An Optimality Criterion to Determine Areas of Endemism

CLAUDIA A. SZUMIK,[1,2] FABIANA CUEZZO,[2] PABLO A. GOLOBOFF,[1,2]
AND ADRIANA E. CHALUP[2]

[1]*Consejo Nacional de Investigaciones Científicas y Técnicas, Miguel Lillo 205, 4000 San Miguel de Tucumán,
Tucumán, Argentina*
[2]*Instituto Superior de Entomología, Miguel Lillo 205, 4000 San Miguel de Tucumán, Tucumán, Argentina;
E-mail: insue@infovia.com.ar*

Abstract.—A formal method was developed to determine areas of endemism. The study region is divided into cells, and the number of species that can be considered as endemic is counted for a given set of cells (= area). Thus, the areas with the maximum number of species considered endemic are preferred. This is the first method for the identification of areas of endemism that implements an optimality criterion directly based on considering the aspects of species distribution that are relevant to endemism. The method is implemented in two computer programs, NDM and VNDM, available from the authors. [Biogeography; endemicity; optimality criterion.]

Identification of areas of endemism is important for both historical biogeography and conservation. Although there are many formalized methods for determining the relationships between areas of endemism in vicariance biogeography (e.g., Nelson and Platnick, 1981; Brooks, 1990; Page, 1994; Nelson and Ladiges, 1996; Ronquist, 1997) and for determining conservation priorities (e.g., Vane-Wright et al., 1991; Faith, 1992; Pressey et al., 1993; Williams, 1996; Rodrigues et al., 2000), there are almost no equivalent methods for the identification of the areas of endemism themselves.

In contrast to species (which normally have discrete boundaries), areas of endemism are difficult to recognize because the basic biogeographic patterns are easily obscured by many factors (dispersal, extinction, etc.). Thus, a formalization of the criteria used for recognition of areas of endemism is clearly needed.

An explicit method to identify areas of endemism should relate relevant evidence and conclusions. With this method, an investigator should be able to evaluate a potential area independently of how the area was defined. Acceptance of those conclusions (i.e., boundaries of areas) that are best supported by available evidence requires (in principle, at least) evaluation of all possible conclusions, selecting the ones judged as optimal based on the established criterion.

Harold and Mooi (1994), Morrone (1994), and Linder (2001) have discussed identification of areas of endemism. However, an explicit criterion of optimality was either lacking (Harold and Mooi) or was used only a posteriori to select among the conclusions found by other less appropriate means (Morrone and Linder).

A method used to identify areas of endemism must consider the taxa occurring in a given area and their positions in space. This spatial component has not been included in preexisting clustering methods, and thus those methods (designed only to recover hierarchy) cannot be adopted for identification of areas of endemism. An attempt to produce such a formalization, taking into account the spatial component of endemism, is presented here.

GENERAL CONSIDERATIONS

An endemic taxon is restricted to a region and is found nowhere else. The range of distribution of a taxon is determined by both historical and current factors. Whatever the factors are, if they affect (or have affected) in a similar way different taxonomic groups, there will be congruence in the patterns of endemicity in different groups. Thus, areas that have many different groups found there and nowhere else can be defined as areas of endemism. Such a situation would, of course, indicate that the speciation processes in the different groups have been caused by common factors, but knowledge of the factors is not a prerequisite to identifying the existence of the area of endemism itself.

This notion of area of endemism has several implications as to factors to be considered when proposing a formalized identification method, particularly regarding the
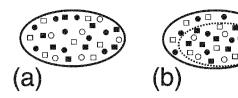
FIGURE 1. Examples of ideal (a) and realistic (b) distributions of species endemic to an area. (a) The two species (black squares and white circles) occupy all the area. (b) Both species are confined to some sector of the area (stippled).

limits of the area, the widely and narrowly distributed taxa, and the use of grids.

Ideally, the limits of the area of endemism would be inviolable; none of its species would be found outside the area. Additionally, under ideal conditions, all the species should be found in every part of the area of endemism (Fig. 1a). However, not all taxonomic groups will respond in exactly the same way to the factors that either cause or modify the area of endemism (e.g., not all species expand or contract their distributions in exactly the same way). A consequence is that the limits of the area will often be diffuse, with borders of areas possibly supporting some of the endemic species but lacking others (Fig. 1b).

