Comparing the Gap Excess Ratio and the Retention Index of the Stratigraphic Character

JOHN A. FINARELLI AND WILLIAM C. CLYDE

Department of Earth Sciences, 56 College Road, University of New Hampshire, Durham, New Hampshire 03824, USA; E-mail: dryopithecus@excite.com, will.clyde@unh.edu

Because phylogenetic hypotheses make implicit statements about the temporal ordering of branching events, various metrics have been developed to reliably quantify the fit of observed temporal data to inferred phylogenies (Norell and Novacek, 1992; Huelsenbeck, 1994; Wagner, 1995; Clyde and Fisher, 1997; Benton et al., 1999; Wills, 1999). Several of these metrics attempt to quantify fit by evaluating the amount of range extension necessary to reconcile the implied branching order of a phylogenetic hypothesis with the observed sequence of appearances in the fossil record. Two of these, the Retention Index (RI; Farris, 1989a,b) calculated for the stratigraphic character (RI_{strat}; Clyde and Fisher, 1997) and the Gap Excess Ratio (GER; Wills, 1999), normalize the amount of implied range extension to the maximum and minimum possible extensions. Such normalized metrics are useful because they provide a way to quantify the fit of hypothesis to observation and to compare the degree of fit between hypotheses derived from datasets of various sizes. Although computationally the RI_{strat} and the GER are essentially identical, the data they use to evaluate hypothesis fitness differ. In this study, we reveal the similarity of RI_{strat} and GER for a set of fossil taxa with an observed stratigraphic distribution, and we compare the behavior of RI_{strat} and GER under controlled simulations to identify important differences in underlying assumptions of the two methods.

The Retention Index of the Stratigraphic Character

As formulated by Farris (1989a,b), the RI normalizes the observed number of state changes in a character relative to the maximum and minimum possible number of state changes for that character. The formula for the RI of a given character is:

\[
RI_i = \frac{\text{MAX}_i - \text{Obs}_i}{\text{MAX}_i - \text{MIN}_i}
\]

where \(\text{MAX}_i\) is the maximum possible number of steps and \(\text{MIN}_i\) is the minimum possible steps, respectively, for the \(i\)th character, and \(\text{Obs}_i\) is the observed number of steps for this character for the phylogenetic hypothesis under evaluation.

Clyde and Fisher (1997) used RI_{strat} (Fisher, 1991, 1992) to evaluate the fit of temporal data to phylogenetic hypotheses. In that study, stratigraphic character states were coded by using the observed succession of first and last appearance events (FAEs and LAEs). Clyde and Fisher (1997) used these appearance events to delimit “levels” in the stratigraphic record defined by the presence of some subset of the ingroup taxa (Fig. 1). Each stratigraphic level in which the existence of a particular taxon is implied by a phylogenetic hypothesis but is not observed in the fossil record is treated as an ad hoc stratigraphic statement, because at least one other hypothesis exists that does not make such a statement (Fisher, 1991). Each ad hoc stratigraphic statement incurred by a particular hypothesis is considered one unit of stratigraphic parsimony debt and is counted by MacClade's stratigraphic character as one “step.” This use of the term step differs somewhat from the traditional morphological and molecular use, where a step refers to any character-state transition, not just those that constitute parsimony debt (see discussions in Fisher, 1992; Clyde and Fisher, 1997:4).

The number of steps for the stratigraphic character takes on its maximum possible value for the phylogenetic hypothesis that joins all taxa to the level of the first appearance of the earliest taxon in the group. The minimum possible number of steps is realized when the taxa are arranged in an anagenetic lineage according to their order
of appearance in the fossil record. This makes the minimum number of steps for the stratigraphic character equal to zero, because at least one evolutionary hypothesis can account for the stratigraphic distribution of taxa with no implied sampling gaps in the fossil record. Computation of $RI_{\text{strat}}$ then follows the above formula for $RI$. The stratigraphic character can be coded in various ways (Fisher, 1992); we are restricting this study, however, to the coding method used by Clyde and Fisher (1997) because of its general applicability. Additionally, we note that the original formulation of the stratigraphic character, and therefore the $RI_{\text{strat}}$, was devised for the evaluation of phylogenetic trees, where ancestral taxa can explicitly be identified.

