Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 378 (2006) 65–73
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

# New candidate species most closely related to penguins

Maiko Watanabe [a], Masato Nikaido [a], Tomi T. Tsuda [b,c], Takanori Kobayashi [d], David Mindell [e], Ying Cao [f,h], Norihiro Okada [a,g,*], Masami Hasegawa [f,h]

[a] *Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Yokohama, Kanagawa 226-8501, Japan*
[b] *Faculty of Human Life Sciences, Tokushima Bunri University, Nishihama, Yamashiro-cho, Tokushima 770-8514, Japan*
[c] *Department of Molecular Life Science, Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Bouseidai, Isehara, Kanagawa 259-1193, Japan*
[d] *National Research Institute of Aquaculture, Fisheries Research Agency, Nansei, Watarai, Mie 516-0193, Japan*
[e] *Department of Biology and Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109, USA*
[f] *Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan*
[g] *Department of Evolutionary Biology and Biodiversity, National Institute for Basic Biology, Nishigonaka 38, Myodaiji, Okazaki 444-8585 Aichi, Japan*
[h] *Department of Biosystems Science, Graduate University for Advanced Studies, Shonan Village, Hayama, Kanagawa 240-0193, Japan*

Received 14 January 2006; received in revised form 3 May 2006; accepted 8 May 2006
Available online 17 May 2006
Received by Takashi Gojobori

**Abstract**

The phylogenetic position of the order Spenisciformes in Aves remains unclear despite several independent analyses based on morphological and molecular data. To address this issue, we determined the complete mtDNA sequence of rockhopper penguins. The mitochondrial genome, excluding the region from the D-loop to 12SrRNA, was also sequenced for petrel, albatross, frigatebird, loon and grebe, which previous studies suggest are related to penguins. A maximum likelihood analysis of the phylogenetic placement of penguins with 23 birds, including 17 species whose mtDNA sequences were previously reported, suggested that storks are the closest extant relatives of penguins, with 78% and 56% bootstrap supports, depending on the choice of outgroup species. Thus, ciconiiform birds constitute new candidates as the closest extant relatives of penguins (previously proposed candidates were either gaviiform, podicipediform, or procellariiform birds). In addition to this new evidence, our analysis gave evidence to some of ambiguous relationships in the avian tree: our analysis supported a basal split between passerines and other neoavians within Neoaves, and rejected the monophyly of Falconiformes as well as that of loons and grebes.


*Keywords:* Storks; Phylogeny; Complete sequences of mitochondria

## 1. Introduction

The long-standing issue of the evolutionary origin of penguins, together with the subsequent remarkable transformation that led to their adaptation to the aquatic environment, has been investigated extensively by morphologists, biogeographers and molecular phylogeneticists. The phylogenetic position of the order Sphenisciformes (penguins) in Aves, however, remains unclear.

Penguins have a good fossil record and the known oldest fossil of penguins was found from Late Paleocene or Early Eocene sediments of New Zealand (Fordyce and Jones, 1990). Olson and Hasegawa (1979) reported "giant penguins" from Oligocene, the fossils of extinct penguin-like birds belonging to Pelecaniformes, and the group and its relationship with sphenisciform birds were argued by several researchers (Olson, 1985; Goedert, 1988; Fesuccia, 1999). However, no fossil has yet been described that clearly suggests the origin of penguins, thus, many presumptions have been made regarding the immediate ancestor of modern penguins.

Initial studies on penguin evolution grouped ancestral penguins with flightless ratites (ostriches, rheas, emus, kiwis, tinamous, etc.; Lowe, 1939). However, this idea has been discounted by major

*Abbreviations:* mt, mitochondria.

* Corresponding author. Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan. Tel.: +81 45 924 5742; fax: +81 45 924 5835.
*E-mail address:* nokada@bio.titech.ac.jp (N. Okada).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.05.003

studies, and many recent studies suggest that penguins evolved from a flying ancestor and that loss of flight occurred independently in penguins and ratites. Based on morphological studies, Simpson (1946, 1975) proposed that penguins are closely related to Procellariiformes (albatross, petrel, and shearwaters). Olson (1985) proposed that the closest relatives of penguins are Gaviiformes (loons). Cracraft (1982), using behavioral and morphological characteristics, proposed a new clade, Sphenisciformes/"Gaviiomorphae" (loons, grebes, and the Cretaceous diving birds), as a monophyletic group based on a cladistic analysis of skeletons. Even though he afterward retracted his idea, he consistently has made a point that the loons/grebes clade has a sister relationship with penguins (Cracraft, 1988; Cracraft et al., 2004). Molecular studies have also yielded different conclusions regarding penguin origins that are sometimes inconsistent each other. Immunological distances of proteins indicated that each of the orders Gaviiformes, Procellariiformes, Ciconiiformes (herons, ibises, storks, hammerhead storks, and flamingos), and Podicipediformes (grebes) is closely related to penguins (Ho et al., 1976). Sibley and Ahlquist (1990) used DNA–DNA hybridization studies to propose that Spheniscidae (penguins) be placed in a superfamily, Procellariidea, with Gaviidae (loons), Procellariidae (petrels and Shearwaters) and Fregatidae (frigatebirds). Later, van Tuinen et al. (2001) constructed a new distance matrix that included hybridization and mitochondrial (mt) DNA sequence data and suggested that the clade, Spheniscidae/Gaviidae/Procellariidae, clusters with Ciconiidae (storks) rather than Fregatidae. In any case, all of those distance data indicated that the Procellariiformes/Gaviifomes clade is the most closely related to Sphenisciformes. Nuclear and mitochondrial DNA sequence analyses have also yielded inconsistent conclusions about the phylogenetic position of penguins in the avian tree. van Tuinen et al. (2001) proposed that shearwaters (Procellariidae), rather than loons, are the closest relatives of penguins.

Recently, complete mt-genome analyses have been used to resolve the phylogeny of diverse bird species. The complete mt genome of *Eudyptula minor* (little blue penguin) and other bird species has been analyzed (Slack et al., 2003; Harrison et al., 2004), but these studies did not include loons, grebes, albatrosses, shearwaters, and frigatebirds, which are candidate relatives of penguins. Thus, these studies may not accurately describe the closest extant relatives of penguins.

