Methods for comparison of biotic regionalizations: the case of pteridophytes in the Iberian Peninsula

Ana L. Márquez, Raimundo Real and J. Mario Vargas

We made several regionalizations of the Iberian Peninsula based on the distributions of the pteridophyte flora to see whether the regionalization depended on the type and scale of lattice or the similarity index considered. We used five types of lattice in which the scale also varied: river basins, mountains and plains, natural regions, physiographic and geological regions, and administrative provinces; and two similarity indices: those of Jaccard and of Baroni-Urbani and Buser. The regionalizations varied according to the type of lattice, the grain size, and the similarity index used. To assess the different regionalizations we used four methods: 1) the coefficient of variation of the size of sites in each lattice, 2) the bestblock method, which considers as the best lattice that which maximizes the number of matches between presences over all pairwise site comparisons, 3) the Mantel test, to assess the statistical significance of the regionalizations obtained, and 4) mapability, which considers the most contiguous regionalization to be the best. The best regionalization according to our four criteria was that based on administrative provinces and Jaccard’s index. This yielded a small central region and three large regions: northern, western, and eastern.

A major objective of descriptive biogeography is to simplify the complex patterns of contemporary species’ distributions by classifying areas based on their biotic composition. That areas of the world can be characterized on the basis of the presence of certain taxa and the absence of others has long been recognised by biogeographers (Buffon 1761), and is a deeply entrenched component of much contemporary biogeographical thinking. In addition, classifying geographical areas into groups with different species composition is valuable for nature conservation planning (Brown et al. 1993), so as to ensure that all groups are represented in the selection of natural reserves (Margules 1986), and to evaluate the biological resources of an area in a regional and global context (Carey et al. 1995). However, currently administrators are often confused by the number of different classifications of their country proposed by ecologists and phytogeographers. In addition, as Thaler and Plowright (1973) pointed out, phytogeographers have disagreed about whether floristic areas represent real entities in nature or whether they are simply convenient, subjective constructs that lack any objective and unambiguous basis.

Operationally, it is possible to identify biotic boundaries when a group of areas with similar biota shares fewer species than expected at random with other group of areas with similar biota. Then the Operational Geographic Units (OGUs; Crovello 1981) of this territory can be grouped into biotic regions (Real et al. 1992a, Myklestad and Birks 1993, McLaughlin 1994). Clustering methods represent one approach to this problem (Birks 1987, Legendre 1990), but only when a probabilistic procedure allows the researcher to distinguish between statistically significant clusters and those indistinguishable from random expectation (Márquez et al. 1997). Area clusters may be
considered as biotic regions when a group of OGUs is statistically significant. Legendre (1987) advocated constraining the clustering to respect spatial contiguousites. However, in some cases, biotic regions may reflect a climatic or ecological pattern that may not be strictly contiguous in space, and then they may be better detected with clusters only defined by the similarity of their species composition. Then, area clusters can be related to environmental variables by techniques such as regression or canonical variates analysis (Carey et al. 1995, Olivero et al. 1998) to investigate possible causes of the regionalization pattern.

However, to analyse biogeographic patterns, any territory must be subdivided initially into different types of OGUs (Crovello 1981, Kenkel et al. 1989, Mandrak 1995). This establishes a lattice that influences the biogeographic regions subsequently obtained, because any biotic boundary necessarily has to follow along the edges of the sites used to partition the study area. Cressie (1993) defined a lattice as a countable collection of spatial sites that are spatially regular or irregular, represent points or regions, and are indices for continuous or discrete variables. In biogeographic regionalization the concept of lattice corresponds to any kind of area subdivision, irrespective of its size and the criteria used to define it, and the sites correspond to the OGUs. The lattices can be natural (e.g., watersheds, Vargas et al. 1998), arbitrary (e.g., latitude-longitude cells, Schall and Pianka 1977, or equal-area quadrats, Márquez et al. 1997). The sites can have the same size when the lattice is regular, such as those based on grid-squares, or different sizes when the lattice is irregular, such as those made according to, for example, physical, geological, or political viewpoints. Variation in size and shape may affect the species richness of each site and hence the results of the subsequent regionalization. Square equal-area quadrats are considered not to be subject to these problems as quadrat size and shape are uniform (Mandrak 1995), although coastal quadrats may vary in area when terrestrial organisms are studied, and mountainous areas have more surface area than flat areas in a “equal-area quadrat”.

