

Alignment and Topological Accuracy of the Direct Optimization approach via POY and Traditional Phylogenetics via ClustalW + PAUP*

T. HEATH OGDEN^{1,2} AND MICHAEL S. ROSENBERG²

¹Idaho State University, Department of Biological Sciences, Pocatello, Idaho 83209, USA; E-mail: ogdet@isu.edu

²Center for Evolutionary Functional Genomics, The Biodesign Institute, and the School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501, USA; E-mail: msr@asu.edu (M.S.R.)

Abstract.—Direct optimization frameworks for simultaneously estimating alignments and phylogenies have recently been developed. One such method, implemented in the program POY, is becoming more common for analyses of variable length sequences (e.g., analyses using ribosomal genes) and for combined evidence analyses (morphology + multiple genes). Simulation of sequences containing insertion and deletion events was performed in order to directly compare a widely used method of multiple sequence alignment (ClustalW) and subsequent parsimony analysis in PAUP* with direct optimization via POY. Data sets were simulated for pectinate, balanced, and random tree shapes under different conditions (clocklike, non-clocklike, and ultrametric). Alignment accuracy scores for the implied alignments from POY and the multiple sequence alignments from ClustalW were calculated and compared. In almost all cases (99.95%), ClustalW produced more accurate alignments than POY-implied alignments, judged by the proportion of correctly identified homologous sites. Topological accuracy (distance to the true tree) for POY topologies and topologies generated under parsimony in PAUP* from the ClustalW alignments were also compared. In 44.94% of the cases, Clustal alignment tree reconstructions via PAUP* were more accurate than POY, whereas in 16.71% of the cases POY reconstructions were more topologically accurate (38.38% of the time they were equally accurate). Comparisons between POY hypothesized alignments and the true alignments indicated that, on average, as alignment error increased, topological accuracy decreased. [ClustalW; direct optimization; multiple sequence alignment; parsimony; phylogenetics; POY; sensitivity analysis; simulation; tree reconstruction.]

There are many approaches available for phylogenetic analysis of DNA sequence information. Traditionally, most analyses proceed as a two-step process, first moving from raw sequence data to a multiple sequence alignment and then, secondarily, using the multiple sequence alignment to estimate a phylogeny. For the first step, there are dozens of programs and algorithms available that will perform multiple sequence alignment. Once the multiple sequence alignment (data matrix) is created, the data are analyzed by tree building methodologies resulting in hypothesized trees that are used to infer the evolution of the sequences (and ultimately the organisms themselves). Among the most widely used tree reconstruction approaches are neighbor joining, parsimony, maximum likelihood, and Bayesian approaches. Recently, however, other approaches have surfaced that can analyze DNA sequence data in a different framework.

Direct optimization (DO) is an alternative approach for phylogenetic analysis in which no prior multiple sequence alignment is required. This idea has attracted much attention due to the fact that empirically derived phylogenies may be more dependent upon the alignment method 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). Although some related ideas concerning optimization heuristics for raw sequence data existed prior to DO (Hein, 1989a, 1989b; Hogeweg and Hesper, 1984; Sankoff, 1975; Thorne et al., 1991), Wheeler (1996) developed an automated cladistic or parsimony approach. DO, formerly known as optimization alignment, was devised to counter the lack of interaction between topology and putative homology (Wheeler, 2001) and to assess directly

the transformations, indels, or other evolutionary events simultaneously in a topological framework without the use of multiple sequence alignment (Wheeler, 1996). The strategy is implemented in the program POY (Wheeler et al., 2003), which can invoke both parsimony and likelihood as optimality criteria (Wheeler, 2006). Recently, a number of similar approaches to DO that implement Bayesian or likelihood models in a combined analysis framework have also been proposed (Fleissner et al., 2005; Lunter et al., 2005; Redelings and Suchard, 2005). DO is a novel theoretical approach to phylogenetic estimation that attempts to avoid the problems of alignment by generalizing phylogenetic character analysis to include insertion/deletion events.

Although theoretical arguments for and against DO, and POY in particular, exist (Giribet, 2001, 2005; Kjer, 2004; Ogden et al., 2005; Phillips et al., 2000; Simmons, 2004; Simmons and Ochoterena, 2000; Wheeler, 2001, 2003), there have been no accuracy comparative analyses performed to evaluate “alternative phylogenetic methods or models” (de Queiroz and Poe, 2001), although some studies have used other forms of comparison (for example, ILD and likelihood scores) to examine the performance of POY versus other alignment methods (Terry and Whiting, 2005; Whiting et al., 2006). During revision of this paper we became aware of an accepted paper that investigates a comparison of POY versus structural alignment (Kjer et al., 2007). We were interested in using simulation to directly compare ClustalW and subsequent parsimony analysis in PAUP* with DO via POY. Many 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). Only recently have alignments been simulated that include indels (Blanchette et al., 2004a, 2004b; Fleissner et al., 2005; Hall, 2005; Keightley and Johnson, 2004; Pollard et al., 2004; Rosenberg, 2005a, 2005b; Stoye et al., 1998). However, none of these studies (except for Fleissner et al., 2005) compare traditional two-step phylogenetic analysis with DO or combined analysis approaches.

The main objective of this paper is to directly compare, within a parsimony framework, the performance of DO (via POY) to phylogenetic methods that first generate a multiple sequence alignment with subsequent tree reconstruction (via ClustalW and PAUP* parsimony). This will be accomplished by (1) calculating and comparing the alignment accuracy score for the implied alignment from POY and the multiple sequence alignment from ClustalW; (2) calculating and comparing the topological accuracy (distance to the true tree) for POY topologies and topologies generated under parsimony in PAUP* from the ClustalW alignments; and (3) investigating the interaction of tree shape (length and branching pattern) with alignment and topological accuracy in relation to the two approaches of sequence analysis.

MATERIAL AND METHODS

Data Simulation

We used the simulated data sets, consisting of seven 16-taxon topologies, which we have previously analyzed (Ogden and Rosenberg, 2006). The simulations were done 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 across the thousands of data sets. The seven base topologies (Fig. 1) consist 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), where the probability for each speciation event is equal for all tips. The relative branch lengths of each topology were set under 11 different conditions: Ultrametric equal-branch length, clocklike random branch length (5 sets), and non-clocklike random branch lengths (5 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 7 topologies was used to create 22 model trees (Fig. 2). All simulations were conducted under identical conditions in the program MySSP (Rosenberg, 2005c). For this study, many potentially variable parameters were held constant in order to gain simplicity. Thus, aside from the different conditions explained above, the initial sequence length was set to 2000 base pairs and noncoding 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 ex-

pected 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, 2005b). 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). As an aside, it is important to point out that Ophir and Graur's (1997) paper was based on indels in pseudogenes and that our conclusions may be limited to evolutionary patterns more similar to these types of data. The intent of associating our indel model to these types of data was to initiate this line of research with a more general case. We are currently working on methods for more specific cases, such as rDNA. 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). Although the underlying mechanisms and frequencies of indels is not understood as well as base substitution processes (Hall, 2005), efforts to remedy this lack of models and methods are underway (Holmes, 2003, 2004, 2005; Holmes and Bruno, 2001; Knudsen and Miyamoto, 2003; Mitchison and Durbin, 1995; Mitchison, 1999).

