

PREPRINT

© 1999 Academic Press

© 1999 Willi Hennig Society

Two requirements for obtaining valid common patterns under different assumptions in vicariance biogeography

Marco G.P. van Veller*, M. Zandee, D.J. Kornet

Section Theoretical Biology and Phylogenetics, Institute of Evolutionary and Ecological Sciences, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands

*Correspondence to: M.G.P. van Veller, Section Theoretical Biology and Phylogenetics, Institute of Evolutionary and Ecological Sciences, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands. E-mail: vanveller@RULSFB.LeidenUniv.NL. Fax: +31 71 527 4900.

Running title: Assumptions in biogeography: two requirements

Keywords: Vicariance biogeography, Assumptions, Inclusion, Solution sets, Common pattern, Single assumption comparison, Explanatory power

ABSTRACT

In vicariance biogeography, widespread or sympatric taxa can be dealt with under assumptions zero, 1 and 2. Data from cladogenetic relationships among taxa of a monophyletic group and their distribution over areas are assumed, in the order zero \rightarrow 1 \rightarrow 2, to represent decreasing information about vicariance events. A less strict assumption carries a larger solution set, *i.e.* the number of possible area cladograms increases with the decrease in strictness of the assumption applied. We formulate two requirements for obtaining valid general area cladograms from data of several monophyletic groups of taxa. First, the assumptions, and with them the sets area cladograms derived under these assumptions, should be inclusive. Second, sets of single group area cladograms should be compared for different monophyletic groups under a single assumption. When these two requirements are met, area cladograms become consistent with respect to the processes (vicariance, extinction, and dispersal) that are *a priori* assumed. The explanatory power increases for any particular monophyletic group of taxa when the set of valid general area cladograms contains a subset of area cladograms derived under a less strict assumption. We discuss examples from literature of how violation of these two requirements affects the results.

INTRODUCTION

Vicariance biogeography seeks to explain the distribution of taxa of a monophyletic group over areas by the reconstruction of the historical relationships between these areas.

Vicariance is the first-order explanation for correspondence between cladogenetic relationships among taxa and historical relationships among areas in which the taxa occur. A vicariance event (*e.g.* the origin of a barrier or the break up of an area) is assumed to trigger speciation events in species of all monophyletic groups present in the area concerned. Consequently, the histories of the areas and monophyletic groups concerned become associated. So there is *a priori* confidence in cladogenetic and distribution data of taxa as being informative regarding the historical relationships between the areas over which the taxa of each monophyletic group are distributed. These relationships form the basis to infer common patterns that are represented in a general area cladogram. We cannot observe common patterns directly but must infer them from congruence among the single patterns obtained for each monophyletic group.

A single pattern is a hypothesis of area relationships, represented by an area cladogram. An area cladogram is inferred from two types of data: the distribution of taxa over areas and the cladogenetic relationships of the taxa concerned.

In the straightforward case, each taxon is endemic to a unique area and each area harbors a single taxon of a particular group. In such cases, one obtains area cladograms by replacing taxa in the taxon cladogram by the areas in which they occur (Rosen, 1978; Nelson and Platnick, 1981; Page, 1988; Morrone and Carpenter, 1994). Since an area cladogram represents a hypothesis of a unique historical pattern and areas are delimited as separate entities, the areas must have their own unique place. Morrone and Carpenter (1994) and Enghoff (1996) call such area cladograms “resolved area cladograms”.

However, data on distribution and cladogenetic relationships of taxa are not always that straightforward with respect to the reconstruction of area relationships. A lack of response to a vicariance event, extinction and dispersal events as well as sympatric speciation, lead to

widespread or sympatric taxa. These phenomena generate no or misleading information on historical relationships between areas as represented in taxon-area cladograms that are obtained after replacing the taxa in the taxon cladogram by their areas of distribution. Sympatric speciation (*i.e.* singular or subsequent speciation event(s) within one area) results in two or more species occurring in the same area. However, such species do not pose a problem in the derivation of an area cladogram. When they are replaced by their areas of distribution, the two or more terminal nodes representing the same area can be safely collapsed without changing the topology of the area cladogram (Fig. 1).

For the non-straightforward cases the consequences of extinction, lack of response to vicariance, and dispersal are dealt with under different assumptions, dubbed zero (A0), 1 (A1) and 2 (A2).

Under A0 the distribution data for the taxa concerned are *a priori* assumed to represent the effects of vicariance alone. The occurrence of widespread and sympatric taxa is thought *a priori* to be the result of a failure to respond to a vicariance event. Homoplasies, *i.e.*, multiple appearances of taxa or their absence in the area cladogram, are explained *a posteriori* by dispersal and/or extinction.

Under A1 the distribution data for the taxa concerned are *a priori* assumed to represent the effects of vicariance and extinction. The occurrence of widespread and sympatric taxa is thought *a priori* to be the result of either the lack of response to a vicariance event or extinction. The presence of dispersals (homoplasies in the area cladogram) can be inferred only *a posteriori*.

Under A2 the distribution data for the taxa concerned are *a priori* assumed to represent the effects of vicariance, extinction and dispersal. The occurrence of widespread and sympatric taxa is thought *a priori* to be the result of either a failure to respond to a vicariance event, or extinction or dispersal. As a consequence, homoplasies in the area cladogram have no *a posteriori* explanation.

In this paper we show that in order to explain all data, one should solve redundancy (*i.e.* sympatric taxa) only "as is", and we argue that absence or multiple presence of taxa in the

area cladogram caused by extinction or dispersal should be accounted for only *a posteriori*, contrary to, *e.g.* Page (1988, 1990) and Nelson and Platnick (1981).

Common patterns in the history of areas are derived by comparing resolved area cladograms obtained for different groups of taxa. These common patterns are represented in a general area cladogram. The received view is that confidence in this cladogram increases when a larger number of different monophyletic groups of taxa shows the same pattern(s) in historical relationships of the areas (Wiley, 1988a,b). This is analogous to the increase of confidence in a taxon cladogram when one observes congruence among a large number of independent characters.

We argue that, for common patterns to be revealed, the area cladograms for different monophyletic groups of taxa should all be obtained under the same assumption. This is contrary to the suggestions made by Morrone and Crisci (1995), Enghoff (1996) and Page (1990). We do not allow a common pattern to be *a priori* explained for one group of taxa by one particular set of processes (*e.g.* vicariance plus extinction) and for another group of taxa by a different set of processes (*e.g.* vicariance plus extinction as well as dispersal). In this way, we derive common patterns with the same amount of *a priori* confidence in the data of different groups of taxa as representing information on vicariance events.

Finally, we argue that the sets of area cladograms obtained under A0, A1 and A2 for one group of taxa should be inclusive, contrary to, *e.g.* Nelson and Platnick (1981) and Nelson and Ladiges (1991b). That is, area cladograms derived under A0 should be contained in the set of area cladograms derived under A1 and these in turn should be contained in the set of area cladograms derived under A2. The processes that one considers *a priori* to result in historical relationships of areas as recovered under A0, A1, or A2 are taken to be largely independent of each other. This means that their effects are additive and as a consequence the resulting patterns are inclusive. Moreover, common patterns revealed in this way have more explanatory power than those based on non-inclusive sets of area cladograms.

DEALING WITH WIDESPREAD OR SYMPATRIC TAXA

Widespread and/or sympatric taxa are the result of processes deviating from vicariance. When an area cladogram is constructed by replacing the taxa by their distributions, the areas over which a widespread taxon is distributed occupy a single terminal node (Fig. 2a). This leads to the violation of the requirement that each area should occupy its own unique terminal node in the resolved area cladogram.

Sympatric taxa of a monophyletic group occur jointly in a single area. This may result in an area appearing more than once in the taxon-area cladogram (Fig. 2a). The distribution of the taxa concerned, therefore, is called redundant. Consequently, the requirement for occupying one terminal node is violated.

