Letter to the Editor

18S rRNA Sequences and Amniote Phylogeny: Reply to Marshall[1]

S. Blair Hedges and Linda R. Maxson
Department of Biology and Institute of Molecular Evolutionary Genetics, Pennsylvania State University

Marshall (1992) presents a reanalysis of some 18S rRNA data (Hedges et al. 1990) bearing on tetrapod relationships. Our phylogenetic analyses of these data for 26 tetrapods supported, among other groupings, a bird-mammal relationship. However, Marshall concluded that his reanalysis of the amniote data, using a weighted parsimony method (Williams and Fitch 1989, 1990), supported a tree uniting birds and squamates (lizards and snakes). We believe that his results were due to a misapplication of the weighted parsimony method, which was designed for highly variable (noisy) data—not for highly conserved data.

The controversy over amniote relationships involves not only the paleontological data and our 18S rRNA data but other morphological and molecular data sets (Hedges et al. 1990; Hedges and Maxson 1991). The best estimate of amniote phylogeny will not hinge on any one data set unless such a set is very large; more sequence data will be needed to resolve this important question. However, the available 18S rRNA sequence data (Hedges et al. 1990, fig. A1; present paper, fig. 1) clearly support a bird-mammal relationship.

Weighted parsimony methods were developed by Sankoff and Cedergren (1983) and Williams and Fitch (1989, 1990), to account for substitution biases in phylogenetic analysis. Marshall (1992) used the latter method to reanalyze our 18S rRNA data. The basic principle is that rare substitution types should be weighted more heavily than common substitution types because common substitutions are more likely to occur multiple times at the same site, thus obscuring phylogenetic information. The implementation of the weighting is done a posteriori by observing the frequency of the different substitution types in an initial tree and then weighting inversely to those frequencies. Each site also can be weighted inversely to the number of changes at that site in the initial topology.

That there is substitution bias in the 18S rRNA data is not surprising, because it is present in most nucleotide sequence data sets. Mechanisms have been proposed for some types of biases (e.g., transition-transversion bias and codon bias), but the reason for the unequal frequencies of certain substitution types in the 18S rRNA data is presently unknown.

One potential problem with weighted parsimony involves the basic assumption that rare substitution types are more reliable indicators of phylogenetic relationships. This concept is more useful when there is a high probability of multiple changes per site (multiple hits). Weighting rare changes more heavily in a highly conserved set of sequences (such as the 18S rRNA data) is unwarranted because all substitution types convey the same phylogenetic information (i.e., they are equally detectable) regardless of relative frequency. There is a wide "gray zone" where the relative information content of rare changes increases as the probability of multiple hits increases. It is unclear at what point (if any) an inverse weighing scheme proportionately compensates for the increasing noise in the data. Application of such a weighting scheme to a data set which has not reached this noise level constitutes a bias.

Williams and Fitch (1989, 1990) intended their method to be used with noisy data sets. Although the presence of homoplasy in the 18S rRNA data indicates that

1. Key words: amniota, reptilia, aves, mammalia, systematics.

Address for correspondence and reprints: S. Blair Hedges, Department of Biology, 208 Mueller Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802.

Mol. Biol. Evol. 9(2):374–377. 1992.

0737-4038/92/0902-0015$02.00

	BIRDS + MAMMALS							BIRDS + CROCODILIANS		BIRDS + SQUAMATES
				1	1	1	1	1	1	
	4	8	9	3	3	6	6	3	8	
SITE:	0	6	1	6	9	4	5	1	9	7
	7	7	6	0	4	5	0	7	9	2
Coelacanth	U	U	U	A	U	A	U	A	U	U
Frog1	A		U	A	U	A				C
Frog2	A	C	U	A	U	A	U	A	U	C
Frog3	A		U	A		A				C
Frog4	A	C		A		A	U			C
Frog5	C	C		A		A	U			C
Frog6	C	C	U	A	U	A	U			C
Frog7	G	C	U	A	U	A	U		U	C
Frog8	A	C	U	A		A	U			A
Salamander1	G		U	A	U	A				C
Salamander2			U	A	U	A				C
Salamander3	G	C	U	A	U	A	U			C
Salamander4		U	U	A	U	A	U			C
Caecilian1	A	C	U	A		A	U			C
Caecilian2	A		U	A		A				C
Caecilian3	A		U	A		A	U			
Caecilian4	G	U	U	A	U	A	U	A		C
Turtle	G	U	U	A	U	A	U			C
Snake	G		U	A	U	A	U			G
Lizard	G	U	U	A		A	U			G
Crocodilian	G	U	U	A	U	A	U	U	C	C
Bird1	C	C	C	C	C	G	C	U	C	G
Bird2	C		C	C	C	G	C			G
Mammal1	C	C	C	C	C	G	C	A	U	C
Mammal2	C	C	C	C	C	G	C	A	U	C
Mammal3	C	C	C	C	C	G	C	A	U	C
Mammal4	C	C	C	C	U	G	C	A	U	C
TRANSITION		*	*		*	*	*		*	
TRANSVERSION	*			*				*		*

FIG. 1.—Informative sites under parsimony criterion pertaining to sister group of birds, from 18S rRNA sequences of Hedges et al. (1990, fig. A1). Derived sites in the amniotes are in boldface type and underlined. Blanks represent missing data.

some multiple changes have occurred, the low level of sequence divergence (4.4% between amphibians and amniotes) and the low level of three-variant (1.0%) and four-variant (0.2%) sites across 26 taxa indicates that the data set is not noisy. Therefore, the weighting scheme imposed by Marshall (1992) is unwarranted.

