

New solutions to old problems: widespread taxa, redundant distributions and missing areas in event-based biogeography

I. Sanmartín* & F. Ronquist

Sanmartín, I. & Ronquist, F., 2002. New solutions to old problems: widespread taxa, redundant distributions and missing areas in event-based biogeography. *Animal Biodiversity and Conservation*, 25.2: 75–93.

Abstract

New solutions to old problems: widespread taxa, redundant distributions and missing areas in event-based biogeography.— Area cladograms are widely used in historical biogeography to summarize area relationships. Constructing such cladograms is complicated by the existence of *widespread taxa* (terminal taxa distributed in more than one area), *redundant distributions* (areas harboring more than one taxon) and *missing areas* (areas of interest absent from some of the compared cladograms). These problems have traditionally been dealt with using *Assumptions 0, 1, and 2*, but the assumptions are inapplicable to event-based methods of biogeographic analysis because they do not specify the costs of alternative solutions and may result in non-overlapping solution sets. The present paper presents the argument that only widespread terminals pose a problem to event-based methods, and three possible solutions are described. Under the *recent option*, the widespread distribution is assumed to be the result of recent dispersal. The *ancient option* assumes that the widespread distribution is the result of a failure to vicariate, and explains any mismatch between the distribution and the area cladogram by extinction. The *free option* treats the widespread taxon as an unresolved higher taxon consisting of one lineage occurring in each area, and permits any combination of events and any resolution of the terminal polytomy in explaining the widespread distribution. Algorithms implementing these options are described and applied to Rosen (1978)'s classical data set on *Heterandria* and *Xiphophorus*.

Key words: Historical biogeography, Widespread taxa, Missing areas, Redundant distributions, Assumptions 0, 1, and 2.

Resumen

Nuevas soluciones a viejos problemas: taxones de amplia distribución, distribuciones redundantes y áreas ausentes en la biogeografía cladista de procesos.— El análisis biogeográfico cladista se basa en la comparación de *cladogramas de áreas* de organismos que habitan una misma región (sustituyendo el nombre de los taxones en la filogenia por las áreas que éstos ocupan) para obtener un patrón común, el *cladograma general de áreas*. La construcción del *cladograma de áreas* se complica cuando existen taxones presentes en más de un área de distribución ("taxones de amplia distribución"), áreas que albergan más de un taxón ("distribuciones redundantes"), o áreas que no están presentes en alguno de los grupos ("áreas ausentes"). En biogeografía cladista de procesos, los taxones de amplia distribución se resuelven aplicando las *Asunciones: 0, 1, y 2*, que difieren en la relación cladogenética permitida entre las áreas donde se distribuye el taxón. Se proponen tres nuevas soluciones para abordar este problema dentro de un nuevo enfoque en biogeografía cladista que incorpora los procesos al análisis biogeográfico: "biogeografía cladista de procesos". Estas opciones difieren no sólo en las relaciones entre las áreas implicadas sino también en los procesos biogeográficos que pudieron haber dado lugar a la distribución. La opción *recent* considera la amplia distribución como si fuera de origen reciente y la explica por dispersión. La opción *ancient* considera que la amplia distribución es ancestral y la explica mediante vicarianza y extinción. La opción *free* considera la amplia distribución como un taxón de alto rango con un linaje en cada una de las áreas implicadas y cuyas relaciones no han sido establecidas, permitiendo cualquier combinación de procesos biogeográficos y cualquier solución de la politomía para explicar la distribución. Se comparan estas opciones utilizando el famoso análisis de Rosen (1978) sobre *Heterandria* y *Xiphophorus*. También se discute brevemente como tratar las distribuciones redundantes y las áreas ausentes dentro de este nuevo enfoque.

Palabras clave: Biogeografía histórica, Taxones de amplia distribución, Áreas ausentes, Distribuciones redundantes, Asunciones 0, 1 y 2.

(Received: 11 I 02; Conditional acceptance: 25 IV 02; Final acceptance: 6 VI 02)

Isabel Sanmartín & Fredrik Ronquist, Dept. of Systematic Zoology, Evolutionary Biology Centre, Uppsala Univ., Norbyvägen 18D, SE-752 36 Uppsala, Sweden.

* Corresponding author: Isabel Sanmartin, Dept. of Systematic Zoology, Evolutionary Biology Centre, Uppsala Univ., Norbyvägen 18D, SE-752 36 Uppsala, Sweden.

E-mail: isabel.sanmartin@ebc.uu.se

Introduction

Cladistic biogeography seeks to summarize information on distribution and phylogenetic relationships of organisms in area cladograms, branching diagrams that express the inter-relationships of areas based on their biotas (fig. 1a). The analysis usually starts with *taxon-area cladograms* (TAC) (ENGHOFF, 1993; MORRONE & CRISCI, 1995), which are constructed by replacing the terminal taxa in a phylogeny with the areas in which they occur. Comparing area cladograms of different organisms that occur in the same region may reveal common biogeographic patterns that can be represented in a general area cladogram (GAC).

If every terminal taxon is endemic to a unique area and every area harbors only one terminal taxon, the TAC represents a valid hypothesis about area relationships. However, the situation becomes more complicated when the "one-area-one-taxon" assumption is violated, in which case the TAC may be incomplete or indicate conflicting area relationships. The sources of these problems are often divided into three categories: *widespread taxa* (taxa present in more than one area, fig. 1b), *redundant distributions* (areas harboring more than one taxon, fig. 1c), and *missing areas* (areas of interest absent from some of the compared taxon-area cladograms, fig. 1d). The latter problem is only relevant when several TACs are analyzed simultaneously.

Problematic TACs can be converted into *resolved area cladograms* (RACs; that is, taxon-specific GACs), in which each area is represented by only one terminal (ENGHOFF, 1996), by applying *Assumptions 0, 1, and 2* (fig. 2). These assumptions mainly differ in their treatment of widespread taxa. *Assumption 0* (*A0*, ZANDEE & ROOS, 1987) regards the widespread distribution as the result of a failure to speciate in response to vicariance events affecting other lineages. The areas inhabited by the widespread taxon are considered to form a monophyletic clade (fig. 2: RAC1) and the widespread taxon is thus treated as a synapomorphy of the areas in which it occurs. *Assumption 1* (*A1*, NELSON & PLATNICK, 1981) explains the widespread distribution as the result of a failure to vicariate, possibly in combination with subsequent extinction. The areas inhabited by the widespread taxon are considered to form a monophyletic or paraphyletic group of areas (fig. 2: RACs 1–3) and the widespread taxon is treated as a sympleiomorphy of areas. *Assumption 2* (*A2*, NELSON & PLATNICK, 1981), finally, allows failure to vicariate, extinction, dispersal, or any combination of these events, in explaining the origin of widespread distributions (VAN VELLER et al., 1999). The areas inhabited by the widespread taxon are regarded as constituting a poly-, para- or monophyletic group of areas (fig. 2: RACs 1–7), and the widespread taxon is treated as a possible

convergence of the areas. In practice, *A2* is implemented by locking each of the areas inhabited by the widespread taxon in turn, while the other areas are allowed to "float" on the RAC (ENGHOFF, 1995; MORRONE & CRISCI, 1995).

The solutions allowed under the three assumptions form inclusive sets (PAGE, 1990; VAN VELLER et al., 1999): the *A0* solutions are a subset of the *A1* solutions, and these in turn are a subset of the *A2* solutions (fig. 2). Usually, there are also solutions that violate all three assumptions, namely those in which none of the areas of a widespread taxon occurs in the RAC in the position predicted by the place of the widespread taxon in the TAC (fig. 2: RACs 8–15). Thus, the *A2* solutions are usually a small subset of the "Full Solution Set" (all possible branching arrangements of the studied areas).

Redundant distributions (sympatric taxa) are essentially handled in the same way as widespread distributions. Under *A0* and *A1*, each occurrence of the redundant area is considered as equally valid, i.e., as representing duplicated area patterns. *A2* also considers the possibility that the redundant distributions are the result of dispersal, that is, each occurrence of the redundant distribution is considered separately (ENGHOFF, 1995). *Missing areas* are treated as missing data under *A1* and *A2*, and explained by primitive absence (the taxon has never been in the area), extinction (the taxon went extinct in the area) or inadequate sampling. Under *A0*, missing areas are considered as observations of true absence and explained as due to primitive absence or extinction (ENGHOFF, 1995; MORRONE & CRISCI, 1995).

Application of these assumptions to empirical data has been controversial, as results can differ greatly when the same data set is processed under different assumptions (MORRONE & CARPENTER, 1994; ENGHOFF, 1995; DE JONG, 1998; VAN VELLER et al., 2000). *A0* (and *A1*) has been criticized as being too restrictive and unrealistic because it does not consider the possibility of dispersal in explaining widespread distributions, which means that areas may be grouped together solely based on recent range expansion involving geographically adjacent areas (NELSON & PLATNICK, 1981; HUMPHRIES & PARENTI, 1986; PAGE, 1989, 1990; MORRONE & CARPENTER, 1994). *A2*, on the other hand, has been considered as uninformative or indecisive in that it allows many more solutions than the stricter *A0* and *A1*, and therefore often gives a less resolved result (ENGHOFF, 1995; VAN VELLER et al., 1999). It has also been argued that *A2* (and *A1*) distort the historical (phylogenetic) relationships established in the original taxon cladogram from which the area cladogram is derived (ZANDEE & ROOS, 1987; WILEY, 1988; ENGHOFF, 1996; VAN VELLER et al., 1999, 2000) but this claim seems to arise from a confusion on the meaning of the assumptions: *A2* and *A1* are interpretations of

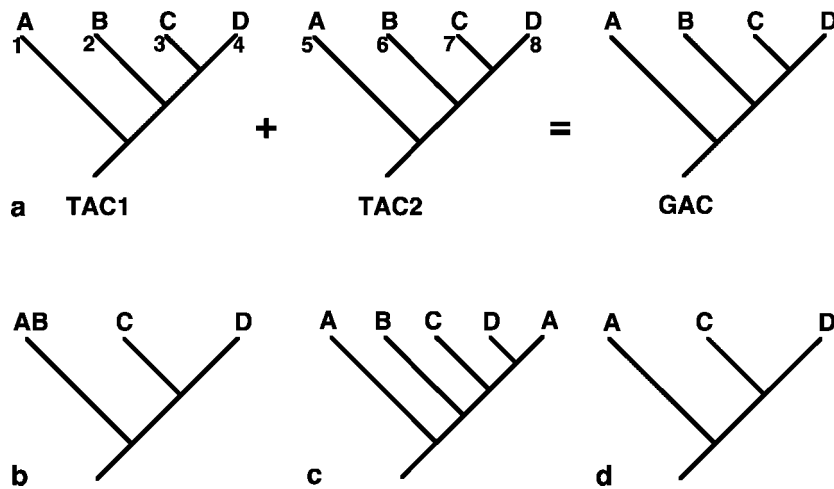


Fig. 1. a. Steps of a cladistic biogeographic analysis: a taxon–area cladogram (TAC) is constructed by replacing the taxa in the phylogeny with the areas in which they occur. Comparing the taxon–area cladograms for different groups reveals the existence of a general biogeographic pattern (GAC). Conflicting area relationships (incongruence) may be indicated by the TAC if this includes: b. Widespread taxa; c. Redundant distributions; d. Missing areas.

