Letter to the Editor

Reanalysis of Published DNA Sequence Amplified from Cretaceous Dinosaur Egg Fossil
Hai-Lin Wang, Zi-Ying Yan, and Dong-Yan Jin

National Key Laboratory of Molecular Virology and Genetic Engineering, Institute of Virology, Chinese Academy of Preventive Medicine, Beijing, People’s Republic of China

The PCR amplification of DNA sequences from ancient specimens has paved the way for research in the molecular evolution of extinct species. However, PCR is extremely sensitive and sources of ancient DNA are extremely rare. Therefore, extreme caution should be observed to eliminate the possibility of contamination. It is generally accepted that the demonstration of the uniqueness of a sequence amplified from an ancient specimen cannot be taken as an indication of the ancient origin of that sequence. Instead, several lines of evidence should be offered to support its validity (Handt et al. 1994). In this regard, a possible dinosaur DNA find from Cretaceous bone fragments by Woodward, Weyand, and Bunnell (1994) was greeted with skepticism (Gibbons 1994) and was later shown by at least five independent laboratories to be derived almost certainly from human contamination (Hedges et al. 1995; Collura and Stewart 1995).

Among the hunters for the first dinosaur DNA (Morell 1993), a group of young Chinese scientists from Peking University claimed the cloning and sequencing of dinosaur DNAs (six pieces of 18S rDNAs and another piece of 191-bp DNA) from the flocculent inclusion of a Cretaceous dinosaur egg fossil found in Xixia Basin, Henan, China. Their findings had been the hot news of the public press, but were published in a very unusual Chinese journal (An et al. 1995; Li et al. 1995) in the absence of strong evidence supporting the authenticity of the ancient origin of the “dinosaur” sequence. The sequence comparisons were discussed in terms of percent homology. The similarity alignments were limited to arbitrarily selected sequences. Therefore, a more conservative and informative reanalysis of these published sequences is required to determine whether they are authentic copies of dinosaur DNA.

We performed an independent analysis of the published “dinosaur” DNA sequence by similarity searching and by phylogenetic clustering using distance matrix or parsimony. Our results showed that the two representative 18S rDNAs (other pieces are highly homologous to these two) cloned from the dinosaur egg fossil (DA18S1, GenBank accession number U41317; DA18S7, U41318) share striking homology of more than 85% with the 18S rDNAs from fungi and from flowering plants, respectively. Among the best matches to clone DA18S1 are rDNAs from fungal symbionts (e.g., GenBank U09535 and U09536) of some leaf-cutting ants (Hinkle et al. 1994) and of some lichen-forming algae (Gargas et al. 1995), as well as other free-living fungi (e.g., M59760). Sequences most homologous to DA18S7 include rDNAs from thale cress (Arabidopsis thaliana; T76203), eggplant (Solanum melongena; X63311), and potato (Solanum tuberosum; X67238). Notably, the sequence homology between DA18S7 and rDNAs from some of these dicotyledons is exceedingly high. For example, there are only six nucleotide substitutions between DA18S7 and the 18S rDNA from Akebia quinata (L31’95; angiosperm-1 in fig. 1).

A similarity alignment of 18S rDNAs from different species (fig. 1) was created using the PILEUP program in the Wisconsin software package (Version 8.1, Genetics Computer Group, Inc., Madison, Wis.). Since this progressive pairwise alignment is supported by the known structure of 18S rRNA, its reliability level is high. As shown in the alignment, the nucleotide difference either between DA18S7 and the two angiosperms or between DA18S1 and the two fungi is very limited. On the other hand, both DA18S1 and DA18S7 are highly divergent from duck, human, alligator, and other animal rDNAs. Thus, it is very clear that fungi and higher plants, rather than amphibians, birds, and human as described by An et al. (1995), are the closest relatives to the species from which DA18S1 and DA18S7 were derived. In particular, DA18S7 is highly likely derived from an unknown angiosperm, taking into account the aforementioned high level of sequence homology. Since evidence from morphology indicates that birds and crocodiles are the closest living relatives of dinosaurs, it is very unlikely that DA18S1 and DA18S7 would be authentic copies of dinosaur DNA.

As valuable as it is, similarity searching is not the only strategy to determine the relatedness of DNA sequence. It is widely accepted that phylogenetic analysis is more informative and should be used whenever sufficient sequence data are obtained. While we agree that the rDNA sequence of about 150 bp is too short and would be inappropriate for phylogenetic clustering, it is the only available information, and a phylogenetic tree is still helpful for finding closest relatives. In this regard, phylogenetic evidence has already proved valuable in the identification of short and unknown “ancient” sequences (Hedges et al. 1995). Here we generated a distance matrix tree (fig. 2, a similar tree based on parsimony is not shown) with the help of the DNADIST and DRAWGRAM programs in the PHYLIP package (Felsenstein 1996). In this tree, DA18S7 clusters with the two angiosperm sequences (angio-1 and angio-2), whereas DA18S1 groups with the fungal rDNAs. Neither of them joins with the duck, alligator, or human sequences. Although the short length of comparable se-
quency may render the genealogical relationships statistically uncertain, the evidence is overwhelming that DA18S1 and DA18S7 are much more closely related to their plant and fungal homologs than to those from bird, alligator, and human. We also note that this tree is generally consistent with the consensus tree of life. This lends further support to its reliability. Taken together, our results from both the similarity alignment and the phylogenetic analysis clearly show that the Peking University scientists accidentally amplified fungal and plant rDNAs in their PCR experiments attempting to get dinosaur DNA.
PCR and similarity search should not be used beyond their real effectiveness.

Acknowledgments

We gratefully acknowledge Drs. Chen-Lu Tsou and Yun-De Hou for critical readings of the manuscript and for helpful discussions. This work was funded in part by the National Key Laboratory of Molecular Virology and Genetic Engineering.

LITERATURE CITED


NAOYUKI TAKAHATA, reviewing editor

Accepted January 24, 1997