Why One-Kilobase Sequences from Mitochondrial DNA Fail to Solve the Hoatzin Phylogenetic Enigma

JOHN C. AVISE, * 1 WILLIAM S. NELSON, * AND CHARLES G. SIBLEY†

* Department of Genetics, University of Georgia, Athens, Georgia 30602; and 1433 Woodley Place, Santa Rosa, California 95409

Received October 22, 1993; revised February 14, 1994

INTRODUCTION

The Hoatzin (Opisthocomus hoazin) is a medium-sized, long-tailed bird with a ragged crest, brownish plumage, and an area of bright blue skin on the face. It is similar to the Guira Cuckoo (Guira guira) in color and plumage pattern, but is much larger. The Hoatzin lives in tropical riparian habitats in northern and central South America and feeds mainly on leaves. Digestion begins in a large crop by microbial foregut fermentation, the only known example of this process in birds, which occurs also in several groups of mammals. Hoatzins form large flocks when not breeding; nesting is communal, as in the anis (Crotophagidae). Groups of two to seven birds build a single nest over water, but only one pair breeds. The others, usually the offspring of previous years, help to incubate the eggs and feed the young. The young have two claws on each wing which are used when climbing near the nest before fledging. When alarmed, the young may drop into the water, then climb back to the nest when the danger is past. These unusual anatomical and behavioral characteristics have made the Hoatzin a taxonomic enigma since its discovery over 200 years ago (color plate in Sibley and Ahlquist, 1973; review in Sibley and Ahlquist, 1990). Indeed, Bock (1992) suggested that “Perhaps of all problems facing avian macrosystematics, the affinities of the hoatzin... is the most vexing.”

Most classifications have placed the Hoatzin in the Galliformes (ducks, Gees), the Galliformes (cuckoos, anis, etc.), or a monotypic order, Opisthocomiformes. From electrophoretic comparisons of proteins (1973) and DNA–DNA hybridization comparisons (1990), Sibley and Ahlquist concluded that the Hoatzin is a member of the Galliform radiation, most closely related to the neotropical Guira Cuckoo, the anis, and the roadrunners (Geococcyx). This assignment has been questioned by Bock (1992), who doubted that the anisodactyl (three toes forward, one back) foot structure of the Hoatzin could have been derived from the zygodactyl foot (two toes forward, two back) of the typical cuckoos. However, the Hoatzin may have shared a common ancestor with the cuckoos and not be “de-
scended" from any living cuckoo. Thus, the phylo-
genic relationships of the Hoatzin remain contro-
versial.

In this paper, we report the results of comparisons
of nucleotide and amino acid sequences from a protein-
coding mitochondrial (mt) gene (cytochrome b). In ad-
dition to the phylogenetic relationships of the Hoatzin,
this study permits an assessment of the window of res-
olution by mtDNA sequences in avian evolution. Re-
striction sites and sequences in the rapidly evolving
mtDNA genome are well-suited to comparisons among
closely related taxa (populations, species, genera)
(Avise et al., 1987; Avise and Ball, 1991), but recent
sequence analyses of the mitochondrial cytochrome b
gene have shown that conservative sites within the
molecule also can retain phylogenetic signal over
longer periods of time (Bowen et al., 1993; Irwin et al.,
1991), extending to deep intraordinal levels in at least
some avian groups (Kornegay et al., 1993).

This study inaugurates a contemplated series in
which we hope to use direct nucleotide sequence data
to examine unclear or controversial issues in avian
phylogeny resulting from the qualitatively different
method of DNA–DNA hybridization (Sibley and Ahl-
quist, 1990).

MATERIALS AND METHODS

Laboratory Procedures and Data Analyses

The DNA samples were prepared as sonicated
("sheared") preparations of "single-copy" DNA, which
includes the mtDNA fraction. The same samples were
used in DNA–DNA hybridization studies conducted in
Sibley's laboratory at Yale University. Mitochondrial
DNA fragments containing cytochrome b gene se-
quencies were amplified via the PCR method, using
primers numbered 3, 5, 7, 8, and 10 as described in
Kornegay et al. (1993). The amplified fragments were
sequenced directly, either in the Avise laboratory (via
dideoxy chain termination using T7 DNA polymerase
and 32P labeling) and/or by the Molecular Genetics In-
strumentation Facility at the University of Georgia
(fluorescent-dye sequencing). Heavy and light strands
were sequenced for each specimen.

