Oecologia
DOI 10.1007/s00442-009-1474-y

COMMUNITY ECOLOGY - ORIGINAL PAPER

The empirical Bayes approach as a tool to identify non-random species associations

Nicholas J. Gotelli · Werner Ulrich

Received: 17 December 2008 / Accepted: 18 September 2009

Abstract A statistical challenge in community ecology is to identify segregated and aggregated pairs of species from a binary presence–absence matrix, which often contains hundreds or thousands of such potential pairs. A similar challenge is found in genomics and proteomics, where the expression of thousands of genes in microarrays must be statistically analyzed. Here we adapt the empirical Bayes method to identify statistically significant species pairs in a binary presence–absence matrix. We evaluated the performance of a simple confidence interval, a sequential Bonferroni test, and two tests based on the mean and the confidence interval of an empirical Bayes method. Observed patterns were compared to patterns generated from null model randomizations that preserved matrix row and column totals. We evaluated these four methods with random matrices and also with random matrices that had been seeded with an additional segregated or aggregated species pair. The Bayes methods and Bonferroni corrections reduced the frequency of false-positive tests (type I error) in random matrices, but did not always correctly identify the non-random pair in a seeded matrix (type II error). All of the methods were vulnerable to identifying spurious secondary associations in the seeded matrices. When applied to a set of 272 published presence–absence matrices, even the most conservative tests indicated a fourfold increase in the frequency of perfectly segregated "checkerboard" species pairs compared to the null expectation, and a greater predominance of segregated versus aggregated species pairs. The tests did not reveal a large number of significant species pairs in the Vanuatu bird matrix, but in the much smaller Galapagos bird matrix they correctly identified a concentration of segregated species pairs in the genus *Geospiza*. The Bayesian methods provide for increased selectivity in identifying non-random species pairs, but the analyses will be most powerful if investigators can use a priori biological criteria to identify potential sets of interacting species.

Keywords Biogeography · Null model · C score · Presence–absence matrix · Statistical test

Communicated by Wolf Mooij.

Electronic supplementary material The online version of this article (doi:10.1007/s00442-009-1474-y) contains supplementary material, which is available to authorized users.

N. J. Gotelli (✉)
Department of Biology, University of Vermont,
Burlington, VT 05405, USA
e-mail: ngotelli@uvm.edu

W. Ulrich
Department of Animal Ecology, Nicolaus Copernicus University
in Toruń, Gagarina 9, 87-100 Torun, Poland
e-mail: ulrichw@uni.torun.pl

Introduction

A major research focus in community ecology and biogeography has been the identification of non-random species associations in binary presence–absence matrices (Simberloff and Connor 1979; Gotelli and Graves 1996; Sfenthourakis et al. 2006). In these matrices, each row represents a species or taxon, each column represents a site or sample, and the entries indicate the presence (1) or absence (0) of a species in a site (McCoy and Heck 1987). There are patterns in such matrices that can be summarized by a single univariate metric, such as the nestedness of the matrix

Published online: 15 October 2009

(Patterson and Atmar 1986), or the C score (Stone and Roberts 1990), a measure of average pairwise species segregation.

In null model analysis (Gotelli 2001), the observed matrix is randomized or reshuffled, and the metric is recalculated for the null assemblage. A large number of null assemblages (typically 1,000) are created this way (Manly 1991), and the tail probability for the observed matrix is estimated in a classic frequentist test of the *P*(observed data | null distribution).

The community metric, such as the number of species pairs with exclusive distributions [checkerboard distributions of Diamond (1975)] or the C score (Stone and Roberts 1990), is often an aggregate sum or an average of an index that is calculated for each unique pair of species in the matrix. However, there is great interest in identifying particular pairs of species that contribute to the pattern of non-randomness (Burns 2007), or perhaps in detecting a few significant pairs that might be embedded in a matrix of mostly random associations (Sfenthourakis et al. 2006; Veech 2006).

In theory, the same null model methodology that is applied to the entire matrix could be used to estimate the statistical significance of each individual species pair. The problem is that, with n species in the matrix, there are $(n)(n-1)/2$ such possible pairs. Thus, for a matrix with 50 species, there are 1,225 unique species pairs to be tested. Many of these pairs may not be biologically or statistically independent of one another. Exactly the same problem has arisen in the fields of genomics and proteomics. With microarrays, it is now possible to rapidly screen the expression levels of thousands of different genes, at least some of which are not independent of one another (Kammenga et al. 2007). How do researchers decide which of these gene products are "interesting" enough to warrant further analysis? The empirical Bayes approach (Efron 2005) has been a very useful statistical tool for this purpose. In a nutshell, the empirical Bayes approach uses a bootstrapped or randomized distribution to estimate the priors in a Bayesian analysis and control for false discovery rates.

In this paper, we apply the empirical Bayes method to the ecological problem of detecting non-random species pairs in binary presence–absence matrices. Of course, any such method imposes an arbitrary cut-point for recognizing "significant" or "interesting" cases from a large ranked list of species pairs that range from strongly aggregated through random to strongly segregated. However, by comparing the results of different methods applied to artificial matrices that have specified levels of structure or randomness, we can assess their performance and make useful interpretations of the results when they are applied to published data sets.

We begin by describing four possible screening tests for non-randomness, ranging from simple frequentist tests of all possible species pairs, through a sequential Bonferroni correction, to two variants of the empirical Bayes approach. We apply these tests to two benchmark data sets of 100 random matrices each that have been created by stochastic sampling processes. We also apply the tests to random matrices that have been supplemented with the addition of a single pair of aggregated or segregated species. We then analyze a set of 272 published matrices that have been used in previous analyses of community-wide co-occurrence (Gotelli and McCabe 2002) and nestedness (Ulrich and Gotelli 2007a). Finally, we analyze in detail data matrices for land birds of the Vanuatu Islands (Diamond and Marshall 1976) and finches of the Galapagos Islands (Sanderson 2000), both of which have figured prominently in the species co-occurrence literature.

Materials and methods

We consider four methods that can be used to decide whether a particular pair of species is aggregated, segregated, or random in occurrence.

Ninety-five percent confidence limit (confidence limit criterion)

The simplest way to test whether two species are aggregated or segregated in occurrence is to compare a co-occurrence metric with either a theoretical or a simulated random distribution (the null model) that provides the necessary mean and confidences limits (CL) to be used in frequentist inference (CL criterion). In the case of a species $\times$ sites matrix, we would have to screen all species pairs for those with scores above or below their respective 95% CL of the random distribution (Sfenthourakis et al. 2004). The problem with the CL criterion is that a certain number of species pairs are expected to show significant patterns even if the matrix as a whole is random. Using the traditional 95% CL benchmark, a matrix with 30 species contains $30 \times 29/2 = 435$ pairs, of which 22 are expected to fall outside 95% CL just by chance. If there were three truly non-random associations in the matrix, the false detection error rate (FDER) would be $(22)/(22 + 3) = 88\%$. Sanderson (2004) and Sfenthourakis et al. (2006) both advocate eliminating the weakest 5% or 10% of the significant pairs to safeguard against this problem.

A second problem with the CL criterion is that, even if the individual pairs of species are biologically independent, they might not be statistically independent in the null model analysis. For co-occurrence and nestedness analyses, the most commonly used randomization is the "fixed–fixed"

algorithm (Gotelli 2000), in which species occurrences are randomized, but the row sums (=species incidences) and column sums (=species richness per site) of the observed matrix are preserved. In benchmark tests for community metrics of nestedness (Ulrich and Gotelli 2007a) and co-occurrence (Gotelli 2000), this algorithm correctly identifies random matrices and structured matrices with acceptable type I and type II error frequencies. However, the constraint of fixed row and column totals introduces associations between species and sites that might distort the number of species pairs that fall outside the 95% CL (Ulrich and Gotelli 2007a).

