

Contribution of restricted and widespread species to diversity: the effect of range cohesion

Héctor T. Arita and Gerardo Rodríguez-Tapia

H. T. Arita (arita@oikos.unam.mx), Inst. de Ecología and Centro de Investigaciones en Ecosistemas, Univ. Nacional Autónoma de México, Apartado Postal 27-3, CP 58090, Morelia, Michoacán, México. – G. Rodríguez-Tapia, Inst. de Ecología, Univ. Nacional Autónoma de México, CP 04510, México, D. F., México.

One of the most noticeable patterns in biogeography is the non-uniform distribution of species richness across continents. This pattern is ultimately determined by the size, shape, and location of species ranges and the concomitant way in which they overlap to set the number of species potentially occurring in a site. Several recent papers have shown a relationship between the size of a species range and its contribution to the geographic patterns of diversity (reviewed in Gaston 2008). For the birds of sub-Saharan Africa, the correlation between overall species richness and the richness of narrow-ranging taxa was found to be relatively low, but to increase when the comparison was made with widespread species (Jetz and Rahbek 2002). Similar patterns have been shown for African and British birds (Lennon et al. 2004), Mexican mammals (Vazquez and Gaston 2004), Pacific fish (Mora and Robertson 2005), Neotropical palms (Kreft et al. 2006), and several taxa in Switzerland (Pearman and Weber 2007). These findings have been extended to the species richness-energy relationship (Evans et al. 2005), and have been considered in optimization algorithms for the prioritization of areas for conservation purposes (Wilhere et al. 2008).

Lennon et al. (2004) examined the relative contribution of restricted and widespread African and British birds to patterns of diversity by correlating, site by site, overall richness with richness values generated by subsets of taxa ranked in range size order. Subsets including the most widespread species showed stronger correlations with overall richness than subsets of the most restricted species. Lennon et al. (2004) demonstrated this tendency by plotting correlation values against cumulative numbers of species for sequences starting with the most widespread or with the most restricted species, finding that in most cases the line for the former were above the line for the latter. They concluded that widespread species exert a stronger influence than restricted taxa do on overall patterns of diversity, and

proposed the search for mechanistic explanations for this general tendency.

Here we show that the correlative patterns discovered by Lennon et al. (2004) are, under some circumstances, statistical consequences of two traits of natural geographic ranges: 1) their cohesiveness, and 2) the shape of the frequency distribution of their sizes.

A one-dimensional case

Lennon et al. (2004) measured the information content of subsets of species using the cumulative binomial variance. The variance in species richness for a set of sites generated by the presence-absence pattern of a given species is $\text{Var}(i) = p_i(1 - p_i)$ where p_i is the proportion of sites in which species i occurs. Lennon et al. (2004) described the distribution of ten species among ten sites with a 10×10 presence-absence matrix. A species occurring in five of the sites would generate the maximum variance in species richness among sites ($0.5 \times 0.5 = 0.25$), whilst a species found in the ten sites would have the lowest variance (0.0). As pointed out by Lennon et al. (2004), a species occurring in only one of the sites would generate the same amount of variance as one occurring in nine ($0.1 \times 0.9 = 0.9 \times 0.1 = 0.09$). This pattern is mirrored by the number of ways in which a range can be accommodated among the ten sites. There is only one configuration for a range size equal to ten: all sites occupied. There are $10!/(5! \times (10 - 5)!) = 252$ possible configurations for a five-celled range, but only 10 combinations for a one-celled or a nine-celled range. Thus, more information is needed to define the position of a non-cohesive five-celled range ($\log_2 252 = 7.98$ bits) than to describe the configuration of a one-celled or a nine-celled range ($\log_2 10 = 3.32$ bits). A bit is the amount of information contained in one binary digit, that is, a choice between 0 and 1 (Shannon and Weaver 1949, Pierce 1980).

Figures so far correspond to non-cohesive ranges, that is, sets of sites that can be scattered across the continent, with no restriction in regards of their closeness between each other.

Now imagine that the ten cells are arranged linearly, forming a one-dimensional domain and that ranges are cohesive. In this case, sites constituting the range must form a continuous line, as in mid-domain models of species richness (Colwell and Lees 2000). Because of this spatial restriction, there are only $10 - (r_i - 1) = 11 - r_i$ ways of accommodating a range of size r_i . Thus, there are still 10 possible configurations for a one-celled range (3.32 bits), but only six for a five-celled range (2.58 bits) and two (1.0 bits) for a nine-celled range. Modelled as cohesive extents, larger ranges show less uncertainty, so they require less information to define their position. Notice that the metric of information proposed by Lennon et al. (2004) is invariant to cohesiveness, as the binomial variance for a given range size is the same in both cases.

