Evolution, 58(2), 2004, pp. 404–415

BIRDS IN A BUSH: FIVE GENES INDICATE EXPLOSIVE EVOLUTION OF AVIAN ORDERS

STEVEN POE[1] AND ALISON L. CHUBB[2]
[1]*Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico 87131*
E-mail: anolis@unm.edu
[2]*Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Sciences Building, University of California, Berkeley, California 94720*
E-mail: alisonc@socrates.berkeley.edu

Abstract.—All recent studies of bird phylogeny have produced poorly resolved relationships among the orders of Neoaves, the lineage that includes most modern birds. This ''bush'' result suggests the possibility of an explosive and potentially unresolvable evolutionary radiation. However, simultaneous radiations of multiple lineages are thought to be rare or nonexistent in nature and difficult to corroborate empirically because lack of phylogenetic resolution can also be caused by analytical artifacts. Here we examine the predictions of the explosive radiation hypothesis for five independent genetic datasets for Neoaves. We propose a methodology for testing for polytomies of evolutionary lineages, perform likelihood-ratio tests to compare trees with zero-length branches to more resolved trees, compare topologies between independent gene trees, and propose a power test for the SOWH test. The evidence of (1) extremely short (in some cases zero-length) branches for interordinal relationships across independent gene trees and (2) topological incongruence among gene trees suggests that the bird tree includes essentially simultaneous radiation of multiple lineages. This result explains why a robust phylogeny of birds has not been produced despite much effort on the part of avian systematists.

Key words.—Avian evolution, birds, Neoaves, phylogeny, polytomy, power test, rapid radiation.

Received January 15, 2003. Accepted September 25, 2003.

The relationships among most orders of modern birds have been notoriously difficult to resolve. This lack of resolution is typified, historically, by numerous conflicting evolutionary classifications existing before the advent of modern phylogenetic methods (reviewed in Sibley and Ahlquist 1990), and currently, by the ongoing output of conflicting or unresolved higher-level molecular trees. Diverse morphological (Cracraft 1988; Cracraft and Mindell 1989) and molecular datasets have been analyzed (Sibley and Ahlquist 1990; Groth and Barrowclough 1999; Mindell et al. 1999; Stanley 1999; van Tuinen et al. 2000; Johansson et al. 2001; Johnson 2001; Chubb 2002, 2004a; Sorenson et al. 2003), yet a robust phylogeny of Neoaves, the main group of modern birds, has not been produced. The repeated finding of poor resolution raises the possibility that the major lineages of Neoaves differentiated so rapidly that the radiation might be considered essentially simultaneous. Alternatively, the lineages of birds may have speciated dichotomously, but some artifact of the data or methods has prevented accurate resolution of relationships. How can these two possibilities be distinguished empirically?

Demonstrating that a polytomy of evolutionary lineages has occurred is difficult because such a pattern is the null hypothesis for phylogeny reconstruction (Walsh et al. 1999), and true multichotomous evolution is thought to occur rarely or not at all. When a phylogenetic analysis results in a polytomy rather than a dichotomously branching tree, artifactual causes—indicating a ''soft'' polytomy—must be eliminated before a hypothesis of simultaneous radiation—or ''hard'' polytomy—can be tentatively accepted (Maddison 1989). Here, we apply phylogenetic tests to multiple datasets to address the hard versus soft polytomy question for the evolution of the orders of Neoaves.

Definitions, Approaches, and Predictions

In testing for multichotomous evolutionary radiation, it is important to distinguish between gene-tree polytomies and lineage-level polytomies. A gene-tree polytomy occurs when a single allele gives rise to three daughter alleles that are maintained in separate lineages (Slowinski 2001). Although theoretically possible, such an occurrence is probabilistically extremely unlikely. We tested for gene-tree polytomies in this study (see below), but our focus is on whether a lineage-level polytomy has occurred in birds.

To define a lineage-level polytomy, we adopt Maddison's (1997) view of phylogeny as a distribution of gene trees. Evolutionary lineages are composed of multiple gene trees that reflect the history of those lineages if time between evolutionary splits is sufficiently large and effective population sizes are sufficiently small (Tateno et al. 1982; see Fig. 1A). Thus, like a statistical distribution, a ''cloud of gene histories'' may display a central tendency that could be interpreted as the species tree (Maddison 1997, p. 523). However, as the number of generations between splits decreases and the effective population size becomes larger, the probability of discordant gene trees due to lineage sorting increases (Pamilo and Nei 1988). At the limit of nondichotomous evolution, gene trees are expected to either track lineage-level phylogeny as polytomies or be no more similar than chance expectation because of randomly varying coalescence times (Slowinski 2001; see Fig. 1B). We define a lineage-level polytomy as occurring when rapid evolutionary splitting has caused a lack of central tendency in the constituent gene trees. In such a case, a polytomy may be a more biologically accurate representation of the evolutionary pattern than a dichotomous tree. Unlike a gene-tree polytomy, a lineage-level polytomy may be expected under plausible evolutionary circumstances.

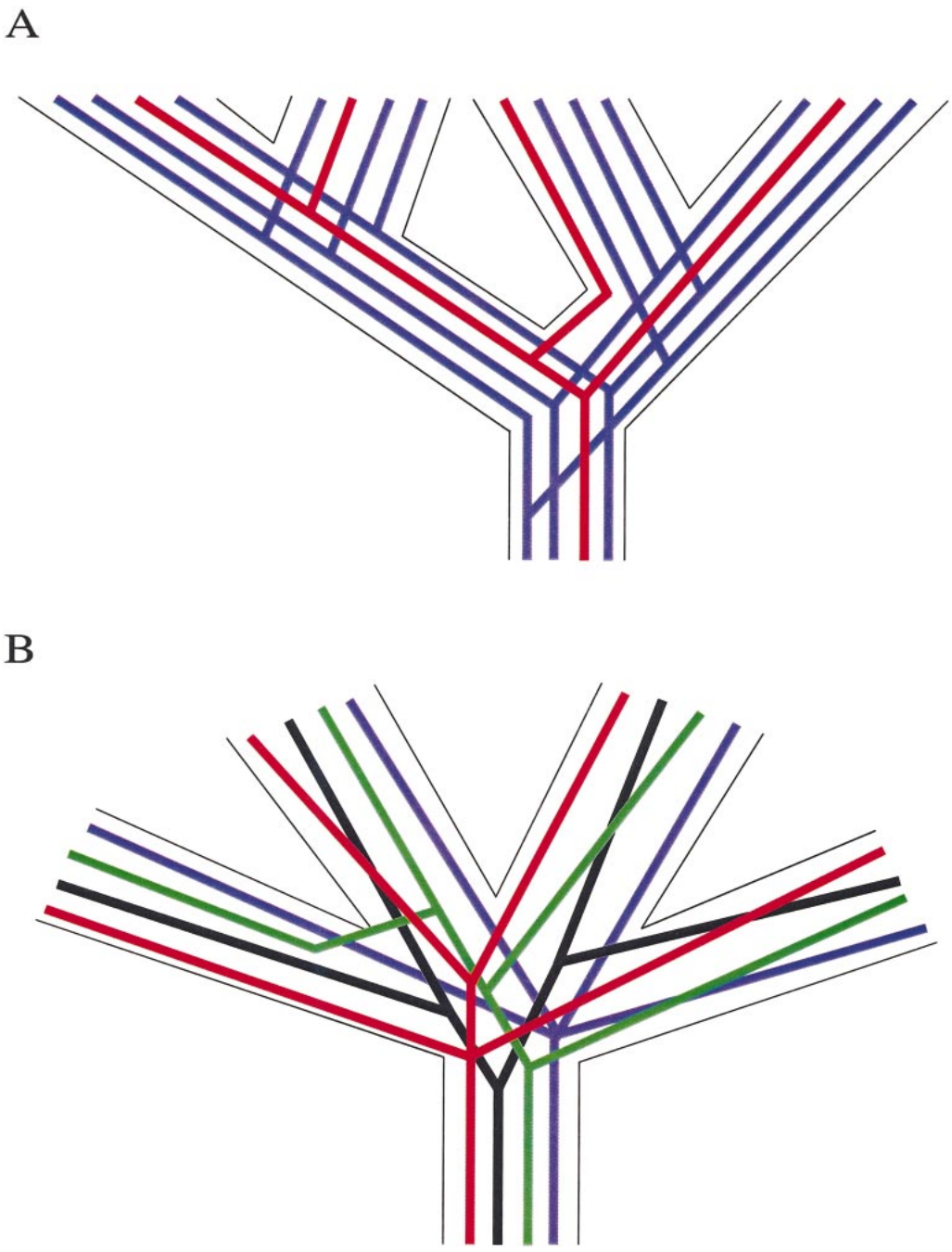

FIG. 1. Expected phylogenetic patterns for four independent genes under (A) rapid dichotomous evolutionary divergence, and (B) simultaneous evolutionary divergence. Gene trees of the same color share the same history. The outlined tree represents the lineage-level history.