Sequential Bonferroni correction (Benjamini and Yekutieli criterion)

The Bonferroni correction is a simple metric to reduce the FDER by dividing the significance level α by the total number of tests r. However in the case of pair-wise species associations, the Bonferroni correction will often result in much too conservative estimates because of the large number of species pairs in a typical matrix. Moreover, the Bonferroni method assumes the tests to be independent (Gotelli and Ellison 2004) which is not generally true for presence–absence matrices. Benjamini and Yekutieli (2001) developed a less conservative sequential FDER correction for dependent tests [Benjamini and Yekutieli (BY) criterion]. This refinement modifies the test-wise H_0 probability benchmark α from the ordered sequence (largest to smallest) of H_0 probabilities P_k to

$$P_k^* = \alpha \frac{k}{r} \frac{1}{\sum_{i=1}^{r} \frac{1}{i}} \quad (1)$$

where the $k = 1$ to r probability values P_k are ordered from largest to smallest, and is P_k^* the adjusted probability benchmark. When a relatively small number of tests are being conducted, as in most experimental and correlative analyses in ecology, there are good philosophical and statistical reasons for avoiding corrections or adjustments to P-values (Gotelli and Ellison 2004; Moran 2003). But for the analysis of species pairs, there are typically hundreds and thousands of comparisons being made, so some adjustment of the standard P-value is prudent.

Empirical Bayes mean based (Bayes mean-based criterion)

Another way to reduce FDERs is to use an empirical Bayes approach (Efron 2005). Assume we have a species $\times$ sites matrix with n species and m sites. We use an appropriate metric for non-random species associations and compare the scores for each species pair with those obtained for the same species pair in randomized matrices. Instead of comparing observed and expected scores of all $n(n-1)/2$ species pairs directly, we first plotted the frequency distribution of pairwise scores generated by the null model. This gives us an impression of how often certain scores are expected by chance, irrespective of whether those scores are statistically significant when analyzed by themselves. We then compare the observed distribution of scores with this null distribution.

For example, Fig. 1 shows a distribution of rescaled C score values based on the occurrence matrix of 71 ground beetle species on 17 lake islands in northern Poland (Ulrich and Zalewski 2006). The C score index was rescaled on a 0.0–1.0 interval (0.0 = complete overlap, 1.0 = complete segregation), and the 2,485 pairwise score values were grouped into 22 classes of evenly spaced bins between 0.0 and 1.0. Within each bin, the scores were ranked from smallest to largest. Next, the original matrix was randomized according to a standard procedure in which the row and column totals of the observed matrix were preserved (Gotelli 2000). The randomization was repeated 1,000 times, and the average number of species pairs in each bin was calculated to generate a null distribution of the frequencies of species pairs with different scores. In bin A (see Fig. 1), 35 species pairs were observed with a score between 0.35 and 0.40, whereas only 32.2 such pairs, on average, were generated in the 1,000 null matrices. In bin B, 24 species pairs were observed with a score between 0.60 and 0.65, but only 20 (19.85) were expected from the null distribution. If we were to retain all of these pairs, the FDER would be 32.2/35 = 92% and 20/24 = 83%, regardless of whether the individual pairs were statistically significant or not. Therefore, we considered only the largest scores within each bin that were above this null expectation. These scores constituted 4 and 6% of pairs, respectively, within each bin. Within this group of species pairs, we retained only those for which the simple null hypothesis was rejected by the standard randomization procedure (i.e., the C score for that specific pair was larger than 95% of the C scores generated for that same species pair in the simulation). In this way, we substantially reduced the number of false positives and the probability of a type I error. We call this selection criterion *mean-based* (Bayes M criterion) because it counts the number of species pairs in each bin that are greater than the mean number of simulated species pairs.

Empirical Bayes CL based (Bayes CL criterion)

As illustrated in bins C and D of Fig. 1, a more conservative test would be to consider only that fraction of species pairs beyond the upper 95% CL of the corresponding null model score (the expected number of significant

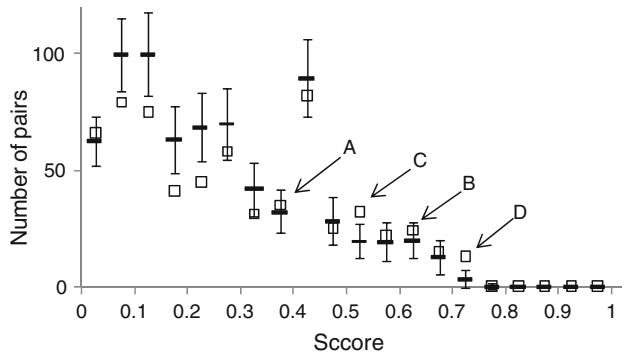


Fig. 1 Observed (*squares*) and expected (*bold lines*) [*error bars* denote the upper and lower 95% confidence limits (CL) of this expectation] distribution of pairwise co-occurrence scores (normalized within the range of 0–1) from a matrix of the occurrence of 71 ground beetle species on 17 lake islands in northern Poland (Ulrich and Zalewski 2006). In bins *A* and *B*, the observed number of species pairs exceeds the expected number calculated from the average of 1,000 null assemblages (Bayes M criterion). In bins *C* and *D*, the observed number of species pairs is even higher than the upper 95% CL (Bayes CL criterion)

scores). In bin C, 32 species pairs had scores between 0.50 and 0.55, whereas 19.5 ± 4.0 (mean $\pm$ 95% CL) were expected (26 pairs marks the upper 95% CL). In bin D, 13 species pairs were observed with scores between 0.70 and 0.75, but only 3.1 ± 1.9 pairs were expected (eight pairs mark the upper 95% CL). The best candidate pairs for non-random association are the 1–26/32 = 19% (bin C) and 1–8/13 = 38.5% (bin D) of species pairs that are the furthest outside their respective CLs (Bayes CL criterion).

For many other score classes in Fig. 1, the number of observed species pairs was less than expected, but our interpretations are restricted to the positive deviations in the different bins. Both of the Bayesian selection criteria (mean-based and CLs) greatly reduce the number of potential candidate pairs for non-randomness. Using our Bayesian selection criteria reduced the number of pairs that are the best candidates for non-random association to 62 (Bayes M criterion) and 16 (Bayes CL criterion). The next step of analysis is to check whether the individual scores of any of these 62 or 16 pairs deviated significantly from the null distribution by the simple CL criterion. In the present case the Bayes M criterion identified 6 pairs and the Bayes CL criterion one pair as being significantly segregated. For comparison, of the 2,485 species pairs, 42 were significant by the simple CL criterion, and 15 after the Bonferroni sequential correction (BY criterion).

Note that pairs identified by the Bayes CL criterion (the most conservative method) are a subset of those identified by the Bayes M criterion, which are a subset of those identified by the simple CL criterion (the most liberal method). The species pairs identified by the BY criterion are also a subset of those identified by the simple CL criterion, although the same pairs are not necessarily identified by the Bayes' criteria.

Summary of empirical Bayes methods

To summarize, the steps in the empirical Bayes methods are:

1. Rescale the C score index to a range from 0 to 1.
2. Calculate the rescaled C score index for all $n(n-1)/2$ species pairs and assign each score to a histogram bin.
3. Create 1,000 null assemblages (using the fixed–fixed or other randomization algorithm) and calculate the mean and confidence interval for the number of species pairs that are within each bin.
4. Within each bin, order the species pairs by their scores and retain the species pairs with the largest scores that place them above the mean (Bayes M criterion) or above the 95% confidence interval (Bayes CL criterion) for the number of species pairs expected from the simulated distribution.
5. Further reduce this set by retaining only those species pairs that are statistically significant in an individual test (simple CL criterion).
6. Classify each non-random species pair as segregated or aggregated.

