---

Heikinheimo, Hannes; Eronen, Jussi T.; Sennikov, Alexander; Preston, Christopher D.; Oikarinen, Emilia; Uotila, Pertti; Mannila, Heikki; Fortelius, Mikael. 2012 Convergence in the distribution patterns of Europe's plants and mammals is due to environmental forcing. *Journal of Biogeography*, 39 (9). 1633-1644. 10.1111/j.1365-2699.2012.02723.x

Please cite as: Heikinheimo, H., Eronen, J.T., Sennikov, A., Preston, C.D., Uotila, P., Mannila, H., Fortelius, M. 2012. Converge in distribution patterns of Europe's plants and mammals is due to environmental forcing. Journal of Biogeography 39, 1633-1644

**Strapline: Original Article**

**Running header: Convergence in European plant and mammal distribution patterns**

**Title:**

# Convergence in the distribution patterns of Europe's plants and mammals is due to environmental forcing

**Authors:**

Hannes Heikinheimo 1, Jussi. T. Eronen* 2, 3 Alexander Sennikov 4, Christopher D. Preston 5, Emilia Oikarinen 1, Pertti Uotila 4, Heikki Mannila 1, and Mikael Fortelius 2, 4

*1. HIIT, Department of Computer and Information Science, Aalto University, P.O. Box 15400, FI-00076 Aalto, Espoo, Finland; 2. Department of Geosciences and Geography, P.O. Box 64, FI-00014 University of Helsinki, Helsinki, Finland; 3. Biodiversity and Climate Research Centre LOEWE BiK-F, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany; 4. The Finnish Museum of Natural History, P.O. Box 26, FI-00014 University of Helsinki, Helsinki, Finland; 5. NERC Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK.*

* Correspondence: J. T. Eronen, Department of Geosciences and Geography, P.O. Box 64, FI-00014 University of Helsinki, Helsinki, Finland. E-mail: Jussi T.Eronen@helsinki.fi

## ABSTRACT

**Aim** Our aims were to test: (1) the extent to which vascular plant associations are related in space to mammalian associations, and (2) whether the plant associations are more closely related than the mammalian associations to climate and to a published environmental stratification of Europe.

**Location** Europe, as defined by the following boundaries: 11˚ W, 32˚ E, 71˚ N and 35˚ N.

**Methods** The analysis is based on presence/absence records of mammal species and plant species with a resolution of 50 km × 50 km. The similarity of the overall spatial structure was tested using a partial Mantel test while controlling for the effect of geographical proximity. To further identify the main spatial components in the datasets, we used k-means clustering and principal components analysis. The resulting geographical patterns were compared with one another, with climate variables, and with the environmental stratification of Europe.

**Results** The clustering of the plant data forms coherent areas that can be interpreted as reflections of floristic regions that are controlled to a large extent by climate and topography. In terms of the correlation between distance matrices, the relationship between plants and mammals is relatively strong. The relationships between mammals and climate, and between plants and climate, are more complex but always statistically significant. There is no evidence that the plant clusters are more closely related to climate than are the mammalian clusters, although plant clusters are closer to environmental data than to climate.

**Main conclusions** The clustering patterns of mammals and plants form groups that agree with one another in their spatial extent. The forcing of floristic patterns into coherent entities appears mainly to be caused by climatic variables (temperature, temperature range and rainfall), mediated by elevation differences. The formation of individual plant clusters is also related to species numbers and to local and regional floristic differences. The close correlation between the floristic and faunal patterns suggests that the mammal and plant distributions are controlled by the same environmental variables, although the extent to which the mammals are controlled directly by climate or through the influence of vegetation requires more detailed study.

**Keywords**

**Climate, cluster analysis, Europe, mammals, presence/absence data, principal components analysis, species distribution, vascular plants.**

## INTRODUCTION

The main link between ecology and biogeography lies in the factors and processes that control the spatial distribution of populations and species (e.g. Rosenzweig, 1995; Hanski, 1998; Polis, 1999; Olff *et al.*, 2002). Awareness of accelerated global change has brought the empirical study of distribution boundaries of individual species into sharp and urgent focus, and predicting changes in the spatial distribution of threatened species is becoming a routine part of applied conservation biology (e.g. Malcolm *et al.*, 2002, 2006; Midgley *et al.*, 2002; Thomas *et al.*, 2004). The question of whether associations of species, such as communities or metacommunities (see Leibold *et al.*, 2004; Holyoak *et al.*, 2005), can be meaningfully regarded as entities with distinct spatial boundaries is less clear. Here we are not interested in specific boundaries but on how associations of species are related to each other across Europe. For a recent review of spatial boundaries, see Whittaker *et al.* (2005).

In a previous study (Heikinheimo *et al.*, 2007), we used presence/absence data on European land mammals to show that two independent clustering methods produced highly coherent spatial patterns based on taxonomic occurrence alone. The results were especially interesting as the clustering methods used did not take geography (spatial adjacency of grid cells) into account in any way but still produced spatially coherent clusters. The pattern was shown to be strongly related to climate variables and similar to an independently derived environmental zonation of the same area (Metzger *et al.*, 2005). It was similar for multiple subsets of mammalian data but showed minor differences that could be related to trophic level and dispersal characteristics. We interpreted the pattern of clusters as reflecting the spatial expression of biologically distinct, metacommunity-like entities (see Leibold *et al.*, 2004; Holyoak *et al.*, 2005) and concluded that their boundaries are mainly defined by factors that are related to the physical environment.

The results from our previous study suggested that the herbivore subset produces the pattern that is most similar to the environmental zonation described by Metzger *et al.* (2005), whereas the patterns for carnivores and omnivores showed markedly lower similarity (Table 2 in Heikinheimo *et al.*, 2007). An obvious interpretation of this result is that herbivores have the most direct dependence on vegetation, which, in turn, depends on climate.

In a parallel study using a different method, but again one which did not take geography explicitly into account, Finnie *et al.* (2007) showed that a sample of European vascular plants also clustered into floristic elements which had spatially coherent concentrations. They classified species rather than grid cells, distinguishing elements in northern and temperate Europe dominated by wide-ranging species and including few European endemics, and more geographically restricted elements in southern Europe which were rich in endemic species. They did not explore the relationship of the clusters to environmental factors in any detail.

Here, we use the same methods as those employed by Heikinheimo *et al.* (2007) to analyse presence/absence data for a somewhat larger sample of the vascular plants of Europe. We then test: (1) how strongly the mammalian associations are related in space to plant associations, and (2) whether the plant associations are more closely related than the mammalian associations to climate and to a published environmental stratification of Europe (Metzger *et al.*, 2005).

## MATERIALS AND METHODS

### Plant data

The plant dataset used in this study was derived from volumes 1–13 of *Atlas Florae Europaeae* (Jalas & Suominen, 1972–1994; Jalas *et al.*, 1996, 1999; Kurtto *et al.*, 2004). The details of the Atlas project (Uotila *et al.*, 2005) and the associated database (Lahti & Lampinen, 1999) are given elsewhere. The *Atlas*

*Florae Europaeae* (AFE) grid system was changed in 2000 (after AFE volumes 1–12). The new grid system is modified from the Universal Transverse Mercator (UTM) co-ordinates and the Military Grid Reference System (MGRS), as defined in the official documents of the U.S. National Imagery and Mapping Agency (see http://www.luomus.fi/english/botany/afe/map/grid.htm). The data from AFE volumes 1–12 have been transformed to the new grid system of the Atlas on which volume 13 was based.

The taxonomy employed follows the original taxonomy and nomenclature of *Flora Europaea* (Jalas & Suominen, 1972–1994; Jalas *et al.*, 1996, 1999; Kurtto *et al.*, 2004), which was updated and revised in the course of mapping. We processed the data at the species level and aggregated records of segregates that were not accepted by *Atlas Florae Europaeae* and of subspecies into the appropriate species. We included a few aggregate species or species complexes in cases where the distributions of the component species were not mapped or were particularly imperfectly recorded. The database includes a total of 3086 species or aggregates, comprising approximately 20% of the European flora, of which only the 2924 species or aggregates that had been mapped as natives or archaeophytes were further considered in our study. Whereas the original plant dataset (4750 grid cells) covers the whole of Europe up to the Urals watershed, the alignment of plant data with mammal and climate datasets (see below) retained 2549 plant taxa in 2179 grid cells for inclusion in the analysis.