Authors have dealt with widespread taxa and redundant distributions in various ways (Table 1). Rosen (1978) and Kluge (1988) *a priori* excluded data from the analysis and Van Soest (1996) dealt with these phenomena by *a priori* assuming multiple histories for the areas with widespread taxa. However, in order to deal with widespread taxa and redundant distributions without *a priori* excluding data or assuming multiple histories of areas, other authors (Table 1) have distinguished three different assumptions.

In the next three sections, we describe how authors (Table 1) originally defined the assumptions and we describe which different processes are *a priori* assumed under A0, A1 and A2 and how these processes have resulted in today's distribution of taxa within a monophyletic group.

ASSUMPTION ZERO

Zandee and Roos (1987) and Wiley (1988a) introduced A0. Under this assumption, these authors interpret widespread taxa as “synapomorphies” of the areas in which they occur, uniting these areas into one component. They do not *a priori* assume extinction or dispersal

to have taken place. The widespread distribution of the taxon is considered the result of isolation or break up of areas without triggering speciation yet. Analysis results in a single area cladogram for a four area case with one widespread taxon (T3) in two areas (Fig. 2b). According to these authors, under A0, redundancy should be interpreted “as is”.

There are two candidate analogies with cladistic character analysis. First, we can compare redundancy with the occurrence of more than one autapomorphy in a taxon cladogram (Zandee and Roos, 1987; Brooks, 1990). Second, in our opinion, one can also compare redundancy with polymorphism. The occurrence of two or more taxa of the same monophyletic group in an area cladogram is then compared with the occurrence of two or more states of the same character in a taxon cladogram. In either case, the occurrence of two (or more) taxa in the same single area (Fig. 3a) is accounted for as just another step in the area cladogram (Fig. 3b). In the case where the taxa are actually sister taxa, the analogy of redundancy with polymorphism may be more appropriate. The steps on the branches leading to the redundant areas can be explained *a posteriori* as a result of either extinction and/or dispersal of taxa (Zandee and Roos, 1987; Wiley, 1988a; Brooks, 1990).

We think under A0 most *a priori* confidence is placed on the data as representing vicariance events since only processes that represent historical relationships of areas (vicariance) are initially admitted. Dispersal and extinction of taxa are in general thought to be random processes that do not produce information on the history of areas.

ASSUMPTION 1

A1 is originally described by Nelson and Platnick (1981, p. 421). With respect to widespread taxa, they stated under A1: “... whatever is true of the one occurrence is also true of the other occurrence”. However, in our view this statement is ambiguous because it is unclear whether they refer to relationships among areas or among taxa.

For widespread taxa, Humphries (1982, pp. 453,454) interpreted A1 to pertain to taxon

relationships: “the implications are that under assumption 1 the taxon occupying area CD [*i.e.* a taxon widespread in both areas C and D] will never be split into separate taxa”. Humphries and Parenti (1986) and Zandee and Roos (1987) use the same taxon relationship approach to deal with distributions of widespread taxa. On the other hand, Page (1989, p. 167) advocates that a crucial distinction must be drawn between the relationships of taxa and the relationships of areas. According to Page A0, A1 and A2 are interpretations of the relationship between areas, not between taxa. Platnick (1988) also maintains this area relationship approach.

The assumptions determine the amount of *a priori* confidence in cladogenetic and distribution relationships of a group of taxa as representing vicariance events. Therefore, we agree with various authors (Humphries, 1982; Humphries and Parenti, 1986; Zandee and Roos, 1987) that A0, A1 and A2 should be used in a taxon relationship approach.

Under A1, a widespread taxon present in two areas (*e.g.* taxon T3 in areas C and D in Fig. 4a) is not split into separate taxa. The joint area C plus D is not considered as such because terminal taxa do not contribute components (Nelson and Platnick, 1981, p. 422). Fig. 4b shows a single taxon-area cladogram constructed under A1. The only components that can be derived are (ABCD) and (BCD). Resolving the polytomy for BCD (in Fig. 4b), by inserting an extra internode gives rise to three different area cladograms (Fig. 4c). The implied relationships for the widespread taxon with the other taxa are not disturbed in any of the three area cladograms. By removing one internode, the taxon-area cladogram may be recovered. The presence of a widespread taxon is assumed *a priori* to be the result of either isolation or break up of areas without speciation (scenario as under A0; “synapomorphy” in Fig. 4c) or extinction of the widespread taxon (T3) in the areas that branch off after the areas in which the widespread taxon is distributed (“extinctions” in Fig. 4c).

With respect to redundancy, under A1 Nelson and Platnick (1981, p. 456) consider all sympatric taxa to be informative for historical relationships between areas. In contrast with assumption zero, sympatry is not solved *a posteriori*, but *a priori* processes deviating from vicariance are assumed to explain the presence of more than one taxon of a monophyletic

group in the same area (*i.e.* redundant distribution). Nelson and Platnick (1981) explain the areas with redundant distributions (*e.g.* area A in Fig. 5a) by *a priori* assuming extinction of (hypothetical) taxa (in addition to vicariance) in areas without redundant distributions (“extinctions” in areas B and C in Fig. 5b). Thereby, they assume an area with redundant distribution to be the remainder of a larger pattern (Fig. 5b).

With the derivation of larger patterns according to the protocol of Nelson and Platnick (1981), taxa present in the same area retain their cladogenetic relationships with the other taxa (Fig. 5b). However, as soon as resolved area cladograms are extracted from the larger pattern, taxa present in areas with redundant distributions are removed from the larger pattern. With the removal of one of the sympatric taxa (*e.g.* taxon T1 in area A in Fig. 5c) its relationships with the other taxa are disturbed. We point out that this is inconsistent with Nelson and Platnick’s (1981, p.456) original point of view that both occurrences of the taxa in the same area are considered to be informative. Solving redundancy under A1 in such a way that both sympatric taxa keep their relationships with the other taxa is not possible. Their relationships are forced to change in order to place each area at its own unique terminal node on the area cladogram. To prevent *a priori* disturbance of the relationships of the sympatric taxa with the other taxa we recommend solving redundancy under A1 *a posteriori*, in the same way (“as is”) as under A0, in accordance with the implementation by Zandee and Roos (1987).

Since apart from vicariance events, extinction is considered *a priori* under A1, relative to A0, less *a priori* confidence is placed on distribution and cladogenetic data as representing vicariance events. Consequently, the number of possible area cladograms that can be derived for a group of taxa under A1 increases.

ASSUMPTION 2

A2 is originally described by Nelson and Platnick (1981, p.432). With respect to

widespread taxa, they state under A2 “... whatever is true of the one occurrence might not be true of the other occurrence”. This statement shows the same ambiguity with regard to a taxon *versus* an area relationship approach as described for A1.

For a widespread taxon (T3 in Fig. 6a), under the taxon relationship approach (Humphries, 1982; Humphries and Parenti, 1986 and Zandee and Roos, 1987), A2 implies that a widespread taxon is allowed to be split into separate taxa. *A priori*, besides vicariance and extinction, random dispersal is allowed as well as an explanation for the widespread distribution. As to area relationships a widespread taxon can “speak the truth” only for one of its occurrences. Consequently, the areas in which the widespread taxon occurs are in turn allowed to float while one area stays in a fixed position (Fig. 6b). Thus the widespread taxon is split according to its occurrences and each of these occurrences is successively removed from the taxon-area cladogram. By replacing them in turn on the area cladogram on all possible positions, new resolved area cladograms are constructed (Fig. 6c).