There are some points to note about our application of the empirical Bayes approach. First, significant species pairs are defined only as positive deviations from the null distribution. Bins in which there is a deficit of species pairs cannot be readily interpreted because we cannot directly identify the "missing species pairs" that are responsible for the deficit. However, because the total count of species pairs in the observed and the average of the simulated distribution is the same, deficits in some bins must be compensated for by excesses in other bins. If those positive deviations are concentrated in one or a small number of bins, they will be scored as statistically significant.

Second, the results may be sensitive to the number and size of bins used, as would be the case for any statistical method that constructs a histogram of samples from a continuous distribution. If too few bins are used, we cannot distinguish between species pairs that exhibit strong versus weak segregation or aggregation. If too many bins are used, the precision of the estimate diminishes. There is also a risk that, with many bins, some may show an excess of species just by chance. However, this is less of a problem for the Bayes CL method because the confidence intervals will also become larger as the number of bins increases. In general, effects of bin number will diminish as the sample size increases. Fortunately, even modest numbers of

species generate many species pairs: with species numbers from only 15–45, the resulting distribution will have between 100 and 1,000 unique species pairs.

Third, the Bayes methods (and the standard confidence interval methods) identify which species pairs are non-random, but they do not specify whether the pattern is one of segregation or aggregation. To classify species pairs as aggregated or segregated, we compared the observed C score with the mean of the simulated C scores for a particular species pair. Segregated pairs are those for which the observed C score was greater than the average simulated C score, and aggregated pairs are those for which the observed C score was less than the average simulated C score. Naturally, the majority of the segregated species pairs occur in the bins that are close to 1.0, and the majority of the aggregated species pairs occur in bins that are close to 0.0. These pairs represent cases of very strong segregation (perfect or near perfect checkerboard distributions) or very strong aggregation (complete or nearly complete overlap). However, as seen in Fig. 3, there is also a collection of aggregated and segregated species pairs that occur over a range of bin values, with the weakest patterns being for those bin values closest to 0.5. These represent cases in which the rescaled overlap is approximately 50%, but the observed C score is still more extreme than 95% of the simulated values. These examples correspond to species pairs exhibiting relatively weak, but statistically significant, segregation or aggregation.

Finally, it might not be apparent to readers why these methods are "Bayesian" since they appear to be based on a frequentist computation of false discovery rates (Benjamini and Hochberg 1995). However, Efron (2005) notes that false discovery rate calculations have a good Bayesian rationale. We assume that some proportion, p_0, of species pairs overlap randomly in occurrence, and the rescaled C scores of these pairs follow a null distribution, f_0. The remaining species pairs do not overlap randomly and follow a distribution f_1. If we know p_0, f_0, and f_1, Bayes' rule yields the probability that a pair is random or non-random, given its rescaled C score value. With large sample sizes, we can use the data to get a frequentist estimate of the prior quantities, and then use these estimates to approximate Bayes' rule. This empirical Bayes method turns out to give results that are very similar to the false discovery rate calculations. See Efron (2005) for further discussion.

Construction of random benchmark matrices

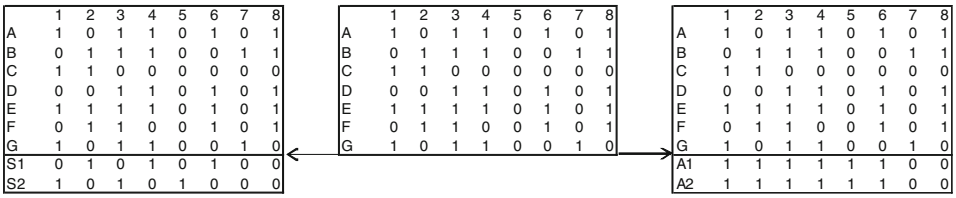
To assess the utility of these four procedures (or of any null model procedure), it is first necessary to benchmark their performance with artificial matrices that contain predetermined amounts of randomness and structure (Gotelli 2001). Matrices that are constructed by random processes should not yield an excessive number of non-random species pairs (type I statistical error). Matrices that incorporate segregated or aggregated species pairs should not yield random patterns for those species pairs (type II statistical error). Following the method of Ulrich and Gotelli (2007a), we generated two types of random presence–absence matrices (100 matrices each) to study the performance of the four selection criteria. We created the first type of presence–absence matrix (M_N) by randomly sampling individuals from a metacommunity in which population sizes of the species were distributed according to a lognormal species rank-order distribution:

$$S = S_0 e^{[-a(R-R_0)^2]} \quad (2)$$

in which S is the number of species, R is the abundance octave, S_0 is the number of species in the modal octave R_0, and a is the shape-generating parameter of the lognormal distribution. Individuals were randomly sampled until a specified number of species per sites was achieved. For each matrix, the shape-generating parameter a was sampled randomly from a uniform distribution between 0.1 and 2.0 (a canonical lognormal has $a = 0.2$; May 1975). Total species numbers m and total sites numbers n per matrix were also sampled from uniform distributions ($10 \leq m \leq 100$ and $10 \leq n \leq 100$). This sampling protocol produced random matrices with relatively high matrix fills that were moderately to strongly nested due to passive sampling (Higgins et al. 2006). In the second type of random matrix (M_E), species occurrences were sampled from a uniform random distribution until a randomly chosen number of species (again $10 \leq m \leq 100$ and $10 \leq n \leq 100$) and matrix fill ($0.1 < \text{fill} < 0.9$) was reached. This sampling protocol produced random matrices with relatively low matrix fills that exhibited little or no nested structure.

Construction of seeded benchmark matrices

In three series of diagnostic tests, we modified our original M_N and M_E matrices by adding to each matrix two additional species to create a single non-random species pair embedded in the matrix ("seeded" matrices). This additional "primary pair" of species contained 50% (M_NS50), 75% (M_NS75), or 90% (M_NS90) of the sites in a perfectly segregated (checkerboard) pattern, or 50% (M_EA50), 75% (M_EA75), or 90% (M_EA90) of the sites in a perfectly aggregated pattern (Fig. 2). Note that a perfectly aggregated pair in which both species occur in all sites cannot be detected with the fixed–fixed null model because this pair of species would also have the same distribution in all of the null matrices. We counted how many new statistically significant "secondary pairs" were introduced through this manipulation and how the addition



	1	2	3	4	5	6	7	8
A	1	0	1	1	0	1	0	1
B	0	1	1	1	0	0	1	1
C	1	1	0	0	0	0	0	0
D	0	0	1	1	0	1	0	1
E	1	1	1	1	0	1	0	1
F	0	1	1	0	0	1	0	1
G	1	0	1	1	0	0	1	0
S1	0	1	0	1	0	1	0	0
S2	1	0	1	0	1	0	0	0

	1	2	3	4	5	6	7	8
A	1	0	1	1	0	1	0	1
B	0	1	1	1	0	0	1	1
C	1	1	0	0	0	0	0	0
D	0	0	1	1	0	1	0	1
E	1	1	1	1	0	1	0	1
F	0	1	1	0	0	1	0	1
G	1	0	1	1	0	0	1	0

	1	2	3	4	5	6	7	8
A	1	0	1	1	0	1	0	1
B	0	1	1	1	0	0	1	1
C	1	1	0	0	0	0	0	0
D	0	0	1	1	0	1	0	1
E	1	1	1	1	0	1	0	1
F	0	1	1	0	0	1	0	1
G	1	0	1	1	0	0	1	0
A1	1	1	1	1	1	1	0	0
A2	1	1	1	1	1	1	0	0

Fig. 2 Adding a segregated species pair (*left matrix*) or an aggregated species pair (*right matrix*) to an initially random matrix (*center matrix*) can generate statistically significant secondary associations with other species (Table 1). The seeded species pair was aggregated or segregated in 75% of the sites (M_E75; see text for details)

of the primary pair changed the C score of the entire matrix. These secondary pairs represent associations between one species in the original matrix and one of the two new species added to the seeded matrix. We also counted how often C score identified all pairs as being non-random using the upper and lower 95% CLs of the null model as benchmarks. The observed fraction of significant scores at the 5% error level was then tested with the Bayesian criteria (Bayes M and Bayes CL) to decide which of these significant scores might be the best candidates for non-randomness. We also tested for correlations between matrix properties (size, fill) and the frequency of non-random species associations.