The potential geographical bias involving the use of plant data from the first volume of *Atlas Florae Europaeae* was discussed by Finnie *et al.* (2007). They showed that, for most territories, there was no significant difference between the proportion of species in this volume and that in the European flora as a whole. However, species in the first volume were over-represented in parts of northern and eastern Europe [Faroe Islands, Finland, Iceland, Russia (East and North) and especially Svalbard]. In our analysis, all of these territories except Finland have been excluded from the dataset, so the bias is considerably reduced. The 99 species that were excluded as a result of the alignment of the plant and mammal datasets occur in 1–2 (rarely 3–5) Flora Europaea territories. Most are from the eliminated territories of Svalbard (17), Russia (North) (12) and the Azores (16). Within the area analysed, most losses were from the Mediterranean islands of Crete (9), Greece (25), Sicily (9) and the Balearics (8); these are species-rich areas, and the loss of these few species is unlikely to have affected the results of the analysis. It is more difficult to analyse the potential ecological biases in the subset of species included in the analysis.

**Mammal data**

We used mammal data collected by the Societas Europaea Mammalogica (http://www.european-mammals.org/) to prepare the *Atlas of European mammals* (Mitchell-Jones *et al.*, 1999). The data consist of presence/absence records of 194 mammal species for a set of 2670 grid cells covering Europe. The grid system is the same as that used by *Atlas Florae Europaeae*.

In the present study, we exclude all records of bats, aquatic mammals, *Rattus* and *Mus* and all mammals not native to Europe except *Nyctereutes*, following Heikinheimo *et al.* (2007). We also exclude all grid cells with fewer than eight species in the original dataset.

**Climate and environmental data**

For comparison with the biological data, we use climate data (Hijmans *et al.*, 2005) and environmental zonation (Metzger *et al.*, 2005), as in Heikinheimo *et al.* (2007). The climate data are available online at http://www.worldclim.org. The data consist of global climate layers with four different spatial cell resolutions: 30 arc-seconds (0.93 km × 0.93 km = 0.86 km$^2$ at the equator) and 2.5, 5 and 10 arc-minutes (18.6 km × 18.6 km = 344 km$^2$ at the equator). We associated the climate values with the UTM grid by taking an average over the 10arc-minute cells occurring within each UTM grid cell.

The data include the monthly averages of four basic climate variables for all 12 months. The variables are mean temperature, precipitation, minimum temperature, and maximum temperature. The data include 19 additional bioclimatic variables derived from these basic climate variables: annual mean temperature, mean diurnal range, isothermality, temperature seasonality, maximum temperature of the warmest month, minimum temperature of the coldest month, annual temperature range, mean temperature of the wettest quarter of the year, mean temperature of the driest quarter, mean temperature of the warmest quarter, mean temperature of the coldest quarter, precipitation of the wettest month, precipitation of the driest month, precipitation seasonality, precipitation of the wettest quarter, precipitation of the driest quarter, precipitation of the warmest quarter, and precipitation of the coldest quarter. Hence, the final dataset included a total of 67 climate variables. The records are from the period 1950–2000.

In addition to the climate data, we also used the environmental zonation (EnZ) of Metzger *et al.* (2005), which is available in the UTM grid format for comparison with the other datasets. This zonation is based on records of the minimum temperature, maximum temperature, precipitation, and percentage of sunshine for the months of January, April, July and October, in addition to values of altitude, slope, latitude, and oceanity. For further details, see Metzger *et al.* (2005). In the following text, we refer to the Hijmans *et al.* (2005) dataset as climate data and to the Metzger *et al.* (2005) dataset as environmental data.

### Elevation data

The original climate data (Hijmans *et al*., 2005) also include altitude information (elevation above sea level). For interpretation purposes, we are interested in the geographical complexity of each grid cell. To estimate this, we computed the standard deviation of the 30-arc-second cells that occur within each UTM grid cell. We also calculated an average mean elevation for each grid cell. We note that for the studied geographical area, the two variables are highly correlated (a Pearson correlation coefficient of 0.82).

### Dataset alignment

A primary aim of this study is to compare the results obtained for mammals by Heikinheimo *et al.* (2007) with the patterns observed from the plant data. Hence, we selected only the subset of 2179 grid cells that contained data for all of our datasets (plant, mammal, climate, environment). The dataset alignment with plant data removed four grid cells from the mammal data used in Heikinheimo *et al*. (2007), covering 2183 grid cells in total. Therefore, we calculate all of the metrics and perform cluster analysis directly using this aligned data for all of the datasets.

**Data transformations and distance measures**

For the analysis, each climate variable was standardized to unit-less variables with a mean of 0 and variance of 1 to cancel out the scales of different units of measurement (Legendre & Legendre, 1998). For both the mammal and the plant data, we used the species presence/absence data directly, that is, each grid cell was represented as a 0–1 vector. For this study, we calculated all distance matrices de novo for all data.

All analyses were performed using the Euclidean distance. In view of recent correspondence concerning the use of different distance measures (Gagné & Proulx, 2009; Heikinheimo *et al.*, 2009), we also performed the analysis using the Hellinger distance (Rao, 1995). There is a close concordance between the results obtained in analyses using Euclidean and Hellinger distances (Heikinheimo *et al.*, 2009), which we also confirmed in our current analyses (both similarity matrices and cluster maps).

**Analysis of similarity in overall spatial structure**

We used the Pearson correlation between the similarity of grid cell pairs to compare the overall spatial structure in the mammal, plant and climate datasets. Thus, we first computed a distance matrix for each of the three datasets and then calculated the correlation coefficients for each distance matrix pair. For the mammal and plant datasets, each value in the distance matrix was based on species occurrence vectors. In the case of the climate data, the values of the distance matrix were based on the values of the standardized climate variables.

To control for the effect of geographical proximity (spatial autocorrelation), we computed a fourth distance matrix based on geographical distance. A distance value in kilometres was obtained for each grid cell pair by applying the Haversine formula (Sinnott, 1984) to the UTM latitude and longitude co-ordinates of the centres of the grid cells. We then recalculated the correlations between the mammal, plant and climate distance matrices while controlling for geographical distance using the partial correlation coefficient (Legendre & Legendre, 1998). The partial correlation observed between two variables, *X* and *Y*, while controlling for variable(s) *Z*, can be thought of as the correlation between the residuals of *X* and *Y* when regressing with *Z*. Furthermore, we also computed the correlation between mammals and plants while controlling for both geographical distance and climate.

To test the significance of the distance matrix-based correlations, we performed a partial Mantel test (Legendre & Legendre, 1998) using 100 permutation rounds for each of the distance matrix pairs while controlling for geographical distance and the additive combination of geographical distance and climate. All calculations were performed with Matlab (MATLAB 2009)

## Principal components analysis

We used principal components analysis (PCA; Sharma, 1996) to study the dominant features of the datasets in terms of variance. We computed the first three principal components for the mammal, plant and climate datasets and then computed the Pearson correlations between the components and some known variables, such as species counts, elevation, and certain climatic variables. Because PCA can potentially suffer from a horseshoe effect on species presence/absence data, we used non-metric multidimensional scaling (NMDS; Shepard 1962a,b; Kruskal, 1964) as an alternative method to confirm the PCA results. We computed NMDS in three dimensions as implemented in Matlab (MATLAB 2009) using Euclidean distance and the initial configuration of a classical multidimensional scaling solution. To allow for zero distances, we used sstress (squared stress normalized with the sum of the 4th powers of the inter-point distances) as the goodness of fit function.

## Clustering methods

We used the k-means (Duda *et al.*, 2000; Theodoridis & Koutroumbas, 2003) clustering method to obtain a clustering of the grid cells for the plant, mammal and climate data. The k-means clustering method is based on an iterative process, and the final clusterings for each dataset were selected as the best out of 100 clustering runs in terms of squared error (sum over the distances of data points from their corresponding cluster centre) to avoid problems of local minima. All calculations were performed with Matlab (MATLAB 2009)

The similarity of the clusterings was compared using the Kappa statistic (Monserud & Leemans, 1992). To evaluate the Kappa statistic, we used the qualitative guidelines of Monserud and Leemans as implemented in Metzger *et al.* (2005): a Kappa value of less than 0.2 represents very poor agreement, 0.2–0.4 poor, 0.4–0.55 fair, 0.55–0.7 good, 0.7–0.85 very good, and greater than 0.85 excellent agreement. As a technical detail, note that before the Kappa statistic can be computed, it must be decided which clusters correspond to one another in the two clusterings of the different groups being compared. This matching was performed so that the aggregate geographical overlap between the matched clusters was maximized. For this, we used the minimum-cost perfect matching algorithm described in detail by Kleinberg & Tardos (2005).

