

COMPONENT-COMPATIBILITY IN HISTORICAL BIOGEOGRAPHY

M. ZANDEE¹ AND M. C. ROOS²

¹*Institute for Theoretical Biology, University of Leiden, The Netherlands.*

²*Institute for Systematic Botany, University of Utrecht, The Netherlands.*

Abstract—The problems of reconstructing historical relationships for areas of endemism from distributional data for groups of taxa and the cladistic relationships among the members of those groups can be solved by applying the two principles of parsimony and mutual inclusion or exclusion (compatibility) of components. Components can be extracted from a data matrix by means of transcription into partial monothetic sets. The data matrix thus derived represents the distribution over areas for the monophyletic groups in one or more cladograms. It is derived from two different matrices by boolean multiplication. The first matrix gives the binary representation of distributions of taxa over areas of endemism; the second describes the cladogram for the same taxa, in terms of character states converted into binary form by additive binary coding. The derived data matrix can be used in historical biogeography to represent the given phyletic data (*Assumption 0* here newly defined), and can be amended to reflect *Assumptions 1* or *2* to accommodate the problems of wide-spread taxa and missing areas. Area-cladograms are determined from the derived matrix by searching for the largest sets of mutually compatible components. Area-cladograms are evaluated in terms of support (vicariance) and contradiction (ad hoc interpretations such as dispersal and extinction). Area-cladograms that best fit the data matrix regarding the balance between support and contradiction are selected as the best possible reconstructions of relationships among the areas of endemism. The procedure is illustrated by the example of the poeciliid fish genera *Heterandria* and *Xiphophorus*, and several other standard examples.

Introduction

Historical biogeography assumes that there is a correspondence between species-relationships and area-relationships. Comparisons between the cladistic relations within various groups of organisms occurring in a certain region might elucidate general patterns which can be used to develop hypotheses on the historical relationships of biota (Rosen 1978, Nelson and Platnick 1981, Humphries 1982, Humphries and Parenti 1986).

For this purpose, species cladograms are transformed into area-cladograms by replacing each species represented at the terminals with its distributional area (described in terms of the smallest relevant biogeographical entity, the area of endemism). Congruencies between area-cladograms are then depicted in general area-cladograms, which represent the general patterns, the cladistic relationships between areas of endemism.

At present two types of analysis with different purposes and different lines of reasoning can be recognized in historical biogeographic research. The first type aims at reconstructing the historical development of biota ('cladistic biogeography'). For this purpose, the delimitation of areas of endemism is a starting-point (Humphries and Parenti 1986: 1). The main problems concern the manipulation of wide-spread species and the absence of groups from one or more areas under consideration. The methods of Rosen (1976) and Nelson and Platnick (1978, 1981) belong here. The second type concerns the search for data for constructing evolutionary scenarios of particular groups under study (e.g. 'vicariance biogeography'). Here, emphasis is put on vicariance patterns as a framework for postulating speciation models. An essential element is the search for geographic separation between sister taxa (Wiley and Mayden 1985: 598-600). In this case the main problem is coping with overlapping ranges. Wiley's 'ancestral species map' method belongs to this category, as also his parsimony method outlined recently (Wiley, in press).

To date, the general treatments of cladistic biogeography (Nelson and Platnick 1981,

Humphries 1982, Humphries and Parenti 1986) have been concerned with obtaining unambiguous area-cladograms. All authors discuss the problem in terms of examples in which species occur in two or more areas of endemism, or in which the groups under study is absent from one area under consideration. Nelson and Platnick (1981), for example, concentrate on the transformation of 3-species-statements (cladograms), including at least one wide-spread species, into 3 or 4 area statements (area-cladograms). The next step, of converting several area-cladograms into general area-cladograms, still needs further elucidation (e.g. the unexplained jumps from Figs. 8.25-27 to 8.28 in Nelson and Platnick, 1981, or from Figs. 2.30 to 2.31 in Humphries and Parenti, 1986).

This paper aims to provide a method for constructing and selecting general area-cladograms based on the principles of group-compatibility and parsimony. Both principles have been used earlier by Zandee (1985) to develop a method for cladistic analysis, recently applied by Geesink (1984), Roos (1985, 1986), and Zandee and Geesink (1987).

We shall argue that the steps from cladograms to area-cladograms and area-cladograms to general area-cladograms are ruled by the same methodological principles. The difference between the two steps is one of degree, not of kind. Therefore, both steps can be integrated into one, general procedure. The intricacies of this procedure will be exemplified in a biogeographical reconstruction of the Poeciliid fishes (see Rosen 1978, Platnick and Nelson 1978, Wiley 1981, Mickevich 1981, Humphries and Parenti 1986). We will show also that some identical examples given by Humphries (1982) and by Humphries and Parenti (1986) lead to conflicting conclusions because, to our knowledge, these authors neither integrate the two steps nor make the procedure for the latter step explicit.

General Area-Cladograms

The cladistic analysis can be summarized as four steps:

1. DATA MATRIX

The data matrix shows the distribution pattern of the various intrinsic character states amongst the terminal taxa. These character states may be coded as unordered (neutral), uniquely ordered (additive binary), or exhaustively ordered (all additive combinations) with respect to the sequence of their transformations, and the data matrix may consist of one or either type or a mixture.

2. MONOTHETIC SETS

Character states, either separately or in combination, uniquely represented by a group of terminal taxa define this group as a cladogenetic unit and a building block for cladograms (cladon: i.e. a set of terminal taxa, without rank, name, or phyletic structure). Clada are extracted from the data matrix by means of transcription into monothetic sets (Beckner, 1959; Sharrock and Felsenstein, 1975; Farris, 1978). When a cladon is defined in terms of *unique* character states only then is it called a *partial* monothetic set. When it is defined in terms of a *unique combination* of character states, none of which need be unique, but some may be so, then it is called a *strict* monothetic set (Zandee, 1984). Character state distributions defining the same cladon represent a character type (*sensu* Nelson and Platnick 1981). The rationale behind the use of monothetic sets of taxa as building blocks for cladograms is the idea that a series of homologous character states (i.e. a character) must be kept as an integrated whole as long as this is tenable in the context of parsimony, rather than broken up into different characters (a non-homologous series of states). In other words, convergence is rejected as a first level explanation for similarities observed among taxa. This implies that we try not to fit a minimal number of singular step to a tree, but build a tree from sets of integrated series of steps.

3. CLIQUES

A search is made for the largest clique present in the list of clada. The recognition of cliques is based on the concept of group-compatibility (not character compatibility). Cladons are compatible with other cladons when they include or exclude each other. These inclusion and exclusion relations can be represented by a graph (network). In the case when all clada in a set are mutually compatible, the set is a clique. In the graph representing the relations, the sets in the clique are all mutually connected. Clique as used here in its original unambiguous meaning is a concept from graph theory, and stands for a maximal complete subgraph (Garey and Johnson 1979). This concept, and other elements from graph theory, are used in the implementation of the method in computer algorithms.

4. CLADOGRAMS

The largest cliques are transcribed into cladograms. This is followed by interpreting and evaluating their implications in terms of the character states in the data matrix and choosing one or a few of them. For this purpose two criteria are used in association. The first (quantitative) criterion is the maximization of the value of support minus contradiction (or, equivalently, minimization of the value of contradiction minus support). Support means the number of fitting character states, i.e. those in a distribution pattern which can be explained by assuming a single origin (synapomorphies). Contradiction (homoplasy) means the number of independent multiple origins and/or reversals of the remaining character states. The second criterion, used supplementarily to the first, pertains to the total number of synapomorphies for each cladogram as determined by local outgroup comparison in the context of all its 3-cladon statements (Zandee, 1984). The latter analysis might elucidate synapomorphies among those states with multiple origins which were homoplasious in the first instance. As a consequence these states are considered to belong to *different* transformation series.

Eventually, one or several cladogram(s) are selected on the basis of these criteria as the most likely representation(s) of a phylogeny.

General area-Cladogram Construction

Humphries (1982: 445) stated that cladistic biogeography is the pursuit of a method which encompasses a code comparable to cladistics. The analogy involves exchanging species with areas of endemism and homologies with sister groups. Following this analogy, the method of historical biogeography proposed here can be summarized in a way similar to the cladistic analysis described above. However, it must be noted that biogeographic analyses can be carried out under three assumptions, i.e. *Assumption 0*, *Assumption 1*, and *Assumption 2* (see also Wiley, in press). The latter two were proposed by Nelson and Platnick (1981), but they did not give full credit to the possibilities of *Assumption 0*.

Assumption 0 begins with the idea that the reconstructed phylogeny is the best possible estimate of the true phylogeny. *Assumption 0* implies that if a species appears to be wrongly delimited and actually represents two or more species, the resulting species are assumed to be sister groups and should together represent a monophyletic group. Consequently, the areas inhabited by each wide-spread species represent undisputable components. This means that in Fig. 1d the identity of species 1 might be misconstrued but not its phylogenetic relations, and that from the distribution of species 1 the component, A+B can be derived (Fig. 1c).

Under *Assumption 1* (Nelson and Platnick 1981: 421), it is assumed that whatever is true of a wide-spread taxon in one part of its range must also be true of the taxon in other parts of its range (compared to other areas inhabited by the other species of the monophyletic group).

In Fig. 1a the wide-spread species 1 (occurring in areas A and B) is more closely related to species 2 (occurring in area C) than to species 3 (occurring in area D). *Assumption 1* implies that area A is more closely related to area C than to D and that area B is more closely related to area C than to area D. This means that if species 1 in the future is regarded as actually representing more than one species, the newly delimited species should be sister species, or branch off from the cladogram sequentially (Platnick 1981: 224). The same holds for the areas they inhabit (Fig. 1c-e). Thus, the wrongly delimited species might actually comprise a mono- or paraphyletic group, and the component defining A+B is in doubt, but the component defining A+B+C is considered indisputable (Fig. 1b).

