

- Tarrio, R., F. Rodriguez-Trelles, and F. J. Ayala. 2001. Shared nucleotide composition biases among species and their impacts on phylogenetic reconstructions of the Drosophilidae. *Mol. Biol. Evol.* 18:1464–1473.
- Thalmann, O., J. Hebler, H. N. Poinar, S. Pääbo, and L. Vigilant. 2004. Unreliable mtDNA data due to nuclear insertions: A cautionary tale from analysis of humans and other great apes. *Mol. Ecol.* 13:321–335.
- Waddell, P. J. 1995. Statistical methods of phylogenetic analysis: Including Hadamard conjugations, LogDet transforms, and maximum likelihood. Ph.D. thesis, Massey University, New Zealand.
- Waddell, P. J., H. Kishino, and R. Ota. 2000. Rapid evaluation of the phylogenetic congruence of sequence data using likelihood ratio tests. *Mol. Biol. Evol.* 17:1988–1992.
- Wahlberg, N. 2003. The breaking of the sequencer: *Phyciodes* and the fallacy of DNA barcodes. Presentation, Meeting of the Lepidopterist's Society.
- Wahlberg, N., R. Oliveira, and J. A. Scott. 2003. Phylogenetic relationships of *Phyciodes* butterfly species (Lepidoptera: Nymphalidae): Complex mtDNA variation and species delimitations. *Syst. Entol.* 28:257–273.
- Whitfield, J. 2003. DNA barcodes catalogue animals. [http://www.nature.com/nsu/nsu\\_pf/030512/030512-7.html](http://www.nature.com/nsu/nsu_pf/030512/030512-7.html). Nature Science Update.
- Will, K. W., and D. Rubinoff. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20:47–55.
- Yoder, A. D., J. A. Irwin, and B. A. Payseur. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. *Syst. Biol.* 50:408–424.
- Zakharov, E. V., M. S. Caterino, and F. A. H. Sperling. 2004. Molecular phylogeny, historical biogeography and divergence time estimates for swallowtail butterflies of the genus *Papilio* sensu lato. *Syst. Biol.* 53:193–215.
- Zhang D. X., and G. M. Hewitt. 1996a. Nuclear integrations: Challenges for mitochondrial DNA markers. *Trends Ecol. Evol.* 11:247–251.
- Zhang, D. X., and G. M. Hewitt. 1996b. Highly conserved nuclear copies of the mitochondrial control region in the desert locust *Schistocerca gregaria*: Some implications for population studies. *Mol. Ecol.* 5:295–300.
- Zhang, D. X., and G. M. Hewitt. 2003. Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. *Mol. Ecol.* 12:563–584.

First submitted 1 January 2005; reviews returned 4 March 2005;  
final acceptance 20 April 2005

Associate Editor: Vincent Savolainen

*Syst. Biol.* 54(6):961–965, 2005  
Copyright © Society of Systematic Biologists  
ISSN: 1063-5157 print / 1076-836X online  
DOI: 10.1080/10635150500354670

## Hastings Ratio of the LOCAL Proposal Used in Bayesian Phylogenetics

MARK T. HOLDER,<sup>1</sup> PAUL O. LEWIS,<sup>2</sup> DAVID L. SWOFFORD,<sup>1</sup> AND BRET LARGET<sup>3,4</sup>

<sup>1</sup>School of Computational Science, Florida State University, Tallahassee, Florida 32306-4120, USA; E-mail: mholder@csit.fsu.edu (M. T. H.)

<sup>2</sup>Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, Unit 3043, Storrs, Connecticut 06269-3043, USA

<sup>3</sup>Department of Botany, University of Wisconsin–Madison, 430 Lincoln Drive, Madison, WI 53706-1685, USA

<sup>4</sup>Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, Madison, WI 53706, USA

As part of another study, we estimated the marginal likelihoods of trees using different proposal algorithms and discovered repeatable discrepancies that implied that the published Hastings ratio for a proposal mechanism used in many Bayesian phylogenetic analyses is incorrect. In this article, we derive the correct Hastings ratio for the (Larget and Simon, 1999) “LOCAL move without a molecular clock.” The derivation illustrates how a recently described method for determining the acceptance probabilities for proposals in Markov chain Monte Carlo (Green, 2003) provides an intuitive method for calculating Hastings ratios. Although the use of the previously reported Hastings ratio could result in a bias toward shorter branch lengths, the effect is very minor and is overwhelmed by the information contained within even small data sets.

