

PREPRINT

© 2001 Academic Press
© 2001 Willi Hennig Society

Measures for obtaining inclusive sets of area cladograms under assumptions zero, 1, and 2 with different methods for vicariance biogeography

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

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

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

Running title: Obtaining inclusive solution sets in vicariance biogeography.

Keywords: Vicariance biogeography, assumptions zero, 1 and 2, Inclusion,
Component Compatibility Analysis, Brooks Parsimony Analysis, Component
Analysis, Reconciled Tree Analysis, Three Area Statements, CAFCA, PAUP,
Component 1.5, Component 2.0, TAS, Requirement I, Requirement II.

Abstract

We present modifications to computer programs of *a posteriori* and *a priori* methods for vicariance biogeography which enable them to obtain inclusive sets of area cladograms under assumptions zero, 1, and 2. With CAFCA (Component Compatibility Analysis) an upper limit for area cladogram selection by the number of steps is not always sufficient for inclusive sets. CAFCA needs additionally a lower limit for the number of components used to derive area cladograms when noninclusion arises because CAFCA selects area cladograms with different resolutions. PAUP (Brooks Parsimony Analysis) derives inclusive sets when it selects area cladograms under assumptions zero, 1, and 2 by using an upper limit for the number of steps and not collapsing unsupported nodes. For the computer programs Component 1.5 (Component Analysis), Component 2.0 (Reconciled Tree Analysis), and TAS (Three Area Statement Analysis) we suggest a two-step procedure for inclusive sets. The first step involves dealing with widespread taxa *a priori* under assumptions zero, 1 or 2. The second step involves dealing with sympatric taxa “as is” (Component 1.5 and TAS) or by tree reconciliation using an upper limit for the number of losses (Component 2.0).

4.1 Introduction

In vicariance biogeography, resolved area cladograms (*i.e.* area cladograms with unique terminal nodes for all areas; Morrone and Carpenter, 1994; Enghoff, 1996) are obtained from data on the distribution and cladogenetic relationships of monophyletic groups of taxa. In straightforward cases each taxon of a monophyletic group is

endemic to an area and each area harbours a single taxon. In such cases, one obtains resolved area cladograms by replacement of taxa in the taxon cladogram by the areas in which these taxa occur.

Whenever cases are not straightforward, taxa of monophyletic groups are widespread distributed or occur together in the same area with taxa of the same monophyletic group (*i.e.* sympatric taxa). It is in these cases that resolved area cladograms are derived under one of three different assumptions: assumption zero (A0), 1 (A1), or 2 (A2). These assumptions determine how deviations from a straightforward vicariance model are analysed and interpreted.

Under A0 the distribution data for the taxa concerned are *a priori* assumed to represent the effects of vicariance alone. Under A1 the distribution data for the taxa concerned are *a priori* assumed to represent the effects of vicariance and extinction. Under A2 the distribution data for the taxa concerned are *a priori* assumed to represent the effects of vicariance, extinction and dispersal (Van Veller *et al.*, 1999).

In Van Veller *et al.* (1999) we argued that two requirements are relevant when these assumptions are applied.

First, the sets of area cladograms (*i.e.* solution sets S) obtained under A0, A1 and A2 for one group of taxa should be inclusive, *i.e.*, $S_0 \subseteq S_1 \subseteq S_2$ (Requirement I). That is, area cladograms derived under A0 should be contained in the set of area cladograms derived under A1 and these in turn should be contained in the set of area cladograms derived under A2. The processes that one considers *a priori* to result in historical relationships of areas as recovered under A0, A1, or A2 are taken to be largely independent of each other. This means that their effects are additive and as a consequence the resulting patterns are inclusive.

Second, when solution sets for different monophyletic groups of taxa are compared in order to find a common pattern (*i.e.* a general area cladogram), these sets should be obtained under the same assumption (Requirement II). We do not allow a common pattern to be *a priori* explained for one group of taxa by one particular set of processes (*e.g.* vicariance plus extinction) and for another group of taxa by a different set of processes (*e.g.* vicariance plus extinction as well as dispersal).

When computer programs that implement A0, A1, and A2 under the different methods are assessed for the extent to which they meet these requirements, it appears that they never violate Requirement II (Van Veller *et al.*, 2000). However, none of the computer programs assessed satisfies Requirement I.

In this paper we present and discuss modifications to the following computer programs which allow them to obtain inclusive solution sets under A0, A1, and A2: CAFCA vs. 1.5j (Zandee, 1999) as an implementation of Component Compatibility Analysis (CCA; Zandee and Roos, 1987); PAUP vs. 3.11 (Swofford, 1990) as an implementation of Brooks Parsimony Analysis (BPA; Brooks, 1990); Component 1.5 (Page, 1990a) as an implementation of Component Analysis (CA; Nelson and Platnick, 1981); Component 2.0 (Page, 1993) as an implementation of Reconciled Tree Analysis (RTA; Page, 1994); and TAS (Nelson and Ladiges, 1991a) as an implementation of Three Area Statement Analysis (TAS; Nelson and Ladiges, 1991b).

4.2 How to deal with widespread and sympatric taxa to obtain inclusion

In an assessment (Van Veller *et al.*, 2000) of the computer programs (CAFCA, PAUP, Component 1.5 and Component 2.0, and TAS) that implement A0, A1, and A2 under different methodologies (CCA, BPA, CA, RTA, and TAS), we showed that dealing with widespread taxa under A0, A1, and A2 almost never results in violation of Requirement I. In such cases, taxon-area cladograms are derived with a single area at each terminal node. When no sympatric taxa are present in the data set, these taxon-area cladograms are equivalent to resolved area cladograms (*sensu* Morrone and Carpenter, 1994 and Enghoff, 1996), and the solution sets obtained under A0, A1, and A2 mostly show inclusion (Van Veller *et al.*, 2000).

However, when sympatric taxa are present as well, not all areas have a unique place in the taxon-area cladogram. To obtain resolved area cladograms, the sympatric taxa have to be dealt with by additional steps.