For example, a lineage-level polytomy might be expected if a vicariant event such as rising sea level caused geologically rapid separation of multiple (i.e., greater than two) subpopulations in a widespread species with a large effective population size.

Our approach is first to identify the predictions of a hard lineage-level polytomy. Then, we test the bird data to see whether these data fit these predictions. Finally, if the data are found to fit these predictions, we attempt to account for any artifactual effects that might cause similar results, and we evaluate potential alternative hypotheses that might result in the same patterns. Because we are, in essence, suggesting a null hypothesis as potentially true, these last steps are crucial.

Figure 2 illustrates our view of the steps that must be taken to demonstrate a hard lineage-level polytomy (see also Jackman et al. 1999). We expect two potential patterns of data under a hard polytomy, both of which require multiple independent datasets to detect. First, if there is topological structure in independent genetic datasets, this structure

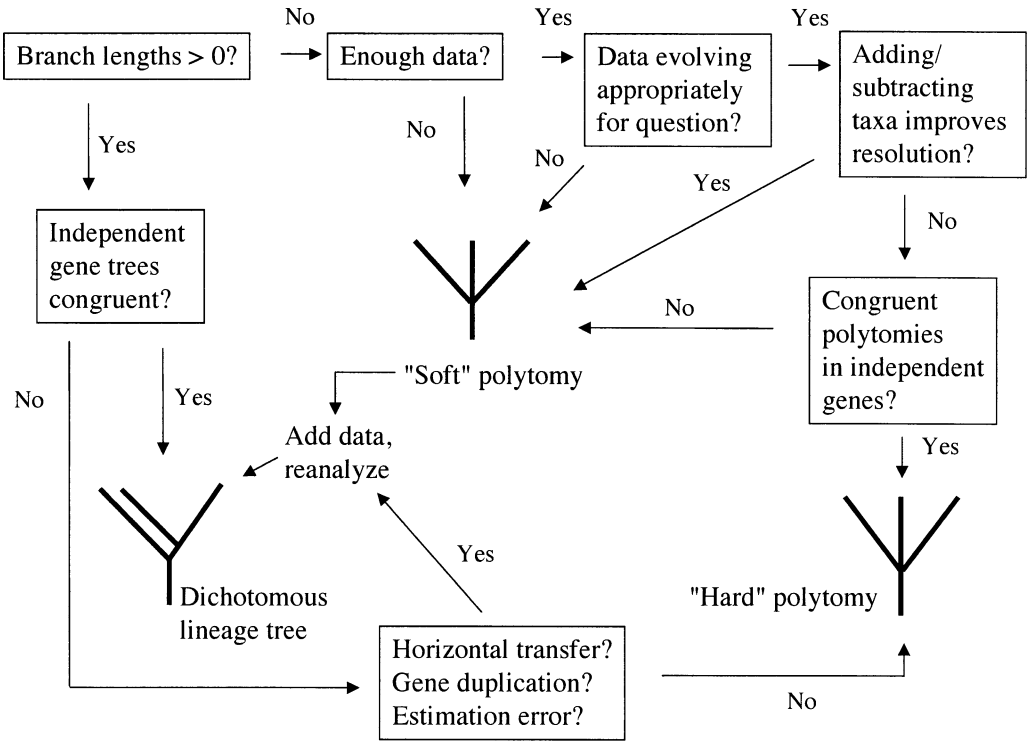

FIG. 2. Flow chart depicting necessary steps for establishing a hard versus soft lineage-level polytomy. The flow chart begins in the upper left and is preceded by phylogenetic analysis.

should not be congruent if a lineage- level hard polytomy has occurred. That is, gene trees should be no more congruent than chance expectation. If, on the other hand, speciation has been rapid (but not polytomous), there should be some ''majority'' signal (i.e., some congruence) among independent genes (Hasegawa et al. 1985). Second, interior branches of gene trees should be extremely short for relationships involving the hypothesized hard polytomy. If multiple independent genes show extremely short interior branches (i.e., branches indistinguishable from zero length) for the same relationships, this is evidence that speciation happened extremely rapidly among those lineages. In this study we test for congruence of independent gene trees and for zero-length branches occurring in the same relationships across independent datasets.

Assuming that datasets meet the predictions of a hard polytomy, several potential alternative hypotheses and artifactual causes must be eliminated before a hard polytomy can be tentatively concluded. Perhaps the most obvious potential cause of a lack of resolution is insufficient data. Here we incorporate a power test for the SOWH test (Swofford et al. 1996) to assess whether enough data have been collected in the case of birds. In addition to insufficient characters, a lack of resolution could be caused by inappropriate taxonomic sampling. Although adding taxa seems more likely to increase accuracy (e.g., Hendy and Penny 1989) rather than to improve support for short branches, we nevertheless performed taxon subsampling experiments with the bird datasets to explore the possibility of a taxon sampling effect. Finally, a common cause of poor resolution is the use of genes that are inappropriate for the divergence level of interest. For example, rapidly evolving mitochondrial genes may be essentially randomized with respect to deep divergences (see Hillis et al. 1996). Thus, it is important that the datasets resolve relationships both deeper and shallower than the hypothesized hard polytomy. We discuss all of these issues with respect to the bird data.

METHODS

Data and Estimation of Phylogeny

We searched the literature for phylogenetic datasets that met four criteria that we deemed useful for testing for a hard polytomy. First, we sought only DNA sequence datasets because these are most amenable to the statistical testing we believe is necessary to rigorously address these questions (e.g., reasonable stochastic models for sequence evolution have been proposed, and such datasets generally possess many characters). Second, we wanted datasets that shared representatives from at least 10 orders of Neoaves. Many datasets include one or a few Neoaves lineages as exemplars in studies of bird or vertebrate phylogeny. We sought datasets for which meaningful comparisons across genes could be made using taxa from the same lineages. Third, we looked for datasets with at least two shared taxa basal to Neoaves to root the trees with the same taxa. Finally, we required that datasets include additional taxa within Neoaves so that we could perform taxon subsampling experiments to assess the sensitivity of our results (see below). We compiled existing Neoaves sequences from six genes that met these criteria: three ribosomal RNAs (mitochondrial 12S,16S, and nuclear 18S; van Tuinen et al. 2000) and three nuclear protein-coding

genes [ZENK (Chubb 2004a,b), c-*myc* (Johansson et al. 2001) and RAG-1 (Groth and Barrowclough 1999; Johansson et al. 2001)]. Other molecular sequence datasets that focus on birds (Johnson 2001: cytochrome *b*; Sorenson et al. 2003: PEPCK nuclear intron; and Prychitko and Moore 2003: β-fibrinogen intron 7) do not meet one or more of these criteria. We discuss results from these studies with reference to our results.

