Syst. Biol. 55(2):314–328, 2006

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150500541730

Multiple Sequence Alignment Accuracy and Phylogenetic Inference

T. HEATH OGDEN AND MICHAEL S. ROSENBERG

Center for Evolutionary Functional Genomics, The Biodesign Institute, and the School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501, USA; E-mail: heath_ogden@asu.edu, msr@asu.edu

Abstract.—Phylogenies are often thought to be more dependent upon the specifics of the sequence alignment rather than on the method of reconstruction. Simulation of sequences containing insertion and deletion events was performed in order to determine the role that alignment accuracy plays during phylogenetic inference. Data sets were simulated for pectinate, balanced, and random tree shapes under different conditions (ultrametric equal branch length, ultrametric random branch length, nonultrametric random branch length). Comparisons between hypothesized alignments and true alignments enabled determination of two measures of alignment accuracy, that of the total data set and that of individual branches. In general, our results indicate that as alignment error increases, topological accuracy decreases. This trend was much more pronounced for data sets derived from more pectinate topologies. In contrast, for balanced, ultrametric, equal branch length tree shapes, alignment inaccuracy had little average effect on tree reconstruction. These conclusions are based on average trends of many analyses under different conditions, and any one specific analysis, independent of the alignment accuracy, may recover very accurate or inaccurate topologies. Maximum likelihood and Bayesian, in general, outperformed neighbor joining and maximum parsimony in terms of tree reconstruction accuracy. Results also indicated that as the length of the branch and of the neighboring branches increase, alignment accuracy decreases, and the length of the neighboring branches is the major factor in topological accuracy. Thus, multiple-sequence alignment can be an important factor in downstream effects on topological reconstruction. [Bayesian; maximum likelihood; maximum parsimony; multiple sequence alignment; neighbor joining; phylogenetics; simulation; tree reconstruction.]

Multiple-sequence alignment is an important tool in biological research and may be used for a variety of purposes ranging from secondary structure identification (Coventry et al., 2004; Dowell and Eddy, 2004; Holmes, 2005a; Knudsen and Hein, 1999), noncoding functional RNA (ncRNA) detection (di Bernardo et al., 2003; Rivas and Eddy, 2001), and phylogenetic inference. Although the overall goal of phylogenetic analysis is to most accurately infer the relationships of the terminal taxa given the data, little attention has been given to the role that alignment error may play in tree reconstruction. It has been concluded that the resulting phylogeny may be more dependent upon the methods of alignment than on the mode of phylogenetic reconstruction (Cammarano et al., 1999; Hwang et al., 1998; Kjer, 1995, 2004; Lake, 1991; Morrison and Ellis, 1997; Mugridge et al., 2000; Ogden and Whiting, 2003; Thorne and Kishino, 1992; Titus and Frost, 1996; Xia et al., 2003). Given the presumed importance of accuracy of multiple-sequence alignments, it is surprising that very few studies have specifically addressed this issue. As Hall (2005) writes, "It is a truism that the quality of a tree is no better than the quality of the alignment used to estimate that tree."

In order to determine the accuracy of an estimate or hypothesis, one must know the truth. Numerous studies have used simulated fixed data sets (usually with no insertions or deletions) to examine topological accuracy of phylogenetic reconstruction methods (e.g., Hillis, 1995; Huelsenbeck and Rannala, 2004; Nei, 1996; Rosenberg and Kumar, 2003; Takahashi and Nei, 2000, just to name a few). The typical modus operandi in simulation studies is to generate a fixed alignment created from the true tree. Then, different tree-building methodologies are used to reconstruct hypothesized trees, and finally, these are compared to the true tree to evaluate topological accuracy. Although this process may enable the comparison of different tree-building methods and models, it says nothing about the effect that alignment error may contribute. Hall (2005) overcomes some of these deficiencies by introducing insertion and deletion events into simulated alignments in order to make the data sets more biologically realistic. Nevertheless, his analysis does not take advantage of a comparison between the true alignments and the hypothesized alignments. Furthermore, all of the true trees he used during simulation were "strictly bifurcating, cladistically symmetric" trees (Hall, 2005); pectinate tree shapes and their effects and interactions with alignment were not examined. Recent studies (Keightley and Johnson, 2004; Pollard et al., 2004) have simulated alignments with insertions and deletions in order to compare and benchmark different alignment methods and approaches, yet none of these studies has examined phylogenetic accuracy. In a similar study (Rosenberg, 2005a), the relationship of pairwise sequence alignment and evolutionary distance was investigated. It was shown that "when sequence identity exceeded 80%, essentially all aligned sites (>99%) were truly homologous . . . [and] As identity declined, the proportion of correctly aligned sites rapidly decreased." Notwithstanding these latest contributions, the question of how various phylogenetic methods respond to alignment error remains open.

Multiple sequence alignment is a procedure to convert sequences of unequal length into sequences of equal length by inferring the placement of gaps, with the goal to infer homology among characters (note, however, that sequences of equal length may also require alignment). Insertion and deletion events (indels) are treated in a variety of ways during multiple-sequence alignment and phylogenetic reconstruction. When sequences require alignment, the investigator is obligated to decide how he or she accounts for insertions, deletions, and mutations.

The question then of which A's, T's, C's, G's, and indels to compare becomes fundamental in DNA systematic analysis (Wheeler, 2001). Indels may be treated as an additional residue (gap characters) in a substitution matrix. Treating gaps as an extra character in a substitution matrix is essentially equivalent to explicitly or implicitly assuming a linear gap cost model, although more complicated weighting schemes allow for different gap initiation and gap extension costs. Alternatively, probabilistic evolutionary models for insertion and deletions may be used, such as HMMs or SCFGs (Metzler, 2003; Miklos et al., 2004; Thorne et al., 1991, 1992) or gaps may be treated using Felsenstein wildcards (Holmes and Bruno, 2001). Although the underlying mechanisms and frequencies of indels is not understood as well as base substitutions processes (Hall, 2005), efforts to remedy this lack of models and method are underway (Holmes, 2003, 2004, 2005b; Holmes and Bruno, 2001; Knudsen and Miyamoto, 2003; Mitchison and Durbin, 1995; Mitchison, 1999). We recognize that new methods that combine the alignment and tree reconstruction processes have been suggested (Fleissner et al., 2005; Lunter et al., 2005; Redelings and Suchard, 2005; Wheeler, 2001), but this study will not address these ideas.

Independent of the means by which multiple sequence alignments are generated, they are in their simplest form statements of putative homology or "primary homology" (de Pinna, 1991; Phillips et al., 2000). Only after subjecting these primary homologies to a test (the reconstructed topology), secondary homologies, or what is usually termed homologous characters, may be inferred. Thus, homologous features can be identified when their origins are traced to a transformation on a branch leading to the most recent common ancestor (Ogden et al., 2005). Character transformation ratios (base substitutions and indels) are generally not directly measurable and can only be inferred or estimated from predetermined phylogenetic patterns. This produces a problem in phylogenetic analysis: that the "interaction between the specification of values a priori and their inference a posteriori" is circular in nature (Wheeler, 1995), accentuating the need for a better understanding of the effects that alignment inaccuracies may contribute to topological reconstruction.

The objectives of this study are: (1) simulate noncoding DNA alignments with indels and compare true alignments to hypothesized alignments through the calculation of alignment accuracy scores; (2) examine the relationship between alignment accuracy and topological accuracy under different methods of tree reconstruction (neighbor joining, parsimony, likelihood, and Bayesian); and (3) investigate the interaction of alignment accuracy with tree shape (length and branching pattern).

Material and Methods

Data Simulation

We simulated data sets for seven 16-taxon topologies under a variety of different conditions in order to cover a reasonable amount of the error space representing alignment inaccuracy. We believe that 16 terminals are sufficient to provide reasonable tree shape diversity and complexity in order to investigate the effects of alignment inaccuracies and tree reconstruction, while at the same time not requiring enormous amounts of computational time to perform reasonable searches under different reconstruction methods (particularly likelihood and Bayesian). The seven topologies (Fig. 1) consisted of a balanced tree, a pectinate tree, and five random trees (A to E) generated under a Yule model in Mesquite (Maddison and Maddison, 2004). The relative branch lengths of each topology were set under 11 different conditions: ultrametric equal branch length, ultrametric random branch length (five sets), and nonultrametric random branch lengths (five sets). Each of these 11 conditions was scaled such that the maximum evolutionary distance between a pair of sequences was equal to 1.0 or 2.0. Thus, each of the seven topologies was used to create 22 model trees (Fig. 2). All simulations were conducted under identical conditions using MySSP (Rosenberg, 2005c). The initial sequence length was 2000 base pairs. For this study, many potentially variable parameters were held constant in order to gain simplicity. Thus, aside from the different conditions explained above, DNA evolution was simulated under the Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al., 1985), with transition-transversion bias $\kappa = 3.6$ (Rosenberg and Kumar, 2003) and initial and expected base frequencies of A and T = 0.2; and G and C = 0.3.

Insertion and deletion events were modeled as a Poisson process, following Rosenberg (2005a). Expected numbers of insertions and deletions (modeled separately) for a given branch were determined as a function of the realized number of substitutions (itself a Poisson process) that occurred on that branch. Expected rates were based on observed values from primates and rodents, with one insertion event for every 100 substitutions and one deletion event for every 40 substitutions (Ophir and Graur, 1997). The realized number of insertion and deletion events was drawn from a Poisson distribution with mean equal to the expectation. The actual size of each insertion and deletion event was independently determined from a truncated (so as not to include zero) Poisson distribution with mean equal to four bases (as observed in primates and rodents) (Ophir and Graur, 1997; Sundstrom et al., 2003).

