*Syst. Biol.* 66(4):857–879, 2017

DOI:10.1093/sysbio/syx041
Advance Access publication March 27, 2017

# Why Do Phylogenomic Data Sets Yield Conflicting Trees? Data Type Influences the Avian Tree of Life more than Taxon Sampling

Sushma Reddy[1], Rebecca T. Kimball[2], Akanksha Pandey[2], Peter A. Hosner[2,3], Michael J. Braun[4,5], Shannon J. Hackett[6], Kin-Lan Han[2], John Harshman[7], Christopher J. Huddleston[8], Sarah Kingston[4,5,9], Ben D. Marks[6], Kathleen J. Miglia[10], William S. Moore[10], Frederick H. Sheldon[11], Christopher C. Witt[12], Tamaki Yuri[2,13], and Edward L. Braun[2,14,*]

[1]*Biology Department, Loyola University Chicago, 1032 West Sheridan Road, Chicago, IL 60660, USA;* [2]*Department of Biology, University of Florida, Gainesville, FL 32607, USA;* [3]*Florida Museum of Natural History, University of Florida, Gainesville, FL 32607, USA;* [4]*Behavior, Ecology, Evolution, and Systematics Program, University of Maryland, College Park, MD 20742, USA;* [5]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution—MRC 163, PO Box 37012, Washington, DC 20013-7012, USA;* [6]*Zoology Department, Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, IL 60605, USA;* [7]*4869 Pepperwood Way, San Jose, CA 95124, USA;* [8]*Collections Program, National Museum of Natural History, Smithsonian Institution, 4210 Silver Hill Road, Suitland, MD 20746, USA;* [9]*Bowdoin College, Department of Biology and Coastal Studies Center, 6500 College Station, Brunwick, ME 04011, USA;* [10]*Department of Biological Sciences, Wayne State University, 5047 Gullen Mall, Detroit, MI 48202, USA;* [11]*Museum of Natural Science and Department of Biological Sciences, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803, USA;* [12]*Department of Biology and Museum of Southwestern Biology, University 15 of New Mexico, Albuquerque, New Mexico 87131, USA;* [13]*Sam Noble Museum, University of Oklahoma, 2401 Chautauqua Avenue, Norman, OK 73072, USA; and* [14]*Genetics Institute, University of Florida, Gainesville, FL 32607, USA;*

[*]*Correspondence to be sent to: Department of Biology, University of Florida, Gainesville, FL 32607, USA; E-mail: ebraun68@ufl.edu*

*Sushma Reddy, Rebecca T. Kimball, and Edward L Braun contributed equally to this article.*

*Received 15 October 2016; reviews returned 6 January 2017; accepted 22 March 2017*
*Associate Editor: John McCormack*

*Abstract*.—Phylogenomics, the use of large-scale data matrices in phylogenetic analyses, has been viewed as the ultimate solution to the problem of resolving difficult nodes in the tree of life. However, it has become clear that analyses of these large genomic data sets can also result in conflicting estimates of phylogeny. Here, we use the early divergences in Neoaves, the largest clade of extant birds, as a "model system" to understand the basis for incongruence among phylogenomic trees. We were motivated by the observation that trees from two recent avian phylogenomic studies exhibit conflicts. Those studies used different strategies: 1) collecting many characters [~42 mega base pairs (Mbp) of sequence data] from 48 birds, sometimes including only one taxon for each major clade; and 2) collecting fewer characters (~0.4 Mbp) from 198 birds, selected to subdivide long branches. However, the studies also used different data types: the taxon-poor data matrix comprised 68% non-coding sequences whereas coding exons dominated the taxon-rich data matrix. This difference raises the question of whether the primary reason for incongruence is the number of sites, the number of taxa, or the data type. To test among these alternative hypotheses we assembled a novel, large-scale data matrix comprising 90% non-coding sequences from 235 bird species. Although increased taxon sampling appeared to have a positive impact on phylogenetic analyses the most important variable was data type. Indeed, by analyzing different subsets of the taxa in our data matrix we found that increased taxon sampling actually resulted in increased congruence with the tree from the previous taxon-poor study (which had a majority of non-coding data) instead of the taxon-rich study (which largely used coding data). We suggest that the observed differences in the estimates of topology for these studies reflect data-type effects due to violations of the models used in phylogenetic analyses, some of which may be difficult to detect. If incongruence among trees estimated using phylogenomic methods largely reflects problems with model fit developing more "biologically-realistic" models is likely to be critical for efforts to reconstruct the tree of life. [Birds; coding exons; GTR model; model fit; Neoaves; non-coding DNA; phylogenomics; taxon sampling.]

## Introduction

Advances in genomic data acquisition and bioinformatics have led to the genesis of a novel field of evolutionary biology, phylogenomics (Delsuc et al. 2005). Phylogenomics was expected to signal the "end of incongruence" in phylogenetics (Gee 2003) and, potentially, to represent the ultimate solution for resolving "bushes" (soft polytomies) in the tree of life (Delsuc et al. 2005; Rokas and Carroll 2006). Recent advances in sequencing technologies have made it possible to obtain very large phylogenetic data sets by using methods such as sequence capture (Faircloth et al. 2012; Prum et al. 2015; Hosner et al. 2016), transcriptome sequencing (Misof et al. 2014; Wickett et al. 2014), and whole-genome sequencing (Jarvis et al. 2014). Thus, phylogenomics seemed poised to fulfill this promise to resolve the tree of life. However, analyses of large data matrices have sometimes yielded incongruent topologies, emphasizing that data collection alone is not sufficient to reach this goal. Herein, we assess the potential of phylogenomics to accurately resolve phylogenetic bushes by examining an extremely difficult phylogenetic problem, the early evolution of birds.

The ability of phylogenomics to resolve the tree of life may be limited by the existence of bias in phylogenetic estimation (Jeffroy et al. 2006). Sources of bias include long-branch attraction (Felsenstein 1978a; Hendy and Penny 1989), base composition convergence (Jeffroy et al. 2006; Katsu et al. 2009), and topological errors that can emerge when the sequences being analyzed

evolved on a mixture of gene trees (Kubatko and Degnan 2007; Matsen and Steel 2007; Roch and Steel 2014). If bias exists, phylogenetic analyses will converge on an incorrect result with increasing certainty as data are added (Swofford et al. 2001). Increased taxon sampling is one way to improve phylogenetic accuracy, helping to overcome some biases, and there are many examples of "problem" trees that have been improved by adding taxa (Hillis 1996; Pollock et al. 2002; Zwickl and Hillis 2002; Soltis et al. 2004); however, increased taxon sampling is not a panacea (Poe and Swofford 1999; Braun and Kimball 2002), and there are cases where appropriate taxa simply do not exist (e.g., no organisms exist that can break up the long branches leading to *Amborella* in angiosperms, to coelacanths in vertebrates, or to the hoatzin in birds).