For the taxon-area cladogram (Fig. 6a), this assumption yields two possibilities: A(BC) and A(BD) when taxon T3b in area D or taxon T3a in area C, respectively, is removed from the analysis. Each of these possibilities includes three of the four areas concerned and further steps are necessary to place the taxon occurrences with missing areas in the cladogram. Placing taxon T3b in area D in cladogram A(BC) yields five different cladograms with six different components ((ABC), (BCD), (AD), (BC), (BD), (CD)) and one component in common (ABCD). Placing taxon T3a in area C in cladogram A(BD) also yields five different cladograms with six different components ((ABD), (BCD), (AC), (BC), (BD), (CD)) and one component in common (ABCD). These two outcomes have nine different components in common. Based upon these common components seven different resolved area cladograms can be constructed (Fig. 6c). When these area cladograms are compared with the original taxon cladogram, it appears that for most of the occurrences of the widespread taxon relationships with the other taxa are disturbed. The occurrences of the widespread taxon need not branch off successively and the taxon-area cladogram cannot always be recovered by the removal of a single internode, as is the case under A1. Apart from the *a priori* explanation of

the presence of a widespread taxon (T3) by assuming isolation or break up of areas without speciation (Fig. 6c; “synapomorphy”) or extinction of the widespread taxon (T3) in part of its range (Fig. 6c; “extinction”), the presence of the widespread taxon (T3) can also be explained by *a priori* assuming dispersal (Fig. 6c; “dispersal”).

With respect to redundancy under A2, Nelson and Platnick (1981, p. 457) state “both pieces of information [regarding an area represented more than once in an taxon-area cladogram] need not be true”. This implies that each sympatric taxon is to be considered separately. They thus construct different resolved area cladograms by retaining one occurrence at a time of the sympatric taxa present in the same area as shown in Fig. 7 (Nelson and Platnick, 1981; Morrone and Crisci, 1995). In solving redundancy (Fig. 7a) under A2, two area cladograms B(AC) and A(BC) are derived by the removal of either taxon T1 in area A (Fig. 7b, first cladogram) or taxon T3 in area A (Fig. 7b, second cladogram). According to Nelson and Platnick’s approach only one of these occurrences of a taxon in area A is considered to be informative for the historical relationships among the areas.

Because taxa are excluded from analysis *a priori*, resolved area cladograms derived under A2 are based upon incomplete data. We have shown above that A1 cannot be used to solve redundancy because of the impossibility of deriving resolved area cladograms in which all occurrences of taxa in the same area are considered to be informative. In our view, the way in which A2 solves redundancy should be rejected because information is removed from the analysis and taxa are excluded from the analysis *a priori*. Our recommendation, following Zandee and Roos (1987), is to solve redundancy under A2 “as is”, similarly to A0.

Apart from vicariance events and extinction also dispersal is considered *a priori* under A2. Since random dispersal cannot generate information on the history of areas, we consider information on the distribution and cladogenetic relationships among taxa to represent vicariance events less strictly under A2 than under A0 and A1. Consequently, the number of possible area cladograms increases when (additionally) dispersal is allowed under A2 as a first-order explanation.

TWO REQUIREMENTS FOR OBTAINING VALID COMMON PATTERNS

So far we have described how A0, A1 and A2 are applied in obtaining sets of resolved area cladograms (solution sets S_i) for a single monophyletic group of taxa. In order to infer the general historical pattern for area relationships from several groups of taxa, represented in a general area cladogram, their solution sets must be compared in some way.

In this paper what concerns us are the methodological requirements for a valid comparison. We see a twofold requirement. First, the assumptions, and with them the sets of solutions generated by their protocols, should be inclusive. Second, the area cladograms obtained for the different monophyletic groups should be compared under the same assumption. These two requirements are individually necessary and jointly sufficient. For a particular monophyletic group, a valid comparison allows us to evaluate the relative importance of the various processes that have actually been affecting the group's distribution patterns.

First, the requirement for inclusion. We have described assumptions A0, A1 and A2, without paying attention to the possibility that they show interaction(s) that may affect their solution sets. We consider the different processes underlying the assumptions *a priori* to be independent in their effect. That is, for any particular group we preclude *a priori* the possibility that, for instance, its possible susceptibility to extinction is related to, affected by, or dependent on its dispersal capabilities. Of course, there is a possibility that a poor disperser has a higher chance of extinction. However, we think that this dependence should not be assumed *a priori*, but be revealed from the data *a posteriori*.

In summary, the effects of the separate processes are *a priori* considered to be additive and not multiplicative. In this way, the number of possible area cladograms increases when additional processes are *a priori* assumed under a less strict assumption. The area cladograms already derived under a strict assumption (*e.g.* S_11 in Fig. 8b) are found also under a less strict assumption (*e.g.* S_12 in Fig. 8b). The additional area cladograms derived under the less

strict assumption are the result of the additional *a priori* assumed process (e.g. dispersal under A2 compared to A1 in Fig. 8b). The same processes *a priori* assumed (e.g. vicariance and extinction in Fig. 8b) under both the strict assumptions (e.g. A1 in 8b) and less strict assumptions (e.g. A2 in Fig. 8b) affect the derivation of area cladograms in the same way. Therefore, these identical processes should result in the same area cladograms under both assumptions. Any situation for which this inclusion relation for assumptions does not hold shows inconsistency. In case of non-inclusion (Fig. 8a), the effects of the common processes (e.g. vicariance and extinction in Fig. 8a when comparing S_{11} with S_{12}) *a priori* considered for both the strict assumptions (e.g. A1 in Fig. 8a) and less strict assumptions (e.g. A2 in Fig. 8a) do not result in the same area cladograms. As a result, the common processes are no longer common with respect to the area cladograms that are obtained and the independence of the processes *a priori* assumed is *a priori* precluded. Again, we think this should not be *a priori* assumed but revealed from the data after area cladograms are obtained.

When we compare solution sets, a requirement for a single monophyletic group is also a requirement for a collection of monophyletic groups. Therefore, inclusion of sets of general area cladograms derived under A0, A1 and A2 is also required.

Second, the requirement for comparison under a single assumption. This requirement is illustrated in Fig. 9. Here solution sets for two groups of taxa (G_1, G_2) are obtained under A0 and A1. Common patterns (CP) are found under A1 in the intersection of S_{11} and S_{21} (Fig. 9a). The common patterns are valid because only a single set of processes (A1: vicariance and extinction) is *a priori* assumed to have affected both groups. Of course, different (sets of) processes may have governed the pattern of distribution for the taxa of the separate monophyletic groups. Since we have no *a priori* knowledge of the relative importance of these processes, we can only test our hypotheses on extinction or dispersal as forces generating distribution patterns. All taxa are considered to have responded equally to the possible processes of vicariance, extinction, or dispersal when comparing patterns of distribution of monophyletic groups. The degree to which this actually makes sense for a particular group of taxa is indicated by the degree of fit of its particular cladogenetic and

distribution history to the area history represented in the general area cladogram. For a particular monophyletic group a fair amount of dispersal, for instance, may well have to be assumed *a posteriori* in order to explain the deviation of its pattern from the general pattern. The general area cladogram thus serves as a framework for testing the relative contribution of a particular set of processes (capability for dispersal or susceptibility to extinction) to the pattern of distribution over areas for a group of monophyletic taxa. It can play this role only when the requirement of comparison under a single assumption is met.

Patterns common to S_10 and S_21 (Fig. 9b) as well as S_11 and S_20 (Fig. 9c) are invalid as they would force us to assume *a priori* that the groups differ in their susceptibility to vicariance (S_10 or S_20) and extinction (S_21 or S_11). This would preempt the possibility of testing any hypothesis regarding *e.g.* an inclination towards extinction for G_2 in the comparison of S_11 versus S_20 (Fig. 9c).

THE INCREASE IN EXPLANATORY POWER

In the previous section, we have shown that the assumptions, including their solution sets, should be inclusive for a general area cladogram to serve as framework for the evaluation of a group's actual capability for dispersal and susceptibility to extinction. Moreover, the solution sets to be compared should be obtained under a single assumption.

