Detecting Dinosaur DNA

The fact that DNA sequence can be obtained from fossil organisms has opened new windows of opportunity for research in organismal and molecular evolution (*1*). Among these is the possibility of obtaining genetic information from major groups of organisms now extinct. Recently, S. R. Woodward *et al.* sequenced DNA from a portion of the mitochondrial cytochrome *b* gene from Cretaceous bone fragments apparently from a dinosaur that lived 80 million years ago (*2*). However, the likely source of those DNA sequences appears to be human contamination.

In addition to experimental controls, a major line of evidence normally used to support a finding concerning ancient DNA is the phylogenetic relationship of the putative ancient sequence to those from the closest living relatives of the fossil organism (*1*). In the case of a possible dinosaur sequence, there is strong evidence from morphology that birds represent the closest living organisms to dinosaurs, and morphological and molecular evidence indicate that crocodilians are the closest living relatives of birds (*3*–*4*). Also, the fossil record indicates that, after splitting with mammals, at least 100 million years of evolution occurred on the lineage leading to dinosaurs and birds before the latter groups diverged (*3*). Therefore, a putative dinosaur sequence would be expected to cluster with birds and crocodilians in a phylogenetic analysis of amniotes.

Woodward *et al.* (*2*) do not present an evolutionary tree, but discuss their sequences in terms of percent sequence difference, noting that these cytochrome *b* sequences differed from all others in the databases. We also performed a BLAST search using the majority rule consensus sequence [figure 6 in (*2*)] and obtained matches to 130 cytochrome *b* sequences of vertebrates (*5*). As reported by Woodward *et al.* (*2*), the consensus sequence differs by about 30% (26% to 52%) from those vertebrate sequences in the databases. However, 87 of the most similar sequences (closest matches) are mammals, including all nine eutherian orders represented, whereas birds, amphibians, and fish comprise nearly all of the remaining sequences and have the lowest similarity to the consensus sequence. Among the mammal sequences, the closest matches are to whales (99/133 = 74% similarity). However, among the nucleotide sites showing similarity to the human sequence (93/133 = 69%), four are rare variants in the other 129 vertebrate sequences (*6*).

A phylogenetic analysis (*7*) with all tetrapod sequences obtained from the BLAST search joins the putative dinosaur DNA sequence (*2*) with human (Fig. 1). Although statistical support for most nodes in the tree is low as a result of the short length of this region (133 base pairs), bootstrap support for this cluster (91%) is relatively high. Furthermore, a consensus sequence of the nine bone sequences which maximizes similarity to human (118/133 = 88% similarity) clusters with the human sequence at a statistically significant bootstrap *P* value of 100%. Consensus sequences with similarity maximized to each of the other taxa yield considerably lower (0 to 46%) probabilities for clustering with the taxon to which similarity was maximized (*8*).

Despite meticulous care, contamination of polymerase chain reaction (PCR) experiments with foreign DNA, often of human origin, is an ever-present aspect of ancient DNA research because of the sensitivity of the methodology and rarity of the target molecules (*1*). The suggestion by Woodward *et al.* (*2*) that variation among the nine sequences (seven from the same bone fragment) is a result of damaged template may be correct. However, our results suggest that the DNA template was not from a Cretaceous organism such as a dinosaur, but rather from an extant organism, most likely a human.