Fig. 1. a. Un análisis biogeográfico cladista comprende dos pasos: construcción del cladograma de áreas (TAC) sustituyendo el nombre de los taxones en la filogenia por el área que ocupan y derivación del cladograma general de áreas (GAC). El cladograma de áreas puede indicar relaciones conflictivas entre las áreas si existen: b. Taxones de amplia distribución; c. Distribuciones redundantes; d. Áreas ausentes.

the relationships between areas, not between taxa (PAGE, 1989, 1990).

The problems of widespread taxa, redundancy, and missing areas have mainly been discussed within the traditional *pattern-based approach* to historical biogeography. Pattern-based methods search for general patterns of area relationships (general area cladograms) allegedly without making any assumptions about evolutionary processes (RONQUIST, 1997; 1998a). Biogeographic processes, such as dispersal or extinction, are only considered a posteriori (or using *ad hoc* procedures) in interpreting incongruence between the general area cladogram and the taxon–area cladograms (WILEY, 1988; PAGE, 1994). However, several different combinations of events can usually explain each case of incongruence, leaving the choice of a specific set of events that could explain the observations to the investigator. Pattern-based methods may also give counter-intuitive results in some cases because they do not necessarily favor reconstructions implying likely events over those implying improbable events (RONQUIST, 1995).

Event-based methods, which are explicitly derived from models of biogeographic processes, have gained in popularity recently (RONQUIST & NYLIN, 1990; PAGE, 1995; RONQUIST, 1995, 1998a, 1998b). Unlike pattern-based methods, the event-based reconstructions directly specify the ancestral distributions and the biogeographic events responsible for those distributions, and no *a posteriori* interpretation is necessary. Each type of biogeographic event in the reconstruction is associated with a cost that should be inversely related to the likelihood of that event occurring in the past: the more likely the event, the lower the cost. The optimal biogeographic reconstruction is found by searching for the reconstruction that minimizes the total cost of the implied events (RONQUIST, 1998a, 1998b, in press).

The purpose of this paper is to reexamine the problems of widespread taxa, redundant distributions and missing areas in the light of the event-based approach to historical biogeography. We find that it is only widespread terminal distributions that cause problems in the event-based approach. Because the pattern-based A0,

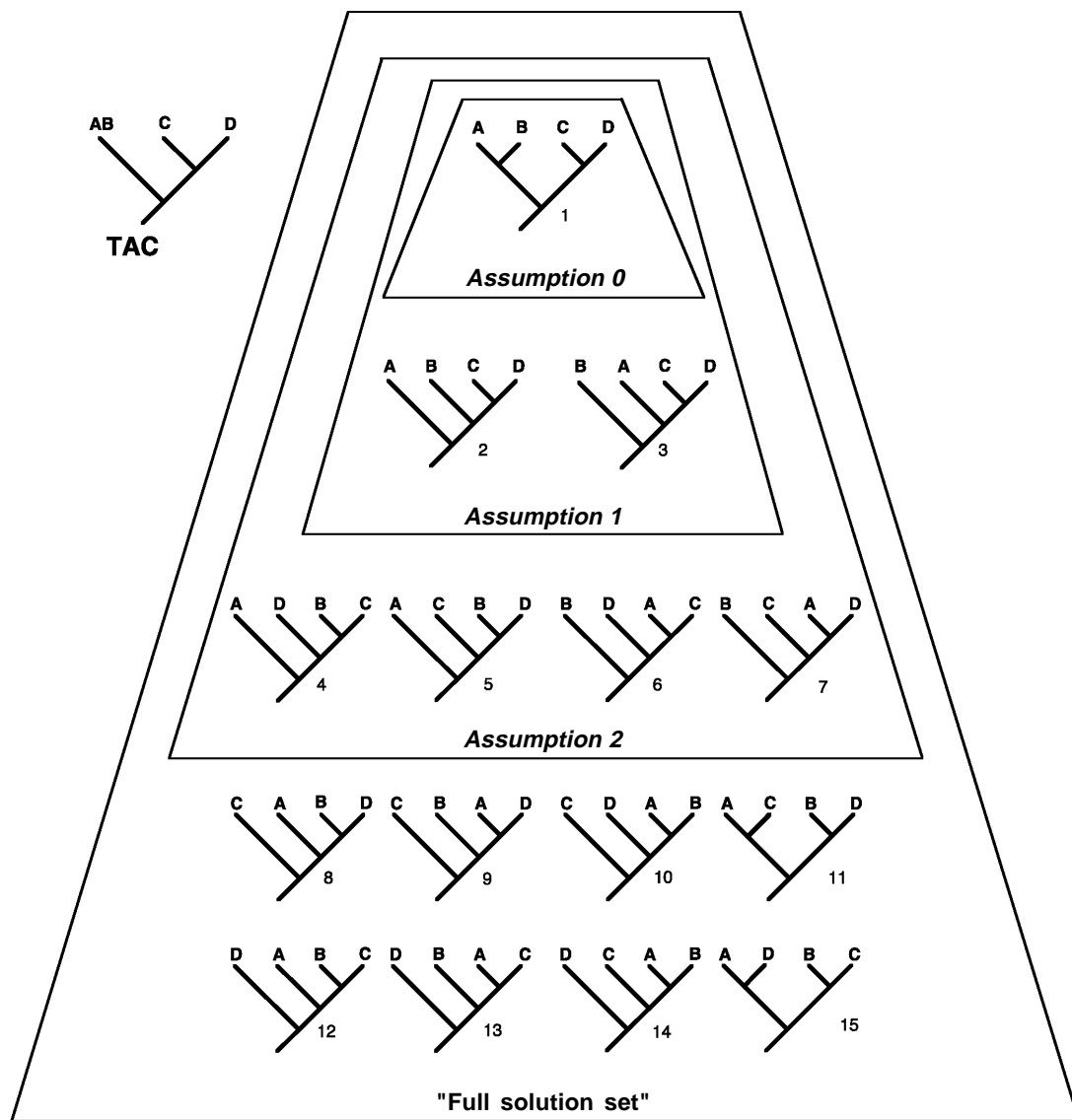


Fig. 2. Taxon–area cladogram with a widespread taxon and three alternative methods of resolution under the pattern–based approach: *Assumptions 0, 1, and 2 (A0, A1, and A2)*. Note that A2 excludes some solutions that are part of the “Full solution set” (all 15 possible rooted binary trees on 4 taxa). (Modified from MORRONE & CRISCI, 1995.)

Fig. 2. Aplicación de las asunciones 0, 1 y 2 (A0, A1 y A2) a la resolución de un cladograma de áreas con un taxón de amplia distribución. Obsérvese que A2 excluye algunas de las soluciones que son parte del “Full solution set” (los 15 posibles cladogramas dicotómicos para 4 áreas). (Adaptado de MORRONE & CRISCI, 1995.)

A1 and A2 only define the set of allowed solutions but not the cost of each solution nor the implied events, they cannot be applied to event-based analyses. Instead, this paper describes three event-based options that may be used to reconcile the occurrence of widespread terminals with the

common assumption of each lineage being restricted to a single area at a time: the *recent, ancient* and *free* options. We give algorithms that implement these options and illustrate their properties by reexamining a classical biogeographic data set, that of ROSEN (1978).

The event-based method

Event-based biogeographic methods rely on explicit models with states (distributions) and transitions between states (biogeographic processes). The most commonly used model includes four different processes (PAGE, 1995): vicariance, duplication, extinction and dispersal. *Vicariance* (v) is allopatric speciation in response to a general dispersal barrier (i.e., a barrier affecting many organisms simultaneously). *Duplication* (d) is sympatric speciation or, alternatively, allopatric speciation due to idiosyncratic events such as a temporary dispersal barrier affecting only a single organism lineage. *Extinction* (e) may simply mean that organisms become extinct in an area but it can also result from the organisms occupying only part of a large ancestral area and therefore being absent in one of the fragments resulting from division of this area. *Dispersal* (i) occurs when organisms colonize a new area separated from their original distribution by a dispersal barrier; this is assumed to be followed by allopatric speciation separating the lineages in the new and old areas.

Once each event type is associated with a cost, the cost of fitting a TAC to a particular GAC can be found by simply summing over the implied events. The GAC with the lowest cost, the most parsimonious GAC, is that which best explains the taxon distributions in the TAC. This optimal GAC can be found, for instance, by explicit enumeration of all possible GACs or by heuristic search for the best GAC. Because inference is based on cost minimization, this approach may be referred to as parsimony-based tree fitting. Similar methods are applicable to problems in coevolutionary inference and in gene tree-species tree fitting (RONQUIST, 1995, 1998a; PAGE & CHARLESTON, 1998).