To circumvent the requirement in k-means clustering to set the number of clusters a priori, we varied the number of clusters from 2 to 13 for plant data. We also computed an agglomerative clustering using Ward's linkage for plant data. For 3–12 clusters, the k-means clusterings and hierarchical clustering are in good or very good agreement based on the Kappa statistic (0.64–0.81); for 13 clusters, the Kappa value of 0.44 indicates fair agreement; and for 2 clusters, we obtained poor agreement (Kappa value of 0.35).

Thus, k-means produces clusterings that are comparable to those obtained using a hierarchical clustering method for our data.

For the comparison between the clusterings obtained for the mammal and climate datasets, we chose 12 as the number of clusters. This allowed a comparison with the environmental zonation (EnZ) of Metzger *et al.* (2005). In their study, the geographical region is divided into 13 environmental zones, but the areas covered by their Anatolian zone are not included in our data, leaving 12 environmental zones in total.

## RESULTS

### Similarity in overall spatial structure

There is a statistically significant relationship (Pearson's $r = 0.60$) in the overall spatial structure between the plant and mammal data when assessed by the between-grid cell pair similarity while controlling for the effect of geographical proximity (Table 1). A similar result (Pearson's $r = 0.61$) is obtained when using the Hellinger distance for the similarity assessment between grid cell pairs. When further controlling for the additive effect of both geographical distance and climate, the correlation drops only very slightly: to 0.59 using Euclidean and to 0.56 using Hellinger distances. The distance matrix that was computed from the patterns in the climate data is related to geographical proximity (Pearson's $r = 0.68$; Table 1).

There is also a statistically significant, albeit weak, relationship in the correlation between the spatial patterns of the climate data and both the mammal and plant data (Table 1) after controlling for geographical proximity.

### Principal components analysis

The three first principal components explain 32% and 40% of the total variation for plants and mammals, respectively (Table 2). For climate, this value is 89%; however, it should be noted that the climate dataset has 67 dimensions (variables), which is much lower than the 2549 dimensions (variables, i.e. individual species) in the plant dataset and 124 dimensions in the mammal dataset. Furthermore, the 67 dimensions in the climate data are based on only four basic climate variables and their annual variation.

The first principal component (PC1) for both plants and mammals correlates with the respective species counts (0.81 and 0.83 for plants and mammals, respectively) in the grid cells (Table 3, Fig. 1). Comparison of the plant PC2 and mammal PC2 with the climate variables shows that they are correlated with mean annual temperature. Of all the climate variables, PC3 of both plants and mammals is most strongly correlated with the annual temperature range (Pearson's $r$ = 0.54 and 0.45 for plants and mammals, respectively). Furthermore, the plant PC3 correlates with the mean elevation (Pearson's $r$ = 0.57), whereas for mammals the correlation with elevation is weak.

PC1 of the climate data almost exactly reflects the mean annual temperature (Pearson's $r$ = 0.99). This is a similar result to that obtained by Metzger *et al.* (2005). Indeed, our PC1 computed from the climate data closely resembles the PC1 of Metzger *et al.* (2005). Additionally, our PC2 almost exactly reflects the pattern of mean annual precipitation (Pearson's $r$ = 0.97), which is again similar to the PC3 of Metzger *et al.* (2005). The relationship with annual temperature range is also fairly strong for PC2 (Pearson's $r$ = 0.70). The interpretation of PC3 for the climate data is less clear; however, it is correlated with the PC1 for mammals (Pearson's $r$ = 0.64) and plants (0.68), as well as the PC3 for plants (0.33) (Table 4).

The correlations between the respective principal component coefficients of the plant and mammal data (that is, the mammal PC1 vs. the plant PC1, the mammal PC2 vs. the plant PC2, and the mammal PC3 vs. the plant PC3) are high (Table 4, Fig. 1). Furthermore, both the mammal PC2 and plant PC2 are highly correlated with the climate PC1, and both the mammal PC1 and plant PC1 are correlated with the climate PC3 (Table 4). The remaining correlations are weaker.

The PCA results obtained using the Hellinger distance are highly concordant with the results reported here using the Euclidean distance (see Appendix S1 in Supporting Information). The percentage of variation explained for the first three components is larger when using the Euclidean distance (for comparison, using the Hellinger distance, the first three principal components explain 31% and 35% of the total variation for plants and mammals, respectively). We also observed that the first and second principal components swap places when using the Hellinger distance (Appendix S1), and that the correlation with the species counts becomes weaker (Appendix S2).

The results from the NMDS confirm those of the PCA. Each of the principal components has quite a high correlation with one of the three NMDS axes that were computed using the Euclidean distance both for mammals (Pearson's $r = 0.95$, 0.95 and 0.93) and plants (Pearson's $r = 0.79$, 0.80 and 0.58). The sstress values for the NMDS for the Euclidean distance are 0.15 and 0.16 for mammals and plants, respectively, indicating a fair representation of data using NMDS with three dimensions.

**Clusterings**

The plant clusters are spatially very coherent (well connected), even though the clustering methods use only presence/absence data. This is very similar to the mammal pattern (Heikinheimo *et al*., 2007). By comparing the plant clustering results and the clustering results for mammals computed as in Heikinheimo *et al.* (2007) for 12 clusters, we obtain Kappa values indicating a fair agreement for most correlations (Table 5). Moreover, the data for herbivorous mammal species yields a larger Kappa value (0.51) than that for non-herbivorous mammal species (0.42). While the correlations between the principal components are much stronger than the correlation values between the Euclidean distance matrices, the Kappa values are often only 'fair' (Table 5). When comparing the plant clustering with the environmentally based clustering of the environmental zones (EnZs; Metzger *et al.,* 2005), we also obtain a fair agreement (0.51) for the Kappa value. For the clustering based on climate data, a Kappa value of 0.45 is obtained. When using the Hellinger distance, the results are similar; see Appendix S3.

## DISCUSSION

### Plant clusters

The PCA together with the Kappa comparisons forms a strong basis for the interpretation of the plant clusters. The clustering of the plant data forms coherent areas that can be interpreted as reflections of floristic regions that are controlled to a large extent by climate and topography and perhaps also by historical factors, resulting in “a patchy” pattern in some regions. The general pattern of the clusters remains almost the same as the initial clusterings of 3 and 4 clusters with continuous splitting of the original large territories in the subsequent increases of cluster numbers. The extensive, mostly lowland territories of central and northern Europe display the prominently zonal character of the spatial classification. In more empirical terms, the patterns in clusters 2 and 3 indicate a large land mass with many species presenting similar distributions (Fig 2). Species with Mediterranean affinities are absent from these areas. In addition, the more oceanic parts of western Europe and the Mediterranean area show a regional patchy character of the classification. This mirrors the analysis of floristic elements performed by Finnie *et al.* (2007), which identified wide-ranging central and northern European elements and much more geographically restricted elements in southern Europe.

The most significant limit in northern Europe in the present regionalization is Limes Norrlandicus (Fransson, 1965), which separates the areas with species-poor northern floras from the southern floras that are rich in species associated with deciduous forests. The position of the southern limit of the Fennoscandian cluster varies depending on the number of clusters specified, fluctuating in Finland from the oak line (Kalela, 1958) to the northern limit of the herbaceous “oak flora” (Lippmaa, 1940). This same limit is visible in the mammal clustering of Heikinheimo *et al.* (2007).

Our classification appears to follow national boundaries to a greater extent than most phytogeographical classifications, notably on the eastern boundary of the primarily “French cluster” that is visible in clustering with 8 clusters and onwards (Fig. 2). The underlying causality is potentially complex and involves several factors, among which national differences in recording practices is only one. Political boundaries often follow natural barriers (e.g. mountainous regions, infertile lands), which are also reflected in species level patterns (e.g. *Carpinus betulus, Pulsatilla vulgaris*; see also the discussion of animal hybrid zones in Heikinheimo *et al.*, 2007, p. 1057). A full discussion of this issue is beyond the scope of the present paper, but the fact that the boundaries follow national boundaries only approximately and in part is clear evidence for primarily natural causation.

