PREPRINT

Methods in vicariance biogeography: assessment of the implementations of assumptions zero, 1 and 2

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

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 5274900.

Running title: Assessing methods in vicariance biogeography

Keywords: Vicariance biogeography, Assumptions, Inclusion, Methods, Programs, Component Compatibility Analysis, Brooks Parsimony Analysis, Component Analysis, Reconciled Tree Analysis, Three Area Statement Analysis, CAFCA, Component 1.5, Component 2.0, TAS, *Heterandria*, *Xiphophorus*, *Cyttaria*, *Eriococcus/Madarococcus*.

ABSTRACT

As we argued earlier, for the valid derivation of general area cladograms in vicariance biogeography, two requirements should be met. First, sets of areas cladograms derived under assumptions zero, 1 and 2, should be inclusive (requirement I). Second, general area cladograms should be based on area cladograms, for different monophyletic groups, derived under the same assumption (requirement II). We now asses for their actual implementation of assumptions A0, A1, and A2 and for the extent to which they meet requirements I and II, the following methods (and correlated programs): Component Compatibility Analysis (CAFCA), Brooks Parsimony Analysis (PAUP), Component Analysis (Component 1.5), Reconciled Tree Analysis (Component 2.0) and Three Area Statement Analysis (TAS). For this purpose we use empirical (*Heterandria*, *Xiphophorus*, *Cyttaria*, *Eriococcus*/*Madarococcus*) and theoretical data sets. All programs appear to violate, to a different degree, requirement I (deriving inclusive sets of area cladograms under assumptions) when dealing with sympatric taxa under A1 or A2. Dealing with sympatric taxa *a posteriori* only prevents this violation. All programs examined appear to meet requirement II (deriving general area cladograms under a single assumption).

INTRODUCTION

In vicariance biogeography the first-order explanation for the distribution of taxa over areas is by vicariance events triggering speciation in species of various monophyletic groups. This implies that the cladogenetic and distribution data of the taxa of these monophyletic groups are considered informative for the reconstruction of the historical relationships among their areas of distribution. To obtain an initial hypothesis on the historical relationships of the areas, the taxa of a taxon cladogram (the hypothesis of the historical taxon relationships) are replaced by their areas of distribution, resulting in a taxon-area cladogram (Morrone and Carpenter, 1994; Enghoff, 1996).

If the distribution of taxa from a particular monophyletic group over areas indeed results from vicariance events only, each taxon will be present in a single area and each area will harbour a single taxon. Data without widespread or sympatric taxa render taxon-area cladograms with for each area an own and unique terminal node. In such straightforward cases the area cladogram is identical to the taxon-area cladogram (Rosen, 1978; Nelson and Platnick, 1981; Page, 1988; Morrone and Carpenter, 1994).

However, the distribution of taxa from a particular monophyletic group can as well be caused by processes such as the origin of a barrier or the break up of an area without speciation, sympatric speciation, extinction and dispersal, which may lead to widespread and sympatric distributions of taxa. The methods we examine all aim to conjecture area cladograms with own and unique terminal nodes for each area, so each of them has to deal with the multiple areas at a terminal node (resulting from widespread taxa) and redundancy of areas (resulting from sympatric taxa). In addition to the assumption of vicariance as first-order explanation (A0: Zandee and Roos, 1987; Wiley, 1988a; Brooks, 1990), more relaxed assumptions (A1 and A2, Nelson and Platnick, 1981; Page, 1988) have been proposed to derive area cladograms. In this paper we will assess the precise implementation of assumptions A0, A1, and A2 by various methods. The methods vary in two main ways regarding how they infer a list of components (each consisting of an area or a combination of areas) from the taxon area cladograms under A0, A1 and A2. From these lists of components they subsequently derive area cladograms with own and unique terminal nodes for all areas.

The *a posteriori* methods Component Compatibility Analysis (CCA; Zandee and Roos, 1987) and Brooks Parsimony Analysis (BPA; Brooks, 1990; Wiley, 1988a,b) do not allow a distortion of the historical relationships in the taxon-area cladogram for which the already established original taxon cladogram was the source. To deal with two areas at a single terminal node of a taxon-area cladogram, nodes (*i.e.*

components) are just added to interpret the as yet unhypothesized historical relationship between these two areas. Under A0, A1 and A2, the possible places of the additional nodes are interpreted with increasing and inclusive degrees of freedom. In cases of redundancy of areas the *a posteriori* methods refrain from additions to the data all together. Incongruent data that remain are explained *a posteriori* (taxon history protocol *sensu* Hovenkamp, 1997 or taxon relationship approach *sensu* Van Veller *et al*., 1999).

A priori methods such as Component Analysis (CA; Nelson and Platnick, 1981; Page, 1988, 1990), Reconciled Tree Analysis (RTA; Page, 1993a, 1994) and Three Area Statement Analysis (TAS; Nelson and Ladiges, 1991a,b,c), on the other hand, allow under A1 and A2 pruning and adding of taxa and taxon relationships. Thereby, these methods fit an explanation by the assumed processes under consideration (extinction, dispersal, etc.) only and overrule the historical relationships that were already established (to fit an explanation by the assumed processes only).

Morrone and Carpenter (1994) evaluated different methods used in vicariance biogeography. They compared area cladograms for different data sets obtained by CA, RTA, TAS and BPA on basis of items of error or agreement and found themselves unable to prefer one method over another. This outcome is perhaps not surprising since CCA, BPA and TAS do not use at all items of error for the selection of area cladograms, but minimal number of steps. Lack of agreement (between the obtained area cladograms) is as well to be expected since, as we will show in this paper, not all methods obtain area cladograms validly.

In a previous paper (Van Veller *et al*., 1999) we developed a methodological framework for the valid derivation of general area cladograms. As we showed, two requirements have to be met:

I. Solution sets containing the area cladograms derived under A0, A1 and A2 for one group of taxa should be inclusive.

II. Solution sets for two or more groups of taxa should be compared under the same assumption.

This paper assesses to what extent requirements I and II are met by the implementations of the assumptions in the following five methods: CCA, BPA, CA, RTA and TAS.

IMPLEMENTATIONS OF A0, A1 AND A2 IN METHODS FOR VICARIANCE BIOGEOGRAPHY

In this part we discuss how CCA, BPA, CA, RTA and TAS (and their implementations in software) obtain area cladograms from cladogenetic and distribution data of the taxa of a monophyletic group under the different assumptions. For a schematic overview of the procedures followed in application of the methods, we refer to appendix i.

Component Compatibility Analysis (CCA)

In CCA, the data matrix that is used to derive area cladograms comprises a mapping of the cladogenetic relationship of the taxa onto the areas in which they occur (*i.e.* a representation of the taxon-area cladogram). This mapping is obtained by combining the taxon cladogram with the taxon distribution over the areas ("inclusive ORing", O'Grady and Deets, 1987; "Boolean inner product", Zandee and Roos, 1987). The part of the matrix that represents the inner nodes of the taxon cladogram is given as a single multistate character. The states of this character represent the additive binary codes of the inner nodes of the taxon cladogram, and are treated accordingly during the cladogram optimization.

In CCA the nodes of a cladogram represent components. The components are defined as partial monothetic sets of areas (Zandee and Roos, 1987) and are characterised by unique character states. Components are extracted from the binary representation of the area by node data matrix by applying this definition. Area cladograms are derived from the list of components by letting a branch and bound algorithm (Bron and Kerbosch, 1980) search for the largest sets of mutual compatible components (maximal cliques). Components are compatible when they either include or exclude each other and do not overlap (Nelson, 1979). Each of these maximal cliques corresponds with an area cladogram. Parsimony mapping of the area by node data matrix finds the area cladograms of minimum length (number of steps).

In CCA, areas with sympatric taxa (redundancy) are considered to be analogous to taxa with more than one autapomorphic character. In a standard cladistic character analysis such taxa are not considered to present a problem that needs to be solved *a priori*. Therefore, by analogy, occurrence of two or more (sympatric) taxa in one area is interpreted "as is" and dealt with by the derivation of a cladogram from the data matrix (Zandee, 1999; Zandee and Roos, 1987). Thus, in CCA no special procedure is applied to deal with redundancy, neither under A0, nor under A1 or A2.

In CCA widespread taxa can be dealt with under either A0, A1 or A2. Under A0 the areas of the widespread taxa are considered to be sister areas. No additional provisions to the data matrix are made under A0. The implementation of A1 implies the derivation of additional columns for the data matrix by combining all subsets of areas of a widespread taxon with the areas of its sister group. These columns are used to extract additional components. As these columns represent assumptions and not observations they are not used in the calculation of the cladogram length. The implementation of A2 also implies the derivation of additional columns for the data matrix. These columns are obtained by letting the areas of a widespread taxon be able to float over the cladogram by combining all subsets of these areas with the distributions of all other clades in the cladogram (Zandee and Roos, 1987). Again, these columns are only used to derive additional components and do not enter the computation of cladogram length.

The program used in this study to perform CCA is CAFCA (vs. 1.5j; Zandee, 1999).

Brooks Parsimony Analysis (BPA)

In BPA, data on the distribution of the taxa and the taxon cladogram are combined in a binary area by node data matrix through inclusive ORing (O'Grady and Deets, 1987). This matrix represents a taxon-area cladogram, obtained when the taxa at the terminal nodes of a taxon cladogram are replaced by their areas of distribution. However, in contrast to CCA, and to polarise the data, a hypothetical outgroup with all zeros (and a one for the root of each taxon cladogram; Brooks, pers. com.) is added. This data matrix is used to derive area cladograms of minimal length (number of steps) under A0, using a standard maximum parsimony approach as implemented in PAUP (vs. 3.11; Swofford, 1990) or Hennig86 (Farris, 1988).

Like in CCA, in BPA, areas with sympatric taxa (redundancy) are interpreted "as is" and the incongruencies they pose are explained *a posteriori* via extinction or dispersal (Brooks, 1990; Van Veller *et al*., 1999).

The BPA protocol (Brooks, 1990) does not provide instructions for the implementation of A1 and A2. However, in order to assess BPA for its implementation of A0, A1, and A2, data matrices have to be derived under A1 and A2 as well. In this study, we derive these data matrices by using CCA's protocol with an all-zero (except for a single 1 for the root of each taxon cladogram; Brooks, pers. com.) outgroup added. However, in contrast to CCA, in BPA the additional extra columns derived to implement both A1 and A2 are treated as real data and are therefore included in the computation of cladogram length.

Component Analysis (CA)

In CA (as implemented by Page, 1988), area cladograms can be derived under one of three different assumptions (A0, A1 or A2) conditional *a priori* on the nature of the processes (vicariance, vicariance and extinction or vicariance, extinction and dispersal) one assumes to have resulted in the pattern of distribution of the taxa involved. Under A0, a binary coded area by node data matrix (representation of the taxon-area cladogram) obtained via Brooks' (1981) coding method is used in the analysis. As in BPA, a hypothetical outgroup (a row with zeros) is added for polarisation. A branch and bound algorithm (Hendy and Penny, 1982) is used to find area cladograms of minimal length (number of steps) (Page, 1988, 1990). Like in BPA and CCA, areas with sympatric taxa (redundancy) are interpreted "as is". Consequently, an A0 analysis with CA is expected to be similar to a BPA analysis in both procedure and results.

For an analysis under A1, the binary coded area by node data matrix constructed under A0 is adjusted for nodes in the taxon-area cladogram with widespread or sympatric taxa. The adjustments for terminal and internal nodes are as follows.

Terminal nodes containing two or more areas that are not redundant are excluded from further analysis since widespread taxa do not contribute components (*sensu* Nelson and Platnick, 1981). The two or more areas present at the excluded terminal node collapse with their sister areas to polytomies when area cladograms are derived from the adjusted data matrix via a standard maximum parsimony approach (as is used in BPA; Fig. 1). However, the branch and bound algorithm used in CA solves these polytomies in all possible dichotomies (Fig. 1).

Internal nodes with descendant lineages with overlapping sets of areas are identified as redundant nodes. Areas only present at one descendant lineage of a redundant node and not in the other descendant lineage of a redundant node are assumed to be missing due to extinction, failure of collection or incorrect identification of one or more taxa (Page, 1988). These areas are identified and coded as question marks in the adjusted data matrix (Fig. 2). These question marks are optimised to either zeros or ones, whichever is most parsimonious (Fig. 2). Via this optimisation, certain columns in the A1 matrix are adjusted and different from the corresponding columns in the A0 matrix. The different components that these adjusted columns define stand for the extinctions, failures of collection or incorrect identifications of taxa that explain the redundancy in the data (Fig. 2).

For an A2 analysis with CA, in contrast to CCA and BPA as well as the analysis under A0 or A1 with CA, no area by node data matrix is used. Adjustments needed to allow for widespread taxa and sympatric taxa under A2 are made by direct manipulation of the taxon-area cladogram. This taxon-area cladogram is obtained in the standard fashion by replacing taxa by areas of distribution in the taxon cladogram. In this taxon-area cladogram terminal nodes containing two or more areas (resulting from widespread taxa) are reduced (for areas) by removing redundant areas from these nodes that are present at own terminal nodes as well (Page, 1990). After dealing with this combination of widespread and sympatric taxa, the taxon-area cladograms are further reduced by removing areas from terminal nodes with more than one area (widespread taxa) and removing redundant occurrences of areas at different terminal nodes. The areas not present in these reduced area cladograms, are the result of the removal of areas with widespread taxa. By placing these areas back in the cladogram at different positions, the areas of the widespread taxon are allowed to float over the whole cladogram thereby deriving (non-reduced) area cladograms (Page, 1988, 1990).

The program used in this study to perform CA is Component (vs. 1.5; Page, 1990)

Reconciled Tree Analysis (RTA)

In RTA, in contrast to CCA, BPA and CA (under A0 and A1), no data matrix is used to derive area cladograms. A taxon-area cladogram is obtained in the standard fashion by replacing the taxa in the taxon-cladogram by their areas of distribution. Each node in this taxon-area cladogram corresponds with a component (Page, 1993a). However, in contrast to BPA, CCA and CA, these components are not represented as columns in a matrix. In RTA, the area cladograms are derived by comparison of estimated area cladograms with the taxon-area cladogram.

RTA, in contrast to BPA, CCA and CA, deals with widespread taxa by means of a procedure that is rather different from the one that it uses to deal with sympatric taxa, although in both cases one of the three different assumptions (A0, A1 or A2) is invoked. Also, as in CA, dealing with widespread taxa has precedence over dealing with sympatric taxa.

First, widespread taxa are dealt with by mapping the taxon cladogram on the taxon-area cladogram. Under A0, each area of a widespread taxon gets its own terminal node by replacing the widespread taxon by its areas of distribution and introducing extra branches for each area. These branches (leading each to an area of a widespread taxon) are connected via internal nodes in such way that the areas form a monophyletic group that corresponds with a component. As a result, an A0 analysis for widespread taxa with RTA is expected to be similar in results to a CCA, BPA or CA analysis under A0. Under A1, the areas of a widespread taxon are not mapped separately, but are included in the range of the ancestor of the widespread taxon (Page, 1994). As a result, no component is defined for these areas (*sensu* Nelson and Platnick, 1981). The areas of the range of the ancestor of the widespread taxon are connected via internal nodes and full dichotomous area cladograms are derived. The results are expected to be similar to the results of a CA analysis under A1. However, Enghoff (1998) criticises results obtained with RTA under A1 when dealing with widespread taxa at basal positions in taxon-area cladograms. Under A2, only one of the areas of the widespread taxon is mapped and the other areas are given the ability to float over the whole cladogram.

Second, sympatric taxa are dealt with in RTA under A0, A1 and A2 via tree reconciliation (Page, 1993b, 1994). Therefore we call this method Reconciled Tree Analysis (RTA). However, when widespread taxa are combined with sympatric taxa, under A2, the distribution of the widespread taxa is reduced in favour of endemics. This is conform Nelson and Platnick's (1981) and Page's (1988) handling of combinations of widespread and sympatric taxa.

The presence of the same area at different terminal nodes in the taxon-area cladogram (redundancy) is explained by reconciling the taxon-area cladogram with an intial area cladogram (which is estimated from the taxon-area cladogram by pruning redundant areas). To measure the degree of fit between the taxon-area cladogram and an area cladogram, Page (1988, 1993a, 1994) suggests three criteria:

- duplications (number of times a lineage of the taxon cladogram has to be duplicated for reconciliation),
- areas added (*i.e.* half the number of items of error); or
- independent losses (number of areas or complete monophyletic groups of areas that have to be assumed for reconciliation).

Different reconciliations can be obtained by changing the area cladogram via branch swapping. Reconciliations between the taxon-area cladogram and an area cladogram that need a minimum of duplications or independent losses are preferred. The area cladograms used for these (minimal) reconciliations are selected.

The area cladograms obtained with RTA can differ from the area cladograms selected via minimisation of items of error (CA *sensu* Nelson and Platnick, 1981) because duplications are not counted in terms of items of error and because a single loss can stand for several areas (together in one clade) that have to be added to the reconciliation.

The program used in this study to perform RTA is Component 2.0 (Page, 1993a). For selection of minimal reconciliations, both duplications and independent losses are counted.

Three Area Statement Analysis (TAS)

In TAS, just like with CCA, BPA and CA (under A0 and A1), a data matrix is used to derive area cladograms. However, in contrast to these methods the data matrix is not a direct representation of the taxon-area cladogram but consists of three area statements. The taxon-area cladogram is derived in the same way as with CCA, BPA, CA or RTA by replacement of the taxa at the terminal nodes of the taxon cladogram with the areas in which they are distributed. From this cladogram the matrix with three area statements is derived.

Three area statements (Nelson and Ladiges, 1991a,b) in area cladistics are analogous to three taxon statements (Nelson and Platnick, 1991) in taxon cladistics.

Three area statements are derived for each node of the taxon-area cladogram by coding which two areas are more related to each other than a third and coding the rest of the areas as question marks. By combining all different three area statements for each node and by transforming all nodes into combinations of three area statements, the taxon-area cladogram is translated into a matrix of three area statements. By recognition of different nodes in the taxon-area cladogram, a matrix of three area statements is derived under A0, A1 and A2.

Under A0 three area statements are derived by recognition of the widespread taxon as an extra node. Under both A1 and A2 only the internal nodes are used to derive three area statements.

Sympatric taxa are dealt with under both A0 and A1 by the derivation of three area statements and construction of area cladograms from these three area statements. As a result, under both A0 and A1, for sympatric taxa, the same area cladograms are expected to be obtained. Dealing with sympatric taxa under these two assumptions is similar to dealing with sympatric taxa "as is" with BPA (under A0, A1 or A2), CCA (under A0, A1 or A2) or CA (under A0). Under A2, in case of a widespread taxon occurring together with another sympatric taxon in one of its areas, Nelson and Ladiges (1991a,c) recommend to remove the sympatric occurrence of the widespread taxon and thereby to deal with redundancy *a priori*. This is similar to the procedure followed in CA and RTA to deal with such combinations of widespread and sympatric taxa *a priori* under A2.

After a matrix of three area statements is obtained under one of the assumptions, parsimony analysis, with an all-zero outgroup, is applied to derive area cladograms.

For the derivation of a matrix with three area statements, in this study the program TAS (Nelson and Ladiges, 1991c) is used. The matrices are analysed with Hennig86 to find most parsimonious area cladograms.

ASSESSMENT OF REQUIREMENTS I AND II IN IMPLEMENTATIONS OF METHODS

Theoretical and empirical data sets

The theoretical data sets in this study consist of:

- all possible topologies for taxon cladograms with three (1 topology), four (2 topologies) and five (3 topologies) taxa,
- the distributions of these taxa over their areas (none of the taxa of the different taxon cladograms overlap in their distribution); and
- all possible arrangements, over these topologies, of the presence of a single widespread taxon, two sympatric taxa and a combination of one widespread taxon and two sympatric taxa (see appendix ii).

