NULL MODEL ANALYSIS OF SPECIES CO-OCCURRENCE PATTERNS

NICHOLAS J. GOTELLI[1]

Department of Biology, University of Vermont, Burlington, Vermont 05405 USA

Abstract. The analysis of presence–absence matrices with "null model" randomization tests has been a major source of controversy in community ecology for over two decades. In this paper, I systematically compare the performance of nine null model algorithms and four co-occurrence indices with respect to Type I and Type II errors. The nine algorithms differ in whether rows and columns are treated as fixed sums, equiprobable, or proportional. The three models that maintain fixed row sums are invulnerable to Type I errors (false positives). One of these three is a modified version of the original algorithm of E. F. Connor and D. Simberloff. Of the four co-occurrence indices, the number of checkerboard combinations and the number of species combinations may be prone to Type II errors (false negatives), and may not reveal significant patterns in noisy data sets. L. Stone and A. Robert's checkerboard score has good power for detecting species pairs that do not co-occur together frequently, whereas D. Schluter's V ratio reveals nonrandom patterns in the row and column totals of the matrix. Degenerate matrices (matrices with empty rows or columns) do not greatly alter the outcome of null model analyses. The choice of an appropriate null model and index may depend on whether the data represent classic "island lists" of species in an archipelago or standardized "sample lists" of species collected with equal sampling effort. Systematic examination of a set of related null models can pinpoint how violation of the assumptions of the model contributes to nonrandom patterns.

Key words: assembly rules; checkerboard distribution; coexistence; community structure; competition; co-occurrence; Monte Carlo simulation; null model; presence–absence matrix; randomization test; species combinations.

INTRODUCTION

The presence–absence matrix is the fundamental unit of analysis in community ecology and biogeography (McCoy and Heck 1987). In such a matrix, rows are species, columns are sites or samples, and entries are the presence (1) or absence (0) of a species in a site. Few issues in community ecology have been more contentious than the statistical analysis of these data matrices. Diamond (1975) described "checkerboard" distributions of avian species in the Bismarck Archipelago that never co-occurred, and argued that these patterns reflected community assembly rules. Connor and Simberloff (1979) compared presence–absence matrices with those generated by null models: Monte Carlo randomizations of real data that produced the number of checkerboards expected in the absence of biological interactions (Gotelli and Graves 1996). Connor and Simberloff (1979) found that patterns in many presence–absence matrices were similar to those generated by null models, and they concluded that assembly rules could not be inferred from the patterns. These analyses touched off a debate on null models that has lasted over 20 years, with no end in sight (reviewed in Harvey et al. 1983, Wiens 1989, Gotelli and Graves 1996).

I do not intend to defend here the philosophical basis of null model analysis or to recount the ongoing con-

troversy over its use (Gotelli and Graves 1996). Instead, I accept the necessity of confronting models with data (Hilborn and Mangel 1997) and consider null models a useful statistical tool for this purpose. In this paper, I systematically compare a number of null models for species co-occurrence patterns to data matrices of known structure to address three questions:

1) What is the most appropriate null model algorithm? Although most attention and controversy has focused on the Connor and Simberloff (1979) algorithm of randomizing matrices while simultaneously maintaining row and column sums, a variety of null model algorithms are possible (e.g., Jackson et al. 1992). I compared the behavior of nine simulation algorithms that range from those that incorporate almost no constraints to those that are highly constrained by the structure of the original data matrix.

2) What index should be used to summarize species co-occurrence patterns? Once the null model simulation has been completed, a single number is used to summarize the pattern in the data. I compared four indices, ranging from Pielou and Pielou's (1968) early use of the number of unique species combinations to the more recent "checkerboard score" proposed by Stone and Roberts (1990).

3) How susceptible are the different algorithms and indices to Type I and Type II errors? In other words, what are the chances that a random matrix produces a statistically significant pattern (Type I error), and what are the chances that a structured matrix produces a

nonsignificant pattern (Type II error)? To answer these questions, I tested each of the nine simulation algorithms and four co-occurrence indices with matrices of known structure. For Type I errors, the algorithms and indices were tested against random matrices. For Type II errors, the algorithms and indices were tested against structured matrices that had increasing amounts of random "noise" added (Gotelli et al. 1997).

METHODS

Presence–absence matrices

The data are organized as a presence–absence matrix with $i = 1$ to R rows and $j = 1$ to C columns. Each row represents a species and each column represents a site or sample. Entry a_{ij} in the matrix represents the presence (1) or absence (0) of species i in sample j. Let S_i equal the row total for row i, that is, the total number of occurrences of species i across the sites. Let T_j equal the column total for column j, that is, the total number of species occurring in site j. Let N equal the total number of all species occurrences in the matrix. Thus:

$$S_i = \sum_{j=1}^{C} a_{ij}$$

$$T_j = \sum_{i=1}^{R} a_{ij}$$

$$N = \sum_{i=1}^{R} S_i = \sum_{j=1}^{C} T_j = \sum_{i=1}^{R} \sum_{j=1}^{C} a_{ij}.$$

Null model procedures

All of the null model procedures described follow the steps of classical statistical randomization tests (Edgington 1987):

1) Condense model predictions from theoretical ecology into a simple prediction about the structure of a presence–absence matrix. For the entire body of classical MacArthur competition theory, the usual prediction is that species co-occurrence is nonrandom and is less than would be expected in a community in which species colonized sites independently of one another (Gotelli and Graves 1996).

2) Define an index X that describes, with a single number, the co-occurrence pattern in a presence–absence matrix.

3) Measure X_{obs} for the observed data matrix.

4) Randomize the observed data matrix with a null model and record X_{sim} for the randomized matrix.

5) Repeat step (4) many times (typically 1000) to generate a frequency histogram of X_{sim}. This histogram represents the range of values expected for X_{obs} given that the null model is true. The histogram is a frequency distribution for testing the null hypothesis that X_{obs} was drawn at random from the distribution of X_{sim}.

6) Use classical methods of statistical inference to interpret X_{obs}. The probability that $X_{sim} = X_{obs}$ is the frequency of X_{sim} in the histogram, that is, the frequency of the bin located at X_{obs} in the histogram. Calculate tail probabilities from the cumulative frequency of simulated observations $\leq$ or $\geq X_{obs}$. By convention and tradition, accept the null hypothesis if $0.05 < P < 0.95$, and reject the null hypothesis if P exceeds these bounds.

7) Use the results of the statistical test to evaluate the original hypothesis. For some ecologists, this has been the most controversial aspect of null model analysis (Gotelli and Graves 1996), but I will not consider it further in this paper.

Selecting the co-occurrence index

I have analyzed the behavior of four indices that have been proposed for the analysis of co-occurrence patterns: (1) the number of species pairs forming perfect checkerboards (CHECKER); (2) the checkerboard score of the matrix (C score); (3) the variance ratio (V ratio); (4) the number of species combinations (COMBO). Each of these indices is a single number that summarizes co-occurrence patterns in a presence–absence matrix. Table 1 gives the details of how these indices are calculated and what the expected results are for a community that is structured by competition.

A universe of null models

Connor and Simberloff (1979) compared the observed pattern in a presence–absence matrix to the pattern present in random matrices that had the same row and column totals as the original matrix. Retaining the row and column totals preserves differences in species richness among sites (column totals) and differences in occurrence frequencies among species (row totals). Some authors have argued that these constraints make the null model overly conservative so that the null hypothesis is accepted more often than it should be (Type II error; Alatalo 1982, Diamond and Gilpin 1982, Gilpin and Diamond 1984; but see Connor and Simberloff 1983, 1984, 1986).

This is not the only way that row and column totals can be treated in a null model of species co-occurrence. If the sites are of similar size and quality (as in quadrat samples), a better null hypothesis might be that the sites (=columns) are equiprobable, so that the number of species in a site (column total) is allowed to vary, but all sites have the same average number of species. A compromise algorithm between fixed column totals and equiprobable columns would be to make the probabilities of species placement proportional to the observed column totals. This would allow column totals to vary, as in the equiprobable model, but would reflect observed differences among sites, as in the fixed sum model. This sort of random model is similar to one in which species passively colonize islands or "targets" of different area, as in Coleman et al.'s (1982) random placement model for species richness. Similar reason-

TABLE 1. Summary of four co-occurrence indices.

