

# Molecular evidence for the origin of birds

(evolution/tetrapod/amniote/phylogeny/systematics)

S. BLAIR HEDGES

Department of Biology and Institute of Molecular Evolutionary Genetics, 208 Mueller Laboratory, Pennsylvania State University, University Park, PA 16802

Communicated by Charles G. Sibley, December 9, 1993

**ABSTRACT** The major groups of amniote vertebrates appeared during a relatively short time span at the end of the Paleozoic Era, a fact that has caused difficulty in estimating their relationships. The fossil record suggests that crocodylians are the closest living relatives of birds. However, morphological characters and molecular sequence data from living amniotes have repeatedly challenged this hypothesis by indicating a bird–mammal relationship. DNA sequences from four slow-evolving genes (mitochondrial 12S and 16S rRNA, tRNA<sup>Val</sup>, and nuclear  $\alpha$ -enolase) now provide strong statistical support for a bird–crocodylian relationship.

The major groups of living amniote vertebrates are the mammals, birds, crocodylians, turtles, squamates (lizards, snakes, and amphisbaenians), and sphenodontids (tuataras). The first amniotes (reptiles) appear in the fossil record about 300 million years ago (mya), and by 250 mya nearly all of the major groups had diverged (1). Mammals are believed to represent the basal lineage, and their evolution is seen in the fossil record as a long transition through groups of reptiles (synapsids) with progressively more mammalian characteristics, especially in jaw anatomy. The fossil bird *Archaeopteryx* provides strong support for a link between birds and dinosaurs, and crocodylians are seen as the closest living relatives of birds. The relationships of the squamates, sphenodontids, and turtles are less clear, although all have been placed closer to birds than to mammals (1).

This classical phylogeny of amniotes has been challenged by recent morphological studies of living forms (2–4). Traits such as a single aortic trunk, folded cerebellum, scroll-like turbinals, loop of Henle (kidney), adventitious cartilage, and endothermy are found only in birds and mammals and have been proposed as evidence for a close relationship. Analyses of fossil and recent morphological data indicated that support for a bird–crocodylian relationship rests primarily on the fossil data, and specifically with some mammal-like reptile fossils that place mammals at the base of the amniote tree (5). Molecular sequence data from three genes (myoglobin,  $\beta$ -hemoglobin, 18S rRNA) have supported a bird–mammal grouping (6–8), but sequence data from several other genes have not (8, 9). These conflicting results have created uncertainty about our ability to resolve amniote phylogeny. To resolve the phylogenetic relationships of the major amniote groups, especially the origin of birds, I obtained DNA sequences of four slow-evolving genes in representative amniotes.\*

## MATERIALS AND METHODS

Complete sequences of three adjacent mitochondrial genes (12S rRNA, 16S rRNA, tRNA<sup>Val</sup>) were obtained from a crocodylian (*Alligator mississippiensis*), a lizard (*Sceloporus undulatus*), a sphenodontid (*Sphenodon punctatus*), and a turtle (*Trachemys scripta*) and compared with tetrapod se-

quences available in the data bases: bird (*Gallus gallus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), cow (*Bos taurus*), whale (*Balaenoptera physalus*), seal (*Phoca vitulina*), human (*Homo sapiens*), and frog (*Xenopus laevis*). The three published lungfish sequences (10) were included to root the tree.

