ARTICLE

Application and Accuracy of Molecular Phylogenies

David M. Hillis, John P. Huelsenbeck, Clifford W. Cunningham

Molecular investigations of evolutionary history are being used to study subjects as diverse as the epidemiology of acquired immune deficiency syndrome and the origin of life. These studies depend on accurate estimates of phylogeny. The performance of methods of phylogenetic analysis can be assessed by numerical simulation studies and by the experimental evolution of organisms in controlled laboratory situations. Both kinds of assessment indicate that existing methods are effective at estimating phylogenies over a wide range of evolutionary conditions, especially if information about substitution bias is used to provide differential weightings for character transformations.

Over the past few decades, biologists from many disciplines have turned to phylogenetic analyses to interpret variation in biological systems (*1*). This increased interest in evolutionary history has developed partly in response to a new appreciation of the importance of understanding evolutionary constraints when interpreting biological variation and partly in response to developments in phylogenetic methodology. Three developments in particular have been critical to the success of the field: (i) the development of objective criteria and algorithms for discriminating among potential phylogenies, (ii) increased computational power to implement phylogenetic algorithms, and (iii) a rapid increase in the data available for inferring phylogenies, especially from molecular investigations (*2*). As a result of these developments, applications of phylogenetic analysis span the range of biological diversity from questions about the history of life (*3*) to studies of the epidemiology of acquired immune deficiency syndrome (AIDS) (*4*). However, the success of these applications depends on the accuracy of the inferred phylogenies, so it is necessary to ask how well the methods work and to identify the conditions under which they may fail.

The accuracy of methods of phylogenetic analysis can be assessed by the examination of either numerical simulations of phylogenies or phylogenies of organisms whose evolutionary history has been observed directly. Numerical simulations assume a particular model of evolution and then generate characters (typically, nucleotide sequences) according to the model and to a given phylogeny. Thus, an investigator can generate many replicate data sets under specified conditions in order to compare the performance of competing methods. The analysis of known phylogenies adds a reality check to the simulation studies: The history of the lineages is known (or, ideally, controlled by the investigator), but the organisms evolve under real biological constraints rather than idealized model conditions. Known phylogenies may involve laboratory or cultivated strains whose history has been recorded (*5*) or lineages that have been manipulated under controlled experimental conditions for the purpose of generating testable phylogenies (*6, 7*).

The numerical simulation and experimental phylogeny approaches are largely complementary, and both kinds of studies are necessary to evaluate methods of phylogenetic analysis effectively. Simulations can be used to explore virtually any conceivable phylogeny, and phylogenies can be replicated with speed and ease. The primary limitation of numerical simulations is that they always include gross simplifications of biological processes. For instance, most simulations assume that nucleotide positions evolve independently of one another, even though several causes of non-independence have been identified (*8*). Many simulations also assume simple one- or two-parameter substitution models; for instance, all possible substitutions may be assumed to be equally probable (a one-parameter model), or separate probabilities of substitution may be assigned to transitions and transversions (a two-parameter model). However, real substitution biases are known to be much more complex (*9*). Although these complexities can be added to simulation studies, there is rarely sufficient knowledge to estimate the extent of the influence of factors such as non-independence among nucleotide positions or variance of rates of evolution across nucleotide positions. Therefore, results from simulation studies need to be compared to results from studies of real biological organisms to determine the effects of the simplifying assumptions. If results from simulations can be replicated with experimental systems, then greater faith can be placed in the simulation results. However, if departures from the simulation results are discovered, then the processes that are responsible for the differences can be identified and the simulations can be improved. The simulations are likely to suggest conditions that are of interest in the experimental phylogenies, and the experimental phylogenies can provide a test of the simulation results. Thus, a combination of the two approaches is the most effective way to evaluate the performance of methods of phylogenetic analysis (*10*).

Simple Evolutionary Models

Most simulated phylogenies assume a simple one- or two-parameter model of evolution and then test the ability of various methods to reconstruct the evolutionary history of lineages generated under the assumed model (*11, 12*). Several methods are known to be consistent (at least for simple tree topologies) for data generated under such models, which means that they converge on the correct answer, given infinite data. In general, most of the commonly used methods are consistent if corrections are made for superimposed changes (such as multiple substitutions at a single nucleotide site) in accord with the model of evolution used (*13*). For instance, most pairwise distance methods (except the UPGMA method) are consistent under the Jukes-Cantor one-parameter model of evolution if Jukes-Cantor distances are used to infer the phylogeny (*12, 14*). Character-based methods such as parsimony can also be made consistent by using a Hadamard transformation to correct the data (*13*). However, the fact that a method is consistent indicates only that it will converge on the correct answer when given unlimited data, so it is necessary to do power analyses in order to compare the performance of competing methods, given finite data sets.

A common objection made to simulation studies is that it is easy to bias the results in favor of almost any method by choosing conditions to simulate that are most favorable to that method (*15*). Such biases can be avoided only by exhaustively exploring the potential parameters of any given problem. As an example, consider one of the most commonly simulated cases: a simple four-taxon unrooted tree, in which the five lineages (four peripheral branches and a central branch) are evolving at two different rates (Fig. 1). Felsenstein (*16*) used a tree of this type to demonstrate that some methods of phylogenetic reconstruc-

The authors are in the Department of Zoology, University of Texas, Austin, TX 78712, USA.

tion are inconsistent when two of the opposing peripheral branches are evolving much more rapidly than are the remaining three branches. Given a model of evolution (for example, the Kimura two-parameter model, which allows for independent substitution rates for transitions and transversions) (*17*), and given two rates of evolution (one rate for two of the opposing branches and a second rate for the remaining three branches), the universe of possible trees can be examined in a two-dimensional graph (Fig. 1). Instantaneous substitution rates can be varied from zero to infinity along each of the axes, and sequences can be generated in accord with the model of evolution. A power analysis is conducted by generating sequences of given finite length and then inferring the trees from the sequences by the use of competing methods.

