J Mol Evol (2003) 57:27–37
DOI: 10.1007/s00239-002-2443-9

JOURNAL OF MOLECULAR EVOLUTION

Congruent Avian Phylogenies Inferred from Mitochondrial and Nuclear DNA Sequences

Jaime García-Moreno,[1,2] Michael D. Sorenson,[3] David P. Mindell[4]

[1] Max Planck Research Centre for Ornithology—Vogelwarte Radolfzell, Schlossalleé 2, 78315 Radolfzell, Germany
[2] Department of Biology, University of Konstanz, 78457 Constance, Germany
[3] Department of Biology, Boston University, 5 Cummington Street, Boston MA 02215, USA
[4] Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, 1109 Geddes Avenue, Ann Arbor MI 48109-1079, USA

Received: 12 April 2002 / Accepted: 30 December 2002

Abstract. Recent molecular studies addressing the phylogenetic relationships of avian orders have had conflicting results. While studies using nuclear DNA sequences tend to support traditional taxonomic views, also supported by morphological data [(paleognaths (galloanseres (all other birds)))], with songbirds forming a clade within Neoaves (all other birds), analyses with complete mtDNA genomes have resulted in topologies that place songbirds as one of the earliest-diverging avian lineages. Considering that over half of the extant bird species are songbirds, these different results have very different implications for our understanding of avian evolution. We analyzed data sets comprising nearly 4 kb of mitochondrial DNA (mtDNA) (complete 12S, ND1, ND2, and cytochrome *b*) plus 600 bp of the nuclear gene c-mos for 15 birds that were chosen to represent all major avian clades and to minimize potential long-branch attraction problems; we used a partition-specific maximum likelihood approach. Our results show congruence with respect to the ingroup among phylogenies obtained with mtDNA and the nuclear gene c-mos, separately or combined. The data sets support a traditional avian taxonomy, with paleognaths (ratites and tinamous) occupying a basal position and with songbirds more derived and forming a monophyletic group. We also show that, for mtDNA studies, turtles may be a better outgroup for birds than crocodilians because of their slower rate of sequence evolution.

Key words: Aves — Avian phylogeny — Phylogeny congruence — mtDNA — c-mos

Introduction

Reconstructing phylogenetic relationships among bird orders remains challenging due to the apparent radiation of many primary avian lineages, yielding short internodes within the tree, and due to uncertainty in rooting the avian tree with distantly related nonavian outgroups.

Recent molecular studies have had strikingly different outcomes. Phylogenies derived from DNA sequences of avian nuclear genes (e.g., RAG-1, CHD) (Groth and Barrowclough 1999; García-Moreno and Mindell 2000) or RNA genes (van Tuinen et al. 2000) have recovered an avian phylogeny consistent with traditional morphological views (e.g., Cracraft 1988), i.e., placing paleognaths (ostrich-like birds and tinamous) and neognaths (all other birds) on either side of the earliest divergence for extant birds and placing songbirds (Passeriformes) as a more derived neognath (Fig. 1a).

Studies based on complete mitochondrial DNA (mtDNA) sequences (Härlid et al. 1998; Mindell et al.

Correspondence to: Dr. Jaime García-Moreno; *email*: Jaime.Garcia-Moreno@uni-konstanz.de

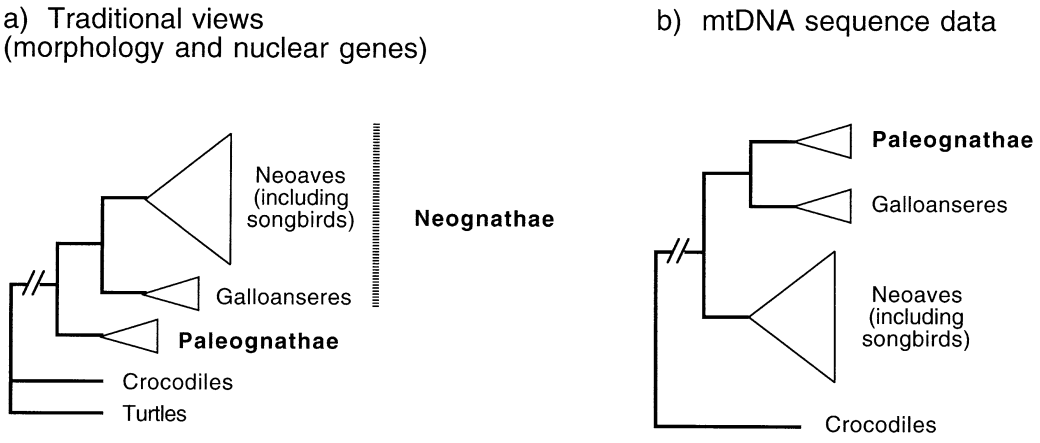


Fig. 1. Major groups in the avian phylogeny. **a** Traditional views based on morphology and nuclear DNA sequences. (**b**) Topology found using mtDNA sequences.

1999; Härlid and Arnason 1999; Haring et al. 2001), on the other hand, have found novel topologies that contradict the traditional views of avian taxonomy by placing songbirds (Passeriformes) as one of the earliest-diverging lineages in the phylogeny (Fig. 1b). These results held up when placing the mtDNA avian phylogeny in the larger contexts of vertebrate phylogeny (Janke et al. 2000) and tetrapod phylogeny (Mindell et al. 1999), though not when the nonavian outgroup taxa were restricted to a crocodile and a turtle (Table 5 of Mindell et al. 1999). Taxon sampling remains sparse for whole-mt genome analyses, and if homoplasious similarity among traits is abundant relative to homologous similarity, this could result in associations of taxa that do not reflect phylogeny (Felsenstein 1978; Hendy and Penny 1989; Philippe and Laurent 1998). Nevertheless, a study by Johnson (2001) that included 916 avian cytochrome *b* (cyt *b*) sequences and crocodile and turtle outgroups concurred with the full-mtDNA studies in placing songbirds at or near the base of the avian tree, demonstrating that the unexpected placement of songbirds is not simply a result of the limited taxon sampling. Mindell et al. (1997), however, pointed out that the ingroup topology of the trees found with mtDNA was consistent with the traditional avian taxonomy.