For the mitochondrial genes, the sequenced region corresponds to sites 648–3229 in the human sequence (11). The 24 primers used for DNA amplification and sequencing are as follows [for each, laboratory name (L = light and H = heavy strand), 3' location on the complete human sequence, and 5' → 3' sequence (IUPAC code) are given]: 12L1, 1091, AAA AAG CTT CAA ACT GGG ATT AGA TAC CCC ACT AT; 12L2, 617, AAA GCA WRG CAC TGA ARA TGC TWA GAT G; 12L3, 1497, TGA RGC RCG YAC ACA CCG CCC GTC ACC CTC; 12L4, 617, CAA AGC AYA GCA CTG AAG ATG; 12L5, 1093, GAT TAG ATA CCC CAC TAT GC; 12H1, 1478, TGA CTG CAG AGG GTG ACG GGC GGT GTG T; 12H2, 1066, GCA TAG TGG GGT ATC TAA TCC CAG TTT G; 12H3, 1194, CGR GGK KTA TCG ATT AYA GAA CAG GCT CCT CTA G; 12H4, 1475, AGG GTG ACG GGC GGT GTG TRC G; 16L1, 2606, CTG ACC GTG CAA AGG TAG CGT AAT CAC T; 16L2, 2204, GGC CTA AAA GCA GCC ACC TGT AAA GAC AGC GT; 16L2a, 2021, CCA AAC GAG CCT AGT GAT AGC TGG TT; 16L3, 1854, AGC AAA GAY YAA MCC TYG TAC CTT TTG CAT; 16L4, 2945, ACC AAG TTA CCC TAG GGA TAA CAG CGC A; 16L8, 2604, TGA CCG TGC RAA GGT AGC ATA ATM A; 16L9, 2511, CGC CTG TTT ATC AAA AAC AT; 16L11, 1802, GTA CCG CAA GGG AAA GAT GAA; 16H1, 3056, CTC CGG TCT GAA CTC AGA TCA CGT AGG; 16H2, 2579, AGT GAT TAC GCT ACC TTT GCA CGG TCA G; 16H3, 2813, GAG GTT TTT TTT TTC TCC AAG GTC GCC CCA; 16H4, 1825, ATG CAA AAG GTA CRA GGK TTR RTC TTT GCT; 16H5, 2170, TTC TTT ATW GGT GGC TGC TTT TAG GCC YAC; 16H6, 3259, GGA TTT GAA CCY CTG RKW WMA AGR KYT TAR GYC TT; 16H10, 2582, TGA TTA CGC TAC CTT TGC ACG GT; 16H11, 1950, TCC CAC TCT TTT GCC ACA GAG A.

The coding region of the nuclear glycolytic enzyme gene  $\alpha$ -enolase from a crocodylian (*Alligator mississippiensis*) was sequenced and compared with the tetrapod sequences available in the data bases: duck (*Anas platyrhynchos*), human (*Homo sapiens*), rat (*Rattus norvegicus*), and frog (*Xenopus laevis*). The sequenced region corresponds to sites 66–1251 in the 1305-bp coding portion of the human  $\alpha$ -enolase gene (12). The six primers used for DNA amplification and sequencing are as follows [for each, name, 3' location on the complete human sequence, and 5' → 3' sequence (IUPAC code) are given]: AE1, 812, GAC TTC AAR TCY CCC GAT GAY CCC AGC AGR TAC AT; AE2, 1252, CTG AAG TTY YTK CCR GCA AAG CKG GCC TTG CTG CC; AE3, 65, ATM TTT GAY TCY CGY GGG AAY CCY ACW GTK GAG GT; AE4, 472, ATG AAY TCY TGC ATR GCC AGC TTR TTG

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

\*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L28074–28078).

CCA GCA TG; AE5, 407, GAG AAG GGW GTC CCM YTG TAC CGY CAC ATY GCT GA; AE6, 859, TGG TCA AAK GGR TCY TCR ATR GAY ACC ACT GGR TA. RNA was extracted from liver (8) and further purified (MicroFastTrack, Invitrogen) to obtain full-length mRNAs. Following reverse-transcription, and amplification of the  $\alpha$ -enolase cDNA by RNA PCR (Perkin-Elmer/Cetus), single-stranded DNA was generated in a second round of amplification with one of the two primers as limiting. Purification of the template and sequencing followed methods described elsewhere (13). Electronic copies of the alignments are available from the author.

Alignments were performed with CLUSTAL (14) and ESEE (15). Phylogenetic analyses were performed with maximum parsimony (16) and neighbor-joining (17). The neighbor-joining analyses were used in conjunction with a Kimura two-parameter distance (18). Other distances also used were the Jukes-Cantor (19), Tamura (20), and Tamura-Nei (21). Those analyses and some parsimony analyses were performed with MEGA (22). Maximum parsimony analyses were performed with PAUP. Sites containing alignment gaps were not used in the distance analyses and were treated as missing information in the parsimony analyses. Confidence in the

phylogenetic groupings was assessed by the bootstrap method (23), with 2000 replications (24).