The challenge posed by bias in phylogenetic analyses raises a fundamental question: can phylogenomics actually resolve difficult nodes in the tree of life? This study asks a related but more focused and practical question: are the evolutionary models implemented in available phylogenetic software sufficient to resolve difficult phylogenetic trees accurately using large sequence data sets? Here, we seek insight into this question by testing specific hypotheses about the early evolution of birds, a phylogenetic problem that has remained intractable despite a steady stream of large-scale efforts to resolve relationships among major avian lineages (e.g., Hackett et al. 2008; Kimball et al. 2013; McCormack et al. 2013; Jarvis et al. 2014; Prum et al. 2015).

### *Birds as a Model System for Phylogenomics*

> In view of the continuing absence of trustworthy information on the relationship of the highest categories of birds to each other it becomes strictly a matter of convention how to group them into orders. Science ends where comparative morphology, comparative physiology, comparative ethology have failed us after nearly 200 years of efforts. The rest is silence.—Stresemann (1959).

It has been appreciated for some time that the most species-rich avian clade, Neoaves (Fig. 1), underwent an extremely rapid radiation (Groth and Barrowclough 1999; Cracraft et al. 2004; Ericson et al. 2006). In fact, Poe and Chubb (2004) suggested that the base of Neoaves could represent a hard polytomy (simultaneous speciation events) based on power analyses. The hard polytomy hypothesis was consistent with the observation that ordinal relationships in trees published prior to 2008 showed only slightly greater congruence than expected in random trees (cf. Table 4 in Chojnowski et al. 2008). However, Hackett et al. (2008) falsified the Poe and Chubb (2004) hard polytomy hypothesis, at least partially, when they analyzed 19 unlinked nuclear loci and found a topology (the "Early Bird tree") with many well-supported nodes that was also robust to gene jackknifing. Much of the structure in the Early Bird topology has been corroborated by subsequent analyses that used independent sets of loci (Wang et al. 2012; Kimball et al. 2013; McCormack et al. 2013; Smith et al. 2013), including some analyses based on very large data matrices (e.g., Fig. 1a Jarvis et al. 2014 and Fig. 1b Prum et al. 2015). These advances have led to increased confidence in relationships among the major lineages in Neoaves (see Fig. 1 and Supplementary Fig. S1, available on Dryad at http://dx.doi.org/10.5061/dryad.6536v).

Continued data collection for large-scale phylogenetic studies, however, has not resulted in a consistent resolution of the deep branches of the bird tree. Specifically, the Jarvis et al. (2014) "total evidence nucleotide tree" (TENT; Fig. 1a), based on 42 Mbp of data extracted from 48 complete avian genomes, and the Prum et al. (2015) (Fig. 1b) tree, based on 0.4 Mbp of data from 259 loci obtained by sequence capture (anchored hybrid enrichment) and sampled for 198 bird species, exhibit a number of conflicts. For instance, these two trees differ in the groupings of the earliest split within Neoaves (compare Fig. 1a and 1b) and the presence of a clade of mainly aquatic birds in the Prum tree that directly contradicts the Jarvis TENT. Both Jarvis et al. (2014) and Prum et al. (2015) report strong support for all of their relationships.

The conflicts between the Jarvis TENT and Prum tree are surprising given the size of the data matrices analyzed in each study. Prum et al. (2015) suggested that the differences reflect greater taxon sampling in their tree. If true, this would provide a striking example of the assertion that "…a much broader sample of taxa (perhaps sequenced for far less than full genomes) will result in a much more accurate estimate of phylogeny than will complete genomes of only a small number of taxa" (Hillis et al. 2003). Although there are cases where adding taxa reduces support and/or results in decreased phylogenetic accuracy (e.g., Poe and Swofford 1999; Sanderson and Wojciechowski 2000; Braun and Kimball 2002; Meiklejohn et al. 2014), adding taxa usually improves phylogenetic accuracy (reviewed by Heath et al. 2008). Moreover, one could argue that the observation that many different topologies result from analyses of subsets of the Jarvis et al. (2014) data, despite the use of the same taxa and analytical methods (e.g., Fig. 1c vs. Fig. 1d), reflects instability due to poor taxon sampling. This conclusion might lead one to embrace the Prum et al. (2015) topology given its denser taxon sampling. However, we believe it is important to examine alternative hypotheses that can explain the differences between the Jarvis TENT and Prum tree.

$H_1$: Insufficient signal—The size of the Prum et al. (2015) data matrix (the number of sites, the number of genes, or both) is insufficient to converge on the topology expected given complete genomes. We note that it is possible that the Jarvis et al. (2014) TENT data matrix (~3.2% of the typical avian genome) may not be large enough for analyses to converge on the hypothetical "whole-genome