The theoretical data sets are used to evaluate the performance of the implementations of the methods with respect to meeting requirement I (inclusive solution sets derived under A0, A1 and A2).

The empirical data sets in this study consist of:

- taxon cladograms and the distribution of two monophyletic poeciliid fish genera *Heterandria* and *Xiphophorus* (Rosen, 1978) (see appendix iii) occurring in overlapping areas.
- taxon cladograms and the distribution of two monophyletic genera Cyttaria (Crisci, 1988) and Eriococcus/Madarococcus (Humphries *et al*., 1986) (see appendix iv) occurring in the same areas.

The empirical data sets are used to evaluate the performance of the methods with respect to meeting requirement I. By virtue of the overlap in the distribution of the taxa of the different genera, the empirical data sets allow as well (contrary to the theoretical data sets) to evaluate the methods for their meeting of requirement II (obtaining general area cladograms under a single assumption).

CAFCA (Component Compatibility Analysis)

Theoretical data sets for single groups

For the single group data sets containing widespread taxa all solution sets show inclusion under A0, A1, and A2 when obtained with CAFCA (Zandee, 1999). Only occasionally the number of area cladograms is larger under A1 and A2, but in most cases the same area cladograms are derived under each assumption.

Under A0, a widespread taxon is interpreted as a “synapomorphy” of the areas in which it occurs (Zandee and Roos, 1987; Wiley, 1988a). Thus, no character conflict is induced and therefore dealing with widespread taxa under A0 occurs most parsimoniously (Fig. 3). Under A1 and A2 only the data columns (*i.e.* the A0 data matrix) are used to establish cladogram length. Since cladograms derived (for widespread taxa) under A0 already represent the most parsimonious solutions, new cladograms derived under A1 or A2 can never be preferred over those derived under A0. It can happen that for taxon-area cladograms with more than one area at a terminal node (widespread taxa), additional cladograms are derived under A1 or A2 that are equally parsimonious to the cladograms derived for widespread taxa under A0. However, in most cases of data sets with widespread taxa in this study, no additional equally parsimonious cladograms are derived under A1 or A2 and the same area cladogram is derived under A0, A1 and A2.

CAFCA makes no special provisions for sympatric taxa, but deals with them "as is" and obtains the same area cladograms under all three assumptions, thus fulfilling requirement I.

Only some data sets containing combinations of a widespread taxon and sympatric taxa in a single monophyletic group generate area cladograms under A1 and A2 that are more parsimonious than those that are found under A0 (Fig. 4). As a result requirement I is broken in these cases (table 1).

One can use the number of steps for the cladogram derived under A0 as an upper limit for cladogram selection under A1 or A2. By applying this *ad hoc* procedure, CAFCA not only finds the most parsimonious area cladograms under A1 or A2 but also the less parsimonious area cladograms which already have been derived under A0. As a result, inclusion is obtained (requirement I).

Heterandria/*Xiphophorus*

Analysis with CAFCA of the data matrix for the poeciliid fishes results in one cladogram of 30 steps derived under A0. Under A1 the same cladogram is found (Fig. 5a). Under A2 however, new area cladograms are found that need 29 steps, none of them present under A0 and A1 (Fig. 5b). As a result, the set of general area cladograms derived for *Heterandria* and *Xiphophorus* under A2 break requirement I. However, we can use the number of steps for the cladogram derived under A0/A1 (30 steps) as an upper limit for cladogram selection under A2. Consequently, the solution set under A2 becomes larger and inclusion is obtained (requirement I).

The general area cladograms are obtained from a data matrix consisting of a combination of the data of *Heterandria* and *Xiphophorus*. Since this data matrix is derived per assumption for both genera together, general area cladograms are obtained under a single assumption (requirement II).

Cyttaria/*Eriococcus*/*Madarococcus*

Under A0 and A1 the solution sets derived with CAFCA from the *Cyttaria* and *Eriococcus*/*Madarococcus* (Crisci, 1991) data matrix are identical and contain one cladogram that needs 16 steps (Fig. 6a). Because a more parsimonious cladogram (15 steps) is found under A2 (Fig. 6b), inclusion of solution sets with general area cladograms derived under A0/A1 and A2 is not obtained.

When under A2 cladograms are selected that need 16 steps or less, *i.e.* by using the upper limit indicated by the A0 result, besides the single most parsimonious cladogram (15 steps) another 4 cladograms (16 steps) are selected (Fig. 6c). One of these cladograms is equal to the one already found under A0 and A1 and therefore by using the number of steps for the cladogram derived under A0/A1 as an upper limit, inclusion of solution sets is obtained (requirement I).

The general area cladograms are obtained from a data matrix consisting of a combination of the data of *Cyttaria* and *Eriococcus*/*Madarococcus*. Since this data matrix is derived per assumption for both genera together, general area cladograms are obtained under a single assumption (requirement II).

Brooks Parsimony Analysis

Theoretical data sets for single groups

With BPA all solution sets derived from data sets of a single group of taxa with a widespread taxon show inclusion. As opposite to CAFCA, all columns in the data matrix are used with BPA to count the number of steps under A1 or A2. New columns result in new clades for the areas and frequently result in an increasing number of equally (most) parsimonious cladograms.

Like with CAFCA, redundancy for areas (caused by sympatric taxa) is dealt with *a posteriori* in BPA. As a result, the same area cladograms are obtained under all three assumptions and thus requirement I is met.

For the single group data sets with combinations of a widespread taxon and sympatric taxa, non-inclusive solution sets are produced when under either A1 or A2 more parsimonious cladograms are obtained than the cladograms derived under A0 or A1 (with the data set optimised on them; Fig. 7). Non-inclusive results for the analysis of single group data sets are represented in Table 2.

Inclusive solution sets can be obtained by optimising the data matrix, derived under A1, on the area cladograms obtained under A0. By using the highest number of steps for this optimisation as an upper limit for cladogram selection from the data matrix derived under A1, area cladograms obtained under A0 are also found in the set of area cladograms derived under A1. By repeating this procedure with the data matrix derived under A2, area cladograms already obtained under A0 and A1 are obtained under A2 as well and thus inclusion is obtained (requirement I).

Heterandria/Xiphophorus

Parsimony analysis of the data matrix derived for *Heterandria/Xiphophorus* results under A0 in a solution set of four general area cladograms that need 35 steps (Fig. 8a). Under A1 (columns for widespread taxa derived with CAFCA) three general area cladograms of 47 steps are derived (Fig. 8b). Two of the general area cladograms derived under A1 are derived under A0 as well. However, because the number of general area cladograms derived under A1 is smaller than the number of general area cladograms derived under A0, inclusion is not obtained. Under A2, 11 general area cladograms of 157 steps are derived of which two are already derived under A0 (*i.e.* partly overlap; Fig. 8c).

The way to obtain inclusion with this data set is by optimising the A1 data matrix on the set of general area cladograms derived under A0 and using the maximum number of steps as an upper limit for cladogram selection under A1. When this is done, optimisation of the A1 matrix on one of the general area cladograms derived under A0 results in a maximum number of 49 steps. Using that as an upper limit results in 222 cladograms under A1 with the four derived under A0 included.

When the data matrix for A2 is optimised on the 222 general area cladograms derived under A1, it appears that the maximum number of steps is 167. Using this as an upper limit results in an explosion of the number of general area cladograms (11698) under A2. Analysis of this data set shows that, depending on the number of extra steps necessary to include all area cladograms derived under a stricter assumption, solution sets can explode in the number of possible area cladograms they encompass.

Just like with CAFCA, the general area cladograms are obtained in BPA from a data matrix consisting of a combination of the data of *Heterandria* and *Xiphophorus*.

The data matrices are derived per assumption for both genera together and therefore general area cladograms are obtained under a single assumption (requirement II).

Cyttaria/Eriococcus/Madarococcus

With the datamatrix derived for *Cyttaria* and *Eriococcus/Madarococcus* (Crisci, 1991), the same general area cladogram was derived under A0, A1, and A2 with BPA (Fig. 9). The solution sets derived under the three assumptions are equal and so inclusion is obtained (requirement I).

The general area cladogram for *Cyttaria* and *Eriococcus/Madarococcus* is obtained under a single assumption (requirement II) because the data matrices for BPA are derived per assumption for both genera together.

Component 1.5 (Component Analysis)

Theoretical data sets for single groups

With respect to widespread taxa, CA (as implemented in Component 1.5) results in inclusive solution sets for the single group data sets under A0, A1, and A2. Because the number of components recognised increases from A0 to A1 to A2, the number of area cladograms derived under these assumptions increases as well.

Solutions derived under A0, A1, or A2 with Component 1.5 can differ because, as we showed before, sympatric taxa are dealt with in different ways under these assumptions. When sympatric taxa are dealt with by Component 1.5, often more or other area cladograms are obtained under A1 than the area cladograms that are obtained by *a priori* removal of areas under A2. As a result, solution sets obtained under A2 do not include all (or none) of the area cladograms already obtained under A1 and therefore inclusion of solution sets is then broken.

Inclusion of solution sets is most frequently broken when combinations of a widespread taxon and sympatric taxa are dealt with *a priori*. This is caused by the specific solution (suggested by Nelson and Platnick, 1981 and Page, 1988) of combinations of a widespread taxon and sympatric taxa under A2. *A priori* removal of one of the redundant areas (*i.e*. pruning of the widespread taxon in the redundant area) often results in a limitation of the number of solutions derived under A2. As a result, solution sets obtained under A2 are smaller than solution sets obtained under A1 and inclusion is broken. Non-inclusive results for the analysis of single group data sets are represented in Table 3.

Heterandria/Xiphophorus

For *Heterandria*, the same area cladogram is derived under A0, A1, and A2 (Fig. 10a) with Component 1.5 because no widespread or sympatric taxa are present in the *Heterandria* data set. Inclusion is thus not broken with this data set. Area C does not inhabit one of the taxa of this group and therefore is not present in the area cladogram.

For the *Xiphophorus* group, a single area cladogram is obtained under A0 (Fig. 10b) with Component 1.5. Under A1, the widespread taxa T5 and T6 are split in different occurrences with a trichotomy at the basis. Resolving these trichotomies results in nine area cladograms under A1 with Component 1.5 (the strict consensus of these cladograms is represented in Fig. 10c). Under A2 each occurrence of a widespread taxon is removed from the cladogram and reconnected in such a way that it floats over the whole cladogram. In this way, for *Xiphophorus* 1165 area cladograms are obtained with Component 1.5. Comparison of the solution sets derived under A0, A1, and A2 shows that inclusion is obtained (requirement I). The obtained area cladograms are without area G since no taxa of *Xiphophorus* are found within this area.

Comparing the sets of area cladograms derived for both genera under A0 or A1 reveals no general area cladograms since the intersections are empty. Only under A2, three general area cladograms are found in the intersection (the strict consensus of these cladograms is represented in Fig. 10d). These three general area cladograms agree on the position of the areas present in the *Heterandria* data set. Only area C (no *Heterandria* taxa present) is found on three different positions based on the

Xiphophorus data. The position of area G in the general area cladograms is derived from its position in the area cladogram derived for *Heterandria*. Since this is only a single area cladogram, the position of area G is stable in the three general area cladograms obtained under A2.

Because with Component 1.5 general area cladograms are only found by comparison of sets of area cladograms obtained for *Heterandria* under A2 and obtained for *Xiphophorus* under A2, general area cladograms are obtained under a single assumption (requirement II).

Cyttaria/Eriococcus/Madarococcus

For *Cyttaria* (Crisci, 1991), a single area cladogram is derived (Fig. 11a) with Component 1.5 under A0. This area cladogram is the same as already derived with CAFCA and BPA under A0. Under A1 three area cladograms are obtained (Fig. 11b) and under A2 only two of these area cladograms are derived (Fig. 11c). As a consequence, requirement I is broken for the data on *Cyttaria*.

The data on *Eriococcus-Madarococcus* (Crisci, 1991) do not cause any problems with respect to inclusion of solution sets. Under A0, a single area cladogram is derived (Fig. 12a). Besides this area cladogram, under A1 two additional area cladograms are derived (Fig. 12b). Under A2, eleven area cladograms are derived (Fig. 12c), including the ones derived under A0 and A1.

With Component 1.5, no general area cladograms are found in the intersection of the sets of area cladograms derived under A0 for both data sets. Under A1 in the intersection of the sets of area cladograms, three general area cladograms are found. Comparison of the sets of area cladograms derived under A2 results in two general area cladograms (also derived under A1) in common. Since the number of general area cladograms derived under A2 is smaller than the number of general area cladograms derived under A1, inclusion of sets of general area cladograms under A0, A1, and A2 is broken (requirement I).

For *Cyttaria* and *Eriococcus-Madarococcus* general area cladograms are obtained by comparison of sets of area cladograms obtained for each group under A1 or A2. By only comparing the sets of area cladograms derived for each group under the same assumption, requirement II is met.

Component 2.0 (Reconciled Tree Analysis)

Theoretical data sets for single groups

RTA with Component 2.0 results in inclusive solution sets derived for single group data sets under A0, A1, and A2 when dealing with widespread taxa. Sympatric taxa can only be dealt with by the derivation of reconciled trees. Area cladograms that need the least independent losses and duplications for reconciliation with the taxon-area cladogram are selected. Sympatric taxa cannot be dealt with under A0 (*i.e. a posteriori*) with this computer program and under A2 sympatric taxa are dealt with in the same way as under A1 (*i.e.* via tree reconciliation). As a result the same area cladograms are obtained with Component 2.0 when sympatric taxa are dealt with under A0, A1 or A2 and inclusion is not broken.

With Component 2.0, inclusion is broken when combinations of a widespread taxon and sympatric taxa are dealt with under A2 by *a priori* removal of areas in favour of endemics. As a result of this operation, often, smaller solution sets are obtained under A2 than under A0/A1 and consequently inclusion is broken. We show non-inclusive results for the analysis of single group data sets in Table 4.

Heterandria/Xiphophorus

With Component 2.0, the data sets from *Heterandria* and *Xiphophorus* are used together for the selection of general area cladograms. Under A0 one general area cladogram (via a reconciliation that need 12 losses and 3 duplications) is obtained for *Heterandria* and *Xiphophorus* (Fig. 13a). Under A1 also a single general area cladogram (via a reconciliation that needs 9 losses and 2 duplications) is obtained that is different from the one derived under A0 (Fig. 13b). Analysis of the data under A2 obtains three general area cladograms (Fig. 13c) (via reconciliations that need 7 losses and 2 duplications). These three general area cladograms are similar to the three general area cladograms that are found in the intersection of solution sets derived for *Heterandria* and *Xiphophorus* with Component 1.5 under A2. The general area cladograms derived with Component 2.0 under A0, A1 and A2 are all different and therefore inclusion is broken (requirement I).

With Component 2.0, general area cladograms are obtained via reconciliation of area cladograms with the taxon-area cladograms of both *Heterandria* and *Xiphophorus*. The area cladograms that need (in total) a minimum amount of independent losses and duplications for reconciliation are selected as general area cladograms. Because widespread or sympatric taxa in the data of both poeciliid fish genera are dealt with in the same way, general area cladograms are derived under the same assumption (requirement II).

Cyttaria/*Eriococcus*/*Madarococcus*

The data sets of *Cyttaria* and *Eriococcus*/*Madarococcus* (Crisci, 1991) are used together for the selection of general area cladograms under A0, A1 and A2. Under A0 and A1 the same general area cladogram is derived (Fig. 14a) (via a reconciliation that needs 6 losses and 5 duplications). Under A2, the data sets are manually edited by removing areas in the distribution of *Cyttaria* and *Eriococcus*/*Madarococcus* in favour of endemics. Analysis of the data sets for both groups under A2 reveals three general area cladograms (Fig. 14b) (via reconciliations that need 6 losses and 4 duplications). The area cladogram derived under A0 and A1 is also found in the solution set derived under A2 and therefore inclusion is not broken.

Just like with the data from *Heterandria*/*Xiphophorus*, general area cladograms for *Cyttaria* and *Eriococcus*/*Madarococcus* are obtained via reconciliations between area cladograms and taxon-area cladograms from both groups. Widespread taxa in

Cyttaria and *Eriococcus*/*Madarococcus* are dealt with in the same way per analysis (*i.e.* under a particular assumption) and therefore requirement II is met.

TAS (Three Area Statement Analysis)

Theoretical data sets for single groups

In order to solve the problem of non-inclusive solution sets, Nelson and Ladiges (1991a,b) described three area statements derived from internal nodes in a taxon-area cladogram with multiple areas at one terminal node (resulting from widespread taxa) or the same (redundant) area at different terminal nodes (resulting from sympatric taxa). For dealing with widespread taxa, the same nodes are considered under A1 and A2. As a result identical solution sets are derived. Dealing with widespread taxa under A1/A2 mostly results in a polytomy for the areas of the widespread taxon together with their sister-areas (Fig. 15a). Because under A0 a node (*i.e.* component) is recognised for the widespread taxon, only a single area cladogram with the areas of the widespread taxon as sister-areas is obtained (Fig. 15b). This area cladogram can be derived by solving the trichotomy of one of the area cladograms derived under A1/A2, but is not recognised under A1/A2. As a result, the solution set derived under A0 is not included in the solution set derived under A1/A2 (Table 5).

Redundancy is dealt with “as is", *i.e.* by the derivation of three area statements under A0 and A1. Under A2, Nelson and Ladiges (1991a,c) recommend to remove redundant areas *a priori* in favour of endemics. These areas are unable to float over the area cladogram and often less area cladograms are derived under A2 than under A0. As a result, requirement I is violated (Table 5).

Heterandria/*Xiphophorus*

Under both A0 and A1/A2 the same general area cladogram is derived for the data sets derived for the poeciliid fishes (Fig. 16). No *a priori* removal of areas from one of the taxon-area cladograms is necessary because no combinations of widespread taxa with sympatric taxa are present in the data sets. Therefore, both A1 and A2 use the same nodes for derivation of the same matrices with three area statements from which the same solution sets are derived. For these data sets, inclusion of solution sets is not broken with TAS.

Just like with CAFCA and BPA, with TAS the general area cladograms for *Heterandria* and *Xiphophorus* are obtained by combining the data of both genera in one single matrix (but now consisting of three area statements). This is done under each assumption and therefore requirement II is met.

Cyttaria/*Eriococcus*/*Madarococcus*

For the data matrix derived for *Cyttaria* and *Eriococcus-Madarococcus*, 73 three area statements are derived with TAS under A0. Analysis of this data matrix with Hennig86 results in a single most parsimonious general area cladogram under A0 (Fig. 17a). Under A1, 55 and under A2, 52 statements are derived. Under both assumptions, analysis of the data matrix results in two most parsimonious (general) area cladograms (Fig. 17b) that are both different from the cladogram which is derived under A0. Therefore, recognition of three area statements under A0, A1 and A2 for this data set results in violation of requirement I.

General area cladograms for *Cyttaria and Eriococcus-Madarococcus* are obtained from a single data matrix. This matrix either is derived under A0, A1 or A2. Because the same assumption is applied in the derivation of this data matrix (with three area statements) for both genera, general area cladograms are obtained under a single assumption (requirement II).

CONCLUSIONS

In this study we assessed the precise implementation of A0, A1 and A2 in the programs of five different methods that are used in vicariance biogeography for inference of area cladograms. By construction of area cladograms for theoretical data sets of single groups of taxa and empirical data sets of *Heterandria*/*Xiphophorus* (Rosen, 1978) and *Cyttaria*/*Eriococcus*-*Madarococcus* (Crisci, 1991), we assessed the performance of the methods for two requirements:

I. Solution sets under A0, A1 and A2 should be inclusive.
II. Solution sets for two or more groups of taxa should be compared under the same assumption.

