

# Points of View

*Syst. Biol.* 50(1):128–136, 2001

## Taxon Sampling and the Phylogenetic Position of Passeriformes: Evidence from 916 Avian Cytochrome *b* Sequences

KEVIN P. JOHNSON<sup>1</sup>

*Department of Biology, University of Utah, Salt Lake City, UT 84112-0840, USA*

Attempts to resolve relationships among major groups of birds from DNA sequences have been plagued by the problem of apparent long branches connected by short internodes (Sibley and Ahlquist, 1990) and the difficulties of sampling a large number of taxa in one study. A recent study of relationships among major avian lineages based on complete mitochondrial DNA sequences (Mindell et al., 1997, 1999) produced a striking result of a basal position for passerine birds (Aves: Passeriformes), placing them as sister to all other birds. This arrangement is novel and differs from Sibley and Ahlquist's (1990) tree based on DNA–DNA hybridization data and from virtually all previous classification schemes for birds (Fig. 1). Most classifications of birds place Passeriformes well within other neognath lineages (e.g., Morony et al., 1975; Howard and Moore, 1980; del Hoyo et al., 1992; Groth and Barrowclough, 1999) and place paleognaths (ratites and tinamous) as the sister to neognaths (all other birds).

One potential criticism of Mindell et al.'s (1997, 1999) studies is that taxon sampling is very sparse (five and seven avian species, respectively). A study with greater taxonomic coverage (21 avian species) based on only mitochondrial cytochrome *b* (cyt *b*) DNA sequences (Harlid et al., 1998) still resulted in a sister relationship between passerines and all other birds. Nonetheless, all of these DNA sequence studies include only a very small frac-

tion of the >9,000 bird species. The basal position of Passeriformes with respect to other birds in these studies might result from long branch attraction (Felsenstein, 1978) rather than phylogenetic relationship. Phylogenetic accuracy can increase with increased taxon sampling, even at the expense of a decrease in the number of characters sampled per taxon (Hillis, 1996, 1998; Graybeal, 1998; Poe, 1998). Increased taxon sampling in phylogenetic studies generally breaks up long branches, thus increasing detected homoplasy (Graybeal, 1998). Improvement of phylogeny estimation through increased taxon sampling can occur even when the number of base pairs sequenced per taxon is greatly decreased (Graybeal, 1998). However, in certain cases, increased taxon sampling may not aid phylogeny reconstruction if taxa are added to branches in particular ways (Kim, 1998; Poe and Swofford, 1999).

If the position of Passeriformes in the mitochondrial genome tree of Mindell et al. (1997, 1999) results from long branch attraction, then increased taxon sampling can potentially provide a better estimate of the phylogeny. To assess the sensitivity of the basal position of Passeriformes to increased taxon sampling, I analyzed sequences of the mitochondrial cyt *b* gene for 916 species of birds from GenBank. This data set includes nearly 10% of all avian species. Because published cyt *b* sequences are not complete for all of these taxa, I analyzed two combinations of the data set: (1) all 916 sequences regardless of length and (2) only the 703 sequences that made up >50% of the total length of the gene. These treatments are referred to as the 916- and 703-taxon data sets. I report

<sup>1</sup>Present address (and address for correspondence): Illinois Natural History Survey, 607 East Peabody Drive, Champaign, IL 61820-6970. E-mail: Kjohnson@inhs.uiuc.edu

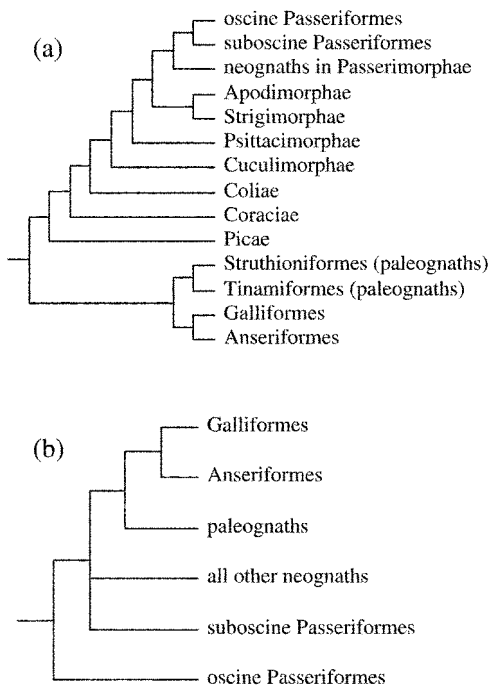


FIGURE 1. Phylogeny of avian major lineages from (a) Sibley and Ahlquist (1990) and (b) Mindell et al. (1999). The sister relationship between paleognaths and Galliformes + Anseriformes is not included in Sibley and Ahlquist's final classification because they found other evidence conflicting with this relationship.

on the phylogenetic position of the large avian order Passeriformes in each analysis and compare my results with those of Mindell et al. (1997, 1999), Harlid et al. (1998), and Sibley and Ahlquist (1990).