Markov chain Monte Carlo (MCMC) methods are widely used to explore posterior probability densities by simulating a walk through tree/model space (Simon and Larget, 2001; Huelsenbeck and Ronquist, 2001). Many MCMC simulations employ the Metropolis-Hastings algorithm, which uses a stochastic function to propose a new state,  $x'$ , for the chain based upon the current state,  $x$ . The “state of the chain” refers to the values of all of the parameters in the model (including branch lengths

for the tree). Let  $q(x, dx')$  denote the probability density of proposing a move from  $x \rightarrow x'$ . The Metropolis algorithm (Metropolis et al., 1953) is limited to simulation schemes in which  $q(x, dx') = q(x', dx)$ . Hastings (1970) significantly eased the task of implementing MCMC methods by modifying the Metropolis algorithm to allow for the use of asymmetric proposal densities. If one is sampling the posterior density (which is proportional to the product of the likelihood,  $\mathcal{L}$ , and the prior probability density,  $p$ ), then the probability of accepting a proposal,  $\alpha(x, x')$ , in the Metropolis-Hastings algorithm is:

$$\alpha(x, x') = \min \left\{ 1, \left[ \frac{\mathcal{L}(x')}{\mathcal{L}(x)} \right] \left[ \frac{p(x')}{p(x)} \right] \left[ \frac{q(x', dx)}{q(x, dx')} \right] \right\} \quad (1)$$

The factor  $q(x', dx)/q(x, dx')$  is referred to as the Hastings ratio.

### SUMMARY OF GREEN'S CONSTRUCTIVE METHOD FOR CALCULATING ACCEPTANCE PROBABILITIES

Green (2003: Section 2.2) generalized the calculation of acceptance ratios to cover both the Metropolis-Hastings algorithm as well as proposals that implement

the reversible-jump techniques to sample models of differing dimensions (Green, 1995). Green (2003: Section 2.2) describes the calculation of the acceptance probability in terms of every parameter and random number that must be drawn when performing a move. This description makes the connection between MCMC theory and the implementation of a proposal in software more explicit. Starting from state  $\mathbf{x}$ , a set of random number(s),  $\mathbf{u}$ , is generated using a probability distribution with the joint probability density  $g(\mathbf{u})$ . Given these random number(s), a deterministic function generates the proposed state:  $\mathbf{x}' = h(\mathbf{x}, \mathbf{u})$ . To calculate  $\alpha(\mathbf{x}, \mathbf{x}')$  we must consider the move that would exactly reverse the effects of the forward move. To propose  $\mathbf{x}' \rightarrow \mathbf{x}$ , a new set of random numbers,  $\mathbf{u}'$ , are generated according to a (potentially different) distribution with density  $g'(\mathbf{u}')$ . These random numbers and the state  $\mathbf{x}'$  are transformed by another deterministic function to produce a proposed state identical to the original state:  $\mathbf{x} = h'(\mathbf{x}', \mathbf{u}')$ . In Green's (2003) formulation, the Hastings ratio is replaced by factors that depend upon  $g(\mathbf{u})$ ,  $g'(\mathbf{u}')$ , and the absolute value of the Jacobian of the transformation from  $\{\mathbf{x}, \mathbf{u}\}$  to  $\{\mathbf{x}', \mathbf{u}'\}$

$$\alpha(\mathbf{x}, \mathbf{x}') = \min \left\{ 1, \left[ \frac{\mathcal{L}(\mathbf{x}')}{\mathcal{L}(\mathbf{x})} \right] \left[ \frac{p(\mathbf{x}')}{p(\mathbf{x})} \right] \left[ \frac{g'(\mathbf{u}')}{g(\mathbf{u})} \right] |J| \right\} \quad (2)$$

where  $J = \det \left[ \frac{\partial(\mathbf{x}', \mathbf{u}')}{\partial(\mathbf{x}, \mathbf{u})} \right]$ . This presentation is easier to follow than previous formulations because the variables in the equation mirror those that one would use when implementing the algorithm in a computer program.