Alignments used were those used by the original authors. We found minimum-evolution trees (Kidd and Sgaramella-Zonta 1971; Rzhetsky and Nei 1992) for each of these datasets (12S and 16S were analyzed together because of their physical linkage) using maximum likelihood distances under the model of evolution determined by Modeltest (Posada and Crandall 1998). Although we would prefer to use likelihood as an optimality criterion, the time needed for likelihood searches is prohibitive for the large number of analyses we are performing. Thus, we adopted minimum evolution as a faster but still model-based proxy for likelihood results. We constructed initial trees with neighbor-joining and 100 replicates of random taxon addition and then perturbed these trees using the tree-bisection-reconnection (TBR) algorithm to find the optimal tree. To facilitate comparison across trees (see below), we also pared each of these datasets to representatives of 10 shared Neoaves orders and a constrained chicken + duck outgroup (Groth and Barrowclough 1999; van Tuinen et al. 2000) and analyzed these reduced-sample trees using maximum likelihood under the model of evolution determined by Modeltest (Posada and Crandall 1998). Taxonomic equivalents across the five datasets were from the same monophyletic lineages. Optimal likelihood trees were obtained by neighbor joining followed by one round of TBR swapping. For both the complete and reduced-sample trees, branch lengths were determined using likelihood. We also estimated a parsimony tree of the combined dataset of all genes for use in the polytomy tests described below. We used PAUP* (Swofford 1999) for all phylogenetic analyses and PAML (Yang 2000) for all simulations.

Tests for Hard Polytomies in the Gene Trees

Do the gene trees contain zero-length branches? And if so, are the zero-length branches congruent across the different datasets? We tested the hypothesis of nonzero length for each branch of interordinal Neoaves relationships in each gene tree using a likelihood-ratio test. The likelihood of the optimal tree was compared to the likelihood of the optimal tree with one branch of interordinal relationships collapsed. A nonsignificant difference between likelihood scores indicates that the tested branch is statistically indistinguishable from zero length. This test was performed for each branch of interordinal relationships, as implemented in PAUP (under Describe Trees) and Bonferroni-corrected for the number of interordinal branches.

The above approach is useful for testing whether zero-length branches exist at all, but it is not able to assess whether multiple lineages have been involved in polytomies (because the test examines one branch at a time), nor can it test whether the same branches are zero-length in all genes (because the datasets return different optimal trees). These aspects are difficult to test for several reasons, especially with large datasets of heterogeneous taxonomic composition. First, the likelihood-ratio test statistic is unlikely to follow a chi-square distribution when the optimal tree is compared to trees with more than one clade collapsed simultaneously (Goldman and Whelan 2000). Calculation of significance levels thus requires a simulation approach, which would be prohibitively time consuming for the larger datasets. Second, differences in taxonomic composition confound comparisons between particular clades. For example, it makes no sense to ask whether the branch uniting *Trogon* and *Rallus* in the ZENK tree is indistinguishable from zero length in all datasets, because *Rallus* is not present in the other datasets. Third, even if taxonomic composition is identical between datasets (see below), different optimal tree topologies make comparison of clades difficult. A zero-length branch in a clade composed of, for example, a turaco, a mousebird, and a nighthawk may be checked for zero length in the other datasets only if these three taxa are grouped together in the trees from those other datasets.

We attempted to address these difficulties by paring the datasets to representatives of orders present in all datasets and, after initial analyses, by constraining identical relationships across genes to reflect plausible lineage-level relationships. We used pared 10-taxon datasets to perform a series of tests to determine whether (1) multiple branches could be collapsed simultaneously and still maintain nonsignificance, and/or (2) there existed congruent zero-length branches across all gene trees, assuming some true underlying lineage tree. First, we assumed one of the gene trees or the combined parsimony tree to be the true lineage-level tree and imposed this topology on each of the datasets. Then we collapsed sets of one to eight (= all) ingroup branches and compared likelihoods between the collapsed and noncollapsed trees using each gene. Specifically, we first collapsed all eight ingroup clades and tested for a significant difference in likelihood from the assumed lineage tree using each gene (see below). If a significant result was obtained, we collapsed sets of seven clades and again tested for significant likelihood difference from the lineage tree. We continued this procedure, collapsing smaller numbers of clades, until a nonsignificant result was obtained. The number of clades that could be collapsed while maintaining nonsignificance gives an estimate of the number of congruent gene-tree polytomies across all genes, assuming some identical underlying topology. Although this approach clearly is ad hoc, we are resigned to it because there is no clear a priori hypothesis of what lineage tree should be assumed or of which sets of clades to hypothesize as potentially zero-length. In spite of the potentially large number of tests, we held the *P*-value at 0.01 so as not to bias the results toward the conclusions of this study, which require nonsignificance.

We created null distributions for the tree comparison test statistics by simulating data on the null trees (the SOWH test of Swofford et al. 1996). The null tree is a tree with the hypothesized zero-length branches (simultaneously) collapsed. For example, the null tree for the 12S/16S results is resolved as in Figure 4D except that the two clades that may be simultaneously collapsed and still maintain nonsignificant difference from zero-length across all datasets (woodpecker + owl, cuckoo + trogon) have been collapsed to trichoto-

mies. For comparisons involving the gene that produced the tested lineage level topology—for example, using the ZENK data on the optimal tree for ZENK—the test statistic was the difference in likelihood between the null (collapsed) tree and the optimal tree under the data that produced those trees. For these cases we created a null distribution by simulating 99 datasets on the null topology using the same model and parameter values for which the likelihoods of the original trees were estimated. For each simulated dataset we compared the likelihood of the null tree to the likelihood of the tree that was optimal for that replicate. This distribution of 99 values was then compared to the value for the test statistic.

A different comparison was undertaken when a gene was tested using independently determined trees from one of the other genes or from the combined analysis, for example, testing the ZENK data on the combined parsimony topology or on the optimal 18S topology. In this case, the test statistic was the difference in likelihood between the null topology (where null topology is determined as above) and the optimal topology obtained under one dataset but inferred using the sequence data of an independent gene and its associated model parameters. Thus, we compared two predetermined trees rather than a predetermined null tree and an estimated optimal tree. For each of these comparisons, we created a null distribution by simulating 99 datasets under the null topology with branch lengths obtained under the parameter values for the tested gene. For each simulation replicate, we recorded the difference in likelihood between the null tree and the original (i.e., not estimated) optimal tree from which the test statistic was obtained. This distribution of 99 values was then compared to the value for the test statistic.

Power Test for Sufficient Data

Are the current datasets large enough to resolve extremely short branches? Power is defined as the probability of rejecting a null hypothesis when that null hypothesis is false. Previously proposed statistical power tests for multichotomous radiation (Walsh et al. 1999) have not taken phylogenetic methods and tree structure into account. We developed a statistical power approach for the SOWH test (Swofford et al. 1996) of the hypothesis of nonzero branch length. To calculate phylogenetic statistical power for hypothesis testing, one needs values for acceptable Type-I error, effect size (branch length), and sample size (sequence length). Because a likelihood-based comparison of resolved and collapsed trees does not necessarily follow any standard statistical distribution under the tested conditions, calculating power necessitates a simulation approach (Goldman and Whelan 2000).

We first selected a Type-I error (*P*-value) of 0.05 and recorded the critical value for data simulated under the null hypothesis of zero-length branches. To do this for a given dataset, we created 100 datasets of the same size as the original using the optimal tree and Modeltest model but with the tested branch lengths constrained to zero. For example, we simulated 12S/16S data on the tree in Figure 4D but with the woodpecker + owl and trogon + cuckoo clades collapsed to form trichotomies. We analyzed each of these simulated datasets using likelihood under the known model and recorded the difference in likelihoods between the null tree and the optimal tree for each simulation replicate. The 5% largest values were noted.

Next, we simulated datasets of the same size and with the same parameter values but using various branch lengths (effect sizes) for the hypothesized zero-length branches. We added equal length to each collapsed branch simultaneously. For example, we simulated 12S/16S data on the tree in Figure 4D with the branch lengths for the woodpecker + owl and trogon + cuckoo clades simultaneously set to (for example) 0.0001 rather than the zero length that was used above to get values under the null hypothesis. We simulated data over the range of branch lengths that produced power values between 0 and 1.0 using both the original number of sites for each gene and using 15,000 sites. This range of informative branch lengths was identified through trial and error (for each gene, only a narrow range of branch lengths produced power values between 0 and 1.0). For each simulation replicate we performed the same comparison of the optimal and null (collapsed) tree. We recorded the percentage of values that exceeded the 5% largest values observed in the null comparison for each set of nonzero branch lengths. This result gave an estimate of Type-II error and allowed calculation of power curves. Power tests were performed on the optimal topologies for each gene using the parameter values for the gene that produced the tree.

