Biological Journal of the Linnean Society (1998), *64:* 493–511. With 9 figures

Article ID: bj980229

Optimal outgroup analysis

JAMES LYONS-WEILER[1]*, GUY A. HOELZER[1,2] AND ROBIN J. TAUSCH[1,3]

[1]*Doctoral Program in Ecology, Evolution and Conservation Biology, University of Nevada, Reno, Reno, NV 89512-0013, U.S.A.*
[2]*Departments of Biology and Environmental Resource Science, University of Nevada, Reno, NV, U.S.A.*
[3]*Rocky Mountain Research Station, Forest Service, USDA, Reno, NV, U.S.A.*

Received 12 September 1997; accepted for publication 10 March 1998

We present and critically examine a statistical criterion for the selection of outgroup taxa for rooting evolutionary trees. The criterion is the amount of phylogenetic signal for the ingroup when the states of the candidate outgroup taxa are assumed to be plesiomorphic relative to the ingroup for the purpose of measuring plesiomorphy content of the outgroup taxon. A statistical measure of rooted, ingroup signal was subjected to a suite of critical tests which indicate that it provides a proxy measure of plesiomorphy content. As the evolutionary distance between the ingroup ancestral node and outgroup taxa increases, the tree-independent measure of signal decreases, tracking the decay in plesiomorphy content and the increase in convergence to the ingroup states. We show that *a priori* generalizations about optimal outgroup taxon sampling strategies are likely to be misleading, and that testing for the suitability of available outgroup taxon sampling in specific instances is warranted. Software for optimal outgroup analysis is available.

ADDITIONAL KEY WORDS:—phylogenetics – RASA – phylogenetic signal – outgroup – rooting – plesiomorphy content.

CONTENTS

*Correspondence to: J. Lyons-Weiler. Present address: Institute of Molecular Evolutionary Genetics, 328 Mueller Laboratory, University Park, PA 16802, U.S.A.

0024–4066/98/080493+19 $30.00/0

INTRODUCTION

The problem: not all outgroups are created equal

The most commonly applied method for rooting evolutionary trees is outgroup analysis (Stevens, 1980; Eldredge & Cracraft, 1980; Arnold, 1981; Nelson & Platnick, 1981). One of the most important aspects of outgroup analysis is the selection of outgroup taxa, but the decision is often apparently arbitrary, or based on poorly understood theoretical considerations. For example, it is not generally appreciated that all of the details of an evolutionary tree, such as topology (branching order and branch length), and convexity (loosely equivalent to the direction of the graph) can vary with the choice of outgroup (e.g. Milinkovitch *et al.*, 1996; Milinkovitch & Lyons-Weiler, in press). This sensitivity translates into variance in biological inferences such as clade monophyly, character polarity, character state transformations, estimated dates of divergence and rates of evolution. Some criteria have been proposed to provide objective means for selecting among the millions of potential outgroup taxa on earth (e.g. Maddison *et al.*, 1984; Mooi, 1989; Donoghue & Cantino, 1984; Watrous & Wheeler, 1981). Watrous & Wheeler (1981), Donoghue & Cantino (1984) and more recently Smith (1994) suggested that outgroup taxa should be shown to be monophyletic with the ingroup in a larger study, and that sister taxa should be used when possible. Donoghue & Cantino (1984) also suggested that confidence should only exist for nodes in the ingroup that are robust to outgroup choice. Although the sister taxon criterion, and other options (such as the use of all available taxa) may appear reasonable and objective, they only serve as general rules of thumb for outgroup choice. We will demonstrate that, in spite of the appeal of such guidelines, such rules of thumb do not usually hold for specific cases. This is possibly why none of the criteria proposed to date have been widely adopted, and the decision processes used to select particular taxa for rooting trees often remains ambiguous.

The purpose of this paper is to define and critically examine a statistical criterion for the justification and optimal selection of outgroups. Our criterion for accepting an outgroup (which may be comprised of one or more taxa) is that phylogenetic signal in a rooted analysis must exceed that found for the unrooted ingroup alone. The criterion is based on the expectation that different outgroup lineages will have converged differentially to states that are synapomorphic (shared, derived) within an ingroup of interest, and that different outgroup lineages will differentially exhibit states that are plesiomorphic for the specified ingroup. The criterion invokes the operational assumption that the character states of an outgroup are plesiomorphic, to the end of measuring the effects of that working assumption in terms of phylogenetic

signal. The assumption of plesiomorphy for a state is implemented by constraining (ignoring) that state in that character during ingroup comparisons while determining the amount of signal in the set of ingroup taxa. The degree to which the operational assumption that the character states of an outgroup are plesiomorphic is violated is measurable as a change in amount of phylogenetic signal in rooted analyses compared to unrooted analysis. For various positions how outgroup taxa may be used when performing tree-based phylogenetic analyses, which is an interesting, unsettled, and largely distinct issue, see Watrous and Wheeler (1981), Wheeler (1981), Wiley (1981) Farris (1982), Donoghue & Cantino (1984), Maddison *et al.* (1984), Mooi (1989), Nixon & Carpenter (1993), and Smith (1994). The essence of outgroup analysis is to define the root of the ingroup. The use of the best possible taxa to root a tree is the essence of optimal outgroup analysis; regardless of how one eventually uses an outgroup, the choice of any particular outgroup should be justified.

Plesiomorphy content

Methods of phylogenetic inference, such as maximum parsimony (Camin & Sokal, 1965; Cracraft & Helm-Bychowski, 1991) will perform better (with respect to accuracy) when they are applied to data that contain phylogenetic information (signal). It is also important that unambiguous phylogenetic signal exist for each branch of the true tree. For example, there should be no 'long branches' on which anagenesis has erased too much evidence of common ancestry, and has lead to convergence on character-states of other taxa in the analysis (Felsenstein, 1978; Lyons-Weiler & Hoelzer, 1997). Outgroup selection also influences accuracy. A general expectation is that the accuracy of root placement via outgroup comparison will be highest when the plesiomorphy content of the outgroup used is highest, especially under the criterion of maximum parsimony (Wheeler, 1990). A direct measure of plesiomorphy content is the proportion of characters in an outgroup taxon that exhibit the ancestral state, in spite of independent anagenetic evolution of outgroup lineages. Given that such a measure would require complete knowledge of evolutionary history, a proxy measure of plesiomorphy content is needed.

The general expectation outlined above is modified in specific cases because no outgroup taxon is likely to have retained complete plesiomorphy. Outgroup taxa are no different than ingroup taxa in that outgroup characters have presumably continued to evolve since the divergence from the most recent common ancestor (MRCA) with an ingroup. In general, therefore, it can be expected that plesiomorphy of any specified ingroup will be variously distributed among candidate outgroup taxa.

