

Analysis and Visualization of Tree Space

DAVID M. HILLIS,¹ TRACY A. HEATH,¹ AND KATHERINE ST. JOHN²

¹*Section of Integrative Biology and Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, Texas 78712, USA;
E-mail: dhillis@mail.utexas.edu (D.M.H.)*

²*Department of Mathematics and Computer Science, Lehman College—City University of New York, Bronx, New York 10468, USA;
E-mail: stjoh@lehman.cuny.edu*

Abstract.—We explored the use of multidimensional scaling (MDS) of tree-to-tree pairwise distances to visualize the relationships among sets of phylogenetic trees. We found the technique to be useful for exploring “tree islands” (sets of topologically related trees among larger sets of near-optimal trees), for comparing sets of trees obtained from bootstrapping and Bayesian sampling, for comparing trees obtained from the analysis of several different genes, and for comparing multiple Bayesian analyses. The technique was also useful as a teaching aid for illustrating the progress of a Bayesian analysis and as an exploratory tool for examining large sets of phylogenetic trees. We also identified some limitations to the method, including distortions of the multidimensional tree space into two dimensions through the MDS technique, and the definition of the MDS-defined space based on a limited sample of trees. Nonetheless, the technique is a useful approach for the analysis of large sets of phylogenetic trees. [Bayesian analysis; multidimensional scaling; phylogenetic analysis; tree space; visualization.]

Systematists are often faced with the need to analyze a large collection of phylogenetic trees. These trees may represent a collection of equally parsimonious solutions to a phylogenetic problem, or a set of trees of similar likelihood, or a sampled set of trees from a Markov chain Monte Carlo (MCMC) Bayesian analysis. In any of these cases, a common approach for expressing the results is to make a consensus tree from the large collection of potential solutions (see Swofford, 1991, for a discussion of consensus methods). Consensus trees are produced to distill a large amount of information into a single summary tree, because it is often impractical to examine or display all of the individual solutions. In the case of MCMC Bayesian analysis, a consensus tree is usually used to summarize information about the posterior probabilities of the individual inferred branches. Although these uses of consensus trees may be appropriate for many purposes, a great deal of information about the individual solutions is usually lost. It is possible that two or more distinct but different biological explanations are represented among different “islands” of solutions (e.g., see Maddison, 1991), but that a consensus of these solutions produces little or no resolution. Although many other solutions among the universe of possible trees may be excluded by the available data, this information can be lost in a consensus tree.

This article describes an alternative method of exploring a set of phylogenetic trees. Geographical topology is often used as an analogy to describe the solution space of phylogenetic trees, and to describe or discuss the behavior of various tree-search strategies (e.g., see Swofford et al., 1996). However, systematists have not usually moved beyond analogy to attempt to visualize this solution space. Here we suggest one approach to this problem and describe a program that allows the rapid and efficient visualization of commonalities and relationships among a large set of phylogenetic trees.

MULTIDIMENSIONAL SCALING OF TREE SPACE

Biologists usually think of tree space in terms of the topological distance among different trees, which may be defined in terms of common measures such as weighted or unweighted Robinson-Foulds (RF) distance (Robinson and Foulds, 1979, 1981; see also Buneman, 1971). The unweighted RF distance merely sums the number of internal edges (branches) that must be collapsed or expanded to move from one tree topology to another, without any effect on the measure of the length of those edges. The weighted version of RF distance weights the edges by their length, so that two trees are considered to move further apart in tree space as the differences in their branch lengths increase. Thus, weighted RF distances can be useful for distinguishing among trees of identical topology, but with different branch lengths. Given a collection of trees, it is straightforward to calculate pairwise RF-distance matrices (weighted and unweighted) among all the pairs of trees, and this can be done efficiently (linear time with respect to number of taxa) using an algorithm developed by Day (1985). However, an exact representation of the relationships among all possible trees typically would require a large multidimensional space. A common method of visualizing and analyzing a large matrix of distances among points is multidimensional scaling (MDS; see Lingoes et al., 1979; Young and Hamer, 1987; Borg and Groenen, 1997). In MDS, a new space of only a few (typically two) dimensions is created in a manner to minimize the distortions of the observed distance matrix. An optimal solution to MDS involves minimizing a stress function, such as the Kruskal-1 function (Borg and Groenen, 1987), defined as

$$S_D(x_1, \dots, x_n) = \left(\sum_{i \neq j=1}^n (D_{ij} - |x_i - x_j|)^2 \right)^{1/2}$$

where D_{ij} is the matrix of distances between trees, (x_1, \dots, x_n) are the point locations in redefined space,

and S_D is the stress function. Thus, in an MDS representation of trees, the distances between different solutions approximate the observed distances between trees, with some distortion (which is quantified by the stress function).

In this article, we explore the use of multidimensional scaling to compare and examine large sets of phylogenetic trees and present several example applications. We also briefly describe software (distributed without cost to the user) that can be used to visualize sets of phylogenetic trees in MDS space.

METHODS

Tree Set Visualization Module for Mesquite

To visualize and examine large sets of phylogenetic trees, we used the Tree Set Visualization module for Mesquite (<http://comet.lehman.cuny.edu/treeviz/>). Maddison and Maddison (2004) developed the software package Mesquite as a modular system for evolutionary analysis. This project complements and extends their popular software tool, MacClade (Maddison and Maddison, 1992), but has the extra flexibility to allow

new methods and visualization techniques to be easily added to the underlying phylogenetic tools. The package is written in Java, and, as such, runs under Macintosh, Windows, and Unix operating systems. The Tree Set Visualization module (Amenta and Klingner, 2002) takes sets of phylogenetic trees and uses MDS to display the trees as points in two-dimensional space, such that the distortion between the true distance between pairs of trees and the screen distance is minimized (based on the stress function described above). Sets of trees can be colored according to their score on an optimality criterion (e.g., a parsimony or maximum likelihood score) or color can represent subsets of trees (e.g., different colors for trees from different Bayesian analyses). Points on the screen can be selected to show the corresponding tree, or consensus of a set of trees (see Fig. 1).

We used this software package to study several potential applications of the visualization of phylogenetic tree space. These examples are meant to stimulate additional work on the visualization of tree space, and are not meant to be an exhaustive description of the potential uses of the software.