In Van Veller *et al.* (1999) we argued that sympatric taxa should always be dealt with *a posteriori*. Assumptions 1 and 2 as formulated by Nelson and Platnick (1981) are adopted by the authors of the *a priori* methods and implemented in Component 1.5, Component 2.0, and TAS. In Van Veller *et al.* (1999) we showed that dealing with sympatric taxa *a priori* (as in these computer programs) results either in area cladograms that are inconsistent with the original formulation of the assumption 1 or area cladograms based upon ignoring available data by the removal of taxa from the taxon-area cladogram (assumption 2). To overcome this inconsistency of the results with the original definition of the assumption or *a priori* ignorance of data we recommended dealing with sympatric taxa only “as is”. By this, we mean that no

attempt should be made to deal with sympatric taxa *a priori*, but that an explanation for the sympatric distribution of taxa must be revealed *a posteriori* from the area cladograms.

4.3 Taxon-area cladogram rew515

To evaluate the performance of the computer programs for the methods with respect to their meeting of Requirement I, we used theoretical data sets in Van Veller *et al.* (2000). These data sets consist of taxon-area cladograms with a single widespread taxon, two sympatric taxa or a combination of a widespread taxon and two sympatric taxa in all possible arrangements over all possible topologies for three, four, and five taxa. Taxon-area cladogram rew515 (Fig. 1) is based on a topology of five taxa (with a combination of a widespread taxon (T1) that is also sympatric with another taxon (T5) in one of the two areas in which it is present). It is the smallest data set that results in noninclusive solution sets when analyses under A0, A1, and A2 with *all* different computer programs are performed (Van Veller *et al.*, 2000).

In the next part of this paper we discuss modifications to computer programs (CAFCA, PAUP, Component 1.5, Component 2.0, TAS) that are the implementations of different methodologies (CCA, BPA, CA, RTA, and TAS) which enable them to obtain inclusive solution sets under A0, A1, and A2 (*i.e.* meeting Requirement I). To show that inclusion is obtained when applying the modifications, we use taxon-area cladogram rew515 as a benchmark. Further, we discuss general expectations with respect to obtaining inclusive solution sets under A0, A1, and A2 when our suggestions are applied to the computer programs.

4.4 Methods, computer programs, and modifications

4.4.1 CAFCA (CCA)

In Van Veller *et al.* (2000) we showed that, for certain theoretical data sets, some combinations of a widespread taxon and sympatric taxa generate area cladograms under A1 and A2 that are more parsimonious than those found under A0 with CAFCA. To obtain inclusive solution sets and to meet Requirement I we proposed to use the (larger) number of steps for the cladogram(s) derived under A0 as an upper limit for cladogram selection under A1 or A2. By this *ad hoc* procedure, CAFCA found not only the most parsimonious area cladograms under A1 or A2 but also the less parsimonious area cladograms that it had already derived under A0.

In this paper, we apply the suggested *ad hoc* procedure for derivation of area cladograms from taxon-area cladogram rew515 with CAFCA. However, because CAFCA selects cliques based on the largest number of mutually compatible components before it selects most parsimonious area cladograms, our application of an upper limit appears not to be sufficient to enable CAFCA to meet Requirement I.

Under A0, CAFCA finds two cliques. These cliques correspond to the two area cladograms (eight and nine steps respectively) that we list in Table 1 (no. 1 and 2). CAFCA selects these cliques from the list of components derived under A0 (B, C, D, E, AE, ABE, CDE, ABCDE) by searching for the largest sets of mutually compatible components. Because CAFCA does not derive the components ABCE, ABDE, CD, ACDE, BCDE or AB under A0, both cliques have a basal polytomy.

Under A1, CAFCA derives additional columns for the data matrix by combining either areas A with B or areas B with E. These columns correspond to the additional components AB and BE. From the list of components derived under A1 (B, C, D, E,

AE, AB, BE, ABE, CDE, ABCDE), CAFCA finds one maximal clique with (compared to the two cliques derived under A0) one extra component (AB). This maximal clique (with additional component) has no polytomies and therefore does not correspond to any of the cliques that CAFCA derives under A0. The maximal clique that CAFCA derives under A1 corresponds with a single area cladogram (Table 1; no. 3) that needs nine steps for parsimony mapping of the area-by-node matrix.

Under A2, CAFCA combines either area A or area E with the distributions of all other monophyletic groups in the cladogram in additional columns. The additional columns correspond to additional components AB, BE, CE, AC, AD, ADE, and ACDE. From the list of components derived under A2 (B, C, D, E, AE, AB, BE, CE, AC, AD, ABE, CDE, ADE, ACDE, ABCDE), CAFCA finds nine maximal cliques that are completely dichotomous. The clique derived under A1 is one of these nine cliques. Parsimony mapping of the area-by-node matrix on these nine cliques results in the selection of four most parsimonious area cladograms of nine steps (Table 1; no. 3-6).

From our analysis with CAFCA of taxon-area cladogram rew515, it appears that application of an upper limit is sufficient for obtaining inclusive solution sets only when the area cladograms in the solution sets S0, S1, and S2 are equally resolved, *i.e.*, have the same number of components. If, however, the area cladograms in the solution set obtained under a stricter assumption (*e.g.*, S0 compared to S1, or S1 compared to S2) are not fully resolved, *i.e.*, still contain polytomies, and these area cladograms are more parsimonious than the area cladograms in the other solution set obtained under a less strict assumption, than a lower limit for the number of components constituting the area cladogram is needed. That is to say, if S0 contains area cladograms that are not fully dichotomous -for instance, if they consist of eight

components while a completely resolved area cladogram contains nine components- then under S1 and S2 all area cladograms containing eight or more components should be considered when applying the parsimony criterion to derive the respective solution set. In its present implementation, *i.e.*, when not remedied, CAFCA will discard all area cladograms that are less resolved than the maximal ones allowed by the respective list of components.

We suggest deriving cliques under A1 or A2 by application of a lower limit for clique size. After derivation of the cliques, we suggest selecting area cladograms under A1 or A2 by using the number of steps under A0 as an upper limit. Application of these lower and upper limits on the analysis of rew515 results in the selection of the same area cladogram (Table 1; no. 1) under A0, A1, and A2.