#### LARGET-SIMON LOCAL MOVE

Larget and Simon (1999) introduced four MCMC proposals for moving through the space of phylogenetic trees: GLOBAL and LOCAL, on clocklike and nonclock trees. Their LOCAL, nonclock move starts by selecting an internal branch in the tree at random (all internal branches have an equal probability of being selected). One node at the end of the branch is randomly assigned the label  $i$ , and the other is referred to as  $j$  (Fig. 1; note that these labels correspond to  $u$  and  $v$  in the original description, but we will instead use  $\mathbf{u}$  to refer to the random numbers generated during the proposal algorithm). One of the nodes adjacent to  $i$  (but not node  $j$ ) is randomly chosen and labeled  $a$ . Similarly, a node adjacent to  $j$  is labeled  $c$ . The associated branch lengths are denoted  $w_{ai}$ ,  $w_{ij}$ , and  $w_{jc}$ . Calculation of the acceptance probability for  $\mathbf{x} \rightarrow \mathbf{x}'$  is most easily accomplished if the fundamental parameters are taken to be the path lengths from  $a$  to each of the other 3 nodes:  $\mathbf{x} = \{w_{ai}, w_{aj}, w_{ac}\}$  and  $\mathbf{x}' = \{w'_{ai}, w'_{aj}, w'_{ac}\}$ . Note that  $w_{aj} = w_{ai} + w_{ij}$  and  $w_{ac} = w_{ai} + w_{ij} + w_{jc}$  (the  $w_{ac}$  path length was referred to as  $m$  by Larget and Simon). This reparameterization does not affect the calculations (because the Jacobian for this reparameterization is 1).

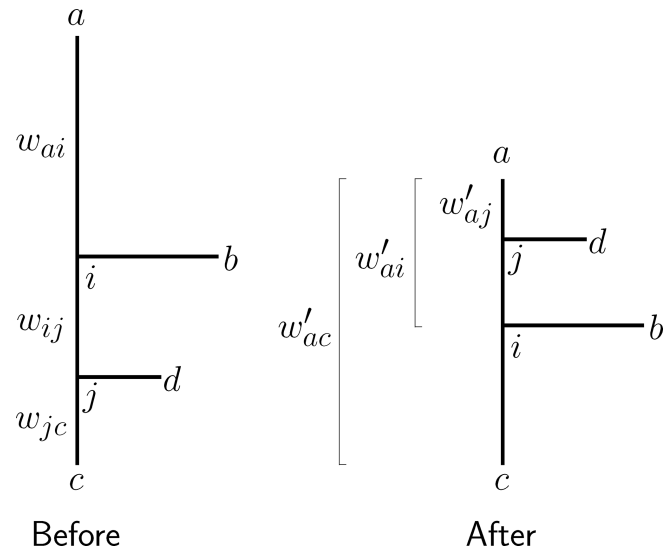


FIGURE 1. Illustration of the notation used to describe the LOCAL move. The panel labeled "Before" shows a tree before a LOCAL move. The panel labeled "After" shows one of the possible trees that can be produced by a LOCAL move if node  $j$  is moved. In this example, the topology of the tree has changed.

The LOCAL move uses two Uniform (0, 1) random variables,  $\mathbf{u} = \{u_1, u_2\}$  and  $\mathbf{u}' = \{u'_1, u'_2\}$ . The densities  $g(\mathbf{u})$  and  $g'(\mathbf{u}')$  will both be equal to 1, and, thus, will cancel out in the calculation of  $\alpha(\mathbf{x}, \mathbf{x}')$ .