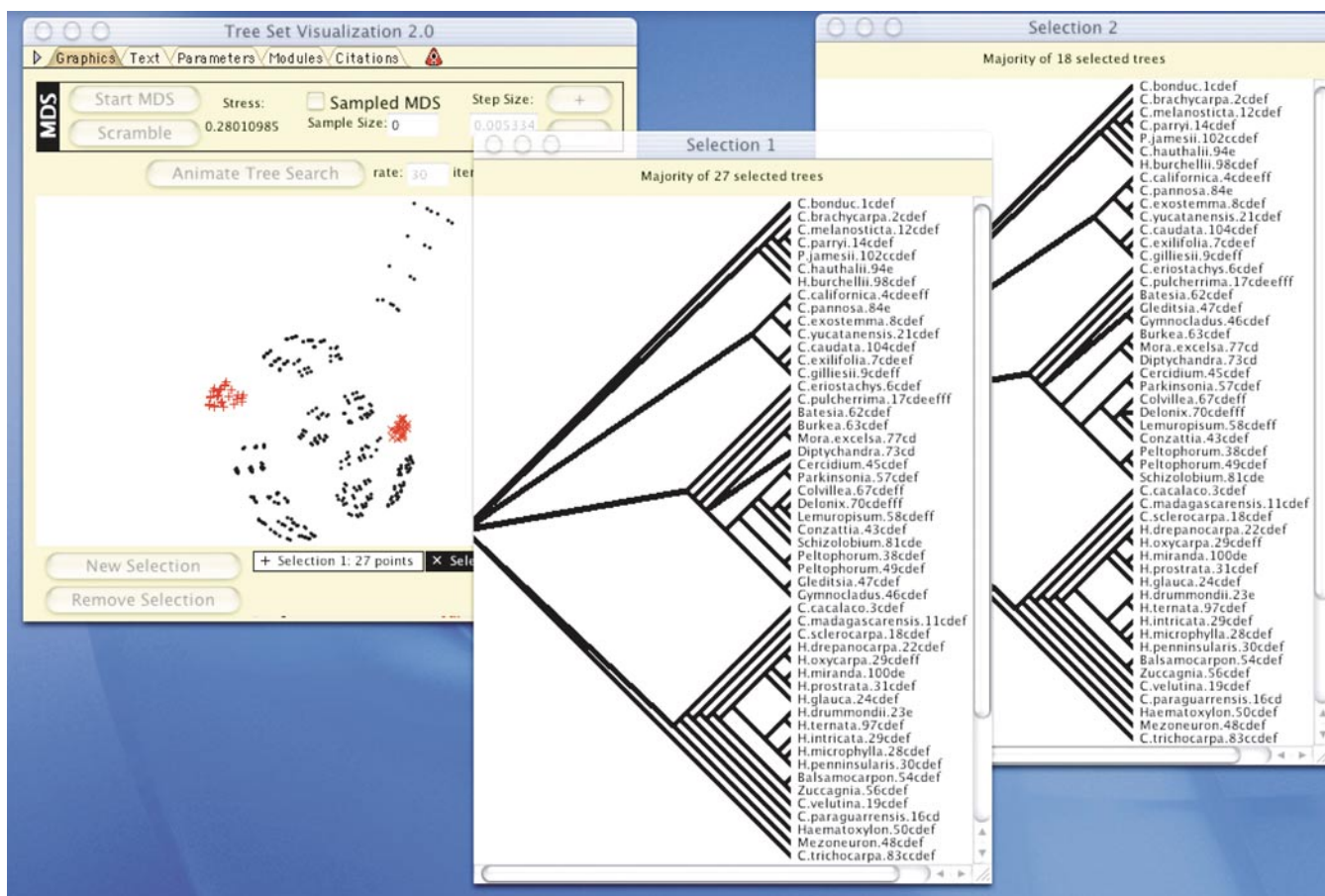


FIGURE 1. Screen shot of output from the Tree Set Visualization module of Mesquite. Windows labeled "Selection 1" and "Selection 2" show the strict consensus trees for the tree islands that have been selected in the main window.

Bayesian Samples versus Bootstrap Samples

One of our analyses consisted of a comparison of a sample of trees from a Bayesian MCMC analysis (Larget and Simon, 1999) with a sample of trees obtained through nonparametric bootstrapping (Felsenstein, 1985). These two sampling methods have been widely used to obtain credible sets of trees and for estimating confidence limits on phylogenies. Comparisons of these methods have largely concentrated on comparing the posterior probabilities for particular internal branches calculated from the Bayesian analyses with the bootstrap proportions for these same branches (e.g., Wilcox et al., 2002; Suzuki et al., 2002; Alfaro et al., 2003; Cummings et al., 2003; Erixon

et al., 2003). We explored an alternative approach that involved the comparison of the sets of sampled trees to the true (model) tree in tree-space.

The data used to generate the trees for comparing the results of Bayesian and bootstrapping analyses were obtained by simulating sequences of genes on the topology and branch lengths of a 44-taxon mammalian tree (Murphy et al., 2000; Fig. 2). This tree was the result of an analysis of 22 different genes and was chosen as a model for our simulations because of the large number of genes and the well-supported resulting phylogenetic estimate (this tree has been discussed in several recent theoretical studies of phylogeny;

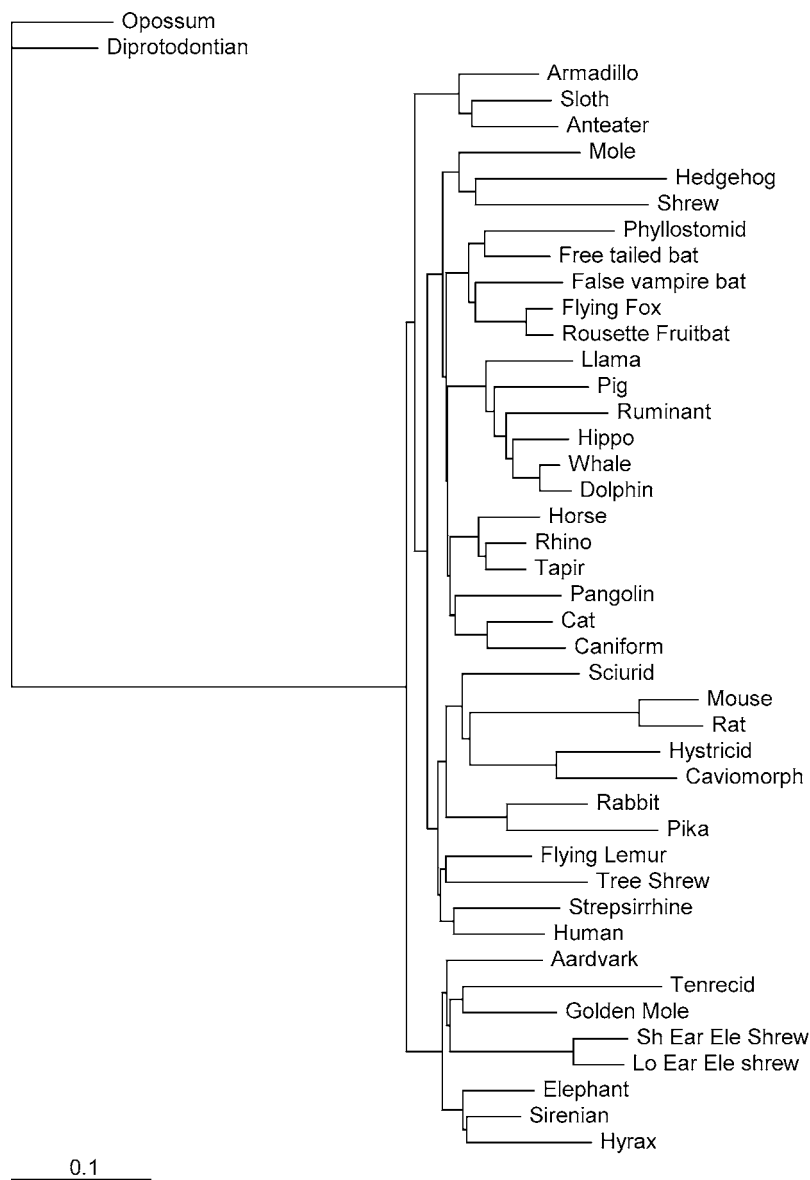


FIGURE 2. Phylogram of mammals by Murphy et al. (2001). This tree was used in the simulations to obtain sets of trees from Bayesian and bootstrap analyses for comparison to the true tree in tree-space. The details of the branch lengths for this tree are presented in Appendix 1.

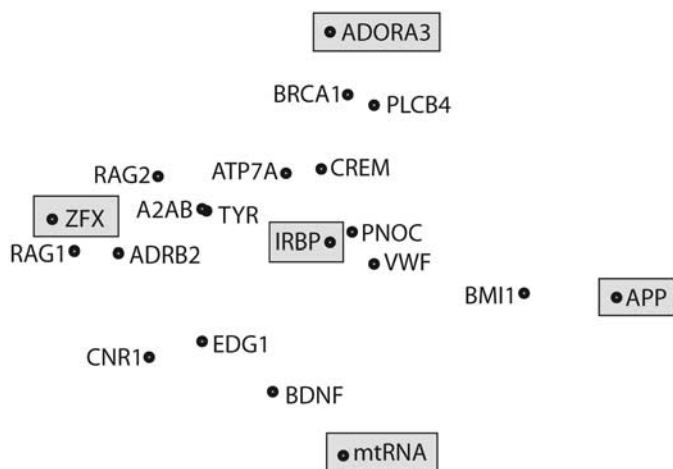


FIGURE 3. Multidimensional scaling of the pairwise Euclidean distances between the model parameters for each gene in the Murphy et al. (2001) data set. The five genes shown in boxes were selected to represent the widest range of parameter conditions.

e.g., Suzuki et al., 2002; Alfaro et al., 2003; Douady et al., 2003). For each gene, we conducted a likelihood-ratio test to select an appropriate model of evolution (Posada and Crandall, 2001) and used PAUP* 4.0b10 (Swofford, 2000) to estimate the optimal model parameters. Then, we computed pairwise Euclidean distances between each set of model parameters by taking the square root of the sum of the squared deviations of each parameter in the model. These distances were then analyzed using multidimensional scaling (Fig. 3). From this MDS analysis, we chose the model parameters of the gene closest to the centroid (IRBP) as simulation conditions for the single gene analysis (other genes were selected for multigene analyses, as described below). The preferred model for this gene was a general time-reversible model with discrete gamma-distributed rate heterogeneity and a proportion of invariant sites (Appendix 2). Seq-Gen version 1.2.5 (Rambaut and Grassly, 1997) was used to simulate 1000 nucleotides per taxon under this model of evolution on the tree shown in Figure 2.