There are also problems in defining the scale of the grain used (Jardine 1972, Birks 1987, Carey et al. 1995). Scale is dependent on both the extent of the study area and the size of organism studied (Wiens 1989). If too coarse a lattice is used, biotic regions will be correspondingly coarse. Conversely, the use of too fine a lattice may fail to reveal major biotic regions and generate many spurious small regions where recordings may have been uneven. Results obtained using small sites may be different from those obtained using larger sites, although these different patterns might be interrelated (Blondel 1985, Vargas 1993). However, this does not mean that there are no criteria for comparing the results obtained using different types of sites or different scales (Levin 1992). These criteria, however, need proper definition and testing.

Ferns could provide a valuable source of biogeographical information that is distinct from, or at least complementary to, that provided by other types of plant (Márquez et al. 1997). Pteridophytes are a group of species with a great dispersal capacity due to the lightness of their spores (Jermy 1984). However, most fern species are restricted in distribution and only a limited number have wide ranges (Kato and Iwatsuki 1983), because their distribution is highly related to climate and substratum (Barrington 1993, Given 1993, Kornas 1993). The 113 Iberian pteridophyte species are well recorded and well understood taxonomically in the Iberian Peninsula (Salvo et al. 1984, Castroviejo et al. 1986), and are an important number of species in a European context.

In this paper we have used several criteria to select the best biotic regionalization among several partitions obtained using a quantitative method of classification. With this aim, we have used fern distributions in the Iberian Peninsula a) to establish whether the Iberian Peninsula may be divided into pteridogeographic regions or whether the pteridophyte species distributions follow a mainly gradual pattern; b) to verify whether the resulting regionalization patterns are different according to the type of lattice, the number of sites considered, and the similarity index used; and c) to apply several criteria to select the best regionalization pattern.

Material and methods

Study area

The Iberian Peninsula is situated on the south-western end of Europe, and presents a stronger peninsular condition than other large Mediterranean peninsulas because its isthmus is relatively narrow. In addition, the presence of the Pyrenees in the isthmus hinders the spread of propagules between the Iberian Peninsula and the rest of the European continent. The interchange of biota with Africa is also hindered by the Straits of Gibraltar. Climatically, the Iberian Peninsula is placed in the southern band of the temperate zone.

Species distribution and Operational Geographic Units (OGUs)

The distribution of the 113 species of pteridophytes were obtained from Salvo et al. (1984), supplemented with new records published later.

We considered five types of OGUs in the Iberian Peninsula: river basins, mountains and plains, natural regions, physiographic and geological regions, and administrative provinces (Anon. 1980) (Fig. 1). The province lattice is political-administrative, although is
partially based on natural partitions, whereas the other four lattices have natural borders. Natural regions and physiographic and geological regions are partially based on vegetation, and so some circular reasoning may be involved in their use as phytogeographic regions (Birks 1987, Real and Ramírez 1992, Vargas 1993).

Classification analysis
Initially we constructed a presence-absence matrix for the 113 species in each lattice considered in the area. We thus had five presence-absence matrices, in which OGU size and number varied: river basins ($11 \times 113$), mountains and plains ($12 \times 113$), natural regions ($17 \times 113$), physiographical regions ($22 \times 113$), and administrative provinces ($56 \times 113$).

We applied two similarity indices to each pair of OGUs: Jaccard’s (1908) index and Baroni-Urbani and Buser’s (1976). The Jaccard index considers only double presences, whereas the Baroni-Urbani and Buser index incorporates double absences. Jaccard values may be considered as absolute measures of biotic similarity, because they are not affected by other species absent from the two OGUs compared, while the Baroni-Urbani and Buser coefficient measures similarity in the context of the Iberian Peninsula, because species absent from the two OGUs but present in other Iberian OGUs increase the similarity between the two OGUs analysed.

![Diagrams of five different lattices considered in the study](Anon. 1980).

