

The following resources related to this article are available online at www.sciencemag.org (this information is current as of January 3, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/312/5772/367b>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/312/5772/367b#related-content>

This article **cites 2 articles**, 1 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/312/5772/367b#otherarticles>

This article appears in the following **subject collections**:

Technical Comments

http://www.sciencemag.org/cgi/collection/tech_comment

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Response to Comment on “Phylogenetic MCMC Algorithms Are Misleading on Mixtures of Trees”

Elchanan Mossel^{1*} and Eric Vigoda^{2*}

We presented a tree mixture in which Markov chain Monte Carlo (MCMC) methods have an exponentially slow convergence rate. We expect that many other mixture scenarios will show slow convergence. Ronquist *et al.* show that Metropolis-coupled MCMC (MC³) converges quickly on our mixture. However, they presented no theoretical or systematic experimental evidence determining the type of mixtures where MC³ or other methods are efficient.

Ronquist *et al.* (1) claim that our results (2) depend critically on having exactly equal mixtures, but this is not correct. For a range of proportions of the two trees, there will be multiple local maxima that are not connected by a nearest neighbor interchange (NNI) transition, and the mixing time will be exponentially slow.

We agree with Ronquist *et al.* that mixtures present a challenge to most phylogenetic approaches. However, there is an important difference between methods that return “Fail” when the model is specified incorrectly and methods that find an incorrect tree, especially if this incorrect tree is assigned a high

“confidence value.” We believe that distance-based methods like those described in (3–5) will not be misled by mixtures of two trees. In such cases, the methods should output “Fail” instead of any specific tree or a distribution on trees.

Ronquist *et al.* consider standard heuristic approaches, also suggested in (2), for overcoming the possible perils of Markov chain Monte Carlo (MCMC) algorithms on mixtures, namely multiple starting points, Metropolis-coupled MCMC, or specifying a mixture model. The experimental results reported in (1) suggest that these methods may be adequate to tackle mixtures in some scenarios.

However, the applicability of these methods on some small examples does not guarantee their success in other settings. In particular, these methods might fail for some range of branch lengths or for large trees. We believe that theoretically provable results should be weighted

more heavily compared with limited experiments. We thus argue that much more theoretical and experimental work is needed before MCMC methods can be safely used in mixture settings.

Our tree mixture example was the first result on the efficiency or inefficiency of MCMC methods for phylogenetic reconstruction. Currently, there are no results showing fast convergence of MCMC methods or Metropolis-coupled MCMC for any class of examples. Even in the idealized setting where character data is generated from a pure distribution (i.e., no mixture), it is unclear whether MCMC methods are always efficient.

Building on our work, Stefankovic and Vigoda (6) recently showed refined mixture examples with slow convergence. In their example, the mixture has a common topology and only varies in the substitution rates. They also show a simple mixture example of two trees with a common topology, which generates a distribution that is identical to a mixture distribution from a different topology. Hence, no methods can determine the correct topology, not even those that infer a mixture.

References

1. F. Ronquist *et al.*, *Science* **312**, 367 (2006); www.sciencemag.org/cgi/content/full/312/5772/367a.
2. E. Mossel, E. Vigoda, *Science* **309**, 2207 (2005).
3. P. L. Erdos, M. A. Steel, L. A. Szekely, T. Warnow, *Theor. Comput. Sci.* **221**, 77 (1999).
4. E. Mossel, *IEEE Comp. Bio. and Bioinformatics*, in press; available at arxiv.org/abs/math.CO/0403508.
5. C. Daskalakis *et al.*, *RECOMB 2006*, in press; available at www.eecs.berkeley.edu/~costis/papers/distpaper.pdf.
6. D. Stefankovic, E. Vigoda, available at www.cc.gatech.edu/~vigoda.

5 January 2006; accepted 23 March 2006
10.1126/science.1124180

¹Department of Statistics, University of California at Berkeley, Berkeley, CA 94720, USA. ²College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA.

*To whom correspondence should be addressed. E-mail: mossel@stat.berkeley.edu; E-mail: vigoda@cc.gatech.edu