The latter requirement, however, does not imply that it is forbidden subsequently to consider the intersection of the valid common patterns obtained under a particular set of assumptions with the solution sets of particular groups of taxa under different sets of assumptions. It may occur that the intersection of the solution sets obtained under a single set of assumptions for a number of monophyletic groups (delivering valid common patterns, thus general area cladograms) contain as subsets the intersection with the solution sets obtained under different sets of assumptions for the particular monophyletic groups. This is illustrated in Fig. 10a where the intersection of S_12 with S_22 includes part of S_11 . Though the occurrence

of such non-empty intersections does not contribute to the inference of the general area cladograms, it does enable us to evaluate generically whether, for particular monophyletic groups, an explanation by a smaller set of assumption is sufficient. In Fig. 10a, for instance, we can conclude that, though the general area cladogram(s) demand assumption of all three processes of vicariance, extinction and dispersal, we can refine the explanation of the distribution pattern of G_1 by vicariance and extinction only. In other words, we have gained explanatory power.

The gain in explanatory power with respect to particular groups can be complex (Fig. 10b). There can be more than one solution set obtained under different sets of assumptions for more than one particular monophyletic group which intersect with the set of valid common patterns. The generic evaluations suggested by these intersections cannot, however, all be true simultaneously. For instance, we may infer that for G_1 vicariance is sufficient (S_10) to explain its pattern of distribution, but only when it is simultaneously true for G_2 that the combination of all three processes is required (S_22). Conversely, we may infer that for G_2 vicariance is sufficient (S_20), but only when for G_1 all three processes are required (S_12). However, these two inferences are incompatible. Under *a priori* reasoning the gain in explanatory power for particular groups is likely to decrease when the number of groups considered increases. This can become important when authors add more monophyletic groups to their biogeographical analysis.

VIOLATING REQUIREMENTS: EXAMPLES FROM THE LITERATURE

The requirements as formulated in this paper enable us to discuss examples from the literature (Table 1) with respect to their treatment of issues related to inclusion and single assumption comparison.

Page (1990) and also Morrone and Crisci (1995) state that, because the protocols under the different assumptions are not mutually exclusive, devising other protocols that combine

different aspects of the original protocols is allowed: “For a given set of monophyletic groups we could treat widespread taxa under A2, but redundant distributions under A0” (Page 1990, p. 120). Enghoff (1996) suggests the opposite and deals with widespread taxa under A0 and redundant distributions under A2. As illustrated in Fig. 9, these examples are a violation of the requirement of comparison of solution sets under a single assumption. Consequently, they break down the framework for testing preconceived ideas with respect to, for instance, dispersal capabilities of a particular group.

We have shown above (Fig. 5) that dealing with redundancy via larger patterns as described by Nelson and Platnick (1981) leads to internal inconsistency with A1 as originally described by them. However, there are problems that are more fundamental. When redundancy is solved *a priori*, solution sets derived under A0, A1 and A2 are no longer inclusive (Nelson and Platnick 1981, p. 462, 463). As we illustrated in Fig. 8, non-inclusive assumptions lead to invalid solution sets.

In an attempt to obtain inclusive solution sets, Nelson and Ladiges (1991a,b) use A0, A1 and A2 from another point of view. Based on the cladogenetic and distribution information for taxa of a monophyletic group they derive three-area statements to obtain area cladograms. To make assumptions and with them the solution sets inclusive, Nelson and Ladiges (1991a,b) restrict A1 and A2 to apply to data specified by the nodes of the taxon cladogram. They do that in such a way that the informative node for a three-area statement corresponds to an informative node of the taxon cladogram. By doing this they claim to remove any contradiction between A1 and A2 and obtain the same results under both assumptions. According to Nelson and Ladiges (1991b), three-area statements derived under A2 are now encompassed by those derived under A1 and these in turn are encompassed by the three-area statements derived under A0. However, the area cladograms they derived under A0 need not be a subset of those derived under A1 (Nelson and Ladiges, 1991a, p. 474). The implementation of A0, A1 and A2 by Nelson and Ladiges (1991a,b), therefore, still violates the requirement of inclusive solution sets.

More recently, Nelson and Ladiges (1996) developed another method for derivation of

area cladograms as paralogy-free subtrees. Paralogy is a term used in molecular biology to refer to copies of the same gene in a genome (Fitch, 1970). Duplication of genes in phylogeny obscure information on relationships of the taxa in which these genes occur because multiple copies of the same gene can show different phylogenies deviating from the species phylogeny (Patterson, 1987).

The presence of multiple copies of one gene in a gene tree is analogous to the presence of redundant distributions in a taxon-area cladogram (Page, 1993; Nelson and Ladiges, 1991a). Nelson and Ladiges (1996) describe redundancy in a taxon-area cladogram as geographic paralogy, which they define as duplication or overlap in geographic distribution among related taxa.

For the analysis that they perform, Nelson and Ladiges (1996) first replace the taxa in a taxon cladogram by their areas of distribution. Before subtree analysis, they use A2 to reduce widespread taxa in favor of endemics (Nelson and Ladiges, 1996; Ladiges, 1998). We believe that with this *a priori* operation they assume that the presence of the widespread taxa in these areas is the result of dispersal. With the *a priori* removal of these taxa, they preclude the possibility of the presence of the removed taxa in the areas as a result of non-response to a vicariance event (A0) or widespread presence with extinction in part of its range (A1). In addition, because of the removal of the taxa, the area cladograms derived in this way are less informative because they are based upon incomplete data.

After application of A2, Nelson and Ladiges (1996) use subtree analysis to obtain area cladograms. They start from each terminal node and build subtrees by progressing to the base of the cladogram. The subtrees are coded in three-item statements that are combined in a matrix. Via a parsimony analysis on this three-item matrix, they obtain area cladograms. However, due to reducing widespread taxa in favor of endemics the area cladograms obtained under A0 or A1 are not always found under A2 and the requirement of inclusion is violated.

CONCLUSION

Under A0, A1 and A2 the amount of *a priori* confidence is limited as to the degree in which cladogenetic relationships of taxa and their distribution over areas reflect historical relationships of areas caused by vicariance. Because under these assumptions relationships of areas are inferred by *a priori* assumed processes (*viz.* vicariance, extinction, or dispersal) that apply to taxa they should, in our view, be applied in a taxon relationship approach.

When resolved area cladograms are obtained according to the protocols under A0, A1 or A2, inclusive solution sets can be produced if and only if sympatric taxa (redundancy) are dealt with *a posteriori*. These inclusive solution sets are in agreement with the inclusive sets of processes *a priori* assumed under A0, A1 and A2:

- sets of processes are related similarly to vicariance (A0) \subset vicariance + extinction (A1) \subset vicariance + extinction + dispersal (A2); and
- solution sets are related similarly to $S_{i0} \subseteq S_{i1} \subseteq S_{i2}$.

Two requirements should be met to make valid comparisons of solutions sets of different monophyletic groups aimed at obtaining general area cladograms (common patterns):

- inclusion of assumptions and with them of the solution sets; and
- comparison of solution sets under a single assumption.

As a bonus, the valid common pattern offers an increase in explanatory power for explaining the distribution of the taxa from any particular monophyletic group for which it holds that its solution set for a more strict assumption is part of the common pattern.

In a companion paper we shall investigate how the different methods have implemented A0, A1 and A2 and whether the requirement of inclusive solution sets is met when the

different methods are applied on theoretical and empirical datasets.

ACKNOWLEDGEMENTS

We thank three anonymous reviewers for their comments on a previous version of the manuscript. This work was supported by the Life Sciences Foundation (SLW), which is subsidized by the Netherlands Organization for Scientific Research (NOW); grant 805-33.193 to MvV.

