Acta Bot. Neerl. **41**(1), March 1992, p. 75–87

Interpretation of historical biogeographic results

P. C. VAN WELZEN

Rijksherbarium/Hortus Botanicus, P.O. Box 9514, 2300 RA Leiden, The Netherlands

SUMMARY

Historical biogeographic analyses of abstract examples obtained by the parsimonious technique Component Compatability (CC) are interpreted. Subjects such as Brooks' Parsimony Analysis (BPA), vicariance, dispersal, extinction, primitive absence, non-reaction to speciation events, coding techniques, missing areas, interdependence between contemporary species and ancestral species, hybrid areas, outgroup areas, and peripheral areas are treated. Distributions are explained in principle as vicariance events; homoplasies give *ad-hoc* explanations such as dispersal, extinction, etc. Methods still need to be improved as several problems in the interpretation of areagrams are due to their inadequacies.

Key-words: cladistics, component compatability, *Guioa*, historical biogeography, parsimony, phylogeny.

INTRODUCTION

Historical biogeography, a rapidly developing scientific area, is nowadays based on cladistic techniques. The method aims to provide an historical explanation of animal and plant distributions. By comparing the phylogenies and distributions of several (unrelated) groups of animals and plants, a branching pattern (areagram) may be found which shows the historical relationships of the distribution areas involved. The method is based on three assumptions.

1. Evolution does exist.
2. The phylogeny of a group can be reconstructed with the aid of newly developed character states, the so-called apomorphies.
3. In principle, distributions of plants and animals are explained as results of vacariance events because alternative explanations will show up as homoplasies in the areagrams.

Geological events probably both influenced directly (tectonic plate movements) and indirectly (changing climates and ecological conditions) different groups of plants and animals simultaneously by splitting their distributions. The species reacted to permanent separation of the populations with speciation (Geesink & Kornet 1989). Consequently, vicariance (the splitting of ancestral distributions into descendant distributions) can be used as a general explanation of the different distributions. Of course, extinctions and dispersals will have happened, but these are usually typical for single taxa and cannot serve as general explanations; they will be used as *ad-hoc* explanations only when the vicariance explanation fails.

The method used in this paper is one of the Parsimony techniques (as opposed to the Consensus techniques: Nelson & Platnick 1981; Page 1988): The Component Compatability method (CC; Zandee & Roos 1987). The method is implemented in the

computer program CAFCA (Zandee 1988). At the moment two parsimony methods are available, CC and Brooks' Parsimony Analysis (BPA; Wiley 1988a). Both use the same datamatrix but a completely different set of computer algorithms to calculate the areagrams. Usually they end up with the same results. However, CC is preferred over BPA, as the latter method sometimes groups areas on the basis of absence of taxa (van Welzen 1989; P. C. van Welzen & M. Zandee, in prep.).

There is another reason to prefer CC. BPA needs the designation of an outgroup area, which is not necessary with CC. Outgroup areas are not always easy to find as several species may inhabit the same area as the outgroup (*Guioa* in New Guinea; van Welzen 1989); or, when several genera are analysed simultaneously, their outgroups may be found in different areas (e.g. the Australian birds; Cracraft 1986). Consequently Wiley (1988a,b) and Brooks (1990) use an artificial outgroup area which only contains 0's (absenses) in the datamatrix. This procedure may result in splitting off those areas that contain the lowest amount of 1's (presences), which may be incorrect.

Component Compatability (but also Brooks' Parsimony Analysis) is a very useful method for finding area relationships but the interpretation of the results is still difficult. The aim of this paper is to show how several of these difficulties can be solved. With the aid of several abstract examples it will be demonstrated in which way historical biogeographic analyses can be interpreted. The examples are arranged in an increasing degree of complexity. The first example presents no problems in the interpretation of the distributions, it simply demonstrates the method; examples 2–4 show that homoplasies indicate *ad-hoc* explanations such as dispersal, extinction, etc; the last five examples demonstrate computational difficulties and possible solutions.

The alternative to the two parsimony methods, the consensus method or Component Analysis (Nelson & Platnick 1981; Page 1988), differs largely in that, as opposed to the parsimony methods, assumptions about especially widespread distributions are made *a priori* (the so-called assumptions 1 and 2). Therefore, with this method it seems that *a postiori* explanations, very necessary and very troublesome with both parsimony methods, are almost absent. In this paper the choice was made to use the method which needs the minimum number of *a priori* assumptions: CC (assumption 0: take distribution data at face value).

At the present time, the historical biogeography of the islands in the Malesian archipelago is one of the main research areas of several Dutch research groups because of its importance in phylogenetic research. CC features as one of the main methods of analysis.

DATA

Before an historical biogeography of analysis can be made three basic items are needed.

1. Several unrelated monophyletic groups.
2. Their distribution data.
3. The phylogenies of those groups.

Several groups are needed as one group will never show all of its homoplasies (extinction, dispersal, etc.) when used singly in an analysis; the groups have to be unrelated as related groups are more likely to react similarly to speciation events. However, for the sake of simplicity, several examples will only show an analysis of one group. The distributions will show areas of endemism. Van Welzen (1989) used the smallest endemic distributions as units for analysis. Other examples can be found, for instance, Duffels (1986) examined areas of endemism in the Pacific and E. Malesia. The phylogenies are

used to reconstruct the distributions of the ancestral species, because both the distributions of contemporary taxa and those of the ancestral species will be used as data.

In historical biogeography, the rows of the data matrix contain the areas of endemism and the columns contain the species and ancestral species (respectively analogous to the taxa and the characters in cladistic analysis). The distributions of the ancestral species are always the sums of the distributions of their descendants. As the data of the ancestral species are not independent of those of the descendant species, some false homoplasies can be obtained in the resulting areagrams. This will be demonstrated in a few of the examples. The data of the different groups are united into one data matrix by placing them after each other.

