

Dispersion fields, diversity fields and null models: uniting range sizes and species richness

Michael Krabbe Borregaard and Carsten Rahbek

M. Krabbe Borregaard (mkrborregaard@bio.ku.dk) and C. Rahbek, Center for Macroecology, Evolution and Climate, Dept of Biology, Univ. of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark.

One of the cornerstones of macroecological research is the ongoing effort to understand large-scale patterns of species richness. These patterns are emergent properties of the distributional ranges of individual species, as they are formed by the overlap of species ranges in a given area. As such, the sizes of ranges, and the processes controlling their geographical location, are key determinants of richness patterns. However, a satisfactory link between range size distributions, the spatial location of individual ranges, and species richness has been slow to emerge.

To forge this link, a central part is the role of species associations in determining the composition of species in a defined area. Overlap between species ranges may arise because species have similar ecologies (Webb 2000), for historical reasons of dispersal (Svenning et al. 2008), or may simply be random, because ranges are constrained by the shape of the geographical domain (Colwell and Lees 2000, Jetz and Rahbek 2001). However, although species interactions have consistently been shown to affect species co-occurrence at local scales (Gotelli and McCabe 2002), the importance of species associations for distributional overlap at biogeographical scales remains a central question for macroecology (Gotelli et al. 1997, in press).

Species' ranges are usually continuous at large scales, and as a consequence, the species richness values of closely located sites are not independent. This means that the spatial pattern of species richness cannot be explained by analyzing sites as a set of independent points in a regression (Legendre 1993). Spatial regression methods solve the statistical issue of autocorrelation (Rangel et al. 2006), but a simplistic use of these statistics risks missing the main point. Spatial patterning is not a statistical issue – it is an inherent quality of biogeographical data (Rahbek and Graves 2000, Diniz-Filho et al. 2003). In the light of this, recent attention has focused on developing conceptual and analytical tools for macroecological analysis that deal explicitly with species' ranges.

One important advance is the concept of the “dispersion field”, developed by Graves and Rahbek (2005). The dispersion field is the set of geographical ranges of all species that occur in a given site. Just as the continental

species richness pattern is created by the overlap of all species in a continent, the dispersion field can be visualized as the pattern created by overlapping the ranges of all species occurring in a given cell (Fig. 1). These dispersion fields have striking geometric shapes, and have a number of promising applications.

First, it has been argued that the geometric shape of dispersal fields are an approximation to the regional species source pool (Graves and Rahbek 2005). The source pool plays a key role in theories of community assembly, but the concept has been consistently difficult to pin down (Gotelli and Graves 1996). Even more importantly, dispersion fields visualize the species associations that create richness patterns. Hence, they provide an opportunity for more stringent tests of ecological hypotheses for species richness than standard regression methods. A growing research paradigm in macroecology is to replace curve-fitting methods with mechanistic models of range placement (Rahbek et al. 2007, Rangel et al. 2007, Gotelli et al. 2009). Such models also generate predictions on the structure of dispersion fields. Comparing both richness patterns and dispersion fields to modeled patterns thus constitutes an opportunity for validation of these models at two hierarchical levels, a standard for pattern-oriented modeling (Grimm et al. 2005).

A promising approach for investigating the link between range sizes and richness patterns has recently been developed by Arita et al. (2008). This approach is based on dispersion fields, and starts with the presence–absence matrix of sites versus species. In this matrix, the columns are sites, rows are species, and the matrix elements represent the presence (1) or absence (0) of a given species in a particular site (Gotelli 2000). The strength of the presence–absence matrix is that it combines information on species richness (which are the column sums), range sizes (which are the row sums), and the co-occurrence of species (which can be measured by the degree of co-variance in the matrix). However, the presence–absence matrix is not easy to visualize graphically.