Each simulation was replicated 100 times. The fate of every insertion and deletion event was tracked throughout the simulations, such that the columns, including those with gaps in the final alignment, represented the true homologies (Rosenberg, 2005a).

Alignment

These simulations resulted in 15,400 unique data sets (alignments) containing gaps representing either insertion or deletion events during the simulation process, and will be referred to as the True Alignments (TA). Each of the TA were then stripped of their gaps and were realigned via ClustalW version 1.83 (Thompson

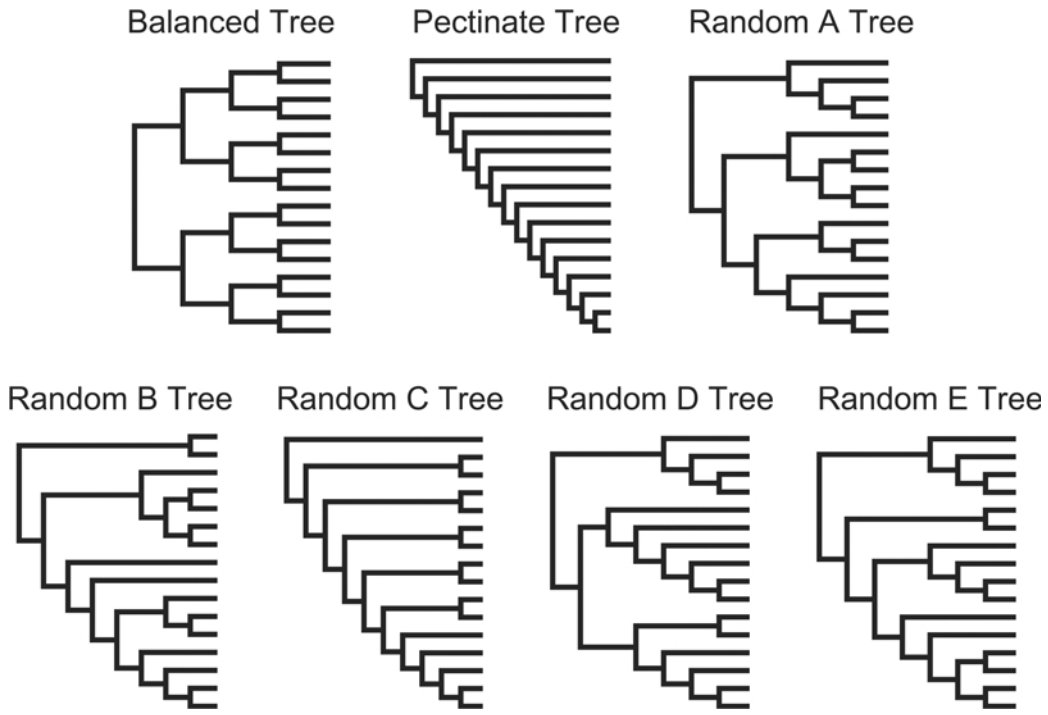

FIGURE 1. The seven topologies used to explore the effect of tree shape on alignment accuracy and tree reconstruction consisted of a balanced tree, a pectinate tree, and five random Yule trees (A–E).

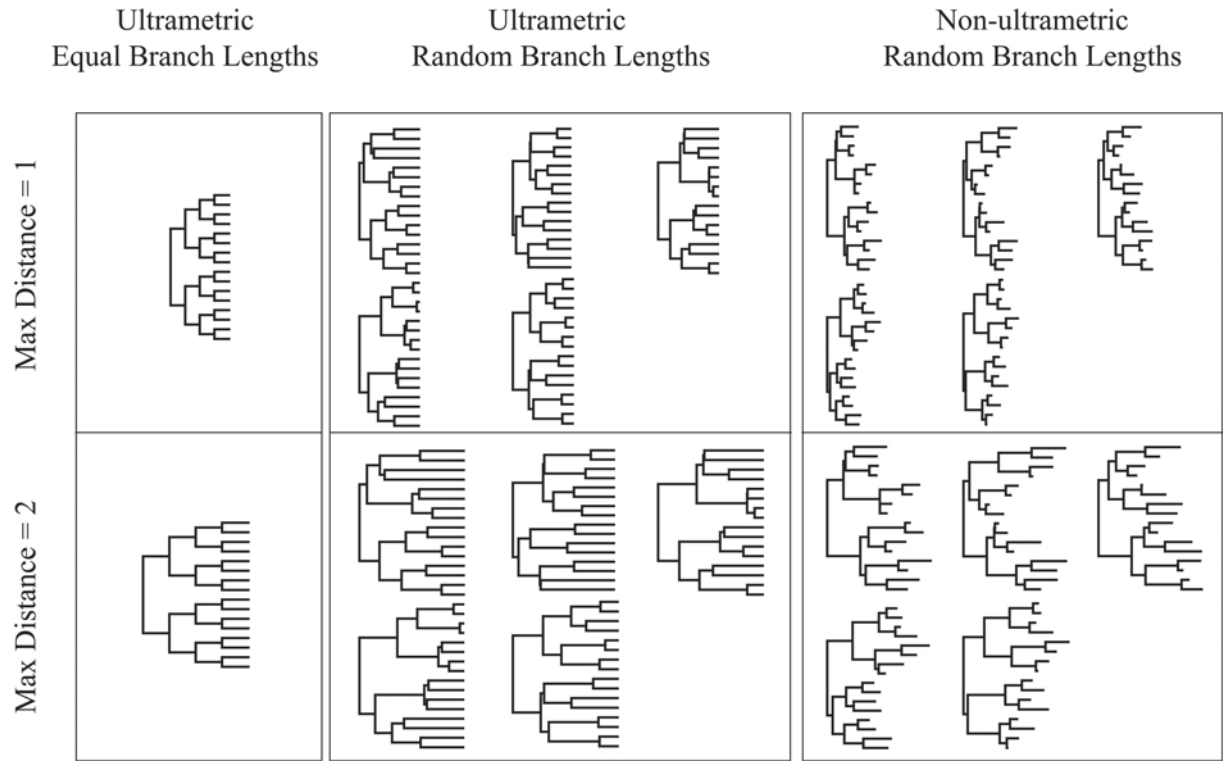

FIGURE 2. An example of the 22 model trees for the balanced tree shape, consisting of ultrametric equal branch length, ultrametric random branch length (five sets), and nonultrametric random branch lengths (five sets). Each of these 11 conditions was scaled such that the maximum evolutionary distance between a pair of sequences was equal to 1.0 or 2.0.

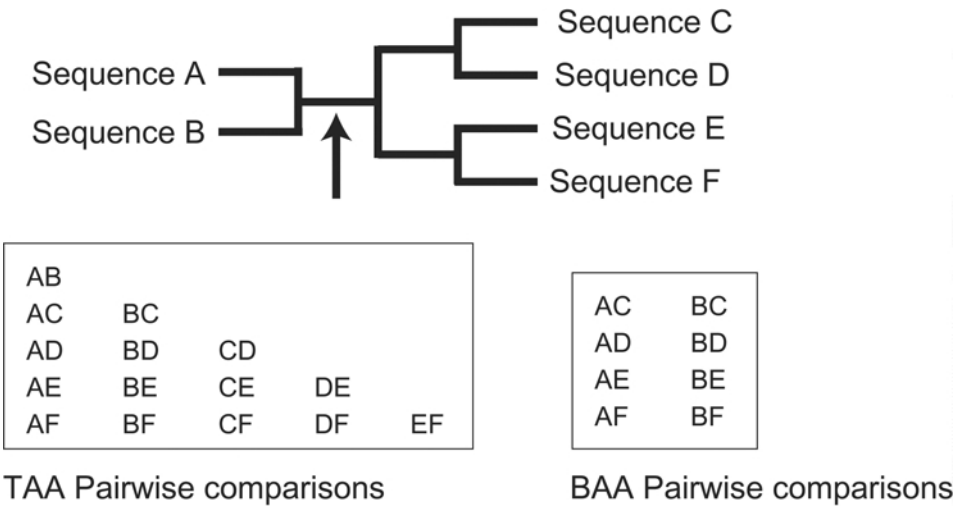

FIGURE 3. The pairwise comparisons (given the example six-taxon tree) that would be used to calculate the TAA and BAA values.

et al., 1994) using default parameters. We will refer to these alignments as the Hypothesized Alignments (HA). Although one could examine the resulting topological effects of varying parameters in ClustalW (Ogden and Whiting, 2003), the focus of this study was not to try to estimate the optimal parameter settings that would generate an alignment and the most accurate reconstructed topology. Rather, we wanted to produce a reasonable and realistic amount of alignment error across the alignment inaccuracy space.

Alignment Accuracy

Alignment accuracy, calculated as the proportion of pairwise ungapped aligned sites that are truly homologous (Rosenberg, 2005a), was summarized by two different measures: (1) the Total Alignment Accuracy (TAA) score and (2) the Branch Alignment Accuracy (BAA) score. The TAA for a data set was calculated from the average accuracy of all pairwise sequence comparisons in the multiple alignment. For example, given a tree with six sequences denoted as A, B, C, D, E, and F, all possible pairwise comparisons would be averaged to calculate TAA (Fig. 3). Similarly, the BAA was calculated from the average of all pairwise comparisons that cross a particular branch. For example, the set of pairwise comparisons that would be averaged to calculate BAA for the branch separating A + B from the remaining taxa (indicated by the arrow) would be: AC, AD, AE, AF, BC, BD, BE, and BF (Fig. 3). It is important to realize that only the aligned sites in the pairwise comparisons are examined; any site consisting of a nucleotide that is aligned with a gap, or a gap with a gap, is not included in the pairwise score.