## RESULTS

Combined sequences of the three contiguous mitochondrial genes ( $\approx 2990$  bp) were analyzed separately from the  $\alpha$ -enolase sequences (1148 bp). Phylogenetic analyses show strong support for a relationship between birds and reptiles (Fig. 1A). Within that group, the mitochondrial DNA data further show strong support for a bird-crocodylian relationship. Separate analysis of the  $\alpha$ -enolase sequence data also supports a bird-crocodylian relationship (Fig. 1B), although the sequences of a turtle, squamate, and sphenodontid are not yet available. For both data sets, these results were not affected by the use of different distance measures or methods of analysis.

Amino acid sequences from 10 nuclear genes that include at least a bird, mammal, crocodylian, and an outgroup now are available (Table 1). Several additional genes in the sequence data bases meet this criterion (e.g., bradykinins, gonadotropin-releasing hormone, and zinc-finger genes) but homologies are unclear. Combined analysis of all amino acid data

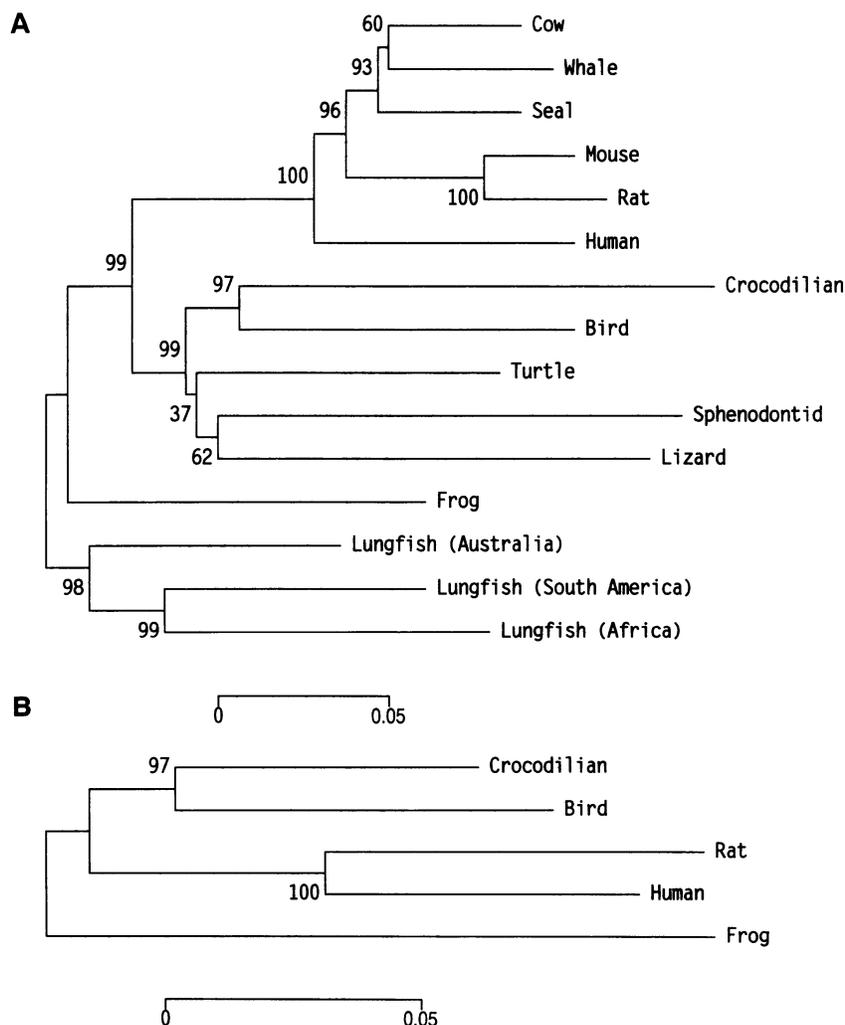


FIG. 1. Amniote relationships inferred from DNA sequences of the mitochondrial 12S rRNA, 16S rRNA, and tRNA<sup>Val</sup> genes (A) and the nuclear glycolytic enzyme gene  $\alpha$ -enolase (B). Bootstrap *P* values obtained with 2000 iterations are indicated. There are 1970 alignable sites in the mitochondrial DNA data set, 746 of which are variable, including 109 parsimony sites (those with two or more variants present in two or more taxa); the tree is rooted with the three lungfish. There are 1186 alignable sites in the  $\alpha$ -enolase data set, 357 of which are variable, including 68 parsimony sites; the tree is rooted with the frog (*Xenopus*). The tree was constructed by the neighbor-joining method using the Kimura distance. Maximum parsimony analyses resulted in identical phylogenetic trees, except that the bird+reptile node in A was resolved as ((bird,crocodylian),(sphenodontid,lizard))turtle).

Table 1. Molecular evidence for the sister group of birds

Gene	Number of sites*			Sister group,† %	
	Total	Variable	Parsimony	Mammal	Crocodylian
<b>Amino acid data</b>					
$\alpha$ -Crystallin A	173	46	7	<1/<1	<b>99/97</b>
$\alpha$ -Enolase	395	53	9	<1/<1	<b>100/100</b>