Three factors will determine the distribution of plesiomorphy in any outgroup, and, in turn, the effects of this distribution on rooted phylogenetic analyses: (1) the number of characters that have changed, (2) which characters have changed, and (3) how those characters have changed. The first factor is the evolutionary distance between a particular outgroup and the MRCA of the ingroup. Evolutionary distance itself is a function of anagenetic rate and time. When the evolutionary distance between an outgroup and the MRCA of the ingroup is great, plesiomorphy content of that outgroup will have (usually) decayed. If rates vary among lineages, the sister taxon may not have the shortest evolutionary distance to the ingroup, reducing the chance that it is the optimal candidate for use in estimating the ingroup ancestral

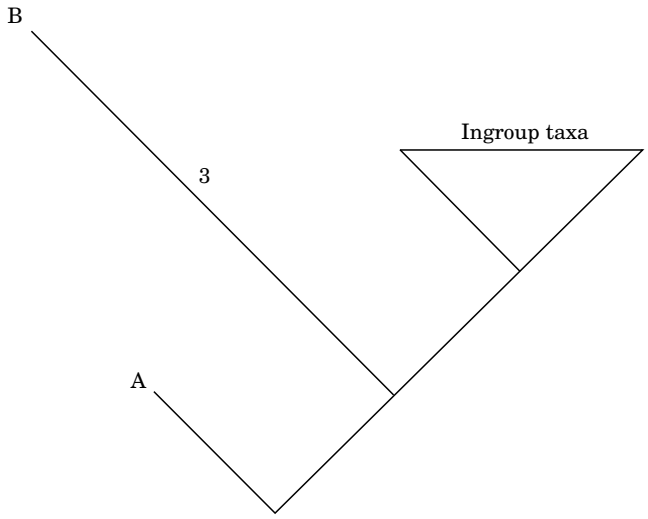

Figure 1. The validity of the assumption that the sister taxon (B) will provide the best estimate of plesiomorphy for the ingroup taxa is rate-dependent. Taxon A is likely to be of more use in assessing the ancestral states of characters. Therefore, both the position of the root (relative to the ingroup taxa) and the polarity of the characters may be better served by using taxon A instead of the sister taxon B.

node, regardless of which outgroup rooting procedure is used (e.g. whether ancestral states are estimated or are taken directly from the outgroup: Farris, 1972; Maddison *et al.*, 1984). For example, in Figure 1, the mutation rate for the sister taxon (B) has increased threefold relative to branches of equal duration, while the rate of evolution in the next basal taxon (A) has slowed considerably. Given that increases in the amount of independent evolutionary history will lead ancestral state inferences farther and farther afield, the sum of the branch lengths between A and the ingroup node (3μ) suggests that it would tend to yield a more accurate estimate of the ancestral node than would the sister taxon (B), which is 4μ away from the ingroup node.

The second factor is *which* characters have changed during the course of evolution in the ingroup and outgroup candidates. The relative importance of change in any given character in an outgroup is determined by the details of the history of character evolution in the ingroup. The representation of plesiomorphy in some characters will be more important for accurate root placement than for other characters. For example, the set of characters for which plesiomorphic representation in outgroup taxa is not relevant to parsimony includes invariant characters and those with misleading distributions of character states (e.g. sites in a biological sequence that are saturated by mutation). The set of characters for which outgroup plesiomorphy is important are those that exhibit variance in states representing shared geneaological relationships (i.e. characters carrying phylogenetic signal). No *a priori* assumption that all characters will carry phylogenetic signal is required for phylogenetic analyses (Lyons-Weiler *et al.* 1996). We therefore define those characters for which outgroup plesiomorphy is important for accurate rooting as 'the set of relevant characters'.

In some cases, many or most of the characters in an outgroup taxon may have

evolved to new states. However, if the set of relevant characters has not evolved in the outgroup, that taxon may provide a useful estimate of ingroup plesiomorphy, regardless of total evolutionary distance. Hence, one can expect variance in relevant plesiomorphy content to result even when the evolutionary distances between the ingroup and different outgroup taxa are identical. A measure of overall plesiomorphy content may not provide the best indicator of outgroup suitability, while a measure of plesiomorphic content that accounts for the differential importance of ancestral information in the outgroup states for the set of relevant characters would be more useful.

The third factor is *how* the evolving characters in the outgroup have changed. The same characters may change in different outgroup taxa, but which states they exhibit will be of primary importance. In particular, the proportion of the set of relevant characters that has converged to states that represent unique synapomorphies within the ingroup, and the proportion of of the set of relevant characters that have converged to evolutionary homoplasy within the ingroup will help determine the suitability of an outgroup. Convergence in an outgroup to non-plesiomorphic ingroup states will tend to be most frequent when the evolutionary distance between the outgroup candidate and the ingroup is great; therefore, taxa that exhibit both the highest amount of plesiomorphy, and that have also converged the least on ingroup synapomorphies will tend to most resemble the MRCA of the ingroup for the set of relevant characters. It is desirable to use an outgroup that exhibits both the highest amount of plesiomorphy, and that have also converged the least on ingroup synapomorphies.

These three factors can interact in counterintuitive ways. Even with equal evolutionary distances, different outgroup taxa may vary considerably with respect to plesiomorphy content. Outgroup lineages with higher amounts of character evolution may nevertheless retain higher plesiomorphy content than slower-evolving lineages, depending on which sets of characters have changed in the course of the evolution of the outgroup lineages *and* the ingroup taxa, and how those characters have changed. The retention of informative plesiomorphy in an outgroup depends heavily on the degree to which these sets of characters overlap. From this perspective, it may be fruitless to attempt to generalize about which type of outgroup might be suitable or best to use in any instance.

RASA

Given that methods of phylogenetic estimation will tend to more accurately place the root and achieve more accurate evolutionary tree topologies when outgroups with greater and more relevant plesiomorphy content are used, a proxy measure of plesiomorphy content for the set of relevant characters would be useful, in specific instances, as a guide to choice among the outgroup combinations that exist. A diversity of tree-based measures of plesiomorphy content are possible, but tree-based measures of information have numerous undesirable properties, such as requiring solutions to a computationally complex problem (solving for trees yields NP-Completeness; Garey & Johnson 1979). More importantly, errors in the ingroup tree estimate, caused by any number of factors, including homoplasy, would lead to errors in any such measure of plesiomorphy content. The potential for confounding between errors in tree estimates and tree-based information measures makes interpretation difficult. Archie (1996) provided a comprehensive review of tree-based measures of phylogenetic information.