The data are binary, a '1' denotes presence, a '0' absence. Widespread species, present in several areas of endemism, have to be scored polytypically in the matrix with a 1 for each area in which they are found.

RESULTS AND DISCUSSION

The very simple example of Figure 1 shows the way in which an historical biogeographic analysis is executed. Genus *Figure* with the species 1–4 is distributed over the areas of endemism A–D respectively; each species inhabits one area only (Fig. 1a). Figure 1b shows the phylogeny of the group, with 5–7 as ancestral species, e.g. 6 is the ancestral species of 2 and 5. Figure 1c shows the data matrix with the species as characters; the ancestral species are taken to have inhabited the sums of the distributions of their descendants. The areagram in Figure 1d is the result: Areas A–D were probably once united with ancestral species 7 living in all of them; area A was split off first, which led to speciation in A (species 1) and B–D (species 6); area B was the second to be split off, followed again by speciation (species 2 and 5 in B and C + D respectively); C and D were the last to split, with the resulting speciation of species 5 into species 3 and 4. In conclusion, the distributions of contemporary species 1–4 can be explained historically by vicariance: the splitting of ancestral distributions initiated speciation with the descendant species living in separate parts of the area inhabited by the ancestral species.

Example 2 (Fig. 2) again shows the genus *Figure* with one species present in two areas: species 1 in areas A and C. There is a widespread distribution and species 1 is scored for 2 areas of endemism, area A and area C. Areas B–D are chosen as areas of endemism because they contain endemic species, area A just remains as a remnant and is therefore recognized as an 'area of endemism' (see also example 8). The areagram shows a homoplasy, a parallel distribution for species 1. In this case, the parallel development is most parsimoniously explained as dispersal of species 1 from area A to area C.

Example 3 in Figure 3 is somewhat more complicated. Here, three genera: *Figure* (species 1–4), *Alpha* (species a–d), and *Roman* (species I–III) are involved. *Figure* and *Alpha* only have endemic species, *Roman* has one species (II) which inhabits area B and C (Fig. 3a). The phylogenies are shown in Figure 3b and the data matrix in Figure 3c. Two areagrams (Fig. 3d) are obtained which only differ in the interpretation of the homoplasious distribution of species II; the left areagram explains the distribution as a parallel development, the right one as a reversal. Usually the left areagram is preferred, with the homoplasy depicted as a parallel. Wiley (1988a) always uses the option DELTRAN with analyses in the computer program PAUP; this only produces the left areagram.

Examples 2 and 3 show parallel developments. Both can be explained by dispersal events, but the situation in 3 can also be explained by indifference to a potential speciation

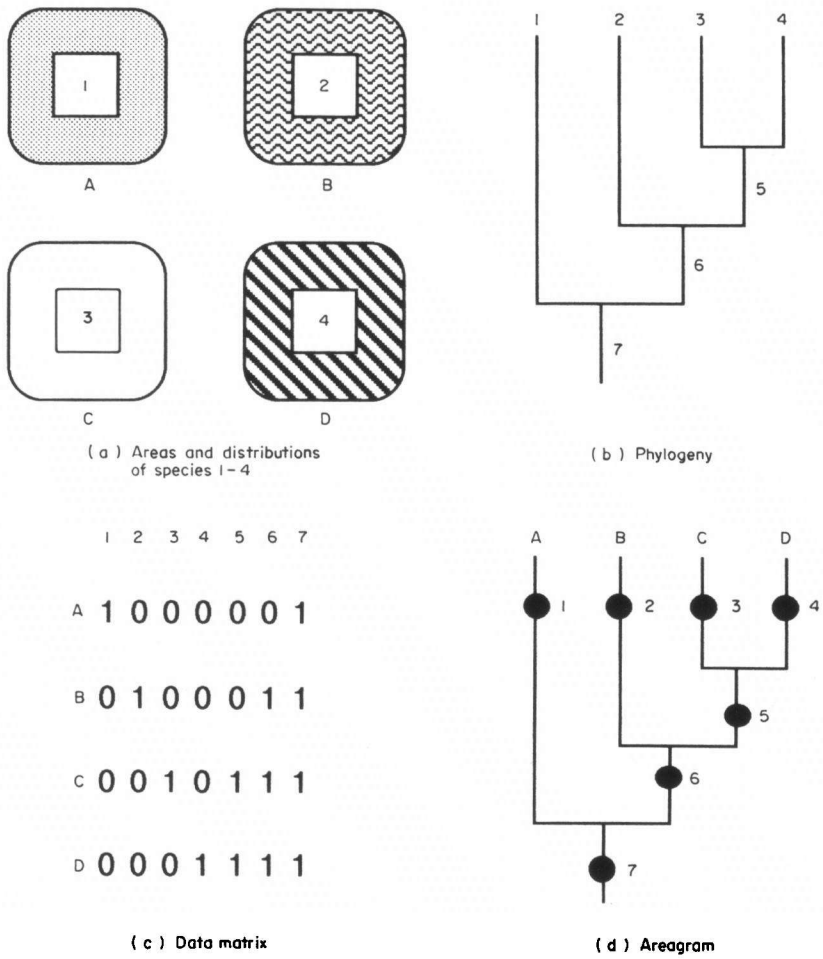


Fig. 1. Genus *Figure* with species 1–4 distributed over areas A–D. The areagram shows no homoplasies. ● = apomorphy.

event: *Roman* did not speciate (or better, did not show morphological difference) with a distinct species on B and one on C+D after area B was split off. The latter situation, indifference, is comparable to parallelisms among species, paraphyletic situations: The same 'ancestral' area, but not all 'descendant' areas included (area B–D acts as 'ancestral' area in the left areagram of Figure 3d, but area D is not included). The first situation (Figure 2) is comparable to convergences among species, polyphyletic situations with different 'ancestral' areas: Area A–D and area C+D in Figure 2d are 'ancestral' areas; missing is the 'ancestral' area inbetween, B–D, this renders the other two polyphyletic. By now it may be clear why the left areagram in Figure 3d is preferred, dispersal (parallel in an areagram) will probably have occurred more often than extinction (reversal in an areagram).