As has been suggested (Platnick, 1991, 1992), the taxa to be used should be those that are maximally endemic, i.e., those for which the ranges of distribution are small, compared with the study region. The range size, of course, is relative to the size of the study region. A species distributed in all of the dry Chaco is widely distributed if the study region is the Chaco but narrowly distributed if the study region is South America.

The use of grids seems unavoidable, because the series of dots that makes up the actual records for a species must be converted into ranges in some explicit way. The size of the grid cells will, obviously, affect the results, perhaps in a deterministic manner. For example, use of very small grid cells will render all distributions entirely discontinuous, and then only very small areas of endemism (or none at all) would be recognized. Alternatively, use of very large grid cells will probably cause very large areas to be recognized, with many species appearing as endemic to each area.

PROBLEMS WITH PREVIOUS PROPOSALS

Harold and Mooi (1994) stated that sympatry is not a prerequisite for the recognition of an area of endemism. Although no one would expect exact congruence in the distributional limits of two or more species at every possible scale of mapping, some extensive sympatry must exist at the relevant level (Platnick, 1991; Morrone, 1994). Harold and Mooi (1994:265) argued that "nonoverlapping distributions need not be considered separate historical entities if there is independent evidence that the areas could be considered as one." They used as example several islands (A, B, and C; Fig. 2), with two species present in each of them; the species in islands B and C are sympatric, and the species in island A are allopatric. The two species that inhabit the island do not coexist; if anything, the distribution of the species in island A argues for recognizing two areas of endemism within island A. However, Harold and Mooi argued that in such a situation, A must be considered an area of endemism instead of a composite, and the validity of this assumption will be tested with information from other groups of organisms. The testing, however, could hardly be considered significant when all of the data are entered in this way or when A is considered (a priori) as a single unit; testing with information from other groups of organisms simply cannot correct for mistakes like these. Not surprisingly, Harold and Mooi's approach does not provide a strict formalization. In the absence of formalization, the criteria proposed by Harold and Mooi are not operational.
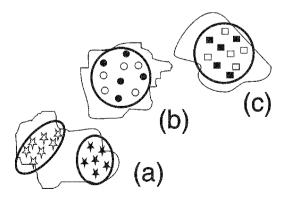

FIGURE 2. Distributions of species inhabiting islands (a, b, c). Redrawn from Harold and Mooi (1994: Fig. 2).

Morrone (1994) and Linder (2001) proposed more detailed operational procedures, scoring presence/absence of each species in each cell of a grid. Both authors proposed to use counts of endemic species as a criterion to evaluate possible areas, although they were not completely specific on how to decide whether or not a species can be considered endemic. Using counts of endemic species to select from among all possible conclusions presents considerable computational difficulties. Instead of using those counts to select from among all possible sets of cells, Morrone and Linder used them to select from among the sets of cells produced by a parsimony analysis or UPGMA clustering using the Jaccard similarities. As both Morrone and Linder were well aware, not all the species appearing as "synapomorphies" of a given set of cells will correspond to endemic species, because they may also be synapomorphies of many other (not closely related and geographically distant) groups. This possibility violates the main requirement for endemicity, that of being restricted to the area. Parsimony is indeed an appropriate criterion for phylogenetic reconstruction, but it cannot be adapted to a field with completely different goals and premises. Likewise, UPGMA may prefer groups of cells with no endemic species over groups with several endemics. Thus the counts of species endemic to different sets of cells should be used to select from among all possible sets, not only those sets that parsimony or UPGMA happen to produce.

AN OPTIMALITY CRITERION

A method to determine areas of endemism based on an optimality criterion must provide a way to assign a value of endemicity, or score, to a given area (= set of grid cells) regardless of how that area was found or hypothesized. For different definitions of an area, there will be different numbers of species that can be considered endemic. For example, a species will satisfy the requirement for endemicity if the area comprises the same cells where it is distributed but will not satisfy the requirement if the area comprises half those cells. Thus, for different sets of cells, there will be different numbers of species that can be considered as "endemic," i.e., they will have different scores

of endemicity. A natural criterion of optimality is thus provided by counting the species that can be considered as endemic, given the area (and the species distributions). Obviously, from among possible areas, those with the highest scores of endemicity should be preferred.