**THE GAP EXCESS RATIO**

Wills’ (1999) GER uses the absolute ages of ingroup FAEs to measure the fit of observed temporal data to the order of branching events implied by cladistic hypotheses. Given the $RI$ (Farris, 1989a) and the homoplasy excess ratio maximum (Archie, 1989), Wills proposed the formula:

$$GER = 1 - \left( \frac{MIG - G_{\min}}{G_{\max} - G_{\min}} \right).$$

(MIG is the minimum implied gap of a cladistic hypothesis as determined by the ordering of branching events. $G_{\max}$ and $G_{\min}$ are the maximum and minimum possible gaps, given the observed temporal distribution. $G_{\max}$ and $G_{\min}$ are calculated in a similar fashion to the MAX and MIN values of the stratigraphic character, where $G_{\max}$ is attained in the cladogram that joins all taxa at the first appearance of the earliest taxon, and $G_{\min}$ is attained in the cladogram that lines up the taxa in their order of appearance (Wills, 1999; Fig. 2). The GER is explicitly designed to evaluate cladograms and is unable (in its current form) to evaluate phylogenetic trees, for which ancestral taxa are explicitly identified.

Although absolute time is continuous, the gap data used in coding the GER are not continuous variables. The temporal information used to calculate the GER may be more or less precise ($10^6$, $10^5$, $10^0$ years, and so

---

**FIGURE 1.** Given four hypothetical fossil taxa (A, B, C, and D) with the stratigraphic ranges shown, and the given phylogenetic hypothesis, the GER codes the observed gaps as 3, 5, and 6 million years. The stratigraphic character codes each span of time as a single unit of stratigraphic parsimony debt.
FIGURE 2. A comparison of $R_{I_{\text{strat}}}$ and the GER for the exhaustive set of 10,395 cladograms relating seven species of the Eocene mammal Ectocion (original data from Thewissen, 1992). A strong linear relationship between the two metrics is observed, although $R_{I_{\text{strat}}}$ scores are consistently lower than the corresponding GER for a given cladogram. This is related to differences between the two methods in the way the minimum value is calculated. See text for discussion. Dashed line is the equation $y = x$.

Therefore, computationally, the GER is a retention index and the difference between the methods lies in the coding of the temporal data.

**Comparing $R_{I_{\text{strat}}}$ with GER**

To compare the behavior of the two metrics, we obtained the exhaustive set of 10,395 cladograms for seven species of the Eocene mammal Ectocion [see Thewissen (1992) and Clyde and Fisher (1997) for original morphological and temporal data] by using PAUP* v4.0b8 (Swofford, 1999). Because the GER in its current form can evaluate hypotheses only at the level of the cladogram, here we restrict our implementation of the stratigraphic character to cladograms to facilitate comparison of the two methods. The

\[
\text{GER} = \frac{G_{\text{max}} - \text{MIG}}{G_{\text{max}} - G_{\text{min}}}.
\]

\[
\text{Therefore,} \quad \text{GER} = \frac{G_{\text{max}} - \text{MIG}}{G_{\text{max}} - G_{\text{min}}}.
\]
sequence of appearance events was used to code the stratigraphic character, as presented by Clyde and Fisher (1997), and the number of stratigraphic steps was calculated for all cladograms by using MacClade v3.08 (Maddison and Maddison, 1999). Maximum and minimum values were then used to calculate RI\textsubscript{strat} for each cladogram. Absolute ages were obtained for the appearance events used to code the stratigraphic character (Wing et al., 2000; Clyde, 2001) and these ages were used to calculate GER values for all cladograms by using the program Ghost2 (Wills, 1999).

Comparison of RI\textsubscript{strat} and GER for the set of Ectocion cladograms reveals a strong positive correlation (Fig. 2). This correlation, however, does not fall about the line of identity \((y = x)\), as would be implied by the similarity of the formulae for the two metrics; rather, the values for RI\textsubscript{strat} tend to be somewhat lower than the corresponding GER values, because of the differences in calculating the minimum number of steps. The RI\textsubscript{strat} assesses the minimum possible number of steps by using the set of all possible phylogenetic trees. As such, the minimum number of steps in the calculation of the RI\textsubscript{strat} is zero, corresponding to the anagenetic lineage in which the taxa are arranged in order of appearance. The GER assesses the minimum implied gap by using only the set of possible cladograms; accordingly, the minimum for the GER will always be greater than zero. Restricting the analysis to consider only cladograms leads to RI\textsubscript{strat} values that appear somewhat lower. However, if this difference in the treatment of the minimum number of steps is accounted for in the calculation of the RI\textsubscript{strat}, and the temporal data are coded such that each stratigraphic character state is equivalent in length to one arbitrary time unit, the GER, as calculated by Ghost2, is identical to the RI\textsubscript{strat}, as calculated in MacClade, for every cladogram. The observed deviation from a perfect linear correlation between the two metrics for the Ectocion data set, therefore, must originate from the different methods of coding the temporal data. Using simple simulations, we have isolated several reasons why GER and RI\textsubscript{strat} will differ from one another in real world examples; in doing so, we can identify different assumptions that each approach makes about the fossil record.

