Evaluating the Performance of a Successive-Approximations Approach to Parameter Optimization in Maximum-Likelihood Phylogeny Estimation

Jack Sullivan,† Zaid Abdo,†‡ Paul Joyce,†‡ and David L. Swofford§*

*Department of Biological Sciences, †Initiative in Bioinformatics and Evolutionary Studies and Program in Bioinformatics and Computational Biology, and ‡Department of Mathematics, University of Idaho; and §School of Computational Science and Department of Biological Science, Florida State University

Almost all studies that estimate phylogenies from DNA sequence data under the maximum-likelihood (ML) criterion employ an approximate approach. Most commonly, model parameters are estimated on some initial phylogenetic estimate derived using a rapid method (neighbor-joining or parsimony). Parameters are then held constant during a tree search, and ideally, the procedure is repeated until convergence is achieved. However, the effectiveness of this approximation has not been formally assessed, in part because doing so requires computationally intensive, full-optimization analyses. Here, we report both indirect and direct evaluations of the effectiveness of successive approximations. We obtained an indirect evaluation by comparing the results of replicate runs on real data that use random trees to provide initial parameter estimates. For six real data sets taken from the literature, all replicate iterative searches converged to the same joint estimates of topology and model parameters, suggesting that the approximation is not starting-point dependent, as long as the heuristic searches of tree space are rigorous. We conducted a more direct assessment using simulations in which we compared the accuracy of phylogenies estimated using full optimization of all model parameters on each tree evaluated to the accuracy of trees estimated via successive approximations. There is no significant difference between the accuracy of the approximation searches relative to full-optimization searches. Our results demonstrate that successive approximation is reliable and provide reassurance that this much faster approach is safe to use for ML estimation of topology.

Introduction

The importance of incorporating information on the process of nucleotide substitution into comparative analyses of molecular sequences has been acknowledged since the inception of the discipline (e.g., Jukes and Cantor 1969). The reason is well known; multiple substitutions at a site obscure the historical pattern of nucleotide substitutions. Because there are only four possible character states for DNA sequence data, molecular systematists are unable to reassess putative character homologies through the detailed character examination that is available to morphological systematists. Thus, the accurate estimation of homoplasy induced by multiple substitutions is particularly critical to molecular systematics studies and is usually achieved through use of probabilistic models of nucleotide substitution (see Felsenstein 2004 for a recent review). As molecular systematists have begun to understand the influence of such processes as unequal nucleotide composition (Felsenstein 1981), transformation bias (e.g., transition bias; Kimura 1980), and among-site rate variation (e.g., Yang 1994) on phylogenetic analyses of DNA sequence data, models describing the process of nucleotide substitution have become increasingly complex.

Nevertheless, a potential limitation is that maximum likelihood (ML) estimates of substitution-model parameters vary across tree topologies (e.g., Sullivan, Holsinger, and Simon 1996), which usually is the "parameter" of greatest interest. This realization implies that all model parameters (including the rate matrix, base frequencies, rate-heterogeneity parameters, and branch lengths) must be optimized for each topology examined during a tree search (see e.g. Yang, Goldman, and Friday 1995). Thus, it is theoretically possible to identify the combination of topology, branch lengths, and parameters of the substitution model that optimizes the likelihood.

In practice, however, the situation is somewhat different. Because of both the computational burden of optimizing model parameters on each tree and the astronomical number of possible candidate trees for even a modest number of taxa, simultaneous optimization of all model parameters for every tree that is examined during a search is not practical for most data sets. One strategy that has been widely used to circumvent this problem takes advantage of what has been learned from studies of the nature of the variation in estimates of model parameters across topologies (Yang 1994; Sullivan, Holsinger, and Simon 1996; Swofford et al. 1996). The early conclusion of Yang (1994), that estimates of model parameters were highly stable across topologies, now appears not to be entirely true (Sullivan, Holsinger, and Simon 1996). However, the nature of the dependence of parameter estimates on topology is fairly well understood. Yang (1994) suggested that accurate estimates of model parameters may be obtained using any topology that is not "too wrong." Sullivan, Holsinger, and Simon (1996) explored the nature of variation in estimates of two parameters, the gamma distribution shape parameter and the ratio of transition rate to transversion rate, across topologies more thoroughly. They demonstrated that accurate estimates can be obtained by any topology that maintains bipartitions of taxa that the data strongly support (i.e., long internal branches). That is, strongly biased estimates of model parameters are typically obtained only when trees used to estimate these parameters incorrectly break up long internal branches. This point can be illustrated by comparing parameter estimates optimized on 100 random trees versus those optimized on the ML tree and the 100 best-parsimony trees (fig. 1). The estimates from the 100 MP trees form a cloud around the estimates from the optimum topology (fig. 1*A*), whereas the estimates derived from 100 random trees exhibit much a larger range of variation (fig. 1*B*).