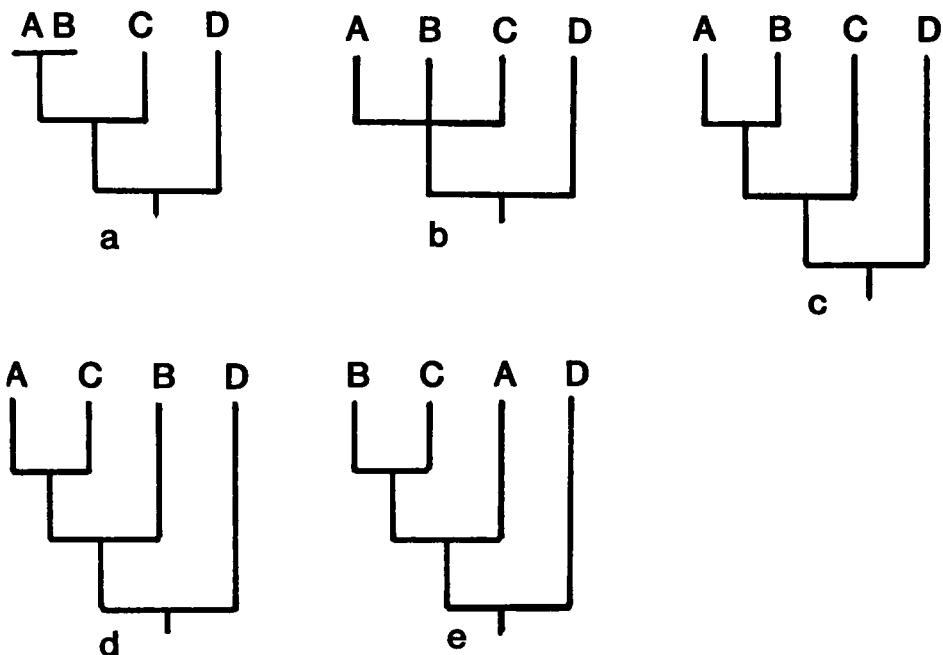


Fig. 1(a). Cladogram showing the relations between species 1, 2, and 3, and their distribution areas (A-D). (b). consensus area-cladogram under *Assumption 1*. (c-e) fully resolved area-cladograms implied by the consensus tree. (c) comprises the components derived under *Assumption 0*.

Assumption 2 (Nelson and Platnick 1981: 432) indicates that whatever is true of a widespread taxon in one part of its range might not be true of the taxon in other parts of its range.

The implications of this assumption are illustrated in Fig. 2. In Fig. 2a, *Assumption 2* implies that area A is more closely related to area C than to D (Fig. 2b) and/or that area B is more closely related to area C than to area D (Fig. 2c), but this is not necessarily true for both areas A and B. If the relations of A compared to C and D are true, then B can take one of all possible positions (white circles in Figs. 2b, c) on the area-cladogram (Fig. 2b). If the relations of B compared to C and D are true, A can take one of all possible positions to branch off in the area-cladogram (Fig. 2c). In other words, both components A+B and A+B+C may be in doubt. This means that both the identity and the relationships of species 1 might be misconstrued. Furthermore, this wrongly delimited

species might actually comprise a mono-, para-, or polyphyletic group with all the consequences for the resulting area-cladograms.

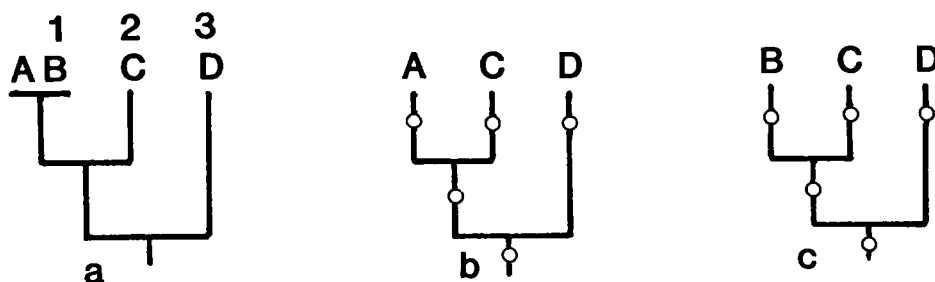


Fig. 2(a). Cladogram showing the relations between the species 1, 2, and 3, and their distribution areas A—D. (b,c) the two area-cladograms required for the application of *Assumption 2* derived from the cladogram. O indicates the positions which the missing areas (B and A, respectively) can take.

DATA MATRIX

The derivation of a data matrix depends on which of the three assumptions are made. This paper is not primarily concerned with a full evaluation of these assumptions, but it provides the procedures to derive the data matrices appropriate for analyses under each of the three assumptions.

Data matrix under Assumption 0

The data matrix comprises the distribution patterns in terms of areas of endemism of the terminal taxa and all corresponding monophyletic groups present in the cladograms (clada \times areas). This data matrix is derived from two sets of raw data in a stepwise manner. First, for each group of taxa we have a binary matrix with distributional data for taxa, comprising only terminal taxa and areas (taxa \times areas). Second, we have a cladogram for each study group of taxa, converted into a binary matrix by additive binary coding (taxa \times nodes; Farris et al. 1970). The combination (boolean inner-product) of each distributional matrix with each cladogram matrix gives a biogeographical data matrix (areas \times clada from one cladogram). Each column in the final data matrix represents the distribution over a terminal taxon or monophyletic group for each group of analyzed taxa. Identical distributions belong to the same distributional type (character type: Nelson and Platnick, 1981). From this data matrix area-cladograms can be derived under *Assumption 0*. When the same set of areas obtains for several groups of taxa, the resulting separate data matrices can be joined (column wise) to form a larger data matrix (areas \times clada from several cladograms) from which general area-cladograms can be derived.

A data matrix to derive a general area-cladogram for several groups of taxa can also be obtained via a kind of feed-back loop. For each separate group of taxa area-cladograms are extracted and evaluated. During this evaluation one area-cladogram is selected for each group of taxa. The distributional types corresponding to all of the nodes of this area-cladogram are selected from the data matrix based on *Assumption 0* in order to obtain a reduced data matrix. Reduced matrices for each group of taxa can be then joined together to form a compound reduced matrix suitable for analysis to derive a general area-cladogram.

The derivation of a data matrix under *Assumption 0* is equivalent to the procedure

described by Brooks (1981: 232–234) which he used to extract a data matrix from parasite distributions over hosts and parasite cladograms in order to infer host relationships from the parasite data. In fact we suggest that the method described here is just as suitable to derive general host cladograms from parasite distributions over hosts and parasite cladograms by simply substituting hosts with areas.

Data matrix under Assumption 1

The derivation of a data matrix fit for an analysis under *Assumption 1* is more complex. We start from the same two sets of raw data, i.e. the distributions for terminal taxa and their cladogram(s). A combination between the two is made as described for *Assumption 0*, but with one proviso; the distributions that correspond with all possible subsets of areas of each wide-spread taxon (occupying 2 or more areas) are combined with the areas of its sister group. These combinations together with the subset distributions are scored in the data matrix. In the actual analysis, this proviso will cause two possible results which together represent the principle of *Assumption 1* i.e. the areas occupied by a wide-spread taxon might either branch off sequentially or occur as sister areas in the area-cladograms.

The derivation of general area-cladograms for several groups of taxa can then proceed on the basis of joined data matrices in the same way as described for *Assumption 0*. When the data matrix for a general analysis is obtained via a feed-back loop, the provisions for wide-spread taxa were already made in the analyses for the separate groups of taxa, thus only selected distributional types are joined as columns.

Data matrix under Assumption 2

The derivation of a data matrix suitable for an analysis under *Assumption 2* is even more complex. Starting from the same two sets of raw data a combination between the two is made in the same way as described above for *Assumption 0* but now followed by two extra steps.

In the first step, we combine the distributions corresponding with all possible subsets of areas of each wide-spread taxon with the distributions of *all* other monophyletic groups in the cladogram(s). These combinations as well as those from the subset distributions themselves are added to the data matrix.

In the second step, we take all the separate missing areas and all of their possible combinations and we combine each of these with all of the columns of a duplicate of the data matrix derived under *Assumption 0*. We take a separate duplicate for each of these operations. To compile the complete data matrix we consequently join all of the duplicates, now amended for missing areas, with the original as amended for wide-spread taxa in the first step. In the actual analysis, these steps allow for many possible results as implied by *Assumption 2* (see above and Figs. 2, 4), i.e. all but one of the areas occupied by a wide-spread taxon might branch off from any other branch in the area-cladogram. From *Assumption 2*, as defined by Nelson and Platnick (1981) and exemplified by their thought experiments (ibid. p. 462 et seq.), it does *not* follow that distributions of taxa over areas should be weighted differentially including a possible zero weight. This means that all distributions of taxa over areas implied by this assumption must be included in the data matrix which must be considered as a whole. The matrix should not be broken down into several parts prior to analysis because of incongruencies present within it, a principle which also applies to *Assumption 1*.

Although in this implementation of *Assumption 2* the data matrix is expanding rapidly, as compared with a matrix analysed under *Assumption 0*, it is conservative nevertheless. For a more elaborate implementation, in the second extra step we can amend for missing areas on the basis of the data matrix as derived in the first extra step, i.e. after amending

for wide-spread terminal taxa. Neither Nelson and Platnick (1981) nor Humphries and Parenti (1986) indicated which implementation should be preferred, because in their examples they deal either with missing areas or with wide-spread taxa but not with both simultaneously.

The data matrix for a general analysis can be obtained by joining the data matrices for the separate groups of taxa. It can also be obtained via a feed-back loop, i.e. after the evaluation and selection of an area-cladogram for each separate group of taxa (see appendix). In that case the provisions for wide-spread taxa and missing areas were already made in the analysis of each separate group of taxa, thus only selected distributional types are joined together.

It should be noted that the coding used in the data matrix is necessarily unordered (or neutral) with respect to areas. Any kind of additive coding is superfluous because the clada among themselves already constitute interested sets of terminal taxa. If areas are to be connected in a sequence, it is because the terminal taxa or groups of taxa occurring in these areas are connected in a hierarchical sequence provided by the cladograms.