Figure 1 shows a power analysis for three common methods of phylogenetic inference and the effects of two common methods of data transformation under the model of evolution outlined above (*18*). For nontransformed data, all three methods are inconsistent in parts of the graph space; use of Kimura-corrected distances (which exactly match the model of evolution) makes the neighbor-joining method consistent across the graph (*12*). Another common type of data transformation involves character weighting (*19, 20*). In character methods such as parsimony, differential weights are often assigned to the different character-state changes, depending on their observed frequency of occurrence. Thus, in the Kimura model simulated in Fig. 1, transitions are 10 times more likely to occur than are transversions, so the weighted-parsimony analysis weights the transversions 10 times more heavily than transitions (in practice, a wide range of weights of transversions over transitions produces identical results) (Fig. 2). Such weighting is not equivalent to transforming the data to account for superimposed changes, so weighted parsimony is not consistent across the entire graph space (*12*). However, the power analysis shown in Fig. 1 indicates that weighting of characters has a much greater effect on performance than does correction for superimposed changes, especially at high rates of change. Although the weighted-parsimony method is more likely to be misleading at extreme differences in the two rates (that is, in the upper left corner of the graph space), it is more likely to find the correct tree at high rates of change (Fig. 1). The Kimura corrections do improve the performance of the neighbor-joining method in regions that are inconsistent for the uncorrected data but do not improve performance when rates are uniformly higher (as does character weighting). The Kimura corrections actually reduce the performance of distance methods under conditions of equal rates of change (Fig. 1).

Some authors have argued that methods such as parsimony should be avoided because they are inconsistent for some trees (for example, those in the upper left corner of the graphs in Fig. 1) when they evolve under simple models of evolution (*21*). However, all methods become inconsistent for some trees when their assumptions are violated (*12*), and the cost of complete consistency under simple models of evolu-

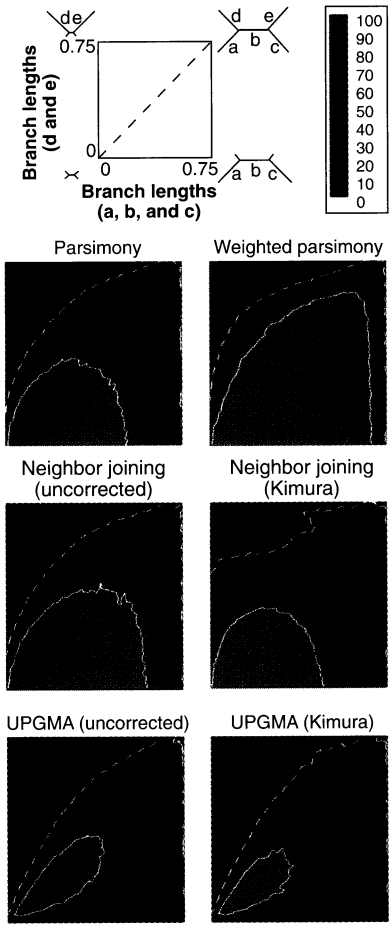

Fig. 1. Performance of three methods of phylogenetic analysis on the basis of simulation of four-taxon trees under the Kimura model of evolution (*18*). Two rates of evolution were simulated: one rate for branches a, b, and c (horizontal axis of each graph) and a second rate for branches d and e (vertical axis). The diagonal (dashed line, top left) represents equal rates of evolution along all lineages. Branch lengths are shown in expected frequency of divergent nucleotides at the two ends of the respective branches. At infinite rates of change, DNA sequences with equal base compositions are expected to differ at 75% of their positions. Blue indicates that the method estimates the correct tree a high percentage of the time under the simulated conditions; red indicates poor performance of the method (see color bar, top right). The solid white lines circumscribe the regions in which each method estimates the correct tree over 95% of the time. In the regions above the dashed white lines, the methods estimate the correct tree less than one-third of the time (a rate worse than that obtained by choosing a tree at random). The three colored graphs on the left were based on nontransformed data; the three graphs on the right show the effects of character-state weighting (for parsimony, top) and distance correction (for neighbor joining and UPGMA, middle and bottom).

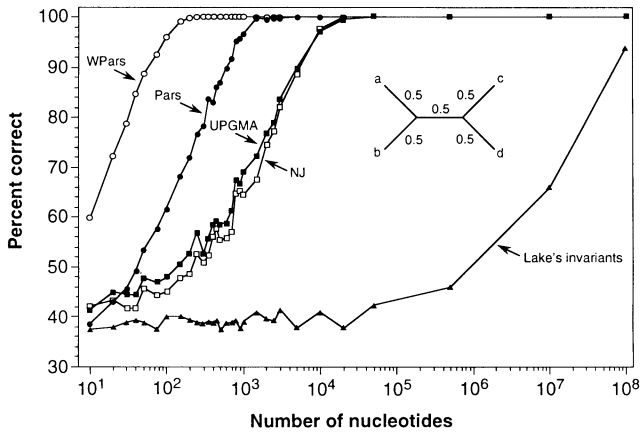

Fig. 2. Efficiency of five methods of phylogenetic analysis for a four-taxon tree with equal rates of evolution, evolving under a Kimura model of evolution and a 10:1 transition:transversion ratio. The branch lengths shown on the tree indicate that 50% of the nucleotide sites are expected to change along each branch. Although all five methods are consistent under these conditions (they all eventually converge on the correct solution), the methods differ markedly in the number of nucleotides needed to find the correct solution. All points are based on 1000 simulated trees. WPars is weighted parsimony (*45*) (any weighting of transversions over transitions from 5:1 to infinity produces results indistinguishable from those shown); Pars is uniformly weighted parsimony (*45*); NJ is neighbor joining with Kimura distances (*38*); UPGMA is the unweighted pair-group method of averages with Kimura distances (*40*); Lake's invariants is the method also known as evolutionary parsimony (*22*).