Although molecular/morphological phylogenetic analyses suggest that aquatic birds, such as Gaviiformes, Podicipediforms, Procellariiformes, Fregatidae, and Ciconiidae, are candidates for the closest extant relative of penguins, the phylogenetic relationships among these birds remain unclear due to discrepancies between the studies described above. Hence, the issue of the closest extant bird relative of penguins remains a mystery.

In addition to the issue of penguin origins, there are several problems with avian phylogeny. Recent molecular and morphological studies have yielded a consensus that the basal split among extant birds resides between palaeognath (ratites and tinamous) and neognaths (all other extant birds) (Cracraft, 1988, 2001; Cracraft and Mindell, 1989; Livezey and Zusi, 2001; van Tuinen et al., 2000; Edwards et al., 2002; Garcia-Moreno et al., 2003; Sorenson et al., 2003; Chubb, 2004; Cracraft et al., 2004; Fain and Houde, 2004; Harrison et al., 2004; Suzuki et al., 2004). This basal divergence in the avian tree is now well established, but the higher-order relationships and those among orders of neognaths birds remain unclear even among recent molecular studies, necessitating a reexamination of the phylogeny by including data from diverse bird species.

To address these issues directly, we determined the complete mt-genome sequences for the rockhopper penguin and five putative relatives, namely the grey petrel, grey-headed albatross, pacific loon, lesser frigatebird, and great-crested grebe. We also conducted three independent analyses using the maximum-likelihood method, focusing especially on clarifying the closest extant relatives of penguins.

## 2. Materials and methods

### 2.1. Sequence data

The complete genome sequences of mitochondria examined in this study were from the following 23 OTUs: *Rhea americana* (greater rhea; GenBank accession no. NC0000846), *Struthio camelus* (ostrich; NC002785), *Gallus gallus* (chicken; NC001323), *Coturnix chinensis* (Chinese blue quail; NC004575), *Coturnix japonica* (Japanese quail; NC003408), *Aythya americana* (redhead duck; NC000877), *Anser albifrons* (white-fronted goose; NC004539), *Corvus frugilegus* (rook; NC002069), *Vidua chalybeata* (steelblue widowfinch; NC000880), *Smithornis sharpei* (grey-headed broadbill; NC000879), *Ciconia ciconia* (white stork; NC002197), *Ciconia boyciana* (oriental white stork; NC002196), *Falco peregrinus* (peregrine falcon; NC000878), *Buteo buteo* (common buzzard; NC003128), *Haematopus ater* (blackish oystercatcher; NC003713), *Arenaria interpres* (turnstone; NC003712), *Eudyptula minor* (little blue penguin; NC004538), *Eudyptes chrysocome* (rockhopper penguin; AP009189), *Diomedea chrysostoma* (grey-headed albatross; AP009193), *Fregata* sp. (frigatebird sp.; AP009192), *Procellaria cinerea* (brown petrel; AP009191), *Gavia pacifica* (Pacific loon; AP009190), and *Podiceps cristatus* (great-crested grebe; AP009194).

### 2.2. Preparation of genomic DNA samples

Fresh samples of liver, muscle, or blood were obtained from aves: rockhopper penguin, grey-headed albatross, brown petrel, Pacific loon, great-crested grebe, frigatebird sp., common buzzard, peregrine Falcon, and oriental white stork. Total genomic DNAs were isolated from each sample using phenol/chloroform extraction and ethanol precipitation (Sambrook, 1989) and stored at 4 °C.

### 2.3. Sequencing methods

We determined the complete mt-genome sequences of six birds using the primer walking method and/or shotgun sequencing. For both methods, each complete mt genome, divided into two fragments, was first amplified by long and accurate (LA) PCR (TaKaRa, Shiga, Japan) using two sets of primers. The primer walking procedure has been described (Nikaido et al., 2000,

2001). Sets of universal primers for first direct sequencing (designed around regions that are highly conserved among avian mt sequences) were derived from Sorenson et al. (1999). Analysis of the newly obtained sequences yielded a second set of primers, and this primer walking procedure was repeated until the entire mt genome was determined. The shotgun sequencing method has been described (Murata et al., 2003). Each of the two mt-genome fragments amplified by LA-PCR was partially digested by DNase I, and the randomly sequenced fragments of them were then assembled to yield a complete mt genome.

### 2.4. Phylogenetic analyses

In order to estimate the phylogenetic relationships, the 12 protein genes encoded on the same strand of mtDNA were analyzed using the ML method (Felsenstein, 1981; Kishino et al., 1990) both in the amino acid and nucleotide sequence levels. ND6 gene on the opposite strand was excluded from the analyses because its nucleotide and amino acid compositions differ from those of the other genes and because this gene contains only a small amount of phylogenetic information (Cao et al., 1988). Sequence alignments were carefully inspected by eye. We excluded all positions with gaps or ambiguous alignments as well as overlapping regions. We used the ProtML program (MOLPHY package, ver.2.3) (Adachi and Hasegawa, 1996a), the TREE-PUZZLE program for quartet-puzzling (QP) analysis (Strimmer and von Haeseler, 1996), and the CodeML program (PAML package, ver.3.14) (Yang, 1997) for the analysis of both the amino acid sequence of proteins and the nucleotide sequence of mt-protein encoding genes. Appropriate models were applied to each analysis, namely the mtREV-F model (Adachi and Hasegawa, 1996b) for amino acid sequences of mt-protein genes, or the codon-substitution model (Goldman and Yang, 1994; Yang et al., 1998) for the nucleotide sequences of mt-protein genes. For the codon-substitution model with the CodeML program, we used Miyata, Miyazawa and Yasunaga's distance (Miyata et al., 1979) with geometric formulae, as described by Yang et al. (1998). The discrete $\Gamma$ distribution (with 8 categories for the amino acid analysis, but with 4 categories for the codon analysis because of the computational burden) for the site-heterogeneity (Yang, 1996) was adopted, and the shape parameter ($\alpha$) of the model was optimized. Bootstrap probabilities (BPs) were estimated by the RELL (resampling of estimated log-likelihoods) method (Kishino et al., 1990) with 10,000 bootstrap resamplings. The RELL method is efficient in estimating BPs without performing ML estimation for each of the resampled data (Hasegawa and Kishino, 1994). Preliminary analysis of concatenated 12 mt proteins using TREE-PUZZLE enabled us to fix some clades if they had high QP supports and if there was no biological controversy concerning these clades. This procedure can reduce the number of candidate trees that would be subsequently analyzed with a more sophisticated approach. Still, the number of candidate trees might be too large to allow exhaustive analysis with a computationally intensive method. In such cases, we performed an approximate likelihood analysis with the ProtML program to reduce the number of candidate trees. The most serious problem with the ML method, when applied to data from many species, is that the number of possible trees increases explosively, and most of these trees are poor and unpromising. In estimating the branch lengths for each tree topology by ML, we have formerly used the time-consuming Newton–Raphson method. The approximate likelihood option implemented in ProtML avoided this process and estimated an "approximate likelihood" from the initial values for the Newton–Raphson method given by the ordinary least squares. We could examine all the possible trees with the approximate likelihood method, and, by excluding unpromising trees by this approximate criterion, we could select the best 20,000 trees for the full likelihood analysis. There is a strong correlation between the approximate likelihood and the maximum likelihood, and, in a practical sense, this is a good method to reduce the computational burden (Adachi and Hasegawa, 1996a). However, even 20,000 trees may be too many for most sophisticated models. Therefore, we further reduced the number of candidate trees by selecting those having log-likelihood scores differing by less than 2SEs for the analysis of the amino acid sequences and less than 1SE for the nucleotide sequence analysis from that of the highest likelihood tree with a simpler model.