$\alpha$ -Hemoglobin‡	142	77	7	20/18	<b>52/76</b>
$\beta$ -Hemoglobin‡	146	104	13	<b>83/100</b>	16/<1
Cytochrome <i>c</i> ‡	104	20	2	<b>50/90</b>	<1/4
Histone H2B	57	4	2	12/13	<b>86/87</b>
Insulin‡	55	16	1	28/71	71/14
Myoglobin	154	97	6	<b>53/62</b>	31/9
Pancreatic polypeptide	36	25	9	<1/<1	<b>99/98</b>
Prolactin	199	83	7	<1/18	<b>99/77</b>
<b>Nucleic acid data</b>					
18S rRNA	1770	99	10	<b>80/74</b>	19/26
28S rRNA	354	22	6	<1/<1	<b>93/87</b>
12S rRNA	756	280	28	<1/<1	<b>88/97</b>
16S rRNA	1152	432	76	<1/11	<b>99/83</b>
tRNA <sup>Val</sup> ‡	62	34	5	34/67	46/28
$\alpha$ -Enolase§	1186	357	68	4/1	<b>93/92</b>
Combined amino acid	1461	525	63	<1/4	<b>100/96</b>
Combined nucleic acid	5280	1224	193	<1/<1	<b>100/100</b>

\*Includes only alignable sites.

†Bootstrap *P* values (2000 replications) for neighbor-joining (left of slash mark) and maximum parsimony (right) analyses; values in boldface type correspond to the topology supported by the method. For protein data, neighbor-joining analyses are with a Poisson-corrected distance; the Kimura distance was used with the nucleotide data. For genes with multiple taxa per major group, chicken, human, alligator, and frog (*Xenopus*) were selected.

‡Alternative phylogenetic relationships are obtained when additional amniote sequences are analyzed for this gene.

§The amino acid translations for these sequences were analyzed separately above.

(1461 sites) strongly supports (bootstrap *P* value = 100%) a bird–crocodilian relationship. Combined analysis of available nucleotide sequence data pertaining to amniote relationships (5280 alignable sites), including published data from two additional nuclear genes (18S and 28S rRNA), also strongly supports (bootstrap *P* value = 100%) a bird–crocodilian relationship when compared with mammals and an outgroup (Table 1).

When sequences from each gene are analyzed separately (Table 1), several genes are found to support a bird–mammal relationship as indicated in the earlier studies. In the case of cytochrome *c*, insulin, and tRNA<sup>Val</sup>, the small number of variable (and parsimony) sites cannot be expected to give reliable results in a phylogenetic analysis, as indicated by the alternative relationships obtained when additional taxa are included.