One of the earliest divisions, occurring at the stage of 5 clusters, separates western and central Europe. In floristic classifications, this limit was originally defined along the eastern border of the distribution area of *Ilex aquifolium* (e.g. Takhtajan, 1986). This division has been a familiar feature of many phytogeographical classifications since the term ‘Atlantic’ was first used by Watson (1835) to describe the distribution of plants in the western area (Dupont, 1962). Many phytogeographical treatments (e.g. Braun-Blanquet, 1923; Meusel *et al.*, 1965; Takhtajan, 1986; Preston & Hill, 1997) identify an area that is very similar to that mapped between 5 and 7 clusters in the British Isles, France and the Low Countries (Fig. 2); however, unlike the current analysis, the traditional concept of the Atlantic zone extends south into the Iberian peninsula and north along the Norwegian coast.  No limit corresponding to the border between the Atlantic and Subatlantic provinces of Meusel *et al*. (1965) appears in the clustering maps (Fig. 2), whereas the highly oceanic part of the Atlantic province delineated at the level of 8 clusters corresponds to the *Cochlearia danica* element that was recognized in the analysis of plant data carried out by Finnie *et al.* (2007).

It is interesting to note that the highest mountains of central Europe (the Pyrenees, Alps and Western Carpathians) are uniformly recognized as a single cluster of mountainous areas, visible from clustering with 7 clusters onwards (Fig. 2). Despite the existence of numerous local and regional endemic species (Pawłowski, 1970; Gómez *et al.*, 2003) that survived the glacial period *in situ* (e.g. Stehlik, 2003; Tribsch, 2004), the flora of these mountains includes many common species that are typical of alpine habitats due to post-glacial recolonization by widespread arctic-montane species from the periglacial zone (e.g. Stehlik, 2003). The presence of plants that are adapted to montane conditions in greater numbers in these mountain regions and the absence of frost-intolerant plants cause these areas to cluster together.

Although the Mediterranean cluster is separated at the earliest stage of splitting as a single area, which approximately corresponds to the European part of the Ancient Mediterranean floristic area in Takhtajan (1986), it is later subdivided. The five clusters that eventually emerge mirror (from west to east) the *Silene scabrifolia, Sarcocapnos enneaphylla, Ostrya carpinifolia*, *Dianthus moesiacus* and combined *Ranunculus psilostachys* and *Brassica cretica* elements that were identified by Finnie *et al.* (2007), and they emphasize the floristic heterogeneity of the Mediterranean area. The territories in south-east Europe, southward from the West Carpathians to the Mediterranean border, tend to group together at the first stages of clustering, further splitting into poorly delimited regions that approximately correspond to the Balkan and Illyric provinces of Meusel *et al.* (1965).

The pattern of plant clustering (Fig. 2) revealed in the present analysis reflects the major factors in the principal components analysis (Fig. 1, Tables 2, 3). The first factor affects the separation of the largest territories in northern Europe, the Mediterranean region and the British Isles. This factor is strongly connected with the species number but might also have a connection to floristic similarity; the large land mass of central Europe presents similar distribution patterns for many species, and the additional areas to the north and south share a similar number of species but differ in species composition. In addition, the low total numbers of species for some Mediterranean territories probably reflect under-recording (lower representation) rather than species poverty (see Finnie *et al.*, 2007). The partial correlation with the temperature pattern and a very good match with a temperate floristic sub-element of Dahl (1998) indicate that an influence of very high and low temperature values is affecting the pattern, as is landscape heterogeneity (see above). The second factor is driven by temperature and affects the segregation of the highly oceanic territories in the west and zonal division in the east. The third factor, which is interpreted as the annual temperature range and topography, serves as the basis for the regional clusters, which approximately correspond to the floristic regions and their agglomerations.

### Relationships between plants, mammals, climate and environment

In terms of the correlation between distance matrices, there is a strong relationship between mammals and plants (Pearson's $r = 0.6$; Table 1). The relationships between mammals and climate (Pearson's $r = 0.15$–0.36; Table 1) and between plants and climate (Pearson's $r = 0.13$–0.29; Table 1) are more complex but always statistically significant. The connection is further highlighted by the remarkably high correlations between some of the principal components scores (Table 4, Fig. 1). In particular, the scores for the PC1 and PC2 of the plant and mammal datasets are highly correlated, and good correlations are also obtained when comparing the scores for the PC1 and PC3 of the climate dataset. These results, together with the Kappa comparison, show that the mammal and plant patterns are similar. The plant and mammal clusters also present higher Kappa values than those of plant or mammal data in relation to either the environmental (EnZ) or climate datasets. This answers to our first question, how strongly the mammalian associations are related in space to plant associations.

A comparison of the principal components scores and some known variables (Table 3) suggests that the climatic variables are the main factors forcing the plant and mammal cluster distributions. The mean annual temperature and PC2 of both mammals and plants are correlated. It is also known (see e.g. Hawkins *et al.*, 2003; Field *et al.*, 2009) that available energy is one of the main components controlling species diversity, which is correlated with PC1 in both the plant and mammal data. Furthermore, the third

PCs of both the mammal and plant data are correlated with the annual temperature range, and the PC3 of plants is correlated with elevation. These results suggest that both mammal and plant associations have similar relationship to climate. We therefore reject our hypothesis that the plant associations would be more closely related to climate than the mammalian associations.

For the climate dataset, a possible reason that the environmental data (EnZ) show a slightly closer similarity to the plant clusterings than to the clustering of the climate data could be the inclusion in EnZ of geomorphology components (elevation and slope); it is known that elevation is linked to species diversity in plants (e.g. Bruun *et al.*, 2006; Kreft *et al.*, 2006; Kreft & Jetz, 2007). However, this relationship is not much stronger, and we must reject our hypothesis that the plant associations are more closely related to environment than the mammalian associations.

## CONCLUSIONS

There is a statistically significant connection between the climate, plant and mammal datasets. The correlation between climate variables and the plant dataset indicates that vegetation is controlled by temperature, precipitation, and elevation. Furthermore, the vegetation and mammal distribution patterns appear to be closely related. Our results suggest that the first PCs of both mammals and plants are related to species diversity, whereas the second PCs reflect latitudinal gradient patterns, that is, temperature differences and available energy. PC3 in each case appears to reflect seasonality in the form of the annual temperature range. The pattern of PC3 for plants is also reasonably similar to the topography, with an observed correlation of 0.5 with elevation. Each of the first three principal components are highly correlated between mammal and plant data.

Our results show that both mammal and plant data form coherent spatial clusters at a continental level and that they are statistically correlated. The clustering patterns of mammals and plants form groups that agree with one another in their spatial extent. The forcing of floristic patterns into coherent entities appears mainly to be caused by climatic variables (temperature, temperature range and rainfall), mediated by elevation differences. The formation of individual plant clusters is also related to species numbers and local and regional floristic differences. The close correlation between the floral and faunal patterns suggests that the mammal and plant distributions are controlled by the same environmental variables. It is, however, difficult to assess from studies at the European scale whether the distribution of mammals is controlled directly by climate, or via interactions with the plants, or by a combination of the two processes. Indeed, there appears to be no a priori reason why a mammal species should be less influenced by general vegetation character than a plant species. These questions might be tackled by an analysis of plant and mammal distributions at a finer scale, coupled with physiological and ecological studies of the plant and mammal species in particular assemblages.

## ACKNOWLEDGEMENTS

We thank the Societas Europaea Mammalogica and Tony Mitchell-Jones for providing the distributional data that were used to prepare the *Atlas of European mammals*. We are grateful to Marc Metzger for resampling the EnZ data for the mammal grid. We also thank Kai Puolamäki, Pauli Miettinen and Kari Lintulaakso for discussions on the implementation of the data analysis and Marco Girardello for suggesting the use of the Mantel test. This study was supported by the Academy of Finland and the Finnish Cultural Foundation.

## REFERENCES

Braun-Blanquet, J. (1923) *L'origine et le développement des flores dans le Massif central de France, avec aperçu sur les migrations des flores dans l'Europe sud-occidentale*. Lhomme, Paris & Beer et Cie, Zurich.

Bruun, H.H., Moen, J., Virtanen, R., Grytnes, J.-A., Oksanen, L. & Angerbjörn, A. (2006) Effects of altitude and topography on species richness of vascular plants, bryophytes and lichens in alpine communities. *Journal of Vegetation Science,* **17**, 37-46.

Dahl, E. (1998) *The phytogeography of northern Europe: British Isles, Fennoscandia and adjacent areas*. Cambridge University Press, Cambridge.

