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 (*Solunum 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* (L31795; 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 abovementioned 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 inappropiate 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-

Key words: ancient DNA, PCR, 18S rDNA, dinosaur, similarity searching, phylogenetic analysis.

Address for correspondence and reprints: Dong-Yan Jin, Institute of Virology, 100 Yingxin Street, Beijing 100052, China. E-mail: yanzy@ccs.capm.ac.cn.

Mol. Biol. Evol. 14(5):589–591. 1997

```
                1                                                      50
angiosperm-1    AACCGTAGTA ATTCTAGAGC TAATACGTGC AA.CAAACCC CGACTTCT..
angiosperm-2    AACCGTAGTA ATTCTAGAGC TAATACGTGC AA.CAATCCC CGACTTCT..
      da18s7    .......... ...CTAGAGC TAATACGTGC AA.CAAACCC CGACTTCT..
    fungus-1    AACTGTGGTA ATTCTAGAGC TAATACATGC AATCAAGCCC CGACTTCT..
    fungus-2    AACTGTGGTA ATTCTAGAGC TAATACATGC AATCAAGCCC TGACTTCT..
      da18s1    .......... ...CTAGAGC TAATACATGC ATTCAAGCCC CGACTTCT..
     ciliate    AACCGTGGTA ATTCTAGAGC TAATACATGC TGTTAAGCCC AACTTTGA..
        frog    AACTGTGGTA ATTCTAGAGC TAATACATGC CGAC.GAGCG CTGACCCCC.
       shark    AACTGTGGTA ATTCTAGAGC TAATACATGC CAAC.GAGCG CTGACCCTC.
   alligator    AACTGTGGTN NTTCTANAGC TAATACATGC CGAC.GAGCG CTGACCTNC.
       human    AACTGTGGTA ATTCTAGAGC TAATACATGC CGAC.GGGCG CTGACCCCCT
        duck    AACTGTGGTA ATTCTAGAGC TAATACATGC CGAC.GAGCG CCGACCTCC.
      weevil    AACTGTGGTA ATTCTAGAGC TAATACATGC AAACAGAGCT CCGACC....
                51                                                    100
angiosperm-1    ........GG AAGGGATGCA TTTATTAGAT AAAAGGTCGA CACAGGCTAT
angiosperm-2    ........GG AAGGGATGCA TTTATTAGAT AAAAGGTCGA CACAGGCTTT
      da18s7    ........GG AAGGGATGCA TTTATTAGAT TAAAAGGCGA .ACNGGGTCT
    fungus-1    ........GG AAGGGGTGTA TTTATTAGAT AAAAAACCAA CGCGG...CT
    fungus-2    ........GG AAGGGGTGTA TTTATTAGAT AAAAAACCAA CGCGG...CT
      da18s1    ........GG AAGGGGTGTA TTTATTAGAT TAAAATCAA. ..........
     ciliate    ........GG AAGGGCTGTA TTTATTAGAT AACAAATCAA TATTC...CT
        frog    .......AGG GATGCGTGCA TTTATCAGAC CAAAACCAAT CCGGGGCCCC
       shark    .......CGG GATGCGTGCA TTTATCAGAC CAAAACCAAT CGGG......