Here we conduct analyses on a phylogenetically balanced set of taxa and characters (comprising more taxa but fewer characters than the full-mtDNA studies) to see if the incongruence between nuclear and mt data for early avian divergences remains. We find our data sets to be congruent, in both separate and combined analyses, in supporting the more traditional phylogenetic hypothesis, with Passeriformes occupying a nonbasal position among the study taxa. This report of congruence for these avian taxa suggests that the difficulties encountered in interpretation of disparate molecular characters are tractable using existing approaches to modeling sequence evolution and a more balanced taxon and character sampling.

Materials and Methods

Taxonomic Sampling

We built data sets containing representatives of three primary avian clades: Paleognaths, Galloanseres (chickens and ducks), and Neoaves (all other birds, including songbirds). We included representatives of both major groups of songbirds (oscines and suboscines). Additionally, we included other representatives of Neoaves. We chose representatives of penguins, loons, and tube-nosed birds because these groups of birds appear to be related to one another (Sibley and Ahlquist 1990; Nunn et al. 1996; Cooper and Penny 1997). By including several species in each clade we attempted to overcome long-branch attraction problems, which can be most acute when single species with no close relatives are included in a phylogenetic analysis. We used two turtles and two crocodilians as outgroups. We used new and published sequences for both nuclear and mtDNA. Our complete sampling scheme, along with GenBank accession numbers, is shown in Table 1.

DNA Extraction, PCR Amplification, and Sequencing

We extracted DNA from tissues using the QIAGEN DNA-Easy tissue extraction kit, following the manufacturer's instructions. We obtained DNA sequences for three mitochondrial protein genes, NADH dehydrogenase subunits I (ND1) and II (ND2) and cyt *b*, using standard protocols and the primers described by Sorenson et al. (1999). We also obtained DNA sequences for the small-subunit ribosomal RNA gene (12S) and for a 608-bp fragment of the nuclear gene c-mos, (Cooper and Penny 1997; Hughes and Baker 1999; Lovette and Bermingham 2001). PCR products were cleaned using the QIAquick PCR purification kit and cycle-sequenced using the ABI BigDye sequencing kit in 10-μl sequencing reactions with 2 μl of the reaction mix. Sequencing reaction products were cleaned with Sephadex G50 columns and resolved in an ABI PRISM 377 DNA Sequencer.

Sequence Analysis

Alignment for the protein coding genes was straightforward and could be done by eye. We aligned 12S sequences based on a larger

Table 1. Taxonomic sampling for mitochondrial and nuclear sequences[a]

	12S	ND1 & ND2	Cyt *b*	c-mos
Paleognaths				
Tinamiformes (tinamous)				
Crypturellus undullatus	AY139627*	AY1396218*	AY139629*	AF478184*
Ratites (ostriches and allies)				
Struthio camelus	NC_002785	NC_002785	NC_002785	U88429
Rhea americana	NC_000846	NC_000846	NC_000846	—
Rhea pennata	—	—	—	U88430
Galloanseres				
Galliformes (chickens)				
Gallus gallus	NC_001323	NC_001323	NC_001323	M194112
Acryllium vulturinum	AF536739*	AF536745*	AF536742*	—
Numida meleagris	—	—	—	U88425
Anseriformes (ducks)				
Aythya americana Anas	NC_000877	NC_000877	NC_000877	—
Anas platyrhynchus	—	—	—	AF478185*
Dendrocygna arcuata	AF536740*	AF536746*	AF536743*	AF478186*
Passeriformes (songbirds)				
Suboscines				
Sayornis phoebe	AF5367411*	AF536747*	AF536744*	AF478187*
Grallaria squamigera	AY139636*	AY139637*	AY139638*	AF478188*
Oscines				
Hemispingus frontalis	AY139639*	AY139640*	AY139641*[b]	AF478191*
Corvus frugilegus	NC_002069	NC_002069	NC_002069	—
Corvus brachyrhynchus	—	—	—	AF478192*
Other birds				
Sphenisciformes (penguins)				
Aptenodytes patagonicus	AY139621*	AY139622*	AY139623*	AF478193*
Eudyptes chrysocome	AY139630*	AY139631*	AY1 39632*	—
Eudyptes pachyrhynchus	—	—	—	U88420
Gaviiformes (loons)				
Gavia arctica	AY139633*	AY139634*	AY139635*	U88423
Procellariiformes (tube-nosed birds)				
Calonectris diomedea	AY139624*	AY139625*	AY139626*	—
Puffinus griseus	—	—	—	U88421
Crocodylians				
Alligator mississippiensis	NC_001922	NC_001922	NC_001922	—
Gavialis gangeticus	—	—	—	AF478194*
Caiman crocodilus	NC_002744	NC_002744	NC_002744	—
Caiman yacare	—	—	—	AF478195*
Testudines (turtles)				
Chrysemys picta	NC_002073	NC_002073	NC_002073	—
Chelodina rugosae	—	—	—	AF039486[c]
Pelomedusa subrufa	NC_001947	NC_001947	NC_001947	AF109208

[a] Genes sequenced for this study are marked with a superscript asterisk.
[b] First 630 bp of cyt *b* from *H. atropileus* (AF006234).
[c] Only 350 bp in length.

structural alignment (Mindell et al. 1997), including well over 100 avian species, using Clustal X (Thompson et al. 1997). Areas of three or more consecutive characters scoring below 50% in the Clustal X alignment were considered ambiguous and excluded from the analysis, leaving a total of 814 bp in the analysis. All alignments are available from the authors upon request. We tested for homogeneity of base frequencies across taxa using PAUP* 4.08b (Swofford 1999). We did this individually for the five genes and, also, for different gene combinations (Table 2).