Key words: maximum likelihood, models, phylogeny, successive approximations, parameter estimation.

E-mail: jacks@uidaho.edu.

Mol. Biol. Evol. 22(6):1386–1392. 2005
doi:10.1093/molbev/msi129
Advance Access publication March 9, 2005

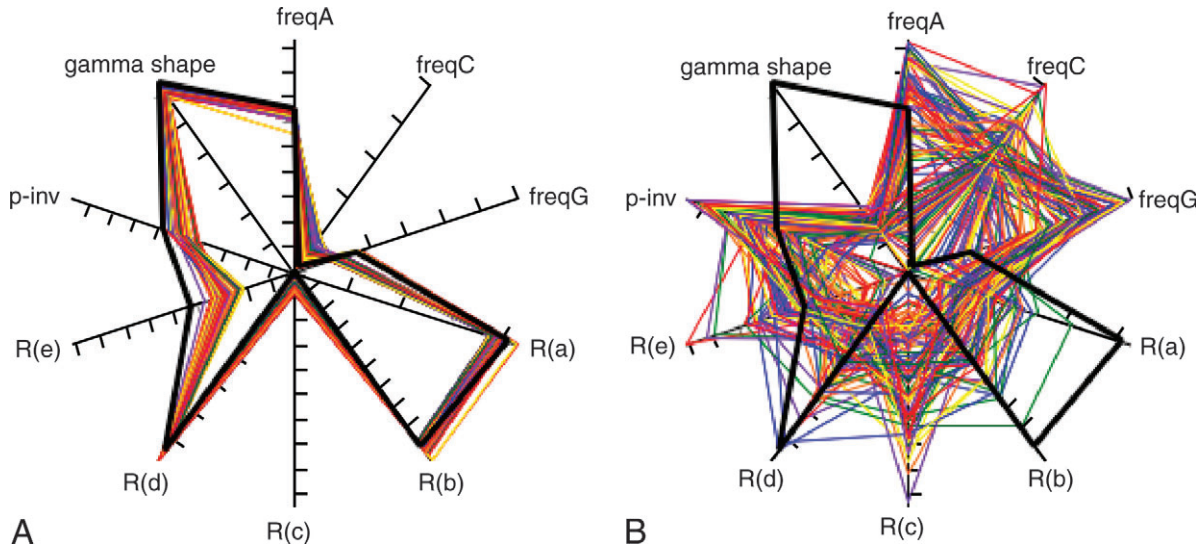

FIG. 1.—The variation in estimates of parameters of the HKY+I+γ model across topologies for the grass waxy data set (each line corresponds to the parameter estimates for one tree). (A) Parameter estimates derived from the ML tree (bold black line) and from the 100 most parsimonious trees (colored lines). (B) Parameter estimates derived from 100 random trees. The axes are scaled identically in both plots, with the minimum value for each parameter on any tree at the center and the maximum value at the tip. There is substantial variation in parameter estimates across random topologies, but all the 100 MP trees provide quite similar estimates as does the ML tree.

Based on these and similar observations, Sullivan, Holsinger, and Simon (1996) suggested that parameters should be estimated on topologies produced by initial simple searches of tree space and Swofford et al. (1996) proposed a search strategy explicitly based on a successive-approximations approach. The logic of this strategy is as follows. First, a set of topologies that will provide reasonable estimates of model parameters is identified through an initial tree search conducted using an approximate approach such as neighbor-joining or parsimony. This set of initial topologies is then used to select a model, usually using some statistical evaluation of a set of candidates using likelihood-ratio tests (e.g., Modeltest, Posada and Crandall 1998) or some alternative (e.g., decision theory; Minin et al. 2003). The process of model selection also provides initial estimates of the parameters of the selected models, and a second tree search is conducted using likelihood as the optimality criterion with a fully defined model of substitution (i.e., parameters of the substitution model are fixed to the previously estimated values). If the ML tree is not a subset of the trees found by the initial search, the new tree should be used as the starting tree for a subsequent iteration. Model parameters are then reoptimized on the new tree, and a second search is conducted; the process continues until the same tree is found in two successive iterations.