REFERENCES

- Brooks, D.R. (1990). Parsimony analysis in historical biogeography and coevolution: Methodological and theoretical update. *Syst. Zool.* **39**, 14-30.
- Enghoff, H. (1996). Widespread taxa, sympatry, dispersal and an algorithm for resolved area cladograms. *Cladistics* **12**, 349-364.
- Fitch, W.M. (1970). Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**, 99-113.
- Humphries, C.J. (1982). Vicariance biogeography in Mesoamerica. *Ann. Missouri Bot. Gard.* **69**, 444-463.
- Humphries, C.J. and Parenti, L.R. (1986). "Cladistic biogeography". Clarendon Press. Oxford.
- Kluge, A.G. (1988). Parsimony in vicariance biogeography: A quantitative method and a Greater Antillean example. *Syst. Zool.* **37**, 315-328.
- Ladiges, P.Y. (1998). Biogeography after Burbidge. *Australian Systematic Botany* **11**, 231-242
- Morrone, J.J. and Carpenter, J.M. (1994). In search of a method for cladistic biogeography: an empirical comparison of Component Analysis, Brooks Parsimony Analysis and Three-

- area statements. *Cladistics* **10**, 99-153.
- Morrone, J.J. and Crisci, J.V. (1995). Historical biogeography: Introduction to methods. *Ann. Rev. Ecol. Syst.* **26**, 373-401.
- Nelson, G. and Ladiges, P.Y. (1991a). Three area statements: standard assumptions for biogeographic analysis. *Syst. Zool.* **40**, 470-485.
- Nelson, G. and Ladiges, P.Y. (1991b). Standard assumptions for biogeographic analysis. *Aust. Syst. Bot.* **4**, 41-58.
- Nelson, G. and Ladiges, P.Y. (1996). Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *American Museum Novitates* **3167**, 58pp.
- Nelson, G. and Platnick, N.I. (1981). "Systematics and biogeography: Cladistics and vicariance". Columbia University Press. New York. 567 pp.
- Page, R.D.M. (1988). Quantitative cladistic biogeography: Constructing and comparing area cladograms. *Syst. Zool.* **37**, 254-270.
- Page, R.D.M. (1989). Comments on component compatibility in historical biogeography. *Cladistics* **5**, 167-182.
- Page, R.D.M. (1990). Component analysis: A valiant failure ? *Cladistics* **6**, 119-136.
- Page, R.D.M. (1993). Genes, organisms and areas: The problem of multiple lineages. *Syst. Biol.* **42**, 77-84.
- Patterson, C. (1987). "Molecules and morphology in evolution: conflict or compromise?". Cambridge University Press. Cambridge, London, New York, New Rochelle, Melbourne, Sydney. 229 pp.
- Platnick, N.I. (1988). Systematics, evolution and biogeography: A Dutch treat. *Cladistics* **4**, 308-313.
- Rosen, D.E. (1978). Vicariant patterns and historical explanation in biogeography. *Syst. Zool.* **27**, 159-188.
- Soest, R.W.M. van (1996). Recoding widespread distributions in general area cladogram construction. *Vie Milieu* **46**, 155-161.
- Wiley, E.O. (1988a). Parsimony analysis and vicariance biogeography. *Syst. Zool.* **37**, 271-

Wiley, E.O. (1988b). Vicariance biogeography. *Ann. Rev. Ecol. Syst.* **19**, 513-542.

Zandee, M. and Roos, M.C. (1987). Component-compatibility in historical biogeography.
Cladistics **3**, 305-332

LIST OF FIGURES AND TABLES

FIG. 1. Solution of redundant distributions (two species in area A) that are the result of sympatric speciation (species T2 and T3).

FIG. 2. Solution of a widespread distribution of taxon T3 in areas C and D under A0 (a: taxon cladogram with areas, b: area cladogram under A0 with a “synapomorphy”).

FIG. 3. Solution of a redundant distribution of both taxon T1 and T3 in area A under A0 (a: taxon cladogram with areas, b: area cladogram under A0 with an “extra step”).

FIG. 4. Solution of a widespread distribution of taxon T3 in areas C and D under A1 (a: taxon cladogram with areas, b: area cladogram with a polytomy, c: area cladograms under A1 with a “synapomorphy” or “extinctions”).

FIG. 5. Solution of a redundant distribution of both taxon T1 and T3 in area A under A1 (a: taxon cladogram with areas, b: “larger” pattern derived by hypothesizing “extinctions”, c: area cladograms under A1 extracted from larger pattern).

FIG. 6. Solution of a widespread distribution of taxon T3 in areas C and D under A2 (a: taxon cladogram with areas, b: removal and “floating” of one of the widespread occurrences hypothesizing dispersal, c: area cladograms under A2 with a “synapomorphy”, “extinctions” or “dispersals”).

FIG. 7. Solution of a redundant distribution of both taxon T1 and T3 in area A under A2 (a: taxon cladogram with areas, b: area cladograms under A2 by removal of one of the sympatric taxa hypothesizing dispersal).

FIG. 8. Solution sets derived under assumptions zero (S_10), 1 (S_11) and 2 (S_12) for a single group of taxa (a: non-inclusive solution sets, b: inclusive solution sets, V = Vicariance, E = Extinction, D = Dispersal).

FIG. 9. Search for the intersection in solution sets (S_10 , S_11 , S_12) to find common patterns (CP_{AiAj}) for two groups of taxa under the same assumption (a: valid common patterns in the intersection of S_11 and S_21 , b: invalid common patterns in the intersection of S_10 and S_21 , c: invalid common patterns in the intersection of S_11 and S_20).

FIG. 10. Complex common patterns (CP_{AiAj}) allowing an increase in explanatory power (a: CP_{22} under A2 for both G_1 and G_2 and as a subset CP_{12} for G_1 under A1 and G_2 under A2, b: CP_{22} under A2 for both G_1 and G_2 and several subsets CP_{12} , CP_{21} , CP_{02} , CP_{20} for G_1 and G_2 under different

assumptions).

TABLE 1: Studies in which authors deal with widespread and sympatric taxa for the derivation of area cladograms.

Author (year)	<i>A priori</i> dealing with widespread taxa	<i>A priori</i> dealing with sympatric taxa	Area cladograms
Rosen (1978) [1]	excluded from analysis	excluded from analysis	reduced for areas with widespread or sympatric taxa
Kluge (1988) [1]	coded as missing data	weighted according to a minimal number of vicariance events	only based on non-widespread taxa
Van Soest (1996) [2]	areas with widespread taxa form a single area with a history different from its constituent areas	coded “as is”	with extra areas (multiple histories) for widespread taxa
Zandee and Roos (1987) [3]	A0: occurrences of widespread taxon form a sister group A1: occurrences of widespread taxon form a sister group or branch off sequentially A2: one occurrence of the widespread taxon fixed position, other occurrences “float”; and vice versa	A0: coded “as is” A1: coded “as is” A2: coded “as is”	A0: based on all taxa A1: based on all taxa A2: based on all taxa
Wiley (1988a,b), Brooks (1990) [3]	A0: occurrences of widespread taxon form a sister group	A0: coded “as is”	A0: based on all taxa
Nelson and Platnick (1981) [3]	A1: occurrences of widespread taxon form a sister group or branch off sequentially A2: one occurrence of the widespread taxon fixed position, other occurrences “float”; and vice versa	A1: all sympatric taxa part of larger pattern A2: all but one of the sympatric taxa removed	A1: based on all taxa plus additional assumed (extinct) taxa A2: based on all non-sympatric taxa and only one of the sympatric taxa
Humphries (1982), Humphries and Parenti (1986) [3]	A1: occurrences of widespread taxon form a sister group or branch off sequentially A2: one occurrence of the widespread taxon fixed position, other occurrences “float”; and vice versa	A1: not discussed A2: not discussed	A1: based on all taxa A2: based on all taxa
Page (1988, 1989, 1990, 1993), Morrone and Carpenter (1994), Morrone and Crisci (1995) [4]	A0: areas with widespread taxon have monophyletic relationship A1: areas with widespread taxon have mono- or paraphyletic relationship A2: areas with widespread taxon have mono- or para- or	A0: not discussed A1: extinction of taxa in areas without sympatric taxa A2: distribution of all but one sympatric taxa result of	A0: based on all taxa A1: based on all taxa plus additional assumed (extinct) taxa A2: based on all non-sympatric taxa and only one of the sympatric