All sequences were unambiguously aligned by eye,
without gaps (Appendix). For most species, scored se-
quencies began at position 14,993 [numbered as in the
complete sequence of the mtDNA genome of the Do-
mestic Fowl (Gallus gallus)] (Desjardins and Morais,
1990) and extended to position 15,953, thus covering
961 bp or 84% of the cytochrome b gene.

Sequences were coded in four formats: (a) original
nucleotide sequences; (b) purines versus pyrimidines,
such that only transversional substitutions were re-
corded; (c) first and second positions of codons; and
(d) amino acid sequences inferred from the mtDNA
code (Li and Graur, 1991), such that only replacement
substitutions were recorded. These codings provide dif-
f erent, but overlapping, classes of information. For ex-
ample, many transversions and a few third-position
changes in codons produce amino acid substitutions,
and some first- and second-position changes in codons
involve transversions and/or amino acid substitutions.

Three phylogenetic analyses were applied to each of
the four databases: (1) maximum parsimony, using
heuristic searches and bootstrapping in PAUP (Swoff-
oford, 1990); (2) UPGMA of genetic distances (Sneath
and Sokal, 1973); and (3) neighbor-joining (N-J) of
genetic distances (Saitou and Nei, 1987). The matrices
of genetic distances in the latter two clustering pro-
duced were either the observed differences between rel-
levant sequences for each coding method ((a–d) above)
or inferred genetic distances between the nucleotide
sequences after correction for presumed multiple sub-
stitutions at a site, using Kimura's (1980) two-
parameter model and varying assumptions about tran-
sition/transversion ratios. The difference and distance
calculations were performed with PAUP and PHYLIP,
respectively, and the clustering algorithms were im-
plemented by PHYLIP (Felsenstein, 1991).

Taxa Examined

New cytochrome b sequences were obtained from 12
individuals representing 10 species, as follows: Hoa-
tzin (3 samples), Cuculiformes: Groove-billed Ani (Cro-
tophaga sulcirostris), Guira Cuckoo (only 715 bp
scored), Pallid Cuckoo (Cuculus pallidus), Squirrel
Cuckoo (Piaya cayana), Chestnut-breasted Cuckoo
(Phaenicophaeus curvirostris), Yellow-billed Cuckoo
(Coccyzus americanus), Black-billed Cuckoo (Coccyzus
erythrophalus), Galliformes: California Quail (Cal-
lipepla californica) and Silver Pheasant (Lophura nycti-
themera). Also included were published sequences of
the homologous 961-bp cytochrome b segment from
nine galliform species (Kornegay et al., 1993): Domes-
tic Fowl ("Chicken"), Silver Pheasant, Chukar Par-
tridge (Alectoris chukar), Japanese (Common) Quail
(Coturnix coturnix), Turkey (Meleagris gallopavo), In-
dian Peafowl (Pavo cristatus), Gambel's Quail (Calli-
ipepla gambelii), Helmeted Guineafowl (Numida me-
leagris), and Plain Chachalaca (Ortalis vetula). The
chachalacas are usually placed in the Galliformes, but
Sibley and Ahlquist (1990) assigned them, and other
cracids, to the order Craciformes, the closest relatives
of the Galliformes (Table 1).

A search of GenBank yielded partial cytochrome b
sequences from several passerines (Passeriformes), a
woodpecker (Piciformes), and a duck (Anseriformes),
which exhibited at least 900 bp of overlap with homolo-
gous cytochrome b positions from the Hoatzin se-
quencies, and they were included in some of the analy-
yses. Finally, a cytochrome b sequence from a mammal
(Homo sapiens; see Irwin et al., 1991) also was com-
pared against the birds.
TABLE 1

Higher-Level Classifications for the Galliformes, Craciformes, and Cuculiformes as Presented by Sibley and Ahlquist (1990), Based on DNA–DNA Hybridization Studies

<table>
<thead>
<tr>
<th>Taxonomy</th>
<th>Galliformes</th>
<th>Craciformes</th>
<th>Cuculiformes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parvorder Phasianida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superfamily Phasianoidea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Phasianidae: Old World pheasants, grouse, turkeys, and partridges</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superfamily Numidoidae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Numididae: Guineafowl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parvorder Odontophorida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Odontophoridae: New World quail</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Cracidae: chachalacas and allies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuculiformes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infraorder Cuculiformes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parvorder Cuculida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superfamily Cuculoidea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Cuculidae: Old World cuckooos</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superfamily Centropodoidae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Centropodidae: Coucals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parvorder Coccyzida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Coccyzida: American cuckooos</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infraorder Crotaphagida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parvorder Opisthocomida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Opisthocomidae: Hoatzin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parvorder Crotaphagida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Crotaphagidae: Anis and Guira Cuckoo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parvorder Neomorphida</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Neomorphidae: Roadrunners, ground cuckooos</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Species represented in the current cytochrome b sequence analyses are indicated in boldface.