Until now, sequences from  $\beta$ -hemoglobin have provided some of the strongest molecular support for a bird–mammal grouping (6–8, 25). However, amino acid sequences from a sphenodontid, two squamates, and a turtle recently became available in the data bases, and phylogenetic analyses of those data do not support a bird–mammal grouping. Neighbor-joining analysis with a Poisson-corrected distance resulted in this set of relationships (bootstrap values, 2000 replications, for groups contained within parentheses are indicated by superscript on right-hand parenthesis): ((((((tu-atara, squamates)<sup>56</sup>mammals)<sup>57</sup>birds)<sup>41</sup>turtle)<sup>91</sup>crocodilians)<sup>89</sup>frog)lungfish).

## DISCUSSION

These sequence data show strong statistical support for a close relationship between birds and crocodilians. Sequences now are available from a sufficient number of genes (14 genes) to conclude that the initial molecular evidence for a bird–mammal relationship (6–8) was unrepresentative. Fur-

thermore, analysis of newly available sequences from a gene that has long supported a bird–mammal relationship,  $\beta$ -hemoglobin, no longer supports that relationship. Instead, the anomalous position of the alligator agrees with the suggestion that the unusual allosteric properties of crocodilian  $\beta$ -hemoglobin may have obscured an ancestry with birds (25, 26).

Only two genes, myoglobin and 18S rRNA, now provide unambiguous support for a bird–mammal relationship. There are several possible explanations for their support of a bird–mammal rather than bird–crocodilian relationship. Because birds and mammals are endotherms, it has been suggested that molecular convergence has resulted from selection acting on the oxygen-binding capabilities of globins (27). Such an explanation involving convergence is less obvious for 18S rRNA because that gene product plays a more general role (protein assembly) in the cell.

Two other possible explanations involve chance convergence, as a result of (i) the higher proportion of G and C nucleotides present in bird and mammal genomes (28) and (ii) the “attraction” of long branch lengths on trees (29). For the 18S rRNA data, an inspection of the sites supporting a bird–mammal relationship (30) shows that all seven possess G and C nucleotides in those two groups. As noted in the original study (8), the branches leading to birds and mammals also are relatively long, although the sites supporting a bird–mammal grouping do not appear saturated by substitutions. It is possible that a combination of these two factors (a higher rate of change in the lineages leading to birds and mammals and a G+C substitution bias) is responsible.

In the case of myoglobin, there are no nucleotide sequences of the gene available in amniotes other than birds and mammals and therefore base compositional bias is difficult to assess. However, the amino acid data can provide some insights. There are 12 sites that have amino acids shared by birds and mammals but that are not present in other amniote sequences in the protein data bases (taxa examined were

hedgehog, aardvark, mouse, human, pig, rabbit, elephant, cow, seal, whale, platypus, echidna, kangaroo, opossum, penguin, chicken, monitor, alligator, sea turtle, map turtle, tuna, carp, and shark). Myoglobin sites at which some or all birds and mammals share an amino acid not present in other amniotes or fish are 8 (shared residue = Q), 22 (A), 34 (H), 40 (L), 44 (D), 74 (A), 78 (K), 110 (A), 112 (I), 120 (A), 122 (N), and 127 (A). The nonsynonymous nucleotide change can be inferred, and in three cases there was a change from A or T to C or G, in four cases there was a change from C or G to A or T, and in five cases there was no change in G+C content. Thus, there is no evidence that the joining of bird and mammal myoglobin sequences is due to a G+C substitution bias. Nucleotide sequences of myoglobin from additional amniotes may help to understand this apparent case of molecular convergence. However, neither of these genes (18S rRNA and myoglobin) favors a bird-mammal relationship with a high degree of statistical confidence.

Molecular evidence for amniote relationships now supports a relationship between birds and reptiles. Among the reptiles, a sister-group relationship with crocodylians is strongly supported by the large mitochondrial DNA data set that includes representatives of the major groups of amniotes. When considered with the fossil evidence, it appears to be a correct interpretation of this branch of the evolutionary tree of amniotes.

I thank C. Daugherty and L. R. Maxson for the tissue sample of *Sphenodon*; I. Ruvinsky and C. Youngblood for laboratory assistance; C. A. Hass, S. Kumar, and L. R. Maxson for comments on the manuscript; S. Kumar, K. Tamura, and M. Nei for the sequence analysis program MEGA; and the Center for Computational Biology (Biotechnology Institute, Pennsylvania State University) for providing access to database search software. This research was supported in part by a grant from the National Science Foundation (BSR-8918926 to S.B.H. and Linda R. Maxson).