The application of a lower limit clique size and an upper limit for clique length under A1 and A2 with CAFCA results in the finding of inclusive solutions sets for taxon-area cladogram rew515 and the example that we presented in Van Veller *et al.* (2000). Moreover, we claim to have found a general way for finding inclusive solution sets with CAFCA under A0, A1, and A2 by applying these upper and lower limits. Since under A1 and A2 only additional components are extracted, the lists of components extracted with CAFCA under A0, A1, and A2 are inclusive. When the number of components for the cliques derived under A0 is used as a lower limit for the number of components for cliques derived under A1 and A2, inclusion of cliques is obtained. In Van Veller *et al.* (2000) we showed that Requirement I is violated when parsimony mapping of the area-by-node data matrix on the cliques results in shorter area cladograms under A1 (or A2) than under A0 (or A1). Application of an upper limit forces CAFCA not only to select the shortest area cladograms but also to

select the area cladograms that have been derived under a more strict assumption (*e.g.* A0 in comparison to A1).

4.4.2 PAUP (BPA)

We showed in Van Veller *et al.* (2000) that for the theoretical data sets only a few of the analyses of the combinations of a widespread taxon and sympatric taxa result in problems with respect to inclusion of solution sets obtained under A0, A1, and A2 using PAUP as an implementation of (modified) BPA.

Because this violation of Requirement I is caused by selection of more parsimonious cladograms under A1 or A2, in Van Veller *et al.* (2000) we suggested constraining cladogram selection by using an upper limit for cladogram length, as proposed for CAFCA.

In this paper, we analyse the data matrix obtained from taxon-area cladogram *rew515* with PAUP under A0 and find two most parsimonious area cladograms (11 steps), listed in Table 2 (no. 1 and 2). Optimisation of the A1 data matrix on these two area cladograms requires 16 and 15 steps, respectively, and so we use 16 steps as an upper limit for cladogram selection under A1 (for reasons explained in Van Veller *et al.*, 2000). By applying this upper limit for area cladogram selection with parsimony analysis, we find 10 area cladograms under A1 (Table 2; no. 1-10).

If we collapse unsupported nodes during the parsimony analysis with PAUP, three (of the 10) area cladograms will have a polytomy. These less resolved area cladograms are not selected with PAUP under A2 because additional components are derived under this assumption. As a result, the solution set obtained under A1 is not included in the solution set obtained under A2 and Requirement I is violated.

To prevent violation of Requirement I by the derivation of area cladograms with different resolution we suggest not collapsing unsupported nodes in the computer program used for (modified) BPA. This option is available in PAUP but not in Hennig86 (Farris, 1988). If we do not collapse unsupported nodes with PAUP, we obtain 10 completely dichotomous area cladograms under A1. Among these 10 area cladograms, the two area cladograms obtained under A0 are selected.

Optimisation of the A2 data matrix on the 10 area cladograms requires maximally 27 steps. When we perform a parsimony analysis on the A2 data matrix with 27 steps as an upper limit for cladogram selection, we select 46 area cladograms (Table 2; no. 1-46). The 10 area cladograms that we find under A1 are present among these 46 area cladograms as well, and Requirement I is met.

We expect to find inclusive sets of completely dichotomous area cladograms not only for taxon-area cladogram rew515, but in general, by the application of an upper limit to area cladogram selection with PAUP. Since unsupported nodes are not collapsed, cladograms are completely dichotomous. By parsimony analysis of the data matrix, most parsimonious area cladograms are obtained under A0. Under A1 and A2, columns are added (by exporting a binary data matrix with these A1 or A2 columns added by CAFCA) to the data matrix used under A0. Parsimony analysis of these modified (A1 or A2) data matrices can result in the selection of other area cladograms and violation of Requirement I (Van Veller *et al.*, 2000). However, by optimising the A1 (or A2) data matrix on the (completely dichotomous) area cladograms obtained under A0 (or A1) and using the largest number of steps for area cladogram selection under A1 (or A2), parsimony analysis is forced to find the same area cladograms as under A0 (or A1) plus more parsimonious area cladograms.

4.4.3 Component 1.5 (CA)

In Van Veller *et al.* (2000) we showed that for the theoretical data sets Requirement I is violated if sympatric taxa or combinations of widespread and sympatric taxa are dealt with *a priori* by pruning or adding taxa and taxon relationships to the taxon-area cladogram with Component 1.5.

To obtain inclusive solution sets we propose here to deal with widespread and sympatric taxa in a two-step procedure. In the first step, we deal with the widespread taxa under A0, A1 or A2. For this paper, in the first step, we deal with the widespread taxon (T1) in taxon-area cladogram rew515 and give the different places for area E (due to sympatric taxa T1 and T5) different codings (*e.g.* E1 and E2). We deal with the widespread taxon and obtain one, three, and 13 taxon-area cladograms under A0, A1, and A2 respectively. In all these taxon-area cladograms we give the different places for area E (E1 and E2) the same code (plain E). In the second step, we deal with the sympatric taxa “as is”. For the single taxon-area cladogram that we obtain after dealing with the widespread taxon under A0, we find two area cladograms after a parsimony analysis under A0 (Table 3; no. 1 and 2). For the three taxon-area cladograms that we obtain after dealing with the widespread taxon under A1, we find three area cladograms after a parsimony analysis under A0 and removal of duplicate area cladograms (Table 3; no. 1-3). For the 13 taxon-area cladograms that we obtain after dealing with the widespread taxon under A2, we find 10 area cladograms after a parsimony analysis under A0 and removal of duplicate area cladograms (Table 3; no. 1-10).

Both the sets of taxon-area cladograms that we obtain in the first step and the sets of area cladograms that we obtain in the second step show inclusion ($S_0 \subset S_1 \subset S_2$) and therefore Requirement I is met.