	Index			
	CHECKER	C score	V ratio	COMBO
Description	Number of species pairs forming perfect checkerboard distributions	Checkerboard score	Variance ratio	Number of unique species combinations
Calculation	Scan matrix rows for species pairs forming checkerboards	$\Sigma(S_i - Q)(S_k - Q)/((R)(R-1)/2)$	σ^2(column sums)/Σ row σ^2	Scan matrix columns for unique species combinations
Source	Diamond (1975)	Stone and Roberts (1990)	Robson (1972); Schluter (1984)	Pielou and Pielou (1968)
Theoretical range	0 to $R(R-1)/2$	0 to $\Sigma S_i S_k/((R)(R-1)/2)$	0 to ∞	1 to 2^R
Pattern expected in a competitively structured community	Observed > simulated	Observed > simulated	Observed < simulated	Observed < simulated
Comments	Most readily testable prediction of Diamond's (1975) assembly rules	Measures species segregation, but does not require perfect checkerboard distributions	Measures pattern in marginal totals of matrix	May reflect "forbidden species combinations" (Diamond 1975)

Notes: S_i = total for row i; R = number of rows (=species) in the matrix; Q = number of sites in which both members of a species pair are present.

ing could be applied to the rows of the species occurrence matrix, which represent occurrence frequencies of each species. The null model could retain observed row totals, allow rows to be equiprobable, or set probabilities proportional to observed row totals.

Thus, a universe of $3^2 = 9$ simple null models (SIM1 to SIM9) could be constructed using only information contained in the row and column totals. These models differ in whether rows and columns are treated as fixed sums, equiprobable, or proportional. All of the algorithms are plausible candidates for a null model, and some have been used in other analyses in biogeography and ecology (Table 2). Table 3 gives a hypothetical data matrix, and Fig. 1 illustrates the simulation procedure for each algorithm with this matrix. I have systematically compared the behavior of all nine algorithms with the four different indices.

Kinds of data

Two types of data sets are used in co-occurrence analysis. The first type I call "island lists." These represent nearly complete lists of species from islands or well-defined habitat patches. The lists are often accumulated from many thorough censuses, and can be found in the literature for well-studied vertebrate taxa and "popular" invertebrate taxa such as butterflies. Island lists are typically nondegenerate matrices, because missing species and empty islands are often not recorded or reported.

The second type of data set I call "sample lists." These represent lists of species taken from standardized samples within areas of relatively homogeneous habitat. These lists are generated from short-term ecolog-

ical censuses, such as pitfall traps, cores, bait samples, sweep samples, line, belt, and quadrat samples, point counts, and timed censuses. Often these lists are incomplete, and may not include rare or undersampled species, so that total species richness must be estimated by extrapolation (Colwell and Coddington 1994). Sample lists often form degenerate matrices, because some samples may contain no species.

Fig. 2 is an example of a matrix based on island lists: the 17 species of finches (Fringillidae) from 19 of the largest islands in the West Indies. The islands have been well censused for the past century, although extensive habitat change and species extinctions have been recorded. The matrix contains a range of species occurrence frequencies from single-island endemics (*Loxia leucoptera*) to widespread species (*Tiara bicolor*). Similarly, the total number of finch species on the islands ranges from 1 (Antigua) to 7 (Hispaniola). Total bird species richness on these islands ranges from 16 (St. Martin, Barbados) to 79 (Cuba, Hispaniola).

Fig. 3 is an example of a matrix based on sample lists: the species of ground-foraging ants collected in 25 pitfall trap samples in an open field in Prince Edward County, Virginia. These data are part of a regional census of ants at 33 sites in the eastern United States. At each site, 25 pitfall traps (50-mL plastic centrifuge tubes, 27 mm diameter) were established in a 5 × 5m grid. Traps were opened for 48 h, and the ants from each trap were identified to species when possible (details in Arnett 1998). Most of the species in the collection occurred in only a single sample, and 13 of the 25 traps contained no ants. The most common taxon (*Aphaenogaster rudis* complex) occurred in 5 of the

TABLE 2. Summary of nine null model algorithms for species co-occurrence analysis.

Constraint	Constraint		
	Columns equiprobable	Columns proportional	Column sums fixed
Rows equiprobable	SIM1 $P(a_{ij}) = 1/RC$ All species and sites are equiprobable. Classic randomization test in which all matrix rearrangements are equally likely. Sokal and Rohlf (1995:803)	SIM6 $P(a_{ij}) = T_j/NR$ All species are equiprobable. Sites differ in suitability. Probabilities of occurrence in sites proportional to observed species richness/site. Inverse of SIM7.	SIM3 $P(a_{ij}) = 1/R$ All species are equiprobable. Species number per site fixed. Inverse of SIM2. Johnson (1974), Simberloff (1978), Patterson and Atmar (1986)
Rows proportional	SIM7 $P(a_{ij}) = S_i/NC$ All sites are equiprobable. Species differ in occurrence. Probabilities of occurrence of species proportional to observed species occurrence frequencies. Inverse of SIM6.	SIM8 $P(a_{ij}) = S_iT_j/N^2$ Both species and sites differ in suitability. Probabilities of occurrence are conditional on both site and species marginal totals. Gilpin and Diamond (1982), Gotelli and Graves (1996)	SIM5 $P(a_{ij}) = S_i/N$ Species richness/site fixed. Species differ in occurrence. Probabilities of occurrence of species proportional to observed species occurrence frequencies. Inverse of SIM4. Abele and Patton (1976), Connor and Simberloff (1978), Patterson and Atmar (1986)
Row sums fixed	SIM2 $P(a_{ij}) = 1/C$ Species occurrence totals fixed. All sites are equiprobable. Inverse of SIM3. Sale (1974), Inger and Colwell (1977), Winemiller and Pianka (1990)	SIM4 $P(a_{ij}) = T_j/N$ Species occurrence totals fixed. Sites differ in suitability. Probabilities of occurrence in sites proportional to observed species richness/site. Inverse of SIM5. Coleman et al. (1982), Graves and Gotelli (1993)	SIM9 $P(a_{ij}) = $ [not applicable] Cannot be simulated by filling an empty matrix. Row and column sums simultaneously maintained. Degenerate matrices not possible. Connor and Simberloff (1979)

Notes: Each entry gives the abbreviation for the model, a formula for calculating the probability of cell occupancy for the first cell in the matrix [$P(a_{ij})$; see Fig. 1], a description of the algorithm, and citations of other studies in community ecology or biogeography that have used this model or one very similar. N = matrix total, R = number of rows, S_i = total for row i; C = number of columns; T_j = total for column j.

traps, and species richness in the traps ranged from only 1 to 3 species.

Both the finch matrix and the ant matrix were tested with all combinations of algorithms and indices. I also used the dimensions and marginal totals of the finch matrix to construct random matrices for Type I error analysis.

The special problem of degenerate matrices

All of the algorithms except the Connor and Simberloff (1979) procedure (SIM9; see Table 2) can potentially create degenerate matrices, that is, matrices in which one or more rows or columns is empty. Bi-

ologically, these correspond to islands with no species and species that occur on no islands. How should these degenerate matrices be handled? Connor and Simberloff (1983) warned that these matrices might be problematic for null model comparisons, and for this reason they advocated the use of SIM9. On the other hand, it is biologically realistic to imagine that degenerate matrices might arise during random colonization. Haefner's (1988) analyses suggest that degenerate matrices may not greatly affect the analysis. Although I do not report the results in detail here, I examined the statistical properties of both degenerate and nondegenerate matrices to understand how they might affect null mod-

TABLE 3. Hypothetical presence–absence matrix used in null model analysis of species co-occurrence patterns.

Species	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Row total (S_i)
Species A	1	1	0	1	0	0	3
Species B	0	1	0	0	0	0	1
Species C	0	0	0	0	1	1	2
Species D	0	1	0	0	0	0	1
Species E	1	1	1	0	0	0	3
Column total (T_j)	2	4	1	1	1	1	10 Grand total (N)

Note: Fig. 1 illustrates simulations procedures with this matrix.