## 3. Results and discussion

### 3.1. Features of mitochondrial genomes

The complete mt genome of the rockhopper penguin was determined in this study. The nearly complete mt genomes of five representatives of candidate penguin relatives (grey-headed albatross, lesser frigatebird, grey petrel, Pacific loon, and great-crested grebe) were also determined; portions of the D-loop region were not determined in these species due to difficulties in amplification and the likely presence of repeat sequences. The complete mt genome of the little blue penguin has a relatively long mt control region compared with other avian mt genomes and contains two sets of repeats (Slack et al., 2003). The present work shows that the control region of rockhopper penguin mtDNA also contains two such sets of repeats, one of which contains 81-bp units, whereas the other contains 7-bp units. The number of each of these two repeated motifs in the control region varies, causing length differences in the mt genome due to heteroplasmy.

The content and arrangement of the mt genes determined in this study are consistent with those of the great majority of aves, such as *Struthio*, *Gallus*, *Aythaya*, *Arenaria*, *Ciconia*, *Vidua*, *Corbus* and *Eudyptula* (Haring et al., 2001). While aligning the mt genes, we found variations in the initiation and termination codons and in the length of the genes (Table 1). All of the initiation and termination codons of penguins and five relatives were identical to those of other aves, although the CO2 initiation codon, ND3 initiation codon, ND4 termination codon and ND5 termination codon varied among penguins and the five relatives.

In mtDNA of mammals and birds, initiation codons sometimes vary from ATG to ATA or GTG. This phenomenon is exemplified by the CO1, CO2 and ND5 genes of penguins and their relatives, which use GTG. Among aves, the standard initiation codon for ND3 is ATG, but ND3 of the species determined here uses ATC or ATT. To date, this unusual initiation codon (for isoleucine) has been found only in ND3 of passerines.

Table 1
Length and start/stop codons of mitochondrial protein-encoding genes of birds, as determined in this study

| Species | *Eudyptes chrysocome* | *Gavia Pacifica* | *Podiceps cristatus* | *Diomedea chrysostoma* | *Procellaria cinerea* | *Fregata* sp. |
|---|---|---|---|---|---|---|
| ND1 | | | | | | |
| Length (bases/amino acids) | 978/325 | 978/325 | 979/325 | 978/325 | 978/325 | 978/325 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) |
| Stop codon | AGG | AGG | AGG | AGG | AGG | AGG |
| ND2 | | | | | | |
| Length | 1041/346 | 1041/346 | 1041/346 | 1041/346 | 1041/346 | 1041/346 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) |
| Stop codon | TAG | TAG | TAG | TAG | TAG | TAG |
| CO1 | | | | | | |
| Length | 1551/516 | 1551/516 | 1551/516 | 1551/516 | 1551/516 | 1551/516 |
| Start codon | GTG(Met?) | GTG(Met?) | GTG(Met?) | GTG(Met?) | GTG(Met?) | GTG(Met?) |
| Stop codon | AGG | AGG | AGG | AGG | AGG | AGG |
| CO2 | | | | | | |
| Length | 684/227 | 684/227 | 684/227 | 684/227 | 684/227 | 684/227 |
| Start codon | ATG | ATG | ATG | GTG(Met?) | GTG(Met?) | ATG |
| Stop codon | TAA | TAA | TAA | TAA | TAA | TAA |
| ATP6 | | | | | | |
| Length | 227/684 | 227/684 | 227/684 | 227/684 | 227/684 | 227/684 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) |
| Stop codon | TAA | TAA | TAA | TAA | TAA | TAA |
| ATP8 | | | | | | |
| Length | 165/54 | 168/55 | 168/55 | 168/55 | 168/55 | 168/55 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) |
| Stop codon | TAA | TAA | TAA | TAA | TAA | TAA |
| CO3 | | | | | | |
| Length | 784/261 | 784/261 | 784/261 | 784/261 | 784/261 | 784/261 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) |
| Stop codon | T-- | T-- | T-- | T-- | T-- | T-- |
| ND3 | | | | | | |
| Length | 352/116 | 352/116 | 352/116 | 352/116 | 352/116 | 352/116 |
| Start codon | ATT(Ile) | ATC(Ile) | ATC(Ile) | ATC(Ile) | ATT(Ile) | ATC(Ile) |
| Stop codon | TAA | TAA | TAA | TAA | TAA | TAA |
| ND4L | | | | | | |
| Length | 297/98 | 297/98 | 297/98 | 297/98 | 297/98 | 297/98 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) |
| Stop codon | TAA | TAA | TAA | TAA | TAA | TAA |
| ND4 | | | | | | |
| Length | 1380/459 | 1377/458 | 1368/455 | 1378/459 | 1378/459 | 1378/459 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) |
| Stop codon | TAG | TAA | AGA | T-- | T-- | T-- |
| ND5 | | | | | | |
| Length | 1818/605 | 1815/604 | 1815/604 | 1815/604 | 1815/604 | 1818/605 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | GTG(Met?) | ATG(Met) |
| Stop codon | TAA | TAA | TAA | AGA | AGA | TAA |
| cyt *b* | | | | | | |
| Length | 1146/381 | 1143/380 | 1143/380 | 1143/380 | 1143/380 | 1143/380 |
| Start codon | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) | ATG(Met) |
| Stop codon | TAA | TAA | TAA | TAA | TAA | TAA |

Gray shading: these features are characteristic of length. Oblique lined: these features are characteristic of codons.