Thus, any study that uses automated model-selection procedures (e.g., Modeltest, Posada and Crandall 1998; DT-ModSel, Minin et al. 2003) employs successive approximation, even if it is an abbreviated version. Although this approach has perhaps become the most widely used strategy for estimating ML trees (although very few studies iterate sufficiently), there are unanswered questions about its behavior. Unlike many applications of successive approximations in numerical optimization, it is not guaranteed either to converge to an optimal solution or to provide an indication that convergence will not occur. Some satisfaction that the procedure actually works would be derived by an empirical demonstration that the iterative approach consistently arrives at the same combination of substitution-model parameter values, branch lengths, and topology as those obtained under full optimization on all trees examined during a tree search. Sullivan and Swofford (1997) used the successive-approximations approach on a 16-taxon data set containing mitochondrial DNA (mtDNA) genomes of several mammalian orders (D'Erchia et al. 1996) and demonstrated that the method is able to escape the long-branch attraction problems that plagued parsimony analyses (and likelihood under equal rates models) of that data set. This suggests that the approach may be relatively insensitive to the initial topology used for estimating model parameters. If such starting-point insensitivity applies generally, the iterative approach should provide a very useful speedup for ML estimation of phylogeny. Here, we address the performance of the iterative search strategy in two ways. First, we address the issue indirectly by examining the degree to which the approach is starting-point dependent for several real data sets. We then use one of the data sets to address the issue directly by conducting full, simultaneous-optimization runs, both on real data and on data simulated using conditions estimated from the real data. These analyses demonstrate that the successive-approximation search strategy performs quite well and can be expected to yield results identical to full-optimization searches in most cases.

Data Sets

We examined several data sets in order to assess the starting-point dependence of the successive-approximations approach. These are summarized in table 1 and were

Table 1
Data Sets Used to Assess Starting-Point Dependence of the Iterative Search Strategy

Taxonomic Group	Gene	Genome	Length (bp)	NTax	References[a]	Model
Harvest mice	Cyt b	mtDNA	1,140	29	1	GTR+I+Γ
Sigmodontine rodents	Cyt b	mtDNA	720	22	2	HKI+I+Γ
Collembola (Insecta)	COII	mtDNA	456	19	3	HKY+Γ
Basal vertebrates	rRNA	Nucleus	4,155	9	4	GTR+I+Γ
Mammals	Protein genes	mtDNA	11,571	16	5	GTR+I+Γ
Grass	*waxy*	Nucleus	773	34	6	GTR+Γ

[a] References are as follows: 1, Sullivan, Arellano, and Rogers (2000); 2, Rinehart et al. (unpublished; GenBank accession numbers AY041185–AY041206); 3, Frati et al. (1997); 4, Mallatt and Sullivan (1998); 5, D'Erchia et al. (1996) and Sullivan and Swofford (1997); and 6, Mason-Gamer, Weil, and Kellogg (1998).

chosen to represent a range of divergence times, taxonomic groups, genes, sizes, and sequence characteristics (as indicated by best-fit model). Thus, we have included the Collembola (Insecta) COII data set (mtDNA) of Frati et al. (1997; this was one of the first studies to use the iterative approach); a harvest mouse Cyt b data set (mtDNA) of Sullivan, Arellano, and Rogers (2000); the vertebrate combined 18S and 28S rRNA data (nuclear) of Mallatt and Sullivan (1998); a grass *waxy* data of Mason-Gamer, Weil, and Kellogg (1998); the mammalian mtDNA data of D'Erchia et al. (1996); and a 22-taxon sigmodontine rodent Cyt b data set (Rinehart et al. unpublished; GenBank accession number AY041185–AY041206) that has been used in exploring novel model-selection methods (Minin et al. 2003; Abdo et al. 2005). This last data set is particularly interesting because it has many suboptimal peaks across tree space, and we therefore also conducted full-optimization searches on it (see below).

