REVERSALS AS BRANCH SUPPORT IN BIOGEOGRAPHICAL PARSIMONY ANALYSIS

H Enghoff

REVERSALS AS BRANCH SUPPORT
IN BIOGEOGRAPHICAL PARSIMONY ANALYSIS

H. ENGHOFF

Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark
e-mail: henghoff@zmuc.ku.dk

HISTORICAL BIOGEOGRAPHY
PARSIMONY
REVERSALS
BROOKS PARSIMONY ANALYSIS
PARSIMONY ANALYSIS OF ENDEMICITY
MATRIX REPRESENTATION

ABSTRACT. – When Brooks Parsimony Analysis (BPA) or Parsimony Analysis of Endemicity (PAE) is performed under Wagner Parsimony, some branches in the resulting resolved area cladogram(s), may be supported by reversals, interpreted as extinctions. Such support is meaningless, so one may perform the analysis with irreversible characters. This, in turn, will increase the number of independent forward character changes on the resolved area cladogram(s) which can also be problematical in BPA, but not in PAE.

BIOGÉOGRAPHIE HISTORIQUE
PARCIMONIE
RÉVERSIONS
ANALYSE DE PARCIMONIE DE BROOKS
ANALYSE DE PARCIMONIE DE
L'ENDÉMICITÉ
MATRICE

RÉSUMÉ. – La méthode de "Brooks Parsimony Analysis (BPA") ou celle de "Parsimony Analysis of Endemicity (PAE)", avec la parcimonie de Wagner, peuvent aboutir à des branches de cladogramme(s) résolu(s) des aires résultantes, dues à des réversions qui sont interprétées comme des extinctions. Ce type de détermination est sans signification, et il vaut mieux réaliser l'analyse avec des caractères irréversibles. Cependant, cette dernière procédure peut augmenter le nombre des changements indépendants des caractères sur le(s) cladogramme(s) résolu(s) des aires, ce qui peut être un problème dans l'analyse de BPA, mais pas dans l'analyse de PAE.

INTRODUCTION

Derivation of area cladograms from distributional and phylogenetic data on taxa of living or fossil organisms has become a major objective of historical biogeography. Among the methods for translating one (or more) taxon-area cladogram(s)/ **TAC**(s) into one (or more) resolved area cladogram(s)/**RAC**(s) which have been proposed (for reviews, see Morrone & Carpenter 1994, Morrone & Crisci 1995, Humphries & Parenti 1999), some include parsimony analysis of matrices. Brooks Parsimony Analysis, or BPA (Wiley 1988, Brooks 1990) uses area x (taxa+ancestors) matrices. Parsimony Analysis of Endemicity, or PAE (Rosen 1988, 1992) uses area x taxa matrices without considering phylogeny. In both methods, relationships between areas are deduced from shared "apomorphies" as indicated by the parsimony analysis: areas which come out as "sister groups" in the resulting RAC(s) are supposed to share a history not shared by the other areas in the analysis.

The methods of three-area statements (TAS: Nelson & Ladiges 1991) and paralogy-free subtrees (Nelson & Ladiges 1996) also work by parsimony analysis of matrices, but the matrices

are of a more derived nature than simple area x taxa or area x (taxa+ancestors) ones. The present paper is about BPA and PAE exclusively.

In BPA, the TAC is transformed into a matrix in which the areas take the role of taxa, and the taxa and their hypothetical ancestors take the role of characters. Each taxon, and each ancestor (node) is coded as a character for the areas involved; presence is coded as 1, absence as 0. BPA proceeds under "Assumption 0" according to which widespread taxa (i.e., taxa present in more than one of the areas of distribution under consideration) are regarded as valid evidence of area relationships (see discussion in Enghoff 1996). When more than one TAC are involved (in which case the outcome of the analysis is not only one or more RAC(s) but at the same time **(a) general area cladogram(s)** (Morrone & Carpenter 1994, Morrone & Crisci 1995), areas missing from some of the TACs are coded with question marks (Wiley 1988). Under the BPA protocol, the TAC in Fig. 1 transforms into the matrix in Table I. The informative characters in Table I are provided by the widespread taxa and those ancestors which are not scored as present in all areas. It is possible to eliminate the evidence from widespread taxa either by adding the distribution of the sister-group of a widespread taxon to its own distribution (*cf.* Zandee & Roos 1987) or by deleting the

Table I. – Character matrix from the TAC in Fig. 1. Character numbers correspond to taxa 1-5 and hypothetical ancestors 6-9 in Fig. 1. An all-zero outgroup has been added.