Published matrices

In addition to the analysis of artificial data matrices, we also analyzed 272 published data matrices (with transposed matrix rows and columns to match the format used here) from the set of 294 matrices that were originally compiled from the literature by Atmar and Patterson (1995). We excluded 12 matrices from the Atmar and Patterson (1995) compilation because three of them contained only one row or one column, and nine others were overly constrained because the matrix was either too full or too empty, so it could not be effectively randomized with the fixed–fixed algorithm.

Species associations

We used the C score (Stone and Roberts 1990) to measure species co-occurrences. The C score measures the average pairwise species segregation for the entire matrix (Gotelli 2000). To compare single species pairs with different numbers of occurrences directly in the Bayesian analyses, we normalized the C score (CS) for each species pair *ij* into a range from 0 (perfect overlap) to 1 (no overlap):

$$\mathrm{CS}_{ij} = \frac{(k_i - K)(k_j - K)}{k_i k_j} \qquad (3)$$

where k_i and k_j denote the numbers of occurrences of species *i* and *j*. *K* is the number of co-occurrences of both species. The same normalization was used on all of the species pairs in the null matrices.

We used the fixed row–fixed column randomization algorithm that maintains both observed row and column totals (Connor and Simberloff 1979; Gotelli 2000) to generate randomized matrices that serve as null models to obtain CLs for the C score. The biological justification for this algorithm is that it preserves in the null matrices observed differences among sites in species richness (column totals) and observed differences among species in their frequency of occurrence (row totals). Co-occurrence patterns are detected above and beyond those introduced by these constraints. The performance of the fixed–fixed algorithm is relatively insensitive to matrix properties such as the size and fill of the matrix and the frequency distribution of species occurrences (Gotelli 2000; Ulrich and Gotelli 2007a, b). We implemented this null model with a variation of the "sequential swap algorithm" (Manly 1995; Gotelli and Entsminger 2001), in which we sequentially reshuffled 25,000 randomly sampled 2 × 2 submatrices that have the same row and column totals after their elements were swapped. Matrices created this way have the same row and column totals as the original matrix. Each subsequent matrix was created with an additional 1,000 swaps. The sequential swap algorithm does not sample all possible matrices with fixed row and column totals with equal probability (Zaman and Simberloff 2002; Miklós and Podani 2004). However, the bias is small (Lehsten and Harmand 2006), and this algorithm has performed well in numerous benchmark tests (Stone and Roberts 1990; Gotelli 2000; Gotelli and Entsminger 2001, 2003; Ulrich 2004; Ulrich and Gotelli 2007a).

For each of the published matrices, we compared the observed number of non-random pairs identified by the four methods with the expected numbers obtained from 100 randomized matrices (using the sequential swap algorithm; number of swaps: 10 nm). Results were stable and would not have differed qualitatively if we had used 1,000 instead of 100 randomized matrices.

Null models and co-occurrence indices were calculated with the software applications Pairs and Matrix (Ulrich

2008). The online appendix provides a spreadsheet with the original data matrix (Ulrich and Zalewski 2006) and fully documented output from the Pairs analysis illustrated in Fig. 1.

Results

Benchmark random matrices

Between 3.72 and 4.40% of the original M_N and M_E matrices had significantly aggregated or segregated species pairs as judged by the 95% CL benchmark (CL criterion) of the fixed–fixed null model (Table 1). The BY criterion reduced this fraction to 0.97% for the M_E and 1.65% for the M_N matrices. The Bayes M criterion returned 0.58% significant pairs for the M_E and 1.67% for the M_N matrices. The more conservative Bayes CL criterion identified only 0.03% (M_E) and 0.48% (M_N) as being significant. That means that for an M_E random matrix of 50 species (1,225 species pairs) approximately 46 pairs are expected to be significant at the 5% error benchmark from the CL criterion, but only 11, seven, and no pairs, respectively, from the BY, the Bayes M, and the Bayes CL criteria.

Irrespective of matrix type (M_N or M_E), for all four diagnostic criteria, the proportion of statistically significant pairs was correlated with the number of sites in the matrix, but not with the percentage fill of the matrix (Table 2). Surprisingly, the matrix-wide C score was only weakly correlated with the number of significant pairs within the matrix ($r = 0.08$–0.26).

Seeded matrices

The addition of one perfectly aggregated or segregated species pair to our matrices caused associations with other species of the matrix that C score identified as being non-random (secondary pairs). Adding one aggregated pair to the M_E matrices added, on average, between five (M_EA90) and ten (M_EA50) aggregated secondary pairs as detected by the CL criterion. For the M_N matrices, between three (M_NA90) and 22 (M_NA50) significantly aggregated secondary species pairs emerged (Table 3). Adding a segregated primary pair introduced between five (M_ES50) and 44 (M_NS90) segregated secondary pairs according to the CL criterion. The BY, Bayes M, and Bayes CL criteria reduced these numbers significantly and identified between 0 (M_ES90) and 28 (M_NS90) non-random secondary pairs. More than ten secondary pairs appeared only for the M_NS90 matrices.

For matrices seeded with a segregated primary species pair, the CL criterion correctly identified over 89% (M_NS50) of these species pairs as non-random (Table 4). The Bayes M and Bayes CL criteria identified 41% (M_ES50) to 93% (M_NS90) and the BY criterion 59% (M_ES50) to 100% (M_ES90) as being non-random. For matrices seeded with an aggregated species pair, the Bayes M and Bayes CL criteria were more conservative and correctly identified for the M_E matrices between 17% (M_EA90) and 79% (M_EA50) of the added pairs as being non-random. In the case of the M_N matrices, the Bayes CL criterion usually failed to recover the non-random primary pair (<10% identified). The most liberal CL criterion

Table 1 Percentages of species pairs per matrix ranging outside the 95% confidence limit (CLs) of the fixed–fixed null model for a set of random M_E and M_N matrices before and after the introduction of one perfectly aggregated or segregated pair occurring on 50% (M_{50E}), 75% (M_E75), and 90% (M_E90) of the sites

Percentage of significant pairs							
Original matrix (%)	Added species pair	M_E50		M_E75		M_E90	
		All pairs (%)	Secondary pairs only (%)	All pairs (%)	Secondary pairs only (%)	All pairs (%)	Secondary pairs only (%)
3.72	Aggregated	4.04	0.48	4.05	0.49	3.82	0.25
3.84	Segregated	3.97	0.30	4.00	0.34	3.95	0.29
Percentage of significant pairs							
Original matrix (%)	Added species pair	M_N50		M_N75		M_N90	
		All pairs (%)	Secondary pairs only (%)	All pairs (%)	Secondary pairs only (%)	All pairs (%)	Secondary pairs only (%)
4.14	Aggregated	5.24	1.29	4.81	0.86	4.13	0.18
4.40	Segregated	4.53	0.33	5.07	0.87	6.45	2.24

Secondary pairs Include one member of the new species pair and one of the original species