Test for topological congruence

Are the five gene trees more congruent than chance expectation? To address this question we summed the number of taxon bipartitions shared across the optimal ingroup trees for the five datasets (Penny et al. 1982). For example, if the only similarities between trees are that three of the five gene trees have a turaco plus mousebird clade and two of the gene trees group trogon, passeriform, and owl, the sum of shared bipartitions is five. This test statistic has a maximum of 35 if all five trees have the same seven ingroup bipartitions and a minimum of zero if no groups are shared across the five datasets. This value was compared to the number of bipartitions shared in sets of five trees chosen from a null distribution of Markovian trees obtained from PAUP.

RESULTS

Phylogenetic Analyses and Tests for Nonzero-Length Branches

Optimal trees from the analyses using all taxa available for the largest number of Neoaves orders are shown in Figure 3. For each of these trees, a large percentage of Neoaves interordinal relationships could be collapsed with nonsignificant difference from the optimal tree: 19 of 19 interordinal relationships in c-*myc*, 10 of 11 in RAG-1, 22 of 27 in ZENK, 16 of 33 in 12S/16S, 19 of 19 in 18S. If a Bonferroni correction is not used, values are: 15 of 19 interordinal relationships in c-*myc*, eight of 11 in RAG-1, 14 of 27 in ZENK, six of 33 in 12S/16S, 13 of 19 in 18S. The RAG-1 tree shown in Figure 3 is based on a data matrix with 2.9 kb, but we also performed the same test on a RAG-1 tree that samples more operational taxonomy units (46), but only 930 base

pairs. In that case, 13 of 13 interordinal branches could be collapsed to yield a likelihood not significantly different from the optimal tree (nine of 13 without the Bonferroni correction). These results show that each of these datasets is characterized by extremely short interordinal Neoaves branches.

Optimal likelihood trees for the pared datasets are shown in Figure 4. Regardless of which of the six (five optimal gene trees, optimal combined parsimony tree) candidate lineage-level trees is assumed, at least two, and up to seven, clades could be collapsed simultaneously with no significant change in likelihood in all five datasets. This result suggests that zero-length branches may occur in the same relationships across all genes.

Power Results

Results of the power tests suggest that enough data have been collected (see Fig. 5). Each dataset includes sufficient characters to reject the null hypothesis of zero length for the tested branches even if actual branch lengths are extremely short. Increasing data to 15,000 sites has virtually no effect on power in three of the five datasets (ZENK, RAG-1, 18S) and only slight improvement in the others (Fig. 5). For example, increasing the number of ZENK characters increases power by more than 30% only for branch lengths from about 0.0002 to 0.0005. For branch lengths below 0.0002 or above 0.0005, power is nearly identical for 1763 and 15000 characters.

Topological Congruence Results

The sum of the number of shared ingroup clades in the five gene trees is eight. Figure 6 shows the distribution of the test statistic under the null hypothesis of Markov-generated trees. Eight is not greater than chance expectation under the Markov model ($P = 0.24$), and thus the null hypothesis of independent histories cannot be rejected.

Discussion

Biological Interpretation of the Gene-Tree Polytomies

There is a minimum of two congruent gene-tree trichotomies among these Neoaves orders (Fig. 4). This number is conservative for two reasons. First, some excluded Neoaves orders may fall within zero-length branches. Second, alternative trees that are suboptimal for all (rather than for just four of the five) datasets may best represent the lineage-level tree. In such a case, achieving nonsignificance for multiple collapsed clades across all datasets may be easier because relationships have not been optimized for any of them. Although many studies discuss rapid radiations, we know of no other data that fail to reject a null hypothesis of zero-length branches using this SOWH test (see Fishbein et al. 2001).

Because we are unable to determine which of the many thousands of candidate lineage trees is most appropriate, we cannot determine exactly which bird relationships to present as multichotomous. The goal of examining six potential lineage trees was to investigate whether the hypothesis of multichotomous evolution is sensitive to choice of lineage topology. Because all six tested trees resulted in some congruent zero-length branches across all gene trees, we are comfortable suggesting that at least some interordinal relationships are multichotomous, even though we are unable to identify these relationships by name.

Although it is clear that most of the interordinal bird branches are in some sense short, we find it difficult to believe that the hypothesized zero-length branches represent true gene-tree polytomies in the sense of three alleles arising from one. Below, we consider alternative hypotheses that involve Type-II error, some of which are compatible with the hypothesis of a hard lineage-level polytomy.

First, it is possible that dichotomous evolution of genes has occurred, but that no fixed mutations occurred in the tested genes during the time between splits. In such a case, the true likelihood branch length is nonzero because branch length is a function of both time and mutation rate (Swofford et al. 1996). However, the estimated branch length is likely to be zero because we estimate branch lengths wholly from mutations. Because branches like these are un likely to be estimated correctly if no data exist to support them, such branches may be best interpreted as zero length al though their actual length, in terms of time, is nonzero. Under this interpretation, hard lineage-level polytomies are indicated by the congruent zero-length branches across the gene trees.

A second possibility is that mutations did occur in some of the genes in the duration between the short dichotomous splits, and the reconstructed branch lengths, however short, represent the actual branch lengths (it must be noted that three of the five optimal gene trees have at least one interordinal branch reconstructed as exactly zero length). Under this interpretation, the lack of nonrandom congruence between gene trees (Fig. 6) fulfills the expectation of a hard lineage-level polytomy (Fig. 1). Similar to this interpretation is the possibility that there is some common lineage-level phylogeny for part or all of the gene trees, but stochastic variation and short branches have obscured the signal. Additional data will determine whether this interpretation should be favored. We do not favor it currently for two reasons. First, the power tests suggest that additional data will not be helpful in resolving these branches, if current branch length estimates are close to accurate. Second, if there is some true underlying lineage structure, we would expect at least some nonrandom congruence between the five independent markers, and this does not occur (Fig. 6).

A third possibility is that the failure to reject the star tree hypothesis may be caused by a combination of factors that are difficult to disentangle—for example, stochastic variation in the evolutionary process (i.e., fixed mutations failed to occur in some genes) and stochastic error in the estimation process (i.e., some topological relationships and branch lengths are reconstructed incorrectly)—and which affect part or all of the interordinal lineage relationships and potentially obscure a common lineage-level signal. This interpretation is difficult to address with current data, but we note, again, that at least some nonrandom congruence of topologies would be expected under these circumstances if a lineage-level polytomy did not occur.

Potential Methodological and Artifactual Causes of Zero-Length Branches

The independent genetic datasets each possess branch lengths that are indistinguishable from zero, which may re-