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

Alignment

These simulations resulted in 15,400 unique data sets (alignments) containing gaps representing either insertion or deletion events during the simulation process, which will be referred to as the True Alignments (TA). Each of the TA was then stripped of their gaps and realigned via ClustalW version 1.83 (Thompson et al., 1994) using default parameters. The default parameters in Clustal for DNA sequence alignment are gap opening = 15, gap extension = 6.66, delay divergent % = 30, DNA transition weight = 0.50, and DNA weight matrix = IUB. These parameters were chosen because they are the most commonly used settings by most investigators who implement Clustal. Our main purpose was to compare ClustalW to POY using the most common approach employed for each program and not to try to match exact parameters between the programs, which may not even be possible. We will refer to these alignments as the Clustal Hypothesized Alignments (Clustal HA). Although POY does not require an alignment, we a posteriori generated the implied alignment from the resultant most parsimonious tree, and these alignments will be

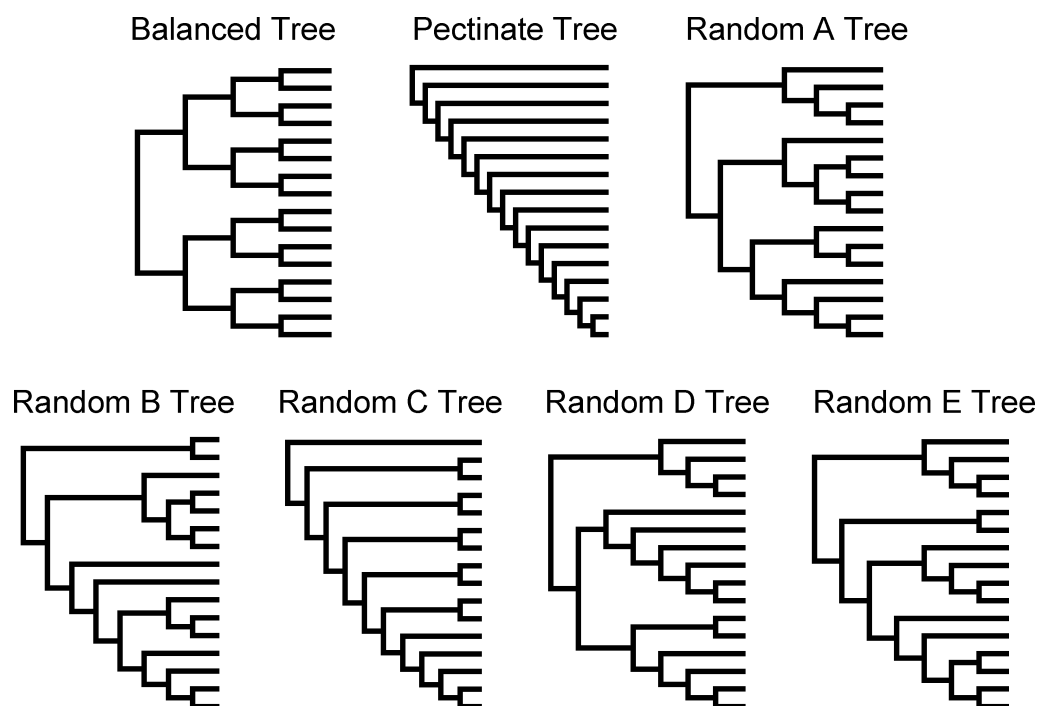


FIGURE 1. The seven basic topologies used to simulate the data. The different tree shapes consisting of a balanced tree, a pectinate tree, and five random Yule trees (A–E). Same as figure 1 from Ogden and Rosenberg (2006).

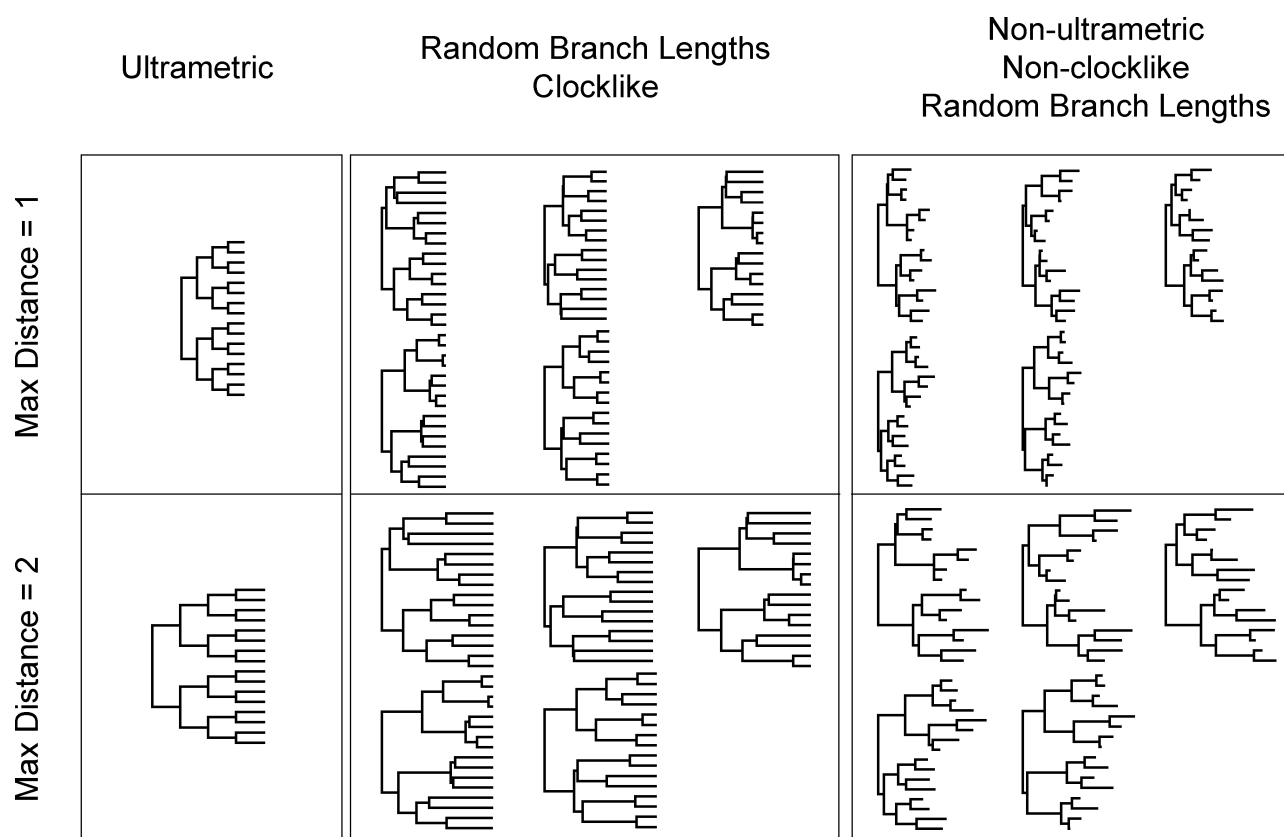


FIGURE 2. An example of the 22 model trees for the balanced tree shape. Consisting of: Ultrametric equal-branch length, clocklike random branch length (5 sets), and non-clocklike random branch lengths (5 sets). Each of these 11 conditions was scaled such that the maximum evolutionary distance between a pair of sequences was equal to 1.0 or 2.0. Same as Fig. 2 from Ogden and Rosenberg (2006).

referred to as POY Hypothesized Alignments (POY HA). The gap:transversion:transition ratio in POY was 1:1:1 for the majority of the analyses. This particular parameter set was chosen because it has been the most commonly selected in phylogenetic studies in the top systematic journals, based on a literature review from January 2004 to February 2006, where 53% of the studies selected the 1:1:1 parameter set, 31% of the studies did not select the 1:1:1 ratio, and 16% did not report the specific parameter settings used in POY. Notwithstanding, we also performed analyses under varying cost schemes for the pectinate tree (1, 2, 4, and 10 for the gap:tv and tv:ts ratios) in order to examine the possibility that other commonly used cost ratios in POY may recover more accurate topologies and alignments. We chose to do this on the pectinate tree because it was the tree shape where POY performed the best, in terms of topological accuracy, and because the outgroup was fixed (see below for discussion on how POY must search using one outgroup). Even though we examined the resulting topological effects of varying parameters in POY, similar to sensitivity analysis approaches (Aagesen et al., 2005; Laamanen et al., 2005; Ogden and Whiting, 2003; Terry and Whiting, 2005), the focus of this study was not to try to estimate the optimal parameter settings that would generate the most accurate alignment and reconstructed topology. Rather, we wanted to compare the performance of the most common settings that are used in phylogenetic analyses for POY and Clustal.