To determine how many species appear as endemic, endemicity itself must be determined for each species in a formalized manner, which can be done in several ways. Four possible criteria have been examined, from a very strict or ideal concept of endemism (criteria 1 and 2, with a very high congruence required between the species distribution and the area) to less rigorous but more realistic requirements (criteria 3 and 4, which allow for some incongruence). Because each of the criteria is a relaxation of the preceding one(s), the score under each criterion will always be equal to or greater than the score under the preceding criteria. The data entry is done following the steps outlined by Morrone (1994), by plotting species localities on a map with a grid, except that the spatial location of the cell in the grid (as row, column; see Fig. 3) must also be considered. The method has been implemented in two computer programs, NDM and VNDM (Goloboff, 2001). NDM is the basic search engine, and VNDM is a program that helps viewing and diagnosing (e.g., finding out which species contribute to the score). Optionally, the data can be read as coordinates, and internally converted by the
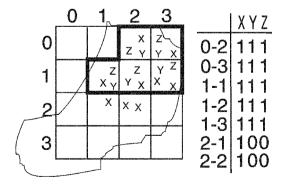

FIGURE 3. An area (including five grid cells) with score 2 under criterion 1 and score 3 under criterion 2. Under criterion 1, species X, even if occurring in each cell of the area, does not contribute to the score because it is also found in cells outside the area. Under criterion 2, species X contributes to the score. All the cells in the area have identical species composition.

programs to a grid of a specified size. In that case, the programs will consider a species as present in a cell if it is present in at least one point (= locality record) inside the cell; a point on the edge (or corner) of a cell indicates presence of the corresponding species in the two (or four) adjacent cells. Optionally, it is also possible to consider each point as having a "radius" equal to some (user defined) percentage of the cell width or height, so that a point very close to the edge (or corner) of a cell can be considered as also present in the adjacent cell(s).

As the criteria are defined here, they cannot be applied to disjunct areas; only areas where all cells are contiguous can be evaluated. Although it would of course be desirable to have a criterion to evaluate disjunct areas, this first approximation to the problem does not allow such an evaluation.

For a more explicit definition of the criteria, a simple notation is used:

A = an area (= set of cells);
C_N = nth cell that belongs to A;
C_n = nth cell that does not belong to A;
N_A = set of cells not adjacent to A;
S_N = set of species present in C_N;
S_n = set of species present in C_n;
X_A = set of species that contribute to the
    score of area A.

In all cases, the score of an area A will be the cardinality of X_A. The complement of a set S is denoted as $\sim$S.

First Criterion (E_1)

This criterion assumes that the distribution of a species must adjust perfectly to the area to contribute to the score. For all C_I in A, S_I must be identical; if some $S_I \neq S_J$, then $X_A = \emptyset$; otherwise $X_A = (S_J \cap S_K \cap \ldots \cap S_N)$ $\cap \sim (S_j \cup S_k \cup \ldots \cup S_n)$. That is, a species contributes to the score if it is found in the area and nowhere else, and each of the cells in the area has exactly the same species composition. Figure 3 is an example; the area formed by the cells 0–2, 0–3, 1–1, 1–2, and 1–3 has an endemicity score $E_1 = 2$.

Second Criterion (E_2)

This criterion is similar to the preceding one, but a species can contribute to the score if present in some cell outside the area as long as the cell is adjacent to the area. Thus, like before, S_I must be identical for all C_I in A; if some $S_I \neq S_J$, then $X_A = \emptyset$; otherwise, define

$B = (S_J \cap S_K \cap \ldots \cap S_N)$;
$V = (S_j \cup S_k \cup \ldots \cup S_n)$ (for all j, k, and n
    that belong to N_A);
$X_A = B \cap \sim V$.

Under this criterion therefore it is not required that all the species contributing to the score have identical distributions. The example of Figure 3 will have a score $E_2 = 3$, contributed by the distributions of species X, Y, and Z; X contributes to the score because it is found outside the area but only in neighboring cells (2–1 and 2–2).

Third Criterion (E_3)

This criterion is similar to the preceding criterion but drops the requirement that S_I must be identical for all C_I in A. Thus, it is not required that all cells in A have identical species composition. However, because X_A is determined as with the previous criterion, only species occurring in each and every one of the cells in A will contribute to the score. Figure 4 shows an example; the area formed by cells 0–1, 0–2, 0–3, 1–1, and 1–2 has a score $E_3 = 2$ (by species X and Y).