Example 4 in Figure 4 shows the distribution of three genera over four areas; one genus (*Roman*) lacks a species in area D. The result is two reversals in the areagram for the two *Roman* ancestral species IV and V. A reversal may be explained in three ways:

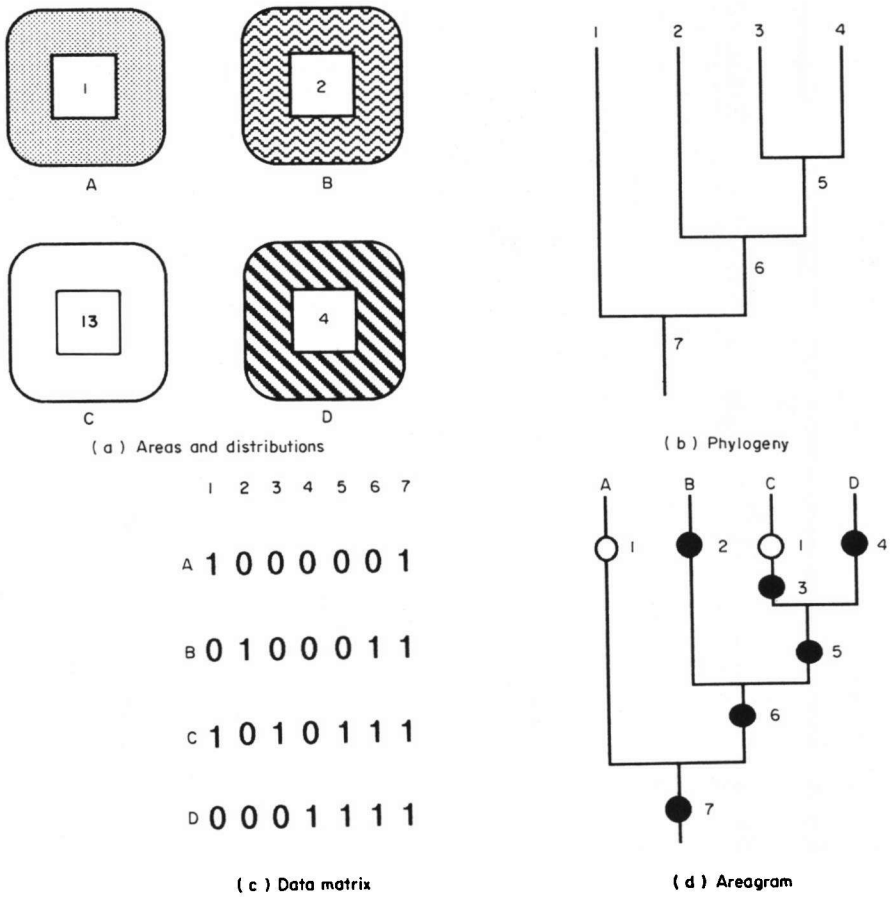


Fig. 2. The distribution of genus *Figure* and the resulting areagram which shows dispersal. ● = apomorphy, ○ = homoplasy.

1. It can be the result of extinction: species IV went extinct in area D.
2. It can be the result of primitive absence: species V was never present in area D.
3. Or, it is an artificial result caused by insufficient collecting in area D.

A real distinction between these three explanations cannot be made. Example 4 shows one of the artificial side-effects of the interdependence between contemporary species and ancestral species: A reversal in species IV also causes a reversal in (part of) its ancestors, i.e. species V. The double reversal has to be explained as one reversal, either species IV went extinct or species IV or V was primitively absent (ignoring deficient collecting).

Example 5 (Fig. 5) shows two genera which overlap only partly: *Figure* is present in area A–D, *Alpha* is present in area D–G, with area D as overlap for both genera (Fig. 5a). The phylogenies of both groups are presented in Fig. 5b and the data matrix in Figure 5c. The combined datamatrix results in two areagrams which are both equally parsimonious and are in a way each other's mirror image; in the left areagram ancestral species e, f, and g are parallel, in the right areagram ancestral species 5, 6, and 7. Note again the influence of the interdependence between contemporary species and ancestral species. All ancestral species are parallel, while the dispersal of only one of them is sufficient to explain the