#### METHODS

I downloaded sequences of *cyt b* from GenBank for 916 species of birds (including 9 paleognaths [ratites and tinamous], 352 Passeriformes, and 555 other neognaths) and 4 outgroup species (alligator, iguana, and two turtles). Only single representatives of each species were included. This data set was uneven in terms of the length of *cyt b* sequenced for each taxon, which ranged from ~200 bases to >1,100 bases. Missing data were treated as ambiguous character states (N's). Parsimony analysis of 916 taxa is computationally inefficient; however, the performance of neighbor joining (NJ) in recovering the correct phylogeny is similar to that of parsimony (Huelsenbeck and Hillis, 1993). Thus, I first analyzed this data set by NJ the uncorrected p-distances with PAUP\* (Swofford,

1999) and rooted the entire tree on the two turtle sequences. This is the 916-taxon data set. I also performed a NJ analysis by excluding taxa that did not have at least 50% of the total *cyt b* sequence length. All the included taxa had  $\geq 622$  bp-sequenced and this was the case for 703 species of birds (the 703-taxon data set). For both data set combinations, I repeated the analysis by excluding the outgroup and used midpoint rooting to assess whether there might be differential rates between the two groups at the root. I repeated the NJ analysis for data combinations and used both Kimura two-parameter (Kimura, 1980) and the HKY 85 (Hasegawa et al., 1985) models of sequence substitution to assess the sensitivity of the results to models of sequence evolution. These analyses were also repeated to include rate heterogeneity according to a gamma distribution ( $\alpha = 0.20$ ). To assess general support for the neighbor joining tree, I conducted a bootstrap analysis of the 703-taxon NJ tree.

In addition to NJ analysis, I used PAUP\* to conduct parsimony analysis of the 703-ingroup-taxon data set with outgroup rooting. For this data set, there were 699 informative characters and 191 variable uninformative characters. First, I performed heuristic parsimony searches, using 1,000 random addition replicates. For each replicate I held five trees. As an alternative to random addition, I also used the NJ tree from above as a starting tree for heuristic parsimony searches (20 trees were held from this search). To assess relative support for various nodes under parsimony, I conducted fast parsimony jackknife (Farris et al., 1996; 50% character deletion) and bootstrap (Felsenstein, 1985) replicates (100 replicates). Using the above parsimony procedure, I also conducted 300 random addition replicates based on amino acid residues.

For each NJ analysis, I assessed the relationships between major bird lineages as identified by Sibley and Ahlquist (1990), including ratites, tinamous, suboscine passerines, oscine passerines, gallinaceous birds, waterfowl, and the remainder of neognaths (all other birds). To determine the congruence of trees from both data sets with the previous classification, I computed the fraction of 21 major groupings (orders and families of birds) that were monophyletic in each analysis. I restricted these assessments to large groups (more than four species

for all but tinamous) and considered only arrangements involving species that were in common to both of the data set combinations. These 21 groups include Struthioniformes (ratites), Tinamiformes (tinamous), oscine Passeriformes (songbirds), suboscine Passeriformes (nonoscine songbirds), Galliformes (fowl), Anseriformes (waterfowl), Piciformes (woodpeckers), Psittaciformes (parrots), Columbiformes (pigeons and doves), Charadriiformes (shorebirds and allies), Caprimulgiformes (nightjars), Trogoniformes (trogons), Cuculiformes (cuckoos), Procellariiformes (tubenoses), Sphenisciformes (penguins), Gruidae (cranes), Apodidae (swifts), Falconidae (falcons), Ciconiidae (storks), Accipitridae (hawks, eagles, and Old World vultures), and Cathartidae (New World vultures).

For the 703-taxon combination, I assessed the tree lengths under parsimony for (1) the outgroup rooted NJ tree (Fig. 2c), (2) the same tree with Passeriformes monophyletic, (3) a tree moving Passeriformes inside Anseriformes plus Galliformes, and (4) a tree making Passeriformes sister to Columbiformes (see Fig. 3), similar to Sibley and Ahlquist's (1990) result. For each of these trees, the relationships within all nonmanipulated taxa were kept as in the NJ tree, while I moved branches as indicated above by using MacClade (Maddison and Maddison, 1992). I performed Templeton's one-tailed tests (Templeton, 1983) in PAUP\* on these trees to test for significantly worse topologies under the cyt *b* data.