Tree Reconstruction Analyses

Each of the data sets (15,400 TA and 15,400 HA) were analyzed under the four most widely used strategies for phylogenetic tree reconstruction: neighbor joining (NJ), maximum parsimony (MP), maximum likelihood (ML), and Bayesian (B) using PAUP* version 4.b10 Windows (Swofford, 2002) and MrBayes version 3.0b4 (Huelsenbeck and Ronquist, 2001). We were interested in looking at the effects of alignment error on reconstruction accuracy by comparing the TA tree reconstructions to the HA tree reconstructions. Thus, it is not the purpose of this study to try to optimize any specific parameters during the tree reconstruction phase. The crucial point is that both the TA and HA be analyzed identically under the different tree-building methods, allowing for direct comparisons. Analyses performed under NJ, ML, and B were implemented under the HKY model and other default settings. For the likelihood analyses, transition/transversion ratios were estimated, nucleotide frequencies were assumed from empirical frequencies, and distribution of rates at variable sites was set to equal. In MP, the analyses carried out consisted of 100 random additions with TBR swapping and all other default settings. When multiple trees were recovered using MP or (rarely) ML, the strict consensus of these trees was used as the result. For the B analyses, 100,000 generations were performed sampling every 100 generations, and the first 250 trees were then discarded as the burn-in. A majority-rule consensus topology of the remaining 750 trees was constructed (nodes that were present in at least 50% of the topologies were retained) and saved as the resultant topology for each B analysis. In summary, each of the 15,400 TA and the 15,400 HA were analyzed identically for each of the tree building methodologies and the resulting topologies (consensus in some cases) were used to compare the TAA and BAA measures to topological accuracy.

Each reconstructed tree was compared to the true model tree using the Robinson-Foulds (1981) measure to estimate topological accuracy; these are referenced as TA_{dist} and HA_{dist}, respectively, for the TA and HA data sets. The difference between these values ($HA_{dist} - TA_{dist}$) therefore represents the difference in

TABLE 1. Mean, median, maximum, and minimum TAA values for each of the different tree shapes.

	Balanced	Pectinate	Random A	Random B	Random C	Random D	Random E
Mean	0.720	0.781	0.726	0.716	0.761	0.727	0.722
Maximum	0.966	0.978	0.976	0.965	0.960	0.973	0.965
Minimum	0.191	0.365	0.173	0.182	0.330	0.229	0.217
Median	0.815	0.844	0.818	0.809	0.816	0.800	0.801

topological accuracy of trees reconstructed from the true and hypothesized alignments; this value is referred to as the Tree Distance Difference (TDD). When the TA tree is topologically more accurate than the HA tree, TDD will be a positive number; if TDD is negative, the HA tree is more accurate that the TA tree. Note that TDD is not itself a measure of topological accuracy, but rather a comparison of the accuracies of the TA tree and HA tree reconstructions. Hence, TA and HA trees could both be completely accurate, with a distance to the true tree of 0, and thus, a TDD equal to 0. Alternatively, they could both be equally inaccurate, with large distances relative to the true tree, and again TDD may also be 0 (the reconstructed trees could be completely different, but also completely wrong).

RESULTS

Results from the numerous analyses can be looked at in many different ways. In order to simplify, we have broken down the main results to the two methods of calculating alignment accuracy.

TAA (Total Alignment Accuracy)

TAA values (Table 1) ranged from a minimum of 0.173 to a maximum of 0.978, with a mean of 0.736 across all shapes. The pectinate tree (0.781) and random C tree (0.761) had higher TAA means (although not medians) than the remaining more balanced topologies. This result at first seems misleading, because one would expect balanced trees to produce more accurate alignments than pectinate trees. However, we know that alignment accuracy is largely dependent on the distances among sequences (Rosenberg, 2005a) and the pectinate trees in this study have fewer long-distance pairs (pairs that cross the root) and more short-distance pairs than balanced trees due to the manner by which the trees were scaled to similar maximum depth. Because the mean pairwise distance among taxa is smaller for the pectinate trees than the balanced trees, the accuracy of all possible pairwise alignments is greater in the pectinate case. It should be noted that the accuracies of the most distant pairs in pectinate trees is less than that of balanced trees because balanced trees lead to more accurate alignments of distant sequences (Rosenberg, 2005b); this emphasizes the contrast between examining specific pairs of sequences and entire data sets. The difference in mean TAA between alignments simulated under a max distance of 1 and 2 is 0.370. In other words, the doubling of the evolutionary distance caused an absolute average decrease in alignment accuracy of 37% (a relative decrease in accuracy of 29%). These results clearly indicate that the simulations produced alignment error across the reasonable large majority of the realistic alignment space (detailed results are found in the online appendix, Table A1, at http://systematicbiology.org).

In order to evaluate possible trends and relationships, the results from the TAA calculations were graphed individually for each tree shape and across the different methods. When looking at each of the tree shapes, all 22 conditions (2 ultrametric equal branch length, 10 ultrametric random branch length, and 10 nonultrametric random branch length) were pooled (unpooled results are provided in the online appendix).

General trends in topological accuracy of the true and hypothesized alignments appear to be very similar at first glance (online appendix, Figs. A1 and A2). In general, pectinate trees were much more difficult to reconstruct than balanced trees, regardless of alignment accuracy. Directly comparing these topological accuracy plots for TA and HA can be difficult; we therefore concentrate most of our discussion on Tree Distance Difference (TDD). The relationship between TDD and TAA for the balanced tree simulations is shown in Figure 4. Points that are to the right have very accurate alignments and become less accurate as they move left, and points that are above the y-axis zero line are cases where the TA tree reconstructions were more accurate than the HA tree reconstructions.

It should be noted that, for any given accuracy, there are many points that fall on the zero line or above and below in a vertical spread over many parts of the graph. Points with negative TAA indicate replicates for which the HA led to more accurate tree reconstruction than the TA. Thus, the symmetric spread of points above and below TAA of zero indicates stochastic variation in tree reconstruction due to alignment difference. It should also be recognized that many points are superimposed upon one another. In order to summarize the average distribution of the points and to discern if any relationship exists between TAA and TDD, a moving average (based on an overlapping sliding window of 50 consecutive points) is shown on the graph.

Figure 4 shows the results of each tree reconstruction method for the balanced tree shape. As mentioned above, many of the points are superimposed; however, one can see that for the more accurate alignments there is a balanced spread of data points above and below the zero-TDD line (except for the neighbor-joining cases); as alignments become less accurate (moving left), the spread increases with an upward trend towards a higher TDD value. The moving average lines were nearly flat and essentially followed the zero-TDD line down to

about 55% alignment accuracy (moving from right to left) for all methods except NJ, which begins to rise at about 60% accuracy. TAA values for all methods below 55% show an increase in TDD, indicating that the TA tree reconstructions were more accurate than the HA tree reconstructions.

In contrast to the balanced tree shape, the pectinate tree shape (Fig. 5) resulted in an immediate increasing trend in TDD as alignment accuracy decreased. An initial peak is reached by MP, ML, and Baysian methods at about 85% alignment accuracy and then a small decline is seen until about 70% alignment accuracy. NJ presents an initial peak at about 78% accurate and then a similar increase is seen until about 70% alignment accuracy as well. Maximum parsimony appears to be less susceptible to large TDD values across the alignment error space than the other methods. This does not necessarily mean that MP reconstructs trees more accurately, only that the effect of less accurate alignments is not as great for MP as it is for ML and Bayesian. In fact, ML and Bayesian outperform MP using the TA data sets (see method comparison below) and therefore these methods have more to lose as alignments become less accurate. Moreover, NJ is more sensitive to alignment error, as seen by an initial increase in TDD with decreasing TAA, and also seems to be more affected by very inaccurate alignments (less that 50% accurate) than the other methods. Therefore, although pectinate trees, on average, contain less alignment error than more balanced topologies (Table 1), these errors have a larger effect on tree reconstruction then the same amount of error in a balanced tree.

In order to examine the random tree shapes and compare them to the balanced and pectinate shapes, we plot the moving average lines separated by method for each tree shape (Fig. 6). Because the same general trends are seen with respect to data point spread, we only show the moving average lines for these graphs. The random tree shape simulations follow the same general curve as the balanced tree, with the obvious exception of the random C tree (Fig. 1). This tree has a much more pectinate shape than the other random trees and therefore it is not surprising that its trend falls in between the pectinate tree and the remaining more balanced tree shapes. MP, ML, and B methods support this same basic result; the more pectinate-like a tree is, the larger the effect of alignment error on topological accuracy, particularly for alignments over 60% accurate. NJ, on the other hand, is slightly more sensitive to alignment error for all tree shapes relative to the other methods.

BAA (Branch Alignment Accuracy)

In order to examine the effect of BAA on topological reconstruction, we counted how many times the correct branch was identified in each of the 100 simulation replicates. We calculated the difference in the number of replicates that recovered the branch between TA and HA tree reconstructions ($TA_{rep} - HA_{rep}$). This number is generally positive; however, there are also a number of cases where the HA tree recovered a branch more often than the TA tree. Figure 7 depicts the resulting data points and moving averages for all of the different tree shapes and conditions pooled together, separated by the four methods of tree reconstruction. This graph represents an average across all trees. Except for a small jump around 87%, the TA and the HA data sets show little difference in branch reconstruction accuracy. Below about 70% alignment accuracy, a general trend of increase is seen in all the methods, with ML and B maximums of over 20 replicate differences at a BAA score of around 34%. However, similarly as above, MP appears to be less affected, as measured by the ($TA_{rep} - HA_{rep}$) value, for BAA values between 30% and 60%.