Duda, R.O., Hart, P.E. & Stork, D.G. (2000) *Pattern classification*, 2nd edn. John Wiley & Sons, New York.

Dupont, P. (1962) *La flore atlantique européenne*. Faculté des Sciences, Toulouse.

Field, R., Hawkins, B.A., Cornell, H.V., Currie, D.J., Diniz-Filho, J.A.G., Guégan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M. & Turner, J.G.R. (2009) Spatial species richness gradients across scales: a meta-analysis. *Journal of Biogeography*, **36**, 132-147.

Finnie, T.J.R., Preston, C.D., Hill, M.O., Uotila, P. & Crawley, M.J. (2007) Floristic elements in European vascular plants: an analysis based on *Atlas Florae Europaeae*. *Journal of Biogeography,* **34**, 1848-1872.

Fransson, S. (1965) The plant cover of Sweden. The borderland. *Acta Phytogeographica Suecica,* **50**, 167–175.

Gagné, S. & Proulx, R. (2009) Accurate delineation of biogeographical regions depends on the use of an appropriate distance measure. *Journal of Biogeography,* **36**, 561–562.

Gómez, D., Sesé, J.A. & Villar, L. (2003) The vegetation of the alpine zone in the Pyrenees. *Alpine biodiversity in Europe* (ed. by L. Nagy, G. Grabherr, C. Körner and D.B.A. Thompson), pp. 85–92. Springer-Verlag, Berlin & Heidelberg.

Hanski, I. (1998) Metapopulation dynamics. *Nature* **396**, 41-49.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guégan, J.-F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., O'Brien, E.M., Porter, E.E. & Turner, J.G.R. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology,* **84**, 3105–3117.

Heikinheimo, H., Fortelius, M., Eronen, J.T. & Mannila, H. (2007) European land mammal species cluster into environmentally distinct and spatially coherent entities. *Journal of Biogeography,* **34**, 1053-1064.

Heikinheimo, H., Fortelius, M., Eronen, J.T. & Mannila, H. (2009) Response to Gagné & Proulx: clusterings should not be compared by visual inspection. *Journal of Biogeography,* **36**, 563–565.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Holyoak, M., Leibold, M.A. & Holt, R.D. (2005) *Metacommunitites: spatial dynamics of ecological communities*. University of Chicago Press, Chicago.

Jalas, J. & Suominen, J. (eds) (1972–1994) *Atlas Florae Europaeae*, Vols 1 (1972), 2 (1973), 3 (1976), 4 (1979), 5 (1980), 6 (1983), 7 (1986), 8 (1989), 9 (1991), 10 (1994). Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki.

Jalas, J., Suominen, J. & Lampinen, R. (eds) (1996) *Atlas Florae Europaeae*, Vol. 11. Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki.

Jalas, J., Suominen, J., Lampinen, R. & Kurtto, A. (eds) (1999) *Atlas Florae Europaeae*, Vol. 12. Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki.

Kalela, A. (1958) Über die Waldvegetationszonen Finlands. *Botaniska Notiser,* **111**, 353–368.

Kleinberg, L. & Tardos, E. (2005) *Algorithm design*. Addison-Wesley, Boston, MA.

Kreft, H. & Jetz, W. (2007) Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences USA,* **104**, 5925-5930.

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

Kreft, H., Sommer, J. H. & Barthlott, W. (2006) The significance of geographic range size for spatial diversity patterns in Neotropical palms. *Ecography*, **29**, 21–30.

Kruskal, J. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27.

Kurtto, A., Lampinen, R. & Junikka, L. (2004) *Atlas Florae Europaeae*, Vol. 13. Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki.

Lahti, T. & Lampinen, R. (1999) From dot maps to bitmaps: *Atlas Florae Europaeae* goes digital. *Acta Botanica Fennica,* **162**, 5–9.

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

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. & Gonzalez, A. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters,* **7**, 601-613.

Lippmaa, T. (1940) Areal und Alterbestimmung einer Union (Galeobdolon-Asperula-Asarum-U) sowie das Problem der Characterarten und der Konstanten. *Acta Instituti et Horti Botanici Universitatis Tartuensis,* **6**, 1–152.

Malcolm, J.R., Markham, A., Neilson, R.P. & Garaci, M. (2002) Estimated migration rates under scenarios of global climate change. *Journal of Biogeography,* **29**, 835-849.

Malcolm, J.R., Liu, C., Neilson, R.P., Hansen, L. & Hannah, L. (2006) Global warming and extinctions of endemic species from biodiversity hotspots. *Conservation Biology,* **20**, 538-548.

MATLAB version 7.8 (2009). Natick, Massachusetts: The MathWorks Inc (build R2009a).

Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Mücher, C.A. & Watkins, J.W. (2005) A climatic stratification of the environment of Europe. *Global Ecology and Biogeography,* **14**, 549-563.

Meusel, H., Jäger, E. & Weinert, E. (1965) *Vergleichende Chorologie der Zentraleuropäischen Flora. Karten*. Gustav Fischer Verlag, Jena.

Midgley, G.F., Hannah, L., Millar, D., Rutherford, M.C. & Powrie, L.W. (2002) Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecology and Biogeography,* **11**, 445-451.

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

Monserud, R.A. & Leemans, R. (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling,* **62**, 275-293.

Olff, H., Ritchie, M. & Prins, H.H.T. (2002) Global environmental controls of diversity in large herbivores. *Nature*, **415**, 901–904.

Pawłowski, B. (1970) Remarques sur l'endémisme dans la flore des Alpes et des Carpates. *Vegetatio,* **21**, 181–243.

Polis, G.A. (1999) Why are parts of the world green? Multiple factors control productivity and the distribution of biomass. *Oikos*, **86**, 3-15.

Preston, C.D. & Hill, M.O. (1997) The geographical relationships of British and Irish vascular plants. *Botanical Journal of the Linnean Society,* **124**, 1-120.

Rao, C. (1995) A review of canonical coordinates and an alternative to corresponence analysis using Hellinger distance. *Questiio (Quaderns d'Estadistica i Investigacio Operativa)*, **19**, 23–63.

Rosenzweig, M.L. (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge.

Sharma, S.C. (1996) *Applied multivariate techniques*. John Wiley & Sons, New York & Chichester.

Shepard, R. (1962a) The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, **27**, 125-140.

Shepard, R. (1962b) The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, **27**, 219-246.

Sinnott, R.W. (1984) Virtues of the Haversine. *Sky and Telescope,* **68**, 159.

Stehlik, I. (2003) Resistance or emigration? Response of alpine plants to the ice ages. *Taxon,* **52**, 499–510.

Takhtajan, A. (1986) *Floristic regions of the world*. University of California Press, Berkeley, CA.

Theodoridis, S. & Koutroumbas, K. (2003) *Pattern recognition*, 2nd edn. Elsevier Academic Press, New York.

Thomas, C.D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, T. A., Phillips, O. L. & Williams, S. E. (2004) Extinction risk from climate change. *Nature,* **427**, 145-148.

Tribsch, A. (2004) Areas of endemism of vascular plants in the Eastern Alps in relation to Pleistocene glaciation. *Journal of Biogeography,* **31**, 747–760.

Uotila, P., Lahti, T. & Suominen, J. (2005) Erfahrung über die langjährige Kartierung von Pflanzen: Atlas Florae Europaeae. *Hoppea,* **66**, 119–132.

Watson, H.C. (1835) *Remarks on the geographical distribution of British plants*. Longman, Rees, Orme, Brown, Green, & Longman, London.

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.

## SUPPORTING INFORMATION

**Appendix S1** Correlation between the first three principal component scores of the mammal, plant and climate datasets using the Hellinger distance.

**Appendix S2** Correlation between the principal component scores using the Hellinger distance with species numbers, temperature, precipitation and elevation.

**Appendix S3** Spatial agreement between the clusterings of the plant and mammal datasets using the Hellinger distance, in addition to the climate and environmental datasets.

## BIOSKETCHES

**Hannes Heikinheimo** is part of a multidisciplinary research group that uses data mining techniques to study biological and palaeontological data. Research topics of the group range from pattern discovery, sequence segmentation and spatio-temporal data analysis to biogeography, climate–environment interactions, and community structure at evolutionary time-scales.

Author contributions: H.H., J.T.E., M.F and H.M. conceived the ideas; J.T.E., A.S., C.D.P., P.U. and M.F. collected the data; H.H., J.T.E., E.O., A.S., H.M. and M.F. analysed the data; and H.H., J.T.E and M.F. led the writing, with input from A.S., C.D.P., E.O. and P.U.

