NEW TRENDS IN PLANT SYSTEMATICS

Bayesian inference of phylogeny: a non-technical primer

Jenny K. Archibald[1], Mark E. Mort[2] & Daniel J. Crawford[2]

[1] *Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, Ohio 43210, U.S.A. archibald.7@osu.edu*
[2] *Department of Ecology and Evolutionary Biology & Museum of Natural History and Biodiversity Research Center, University of Kansas, Lawrence, Kansas 66045, U.S.A. memort@ku.edu (author for correspondence); dcrawfor@ku.edu*

In our initial contribution to this column, we briefly commented upon the impact that phylogeny reconstruction has had on systematic botany. We further indicated that we feel that plant systematics is currently in a period of reevaluation of the data we have used as well as the methodology employed to estimate phylogeny. This is not to say that issues such as the use of morphological data versus molecular data or whether or not to combine data for phylogenetic analyses have not been debated in the past. Clearly, these and many other issues have received a considerable amount of debate in the phylogenetic literature.

Possibly one of the most rigorously debated topics is the choice of an optimality criterion. In general, there are three basic methods that have been used to estimate phylogeny, including distance, maximum parsimony (MP), and maximum likelihood (ML). The relative merits and shortcomings of these methods have been debated for a number of years (e.g., Faith, 1985; Swofford & Olsen, 1990; Kunhner & Felsenstein, 1994; Huelsenbeck, 1995; Farris & al., 1996; Lewis, 1998; Steel & Penny, 2000), and it is not within the scope of this column to reiterate these discussions. However, it is noteworthy that numerous comparative studies employing both known phylogenies and simulated data have been very useful in determining under what set of conditions each of the methods "out performs" the others. For example, it is now generally accepted that when rates of change along branches vary greatly, employing a parsimony optimality criterion may be misleading due to "long branch attraction" (Felsenstein, 1978; but see Siddall, 1998); whereas additional studies have shown that ML may be inconsistent in other situations, such as when the chosen model of evolution is inappropriate (e.g., Farris, 1999). Simulation studies indicate that distance methods (especially UPGMA) are highly susceptible to variations in evolutionary rates and typically perform more poorly than either MP or ML (e.g., Huelsenbeck & Hillis, 1993). Studies such as these have been important in laying a theoretical foundation for making decisions on how best to estimate phylogeny given the data in hand. However, under most sets of realistic conditions, comparison of ML and MP indicates that these methods perform similarly and often result in highly concordant topologies (e.g., Reed & al., 2002; Kimball & al., 2003). Recently, another round of comparative studies has begun to address a new approach for phylogeny reconstruction (e.g., Suzuki & al., 2002; Wilcox & al., 2002; Alfaro & al., 2003; Douady & al., 2003).

This new approach, Bayesian analyses, was proposed in 1996 (Rannala & Yang, 1996; Mau, 1996; Li, 1996) and is now receiving much attention in the literature [e.g., see *Systematic Biology* 51 (5)]. Several excellent technical reviews have recently been provided by Huelsenbeck & al. (2001, 2002) and Lewis (2001). Although this approach is now a "hot" topic in systematics, Bayesian statistics actually dates back to the 18th century and its utility for reconstructing phylogeny was suggested initially in 1968 by Felsenstein (see Huelsenbeck & al., 2002). It is only recently, however, that these methods have become more widely known and that relevant computer programs have become available. Internet links for downloadable programs for Bayesian analyses (and other methods) are available from the websites of J. Felsenstein (http://evolution.genetics.washington.edu/phylip/software.html), and P. Lewis (http://lewis.eeb.uconn.edu/lewishome/), or directly from the MrBayes (http://morphbank.ebc.uu.se/mrbayes/; Huelsenbeck & Ronquist, 2001) or BAMBE (http://www.mathcs.duq.edu/larget/bambe.html; Simon & Larget, 1998) websites. A helpful introduction on how to use MrBayes is provided by Hall (2001). Here we attempt to provide a basic introduction to Bayesian approaches to phylogeny reconstruction. In doing so, we point out what we feel are some of the most significant attributes of this new approach. It is important to note that we do not consider ourselves to be experts on this topic, but merely are interested in how this approach differs from other methods and how to implement this methodology into our own research, should we feel it appropriate. Thus, we hope this column will serve as a primer for those of you curious about Bayesian methods.