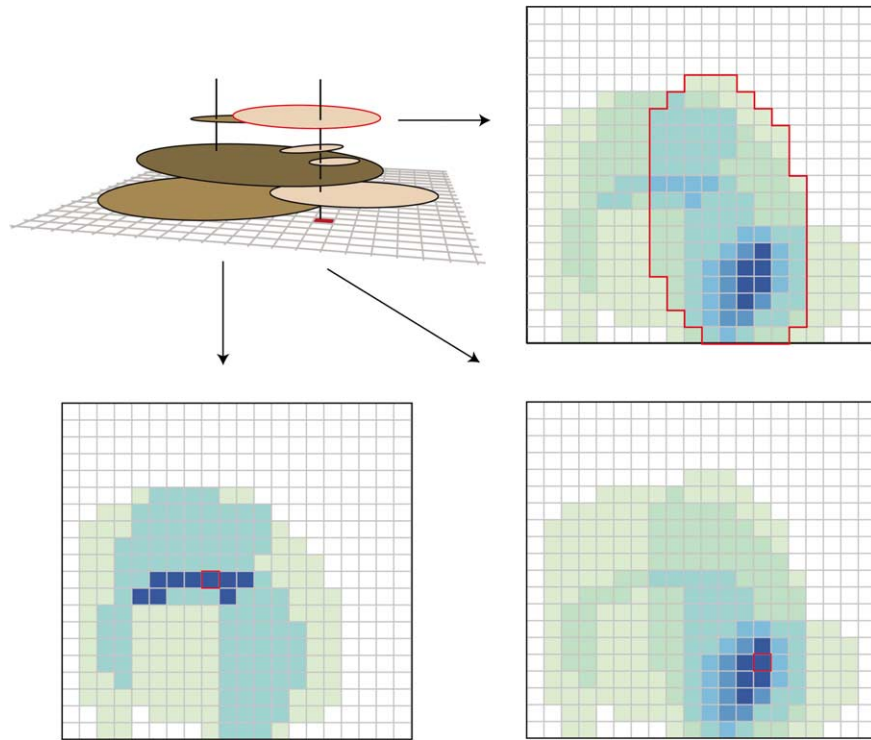


Figure 1. Illustration of the dispersal fields and diversity fields of Graves and Rahbek (2005) and Arita et al. (2008), respectively. Top left: a schematic illustration of elliptical ranges. Vertical lines indicate two focal cells (marked as red squares) for the dispersion fields shown in the bottom panels. A red ellipse illustrates a focal range for one diversity field. Ranges of the same color belong to the same diversity field. One range belongs to both diversity fields. Top right: the richness map resulting from overlap of the ranges in top left panel. One range is illustrated as a red outline. The richness values within this outline constitute the diversity field. Bottom left: the dispersion field for the leftmost focal cell in the top left panel. Bottom right: the dispersion field for the rightmost focal cell in the top panel.

One approach is to create “range-diversity plots” (Arita et al. 2008), which are scatter plots that combine information from the columns and rows of the presence-absence matrix (Fig. 2, from Arita et al. 2008). The dispersion field of a site can be calculated from the presence-absence matrix, as the mean range size of all species that occur at the site. A “by sites” range-diversity plot is then created by plotting this value against the species richness of the site (similar to Fig. 2 of Graves and Rahbek 2005). Because dispersion fields allow range sizes

to be expressed in the same currency as richness values (i.e. individual sites), the range-diversity plot makes it possible to investigate their relationship directly.

Analogously to the “dispersion field” of Graves and Rahbek (2005), Arita et al. (2008) developed the concept of the “diversity field”, which is the set of richness values of sites within the range of a given species (Fig. 1 upper right). The diversity field is illustrated in a “by species” range-diversity plot, which plots the mean species richness of sites occupied by a species against the range size of that

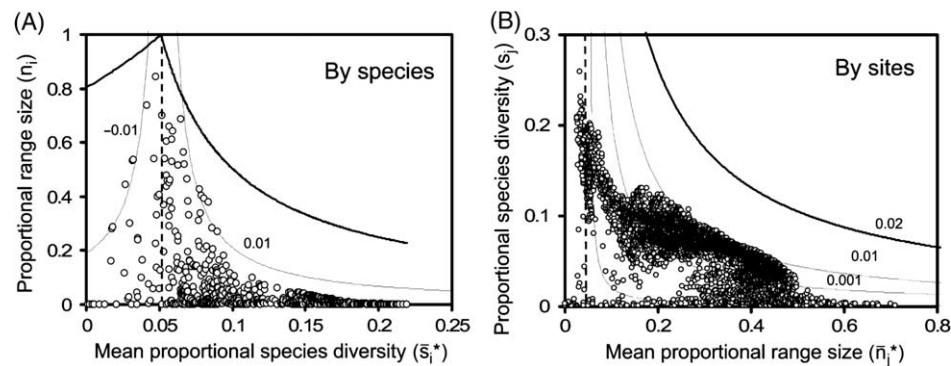


Figure 2. The range-diversity plot presented by Arita et al. (2008). Left: the range-diversity plot by species for North American mammals. See text or (Arita et al. 2008) for a description of the constraint lines in the plot. Right: the range-diversity plot for sites; same dataset as left panel. Whole plot taken directly from Arita et al. (2008).