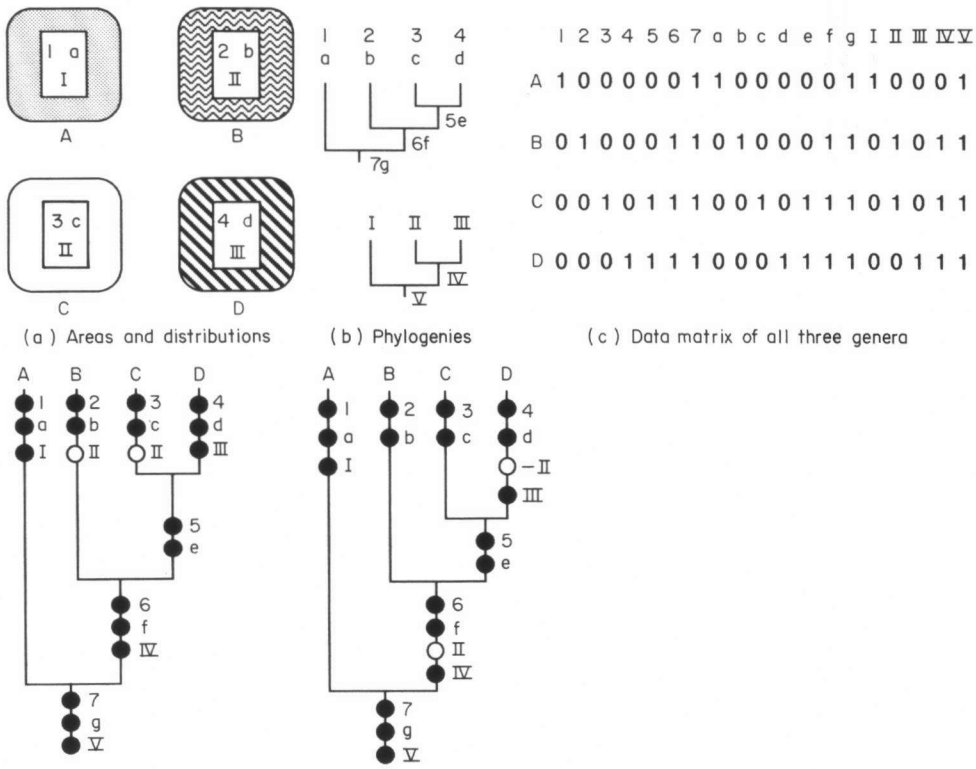


Fig. 3. The distribution of genera *Figure*, *Alpha* and *Roman* over areas A–D, their phylogenies, data matrix and resulting areagrams. Species II does not react to a speciation event. ● = apomorphy, ○ = homoplasy.

areagrams. If species e in the left areagram or species 5 in the right areagram dispersed to area D, then only one dispersal event explains everything; the other ancestral species are just forced by the computer method to become homoplasious too. Both areagrams are an artificial result because both parsimony methods search for one pattern, a so-called generalized areagram, and are unable to find more than one pattern. In the example, area D plays a part in two different processes which should not be united as is done in this example. Figure 5a shows vicariance for the genus *Figure* and dispersal from area G to D for the genus *Alpha* (N.B. the areagrams do not show this, it may as well be *vice versa*, only a comparison with more groups will reveal which taxa dispersed or did show vicariance). Both processes are probably unrelated, may even have occurred in different geological times; this calls for separate analyses for both genera. When analysed separately, the areagrams show no homoplasies (Fig. 5e). The different patterns can only be recognized (a) when all taxa involved are first analysed separately, and (b) when more than one taxon shows one of the patterns. Duffels & de Boer (1990) found two patterns for the Moluccas and New Guinea, one involving the Vogelkop and one without the Vogelkop. Cracraft (1986) had the same problem during his analysis of the distributions of some Australian bird genera. He had two groups of areas with an overlap for only two areas. Both groups were analysed separately, resulting in satisfactory areagrams. Wiley (1988b) united all areas and ended with the computer program PAUP's default number of 50 areagrams, an