From our assessment we conclude that requirement II is never violated since the programs of all methods appear to obtain general area cladograms under a single assumption.

The programs Component 1.5, Component 2.0 and TAS of the *a priori* methods Component Analysis, Reconciled Tree Analysis and Three Area Statement Analysis respectively, implement A0, A1 and A2 in such way that requirement I is bound to be violated. Each of these *a priori* methods adds or prunes taxa and taxon relationships to the original taxon cladogram from which the initial hypothesis of area relationships is obtained. Each of these *a priori* methods prune under A2 taxa and taxon relationships from the taxon cladogram in such way that the initial historical relationships of taxa under A0 may become excluded unavoidably resulting in non-inclusion.

In this study we show that violation of requirement I arises when implementations of CA, RTA or TAS deal with sympatric taxa or combinations of widespread and sympatric taxa *a priori*. In Van Veller *et al.* (1999) we recommend that in order to explain all data, one should deal with sympatric taxa) only “as is". With this we mean that no *a priori* steps should be taken to deal with sympatric taxa and that absence or multiple presence of taxa in the area cladogram, caused by extinction or dispersal, should be accounted for only *a posteriori*. For Component 1.5 this can be achieved by dealing with sympatric taxa under A0. For TAS, we showed in this paper that sympatric taxa are dealt with "as is" under A0 and A1. By not *a priori* removing sympatric occurrences of widespread taxa (contrary to the recommendation by Nelson and Ladiges, 1991a,c), sympatric taxa can be dealt with "as is" with TAS as well. For Component 2.0 we also recommend not to deal with sympatric taxa *a priori* by reducing widespread taxa in favour of endemics. However, for an implementation of RTA it is not possible to deal with sympatric taxa "as is" since no data matrices are constructed, but trees are compared and reconciled used to deal with sympatric taxa *a priori*. In a future paper we will suggest additional procedures for dealing with widespread and sympatric taxa with Component 1.5, TAS and Component 2.0 in such way that violation of requirement I is remedied.

In this paper we also show that implementations of *a posteriori* methods are neither free from violation of requirement I. Non-inclusion of solution sets may emerge when the data contain incongruencies. We found that this can be remedied by applying an upper limit for the number of steps (for selection of area cladograms) under a less strict assumption (*e.g.* A1). This upper limit is equal to the number of steps of the most parsimonious area cladogram(s) found under a more strict assumption (*e.g.* A0). Application of these *ad hoc* procedures results in inclusive solution sets obtained with CAFCA or via BPA. For BPA, however, we showed that application of an upper limit for selection of area cladograms under a less strict assumption can result in an explosion of solution sets. A more fundamental solution to the violation of requirement I with the implementations of *a posteriori* methods might be sought in a quality assessment of the nodes of the area cladograms ahead of the derivation of solution sets. However, this remains to be examined in a future study.

ACKNOWLEDGEMENTS

We thank D.R. Brooks for his comments on a previous version of the manuscript. This work was supported by the Life Sciences Foundation (SLW), which is subsidised by the Netherlands Organization for Scientific Research (NOW); grant 805-33.193 to MvV.

REFERENCES

Bron, C. and Kerbosch, J. (1980). Algorithm 457: Finding all cliques of an undirected graph. *In* "Collected algorithms from ACM". Volume II: Algorithms 221-492. pp. 457-P1-P3. Association for Computing Machinery. New York.

Brooks, D.R. (1981). Hennig's parasitological method: A proposed solution. *Syst. Zool.* **30**, 229-249.

Brooks, D.R. (1990). Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Syst. Zool.* **39**, 14-30.

Crisci, J.V., Gamundi, I.J. and Cabello, M.N. (1988). A cladistic analysis of the genus *Cyttaria* (Fungi-Ascomycotina). *Cladistics* **4**, 279-290.

Crisci, J.V., Cigliano, M.M. and Morrone, J.J. (1991). Historical biogeography of southern South America. *Syst. Zool.* **40**, 152-171.

Enghoff, H. (1996). Widespread taxa, sympatry, dispersal, and an algorithm for resolved area cladograms. *Cladistics* **12**, 349-364.

Enghoff, H. (1998). Widespread taxa and Component 2.0. *Cladistics* **14**, 383-386.

Farris, J.S. (1988). Hennig86 vs 1.5. Distributed by the author. Port Jefferson Station. New York.

Hendy, M.D. and Penny, D. (1982). Branch and bound algorithms to determine minimal evolutionary trees. *Math. Biosc.* **59**, 277-290.

Hovenkamp, P. (1997). Vicariance events, not areas, should be used in biogeographical analysis. *Cladistics* **13**, 67-79.

Humphries, C.J., Cox, J.M. and Nielsen, E.S (1986). Nothofagus and its parasites: a cladistic approach to coevolution. *In* "Coevolution and systematics". (A.R. Stone and D.L. Hawksworth, Eds), pp. 55-76. Clarendon Press. Oxford.

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. (1979). Cladistic analysis and synthesis: principles and definitions, with a historical note on Adanson's Familles des Plantes (1763-1764). *Syst. Zool.* **28**, 1-21.

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. (1991c). TAS: Three Area Statements. Program and user's manual. Published by authors. Melbourne, New York.

Nelson, G. and Platnick, N.I. (1981). "Systematics and biogeography; cladistics and vicariance". Columbia University Press. New York. 567 pp.

Nelson, G. and Platnick, N.I. (1991). Three-taxon statements: a more precise use of parsimony? *Cladistics* **7**, 351-366.

O'Grady, R.T. and Deets, G.B. (1987). Coding multistate characters, with special reference to the use of parasites as characters of their hosts. *Syst. Zool.* **28**, 1-21.

Page, R.D.M. (1988). Quantitative cladistic biogeography: constructing and comparing area cladograms. *Syst. Zool.* **37**, 254-270.

Page, R.D.M. (1990). Component 1.5. Program and user's manual. University of Auckland. Auckland.

Page, R.D.M. (1993a). Component 2.0 for Windows. Program and user's manual. Natural History Museum. London.

Page, R.D.M. (1993b). Genes, organisms, and areas: the problem of multiple lineages. *Syst. Biol.* **42**, 77-84.

Page, R.D.M. (1994). Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* **43**, 58-77.

Platnick, N.I. (1981). Widespread taxa and biogeographic congruence. *In* "Proceedings of the First Meeting of the Willi Hennig Society". (V.B. Funk and D.R. Brooks, Eds), pp. 223-227. New York Botanical Garden. Bronx, New York.

Rosen, D.E. (1978). Vicariant patterns and historical explanation in biogeography. *Syst. Zool.* **27**, 159-188.

Swofford, D.L. (1990). PAUP vs. 3.11: Phylogenetic Analysis Using Parsimony. Illinois Natural History Survey. Champaign.

Van Veller, M.G.P., Zandee, M. and Kornet, D.J. (1999). Two requirements for obtaining valid common patterns under assumptions zero, 1 and 2 in vicariance biogeography. *Cladistics* **15**, 393-406.

Wiley, E.O. (1988a). Parsimony analysis and vicariance biogeography. *Syst. Zool.* **37**, 271-290.

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

Zandee, M. (1999). CAFCA vs. 1.5j: a collection of APL functions for cladistic analysis. Program and user's manual. Institute of Evolutionary and Ecological Sciences, Leiden University. Leiden.

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

LIST OF FIGURES, APPENDICES AND TABLES

FIG. 1. Adjustment of area by node data matrix for dealing with a widespread taxon and derivation of area cladograms under A1 with CA.

FIG. 2. Adjustment of area by node data matrix for dealing with sympatric taxa and derivation of area cladograms under A1 with CA.

FIG. 3. Area cladograms derived under A0 and A1 with CAFCA for a cladogram with a widespread taxon (in areas C and D).

■ = non-homoplaseous □ = homoplaseous

FIG. 4. Derivation of a more parsimonious area cladogram under A2 with CAFCA for a cladogram with a combination of widespread and sympatric taxa.

■ **m =** multistate character ■ = non-homoplaseous □ = homoplaseous

FIG. 5. General area cladograms derived under A0, A1 and A2 with CAFCA for *Heterandria*/*Xiphophorus* (a: general area cladograms under A0 and A1, b: genera area cladograms under A2).

FIG. 6. General area cladograms derived under A0, A1 and A2 with CAFCA for *Cyttaria*/*Eriococcus*/*Madarococcus* (a: general area cladograms under A0 and A1, b: genera area cladograms under A2, c: additional general area cladograms under A2 when applying an upper limit).

FIG. 7. Non-inclusion with BPA. Derivation of other area cladograms, under A1, that are more parsimonious than the area cladograms derived under A0.

FIG. 8. General area cladograms derived with BPA under A0, A1 and A2 for *Heterandria*/*Xiphophorus* (a: general area cladograms under A0, b: general area cladograms under A1, c: general area cladograms under A2).

FIG. 9. General area cladogram derived with BPA under A0, A1 and A2 for *Cyttaria*/*Eriococcus*/ *Madarococcus*.

FIG. 10. Area cladograms derived with Component 1.5 under A0, A1 and A2 for *Heterandria* and *Xiphophorus* (a: area cladogram for *Heterandria* under A0, A1, and A2, b: area cladogram for *Xiphophorus* under A0, c: strict consensus of the nine area cladograms for *Xiphophorus* under A1, d: strict consensus of the three general area cladograms for *Heterandria*/*Xiphophorus* in the intersection of solution sets under A2).

FIG. 11. Area cladograms derived with Component 1.5 under A0, A1 and A2 for *Cyttaria* (a: area cladogram under A0, b: area cladograms under A1, c: area cladograms under A2).

FIG. 12. Area cladogram derived with Component 1.5 under A0, A1 and A2 for *Eriococcus*/*Madarococcus* (a: area cladogram under A0, b: area cladograms under A1, c: area cladograms under A2).

FIG. 13. General area cladograms derived with Component 2.0 under A0, A1 and A2 for *Heterandria*/*Xiphophorus* (a: general area cladogram under A0, b: general area cladogram under A1, c: general area cladograms under A2).

FIG 14. General area cladograms derived with Component 2.0 under A0, A1 and A2 for *Cyttaria* and *Eriococcus*/*Madarococcus*. (a: general area cladogram under A0 and A1, b: general area cladograms under A2).

FIG. 15. Solving widespread taxa with TAS under A0 and A1/A2 (a: area cladograms under A1/A2, b: area cladogram under A0).

FIG. 16. General area cladogram derived with TAS under A0 and A1/A2 for *Heterandria*/*Xiphophorus*.

FIG. 17. General area cladograms derived with TAS under A0 and A1/A2 for *Cyttaria* and *Eriococcus*/*Madarococcus*. (a: general area cladogram under A0, b: general area cladograms under A1/A2).

Appendix i.

Steps when applying implementations of CCA, BPA, CA, RTA and TAS to obtain area cladograms from cladogenetic and distribution data under different assumptions.

Appendix ii.

Taxon cladograms for up to five taxa in a single group with one widespread taxon (wid), two sympatric taxa (red) and a combination of one widespread taxon and two sympatric taxa (rew).

Appendix iii.

Taxon cladograms for *Heterandria* and *Xiphophorus* (Rosen, 1978) with area of distribution superimposed. Areas D and E are combined in accordance to Platnick (1981) and Page (1988).

Appendix iv.

Taxon cladograms for *Cyttaria* (Crisci *et al*., 1988) and *Eriococcus/Madarococcus* (Humphries *et al*., 1986) with areas of distribution superimposed. Both cladograms also represented in Crisci (1991).

TABLE 1

Non-inclusion of solution sets derived under A0, A1 and A2 with CAFCA for a single group with 3 to 5 taxa (cladograms listed in appendix ii).

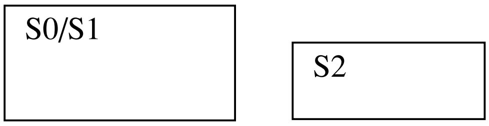
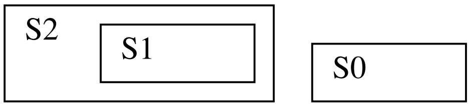
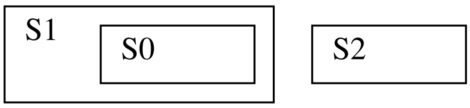
Cladogram	Solution sets under A0, A1, and A2
rew406, rew509, rew511, rew524	
rew515	
rew516	

TABLE 2

Non-inclusion of solution sets derived under A0, A1 and A2 with BPA for a single group with 3 to 5 taxa (cladograms listed in appendix ii).

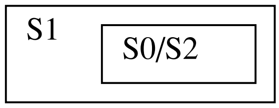
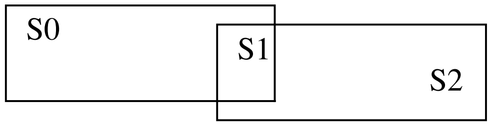
Cladogram	Solution sets under A0, A1, and A2
rew514	
rew515	

TABLE 3

Non-inclusion of solution sets derived under A0, A1 and A2 with Component 1.5 for a single group with 3 to 5 taxa (cladograms listed in appendix ii).

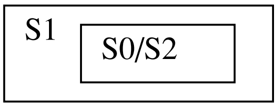
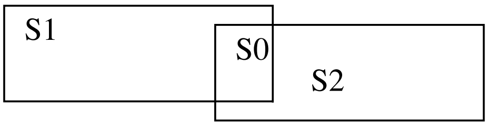
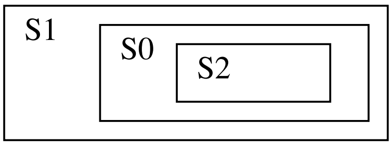
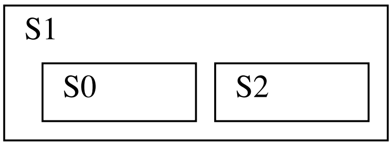
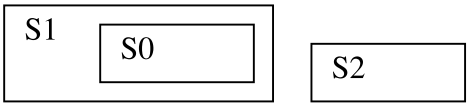
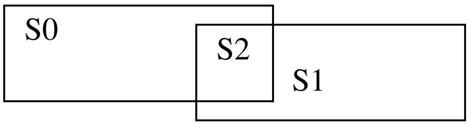
Cladogram	Solution sets under A0, A1, and A2
red406, red510, red515, rew406, rew509, rew524, rew525	
red512	
rew301, rew402, rew502, rew503, rew505, rew511, rew512, rew520	
rew302, rew401, rew403, rew409, rew501, rew504, rew506, rew516, rew517, rew519, rew523	
rew407, rew508, rew510, rew513, rew522	
rew515	

TABLE 4

Non-inclusion of solution sets derived under A0, A1 and A2 with Component 2.0 for a single group with 3 to 5 taxa (cladograms listed in appendix ii).

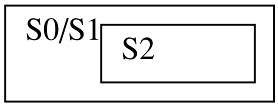
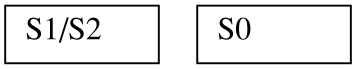
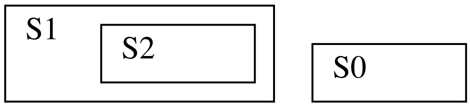
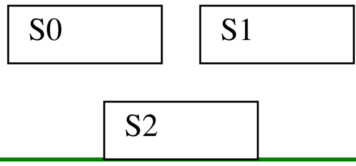
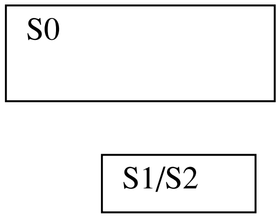
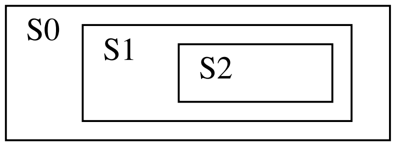
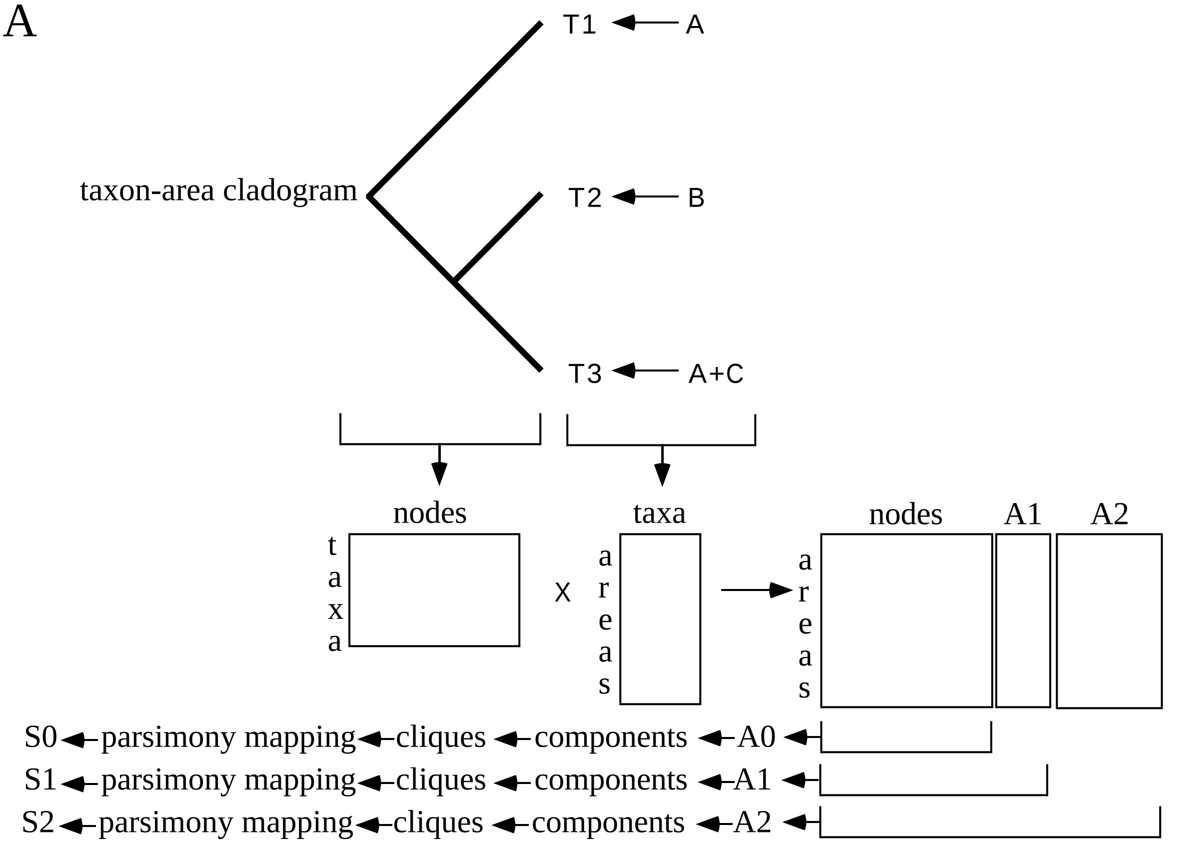
Cladogram	Solution sets under A0, A1 and A2
rew301, rew402, rew404, rew406, rew502, rew503, rew505, rew509, rew511, rew512, rew520, rew524	
rew302, rew403, rew506, rew519	
rew401, rew409, rew501, rew504, rew515, rew516, rew517, rew523	
rew407, rew508, rew510, rew513, rew514, rew522	

TABLE 5

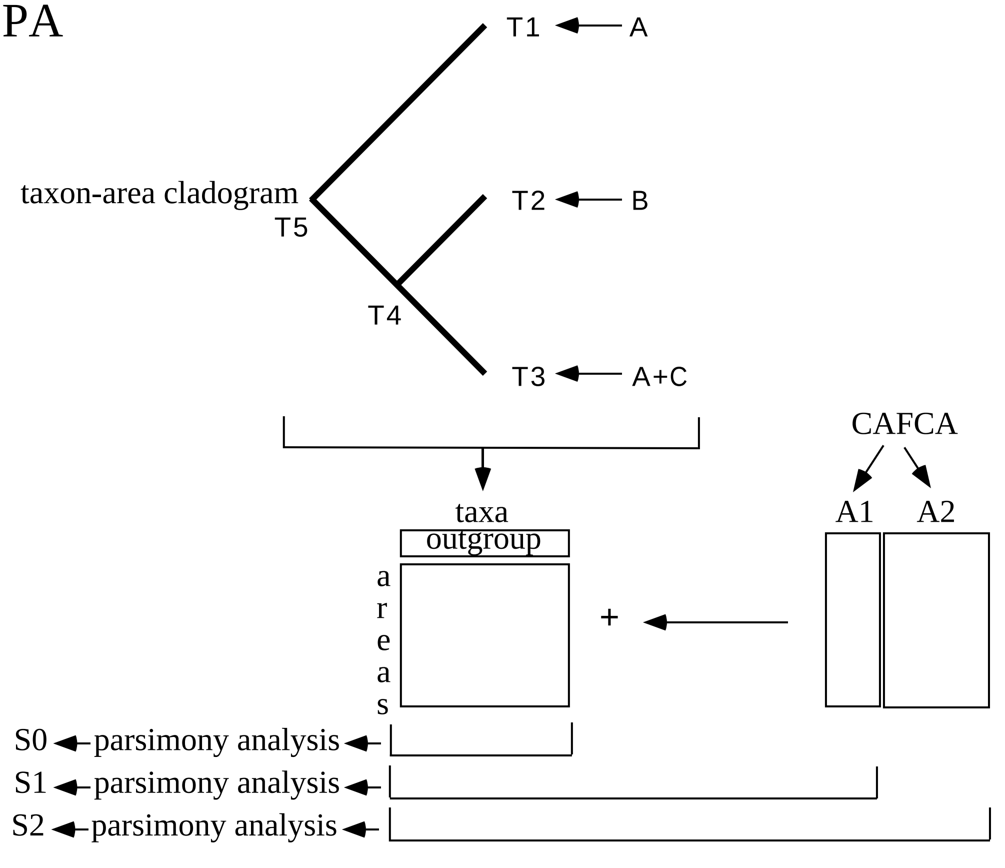
Non-inclusion of solution sets derived under A0, A1 and A2 with TAS for a single group with 3 to 5 taxa (cladograms listed in appendix ii).