## RESULTS

Uncorrected pairwise divergences between taxa ranged from <1% between closely related species to ~32% between alligator and birds. Within birds, the divergence between Passeriformes and other birds ranged between 20% and 23%. NJ analysis of both data sets produced trees in which Passeriformes are not embedded within other neognath birds (Figs. 2a–d). In all four cases (data and rooting combinations) oscine passerines are monophyletic, but in three of four cases Passeriformes (oscine and suboscine passerines) are paraphyletic. Suboscines are sister to all other nonpasserine neognath birds in these three cases. Rooting on only *Alligator* for the 703-taxon data set also resulted in identical relationships

among the major groups (as in Fig. 2c). Moreover, the 703-taxon data set produced the largest fraction of monophyletic traditional groups (Table 1). NJ bootstrap replicates supported monophyly of many groups of birds; however, the arrangement among the major lineages (paleognaths, Passeriformes, other neognaths) was not supported at >50% (Fig. 3).

Parsimony tree lengths for the 703-taxon data set revealed that a tree making Passeriformes monophyletic was two steps shorter than the NJ tree, but the two trees placing Passeriformes within other neognaths were significantly worse under the Templeton test ( $P < 0.05$ ). The complete tree from this data set (Fig. 3) indicates relatively even branches (in terms of the lengths of the tips).

Parsimony analyses of the 703-taxon data set also produced a relatively basal position for Passeriformes (Fig. 2e). Both the analysis using 1,000 random addition replicates and the analysis using the NJ tree as a starting tree placed Passeriformes as the sister group to all other birds. By the stepwise addition algorithm, the island of shortest trees (length = 37,393; rescaled consistency index [RC] = 0.028) was found on only one of the 1,000 random addition replicates. Branch swapping on the NJ tree produced a tree length of 37,425. These parsimony analyses suggest that trees probably exist that are substantially shorter than the ones recovered in these analyses. Outgroup-rooted parsimony analysis of amino acids (Fig. 2f) places oscines at the base of the tree with a lack of resolution for other higher-level relationships. The shortest tree for amino acids was found in 4 of the 300 replicates, so these searches were probably more efficient than the parsimony analysis of nucleotides. Fast jackknife replicates (jackknifing characters) recovered 327 of 702 ingroup nodes supported at >50%. Fast bootstrap replicates produced a similar number (325 of 702) of nodes supported at the 50% level. In both cases the relationships among the major bird lineages were not supported above the 50% level. In parsimony jackknife replicates, the monophyly of birds was supported at 73% and the monophyly of oscine passerines was supported at 60%. In addition, the monophyly of many other orders and families of birds was supported in >50% of replicates.

