THE PRESENCE-ABSENCE MATRIX RELOADED: THE USE AND INTERPRETATION OF RANGE-DIVERSITY PLOTS

Héctor T. Arita[1], Andrés Christen[2], Pilar Rodríguez[3] and Jorge Soberón[4]

[1]Centro de Investigaciones en Ecosistemas, Universidad Nacional Autónoma de México, Apartado Postal 27-3, CP 58090, Morelia, Michoacán, México, [2]Centro de Investigación en Matemáticas, Apartado Postal 402, CP 36000, Guanajuato, México, [3]Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Liga Periférico-Insurgentes Sur 4903, CP 14010 México, DF, México, [4]Natural History Museum, University of Kansas, Lawrence, Kansas 66045, USA

Correspondence: Héctor T. Arita, Centro de Investigaciones en Ecosistemas, Universidad Nacional Autónoma de México, Apartado Postal 27-3, CP 58089, Morelia, Michoacán, México.
E-mail: arita@oikos.unam.mx

Submitted to GLOBAL ECOLOGY AND BIOGEOGRAPHY

ABSTRACT

Aim A great deal of information on distribution and diversity can be extracted from presence-absence matrices (PAMs), the basic analytical tool of many biogeographic studies. This paper presents numerical procedures that allow the analysis of such information by taking advantage of mathematical relationships within PAMs. In particular, we show how range-diversity plots summarize much of the information contained in the matrices by the simultaneous depiction of data on distribution and diversity.

Innovation We use matrix algebra to extract and process data from PAMs. Information on distribution of species and on species richness of sites is computed using the traditional *R* (by rows) and *Q* (by columns) procedures, as well as the new *Rq* (by rows, considering the structure of columns) and *Qr* (by columns, considering the structure by rows) methods. Matrix notation is particularly suited for summarizing complex calculations using PAMs, and the associated algebra allows the implementation of efficient computational programs. We show how information on distribution and species richness can be depicted simultaneously in range-diversity (RD) plots, allowing a direct examination of the relationship between those two aspects of diversity. We explore properties of RD plots with a simple example, and use null models to show that while parameters of central tendency are not affected by randomization, the dispersion of points in RD plots do change, showing the significance of patterns of co-occurrence of species and of similarity among sites.

Main conclusion Species richness and range size are both valid measures of diversity that can be analyzed simultaneously with RD plots. A full analysis of a system requires measures of central tendency and dispersion for both distribution and species richness.

Running title: Range-diversity plots

Keywords

Distribution, diversity, Mexico, mammals, presence-absence matrix, range-diversity plots, variance-ratio test

INTRODUCTION

Research on macroecology focuses on the analysis of spatial patterns and processes at the regional, continental, and global scales. Patterns of interest are based on variables showing geographic variation, from intraspecific and interspecific traits to attributes of whole assemblages (Gaston *et al.*, 2008). Most of these patterns can be summarized in species x sites matrices, in which rows represent taxa, columns correspond to localities, and each element shows some attribute of a particular species at a given site (Bell, 2003; Gaston *et al.*, 2008). The most basic form of such matrices is the presence-absence matrix (PAM), in which elements acquire binary values that represent the presence (1) or absence (0) of a particular species in a given site (Gotelli, 2000; Arita *et al.*, 2008). Matrices can be analyzed by columns (*Q*-mode) or by rows (*R*-mode), yielding different kinds of information from the same data (Williams & Lambert, 1961; Sneath & Sokal, 1973; Simberloff & Connor, 1979; Legendre & Legendre, 1983). In large-scale studies, an analysis of PAMs by rows produces information on the range size of species, whilst an equivalent analysis by columns yields data on the species richness of sites.

Additional information can be extracted from PAMs by using *Rq*- and *Qr*-mode analyses (Arita *et al.*, 2008). In *Qr*-mode, data are computed by columns (by sites), but considering the structure of the rows that intersect a given column with a non-zero entry (that is, species occurring in the focal site). This procedure generates the “dispersion field” of a site, which is the set of ranges of all species that occur in that locality (Graves & Rahbek, 2005; Arita *et al.*, 2008). A comparison of dispersion fields for several sites allows in turn the analysis of the geographic variation of range sizes (Lutz, 1921; Anderson & Koopman, 1981; Rapoport, 1982; Hawkins & Diniz-Filho, 2006; Orme *et al.*, 2006). Equivalently, *Rq*-mode analyses are performed by rows (by species), but incorporating information of the columns that intersect the focal row with a non-zero entry. The resulting set of species richness values of the sites that form the range of a species is the “diversity field” of that species (Arita *et al.*, 2008; Villalobos & Arita, 2010).

The properties of dispersion and diversity fields can be envisioned using range-diversity plots (RD plots, Fig. 1), in which information on range size and species richness is depicted simultaneously (Arita *et al.*, 2008; Borregaard & Rahbek, 2010). RD plots can be built by species or by sites, and a complete understanding of a system consisting of several species occurring in a number of sites would normally require the use of both types of plots. The location of points in RD plots by species depends on the covariation among species, which is ultimately defined by the patterns of co-occurrence. Because variance in species richness can be partitioned into components determined by the distribution of species (Schluter, 1984; Bell, 2003; Legendre *et al.*, 2005), RD plots can be used as a visual tool for examining such decomposition, which can be tested quantitatively with a variance ratio test (Schluter, 1984).

Rq and *Qr* procedures, by combining information on species richness and distribution, allow analyses that go beyond the standard studies that consider each variable separately. Thus, RD plots can be useful tools in studies that require the simultaneous consideration of patterns of diversity and distribution, for example when examining patterns of beta diversity or nestedness (Arita *et al*., 2008; Christen & Soberon, 2009). RD plots and associated parameters can also be useful in the validation of dynamic models of continental diversity (Gotelli *et al*., 2009; Borregard & Rahbek, 2010; Villalobos & Arita, 2010) and in the identification of priority areas and species for conservation initiatives.

In this paper, we discuss the use of RD plots in extracting and interpreting information from PAMs. In particular, we examine the role of covariance in determining the position of points in the graphs, and explore the use of variance ratios in detecting association among species, as proposed by Schluter (1984), or among sites, as proposed here. We use matrix algebra to derive the mathematical relationships between diversity and distribution, and show how this procedure enables fast and efficient computer algorithms. We also present empirical examples and null models to illustrate the analytical power of RD plots.

INNOVATION

In this section we present a step-by-step guide to build and interpret RD plots by extracting information from a PAM. We employ the mathematical relationships between diversity and distribution that have been demonstrated by Arita *et al*. (2008) and present an alternative notation based on matrix algebra (Christen & Soberón 2009). We use a worked example to show simple ways in which parameters can be readily calculated and provide a fully functional R script (R Development Core Team, 2008) that should enable readers to manage large datasets (Appendix S1 in Supporting Information). Most potential users of RD plots not interested in the mathematical details might find all the information that they require by following the step-by-step introductory example and by analyzing their own data with the R script as it is. Other users might want to go through the mathematical derivations to be able to modify the R script to suit their particular datasets or analytical needs.