Fourth Criterion (E_4)

Under criteria 1 through 3, a species can contribute to the score only if it is present in each and every one of the cells of the area. A more realistic criterion, however, must take into account the fact that a species may be absent from a given cell because of poor collecting effort or partial extinction (as in urban

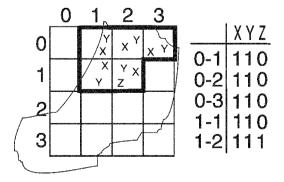|	X Y Z
0-1	1 1 0
0-2	1 1 0
0-3	1 1 0
1-1	1 1 0
1-2	1 1 1

FIGURE 4.   An area with score 2 under criterion 3. Not all cells in the area have identical species composition. Species X and Y contribute to the score; species Z does not because it is found in only some cells of the area.

areas). A species, therefore, should be able to contribute to the score even if absent from some cells. However, the mere number (or percentage) of cells in which the species is present (as proposed by Linder, 2001) would be a poor indicator, because the species could satisfy this requirement by being confined to some part of the area (e.g., very common in the right half of the cells, really absent in the left half), which is clearly undesirable. Some indicator of whether the species is more evenly distributed in the area is needed. One can be provided by considering that a species must satisfy three conditions for endemicity: (1) It is present in at least two of the cells that form the area, (2) it is present either in C_I itself or in one of the adjacent cells that belongs to A for each of the C_I, and (3) it is absent from no more than Q (where $0 < Q < 8$) of the cells around C_I that belong to A ($Q = 0$ is equivalent to criterion 3). Only species that are more or less evenly distributed in the area will satisfy this requirement, and then X_A will include all those species as long as they are not found in any of the cells in N_A (in which case, they are widespread taxa). Additionally, each of the cells in A must have at least one of the species in X_A; if one of the cells lacks each of the species in X_A (i.e., if $X_A \cap C_I = \emptyset$, for some I), then make $X_A = \emptyset$. (Without this proviso, adding a strip of empty cells on the side of the area would sometimes maintain the same score and will produce areas a little larger than actually indicated by the data.) The check for empty surrounding cells can also be done for occupied cells. (Checking around occupied cells provides for a more stringent requirement and thus a score equal to or less than that obtained without checking.)

Under this criterion, not all cells in A are required to have the same species composition. An example is shown in Figure 5, where cells 0–2, 0–3, 1–2, and 1–3 have $E_4 = 4$ (given by species W, X, Y, Z); each of the species contributing to the score is present in only some cells in the area but is present in at least one adjacent cell.

SEARCHING FOR OPTIMAL AREAS

As the criteria are defined above, one can simply evaluate all possible sets of cells and select those with the highest scores. However, this approach is computationally very intensive and is intractable for even modest
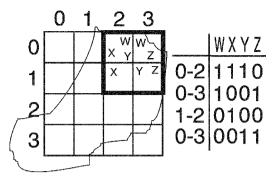

FIGURE 5. An area with score 4 under criterion 4. None of the species are found in every cell of the area, but all satisfy the requirement of having no more than seven empty cells around a given cell in the area and having at least one adjacent cell occupied.

numbers of cells. Equally difficult problems have been posed for reserve selection criteria (reviewed by Rodrigues et al., 2000), also based on evaluating possible sets of cells. However, the reserve selection algorithms implemented so far do not take into account whether a species occurring inside the study area also occurs outside, which is a key component for evaluating endemism. The branch-and-bound implementation of those reserve selection criteria is about as time consuming as our present implementation (see Rodrigues et al., 2000, for details), and the heuristic algorithms used for that problem are not applicable in the present case.

NDM, the program used here to explore the method, uses a branch-and-bound implicit enumeration of areas. Such an approach guarantees the correctness of the results; however, although useful to explore the properties of the method and to analyze small or medium sized examples, it is not applicable to very large data sets. The strategy used by NDM to find the sets that actually maximize the score under a given criterion is detailed here. To facilitate description, an absolute numbering of the cells will be used; the absolute number of a cell with coordinates x and y in a grid of C columns is defined as $(y * C) + x$. Internally, NDM uses a bitwise representation of the species distributions and the sets of cells; this approach allows calculation of unions or intersections easily for 32 species or cells at a time. (Analogous procedures have already been used in several parsimony programs; Allard et al., 1999; Moilanen, 1999; see also Goloboff, 2002, for a generalization to polytomies.)