Using PAUP* we performed heuristic searches under the maximum likelihood (ML) criterion and under the maximum parsimony criterion with several weighting schemes, including removal of third codon positions. The appropriate models for the different data partitions were estimated with the program Modeltest v. 3.04 (Posada and Crandall 1998) (Table 3), which compares models using likelihood-ratio tests (Huelsenbeck and Crandall 1997). Searches were performed separately for each of the five genes and for different combinations as follows: total evidence (all five genes together), all protein genes, all mt genes, and mt protein genes only (i.e., excluding 12S and c-mos). Note that for some of these combinations, the data set includes chimeric taxa, with mtDNA sequences from one species and nuclear sequences from another within the same genus (*Corvus* and *Eudyptes*) or family (*Acryllium* and *Numida*, *Aythya* and *Anas*, *Calonectris* and *Puffinus*, *Alligator* and *Gavialis*; see Table 1). The inherent assumption of monophyly for each of these chimeric taxa is justified based on independent data and analyses (Sibley and Ahlquist 1990; del Hoyo et al. 1994; Nunn et al. 1996. Mindell et al. 1997; Johnson and Sorenson 1998). Branch support was estimated using Puzzle (Strimmer and von Haeseler 1996) and bootstrap as implemented in PAUP*.

Tree topologies were compared using the Shimodaira–Hasegawa (1999) test (see also Goldman et al. 2000). This test allows for

Table 2. Base composition of the different gene partitions used in this study, and chi-square homogeneity test for base composition

	bp	A	C	G	T	Avg. bp	χ^2(df = 42)	*p*
c-mos	608	0.228	0.259	0.310	0.204	602	23.936	0.989
12S	814	0.313	0.267	0.222	0.198	781	13.088	0.999
ND1	993	0.272	0.330	0.133	0.264	976	41.971	0.472
ND2	1077	0.309	0.346	0.103	0.242	1040	69.152	0.005
Cyt *b*	1158	0.274	0.346	0.129	0.251	1123	56.574	0.066
mtDNA	3976	0.291	0.326	0.142	0.242	3920	115.492	$\ll$0.001
mt prots	3162	0.285	0.341	0.122	0.252	3139	125.065	$\ll$0.001
All prots	3740	0.275	0.328	0.153	0.244	3711	112.914	$\ll$0.001
All chars	4584	0.282	0.317	0.164	0.236	4522	106.699	$\ll$0.001

Table 3. Parameters of the maximum likelihood models estimated on the different gene partitions used in this study[a]

	A↔C	A↔G	A↔T	C↔G	C↔T	α[b]	Invar.
12S	1.000	6.336	1.000	1.000	12.609	1.260	0.596
ND1	0.275	6.991	0.672	0.280	6.991	0.377	0.316
ND2	0.210	2.598	0.134	0.144	1.419	0.814	0.301
Cyt *b*	0.456	8.773	0.999	0.667	8.773	0.420	0.398
c-mos	1.000	4.856	1.000	1.000	6.973	0.317	0.000
mt prots	0.282	4.896	0.312	0.263	2.752	0.624	0.374
All mt	0.607	5.703	0.697	0.334	5.703	0.579	0.391
All prots	0.835	4.353	0.614	0.307	4.353	0.744	0.386
All chars	1.723	4.936	1.266	0.339	9.110	0.802	0.408

[a] Parameters are reversible and measured against the G↔T parameter with a value of 1, as implemented in PAUP* (Swofford 1999).
[b] Shape parameter for the Γ distribution.

the statistical comparison between an optimal tree and other topologies. We used the test as implemented in PAUP*, with one-tailed probabilities and 1000 RELL bootstrap replicates. We performed this test for each partition using the ML models identified as most appropriate with Modeltest.

After finding optimal ingroup topologies (i.e., avian taxa only), we constrained the stable ingroup relationships to assess the attachment point of different outgroup combinations to the ingroup topology that we found to be optimal in previous unrooted analyses. Relative rate tests were performed following the method proposed by Mindell and Honeycutt (1990).

Results

The results of the base composition homogeneity test in the different genes are presented in Table 2. ND2 and cyt *b* were significantly different in their base composition (due to differences in composition of third base codon positions), suggesting potential for biases in phylogenetic analyses. We report our results based on ML models; results of parsimony analyses were nevertheless generally congruent with those of ML. The ML model parameters estimated for each of the five genes used in this study (Table 3) suggest three data partitions that make sense from a biological perspective. One partition encompasses the three mitochondrial protein genes (cyt *b*, ND1, and ND2), all of which were best explained by a GTR + Γ + *I* model, with a similar number of invariant positions, and similar shape parameters for the Γ distribution (evolutionary rate heterogeneity) of about 0.6 ± 0.25. The c-mos nuclear sequences were best explained by the TrN + Γ model (having equal transversion ratios, different transition ratios, and no invariants), and a Γ distribution shape parameter (α) of 0.3. The mt ribosomal 12S gene also fit a TrN + Γ model, but with a Γ distribution of 1.3. We provide additional results for analyses of individual genes and/or gene combinations.

Of the three biologically relevant data partitions described above, analyses of the mt protein genes and c-mos yield trees with almost-identical topologies (Fig. 2). They differ in that suboscine birds are paraphyletic in the c-mos tree, with the phoebe (*Sayornis phoebe*) basal to other songbirds, and in the ostrich (*Struthio camelus*) pairing with tinamou (*Crypturellus undulatus*) rather than rhea (*Rhea americana*) in the mt protein tree. Neither of these topologies is significantly different from each other (Table 4) or from the topology of the tree recovered with all characters (and also with all four protein genes or all four mitochondrial genes; Fig. 2b). Parsimony analysis of the mt protein genes excluding third codon positions results in a single tree showing paleognaths nested inside the loon/penguin/shear-