Table 2 Spearman's rank order correlations between the proportion of significant non-random species pairs (both aggregated and segregated pairs) and basic matrix properties

	M_E				M_N			
	CL criterion	Bayes M criterion	Bayes CL criterion	BY criterion	CL criterion	Bayes M criterion	Bayes CL criterion	BY criterion
Species	0.09	0.53***	0.05	0.15	0.22*	0.04	0.38***	0.21*
Sites	0.69***	0.47***	0.09	0.50***	0.74***	0.61***	0.50***	0.60***
Matrix fill	0.27**	−0.09	−0.02	0.25*	−0.08	−0.05	−0.15	−0.04
Matrix size	0.54***	−0.08	0.09	0.48***	0.68***	0.48***	0.67***	0.55***
Matrix shape	−0.38***	−0.69***	−0.01	−0.22*	−0.34***	−0.40***	0.09	−0.28**
C score	0.19	0.08	0.26*	0.23*	0.24*	0.21*	0.08	0.24*

See text for details on matrix structure (M_E and M_N) and testing criteria [CL, Bayes mean-based *(Bayes M)*, Bayes CL, Benjamini and Yekutieli (*BY*)]

Matrix fill Percentage of matrix cells occupied, *Matrix size* number of species in the matrix (*m*) × number of sites in the matrix (*n*), *Matrix shape m/n*, *C score* C score for entire matrix

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

performed best in recovering the non-random aggregated seeded pair (Table 4) while the CL and the BY criterion did equally well in recovering the segregated seeded pair.

In the seeded matrices, all of the non-random pairs identified by the Bayes and BY criteria were (by definition) a subset of those pairs identified by the CL criterion. However, the Bayes and the BY criteria did not necessarily identify the same secondary species pairs as being non-random. The percentage of commonly identified pairs for the M_E matrices was only 15% (Bayes M–BY) and 5% (Bayes CL–BY) and for the M_N matrices 36% (Bayes M–BY) and 19% (Bayes CL–BY). As a consequence, the joint application of criteria reduced the fraction of significant pairs in the M_E and M_N matrices compared to the single criterion by a factor of 2–3 (not shown). Detection probabilities of the added pairs remained at the level of the respective Bayes criteria.

The addition of one perfectly segregated or aggregated pair generated segregated or aggregated secondary pairs with even more extreme scores than those of the seeded pair (Table 5). This effect was particularly strong for the matrices with an aggregated pair added: 60 and 247 perfectly aggregated secondary pairs were then detected. For the M_NA90 matrices, the Bayes M and Bayes CL criteria did not identify any of the primary pairs, but instead identified 41 and 14, respectively, of the secondary pairs as being non-random (Table 5). For matrices that were seeded with a segregated primary pair, spurious detection of significant secondary pairs was much weaker: The Bayes M and Bayes CL criteria detected at most five species pairs with scores higher than any of the scores of the added pairs. The BY criterion performed worse and identified between three and 11 such pairs.

Published matrices

The 272 published matrices contained a total of 463,768 unique species pairs. Of these, the simple CL criterion identified 0.73% as segregated and 0.48% as aggregated. By chance, approximately 5% should have been significant. However, even the most stringent Bayes CL and BY criteria identified significantly more segregated and aggregated pairs than expected (Fig. 3). In particular, the number of perfect checkerboard pairs exceeded the expectation by a factor of more than fourfold for BY and 50-fold for Bayes CL. On the other hand, the total frequency of significantly segregated and aggregated pairs identified by the Bayes CL criterion was only 0.18% (segregation) and 0.03% (aggregation), which was similar in magnitude to the randomized matrices. Regardless of the criterion used, consistently more segregated than aggregated species pairs were detected. The overall C score of the matrices was significantly positively correlated with the number of segregated pairs per matrix (Bayes CL criterion: Spearman's $r = 0.47$; $P < 0.0001$), but was also weakly positively correlated with the number of aggregated pairs (Spearman's $r = 0.35$; $P < 0.0001$). This seeming contradiction reflects the fact that a perfectly segregated primary pair can also generate perfectly or partly aggregated secondary pairs. Numbers of aggregated and segregated pairs per matrix were indeed positively correlated with one another (Spearman's $r = 0.42$; $P < 0.0001$).

A detailed analysis of the matrices revealed that only 15 matrices (Table 6) accounted for more than 83.9% of all identified segregated pairs. Ten matrices with significantly higher numbers of aggregated pairs accounted for 83.6% of all of these pairs (Table 6). In only 54 matrices did the

Table 3 Mean number of species pairs per matrix ranging outside the 95% CLs of the fixed–fixed null model for each combination of matrix (M_E and M_N) and added species pairs (segregated and aggregated)

	M_E50								M_N50							
	CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion		CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion	
	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only
Aggregated	85	10	17	4	4	2	23	2	90	22	38	10	13	4	33	7
Segregated	71	5	12	0	2	0	19	1	88	6	30	0	9	0	34	2
	M_E75								M_N75							
	CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion		CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion	
	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only
Aggregated	86	10	20	7	5	3	23	2	83	15	32	4	11	2	31	4
Segregated	72	6	14	0	2	0	19	2	99	17	42	8	14	4	41	9
	M_E90								M_N90							
	CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion		CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion	
	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only	All pairs	Sec. pairs only
Aggregated	81	5	15	3	2	0	22	1	71	3	24	0	8	0	28	1
Segregated	71	5	14	1	2	0	18	1	126	44	58	24	26	15	60	28

See text for details on matrix structure (M_E and M_N) and testing criteria (CL, Bayes M, Bayes CL, BY). *Sec. pairs* Secondary pairs (significant pairs containing one of the two added species); for other abbreviations, see Table 2

The mean M_N and M_E matrices both had 55 species (=1,485 species pairs) and 55 sites

Table 4 Percentage of species pairs ranging outside the 95% CLs of the fixed–fixed null model for each combination of matrix (M_E and M_N) and seeded species pairs (segregated and aggregated) that were added to a random matrix

	M_E50								M_N50							
	CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion		CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion	
	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)
Aggregated	4.0	100.0	0.8	79.0	0.2	56.0	1.1	79.0	5.2	100.0	2.2	52.0	0.8	10.0	1.9	46.0
Segregated	4.0	69.0	0.7	61.0	0.1	41.0	1.0	59.0	4.5	89.0	1.6	84.0	0.4	72.0	1.7	89.0
	M_E75								M_N75							
	CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion		CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion	
	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)
Aggregated	4.1	100.0	0.9	57.0	0.2	39.0	1.1	77.0	4.8	99.0	1.9	11.0	0.6	1.0	1.8	31.0
Segregated	4.0	94.0	0.8	81.0	0.1	56.0	1.0	94.0	5.1	99.0	2.1	91.0	0.7	78.0	2.1	99.0
	M_E90								M_N90							
	CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion		CL criterion		Bayes M criterion		Bayes CL criterion		BY criterion	
	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)	All pairs (%)	Seeded pairs (%)
Aggregated	3.8	100.0	0.7	26.0	0.1	17.0	1.0	83.0	4.1	91.0	1.4	0.0	0.4	0.0	1.6	32.0
Segregated	4.0	100.0	0.8	89.0	0.1	53.0	1.0	100.0	6.4	99.0	3.0	93.0	1.3	78.0	3.1	99.0

All pairs Includes the seeded pair plus any secondary pairs that were also significant; for other abbreviations, see Table 2

Table 5 Total numbers (out of 100 seeded matrices) of perfectly segregated or aggregated secondary species pairs having a C score higher than at least one of the primary segregated or aggregated species pairs

Matrix	CL criterion	Bayes M criterion	Bayes CL criterion	BY criterion
M_EA50	60	28	3	11
M_EA75	84	29	3	9
M_EA90	95	6	0	5
M_ES50	9	5	1	5
M_ES75	8	3	0	8
M_ES90	5	3	0	8
M_NA50	174	52	7	9
M_NA75	247	13	3	12
M_NA90	137	41	14	9
M_NS50	3	1	0	3
M_NS75	5	1	0	5
M_NS90	4	1	1	4

For abbreviations, see Table 2

number of aggregated or segregated species pairs exceed the expected numbers (positive effect sizes). Hence for at least 218 matrices we did not observe a significant trend towards species segregation or aggregation when using a species pair approach. However, in 107 matrices we did find a significant matrix-wide C score (cf. Gotelli and McCabe 2002), even when no individual cases of strongly segregated species pairs could be detected.