Among neognath birds, the standard termination codon for ND4 is T–, an incomplete codon for TAA. However, this codon differed for the ND4 genes of penguins (TAG), loons (TAA), and grebes (AGA).

An extra nucleotide is present in the ND3 gene of several birds and turtles, and is thought not to be translated (Mindell et al., 1998). The ND3 gene of the six birds determined in this study also has this extra nucleotide. Moreover, an extra nucleotide was found in the ND1 gene of grebes (this study), although this nucleotide has not been reported in other aves. We also confirmed this extra nucleotide in the ND1 gene of another grebe (*Podiceps grisegena*; red-necked grebe).

The ND4 genes of the Pacific loon and the great-crested grebe have 458 aa and 455 aa, respectively, whereas ND4 of all neognath birds has 459 aa (Slack et al., 2003).

### 3.2. TREE-PUZZLE analysis

Since there was a large number of sequences to examine from 23 species, it was impossible to investigate the topologies of all possible trees by the ML method because of the explosive increase in the number of possible topologies. TREE-PUZZLE (Strimmer and von Haeseler, 1996) provides an approximation method that can partially overcome such difficulties.

It is well established that the extant birds are classified into four large groups, palaeognaths, galloanserines, passerines, and other neoaves. The basal offshoot of these extant birds is, however, still ambiguous despite extensive studies on this issue (Cracraft and Mindell, 1989; Sibley and Ahlquist, 1990; Mindell, 1997; Harlid et al., 1998; Harlid and Arnason, 1999; Mindell et al., 1999; Waddell et al., 1999; Johnson, 2001; Livezey and Zusi, 2001; van Tuinen et al., 2000; Edwards et al., 2002; Feduccia, 2003; Garcia-Moreno et al., 2003; Slack et al., 2003; Sorenson et al., 2003; Chubb, 2004; Cracraft et al., 2004; Harrison et al., 2004). The current, conventional hypothesis is that the palaeognaths belong to the oldest lineage of birds, as deduced from both morphological and molecular studies. However, several molecular studies are not consistent with this hypothesis. Some of the analyses based on mtDNA sequences support the basal divergence for passerines within an avian tree. These unconventional trees, deduced by several molecular studies were, however, ambiguous and fragile because of the distant outgroups,

sparse taxon sampling, differences in evolutionary rates among taxa, and unrealistic models, which can seriously affect the rooting position of birds (Mindell, 1997; Mindell et al., 1999; Waddell et al., 1999; Garcia-Moreno et al., 2003; Slack et al., 2003). The large majority of the previous molecular and morphological studies, including the complete mtDNA sequences in the studies by Mindell et al. (1999) and Waddell et al. (1999), who have extensively explored the phylogenetic position of Passeriformes, placed ratites and tinamous at the basal position of all aves, and this basal offshoot in the avian tree is also supported by the recent mt study by Harrison et al. (2004), which included the greatest number of bird species. After considering these data, we used two palaeognaths, *S. camelus* (ostrich) and *R. americana* (greater rhea), as outgroups for our analysis.

TREE-PUZZLE was used to deduce the amino acid sequences of the 12 concatenated mt proteins, and the results are summarized as follows: (1) the monophyly of Neognathae was confirmed; (2) the sister-group relationship of Anseriformes (duck and goose) and Galliformes (pheasant and chicken), and the basal split of galloanserines (Anseriformes and Galliformes) from other neognaths were confirmed; (3) the monophyly of each of the following orders, Sphenisciformes, Ciconiiformes, Charadriiformes, Procellariiformes, and Passeriformes, was confirmed. The monophyly of these orders is consistent with previous studies (Sphenisciformes: Jouventin, 1981; O'Hara, 1989; Charadriiformes: Sibley and Ahlquist, 1990; Paton et al., 2003; Sorenson et al., 2003; Harrison et al., 2004; Passeriformes: Raikow, 1982; Sibley and Ahlquist, 1990; Groth and Barrowclough, 1999; Mindell et al., 1999; Sorenson et al., 2003; Harrison et al., 2004). In this TREE-PUZZLE analysis, the QP value of each clade of the above orders was: Sphenisciformes (63), Ciconiiformes (89), Charadriiformes (70), Procellariiformes (87), and Passeriformes (61), even though the reliability of some clades was not very high. (4) The monophyly or paraphyly of the order Falconiformes could not be settled by this analysis since two falconiform birds (buzzard and falcon) remained as multifurcations together with eight other bird lineages.

### 3.3. ML analysis

Our TREE-PUZZLE analysis, using ratites as outgroups, showed the monophyly of Neoaves (neognaths, excluding galloanserines). Namely, splitting occurred between galloanserines and all other neognaths (Neoaves), which is consistent with the majority of previous studies. Several molecular studies based on mtDNA have suggested that passerines diverged earlier than galloanserines, implying the paraphyly of Neoaves; however, the unexpected placement of passerines might be an artifact produced by the limited taxon sampling and insufficient outgroups, as described above (Mindell, 1997; Mindell et al., 1999; Waddell et al., 1999; Garcia-Moreno et al., 2003; Slack et al., 2003). Since the monophyly of Neoaves (including passerines) was recovered by our TREE-PUZZLE analysis, in the subsequent detailed analysis we added the galloanserines (chicken and redhead duck) to outgroups in order to focus on the relationships among neoaves, particularly the penguin relatives. The inter-species relationships within Galloanserae and Passeriformes were fixed, as deduced by the TREE-PUZZLE tree, and the lineages of Sphenisciformes, Ciconiiformes, Charadriiformes and Procellariiformes were treated as one outgroup in the subsequent analysis.

