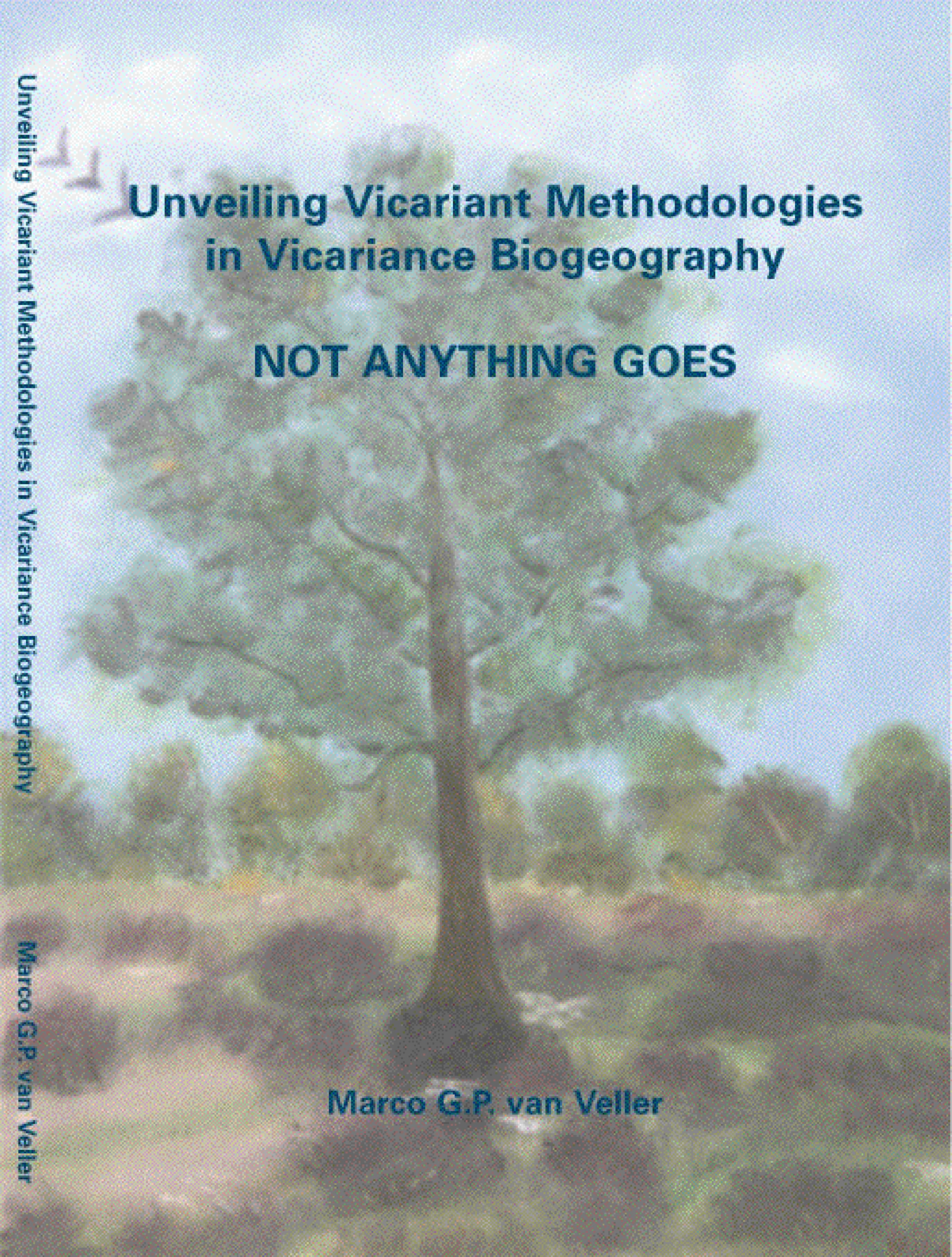
Unveiling vicariant methodologies in vicariance biogeography : not anything goes

Veller, M.G.P. van

Citation

Veller, M. G. P. van. (2000, November 29). *Unveiling vicariant methodologies in vicariance biogeography : not anything goes*. Retrieved from https://hdl.handle.net/1887/12439

Version: Corrected Publisher’s Version



Unveiling Vicariant Methodologies in Vicariance Biogeography

NOT ANYTHING GOES

Marco G.P. van Veller

Unveiling Vicariant Methodologies in Vicariance Biogeography Marco G.P. van Veller

Unveiling Vicariant Methodologies in Vicariance Biogeography

NOT ANYTHING GOES

Marco G.P. van Veller

Illustratie omslag: Rita van Veller-Keizer

Veller, Marco G.P. van
Unveiling vicariant methodologies in vicariance biogeography.
Not anything goes.
Thesis Universiteit Leiden – With ref. – Summary in Dutch
ISBN 90-9014087-5
NUGI 821

Unveiling Vicariant Methodologies in Vicariance Biogeography

NOT ANYTHING GOES

PROEFSCHRIFT

ter verkrijging van de graad van Doctor
aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. W.A. Wagenaar,
hoogleraar in de faculteit der Sociale Wetenschappen
volgens besluit van het College voor Promoties
te verdedigen op woensdag 29 november 2000
te klokke 14.15 uur

door
MARCO GERARDUS PETRUS VAN VELLER
geboren te Gouda in 1970

Promotiecommissie:

Promotor: Prof. dr. D.J. Kornet
Co-promotor: Dr. M. Zandee
Referent: Prof. dr. F. Ronquist (Uppsala Universitet)
Overige leden: Prof. dr. P. Baas
Prof. dr. D.R. Brooks (University of Toronto)
Prof. dr. J.L. Olsen (Rijksuniversiteit Groningen)
Prof. dr. F.R. Schram (Universiteit van Amsterdam)

Het onderzoek beschreven in dit proefschrift werd bekostigd door het gebied Aard- en Levenswetenschappen (ALW) van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), project 805-33.193.

Aan mijn ouders

Table of contents

1. General introduction 9
   - 1.1 A survey of the scientific discipline of biogeography 11
   - 1.2 Evaluation of *a priori* and *a posteriori* methods in vicariance biogeography 13
   - 1.3 The future of *a posteriori* methods 14

2. Two requirements for obtaining valid common patterns under different assumptions in vicariance biogeography 19
   - 2.1 Introduction 21
   - 2.2 Dealing with widespread or sympatric taxa 23
     - 2.2.1 Assumption zero 24
     - 2.2.2 Assumption 1 26
     - 2.2.3 Assumption 2 29
   - 2.3 Two requirements for obtaining valid common patterns 32
   - 2.4 The increase in explanatory power 35
   - 2.5 Violating requirements: examples from the literature 36
   - 2.6 Conclusions 40

3. Methods in vicariance biogeography: assessment of the implementations of Assumptions zero, 1, and 2 43
   - 3.1 Introduction 45
   - 3.2 Implementations of A0, A1, and A2 in methods for vicariance biogeography 46
     - 3.2.1 Component Compatibility Analysis (CCA) 47
     - 3.2.2 Brooks Parsimony Analysis (BPA) 48
     - 3.2.3 Component Analysis (CA) 48
     - 3.2.4 Reconciled Tree Analysis (RTA) 50
     - 3.2.5 Three Area Statement Analysis (TAS) 53
   - 3.3 Assessment of Requirements I and II in implementations of methods 54
     - 3.3.1 Theoretical and empirical data sets 54
     - 3.3.2 CAFCA (Component Compatibility Analysis) 54
     - 3.3.3 PAUP (Brooks Parsimony Analysis) 59
     - 3.3.4 Component 1.5 (Component Analysis) 64
     - 3.3.5 Component 2.0 (Reconciled Tree Analysis) 70
     - 3.3.6 TAS (Three Area Statement Analysis) 73
   - 3.4 Conclusions 76

4. Measures for obtaining inclusive solution sets under Assumptions zero, 1, and 2 with different methods for vicariance biogeography 89
   - 4.1 Introduction 91
   - 4.2 How to deal with widespread and sympatric taxa to obtain inclusion 91

4.3 Taxon-area cladogram rew515 92
4.4 Methods, programs, and modifications 94
  4.4.1 CAFCA (CCA) 94
  4.4.2 PAUP (BPA) 96
  4.4.3 Component 1.5 (CA) 98
  4.4.4 Component 2.0 (RTA) 99
  4.4.5 TAS (TAS) 102
4.5 Discussion and conclusions 104
  4.5.1 *A posteriori* methods 104
  4.5.2 *A priori* methods 105
  4.5.3 Comparing *a posteriori* and *a priori* methods 105

5. *A posteriori* and *a priori* methodologies for testing hypotheses of causal processes in vicariance biogeography 109
5.1 Introduction 111
5.2 The *a posteriori* methodology for vicariance biogeography 114
  5.2.1 Process assumptions 114
  5.2.2 Formulating the null hypothesis 114
  5.2.3 Testing the null hypothesis 115
  5.2.4 Why A0 is sufficient for *a posteriori* methods and Requirements I and II therefore do not apply 115
5.3 The *a priori* methodology for vicariance biogeography 117
  5.3.1 Process assumptions 117
  5.3.2 Formulating and testing hypotheses 117
  5.3.3 Why and when Requirements I and II apply in the *a priori* methodology 119
5.4 Conclusions 121

Nederlandse samenvatting 127

Nawoord 143

Curriculum Vitae 145

CHAPTER 1

GENERAL INTRODUCTION

1.1 A survey of the scientific discipline of biogeography

Biogeography is the scientific discipline in which one tries to explain the distribution of groups of organisms, or taxa, over the surface of the earth. Depending on whether these distributions are explained on a short-term ecological or on a long-term evolutionary scale, the discipline of biogeography is subdivided into ecological and historical biogeography (Myers and Giller, 1988). The structure of the discipline of biogeography is summarized in Fig. 1.

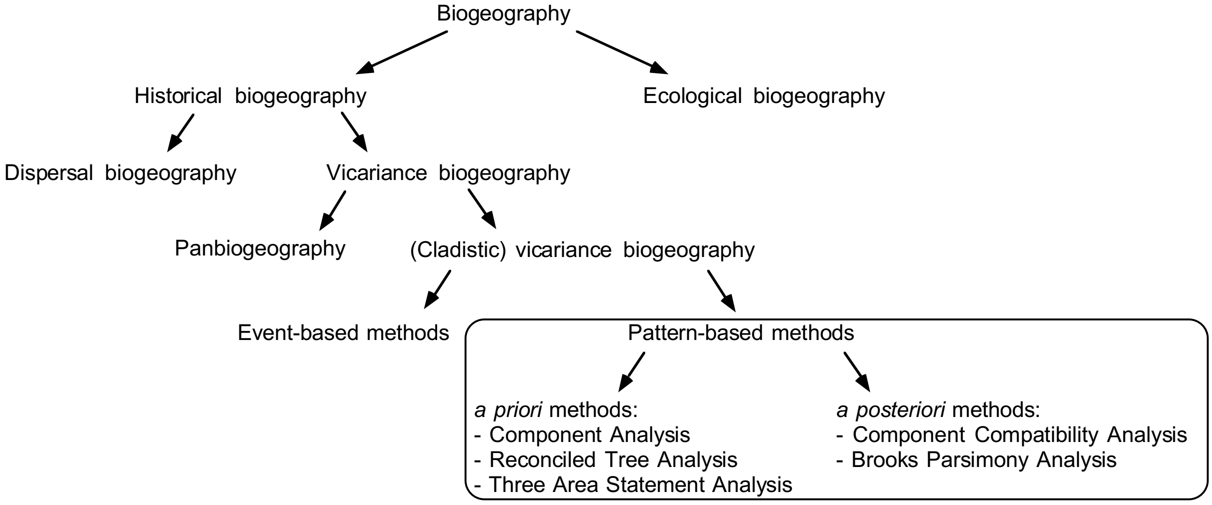


Fig. 1. Outline of the scientific discipline of biogeography.

Ecological biogeography examines the distribution of taxa to find out why they are at present restricted to that particular distribution (Nelson and Platnick, 1981). In historical biogeography, by contrast, the present-day distribution of taxa is combined with hypotheses on cladogenetic relationships between these taxa to infer historical connections among the areas or biota[*] over which they are distributed (Rosen, 1978) and to explain how the taxa became distributed over these areas by processes including allopatric speciation, sympatric speciation, extinction, and dispersal.

Within historical biogeography, dispersal biogeography explains disjunct distribution ranges of taxa by the dispersal of ancestor taxa over pre-existing barriers that originated as a result of (a)biotic events. If sympatric speciation takes place after dispersal, descendant taxa are distributed over different areas separated by the (a)biotic barrier. Dispersal biogeographers start with a centre of origin for ancestor taxa and explain the distribution of descendant taxa by dispersal across pre-existing barriers (Myers and Giller, 1988; Bremer, 1992; Ronquist, 1994; Hausdorf, 1998).

Vicariance biogeography, by contrast, explains disjunct distribution ranges of taxa by the process of vicariance, *i.e.* the process whereby allopatric speciation is triggered by an abiotic or a biotic event (Myers and Giller, 1988). As a result of such an event an ancestor taxon is divided into two (or more)

---

[*] In this thesis only areas will be considered although also biota can be studied in historical biogeography.

disjunct populations that differentiate into two (or more) allopatric descendant taxa (Nelson and Platnick, 1981).

Despite the fact that historical biogeography falls into two sub-disciplines, neither the process of dispersal nor the process of vicariance can be discounted *a priori* as irrelevant for the explanation of the distribution of any particular group of taxa.

Widespread and sympatric taxa indicate that not all biogeographical patterns result from vicariance alone, and suggest that dispersal takes place too. On the other hand, endemic taxa suggest vicariance as an explanation for biogeographical patterns. Though vicariance biogeography *a priori* assumes vicariance, the methodologies of its various approaches also include explanations in terms of dispersal.

The discipline of vicariance biogeography can be divided into two different approaches depending on how distribution histories are reconstructed.

Panbiogeography, founded by Croizat (1952, 1958), reconstructs distribution histories by drawing lines on a geographical map. These lines, called “tracks”, connect known distributions of related taxa in different areas. If two or more tracks of unrelated taxa coincide, they are called generalized tracks. These generalized tracks indicate the preexistence of widespread distributions of ancestral taxa that are subsequently fragmented by (a)biotic events (Morrone and Crisci, 1995). In panbiogeography, the cladogenetic relationships of the taxa from which distribution the tracks are drawn play a subordinate role.

The role of cladogenetic relationships between taxa is more important in the form of vicariance biogeography that dominates the disciple at present: cladistic vicariance biogeography. This approach combines cladograms of taxa and their distributions in areas to derive area cladograms. Such area cladograms are hypotheses of historical relationships between areas. Explanations of the distribution of taxa over areas by the processes of vicariance, extinction, and dispersal are obtained from these area cladograms by the application of various methods. In this thesis, this approach will be referred to as vicariance biogeography.

Two approaches can be distinguished within vicariance biogeography. The first set of approaches are methods that derive divergent patterns (*i.e.* area cladograms) on the basis of the assumption of vicariance. These methods are called biogeographic pattern analysis methods (Cracraft, 1988) or pattern-based methods (Ronquist and Nylin, 1990). This thesis concerns an evaluation of such pattern-based methods (see box, Fig. 1). The second set of approaches are event-based methods (Ronquist and Nylin, 1990) and are proposed by Ronquist (1997, 1998) and Charleston (1996) to accommodate reticulate biogeographical scenarios, representing the accretion of areas. In event-based methods, reticulate biogeographical scenarios are obtained by assigning differential costs to different processes (vicariance, dispersal, extinction, and sympatric speciation). An event-based protocol that reveals historical sequences of vicariance events (and not historical relationships among areas) has been developed by Hovenkamp (1997).

1.2 Evaluation of *a priori* and *a posteriori* methods in vicariance biogeography

The research reported in this thesis deals with the pattern-based methods in vicariance biogeography. All the pattern-based methods derive divergent (general) area cladograms that hypothesize historical relationships between areas based on cladogenetic and distributional data of taxa that inhabit these areas. In order to explain the distribution of the taxa, explanations by the processes of vicariance, extinction or dispersal are inferred from the data and the (general) area cladogram(s) obtained.

According to the way in which the pattern-based methods deal with the data in order to obtain explanations for distributions of taxa by the processes, two categories of methods are recognized in this thesis: *a priori* and *a posteriori* methods. *A priori* methods include Component Analysis (CA: Nelson and Platnick, 1981; Page, 1988, 1990), Reconciled Tree Analysis (RTA: Page, 1993, 1994), and Three Area Statement Analysis (TAS: Nelson and Ladiges, 1991a,b,c). *A posteriori* methods include Component Compatibility Analysis (CCA: Zandee and Roos, 1987) and Brooks Parsimony Analysis (BPA: Brooks, 1990; Wiley, 1988a,b). Each of these methods has its proponents (*e.g.* Andersson, 1996; Page, 1989, 1990; Wiley, 1988a,b; Zandee and Roos, 1987; Ladiges, 1998). Several studies (Cracraft, 1988; Crisci *et al.*, 1991; Enghoff, 1995; Morrone and Carpenter, 1994; Van Soest and Hajdu, 1997) have shown that the *a priori* and *a posteriori* methods in vicariance biogeography may infer different (general) area cladograms from identical empirical or theoretical data sets.

The aim of this thesis is to compare the *a priori* and *a posteriori* methods for their protocol of deriving (general) area cladograms and their explanation of taxon distributions. Further, the research aims to evaluate the methods for their methodological requirements and assumptions in order to gain insight into the disagreement of the outcomes of the different methods.

In an earlier evaluation of different methods used in vicariance biogeography, Morrone and Carpenter (1994) compared area cladograms obtained by CA, RTA, TAS, and BPA for different data sets. They found that the methods often do not agree in their results and concluded that agreement among the obtained area cladograms cannot serve as a criterion for prefering one of the methods over another. Furthermore, they found that even their criterion of counting the number of so-called “items of error” (Nelson and Platnick, 1981) for the area cladograms obtained with the various methods failed to identify any of the methods as preferable. Their use of items of error as an evaluation criterion, however, is questionable, because items of error may serve in CA and RTA as an optimality criterion for the selection of area cladograms. As a result, the criterion for the evaluation of methods used by Morrone and Carpenter (1994) was not independent of the optimality criterion within methods. The methods use different inputs from the same data set for the derivation of area cladograms. An optimality criterion such as items of error (used with CA and RTA) or number of steps (used with TAS and BPA) cannot be used for comparison between all methods because it is restricted to a particular method.

In chapter 2 of this thesis, the need for two methodological requirements (Requirements I and II) for the methods in vicariance biogeography is established and explained. These requirements concern the use of processes assumed *a priori* under Assumptions zero (A0: vicariance), 1 (A1: vicariance + extinction), and 2 (A2: vicariance + extinction + dispersal). The applicability of the assumptions to all examined methods allows the generalization of the requirements over all the methods.

The *a priori* and *a posteriori* methods are implemented in different computer programs. In chapter 3 it is first discussed how CCA, BPA, CA, RTA ,and TAS (and their implementations in software) obtain area cladograms from cladogenetic and distribution data of the taxa of monophyletic groups under the different assumptions. Second, the implementations of the methods are used to derive sets of area cladograms *(i.e.* solution sets) for both theoretical and empirical data sets (*Heterandria*, *Xiphophorus*, *Cyttaria*, *Eriococcus*/ *Madarococcus*) under A0, A1, and A2. The results are examined to assess whether Requirements I and II are met by *a priori* and *a posteriori* methods.

Since the evaluation of methods performed in chapter 3 shows that Requirement II is never violated, no modifications to the methods are necessary to meet this requirement. All the methods, however, may under certain circumstances violate Requirement I. In chapter 4 remedial measures are suggested which ensure that each of the *a priori* and *a posteriori* methods meets Requirement I.

By the conclusion of chapter 4 it will appear that, even if all methods meet Requirements I and II, *a priori* and *a posteriori* methods do not always deliver the same general area cladograms. The reason is that the two categories of methods represent different research methodologies in the testing of hypothesis concerning the causal processes that yield the distribution of taxa over areas. In chapter 5 the formulation and testing of null hypotheses for both *a priori* and *a posteriori* methods is described. There, it is evaluated whether Requirements I and II are necessary for both *a priori* and *a posteriori* methods. It is established that *a posteriori* methods are preferable because they need less methodological requirements.

1.3 The future of *a posteriori* methods

The research described in the five chapters of this thesis leads us to attach a methodological preference to *a posteriori* methods. The two methods in this category, BPA and CCA, code cladogenetic and distributional data of taxa of monophyletic groups in one area-data matrix that they use as input in a parsimony analysis.

Whereas the parsimony analysis in CCA is profitably constrained by the components and cliques that are extracted from the data matrix (Zandee, 1999), BPA uses only a standard (unconstrained) parsimony algorithm as implemented in PAUP (Swofford, 1990) or Hennig86 (Farris, 1988). As a result, BPA sometimes finds more parsimonious (general) area cladograms than CCA, which, however, may contain nodes that are supported only by homoplaseous components (*i.e.* reversals or parallelisms). This has already been reported by Van Welzen (1990). If one requires that all nodes in (general) area cladograms are supported by at least one non-homoplaseous component, this may count as a reason for choosing the less parsimonious cladograms obtained by constrained parsimony analysis (as in CCA) rather than the more parsimonious cladograms obtained via standard parsimony analysis (as in BPA).

Differences in (general) area cladograms obtained via CCA or BPA are possibly caused by their different coding of the cladogenetic relationships of the taxa distributed over the areas. BPA uses additive binary coding of the inner nodes of the taxon cladogram to obtain an area-data matrix with a separate column for each inner node. By contrast, CCA represents these columns 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. Further research and agreement on the coding and optimization of the inner nodes of taxon cladograms that are used as input in biogeographical analyses with CCA and BPA are expected to make these methods more similar in their results.

Pattern-based methods (the *a posteriori* as well as the *a priori* methods) can be used only to reconstruct divergent patterns. However, areas should be expected not only to break up but also to collide, yielding reticulate patterns (Cracraft, 1988; Craw, 1982, 1983; Hovenkamp, 1997). At present, pattern-based methods are unable to represent such reticulate patterns in (area) cladograms. Future research might result in modifications to *a posteriori* methods (CCA or BPA) to enable them to infer reticulation events and to explain these events in biogeographic scenarios.

Further comments on pattern-based methods have been provided by Hovenkamp (1997) and Ronquist and Nylin (1990). Hovenkamp criticizes pattern-based methods (such as BPA, TAS, and CA) for their assumption of an analogy between areas (and taxa) in a historical biogeographic analysis and taxa (and characters) in a phylogenetic systematic analysis. He argues not only that the history of areas is rarely exclusively divergent but also that taxa are less reliable indicators for the history of areas because of (putative) dispersal. Though the assumed analogy of these methods in vicariance biogeography might not always hold, this is not a reason to eliminate this analogy *a priori*. In the pattern-based methods, the analogy is a starting point that is tested in the analysis. Whenever the analogy does not hold, alternative explanations are provided.

Ronquist and Nylin (1990) discuss four problems with the use of pattern-based (parsimony mapping) methods. First, they state that it is not sufficient to investigate the congruence between the phylogenies of a single host and a single parasite group to test for coevolutionary process models. This practise can be considered analogous to the derivation of an area cladogram from data provided by a single monophyletic group of taxa in vicariance biogeography. But, just as in coevolutionary studies, several monophyletic groups of taxa are necessary to obtain a general area cladogram. In this general area cladogram, common speciation events for the several monophyletic groups that are triggered by a single (a)biotic event are explained as vicariance events. The absence of taxa of a monophyletic group in any of the areas studied (the problem of missing taxa) is explained *a posteriori* by optimization of its distributional and cladogenetic data on the general area cladogram.

Second, Ronquist and Nylin (1990) discuss the assumption of allopatric cospeciation as the null model in coevolutionary studies. This assumption is analogous to the assumption of vicariance as the null model in vicariance biogeography. It may happen that taxa of a particular monophyletic group failed to speciate when vicariance took place. However, this is not a problem for the methodology of pattern-based methods, as such methods use data from several monophyletic groups to derive one or more general area cladogram(s). The vicariance event (that did not result in speciation for a particular monophyletic group) is inferred from the general area cladogram when other monophyletic groups speciated as a result of the vicariance event.

The third problem for pattern-based methods that Ronquist and Nylin discuss corresponds to the problem of the interpretation of (general) area cladograms in vicariance biogeography. The *a posteriori* methods CCA and BPA sometimes differ in their results because of their different coding of the cladogenetic relationships of the taxa distributed over the areas. As mentioned above, however, future research is expected to result in greater agreement on coding and optimization of the data and to result in less ambiguous interpretations of the results of CCA or BPA.

Ronquist and Nylin's fourth problem with pattern-based methods corresponds to the treatment of widespread taxa that occur in more than one area. To explain these distributions they need a general method that *a priori* assumes not only vicariance but also dispersal. This idea inspired Ronquist (1996) and Charleston (1996) to develop new methods such as Dispersal-Vicariance Analysis (DIVA: Ronquist, 1996) and Jungles (Charleston, 1996). However, all these methods use *a priori* assigned costs to processes such as vicariance, extinction, dispersal or sympatric speciation in order to arrive at hypotheses that represent histories of areas and process explanations for the distribution of taxa in these areas. Such *a priori* assigned costs make these methods immune to test the relative importance of the different explanatory processes.

Future research might result in a further development of *a posteriori* methods (such as CCA or BPA) that can deal with divergent and reticulate patterns without assigning any costs to particular processes (vicariance, extinction, dispersal or sympatric speciation) *a priori*.

References

Andersson, L. (1996). An ontological dilemma, epistemology and methodology of historical biogeography. *Journal of Biogeography* **23**, 269-277.

Bremer, K. (1992). Ancestral areas: a cladistic reinterpretation of the center of origin concept. *Syst. Biol.* **4**, 436-445.

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

Charleston, M.A. (1996). Jungles: a new solution to the host/parasite phylogeny reconcilation problem. *Math. Bio. Sci.* **149**, 191-223.

Cracraft, J. (1988). Deep-history biogeography: retrieving the historical pattern of evolving continental biotas. *Syst. Zool*. **37**, 221-236.

Craw, R.C. (1982). Phylogenetics, areas, geology and the biogeography of Croizat: a radical view. *Syst. Zool.* **31**, 304-316.

Craw, R.C. (1983). Panbiogeography and vicariance cladistics: are they truly different? *Syst. Zool.* **32**, 431-438.

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

Croizat, L. (1952). “Manual of Phytogeography”. Junk. The Hague.

Croizat, L. (1958). “Panbiogeography. Vol. 1, 2a, 2b”. Published by author. Caracas, Venezuela.

Enghoff, H. (1995). Historical Biogeography of the Holarctic: area relationships, ancestral areas, and dispersal of non-marine animals. *Cladistics* **11**, 223-263.

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

Hausdorf, B. (1998). Weighted ancestral area analysis and a solution of the redundant distribution problem. *Syst. Biol.* **47**, 445-456.

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

Ladiges, P.Y. (1998). Biogeography after Burbidge. *Austral*. *Syst*. *Bot*.**11**, 231-242

Morrone, J.J. and Carpenter, J.M. (1994). In search of a method for cladistic biogeography: an empirical comparison of Component Analysis, Brooks Parsimony Analysis, and Three area statements. *Cladistics* **10**, 99-153.

Morrone, J.J. and Crisci, J.V. (1995). Historical Biogeography: introduction to methods. *Ann. Rev. Ecol. Syst.* **26**, 373-401.

Myers, A.A. and Giller, P.S. (1988). Process, pattern and scale in biogeography. In: “Analytical biogeography: an integrated approach to the study of animal and plant distributions”. Myers, A.A. and Giller, P.S. (eds.). Chapman and Hall. London, New York. pp. 3-10.

Nelson, G. and Ladiges, P.Y. (1991a). Standard assumptions for biogeographic analysis*.* *Austral. Syst. Bot.* **4**, 41-58.

Nelson, G. and Ladiges, P.Y. (1991b). Three area statements: standard assumptions for biogeographic analysis. *Syst. Zool.* **40**, 470-485.

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 Univ. Press. New York.

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

Page, R.D.M. (1989). Comments on component-compatibility in historical biogeography. *Cladistics* **5**,167-182.

Page, R.D.M. (1990). Component analysis: a valiant failure? *Cladistics* **6**, 119-136.

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

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

Ronquist, F. (1994). Ancestral areas and parsimony. *Syst. Biol.* **43**, 267-274.

Ronquist, F. (1996). DIVA. Program and user's manual. Department of Zoology, University of Uppsala. Uppsala.

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst*. *Biol*. **46**, 195-203.

Ronquist, F. (1998). Three-dimensional cost-matrix optimization and maximum cospeciation. *Cladistics* **14**, 167-172.

Ronquist, F. and Nylin, S. (1990). Process and pattern in the evolution of species associations. *Syst. Zool.* **39**: 323-344.

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 Soest, R.W.M. and Hajdu, E. (1997). Marine area relationships from twenty sponge phylogenies. A comparison of methods and coding strategies. *Cladistics* **13**, 1-20.

Van Welzen, P.C. (1990). “Guioa Cav. (Sapindaceae): taxonomy, phylogeny, and historical biogeography”. PhD Thesis. Rijksherbarium/Hortus Botanicus. Leiden.

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.

CHAPTER 2

TWO REQUIREMENTS FOR OBTAINING VALID COMMON PATTERNS UNDER DIFFERENT ASSUMPTIONS IN VICARIANCE BIOGEOGRAPHY

Abstract

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

This chapter has been published in a modified form as: Veller, M.G.P. van, Zandee, M. & Kornet, D.J. (1999). Two requirements for obtaining valid common patterns under different assumptions in vicariance biogeography. *Cladistics* **15**, 393-406.

2.1 Introduction

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

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

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

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

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

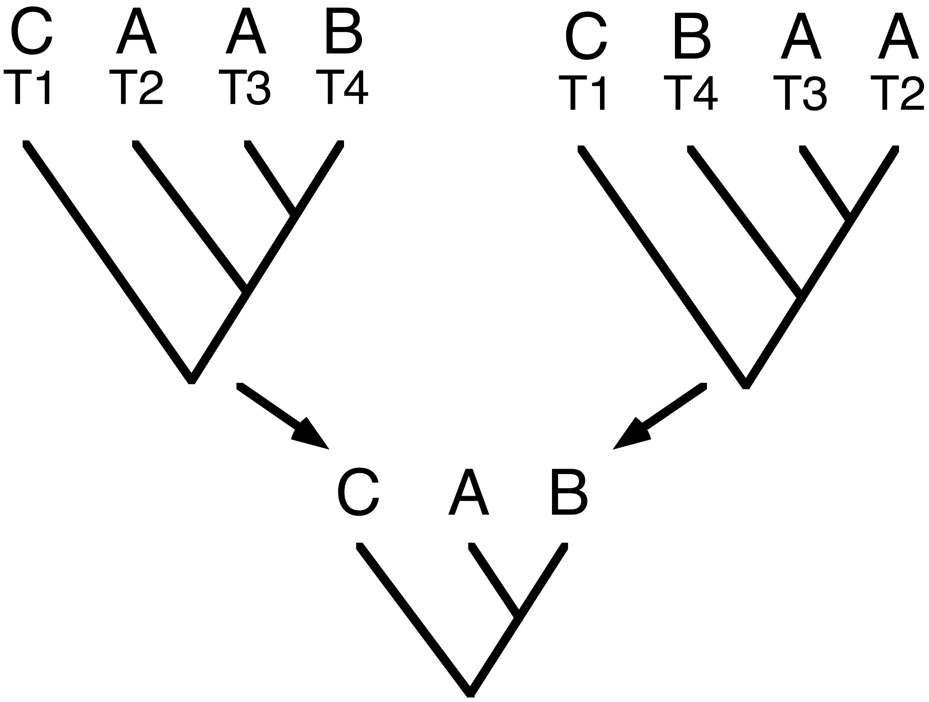


Fig. 1. Solution of redundant distributions (two species in area A) that are the result of sympatric speciation (species T2 and T3).

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

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

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

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

In this chapter we show that in order to explain all data, one should solve redundancy (*i.e.* sympatric taxa) only “as is”, and we argue that absence or multiple presence of taxa in the area cladogram caused by extinction or dispersal should be accounted for only *a posteriori*, contrary to, *e.g.* Page (1988, 1990) and Nelson and Platnick (1981).

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

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

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

2.2 Dealing with widespread or sympatric taxa

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

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

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

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

2.2.1 Assumption zero

Zandee and Roos (1987) and Wiley (1988a) introduced A0. Under this assumption, these authors interpret widespread taxa as “synapomorphies” of the areas in which they occur, uniting these areas into one component. They do not *a priori* assume extinction or dispersal to have taken place. The widespread distribution of the taxon is considered the result of isolation or break-up of areas without triggering speciation yet. Analysis results in a single area cladogram for a four area case with one widespread taxon (T3) in two areas (Fig. 2b). According to these authors, under A0, redundancy should be interpreted “as is”.

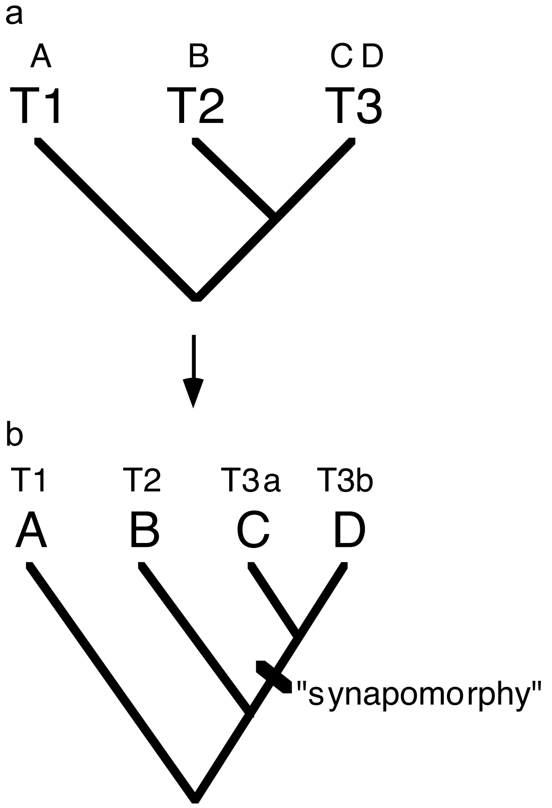


Fig. 2. Solution of a widespread distribution of taxon T3 in areas C and D under A0 (a: taxon cladogram with areas, b: area cladogram under A0 with a “synapomorphy”).