After the random numbers are generated, the next step is to rescale the  $a$  to  $c$  path length:

$$w'_{ac} = w_{ac} e^{\lambda(u_1 - 0.5)} \quad (3)$$

where  $\lambda$  is a user-specified tuning parameter for the move. Next, either  $i$  or  $j$  is selected as a node to move. For simplicity we will consider the case in which  $j$  is the node that is repositioned (all results hold in the case of  $i$  being selected as well). The other central node ( $i$  in our case) will maintain its relative position on the path from  $a$  to  $c$ :

$$w'_{ai} = w_{ai} e^{\lambda(u_1 - 0.5)} \quad (4)$$

The new path distance from  $a$  to the "sliding" node is determined by repositioning the node uniformly along the newly proposed  $w'_{ac}$  path:

$$w'_{aj} = u_2 w_{ac} e^{\lambda(u_1 - 0.5)} \quad (5)$$

If the new path length from  $a$  to  $j$  is less than  $w'_{ai}$ , then a topological change has been proposed and the proposed branch lengths can be recovered from the newly proposed path lengths given in Equations 3–5:

$$w'_{ji} = w'_{ai} - w'_{aj} \quad (6)$$

$$w'_{ic} = w'_{ac} - w'_{ai} \quad (7)$$

Note that, in the case of a topological move, the reverse move requires the selection of the same set of nodes, but node  $i$  must be chosen as the node to reposition in the second step of the LOCAL move. This does not affect the acceptance probability because both  $i$  and  $j$  have equal probability of being selected for repositioning. If  $w'_{aj} > w'_{ai}$ , then the topology does not change and the proposed branches are:

$$w'_{ij} = w'_{aj} - w'_{ai} \quad (8)$$

$$w'_{jc} = w'_{ac} - w'_{aj} \quad (9)$$

#### Acceptance Probability Calculation for the LOCAL Move

The Jacobian that appears in Equation 2 is the determinant of the matrix of partial derivatives of  $\mathbf{x}'$  and  $\mathbf{u}'$  with respect to  $\mathbf{x}$  and  $\mathbf{u}$ . Thus, we must derive equations for  $\mathbf{u}'$  in terms of  $\mathbf{x}$  and  $\mathbf{u}$ . The rescaling for the reverse move must recreate the original  $a$  to  $c$  path length:

$$w_{ac} = w'_{ac} e^{\lambda(u'_1 - 0.5)} \quad (10)$$

$$= w_{ac} e^{\lambda(u_1 - 0.5)} e^{\lambda(u'_1 - 0.5)} \quad (11)$$

thus:

$$u'_1 = 1 - u_1 \quad (12)$$

The value  $u'_2$  required for the  $\mathbf{x}' \rightarrow \mathbf{x}$  proposal is simply the original path length from  $a$  to  $j$  expressed as a proportion of the  $a$  to  $c$  path length:

$$u'_2 = \frac{w_{aj}}{w_{ac}} \quad (13)$$

For the sake of brevity, we can denote the multiplier of the  $a$  to  $c$  path length as  $r$ ; note that  $r = e^{\lambda(u_1 - 0.5)} = (\frac{w'_{ac}}{w_{ac}})$ . Differentiating Equations 3, 4, 5, 12, and 13 yields the necessary Jacobian:

$$J = \begin{vmatrix} \frac{\partial w'_{ai}}{\partial w_{ai}} & \frac{\partial w'_{aj}}{\partial w_{ai}} & \frac{\partial w'_{ac}}{\partial w_{ai}} & \frac{\partial u'_1}{\partial w_{ai}} & \frac{\partial u'_2}{\partial w_{ai}} \\ \frac{\partial w'_{ai}}{\partial w_{aj}} & \frac{\partial w'_{aj}}{\partial w_{aj}} & \frac{\partial w'_{ac}}{\partial w_{aj}} & \frac{\partial u'_1}{\partial w_{aj}} & \frac{\partial u'_2}{\partial w_{aj}} \\ \frac{\partial w'_{ai}}{\partial w_{ac}} & \frac{\partial w'_{aj}}{\partial w_{ac}} & \frac{\partial w'_{ac}}{\partial w_{ac}} & \frac{\partial u'_1}{\partial w_{ac}} & \frac{\partial u'_2}{\partial w_{ac}} \\ \frac{\partial u_1}{\partial u_1} & \frac{\partial u_1}{\partial u_1} & \frac{\partial u_1}{\partial u_1} & \frac{\partial u_1}{\partial u_1} & \frac{\partial u_1}{\partial u_1} \\ \frac{\partial w'_{ai}}{\partial u_2} & \frac{\partial w'_{aj}}{\partial u_2} & \frac{\partial w'_{ac}}{\partial u_2} & \frac{\partial u'_1}{\partial u_2} & \frac{\partial u'_2}{\partial u_2} \end{vmatrix}$$

