TREE-BASED APPROACHES TO UNDERSTANDING HISTORY: COMMENTS ON RANKS, RULES, AND THE QUALITY OF THE FOSSIL RECORD

MARK A. NORELL
Department of Vertebrate Paleontology,
American Museum of Natural History,
79th Street at Central Park West, New York, New York 10024

ABSTRACT. How do we understand the history of life? Traditionally patterns of preservation have been used as direct indicators of taxic and faunal origin and demise. However, the fossil record is rife with inconsistencies, and patterns of fossil preservation, ordered superpositionally, give at best a crude (and sometimes even misleading) reflection of phylogenetic relationship and faunal diversity. The application of phylogenetic systematics (cladistics) to patterns in the fossil record offers an alternative to the direct reading of life’s history in the rocks. Paleontologic information can be integrated within this system and interpreted in a logical and consistent fashion. The results of such an approach is a new taxic and faunal history, one that is predicted by phylogeny and is to varying degrees incongruous with a direct reading of the fossil record.

The claim of the phylogenetic system to elevation into the universal reference system of biology has a logical, even if not historical, foundation, and arises because few areas of research can be conceived which do not bear fruit and lead to more profound conclusions through a knowledge of the phylogenetic relationship of its objects, and which cannot, in turn, lead to the discovery of hitherto unknown relationships in the course of mutual exchange of information.

Willi Hennig (1965, p. 101)

INTRODUCTION

The occurrence of fossils in sedimentary rocks has been recognized for hundreds of years as evidence of the antiquity of life. During the early 1800’s, the birth of modern paleontology, systematics and geology nurtured a belief that the successional pattern of fossils represents life’s sequence of appearance on the Earth. Since this era, understanding the temporal history of organisms has traditionally been the domain of paleontologists studying preservational patterns.

Patterns of preservation have been considered an accurate indicator or approximation of taxic diversification (Bown and Rose, 1987; Valentine, 1969), extinction (Sepkoski, 1982; Raup and Sepkoski, 1984; Raup and Boyajian, 1988), and the temporal (Van Valen, 1985; Gingerich, 1977) and spatial (Simpson, 1965; Flynn, 1986) aspects of evolution. Superpositional ordering of fossil occurrences has also been assumed to be indicative of direct ancestor descendant relationships (Bown and Rose, 1987; Wei and Kennet, 1988, among others). Over the last two decades this research program has come under attack (see comments in Nelson and Platnick, 1981; Patterson, 1981; Smith and Patterson, 1988; Patterson and Smith, 1987; and earlier by McKenna and others, 1977), and the need for more logical methods addressing historical pattern is

In an earlier paper I proposed a method combining the fossil record with cladistic hypotheses of relationship (Norell, 1992). This method phylogenetically “corrects” observed stratigraphic ranges of individual taxa to conform with predictions of phylogeny, changing the focus of studies based on the temporal origin and duration of taxa. Instead of relying exclusively on the tabulations of stratigraphic ranges, a tree-based approach provides a synthesis of fossils and phylogeny by calibrating testable cladistic hypotheses with paleontologic information. This approach centers on phylogeny as the key to understanding the timing of taxic evolution. Fossils allow absolute time to be added to this framework by calibrating tree branches.

Such analyses influence the perception of the past by extending the histories of taxa. When times of origin of taxa are extended, temporal taxic diversities are inflated. Put simply, our phylogenies predict that there were more taxa present during specific intervals than are observed in the fossil record, and because they directly reflect the phylogenetic hierarchy, they also tell us what these taxa were.

A REVIEW OF THE METHOD

In 1953 the German entomologist Willi Hennig laid out the framework of modern systematics by defining the concept of phylogenetic relationship as “species B, is more closely related to species, C, than to another species, A, when B has at least one ancestral species source in common with species C which is not the ancestral source of species A” (Hennig, 1965, p. 97). In defining phylogenetic relationship as a relative concept, Hennig provided a structure to examine phylogenetic relationships that has implications regarding the ages of taxa.

This structure allows paleobiologists to go beyond the traditional practice of basing a group’s absolute age on the oldest representative of that group in the stratigraphic column (Simpson, 1961). Instead, Hennig suggested “When, however, it has been firmly established that a fossil belongs to a given monophyletic group, that fossil can then be of importance not only for determining the minimum age of the group to which it belongs, but also for determining the minimum age of related groups, of which no fossil finds are available.” These sentiments are best expressed in Hennig’s figure 4 (fig. 1), where he recognized three times in the origin of a taxon: t-1, the time of origin of a group, t-2, the first appearance of a group’s “typical” characters (quotes are Hennig’s), and
t-3, the age of division of a group. This figure forms the basis for a procedure (Norell, 1992) that corrects the stratigraphic ranges of individual taxa to conform with Hennig’s assertion that sister-taxa must have equal temporal durations (Paul, 1982).