We claim that we can find inclusive solution sets not only for taxon-area cladogram rew515, but in general when applying our suggested two-step procedure to Component 1.5 to derive area cladograms under A0, A1, and A2. In the first step, we deal with widespread taxa without yet dealing with sympatric taxa. In Van Veller *et al.* (2000) we have already shown that dealing with widespread taxa under A0, A1, and A2 always results in inclusive solution sets (also in Nelson and Platnick, 1981 and Page, 1990b). Therefore, the sets of taxon-area cladograms that we obtain in the first step are always expected to be inclusive. In the second step, sympatric taxa are dealt with in the same way for all taxon-area cladograms and no *a priori* ignorance of data is assumed. As a result, in accordance with the inclusion of the sets of taxon-area cladograms after the first step, the sets of area cladograms (*i.e.* solution sets) after the second step should be inclusive too.

4.4.4 Component 2.0 (RTA)

In Van Veller *et al.* (2000) we showed that for the theoretical data sets Requirement I is violated when Component 2.0 deals with combinations of widespread and sympatric taxa by tree reconciliations that need a smaller number of losses under A1 (or A2) than under A0 (or A1).

Above, as well as in Van Veller *et al.* (1999), we argued that sympatric taxa should be dealt with *a posteriori* to prevent any inconsistency of the results with the formulation of A1 or *a priori* ignorance of data under A2. However, departure from tree reconciliation is not possible when applying RTA. Therefore we make suggestions here for meeting Requirement I when applying tree reconciliation for dealing with sympatric taxa.

For obtaining inclusive solution sets, we suggest dealing with widespread taxa under A0, A1 or A2. We suggest dealing with sympatric taxa by minimising the number of duplications. For inclusive solution sets, we suggest using the number of losses needed for tree reconciliation after dealing with widespread taxa under A0 (or A1) as an upper limit for further selection of tree reconciliations under A1 (or A2).

For taxon-area cladogram rew515, dealing with the widespread taxon (T1 in area A and E) under A0 results in 13 area cladograms that need one duplication and four losses for dealing with the two sympatric taxa (T1 and T5 in area E) by tree reconciliation (Table 4; no. 1-13). Dealing with the widespread taxon under A1 results in 35 area cladograms with one duplication and four losses or fewer (upper limit for losses is four: Table 4; no. 1-35) for dealing with the two sympatric taxa by tree reconciliation. Dealing with widespread taxa under A2 is not implemented directly in Component 2.0. If an *a priori* modification of the input file for Component 2.0 is applied, only one of the areas of the widespread taxon is mapped while the other area is given the ability to float over the whole cladogram. For an analysis of theoretical taxon-area cladogram rew515 under A2, area E (with sympatric taxa T1 and T5) is given different codings for its different positions in the taxon-area cladogram (*e.g.* E1 and E2). After a first analysis (in which the widespread taxon (T1) is dealt with under A2), taxon-area cladograms with areas at own terminal nodes are obtained. In these cladograms, the areas with sympatric taxa get the same code (plain E). In a second analysis, dealing with the sympatric taxa in area E by tree reconciliation (with one duplication and four losses) results in the same 35 area cladograms as have already been selected under A1. (Table 4; no. 1-35). The first 13 of these 35 area cladograms are selected under A0 as well. As a result, inclusion is obtained ($S_0 \subset S_1 = S_2$) and Requirement I is met.

We expect to find inclusive solution sets not only for taxon-area cladograms rew515, but in general when applying an upper limit when dealing with sympatric taxa by tree reconciliation. In Van Veller *et al.* (2000) we described how Component 2.0 deals with widespread taxa by mapping the taxon cladogram on the taxon-area cladogram and giving each area its own terminal node. This mapping occurs with increasing and inclusive arrangements of the relationships of the areas of the widespread taxa and results in inclusive solution sets for taxon-area cladograms with widespread taxa but without sympatric taxa (Van Veller *et al.*, 2000).

To deal with sympatric taxa, Component 2.0 uses tree reconciliation after potential widespread taxa have been dealt with. Here, we suggest applying an upper limit in the selection of tree reconciliations under A1 (or A2), constrained by the number of losses needed for tree reconciliation after dealing with widespread taxa under A0 (or A1). Application of this upper limit for selection of tree reconciliations under A1 (or A2) forces Component 2.0 to select not only the tree reconciliations that need the smallest number of losses, but also the tree reconciliations that are selected under A0 (or A1). In this way, the sets of tree reconciliations derived under A0, A1 and A2 show inclusion. Because each tree reconciliation corresponds to a unique area cladogram, inclusion of sets of area cladograms is obtained with this inclusion of sets of tree reconciliations.

Component 2.0 does not offer the option of imposing an upper limit for the selection of tree reconciliations. For application of an upper limit (for the number of losses) we first select tree reconciliations by minimising of the number of duplications. For each tree reconciliation we count the number of losses that it needs when it is compared with the taxon-area cladogram. By applying an upper limit on this number of losses, we select the area cladogram(s) (from the tree

reconciliation(s)). A first selection by minimisation of duplications, however, can result in noninclusion when tree reconciliations obtained under A1 (or A2) differ from those obtained under A0 (or A1) because the tree reconciliations need different numbers of duplications. At this moment, however, we do not see another modification to select inclusive sets tree reconciliations with Component 2.0 by applying an upper limit.

4.4.5 TAS (TAS)

In Van Veller *et al.* (2000) we showed that for the theoretical data sets Requirement I is violated when dealing with widespread taxa or combinations of a widespread taxon and sympatric taxa with TAS. When dealing with widespread taxa under A0, TAS recognises three area statements, uniting the areas in which the widespread taxon is distributed. However, these three area statements are recognised by TAS neither under A1 nor under A2. As a result, the completely dichotomous area cladograms (with the areas of the widespread taxon as sister areas) obtained under A0 are not obtained with TAS under A1 and A2. For combinations of a widespread taxon and sympatric taxa, TAS *a priori* prunes taxa and taxon relationships from the taxon-area cladogram under A2, resulting in smaller solution sets and thereby violating Requirement I (Van Veller *et al.*, 2000).

