Reply to Gatesy and Springer: The multispecies coalescent model can effectively handle recombination and gene tree heterogeneity

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

Gatesy and Springer cite two papers to claim that results of recent mammal coalescent analyses differ from ours because of how we sampled genes (2). However, differences with these studies are more likely a result of their use of Bayesian methods unsuitable for phyologenomic data, or of genomic markers (ultraconserved elements) different from ours. Different genomic markers likely carry different signals (3), but our use of exonic data has proven effective in many contexts.

Gatesy and Springer (1) claim that the existence of recombination in our dataset 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).

Gatesy and Springer (1) claim that the use of in silico concatenated exonic data in our analysis violated a basic assumption of coalescent methods. We did in fact "concatenate" exonic sequences separated in the genome, but in calling that problematic, Gatesy and Springer overlooked 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 Gatesy and Springer, and the practice of concatenation in general, that "ignore 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. (6) 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 Gatesy and Springer's work (1), loci have been undersampled at the expense of more taxa, which can lead to problems in analysis.

Overall, Gatesy and Springer (1) ignore the basic difference between concatenation and coalescent methods, namely in how they handled 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.

ACKNOWLEDGMENTS. This work was supported in part by the startup funds of L.L. provided by the University of Georgia; the Collaboration Grant from Tsinghua National Laboratory for Computation; and National Science Foundation Grant DEB-0743616 (to S.V.E).

Shaoyuan Wu[a,1], Sen Song[b], Liang Liu[c], and Scott V. Edwards[d,1]

[a]Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China; [b]Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China; [c]Department of Statistics and Institute of Bioinformatics, University of Georgia, Athens, GA 30606; and [d]Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Cambridge, MA 02138

1 Gatesy J, Springer MS (2013) Concatenation versus coalescence versus "concatalescence". *Proc Natl Acad Sci USA* 110:E1179.
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(37): 14942–14947.
3 Dolgin E (2012) Phylogeny: Rewriting evolution. *Nature* 486(7404):460–462.
4 Lanier HC, Knowles LL (2012) Is recombination a problem for species-tree analyses? *Syst Biol* 61(4):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 Biol* 10:65.
6 Meredith RW, et al. (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* 334(6055):521–524.

Author contributions: S.W. designed research; S.W., S.S., L.L., and S.V.E. performed research; S.W., S.S., and L.L. analyzed data; and S.W. and S.V.E. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: shaoyuanwu@outlook.com or sedwards@fas.harvard.edu.