Alignment Accuracy

Alignment accuracy, calculated as the proportion of aligned sites that are truly homologous (Rosenberg, 2005a), was summarized as the Total Alignment Accuracy (TAA) score. The TAA for a data set was calculated as the average accuracy of all pairwise sequence comparisons in the multiple alignments as judged against the corresponding homologous sites of the true alignments. Thus, we did not compare the Clustal-generated hypothesized alignment directly to the POY-implied alignment, as some have suggested these may not be comparable (Giribet, 2005; Wheeler, 2003). Rather, we directly compared the Clustal-generated hypothesized alignments (HA) to the true alignments; likewise, we independently compared the POY-implied alignments (POY HA) to the true alignments. In other words, we compared the primary homology sites of the Clustal hypothesis to the true secondary homology sites of the simulated data set; and we compared the secondary homology sites from the POY-implied alignment hypotheses to the true secondary homology sites of the simulated data set (de Pinna, 1991; Giribet, 2005). Finally, an indirect comparison of the two hypothesized alignment TAA scores was matched up, head-to-head (Clustal HA versus POY HA), for every simulation replicate for all tree shapes.

Tree Reconstruction

The Clustal HA and TA were analyzed under parsimony using PAUP* version 4.b10 for Windows

(Swofford, 2002) consisting of 100 random additions with TBR swapping and all other default settings. In order to compare the performance of POY HA to POY TA (which are the same data sets as Clustal TA), we additionally analyzed all of the TA with gaps as a fifth state character (GapMode=NewState) in PAUP*. This was done because POY HA analyses treated gaps as such, using the 1:1:1 gap:tv:ts ratio. These four different approaches to treating the sequences (HA and TA with gaps as missing in Clustal and PAUP, and POY HA and TA with gaps treated as fifth) allowed direct comparisons of topological accuracy of the resulting trees (consensus in some cases).

We looked at the effects of alignment error on reconstruction accuracy by comparing the TA tree reconstructions to the HA tree reconstructions. 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 topological accuracy of trees reconstructed from the true and hypothesized alignments. When the TA tree is topologically more accurate than the HA tree, ($HA_{dist} - TA_{dist}$) will be a positive number; if ($HA_{dist} - TA_{dist}$) is negative, the HA tree is more accurate than the TA tree. Note that ($HA_{dist} - TA_{dist}$) itself is not 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 ($HA_{dist} - TA_{dist}$) equal to 0. Alternatively, they could both be equally inaccurate, with large distances relative to the true tree, and again ($HA_{dist} - TA_{dist}$) may also be 0 (the reconstructed trees could be completely different, but also completely wrong).

In addition to analyzing the data for a cost ratio of 1:1:1 (gap:transversion:transition) in POY, a sensitivity analysis was performed across the pectinate cases to examine if other parameter sets may produce more accurate hypotheses of homology and more accurate topologies. For the 1:1:1 analyses across all tree shapes (all the data sets) and for the sensitivity analysis data sets (pectinate tree shape only), the following search strategy was used in POY: -nooneasis-noleading-norandomizeoutgroup-quick-staticapprox-notbr-replicates 4-. Although this strategy does not represent a very thorough search, it was necessary in order to be able to analyze the thousands of data sets in a reasonable amount of time. However, much more extensive analyses (-nooneasis-noleading-norandomizeoutgroup-quick-staticapprox-notbr-replicates 4-buildmaxtrees 2-sprmaxtrees 1-nospr-tbr-tbrmaxtrees 5-maxtrees 5-holdmaxtrees 50-slop 5-checkslop 10-stopat 25-treefuse-fuse-limit 10-fusemingroup 5-fusemaxtrees 50-ratchetspr 2-ratchettbr 2-checkslop 10) on a subset of the data (200 data sets) showed that there was no significant difference in topological accuracy; in fact, the less extensive analyses recovered, on average, more accurate topologies by 0.14 ($HA_{dist} - TA_{dist}$) distance. Furthermore,

alignment accuracy actually decreased from an average of 70% down to 68% for the more extensive searches. Both of these values are much less than the Clustal alignment accuracy average of 86%. Therefore, further searching, at least on these data, does not change the results and conclusions, and we are justified in using the results from the less extensive searches to make general comparisons. Although, again, it should be pointed out that the results from any one single data set might be drastically changed with different efforts of searching.

It is appropriate to summarize the results from our previous publication (Ogden and Rosenberg, 2006) in order to better understand the current study. In general, as alignment error increased, topological accuracy decreased. This trend was much more pronounced for data sets derived from more pectinate topologies. On the other hand, for balanced, clocklike, and equal-branch-length tree shapes, alignment inaccuracy had little average effect on tree reconstruction. These conclusions are based on averages across the different simulation conditions pooled together. However, it should be pointed out that any one specific analysis, independent of the alignment accuracy, was recovered more accurate, less inaccurate, or equally accurate as the true topology. Maximum likelihood and Bayesian, in general, outperformed neighbor joining and parsimony in terms of tree reconstruction accuracy. The results also indicated that as the length of the branch and of the neighboring branches increased, alignment accuracy decreased, indicating that neighboring branches may be a major factor in topological accuracy. These basic conclusions will be drawn upon throughout this paper; for further details see Ogden and Rosenberg (2006).

RESULTS

Of the 15,400 head-to-head comparisons of Clustal HA and POY HA alignments, in only 7 cases were POY alignments more accurate than Clustal alignments, as measured by TAA scores (Table 1 and Fig. 3); in the remaining 15,393 cases, the Clustal HA were more accurate. In other words, 99.95% of the time, ClustalW generated alignments that were more accurate, as judged by the comparison to the truly homologous sites of the true alignment, than POY implied alignments. Clustal alignments ranged from 97.79% to 17.33% accurate with an average of 72.19%, whereas POY-implied alignments ranged from 87.78% to 5.05% accurate with an average of 50.76%. Of the few POY HA cases that outperformed the associated Clustal HA, the maximum TAA difference was 0.07. However, in the numerous cases where Clustal HA outperformed POY HA, the maximum TAA difference was 0.74.

Although not as drastic as the alignment accuracy comparisons, Clustal HA topological accuracy also outperformed POY tree reconstructions in 6920 cases (44.94%). POY did better in 2573 cases (16.71%), and both performed equally well (or poorly) in 5907 cases (38.36%) (Table 2). Therefore, across all data sets pooled together, Clustal HA reconstructions were, on average, more accu-

TABLE 1. Comparison of Total Alignment Accuracy (TAA) for Clustal-hypothesized alignments and POY-hypothesized alignments. The Hypothesized Alignments (HA) are derived by stripping out the gaps in the truly homologous alignments (generated by the simulation program) and then using Clustal to construct a multiple sequence alignment or retrieve the POY implied alignment of the gapless sequences. Comparisons are made across all tree shapes (15,400 cases) and for just the balanced (2200 cases) and pectinate (2200 cases) tree shapes.