1. Carroll, R. L. (1988) *Vertebrate Paleontology and Evolution* (Freeman, New York).
2. Gardiner, B. G. (1982) *Zool. J. Linnean Soc.* **74**, 207-232.
3. Løvtrup, S. (1985) *Syst. Zool.* **34**, 463-470.
4. Jamieson, B. G. M. & Healy, J. M. (1992) *Phil. Trans. R. Soc. London B* **335**, 207-219.
5. Gauthier, J., Kluge, A. G. & Rowe, T. (1988) *Cladistics* **4**, 105-209.
6. Dene, H., Sazy, J., Goodman, M. & Romero-Herrera, A. E. (1980) *Biochim. Biophys. Acta* **624**, 397-408.
7. Goodman, M., Miyamoto, M. M. & Czelusniak, J. (1987) in *Molecules and Morphology in Evolution: Conflict or Compromise?*, ed. Patterson, C. (Cambridge Univ. Press, Cambridge, U.K.), pp. 141-176.
8. Hedges, S. B., Moberg, K. D. & Maxson, L. R. (1990) *Mol. Biol. Evol.* **7**, 607-633.
9. Hedges, S. B. & Maxson, L. R. (1991) *Mol. Biol. Evol.* **8**, 888-891.
10. Hedges, S. B., Hass, C. A. & Maxson, L. R. (1993) *Nature (London)* **363**, 501-502.
11. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. C. (1981) *Nature (London)* **290**, 457-465.
12. Giallongo, A., Feo, S., Moore, R., Croce, C. M. & Showe, L. C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6741-6745.
13. Hedges, S. B., Bezy, R. L. & Maxson, L. R. (1991) *Mol. Biol. Evol.* **8**, 767-780.
14. Higgins, D. G. & Sharp, P. M. (1988) *Gene* **73**, 237-244.
15. Cabot, E. L. & Beckenbach, A. T. (1989) *Comput. Appl. Biosci.* **5**, 233-244.
16. Swofford, D. L. (1993) PAUP, Phylogenetic Analysis Using Parsimony (Illinois Nat. Hist. Surv., Champaign, IL), Version 3.1.1.
17. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406-425.
18. Kimura, M. (1980) *J. Mol. Evol.* **16**, 111-120.
19. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munroe, H. N. (Academic, New York), pp. 21-132.
20. Tamura, K. (1992) *Mol. Biol. Evol.* **9**, 678-687.
21. Tamura, K. & Nei, M. (1993) *Mol. Biol. Evol.* **10**, 512-526.
22. Kumar, S., Tamura, K. & Nei, M. (1993) MEGA, Molecular Evolutionary Genetic Analysis (Penn. State Univ., University Park, PA), Version 1.01.
23. Felsenstein, J. (1985) *Evolution* **39**, 783-791.
24. Hedges, S. B. (1992) *Mol. Biol. Evol.* **9**, 366-369.
25. Bishop, M. J. & Friday, A. E. (1988) in *The Phylogeny and Classification of the Tetrapods*, ed. Benton, M. J. (Clarendon, Oxford), pp. 33-58.
26. Perutz, M. F., Bauer, C., Gros, G., Leclercq, F., Vandecastrie, C., Schnek, A. G., Braunitzer, G., Friday, A. E. & Joysey, K. A. (1981) *Nature (London)* **291**, 682-684.
27. Dickerson, R. E. & Geis, I. (1983) *Hemoglobin: Structure, Function, Evolution, and Pathology* (Benjamin Cummings, Menlo Park, CA).
28. Bernardi, G., Olofsson, B., Filipinski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. & Rodier, F. (1985) *Science* **228**, 953-958.
29. Felsenstein, J. (1978) *Syst. Zool.* **27**, 401-410.
30. Hedges, S. B. & Maxson, L. R. (1992) *Mol. Biol. Evol.* **9**, 374-377.