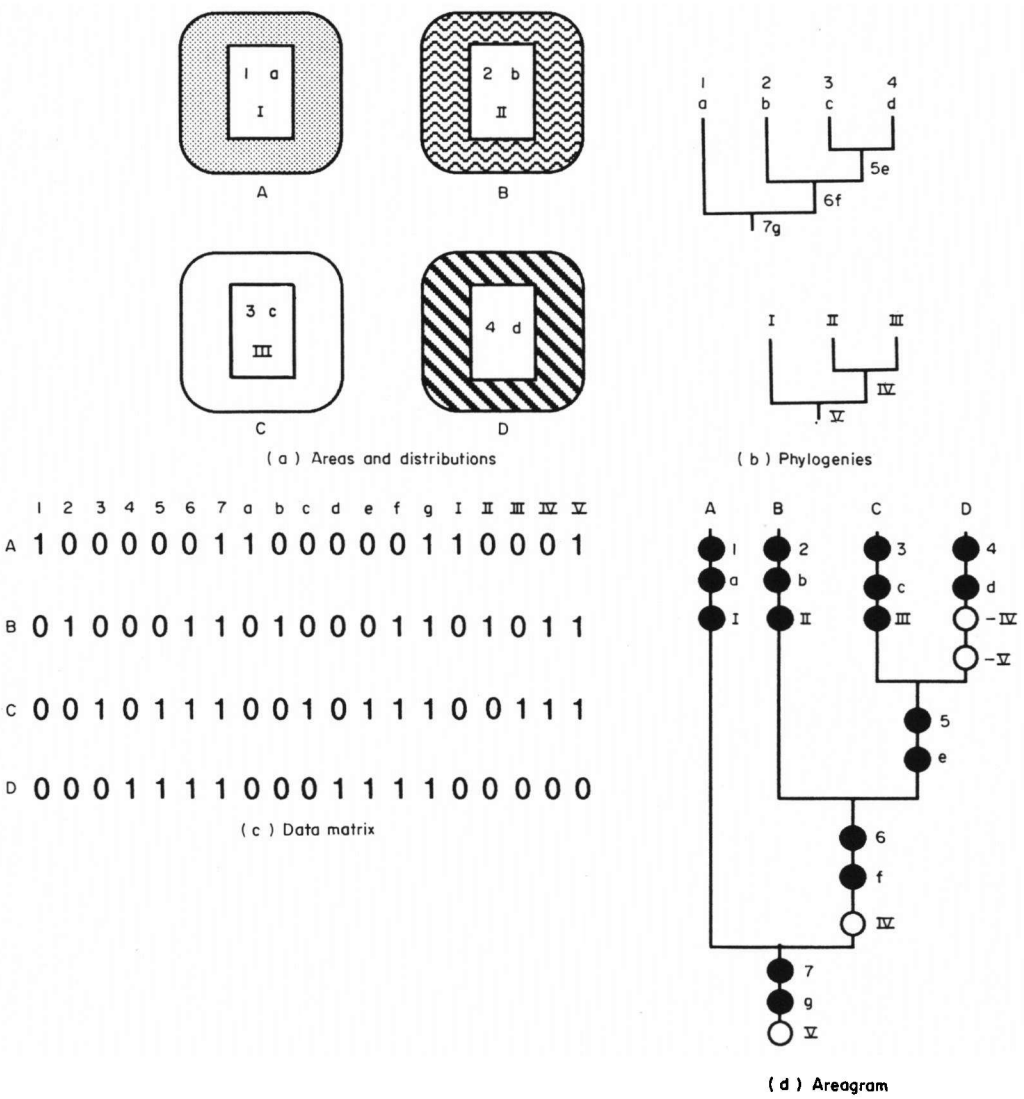


Fig. 4. The distributions and phylogenies of the genera *Figure*, *Alpha* and *Roman*, the data matrix and areagram of the areas A–D. The areagram shows absence of *Roman* for area 5, this can be extinction, primitive absence or incomplete sampling. ● = apomorphy, ○ = homoplasy.

unsatisfactory result. Cracraft (1988) showed in another example for South American groups that areas can be part of different contemporary patterns; in this case the patterns were treated separately too. Brooks (1990) introduced another solution to this problem. He divided area D into two areas, D1, taking part in the analysis of *Figure*, and D2, taking part in the analysis of *Roman*; this is like coding seemingly homologous character states as separate analogous characters after a failing initial analysis which showed them to be convergences. All genera are then still combined into one data matrix and are not analysed separately. N.B. area D can be regarded as a hybrid area, taking part in two patterns.

The right areagram of Figure 5e shows no homoplasies, while Figure 5a shows that dispersal accounts for the distribution of genus *Alpha*. One of the reasons for this result is

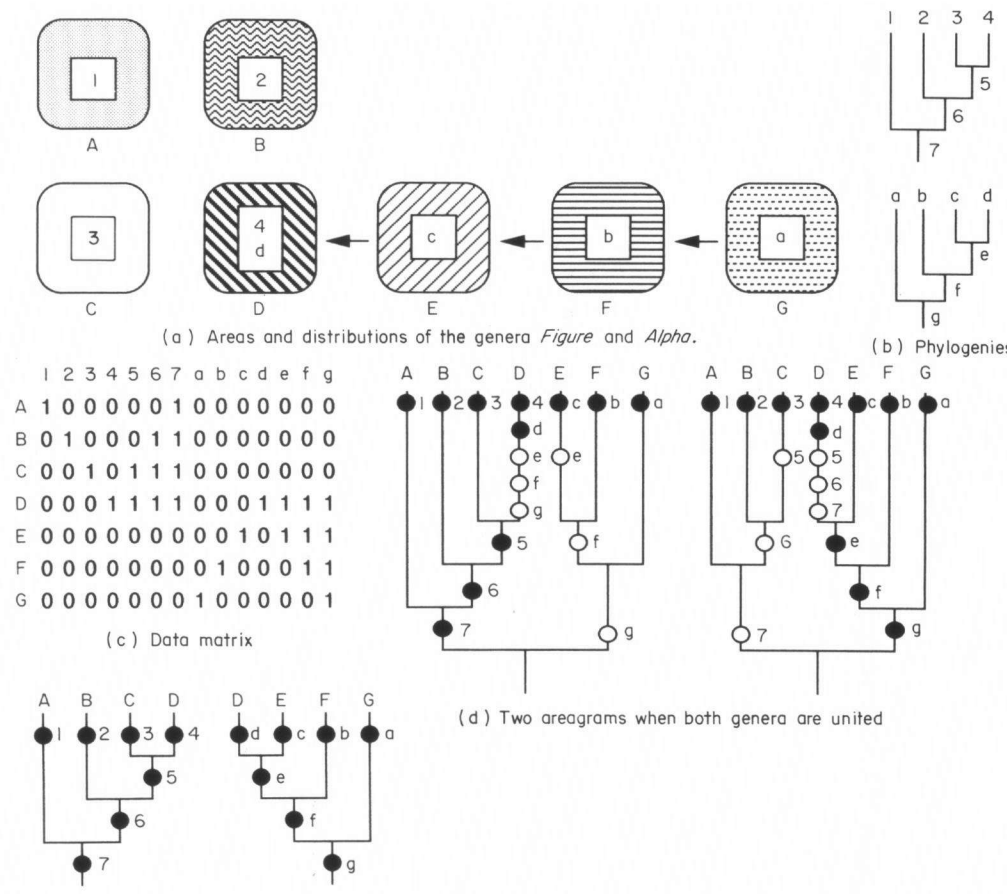


Fig. 5. The distributions of the genera *Figure* and *Alpha* overlap only partly. Combining the two, results in two unacceptable areagrams with many homoplasies. Analysing them separately results in two acceptable areagrams, both Hennigian combs, the left one the result of vicariance events, the right one of dispersal. ● = apomorphy, ○ = homoplasy.

that only one genus is analysed, the other reason is that all species are the result of speciation during dispersal. Similar cases, dispersal camouflaged as vicariance, were found by van Welzen (1989) for the genera *Guioa* (plants) and *Aceropyga* (insects; after Duffels, 1977) over several Pacific island arcs and also for *Guioa* in W. Malesia. The dispersal only becomes apparent when the obtained vicariance events do not coincide with known geological data.