In order to obtain inclusive sets of area cladograms, when TAS deals with widespread taxa for analyses under A1 or A2, we suggest adding a column to the matrix with three area statements that unites the areas for each widespread taxon. For dealing with combinations of widespread and sympatric taxa, we suggest using a two-step procedure as we suggest for Component 1.5. In the first step, one deals with the widespread taxa by deriving of a matrix of three area statements. In the second step,

one deals with sympatric taxa “as is”. To prevent TAS from dealing with sympatric taxa already in the first step, we suggest giving the areas with sympatric taxa separate codings before dealing with the widespread taxa.

We apply the suggested two-step procedure to taxon-area cladogram rew515. First, we give area E different codings (E1 and E2) for the different terminal nodes that it occupies in taxon-area cladogram rew515. Under A0, A1, and A2 we derive matrices with three area statements by applying TAS. However, to the matrices that we derive under A1 and A2 we add a column of “ones” that unite areas A and E1 (*i.e.* the areas of the widespread taxon T1), and question marks for the other areas. After parsimony analysis, we obtain under A0 one completely dichotomous taxon-area cladogram and under A1 and A2 three completely dichotomous taxon-area cladograms. Areas E1 and E2 get the same code back (plain E) and we deal with the sympatric taxa in a parsimony analysis. Under A0 we obtain four area cladograms (Table 5; no. 1-4) and under A1 and A2 we obtain five area cladograms (Table 5; no. 1-5).

Both the sets of taxon-area cladograms that we obtain in the first step and the sets of area cladograms that we obtain in the second step show inclusion ($S_0 \subset S_1 = S_2$) and Requirement I is therefore met.

We claim that applying the suggested two-step procedure to TAS for obtaining area cladograms under A0, A1, and A2 delivers inclusive solution sets not only in the case of taxon-area cladogram rew515 but in general. In Van Veller *et al.* (2000) we showed that noninclusive sets of area cladograms arise when TAS deals with widespread taxa under A1 and A2 and when a parsimony analysis is incapable of finding the completely dichotomous area cladogram obtained under A0. Addition of a column (that unites the areas of a widespread taxon) to the matrix with three area statements in a parsimony analysis forces the optimisation of the questions marks in

such a way that the area cladogram obtained under A0 is obtained under A1 and A2 as well. As a result of the addition of this column, inclusive sets of area cladograms are obtained by parsimony analysis of matrices of three area statements derived with TAS to deal with widespread taxa. Sympatric taxa are always dealt with “as is” when applying TAS. Here we suggest dealing with sympatric taxa in a second step, after widespread taxa have been dealt with in a first step. Since the first step (dealing with widespread taxa) results in inclusive sets of cladograms, and the second step (dealing with sympatric taxa “as is” by a parsimony analysis) is identical under each assumption, inclusive sets of area cladograms should be obtained.

Further, by applying these modifications to TAS, differences between analyses under A1 and A2 disappear completely: the same three area statements, and thus the same solution sets, are obtained under both assumptions.

4.5 Discussion and conclusions

In order to satisfy inclusion of solution sets under assumptions zero, 1, and 2 (Requirement I), we suggest modifications to the computer programs that implement A0, A1, and A2 for various methods used for inferring resolved area cladograms.

We use taxon-area cladogram *rew515* as a benchmark to show that the computer programs for CCA, BPA, CA, RTA, and TAS can be modified to deliver inclusive solution sets under A0, A1, and A2. On the grounds that Requirement I (this paper) and Requirement II (established for these computer programs in Van Veller *et al.*, 2000) are met, we claim (Van Veller *et al.*, 1999) that the computer programs derive valid common patterns (*i.e.* general area cladograms).

4.5.1 *A posteriori* methods

In Van Veller *et al.* (2000) we showed that the computer programs for *a posteriori* methods (CAFCA and PAUP) sometimes obtain noninclusive results when more parsimonious area cladograms are obtained under a less strict assumption (*e.g.* A1 compared to A0). To obtain inclusive solution sets, we suggested (Van Veller *et al.*, 2000) applying an upper limit to the number of steps (for selection of area cladograms) under a less strict assumption (*e.g.* A1). This limit is determined by the number of steps of the most parsimonious area cladogram(s) found under a more strict assumption (*e.g.* A0).

Since CAFCA selects most parsimonious area cladograms from maximal cliques, less resolved area cladograms derived under A0 are not derived under A1 or A2 when CAFCA derives additional components under A1 or A2. To obtain inclusion we therefore suggest (in addition to the upper limit for cladogram length) applying a lower limit for clique size (determined by the number of components for maximal cliques derived under A0) under A1 or A2.

Application of PAUP or Hennig86 for (modified) BPA can result in noninclusive solution sets when these computer programs derive area cladograms under A0 (or A1) that are less resolved than the area cladograms that these computer programs derive under A1 (or A2). To prevent the computer programs from selecting noninclusive solution sets we recommend (in addition to applying the upper limit for cladogram length) not collapsing unsupported nodes. This is implemented in PAUP, but not in Hennig86.

When performing (modified) BPA under A1 or A2, PAUP can select area cladograms that contain either unsupported nodes or nodes that are supported not by the data but by additional components that are the result from the assumption applied.

We think therefore that a more fundamental solution must be sought in a quality assessment of the nodes of the area cladograms before the derivation of solution sets.

For *a posteriori* methods (CCA and BPA) the most parsimonious way to deal with a widespread taxon in an area cladogram is by a single synapomorphy, uniting the areas of the widespread taxon in one monophyletic group. If a most parsimonious area cladogram, obtained under A0, is completely dichotomous, all nodes are supported by data. Any nodes not supported by data collapse to polytomies (see above with CAFCA results for taxon-area cladogram rew515 under A0). When dealing with widespread taxa under A1 or A2 with CAFCA or PAUP, additional components are derived that can result in the selection of additional area cladograms besides the one(s) derived under A0. These additional area cladograms are built from additional components, not supported by the data, and contain unsupported nodes.