COLUMNS EQUIPROBABLE

SIM1

		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	
Rows Equiprobable	Species A	0.033	0.033	0.033	0.033	0.033	0.033	
	Species B	0.033	0.033	0.033	0.033	0.033	0.033	
	Species C	0.033	0.033	0.033	0.033	0.033	0.033	
	Species D	0.033	0.033	0.033	0.033	0.033	0.033	
	Species E	0.033	0.033	0.033	0.033	0.033	0.033	(10)

Grand Total (*N*)

SIM7

		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	
Rows Proportional	Species A	0.050	0.050	0.050	0.050	0.050	0.050	
	Species B	0.017	0.017	0.017	0.017	0.017	0.017	
	Species C	0.033	0.033	0.033	0.033	0.033	0.033	
	Species D	0.017	0.017	0.017	0.017	0.017	0.017	
	Species E	0.050	0.050	0.050	0.050	0.050	0.050	
		0.167	0.167	0.167	0.167	0.167	0.167	(10)

Grand Total (*N*)

SIM2

		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Row Total (S_i)
Row Sums Fixed	Species A	0.167	0.167	0.167	0.167	0.167	0.167	(3)
	Species B	0.167	0.167	0.167	0.167	0.167	0.167	(1)
	Species C	0.167	0.167	0.167	0.167	0.167	0.167	(2)
	Species D	0.167	0.167	0.167	0.167	0.167	0.167	(1)
	Species E	0.167	0.167	0.167	0.167	0.167	0.167	(3)

COLUMNS PROPORTIONAL

SIM6

		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	
Rows Equiprobable	Species A	0.040	0.080	0.020	0.020	0.020	0.020	
	Species B	0.040	0.080	0.020	0.020	0.020	0.020	
	Species C	0.040	0.080	0.020	0.020	0.020	0.020	
	Species D	0.040	0.080	0.020	0.020	0.020	0.020	
	Species E	0.040	0.080	0.020	0.020	0.020	0.020	(10)

Grand Total (*N*)

SIM8

		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	
Rows Proportional	Species A	0.060	0.120	0.030	0.030	0.030	0.030	
	Species B	0.020	0.040	0.010	0.010	0.010	0.010	
	Species C	0.040	0.080	0.020	0.020	0.020	0.020	
	Species D	0.020	0.040	0.010	0.010	0.010	0.010	
	Species E	0.060	0.120	0.030	0.030	0.030	0.030	(10)

Grand Total (*N*)

SIM4

		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Row Total (S_i)
Row Sums Fixed	Species A	0.200	0.400	0.100	0.100	0.100	0.100	(3)
	Species B	0.200	0.400	0.100	0.100	0.100	0.100	(1)
	Species C	0.200	0.400	0.100	0.100	0.100	0.100	(2)
	Species D	0.200	0.400	0.100	0.100	0.100	0.100	(1)
	Species E	0.200	0.400	0.100	0.100	0.100	0.100	(3)

FIG. 1. Probability calculations for nine null model algorithms (SIM1–SIM9). The algorithm is defined by whether rows and columns are treated as equiprobable, proportional to observed totals, or fixed sums. For each simulation, the circled values indicate marginal totals that are used to randomly fill the matrix. Each cell entry is the probability of occurrence for the first cell that is filled in the matrix. Probability values were calculated from the hypothetical data set in Table 3, using the formulae in Table 2. No entries are shown in the cells of SIM9, because this model requires maintenance of both row and column totals, and is created by a transposition algorithm (see the discussion in *Simulation procedures* for details).

COLUMN TOTALS FIXED

SIM3

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Species A	0.200	0.200	0.200	0.200	0.200	0.200
Species B	0.200	0.200	0.200	0.200	0.200	0.200
Species C	0.200	0.200	0.200	0.200	0.200	0.200
Species D	0.200	0.200	0.200	0.200	0.200	0.200
Species E	0.200	0.200	0.200	0.200	0.200	0.200
Column Total (T_i)	2	4	1	1	1	1

Rows Equiprobable

SIM5

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Species A	0.300	0.300	0.300	0.300	0.300	0.300
Species B	0.100	0.100	0.100	0.100	0.100	0.100
Species C	0.200	0.200	0.200	0.200	0.200	0.200
Species D	0.100	0.100	0.100	0.100	0.100	0.100
Species E	0.300	0.300	0.300	0.300	0.300	0.300
Column Total (T_i)	2	4	1	1	1	1

Rows Proportional

SIM9

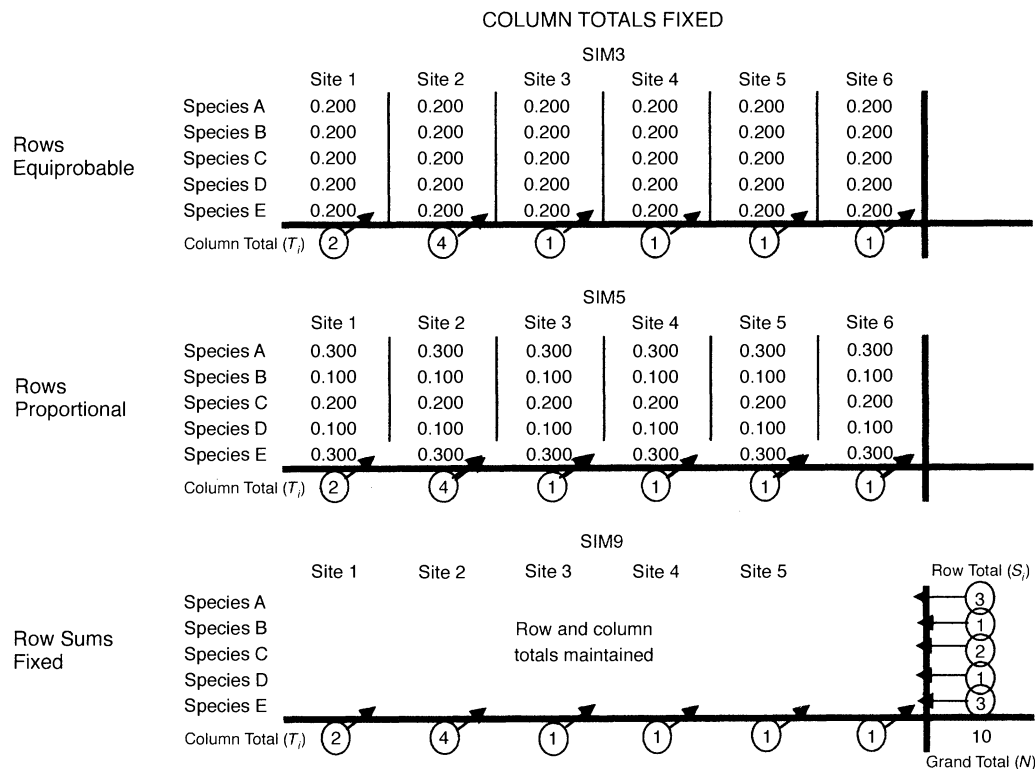|	Site 1	Site 2	Site 3	Site 4	Site 5	Row Total (S_i)
Species A						3
Species B		Row and column				1
Species C		totals maintained				2
Species D						1
Species E						3
Column Total (T_i)	2	4	1	1	1	1	10

Row Sums Fixed

Grand Total (N)

Fig. 1. Continued.

el tests. The general result was that degenerate matrices did increase the frequency with which the null hypothesis was rejected. However, for well-behaved algorithms, this increase was usually <10%. The effect of degenerate matrices was usually much less than the effect of the algorithm or index selected.

Estimating Type I errors

Estimating Type I errors means using each algorithm and index to statistically evaluate a "random" test matrix, which is presumably random with respect to species interactions. For each combination of index and algorithm, I created 100 such test matrices, and kept track of the 100 upper- and lower-tail probabilities for each test. If the test is robust to Type I error, ~10 of the 100 test matrices should have been nonrandom at $P < 0.05$ (in either tail). On the other hand, if the test is prone to Type I error, the number of times the null hypothesis is rejected will be much greater than 10. Algorithms or indices that are prone to Type I errors should not be used because there is a danger that we will incorrectly reject the null hypothesis for a data set that is random.

How, exactly, should a "random" matrix be constructed for such a test? I used four procedures to create different kinds of test matrices. Each matrix was created using the observed marginal totals from the West Indian finch matrix (Fig. 2). The four kinds of test matrices were:

Test 1—Complete randomization. The 55 species occurrences were completely randomized across the 19 × 17 = 323 cells of the matrix. This is equivalent to SIM1.

Test 2—Randomize each row, columns equiprobable. The species occurrences in each row were randomized among the 19 sites. This is equivalent to SIM2.

Test 3—Randomize each row, columns proportional. The species occurrences in each row were randomized among the 19 sites, with the probability of occurrence being proportional to the column total in the matrix. This is equivalent to SIM4.

Test 4—Randomize each row, columns proportional to the logarithm of island areas. The species occurrences in each row were randomized among the 19 sites, with the probability of occurrence proportional to the logarithm of island area.