### Simulations

Beyond differences in calculating the minimum possible number of steps, RI\textsubscript{strat} and GER behave in substantially different ways under different stratigraphic conditions. Various stratigraphic factors were examined for their differential effect on RI\textsubscript{strat} and GER. The exhaustive set of 105 cladograms for five hypothetical taxa was generated in PAUP* and the fit of each cladogram to a given stratigraphy was measured by both RI\textsubscript{strat} and GER. A pairwise comparison of each metric was performed for all cladograms. The simulations were organized into three cases. In the first case, we examined the effect of repositioning a relatively long stratigraphic interval. In the second case, we examined the effect of repositioning a last appearance event. In the third case, we examined the effect of sampling gaps in the fossil record and their position in the stratigraphic record. In each case we observed a strong correlation between GER and RI\textsubscript{strat}. However, individual cladograms were often characterized by significantly different values of stratigraphic fit depending on the metric used.

#### Case I—The Effect of Relatively Long Time Intervals

A stratigraphy was created for the five hypothetical taxa such that the time difference between the FAE of the first taxon to appear, taxon A, and that for the next taxon to appear, taxon B, was five units (Fig. 3a; Stratigraphy 1). The remaining taxa appear in sequence at uniform intervals of a single time unit. This creates a scenario in which coding for the stratigraphic character and the GER deviate from one another only for the long interval between the FAEs of taxa A and B. To avoid other complicating factors (see Cases II and III), all taxa were coded as ranging to “extant.” To test the effect of the position of this long time interval, we constructed an alternative stratigraphy identical to Stratigraphy 1 except the long interval was placed after the origination of taxon B (Fig. 3a; Stratigraphy 2).

Comparing the RI\textsubscript{strat} and GER between the two stratigraphies demonstrates the relative effect of an interval of variable length and position on the two metrics. Because the relative order of appearance events does not change between the stratigraphies, the
FIGURE 3. (a) Comparison of two stratigraphies for five hypothetical taxa containing a single long time span between the first appearances of two of the taxa. The position of the long time span (5 time units) is between the FAEs of taxon A and taxon B in Stratigraphy 1 and between the FAEs of taxon B and taxon C in Stratigraphy 2. Because the ordering of appearances does not change between the two stratigraphies, the coding of the stratigraphic character, and therefore $RI_{strat}$, remain constant. However, this longer interval is incorporated into the computation of the GER, changing the scale of the metric. Depending on the topology of the cladogram, GER is either increased or decreased. (b) $RI_{strat}$ and GER are calculated for two topologies. For the topology to the left, the GER is greater for Stratigraphy 1 and less for Stratigraphy 2. The opposite is the case for the topology to the right. $RI_{strat}$ is the same for both stratigraphies in all cases.
coding of the stratigraphic character and therefore $R_{I_{\text{strat}}}$, does not change for any given cladogram (Fig. 3b). However, the two stratigraphies generate different GER values for a given cladogram. The first cladogram in Figure 3b is characterized by a higher GER for Stratigraphy 1 and a lower GER for Stratigraphy 2, whereas the reverse is true for the second topology. The main difference between the two cladograms is that taxa A and B are not sister taxa in the first cladogram, but are in the second cladogram. With respect to the GER, the first cladogram fits the observed stratigraphy better than the second when the time period between taxa A and B is relatively long (Stratigraphy 1). However, the second cladogram fits the observed stratigraphy better when the long period serves to cluster these two taxa to the exclusion of the rest of the ingroup (Stratigraphy 2).

*Case II—The Effect of Last Appearances*

The scenario in Case I demonstrates that adding time intervals of various lengths to a stratigraphy will alter the GER values obtained for a set of cladistic hypotheses relative to a fixed $R_{I_{\text{strat}}}$ value. This is not the only difference in the relative behavior of the two metrics. A second difference in the coding of temporal data is the stratigraphic character’s treatment of last appearance events. The GER is unaffected by any change in last appearance events, being calculated solely from the difference in absolute time between first appearance events. As used by Clyde and Fisher (1997), the stratigraphic character makes use of all appearance events to define the boundaries of character states. For example, in Stratigraphy 3 (Fig. 4a), each of the five taxa appear at different times in the fossil record, but most importantly, no last appearance event occurs before a first appearance event. As altered in Stratigraphy 4 (Fig. 4a), the last appearance of taxon B occurs before the first appearance of taxon C, thus creating an additional character state for the stratigraphic character.