During the search stage, NDM actually examines only the areas having more than a single cell (those with single cells can be easily examined later). To enumerate all possible combinations of cells, NDM starts with an empty set. To this set, it adds first cell number 0 (upper left corner) and tries all combinations of the remaining cells together with cell 0. Then it eliminates cell 0, adds cell 1, and tries all possible combinations of the remaining cells. This procedure is repeated until the first cell included in the set is the one before the last cell in the grid (lower right corner), in which case only one two-cell set can be generated. The possible combinations of cells are always examined in the same orderly fashion.

The procedure described allows generation of all possible sets of cells. Each of the combinations must be evaluated for continuity (disjunct areas are ignored) and, if continuous, assigned a score under the criterion (or criteria) in effect. Actual examination of each possible combination in this fashion is extremely time consuming (requiring several hours even for small data sets), but many of the sets can be implicitly rejected by predicting that they will be discontinuous or that they will have a low number of endemic species.

Discontinuous sets will have gaps. For example, in a grid with eight columns, the set formed by cells 0, 1, 5, and 6 is discontinuous. The mere existence of some gap (such as 2, 3, and 4 in the example) is not enough to deduce that any resulting set will be discontinuous; e.g., adding cells 10, 11, and 12 to the original set will make it continuous. However, whenever a gap is longer than the number of columns plus 2, any resulting set produced by adding subsequent cells will be discontinuous (e.g., in an eight-column grid, the set formed by 0 and 10 is discontinuous, and no possible addition of cells beyond 10 can make it continuous). When a partial combination of cells contains such a long gap, all the sets that result from adding further cells to that partial combination are ignored.

For criteria 1 through 3, predicting which species can potentially contribute to the score is easy because these criteria require that a species be found in each and every cell of the area to count as endemic. As each cell is added to the set, the intersection of the species contained in the new cell with those previously included is calculated. If the in-

tersection for a set of cells is empty, it follows that the intersection of any possible additional combination of cells will also be empty, and then those additional cells are never added. Actually, NDM checks whether the partial intersection has fewer members than a given minimum score; obviously, searching for areas with larger scores speeds up calculations because it interrupts calculations earlier.

For criterion 4, the calculations are more difficult because a species can contribute to the score even if it does not occur in each cell of the area. A good lower bound on the score can be obtained by calculating an enlarged distribution for each species (done before the search itself starts and stored in memory). For such enlarged distribution, a cell is considered as having the species present if the species satisfies the requirements of criterion 4 in that cell (i.e., actually present in at least one adjacent cell, absent in no more than Q cells). The intersection of the species (with enlarged distributions) in the cells of a given set of areas will be a superset of the set of species actually giving a score under criterion 4 for that area. Thus, if the number of members in the intersection of the species occurring in the enlarged distribution in a set of cells is less than the minimum score, it follows that no set formed by adding further cells can have an E_4 equal to or greater than the minimum score. This is true as long as the distributions have been enlarged by allowing up to seven empty cells around a given cell and not checking around cells actually occupied. If the number of allowed empty cells is less, the number of species contributing to the score can be underestimated because a cell may be surrounded by some number of empty cells in the full grid but by a smaller number when an area is defined (if the area excludes some of the cells that did not have the species; only the empty cells belonging to the area are counted). Thus, some areas with a positive score (optimal or not) may be missed during the search. The likelihood of missing positive areas in a given case depends on the relative numbers of empty cells used to create the enlarged species distributions and to evaluate the areas. Allowing up to five empty cells when enlarging species distributions is unlikely to create errors if the areas are to be evaluated allowing up to two empty cells but is very likely to create errors if the areas are

to be evaluated allowing up to five empty cells.

The enlargement of species distributions allowing for fewer empty cells can find areas that are contractions of the actual optimal areas, i.e., areas that are produced by eliminating some cells from the actual optimal area. Some of these errors (not necessarily all) will be remedied if a heuristic addition of cells, one at a time, is done for each of the cells found, retaining (and submitting to the same procedure) each of the enlarged areas that has a positive score.