Discussion

Our results confirm many ideas concerning the affect of alignment accuracy on topological accuracy (Hall, 2005; Lake, 1991; Morrison and Ellis, 1997; Ogden and Whiting, 2003; Thorne and Kishino, 1992). For example, we find that alignment accuracy can have a profound effect on any one single data set. This is evidenced by the fact that many data points are found above and below the y-axis zero lines for identical or nearly identical alignment accuracy scores (Figs. 4 and 7). Therefore, any hypothesized alignment (whether correct homologies were recovered or not) may give you a topology that is very accurate, very inaccurate, or something in between. It is difficult to predict this on a case by case basis, but this study does confirm that the alignment can drive the resultant accuracy.

However, there are also many cases where one is no better or worse off with an inaccurate alignment. This does not mean that one is necessarily reconstructing the topology correctly, only that any hypothesized alignment may perform about the same as the true alignment. It also does not mean the true and hypothesized alignments are reconstructing the same trees. They could each get half of the branches wrong, but not the same half. This is apparent from our results because many data points fall essentially on the y-axis zero lines (Figs. 4 to 7). These results are not surprising because there are tree shapes and data sets that are inherently hard to reconstruct, and any hypothesized alignment may reconstruct the topology as accurately (or as inaccurately) as the unknown (for empirical data) true alignment. So although we know that alignment may have a huge downstream effect on topological accuracy, we also know that in some cases inaccurate alignments may have (on the average) no reasonable noticeable negative effect on tree reconstruction (stochastically, some inaccurate alignments produce better trees than the correct alignment). Thus, Hall's (2005) "truism" may be true in some cases, but there are also certainly cases where the quality of the tree is essentially independent from the quality of the alignment.

Despite the intricacies of the behavior of any one particular data set, we confirm that, in general, more accurate alignments give you more accurate topologies. This is demonstrated through the moving average

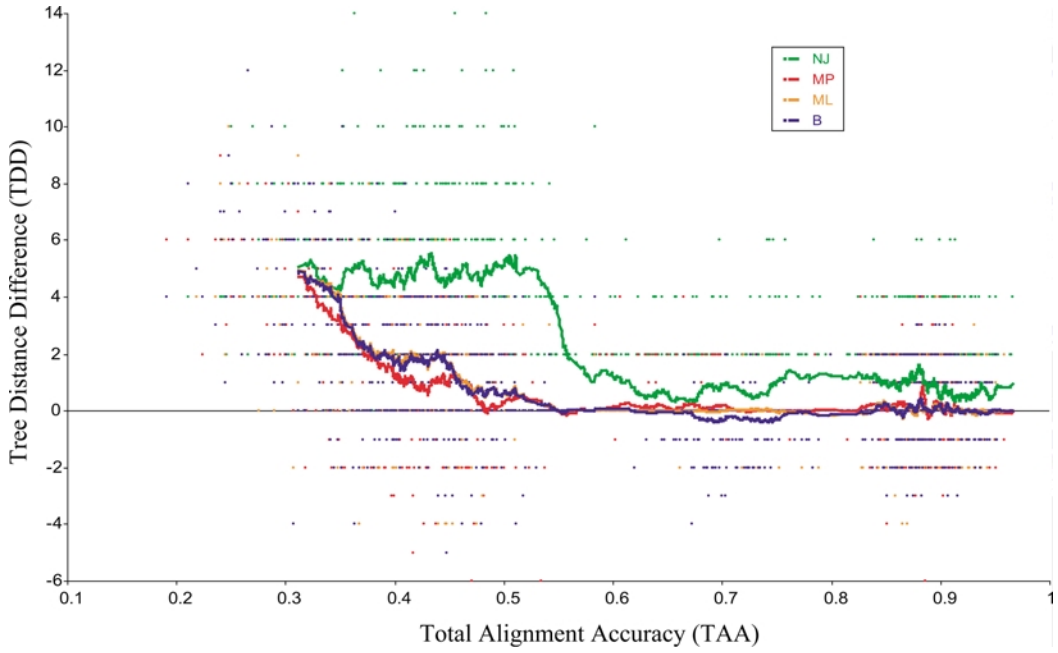

FIGURE 4. Relationship of Total Alignment Accuracy (TAA) and Tree Distance Difference (TDD) for balanced tree shape (all 22 model conditions pooled). Points to the far right are the most accurate alignments, whereas points to the left are the least accurate alignments. Points above the 0 TDD line are cases where the TA reconstructed tree was more accurate than the HA reconstructed tree, and the opposite is true for points below the 0 TDD line. Many points may be superimposed upon one another. The lines are moving averages based on an overlapping sliding window of 50 consecutive points. Note that the likelihood moving average line is essentially coincident and hidden by the Bayes line.

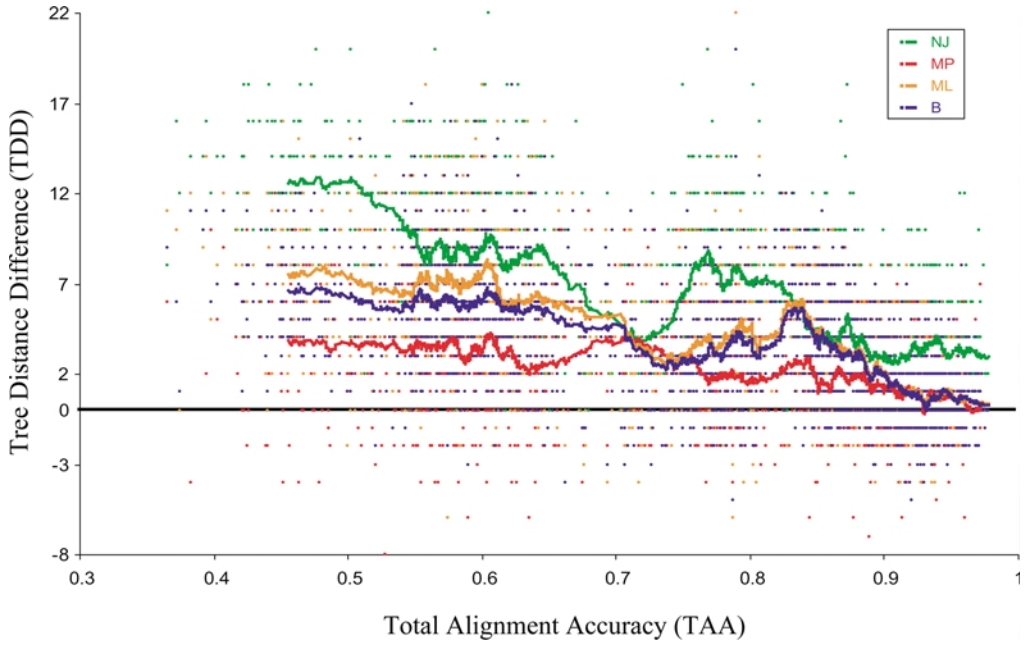

FIGURE 5. Relationship of Total Alignment Accuracy (TAA) and Tree Distance Difference (TDD) for pectinate tree shape (all 22 model conditions pooled). Points to the far right are the most accurate alignments, whereas points to the left are the least accurate alignments. Points above the 0 TDD line are cases where the TA reconstructed tree was more accurate than the HA reconstructed tree, and the opposite is true for points below the 0 TDD line. Many points may be superimposed upon one another. The lines are moving averages based on an overlapping sliding window of 50 consecutive points.