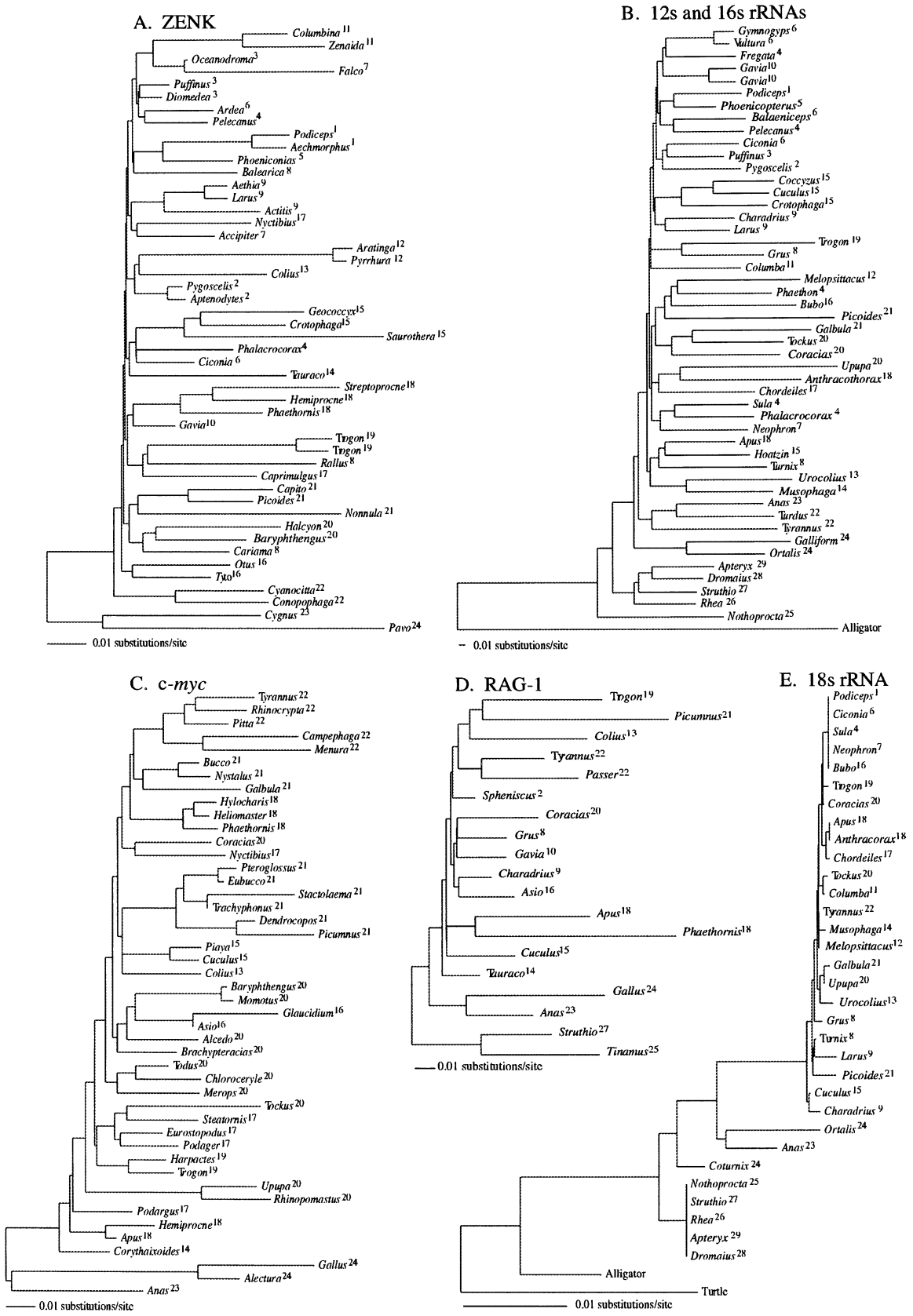

FIG. 3. Optimal distance trees with maximum-likelihood branch lengths and complete taxon sampling for the most avian orders and base pairs available. (A) ZENK (Genbank nos. AF490141–AF490243, unreleased), (B) 12s and 16s rRNAs (van Tuinen et al. 2000), (C) c-*myc* (Johansson et al. 2001), (D) RAG-1 (Groth and Barrowclough 1999; Johansson et al. 2001), (E) 18s rRNA (van Tuinen et al. 2000).

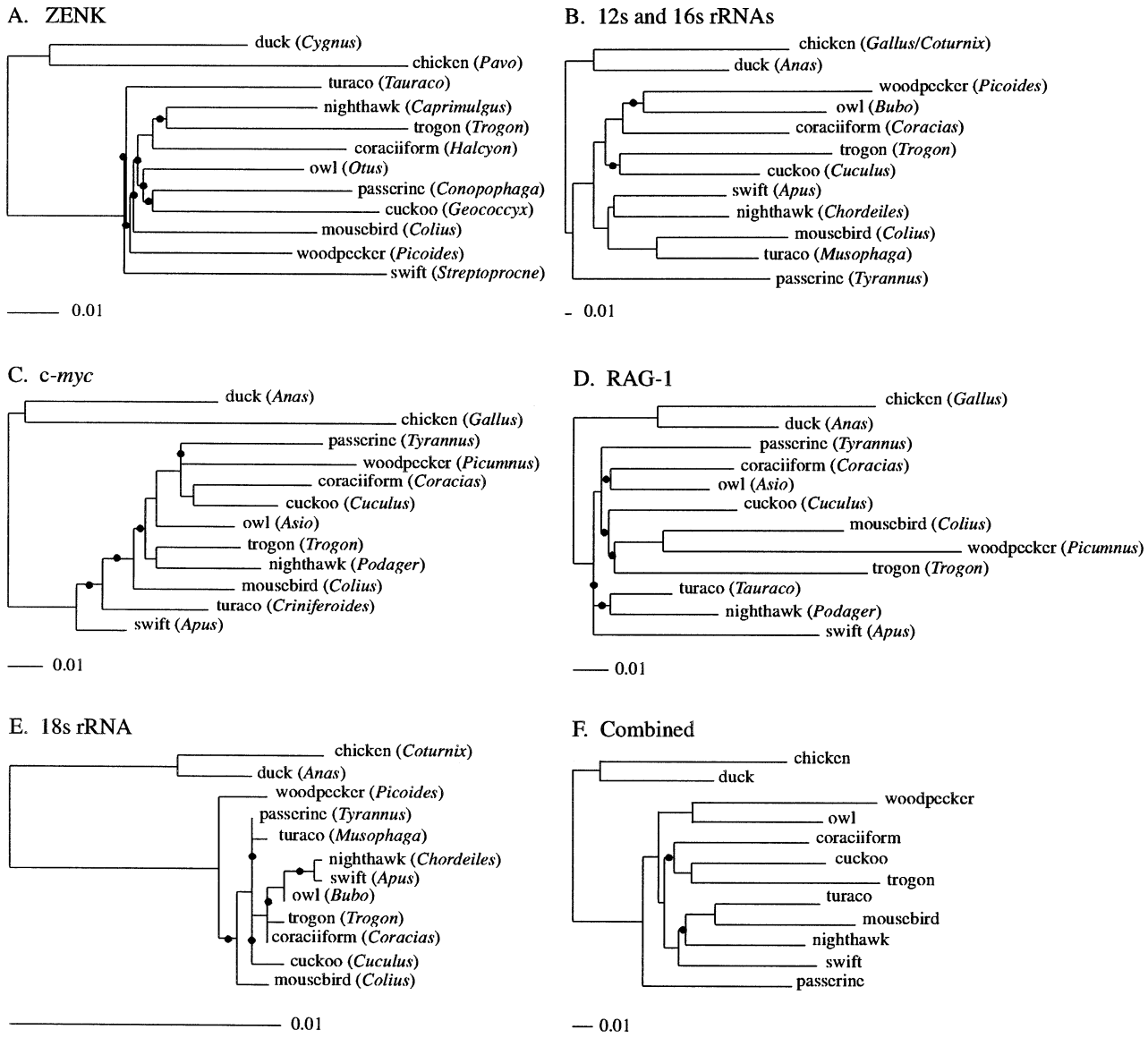

FIG. 4. Optimal maximum-likelihood trees from reduced taxon samples, containing orders shared by all five datasets. Clades that were congruently zero length across all datasets when the tree from a particular gene tree or the combined tree was assumed are identified with dots (not significantly different from zero length): (A) ZENK tree, (B) 18s rRNA tree, (C) RAG-1 tree, (D) 12s and 16s rRNA, (E) c-*myc*, (F) combined maximum-parsimony tree, shown with likelihood branch lengths.

flect either the real evolution of these organisms or some artifact of these analyses. The most obvious potential artifact is insufficient data; however, the power results suggest that this is not a problem for these genes (Fig. 5). An alternative potential problem is that some taxon sampling artifact has biased each analysis towards a multichotomy (we have excluded approximately 9500 Neoaves species). However, the result that zero-length branches are obtained with both larger (Fig. 3) and smaller (Fig. 4) numbers of taxa suggests that these conclusions are robust to taxon sampling artifacts. In