In this case, the value of $R_{I_{\text{strat}}}$ is variable because the coding of the stratigraphic character changes between the two stratigraphies. However, the ages for the FAEs remain unchanged in absolute time between the stratigraphies; therefore, the GER calculated for any cladogram is unaffected (Fig. 4b). The presence of an additional character state in the stratigraphic character favors topologies that cluster the taxa appearing after the decreased LAE of taxon B. Variable positions of LAEs, such that one or many appear before any of the FAEs, cause the $R_{I_{\text{strat}}}$ to be increased or decreased, the change being topology-dependent.

*Case III—Gaps in the Fossil Record*

A final difference between the two coding schemes can be observed in the treatment of sampling gaps in the fossil record. In some periods of time there may be no recovered fossils for the study group. Such sampling gaps—not to be confused with the term “gap” as used by Benton et al. (1999) or Wills (1999) to mean a span of time occupied by a ghost lineage—can be left uncoded for the stratigraphic character (see Clyde and Fisher, 1997; Fig. 4) and treated as “no data.” Left uncoded, these sampling gaps do not figure into the calculation of $R_{I_{\text{strat}}}$. However because GER codes the temporal data using absolute time, discounting these gaps is not possible. Some observed amount of time has always passed from the beginning to the end of the sampling gap, an amount that will be incorporated into the minimum, maximum, and observed values used to calculate the GER.

The effect of these sampling gaps in the fossil record was examined by producing a stratigraphy for five hypothetical taxa to include such a gap between the last appearance of taxon A and the first appearance of taxon B (Fig. 5a; Stratigraphy 5). To eliminate confounding factors, the intervals of time separating the remaining taxa were set to one time unit each and the length of the gap was also set to one time unit. A second stratigraphy was constructed identical to Stratigraphy 5 except that the sampling gap was placed between the last appearance of taxon B and the first appearance of taxon C (Fig. 5a; Stratigraphy 6).

Comparing the values of the GER with those of the $R_{I_{\text{strat}}}$ for the two stratigraphies clearly shows that sampling gaps in the fossil record affect the values of the GER relative to a fixed $R_{I_{\text{strat}}}$ (Fig. 5b). Once again, GER values vary depending on the topology of the cladogram. For the two cladograms considered in Figure 5b, the value of $R_{I_{\text{strat}}}$ does not vary between the two stratigraphies, because the sequence of appearance events does not change and the gap is not coded in either case. The GER, however, is affected as in
Figure 4. (a) Comparison of $RI_{\text{strat}}$ and GER when the position of a single LAE is changed. In Stratigraphy 3, five hypothetical taxa appear in succession. In Stratigraphy 4, the absolute ages remain the same for all appearance events, except for the LAE of taxon B, which has been placed below the FAE of taxon C. Because the ages associated with the FAEs of all taxa are unchanged between the two stratigraphies, the GER will not differ for a given cladogram. However, $RI_{\text{strat}}$ is affected by the change in position of the LAEs because it serves to define stratigraphic character states. (b) $RI_{\text{strat}}$ and GER are calculated for two different cladograms. For Stratigraphy 3, the $RI_{\text{strat}}$ value for the first topology is greater than the value for Stratigraphy 4, whereas this pattern is reversed for the second topology. This results because the second topology groups into a clade the taxa appearing after the decreased LAE of taxon B. The position of the altered LAE in the stratigraphic record, and thus the position of the added stratigraphic character state, will vary the values of $RI_{\text{strat}}$ depending on the topology of a given cladogram, but will not affect the value of the GER.
Figure 5. (a) Comparison of the effect of sampling gaps in the fossil record at different positions relative to the succession of taxa. A sampling gap exists between the LAE of taxon A and the FAE of taxon B for Stratigraphy 5. For Stratigraphy 6, the sampling gap occurs above the LAE of taxon B and below the FAE of taxon C. Sampling gaps are highlighted in gray. If the sampling gap is left uncoded, it will not figure into the calculation of $R_{\text{strat}}$, and because the order of appearance events is unchanged between the two stratigraphies, the value of $R_{\text{strat}}$ is not affected. However, the sampling gap will be factored into the GER metric because, some observed amount of time has passed between the beginning and end of the sampling gap. (b) GER and $R_{\text{strat}}$ values are compared for two topologies relative to the two stratigraphies. $R_{\text{strat}}$ is the same for both stratigraphies in all cases. The GER for first cladogram do greater for Stratigraphy 5 and less for Stratigraphy 6. The reverse is true for the second cladogram GER.
Case I. Again, the main difference between the two cladograms is the relationship between taxa A and B, which are not sister taxa in the first cladogram but are in the second cladogram. The first cladogram fits the observed stratigraphy relatively better than the second when the sampling gap separates taxa A and B (Stratigraphy 5), whereas the second cladogram fits the observed stratigraphy relatively better when the gap separates these taxa from the remainder of the ingroup (Stratigraphy 6). Because MacClade evaluates the coded stratigraphic character states in alphanumeric order, a gap can be coded in the stratigraphic character by skipping a letter or number in the coding scheme (Fisher, 1992). If treated in this manner, sampling gaps then become a special instance of Case I discussed above, where the $R_{I_{strat}}$ treats the sampling gap as a single stratigraphic character state, and the GER will treat the sampling gap as some variable amount of elapsed time.

