Substitution Bias, Weighted Parsimony, and Amniote Phylogeny as Inferred from 18S rRNA Sequences

Charles R. Marshall
Department of Earth and Space Sciences, University of California

Recently published rRNA sequences (Hedges et al. 1990) provide important new information on tetrapod relationships. Specifically Hedges et al.'s (1990) data support a monophyletic Lissamphibia, Amniota, Squamata, Mammalia, and Aves. It is surprising that the 18S rRNA sequences also support a close relationship between the two birds and the four mammals sequenced. Paleontological data provide considerable evidence for a close bird-crocodile relationship (e.g., see Gauthier et al. 1988; Donoghue et al. 1989), and the strength of these data suggested the possibility that the support for a bird-mammal relationship shown by the 18S rRNA sequences may be an artifact created by substitution biases.

Table 1 shows that there are indeed pronounced substitution biases in the amniote 18S rRNA sequences. In particular, there is both a significant overrepresentation of T-to-C substitutions and a significant underrepresentation of A-to-T and T-to-A substitutions. It is interesting that a disproportionately large number of sites that support a bird-mammal relationship are T-to-C substitutions (table 1). The only site that involves T-to-A and/or A-to-T substitutions is site 1317 in Hedges et al.'s (1990) figure A1 (note that there are 140 nucleotides/line in their fig. A1, not 150 as indicated), and on the most parsimonious tree this site is homoplasious, requiring both a T-to-A substitution in the branch leading to the four mammals and either a T-to-A substitution in the Coelacanth or an A-to-T substitution at the base of the amniotes. However, when viewed in isolation, site 1317 supports a clade including birds and crocodiles, to the exclusion of mammals.

Note that the substitution biases are quite complex. The T/C transitions are vastly overrepresented, yet A/G transitions are not. Similarly, A/T transversions are underrepresented, but C/G transversions are not (table 1).

Dynamically Weighted Parsimony

Williams and Fitch's (1990) dynamically weighted parsimony method provides a way of determining whether the substitution biases in the 18S rRNA sequences might be responsible for the unexpected 18S rRNA amniote phylogeny. Their algorithm weights sites according to the relative number of substitutions each incurs on the shortest tree, and it weights substitution types according to number of times they appear on the shortest tree; a site involving relatively few substitutions and relatively rare substitution types will be assigned a higher weight than will a site requiring many substitutions involving frequently seen substitution types. Of course, both the number of substitutions that a site incurs and the frequency of each substitution type depend in part on the topology of the tree being examined. The algorithm assigns weights on the basis of a predefined topology and then uses these weights to search for a new most parsimonious tree. If a new topology is found, then the weights are recalculated, and a new shortest tree is searched for, etc, i.e., the weights are assigned dynamically.

1. Key words: 18S rRNA, amniote phylogeny, birds, crocodiles, mammals, weighted parsimony, substitution bias.

Address for correspondence and reprints: Charles R. Marshall, Department of Earth and Space Sciences, University of California, Los Angeles, Los Angeles, California 90024-1567.

© 1992 by The University of Chicago. All rights reserved.
0737-4038/92/0902-0014+$02.00

370
As suspected, when the amniote 18S rRNA sequences were analyzed with Williams and Fitch’s weighted parsimony program (WTSUBS), the shortest tree differed from the tree produced by unweighted parsimony and is almost identical to the “paleontological” tree, with just the birds and squamates switched in position (fig. 1). The weighted parsimony tree (fig. 1) was found by WTSUBS in all 21 of its different combinations of parameters, when given the “paleontological” tree as the starting tree. WTSUBS became trapped in a local minimum when given the most parsimonious tree (unweighted) as the starting topology and did not find a shorter tree even though this starting topology is ~6% longer than the weighted parsimony tree shown in figure 1.

The difference between the dynamically weighted parsimony tree and the paleontological tree is troubling, but the discrepancy may simply be due to the small number of informative sites available for the Amniota. At site 1317 (discussed above) the sequence is only known for the four mammals (A), for one of the birds (T), for the crocodile (T), and for the outgroup (A). If the lizard, snake, and turtle are found to have the outgroup condition (A) at this site, then weighted parsimony may well unite the birds and crocodile as closest relatives. However, the fact that the topology of the shortest tree may be sensitive to the sequences at just one site suggests that the 18S rRNA molecule is really too highly conserved to be particularly informative when one is trying to assess the relationships between the amniote classes.
"Paleontological" phylogeny

Coelacanth

Mammals

Turtle

Squamates

Crocodilian

Birds

18S rRNA maximum parsimony phylogenies

Unweighted

Coelacanth

Turtle

Crocodilian

Squamates

Mammals

Birds

Weighted

Coelacanth

Mammals

Turtle

Crocodilian

Squamates

Birds

FIG. 1.—Alternative phylogenies of major amniote classes. The "paleontological" tree is based on a maximum-parsimony analysis of morphological characters drawn from fossil and extant taxa (Gauthier et al. 1988). The unweighted parsimony tree is paleontologically highly implausible, whereas the weighted parsimony tree is almost identical to the "paleontological" tree. The major controversy in amniote relationships centers around the groups shown in boldface type. The 18S rRNA sequences analyzed were from \textit{Latimeria chalumnae} (coelacanth), \textit{Pseudemys scripta} (turtle), the two squamates \textit{Sceloporus undulatus} (lizard) and \textit{Heterodon platyrhinos} (snake), \textit{Alligator mississippiensis} (crocodilian), the two birds \textit{Gallus gallus} (galliform) and \textit{Turdus migratorius} (passeriform), and the four mammals \textit{Oryctolagus cuniculus} (rabbit), \textit{Rattus norvegicus} (rat), \textit{Mus musculus} (mouse), and \textit{Homo sapiens} (human).

Bootstrapping in the Light of Substitution Biases

Bootstrapping allows one to gauge the relative strengths of mutually contradictory phylogenetic signals in a data set. However, if the data are biased, then the bootstrapping support may also be biased and should not be interpreted as a statistical measure of confidence in the hypothesis being evaluated. In light of the nonrandomness in the amniote 18S rRNA substitutions, the fact that 88% of bootstrap maximum-parsimony trees show support for a bird-mammal association (Hedges et al. 1990) may just as likely reflect the strength of the systematic biases in the amniote sequences as reflect the likelihood that birds are the closest relatives of mammals.

Acknowledgments

Thanks to Dr. Hedges for providing the aligned sequences; to Norman Pace, Clint Turbeville, and Rudy Raff for helpful criticisms; to Walter Fitch and Eileen Owens for running WTSUBS; to Walter Fitch for making numerous suggestions for improving the manuscript; and to Rudy Raff and the Department of Biology, Indiana University, for providing the facilities for preparing the manuscript. This research was supported by NIH NRSA fellowship 5 F32 HD07332-02.

LITERATURE CITED


WALTER M. FITCH, reviewing editor.

Received February 14, 1991; revision received June 20, 1991.

Accepted June 28, 1991