Fig. 1. The five different lattices considered in the study (Anon. 1980).
(Real et al. 1992b). However, the possibility that two OGU is considered similar exclusively because of double absences is avoided in the Baroni-Urbani and Buser index by multiplying double absences by double presences.

We used UPGMA (Unweighted Pair-Group Method using Arithmetic Averages) as our clustering method (Sneath and Sokal 1973). Agglomerative techniques are preferred to divisive techniques because the former maximizes the similarity within a group, which defines biotic regions. We represented the results as dendrograms.

To test the groups obtained in any association analysis several authors use a null model based on a Monte Carlo randomization procedure (Connor and Simberloff 1979, Oden et al. 1993). However, this model uses the observed data matrix to generate a null distribution, and so observed and null distributions lack statistical independence (Grant and Abbot 1980). Nonetheless, it is feasible to determine all the possible distributions of N species in any of the OGUs compared for each value of N, and so an exact randomization test (Sokal and Rohlf 1981, p. 788) can be performed to determine whether an observed value of the Jaccard index or the Baroni-Urban and Buser index is significantly different from those expected at random. We obtained the statistical significance of the observed similarities values, according to the exact randomization tests in Real and Vargas (1996) for the Jaccard index, and in Baroni-Urbani and Buser (1976) for the Baroni-Urbani and Buser index. In this way, the similarity matrices were transformed into matrices of statistical significance (Marquez et al. 1997) that consisted of pluses, for similarity values significantly higher than those expected at random, minuses, for similarity values significantly lower than those expected at random, and zeros, for similarity values compatible with random expectation.

Our null hypothesis is that biotic boundaries do not exist, so that we started searching for boundaries at the level of the dendrogram with the lowest similarity value, and then moved to the higher similarity nodes, finishing when no more significant boundaries appeared. This divisive procedure does not contradict the use of an agglomerative classification, because significant boundaries divide the space into biotic regions that require not only the existence of a biotic boundary, but also a biotic affinity between OGUs in the region, which is favoured by the agglomerative method.

We tested at every node of the dendrogram for the existence of significant biotic boundaries between the OGUs separated by the node using the method of McCoy et al. (1986), as modified by Real et al. (1992b) for application to dendrograms, as indicated in Appendix 1.

To determine which lattice and similarity index best define the biotic regions we used four methods: 1) Coefficient of variation of the OGU’s size in each lattice (Sokal and Rohlf 1981). Differences in the size of the OGUs within a lattice produce area-related distortions, such as differences in the number of species in different parts of the study area. So, the lattice with the lowest coefficient of variation of the OGUs’ size is considered the best one. This method selects the best lattice regardless of the distribution data. 2) Bestblock (Phipps 1975), which considers the best lattice to be that which maximizes the number of matches between presences over all pairwise OGU comparisons. This method selects the best lattice after taking into account the distribution data regardless of the regionalization obtained. 3) We used the Mantel test of matrix comparison (Mantel 1967) to compare each original similarity matrix with each matrix of hypothetical similarities according to the resulting regionalizations. The matrices of hypothetical similarities were obtained by considering all the similarities within a biotic region to be 1, all the similarities between OGUs separated by weak boundaries to be the random expectation (0.33 for Jaccard’s index and 0.5 for Baroni-Urbani and Buser’s index), and all the similarities between OGUs separated by strong boundaries to be 0. Only significant regionalizations were considered. The most significant regionalization is that which is most consistent with the original similarities and is thus considered the best. 4) Mapability (Harper 1978). Contiguity is preferred to discontinuity of the biotic regions obtained using each similarity index in each lattice. We measured the mapability in terms of the minimum number of quadrats UTM 50 x 50 km necessary to join together the parts of all fragmented regions. The lower this number of quadrats is, the more mapable the regionalization is.

We applied the two former criteria to the five lattices and the two latter criteria to the ten regionalizations resulting from using each similarity index in each lattice.