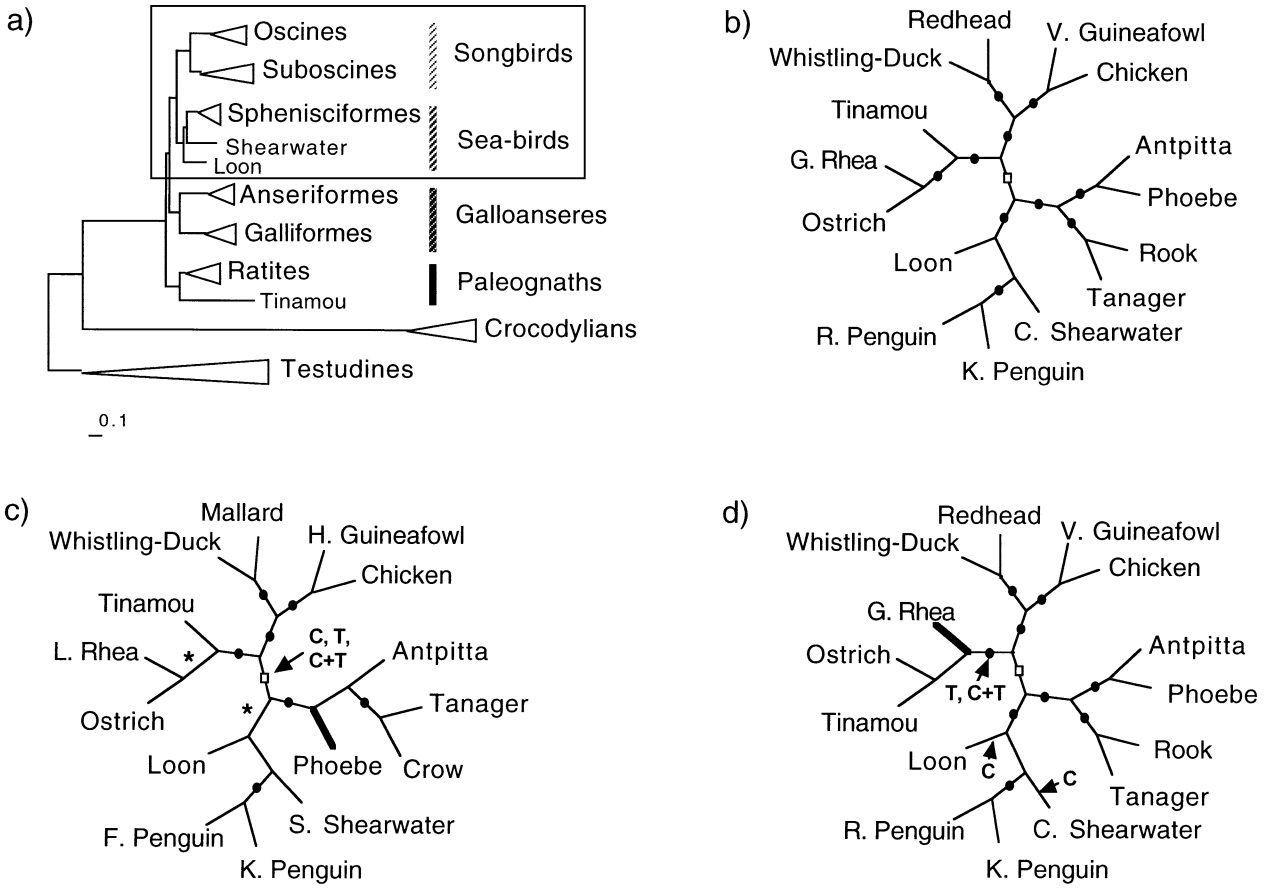


Fig. 2. Phylogenetic trees retrieved with our mitochondrial and nuclear sequences using maximum likelihood on nucleotide sequences (see Table 3). (**a**) Rooted tree based on four mt genes (12S, ND1, ND2, cyt *b*) with both turtles and crocodiles as outgroups and showing the four expected avian clades. Branches are drawn in proportion to the amount of change along them. Notice the short internodes in the avian clade and the rate acceleration in the branch leading to crocodiles. *Boxed* taxa constitute the Neoaves (along with all other birds [not included in this study] and excluding Paleognaths and Galloanseres). (**b**) Unrooted cladogram of the avian ingroup from a. Trees with identical branching order were also found using all characters and combining the four protein genes (see Table 4 and Fig. A1). (**c**) Unrooted cladograms obtained with c-mos and (**d**) combining the three mitochondrial protein genes. Thick branches are those that differ from the topology shown in b. Notice the substantial congruence among the three cladograms—none of the three unrooted trees shown is significantly different from the others. *Arrows* in **c** and **d** denote the attachment point for outgroups; C, crocodiles (two equally likely positions in **d**); T, turtles: C + T, both crocodiles and turtles. *Squares*, branches with puzzle and maximum likelihood bootstrap support values between 50 and 65%; *circles*, branches with support values equal to or higher than 80%. The two branches marked with asterisks in c show high puzzle values (> 90%) but only moderate bootstrap values (between 50 and 65%).

water clade (see Figs. A1 and A2). This topology is only two steps shorter than the one depicted in Fig. 2b and is not statistically different from it (Kishino and Hasegawa test under parsimony criterion $p = 0.695$). A similar analysis for c-mos results in 10 equally parsimonious trees, and their majority rule consensus is in complete agreement with the tree depicted in Fig. 2c except that the loon and the shearwater are unresolved at the base of the clade containing songbirds, loon, and penguins.

The topology of the tree retrieved by 12S (see Fig. A1) differs from those recovered by protein genes in many substantial ways and was rejected as significantly worse using all other data sets (Table 4). By excluding portions of the 12S gene that are variable in length among taxa, however, potential phylogenetic information is also excluded. To find trees based on the entire 12S gene, we also used parsimony-based optimization alignment as implemented in the program POY (phylogeny reconstruction via direct optimization [Gladstein and Wheeler 1996]). Trees found in these analyses showed greater congruence with the overall analysis and with trees from other data partitions, suggesting that incongruent 12S topologies found in the ML analyses were the result of limited phylogenetic information rather than strongly conflicting data. The phylogenetic tree retrieved using only cyt *b* was the most different among the protein coding genes. Although not significantly different from the two partitions discussed here (i.e., mt proteins and c-mos), using a more conservative analysis (e.g., $\alpha = 90\%$) its topology would be rejected as a worse arrangement with three ML models whose data partition recover what we assume is the correct

Table 4. Shimodaira–Hasegawa tests for each of the six optimal topologies recovered with our nine data partitions[a]

Tree[b]/model	All OK	All prots	mt genes	mt prots
1	—	—	—	0.893
2	**0.000***	**0.000***	**0.000***	**0.000***
3	0.857	0.863	0.835	0.796
4	0.594	0.809	0.712	—
5	*0.051*[c]	*0.078*[c]	*0.071*[c]	0.164
6	0.324	0.434	0.342	0.440