   alligator    .......GGG GATNCGTGNA TTTATCAGAC CAAAACCAAC GGG.......
       human    TCGCGGGGGG GATGCGTGCA TTTATCAGAT CAAAACCAAC CCGGTCAGCC
        duck    .......GGG GACGCGTGCA TTTATCAGAC CAAAACCAAC CCGGGCTCGC
      weevil    ..GGAGACGG AAGGAGTGCT TTTATTAGAT CAAAACCAAT CGGTGGCGGG
                101                                                   150
angiosperm-1    GC........ .......... .......... CCGTTGCTCT GATGATTCAT
angiosperm-2    GC........ .......... .......... CCGTTGCTCT GATGATTCAT
      da18s7    GC........ .......... .......... CCGTTGCTCT .ATGATTCAT
    fungus-1    CG........ .......... .......... CCGCTCTTTT GGTGATTCAT
    fungus-2    CG........ .......... .......... CCGCTCCATT GGTGAATCAT
      da18s1    .......... .......... .......... ...CTTTGTT GGTGAATCAT
     ciliate    TG........ .......... .......... TGTCTATTGT GATGATTCAT
        frog    CGCGCCC... .......... .......... CGGCCGCTTT GGTGACTCTA
       shark    .CTTGCC... .......... .......... CGGCAGCTTT GGTGACTCTA
   alligator    .CTCGCC... .......... .......... CGGCCGCTNT GGTGACTCTA
       human    CCTCTCCGGC CCCGGCCGGG GGGCGGGCCG CGGCGGCTTT GGTGACTCTA
        duck    C......... .......... .......... CGGCGGCTTT GGTGACTCTA
      weevil    TTTACTC... .......... .....GTCAT CGTACAACTT GGTGACTCTG
                151                                                   200
angiosperm-1    GATAACTCGA CG..GATCGC ACGGCCTTAG TGCTGGCGAC GCATCATTCA
angiosperm-2    GATAACTCGA CG..GATCGC ATGGCCCTCG TGCTGGCGAC GCATCATTCA
      da18s7    GATAAC.CGA CG..GATCGC ATGGCC.TAG TGCTGGCGAC GCATCATT..
    fungus-1    AATAACTTGT CG..AATCGC ATGGCC.TTG TGCCGGCGAT GCTTCATTCA
    fungus-2    AATAACTTCT CG..AATCGC ATGGCC.TTG TGCCGGCGAT GCTTCATTCA
      da18s1    AATAACTTCT CG..GACCGC ATGGCC.TCG TGCTGGCGGT GCTTCATT..
     ciliate    AATAACTGAT CG..AATCGC ATGGGC.TT. TGCTCGCGAT ACATCATTCA
        frog    GATAACCTCG GGCCGATCGC ACG.TCCCCG TGACGGCGAC GATACATTCG
       shark    GATAACCTCG GGCTGATCGC ACG.TCCTCG TGACGGCGAC GACTCATTCG
   alligator    GATAACCNCG GGCCGATCGC AGC.NCCNCG TGGCGGCGAC GACGCATTCG
       human    GATAACCTCG GGCCGATCGC ACGCCCCCCG TGGCGGCGAC GACCCATTCG
        duck    GATAACCTCG AGCCGATCGC ACG.CCCCCG CGGCGGCGAC GACCCATTCG
      weevil    AATAACTTTA CGCTGATCGC ACGGT.CTCG CACCGGCGAC CCATCTTTCA
```