To obtain the bootstrap sample, we analyzed a simulated data set for 1000 nonparametric bootstrap replicates in PAUP* 4.0b10 (Swofford, 2000) with nearest-neighbor-interchange swapping on a stepwise-addition starting tree. Prior to bootstrapping, optimal model parameters were estimated and set for the bootstrapping analysis. Additionally, we analyzed the same data set using MrBayes 3.04b (Huelsenbeck and Ronquist, 2003). The four Markov chains (see Geyer, 1991; Gilks and Roberts, 1996; Huelsenbeck et al., 2001) were run for 5 million generations under the simulation model (GTR+I+ Γ), sampling every 1000 generations. The 1000 trees from the bootstrapping analysis were combined with the sample of every 1000th tree from the last 1,000,000 generations of the Bayesian analysis, together with the single true (model) tree, into a single file for visualization using multidimensional scaling.

Gene Length and Concatenated Datasets

A second problem that we addressed concerned the performance of phylogenetic inference as a function of increasing sequence length and increasing numbers of genes that evolve under different models of evolution. For this analysis, we selected five genes from the MDS analysis of genes from the Murphy et al. (2001) study (Fig. 3). In this case, we included the gene closest to the centroid (IRBP), as well as four genes that were the most divergent from IRBP in the MDS analysis (Fig. 3). For each gene, we simulated independent datasets from 200 to 1000 nucleotides long on the tree shown in Figure 2.

We analyzed each data set under its individual simulation model in MrBayes 3.04b (Huelsenbeck and Ronquist, 2003) for one million generations. Additionally, we concatenated and analyzed all of the genes in a combined analysis. We conducted a hierarchical likelihood ratio test to determine a suitable model of analysis for each concatenated data set (Posada and Crandall, 2001); these tests on the concatenated data sets always selected the GTR+I+ Γ model. The multigene data sets were then analyzed in MrBayes under a single model (GTR+I+ Γ) for one million generations. The concatenated data sets were analyzed in MrBayes under a "composite model" consisting of the five models used to simulate each gene (Nylander et al., 2004). The true tree and one thousand trees from the analyses of each single gene data set and the single model analysis of the concatenated data set, sampled every 100 generations for the last 100,000 generations, were combined for the MDS analysis.

Comparing Two Bayesian Analyses

A third problem that we considered was the use of multidimensional scaling to compare the samples of trees from multiple independent Bayesian analyses, as a means of determining the degree of convergence in the analyses. For this example, we used the same data set described above (for comparing Bayesian and bootstrap samples) to compare the trees generated by two Bayesian analyses. Two runs were conducted using MrBayes; each run consisted of four Markov chains run for 10 million generations sampled every 1000 generations. The model used for analysis was GTR+I+ Γ (the simulation model).

Three thousand trees from the last 3 million generations from each analysis were combined into a single file for visualization using MDS. These trees were plotted using both weighted and unweighted RF distances (Robinson and Foulds, 1979, 1981).

RESULTS AND DISCUSSION

Comparison of Bayesian and Bootstrap Samples of Phylogenetic Trees

Interpretation of the results of bootstrap resampling results from phylogenetic studies has been the focus of considerable discussion. Felsenstein (1985) suggested that

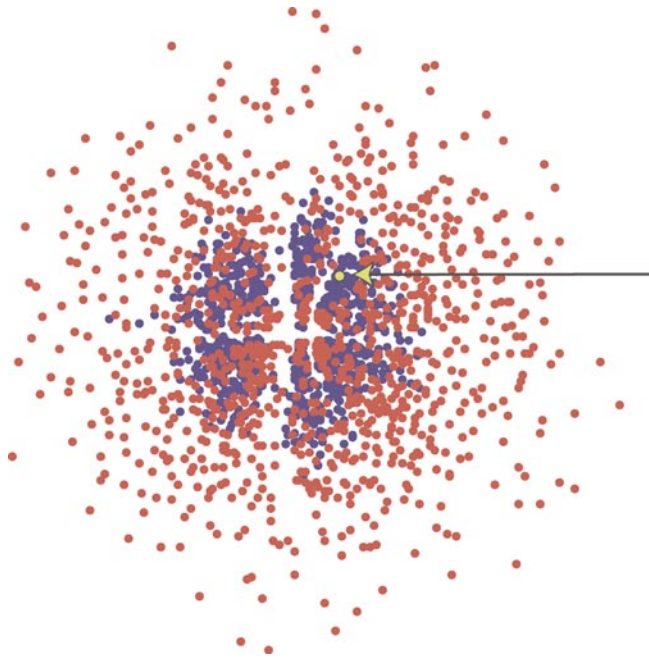


FIGURE 4.

bootstrapping could be used to assess confidence in particular branches of a phylogenetic tree. Felsenstein’s method relies on using the proportion of the bootstrap pseudosamples that support each branch in the tree (the bootstrap proportions). However, several authors have noted that these proportions present a highly biased (and usually conservative) estimates of phylogenetic accuracy (e.g., Zharkikh and Li, 1992a, 1992b, 1995; Hillis and Bull, 1993; Rodrigo, 1993; Li and Zharkikh, 1994; Efron et al., 1996). More recently, Bayesian posterior probabilities (derived from a sample of trees drawn from an MCMC search of tree space after the search has reached equilibrium) have also been used to assess the support

FIGURE 4. Visualization of 1000 trees from a bootstrapping analysis of 1000 replicates and 1000 trees from a Bayesian MCMC analysis (every 1000th tree from generation 4 million to generation 5 million). The true tree is indicated in yellow and by the yellow arrow. The trees are compared using Robinson-Foulds distances. Note the smaller region encompassed by the Bayesian sample.