Editor: Peter Linder

## TABLES

**Table 1** Strength of the Pearson correlation (*r*) between grid cell pair similarity for the European mammal, plant and climate datasets using both the Euclidean and Hellinger distances. For rows marked with “cell proximity controlled for” (or “cell proximity and climate controlled for”), the values have been obtained after controlling for the effect of geographical proximity, that is, spatial autocorrelation (or both autocorrelation and climate, respectively). The number of data points (grid cell pairs) is 2,372,931. All the correlation coefficients are significant according to the partial Mantel test procedure using 100 permutation rounds.

Correlation (Pearson’s *r*)

| | Climate | Plant (Euclidean) | Mammal (Euclidean) | Plant (Hellinger) | Mammal (Hellinger) |
|---|---|---|---|---|---|
| Cell proximity | 0.68 | 0.44 | 0.6 | 0.71 | 0.7 |
| Climate | | 0.4 | 0.49 | 0.67 | 0.63 |
| Plant (Euclidean) | | | 0.7 | | |
| Plant (Hellinger) | | | | | 0.8 |
| Climate (cell proximity controlled for) | | 0.15 | 0.13 | 0.36 | 0.29 |
| Plant (Euclidean) (cell proximity controlled for) | | | 0.6 | | |

| | | | | | |
|---|---|---|---|---|---|
| Plant (Hellinger) (cell proximity controlled for) | | | | | 0.61 |
| Plant (Euclidean) (cell proximity and climate controlled for) | | | 0.59 | | |
| Plant (Hellinger) (cell proximity and climate controlled for) | | | | | 0.56 |

**Table 2** Percentage of variation explained using the three first principal components (PC1–PC3) for the European plant, mammal and climate datasets. The dimensionalities (dims.) of the datasets (the numbers of variables) are denoted in the column headers of the table.

Percentage of variation

| | Plant (dims. 3626) | Mammal (dims. 124) | Climate (dims. 67) |
|---|---|---|---|
| PC1 | 14.5 % | 17.3 % | 60.2 % |
| PC2 | 11.5 % | 14.5 % | 23.3 % |
| PC3 | 6.2 % | 8.0 % | 5.8 % |
| Total | 32.2 % | 39.8 % | 89.3 % |

**Table 3** Strength of the Pearson correlation (absolute value) between the principal component (PC) scores of European mammal, plant and climate data with species numbers, temperature, precipitation and elevation. The highest values are shown in bold.

**Correlation (Pearson's *r*)**

| | Plant | | | Mammal | | | Climate | | |
|---|---|---|---|---|---|---|---|---|---|
| | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 |
| Plant species count | **0.81** | 0.18 | 0.28 | 0.63 | 0.17 | 0.09 | 0.17 | 0.16 | 0.5 |
| Mammal species count | 0.65 | 0.2 | 0.19 | **0.83** | 0.3 | 0.01 | 0.13 | 0.1 | 0.55 |
| Mean annual temperature | 0.24 | **0.79** | 0.35 | 0.26 | **0.82** | 0.03 | **0.99** | 0 | 0.05 |
| Temperature annual range | 0.14 | 0.2 | **0.54** | 0.08 | 0.35 | **0.45** | 0.41 | **0.7** | 0.27 |
| Mean annual precipitation | 0.1 | 0 | 0.03 | 0.03 | 0.07 | 0.16 | 0.01 | **0.97** | 0.15 |
| Mean Elevation | 0.09 | 0.2 | **0.57** | 0.02 | 0.19 | 0.04 | 0.14 | 0.27 | 0.19 |

**Table 4** Strength of the Pearson correlation (absolute value) between the first three principal component (PC) scores of the European mammal, plant and climate datasets. The highest values between two datasets are shown in bold.

**Correlation (Pearson's *r*)**

| | | Mammal | | | Plant | | |
|---|---|---|---|---|---|---|---|
| | | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 |
| Plant | PC1 | **0.88** | 0.04 | 0.05 | | | |
| | PC2 | 0.01 | **0.90** | 0.13 | | | |
| | PC3 | 0.05 | 0.14 | **0.58** | | | |
| Climate | PC1 | 0.22 | **0.83** | 0.03 | 0.20 | **0.80** | 0.35 |
| | PC2 | 0.02 | 0.10 | **0.30** | 0.06 | 0.00 | **0.1** |
| | PC3 | **0.64** | 0.03 | 0.25 | **0.68** | 0.03 | 0.33 |

**Table 5** Strength of the spatial agreement between the clusterings using the European plant, mammal, climate and environmental data that were measured using the Kappa statistic.

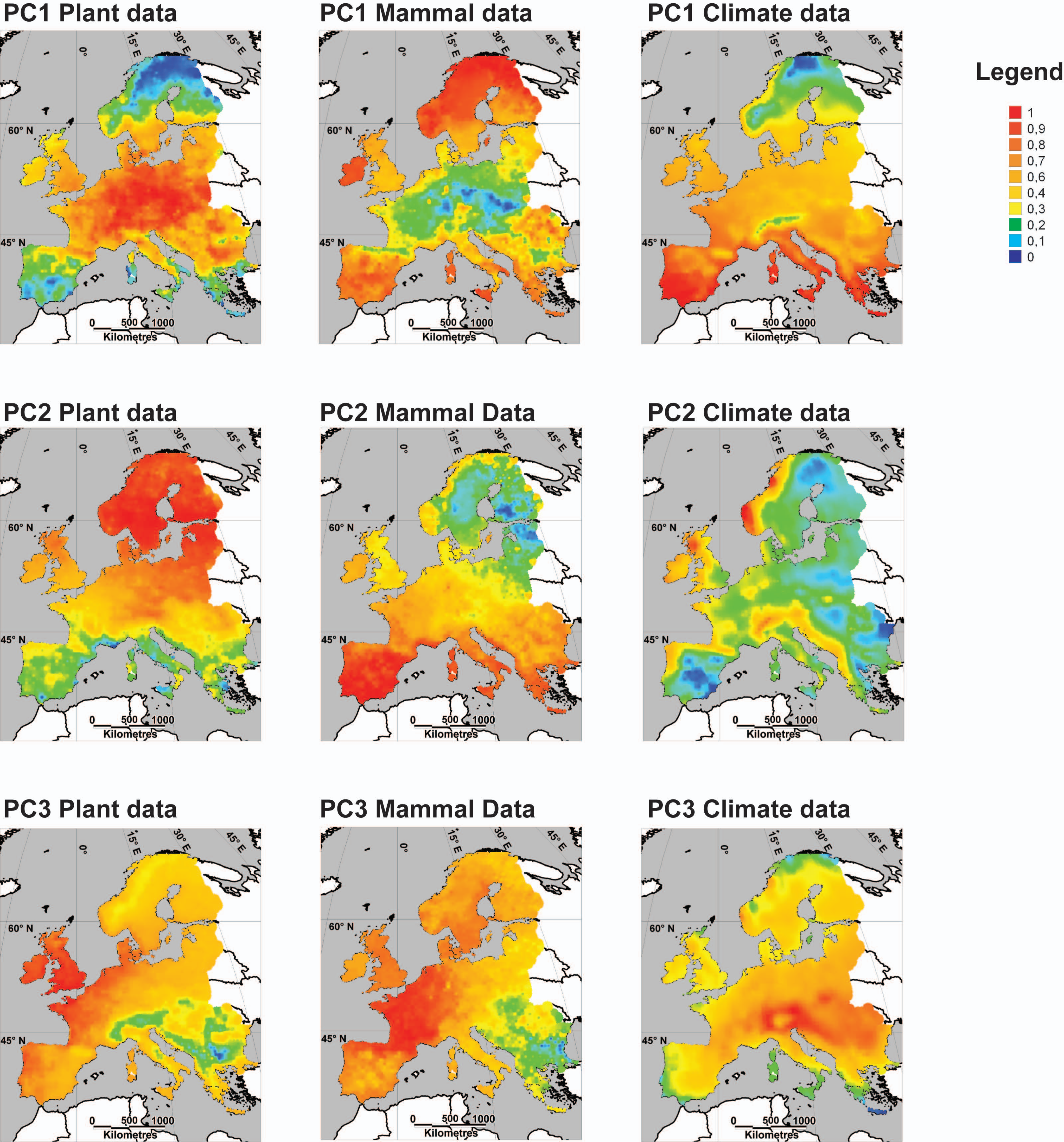
Kappa

| | | Mammal | | | Plant | Environment | Climate |
|---|---|---|---|---|---|---|---|
| | | All | Herbivores | Non-herbivores | | | |
| Mammal | All | 1 | 0.6 | 0.77 | 0.46 | 0.4 | 0.41 |
| | Herbivores | | 1 | 0.5 | 0.51 | 0.4 | 0.37 |
| | Non-herbivores | | | 1 | 0.42 | 0.35 | 0.41 |
| Plant | | | | | 1 | 0.51 | 0.45 |
| Environment | | | | | | 1 | 0.49 |

## FIGURES

**Fig. 1** Spatial distribution of the principal component (PC) scores for the European plant, mammal and climate datasets. The data were smoothed for the illustration with a 70-km radius interpolation (the mean grid size was approximately 50 km × 50 km).