FIG. 1.—Sequence alignment of 18S rDNAs. Dots indicate gaps. Shown below are species names and GenBank accession numbers of the sequences being compared: angiosperm-1, *Akebia quinata* Houtt. Decne., L31795; angiosperm-2, *Champereia manillana* Merr., L24746; da18s7, unidentified angiosperm, U41318; fungus-1, *Basidiomycete* symbiont of *Apterostigma collare* (attine ant), U09535; fungus-2, *Spongipellis unicolor,* M59760; da18s1, unidentified fungus, U41317; ciliate, *Stylonychia pustulata* (hypotrichous ciliate), X03947; frog, *Xenopus laevis* (African clawed frog), X04025; shark, *Notorynchus cepedianus* (cow shark), M91183; alligator, *Alligator mississipiensis* (American alligator) M59383; human, *Homo sapiens,* M10098; duck, *Anas platyrhynchos,* D38362; weevil, *Lebanorhinus succinus,* L08072.

quence 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.

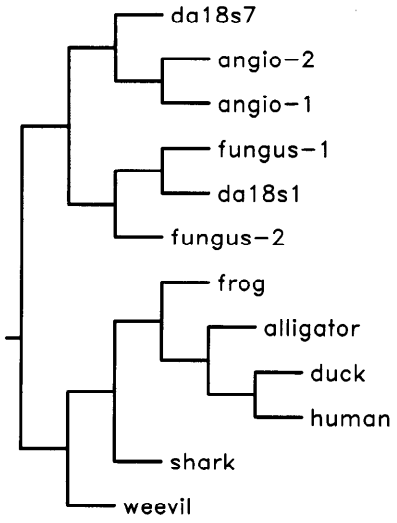

FIG. 2.—Phylogenetic tree of partial 18S rDNAs from different species. The tree is based on a matrix of pairwise evolutionary distances expressed as substitutions per 100 bases and corrected by the Jin-Nei gamma distance method (Jin and Nei 1990) for multiple substitutions at a site. The tree reconstruction was performed using a neighbor-joining algorithm. A ciliate rDNA sequence was used for rooting. See legend to figure 1 for species names and GenBank accession numbers of the sequences in the tree; angio-1 = angiosperm-1; angio-2 = angiosperm-2.

As for the other piece of 191-bp DNA (GenBank U41319), amplified from the same fossil and previously claimed to encode a 37-amino-acid peptide closely related to cadherin, we found that neither the nucleotide nor the deduced amino acid sequence was significantly homologous to any sequence in the current databases and that the sequence information would add little to support its ancient origin. Obviously, there is more than one possible explanation for this anomalous fragment. Among them, we do not want to rule out that it could be an authentic copy of a broken piece of dinosaur gene or any other ancient DNA hidden in the fossil. Neither can we eliminate that it was derived from an unknown source contaminated in the template. But in consideration of the fact that random primers were added to the PCR reaction (Li et al. 1995), we prefer to base our explanation on a more plausible but less stimulating hypothesis that the fortuitously cloned fragment was simply a PCR artifact, unless it can be recloned from the same source using specific primers.

In conclusion, our findings strongly suggest that the scientific evidence for dinosaur egg DNA is still lacking. Again, as versatile and helpful as they are, the tools of 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

AN, C.-C., Y. LI, Y.-X. ZHU et al. (13 co-authors). 1995. Molecular cloning and sequencing of the 18S rDNA from specialized dinosaur egg fossil found in Xixia Henan, China. Acta Sci. Nat. Univ. Pekinensis **31**:140–147.

COLLURA, R. V., and C.-B. STEWART. 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature **378**:485–489.

FELSENSTEIN, J. 1996. Inferring phylogenies from protein sequences by parsimony, distance and likelihood methods. Methods Enzymol. **266**:418–427.

GARGAS, A., P. T. DEPRIEST, M. GRUBE, and A. TEHLER. 1995. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. Science **268**:1492–1695.

GIBBONS, A. 1994. Possible dino DNA find is greeted with skepticism. Science **266**:1159.

HANDT, O., M. HOSS, M. KRINGS, and S. PAABO. 1994. Ancient DNA: methodological challenges. Experientia **50**: 524–529.

HEDGES, S. B., M. H. SCHWEITZER, S. HENIKOFF et al. (14 co-authors). 1995. Detecting dinosaur DNA. Science **268**: 1191–1194.

HINKLE, G., J. K. WETTERER, T. R. SCHULTZ, and M. L. SOGIN. 1994. Phylogeny of the attine ant fungi based on analysis of small subunit ribosomal RNA gene sequences. Science **266**:1695–1697.

JIN, L., and M. NEI. 1990. Limitations of evolutionary parsimony method of phylogenetic analysis. Mol. Biol. Evol. **7**: 82–102.

LI, Y., C.-C. AN, Y.-X. ZHU et al. (12 co-authors). 1995. DNA isolation and sequence analysis of dinosaur DNA from Cretaceous dinosaur egg in Xixia Henan, China. Acta Sci. Nat. Univ. Pekinensis **31**:148–152.

MORELL, V. 1993. Dino DNA: the hunt and the hype. Science **261**:160–162.

WOODWARD, S. R., N. J. WEYAND, and M. BUNNELL. 1994. DNA sequence from Cretaceous period bone fragments. Science **266**:1229–1232.

NAOYUKI TAKAHATA, reviewing editor

Accepted January 24, 1997