	12S	ND1	ND2	Cyt *b*	c-mos
1	0.172	0.909	0.912	0.587	0.717
2	—	**0.014***	**0.001***	**0.014***	**0.022***
3	0.170	—	0.892	0.506	0.717
4	0.075	0.579	—	0.872	0.679
5	**0.018***	0.108	0.238	—	0.329
6	0.070	0.529	0.518	0.543	—

[a] Each topology was tested using the data partition listed in each column and its associated maximum likelihood model. One-tailed tests.
[b] Tree topology 1 was retrieved with all characters, all proteins, and mt genes partitions; topology 2, with 12S characters partition; topology 3, with ND1 partition; topology 4, with ND2 partition and mt proteins partition; topology 5, with cyt *b* partition; and topology 6, with c-mos partition. Topologies are described in Fig. A1.
[c] Italics show topologies that would be rejected under a more conservative test (i.e., $\alpha = 90\%$).
* Significantly worse topology at 95% confidence level (one-tailed test) under the model being used.

topology (Fig. 2b), namely, all alignable characters ($p = 0.051$), all mt genes ($p = 0.071$), and all protein genes ($p = 0.078$) (see Table 4).

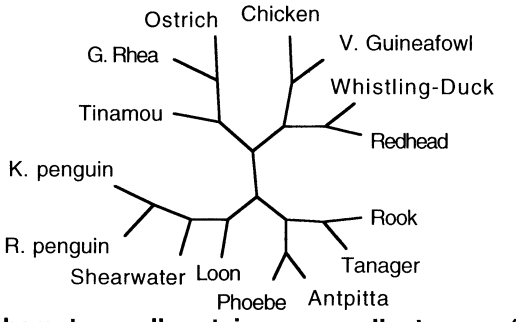
To explore the attachment point of different outgroups to the optimal ingroup phylogeny found in unrooted analyses, we constrained the topology of the ingroup using a backbone constraints tree and then repeated the analyses using the same partitions of data (including 12S) and the appropriate likelihood models. The backbone constraints tree was a strict consensus of five of six optimal topologies found using all nine data partitions (i.e., all topologies except for the one obtained with the 12S partition, which is significantly different from the others; see Table 4).

The different outgroups tested attached to different branches. When we used two crocodilians as the outgroup (*Alligator mississippiensis* and *Caiman crocodylus* for mtDNA, *Gavialis gangeticus* and *Caiman yacare* for c-mos), they attached to the branch leading to Neoaves, with the c-mos nuclear data set (Fig. 2c) resulting in a sister relationship between galloanseres and paleognaths. The two crocodilians attached to the branch leading to the loon/penguin/shearwater clade with the mt protein data set (Fig. 2d) and to the branch between tinamou and ratites with 12S. They attached to the branch leading to oscine songbirds when we used only the cyt *b* data. When we used testudines as outgroup (*Chrysemys picta* and *Pelomedusa subrufa*), we recovered a traditional branching order with the mt protein genes, i.e., (paleognaths (galloanseres (other birds))), with songbirds grouped together in a clade within the Neoaves (Fig. 2d). Turtles attached to the branch leading to Neoaves with the c-mos data set (Fig. 2c), again resulting in a sister relationship of paleognaths and galloanseres. The turtle c-mos fragment we used, however, is only about half as long as the avian and crocodylian sequences (350 vs 608 bp) (Saint et al. 1998). Using cyt *b* only, the turtles also attached to the branch leading to Neoaves. Using both outgroups combined (i.e., turtles and crocodiles) resulted in either the paleognath–galloanseres sister relationship (c-mos) or a traditional branching order (12S, mt proteins).

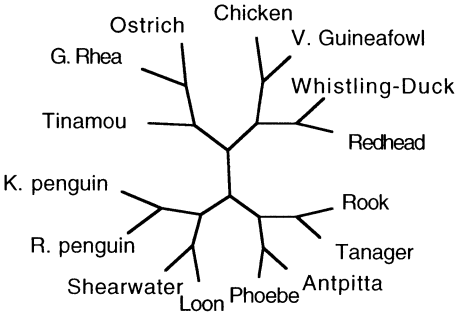
We used the test of Shimodaira and Hasegawa (1999) to test the robustness of the outgroup attachment points. We compared topologies where the songbirds or the loon/penguin/shearwater clades were forced to be basal among birds with the topologies of the nine optimal trees found with each gene partition and with four outgroup taxa (two crocodilians and two turtles). We ran the test using each of the partitions with its associate ML model (the outcome was similar regardless of whether the model parameters were estimated with or without the four outgroup taxa). The only tree that was consistently rejected was the one retrieved using cyt *b*, where the outgroup attached to the branch leading to oscine birds.

We performed relative rate tests on the whole-mtDNA sequences and on the c-mos sequences. We detected a rate difference among birds only between the redhead (*Aythya americana*) and the chicken (*Gallus gallus*), the chicken being slower ($p = 0.02$ rooting with crocodiles, and $p = 0.043$ rooting with turtle). In all comparisons between turtles and crocodiles, rooting with mammals (seal, GenBank acces-

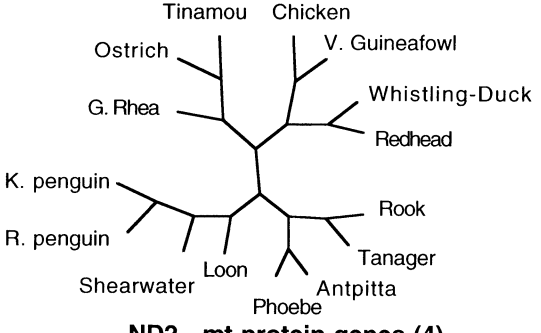
Unrooted trees found in the ML analyses without outgroup and in the POY analyses for 12S. Numbers in parenthesis refer to the number of each tree in Table 4



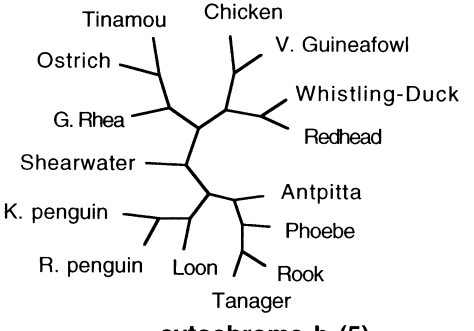
All characters - all protein genes - all mt genes (1)



