Molecular Phylogenetics of Mastodon and *Tyrannosaurus rex*

Chris L. Organ,[1,2] Mary H. Schweitzer,[3,4] Wenxia Zheng,[3] Lisa M. Freimark,[5] Lewis C. Cantley,[5,6] John M. Asara[5,6]*

Protein sequences from bone-derived collagen as old as 68 million years have been detected by using mass spectrometry (*1*, *2*). A BLAST search of the resulting peptide sequences found the highest similarities between the collagen peptides obtained from extracts of *Tyrannosaurus rex* fossil bone and those of birds (*Gallus gallus*) and between mastodon (*Mammut americanum*) and other mammals, including elephant (*Loxodonta africana*). We performed phylogenetic analyses to infer the evolutionary relationships of the *T. rex* [Museum of the Rockies (MOR) 1125] and mastodon (MOR 605). The results extend our knowledge of trait evolution within nonavian dinosaurs into the macromolecular level of biological organization.

We used 21 extant organisms in the analyses. Collagen α1(I) and α2(I) protein sequences from 19 extant organisms (*3*) were obtained from the National Center for Biotechnology Information (NCBI) and ENSEMBL databases. Collagen α1(I) and α2(I) peptide sequences from a metacarpal of *Alligator mississippiensis* were obtained by microcapillary liquid chromatography tandem mass spectrometry (LC/MS/MS) with an ion trap mass spectrometer using the Sequest algorithm (*1*, *2*) and the Paragon algorithm (*4*). Ostrich collagen sequences were determined elsewhere (*1*, *2*). Sequences from alligator (Crocodilia) and ostrich (Aves) represent 63% and 43% of the full-length collagen α1(I) protein available for other organisms, respectively. The collagen α1(I) and α2(I) sequences for elephant (*L. africana*), tenrec (*Echinops telfairi*), and green anole (*Anolis carolinensis*) were obtained by using gene translations from online databases (*5*).

Bayesian, likelihood, parsimony, and distance methods were used to generate evolutionary trees (*6*). In the Bayesian analysis, the posterior distribution of trees reconstructed all extant groups in generally agreed-upon relationships (the posterior probability of clades ranged from 0.80 to 1.00), with the exception of green anole (*A. carolinensis*), which is inferred here to lie at the base of amniotes instead of grouping as the sister taxon to alligator and birds (archosaurs) (Fig. 1). LC/MS/MS from tryptic digests produces fragmentary protein sequence data; however, we found unequivocal support (posterior probability of 1.00) uniting mastodon with elephant as members of Elephantinae, which together group with tenrec (*E. telfairi*) as members of the mammalian group Afrotheria (*7*). Maximum likelihood produces the same groupings, although with less support (approximate likelihood ratio test; aLRT = 0.855 for Elephantinae and 0.872 for Afrotheria). Maximum parsimony analysis also groups mastodon, elephant, and tenrec together (fig. S1, B to D). For the *T. rex* sample, we used five peptide sequences from collagen α1(I) and one from collagen α2(I) (*1*, *2*, *8*) for a total of 89 amino acids (Fig. 1). The *T. rex* clusters within the Archosauria (posterior probability of 0.92), more closely related to birds (chicken and ostrich, 0.9) than alligator, although a lack of informative sites in the ostrich and *T. rex* leaves Dinosauria unresolved. The likelihood tree is identical to the Bayesian tree, except for higher support at these locations in the tree (aLRT = 0.969 for Archosauria and 0.907 for Dinosauria). Branch lengths (expected rates of change per site) indicate a relatively stable and uniform rate of evolution, lacking evidence for a deviation from a molecular clock. Maximum parsimony analysis also groups the *T. rex* with the chicken and ostrich, although bootstrap support is low (fig. S1, B to D). Neighbor joining groups the *T. rex* with the birds, but miscalculates the branching order and misplaces alligator, mastodon, and several extant organisms (fig. S1, B to D).

The slight disagreement between the distance results compared with the Bayesian, likelihood, and parsimony results (all three of which are concordant) are predictable given that distance methods perform poorly for taxa with large amounts of missing data (*9*). Nevertheless, there is congruence between three out of the four methods for the mastodon and four out of four methods for the *T. rex* despite the problem of missing sequence data. The recovered sequences contain informative phylogenetic signal consistent with predictions based on genetic and morphological data for mastodon (*10*, *11*) and on morphological data for *T. rex* (*12*, *13*).

