



ELSEVIER

Palaeogeography, Palaeoclimatology, Palaeoecology 2726 (2002) 1–14

PALAEO

www.elsevier.com/locate/palaeo

Phylogenetic biogeography with and without the fossil record: gauging the effects of extinction and paleontological incompleteness

Bruce S. Lieberman *

Departments of Geology and Ecology and Evolutionary Biology, 1475 Jayhawk Blvd., Lindley Hall, Room 120, University of Kansas, Lawrence, KS 66045, USA

Received 6 July 2000; accepted 29 August 2001

Abstract

Biogeography played an important role in early developments in evolutionary theory and continues to play an important role in evolutionary studies and paleogeographic reconstructions. The development of a phylogenetic approach to biogeographic analysis has been important; however, fossil taxa have not always played a role in phylogenetic biogeographic studies and their role has been criticized by some phylogenetic biogeographers. Here, simulation studies are used to show that phylogenetic biogeographic studies on extant organisms that do not include fossil taxa can often be artificially incongruent and inaccurate. This is because area cladograms for extant taxa alone may differ from those that also include extinct taxa, implying different patterns of biogeographic relationship between areas, and area cladograms are the fundamental data of phylogenetic biogeographic analysis. This finding is analogous to what is known about how including fossil taxa in phylogenetic analyses along with extant taxa can improve resolution and accuracy.

The incompleteness of the fossil record can also lead to artificial incongruence in phylogenetic paleobiogeographic studies because not all taxa that have ever lived are preserved in the fossil record, and thus those missing taxa cannot be incorporated into area cladograms. Therefore, area cladograms based on all preserved fossil species may also differ fundamentally from the true area cladograms which would be available if all taxa were preserved in the fossil record. Again, this can lead to artificial incongruence and inaccurate results. Simulations are used to show that under average preservational regimes for skeletonized marine invertebrates in the fossil record, and with average extinction probabilities, phylogenetic biogeographic studies of the extant biota are typically more likely to be artificially incongruent than those relying on fossil biotas. Further, in phylogenetic and biogeographic studies extant groups that have persisted for a very long time, have high extinction rates, and low diversity should be avoided. Phylogenetic biogeographic studies of fossil taxa should avoid groups with low diversity and a poor fossil record; these studies should also avoid time periods or regions with a poor fossil record. © 2002 Published by Elsevier Science B.V.

Keywords: phylogenetic analysis; simulations; biogeography; paleontology

* Corresponding author. Tel.: +1-785-864-2741; Fax: +1-785-838-9664.

E-mail address: blieber@ku.edu (B.S. Lieberman).

1. Introduction

A goal of biogeography is to reconstruct how

episodes of global change, be they climatic or tectonic, influence the evolution of biotas. The discipline has a long and rich intellectual tradition in western science, stretching back into the 18th century (Browne, 1983; Hallam, 1983a, 1994; Bowler, 1996; Brown and Lomolino, 1998; Lieberman, 2000). Associated with the growth of a research program in macroevolution and phylogenetic analysis (e.g. Eldredge and Cracraft, 1980; Vrba, 1980; Brooks and McLennan, 1991) there has been a concomitant development of biogeographic approaches that utilize phylogenetic methods to study biogeographic patterns (e.g. Brooks, 1985; Wiley, 1988a; Brooks and McLennan 1991; Michaux, 1991, 1996; Wiley et al., 1991; De Boer and Duffels, 1996). These approaches will be referred to as phylogenetic biogeography.

Palaeontologists have carried out numerous important biogeographic studies (e.g. Schuchert, 1910; Hallam, 1967, 1977, 1983b; Rowell et al., 1973; Jell, 1974; McKenna, 1975, 1983; Fortey and Cocks, 1992; Babcock, 1994; and the collected papers in McKerrow and Scotese, 1990). Thus, it is paradoxical that some biogeographers have argued that paleontology can at best be only the handmaiden of phylogenetic biogeographic studies in modern organisms (e.g. Croizat et al., 1974; Nelson and Platnick, 1981; Patterson, 1983). Unfortunately, most phylogenetic biogeographic studies concentrate on extant organisms. This should thus give us cause for concern for two reasons, the first philosophical, the second methodological. First, the fossil record is our one true chronicle of the history of life. As biogeography relates evolution to earth history change, one would predict that the fossil record should be the repository of a host of potentially interesting biogeographic patterns. The second, methodological reason will be the focus of the remainder of this paper; it involves extinction and its role in producing a biased, pruned, and, in a critical sense, biologically incomplete sample of all the organisms that have ever lived, in the same sense that the fossil record is an incomplete sample of the organisms alive at any one time in the distant past. Here, simulation studies are used to consider the potential significance of paleontology to phy-

logenetic biogeography. The role of extinct taxa and the incompleteness of the fossil record is studied to see how these affect our ability to retrieve biogeographic patterns in studies of the extant and fossil biotas, respectively. It is argued that possible fossil taxa should be included in phylogenetic biogeographic studies because including fossil taxa can increase the accuracy of phylogenetic biogeographic studies. For some groups this may not be possible, but for those groups with a good fossil record this would seem worthwhile.

2. Palaeontology and phylogenetic analysis

Cuvier's documentation of extinction as a pervasive phenomenon was paleontology's first great theoretical contribution to biology (Rudwick, 1976). Much more recently, Patterson (1981) and Gardiner (1982) asserted that fossil taxa did not contribute in an important way to the study of phylogenetic relationships. In short, they denied that the extinction of taxa affects our ability to retrieve phylogenetic patterns among extant organisms. This essentially obviated paleontology's relevance to phylogenetic studies because this assertion, if true, means fossil organisms need not be considered in phylogenetic studies of extant groups. However, Gauthier et al. (1988) and Donoghue et al. (1989) demonstrated that this was not true. They convincingly argued that the phenomenon of extinction is something that critically impacts phylogenetic analyses of extant organisms because it determines a priori which taxa are excluded from study by neontologists. Including more taxa in a phylogenetic study, especially fossil taxa which may have novel character combinations, can produce a phylogenetic tree topology closer to the underlying evolutionary pattern (Gauthier et al., 1988; Donoghue et al., 1989; Huelsenbeck, 1991; Novacek, 1992). Various authors, including Benton et al. (1999), have also commented on the relationship between phylogenetic patterns and the distribution of species in the fossil record.