The Presence-absence matrix

The basic source of information for building RD plots is an $S \times N$ presence-absence matrix $\mathbf{\Delta}$ that summarizes the distribution of S species among N sites (we follow the convention of denoting matrices and vectors with bold characters). Each row represents a species, each column represents a site, and the elements of the matrix are $\delta(i,j)=1$ if species i is present in site j, and $\delta(i,j)=0$ otherwise. The sum of elements along a row yields the number of sites in which the corresponding species occurs (that is, its range size n_i), and the equivalent sum along a column equals the total number of species present in a site (that is, its species richness s_j). The vectors containing the S range size values and the N species richness values can be easily calculated as $\mathbf{n}=\mathbf{\Delta 1}_N$ and $\mathbf{s}=\mathbf{\Delta}^T\mathbf{1}_S$, where $\mathbf{1}_N$ and $\mathbf{1}_S$ are vectors of ones of length N and S, respectively, and the superscript T indicates transpose.

Table 1 is a PAM showing the distribution of S = 10 species of mammals in N = 18 islands of the Thousand Islands region of New York (Lomolino, 1986): The right-hand column in bold marked as n_i is $\mathbf{n}$, the vector of range size values (n_i, which in this example is the number of islands in which each species occurs). The first row in bold is the transpose of $\mathbf{s}$, the vector of species richness values for each island, s_j). The averages of these vectors are $\bar{n}$ = 5.0 sites and $\bar{s}$ = 2.78 species, respectively.

The fill of the matrix is the total of occurrences (ones) in it, $f = \mathbf{1}_N^T \mathbf{\Delta}^T \mathbf{1}_S$ = 50 in this example. The fill can be easily computed as the sum of all species richness values $f = \sum_{j=1}^{N} s_j$ or of all range size values $f = \sum_{i=1}^{S} n_i$, that is, $f = \mathbf{1}_N^T \mathbf{s} = \mathbf{1}_S^T \mathbf{n}$. If the $\bar{n}$ and $\bar{s}$ values are divided by the total number of sites or the total number of species, respectively, we obtain their proportional values $\bar{n}^* = \bar{n} / N$ and $\bar{s}^* = \bar{s} / S$. It is easy to show that $\bar{n}^* = \bar{s}^* = f^* = f / SN$, that is, the proportional fill of a matrix is always equal to the proportional mean richness or the proportional mean range size in the system. In the Thousand Islands example, $f^* = \bar{n}^* = \bar{s}^*$ = 0.278, meaning that on average each island contains 27.8 % of the species and that the average species occurs in 27.8% of the islands. Whittaker's index of beta diversity equals the inverse of the proportional fill, $\beta = (f^*)^{-1}$ = 3.6 in the present example, so it can be envisioned either as the factor relating the total species richness with average local richness, $\beta = S / \bar{s}$ (Whittaker, 1960) or as the ratio of the total area of the region and the average range-extent of species, $\beta = N / \bar{n}$ (Routledge, 1977; Arita *et al.*, 2008).

Rq **and** ***Qr*** **analyses**

The diversity field volume (D_i) of species i is the summation of species richness values of sites within its range. In the example, the diversity field volume of species 1 (first row) is the sum of the richness values of sites (columns) 1, 2, 4, 5, 6, 7 and 8, that is $D_1 =$ 34 species. The dispersion field volume (R_j) of site j is the summation of range sizes of the species occurring in that site (Graves & Rahbek, 2005). The dispersion field volume of the 18th site (last column) in the example is the sum of the range sizes of species 4 and 5, that is R_{18} = 7 sites. The vectors of the S diversity field volumes and the N dispersion field volumes can be computed as $\mathbf{D} = \mathbf{\Delta s} = \mathbf{\Delta \Delta}^T \mathbf{1}_S$ and $\mathbf{R} = \mathbf{\Delta}^T \mathbf{n} = \mathbf{\Delta}^T \mathbf{\Delta 1}_N$, respectively.

Dividing the diversity field volumes by the corresponding range size of each species, we obtain the vector of average species richness values within each range. Equivalently, dividing the dispersion field volumes by the corresponding species richness of each site, we can compute a vector of the mean range sizes of species occurring in the site. We call these parameters the mean range richness of a species ($\overline{s}_i = D_{i/} / n_i$) and the mean per-site range size of a locality ($\overline{n}_j = R_j / s_j$). Notice that the first one is a richness value that can be assigned to a species, and the second one is a range size variable assigned to a site. Dividing these variables by the total number of species or by the total number of sites, we obtain the proportional range richness of a species ($\overline{s}_i * = \overline{s}_i / S$) and the proportional per-site range size of a site ($\overline{n}_j * = \overline{n}_j / N$).

RD plots

In RD plots by species, abscissas are the proportional range richness values ($\overline{s}_i *$) and ordinates are the proportional range sizes of species ($n_i *$, Fig. 1A). In RD plots by sites, abscissas represent the proportional per site range size ($\overline{n}_j *$) and the proportional species richness values correspond to the ordinates ($s_j *$, Fig. 1B). In both cases, a vertical line is drawn to coincide, along the x-axis, with the proportional fill of the PAM, which equals both the average proportional range size and the average proportional species richness of the system; in this example, $f* = \overline{n}* = \overline{s}*$ = 0.278.

In Fig. 1A and B, the dark curved lines represent mathematical constraints that mark a limit to the possible values of points in RD plots. Their shape and position depend on the minimum and maximum richness and range size values (Arita *et al*., 2008). These limits can be explained verbally using the law of the large numbers. In the plot by species the larger the “sample size” (the number of sites forming a range), the closer the range richness value has to be to the overall mean. In the limit, the average richness of the sites forming the range of a species occurring everywhere is identical to the overall mean richness, so a point lying on the very top of the RD plot will necessarily be located on the vertical dashed line. By contrast, points corresponding to species occurring in a few sites (representing “small samples”) can vary widely along the abscissa, as shown by the larger dispersion of points in the bottom part of the RD plots in Fig. 1A. With a similar reasoning, the point corresponding to a site containing all species will necessarily be located on the top of the plot and on the vertical dashed line, because the average range size of species occurring there is identical to the overall mean. Sites with low diversity values, in contrast, will show more variation in average range size.

In Fig. 1A the point close to the top corresponds to species 2, which occurs in 16 of the 18 islands (n_i* = 0.89) and co-occurs, on average, with 2.44 species in each island ($\overline{s}_i*$ = 0.244). A species occurring in all islands would necessarily co-occur with an average of 2.78 species and its point would be at the top of the graph, exactly on the vertical dashed line. In contrast, species occurring in only one island could in principle have proportional range richness values ($\overline{s}_i*$) from 1/10 = 0.1 (the focal species being the only one in the island) to 10/10 = 1.0 (the focal species sharing the island with all other species). In the example, points corresponding to species occurring in a few islands are all located on the right hand side of the plot, indicating a tendency of restricted species to occur only in high-richness sites. In the plot by sites (Fig. 1B, island 1 harbours the ten species (s_j* = 1.0), so its point lies on top of the plot and exactly on the vertical line, indicating that species occurring there have an average range of 5.0 islands ($\overline{n}_j*$ = 0.278). Almost all islands with low or intermediate species richness harbour species occurring in many sites, so their points lie on the right-hand part of the plot. The only exception is island 18, which contains two species, occurring in only 3 and 4 islands. The point corresponding to this island is the one on the lower left part of the plot.