How does Bayesian inference differ from other methods of phylogenetic inference? Perhaps this question is best answered by first noting the major differences between Bayesian and classical statistics. We are all familiar with the use of classical statistics to examine a current set of data to test specific hypotheses; Bayesian statistics differs in that in addition to the current data, prior knowledge is included in the testing of the hypothesis. To illustrate how prior knowledge can affect one's conclusions, we will use an hypothetical example involving a medical test for an illness (C. P. Randle, pers. comm.). Assume that previous studies have evaluated the accuracy of this test and have shown that, if you are in fact ill, there is a 99% likelihood that the test will give a true positive result (and thus, a 1% likelihood that the test will give a false negative). It was also found that if you are healthy, there is a 0.1% likelihood of a false positive result from the test. If we were simply using the "data" (i.e., the test result), we would then conclude that a positive test result had approximately a 99% chance of being correct. How does this conclusion change with the incorporation of prior knowledge? If we were to examine this question in a Bayesian framework, we could incorporate prior knowledge—in this case that other studies have shown that the base rate of this illness is 0.1% in the population. Thus, of a population of 100,000 individuals, 100 would be ill and 99,900 would be healthy. Using the likelihood values mentioned above, we could conclude that a positive test result would be seen in 99% of the ill individuals (99 true positives) and 0.1% of the healthy individuals (approximately 100 false positives). This leaves us with a conclusion that if a person has a positive test result, there is a 99/199 or approximately 50% chance that the test is correct and this person is actually ill. Therefore, by including prior knowledge of the base rate of the illness in the population, the perceived chance that a positive result indicates that an individual actually has the illness drops from 99% to 50%. Thus, the incorporation of prior knowledge has the potential to greatly influence our interpretation of results.

When applied to phylogeny reconstruction, Bayesian inference is similar to maximum likelihood (ML) in that it employs a likelihood function and an explicitly stated model of nucleotide substitution. Thus, this approach most likely exhibits many of the positive and negative attributes of ML. However, as noted above, to the best of our knowledge, the performance of Bayesian analyses is only beginning to be rigorously tested in a manner similar to ML or MP. An attribute of Bayesian inference that sets it apart from ML is the ability (at least theoretically) to include prior information regarding relationships into the process of phylogeny estimation. This is accomplished by stating a prior probability distribution of trees and can be viewed as either a positive or negative attribute, depending upon the strength and legitimacy of the prior expectation. In practice, implementing realistic priors is not yet possible, so most analyses use a simplistic prior, such as according all trees equal prior probabilities ("flat priors"). Ironically, it is the ability to incorporate prior information that makes this approach truly Bayesian in nature, but it is not yet possible to incorporate this information. However, it is likely that in the future, inclusion of prior beliefs will be possible, and it should be noted that this has the potential to affect the resulting "best" topology. In our opinion the ability to state priors is a double-edged sword in that valid priors could assist in obtaining the "true" phylogeny, whereas invalid priors could lead to an inaccurate estimate of phylogeny. The degree to which incorrect priors could affect the outcome of phylogenetic analyses remains to be tested rigorously (discussed below).

Bayesian inference of phylogeny utilizes Markov Chain Monte Carlo (MCMC) simulation (Metropolis & al., 1953; Green, 1995) in combination with the chosen model and data to produce a posterior probability distribution of trees. A technical description of this method is outside the scope of this paper, but an excellent explanation is given in Huelsenbeck & al. (2002). While the prior probability distribution describes the probability of different trees given previous knowledge, the posterior probability distribution describes the probability of trees considering the prior distribution, the model, and, hopefully most importantly, the data. This distribution of trees is the main product of Bayesian phylogenetic analyses. There are several ways to summarize the posterior probability distribution into a tree-like form that is more easily interpreted. One option is to show the maximum posterior probability estimate of phylogeny (MAP). This is the single tree that was deemed most probable. Another of the several options that have been used is to summarize the distribution of trees as a majority rule consensus tree. Regardless of the exact tree displayed, the majority rule consensus is used to determine support values. Thus the values seen on Bayesian phylogenies are the posterior probabilities for a particular clade, that is, the probability that the clade is "true" given the priors, model, and data.