Starting-Point Dependence

For each data set, iterative searches were conducted with PAUP* (Swofford 1998), either using the model selected by the original authors or a model that we chose using DT-ModSel (Minin et al. 2003; Abdo et al. 2005). In the initial runs, we used an MP tree as the tree from which to derive initial estimates of model parameters. We then conducted 104 replicate successive approximations per data set. Each replicate used a different random tree to initiate estimation of model parameters. A minimum of four and a maximum of six iterations were needed to achieve convergence to a single topology. (A perl script that automates the successive approximations is available at http://www.webpages.uidaho.edu/~jacks/DTModSel.html.) We examined three approaches for the heuristic searches that varied in degree of rigor. For most of the analyses, we used the moderately rigorous implementation (table 2) in which heuristic searches involved construction of a starting tree by stepwise addition of taxa in the order they were found in the data matrix, followed by tree bisection-reconnection (TBR) branch swapping. For the grass *waxy* data set and the sigmodontine Cyt b data sets, we also used the more rigorous strategy (table 2) in which heuristic searches were started from 10 stepwise-addition trees obtained using random-addition sequences, followed by TBR branch swapping. In addition, we applied the very approximate method (table 2) in which we replaced construction of the stepwise-addition starting trees with the tree that was estimated during the immediately previous iteration. By avoiding the stepwise addition prior to branch swapping in each iteration, substantial time savings that are useful or even necessary for the analysis of very large data sets may be achieved. However, not recomputing the starting tree after parameter reoptimization may increase susceptibility to entrapment in local nonglobal optima. In order to address this possibility, we reran the iterative searches on the grass *waxy* data using this third more approximate approach.

Somewhat surprisingly even to us, for each of the six data sets, all the replicate iterative searches converged to the same topology; there appears to be no starting-point dependence for these data sets (fig. 2). Furthermore, in each of these data sets, the chosen tree is identical to the one found when the iteration is started from a better nonrandom tree (e.g., neighbor-joining or parsimony). It is worth noting that the polytomy in the harvest mouse Cyt b data set (fig. 2*C*) represents a zero-length internal branch (i.e., a hard polytomy under the chosen model: HKY + I + Γ) rather than different trees being found across replicates. This polytomy was also found by Sullivan, Arellano, and Rogers (2000) in their analysis using parsimony trees to initiate searches. Also noteworthy is the result for the mammalian mtDNA data set. Sullivan and Swofford (1997) demonstrated that parsimony analyses of this data set is plagued by long-branch attraction that is overcome by ML estimation under an adequate model, even when the parsimony tree is used to derive initial parameter estimates. Long-branch attraction is also avoided by the iterative search strategy when initial parameter estimates are derived using random trees (although more iterations are usually required to achieve convergence).

We did, however, see apparent starting-point dependence in the grass *waxy* data set when we used a single addition sequence to construct a stepwise-addition starting tree during the starting-point dependence analyses. Each analysis using stepwise-addition trees constructed with

Table 2
Three Implementations of the Successive-Approximation Strategies Employed Here

Rigor	Starting Tree	Addition Sequence	Data Sets
High	Stepwise addition	Random: nreps = 10	2 and 6
Moderate	Stepwise addition	As is	All
Low	Previous search	N/A	6

N/A, not applicable.