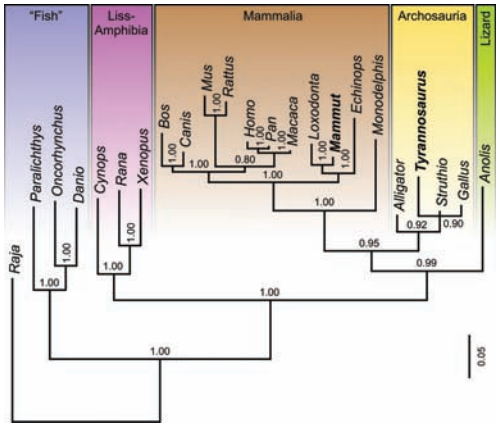

Fig. 1. Inferred evolutionary relationships of major vertebrate groups hypothesized from collagen α1(I) and α2 (I) protein data by using a Bayesian approach. The node (bifurcation) labels are measures of support, which indicate the proportion of trees in the posterior distribution to containing the node. Branch lengths are in expected changes per site.

These results support the endogenous origin of the preserved collagen molecules and confirm the prediction based on morphology that, if biomolecules could be retrieved from a nonavian dinosaur, they would share a higher degree of similarity with birds than with other extant vertebrates. Our findings suggest that molecular data from long-extinct organisms may have the potential for resolving relationships at critical areas of the vertebrate evolutionary tree that have, so far, been intractable. The findings presented here also bolster the use of morphology in phylogenetics because our results are consistent with studies on the evolutionary relationships of fossil forms that rely on morphology (*12*, *13*).

References and Notes

1. J. M. Asara, M. H. Schweitzer, L. M. Freimark, M. Phillips, L. C. Cantley, *Science* **316**, 280 (2007).
2. J. M. Asara *et al.*, *Science* **317**, 1324 (2007).
3. Collagen α1(I) and α2(I) protein sequences were obtained by querying the nonredundant all-taxa protein database from the NCBI and through genome alignments at ENSEMBL (www.ensembl.org).
4. I. V. Shilov *et al.*, *Mol. Cell. Proteomics* **6**, 1638 (2007).
5. The peptide sequence translations for were obtained from ENSEMBL. See (*6*) for accession numbers.
6. Materials and methods are available on *Science* Online.
7. M. S. Springer, W. J. Murphy, E. Eizirik, S. J. O'Brien, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1056 (2003).
8. J. M. Asara, M. H. Schweitzer, *Science* **319**, 33 (2008).
9. J. J. Wiens, *Syst. Biol.* **52**, 528 (2003).
10. N. Rohland *et al.*, *PLoS Biol.* **5**, 1663 (2007).
11. M. G. Thomas *et al.*, *Proc. R. Soc. London Ser. B* **267**, 2493 (2000).
12. M. J. Benton, in *The Dinosauria*, D. B. Weishampel, P. Dodson, H. Osmólska, Eds. (Univ. of California Press, Berkeley, 2004), pp. 7–19.
13. T. R. Holtz, H. Osmólska, in *The Dinosauria*, D. B. Weishampel, P. Dodson, H. Osmólska, Eds. (Univ. of California Press, Berkeley, 2004), pp. 21–24.
14. M. S. Springer *et al.*, *Syst. Biol.* **56**, 673 (2007).

15. We thank S.V. Edwards, C. Marshall, C. Balakrishnan, M. Baldwin, P. Brito, N. Hobbs, M. Phillips, S. Nagpal, S. Seymour, G. Poulogiannis, B. Zheng, M. Vander Heiden, J. Horner, N. Myhrvold, and C. Hill for assistance. This research was supported by an NIH National Service Research Award Postdoctoral Fellowship granted to C.L.O., the NSF (awards EAR-0634136 and EAR-0541744), the Paul F. Glenn Foundation, and the David and Lucile Packard Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5875/499/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
References and Notes
Appendices 1 and 2

17 December 2007; accepted 25 February 2008
10.1126/science.1154284

[1]Harvard University, Cambridge, MA 02138, USA. [2]Museum of Comparative Zoology, Cambridge, MA 02138, USA. [3]North Carolina State University, Raleigh, NC 27695, USA. [4]North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA. [5]Beth Israel Deaconess Medical Center, Boston, MA 02115, USA. [6]Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: jasara@bidmc.harvard.edu