The position of points in RD plots in relation to the vertical line is also related to the average covariance of species or sites. In general, $\rho_i = n_i*\left(\overline{s}_i*-\overline{s}*\right)$ is the average covariance of species i with all species, and $\tau_j = s_j*\left(\overline{n}_j*-\overline{n}*\right)$ is the average covariance of site j with all sites (Arita *et al*., 2008). Hence, the covariance of a species depends on the number of species with which it shares its distribution, and the covariance of a site is determined by the number of sites with which it shares species (Arita *et al.*, 2008). Points located along the hyperbolic dashed curves in Fig. 1 have the same covariance (the ± 0.01, 0.05, and 0.1 isocovariance lines are shown, negative covariances to the left, positive covariances to the right). In the plot by species, the farther a point is from the dashed line, the higher is the absolute value of the covariance of the corresponding species with the complete biota. These lines are drawn using the equations $n_i * = \rho / \left(\overline{s}_i * - \overline{s} * \right)$ and $s_j * = \tau / \left(\overline{n}_j * - \overline{n} * \right)$ where ρ and τ are particular covariance values for species and for sites, respectively.

Histograms on top of RD plots in Fig. 1 show that the points for most species and most islands lie to the right of the vertical dashed line, that is, their average covariance is >0. In fact, nine of the ten species and 15 of the 18 sites have average covariances between +0.05 and +0.1. This pattern, in which points of most species and most sites fall on the right-side sector of RD plots is characteristic of highly nested assemblages, in which if a species occurs in only a few sites, these sites tend to be areas of high species richness. Equivalently, low-richness sites are populated by species that are widespread.

Variance partitioning and variance-ratio tests

The $N \text{ x} N$ matrix of variance-covariance among sites is computed as $\mathbf{C}_{si} = \frac{1}{S}\left(\mathbf{\Delta}^T \mathbf{\Delta} - \frac{1}{S} \mathbf{s}\mathbf{s}^T \right) = \left[c_{si}(j,m) \right]$ for j and m = 1 to N, where $c_{si}(j,m)$ is the covariance between sites j and m. The equivalent $S \text{ x} S$ variance-covariance matrix for species is $\mathbf{C}_{sp} = \frac{1}{N}\left(\mathbf{\Delta}\mathbf{\Delta}^T - \frac{1}{N} \mathbf{n}\mathbf{n}^T \right) = \left[c_{sp}(i,l) \right]$ for i and l = 1 to S, where $c_{sp}(i,l)$ is the covariance between species i and l. The elements along the diagonals are the binary variances $v_{si}(j) = s_j * (1 - s_j *)$ for site j and $v_{sp}(i) = n_i * (1 - n_i *)$ for species i. Notice that $\mathbf{\Delta}\mathbf{\Delta}^T$ and $\mathbf{\Delta}^T \mathbf{\Delta}$ are the $S \text{ x} S$ matrix of co-occurrence of species and the $N \text{ x} N$ matrix of the number of species shared by sites, respectively. The diagonal of the first matrix is equal to the vector $\mathbf{n}$ of range sizes, the diagonal of the second matrix is equal to the vector $\mathbf{s}$ of species richness values, and the trace of either of these matrices equals f, the fill of the matrix. The average covariance of any given site j with all sites is τ_j, and the N x1 vector of such values for all sites is given by $\boldsymbol{\tau} = \frac{1}{N} \mathbf{C}_{si} \mathbf{1}_N$. By symmetry, the average covariance of species i with all species is ρ_i, and the S x1 vector of such values is $\boldsymbol{\rho} = \frac{1}{S}\mathbf{C}_{sp}\mathbf{1}_S$.

In any PAM the variance in species richness of sites (s_j) equals the sum of the variances in range size for all species plus twice the sum of covariances among species (Schluter, 1984; Bell, 2005):

$$Var(s) = \sum_{i=1}^{S} v_{sp}(i) + \sum_{i=1}^{S}\sum_{l=1,l\neq i}^{S} c_{sp}(i,l) \tag{1}$$

Notice that $\sum_{i=1}^{S} v_{sp}(i)$ is the trace of the matrix $\mathbf{C}_{sp}$, that is, the summation of the variances of species, and that $\sum_{i=1}^{S}\sum_{l=1,l\neq i}^{S} c_{sp}(i,l)$ is the summation of the non-diagonal elements of the matrix $\mathbf{C}_{sp}$, that is, twice the summation of all pair-wise covariances. In other words, the right-hand part of equation (1) is simply the summation of all elements of $\mathbf{C}_{sp}$, that is:

$$Var(s) = \mathbf{1}_S^T \mathbf{C}_{sp} \mathbf{1}_S \tag{2}$$

This leads to the important result that the variance in species richness among sites depends on the variance and covariance of distributional values for species. This property can be used for testing the hypothesis of a possible association of species in terms of co-occurrence patterns (Robson, 1972; Schluter, 1984; Bell, 2005). From equation (1), if the sum of covariances of all species is zero (meaning that on average there is no association among them), then the ratio $V_{sp} = Var(s) / \sum_{i=1}^{S} v_{sp}(i)$ must be equal to 1. An observed V_{sp} that is < an expected value generated by some null model would indicate a negative total covariance, which might point to a possible mechanism of avoidance or exclusion between species at local scales (Gotelli, 2000; Bell, 2005), or at spatial segregation due to contrasting climatic or habitat preferences at biogeographical scales.

Following the same logic, the variance in range size among species is determined by the variance and covariance of sites in terms of species richness:

$$Var(n) = \sum_{j=1}^{N} v_{si}(j) + \sum_{j=1}^{N} \sum_{m=1, m \neq j}^{N} c_{si}(j,m) \qquad (3)$$

$$Var(n) = \mathbf{1}_N^T \mathbf{C}_{si} \mathbf{1}_N\,. \qquad (4)$$

In equation (3) the first term on the right side is the sum of variances and the second term is twice the sum of covariances among sites, so $Var(n)$ equals the summation of all elements of the matrix $\mathbf{C}_{si}$, as shown by equation (4). A variance ratio test equivalent to the one proposed by Schluter (1984) can be used for sites to test for significant similitude in terms of shared species, $V_{si} = Var(n) / \sum_{j=1}^{N} v_{si}(j)$. V_{si} can be used for testing a possible clustering of sites in terms of shared species. In principle, V_{sp} and V_{si} are related, through the relationships between variance among sites and among species, but not totally dependent on each other. A full description of a system, including patterns by species and by sites, could be achieved through the use of both parameters.

Table 2 shows the variance-covariance matrix by species ($\mathbf{C}_{sp}$) of the Thousand Islands example. The diagonal of the matrix includes the binary variances generated by the range size of each species, so the sum of these S = 10 elements is $\sum_{i=1}^{S} v_{sp}(i) = \sum_{i=1}^{S} n_i{*}\left(1 - n_i{*}\right)$ = 1.518. The sum of the $S(S-1)$ = 90 non-diagonal elements equals twice the sum of all pair-wise covariances, $\sum_{j=1}^{N} \sum_{m=1, m \neq j}^{N} c_{si}(j,m)$ = 5.654. From Table 1, we can calculate the population variance in richness $Var(s)$ =

$\frac{1}{N}\sum_{i=1}^{S}\left(n_i - \overline{n}\right)^2$ = 7.173. The partitioning of variance defined by Equation (1) is readily corroborated: 7.173 = 1.518 + 5.654. Equivalently, the variance in range-size values can be partitioned into two components from the variance-covariance matrix by sites (Equation 3): $\mathrm{var}(n) = \frac{1}{S}\sum_{i=1}^{S}\left(n_i - \overline{n}\right)^2$ = 15.80 = 2.32 + 13.48. Schluter's (1984) variance-ratio parameter is V_{sp} = 7.173/1.518 = 4.72, and the equivalent ratio for sites is V_{si} = 15.80/2.32 = 6.81.