Phylogenetic analysis and phylogenetic biogeography are two related fields (Brooks and

McLennan, 1991), and thus there are analogies between several aspects of phylogenetic analysis and phylogenetic biogeography. These include how incorporating fossil taxa influences our ability to resolve patterns. However, analogy between the fields of phylogenetic analysis and phylogenetic biogeography should not be mistaken for identity. Just as these two fields differ in important ways (see Brooks and McLennan, 1991; Lieberman, 2000), the affects of including or excluding fossil taxa in phylogenetic analysis and biogeography differ. The former topic has already been considered by Gauthier et al. (1988), Donoghue et al. (1989), Huelsenbeck (1991), Novacek (1992), and others. The latter topic is the focus of this paper. The different roles that fossil taxa play in the two fields is partly determined by the different aims of phylogenetic analysis and phylogenetic biogeography. The aim of phylogenetic analysis is to determine the evolutionary relationships of groups of taxa. The aim of phylogenetic biogeography is to determine if different taxa show congruence in their evolutionary relationships in conjunction with their geographic distributions (Brooks, 1985; Wiley, 1988a,b; Brooks and McLennan, 1991; Lieberman, 2000). Excluding fossil taxa from phylogenetic analysis generally decreases our ability to retrieve the underlying pattern of evolutionary relationship. Excluding fossil taxa in phylogenetic biogeography generally decreases our ability to retrieve biogeographic congruence, as shall be described more fully below.

3. Extinction, paleontological incompleteness, and phylogenetic biogeography

3.1. Extinction and phylogenetic biogeography

A detailed discussion of phylogenetic approaches to biogeographic analysis is not included here for the purposes of brevity and clarity, and the interested reader is referred to Brooks (1985), Wiley (1988a, 1988b), Brooks and McLennan (1991), Lieberman and Eldredge (1996), and Lieberman (2000); however, at base all phylogenetic biogeographic methods use information about

how groups of organisms evolve (i.e. phylogenetic patterns) as well as where these groups occur, to determine whether several different groups of organisms show a similar pattern of evolution across geographic space (Fig. 1).

Since phylogenetic biogeographic methods rely on phylogenies of organisms, and recognizing the problem identified by Gauthier et al. (1988) and Donoghue et al. (1989) that excluding or including taxa can dramatically change tree topology, we should expect this to feed back on biogeographic analyses. Biogeographic studies of extant organisms must by necessity exclude extinct taxa, and thus risk retrieving an inaccurate pattern. In addition, the set of excluded or extinct taxa have geographic range data that would also play a role in reconstructing biogeographic patterns, but which would not be considered in a study of extant organisms alone. This problem confronting phylogenetic biogeographers who study extant organisms can be thought of partially as a sampling problem: extinct taxa will not be sampled by neontologists.

3.2. The fossil record, incompleteness, and phylogenetic biogeography

The problem that biogeographic studies concentrating on extant organisms face is essentially analogous to the problem paleontologists face in their studies of the fossil record: it is the problem of incompleteness. The nature of the fossil record,

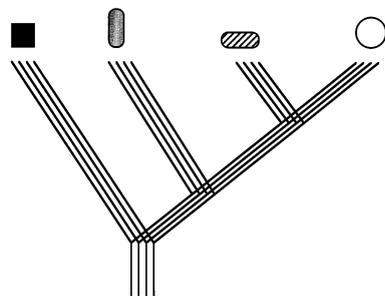


Fig. 1. Four overlain area cladograms with the geographic distribution of the terminal taxa represented by different shapes and substituted for the species names. These four clades show the same congruent pattern of evolution across geographic space, and have the same pattern of biogeographic relationship among areas.

and the phenomenon of paleontological incompleteness is well known, and again for the purposes of brevity and clarity it is not discussed in detail here. The interested reader should see Dings and Sadler (1982), Paul (1982, 1998), Marshall (1990), Benton and Storrs (1994) and the references cited therein for a detailed discussion. What matters for phylogenetic paleobiogeographic studies is that not every taxon that has ever lived is preserved in the fossil record. Thus, phylogenetic analyses with the biased set of taxa preserved may be inaccurate. Moreover, the geographic range data from these missing taxa also cannot be incorporated into biogeographic studies. In either case, this will influence the results of biogeographic analysis, and may lead to an erroneous result.

3.3. Comparing the relative performance of the fossil record and the extant biota in biogeographic studies

A potentially pertinent question becomes: does the fossil record or the extant biota perform better in phylogenetic biogeographic studies on average? That is, with the problem biogeographic studies of the extant biota face with the pruning of taxa through extinction, and the problem biogeographic studies of the fossil biota face with the pruning of taxa via paleontological incompleteness, which should we predict is generally more accurate?

4. Biogeographic congruence

There are a variety of ways of thinking about how extinction and incompleteness affect the results of any biogeographic study but one is to see how they influence what is referred to as biogeographic congruence. Biogeographic congruence is the agreement between different sets of data about a specific sequence of geological events and a concomitant series of evolutionary events. Biogeographic congruence can be conceived of in two ways. First, it can be thought of as compared to some idealized pattern. This is referred to as the ‘idealized case’. Imagine that there were a series of

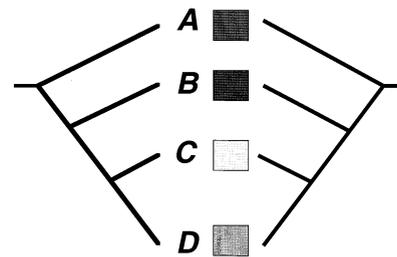


Fig. 2. Biogeographic congruence in the ‘idealized case’ (see text). The geological history of four hypothetical areas is shown as a branching diagram or cladogram on the right. Here the areas are the shaded squares. The closer two areas are on the cladogram, the more recently they shared a common geological history. The evolutionary history of four hypothetical taxa distributed in those areas is shown in the cladogram on the left. The evolution of this clade can be related to the geological history of the region.

geological changes that isolated a series of regions in a specific order; these can be expressed as a geological area cladogram (see De Boer and Duffels, 1996). These changes would be predicted to affect a clade in a specific way (Fig. 2). Biogeographic congruence can also be viewed as a similar pattern of phylogenetic differentiation across geographic space and thus a common biogeographic history and pattern shared between two (or more) clades (Fig. 3). This will be referred to as the ‘two-clade case’ of congruence.