tion can be high. Figure 2 shows a set of simulations from a single point of the tree space described in Fig. 1, where all branches have a length of 0.5 (that is, 50% of the nucleotide positions are expected to differ at either end of each branch) (*18*). At these high but equal rates of evolution across all branches, all methods of phylogenetic analysis are consistent, but they differ dramatically in the number of nucleotides that are required to find the correct tree with high probability. For instance, Lake's method of invariants (*22*), which is preferred over parsimony by some workers because of its consistency under the Kimura model for four-taxon trees (*21*), requires more nucleotides to find the correct solution (at a probability greater than 99%) than are present in many genomes. Neighbor joining with Kimura distances, another consistent method, is much more efficient but still requires approximately 50,000 nucleotides to find the correct tree with equally high probability. In contrast, uniformly weighted parsimony requires only 2000 nucleotides and weighted parsimony requires only 200 nucleotides to achieve the same performance. Because of the current limitations on collection of sequence data, these differences in efficiency need to be weighed against considerations of consistency when an analytical method is chosen.

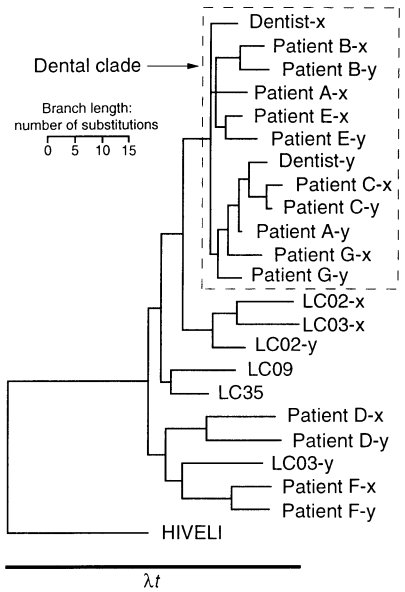

Fig. 3. Estimated phylogeny of HIV sequences from a Florida dentist, seven of his HIV-seropositive patients, and four individuals from the local population (LC) whose HIV sequences were most similar to those of the dentist (*47*). The outgroup (HIVELI) is an African HIV-1 sequence. Two divergent HIV sequences (labeled x and y) were examined from most individuals. The dental clade consists of patients whose HIV sequences are closer to those of the dentist than to those of any of the local controls. Branch lengths are proportional to the number of inferred evolutionary changes averaged across all possible character reconstructions (from *MacClade*) (*20*). The bar labeled λt is the distance from the root to the most divergent tip; it also indicates the divergence scale for the simulations in Fig. 4.

Simulations such as those in Figs. 1 and 2 provide comparisons of methods under idealized conditions, when the assumptions of some methods match the evolutionary model exactly. Such simulations are useful for identifying general patterns in the performance of various methods or types of data transformation. For instance, the simulations in Figs. 1 and 2 show that appropriate corrections for superimposed changes can increase the performance of methods when the rates of evolution are highly variable but that differential weighting of character states is more effective in increasing performance at high rates of evolution.

Although simple trees and simulations of simple evolutionary models can provide insights into the performance of different methods, the generality of the conclusions is not always obvious. Real evolving nucleotide sequences differ from these simple simulations in many important ways, the most obvious of which are that substitution biases are unlikely to fit a simple one- or two-parameter model very closely and that the phylogenies that are of interest are rarely as simple or as uniform as the four-taxon case. Therefore, it is necessary to increase the complexity of the evolutionary models and to model particular phylogenies that are of interest.

Modeling Complex Phylogenies: HIV and the Florida Dental Case

Consider the recently reported case of the Florida dentist suspected of transmitting human immunodeficiency virus (HIV) to some of his patients (*4*). After a probable case of HIV transmission from dentist to patient was identified, the dentist wrote an open letter to his other patients in which he encouraged them to be tested for HIV infection. To date, 10 seropositive patients have been identified (*23*). However, some of these patients have other risk factors for HIV, so the question arises as to which, if any, of the patients were infected by the dentist rather than from another source. Sequences of the *gp120* gene of HIV encoding the C2-V3 domains were obtained from DNA amplified from peripheral mononuclear cells from the dentist, from seropositive patients, and from control individuals from the local population. Phylogenetic analyses of these sequences are consistent with the dental-transmission hypothesis for 6 of the 10 patients (*4, 23*). In the original study, a single evolutionary lineage was identified that contained only viruses from the dentist and five patients he purportedly infected (the sixth infected patient was discovered later); this lineage has been termed the dental clade (Fig. 3). The inferred phylogeny was consistent with the independent epidemiologic data: The dental clade contained all of the patients without any other identified risk factors for HIV and excluded all of the patients with other confirmed risk factors. However, the study reporting these results has been criticized because of questions concerning the reliability of the inferred dental clade (*24*), so it is of interest to assess the probability of correctly inferring this phylogeny.