CONTINENTAL EXAMPLES: MAMMALS IN THREE MEXICAN REGIONS

In this section, we present data on the distribution and richness patterns of the mammals of three contrasting regions of Mexico to illustrate the analytical power of RD plots (Fig. 2), and present the results of three different null models to show how RD plots and variance-ratio tests can be used in combination to dissect the variance components of the distribution and diversity parameters (Figs. 3 and 4).

Each region consists of a set of ½-degree quadrats in which the distribution of mammals was used to build the corresponding presence-absence matrices. The first region was located in Central Mexico, a very heterogeneous area located in the transitional zone between the Nearctic and Neotropical biogeographic realms; the second region included parts of the Isthmus of Tehuantepec in Southeastern Mexico, also a highly heterogeneous area lying on the transitional zone but with a stronger component of Neotropical influence; the third square included the Yucatan Peninsula, a relatively homogeneous area of full Neotropical composition. The Central Mexico region included 62 ½-degree quadrats, while the other two regions included 50 quadrats each. Distributional data were extracted from the database described in Arita *et al.* (1997), and more details on the scaling of diversity patterns in these three regions can be found in a previous publication (Arita & Rodriguez, 2002).

In Fig. 2, the three regions are shown in order of their f^* value (or, equivalently, in order of decreasing beta diversity), from Central Mexico (Fig. 2A and B) to Yucatan (Fig. 2E and F). The fill of the PAM equals the summation of all range size values or the summation of all species richness values; consequently, its magnitude is closely tied to the range-size and species-richness frequency distributions, which are shown in the right-hand panels of RD plots in Fig. 2. Histograms for the Yucatan region, for example, show a predominance of widespread species and species-rich sites, a fact that is reflected on the high f^* value (Fig. 2E and F). The Central Mexico region shows a more even distribution of range-size values and lower values of species richness for its sites, all of which reflects in a lower f^* (Fig. 2A and B). The Isthmus region is intermediate between the Yucatan and the Central Mexico cases, with a range-size frequency distribution skewed to small ranges, but not as extreme as for the Yucatan region (Fig. 2C and D).

The position of points relative to the dashed vertical line depends on the degree of association among species or the degree of similitude among sites. In the RD plots by sites for the three regions, points are located to the right of the vertical line, with several points going farther than the +0.1 isocovariance line (Fig 2B, D and F), which indicates that all sites show a positive average covariance with other sites. This is also shown by the high variance ratio by sites (V_{si} > 24 in the three cases, Table 1). In the plots by species, points tend to lie to the right of the vertical line, but the tendency is much stronger in the Isthmus region (Fig. 2C) than in the Central Mexico or Yucatan regions (Fig. 2A and E). Notice that in the case of the Yucatan (Fig. 2E), there are several points lying at the very top of the plot, coinciding with the vertical line. This pattern is corroborated by the variance ratio values by species (V_{sp}), which are >1.0 in the three cases, but higher for the Isthmus region (Table 1).

The Central Mexico mammal fauna (Fig. 2A and B) is a combination of widespread and restricted taxa that generates a pattern of low average local richness but high regional richness (*i.e.*, a high β diversity). Covariance among species (association) is positive but low and covariance among sites (similitude) is also low. Several mammalian species in the Isthmus region are widespread, but the region also harbours many species with restricted distributions. This pattern generates sites with local species richness values that are higher than those for Central Mexico but whose aggregate richness is lower, indicating a lower β diversity (Fig. 2C and D). Finally, the Yucatan region consists of sites containing mostly occurring all over the peninsula, generating a pattern of high local species richness, but a very low β diversity (Fig. 2E and F).

Null models and the effect of range cohesion

Null models have the purpose of contrasting real-world assemblages against hypothetical patterns generated by randomizing some variables while retaining the empirical values for other parameters (Gotelli & McGill, 2006). We used three null models that have been shown to generate contrasting patterns when examined with RD plots (Borregaard & Rahbek, 2010; Villalobos & Arita, 2010).

In our first null model, we maintained the empirical column sums, that is, we retained the original species-richness frequency distribution, but assigned sites to species at random with no replacement. We did this by permutating the zeroes and ones in each column, so we conserved the empirical fill of the matrix and, consequently, the original Whittaker's beta diversity. The range size values, in contrast, changed with this procedure, and so did the variance-covariance matrices both for species and for sites.

Figs. 3A and B show the results of one run of this null model applied to the Central Mexico region. Notice that the position of the vertical line is identical to that in Figs. 2A and B (corresponding to f^* = 0.44), and that the species-richness frequency distribution is the same in both cases (right-hand panel in Fig. 2B and Fig. 3B). In contrast, the range size frequency distribution (Fig. 3A, right panel), the frequency distributions of the range richness and per site range parameters (Fig. 3A and B, top panels), and the position of points are all changed. The randomization process generated a system in which the covariance among sites was zero, a pattern shown in the RD plot by sites in the arrangement of points along the vertical dashed line (Fig. 3B) and by the value of the variance-ratio parameter by sites, whose average across 1,000 iterations of the model was practically equal to 1.0 (V_{si} = 0.998, with variance = 0.009), contrasting with the empirical value (V_{si} = 24.08, Table 1). Species also showed a lower variance ratio in the simulations than in the real world system (V_{sp} = 4.69, mean for 1,000 iterations, variance = 5.13 x 10^{-5}; V_{sp} = 7.44, empirical value, Table 1). This means that in the simulations species had a tendency to overlap less than in the real world, but never attaining a total independence.

Our second null model was mathematically identical to the first one, but inverting the role of sites and species. We retained the empirical range size frequency distribution (row totals) and generated permutations of rows of the PAM to simulate the random assignment, without replacement, of species to sites. This model also retains the empirical f^*, so the position of the vertical line is not changed (Fig. 3C and D, corresponding to the Central Mexico region). In the RD by species (Fig. 3C), the range-size frequency distribution (right panel) is unchanged, but points arrange along the vertical line, showing that the average covariance among species is close to zero, as a consequence of the randomization procedure. This pattern is also shown by the variance-ratio being practically equal to one (V_{sp} = 0.994 for 1,000 iterations, Fig. 4A left hand histogram).

Sites showed less variation in species richness than in the real world system (histograms in the right panel of Fig. 2B and Fig. 3D) but had a strong positive covariance (similitude in species composition), as shown by the points in Fig. 3D being concentrated on the right side of the plot and by the value of the variance ratio for sites (V_{si} = 23.62, mean for 1,000 iterations, Fig. 4B left hand histogram). However, these V_{si} values are less than 24.08, the empirical value for the system, meaning that species in the simulations show less overlap in their distributions than in the empirical system (Fig. 4B).

In the third null model we retained the empirical range size frequency distribution but simulated ranges as random cohesive units by using the spreading-dye algorithm as described by Jetz & Rahbek (2002). For each species, we started with a randomly located site; then, we filled the available adjacent cells until the count of sites equalled the empirical range size of the species. The model generated higher covariance values among species than in the real world systems. This is shown by the significantly higher variance ratio in the simulations (V_{sp} = 14.02, mean of 1,000 iterations) than in the real world system (V_{sp} = 7.44, Fig. 4A, right hand histogram). This tendency can be also seen in the RD plot by species (Fig. 3E). Sites also showed a higher covariance in the simulations than in the empirical dataset. In the simulations, the variance ratio was significantly higher (V_{si} = 24.56, average for 1,000 iterations of the model, V_{si} = 24.08, empirical data) and points aggregated to the right in the RD plot by sites (Fig. 3F). These patterns show that real-world species tend to co-occur less frequently than expected if ranges are modelled as cohesive units, but more frequently than expected from scattered-ranges models (Arita & Rodríguez-Tapia, 2009).