#### 3.3.1. Close relationships between Spheniscidae and Ciconiidae

Although the close relationships among Spenisciformes, Procellariiformes, Gaviiformes, Podicipediformes, and Pelecaniformes have been reported in several independent analyses, the closest extant relatives of penguins remain unclear. Until now, few molecular phylogenetic analyses have included a sufficient number of penguin relatives. Indeed, none of the complete mtgenome analyses have included candidates of penguin relatives (except storks; Slack et al., 2003; Harrison et al., 2004). Our present analysis included a larger set of putative penguin relatives.

The close relationship between penguins and storks was supported by 78% bootstrap value in our ML analysis, which included almost all candidates of penguin relatives, such as gaviiform, podicipediform, procellariiform and pelecaniform birds (Fig. 1). The ciconiiform birds constitute a new candidate

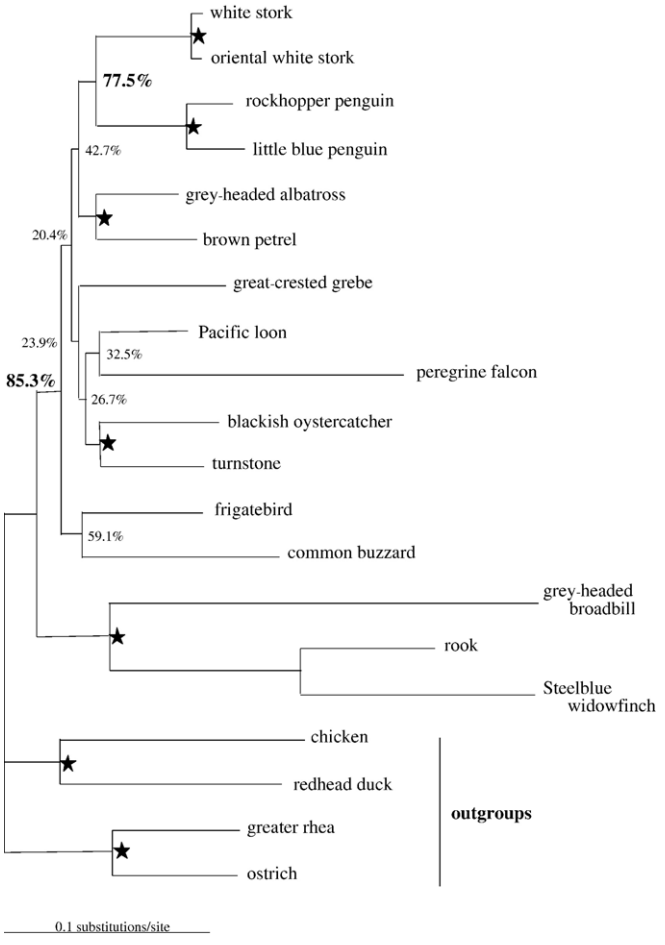


Fig. 1. An ML tree of the concatenated amino acid sequences of 12 mt-proteins with the mtREV-F model (Adachi and Hasegawa, 1996b; Yang, 1996) with four species for outgroups. The horizontal length of each branch is proportional to the estimated number of amino acid substitutions. Numbers indicate percent BPs estimated by the RELL method (Kishino et al., 1990; Hasegawa and Kishino, 1994) using 10,000 replications with the Γ model. The fixed clades in this ML analysis was indicated by asterisks.

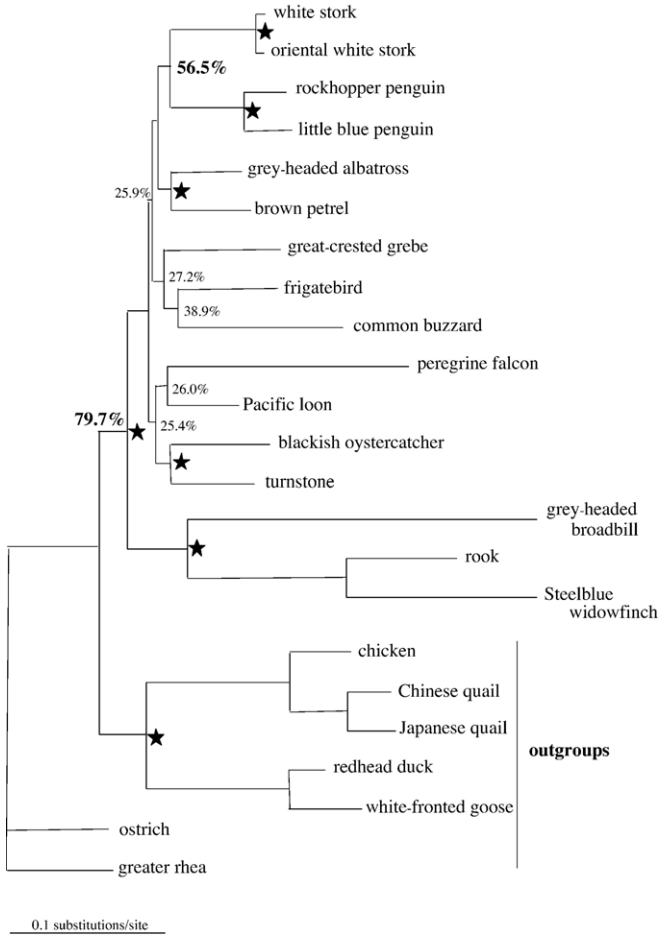


Fig. 2. An ML tree of the concatenated amino acid sequences of 12 mt-proteins with the mtREV-F model with seven species for outgroups. The horizontal length of each branch is proportional to the estimated number of amino acid substitutions. Numbers indicated percent BPs estimated by the RELL method using 10,000 replications with the Γ model. The fixed clades in this ML analysis was indicated by asterisks.

for the closest extant-relatives of penguins since the previously proposed candidates were either gaviiform, podisipediform, or procellariiform birds. Another analysis with seven outgroup species (i.e., three galloanserines: Chinese quail, Japanese quail and white fronted goose plus the four species described above) is shown in Fig. 2. In this tree, the support value of the penguin/stork clade falls to 57%, but the sister relationship between ciconiiforms and spenisciforms remains the most likely relationship. We also performed an analysis of the nucleotide sequences of the protein-encoding genes using the seven outgroup species (codon-substitution model). This analysis also supported the sister relationship between penguins and storks by 79% BP (data not shown). Higher BP support values from the nucleotide analysis than from amino acid analysis is probably because synonymous substitutions are taken into account in the former analysis.