Another problem with weighted parsimony, recognized by Williams and Fitch (1989, 1990), is seed-tree topology bias: the output tree depends, to a varying degree, on the input tree. Williams and Fitch (1989) recommended using several different seed trees if one is unsure of the true phylogeny. Marshall used only two seed trees. With the "paleontological" seed tree (birds + crocodilians) he obtained a tree joining birds and squamates, and with the unweighted seed tree (birds + mammals) he obtained the seed topology. Because the birds + squamates tree was 6% shorter than the birds + mammals tree, Marshall concluded that weighted parsimony supports a bird-squamate relationship.

We interpret these results differently. Tree length in a weighted parsimony analysis does not have the same meaning as tree length in a conventional (unweighted) parsimony analysis. It is not a measure of the actual number of substitutions in a tree; rather, it is a compounded value that depends on relative frequencies both of substitution types and of changes per site. It is debatable whether the two trees obtained by

Marshall, each using a different seed tree, can be validly compared. We believe that they cannot be compared, because of the influence of the seed-tree topology on tree length. Even if one argues that they are comparable, it must be shown that the two trees are significantly different in length. Given the results (two input trees, two output trees), one wonders whether every different input tree will result in a different output tree—and whether some third seed-tree topology might result in yet a shorter output tree!

Because of the recognized seed-tree topology bias of weighted parsimony, we believe that the only seed tree that is justifiable is the initial unweighted tree, which in this case is the bird + mammal tree. In that no better tree could be found by using that tree as the seed tree, we interpret Marshall's results as affirming that the 18S rRNA data support a bird-mammal relationship. However, other problems—including the use of the coelacanth, rather than the closer lineage (amphibians), as the outgroup—preclude any interpretations from Marshall's reanalysis.

The influence of the "paleontological" seed-tree topology on the results obtained by Marshall underscores how previous hypotheses of relationships can effect phylogenetic analysis. Our analyses of amniote relationships are unbiased, and we have found support for either mammals (Hedges et al. 1990) or crocodilians (Hedges and Maxson 1991) as the sister group to birds. If future studies show overwhelming support for a bird + crocodilian relationship, then it will be interesting to determine why and how the bird and mammal sequences of several genes (beta hemoglobin, myoglobin, and 18S rRNA) have converged. On the other hand, if a bird + mammal relationship is later confirmed, then it will be interesting to determine how the bird and crocodilian sequences (histone H2B and pancreatic polypeptide) have converged—and how the fossil record can be reinterpreted. Whatever final consensus is obtained, phylogenetic analysis should be independent of previous hypotheses.

Despite our criticisms of weighted parsimony as applied by Marshall, we recognize the importance of understanding and accounting for the biases inherent in sequence data. At present, the biases and the mechanisms responsible for those biases are not well understood. Methods of sequence analysis will surely improve when those mechanisms become better known.

Acknowledgments

This research was supported by a grant from the National Science Foundation (BSR 8918926 to L.R.M. and S.B.H.).

LITERATURE CITED

HEDGES, S. B., K. D. MOBERG, and L. R. MAXSON. 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. Mol. Biol. Evol. 7:607–633.

HEDGES, S. B., and L. R. MAXSON. 1991. Pancreatic polypeptide and the sister group of birds. Mol. Biol. Evol. **8**:888–891.

MARSHALL, C. R. 1992. Substitution bias, weighted parsimony, and amniote phylogeny as inferred from 18S rRNA sequences. Mol. Biol. Evol. 9:000–000.

SANKOFF, D., and R. J. CEDERGREN. 1983. Simultaneous comparison of three or more sequences related by a tree. Pp. 253–263 *in* D. SANKOFF and J. B. KRUSKAL, eds. Time warps, string edits, and macromolecules: the theory and practice of sequence comparison. Addison-Wesley, London.

WILLIAMS, P. L., and W. M. FITCH. 1989. Finding the minimal change in a given tree. Pp. 453–470 *in* B. FERNHOLM, K. BREMER, and H. JÖRNVALL, eds. The hierarchy of life. Elsevier Science, Stockholm.

———. 1990. Phylogeny determination using dynamically weighted parsimony method. Methods Enzymol. **183**:615–626.

WALTER M. FITCH, reviewing editor

Received August 28, 1991; revision received September 3, 1991

Accepted September 3, 1991