In this regard, another aspect of Bayesian inference that has received attention involves the apparent discrepancy between posterior probabilities and non-parametric bootstrap values (Suzuki & al., 2002; Wilcox & al., 2002; Alfaro & al., 2003; Douady & al., 2003). Regardless of one's personal opinion on their utility, bootstrap analyses (Felsenstein, 1985) have become one of the standard ways in which relative support values for clades are estimated. Multiple studies using both methods have shown that posterior probabilities are often higher than bootstrap values estimated under either MP or ML (e.g.,

Reed & al., 2002; Wilcox & al., 2002; Alfaro & al., 2003). This has been alternatively interpreted to mean that bootstrap values are overly conservative (Wilcox & al., 2002) or posterior probabilities are overly liberal (Suzuki & al., 2002). As is often the case with hypothesis testing, this appears to be a choice between accepting increased type I or type II errors. Although bootstrap values might be more likely to fail to strongly support a true node (type I error), posterior probabilities might instead fail to reject a false node (type II error). Regardless of the interpretation, it is important to note that these two measures of support are not directly comparable. A relatively lower bootstrap value could indicate the same level of support as a higher posterior probability. Thus, it is not appropriate to claim that you have a more strongly supported Bayesian phylogeny than a previous ML or MP phylogeny simply because the posterior probabilities are higher than the bootstrap values. As noted by Alfaro & al. (2003) and Douady & al. (2003), it appears that both measures of support can be useful as long as their respective limitations and differences are kept in mind. The differences between these two measures, and reasons for those differences, will no doubt continue to be investigated. However, if one is inclined to use model-based phylogeny reconstruction, posterior probabilities have the clear benefit of being much more temporally feasible (discussed below under Computational Abilities). When interpreting posterior probabilities, in addition to the lack of comparability to bootstrap values, it is also important to remember that they are not simply the probability that a clade is "true", but rather the probability that it is true *given* the model and parameters used, the priors, and the data. This becomes evident in studies (e.g., Rydin & Källersjö, 2002) that show alternate datasets or models strongly supporting the "truth" of conflicting clades, for example 100% posterior probability support for ((A,B)C) using model 1 and 100% support for ((B,C)A) using model 2.

GENERAL CONSIDERATIONS

As with any new (or even well established) methodology, there are numerous potentially positive and negative attributes that must be considered. Bayesian approaches to phylogeny reconstruction are clearly in their infancy, and thus it is difficult to thoroughly review the pluses and minuses at this time. In the future, many concerns will undoubtedly be addressed; however, given our collective history as a systematic community, it seems safe to say that as these concerns are set aside, others will arise. At this point, several aspects of Bayesian analyses (although not all are unique to this approach) deserve mention. Other aspects of Bayesian analyses that are currently being discussed in the literature, such as problems with convergence and mixing, will likely be the topics of future columns. For now, we will limit our discussion to three of the considerations that we feel are important.

Model choice. — As with any model-based method of phylogenetic inference, Bayesian analyses could be negatively affected if an invalid model of evolution is implemented. A number of methods are currently available to assist in estimating the "best" model to use for ML analyses (e.g., Modeltest; Posada & Crandall, 1998) and many of these same approaches are being used for Bayesian analyses (Huelsenbeck & al., 2002; Reed & al., 2002). However, whether or not it is appropriate to use these tests in a Bayesian framework has not yet been clearly addressed. Bayesian methods of model testing are also being investigated. These methods have several potential benefits such as not requiring nested hypotheses and having the ability to incorporate uncertainty by integrating over model parameters (Huelsenbeck & al., 2001, 2002; Bollback, 2002). There is no current consensus on the best way to choose a model; however, this issue is worthy of future work because, among other concerns, it appears that posterior probability values could be especially sensitive to model choice (Buckley, 2002; Douady & al., 2003).