Thus, in the first three kinds of test matrices, there is an identical simulation procedure that is used to evaluate the matrix. Test 4 incorporates independent data on island areas in order to construct test matrices. Only nondegenerate test matrices were used in these analyses. Once each test matrix was created, it was evaluated statistically with 1000 randomizations of each of the 36 combinations of algorithms (9) and indices (4). I repeated this process for 100 test matrices of each combination. The computer program kept track of the number of times the null hypothesis was rejected (either

Species	Cuba	Hispaniola	Jamaica	Puerto Rico	Guadeloupe	Martinique	Dominica	St. Lucia	Barbados	St. Vincent	Grenada	Antigua	St. Croix	Grand Cayman	St. Kitts	Barbuda	MonSt.errat	St. Martin	St. Thomas	Row Sum
Carduelis dominicensis	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxia leucoptera	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Volatinia jacarina	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Sprophila nigricollis	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Melopyrrha nigra	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2
Loxigilla portoricensis	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxigilla violacea	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Loxigilla noxis	0	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	12
Melanospiza richardsoni	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Tiara olivacea	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	5
Tiara bicolor	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	17
Tiara canora	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxipasser anoxanthus	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Saltator albicollis	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4
Torreornis inexpectata	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ammodramus savannarum	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Zonotrichia capensis	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Column Sum	4	7	5	4	3	3	3	4	2	2	4	2	1	2	2	2	2	2	1	55

FIG. 2. An example of an "islands list" matrix of West Indian finches (Fringillidae). Data from Gotelli and Abele (1982).

tail) and the average, or expected value of the index for each simulation. Because there is no a priori way to decide which of the four kinds of random matrices (Test 1 through Test 4) are most valid, I averaged the results across all four in order to compare the different simulation algorithms (SIM1 through SIM9) and different co-occurrence indices (CHECKER through COMBO).

Estimating Type II errors

Estimating Type II errors means using each algorithm and index to statistically evaluate a nonrandom test matrix. Such a nonrandom matrix should reflect the effects of species interactions. A test that was vulnerable to Type II error would fail to detect nonrandom patterns in such a structured matrix. How should such a nonrandom matrix be constructed? One strategy would be to build a specific model of species inter-

actions and use this model to create the test matrix. Examples of this strategy include Case and Sidell (1983), Colwell and Winkler (1984), Hastings (1987), and Kelt et al. (1995).

However, a weakness of this approach is that the results are always contingent on the particular model chosen for comparison. In addition, it may not represent a fair test of the null model. As some critics have pointed out (Roughgarden 1983), the null model does not explicitly describe the colonization and extinction of a noninteractive fauna (but see Caswell 1976), so perhaps it should not be evaluated against an explicit competition model. Rather, the null model can be evaluated by using a matrix that has a *pattern* that is expected to result from competition, without specifying a mathematical model to create this pattern. Diamond and Gilpin (1982) give two such examples of nonrandom matrices (Fig. 4) that they believe would reflect patterns

Species	Pitfall Trap Number																									Row sum
	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	
Aphaenogaster rudis complex	1	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5
Aphaenogaster treatae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	3
Camponotus castaneus	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Crematogaster lineolata	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Formica inserta	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Formica subserecia	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Myrmecina americana	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Pheidole morrisi	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Pheidole vinelandica	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Smithistruma sp	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Solenopsis texana	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Column Sum	2	0	1	0	2	1	0	3	0	3	1	1	0	0	1	1	1	0	0	0	0	0	1	0	0	18

FIG. 3. An example of a "sample list" matrix of Virginia ants. Source: unpublished data from A. E. Arnett and N. J. Gotelli (see Arnett 1998).

caused by interspecific interactions. Each matrix is 20 × 20. The first contains 100 of 190 possible species pairs forming perfect checkerboards (Fig. 4A), and the second contains 10 of 190 pairs in a perfect checkerboard arrangement (Fig. 4B). Stone and Roberts (1990) used these matrices to evaluate their *C* score index, and I follow their lead in these analyses.

However, with these matrices, the null hypothesis is rejected at $P < 0.001$ for nearly all of the algorithms and indices. These two matrices are so extreme that they do not reveal the relative power of the different tests. Moreover, real data sets will never have distributions this extreme, which limits the usefulness of the comparison. To overcome these problems, I used a "noise test" to evaluate the power of different null models (Gotelli et al. 1997).

The noise test can be understood with an analogy (suggested by Gary Entsminger). Imagine that I open a new deck of playing cards, in which the cards are arranged in numerical order within each suit. Next, I randomly choose two cards in the deck and switch them. If I showed you this deck, you would immediately recognize that it was unshuffled. In fact, I could probably transpose many such pairs of cards and you would still have no trouble detecting the original ordering. However, if I continued to switch cards, at some point you could no longer distinguish the original ordering, and you would conclude that the deck had been previously shuffled. Eventually, the "noise" from the transpositions will have swamped the "signal" of the original ordering.

When applied to co-occurrence matrices, repeated random swapping of cell contents reduces the effect size in the matrix. For power tests of the different algorithms, I began with Diamond and Gilpin's (1982) perfect checkerboard matrix (Fig. 4A), and then began by swapping the contents of two randomly chosen cells within each row of the matrix. After the two cells were swapped, I tested the new matrix against each of the null models, and recorded the tail probabilities. Next, I swapped an additional pair of cells in the matrix, and repeated the test. This sequence continued until a total of 20 pairs of cells in each row had been swapped. At that point, the matrix had been entirely randomized, following the procedure of SIM2. I created five such sequences of random matrices and recorded the average tail probability at each of the 10 noise levels. To evaluate the power of the test, I plotted the average tail probability against the number of cells swapped. I used a one-tailed test for the pattern that would be expected with competition (Table 1).

What should the plot of tail probability vs. noise level look like? If the test is prone to Type II error, the curve would rise very quickly above the 0.05 rejection level. In other words, it would take very little noise to obscure the statistical significance of the pattern. Conversely, the test is prone to Type I error if the curve never rises above the 0.05 mark, even at very high

A

```
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
```

B

```
0 0 1 0 1 0 1 1 1 1 0 1 0 0 0 0 1 1 1 0
1 1 0 1 0 1 0 0 0 0 1 0 1 1 1 1 0 0 0 1
0 0 1 0 1 1 1 1 0 0 1 0 0 1 0 0 1 1 1 0
1 1 0 1 0 0 0 0 1 1 0 1 1 0 1 1 0 0 0 1
1 0 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 0 1 1
0 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 1 0 0
1 0 0 1 1 1 0 1 0 1 1 0 0 1 0 0 1 1 0 0
0 1 1 0 0 0 1 0 1 0 0 1 1 0 1 1 0 0 1 1
0 0 1 0 1 0 1 1 1 0 0 1 0 0 1 1 0 1 1 0
1 1 0 1 0 1 0 0 0 1 1 0 1 1 0 0 1 0 0 1
1 1 0 1 0 1 1 1 1 0 0 0 0 1 0 1 1 0 0 0
0 0 1 0 1 0 0 0 0 1 1 1 1 0 1 0 0 1 1 1
1 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 1 1 1 0
0 1 1 1 1 0 0 0 0 1 1 1 0 0 1 1 0 0 0 1
1 1 0 0 1 1 0 0 1 0 0 1 1 1 0 0 0 1 1 0
0 0 1 1 0 0 1 1 0 1 1 0 0 0 1 1 1 0 0 1
0 0 0 1 0 1 0 1 1 0 1 0 1 1 1 0 1 0 1 0
1 1 1 0 1 0 1 0 0 1 0 1 0 0 0 1 0 1 0 1
0 1 1 0 1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1
1 0 0 1 0 0 1 0 1 0 1 1 1 0 1 0 1 0 1 0
```

FIG. 4. Two highly structured presence–absence matrices. Diamond and Gilpin (1982) presented these matrices as examples of the sort of pattern that should be revealed as nonrandom by a null model test. Each row is a species and each column is a site. The matrix on the left (**A**) was used as the basis for the Type II "noise test" illustrated in Fig. 6.

levels of noise. Such a test will continue to reject the null hypotheses even though the pattern in the data had been entirely obliterated by noise. A desirable profile would be one in which the test statistic is well below the 0.05 mark for low levels of noise, then well above the 0.05 mark for high levels of noise (Figure 5).

Comparisons of different indices

Because they all quantify "co-occurrence," analyses based on these four indices (CHECKER, C score, V ratio, and COMBO) might produce similar results. To evaluate the extent to which each index measures a different type of pattern, I began with the West Indian finch matrix and started transposing submatrices according to the recipe for SIM9. I carried out 10 000 such transpositions, and retained the one matrix that generated the greatest score for CHECKER. Next, I tested this matrix, using SIM9, with the C score and COMBO indices. I repeated this procedure of generating an extreme matrix for each index and testing it against the other three indices. If the four indices are

measuring similar sorts of patterns, then an extreme matrix for one should be extreme for the other three. On the other hand, if the indices detect different kinds of patterns, a matrix that is extreme for one index will not generate a significant result using other indices. The V ratio was not used in this analysis because retaining row and column totals as in SIM9 does not ever change the index, so all of the null matrices would have the same V ratio as the original matrix.

Simulation procedures

I used three different strategies for constructing randomized matrices:

1) "Reshuffling." For SIM1, SIM2, and SIM3, I began with the original input matrix and reshuffled species occurrences within rows (SIM2), columns (SIM3), or across the entire matrix (SIM1).