COMPONENTS

In this step the data matrix is transcribed into a list of (partial) monothetic sets of areas. This transcription is equivalent to the derivation of components (Nelson and Platnick, 1981). A component (analogous to a cladon) is characterized by a particular distribution. Components serve as building blocks for area-cladograms as well as for general area-cladograms (as do clada for cladograms).

CLIQUEs

Inclusion and exclusion relations (compatibilities) are recognized among the components and maximal cliques are sought in the network expressing these relations. Cliques comprise components which mutually include or exclude each other.

AREA-CLADOGRAMS AND GENERAL AREA-CLADOGRAMS

The maximal cliques are transcribed into area-cladograms or general area-cladograms. If only one group of taxa is used to compile the data matrix, one or more area-cladograms will be obtained. If more than one group of taxa is used in a compound matrix (based on several cladograms pertaining to the same areas) general area-cladograms will be obtained. As more unrelated groups of taxa are involved, the resulting general area-cladograms will be more 'general'.

These diagrams can then be interpreted and evaluated in terms of vicariance, primitive absence, dispersal, and extinction with regard to every taxon involved in the analysis. Synapomorphy, in a strict sense, as used in cladistic analysis, does not really apply to character states of areas or groups of areas. General area-cladograms may share monophyletic groups of taxa, but not *intrinsic* characters, at least not directly since areas do not have members with genes by which their descendants inherit and consequently share monophyletic groups of taxa. It follows that 'synapomorphy' or homology in biogeography is a monophyletic group of taxa uniquely sharing a group of areas as a consequence of shared history. Indicators of non-shared history, such as dispersal and extinction, are ad-hoc statements and thus are analogues of homoplasy. Evaluation by means of the parsimony criterion leads to a choice for one or a few area-cladograms.

Columns of the data matrix indicating monophyletic groups responding to vicariance events fit the area-cladogram and as such represent support. Columns indicating dispersal or extinction do not fit the area-cladogram and represent contradiction. As in cladistic

analysis, character states on internal nodes of an area-cladogram are estimated by optimization (Farris, 1970). After optimization the number of state changes for each column of the data matrix can be computed for each area-cladogram. All columns with zero or one state change represent support. Character states which are present from the root onwards are considered to have zero changes. Choosing an appropriate outgroup might tell us whether the state really originated at the root. Until this can be examined such states are taken as support without assuming a single origin at the root. Columns indicating multiple parallel origins and reversals, represent contradiction. Contradiction minus support is chosen as a measure to express the degree of best fit for area-cladograms with respect to the data matrix, as it serves as a parsimony criterion for the evaluation of area-cladograms. This criterion is chosen in preference to the total number of state changes (steps) because it enables us to make a distinction between area-cladograms of equal length.

We question whether all of the columns in the data matrix should be used as a means of evaluating area-cladograms by the parsimony criterion. We can certainly use them all in *Assumption 0* when all columns, from the respective cladograms, are based on real empirical data. For *Assumptions 1* or *2* it is doubtful that the extra columns refer to real data but to assumptions expressing doubt with respect to some of the data used to reconstruct phylogenies. It is our opinion that only real observations rather than assumptions that can be used to evaluate hypotheses. This includes hypotheses regarding biogeographical events in terms of parsimony. This point of view is not without practical consequences. It implies that if an analysis using *Assumption 0* on at least two groups of taxa results in at least one fully resolved general area-cladogram, this general area-cladogram will also prove to be among the best ones resulting from an analysis with *Assumptions 1* or *2*. We show this in an example dealing with the poeciliid fish (p. 317).

As a consequence analysis under *Assumption 1* or *2* may be unneeded if an analysis under *Assumption 0* already results in fully resolved general area-cladograms, for it can only produce more equally parsimonious explanations. It is only when analysis with *Assumption 0* produces general area-cladogram(s) that are not completely dichotomous, that further analyses under *Assumptions 1* or *2* make any sense.

There is yet another problem when analysing all columns at once using *Assumption 0*. Given a contradiction regarding a certain column in the data matrix with respect to a particular area-cladogram, the same contradiction will occur in all other columns that include the areas indicated by the affected column. It follows that contradictions shown by a data matrix with regard to a particular area-cladogram may not be independent. Nevertheless, they are all counted in the inventory of state changes. This may result in exaggerating the degree of bad fit, especially in area-cladograms in which small (i.e. less inclusive) monophyletic groups fail to respond to vicariance events. The problem is caused by the nested mutual interdependence of columns referring to monophyletic taxa, i.e. the internal nodes of a cladogram. Columns referring to terminal taxa are not affected because these distributional types always fit an area-cladogram. As yet, no solution to this problem is implemented in the algorithms used so far.

Results

In the analyses given here, two elements are emphasized. First, we explore the relationships of components with respect to their compatibilities in order to compile area-cladograms or general area-cladograms. Second, we interpret area-cladograms or general area-cladograms in terms of parsimony, i.e. the balance of implied support and contradiction. Monophyletic taxa responding to vicariance events imply support for the area-cladogram; ad hoc hypotheses necessary to explain distribution in terms of dispersal or extinction imply contradiction.

As a corollary, we note that a chosen general area-cladogram can be used to interpret the area-cladograms in terms of incongruencies shown between them. Congruency of area-cladograms with the general area-cladogram can be interpreted as due to a common cause, i.e. analogous to synapomorphy. Incongruency is analogous to homoplasy.

However, it must be stressed that vicariance is not the only mechanism for explaining biogeographic patterns. As a consequence, the analogy implies that 'positive' incongruencies (occurrences incongruent with the general area-cladogram) might be caused by dispersal and the 'negative' incongruencies (the absence of taxa in particular areas i.e. 'reversals') might be evidence of extinctions or primitive absence. We use this evaluation for general area-cladograms in cases from the literature, where original data matrices are unavailable.

CRITICAL EXAMPLES UNDER ASSUMPTION 1

The problems we encountered using *Assumption 1* and 2 are illustrated by comparing some hypothetical examples discussed by Humphries (1982) and Humphries and Parenti (1986).

Figure 3

Figures 3a, b, e present Figs. 13.iii, iv of Humphries (1982 and Figs. 2.28.c and 2.28.d of Humphries and Parenti, 1986. The coding is modified as follows: AUS (Australia) = A, NG (New Guinea) = B, SA (South America) = C, and AF (Africa) = D. It is clear that the conclusion of Humphries (1982, his Fig. 13.v = our Fig. 3d) differs from that of Humphries and Parenti (1986, their Fig. 2.28e = our Fig. 3e). Moreover, in both cases the conclusions are counter-intuitive. Systematists would be pleased to find character states among A, B, C and D, one that defines A+B (cf. their component 3) and another for defining A+B+D (cf. their component 0). This would yield a fully resolved cladogram (Fig. 3e).

Taking *Assumption 1* (Nelson and Platnick, 1981) for granted, this example can be analysed using the method outlined in the previous paragraph. Area-cladogram 1 (Fig. 3a) yields the components given in Table 1. One component (A+B) is undisputable. The other three are disputable in the sense that they represent components from different (alternate) area-cladograms. Fig. 3a represents the consensus situation harmonizing the contradictions among these components. Consequently, from these remaining three components only one can be supported by this area-cladogram in the final general area-cladogram. In a similar way, area-cladogram 2 (Fig. 3b) generates the components given in Table 2.

Taken together, Table 3 lists the components, including the sequence number of their supporting area-cladograms (u = undisputable). From this list, the cliques present are given in Table 4.

Evaluation of these four possible cliques (general area-cladograms) with regard to support minus contradiction leads to the choice for clique 2 (general area-cladogram Fig. 3c) and differs from the conclusions given by Humphries (1982; our Fig. 3d as well as Humphries and Parenti (1986; our Fig. 3e).

Figure 4

Figures 4a-c represent Fig. 14.iv-vi of Humphries (1982; = 2.29.d-f of Humphries and Parenti 1986; see coding in Fig. 3). In this case also, the conclusions reached by both publications differ (his Fig. 14.vii = our Fig. 4g and their Fig. 2.29g = our Fig. 4h, respectively).

The three area-cladograms (Figs. 4a-c) yield the components presented in Table 5 (in this case also, one component is undisputable, whereas of the other three only one

can be supported by the respective area-cladograms). Taken together, the list of the components is given in Table 6. From this list, eight maximal cliques are available (Table 7).

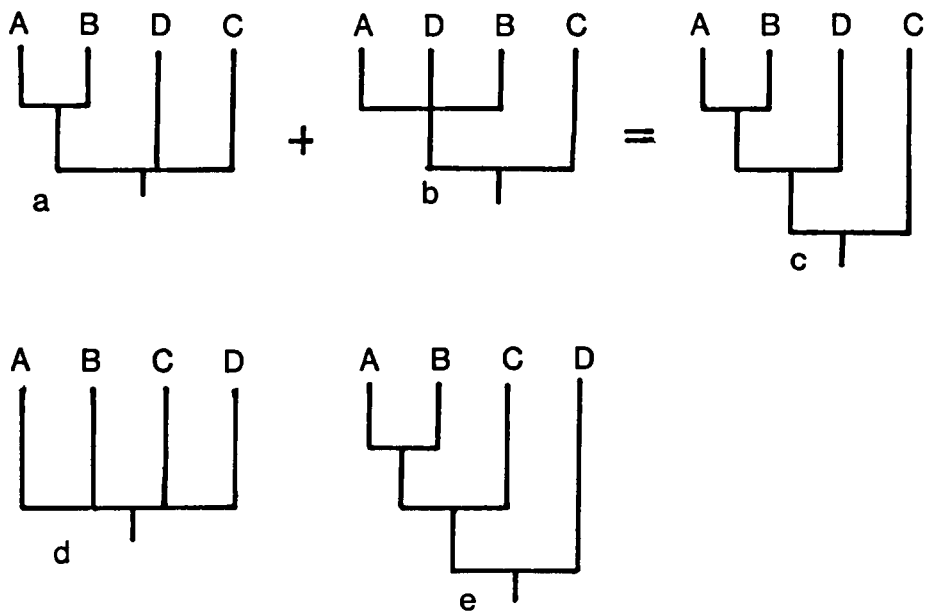


Fig. 3(a). Consensus area-cladogram based on a particular monophyletic group. (b). consensus area-cladogram for a different group. (c). general area-cladogram derived from these two area-cladograms under *Assumption 1*. (d). general area-cladogram presented by Humphries (1982) for the same data. (e). general area-cladogram presented by Humphries and Parenti (1986) for the same data.