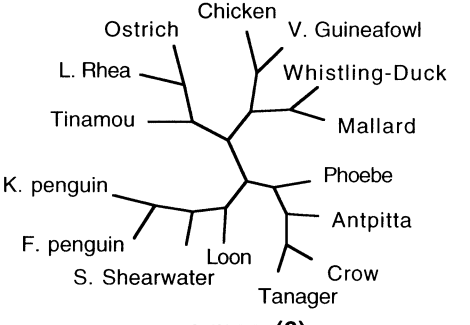
ND1 (3)



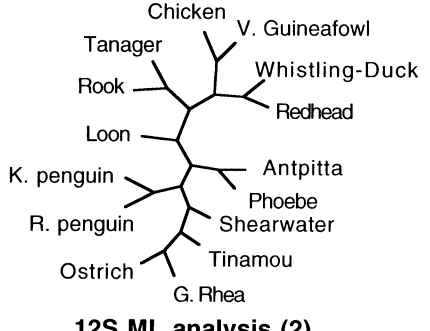
ND2 - mt protein genes (4)



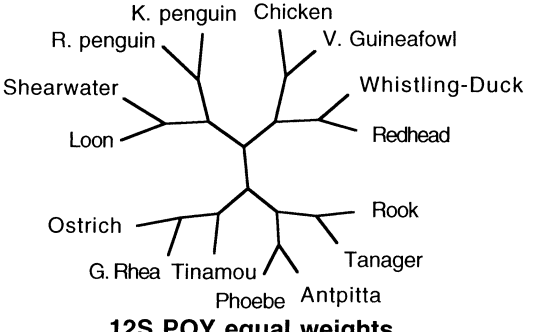
cytochrome b (5)



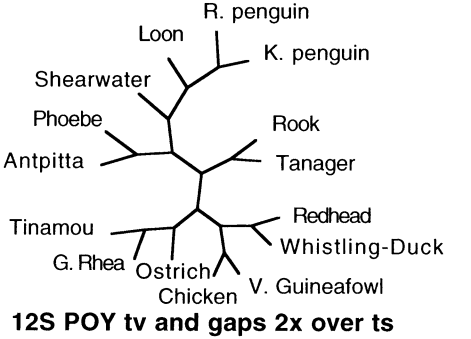
c-mos (6)



12S ML analysis (2)



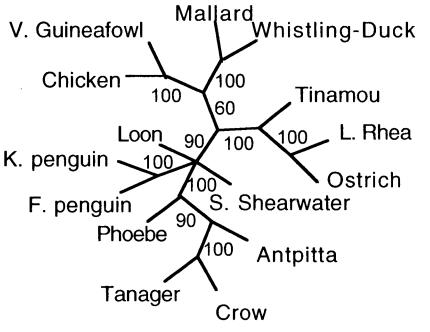
12S POY equal weights



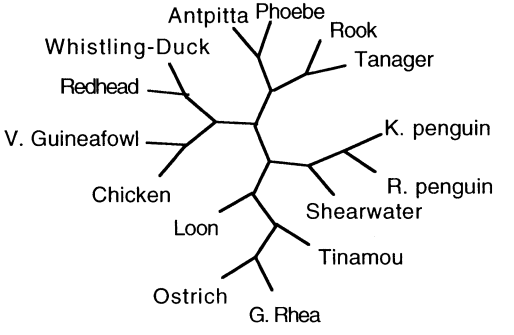
12S POY tv and gaps 2x over ts

Fig. A1. Unrooted trees found in the ML analyses without an outgroup and in the POY analyses for 12S. Numbers in parentheses refer to the tree numbers in Table 4.

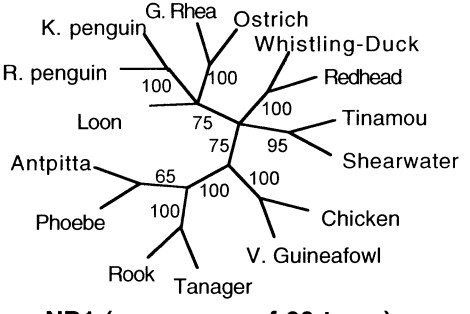
Unrooted trees found in the Maximum Parsimony analyses without outgroup excluding third codon positions and with equal weights. Majority rule consensus trees shown where multiple equally parsimonious solutions were found.



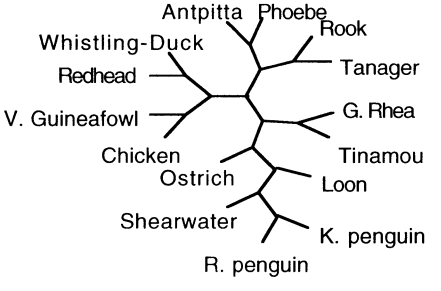
c-mos (consensus of 10 trees)



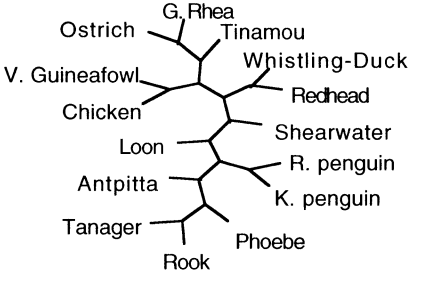
mt protein genes



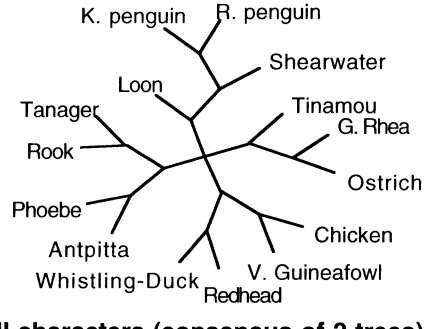
ND1 (consensus of 20 trees)