Vanuatu and Galapagos matrix analyses

The Vanuatu bird matrix contains 56 species and 28 sites (Diamond and Marshall 1976). As demonstrated in other analyses (Stone and Roberts 1990; Gotelli and Entsminger 2001; Zaman and Simberloff 2002; Miklós and Podani 2004), this matrix is significantly segregated using the matrix-wide C score and the fixed–fixed null model (Z-score = 4.23, $P < 0.001$). The simple CL criterion identified 26 significantly segregated (1.7%) and 22 (1.4%) significantly aggregated pairs. These results are roughly comparable to those of Zaman and Simberloff (2002), who used a corrected swap algorithm and a different statistical method for their CL criterion.

In our analyses, none of the species pairs identified with the CL criterion were significant with the BY criterion. The Bayes M criterion identified three species pairs [*Megapodius freycinet layardi* (scrubfowl)–*Poliolimnas cinereus tannensis* (white-browed Crake); *Clytorhynchus pachycephaloides* (southern shrikebill)–*Porzana tabuensis tabuensis* (spotless crake); *Lichmera incana flavotincta* (silver-eared honeyeater)–*Dukula bakeri* (Baker's imperial pigeon)] as being significantly segregated and another three pairs [*Anas superciliosa pelewensis* (Pacific black

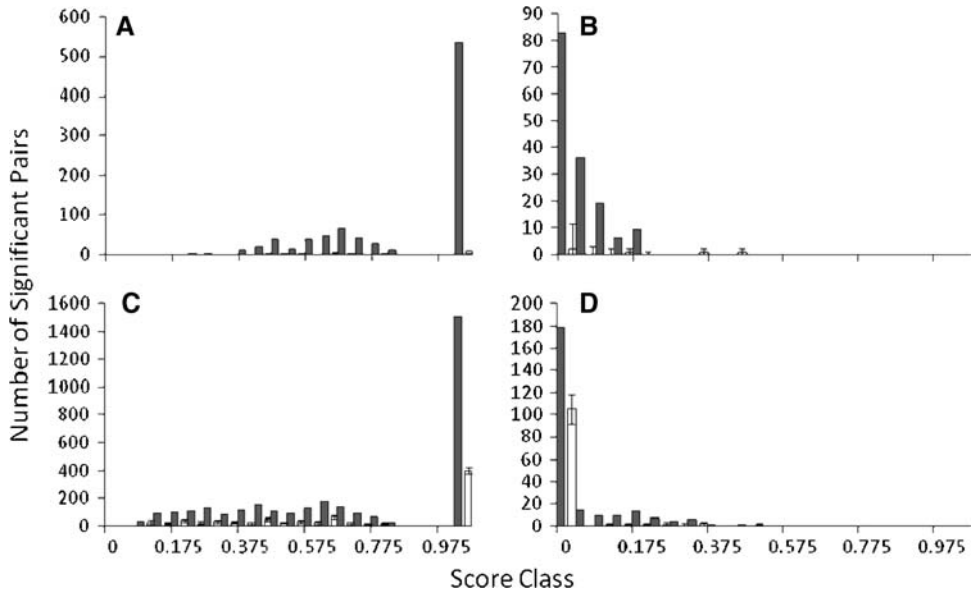


Fig. 3 Numbers of significant species pairs in 272 published data matrices (in total 463,768 species pairs) identified by the Bayes CL (**a**, **b**) and the Benjamini and Yekutieli (**c**, **d**) criterion for each class of the rescaled C score. The rescaled C score ranges from 0.0 (complete aggregation) to 1.0 (complete segregation). **a**, **c** Significant ($P < 0.05$) segregation; **b**, **d** significant aggregation. *Solid bars* are the number of species pairs (summed over all matrices) in each C score class, *open bars* are the sum of the average number of species pairs from 100 null assemblages created for each original matrix. The *vertical line* represents 1 SD for each bin. Most of the segregated distributions are pairs of species that are completely non-overlapping (checkerboard distributions; rescaled C score = 1.0), and most of the aggregated distributions are pairs of species that overlap completely (rescaled C score = 0). However, there are also statistically significant species pairs from a range of bin values that show relatively weak patterns of segregation or aggregation

Table 6 Presence–absence matrices of the Atmar and Patterson (1995) data set with the greatest numbers of significant segregated (at least 16) and aggregated (at least four) species pairs (Bayes CL criterion)

Matrix	Species	Sites	Matrix fill	Percentage of pairs	Matrix-wide C score	Reference
Segregated pairs						
Amazonian bats	82	19	0.47	4.03	66.52	Patterson et al. (1996)
Amazonian bats (part)	82	19	0.34	3.67	55.31	Patterson et al. (1996)
Pacific shore fish	179	16	0.58	0.42	9.92	Springer (1982)
North American water birds	152	15	0.26	0.47	6.10	Hatt et al. (1948)
Australian island plants	147	49	0.08	0.46	7.99	Abbott and Black (1980)
Antillean trees (part)	61	11	0.49	2.57	31.37	Beard (1948)
Antillean trees (part)	112	14	0.64	0.74	25.31	Beard (1948)
Antillean trees (part)	102	12	0.43	0.83	10.81	Beard (1948)
Finnish island birds	82	16	0.39	0.87	15.10	Haila et al. (1980)
Andean butterflies	87	13	0.17	0.64	7.44	Descimon (1986)
North American prairie plants	39	102	0.38	3.24	12.31	Glass, unpublished data
Canary Island birds	78	7	0.51	0.80	16.94	Bacallado (1976)
New Zealand birds	57	31	0.26	1.44	3.28	Patterson (1987)
Canary Island birds (part)	61	7	0.66	1.04	15.47	Bacallado (1976)
New Zealand birds (part)	53	22	0.33	1.16	4.49	Patterson (1987)
Aggregated pairs						
Amazonian bats	82	19	0.47	0.54	66.52	Patterson et al. (1996)
West Australian snails	35	55	0.22	2.18	0.99	Cameron (1992)
African ostracods	104	38	0.21	0.24	5.41	Cohen, unpublished data
North American weeds	128	26	0.25	0.14	5.66	Crowe (1979)
Antillean trees (part)	112	14	0.64	0.16	25.31	Beard (1948)
Antillean trees (part)	61	11	0.49	0.49	31.37	Beard 1948
North American desert mice	29	129	0.12	2.22	1.23	Brown and Kurzius (1987)
Amazonian bats (part)	82	19	0.34	0.24	55.31	Patterson et al. (1996)
North American prairie plants	39	102	0.38	1.08	12.31	Glass, unpublished data
Baja Islands herbaceous plants	84	48	0.08	0.23	3.48	Murphy (1983)
West Australian snails	20	55	0.11	4.21	9.24	Cameron (1992)
North American fish	78	48	0.06	0.17	1.42	Smith, unpublished data
North American fish	35	30	0.29	0.67	7.90	Hocutt et al. (1978)
North American subtidal invertebrates	37	18	0.40	0.60	−0.03	Sutherland and Karlson (1977)
North American desert rodents	14	48	0.19	2.20	5.03	Brown and Kurzius (1987)

The 15 segregated matrices accounted for 721 of the 859 identified segregated pairs within the entire data set of 272 matrices

The 15 aggregated matrices accounted for 128 of the 153 identified aggregated pairs

duck)–*Aythya australis* (hardhead); *Aytyia australis* (hardhead)–*Tachybaptus novaehollandiae leucosternos* (Australasian grebe); *Coracina caledonica seiuncta* (cuckoo-shrike)–*Halcyon farquhari* (chestnut-bellied kingfisher)] as being significantly aggregated. Using the more stringent Bayes CL criterion the pair *Lichmera incana flavotincta* (silver-eared honeyeater)–*Dukula bakeri* (Baker's imperial pigeon) remained significantly segregated and the pair *Coracina caledonica seiuncta* (cuckoo-shrike)–*Halcyon farquhari* (chestnut-bellied kingfisher) significantly aggregated.