Table 1.
Components derived from Fig. 3a under Assumption 1.

A + B	undisputable
A + B + C	
A + B + D	only one
C + D	

Table 2.
Components derived from Fig. 3b under Assumption 1.

A + B + D	undisputable
A + B	
A + D	only one
B + D	

General area-cladograms 1 and 2 (Fig. 4d and 4e) are alike with regard to support minus contradiction. General area-cladogram 3 is second best (Fig. 4f). These three

possibilities are implied in the conclusion of Humphries and Parenti (1986, Fig. 2.29g). However, they do not recognize the difference in likelihood. Moreover, Humphries (1982: Fig. 14vii) concludes that all 15 possible general area-cladograms (of these four areas) are equally likely. Using the present analysis we fail to confirm this conclusion.

Table 3.
Components derived from Fig. 3a, b under Assumption 1.

A + B 1(u), 2	B + D 2	A + B + C 1
A + D 2	C + D 1	A + B + D 1, 2(u)

Table 4.
Cliques from the components derived from area-cladograms in Fig. 3a and 3b. "+" = support: component found in one of the original area-cladograms. "-" = contradiction: component absent in one of the original area-cladograms.

Component	Support	Contradiction	Summary
1 A + B 1 x u;	2 x +		
A + B + C	1 x +,	1 x -	3 x +, 1 x -
2 A + B 1 x u;	2 x +		
A + B + D 1 x u;	2 x +		4 x +
3 A + D	1 x +,	1 x -	
A + B + D 1 x u;	2 x +		3 x +, 1 x -
4 B + D	1 x +,	1 x -	
A + B + D 1 x u;	2 x +		3 x +, 1 x -

Table 5.
Components derived from Fig. 4a-c under Assumption 1.

5.1 A + B u	5.2 A + B	5.3 A + B
	A + C	A + D
	B + C	B + D
A + B + C	} only one	
A + B + D		
C + D	A + B + C u	A + B + D u

EXAMPLES UNDER ASSUMPTION 2

The next example (Fig. 5), determined under *Assumption 2* is also taken from Humphries (1982: Fig. 14.i—iii). For each of the given area-cladograms (Fig. 5a-c), the components found are presented in Table 8 (none of them being undisputable). Taken all together, the components are presented in Table 9. Within this list all 15 cliques possible for four areas are found, only one of which is presented by Humphries (1982) and Humphries and Parenti (1986). Moreover, these authors do not mention alternative solutions. The distribution of support and contradiction for the general area-cladogram is given in Table 10. The general area-cladogram in Fig. 5d is the most likely hypothesis as it lacks (unambiguous) contradiction. There are seven possibilities showing 4 as the value of support minus contradiction.

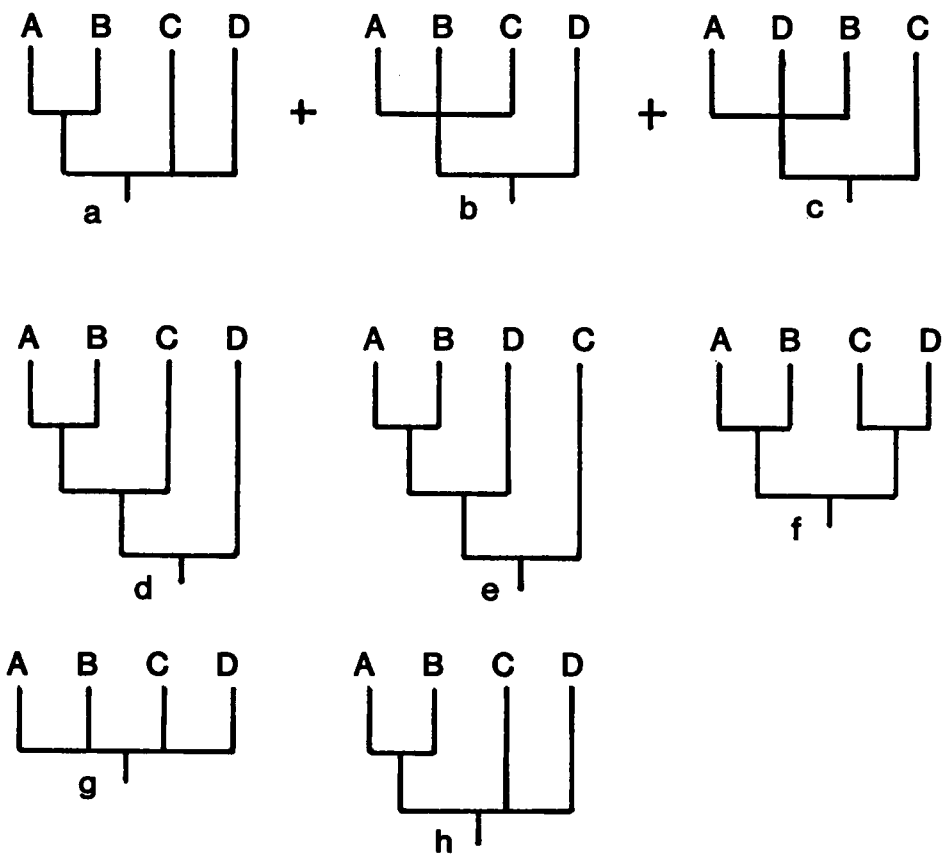


Fig. 4(a-c). Three consensus area-cladograms of three different monophyletic groups; (d,e). the two best general area-cladograms derived from these area-cladograms under *Assumption 1*. (f). the second best general area-cladogram under *Assumption 1*. (g). general area-cladogram presented by Humphries (1982) for the same data and assumption. (h). general area-cladogram presented by Humphries and Parenti (1986) for the same data and assumption.

Table 6
Components derived from Fig. 4a-c under Assumption 1.

A + B 1 (u), 2, 3	B + C 2	A + B + C 1, 2(u)
A + C 2	B + D 3	A + B + D 1, 3(u)
A + D 3	C + D 1	

What we have shown here is that by building from a list of possible components present in, or on the basis of additional assumptions derived from each area-cladogram, and judging their compatibilities and building general area-cladograms, we can arrive at one (or a few) likely general area-cladograms on the basis of a parsimony criterion. A comparable, equally explicit procedure cannot be derived from the explanations given by Nelson and Platnick (1981), Humphries (1982), and Humphries and Parenti (1986), especially in those examples in which alternate, conflicting possibilities are comprised into one (or a few) general area-cladogram(s).

Table 7.
Cliques from the components derived from area-cladograms given in Fig. 4a-c.

1 A+B	1x u;	3x +	
A+B+C	1x u;	2x +, 1x -	5x +, 1x -
2 A+B	1x u;	3x +	
A+B+D	1x u;	2x +, 1x -	5x +, 1x -
3 A+B	1x u;	3x +	
C+D		1x +, 2x -	4x +, 2x -
4 A+C		1x +, 2x -	
A+B+C	1x u;	2x +, 2x -	3x +, 3x -
5 A+D		1x +, 2x -	
A+B+D	1x u;	2x +, 1x -	3x +, 3x -
6 B+C		1x +, 2x -	
A+B+C	1x u;	2x +, 1x -	3x +, 3x -
7 A+C		1x +, 2x -	
B+D		1x +, 2x -	2x +, 4x -
8 A+D		1x +, 2x -	
B+C		1x +, 2x -	2x +, 4x -

Table 8
List of components derived from area-cladograms given in Fig. 5a-c.

5 (a)	5 (b)	5 (c)
A+B	A+B	A+B
A+C	A+C	A+C
A+D	A+D	A+D
B+C	B+C	B+D
B+D	B+D	C+D
C+D	A+B+C	A+B+C
A+B+C	A+C+D	A+B+D
A+B+D	B+C+D	B+C+D
(Neither A+C B+D nor A+D B+C can occur together)		

Table 9
Components derived from Fig. 5a-c under Assumption 2.

A+B 1, 2, 3	B+C 1, 2	A+B+C 1, 2, 3
A+C 1, 2, 3	B+D 1, 2, 3	A+B+D 1, 3
A+D 1, 2, 3	C+D 1, 3	A+C+D 2
		B+C+D 2, 3

ROSEN'S POECILIID FISHES AS A BENCHMARK

Assumption 0

The theoretical and empirical implications of *Assumption 0*, in terms of the complexity

of the data matrix and of the number of area-cladograms to be evaluated, are simpler than those of the other two assumptions. The analysis has been carried out without area 11 (for coding see Rosen, 1976) because it is assumed to be of hybrid origin (Wiley 1981).

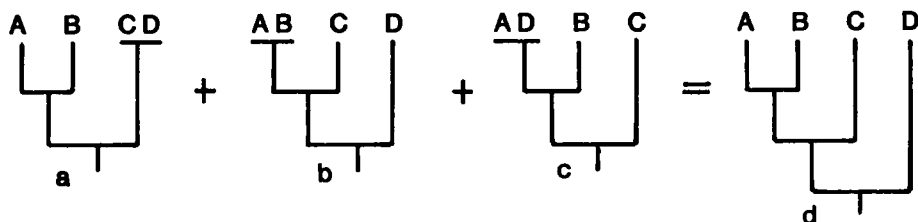


Fig. 5(a-c). Three area-cladograms of three different monophyletic groups; (d), the best general area-cladogram derived from these area-cladograms under *Assumption 2*.

Table 10.
Cliques based on the components derived from the area
cladograms in Fig. 5a-c.