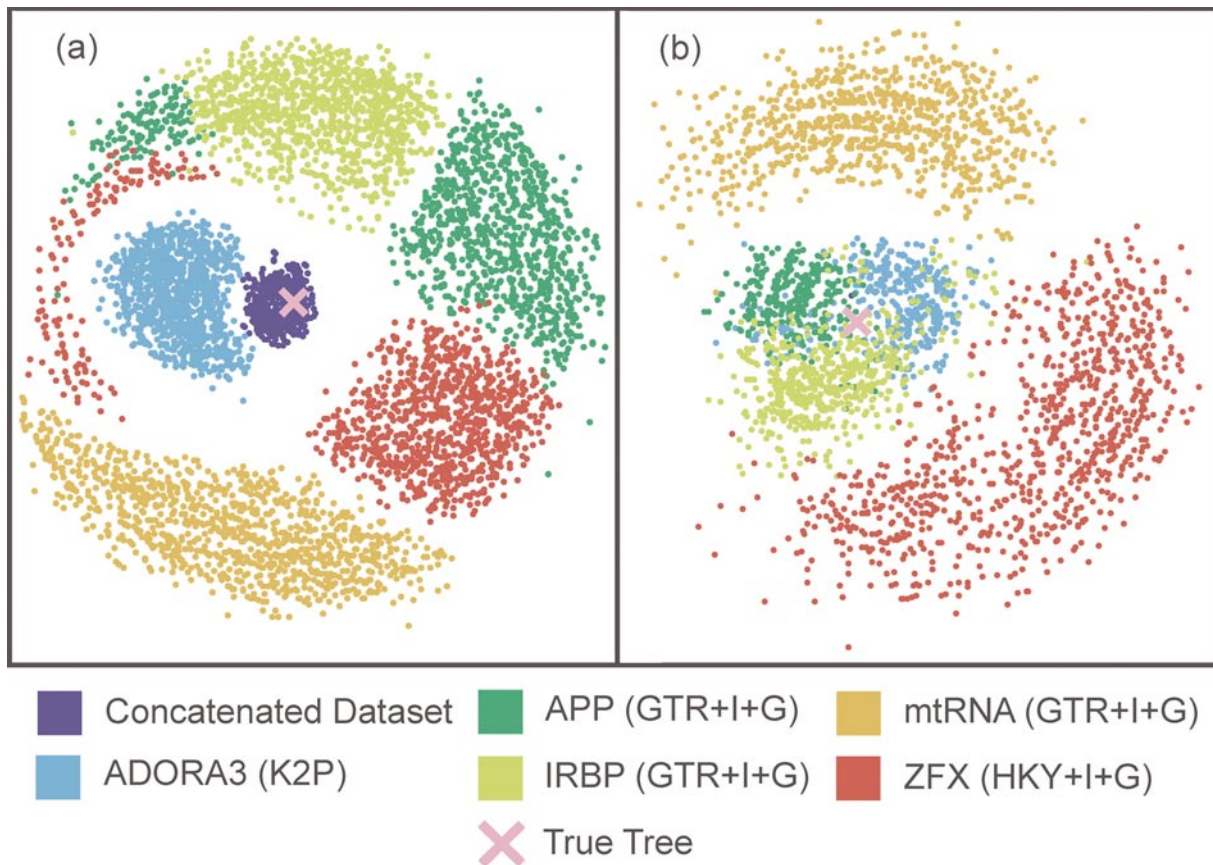


FIGURE 5. Visualization (using MDS) of the tree space occupied by six sets of trees sampled in Bayesian analyses of five different data sets, plus the concatenated data set (dark blue). Each of the five genes was simulated on the same tree under different model parameters (see text and Appendices 1–3). The true tree is indicated by the pink “X.” All trees were compared using Robinson-Foulds distances. (a) Results from a simulation of 200 bp for each gene; none of the individual data sets found the correct tree, but the trees from the concatenated are centered around the true tree. (b) Results from a simulation of 1000 bp for each gene; many of the trees from the concatenated analysis were identical to the true tree in this example.

or confidence of individual branches in a phylogenetic tree. Many authors have suggested that these Bayesian posterior probabilities are more reasonable estimates of phylogenetic accuracy compared to bootstrap proportions (given the assumptions of the models used in the estimation procedure), although other authors disagree (e.g., see Wilcox et al., 2002; Suzuki et al., 2002; Alfaro et al., 2003; Cummings et al., 2003; and Huelsenbeck and Rannala, 2004, for a diversity of opinions). However, both of these methods (nonparametric bootstrapping and Bayesian MCMC analysis) can be used in a different manner to assess phylogenetic results; namely, they can be used to produce a "credible set" of trees that are considered consistent with the observed data. Bootstrap proportions and Bayesian posterior probabilities are both summary statistics that are extracted from these respective credible sets of trees, and both statistics result in potential loss of information. An alternative approach might be to visualize the tree space that is defined by these credible sets and then use these credible sets to envision actual "confidence limits" in this space around the best estimate. The space within these confidence limits could be used to test a particular hypothesis that might depend upon many different branches in a phylogenetic tree, rather than on the support of a single branch.

An example of this approach is illustrated in Figure 4. In this figure, the true (model) tree is represented by the yellow dot, the trees that are estimated from bootstrap pseudosamples are shown in red, and the trees that are sampled from the MCMC analysis are shown in blue. One might be tempted to draw confidence ellipses around these samples to include 95% of the results, which could indicate the "confidence limits" of the result in tree space. However, there are problems with this approach. First, MDS space is defined by a limited sample of trees (rather than the universe of possible trees). So the "confidence limits" would be dependent on the sample used to define the space. Second, there is distortion of multidimensional space in two dimensions (see section on potential limitations, below). Therefore, we do not think that this approach to defining confidence limits of trees is likely to be practical. Nonetheless, we see the visualization shown in Figure 4 as a useful way to illustrate the relative distributions of trees drawn from bootstrap resampling to trees sampled from a MCMC Bayesian analysis. The tighter clustering of the Bayesian sample compared to the bootstrap pseudosample corresponds to the broader confidence limits typically obtained from nonparametric bootstrap analyses (e.g., Wilcox et al., 2002; Huelsenbeck and Rannala, 2004).

Comparing Results from Multiple Genes and Concatenated Data Sets

Another potential application for the visualization of tree space is in studies of multiple data sets and the combination of data sets. The MDS analysis of the results from

our simulation of five of the genes studied by Murphy et al. (2001), as well as the analysis of the concatenation of those five genes, is shown in Figure 5. When we compared the RF distances of the trees in Figure 5a to the true tree by computing the distance matrix in PAUP* 4.0b10 (Swofford, 2000), we found that the true tree was an RF distance of 18 to 60 from the estimated trees for the separate genes, although the single model analysis of the concatenated dataset provided much closer estimates of the true tree (RF = {4, 6, 8, ..., 22}). With longer sequences (1000 nucleotides for each gene), three of the single gene analyses included the true tree, whereas the space defined by trees from the other two genes did not include the true tree (Fig. 6). However, both the single model and composite model analyses of the concatenated data set produced similar results and found the true tree or trees that were only a few rearrangements from the true tree (maximum RF = 6; Fig. 6). In this example, all of the genes were simulated on the same tree, and the simulations differed only in the underlying model of evolution of the gene. The analysis suggests that a combined approach to data analysis was highly successful in this case, using either a homogeneous or heterogeneous model for analysis. We do not wish to generalize this result to suggest that all genes should be combined in a joint analysis, as we suspect that the specific results are likely to differ depending on the degrees of (and reasons for) differences among the genes (or other data partitions). Rather, we use this example to show how the visualization approach might be used to investigate the utility of combined versus separate analysis of datasets.

Comparing Results from Multiple Bayesian Analyses

In conducting a Bayesian phylogenetic analysis, it is important to assess whether or not a sample from a Markov chain has converged on the joint posterior distribution of the various parameters (Huelsenbeck and Ronquist, 2003). However, the techniques for assessing convergence have not been well developed, and most approaches do not examine convergence on the *joint* posterior distribution of the parameters, but instead examine convergence on the posterior distributions of individual parameters (the *marginal* posterior distributions). For instance, a common approach is to examine the marginal posterior probabilities of individual bipartitions of the tree across several Bayesian MCMC runs, and then to assume that the samples from the Markov chains have converged if these posterior probabilities are consistent. Many phylogenetic hypotheses, however, do not depend on particular bipartitions in the tree, but rather on larger features of tree topology or combinations of branch lengths, and convergence on marginal posterior distributions does not guarantee convergence on the joint posterior distribution. Here we suggest a method for detecting non-convergence in Bayesian analyses on the joint posterior distribution of topologies and branch lengths.

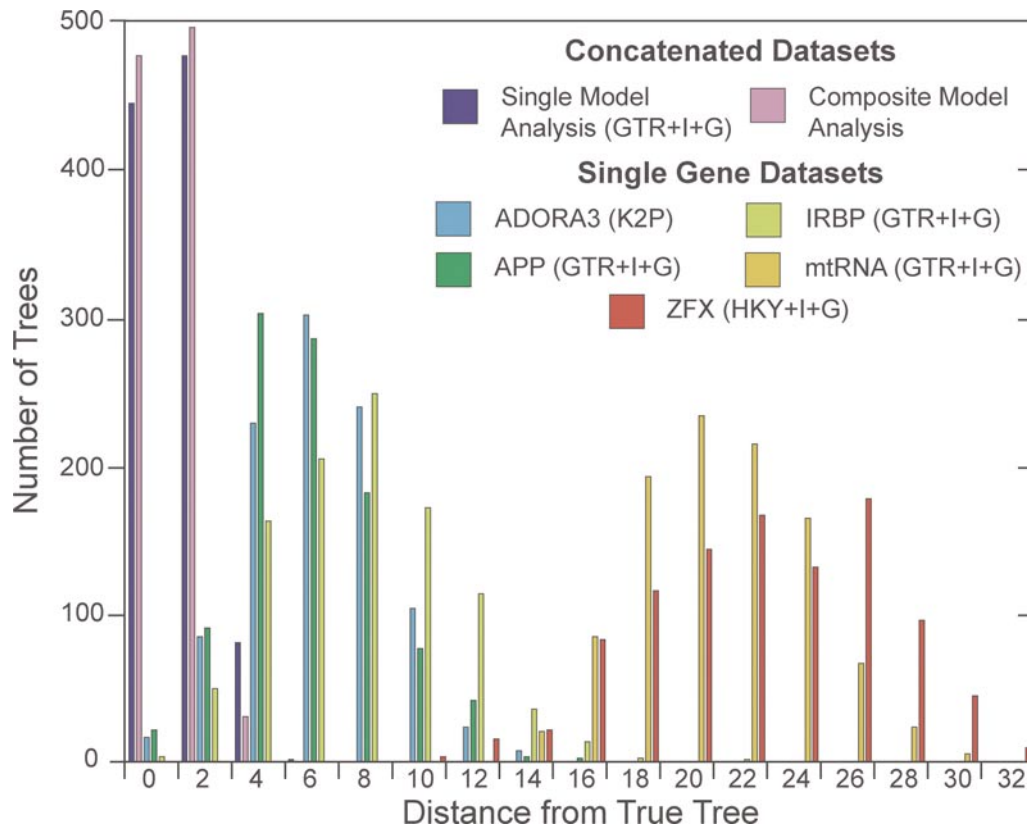


FIGURE 6. Histogram of the distance (in RF distance) from the true tree for trees sampled from a Bayesian analysis of five genes, each 1000 nucleotides in length, as well as the homogeneous model analysis and the heterogeneous model analysis of the concatenated sequences.