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

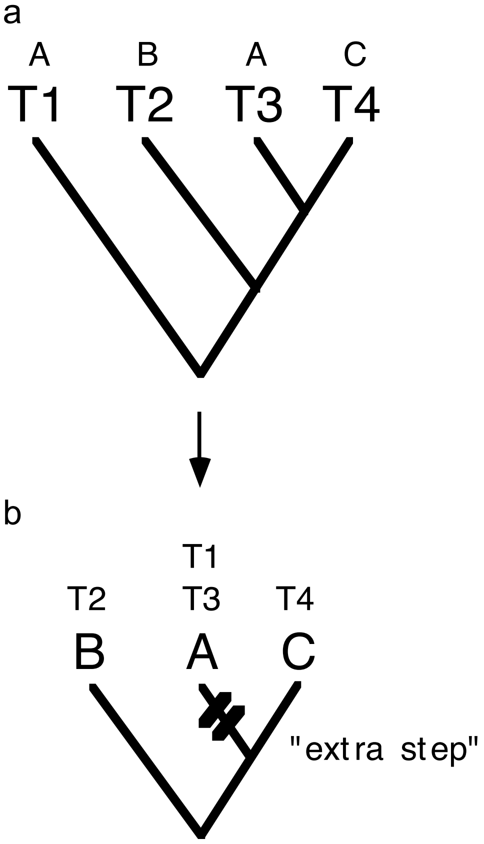


Fig. 3. Solution of a redundant distribution of both taxon T1 and T3 in area A under A0 (a: taxon cladogram with areas, b: area cladogram under A0 with an “extra step”).

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

2.2.2 Assumption 1

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

For widespread taxa, Humphries (1982, pp. 453,454) interpreted A1 to pertain to taxon relationships: "the implications are that under Assumption 1 the taxon occupying area CD [*i.e.* a taxon widespread in both areas C and D] will never be split into separate taxa". Humphries and Parenti (1986) and Zandee and Roos (1987) use the same taxon relationship approach to deal with distributions of widespread taxa. On the other hand, Page (1989, p. 167) advocates that a crucial distinction must be drawn between the relationships of taxa and the relationships of areas. According to Page A0, A1, and A2 are interpretations of the relationship between areas, not between taxa. Platnick (1988) also maintains this area relationship approach.

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

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

With respect to redundancy, under A1 Nelson and Platnick (1981, p. 456) consider all sympatric taxa to be informative for historical relationships between areas. In contrast with Assumption zero, sympatry is not solved *a posteriori*, but *a priori* processes deviating from vicariance are assumed to explain the presence of more than one taxon of a monophyletic group in the same area (*i.e.* redundant distribution). Nelson and Platnick (1981) explain the areas with redundant distributions (*e.g.* area A in Fig. 5a) by *a priori* assuming extinction of (hypothetical) taxa (in addition to vicariance) in areas without redundant distributions ("extinctions" in areas B and C in Fig. 5b). Thereby, they assume an area with redundant distribution to be the remainder of a larger pattern (Fig. 5b).

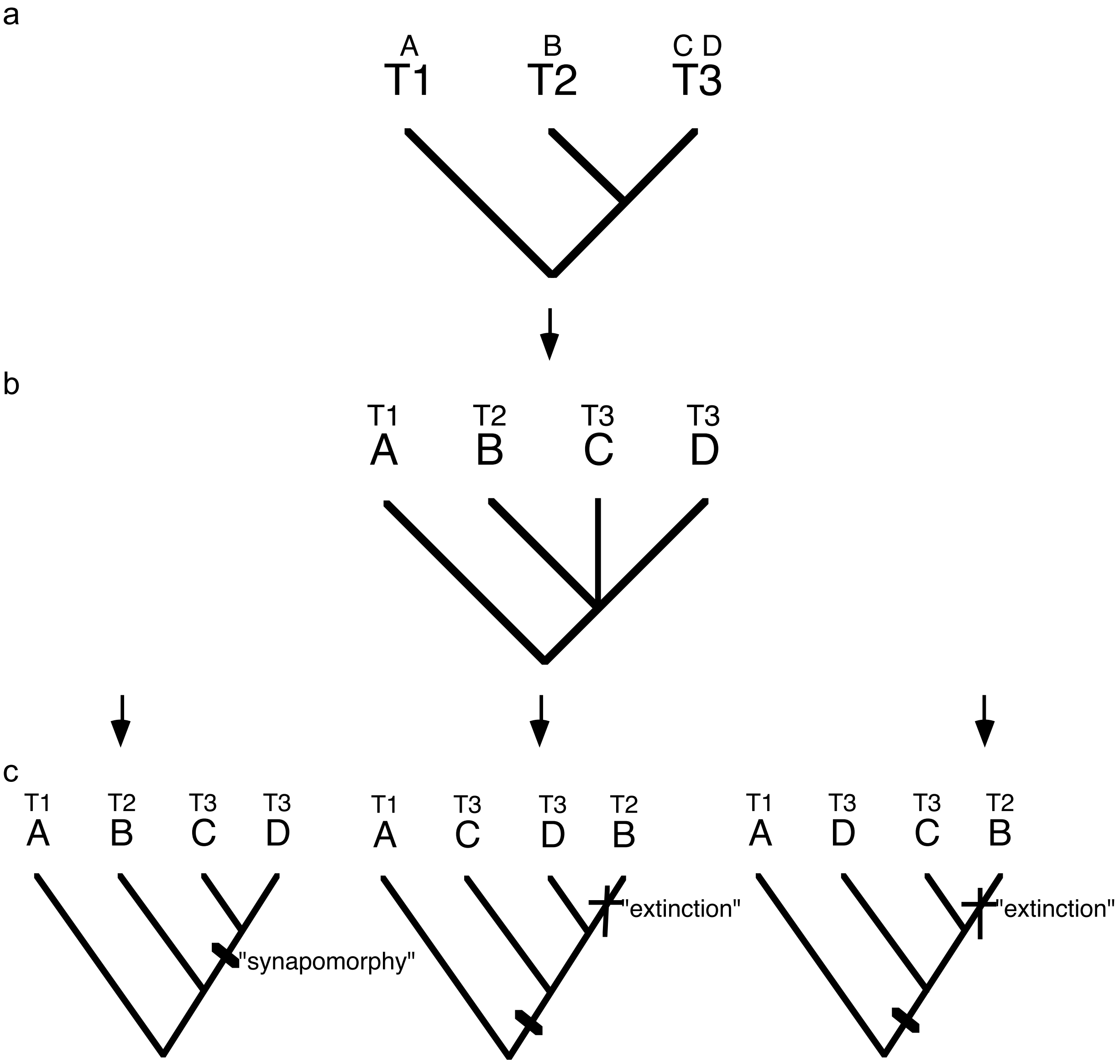


Fig. 4. Solution of a widespread distribution of taxon T3 in areas C and D under A1 (a: taxon cladogram with areas, b: area cladogram with a polytomy, c: area cladograms under A1 with a "synapomorphy" or "extinctions").

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

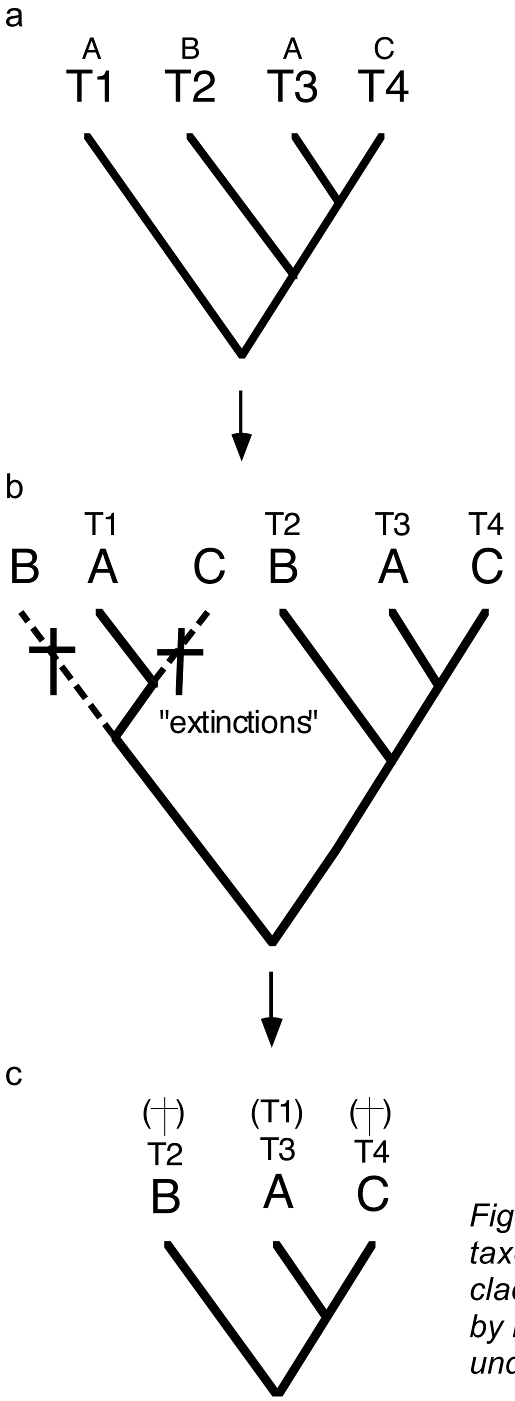


Fig. 5. Solution of a redundant distribution of both taxon T1 and T3 in area A under A1 (a: taxon cladogram with areas, b: "larger" pattern derived by hypothesizing "extinctions", c: area cladograms under A1 extracted from larger pattern).

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

2.2.3 Assumption 2

A2 is originally described by Nelson and Platnick (1981, p.432). With respect to widespread taxa, they state under A2 "... whatever is true of the one occurrence might not be true of the other occurrence". This statement shows the same ambiguity with regard to a taxon *versus* an area relationship approach as described for A1.

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

For the taxon-area cladogram (Fig. 6a), this assumption yields two possibilities: A(BC) and A(BD) when taxon T3b in area D or taxon T3a in area C, respectively, is removed from the analysis. Each of these possibilities includes three of the four areas concerned and further steps are necessary to place the taxon occurrences with missing areas in the cladogram. Placing taxon T3b in area D in cladogram A(BC) yields five different cladograms with six different components ((ABC), (BCD), (AD), (BC), (BD), (CD)) and one component in common (ABCD). Placing taxon T3a in area C in cladogram A(BD) also yields five different cladograms with six different components ((ABD), (BCD), (AC), (BC), (BD), (CD)) and one component in common (ABCD). These two outcomes have nine different components in common. Based upon these common components seven different resolved area cladograms can be constructed (Fig. 6c). When these area cladograms are compared with the original taxon cladogram, it appears that for most of the occurrences of the widespread taxon relationships with the other taxa are disturbed. The occurrences of the widespread taxon need not branch off successively and the taxon-area cladogram cannot always be recovered by the removal of a single internode, as is the case under A1. Apart from the *a priori* explanation of the presence of a widespread taxon (T3) by assuming isolation or break-up of areas without speciation (Fig. 6c; "synapomorphy") or extinction of the widespread taxon (T3) in part of its range (Fig. 6c; "extinction"), the presence of the widespread taxon (T3) can also be explained by *a priori* assuming dispersal (Fig. 6c; "dispersal").

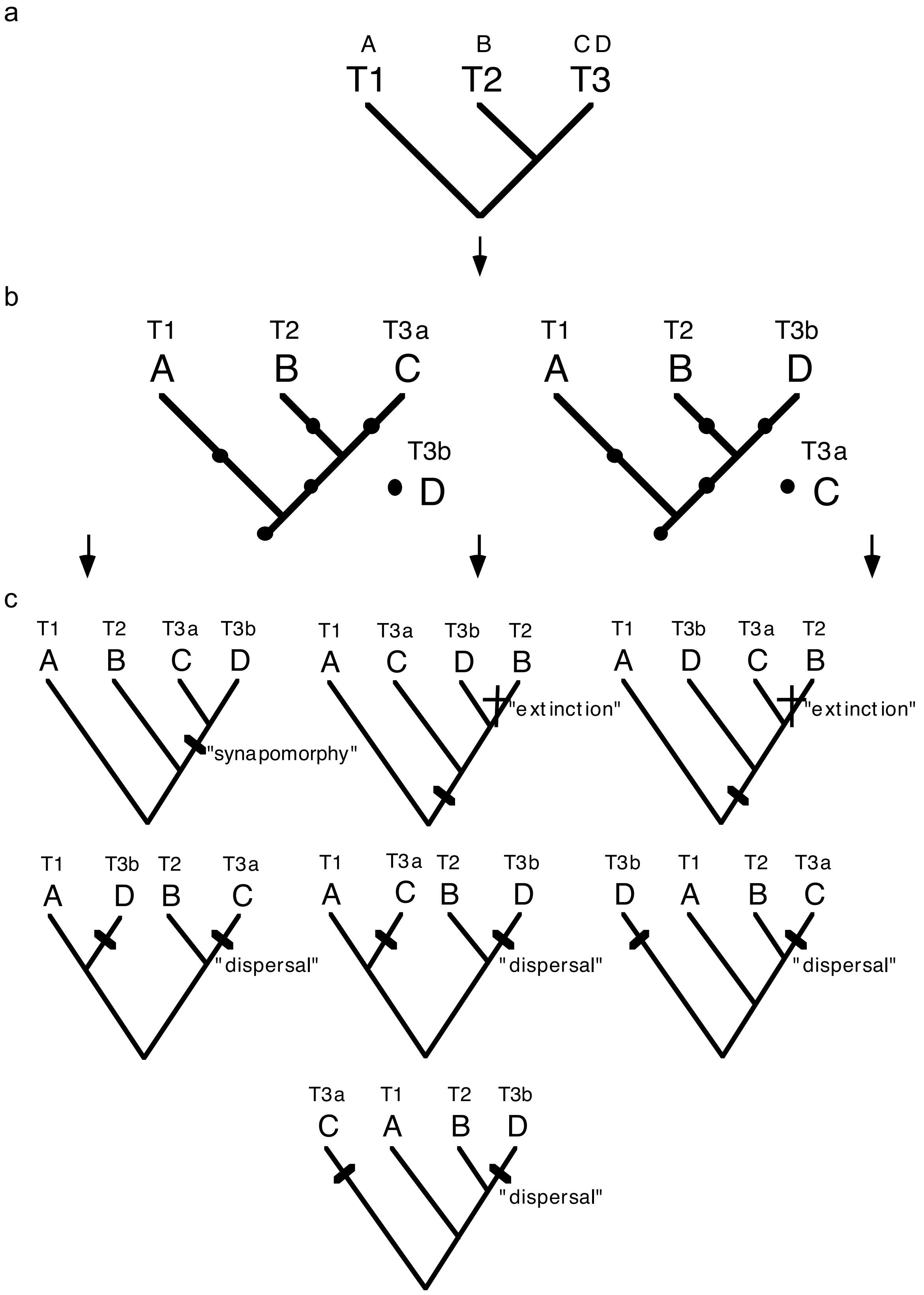
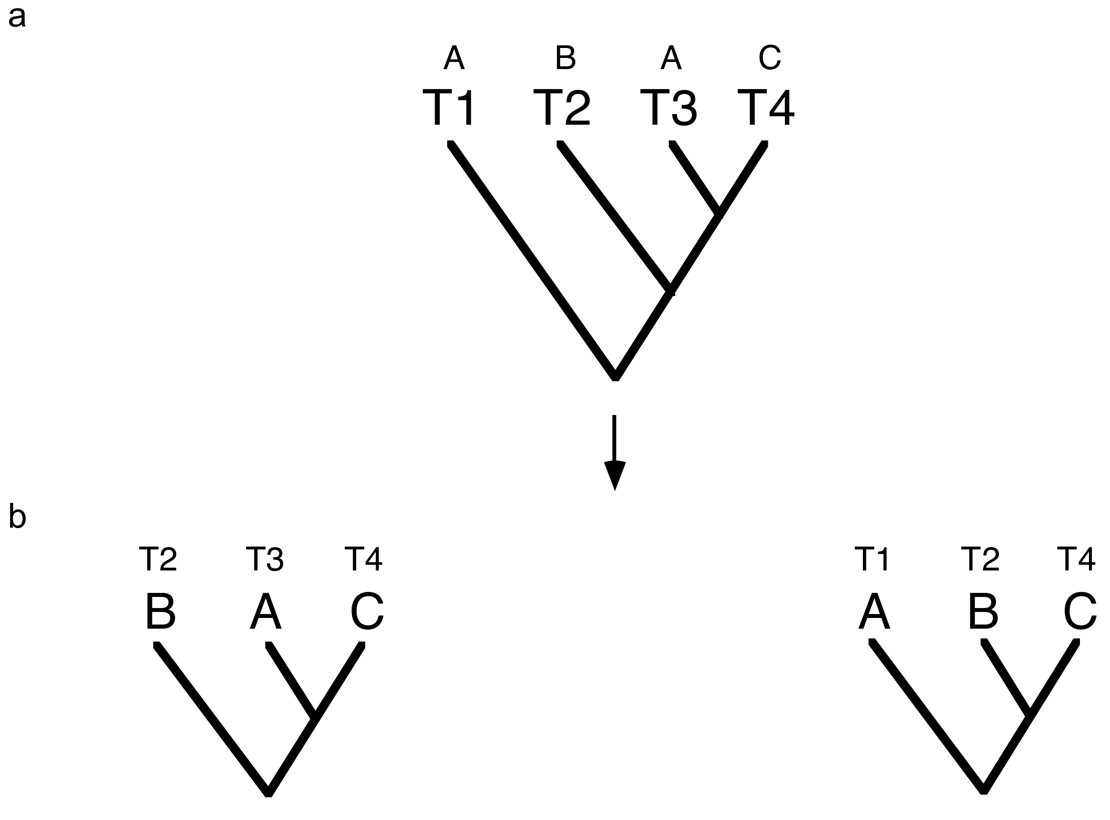


Fig. 6. Solution of a widespread distribution of taxon T3 in areas C and D under A2 (a: taxon cladogram with areas, b: removal and "floating" of one of the widespread occurrences hypothesizing dispersal, c: area cladograms under A2 with a "synapomorphy", "extinctions" or "dispersals").

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



removal of T1 in area A ("dispersal")

removal of T3 in area A ("dispersal")

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

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

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

2.3 Two requirements for obtaining valid common patterns

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

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

First, the requirement for inclusion. We have described Assumptions zero, 1, and 2, without paying attention to the possibility that they show interaction(s) that may affect their solution sets. We consider the different processes underlying the assumptions *a priori* to be independent in their effect. That is, for any particular group we preclude *a priori* the possibility that, for instance, its possible susceptibility to extinction is related to, affected by, or dependent on its dispersal capabilities. Of course, there is a possibility that a poor disperser has a higher chance of extinction. However, we think that this dependence should not be assumed *a priori*, but should be revealed from the data *a posteriori*.

In summary, the effects of the separate processes are *a priori* considered to be additive and not multiplicative. In this way, the number of possible area cladograms increases when additional processes are *a priori* assumed under a less strict assumption. The area cladograms already derived under a strict assumption (*e.g.* $S_1$1 in Fig. 8b) are found also under a less strict assumption (*e.g.* $S_1$2 in Fig. 8b). The additional area cladograms derived under the less strict assumption are the result of the additional *a priori* assumed process (*e.g.* dispersal under A2 compared to A1 in Fig. 8b). The same processes *a priori* assumed (*e.g.* vicariance and extinction in Fig. 8b) under both the strict assumptions (*e.g.* A1 in Fig. 8b) and less strict assumptions (*e.g.* A2 in Fig. 8b) affect the derivation of area cladograms in the same way. Therefore, these identical processes should result in the same area cladograms under both assumptions. Any situation for which this inclusion relation for assumptions does not hold shows inconsistency. In case of noninclusion (Fig. 8a), the effects of the common processes (*e.g.* vicariance and extinction in Fig. 8a when comparing $S_1$1 with $S_1$2) *a priori* considered for both a strict assumption (*e.g.* A1 in Fig. 8a) and a less strict assumption (*e.g.* A2 in Fig. 8a) do not result in the same area cladograms. As a result, the common processes are no longer common with respect to the area cladograms that are obtained and the independence of the processes *a priori* assumed is *a priori* precluded. Again, we think this should not be *a priori* assumed but revealed from the data after area cladograms are obtained.

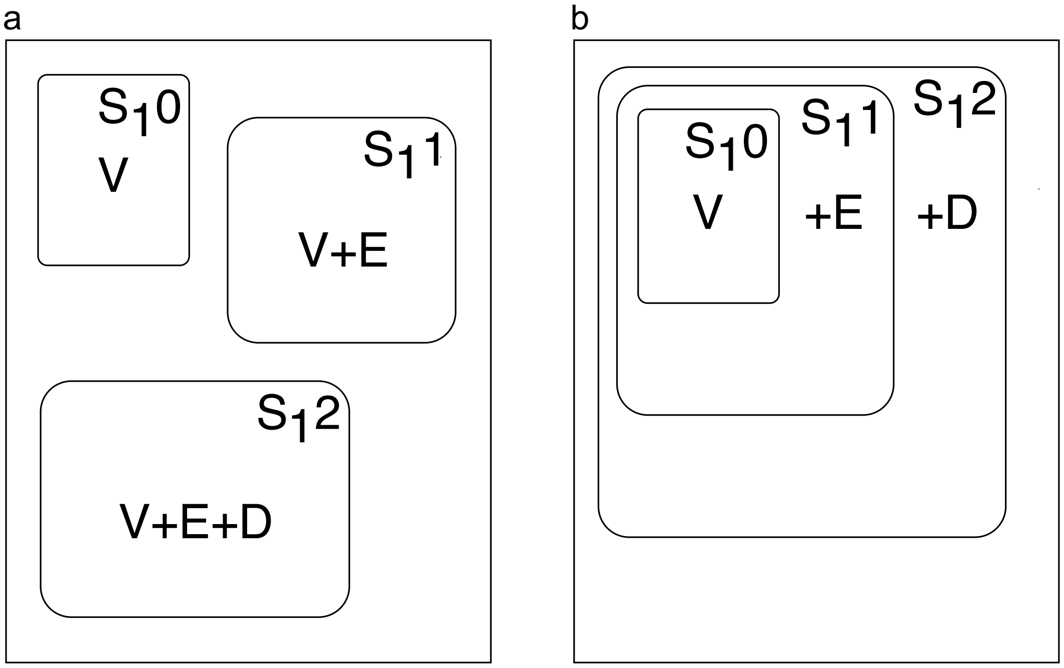

a
b

Fig. 8. Solution sets derived under Assumptions zero ($S_1$0), 1 ($S_1$1) and 2 ($S_1$2) for a single group of taxa (a: noninclusive solution sets, b: inclusive solution sets, V = Vicariance, E = Extinction, D = Dispersal).

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

Second, the requirement for comparison under a single assumption. This requirement is illustrated in Fig. 9. Here solution sets for two groups of taxa (G_1, G_2) are obtained under A0 and A1. Common patterns (CP) are found under A1 in the intersection of S_11 and S_21 (Fig. 9a). The common patterns are valid because only a single set of processes (A1: vicariance and extinction) is *a priori* assumed to have affected both groups. Of course, different (sets of) processes may have governed the pattern of distribution for the taxa of the separate monophyletic groups. Since we have no *a priori* knowledge of the relative importance of these processes, we can only test our hypotheses on extinction or dispersal as forces generating distribution patterns. All taxa are considered to have responded equally to the possible processes of vicariance, extinction, or dispersal when comparing patterns of distribution of monophyletic groups. The degree to which this actually makes sense for a particular group of taxa is indicated by the degree of fit of its particular cladogenetic and distribution history to the area history represented in the general area cladogram. For a particular monophyletic group a fair amount of dispersal, for instance, may well have to be assumed *a posteriori* in order to explain the deviation of its pattern from the general pattern. The general area cladogram thus serves as a framework for testing the relative contribution of a particular set of processes (capability for dispersal or susceptibility to extinction) to the pattern of distribution over areas for a group of monophyletic taxa. It can play this role only when the requirement of comparison under a single assumption is met.

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

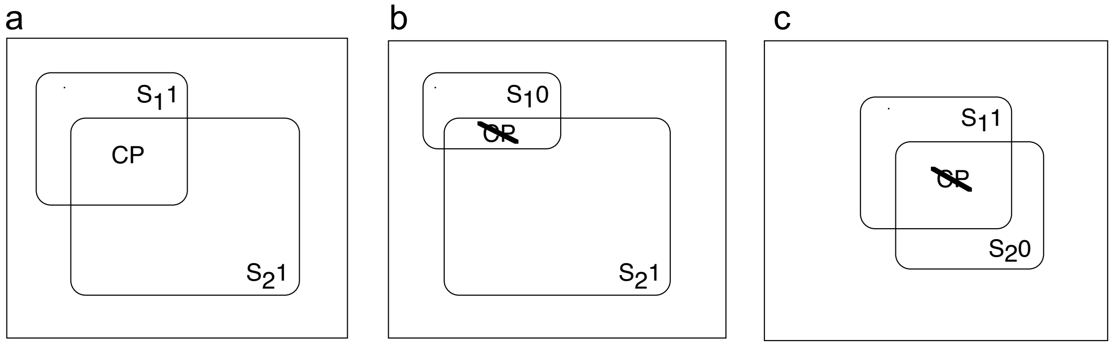


Fig. 9. Search for the intersection in solution sets (S_i0, S_i1, S_i2) to find common patterns (CP) for two groups of taxa under the same assumption (a: valid common patterns in the intersection of S_11 and S_21, b: invalid common patterns in the intersection of S_10 and S_21, c: invalid common patterns in the intersection of S_11 and S_20).

2.4 The increase in explanatory power

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

The latter requirement, however, does not imply that it is forbidden subsequently to consider the intersection of the valid common patterns obtained under a particular set of assumptions with the solution sets of particular groups of taxa under different sets of assumptions. It may occur that the intersection of the solution sets obtained under a single set of assumptions for a number of monophyletic groups (delivering valid common patterns, thus general area cladograms) contain as subsets the intersection with the solution sets obtained under different sets of assumptions for the particular monophyletic groups. This is illustrated in Fig. 10a where the intersection of $S_1 2$ with $S_2 2$ includes part of $S_1 1$. Though the occurrence of such non-empty intersections does not contribute to the inference of the general area cladograms, it does enable us to evaluate generically whether, for particular monophyletic groups, an explanation by a smaller set of assumption is sufficient. In Fig. 10a, for instance, we can conclude that, though the general area cladogram(s) demand assumption of all three processes of vicariance, extinction and dispersal, we can refine the explanation of the distribution pattern of G_1 by vicariance and extinction only. In other words, we have gained explanatory power.

The gain in explanatory power with respect to particular groups can be complex (Fig. 10b). There can be more than one solution set obtained under different sets of assumptions for more than one particular monophyletic group which intersect with the set of valid common patterns. The generic evaluations suggested by these intersections cannot, however, all be true simultaneously. For instance, we may infer that for G_1 vicariance is sufficient ($S_1 0$) to explain its pattern of distribution, but only when it is simultaneously true for G_2 that the combination of all three processes is required ($S_2 2$). Conversely, we may infer that for G_2 vicariance is sufficient ($S_2 0$), but only when for G1 all three processes are required ($S_1 2$). However, these two inferences are incompatible. Under *a priori* reasoning the gain in explanatory power for particular groups is likely to decrease when the number of groups considered increases. This can become important when authors add more monophyletic groups to their biogeographical analysis.

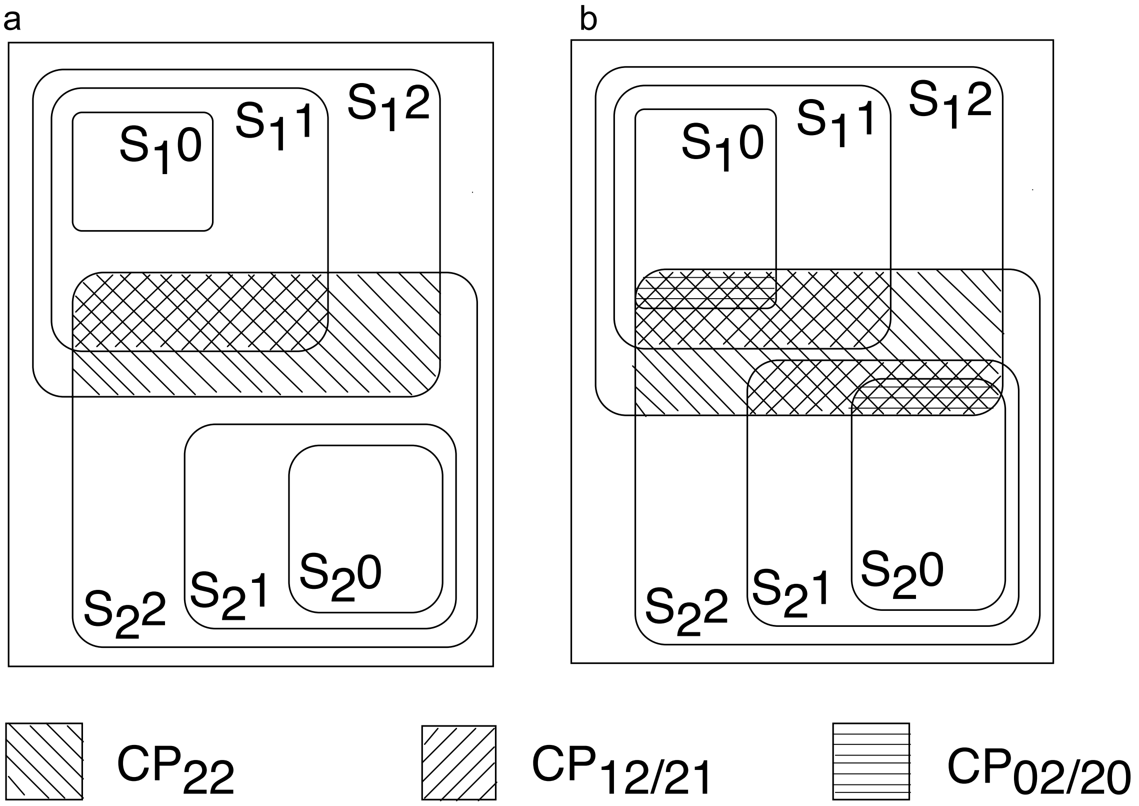


Fig. 10. Complex common patterns (CP_{AiAj}) allowing an increase in explanatory power (a: CP_{22} under A2 for both G_1 and G_2 and as a subset CP_{12} for G_1 under A1 and G_2 under A2, b: CP_{22} under A2 for both G_1 and G_2 and several subsets CP_{12}, CP_{21}, CP_{02}, CP_{20} for G_1 and G_2 under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

2.5 Violating requirements: examples from the literature

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

Page (1990) and also Morrone and Crisci (1995) state that, because the protocols under the different assumptions are not mutually exclusive, devising other protocols that combine different aspects of the original protocols is allowed: “For a given set of monophyletic groups we could treat widespread taxa under A2, but redundant distributions under A0” (Page 1990, p. 120). Enghoff (1996) suggests the opposite and deals with widespread taxa under A0 and redundant distributions under A2. As illustrated in Fig. 9, these examples are a violation of the requirement of comparison of solution sets under a single assumption. Consequently, they break down the framework for testing preconceived ideas with respect to, for instance, dispersal capabilities of a particular group.

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

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

More recently, Nelson and Ladiges (1996) developed another method for derivation of area cladograms as paralogy-free subtrees. Paralogy is a term used in molecular biology to refer to copies of the same gene in a genome (Fitch, 1970). Duplication of genes in phylogeny obscure information on relationships of the taxa in which these genes occur because multiple copies of the same gene can show different phylogenies deviating from the species phylogeny (Patterson, 1987).

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

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

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

Author (year)	*A priori* dealing with widespread taxa	*A priori* dealing with sympatric taxa	Area cladograms
Rosen (1978) [1]	excluded from analysis	excluded from analysis	reduced for areas with widespread or sympatric taxa
Kluge (1988) [1]	coded as missing data	weighted according to a minimal number of vicariance events	only based on non-widespread taxa
Van Soest (1996) [2]	areas with widespread taxa form a single area with a history different from its constituent areas	coded "as is"	with extra areas (multiple histories) for widespread taxa
Zandee and Roos (1987) [3]	A0: occurrences of widespread taxon form a sister group	A0: coded "as is"	A0: based on all taxa
	A1: occurrences of widespread taxon form a sister group or branch off sequentially	A1: coded "as is"	A1: based on all taxa
	A2: one occurrence of the widespread taxon fixed position, other occurrences "float"; and vice versa	A2: coded "as is"	A2: based on all taxa
Wiley (1988a,b), Brooks (1990) [3]	A0: occurrences of widespread taxon form a sister group	A0: coded "as is"	A0: based on all taxa
Nelson and Platnick (1981) [3]	A1: occurrences of widespread taxon form a sister group or branch off sequentially	A1: all sympatric taxa part of larger pattern	A1: based on all taxa plus additional assumed (extinct) taxa
	A2: one occurrence of the widespread taxon fixed position, other occurrences "float"; and vice versa	A2: all but one of the sympatric taxa removed	A2: based on all non-sympatric taxa and only one of the sympatric taxa

Humphries (1982), Humphries and Parenti (1986) [3]	A1: occurrences of widespread taxon form a sister group or branch off sequentially	A1: not discussed	A1: based on all taxa
	A2: one occurrence of the widespread taxon fixed position, other occurrences “float”; and vice versa	A2: not discussed	A2: based on all taxa
Page (1988, 1989, 1990, 1993), Morrone and Carpenter (1994), Morrone and Crisci (1995) [4]	A0: areas with widespread taxon have monophyletic relationship	A0: not discussed	A0: based on all taxa
	A1: areas with widespread taxon have mono- or paraphyletic relationship	A1: extinction of taxa in areas without sympatric taxa	A1: based on all taxa and additionally assumed taxa
	A2: areas with widespread taxon have mono- or para- or polyphyletic relationship	A2: distribution of all but one sympatric taxa result of dispersal and removed	A2: based on all non-sympatric taxa and only one of the sympatric taxa
Enghoff (1996)	A0: occurrences of widespread taxon form a sister group	A2: distribution of all but one sympatric taxa result of dispersal and removed	A02: based on all non-sympatric taxa and only one of the sympatric taxa
Nelson and Ladiges (1991a,b)	A0: widespread taxon contributes an extra three area statement	A0: not discussed	A0: based on all taxa
	A12: widespread taxa do not contribute three area statements	A12: not discussed	A12: based on all non-widespread taxa
Nelson and Ladiges (1996), Ladiges (1998)	widespread taxa removed in favor of endemics	subtrees derived by removal of all but one of the sympatric taxa	based on all non-sympatric taxa and only one of the sympatric taxa

[1] = A0, A1 or A2 are not applied; widespread or sympatric taxa are dealt with by a priori excluding these data from analysis.
[2] = A0, A1 or A2 are not applied; widespread taxa are dealt with by a priori assuming multiple histories for the areas with widespread taxa and thereby excluding these data from analysis.
[3] = Taxon relationship approach; A0, A1 and A2 are interpreted to pertain to taxon relationships.
[4] = Area relationship approach; A0, A1 and A2 are interpreted to pertain to area relationships.

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

2.6 Conclusions

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

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

- sets of processes are related similarly to vicariance (A0) $\subset$ vicariance + extinction (A1) $\subset$ vicariance + extinction + dispersal (A2); and
- solution sets are related similarly to $S_i0 \subseteq S_i1 \subseteq S_i2$.

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

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

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

In the next chapter we shall investigate how the different methods have implemented A0, A1, and A2 and whether the requirement of inclusive solution sets is met when the different methods are applied on theoretical and empirical datasets.

Acknowledgements

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

Fitch, W.M. (1970). Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**, 99-113.