Measures of central tendency and dispersion

When quantifying species richness of sites and range size of species through a PAM, mathematical relationships determine limits to the possible values that diversity and distribution components can attain. Our theoretical developments and null models, however, show that the species richness frequency distribution cannot be fully predicted if only the range-size frequency distribution is known. The same is true the other way around; the species richness frequency distribution sets limits to but do not fully determine the range size frequency distribution.

The proportional fill of a PAM ($f^* = \overline{n}^* = \overline{s}^*$), or equivalently, Whittaker's beta $\beta = (f^*)^{-1}$, determines the central tendency of points in RD plots when the mean covariance is zero, but not their dispersion. A way to envision this is to imagine a system in which the general parameters of the system $(\bar{n}^*,\ \bar{s}^*,\ f^*,\ \beta)$ are completely determined. Imagine now that we can move, distort, and even fragment the ranges of species with the only restriction that we retain their size. No matter how extreme our actions are, the values of the parameters mentioned above do not change. A direct consequence of this thought experiment is that Whittaker's index, despite being the most commonly used measure of beta diversity (Koleff *et al.*, 2003; Tuomisto, 2010a, b; Anderson *et al*., in press), is insensitive to transformations of the PAM that leave its dimension and fill constant (Arita & Rodríguez 2002). In contrast, the manipulation of ranges implies changes in the parameters of variation around the mean, for example the variance-covariance matrices, the shape of the species richness frequency distribution, the horizontal location of points in RD plots, and the Schluter's variance-ratio parameters.

CONCLUSION

Species richness and range size are two sides of the same coin, that is, they are equally valid parameters to measure biological diversity. A complete comprehension of the assemblage will require an analysis of parameters of central tendency and dispersion for both species richness and range size. Range-diversity plots and their associated parameters can be a powerful instrument in such endeavor.

ACKNOWLEDGEMENTS

We thank R. K. Colwell, N. J. Gotelli, T. Rangel, P. Trejo, and F. Villalobos for helpful discussion. G. Rodríguez-Tapia provided efficient technical support and crucial help in computer-based analyses. Financial support was provided by DGAPA-UNAM, PAPIIT program and by Microsoft Research KUCR # 47780 for J. Soberón.

REFERENCES

Anderson, M. J., Crist, T. O., Chase, J . M., Vellend, M., Inouye, B. D., Freestone, A. L., et al. (in press) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters*.

Anderson, S. & Koopman, K. F. (1981) Does interspecific competition limit the sizes of ranges of species? *American Museum Novitates. New York NY*, **2716**, 1-10.

Arita, H. T., Christen, A., Rodríguez, P. & Soberón, J. (2008) Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications. *American Naturalist*, **112**, 519-532.

Arita, H. T., Figueroa, F., Frisch, A., Rodriguez, P. & Santos del Prado, K. (1997) Geographical range size and the conservation of Mexican mammals. *Conservation Biology*, **11**, 92-100.

Arita, H. T. & Rodriguez, P. (2002) Geographic range, turnover rate and the scaling of species diversity. *Ecography*, **25**, 541-550.

Arita, H. T. & Rodríguez-Tapia, G. (2009) Contribution of restricted and widespread species to diversity: the effect of range cohesion. *Ecography* **32**,210-214.

Bell, G. (2003) The interpretation of biological surveys. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 2531-2542.

Bell, G. (2005) The co-distribution of species in relation to the neutral theory of community ecology. *Ecology*, **86**, 1757-1770.

Borregaard, M. K. & Rahbek, C. (2010) Dispersion fields, diversity fields and null models: uniting range sizes and species richness. *Ecography*, **33**, 402-407.

Christen, A. & Soberon, J. (2009) Anidamiento y los análisis Rq y Qr en PAMs. *Miscelánea Matemática*, **49**, 51-61.

Gaston, K. J., Chown, S. L. & Evans, K. L. (2008) Ecogeographical rules: elements of a synthesis. *Journal of Biogeography*, **35**, 483-500.

Gotelli, N. J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606-2621.

Gotelli, N. J. & McGill, B. J. (2006) Null versus neutral models: what's the difference? *Ecography*, **29**, 793-800.

Gotelli, N. J., Anderson, M. J., Arita, H. T., Chao, A., Colwell, R. K., Connolly, S. R., *et al*. (2009) Patterns and causes of species richness: A general simulation model for macroecology. *Ecology Letters*, **12**, 873-886.

Graves, G. R. & Rahbek, C. (2005) Source pool geometry and the assembly of continental avifaunas. *Proceedings of the National Academy of Sciences, USA*, **102**, 7871-7876.

Hawkins, B. A. & Diniz-Filho, J. A. F. (2006) Beyond Rapoport's rule: Evaluating range size patterns of New World birds in a two dimensional framework. *Global Ecology and Biogeography*, **15**, 461-469.

Jetz, W. & Rahbek, C. (2002) Geographic range size and determinants of avian species richness. *Science*, **297**, 1548-1551.

Koleff, P., Gaston, K. J. & Lennon, J. J. (2003) Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, **72**, 367-382.

Legendre, L. & Legendre, P. (1983) *Numerical ecology: Developments in Environmental Modelling, v. 3*, edn. Elsevier Scientific Publixhing Company, Amsterdam.

Legendre, P., Bocard, D. & Peres-Neto, P. R. (2005) Analyzing beta diversity: Partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435-450.

Lomolino, M. V. (1986) Mammalian community structure on islands: The importance of immigration, extinction and interactive effects. *Biological Journal of the Linnean Society*, **28**, 1-21.

Lutz, F. E. (1921) Geographic average, a suggested method for the study of distribution. *American Museum Novitates. New York NY*, **5**, 1-7.

Orme, C. D. L., Davies, R. G., Olson, V. A., Thomas, G. H., Ding, T.-S., Rasmussen, P. C., Ridgely, R. S., Stattersfield, A. J., Bennett, P. M., Owens, I. P. F., Blackburn, T. M. & Gaston , K. J. (2006) Global patterns of geographic range size in birds. *PLoS Biology*, **4**, 1276-1283.

Patterson, B. D. (1987) The principle of nested subsets and its implications for biological conservation. *Conservation Biology*, **1**, 323-334.

R Development Core Team. (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rapoport, E. H. (1982) *Areography: Geographical strategies of species*, edn. Pergamon, New York.

Robson, D. (1972) Appendix: statistical tests of significance. *Journal of Theoretical Biology*, **34**, 350-352.

Routledge, R. D. (1977) On Whittaker's components of diversity. *Ecology*, **58**, 1120-1127.

Schluter, D. (1984) A variance test for detecting species associations, with some example applications. *Ecology*, **63**, 998-1005.

Simberloff, S. & Connor, E. F. (1979) Q-mode and R-mode analyses of biogeogrpahic distributions: null hypotheses based on random colonization. *Contemporary quantitative ecology and related ecometrics* (ed. by G.P. Patil & M.L. Rosenzweig), pp 123-138. International Cooperative Publishing House, Fairland, Maryland, USA.

Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical taxonomy*. Freeman, San Francisco.

Tuomisto, H. (2010a) A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alphja and gamma diversity. *Ecography*, **33**, 2-22.

Tuomisto, H. (2010b) A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography*, **33**, 23-45.

Ulrich, W., Almeida, M. & Gotelli, N. J. (2009) A consumer's guide to nestedness analysis. *Oikos*, **118**, 3-17.