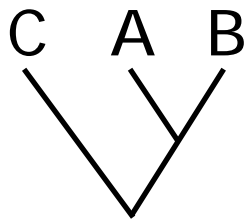
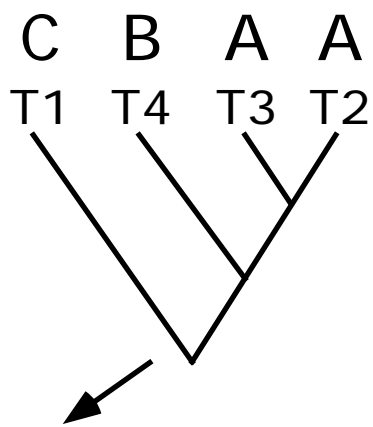
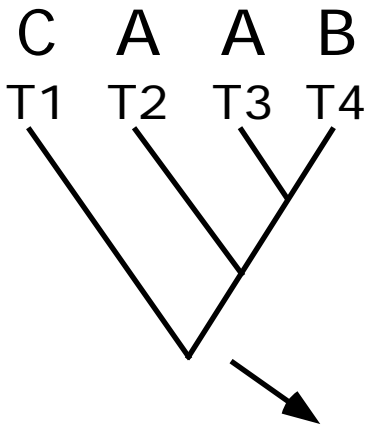
	polyphyletic relationship	dispersal and therefore removed	taxa
Enghoff (1996)	A0: occurrences of widespread taxon form a sister group	A2: distribution of all but one sympatric taxa result of dispersal and therefore removed	A02: based on all non-sympatric taxa and only one of the sympatric taxa
Nelson and Ladiges (1991a,b)	A0: widespread taxon contributes an extra three area statement A12: widespread taxa do not contribute three area statements	A0: not discussed A12: not discussed	A0: based on all taxa A12: based on all non-widespread taxa
Nelson and Ladiges (1996), Ladiges (1998)	widespread taxa removed in favor of endemics	subtrees derived by removal of all but one of the sympatric taxa	based on all non-sympatric taxa and only one of the sympatric taxa

[1] = A0, A1 or A2 are not applied; widespread or sympatric taxa are dealt with by *a priori* excluding these data from analysis.

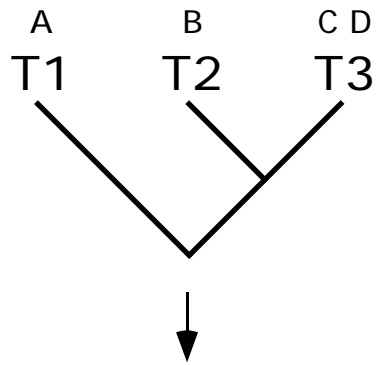
[2] = A0, A1 or A2 are not applied; widespread taxa are dealt with by *a priori* assuming multiple histories for the areas with widespread taxa and thereby excluding these data from analysis.

[3] = Taxon relationship approach; A0, A1 and A2 are interpreted to pertain to taxon relationships.

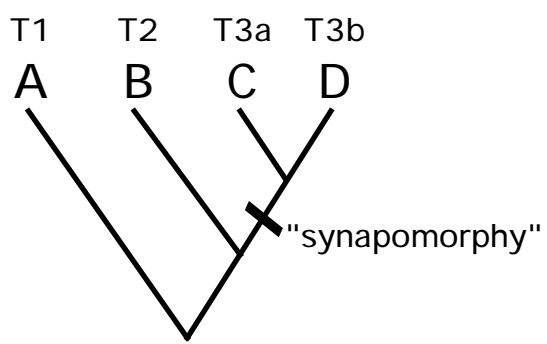
[4] = Area relationship approach; A0, A1 and A2 are interpreted to pertain to area relationships.