	Alignment accuracy			
	No. cases where TAA POY > TAA Clustal	%	No. cases where TAA Clustal > TAA POY	%
Across all tree shapes	7	0.05	15393	99.95
Balanced tree shapes	5	0.23	2195	99.77
Pectinate tree shapes	0	0.00	2200	100.00

rate (in terms of distance to the true tree) than POY reconstructions. As POY TAA decreased, this average difference became larger (Fig. 4). In the 2573 cases where POY reconstructed more accurate topologies than Clustal HA reconstructions, nearly half (1009 cases) were observed in the pectinate tree shape data sets (Table 2). For the pectinate tree shape data, Clustal HA reconstructions recovered only 744 trees that were more accurate than the associated POY HA reconstructions (in 447 cases no difference was seen). Therefore, for the pectinate topologies, over 45% of the time POY reconstructed trees more accurately as compared to 33% where Clustal recovered more accurate trees. (In 20% of the pectinate cases,

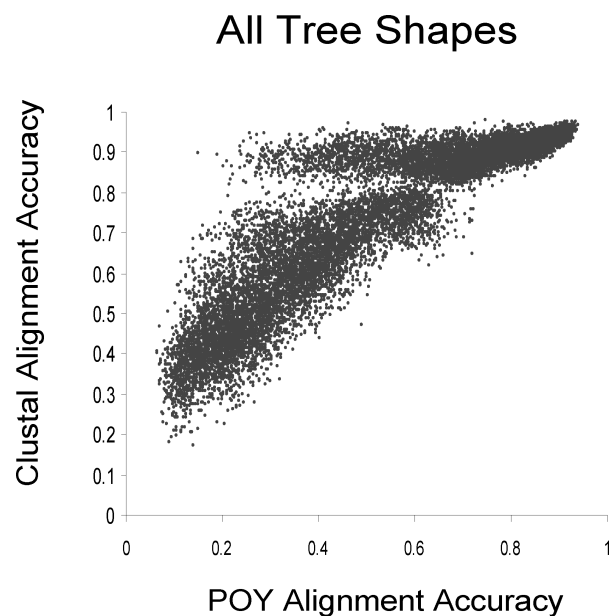


FIGURE 3. Comparison of total alignment accuracy (TAA) for Clustal alignments and POY TAA of implied alignments. Comparisons are made across all tree shapes (15,400 cases). The points in the upper cluster and lower cluster correspond to the maximum evolutionary distance between a pair of sequences that was equal to 1.0 and 2.0, respectively.

TABLE 2. Comparison of topological accuracy, as measured by the Robinson-Foulds (1981) distance to the true tree, for Clustal HA reconstructions and POY HA reconstructions. Comparisons are made across all tree shapes (15,400 cases) and for just the balanced (2200 cases) and pectinate (2200 cases) tree shapes.

	Topological accuracy					
	No. cases where POY HA reconstructions recovered more accurate topologies than Clustal HA	%	No. cases where Clustal HA reconstructions recovered more accurate topologies than POY HA	%	No. cases where POY and Clustal were equally accurate	%
Across all tree shapes	2573	16.71	6920	44.94	5907	38.36
Balanced tree shapes	181	8.23	897	40.77	1122	51.00
Pectinate tree shapes	1009	45.86	744	33.82	447	20.32

both approaches recovered equally accurate topologies.) The moving average of all the pectinate tree shape cases showed an interesting trend. There was a noticeable change in the curve for POY TAA accuracies between

0.4 and 0.6 (Fig. 5). In this span of POY TAA, Clustal HA reconstructions were more accurate than POY HA reconstruction (on average). By breaking down the different types of pectinate trees, the results show that the

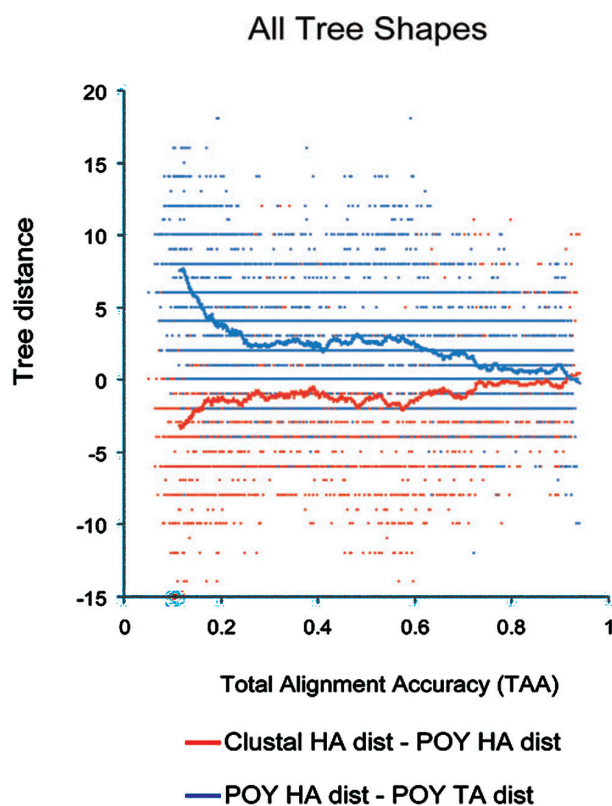


FIGURE 4. Relationship of POY TAA (total alignment accuracy measured from implied alignments; results follow similar pattern shifted to the right if Clustal TAA are used) and the relative performance of topological accuracy of Clustal versus POY for all cases of varying tree shapes and conditions. Points to the far right are the most accurate alignments while points to the left are the least accurate alignments. For the (Clustal HA_{dist} - POY HA_{dist}) data points, negative values on the y-axis are cases where Clustal HA reconstructions were more topologically accurate than the POY HA reconstructions, and the opposite is true for positive values. For the (POY HA_{dist} - POY TA_{dist}) data points, positive values on the y-axis are cases where the POY TA-reconstructed trees were more accurate than the POY HA-reconstructed trees, and the opposite is true for negative values. Many points may be superimposed upon one another, particularly when the y-axis value = 0 (cases where no differences were observed). The lines are moving averages based on an overlapping sliding window of 250 consecutive points.