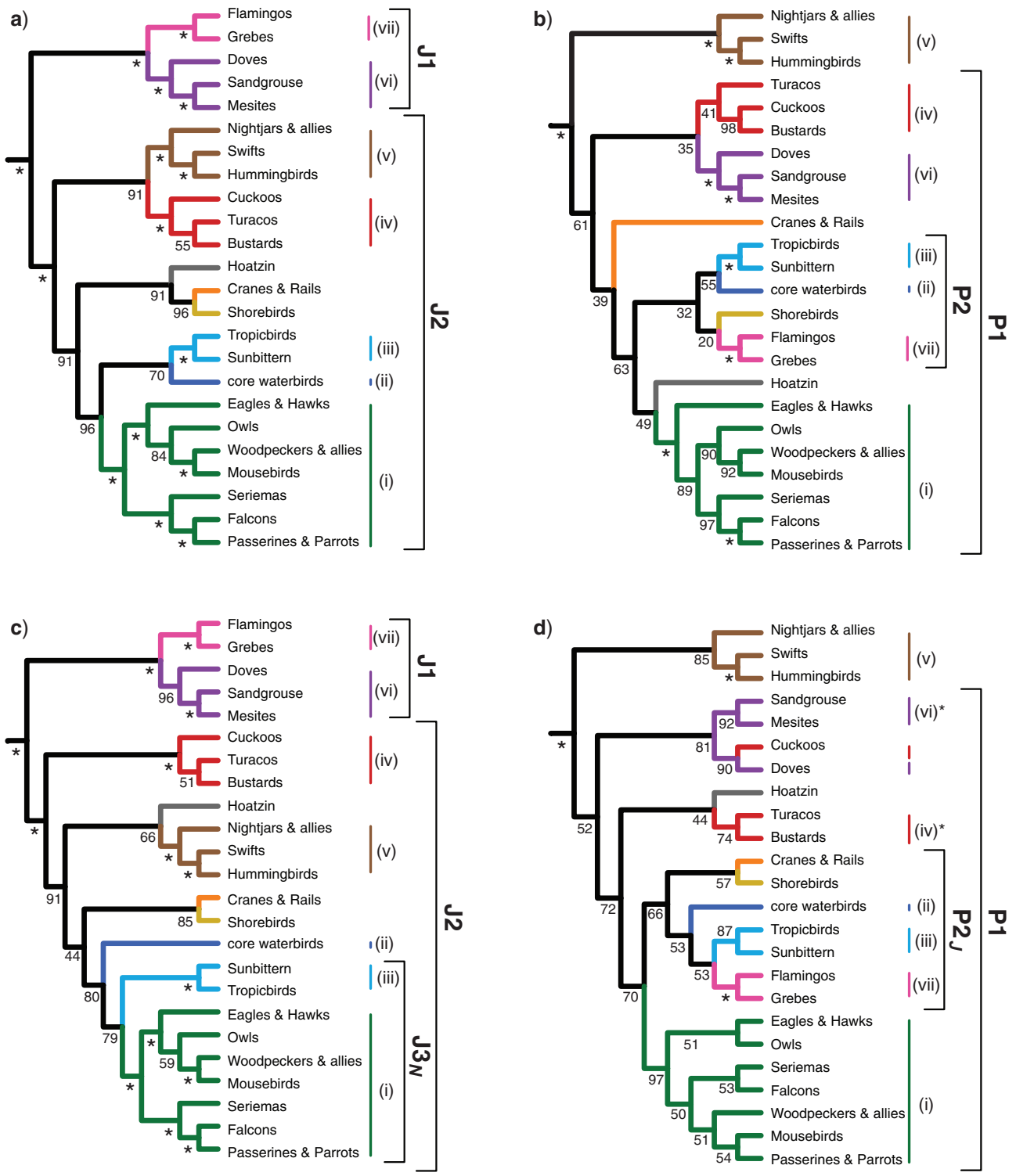


FIGURE 1. Recent estimates of Neoaves phylogeny. a) The Jarvis et al. (2014) TENT, based on 42 Mbp (46% intron, 32% exon, 22% non-coding UCEs) extracted from complete genomes. The "magnificent seven" (clades i–vii) and the "Jarvis indicator clades" (J1 and J2) are highlighted. b) The Prum et al. (2015) tree, based on 0.4 Mbp (82.5% exonic) obtained by sequence capture with the "Prum indicator clades" (P1 and P2) indicated. c) Jarvis et al. (2014) intron tree, with the "Jarvis non-coding indicator clade" ($J3_N$) identified. d) Jarvis et al. (2014) c12 coding exon tree with a Prum indicator clade (P1) and an aquatic/semiaquatic clade similar to the clade P2 ($P2_J$). The simplified names (J and P) are used here to indicate clades that appear to be diagnostic for each of the two topologies; we provide full clade names from Jarvis et al. (2014) and Prum et al. (2015) in Supplementary File S3. We present ordinal names (based on Cracraft 2013) in Table 1 and in Supplementary File S3. Bootstrap support (based on the original publications) is below branches; "*" indicates 100% support.

topology". Indeed, it is possible that even whole-genome data are insufficient to resolve avian phylogeny (this would be indistinguishable from a hard polytomy).

$H_2$: Hard polytomy—The base of Neoaves includes at least some major lineages that form a hard polytomy, as suggested by Poe and Chubb (2004). This hypothesis is important to reconsider because Suh (2016) recently suggested that a hard polytomy (albeit one involving fewer lineages than suggested by Poe and Chubb 2004) could explain the observed conflicts among phylogenomic studies. The expected set of gene trees given a hard polytomy reflects only the random sorting of alleles into descendant lineages (cf. Slowinski 2001).

$H_3$: Taxon sampling—The increased taxon sampling in the Prum et al. (2015) study relative to Jarvis et al. (2014) results in different estimates of the bird tree. This hypothesis predicts that analyses of data matrices with increased taxon sampling should converge on the Prum tree topology.

$H_4$: Data-type effects—The differences between the Prum tree and Jarvis TENT reflect the use of different data types [i.e., primarily coding exons for the Prum tree vs. a mixture of exons, introns, and non-coding ultraconserved elements (UCEs) in the Jarvis TENT]. Although this hypothesis has not been articulated explicitly in previous studies, Jarvis et al. (2014) did note differences among the trees resulting from analyses of different data types (e.g., Fig. 1c and 1d). This hypothesis predicts that estimates of phylogeny obtained by analyzing the same data type will converge on the same topology regardless of taxon sampling.

Given the interest in the multispecies coalescent methods for phylogenetic analyses (e.g., Edwards 2009; Edwards et al. 2016) it may seem desirable to add an $H_5$ that invokes gene tree–species tree discordance. However, this hypothesis is not viable as an explanation for the observed differences between the Prum tree and Jarvis TENT. Both trees reflect analyses of large samples of concatenated genes from many locations in the genome. Even if the avian species tree lies in the anomaly zone, that part of treespace where the most common gene tree disagrees with the species tree (Degnan and Rosenberg 2006; Rosenberg 2013), analyses of concatenated data are still expected to converge on the same estimate of phylogeny as data are increased. Of course, the tree resulting from analyses of concatenated data would be an incorrect estimate of phylogeny if the true species tree lies in the anomaly zone. But the observation that motivated this study (incongruent trees generated by analyses of concatenated data) would not be expected because the same topology is still expected to emerge from concatenated analyses whenever a sufficient number of loci are sampled (Kubatko and Degnan 2007). It is possible that the 259 loci sampled by Prum et al. (2015) are insufficient to converge on the same tree as the Jarvis TENT, which is based on more than 12,000 loci. However, that would be a special case of $H_1$. Under that scenario, the underlying reason for the observed incongruence would indeed be discordance among gene trees, but the only reason we observe that incongruence is that the sample of gene trees for the 259 loci in the Prum et al. (2015) data set do not provide sufficient information to converge on the tree expected given all orthologous genes in bird genomes. For this reason, we do not consider discordance among gene trees to be a distinct explanation for the differences between the Prum and Jarvis trees.