Figure 7a shows the visualization of tree space sampled from two different Bayesian analyses of the same data set, with the trees compared using unweighted Robinson-Foulds distance. In this case, the sampled space is nearly indistinguishable in the two separate runs (even though the same trees are rarely sampled in the two analyses), indicating that the sampled space of tree topologies (without consideration of branch lengths) was very similar in the two replicates. Figure 7b shows the results of a similar analysis, but using weighted Robinson-Foulds distances to compare the trees, and in this case the sampled space of the two runs is clearly distinguishable. We interpret this result to mean that there is a greater degree of branch-length similarity within independent runs than between runs, suggesting that the chains have either not reached stationarity with respect to branch lengths, or more likely they have not been sampled for a long enough time (after stationarity was reached) to ensure convergence to the joint posterior distribution. In contrast, Figure 7a suggests that the space of tree topologies (independent of branch lengths) has been consistently sampled in the two analyses, although one would probably want to examine additional runs to ensure this result. Thus, this approach to visualizing the results of Bayesian analy-

ses may prove to be a fruitful heuristic for assessing appropriate chain lengths and sampling strategies in MCMC Bayesian analyses of phylogeny. We have also used this approach with several empirical data sets (results not shown) and have found it to be a useful approach for assessing convergence among independent analyses.

Some Other Potential Uses of MDS for Visualizing Tree-Space

Another potential use of multidimensional scaling as a means of visualizing tree-space is for examining the progress of a Bayesian MCMC analysis. This visualization may be useful mostly as a tool for describing the method (as in a class or workshop on phylogenetic methods). As an example, we selected a sample of every 10th tree from the first 100,000 generations of an analysis of frog phylogeny (Hillis and Wilcox, 2005), and analyzed the trees under MDS (Fig. 8). These trees represent the early samples from the Markov chain, as the chain begins to sample trees across tree-space, moving to regions of progressively higher optimality scores. This progress can be animated using the Tree Set Visualization module. We have found this example of tree-set visualization to

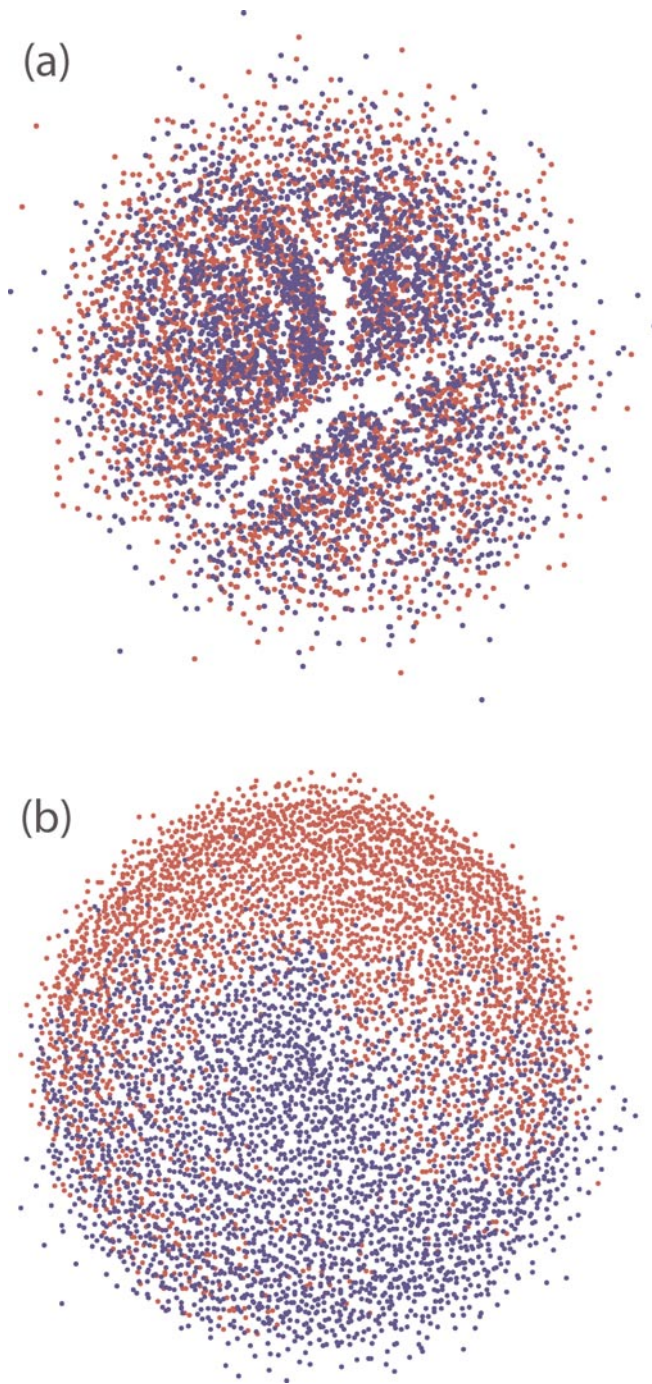


FIGURE 7. Comparison of trees generated from two different Bayesian MCMC analyses of the same data set. Both Bayesian analyses were conducted under the same conditions (see text). Each color represents the trees from a different analysis. (a) Analysis based on unweighted Robinson-Foulds distances; (b) analysis based on weighted Robinson-Foulds distances.

be a useful means for describing the Bayesian MCMC approach in classes and workshops on phylogenetics (see <http://lewis.eeb.uconn.edu/lewishome/software.html> for another useful MCMC instruction tool).

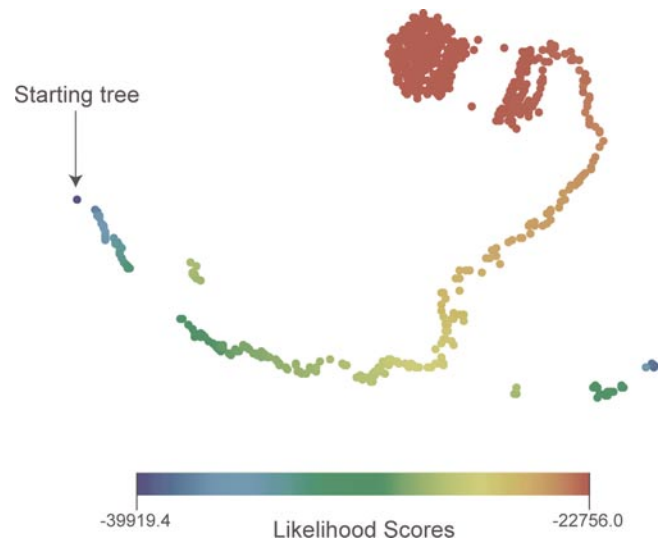


FIGURE 8. Progress in a Bayesian MCMC analysis. The progress in the search can be visualized in the Tree Set Visualization program, as a demonstration of how an MCMC analysis functions. In the visualization, the progress of the chain through tree-space moves from regions of low optimality scores (blue) to regions of high optimality scores (red).

Another common problem in phylogenetics is the discovery of several distinct “tree islands” of equally optimal or near-optimal phylogenetic solutions for a given dataset (Maddison, 1991). In analyzing a particular data set, one might discover that there are a large number of solutions that fit the data equally well. A consensus of these trees may show little or no resolution. However, an unresolved consensus tree does not necessarily indicate that all potential solutions fit the data equally well. Separate summaries of each of the tree islands is likely to show a much higher degree of resolution, and the separate tree islands may represent alternative phylogenetic solutions for the data set. Tree Set Visualization can be used to identify and analyze these tree islands, as shown in Figure 9.