Taxa represented by only a single species tended to change positions most

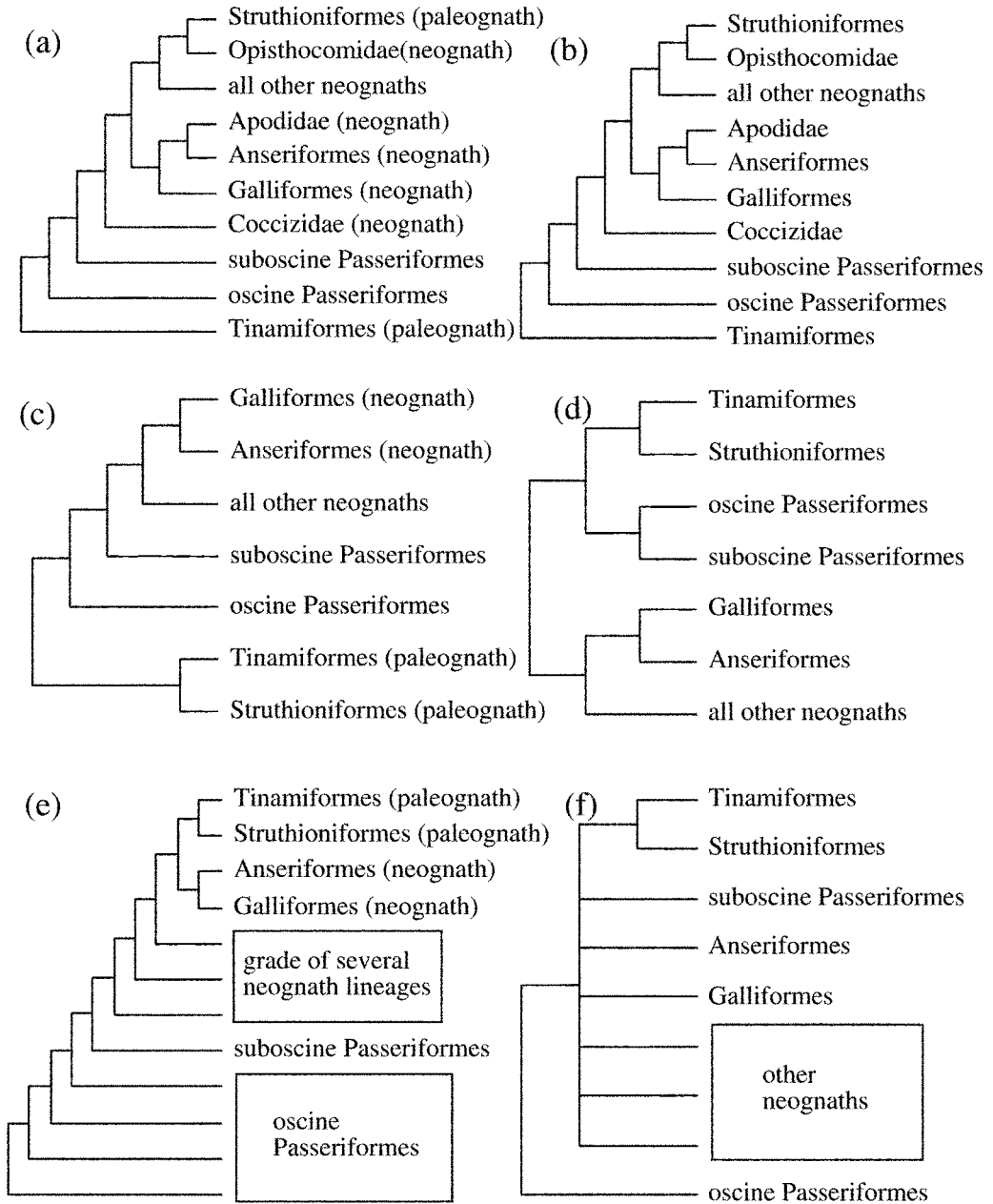


FIGURE 2. Relationships between major lineages of birds for various combinations of data sets and rooting when using NJ (a–d) and parsimony (e–f). Major lineages are Struthioniformes (ratites), Tinamiformes (tinamous), oscine Passeriformes, suboscine Passeriformes, Galliformes, Anseriformes, and all other neognaths. Lineages within other neognaths are shown when all other neognaths are not monophyletic under a particular analysis. (a) 916-taxon data set, outgroup rooting; (b) 916-taxon data set, midpoint rooting; (c) 703-Taxon data set, outgroup rooting; (d) 703-taxon data set, midpoint rooting; (e) parsimony analysis of 703-taxon data set (length = 37,393, RC = 0.028), outgroup rooting; (f) parsimony analysis of amino acids for 703-taxon data set (length = 6899, RC = 0.401), outgroup rooting. In the outgroup-rooted phylogenies, the outgroup is not shown to facilitate comparison of topologies. The outgroup taxa are *Clemmys marmorata*, *Kinosteron odoratus*, *Brachylophus fasciatus*, and *Alligator mississippiensis*.