species (Fig. 2 left). “By sites” and “by species” range-diversity plots are complementary, and together visualize the presence–absence matrix.

The main strength of range-diversity plots is that they clarify the connection between richness values and range sizes: range sizes and species richness values of a community are linked because they are the marginal sums of the presence–absence matrix. Because of this link, the points in range-diversity plots become constrained to certain regions of the plot area. Thus, the dispersion of points over the plot region can be interpreted to yield information on the ecological processes structuring the assemblage.

To facilitate the interpretation of range-diversity plots, Arita et al. (2008) mathematically developed a set of constraint lines for point dispersion. First, they use a thick, solid line to mark regions of the plot that cannot be occupied by points. These areas represent impossible combinations of e.g. mean range and species richness, and are calculated as a mathematical function of the minimum and maximum values of range size or species richness (Fig. 2).

Additionally, Arita et al. (2008) added thin lines that connect areas of equal covariance. They argue that what creates the dispersion of points in range-diversity plots is covariance in the presence-absence matrix, which reflects associations between species (in the “by species” plot) or similitude between sites (in the “by sites” plot). Accordingly, species with a similar degree of covariance between their geographic distribution and the distributions of all other species should align along these lines, when observing a “by species” plot.

In the range-diversity plots presented by Arita et al. (2008), the points are widely dispersed across the plot area (Fig. 2). The point clouds have characteristic shapes, and the points all lie within a region that is clearly smaller than the permissible area delineated by the thick line. Also, most points are located to the right of the “fill” line (showing the grand mean), which indicates positive covariance for both sites and species.

Arita et al. (2008) interpret the patterns of point dispersion in range-diversity plots as the results of ecological processes. For instance, they argue that the general orientation of points in the “by sites” plot (Fig. 2 right) is created by a combination of Rapoport’s rule and the latitudinal gradient of species richness. However, to argue that patterns are created by ecological processes, and not by mathematical constraints on point dispersion, requires that the mathematical constraints are well described.

Also, for range-diversity plots to be a useful tool, they should reveal patterns that result from a relationship between range sizes and richness values. If the patterns in range-diversity plots are just functions of the range size and richness frequency distributions themselves, it would be simpler to investigate these distributions separately. Thus, to evaluate the analytical power of range-diversity plots, the pertinent questions are: a) is the entire area within the solid line available for points, or is point dispersion constrained by other factors? And b) is the pattern of point dispersion created by associations among species/sites, or does it result from some other aspect of the calculation of range-diversity plots?

To answer these questions, we constructed range-diversity plots for a high-quality dataset of the birds of South America (Rahbek and Graves 2001). This dataset contains 2869 species in 1676 one-degree grid cells. Thus, they contain fewer sites but more species than the dataset for North American mammals used by Arita et al. (2008). Of these 2869 species, 643 have ranges that extend into Central and North America (this was generally only a small part of their range). For these species, we considered only the range within South America, even though this leads to a, for most species, minor underestimation of their actual range. Repeating the analyses using only South American endemics does not affect any conclusions here (unpubl.).

The range-diversity plot by species

The “by species” range-diversity plot shows a pattern similar to Arita et al.’s (2008) pattern for North American mammals, although there are more points in the lower left corner, indicating a group of small-ranged species that occur in grid cells with low overall diversity (compare Fig. 2 left with Fig. 3 top left). The points fall within a conical shape, with most points located towards the right part of the plot. A general difficulty for the interpretation of range-diversity plots is that no standard statistical tests exist for testing point dispersion or the degree to which covariance lines explain the location of points. Still, for South American birds the points do not appear to follow the iso-covariance lines even at visual inspection. Instead the points appear to be constrained along a straight line at the right edge of the cloud of points. However, this constraint lies far from the solid line indicating the permissible area.