Villalobos, F. & Arita, H. T. (2010) The diversity field of New World leaf-nosed bats (Phyllostomidae). *Global Ecology and Biogeography*, **19**, 200-211.

Whittaker, R. H. (1960) Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs*, **30**, 279-338.

Williams, W. T. & Lambert, J. M. (1961) Multivariate methods in plant ecology: III. Inverse association-analysis. *Journal of Ecology*, **49**, 717-729.

SUPPLEMENTARY MATERIAL

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

Appendix S1 R script and instructions to construct range-diversity plots from a presence-absence matrix.

BIOSKETCHES

Héctor T. Arita is research professor at the National University of Mexico, where he teaches community ecology, ecological statistics, and conservation biology. He is a macroecologist specialized in mathematical models of diversity and distribution.

Andrés Christen is senior researcher at the Center for Mathematical Research (CIMAT) in Guanajuato, Mexico, where he applies statistical and probabilistic models to biological and geological problems.

Pilar Rodríguez is an analyst in bioinformatics at the Mexican National Commission of Biodiversity (CONABIO). Her research on biogeography and macroecology, has focused on patterns of diversity and distribution, scales, and beta diversity.

Jorge Soberón is professor of ecology and. evolutionary biology and senior scientist at the Musem of Natural History and Biodiversity Research Center of the University of Kansas. His interests include the theoretical and applied aspects of niche modelling in biodiversity studies.

Table 1. Presence-absence matrix (PAM) showing the distribution of 10 mammal species among 18 islands in the Thousand Islands Region of New York (data from Lomolino 1986). n_i are the range size (occupancy) values for each species; D_i and $\bar{s}_i$ are the corresponding diversity field and range-diversity values. s_j represents the species richness values for islands, and R_j and $\bar{n}_j$ are the corresponding dispersion field volume and per-site range size values.

	si 1	**si 2**	**si 3**	**si 4**	**si 5**	**si 6**	**si 7**	**si 8**	**si 9**	**si 10**	**si 11**	**si 12**	**si 13**	**si 14**	**si 15**	**si 16**	**si 17**	**si 18**	n_i	D_i	$\bar{s}_i$
sp 1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	**7**	**34**	**4.86**
sp 2	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	**16**	**39**	**2.44**
sp 3	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	**5**	**31**	**6.20**
sp 4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**3**	**21**	**7.00**
sp 5	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**4**	**26**	**6.50**
sp 6	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**	**32**	**6.40**
sp 7	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	**28**	**7-00**
sp 8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	**19**	**9.50**
sp 9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	**19**	**9.50**
sp 10	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	**19**	**9.50**
s_j	**10**	**9**	**5**	**4**	**4**	**2**	**2**	**3**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**			
R_j	**50**	**34**	**34**	**32**	**33**	**23**	**23**	**28**	**16**	**16**	**16**	**16**	**16**	**16**	**16**	**16**	**16**	**7**			
$\bar{n}_j$	**5.00**	**3.78**	**6.80**	**8.00**	**8.25**	**11.50**	**11.50**	**9.33**	**16.00**	**16.0**	**16.0**	**16.0**	**16.0**	**16.0**	**16.0**	**16.0**	**16.0**	**3.5**			

Table 2. Variance-covariance matrix (by species) for the mammals of the Thousand Islands Region, calculated from Table 1. Shaded cell show the diagonal containing the binary variance values for each species. Non-diagonal elements are the covariance values.

	sp 1	**sp 2**	**sp 3**	**sp 4**	**sp 5**	**sp 6**	**sp 7**	**sp 8**	**sp 9**	**sp 10**
sp 1	0.24	-0.01	0.11	0.05	0.02	0.11	0.08	0.07	0.07	0.07
sp 2	-0.01	0.10	-0.02	-0.09	-0.09	-0.02	-0.03	-0.04	-0.04	-0.04
sp 3	0.11	-0.02	0.20	0.06	0.10	0.15	0.10	0.08	0.08	0.08
sp 4	0.05	-0.09	0.06	0.14	0.13	0.06	0.07	0.09	0.09	0.09
sp 5	0.02	-0.09	0.10	0.13	0.17	0.10	0.12	0.09	0.09	0.09
sp 6	0.11	-0.02	0.15	0.06	0.10	0.20	0.16	0.08	0.08	0.08
sp 7	0.08	-0.03	0.10	0.07	0.12	0.16	0.17	0.09	0.09	0.09
sp 8	0.07	-0.04	0.08	0.09	0.09	0.08	0.09	0.10	0.10	0.10
sp 9	0.07	-0.04	0.08	0.09	0.09	0.08	0.09	0.10	0.10	0.10
sp 10	0.07	-0.04	0.08	0.09	0.09	0.08	0.09	0.10	0.10	0.10

Table 3 Parameters of diversity and distribution of the mammals of three Mexican regions.

	REGION		
	Central Mexico	**Isthmus**	**Yucatan**
Parametes of the region			
Quadrats	62	50	50
Species	212	206	111
Fill of PAM	5770 (0.44)	6601 (0.64)	4265 (0.77)
Whittaker's Beta	2.28	1.56	1.30
Parameters of species			
Mean range size	27.22 (0.44)	32.04 (0.64)	38.42 (0.77)
Mean range richness	98.14 (0.46)	137.07 (0.67)	88.18 (0.79)
V_{sp}	7.44	15.88	9.65
Parameters of sites			
Mean species richness	93.06 (0.44)	132.02 (0.64)	85.3 (0.77)
Mean per-site range size	40.86 (0.66)	41.62 (0.83)	45.14 (0.90)
V_{si}	24.08	26.77	28.97