Lennon et al. (2004) presented a simple example in which overall species richness correlated stronger with subsets of restricted species than with subsets of widespread ones (their Fig. 3). They contrasted this single example to their empirical results, in which overall richness showed higher correlations with widespread species than with restricted taxa. However, they did not perform a statistical analysis of how representative was their example. To explore such a question, we developed a null model to examine the patterns of richness correlation among sets of species that are randomly distributed.

We defined a “continent” as a ten-cell linear domain, and created sets of 10 species with range sizes corresponding to those in the example of Lennon et al. (2004): 1, 2, 3, 4, 5, 6, 7, 8, 9, and 9 cells. First we modelled non-cohesive ranges, assigning occurrences by sampling (with no replacement) from the pool of ten cells until reaching the corresponding range size for each species. Then, we correlated the overall species richness with the patterns generated by each species alone. We then subtracted the coefficient corresponding to the most restricted species (occurring in one cell) from the coefficient corresponding to the most widespread species (occurring in nine cells), to see which of the correlation coefficients was higher (a positive difference in coefficients meaning that the correlation is stronger for the widespread species). We repeated the procedure 20 000 times to sample the universe of 3.2×10^{17} possible permutations. We repeated the procedure for cohesive ranges; in this case, each simulation sampled randomly from the pool of possible arrangements corresponding to each range size: 10 continuous lines for the one-celled range, 9 for the two-celled range, and so on. We ran 20 000 simulations to sample from the universe of 36.3 million possible permutations.

Simulations with non-cohesive ranges showed stronger correlations for the ranges of intermediate size (five cells), whereas cohesive ranges generated a pattern in which larger ranges produced higher correlation coefficients (Fig. 1). In the latter case, correlation coefficients for the largest ranges were not the highest, but still they were higher than those for the most restricted species. Figure 1 clearly shows an

effect of range cohesion on the way overall species richness correlates with patterns generated species by species.

Simulations also showed that cases in which the correlation corresponding to the most restricted species is higher than that of the most widespread species are not uncommon. In fact, 10 979 of the 20 000 simulations ($p = 0.55$) generated such a trend, whereas in 8745 cases the correlation was higher for the most widespread species ($p = 0.44$). In the remaining 276 cases ($p = 0.014$) the difference in correlations was exactly 0.0 (Fig. 2A). Thus, the example of Lennon et al. (2004) is not unexpected in terms of showing a stronger correlation for the most restricted species (a negative value in Fig. 2A). It is, however, highly improbable in terms of the magnitude of the difference between the correlations. In the simulations, the probability of observing a value equal or less than the one of the example (-1.083) is $p = 0.0063$.

Results modelling cohesive ranges showed a different pattern in which only 3076 cases ($p = 0.15$) yielded a higher correlation for the restricted species and 16 924 cases ($p = 0.85$) generated a higher correlation for the widespread species (Fig. 2B). A pattern as extreme as the example in Lennon et al. (2004) cannot be generated with cohesive ranges, or at least was not observed among the 20 000 simulations ($p < 5 \times 10^{-5}$). In fact, it is not even possible to generate a pattern in which the correlation using the most widespread species is negative, as in the example of Lennon et al. (2004).