The pattern shown in Fig. 1 is a hypothetical case of complete congruence in four different groups of organisms, where the two most closely related species in each group occur in the circular area and the hatched elliptical area. Such congruence suggests that the different groups speciated in response to the same geological or climatic events. Unfortunately, the real world is typically much more complex than what is shown in Fig. 1, and different clades may differ in important respects in terms of their implicit biogeographic signal. For example, biogeographic congruence will fail to emerge naturally if different groups fail to speciate at roughly the same time or in response to the same geological event. In addition, the emergence of geographic barriers, which play an important role in mediating speciation (e.g. Mayr, 1942; Vrba, 1985, 1992, 1996), may affect one group and not another. Finally, dispersal of individual lineages across large distances can lead to biogeo-

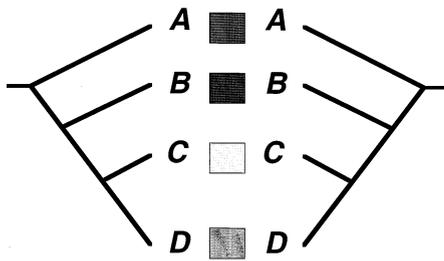


Fig. 3. An illustration of biogeographic congruence in the 'two-clade case'. Here two clades have congruent biogeographic patterns. They occur in a set of areas represented by the shaded squares and have differentiated across geographic space in concert.

graphic incongruence (Brooks and McLennan, 1991; Michaux, 1991). However, what biogeographers also need to be concerned with is when incongruence emerges as an artifact in any biogeographic study of the extant biota or the fossil record, due to extinction and paleontological incompleteness, respectively. There are a variety of analytical methods in phylogenetic biogeography, but all start out with area cladograms: phylogenies that have the geographic distribution of taxa substituted for the taxon names. Also, all analytical methods aim to retrieve congruence out of whatever noise and incongruence are implicit in the data due to disagreement among area cladograms or between an area cladogram and a known geological history. Finally, all analytical methods treat the absence of congruence among different clades as the absence of coherent underlying biogeographic patterns.

With these views of biogeographic congruence developed thus far, we can think about the effects of extinction and paleontological incompleteness on biogeographic congruence. First, in the 'idealized case' of congruence, if a biologist only studied a clade with living species, and one of the species in that clade went extinct such that it could not be sampled by a biologist, a divergent biogeographic pattern, incongruent from the ideal one would be recovered, simply as an artifact of the biologist's inability to sample extinct species (Fig. 4). Here, biogeographic congruence is considered specifically with reference to Brooks Parsimony Analysis (BPA) (Brooks, 1985; Brooks and McLennan, 1991; Lieberman and Eldredge,

1996). For example, the resulting area cladogram produced from BPA coding of the area cladogram on the right is $(D(A(B(C))))$, assuming the existence of area D is known, whereas the resultant area cladogram produced from BPA coding of the area cladogram on the left is $(A(B(C(D))))$. Other methods like components analysis were not considered here in the evaluation of incongruence because depending on the different assumptions used (see discussion in Enghoff, 1995) these area cladograms may or may not be incongruent.

A similar problem with incongruence might affect a paleontologist studying biogeographic patterns in the fossil record. Species in a clade might have shown a biogeographic pattern that was congruent with some idealized geological history, but due to the vagaries of the fossil record, one or more species in that clade were not preserved (Fig. 4).

In the 'two-clade case' of biogeographic congruence, the same principles apply. Two area cla-

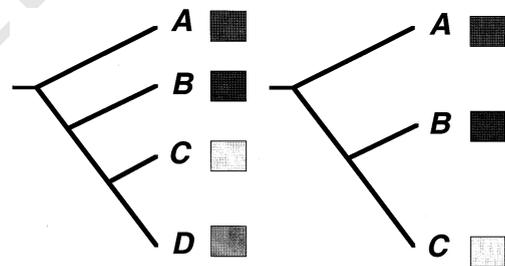


Fig. 4. A case of artificial incongruence in the 'idealized case'. Shown on the left is a clade of four hypothetical taxa A–D distributed in four different areas represented by the shaded squares. In this case, species C from the lightly shaded region is the sister group of species D from the intermediately shaded region, implying potentially that these areas share a recent common history. On the right is the same hypothetical clade but species D from the intermediately shaded area has been pruned. If this were an extant clade, this species might have gone extinct and could not be studied by a biologist that did not or could not sample the fossil record. If this were a fossil clade, species D might not have been preserved due to paleontological incompleteness. In either case, the pattern of area relationship implied by the area cladogram on the right is not entirely congruent, in a biogeographic framework based on Brooks Parsimony Analysis (BPA), with the area cladogram on the left. Area B appears to be the sister area of area C on the area cladogram on the right, but actually they really do not share a sister area relationship.

dograms might have an underlying pattern of biogeographic congruence, but extinction of taxa or the failure of certain taxa to be preserved will result in artificial biogeographic incongruence (Fig. 5). For example, when BPA is performed on each of the area cladograms in Fig. 5, the result for the area cladogram on the right is (B(A(C(D)))) whereas the result for the area cladogram on the left is (C(A(B(D)))).

Biogeographic incongruence needs not imply by itself a completely irretrievable biogeographic pattern, and therefore the artificial introduction of incongruence into biogeographic analysis by extinction or an incomplete fossil record does not obviate the utility of phylogenetic biogeography. In fact, various phylogenetic biogeographic methods are adept at retrieving meaningful biogeographic signal in the presence of biogeographic incongruence, as long as the incongruence is not so severe as to swamp the biogeographic signal. Still, in biogeographic analyses artificial biogeographic incongruence should be avoided.

5. Simulations: statistical tests

5.1. Methods

To test whether paleobiogeographic patterns in the fossil record or biogeographic patterns in the extant biota were more severely affected by artificial incongruence, a series of simulation studies were performed. There are several interesting biogeographic processes that lead to biogeographic incongruence, including dispersal, that can be modeled in a simulation study. However, here the primary focus is on extinction and incompleteness rather than dispersal because dispersal will have an equivalent effect on biogeographic incongruence in extant and fossil taxa.

Hypothetical pectinate clades of six taxa each were generated to determine whether a biogeographic study of modern organisms or of fossil organisms (of Pangaean age) could be most informative about the biogeographic events related to the breakup of Pangaea. These clades were then converted to area cladograms by substituting the geographic distribution of the taxon for its name.

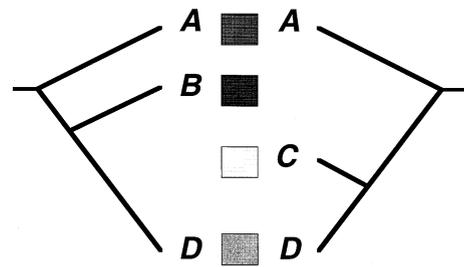


Fig. 5. A case of artificial incongruence in the 'two-clade case'. Imagine that two area cladograms were initially completely congruent. However, suppose that with the clade on the left the species from the lightly shaded area could not be sampled. If this were a neontological study, a biologist could not sample it because it went extinct. If it were a paleontological study, a paleontologist could not sample it because it was not preserved in the fossil record. By contrast, suppose similar phenomena affected the clade on the right but this time caused the species from the darkly shaded area not to be preserved. These area cladograms which are truly congruent no longer appear to be fully congruent, in a biogeographic framework based on BPA, because they imply different patterns of area relationship.

Six taxa were chosen as a compromise between the need for generating a clade of moderately complex cladistic structure while recognizing the difficulties of assessing congruence between clades of large size.