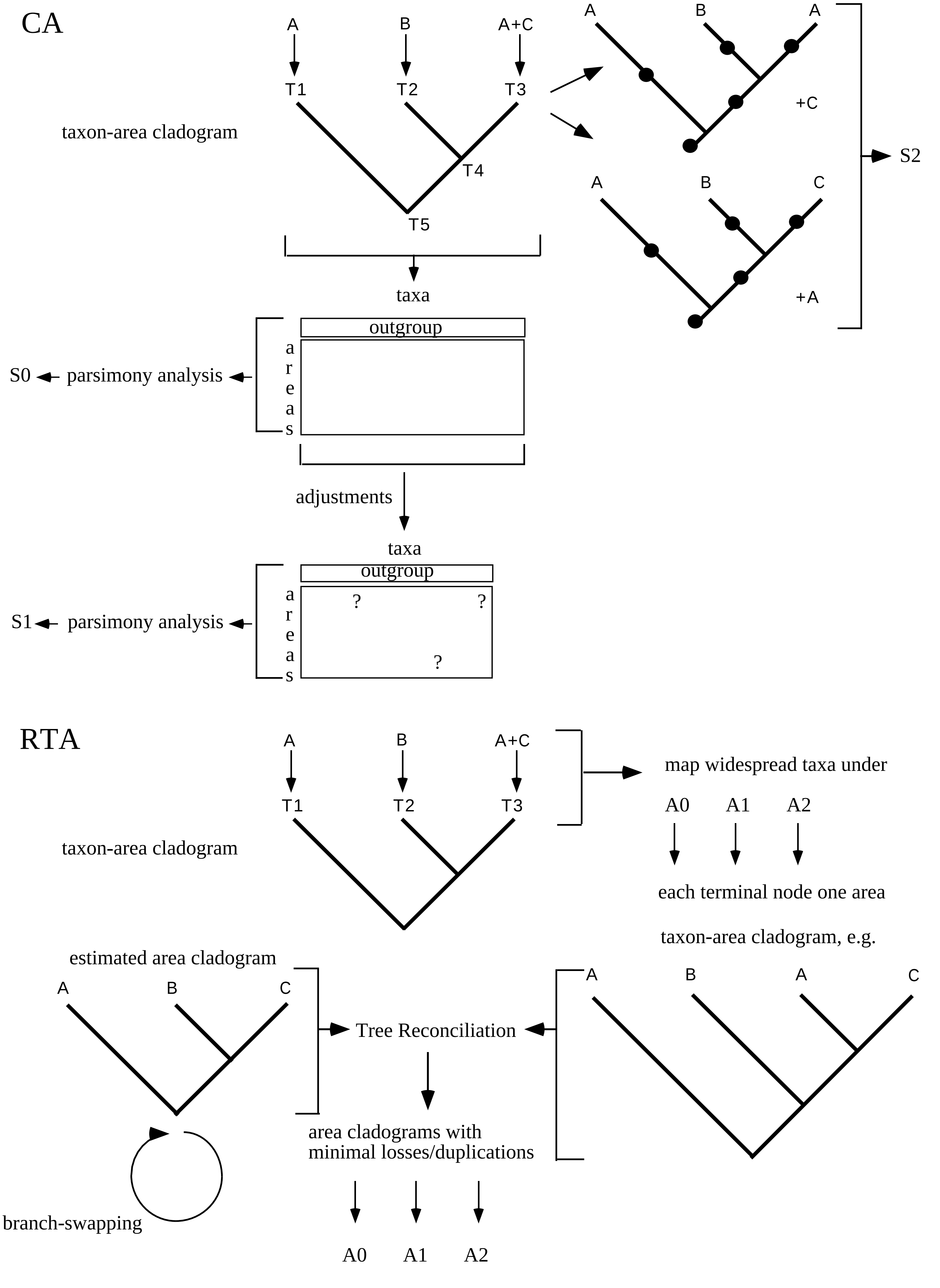
Cladogram	Solution sets under A0, A1 and A2
wid401, wid402, wid403, wid404, wid501, wid502, wid503, wid504, wid505, wid506, wid507, wid508, wid509	
rew301, rew302, rew401, rew402, rew403, rew404, rew407, rew409, rew501, rew502, rew503, rew504, rew505, rew506, rew508, rew510, rew513, rew514, rew516, rew517, rew519, rew520, rew522, rew523	
rew406, rew509, rew511, rew512, rew515, rew524	

CCA

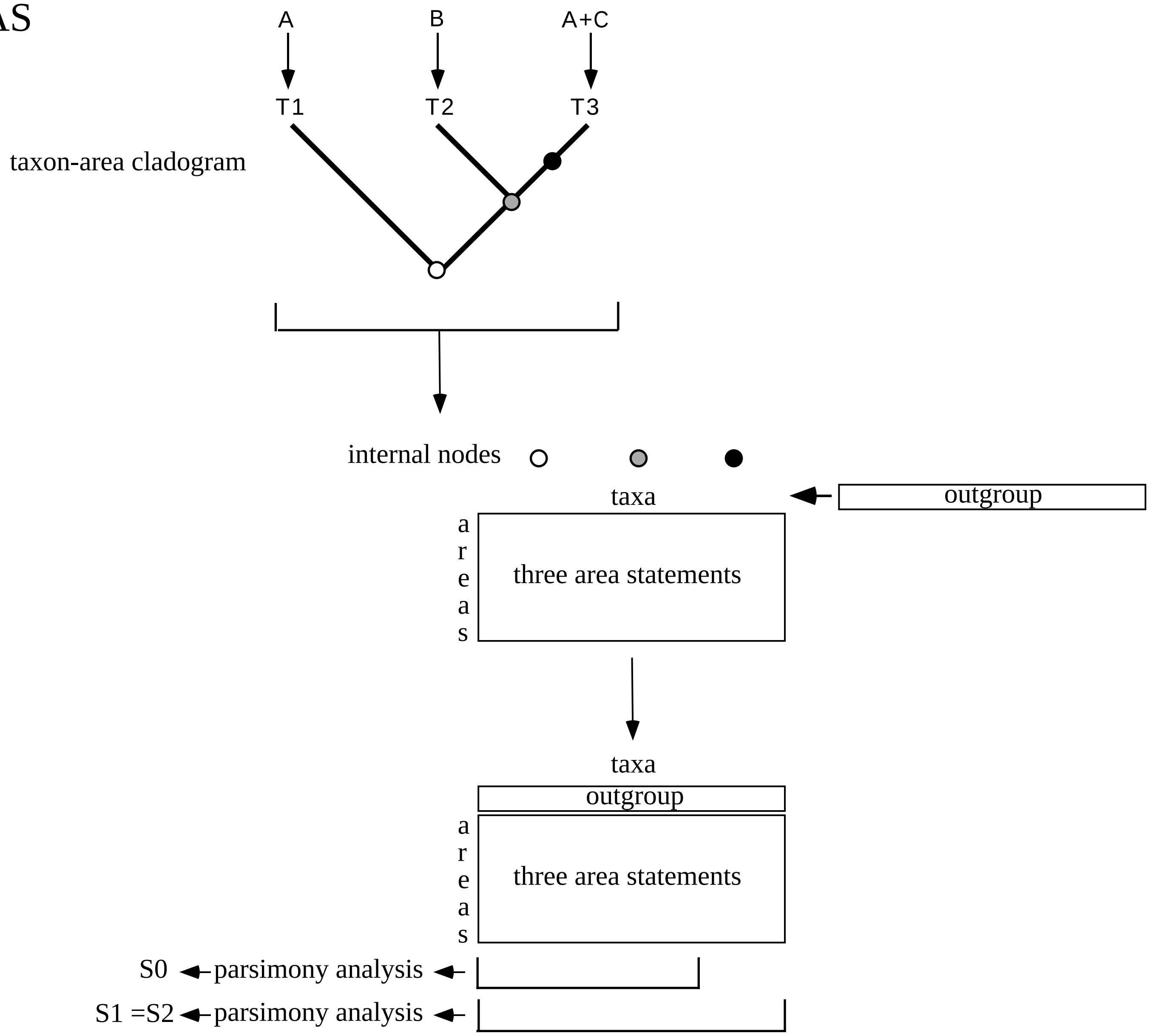

T1 A
T2 B
T3 A+C
taxon-area cladogram
nodes
taxa
X
areas
nodes
A1
A2
S0 parsimony mapping cliques components A0
S1 parsimony mapping cliques components A1
S2 parsimony mapping cliques components A2

BPA


T1 A
T2 B
T3 A+C
taxon-area cladogram
T5
T4
CAFCA
taxa
outgroup
areas
+
A1
A2
S0 parsimony analysis
S1 parsimony analysis
S2 parsimony analysis

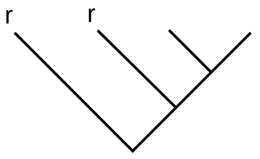

CA
A
B
A+C
T1
T2
T3
T4
T5
taxon-area cladogram
A
B
A
+C
A
B
C
+A
S2
taxa
outgroup
areas
S0
parsimony analysis
adjustments
taxa
outgroup
areas
?
?
?
S1
parsimony analysis
RTA
A
B
A+C
T1
T2
T3
taxon-area cladogram
map widespread taxa under
A0
A1
A2
each terminal node one area
taxon-area cladogram, e.g.
A
B
A
C
estimated area cladogram
A
B
C
Tree Reconciliation
area cladograms with
minimal losses/duplications
A0
A1
A2
branch-swapping

TAS

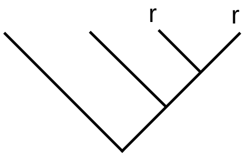

A
B
A+C
T1
T2
T3
taxon-area cladogram
internal nodes
taxa
outgroup
areas
three area statements
taxa
outgroup
areas
three area statements
S0
parsimony analysis
S1 =S2
parsimony analysis

SYMPATRIC TAXA
(r = sympatric taxa present in redundant area r)

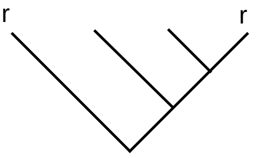
4 taxa



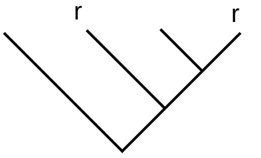
red401



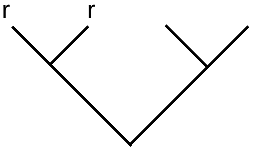
red402



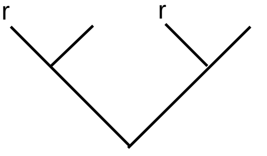
red403



red404

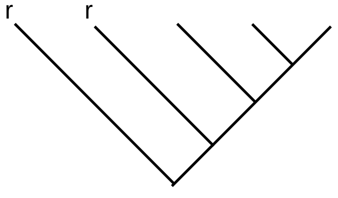


red405

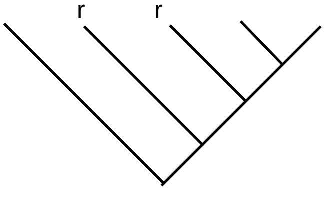


red406

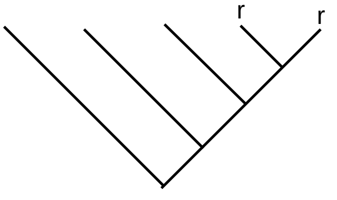
5 taxa



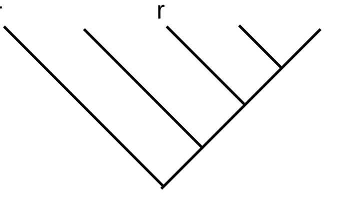
red501



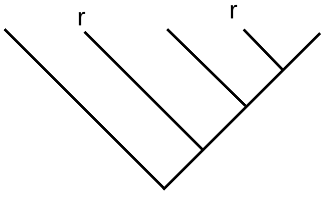
red502



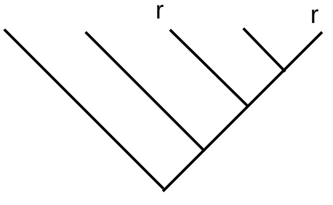
red503



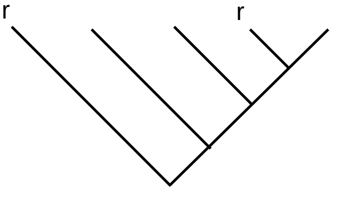
red504



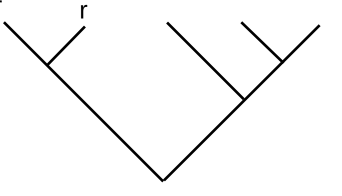
red505



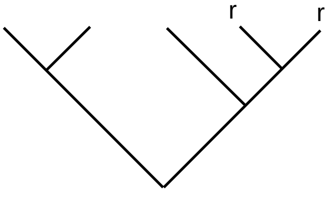
red506



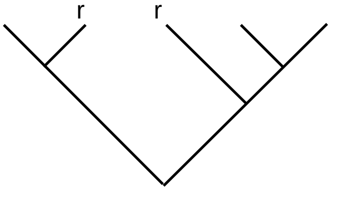
red507



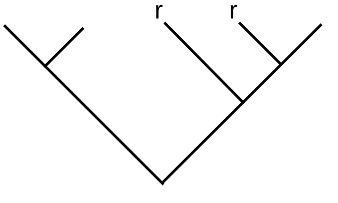
red508



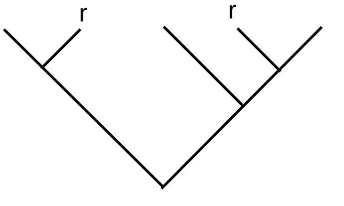
red509



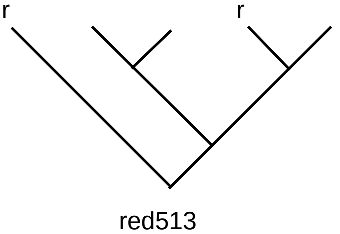
red510



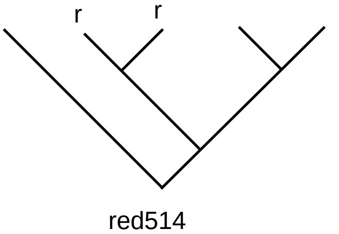
red511



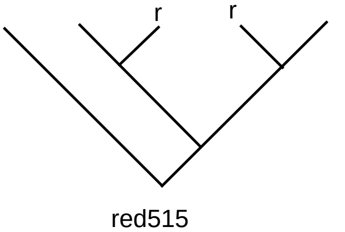
red512



red513



red514

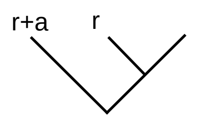


red515

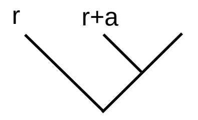
SYMPATRIC & WIDESPREAD TAXA

(r = sympatric taxa present in redundant area r;
r+a = widespread taxon present in both redundant area r and another non-redundant area a)