An important problem in event-based methods is to find the cost for each type of biogeographic process. The most common approach is to work with simple event-cost assignments that focus on one or two of the events and ignore the others (RONQUIST & NYLIN, 1990; RONQUIST, 1995). An example of this is *Maximum vicariance* (or *Maximum cospeciation*; PAGE, 1995; RONQUIST, 1998a, 1998b), in which vicariance events are maximized by associating them with a negative cost (a "benefit", $v = -1$), whereas the other events are not considered in the calculations (duplication (d) = extinction (e) = dispersal (i) = 0). The other approach is to set the cost assignments according to some optimality criterion. A reasonable optimality criterion is to maximize the likelihood of finding phylogenetically conserved distribution patterns (RONQUIST, 1998a, 1998b, in press). Assume that we test for conserved distribution patterns by randomly permuting the terminal taxa of the TAC and comparing the cost of the permuted

data sets with the cost of the original data set. Examination of simulated and real data suggests that, in most cases, chances of finding conserved patterns are best when duplication and vicariance events carry a small cost relative to extinctions and dispersals (RONQUIST, in press). This occurs because both vicariance and duplication are phylogenetically constrained processes, whereas dispersal and extinction are not. In practice, it is often found that the optimal solution is the same under a relatively wide range of event-cost assignments. In the examples discussed in this paper, the cost of vicariance (v) and duplication (d) events are arbitrarily set to 0.01; extinction events (e) to 1.0; and dispersal events (i) to 2.0.

A simple example may illustrate parsimony-based tree fitting in historical biogeography. Consider a TAC with four terminals distributed in four areas (fig. 3a). Each possible GAC for the four areas (there are 15 in all) is fitted in turn to the TAC. For example, only three vicariance events are needed to fit the TAC to GAC1 (fig. 3b), whereas extra dispersal and extinction events must be postulated to explain the observed TAC on GAC2 and GAC3 (figs. 3c–d), and extra duplication and extinction events are needed for GAC4 (fig. 3e). Clearly, GAC1 will be the most parsimonious solution among those considered in figure 3 given the chosen event-cost assignments. Actually, GAC1 will remain optimal under a much wider range of cost assignments: as long as dispersals and extinctions cost more than vicariance events, the optimal solution will be the same. By explicitly enumerating all the 15 GACs and finding the cost of fitting each of them to the given TAC, it can also be demonstrated that GAC1 is the optimal solution.

The optimal reconstruction and the cost for any TAC–GAC combination can be found using fast dynamic programming algorithms (RONQUIST, 1998b). This means that a particular GAC can be fitted to a large set of TACs quickly. Nevertheless, searching for the best GAC using exhaustive algorithms is impractical for problems with more than around 10 areas, in which case heuristic algorithms or other types of exact algorithms should be used instead.

Widespread taxa

Cladistic biogeography focuses on hierarchical ("branching") patterns, in which a sequence of vicariance events successively divides a continuous ancestral area and its biota into smaller components (fig. 4a). This history is described by the GAC (fig. 4b). The terminal branches in the GAC correspond to *present areas* (A, B, C) and the internal branches to *ancestral areas* (E, D), which are combinations of present areas.

In event-based methods (and in pattern-based

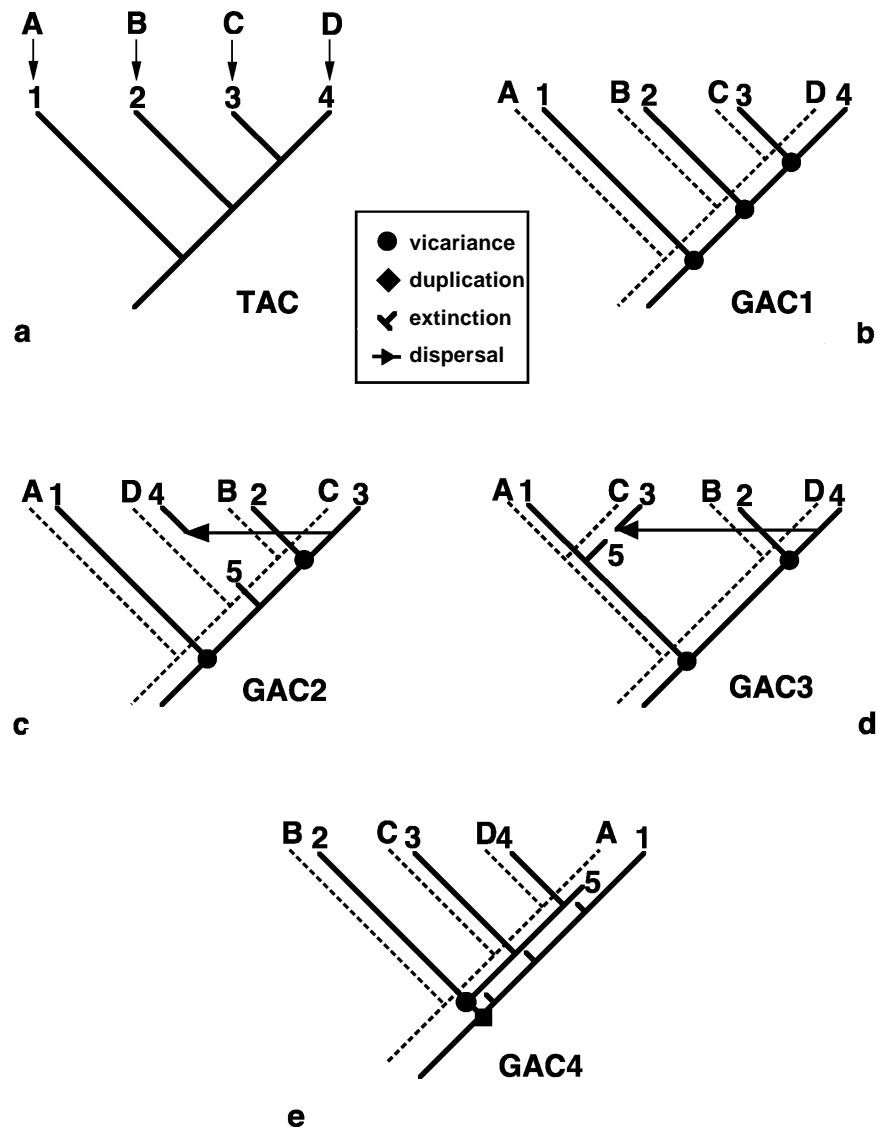


Fig. 3. Example of a reconstruction of the distribution history of a group of organisms in event-based methods: a. A taxon-area cladogram (TAC) distributed in four areas; b–e. Biogeographic reconstruction in which the TAC is fitted in turn into four GACs with alternative resolutions of the relationships between the areas occupied by the TAC. Fitting is evaluated as the cost of the biogeographic events that must be postulated to explain the observed distributions in the TAC according to the GAC. Four types of events are considered in the reconstruction: vicariance, duplication (sympatric speciation), extinction, and dispersal.

Fig. 3. Cómo reconstruir la historia biogeográfica de un grupo de organismos incorporando los procesos en la reconstrucción biogeográfica ("biogeografía cladista de procesos"). a. Cladograma de áreas (TAC) distribuido en cuatro áreas; b–e. Reconstrucción biogeográfica en la que se muestra el grado de congruencia ("ajuste") entre el TAC y cuatro cladogramas generales de áreas (GAC), que difieren en la relación cladogenética entre las áreas que forman parte de la amplia distribución. El "ajuste" se evalúa como el coste de los procesos biogeográficos que deben asumirse para explicar la distribución de los taxones en el TAC de acuerdo con las relaciones entre áreas establecidas por el GAC (ver texto). Se consideran cuatro tipos de procesos biogeográficos: vicariancia, duplicación o especiación simpátrica, extinción y dispersión.

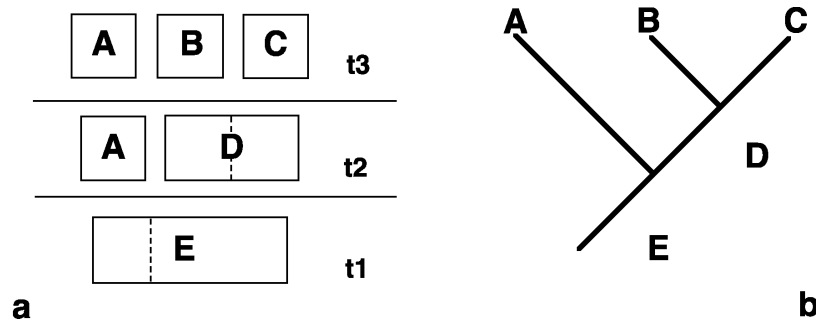


Fig. 4. a. Hierarchical scenario illustrated as a sequence of vicariance events successively subdividing a continuous ancestral area into smaller components; b. The same scenario represented in the form of a tree-shaped diagram, the general area cladogram (GAC).

Fig. 4. a. Escenario biogeográfico jerárquico en el que sucesivos eventos de vicariancia dividen un área ancestral continua en fragmentos más pequeños; b. El mismo escenario, pero representado como un diagrama de árbol, el cladograma general de áreas (GAC).

methods), organism lineages are commonly assumed to be restricted to a single area at a time (for an exception see RONQUIST, 1997); that is, an ancestral distribution must be either a single present area or one of the ancestral areas (combinations of present areas) specified by the GAC. The one area–one lineage assumption makes parsimony-based tree fitting mathematically more tractable but it is also biologically sound: evolving lineages are not normally expected to maintain their coherence over long time periods across major dispersal barriers. However, the assumption causes problems with widespread terminals: how do we reconcile the observation of widespread terminals with the assumption of one area per lineage? The problem is analogous to that of treating polymorphic characters in standard parsimony analysis, in which ancestors are normally assumed to be monomorphic (MADDISON & MADDISON, 1992).