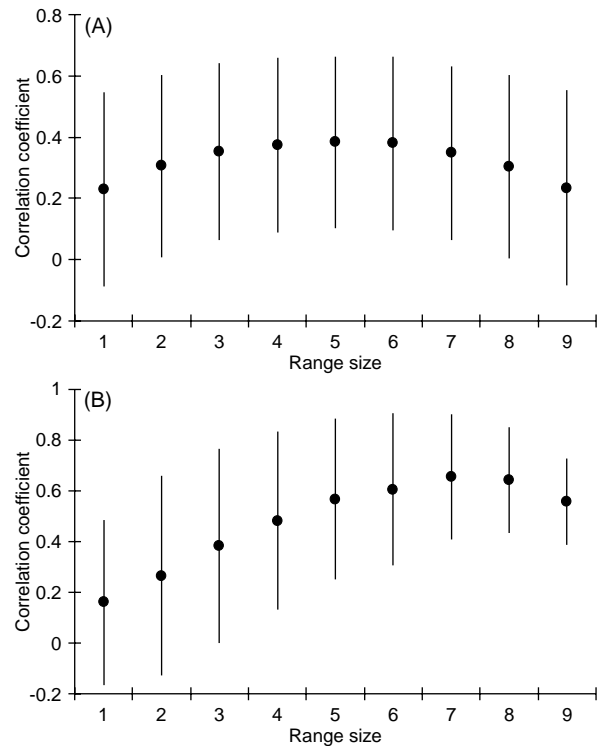


Figure 1. Pearson's correlation coefficients (mean \pm one standard deviation of 20 000 repetitions) between overall species richness and richness generated by one species at a time for a set of ten species distributed in ten sites. Correlations are shown for non-cohesive (A) and cohesive (B) one-dimensional ranges spanning in size from one to ten cells.

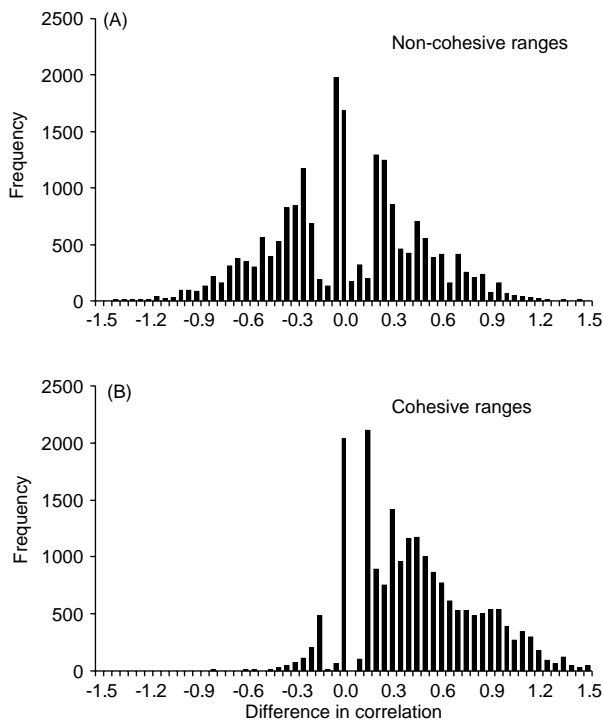


Figure 2. Difference in correlation coefficients between a widespread species (occurring in nine cells) and a restricted one (occurring in a single site). Correlations are between overall species richness and the richness pattern generated by one species at a time. Positive values mean that the correlation for the widespread species is higher than that of the restricted one. Results are for 20000 repetitions of a simulation involving species with non-cohesive (A) and cohesive (B) ranges.

A two dimensional case

We modelled a two dimensional case with a “continent” consisting of 400 quadrats arranged in a 20×20 square. As in the one-dimensional case, the highest number of combinations occurred when non-cohesive ranges were about half the size of the continent. There are ca 10^{119} configurations for a 200-quadrat range (171.7 bits) but a single-cell range can be accommodated in only 400 different forms (8.64 bits). Enforcing cohesion, there are only 49 configurations for a square-shaped 14×14 (196-celled) range (5.6 bits) and four possible ways of arranging a 19×19 (361-celled) range (2 bits). The number of possible configurations in these cases is actually higher if shapes other than squares are considered, but still the number of possible arrangements is much lower than with non-cohesive ranges and, as in the one-dimensional case, the number of possible cohesive configurations is lower as the size of the range increases.

To explore the null distribution of correlation values we simulated cases with and without range cohesion. To simulate the distribution of non-cohesive ranges, cells were drawn randomly, without replacement, from the pool of 400 cells until attaining the corresponding range size. For cohesive ranges, we used the spreading-dye algorithm (Jetz and Rahbek 2001): a starting cell was chosen at random, and new cells were added, also randomly, in the periphery of the simulated range until

attaining the corresponding range size. We used sets of one hundred species, considering two cases. In the first one, the distribution of range size was randomly uniform in the interval (1, 400); in the second we used the range size frequency distribution (RSFD) of North American mammals, re-scaling it to the interval (1, 400) to generate 100 ranges. As in other continental assemblages, the North American mammal fauna follows a highly right-skewed distribution of range sizes, with many restricted species and a few ones with wide distribution (Arita et al. 2005). Calculation of correlation coefficients and of the difference between the values corresponding to the most widespread and the most restricted species were done as in the one-dimensional case. For the two-dimensional case, however, we also calculated the parameters for sets of fifty species, that is, for the most restricted and the most widespread halves. We ran 1000 repetitions of each combination of cases.