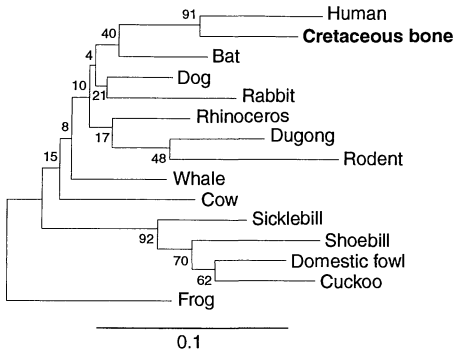

Fig. 1. Phylogenetic tree of partial cytochrome *b* DNA sequences in representatives of extant tetrapod groups and putative dinosaur DNA sequence (majority rule consensus) derived from Cretaceous bone fragments (*2*). Numbers on nodes are bootstrap confidence probabilities. Inclusion of all nine putative dinosaur sequences (*2*) resulted in an identical tree in which those sequences clustered together with human. A frog was included to root the tree. Tree shown is neighbor-joining with transversion distance; parsimony analyses (transversions only and weighted transversions) also clustered the putative dinosaur DNA sequence with the human sequence; 133 sites total, 88 variable, and 66 parsimony. GenBank accession numbers: human (V00662), bat (L28943), rhinoceros (X56283), dugong (U07564), cow (J01394), dog (L29416), rabbit (U07566), whale (X75581), rodent (L11902), sicklebill (X74253), domestic fowl (X52392), cuckoo (U09262), shoebill (U08937), and frog (U02890).

Determining the authenticity of an ancient DNA sequence often can be diffcult, and criteria for this have been discussed elsewhere (*1*, *9*). Two criteria that are important, and that were not fulfilled in the study by Woodward *et al.*, are phylogenetic context and independent replication. Although phylogenetic support has been presented for other findings of DNA surviving for millions of years (*10*), real advance in this field will come only when it is demonstrated that those studies can be replicated in independent laboratories.

S. Blair Hedges
Department of Biology and
Institute of Molecular Evolutionary Genetics,
208 Mueller Laboratory,
Pennsylvania State University,
University Park, PA 16802, USA
Mary H. Schweitzer
Paleontology Department,
Museum of the Rockies,
Bozeman, MT 59717, USA,
and *Department of Biology,*
Montana State University,
Bozeman, MT 59717

REFERENCES AND NOTES

1. S. Pääbo, R. G. Higuchi, A. C. Wilson, *J. Biol. Chem.* **264**, 9709 (1989); S. Pääbo, *Sci. Am.* **269**, 86 (November, 1993).
2. S. R. Woodward, N. J. Weyand, M. Bunnell, *Science* **265**, 1229 (1994).
3. M. J. Benton, *J. Mol. Evol.* **30**, 409 (1990).
4. S. B. Hedges, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2621 (1994).
5. BLAST version 1.4.7MP [S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403, (1990)]; some sequences were of multiple individuals of the same species.
6. Sites 15646, 15687, 15703, and 15706; S. Anderson *et al.*, *Nature* **290**, 457 (1981).
7. The DNA sequences were analyzed with MEGA version 1.01 [S. Kumar, K. Tamura, M. Nei, *MEGA: Molecular Evolutionary Genetics Analysis*, (Pennsylvania State University, University Park, PA, 1993)] for distance analyses and PAUP [D. L. Swofford, *Phylogenetic Analysis Using Parsimony*, Version 3.1 (University of Illinois, Champaign, IL (1993)] for parsimony analyses. Average pairwise Jukes-Cantor (T. H. Jukes and C. R. Cantor, in *Mammalian Protein Metabolism*, H. N. Munroe, Ed. (Academic Press, New York, 1969, pp. 21–132) corrected distances were large (0.3 to 0.5), and therefore a transversion distance (M. Kimura, *J. Mol. Evol.* **16**, 111, 1980) was used with neighbor-joining [N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987)]; and transversion only, or transversions weighted 10 times transitions, were used with parsimony. The marsupial sequence was excluded from the phylogenetic analyses because of anomalous results. Statistical significance (>95%) was assessed with the bootstrap method [J. Felsenstein, *Evolution* **39**, 783 (1985)], with 2000 replications.
8. Bootstrap probabilities for clustering with the taxon-specific maximized consensus in neighbor-joining analyses (transversion distance) are as follows: human (100%), bat (17%), whale (8%), rabbit (46%), dog (6%), rhinoceros (31%), cow (22%), dugong (9%), rodent (26%), sicklebill (1%), domestic fowl (21%), cuckoo (0%), shoebill (1%), and frog (6%).
9. T. Lindhahl, *Nature* **365**, 700 (1993); O. Handt *et al.*, *Experentia* **50**, 524 (1994).
10. E. M. Golenberg *et al.*, *Nature* **344**, 656 (1990); P.