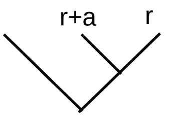
3 taxa



rew301

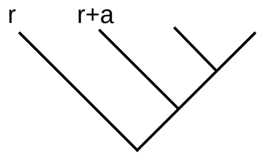


rew302

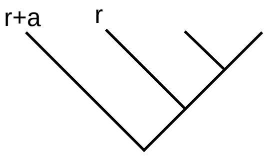


rew303

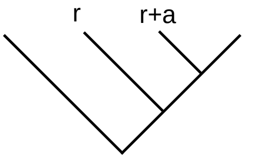
4 taxa



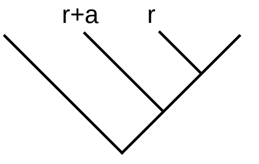
rew401



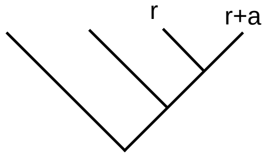
rew402



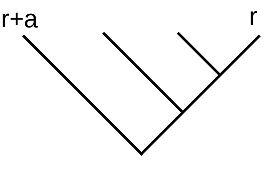
rew403



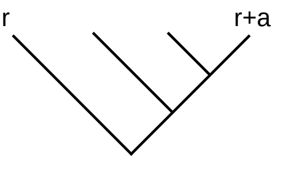
rew404



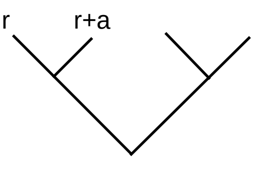
rew405



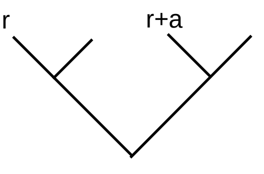
rew406



rew407

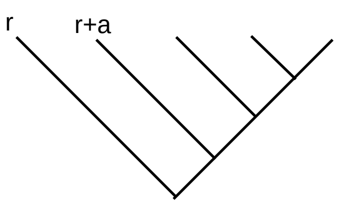


rew408

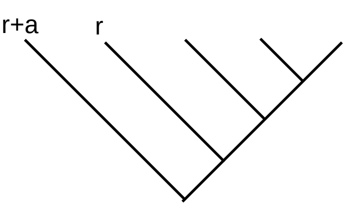


rew409

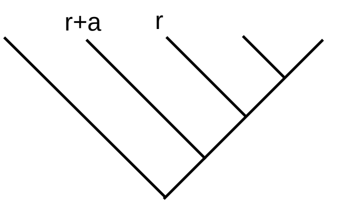
5 taxa



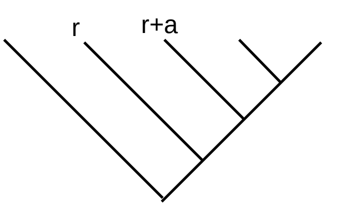
rew501



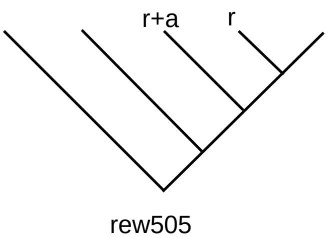
rew502



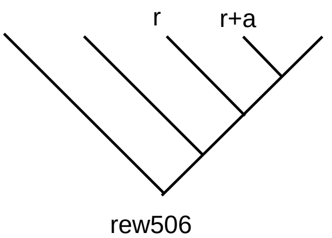
rew503



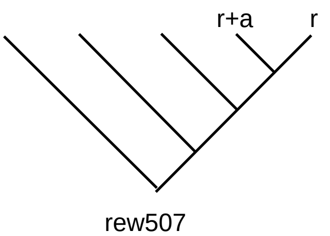
rew504



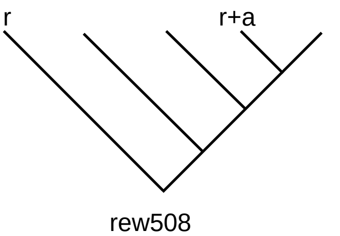
rew505



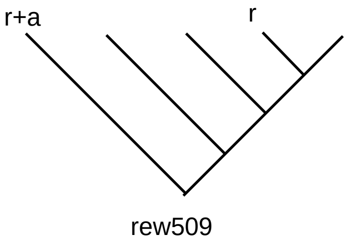
rew506



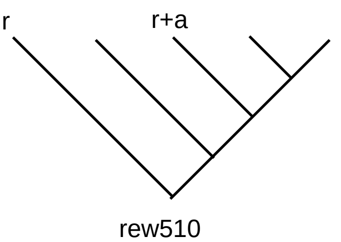
rew507



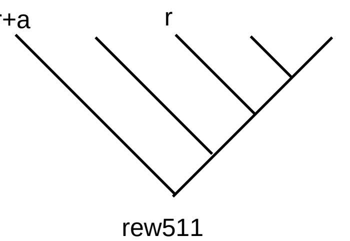
rew508



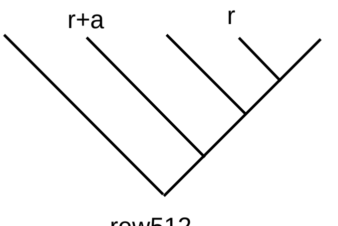
rew509



rew510

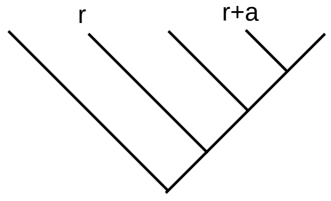


rew511

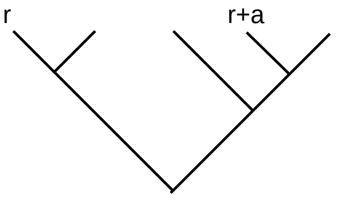


rew512

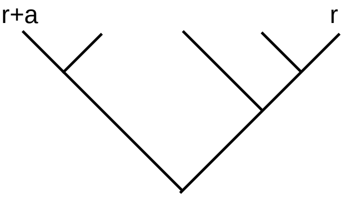
5 taxa



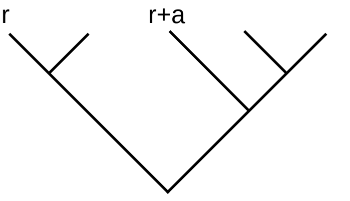
rew513



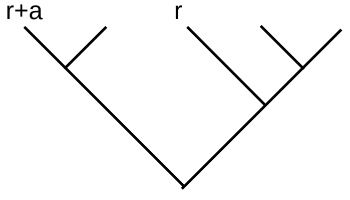
rew514



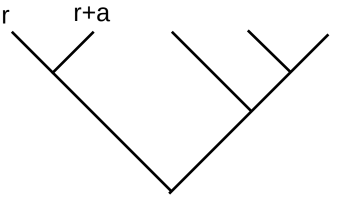
rew515



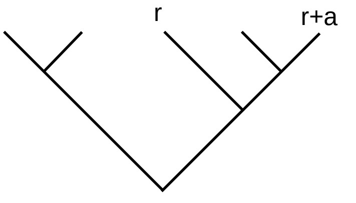
rew516



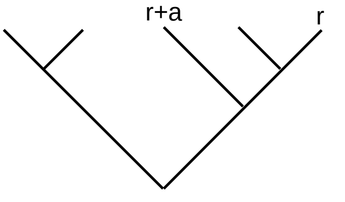
rew517



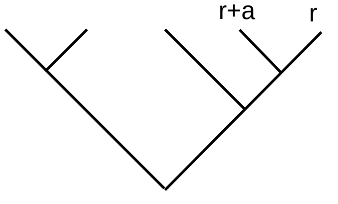
rew518



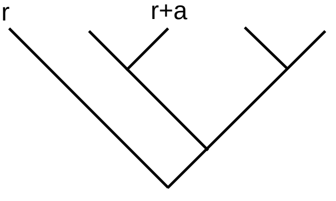
rew519



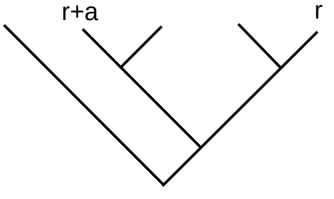
rew520



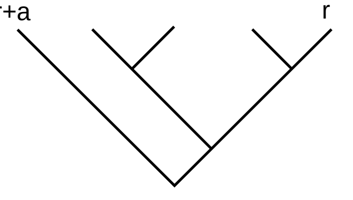
rew521



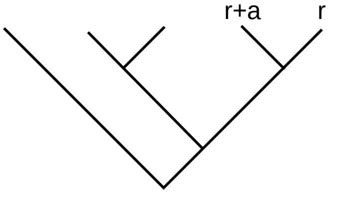
rew522



rew523



rew524

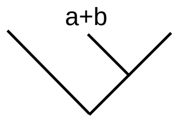


rew525

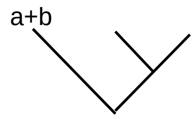
WIDESPREAD TAXA

(a+b = widespread taxon present in both area a and area b)