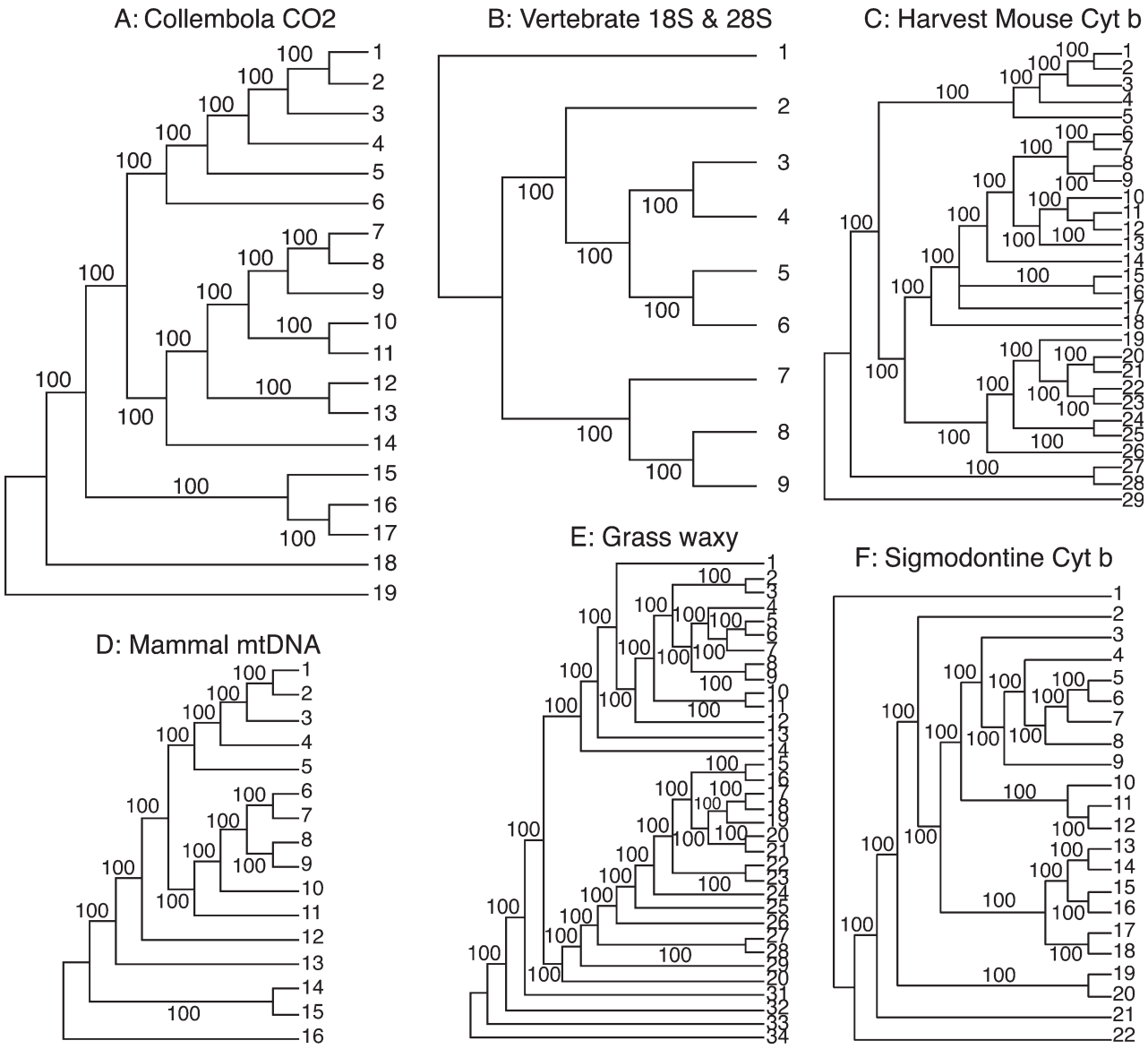

FIG. 2.—Results of starting-point dependence runs for six data sets. For each data set, the majority rule consensus tree is shown. The trees were produced by 104 replicates of the successive-approximation strategy, each started with a different random topology to provide initial estimates of model parameters. We have not been able to find starting-point dependence when the runs involve rigorous heuristic searches of tree space (i.e., stepwise-addition starting trees with 10 random-addition sequences).

"as is" addition sequence in the heuristic searches converged to one of two different trees (not shown). This data set exhibits tree islands; a parsimony search with 100 replicate addition sequences found three islands that contain equally parsimonious trees, whereas there appear to be two islands across the tree space under the likelihood criterion. However, when the starting-point dependence runs involved heuristic searches with 10 random addition sequences, the failure to converge on a single topology regardless of what tree was used to derive initial estimates of model parameters disappeared (fig. 2*E*). Thus, just as for any search strategy, multiple peaks across tree space can trap the successive approximation if the heuristic searches of tree space are not sufficiently rigorous. A similar pattern occurred in the starting-point dependence replicates of the sigmodontine Cyt b data set; when the heuristic searches used a single (as is) addition sequence to construct stepwise-addition trees, not all the replicates converged to the same tree. However, the more rigorous analyses (10 random-addition sequences) showed 100% convergence to the same tree, regardless of what tree was used to initiate the successive searches.