Potential Limitations of MDS for Visualizing Tree-Space

Multidimensional scaling based on RF distances is clearly not the only way (and is not necessarily even the best way) to visualize and represent tree-space. We have found this approach to the problem to be useful for exploring large sets of phylogenetic trees, but we also recognize that the approach has some limitations. For instance, any reduction of high-dimensional space into two dimensions necessarily will result in some distortions. As an example of distortion, consider the MDS visualization of trees shown in Figure 10. In this case, a reference tree is shown in blue, and a series of trees that differ from the reference tree by one bipartition each ($RF = 2$) are shown in red. All of the trees are equally distant from the reference tree in tree-space, and in multiple dimensions would form a “multidimensional sphere” around the

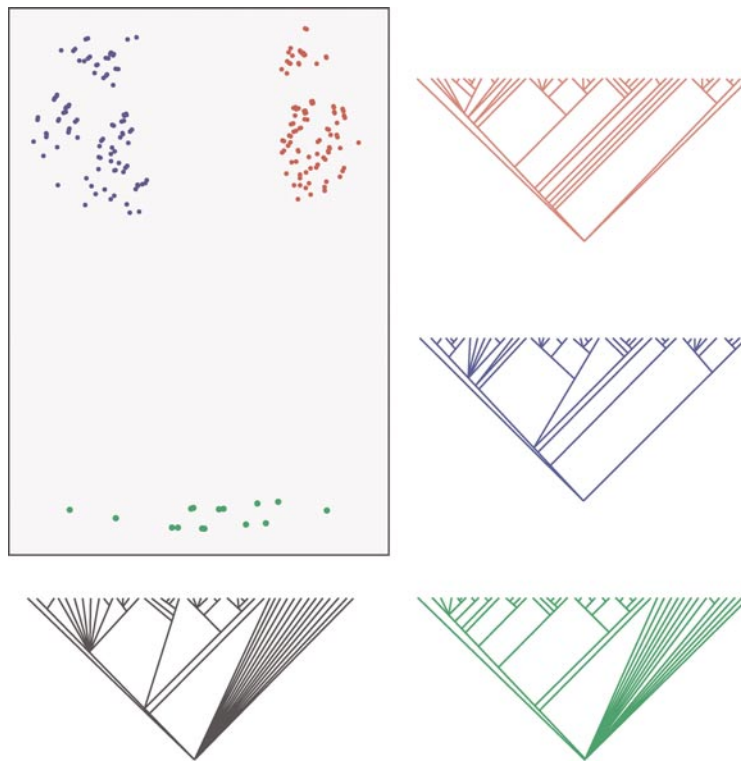


FIGURE 9. The use of multidimensional scaling to visualize “tree islands” of trees with high optimality scores. The three discrete clusters of dots (shown in green, blue, and red) represent three tree islands. The Tree Set Visualization software can be used to find and examine separate consensus trees for each tree island (all of which are much more highly resolved than is the consensus tree of all the trees combined). Each strict consensus tree is colored corresponding to the cluster of trees it represents. The black tree is a strict consensus of all the trees in the three separate tree islands.

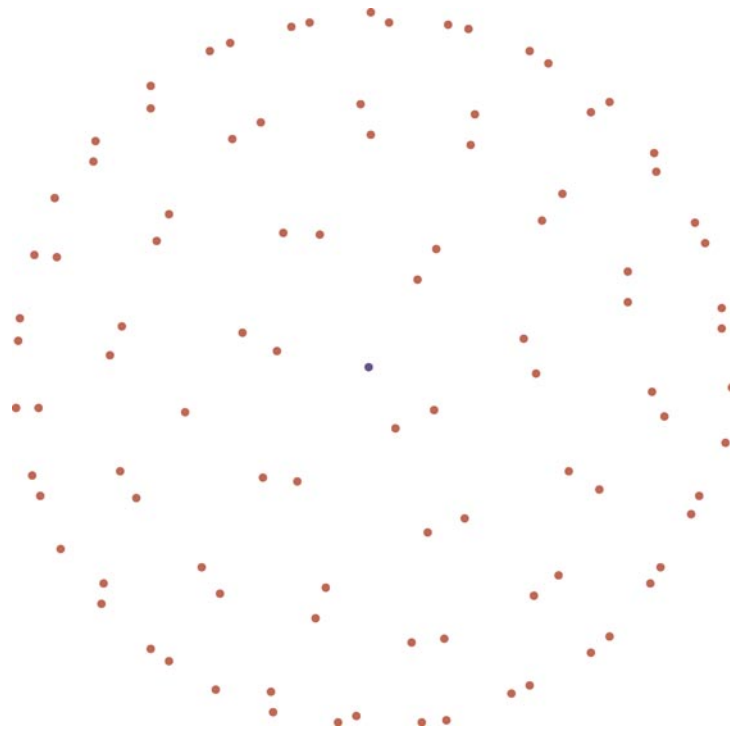


FIGURE 10. Distortion of tree space in two dimensions. In this example, all of the trees indicated by red dots differ from the reference tree (shown in blue) by exactly one bipartition ($RF = 2$). In multidimensional space, they form a sphere around the reference tree. In two dimensions, this sphere is collapsed into a circle, and some of the trees appear to be closer to the reference tree than others, despite the fact that they are all equally distant from the reference tree. Pairs of trees in close proximity to each other also differ from each other by $RF = 2$, and all other trees represented by red dots are $RF = 4$ from one another.

reference tree. In addition, pairs of trees in close proximity in Figure 10 differ from each other by one bipartition, and all other trees represented by red dots differ from one another by two bipartitions ($RF = 4$). The stress function in the MDS analysis is minimized in two dimensions with the representation shown in Figure 10. It would be easy to misinterpret this diagram, and one might think that the trees closer to the center of the figure were closer in RF distance to the reference tree than those that are represented on the outside of the circle. Therefore, it is important to understand the potential for such distortions, and to check the primary distance matrix before interpreting too much about the spatial relationships of trees shown in the MDS analyses.

A second caveat that must be considered is that the tree space in these examples is entirely dependent on the sample of trees examined. In other words, the space is redefined and redistorted with the addition of any new tree, and the space visualized does not exist in the absence of the sample at hand. Therefore, care should be taken to avoid over-interpretation of results. Ideally, one might want to define the tree space by considering all possible solutions, and then plot the results of a particular analysis within the universe of possible solutions. This may be possible for smaller data sets, although the large number of possible tree topologies for even a modest number of taxa presents a serious challenge to this approach.

A third limitation concerns the use of RF distances to define the tree-space. This is appropriate when the related trees are connected to one another by nearest-neighbor interchanges (Penny and Hendy, 1985). However, two trees may differ only in the placement of a single taxon, and yet exhibit the maximum possible RF distance from one another. It would be useful to explore other methods for defining the pairwise distances between trees and using these distances to define and analyze tree space.

A fourth limitation involves the use of multidimensional scaling. In some cases, it is possible for the MDS analysis to become trapped in local optima. This can be avoided, however, by altering the initial states, conducting multiple restarts, and changing the step size used by the MDS algorithm (Borg and Groenen, 1997). It is also important to consider that visualization by multidimensional scaling relies on human pattern recognition skills to identify clusters of trees. Stockham et al. (2002) have developed a more formal and quantifiable method for clustering sets of phylogenetic trees.

CONCLUSIONS

Our intent in this article was to present an approach for multidimensional scaling of tree-space and to suggest some of its potential applications. We acknowledge that we have not thoroughly developed any of the suggested possible uses of this approach in this article. However, we

hope that this article will stimulate additional research on the applications that we have suggested, and that the availability of a program for conducting these analyses will aid in these investigations. At the minimum, it appears that the approach presents a useful means for visualizing the results of large sets of phylogenetic trees. We expect that it will lead to new ways of thinking about samples of trees, beyond the usual consensus summaries.