1	A + B	3 x +,	
	A + B + C	3 x +	6 x +
2	A + C	2 x +, 1 x -	
	A + B + C	3 x +	5 x +, 1 x -
3	A + B	3 x +	
	C + D	2 x +, 1 x -	5 x +, 1 x -
4	A + B	3 x +	
	A + B + D	2 x +, 1 x -	5 x +, 1 x -
5	A + C	2 x +, 1 x -	
	A + C + D	2 x +, 1 x -	4 x +, 2 x -
6	A + C	3 x +	
	B + D	2 x +, 1 x -	
	or	2 x +, 1 x -	
		3 x +	5 x +, 1 x -
7	A + D	3 x +	
	A + B + D	2 x +, 1 x -	5 x +, 1 x -
8	A + D	3 x +	
	A + C + D	1 x +, 2 x -	4 x +, 2 x -
9	B + C	1 x +, 2 x -	
	A + D	3 x +	
	or	2 x +, 1 x -	
		2 x +, 1 x -	4 x +, 2 x -
10	B + C	2 x +, 1 x -	
	A + B + C	3 x +	5 x +, 1 x -
11	B + C	2 x +, 1 x -	
	B + C + D	1 x +, 2 x -	3 x +, 3 x -
12	B + D	3 x +	
	A + B + D	2 x +, 1 x -	5 x +, 1 x -
13	A + D	3 x +	
	A + C + D	1 x +, 2 x -	4 x +, 2 x -
14	C + D	2 x +, 1 x -	
	A + C + D	1 x +, 2 x -	3 x +, 3 x -
15	C + D	2 x +, 1 x -	
	B + C + D	2 x +, 1 x -	4 x +, 2 x -

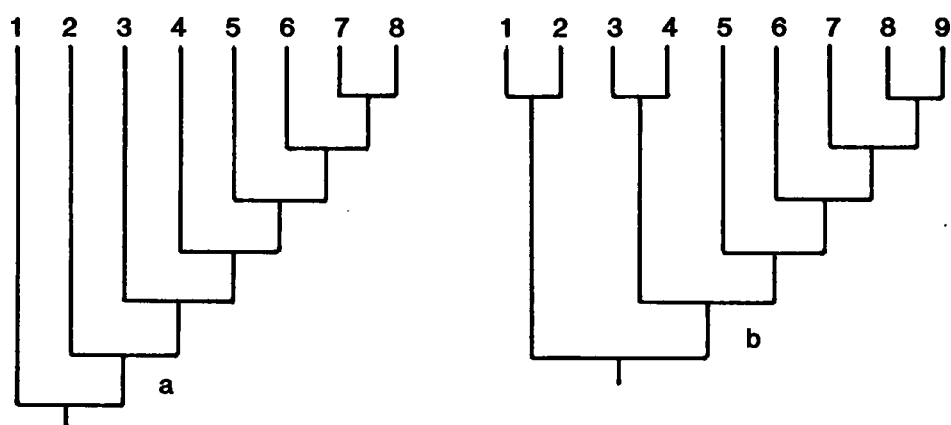


Fig. 6(a). Cladogram of *Heterandria*. (b). cladogram of *Xiphophorus*; adapted from Rosen (1976, 1978).

Table 11.

Species names and numbers used in text for *Heterandria* and *Xiphophorus*.

<i>Heterandria</i>	<i>Xiphophorus</i>
1 attenuata	1 pigmaeus
2 jonesi	2 nigrensis
3 litoperas	3 montezumae
4 obliqua	4 cortezi
5 anzuetoii	5 clemenciae
6 cataractae	6 alverezi
7 dirempta	7 "PMH"
8 bimaculata	8 signum
	9 helleri

The distribution over areas is coded as a binary matrix in Table 14 (for species names see Table 11). The phylogeny of *Heterandria* as given by Rosen (1975: Fig. 6a), shows 8 terminal taxa and 7 (internal) clada, and is also coded in binary form (Table 12). From these two binary matrices a data matrix is derived, giving distribution over areas for each claxon in the cladogram (Table 16a). The data matrix represents the boolean product of the matrices in Tables 12 and 14. It is compiled from combining distributions over areas (Table 14) for all terminal taxa indicated in each column of Table 12.

From this data matrix, 19 partial monothetic sets (components) can be extracted (Table 17). A search for the largest set of mutual compatible components reveals one maximal clique and one completely resolved area-cladogram (Fig. 7d). The wide-spread species 8 represents a true component under *Assumption 0* and, therefore, areas 2 and 3 are 'sister areas'. The relations among the other areas are similar to those postulated by previous authors (e.g. Wiley, 1981).

The phylogeny of *Xiphophorus* as given by Rosen (1976: Fig. 6b), shows 9 terminal taxa and 8 clada, and is represented in binary form in Table 13. The distribution over areas for the species of *Xiphophorus* is given in Table 15 (for species names see Table 11). The combination of these two matrices gives another matrix indicating the distribution over

areas for each cladon in the cladogram (Table 16b). From this data matrix, 18 partially monothetic sets of areas can be derived (Table 18).

Table 12.
Binary matrix for *Heterandria* cladogram (Fig. 6a).

SPECIES	clada														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1
3	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1
4	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1
5	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1
6	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1
7	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1
8	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

Table 13.
Binary matrix for *Xiphophorus* cladogram (Fig. 6b).

SPECIES	Clada																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
2	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
3	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1
4	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1
6	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1
7	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1
8	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1
9	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1

Table 14.
Distributions of *Heterandria* species (after Rosen, 1976; 1978) but excluding area 11.

SPECIES	areas									
	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	1	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	1	0
4	0	0	0	1	1	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	1	0	0
8	0	1	1	0	0	0	0	0	0	0

Table 15.
Distributions of *Xiphophorus* species (after Rosen, 1976, 1978) but excluding area 11.

SPECIES	areas									
	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0
6	0	0	0	1	1	1	0	0	0	0
7	0	0	0	0	0	0	0	0	1	1
8	0	0	0	0	0	0	0	1	0	0
9	0	1	0	0	0	0	0	0	0	0

a ((((((((2,3) 8) 7) 10) (4,5) 9) 1) 6)
b ((((((((2,8) (9,10) (4,5,6) 3) 1) 7)

Fig. 7(a). Area-Cladogram based on *Heterandria*. (b) area-cladogram based on *Xiphophorus*; both derived under Assumption 0.

As with *Heterandria*, the analysis results in one, partially resolved area-cladogram (Fig. 7b). It differs in several respects from that of *Heterandria*, in the positions of areas 3, 6, and 9. Furthermore, area 7 appears at the root because *Xiphophorus* is absent from this area. This area-cladogram is basically similar to those presented by previous authors (e.g. Wiley, 1981).

When the two data matrices are combined (Tables 16a and b) and analyzed together, 26 components can be defined (Table 19). From these components, three cliques of maximum size were found (the general area-cladograms in Fig. 8).

A joint analysis of two or more groups of taxa can start also from a combination of two reduced data matrices. In this particular case, where separate analyses of each genus yield only one area-cladogram for each genus, there is no need for a selection of distributional types. As a consequence, the respective data matrices need not be reduced but joined directly. We meet a different situation when analysing under Assumption 1 or 2.

a ((((((((2,3) 8) 7) 10) (4,5) 9) 1) 6)
b (((((2,8) (9,10)) (4,5) 6)) 3) 1) 7)
c ((((((((2,8) 3) 7) 10) (4,5) 9) 1) 6)

Fig. 8. Three general area-cladograms based on *Heterandria* and *Xiphophorus* under Assumption 0.

The general area-cladogram in Fig. 8a is identical to the area-cladogram for *Heterandria* (Fig. 6a), whereas the general area-cladogram in Fig. 8b is almost similar to the *Xiphophorus* area-cladogram (Fig. 6b) except for the subclade of areas 4-6 which is fully resolved. The general area-cladogram in Fig. 8c differs from the general area-cladogram in 8a only in terms of the subclade for areas 2, 3, and 8.

A quantitative evaluation of the three general area-cladograms shows that Fig. 8a implies 26 supportive and 15 contradictory events, to give an overall value of 11. The general area-cladograms in figs. b and c imply overall values of 9, and 12. Thus Fig.

8c has the best value of support minus contradiction and also has the minimum number of 39 state changes. The other two both have 40 state changes.

Table 16,

Transposed data matrix under Assumption 0; *Heterandria*: 1-8 species, 9-15 clada; *Xiphophorus*: 1-9 species, 9-17 clada; the numbers on the right refer to the constituting species of the respective clada. N.B. For technical reasons columns are horizontal and rows vertical.

(a) species											
/clada											
column nos.	areas										clades
	1	2	3	4	5	6	7	8	9	10	
1	0	0	0	0	0	1	0	0	0	0	
2	1	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	1	0	
4	0	0	0	1	1	0	0	0	0	0	
5	0	0	0	0	0	0	0	0	0	1	
6	0	0	0	0	0	0	1	0	0	0	
7	0	0	0	0	0	0	0	1	0	0	
8	0	1	1	0	0	0	0	0	0	0	
9	0	1	1	0	0	0	0	1	0	0	7 + 8
10	0	1	1	0	0	0	1	1	0	0	6 + 7 + 8
11	0	1	1	0	0	0	1	1	0	1	5 + 6 + 7 + 8
12	0	1	1	1	1	0	1	1	0	1	4 + 5 + 6 + 7 + 8
13	0	1	1	1	1	0	1	1	1	1	3 + 4 + 5 + 6 + 7 + 8
14	1	1	1	1	1	0	1	1	1	1	2 + 3 + 4 + 5 + 6 + 7 + 8
15	1	1	1	1	1	1	1	1	1	1	1 + 2 + 3 + 4 + 5 + 6 + 7 + 8
(b)											
1	1	0	0	0	0	0	0	0	0	0	
2	1	0	0	0	0	0	0	0	0	0	
3	1	0	0	0	0	0	0	0	0	0	
4	1	0	0	0	0	0	0	0	0	0	
5	0	0	1	0	0	0	0	0	0	0	
6	0	0	0	1	1	1	0	0	0	0	
7	0	0	0	0	0	0	0	0	1	1	
8	0	0	0	0	0	0	0	1	0	0	
9	0	1	0	0	0	0	0	0	0	0	
10	1	0	0	0	0	0	0	0	0	0	1 + 2
11	1	0	0	0	0	0	0	0	0	0	3 + 4
12	0	1	0	0	0	0	0	1	0	0	8 + 9
13	0	1	0	0	0	0	0	1	1	1	7 + 8 + 9
14	0	1	0	1	1	1	0	1	1	1	6 + 7 + 8 + 9
15	0	1	1	1	1	1	0	1	1	1	5 + 6 + 7 + 8 + 9
16	1	1	1	1	1	1	0	1	1	1	3 + 4 + 5 + 6 + 7 + 8 + 9
17	1	1	1	1	1	1	0	1	1	1	1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9

Choosing Fig. 8c (also Fig. 8a) implies that the phylogeny of *Heterandria* can be explained by assuming a sequence of vicariance events and allopatric speciation events in which only the species in area 8 has not responded. For *Xiphophorus* the implications require more *ad hoc* hypotheses. Apart from three sympatric speciation events for the species in areas 1-4, the absence of an endemic species in area 7 might be primitive absence or extinction. We may tip the balance by assuming the first hypothesis and deleting

the latter as it supposes one extra step and is therefore less parsimonious. The widespread species in area 7 might not have responded to the vicariance event separating area 9, or its present distribution might be the result of a later dispersal into area 9 implying primitive absence in this area. Species 5 has either dispersed into area 3 thus implying primitive absence in the latter or it might have become extinct in areas 2+4-6+8+10.