The high level of support that Prum et al. (2015) report for their tree suggests we should reject $H_1$ (insufficient signal). However, the high degree of support reported in Fig. 1 of Prum et al. (2015) reflects the use of Bayesian posterior probabilities. Bayesian MCMC methods can overestimate support (Alfaro et al. 2003; Simmons et al. 2004), especially when the model of evolution used in the analysis is misspecified (Buckley 2002) or incorrect priors are used (Yang and Rannala 2005). Maximum-likelihood (ML) bootstrap analysis of the Prum et al. (2015) data (presented in their Supplementary Information) yields much more limited (often <50%) support for many important clades (Fig. 1b). This limited support includes some clades that correspond to critical differences between the Prum tree and Jarvis TENT (i.e., clades P1 and P2 in Fig. 1b). Thus, the degree of confidence in the clades recovered by Prum et al. (2015) that contradict the deep clades in the Jarvis TENT (i.e., clades J1 and J2 in Fig. 1a) depends on whether the priors and the fit of the model used for the Bayesian analysis were sufficient to yield accurate posterior probabilities.

Congruence offers a more appropriate test of $H_1$. $H_1$ postulates that the number of sites and/or loci in a specific data set is not sufficient for analyses of those data to converge on the topology expected given all genes. Thus, trees based on two different data sets are expected to differ for trivial reasons if $H_1$ is true for one (or both) the data sets. We can reject $H_1$ for a specific data set if estimates of phylogeny contain one or more specific clades that were defined *a priori* and the probability of observing the specified clade (or clades) by chance alone is low (e.g., finding a specific clade given a rooted three-taxon tree is unsurprising but finding a specific five-taxon clade in a rooted 10-taxon tree would be highly significant). The early evolution of Neoaves can be viewed as a rooted 10-taxon tree. This reflects the fact that there are seven superordinal clades, which we call the "magnificent seven," in a strict consensus of the Jarvis TENT and Prum tree (Fig. 1 and Supplementary Fig. S1, available on Dryad). This places 10 lineages in play when the magnificent seven are combined with the three "orphan orders" (shorebirds, cranes, and the monotypic hoatzin) that cannot be placed in a strongly corroborated superordinal clade. There are more than 34 million possible rooted 10-taxon trees (Felsenstein 1978b), and very few of those trees split Neoaves into either clades

J1 and J2 or clades P1 and P2 (Supplementary File S1 and Supplementary Fig. S1, available on Dryad). Thus, recovery of a tree with either of those sets of clades would be highly unlikely to occur by chance. We proposed these "indicator clades" based on comparisons of the trees in Jarvis et al. (2014) and Prum et al. (2015); thus, their use to test $H_1$ requires a novel data set. If an estimate of phylogeny obtained by analyzing a novel data matrix contains clades J1 and J2 or clades P1 and P2 we can argue, based on congruence with either the Jarvis TENT or the Prum tree, that $H_1$ is unlikely to be correct.

A sufficient test of $H_2$ (the hypothesis that the existence of a hard polytomy explains the differences between the Jarvis TENT and Prum tree) is difficult to undertake. The existence of a hard polytomy can be corroborated if one finds an unresolved topology and establishes that the amount of data analyzed is sufficient to resolve the relevant internal branches (see Braun and Kimball 2001). However, analyses of simulated data generated on "star phylogenies" (trees with hard polytomies) sometimes yield trees that include clades with high support (Suzuki et al. 2002; Steel and Matsen 2007). This "star tree paradox" has been observed in both Bayesian (e.g., Lewis et al. 2005) and ML analyses (Susko 2008). However, $H_2$ is certainly not untestable. Moreover, our articulation of $H_2$ actually has two parts: 1) the base of Neoaves is a hard polytomy for at least some lineages (e.g., the magnificent seven and the orphan orders); and 2) the hard polytomy explains the incongruence among phylogenomic estimates of the bird tree. The star tree paradox occasionally results in high support for a random resolution but analyses of different samples of the genome are expected to yield contradictory topologies (often without support). Thus, the recovery of many different trees when independent subsets of the genome are analyzed, even if some clades appear to be strongly supported in some (but never all) of the trees, would be consistent with the hard polytomy hypothesis. The argument from congruence with prior estimates of phylogeny (i.e., the presence of indicator clades based on the Jarvis TENT and Prum tree in this study) that can be used to assess $H_1$ can also be used to test $H_2$. Thus, $H_2$ is unlikely to be correct if we analyze another data matrix comprising an arbitrarily chosen and independent set of loci and it yields a tree with either J1 and J2 or P1 and P2.

In contrast to $H_1$ and $H_2$, for which tests are subtle, assessing $H_3$ (the taxon sampling hypothesis) is straightforward. All that is necessary is to subsample taxa from the study with more extensive taxon sampling. Prum et al. (2015) performed such a test and found that the resulting reduced-taxon tree [hereafter "Prum (Javis taxon set)"] lacked clade P1. Although Prum et al. (2015) interpreted this as support for $H_3$ it is important to note that clade P2 was present in the Prum (Jarvis taxon set) tree. Since the presence of clade P2 is incompatible with the existence of clades J1 and J2 (see Fig. 1a), simply reducing the number of taxa included in analyses of the Prum et al. (2015) data matrix is not sufficient to shift the topology of the Prum tree to that of the Jarvis TENT (if we view the presence of clades J1 and J2 as markers for a "Jarvis-like" tree). Since the taxon-sampling experiment conducted by Prum et al. (2015) is ambiguous, a different test of $H_3$ is necessary. It is possible to conduct such a test and corroborate (or falsify) $H_3$, at least in principle, by examining the results of analyses using an independent data matrix with a more extensive taxon sampling.

The final hypothesis, $H_4$, postulates that data type is a critical variable. Both our group and others (Jarvis E., personal communication) observed that the Prum tree exhibits striking similarities, such as the presence of clade P1, to the Jarvis et al. (2014) coding exon tree (Fig. 1d). In addition to clade P1, the Jarvis et al. (2014) exon tree contains an aquatic/semiaquatic clade ($P2_J$ in Fig. 1). There are *a priori* reasons to believe that the aquatic/semiaquatic clades P2 and $P2_J$ are effectively the same. Jarvis et al. (2014) sampled a single crane and Prum et al. (2015) specifically highlighted two orders as undersampled by Jarvis et al. (2014): cranes and rails (Gruiformes) and shorebirds (Charadriiformes). Thus, $P2_J$ could reflect rogue behavior of the single crane included in the Jarvis et al. (2014) study. The Prum et al. (2015) data matrix largely (>80%) comprises coding data (Supplementary File S2, available on Dryad) and is therefore largely the same data type as the Jarvis et al. (2014) exon tree. In contrast, the data used to generate the Jarvis TENT (Fig. 1a) are 68% non-coding, a fact that is especially important when we consider that the Jarvis TENT is much more congruent with those trees in Jarvis et al. (2014) that were estimated using only non-coding data [i.e., the intron tree (Fig. 1c) and the UCE tree (see Fig. 4b in Jarvis et al. (2014)]. These observations are the basis for our hypothesis ($H_4$) that data type could be an important variable.