3 taxa

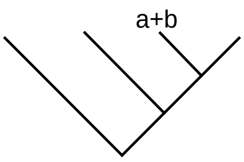


wid301

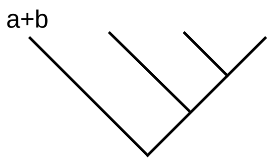


wid302

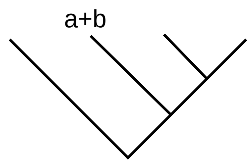
4 taxa



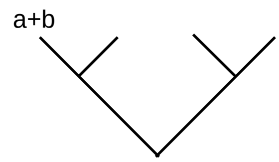
wid401



wid402

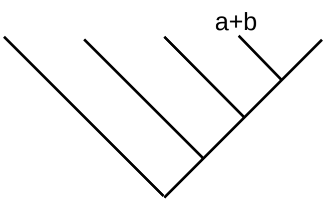


wid403

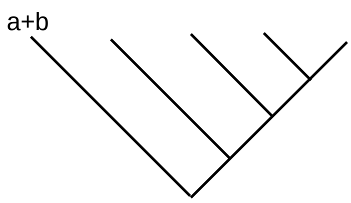


wid404

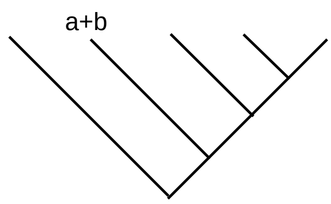
5 taxa



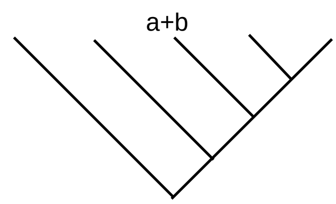
wid501



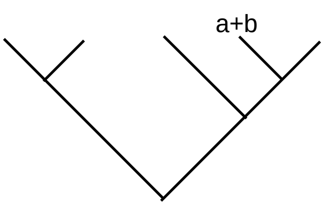
wid502



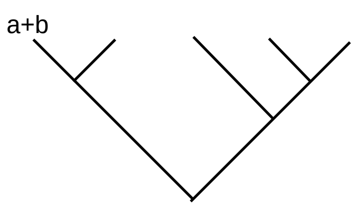
wid503



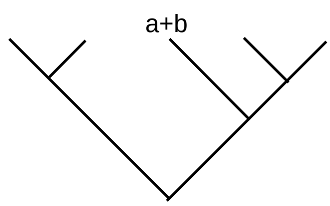
wid504



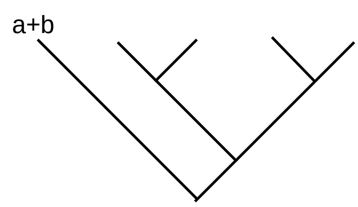
wid505



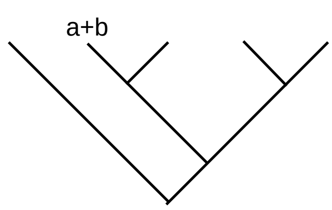
wid506



wid507

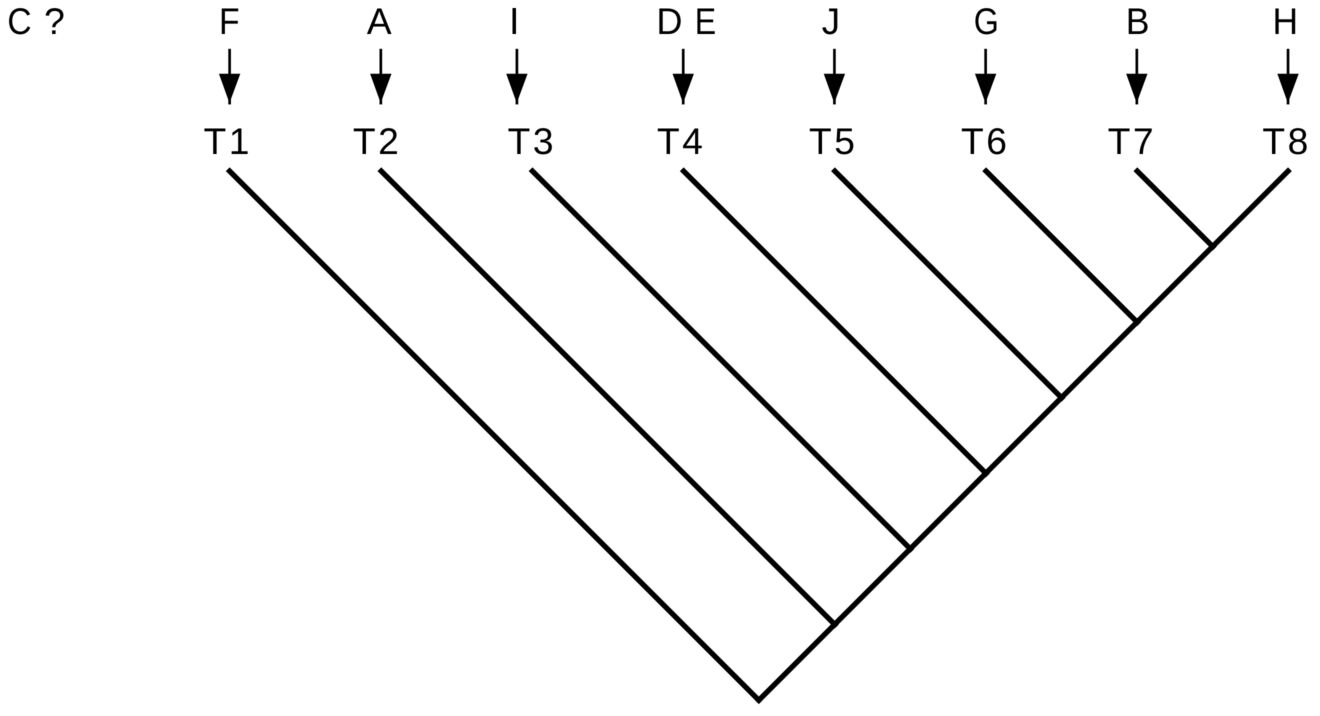


wid508

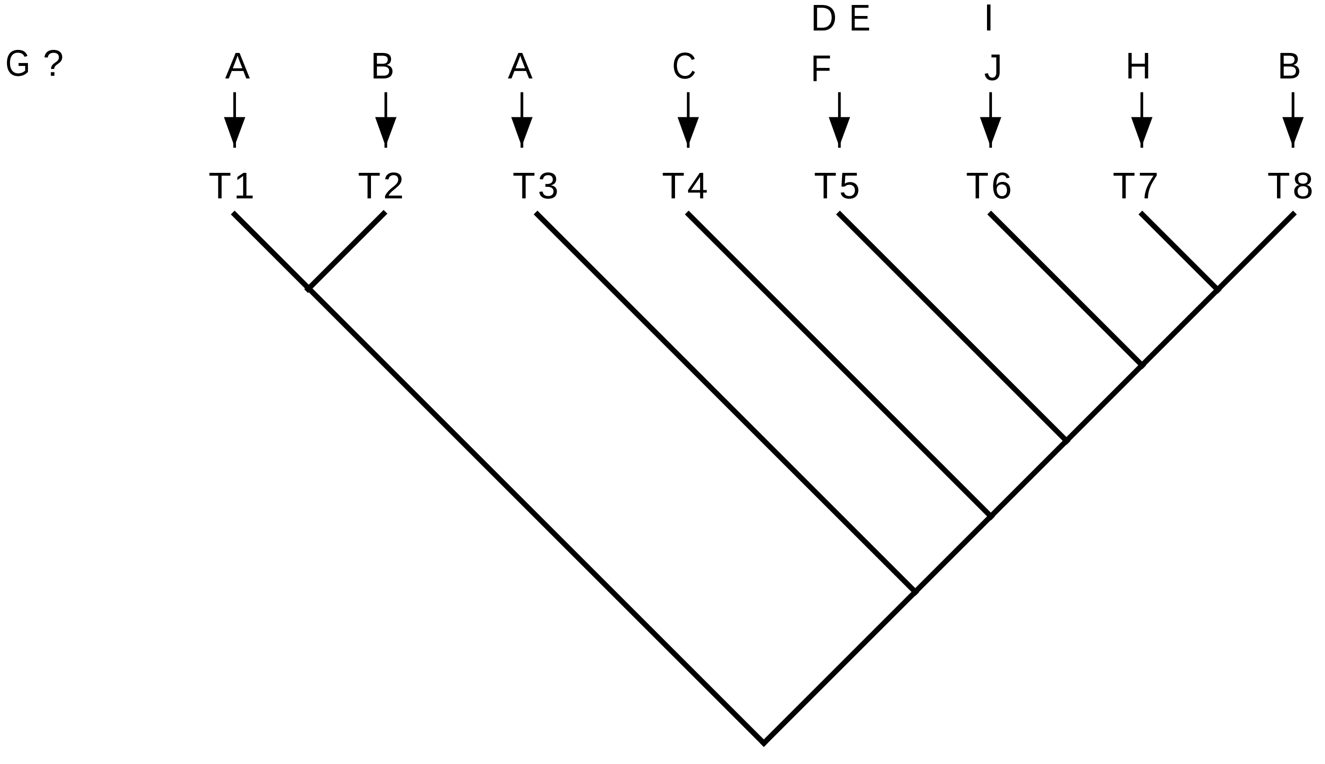


wid509

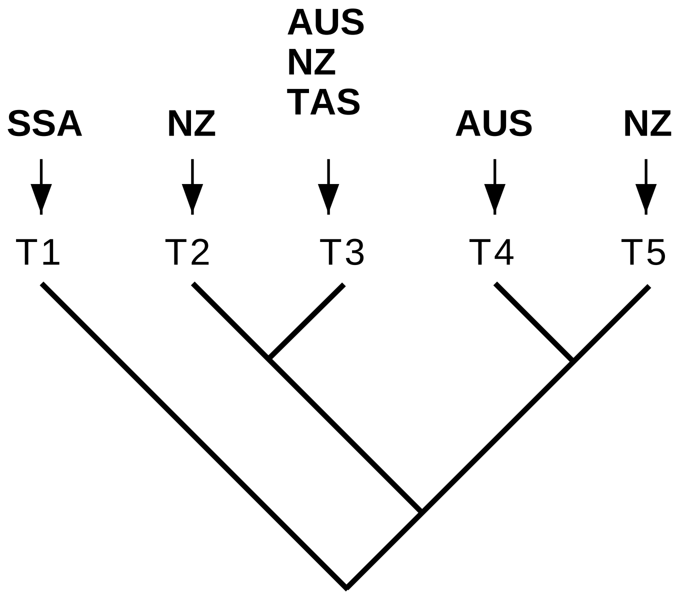
Heterandria



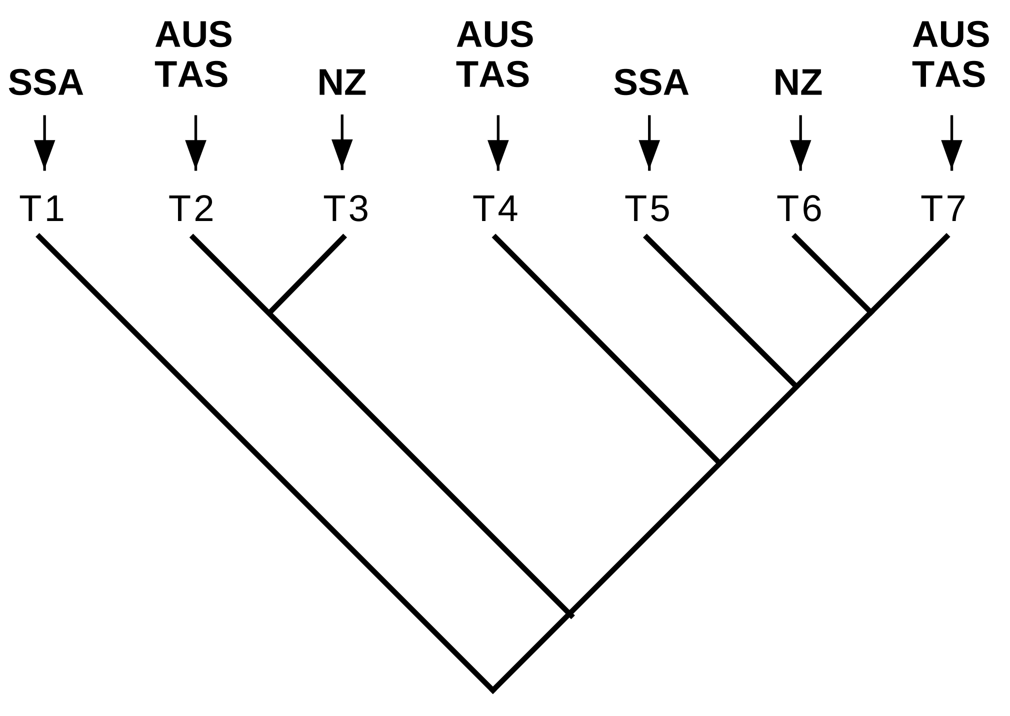
Xiphophorus



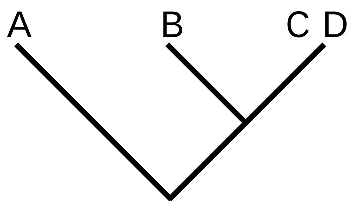
Cyttaria



Eriococcus/Madarococcus



taxon-area cladogram



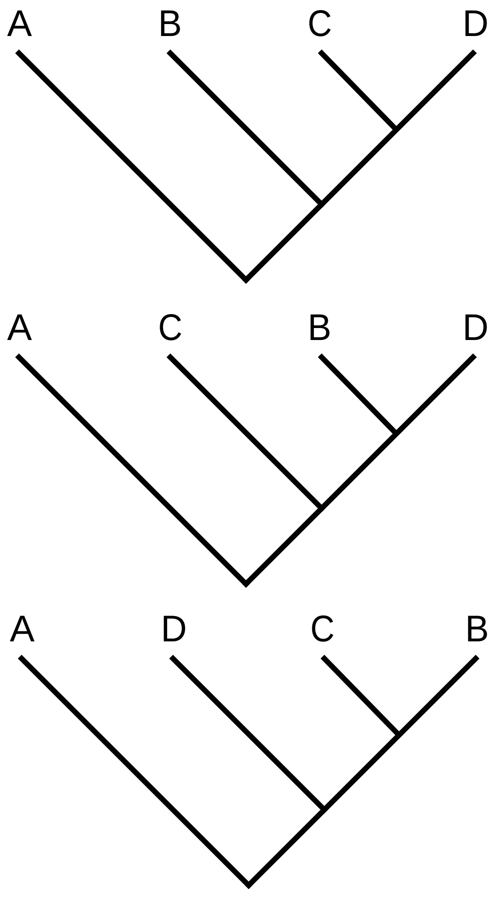
A0 matrix

out	0	0	0	0	0
A	1	0	0	0	1
B	0	1	0	1	1
C	0	0	1	1	1
D	0	0	1	1	1

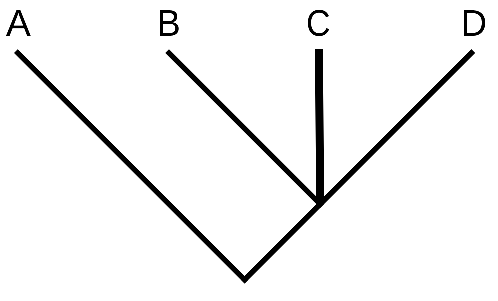
adjusted matrix: A1

out	0	0	0	0
A	1	0	0	1
B	0	1	1	1
C	0	0	1	1
D	0	0	1	1

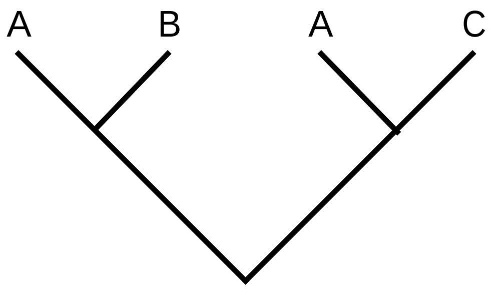
solutions with CA under A1



solution via standard maximum parsimony approach



taxon-area cladogram



A0 matrix

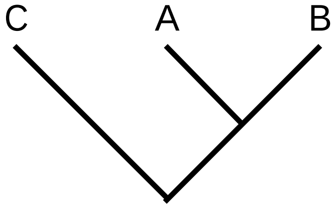
out	0	0	0	0	0	0	0
A	1	0	1	0	1	1	1
B	0	1	0	0	1	0	1
C	0	0	0	1	0	1	1

adjusted matrix: A1

out	0	0	0	0	0	0	0
A	1	0	1	0	1	1	1
B	0	1	?	?	1	?	1
C	?	?	0	1	?	1	1

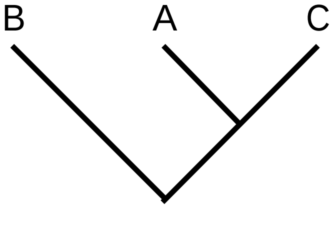
solutions with
CA under A1

optimised matrix



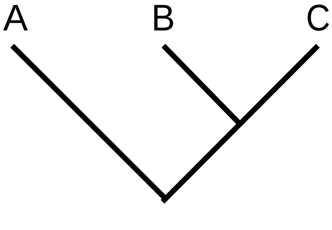
out	0	0	0	0	0	0	0
A	1	0	1	0	1	1	1
B	0	1	1	0	1	1	1
C	0	0	0	1	0	1	1

2 columns adjusted
2 different components
2 extinctions



out	0	0	0	0	0	0	0
A	1	0	1	0	1	1	1
B	0	1	0	0	1	0	1
C	1	0	0	1	1	1	1

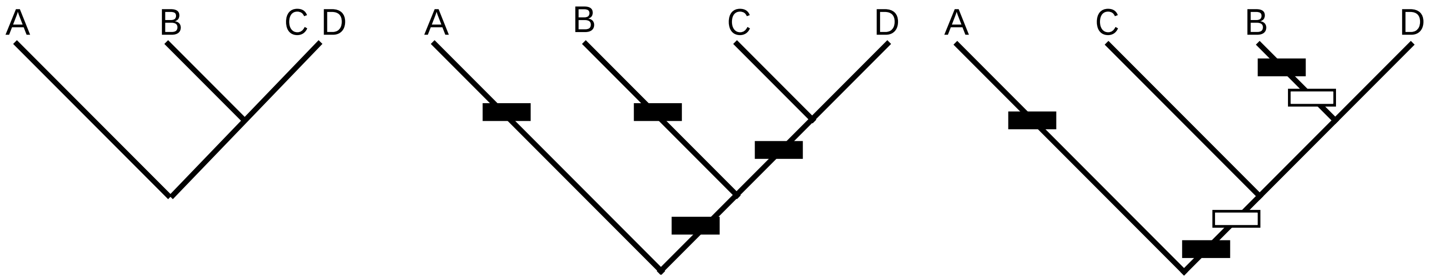
2 columns adjusted
2 different components
2 extinctions



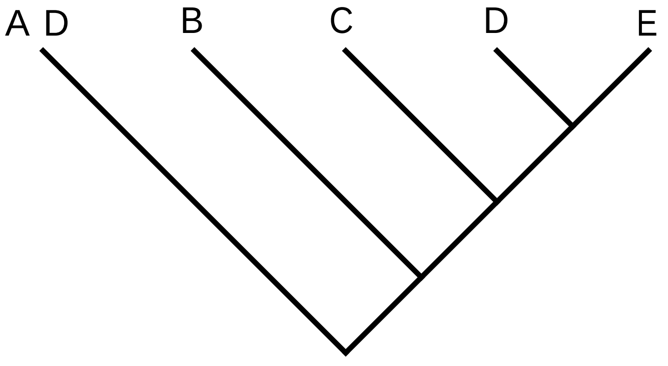
out	0	0	0	0	0	0	0
A	1	0	1	0	1	1	1
B	0	1	0	1	1	1	1
C	0	1	0	1	1	1	1

4 columns adjusted
2 different components
2 extinctions

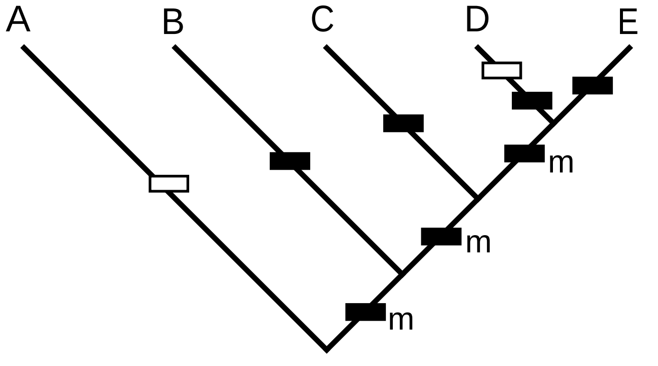
taxon-area cladogram | A0: 4 steps | A1: 5 steps



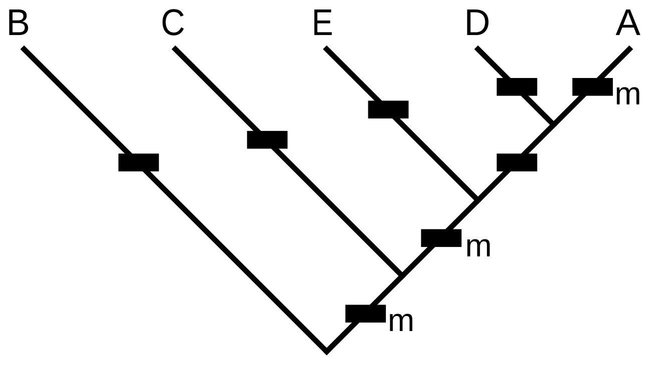
taxon-area cladogram



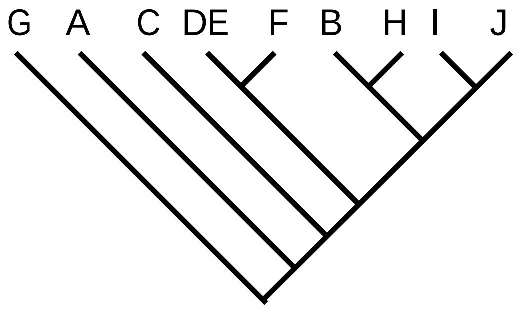
A0/A1: 9 steps



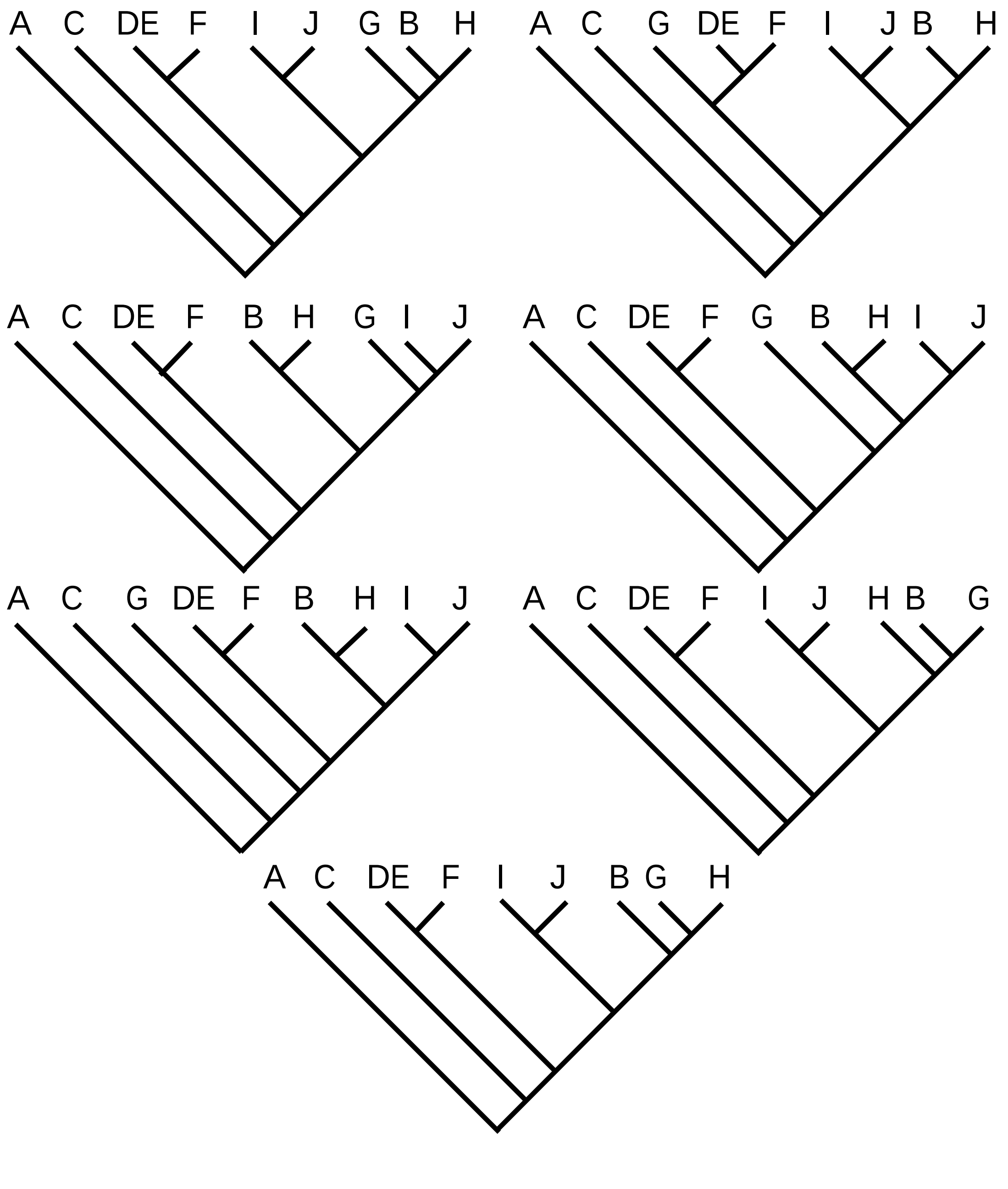
A2: 8 steps



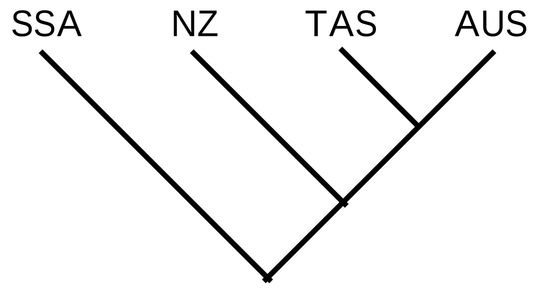
a


G A C DE F B H I J

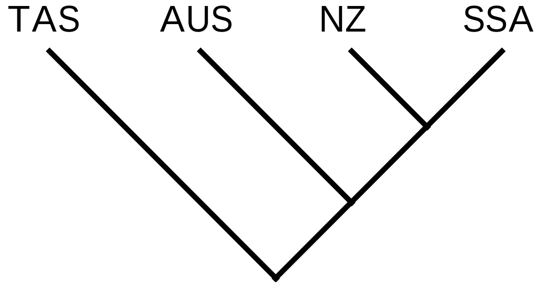
b


A C DE F I J G B H
A C G DE F I J B H
A C DE F B H G I J
A C DE F G B H I J
A C G DE F B H I J
A C DE F I J H B G
A C DE F I J B G H