Previously, I distinguished between two internested classes of a phylogenetic branch’s duration: groups and lineages (Norell, 1992). The less inclusive class is groups. Groups are the typical commodities of systematic analysis, and my usage is equivalent to Hennig’s t-2. That is, groups are phylogenetic branches that have acquired apomorphies (Hennig’s typical characters) and are therefore diagnostic (Rowe, 1987). The determination of the minimum age of a group, therefore, requires the discovery of apomorphies in a fossil taxon. These apomorphies indicate a
group's monophyly and therefore eliminate it from the possibility of being an ancestral taxon. Monophyletic taxa are the only biological units that exist in nature “independent of the taxonomists ability to recognize them” (Smith and Patterson 1988, p. 129). Hence, it is only monophyletic groups that depict evolutionary and historical processes (Schoch, 1986).

In my earlier paper I recognized a second class—the lineage. Lineages represent the entire history of a phylogenetic branch since it split from its sister taxon. The origin of a lineage conforms to the time of cladogenesis. Empirically this point corresponds with the oldest group in the lineage or the oldest member of its sister group, depending on which is more ancient. Lineages, therefore, have an antiquity equal to or older than their included groups. If lineages are older than groups the difference between the lineage age and group age (in Hennig’s sense t1-t2) is termed the ghost lineage. Ghost lineages are segments of evolutionary history that can be recovered not by the fossil record alone but through paleontologically calibrated phylogenetic analysis (fig. 2).

Ghost lineages are not limited to extending the history of groups observable in the fossil record or those alive today—the terminal taxonomic twigs on phylogenetic trees. They can also be observed within the deep structure of phylogenetic hypotheses. The procedure for uncovering ghost taxa is outlined in figure 2 and is no different than determining the duration of a ghost lineage, except ghost taxa are embedded within a

---

Fig. 2. A resolved phylogeny is shown in A, and fossil ranges for taxa A, B, C, and D are indicated by the solid lines in C. The phylogeny in A can be divided up into all of its separate component taxa. Taxa E and F are species that were ancestral to groups A-B and C-D respectively as predicted by the phylogeny in A. This tree can be calibrated using the fossil record in C. Therefore the minimum time of divergence of A and B is 4, and C and D is 10. The range of the lineage of taxon A has a corrected range of 4, 1 from the group plus 3 from the ghost lineage. Similarly the lineage origin for D is extended to 10, a group origin at 6 plus a ghost lineage duration of 4. Group A-B split from group C-D at 10. Because A or B do not appear until 4, a ghost taxon E is required. The minimum time of origin of this ghost taxon was at time 10, as calibrated by taxon C. Therefore the diagram in C can be used to indicate diversity. At time 9 the cladogram and the fossil record predict that a minimum of three taxa were present even though only one has been recovered in the fossil record.
tree's internal structure. As a point of clarification, in my original paper (Norell, 1992), my use of ghost lineages and ghost taxa was less than clear. Ghost lineages are the corrected extensions of groups. Ghost taxa are the kinds of ghost lineages that are embedded in the internal structure of trees, such as E in figure 2. These taxa become extinct by speciating into terminal taxa. Although operationally slightly different, theoretically they are the same thing; ghost taxa are just the ghost lineages of resolved groups that contain two or more taxa. For instance, if taxa A and B in figure 2 were considered a single group, the ghost taxon E would be the ghost lineage of taxon A B.

This procedure casts the temporal record in a new perspective: the temporal durations of taxa can be predicted by specific phylogenetic trees. The outcome is that temporal ranges are extended: for ghost lineages, it is the ranges of single monophyletic terminal taxa; for ghost taxa, it is the range of larger monophyletic assemblages. Temporal range extension has portentous consequences. For instance, temporal diversities may be strikingly modified (fig. 2; and Norell, 1992, fig. 3.10 and 3.11). These extensions produce an increase in numeric diversity, and often they severely modify temporal diversity patterns.

Alterations to traditional diversity patterns based on a direct reading of the fossil record require comment. From a phylogenetic perspective it would be foolish to demand an explanation for every peak and valley of numeric diversity through time. Corrected temporal ranges and temporal diversities based on them are only minimum estimates in a phylogenetic approach, estimates based on specific phylogenetic trees. Therefore the most constructive use of phylogenetic methods is in evaluating specific hypotheses of temporal change. As minimum estimates, actual diversities and patterns may be much different.

The implementation of the procedure outlined in the 1992 paper (Norell, 1992) and above to empirical data sets has just begun (Edgecombe, 1992). It is too early to tell what modifications need be made to make it more logically applicable to actual data sets and associated problems. Understanding how different kinds of data sets are modified by a phylogenetic correction, as well as the limits to such an approach, have yet to be investigated. As a prelude to this ongoing process a few points are developed below.