Table 17.
Partial monothetic sets of areas for *Heterandria*
under Assumption 0.

distributional types	components
1	2
6	1
7	6
8	7
9	3
10	5
2, 3	8
4, 5	4
2, 3, 8	9
2, 3, 7, 8	10
2, 3, 7, 8, 10	11
2, 3, 4, 5, 7, 8, 10	12
2, 3, 4, 5, 7, 8, 9, 10	13
1, 2, 3, 4, 5, 7, 8, 9, 10	14
1, 2, 3, 4, 5, 6, 7, 8, 9, 10	15

Table 18.
Partial monothetic sets of areas for *Xiphophorus* under
Assumption 0.

distributional types	components
1	1, 2, 3, 4, 10, 11
2	9
3	5
8	8
2, 8	12
9, 10	7
4, 5, 6	6
2, 8, 9, 10	13
2, 4, 5, 6, 8, 9, 10	14
2, 3, 4, 5, 6, 8, 9, 10	15
1, 2, 3, 4, 5, 6, 8, 9, 10	16, 17

Choosing the general area-cladogram in Fig. 8b implies that the *Xiphophorus* phylogeny can be explained by assuming a sequence of vicariance events in which species in areas

6 and 7 have not responded, and thus a primitive absence in area 7 can be assumed. For *Heterandria*, primitive absence in area 3 can be assumed, implying that the present distribution of the species in area 8 must be the result of dispersion into area 3. The species in area 1 might have become extinct in areas 1-5+8-10 or it has dispersed into area 6, implying primitive absence in this area. The species in area 3 might have become extinct in areas 2-5+8+10, or it has dispersed into area 9, again implying primitive absence in this area.

Table 19.
Partial monothetic sets of areas based on *Heterandria* (H) and
Xiphophorus (X) under Assumption 0.

distributional types	components
1	2 (H), 1, 2, 3, 4, 10, 11 (X)
2	9 (X)
3	5 (X)
6	1 (H)
7	6 (H)
8	7 (H), 8 (X)
9	3 (H)
10	5 (H)
2, 3	8 (H)
4, 5	4 (H)
2, 8	12 (X)
9, 10	7 (X)
4, 5, 6	6 (X)
2, 3, 8	9 (H)
2, 3, 7, 8	10 (H)
2, 8, 9, 10	13 (X)
2, 3, 7, 8, 10	11 (H)
2, 3, 4, 5, 7, 8, 10	12 (H)
2, 4, 5, 6, 8, 9, 10	14 (X)
2, 3, 4, 5, 6, 8, 9, 10	15 (X)
2, 3, 4, 5, 7, 8, 9, 10	13 (H)
1, 2, 3, 4, 5, 6, 8, 9, 10	16, 17 (X)
1, 2, 3, 4, 5, 7, 8, 9, 10	14 (H)
1, 2, 3, 4, 5, 6, 7, 8, 9, 10	15 (H)

An important difference with Wiley's (1981) analysis is that he infers unique events for all incongruent parts of the area-cladograms of the two genera without producing a general area-cladogram. However, a choice here for one of the possible general area-cladograms determines the unique events to be explained. For instance, choosing the general area-cladogram in Fig. 8c implies only one unique event affecting *Xiphophorus*, the occurrence of the species from area 5 in area 3. Choosing the general area-cladogram in Fig. 8b implies 3 unique events with regard to *Heterandria*, i.e. occurrences in areas 1, 9 and 7 for the species from areas 1, 3 and 6, respectively.

Assumption 1

Platnick (1981) considered the reduced area-cladogram obtained by Rosen equivalent to analysis under Assumption 1. However, removal of incongruent subtrees from area-cladograms is not part of this Assumption but a form of consensus tree analysis. It is

a misconception to suggest that incongruencies can lead only to incomplete or only partially resolved general area-cladograms and that they cannot be resolved into fully informative (dichotomous) hypotheses. To be able to do this, the component compatibility method is required and parsimony must be applied as shown in the next example.

Under *Assumption 1* the data matrix for *Heterandria* comprises the same columns as given in the data matrix under *Assumption 0* (Table 16a) plus those given in Table 20a. The columns in Table 20a are derived from those in Table 16a in the following manner. From those columns with wide-spread terminal taxa (table 16a, column numbers 4 and 8) we extract the distributions corresponding to all subsets of areas i.e. the singletons (Table 20a, column nos 3, 4, 7, and 8) and combine them with the distributions of prospective sistergroups (Table 16a, species/cladon column numbers 7 and 11) of the wide-spread terminal taxa (Table 20a column numbers 1, 2, 5, and 6). An analysis of *Heterandria* yields the components given in Table 21. In comparison to Table 17 (the components derived under *Assumption 0*), four new distributional components are present (i.e. areas 2+8, 3+8, 2+3+4+7+8+10, 2+3+5+7+8+10).

Table 20.

Additional columns (transposed horizontally) for Table 16 for analysis under Assumption 1, *Heterandria*, *Xiphophorus*.

											species
	1	2	3	4	areas		7	8	9	10	
					5	6					
(a)	<i>Heterandria</i>										
1	0	1	1	0	1	0	1	1	0	1	based on species 4
2	0	1	1	1	0	0	1	1	0	1	
3	0	0	0	1	0	0	0	0	0	0	
4	0	0	0	0	1	0	0	0	0	0	
5	0	0	1	0	0	0	0	1	0	0	based on species 8
6	0	1	0	0	0	0	0	1	0	0	
7	0	1	0	0	0	0	0	0	0	0	
8	0	0	1	0	0	0	0	0	0	0	
(b)	<i>Xiphophorus</i>										
1	0	0	0	0	0	1	0	0	0	0	based on species 6
2	0	0	0	0	1	0	0	0	0	0	
3	0	0	0	0	1	1	0	0	0	0	
4	0	0	0	1	0	0	0	0	0	0	
5	0	0	0	1	0	1	0	0	0	0	
6	0	0	0	1	1	0	0	0	0	0	
7	0	1	0	0	0	1	0	1	1	1	
8	0	1	0	0	1	0	0	1	1	1	
9	0	1	0	0	1	1	0	1	1	1	
10	0	1	0	1	0	0	0	1	1	1	
11	0	1	0	1	0	1	0	1	1	1	
12	0	1	0	1	1	0	0	1	1	1	
13	0	0	0	0	0	0	0	0	0	1	based on species 7
14	0	0	0	0	0	0	0	0	1	0	
15	0	1	0	0	0	0	0	1	0	1	
16	0	1	0	0	0	0	0	1	1	0	

From this component array, 9 maximal cliques can be extracted, each showing equal values of support minus contradiction for the complete data matrix (Fig. 9). Figure 9d

is the area-cladogram derived under *Assumption 0*, by ignoring area-cladograms 2, 3, 5, 6, 8, and 9, because they separate areas 4 from 5 (which previous authors considered to be one area of endemism), we end up with three possible general area-cladograms (Figs. 9a, 9d, and 9g). These differ from one other by having three possible alternatives for areas 2, 3, and 8. When evaluated against the data matrix for *Assumption 0* (see p. 307 for a justification) one area-cladogram shows a best fit (Fig. 9a). It lacks homoplasies entirely while Figs. 9d and 9g each show one homoplasy (reversal = extinction for the species in area 8) and the same support as Fig. 9a.

Table 21.
Partial monothetic sets of areas based on *Heterandria* under Assump-
tion 1; (16) refers to Table 16a (20) to Table 20a.

distributional types	components
1	2 (16)
6	1 (16)
7	6 (16)
8	7 (16)
9	3 (16)
10	5 (16)
2, 3	8 (16)
4, 5	4 (16)
2, 8	4 (20)
3, 8	3 (20)
2, 3, 8	9 (16)
2, 3, 7, 8	10 (16)
2, 3, 7, 8, 10	11 (16)
2, 3, 4, 7, 8, 10	2 (20)
2, 3, 5, 7, 8, 10	1 (20)
2, 3, 4, 5, 7, 8, 10	12 (16)
2, 3, 4, 5, 7, 8, 9, 10	13 (16)
1, 2, 3, 4, 5, 7, 8, 9, 10	14 (16)
1, 2, 3, 4, 5, 6, 7, 8, 9, 10	15 (16)

- a ((((((((2,3) 8) 7) 10) (4,5) 9) 1) 6)
- b (((((((((2,3) 8) 7) 10) 4) 5) 9) 1) 6)
- c (((((((((2,3) 8) 7) 10) 5) 4) 9) 1) 8)
- d (((((((((2,8) 3) 7) 10) (4,5)) 9) 1) 6)
- e (((((((((2,8) 3) 7) 10) 4) 5) 9) 1) 6)
- f (((((((((2,8) 3) 7) 10) 5) 4) 9) 1) 6)
- g (((((((((3,8) 2) 7) 10) (4,5)) 9) 1) 6)
- h (((((((((3,8) 2) 7) 10) 4) 5) 9) 1) 6)
- i (((((((((3,8) 2) 7) 10) 5) 4) 9) 1) 6)

Fig. 9. The nine area-cladograms of *Heterandria* derived under *Assumption 1*.