Char. no.	12345	6789
OUT	00000	0000
AREA	10000	1001
BREA	10100	1011
CREA	10011	1111
DREA	01001	1111
EREA	00001	0111

Fig. 1. – Taxon-area cladogram for taxa 1-5 in areas AREA-EREA. Ancestors are numbered 6-9.

Table II. – Occurrence of taxa 1-9 in areas AREA-EREA.

area/taxon	1	2	3	4	5	6	7	8	9
AREA	+	-	-	-	-	+	-	-	+
BREA	+	-	+	-	-	+	-	+	+
CREA	+	-	+	+	+	+	+	+	+
DREA	-	+	-	+	+	+	+	+	+
EREA	-	-	-	-	+	-	+	+	+

character provided by the widespread taxon (*cf.* Kluge 1988). BPA has been used by many biogeographers (e.g., Crisci *et al.* 1991, Pollock 1995, Härlin 1996, van Soest & Hajdu 1997, de Jong 1998, Hugot & Cosson 2000). The last-mentioned authors called their method "Constructing general area cladograms by matrix representation with parsimony" but as applied by Hugot and Cosson (2000), this method is identical to BPA. Andersson (1996) recommended BPA although it has been criticized by proponents of alternative methods for RAC construction (see references in Morrone & Crisci 1995).

PAE can be seen as an extreme Assumption 0 approach, because with this method *only* the widespread taxa provide evidence of area relationships; cladistic relationships between taxa are ignored. Under the PAE protocol, the distribution table in Table II transforms into the same matrix as the BPA matrix in Table I. PAE is not a cladogram-based method and thus strictly spoken falls outside the scope of cladistic biogeography although it employs cladistic methodology to detect area relationships. PAE has been used by several biogeographers and palaeobiogeographers (e.g. Fortey & Mellish 1992, Morrone 1994a, da Silva & Oren 1996), and a very similar approach has found its way into synecology (Nel *et al.* 1998). A PAE-like method was suggested already by Legendre (1986), and more recently PAE has been recommended for delimiting the "areas of endemism" forming the units in cladistic biogeographical analysis (Morrone 1994b).
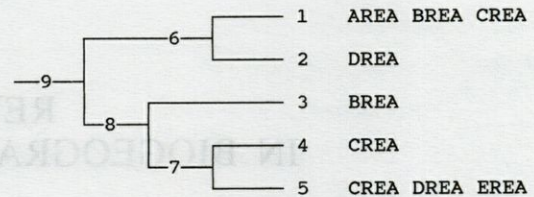
The following discussion applies to both BPA and PAE. For parsimony analysis, an all-zero outgroup is added to the matrix, corresponding to a hypothetical area where the organisms in question do not occur at all. A non-zero root has been used by some authors, e.g. van Welzen (1989) and Härlin (1996), but this "subspecies" of BPA is not considered here. To derive resolved are cladograms from the matrix, authors have routinely employed the simplest parsimony criterion, Wagner parsimony, in which characters are free to change back and forth with equal probabilities. Characters in biogeographical matrices are usually binary (but see Mayden 1988), issues regarding ordered characters are therefore not relevant.

HOMOPLASY IN BPA AND PAE

If characters (or, to be more specific: taxa and ancestors) are optimized on a RAC derived using BPA or PAE they may show homoplasy. Some taxa/ancestors may appear more than once on the RAC, and some may show reversals.

Taxa or ancestors appearing several times on the RAC can be interpreted in terms of dispersal (Wiley 1988, Brooks 1990, Page 1990). It should, however, be borne in mind that such interpretations, as far as BPA is concerned, founder on the lack of independence between the characters in the matrix (Page 1987, 1990, see also Mayden 1988): a character derived from an ancestor is not independent from characters derived from that ancestor's descendants. PAE does not suffer from this weakness because the taxa providing the characters can be regarded as independent evidence.