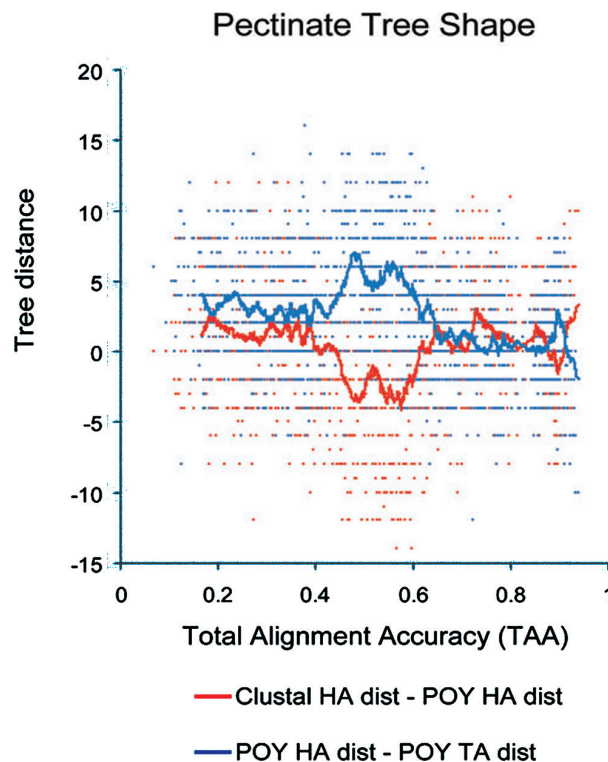


FIGURE 5. Relationship of POY TAA (total alignment accuracy measured from implied alignments; results follow similar pattern shifted to the right if Clustal TAA are used) and the relative performance of topological accuracy of Clustal versus POY for the pectinate tree shapes. Points to the far right are the most accurate alignments while points to the left are the least accurate alignments. For the (Clustal HA_{dist} - POY HA_{dist}) data points, negative values on the y-axis are cases where Clustal HA reconstructions were more topologically accurate than the POY HA reconstructions, and the opposite is true for positive values. For the (POY HA_{dist} - POY TA_{dist}) data points, positive values on the y-axis are cases where the POY TA-reconstructed trees were more accurate than the POY HA-reconstructed trees, and the opposite is true for negative values. Many points may be superimposed upon one another, particularly when the y-axis value = 0 (cases where no differences were observed). The lines are moving averages based on an overlapping sliding window of 50 consecutive points.

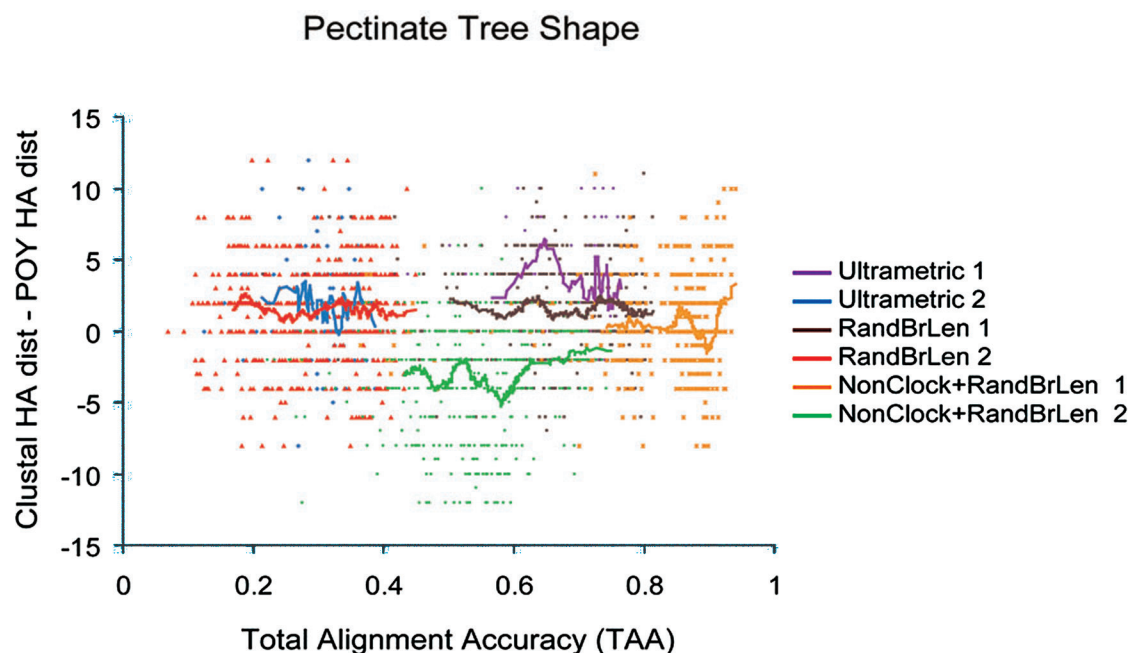


FIGURE 6. Relationship of POY TAA (total alignment accuracy measured from implied alignments; results follow similar pattern shifted to the right if Clustal TAA are used) to the relative performance of topological accuracy of Clustal versus POY broken down into the six main types of pectinate tree shapes. The 1 and 2 in the legend indicate that the maximum evolutionary distance between a pair of sequences was equal to 1.0 and 2.0, respectively. The different tree types used for simulation are indicated: Ultrametric = ultrametric equal branch length; RandBrLen = clocklike random branch length; and NonClock+RandBrLen = non-clocklike random branch lengths. Points to the far right are the most accurate alignments while points to the left are the least accurate alignments. The data points with positive values on the y -axis are cases where POY HA reconstructions were more topologically accurate than the Clustal HA reconstructions, and the opposite is true for negative values. The lines are moving averages based on an overlapping sliding window of 50 consecutive points.

region of POY TAA between 0.4 and 0.6 is predominantly represented by the pectinate, random branch length, and non-clocklike condition. For these cases, the Clustal HA reconstructions vastly outperform the POY reconstructions (on average). Contrastingly, most of the cases where POY HA tree reconstructions outperform Clustal HA tree reconstructions are for data sets derived from pectinate, nonrandom branch length, and more clocklike tree shapes that fall within the POY TAA range less than 0.4 and greater than 0.6 (Fig. 6). Except for these few cases, Clustal+PAUP generally outperformed POY in topological accuracy across all other tree shapes and conditions. Finally, it should be reemphasized that these conclusions are based on average trends of many analyses under different conditions, and any one specific analysis, independent of the alignment accuracy and methodology, may recover very accurate or inaccurate topologies.

The results across the pectinate tree data sets from the varied POY cost ratios (sensitivity analysis) indicate a couple of interesting points (Table 3). First, only 3 of the 35,200 cases of the POY parameter sets investigated recovered more accurate hypotheses of secondary homology than the primary homology hypotheses of Clustal. In other words, Clustal-hypothesized alignments essentially always outperformed POY-implied alignments, regardless of the parameter set used in POY. Second, our results show that, on average, 2:2:1 cost ratio was the

most accurate of the POY parameter sets investigated. And four other parameter sets (4:4:1, 4:2:1, 10:10:1, and 2:1:1) also recovered more accurate alignments than the 1:1:1 set. Therefore, although 1:1:1 is the most commonly used (often selected from an ILD test), for our data it did not produce the most accurate homologies. Related to this is the result that the 1:1:1 parameter set did not recover the most accurate topologies either (Table 4). Once again, 2:2:1 recovered more accurate topologies more often than any other cost ratio, followed by 4:4:1 and then 1:1:1. There is a fairly good correlation ($r^2 = -0.675$) between the parameter sets that recover accurate topologies more often than other parameter sets as compared to the parameter sets from which more accurate implied alignments are produced. As an aside, 2:1:1 (gap cost of two relative to nucleotide changes) is the default setting in POY.

We also examined the effect of alignment accuracy on topological accuracy for the POY analyses. Across all data sets, as POY TAA decreased, ($HA_{dist} - TA_{dist}$) increased from around 0 (no difference between the POY TA and POY HA) for very accurate alignments to more than 7 for very inaccurate alignments (Fig. 4, blue points and line). This trend was consistent across all tree shapes, except for the pectinate trees, where the maximum ($HA_{dist} - TA_{dist}$) values were found between 0.40 and 0.60 ($HA_{dist} - TA_{dist}$) distance (Fig. 5).

TABLE 3. Relative performance for alignment accuracy of each possible pair of cost ratios for the pectinate tree cases. For each pair of ratios, the table shows the number of cases in which the cost ratio in the left column resulted in more accurate alignments than the cost ratio in the top row. Accuracy was measured as the proportion of truly homologous non-gapped sites. A ranking (1–16), based on the average of any ratio compared to all other ratios, indicates the order of parameter sets from most accurate (1) to less accurate (16) in identifying truly homologous sites. The number of cases where Clustal alignments were more accurate is included in the bottom row.