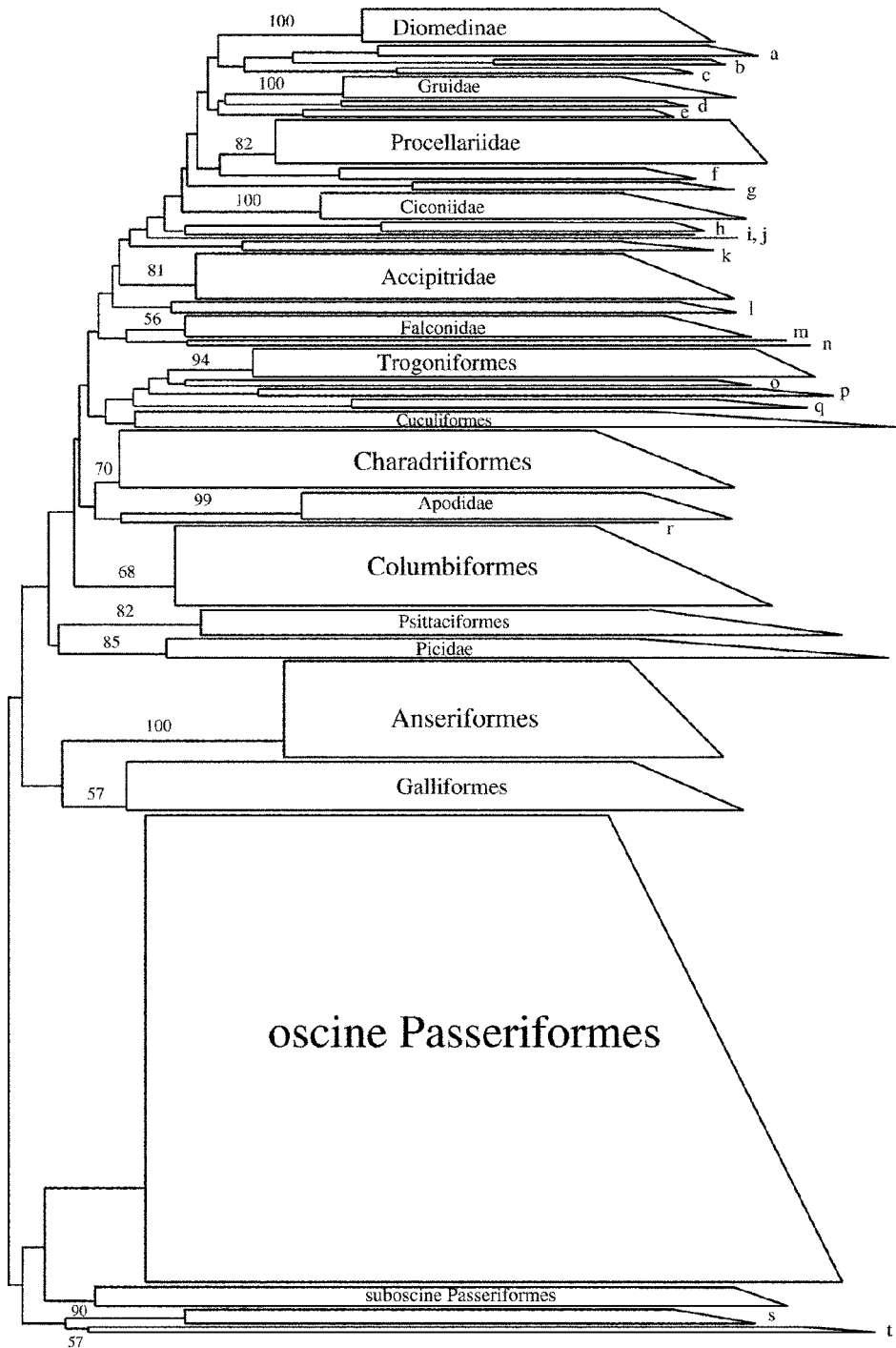


FIGURE 3. Tree topology from NJ of uncorrected p-distances using the 703-taxon data set with midpoint rooting. The length of the branches in each group indicates the shortest and longest branches of species in that group. Numbers on major nodes indicate groups that were supported in >50% of NJ bootstrap replicates (no groupings of major lineages were supported at >50%). The width of each group is proportional to the number of species from that group included in the analysis. Taxa not named on the tree are (a) Hydrobatidae (group 1), (b) Sphenisciformes (group 1), (c) Sphenisciformes (group 2), (d) Threskiornithidae, (e) Hydrobatidae (group 2), (f) Cathartidae, (g) Phaenicopteridae, (h) Aegothelidae, (i) *Pelecanus* (Pelecanidae), (j) *Balaeniceps* (Pelecanidae), (k) Caprimulgidae (Caprimulgiformes), (l) Nyctibiidae (Caprimulgiformes), (m) Opisthocomidae, (n) *Alcedo* (Alcedinidae), (o) Eurostopodidae + Steatornithidae (Caprimulgiformes), (p) Podargidae + Batrachostomidae (Caprimulgiformes), (q) Coliiformes, (r) Scopidae, (s) Struthioniformes (ratites), (t) Tinamiformes (tinamous).

TABLE 1. Monophyly of 21 select groups.

Data set	Percent monophyletic	Paraphyletic groups
916 taxa	76	Cuculiformes, Caprimulgiformes, Procellariiformes, Charadriiformes, Sphenisciformes
703 taxa	86	Caprimulgiformes Procellariiformes, Sphenisciformes

dramatically across NJ analyses. For example, *Alcedo*, the sole representative in this data set of Coraciiformes (153 species), grouped with the monotypic family Opisthocomidae in the 703-taxon combination and with Falconidae in the 916-taxon combination. Two of the oldest traditional orders among birds as determined by Sibley and Ahlquist (1990)—Cuculiformes (not shown) and Caprimulgiformes (Fig. 3)—tended to be paraphyletic in most analyses. This result suggests that long branches are relatively unstable in this analysis.

Given the ability of additional taxa to break up long branches, one might be surprised that the 916-taxon data set performed worse, in terms of recovering monophyly for orders and families of birds, than the 703-taxon data set. Besides finding a greater fraction of monophyletic groups using the 703-taxon data set, other results indicate that this data set may be performing better than the 916-taxon combination. Whereas the 703-taxon combination places paleognaths near the base of the tree and sister to all other birds (Fig. 2c), the 916-taxon combination (Fig. 2a) places ratites within other neognaths (in conflict with Sibley and Ahlquist, 1990; Harlid et al., 1998; and traditional classification, but see Mindell et al., 1999).