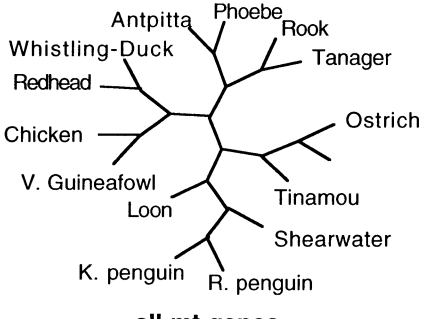
ND2



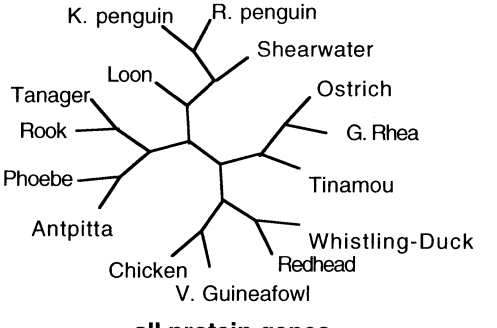
cytochrome b



all characters (consensus of 2 trees)



all mt genes



all protein genes

Fig. A2. Unrooted trees found in the maximum-parsimony analyses without an outgroup excluding third codon positions and with equal weights. Majority-rule consensus trees shown where multiple equally parsimonious solutions were found.

sion number NC_001325; horse, NC_001640) or *Xenopus* (NC_001573), the crocodiles showed an accelerated rate of mutation ($p < 0.001$) relative to the turtles or birds, and birds showed an accelerated rate relative to turtles ($p \leq 0.034$). For c-mos we found that the ducks, chicken, and tinamou showed an accelerated rate in comparison with the loon, penguins, shearwater, guineafowl, and songbirds. However, we did not find significant differences between turtles and crocodiles for a 350-bp fragment of this nuclear gene.

Discussion

Debate about higher-level avian phylogenetics has recently centered on the different results obtained from nuclear and mt sequence data. Nuclear data are generally congruent with traditional views of avian classification (Sibley and Ahlquist 1990; Groth and Barrowclough 1999; García-Moreno and Mindell 2000), whereas mt data are not (Härlid et al. 1998; Härlid and Arnason 1999; Mindell et al. 1999; Johnson 2001). Our results show remarkable congruence in the ingroup topology between three mitochondrial protein genes and an independent nuclear protein gene (Fig. 2). Aside from the odd placement of the rhea in the mt protein tree, the only difference between the two ingroup topologies is whether suboscine songbirds are monophyletic (mt proteins) or paraphyletic (c-mos); recent studies focusing on this particular issue corroborate the monophyly of both songbirds and oscine and suboscine birds (Irestedt et al. 2001; see also García-Moreno and Mindell 2000). Despite the differences in base composition, we found similar results when we applied a parsimony approach to our data sets. This is probably because the data sets are very well balanced, with no long branches. The partial disagreement seen when third codon positions are removed may be due to lack of enough informative characters left in the mt protein genes. This results in some very short internal branches, which in turn make it difficult to resolve the phylogeny unequivocally.

It is interesting to note that all genes and gene combinations, with the exception of 12S on its own, recovered the basic four clades of the tree (paleognaths, galloanseres, loon/penguin/shearwater, and songbirds). This congruence between data sets, and with other studies (Sibley and Ahlquist 1990; Groth and Barrowclough 1999; García-Moreno and Mindell 2000; van Tuinen et al. 2000), suggests that this is the correct overall topology of the avian phylogeny (see Miyamoto and Fitch 1995). Why is it, then, that other studies addressing the avian phylogeny with mtDNA characters reached different conclusions? Although all mt genome-based studies had a very thorough character sampling (full mt genomes), their taxon sampling was limited to between 4 and 10 avian taxa (Härlid et al. 1998; Mindell et al. 1999; Härlid and Arnasson 1999)—some of which appear to have accelerated rates of evolution (e.g., *Falco*, *Smithornis*). Even though our taxon sampling is still very limited (15 birds), we chose taxa in a way that minimizes long-branch attraction problems. This may be the reason why our parsimony analyses are congruent with our likelihood ones. In any case, it appears that more complex model-based likelihood analyses have the power of retrieving optimal topologies from mtDNA data that are compatible with more traditional views and with nuclear DNA studies. In fact, Mindell et al. (1999) found optimal trees compatible with traditional views of avian taxonomy when more complex likelihood models were used (their Table 5).

More difficult to establish is the branching order among these clades because of an inherent outgroup problem. Crocodilians are generally considered to be the extant group most closely related to birds (Cao et al. 2000), but the internode between ingroup and outgroup is so long that it is prone to artifacts (Felsenstein 1978; Philippe et al. 2000) (see Fig. 2a), and it is therefore unclear whether the results obtained are correct. Although in theory it is possible to find a better outgroup than the crocodilians by means of gene duplications (see García-Moreno and Mindell 2000), in practice we have not found a system that can truly bypass this situation. Therefore our study, like all avian molecular analyses, suffers from the same uncertainty regarding the correct branching order of the avian clades. This is reflected in the fact that we are unable to discriminate statistically between alternative hypotheses regarding the most basal avian group. When we use the crocodilians as outgroup, the results with the nuclear gene are consistent with the results obtained with DNA–DNA hybridization (Sibley and Ahlquist 1990), i.e., a sister relationship of paleognaths and galloanseres. This relationship, however, remains uncertain; other characters, such as heteromorphic sex chromosomes and apparent barriers to sex chromosome recombination (García-Moreno and Mindell 2000), appear to link galloanseres more closely to neognaths than to paleognaths. Completely novel would be the topology that places loon/penguin/shearwater as the basal taxon among birds, although some of the oldest avian fossils known belong to these groups (e.g., Ref. 33 in Cooper and Penny 1997). If turtles are used as outgroup with c-mos sequences, we get the paleognath–galloanseres relationship, just as with a crocodilian outgroup. We get a traditional avian topology, however, with the mt protein genes and a turtle outgroup. Using both outgroups combined (i.e., turtles and crocodilians) yields either the paleognath–galloanseres sister relationship (c-mos) or a traditional topology (12S, mt proteins). These differences in the attachment of the root may derive from the fact that the deepest branches of the

ingroup are very short, thus small changes in the outgroup may result in its attachment to a different (contiguous) branch. Moreover, the crocodilian lineage has a significantly faster rate of mtDNA sequence evolution relative to turtles and birds (long-branch leading to crocodiles in Fig. 2a [see also Janke et al. 2000]), and this may explain the different attachment positions for crocodilian and turtle outgroups (Fig. 1). Despite crocodiles being the closest extant relatives of birds, turtles may be more useful as avian outgroups, as their sequences are likely more similar to those of their common ancestor with birds.