Exclusion of the additional area cladograms with unsupported nodes can be performed by leaving the additional components (derived under A1 and A2) out of consideration. By dealing with widespread taxa and sympatric taxa only under A0 with CAFCA and PAUP (also recommended by Zandee and Roos, 1987 and Brooks, 1990), no unsupported nodes are introduced in the area cladograms. Whenever nodes are not supported by the data, CAFCA collapses these nodes to polytomies and indicates that additional data are necessary. With both CCA and BPA, general area cladograms are obtained by combining data sets of several monophyletic groups (obtained under A0) and selection of general area cladograms by a total evidence approach (Zandee and Roos, 1987; Brooks, 1990). Incongruent data are explained *a posteriori* and analyses under A1 and A2 become superfluous.

4.5.2 *A priori* methods

For the computer programs that implement *a priori* methods (Component 1.5, Component 2.0, and TAS) we suggest a two-step procedure in which, first, widespread taxa are dealt with under A0, A1 or A2 and, second, sympatric taxa are dealt with “as is”. For both Component 1.5 and TAS we suggest modifications that ensure that widespread taxa and sympatric taxa are dealt with in this way and that inclusion of solution sets is satisfied. Thereby, differences in the procedure and the results of TAS under A1 and A2 completely disappear. With Component 2.0, however, sympatric taxa cannot be dealt with “as is” since this computer program does not allow any departure from tree reconciliation when dealing with sympatric taxa. Therefore, with this computer program we suggest first dealing with widespread taxa under A0, A1, and A2 and second dealing with sympatric taxa via tree reconciliation. To derive inclusive sets of tree reconciliations, and thereby inclusive solution sets, we suggest using an upper limit for the number of losses.

4.5.3 Comparing *a posteriori* and *a priori* methods

In Table 6 we compare the solution sets obtained under A0, A1, and A2 after the suggested modifications have been applied to the computer programs for CCA, BPA, CA, RTA, and TAS for taxon-area cladogram rew515. The bold numerals in the diagonal row of the table represent the solution sets obtained with the computer programs under A0, A1, and A2. For each computer program the table shows area cladograms in common and thereby inclusion of solution sets (with the exception of CCA when S0 is compared with either S1 or S2) when solution sets obtained under different assumptions for each computer program (*i.e.* the other bold numerals) are compared.

Table 6 also compares solution sets obtained by computer programs for different methods (*e.g.* CAFCA for CCA and Component 2.0 for RTA) under the same assumption. The non-bold numerals represent the area cladograms in common that the computer programs derive under the same assumption. As we show in Table 6, the computer programs for CCA, BPA, CA, RTA, and TAS may vary with respect to the solution sets that they obtain under A0, A1 or A2. For instance, none of the 13 area cladograms obtained with Component 2.0 (RTA) under A0 is obtained with CAFCA (CCA), PAUP (BPA), Component 1.5 (CA) or TAS (TAS) under A0.

We think that this lack of agreement (between the obtained area cladograms) has to be retraced to the *a priori* and *a posteriori* methodologies on which the computer programs are based. The difference between area cladograms obtained with Component 2.0 and those obtained with all other computer programs for taxon-area cladogram rew515 under A0 illustrates such differences in methodology. In order to obtain inclusive solution sets, we modify CAFCA, PAUP, Component 1.5 and TAS in such a way that they deal with sympatric taxa only “as is”. However, a modification of Component 2.0 to deal with sympatric taxa also “as is” is not possible since its method (RTA) does not allow any departure from tree reconciliation. In a next paper we will discuss the lack of agreement in results due to the differences between methodologies.

Acknowledgements

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

References

- Brooks, D.R. (1990). Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Syst. Zool.* **39**, 14-30.
- Enghoff, H. (1996). Widespread taxa, sympatry, dispersal, and an algorithm for resolved area cladograms. *Cladistics* **12**, 349-364.
- Farris, J.S. (1988). Hennig86 vs. 1.5. Distributed by the author. Port Jefferson Station. New York.
- 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.
- Nelson, G. and Ladiges, P.Y. (1991a). TAS Three Area Statements. Program and user's manual. Published by authors. Melbourne, New York.
- Nelson, G. and Ladiges, P.Y. (1991b). Three area statements: standard assumptions for biogeographic analysis. *Syst. Zool.* **40**, 470-485.
- Nelson, G. and Platnick, N.I. (1981). "Systematics and biogeography; cladistics and vicariance". Columbia University Press. New York. 567 pp.

- 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. (1990a). Component 1.5. Program and user's manual. University of Auckland. Auckland.
- Page, R.D.M. (1990b). Component analysis: a valiant failure? *Cladistics* **6**, 119-136.
- Page, R.D.M. (1993). Component 2.0 for Windows. Program and user's manual. Natural History Museum. London.
- Page, R.D.M. (1994). Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* **43**, 58-77.
- 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 different assumptions in vicariance biogeography. *Cladistics* **15**, 393-406.
- Van Veller, M.G.P., Kornet, D.J. and Zandee, M. (2000). Methods in vicariance biogeography: assessment of the implementations of assumptions zero, 1 and 2. *Cladistics* **16**, 319-345.
- 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 and tables