An alternative to tree-based measures of phylogenetic information is the regression-based, Relative Apparent Synapomorphy Analysis (RASA; Lyons-Weiler *et al.*, 1996). Tests using RASA have been devised for differential lineage sorting (Lyons-Weiler & Milinkovitch, 1997), for the detection and identification of long-branch taxa (Lyons-Weiler & Hoelzer, 1997), and to provide two orthogonal measures of nestedness in species' distribution via Lundberg and AntiLundberg rooting (Lyons-Weiler & Tausch, 1996). In the unrooted form, the method provides a statistical measure of the degree to which the rate of increase in apparent cladistic similarity (measured by relative apparent synapomorphy, RAS) per unit phenetic similarity among pairs of taxa differs from that expected by an equiprobable (null) model. The null model used by Lyons-Weiler *et al.* (1996) is that apparent cladistic similarity increases at a rate that is proportional to overall similarity when each is reciprocally redistributed among pairs of taxa. The difference between the observed and null slopes is measured by a test statistic (t_{RASA}) that has a distribution similar to the *t*-distribution. As a measure of phylogenetic signal, it is designed to prevent the construction of trees from spurious or misleading character state distributions. During the course of analysis of published data, we have observed dozens of instances where the null hypothesis cannot be rejected. Unlike other measures of information, the test can provide a mechanistic explanation for the apparent absence of signal, such as the presence and identity of problematic taxa (Lyons-Weiler & Hoelzer, 1997). Moreover, the test does not depend on *ad-hoc* strategies, such as character weighting, although encoding characters in classes, such as nucleotides into purines and pyrimidines, may reveal hidden signal.

In an unrooted analysis, the apparent synapomorphy shared by any pair of taxa is equal to the number of times any other taxon exhibits a state that is different from that shared by that pair of taxa (Fig. 2; RAS – upper diagonals). This measure of pairwise cladistic similarity (RAS) indicates the degree of uniqueness in the similarity for a given taxon pair, all characters considered. This measure is regressed on a pairwise measure of phenetic similarity (E), which corresponds to the number of variable characters for which that pair of taxa shares a state; Fig. 2; E – lower diagonals).

Phylogenetic signal is considered to be present, or at least not obscured, when the observed slope (βobs) of the regression model (1) is significantly greater than that expected under the equiprobable model ($\beta null$ = null slope in Lyons-Weiler *et al.*, 1996, or βp if a permutation estimate of the null slope is used; Lyons-Weiler and Hoelzer, in prep). The homogeneity of slopes test is used to compare the observed slope in the model.

$$\mathbf{R\hat{A}S}_m = \beta obs\mathbf{E}_m + \beta_{0+}\varepsilon_m \quad (1)$$

to the null slope, where **RÂS** is the predicted value of the pairwise measure of cladistic similarity (**RAS**) for the *mth* taxon pair, **E** is the phenetic measure of similarity, β_0 is the intercept in the RASA regression, and ε is the error with which **RAS** is estimated by the model.

In a rooted analysis, the ingroup matrix is reduced by eliminating, for the purpose of data exploration only, the character-states of the candidate outgroup (one or more taxa) during the calculation of **RAS** and **E** (Fig. 2; lower left matrices). This prevents comparisons among taxa on the basis of putative plesiomorphy for the purpose of testing the purely operational assumption that outgroup states are ancestral for the set of relevant characters.

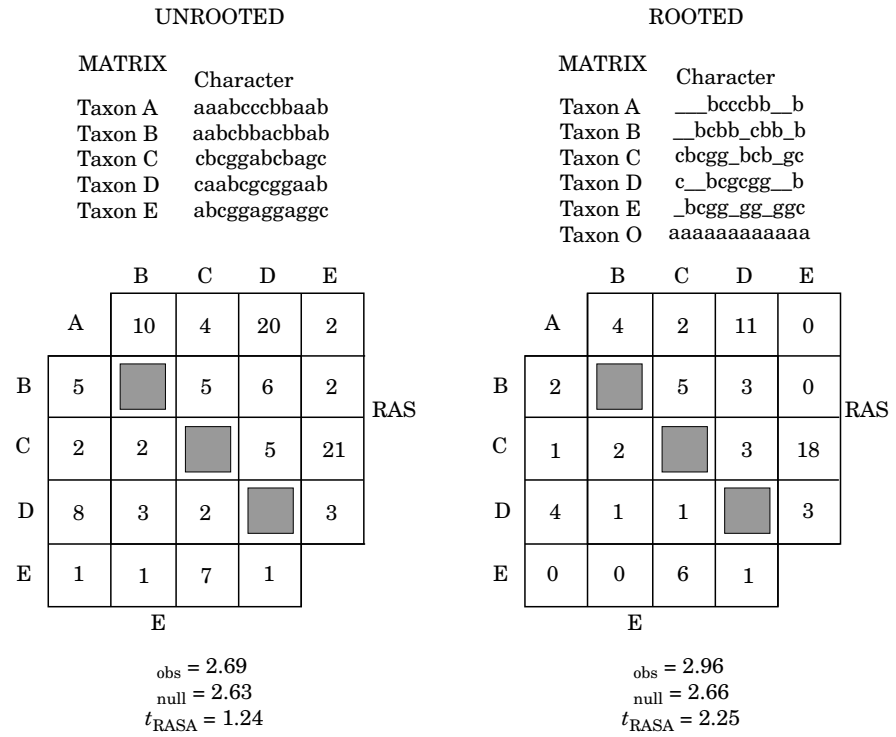

Figure 2. Comparison of all pairs of taxa to all other individual taxa for all characters results in the pairwise measure of similarity **RAS**. Here an unrooted (left) example is shown; the upper right diagonal is **RAS**; the lower left diagonal is the pairwise phenetic measure **E**. On the right, the same character state matrix is rooted on taxon O, resulting in lower **RAS** and **E** scores, and an increase in signal over the unrooted analysis. βobs = observed slope, $\beta null$ = null slope.

When an outgroup without relevant plesiomorphy content is used, a diversity of ingroup states are constrained for the ingroup comparisons, including plesiomorphy, synapomorphy, and homoplasy. When plesiomorphy alone is constrained, the proportion of character comparisons among ingroup taxa that conflate symplesiomorphy with synapomorphy is reduced. The expectation when rooting with plesiomorphy, therefore, is an increase in signal. Signal is expected to be reduced in arbitrarily rooted analyses compared to the unrooted analysis because true synapomorphies will be constrained at random.

MATERIAL AND METHODS

Bacteriophage T7 data

In our evaluation of the behavior of t_{RASA} as a criterion for optimal outgroup choice, we performed a variety of unrooted and rooted RAS analyses on the matrix of restriction site characters for the experimentally evolved phylogeny of bacteriophage T7 from Hillis *et al.* (1991). The data resulting from this experimental

phylogenesis were determined to lead to the correct tree topology, regardless of the method of phylogenetic inference used (Hillis *et al.*, 1991). In addition, Lyons-Weiler *et al.* (1996) found that the matrix of ingroup character states alone contained phylogenetic signal ($t_{RASA} = 11.3$; $P < 0.001$), and when rooted with the sister taxon, signal increased ($t_{RASA} = 12.7$; $P < 0.001$).