Cost ratio	Rank	Average	1:1:1	2:1:1	4:1:1	10:1:1	1:2:1	2:2:1	4:2:1	10:2:1	1:4:1	2:4:1	4:4:1	10:4:1	1:10:1	2:10:1	4:10:1	10:10:1
1:1:1	6	1347.4375	0	224	2157	2200	2199	30	144	2199	2200	2151	79	1134	2200	2200	2199	243
2:1:1	5	1627.5	1976	0	2198	2200	2200	299	585	2200	2200	2196	546	1907	2200	2200	2200	933
4:1:1	9	958.8125	43	2	0	2200	1449	1	0	2121	2152	1132	2	30	2152	1988	2062	7
10:1:1	16	6.0625	0	0	0	0	0	0	0	1	22	0	0	0	22	17	35	0
1:2:1	10	885.4375	1	0	751	2200	0	0	0	1969	2198	607	0	5	2197	2089	2150	0
2:2:1	1	1929.6875	2170	1901	2199	2200	2200	0	1523	2199	2200	2200	1478	2158	2200	2200	2200	1847
4:2:1	3	1792.6875	2056	1615	2200	2200	2200	677	0	2200	2200	2200	974	2126	2200	2200	2200	1435
10:2:1	12	518.1875	1	0	79	2199	231	1	0	0	1496	158	1	1	1520	1165	1438	1
1:4:1	13	350.3125	0	0	48	2178	2	0	0	704	0	0	0	0	1126	396	1151	0
2:4:1	8	986.1875	49	4	1068	2200	1593	0	0	2042	2200	0	0	34	2199	2191	2199	0
4:4:1	2	1836.5	2121	1654	2198	2200	2200	722	1226	2199	2200	2200	0	2135	2200	2200	2200	1729
10:4:1	7	1341.375	1066	293	2170	2200	2195	42	74	2199	2200	2166	65	0	2200	2199	2200	193
1:10:1	15	342.3125	0	0	48	2178	3	0	0	680	1066	1	0	0	0	371	1130	0
2:10:1	11	566.25	0	0	212	2183	111	0	0	1035	1804	9	0	1	1829	0	1876	0
4:10:1	14	347.5	1	0	138	2165	50	0	0	762	1049	1	0	0	1070	324	0	0
10:10:1	4	1663.25	1957	1267	2193	2200	2200	353	765	2199	2200	2200	471	2007	2200	2200	2200	0
No. cases Clustal more accurate			2200	2200	2200	2200	2200	2199	2200	2200	2200	2200	2199	2200	2200	2200	2200	2199

DISCUSSION AND CONCLUSION

The results clearly indicate that, under the data simulation conditions of this study, Clustal alignments (under default parameters) with subsequent parsimony tree building approach is superior to DO parsimony in POY (under a wide range of parameter sets). This is especially true for the comparisons of Clustal alignments and POY-implied alignments in relation to the true alignments, where over 99% of the time Clustal produces more accurate alignments; i.e., more accurate proportion of truly homologous sites. Similar results have recently been reported for comparisons of Clustal versus the newer likelihood DO-like methods (for example, Fleissner et al., 2005). Therefore, traditional multiple sequence align-

ment approaches appear to vastly outperform direct optimization-like approaches in terms of alignment accuracy, at least for the data sets and parameter settings that have been examined thus far.

Even though Clustal outperforms POY in terms of alignment accuracy, we know that there are many cases where less accurate alignments will recover more accurate topologies. These results confirm that alignment accuracy is not directly tied to topological accuracy for any one specific data set, but that on average more accurate alignments do lead to more accurate topologies. Furthermore, these results indicate that for these data simulation conditions, Clustal, and most likely other multiple sequence alignment programs, with subsequent

TABLE 4. Relative performance for topological accuracy of each possible pair of cost ratios for the pectinate tree cases. For each pair of ratios, the table shows the number of cases in which the cost ratio in the left column resulted in less accurate topologies than the cost ratio in the top row. Topological accuracy was measured as the distance to the true tree. A ranking (1–16), based on the topological accuracy averages, indicates which parameter set resulted in the most accurate topological reconstructions. (Note that, as opposed to Table 3, here the lower value indicates which ratio was most accurate.) The number of cases where the POY reconstruction (for the different cost ratios) was more accurate than the PAUP reconstruction of the Clustal alignments is included in the bottom row.

Cost ratio	Rank	Average	1:1:1	2:1:1	4:1:1	10:1:1	1:2:1	2:2:1	4:2:1	10:2:1	1:4:1	2:4:1	4:4:1	10:4:1	1:10:1	2:10:1	4:10:1	10:10:1
1:1:1	3	412.6875	0	724	298	25	557	876	654	125	135	668	838	458	82	125	267	771
2:1:1	4	462.6875	885	0	328	19	644	986	700	116	170	784	898	485	98	144	279	867
4:1:1	10	952	1573	1486	0	59	1284	1599	1387	367	498	1390	1562	1070	337	438	715	1467
10:1:1	16	1853.375	2132	2128	2001	0	2101	2135	2118	1701	1738	2113	2143	2065	1584	1697	1875	2123
1:2:1	8	605.5625	1139	1031	482	29	0	1182	944	192	189	931	1141	667	132	184	386	1060
2:2:1	1	370	762	667	260	23	493	0	571	115	123	624	748	401	97	113	207	716
4:2:1	7	544	1044	940	390	20	766	1085	0	164	214	895	1040	546	130	174	345	951
10:2:1	14	1382.1875	1893	1883	1457	233	1774	1922	1813	0	1006	1817	1892	1603	821	904	1216	1881
1:4:1	12	1302.75	1859	1777	1332	237	1686	1899	1753	792	0	1755	1861	1535	649	738	1132	1839
2:4:1	6	530.125	990	923	411	37	688	1069	830	176	203	0	1021	577	137	178	308	934
4:4:1	2	391.6875	791	715	288	17	551	830	608	122	131	654	0	410	86	112	227	725
10:4:1	9	784.6875	1348	1274	649	49	1050	1382	1162	298	364	1203	1382	0	238	321	530	1305
1:10:1	15	1459.875	1966	1926	1515	329	1834	1987	1887	1004	1068	1883	1986	1686	0	998	1350	1939
2:10:1	13	1374.625	1890	1865	1425	269	1777	1939	1810	882	951	1822	1930	1591	752	0	1220	1871
4:10:1	11	1110.875	1681	1623	1085	163	1473	1721	1555	615	627	1580	1698	1287	452	575	0	1639
10:10:1	5	464.5	943	829	350	26	634	972	736	136	155	767	883	476	116	144	265	0
No. cases POY more accurate			1045	967	439	24	759	1113	866	162	148	893	1075	598	84	132	287	992

phylogenetic analysis (at least in a parsimony framework, but most likely true for others as well) will often lead to more accurate topologies than POY and possibly other direct optimization approaches (Fleissner et al., 2005). This is evident from the result that POY, on average, recovered less accurate alignments than Clustal+PAUP across nearly the entire spread of alignment accuracy (Fig. 4). Only for a small span of the most highly accurate topologies did POY outperform Clustal+PAUP. This area where POY apparently does better may just be an artifact of gap treatment, as POY treated gaps using a 1:1:1 (gap:tv:ts) ratio, which is the PAUP equivalent of treating gaps as a new or fifth state character. Our Clustal+PAUP approach treated gaps as missing. Some of our other work (Ogden and Rosenberg, 2007) indicates that treating gaps as fifth state character is generally a better approach than treating gaps as missing data. Thus, the small area where POY recovers more accurate topologies turns out to be essentially nonexistent when we compared POY to Clustal+PAUP with gaps treated as a fifth state character instead of missing. The difference, across all tree shapes, in topological accuracy between the Clustal+PAUP cases (more accurate) and POY cases (less accurate) was even more extreme when we compared POY to fifth state gap treatment in PAUP. Therefore, although treating gaps as missing was a more conservative approach (as well as representing the more commonly used approach to parsimony-based phylogenetic analysis), it still recovered more accurate topologies on average across all tree shapes and conditions pooled together.