In addition to the permissible area constraint developed by Arita et al. (2008), the potential combinations of range size and mean diversity will also be constrained by the empirical distribution of richness values. A simple way of describing the mechanism for this is to start with a species with a range size of only one grid cell. The highest possible value of mean species richness at occupied sites (the x axis of the range-diversity plot) will then equal the highest species richness value on the map. If the range size is two grid cells, the highest mean richness value is the mean of the two most species-rich areas, and so forth. The highest richness reflected in our bird data is a grid cell close to Quito in Ecuador, and it is extremely high: 845 bird species within a one-degree latitudinal-longitudinal square. In the second most speciose cell, richness already decreases to 782 species, and the maximum value for mean species richness quickly decreases as more cells are included. We added this constraint to the plot (blue and red in Fig. 3). The fit of the constraint line to the point cloud is visually striking. This indicates that the richness frequency distribution is what constrains the point dispersion in the “by species” plot for South American birds.

The range size distribution is also very clearly visible, as it is the distribution of points along the y-axis. As Arita et al. (2008) clearly point out, ranges are proportions, and proportion data tend to stack up as they near the limits (0 and 1). At continental scales, this effect results in a highly skewed range-size distribution, with a large number of very

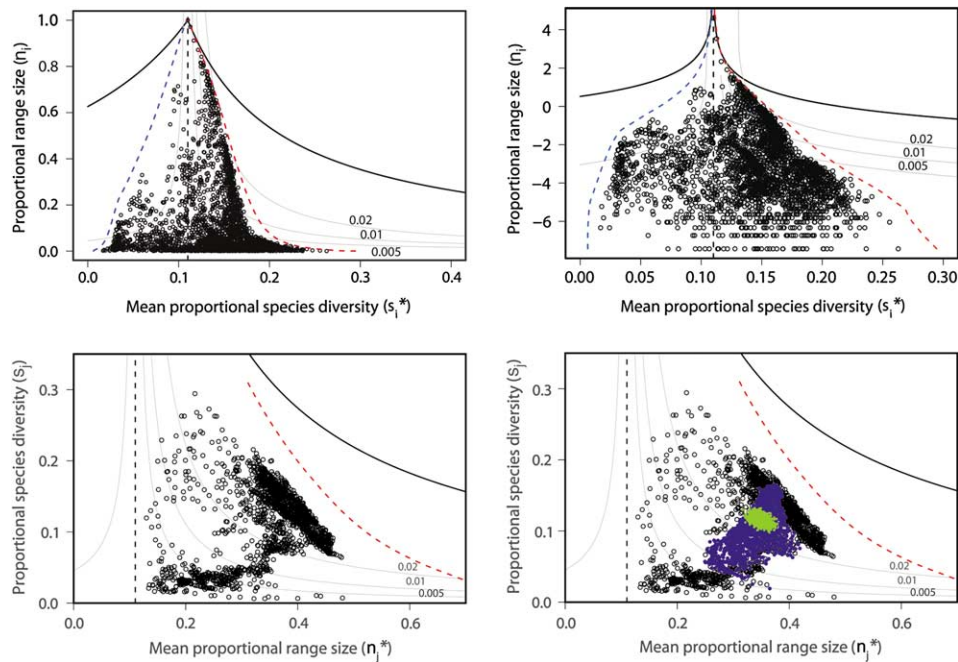


Figure 3. Top left: the range-diversity plot by species for South American birds. Plots show isoclines connecting areas of equal covariance and the mathematical constraint lines developed by Arita et al. (2008). An envelope delimiting the possible values is added in blue and red (see text for explanation of this null constraint). Top right: same as top left, except that range size values have been logit-transformed to remove the effect of skew in the range size frequency distribution. Bottom left: the range-diversity plot by sites for South American birds. The maximal constraint line of a null model similar the one shown in the top left panel is shown in red. Bottom right: same as bottom left, with points added indicating the results of two null models. The green points indicate a simple null model, accounting for only the higher contribution of larger-ranging species to the calculations. The blue points indicate a “spreading dye” model, which also incorporates the effects of the continuous nature of individual ranges.

small-ranged species (Gaston 1996, Graves and Rahbek 2005). The strong skew moves most points towards the bottom of the range-diversity plot, which potentially obscures any pattern between the points. To increase the linearity of proportion data, logit-transformation is often recommended prior to analysis (Sokal and Rohlf 1995). For South American birds, logit-transforming the range sizes yields a much more uniform pattern (Fig. 3 top right). The fit of the null model is even more apparent, and the points are evenly dispersed within the area between the null model lines.