We also used these data to determine whether rooted analyses would correctly identify an outgroup that threatened the accuracy of rooting the tree (Hillis *et al.*, 1991). First, we rooted the matrix of ingroup states on each ingroup taxon in the bacteriophage T7 matrix, expecting that signal would decrease relative to the unrooted ingroup. In this case, if rooting the ingroup matrix on any of the ingroup taxa resulted in an increase in t_{RASA}, the criterion would have failed.

As a second critical test, we performed 2100 rooted analyses wherein the character states of the true outgroup were randomized without replacement across characters. If an unacceptable proportion (e.g. >5%) of the permuted analyses resulted in a test statistic ≥ 11.3 (the value for the unrooted analysis), we would be forced to conclude that the action of rooting informative data with virtually any outgroup might result in a spurious increase in the test statistic, once again potentially indicating a flaw in the criterion.

Random data

To further evaluate the possibility that increases in signal were expected by chance alone when a character state matrix is arbitrarily rooted, we generated 1000 random quaternary state data matrices with even character state probabilities for seven taxa, and performed, in tandem, rooted and unrooted RASA. Outgroup strings were also randomly constructed. A paired *t*-test was performed to compare the two distributions.

Decay in plesiomorphy content

As the branch connecting an ingroup to an outgroup becomes longer, the information on plesiomorphy is eroded. To determine if t_{RASA} tracks the decay in plesiomorphy content as the evolutionary distance to an outgroup increases, we performed a series of simulations in which the ingroup topology was held constant, but the depth of the ingroup-outgroup internode (D) was serially increased. We simulated the evolution of 50 quaternary state characters on a specified topology (Fig. 3) at a mutation rate of $\mu = 0.08$, where μ is the proportion of characters changing on each branch of one standard length. This mutation rate has been shown to yield the greatest amount of signal in these simulated conditions (Lyons-Weiler *et al.*, 1996). Four-hundred iterations at each level of D, which varied from μ to 20 μ, were conducted. If t_{RASA} does not, on average, decrease monotonically with increasing D, its utility as a criterion would be questionable.

Are generalizations useful?

Optimal outgroup analysis is well suited to circumstances in which multiple, putative, outgroup candidates are available. However, if one particular outgroup

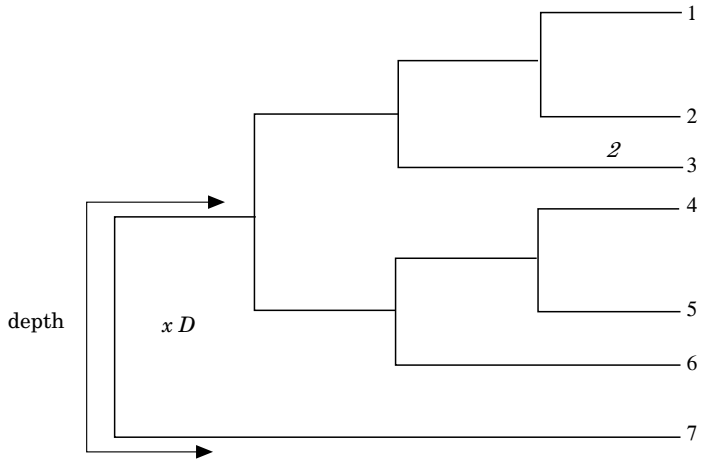

Figure 3. The topology used to generate character state matrices with 7 taxa. The outgroup taxon (7) was used in rooted analyses of signal (RASA; Lyons-Weiler *et al.* 1996), where the depth of the branch connecting the outgroup taxon and the ingroup varied from $D = \mu$ to $D = 20\mu$.

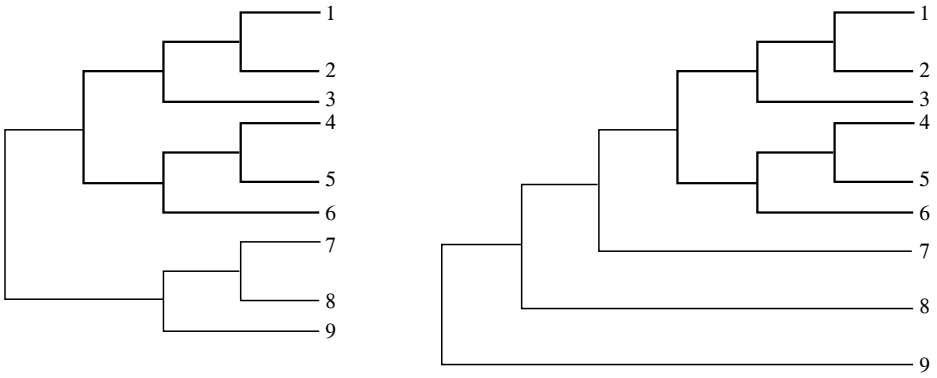

Figure 4. The same topology (taxa 1–6) rooted with a monophyletic (A) and a pectinate (B) outgroup. No comparison is made between A and B; the comparisons made were the same topology rooted with all-merous combinations of the outgroup taxa.

sampling strategy is optimal most of the time, then case-by-case optimal outgroup analysis would be of limited utility. For instance, it may be best to sample three outgroup taxa, as long as one is the sister taxon. Alternatively, it may always be most appropriate to sample as many outgroup taxa as possible. Smith (1994) concluded that it should, in general, be better to sample more sister taxa than other taxa (i.e. to prefer *n* taxa from a monophyletic sister taxon to *n* taxa from a pectinate outgroup). To determine if the test would support any generalizations that might be useful for an optimal strategy for taxon sampling, we simulated character evolution on two nine-taxon topologies (Fig. 4) under the conditions described above (50 characters, four states, basal $\mu = 0.08$; 1000 trials). Three outgroup taxa were used. The outgroups were evolved both as a monophyletic group (Fig. 4A), and as a

TABLE 1. Signal in restriction site matrix (T7 Bacteriophage; Hillis, 1991) when rooted on each taxon

ingroup	t_{RASA}	'out-group'	t_{RASA}
all-J&R	10.145	J	4.664[a]
all-K&R	8.749	K	3.474[a]
all-L&R	8.945	L	4.641[a]
all-M&R	9.899	M	3.556[a]
all-N&R	4.834	N	3.150[a]
all-O&R	8.305	O	4.225[a]
all-P&R	5.266	P	3.631[a]
all-Q&R	6.114	Q	4.051[a]
all-R	11.3	R	12.7 [b]

[a] ingroup
[b] outgroup
See Hillis *et al.* (1994), text, and Lyons-Weiler *et al.* (1996) for details.

pectinate, paraphyletic group (Fig. 4B). In each simulation, all possible combinations of outgroups were used to root the ingroup matrices (i.e. taxon 7, taxon 8, and taxon 9 alone; taxa 7+8, taxa 7+9, taxa 8+9; and taxa 7+8+9). When the outgroup taxa were monophyletic, the depth of the first outgroup node was arbitrarily predetermined to occur at 1/2 of the anagenetic branch length leading from any ingroup taxon to the outgroup-ingroup node. The internodes of the outgroup were held constant to topology (*sensu* Lyons-Weiler *et al.*, 1996) for both monophyletic and pectinate outgroup scenarios.