Humphries, C.J. (1982). Vicariance biogeography in Mesoamerica. *Ann. Missouri Bot. Gard.* **69**, 444-463.

Humphries, C.J. and Parenti, L.R. (1986). “Cladistic biogeography”. Clarendon Press. Oxford.

Kluge, A.G. (1988). Parsimony in vicariance biogeography: A quantitative method and a Greater Antillean example. *Syst. Zool.* **37**, 315-328.

Ladiges, P.Y. (1998). Biogeography after Burbidge. *Austral. Syst. Bot.* **11**, 231-242.

Morrone, J.J. and Carpenter, J.M. (1994). In search of a method for cladistic biogeography: an empirical comparison of Component Analysis, Brooks Parsimony Analysis and Three area statements. *Cladistics* **10**, 99-153.

Morrone, J.J. and Crisci, J.V. (1995). Historical biogeography: Introduction to methods. *Ann. Rev. Ecol. Syst.* **26**, 373-401.

Nelson, G. and Ladiges, P.Y. (1991a). Three area statements: standard assumptions for biogeographic analysis. *Syst. Zool.* **40**, 470-485.

Nelson, G. and Ladiges, P.Y. (1991b). Standard assumptions for biogeographic analysis. *Austral. Syst. Bot.* **4**, 41-58.

Nelson, G. and Ladiges, P.Y. (1996). Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *American Museum Novitates* **3167**, 58pp.

Nelson, G. and Platnick, N.I. (1981). “Systematics and biogeography: Cladistics and vicariance”. Columbia University Press. New York. 567 pp.

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

Page, R.D.M. (1989). Comments on component compatibility in historical biogeography. *Cladistics* **5**, 167-182.

Page, R.D.M. (1990). Component analysis: A valiant failure? *Cladistics* **6**, 119-136.

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

Patterson, C. (1987). “Molecules and morphology in evolution: conflict or compromise?” Cambridge University Press. Cambridge, London, New York, New Rochelle, Melbourne, Sydney. 229 pp.

Platnick, N.I. (1988). Systematics, evolution and biogeography: A Dutch treat. *Cladistics* **4**, 308-313.

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

Soest, R.W.M. van (1996). Recoding widespread distributions in general area cladogram construction. *Vie Milieu* **46**, 155-161.

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

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

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

CHAPTER 3

METHODS IN VICARIANCE BIOGEOGRAPHY: ASSESSMENT OF THE IMPLEMENTATIONS OF ASSUMPTIONS ZERO, 1, AND 2

Abstract

Two requirements should be met if general area cladograms in vicariance biogeography are to be derived validly. First, sets of area 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 assess the following methods (and correlated computer 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 their implementation of Assumptions zero, 1, and 2 and for the extent to which they meet Requirements I and II. For this purpose we use empirical (*Heterandria*, *Xiphophorus*, *Cyttaria*, *Eriococcus*/*Madarococcus*) and theoretical data sets. We find that all programs violate, to different degrees, Requirement I when dealing with sympatric taxa under A1 or A2. This violation is prevented only by dealing with sympatric taxa *a posteriori*. All programs examined meet Requirement II.

This chapter has been published in a modified form as: Veller, M.G.P. van, Kornet, D.J. & Zandee, M. (2000). Methods in vicariance biogeography: assessment of the implementations of Assumptions zero, 1 and 2. *Cladistics* **16**, 319-345.

3.1 Introduction

In vicariance biogeography the first-order explanation for the distribution of taxa over areas is that vicariance events triggered 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 of 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 have a single taxon. Data sharing no widespread or sympatric taxa result in taxon-area cladograms with a unique terminal node for each area. 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 of a particular monophyletic group can also be due to 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 that we examine all aim to hypothesize area cladograms with unique terminal nodes for each area, and so each of them has to deal with 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), less strict assumptions (A1 and A2, Nelson and Platnick, 1981; Page, 1988) have been proposed to derive area cladograms. In this chapter we assess the precise implementation of Assumptions zero, 1, and 2 by various methods. The methods vary in two main ways according to 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 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 of 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 altogether.

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 pruning and adding of taxa and taxon relationships under A1 and A2. These methods thereby fit an explanation only by the assumed processes under consideration (extinction, dispersal, etc.) and overrule the historical relationships that were previously 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 the 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 items of error for the selection of area cladograms at all, but rather minimal number of steps. Lack of agreement (between the area cladograms obtained) is also to be expected since, as we show in this chapter, 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 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.

3.2 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 computer programs) 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, see appendix i.

3.2.1 Component Compatibility Analysis (CCA)

In CCA, the data matrix that is used to derive area cladograms comprises a mapping of the cladogenetic relationships 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 corresponds to 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 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 mutually 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 either under A0, or 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 are made under A0 to the data matrix. 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 allowing the areas of a widespread taxon 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 used only to derive additional components and do not enter into the computation of cladogram length.

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

3.2.2 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 matrix is added corresponding to a hypothetical outgroup. This matrix has the value 1 for the root of each taxon cladogram (Brooks, pers. com.) and the value zero elsewhere. The 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).

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

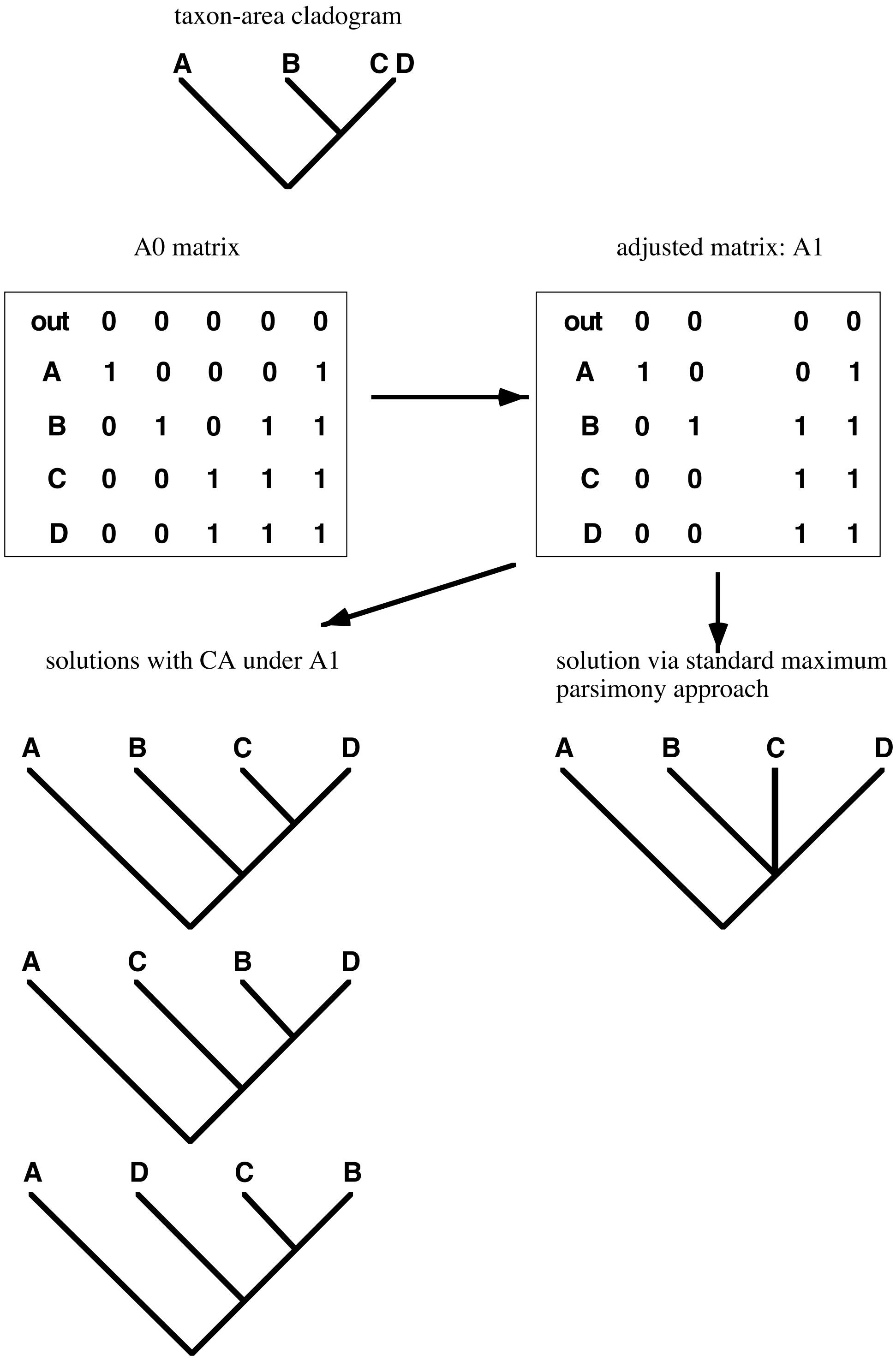
The BPA protocol (Brooks, 1990) does not provide instructions for the implementation of A1 and A2. However, in order to assess BPA for its implementation of all three assumptions, data matrices must be derived under A1 and A2 as well. In this study, we derive these data matrices by using CCA's protocol with a matrix corresponding to a hypothetical 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.

3.2.3 Component Analysis (CA)

In CA (as implemented by Page, 1988), area cladograms can be derived under one of the three different assumptions that are conditional *a priori* on the nature of the processes (vicariance, vicariance + extinction or vicariance + extinction + dispersal) that 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 the coding method of Brooks (1981) is used in the analysis. As in BPA, a hypothetical outgroup (a row of 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).



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

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

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

Internal nodes with descendant lineages with overlapping sets of areas are identified as redundant nodes. Areas present at only one descendant lineage of a redundant node and not at the other 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). Each question mark is optimised to either zero or 1, depending on which is more parsimonious (Fig. 2). Via this optimisation, certain columns in the A1 matrix are adjusted and differ 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).

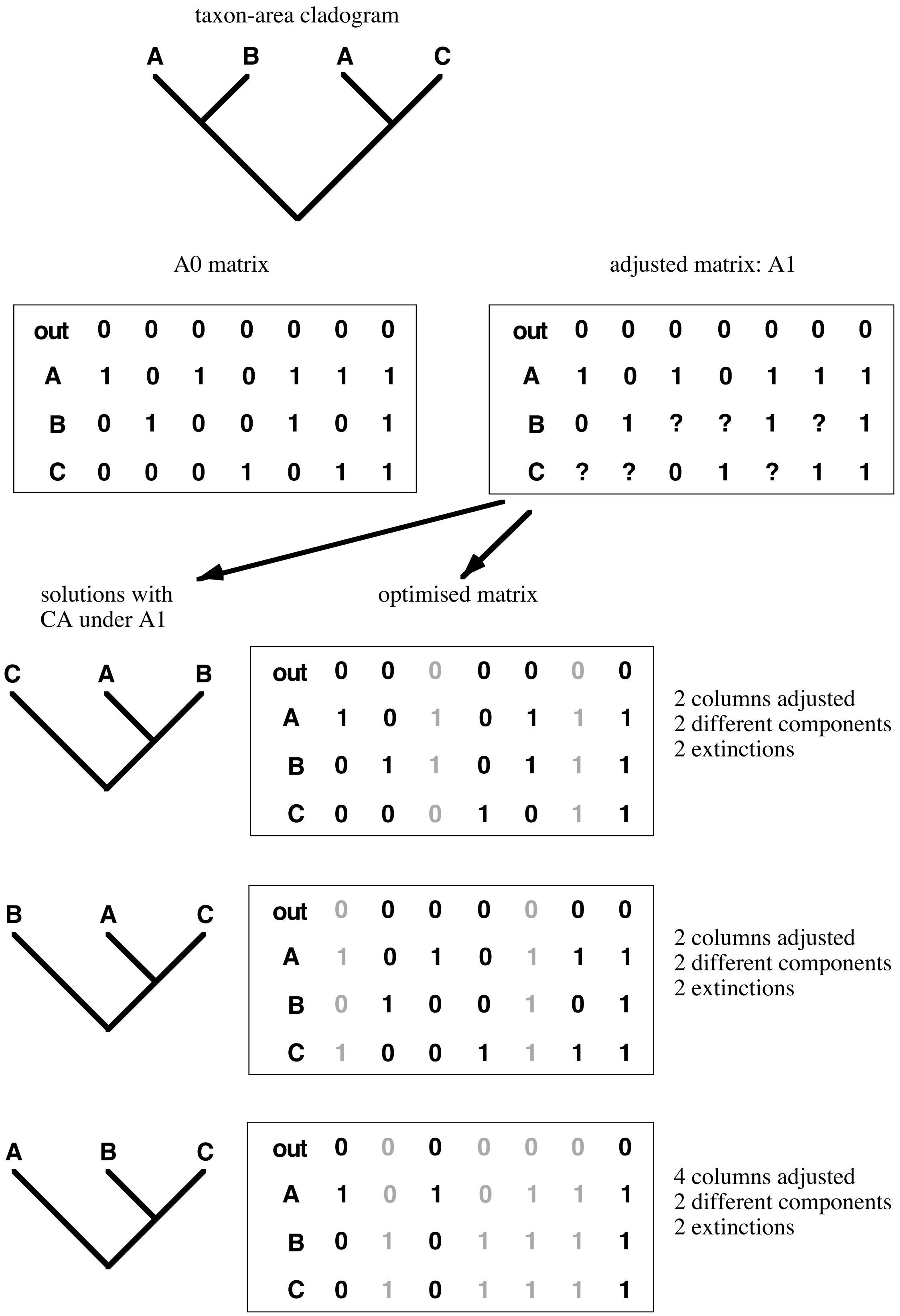
For an A2 analysis with CA, in contrast to CCA and BPA as well as the analysis under A0 or A1 with CA, no area-by-node data matrix is used. Adjustments needed to allow for widespread taxa and sympatric taxa under A2 are made by direct manipulation of the taxon-area cladogram. This taxon-area cladogram 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 their 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 allowing (non-reduced) area cladograms to be derived (Page, 1988, 1990).

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

3.2.4 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 to 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 by which it deals with sympatric taxa, although in both cases one of the three different assumptions is invoked. Also, as in CA, dealing with widespread taxa has precedence over dealing with sympatric taxa.



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

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

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

First, widespread taxa are dealt with by mapping the taxon cladogram on the taxon-area cladogram. Under A0, each area of a widespread taxon is given its own terminal node by replacing the widespread taxon by its areas of distribution and introducing extra branches for each area. These branches (each leading to an area of a widespread taxon) are connected via internal nodes in such a 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 give similar results to a CCA, BPA or CA analysis under A0. Under A1, the areas of a widespread taxon are not mapped separately, but 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 those 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; 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). We therefore 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 in accordance with 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); and
- 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 for the reconciliation.

The computer 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.

3.2.5 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 replacing 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 remainder 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, for sympatric taxa the same area cladograms are obtained under both A0 and A1. 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 removing 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.

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

3.3 Assessment of Requirements I and II in implementations of methods

3.3.1 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 (no two 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 distribution of two monophyletic poeciliid fish genera *Heterandria* and *Xiphophorus* (Rosen, 1978) (see appendix iii) occurring in overlapping areas.
- taxon cladograms and 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 implementations 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 unlike the theoretical data sets allow us also to evaluate the implementations of the methods for the extent to which they meet Requirement II (obtaining general area cladograms under a single assumption).

3.3.2 CAFCA (Component Compatibility Analysis)

Theoretical data sets for single groups

For the single-group data sets containing widespread taxa, all solution sets obtained with CAFCA (Zandee, 1999) show inclusion under A0, A1, and A2. The number of area cladograms is occasionally 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 introduced 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 as parsimonious as the cladograms derived for widespread taxa under A0. However, for most data sets with widespread taxa analyzed 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.

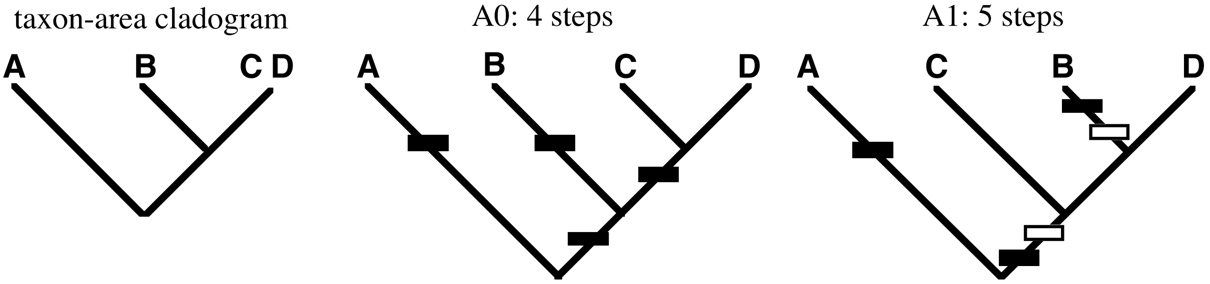


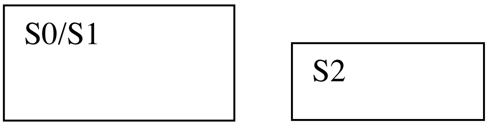
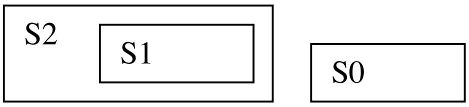
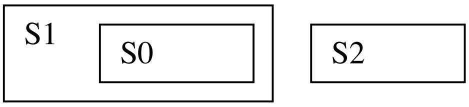
Fig. 3. Area cladograms derived under A0 and A1 with CAFCA for a cladogram with a widespread taxon (in areas C and D).

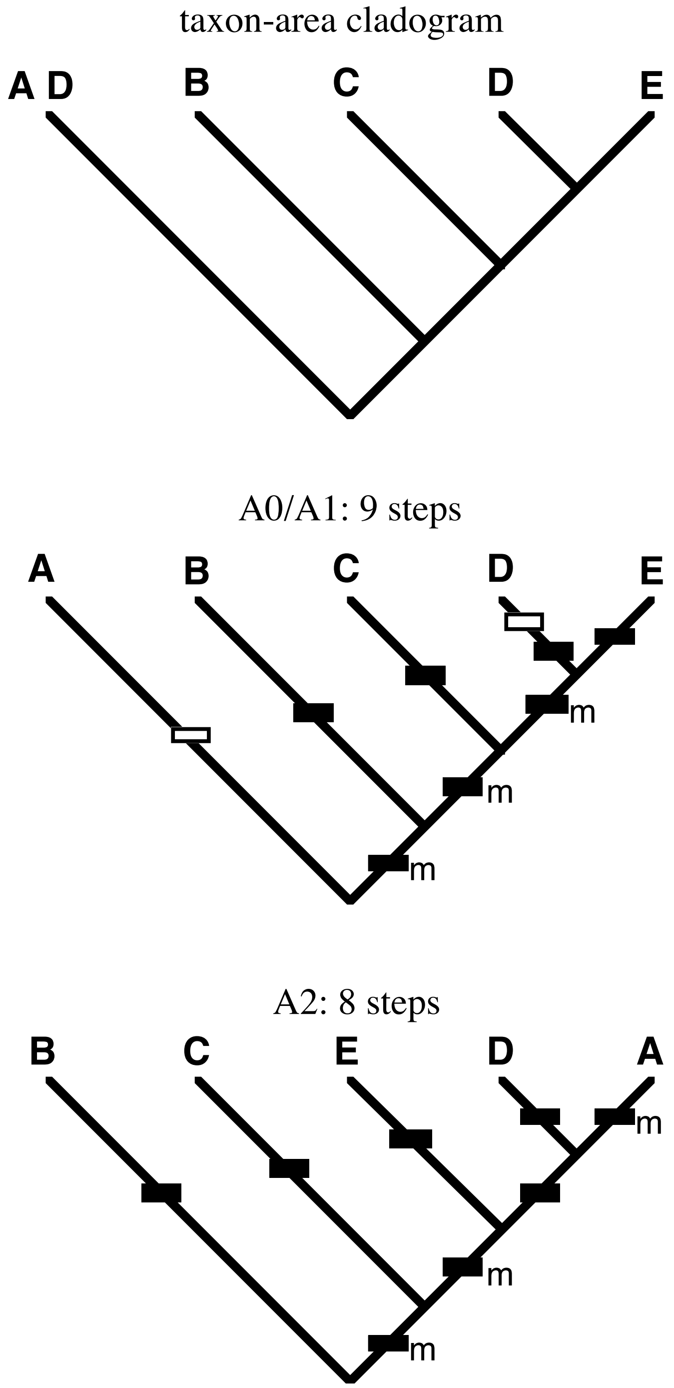
▬ = *non-homoplaseous* ▭ = *homoplaseous*

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 violated in these cases (Table 1).

Table 1: Noninclusion 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	



taxon-area cladogram

A0/A1: 9 steps

A2: 8 steps

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*

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 finds besides the most parsimonious area cladograms under A1 or A2 also the less parsimonious area cladograms that already have been derived under A0. As a result, inclusion is obtained (Requirement I).

Heterandria/Xiphophorus

Analysis of the data matrix for the poeciliid fishes with CAFCA results in one cladogram of 30 steps derived under A0. Under A1 the same cladogram is found (Fig. 5a). Under A2, by contrast, 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 violates Requirement I. However, we can use the number of steps for the cladogram derived under A0 or A1 as an upper limit for cladogram selection under A2. Consequently, the solution set under A2 becomes larger and inclusion is obtained (Requirement I).

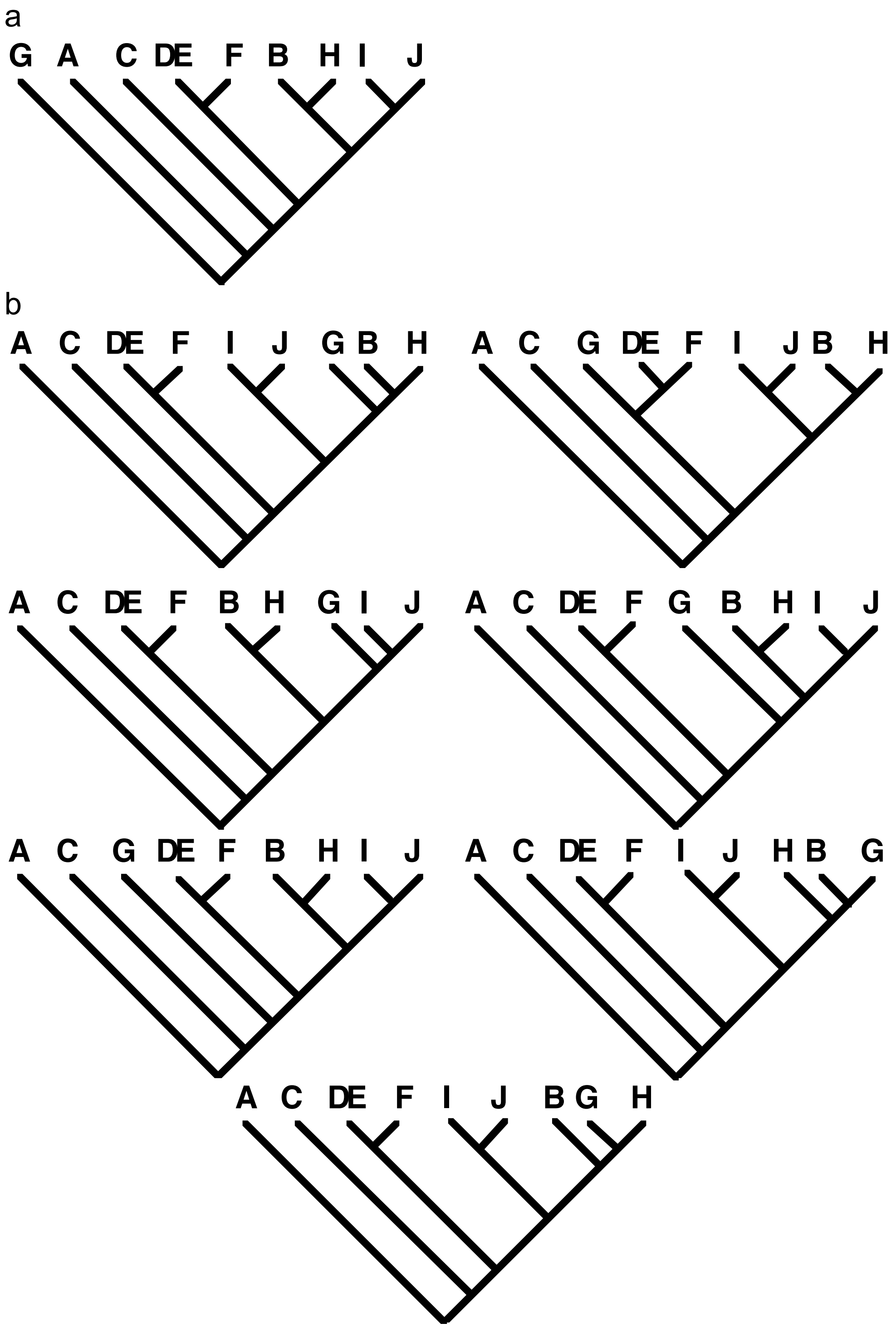


Fig. 5. General area cladograms derived under A0, A1 and A2 with CAFCA for Heterandria/Xiphophorus (a: general area cladograms under A0 and A1, b: general area cladograms under A2).

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 under each 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.

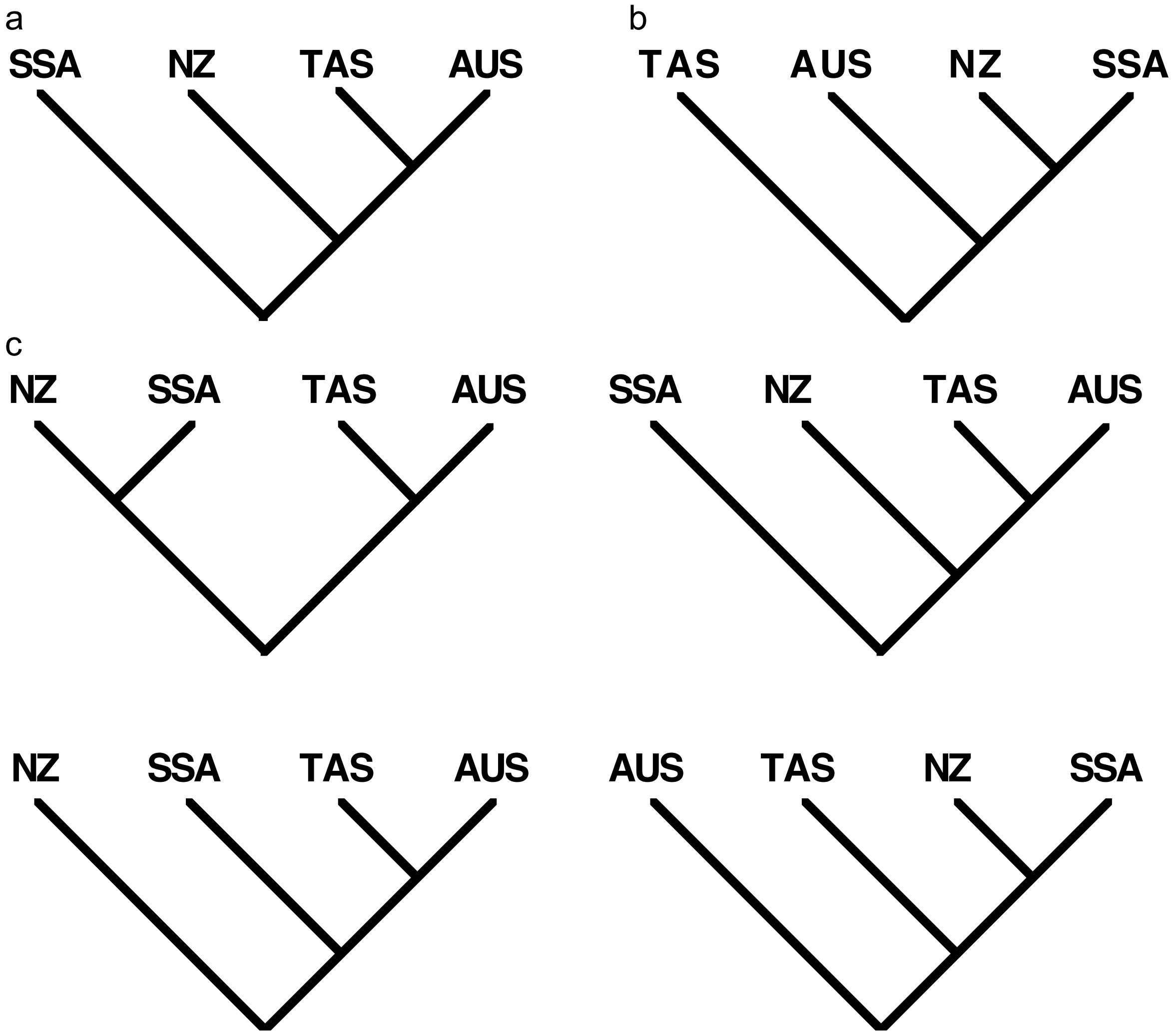


Fig. 6. General area cladograms derived under A0, A1 and A2 with CAFCA for Cyttaria/Eriococcus/Madarococcus (a: general area cladograms under A0 and A1, b: genera area cladograms under A2, c: additional general area cladograms under A2 when applying an upper limit).

The procedure of selecting cladograms under A2 that need 16 steps or fewer -*i.e.*, no more steps than the upper limit posed by the result for A0- yields not only the single most parsimonious cladogram (15 steps) but also 4 further cladograms of 16 steps (Fig. 6c). One of these cladograms is identical to the one found under A0 and A1. Therefore, if the number of steps for the cladogram derived under A0 or A1 is used 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 under each assumption for both genera together, general area cladograms are obtained under a single assumption (Requirement II).

3.3.3 PAUP (Brooks Parsimony Analysis)

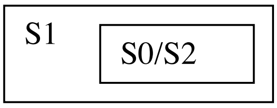
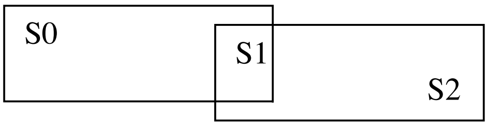
Theoretical data sets for single groups

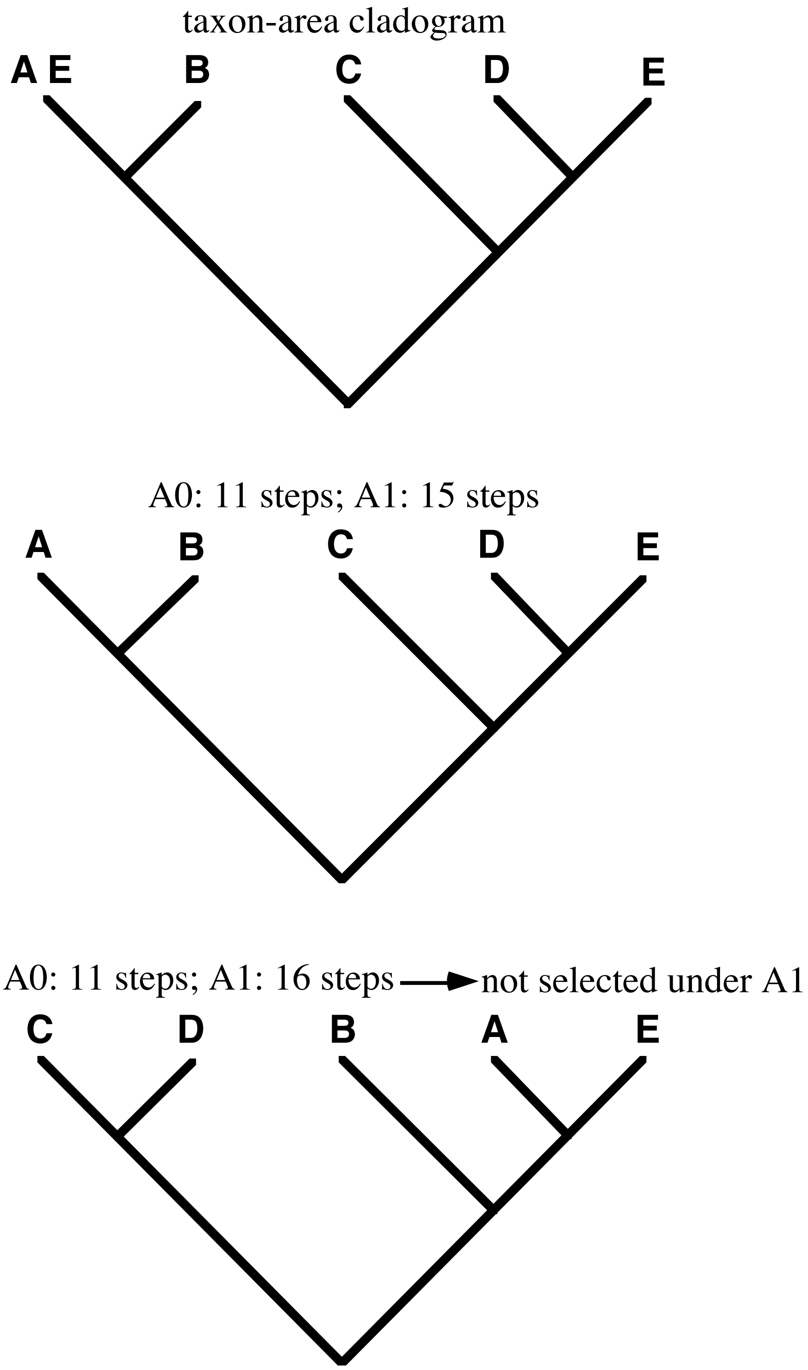
With Brooks Parsimony Analysis (BPA) all solution sets derived from data sets of a single group of taxa with a widespread taxon show inclusion. Contrary to the practise with CAFCA, all columns in the data matrix are used to count the number of steps under A1 or A2 with BPA. New columns result in new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

As in 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, noninclusive 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). Noninclusive results for the analysis of single group data sets are represented in Table 2.

Table 2: Noninclusion 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	



taxon-area cladogram

A0: 11 steps; A1: 15 steps

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

Fig. 7. Noninclusion with BPA. Derivation of other area cladograms, under A1, that are more parsimonious than the area cladograms derived under A0.

Inclusive solution sets can be obtained by optimising the data matrix, derived under A1, on the area cladograms obtained under A0. By using the greatest 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 found also 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 not violated (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 (partly overlap; Fig. 8c).