#### DISCUSSION

The inclusion of 916 birds in a phylogenetic analysis of *cyt b* represents one of the largest across-species analyses ever conducted for any group of organisms (see Chase et al., 1993, for an analysis of 499 seed plants; Soltis et al., 1997, for an analysis of 223 seed plants; and Källersjö et al., 1998, for an analysis of 2,538 plants). Sibley and Ahlquist (1990) present a tree for birds with 1,034 terminals, but they constructed this tree from an incomplete distance matrix, using only a fraction of all possible pairwise comparisons. I analyzed the *cyt b* data with NJ in two different data set combinations (916 and 703

taxa) with and without outgroup rooting. In all NJ analyses, Passeriformes were not embedded as a lineage within other neognaths, but rather were sister to all other neognaths (with oscines and suboscines often being paraphyletic) or were sister to paleognaths and this clade in turn was sister to all other birds (Fig. 2). In addition, all parsimony analyses placed Passeriformes at the base of the avian tree. The results of the present study are in general agreement with more taxonomically restricted parsimony analyses of *cyt b* (Harlid et al., 1998) and of complete mitochondrial sequences (Mindell et al., 1997, 1999). Thus, it is unlikely that the controversial results of Mindell et al. (1997, 1999) and Harlid et al. (1998) are merely the result of inadequate taxon sampling. In addition, some of the analyses indicated paraphyly for Passeriformes, with suboscines and oscines being separate from each other (also found by Mindell et al., 1999). Whether or not Passeriformes are paraphyletic, the divergence between suboscines and oscines is large and appears to predate most other radiations within birds (Fig. 3).

Although a position for Passeriformes outside of other neognaths was generally consistent across analyses, this result did not receive strong support (i.e., >50%) in either bootstrap or jackknife analyses (Fig. 3). The extent of homoplasy is high in this data set (RC < 0.03), and high homoplasy can reduce the power of bootstrap analyses (Zarkikh and Li, 1992, 1995). Although low bootstrap and jackknife support indicate that caution should be applied to the results, the consistency of the general position of Passeriformes (even though some of the details of the topology change) increases confidence that this result may be stable to the addition of more data. Jackknife and bootstrap analyses did not strongly support arrangements of major clades, but they did support the monophyly of many groups (including many orders and families). In parsimony jackknife replicates, 327 of 702 ingroup nodes (46.6%)

were supported at the 50% level. In the parsimony jackknife analysis of plant *rbcL* sequences by Källersjö et al. (1998), 1400 of 2537 nodes (55.2%) were supported at >50%.

Because monophyly for most of the major groups of birds was generally supported in all analyses, *cyt b* is likely to contain phylogenetic information relevant for deeper avian divergences. Although some have argued that *cyt b* may have difficulty recovering certain phylogenetic problems (Meyer, 1994; Moore and DeFilippis, 1997), this marker appears to recover several traditionally accepted deep relationships in most analyses (e.g., monophyly of oscine Passeriformes, sister relationship between Anseriformes and Galliformes).

Whereas most relationships inferred from analysis of *cyt b* were consistent with some previous assessments, taxa that appeared in unusual positions fell into two major categories: old, poorly sampled lineages, and lineages represented by only one or two species. Two old lineages of neognaths recognized by Sibley and Ahlquist (1990)—the cuckoos (traditional Cuculiformes) and the nightjars and allies (traditional Caprimulgiformes)—were paraphyletic in many analyses. This is probably a result of the relatively poor sample of these two groups and the fact that most of the branches in these groups are very long. However, in many analyses most of the species in these groups clustered together, with only a few lineages being responsible for the paraphyly. In addition, families or orders represented by single species tended to change positions readily and be involved in causing paraphyly in other larger groups (by inserting within these groups). These observations suggest that long branches have a strong potential to confound phylogenetic analyses and that phylogeny reconstruction might be improved by breaking long branches.

Rooting of molecular phylogenies can be an extremely problematic exercise (Wheeler, 1990; Smith, 1994). Birds have no close living relatives, making identification of the root of the avian phylogeny difficult. Although one could argue that rooting with alligator and other outgroups merely attracts the root to the longest branch (Wheeler, 1990), no rooting of any of the NJ topologies from the data set combination analyses (Fig. 2) resulted in Passeriformes being derived from other neognaths with paleognaths being outside

neognaths. Rerooting the parsimony trees (Figs. 2e, 2f) might provide such a result. However, adding more data of the same type as *cyt b* seems unlikely to do so in most analyses. Finding the root for an avian phylogeny apparently is still a problem because rooting changed slightly across some of the analyses. In particular, more data are needed to determine whether the avian tree should be rooted between tinamous and all other birds (Fig. 2a), between paleognaths and neognaths (Fig. 2c), between paleognaths plus Passeriformes and all other neognaths (Fig. 2d), or in some other way.