Although there are points everywhere between the red and blue lines, there remains an overweight of points to the right of the “fill” line (Fig. 3 top right). Still, there is no strong evidence for an effect of biological inter-dependence between range size and mean species richness. Thus, the results indicate that the “by species” range-diversity plot for the South American bird fauna does not convey any additional information above that of the range size distribution.

The range-diversity plot by sites

The “by sites” plot (Fig. 3 bottom) for South American birds is very different, both from the “by species” plot and from Arita et al.’s (2008) plot for North American mammals (Fig. 2). The most striking differences are the absence of points around and to the left of the “fill” line,

and the pattern of point dispersion: the points fall into two separate point clouds with diverging shapes. In Fig. 3 (bottom left), we have also added constraint lines based on the range-size frequency distribution. These constraints follow a rationale similar to that presented for the “by species” plot: if a site contains only one species, the highest possible mean range value is the range size of the largest-ranging species of the assemblage; if it contains two species, the highest value is the mean of the two largest range sizes, and so on. The right edge of the right-most group of points seems to follow the null constraint line (though at a small distance), but the fit is less convincing than for the “by species” plot.

All of the points in the “by sites” plot for South American birds occur to the right of the “fill” line. Arita et al. (2008) argued that all points are expected to cluster around this line in the absence of biological processes, which means that deviations from this line indicate similarity between sites. However, unless the range size distribution is completely symmetric, most points are expected to lie to the right of the “fill” line, simply as a consequence of sampling effects. The reason is that larger-ranging species exist at more sites, and thus contribute a range size value to more data points in the “by sites” plot (for a discussion of the statistical consequences of range sizes see Jetz and Rahbek 2002, Lennon et al. 2004). As a result, each site does not randomly sample the range size frequency distribution of the entire assemblage. Thus, the mean range size of each site will generally be higher than the grand mean range size of the assemblage. Figure 3 (bottom right)

demonstrates this effect using a simple null model, which allocates sites randomly to each species while maintaining the empirical range size distribution (shown as green points). Even though the effect seems to explain why the points are located in the right side of the plot, it cannot account for the dispersion of points. The empirical points are widely dispersed in the plots, whereas the null model points are tightly clustered.

The wide dispersion of points in the “by sites” plot reveals a high level of spatial structure of the assemblage. Sites with high similarity are grouped together, whereas sites with a very different species composition are located further apart. The complex pattern of points indicates a possible role for ecological assembly processes in structuring species composition. However, a high degree of site similarity, and thus point dispersion, could also be generated simply by range cohesiveness. Species ranges usually consist of several closely located cells, and thus random overlap is expected to lead to adjacent cells being highly similar.

To investigate this effect, we created 2869 random species range maps using a spreading dye algorithm (Jetz and Rahbek 2001). This algorithm randomly places cohesive ranges on the geographic domain, while maintaining the empirical range size frequency distribution. We then generated range-diversity plots for this dataset, adding the points to Fig. 3 (blue points, bottom right panel). The set of points generated by this null model are much more dispersed across the plot. The center of gravity for the points is close to the points created without range cohesion (in green), and thus it does not seem that range cohesion in itself is responsible for the covariance of sites (for a discussion of range cohesion and the “by species” range-diversity plot, see Villalobos and Arita in press).

Incorporating range cohesion clearly generates a more realistic level of point dispersion. However, it still does not capture the empirical pattern. The empirical pattern thus probably reflects historical or ecological processes that have generated two disparate areas in South America: one large area where species demonstrate a high degree of nestedness in their distribution, and one in which there is very little nestedness.