Quite often one or several of the genera are not present in all areas; this is demonstrated by example 6 and 7 in Figures 6 and 7. In Figure 6a, the genus *Alpha* is missing in area A, while area D is a missing area for the genus *Roman*; in Figure 7a *Figure* is missing from area A and *Alpha* from area D. In Figure 7, no genus is present which is distributed over all areas like *Figure* is in Figure 6a. Two options can be used during coding of the distributions in the data matrix. The genera can be scored as absent (0) in the data matrix or as unknown (?). A question mark forces the computer to make several calculations, alternating between absent and present. Wiley (1988a and b) and Brooks (1990) prefer

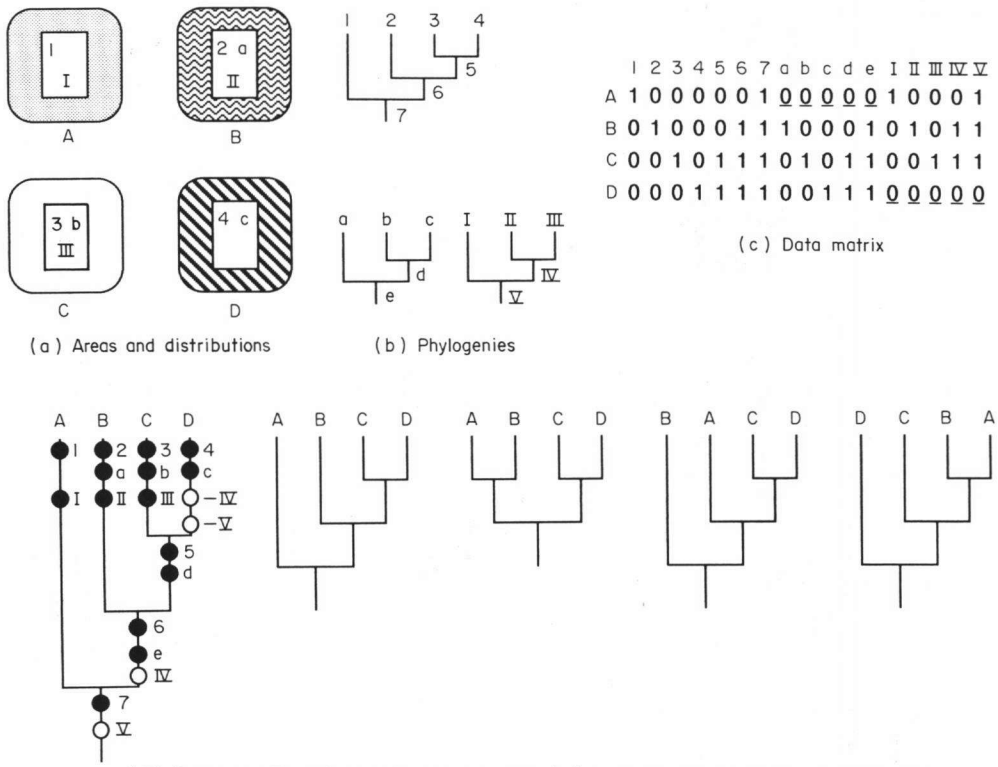


Fig. 6. The genera *Alpha* and *Roman* are both absent from one area, the genus *Figure* is present in all areas. Coding the missing areas alternatively as absent (0) or as unknown (?) provides different results. 0 = alternatively absent and unknown, ● = apomorphy, ○ = homoplasy.

the unknown option, van Welzen (1989) and P. C. van Welzen & M. Zandee (in prep.) prefer the absent option. Both are used in Figures 6 and 7. In Figure 6 the absent option results in one tree (Fig. 6d, left areagram), in Figure 7 the absent option results in three parsimonious areagrams (Fig. 7d, upper left three areagrams). The unknown option results in the right four areagrams in Figure 6d and all 12 areagrams in Figure 7d (almost all theoretically possible areagrams with four areas). The missing option produces more areagrams and prevents the selection of the correct areagram. Figure 7 also demonstrates that with all genera absent in some areas the correct areagram may never be found. Using the unknown option is analogous to using a ? for unknown character states in a cladistic analysis. However, I do not think that missing species can be treated similar to missing character states. Missing character states are scored for, for instance, incompletely known species (e.g. fruit lacking or flowers lacking) or for dependent characters like hairs on leaves when the leaves are absent in one species. In these cases one is certain that the missing characters should be coded as unknown; this will never be the case with missing areas, because one is never certain whether a group is primitively absent in an area or not. One might argue that when a missing area is scored as unknown, all other absences in the data matrix should also be coded as unknown; e.g. species 1 in Figure 6a is absent in areas B–D, this may be due to insufficient collecting, therefore species 1 may be present in all