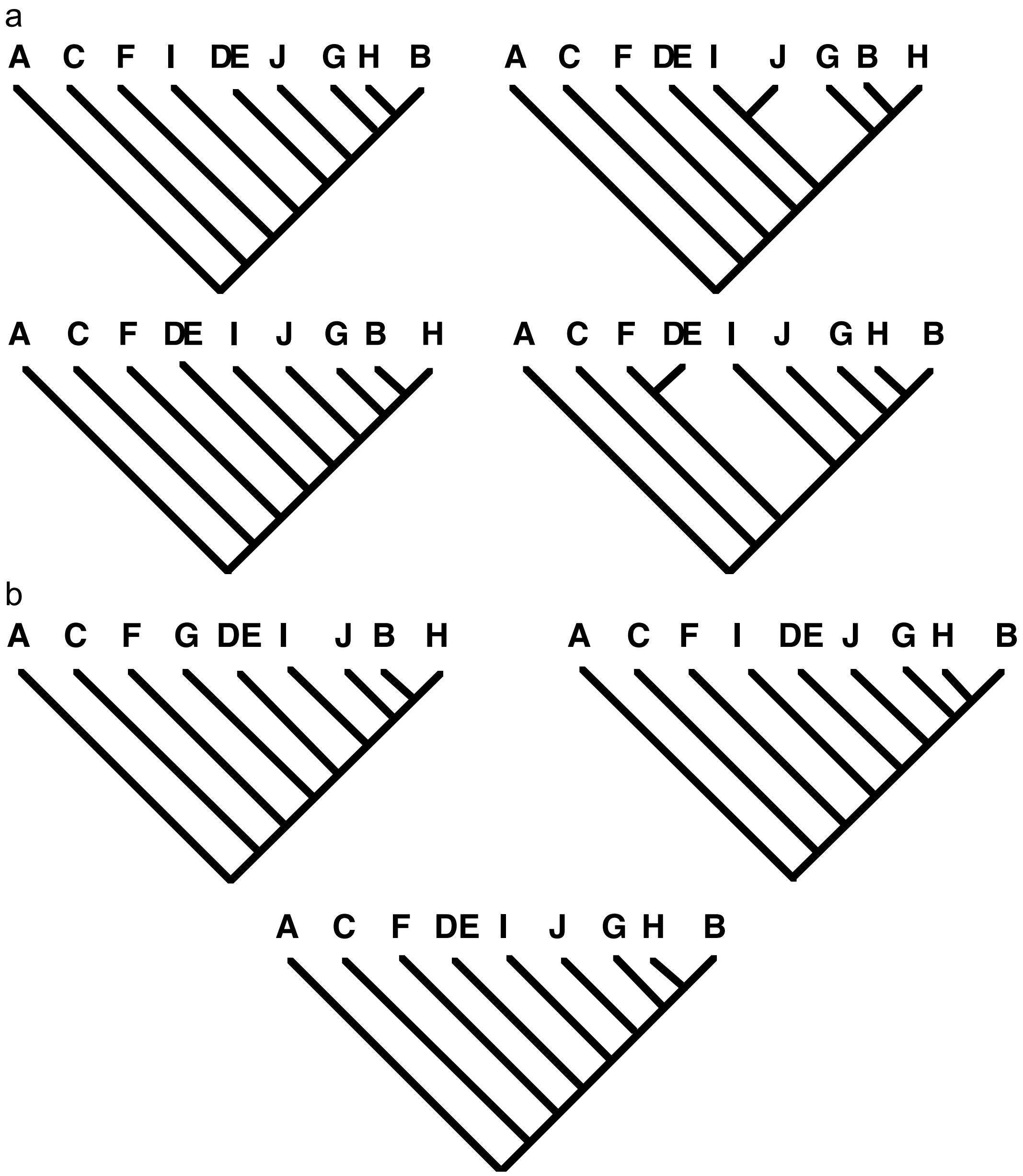


Fig. 8 (continues on next page).

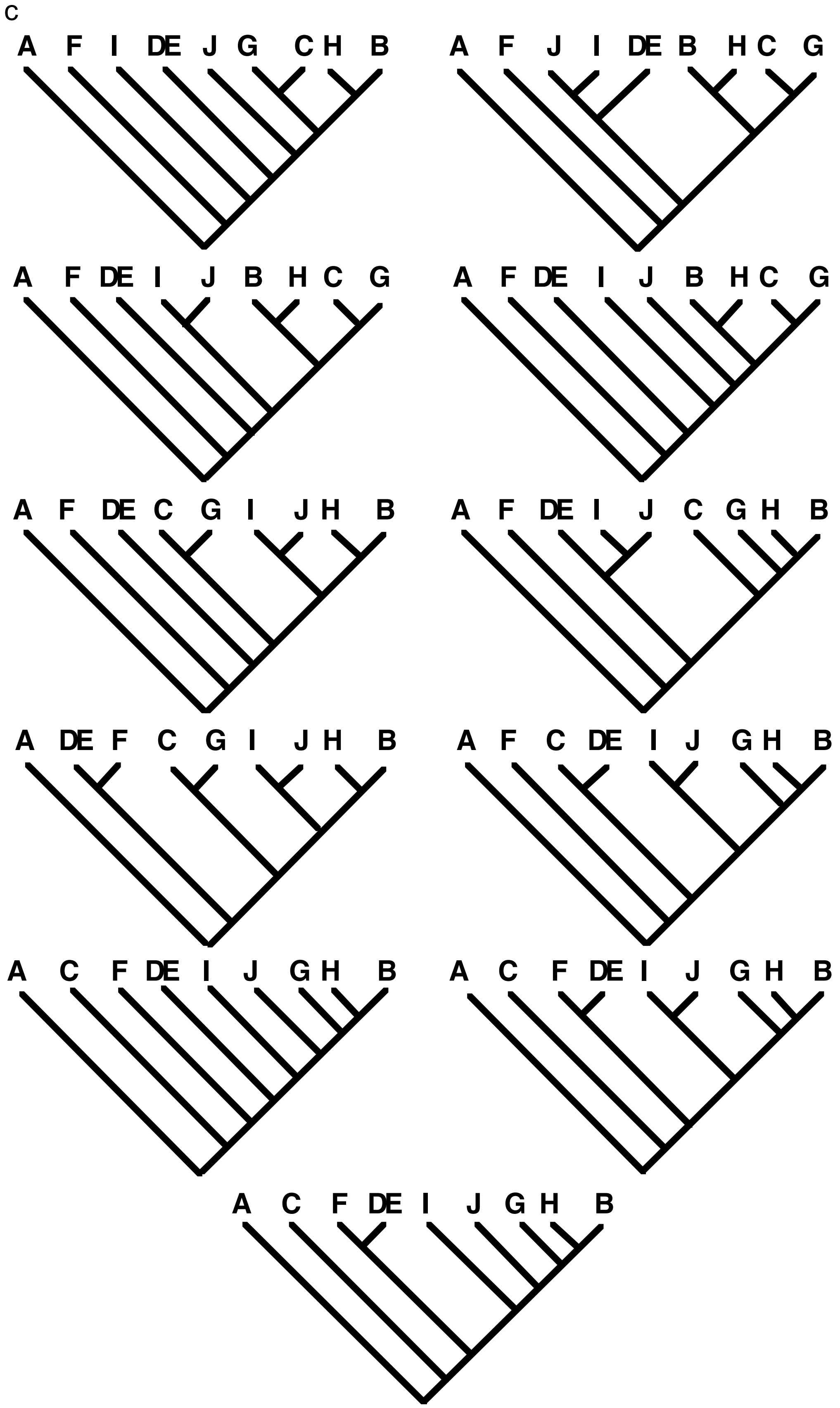


Fig. 8. General area cladograms derived with BPA under A0, A1 and A2 for Heterandria/Xiphophorus (a: general area cladograms under A0, b: general area cladograms under A1, c: general area cladograms under A2).

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 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 as 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 under each assumption for both genera together and therefore general area cladograms are obtained under a single assumption (Requirement II).

Cyttaria/Eriococcus/Madarococcus

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

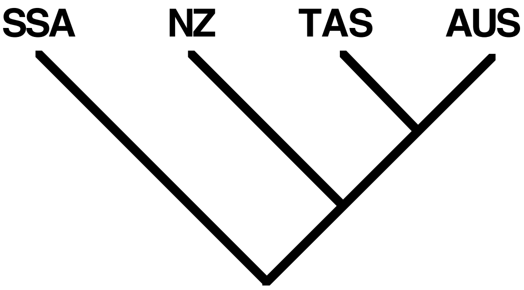


Fig. 9. General area cladogram derived with BPA under A0, A1 and A2 for Cyttaria and Eriococcus/ Madarococcus.

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 under each assumption for both genera together.

3.3.4 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, the result is often a larger number or a different range of area cladograms compared to those 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 any) of the area cladograms already obtained under A1 and inclusion of solution sets is violated.

Inclusion of solution sets is most frequently violated 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 violated. Noninclusive results for the analysis of single group data sets are represented in Table 3.

Heterandria/Xiphophorus

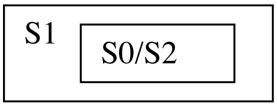
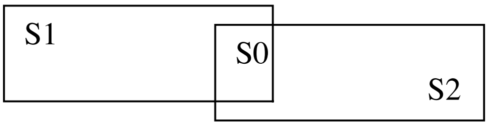
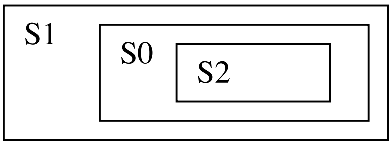
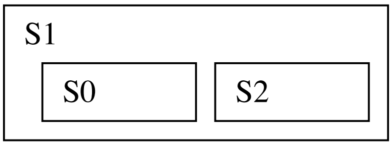
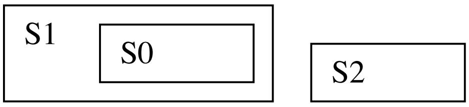
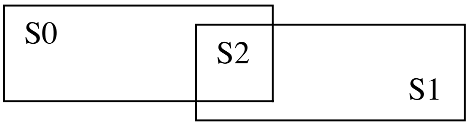
For *Heterandria*, the same area cladogram is derived under A0, A1, and A2 (Fig. 10a) with Component 1.5 because no widespread or sympatric taxa are present in the *Heterandria* data set. Inclusion is thus not violated with this data set. Area C does not inhabit any of the taxa of this group and therefore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

For *Xiphophorus*, 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 met (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.

Table 3: Noninclusion 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	

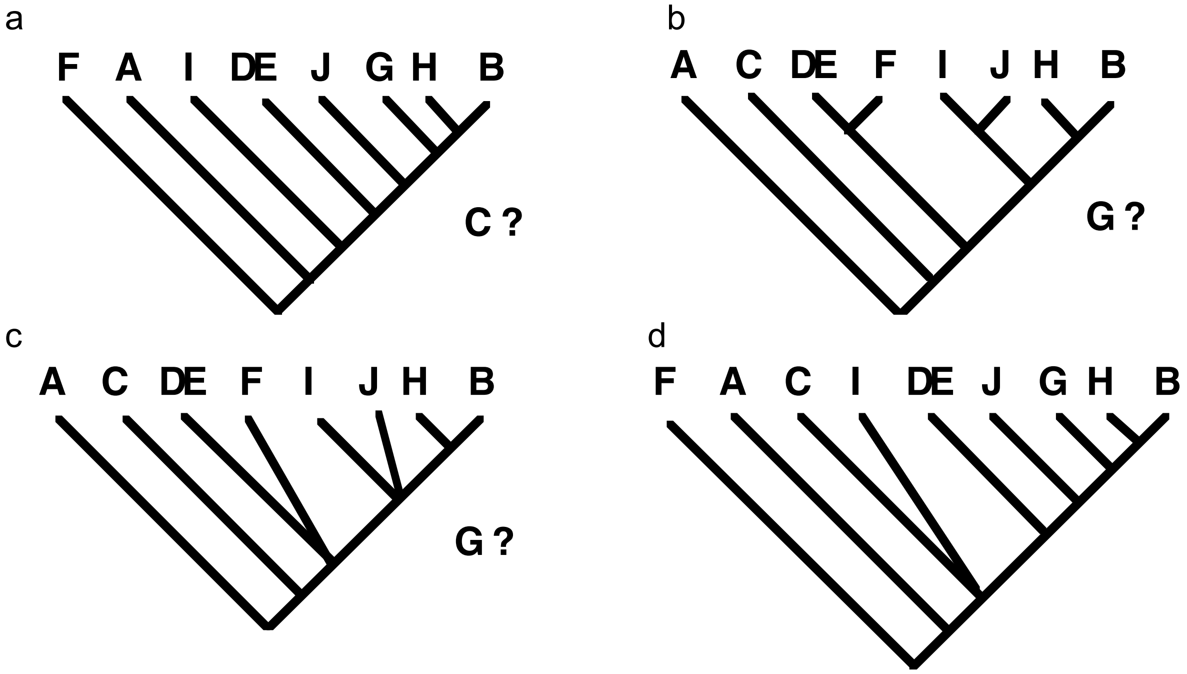


Fig. 10. Area cladograms derived with Component 1.5 under A0, A1 and A2 for Heterandria and Xiphophorus (a: area cladogram for Heterandria under A0, A1, and A2, b: area cladogram for Xiphophorus under A0, c: strict consensus of the nine area cladograms for Xiphophorus under A1, d: strict consensus of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

Because with Component 1.5 general area cladograms are found only by comparison of sets of area cladograms obtained for *Heterandria* under A2 and cladograms 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 one already derived with CAFCA and BPA under A0. Three area cladograms are obtained under A1 (Fig. 11b) and only two of these are derived under A2 (Fig. 11c). As a consequence, Requirement I is violated for the data on *Cyttaria*.

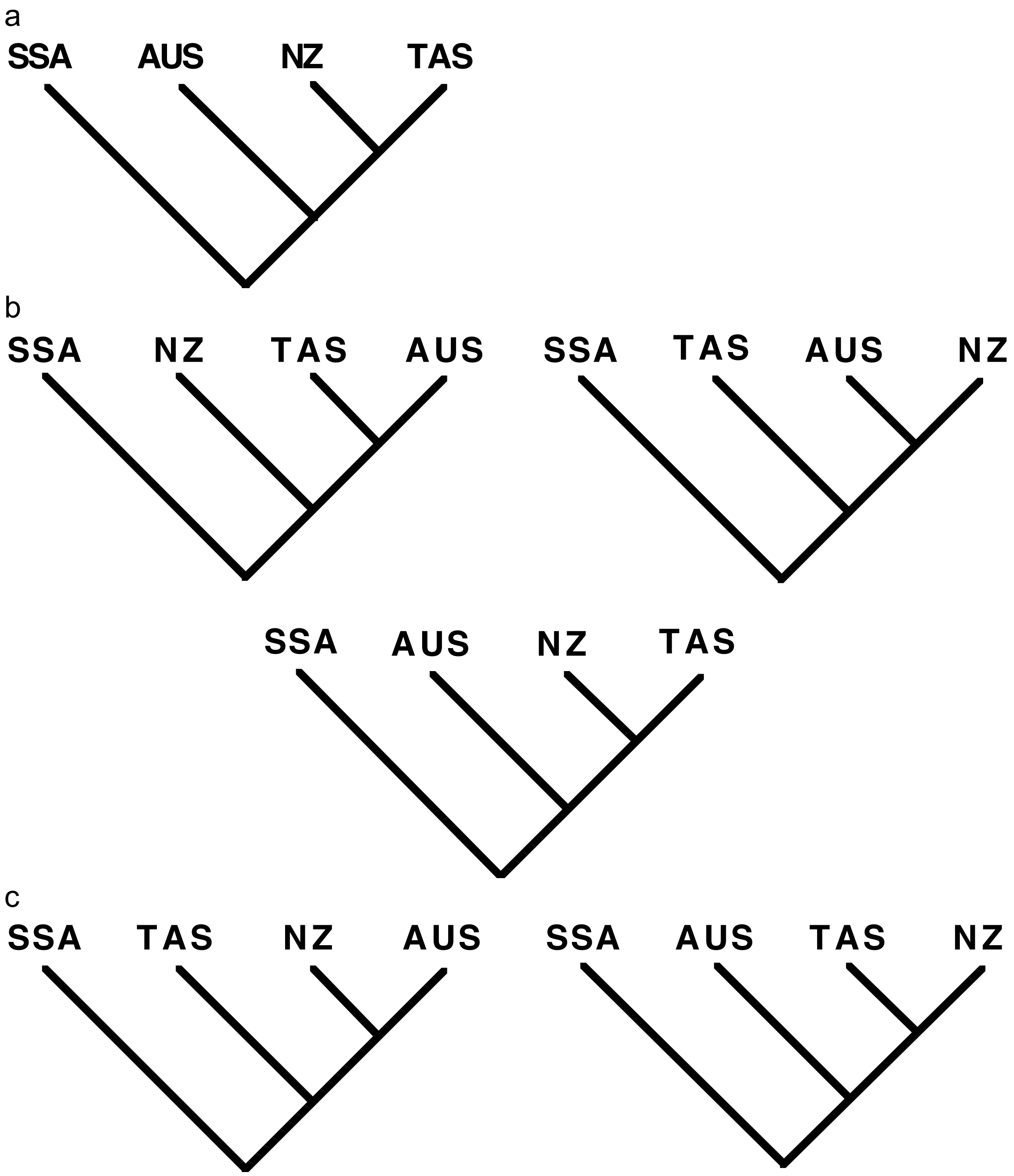


Fig. 11. Area cladograms derived with Component 1.5 under A0, A1 and A2 for Cyttaria (a: area cladogram under A0, b: area cladograms under A1, c: area cladograms under A2).

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). This area cladogram and two additional area cladograms are derived under A1 (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 violated (Requirement I).

For *Cyttaria* and *Eriococcus/Madarococcus* general area cladograms are obtained from the comparison of sets of area cladograms obtained for each group under A1 or A2. Since the sets of area cladograms are derived for each group under the same assumption, Requirement II is met.

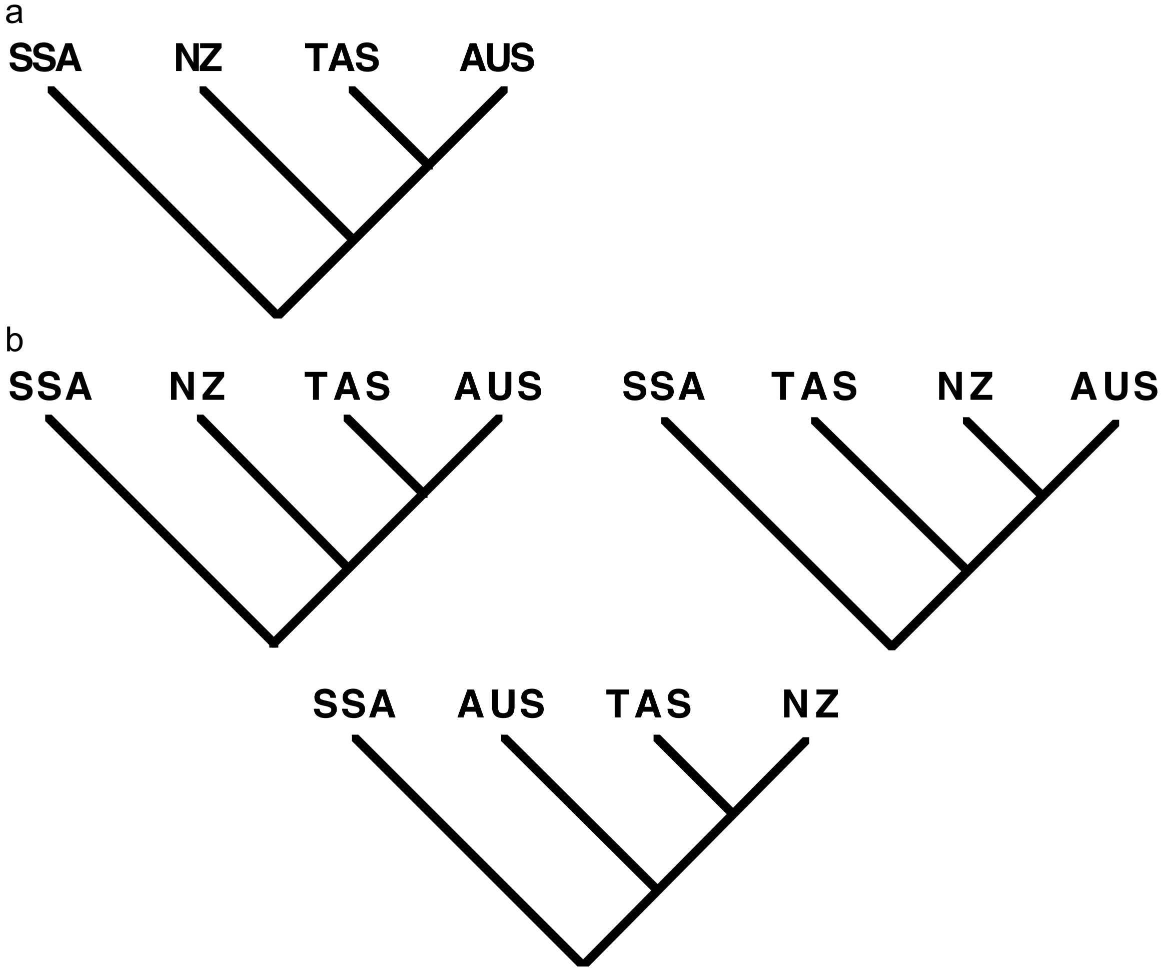


Fig. 12 (continues on next page).

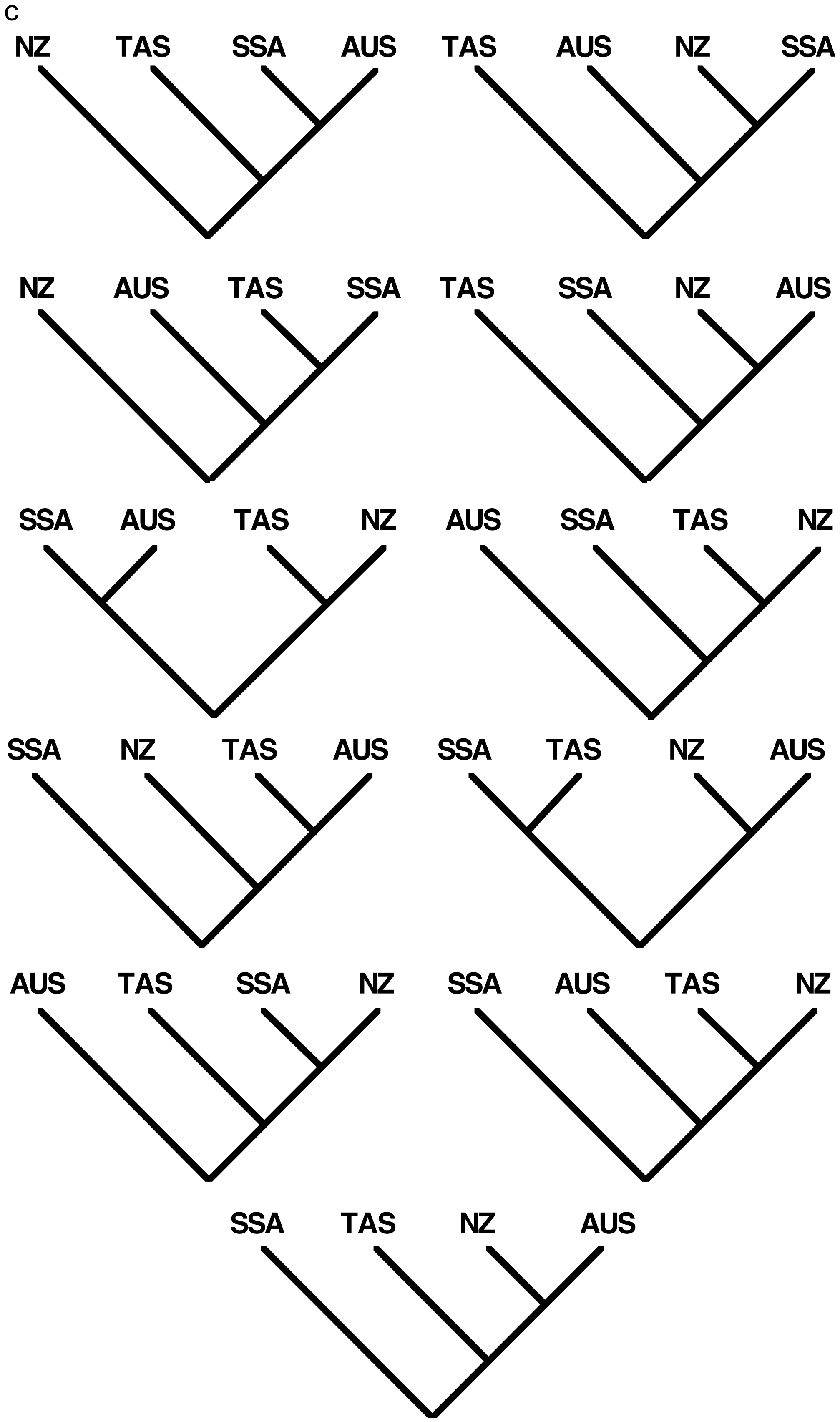


Fig. 12. Area cladogram derived with Component 1.5 under A0, A1 and A2 for Eriococcus/Madarococcus (a: area cladogram under A0, b: area cladograms under A1, c: area cladograms under A2).

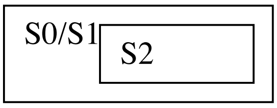
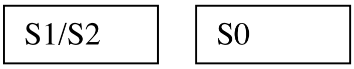
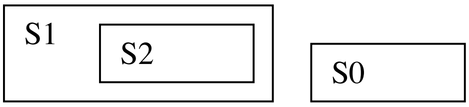
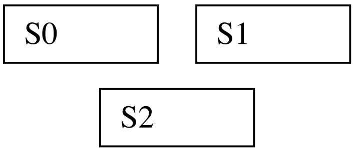
3.3.5 Component 2.0 (Reconciled Tree Analysis)

Theoretical data sets for single groups

RTA with Component 2.0 results in inclusive solution sets derived for the 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 fewest 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 violated.

With Component 2.0, inclusion is violated 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 smaller solution sets are often obtained under A2 than under A0 or A1, and consequently inclusion is violated. We show noninclusive results for the analysis of single group data sets in Table 4.

Table 4: Noninclusion 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	

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 needs 12 losses and 3 duplications) is obtained for *Heterandria* and *Xiphophorus* (Fig. 13a). Under A1 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 yields 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 violated (Requirement I).

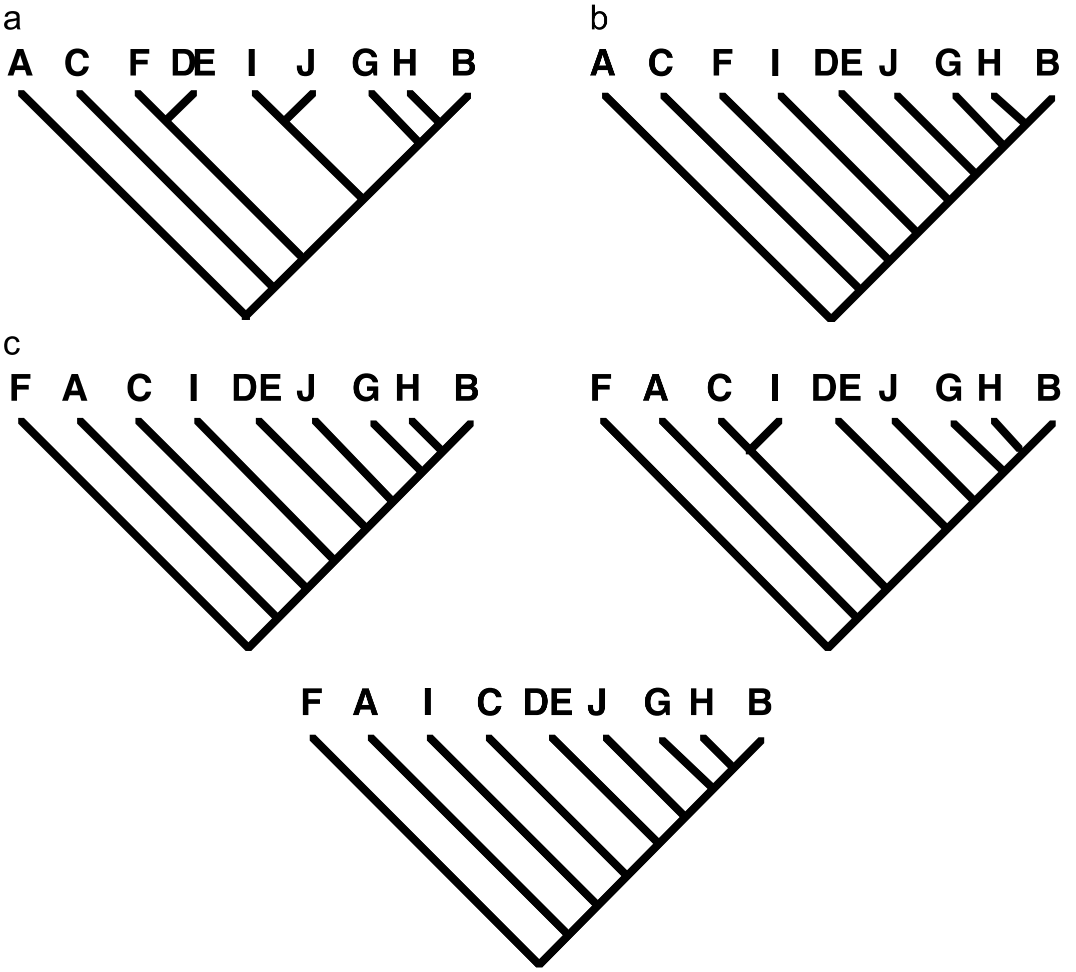


Fig. 13. General area cladograms derived with Component 2.0 under A0, A1 and A2 for Heterandria/Xiphophorus (a: general area cladogram under A0, b: general area cladogram under A1, c: general area cladograms under A2).

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) the least number 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, for each analysis 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 to remove 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 met.

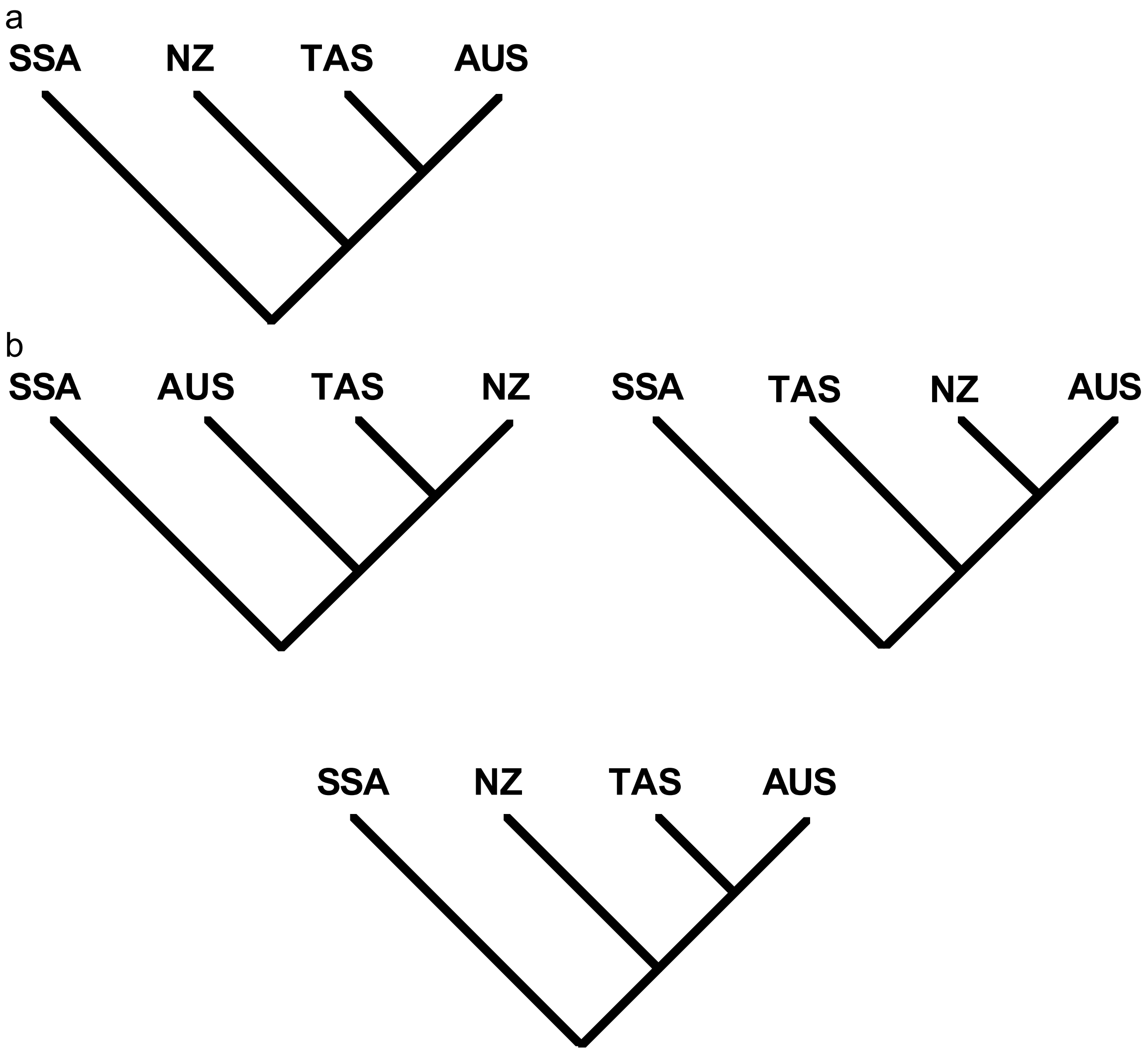


Fig. 14. General area cladograms derived with Component 2.0 under A0, A1 and A2 for Cyttaria and Eriococcus/Madarococcus. (a: general area cladogram under A0 and A1, b: general area cladograms under A2).

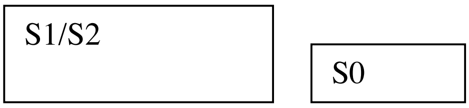
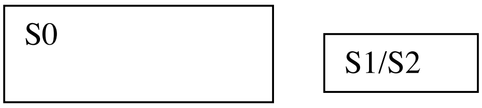
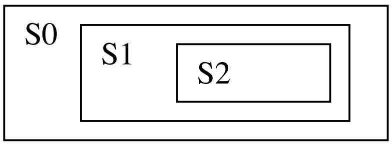
As with the data on *Heterandria*/*Xiphophorus*, general area cladograms for *Cyttaria* and *Eriococcus*/*Madarococcus* are obtained via reconciliations between area cladograms and taxon-area cladograms from both groups. Widespread taxa in *Cyttaria* and *Eriococcus*/*Madarococcus* are dealt with in the same way for each analysis (*i.e.* under a particular assumption) and therefore Requirement II is met.

3.3.6 TAS (Three Area Statement Analysis)

Theoretical data sets for single groups

In order to solve the problem of noninclusive 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).

