibians; Table 1), many of which appear to largely reflect variation in contemporary climate and vegetation types that define subregions rather than regions as defined by Wallace (Fig. 2), whereas the familial rank results in a small number of regions (five for mammals and birds and two for amphibians), generally quite broad, with boundaries that probably reflect evolutionary processes at a range of very deep geological times (see Appendix S1). We present the familial rank results for completeness, but given that neither Wallace nor anyone else has advocated using them to define biogeographical regions we do not consider them further.

DISCUSSION

The six regions found for amphibian genera were strikingly similar to Wallace's if we do not consider areas where most or all genera have been extirpated by Pleistocene and Holocene climates, whereas for mammals and birds we obtained a few new regions (Fig. 1b,c). In the New World, the Nearctic is split in two and a 'subarctic' region appears for birds, whereas for both mammals and birds, temperate South America separates from the tropical zone forming an 'austral' region; we do not define these formally using capitalizations for reasons we will come to shortly. But both the subarctic and austral regions closely resemble subregions defined by Wallace (Fig. 2a), the Canadian and Chilean. In the Old World, a new region emerges for mammals, extending from the Sahara and Arabian deserts across Iran and Pakistan to the Great Indian Desert, clearly due to genera associated with arid conditions (Fig. 1b) and somewhat similar to Wallace's Mediterranean subregion of the Palaearctic, although Wallace's subregion is circum-Mediterranean (Fig. 2a). It is also very similar to the proposed 'Saharo-Arabian Realm' of Holt *et al.* (2013a), formalized after they clustered phylogenetic turnover of amphibian, bird and mammal species together.

It is noteworthy that, setting aside the new 'regions' found for mammals and birds, the major boundaries largely match across the three taxonomic groups and are close to those proposed by Wallace. There are, however, some differences. For example, the northern Ethiopian (now Afrotropics) boundary is located farther south, excluding the southern Arabian Peninsula, whereas the Palaearctic–Oriental boundary across China is extended farther north (Fig. 1). Wallace (1876, Vol. 1, p. 180; 1880, p. 39) explained the difficulty in accurately defining southern Palaearctic boundaries because these often represented regions less well known at that time. Accepting the impossibility of placing the boundaries accurately, he drew the southern Palaearctic borders arbitrarily, considering the Tropic of Cancer as the Palaearctic–Ethiopian boundary and setting the Palaearctic–Oriental boundary in eastern China slightly to the south of Shanghai on the coast. However, more recent taxonomically based studies (Zhang, 1979; Zhang *et al.*, 1989; Zhang & Chen, 1997) have moved the Oriental boundary in China farther north than did Wallace, more in line with the Palaearctic–Oriental boundary shown in our maps. Also, in our results, Wallace's Line is shifted, placing Sulawesi in the Oriental Region instead of the Australian. Again, Wallace (1876, Vol. 1, p. 387; 1880, pp. 421–434) was unsure where to locate the boundary between the Australian and Oriental regions owing to the strong faunal relationships of the island of Sulawesi with both Asia and Australia. In the end he included Sulawesi in the Australian Region, although he pointed out that it could almost as easily belong to the

Oriental. This uncertainty about where to draw the line has continued, with the exact location of Wallace's Line attracting a great deal of discussion (e.g. Mayr, 1944; Simpson, 1977; Vane-Wright, 1991; Cox, 2001; Esselstyn *et al.*, 2010). This is outside the brief of this paper.

Interestingly, for mammals the Palaearctic extends into the northern part of the Western Hemisphere (Fig. 1b), similar to the results of Kreft & Jetz (2010) for mammal families and Holt *et al.* (2013a) using phylogenetic turnover of species. More unusual is the inclusion of Madagascar within the Oriental region (Fig. 1b). Kreft & Jetz (2010) also found Madagascar oddly grouped, but with Australia in their case. They and Cox (2010) argued that this result is not due to the existence of a deep linkage between both areas, but rather to the fact that Australia and Madagascar are very different from the other regions, each having a high level of endemism at higher taxonomic ranks. This is true, and biogeographers have offered alternative explanations for why Madagascar is difficult to classify biogeographically. Wallace (1880, pp. 384–420) described the anomalies in the distributions of mammals in Madagascar and its links with Asia, arguing that the complexity of Madagascar may be because the fauna is derived from two distinct sources: from the African continent through a direct land connection before the greater part of its existing fauna had reached Africa (in the Eocene or Late Cretaceous), and from Asia by means of intervening large islands, now submerged, that connected north-eastern Madagascar with India. Recent studies (e.g. Voelker & Outlaw, 2008, and references therein) have explained current lineage distributions in Madagascar based on two similar suppositions: the so-named Lemurian stepping-stones (Eocene–Oligocene islands or land bridges), which connected India and Sri Lanka to Madagascar via the Seychelles; and vicariance events postdating the breakup of Gondwana. Recently, Parenti & Ebach (2010, 2013a; but see also de Bruyn *et al.*, 2013; Parenti & Ebach, 2013b) provided an explanation based on comparative biogeography, suggesting that Madagascar is a composite of two old biogeographical subregions – the Indo-Malayan and Pandora – that were joined in the early Oligocene (30 Ma). Madagascan freshwater fishes seem to reflect this paradoxical distribution (Sparks & Smith, 2004).