Reversals are interpreted as extinctions in BPA (Wiley 1988, Brooks 1990, Brooks & McLennan 1991, see also Ronquist & Nylin 1990): a taxon may be absent from an area because of extinction, and such an extinction may appear as a reversal on the RAC (but see Mayden 1988: 349-350 for a discussion of "apparent extinctions"). To minimize the number of reversals, Wiley (1988) recommended using DELTRAN optimization of characters.
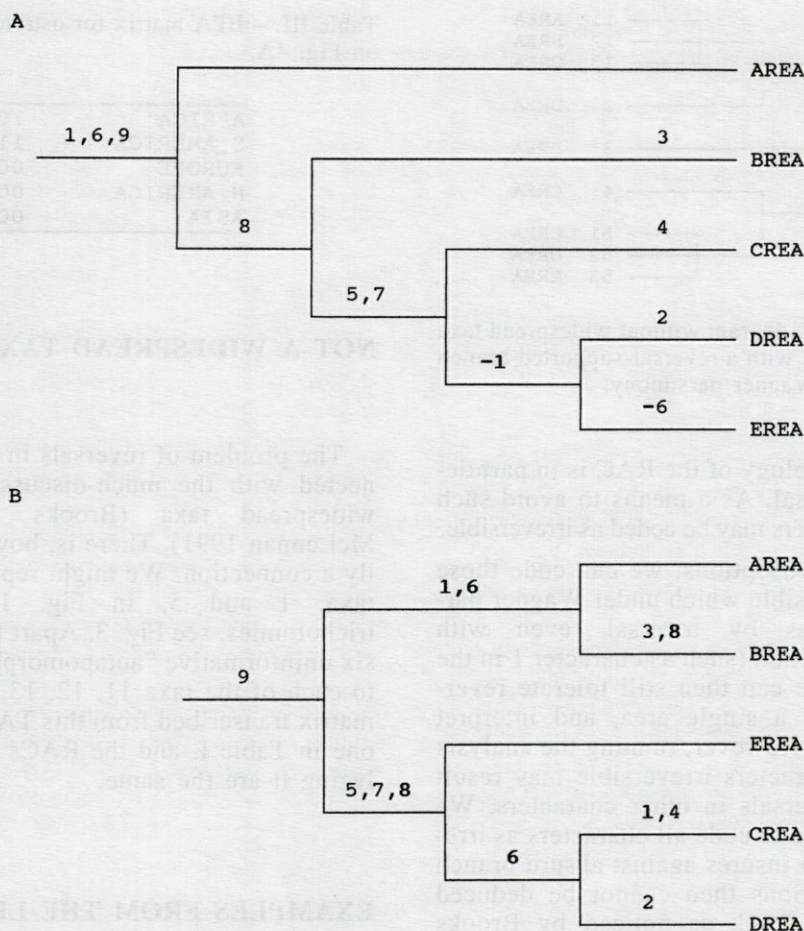
Fig. 2. – Most parsimonious trees (RAC) derived by BPA from the matrix in Table I. Numbers indicate characters supporting the branch below; negative numbers : reversals, bold numbers : characters showing convergence, negative bold number : non-terminal reversal. A, Single most parsimonious RAC found with Wagner parsimony. B, Single most parsimonious RAC found with all characters coded as irreversible.

REVERSALS AS SUPPORT FOR AREA RELATIONSHIPS

If a reversal occurs on a terminal branch of the RAC, it has no influence on RAC topology, irrespective of how it is interpreted *a posteriori*. If it occurs on a non-terminal branch, it may, however, contribute to determining RAC topology. Specifically, a reversal on a non-terminal branch of a RAC supports relationship among the areas united by this branch. This does not make sense because although such a non-terminal reversal might indeed represent extinction, and although one extinction in an "ancestral" area might be considered more parsimonious than independent extinctions in independent areas, the case for discarding independent extinction is much weaker than the opposite: discarding independent immigration in favour of shared ancestral presence in two or more areas: "A shared absence could be local extinction as a result of very different processes due to different

causes...." (Nel *et al.* 1998: 175). This problem with BPA and PAE seems not to have been addressed so far but certainly deserves attention. Van Welzen (1989) discarded BPA in favour of component compatibility analysis (Zandee & Roos 1987) because some branches in his BPA-RACs were supported by absences of species or ancestors. Because of the non-zero rooting employed by van Welzen, these absences do, however, not correspond to reversals

Consider the TAC in Fig. 1. If the corresponding BPA matrix (which could equally well be a PAE matrix, *cf.* above) is analyzed with Wagner parsimony, e.g., using the branch and bound option of PAUP, version 3.1.1. (Swofford 1993), one tree, or RAC, is obtained (length 11, ci 0.818, ri 0.750), see Fig. 2A.