FIGURE LEGENDS

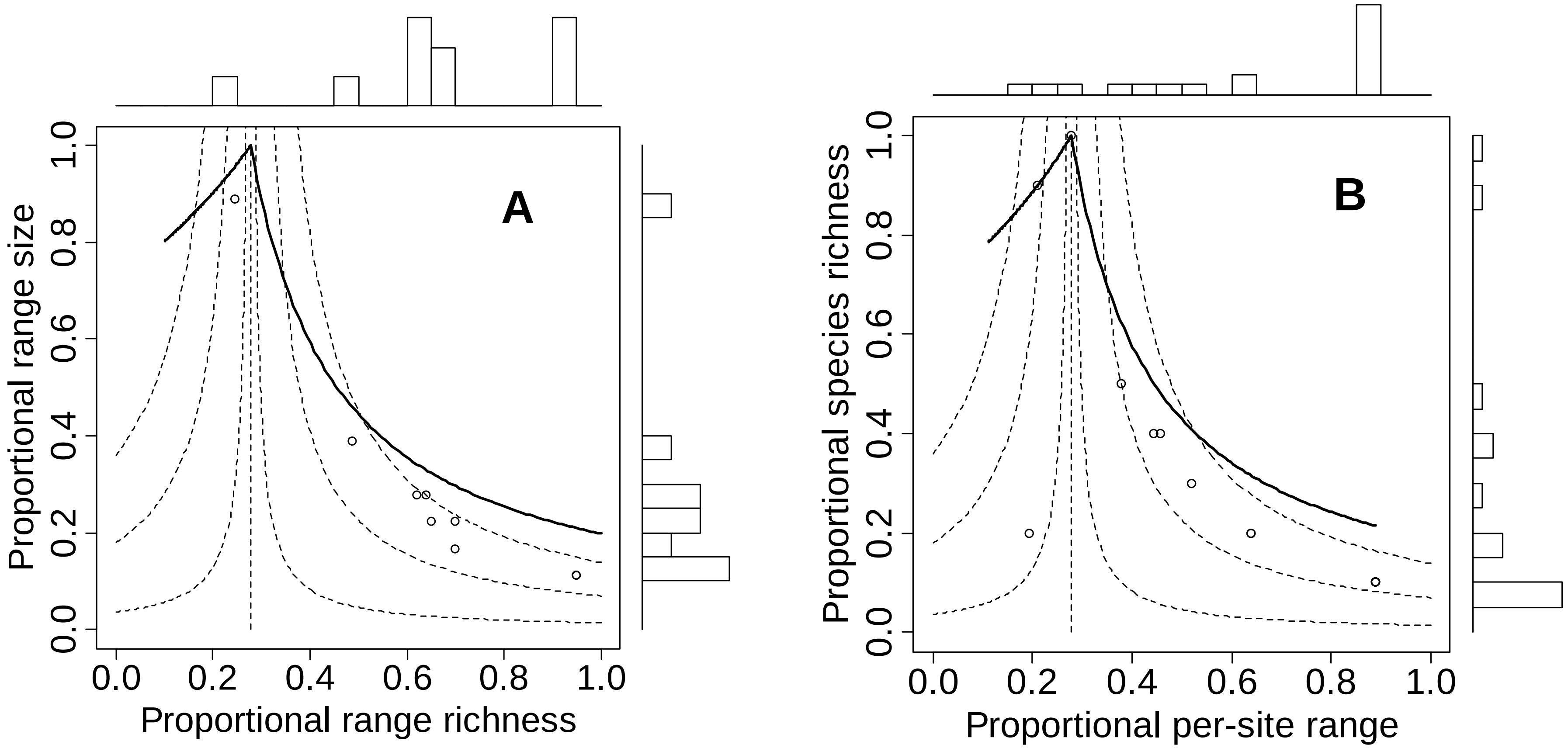
Figure 1 Range-diversity plots for 10 species of mammals in 18 islands of the Thousand Islands region in New York (Lomolino 1986). (A) By species, showing their proportional range sizes (ordinates) and the average species richness within their ranges (abscissas); histograms on top and on the right side show the frequency distribution of those variables; the solid curved line marks the upper theoretical limit for points; the vertical dashed line corresponds with the mean proportional species richness of the 18 sites, and the hyperbolic dashed curves are lines of equal covariance among species. (B) By sites, showing their proportional species richness (ordinates) and the average proportional range size of species occurring in the sites; the other elements of the graph correspond to those described for (A).

Figure 2 Range-diversity plots for the mammal fauna of three regions in Mexico, by species (A, C, and E), and by sites (B, D, and F). (A) and (B) Central Mexico. (C) and (D) the Isthmus of Tehuantepec. (E) and (F) the Yucatan Peninsula. Elements of plots as in Fig. 1.

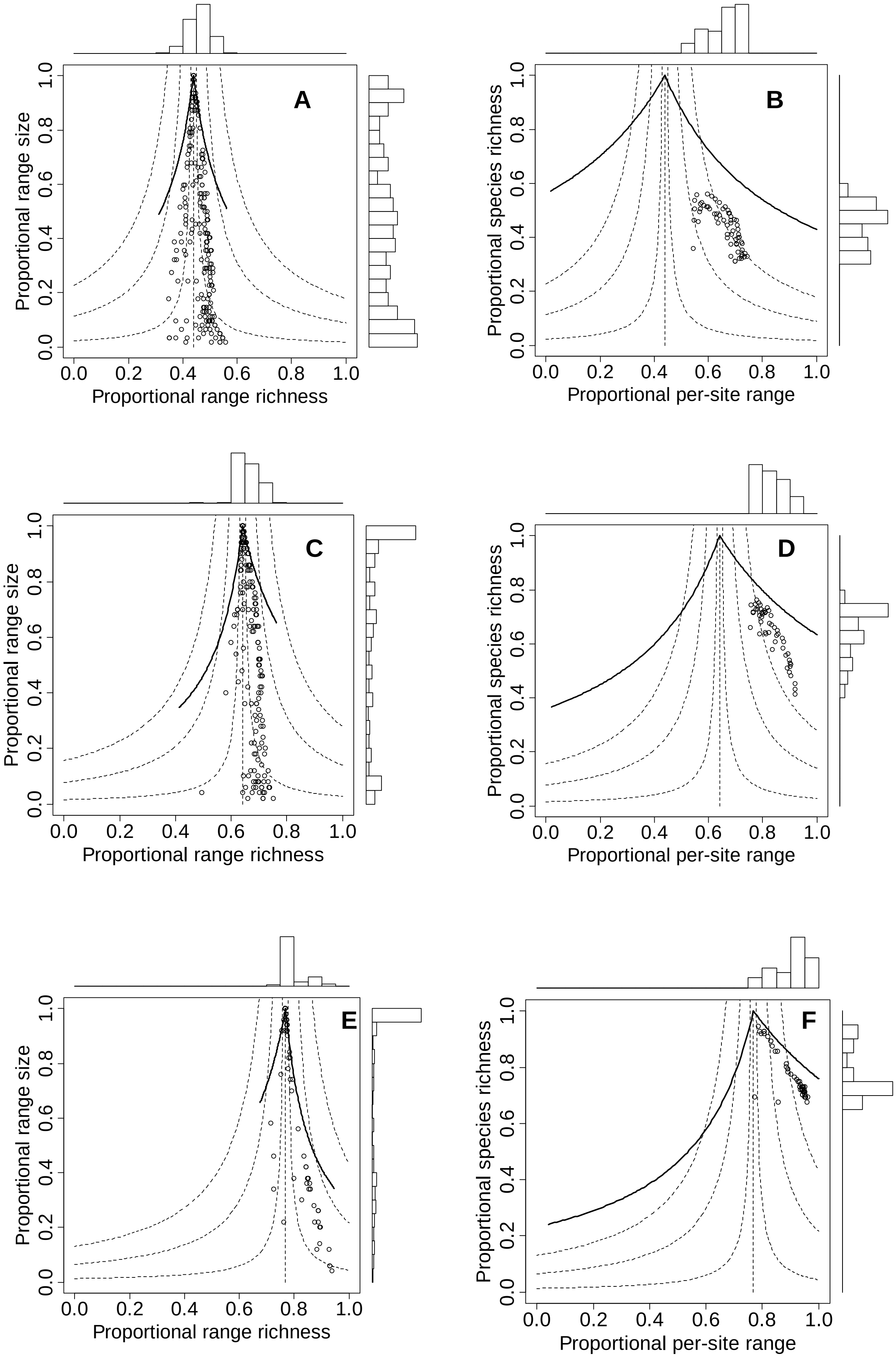
Figure 3 Range-diversity plots for three null models using empirical data for the mammal fauna of Central Mexico, by species (A, C, and E), and by sites (B, D, and F). (A) and (B) Scattered ranges simulation retaining the empirical species richness frequency distribution. (C) and (D) Scattered ranges simulation retaining the empirical range size frequency distribution. (E) and (F) Cohesive ranges simualtion using the spreading dye algorithm, retaining the empirical range size frequency distribution. Elements of plots as in Fig. 1.

Figure 4 Frequency distribution of the values of Schluter's variance ratio parameter by species (A) and by sites (B) corresponding to two null models using data for the mammals of Central Mexico. The left-hand slim histogram in each panel corresponds to the simulations using scattered ranges and retaining the empirical range size frequency distribution; the right-hand histogram in each case corresponds to the simulations using the spreading-dye algorithm to model cohesive ranges. Numbers and arrows show the value and location of the empirical values. Histograms show the results of 1,000 iterations of each simulation.