Although Clustal+PAUP recovered, on average, more accurate topologies than POY across all tree shapes, POY outperformed Clustal+PAUP in some single-data set cases (as indicated by points with positive *y*-axis values in Fig. 4) and numerous pectinate tree shape cases (Table 2). Specifically, our data showed that the pectinate, nonrandom branch length tree shapes were, on average, more accurately reconstructed topologically in POY, even though none of the POY implied alignments were more accurate than the Clustal alignments (Tables 1 and 2, and Fig. 6). It is interesting that the pectinate, nonrandom branch length topologies were essentially the only tree shape type that, on average, resulted in more accurate topologies under POY. This may have something to do with the rooting issue described below, but it may also mean that direct optimization approaches are better at reconstructing topologies with more pectinate shapes and clocklike evolution. However, because we do not know the tree shape for empirical data beforehand, it would generally be better to use approaches that recover relationships more accurately under a variety of tree shapes instead of being biased toward a specific tree shape.

The cost ratios of the sensitivity analysis, on average, did not perform as well as the Clustal+PAUP analyses. To reiterate the example, only 3 out of the 35,200 implied alignments identified more truly homologous sites than the Clustal hypotheses of homology. So even though POY was permitted to explore the parameter cost space, it still did not produce more accurate alignments. Although only 9584 POY reconstructions across the pa-

rameter space were more accurate than Clustal HA reconstructions, the 2:2:1 was more accurate in 1113 of the possible 2200 cases. In other words, for the pectinate tree shape cases, more often than not, using a 2:2:1 parameter set in POY recovered more accurate topologies. This is in spite of the fact that there was only one case where the implied alignment actually had more homologous sites hypothesized. Another interesting trend that is apparent from the sensitivity analysis is that cases where the ratio of the gap and transversion cost is equal (2:2:1, 4:4:1, 1:1:1, and 10:10:1) resulted in the most accurate topologies and implied alignments within the POY analyses. As many research studies use congruence to select among parameters, it would be interesting to see if the ILD test actually selects the best parameter set (as measured by the set that gives the highest accuracy) in a future simulation study.

Critics may have concerns with our experimental design and conclusions that we make based on the results. One possible criticism of the current study could be that with an increased taxon sampling, POY may outperform traditional methods. In order to address this issue, we also performed 100 simulations on a 64-taxon clocklike random branch length topology. Analyses of these simulations showed that while in the 16-taxon trees only around 44% of the time Clustal+PAUP recovers more accurate topologies, in the 64-taxon cases, 93% of the time Clustal+PAUP recovers more accurate topologies. Apparently, the benefit of adding more taxa is greater in Clustal+PAUP (alignment, branch swapping, tree searching, etc.) than in POY.

Another possible criticism could be that we do not use tree length to compare among competing topologies from POY and Clustal+PAUP. The tree length measure is not reported because we do not think that tree length is an appropriate optimality criterion for different data sets (different alignments). Once you have different hypotheses of homology (different alignments), you have different data sets. By the same reasoning, you cannot compare distance, likelihood, or Bayesian measures for noncongruent or nonhomologous data sets. This is true even if only one column in the matrix is different. If the matrix is not the same, the resulting topologies should not be compared based on a tree length optimality score. POY claims to circumvent this issue because it goes directly from the sequences to the topology. However, just as it may not be appropriate to take a POY-implied alignment and then run other analyses like bootstraps or likelihood, it likewise may not be appropriate to take a Clustal alignment + PAUP tree reconstruction and run it through POY to output a tree length. In a simulation framework such as ours, what one can do is compare the hypotheses of homologous sites (alignments) to the true homologous sites (simulated data set) and one can compare the hypothesis of branching pattern (tree) to the true branching pattern (simulated tree). But one should not compare tree lengths that have been derived from different hypotheses of homologous sites; it can be done, but in our view it should not be done.

These results support the use of Clustal+PAUP or other two-step approaches as opposed to DO and POY

for data sets similar to the ones we simulated. However, it is possible that for alternate types of data sets, the putatively positive aspects of POY may override the problems identified in this study. For example, our data do not include morphology (simulation of morphology is problematic in practice and theory) and we included only one partition or "gene." DO through POY may perform better on multiple partitioned data sets, particularly when simulated under differing evolutionary rates and models (Giribet, 2001; Terry and Whiting, 2005). Moreover, it would seem that the large majority of empirical analyses that have used POY contained at least one ribosomal gene (12S, 16S, 18S, or 28S). These genes are made up of conserved (stem) and unconserved (loop or expansion) regions, and it remains unclear how accurately POY or other combined approaches may perform by combining these data together in simultaneous analysis than independent alignment of each individual region. This study can only suggest that the demonstrated level of superiority of traditional methods of phylogenetic analysis over direct optimization methods for the examined conditions is very convincing.

Aside from any biases that may exist due to the particular nature of the data sets, there may be some alternative explanations for these disparate results. It is possible that the program POY, and not necessarily the theoretical framework of direct optimization-like approaches (but see Fleissner et al., 2005), may have some implementation limitations that bias the results. One issue that appears to be problematic is that POY searches across rooted topologies and thus requires a single specified outgroup. Most other methods of tree reconstruction search across unrooted topologies and rooting is not required until searching has completed (if at all). For our data sets, only the pectinate tree and the random C tree contained the same single outgroup rooting in POY and in the rooted Clustal+PAUP reconstructions (Fig. 1), which is one of the reasons we chose to do the additional extensive analyses and the sensitivity analysis on the pectinate tree shape. Thus, there may be a bias of tree distance to the true topology for the other tree shapes. If we consider only the pectinate and the random C trees, Clustal alignments were always more accurate than POY-implied alignments, but POY HA reconstructions were topologically more accurate 31% of the time (1373 out of 4400 cases), whereas Clustal+PAUP reconstructions were more topologically accurate in 40% of the cases (1761 out of 4400). Still, the outgroup, by definition, is the least closely related taxa to any other taxa in the data set and most likely will also be the most evolutionary distant. Therefore, the comparison of the most basal hypothetical taxon unit (the hypothesized ancestor of the ingroup) to the outgroup may be difficult to correctly align and optimize. Of course, the evolutionary distances of the ingroup and outgroup can be very small, but as they increase, the error in each of the hypothetical taxon units (hypothesized ancestors on each internal node) will most likely also increase. Still, evolutionary distance does bias Clustal alignment accuracy (Pollard et al., 2004;

Rosenberg, 2005a, 2005b), but perhaps in a different way than seen in POY.

Despite the above issues and results, the framework of direct optimization may still be a useful way to analyze data under different data sets and/or implementations (i.e., modified parsimony, likelihood, Bayesian, etc.), but further development, exploration, and testing are required. Although our data represents a fairly simple case, for data sets similar to these the traditional two-step approach will almost always give a more accurate alignment and will most likely recover equally or more accurate phylogenetic relationships than direct optimization as implemented in POY.