If we examine the support for the branches of this RAC, we find that the branch uniting areas DREA and EREA is supported only by a reversal in character 1. Areas DREA and EREA thus appear as sister areas only because they *lack* Taxon 1. In
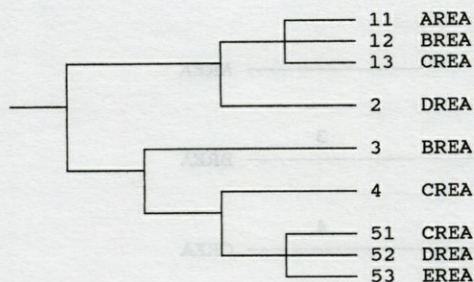
Fig. 3. – Taxon-area cladogram without widespread taxa but resulting in a RAC with a reversal-supported branch when analyzed with Wagner parsimony.

Table III. – BPA matrix for ostariophysan fishes, based on Fig. 4A.

AFRICA	101011111
S_AMERICA	110111111
EUROPE	001010111
N_AMERICA	001010111
ASIA	001010111

other words, the topology of the RAC is in part determined by a reversal. As a means to avoid such absurd RACs, characters may be coded as irreversible.

There now are two options: we can code those characters as irreversible which under Wagner parsimony unite areas by reversal, even with DELTRAN optimization (such as character 1 in the above example). We can then still tolerate reversals involving only a single area, and interpret them as extinctions. However, running the analysis with only some characters irreversible may result in non-terminal reversals in other characters. We may therefore choose to code all characters as irreversible. This option insures against absurd branch support, but extinctions then cannot be deduced from the resulting RAC, as noticed by Brooks (1990: 20).

Applying the first option to the matrix in Table I, coding only the reverting character 1 as irreversible, PAUP (branch and bound) finds 28 RACs (length 12 ci 0.750 ri 0.625) with a fully unresolved consensus, so in this case, this option is uninformative.

Applying the second option, coding all characters as irreversible, PAUP (branch and bound) finds one RAC (length 12, ci 0.750 ri 0.812), see Fig. 2B. This RAC is quite different from the one found with Wagner parsimony. In the "irreversible" RAC, the clade AREA+BREA is supported by character 1, which is parallelled in CREA, and by character 6, which is parallelled in CREA+DREA, the latter clade being only supported by character 6.

Whereas irreversible coding avoids branch support by reversals, it may thus increase the number of other problematical character optimisations in BPA. These problems with BPA were discussed by Wiley (1988), Brooks (1990), Brooks McLennan (1991) and Page (1990); they do not apply to PAE.

Irreversible coding in BPA has been suggested previously (Wiley 1988, Brooks 1990) but these authors did not specifically refer to the problem posed by branch support by non-terminal reversals in this connection.

NOT A WIDESPREAD TAXON PROBLEM

The problem of reversals in BPA has been connected with the much-discussed phenomenon of widespread taxa (Brooks 1990, Brooks & McLennan 1991). There is, however, not necessarily a connection. We might replace the widespread taxa, 1 and 5, in Fig. 1 with unresolved trichotomies, see Fig. 3. Apart from the addition of six uninformative "autapomorphies" corresponding to each of the taxa 11, 12, 13, 51, 52 and 53, the matrix transcribed from this TAC is identical to the one in Table I, and the RACs resulting from analyzing it are the same.

EXAMPLES FROM THE LITERATURE

Mayden (1988) analyzed relationships of North American drainage systems based on distributions and relationships of freshwater fishes. Two of his area clades (clade 15 and 16 in his fig. 5) are each supported by one reversal and one "positive" character (presence of a particular species of fish). Reanalysis with PAUP of Mayden's data matrix for the 15 areas forming his clade 10, using the original data coding and weighting scheme, shows that clade 15 is robust to irreversible coding of all characters but that clade 16 is not. With only the two reverting characters coded as irreversible, neither clade 15 nor 16 appears in the strict consensus. Reversals thus do play a role in determining the tree topology in this case.