There are several considerations when using indicator clades. First, a general consideration is that indicator clades should be chosen before analyses of a new data matrix and they should reflect comparisons of "independent trees" (i.e., trees estimated using data from non-overlapping subsets of the genome). Clade J2 is present in two independent trees (the intron and UCE trees from Jarvis et al. 2014), as was a second clade ($J3_N$), which is present only in trees that are based exclusively on non-coding data (e.g., Fig. 1c). In contrast, the other indicator clades (J1, P1, and $P2/P2_J$) were not identified using comparisons of completely independent trees. However, those indicator clades are present in multiple published trees with large amounts of non-overlapping data (Supplementary File S1, available on Dryad). Specifically, clade J1 is present in three different Jarvis trees that are dominated by non-coding data, the intron tree, TENT, and a "whole-genome tree" based on 322-Mbp of many different data types (see Supplementary File S1, available on Dryad). Likewise, P1 and $P2/P2_J$ are present in the Prum tree and two different Jarvis exon trees. Second, a consideration specific to this study is that one must assume at least a weak data-type effect in addition to a taxon-sampling

effect to use P1 and $P2_J$ as indicator clades for $H_3$ since both are present in the Jarvis et al. (2014) exon trees (e.g., Fig. 1d). This data-type effect could be the lower rate of substitution in exons, which could make analyses of exons less susceptible to long-branch attraction (thus reducing the need for increased taxon sampling). We can address both of these concerns by examining the tree distances among the estimates of phylogeny to see if trees cluster by taxon sampling ($H_3$) or data type ($H_4$). Thus, we view the indicator clade and tree distance approaches as complementary.

Here, we test the hypotheses described above, with a special focus on the taxon sampling ($H_3$) and data type ($H_4$) hypotheses, by evaluating analyses of a novel data matrix in light of the results published by Jarvis et al. (2014) and Prum et al. (2015). Specifically, we recognized that performing analyses of a matrix that primarily comprises non-coding data (comparable to the Jarvis TENT) but with taxon sampling comparable to Prum et al. (2015) would result in a set of analyses analogous to a 2 × 2 factorial experiment (Fig. 2) to assess the impact of taxon sampling and data type. To produce the novel data matrix, we generated a largely intronic data set by combining published data matrices (from Hackett et al. 2008; Braun et al. 2011; Wang et al. 2012; Kimball et al. 2013; Smith et al. 2013) and adding novel genes and taxa to the matrix. This resulted in a data set we refer to as Early Bird II (EB2) that comprises 54 loci sampled from up to 235 bird species (Table 1 and Supplementary File S3, available on Dryad). We also performed taxon sampling experiments, reducing the EB2 taxon set to the same 48 taxa as Jarvis et al. (2014) and to a second taxon set of intermediate size (120 taxa) that bisected long branches as much as possible. We examined the presence or absence of the mutually exclusive indicator clades present in the Jarvis TENT (clades J1 and J2) or in the Prum tree (clades P1 and $P2/P2_J$). We also assessed the distances among the estimates of phylogeny obtained as part of this study and the Jarvis et al. (2014) trees [including the Suh et al. (2015) transposable element (TE) insertion tree based on the Jarvis data] and the Prum et al. (2015) trees. The results of these analyses have implications for the theory and practice of phylogenomics that extend beyond the information they provide about the avian tree of life.

## Materials and Methods

### *Data Collection and Sequence Alignment*

We assembled a 54-locus data set (Supplementary File S4, available on Dryad) by producing new sequences and extracting data from draft genome assemblies (listed in Supplementary File S3, available on Dryad) and combining those novel data with sequences collected in previous studies (Hackett et al. 2008; Braun et al. 2011; Wang et al. 2012; Kimball et al. 2013; Smith et al. 2013). We added four loci that were not sampled prior to this study and 66 species (both for novel loci and the previously sampled loci). We generated PCR amplicons for the new sequences using standard methods described by Kimball et al. (2009) and assembled contigs using Sequencher™ 4.1 (Gene Codes Corp.). The new sequences generated by PCR have been deposited in GenBank with accession numbers KY762311-KY763958. We also extracted data from genome sequences using a custom pipeline. Briefly, we used nhmmer (Wheeler and Eddy 2013) to search genome sequences using profile hidden Markov models (HMMs) for each gene region as queries. We generated the query HMMs using the alignments of data generated by PCR and Sanger sequencing. This pipeline (available from https://github.com/aakanksha12/Extract_seq) allowed us to extract sequences from genome assemblies regardless of the quality of the annotation for those genome sequences. Although many published avian genomes, like those used to generate the data set used in Jarvis (described by Zhang et al. 2014), are relatively well annotated this is not the case for all avian genome assemblies. Our pipeline allowed us to include these low-coverage genome assemblies, like the Gunnison sage-grouse (*Centrocercus minimus*) and Clark's nutcracker (*Nucifraga columbiana*) genome sequences (Card et al. 2014).

Since some taxa with genomic data were also in published matrices, our alignment had sequence data from 258 individuals of 235 species (Supplementary File S3, available on Dryad). Our data set sampled diverse taxa representing members of all orders of birds as circumscribed in all modern checklists (Cracraft 2013; Clements et al. 2015; Gill and Donsker 2016); we used the Howard and Moore 4th edition checklist for species names and other taxonomic information (Dickinson and Remsen 2013; Dickinson and Christidis 2014). We assembled four data matrices, each of which included data from all 54 loci: 1) ALL (258 taxa, 137,463 bp): all data, including multiple individuals of some species; 2) EB2 (235 taxa, 137,324 bp): all species with duplicate individuals removed (we selected sequences extracted from genome sequences whenever available); 3) KIM (120 taxa, 118,233 bp): data from species sampled by Kimball et al. (2013), which attempted to maximally subdivide long branches in the avian tree, supplemented with data extracted from genome sequences; and 4) JAR (48 taxa, 91,483 bp): the species used by Jarvis et al. (2014). All of these data matrices contain substantial missing data, reflecting the large number of insertions and deletions (indels) that accumulate in introns and untranslated regions (UTRs) and the limitations of PCR-based sampling. The specific taxa included in each taxon set are listed in Supplementary File S3, available on Dryad. We also conducted analyses after excluding the limited amount of coding exon data in the EB2 matrix.