The Galapagos data set (13 species, 17 islands) also had a highly significant matrix-wide C score ($P < 0.0001$), but neither the BY nor the Bayes CL criteria identified any of the species pairs as being non-random. However, the simple CL criterion and the Bayes M criterion both pointed to nine and seven pairs, respectively, as being significantly non-random (Table 7). All five of the segregated species pairs are in the genus *Geospiza*. From the total number of 15 *Geospiza* pairs (six species) at most one is expected just by chance at the 5% level. The matrix of 13 species contains 78 species pairs, of which only 19 are congeneric

Table 7 Significant non-random species pairs (simple CL criterion) of Galapagos finches, numbers of occurrences on the 17 islands, numbers of joint occurrences, and the probability levels for the null hypothesis of random association

Species1	Occurrences	Species 2	Occurrences	Joint occurrences	*P*
Geospiza fulignosa[a]	14	*Geospiza difficilis*[a]	10	7	<0.0001
Geospiza fortis[a]	13	*Geospiza difficilis*[a]	10	7	<0.001
Geospiza fortis[a]	13	*Geospiza conirostris*[a]	2	0	<0.0001
Geospiza scandens[a]	12	*Geospiza difficilis*[a]	10	6	<0.0001
Geospiza scandens[a]	12	*Geospiza conirostris*[a]	2	0	<0.0001
Platyspiza crassirostris	11	*Camarhynchus psittacula*	10	10	<0.05
Platyspiza crassirostris[a]	11	*Geospiza conirostris*[a]	2	0	<0.0001
Camarhynchus parvulus[a]	10	*Geospiza conirostris*[a]	2	0	<0.0001
Geospiza difficilis	10	*Camarhynchus psittacula*	10	6	<0.001

Except for the *Platyspiza crassirostris–Camarhynchus psittacula* pair, all species pairs are significantly segregated

[a] Pairs were identified by the Bayes M criterion

(24%). In contrast, five of the seven significant segregated pairs identified by the Bayes M criterion were congeners (71%) (χ^2 contingency test: $\chi^2 = 3.03$, $P = 0.08$). Sanderson (2000) and Sfenthourakis et al. (2006) reported similar results for this data set using the CL criterion, although Sanderson (2000) used a different null model algorithm.

Discussion

Pairwise tests of species co-occurrence patterns invariably reveal statistically significant associations in random matrices using the simple 95% CL criterion (Table 1). The sequential Bonferroni, Bayes M, and Bayes CL criteria substantially reduce such occurrences, although they do not entirely eliminate them from random matrices. However, these analyses reveal the unavoidable trade-off between type I and type II statistical errors. For random matrices that were seeded with a non-random pair of species, the simple CL criterion did the best job of recovering these patterns, whereas the Bonferroni and Bayes methods did not detect the non-random pair in a substantial number of cases.

One difficulty is that all four of the methods detected false "secondary pairs" of species associations that emerged when a single non-random association was added to the matrix (Table 1). This result probably reflects, in part, the complex non-independence among all species pairs when the null model preserves fixed row and column totals. However, these statistically significant secondary pairs were more of a problem for aggregated than segregated distributions. Previous authors have discussed the possibility of a "dilution effect" in null model analysis in which significantly segregated species pairs are not detected because too many pairwise comparisons are made between pairs of species that are not interacting (Diamond and Gilpin 1982; Colwell and Winkler 1984). However, our results suggest there may well be a "concentration effect" because the addition of a single non-random species pair to a random matrix may generate a number of significant secondary pairs.

The analysis of the 272 published matrices revealed that the majority of significant species pairs showed segregated, rather than aggregated distributions. There was a strong concentration of both highly segregated and highly aggregated species pairs, but also a set of species pairs that showed weaker, but still highly non-random, patterns of overlap (Fig. 3). However, most of the significant pairs were concentrated in a relatively small number of matrices (Table 6). In many cases, the overall C score of the matrix may be highly significant even though few or no individual pairs of species show non-random patterns. This result might reflect widespread, but weak species interactions ("diffuse competition"; Diamond 1975), or mechanisms of species segregation that are not related to species interactions (Gotelli and McCabe 2002; Sfenthourakis et al. 2006).

It is noteworthy that six of the 15 most aggregated matrices (40%) were for poikilotherm groups (snails, ostracods, and fish), whereas only two of the 15 most segregated matrices (13.3%) were for poikilotherms (Table 6). These results are consistent with matrix-wide patterns of species segregation (Gotelli and McCabe 2002), which were much stronger for homeotherms (and ant and plant matrices) than for poikilotherms (invertebrate, amphibian, reptile, and fish matrices). Although all matrices contain segregated, random, and aggregated species pairs, the frequencies of these pairwise patterns are consistent with the overall matrix score.

However, pairwise analyses may not always be concordant with overall matrix scores. Both the Galapagos and

Vanuatu bird matrices have significant matrix-wide segregation, but the pairwise analysis of the Vanuatu matrix revealed very few significant pairs of species, which are ecologically and phylogenetically heterogeneous. In contrast, the Bayes M criterion identified seven significant pairs in the much smaller Galapagos matrix (Table 7). Five of these seven species pairs were concentrated within the genus *Geospiza*, which is one of the few examples of a competitively structured community that has been supported by extensive null model analysis (Simberloff and Connor 1981; Schluter and Grant 1984; Sanderson 2000).

Although the Bayes criteria and the sequential Bonferroni test do a better job of guarding against type I errors than the simple CL criterion, all of the methods proposed here must be used with caution. Even the most stringent criteria still detected a small number of unusual pairs in a large random matrix, and random matrices that were seeded with significant species pairs generated spurious statistical associations with other species in the matrix.

Perhaps it is asking too much of a statistical analysis to reveal biologically meaningful pairwise associations with no other information than a binary presence–absence matrix. A similar limitation has emerged in regression analyses and model selection. Whereas ecologists often use stepwise criteria to select a subset of meaningful predictor variables, these methods do not always identify the correct underlying model. A more powerful approach is to specify a priori a set of potential biological models, fit them to the data, and then use model selection criteria to rank them or distinguish between them (Burnham and Anderson 2002; Shipley 2002). For presence–absence matrices, the best strategy might be to identify ahead of time guilds or subsets of potentially interacting species and restrict the analysis to these pairs.

Acknowledgments This work was supported in part by a grant from the Polish Ministry of Science to W. U. (KBN, 2 P04F 039 29). N. J. G. was supported by NSF grant 0541936. We thank Aaron Ellison for calling our attention to Efron (2005). The manuscript was improved by comments from two anonymous reviewers and associate editor W. M. Mooij.