Table 1. Number of nucleotide substitutions across the tree shown in Fig. 3, as estimated from HIV sequence data (*4*). Values were derived from the averages across all equally parsimonious character-state reconstructions by use of the program MacClade (*20*); minimum and maximum number of substitutions across reconstructions are shown in parentheses.

From	To			
	A	G	C	T
A	—	80.00 (66, 94)	40.62 (29, 53)	17.67 (13, 22)
G	41.90 (28, 56)	—	4.46 (3, 6)	3.00 (0, 6)
C	23.08 (12, 35)	1.93 (1, 3)	—	21.83 (15, 29)
T	10.34 (5, 15)	12.67 (9, 16)	23.50 (16, 31)	—

Even a cursory examination of HIV evolution shows an important departure from the assumptions of the Kimura model of evolution (*25*) (Table 1). For instance, although A to G nucleotide transitions are the most common type of change observed in HIV sequences, A to C transversions are more common than C to T transitions and all of these types of change are several times more common than C to G transversions. Moreover, the substitution matrix is highly asymmetric (Table 1). Therefore, to evaluate the ability of various methods to infer HIV phylogenies, we need to take the relative frequencies of all 12 types of nucleotide change into account.

We can take the estimated phylogeny of the HIV viruses (Fig. 3) and the observed substitution matrix (Table 1) as our model and simulate the phylogeny at varying rates of overall change (*26*). Figure 4 shows the percentage of branches in the simulated trees that are inferred correctly by several common methods of analysis (*27*), as well as the probability of resolving the dental clade, as the overall amount of change is varied from one-fifth to 20 times the amount observed in the original study. At the level of change seen in the original study, 90 to 94% of all the branches in the tree were resolved in the simulations by

every method except UPGMA, and the dental clade was resolved 100% of the time with every method except UPGMA. The positive effects of character weighting are not seen unless the overall rate of change is much greater than was originally observed (Fig. 4), which indicates that the uniform weighting used in the original study was justified. The 100% recovery of the dental clade in the simulations provides significant support for a phylogeny that is consistent with the dental transmission hypothesis.

Part of the reason that the dental clade is so easy to infer is that the HIV sequences were obtained within 2 to 3 years after the transmission from dentist to patients (*28*). If the time lag had been greater, it is likely that the probability of successfully recovering the dental clade would have been lower. This effect is shown in Fig. 5, in which the terminal lineages of the model phylogeny were extended by simulation for up to 20 additional years. Six years after transmission, the probability of recovering the dental clade drops below 95% for all the methods. Thus, despite the high confidence that can be placed in the veracity of the dental clade ($P > 99\%$ for every method except UPGMA), the relevant part of the phylogeny could not have been recovered with high confidence by use of these sequences if the problem had not been detected soon after transmission.

In general, the clustering methods based on Kimura distances (UPGMA and neighbor joining) perform more poorly than the parsimony methods, especially with higher levels of divergence. At the maximum level of divergence simulated (about 20 times the observed level), all of the parsimony methods (as well as UPGMA) recovered the dental clade 100% of the time, whereas neighbor joining found this clade only 79% of the time (Fig. 4). Overall, neighbor joining recovered 72.6% of the clades and UPGMA recovered 71.1% at the high level of divergence, compared with 83.4 to 89.4% clade recovery for the various parsimony methods. At the lowest level of divergence simulated (one-fifth of that observed), neighbor joining did about as well as the parsimony methods but UPGMA did considerably worse (Fig. 4). In the simulated extension of the terminal lineages (Fig. 5), the performance of UPGMA (and to a lesser extent neighbor joining) fell off more quickly than that of any of the parsimony methods. This decrease in performance is likely to be at least partly the result of departures by HIV from the Kimura model of evolution in the case of neighbor joining, and by the sensitivity of the method to departures from equal rates of change in the case of UPGMA.

These simulations of the Florida dentist case still do not take all of the complexities of HIV evolution into account. For instance, the simulations do not model the changes in HIV that result from the duration of infection (*29*), pressure from the host's immune system (*30*), stage of the disease (*31*), or therapy (*32*), any one of which could result in parallel evolution across lineages and thereby reduce the probability of correct estimation of phylogeny. There also are likely to be interactions among sites in the HIV genome that reduce their evolutionary independence. Rates of change also vary considerably across sites in the HIV genome, although the parsimony methods appear to be effective across a wide spectrum of rates of change (Fig. 4). However, it is unlikely that any simulation will ever be complex enough to incorporate all the details of the evolutionary process. Even if such knowledge were readily available, a simulation that incorporated all the relevant details would have to be as complex as a real organism. Therefore, one option is to build phylogenies through controlled laboratory evolution of actual organisms.

Experimental Phylogenies with Real Organisms

The observed evolution of HIV sequences over the course of just a few years (Fig. 3) suggests that controlled phylogenies of viruses could be generated in the laboratory and used to test methods of phylogenetic analysis. Indeed, such phylogenies have been produced (*6*). In experimental phylogenies, the shape of the phylogeny (order of branching events and time between branching events) and some details such as population size and mutagenic environment are controlled by the investigator, but the evolutionary changes incorporated depend on the constraints imposed by the experimental organisms. Experimental phylogenies can provide a reality check on simulation studies and provide a test of the fallibility of analysis methods.