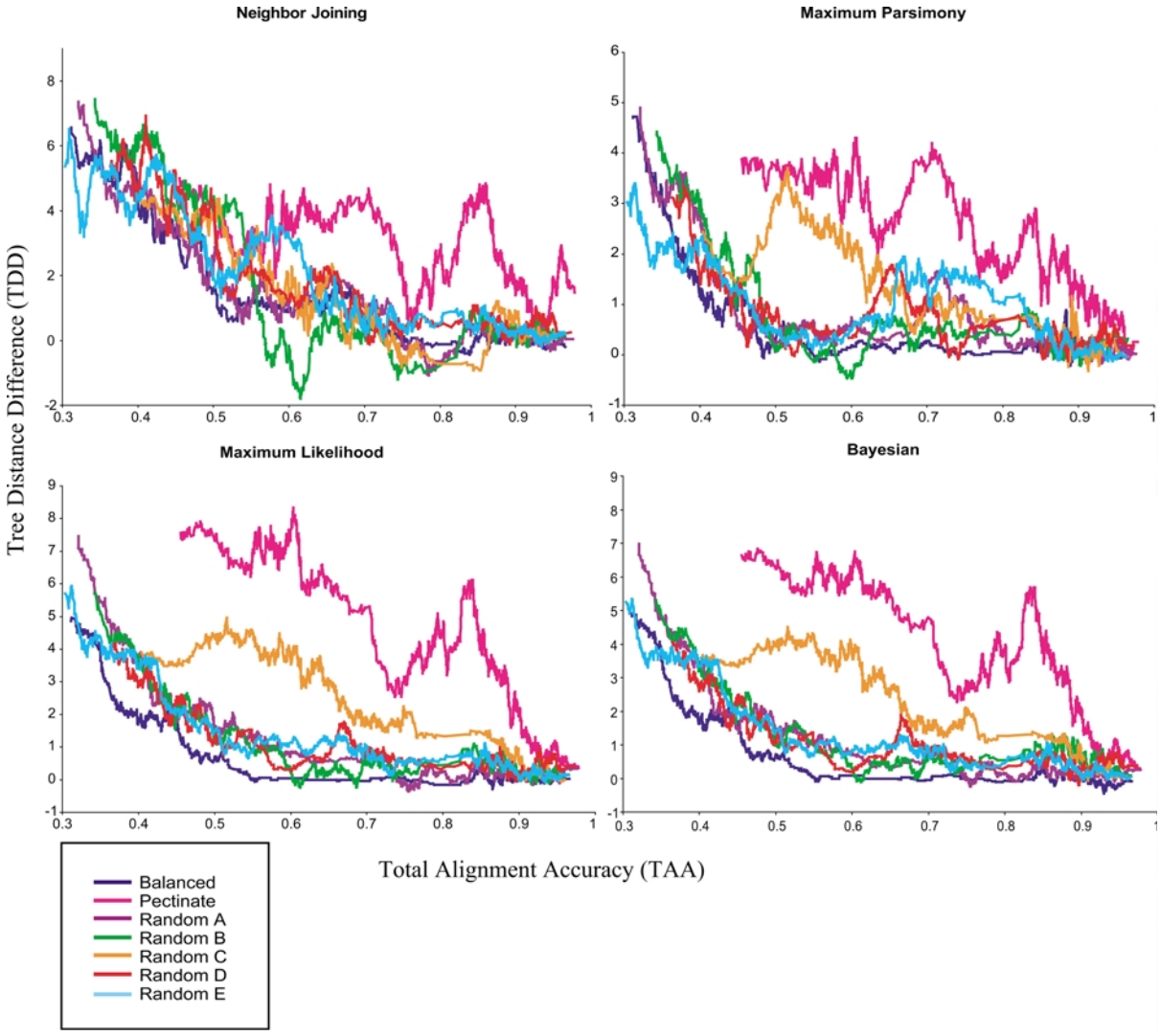

FIGURE 6. Moving average lines showing the relationship of Total Alignment Accuracy (TAA) and Tree Distance Difference (TDD) separated by the four methods of tree reconstruction and tree shape.

lines. Across more "realistic topologies" (i.e., not fully balanced or fully pectinate), as alignment error increases the TA reconstructions outperform the HA reconstructions. Although this notion is based on an average across all of the analyses performed (or subsets of the analyses), it can still be adhered to as a good rule of thumb. This result is not particularly surprising and is logically attractive; however, until this study this obvious assumption had never been formally tested.

Our results strongly demonstrated that balanced topologies are much less affected by alignment error than pectinate topologies. This trend is not surprising as balanced tree branch lengths tend not to be as long or short as pectinate tree branch lengths for trees of identical depth. These factors and maybe others not fully understood may elucidate questions as to why balanced tree shapes seem to be just easier to reconstruct than pectinate ones. As an aside, it has been suggested that certain methodologies or data sets are biased toward producing more pectinate trees (Colless, 1996; Harcourt-Brown et al., 2001; Heard and Mooers, 1996; Huelsenbeck and Kirkpatrick, 1996; Mooers et al., 1995), yet arguments exist against this idea as well (Farris and Kallersjo, 1998; Wenzel and Siddall, 1999), and future studies are needed to further investigate the role of methodological biases in alignment and tree reconstruction. Nevertheless, the degree to which the balanced trees were robust to alignment inaccuracy was unexpected. Essentially, alignments that were 50% inaccurate showed no average disadvantage as compared to the true alignments. It also calls to question an aspect of Hall's (2005) recent study; although he simulated sequences in an extremely realistic fashion, including indels and alignment, he only used strictly balanced tree topologies, which likely mitigated much of the effect of alignment on his results. For many cases, it

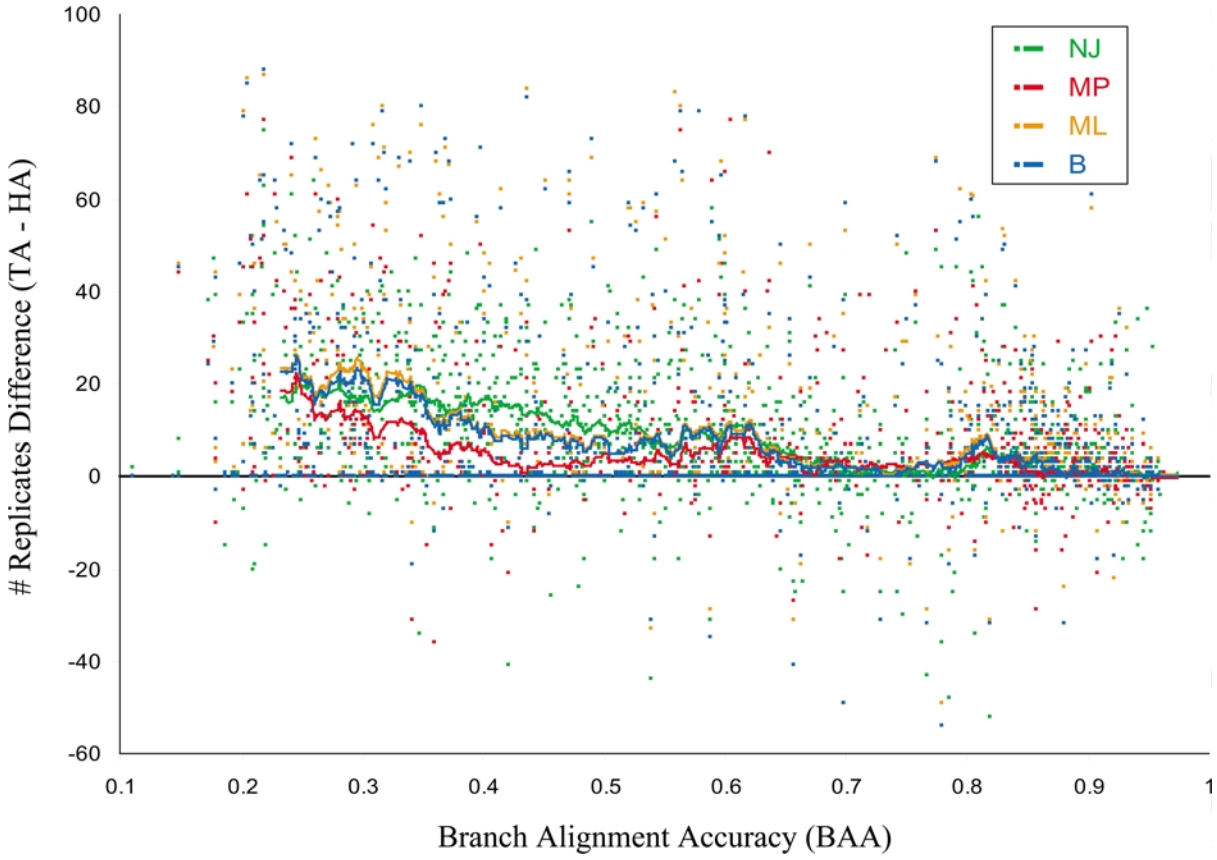

FIGURE 7. Relationship of Branch Alignment Accuracy (BAA) and number of replicates difference. Points to the far right are the most accurate alignments by branch, whereas points to the left are the least accurate alignments by branch. Points above the 0 line are cases where the TA reconstructed tree recovered the particular branch in more replicates than the HA reconstructed tree, and the opposite is true for points below the line. Many points may be superimposed upon one another. The lines are moving averages based on an overlapping sliding window of 50 consecutive points.

might not matter if your alignment is poor, and any of the available alignment programs may "do the job" well enough. However, it should cautiously be remembered that this conclusion is based on the average of many analyses, and for any one analysis it could matter a great deal. This is particularly applicable if one is interested in a specific relationship where branches are very short or very long, or the node of interest falls along a pectinate backbone (see Ogden and Whiting, 2003, for an empirical example).

The Indel Model

One potentially important issue in this study is the accuracy of the indel model used as the basis of our simulations. Although very simple, the model is not tremendously unrealistic, particularly for noncoding DNA. Insertions and deletions were independently modeled as Poisson processes, with frequency of occurrence on each branch based (indirectly) on the branch length and general rate parameters obtained from empirical studies (Ophir and Graur, 1997; Sundstrom et al., 2003). Although the decision to model insertion and deletion events separately was likely inconsequential to this study, it could have importance for future work because advances in multiple-sequence alignment have found advantages to treating them as separate processes (Löytynoja and Goldman, 2005). Unlike some commonly employed indel models (e.g., Thorne et al., 1991), in our simulations individual indel events were not restricted to single base pairs but were drawn from a size distribution. In this case, the Poisson distribution we used for indel sizes appears to be a poor fit to empirically derived size distributions estimated from entire genome alignments. (Chimpanzee Sequencing and Analysis Consortium, 2005); however, it should be noted that this and other empirically determined patterns of indel size (from pairwise comparisons of mammalian genomes) cannot easily be modeled by any standard theoretical distribution. Despite the limitations and simplicity of our model, the produced alignment accuracies are very similar to those found by other researchers using alternate indel models (Keightley and Johnson, 2004; Pollard et al., 2004).

