



# Reply to Gatesy and Springer: The Multispecies Coalescent Model Can Effectively Handle Recombination and Gene Tree Heterogeneity

## Citation

Wu, Shaoyuan, Sen Song, Liang Liu, and Scott V. Edwards. 2013. Reply to Gatesy and Springer: The Multispecies Coalescent Model Can Effectively Handle Recombination and Gene Tree Heterogeneity. *Proceedings of the National Academy of Sciences* 110, no. 13: E1180–E1180.

## Published Version

doi:10.1073/pnas.1300129110

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12763607>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Classification: Biological Sciences

**The multispecies coalescent model can effectively handle recombination and gene tree heterogeneity**

Shaoyuan Wu<sup>a,b,1</sup>, Sen Song<sup>c</sup>, Liang Liu<sup>b</sup>, Scott V. Edwards<sup>d,1</sup>

<sup>a</sup>Institute of Paleontology, Shenyang Normal University, Shenyang, Liaoning 110034, China

<sup>b</sup>Department of Statistics and Institute of Bioinformatics, the University of Georgia, Athens, GA 30606, USA

<sup>c</sup>Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China.

<sup>d</sup>Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Cambridge, MA 02138, USA

<sup>1</sup>To whom correspondence should be addressed. E-mail: shaoyuanwu@outlook.com or sedwards@fas.harvard.edu

Gatesy and Springer (1; “GS”) suggest that coalescent methods cannot build a reliable phylogeny from gene trees, particularly from exonic sequences concatenated *in silico*. However, they confuse several key issues in coalescent theory and phylogenetic reconstruction, and their criticisms are unfounded.

GS cited two papers to claim that results of recent mammal coalescent analyses differ from ours due to how we sampled genes (2). However, differences with these studies are more likely due to their use of Bayesian methods unsuitable for phylogenomic data, or of genomic markers (UCEs) different from ours. Different genomic markers likely carry different signals (3), but our use of exonic data has proven effective in many contexts.

GS claim that the existence of recombination in our data set might mislead coalescent analysis. However, a recent study (4) shows that recombination is not a primary factor influencing the accuracy of coalescent models. It is understandable that systematists might think that recombination would ruin a species tree or even a concatenation analysis, but in fact, because recombination only occurs within species, it is unlikely to be problematic for any but the most extreme cases of incomplete lineage sorting (4).

GS claim that our use of *in silico* concatenated exonic data in our analysis violated a basic assumption of coalescent methods. We did in fact ‘concatenate’ exonic sequence separated in the genome, but in calling that problematic, GS overlook the wealth of phylogenomic studies that do the same thing when using transcriptome data. Our use of concatenated exonic data is actually the same as the use of cDNA sequences from transcriptomes, which has been widely and successfully used in phylogenomic studies (5). It is GS, and the practice of concatenation in general, that “ignore[s] a fundamental tenet of molecular biology”, when, by using concatenation methods, which typically not only combine sequences from the same gene, but also from genomically dispersed genes together, they introduce a process that never occurs in nature. Thus, our “concatenation” in constructing exonic sequences is in fact widely used and is not the same as the concatenation typically practiced in phylogenetics.

Finally, in suggesting that the data of Meredith et al (cited in 1) was insufficient for their analysis, we do not imply that the many systematic studies that have sampled fewer base pairs are invalid. Indeed, we have shown extensively elsewhere that coalescent methods can work well with as few as two loci depending on the degree of hemiplasy. The number of loci or base pairs required scales with the number of taxa sampled; in GS’s work, loci have been undersampled at the expense of more taxa, which can lead to problems in analysis.

Overall, GS ignore the basic difference between concatenation and coalescent methods, namely how they handle gene tree heterogeneity. As shown by our subsampling analysis (2) and many other studies, violation of the assumption of identical gene trees can lead concatenation into inconsistency. When analyzing phylogenomic datasets with many taxa, gene tree heterogeneity will be substantial and coalescent methods will be more consistent.

## **Acknowledgements**

The startup funds of L.L. provided by the University of Georgia, the Collaboration Grant from Tsinghua National Laboratory for Computation, the startup fund of S.W. from Shenyang Normal University, and NSF DEB-0743616 (to S.V.E) supported this research.

## **References**

1. Gatesy J, Springer MS. Reply to Song et al.: Concatenation versus Coalescence versus “Concatalescence”. *Proc Natl Acad Sci USA* (in review).
2. Song S, Liu L, Edwards SV, Wu S (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci. USA* 109:14942-14947
3. Dolgin E (2012) Rewriting evolution. *Nature* 486: 460-462.
4. Lanier HC, Knowles LL (2012) Is recombination a problem for species-tree analyses? *Syst. Biol.* 61: 691-701.
5. Chiari Y, Cahais V, Galtier N, Delsuc F (2012) Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biology* 10: 65.