In the simulations with a uniform random distribution of range sizes, patterns generated by cohesive ranges differed markedly from those corresponding to scattered ranges (Fig. 3A, C). When considering the most widespread and the most restricted species (Fig. 3A), the difference in correlation coefficient had an average of 0.021 for non-cohesive ranges, with 389 cases showing a difference < 0.0 . For cohesive ranges, the mean value of the difference in correlation was 0.20, and only six cases produced a difference < 0.0 . A similar pattern emerged when considering the fifty most widespread and the fifty more restricted species (Fig. 3C): the mean for non-cohesive ranges was 0.007, with 559 cases being < 0.0 , whereas the mean for cohesive ranges was 0.06, with no case generating a value < 0.0 .

The difference between the patterns for cohesive and non-cohesive ranges blurred a little when examining range sizes derived from the empirical distribution, but still showed a predominance of larger differences in correlation for cohesive ranges (Fig. 3B, D). When considering only the most widespread and the most restricted species, the mean for non-cohesive ranges was 0.17, with 15 cases < 0.0 ; the mean for cohesive ranges was 0.33, with 71 cases < 0.0 (Fig. 3B). When comparing groups of fifty species, the mean was 0.73 for non-cohesive ranges and 0.77 for cohesive ranges; in both cases, no simulation yielded a difference in correlation < 0.0 (Fig. 3D).

The effect of cohesion and range-size frequency distribution

Results of the simulations show a distinctive effect of cohesive ranges on correlations of species richness. The relative contribution of restricted and widespread species to the overall patterns of richness depends on the way ranges are represented, either as sets of scattered cells or as cohesive ranges. This difference is apparently related to the degree of uncertainty in the location of a range, given only its size. For scattered ranges, the amount of information needed to define their location (uncertainty) is highest for medium-sized ranges, whereas for cohesive ranges the amount of uncertainty is highest for the smallest ranges. This difference makes highly unlikely the cases where small cohesive