ACKNOWLEDGEMENTS

We thank Derrick Zwickl for comments and assistance, and for suggesting Figure 10. Nina Amenta and Wayne Maddison provided considerable assistance and advice on the use of Tree Set Visualization and Mesquite. We also thank the other individuals who helped develop the Tree Set Visualization module, including Jeff Klingner, Fred Clarke, Denise Edwards, Silvio Neris, and Ruchi Mahindru. Andy Anderson, Peter Foster, Tamara Munzner, Rod Page, and Alisha Holloway provided helpful suggestions on the manuscript and software. Funding for this project was provided by the National Science Foundation (NSF ITR 0121682/0121651 to Nina Amenta and David Hillis at University of Texas and Katherine St. John at City University of New York), and Tracy Heath was supported on an NSF IGERT fellowship in Computational Phylogenetics and Applications to Biology at the University of Texas.

REFERENCES

- Alfaro, M. E., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* 20:255–266.
- Amenta, N., and J. Klingner. 2002. Case study: Visualizing sets of evolutionary trees. 8th IEEE Symposium on Information Visualization 2002:71–74.
- Borg, I., and P. Groenen. 1997. *Modern multidimensional scaling*. Springer-Verlag, Heidelberg.
- Buckley, T. R., P. Arensburger, C. Simon, and G. K. Chambers. 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Syst. Biol.* 51:4–18.
- Buneman, P. 1971. The recovery of trees from measures of dissimilarity. Pages 387–395 in *Mathematics in archaeological and historical sciences*. (F. R. Hodson, D. G. Kendall, and P. Tautu, eds.). Edinburgh University Press, Edinburgh.
- Cummings, M. P., S. A. Handley, D. S. Myers, D. L. Reed, A. Rokas, and K. Winka. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* 55:477–487.
- Day, W. H. E. 1985. Optimal algorithms for comparing trees with labeled leaves. *J. Classification* 2:7–28.
- Efron, B., E. Halloran, and S. Holmes. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 93:7085–7090.
- Erixon, P., B. Sennblad, T. Britton, and B. Oxelman. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52:665–673.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- Geyer, C. J. 1991. Markov chain Monte Carlo maximum likelihood. Pages 156–163 in *Computing science and statistics: Proceedings of the 23rd Symposium on the Interface* (E. M. Keramidas, ed). Interface Foundation, Fairfax Station.
- Gilks, W. R., and G. O. Roberts. 1996. Strategies for improving MCMC. Pages 89–114 in *Markov Monte Carlo in practice* (W. R. Gilks, G. O. Roberts, and D. J. Spiegelhalter, eds). Chapman and Hall, London.
- Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42:182–192.
- Hillis, D. M., and T. P. Wilcox. 2005. Phylogeny of the New World true frogs (*Rana*). *Mol. Phylogenet. Evol.* 34:299–314.
- Huelsenbeck, J. P., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.* 53:904–913.

- Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Huelsenbeck, J. P., F. Ronquist, R. Nielson, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314.
- Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750–759.
- Li, W.-H., and A. Zharkikh. 1994. What is the bootstrap technique? *Syst. Biol.* 43:424–430.
- Lingoes, J. C., E. E. Roskam, and I. Borg. 1979. *Geometric representations of relational data*, 2nd edition. Mathesis Press, Ann Arbor, Michigan.
- Maddison, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Syst. Zool.* 40:315–328.
- Maddison, W. P., and D. R. Maddison. 1992. *MacClade: Analysis of phylogeny and character evolution*. Sinauer, Sunderland, Massachusetts (see also <http://www.macclade.org>).
- Maddison, W. P., and D. R. Maddison. 2004. *Mesquite: A modular system for evolutionary analysis*. Version 1.02. <http://mesquiteproject.org>.
- Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.
- Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53:47–67.
- Penny, D., and M. D. Hendy. 1985. The use of tree comparison metrics. *Syst. Zool.* 34:75–82.
- Posada, D., and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* 50:580–601.
- Rambaut, A., and N. C. Grassly. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* 13:235–238.
- Robinson, D. F., and L. R. Foulds, 1979. Comparison of weighted labeled trees. *Lect. Notes Math.* 748:119–126.
- Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53:131–147.
- Rodrigo, A. G. 1993. Calibrating the bootstrap test of monophyly. *Int. J. Parasitol.* 23:507–514.
- Stockham, C., L.-S. Wang, and T. Warnow. 2002. Statistically based postprocessing of phylogenetic analysis by clustering. *Bioinformatics* 18:S285–S293.
- Suzuki Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. USA* 99:16138–16143.
- Swofford, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pages 295–333 in *Phylogenetic analysis of DNA sequences* (M. M. Miyamoto and J. Cracraft, eds.). Oxford University Press, New York.
- Swofford, D. L. 2000. *PAUP*: Phylogenetic analysis using parsimony (and other methods)*. Sinauer Associates, Sunderland, Massachusetts.
- Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. *Phylogenetic inference*. Pages 407–514 in *Molecular systematics*, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.
- Whittingham, L. A., B. Slikas, D. W. Winkler, and F. H. Sheldon. 2002. Phylogeny of the tree swallow genus *Tachycineta* (Aves: Hirundinidae), by Bayesian analysis of mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 22:430–441.
- Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* 25:361–371.
- Young, F. W. and R. M. Hamer. 1987. *Multidimensional scaling: History, theory and applications*. Erlbaum, New York.
- Zharkikh, A., and W.-H. Li. 1992a. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Mol. Biol. Evol.* 9:1119–1147.
- Zharkikh, A., and W.-H. Li. 1992b. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. II. Four taxa without a molecular clock. *J. Mol. Evol.* 35:356–366.
- Zharkikh, A., and W.-H. Li. 1995. Estimation of confidence in phylogeny: The complete-and-partial bootstrap technique. *Mol. Phylogenet. Evol.* 4:44–63.

First submitted 23 May 2004; reviews returned 3 August 2004;
final acceptance 26 August 2004
Associate Editor: Frank Anderson

APPENDIX 1

TREE DESCRIPTION OF THE TREE FROM MURPHY ET AL. (2000)

((Opossum: 0.072454, Diprotodontian: 0.061694):0, (((Sloth: 0.056950, Anteater: 0.061637):0.009169, Armadillo: 0.056660):0.032179, (((Hedgehog: 0.137379, Shrew: 0.124147):0.011789, Mole: 0.086828):0.011954, (((Phyllostomid: 0.093178, Free tailed bat: 0.046665):0.011564, (False vampire bat: 0.062583, (Flying Fox: 0.018553, Rousette Fruitbat: 0.018931):0.036729):0.004788):0.016400, (((((Whale: 0.013788, Dolphin: 0.021978):0.019568, Hippo: 0.039894):0.004885, Ruminant: 0.073210):0.008450, Pig: 0.067448):0.005893, Llama: 0.061851):0.027757, ((Horse: 0.043682, (Rhino: 0.028867, Tapir: 0.028638):0.005116):0.020583, ((Cat: 0.046372, Caniform: 0.055840):0.023068, Pangolin: 0.075956):0.003871):0.001685):0.001155):0.002432):0.011058, ((Sciurid: 0.083962, (Mouse: 0.042059, Rat: 0.045451):0.122018, (Hystricid: 0.074622, Caviomorph: 0.086677):0.062121):0.005432):0.011864, (Rabbit: 0.057873, Pika: 0.108683):0.043771):0.005743, ((Flying Lemur: 0.061380, Tree Shrew: 0.101818):0.003958, (Strepsirrhine: 0.076186, Human: 0.065099):0.009553):0.001707):0.007711):0.009175):0.005977, (((Tenrecid: 0.142758, Golden Mole: 0.067180):0.009411, (Short Eared Elephant Shrew: 0.039055, Long Eared Elephant Shrew: 0.036033):0.088816):0.002240, Aardvark: 0.068518):0.003248, ((Sirenian: 0.038154, Hyrax: 0.089482):0.002916, Elephant: 0.050883):0.014801):0.025967):0.284326)