An obvious way of solving the dilemma is to assume that the widespread terminal is in reality not a homogeneous evolutionary lineage but an unresolved higher taxon consisting of a number of lineages, each occurring in a single area (fig. 5a). This does not necessarily imply that the widespread taxon actually comprises different species that have failed to be distinguished (HUMPHRIES & PARENTI, 1986; WILEY, 1988; ENGHOFF, 1996; ZANDEE & ROOS, 1987; VAN VELLER et al., 1999) but it suggests that the widespread distribution is a temporary condition. Now, assuming that the widespread taxon is a soft (unresolved) terminal polytomy with one lineage for each area occupied by the

taxon, we can obtain the minimum cost over all possible resolutions of the polytomy for each ancestral distribution at the base of the polytomy (the node marked with a black dot in the TAC, fig. 5a). For each possible ancestral distribution (i.e., each area in the GAC; fig. 5b), the terminal polytomy is resolved such that the cost of that distribution being ancestral is minimized (fig. 5c). This cost, in turn, is used in the subsequent fitting of the TAC to the GAC. The cost will depend on the GAC because the same ancestral distribution of a widespread taxon may have different costs on different GACs (see fig. 6, table 1).

In determining the possible ancestral distributions of the widespread taxon, we suggest three different options: the *recent*, *ancient* and *free* options. These options constrain the possible ancestral distributions of the widespread taxon in different ways, just like the traditional *Assumptions A0*, *A1* and *A2*. However, unlike the traditional assumptions, the event-based options constrain the solutions by explicitly specifying the processes allowed in explaining the origin of the widespread distribution. Furthermore, each allowed solution is associated with a specific set of events and a specific cost. When many solutions are allowed, they often differ in cost such that they still convey useful information about the grouping of areas in the GAC. On continuation, the event-based options are described in more detail and compared with *Assumptions 0*, *1*, and *2*, both in terms of how they explain the widespread distribution (fig. 5) and how they affect the testing of alternative GACs (fig. 6).

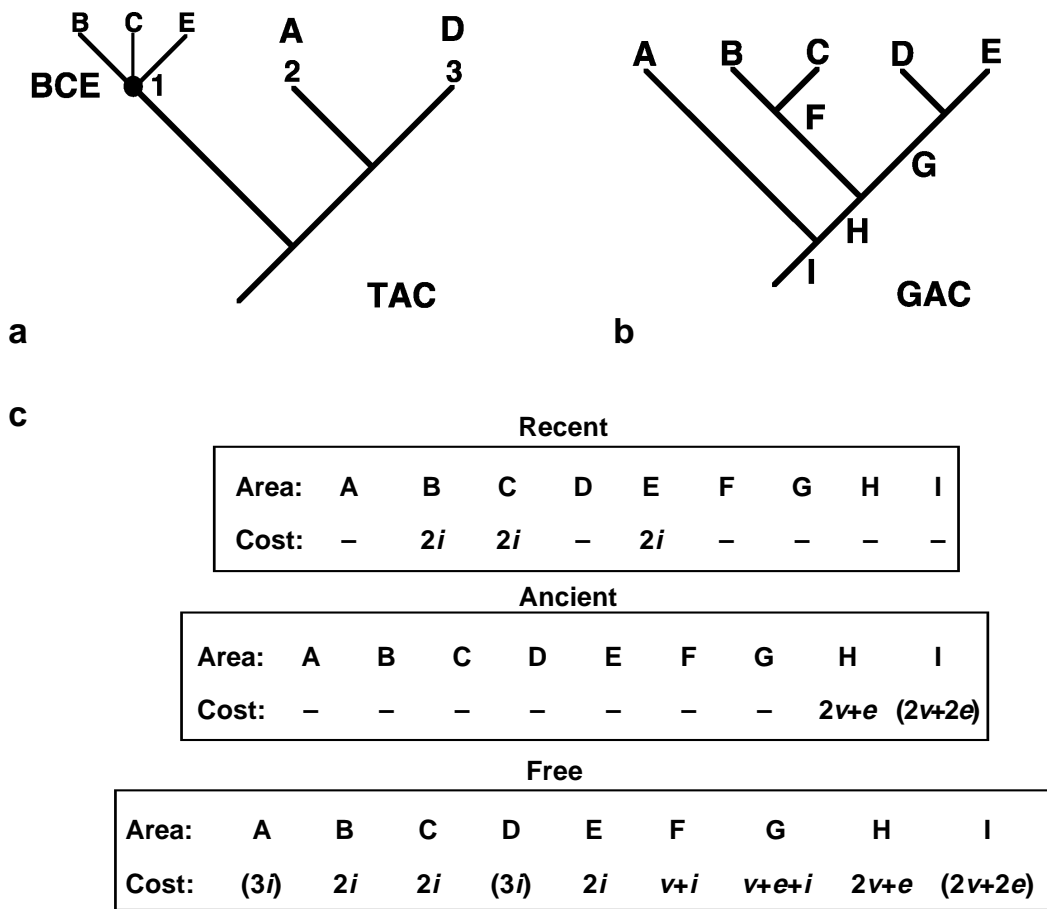


Fig. 5. Resolving the problem of widespread taxa under the event-based approach: a. Taxon area cladogram with a widespread taxon represented as a terminal polytomy of single-area lineages; b. General area cladogram; c. The down-pass cost of each area in the GAC being the ancestral distribution of the widespread taxon (the state at the node marked with a black dot in the TAC) is found by resolving the terminal polytomy in congruence with area relationships in the GAC and then running a down-pass optimization in this subtree. The *recent*, *ancient* and *free* options allow various sets of possible ancestral distributions. The *recent* option only allows the GAC terminals occupied by the taxon (B, C and E); the *ancient* option only allows the immediate ancestor in the GAC of these areas (H), and the *free* option allows both these possibilities plus all intermediates between them (F and G). (Symbols: - Areas not considered as possible ancestral distributions and associated with infinite cost; () Areas that are allowed as ancestral distributions but that will never occur in optimal reconstructions because there will always be more parsimonious solutions; Abbreviations: d-duplication, e-extinction, i-dispersal, v-vicariance).

Fig. 5. Tratamiento de taxones de amplia distribución en "biogeografía cladista de procesos": a. Cladograma de áreas (TAC) con un taxón de amplia distribución representado como una politomía terminal formada por varios linajes, cada uno distribuido en un área; b. Cladograma general de áreas (GAC); c. Para calcular el coste de cada una de las áreas del GAC como posible distribución ancestral del taxón de amplia distribución (el estado ancestral en el nodo señalado con un punto negro en el TAC), se resuelve la politomía terminal de acuerdo con las relaciones entre áreas establecidas por el GAC, y luego se realiza una optimización del coste moviéndose desde los terminales hacia la raíz del subárbol. Las opciones *recent*, *ancient* y *free* permiten diferentes soluciones de las posibles distribuciones ancestrales. La opción *recent* sólo permite las áreas del GAC ocupadas por el taxón (B, C y E); la opción *ancient* sólo permite el área que representa el "ancestro común más reciente" de esas áreas en el GAC (H), mientras que la opción *free* permite cualquiera de estas posibilidades más todas las áreas intermedias entre ellas (F y G). (Para los símbolos y abreviaturas ver arriba).

Recent option

This option is applicable when the widespread distribution can be assumed to be of recent origin. One of the areas inhabited by the widespread taxon is considered the true ancestral area (the center of origin of the taxon) and the others are treated as if added by recent, independent dispersal.

The possible ancestral distributions of the widespread taxon are only those terminal areas occupied by the taxon (B, C, E in fig. 5b). Regardless of whether we are using Maximum Vicariance or any other set of cost assignments is used, the cost C of a present area being the ancestral distribution is simply determined by:

$$C = (n - 1)i$$

where n is the number of areas inhabited by the widespread taxon and i is the dispersal cost (e.g., $C = 2i$ in fig. 5c). The cost of all other GAC areas (terminal areas A, D and ancestral areas F–I in fig. 5b) is set to infinity (an arbitrary high cost) (fig. 5c), since they are not allowed as ancestral distributions.

In terms of explaining the widespread distribution, the recent option (“only dispersal allowed”) is not directly comparable to any of the traditional assumptions. In the context of testing alternative GACs, it will weight against $A0$ solutions in which the areas inhabited by the widespread taxon form a monophyletic clade (fig. 6b; table 1). It will also weight against “Full set” solutions in which all areas harboring the widespread taxon occur in the GAC in positions other than that predicted by the place of the widespread taxon in the TAC (fig. 6e, table 1). These solutions, of course, violate $A2$.

Ancient option

This option is applicable when the widespread distribution can be assumed to be of ancient origin. All areas inhabited by the widespread taxon are considered part of the ancestral distribution. Any mismatch between this distribution and the GAC is then explained as due to extinction; dispersals are not allowed.

Under the ancient option, the only possible ancestral distribution of the widespread taxon is the *most recent common ancestor* in the GAC (“MRCA”) of all of the areas inhabited by the widespread taxon (H in fig. 5b). The GAC areas that are not ancestral to all of the recent areas inhabited by the taxon (A–G in fig. 5b) will require at least one dispersal and are therefore disallowed under the ancient option and are assigned infinite cost (fig. 5c). Areas in the GAC that are ancestral to the MRCA (I in fig. 5b) are allowed but will never occur in optimal reconstructions, as they will always be more costly than the MRCA (fig. 5c).

The cost of the MRCA is calculated assuming that the terminal polytomy is resolved so that the topology fits the GAC perfectly. Under these conditions, only extinction and vicariance events need to be considered because duplications are not required and dispersals are, of course, not allowed. The cost (C) of the MRCA is then given by

$$C = pe + (n - 1)v$$

where p is the number of required extinction events, n is the number of areas inhabited by the widespread taxon, and e and v the costs of the extinction and vicariance events, respectively. The number of required extinction events (p) is computed as follows:

In the GAC, focus on the subtree subtended by the MRCA: ((B, C), (D, E)) in fig. 5b. Assign 1 to the areas harboring the taxon (B, C, E) and 0 to the other areas (D). Then, find the number of losses (p) in this presence/absence character assuming irreversibility ($1 \rightarrow 0$). In fig. 5b, there would be only one loss in area D so the cost is

$$C = 1e + (3 - 1)v = 2v + e$$

In terms of explaining the widespread distribution, the ancient option is similar to $A1$ in that it allows extinctions but not dispersals. In the context of testing alternative GACs, however, it will strongly favor $A0$ solutions in which the areas inhabited by the widespread taxon form a monophyletic clade (fig. 6b; table 1). Thus, widespread taxa provide strong evidence for grouping the areas inhabited by them under the ancient option.