Multiple topologies.—Often more than a single parsimonious tree is encountered during phylogenetic analysis of a group. A common procedure to summarize congruent phylogenetic information is to summarize these trees into consensus trees (Adams, 1972; Wiley and others 1991). Consensus trees are not to be confused with phylogenies, however, because character information supporting individual groupings may not be common among the original fundamental cladograms summarized into the consensus tree (Miyamoto, 1985).

My 1992 paper used consensus trees in the correction of temporal ranges. In cases where nodes were incompletely resolved, all possible resolutions of that node were examined, and the temporal point of origin
of ghost lineages and taxa were calculated accordingly. This resulted in a range for the origin of individual ghost lineages, a range whose boundaries were specified by all possible dichotomous resolution solutions of the multichotomous consensus nodes. This is flawed because all possible dichotomous solutions are not required to specify a multichotomous branch point on a consensus tree. Consider the example in figure 3. Four taxa (A, B, C, and D) have the stratigraphic distribution shown. Cladistic analysis of these taxa results in two most parsimonious trees, one preferring a monophyletic grouping of A and B, the other A and C. In both cases A, B, and C are monophyletic relative to D (fig. 3A and C). The strict consensus tree (Fitch, 1971) of these two solutions is shown in figure 3D. A comparison can be made between ghost lineage origination points in each of the fundamental cladograms. These results are summarized in the table in figure 3. However, another dichotomous solution that could contribute to the consensus topology in figure 3D is shown in figure 3C. Because this cladogram is less parsimonious than those in 3A and B, there is no reason to include it in calculating range extensions.

Ghost lineage durations can be compared to the group durations in the upper left. These group durations are also listed in the row of the table (top left). For both cladogram A and B (the shortest topologies) ghost lineages for each of the taxa A, B, and C are 10. In topology C, lineage lengths for taxa B and C are only 6. Topology C is, however, longer than topologies A and B. Yet it is a possible dichotomous resolution of the multichotomy shown in the consensus solution. Considering it as a possible topology causes underestimation of the temporal extent of ghost lineages B and C predicted by the shortest solutions.

When more than one most parsimonious tree is found during cladistic analysis, the best technique is to treat each tree independently and present separate interpretations as equivalent alternatives. This procedure avoids the calculation of ghost lineage durations on trees that are not supported evidentially.

Boundaries.—Under what conditions do the stratigraphic records of taxa require correction? If the fossil record were extraordinarily precise and all taxa were the same age, no phylogenetic correction would be required, and no ghost taxa would be demanded. Such a result requires a polytomic diversification burst of all species from a single common ancestor—or special creation. Because this is not true of the fossil record, some taxa will always require a ghost lineage to be tacked onto their temporal range. This is even the case for groups that have no fossil record. If life is monophyletic, somewhere in nature all organisms have a sister-taxon for which a fossil record exists. No matter how remotely related, these fossils calibrate the date of divergence of those two groups.

Aside from sampling factors such as preservability of fossils and their numerical abundance, other factors influence the magnitude of phylogenetic correction on members of specific monophyletic groups. Consider the variable effect of phylogenetic correction on the duration of taxa. It may seem that taxa with short durations would be effected more by the
Fig. 3. A fossil record for four taxa (A, B, C, and D) is shown in the upper left. Below are four phylogenetic trees. Trees A and B are minimum length trees based on a phylogenetic analysis of taxa A, B, C, and D. Tree C is longer than trees A and B. D is a strict consensus tree for A and B; however, tree C could be added to this tree without topological modification making it one possible dichotomous resolution of tree D. Lineage lengths (the length of a group from the fossil record + the length of a ghost lineage determined from calibration of the phylogenetic hypotheses) are indicated in the table for each of the three fundamental topologies (top right). The asterisk row marks groups A, B, C, and D's first appearance in the fossil record.
phylogenetic correction than would long lived taxa. This is not the case. It is not the length of a taxon range that is important, it is time of origin relative to those of other members of its monophyletic group.

This leads to a general rule. The oldest member of a clade will never have its range extended until new data (for example, an older sister-taxon or additional older fossil of previously considered taxa) are added. It can only form the calibration point for all other ranges in the group.

Cladograms can be divided into two classes: pectinate and those with secondary topologic structure. Pectinate cladograms are those where all the terminal taxa branch off of a single phylogenetic axis and every terminal taxon is the sister-group to a sequentially less inclusive crown group (such as those in figure 3A, B, and C). In topologically complex cladograms, lineages branch off of the main phylogenetic axis only to branch again into terminal taxa (fig. 2A). In a pectinate cladogram or a topologically complex cladogram’s pectinate reduction (see discussion in Norell and Novacek, 1992 and submitted), where the oldest member is the sister-taxon to another taxon, and they form the crown group, the oldest member adds ghost lineages to all the taxa back to its point of origin. In these cases no additional ghost taxa are predicted by the cladogram. For the penultimate oldest taxon, the effect is much the same. A ghost lineage will never be added, unless it is a sister-taxon to the oldest.