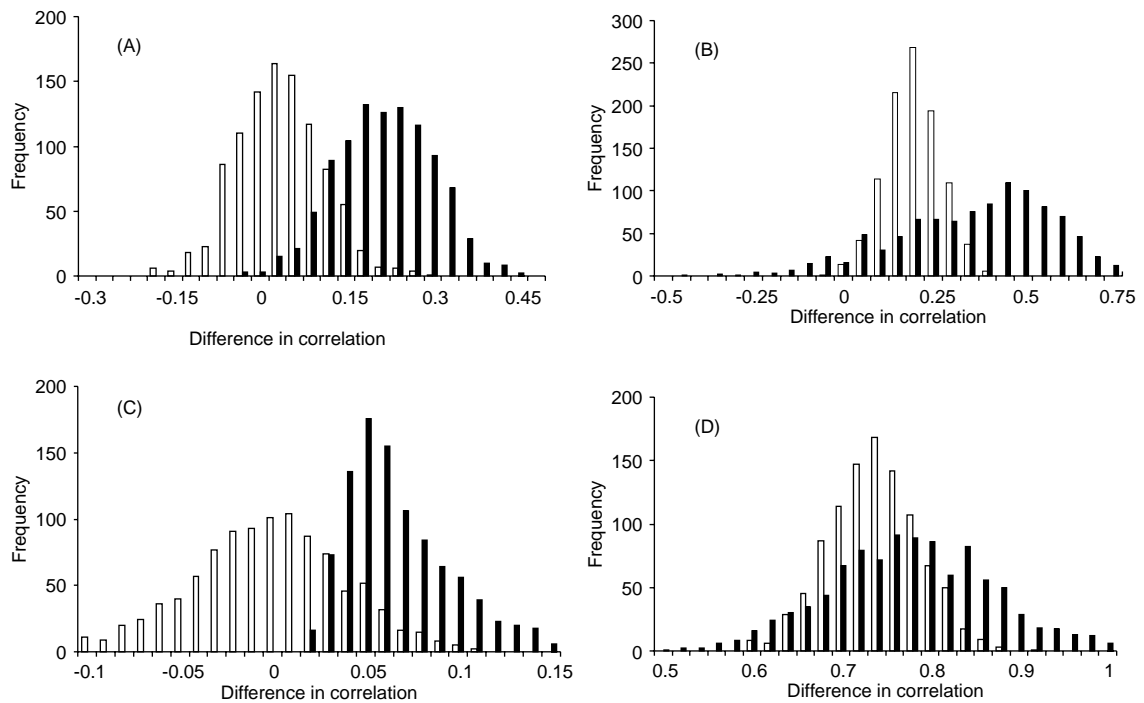


Figure 3. Difference between correlation coefficients of widespread and restricted species richness with overall richness in a 20×20 -quadrat continent. Results are for 1000 simulations. In A and C, the range size of 100 species follows a random uniform distribution in the interval (1, 400). In B and D, the frequency distribution was taken from that of North American mammals, but scaled to the interval (1, 400). (A) and (B) show the difference in correlation between the most widespread and the most restricted species. (C) and (D) show the difference between the correlation of the 50 most widespread species and the 50 most restricted ones. In all cases, open bars correspond to non-cohesive ranges and filled bars are for cohesive ranges.

ranges correlate better than large ranges with overall species richness (Fig. 2 and 3A, D). Moreover, skewed RSFDs seem to exacerbate this tendency, in such a way that higher correlations for restricted species constitute extremely unlikely cases that can be considered impossible for all practical purposes (Fig. 3B, D).

The two cases discussed here (cohesive vs randomly scattered ranges) are extremes in a continuum of possible arrangements, determined by the degree of spatial autocorrelation. We acknowledge that, strictly speaking, totally cohesive ranges do not exist, so most large-scale analyses of species distribution suffer from “errors of commission” caused by calculated geographic extents (“range areas”) being larger than the actual occupancy within the range (La Sorte and Hawkins 2007). However, at the continental and global scales, ranges tend to show a very high degree of cohesion. Although showing some degree of internal structure and geographic variation (Brown et al. 1996, Gaston 2003, Hurlbert and Jetz 2007), they are generally modelled as continuous one- or two-dimensional extents (Colwell and Lees 2000, Jetz and Rahbek 2001). At smaller scales, ranges generally increase their degree of non-cohesiveness. For example, the distribution of species among islands in an archipelago or the patterns of occupancy of plants and animal in small-scale sampling sites can be highly scattered.

Our measure of information content takes into account those differences, mirroring patterns found in examples from other disciplines. The information content of written messages depends on the correlation between letters. In English, for instance, there is a much higher probability of

finding the letter sequence “TH” than it is of finding the combination “QU” (Pierce 1980). This and other correlative patterns determine that less information is required to define a text written in English than it is to describe a sequence of random letters. In the same manner, the information content of ranges depends on the correlation between sites; in cohesive ranges, the probability that contiguous sites contain the same species is high, so the amount of information required to define a cohesive range is less than that needed to describe a set of scattered sites. The binary variance metric (Lennon et al. 2004) is insensitive to these differences.

Simulations presented here also document an effect of skewed RSFDs on the correlation of overall species richness with patterns generated by subsets of species. Such frequency distributions are typical of continental and global assemblages of species (Gaston 2003), although they vary with scale and tend to be more uniform at regional scales (Arita and Rodriguez 2002). There might be also a correspondence between cohesive ranges and the shape of the RSFD. In one-dimensional domains, the random placement of cohesive ranges generates a triangular RSFD with many restricted and few widespread species (Colwell and Lees 2000, Arita 2005). In two dimensions, the same random procedure yields a right-skewed RSFD similar to the one observed for continental assemblages of species (Arita 2005).

An effect of range cohesion has been shown for other patterns of diversity. In neutral models, incorporating a short-distance migration parameter generates distributional patterns for species that resemble cohesive ranges (Rangel

and Diniz-Filho 2005). Cohesive ranges, in turn, generate species richness patterns similar to those predicted by two-dimensional mid-domain effect model, whereas scattered ranges generated by long-distance dispersal produce no discernible geographic pattern of species richness (Rangel and Diniz-Filho 2005). The correlation of species richness patterns with environmental variables has been shown to be higher for widespread species than for restricted ones, both in static (Jetz and Rahbek 2002, Mora and Robertson 2005) and dynamic models (Rahbek et al. 2007) of range location. Null models in these studies corroborate the tendency of widespread species to show stronger correlations than restricted species, suggesting the existence of statistical constraints similar to those documented here.