Results
We found significant biotic boundaries for all types of OGU using both similarity indices. All the regionalization patterns obtained using the Jaccard index (Table 1) and most of the regionalizations obtained using Baroni-Urbani and Buser’s index (Table 2) were statistically significant according to the Mantel test of matrix comparison (Table 3). Only the regionalization obtained using mountains and plains and the Baroni-Urbani and Buser index was found to be non-significant (p > 0.01) using the Mantel test.
Table 1. Statistically significant boundaries between groups of OGUs formed by UPGMA using the Jaccard similarity index. DW > 0 and significant GW indicate weak boundary between the groups; DS > 0 and significant GS indicate a strong boundary between the groups. N.S.: p > 0.05; *: p < 0.05; **: p < 0.01; ***: p < 0.005.

<table>
<thead>
<tr>
<th>Groups set up by UPGMA</th>
<th>Biotic boundary</th>
<th>Weak</th>
<th>Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DW (A × A)</td>
<td>DW (B × B)</td>
<td>DW</td>
</tr>
<tr>
<td>River basins 3–5</td>
<td>0.206</td>
<td>0.168</td>
<td>0.187</td>
</tr>
<tr>
<td>Mountains and plains 2–9</td>
<td>0.562</td>
<td>0.627</td>
<td>0.595</td>
</tr>
<tr>
<td>Natural regions 2–13</td>
<td>0.219</td>
<td>0.209</td>
<td>0.214</td>
</tr>
<tr>
<td>Physiographical regions 7–2</td>
<td>0.372</td>
<td>0.571</td>
<td>0.472</td>
</tr>
<tr>
<td>Administrative provinces 16–54</td>
<td>0.189</td>
<td>0.707</td>
<td>0.448</td>
</tr>
</tbody>
</table>

Table 2. Statistical significant boundaries between groups of OGUs formed by UPGMA using the Baroni-Urbani and Buser similarity index. DW > 0 and significant GW indicate weak boundary between the groups; DS > 0 and significant GS indicate a strong boundary between the groups. N.S.: p > 0.05; *: p < 0.05; **: p < 0.01; ***: p < 0.005.

<table>
<thead>
<tr>
<th>Groups set up by UPGMA</th>
<th>Biotic boundary</th>
<th>Weak</th>
<th>Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DW (A × A)</td>
<td>DW (B × B)</td>
<td>DW</td>
</tr>
<tr>
<td>River basins 3–5</td>
<td>0.116</td>
<td>0.116</td>
<td>0.116</td>
</tr>
<tr>
<td>Mountains and plains 2–4</td>
<td>0.307</td>
<td>0.307</td>
<td>0.307</td>
</tr>
<tr>
<td>Natural regions 6–14</td>
<td>0.114</td>
<td>0.114</td>
<td>0.114</td>
</tr>
<tr>
<td>Physiographical regions 7–12</td>
<td>0.250</td>
<td>0.191</td>
<td>0.244</td>
</tr>
<tr>
<td>Administrative provinces 16–12</td>
<td>0.320</td>
<td>0.392</td>
<td>0.356</td>
</tr>
</tbody>
</table>

Table 3. Results of the application of the four criteria used to determine which is the best regionalization. CV: Coefficient of variation of the OGUs’ size; TM: Total matches between presences in all pair-wise comparisons; SQ: Number of squares 50 × 50 km necessary to join together each fragmented region.

<table>
<thead>
<tr>
<th>OGU Systems</th>
<th>Coefficient of variation</th>
<th>Best block</th>
<th>Mantel test</th>
<th>Mapability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CV</td>
<td>TM</td>
<td>z</td>
<td>p(z obs &lt; z exp)</td>
</tr>
<tr>
<td>River basins</td>
<td>0.481</td>
<td>20 983</td>
<td>10.17</td>
<td>0.002</td>
</tr>
<tr>
<td>Mountains and plains</td>
<td>0.445</td>
<td>21 754</td>
<td>13.72</td>
<td>0.001</td>
</tr>
<tr>
<td>Natural regions</td>
<td>0.360</td>
<td>26 885</td>
<td>24.92</td>
<td>0.001</td>
</tr>
<tr>
<td>Physiographical regions</td>
<td>0.447</td>
<td>28 977</td>
<td>60.75</td>
<td>0.001</td>
</tr>
<tr>
<td>Administrative provinces</td>
<td>0.359</td>
<td>47 881</td>
<td>544.9</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Area classification using the Jaccard similarity index

Figure 2 shows the biotic boundaries found using each lattice and the Jaccard similarity index. The lattices of river basins and natural regions present only a weak biotic boundary that divides the Iberian Peninsula into a northern and a southern region, whereas two or more boundaries are obtained using the other lattices (Fig. 2, Table 1).