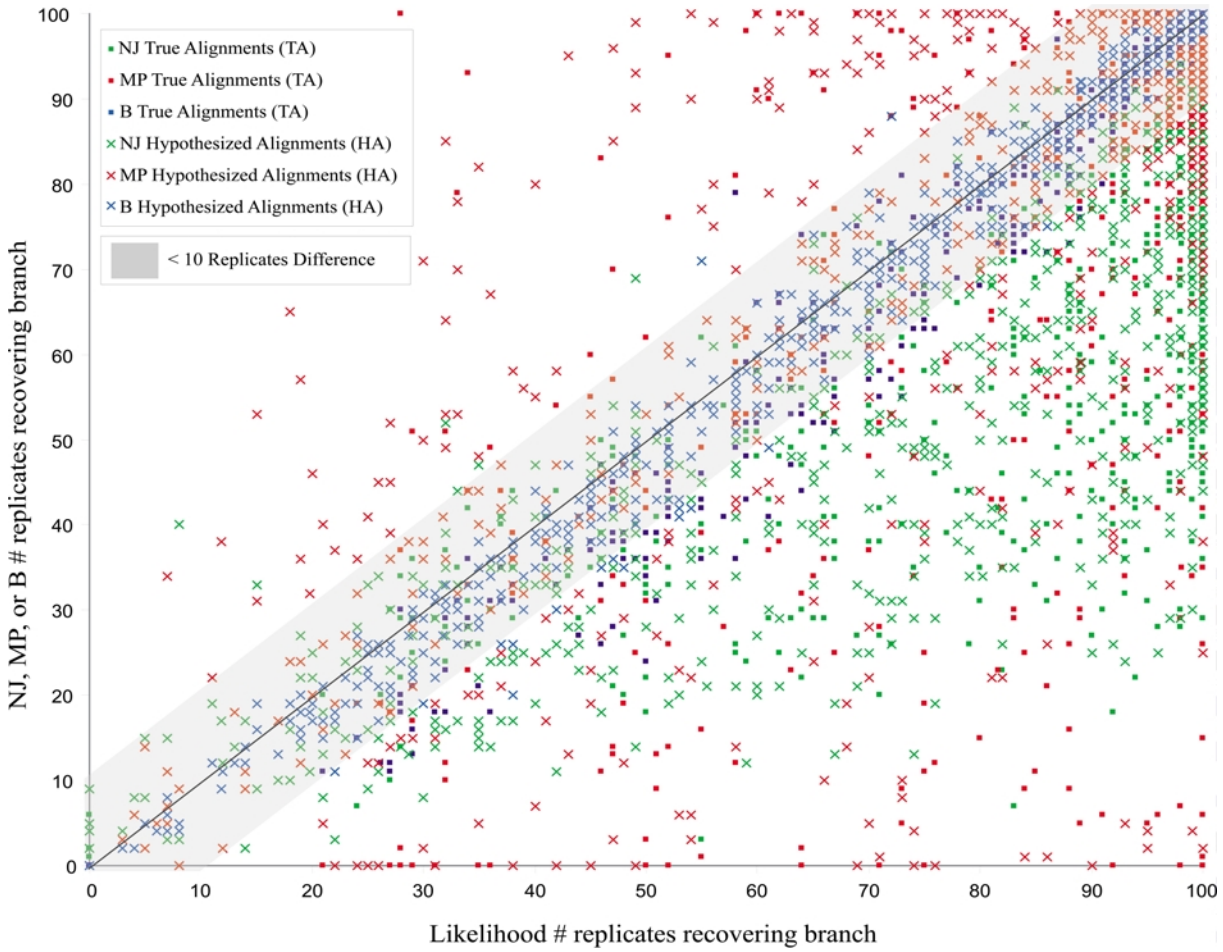

FIGURE 8. Topological accuracy comparison of ML versus the other tree reconstruction methods (NJ, MP, and B). The shaded area represents 86% of all comparisons between ML and the other methods where there was a difference <10 in the number of replicates that recovered the branch. Outside the shaded area are the remaining 14% of the contrasts where there was a difference ≥10 in the number of replicates that recovered the branch. See Table 2 for a detailed breakdown of the cases outside the shaded area.

From a pure phylogenetic perspective, the details of the indel model are likely to be of less importance than one might suspect, as long as the total number of gapped sites found between any pair of taxa is reasonable given their evolutionary distance. The simple reason for this is that the phylogenetic methods used in this study completely ignore indel size and distribution; basic phylogenetic methods treat each site independently. Gaps are either treated as missing/unknown (e.g., in maximum parsimony) or the sites are ignored completely (e.g., in the distance calculations of neighbor joining); once aligned it does not matter whether gaps are randomly distributed throughout the sequence or are clustered. One could randomly rearrange the columns of the aligned matrix without affecting the phylogenetic reconstruction. Obviously, more complicated methods and models of phylogeny reconstruction do take order and placement of gaps into account, but for the analyses performed in this study, the specifics of gap size and distribution are of little consequence.

Clearly, the specifics of the indel model do affect alignment accuracy, but as already stated, our simulations provide accuracy curves similar to those found in other recent studies based on completely different models. Because the goal of this study was to contrast phylogeny reconstructions based on correctly and incorrectly aligned data sets, any alignment errors due to incorrect model specification simply enhance this contrast.

Method Comparison

Although our primary objective was to examine the relationships between alignment accuracy and topological accuracy, and the interaction of alignment accuracy with tree shape, our results permitted performance comparisons of the different methods of tree reconstruction. Although this has been done many times for fixed alignments (e.g., Hillis, 1995; Huelsenbeck and Rannala, 2004; Nei, 1996; Rosenberg and Kumar, 2003; Takahashi and Nei, 2000), the current study differs in that it takes into account the additional variable of data sets that require

TABLE 2. Comparison of topological accuracy of ML versus the other three methods. For each method comparison, the first column represents the percent of all the analyses that have a reasonable difference (≥10); the next two columns show which method reconstructs the topology more correctly, of the cases (from the first column) where there was a difference. TA = True Alignment; HA = Hypothesized Alignment (see text).

	ML vs. NJ			ML vs. MP			ML vs. Bayesian		
	% Of all the analyses with a reasonable difference (≥10)	Reconstructs the topology more accurately		% Of all the analyses with a reasonable difference (≥10)	Reconstructs the topology more accurately		% Of all the analyses with a reasonable difference (≥10)	Reconstructs the topology more accurately	
		ML	NJ		ML	MP		ML	B
TA	24.38	100.00%	0.00%	13.39	83.96%	16.04%	2.95	98.31%	1.69%
HA	28.22	98.76%	1.24%	13.94	62.01%	37.99%	0.80	87.50%	12.50%

alignment (Hall, 2005). Figure 8 shows that for 86% of the cases examined in our study (on a branch by branch comparison), there was no reasonable difference (defined arbitrarily as a <10 difference in the number of replicates that recovered the branch, as indicated by the shaded region on the graph) between ML and any of the other methods. However, for the other 14%, ML outperformed the remaining methods in the vast majority of cases. Table 2 contains a breakdown of these cases where there was a reasonable difference (≥10 replicates that recovered the branch) in topological accuracy between each method and ML. The measure comes from the number of replicates (out of 100) that recovered the correct node(s).

ML versus NJ.—ML outperformed NJ in 100% of the TA cases where there was a difference. Of the 28.22% HA cases with a reasonable difference, ML was more accurate 98.76% of the time.

ML versus MP.—When compared to MP, ML is more accurate for both the TA and the HA (83.96% and 62.01% of the time, respectively). Nevertheless, we know that as alignments become less accurate, MP is less affected, and these data show that for HA, there are many cases where MP does outperform ML (37.99% of the time). Thus, in general, ML will outperform MP; however, for data sets that contain more alignment error, MP; may also outperform ML in many cases as well (almost 40% of the cases in our study). These results also confirm many empirical results demonstrating that alignment error can influence the relationships of certain nodes (Lake, 1991; Morrison and Ellis, 1997; Mugridge et al., 2000; Ogden and Whiting, 2003; Thorne and Kishino, 1992). It could be argued that our study is in an unfair comparison given that the exact model that was used to simulate the data was also given to ML during tree reconstruction, whereas no model (or unweighted parsimony) was used in MP. Clearly this is a simple case and future studies need to address the issue of performance where both methods have the "wrong model." This would be particularly important as real data are most likely much more complicated than any currently constructed model.

ML versus B.—Of the very few cases where a reasonable difference was found between ML and B (2.95% for the TA and 0.80% for the HA), likelihood was more accurate most of the time (98.31% for the TA and 87.50% for the HA). This is interesting in that, although ML and B have essentially the same level of performance (around 97% of the cases with no reasonable difference), when there is a difference, ML does a better job at reconstructing the correct topology. It must be noted that we did not examine bootstrapping or posterior probabilities, rather these comparisons were based on whether a branch was recovered or not (strict consensus in MP and ML, and majority rule in B). These results are similar to those Hall (2005) found when comparing across the same four methods of tree reconstruction.

Alternative Branch Error (BE) Measure

Our method of estimating alignment accuracy for each branch (BAA) is an average measure of the alignment inaccuracy that crosses the branch, rather than an estimate of the error directly associated with the branch. As an alternate to BAA, we also estimated the amount of alignment error generated by each branch. To do this, we constructed a matrix containing the pairwise alignment errors for every pair of sequences. This matrix was fit onto the true simulated topology using a

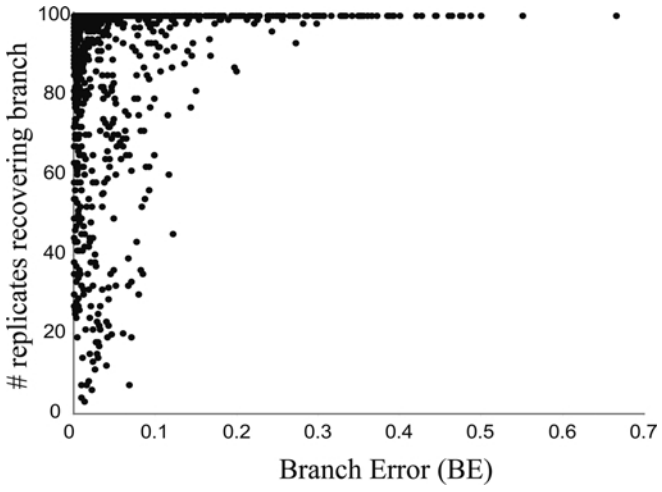

FIGURE 9. Relationship of Branch Error (BE) and topological accuracy, as measured by the number of replicates (out of 100) that recovered each branch. Each point represents a single internal branch from a specific model tree. Points to the far left are data sets that contained little or no branch alignment error, whereas points to the right contained more alignment error. Only the ML analyses are shown on this graph, but the other methods show the same general trend.