To evaluate artificial biogeographic incongruence in the extant biota a QuickBASIC Version 4.50 (1988) program (available from the author on request) was written to evolve a clade over 160 Myr at discrete 1 Myr steps (where 160 Ma was taken to approximate the breakup of the Atlantic margin of Pangaea following Hallam (1992, 1994) and Scotese (1997)). The clade was assumed to be in existence and biogeographically differentiated immediately after the initial continental breakup. Evolution was modeled to occur by a time homogeneous model of cladogenesis and extinction, and extinction probabilities were applied prior to speciation probabilities (though with adjustments for conditional probabilities, these could have been readily modeled as occurring simultaneously). The extinction probability used was 0.21 per Myr, which is the mean value found in a diverse set of clades. (The mean probability value was obtained using data from Stanley (1979), Walker and Valentine (1984), Vrba

(1987), Hulbert (1993), and Lieberman (1999). These studies were chosen because they used some phylogenetic information, along with a paleontological approach, to constrain probabilities of speciation and extinction.) Speciation probability was set to extinction probability plus 0.1 per Myr because if extinction probability is greater than or equal to speciation probability a clade will be ephemeral and thus is unlikely to be sampled. By contrast, if it is much less than speciation probability, the clade will grow inexorably, which may not be realistic; therefore, an upper limit on clade size of 25 taxa per individual branch was set. (These values for extinction and speciation probabilities are also in the range used in a variety of simulations studies involving phylogenies and paleontological data including those of Sepkoski et al. (1993) and Robeck et al. (2000).) The program evolved 100 clades and calculated the percentage of area cladograms that were artificially incongruent for both the ‘idealized case’ and the ‘two-clade case’ (in the ‘two-clade case’ two clades were evolved simultaneously). This procedure was then iterated 100 times.

Incongruence is tabulated in the ‘idealized case’ as the percentage of area cladograms incongruent relative to the idealized geological history, and in the ‘two-clade case’ as the percentage of area cladograms that are mutually incongruent. However, with sufficient elimination of taxa due to extinction or paleontological incompleteness, area cladograms are not properly thought of as incongruent but are rather best viewed as uninformative. For example, the two area cladograms in Fig. 6 are not incongruent, they merely consider the relationships of different areas. (It is conceivable, however, that if we consider each of these area cladograms to be pruned area cladograms that once included partially overlapping areas, they may actually have originally been biogeographically incongruent, though such a proposition is typically untestable.) Therefore, the percentage of incongruent area cladograms was always scaled relative to the total number of comparisons that were actually informative. In the simulations initial diversity values in each of the six taxa were also varied to equal one, five, and 10, treating

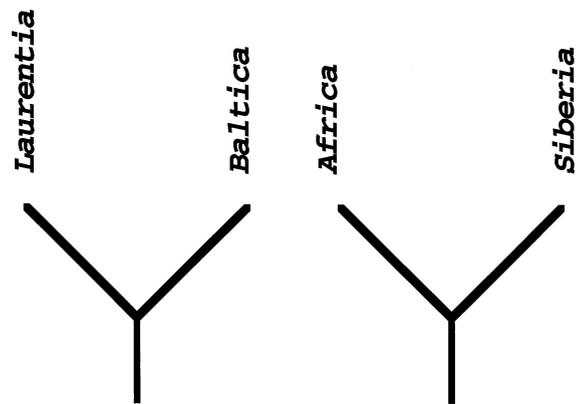


Fig. 6. These two area cladograms have been excessively pruned, due to either extensive extinction or a poor fossil record. They are neither biogeographically congruent or incongruent either. Rather, because they contain mutually exclusive areas the biogeographic patterns they each imply are best viewed as uninformative.

each of the six taxa essentially as higher taxa with several species in each area of endemism.

Then, simulations were used to look at how artificial biogeographic incongruence affects phylogenetic biogeographic studies done on fossil organisms. Again, a different QuickBASIC Version 4.50 (1988) program (available from the author on request) was used. Completeness estimates were chosen to approximate the average value of paleontological completeness for marine invertebrates with mineralized exoskeletons and a good fossil record and were taken from Foote and Raup (1996) and set at 0.75 per taxon. (Other fossil groups such as vertebrates and insects probably have lower completeness values; estimates for these groups have not yet been derived though, and thus they were not considered.) The completeness estimate was treated as a probabilistic proxy that a given taxon among each of the six would be preserved in the fossil record. This is the sense in which Foote and Raup (1996), p. 128, Solow and Smith (1997), and Fox et al. (1999), for example, treated it. The Foote and Raup (1996) study represents the most detailed study available that is broadly comparable to the type of analysis conducted here. Further, their conclusions on completeness were largely supported by Foote (1997), Paul (1998), and Fox et al. (1999).

Other detailed model-based approaches to completeness in the fossil record do exist, such as those of Holland and Patzkowsky (1999), but these did not produce explicit values of completeness in the sense needed here, and thus could not be used.

The simulations worked by applying the preservation probability given above to each of the taxa in hypothetical pectinate clades of six taxa. Taxa that were not ‘preserved’ were pruned from the area cladogram. 100 clades were considered, and the percentage of artificially incongruent area cladograms was calculated, both in the ‘idealized case’ and the ‘two-clade case’ (as described above, in the ‘two-clade case’ two clades were treated simultaneously). This procedure was iterated 100 times. Again, initial diversity values in each of the six taxa were also varied, with values of one, five, and 10 species per taxon used as described above.

Then, for both the ‘idealized case’ and the ‘two-clade case’ statistical comparisons were used to determine if artificial biogeographic incongruence differed between the fossil record and the extant biota. This involved six comparisons using the parametric *t*-test and the non-parametric Mann–Whitney *U*-test, which were implemented using Minitab 10.5Xtra (Minitab, 1995). These tests were then corrected for the potential of Type I error associated with multiple comparisons using the Bonferroni correction. This correction requires that for a hypothesis involving six comparisons to be globally significant at the 0.05 level, each comparison must be individually significant at the $1 - \sqrt[6]{0.05} \approx 0.01$ level.

5.2. Results

In every case, simulated phylogenetic biogeographic studies that use the fossil record have much less artificial biogeographic incongruence than those that consider just the extant biota (Table 1). Significant differences emerge at the 0.0001 level, meaning the fossil record of frequently preserved marine invertebrates exceeds the extant biota for biogeographic accuracy when standard extinction rates and standard preservation probabilities are assumed, and the simulation approach described herein is implemented.

Table 1

Test of H_0 that the degree of artificial biogeographic incongruence did not differ between phylogenetic biogeographic studies of the extant biota and the fossil record

Number of taxa	<i>P</i> -value
Idealized case	
1	< 0.0001
5	< 0.0001
10	< 0.0001
Two-clade case	
1	< 0.0001
5	< 0.0001
10	< 0.0001

Pectinate clades of six taxa were considered in both the ‘idealized case’ and the ‘two-clade case’ of congruence. Number of taxa refers to number of species placed on each of the six branches of the cladogram. Individual significance values are for the *t*-test and the Mann–Whitney *U*-test for all cases considered under the *P*-value column. If $P < 0.01$ the null hypothesis was rejected. This differs from the standard 0.05 level because of the correction for multiple comparisons (see text). In every case the null hypothesis was rejected because artificial incongruence was much lower in the fossil record.