In sum, it appears that the somewhat minor differences we obtained from Wallace, including the positions of the boundaries of the Palaearctic–Afrotropics, Palaearctic–Oriental and Oriental–Australian regions, as well as the placement of Madagascar, were anticipated and discussed by him.

In general, and according to Wallace (1876), both Madagascar and the new regions found for mammals and birds should probably be considered subregions belonging to the broader regions which they are embedded within or adjacent to, because even if they have unique genera, they are too small or species poor (especially in the latter cases) to be considered zoological regions as such. It can be argued that the Saharo-Arabian Realm (*sensu* Holt *et al.*, 2013a) found for mammals is large enough to form a distinct region, but an alternative interpretation is that it is more a shifting transition zone between the Ethiopian, Palaearctic and Oriental regions than a distinct region with many unique groups. Wallace (1876, Vol. 1, p. 70) was well aware of the issues related to the Old Word deserts, claiming that although this expansive area, extending from the Atlantic shores of the Sahara across Arabia to Central Asia, contains a number of desert taxa largely or entirely restricted to it, any attempt to define it as a separate region would introduce confusion. He argued that it is not possible to place geographically or zoologically defined limits to this region, probably because its climate has been unstable over recent time in comparison to the great masses of land with more or less well-defined geographical barriers that constitute the major regions. He reasoned that the main regions had retained considerable geological and climatic stability over a sufficient period of time to allow the development and co-adaptations of faunal assemblages. This instability would also account for why he could not place the Palaearctic–Ethiopian boundary with confidence. It is now known that the border of the Sahara and Sahel has repeatedly shifted southwards and northwards (Foley *et al.*, 2003) and that in some periods (e.g. the early Holocene), the Sahara was replaced entirely by wooded savanna in which giraffe, hippopotamus, elephant and bovids were abundant (de Vivo & Carmignotto, 2004). Thus, as Wallace argued, there is considerable evidence that this area does not have the distinctiveness in its fauna that other, larger regions have and should not be considered as being equivalent to them. Indeed, only 12 of the 128 genera found there are endemic, most of them having narrow ranges and belonging to a few families (e.g. Muridae, Bovidae) or orders (e.g. Rodentia). This also holds for the division of the Nearctic into two regions for birds and mammals. The fauna of Canada is the product of repeated recolonizations following the major glaciations of the Pleistocene (see Dyke *et al.*, 2003, for maps of ice coverage in North America for several of the more recent glaciations), with little or no unique, uninterrupted biogeographical signal dating from the Palaeogene or earlier. Thus, it seems to make little sense to consider this constantly re-mixed fauna, lacking any *in situ* diversification at the generic rank, to be equivalent to the centres of evolution associated with the major biogeographical regions. And because a substantial portion of Canada's mammalian fauna is derived from groups migrating into North America from East Asia during the late Cenozoic (e.g. Morales-Castilla *et al.*, 2012), it is not surprising that it sometimes clusters with the Palaearctic. Finally, the austral region (Chilean subregion of Wallace), which is sometimes distinguished from the rest of the Neotropics (see Morrone, 2002), is relatively small, depauperate and lacks any obvious geographical barrier, and probably largely appears due to the sharp climatic gradient from tropical to temperate latitudes that currently exists and which partitions the fauna into lowland tropical and cold-adapted groups. Wallace clearly realized that southern South America differed from the rest of the continent (Wallace, 1876, Vol. 2, pp. 36–48), but he did not consider these differences sufficient to merit regional status.

As argued by Wallace (1876), results at the species rank, with variable numbers of regions and boundaries, indicate that regions derived from species distributions are to some extent dependent on smaller-scale and shorter-term environmental factors – contemporary climate and vegetation type – constraining the distributions of low-rank taxa. Previous results at the continental scale have also demonstrated that biogeographical units for vertebrate groups based on species are partly determined by climate and habitat (Heikinheimo *et al.*, 2007; Rueda *et al.*, 2010), which also suggests linkages between biological traits and patterns of regionalization at finer spatial and taxonomic scales. For example, limited dispersal ability and ectothermy probably make amphibians more susceptible to climatic conditions and generate more localized clusters of species than for birds or mammals (Rueda *et al.*, 2010). Thus, regionalization patterns obtained at the species rank are too dependent on the taxonomic group, rendering many small regions, and should not be used to assess Wallace's scheme, although this rank possesses the highest conservation value (Whittaker *et al.*, 2005).