### 3.3.2. The basal split of Passeriformes from Neoaves

Passerines have traditionally been considered as modern groups among the extant birds, mainly from morphological perspectives (Sibley and Ahlquist, 1990; Livezey and Zusi, 2001; Feduccia, 2003; Sorenson et al., 2003; Fain and Houde, 2004; Suzuki et al., 2004). However, as described above, the phylogenetic position of these birds is still ambiguous. The present ML analyses, using palaeognaths and galloanserines as outgroups, showed that passerines diverged first among neoaves; that is, they are the oldest lineages among neoaves. The support value for the basal position of passerines within Neoaves is 85% (Fig. 1), and 80% (Fig. 2), depending on which outgroups are used. This divergence was also suggested by the nucleotide sequence analysis of mt-protein encoding genes using the codon-substitution model (84%; data not shown).

Our ML trees in this study indicate the basal split of Passeriformes from Neoaves. However, we should consider the effect of the long-branch attraction on our phylogenetic tree because of the long branches of the passerines in all of our analyses (e.g. Fig. 1). Furthermore, previous analyses based on partial sequences of both nuclear genome and mt genome including more taxa than ours suggested that some species, for example mousebirds, woodpeckers, parrots, buttonquails, owls, hoatzins, and hornbills, branched basally within Neoaves in their analyses (Cracraft, 2001; Johnson, 2001; van Tuinen et al., 2000; Sorenson et al., 2003; Cracraft et al., 2004; Fain and Houde, 2004; Suzuki et al., 2004). Therefore, our results must be confirmed by extensive analyses with additional taxa, which reduce these long branches to shorter ones and are expected to be basal lineage of Neoaves.

### 3.3.3. Rejection of the monophyly of Falconiformes

The group of raptors is a conventional taxon that is classified based on morphological features and behavioral similarities. This group contains falconiform (hawks, eagles, vultures, falcons, and secretary birds) and strigiform (owls) birds. Some studies based on morphological data suggest that Strigiformes and Falconiformes are closely related (Cracraft, 1988; McKitrick, 1991). However, the majority of morphological and molecular data do not support the monophyly of Falconiformes and Strigiformes (Wetmore, 1960; Brown and Amadon, 1968; Sibley and Ahlquist, 1990; Griffiths, 1994; Livezey and Zusi, 2001). Furthermore, the monophyly of the order, Falconiformes, itself is disputable. Some studies based mainly on morphological data support the monophyly of this order (Griffiths, 1994; Livezey and Zusi, 2001). Other studies suggest that the order is not monophyletic but rather a polyphyletic group. For example, some families of this group

Table 2
RELL bootstrap probabilities supporting monophyletic clades of buzzards/falcons or loons/grebes in our three independent analyses

| Model | Bootstrap probabilities | |
|---|---|---|
| | Buzzard/falcon (%) | Loon/grebe (%) |
| A | 0.56 | 1.48 |
| B | 0.72 | 1.37 |
| C | 0.04 | 0.01 |

Bootstrap probabilities were estimated by the RELL (resampling of estimated log-likelihoods) method (Kishino et al., 1990) with 10,000 bootstrap resamplings.
A: Amino acid sequence analysis by the mtREV-F model with four outgroup species, B: Amino acid sequence analysis by the mtREV-F model with seven outgroup species, C: Nucleotide sequence analysis by the codon-substitution model with seven outgroup species.

are classified near Gruiformes (cranes and rails), Ciconiiformes, or Strigiformes (Friedmann, 1950; Mayr and Amandon, 1951; Ligon, 1967; Rea, 1983; Sibley and Ahlquist, 1990; McKitrick, 1991; Avise et al., 1994; Hedges and Sibley, 1994; Mindell, 1997; Haring et al., 2001; Johnson, 2001; Cracraft et al., 2004). The results from those studies differ depending on the methods, genes, and taxon sampling. That is, the phylogenetic relationships of the birds of prey remain very controversial.

In our ML tree, the peregrine falcon and the common buzzard, both of which are grouped in Falconiformes, do not form a monophyletic clade. Our test for the monophyly of Falconiformes given in Table 2 indicates that this clade is supported only with 0.56% BP, 0.72% BP, or 0.04% BP (codon-substitution model based on nucleotide sequences), although phylogenetic relationships shown in Figs. 1 and 2 are not reliable due to the low support values for some nodes of the relationships concerning either buzzard or falcon. The results strongly support the rejection of the monophyly of Falconiformes with very high BP value. The long branch of the peregrine falcon may have placed this species in the wrong place, and the loon/falcon clade suggested by our analysis may not be real (BP value is also very low). It is noteworthy that the common buzzard has a longer branch than the frigatebird. Therefore, if the long-branch attraction plays an important role in our analysis, and if the monophyly of Falconiformes is real, then falcons would be expected to group with buzzards. However, our analysis shows just the opposite result, thereby strengthening our hypothesis that Falconiformes might not be a monophyletic group.

Our hypothesis of non-monophyly of falconiform birds is consistent with several studies (Mindell, 1997; Johnson, 2001; Sorenson et al., 2003). Harrison et al. (2004) reported that the monophyly of Falconiformes was difficult to recover, and suggested that additional taxa from Falconiformes and Strigiformes be included in future analyses.

#### *3.3.4. Rejection of the sister relationship between loons and grebes*

Cracraft (1988), Cracraft and Mindell (1989) and McKitrick (1991) independently regarded loons and grebes as a monophyletic group based on behavior and common morphological characteristics, such as hindlimb musculature and the skeleton (e.g., squamosal, sternum). On the other hand, Livezey and Zusi (2001) focused on the cranial and vertebral traits and suggested that loons, grebes, and the penguins/procellariiform birds are monophyletic, in which loons and grebes are paraphyletic.