Additional speed can be obtained by identifying in advance species that cannot contribute to the score of a given area by virtue of occurring in nonadjacent cells. A cell that is columns + 2 positions before the first cell in a set and a cell that is columns + 2 positions beyond the last cell in the set will by necessity be discontinuous (i.e., nonadjacent) to the area. For each cell i, a set F_i can be calculated as $F_i = F_{i-1} \cup S_i$ and a set B_i as $B_i = B_{i+1} \cup S_i$ (where S_i is the set of species occurring in cell i); this calculation is done before the search starts. Then, during the search, if the first cell in the set is in position i, any species occurring in the set $F_{i-(columns+2)}$ cannot contribute to the score and can be eliminated from the set of species potentially contributing to the score. (As before, if fewer species than the minimum score occur in that set, there is no need to form all the areas that result from adding further cells to the present set of cells.) Before evaluating a given area, all the species in $B_{j+columns+2}$ (where j is the last cell of the set) can be eliminated from the set of species potentially contributing to the score. (This saves less time than checking against F but still saves some time because some areas can be rejected easily without further evaluation.)

Because higher minimum scores allow for a quicker rejection of many areas, they produce faster searches. Using all the shortcuts described above, NDM can analyze data sets of medium size in reasonable times. On a 266-MHz pentium II machine, the areas with score ≥ 2 for a real matrix of carabid beetles, with a grid of 10×15 and 33 species (actually occurring in 42 cells) can be found in 664 sec, the areas with score ≥ 3 in 1.97 sec, and the areas with score ≥ 4 in 0.69 sec. The areas with score ≥ 2 can be found in only 1.05 sec if the enlarged species distributions are calculated allowing for up to four empty cells, and the differences from the correct results

are minimal. For larger problems, it is possible to find good solutions by constraining the search to a given region; only those sets contained within the region are evaluated. The candidate regions can be selected by analyzing the data with enlarged grid cells (e.g., reducing the number of rows and columns to a half or a third) and then constraining the search to the corresponding region of the larger data set.

FURTHER CONSIDERATIONS

A possibility that has not been discussed so far is that of conflict between the areas with a positive score under some of the criteria. It is of course possible, given conflicting distributions, that two sets of cells, where one is a subset of the other, both have positive scores (under criteria 2 through 4). The one with the largest score is the one more strongly supported by the evidence. If two partially overlapping areas have the same score, either the evidence is ambiguous regarding which of the areas is an area of endemism or both represent real phenomena. (If each is supported by the congruent distribution of many taxa, the taxa may simply be responding to different factors, such as terrestrial vs. aquatic organisms.) Another possibility is that several subsets of an optimal area will also have some positive score. This result does not really represent conflict but simply reflects the fact that some species may have their ranges further contracted. As implemented in NDM, such smaller areas will not be considered; the program eliminates them. The situation is different, of course, if the smaller area has a larger score (under some criteria), in which case both areas are saved. Ideally, the comparison should take into account whether the scores for the larger and smaller areas are given by different sets of species, and if so, it should retain both areas (this option has not yet been implemented).

Whether an area X in conflict with another area Y of higher score is reported by the program or not may depend in turn on whether area Y itself is in conflict with another area (e.g., Z) of even higher score. If so, area Y must be eliminated (because it loses against Z), and X will be retained. Thus, NDM cannot check for conflict between the areas as it finds them. If it did so, finding first Y, then X, then Z, it would miss area X; when X is found and compared to Y, it is discarded, and when

Z is found, it discards Y. Only finding X after both Y and Z are found would produce the correct result. To avoid this problem, NDM stores all the areas with positive score that it finds during the search, and only when the search is finished does it globally compare all the areas for conflict.

The four criteria for scoring can be used simultaneously during a search. Because each criterion is a relaxation of the preceding one(s), the criteria do not actually contradict each other but give instead complementary information.

A REAL EXAMPLE: REANALYSIS OF *SCIOBIUS* SCHÖNHERR

Morrone (1994) analyzed, using parsimony, a matrix of 47 species of *Sciobius* (Coleoptera: Curculionidae) from South Africa in a grid with 21 occupied cells. On the consensus from 289 optimal trees, Morrone (1994) proposed three areas of endemism (Fig. 6). Area 1 (cells I, J, L, and M) is defined by having five species; there are seven species as synapomorphies of this area, but Morrone indicated only five, presumably by considering that only these five were endemic. Area 2 (cells N, O, R, S, and T) is defined by having two species, and area 3 (cell P) is defined by having seven species (here, Morrone counted only the autapomorphies). The same matrix analyzed under criteria 3 and 4 with NDM (allowing for up to two empty cells around each cell in the area) obtained a total of 16 areas (in 1.17 sec running on a 266-MHz Pentium II machine), as shown in Figure 7 (the two single cell areas, 4-3 and 4-5, N and P in Morrone's grid, are not shown in that