$$= \begin{vmatrix} r & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{w_{ac}} \\ 0 & u_2 r & r & 0 & \frac{-w_{aj}}{w_{ac}^2} \\ \lambda w_{ai} r & u_2 \lambda w_{ac} r & \lambda w_{ac} r & -1 & 0 \\ 0 & w_{ac} r & 0 & 0 & 0 \end{vmatrix}$$

$$= r \times \frac{-1}{w_{ac}} \times w_{ac} r \times -r$$

$$J = r^3 \quad (14)$$

Because the uniform  $g(\mathbf{u})$  and  $g'(\mathbf{u}')$  do not contribute to the acceptance probability, the Hastings ratio for the move is  $r^3$ . The previous derivation can be made algebraically simpler if the path-length multiplier,  $r$ , is treated as a random variable directly with density  $\lambda/r$  on the interval  $(e^{-\lambda/2}, e^{\lambda/2})$ . The tradeoff for a simpler algebraic expression for the Jacobian is the need to derive the density of  $r$ .

#### IMPLICATIONS

The use of an incorrect Hastings ratio has the potential to profoundly affect MCMC analyses. The LOCAL proposal is used in BAMBE (Simon and Larget, 2001) and accounts for roughly 38% of the tree-changing proposals by default in MRBAYES (Huelsenbeck and Ronquist, 2001). Note that other software packages for Bayesian phylogenetics, such as BEAST (Drummond and Rambaut, 2003) or BAli-Phy (Redelings and Suchard, 2005), do not rely on the LOCAL move. The Hastings ratio that was originally reported for the LOCAL move was too small by a factor of  $r$ . Because  $r$  is greater than 1 whenever a proposal increases the length of the path from node  $a$  to  $c$ , the acceptance probability for proposals that increase the tree length will be too low. These moves will be rejected too often. Conversely, using the incorrect Hastings ratio will cause moves that shorten the tree to be accepted more frequently than in a correct implementation. Thus, using the incorrect Hastings ratio is similar to performing MCMC using the correct Hastings ratio, but with an altered prior on branch lengths which favors short branch lengths more strongly than the intended prior (the effects of the incorrect Hastings ratio cannot be emulated exactly by specifying an altered prior because the three edges involved in the LOCAL move change with each iteration of the MCMC). These effects can be detected by running an MCMC chain without data and examining the marginal distribution of the branch lengths; software that implements the incorrect Hastings ratio will result in biased (low) estimates of the branch lengths when compared to the prior distribution on branch lengths. Because the effects are similar to the choice of an alternate, but still vague, prior,

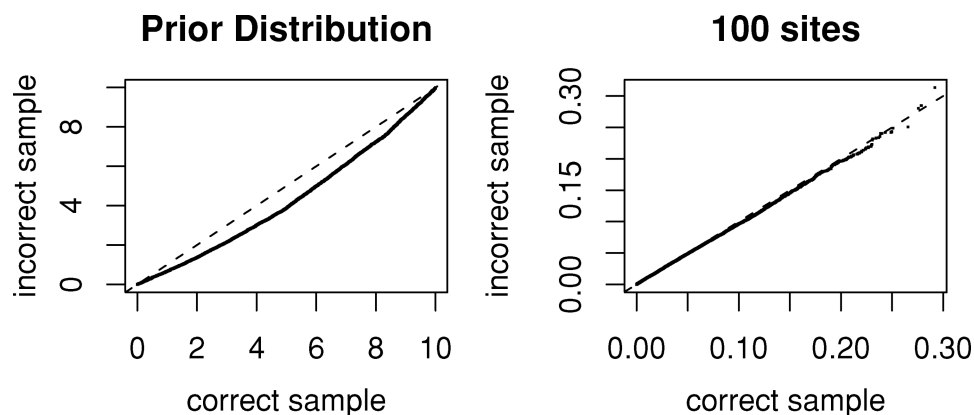


FIGURE 2. Quantile-quantile plots comparing correct and incorrect posterior samples. The plot on the left compares the posterior distributions of the internal branch length from a four-taxon tree for data sampled from the prior distribution. The plot on the right compares the posterior distribution of the internal branch length for a data set with 100 simulated sites.