CONCLUSIONS

Although computationally similar, $R_{I_{strat}}$ and the GER differ from one another when applied to a real group of taxa with observed stratigraphic ranges. $R_{I_{strat}}$ values are generally lower than GER values for a given cladogram, because of differences in calculating the minimum number of transitions. This scaling issue aside, simple simulations show that the remaining differences in the scoring between $R_{I_{strat}}$ and GER result from the different ways that each method codes temporal data. $R_{I_{strat}}$ and GER vary relative to each other based on three characteristics of the fossil record for a particular study group of taxa. Time spans of variable length between first appearances will cause the GER to vary relative to $R_{I_{strat}}$, because the absolute amount of time that has passed between appearance events is ignored in coding the stratigraphic character. The position of last appearances of ingroup taxa affects the score of $R_{I_{strat}}$ relative to the GER, because LAEs serve to define the boundaries of stratigraphic character states but are not used in the calculation of GER, which relies solely on FAEs. Lastly, the existence and position of sampling gaps in the fossil record of the ingroup will affect GER relative to $R_{I_{strat}}$ if the coding scheme utilized by Clyde and Fisher (1997) is used, because these gaps are factored into the calculation of the GER but not the $R_{I_{strat}}$. For each of these cases, the effect of the change in stratigraphy depends on the topology of the cladogram under consideration.

The different coding methods used for the $R_{I_{strat}}$ and GER reveal different assumptions about the fossil record. For the stratigraphic character, each instance in which a phylogenetic hypothesis postulates a lineage passing unrepresented through a stratigraphic level that preserves other ingroup taxa is treated as an ad hoc statement of nonpreservation. Each ad hoc statement is equally weighted for evaluating the stratigraphic fitness of a hypothesis regardless of how much absolute time it represents. In computing the GER, however, a mismatch between hypothesis and the fossil record is quantified as some integer number of units of absolute time. In this case, all absolute time units in which the existence of a taxon is implied by a particular hypothesis but not actually observed are essentially treated as an ad hoc statement and given equal weight. The fundamental difference between these approaches, then, boils down to whether the rock record is treated in terms of time (GER) or in terms of stratigraphy ($R_{I_{strat}}$).

The approach of the GER parses the rock record by using increments of absolute time. Here, the unit of measure is arbitrary, constant, and independent of the fossil record being considered. Because the GER treats each integer unit of time equally, the GER makes the assumption of horizontal (across taxa) equivalence of preservation potential and also includes an additional assumption of equivalence of preservation potential for each integer time unit. However, given that sampling intensity has been shown to vary greatly through time for a wide variety of taxonomic groups (Alroy, 1998; 2000; Marwick, 1998; Budd and Johnson, 1999; Aguirre et al. 2000), this observed variability in sampling intensity violates any assumption that the preservation potential for any one unit of absolute time can be treated as equivalent to the preservation potential for any other unit.