←

Taxa belong to the following Neoaves orders, as marked: 1. Podicipediformes (grebes), 2. Sphenisciformes (penguins), 3. Procellariiformes (tube-nosed seabirds), 4. Pelecaniformes (pelicans and allies), 5. Phoenicopteriformes (flamingos), 6. Ciconiiformes (herons, storks, and allies), 7. Falconiformes (birds of prey), 8. Gruiformes (cranes, rails, and allies), 9. Charadriiformes (shorebirds, gulls, and allies), 10. Gaviiformes (loons), 11. Columbiformes (pigeons and doves), 12. Psittaciformes (parrots), 13. Coliiformes (mousebirds), 14. Musophagiformes (turacos), 15. Cuculiformes (cuckoos), 16. Strigiformes (owls), 17. Caprimulgiformes (nightjars and allies), 18. Apodiformes (swifts and hummingbirds), 19. Trogoniformes (trogons), 20. Coraciiformes (kingfishers, rollers, and allies), 21. Piciformes (woodpeckers and allies), 22. Passeriformes (perching birds). Galloanserae orders include: 23. Anseriformes (waterfowl), 24. Galliformes (fowl-like birds). Paleognath orders include: 25. Tinamiformes (tinamous), 26. Rheiformes (rheas), 27. Struthioniformes (ostriches), 28. Casuariiformes (emus and cassowaries), 29. Dinornithiformes (kiwis).

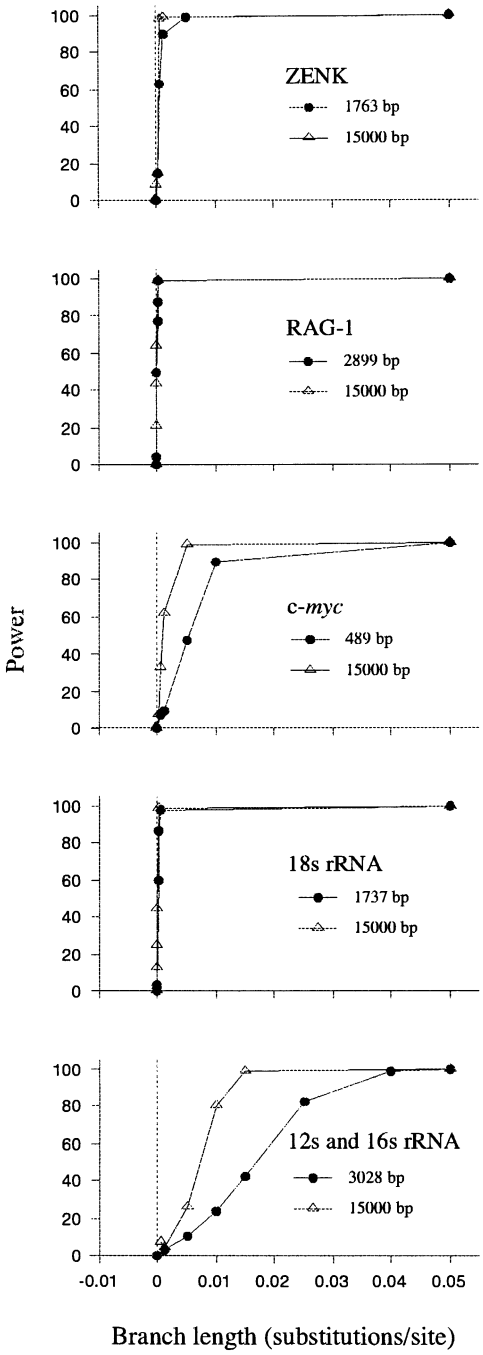

FIG. 5. Power curves for each gene, comparing original datasets (crosses) with simulated datasets of 15,000 bases (boxes). Note differences are branch-length scale.

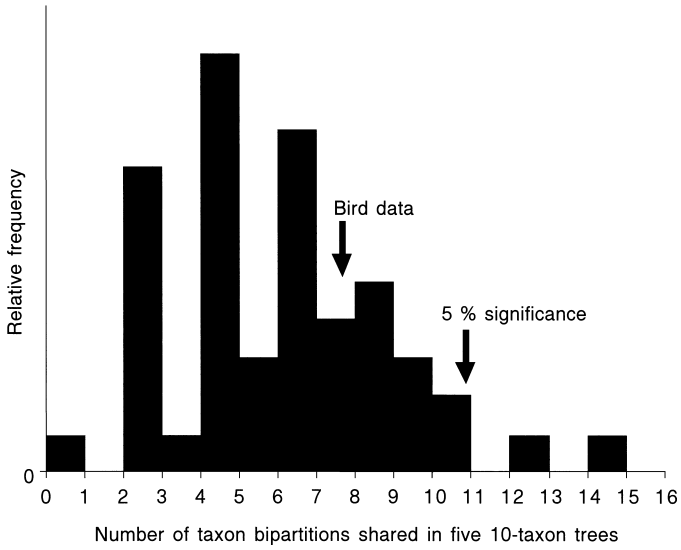

FIG. 6. Null distribution for significance test of the sum of taxon bipartitions shared in sets of five Markov-generated trees of 10 taxa.

addition to taxon sampling issues, it might be thought that each of the five independent genes is evolving at an inappropriate rate to resolve Neoaves ordinal relationships. We reject this explanation because each gene (except for 18S) resolves both older and more recent relationships without saturation (Groth and Barrowclough 1999; van Tuinen et al. 2000, 2001; Johansson et al. 2001; Chubb 2004a,b). It remains possible that these genes are able to resolve ''old long'' internodes but not ''old short'' internodes. That is, each of these genes may be evolving slowly enough to be informative at deep splits but too slowly to have undergone mutation in the short periods of interordinal diversification. Whether these five genes are such an unfortunate sample can only be determined by collection of additional data. If such an interpretation is correct, we note that identification of appropriate genes for Neoaves ordinal relationships is likely to be extremely difficult. Finally, our results do not exclude the possibility that the genes have simultaneously undergone a deceleration in rate of molecular evolution at the point of ordinal diversification followed by a simultaneous acceleration after ordinal diversification. However, in light of the additional evidence presented here (e.g., topological incongruence) and our belief in the improbability of this explanation, we consider this possibility less likely than the ''common cause'' explanation of a lineage-level polytomy.

Biological Interpretation of Topological Incongruence

The failure to reject a null hypothesis of independent evolution for these gene trees (Fig. 6) is compatible with a hypothesis of a hard lineage-level polytomy (Fig. 1). If there was some congruent lineage-level relationship, we would expect to see some ''majority'' signal from the independent genes. We consider this evidence compelling because the hypothesis of nonrandom tree similarity is extremely conservative. If, for example, even two clades were shared across the five trees, or if two of the five trees were identical and the other three each unique, a significant result would be obtained under this test. We performed this test using only the ingroup taxa. An additional test including an outgroup and comparing rooted (11-taxa) trees also returns a nonsignificant result (data not shown).

There are other factors that could cause a lack of congruence between gene trees. One of these is gene duplication followed by subsequent loss of one of the gene copies (Fitch 1970). If gene duplication is a factor in the bird data, the utter lack of congruence between trees suggests that some

process is simultaneously affecting each of the five putatively independent genes. If only one or a few genes were affected or if the gene duplications occurred late in bird ordinal phylogeny, we would expect some congruence between topologies. The independent duplication of five randomly selected genes is probabilistically unlikely over the apparently short timeframe of bird ordinal diversification (Lynch and Conery 2000). Large-scale (i.e., whole-genome) duplications are not unheard of in nature (e.g., Wolfe and Shields 1997), and additional data may shed light on this possibility. For now, we note that the hard lineage polytomy hypothesis predicts short gene-tree branches, which are present in these data, whereas the gene duplication hypothesis does not (although it does not exclude them).

Hybridization may also cause discordance between gene trees (e.g., Doyle 1992). However, this phenomenon would have to be rampant to produce the extreme topological dissimilarity observed in these data. Furthermore, like the gene duplication hypothesis, short internodal branches are not necessarily expected under hybridization as they are under a hard lineage-level polytomy.