5.3. Discussion

These results can be taken as a cautionary tale. Potentially, biogeographic studies of the extant biota may be more predisposed to artificial incongruence than biogeographic studies that consider the most frequently preserved elements of the fossil biota. This suggests that whenever possible biogeographers should include extinct taxa in their phylogenies, and further, that it might be worthwhile for paleontologists and neontologists to link up their biogeographic studies. This conclusion is analogous to the conclusions made by the studies of Gauthier et al. (1988) and Donoghue et al. (1989) in the sense that they all have shown how important it is to incorporate extinct taxa into evolutionary studies.

6. Simulations: general patterns

6.1. Methods

After these statistical tests, the general performance of the fossil record versus the extant biota was assessed with regards to predilection towards

artificial biogeographic incongruence. Specifically, different conditions that might lead to artificial biogeographic incongruence were varied. The analyses were performed to point out broad patterns rather than test specific hypotheses. Simulations such as these describe a hypothetical scenario where one condition is varied while others remain fixed. The advantage of this approach is that it allows investigators to isolate causality. Special situations like these do not generally occur in the real world. However, the simulations can at least give biogeographers an idea of how varying certain key parameters that pertain to clades can affect artificial biogeographic incongruence. This may aid biogeographers in choosing those clades most or least ideally suited to biogeographic analysis.

The same approach to simulations described above was used. This time though, simulations focused on four questions: (1) in general how does changing the number of taxa in any one branch of a six-taxon area cladogram change the degree of artificial biogeographic incongruence in both neontological and paleontological studies?; (2) how is artificial biogeographic incongruence affected in neontological studies as extinction probabilities and speciation probabilities rise and fall?; (3) how in neontological studies does clade duration influence artificial incongruence?; and finally, (4) how is artificial incongruence affected in paleontological studies when the completeness of the fossil record varies? Again, distinct QuickBASIC Version 4.50 (1988) subroutines were created to perform each of these simulations (available from the author on request). The percentage of biogeographically incongruent area cladograms for both the 'idealized case' and the 'two-clade case' were calculated as described above.

6.2. Results

The results are grouped into answers to the questions given in the preceding methods section. (1) Artificial incongruence declines as the number of taxa within each of the six higher taxa increases (Fig. 7). (2) Artificial biogeographic incongruence increases in neontological studies as extinction

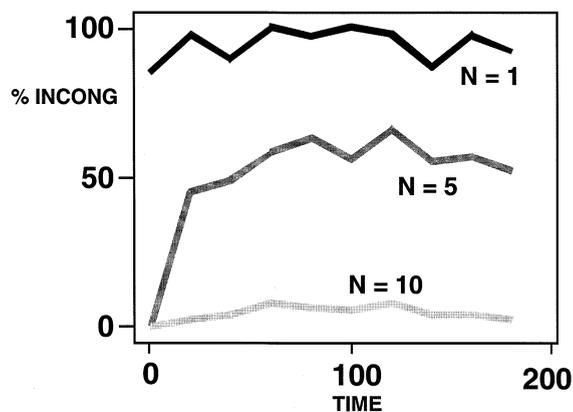


Fig. 7. A simulation analysis that quantified the effect that changing the number of taxa within any given branch of the six taxon area cladogram had on artificial biogeographic incongruence. An extant clade was simulated, so clade duration (TIME) is graphed on the *x*-axis, but identical results are retrieved with simulated fossil clades. The percentage of artificially incongruent phylogenies is graphed on the *y*-axis. Clades with one, five, and 10 taxa per each of the six branches on the original six branched area cladograms were considered. A different line shading is shown for each taxon value. Artificial incongruence declines as the number of taxa increases. The particular simulation shown used pectinate clades and tested for incongruence in the 'idealized case', with $\mu=0.41$ and $\lambda=0.51$.

probability increases (Fig. 8). (3) Artificial biogeographic incongruence also increases in neontological studies as the duration of a clade increases (Figs. 7, 8). (4) Finally, in paleontological studies biogeographic incongruence increases as paleontological incompleteness increases (Fig. 9). These results are summarized in Table 2. No consistent, significant differences in biogeographic incongruence were found between simulations that evaluated both symmetrical and pectinate clades and the 'idealized case' and 'two-clade case' over a variety of parameters. Thus, these results are not shown for the purposes of brevity and clarity.

6.3. Discussion

The results described above point out certain conditions when biogeographic analyses are most likely to be burdened with excess artificial incongruence. For example, the first result suggests that diverse clades may be better subjects

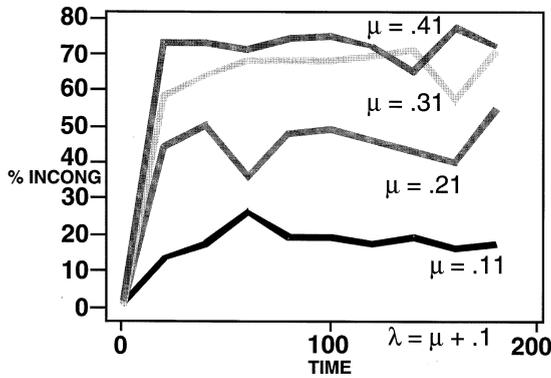


Fig. 8. A simulation that quantified the effect that increasing extinction rates had on artificial biogeographic incongruence. Clade duration is plotted on the x-axis and again the percentage of artificially incongruent phylogenies is graphed on the y-axis. Results for different extinction probabilities, represented by the symbol μ , are graphed, and a different line shading is used for each μ value. Speciation probability, λ , was set to $\mu+0.1$ as described in the text. The percentage of artificially incongruent area cladograms increases as the extinction probability increases and also as the age of the clade increases. The particular simulation shown used symmetrical clades, tested for incongruence in the ‘two-clade case’, and each branch of the six taxon area cladogram had five taxa.

for biogeographic analysis than depauperate ones, both for neontologists and paleontologists. This is not surprising, because as long as at least one representative in any given area of endemism is still present, artificial biogeographic incongruence will not occur, and the more taxa in any given area of endemism, the less likely they will have all gone extinct or all not have been preserved

Table 2
Summary of predictions about artificial biogeographic incongruence based on general simulations for phylogenetic biogeographic studies involving the fossil record and the extant biota

Fossil record	Artificial incongruence
Diversity increases	declines
Quality of fossil record decreases	increases
Extant biota	Artificial incongruence
Diversity increases	declines
Extinction probability increases	increases
Clade duration increases	increases