Table 5: Noninclusion 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	

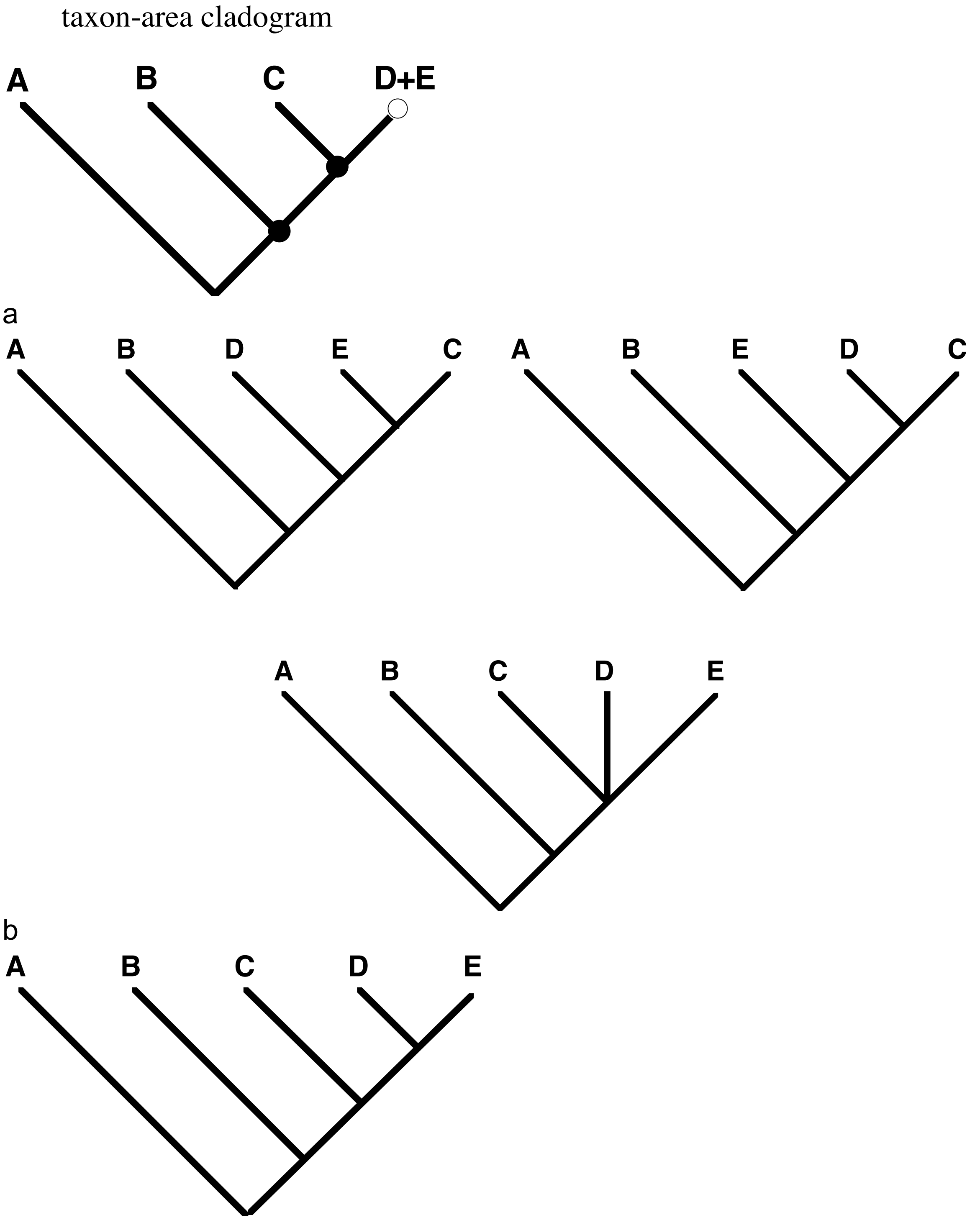


Fig. 15. Solving widespread taxa with TAS under A0 and A1/A2 (a: area cladograms under A1/A2, b: area cladogram under A0).

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 removing redundant areas *a priori* in favour of endemics. These areas are unable to float over the area cladogram and often fewer 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 violated with TAS.

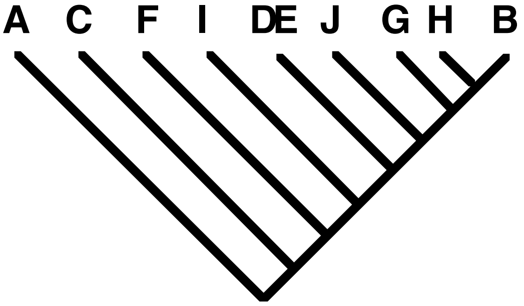


Fig. 16. General area cladogram derived with TAS under A0 and A1/A2 for Heterandria/Xiphophorus.

As 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 derived under A0. Therefore, recognition of three area statements under A0, A1, and A2 for this data set results in violation of Requirement I.

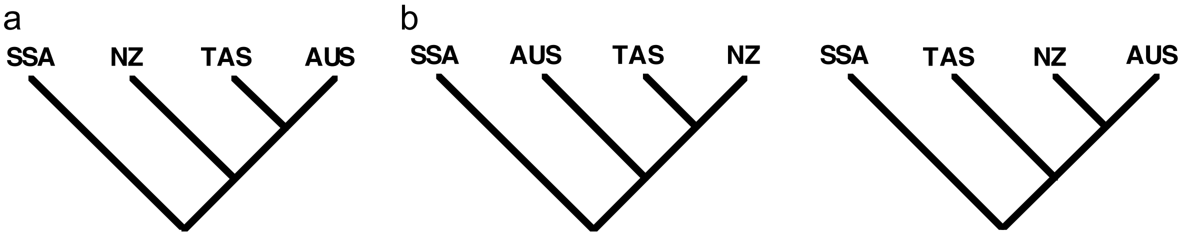


Fig. 17. General area cladograms derived with TAS under A0 and A1/A2 for Cyttaria and Eriococcus/Madarococcus. (a: general area cladogram under A0, b: general area cladograms under A1/A2).

General area cladograms for *Cyttaria and Eriococcus/Madarococcus* are obtained from a single data matrix. This matrix is derived under either 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).

3.4 Conclusions

In this study we assessed the precise implementation of A0, A1, and A2 in the computer programs of five different methods 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 derived 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: the computer programs of all methods obtain general area cladograms under a single assumption.

The computer 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 a 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. Under A2 each of these *a priori* methods prune taxa and taxon relationships from the taxon cladogram in such a way that the initial historical relationships of taxa as derived under A0 may become excluded, unavoidably resulting in noninclusion.

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”. By this we mean that no *a priori* steps should be taken in order 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*. With Component 1.5 this can be achieved by dealing with sympatric taxa under A0. With TAS, as we showed in this , 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 of Nelson and Ladiges, 1991a,c), sympatric taxa can be dealt with “as is” in TAS as well. For Component 2.0 we also recommend not dealing with sympatric taxa *a priori* by reducing widespread taxa in favour of endemics. However, in 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 to deal with sympatric taxa *a priori*. In the next chapter we will suggest additional procedures for dealing with widespread and sympatric taxa with Component 1.5, TAS, and Component 2.0 in such a way that violation of Requirement I is remedied.

In this chapter we have also shown that implementations of *a posteriori* methods do not avoid violation of Requirement I. Noninclusion 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 stricter 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 (NWO); 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. *Austral. 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.

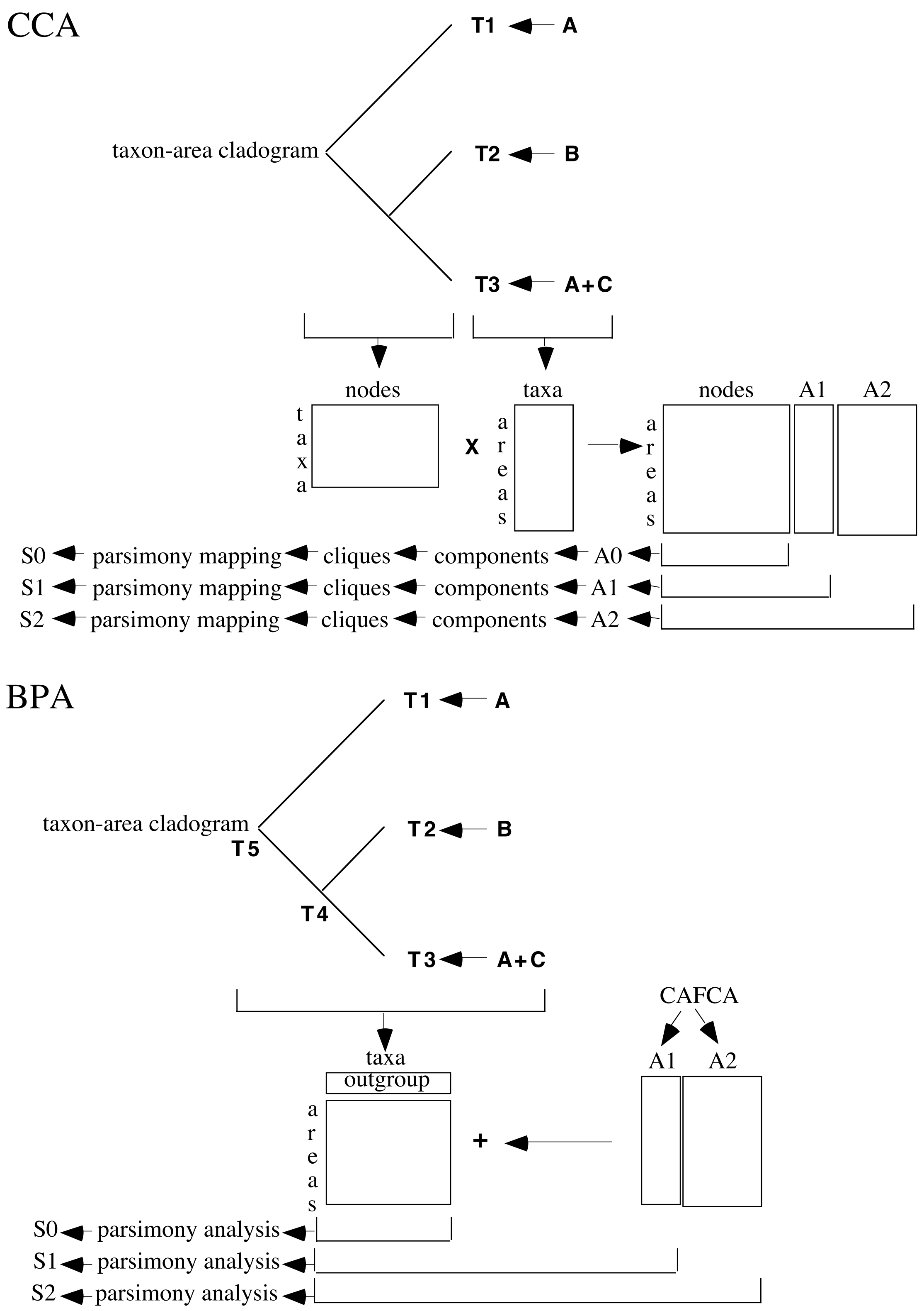
Wiley, E.O. (1988a). Parsimony analysis and vicariance biogeography. *Syst. Zool.* **37**, 271-290.

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

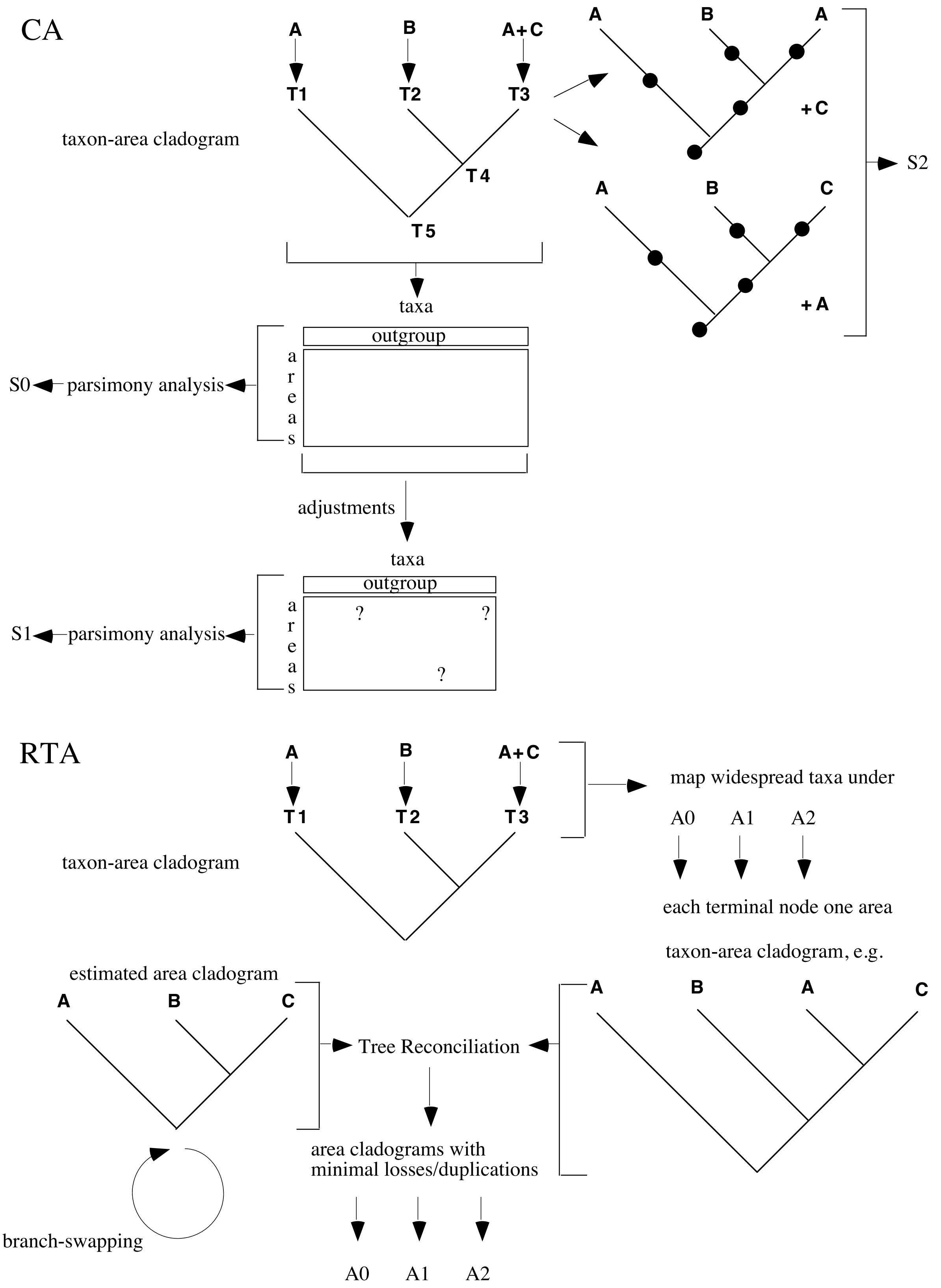
Zandee, M. (1999). CAFCA vs. 1.5j: a collection of APL functions for cladistic analysis. Program and user's manual. Institute of Evolutionary and Ecological Sciences, Leiden University. Leiden.

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

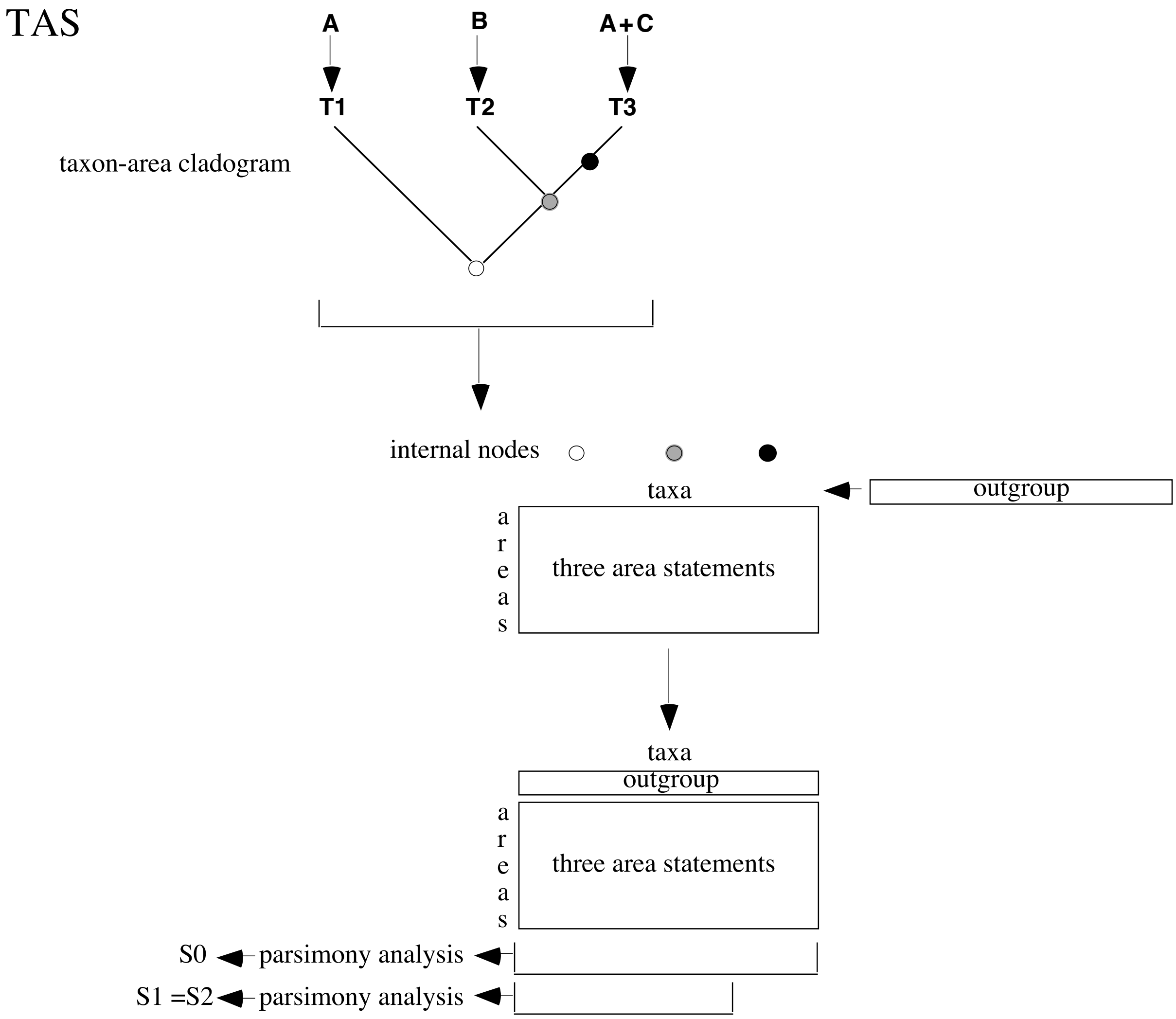
Appendices



Appendix i (continues on next page).



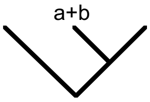
Appendix i (continues on next page).



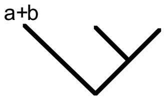
Appendix i.
Steps when applying implementations of CCA, BPA, CA, RTA and TAS to obtain area cladograms from cladogenetic and distribution data under different assumptions.

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

3 taxa

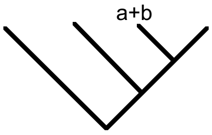


wid301

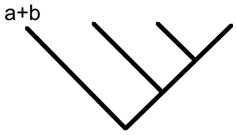


wid302

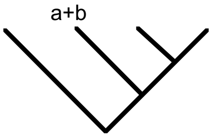
4 taxa



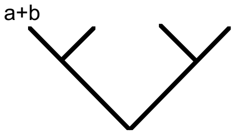
wid401



wid402

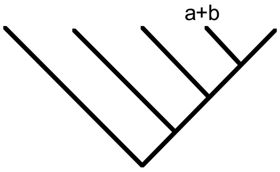


wid403

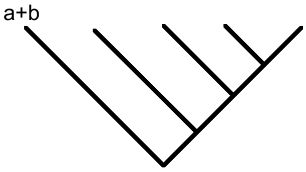


wid404

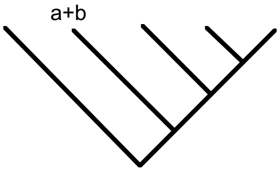
5 taxa



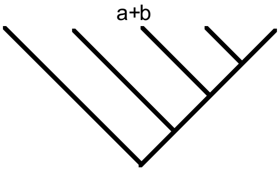
wid501



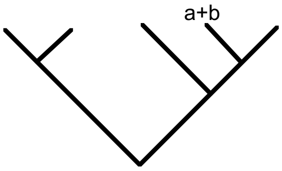
wid502



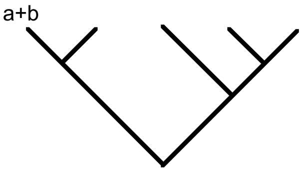
wid503



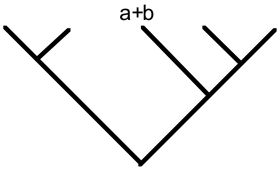
wid504



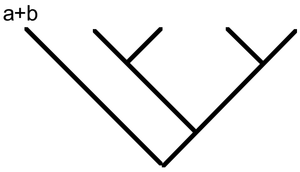
wid505



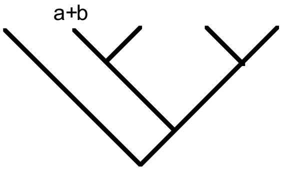
wid506



wid507



wid508

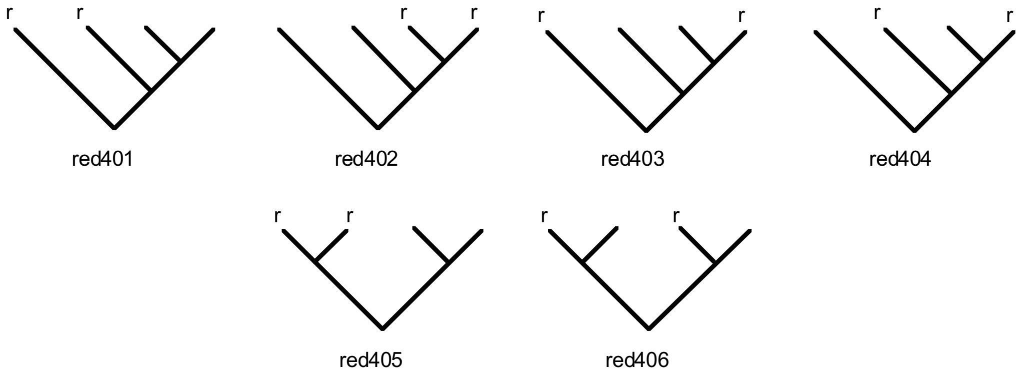


wid509

Appendix ii (continues on next page).

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

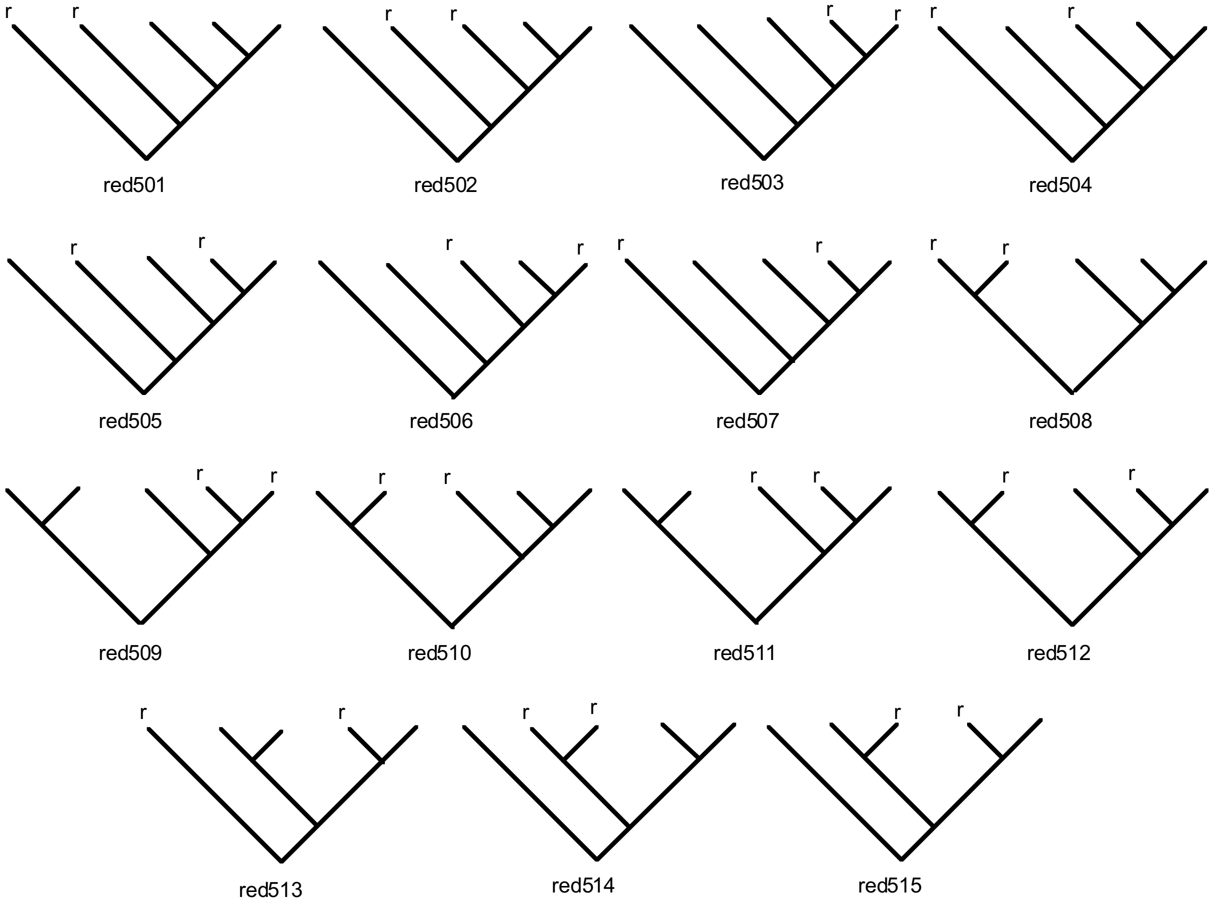
4 taxa



red401 red402 red403 red404

red405 red406

5 taxa



red501 red502 red503 red504

red505 red506 red507 red508

red509 red510 red511 red512

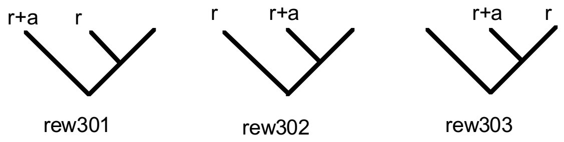
red513 red514 red515

Appendix ii (continues on next page).

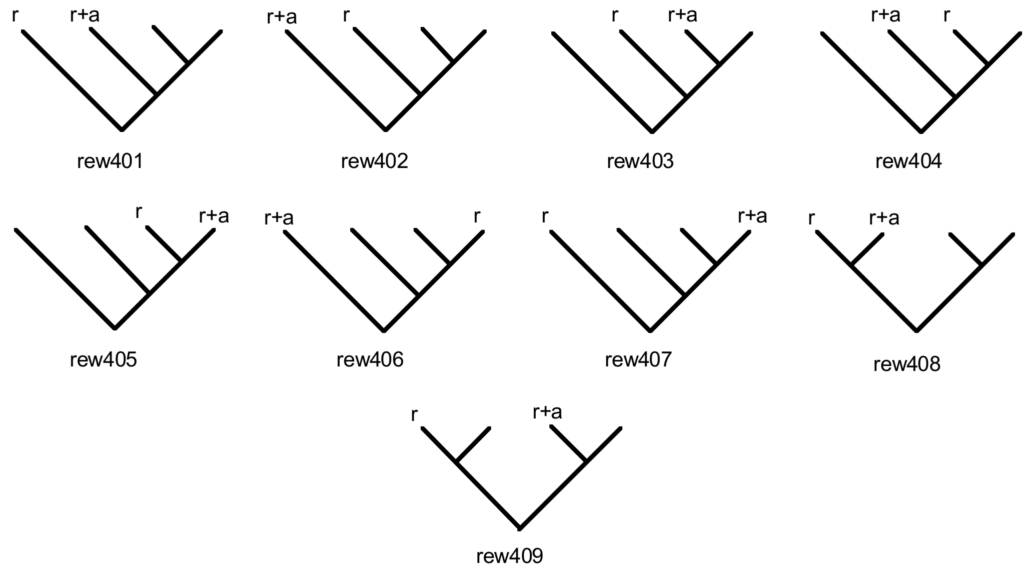
SYMPATRIC & WIDESPREAD TAXA

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

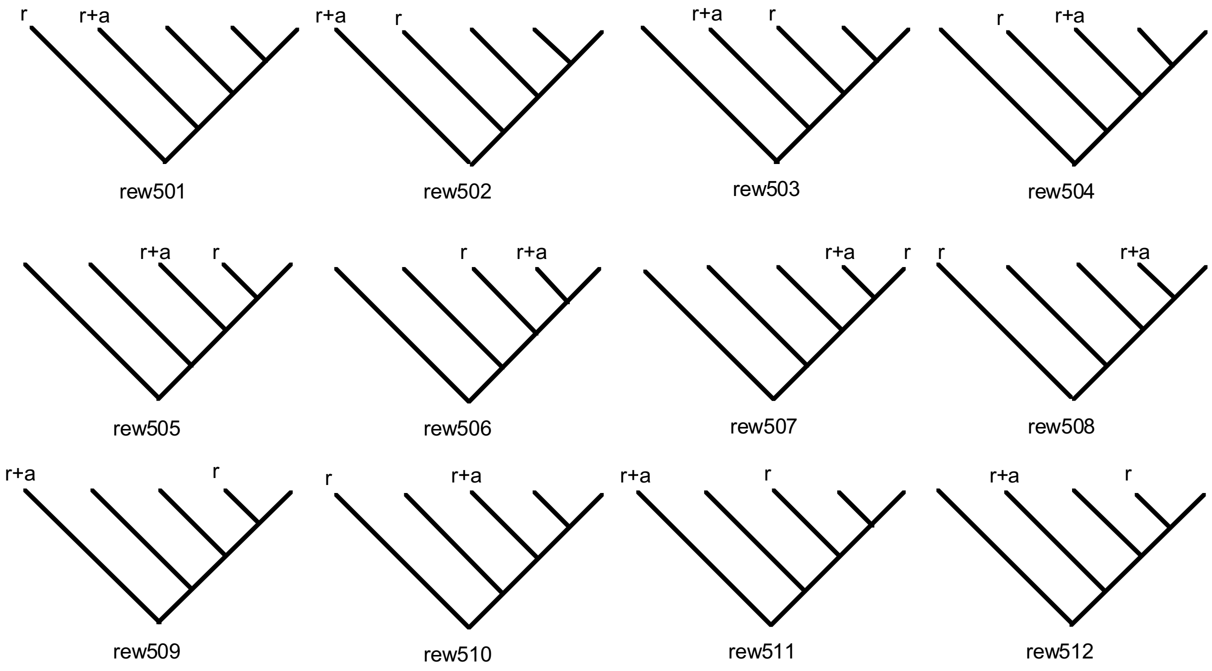
3 taxa



4 taxa



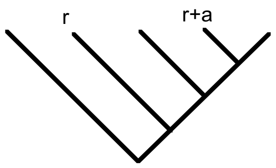
5 taxa



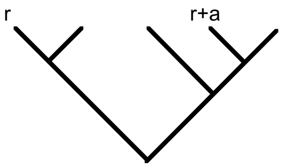
Appendix ii (continues on next page).

SYMPATRIC & WIDESPREAD TAXA (CONTINUATION)

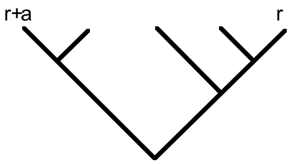
5 taxa



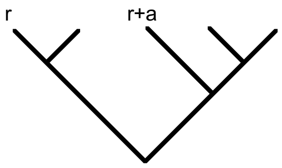
rew513



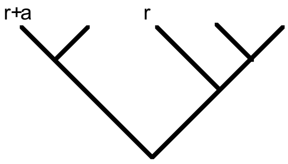
rew514



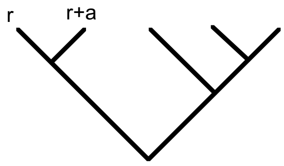
rew515



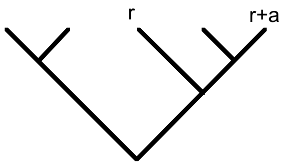
rew516



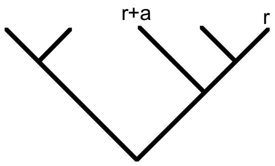
rew517



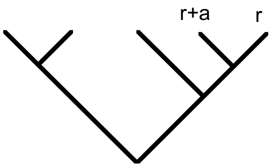
rew518



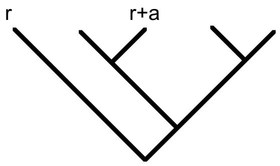
rew519



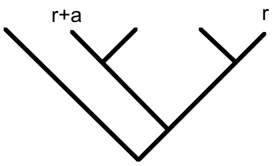
rew520



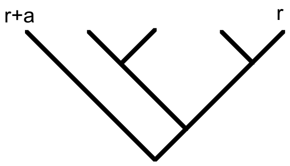
rew521



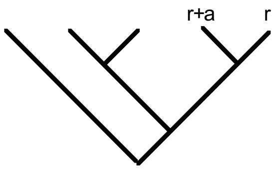
rew522



rew523



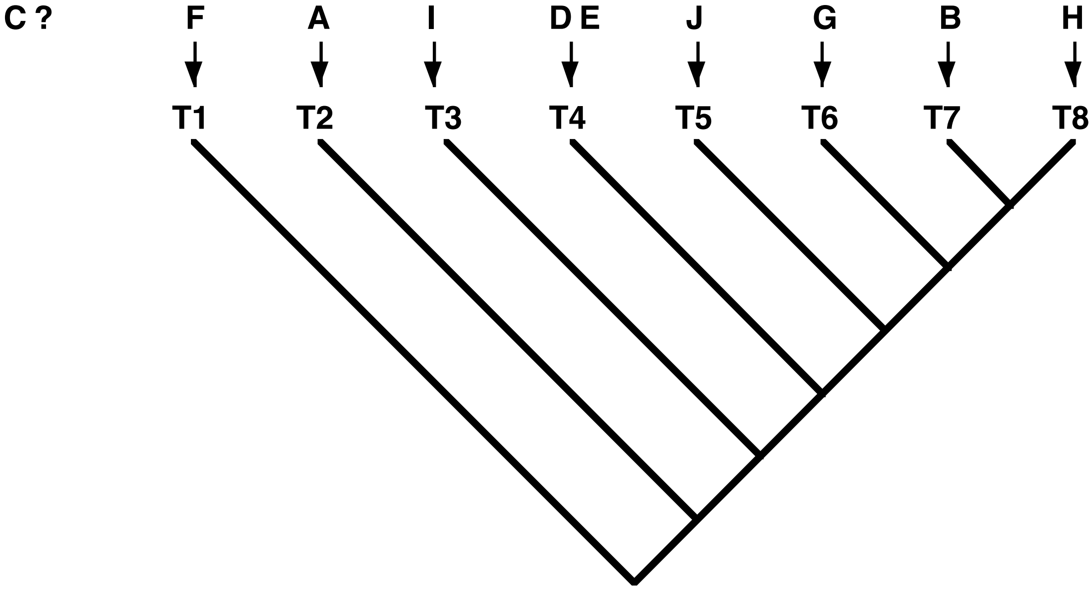
rew524



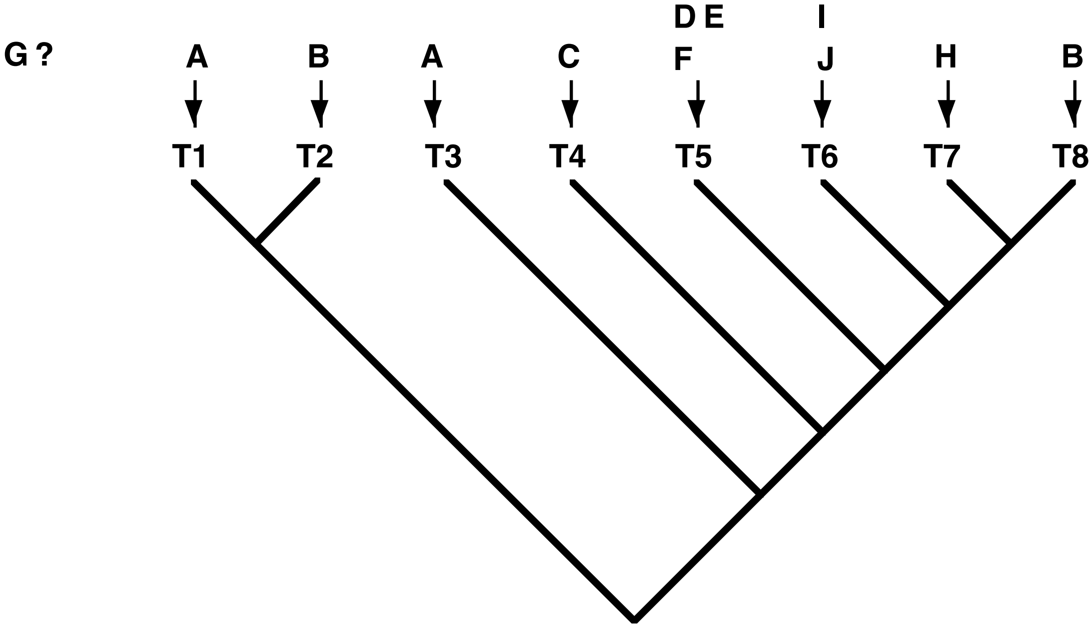
rew525

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).