2) "Filling." For SIM4 through SIM8, I started with an empty matrix, and then began filling rows (SIM4), columns (SIM5), or the entire matrix (SIM6 through SIM8), with cell probabilities being either equiprobable

TABLE 4. Summary of Type I error tests for null models.

Simulation algorithm	Co-occurrence index				
	CHECKER	C score	V ratio	COMBO	Average
SIM1	0.72	0.76	0.14	0.34	0.49
SIM2	**0.07**	**0.08**	**0.10**	**0.00**	**0.06**
SIM3	0.77	0.77	0.77	0.49	0.70
SIM4	**0.16**	**0.27**	**0.10**	**0.02**	**0.14**
SIM5	0.74	0.76	0.62	0.16	0.57
SIM6	0.89	0.73	0.12	0.11	0.43
SIM7	0.77	0.73	0.12	0.11	0.43
SIM8	0.91	0.56	0.08	0.06	0.40
SIM9	**0.08**	**0.10**	**n.a.**	**0.01**	**0.06**
Average	0.57	0.53	0.26	0.15	

Notes: Each entry is the average proportion of simulations for which the null hypothesis was rejected ($P < 0.05$; two-tailed) when tested against a random matrix. Boldface rows indicate simulation algorithms for which the average proportion was <0.15. By chance, the expected proportion is 0.10. SIM9 is not applicable (n.a.) to the V ratio. See Table 2 for details of simulation algorithms, and Table 1 for details of co-occurrence indices.

or proportional to row and/or column totals. Different cell probabilities were calculated with a subroutine that selected a random point along a number line. The probabilities of occurrence corresponded to segments of different length along the number line. If a cell was already occupied, another random placement was made along the number line until an empty cell was selected.

3) "Transposing." For SIM9, I used a modified version of Manly's (1995) method. This method involves transposing randomly chosen submatrices of the form:

$$\begin{bmatrix} \cdots & 1 & \cdots & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & \cdots & 1 & \cdots \end{bmatrix} \text{ to } \begin{bmatrix} \cdots & 0 & \cdots & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 1 & \cdots & 0 & \cdots \end{bmatrix}.$$

Submatrices do not have to be in physically adjacent rows or columns. This transposition results in a new matrix, but one that retains row and column totals of the original matrix. I first used 1000 transpositions to remove transient effects and make sure that the randomized matrix was very different from the original. Next, I retained *each consecutive matrix* produced by a single swapping of submatrices. Manly (1995) established that co-occurrence indices derived from sequences of these matrices are independent with respect to the original matrix. I checked this procedure against results presented by Stone and Roberts (1990), who used two other methods for creating random matrices with fixed row and column totals, and was able to generate identical results when testing the same matrix.

All of the algorithms described in this paper are available in a user-friendly application of null models software, EcoSim (Gotelli and Entsminger 1999).

RESULTS

Type I errors

Table 4 summarizes the average proportion of simulations for which the null hypothesis was rejected

when it was tested against a random matrix. If the test is not prone to Type I error, the null hypothesis should have been rejected ~10% of the time (5% in each tail). There were strong, consistent differences in the error frequencies of the different algorithms. Only SIM2 and SIM9 had average error frequencies of <10%, although SIM4 also performed reasonably well, with an average error frequency of 14%. SIM3 and SIM5 had the worst performance, with the null hypothesis being rejected >50% of the time for randomly constructed matrices. The other algorithms rejected the null hypothesis <50% of the time, but they were still well above the expected error frequency of 10%.

Comparing the four indices on the basis of Type I errors is difficult because of the variation among algorithms. Overall, COMBO and the V ratio had lower Type I error frequencies than CHECKER or the C score. For each index, there were at least two algorithms that had acceptable Type I error frequencies.

Type II errors

Fig. 6 illustrates the effect of adding progressive amounts of noise to the perfect matrix in Fig. 4A. As more and more pairs of elements in each row of the matrix are randomly reshuffled, the statistically significant pattern breaks down, and the P value rises above 0.05. If the P value rises very quickly with increasing noise, the test is prone to Type II error, because only a slight rearrangement of the data renders the test nonsignificant. Conversely, some combinations of indices and algorithms were prone to Type I error, and these curves never rose above the 0.05 level, even when the entire matrix had been randomly reshuffled.

Of the four indices, COMBO and CHECKER were most prone to Type II error, and in many simulations, the P value curves immediately rose above the 0.05 level. The least sensitive index was the C score, although this index was prone to Type I error for some of the algorithms. The V ratio had an intermediate response, usually rising above the curve for the C score but falling below the curves for COMBO and CHECKER.

Comparisons of the nine algorithms are less clear cut. COMBO and CHECKER were sensitive to noise for all the algorithms, whereas C and V seemed to perform best with SIM2, SIM4, and SIM9. However, other combinations also gave acceptable results. Very few of the trials generated the idealized noise curve of Fig. 5. Overall, the best combination was using the C score with a randomization that held row and column totals fixed (SIM9). This test generated statistically significant results until approximately half of the cell values in the matrix had been randomly rearranged, then gave nonsignificant results as more noise was added.

Performance of different indices

Table 5 illustrates the effect of using different indices to evaluate extreme patterns. In this table, the original
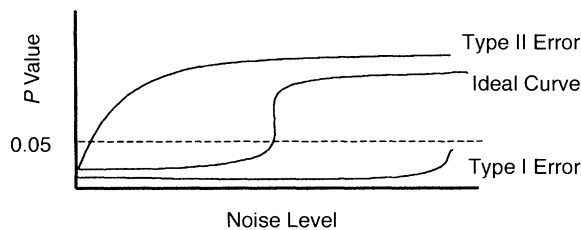
FIG. 5. Hypothetical error curves for a Type II noise test. As more and more site occurrences of species are randomized in the matrix, the P value will rise from significant to non-significant levels. A test is susceptible to Type I error if it never rises above the 0.05 mark, even when the matrix is entirely randomized. A test is susceptible to Type II error if it rises too quickly above the 0.05 mark, even when very little of the original pattern has been degraded by random transpositions. A test has a good balance of Type I and Type II errors if it stays below the 0.05 mark until approximately half of the matrix has been randomized, then quickly rises well above the 0.05 mark as more noise is added.

West Indian finch matrix was rearranged using SIM9, and the most extreme of 10 000 matrices was retained. For example, using the number of checkerboards as the index, the computer generated one matrix out of 10 000 that had 98 species pairs in perfect checkerboards. The average number of checkerboards for all the simulated matrices was only 89.73. However, when this extreme matrix was then tested with the C score and COMBO, the patterns for these indices were nonsignificant. The same result was obtained when extreme matrices were first generated for the C score and COMBO, and then tested with the other two indices. Only the extreme matrix generated for COMBO gave a similar result when tested with the C score. Thus, although all the indices measure species co-occurrence, they reveal different aspects of co-occurrence pattern, and do not always give congruent results.

Empirical comparisons

Table 6 illustrates the results of testing the original West Indian finch matrix against all nine algorithms and four co-occurrence indices. The number of species combinations in this matrix is 10, which was significantly less than expected for all of the simulation algorithms. In contrast, the observed V ratio of 1.23 was never less than expected by chance. The C score was greater than expected for SIM2, SIM4, and SIM9. These are the three algorithms that were not prone to Type I errors (false positives; Table 4). There were 91 species pairs that formed perfect checkerboards (CHECKER), and this was greater than expected for SIM1, SIM3, SIM5, SIM6, SIM7, and SIM8. However, these algorithms are all very prone to Type I errors (Table 4) when used with CHECKER. Overall, there is good evidence that there were fewer species combinations than expected by chance in the West Indian finch matrix, and that the average number of checkerboard units per species pair was greater than ex-

TABLE 5. Extreme matrix analysis.

Extreme matrix tested with:	Extreme matrix generated for:		
	CHECKER	C score	COMBO
CHECKER	**98** [89.73] (0.0001)	86 [89.68] (0.986)	91 [89.18] (0.258)
C score	2.65 [2.73] (0.670)	**3.98** [2.77] (0.0001)	3.79 [2.77] (0.001)
COMBO	17 [15.50] (0.987)	14 [15.28] (0.219)	**12** [15.23] (0.0001)

Notes: Each column represents a different co-occurrence index that was used to generate an extreme matrix with SIM9. Each row represents a co-occurrence index that was used to test the matrix generated for each column. The first entry is the observed co-occurrence index. The entry in square brackets is the expected value of the index from 1000 simulations (SIM9). The entry in parentheses is the tail probability for this pattern. Diagonal entries (in boldface) represent the original extreme matrix that was generated by 10 000 random transpositions using SIM9.

pected. Both patterns are consistent with the hypothesis that local coexistence of congeneric species of finches is limited (Terborgh 1973; but see Gotelli and Abele 1982). Although species pairs showed less co-occurrence on islands than expected by chance as measured by the C score, there was no tendency for species to form perfect checkerboard distributions, as in Diamond (1975) or Graves and Gotelli (1993).