In the administrative provinces lattice, three strong boundaries divide the Iberian Peninsula into four supraregions: northern supraregion, western supraregion, eastern supraregion, and a small central supraregion. The western and eastern supraregions are divided by weak biotic boundaries into four and three regions, respectively (Fig. 2).

Area classification using the Baroni-Urbani and Buser similarity index

Figure 3 shows the biotic boundaries found using each lattice and the Baroni-Urbani and Buser index. In four lattices (river basins, mountains and plains, natural regions, and physiographical regions) there are only weak biotic boundaries (Table 2) that divide the Iberian Peninsula into northern and southern regions, although using mountains and plains the regionalization was not statistically significant (Table 3). However, using the provinces there are three big regions (northern, western, and eastern) and three small regions (Fig. 3).

Which lattice defines best the biotic regions?

The administrative provinces have the lowest coefficient of variation for the size of the sites (Table 3). That means that using the provinces the Iberian territory is more evenly distributed than using the other lattices. According to the bestblock method, the administrative provinces lattice maximizes the number of positive matches over all pairwise species comparisons (Table 3). The significance values associated with the Mantel test are lower for river basins and nonsignificant for mountains and plains when using the Baroni-Urbani and Buser’s index, whereas the other regionalizations all have the same high probability value (Table 3). The regions obtained from physiographical regions using the Jaccard index and from administrative provinces using the Baroni-Urbani and Buser index are fragmented (Table 3, Figs 2 and 3), whereas the other lattices yield continuous regions.

Therefore, the only regionalization that obtains the best scores according to the four criteria is the one obtained using administrative provinces and the Jaccard index (Fig. 2).

Discussion

Although there is no absolute regionalization of the Iberian Peninsula according to the distribution of pteridophytes, the concept of the biotic region is not an useless artefact. Regionalizations vary according to the type of OGU, the grain size, and the similarity index used. In this way, any regionalization must indicate...
those characteristics, which are inseparable from the pattern obtained. This is especially true given the exploratory nature of current approaches to biotic regionalization (Carey et al. 1995). However, the biogeographic regions and boundaries are useful for understanding the mechanisms underlying the distribution of organisms, provided that each regionalization reflects a distribution pattern that is statistically significant and may be related to environmental conditions, although detected through the filters of the OGU’s, the scale of grain, and the similarity index.

The indices of similarity

Oden et al. (1993) modified the method of Womble (1951) to detect intervals of marked change in categorical variables between contiguous sites. When the vari-
able analysed is the presence or absence of species this method could be used to obtain biotic boundaries and regions. However, the basic statistic that results from the method of Oden et al. (1993) is the complement of the simple matching coefficient, which is not recommended for biogeographical regionalizations, since regions may be defined only on the basis of shared absences, even with no single species in common (Baroni-Urbani and Buser 1976).

We opted for using the Jaccard and Baroni-Urbani and Buser similarity indices to estimate the associations between the OGUs and, later, to test the statistical significance of such associations, because they are commonly used and are somewhat complementary. The Jaccard index only takes into account the double presences of species in the two OGUs compared, and the similarity between two OGUs is not affected by other species present in other OGUs in the study area. The Baroni-Urbani and Buser index takes into account double absences, which offers potentially useful information because if a species is absent from two OGUs, it may be due to ecological or historical factors impacting in a similar way on the species in the context of the area studied. If the pattern obtained considering the similarities between each pair of OGUs only (using the Jaccard index) is maintained when the similarities take into account all the study area (using the Baroni-Urbani and Buser index), then the lattice used may be considered suitable for revealing a consistent distribution pattern of the species.

The regionalizations obtained using the two indices are identical when based on river basins and on natural regions, partially similar using physiographical regions and provinces, and very different using mountains and plains. The differences seem to be more related to the type of OGU than to the scale of grain. It may be that when using river basins and natural regions the distribution of pteridophytes is very well defined, because the double absences do not modify the overall relationship pattern obtained using only the presences, whereas mountains and plains are very unsuitable units for recording the distribution of pteridophytes, so that double absences greatly alter the relationship pattern between the OGUs, with physiographical regions and provinces in an intermediate position.