In conclusion, for large-scale sets of species, a higher correlation between overall species richness and richness generated by widespread species is the null expectation, generated by the cohesion of species ranges and the shape of the RSFD. This fact has to be taken into account in the search for mechanistic explanations for the patterns recently reported in the literature (Vazquez and Gaston 2004, Lennon et al. 2004, Kreft et al. 2006, Pearman and Weber 2007). Similar patterns have been also found for non-cohesive distributions (Lennon et al. 2004), and reversed patterns have been documented for continental sets (Vazquez and Gaston 2004), which highlights the relevance of the continuing search for the biological mechanisms that determine the relationship between diversity and distributions.

Acknowledgements – We thank J. Navarro, P. Rodríguez, E. Vázquez-Domínguez, and F. Villalobos for comments on an early version of the manuscript. Valuable comments and suggestions by H. Kreft, J. J. Lennon, and one anonymous reviewer helped to improve this work. Support for this project was provided by the Inst. of Ecology, National Autonomous Univ. of Mexico (UNAM).

References

- Arita, H. T. 2005. Range size in mid-domain models of species diversity. – *J. Theor. Biol.* 232: 119–126.
- Arita, H. T. and Rodríguez, P. 2002. Geographic range, turnover rate and the scaling of species diversity. – *Ecography* 25: 541–550.
- Arita, H. T. et al. 2005. Continental and regional ranges of North American mammals: Rapoport's rule in real and null worlds. – *J. Biogeogr.* 32: 961–971.
- Brown, J. H. et al. 1996. The geographic range: size, shape, boundaries, and internal structure. – *Annu. Rev. Ecol. Syst.* 27: 597–623.
- Colwell, R. K. and Lees, D. C. 2000. The mid-domain effect: geometric constraints on the geography of species richness. – *Trends Ecol. Evol.* 15: 70–76.
- Evans, K. L. et al. 2005. Relative contribution of abundant and rare species to species-energy relationships. – *Biol. Lett.* 1: 87–90.
- Gaston, K. J. 2003. Structure and dynamics of geographic ranges. – Oxford Univ. Press.
- Gaston, K. J. 2008. Biodiversity and extinction: the importance of being common. – *Prog. Phys. Geogr.* 32: 73–79.
- Hurlbert, A. H. and Jetz, W. 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. – *Proc. Nat. Acad. Sci. USA* 104: 13384–13389.
- Jetz, W. and Rahbek, C. 2001. Geometric constraints explain much of the species richness pattern in African birds. – *Proc. Nat. Acad. Sci. USA* 98: 5661–5666.
- Jetz, W. and Rahbek, C. 2002. Geographic range size and determinants of avian species richness. – *Science* 297: 1548–1551.
- Kreft, H. et al. 2006. The significance of geographic range size for spatial diversity patterns in Neotropical palms. – *Ecography* 29: 21–30.
- La Sorte, F. A. and Hawkins, B. A. 2007. Range maps and species richness patterns: errors of commission and estimates of uncertainty. – *Ecography* 30: 649–662.
- Lennon, J. J. et al. 2004. Contribution of rarity and commonness to patterns of species richness. – *Ecol. Lett.* 7: 81–87.
- Mora, C. and Robertson, D. R. 2005. Causes of latitudinal gradients in species richness: a test with fishes of the Tropical Eastern Pacific. – *Ecology* 86: 1771–1782.
- Pearman, P. B. and Weber, D. 2007. Common species determine richness patterns in biodiversity indicator taxa. – *Biol. Conserv.* 138: 109–119.
- Pierce, J. R. 1980. An introduction to information theory – Dover.
- Rahbek, C. et al. 2007. Predicting continental-scale patterns of bird species richness with spatially explicit models. – *Proc. R. Soc. B* 274: 165–174.
- Rangel, T. F. L. V. B. and Diniz-Filho, J. A. F. 2005. Neutral community dynamics, the mid-domain effect and spatial patterns in species richness. – *Ecol. Lett.* 8: 783–790.
- Shannon, C. E. and Weaver, W. 1949. The mathematical theory of communication. – The Univ. of Illinois Press.
- Vazquez, L. B. and Gaston, K. J. 2004. Rarity, commonness, and patterns of species richness: the mammals of Mexico. – *Global Ecol. Biogeogr.* 13: 535–542.
- Wilhere, G. F. et al. 2008. Average optimacity: an index to guide site prioritization for biodiversity conservation. – *Biol. Conserv.* 141: 770–781.