the implications of using  $r^2$  as the Hastings ratio for the LOCAL move will typically be negligible if there is a moderate (or large) amount of data. Using the incorrect acceptance probabilities has only an indirect effect on estimated clade probabilities. Chains run under the prior distribution (with no data) using the incorrect acceptance probability still sample all possible tree topologies equally.

We examined the effects of the incorrect acceptance probability in two cases. In both cases, the LOCAL move was the only proposal used (after the burnin stage) to modify the tree and branch lengths. Figure 2 shows quantile-quantile plots of the internal branch length in a four-taxon tree for two data sets based on posterior samples from MCMC runs that used the LOCAL move with the correct ( $r^3$ ) and incorrect ( $r^2$ ) Hastings ratio. Quantile-quantile plots are an effective way to compare two different probability distributions. A quantile-quantile plot of two samples of equal size is simply a scatterplot of the sorted samples. Plots of points drawn from the same distribution have high probability of falling close to the line  $y = x$  (with greater variability in the tails), whereas marked curvature in the plots indicate differences between the two distributions. One data set is empty and the resultant sample is taken from a prior distribution with independent uniform (0,10) branch lengths and a uniform probability on the three unrooted tree topologies, under the Jukes-Cantor likelihood model. The second data set contains 100 sites simulated using the Jukes-Cantor model evaluated using the same prior as the first data set. The plot of points sampled from the prior is markedly curved and is consistent with the  $r^2$  sample being skewed toward 0 relative to the  $r^3$  sample. In contrast, the plot of points from the data set with 100 sites is quite straight and the bias remaining due to the error of the incorrect acceptance probability is quite small. In both cases, the differences in calculated posterior probabilities for the three topologies between the two samples

is much smaller than one percent and can be explained by Monte Carlo error.

Figure 3 plots the clade posterior probability estimates from two MCMC analyses: one using the correct acceptance ratio, and the other using  $r^2$ . The data set for this example is the collection of cytochrome *b* sequences (1140 nucleotides) for 31 whales and artiodactyls. This data set is distributed with the BAMBE (Simon and Larget, 2001) and was used as an example in the original description of the LOCAL algorithm (Larget and Simon, 1999). Clearly the clade posteriors are highly correlated,

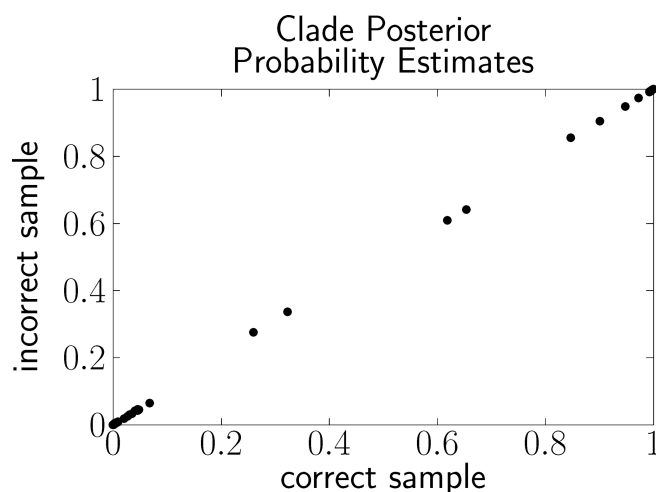


FIGURE 3. Clade posterior correlation plot (Huelsenbeck et al., 2001) for an MCMC analysis with the posterior probability estimated from an incorrect analysis (using incorrect Hastings ratio  $r^2$ ) plotted against the posterior probability for the same clade based on an analysis using the correct Hastings ratio. The data were 31 cytochrome *b* sequences (1140 nucleotides) for whales and artiodactyls (the "whales.dna" file distributed with the BAMBE (Simon and Larget, 2001)). Chains were run from 50 million cycles (starting from the same tree and random number seed). The HKY85 (Hasegawa et al., 1985) model with empirical base frequencies was used. The plot was produced using AWTY (Wilgenbusch et al., 2004).



and use of the incorrect Hastings ratio has a negligible effect on the clade posteriors derived these data; the largest difference in clade posterior probability between the two runs was only 0.016.