**Fig. 2** The sequence of clusterings of the European plant data cells with the number of clusters ranging from 2 to 13. The colours are used only to distinguish the clusters within each image and do not imply a one-to-one matching of clusters between images.



PC1 Plant data

PC1 Mammal data

PC1 Climate data

PC2 Plant data

PC2 Mammal Data

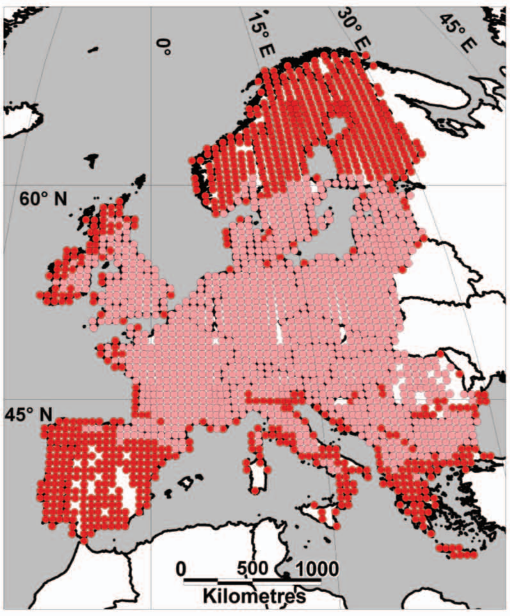
PC2 Climate data

PC3 Plant data

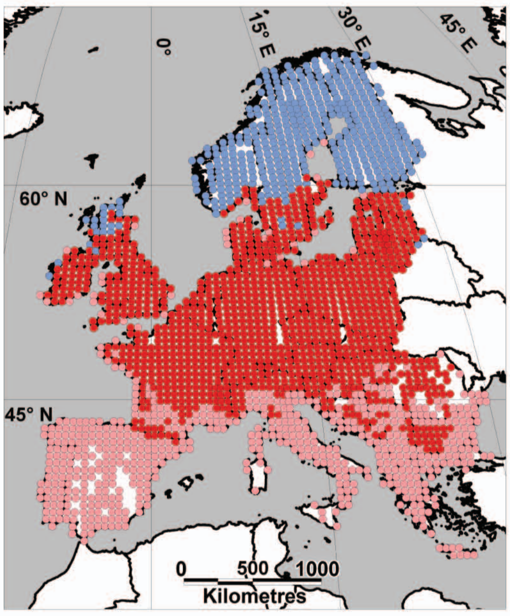
PC3 Mammal Data

PC3 Climate data

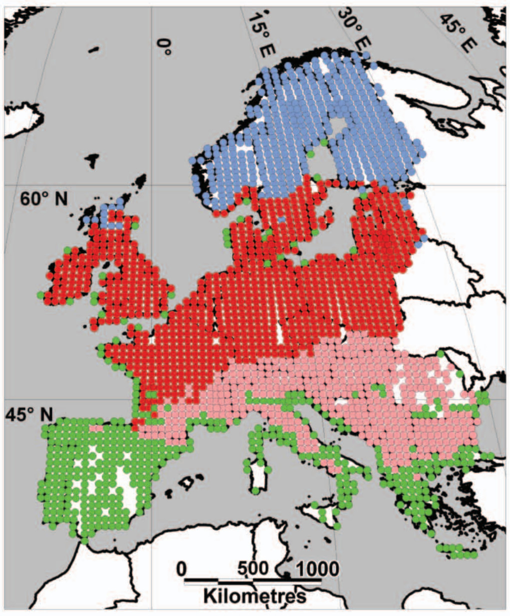
**2 clusters**



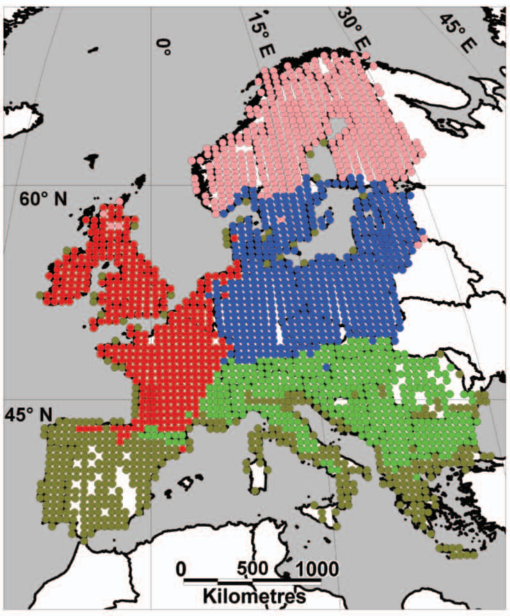
**3 clusters**



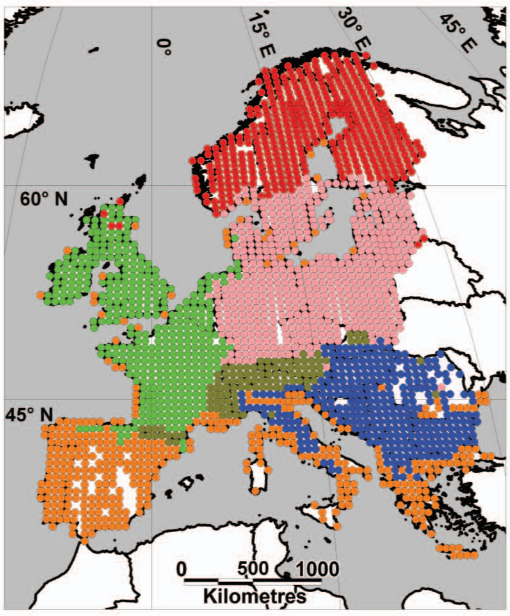
**4 clusters**



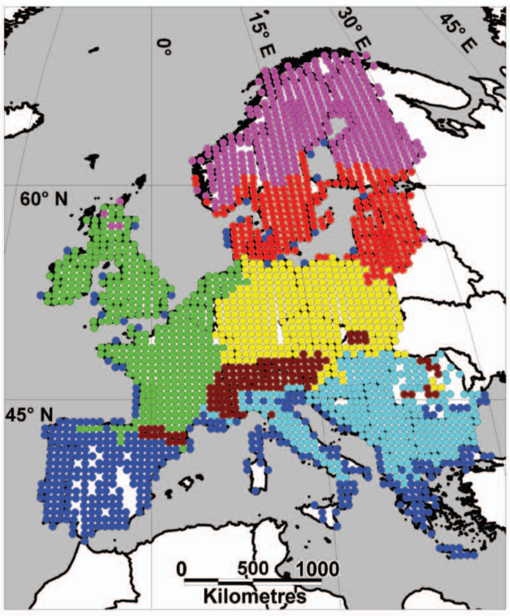
**5 clusters**



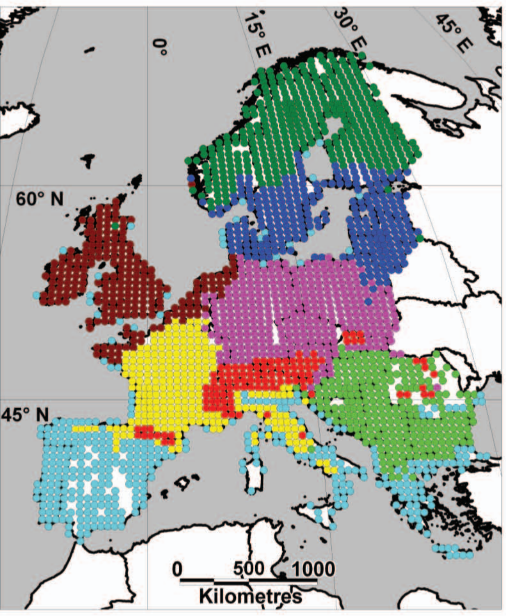
**6 clusters**



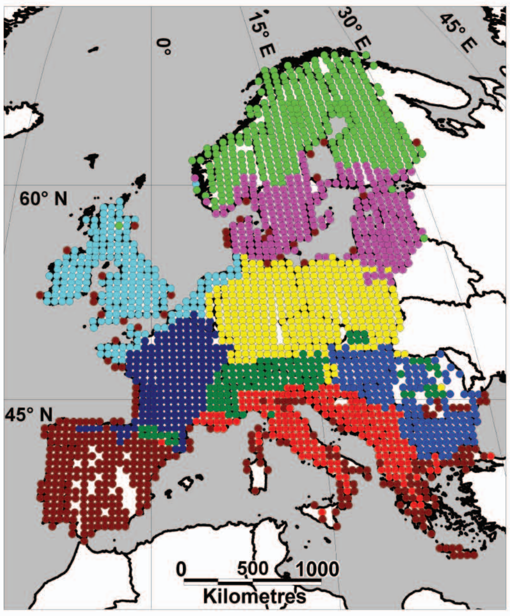
**7 clusters**



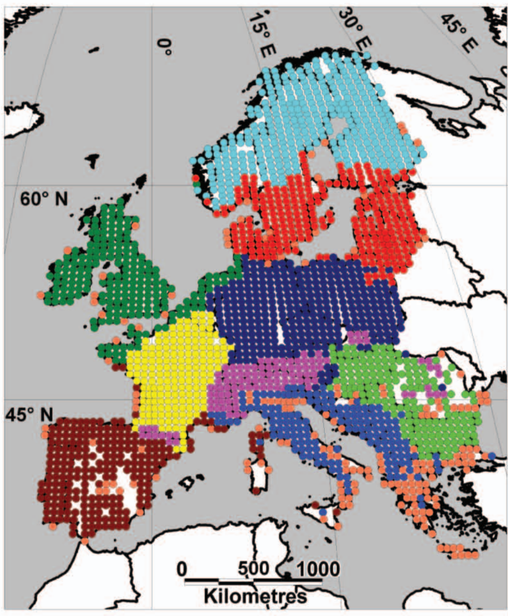
**8 clusters**



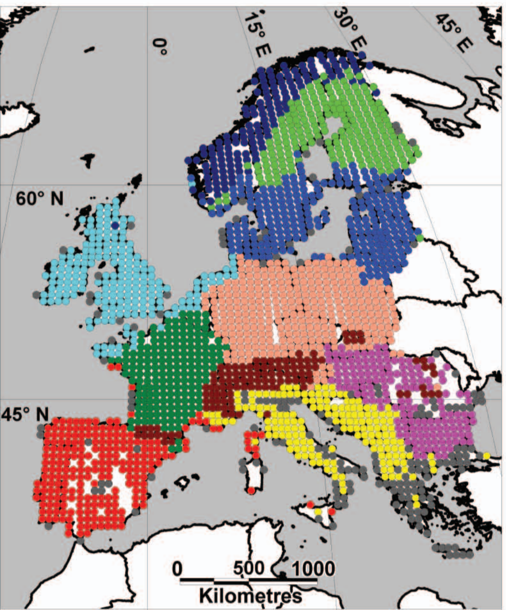
**9 clusters**



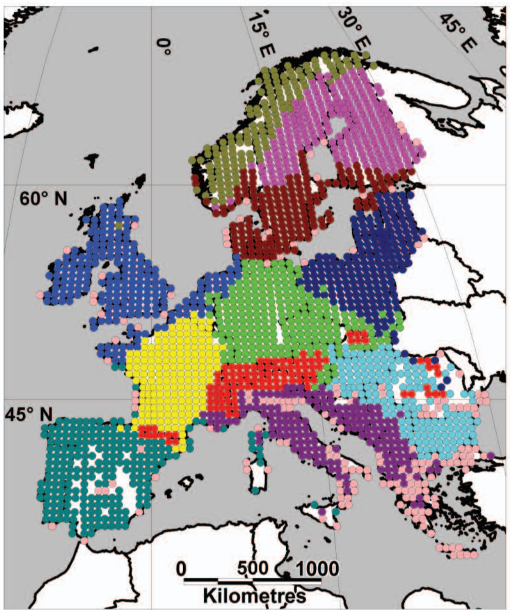
**10 clusters**



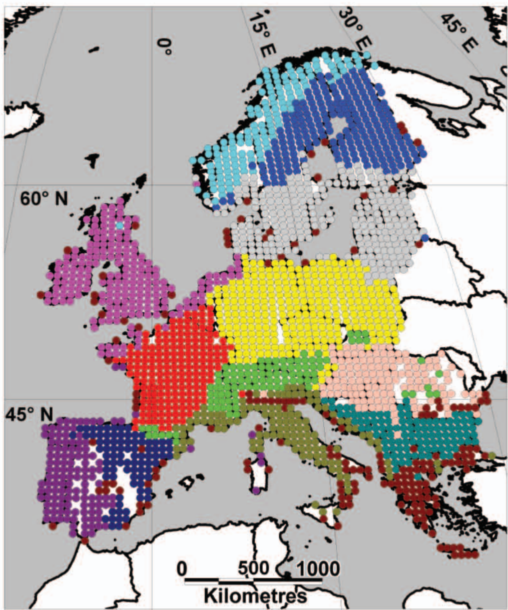
**11 clusters**



**12 clusters**



**13 clusters**



**Appendix S1.** Strength of the Pearson correlation (absolute value) between the first three principal component scores of the mammal, plant and climate datasets using the Hellinger distance.
Compare this information with Table 3 in the main text for correlations using Euclidean distance.

| | | Mammals | | | Plants | | |
|---|---|---|---|---|---|---|---|
| | | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 |
| Plants | PC1 | 0.95 | 0.06 | 0 | | | |
| | PC2 | 0.04 | 0.87 | 0.15 | | | |
| | PC3 | 0.02 | 0.2 | 0.77 | | | |