least-squares fit procedure (in the exact same manner an evolutionary distance matrix may be fit onto a topology in order to evaluate a minimum evolution model). This led to an alignment error estimate for each branch (BE) that indicates the proportion of misaligned sites that can be attributed to the branch itself. This measure had some rather interesting, but otherwise useless properties, but we think it bears mentioning nonetheless. The first is that the fit worked very well, such that large alignment errors were associated with long branches and small alignment errors were associated with short branches. Logically, a branch of zero length cannot have any alignment error associated with it because no mutations would have occurred and identical sequences align perfectly, whereas long branches generate many mutations and lead to greater levels of misalignment (Rosenberg, 2005a). In fact, the correlation between BE and branch length is $r = 0.7323$ across all simulated data sets. The (initially) unexpected result is found when one examines the relationship between BE and branch reconstruction accuracy (Fig. 9). Branches with large amounts of alignment error are reconstructed accurately whereas those with little to no alignment error are often reconstructed poorly. This apparently paradoxical result is explained by the correlation between branch length and BE. Extremely short branches may not lead to alignment error but, as is well known, are difficult to reconstruct. Long branches may lead to more alignment error, but are also easier to reconstruct because of the increased number of character changes associated

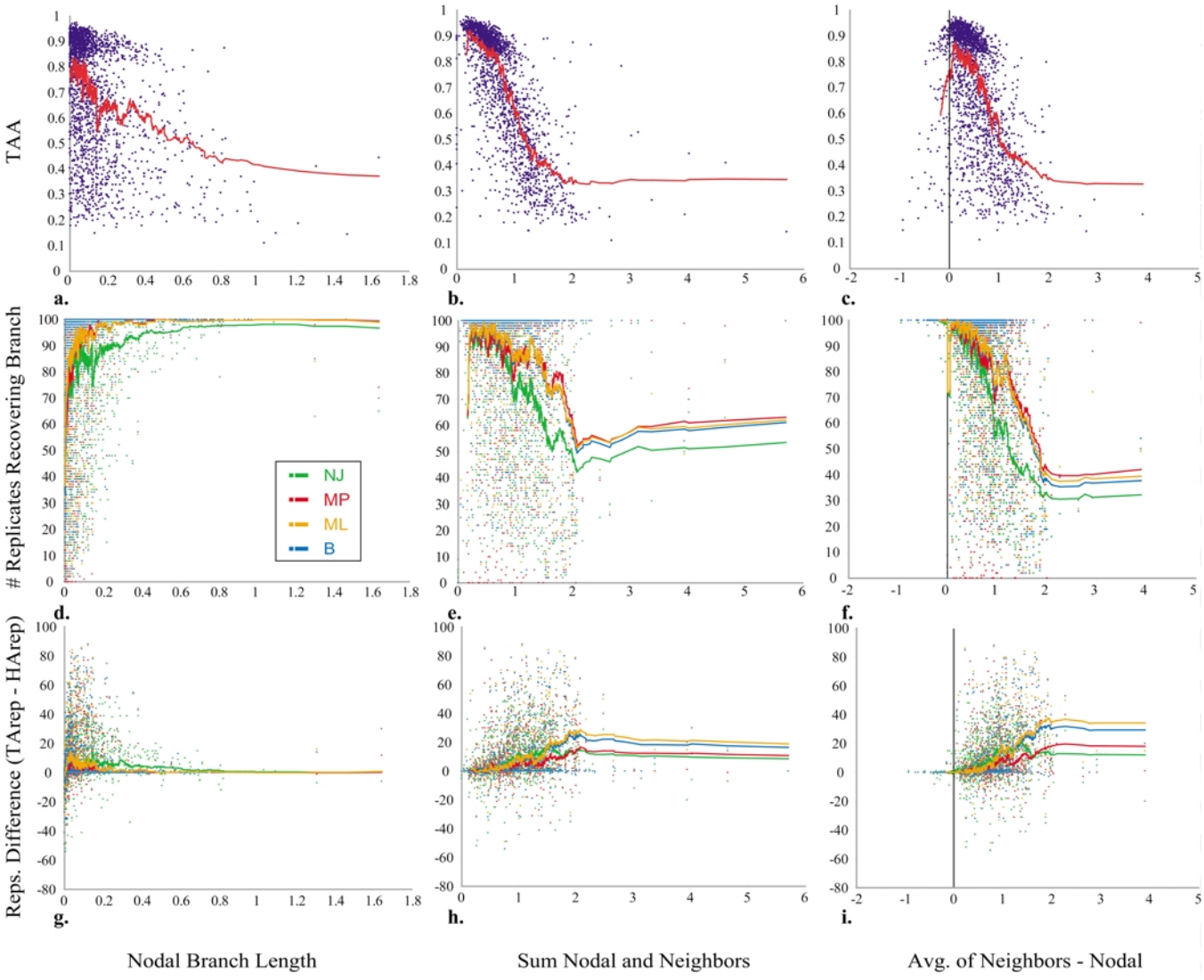

FIGURE 10. Results of the three ways branch length was quantified: the nodal branch length (length of the branch being analyzed); the sum of the nodal branch length and the lengths of the four neighboring branches; and the average length of the four neighboring branches minus the nodal branch length. Each branch length summary is plotted against the TAA (Total Alignment Accuracy), the number of replicates (out of 100) that recovered each branch, and the difference between the True Alignment (TA) reconstructions and the Hypothesized Alignment (HA) reconstructions in the number of replicates (out of 100) that recovered each branch. For the TAA row (a–c), all tree shape cases and methods are pooled together, whereas in the remaining panels (d–i) results are separated by tree reconstruction method.

with the branch. Figure 9 really indicates the relationship between branch length and branch reconstruction accuracy; the alignment error generated by a branch is essentially inconsequential relative to other topological factors. This is why BAA is a more informative measure; it accounts for global trends in alignments associated with a specific branch rather than just local alignment properties.

Branch Length

As demonstrated above with BE, branch length is clearly an important aspect of alignment and topological accuracy. In order to further investigate its role, we quantified branch length three different ways (Fig. 10): (1) nodal branch length (length of the branch that separates the tree into two clades); (2) sum of the nodal branch length and the lengths of the four neighboring branches; and (3) average length of the four neighboring branches minus the nodal branch length. Our results indicate that as nodal branch length increases, TAA decreases. This is true for the nodal branch (Fig. 10a) but even more so as the lengths of the neighboring branches are also considered (Fig. 10b and c).

Given that long branches lead to more alignment error, it is logical to conclude that as branch lengths increase, topological accuracy should decrease. However, our results indicate that as the nodal branch length increases, topological accuracy also increases (Fig. 10d). Of course, if extremely long lengths were used, one would expect the curve to decrease as the sequences become saturated. This counterintuitive result may be explained in the same way as above (BE section). The results for the sum of the five branches (Fig. 10e) and the average length of the four neighboring branches minus the length of the nodal branch (Fig. 10f) indicate that as neighboring branch length increases, topological accuracy decreases. Therefore, the lengths of the neighboring branches may actually be the leading factor that causes low topological accuracy, confirming many ideas in the well-documented phenomenon of long branch attraction (Bergsten, 2005; Felsenstein, 1978; Gadagkar et al., 2005; Huelsenbeck, 1995, 1997; Huelsenbeck and Hillis, 1993; Siddall and Whiting, 1999; Whiting, 1998).

A related result is seen in the replicates difference ($TA_{rep} - HA_{rep}$) value and its relationship to the nodal branch length (Fig. 10g), the sum of the five branches (Fig. 10h), and the average length of the four neighboring branches minus the nodal branch length (Fig. 10i). As length increases on the nodal branch, there is little to no difference between the number of replicates from the TA and HA that recover that node. Yet, as the neighboring branches become longer than the nodal branch, TA tree reconstructions outperform HA tree reconstructions by recovering the nodal branch in more replicates (a difference reaching nearly 40 in ML and B, and around 20 in MP and NJ in Fig. 10i).

Thus, we see that the lengths of both the nodal branch and the neighboring branches influence the alignment accuracy, although the neighboring branch lengths may be more important for topological accuracy.

CONCLUSION

This study represents an initial important step into understanding the influence of alignment accuracy on phylogenetic inference. We presented two measures of alignment accuracy, TAA and BAA, which were used to investigate the relationship of alignment error and tree reconstruction. It has been shown (on the average) that for more balanced tree shapes and shorter branch lengths, alignment error may have little affect on topological reconstruction, and that for more pectinate tree shapes and longer branches, the effect is much more pronounced. However, any one hypothesized alignment may give you a topology that is very accurate, very inaccurate, or something in between. It is difficult to predict this on a case by case basis, but this study does confirm that the alignment can drive the resultant accuracy. Under the studied conditions, maximum likelihood and Bayesian, in general, outperformed the other methods (neighbor joining and maximum parsimony) in terms of tree reconstruction accuracy. Branch length also played an important role in both alignment accuracy and topological accuracy. We recognize that there are many other factors that may play a part in alignment and topological accuracy in addition to the ones studied here. However, "you have to start somewhere" (Strunk and White, 1918), and we choose to begin by primarily investing the affects of tree shape. Hence, many potential variables (such as point substitution and indel models) that could influence both alignment and tree reconstruction were held constant in this study and future efforts will be required to elucidate their effects.

ACKNOWLEDGMENTS

Thanks to D. Morrison, K. Kjer, and an anonymous reviewer for comments and suggestions on an earlier version of this manuscript. This work was partially supported by the NIH R03-LM008637 (MSR) and Arizona State University.

REFERENCES

Bergsten, J. 2005. A review of long-branch attraction. Cladistics 21:163–193.