TABLE 6. Results of null model analyses of the West Indian finches co-occurrence matrix (Fig. 2).

	CHECKER	C score	V ratio	COMBO
Observed	91	3.79	1.23	10
SIM1	**71.23** (0.001)	7.21 (1.000)	0.69 (0.957)	**18.56** (<0.001)
SIM2	94.86 (0.873)	**2.93** (<0.001)	0.76 (0.932)	**17.43** (<0.001)
SIM3	**71.66** (<0.001)	7.30 (1.000)	0.76 (1.000)	**18.75** (<0.001)
SIM4	90.10 (0.552)	**2.52** (<0.001)	1.14 (0.622)	**15.91** (<0.001)
SIM5	**53.09** (0.008)	5.76 (0.983)	1.08 (0.914)	**17.48** (<0.001)
SIM6	**61.85** (<0.001)	6.57 (1.000)	1.28 (0.520)	**17.94** (<0.001)
SIM7	**53.71** (0.007)	5.74 (0.974)	0.78 (0.915)	**17.39** (<0.001)
SIM8	**51.67** (0.004)	5.11 (0.920)	1.25 (0.585)	**17.16** (0.002)
SIM9	89.30 (0.239)	**2.75** (<0.001)	n.a.	**15.12** (0.001)

Notes: Each column is a different co-occurrence index and each row is a different simulation algorithm (see Tables 1 and 2 for details). The first row is the observed co-occurrence index for the matrix in Fig. 2. The other rows give the expected index, averaged over 1000 simulations of the algorithm. The P value for a one-tailed test is given in parentheses. Tail probabilities <0.05 are indicated in boldface. The V ratio cannot be used with SIM9 (n.a.) because it maintains row and column sums (see the discussion in *The variance ratio as an index of nonrandomness* for details).

TABLE 7. Results of null model analyses of Virginia ants co-occurrence matrix (Fig. 3). Entries are as in Table 6.

	CHECKER	C score	V ratio	COMBO
Observed	48	2.07	0.98	10
SIM1	**32.35**	3.31	0.70	11.24
	(0.050)	(0.993)	(0.865)	(0.230)
SIM2	49.58	2.15	0.70	11.39
	(0.860)	(0.682)	(0.878)	(0.195)
SIM3	**30.04**	3.04	0.98	10.81
	(0.014)	(0.994)	(0.569)	(0.353)
SIM4	**41.89**	**1.67**	**1.71**	9.40
	(0.014)	**(0.007)**	**(0.034)**	(0.861)
SIM5	**23.52**	3.35	1.03	10.19
	(0.003)	(0.988)	(0.148)	(0.608)
SIM6	**26.05**	2.39	**1.73**	9.45
	(0.001)	(0.704)	**(0.029)**	(0.859)
SIM7	**24.92**	3.83	0.71	10.37
	(0.012)	(0.993)	(0.852)	(0.559)
SIM8	**20.37**	2.56	**1.72**	9.23
	(<0.001)	(0.767)	**(0.033)**	(0.888)
SIM9	47.18	2.00	n.a.	10.71
	(0.175)	(0.199)		(0.406)

Table 7 illustrates the results of testing the Virginia ant matrix against all nine algorithms and four co-occurrence indices. In contrast to the results of the West Indian finch analysis, few of these tests are statistically significant for well-behaved algorithms. Ten species combinations were observed in this matrix, which was not statistically significant for any of the algorithms. There were more checkerboard species pairs (48) than expected compared with SIM1, and SIM3 to SIM8. However, these algorithms are prone to Type I error for this index (Table 1). The C score was significantly greater than expected only when compared to SIM4, which is also prone to Type I error for this index (Table 1). The only evidence for nonrandomness was in the V ratio, which was significantly less than expected for SIM4, SIM6, and SIM8. However, all of these models assume that there is variation in site quality, so that probabilities of occurrence are proportional to species richness totals for each site. Thus, "empty sites" in the original matrix will not be filled in these simulations. A more appropriate null model for the ant data would be SIM2, in which species occurrences are fixed and all sites are equiprobable. None of the co-occurrence indices showed a significant pattern with SIM2. Overall, there is little evidence that the small-scale co-occurrence of ant foragers in pitfall traps was nonrandom.

DISCUSSION

Choosing the right algorithm

Although all nine algorithms are logically plausible, the analyses reveal that many of the algorithms would be unacceptable choices because they are very prone to Type I error, and would be expected to produce false positives with data sets that have little or no real structure. However, three algorithms consistently had low probabilities of Type I errors when compared with a variety of random data matrices: SIM2, SIM4, and SIM9 (Table 4). These algorithms share in common the property that they all maintain the observed row totals, that is the species occurrence frequencies. The algorithms differ in how the columns (=sites) are treated. In SIM2, the sites are equiprobable, whereas in SIM4, the probability that a species occurs in a site is proportional to the column total for that site. Finally, SIM9 maintains the observed number of species in a site. Conceptually, these algorithms are satisfying, because they correspond to a colonization model in which species colonize an archipelago randomly with respect to one another. Note that the colonization of each species is not "random" with respect to the sites: occurrence frequencies are maintained for each species, and, in SIM4 and SIM9, differences among islands are maintained. But the species occurrences are random with respect to one another, which is an appropriate null model for detecting patterns caused by species interactions.

Critics have pointed out that extinctions are an important outcome of species interactions, so that incorporating species occurrence frequencies may "smuggle in" species interactions in the null model. Although this may be true, my analysis of Type I errors suggests that allowing species occurrence frequencies to vary may generate false positives in a null model test. This conclusion is reinforced by other debates in the null model literature. For example, Gilpin and Diamond's (1982) null model operates on the same principle as SIM8, which allows species occurrence frequencies to vary. Wilson (1987) showed that this model rejects the null hypothesis for random data sets constructed by the recipe for TEST1, although Gilpin and Diamond (1987) contend that Wilson (1987) did not correctly implement their model.

In an analysis of species co-occurrence as measured by "favored states" analysis, Fox and Brown (1993) used a null model that implicitly assumed species occurrence frequencies were equiprobable. Stone et al. (1996) and Wilson (1995) reanalyzed the data and adjusted species occurrences on the basis of observed frequencies and species geographic ranges. In these reanalyses, the co-occurrence patterns were no longer statistically significant.