RESULTS

Among the 961 nucleotide positions typically assayed for the 18 species (galliforms, cuculiforms, Plain Chachalaca, and Hoatzin), 600 exhibited variation, and 395 of these were potentially informative in a cladistic sense, i.e., two or more taxa shared a variant character state. Corresponding counts for the three alternative formats were purines vs pyrimidines, 961 positions, 294 variable and 213 potentially informative; first and second positions of codons, 620 positions, 206 variable and 124 informative; amino acid sequences, 320 positions, 139 variable, 85 informative.

 Parsimony Analyses

Figure 1 summarizes the results of maximum parsimony analyses based on the cytochrome b data coded as purines vs pyrimidines and as amino acid sequences. Groups (or clades) consistently showing strong (>70%) bootstrap support in these analyses were: (1) the two samples of the Silver Pheasant; (2) the three samples of the Hoatzin; (3) the two New World quail (Gambel's and California) vs the other galliforms assayed; (4) the Galliformes (not including the Plain Chachalaca); (5) the Groove-billed Ani and Guira Cuckoo vs the other Cuculiformes; and (6) a group composed of the Squirrel Cuckoo, Yellow-billed Cuckoo, Black-billed Cuckoo, and Chestnut-breasted Cuckoo. These same groups were also recognized and bootstrap support at >70% in parsimony analyses based on total cytochrome b sequence data, except for clade (6), and parsimony analyses based only on the first and second positions of codons. The only other putative clade consistently recognized by the latter two approaches consisted of the seven assayed cuculiforms, at 81 and 96% bootstrap support, respectively.

The clustering of the Chestnut-breasted Cuckoo, a native of southeast Asia, with the three New World cuckoos in group (6) seems surprising because the New World Coccyzus and Piaya species are distantly related to Cuculus, and other Old World cuckoos, at \( \Delta T_{50} \) 14.5 in DNA hybridization comparisons using UPGMA (Sibley and Ahlquist, 1990). However, DNA hybridization measured an even greater gap between the two New World cuckoos in group (5) and the New World cuckoos in group (6) at \( \Delta T_{50} \) 17.6. Thus, there may be a congruence between the mtDNA data and the DNA hybridization data, although the Chestnut-breasted Cuckoo might have been expected to cluster with Cuculus, rather than with Piaya and Coccyzus. Because the Chestnut-breasted Cuckoo was not used in the DNA hybridization comparisons, this question remains open.

Two of the four classes of parsimony analysis tended to group the Hoatzin with the cuculiforms rather than with the galliforms, with levels of bootstrap support 68% for original sequence data and 73% for amino acid sequences. However, a majority-rule consensus tree based on transversions left unresolved a phylogenetic “trichotomy” for the Hoatzin, Galliformes, and Cuculiformes, and a consensus tree based on first and second codon positions placed the Hoatzin just outside a galliform–cuculiform clade that was supported at the 56% bootstrap level.

Phenetic Analyses

Cluster analyses of the cytochrome b distance matrices are summarized in Fig. 2. Again, the six “clades” described above were recognized in the eight phenetic analyses attempted [except that group 4 (Galliformes) included the Plain Chachalaca]. The reason for this level of cluster stability is, no doubt, that the internodal distances involved were relatively large, such that both the distance-based and the parsimony approaches “captured” the same putative clades. For example, in the N-J reconstruction of Fig. 2, branch lengths leading to clusters common to all trees averaged 0.052 (range 0.021–0.121), whereas the branch
FIG. 1. Examples of maximum parsimony analyses of cytochrome b sequences in Hoatzins and putative allies, based on data coded as (left) purines vs pyrimidines and (right) amino acid sequences. Trees were rooted to a supposed outgroup, the Plain Titmouse (Passeriformes; \textit{Parus inornatus}). Each tree is a 50\% majority-rule consensus representation, with levels of bootstrap support across 100 replicates indicated (only bootstrap values greater than 70\% are shown). Shaded boxes surrounding tree branches (left and right) or taxa (center) highlight "clades" that received greater than 70\% bootstrap support in most or all classes of parsimony analysis attempted, including also those based on total sequence data and on first and second positions of codons.