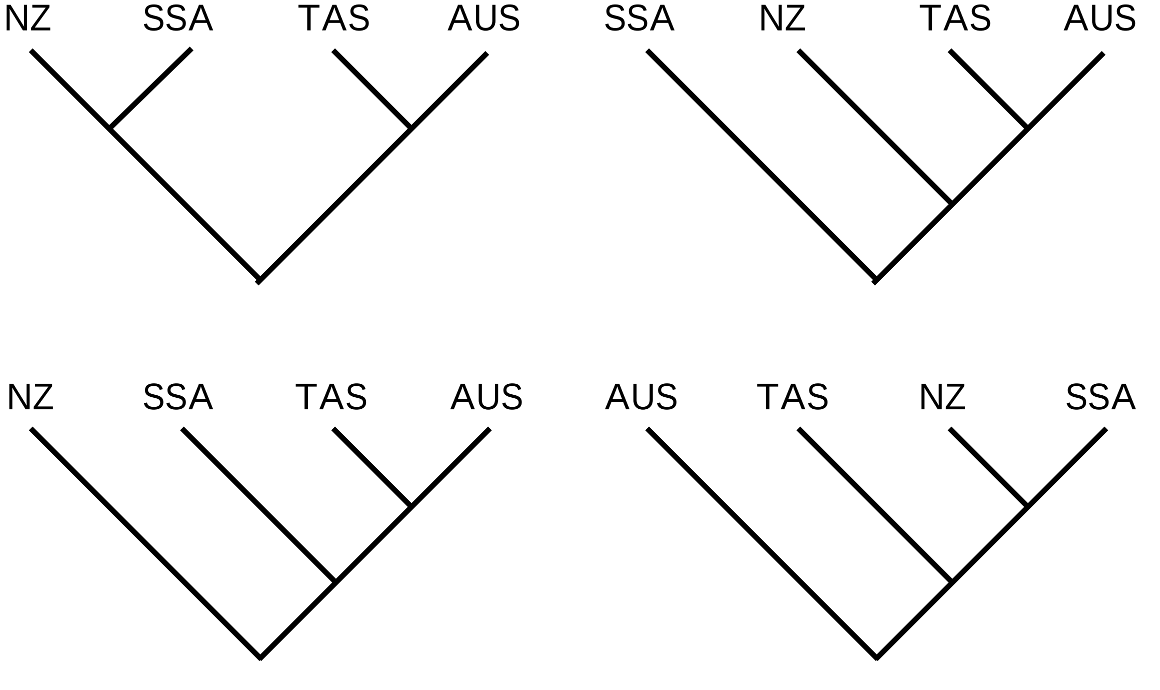
a

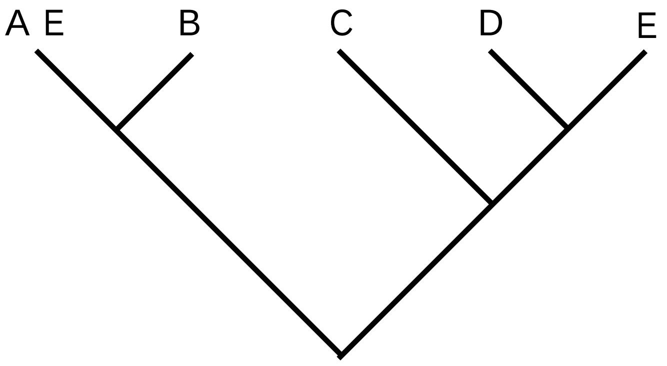

SSA
NZ
TAS
AUS

b

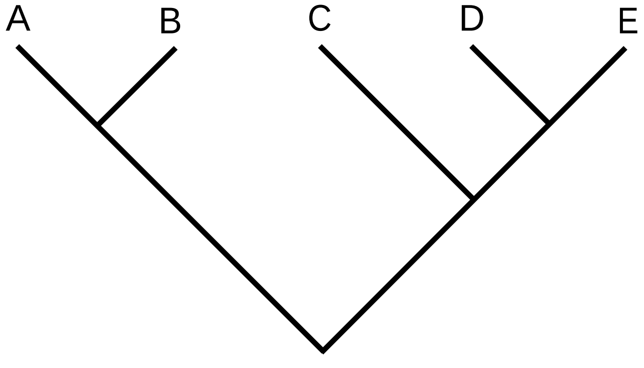

TAS
AUS
NZ
SSA

c

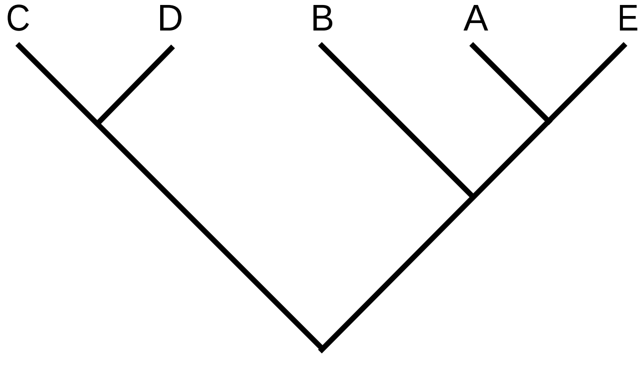

NZ
SSA
TAS
AUS
SSA
NZ
TAS
AUS
NZ
SSA
TAS
AUS
AUS
TAS
NZ
SSA



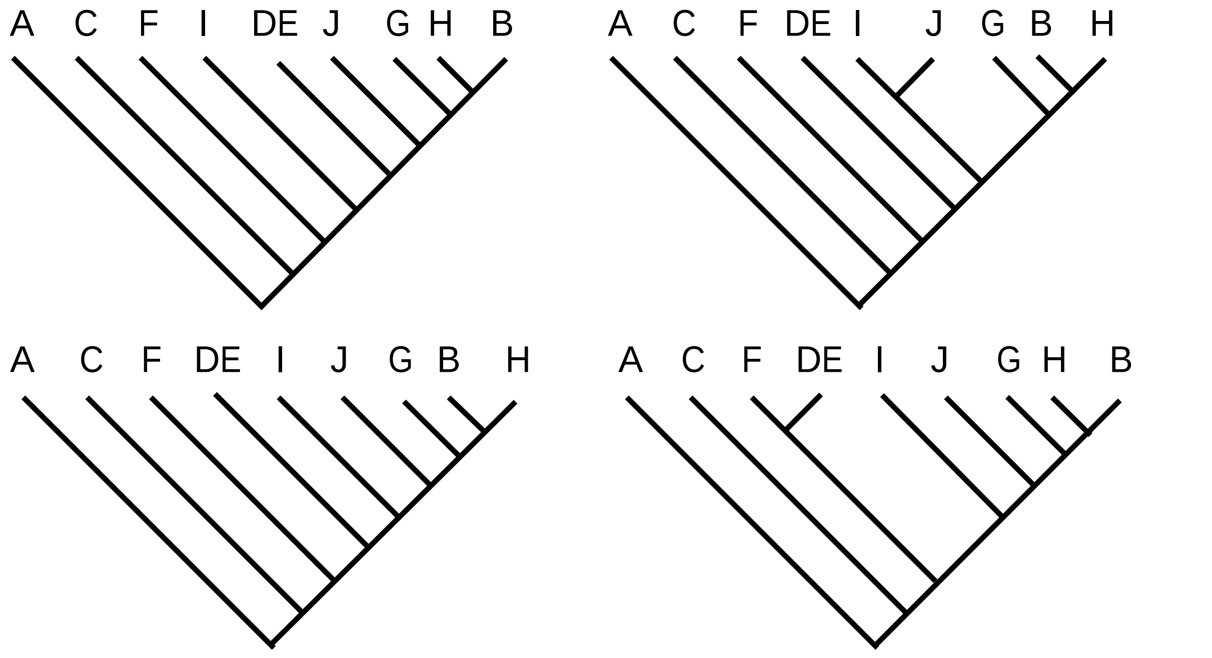
A0: 11 steps; A1: 15 steps



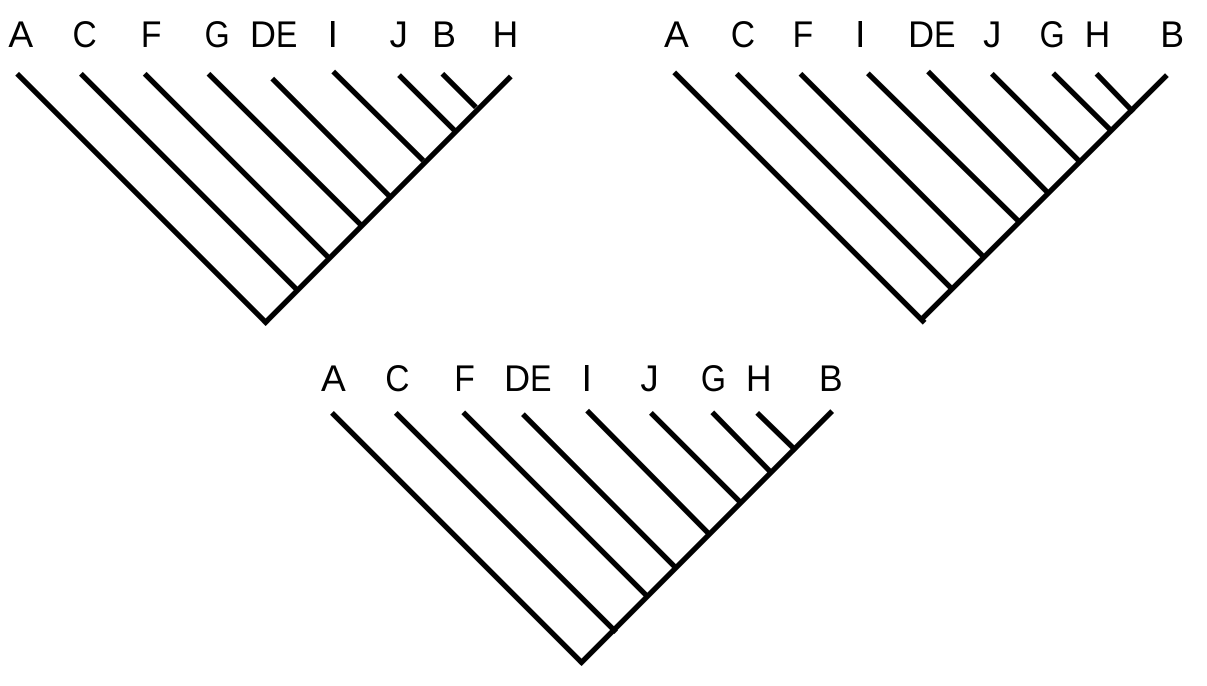
A0: 11 steps; A1: 16 steps → not selected under A1



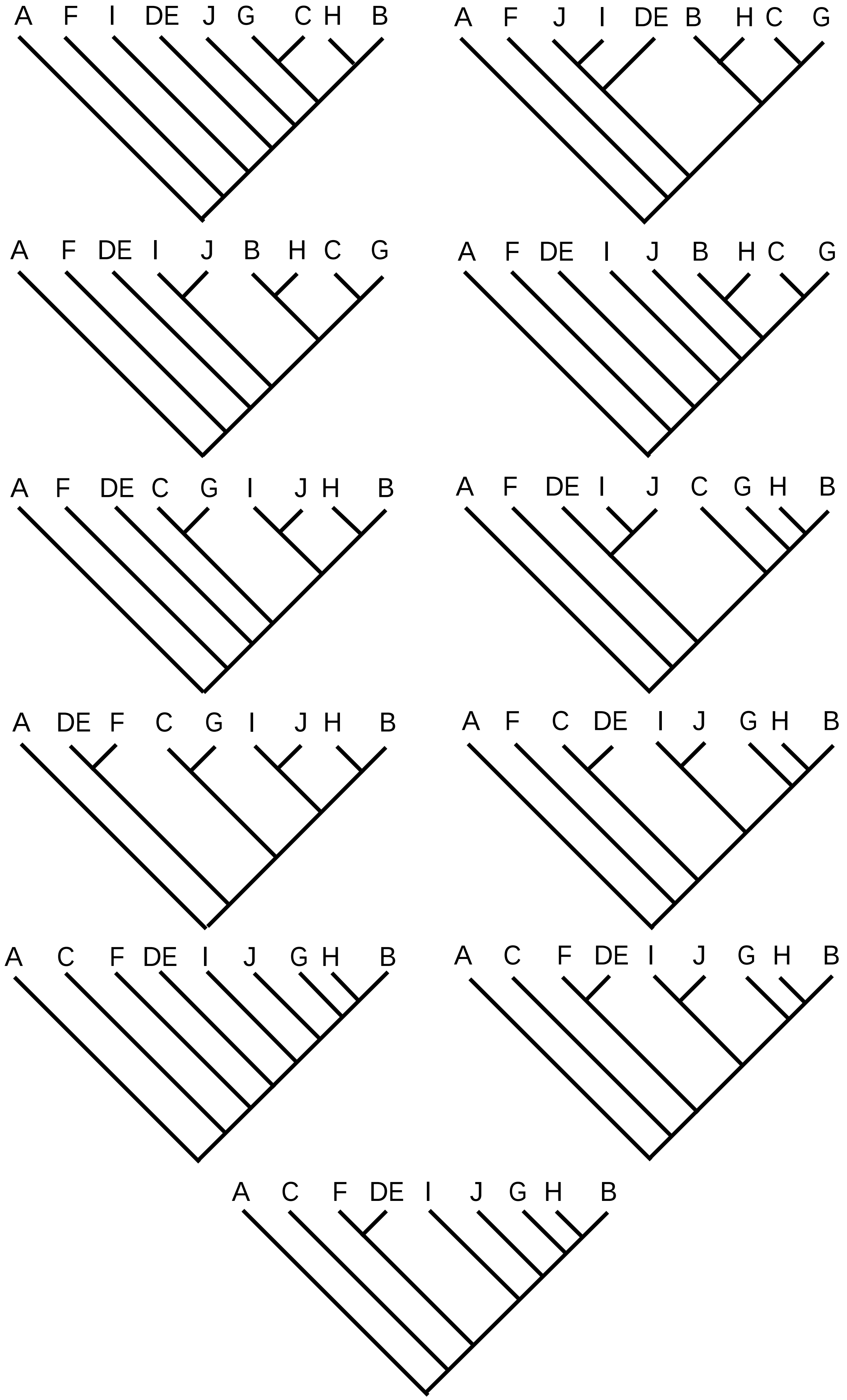
a

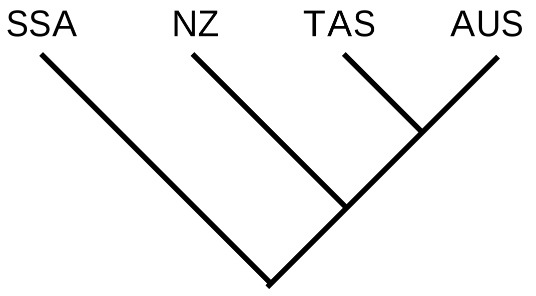

A C F I DE J G H B
A C F DE I J G B H
A C F DE I J G B H
A C F DE I J G H B

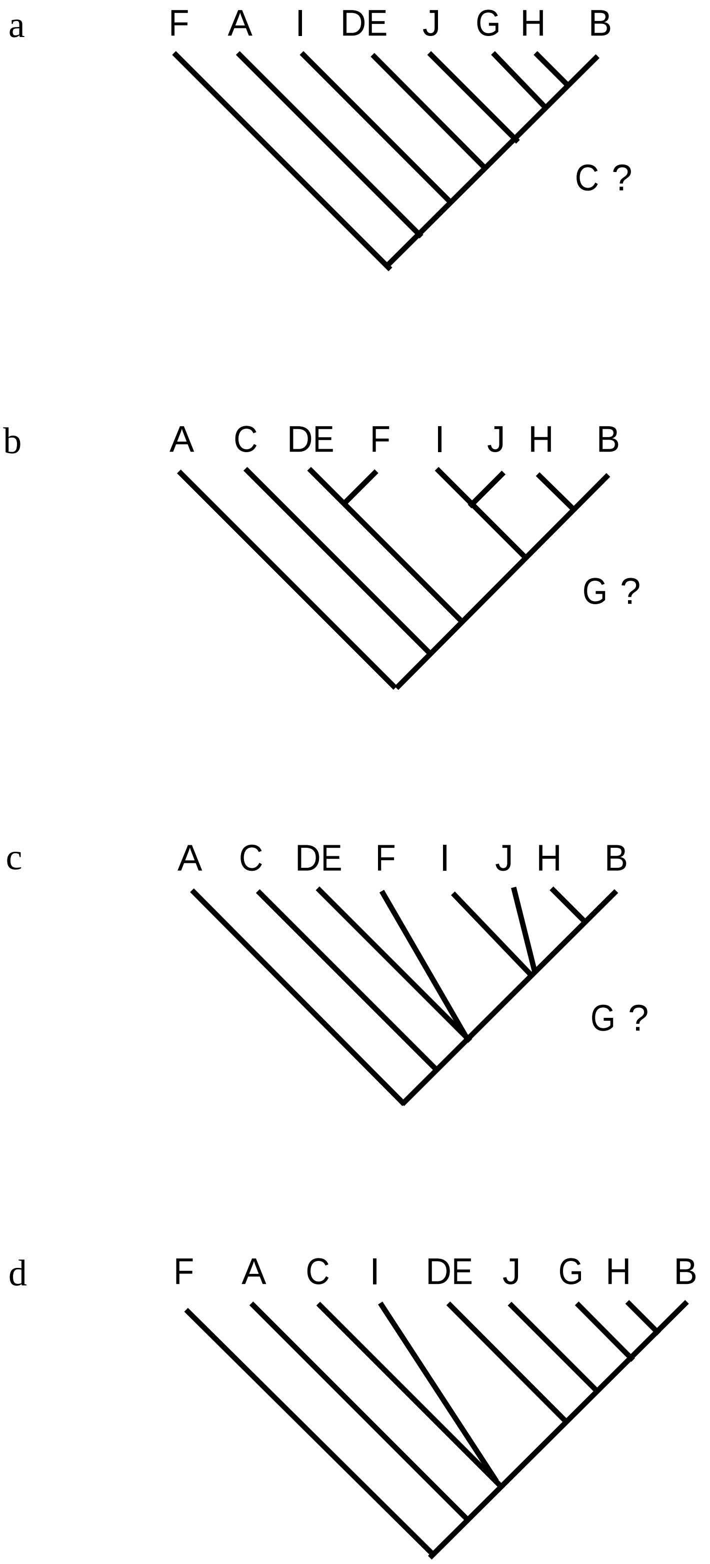
b


A C F G DE I J B H
A C F I DE J G H B
A C F DE I J G H B

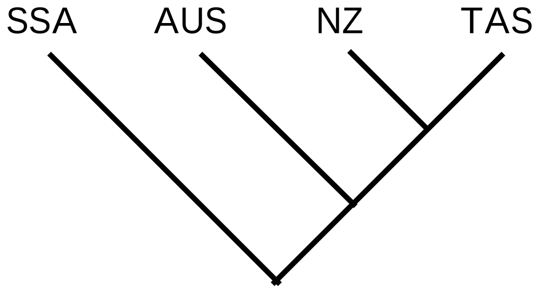
c


A F I DE J G C H B
A F J I DE B H C G
A F DE I J B H C G
A F DE I J B H C G
A F DE C G I J H B
A F DE I J C G H B
A DE F C G I J H B
A F C DE I J G H B
A C F DE I J G H B
A C F DE I J G H B
A C F DE I J G H B