Using modern analytical techniques and our current knowledge of distributions and generic concepts, we have shown that when Wallace's principles are followed it is possible to obtain a map of zoological regions quite similar to the map he generated more than a century ago. Thus, mismatches between the recent 'updates' and Wallace's scheme are not necessarily wholly due to the use of modern analytical techniques per se but also reflect that they have not used the same criteria for classification. Hence, Holt *et al.* (2013a), for example, identified a total of 20 regions nested within 11 large realms. However, their classification was based on species-level distributions and, in the final classification scheme, combined three vertebrate groups. Moreover, they included in their zoogeographical map new realms such as Madagascan, Sino-Japanese and Panamanian, which under Wallace's classification criteria are too small to be considered regions and purposefully compound influences on the distribution of taxa operating over relatively recent time-scales with those operating over deeper time. Procheş & Ramdhani (2012) used generic-level distributions but ecoregions as the geographical units, which are somewhat coarse and ensure that the outcome will include ecological (vegetation types) effects. They obtained 11 regions and again provide as an update of Wallace's scheme new island regions too small to be considered as such by Wallace, including the Madagascan, Wallacean, Caribbean and New Guinean regions. Finally, Kreft & Jetz (2010) compared species, generic and familial rank data for mammals, mapping both volant and non-volant species. However, they only mapped the first 12 groups obtained by their UPGMA (unweighted pair-group method using arithmetic averages) clustering procedure, and although they provide an optimal number of clusters for the multi-rank data, they did not map the optimal generic rank patterns, so direct comparison with Wallace's scheme is not possible. In any case, their final map shows 30 regions generated at the species level, greatly differing in size – from tiny

New Caledonia to the expansive Euro-Siberian region – that are nested in six major realms where North Africa, the Sahara, the Arabian Peninsula and the Middle East are parts of the Afrotropics but from which Madagascar is excluded.

All of the regions generated by the recent spate of analyses are spatially cohesive and computationally accurate, and is not our intention to criticize them. In fact, they provide good examples of some of the issues involved with attempting global regionalizations, showing how results depend on the unit of analysis (e.g. ecoregions versus regular grids), the taxonomic rank, the distance metric (beta diversity versus phylogenetic beta diversity) and/or the clustering method (see also Holt *et al.*, 2013b; Kreft & Jetz, 2013). However, in view of our results they should not be considered updates of Wallace's zoological regions, as they are not tests of his scheme at all. Rather, they are original schemes that seem to have little to do with what Wallace was trying to achieve when he divided the world into biogeographical regions.

So, what was Wallace trying to achieve? He later defined zoological regions as 'those primary divisions of the earth's surface of approximately continental extent, which are characterized by distinct assemblages of animal types' (Wallace, 1894, p. 613), and he considered them useful for facilitating the study of the existing geographical distributions of animals as a consequence of evolution by common descent and the geological history of the Earth. As time passes new challenges arise, and the current aims of global regionalization of vertebrates often have more to do with conservation planning than with evolution or history. In regionalizations with the former goal, many or most of Wallace's tenets may not be relevant, and workers can use any method they feel appropriate to the question. But when the goal of a new analysis is to generate 'biogeographical regions' *sensu* Wallace, we should perhaps be more circumspect in claiming that each new classification scheme generated via a new methodology or turnover metric is an updated version of his scheme. There are at least two reasons for this: first, none of the computer-based approaches have followed Wallace's criteria, and second, even when using state-of-the-art analytical techniques and greatly modified animal classification systems, we find that when identifying global zoological regions the principles devised by Wallace more than 135 years ago can lead to virtually the same patterns today.