figure). The three areas of largest score are the first three in Figure 7. Area 1 completely includes areas 8 and 9 (all of lower E_4, but reported by NDM because they have higher E_3) and is completely included in areas 4, 5, 6, and 7 (all of lower E_4). Area 2 is in conflict with area 10 and completely includes area 11 (both of lower E_4). Area 3 is in conflict with area 13, is included completely in area 12, and includes area 14 (the three with lower E_4). Area 1 of Morrone (1994) is equivalent to our area 1, and area 2 of Morrone (1994) is equivalent to our area 12 (which is suboptimal according to our criterion). Area 3 of Morrone is equivalent to one of our single-cell areas. Morrone's analysis did not recognize any possible equivalent of our area 2 nor any equivalent of the single cell area N.
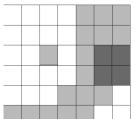
Even for the areas that appear (identical or very similar) in the analysis of Morrone (1994), there are significant differences in the species that define the areas. Area 1 is diagnosed under criterion 4 as having 17 endemic species (see Fig. 8). Of these 17 species, only 7 (6, 7, 10, 12, 22, 23, and 46) appear as synapomorphies of the area when mapped most parsimoniously onto the consensus tree; Morrone (1994) actually showed only 5 species (he did not show 12 and 23). (Morrone [1994] mapped the characters onto the consensus tree; we consider that it is better to map the individual trees, but we use the consensus for comparability with Morrone's results.) Some of the species contributing to the score under criterion 4 do not appear as synapomorphies under parsimony because they are not found in all the cells forming the area. Species 23 (*S. marshalli*) appears as a synapomorphy under parsimony, but because it is also present in the nonadjacent cells A and B it seems illogical to count it as supporting endemicity. Thus, for the distribution of *Sciobius*, the criteria proposed here produce more reasonable results than parsimony.
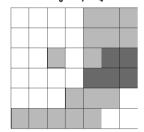
CONCLUSIONS AND PROSPECTS

The method proposed here is only a first approximation of a solution to the problem of identifying areas of endemism. The method could be improved in many ways that would still reflect its general spirit and approach.
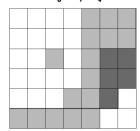
The first aspect is the continuity of the area of endemism; as the criteria are defined here, the areas of endemism resulting from habitat
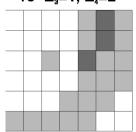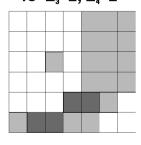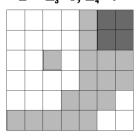

FIGURE 6. Grid used by Morrone (1994) in his analysis of *Sciobius*. The areas marked are the ones selected by Morrone's method. Area 1 = cells I, J, L, and M (medium shading); area 2 = cells N, O, R, S, and T (dark shading); area 3 = cell P (light shading).

1 $E_3=6$, $E_4=17$

2 $E_3=0$, $E_4=4$

3 $E_3=1$, $E_4=4$

4 $E_3=0$, $E_4=10$

5 $E_3=0$, $E_4=10$

6 $E_3=1$, $E_4=10$

7 $E_3=0$, $E_4=10$

8 $E_3=8$, $E_4=14$

9 $E_3=12$, $E_4=12$

10 $E_3=1$, $E_4=2$

11 $E_3=2$, $E_4=2$

12 $E_3=1$, $E_4=2$

13 $E_3=2$, $E_4=2$
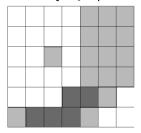
14 $E_3=2$, $E_4=3$

FIGURE 7.   The 14 sets with positive E_3 or E_4 for the data of Morrone (1994).
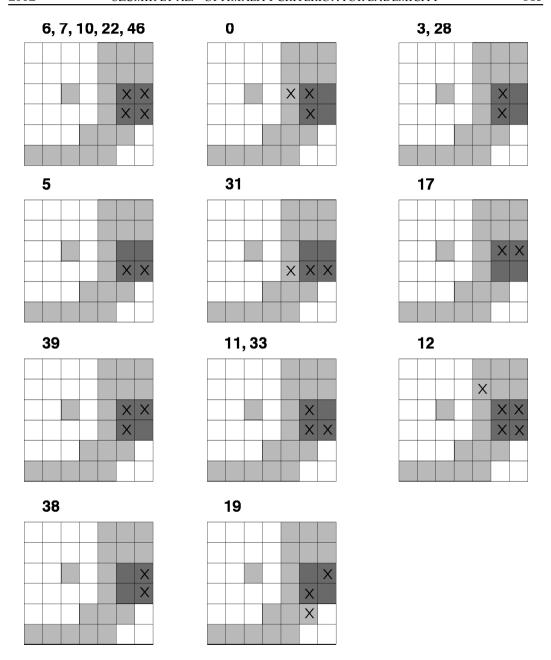
FIGURE 8.   Distributions of the 17 species endemic to area 1 in Figure 7, according to criterion 4.