Cammarano, P., R. Creti, A.M. Sanangelantoni, and P. Palm. 1999. The Archaea monophyly issue: A phylogeny of translational elongation factor G (2) sequences inferred from an optimized selection of alignment positions. J. Mol. Evol. 49:524–537.

Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 436:69–87.

Colless, D. H. 1996. A further note on symmetry of taxonomic trees. Syst. Biol. 45:385–390.

Coventry, A., D. J. Kleitman, and B. Berger. 2004. MSARI: Multiple sequence alignments for statistical detection of RNA secondary structure. Proc. Natl. Acad. Sci. USA 101:12102–12107.

de Pinna, M. C. C. 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7:367–394.

di Bernardo, D., T. Down, and T. Hubbard. 2003. ddbRNA: Detection of conserved secondary structures in multiple alignments. Bioinformatics 19:1606–1611.

Dowell, R., and S. Eddy. 2004. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. BMC Bioinformatics 5:71.

Farris, J. S., and M. Kallersjo. 1998. Asymmetry and explanations. Cladistics 14:159–166.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.
Fleissner, R., D. Metzler, and A. Haeseler. 2005. Simultaneous statistical multiple alignment and phylogeny reconstruction. Syst. Biol. 54:548–561.
Gadagkar, S. R., M. S. Rosenberg, and S. Kumar. 2005. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. J. Exp. Zool. B Mol. Dev. Evol. 304B:64–74.
Hall, B. G. 2005. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. Mol. Biol. Evol. 22:792–802.
Harcourt-Brown, K. G., P. N. Pearson, and M. Wilkinson. 2001. The imbalance of paleontological trees. Paleobiology 27:188–204.
Hasegawa, M., K. Kishino, and T. Yano. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.
Heard, S. B., and A. O. Mooers. 1996. Imperfect information and the balance of cladograms and phenograms. Syst. Biol. 45:115–118.
Hillis, D. M. 1995. Approaches for assessing phylogenetic accuracy. Syst. Biol. 44:3–16.
Holmes, I. 2003. Using guide trees to construct multiple-sequence evolutionary HMMs. Bioinformatics 19:147i–157i.
Holmes, I. 2004. A probabilistic model for the evolution of RNA structure. BMC Bioinformatics 5:166.
Holmes, I. 2005a. Accelerated probabilistic inference of RNA structure evolution. BMC Bioinformatics 6:73.
Holmes, I. 2005b. Using evolutionary expectation maximization to estimate indel rates. Bioinformatics 21:2294–2300.
Holmes, I., and W. J. Bruno. 2001. Evolutionary HMMs: A Bayesian approach to multiple alignment. Bioinformatics 17:803–820.
Huelsenbeck, J. P. 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44:17–48.
Huelsenbeck, J. P. 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46:69–74.
Huelsenbeck, J. P., and D. M. Hillis. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. 42:247–264.
Huelsenbeck, J. P., and M. Kirkpatrick. 1996. Do phylogenetic methods produce trees with biased shapes? Evolution 50:1418–1424.
Huelsenbeck, J. P., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53:904–913.
Huelsenbeck, J. P., and F. R. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755.
Hwang, U. W., W. Kim , D. Tautz, and M. Friedrich. 1998. Molecular phylogenetics at the Felsenstein zone: Approaching the Strepsiptera problem using 5.8S and 28S rDNA sequences. Mol. Phylogenet. Evol. 9:470–480.
Keightley, P. D., and T. Johnson. 2004. MCALIGN: Stochastic alignment of noncoding DNA sequences based on an evolutionary model of sequence evolution. Genome Res. 14:442–450.
Kjer, K. 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53:506–514.
Kjer, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. Mol. Phylogenet. Evol. 4:314–330.
Knudsen, B., and J. Hein. 1999. RNA secondary structure prediction using stochastic context-free grammars and evolutionary history. Bioinformatics 15:446–454.
Knudsen, B., and M. M. Miyamoto. 2003. Sequence alignments and pair hidden Markov models using evolutionary history. J. Mol. Biol. 333:453–460.
Lake, J. A. 1991. The order of sequence alignment can bias the selection of tree topology. Mol. Biol. Evol. 8:378–385.
Löytynoja, A., and N. Goldman. 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. USA 102:10557–10562.
Lunter, G., I. Miklos, A. Drummond, J. Jensen, and J. Hein. 2005. Bayesian coestimation of phylogeny and sequence alignment. BMC Bioinformatics 6:83.
Maddison, W. P., and D. R. Maddison. 2004. Mesquite: A modular system for evolutionary analysis, version 1.05.
Metzler, D. 2003. Statistical alignment based on fragment insertion and deletion models. Bioinformatics 19:490–499.
Miklos, I., G. A. Lunter, and I. Holmes. 2004. A "long indel" model for evolutionary sequence alignment. Mol. Biol. Evol. 21:529–540.
Mitchison, G., and R. Durbin. 1995. Tree-based maximal likelihood substitution matrices and hidden Markov models. J. Mol. Evol. (Hist. Arch.) 41:1139–1151.
Mitchison, G. J. 1999. A probabilistic treatment of phylogeny and sequence alignment. J. Mol. Evol. 49:11–22.
Mooers, A. O., R. D. M. Page, A. Purvis, and P. H. Harvey. 1995. Phylogenetic noise leads to unbalanced cladistic tree reconstructions. Syst. Biol. 44:332–342.
Morrison, D., and J. Ellis. 1997. Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of apicomplexa. Mol. Biol. Evol. 14:428–441.
Mugridge, N. B., D. A. Morrison, T. Jakel, A. R. Heckeroth, A. M. Tenter, and A. M. Johnson. 2000. Effects of sequence alignment and structural domains of ribosomal DNA on phylogeny reconstruction for the protozoan family Sarcocystidae. Mol. Biol. Evol. 17:1842–1853.
Nei, M. 1996. Phylogenetic analysis in molecular evolutionary genetics. Ann. Rev. Genet. 30:371–403.
Ogden, T. H., and M. Whiting. 2003. The problem with "the Paleoptera problem": Sense and sensitivity. Cladistics 19:432–442.
Ogden, T. H., M. F. Whiting, and W. C. Wheeler. 2005. Poor taxon sampling, poor character sampling, and non-repeatable analyses of a contrived dataset do not provide a more credible estimate of insect phylogeny: A reply to Kjer. Cladistics 21:295–302.
Ophir, R., and D. Graur. 1997. Patterns and rates of indel evolution in processed pseudogenes from humans and murids. Gene 205:191–202.
Phillips, A., D. Janies, and W. Wheeler. 2000. Multiple sequence alignment in phylogenetic analysis. Mol. Phylogenet. Evol. 16:317–330.
Pollard, D., C. Bergman, J. Stoye, S. Celniker, and M. Eisen. 2004. Benchmarking tools for the alignment of functional noncoding DNA. BMC Bioinformatics 5:6.
Redelings, B., and M. Suchard. 2005. Joint Bayesian estimation of alignment and phylogeny. Syst. Biol. 54:401–418.
Rivas, E., and S. Eddy. 2001. Noncoding RNA gene detection using comparative sequence analysis. BMC Bioinformatics 2:8.
Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.
Rosenberg, M. S. 2005a. Evolutionary distance estimation and fidelity of pair wise sequence alignment. BMC Bioinformatics 6:102.
Rosenberg, M. S. 2005b. Multiple sequence alignment accuracy and evolutionary distance estimation. BMC Bioinformatics 6:278.
Rosenberg, M. S. 2005c. MySSP: Non-stationary evolutionary sequence simulation, including indels. Evol. Bioinformatics Online 1:51–53.
Rosenberg, M. S., and S. Kumar. 2003. Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. Mol. Biol. Evol. 20:610–621.
Siddall, M. E., and M. F. Whiting. 1999. Long-branch abstractions. Cladistics 15:9–24.
Strunk, W., and E. B. White. 1918. The elements of style, 4th edition. Allyn and Bacon, Boston.
Sundstrom, H., M. T. Webster, and H. Ellegren. 2003. Is the rate of insertion and deletion mutation male biased?: Molecular evolutionary analysis of avian and primate sex chromosome sequences. Genetics 164:259–268.
Swofford, D. L. 2002. PAUP* Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
Takahashi, K., and M. Nei. 2000. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol. Biol. Evol. 17:1251–1258.
Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
Thorne, J. L., and H. Kishino. 1992. Freeing phylogenies from artifacts of alignment. Mol. Biol. Evol. 9:1148–1162.

Thorne, J. L., H. Kishino, and J. Felsenstein. 1991. An evolutionary model for the maximum likelihood alignment of sequence evolution. J. Mol. Evol. 33:114–124.

Thorne, J. L., H. Kishino, and J. Felsenstein. 1992. Inching toward reality: An improved likelihood model of sequence evolution. J. Mol. Evol. 34:3–16.

Titus, T. A., and D. R. Frost. 1996. Molecular homology assessment and phylogeny in the lizard family Opluridae (Squamata: Iguania). Mol. Phylogenet. Evol. 6:49–62.

Wenzel, J. W., and M. E. Siddall. 1999. Noise. Cladistics 15:51–64.

Wheeler, W. 2001. Homology and the optimization of DNA sequence data. Cladistics 17:S3–S11.

Wheeler, W. C. 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst. Biol. 44:321–331.

Whiting, M. F. 1998. Long-branch distraction and the Strepsiptera. Syst. Biol. 47:134–138.

Xia, X., Z. Xie, and K. M. Kjer. 2003. 18S ribosomal RNA and tetrapod phylogeny. Syst. Biol. 52:283–295.

First submitted 14 July 2005; reviews returned 17 October 2005; final acceptance 25 November 2005

Associate Editor: Rod Page