We close by turning the above question towards ourselves. What are we trying to achieve by repeatedly revisiting the issue of global regionalization from a biogeographical perspective? Clearly, it is perceived to be an important issue by biogeographers, but what are we doing when we rely on computers to answer the question? Wallace had relatively limited taxonomic and distributional information and the most basic analytical tools, whereas we can download range maps, select among metrics to generate matrices of 'distance' or 'turnover' and feed this grist into the mill of computerized clustering methods. The outcome of this approach is becoming clear: every methodological decision we make generates one or more possible schemes. Are these really

improvements on Wallace's classification? He utilized keen evolutionary and geological insights based on both extant and extinct groups [the latter being an important source of information missing from all computer-based analyses] to generate a global map. It is possible to very nearly reproduce his map using a clustering algorithm and excluding fossils, but we could also change the taxa, the distance metric or the clustering method and generate a range of scenarios. Would these variations give us any insights into the underlying biogeography or just generate more confusion? Perhaps deepening our understanding of biogeographical regions does not lie in a computer algorithm but requires information on biology, evolution and the history of the Earth; that is, the use of standard biogeographical methods. Even if true, it seems evident that Wallace had the ability to devise a scheme that can be recovered by a computer algorithm more than a century later. This strikes us as being remarkable.

ACKNOWLEDGEMENTS

REFERENCES

Bodenhofer, U., Kothmeier, A. & Hochreiter, S. (2011) AP-Cluster: an R package for affinity propagation clustering. *Bioinformatics*, **27**, 2463–2464.

de Bruyn, M., Stelbrink, B., Page, T.J., Phillips, M.J., Lohman, D.J., Albrecht, C., Hall, R., von Rintelen, K., Ng, P.K.L., Shih, H.-T., Carvalho, G.R. & von Rintelen, T. (2013) Time and space in biogeography: response to Parenti & Ebach (2013). *Journal of Biogeography*, doi:10.1111/jbi.12166.

Cardille, J.A. & Lambois, M. (2009) From the redwood forest to the Gulf Stream waters: human signature nearly ubiquitous in representative US landscapes. *Frontiers in Ecology and the Environment*, **8**, 130–134.

Cox, C.B. (2001) The biogeographic regions considered. *Journal of Biogeography*, **28**, 511–523.

Cox, C.B. (2010) Underpinning global biogeographical schemes with quantitative data. *Journal of Biogeography*, **37**, 2027–2028.

Dyke, A.S., Moore, A. & Robertson, L. (2003) *Deglaciation of North America*. Geological Survey of Canada Open File 1574. Natural Resources Canada, Ottawa.

Esselstyn, J.A., Oliveros, C.H. & Moyle, R.G. (2010) Integrating phylogenetic and taxonomic evidence illuminates complex biogeographic patterns along Huxley's modification of Wallace's Line. *Journal of Biogeography*, **37**, 2054–2066.

Foley, J.A., Coe, M.T., Scheffer, M. & Wang, G. (2003) Regime shifts in the Sahara and Sahel: interactions between ecological and climatic systems in Northern Africa. *Ecosystems*, **6**, 524–539.

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

Frey, B.J. & Dueck, D. (2008) Response to comment on "Clustering by passing messages between data points". *Science*, **319**, 726.

Gagné, S. & Proulx, R. (2009) Accurate delineation of biogeographical regions depends on the use of an appropriate distance measure. *Journal of Biogeography*, **36**, 561–567.

Heikinheimo, J., Fortelius, M., Eronen, J. & Mannila, E. (2007) Biogeography of European land mammals shows environmentally distinct and spatially coherent clusters. *Journal of Biogeography*, **34**, 1053–1064.

Holt, B.G., Lessard, J.-P., Borregaard, M.K., Fritz, S.A., Araújo, M.B., Dimitrov, D., Fabre, P.-H., Graham, C.H., Graves, G.R., Jønsson, K.A., Nogués-Bravo, D., Wang, Z., Whittaker, R.J., Fjeldså, J. & Rahbek, C. (2013a) An update of Wallace's zoogeographic regions of the world. *Science*, **339**, 74–78.

Holt, B.G., Lessard, J.-P., Borregaard, M.K., Fritz, S.A., Araújo, M.B., Dimitrov, D., Fabre, P.-H., Graham, C.H., Graves, G.R., Jønsson, K.A., Nogués-Bravo, D., Wang, Z., Whittaker, R.J., Fjeldså, J. & Rahbek, C. (2013b) Response to comment on "An update of Wallace's zoogeographic regions of the world". *Science*, **341**, 343.

Kreft, H. & Jetz, W. (2010) A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, **37**, 2029–2053.

Kreft, H. & Jetz, W. (2013) Comment on "An update of Wallace's zoogeographic regions of the world". *Science*, **341**, 343.

Legendre, P. & Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.

Legendre, P. & Legendre, L. (1998) *Numerical ecology*, 2nd edn. Elsevier, Amsterdam.

Mayr, E. (1944) Wallace's Line in the light of recent zoogeographic studies. *The Quarterly Review of Biology*, **19**, 1–4.