References

Abbott I, Black R (1980) Changes in species composition of floras on islets near Perth, Western Australia. J Biogeogr 7:399–410

Atmar W, Patterson BD (1995) The nestedness temperature calculator: a visual basic program, including 294 presence absence matrices. AICS Research Incorporate and The Field Museum. http://www.aics-research.com/nestedness/tempcalc.html

Bacallado JJ (1976) Notas sobre la distribucion y evolucion de la avifauna Canaria. In: Kunkel G (ed) Biogeography and ecology in the Canary Islands. Junk, The Hague, pp 13–431

Beard JS (1948) The natural vegetation of the Windward and Leeward Islands. Oxford For Mem 21:1–192

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188

Brown JH, Kurzius MA (1987) Composition of desert rodent faunas: combinations of coexisting species. Ann Zool Fenn 24:227–237

Burnham KP, Anderson DR (2002) Model selection and inference: A practical information-theoretic approach. Springer, New York

Burns KC (2007) Patterns in the assembly of an island plant community. J Biogeogr 34:760–768

Cameron RAD (1992) Land snail faunas of the Napier and Oscar Ranges, Western Australia; diversity, distribution and speciation. Biol J Linn Soc 45:271–286

Colwell RK, Winkler DW (1984) A null model for null models in biogeography. In: Strong Jr, Simberloff D, Abele LG, Thistle AB (eds) Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, pp 344–359

Connor EF, Simberloff D (1979) The assembly of species communities: chance or competition? Ecology 60:1132–1140

Crowe TM (1979) Lots of weeds. J Biogeogr 6:169–181

Descimon H (1986) Origins of Lepidopteran faunas in the high tropical Andes. In: Vuilleumier F, Monasterio M (eds) High altitude tropical biogeography. Oxford University Press, Oxford, pp 500–532

Diamond JM (1975) Assembly of species communities. In: Cody ML, Diamond JM (eds) Ecology and evolution of communities. Harvard University Press, Cambridge, pp 342–444

Diamond JM, Gilpin ME (1982) Examination of the "null" model of Connor and Simberloff for species co-occurrences on islands. Oecologia 52:64–74

Diamond JM, Marshall AG (1976) Origin of the New Hebridean avifauna. Emu 76:187–200

Efron B (2005) Bayesians, frequentists, and scientists. J Am Stat Assoc 100:1–5

Gotelli NJ (2000) Null model analysis of species co-occurrence patterns. Ecology 81:2606–2621

Gotelli NJ (2001) Research frontiers in null model analysis. Glob Ecol Biogeogr 10:337–343

Gotelli NJ, Ellison AE (2004) A primer of ecological statistics. Sinauer, Sunderland

Gotelli NJ, Entsminger GL (2001) Swap and fill algorithms in null model analysis: rethinking the knight's tour. Oecologia 129:281–291

Gotelli NJ, Entsminger GL (2003) Swap algorithms in null model analysis. Ecology 84:532–535

Gotelli NJ, Graves GR (1996) Null models In ecology. Smithsonian Institution Press, Washington

Gotelli NJ, McCabe DJ (2002) Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. Ecology 83:2091–2096

Haila Y, Järvinen O, Vaisanen RA (1980) Habitat distributions and species associations of land bird populations on the Aland Islands, SW Finland. Ann Zool Fenn 17:87–106

Hatt RT, Van Tyne J, Stuart LC, Pope CH, Grobman AB (1948) Island life: a study of the land vertebrates of the islands of eastern Lake Michigan. Cranbrook Institute of Science. Bloomfield Hills, MI

Higgins CL, Willig MR, Strauss RE (2006) The role of stochastic processes in producing nested patterns of species distributions. Oikos 114:159–167

Hocutt CH, Denoncourt RF, Stauffer JR (1978) Fishes of the Greenbriar River, West Virginia, with drainage history of the Central Appalachians. J Biogeogr 5:59–80

Kammenga JE, Herman MA, Ouborg NJ, Johnson L, Breitling R (2007) Microarray challenges in ecology. Trends Ecol Evol 22:273–279

Lehsten V, Harmand P (2006) Null models for species co-occurrence patterns: assessing bias and minimum iteration number for the sequential swap. Ecography 29:786–792

Manly BFJ (1991) Randomization and Monte Carlo methods in biology. Chapman and Hall, London

Manly BFJ (1995) A note on the analysis of species co-occurrences. Ecology 76:1109–1115

May RM (1975) Patterns of species abundance and diversity. In: Cody ML, Diamond JM (eds) Ecology and evolution of communities. Harvard University Press, Cambridge, pp 81–120

McCoy ED, Heck KL (1987) Some observations on the use of taxonomic similarity in large-scale biogeography. J Biogeogr 14:79–87

Miklós I, Podani J (2004) Randomization of presence–absence matrices: comments and new algorithms. Ecology 85:86–92

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. Oikos 100:403–405

Murphy RW (1983) The reptiles: origins and evolution. In: Case TJ, Cody ML (eds) Island biogeography in the Sea of Cortez. University of California Press, Berkeley, pp 130–158

Patterson BD (1987) The principle of nested subsets and its implications for biological conservation. Conserv Biol 1:323–334

Patterson BD, Atmar W (1986) Nested subsets and the structure of insular mammalian faunas and archipelagos. In: Heaney LR, Patterson BD (eds) Island biogeography of mammals. Academic Press, London, pp 65–82

Patterson BD, Pacheco V, Solari S (1996) Distributions of bats along an elevational gradient in the Andes of south-eastern Peru. J Zool 240:637–658

Sanderson JG (2000) Testing ecological patterns. Am Sci 88:332–339

Sanderson JG (2004) Null model analysis of communities on gradients. J Biogeogr 31:879–883

Schluter D, Grant PR (1984) Determinants of morphological patterns in communities of Darwin's finches. Am Nat 123:175–196

Sfenthourakis S, Giokas S, Tzanatos E (2004) From sampling stations to archipelagos: investigating aspects of the assemblage of insular biota. Glob Ecol Biogeogr 13:23–35

Sfenthourakis S, Tzanatos E, Giokas S (2006) Species co-occurrence: the case of congeneric species and a causal approach to patterns of species association. Glob Ecol Biogeogr 15:39–49

Shipley B (2002) Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference. Cambridge University Press, Cambridge

Simberloff D, Connor EF (1979) Q-mode and R-mode analyses of biogeographic distributions: null hypotheses based on random colonization. In: Patil GP, Rosenzweig ML (eds) Contemporary quantitative ecology and related ecometrics. International Cooperative Publishing House, Fairland, pp 123–138

Simberloff D, Connor EF (1981) Missing species combinations. Am Nat 118:215–239

Springer VG (1982) Pacific plate biogeography, with special reference to shorefishes. Smithsonian Institution, Washington

Stone L, Roberts A (1990) The checkerboard score and species distributions. Oecologia 85:74–79

Sutherland JP, Karlson RH (1977) Development and stability of the fouling community at Beaufort, North Carolina. Ecol Monogr 47:425–446

Ulrich W (2004) Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. Oikos 107:603–609

Ulrich W (2008) Pairs—a FORTRAN program for studying pair-wise species associations in ecological matrices. www.uni.torun.pl/~ulrichw

Ulrich W, Gotelli NJ (2007a) Null model analysis of species nestedness patterns. Ecology 88:1824–1831

Ulrich W, Gotelli NJ (2007b) Disentangling community patterns of nestedness and species co-occurrence. Oikos 116:2053–2061

Ulrich W, Zalewski M (2006) Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. Oikos 114:338–348

Veech JA (2006) A probability-based analysis of temporal and spatial co-occurrence in grassland birds. J Biogeogr 33:2145–2153

Zaman A, Simberloff D (2002) Random binary matrices in biogeographical ecology—instituting a good neighbor policy. Environ Ecol Stat 9:405–421