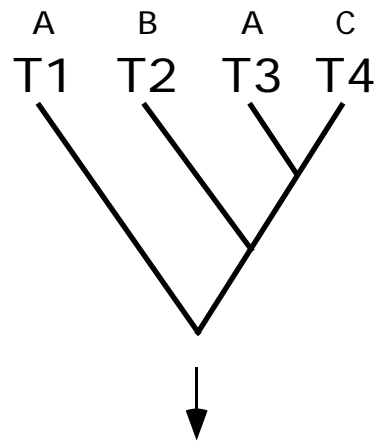
a



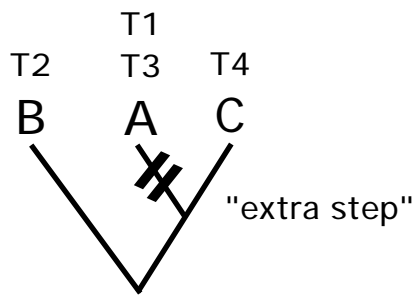
b



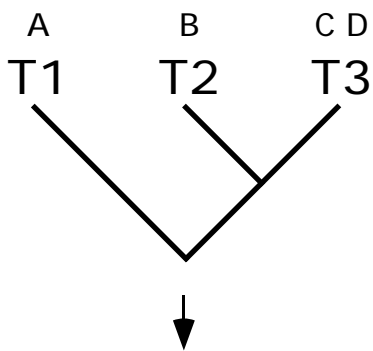
a



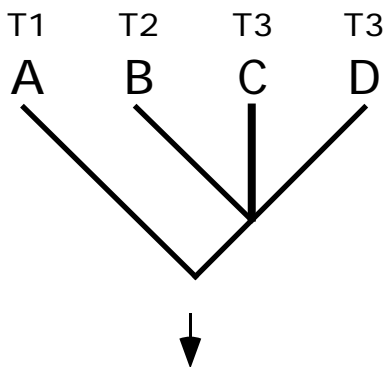
b



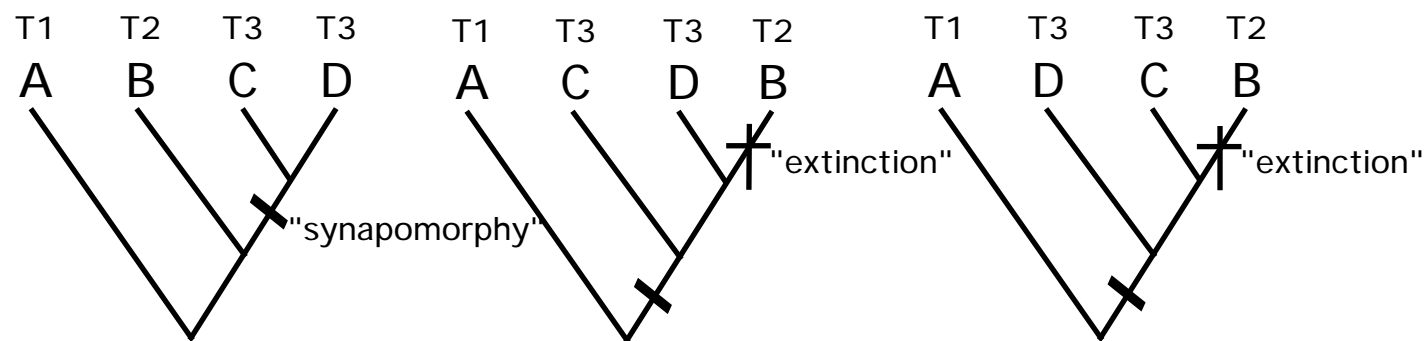
a



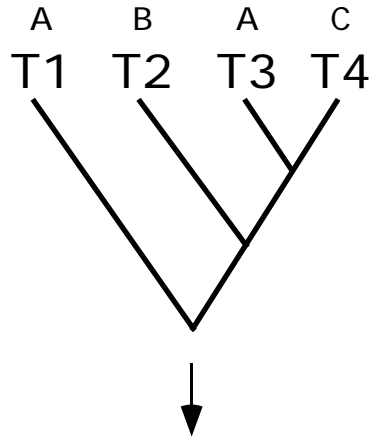
b



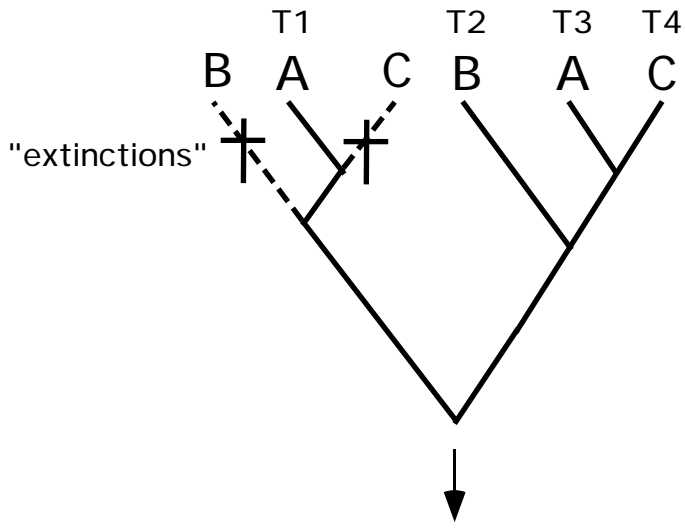
c



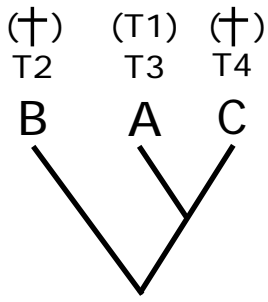
a



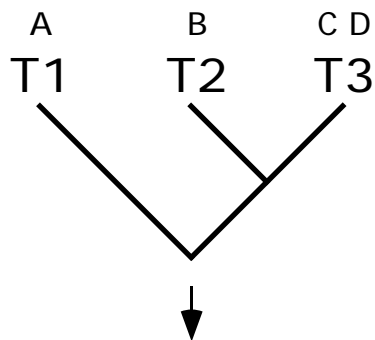
b



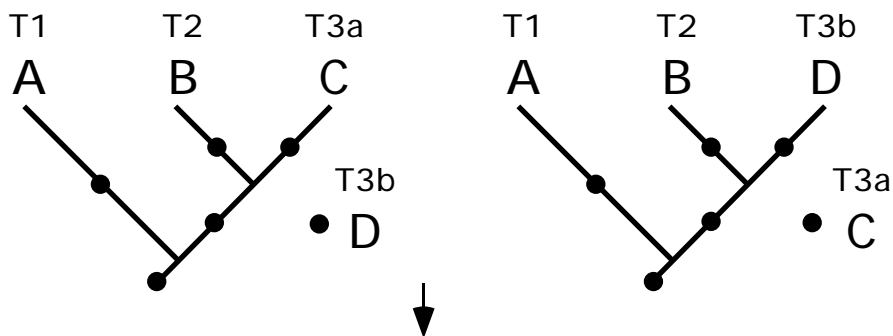
c



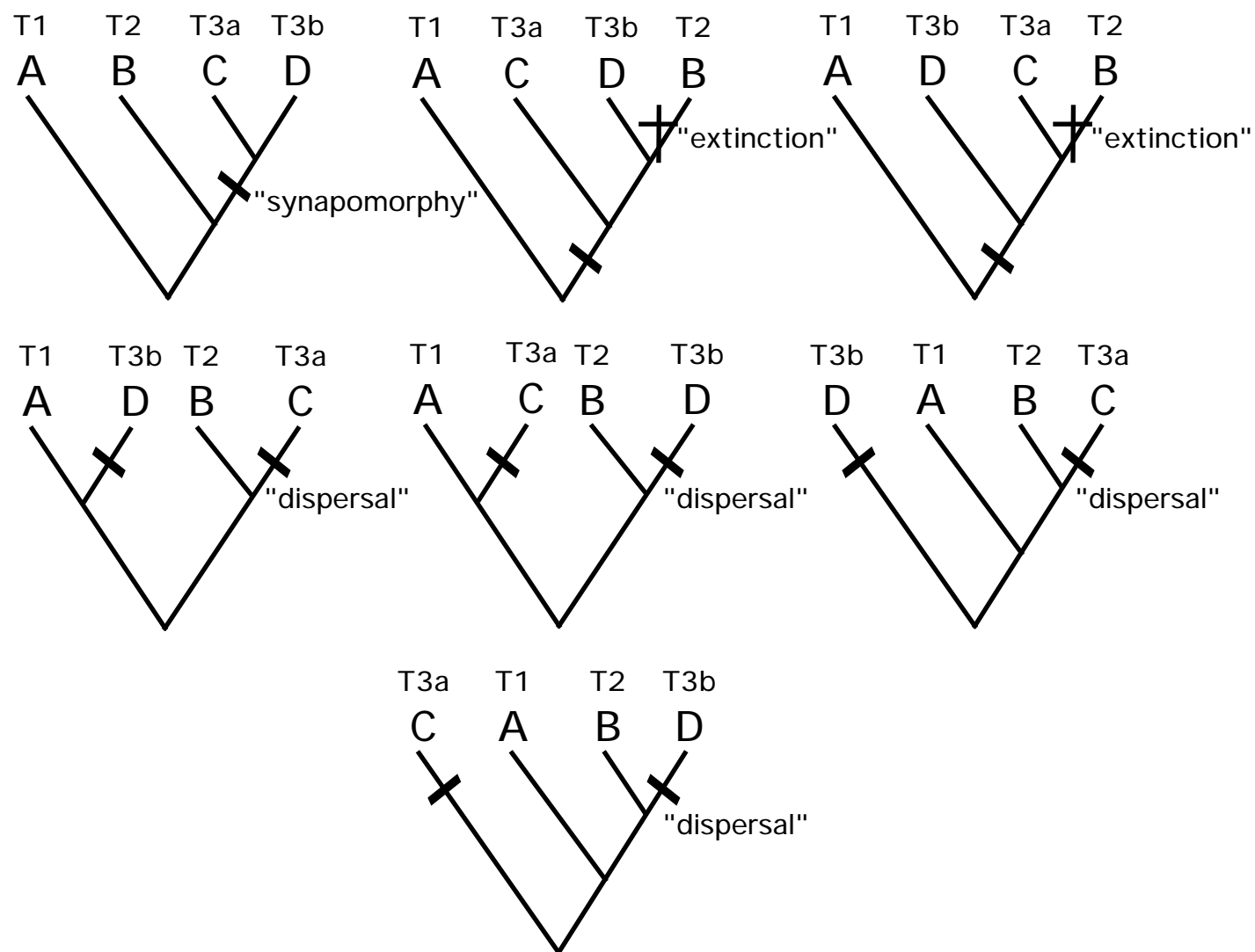
a



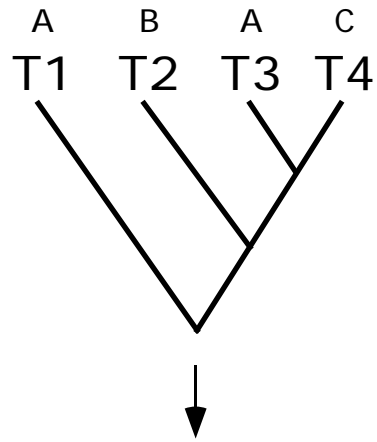
b



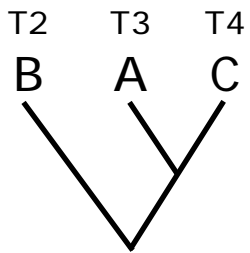
c



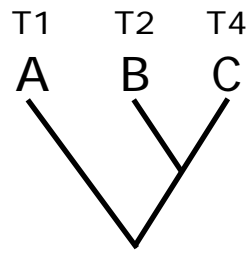
a



b

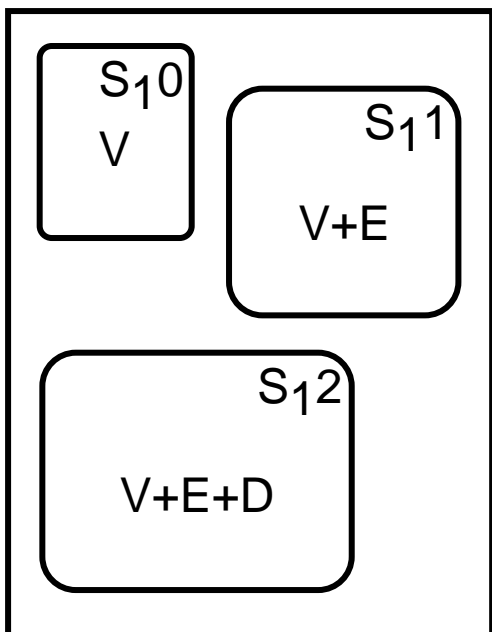


removal of T1 in area A
("dispersal")