We determined the proportion of times each taxon sampling return the highest test statistic ($=P_{max}$). If any general, optimal outgroup taxon sampling strategy exists, we would expect that the optimal outgroup taxon sampling would reveal itself by leading to the highest test statistic a great proportion of the time (e.g. $P_{max} > 95\%$). If no outgroup sampling will achieve the maximum test statistic almost all the time, then no generalization or rule of thumb can be supported.

RESULTS

Bacteriophage T7 data

In the first analysis of the experimental phylogeny data, all but one rooted analysis resulted in a reduction in signal, but none resulted in the complete loss of signal (Table 1), and the sister taxon (Taxon R) alone yielded an increase in signal. This demonstrates both that the use of different putative outgroups results in variance in measured signal content, and that the variance can be used to identify the most appropriate outgroup from a set of candidates. If the outgroup taxon had been in question, Taxon R would have clearly been selected as an obvious outgroup choice, and the increase in signal over that for the unrooted ingroup would have indicated that it was a suitable outgroup.

Permutation of the outgroup states also consistently led to a decrease in signal

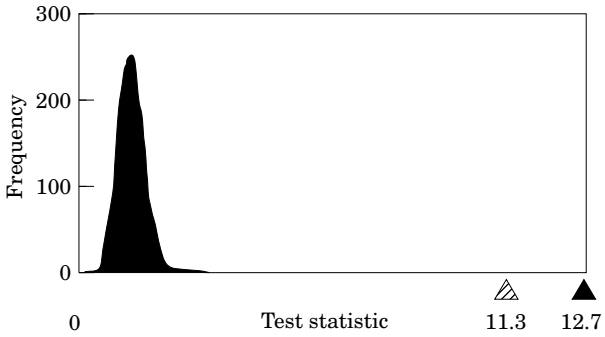

Figure 5. The results of randomization of the outgroup taxon R's character states from the Hillis *et al.* (1991) experimentally derived lineages of the bacteriophage T7. This distribution shows that none of the permuted analyses approached the observed amount of phylogenetic signal when the analysis was unrooted ($t_{RASA}=11.3164$; △), or when it was rooted with the original, non-permuted characters states of the appropriate outgroup ($t_{RASA}=12.68451$; ▲).

relative to unrooted analysis (Fig. 5). These results refute the hypothesis that a spurious increase in t_{RASA} will result when *informative* data are rooted at random. The frequency distribution that resulted from this permutation tail probability test (Fig. 5) indicates that the action of rooting informative matrices with randomized states will, for all practical purposes, never cause an increase in the observed amount of signal.

Random data

As with informative data, there should also be no systematic increase in signal when random matrices are rooted with random data. Both rooted and unrooted distributions show a central tendency near zero, as expected. If the act of rooting on outgroup states causes a spurious increase in measured signal, the rooted analyses would show a greater mean test statistic than the unrooted analyses. A paired *t*-test revealed a slight difference in the opposite direction. The mean test statistic for the unrooted analyses was 0.702, while the mean test statistic for the rooted analyses was 0.365 (Student's $t=-0.2025$; df$=999$; $P=0.043$; Fig. 6). The effect of rooting random data was a minor decrease in the test statistic, as well as a decrease in the standard deviation of the distribution of test statistics (rooted SD$=0.676$; unrooted$=$ 1.305; Fig. 6).

Decay in plesiomorphy content

A useful indicator of plesiomorphy content should show a steady decrease in the amount of signal measured as the ingroup–outgroup branch length is increased. Simulation of evolved data showed that t_{RASA} behaves appropriately under these conditions (Fig. 7), and tracks both the decay in plesiomorphy content in the outgroup, and the increase in the random convergence to ingroup synapomorphy, as the evolutionary distance between the ingroup ancestral node and the extant

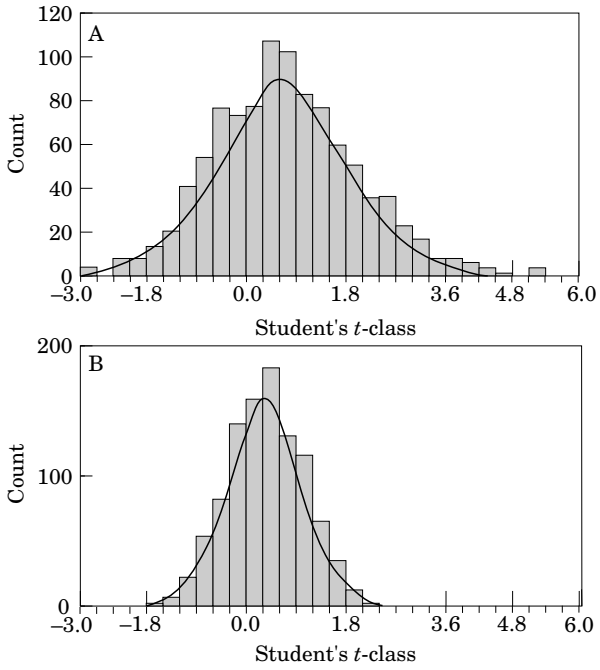

Figure 6. Rooted measures of signal (A) are slightly lower, and are less dispersed, than unrooted measures of signal (B) for random data. This suggests that increased signal obtained by constraining putative plesiomorphies found in outgroup taxa may result in more accurate phylogenetic inference.

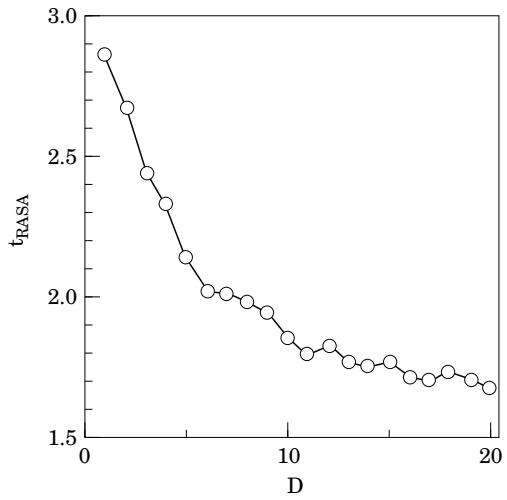

Figure 7. The results of the analysis of the character state matrices generated by the topology in Fig. 3. As the depth of the branch connecting the ingroup and outgroup taxa increases (D), the amount of signal measured in the rooted analyses decreases (t_{RASA}).

outgroup increases. The test thus provides a relative measure of relevant plesiomorphy content among various outgroups. Because the criterion is phylogenetic signal, this relative measure is, by definition, for the set of relevant characters.