S. Soltis, D. E. Soltis, C. J. Smiley, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 449 (1992); R. J. Cano, H. N. Poinar, D. W. Roubik, G. O. Poinar, *Med. Sci. Res.* **20**, 619 (1992); R. DeSalle, J. Gatesy, W. Wheeler, D. Grimaldi, *Science* **257**, 1860 (1992); R. J. Cano, H. N. Poinar, N. J. Pieniazek, A. Acra, G. O. Poinar, *Nature* **363**, 536 (1993); H. N. Poinar, R. J. Cano, G. O. Poinar, *ibid.*, p. 677.

11. We thank C. A. Hass, S. Kumar, and S. Pääbo for helpful comments.

6 December 1994; accepted 7 April 1995

The comparisons reported by Woodward *et al.* (*1*) were limited to identity percentages, whereas more informative comparisons should be possible by scoring each aligned amino acid pair with the use of a log-odds substitution matrix based on homologous protein alignments (*2*). I compiled a database of all 223 cytochrome *b* segments from different species in the combined protein databanks (through 11/94). Each segment was scored for similarity to a consensus representing the seven long bone sequences, with the use of the most frequent predicted amino acid at each position. A range of BLOSUM (*3*) and PAM (*4*) substitution matrices was used for scoring. In addition, each segment was scored using position-specific scoring matrices (*5*) constructed from the seven long bone sequences and from the two rib bone sequences.

All tested scoring systems provided similar results (data not shown). Among the well-represented taxa, the highest mean scores were found for cetaceans and ungulates. In both cases the mean scores are significantly higher than the mean scores for birds. It is notable that all 15 alignments with cetacean segments outscored all 72 alignments with bird segments, even though both groups are diversely represented (*6*). Overall, scores for vertebrates were much higher than for arthropods, which in turn were much higher than for non-animals (plants, fungi, and bacteria), indicating that this method applied to bone sequences provides rankings consistent with known phylogenetic relationships. Moreover, similar results were found for rib sequences analyzed independently of long bone sequences, despite several nucleotide sequence differences (*1*).

I conclude that the bone sequences more closely resemble homologs in mammals than in birds, which are thought to be the closest living relatives to dinosaurs. Furthermore, the significantly higher scores for some mammals (cetaceans and ungulates) than for others (*7*) further suggest either a mammalian origin or convergence of this region of cytochrome *b*. The analysis also contradicts criticisms that the bone sequences resulted from microbial contamination or were seriously affected by PCR-generated errors. Therefore, further PCR-based analysis of the Utah bones is warranted. For such studies, most efficient synthesis should be possible with primers modeled on the mammalian taxa with high alignment scores. Reducing the high PCR failure rate (*1*) in this way should greatly increase the amount of sequence available for phylogenetic analysis.

Steven Henikoff
Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98104, USA

REFERENCES

1. S. R. Woodward, N. J. Weyand, M. Bunnell, *Science* **266**, 1229 (1994).
2. S. F. Altschul, *J. Mol. Biol.* **219**, 555 (1991).
3. S. Henikoff and J. G. Henikoff, *Proc. Natl. Acad. Sci. USA* **89**, 10915 (1992).
4. D. T. Jones, W. R. Taylor, J. M. Thornton, *CABIOS* **8**, 275 (1992).
5. S. Henikoff and J. G. Henikoff, *J. Mol. Biol.* **243**, 574 (1994).
6. U. Arnason and A. Gullberg, *Nature* **367**, 726 (1994).
7. S. Henikoff, unpublished data.