When we used trees from previous searches for branch swapping, the convergence properties of the successive approximation deteriorated further. The results indicate that, although most of the replicate starting-point dependence runs converged to the same topology (the best estimate of the ML tree), a few of the replicates with no stepwise-addition/random-addition sequence became trapped in local optima (fig. 3). The searches converged to one of six trees.

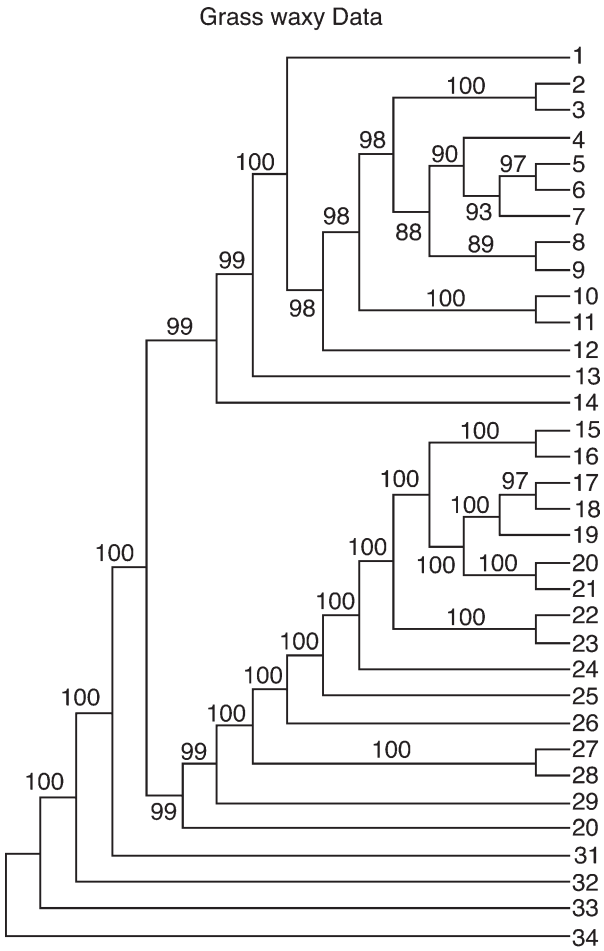

FIG. 3.—The starting-point dependence runs on the grass *waxy* data set with the least rigorous tree searches. During the heuristic searches, the tree found in the previous iteration was used as a starting tree for TBR branch swapping. There is starting-point dependence when this approximation is used.

This indicates that successive approximations should be run with the most rigorous heuristic searches possible (i.e., stepwise-addition with random-addition sequences) for each tree search in the iteration.

Therefore, the approximation appears to be robust to starting points. Even when there are islands in tree space, as in the grass *waxy* data set and sigmodontine Cyt b data, the successive-approximations strategy can avoid becoming trapped on suboptimal peaks, as long as each of the iterative heuristic searches is rigorous. Given the large amount of variation in parameter estimates across random trees (e.g., fig. 1), it is quite remarkable that successive-approximation ML searches under such initially misspecified models exhibit such excellent convergence properties.

Simultaneous Optimization Versus Successive Approximation

In the simultaneous-optimization search on the sigmodontine Cyt b data, we used 10 random-addition sequence replicates and constructed a starting trees stepwise addition under the HKY + I + Γ model (selected using DT-ModSel, Minin et al. 2003). We conducted full TBR branch swapping, and because likelihood settings in PAUP* were set to estimate all parameters, all substitution-model parameters and branch lengths were optimized on each tree examined, both during construction of starting trees and during branch swapping. We limited TBR branch swapping to a maximum of 10 trees (i.e., MAXTREES = 10). We then compared the simultaneous-optimization tree to that found using the most rigorous implementation of the successive-approximations strategy (table 2). The simultaneous-optimization run took 776 h of CPU time on a 750-MHz Sun Workstation and found a single ML tree, with a likelihood score of ln L = −6,448.7137. This is apparently a difficult tree to find because TBR branch swapping was able to find this tree in only one of the 10 random-addition sequence replicates. In comparison, the most rigorous successive-approximations approach was able to find the same tree in just 92 min of CPU time on the same processor.