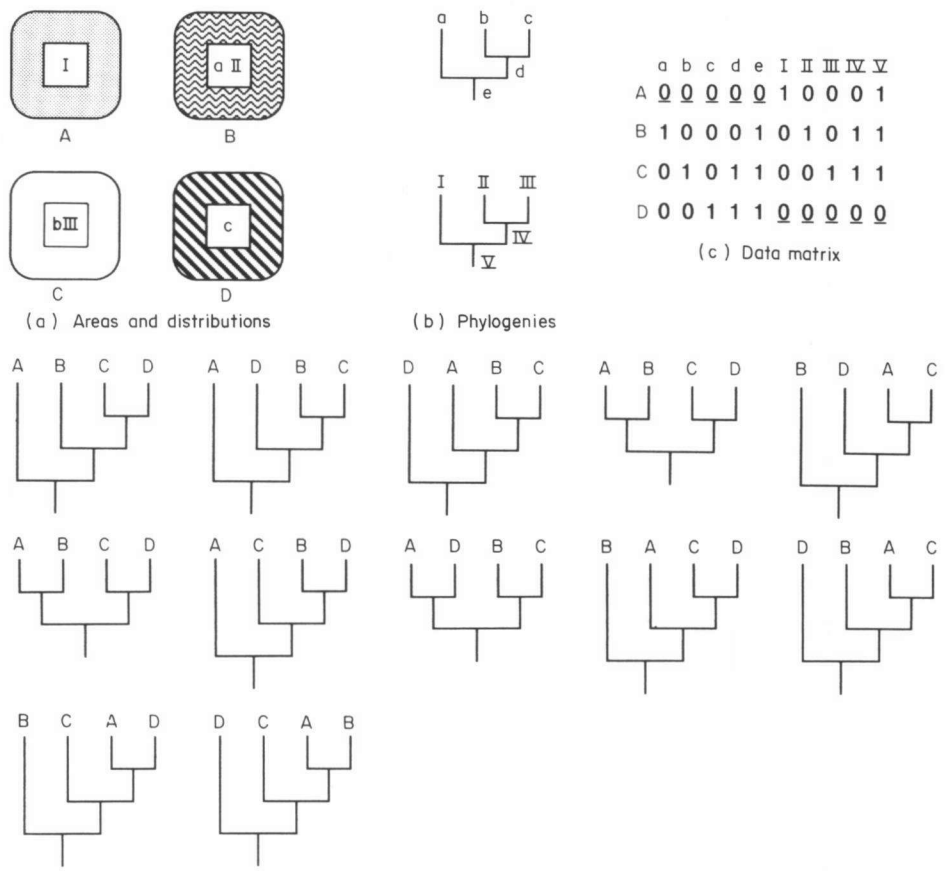


Fig. 7. Both genera, *Roman* and *Alpha*, lack species in one area. The data matrix is alternatively coded as absent (0) for these areas or as unknown (?). 0 = alternatively 0 and ?.

areas and should be coded as unknown for areas B–D. No computer program will be able to analyse the resulting data matrix.

Example 8 in Figure 8a species 3 is widespread over three areas, while area C contains no endemic species. Areas A, B, and D are selected because they contain endemic species, area C remains as a remnant area. Areas like C will end as empty terminal branches in the areagrams (Fig. 8d), while widespread species like species 3 will show up on a lower node, together with ancestral species (sympatric speciation?). The widespread distribution may be a result of dispersal and/or of indifference to potential speciation events. Similar cases can be found in van Welzen (1989) for Central New Caledonia and for Palawan and the S. Philippines.

Example 9 in Figure 9 shows a partial end-result of a historical biogeographic analysis of New Guinea (van Welzen, 1989); the example itself is too large to be presented completely. The recognized areas of endemism can be found in Figure 9a. Of the areagram, only the part relevant to *Guioa rigidiuscula* is shown (Fig. 9b). In the left areagram *G. rigidiuscula* is present with two parallel events and one reversal. The areas in which

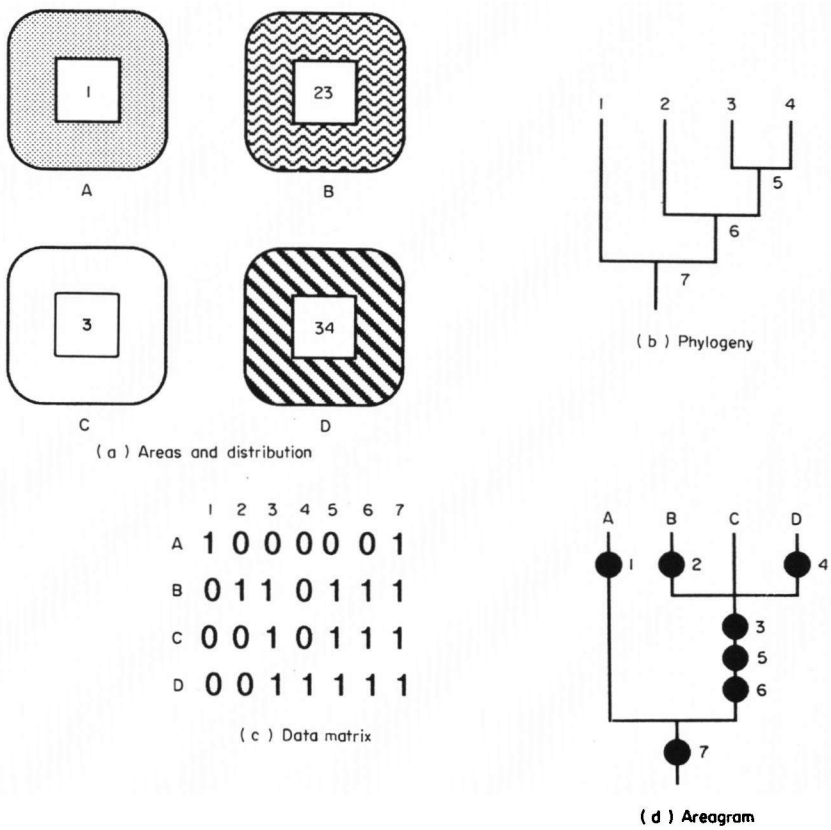
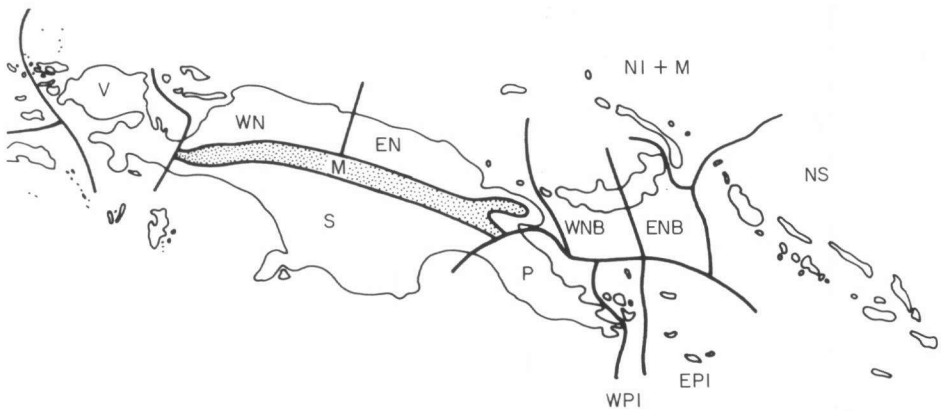