9 December 1994; accepted 7 April 1995

Assuming that each of the published sequences (*1*) are representative of the study by Woodward *et al.*, we chose two for extensive analyses to assess the history of these molecules (3-37 from bone fragment one and 5-37 from bone fragment two). When alignments were determined by comparison against all of the sequences in a current issue of Entrez (NCBI, release 6.0 of GenBank) with the use of the MacVector program (version 4.1.4, Eastman Kodak, Rochester, New York), the best 30 alignments against fragment one were all mammalian cytochrome *b* sequences, with the first nine chosen from the order Artiodactyla (cattle, deer, antelopes, and their relatives). A similar result was obtained for alignments against fragment two, with the best four alignments each to human cytochrome *b* genes. Other vertebrates are not equally divergent from these purported dinosaur sequences. To the contrary, these unknown sequences have closest similarity to the mule deer (*Odocoileus hemionus*, accession number X56291) and to human cytochrome b genes (*Homo sapiens*, accession number V00662), respectively.

The best strategy for determining relatedness of an unknown sequence is not through a similarity search, but rather by a phylogenetic analysis using parsimony (*2*). While we agree with Woodward *et al.* (*1*) that their small fragment of cytochrome *b* sequence is inappropriate for use in a phylogenetic analysis, it is the only available evidence, and parsimony is still the best strategy for determining the closest relative and for identifying these new sequences. We aligned the unknown cytochrome *b* sequences to several mammals (human, cow, rat, and mouse), to chicken, and to clawed frog (*Xenopus*, our outgroup). The most parsimonious solution was one that grouped Cretaceous bone fragment two with the human and next with the other unknown fragment. This resulted when the characters for each codon were numbered and third positions were omitted and when we looked at the more conserved transversions. When amino acids were translated from the original nucleotide sequences and parsimony analysis was conducted, the unknown fragments were closest, then the chicken (supported by two characters). This pattern also resulted when all characters were examined.

Our most conservative and informative analyses point to mammals as the closest relatives to the available "Cretaceous" sequence, an unlikely relation if these are truly dinosaur remains. This contradicts with numerous morphological characters that support birds as the closest living relative to the dinosaurs (*3*). One might ask, how did mammalian DNA get into these samples? At the time that these coal beds were formed, all of the known mammals were smaller than the bone fragments described (*4*). Possibly, either ancient DNA of a smaller mammal was preserved along with these deposits and thus contaminated these bone fragments, or a more recent DNA sample contaminated these tissue samples. Fossil mammals are known from this geological formation, potentially supporting the former hypothesis. We prefer the latter hypothesis because of the great similarity of these sequences to living mammalian genes.

Marc W. Allard
Deshea Young
Yentram Huyen
Department of Biological Sciences, George Washington University, Washington, DC 20052, USA

REFERENCES

1. S. R. Woodward, N. J. Weyland, M. Bunnell, *Science* **266**, 1229 (1994).
2. R. DeSalle, J. Gatesy, W. Wheeler, D. Grimaldi, *ibid.* **257**, 1933 (1992).
3. J. A. Gauthier, *Mem. Calif. Acad. Sci.* **8**, 1 (1986).
4. J. A. Lillegraven and M. C. McKenna, *Amer. Mus. Novitates* **2840** (1986).

19 December 1994; accepted 7 April 1995

Our preliminary phylogenetic analysis of the putative dinosaur sequences in the report by Woodward *et al.* (*1*) showed them to be weakly related to the human cytochrome *b* gene, albeit quite distantly (earlier comment by Hedges *et al.*, data not shown). As nuclear insertions of mitochondrial DNA are known to occur (*2*), and as 12*S* ribosomal DNA sequences amplified from ancient monkey bones have been attributed to insertions of mitochondrial DNA into the human nuclear genome (*3*), the putative dinosaur cytochrome *b* sequences might represent ancient integrations of mitochondrial DNA into the human nuclear genome.