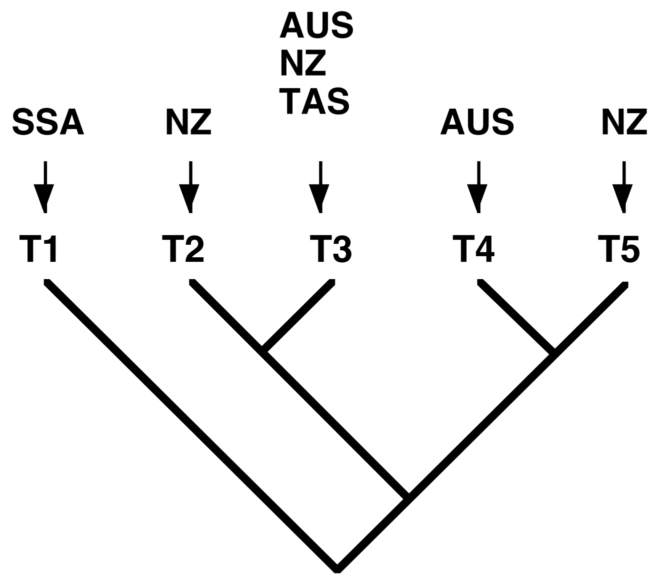
Heterandria



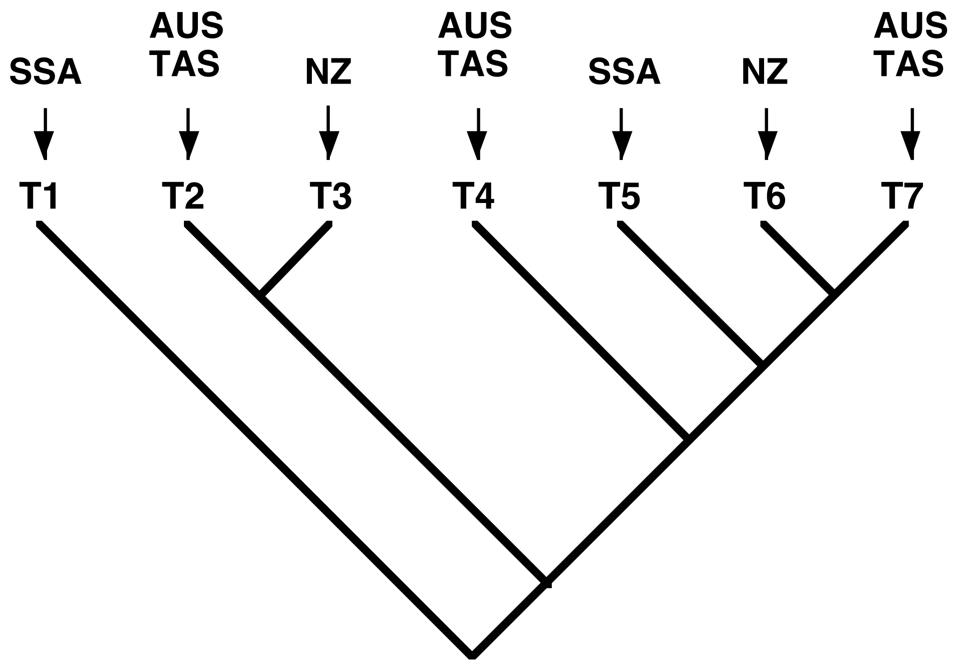
Xiphophorus



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



Cyttaria



Eriococcus/Madarococcus

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).

CHAPTER 4

MEASURES FOR OBTAINING INCLUSIVE SOLUTION SETS UNDER ASSUMPTIONS ZERO, 1, AND 2 WITH DIFFERENT METHODS FOR VICARIANCE BIOGEOGRAPHY

Abstract

We present modifications to computer programs of *a posteriori* and *a priori* methods for vicariance biogeography which enable them to obtain inclusive solution sets under Assumptions zero, 1, and 2. Application of an upper limit for area cladogram selection by the number of steps with CAFCA (Component Compatibility Analysis) is sufficient for inclusion only when the area cladograms in the solution sets derived under different assumptions are equally resolved. CAFCA needs a lower limit for the number of components used to derive area cladograms when noninclusion arises because CAFCA selects area cladograms with different resolutions under different assumptions. PAUP (Brooks Parsimony Analysis) derives inclusive solution 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 solution 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).

This chapter has been submitted in a modified form as: Veller, M.G.P. van, Zandee, M. & Kornet, D.J. Measures for obtaining inclusive solution sets under Assumptions zero, 1 and 2 with different methods for vicariance biogeography. *Cladistics*.

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 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. Two requirements are relevant when these assumptions are applied (Van Veller *et al.*, 1999). First, the solution sets obtained under the different assumptions should be inclusive, *i.e.*, $S0 \subseteq S1 \subseteq S2$ (Requirement I). Second, when we compare solution sets for different monophyletic groups of taxa in order to find a common pattern (*i.e.* a general area cladogram), these sets should be obtained under the same assumption (Requirement II).

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 the second requirement (Van Veller *et al.*, 2000). However, none of the programs assessed satisfies the first requirement.

In this chapter 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 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, 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 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).

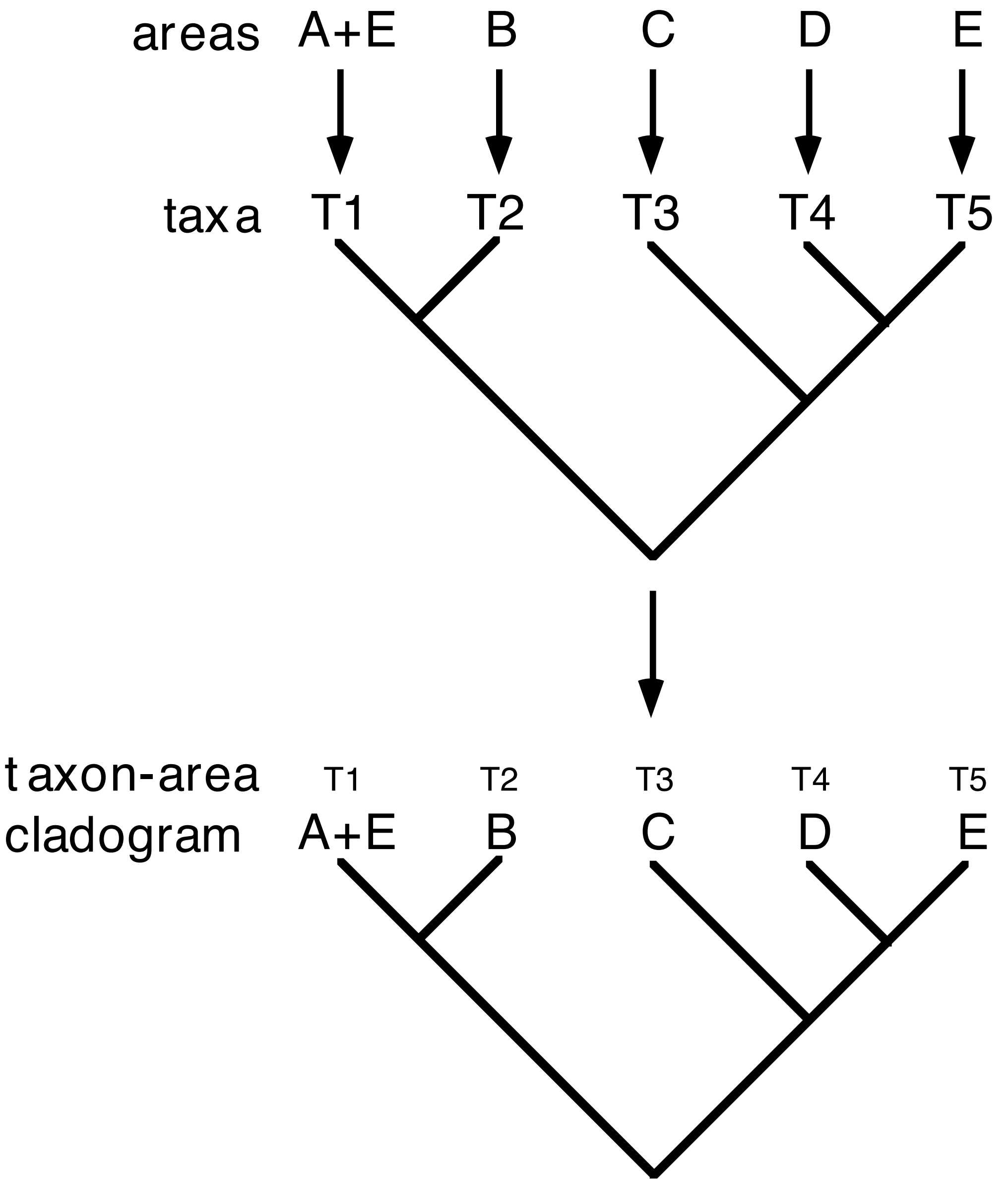


Fig. 1. Theoretical data set rew515 (A, B, C, D, E=areas; T1, T2, T3, T4, T5=taxa).

In the next part of this chapter 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. 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, 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 chapter, 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 inclusion.

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 (nos. 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.

*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

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 inclusion 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 chapter, 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 inclusion is violated.

To prevent violation of inclusion 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.

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

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 inclusion 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 chapter, 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 taxon-area cladograms that we obtain in the first step and the area cladograms that we obtain in the second step show inclusion ($S0 \subset S1 \subset S2$) and therefore Requirement I is met.

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))))

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 minimizing 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 (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 ($S0 \subset S1 = S2$) 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).

*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*			

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 an 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 minimizing 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 (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).

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

Both the taxon-area cladograms that we obtain in the first step and the area cladograms that we obtain in the second step show inclusion (S0 $\subset$ S1 = S2) 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 chapter) and Requirement II (established for these programs in Van Veller *et al.*, 2000) are met, we claim (Van Veller *et al.*, 1999) that the 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 programs for *a posteriori* methods (CAFCA and PAUP) sometimes result in noninclusion 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 secondly 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 program (*i.e.* the other bold numerals) are compared.

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**

Table 6 also compares solution sets obtained by 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 programs derive under the same assumption. As we show in Table 6, the 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 the next chapter 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. 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.

CHAPTER 5

A POSTERIORI AND *A PRIORI* METHODOLOGIES FOR TESTING HYPOTHESES OF CAUSAL PROCESSES IN VICARIANCE BIOGEOGRAPHY

Abstract

Methods used in vicariance biogeography fall into the categories of *a posteriori* methods (*e.g.*, Component Compatibility Analysis and Brooks Parsimony Analysis) and *a priori* methods (*e.g.*, Component Analysis, Reconciled Tree Analysis, and Three Area Statement Analysis). Each category corresponds to a particular methodology that arrives at general area cladograms by testing null hypotheses in a particular way. *A posteriori* methods assume the process of vicariance only (A0) as common cause of the distribution of different monophyletic groups of taxa under the null hypothesis. Whenever a parsimony analysis of combined data from these monophyletic groups results in a general area cladogram with homoplasy, the null hypothesis is rejected and extinction and dispersal are invoked *a posteriori* as *ad hoc* process explanations. *A priori* methods assume not only vicariance (A0) but also combinations of vicariance with the processes of extinction (A1) and dispersal (A2) as possible causes of the distribution of the taxa of different monophyletic groups. Each assumed set of processes corresponds to a different null hypothesis. Under the assumption of independence and thus additivity of the processes involved, the sets of area cladograms obtained from data of each monophyletic group must be inclusive (Requirement I). Whenever no congruent area cladograms are found in the intersection of sets of area cladograms derived under the same assumption for different monophyletic groups (Requirement II), the corresponding null hypothesis is rejected.

5.1 Introduction

Whenever a geographical or ecological barrier arises, species lineages distributed on either side of the barrier may be split into allopatric units (micro-species; Kornet, 1993a). When those splits become irreversible (permanent splits; Kornet, 1993b), vicariant speciation, or vicariance, has occurred. The geographical or ecological event forming the barrier is called a vicariance event. All species produced by vicariance share a property: the history of the areas or biota in which they occur mirrors their history of speciation. All monophyletic groups of species produced by the same vicariance events, therefore, share a general evolutionary history that parallels the history of the areas and biotas in which they occur. Historical biogeographers have taken advantage of this connection between vicariant speciation and earth history to classify biogeographic patterns into two groups: general patterns and unique patterns.

General patterns are those exhibited by multiple monophyletic groups. The general assumption (A0) in vicariance biogeography is that these general patterns have vicariance events as their common cause (Van Veller *et al.*, 1999).

In vicariance biogeography one looks for an explanation of the distribution of taxa of monophyletic groups over areas by the reconstruction of historical relationships between these areas (taxon history approach *sensu* Hovenkamp, 1997 or taxon relationship approach *sensu* Van Veller *et al.*, 1999). The pattern of these historical relationships is represented in an area cladogram with a unique representation of each area (Van Veller *et al.*, 1999).

On the basis of a first-order explanation of vicariance, several biogeographers have studied historical relationships between various areas for different monophyletic groups of taxa (*e.g.* Brooks and McLennan, 1991; Cracraft, 1986, 1988, 1994; Crisci *et al.*, 1991; Enghoff, 1995; Funk and Brooks, 1990; Green *et al.*, submitted; Humphries, 1982; Kluge, 1988; Ladiges, 1998; Linder and Crisp, 1995; Mayden, 1988; Morrone, 1993; Page and Lydeard, 1994; Rosen, 1978; Turner, 1996; Van Soest and Hajdu, 1997). From these empirical studies it appears that processes other than vicariance (extinction, dispersal) give rise to exceptions to the simple association between cladograms of monophyletic groups of taxa and the history of the areas in which they live.

The way to arrive at an area cladogram in vicariance biogeography is by replacing taxa in a taxon cladogram of a monophyletic group by their areas of endemism (Morrone and Crisci, 1995). If each taxon is present in only a single area (*i.e.* no widespread taxa are present) and each area is inhabited by only a single taxon (*i.e.* no sympatric taxa are present), this replacement straightforwardly produces a taxon-area cladogram (resolved area cladogram *sensu* Morrone and Carpenter, 1994; Enghoff, 1996; Van Veller *et al.*, 1999) that can be completely explained by vicariance. However, when sympatric or widespread taxa are present, replacement of taxa by areas results in taxon-area cladograms with either the same area at different terminal nodes (due to sympatric taxa) or more than one area at the same terminal node (due to a widespread taxon), and subsequent analysis to resolve the taxon-area cladogram is needed.

General area cladograms are derived from taxon-area cladograms by the application of different methods of analysis. Van Veller *et al.* (2000) divided these methods into two categories, depending on how they infer area cladograms from the taxon-area cladograms under three sets of process assumptions, dubbed Assumption zero (A0: vicariance), Assumption 1 (A1: vicariance + extinction), and Assumption 2 (A2: vicariance + extinction + dispersal) (see also Zandee and Roos, 1987; Wiley, 1988a,b, and Nelson and Platnick, 1981).

A posteriori methods include Component Compatibility Analysis (CCA; Zandee and Roos, 1987) and Brooks Parsimony Analysis (BPA; Brooks, 1990; Wiley, 1988a,b; modified for analyses under A1 and A2 in Van Veller *et al.*, 2000). They originally deal with widespread taxa under A0 but can also deal with widespread taxa under A1 and A2 via increasing and inclusive degrees of freedom and without making any *a priori* process assumptions (other than vicariance) for explanation. They deal with sympatric taxa (under all assumptions) by combining all data and explaining incongruences *a posteriori* by invoking extinction or dispersal *ad hoc*.

A priori methods include Component Analysis (CA; Nelson and Platnick, 1981; Page, 1988, 1990a,b), Reconciled Tree Analysis (RTA; Page, 1993, 1994), and Three Area Statement Analysis (TAS; Nelson and Ladiges, 1991a,b). They deduce from the *a priori* process assumption(s) the modifications in the data to deal with widespread or sympatric taxa. These *a priori* modifications may overrule the historical relationships established in the taxon cladograms. The modifications are deduced from the assumption under consideration.

The critical difference between *a posteriori* and *a priori* methods lies not in their implementations in software packages but in the different protocols that they use to derive area cladograms from cladogenetic and distributional data. For instance, Brooks Parsimony Analysis (*a posteriori* method by Brooks, 1990) allows no *a priori* modifications of the data to deal with widespread or sympatric taxa. Parsimony analysis with PAUP (Swofford, 1990) or Hennig86 (Farris, 1988) obtains area cladograms from a data matrix that directly represents the distributional and cladogenetic relationships of the data. Three Area Statement Analysis (*a priori* method by Nelson and Ladiges, 1991a,b) also uses parsimony analysis for obtaining area cladograms. However, when dealing with widespread or sympatric taxa, Three Area Statement Analysis allows *a priori* modifications of the data (depending on the assumption under which the analysis is performed) via the derivation of a matrix with three area statements. Due to this difference in methodological protocols that either forbid or allow *a priori* modifications of the data before using the *same* implementation (*i.e.* PAUP or Hennig86) for parsimony analysis, the methods TAS and BPA are categorized as *a priori* and *a posteriori* methods respectively.

The methods mentioned so far have also been referred to as pattern-based methods (Ronquist and Nylin, 1990). Both *a posteriori* and *a priori* methods derive divergent patterns (area cladograms) based on a first-order explanation of vicariance and subordinate explanations by extinction or dispersal. In addition to these pattern-based methods, event-based methods have been proposed by Ronquist (1997, 1998) and Ronquist and Nylin (1990). In event-based methods, (reticulate) biogeographical scenarios are analysed by assigning different costs to different processes (vicariance, dispersal, extinction, and duplication). In this chapter we restrict ourselves to pattern-based methods.

Van Veller *et al.* (1999) developed two criteria for consistency of all (pattern-based) methods with respect to their implementation of the assumptions and their capacity for finding general area cladograms for different monophyletic groups. First, Van Veller *et al.* (1999) claimed that if one *a priori* assumes inclusive sets of independent -and thus additive- processes under A0, A1, and A2, this should result in inclusive sets of area cladograms under these assumptions (Requirement I). Second, Van Veller *et al.* (1999) argued that these sets of area cladograms obtained for different monophyletic groups should be compared under the same assumption to obtain valid general area cladograms (Requirement II).

After assessing the *a posteriori* and the *a priori* methods for the extent to which they satisfy these two requirements, Van Veller *et al.* (2000) find that none of the methods show any problems with respect to Requirement II, but that all the methods may violate Requirement I under certain circumstances.

Van Veller *et al.* (submitted) suggested modifications to the various methods relating to area cladogram selection and succeeded in meeting both Requirements I and II for both *a posteriori* and *a priori* methods. After these modifications are applied, all methods show consistent behavior with respect to the (independent) processes assumed *a priori* and the derivation of general area cladograms. Even after correcting all methods so that they meet Requirements I and II, *a posteriori* and *a priori* methods may still select different (general) area cladograms (Van Veller *et al.*, submitted). We think that these differences between the (general) area cladograms obtained with *a posteriori* and *a priori* methods indicate that the two categories represents different methodologies. These methodologies arrive at general area cladograms by different ways of testing hypotheses of causal processes for taxa of different monophyletic groups. The general area cladograms are explained by processes such as vicariance, extinction, and dispersal in different ways by the two categories.

In this chapter we discuss the differences between *a posteriori* and *a priori* methods in the testing of these hypotheses. We examine the need for A0, A1, and A2 in both *a posteriori* and *a priori* methods and see how these assumptions relate to the need for Requirements I and II for both categories of methods.

5.2 The *a posteriori* methodology for vicariance biogeography

A posteriori methods include Component Compatibility Analysis (CCA; Zandee and Roos, 1987) and Brooks Parsimony Analysis (BPA; Brooks, 1990; Wiley, 1988a,b).

5.2.1 Process assumptions

In *a posteriori* methods the null assumption (A0) (comparable to the null model *sensu* Brooks and McLennan, 1991; Ronquist, 1997) is that the distribution of the taxa of a monophyletic group over areas is caused by the process of vicariance. In this respect, the *a posteriori* methodology does not differ from the *a priori* methodology. Contrary to the *a priori* methods, the *a posteriori* methods do not need to assume extra processes of extinction and dispersal (see below).

5.2.2 Formulating the null hypothesis

Under the null assumption (A0), the null hypothesis is that the cladogenetic relationships between taxa of several monophyletic groups (represented in their taxon cladograms) and their distributions over the same areas result in a single general area cladogram without homoplasy (Fig. 1). This single general area cladogram without any homoplasious components represents the pattern of the historical relationships between areas.

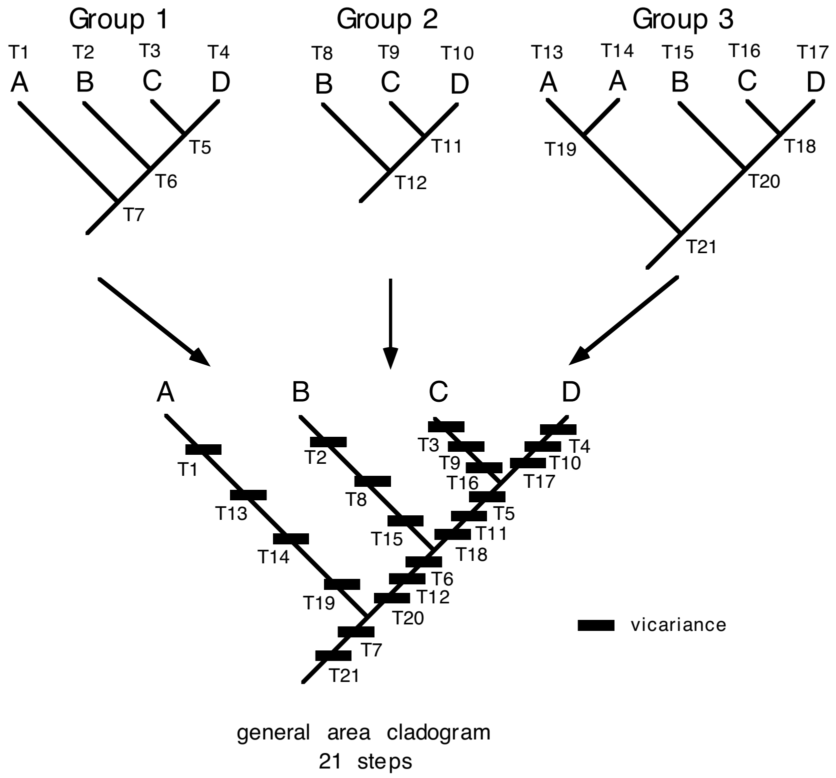


Fig. 1. No rejection of the null hypothesis when parsimony analysis with a posteriori methods results in a general area cladogram without homoplasy (T1 - T21 = taxa; A - D = areas).

5.2.3 Testing the null hypothesis

Data on several monophyletic groups of taxa are provided by the taxon cladograms of all monophyletic groups and the distribution of the taxa of each monophyletic group over the areas (Fig. 2). These data are used to derive an area-by-node data matrix (area-data matrix), either directly or indirectly. The direct derivation is by computing the Boolean inner product of the matrix representing the distribution of the taxa over the areas concerned and the matrix representing the cladogram of the taxa (Zandee and Roos, 1987). The indirect derivation of the area-data matrix uses an intermediate taxon-area cladogram. A taxon-area cladogram is derived from each taxon cladogram of a monophyletic group by replacing the taxa by the areas in which these taxa are distributed. Each taxon-area cladogram of a particular monophyletic group is translated into an area-data matrix via inclusive Oring (O'Grady and Deets, 1987) or additive binary coding with question marks standing in for missing data (Wiley, 1988a,b; Brooks, 1990; Brooks and McLennan, 1991) (Fig. 2).

The *a posteriori* methods CCA (Zandee and Roos, 1987) and BPA (Brooks, 1990) combine the area-data matrices of several monophyletic groups in one combined area-data matrix (Fig. 2). This combination of area-data matrices is analogous to a total evidence approach in systematic phylogenetics (Kluge, 1989, 1998). From the combined area-data matrix, area cladograms are derived via parsimony analysis. The most parsimonious general area cladogram is selected.

If all components are congruent with this general area cladogram, the null hypothesis is not rejected and the distribution of taxa over areas is explained by vicariance. If at least one component is not congruent with the general area cladogram, *i.e.* the component is homoplasious, the null hypothesis is rejected. Homoplasious components are explained *a posteriori* by *ad hoc* process explanations such as extinction or dispersal. In Fig. 2 we show some homoplasious components in the general area cladogram that is obtained with CCA or BPA from the data of three groups. The two parallelisms (open box and circle) are interpreted as two dispersal events of taxon T1 from area A to area C and of taxon T13 from area A to area D. The two reversals and single reversal (grey ovals) obtained by BPA and CCA resepectively are interpreted as two (taxa T11 and T12) and one (taxon T11) extinctions in area C for BPA and CCA respectively.

5.2.4 Why A0 is sufficient for *a posteriori* methods and Requirements I and II therefore do not apply

In the *a posteriori* methods CCA and BPA the area-data matrices of all monophyletic groups distributed in the same areas are combined in one area-data matrix. Parsimony analysis on this combined area-data matrix, always results in one (or more) general area cladograms. In cases where the null hypothesis is not rejected, the common cause explanation is vicariance. *A posteriori ad hoc* explanation of the distribution of taxa of monophyletic groups by processes of extinction or dispersal is needed whenever the null hypothesis is rejected. However, these extra processes (extinction or dispersal) are not assumed *a priori* in arriving at a general area cladogram. As a result, the data for all monophyletic groups are always analyzed under the same assumption (*i.e.* A0) and Requirement II is thus always met (see also Van Veller *et al.*, 2000). The *a posteriori* methods do not (necessarily) obtain sets of area cladograms under A1 or A2 (but only under A0 *sensu* Zandee and Roos, 1987; Wiley, 1988a,b) and, therefore, Requirement I does not apply.

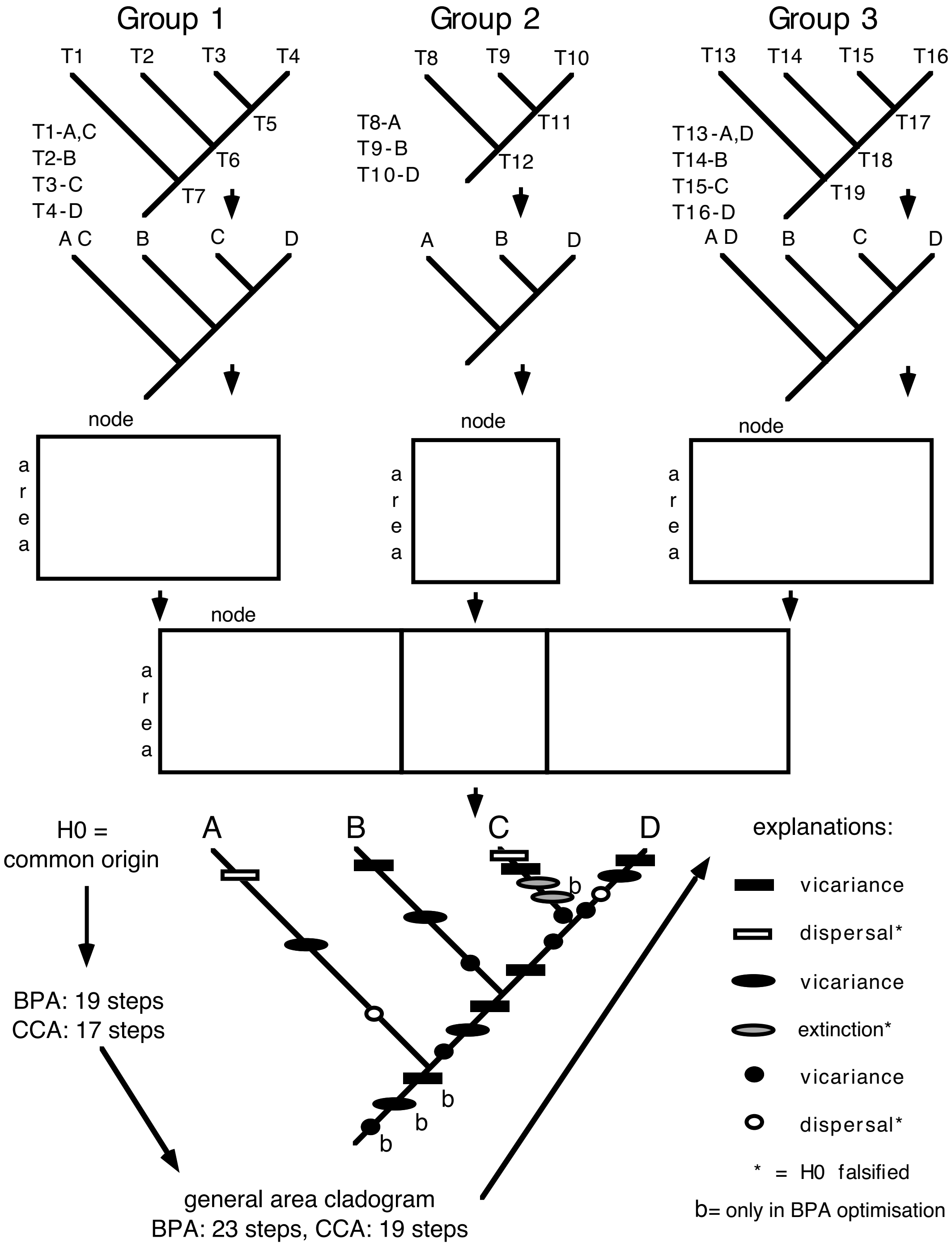


Fig. 2. A framework for testing the null hypothesis (that the distribution of taxa of several monophyletic groups over the same areas is caused by vicariance) with a posteriori methods and ad hoc explanations when this null hypothesis is rejected (by extinction or dispersal) (T1 - T19 = taxa; A - D = areas).

5.3 The *a priori* methodology for vicariance biogeography

A priori methods include Component Analysis (CA; Nelson and Platnick, 1981; Page, 1988, 1990a,b), Reconciled Tree Analysis (RTA; Page, 1993, 1994), and Three Area Statement Analysis (TAS; Nelson and Ladiges, 1991a,b).

5.3.1 Process assumptions

In *a priori* methods the null assumption (A0) is that the distribution of the taxa of a monophyletic group over areas is caused by the process of vicariance. In this respect the *a priori* methodology does not differ from the *a posteriori* methodology. Contrary to the *a posteriori* methods, the *a priori* methods assume combinations with the extra processes of extinction (A1) and dispersal (A2) (see below and in Table 1).

Table 1: A priori assumed processes under different null hypotheses formulated under Assumptions zero, 1, and 2 to explain the distributions of taxa of monophyletic groups with a priori methods.

Assumption	Null hypothesis: distribution is caused by the processes of:
A0	Vicariance
A1	Vicariance + Extinction
A2	Vicariance + Extinction + Dispersal

5.3.2 Formulating and testing hypotheses

In the application of the *a priori* methodology under each assumption zero, 1, and 2, a different null hypothesis is formulated. Thus, rather than using a single null hypothesis (as *a posteriori* methods do), *a priori* methods use different null hypotheses depending on the assumption applied.

Under A0 the null hypothesis states that the cladogenetic relationships between taxa of several monophyletic groups (represented in their taxon cladograms) and their distributions over the same areas result in a general area cladogram (Fig. 1). This general area cladogram is the result of comparing the solution sets of different monophyletic groups for congruent area cladograms. If the intersection of the solutions sets (Fig. 3; S_i0) of the different groups is empty (no congruence), the null hypothesis is rejected.

If the null hypothesis under A0 is rejected, a new null hypothesis of congruence among solution sets is formulated under A1. Under this new null hypothesis extinction is added to vicariance in the set of assumed causes of the distribution of the taxa of monophyletic groups over areas. By dealing with widespread and sympatric taxa via modifications to the data deduced from A1, a larger set of area cladograms (S_i1) is obtained (Fig. 3). The two sets of area cladograms are again compared for congruent area cladograms and the null hypothesis (under the assumption that the distribution was caused by vicariance and extinction) is rejected if none is found.