In addition to rooting problems, incorrect models of DNA substitution may result in the incorrect topology. Although changing the model of substitution to Kimura two-parameter (Kimura, 1980) or HKY 85 (Hasegawa et al., 1985) or incorporating rate heterogeneity did alter the arrangements of some nodes, the placement of Passeriformes and the other major lineages was largely unaffected by use of an alternative model. Parsimony trees also placed Passeriformes in a basal position, sister to all other birds. Templeton (1983) tests on parsimony trees also support the general conclusion that passerines are outside other neognaths, indicating the insensitivity of this result to general method of analysis.

Adding taxa to phylogenetic analysis can possibly aid in many phylogenetic problems (Hillis, 1996, 1998; Graybeal, 1998) but may not be a solution in every case (Kim, 1998; Poe and Swofford, 1999). If the basal position of Passeriformes among birds is spurious because of long branches, increasing the number of taxa from 7 to 916 does not appear to change this result. In addition, the possibility that Passeriformes are an unusually long branch seems unlikely. Although the branch leading to the root of oscine Passeriformes is relatively long, the total length of branches within Passeriformes is largely the same as other branches across other birds (Fig. 3). In fact, most of the longest branches in the tree involve non-passerines.

Another interesting result is that when taxa are well sampled, adding taxa such that sequences become uneven in length also appears to reduce phylogenetic congruence. In this case, missing data may have compromised phylogenetic analyses. In addition, some authors have suggested that a threshold sequence length is needed for certain

phylogenetic problems. Above this threshold, adding taxa can greatly improve accuracy, but below the threshold, accuracy is compromised even when a large number of taxa are added (Cummings et al., 1995; Mindell et al., 1997). For some phylogenetic problems, adding sequences increases accuracy more than does adding taxa (Poe and Swofford, 1999). The results presented here suggest that caution should be exercised in analyzing data sets in which taxa differ markedly in the number of missing characters.

In conclusion, analysis of 916 avian *cyt b* sequences reveals a position for Passeriformes outside other neognath birds, supporting the results of Mindell et al. (1997, 1999) and Harlid et al. (1998). This phylogenetic position for Passeriformes is in contrast to Sibley and Ahlquist's (1990) large DNA-DNA hybridization tree for birds and virtually all previous classification schemes. In addition, recent evidence from nuclear sequences places Passeriformes within other neognaths (Groth and Barrowclough, 1999). The possibility that nuclear and mitochondrial genes are in conflict with regard to the position of Passeriformes within birds merits further investigation.

#### ACKNOWLEDGMENTS

L. Bohs, D. Drown, D. Mindell, K. Omland, R. Olmstead, and an anonymous reviewer provided valuable comments on this manuscript. I was supported by NSF-CAREER award DEB-9703003 to D. Clayton during the writing of this manuscript.

#### REFERENCES

- CHASE, M. W., AND 41 OTHER AUTHORS. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastic gene *rbcL*. *Ann. Mo. Bot. Gard.* 80:528–580.
- CUMMINGS, M. P., S. P. OTTO, AND J. WAKELEY. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* 12:814–822.
- DEL HOYO, J., A. ELLIOTT, AND J. SARGATAL (eds.). 1992. Handbook of the birds of the world, volume 1. Lynx Edicions, Barcelona.
- FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, AND A. G. KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12:99–124.
- FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility will be positively misleading. *Syst. Zool.* 27:401–410.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47:9–17.
- GROTH, J. G., AND G. F. BARROWCLOUGH. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. *Mol. Phylogenet. Evol.* 12:115–123.
- HÄRLID, A., A. JANKE, AND U. ARNASON. 1998. The complete mitochondrial genome of *Rhea americana* and early avian divergences. *J. Mol. Evol.* 46:669–679.
- HASEGAWA, M., H. KISHINO, AND T. A. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- HILLIS, D. M. 1996. Inferring complex phylogenies. *Nature* 383:130–131.
- HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* 47:3–8.
- HOWARD, R., AND A. MOORE. 1980. A Complete Checklist of the Birds of the World. Academic Press, London.
- HUELSENBECK, J. P. AND D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42:247–264.
- KÄLLERSJÖ, M., J. S. FARRIS, M. W. CHASE, B. BREMER, M. F. FAY, C. J. HUMPHRIES, G. PETERSON, O. SEBERG, AND K. BREMER. 1998. Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Systematics and Evolution* 213:259–287.
- KIM, J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. *Syst. Biol.* 47:43–60.
- KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–120.
- MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade: Analysis of phylogeny and character evolution, version 3.04. Sinauer Associates, Sunderland, Massachusetts.
- MEYER, A. 1994. Shortcomings of the cytochrome *b* gene as a molecular marker. *Trends Ecol. Evol.* 9:278–280.
- MINDELL, D. P., M. D. SORENSON, D. E. DIMCHEFF, M. HASEGAWA, J. C. AST, AND T. YURI. 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Syst. Biol.* 48:138–152.
- MINDELL, D. P., M. D. SORENSON, C. J. HUDDLESTON, H. C. MIRANDA, JR., A. KNIGHT, S. J. SAWCHUK, AND T. YURI. 1997. Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. Pages 213–247 in *Avian molecular evolution and systematics* (D. P. Mindell, ed.). Academic Press, San Diego, CA.
- MOORE, W. S., AND V. R. DEFILIPPIS. 1997. The window of taxonomic resolution for phylogenies based on mitochondrial cytochrome *b*. Pages 83–113 in *Avian molecular evolution and systematics* (D. P. Mindell, ed.). Academic Press, San Diego, CA.
- MORONY, J. J., W. J. BOCK, AND J. FARRAND. 1975. Reference list of the birds of the world. American Museum of Natural History, New York.
- POE, S. 1998. Sensitivity of phylogeny estimation to taxonomic sampling. *Syst. Biol.* 47:18–31.
- POE, S., AND D. L. SWOFFORD. 1999. Taxon sampling revisited. *Nature* 398:299–300.
- SIBLEY, C. G., AND J. E. AHLQUIST. 1990. Phylogeny and classification of birds: A study in molecular evolution. Yale Univ. Press, New Haven, CT.