\textbf{UPGMA}

\textbf{N-J}

\textbf{sequence divergence}
lengths leading to the remaining (unstable) clusters averaged only 0.006 (range 0.001–0.018).

In Fig. 3, the Hoatzin’s position relative to the cuculiforms and galliforms varied according to the database used. For the cytochrome b sequence, the Hoatzin falls just outside a cluster joining the Galliformes and Cuculiformes, and the same is true for the analysis based only on first and second codon positions. However, data coded as purines vs pyrimidines group the Hoatzin closest to the cuculiforms, and as amino acid sequences, closest to the galliforms. Thus, depending on the class of sequence characters, all three possible arrangements were observed among lineages leading to the Hoatzin, Cuculiformes, and Galliformes. However, the analyses were consistent in the close positions of the two relevant nodes relative to the ranges of genetic differences observed within the Cuculiformes or Galliformes (Fig. 3). Thus, the most conservative conclusion is that a trichotomy among the Hoatzin, Cuculiformes, and Galliformes remains unresolved from the available cytochrome b sequences.

This lack of resolution is supported by the data in Table 2. Species representing the Passeriformes, Piciformes, and Anseriformes, orders not normally considered to be closely related to the Hoatzin or to one another, have levels of cytochrome b divergence from the Hoatzin virtually identical to mean distances between the Hoatzin and the Cuculiformes and Galliformes! As

![Diagram of phylogenetic relationships]

**FIG. 3.** Simplified versions of UPGMA dendrograms (all drawn to respective scales) summarizing cytochrome b sequence relationships within and among Cuculiformes, Galliformes (arbitrarily including the Plain Chachalaca), and the Hoatzins. Shown are outcomes based on difference matrices derived from cytochrome b data coded in the four different formats indicated. Shaded boxes indicate the observed ranges of respective genetic distances connecting UPGMA nodes within the Galliformes (10 assayed species) and within the Cuculiformes (7 assayed species). Also shown are the respective mean genetic distances at which these two orders join to one another or the Hoatzin.

<table>
<thead>
<tr>
<th>Hoatzins vs</th>
<th>Nucleotide sequences</th>
<th>Transversions only</th>
<th>1st and 2nd codon positions</th>
<th>Amino acid sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuculiformes (7 species)</td>
<td>0.218 (0.204–0.242)</td>
<td>0.103 (0.098–0.113)</td>
<td>0.107 (0.086–0.124)</td>
<td>0.168 (0.132–0.197)</td>
</tr>
<tr>
<td>Galliformes (10 species)</td>
<td>0.224 (0.213–0.245)</td>
<td>0.105 (0.095–0.117)</td>
<td>0.101 (0.091–0.113)</td>
<td>0.162 (0.149–0.190)</td>
</tr>
<tr>
<td>Passeriformes (11 species)</td>
<td>0.222 (0.209–0.235)</td>
<td>0.104 (0.095–0.112)</td>
<td>0.098 (0.088–0.111)</td>
<td>0.157 (0.130–0.182)</td>
</tr>
<tr>
<td>Piciformes (1 species)</td>
<td>0.205</td>
<td>0.099</td>
<td>0.082</td>
<td>0.138</td>
</tr>
<tr>
<td>Anseriformes (1 species)</td>
<td>0.215</td>
<td>0.097</td>
<td>0.092</td>
<td>0.138</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>0.280</td>
<td>0.171</td>
<td>0.191</td>
<td>0.265</td>
</tr>
</tbody>
</table>

*Note.* Also shown, for perspective, are the corresponding cytochrome b genetic distances to a mammal (humans).

**FIG. 2.** Examples of phenetic analyses of cytochrome b gene sequences in Hoatzins and putative allies. (Left) UPGMA analysis of genetic distances based on Kimura’s two-parameter method with assumed transition:transversion ratio of 2:1. The dendrogram is drawn such that the axis at the bottom refers to the joint distances (rather than individual branch lengths) between taxa or nodes. (Right) Neighbor-joining analysis based on the same distance matrix, with individual branch lengths noted. Shaded boxes surrounding tree branches (left and right) or taxa (center) highlight “clades” that appeared consistently in all of the eight phenetic analyses attempted, including also UPGMA and N-J as applied to difference matrices based on transversions only, first and second positions of codons, or amino acid substitutions.
elaborated below, these results raise the possibility that the four classes of sequence data have approached saturation with respect to attainable genetic differences, such that the phylogenetic position of the Hoatzin falls outside the window of resolution for the cytochrome *b* gene.