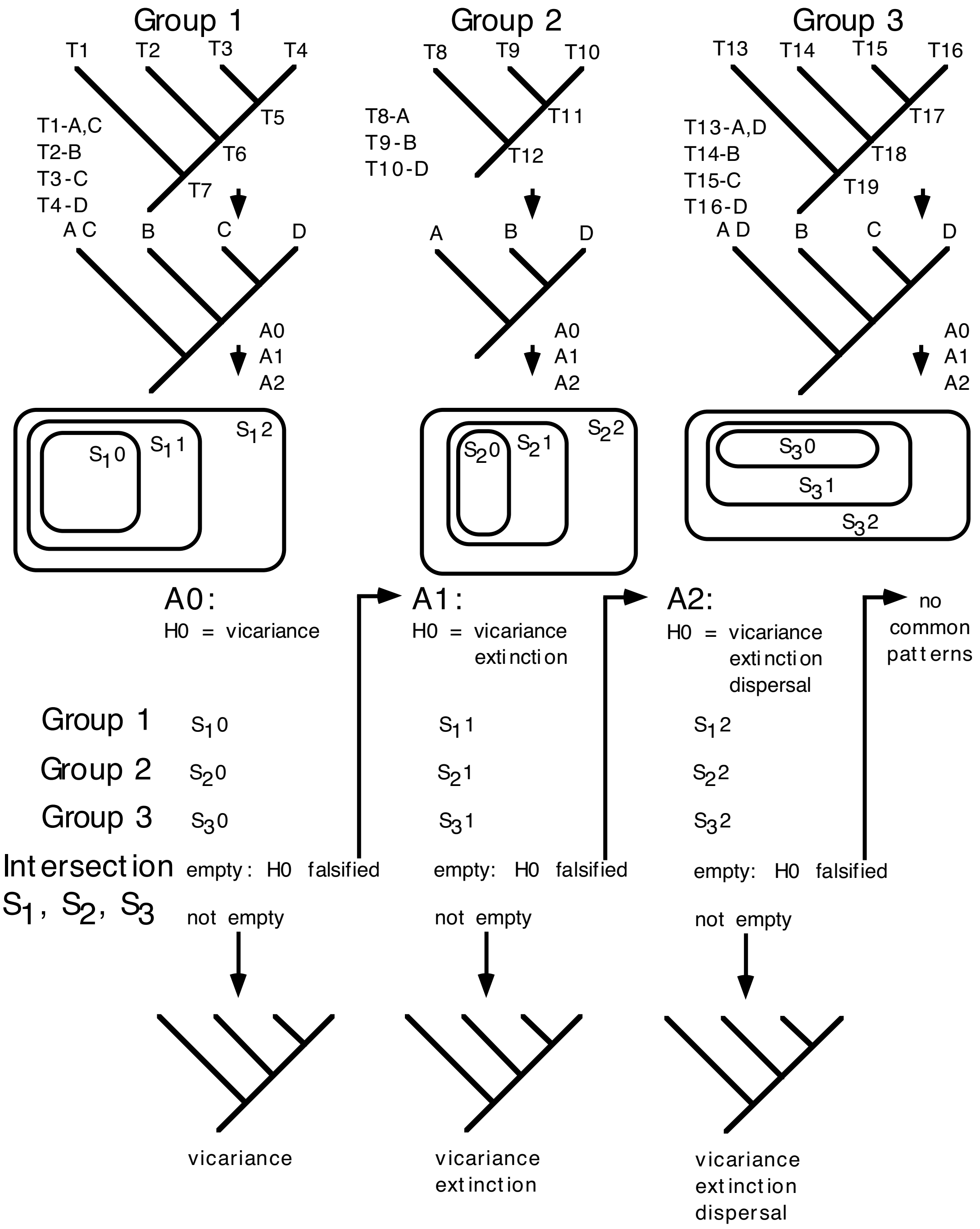


Fig. 3. A framework for testing hypotheses (that the distribution of taxa of several monophyletic groups over the same areas is caused by vicariance, vicariance or extinction, vicariance or extinction or dispersal) with a priori methods (T1 - T19 = taxa; A - D = areas; S_i0, S_i1, and S_i2 = sets of area cladograms derived under A0, A1, and A2 for monophyletic group i).

Finally, if the null hypothesis under A1 is rejected, a new null hypothesis of congruence among solution sets is formulated under A2. Under this new null hypothesis dispersal is added to vicariance and extinction in the set of assumed causes of the distribution of the taxa of monophyletic groups over areas. Under A2, widespread and sympatric taxa are dealt with in such a way that modifications to the data allowed by *a priori* assumption of vicariance, extinction or dispersal result in sets of area cladograms (S_i2) for each monophyletic group (Fig. 3). These sets are again compared for congruent area cladograms that are considered to represent general patterns. However, if no congruent area cladograms are found, the null hypothesis of a general pattern of historical relationships between areas as explained by vicariance, extinction or dispersal in the taxa of all monophyletic groups is rejected. With this rejection of the null hypothesis that encompasses all these processes, the possibility of finding any general pattern for all monophyletic groups is ruled out.

5.3.3 Why and when Requirements I and II apply in the *a priori* methodology

Van Veller *et al.* (submitted) suggested modifications to the *a priori* methods CA, RTA, and TAS that ensure that they obtain inclusive sets of area cladograms under A0, A1, and A2. By dealing with widespread taxa *a priori* under the different assumptions and dealing with sympatric taxa either via tree reconciliation (RTA) or by an *a posteriori* approach (CA and TAS as in CCA and BPA), the sets of area cladograms expand inclusively when analyses are performed under A0, A1, and A2.

Under the different assumptions, different null hypotheses are formulated which are tested by looking for congruence in solution sets of area cladograms that result from the *a priori* assumption of the different sets of causal processes (vicariance; vicariance + extinction; vicariance + extinction + dispersal).

Van Veller *et al.* (1999) showed that under the assumption of the independence of these processes, and thus the additivity of the results obtained, Requirements I and II must be met if any comparison of the solution sets obtained has to be valid.

Fig. 4 shows sets of area cladograms derived under A0 (S_10 and S_20), A1 (S_11 and S_21), and A2 (S_12 and S_22) for two monophyletic groups of taxa. As we showed above, general area cladograms for these two monophyletic groups of taxa are found in the intersection of the sets obtained under the same assumption (Requirement II).

Fig. 4a shows two intersections of sets (the sets derived under A1 and those derived under A2). Despite the fact that Requirement II is met, it is not clear whether the general area cladogram(s) must be selected under A1 or A2. Neither the null hypothesis under A1 (vicariance + extinction assumed) nor the null hypothesis under A2 (vicariance + extinction + dispersal assumed) can be rejected. If, however, the processes are assumed to be dependent on each other, and therefore to be non-additive in their results, an external criterion for rejection is needed to choose between the general area cladograms. If, however, the processes are assumed to be independent, and therefore additive in their results, the solution sets represented in Fig. 4a do not show inclusion and violate Requirement I.

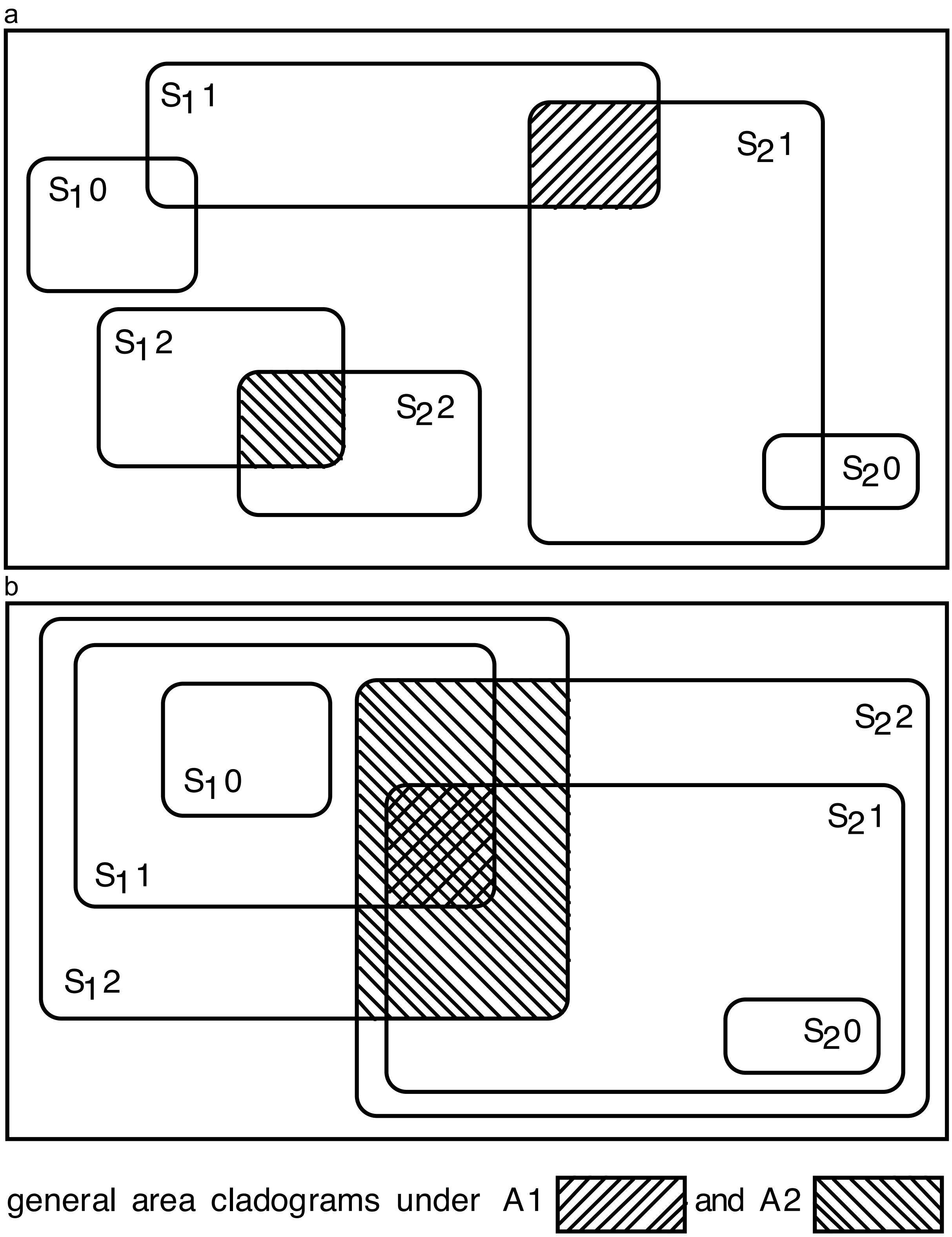


Fig. 4. The case for Requirement I with a priori methods (a: sets of area cladograms and general area cladograms (in intersections) when Requirement I is violated, b: sets of area cladograms and general area cladograms (in intersections) when Requirement I is met) (S_i0, S_i1, and S_i2 = sets of area cladograms derived under A0, A1, and A2 for monophyletic group i).

In Fig. 4b we show again two intersections of sets (the sets derived under A1 and the sets derived under A2). Requirement II is met in this example too nand general area cladograms can be selected under A1 or A2. The null hypotheses cannot be rejected either under A1 (vicariance + extinction assumed) or under A2 (vicariance + extinction + dispersal assumed). If one wishes to choose the general area cladogram(s) that is (are) explained by the smallest set of processes, one will select the general area cladogram(s) under A1 in Fig. 4b. However, because of inclusion of sets of area cladograms (Requirement I), these general area cladogram(s) are selected under A2 also. By selecting these general area cladogram(s) (derived under both A1 and A2) and by giving explanations by vicariance and extinction (under A1) rather than by vicariance, extinction and dispersal (under A2) one gains explanatory power (*i.e.* explanation by fewer processes for the same general area cladogram(s)).

5.4 Conclusions

Above we have discussed a framework under which *a posteriori* and *a priori* (pattern-based) methods test null hypotheses of causes of the distribution of several monophyletic groups of taxa over areas. Both *a posteriori* and *a priori* methods derive area cladograms on the basic idea that cladogenetic and distributional data of monophyletic groups represent historical (divergent) relationships between areas. However, they formulate null hypotheses and test those null hypotheses in very different ways.

A posteriori methods such as CCA and BPA do not make any *a priori* process assumptions beyond A0. Under the null hypothesis (vicariance events as common cause for the distribution of taxa of different monophyletic groups) they assume that the cladogenetic relationships of the taxa of several monophyletic groups are associated with the historical relationships of the areas. For all monophyletic groups of taxa these methods derive area-data matrices which they combine in one matrix. Parsimony analysis of this combined area-data matrix results in a general area cladogram in which homoplasy forces a rejection of the null hypothesis. Whenever the null hypothesis is rejected, *a posteriori* methods use *ad hoc* explanations in terms of extinction or dispersal to explain the deviation from vicariance.

As a result, we have shown that *a posteriori* methods test null hypotheses and derive general area cladograms without using A1 or A2. A0 suffices and both A1 and A2 are superfluous. Neither Requirement I nor Requirement II (Van Veller *et al.*, 1999) can therefore be violated by *a posteriori* methods.

A priori methods such as CA, RTA and TAS use A0 and the extra Assumptions 1 and 2 to deal with widespread and sympatric taxa *a priori*.

In this chapter we discuss a general framework for all *a priori* methods for testing null hypotheses derived under different assumptions in a consistent way. By comparing the sets of area cladograms derived for different monophyletic groups under the same assumption and searching for congruent patterns in the intersection of the different solutions sets, general area cladograms are found. If general area cladograms are found, the null hypothesis is not rejected and the set of processes assumed *a priori* appears to be sufficient for an explanation of the distribution of all taxa of the different monophyletic groups over areas. However, if no general area cladograms can be found, a different null hypothesis must be formulated under another assumption.

In this framework of testing null hypotheses by searching for congruence among area cladograms, Requirement II (Van Veller *et al.*, 1999) should not be violated. Valid general area cladograms are only found for several monophyletic groups when the same null hypothesis under the same assumption (either A0, A1, or A2) for all groups is not rejected. Requirement I (Van Veller *et al.*, 1999), under the assumption of independence of processes, and thus additivity of results, is necessary to enable testing (and possible rejection) of the null hypothesis.

Requirement I is not an optimality criterion like minimal items of error (CA), minimal number of losses/duplications (RTA) or minimal number of steps (TAS). These optimality criteria describe difference in quality of area cladograms within a particular solution set, derived under A0, A1, or A2 for a single monophyletic group of taxa. Requirement I is a methodological necessity over different solution sets (that are derived for a single monophyletic group of taxa under A0, A1 and A2) when independence of the causal processes (vicariance, extinction or dispersal) is assumed. Requirement I thereby does not compete with the description of quality of area cladograms within a particular solution set.

In this chapter we show how *a posteriori* and *a priori* methodologies differ. Each is caplable of deriving general area cladograms consistently. If the two methodologies agree in their results, no dilemma arises. In cases where their results differ, we prefer the results obtained with *a posteriori* methods because this category of methods needs no *a priori* assumption on the independence of the causal processes (vicariance, extinction or dispersal).

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 subsidized by the Netherlands Organization for Scientific Research (NWO); grant 805-33.193 to MvV, and by the Leids Universiteits Fonds for a working visit of MvV to D.R. Brooks.

References

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

Brooks, D.R. and McLennan, D.A. (1991). “Phylogeny, Ecology and Behavior: A Research Program in Comparative Biology”. Univ. Chicago Press. Chicago.

Cracraft, J. (1986). Origin and evolution of continental biotas: speciation and historical congruence within the Australian avifauna. *Evolution* **40**: 977-996.

Cracraft, J. (1988). Deep-history biogeography: retrieving the historical pattern of evolving continental biotas. *Syst. Zool*. **37**: 221-236.

Cracraft, J. (1994). Species diversity, biogeography, and the evolution of biotas. *Amer. Zool*. **34**: 33-47.

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

Enghoff, H. (1995). Historical Biogeography of the Holarctic: area relationships, ancestral areas, and dispersal of non-marine animals. *Cladistics* **11**: 223-263.

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.

Funk, V.A. and Brooks, D.R. (1990). Phylogenetic systematics as the basis of comparative biology. *Smithsonian Contr. Bot.* **73**: 1-45.

Green, M.D., Van Veller, M.G.P. and Brooks, D.R. (submitted). Assessing modes of speciation: range asymmetry and biogeographical congruence. *Cladistics*.

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

Humphries, C.J. (1982). Vicariance biogeography in Mesoamerica. *Ann. Missouri Bot. Gard.* **69**: 444-463.

Kluge, A.G. (1988). Parsimony in vicariance biogeography: A quantitative method and a Greater Antillean example. *Syst. Zool*. **37**: 315-328.

Kluge, A.G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (*Boidae*, *Serpentes*). *Syst. Zool*. **38**: 7-26.

Kluge, A.G. (1998). Total evidence or taxonomic congruence: cladistics or consensus classification. *Cladistics* **14**: 151-158.

Kornet, D.J. (1993a). “Reconstructing species. Demarcations in genealogical networks”. PhD thesis. Leiden University. Leiden.

Kornet, D.J. (1993b). Permanent splits as speciation events: a formal reconstruction of the internodal species concept. *Journal of Theoretical Biology* **164**: 407-435.

Ladiges, P.Y. (1998). Biogeography after Burbidge. *Austral. Syst. Bot.* **11**: 231-242.

Linder, H.P. and Crisp, M.D. (1995). *Nothofagus* and Pacific biogeography. *Cladistics* **11**: 5-32.

Mayden, R.L. (1988). Biogeography, parsimony, and evolution in North American freshwater fishes. *Syst. Zool*. **37**: 329-355.

Morrone, J.J. (1993). Cladistic and biogeographic analyses of the weevil genus *Listroderes* Schoenherr (*Coleoptera*: *Curculionidae*). *Cladistics* **9**: 397-411.

Morrone, J.J. and Carpenter, J.M. (1994). In search of a method for cladistic biogeography: an empirical comparison of Component Analysis, Brooks Parsimony Analysis, and Three area statements. *Cladistics* **10**: 99-153.

Morrone, J.J. and Crisci, J.V. (1995). Historical Biogeography: introduction to methods. *Ann. Rev. Ecol. Syst.* **26**: 373-401.

Nelson, G. and Ladiges, P.Y. (1991a). Standard assumptions for biogeographic analysis*. Austral. Syst. Bot.* **4**: 41-58.

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 Univ. Press. New York.

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.* **36**: 268-279.

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

Page, R.D.M. (1990a). Temporal congruence and cladistic analysis of biogeography and cospeciation. *Syst. Zool.* **39**: 205-226.

Page, R.D.M. (1990b). Component analysis: a valiant failure? *Cladistics* **6**: 119-136.

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

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

Page, R.D.M. and Lydeard, C. (1994). Towards a cladistic biogeography of the Carribean. *Cladistics* **10**: 21-41.

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* **46**: 195-203.

Ronquist, F. (1998). Phylogenetic approaches in coevolution and biogeography. *Zoologica Scripta* **26**: 313-322.

Ronquist, F. and Nylin, S. (1990). Process and pattern in the evolution of species associations. *Syst. Zool.* **39**: 323-344.

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.

Turner, H. (1996). *Sapindaceae* and the biogeography of eastern Australia. *Austral*. *Syst*. *Bot*. **9**: 133-167.

Van Soest, R.W.M. and Hajdu, E. (1997). Marine area relationships from twenty sponge phylogenies. A comparison of methods and coding strategies. *Cladistics* **13**: 1-20.

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.

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., Zandee, M. and Kornet, D.J. (submitted). Measures for obtaining inclusive solution sets under Assumptions zero, 1 and 2 with different methods for vicariance biogeography. *Cladistics*.

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

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

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

NEDERLANDSE SAMENVATTING

Inleiding

In de biogeografie probeert men een verklaring te vinden voor de verspreiding van groepen van organismen (taxa) over het aardoppervlak. Binnen de biogeografie worden twee disciplines onderscheiden. Wanneer men naar de tegenwoordige verspreiding van taxa kijkt en deze probeert te verklaren aan de hand van ecologische processen die op een relatief korte tijdschaal plaatsvinden, spreekt men van ecologische biogeografie. Het kan echter ook zo zijn dat men met behulp van de tegenwoordige verspreidingen van taxa gegevens probeert te achterhalen over de verspreidingen van taxa in het verleden en over een veel langere (evolutionaire) tijdschaal. In het laatste geval spreekt men van historische biogeografie. Dit proefschrift gaat over de sub-discipline vicariantie biogeografie binnen de discipline historische biogeografie.

Vicariantie biogeografie maakt gebruik van informatie over de hedendaagse verspreiding van taxa over een bepaald aantal gebieden (arealen). Door naast informatie over de verspreiding van de taxa over de arealen gebruik te maken van informatie over de verwantschapsrelaties tussen deze taxa, worden hypothesen opgesteld over de historische relaties tussen de arealen waarin de taxa verspreid zijn. Deze hypothesen worden weergegeven in een divergerend vertakkingspatroon dat een areaal-cladogram genoemd wordt. Een areaal-cladogram bestaat uit takken en interne en terminale knopen. De arealen zitten op de terminale knopen en worden verbonden met verwante arealen via takken en interne knopen. Interne knopen die verwant zijn worden via takken en interne knopen verbonden totdat ze aan de basis van het cladogram in een enkele (basale) knoop samenkomen.

Op basis van deze hypothesen over historische relaties tussen arealen wil men vervolgens verklaren door welke processen de taxa hun verspreiding over de arealen hebben gekregen. Binnen de vicariantie biogeografie zijn de volgende processen hierbij van belang:

1. Vicariantie. Men spreekt van het proces van vicariantie wanneer soortsvorming (speciatie) wordt veroorzaakt door een abiotische of biotische gebeurtenis. Deze gebeurtenis splitst de populatie van een bepaalde (voorouder) soort (in één areaal) in twee disjuncte populaties (in twee nieuwe arealen). Als gevolg van de gebeurtenis hebben de twee disjuncte populaties geen contact meer en kunnen ze zich tot twee verschillende (nakomeling) soorten ontwikkelen. In de vicariantie biogeografie wordt aangenomen dat (a)biotische gebeurtenissen soortsvorming bij populaties van meerdere taxa teweegbrengt en dus een algemeen effect hebben. Dankzij het algemene karakter van speciatie door vicariantie, kunnen overeenkomstige areaal-cladogrammen voor meerdere taxa (van verschillende groepen) afgeleid worden.

2. Extinctie. Dit proces is een verklaring voor het niet verspreid zijn van een taxon in een bepaald areaal wanneer men dat op grond van resultaten, afgeleid van data van meerdere groepen van taxa, wel verwacht.

3. Dispersie. Het proces van dispersie kan ten opzichte van het eerder genoemde proces van vicariantie een tegenovergesteld effect op de distributie

van taxa hebben. Bij dispersie verspreiden taxa zich namelijk zonder in hun distributie tegengehouden te worden door de (a)biotische gebeurtenissen die bij vicariantie voor een disjuncte verspreiding van taxa zorgen. In vicariantie biogeografie heeft dispersie als verklarend proces geen algemeen karakter. Aangenomen wordt dat wanneer dispersie van een taxon van één areaal naar een ander areaal plaatsvindt (en een barrière veroorzaakt door een (a)biotische gebeurtenis dus overschreden wordt) dit uniek voor één of enkele taxa gebeurt en niet algemeen is voor alle taxa van verschillende groepen.

Voor het afleiden van areaal-cladogrammen en de verklaring van distributies van taxa over arealen met behulp van deze processen, is een aantal verschillende methoden ontwikkeld. Afhankelijk van hoe deze methoden in de vicariantie biogeografie omgaan met gegevens over de distributie van taxa en verwantschapsrelaties tussen deze taxa (weergegeven in een taxon-cladogram) worden de methoden in dit proefschrift ingedeeld in twee categorieën: *a priori* methoden en *a posteriori* methoden. Onder de *a priori* methoden vallen Component Analysis (CA), Reconciled Tree Analysis (RTA) en Three Area Statement Analysis (TAS). Onder de *a posteriori* methoden vallen Component Compatibility Analysis (CCA) en Brooks Parsimony Analysis (BPA).

Het doel van het onderzoek beschreven in dit proefschrift is drievoudig. De *a priori* en de *a posteriori* methoden worden vergeleken om:

- te beoordelen of het protocol waarmee ze algemene areaal-cladogrammen voor meerdere groepen van taxa afleiden methodologisch correct is;
- te achterhalen of ze op een consistente manier de distributie van taxa over de arealen verklaren met behulp van de processen vicariantie, extinctie en dispersie; en
- inzicht te verkrijgen in tegenovergestelde resultaten die vaak worden verkregen bij de toepassing van verschillende computerprogramma's behorende bij de verschillende methoden.

Twee methodologische vereisten in vicariantie biogeografie

Zoals hierboven is vermeld, wordt een areaal-cladogram afgeleid door een taxon-cladogram te combineren met de distributie van de taxa over arealen. In het taxon-cladogram worden de taxa vervangen door de arealen. Het op deze manier verkregen cladogram wordt een taxon-areaal-cladogram genoemd.

In het geval dat ieder taxon verspreid is over één areaal en ieder areaal één taxon van een bepaalde groep huisvest, bezet ieder areaal in het taxon-areaal-cladogram een eigen en unieke terminale knoop. Dit patroon is in overeenstemming met een eigen en unieke historie voor ieder areaal dat aanwezig is in het cladogram. Een taxon-areaal-cladogram met een eigen en unieke terminale knoop voor ieder areaal wordt in dit proefschrift een areaal-cladogram genoemd.

Maar vaak is niet ieder taxon verspreid over één areaal en huisvest ieder areaal meer dan één taxon. Taxa kunnen wijdverspreid zijn over meerdere arealen en een bepaald areaal kan meerdere verschillende (sympatrische) taxa huisvesten. Wanneer in deze gevallen taxa in het taxon-cladogram worden vervangen door hun arealen van distributie, ontstaan taxon-areaal-cladogrammen met:

- meer dan één areaal op één terminale knoop (als gevolg van wijdverspreide taxa); of
- hetzelfde areaal op twee of meer verschillende terminale knopen (als gevolg van sympatrische taxa).

In deze gevallen zijn extra stappen nodig om van deze taxon-areaal-cladogrammen areaal-cladogrammen te verkrijgen.

In vicariantie biogeografie hebben verschillende auteurs in drie verschillende aannamen beschreven hoe omgegaan moet worden met wijdverspreide taxa en sympatrische taxa om areaal-cladogrammen te verkrijgen. Dit wordt gedaan door *a priori* bepaalde processen aan te nemen en wijdverspreide of sympatrische taxa aan de hand van die processen te verklaren. Als resultaat worden sets van hypothesen (bestaande uit één of meer areaal-cladogrammen) afgeleid onder de drie verschillende aannamen.

Onder Aanname 0 (A0) gaat men *a priori* alleen van het proces vicariantie uit. Met behulp van dit proces worden wijdverspreide of sympatrische taxa verklaard en krijgt ieder areaal een eigen terminale knoop in het cladogram. Naar aanleiding van een analyse via één van de methoden in vicariantie biogeografie kan echter blijken dat de *a priori* aanname van alleen het proces vicariantie niet voldoende was en dat de verspreiding van bepaalde taxa *a posteriori* verklaard moet worden via extinctie of dispersie. De set van areaal-cladogrammen verkregen onder A0 wordt S0 genoemd.

Onder Aanname 1 (A1) gaat men *a priori* naast het proces van vicariantie ook van het proces extinctie uit voor omgang met wijdverspreide en sympatrische taxa. Naar aanleiding van een analyse via één van de methoden kan ook onder deze aanname blijken dat *a posteriori* de verspreiding van bepaalde taxa met behulp van dispersie verklaard moet worden. De set van areaal-cladogrammen verkregen onder A1 wordt S1 genoemd.

Onder Aanname 2 (A2) gaat men *a priori* van de drie processen vicariantie, extinctie en dispersie uit om wijdverspreide of sympatrische taxa te verklaren en areaal-cladogrammen te verkrijgen. Met de *a priori* aanname van deze drie processen zijn er voor *a posteriori* verklaringen geen processen meer over. De set van areaal-cladogrammen verkregen onder A2 wordt S2 genoemd.

In dit proefschrift worden de verschillende processen die *a priori* aangenomen worden onder een bepaalde aanname verondersteld onafhankelijk te zijn in hun effect. Dit betekent bijvoorbeeld dat voor een bepaald taxon de frequentie van extinctie onafhankelijk verondersteld wordt te zijn van bijvoorbeeld de dispersiecapaciteiten van dit taxon. Natuurlijk bestaat de mogelijkheid dat een taxon dat zich slecht verspreidt (dispersie), een hogere kans op uitsterven (extinctie) heeft. Echter, omdat we de afhankelijkheid van deze processen voor een bepaald taxon *a priori* niet kennen, willen we deze afhankelijkheid niet *a priori* aannemen en gaan we uit van onafhankelijkheid van de processen vicariantie, extinctie en dispersie.

Met de *a priori* aanname van een nieuw proces naast een al eerder aangenomen proces (bijvoorbeeld extinctie naast vicariantie zoals bij A1 ten opzichte van A0) worden dezelfde en (eventueel) nieuwe areaal-cladogrammen afgeleid. Het nieuwe proces (additioneel *a priori* aangenomen onder een andere aanname) kan dus leiden tot de eventuele afleiding van nieuwe (extra) areaal-cladogrammen. Dit leidt tot de formulering van methodologische vereiste I:

De inclusieve sets van processen (vicariantie ⊆ vicariantie + extinctie ⊆ vicariantie + extinctie + dispersie) a priori aangenomen onder de drie aannamen (A0, A1, A2) moeten leiden tot inclusieve sets van areaal-cladogrammen (S0 ⊆ S1 ⊆ S2).

Tot nu is alleen gesproken over het afleiden van sets van areaal-cladogrammen met het taxon-cladogram en de distributie over de arealen van een enkele groep van taxa. Wanneer men echter de algemene geschiedenis van bepaalde arealen met betrekking tot meerdere groepen van taxa wil achterhalen, moet een algemeen areaal-cladogram afgeleid worden. Met behulp van een algemeen areaal-cladogram worden de distributies van de taxa van verschillende groepen over dezelfde arealen verklaard met behulp van de processen vicariantie, extinctie en dispersie.

Bij de toepassing van A0, A1 en A2 worden algemene areaal-cladogrammen afgeleid door sets van areaal-cladogrammen afgeleid voor verschillende groepen van taxa met elkaar te vergelijken. In de overlap tussen de sets worden overeenkomstige areaal-cladogrammen gevonden die geselecteerd worden als algemene areaal-cladogrammen. Deze algemene areaal-cladogrammen laten een overeenkomstige geschiedenis van de arealen zien zoals die is afgeleid voor de verschillende groepen van taxa. De processen die bij deze overeenkomstige (algemene) geschiedenis van arealen voor de verspreiding van de taxa hebben gezorgd, worden geselecteerd aan de hand van de aanname waaronder de algemene patronen gevonden worden.

A priori heeft men geen kennis over het relatieve belang van de processen bij de verspreiding van de taxa van de afzonderlijke groepen. Daarom gaat men er *a priori* van uit dat alle taxa hetzelfde hebben gereageerd (via de processen vicariantie, extinctie of dispersie) op de (a)biotische gebeurtenissen die de arealen van elkaar scheidden. Dit leidt tot de formulering van methodologische vereiste II:

Algemene areaal-cladogrammen en informatie over (algemene) processen die de distributie van taxa van verschillende groepen over dezelfde arealen veroorzaakt hebben worden gevonden door sets van areaal-cladogrammen (afgeleid voor de groepen van taxa) te vergelijken die afgeleid zijn onder dezelfde aanname (A0, A1 of A2).

Met de koppeling van de drie aannamen in vicariantie biogeografie aan biologisch relevante processen die de geschiedenis van de distributie van taxa over arealen kunnen verklaren, krijgen de aannamen een betekenis als toetsingskader voor de processen die uniek en algemeen van aard zijn en die voor de distributie van een aantal groepen van taxa over dezelfde arealen verantwoordelijk zijn.

Indien men algemene areaal-cladogrammen wil afleiden op een manier die consistent is met de aannamen van vicariantie biogeografie moeten de methodologische vereisten I en II opgevolgd worden. Door een uitgevoerde literatuurstudie blijkt dat dit lang niet altijd gedaan is. Ook blijkt dat *a priori* omgaan met sympatrische taxa door *a priori* extinctie of dispersie te veronderstellen leidt tot de verbreking van methodologische vereiste I (non-inclusiviteit). Voor een omgang met sympatrische taxa die consistent is met de originele beschrijving van A1 en A2 en ter voorkoming van het breken van methodologische vereiste I moeten sympatrische taxa altijd *a posteriori* verklaard worden via vicariantie, extinctie of dispersie.

Evaluatie van de methoden

Om areaal-cladogrammen af te leiden -met behulp van één of meer taxon-cladogrammen en informatie over distributie van de taxa in de arealen- zijn verschillende methoden ontwikkeld. Al deze methoden starten voor hun analyse met een taxon-areaal-cladogram dat afgeleid wordt door de taxa in een taxon-cladogram te vervangen door de arealen waarin deze taxa verspreid zijn. Afhankelijk van hoe de methoden omgaan met de data wanneer wijdverspreide en sympatrische taxa verklaard moeten worden, kunnen de methoden ingedeeld worden in *a posteriori* en *a priori* methoden.

De *a posteriori* methoden brengen *a priori* geen modificaties in het taxon-areaal-cladogram aan om wijdverspreide en sympatrische taxa *a priori* te verklaren met behulp van vicariantie, extinctie of dispersie. Er zijn twee *a posteriori* methoden ontwikkeld: Brooks Parsimony Analysis (BPA) en Component Compatibility Analysis (CCA).

In Brooks Parsimony Analysis wordt het taxon-areaal-cladogram gecodeerd in een (areaal) x (knoop) matrix via additieve binaire codering. De sympatrische taxa worden gecodeerd op hun aanwezigheid en *a posteriori* verklaard. De wijdverspreide taxa worden ook op hun aanwezigheid gecodeerd en *a posteriori* verklaard onder A0. Onder A1 en A2 kunnen extra kolommen aan de (areaal) x (knoop) matrix toegevoegd worden die verschillende verwantschapsrelaties van de aanwezigheid van het wijdverspreide taxon in verschillende arealen weergeven zoals die onder A1 en A2 aangenomen kunnen worden. Door een standaard maximale parsimonie analyse zoals geïmplementeerd in PAUP of Hennig86 toe te passen op de (areaal) x (knoop) matrix, worden (sets van) meest parsimone (met de kortste lengte) areaal-cladogrammen geselecteerd onder A0, A1 en A2.

Component Compatibility Analysis combineert het taxon-cladogram met de taxon distributie over arealen (net zoals bij BPA) in een (areaal) x (knoop) matrix. Deze matrix verschilt echter van een BPA (areaal) x (knoop) matrix omdat de distributies die horen bij de interne knopen van het taxon-cladogram gecodeerd en gecombineerd worden tot één kenmerk met meerdere toestanden. Van de CCA (areaal) x (knoop) matrix worden componenten (bestaande uit sets van arealen die verondersteld worden verwant te zijn) afgeleid. De grootste sets van wederzijds compatibele componenten worden geselecteerd. Elk van deze sets correspondeert met een areaal-cladogram. Voor wijdverspreide taxa worden onder A1 en A2 extra kolommen aan de (areaal) x (knoop) matrix toegevoegd. Deze kolommen worden gebruikt voor de selectie van extra componenten. Voor sympatrische taxa worden geen extra kolommen afgeleid. Net zoals bij BPA worden de sympatrische taxa *a posteriori* verklaard. Door het plotten van de (areaal) x (knoop) data van het taxon-areaal-cladogram op de grootste sets van componenten, wordt het meest parsimone (met de kortste lengte) areaal-cladogram geselecteerd. Voor deze selectie worden de extra kolommen zoals afgeleid onder A1 en A2 niet meegenomen omdat die geen data weergeven. De methode CCA is geïmplementeerd in het computerprogramma CAFCA.