The number of boundaries obtained using the Baroni-Urbani and Buser coefficient tends to be lower than that obtained using the Jaccard index, probably because the double absences considered in the Baroni-Urbani and Buser index tend to homogenize the regions. This may be seen especially in the south, that has a lower number of species than in the north. Perhaps this tendency should be taken into account when a researcher decides which index to use to reveal a regionalization pattern, according to the underlying knowledge and assumptions.

**The criteria to select the best regionalization**

Even when using an numerical method of classification and a probabilistic procedure to assess the groups of OGUs obtained, several regionalizations may result from the use of different types of OGUs or different similarity indices, so that several criteria are needed to compare the suitability of the regionalizations obtained.
We used two criteria prior to the classification procedure, and two criteria after the classification. The coefficient of variation (Sokal and Rohlf 1981) indicated that the size of each OGU in the provinces is more similar than in the other lattices and is thus to be preferred independently of the distribution data. The "bestblock" method (Phipps 1975) showed that administrative provinces provide the best lattice for establishing the biotic boundaries of the Iberian Peninsula according to its pteridoflora (Table 3), because the finer scale of the partition produced more units, and because the presences are more equitably distributed through the OGUs, thus producing more matches of presences in pairwise comparisons. The regionalization obtained using provinces and the Jaccard index was one of five regionalizations that were suitable according to both the Mantel test and the mapability criterion, because it had the highest statistical significance and was totally mapable. Mantel and mapability tests were applied to the final regionalization patterns, and showed whether the patterns were in accordance with the original similarities between OGUs and with the within-
Administrative provinces as operational geographic units

In the Iberian Peninsula, provinces provide the lattice that yields the most appropriate regionalization according to our four statistical criteria when using Jaccard’s index, and is thus the regionalization to be preferred in attempting to identify the biogeographical processes affecting the distribution of pteridophytes. One reason for this may be that provinces are administrative units of ancient origin and are not completely arbitrary. This is likely to be true elsewhere, at least in Europe, where administrative units are partially based on natural partitions perceived by local people, as is the case, for example, with vice-counties in the British Isles (Baroni-Urbani and Collingwood 1976).

Different types of arbitrary areas within a study region can be used, such as countries, counties, provinces, river basins, latitude/longitude blocks, or equal size U.T.M. squares (Birks 1987). In the last decade regular geographical units, like networks of regular squares, have been commonly used (e.g. Martin and Gurrea 1990, Lausini and Nimis 1991, Myklestad and Birks 1993, Carey et al. 1995, Moreno-Saiz et al. 1998). However, regular squares have artificial borders and, therefore, may include several fragments of different geographic or environmental units, so hindering the search for biogeographical processes that may only act on natural units (Palomo and Antúnez 1992). The division in provinces is also arbitrary, but the administrative borders between provinces follow, in part, natural boundaries such as geographical ranges, so that they may be more appropriate to assess the environmental explanations for species distributions. In addition, provinces may be a suitable type of OGU for analysing human influence on natural distributions, because in most countries human activities follow political-administrative borders.

Acknowledgements – We thank H. J. B. Birks, S. P. McLaughlin and P. L. Nimis for revisions of the manuscript and their valuable suggestions. We are also grateful to M. A. Rendón, J. C. Guerrero and J. Olivero for their help with the statistical procedures.

References


Appendix 1. Calculation of parameters used to test the existence of significant biotic boundaries.

For each dendrogram node we established a submatrix of significant similarities that only included the two groups of OGUs separated by that node, which we named group A and group B, respectively. This submatrix was divided into three zones: zone A × A and zone B × B, which corresponded to the significant similarities between OGUs of group A and between OGUs of group B, respectively; and zone A × B, corresponding to the significant similarities between the two groups of OGUs.