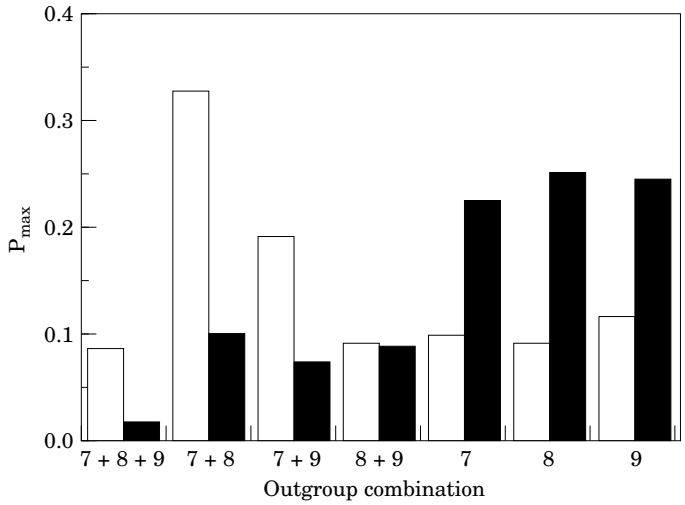

Figure 8. Use of all combinations of outgroup taxa in simulated phylogenies (Fig. 4) reveals effects of the phylogenetic relationships of sampling from two types of outgroup taxa (pectinate (□) vs. monophyletic (■) on the relative estimate of plesiomorphy content. P_{max} is the proportion of times that the outgroup combination yielded the highest plesiomorphy content within the set of simulations for the particular outgroup type in 1000 simulation runs.

Are generalizations useful?

No particular strategy of outgroup sampling was found to be generally optimal. Indeed, the use of any one strategy will be suboptimal most of the time (Fig. 8). Sometimes the use of one, sometimes the use of a few, and sometimes (albeit rarely) the use of all of the members of a monophyletic outgroup will maximize signal (Fig. 8). Note that for a paraphyletic outgroup (Fig. 4B), the sister taxon is not necessarily included in the subset of available outgroups that would (by our criterion) lead to the best estimate of ingroup plesiomorphy. A close examination of the results (Fig. 8) shows that the inclusion of more than one representative taxon from a monophyletic group usually tends to decrease signal. Some apomorphies in the outgroup will have converged on ingroup synapomorphy; it may be that as more representatives of a monophyletic outgroup are included, the probability of sampling such apomorphies will increase. An additional factor is a reduction in the number of effective characters that accompanies RASA rooting, which is discussed in the *Limitations* section (below).

These simulations suggest that the best general strategy, although a poor one, is to use two species from a paraphyletic outgroup, including, if possible, the sister taxon (Fig. 8). In addition, given a choice between using a pair of closely related outgroups and a pair of more distantly related taxa, the latter choice might be the better combination. This finding is in opposition to that of Smith (1994). Finally, the inclusion of all available outgroup candidates may occasionally decrease, rather than increase, the measured estimate of plesiomorphy content.

DISCUSSION

Accidents of history

During the course of evolution, traceable evidence of evolutionary relationships is not always imprinted in the character states observed in extant taxa. The focus is therefore placed on the statistical measurement of phylogenetic signal that is found in the distribution of character states among taxa. After speciation, each taxon embarks on an independent evolutionary trajectory; therefore, plesiomorphy will be distributed variously among available outgroup taxa. This distribution will be influenced by the accidents of history that have 'erased' a greater or lesser proportion of plesiomorphy during the independent anagenesis of any outgroup lineage. In general, no outgroup will lead to a perfect estimate of ingroup plesiomorphy. There will, therefore, almost always be measurement error in the estimated ('reconstructed') MRCA of the ingroup, which will vary as different outgroup candidates are used (cf. Milinkovitch *et al.*, 1996). This error, however, may adopt a counterintuitive structure in many instances: if a deeper taxon has converged less toward the ingroup states than a closer taxon, for the set of relevant characters, use of the more distantly related taxin may lead to fewer errors in the global (ingroup + outgroup) topology.

We have shown that these accidents of history reveal that proposed 'rules of thumb' for the identification of optimal combinations provide ineffective criteria with a finite amount of data. Even under identical and optimal conditions for the generation and retention of phylogenetic signal, no simple prescription for the selection of an outgroup suffices. In addition, even the best choice rule suggested by Figure 8 (i.e. use two taxa from a pectinate outgroup) fails to find the optimal outgroup most of the time. The utility of such operational rules seems dim. They are not required, however, because optimal outgroup choice can be applied in specific instances, for instance by using RASA to measure objectively the influences of different outgroups. This should enable researchers to determine which outgroups have apparently retained enough plesiomorphy in the appropriate characters to reduce, as much as possible, errors in phylogenetic estimation. If the use of all available outgroups consistently obliterates phylogenetic signal that has been previously detected in an unrooted RASA analysis (i.e. no outgroup seems suitable), perhaps the appropriate course of action would be to then estimate an unrooted phylogenetic tree of the ingroup taxa alone.

We have shown that the sister taxon will not always lead to the best estimate of plesiomorphy, even when evolutionary rates are equal. The accidents of history that occur during the anagenesis of any outgroup lineage may have 'erased' a greater or lesser proportion of plesiomorphy for different, but potentially overlapping sets of characters. If a sister taxon has converged more towards an ingroup than a more distant taxon, the best estimate of ingroup plesiomorphy (i.e. the reconstructed MRCA) may not be achieved using the sister taxon. Logical rules of inference (e.g. phylogenetic argumentation) do not help determine if the ingroup trajectories have resulted in unique synapomorphy in the 'right characters' (i.e. the same set of characters that have also remained plesiomorphic in the sister taxon). Evolution itself will have determined the plesiomorphy content of available putative outgroups, and it is clear that the selection of an outgroup can influence plesiomorphy estimates, the placement of the root, and the ingroup tree topology (Milinkovitch *et al.*, 1996; Milinkovitch & Lyons-Weiler, in press.

Optimal outgroup choice requires a search among all possible combinations of putative outgroup lineages. Some candidates will be too far out, making the correct inference of the ingroup–outgroup node difficult, while others may actually be members of the ingroup, leading to errors in the inference of clade convexity. The optimal combination indicated by the RASA criterion may consist of any number of outgroup lineages, with a limitation imposed by the reduction in effective characters, reflecting the increased probability of convergence with a finite number of character-states (see *Limitations*, below). When an optimal outgroup is found, no assertion is made regarding the plesiomorphy of any particular character; this is left to the method of tree estimation. The goal is to screen among available outgroup candidates to reduce the probability of causing errors in the ingroup topology.