Free option

Under the free option, all possible ancestral areas are considered and any mismatch between the areas inhabited by the widespread taxon and the GAC is explained by the most favorable combination of events. The minimum cost of each possible ancestral distribution is calculated without any constraints on the type of assumed events: dispersals, extinctions, duplications and vicariance events are all allowed.

For the Maximum Vicariance method, the optimal cost of each possible ancestral distribution is found if the terminal polytomy is resolved so that it becomes congruent with the GAC. This might hold for more complex event–cost assignments as well, if the cost of the ancestral distributions is found with algorithms ignoring the complexity of dispersals, the so-called lower bound algorithms (RONQUIST, 1995, 1998b, in press). Why the complexity of dispersals should be ignored is because optimal solutions may occasionally require combinations of dispersals that are impossible on terminal trees congruent with the GAC, but it seems that

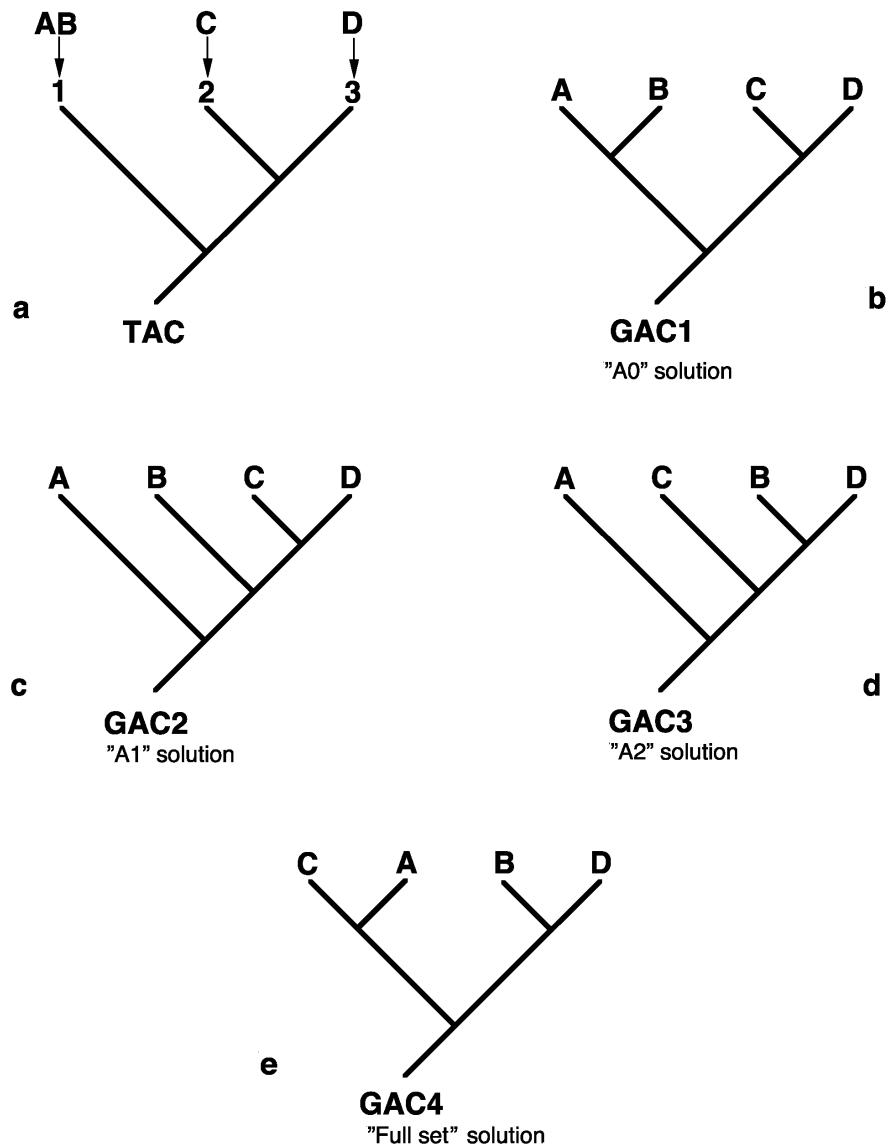


Fig. 6. a. TAC with a widespread taxon; b–e. Four GACs with alternative resolutions of the relationships between the areas occupied by the widespread taxon. These solutions are associated with the traditional pattern-based Assumptions A0, A1, and A2 as follows: A0 would allow only the first solution (b), in which the areas of the widespread taxon form a monophyletic clade; A1 would also allow the second GAC (c), in which the areas are paraphyletic; A2 would allow the third GAC (d), in which the areas are polyphyletic. Finally, the "Full solution set" would include some solutions (e) in which neither of the areas occurred in the position in the GAC predicted by the TAC.

Fig. 6. a. Un TAC con un taxón de amplia distribución; b–e. Cuatro GACs mostrando soluciones distintas de la relación entre las áreas ocupadas por el taxón. Estas soluciones están relacionadas con A0, A1 y A2, de la siguiente manera: A0 permitiría sólo el primer GAC (b), en el que las áreas ocupadas por el taxón de amplia distribución forman un grupo monofilético; A1 permitiría además el segundo GAC (c), en el que las áreas son parafiléticas; A2 permitiría el tercer GAC (d), en el que las áreas son polifiléticas. Finalmente, el "Full solution set" incluiría algunas soluciones (e) en las que ninguna de las áreas en el GAC aparecen en la posición en la que se encuentran en el TAC.

these conflicts can always be solved by rearranging the terminal tree without increasing the total cost (Ronquist, unpublished data). The lower-bound algorithms are computationally extremely efficient so the implementation of the free option is straightforward if this conjecture is true.

In terms of explaining the widespread distribution, the free option is similar to *A2* in that it allows all types of events. However, in the context of comparing alternative GACs, the free option will favor solutions in which the areas inhabited by the widespread taxon form a monophyletic clade, i.e., *A0* solutions (fig. 6b, table 1). The relative cost difference between other solutions will depend on the set of areas inhabited by the widespread taxon and their position in the GAC (table 1). It is interesting to note that, although the free option is similar to *A2* in terms of allowed events, it obviates one of the main criticisms raised against *A2*, namely that it is indecisive. According to the traditional view of *A2*, GACs 1–3 (figs. 6b–6d) would be equally probable solutions, whereas the free option selects GAC 1 (fig. 6b) as the most parsimonious solution. Thus, in this case the free option allows effective selection among alternative GACs.

Missing areas and redundant distributions

In pattern-based methods, missing areas (B in fig. 1d) and redundant distributions (A in fig. 1c) are often identified in the TACs prior to the analysis and different protocols (*A0*, *A1*, and *A2*) are then used to determine the possible RACs. For instance, missing areas can be treated either as missing data or as observations of true absence. If treated as missing data (*A1*, *A2*), absence may be due to primitive absence, extinction, or inadequate sampling and the missing area can thus occupy any position in the RAC. If treated as true absence (*A0*), only primitive absence or extinction are possible explanations. For instance, if several areas are missing from the TAC, this may be taken as evidence that these areas should be grouped in the RAC (extinction) or that the non-missing areas should be grouped (primitive absence). Redundant distributions can be treated under *A0*, *A1* (all occurrences due to ancestry, and any GAC–TAC mismatch explained by duplication and extinction) or under *A2* (some of the occurrences possibly due to dispersal).

In event-based methods, it is difficult to separate potential cases of incongruence that can be identified in TACs prior to analysis (observed) from missing areas and redundant distributions that are introduced during the TAC–GAC fitting process (inferred). If an area is redundant or missing in a TAC simply depends on the general area cladogram (GAC) being analyzed and on the particular events postulated by the reconstruction fitting the TAC to the GAC. The reconstruction

may postulate TAC redundancy that is not apparent before analysis or change the interpretation of which areas are truly missing from the TAC. For instance, a TAC fitted to a congruent GAC will have no missing or redundant areas (figs. 3a, 3b) but if the same TAC is fitted to an incongruent GAC (fig. 3c) one must postulate that some TAC distributions are missing or redundant. A lineage (5) may have become extinct in area D and another taxon (4) may have secondarily re-colonized the same area (fig. 3c). In this reconstruction, there is both a missing area (the absence of taxon 5 in area D) and a redundant distribution (the presence of taxon 4 in area D). However, a different incongruent GAC (fig. 3d) postulates a different set of missing and redundant areas: in this case area C is both the missing area (the absence of taxon 5) and the redundant distribution (the presence of taxon 3). Therefore, *a priori* (observed) and *a posteriori* (inferred) cases of redundancy and missing areas should be treated in the same way in event-based methods; there is no need for special protocols dealing with these cases of incongruence prior to analysis.

The treatment of missing areas in event-based methods is of particular interest. Event-based methods treat missing areas as true absence and explain them as due to primitive absence or extinction. If the missing data interpretation were allowed, then parsimony-based tree fitting would not work because any analysis would be swamped by low-cost solutions postulating events that left no trace in the observed TAC (RONQUIST, in press).

A simple example will illustrate the event-based treatment of missing areas: assume that we have a “two-taxa-two-area” TAC and a four area GAC (fig. 7). GAC 1 (fig. 7a) groups the TAC areas into a monophyletic group (C–D) so a vicariance event is sufficient to explain the history of the organisms; absence of the group in areas A and B is explained as primitive absence. This could mean that the ancestor of the TAC dispersed from an area outside of the considered GAC to the area in the GAC ancestral to C and D, that the outgroups of the TAC occur in areas A and B, or some other alternative. Since we have no information about the outgroups, we cannot distinguish among the alternatives.