We also used this data set to establish conditions for simulation of 1,000 replicate data sets. In this simulation, our approach was to use a much more complex model to generate the data than we used for analysis. This was accomplished by applying a distinct GTR+I+Γ model to each codon position in the real data, with branch lengths for each codon position estimated independently. We then simulated each codon position separately, using the codon-position–specific GTR+I+Γ, and concatenated the positions into single replicate. Simulated data were thus generated with a nucleotide substitution model with 30 parameters (9 independent base frequencies, 15 transformations rates, 3 gamma shape parameters, and 3 invariable sites parameters). We then selected a model for each replicate using DT-ModSel (Minin et al. 2003) and ran both successive-approximation and full-optimization searches for each replicate data set. The simulations were run on a 100-node Beowulf cluster housed at the University of Idaho Bioinformatics Core Facility (each node has four 2.0-GHz Xeon processors). We evaluated the accuracy of estimation using full optimization to that of estimation using the approximate strategy using the distribution of symmetric difference distances (SDDs; Robinson and Foulds 1982) between the ML estimates and the true tree that was used to simulate the data.

The two approaches are extremely similar in their accuracy (fig. 4). The distribution of SDDs for the full optimization had a mean of 5.66 (variance = 12.69), and in 7.6% of replicates the full-optimization searches succeeded in finding the true tree. For the successive approximations, the distribution of SDDs had a mean of 5.59 (variance = 12.32), and in 7.9% of the replicates the iterative searches were able to identify the true tree. There is no evidence that the distributions of SDDs are different ($P = 0.66$). While neither of these approaches succeeded frequently in identifying precisely the true tree, this is an extremely difficult phylogenetic problem, and phylogeny estimation is not hindered by using successive approximation rather than full optimization.

Conclusions

The analyses we present here are extremely encouraging. The successive approach has become the default

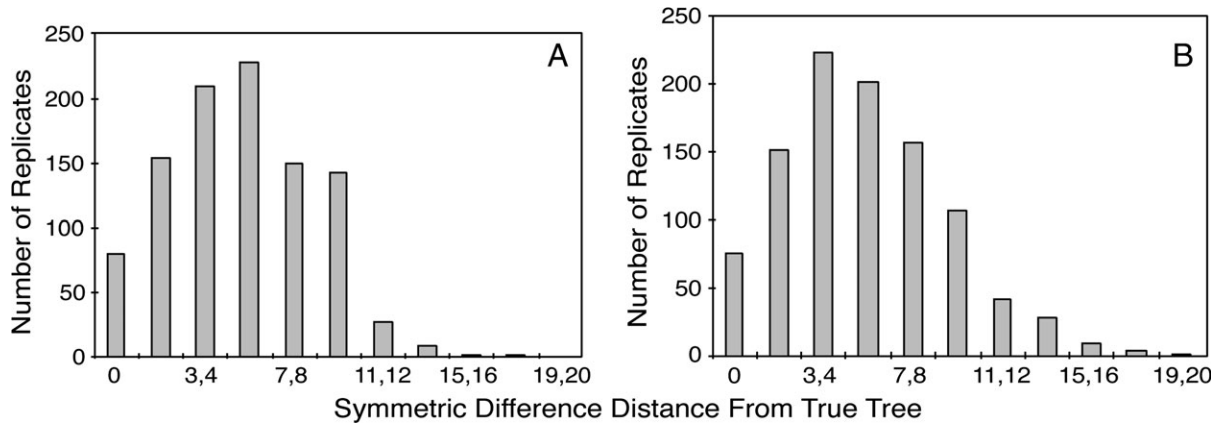

FIG. 4.—Comparison of the accuracy of (*A*) full-optimization searches versus (*B*) successive approximations. One thousand replicate data sets were generated using a separate GTR + I + Γ model applied to each codon position, and each data set was analyzed using the most rigorous approximate search as well as a full-optimization search (where all model parameters were optimized on each tree examined during a single heuristic search). Resulting trees were compared to the true tree used to simulate the data using the SDD. There is no difference in the accuracy of the two methods ($t = 0.436$: $P = 0.66$).