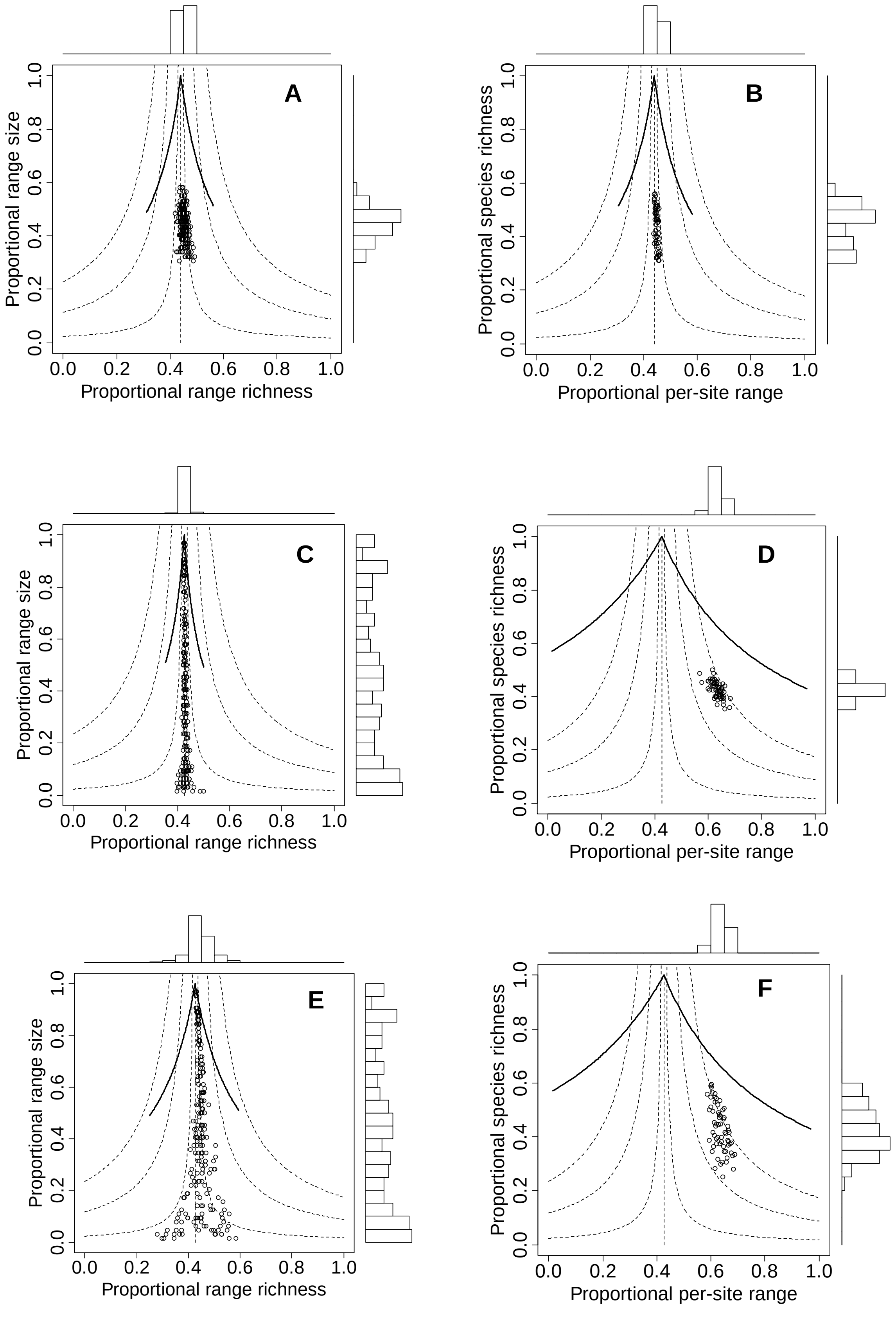
Arita *et al*. Fig. 1



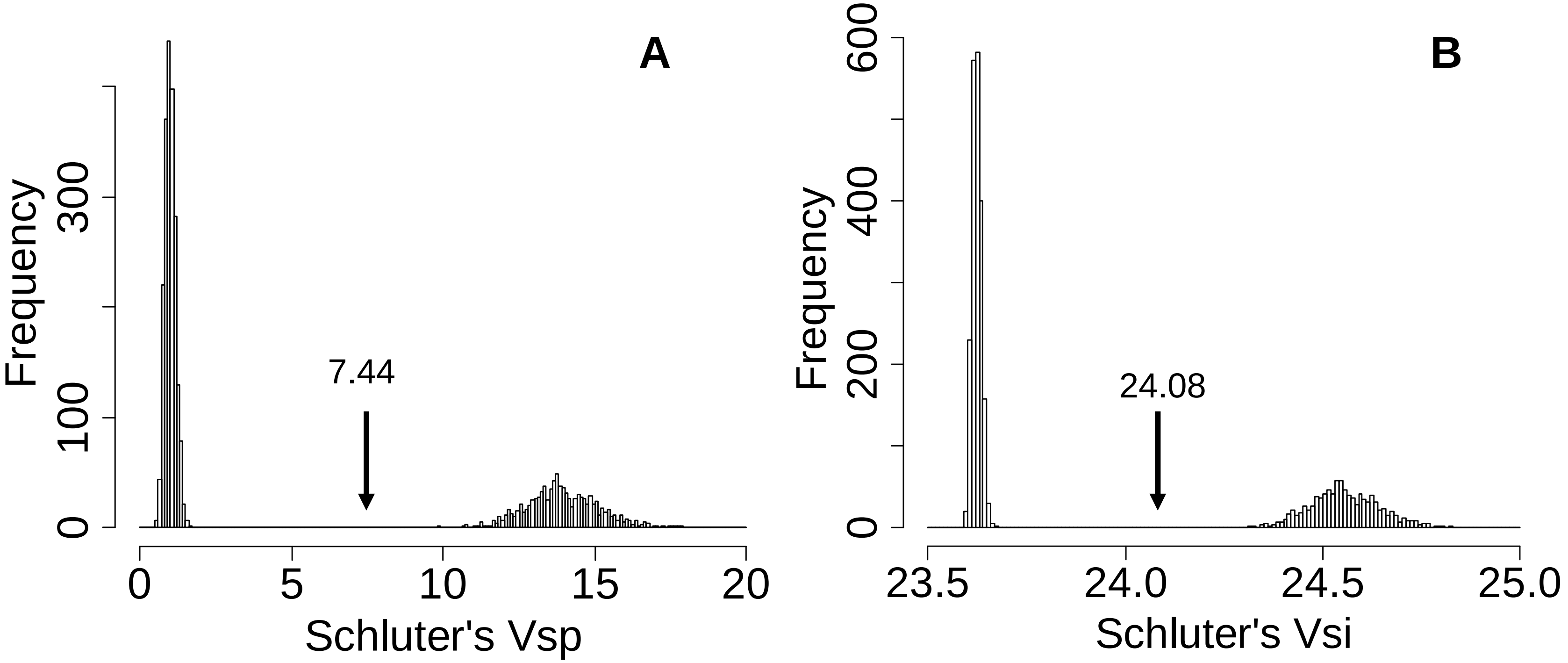
Arita *et al*. Fig. 2



Arita et al. Fig. 3



Arita *et al*. Figure 4


A
Frequency
Schluter's Vsp
B
Frequency
Schluter's Vsi