Because the co-occurrence tests are very sensitive to variation in species occurrence frequencies, row totals should be preserved as a constraint in the null model. On the other hand, the results were surprisingly insensitive to variation in column totals (=number of species per site), so this constraint should be modified to reflect sampling methods or variation in site quality. Specifically, SIM9 seems most appropriate for analyzing "island lists," especially for classic archipelago data in which there are strong species–area effects. On the other hand, SIM2 seems most appropriate for analyzing "sample lists," particularly when comparing standardized samples that have been collected in areas
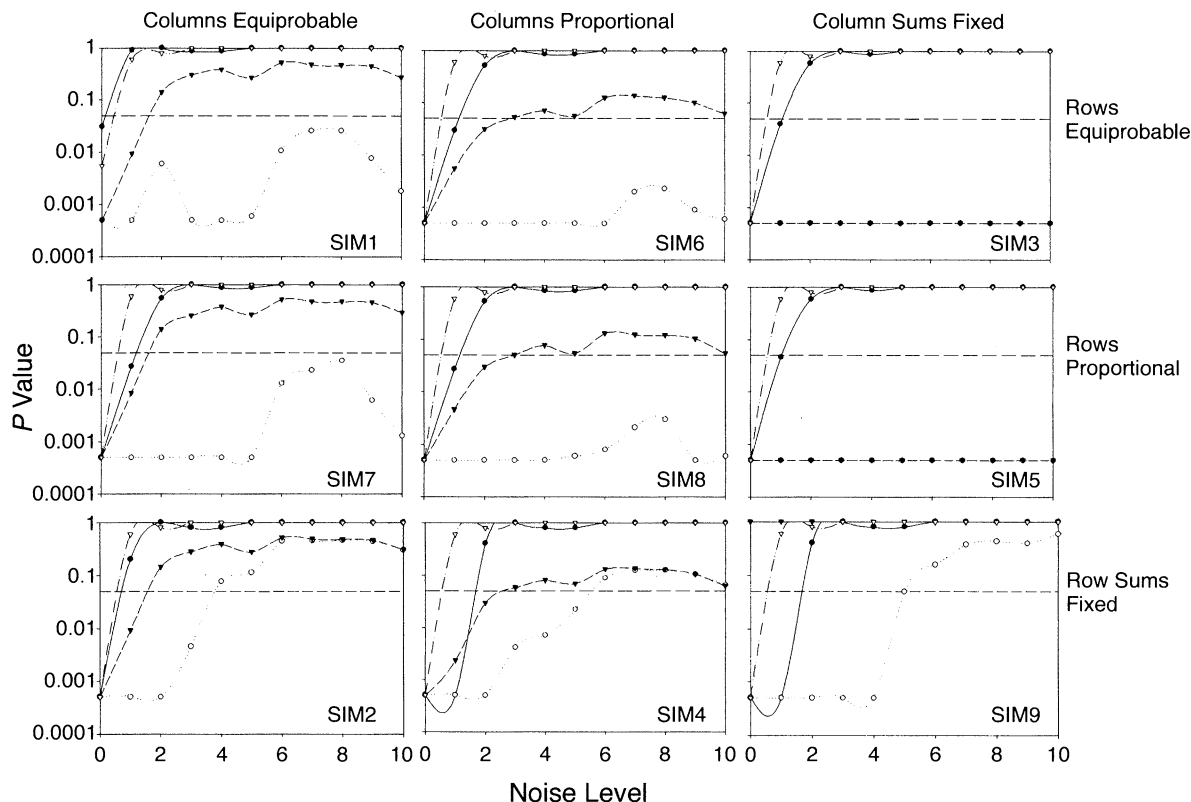
FIG. 6. Type II error tests for SIM1–SIM9. Each panel depicts the four co-occurrence metrics tested against a different simulation algorithm. The x-axis is the noise level, that is, the number of site occurrences that have been randomly transposed within each row of the perfect checkerboard matrix (Fig. 4A). The y-axis is the P value, shown on a log scale, with the 0.05 level indicated by a dashed line. Each point represents the average P value for five independent trials. Key to symbols: open circle = C score; solid diamond = V ratio; open triangle = number of species combinations (COMBO); solid circle = number of species pairs forming perfect checkerboards (CHECKER). Compare these curves to the idealized curves in Fig. 5.

of homogenous habitat. SIM4 is somewhat of a hybrid between these two, because it allows column totals to vary, but in proportion to observed totals. However, it may cause the null hypothesis to be incorrectly rejected, especially if used with the C score or CHECKER.

Choosing the right index

The choice of index is not as clear-cut as the choice of which algorithm to use, in part because each index seems to measure a slightly different aspect of species co-occurrence (Table 5). The number of checkerboards (CHECKER) and the number of species combinations (COMBO) are most relevant to the historical development of ideas on community assembly (Diamond 1975). However, both of these indices may cause the null hypothesis to be incorrectly accepted (Fig. 6), because they are sensitive to rearrangements of species occurrence patterns. A change in a single species occurrence can create or destroy a perfect checkerboard, or add or delete a species combination. Consequently, these measures will also be sensitive to measurement error, which may be common in presence–absence matrices.

In contrast, the C score and the V ratio are based on the average co-occurrence and covariance, respectively, of all species pairs. Therefore, minor changes in the data do not affect these indices as much. The C score, in particular, seems relatively insensitive to noise in the data, and can still detect pattern even when ~50% of the species occurrences in a perfect checkerboard matrix have been randomized (Fig. 6). When used with SIM2 or SIM9, the C score has good statistical properties and is not prone to false positives (Table 4). Finally, this index measures the checkerboard pattern of species mutual exclusion that reflects competitive interactions, but is not as restrictive as a count of perfect checkerboard pairs (CHECKER).

The variance ratio as an index of nonrandomness

The behavior of the V ratio is somewhat unusual, and deserves special comment. Of all the co-occurrence indices, it is the one index that is uniquely determined by the row and column totals, and not by the actual co-occurrence pattern itself. For this reason, it cannot be tested with SIM9, which retains row and column totals, and therefore generates no variation in the V ratio. For the V ratio, "The null hypothesis of no as-

sociation (H_0) states that the sum of the [species'] co-variances is zero" (Schluter 1984: 999). However, this turns out to be a somewhat different measure of co-occurrence than the C score and CHECKER.

What, precisely, does the V ratio measure? It measures the variability in the number of species per site. In a null community, we would expect the number of species per site to vary because of chance colonization. However, if species richness is regulated because of biological interactions, communities should converge on a relatively constant number of species per site. This is the niche limitation hypothesis of Wilson et al. (1987), which states that the variance in species richness per site should be unusually small if the number of species within a guild is limited by competition (Colwell 1979).

If all the sites have exactly the same number of species, there is no variance among sites, and the observed V ratio is zero. The reason the V ratio detects a highly significant pattern for the matrix in Fig. 4A is not because of the large number of checkerboard arrangements of species, but because each site in the matrix contains exactly 20 species. Thus, the V ratio, used with SIM2 or SIM4, is a useful probe for determining whether species interactions are constraining the number of coexisting species. There is no evidence of this for either the West Indian bird matrix (Table 6) or the Virginia ant matrix (Table 7), because in both of these examples, there is considerable heterogeneity in species richness per site. Variation among sites probably also explains Schluter's (1984) finding that most published co-occurrence matrices exhibit V ratios greater than 1.0.

The importance of row and column totals

All of the algorithms presented in this paper make use of information in the row and column totals to constrain the randomizations. Critics of null models have claimed that this procedure is circular because the marginals themselves reflect competitive interactions (Grant and Abbott 1980, Colwell and Winkler 1984). However, my analyses demonstrate that row and column constraints do not prevent the null model from detecting patterns in nonrandom matrices, even when the pattern has been considerably degraded by adding noise (Fig. 6). Moreover, the use of marginal constraints forms the basis for contingency table analysis (Fienberg 1980), which has been widely used in the analysis of species co-occurrence patterns (Whittam and Siegel-Causey 1981). If one wishes to test the hypothesis that marginal constraints are affected by species interactions, the V ratio, perhaps used with SIM2 or SIM4, would be an appropriate model.

Practical advice for the empiricist

Some ecologists may despair at the results of these analyses. After all, the same data matrix may yield random or highly significant patterns, depending on which algorithm or index is used for analysis (Tables 6 and 7). However, the same troubling result can be obtained from conventional parametric analyses. Data transformations, interaction terms, model structure, and designation of fixed and random factors can generate an equally bewildering diversity of outcomes in a "standard" analysis of variance (Scheiner and Gurevitch 1993). Ecologists need to move beyond the idea that there is a single "one-size-fits-all" null model that is appropriate. Rather, the null model and index should be chosen based on the kind of data (island lists vs. sample lists) collected and the question being asked. For a preliminary analysis of co-occurrence patterns of island list data, I suggest using the C score with SIM9. This combination is not vulnerable to false positives, has good power in the face of noisy data, and measures a pattern of co-occurrence that would be consistent with competitive exclusion.

But I would also recommend examining the outcome of other models and other indices, as in Tables 6 and 7. There is great value in exploring the results of several null models that incorporate different degrees of randomness. Statistical purists will not approve of this approach because it undercuts the rigorous interpretation of calculated probability values, and because it may tempt ecologists to go on "fishing expeditions" and search for an analysis that supports their preconceptions. The advantage of this approach is that it pinpoints how changing the assumptions of the model alters the results, which is an essential comparison for any confrontation of a model with real data (Hilborn and Mangel 1997). Certainly the interpretation of the West Indian finch matrix and the Virginia ant matrix was enhanced by consideration of a variety of null models.

The models presented here are by no means exhaustive, and it is easy to imagine other algorithms that could have been used. Ecologists will continue to develop new kinds of null models that incorporate spatial and temporal variability in populations and communities (Thomson et al. 1996, Roxburgh and Chesson 1998). However, the most important progress in null model analysis will probably come from the analysis of new data sets. Much of the older null model literature is dominated by analysis and reanalysis of published islands lists, such as the West Indian finch matrix (Fig. 2). These second- and third-hand data sets have a number of limitations, including species taxonomy, sample standardization, source pool limits, and geographic variation (Gotelli and Graves 1996). Home-grown data sets, such as the Virginia ant matrix (Fig. 3), may offer the best chance for examining species co-occurrence patterns, because the methods of collection and analysis can be tailored to the questions being asked. Of course, such data sets are more time consuming and costly to obtain than using published island lists, but ultimately, they should be more rewarding for understanding co-occurrence patterns in nature.

LITERATURE CITED

Abele, L. G., and W. K. Patton. 1976. The size of coral heads and the community biology of associated decapod crustaceans. Journal of Biogeography 3:35–47.