Computational abilities. — One clear benefit of Bayesian inference of phylogeny is the increased speed of analyses (relative to ML), which permits more extensive searches to be performed than previously possible in a model-based system. This is due to the implementation of MCMC to estimate the posterior probability distribution, which eliminates much of the complex summation and integration and leaves comparatively simple calculations (described in detail by Huelsenbeck & al., 2002). Although the analyses are still slower than MP, they are much faster than traditional ML analyses. For example, Reed & al. (2002) spent approximately nine days conducting Bayesian analyses while it took approximately 93 days to conduct a comparable set of ML analyses (with nearly identical topologies found from each set of analyses). In this same study, the MP topology, inferred using two different weighting schemes, was highly congruent with both the Bayesian and ML topology and was estimated in only two days (including bootstrap analyses). Obviously, the amount of time needed for Bayesian analyses will vary greatly depending on the dataset and methodology, but it is clear that if one wants to use model-based phylogeny reconstruction, Bayesian analyses can be conducted in a much more feasible period of time than ML. This could also allow for a more thorough exploration of tree space (using more complex models, for example) as well as allowing the investigation of types of questions that were not previously practicable.

Priors. — As demonstrated in the medical-test example, including prior information can affect the outcome of statistical analyses. The degree to which this could affect the outcome of phylogenetic inference is still poorly understood. As stated by Huelsenbeck & al. (2002: 681): "In a typical Bayesian analysis of phylogeny, the results are likely to be rather insensitive to the prior". However, as implied by the vague phrasing of this statement, we do not currently know what effect realistic priors will have on phylogenetic analyses since we are not yet able to implement them. As the systematic community is able to include more realistic priors, it seems likely that this issue will become more widely evaluated and discussed. Including valid prior information, if it exists, into the estimation of phylogenetic relationships could be reasonable. In fact, despite our drive for objectivity in systematics, decisions such as which taxa to include, or what data to use, often require consideration of our prior knowledge of the taxa studied. However, these decisions do differ from stating a prior probability distribution of trees. Clearly, several questions need to be addressed in the future. For example, what type of "knowledge" is valid versus invalid for use as priors? How are conflicts among different sets of prior information resolved? Does including priors strongly bias the results of phylogenetic inference? An optimistic view of priors is that including more information is a good thing and that inferences of evolutionary relationships will only be strengthened by including as much biological information as possible. A less optimistic, but equally valid, point of view is that including prior "knowledge" has the potential to bias the resulting topology towards the answer that we "believed" was correct before doing the analyses. The use of priors is the defining characteristic of Bayesian analyses, yet at the moment we are unable to implement realistic priors into phylogenetic analyses and in fact are not certain whether or not it is something that should be done. Clearly, much future work is needed regarding the use of Bayesian analyses in phylogenetics. It is yet to be seen if this currently popular methodology will rise to dominance, fall to obscurity, or settle somewhere in between as one of several methodological options available for phylogenetic analyses.

ACKNOWLEDGEMENTS

We would like to thank Paul Lewis and Chris Randle for helpful discussions, while emphasizing that the opinions expressed in this column are those of the authors.

LITERATURE CITED

Alfaro, M. E., Zoller, S. & Lutzoni, F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molec. Biol. Evol.* 20: 255–266.

Bollback, J. P. 2002. Bayesian model adequacy and choice in phylogenetics. *Molec. Biol. Evol.* 19: 1171–1180.

Buckley, T. R. 2002. Model misspecification and probabilistic tests of topolopy: evidence from empirical data sets. *Syst. Biol.* 51: 509–523.

Douady, C. J., Delsuc, F., Boucher, Y., Doolittle, W. F. & Douzery, E. J. P. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molec. Biol. Evol.* 20: 248–254.

Faith, D. P. 1985. Distance methods and the approximation of most-parsimonious trees. *Syst. Zool.* 34: 312–325.

Farris, J. S. 1999. Likelihood and inconsistency. *Cladistics* 15: 199–204.