There are many issues that remain to be studied concerning the performance of direct optimization methods, such as more complicated data sets and alternative frameworks (i.e., likelihood and Bayesian). This study represents the first analysis to directly compare traditional two-step phylogenetic analysis (via Clustal+PAUP) to direct optimization (via POY) in order to analyze both alignment and topological accuracy. In almost all cases (99.95%), ClustalW produced more accurate alignments than POY-implied alignments. Similarly, in 45% of the cases, Clustal alignment tree reconstructions in PAUP* were more topologically accurate than the POY tree reconstructions, which were only more accurate than Clustal+PAUP in 17% of the cases. Varied cost ratios (sensitivity analysis) in POY also performed worse, on average, than Clustal+PAUP, although within POY 2:2:1 was consistently the most accurate parameter set. Finally, the same trend (on average as alignment accuracy increases, topological accuracy increases) found for multiple sequence alignment with tree reconstruction via neighbor joining, parsimony, likelihood, and Bayesian held true for direct optimization via POY. All 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.

ACKNOWLEDGMENTS

We wish to thank G. Giribet, K. Kjer, and an anonymous reviewer for comments and suggestions on an earlier version of this manuscript. Parts of this work were supported by the NIH R03-LM008637 (MSR) and Arizona State University.

REFERENCES

- Aagesen, L., G. Petersen, and O. Seberg. 2005. Sequence length variation, indel costs, and congruence in sensitivity analysis. *Cladistics* 21:15–30.
- Blanchette, M., E. D. Green, W. Miller, and D. Haussler. 2004a. Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res.* 14:2412–2423.
- Blanchette, M., W. J. Kent, C. Riemer, L. Elntski, A. F. A. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, and W. Miller. 2004b. Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. *Genome Res.* 14:708–715.
- 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.

- de Pinna, M. C. C. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7:367–394.
- de Queiroz, K., and S. Poe. 2001. Philosophy and phylogenetic inference: A comparison of likelihood and parsimony methods in the context of Karl Popper's writings on corroboration. *Syst. Biol.* 50:305–321.
- Fleissner, R., D. Metzler, and A. Haeseler. 2005. Simultaneous statistical multiple alignment and phylogeny reconstruction. *Syst. Biol.* 54:548–561.
- Giribet, G. 2001. Exploring the behavior of POY, a program for direct optimization of molecular data. *Cladistics* 17:S60–S70.
- Giribet, G. 2005. Generating implied alignments under direct optimization using POY. *Cladistics* 21:396–402.
- Hall, B. G. 2005. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Mol. Biol. Evol.* 22:792–802.
- 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.
- Hein, J. 1989a. A new method that simultaneously aligns and reconstructs ancestral sequences for any number of homologous sequences, when the phylogeny is given. *Mol. Biol. Evol.* 6:649–668.
- Hein, J. 1989b. A tree reconstruction method that is economical in the number of pairwise comparisons used. *Mol. Biol. Evol.* 6:669–684.
- Hillis, D. M. 1995. Approaches for assessing phylogenetic accuracy. *Syst. Biol.* 44:3–16.
- Hogeweg, P., and B. Hesper. 1984. The alignment of sets of sequences and the construction of phyletic trees: An integrated method. *J. Mol. Evol.* 20:175–186.
- 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 Bioinform.* 5:166.
- Holmes, I. 2005. 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., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.* 53:904–913.
- 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., J. J. Gillespie, and K. A. Ober. 2007. Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. *Syst. Biol.* 56:133–156.
- 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 M. M. Miyamoto. 2003. Sequence alignments and pair hidden Markov models using evolutionary history. *J. Mol. Biol.* 333:453–460.
- Laamanen, T. R., R. Meier, M. A. Miller, A. Hille, and B. M. Wiegmann. 2005. Phylogenetic analysis of Themira (Sepsidae: Diptera): Sensitivity analysis, alignment, and indel treatment in a multigene study. *Cladistics* 21:258–271.
- Lake, J. A. 1991. The order of sequence alignment can bias the selection of tree topology. *Mol. Biol. Evol.* 8:378–385.
- Lunter, G., I. Miklos, A. Drummond, J. Jensen, and J. Hein. 2005. Bayesian coestimation of phylogeny and sequence alignment. *BMC Bioinform.* 6:83.
- Maddison, W. P., and D. R. Maddison. 2004. Mesquite: A modular system for evolutionary analysis, version 1.05.
- Mitchison, G., and R. Durbin. 1995. Tree-based maximal likelihood substitution matrices and hidden Markov models. *J. Mol. Evol. (Historical Archive)* 41:1139–1151.
- Mitchison, G. J. 1999. A probabilistic treatment of phylogeny and sequence alignment. *J. Mol. Evol.* 49:11–22.
- 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. Heckerroth, 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. *Annu. Rev. Genet.* 30:371–403.
- Ogden, T. H., and M. S. Rosenberg. 2006. Multiple sequence alignment accuracy and phylogenetic inference. *Syst. Biol.* 55:314–328.
- Ogden, T. H., and M. S. Rosenberg. 2007. How should gaps be treated in parsimony? A comparison of approaches using simulation. *Mol. Phylogenet. Evol.* 42:817–826.
- 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 Bioinform.* 5:6.
- Redelings, B., and M. Suchard. 2005. Joint Bayesian estimation of alignment and phylogeny. *Syst. Biol.* 54:401–418.
- 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 Bioinform.* 6:102.
- Rosenberg, M. S. 2005b. Multiple sequence alignment accuracy and evolutionary distance estimation. *BMC Bioinform.* 6:278.
- Rosenberg, M. S. 2005c. MySSP: Non-stationary evolutionary sequence simulation, including indels. *Evol. Bioinform.* 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.
- Sankoff, D. 1975. Minimal mutation trees of sequences. *SIAM. J. Appl. Math.* 28:35–42.
- Simmons, M. P. 2004. Independence of alignment and tree search. *Mol. Phylogenet. Evol.* 31:874–879.
- Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49:369–381.
- Stoye, J., D. Evers, and F. Meyer. 1998. Rose: Generating sequence families. *Bioinformatics* 14:157–163.
- 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.
- Terry, M. D., and M. F. Whiting. 2005. Comparison of two alignment techniques within a single complex data set: POY versus Clustal. *Cladistics* 21:272–281.
- 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.

- 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.
- Wheeler, W. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? *Cladistics* 12:1–9.
- Wheeler, W. 2001. Homology and the optimization of DNA sequence data. *Cladistics* 17:S3–S11.
- Wheeler, W. C. 2003. Implied alignment: a synapomorphy-based multiple-sequence alignment method and its use in cladogram search. *Cladistics* 19:261–268.
- Wheeler, W. C. 2006. Dynamic homology and the likelihood criterion. *Cladistics* 22:157–170.
- Wheeler, W. C., D. Gladstein, and J. De Laet. 2003. POY, version 3.0.11. American Museum of Natural History.
- Whiting, A. S., J. J. W. Sites, K. C. M. Pellegrino, and M. T. Rodrigues. 2006. Comparing alignment methods for inferring the history of the new world lizard genus *Mabuya* (Squamata: Scincidae). *Mol. Phylogenet. Evol.* 38:719–730.
- Xia, X., Z. Xie, and K. M. Kjer. 2003. 18S ribosomal RNA and Tetrapod phylogeny. *Syst. Biol.* 52:283–295.

First submitted 17 November 2005; reviews returned 26 February 2006;

final acceptance 19 October 2006

Associate Editor: Karl Kjer