Figure 6 shows an experimental phylogeny derived from bacteriophage T7 and compares the observed amounts of change in nucleotide sequences to those inferred

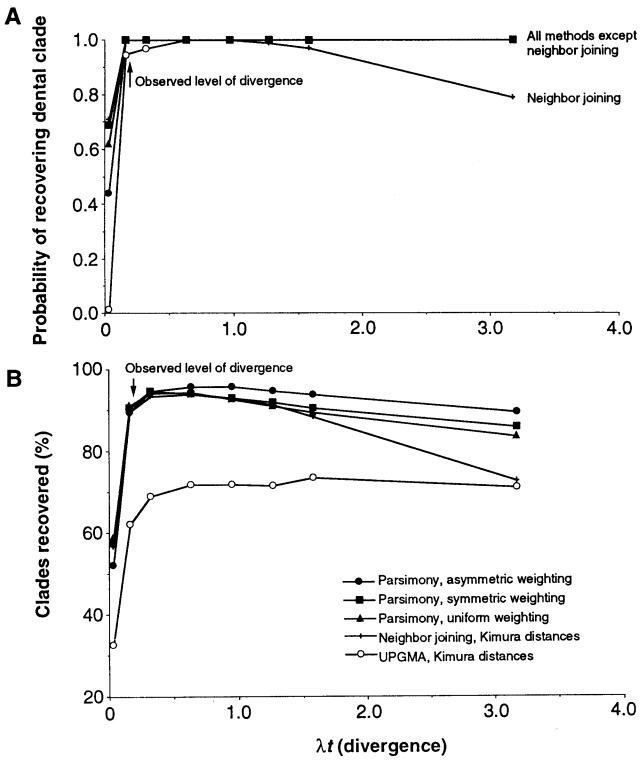

Fig. 4. Simulations of the phylogeny shown in Fig. 3 at various levels of overall divergence (λt, the product of substitution rate and time). One hundred simulated data sets were generated as described (*26*). (**A**) Probability of recovering the dental clade at different levels of divergence for five methods of analysis. (**B**) Overall percentage of the clades in Fig. 3 that were recovered by each method. The level of divergence in the observed data is indicated by arrows, the location of which was estimated on the basis of the length of the original tree as compared with the lengths of the simulated trees.

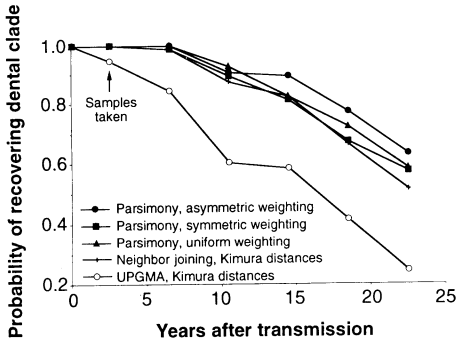

Fig. 5. Simulations of the phylogeny shown in Fig. 3, with the terminal branches extended in time up to approximately 22 years after the last estimated transmission of HIV from the dentist to a patient. Six years after transmission, the probability of detecting the dental clade falls off rapidly, which indicates that the success of the study depended on rapid investigation of the problem.

from parsimony. In the original study, the phylogeny of these lineages was inferred from restriction site maps of the entire viral genome, and all methods tested were successful at recovering the known phylogeny (6). The methods differed significantly in their ability to recover the branch lengths of the phylogeny (7), and the study also indicated a high degree of success in the reconstruction of ancestral restriction maps (>98% accuracy). However, the study did not discriminate among methods on the basis of their ability to find the correct order of branching events, because all methods found the correct tree.

We have now investigated this phylogeny, using two additional data sets: restriction fragments and DNA sequences (33). Some authors recommend using the presence or absence of restriction fragments (rather than the presence or absence of restriction sites) to infer phylogenies, because it is much easier to collect restriction fragment data than restriction site data (34). However, restriction fragments do not evolve independently (a single site gain results in the loss of one fragment and the gain of two others), and deletions can affect the fragments produced by many restriction enzymes simultaneously. Because of these problems, many authors argue that restriction site data should be preferred to restriction fragment data (35). This position is supported by the experimental T7 phylogeny, because all methods estimated an incorrect phylogeny when using high-resolution restriction fragments, but they estimated the correct phylogeny when using restriction sites. This difference in the performance of analyses based on the two types of data has not been apparent in simulation studies, possibly because simulation studies rarely include deletions in their models of evolutionary change.

The sequence data consist of 1091 base pairs across four genes of T7 (36). There are only 63 variable sites across the sequences, or about one-third as many variable characters as are present in the restriction site data (6). Competing methods do not perform as well with the sequence data as they do with the restriction site data. With the sequence data, only parsimony and weighted parsimony estimate the correct tree, although a second tree (that differs by one branch) is equally parsimonious. Maximum likelihood (37), neighbor joining (38), the Fitch-Margoliash method (39), and UPGMA (40) each estimate a single, incorrect tree that differs from the correct tree by one branch rearrangement. The less accurate overall performance of all methods with the sequence data does not necessarily imply that sequences are less reliable than restriction sites for inferring phylogeny, because there are fewer variable sites in the sequence data set. However, if bootstrap samples equal in size to the sequence data set are selected from the complete restriction site data and compared to bootstrap samples of the sequence data, then the restriction site data do appear to be somewhat more reliable for inferring phylogeny for most methods (maximum likelihood is the exception) (Fig. 7). A possible explanation lies in the non-independent evolution of some nucleotides within genes (7, 8); the variable restriction sites are distributed across the entire T7 genome and therefore are more likely to vary independently of one another. For these data, differential weighting of character states does not improve phylogenetic resolution, because rare substitutions are restricted to single terminal lineages and therefore are uninformative under the parsimony criterion. On the basis of the simulated HIV phylogenies discussed earlier, the beneficial effects of weighting are expected only at higher rates of evolution than were observed. The relatively poor performance of maximum-likelihood estimation on the restriction site data may be because the strongly biased substitution matrix violates the assumptions of the method (7).