Farris, J. S., Albert, V. A., Källersjö, M., Lipscomb, D. & Kluge, A. G. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99–124.

Felsenstein, J. 1968. *Statistical Inference and the Estimation of Phylogenies*. Ph.D. thesis, Univ. Chicago, Chicago.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27: 401–410.

Felsenstein, J. 1985. Confidence limits on phylogeny: an approach using the bootstrap. *Evolution* 39: 783–791.

Green, P. J. 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711–732.

Hall, B. G. 2001. *Phylogenetic Trees Made Easy. A How-To Manual for Molecular Biologists.* Sinauer Associates, Sunderland, Massachusetts.

Huelsenbeck, J. P. 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.* 44: 17–48.

Huelsenbeck, J. P. & Hillis, D. M. 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42: 247–264.

Huelsenbeck, J. P., Larget, B., Miller, R. E. & Ronquist, F. 2002. Potential application and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51: 673–688.

Huelsenbeck, J. P. & Ronquist, F. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17: 754–755.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. 2001. Bayesian influence of phylogeny and its impact on evolutionary biology. *Science* 294: 2310–2314.

Kimball, R. T., Crawford, D. J. & Smith, E. B. 2003. Evolutionary processes in the genus *Coreocarpus*: insights from molecular phylogenetics. *Evolution* 57: 52–61.

Kunhner, M. K. & Felsenstein, J. 1994. A simulation comparison of phylogeny—algorithms under equal and unequal evolutionary rates. *Molec. Biol. Evol.* 11: 459–468.

Lewis, P. O. 1998. Maximum likelihood as an alternative to parsimony for inferring phylogeny using nucleotide sequence data. Pp. 132–163 in: Solits, D. E., Soltis, P. S. & Doyle, J. J. (eds.), *Molecular Systematics of Plants II.* Kluwer, Boston.

Lewis, P. O. 2001. Phylogenetic systematics turns over a new leaf. *Trends Ecol. Evol.* 16: 30–37.

Li, S. 1996. *Phylogenetic Tree Construction Using Markov Chain Monte Carlo.* Ph.D. thesis, Ohio State Univ., Columbus.

Mau, B. 1996. *Bayesian Phylogenetic Inference via Markov Chain Monte Carlo Methods.* Ph.D. thesis, Univ. Wisconsin, Madison.

Metropolis, N., Rosenbluth, A. W., Rosenbluth. N., Teller, A. H. & Teller, E. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21: 1087–1091.

Posada, D. & Crandall, K. A. 1998. Modeltest, testing the models of substitution. *Bioinformatics* 14: 817–818.

Rannala, B. & Yang, Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Molec. Evol.* 43: 304–311.

Reed, D. L., Carpenter, K. E. & deGravelle, M. J. 2002. Molecular systematics of the jacks (Perciformes: Carangidae) based on mitochondrial cytochrome *b* sequences using parsimony, likelihood, and Bayesian approaches. *Molec. Phyl. Evol.* 23: 513–524.

Rydin, C. & Källersjö, M. 2002. Taxon sampling and seed plant phylogeny. *Cladistics* 18: 485–513.

Siddall, M. E. 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris Zone. *Cladistics* 14: 209–220.

Simon, D. & Larget, B. 1998. *Bayesian Analysis in Molecular Biology and Evolution (BAMBE), version 1.01 beta.* Department of Mathematics and Computer Science, Duquesne Univ., Pittsburgh.

Steel, M. & Penny, D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Molec. Biol. Evol.* 17: 839–850.

Suzuki, Y., Glazko, G. V. & Nei, M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci U.S.A.* 99: 16138–16143.

Swofford, D. & Olsen, G. 1990. Phylogeny reconstruction. Pp. 411–501 in: Hills, D. & Moritz, C. (eds.), *Molecular Systematics.* Sinauer Associates, Sunderland, Massachusetts.

Wilcox, T. P., Zwickl, D. J., Heath, T. A. & Hillis, D. M. 2002. Phylogenetic relationships of the dwarf boas and a comparison of the Bayesian and bootstrap measures of phylogenetic support. *Molec. Phyl. Evol.* 25: 361–371.