| Climate | PC1 | 0.88 | 0.13 | 0.09 | 0.89 | 0.14 | 0.22 |
| | PC2 | 0.01 | 0.03 | 0.34 | 0.03 | 0 | 0.35 |
| | PC3 | 0.26 | 0.57 | 0.25 | 0.26 | 0.57 | 0.25 |

**Appendix S2.** Strength of the Pearson correlation (absolute value) between the principal component scores using the Hellinger distance with species numbers, temperature, precipitation and elevation. Compare this information with Table 4 in the main text for correlations using Euclidean distance.

| | Plants | | | Mammals | | |
|---|---|---|---|---|---|---|
| | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 |
| Plant species co | 0.44 | 0.39 | 0.32 | 0.41 | 0.41 | 0.08 |
| Mammals speci | 0.08 | 0.55 | 0.34 | 0.11 | 0.68 | 0.07 |
| Climate mean a | 0.9 | 0.09 | 0.21 | 0.88 | 0.09 | 0.1 |
| Climate mean a | 0.03 | 0.01 | 0.2 | 0.01 | 0.01 | 0.2 |
| Climate annual | 0.34 | 0.01 | 0.66 | 0.28 | 0 | 0.59 |
| Elevation | 0.11 | 0.27 | 0.36 | 0.15 | 0.18 | 0.14 |

**Appendix S3.** Strength of the spatial agreement measured using the Kappa statistic between the clusterings using the plant and mammal datasets with the Hellinger distance, in addition to the climate (Hijmans et al., 2005) and environmental (Metzger et al., 2005) datasets.
Compare this information with Table 5 in the main text for clusterings using Euclidean distance.

| Kappa: | Mammals (all) | Mammals (Herbivores) | Mammals (Non-Herbivores) | Plants | Environment | Climate |
|---|---|---|---|---|---|---|
| Mammals (all) | 1 | 0.55 | 0.8 | 0.5 | 0.37 | 0.41 |
| Mammals (Herbivores) | | 1 | 0.49 | 0.55 | 0.45 | 0.43 |
| Mammals (Non-Herbivores) | | | 1 | 0.48 | 0.38 | 0.41 |
| Plants | | | | 1 | 0.54 | 0.49 |
| Environment | | | | | 1 | 0.49 |