When to use a rooted RASA

There are three situations in which this approach to optimal outgroup analysis should be applied:

Situation 1. When a single outgroup is available

A comparison of the amount of signal in the rooted analysis to the unrooted analysis of the ingroup alone will determine whether or not the available outgroup might lead to a reasonable estimate of plesiomorphy in the ingroup character states. If the putative outgroup is truly nested somewhere in the ingroup, some synapomorphies will be erroneously constrained in the analysis, resulting in decreased signal. RASA can therefore test the assumption of ingroup monophyly as well as the utility of rooting with an outgroup that may be too far out, although delineating between these two alternatives may prove to be difficult.

Situation 2. When a set of putative outgroup candidates are selected prior to the measurement of signal

As shown above, there is no efficient general rule of thumb for justified use of particular outgroups. However, signal measurements made in particular instances can be used as a basis for the selection among available outgroup taxa. The set of one or more outgroup taxa that results in the greatest signal is selected because it should allow the best possible estimate of plesiomorphy for the ingroup. Of course, the suitability of any single outgroup taxa may be modified by the inclusion or exclusion of other outgroup taxa.

Situation 3. When no useful outgroup is available and a subset of the ingroup may effectively root the rest of the ingroup taxa

This type of analysis might be applied to determine which among a set of ingroup taxa are most basal and would lead to an effective estimate of plesiomorphy for the rest. Unfortunately, it is possible for this application to lead to increased error in the placement of the root. Under the conditions shown in Figure 9, optimal

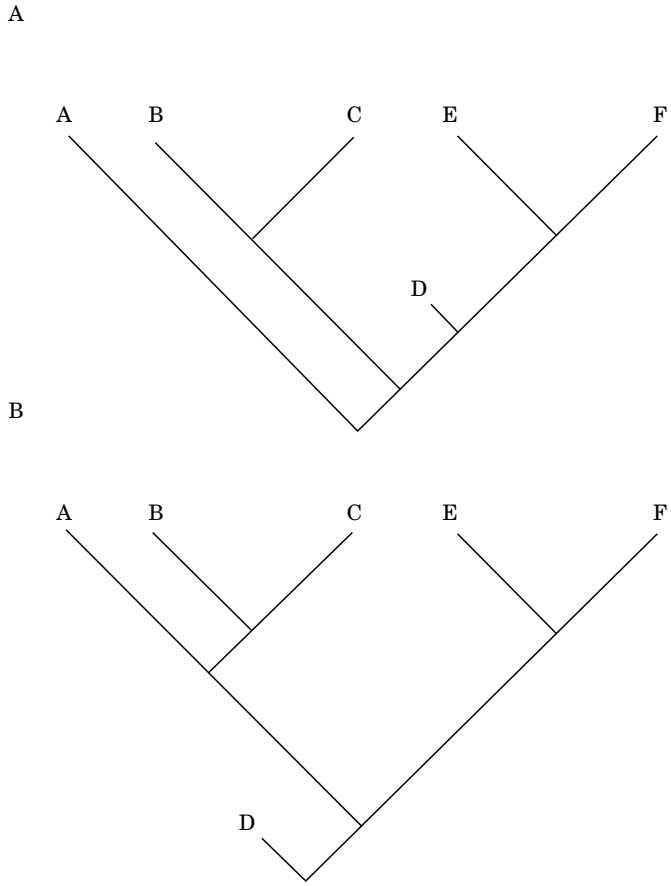

Figure 9. The taxon that might permit the most accurate estimate of plesiomorphy for the ingroup (taxa B–F in A) might actually be an ingroup taxon (taxon D), instead of the sister taxon (taxon A). Unless rates of evolution are known to be constant (they usually are not), the use of RASA to identify outgroup taxa without an *a priori* delineation of an outgroup taxon may be hazardous, resulting in erroneous inferences of monophyly (B).

outgroup analysis will correctly identify taxon D as exhibiting the greatest amount of plesiomorphy. However, taxon D does not derive uniquely and directly from the root. Similarly, a set of taxa that would lead to the best estimate of plesiomorphy may not constitute a monophyletic group. Therefore, although the best estimate of plesiomorphy would be affordable by some set of ingroup taxa, the placement of the root is uncertain in this situation. This need not, however, lead to difficulties with the eventual assessment of character polarities. The resulting topology would be an unrooted tree with polarized character states.

An additional advantage of using t_{RASA} as a criterion to screen out poor outgroup candidates is that the test is sensitive in an informative way to the effects of long branches (Lyons-Weiler & Hoelzer, 1997). It is sometimes considered that any taxon that can be placed outside the ingroup as a suitable outgroup. However, in some instances, the evolutionary distance between the putative outgroup and the ingroup

will be so great that little evidence of ancestry remains, and the result may be branch attraction. When long branches obscure and distort the phylogenetic structure in a matrix of ingroup character states, the effect is to decrease t_{RASA}. Such problematic taxa, including inappropriate outgroups, can be identified in a graphical diagnostic called the taxon variance plot (Lyons-Weiler & Hoelzer, 1997). Paradoxically, outgroups for which the phylogenetic relationship with ingroup taxa has been obscured by anagenetic evolution may nevertheless provide suitable estimates of ingroup plesiomorphy for the set of relevant characters. The information about plesiomorphy in such outgroup taxa may nevertheless be useful when estimating an unrooted tree for the ingroup alone (minus the outgroup) in which the outgroup states are assigned as ancestral, thus potentially avoiding long branch attraction. Trees achieved in this manner would be unrooted with polarized character states. We recommend preferring outgroup taxa that both increase t_{RASA} in the rooted analysis, and are not indicated as long edge taxa in the taxon variance plot (Lyons-Weiler & Hoelzer, 1997) in an unrooted RASA that includes both the ingroup and outgroup taxa.

Limitations

Rooted RASAs constrain outgroup character states throughout the ingroup. Occasionally, non-plesiomorphic states will be constrained in a rooted RASA (e.g. when an outgroup character has converged on an ingroup character state), even when signal is increased overall. However, such errors are clearly preferable to asserting false synapomorphy within the ingroup, which are misleading. Therefore, no certainty regarding plesiomorphy of any particular character is asserted.

One effect of rooting in this way is a reduction in the number of character states for each character by one if the outgroup has one state per character. This causes a reduction in the effective number of characters, because a taxon sharing a state with an outgroup will lose its representation for that character. Typically, when an outgroup comprised of multiple taxa is being evaluated, multiple states will be constrained for some characters. This further reduces the number of residual effective characters. Although the numerical values of **RAS** and **E** are reduced (i.e. the problem is scaled down), the effect may either be a decrease or an increase in signal, depending on the net effects on hierarchical structure in the character-state matrix, which in turn depends upon the juxtaposition of the outgroup states to those in the ingroup. The matrix reduction decreases signal when too many informative character states been eliminated, which becomes increasingly likely when the number of informative characters decreases. For matrices with relatively few characters, therefore, marginal decreases in t_{RASA} may therefore be ambiguous. Determination of the effect of character sampling effort on signal (power analysis) may help delineate cases in which the decrease is beyond that which may be caused by increasing the effects of sampling variance. As long as the criterion of increased signal is followed, any decrease in signal owing to reduced effective character number will be offset by the increase in signal due to reduced conflation of synapomorphy with plesiomorphy.