method for estimating phylogenies under likelihood, but this is the first study that actually examines how well the method approximates the exact approach of optimizing all model parameters on all trees examined during a search of tree space. Under a wide variety of conditions (represented by six disparate data sets), there appears to be no starting-point dependence to successive approximation, as long as the heuristic searches of tree space are sufficiently rigorous. Even more encouraging, the comparison of accuracy of approximation versus full-optimization searches in the simulation indicates that successive approximations are equally accurate to the full-optimization searches. There are several novel approaches to generating a good approximation to the ML tree, including PHYML (Guindon and Gascuel 2003) and IQPNNI (Vinh and von Haeseler 2004). These are especially useful for large data sets, and in many cases a good approximation will be sufficient. However, if one is interested in statistical tests of phylogenetic hypotheses from a frequentist framework, a good estimate of optimal trees (both constrained and unconstrained) assumes much greater importance. While the results reported here may not be universal because we simulated sequences on a single tree shape (albeit a very difficult one), they are sufficiently general to provide confidence that use of the common approximate strategy will not unacceptably compromise ML estimation of phylogeny.

Acknowledgments

This research is part of the University of Idaho Initiative in Bioinformatics and Evolutionary Studies (IBEST). Funding was provided by National Science Foundation (NSF) EPSCoR grant EPS-0080935 (to IBEST), NSF Systematic Biology Panel grant DEB-9974124 (to J.S. and D.L.S.), NSF Probability and Statistics Panel grant DMS-0072198 (to P.J.), NSF EPSCoR grant EPS-0132626 (to P.J. and Z.A.), NSF Population Biology Panel grant DEB-0089756 (to P.J.), and National Institutes of Health (NIH) NCCR grant NIH NCCR 1P20PR016448-01 (to IBEST: PI, L. J. Forney). We thank T. Reinhardt, R. Graham, and H. Wichman for permission to use unpublished sigmodontine Cyt b sequences, which were generated with funds from NIH grant GM38737 (to H. Wichman). We also thank Ken Blair, our System Administrator, for maintaining the University of Idaho Beowulf clusters used in this project.

Literature Cited

Abdo, Z., V. Minin, P. Joyce, and J. Sullivan. 2005. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. Mol. Biol. Evol. **22**:691–703.

D'Erchia, A. M., C. Gissi, G. Pesole, C. Saccone, and U. Arnason. 1996. The guinea pig is not a rodent. Nature **381**: 597–600.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Mass.

Frati, F., C. Simon, J. Sullivan, and D. L. Swofford. 1997. Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. J. Mol. Evol. **44**:154–158.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**:696–704.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. Munro, ed. Mammalian protein metabolism. Academic Press, New York.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

Mallatt, J., and J. Sullivan. 1998. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. Mol. Biol. Evol. **15**:1706–1718.

Mason-Gamer, R. J., C. Weil, and E. A. Kellogg. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. **15**:1658–1673.

Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. **52**:1–10.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Biofinormatics **14**:817–818.

Robinson, D. F., and L. R. Foulds. 1982. Comparison of phylogenetic trees. Math. Biosci. **53**:131–147.

Sullivan, J., E. Arellano, and D. S. Rogers. 2000. Comparative phylogeography of Mesoamerican highland rodents: concerted versus independent response to past climatic fluctuations. Am. Nat. **155**:754–768.

Sullivan, J., K. E. Holsinger, and C. Simon. 1996. The effect of topology on estimates of among-site rate variation. J. Mol. Evol. **42**:308–312.

Sullivan, J., and D. L. Swofford. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. **4**:77–86.

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10a. Sinauer Associates, Sunderland, Mass.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. Hillis, C. Moritz, and B. Mable, eds. Molecular systematics. 2nd edition. Sinauer Associates, Sunderland, Mass.

Vinh, L. S., and A. von Haeseler. 2004. IQPNNI: moving fast through tree space and stopping in time. Mol. Biol. Evol. **21**: 1565–1571.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39**:306–314.

Yang, Z., N. Goldman, and A. Friday. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Syst. Biol. **44**:384–399.

Arndt von Haeseler, Associate Editor

Accepted February 28, 2005