If the outgroups we have chosen are not too misleading, our data set is consistent with a derived position of songbirds, which comprise a monophyletic clade. It is not consistent with suggestions that songbirds are the most basal avian order or that they constitute a paraphyletic assemblage.

Conclusions

In this work we have shown congruence in the avian phylogeny obtained with three mitochondrial protein sequences and the nuclear protein c-mos, as both data sets recover phylogenies with very similar ingroup topologies. We have also shown that turtles perform better as outgroup for birds than crocodiles, at least for mtDNA-based studies, despite the latter being the closest living relatives of birds. Our analyses are congruent with traditional avian taxonomy, placing paleognath birds at or near the base of the tree and songbirds as a more derived monophyletic group.

Acknowledgments. This work was supported by NSF Grant DEB 9726427 and a Max Planck Society postdoctoral fellowship (J.G.-M.).

References

Cao Y, Sorenson MD, Kumazawa Y, Mindell DP (2000) Phylogenetic position of turtles among amniotes: Evidence from mitochondrial and nuclear genes. Gene 259:139–148

Cooper A, Penny D (1997) Mass survival of birds across the Cretaceous–Tertiary boundary: Molecular evidence. Science 275:1109–1113

Cracraft J (1988) The major clades of birds. In: Benton MJ (ed) The phylogeny and classification of the tetrapods, Vol. 1. Clarendon Press, Oxford, pp. 339–361

del Hoyo J, Elliot A, Sargatal J (1994) Handbook of the birds of the world. Vol 2. New World vultures to guinea fowl. Lynx Edicions, Barcelona

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410

García-Moreno J, Mindell DP (2000) Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): A case study using avian CHD. Mol Biol Evol 17:1826–1832

Gladstein D, Wheeler WC (1996) POY: Phylogeny reconstruction via direct optimization. American Museum of Natural History, New York

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Syst Biol 49:652–670

Groth JG, Barrowclough GF (1999) Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol Phylogenet Evol 12:115–123

Haring E, Kruckenhauser L, Gamauf A, Riesing MJ, Pinkster W (2001) The complete sequence of the mitochondrial genome of *Buteo buteo* (Aves, Accipitridae) indicates an early split in the phylogeny of raptors. Mol Biol 18:1892–1904

Härlid A, Arnason U (1999) Analyses of mitochondrial DNA nest ratite birds within the Neognathae: Supporting a neotenous origin of ratite morphological characters. Proc R Soc Lond B 266:305–309

Härlid A, Janke A, Arnason U (1998) The complete mitochondrial genome of *Rhea americana* and early avian divergences. J Mol Evol 46:669–679

Hendy MD, Penny D (1989) A framework for the quantitative study of evolutionary trees. Syst Zool 38:297–309

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Syst 28:437–466

Hughes JM, Baker AJ (1999) Phylogenetic relationships of the enigmatic hoatzin (*Opisthocomus hoatzin*) resolved using mitochondrial and nuclear gene sequences. Mol Biol Evol 16: 1300–1307

Irestedt M, Johansson US, Parsons TJ, Ericson PGP (2001) Phylogeny of the major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. J Avian Biol 32:15–25

Janke A, Erpenbeck D, Nilsson M, Arnason U (2000) The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): Implications for amniote phylogeny. Proc R Soc Lond B 268:623–631

Johnson KP (2001) Taxon sampling and the phylogenetic position of Passeriformes: Evidence from 916 avian cytochrome *b* sequences. Syst Biol 50:128–136

Johnson KP, Sorenson MD (1998) Comparing molecular evolution in two mitochondrial protein coding genes (cytochrome *b* and ND2) in the dabbling ducks (tribe: Anatini). Mol Phylogenet Evol 10:82–94

Lovette IJ, Bermingham E (2000) c-mos variation in songbirds: Molecular evolution, phylogenetic implications, and comparisons with mitochondrial differentiation. Mol Biol Evol 17:1569–1577

Mindell DP, Honeycutt RL (1990) Ribosomal RNA in vertebrates: Evolution and phylogenetic applications. Annu Rev Ecol Syst 21:541–566

Mindell DP, Sorenson MD, Huddleston CJ, Miranda HC Jr., Knight A, Sawchuk SJ, Yuri T (1997) Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. In: Mindell DP (ed) Avian molecular evolution and systematics. Academic Press, San Diego, pp. 213–247

Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, Ast JC, Yuri T (1999) Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst Biol 48:138–152

Miyamoto MM, Fitch WM (1995) Testing species phylogenies and phylogenetic methods with congruence. Syst Biol 44:64–76

Nunn GB, Cooper J, Jouventin P, Robertson CJR, Robertson GG (1996) Evolutionary relationships among extant albatrosses (Procellariiformes: Diomedeidae) established from complete cytochrome-b gene sequences. Auk 113:784–801

Philippe H, Laurent J (1998) How good are deep phylogenetic trees? Curr Opin Genet Dev 8:616–623

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D Müller M, Le Guyader H (2000) Early-branching

or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc R Soc Lond B 267:1213–1221

Posada D, Crandall KA (1998) ModelTest: Testing the model of DNA substitution. Bioinformatics 14:817–818

Saint KM, Austin CC, Donnellan SC, Hutchinson MN (1998) *C-mos*, a nuclear marker useful for squamate phylogenetic analysis. Mol Phylogenet Evol 10:259–263

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116

Sibley CG, Ahlquist JE (1990) Phylogeny and classification of birds: A study in molecular evolution. Yale University Press, New Haven, CT

Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol Phylogenet Evol 12:105–114

Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Swofford DL (1999) PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.08b. Sinauer Associates, Sunderland, MA

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin J, Higgins DG (1997) The Clustal X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

van Tuinen M, Sibley CG, Hedges SB (2000) The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. Mol Biol Evol 17:451–457