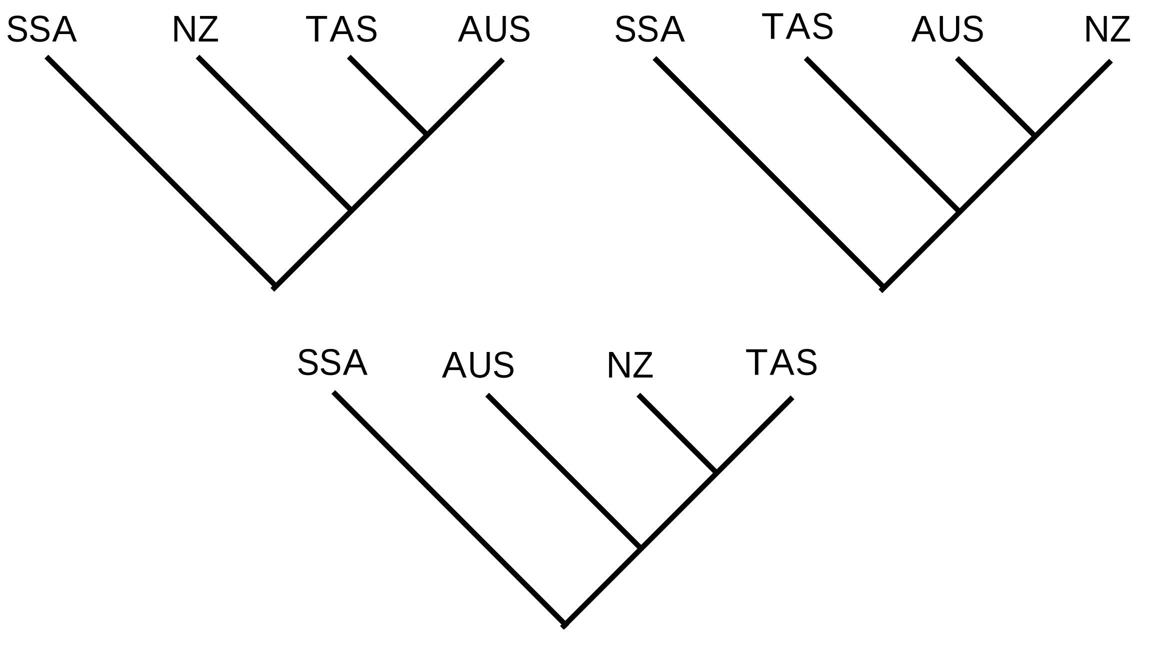

SSA
NZ
TAS
AUS


a
F A I DE J G H B
C ?
b
A C DE F I J H B
G ?
c
A C DE F I J H B
G ?
d
F A C I DE J G H B

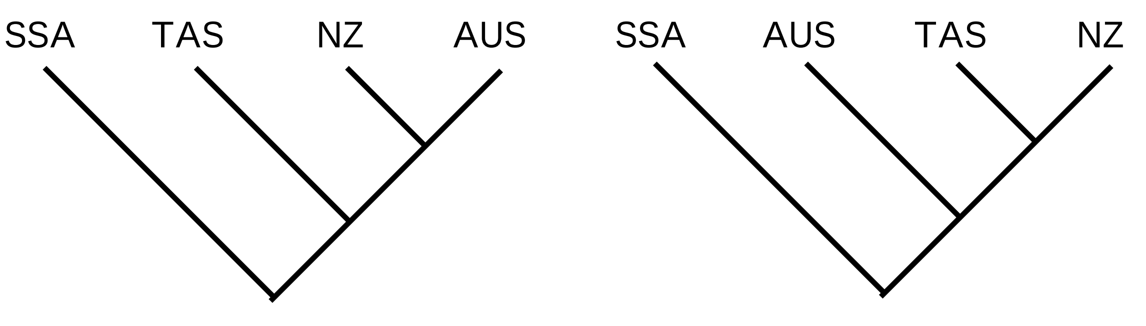
a


SSA
AUS
NZ
TAS

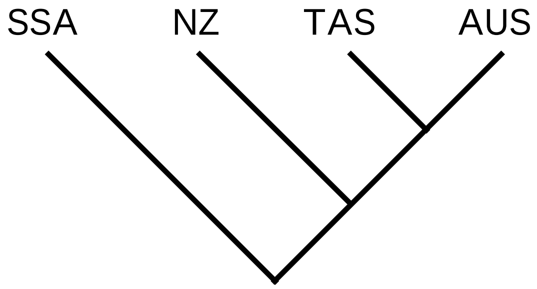
b


SSA
NZ
TAS
AUS
SSA
TAS
AUS
NZ
SSA
AUS
NZ
TAS

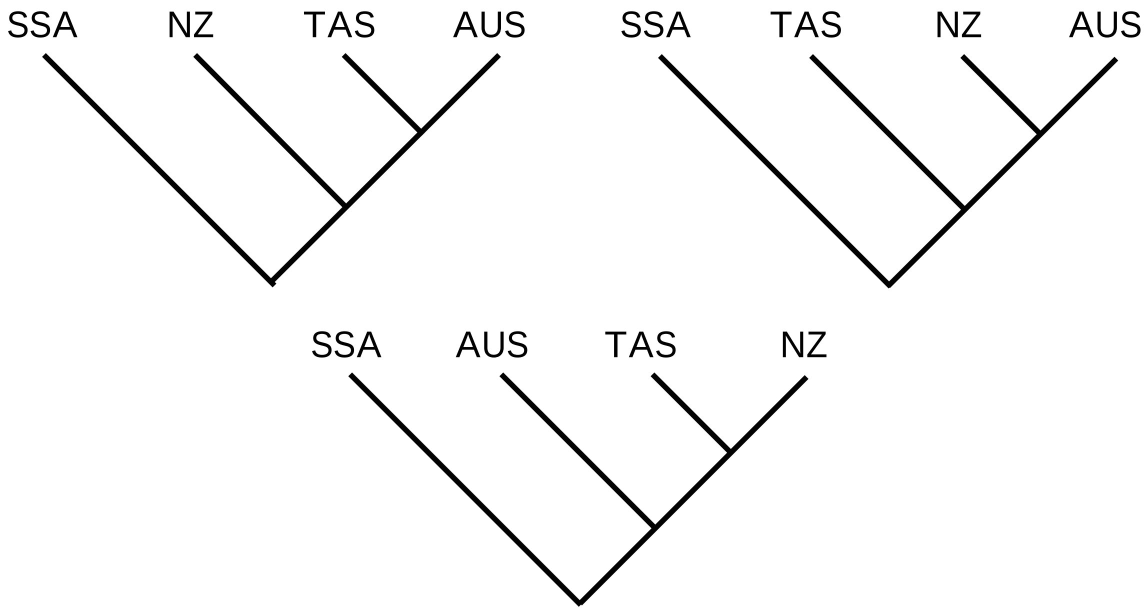
c


SSA
TAS
NZ
AUS
SSA
AUS
TAS
NZ

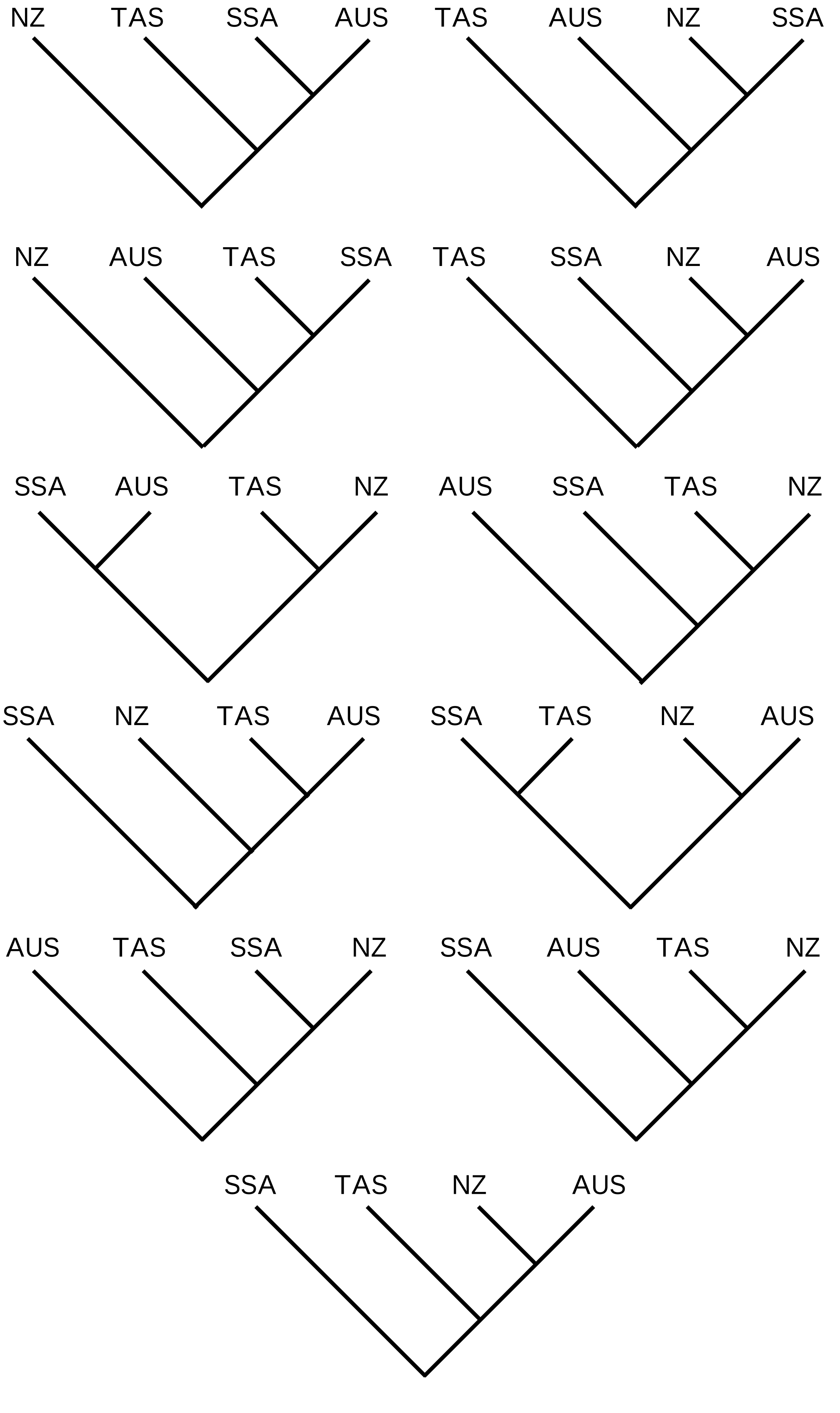
a



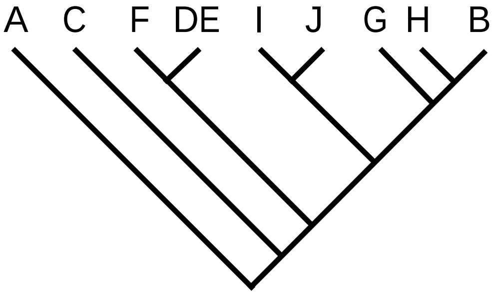
b



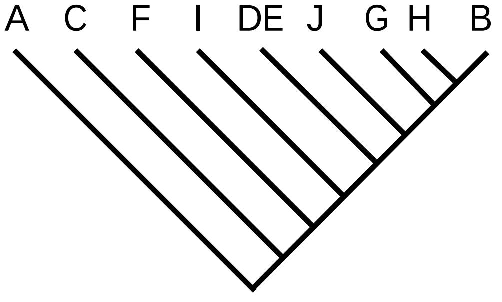
c


NZ TAS SSA AUS
TAS AUS NZ SSA
NZ AUS TAS SSA
TAS SSA NZ AUS
SSA AUS TAS NZ
AUS SSA TAS NZ
SSA NZ TAS AUS
SSA TAS NZ AUS
AUS TAS SSA NZ
SSA AUS TAS NZ
SSA TAS NZ AUS

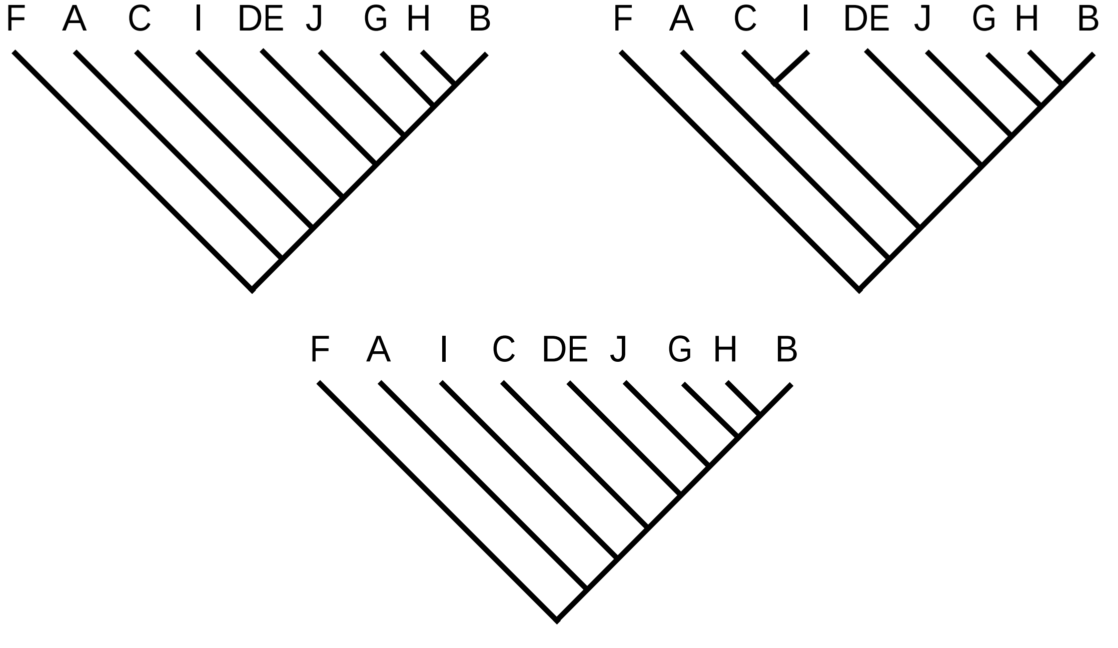
a


A C F DE I J G H B

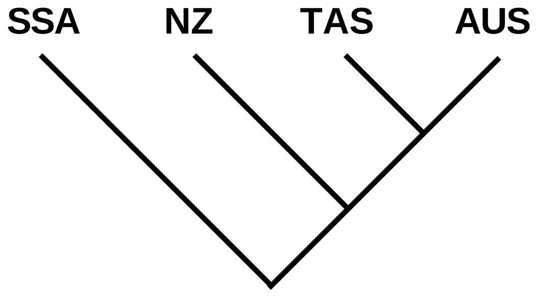
b


A C F I DE J G H B

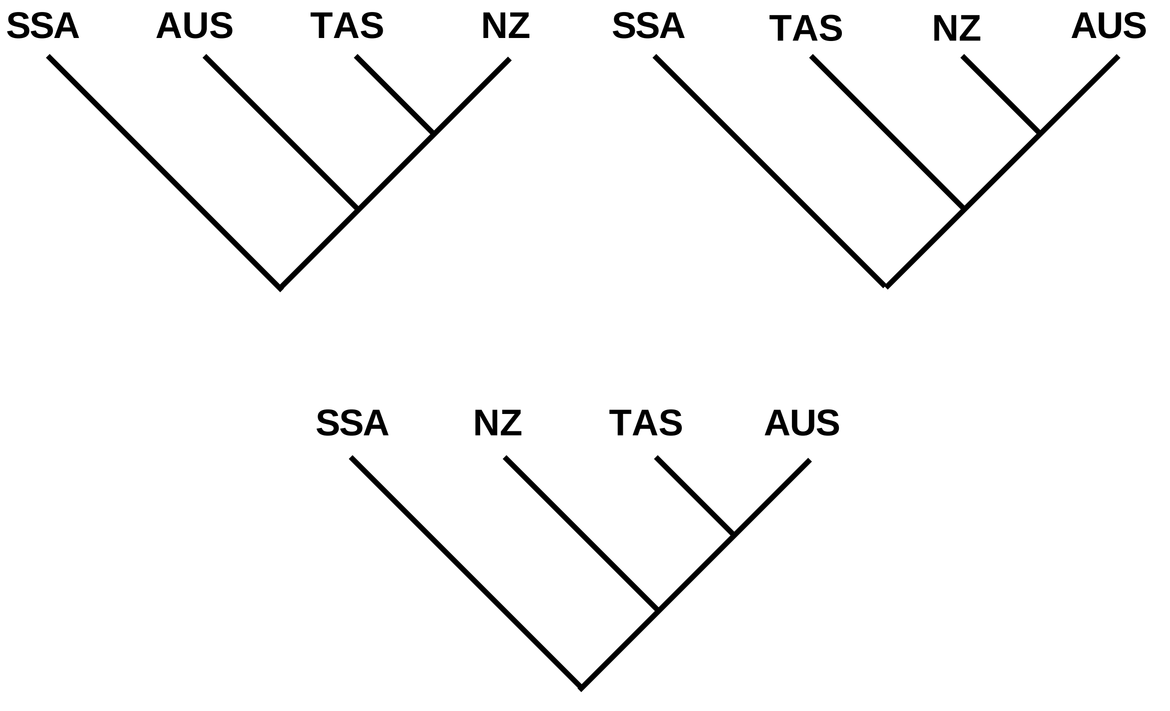
c


F A C I DE J G H B
F A C I DE J G H B
F A I C DE J G H B

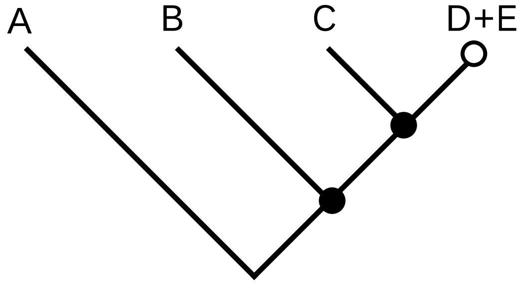
a



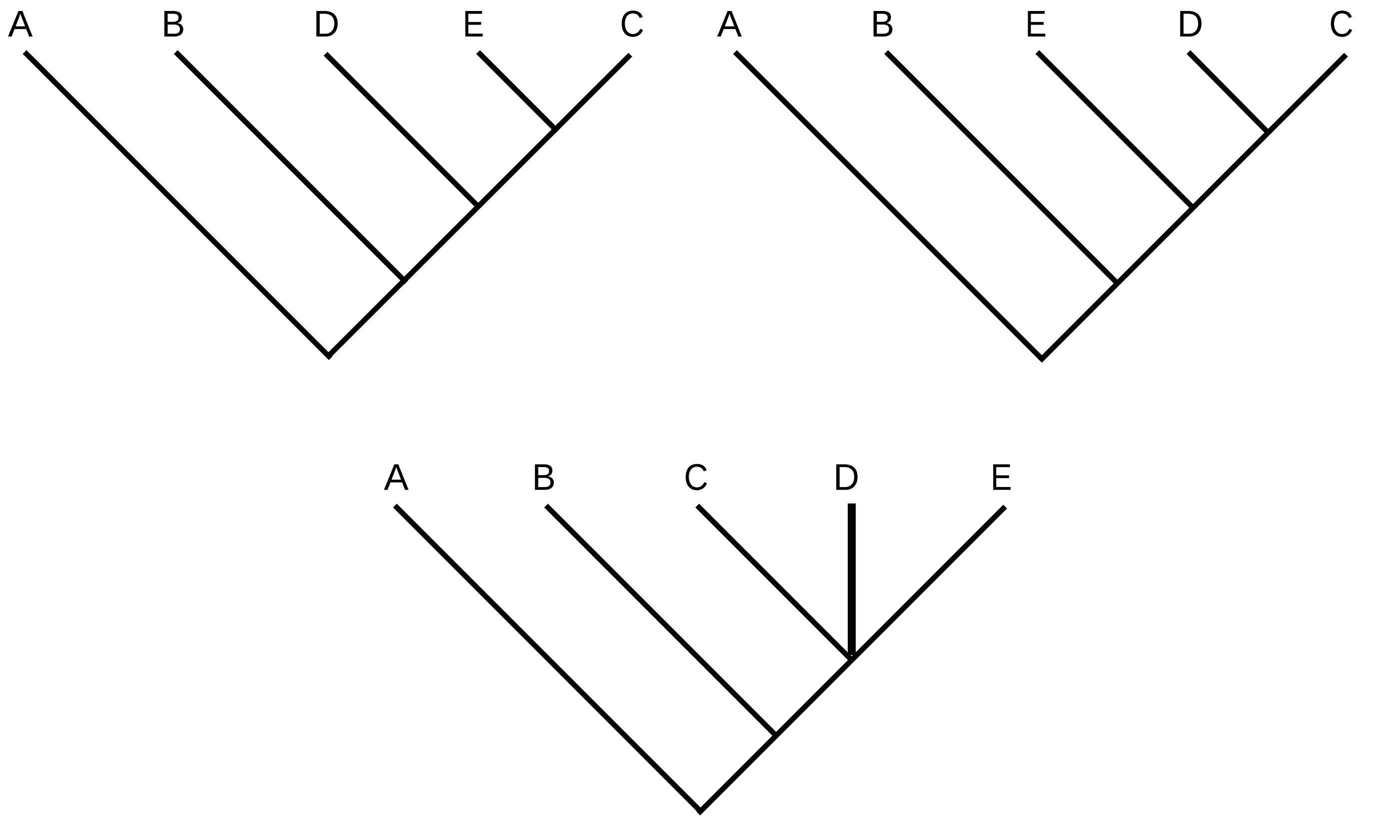
b



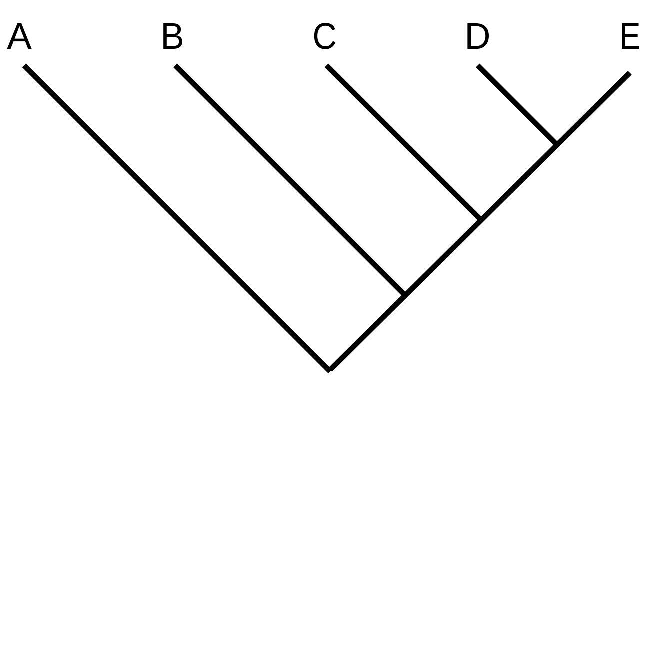
taxon-area cladogram

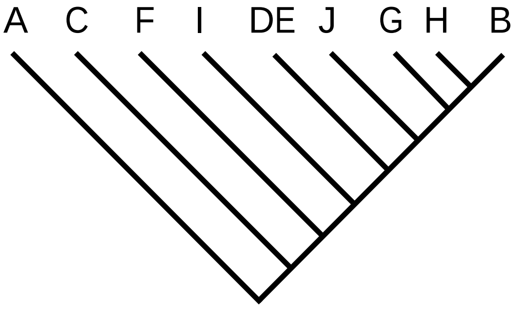


a

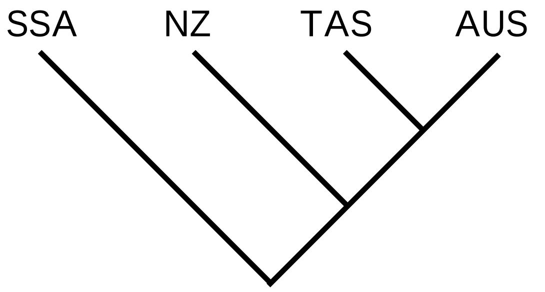


b




A
C
F
I
DE
J
G
H
B

a



b