**DISCUSSION**

All algorithms for phylogenetic reconstruction involve assumptions about how evolution proceeds [see extensive discussions in Hillis and Moritz (1990) and Miyamoto and Cracraft (1991)]. For example, UPGMA assumes equal rates of evolution in all branches, N-J in effect relaxes this assumption by adjusting distances on the basis of mean divergence from all other nodes, and maximum parsimony assumes that evolution has proceeded along pathways involving the smallest total numbers of character-state transformations. Further, as elaborated by Bull et al. (1993), various approaches for partitioning and combining data sets often produce different estimates of phylogeny, the net result being that "Controversy exists over which of these approaches is superior, and advocates of different views openly acknowledge that the matter is unresolved." We claim no special insight into the appropriateness or validity of alternative phylogenetic analysis procedures, but prefer instead to examine a philosophically diverse array of approaches, compare results, and adopt a conservative stance that outcomes robust to alternative analyses and databases should be less controversial than those that are strongly analysis- or data-dependent. We also suspect that if this perspective were adopted more widely, systematists would view with (appropriately) greater skepticism the reliability of trees generated from particular tree-building methods for sequence data.

**Phylogenetic Signals and Noise**

A conservative (relatively unresolved) phylogenetic tree representing the phylogeny of the Hoatzin, cuckoos, and galliforms based on cytochrome *b* sequences is presented in Fig. 4. The data clearly contain phylogenetic information. Thus, regardless of how data were coded, similar or identical phylogenies were produced from generic to ordinal levels, and most or all analyses revealed clades that are supported by other evidence, e.g., DNA hybridization (Table 1). For example, the Groove-billed Ani and the Guira Cuckoo (Crotaphagidae) grouped together, as did all of the Cuculiformes (except the Hoatzin). Also, the two New World Quail (Odontophoridae) formed a clade, as did all of the Galliformes. Thus, phylogenetic signal clearly is present in the data.

The separation of New World quail from the other galliforms is of particular interest. In most classifications of the past century, the New World quail have been included in the Phasianidae with the pheasants, grouse, Old World quail, etc. However, DNA–DNA hybridization comparisons indicated that the New World quail are the sister group of the Phasianidae + Numididae (Sibley and Ahlquist, 1985, 1990). In addition to the present study, Kornegay et al. (1993) also concluded from cytochrome *b* sequence comparisons that the Numididae formed a clade with the Phasianidae, and the New World quail fell outside this group.

On the other hand, the conservative summary of the phylogeny depicted in Fig. 4 fails to resolve many of the lineages examined in this study. For example, the positions of the Pallid Cuckoo in the cuculiform radiation and of the Turkey, Peafowl, Chicken, and Guineafowl in the galliform radiation appeared to change depending on the database and method of analysis. When internodal differences are less than the mean differences between conspecific samples (e.g., Fig. 2), in principle it seems unlikely that polytomies can be resolved by individual loci because, at the relevant divergence times, polymorphic gene-tree lineages in the common ancestor could have been randomly distributed to descendant taxa (Nei, 1987; Avise, 1994).
Discrepancies between a “gene tree” and a “species tree” can arise even in the absence of sampling error caused by a finite number of characters at the assayed locus. In the present study, observed genetic differences among conspecific Hoatzins and Silver Pheasants were greater in magnitude than several of the inferred differences between internal nodes in the cuculiform and galliform reconstructions (Fig. 2).

Ceiling Effects and Limitations on Phylogenetic Resolving Power

Although this study failed to resolve relationships of the Hoatzin (beyond the conclusion that this species does not appear to be deeply imbedded within either the galliform or cuculiform clades), the analyses do provide a frame-of-reference for possible ceiling effects across ancient nodes. Through time, are plateau levels of differentiation approached as particular classes of cytochrome \( b \) sequence become saturated with acceptable molecular changes? If so, genetic distances at deeper nodes might appear artificially bunched at magnitudes near these achievable ceilings. Alternatively, in the absence of ceiling effects, phylogenetic radiations within narrow windows of time could also generate a clustering of nodal depths. In their study of cytochrome \( b \) sequences, Kornegay et al. (1993) attributed the lack of resolution of the major groups of Galliforms to a relatively rapid radiation of lineages at least 30 million years ago (mya).