The discrepancy among these morphological studies might be due to differences in taxon sampling, data sets, and the weighting of characters. Also, the interpretation of synapomorphies or ancestral conditions of some of the characteristics might be confused. That is, there are two possible interpretations of the characters shared between the loon and grebe lineages. First, the shared characters may have derived from an immediate ancestor of these species (synapomorphies). Second, ancestral characters may have been retained in these two lineages but lost in other relatives. In either case, similarities in characters, synapomorphies or ancestral characters might have been attained in common ancestors. It is also possible that these two groups attained their similar traits independently, either in parallel or by convergent evolution. Accordingly, it is quite difficult to infer the phylogenetic relationship of loons and grebes based solely on morphological characters.

The results of our analysis supporting the rejection of the monophyletic relationship between loons and grebes, in agreement with several molecular analyses (Sibley and Ahlquist, 1990; van Tuinen et al., 2001; Table 2), suggesting that the morphological similarities of these birds do not reflect a close relationship. The similarities between these two groups should not be regarded as synapomorphies but rather a result of ancestral or independently acquired characters. The morphological similarities between loons and grebes may have evolved via adaptation to aquatic environments, in which their individual diving behaviors evolved similarly.

Ours is the first molecular analysis to include almost all candidate penguin relatives. The results show that the stork is a new candidate for the group most closely related to penguins, a possibility that was not previously raised by morphological studies. We also discussed data that address several phylogenetic issues that remain ambiguous in the avian phylogenetic tree. Our discussion of these issues is not consistent with traditional hypotheses. These issues require thoughtful consideration because the use of mt data, exclusively, may sometimes be misleading due to sparse taxon sampling and/or model misspecification (e.g., Cao et al., 1988; Lin et al., 2002; Nikaido et al., 2003). In the future, our results must be confirmed independently by other analyses; furthermore, extensive analyses of nuclear DNA are needed. The present results constitute the indispensable groundwork required for future phylogenetic analyses of penguins and their relatives using other genetic markers.

## Acknowledgments

We thank Dr. K. Murata for providing DNA samples of *Fregata* sp. and Dr. Y. Shibata for providing tissue samples of loons and grebes. This work was supported by a grant-in-aid to N.O. from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## References

Adachi, J., Hasegawa, M., 1996a. MOLPHY: Programs for Molecular Phylogenetics, ver. 2.3. Institute of Statistical Mathematics, Tokyo.

Adachi, J., Hasegawa, M., 1996b. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. 42, 459–468.

Avise, J.C., Nelson, W.S., Sibley, C.G., 1994. DNA sequence support for a close phylogenetic relationship between some storks and New World vultures. Proc. Natl. Acad. Sci. U. S. A. 91, 5173–5177.

Brown, L., Amadon, D., 1968. Eagles, Hawks and Falcons of the World. McGraw-Hill, New York.

Cao, Y., Waddell, P.J., Okada, N., Hasegawa, M., 1988. The complete mitochondrial DNA sequence of the shark *Mustelus manazo*: resolving vertebrate phylogeny with mitochondrial genome sequences when all known methods fail completely. Mol. Biol. Evol. 15, 1637–1646.

Chubb, A.L., 2004. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK (i). Mol. Phylogenet. Evol. 30, 140–151.

Cracraft, J., 1982. Phylogenetic relationships and monophyly of loons, grebes, and Hesperornithiform birds, with comments on the early history of birds. Syst. Zool. 31 (1), 35–56.