We aligned our sequences using MUSCLE v. 3.8.31 (Edgar 2004), which we chose because it appeared to generate the best multiple sequence alignments in our previous analyses of intronic data (Wang et al. 2012). Afterward, we examined each alignment manually

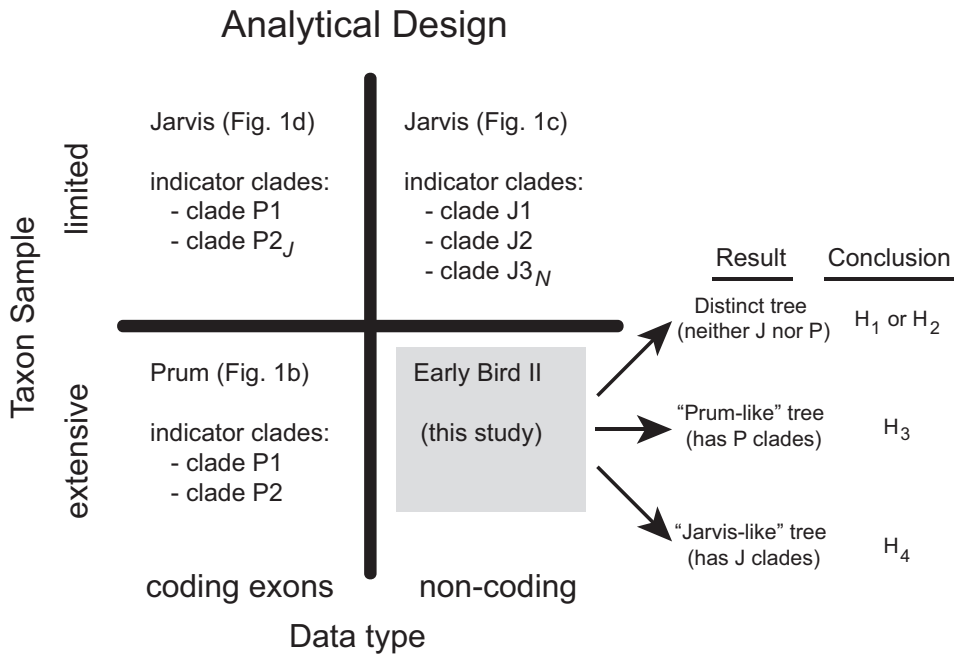


FIGURE 2. Our analyses are analogous to a 2 × 2 factorial examination of the impact of data type (coding exon vs. non-coding) and taxon sampling. We conducted analyses of various versions of the EB2 data matrix. Specifically, we predicted that analyses of EB2 will yield a "Jarvis-like tree" (one with clades J1 and J2) if $H_4$ is true, a "Prum-like tree" (one with clades P1 and $P2/P2_J$) if $H_3$ is true, and a different (and potentially poorly supported) tree if $H_1$ or $H_2$ are correct. $H_4$ also predicts that trees based on the same data type will cluster in treespace whereas $H_3$ predicts that trees with similar taxon sampling will cluster.

to identify exon–intron borders and assembled the alignments of each locus to generate the annotated data matrix (Supplementary File S5, available on Dryad). We confirmed that all sequences sampled from individual species clustered with conspecifics (when available) by analyzing each gene with the ALL taxon set (Supplementary File S6, available on Dryad); subsequent phylogenetic analyses were conducted using the other three data matrices. This allowed us to test the impact of taxon sampling on the estimation of avian phylogeny for these data.

We compared the EB2, Jarvis, and Prum data sets in two ways. For these analyses we downloaded the Jarvis data from http://gigadb.org/dataset/101041 (Jarvis et al. 2015) and we obtained the Prum data from https://zenodo.org/record/28343. First, we assessed the amount of variation in base composition for each locus in all three data sets. We used $\Delta GC_{inf}$, which we define as the difference between the GC-content of the parsimony-informative sites for the most GC-rich taxon and the least GC-rich taxon, to assess GC-content variation among loci. Second, we conducted analyses of the EB2 and Prum data after excluding those sites that overlapped with the Jarvis et al. (2014) data. To do this, we identified the sequences in the EB2 and Prum data sets that were also in the Jarvis TENT data set. The Jarvis TENT data comprise 8251 coding regions, at least some of the introns from 2516 of those protein-coding loci, and 3769 non-coding UCEs. We conducted BLASTN (Camacho et al. 2008) searches of the exons and introns included in the Jarvis TENT data using EB2 and Prum et al. (2015) sequences as queries. There was overlap for 19 of the 54 EB2 loci (Supplementary File S4, available on Dryad), although the overlap was limited to the small amounts of coding data in 11 of those 19 loci. We found that 146 of 250 Prum et al. (2015) loci that include coding data overlapped with the Jarvis TENT data (Supplementary File S2, available on Dryad; note that 9 of the 259 Prum loci are conserved non-coding regions).

### *Phylogenetic Analyses*

We used RAxML v. 8.2.3 (Stamatakis 2014) and IQ-TREE v. 1.3.10 (Nguyen et al. 2015) for ML analysis of concatenated data. We used three different partitioning schemes: 1) full: each region within each locus [i.e., each intron, the UTR (when present), and the coding exons (all exons from a treated as a single partition because the individual exons were short)] served as a partition; 2) PF: the partitions generated by the rcluster algorithm in PartitionFinder v 1.1 (Lanfear et al. 2012) using the Bayesian information criterion (BIC for model selection); and 3) Unpart: unpartitioned analyses. These partitions are available in Nexus files for the sequence alignments we analyzed (Supplementary File S5, available on Dryad). We assessed support using the rapid bootstrap in RAxML (Stamatakis et al. 2008) and ultrafast bootstrap (Minh et al. 2013) in IQ-TREE. We also conducted one RAxML analysis using the standard bootstrap with the GTR+Γ model and PF partitioning

TABLE 1. Taxon sampling for EB2, Jarvis et al. (2014), and Prum et al. (2015).