#### ACKNOWLEDGEMENTS

We thank Marc Suchard, Jeff Thorne, Rod Page, and an anonymous reviewer for many helpful suggestions for improving the manuscript, and we thank Fredrik Ronquist for valuable discussions on this topic. We thank the National Science foundation for financial support (MTH was supported by award DBI-0306047 and DLS, POL, and MTH were funded by EF 03-31495, part of the CIPRES project). BL acknowledges support from NIH grants R01 GM068950-01 and R01 GM069801-01.

#### REFERENCES

- Drummond, A. J., and A. Rambaut. 2003. *Bayesian Evolutionary Analysis Sampling Trees (BEAST)*, v1.0. Available from <http://evolve.zoo.ox.ac.uk/beat/>.
- Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.
- Green, P. J. 2003. Trans-dimensional Markov chain Monte Carlo. Pages 179–198 in *Highly structured stochastic systems* (P. J. Green, N. L. Hjort, and S. Richardson, Eds.). Oxford University Press, Oxford, UK.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.
- Huelsenbeck, J. P., and F. R. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Huelsenbeck, J. P., F. R. Ronquist, R. Nielsen, 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.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.
- Redelings, B. D., and M. A. Suchard. 2005. Joint Bayesian estimation of alignment and phylogeny. *Syst. Biol.* 54:401–418.
- Simon, D., and B. Larget. 2001. Bayesian analysis in molecular biology and evolution (BAMBE), 2.03 beta edition. Department of Mathematics and Computer Science, Duquesne University.
- Wilgenbusch, J. C., D. L. Warren, and D. L. Swofford. 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference, v0.5. Available from <http://ceb.csit.fsu.edu/awty>.

First submitted 6 December 2004; reviews returned 4 February 2005;  
final acceptance 17 May 2005

Associate Editor: Jeff Thorne

*Syst. Biol.* 54(6):965–973, 2005  
Copyright © Society of Systematic Biologists  
ISSN: 1063-5157 print / 1076-836X online  
DOI: 10.1080/10635150500354654

## Characters, States, and Homology

JOHN V. FREUDENSTEIN

Department of Evolution, Ecology and Organismal Biology, Ohio State University Herbarium, 1315 Kinnear Road, Columbus, Ohio 43212, USA;  
E-mail: [freudenstein.1@osu.edu](mailto:freudenstein.1@osu.edu)

Characters are the fundamental units used to formalize hypotheses of homology for all phylogenetic analyses, meaning that the decision about how observations are translated into characters is of paramount importance in systematics. Clearly, the importance of characters also extends beyond systematics, being central in evolutionary process studies (cf. Gould and Lewontin, 1979), physiology, and any branch of biology that is concerned with the attributes of organisms. Therefore, it is important that an internally consistent, nonarbitrary, yet flexible way of viewing characters be available that can accommodate any type of organismal aspect. It is beyond the scope of this contribution to attempt to solve all problems with character delimitation and coding, but one important issue involving the distinction between characters and states remains problematic that might be clarified via review and consideration in the light of current thinking in systematics.

Although the idea of homologous structures among taxa has a long history (cf. Panchen, 1994), the distinction between the terms *character* and *character state* was not introduced until the middle of the 20th century. Mayr (1942), for example, used the term *character* to denote the particular attribute of an organism (e.g., red flowers, backbone, or five petals), not distinguishing between *character* and *state*. It was with the rise of numerical approaches to taxonomy that the *character/state* distinction became common. Maslin (1952) described a “chronocline” that relates a series of characters through time and is equivalent to the current concept of *transformation series*. Michener and Sokal (1957) distinguished between the *character/state* usage (which they employed) and the practice of calling all attributes simply *characters*, but ascribed no conceptual implications to the difference. Cain and Harrison (1958) did not use the term *state*, but did assign different numerical values to characters. Sneath