Factor varied and the effect on artificial incongruence are shown.

in the fossil record. The second result is in line with a theme developed throughout this paper: extinction can lead to artificial biogeographic incongruence in biogeographic studies of the extant biota. Thus, as extinction probabilities climb, artificial incongruence will increase, even in the face of climbing speciation probabilities. This is because clades with higher extinction probabilities are more volatile and thus their members are more likely to go extinct (Stanley, 1990; Gilinsky, 1994). Therefore, extant clades with high extinction probabilities should be avoided in biogeographic studies. However, it is important to note that some fossil clades with high extinction and origination probabilities such as ammonites (e.g. Hallam, 1977, 1983b) and trilobites (e.g. Jell, 1974; Babcock, 1994; Lieberman and Eldredge, 1996; Lieberman, 1997) do make excellent candidates for paleobiogeographic studies. This derives from the fact that the extinct taxa can still be sampled in the paleobiogeographic studies. Wiley and Mayden (1985) and Brooks and McLennan (1991) also have suggested that clades that have accumulated lots of extinction should be avoided

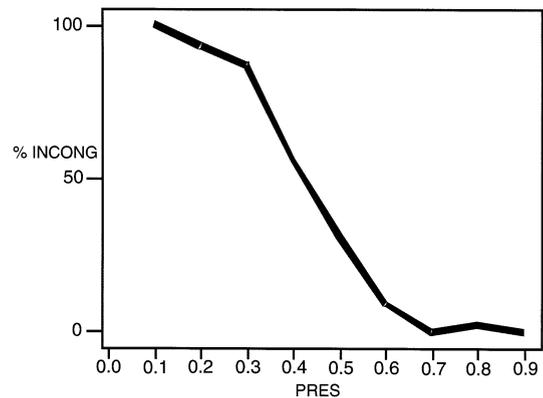


Fig. 9. A simulation that quantified the effect that improving the quality of the fossil record, in this case the preservation probability for a particular taxon, has on artificial biogeographic incongruence. Preservation probability per taxon (PRES) is graphed on the x-axis and again, the percentage of artificially incongruent phylogenies is graphed on the y-axis. The percentage of artificially incongruent area cladograms declines at a roughly constant rate as the quality of the fossil record improves. The particular simulation shown used pectinate clades, tested for incongruence in the ‘two-clade case’, and each branch of the six taxon area cladogram had five taxa.

in biogeographic studies of the extant biota. One reason is that artificial biogeographic incongruence may ensue. The conclusions reached in the third result derive logically from those reached in the second result: a clade that has been around for a long time will have accrued more artificial biogeographic incongruence because there are more opportunities for extinction to occur. Again, Wiley and Mayden's (1985) and Brooks and McLennan's (1991) suggestion that the best subjects for biogeographic analysis are those clades that have accrued minimal extinction applies here. Finally, in the fourth result, the simulations suggest that time periods, organisms, or regions with poor fossil records make poor candidates for biogeographic analysis. This is because as paleontological incompleteness climbs, taxa are more likely to be excluded from an analysis due to taphonomic artifact. When this happens, artificial biogeographic incongruence is likely to emerge.

Thus, to summarize, biogeographic studies by neontologists of extant taxa should avoid species poor groups, groups with high extinction rates, and groups that have persisted for a very long time. For example, lungfish and horseshoe crabs fit at least two of these categories, and thus probably make poor candidates for phylogenetic biogeographic studies. (Note, as discussed by Stanley (1990), those groups with high extinction rates are unlikely to have long durations.) Biogeographic studies by paleontologists should avoid species poor groups and also those groups, regions, or time periods with a poor fossil record.

Using these precepts, certain concrete cutoff points may actually emerge. Assuming that biogeographers would like to avoid artificial incongruence at least 50 percent of the time, clades with extinction probabilities of 0.3 per Myr or higher should be avoided if they have persisted longer than roughly 10 Myr (see Fig. 8). Most extant taxa of moderate rank in the Linnean hierarchy are likely to have a pre-Neogene history, signifying potential problems with biogeographic incongruence in these taxa if they have high extinction probabilities. Palaeontological studies appear to suffer significantly if preservation probability falls below about 0.4. Since current studies (e.g. Foote and Raup, 1996; Foote, 1997; Paul, 1998) put

this value at about 0.75 for well skeletonized marine invertebrates, this indicates some optimism for biogeographic studies conducted on many, though not all kinds of fossil organisms. However, there are other problems that emerge with real world paleontological data which would be impossible to quantify or gauge in a study like this. This again involves paleontological incompleteness, which not only excludes a taxon from the fossil record, but which may also cause it to appear to have a geographic range which is a biased and inaccurate subset of its actual geographic range. Although these effects cannot be realistically quantified at this time, it may be possible to model them, and they do sound an additional, cautionary note for paleobiogeographic studies done on fossil organisms. Furthermore, for those groups with a poor fossil record, the accuracy of biogeographic studies performed solely on extant taxa will likely exceed the accuracy of biogeographic studies performed solely on fossil taxa.

7. Conclusions

There are a variety of processes that conspire to make the data available to natural historians incomplete. These precepts apply for studies of the extant biota as well as the fossil record. Although numerous biological studies, in a tradition that began with Darwin (1859), have criticized the incompleteness of the fossil record, we must recognize that the extant biota also suffers from incompleteness: a large percentage of all the species that have ever lived are extinct. (The precise percentage depends on the extent to which the exponential model of diversification, e.g. Benton (1997), describes the history of life.) It remains an open question whether the failure of the fossil record to preserve every taxon that has ever lived is a greater obstacle to paleontology than the failure of the extant biota to contain much of the total diversity of the history of life is an obstacle to neontology. In truth, the answer to that question depends on the scope of the study. The pruned extant biota is certainly complete enough to address many scientific questions, as is the fossil record. This view-

point is nicely encapsulated by Paul's (1998) approach to the concept of completeness in paleontological studies.

Here, based on the application of simulation studies, which must be treated with the same caveats that apply to results from any study that uses simulations, in phylogenetic biogeography there are going to be many questions that the study of the fossil record is equally or better equipped to address than the study of the extant biota. The statements by certain vicariance biogeographers such as Nelson and Platnick (1981) might suggest that this conclusion is counterintuitive. However, three sets of observations actually suggest that this statement makes perfect sense: first, the effectiveness of a variety of paleogeographic studies too numerous to adequately cite but including, for example, Hallam (1967, 1977, 1983b, 1994), Jell (1974), Rowell et al. (1972), McKenna (1975, 1983), McKerrow and Scotese (1990), Fortey and Cocks (1992), and Babcock (1994); second, the recognition by Brooks and McLennan (1991) that it is important to address the history of life when studying biogeographic patterns and the extent to which the earth and its biota have co-evolved; and finally, the statement by Wiley and Mayden (1985) and Brooks and McLennan (1991) that biogeographic studies of the extant biota must focus on clades that have not been excessively pruned by extinction.