De *a priori* methoden brengen onder de verschillende aannamen (A0, A1 of A2) verschillende *a priori* modificaties in het taxon-areaal-cladogram aan om *a priori* wijdverspreide en sympatrische taxa te verklaren met behulp van de processen die *a priori* aangenomen worden onder de verschillende aannamen. Er zijn drie *a priori* methoden ontwikkeld: Component Analysis (CA), Reconciled Tree Analysis (RTA) en Three Area Statement Analysis (TAS).

Bij Component Analysis wordt het taxon-areaal-cladogram onder A0 gecodeerd in een binaire (areaal) x (knoop) matrix op de manier zoals ook is beschreven voor BPA. Hierbij worden zowel wijdverspreide als sympatrische taxa gecodeerd op hun aanwezigheid in een bepaald areaal en worden er geen extra stappen ondernomen om deze *a priori* te verklaren met behulp van extinctie of dispersie. Via een standaard maximale parsimonie analyse worden areaal-cladogrammen afgeleid onder A0. Onder A1 wordt uitgegaan van de (areaal) x (knoop) matrix zoals afgeleid onder A0. In deze matrix worden echter modificaties aangebracht voor omgang met wijdverspreide en sympatrische taxa onder deze aanname. De kolommen die corresponderen met wijdverspreide taxa worden verwijderd en de kolommen die corresponderen met sympatrische taxa worden zodanig aangepast dat de sympatrische taxa *a priori* via een combinatie van de processen vicariantie en extinctie verklaard kunnen worden. Door de toepassing van een standaard maximale parsimonie analyse op deze gemodificeerde (areaal) x (knoop) matrix worden areaal-cladogrammen geselecteerd onder A1. Voor een analyse onder A2 gebruikt CA (in tegenstelling tot BPA, CCA en analyses onder A0 en A1) geen (areaal) x (knoop) matrix. Onder A2 worden directe modificaties op het taxon-areaal-cladogram aangebracht om wijdverspreide en sympatrische taxa *a priori* te verklaren via een combinatie van vicariantie, extinctie en dispersie. Voor wijdverspreide taxa wordt de aanwezigheid van het taxon in ieder van de arealen verwijderd uit het taxon-areaal-cladogram en vervolgens weer op alle mogelijke plaatsen in het taxon-areaal-cladogram teruggeplaatst. Voor sympatrische taxa wordt ieder van de verschillende taxa die verspreid is in hetzelfde gebied *a priori* verwijderd uit het taxon-areaal-cladogram. Indien wijdverspreide taxa voorkomen in combinatie met sympatrische taxa, worden de wijdverspreide taxa die sympatrisch voorkomen met andere taxa in een bepaald areaal *a priori* verwijderd. Via deze directe modificaties in het taxon-areaal-cladogram, krijgen alle arealen een eigen en unieke terminale knoop en worden areaal-cladogrammen verkregen onder A2. De methode CA is geïmplementeerd in het computerprogramma Component 1.5.

In Reconciled Tree Analysis wordt geen (areaal) x (knoop) matrix gemaakt, maar wordt het taxon-areaal-cladogram direct vergeleken met mogelijke areaal-cladogrammen. RTA maakt gebruik van twee stappen om areaal-cladogrammen te verkrijgen. In de eerste stap wordt zo omgegaan met wijdverspreide taxa dat taxon-areaal-cladogrammen worden verkregen met maar één areaal voor iedere terminale knoop. Het wijdverspreide taxon, dat aanwezig is in verschillende arealen, krijgt voor ieder areaal in deze eerste stap een eigen terminale knoop door extra knopen aan het taxon-areaal-cladogram toe te voegen. De plaats in het taxon-areaal-cladogram waar de extra knopen worden toegevoegd is afhankelijk van de toepassing van A0, A1 of A2 en correspondeert met een *a priori* verklaring van het wijdverspreid taxon door de set van processen aangenomen onder A0, A1 of A2. In de tweede stap wordt omgegaan met sympatrische taxa en worden areaal-cladogrammen verkregen. RTA maakt in deze tweede stap geen onderscheid tussen analyses onder A0, A1 of A2. Sympatrische taxa worden met RTA *a priori* verklaard door het taxon-areaal-cladogram (met één areaal voor iedere terminale knoop) te vergelijken met steeds andere areaal-cladogrammen. Om bij iedere vergelijking het patroon van het taxon-areaal-cladogram en het patroon van een areaal-cladogram identiek te maken, worden interne en terminale knopen toegevoegd. De toegevoegde interne knopen representeren duplicaties die verklaard kunnen worden met sympatrische soortsvorming. De toegevoegde terminale knopen representeren toegevoegde taxa die verspreid waren in bepaalde arealen en daar nu niet meer aanwezig zijn als gevolg van extinctie. Het areaal-cladogram dat voor de (patroon) vergelijking met het taxon-areaal-cladogram een minimale hoeveelheid toegevoegde interne en terminale knopen nodig heeft, wordt geselecteerd. De methode RTA is geïmplementeerd in het computerprogramma Component 2.0.

Three Area Statement Analysis gebruikt net zoals BPA, CCA en CA (onder A0 en A1) een matrix om areaal-cladogrammen af te leiden. In tegenstelling tot deze methoden is echter de matrix in TAS geen directe representatie van het taxon-areaal-cladogram. In TAS wordt gebruik gemaakt van “drie-areaal permutaties”. Dit houdt in dat voor iedere interne knoop (en zijn twee corresponderende terminale knopen) van het taxon-areaal-cladogram gecodeerd wordt welke twee arealen (van de twee terminale knopen) meer verwant met elkaar zijn dan met een derde areaal. In het geval van een wijdverspreid taxon wordt van iedere corresponderende terminale knoop één areaal gekozen en worden de overige arealen in de drie-areaal permutatie gecodeerd als vraagtekens. Door de combinatie van alle mogelijke drie-areaal permutaties per interne knoop en de combinatie van alle drie-areaal permutaties over alle interne knopen, wordt het taxon-areaal-cladogram gecodeerd in een matrix met drie-areaal permutaties. Onder A0 worden drie-areaal permutaties afgeleid voor alle interne knopen en worden extra interne knopen aangenomen voor wijdverspreide taxa. Onder A1 en A2 worden drie-areaal permutaties alleen afgeleid voor alle interne knopen. Onder A2 wordt, voorafgaand aan de afleiding van de drie-areaal permutaties, in het geval van een combinatie van wijdverspreide en sympatrische taxa, de sympatrische aanwezigheid van een wijdverspreid taxon *a priori* verwijderd uit het taxon-areaal-cladogram. Van de matrix met drie-areaal permutaties worden via een standaard maximale parsimonie analyse sets met areaal-cladogrammen afgeleid. De methode TAS is geïmplementeerd in het computerprogramma TAS.

Voor de evaluatie van zowel *a priori* als *a posteriori* methoden met betrekking tot methodologische vereisten I en II zijn theoretische en empirische datasets gebruikt. De theoretische datasets bestaan uit taxon-areaal-cladogrammen voor 3, 4 en 5 taxa met één wijdverspreid taxon, twee sympatrische taxa of een combinatie van één wijdverspreid taxon dat in één van de twee arealen samen voorkomt met een ander (sympatrisch) taxon. De empirische datasets bestaan uit taxon-cladogrammen en de distributie van de genera *Heterandria* en *Xiphophorus* en uit taxon-cladogrammen en de distributie van de genera *Cyttaria* en *Eriococcus/Madarococcus*. Met de theoretische en empirische datasets worden onder A0, A1 en A2 door de verschillende methoden sets van areaal-cladogrammen afgeleid die vergeleken worden op inclusiviteit (methodologische vereiste I). Met de empirische datasets worden algemene areaal-cladogrammen afgeleid met de verschillende methoden en wordt geëvalueerd of deze cladogrammen voor alle groepen van taxa onder dezelfde aanname zijn afgeleid (methodologische vereiste II).

Zowel *a priori* als *a posteriori* methoden blijken methodologische vereiste I onder bepaalde omstandigheden bij de theoretische of empirische datasets te breken.

Bij *a posteriori* methoden wordt methodologische vereiste I soms gebroken bij de theoretische datasets bestaande uit taxon-areaal-cladogrammen met een combinatie van wijdverspreide en sympatrische taxa. De sets van areaal-cladogrammen zijn niet inclusief wanneer onder A1 (of A2) meer parsimone areaal-cladogrammen worden afgeleid dan onder A0 (of A1).

Bij *a priori* methoden wordt methodologische vereiste I vaak gebroken bij theoretische datasets bestaande uit taxon-areaal-cladogrammen met sympatrische taxa of met een combinatie van wijdverspreide en sympatrische taxa. Het *a priori* verwijderen van wijdverspreide of sympatrische taxa in een deel van hun verspreiding onder A2 leidt vaak tot kleinere sets van areaal-cladogrammen dan onder A1 en daarmee tot non-inclusiviteit.

Uit de evaluatie van de *a posteriori* en de *a priori* methoden met de empirische datasets blijkt dat methodologische vereiste II nooit gebroken wordt. Met de *a posteriori* methoden CCA en BPA en de *a priori* methoden RTA en TAS worden algemene areaal-cladogrammen verkregen door data van verschillende groepen van taxa onder één aanname (A0, A1 of A2) te combineren en uit deze gecombineerde data de algemene areaal-cladogrammen af te leiden. Met de *a priori* methode CA worden sets van areaal-cladogrammen (die afgeleid zijn voor verschillende groepen van taxa) met elkaar vergeleken om algemene areaal-cladogrammen te vinden. Door de sets van areaal-cladogrammen alleen onder dezelfde aanname met elkaar te vergelijken wordt methodologische vereiste II niet gebroken.

Reparatie van de methoden

Na de evaluatie van de *a posteriori* en de *a priori* methoden voor methodologische vereisten I en II blijkt dat onder bepaalde omstandigheden non-inclusieve sets van areaal-cladogrammen afgeleid kunnen worden.

Een *ad hoc* remedie om met *a posteriori* methoden inclusieve sets van areaal-cladogrammen te verkrijgen is de toepassing van een bovengrens (voor de lengte van areaal-cladogrammen) bij de selectie van areaal-cladogrammen. Deze *ad hoc* remedie voorkomt dat onder A1 of A2 alleen de meer parsimone cladogrammen worden geselecteerd waardoor methodologische vereiste I gebroken wordt. Met het instellen van een bovengrens worden de areaal-cladogrammen die gevonden zijn onder A0 (of A1) ook geselecteerd onder A1 (of A2).

Bij CCA of BPA blijkt echter dat ondanks de toepassing van deze *ad hoc* remedie er soms toch non-inclusieve sets van areaal-cladogrammen afgeleid kunnen worden. Dit gebeurt wanneer onder A0 areaal-cladogrammen worden geselecteerd die niet volledig dichotoom zijn. Onder A1 of A2 worden door de afleiding van extra componenten de polytomieën in deze areaal-cladogrammen opgelost. Omdat er bij CCA en BPA een voorkeur is voor volledig dichotome cladogrammen, worden de areaal-cladogrammen met polytomieën (afgeleid onder A0) niet meer afgeleid onder A1 of A2. Wanneer echter de originele data geplot worden op deze volledig dichotome areaal-cladogrammen (afgeleid onder A1 of A2) blijkt dat bepaalde interne knopen niet ondersteund worden door de data. Deze niet ondersteunde knopen zijn afgeleid door de toepassing van de aanname.

Een meer fundamentele manier om inclusieve sets van areaal-cladogrammen af te leiden met *a posteriori* methoden is door met CCA en BPA alleen areaal-cladogrammen af te leiden onder A0 en geen extra componenten af te leiden onder A1 of A2. Met zowel CCA als BPA worden taxon-cladogrammen en distributiedata van meerdere groepen van taxa gecombineerd in één (areaal) x (knoop) matrix. Door een parsimonie analyse toe te passen op deze gecombineerde matrix worden algemene areaal-cladogrammen afgeleid onder A0. Incongruenties in de data (die veroorzaakt worden door sympatrische of wijdverspreide taxa) kunnen *a posteriori* verklaard worden met behulp van vicariantie, extinctie of dispersie door de data van de afzonderlijke groepen van taxa te plotten op het algemeen areaal-cladogram.

Met *a priori* methoden kunnen inclusieve sets van areaal-cladogrammen afgeleid worden door wijdverspreide en sympatrische taxa in twee stappen *a priori* te verklaren. In de eerste stap worden de wijdverspreide taxa verklaard onder A0, A1 of A2 en worden taxon-areaal-cladogrammen verkregen met slechts één areaal per terminale knoop. Deze sets van taxon-areaal-cladogrammen, die worden verkregen door met CA en RTA wijdverspreide taxa *a priori* te verklaren, zijn inclusief. TAS kan door de *a priori* behandeling van wijdverspreide taxa met behulp van drie-areaal permutaties resulteren in non-inclusieve sets van taxon-areaal-cladogrammen. Om ook met TAS inclusieve sets van taxon-areaal-cladogrammen af te leiden (waarin wijdverspreide taxa *a priori* zijn opgelost) moet bij de afleiding van de matrix met drie-areaal permutaties onder A1 of A2 een extra kolom toegevoegd worden waarbij de arealen van een wijdverspreid taxon gecodeerd worden als één component.

In de tweede stap worden de nog in de taxon-areaal-cladogrammen aanwezige sympatrische taxa verklaard. Om inclusieve sets van areaal-cladogrammen te verkrijgen en geen data *a priori* bij de analyse uit te sluiten, worden met CA en TAS de sympatrische taxa *a posteriori* verklaard. Dit wordt gedaan door het coderen van het taxon-areaal-cladogram in een binaire matrix of een matrix met drie-areaal permutaties. Door parsimonie analyse toe te passen op deze matrix, worden de areaal-cladogrammen geselecteerd. Met RTA is het onmogelijk om sympatrische taxa *a posteriori* te verklaren omdat er geen matrix afgeleid wordt. Bij RTA worden sympatrische taxa *a priori* verklaard door een minimaal aantal interne en terminale knopen toe te voegen die met extincties en sympatrische soortsvorming kunnen worden verklaard. Inclusieve sets met areaal-cladogrammen worden afgeleid door het aantal toegevoegde interne en terminale knopen onder A0 (of A1) als bovengrens te gebruiken voor het aantal toegevoegde interne en terminale knopen onder A1 (of A2).

Nadat de bovenstaande remedies zijn toegepast op *a posteriori* methoden en *a priori* methoden blijkt echter dat bij analyses onder de verschillende aannamen, CCA, BPA, CA, RTA en TAS toch nog verschillende algemene areaal-cladogrammen kunnen afleiden. Het verschil in resultaten afgeleid met deze verschillende methoden moet gezocht worden in het verschil in methodologieën om algemene areaal-cladogrammen af te leiden en distributies van taxa te verklaren via de processen vicariantie, extinctie of dispersie.

A posteriori en *a priori* methodologieën

Zoals eerder is aangegeven, vallen de methoden in de vicariantie biogeografie in twee verschillende categorieën: *a posteriori* methoden en *a priori* methoden. Ieder van deze twee categorieën correspondeert met een bepaalde methodologie waarmee algemene areaal-cladogrammen worden afgeleid door het formuleren en testen van nulhypothesen over de processen (vicariantie, extinctie of dispersie) die de oorzaak zijn van de verspreiding van taxa over arealen.

De *a posteriori* methoden (CCA en BPA) formuleren één nulhypothese onder A0. Onder deze aanname gaan de *a posteriori* methoden ervan uit dat de verspreiding van taxa over arealen is veroorzaakt door het proces van vicariantie. Extra processen als extinctie of dispersie worden onder deze aanname niet *a priori* aangenomen.

De nulhypothese die correspondeert met deze aanname is dat de verwantschapsrelaties tussen taxa van de verschillende groepen (zoals is weergegeven in de taxon-cladogrammen) en hun distributies over dezelfde arealen resulteren in één enkel algemeen areaal-cladogram. Dit enkele algemene areaal-cladogram representeert het patroon van historische relaties tussen de arealen. Bij het plotten van de gecombineerde (areaal) x (knoop) matrix op dit cladogram zijn alle componenten congruent met de data en wordt de distributie van alle taxa verklaard met vicariantie.

Voor het testen van de nulhypothese wordt het taxon-cladogram en de distributiedata van iedere groep van taxa gecombineerd in een (areaal) x (knoop) matrix. CCA doet dit op een directe (door een taxon- en een distributiematrix te vermenigvuldigen) en BPA doet dit op een indirecte manier (door eerst een taxon-areaal-cladogram af te leiden). De (areaal) x (knoop) matrices voor de verschillende groepen van taxa worden bij zowel CCA als BPA gecombineerd in één grote (areaal) x (knoop) matrix. Van deze gecombineerde (areaal) x (knoop) matrix worden algemene areaal-cladogrammen afgeleid via een parsimonie analyse.

Wanneer in een algemeen areaal-cladogram alle componenten congruent zijn met de data kan de nulhypothese niet verworpen worden en wordt de distributie van de taxa over de arealen verklaard met het proces van vicariantie. Indien er echter één component niet congruent is met de data wordt de nulhypothese verworpen. De niet congruente componenten worden dan *a posteriori* verklaard met *ad hoc* procesverklaringen zoals extinctie of dispersie.

Parsimonie analyse van de gecombineerde (areaal) x (knoop) matrix leidt altijd tot een algemeen areaal-cladogram. Wanneer de nulhypothese niet wordt verworpen, wordt de algemene distributie van verschillende groepen van taxa over dezelfde arealen verklaard door vicariantie. In het geval dat de nulhypothese wordt verworpen, worden *a posteriori* de processen extinctie of dispersie gebruikt als verklaring voor de (unieke) distributie van bepaalde taxa. Algemene areaal-cladogrammen worden altijd afgeleid onder A0 en aan methodologische vereiste II wordt altijd voldaan. Methodologische vereiste I blijkt op de *a posteriori* methoden niet van toepassing te zijn omdat de processen extinctie of dispersie nooit *a priori* aangenomen worden en dus A1 en A2 niet nodig zijn om algemene areaal-cladogrammen af te leiden.

De *a priori* methoden (CA, RTA en TAS) formuleren drie verschillende nulhypothesen onder A0, A1 en A2.

Onder A0 gaan de *a priori* methoden (net zoals de *a posteriori* methoden) ervan uit dat de verspreiding van de taxa over arealen is veroorzaakt door vicariantie. De bijbehorende nulhypothese is dat de taxon-cladogrammen en de distributies van de verschillende groepen van taxa over dezelfde arealen resulteren in één of meerdere algemene areaal-cladogrammen. De nulhypothese wordt getest door sets van areaal-cladogrammen, afgeleid voor verschillende groepen van taxa onder A0 (door *a priori* alleen vicariantie aan te nemen), met elkaar te vergelijken. Indien congruente areaal-cladogrammen worden gevonden, kan de nulhypothese niet worden verworpen en worden de congruente areaal-cladogrammen geselecteerd als algemene areaal-cladogrammen. De distributie van de taxa over de arealen wordt verklaard met vicariantie. Indien er echter geen congruente areaal-cladogrammen worden gevonden, wordt de nulhypothese (onder A0) verworpen en wordt gekozen voor het testen van een alternatieve nulhypothese onder A1.

De nulhypothese onder A1 is dat de verspreiding van taxa over arealen kan worden verklaard met vicariantie of extinctie. Op grond van deze twee processen worden (voor de verschillende groepen van taxa) sets van areaal-cladogrammen afgeleid en vergeleken op congruente areaal-cladogrammen. Indien congruente areaal-cladogrammen (algemene areaal-cladogrammen) worden gevonden, kan de nulhypothese onder A1 niet verworpen worden en wordt de distributie van taxa over arealen verklaard met de processen vicariantie en extinctie. Indien er echter nog steeds geen congruente areaal-cladogrammen worden gevonden, wordt de nulhypothese (onder A1) verworpen en wordt gekozen voor het testen van een alternatieve nulhypothese onder A2.

De nulhypothese onder A2 is dat de verspreiding van taxa over arealen kan worden verklaard met vicariantie, extinctie of dispersie. Op grond van deze drie processen worden (voor de verschillende groepen van taxa) sets van areaal-cladogrammen afgeleid en vergeleken op congruente areaal-cladogrammen. Indien congruente areaal-cladogrammen (algemene areaal-cladogrammen) worden gevonden, kan de nulhypothese onder A2 niet verworpen worden en wordt de distributie van taxa over arealen verklaard met de processen vicariantie, extinctie en dispersie. Indien er echter ook onder A2 geen congruente areaal-cladogrammen worden gevonden, wordt de nulhypothese (onder A2) verworpen. Bij het verwerpen van de nulhypothese onder A2 kunnen geen algemene areaal-cladogrammen voor alle groepen van taxa meer afgeleid worden omdat alle drie de processen (vicariantie, extinctie of dispersie) die de distributie van taxa over arealen onder A0, A1 en A2 kunnen verklaren aangenomen zijn.

Onder de verschillende aannamen (A0, A1 en A2) worden verschillende nulhypothesen geformuleerd en getest door het vergelijken van sets van areaal-cladogrammen afgeleid onder één bepaalde aanname. Omdat algemene areaal-cladogrammen alleen afgeleid kunnen worden wanneer dezelfde nulhypothese onder dezelfde aanname voor alle taxa niet verworpen wordt, is methodologische vereiste II noodzakelijk voor *a priori* methoden.

Wanneer voor twee of meer groepen van taxa onder verschillende aannamen er congruente areaal-cladogrammen verkregen worden, moet er gekozen worden voor één bepaalde aanname om de algemene areaal-cladogrammen te selecteren.

Wanneer men er *a priori* vanuit gaat dat de processen vicariantie, extinctie en dispersie afhankelijk zijn in hun effect, kunnen non-inclusieve sets van areaal-cladogrammen afgeleid zijn. Bij het vergelijken van deze non-inclusieve sets kan het zo zijn dat de congruente areaal-cladogrammen (die afgeleid zijn onder de verschillende aannamen) totaal verschillend zijn. Om de algemene areaal-cladogrammen te kiezen is dan een extra selectiecriterium nodig. Voor de *a priori* methoden is dit selectiecriterium nog niet geformuleerd.

Wanneer men *a priori* ervan uitgaat dat de processen vicariantie, extinctie en dispersie onafhankelijk zijn in hun effect, is methodologische vereiste I noodzakelijk voor *a priori* methoden. Bij inclusieve sets van areaal-cladogrammen per groep van taxa (methodologische vereiste I), zijn ook de sets van congruente areaal-cladogrammen inclusief. De congruente areaal-cladogrammen die geselecteerd worden onder verschillende aannamen zijn gebaseerd op de *a priori* aanname van zo weinig mogelijk processen en worden geselecteerd als de algemene areaal-cladogrammen met het grootste verklarend vermogen.

A posteriori en *a priori* methoden zijn gebaseerd op zeer verschillende methodologieën voor het formuleren en testen van nulhypothesen. Wanneer beide methodologieën overeenkomen in hun resultaten (algemene areaal-cladogrammen), is er geen enkel probleem. Indien er echter verschillende algemene areaal-cladogrammen gevonden worden, wordt aan het eind van dit proefschrift een voorkeur uitgesproken voor de resultaten van de *a posteriori* methoden. De reden hiervoor is dat de methoden binnen de categorie van *a posteriori* methoden geen *a priori* aanname over de onafhankelijkheid van de processen (die de distributie van verschillende groepen van taxa over dezelfde arealen verklaren) nodig hebben.

NAWOORD

Na vier jaar hypothetiseren, produceren, analyseren, interpreteren, discussiëren, concluderen en heel veel leren is het proefschrift dan opeens af. Op één van de laatste bladzijden is er dan nog ruimte om je te richten tot een aantal mensen en dat wil ik bij deze dan ook graag doen.

Methodologisch onderzoek vraagt veel concentratie en daarom dus regelmatig momenten van éénzame opsluiting in je kamertje. Gelukkig was er echter altijd de bel/gong die een uitnodiging vormde voor gezelligheid, maar ook voor goede (vak)inhoudelijke discussies met collega's van de oude groep Theoretische Biologie en de "vernieuwde" secties Theoretische Biologie en Fylogenetische Systematiek en Theoretische Evolutiebiologie.

Andere momenten van "luchten" waren de ontmoetingen met Dirk, Raoul en Paul. Onze besprekingen van fylogenetische en biogeografische literatuur hebben voor mij bijgedragen tot een beter inzicht in de verschillende methoden die gebruikt worden in de (vicariantie) biogeografie.

During my PhD research I have had the oppurtunity to travel and visit the Zoological Department of the University of Toronto. Here, I had a nice time of supportive research and friendship with Dan, Debbie, Michelle, Aryeh and Marc.

Het project *Phylogenetic reconstruction of a neural network underlying theory of mind in primates* met Ruud en Cynthia was een leuke stap buiten mijn eigen promotieonderzoek en was verder goed voor gezelligheid in Leiden en Geldermalsen.

Met Hendrik-Jan en Coline had ik zinvolle discussies over de toepassing van fylogenetische methoden op moleculaire en biogeografische datasets.

Buiten het onderzoek waren er regelmatig de bijeenkomsten met Roberto, Nolanda, Jan, Bas en Ann waarbij we ons steeds voornamen om over een levensbeschouwelijk boek te praten, maar dan meestal over hele andere onderwerpen spraken.

Tijdens de maandelijkse maaltijden met Femmie, afwisselend in Leiden en Amsterdam, was het gezellig bijkletsen.

De weekendjes bij mijn familie in het mooie Drenthe zijn altijd erg ontspannend en verfrissend geweest. Wanneer Haitze er was, hadden we regelmatig tijd om te oefenen voor een judo-examen of voor de Dam tot Dam loop. Gonnie was er altijd voor advies en relativerende gesprekken. Mijn ouders, Peter en Rita, bedank ik voor hun interesse en motiverende woorden over de telefoon, via de e-mail, onder het koken, maar ook tijdens de wandelingen met Dasai en Zenji.

Maria bedank ik voor haar interesse, spontaniteit en opvrolijkende woorden tijdens de afgelopen en in de komende tijd.

CURRICULUM VITAE

Marco Gerardus Petrus van Veller werd op 1 november 1970 geboren te Gouda. Hij bracht zijn jeugd door in Boskoop en bezocht daar de Rehoboth- en de Immanuelschool. Het voortgezet onderwijs werd gevolgd op de Maranatha-Mavo, eveneens te Boskoop. Op 27 mei 1987 ontving hij zijn diploma waarna hij het eerste en tweede jaar van het Middelbaar Laboratorium Onderwijs aan het Van Leeuwenhoek Instituut te Delft volgde. In 1989 maakte hij de overstap naar het Hoger Laboratorium Onderwijs (Hogeschool Rotterdam en Omstreken, vestiging Delft). Na de propaedeuse werd in 1990 begonnen aan de studies Plantenveredeling en Plantenziektekunde van de Landbouwuniversiteit Wageningen. Tijdens deze twee studies werden twee afstudeervakken (Virologie en Plantentaxonomie) en een stage (Nematologie) uitgevoerd. Tijdens het afstudeervak Virologie werd de resistentie tegen het tomatenbronsvlekkenvirus (TSWV) in transgene tabak onderzocht. Voor het afstudeervak Plantentaxonomie werd een literatuurstudie uitgevoerd naar de co-evolutie tussen wilde verwanten van de aardappel (*Solanum sect. Petota*) en aardappelcysteaaltjes (*Globodera*). Van maart tot juli 1994 werd voor een stage Nematologie gewerkt bij het Hilbrandslaboratorium voor Bodemziekten te Assen. Tijdens deze stage werden een literatuurstudie en veld- en kasonderzoek verricht naar de selectie voor virulentie in *Globodera pallida* door de teelt van resistente aardappelrassen. Op 22 augustus 1994 werd het doctoraaldiploma voor de studie Plantenveredeling en op 21 augustus 1995 werd het doctoraaldiploma voor de studie Plantenziektekunde ontvangen. Hierna werd in een Na-Doctoraal Onderzoeksproject onderzoek verricht bij de vakgroep Plantentaxonomie naar de theoretische achtergronden van fylogenetische en fenetische methoden in moleculaire systematiek. Tijdens dit project kreeg hij een aanstelling bij de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) als Onderzoeker in Opleiding. Van 1 september 1996 tot 1 september 2000 werd in het kader van het project *Methods for biogeographic analysis: Comparison and evaluation* het onderzoek verricht dat is beschreven in dit proefschrift. Het onderzoek werd uitgevoerd bij de sectie Theoretische Biologie (Instituut voor Evolutionaire en Ecologische Wetenschappen, Universiteit Leiden) welke in 1999 overging in de sectie Theoretische Biologie en Fylogenetische Systematiek. Tijdens het project werden drie werkbezoeken gebracht aan prof. dr. D.R. Brooks van het Zoological Department (University of Toronto). Op 18 februari 2000 ontving hij de Kuenenprijs voor het artikel “Two requirements for obtaining valid common patterns under different assumptions in vicariance biogeography” tijdens het jaarlijkse symposium van het Instituut voor Evolutionaire en Ecologische Wetenschappen. Vanaf 1997 is hij bestuurslid van de Stichting Universitaire Woonwijk “Boerhaave”. Verder is hij lid van de barcommissie van de Leidse Studenten Ekklesia. In zijn vrije tijd doet hij graag aan sport (Judo, hardlopen en badminton) en leest hij graag een levensbeschouwelijke tekst.

Publicaties

Prins, M., Haan, P. de, Luyten, R., Veller, M.G.P. van, Grinsven, M.Q.J.M. van & Goldbach, R. (1995). Broad resistance to tospoviruses in transgenic plants expressing three tospoviral nucleoprotein gene sequences. *MPMI* **8**, 85-91.

Bos, R. van den, Nolten, C. & Veller, M.G.P. van. (1998). Phylogenetic reconstruction of a neural network underlying theory of mind in primates. *European Journal of Neuroscience* **10** (Suppl. 10: Abstract book), 197.

Veller, M.G.P. van, Zandee, M. & Kornet, D.J. (1999). Two requirements for obtaining valid common patterns under different assumptions in vicariance biogeography. *Cladistics* **15**, 393-406.

Veller, M.G.P. van, Kornet, D.J. & Zandee, M. (2000). Methods in vicariance biogeography: assessment of the implementations of Assumptions zero, 1 and 2. *Cladistics* **16**, 319-345.

Bos, R. van den, Veller, M.G.P. van, Nolten, C.H.M., Veer, M.W. de & Kornet, D.J. (aangeboden). Theory of Mind and neuronal networks: can differences in neuronal networks explain differences in the presence of Theory of Mind in great ape species? *Animal Behaviour*.

Brooks, D.R., Veller, M.G.P. van & McLennan, D.A. (aangeboden). How to do BPA, Really. *Journal of Biogeography*.

Green, M.D., Veller, M.G.P. van & Brooks, D.R. (aangeboden). Assessing modes of speciation: range asymmetry and biogeographical congruence. *Cladistics*.

Veller, M.G.P. van (aangeboden). Methodology and implementation of Component Analysis in coevolutionary and biogeographical studies. In: “All things considered: phylogenetic study of historical associations”. Brooks, D.R. & Veller, M.G.P. van (eds). The University of Chicago Press. Chicago, Il.

Veller, M.G.P. van & Brooks, D.R. (aangeboden). When simplicity is not parsimonious: Inductive and hypothetico-deductive approaches in historical biogeography. *Journal of Biogeography*.

Veller, M.G.P. van, Zandee, M. & Kornet, D.J. (aangeboden). Measures for obtaining inclusive solution sets under Assumptions zero, 1 and 2 with different methods for vicariance biogeography. *Cladistics*.

Presentaties

Veller, M.G.P. van. (1998). Should Assumptions 0, 1 and 2 lead to inclusive solution sets in vicariance biogeography and analogous studies? Texel-bijeenkomst van de Nederlandse Vereniging voor Theoretische Biologie. 12-13 februari 1998. NIOZ. Texel.

Veller, M.G.P. van. (1998). Should Assumptions 0, 1 and 2 lead to inclusive solution sets in vicariance biogeography and analogous studies? EEW Symposium "Comparative Research and Phylogeny". 26-27 maart 1998. EEW. Leiden.

Veller, M.G.P. van. (1998). Assumptions 0, 1 and 2 in vicariance biogeography: an update. IVth European Meeting of Ph.D. students in Evolutionary Biology. 1-3 april 1998. Biology Department, University of Crete. Heraklion, Crete. Greece.

Veller, M.G.P. van. (1998). Empirical properties of different methods under Assumptions zero, 1, and 2. "Assessing historical biogeographic and coevolutionary associations: A workshop". 24 oktober 1998. Department of Zoology, University of Toronto. Toronto, Ontario. Canada.

Veller, M.G.P. van. (1998). Component Analysis (CA). "Assessing historical biogeographic and coevolutionary associations: A workshop". 24 oktober 1998. Department of Zoology, University of Toronto. Toronto, Ontario. Canada.

Veller, M.G.P. van. (1999). Assessment of methods in vicariance biogeography: the requirement to obtain inclusive sets of area cladograms under Assumptions zero (A0), 1 (A1), and 2 (A2). Vth European Meeting of Ph.D. students in Evolutionary Biology. 24-26 maart 1999. Umeå University. Umeå. Sweden.

Veller, M.G.P. van. (1999). Assessment of methods in vicariance biogeography for obtaining inclusive solution sets under Assumptions zero (A0), 1 (A1), and 2 (A2). XVIIIth Meeting of the Willi Hennig Society. 12-17 september 1999. Institut für Zoologie und Anthropologie. Göttingen. Germany.

Veller, M.G.P. van. (2000). Two requirements for detecting common history in co-evolving areas and their species. EEW Symposium "Molecular aspects of evolutionary change". 17-18 februari 2000. EEW. Leiden.

Veller, M.G.P. van. (2000). Requirements for obtaining valid general area cladograms with a posteriori and a priori methods in vicariance biogeography. XIXth Meeting of the Willi Hennig Society. 28 mei –2 juni 2000. Natural History Museum Naturalis and National Herbarium of the Netherlands. Leiden.

Posters

Veller, M.G.P. van. (1997). Why should solutions derived under Assumptions 0, 1 and 2 in vicariance biogeography be inclusive? XVIth Meeting of the Willi Hennig Society. 23-26 oktober 1997. George Washington University. Washington D.C.

Veller, M.G.P. van. (1998). Why should solutions derived under Assumptions 0, 1 and 2 in vicariance biogeography be inclusive? IVth European Meeting of Ph.D. students in Evolutionary Biology. 1-3 april 1998. Biology Department, University of Crete. Heraklion, Crete. Greece.

Veller, M.G.P. van. (1999). Common patterns in vicariance biogeography: a priori versus a posteriori methods. Vth European Meeting of Ph.D. students in Evolutionary Biology. 24-26 maart 1999. Umeå University. Umeå. Sweden.

Veller, M.G.P. van. (1999). Common patterns in vicariance biogeography: a priori versus a posteriori methods. XVIIIth Meeting of the Willi Hennig Society. 12-17 september 1999. Institut für Zoologie und Anthropologie. Göttingen. Germany.