Potential Methodological Causes of Topological Incongruence

Another possible explanation for the discord between gene trees is estimation error. That is, the five datasets may be responding differently to difficult phylogeny estimation conditions involving long unbranched lineages and short internodes (see Felsenstein 1978). Certainly, the reconstruction of these short, deep branches in birds is likely to be an extremely difficult estimation problem. However, if misestimation is a problem, we argue that more congruence should be observed than is seen among these datasets. The reason significant congruence is expected is that congruence can occur due to both shared true relationships and shared bias (see e.g., Poe 2003, fig. 1). If long-branch attraction is a problem because lineages have been evolving independently for a long time, then at least some of these datasets should share the same long branches. Under these conditions, independently analyzed datasets would have two forces contributing to congruence—shared history and the shared bias of placing the same long branches together—and greater congruence than is observed here would be expected. Conversely, for misestimation to cause systematic (rather than stochastic) incongruence due to long-branch attraction, rates of molecular evolution would need to vary greatly over time within and between each dataset such that different long branches underlie each of the five genes. We do not suggest that the misestimation explanation for the incongruence observed here is impossible. But in light of the additional evidence presented here we consider it less likely than the common-cause explanation of a hard polytomy.

Comparison with Other Approaches

Jackman et al. (1999) and Walsh et al. (1999) have also offered procedures for distinguishing hard and soft polytomies. Our approach draws from each of these works.

Walsh et al. (1999) suggested applying the concept of statistical power to the question of how many molecular sequence sites are needed to resolve a given phylogenetic branch. These authors calculated the amount of sequence data required to obtain at least one substitution on a branch of length 100,000 years in a study of auklets. Because a single substitution is the minimum amount of data required to resolve a branch, this approach provides a useful estimate of the minimum amount of data needed for resolution (Braun and Kimball 2001; Walsh and Friesen 2001). However, this measure does not take into account structure in the rest of the tree or the method of reconstruction used, both of which can have profound consequences for the ability to reconstruct a branch of interest (e.g., Felsenstein 1978). The simulation approach to calculating power that we adopt here takes tree structure into account and is usable with different phylogenetic methods (for application of simulation-based power to their likelihood ratio test for monophyly, see Huelsenbeck et al. 1996). For example, one could compare the length in parsimony steps (rather than likelihood score) of the optimal tree to the constrained tree for each simulation replicate when calculating power.

Jackman et al. (1999) were among the first to treat the rapid radiation question quantitatively. They presented a series of tests within a parsimony framework to distinguish hard versus soft polytomies. Our Figure 2 adopts many of the principles of the flowchart of Jackman et al. (1999, fig. 4), including tests for significant internal branches and asking whether sufficient numbers of characters have been collected. Jackman et al. (1999) tested whether certain groups of sites were saturated with substitutions and performed taxon subsampling experiments to test whether short branches in their maximally sampled tree remained short under reduced samples of taxa. Although we discuss the appropriateness of the genes we used, we did not perform saturation tests because the likelihood framework adopted in this study is expected to deal better with differing rates of evolution (Swofford et al. 1996). Like Jackman et al. (1999), we performed taxon subsampling experiments to assess the effect on hypothesized zero-length branches. Our approach differs in adopting a likelihood rather than parsimony framework (although many of our tests could be performed under parsimony), including an explicit statistical assessment of whether sufficient numbers of characters have been collected, and incorporating topology and branch length tests with multiple independent genes. We also discuss the distinction between gene and lineage-level trees, emphasize the necessity of using multiple markers in addressing hard versus soft lineage-level polytomy questions, and examine alternative explanations for polytomous results. Jackman et al. (1999) used a single mitochondrial region and thus were unable to address issues related to multiple loci.

Causes of Rapid Radiation

Several studies have noted the rapid radiation of birds and the difficulty in reconstructing Neoaves ordinal relationships (Cracraft 1988; Sibley and Ahlquist 1990; Hedges 1994; Cooper and Penny 1997; Groth and Barrowclough 1999; Mindell et al. 1999; Stanley and Cracraft 2002; Sorenson et al. 2003; Chubb 2004a). Hard polytomies are thought to be rare or nonexistent in nature, so the continued finding of poor resolution begs for some causal explanation. The traditional

explanation for the explosive radiation of birds is niche expansion following mass extinction at the Cretaceous-Tertiary (K-T) boundary (Romer 1966; Feduccia 1995). This seems a plausible explanation if the diversification of Neoaves can be dated to around 60 million years ago. However, recent molecular studies have suggested that most orders of Neoaves diversified before the K-T boundary (Hedges et al. 1996; Cooper and Penny 1997). This contention could be reconciled with the rapid-radiation results of this paper if continental breakup during the Cretaceous caused (geologically) simultaneous isolation of multiple lineages. Alternatively, perhaps niche expansion occurred but not in response to a mass extinction.

Predictions for Future Data

The great biological interest in birds ensures that phylogenetic data will continue to be collected for resolving avian interordinal relationships. We currently favor a hypothesis of lineage-level polytomy for the relationships of some of the bird orders, but we eagerly await new data that may falsify this hypothesis. Although the data may be interpreted as compatible with a hard polytomy involving many (up to 10) lineages, we consider it unlikely that such a large polytomy would exist in nature. Rather, we expect that some bird ordinal relationships analyzed in this study will be resolved by additional data, whereas other bird ordinal relationships will be found to form true hard lineage-level polytomies. Fortunately, these two conditions should be distinguishable because the predictions of the lineage-level polytomy (as we have defined it) are clear. If data from several independent genes eventually converges on some majority topology, the hypothesis of lineage-level polytomy will be falsified. Alternatively, if independent genes continue to favor disparate gene trees with short internal branches, as is currently the case, the hard-polytomy hypothesis should be favored over one of normal dichotomous evolution. Such testing will require analyzing putatively independent genetic datasets separately to allow them to display separate histories (see Miyamoto and Fitch 1995). We also recommend continued analyses of other characters such as chromosomes, gene order, and morphology. As with the gene sequence data, the expectation under the hard-polytomy hypothesis is that some relationships will not be resolvable dichotomously. The consequences of other potential causes for the patterns seen here (i.e., gene duplication) should also be investigated.

Recent work that did not meet our sampling criteria nevertheless mainly supports, or at least is compatible with, the hard polytomy interpretation. Sorenson et al.'s (2003) study of Hoatzin and other bird relationships and Johnson's (2001) 916-taxon study using cytochrome *b* both found extremely weak support for interordinal Neoaves relationships (i.e., bootstrap values $<50\%$). Cytochrome *b* may not be appropriate for divergences at the level of Neoaves orders (Moore and DeFilippis 1997), but Sorenson et al. (2003) sampled a nuclear intron that appeared useful for older divergences in birds (e.g., strong support for a Galloanserae-Neoaves split). Prychitko and Moore (2003) included six Neoaves orders and found strong support for three clades of Neoaves ordinal relationships (bootstraps of 96, 98, 100%) and weak support for one other (43%). Although results from this limited ordinal-level taxon sample should be treated with caution, the finding of support for at least some interordinal relationships is encouraging.

We suggest that if bird ordinal relationships are actually dichotomous, bird evolution may represent an extremely difficult phylogenetic case. If the true interordinal branch lengths are anything like those reconstructed in the five datasets analyzed here, it will take a huge amount of data and an extremely accurate model of evolution to resolve these splits (Fig. 5). Fortunately, recent technological advances in data collection and analysis push such requirements into the realm of possibility. The next few years should be an exciting time in avian systematics.