The bane of any biogeographic study is biogeographic incongruence. Incongruence can emerge in a biogeographic database for many reasons, but it is certainly worthwhile to minimize biogeographic incongruence that arises in a database for artificial reasons. The pruned extant biota bears mute testament to a host of extinct taxa, and the simple removal of these taxa due to random extinction events may cause two (or more) clades to appear biogeographically incongruent when in fact they actually are not. The failure of some taxa to be preserved in the fossil record causes a similar type of incongruence. The results presented here suggest that based on known probabilities of extinction and known preservation probabilities for the most abundant representatives of the fossil record, the skeletonized marine invertebrates, the extant biota is more likely to

suffer from artificial incongruence than the fossil record.

A framework for thinking about phylogenetic biogeographic studies of clades in the context of key parameters that may be inherent to those clades, like speciation and extinction rates, duration, and probability of being preserved in the fossil record, was also presented. The situations which are apt to introduce excessive artificial incongruence into biogeographic studies should be avoided. Specifically, biogeographic studies of the extant biota should avoid focusing on clades that have persisted for a geologically long period of time, that have high extinction probabilities, or are species poor. Palaeobiogeographic studies should avoid taxa, geographic regions, or time periods with a poor fossil record, and they should also avoid low diversity clades. It will be worthwhile if biogeographic studies of the extant biota work backwards to incorporate extinct taxa, while paleobiogeographic studies interdigitate as much as possible with still extant organisms. This is a pragmatic approach and likely to be more illuminating to biogeographic patterns in general because a biogeographers' 'field of research extends over the whole earth, not only as it now exists, but also during the countless changes it has undergone from the earliest geological epochs' (Wallace, 1857, p. 159).

Acknowledgements

I thank James Cornette, Anthony Hallam, Roger Kaesler, Dick Robison, Alycia Rode, Bert Rowell, Andrew Smith, Ed Wiley, and an anonymous reviewer for comments on this paper, and Rick Prum and Chris Raxworthy for fruitful discussions. This research was supported by NSF EPS-9874732 and OPP-9909302.

References

- Babcock, L.E., 1994. Biogeography and biofacies patterns of Middle Cambrian polymeroid trilobites from North Greenland: palaeogeographic and palaeo-oceanographic implications. *Bull. Grøn. Geol. Unders.* 169, 129–147.

- Benton, M.J., 1997. Models for the diversification of life. *Trends Ecol. Evol.* 12, 490–495.
- Benton, M.J., Hitchin, R., Wills, M.A., 1999. Assessing congruence between cladistic and stratigraphic data. *Syst. Biol.* 483, 581–596.
- Benton, M.J., Storrs, G.W., 1994. Testing the quality of the fossil record: paleontological knowledge is improving. *Geology* 22, 111–114.
- Bowler, P.J., 1996. *Life's Splendid Drama*. University of Chicago Press, Chicago, IL, 525 pp.
- Brooks, D.R., 1985. Historical ecology: a new approach to studying the evolution of ecological associations. *Ann. Miss. Bot. Gard.* 72, 660–680.
- Brooks, D.R., McLennan, D.A., 1991. *Phylogeny, Ecology, and Behavior*. University of Chicago Press, Chicago, IL, 434 pp.
- Brown, J.H., Lomolino, M.V., 1998. *Biogeography*. Sinauer, Sunderland, MA, 691 pp.
- Browne, J., 1983. *The Secular Ark: Studies in the History of Biogeography*. Yale University Press, New Haven, CT, 273 pp.
- Croizat, L., Nelson, G., Rosen, D.E., 1974. Centers of origin and related concept. *Syst. Zool.* 23, 265–287.
- Darwin, C., 1859. *On the Origin of Species*. John Murray, London, 490 pp.
- De Boer, A.J., Duffels, J.P., 1996. Historical biogeography of the cicadas of Wallacea, New Guinea and the West Pacific: a geotectonic explanation. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 124, 153–177.
- Dingus, L., Sadler, P.M., 1982. The effects of stratigraphic completeness on estimates of evolutionary rates. *Syst. Zool.* 31, 400–412.
- Donoghue, M.J., Doyle, J., Gauthier, J., Kluge, A., Rowe, T., 1989. The importance of fossils in phylogeny reconstruction. *Annu. Rev. Ecol. Syst.* 20, 431–460.
- Eldredge, N., Cracraft, J., 1980. *Phylogenetic Patterns and the Evolutionary Process: Method and Theory in Comparative Biology*. Columbia University Press, New York, 349 pp.
- Enghoff, H., 1995. Historical biogeography of the Holarctic: area relationships, ancestral areas, and dispersal of non-marine animals. *Cladistics* 11, 223–263.
- Foote, M., 1997. Estimating taxonomic durations and preservation probability. *Paleobiology* 23, 278–300.
- Foote, M., Raup, D.M., 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 22, 121–140.
- Fortey, R.A., Cocks, L.R.M., 1992. The early Palaeozoic of the North Atlantic region as a test case for the use of fossils in continental reconstruction. *Tectonophysics* 206, 147–158.
- Fox, D.L., Fisher, D.C., Leighton, L.R., 1999. Reconstructing phylogeny with and without temporal data. *Science* 284, 1816–1819.
- Gardiner, B., 1982. Tetrapod classification. *Zool. J. Linn. Soc.* 74, 207–232.
- Gauthier, J., Kluge, A.G., Rowe, T., 1988. Amniote phylogeny and the importance of fossils. *Cladistics* 4, 105–209.
- Gilinsky, N.L., 1994. Volatility and the Phanerozoic decline of background extinction intensity. *Paleobiology* 20, 445–458.
- Hallam, A., 1967. The bearing of certain palaeogeographic data on continental drift. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 3, 201–224.
- Hallam, A., 1977. Jurassic bivalve biogeography. *Paleobiology* 3, 58–73.
- Hallam, A., 1983. *Great Geological Controversies*. Oxford University Press, New York, 182 pp.
- Hallam, A., 1983b. Early and mid-Jurassic molluscan biogeography and the establishment of the central Atlantic seaway. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 43, 181–193.
- Hallam, A., 1992. *Phanerozoic Sea-Level Changes*. Columbia University Press, New York, 266 pp.
- Hallam, A., 1994. *An outline of Phanerozoic biogeography*. Oxford Biogeography Series vol. 10. Oxford University Press, Oxford, 246 pp.
- Holland, S.M., Patzkowsky, M.E., 1999. Models for simulating the fossil record. *Geology* 27, 491–494.
- Huelsenbeck, J.P., 1991. When are fossils better than extant taxa in phylogenetic analysis? *Syst. Zool.* 40, 458–469.
- Hulbert, R.C., Jr., 1993. Taxonomic evolution in North American Neogene horses (subfamily Equinae): the rise and fall of an adaptive radiation. *Paleobiology* 19, 216–234.
- Jell, P.A., 1974. Faunal provinces and possible planetary reconstruction of the Middle Cambrian. *J. Geol.* 82, 319–350.
- Lieberman, B.S., 1999. Turnover Pulse in trilobites during the Acadian Orogeny. *Virg. Mus. Nat. Hist. Spec. Publ.* 7, 99–108.
- Lieberman, B.S., 2000. *Paleobiogeography: Using Fossils to Study Global Change, Plate Tectonics, and Evolution*. Kluwer Academic Press/Plenum Publishing, New York, 208 pp.
- Lieberman, B.S., Eldredge, N., 1996. Trilobite biogeography in the Middle Devonian: geological processes and analytical methods. *Paleobiology* 22, 66–79.
- Marshall, C.R., 1990. Confidence intervals on stratigraphic ranges. *Paleobiology* 16, 1–10.
- Mayr, E., 1942. *Systematics and the Origin of Species*. Dover Press, New York, 334 pp.
- McKenna, M.C., 1975. Fossil mammals and Early Eocene North Atlantic land continuity. *Ann. Miss. Bot. Gard.* 62, 335–353.
- McKenna, M.C., 1983. Holarctic landmass rearrangement, cosmic events, and Cenozoic terrestrial organisms. *Ann. Miss. Bot. Gard.* 70, 459–489.
- McKerrow, W.S., Scotese, C.R., 1990. *Palaeozoic Palaeogeography and Biogeography*. *Geol. Soc. Lond. Mem.* vol. 2.
- Michaux, B., 1991. Distributional patterns and tectonic development in Indonesia: Wallace reinterpreted. *Aust. Syst. Bot.* 4, 25–36.
- Michaux, B., 1996. The origin of southwest Sulawesi and other Indonesian terranes: a biological view. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 122, 167–183.
- Minitab, 1995. *Release 10.5Xtra*. Minitab Inc., State College, PA.
- Nelson, G., Platnick, N.I., 1981. *Systematics and Biogeography: Cladistics and Vicariance*. Columbia University Press, New York, 567 pp.