Cracraft, J., 1988. The major clades of birds. In: Benton, M.J. (Ed.), The Phylogeny and Classification of the TetrapodsAmphibians, Reptiles, Birds, vol. 1, pp. 339–361.
Cracraft, J., 2001. Avian evolution, Gondwana biogeography and the Cretaceous–Tertiary mass extinction event. Proc. R. Soc. Lond., B Biol. Sci. 268, 459–469.
Cracraft, J., Mindell, D.P., 1989. The early history of modern birds: a comparison of molecular and morphological evidence. In: Fernholm, B., Bremer, K., Jornvall, H. (Eds.), The Hierarchy of Life. Elsevier Science Publishers B.V., pp. 389–403.
Cracraft, J., et al., 2004. Phylogenetic Relationships among Modern Birds (Neornithes): Toward an Avian Tree of Life. Oxford University Press, New York.
Edwards, S.V., Fertil, B., Giron, A., Deschavanne, P.J., 2002. A genomic schism in birds revealed by phylogenetic analysis of DNA strings. Syst. Biol. 51, 599–613.
Fain, M.G., Houde, P., 2004. Parallel radiations in the primary clades of birds. Evolution. Int. J. Org. Evol. 58, 2558–2573.
Feduccia, A., 2003. Big bang for tertiary birds? Trends Ecol. Evol. 8, 172–176.
Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.
Fesuccia, A., 1999. The Origin and Evolution of Birds, Second ed. Yale University Press, New Haven.
Fordyce, R.E., Jones, C.M., 1990. Penguin history and new fossil material from New Zealand. In: Davis, L.S., Darby, J.T. (Eds.), Penguin Biology. Academic Press, San Diego, pp. 419–446.
Friedmann, H., 1950. Birds of North and Middle America, Falconiformes. U.S. Natl. Mus. Bull. 50, 1–793.
Garcia-Moreno, J., Sorenson, M.D., Mindell, D.P., 2003. Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. J. Mol. Evol. 57, 27–37.
Goedert, J.L., 1988. A new late Eocene species of Plotopteridae (Aves: Pelecaniformes) from northwestern Oregon. Proc. Calif. Acad. Sci. 45, 97–102.
Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11, 725–736.
Griffiths, C.S., 1994. Monophyly of the Falconiformes based on syringeal morphology. Auk 111, 787–805.
Groth, J.G., Barrowclough, G.F., 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, M.J., Pinsker, 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. Evol. 18, 1892–1904.
Harlid, 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.
Harlid, A., Janke, A., Arnason, U., 1998. The complete mitochondrial genome of *Rhea americana* and early avian divergences. J. Mol. Evol. 46, 669–679.
Harrison, G.L., McLenachan, P.A., Phillips, M.J., Slack, K.E., Cooper, A., Penny, D., 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the Late Cretaceous. Mol. Biol. Evol. 21, 974–983.
Hasegawa, M., Kishino, H., 1994. Accuracies of the simple methods for estimating the bootstrap probability of a maximum likelihood tree. Mol. Biol. Evol. 11, 142–145.
Hedges, S.B., Sibley, C.G., 1994. Molecules vs. morphology in avian evolution: the case of the "pelecaniform" birds. Proc. Natl. Acad. Sci. U. S. A. 91, 9861–9865.
Ho, C.Y.-K., Prager, E.M., Wilson, A.C., Osuga, D.T., Feeney, R.E., 1976. Penguin evolution: protein comparisons demonstrate phylogenetic relationship to flying aquatic birds. J. Mol. Evol. 8, 271–282.
Johnson, K.P., 2001. Taxon sampling and the phylogenetic position of Passeriformes: evidence from 916 avian cytochrome *b* sequences. Syst. Biol. 50, 128–136.
Jouventin, P., 1981. Visual and Vocal Signals in Penguins, their Evolution and Adaptive Characters. Paul Parey Scientific Pub.
Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. J. Mol. Evol. 31, 151–160.
Ligon, J.D., 1967. Relationships of the cathartid vultures. Occas. Papers - Univ. Mich., Mus. Zool. 651, 26.
Lin, Y.-H., et al., 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol. Biol. Evol. 19, 2060–2070.
Livezey, B.C., Zusi, R.L., 2001. Higher-order phylogenetics of modern aves based on comparative anatomy. Neth. J. Zool. 51, 179–205.
Lowe, P.R., 1939. Some additional notes on Miocene penguins in relation to their origin and systematics. Ibis 3, 281–294.
Mayr, E., Amandon, D., 1951. A classification of recent birds. Am. Mus. Novit. 1496, 1–42.
McKitrick, M.C., 1991. Phylogenetic analysis of avian hindlimb musculature. Univ. Mich. Mus. Zool. Misc. Publ. 179, 1–85.
Mindell, E.A., 1997. Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. In: M.D., P. (Ed.), Avian Molecular Evolution and Systematics. Academic Press, San Diego, pp. 213–247.
Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. Mol. Biol. Evol. 15, 1568–1571.
Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., Hasegawa, M., Ast, J.C., Yuri, T., 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. 48, 138–152.
Miyata, T., Miyazawa, S., Yasunaga, T., 1979. Two types of amino acid substitutions in protein evolution. J. Mol. Evol. 12, 219–236.
Murata, Y., et al., 2003. Afrotherian phylogeny as inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. 28, 253–260.
Nikaido, M., Harada, M., Cao, Y., Hasegawa, M., Okada, N., 2000. Monophyletic origin of the order Chiroptera and its phylogenetic position among mammalia, as inferred from the complete sequence of the mitochondrial DNA of a Japanese megabat, the Ryukyu flying fox (*Pteropus dasymallus*). J. Mol. Evol. 51, 318–328.
Nikaido, M., et al., 2001. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. J. Mol. Evol. 53, 508–516.
Nikaido, M., Cao, Y., Harada, M., Okada, N., Hasegawa, M., 2003. Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. Mol. Phylogenet. Evol. 28, 276–284.
O'Hara, R.L., 1989. An estimate of the phylogeny of the living penguins. Am. Zool. 29.
Olson, S.L., Hasegawa, Y., 1979. Fossil counterparts of giant penguins from the North Pacific. Science 206, 688–689.
Olson, S.L., 1985. The fossil record of birds. In: Farner, D.S., J.K., Parkes, K.C. (Eds.), Avian Biology. Academic Press, New York.
Paton, T.A., Baker, A.J., Groth, J.G., Barrowclough, G.F., 2003. RAG-1 sequences resolve phylogenetic relationships within Charadriiform birds. Mol. Phylogenet. Evol. 29, 268–278.
Raikow, R.J., 1982. Monophyly of the Passeriformes: test of a phylogenetic hypothesis. Auk 99, 431–445.
Rea, A., 1983. Cathartid affinities: a brief overview. In: Wilbur, S.A., Jackson, J.A. (Eds.), Vulture Biology and Management. University of California Press, Berkeley, pp. 26–54.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds. Yale University Press, New Haven and London.
Simpson, G.G., 1946. Fossil penguins. Bull. Am. Mus. Nat. Hist. 87, 1–100.
Simpson, G.G., 1975. Fossil penguins. In: Stonhouse, B. (Ed.), The Biology of Penguins. Macmillan, London, pp. 19–41.
Slack, K.E., Janke, A., Penny, D., Arnason, U., 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene 302, 43–52.
Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T., Mindell, D.P., 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol. Phylogenet. Evol. 12, 105–114.
Sorenson, M.D., Oneal, E., Garcia-Moreno, J., Mindell, D.P., 2003. More taxa, more characters: the hoatzin problem is still unresolved. Mol. Biol. Evol. 20, 1484–1498.
Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Suzuki, N., Laskowski Jr., N., Lee, Y.C., 2004. Phylogenetic expression of Galalpha1–4Gal on avian glycoproteins: glycan differentiation inscribed in the early history of modern birds. Proc. Natl. Acad. Sci. U. S. A. 101, 9023–9028.

van Tuinen, M., Sibley, C.G., Hedges, S.B., 2000. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. Mol. Biol. Evol. 17, 451–457.

van Tuinen, M.V., Butvill, D.B., Kirsch, J.A.W., Hedges, S.B., 2001. Convergence and divergence in the evolution of aquatic birds. Proc. R. Soc. Lond., B 268, 1345–1350.

Waddell, P.J., Cao, Y., Hasegawa, M., Mindell, D.P., 1999. Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. Syst. Biol. 48, 119–137.

Wetmore, A., 1960. A classification for the birds of the world. Smithson. Misc. Collect. 139, 1–37.

Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. TREE 11, 367–372.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z., Nielsen, R., Hasegawa, M., 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. 15, 1600–1611.