Increases in the number of outgroup taxa may eventually lead to the constraining of the entire ingroup matrix, and signal will appear to nil. This is clearly not an optimal outcome. When more than one outgroup is specified, the trade-offs among the amounts of synapomorphy, plesiomorphy and homoplasty constrained in the

ingroup comparisons become considerably more complex. In such cases, estimates ('reconstructions') of the MRCA of the outgroup and ingroup may be used to test the suitability of that node for rooting the ingroup (and, conveniently, the outgroup). Rooted RAS analyses that reveal a decrease in signal when estimated ancestral states are assumed to be plesiomorphic would provide sufficient warning that the tree-based phylogenetic analysis may have rendered a misleading result.

Software to perform RASA and optimal outgroup analysis is available via the world-wide-web (Lyons-Weiler, 1998).

The future of outgroup comparisons

Other criteria should be evaluated to see if they track the decaly in plesiomorphy content as well, or better than t_{RASA}. Ways to synthesize matrix reduction and tree topology estimation are currently being explored. We also foresee a synthesis between cladistic analyses and maximum likelihood approaches. For instance, it should be possible, and we suspect rather fruitful, to define models of nucleotide evolution that would explain the distributions of apparently non-plesiomorphic states that could be evaluated using likelihood scores. Due to the errors that would follow a spurious placement of the root, the development of such methods to further test hypotheses of directed graphs may be very useful.

ACKNOWLEDGEMENTS

We thank M. Milinkovitch for thorough analysis of RASA rooting. Robert Espinoza is thanked for insightful comments and reiterated calls for clarification and specification. An anonymous reviewer provided useful comments. In our rendering of this paper, we have learned far more about the challenges of dealing with ingroups than applies to phylogenetic inference. David Cannatella, Dan Faith, and John Trueman are thanked for comments on an earlier draft.

REFERENCES

Archie JW. 1996. Measures of homoplasy. In: Sanderson MJ, Hufford L, eds. *Homoplasy: the recurrence of similarity in evolution*. Academic Press, 153–188.

Arnold EN. 1981. Estimating phylogenies at low taxonomic levels. *Zeitschr. Zool. Syst. Evolutionsforsch.* **19:** 1–35.

Camin JH, Sokal RR. 1965. A method for deducing branching sequences in phylogeny. *Evolution* **19:** 311–326.

Cracraft J, Helm-Bychowski K. 1991. Parsimony and phylogenetic inference using DNA sequences: some methodological strategies. In: Miyamoto MM, Cracraft J, eds. *Phylogenetic Analysis of DNA Sequences*. New York: Oxford University Press, 184–220.

Donoghue MJ, Cantino PD. 1984. The logic and limitations of the outgroup substitution approach to cladistic analysis. *Systematic Botany* **9:** 192–202.

Eldredge N, Cracraft J. 1980. *Phylogenetic patterns and the evolutionary process*. New York: Columbia University Press.

Farris JS. 1972. Estimating phylogenetic trees from distance matrices. *American Naturalist* **106:** 645–66.

Farris JS. 1982. Outgroups and parsimony. *Systematic Zoology* **31:** 328–334.

Felsenstein J. 1978. Cases in which parsimomy and compatibility methods will be positively misleading. *Systematic Zoology* **27:** 401–410.

Garey M, Johnson DS. 1979. *Computers and intractability: a guide to the theory of NP-Completeness*. New York: WH Freeman.

Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ. 1991. Experimental phylogenetics: generation of a known phylogeny. *Science* **255:** 589–591.

Lyons-Weiler J. 1988. *RASA 2.2 Manual for the Macintosh*. http://loco.biology.unr.edu/archives/rasa/rasa.html

Lyons-Weiler J, Hoelzer GA, Tausch RJ. 1996. Relative Apparent Synapomorphy Analysis (RASA) I: the statistical measurement of phylogenetic signal. *Molecular Biology and Evolution* **13:** 749–757.

Lyons-Weiler J, Hoelzer GA. 1997. Escaping from the Felsenstein Zone by detecting long branches in phylogenetic data. *Molecular Phylogenetics and Evolution* **8:** 375–384.

Lyons-Weiler J, Milinkovitch M. 1997. A phylogenetic approach to the problem of differential lineage scoring. *Molecular Biology and Evolution* **14:** 968–975.

Lyons-Weiler J, Tausch RJ. 1996. The demarcation of historical from ecological variance in species patterns. In: *Conservation and Management of Oregon's Native Flora*. Native Plant Society of Oregon and Oregon State University, 209–221.

Maddison WP, Donoghue MJ, Maddison DR. 1984. Outgroup analysis and parsimony. *Systematic Zoology* **33:** 83–103.

Milinkovitch MC, LeDuc RG, Adachi J, Farnir F, Georges M, Hasegawa M. 1996. Effects of character weighting and species sampling on phylogeny reconstruction: A case study based on DNA sequence data in cetaceans. *Genetics* **144:** 1817–1833.

Milinkovitch MC, Lyons-Weiler J. (in press). Finding optimal ingroup topologies and convexities when putative outgroup candidates are not obvious. MPE Special Issue, 'Molecules and Morphology in Systematics', *Molecular Phylogentics and Evolution.*

Mooi R. 1989. The outgroup criterion revisited via naked zones and alleles. *Systematic Zoology* **38:** 283–290.

Nelson GJ, Platnick NI. 1981. *Systematics and biogeography: cladistics and vicariance*. New York: Columbia University Press.

Nixon KC, Carpenter JM. 1993. On outgroups. *Cladistics* **9:** 413–426.

Smith AB. 1994. Rooting molecular trees: problems and strategies. *Biological Journal of the Linnean Society* **51:** 279–292.

Stevens PF. 1980. Evolution of polarity of character states. *Annual Review of Ecology of Systematics* **11:** 333–358.

Watrous LE, Wheeler QE. 1981. The out-group comparison method of character analysis. *Systematic Zoology* **30:** 1–11.

Wheeler QD. 1981. The ins and outs of character analysis: a response to Crisci and Stuessy. *Systematic Botany* **6:** 297–306.

Wheeler WC. 1990. Nucleic acid sequences phylogeny and random outgroups. *Cladistics* **6:** 363–368.

Wiley EO. 1981. *Phylogenetics: the theory and practice of phylogenetic systematics*. New York: John Wiley.