In *Xiphophorus* the data matrix comprises the same columns given under *Assumption 0* (Table 16b), together with those in Table 20b. The newly incorporated columns all define new partial monothetic sets (components) derived under *Assumption 1* (Table 22,

areas 2+8+9, 2+8+10, 2+4+8+9+10, 2+5+8+9+10, 2+6+8+9+10, 2+4+5+8+9+10, 2+4+6+8+9+10, 2+5+6+8+9+10

Table 22.

Partial monothetic sets of areas based on *Xiphophorus* under Assumption 1; (16) refers to Table 16b and (20) to Table 20b.

distributional types	components
1	1, 2, 3, 4, 10, 11 (16)
2	9 (16)
3	5 (16)
4	4 (20)
5	2 (20)
6	1 (20)
8	8 (16)
9	14 (20)
10	13 (20)
4, 5	6 (20)
4, 6	5 (20)
5, 6	3 (20)
2, 8	12 (16)
9, 10	7 (16)
4, 5, 6	6 (16)
2, 8, 9	16 (20)
2, 8, 10	15 (20)
2, 8, 9, 10	13 (16)
2, 4, 8, 9, 10	10 (20)
2, 5, 8, 9, 10	8 (20)
2, 6, 8, 9, 10	7 (20)
2, 4, 5, 8, 9, 10	12 (20)
2, 4, 6, 8, 9, 10	11 (20)
2, 5, 6, 8, 9, 10	9 (20)
2, 4, 5, 6, 8, 9, 10	14 (16)
2, 3, 4, 5, 6, 8, 9, 10	15 (16)
1, 2, 3, 4, 5, 6, 8, 9, 10	16, 17 (16)

Within these sets, 45 maximal cliques can be determined, 36 of which show equally good values of support minus contradiction when using the complete data matrix, and 3 when using the *Assumption 0* data matrix (Fig. 10). By ignoring those cliques in which area 4 and 5 are separated (as for *Heterandria*), we end up with 9 area-cladograms when evaluated on the complete data matrix and one on the *Assumption 0* data matrix (see Fig. 8b).

a (((((2,8) (9,10)) ((4,5) 6)) 3) 1) 7)
 b (((((2,8) (9,10)) ((4,6) 5)) 3) 1) 7)
 c (((((2,8) (9,10)) ((5,6) 4)) 3) 1) 7)

Fig. 10. The three area-cladograms of *Xiphophorus* derived under *Assumption 1*, evaluated on the data matrix used for *Assumption 0*.

The analysis of *Heterandria* and *Xiphophorus* together comprises a data matrix based on Tables 16a, b and 20a, b. The complete range of components are given in Table 23.

Table 23.
Partial monothetic sets of areas based on *Heterandria* (H) and *Xiphophorus* (X) under Assumption 1; (16) refers to Table 16 and (20) refers to Table 20.

distributional types	components
1	2 (H16), 1, 2, 3, 4, 10, 11, (X16)
2	9 (X16)
3	5 (X16)
4	4 (X20)
5	2 (X20)
6	1 (H16), 1 (X20)
7	6 (H16)
8	7 (H16), 8 (X16)
9	3 (H16), 14 (X20)
10	5 (H16), 13 (X20)
2, 3	8 (H16)
4, 5	4 (H16), 6 (X20)
4, 6	5 (X20)
5, 6	3 (X20)
2, 8	4 (H20), 12 (X16)
3, 8	3 (H20)
9, 10	7 (X16)
4, 5, 6	6 (X16)
2, 3, 8	9 (H16)
2, 8, 9	16 (X20)
2, 8, 10	15 (X20)
2, 3, 7, 8	10 (H16)
2, 8, 9, 10	13 (X16)
2, 3, 7, 8, 10	11 (H16)
2, 4, 8, 9, 10	10 (X20)
2, 5, 8, 9, 10	8 (X20)
2, 6, 8, 9, 10	7 (X20)
2, 3, 4, 7, 8, 10	2 (H20)
2, 3, 5, 7, 8, 10	1 (H20)
2, 4, 5, 8, 9, 10	12 (X20)
2, 4, 6, 8, 9, 10	11 (X20)
2, 5, 6, 8, 9, 10	9 (X20)
2, 3, 4, 5, 7, 8, 10	12 (H16)
2, 4, 5, 6, 8, 9, 10	14 (X16)
2, 3, 4, 5, 6, 8, 9, 10	15 (X16)
2, 3, 4, 5, 7, 8, 9, 10	13 (H16)
1, 2, 3, 4, 5, 6, 8, 9, 10	16, 17, (X16)
1, 2, 3, 4, 5, 7, 8, 9, 10	14 (H16)
1, 2, 3, 4, 5, 6, 7, 8, 9, 10	15 (H16)

- a (((((2,8) (9,10)) ((4,5) 6)) 3) 1) 7)
b ((((((2,8) 9) 10) ((5,6) 4)) 3) 1) 7)
c ((((((((2,8) 10) 9) 6) 4) 5) 3) 1) 7)
d ((((((((2,8) 10) 9) 6) 5) 4) 3) 1) 7)

Fig. 11. The four best general area-cladograms of *Heterandria* and *Xiphophorus* derived under Assumption 1 from a complete data matrix.

From this list 54 maximal cliques are present. When evaluated on the basis of the complete data matrix, i.e. including the assumptuonal columns, four of them have the best value of support minus contradiction (=10: Fig. 11a-d), but only two show a distributional pattern of areas 4+5 (Fig. 11a, b). The first is identical with Fig. 8b i.e. one of general area-cladograms derived from *Assumption 0* but the other differs from the first in that it shows areas 9 and 10 branching sequentially.

When evaluated with the *Assumption 0* data matrix, there are two general area-cladograms, one better than the other. The first is identical with Figs. 8c and 9d, the second with Figs. 8d and 9d. They have the same support but differ with respect to 1 homoplasy. The best solution comes from an analysis using *Assumption 0* (Fig. 8c) which is among the best ones from an analysis under *Assumption 1*, even when the evaluation includes or excludes assumptuonal columns in the data matrix.

When we use the distributional types corresponding with Figs. 8a, b to build a reduced data matrix we can subsequently derive 26 components. Exploring their mutual compatibility reveals 3 general area-cladograms. The best two are identical with Figs. 9a, d. Fig. 9a is slightly better because it has one less homoplasy. Humphries (1982), Platnick (1981), and Humphries and Parenti (1986) give only one, incompletely resolved general area-cladogram under this assumption. In contrast, our analysis yields several completely dichotomous general area-cladograms.

Assumption 2

In *Assumption 2* all data concerning wide-spread taxa and missing areas are doubted from the beginning. Nelson and Platnick (1981) and others have a preference for *Assumption 2* which seems to be due to the fact that their method uses consensus trees. Therefore, they must doubt specific aspects of the data in advance, i.e. they regard some aspects of the data to be ad hoc. In our procedure the reverse is true; conflicts can be resolved or they might lead to ad hoc hypotheses. Apart from this criticism, *Assumption 2* as implemented here is hard to apply in actual practice because even in simple cases, it leads to an unmanageable number of possibilities, as illustrated by the following results.

Under *Assumption 2*, the data matrices for *Heterandria* and *Xiphophorus* (Table 16) are extended with columns added for wide-spread taxa and missing areas. These data matrices are far too large to be shown here in their entirety but we will briefly describe how they might be compiled. For a wide-spread taxon we first take the corresponding column from the data matrix derived under *Assumption 0* (e.g. the species column 6 in Table 16b for the wide-spread *Xiphophorus* species in area 6). Then we extract all its possible subsets of areas (Table 20b, columns 1-6) and combine these with the distributions of all other clada (Table 16b) to obtain the columns for *Assumption 2*. This procedure is repeated for all wide-spread taxa in both genera. For missing areas (e.g. area 7 in Table 15 for *Xiphophorus*) we extract all possible subsets (here only the area itself) and combine these with the distributions of all other clada (Table 16b).

The size of the data matrices increases significantly under this assumption compared to those derived under the former two (for *Heterandria* 15 columns under *Assumption 0*, 23 columns under *Assumption 1*, and 79 columns under *Assumption 2*; for *Xiphophorus* 17 columns under *Assumption 0*, 33 columns under *Assumption 1*, and 178 columns under *Assumption 2*; for their combination, 32 columns under *Assumption 0*, 56 columns under *Assumption 1*, and 257 columns under *Assumption 2*). The number of components increases accordingly (For *Heterandria* 19 vs 23 vs 53; for *Xiphophorus* 18 vs 29 vs 85; for their combination 26 vs 39 vs 109). As a consequence, the number of maximal cliques rises almost exponentially (for *Heterandria* 1 vs 9 vs 123; for *Xiphophorus* 1 vs 45 vs 2613; for their combination 3 vs 67 vs 8431!)

It is clear that complex problems analysed under *Assumption 2*, i.e. with several species

having a range of 2 or more areas, leads to an unmanageable number of possibilities.

Independent analysis of *Heterandria* results in 123 completely resolved area-cladograms. When these are evaluated on the basis of the complete data matrix two better ones emerge regarding the balance between support (52) and contradiction (65). These are identical with those in Figs 9b, c but as noted earlier under *Assumption 1*, they cannot be accepted because areas 4 and 5 branch off sequentially rather than as sister areas. There are two second best area-cladograms with 54 supporting states and 68 contradictory events, but these also do not support areas 4 and 5 as sister areas. There are 11 third best area-cladograms. Their support values range from 49–54 and their homoplasies from 64–72. Among them is the one area-cladogram that has the least total number of state changes (108) of all 123 area-cladograms. It is also the only one out of the 11 that supports 4 and 5 as sister areas. It has the same topology as that in Fig. 9.1. When evaluated against the *Assumption 0* data matrix the best area-cladogram has 15 supporting states and no homoplasies which corresponds with Fig. 9a.