The data summarized in Table 2 suggest that ceiling regions for cytochrome \( b \) may have been approached, or reached, at the following levels: total sequences, ca. 20% sequence differences; transversional substitutions, ca. 10%; first and second positions of codons, ca. 8–10%; and amino acid sequences, ca. 13–15%. These values may set approximate upper bounds on windows of phylogenetic resolution for avian cytochrome \( b \) sequences. On the other hand, these ceilings cannot be absolute because the cytochrome \( b \) differences between mammals and birds are somewhat greater (by 1.3- to 1.8-fold) than the largest mean distances among any of the assayed bird groups (Table 2).

In previous analyses of galliform cytochrome \( b \) sequences, Kornegay et al. (1993) separated the data into transitional and transversional differences, each considered separately at the first, second, and third codon positions. For all six categories, the slopes of the regressions relating sequence differences to time were extremely shallow beyond 10–20 mya, such that considerably different times of evolutionary separation would be effectively indistinguishable with current levels of sequencing effort. In the present study, data coded in each of the four formats also yielded divergence curves whose slopes were shallow beyond 10–20 mya (Fig. 5). The deeper separations in the Galliformes and Cuculiformes fall well into these “plateau” regions of genetic differentiation.

Thus, the lack of resolution of Hoatzin relationships with current sequence data is probably attributable to the considerable antiquity of the species, in conjunction with inevitable sampling errors in estimation of sequence divergence and with the shallow slopes in the regressions relating cytochrome \( b \) sequence differences to time. To exemplify the seriousness of the problem, consider the prospects for distinguishing branching orders for three extant taxa whose two nodal separations truly date to 10 million years apart (e.g., 60 mya vs 50 mya). Under the linear portion of the regression presented in the top graph of Fig. 5, only about 0.8%
sequence differences are expected to have accumulated in the cytochrome \( b \) gene over this time interval. With 960 such positions assayed, as in current study, this translates into about 7 nucleotide substitutions that could potentially distinguish between the nodes. This is a best-case scenario, because it assumes a perfectly metronomic and linear clock. ignores the complication of random lineage sorting from a polymorphic ancestor, and assumes that the informative characters can be properly recognized from among the much larger pool of variable character states (600 in our case) observed in the overall study. Furthermore, 10 million years of separation time between nodes is quite large by the standards of avian evolution. Results show the exceptional difficulties to be expected in using particular gene sequences to distinguish ancient but closely spaced nodes. Similar calculations can readily be done for the other data treatments in Fig. 5.

It is known that certain positions in the cytochrome \( b \) sequence are more conserved than others, due in part to functional constraints on protein domains (Edwards et al., 1991). The rationale for using alternative codings of the cytochrome \( b \) data is that conservative sequence characters (replacement substitutions, transversions, slowly evolving first and second positions of codons) might be especially useful for resolving the older nodes. However, this appears to be an oversimplification, because phenetic and parsimony analyses of alternatively coded data often yielded similar phylogenetic conclusions at both recent and intermediate levels of divergence. One explanation is suggested by Fig. 5, which implies that all four classes of data yield genetic differences that accumulate (albeit very gradually) through time. Another possible explanation is that competing factors are involved; although raw sequence data are likely to be afflicted by a higher mean level of homoplaspy per character (due to inclusion of silent changes, third-position changes, and transitions), the phylogenetic “noise” thereby introduced may be offset by the increased information content in data sets of larger numbers of variable characters, some of which may contribute to phylogenetic signal.

Thus, the advantage of using small numbers of conservative characters may be cancelled by the imprecise information from a larger class of variable character states. Indeed, this tradeoff between characters assumed to have special properties versus the larger numbers of less informative characters appears to be at the center of the debate between cladistics and phenetics, as well as between proponents of direct nucleotide sequencing of particular genes versus those who advocate the DNA–DNA hybridization method.

Our initial experience with cytochrome \( b \) sequences in birds suggests at least two modifications for future studies. First, such studies might profitably focus on problems involving more recent levels of divergence, e.g., intraordinal rather than interordinal. Second, if older nodes are to be resolved, much longer sequences from conservative genes or character classes should be considered. Because cytochrome \( b \) is thought to be among the more conservative of mitochondrial genes (but see Kumazawa and Nishida, 1993), such information may have to come from nuclear assays.

ACKNOWLEDGMENTS

Work was supported by funds provided by the University of Georgia and by the NSF.

REFERENCES

Nucleotide sequences across 320 codons of the cytochrome b gene (beginning adjacent to position 14,993) for the 12 newly assayed species representing the Cuculiformes, Galliformes, and Hoatzin. Dots indicate identity to the squirrel cuckoo sequence at the top.