fragmentation (due to many possible causes) cannot be recognized as such. It would be desirable to modify the criteria in such a way that disjunct areas can be recognized. Modifications of the criteria for meaningful evaluation of disjunct areas are currently being investigated.

Another aspect that should be improved is the all-or-none aspect of the method; a given species either contributes to the score or it does not. Ideally, species that adjust well to the expectation of endemicity should contribute to the score more than species that adjust poorly (in a proportion that depends on how well the species adjust to endemicity). A possibility is to weight a species according to the proportion of cells in the area that are effectively occupied by the species or by the ratio of occupied cells inside and outside the area, or by both methods.

Aside from those possible improvements, a better insight into the properties of the method can be gained by testing the method on randomly generated distributional data. Another aspect that must be studied more closely is the effect of the grid cell size on the results (for a brief discussion, see Morrone, 1994). More detailed analyses along these lines are currently being carried out, and their results will be published elsewhere.

REFERENCES

ALLARD, M., J. FARRIS, AND J. CARPENTER. 1999. Congruence among mammalian mitochondrial genes. Cladistics 15:75–84.

BROOKS, D. R. 1990. Parsimony analysis in historical biogeography and coevolution: Methodological and theoretical update. Syst. Biol. 39:14–30.

FAITH, D. P. 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61:1–10.

GOLOBOFF, P. A. 2001. NDM and VNDM: Programs for analysis of endemicity. Distributed by the author, San Miguel de Tucumán, Tucumán, Argentina.

GOLOBOFF, P. A. 2002. Optimization of polytomies: State set and parallel operations. Mol. Phylogenet. Evol. 22:269–275.

HAROLD, A. S., AND R. D. MOOI. 1994. Areas of endemism: Definition and recognition criteria. Syst. Biol. 43:261–266.

LINDER, P. 2001. On areas of endemism, with an example from the African Restionaceae. Syst. Biol. 50:892–912.

MOILANEN, A. 1999. Searching for most parsimonious trees with simulated evolutionary optimization. Cladistics 15:39–50.

MORRONE, J. J. 1994. On the identification of areas of endemism. Syst. Biol. 43:438–441.

NELSON, G., AND P. LADIGES. 1996. Paralogy in cladistic biogeography and analysis of paralogy-free subtree. Am. Mus. Novit. 3167:1–58.

NELSON, G., AND N. I. PLATNICK. 1981. Systematics and biogeography: Cladistics and vicariance. Columbia Univ. Press, New York.

PAGE, R. D. M. 1994. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. Syst. Biol. 43:58–77.

PLATNICK, N. I. 1991. On areas of endemism. Aust. Syst. Bot. 4:xi–xii.

PLATNICK, N. I. 1992. Patterns of biodiversity. Pages 15–24 in Systematics, ecology, and the biodiversity crisis (N. Eldredge, ed.). Columbia Univ. Press, New York.

PRESSEY, R. L., C. J. HUMPHRIES, C. R. MARGULES, R. I. VANE-WRIGHT, AND P. WILLIAMS. 1993. Beyond opportunism: Key principles for systematic reserve selection. Trends Ecol. Evol. 8:124–128.

RODRIGUES, A., J. ORESTES CERDEIRA, AND K. GASTON. 2000. Flexibility, efficiency, and accountability: Adapting reserve selection algorithms to more complex conservation problems. Ecography 23:565–574.

RONQUIST, F. 1997. Dispersal–vicariance analysis: A new approach to the quantification of historical biogeography. Syst. Biol. 46:195–203.

VANE-WRIGHT, R., C. J. HUMPHRIES, AND P. H. WILLIAMS. 1991. What to protect—systematics and the agony of choice. Biol. Conserv. 55:235–254.

WILLIAMS, P. H. 1996. WORLDMAP 4: Program and documentation. Distributed by the author, www.nhm.ac.uk/science/projects/worldmap