GAC 2 (fig. 7b) groups the TAC areas into a paraphyletic group so a vicariance and an extinction event are required to explain the history of the organisms. In GAC 3 (figs. 7c–d), the TAC areas form a polyphyletic group. The TAC can be mapped onto this GAC either by introducing a vicariance and two extinction events (fig. 7c) or one dispersal event (fig. 7d). If vicariance and duplication events are associated with a low cost and dispersal and extinction with high cost, as suggested above, GAC 1 would clearly be favored over GAC 2 and GAC 3. Thus, in searching for the optimal GAC, event-based methods favor scenarios in which the missing areas are explained as being

Table 1. Testing alternative GACs using the event-based approach. For each of the GACs in figs. 6b–e, the minimal cost of fitting them to the TAC in figure 6a is calculated using three different options for treating widespread taxa (*recent*, *ancient* and *free*), and two different event-based methods having different event-cost assignments (parsimony-based tree fitting and maximum vicariance). For parsimony-based tree fitting, the costs of the events are: duplications (d) = vicariance (v) = 0.01, extinction (e) = 1.0 and dispersal (i) = 2.0. For maximum vicariance the event costs are $d = e = i = 0$ and $v = -1$. In calculating the costs, note that the constraints implied by the recent and ancient options ("only dispersals allowed" or "no dispersals allowed", respectively) were applied only to the ancestral distribution of the widespread taxon, not to the distributions of the rest of the nodes in the TAC. Each of the GACs (figs. 6b–e) corresponds to one of the pattern-based *Assumptions 0*, *1* or *2* or the "Full solution set" (see text of fig. 6): * In general, the maximum vicariance reconstruction can be obtained directly from the parsimony-based reconstruction by replacing the cost assignments ($d = i = e = 0$, $v = -1$). However, in this case the least costly reconstruction in parsimony-based tree fitting is cheaper ($v + 2i$) than the one with the maximum number of vicariance events ($d + 2v + 4e$).

*Tabla 1. Comparación de distintos GACs utilizando las opciones recent, ancient y free en la resolución de taxones de amplia distribución. Para cada uno de los GACs en las figuras 6b–6e, calculamos el coste mínimo de "ajuste" al TAC (fig. 6a), utilizando tres opciones diferentes para resolver taxones de amplia distribución (recent, ancient, y free) y dos métodos distintos que difieren en el coste asignado a cada proceso (parsimony-based tree fitting y maximum vicariance). En parsimony-based tree fitting, el coste de cada proceso es: duplicación (d) = vicariancia (v) = 0.01, extinción (e) = 1.0 y dispersión (i) = 2.0. En maximum vicariance, los costes son $d = e = i = 0$ y $v = -1$. Obsérvese que, al calcular los costes, las restricciones impuestas por las opciones reciente y ancestral ("sólo se permiten dispersiones" o "no se permiten dispersiones") se aplican sólo a la distribución ancestral del taxón de amplia distribución, no a las distribuciones del resto de nodos en el TAC. Cada uno de los GACs (figs. 6b–6e) corresponde a una de las tradicionales Asunciones 0, 1 y 2, o al "Full solution set" (ver pie de figura 6): * En general, la reconstrucción con el método de maximum vicariance se obtiene directamente de la reconstrucción con parsimony-based tree fitting reemplazando los costes de los procesos por ($d = i = e = 0$, $v = -1$). Sin embargo, en este caso el coste de la reconstrucción óptima en parsimony-based tree fitting es menor ($v + 2i$) que la que se obtendría considerando el número máximo de eventos de vicariancia ($d + 2v + 4e$).*

Option	GAC1, fig. 6b (Assumption 0)	GAC2, fig. 6c (Assumption 1)	GAC3, fig. 6d (Assumption 2)	GAC4, fig. 6e ("Full solution set")
Parsimony-based tree fitting				
Recent	$2v+e+i = 3.02$	$2v+i = 2.02$	$2v+e+i = 3.02$	$v+2i = 4.01$
Ancient	$3v = 0.03$	$d+2v+3e = 3.03$	$d+2v+4e = 4.03$	$d+2v+4e = 4.03$
Free	$3v = 0.03$	$2v+i = 2.02$	$2v+e+i = 3.02$	$v+2i = 4.01$
Maximum vicariance				
Recent	- 2	- 2	- 2	- 1
Ancient	- 3	- 2	- 2	- 2
Free	- 3	- 2	- 2	- 2*

due to primitive absence, and the rest of the GAC fits the TAC perfectly (GAC1, fig. 7a). The relative cost of other GACs depends on the event-cost assignments. Generally speaking, extinction explanations are favored over dispersal explanations unless the extinction cost is considerably higher than the dispersal cost. Consider, for instance, the two different ways of

fitting the TAC to GAC3 (figs. 7c–d). The extinction explanation (fig. 7c; one vicariance (cost v) and 2 extinctions (cost $2e$) is favored over the dispersal explanation (fig. 7d; one dispersal (cost i)) unless $i < 2e + v$. This is an event-cost assignment scheme that in most cases has a low probability of discovering phylogenetically constrained distribution patterns.

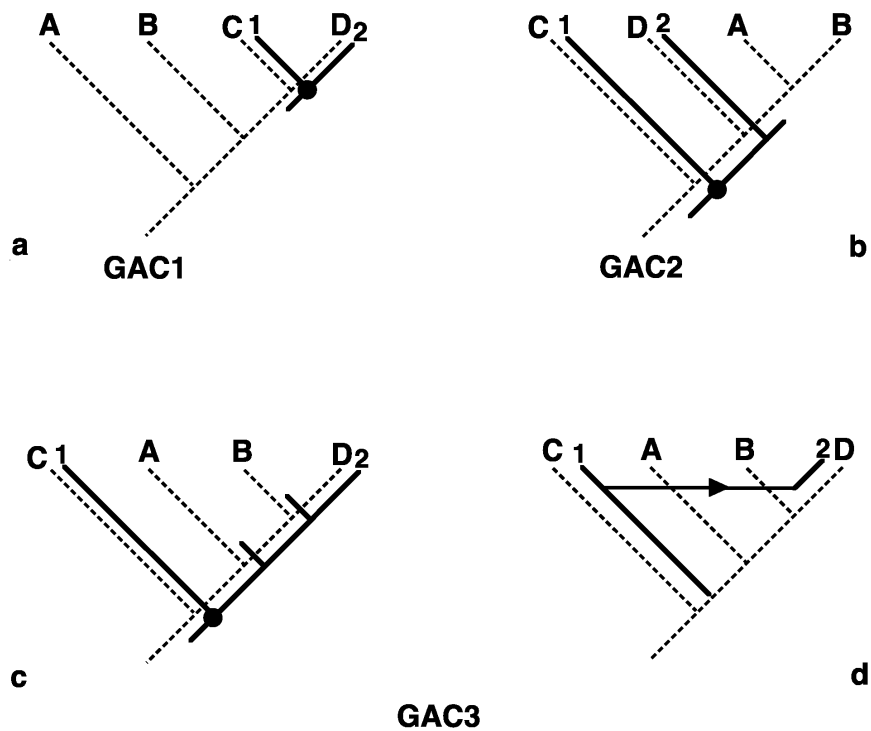


Fig. 7. Treatment of missing areas in the event-based approach. A two-taxon-two-area TAC is fitted to a four-area GAC: a. If the TAC areas are monophyletic in the GAC, only one vicariance event is required; b. If the TAC areas are paraphyletic in the GAC, one extinction and one vicariance event are required; c–d. If the TAC areas are polyphyletic in the GAC, either a vicariance and two extinction events (c) or a dispersal event (d) are required. (Symbols as in figure 3.)

Fig. 7. Tratamiento de las "áreas ausentes" en "biogeografía cladista de procesos". Un cladograma de áreas con dos taxones distribuidos en dos áreas superimpuesto en un GAC de cuatro áreas: a. Si las áreas del TAC son monofiléticas en el GAC, sólo se requiere un evento de vicarianza para explicar las distribuciones en el TAC; b. Si las áreas del TAC son parafiléticas en el GAC, una extinción y un evento de vicarianza son necesarios para explicar la distribución; c–d. Si las áreas del TAC son polifiléticas en el GAC, se necesitan o bien una vicarianza y dos eventos de extinción (c) o un evento de dispersión (d). (Símbolos como en la figura 3.)

As this example clearly demonstrates, absence data are informative in the search for the optimal GAC with event-based methods. The cost of extinction events determines the extent to which absence data influence the search for the GAC: the lower the weight of extinction, the smaller the effect of absence data. A low extinction cost downplays the importance of absence data, regardless of whether this is caused by poor sampling or true absence. Thus, an event-based method with a low extinction cost mimics the missing data treatment of true absences in pattern-based methods. This is a good argument for assigning a lower cost to extinctions than to dispersals in event-based methods of biogeographic absence.

Which assumption should we choose?

Of the three event-based options described above for treating widespread taxa, there is none that is ideally suited to all kinds of problems. Each option has its strengths and weaknesses, and the choice should therefore depend on the nature of the data. The free option is more general in that it allows more processes in explaining widespread terminal distributions. On the negative side, it is computationally more demanding than the other options and because it allows more solutions, it may also be associated with loss of information concerning the optimal GAC. To some extent, however, the potential information loss may be