On the other hand, the stratigraphic character parses the rock record by using a relative succession of observed stratigraphic events, the FAEs and LAEs of the ingroup taxa. If one can demonstrate that event A happened at a stratigraphic level below that for event B, then A and B define a character state for the stratigraphic character. This approach assumes that the presence of any one ingroup taxon serves as a sufficient
condition for expecting the presence of other taxa in the group (i.e., ingroup taxa serve as reliable taphonomic indicators for each other; Fisher, 1991). As such, the RI\textsubscript{strat} assumes only horizontal equivalence in preservation potential for all ingroup taxa for each coded stratigraphic character state. This assumption states that the presence of one ingroup taxon demonstrates the potential for any other ingroup taxon to be preserved at the given level. No temporal assumption is made, and the preservation potential may vary dramatically through the stratigraphic range of the ingroup. This coding of the stratigraphic character represents the most conservative approach to parsing the temporal range of ingroup taxa, in that only the stratigraphic sequence of appearance events of the study group are used in evaluating the internal branching events.

The coding method used here (and in Clyde and Fisher, 1997) is not the only possible way to code the stratigraphic character. For example, sampling gaps in the fossil record can be coded (Fisher, 1992), and under such a scheme, the differences observed in Case III become a special instance of Case I. Additionally, data extrinsic to the ingroup could be utilized to hypothesize periods of equal preservation potential. For example the stratigraphic character for the Ectocion dataset could be coded for each North American Land Mammal Age subunit (e.g., Wasatchian 1, 2, 3, and so forth). In this case, associated faunal communities are used as the taphonomic indicators for ingroup taxa. However, as more data extrinsic to the ingroup taxa are considered in parsing the stratigraphic range of the ingroup taxa, more assumptions are added, although such approaches may be well suited to particular cases.

The GER and RI\textsubscript{strat} represent two computationally equivalent metrics for evaluating the fitness of phylogenetic hypotheses to observations of appearance events in the fossil record. The GER bases its differential weighting of ingroup ranges on absolute time, which has been shown to be problematic. On dropping this reliance on absolute time, the GER and RI\textsubscript{strat} are essentially identical.

**REFERENCES**


A Character-Based Method for Measuring the Fit of a Cladogram to the Fossil Record

KENNETH D. ANGIELCZYK

Department of Integrative Biology and Museum of Paleontology, 1101 VLSB, University of California, Berkeley, California 94720, USA; E-mail: etranger@socrates.berkeley.edu

Although the history of life cannot be observed directly, several sources of information preserve a historical signal that allows inferences about it. The phylogenetic relationships of organisms provide insight into the order of appearance of clades and their morphological evolution. The fossil record also preserves information about the relative and absolute ages of clades and provides documentation of the existence of organisms that otherwise would be unknown. Because only one true history exists, the signal preserved in each dataset ideally should be the same, and so predictions based on one can be tested with observations from the other.

Several techniques are now available to compare the fit of a cladogram to the stratigraphic record (e.g., Gauthier et al., 1988; Norell and Novacek, 1992a,b; Benton and Storrs, 1994; Huelsenbeck, 1994; Siddall, 1998; Wills, 1999), and others use stratigraphic information directly in the construction of phylogenetic trees (e.g., Gingerich, 1979; Fisher, 1988, 1991, 1992, 1994, 1997; Wagner, 1995; Clyde and Fisher, 1997; Huelsenbeck and Rannala, 1997). All of these methods use stratigraphic data associated with taxa, proceeding from the premise that the order of appearance of taxa on a cladogram and in the fossil record ideally should be the same. However, cladograms also make predictions about the order of appearance of character states, predictions that can be compared with the order of appearance of these states in the fossil record. Although to date, a character-based approach has not been implemented, this type of congruence between a cladogram and stratigraphy should be considered explicitly. Characters or complexes of characters can evolve in a mosaic or stepwise fashion, making it possible for the order of appearance of character states on a cladogram to conflict with that of the fossil record, even when the order of appearance of taxa does not.

CHARACTERS, TAXA, AND STRATIGRAPHY

The method proposed in this paper is based on the principle that taxa are hierarchically arranged lineages or clades (or at least should be; the lineage/clade status of all taxa has not been tested). Because lineages and clades cannot be studied directly, the distribution of synapomorphies is the primary evidence used in reconstructing relationships among different lineages or clades. Thus we recognize that Diictodon and Eodicynodon (Fig. 1a) share a more recent common ancestor than either does with Paratanomodon (Fig. 1a), based on a pattern of character-state distribution that we infer reflects genealogy.

A similar process occurs when the stratigraphic range of a taxon is measured. By definition, stratigraphic range implies a series of specimens of a given taxon found to occur in a sequence of rocks. Character states must be used to recognize any individual specimen as a member of a particular taxon. The earliest known occurrence of a unique, diagnostic set of character states is the first appearance datum for the taxon, whereas the latest occurrence provides a lower bound for when the lineage became extinct. Thus,