Brooks & McLennan (1991: 235-237) reworked the analysis by Novacek and Marshall (1976) of the distribution of ostariophysan freshwater fishes. The TAC for this group (Fig. 4A) transcribes into the matrix in Table III. Brooks & McLennan (1991: fig. 7.31) gave two equally parsimonious RACs for this matrix (length 11 steps), see Fig. 4B-C. However, when analyzed with PAUP using Wagner parsimony two trees of length 10 are found, see Fig. 4D-E (inspection of fig. 7.31 of Brooks & McLennan also shows that one step can be saved by connecting AFRICA and S-AMERICA).
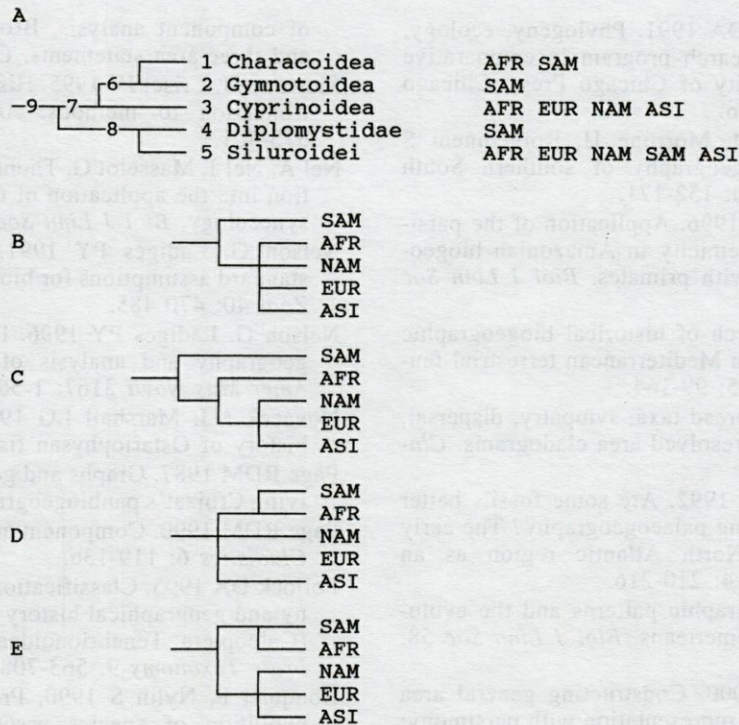
A

```
              ┌── 1 Characoidea      AFR SAM
         ┌─6──┤── 2 Gymnotoidea       SAM
    ─9─┬─7    └── 3 Cyprinoidea       AFR EUR NAM ASI
       └─8──┬── 4 Diplomystidae       SAM
            └── 5 Siluroidei          AFR EUR NAM SAM ASI
```

B
```
        ┌──── SAM
        │──── AFR
        │──── NAM
        │──── EUR
        └──── ASI
```

C
```
        ┌──── SAM
        │──── AFR
        │──── NAM
        │──── EUR
        └──── ASI
```

D
```
        ┌──── SAM
        │──── AFR
        │──── NAM
        │──── EUR
        └──── ASI
```

E
```
        ┌──── SAM
        │──── AFR
        │──── NAM
        │──── EUR
        └──── ASI
```

Fig. 4. – The case of the ostariophysan fishes. A, TAC for the group, without the basal lineage (Gonorhynchiformes). From Novacek & Marshall (1976), also reconstructable from fig. 7.30 and table 7.16 in Brooks & McLennan (1991), AFR = Africa, ASI = Asia, EUR = Europe, NAM= North America, SAM = South America. B-C, RACs of length 11 found by Brooks & McLennan (1991). D-E, RACs of length 10 found here.

Brooks & McLennan (1991: fig. 7.31) showed taxon/ancestor optimizations on their two RACs. Although they are not the shortest RACs it is interesting to note that on fig. 7.31b (= Fig. 4E here), the branch connecting N America, Europe and Asia is supported only by two reversals which seems not to bother the authors.

Van Welzen (1989: 121-124) analyzed relationships between 11 West Malesian areas, based on his own cladogram for the sapindacean shrub/tree genus *Guioa* Cav., using BPA and component compatibility analysis (Zandee & Roos 1987). The resulting RAC, which is the same for the two methods, has two branches exclusively supported by reversals (secondary absences). Curiously, van Welzen did not comment on this although branches supported by absences (which do not represent reversals due to the rooting method employed, *cf.* above) in some of van Welzen's other BPA-RACs were used to criticize BPA as a method.

other problems with BPA, reversal-supported branches also causes problems for PAE, and users of the two methods should consider this difficulty.

To add to the confusion, it should be mentioned that Fortey & Mellish (1992), in order to avoid some types of "impossible" optimizations, suggested to code PAE characters as Dollo parsimonious, i.e. to explain *all* homoplasy by reversals. Their rationale for this procedure was to avoid "re-evolving" taxa on their RAC. Of course, Dollo coding will increase the problem of reversals, just as irreversible coding will increase the problem of independent appearances of taxa or ancestors.