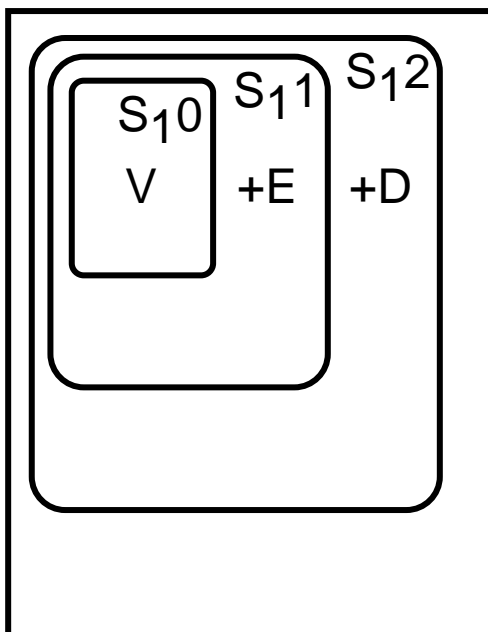


removal of T3 in area A
("dispersal")

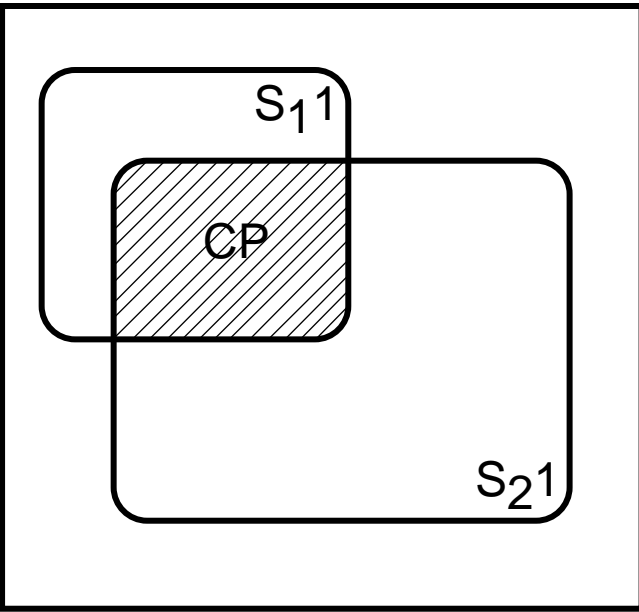
a



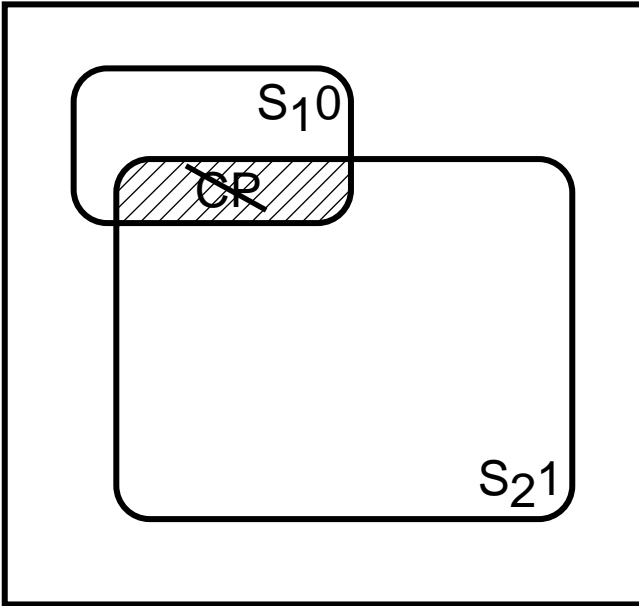
b



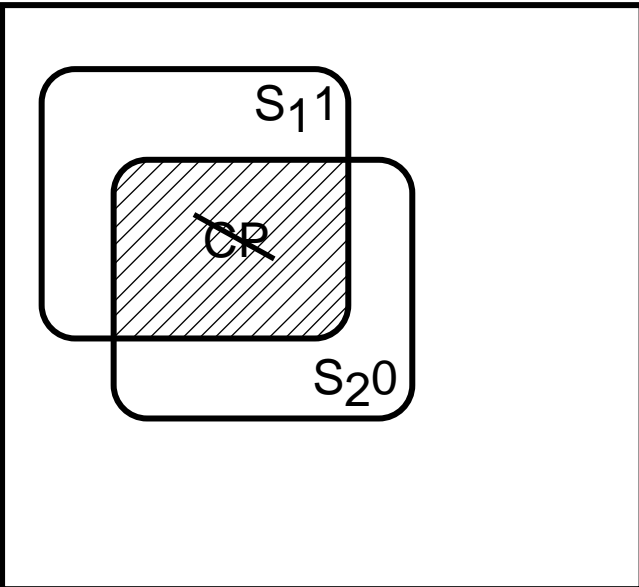
a



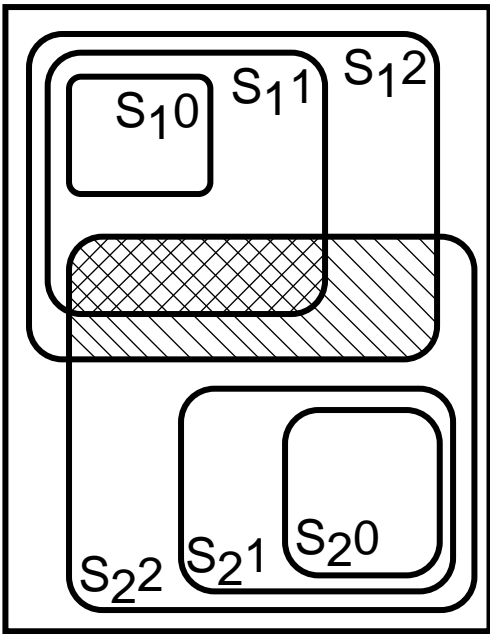
b



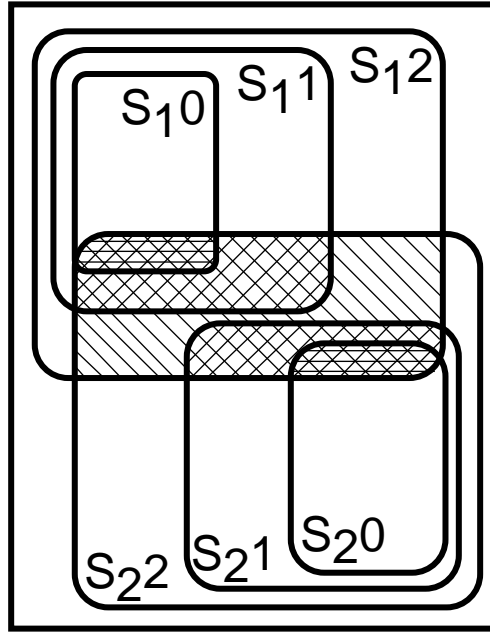
c



a



b



CP_{22}

$CP_{12/21}$

$CP_{02/20}$