- SMITH, A. B. 1994. Rooting molecular trees: Problems and strategies. *Biol. J. Linn. Soc.* 51:279–292.
- SOLTIS, D. E., AND 15 OTHER AUTHORS. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. Mo. Bot. Gard.* 84:1–49.
- SWOFFORD, D. L. 1999. PAUP\*: Phylogenetic analysis using parsimony, version 4.0. Sinauer Associates, Sunderland, Massachusetts.
- TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37:221–244.

- WHEELER, W. C. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6:363–368.
- ZARKIKH, A., AND W.-H. LI. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Mol. Biol. Evol.* 9:1119–1147.
- ZARKIKH, A. AND W.-H. LI. 1995. Estimation of confidence in phylogeny: The complete-and-partial bootstrap technique. *Mol. Phylog. Evol.* 4:44–63.

Received 10 May 1999; accepted 27 December 1999  
Associate Editor: R. Olmstead

*Syst. Biol.* 50(1):136–141, 2001

## Useful Characters in Gastropod Phylogeny: Soft Information or Hard Facts?

CHRISTOFFER SCHANDER<sup>1</sup> AND PER SUNDBERG

Göteborg University, Department of Zoology, P.O. Box 463, SE-405 30 Göteborg, Sweden;  
E-mail: c.schander@zool.gu.se, p.sundberg@zool.gu.se

Traditionally, gastropod classification and phylogeny are often based on hardshell structures, although other anatomical characters and molecular characters are increasingly included in systematic studies (Ponder, 1988, and references therein; Bieler, 1992), together with ultrastructural shell characters (Falniowski and Szarowska, 1995). Some authors have expressed a certain scepticism for using shell characters in cladistic analyses (e.g., Kool, 1993; Robertson, 1996; Wise, 1996), whereas others have based their analyses entirely on such characters (e.g., Michaux, 1989), which in turn has been questioned by Emberton (1995). Some authors (e.g., Robertson, 1978, 1996; Davis, 1979; Kool, 1993), however, have stated that shell characters should be avoided, being more prone to selective processes and hence more homoplastic than other characters. According to those authors, selection will increase the risk that an empirically observed similarity is not the result of common ancestry but rather of parallel or convergent evolution, which may lead to erroneous conclusions.

Although shell shape, sculpture, and coloration are the traditional characters in

molluscan systematics (Bieler, 1992), other information has also been used. For example, Cuvier (1797) established the taxa *Gastropoda* and *Cephalopoda* on the basis of how they moved. Lamarck (1799) opposed this by arguing that in most cases only the shell was known and classification should, therefore, be based on conchological characters; indeed, the name *Bivalvia* stems from the emphasis on shell characters. Not until the middle and late nineteenth century (e.g., Lovén, 1840, 1849; Herdman, 1890; Pelseener, 1899) did comparative anatomy of soft parts, histology, and embryology gain acceptance in classification. Milne-Edwards (1848) divided the gastropods into *Prosobranchia*, *Opisthobranchia*, and *Pulmonata* mainly on the basis of the organization of the respiratory system. Spengel (1881) divided the same animals into *Streptoneura* and *Euthyneura*, according to differences in the nervous system. Mörch (1865) argued that different characters were useful at different taxonomic levels; for example, characters of the heart were better for diagnosing higher categories, whereas characters from the reproductive system could be used at lower taxonomic levels. This is a remnant from Cuvier's distinction of important and less important characters (Geoffroy Saint-Hilaire and Cuvier, 1795), in which characters essential for survival were used

<sup>1</sup>Present address (and address for correspondence): Arctic Station, University of Copenhagen, Box 504, DK-3953 Qeqertarsuaq, Greenland.