We call \( Pp(A \times A) \) the number of pluses within zone A × A divided by the total number of pairwise OGU comparisons in zone A × A. So, \( Pp(A \times A) \) is the proportion of pluses in zone A × A. We call \( Pp(A \times A) \) the number of OGUs in group A that have at least one plus divided by the total number of OGUs in group A. We can then compute \( d_1(A \times A) \) as follows: if the number of pluses in A × A is zero, then \( d_1(A \times A) = 0 \); otherwise,

\[
d_1(A \times A) = \frac{Pp(A \times A) \times Pp(A \times A)}{\sqrt{(Pp(A \times A))^2 + (Pp(A \times A))^2}}
\]

The values of \( d_1(A \times A) \) range from 0 to 0.707, estimating to what extent similarities higher than expected at random (\(+\)) predominate within zone A × A.

We define \( Pm(A \times A) \) and \( Pm(A \times A) \) as the proportion of minuses in zone A × A and the proportion of OGUs in group A with at least one minus, respectively, and these are computed in the same way as \( Pp(A \times A) \) and \( Pp(A \times A) \), but taking into account the minuses. We then define \( d_2(A \times A) \) in the following way: if the number of minuses in zone A × A is zero, then \( d_2(A \times A) = 0 \); otherwise,

\[
d_2(A \times A) = \frac{Pm(A \times A) \times Pm(A \times A)}{\sqrt{(Pm(A \times A))^2 + (Pm(A \times A))^2}}
\]

The values of \( d_2(A \times A) \) range from 0 to 0.707, estimating to what extent similarities lower than expected at random (\(−\)) predominate within zone A × A. We define \( Pp(A \times B) \) and \( Pp(A \times B) \) in a similar way to \( Pp(A \times A) \) and \( Pp(A \times A) \), but with reference to zone A × B. So, \( d_3 \) is zero when the number of pluses in A × B is zero; otherwise,
The values of $d_3$ range from 0 to 0.707, estimating to what extent similarities higher than expected at random (+) pre-dominate within zone $A \times B$.

The parameter $DW(A \times A)$ measures to what extent the similarities that are higher than expected (+) tend to be in zones $A \times A$ but not in $A \times B$ (see McCoy et al. 1986), where $DW(A \times A) = d_1(A \times A) - d_2(A \times A) - d_3$.

Similarly, $DW(B \times B) = d_1(B \times B) - d_2(B \times B) - d_3$, where $d_1(B \times B)$ and $d_2(B \times B)$ are calculated as $d_1(A \times A)$ and $d_2(A \times A)$, but computing the pluses and minuses in zone $B \times B$. The average of $DW(A \times A)$ and $DW(B \times B)$, named $DW$, measures to what extent similarities that are higher than expected (+) tend to be in either zones $A \times A$ or $B \times B$ but not in $A \times B$.

We define $d_4$ in the same way as $d_3$, but compute the minuses in $A \times B$. We then compute the parameter $DS = d_4 - d_3 - d_2(A \times A) - d_2(B \times B)$, which gives a measure of whether the similarities that are lower than expected (−) tend to be located in $A \times B$, but not in $A \times A$ or $B \times B$.

The statistical significance of a node was assessed using a G-test of independence (Sokal and Rohlf 1981, McCoy et al. 1986) of the distribution of the signs ‘+’, ‘−’ and ‘0’ in the three zones of the submatrix, and so we obtained the parameters $GW$, for weak boundaries, and $GS$, for strong boundaries. If similarities higher than expected (+) tend significantly to be in zones $A \times A$ or $B \times B$, but not in $A \times B$, that is, if $DW > 0$ and $GW$ is statistically significant, then there is at least a weak biotic boundary between both groups of OGUs. In this case, if $DW(A \times A) > 0$ then the group of OGUs $A$ constitutes a biotic region, and the same applies for $DW(B \times B)$ and the group of OGUs $B$, because then a group of OGUs with similar biota shares a number of species compatible with random expectation with another group of OGUs that also has a similar biota. If similarities significantly lower than expected (−) tend to be located in $A \times B$, but not in $A \times A$ or $B \times B$, that is, if $DS > 0$ and $GS$ is significant, then a strong biotic boundary exists between the groups of OGUs. In this case a group of OGUs shares fewer species than expected at random with another group of OGUs. The areas delimited by strong boundaries are biotic supraregions. A supraregion can consist of several biotic regions separated by weak boundaries.