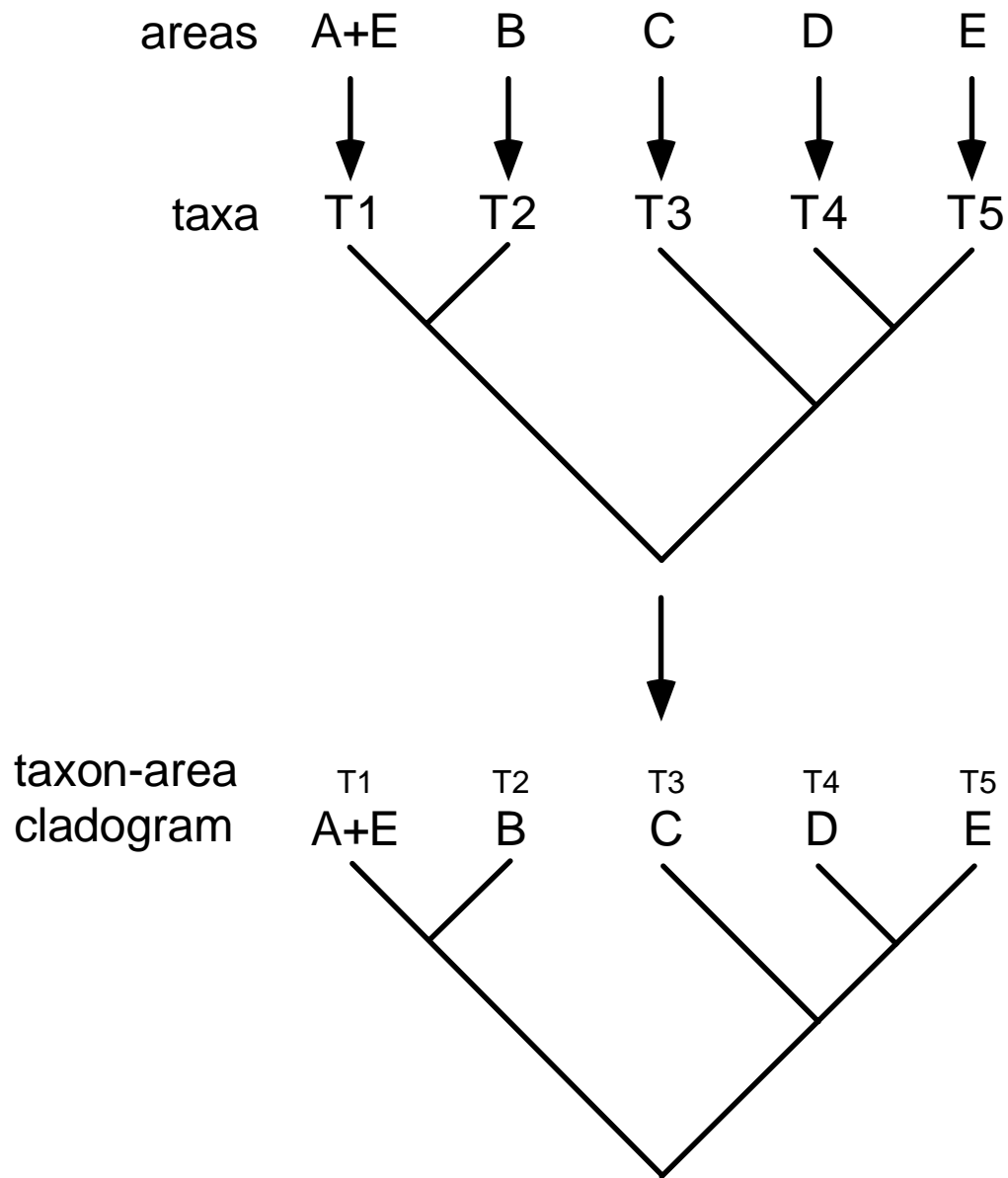


FIG. 1. Theoretical data set rew515 (A, B, C, D, E=areas; T1, T2, T3, T4, T5=taxa)

TABLE 1

Area cladograms obtained with CAFCA under A0, A1 and A2 for taxon-area cladogram rew515 ((A+E)(C(DE))) (no.=cladogram label; *=area cladograms with basal polytomy; +=most parsimonious area cladogram and therefore selected with CAFCA).

no.	widespread taxon dealt with under A0; sympatric taxa dealt with “as is”	# steps
1.	(CD(B(AE)))*+	8
2.	(AB(C(DE)))*	9
no.	widespread taxon dealt with under A1; sympatric taxa dealt with “as is”	# steps
3.	((AB)(C(DE)))+	9
no.	widespread taxon dealt with under A2; sympatric taxa dealt with “as is”	# steps
3.	((AB)(C(DE)))+	9
4.	(B(C(D(AE))))+	9
5.	(B(A(C(DE))))+	9
6.	(B(A(D(CE))))+	9

TABLE 2

Area cladograms obtained with PAUP under A0, A1 and A2 for taxon-area cladogram
rew515 ((A+E)(C(DE))) (no.=cladogram label).

no.	area cladogram A0	# steps	no.	area cladogram A0	# steps
1.	((AB)(C(DE)))	11	2.	(((AE)B)(CD))	11
no.	area cladogram A1	# steps	no.	area cladogram A1	# steps
1.	((AB)(C(DE)))	15	6.	((CD)(A(BE)))	16
2.	(((AE)B)(CD))	16	7.	((AB)(E(CD)))	16
3.	((CD)(E(AB)))	16	8.	(D(C(E(AB))))	16
4.	(C((AB)(DE)))	16	9.	((AB)(D(CE)))	16
5.	(C(D(E(AB))))	16	10.	((DE)(C(AB)))	16
no.	area cladogram A2	# steps	no.	area cladogram A2	# steps
1.	((AB)(C(DE)))	25	24.	((A((CE)D))B)	26
2.	(((AE)B)(CD))	26	25.	(((AE)B)D)C)	27
3.	((CD)(E(AB)))	27	26.	(((AE)B)C)D)	27
4.	(C((AB)(DE)))	26	27.	(((AB)(CE))D)	27
5.	(C(D(E(AB))))	26	28.	(((AB)D)E)C)	27
6.	((CD)(A(BE)))	27	29.	(((AB)D)(CE))	27
7.	((AB)(E(CD)))	27	30.	(((AD)B)(CE))	27
8.	(((AB)C)(DE))	26	31.	((AE)(B(CD)))	27
9.	(((AB)E)C)D)	27	32.	(((AC)B)(DE))	27
10.	((AB)((CE)D))	26	33.	(((AE)D)(BC))	27
11.	(((AE)B)C)D)	27	34.	(((A(BE))D)C)	27
12.	(((AE)D)C)B)	25	35.	(((AD)(BE))C)	27
13.	(((AE)(CD))B)	26	36.	(((AD)C)(BE))	27