Clearly, it will be necessary to construct additional experimental phylogenies that are based on other tree topologies and experimental conditions so that the generality of the results can be checked. In particular, predicted conditions of inconsistency need to be examined experimentally. Nonetheless, there is a high degree of correspondence between the results from simulations and the experimental phylogenies, although the experiments suggest additional complexities that need to be added to simulations. For instance, the comparison of restriction site data with restriction fragment data indicates the need to incorporate insertion-deletion events into simulations as well as methods of analysis, and the sequence analyses confirm the importance of accounting for non-independence among nucleotide sites. In general, however, the experimental phylogenies confirm the relatively high levels of performance of the various methods of phylogenetic analysis under realistic conditions.

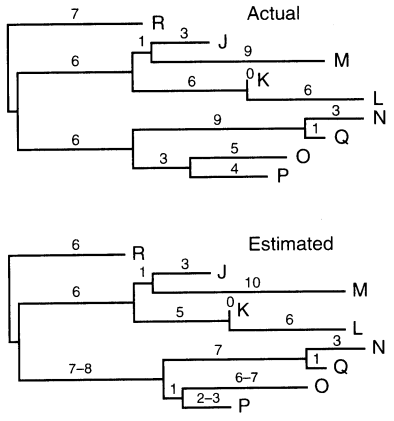

Fig. 6. Comparison of an observed phylogeny of viruses derived from bacteriophage T7 with an estimated phylogeny from the parsimony method, on the basis of analysis of the terminal sequences (J through R). The numbers above the branches indicate the actual or estimated number of substitutions that occurred along the respective lineages. The actual numbers of substitutions were determined by sequencing the ancestral viruses. Ranges of values on the estimated tree indicate that multiple, equally parsimonious reconstructions of character states are possible.

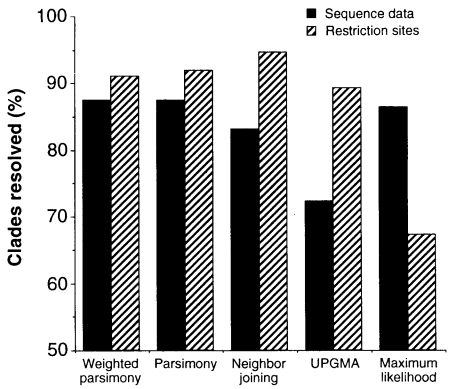

Fig. 7. Comparison of phylogenetic analyses of the viral lineages derived from bacteriophage T7, on the basis of 1000 bootstrap samples of DNA sequences and 1000 bootstrap subsamples of the restriction site data that have the same number of variable sites as are in the sequence data. All methods found the correct tree with the complete restriction site data set; only parsimony and weighted parsimony found the correct tree with the complete sequence data set.

Conclusions

Both simulation studies and experimental phylogenies indicate that many methods of phylogenetic analysis are powerful enough to reconstruct evolutionary histories with a high degree of accuracy, as long as the rates of change of the observed characters are appropriate for analysis. This emphasizes the importance of methods that evaluate whether rates of evolutionary change in target sequences are appropriate for phylogenetic analysis (41). Experimental phylogenies also indicate that many methods may be fairly robust to violations of the underlying assumptions, such as non-independence among nucleotide sites or deviations from simple models of evolution. It also is clear that differential weighting of character-state changes to reflect the observed frequency of the different types of transformations may substantially improve the performance of phylogenetic methods (especially

at high rates of change) and that methods that have a strict assumption of equal rates of change (for example, UPGMA) show poor overall performance.

Although most current methods of phylogenetic analysis perform quite well, there clearly is room for improvement. Further development of methods of accurately estimating the probability of changes among character states is needed in order to improve the incorporation of differential weighting in phylogenetic analyses (*42*). Such weighting methods could also be incorporated into methods of calculating pairwise distances, in order to improve the performance of distance methods. Maximum-likelihood methods are undergoing considerable development (*43*), and their performance is likely to improve markedly as the relevant evolutionary parameters are identified and incorporated into analyses. However, methods of phylogenetic analysis that identify an optimality criterion (so that alternative solutions can be compared and ranked) are limited by computational constraints. For just 50 taxa, there are approximately 3×10^{76} possible rooted bifurcating solutions, and for the estimated 30 million living species on Earth, there are approximately $10^{300,000,000}$ possible bifurcating phylogenies (*44*). Obviously, it is unrealistic to expect exact solutions to problems of this complexity. Although there has been considerable development of tree-searching algorithms in the past few decades (*45*), algorithms that take advantage of parallel processing need to be developed to achieve additional substantial improvement (*46*).

Two decades ago, phylogenetic methods were rarely used except by systematists with a basic interest in the evolutionary history of life. Today, those methods are widely used in biomedical applications, in molecular investigations of genome organization and gene structure, in studies of the origin of new alleles and laboratory strains, in comparative studies of ecology and behavior, in investigations of physiological processes, and in all fields in which biological comparisons are made among organisms. Phylogenetic analyses will play an increasingly important role as molecular biologists work in the coming decades to synthesize the comparative sequence information from the various genome projects. Studies of the accuracy of methods of phylogenetic analysis will be necessary to ensure that methods are developed and implemented that maximize our ability to reconstruct evolutionary history.