Amount of correction required is largely a function of the topology of a phylogenetic tree, and specifically how well that tree matches the fossil record. Novacek and Norell (1992; submitted) outlined a tree-based procedure for examining how well individual fossil records conform with specific phylogenies. Basically this approach examines to what degree ranked patterns of occurrences correspond with ranked stratigraphic occurrences. Yet even in ideal cases, where stratigraphic pattern directly reflects the branching sequence of monophyletic taxa, phylogenetic corrections are still required. Even in the best cases, stratigraphic records underestimate the ages of clades. This is demonstrated by any phylogeny of monophyletic taxa. Even if only one branch is resolved, the phylogeny will contain an internal segment—this internal segment is a species that has not been recovered in the fossil record.

Ranks.—A shortcoming of most temporal studies is their reliance on taxonomic ranks as the units of comparison instead of concentrating on explicitly monophyletic taxa. In a phylogenetic system it is unclear what taxonomic ranks represent (De Queiroz and Gauthier, 1990). At best, ranks represent monophyletic taxa, yet most of the time they are subjective categories erected independent of a detailed knowledge of phylogeny (see comments in Smith and Patterson, 1988, regarding actual cases). While most systematists would agree that ranks are in no sense comparable, this has not stopped paleobiologists from compiling ranked taxa as the basic data for the examination of temporal patterns.

Tree-based approaches deal with monophyletic groups rather than ranks. For those not ready to purge the Linnean ranking system from the
vocabulary of systematics, ranks, if monophyletic, can be integrated into a tree-based approach. For instance, questions regarding the timing of appearance of higher categories (like the orders of eutherian mammals [Norell, 1992]) can be addressed by examination of their corrected ranges.

Yet problems remain especially in regard to ghost taxa. Take a simple example of four monophyletic taxa, that have been traditionally considered to be of equal taxonomic rank (for instance genera). Our question relates to the standing diversity of these genera through time. We want to know the standing generic diversity at a particular instant in time predicted by a particular phylogeny. First we examine the distribution of fossils (fig. 4A). If the phylogeny in figure 4B is preferred, the temporal lineage lengths of the genera A, B, C, and D can be determined. A and D are found to have ghost lineage segments. An additional ghost taxon (E) is recovered in the analysis. This taxon is predicted by the phylogeny, yet should it be classified as a “genus” in the compilation of genera through time even though no specimens of it have been recovered? The enigma of how and at what rank a ghost taxon is classified underscores the problem of subjectivity in rank based approaches in general and characterizes the root of their problem as one of focus. The question that should be asked is: How many taxa are present through time and how does this compare with what is observed in the fossil record? Instead of concentrating on the number and diversity of categories through time and space, attention should be riveted on the predictions of specific phylogenetic trees.

Fig. 4. A fossil record for four monophyletic genera (A, B, C, and D) (A) and a phylogeny of those taxa (B). Fossil occurrences are indicated by solid bars. A phylogenetic correction of the stratigraphic ranges in A results in two ghost lineages (the narrow lines attached to the ranges in A and D) and a single ghost taxon (E).
CONCLUSIONS

Phylogenetic approaches suggest reexamination of many of the basic tenets of paleobiology. For instance, what constitutes a good fossil record? Traditionally records with dense sampling (lots of fossils and finely ordered stratigraphy) have been used to test theories of process and construct phylogeny (Prothero and Lazarus, 1980; Lazarus and Prothero, 1984; Gingerich, 1979; Gould and Eldredge, 1977). Explicit is the assumption that the superpositional pattern of fossils represents direct ancestor-descendant relationships and changes in numbers of taxa, and first and last appearances reflect faunal changes.

Instead of ordering fossils through time, a phylogenetic approach orders fossils through phylogenies and evaluates the quality of fossil records through their congruence with independently constructed phylogeny. Therefore good fossil records are those that contain fossils that are adequately enough preserved to be incorporated into detailed cladistic analysis. Good fossil records are not necessarily those that have a lot of fossils, but ones that contain fossil specimens that are constituents of retrievable and repeatable cladistic phylogenies from which temporal patterns can be deduced.

ACKNOWLEDGMENTS

This paper was substantially improved through discussions with Michael J. Novacek, J. Clark, G. Edgecombe, and M. J. Donoghue. Jim Clark and Bruce Lieberman read the manuscript, and their comments are appreciated. Joan Davis and Mic Ellison aided with production of the manuscript.

REFERENCES