A separate analysis of *Xiphophorus* results in 2613 completely resolved area-cladograms. Evaluation against the complete data matrix (178 columns) gives 6 equally good area-cladograms (231 homoplasies, 90 support, balance 141) and 15 second best (223–231 homoplasies, 81–90 support, balance 142). Among the latter are the three shortest area-cladograms (302 steps). All three have an identical topology to the cladogram in Fig. 10. Evaluation against the data matrix used for *Assumption 0* revealed 3 of the best cladograms, corresponding with those in Fig. 10. By ignoring those without areas 4 and 5 as sister areas, we end up with the cladogram in Fig. 10, which is identical to that in Fig. 8b.

It is obvious that an analysis of *Heterandria* and *Xiphophorus* taken together and based on reduced data matrices will give the same result as obtained under *Assumption 1* because the selected distributional patterns come from the same area-cladograms (Figs 8b and 9a). An analysis of *Heterandria* and *Xiphophorus* with a complete combined data matrix (257 columns) generates 8431 completely dichotomous general area-cladograms. Evaluation against the complete data matrix gives two 'best' general area-cladograms (homoplasy–support = 241, 478 steps). Both show sequential branching for areas 4 and 5. For the next four values (242, 245, 246, 250) there are also four general area-cladograms, all showing areas 4 and 5 as sister areas. One of them (value 250) is the shortest general area-cladogram (451 steps; Fig. 13).

As to the number of area-cladograms our results are most surprising, as they differ remarkably from those obtained by Platnick (1981), Humphries (1982), and Humphries and Parenti (1986). They mention only 3 possible general area-cladograms, but it remains unclear how they obtained their results. Compared with the provisions of our data matrix, their small number of general area-cladograms might be due to the fact that they apparently let only areas 3 and 9 take different positions under *Assumption 2* because they are part of the range of the wide-spread *Heterandria* species in area 8 and the *Xiphophorus* species in area 7. However, when this assumption is strictly applied, the different positions for areas 10 and 2, 9 and 2, or 10 and 3 should also be taken into account. It seems strange to us that these authors do not mention the general area-cladogram in our figure 8a as a valid alternative hypothesis, because under *Assumption 2* the component 2+3, based on the wide-spread species in area 8, might be true.

Discussion and Conclusions

Humphries and Parenti (1986) distinguish three different methods for historical biogeographic analysis, i.e. those of Rosen (1976), Wiley (1980, 1981), and Nelson and Platnick (1978).

Rosen's (1976, 1978) method is the construction of reduced consensus area-cladograms, in which the incongruent parts of the compared area-cladograms are deleted. These

reduced area-cladograms imply that observed congruencies are due to factors that affect all study groups (Wiley, 1981, p. 293). However, reduced consensus area-cladograms lose information. The first aim of historical biogeography should be the determination of the relationships between delineated areas. By omitting areas, their relationships with the other areas cannot be established. Criticism analogous to points raised against character compatibility methods (especially by Farris, 1983 p. 31) obtains here.

Wiley's (1980, 1981) 'ancestral species maps' method (see Humphries and Parenti 1986) aims at reconstructing a sequence of speciation events. For this purpose, reduced area-cladograms are used implying that incongruencies are caused by unique factors. In the sole worked example, only minor incongruencies are met. It is unclear as to what might happen in an example including more groups in which there exists a great degree of incongruence. At the extreme, one wonders what the conclusions might be if an analysis for several groups revealed total incongruence. Would a general area-cladogram derived from a compound data matrix by component compatibility and selected on the basis of parsimony not constitute a more general explanation than only the unique factors? Furthermore, speciation events should be interpreted *a posteriori* based on a well resolved general area-cladogram.

Recently, Wiley (in press) presented a general outline of a formal parsimony method in which he uses Brooks' (1981) method for converting taxon trees to binary codes for areas. In this method he explores different coding strategies, different from ours, for groups with missing taxa and/or with wide-spread species, using the PAUP Program (Swofford 1985) as an analytical tool. His method and its coding strategies are aimed at obtaining improved resolution than current methods allow in answering the two main questions posed in historical biogeography; (1) what are the relationships between areas, and (2) what are the co-evolutionary patterns between taxa with common or partly common distributions. As far as can be judged from the results obtained so far Wiley's new method operates in much the same spirit as ours and as a consequence removes the objections that can be raised against methods aimed solely at estimating a minimum number of unique events.

Nelson and Platnick's (1981) component analysis aims at constructing general consensus area-cladograms when several conflicting biogeographic patterns are encountered in different groups of taxa. It has been amply demonstrated that consensus general area-cladograms are coupled with a loss of information besides being notoriously unpar-simonious explanations of the considered events. We have demonstrated that contradictions arising from incompatible area-components can be resolved perfectly in a general area-cladogram by applying the parsimony criterion.

We show also a fully implementation of Nelson and Platnick's *Assumption 2* can hardly be strictly applied in practice, as it leads to many general area-cladograms. Therefore, because the method described here resolves conflicts present in a data matrix, it is sufficient to undertake historical biogeographic analysis under *Assumption 0*.

In conclusion, our biogeographic method renders full account of all information present in the initial data and gives the most resolved parsimonious general area-cladograms possible.

Acknowledgments

Many thanks are due to Dr. E.O. Wiley, who was a stimulating participant in many discussions during his tenure as a visiting scientist in the Netherlands, and who greatly contributed to our understanding of historical biogeography. We also wish to thank Dr. C.J. Humphries and two anonymous reviewers for their critical remarks and suggestions to improve the manuscript. We also acknowledge the comments made by Dr. C.H. Haufler.

LITERATURE CITED

- BECKNER, M. 1959. The biological way of thought. New York.
- BROOKS, D. R. 1981. Hennig's parasitological method: a proposed solution. *Syst. Zool.* 30: 229-249.
- FARRIS, J. S. 1970. Methods for computing Wagner trees. *Syst. Zool.* 19: 83-92.
- FARRIS, J. S. 1978. An efficient method for finding monothetic groups. *Syst. Zool.* 27: 468-472.
- FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. In Platnick, N. I. and V. A. Funk (eds.) *Advances in cladistics II: Proceedings of the second meeting of the Willi Hennig Society*. Columbia Univ. Press, New York, pp. 7-36.
- FARRIS, J. S., A. G. KLUGE, AND M. J. ECKARDT. 1970. A numerical approach to phylogenetic systematics. *Syst. Zool.* 19: 172-189.
- GAREY, M. R., AND D. S. JOHNSON. 1979. *Computers and intractability*. W. H. Freeman & Co., San Francisco.
- GEESINK, R. 1984. *Scala Millettiarum*. Leiden Bot. Ser. 8., Leiden.
- HUMPHRIES, C. J. 1982. Vicariance biogeography in Mesoamerica. *Ann. Missouri Bot. Gard.* 69: 444-463.
- HUMPHRIES, C. J. AND L. R. PARENTI. 1986. *Cladistic biogeography*. Clarendon Press, Oxford.
- MICKEVICH, M. F. 1981. Quantitative phylogenetic biogeography. In Funk, V. A. and D. R. Brooks (eds.), *Advances in Cladistics I: Proceedings of the first meeting of the Willi Hennig Society*. Columbia Univ. Press, New York, pp. 209-222.
- NELSON, G. AND N. I. PLATNICK. 1978. A method of analysis for historical biogeography. *Syst. Zool.* 27(1): 1-16.
- NELSON, G. AND N. I. PLATNICK. 1981. *Systematics and biogeography*. Columbia Univ. Press, New York.
- PLATNICK, N. I. 1981. Widespread taxa and biogeographic congruence. In: Funk, V. A. and D. R. Brooks (eds.), *Advances in cladistics I: Proceedings of the first meeting of the Willi Hennig Society*. Columbia University Press, New York, pp. 223-227.
- ROOS, M. C. 1985. *Phylogenetic systematics of the Drynarioideae (Polypodiaceae)*. PhD. Thesis, Utrecht.
- ROOS, M. C. 1986. *Phylogenetic systematics of the Drynarioideae (Polypodiaceae)*. *Verh. K.N.A.W. Natuurk. Ser. 2, deel 85*. Amsterdam.
- ROOS, M. C. AND M. ZANDEE. (in prep.). *Historical biogeography of the Drynarioideae revisited*.
- ROSEN, D. E. 1976. A vicariance model of Caribbean biogeography. *Syst. Zool.* 24: 431-464.
- ROSEN, D. E. 1978. Vicariant patterns and historical explanation in biogeography. *Syst. Zool.* 27: 159-188.
- SHARROCK, G. AND J. FELSENSTEIN. 1975. Finding all monothetic subsets of a taxonomic group. *Syst. Zool.* 24: 373-377.
- SWOFFORD, D. L. 1985. *PAUP. Phylogenetic analysis using parsimony. Users manual Version 2.4*. Illinois Natural History Survey.
- WILEY, E. O. 1980. *Phylogenetic systematics and vicariance biogeography*. *Syst. Bot.* 5: 194-220.
- WILEY, E. O. 1980. *Phylogenetics. The theory and practice of phylogenetic systematics*. Wiley, New York.
- WILEY, E. O. 1987. *Methods of vicariance biogeography*. (in press).
- WILEY, E. O. AND R. L. MAYDEN. 1985. Species and speciation in phylogenetic systematics, with examples from the North American fish fauna. *Ann. Missouri Bot. Gard.* 72: 596-635.
- ZANDEE, M. 1984. Three-cladon-statement permutations + local outgroup comparison = cladistic analysis. Paper read at the 4th meeting of the Willi Hennig Society, London.
- ZANDEE, M. 1985. C.A.F.C.A.: A collection of APL functions for cladistic analysis. On-line C.A.F.C.A. documentation (in Dutch).
- ZANDEE, M. AND R. GEESINK. 1987. Legumes and phylogenetics: A desire for the impossible? *Advances in Legume Systematics* 3 (in press).