Fig. 8. No species is endemic in area C. This area ends up with an empty branch in the resulting areagram.

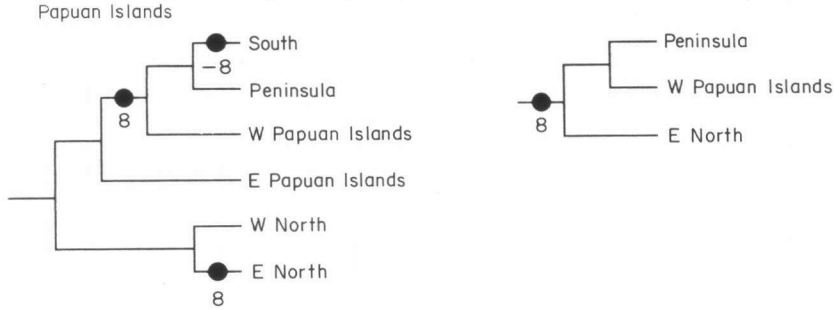
this species is present form one continuous distribution (Peninsula, W. Papuan Islands, and E. North). The parts from which *G. rigidiuscula* is absent are peripheral to this continuous distribution and possibly *G. rigidiuscula* was primitively absent from these areas. Omission in the areagram of the latter areas results in a normal distribution without any homoplasies for *G. rigidiuscula*.

CONCLUSION

Historical biogeographic analyses with either Component Compatability or with Brooks' Parsimony Analysis produce areagrams with quite a number of difficulties in the interpretation of the species distributions. Several of the difficulties are caused by the data themselves (dispersal, extinction, etc.), which show up as homoplasies and indicate the necessity of *ad-hoc* explanations of distributions. Others are due to the methods used, like the interdependence between contemporary species and ancestral species, the inability to accommodate hybrid areas (example 5), the coding of missing areas, and sometimes the concealment of vicariance events because of the presence of peripheral areas (example 9). In conclusion, both methods still need improvement, BPA more so than CC because BPA needs an outgroup area and sometimes distinguishes components on the basis of absence of species.



(a) Division of New Guinea into areas of endemism. EN = E North; ENB = E New Britain; EPI = E Papuan Islands; M = Mountains; NI + M = New Ireland + Manus; NS = N Solomon Islands; P = Peninsula; S = South; V = Vogelkop; WN = W North; WNB = W New Britain; WPI = W Papuan Islands



(b) Part of areagram of New Guinea with relevant data for species 8 (*Guioa rigidiuscula*)

Fig. 9. Optimalization of species 8 (*Guioa rigidiuscula*) in New Guinea. Seemingly this species has a parallel development and a reversal (left areagram). However, when the marginal areas are omitted from the areagram species 8 is present as one apomorphy only (right areagram). After van Welzen 1989.

ACKNOWLEDGEMENTS

Dr R. Geesink and Mr H. Turner are thanked for their critical reading of the manuscript. The idea for this subject was born during discussions with Dr R. Geesink, Dr M. Zandee, and the Theory club of the Rijksherbarium.

REFERENCES

Brooks, D.R. (1990): Parsimony analysis in historical biogeography and coeveolution: Methodological and theoretical update. *Syst. Zool.* **39:** 14–30.

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.

Duffels, J.P. (1977): A revision of the genus *Dicheropygа* Stål, 1870 (Homoptera, Cicadidae). *Monogr. Ned. Entom. Ver.* **8:** 1–227.

Duffels, J.P. (1986): Biogeography of Indo-Pacific Cicadoidea: a tentative recognition of areas of endemism. *Cladistics* **2:** 281–336.

Duffels, J.P. & de Boer, A.J. (1990): Areas of endemism and composite areas in East Malesia. In: Baas, P., Kalkman, K. and Geesink, R. (eds): *The plant diversity of Malesia:* 249–272. Kluwer Academic Publishers, Dordrecht.

Geesink, R. & Kornet, D.J. (1989): Speciation and Malesian Leguminosae. In: Holm-Nielsen, L.B., Nielsen, I.C. & Balslev, H. (eds): *Tropical forests: botanical dynamics, speciation and diversity:* 135–151. Academic Press, London.

Nelson, G. & Platnick, N. (1981): *Systematics and biogeography.* Columbia University Press, New York.

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

Welzen, P.C. van (1989): *Guioa* Cav. (Sapindaceae): Taxonomy, phylogeny, and historical biogeography. *Leiden Bot. Ser.* **12:** 1–315.

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. (1988): *A user manual for CAFCA/PC* (version 1.9.6). Institute of Theoretical Biology, University of Leiden.

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