Morales-Castilla, I., Olalla-Tárraga, M.Á., Purvis, A., Hawkins, B.A. & Rodríguez, M.Á. (2012) The imprint of Cenozoic migrations and evolutionary history on the biogeographic gradient of body size in New World mammals. *The American Naturalist*, **180**, 246–256.

Morrone, J.J. (2002) Biogeographical regions under track and cladistic scrutiny. *Journal of Biogeography*, **29**, 149–152.

Parenti, L.R. & Ebach, M.C. (2010) Wallacea deconstructed. *Beyond cladistics: the branching of a paradigm* (ed. by D.M. Williams and S. Knapp), pp. 303–318. University of California Press, Berkeley, CA.

Parenti, L.R. & Ebach, M.C. (2013a) Evidence and hypothesis in biogeography. *Journal of Biogeography*, **40**, 813–820.

Parenti, L.R. & Ebach, M.C. (2013b) The explanatory power of biogeographical patterns: a reply to de Bruyn *et al. Journal of Biogeography*, doi:10.1111/jbi.12205.

Proches, Ş. & Ramdhani, S. (2012) The world's zoogeographical regions confirmed by cross-taxon analyses. *BioScience*, **62**, 260–270.

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

Rousseeuw, P.J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, **20**, 53–65.

Rueda, M., Rodríguez, M.Á. & Hawkins, B.A. (2010) Towards a biogeographic regionalization of the European biota. *Journal of Biogeography*, **37**, 2067–2076.

Schmidt, K.P. (1954) Faunal realms, regions, and provinces. *The Quarterly Review of Biology*, **29**, 322–331.

Sclater, P.L. (1858) On the general geographical distribution of the members of the class Aves. *Journal of the Proceedings of the Linnean Society: Zoology*, **2**, 130–145.

Simpson, G.G. (1977) Too many lines: the limits of the Oriental and Australian zoogeographic regions. *Proceedings of the American Philosophical Society*, **121**, 107–120.

Smith, C.H. (1983) A system of world mammal faunal regions. I. Logical and statistical derivation of the regions. *Journal of Biogeography*, **10**, 455–466.

Sparks, J.S. & Smith, W.L. (2004) Phylogeny and biogeography of the Malagasy and Australasian rainbowfishes (Teleostei: Melanotaenioidei): Gondwanan vicariance and evolution in freshwater. *Molecular Phylogenetics & Evolution*, **33**, 719–734.

Vane-Wright, R.I. (1991) Transcending the Wallace line: do the western edges of the Australian region and the Australian plate coincide? *Australian Systematic Botany*, **4**, 183–197.

de Vivo, M. & Carmignotto, A.P. (2004) Holocene vegetation change and the mammal faunas of South America and Africa. *Journal of Biogeography*, **31**, 943–957.

Voelker, G. & Outlaw, R.K. (2008) Establishing a perimeter position: speciation around the Indian Ocean Basin. *Journal of Evolutionary Biology*, **21**, 1779–1788.

Wallace, A.R. (1876) *The geographical distribution of animals, Vols 1 & 2*. Harper & Brothers, New York.

Wallace, A.R. (1880) *Island life*. Harper & Brothers, New York, NY.

Wallace, A.R. (1894) What are zoological regions? *Nature*, **49**, 610–613.

Wang, K., Zhang, J., Li, D., Zhang, X. & Guo, T. (2007) Adaptive affinity propagation clustering. *Acta Automatica Sinica*, **33**, 1242–1246.

Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–23.

Zhang, E. & Chen, Y. (1997) Fish fauna in Northeastern Jiangxi Province with a discussion on the zoogeographical division of East China. *Acta Hydrobiologica Sinica*, **21**, 260–266.

Zhang, J.-T., Liu, C.-Y. & Wu, H.-Y. (1989) Studies of the boundary line between Palaearctic and Oriental Regions in central China in relation to the zoogeography of Siphonaptera. *Acta Zootaxonomica Sinica*, **4**, 020.

Zhang, R. (1979) *Physical geography of China: zoogeography*. Science Press, Beijing.

SUPPORTING INFORMATION

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

Appendix S1 Familial rank regions.

BIOSKETCHES

Marta Rueda is a post-doctoral fellow interested in biogeography and global ecological and evolutionary patterns. Her current research interests also include fragmentation theory and community phylogenetics at broad scales.

Miguel Á. Rodríguez is interested in the processes conditioning patterns of biodiversity at regional and global scales. His recent research has involved investigating the effects of habitat destruction and fragmentation on faunas and floras at broad scales.

Bradford A. Hawkins is interested in biogeography and geographical ecology, with a major emphasis on linking ecological and evolutionary processes across spatial scales.

Editor: Michael Patten