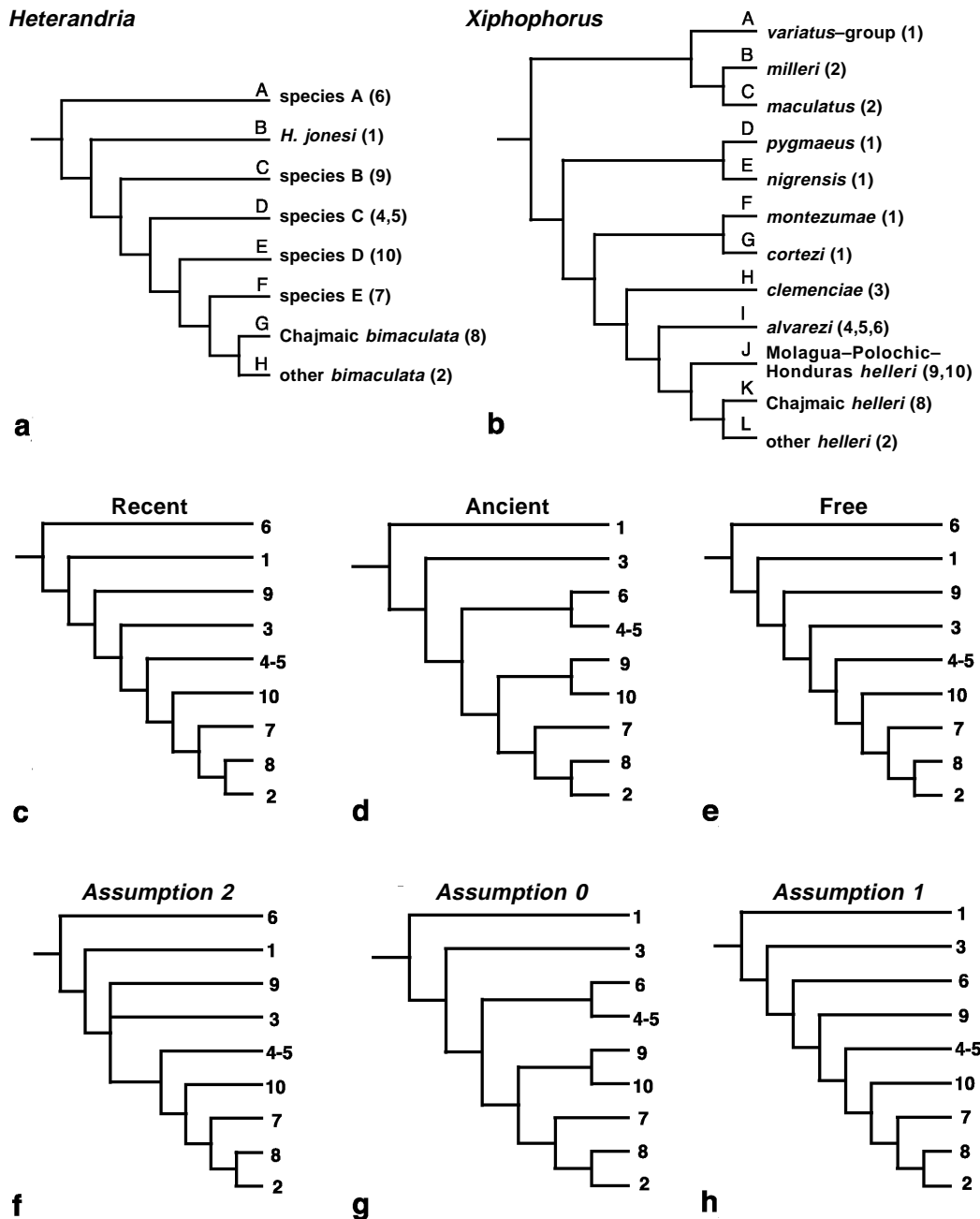


Fig. 8. Example of the application of the event-based options to treat widespread taxa. Taxon-area cladograms for the poeciliid fish genera: *Heterandria* (a) and *Xiphophorus* (b) (After ROSEN, 1978), with areas 4 and 5 combined in accordance with PAGE (1989); c-e. General area cladogram derived with TreeFitter 1.0 (RONQUIST, 2001) for *Heterandria* / *Xiphophorus* under the three different options to treat widespread taxa: *recent*, *ancient*, and *free*; f-h. General area cladograms obtained with Component 2.0 (PAGE, 1993) under A0, A1, and A2 (After VAN VELLER et al., 2000).

Fig. 8. Ejemplo del tratamiento de taxones de amplia distribución en "biogeografía cladista de procesos". Cladograma de áreas (TAC) para los géneros de Poeciliidae: *Heterandria* (a) y *Xiphophorus* (b) (modificado de ROSEN, 1978); las áreas 4 y 5 han sido combinadas en una sola área de distribución como en PAGE (1989); c-e. GACs obtenidos con Tree Fitter 1.0 (RONQUIST, 2001) para *Heterandria* / *Xiphophorus* utilizando tres opciones diferentes para resolver taxones de amplia distribución: *recent*, *ancient* y *free*; f-h. GACs obtenidos con Component 2.0 (PAGE, 1993) utilizando las opciones tradicionales A0, A1 y A2 (adaptado de VAN VELLER et al., 2000).

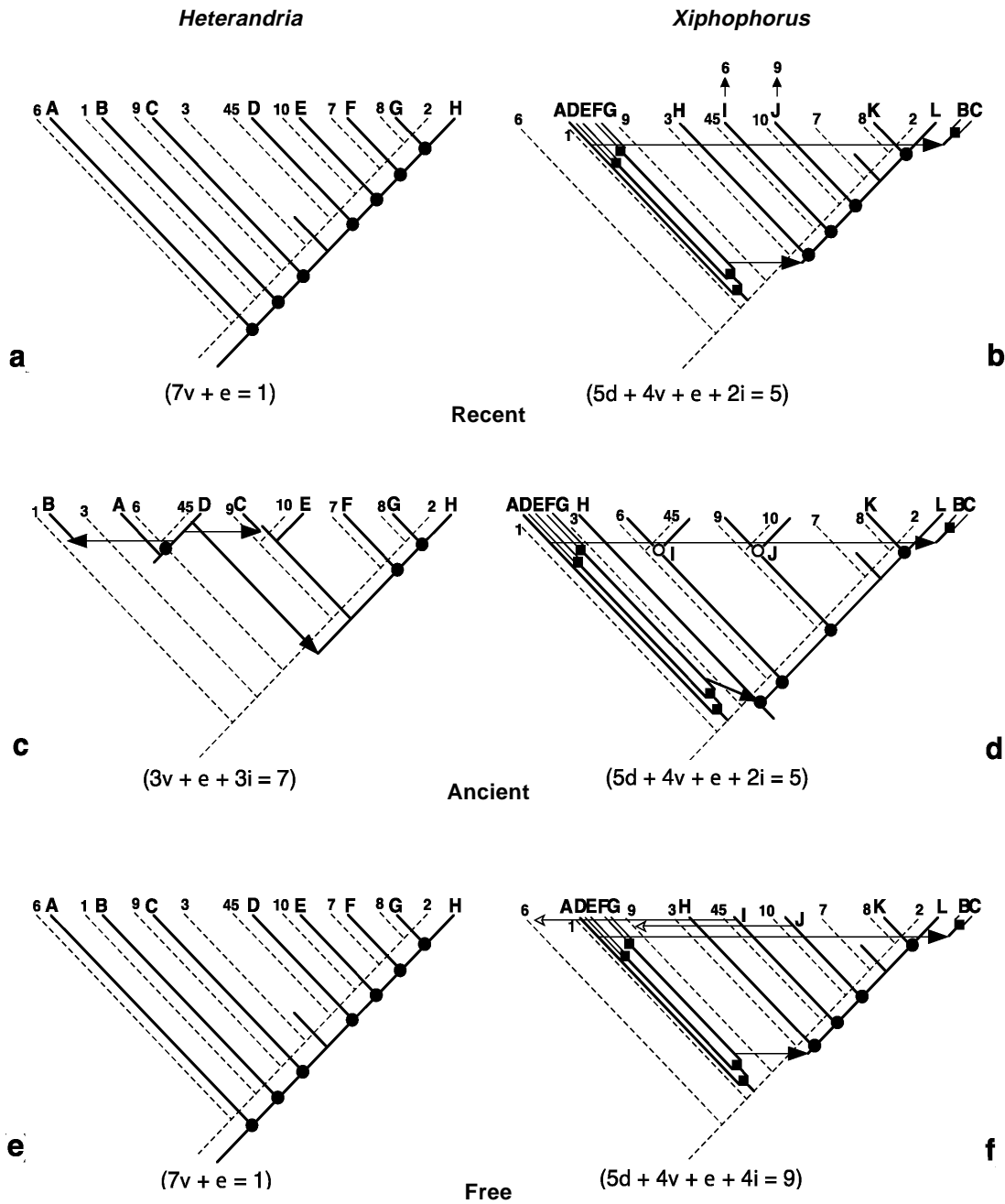


Fig. 9. Event-based reconstructions showing the fit between the TACs of *Heterandria* and *Xiphophorus* (figs. 8a, 8b) and the GACs obtained under the event-based options to treat widespread taxa (figs. 8c–8e). Biogeographic reconstructions of *Heterandria* and *Xiphophorus*: a–b. Under the *recent* option; c–d. Under the *ancient* option; e–f. Under the *free* option. The cost of the reconstruction is indicated under each figure ($d = v = 0$, $e = 1.0$, $i = 2.0$). Symbols as in figure 3. Small black arrows: dispersal events within the terminals that are not considered in the cost of the reconstruction under the *recent* option because they are invariable across the possible GACs (“only dispersal allowed”). Hollow circles: vicariance events within the terminals that are not considered in the cost of the reconstruction under the *ancient* option because they are invariable across the GACs (“only vicariance and extinction allowed”). Hollow arrows: dispersal events within the terminals that are included in the cost of the *free* reconstruction, since this option allows all types of events in explaining the widespread distribution.

counteracted by the differences in the cost associated with the allowed solutions. The ancient option makes the boldest assumptions about the origin of the widespread distributions. If the assumptions are warranted, the search for the optimal GAC should gain in power; if they are not, the result of the analysis may be flawed. For instance, the ancient option might be useful in analyzing the distribution history of old groups that are very unlikely to have dispersed, or in which the widespread taxon has lost the ability to disperse (e.g., a wingless species in a fully winged group).

In many cases, it is quite clear that the widespread terminals are younger than any of the ancestral areas in the GAC, in which case the recent option would be the only defensible choice. The recent option may also be advantageous in the identification of phylogenetically constrained biogeographic patterns because it does not allow vicariance events within widespread terminals, in contrast to the free and ancient options. Assume that we test for constrained distributions by comparing the cost of the observed TAC with that of random TACs obtained either by randomly drawing new TAC topologies or randomly shuffling the TAC terminals. Because the widespread terminals are the same in both the observed and random TACs, the terminal events will not contribute to distinguishing the observed TAC from the random TACs. However, it is quite likely that several of the "terminal" events could be pushed onto the ancestral nodes in the observed TAC but not in the random TACs. This potential support for the GAC is ignored by the free and ancient options. The recent option forces vicariance events onto ancestral nodes in the TAC and is therefore more powerful in separating phylogenetically constrained distribution patterns from random data in this kind of test. For an empirical example, see SANMARTÍN et al. (2001).