Alatalo, R. V. 1982. Bird species distributions in the Galapagos and other archipelagoes: competition or chance? Ecology 63:881–887.

Arnett, A. E. 1998. Geographic variation in life-history traits of the ant lion, *Myrmeleon immaculatus*: evolutionary implications of Bergmann's Rule. Dissertation. Department of Biology, University of Vermont, Burlington, Vermont, USA.

Case, T. J., and R. Sidell. 1983. Pattern and chance in the structure of model and natural communities. Evolution 37: 832–849.

Caswell, H. 1976. Community structure: a neutral model analysis. Ecological Monographs 46:327–354.

Coleman, B. D., M. A. Mares, M. R. Willig, and Y.-H. Hsieh. 1982. Randomness, area, and species richness. Ecology 63: 1121–1133.

Colwell, R. K. 1979. Toward a unified approach to the study of species diversity. Pages 75–91 in J. F. Grassle, G. P. Patil, W. Smith, and C. Tallie, editors. Ecological diversity in theory and practice. International Cooperative Publishing House, Fairland, Maryland, USA.

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London B 345:101–118.

Colwell, R. K., and D. W. Winkler. 1984. A null model for null models in biogeography. Pages 344–359 in D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B. Thistle, editors. Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, New Jersey, USA.

Connor, E. F., and D. Simberloff. 1978. Species number and compositional similarity of the Galapagos flora and avifauna. Ecological Monographs 48:219–248.

Connor, E. F., and D. Simberloff. 1979. The assembly of species communities: chance or competition? Ecology 60: 1132–1140.

Connor, E. F., and D. Simberloff. 1983. Interspecific competition and species co-occurrence patterns on islands: null models and the evaluation of evidence. Oikos 41:455–465.

Connor, E. F., and D. Simberloff. 1984. Neutral models of species' co-occurrence patterns. Pages 316–331 in D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B. Thistle, editors. Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, New Jersey, USA.

Connor, E. F., and D. Simberloff. 1986. Competition, scientific method, and null models in ecology. American Scientist 74:155–162.

Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 in M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Harvard University Press, Cambridge, Massachusetts, USA.

Diamond, J. M., and M. E. Gilpin. 1982. Examination of the "null" model of Connor and Simberloff for species co-occurrences on islands. Oecologia 52:64–74.

Edgington, E. S. 1987. Randomization tests. Marcel Dekker, New York, New York, USA.

Fienberg, S. E. 1980. The analysis of cross-classified categorical data. MIT Press, Cambridge, Massachusetts, USA.

Fox, B. J., and J. H. Brown. 1993. Assembly rules for functional groups in North American desert rodent communities. Oikos 67:358–370.

Gilpin, M. E., and J. M. Diamond. 1982. Factors contributing to nonrandomness in species co-occurrences on islands. Oecologia 52:75–84.

Gilpin, M. E., and J. M. Diamond. 1984. Are species co-occurrences on islands nonrandom, and are null hypotheses useful in community ecology? Pages 297–315 in D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B. Thistle, editors. Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, New Jersey, USA.

Gilpin, M. E., and J. M. Diamond. 1987. Comment on Wilson's null model. Oecologia 74:159–160.

Gotelli, N. J., and L. G. Abele. 1982. Statistical distributions of West Indian land bird families. Journal of Biogeography 9:421–435.

Gotelli, N. J., N. J. Buckley, and J. A. Wiens. 1997. Co-occurrence of Australian land birds: Diamond's assembly rules revisited. Oikos 80:311–324.

Gotelli, N. J., and G. L. Entsminger. 1999. EcoSim. Null models software for ecology. Version 3.0. Acquired Intelligence Incorporated, and Kesey-Bear. [Online: ⟨http://homepages.together.net/~gentsmin/ecosim.htm⟩.]

Gotelli, N. J., and G. R. Graves. 1996. Null models in ecology. Smithsonian Institution Press, Washington, D.C., USA.

Grant, P. R., and I. Abbott. 1980. Interspecific competition, island biogeography and null hypotheses. Evolution 34: 332–341.

Graves, G. R., and N. J. Gotelli. 1993. Assembly of avian mixed-species flocks in Amazonia. Proceedings of the National Academy of Sciences (USA) 90:1388–1391.

Haefner, J. W. 1988. Niche shifts in greater Antillean *Anolis* communities: effects of niche metric and biological resolution on null model tests. Oecologia 77:107–117.

Harvey, P. H., R. K. Colwell, J. W. Silvertown, and R. M. May. 1983. Null models in ecology. Annual Review of Ecology and Systematics 14:189–211.

Hastings, A. 1987. Can competition be detected using species co-occurrence data? Ecology 68:117–123.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey, USA.

Inger, R. F., and R. K. Colwell. 1977. Organization of contiguous communities of amphibians and reptiles in Thailand. Ecological Monographs 47:229–253.

Jackson, D. A., K. M. Somers, and H. H. Harvey. 1992. Null models and fish communities: evidence of nonrandom patterns. American Naturalist 139:930–951.

Johnson, M. P. 1974. Species number, endemism, and equilibrium in the Galapagos biota. AAAS Galapagos Symposium, San Francisco. American Academy of Arts and Sciences, Cambridge, Massachusetts, USA.

Kelt, D. A., M. L. Taper, and P. L. Meserve. 1995. Assessing the impact of competition on community assembly: a case study using small mammals. Ecology 76:1283–1296.

Manly, B. F. J. 1995. A note on the analysis of species co-occurrences. Ecology 76:1109–1115.

McCoy, E. D., and K. L. Heck, Jr. 1987. Some observations on the use of taxonomic similarity in large-scale biogeography. Journal of Biogeography 14:79–87.

Patterson, B. D., and W. Atmar. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. Biological Journal of the Linnean Society 28:65–82.

Pielou, D. P., and E. C. Pielou. 1968. Association among species of infrequent occurrence: the insect and spider fauna of *Polyporus betulinus* (Bulliard) Fries. Journal of Theoretical Biology **21**:202–216.

Robson, D. S. 1972. Appendix: statistical tests of significance. Journal of Theoretical Biology **34**:350–352.

Roughgarden, J. 1983. Competition and theory in community ecology. American Naturalist **122**:583–601.

Roxburgh, S. H., and P. Chesson. 1998. A new method for detecting species associations with spatially autocorrelated data. Ecology **79**:2180–2192.

Sale, P. F. 1974. Overlap in resource use, and interspecific competition. Oecologia **17**:245–256.

Scheiner, S. M., and J. Gurevitch, editors. 1993. Design and analysis of ecological experiments. Chapman and Hall, New York, New York, USA.

Schluter, D. 1984. A variance test for detecting species associations, with some example applications. Ecology **65**:998–1005.

Simberloff, D. 1978. Using island biogeographic distributions to determine if colonization is stochastic. American Naturalist **112**:713–726.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. W. H. Freeman, New York, New York, USA.

Stone, L., T. Dayan, and D. Simberloff. 1996. Community-wide assembly patterns unmasked: the importance of species' differing geographical ranges. American Naturalist **148**:997–1015.

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

Terborgh, J. 1973. Chance, habitat and dispersal in the distribution of birds in the West Indies. Evolution **27**:338–349.

Thomson, J. D., G. Weiblen, B. A. Thomson, S. Alfaro, and P. Legendre. 1996. Untangling multiple factors in spatial distributions: Lilies, gophers, and rocks. Ecology **77**:1698–1715.

Whittam, T. S., and D. Siegel-Causey. 1981. Species interactions and community structure in Alaskan seabird colonies. Ecology **62**:1515–1524.

Wiens, J. A. 1989. The ecology of bird communities. Volume 1. Foundations and patterns. *In* R. S. K. Barnes, H. J. B. Birks, E. F. Connor, and R. T. Paine, editors. Cambridge studies in ecology. Cambridge University Press, Cambridge, UK.

Wilson, J. B. 1987. Methods for detecting nonrandomness in species co-occurrences: a contribution. Oecologia **73**:579–582.

Wilson, J. B. 1995. Null models for assembly rules: the Jack Horner effect is more insidious than the Narcissus effect. Oikos **72**:139–144.

Wilson, J. B., H. Gitay, and A. D. Q. Agnew. 1987. Does niche limitation exist? Functional Ecology **1**:391–397.

Winemiller, K. O., and E. R. Pianka. 1990. Organization in natural assemblages of desert lizards and tropical fishes. Ecological Monographs **60**:27–55.

SUPPLEMENTARY MATERIALS

The software utilized in the analysis presented here (EcoSim) is available in ESA's Electronic Data Archive: *Ecological Archives* E081-022. The software may be downloaded at no cost. All of the algorithms described in the paper are included.