Conclusions

Despite over a century of effort by avian systematists, a robustly-supported avian evolutionary tree has been elusive. Our analyses suggest that this difficulty is not due to a paucity of characters or datasets, but rather to an evolutionary pattern that may violate the assumption of a dichotomously branching tree. We believe the evidence of (1) extremely short (in some cases zero-length) branches for the same relationships in independent gene trees, and (2) lack of nonrandom topological congruence between independent gene trees, is best interpreted as indicating a lineage-level polytomy among some of the orders of Neoaves. Although new data should continue to be collected and analyzed, the avian evolutionary tree may remain more like a bush at the ordinal level. The explosive radiation of birds may be exceptional, or perhaps in the future our increased ability to collect genetic data and the continued application of new phylogenetic tools will allow identification of other cases of simultaneous evolutionary radiation of multiple lineages. The general methodology that we have proposed for testing rapid radiations should be useful in this regard.

Acknowledgments

Thanks to the Miller Institute for funding to SP. Thanks to S. Kuchta, M. Mahoney, L. Rissler, J. Rodriguez-Robles, N. Johnson, D. Wake, C. Moritz, R. Mueller, J. Patton, D. Meyer, J. Huelsenbeck, J. Wiens, and the anonymous reviewers for comments on this work. The authors contributed equally to this research.

Literature Cited

Braun, E. L., and R. T. Kimball. 2001. Polytomies, the power of phylogenetic inference, and the stochastic nature of molecular evolution: a comment on Walsh et al. (1999). Evolution 55: 1261–1263.

Chubb, A. L. 2002. Molecular evolution and phylogenetic utility of the ZENK gene in birds. Ph.D. diss. Department of Integrative Biology; University of California, Berkeley.

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

———. 2004b. Nuclear corroboration of DNA-DNA hybridization in deep phylogenies of hummingbirds, swifts, and passerines: the phylogenetic utility of ZENK (ii). Mol. Phylogenet. Evol. 30:128–139.

Cooper, A., and D. Penny. 1997. Mass survival of birds across the

Cretaceous-Tertiary boundary: molecular evidence. Science 275: 1109–1113.
Cracraft, J. 1988. The major clades of birds. Pp. 339–361 *in* M. J. Benton, ed. The phylogeny and classification of the tetropods. Vol.1. Syst. Assoc. Special Vol. No. 35A. Clarendon, Oxford, U.K.
Cracraft, J., and D. P. Mindell. 1989. The early history of modern birds: a comparison of molecular and morphological evidence. Pp. 389–403 *in* B. Fernholm, K. Bremer, and H. Jornvall, eds. The hierarchy of life. Elsevier, Amsterdam.
Doyle, J. J. 1992. Gene trees and species trees: Molecular systematics as one character taxonomy. Syst. Bot. 17:144–163.
Feduccia, A. 1995. Explosive evolution in Tertiary birds and mammals. Science 267:637–638.
Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27:401–410.
Fishbein, M., C. Hibsch-Jetter, D. E. Soltis, and L. Hufford. 2001. Phylogeny of Saxifragales (Angiosperms, Eudicots): analysis of a rapid, ancient radiation. Syst. Biol. 50:817–847.
Fitch, W. M. 1970. Distinguishing homologous from analogous proteins. Syst. Zool. 19:99–113.
Goldman, N., and S. Whelan. 2000. Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. Mol. Biol. Evol. 17:975–978.
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.
Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.
Hedges, S. B. 1994. Molecular evidence for the origin of birds. Proc. Natl. Acad. Sci. USA. 91:2621–2624.
Hedges, S. B., P. H. Parker, C. G. Sibley, and S. Kumar. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226–229.
Hendy, M. D., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38:297–309.
Hillis, D. M., B. K. Mable, A. Larson, S. K. Davis, and E. A. Zimmer. 1996. Nucleic acids IV: sequencing and cloning. Pp. 321–381 *in* D. M. Hillis, C. Moritz and B. K. Mable, eds. Molecular systematics. Sinauer, Sunderland, MA.
Huelsenbeck, J. P., D. M. Hillis, and R. Jones. 1996. Parametric bootstrapping in molecular phylogenies: applications and performance. Pp. 19–45 in J. D. Ferraris and S. R. Palumbi, eds. Molecular zoology: advances, strategies, and protocols. Wiley-Liss, New York.
Jackman, T. R., A. Larson, K. De Queiroz, and J. B. Losos. 1999. Phylogenetic relationships and tempo of early diversification in Anolis lizards. Syst. Biol. 48:254–285.
Johansson, U. S., T. J. Parsons, M. Irestedt, and P. G. P. Ericson. 2001. Clades within the ''higher land birds'', evaluated by nuclear DNA sequences. J. Zool. Syst. Evol. Res. 39:37–51.
Johnson, K. P. 2001. Taxon sampling and the phylogenetic position of Passeriformes: evidence from 916 avian cytochrome *b* sequences. Syst. Biol. 50:128–136.
Kidd, K. K., and L. A. Sgaramella-Zonta. 1971. Phylogenetic analysis: concepts and methods. Am. J. Hum. Genet. 23:235–252.
Lynch, M., and J. C. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1154.
Maddison, W. P. 1997. Gene trees in species tree. Syst. Biol. 46: 523–536.
———. 1989. Reconstructing character evolution on polytomous cladograms. Cladistics 5:365–377.
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.
Miyamoto, M. M., and W. M. Fitch. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.
Moore, W. S., and V. R. DeFilippis. 1997. The window of taxonomic resolution for avian phylogenies based on mitochondrial cytochrome *b* DNA sequences. Pp. 83–119 *in* D. P. Mindell, ed. Avian molecular evolution and systematics. Academic Press, San Diego.
Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.
Penny, D., L. R. Foulds, and M. D. Hendy. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature 297:197–200.
Poe, S. 2003. Evaluation of the strategy of long branch subdivision to improve accuracy of phylogenetic methods. Syst. Biol. 52: 423–428.
Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.
Prychitko, T. M., and W. S. Moore. 2003. Alignment and phylogenetic analysis of β-fibrinogen intron 7 sequences among avian orders reveal conserved regions within the intron. Mol. Biol. Evol. 20:762–771.
Romer, A. S. 1966. Vertebrate paleontology. Univ. of Chicago Press, Chicago, IL.
Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. 9: 945–967.
Sibley, C. G., and J. E. Ahlquist. 1990. Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven, CT.
Slowinski, J. B. 2001. Molecular polytomies. Mol. Phylogenet. Evol. 19:114–120.
Sorenson, M. D., E. Oneal, J. Garcia-Moreno, and D. P. Mindell. 2003. More taxa, more characters: the hoatzin problem is still unresolved. Mol. Biol. Evol. 20:1484–1498.
Stanley, S. E. 1999. Higher order avian systematics: contributions of mitochondrial and nuclear DNA to questions in avian evolution. Ph.D. diss. Dept. of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY.
Stanley, S. E., and J. Cracraft. 2002. Higher-level systematic analysis of birds: current problems and possible solutions. Pp. 31–43 *in* R. DeSalle, G. Giribet, and W. Wheeler, eds. Molecular systematics and evolution: theory and practice. Birkhauser Verlag, Basel.
Swofford, D. L. 1999. PAUP*: phylogenetic analysis using parsimony (*and other methods). Ver. 4. Sinauer, Sunderland, MA.
Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. Hillis, C. Moritz, and B. K. Mable, eds. Molecular systematics. Sinauer, Sunderland, MA.
Tateno, Y., M. Nei, and F. Tajima. 1982. Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. J. Mol. Evol. 18:387–404.
van Tuinen, M., C. G. Sibley, and S. B. Hedges. 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., D. B. Butvill, J. A. Kirsch, and S. B. Hedges. 2001. Convergence and divergence in the evolution of aquatic birds. Proc. R. Soc. Lond. B 268:1345–1350.
Walsh, H. E., and V. L. Friesen. 2001. Power and stochasticity in the resolution of soft polytomies: a reply to Braun et al. Evolution 55:1264–1266.
Walsh, H. E., M. G. Kidd, T. Moum, and V. L. Friesen. 1999. Polytomies and the power of phylogenetic inference. Evolution 53:932–937.
Wolfe, K. H., and D. C. Shields. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713.
Yang, Z. 2000. Phylogenetic analysis by maximum likelihood (PAML). Ver. 3.0. University College London, London.

Corresponding Editor: J. Wiens