REFERENCES AND NOTES

1. D. R. Brooks and D. A. McLennan, *Phylogeny, Ecology, and Behavior* (Univ. of Chicago Press, Chicago, 1991); P. H. Harvey and M. D. Pagel, *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991); M. M. Miyamoto and J. Cracraft, Eds., *Phylogenetic Analysis of DNA Sequences* (Oxford Univ. Press, New York, 1991).
2. D. M. Hillis and C. Moritz, Eds., *Molecular Systematics* (Sinauer, Sunderland, MA, 1990).
3. G. J. Olsen, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 825 (1987); J. A. Lake, *Nature* **331**, 184 (1988); M. Gouy and W.-H. Li, *ibid.* **339**, 145 (1989).
4. C.-Y. Ou *et al.*, *Science* **256**, 1165 (1992).
5. W. R. Atchley and W. M. Fitch, *ibid.* **254**, 554 (1991); D. M. Hillis and J. J. Bull, *ibid.*, p. 528.
6. D. M. Hillis, J. J. Bull, M. E. White, M. R. Badgett, I. J. Molineux, *ibid.* **255**, 589 (1992).
7. J. J. Bull, C. W. Cunningham, I. J. Molineux, M. R. Badgett, D. M. Hillis, *Evolution* **47**, 993 (1993).
8. W. M. Fitch, *Philos. Trans. R. Soc. London Ser. B* **316**, 317 (1986); W. C. Wheeler and R. L. Honeycutt, *Mol. Biol. Evol.* **5**, 90 (1988); M. T. Dixon and D. M. Hillis, *ibid.* **10**, 256 (1993).
9. T. Gojobori, W.-H. Li, D. Graur, *J. Mol. Evol.* **18**, 360 (1982); W.-H. Li, C.-I. Wu, C.-C. Luo, *ibid.* **21**, 58 (1984).
10. E. Sober, *Syst. Biol.* **42**, 85 (1993); D. M. Hillis, J. J. Bull, M. E. White, M. R. Badgett, I. J. Molineux, *ibid.*, p. 90.
11. M. Nei, in *Phylogenetic Analysis of DNA Sequences*, M. M. Miyamoto and J. Cracraft, Eds. (Oxford Univ. Press, New York, 1991), pp. 90–128.
12. J. P. Huelsenbeck and D. M. Hillis, *Syst. Biol.* **42**, 247 (1993).
13. D. Penny, M. D. Hendy, M. A. Steel, *Trends Ecol. Evol.* **7**, 73 (1992).
14. T. H. Jukes and C. R. Cantor, in *Mammalian Protein Metabolism*, H. Munro, Ed. (Academic Press, New York, 1969), pp. 21–132.
15. Biases in simulation studies may be created by choice of (i) tree topology, (ii) the model of evolution, (iii) the implementation of a method (for example, the adequacy of the search for optimal solutions), and (iv) the method of scoring ties. If a method finds multiple solutions, one of which is correct, some researchers score this result either as complete failure or as complete success. Neither extreme seems reasonable; we use the average number of correctly resolved components among all optimal solutions to score each method. Under this criterion, a method that finds two equally good solutions for an unrooted four-taxon tree (one of which is correct) is scored as 50% correct.
16. J. Felsenstein, *Syst. Zool.* **27**, 401 (1978).
17. M. Kimura, *J. Mol. Evol.* **16**, 111 (1980).
18. Four-taxon simulations assumed a Kimura model of evolution with transitions occurring 10 times more frequently than transversions. The overall probability of change occurring along the branches was varied from 1 to 75% expected internodal change, in 2% increments. For Fig. 1, sequences of 200 sites were simulated, and 1000 four-taxon trees were simulated for each combination of branch-length parameters. Neighbor joining and UPGMA analyses were done with uncorrected distances (the frequency of observed differences) and Kimura distances (*17*). Because Kimura distances can become undefined at very high levels of divergence, the upper limit of Kimura distances was constrained to be 7.5 (equivalent results are obtained by using any arbitrarily large number to replace undefined distances; the only alternative is to consider the results undefined). For all analyses in this paper, all trees were treated as unrooted to avoid biasing the results against methods that produce only rooted trees (such as UPGMA).
19. D. Sankoff, *SIAM J. Appl. Math.* **28**, 35 (1975); P. L. Williams and W. M. Fitch, in *The Hierarchy of Life: Molecules and Morphology in Phylogenetic Analysis*, B. Fernholm, K. Bremer, H. Jörnvall, Eds. (Excerpta Medica, Amsterdam, 1989), pp. 453–470. Weighted parsimony analyses in this paper used weights of character-state changes on the basis of the inverse of the changes' frequency of occurrence, rounded to the nearest integer value and corrected to satisfy the triangle inequality. Symmetric weights were based on the average frequency of the relevant reciprocal character transformations.
20. W. P. Maddison and D. R. Maddison, *MacClade: Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, MA, 1992).
21. A. Sidow, *Nature* **367**, 26 (1994).
22. J. A. Lake, *Mol. Biol. Evol.* **4**, 167 (1987).
23. Centers for Disease Control and Prevention, *Morb. Mortal. Wkly. Rep.* **42**, 329 (1993).
24. R. W. DeBry *et al.*, *Nature* **361**, 691 (1993).
25. E. N. Moriyama *et al.*, *J. Mol. Evol.* **32**, 360 (1991).
26. The substitution matrix for the HIV simulations was derived from Table 1 by dividing the average number of changes for each nucleotide substitution by 281 (the total number of observed changes). Each element of the substitution matrix was taken as the instantaneous rate of change between nucleotides. If we denote the probability that the nucleotide at a given site is an A, C, G, or T by g_1, g_2, g_3, or g_4, respectively, then the frequencies of the nucleotides in the next time interval dt are given by the following differential equations:

$$\frac{dg_1}{dt} = -0.493g_1 + 0.082g_2 + 0.149g_3 + 0.037g_4$$

$$\frac{dg_2}{dt} = 0.145g_1 - 0.167g_2 + 0.160g_3 + 0.084g_4$$

$$\frac{dg_3}{dt} = 0.285g_1 + 0.007g_2 - 0.320g_3 + 0.045g_4$$

$$\frac{dg_4}{dt} = 0.063g_1 + 0.078g_2 + 0.011g_3 - 0.166g_4$$

These equations were integrated numerically with respect to time to determine the probability that a nucleotide of initial identity i has identity j after t units of time. The tree and branch lengths in Fig. 3 were used as the basis of the HIV simulations. The overall amount of time elapsed from the basal node of the phylogeny to sequence LC03-x was fixed at 0.1, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, and 10.0 time units. For the simulations in Fig. 5, the terminal branches from the simulation at 0.5 time units were extended by 0.1, 0.2, 0.3, 0.4, and 0.5 time units (which were converted to years by scaling the original terminal branches to the known times of infection and of HIV sampling) (*28*). The sequence for HIVELI was used as the starting point of the simulations. For each branch of the phylogeny, the probabilities of different nucleotide changes were determined for the amount of time from one end of the branch to the other. A pseudo-random number was used to determine which event occurred. One hundred replicate data sets were simulated for each level of divergence.
27. Parsimony analyses were conducted with the use of PAUP 3.1.1 and 3.1.2; D. L. Swofford, Smithsonian Institution, Washington, DC; with tree bisection and reconnection (TBR) branch-swapping and retaining only a single tree from each replicate. Maximum-likelihood estimates of Kimura distances (using the actual transition transversion ratio), as well as neighbor joining and UPGMA trees, were computed with the use of Phylip 3.5; J. Felsenstein, Department of Genetics, University of Washington, Seattle.
28. C. Ciesielski *et al.*, *Ann. Intern. Med.* **116**, 798 (1992); C. Ciesielski, personal communication, Centers for Disease Control and Prevention, Atlanta, GA.
29. B. H. Hahn *et al.*, *Science* **232**, 1548 (1986); M. S. Saag *et al.*, *Nature* **334**, 440 (1988).
30. J. A. McKeating *et al.*, *AIDS* **3**, 777 (1989).
31. T. Shioda, J. A. Levy, C. Cheng-Mayer, *Nature* **349**, 167 (1991).
32. C. A. B. Boucher *et al.*, *Lancet* **336**, 585 (1990); B. A. Larder, G. Darby, D. D. Richman, J. M. Grimes, S. W. Lagakos, *J. Acquired Immune Defic. Syndr.* **3**, 743 (1990).

33. Sequences have been deposited in GenBank, accession numbers L26366 to L26392.
34. B. Bremer, *Plant Syst. Evol.* **175**, 39 (1991).
35. T. E. Dowling, C. Moritz, J. D. Palmer, in (*2*), pp. 250–317.
36. Sequences were obtained from genes 0.3 (519 nucleotides), 17 and 17.5 (356 nucleotides), and 18 (216 nucleotides). Protein 0.3 inactivates host restriction, protein 17 is a tail fiber protein, and protein 18 functions in DNA maturation. The function of protein 17.5 is unknown.
37. J. Felsenstein, *J. Mol. Evol.* **17**, 368 (1981); *Evolution* **46**, 159 (1992), as implemented in Phylip (*27*).
38. N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987).
39. W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).
40. R. R. Sokal and C. D. Michener, *Univ. Kans. Sci. Bull.* **38**, 1409 (1958).
41. D. M. Hillis and J. P. Huelsenbeck, *J. Hered.* **83**, 189 (1992).
42. J. Felsenstein, *Biol. J. Linn. Soc.* **16**, 183 (1981); W. C. Wheeler, *Cladistics* **6**, 269 (1990).
43. J. Felsenstein and E. Sober, *Syst. Zool.* **35**, 617 (1986); B. Golding and J. Felsenstein, *J. Mol. Evol.* **31**, 511 (1990); N. Goldman, *Syst. Zool.* **39**, 345 (1990); *J. Mol. Evol.* **36**, 182 (1993).
44. Formulas for calculating the number of possible trees were derived by L. L. Cavalli-Sforza and A. W. F. Edwards [*Evolution* **21**, 550 (1967)]. The actual number of possible phylogenies of life would also have to include trees with reticulations and multifurcations and would need to include all the extinct species that have ever existed, so the estimate shown is conservative.
45. D. L. Swofford and G. J. Olsen, in (*2*), pp. 411–501.
46. W. D. Hillis and B. M. Boghosian, *Science* **261**, 856 (1993).
47. Figure 3 was based on a reanalysis of the data from (*4*), by use of uniformly weighted parsimony and TBR branch-swapping with PAUP (*27*). The tree shown is one of eight equally parsimonious trees of 281 steps (consistency index = 0.68), which differ in minor rearrangements within the dental clade. The local control consensus sequence from the original study was not included because it does not represent an observed sequence.
48. Supported by NSF grants BSR 9106746 and DEB 9221052. We thank B. Bowman, J. J. Bull, C. A. Ciesielski, E. Holmes, H. Jaffe, C. Jenkins, M. Kalish, P. MacManus, I. J. Molineux, G. Myers, G. Schochetman, and D. L. Swofford for advice and assistance.