Claims of homology errors and zombie lineages do not compromise the dating of placentalian diversification

Liang Liu*a,b,c, Jin Zhangd, Frank E. Rheindt’e, Fumin Leif, Yanhua Quf, Yu Wangg,h, Yu Zhangg,h, Corwin Sullivani, Wenhui Nie, Jinhuang Wangj, Fengtang Yangg, Jiping Chenl, Scott V. Edwardsn,m,n,1, Jin Mengo, and Shaoyuan Wup,k,l

Gatesy and Springer (1) consider 3 out of 89 nodes in our “preferred STAR tree” (2) unusual, raising suspicions that underlying alignment errors have generated these and other perceived misestimations in our analysis. As in their other critiques of our work, their claims are based on subjective and unrepeatable logic. We acknowledge that our alignments can be improved; in particular, we neglected to align and trim our loci based on more conserved amino acid alignments. However, our alignments still contain substantial phylogenetic information, and our protocols correctly extracted individual codon positions for analysis. Application of a suite of repeatable best practices for quality control in phylogenomics (3, 4) suggests that, after trimming, about 2.5% of individual sequences—a better measure of contamination than percentage of whole alignments—might be compromised. However, analysis of an improved 60-gene subset of our data yields divergence times that correlate strongly (0.9997) with our reported results (95% CI on slope of old versus new dates: 0.9995–0.9998), rejecting the notion that misalignments invalidate our analysis. As expected (5), our interordinal dates are influenced most strongly by the particular fossils we chose as calibrations, rather than sequence features.

Gatesy and Springer assert our data are plagued by “clear-cut homology errors.” “Clear-cut” is not an objective measure of quality, varies among investigators, and is therefore unscientific. Claims of errors in homology must include quantification and a repeatable protocol, neither of which Gatesy and Springer provide. We agree (6) that the manual data curation that Gatesy and Springer favor (1) has become “unfeasible” (3) and is subjective and neither repeatable nor sustainable, rendering its role in phylogenomics highly questionable (7, 8).

Other analyses, including STAR trees of C3 and CDS partitions, and concatenation trees of C12, C3, and CDS partitions, were consistent with the topological arrangements advocated by Gatesy and Springer (1). Our “preferred” tree was preferred not because it captured clades deemed correct by Gatesy and Springer, but because it possessed stability across analyses and broad congruence with previous work. The three clades in question are nowhere presented as new findings of our study and represent a minor issue in what is recognized by others (9, 10) as our paper’s (2) broader context. The claim that ASTRAL results differ from ours is insupportable: one cannot compare the local posterior probabilities produced by ASTRAL (11), which do not

*Jiangsu Key Laboratory of Phylogenomics and Comparative Genomics, School of Life Sciences, Jiangsu Normal University, Xuzhou 221116, Jiangsu, China; aDepartment of Statistics, University of Georgia, Athens, GA 30606; bInstitute of Bioinformatics, University of Georgia, Athens, GA 30606; cKey Laboratory of High Performance Computing and Stochastic Information Processing of the Ministry of Education of China, Department of Computer Science, College of Mathematics and Computer Science, Hunan Normal University, Changsha, Hunan 410081, China; dDepartment of Biological Sciences, National University of Singapore, Singapore 117543; eKey Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; fState Key Laboratory of Reproductive Biology, Chinese Academy of Sciences, Beijing 1000101, China; gInstitute of Zoology, Chinese Academy of Sciences, Beijing 100044, China; hState Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China; iWelcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom; jGuangdong Public Laboratory of Wild Animal Conservation and Utilization, Guangdong Key Laboratory of Integrated Pest Management in Agriculture, Guangdong Institute of Applied Biological Resources, Guangzhou 510260, China; kDepartment of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138; lDivision of Paleontology, American Museum of Natural History, New York, NY 10024; and mDepartment of Biochemistry and Molecular Biology, 2011 Collaborative Innovation Center of Tianjin for Medical Epigenetics, Tianjin Key Laboratory of Medical Epigenetics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China


The authors declare no conflict of interest.

Published under the PNAS license.

To whom correspondence may be addressed. Email: shaoyuanwu@outlook.com or sedwards@fas.harvard.edu.
take gene tree error into account, to the measures of support we used.

Gatesy and Springer are concerned that “zombie lineages” compromise our conclusions. We acknowledged zombie lineages as a reasonable concern and discussed such discrepancies and their likely causes at some length in our study (2). At the same time, our analysis is an advance because many more fossil and molecular divergences, particularly ordinal divergences, are now better reconciled. Hard bounds on priors can work but are also more likely to mislead (5) than the soft bounds we used. Even sophisticated approaches can misestimate divergences in some cases, while uncertainties in the phylogenetic placement and dating of fossils may often yield false assumptions about fossil ages used for calibration.

Acknowledgments
This work was supported by funds from the Priority Academic Program Development of Jiangsu Higher Education Institutions, the National Natural Science Foundation of China (Grant 31470111) (to S.W.).