14.	$((((AE)C)D)B)$	26	37.	$((AD)((BE)C))$	27
15.	$((((AC)E)D)B)$	27	38.	$((AD)(B(CE)))$	27
16.	$((A(CE))D)B)$	27	39.	$((A(B(DE)))C)$	27
17.	$((((AD)E)B)C)$	27	40.	$((AC)(B(DE)))$	27
18.	$((((AD)E)C)B)$	26	41.	$((A(DE))(BC))$	27
19.	$((((AD)C)E)B)$	27	42.	$(A(B(C(DE))))$	27
20.	$((A(CD))E)B)$	27	43.	$((A(DE))C)B)$	26
21.	$((((AC)D)E)B)$	27	44.	$((AC)(DE))B)$	26
22.	$((AD)(CE))B)$	26	45.	$((A(C(DE)))B)$	26
23.	$((A(DE))B)C)$	26	46.	$((A((CD)E))B)$	26

TABLE 3

Area cladograms obtained with Component 1.5 under A0, A1 and A2 for taxon-area cladogram rew515 ((A+E)(C(DE))) (no.=cladogram label).

no.	widespread taxon dealt with under A0; sympatric taxa dealt with under A0
-----	--

1.	((AB)(C(DE)))
2.	((CD)(E(AB)))

no.	widespread taxon dealt with under A1; sympatric taxa dealt with under A0
-----	--

1.	((AB)(C(DE)))
2.	((CD)(E(AB)))
3.	((CD)(A(BE)))

no.	widespread taxon dealt with under A2; sympatric taxa dealt with under A0
-----	--

1.	((AB)(C(DE)))
2.	((CD)(E(AB)))
3.	((CD)(A(BE)))
4.	((AB)(D(CE)))
5.	(A(B(C(DE))))
6.	(B(A(C(DE))))
7.	(B((AC)(DE)))
8.	(B(C(A(DE))))
9.	(B(C(E(AD))))
10.	(B(C(D(AE))))

TABLE 4

Area cladograms obtained with Component 2.0 under A0, A1 and A2 for taxon-area cladogram rew515 ((A+E)(C(DE))) (no.=cladogram label; *=number of losses under resp. A1 and A2).

no.	area cladogram A0	# losses	no.	area cladogram A0	# losses
1.	(C(B(A(DE))))	4	8.	(C(B(D(AE))))	4
2.	(B(A(C(DE))))	4	9.	(B(C(E(AD))))	4
3.	(C(D(B(AE))))	4	10.	(B(C(D(AE))))	4
4.	(B((AC)(DE)))	4	11.	(C((AE)(BD)))	4
5.	(B(C(A(DE))))	4	12.	((BC)(E(AD)))	4
6.	((BC)(A(DE)))	4	13.	((BC)(D(AE)))	4
7.	(C(B(E(AD))))	4			
no.	area cladogram A1 or A2	# losses	no.	area cladogram A1 or A2	# losses
1.	(C(B(A(DE))))	4	19.	((DE)(B(AC)))	4
2.	(B(A(C(DE))))	4	20.	(A((BC)(DE)))	4
3.	(C(D(B(AE))))	4	21.	(A(B(C(DE))))	4
4.	(B((AC)(DE)))	4	22.	(A(C(B(DE))))	4
5.	(B(C(A(DE))))	4	23.	(C(E(B(AD))))	4
6.	((BC)(A(DE)))	4	24.	(A(C(E(BD))))	4
7.	(C(B(E(AD))))	4	25.	(A(C(D(BE))))	4
8.	(C(B(D(AE))))	4	26.	((AC)(E(BD)))	4
9.	(B(C(E(AD))))	4	27.	((AC)(D(BE)))	4
10.	(B(C(D(AE))))	4	28.	((AC)(B(DE)))	4
11.	(C((AE)(BD)))	4	29.	(C(A(E(BD))))	4
12.	((BC)(E(AD)))	4	30.	(C(A(D(BE))))	4

13.	((BC)(D(AE)))	4	31.	(C(A(B(DE))))	4
14.	((DE)(C(AB)))	3/4*	32.	((DE)(A(BC)))	4
15.	(C((AB)(DE)))	3/4*	33.	(C((BE)(AD)))	4
16.	((AB)(C(DE)))	3/4*	34.	(C(D(A(BE))))	4
17.	(C(E(D(AB))))	3/4*	35.	(C(E(A(BD))))	4
18.	(C(D(E(AB))))	3/4*			

TABLE 5

Area cladograms obtained with TAS under A0 and A1/A2 for taxon-area cladogram

rew515 ((A+E)(C(DE))) (no.=cladogram label).

no.	widespread taxon dealt with under A0; sympatric taxa dealt “as is”	# steps
1.	((CD)(E(AB)))	22
2.	((CD)(B(AE)))	22
3.	((AB(E(CD)))	22
4.	((AB(C(DE)))	22
no.	widespread taxon dealt with under A1 or A2; sympatric taxa dealt “as is”	# steps
1.	((CD)(E(AB)))	22
2.	((CD)(B(AE)))	22
3.	((AB(E(CD)))	22
4.	((AB(C(DE)))	22
5.	((CD(A(BE)))	22

TABLE 6

Differences and similarities in solution sets obtained with (modified) computer programs for CCA (CAFCA), BPA (PAUP), CA (Component 1.5), RTA (Component 2.0) and TAS (TAS) under A0, A1 and A2 for taxon-area cladogram rew515 (inclusive solution sets (S) obtained with computer programs under A0, A1 and A2 represented in **bold**).

		CCA			BPA			CA			RTA			TAS		
		S0	S1	S2	S0	S1	S2	S0	S1	S2	S0	S1	S2	S0	S1	S2
C	S0	2	0	0	0	-	-	0	-	-	0	-	-	0	-	-
C	S1		1	1	-	1	-	-	1	-	-	0	-	-	1	-
A	S2			4	-	-	4	-	-	3	-	-	3	-	-	1
B	S0				2	2	2	2	-	-	0	-	-	2	-	-
P	S1					10	10	-	3	-	-	4	-	-	4	-
A	S2						46	-	-	9	-	-	19	-	-	5
C	S0							2	2	2	0	-	-	2	-	-
A	S1								3	3	-	0	-	-	3	-
	S2									10	-	-	7	-	-	3
R	S0										13	13	13	0	-	-
T	S1											35	35	-	1	-
A	S2												35	-	-	1
T	S0													4	4	4
A	S1														5	5
S	S2															5