- Novacek, M.J., 1992. Fossils as critical data for phylogeny. In: Novacek, M.J., Wheeler, Q.D. (Eds.), *Extinction and Phylogeny*. Columbia University Press, New York, pp. 46–88.
- Patterson, C., 1981. The significance of fossils in determining evolutionary relationships. *Annu. Rev. Ecol. Syst.* 12, 195–223.
- Patterson, C., 1983. Aims and methods in biogeography. In: Sims, R.W., Price, J.H., Whalley, P.E.S. (Eds.), *Evolution, Time and Space: The Emergence of the Biosphere*. Academic Press, New York, pp. 1–28.
- Paul, C.R.C., 1982. The adequacy of the fossil record. In: Joysey, K.A., Friday, A.E. (Eds.), *Problems of Phylogenetic Reconstruction*. Systematics Association Special Volume Number 21. Academic Press, New York, pp. 75–117.
- Paul, C.R.C., 1998. Adequacy, completeness and the fossil record. In: Donovan, S.K., Paul, C.R.C. (Eds.), *The Adequacy of the Fossil Record*. John Wiley and Sons, New York, pp. 1–22.
- QuickBASIC Version 4.50, 1988. Microsoft, Seattle, WA.
- Robeck, H.E., Maley, C.C., Donoghue, M.J., 2000. Taxonomy and temporal diversity patterns. *Paleobiology* 26, 171–187.
- Rowell, A.J., McBride, D.J., Palmer, A.R., 1973. Quantitative study of Trempealeauian (latest Cambrian) trilobite distribution in North America. *Bull. Geol. Soc. Am.* 84, 3429–3442.
- Rudwick, M.J.S., 1976. *The Meaning of Fossils*. Science History Publications, New York, 287 pp.
- Schuchert, C., 1910. *Paleogeography of North America*. *Geol. Soc. Am. Bull.* 20, 427–606.
- Scotese, C.R., 1997. *Paleogeographic Atlas*. PALEOMAP Project, University of Texas at Arlington, Arlington, TX.
- Sepkoski, J.J., Jr., Kendrick, D.C., 1993. Numerical experiments with model monophyletic and paraphyletic taxa. *Paleobiology* 19, 168–184.
- Solow, A.R., Smith, W., 1997. On fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 23, 271–277.
- Stanley, S.M., 1979. *Macroevolution: Pattern and Process*. W.H. Freeman, San Francisco, CA, 332 pp.
- Stanley, S.M., 1990. The general correlation between rate of speciation and rate of extinction: fortuitous causal linkages. In: Ross, R.M., Allmon, W.D. (Eds.), *Causes of Evolution*. University of Chicago Press, Chicago, IL, pp. 103–127.
- Vrba, E.S., 1980. Evolution, species and fossils: how does life evolve? *S. Afr. J. Sci.* 76, 61–84.
- Vrba, E.S., 1985. Environment and evolution: alternative causes of the temporal distribution of evolutionary events. *S. Afr. J. Sci.* 81, 229–236.
- Vrba, E.S., 1987. Ecology in relation to speciation rates: some case histories of Miocene-Recent mammal clades. *Evol. Ecol.* 1, 283–300.
- Vrba, E.S., 1992. Mammals as a key to evolutionary theory. *J. Mammal.* 73, 1–28.
- Vrba, E.S., 1996. On the connection between paleoclimate and evolution. In: Vrba, E.S., Denton, G.H., Partridge, T.C., Burckle, L.H., (Eds.), *Paleoclimate and Evolution with Emphasis on Human Origins*. Yale University Press, New Haven, CT, pp. 24–45.
- Walker, T.D., Valentine, J.W., 1984. Equilibrium models of evolutionary species diversity and the number of empty niches. *Am. Nat.* 124, 887–899.
- Wallace, A.R., 1857. On the natural history of the Aru Islands. *Ann. Mag. Nat. Hist.* 2nd Ser. 20, 473–485.
- Wiley, E.O., 1988a. Parsimony analysis and vicariance biogeography. *Syst. Zool.* 37, 271–290.
- Wiley, E.O., 1988b. Vicariance biogeography. *Annu. Rev. Ecol. Syst.* 19, 513–542.
- Wiley, E.O., Siegel-Causey, D., Brooks, D.R., Funk, V.A., 1991. *The Compleat Cladist*. University of Kansas Museum of Natural History Special Publication vol. 19. University of Kansas Press, Lawrence, KS.
- Wiley, E.O., Mayden, R.L., 1985. Species and speciation in phylogenetic systematics, with examples from the North American fish fauna. *Ann. Miss. Bot. Gard.* 72, 596–635.