| Taxon set | | | | | |
|---|---|---|---|---|---|
| Order | Common name | JAR | KIM | EB2 | Prum |
| PALAEOGNATHAE | | | | | |
| Struthioniformes | Ostrich | 1 | 1 | 1 | 1 |
| Rheiformes | Rheas | — | 2 | 2 | 1 |
| Apterygiformes | Kiwis | — | — | 1 | 1 |
| Casuariiformes | Emu and Cassowaries | — | 1 | 2 | 2 |
| Tinamiformes | Tinamous | 1 | 2 | 5 | 4 |
| GALLOANSERES | | | | | |
| Anseriformes | Waterfowl | 1 | 3 | 8 | 7 |
| Galliformes | Landfowl | 2 | 7 | 12 | 9 |
| NEOAVES | | | | | |
| Phoenicopteriformes | Flamingoes | 1 | 2 | 2 | 1 |
| Podicipediformes | Grebes | 1 | 2 | 2 | 1 |
| Columbiformes | Doves | 1 | 3 | 5 | 5 |
| Mesitornithiformes | Mesites | 1 | 1 | 2 | 2 |
| Pteroclidiformes | Sandgrouse | 1 | 2 | 3 | 2 |
| Otidiformes | Bustards | 1 | 3 | 3 | 1 |
| Cuculiformes | Cuckoos | 1 | 3 | 8 | 4 |
| Musophagiformes | Turacos | 1 | 1 | 3 | 2 |
| Gruiformes | Cranes and rails | 1 | 4 | 8 | 9 |
| Charadriiformes | Shorebirds | 1 | 3 | 14 | 16 |
| Opisthocomiformes | Hoatzin | 1 | 1 | 1 | 1 |
| Caprimulgiformes | Nightjars and allies (includes Hummingbirds and Swifts) | 3 | 9 | 19 | 13 |
| Gaviiformes | Loons | 1 | 2 | 2 | 1 |
| Procellariiformes | Tubenoses | 1 | 2 | 6 | 8 |
| Sphenisciformes | Penguins | 2 | 3 | 2 | 1 |
| Pelecaniformes | Pelicans and allies | 4 | 7 | 19 | 13 |
| Eurypygiformes | Sunbittern and Kagu | 1 | 2 | 2 | 1 |
| Phaethontiformes | Tropicbirds | 1 | 1 | 2 | 1 |
| Accipitriformes | Eagles, Hawks, and allies (includes New World vultures) | 3 | 7 | 10 | 7 |
| Strigiformes | Owls | 1 | 4 | 4 | 2 |
| Coliiformes | Mousebirds | 1 | 3 | 3 | 2 |
| Leptosomiformes | Cuckoo roller | 1 | 1 | 1 | 1 |
| Trogoniformes | Trogons | 1 | 3 | 3 | 2 |
| Bucerotiformes | Hornbills and allies | 1 | 4 | 6 | 4 |
| Coraciiformes | Bee-eaters and allies | 1 | 3 | 9 | 7 |
| Piciformes | Woodpeckers and allies | 1 | 4 | 13 | 10 |
| Cariamiformes | Seriemas | 1 | 2 | 2 | 2 |
| Falconiformes | Falcons | 1 | 4 | 6 | 4 |
| Psittaciformes | Parrots | 2 | 4 | 11 | 6 |
| Passeriformes | Passerines | 5 | 14 | 32 | 44 |
| | Total = | 48 | 120 | 235 | 198 |

Note: Numbers reflect the number of species in each group.

scheme to assess bias (if any) in the fast bootstrap. We used the GTR+Γ model in RAxML and the model chosen by the -m TEST with the BIC option in IQ-TREE (-m TEST assigns the best-fitting model to each partition). We limited the set of models examined by IQ-TREE to submodels of the GTR+I+Γ with one (1ST), two (2ST), or six (6ST) substitution types (i.e., we used the -mset mrbayes option); Γ-distributed rates were approximated using a four-category approximation (the +G4 option in IQ-TREE and the default in RAxML). We also tested the free-rates model (Yang 1995) in IQ-TREE (i.e., we used the -m TESTNEW option). Finally, we conducted analyses of our data after excluding the small amount of coding exon data.

We also conducted phylogenetic analyses using MrBayes 3.2.6 (Ronquist et al. 2012) on the CIPRES Science Gateway (Miller et al. 2010) and ExaBayes v1.4.1 (Aberer et al. 2014). We found that the default settings for MrBayes analyses resulted in poor convergence so we used these modified settings: brlenspr = unconstrained:exponential (100); nrun = 1 nchain = 8 temp = 0.06 (Moyle et al. 2012). We used the GTR+I+Γ model and executed six iterations starting from random starting trees and sampled every 10,000 states. We ran each iteration until it reached the CIPRES walltime (168:00:00). The six runs sampled 1,260,000; 1,760,000; 1,480,000; 2,120,000; 1,610,000; and 1,710,000 states, respectively. We examined parameter estimates

to determine their stationarity in Tracer 1.5, and set the burn-in to 200,000 states, except for run 1 which did not reach stationarity until 650,000 states. When combined, these six iterations resulted in 1368 post burn-in posterior samples. Effective sample sizes for the combined parameter estimates were all greater than 200. Because Prum et al. (2015) used ExaBayes for their Bayesian analyses we also conducted two runs in ExaBayes using methods similar to that study; specifically, we used the GTR+Γ model with default tuning and branch swapping parameters and four Metropolis-coupled chains (three of which were heated). We sampled 1,105,500 and 604,000 states, respectively, for the first and second ExaBayes runs. The ExaBayes runs converged rapidly and examination of the runs using Tracer revealed that the post burn-in parameter estimates for each run had effective sample sizes greater than 200. We summarized the samples from the MCMC chains using sumtrees.py 4.1.0 from the DendroPy package (Sukumaran and Holder 2010).

We believe that it is valuable to conduct comparable analyses using at least two different programs to show that numerical optimization does not have an impact on conclusions. Evidence is accumulating that even minor differences in numerical optimization can result in the recovery of fairly different trees if the data matrix has a rough likelihood surface (e.g., Darriba et al. 2015). Meiklejohn et al. (2016) found conflicts with more than 95% bootstrap support in species tree analyses conducted using input trees generated using different programs. It seems unlikely for differences among trees due to this effect to achieve such high support unless the data matrices used for analyses are large, but it is certainly a concern for phylogenomics. This problem can be viewed as another hypothesis to explain topological differences among trees and it can be falsified by finding that different software packages converge on the same (or very similar) solutions.

To address whether our results were influenced by base compositional variation (or any other deviations from the GTR model), we conducted "squangle" analyses, an invariants method for quartets that is consistent given the general Markov model (GMM) (Holland et al. 2013). We conducted squangle analyses using the python program written by Holland et al. (2013) that is available from http://datadryad.org/resource/doi:10.5061/dryad.2k9j0. We used the settings HCQ = 0, CLEAN_EACH = 1, and EST_INV = 1 (the last setting is to use the GMM+I model) in the squangle.py program for quartet inference. We used two of the supertree methods implemented in clann (Creevey and McInerney 2005) for quartet amalgamation [neighbor joining (NJ; Saitou and Nei 1987) and matrix representation with parsimony (MRP; Baum 1992; Ragan 1992)]. clann only produces an MRP matrix; the actual parsimony analysis for the MRP data was conducted in PAUP* 4.0a149 (Swofford 2016). We bootstrapped the quartets to assess support in the squangle analyses. Since the squangle.py program samples quartets exhaustively it was only feasible to apply it to the relatively small JAR taxon set.

We conducted multispecies coalescent (species tree) analyses using ASTRAL v. 4.10.6 (Mirarab et al. 2014). Briefly, we used RAxML and IQ-TREE to obtain the ML tree and a set of bootstrap trees (100 fast bootstrap replicates for RAxML and 1000 ultrafast bootstrap replicates for IQ-TREE) for each locus. The estimates of gene trees were generated by partitioned analyses separating the introns, exons, and/or UTRs within each locus. ASTRAL analyses were conducted using two strategies: 1) using 54 gene trees as input, one for each locus; and 2) using 51 gene trees as input, one for each locus but excluding BDNF, NGF, and NTF3 (the three loci with the greatest compositional variation; see below).