APPENDIX 2. Preferred model and estimated parameters for the genes studied by Murphy et al. (2001):

| Gene | Preferred model | Base frequencies | | | | Relative substitution rates | | | | | | Proportion invariant sites | Alpha |
|--------|-----------------|------------------|------|------|------|-----------------------------|------|------|------|------|----|----------------------------|-------|
| | | A | C | G | T | AC | AG | AT | CG | CT | GT | | |
| ADORA3 | K2P | 0.25 | 0.25 | 0.25 | 0.25 | 1 | 3 | 1 | 1 | 3 | 1 | 0 | |
| ADRB2 | HKY+I+G | 0.2 | 0.33 | 0.25 | 0.22 | 1 | 5.75 | 1 | 1 | 5.75 | 1 | 0.46 | 1.05 |
| APP | GTR+I+G | 0.25 | 0.24 | 0.18 | 0.33 | 1.6 | 3.66 | 0.47 | 0.72 | 2.65 | 1 | 0 | 0.78 |
| ATP7A | GTR+I+G | 0.33 | 0.21 | 0.19 | 0.19 | 1.11 | 5.33 | 0.68 | 0.92 | 4.43 | 1 | 0.2 | 1.56 |
| BDNF | HKY+I+G | 0.21 | 0.33 | 0.28 | 0.17 | 1 | 4.73 | 1 | 1 | 4.73 | 1 | 0.42 | 0.61 |
| BMI1 | GTR+I+G | 0.29 | 0.15 | 0.16 | 0.4 | 2.35 | 7.08 | 0.64 | 1.77 | 5.71 | 1 | 0.14 | 0.82 |
| CNR1 | GTR+I+G | 0.18 | 0.32 | 0.25 | 0.24 | 3.43 | 14 | 1.3 | 2.13 | 14.6 | 1 | 0.53 | 0.7 |
| CREM | GTR+I+G | 0.21 | 0.24 | 0.28 | 0.27 | 1.68 | 3.44 | 0.55 | 0.8 | 2.97 | 1 | 0.18 | 1.6 |
| EDG1 | HKY+I+G | 0.17 | 0.36 | 0.27 | 0.2 | 1 | 4.93 | 1 | 1 | 4.93 | 1 | 0.44 | 0.72 |
| PLCB4 | GTR+I+G | 0.3 | 0.27 | 0.19 | 0.24 | 0.94 | 2.77 | 0.59 | 0.56 | 2.33 | 1 | 0.04 | 2.88 |
| PNOC | GTR+I+G | 0.23 | 0.33 | 0.31 | 0.12 | 0.9 | 2.73 | 0.86 | 0.38 | 4.14 | 1 | 0.15 | 1.09 |
| RAG1 | GTR+I+G | 0.21 | 0.3 | 0.29 | 0.19 | 2.04 | 5.59 | 1.01 | 0.67 | 9.09 | 1 | 0.49 | 1.07 |
| RAG2 | HKY+I+G | 0.28 | 0.24 | 0.22 | 0.27 | 1 | 6 | 1 | 1 | 6 | 1 | 0.35 | 1.63 |
| TYR | GTR+I+G | 0.24 | 0.26 | 0.25 | 0.25 | 2.18 | 7.86 | 1.3 | 0.93 | 8.76 | 1 | 0.32 | 1.27 |
| ZFX | HKY+I+G | 0.35 | 0.23 | 0.18 | 0.23 | 1 | 7.94 | 1 | 1 | 7.94 | 1 | 0.49 | 1.24 |
| VWF | HKY+I+G | 0.2 | 0.34 | 0.28 | 0.18 | 1 | 4.41 | 1 | 1 | 4.41 | 1 | 0.15 | 0.92 |
| BRCA1 | GTR+I+G | 0.33 | 0.22 | 0.23 | 0.22 | 1.15 | 4.38 | 0.75 | 1.17 | 4.75 | 1 | 0.04 | 3.4 |
| IRBP | GTR+I+G | 0.21 | 0.3 | 0.3 | 0.18 | 1.5 | 4.91 | 1.34 | 0.83 | 5.8 | 1 | 0.18 | 1.04 |
| A2AB | GTR+I+G | 0.17 | 0.34 | 0.3 | 0.18 | 1.02 | 3.59 | 0.93 | 0.62 | 3.71 | 1 | 0.3 | 1.29 |
| mtRNA | GTR+I+G | 0.34 | 0.2 | 0.21 | 0.25 | 5.86 | 14 | 3.85 | 0.58 | 29.3 | 1 | 0.41 | 0.53 |

APPENDIX 3. Pairwise Euclidian distances between the models of sequence evolution for each gene studied by Murphy et al. (2001).

| | ADORA3 | ADRB2 | APP | ATP7A | BDNF | BMI1 | CNR1 | CREM | EDG1 | PLCB4 | PNOC | RAG1 | RAG2 | TYR | ZFX | VWF | BRCA1 | IRBP | A2AB | mtRNA | |
|--------|--------|-------|------|-------|------|------|------|------|------|-------|------|------|------|------|------|------|-------|------|------|-------|--|
| ADORA3 | 0 | | | | | | | | | | | | | | | | | | | | |
| ADRB2 | 1.07 | 0 | | | | | | | | | | | | | | | | | | | |
| APP | 1.29 | 0.61 | 0 | | | | | | | | | | | | | | | | | | |
| ATP7A | 0.69 | 0.45 | 0.7 | 0 | | | | | | | | | | | | | | | | | |
| BDNF | 1.69 | 0.69 | 0.6 | 1.03 | 0 | | | | | | | | | | | | | | | | |
| BMI1 | 1.24 | 0.51 | 0.21 | 0.62 | 0.61 | 0 | | | | | | | | | | | | | | | |
| CNR1 | 1.54 | 0.5 | 0.6 | 0.89 | 0.26 | 0.54 | 0 | | | | | | | | | | | | | | |
| CREM | 0.66 | 0.46 | 0.69 | 0.21 | 1.05 | 0.63 | 0.9 | 0 | | | | | | | | | | | | | |
| EDG1 | 1.47 | 0.44 | 0.52 | 0.82 | 0.25 | 0.5 | 0.14 | 0.83 | 0 | | | | | | | | | | | | |
| PLCB4 | 0.36 | 0.75 | 0.94 | 0.36 | 1.35 | 0.9 | 1.21 | 0.35 | 1.13 | 0 | | | | | | | | | | | |
| PNOC | 0.95 | 0.36 | 0.51 | 0.38 | 0.78 | 0.51 | 0.68 | 0.38 | 0.58 | 0.63 | 0 | | | | | | | | | | |
| RAG1 | 1.08 | 0.16 | 0.67 | 0.49 | 0.72 | 0.58 | 0.52 | 0.49 | 0.48 | 0.78 | 0.36 | 0 | | | | | | | | | |
| RAG2 | 0.72 | 0.38 | 0.77 | 0.2 | 1.03 | 0.66 | 0.85 | 0.24 | 0.8 | 0.43 | 0.43 | 0.4 | 0 | | | | | | | | |
| TYR | 0.86 | 0.24 | 0.62 | 0.25 | 0.86 | 0.52 | 0.69 | 0.26 | 0.63 | 0.54 | 0.29 | 0.26 | 0.19 | 0 | | | | | | | |
| ZFX | 0.97 | 0.25 | 0.72 | 0.35 | 0.86 | 0.59 | 0.67 | 0.44 | 0.64 | 0.67 | 0.46 | 0.28 | 0.26 | 0.23 | 0 | | | | | | |
| VWF | 1.1 | 0.34 | 0.35 | 0.49 | 0.61 | 0.37 | 0.53 | 0.49 | 0.42 | 0.76 | 0.21 | 0.4 | 0.54 | 0.37 | 0.5 | 0 | | | | | |
| BRCA1 | 0.32 | 0.8 | 1.01 | 0.4 | 1.4 | 0.95 | 1.26 | 0.4 | 1.19 | 0.13 | 0.67 | 0.81 | 0.46 | 0.58 | 0.7 | 0.82 | 0 | | | | |
| IRBP | 0.99 | 0.3 | 0.44 | 0.39 | 0.71 | 0.42 | 0.6 | 0.38 | 0.51 | 0.66 | 0.11 | 0.33 | 0.42 | 0.25 | 0.42 | 0.14 | 0.7 | 0 | | | |
| A2AB | 0.85 | 0.25 | 0.64 | 0.3 | 0.87 | 0.59 | 0.71 | 0.26 | 0.63 | 0.54 | 0.25 | 0.29 | 0.27 | 0.16 | 0.34 | 0.34 | 0.59 | 0.24 | 0 | | |
| mtRNA | 1.97 | 1 | 0.82 | 1.31 | 0.41 | 0.81 | 0.56 | 1.34 | 0.62 | 1.63 | 1.06 | 0.99 | 1.31 | 1.14 | 1.12 | 0.92 | 1.67 | 1.01 | 1.18 | 0 | |