Software

The Recent, Ancient and Free event-based options have been implemented in the computer program TreeFitter 1.0 (RONQUIST, 2001). TreeFitter is a program for finding the optimal biogeographic reconstruction/s (GACs), given one or more TACs. TreeFitter is available as free software on the website: <http://www.ebc.uu.se/systzoo/research/treefitter/treefitter.html>.

An empirical example: *Xiphophorus* and *Heterandria* (Rosen, 1978)

ROSEN (1978)'s study on the poeciliid fishes *Heterandria* and *Xiphophorus* is probably the most widely used benchmark data set in the development of biogeographic methods. Because the solutions under *Assumptions 0, 1, and 2* for this data set are well known, it provides a useful comparison with the results of the event-based options.

Figure 8 shows the taxon-area cladograms for *Heterandria* (fig. 8a) and *Xiphophorus* (fig. 8b). They include widespread taxa (e.g., *X. alvarezii* in areas 4, 5, 6), redundant distributions (e.g., area 2 in *Xiphophorus*), and missing areas (e.g., area 3 in *Heterandria* or area 7 in *Xiphophorus*). Using TreeFitter 1.0 (RONQUIST, 2001), we searched for the optimal GAC for the two genera treating widespread taxa under the different event-based options.

The recent option (fig. 8c) finds an optimal GAC that basically follows the pattern of area relationships in *Heterandria*. The areas included in widespread (4–5, 6, 9 and 10) or redundant (2) distributions in *Xiphophorus* are positioned in the optimal GAC according to the TAC of *Heterandria*; only area 3, missing in *Heterandria*, is placed according to its position in *Xiphophorus* (basal to areas 4–5). The optimal GAC under the recent option is one of the three GACs found

Fig. 9. Reconstrucción biogeográfica mostrando el grado de ajuste entre los TACs de Herterandria y Xiphophorus (figs. 8a, 8b) y los tres GACs obtenidos con las tres opciones para resolver taxones de amplia distribución: recent, ancient y free (figs. 8c–8e) en la resolución de taxones de amplia distribución (figs. 8c–8e). Reconstrucción de la historia biogeográfica de Heterandria y Xiphophorus: (a–b) con la opción recent, (c–d) la opción ancient y (e–f) la opción free. Debajo de cada reconstrucción se indica su coste en términos de procesos biogeográficos ($v = d = 0$, $e = 1.0$, $i = 2.0$). Símbolos como en la figura 3. Pequeñas flechas negras: dispersiones dentro de los terminales no consideradas en el coste de la reconstrucción bajo la opción recent porque estarían presentes en todos los posibles GAC ("sólo se permite dispersión", ver texto). Círculos blancos: vicarianzas dentro de los terminales no consideradas en el coste de la reconstrucción ancient porque estarían presentes en todos los posibles GAC ("sólo se permite vicarianza y extinción", ver texto). Flechas blancas: dispersiones dentro de los terminales incluidas en el coste de la reconstrucción free porque esta opción permite cualquier tipo de evento (dispersión, vicarianza, o extinción) para explicar la amplia distribución.

under A2 (fig. 8f) by PAGE (1989) and VAN VELLER et al. (2000) but is different from the single GAC obtained under either A0 (fig. 8g) or A1 (fig. 8h).

The optimal GAC under the ancient option (fig. 8d) agrees mainly with the relationships among areas in *Xiphophorus*. Areas 1 and 3 are placed basally in the cladogram, whereas areas 4–5 and 6, and areas 9 and 10, are grouped together as sister-areas. This is the same GAC found by VAN VELLER et al. (2000) using COMPONENT 2.0 (PAGE, 1993) under A0 (fig. 8g), which is not surprising considering that both the ancient option and A0 group areas based on widespread distributions. It is also similar to the GAC obtained under A1 (fig. 8h) except that the areas forming part of the widespread distribution are not monophyletic in A1. This assumption, like the ancient option, considers the widespread distribution to be ancestral and only allows extinction and vicariance events as possible explanations. In this case, treating the widespread taxa as fully informative about area relationships conflicts with the evidence from endemic taxa because for each pair of areas in a widespread terminal in the *Xiphophorus* TAC (e.g., 9 and 10), the corresponding endemic taxa in the *Heterandria* TAC are not closely related (*species B* and *species D*). Nevertheless, the grouping information provided by the widespread taxa is strong enough to override the signal from the endemic taxa.

The free option finds the same optimal GAC as the recent option. Thus, the widespread terminal distributions in *Xiphophorus* are best explained as due to recent dispersal when all processes are allowed and the cost of all implied events, ancestral as well as terminal, is considered (see fig. 9). As mentioned above, this GAC is one of the three solutions found under A2 by PAGE (1989: his fig. 10) and VAN VELLER et al. (2000: their fig. 13c). The other two solutions place area 3 basal to area 9 or areas 3 and 9 in a monophyletic clade, in both cases requiring an extra extinction event in the event-based framework. For these data, A2 is clearly associated with a loss in resolving power compared to A0 and A1 because it allows three instead of one solution. This information loss does not occur for the free option in the event-based analyses.

Our analyses of the Rosen data show some of the similarities and differences between the traditional pattern-based assumptions and the event-based options. Clearly, there is no one-to-one correspondence between the options and assumptions. Both the recent and free options share properties with A2, whereas the ancient option is more similar to A0 and A1. For Rosen's data, the results obtained with the free option support those obtained with the recent option. This suggests that the ancient option may force unrealistic constraints onto the analysis and that the optimal GAC under the free and recent options may be preferable. This is also the GAC

that is better supported by the phylogenetically determined (as opposed to the within-terminal) area relationships in the two TACs.

Conclusions

The controversy surrounding the treatment of widespread taxa, missing areas and redundant distributions in historical biogeography has been difficult to resolve because of the lack of a common theoretical framework. The event-based approach provides such a framework within which the nature of different methodological options and their effect on biogeographic reconstruction can easily be understood. We hope that our exploration of event-based solutions to the resolution of incongruence in biogeographic inference will contribute to a more focused debate on these issues in the future. The event-based solutions described here should be applicable not only to biogeographic analysis but also to coevolutionary inference.

Acknowledgments

We thank Henrik Enghoff and an anonymous reviewer for useful comments on this manuscript. This research was supported by the Swedish Natural Science Research Council (grant to Fredrik Ronquist) and through a European Community Marie Curie Fellowship (Isabel Sanmartín) under the Improving Human Potential programme (Project MCFI-2000-00794).

References

- DE JONG, H., 1998. In search of historical biogeographic patterns in the western Mediterranean terrestrial fauna. *Biol. J. Linn. Soc.*, 65: 99–164.
- ENGHOFF, H., 1993. Phylogenetic biogeography of a Holarctic group: the Julian millipedes. Cladistic subordinateness as an indicator of dispersal. *J. Biogeography*, 20: 525–536.
- 1995. Historical Biogeography of the Holarctic: area relationships, ancestral areas, and dispersal of non-marine animals. *Cladistics*, 11: 223–263.
- 1996. Widespread taxa, sympatry, dispersal, and an algorithm for resolved area cladograms. *Cladistics*, 12: 349–364.
- HUMPHRIES, C. J. & PARENTI, L. R., 1986. *Cladistic Biogeography*. Oxford University Press, Oxford.
- MADDISON, W. P. & MADDISON, D. R., 1992. *MacClade: Analysis of phylogeny and character evolution*, v. 3.0. Sinauer, Sunderland, Massachusetts.
- MORRONE, J. J. & 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. & CRISCI, J. V., 1995. Historical biogeography: Introduction to Methods. *Annu. Rev. Ecol. Syst.*, 26: 373–401.
- NELSON, G. J. & PLATNICK, N. I., 1981. *Systematics and Biogeography: cladistics and vicariance*. Columbia University Press, New York.
- PAGE, R. D. M., 1989. Comments on component-compatibility in historical biogeography. *Cladistics*, 5: 167–182.
- 1990. Component analysis: A valiant failure? *Cladistics*, 6: 119–136.
 - 1993. *COMPONENT user's manual. Release 2.0*. Natural History Museum, London.
 - 1994. Maps between trees and cladistics analysis of relationships among genes, organisms, and areas. *Syst. Biol.*, 43: 58–77.
 - 1995. Parallel Phylogenies: Reconstructing the history of host–parasite assemblages. *Cladistics*, 10: 155–173.
- PAGE, R. D. M. & CHARLESTON, M. A., 1998. Trees within trees: phylogeny and historical associations. *TREE*, 13: 356–359.
- RONQUIST, F., 1995. Reconstructing the history of host–parasite associations using generalised parsimony. *Cladistics*, 11: 73–89.
- 1997. Dispersal–Vicariance analysis: a new biogeographic approach to the quantification of historical biogeography. *Syst. Biol.*, 46: 195–203.
 - 1998a. Phylogenetic approaches in coevolution and biogeography. *Zool. Scripta*, 26: 313–322.
 - 1998b. Three dimensional cost–matrix optimization and minimum cospeciation. *Cladistics*, 14: 167–172.
- 2001. TreeFitter ver. 1.0. Software available from <http://www.ebc.uu.se/systzoo/research/treefitter/treefitter.html>
 - (in press). Parsimony analysis of coevolving species associations. In: *Cospeciation* (R. D. M. Page, Ed.). University of Chicago Press, Chicago.
- RONQUIST, F. & 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.
- SANMARTÍN, I., ENGHOFF, H. & RONQUIST, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biol. J. Linn. Soc.*, 73: 345–390.
- VAN VELLER, M. G. P., KORNET, D. J. & ZANDEE, M., 2000. Methods in vicariance biogeography: Assessment of the implementations of Assumptions 0, 1, and 2. *Cladistics*, 16: 319–345.
- VAN VELLER, M. G. P., ZANDEE, M. & KORNET, D. J., 1999. Two requirements for obtaining valid common patterns under assumptions 0, 1 and 2 in vicariance biogeography. *Cladistics*, 15: 393–406.
- WILEY, E. Q., 1988. Parsimony analysis and vicariance biogeography. *Syst. Zool.*, 37: 271–290.
- ZANDEE, M. & ROOS, M. C., 1987. Component Compatibility in historical biogeography. *Cladistics*, 3: 305–332.
-