CONCLUDING REMARKS

To the problems which have previously been recognized with BPA, that of reversal-supported branches should be added. Unlike some of the

REFERENCES

Andersson L 1996. An ontological dilemma: epistemology and methodology of historical biogeography. *J Biogeography* 23: 269-277.

Brooks DR 1990. Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Syst Zool* 39: 14-30.

Brooks DR, McLennan DA 1991. Phylogeny, ecology, and behaviour. A research program in comparative biology. The University of Chicago Press, Chicago and London. xii+434 p.

Crisci JV, Cigliano MM, Morrone JJ, Roig-Juñent S 1991. Historical biogeography of southern South America. *Syst Zool* 40: 152-171.

da Silva JMC, Oren DC 1996. Application of the parsimony analysis of endemicity in Amazonian biogeography: an example with primates. *Biol J Linn Soc* 59: 427-437.

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

Fortey RA, Mellish CJT 1992. Are some fossils better than others for inferring palaeogeography? The early Ordovician of the North Atlantic region as an example. *Terra Nova* 4: 210-216.

Härlin M 1996. Biogeographic patterns and the evolution of eureptantic nemerteans. *Biol J Linn Soc* 58: 325-342.

Hugot JP, Cosson JF 2000. Constructing general area cladograms by matrix representation with parsimony: a western palearctic example. *Belg J Ent* 2: 77-86.

Humphries CJ, Parenti LR 1999. Cladistic biogeography. Second edition. Interpreting patterns of plant and animal distributions. Oxford University Press, xii+187 p.

Kluge AG 1988. Parsimony in vicariance biogeography: a quantitative method and a Greater Antillean example. *Syst Zool* 37: 315-328.

Legendre P 1986. Reconstructing biogeographic history using phylogenetic tree analysis of community structure. *Syst Zool* 35: 68-80.

Mayden RL 1988. Vicariance biogeography, parsimony and evolution in North American freshwater fishes. *Syst Zool* 37: 329-355.

Morrone JJ 1994a. Distributional patterns of species of Rhytirrhini (Coleoptera: Curculionidae) and the historical relationships of the Andean provinces. *Global Ecol Biogeogr Letters* 4: 188-194.

Morrone JJ 1994b. On the identification of areas of endemism. *Syst Biol* 43: 438-441.

Morrone JJ, Carpenter JM 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 JJ, Crisci JV 1995. Historical biogeography: introduction to methods. *Ann Rev Ecol Syst* 26: 373-401.

Nel A, Nel J, Masselot G, Thomas A 1998. An investigation into the application of the Wagner parsimony in synecology. *Biol J Linn Soc* 65: 165-189.

Nelson G, Ladiges PY 1991. Three-area statements: standard assumptions for biogeographic analysis. *Syst Zool* 40: 470-485.

Nelson G, Ladiges PY 1996. Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *Amer Mus Novit* 3167: 1-58.

Novacek MJ, Marshall LG 1976. Early biogeographic history of Ostariophysan fishes. *Copeia* 1976: 1-12.

Page RDM 1987. Graphs and generalized tracks: quantifying Croizat's panbiogeography. *Syst Zool* 36: 1-17.

Page RDM 1990. Component analysis: a valiant failure? *Cladistics* 6: 119-136.

Pollock DA 1995. Classification, reconstructed phylogeny and geographical history of genera of Pilipalpinae (Coleoptera: Tenebrionoidea: Pyrochroidae). *Invertebrate Taxonomy* 9: 563-708.

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

Rosen BR 1988. From fossils to earth history: applied historical biogeography. Analytical biogeography. An integrated approach to the study of animal and plant distributions. Ed by AA Myers & PS Giller: 437-481. Chapman & Hall, London & New York.

Rosen BR 1992. Empiricism and the biogeographical black box: concepts and methods in marine palaeobiogeography. *Palaeogeogr Palaeoclimatol Palaeoecol* 92: 171-205.

Swofford D 1993. PAUP 3.1.1. Illinois Natural History Syrvey, Champaign.

van Soest RWM, Hajdu E 1997. Marine area relationships from twenty sponge phylogenies. A comparison of methods and coding strategies. *Cladistics* 13: 1-20.

van Welzen PC 1989. *Guioa* Cav. (Sapindaceae): taxonomy, phylogeny, and historical biogeography. *Leiden Bot Ser* 12: 1-315.

Wiley EO 1988. Parsimony analysis and vicariance biogeography. *Syst Zool* 37: 271-290.

Zandee M, Roos MC 1987. Component-compatibility in historical biogeography. *Cladistics* 3: 305-332.