The present analysis identified structural constraints on the point dispersion in range diversity plots, beyond those considered by Arita et al. (2008). For our dataset of South American birds, these constraints were more important for structuring range-diversity plots than were the constraint lines described by Arita et al. (2008). Given the diversity of data sets that can be analyzed with range-diversity plots, a promising research avenue is to investigate how these results generalize. For instance, spatial scale is known to be an important determinant of ecological patterns (Rahbek 2005, Nogues-Bravo et al. 2008). At finer scales, where range cohesion is lower and biotic interactions are more pronounced, it is likely that other processes drive the patterns in range-diversity plots (H. Arita pers. comm.). The diverse patterns observed in “by sites” plots, which are robust even after accounting for mathematical constraints, highlight the potential of these plots for generating and testing hypotheses on how species’ distributions create patterns of species richness.

Acknowledgements – The authors thank the Danish National Research Foundation for its support of the Center for Macroecology, Evolution and Climate. We also thank Hector Arita for insightful comments and constructive criticism on the manuscript, which significantly improved the final version.

References

- Arita, H. T. et al. 2008. Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications. – *Am. Nat.* 172: 519–532.
- Colwell, R. and Lees, D. C. 2000. The mid-domain effect: geometric constraints on the geography of species richness. – *Trends Ecol. Evol.* 15: 70–76.
- Diniz-Filho, J. A. F. et al. 2003. Spatial autocorrelation and red herrings in geographical ecology. – *Global Ecol. Biogeogr.* 12: 53–64.
- Gaston, K. J. 1996. Species-range-size distributions: patterns, mechanisms and implications. – *Trends Ecol. Evol.* 11: 197–201.
- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. – *Ecology* 81: 2606–2621.
- Gotelli, N. J. and Graves, G. R. 1996. *Null models in ecology*. – Smithsonian Inst. Press, Washington DC, USA.
- Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond’s assembly rules model. – *Ecology* 83: 2091–2096.
- Gotelli, N. J. et al. 1997. Co-occurrence of Australian land birds: Diamond’s assembly rules revisited. – *Oikos* 80: 311–324.
- Gotelli, N. J. et al. 2009. Patterns and causes of species richness: a general simulation model for macroecology. – *Ecol. Lett.* 12: 873–886.
- Gotelli, N. J. et al. in press. Macroecological signals of species interactions in the Danish avifauna. – *Proc. Nat. Acad. Sci. USA*.
- Graves, G. R. and Rahbek, C. 2005. Source pool geometry and the assembly of continental avifaunas. – *Proc. Nat. Acad. Sci. USA* 102: 7871–7876.
- Grimm, V. et al. 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. – *Science* 310: 987–991.
- 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.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? – *Ecology* 74: 1659–1673.
- Lennon, J. J. et al. 2004. Contribution of rarity and commonness to patterns of species richness. – *Ecol. Lett.* 7: 81–87.
- Nogues-Bravo, D. et al. 2008. Scale effects and human impact on the elevational species richness gradients. – *Nature* 453: 216–219.
- Rahbek, C. 2005. The role of spatial scale and the perception of large-scale species-richness patterns. – *Ecol. Lett.* 8: 224–239.
- Rahbek, C. and Graves, G. R. 2000. Detection of macroecological patterns in South American hummingbirds is affected by spatial scale. – *Proc. R. Soc. B* 267: 2259–2265.
- Rahbek, C. and Graves, G. R. 2001. Multiscale assessment of patterns of avian species richness. – *Proc. Nat. Acad. Sci. USA* 98: 4534–4539.
- 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. et al. 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. – *Global Ecol. Biogeogr.* 15: 321–327.
- Rangel, T. F. L. V. et al. 2007. Species richness and evolutionary niche dynamics: a spatial pattern-oriented simulation experiment. – *Am. Nat.* 170: 602–616.
- Sokal, R. R. and Rohlf, F. J. 1995. *Biometry: the principles and practice of statistics in biological research.* – W. H. Freeman.
- Svenning, J. C. et al. 2008. Postglacial dispersal limitation of widespread forest plant species in nemoral europe. – *Ecography* 31: 316–326.
- Villalobos, F. and Arita, H. T. in press. The diversity field of new world leaf-nosed bats (phyllostomidae). – *Global Ecol. Biogeogr.*
- Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. – *Am. Nat.* 156: 145–155.