We measured distances among trees using symmetric RF distances [Robinson and Foulds (1981) distances multiplied by two] that were calculated using PAUP*. We compared "backbone trees," which we defined as trees reduced to the taxa used Jarvis et al. (2014), further reducing three orders (Passeriformes, Psittaciformes, and Pelecaniformes) that are clearly monophyletic based on extensive data to a single lineage. The use of backbone trees allowed us to focus on the relationships among major lineages instead of rearrangements within clades, and it facilitated comparisons of trees with different taxon samples. We used NJ (Saitou and Nei 1987) to cluster the symmetric RF distances among backbone trees.

## Results and Discussion

### *Analyses of Early Bird II Data Matrix*

Analyses of the 54-locus, 235-taxon EB2 matrix resulted in a tree with more than 70% ML bootstrap support for many deep nodes (Fig. 3). This suggests the data set has sufficient signal to provide information about basal avian relationships and that it will allow us to test our hypotheses. Most recovered nodes agreed with published studies (e.g., Hackett et al. 2008; Jarvis et al. 2014; Prum et al. 2015), including several studies that used different taxon samples and genomic regions (e.g., Harshman et al. 2008; Wang et al. 2012; Kimball et al. 2013; McCormack et al. 2013; Smith et al. 2013), indel data (Yuri et al. 2013), and TE insertions (Suh et al. 2011, 2015). This included six of the magnificent seven clades (the exception was clade iv; cuckoos and bustards formed a clade, but turacos were placed elsewhere). Regardless, it is clear that the EB2 tree corroborates a number of clades that were viewed as intractable less than a decade ago. The EB2 tree and the Jarvis TENT exhibited similar degrees of branch-length heterogeneity among taxa (Fig. 4) and identical branching patterns within the large superordinal clades (landbirds, clade i, and waterbirds, clade ii, in Fig. 1). More significantly, the two indicator clades of Jarvis et al. (2014) (J1 and J2; see Fig. 1a) were recovered with a high degree of bootstrap support (>95% and >70%, respectively; Fig. 3).

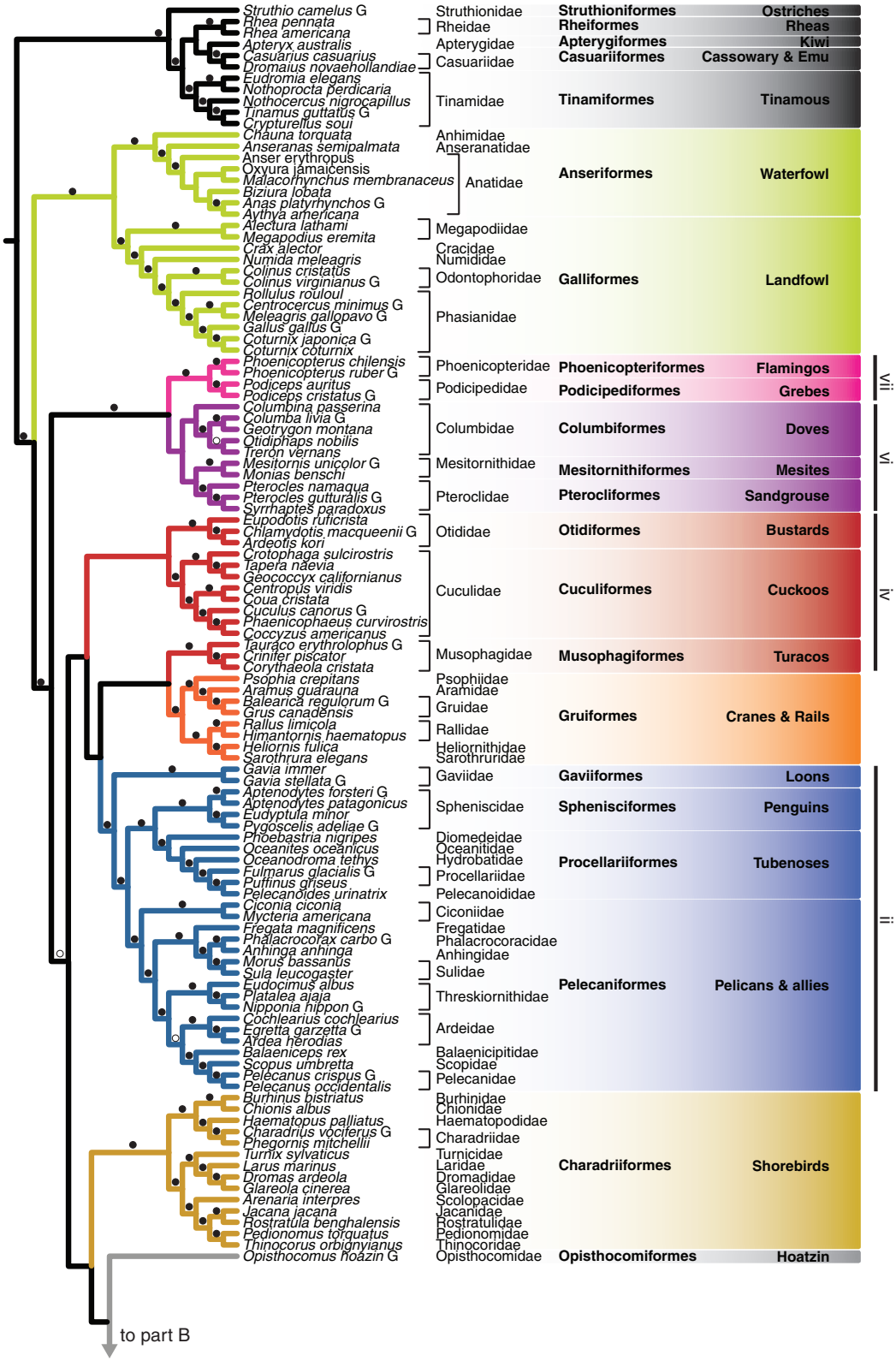


FIGURE 3. The "Early Bird II" (EB2) ML tree of 235 birds using 54 loci. This cladogram is based on a RAxML analysis with the maximum number of partitions. Filled circles indicate ⩾95% support and open circles indicate ⩾70% support. We have indicated the data derived from draft genome sequences by adding a "G" after the taxon name. We obtained similar trees using a variety of partitioning schemes in ML analyses (Supplementary File S7) and in Bayesian analyses (although support values were much higher with the latter).

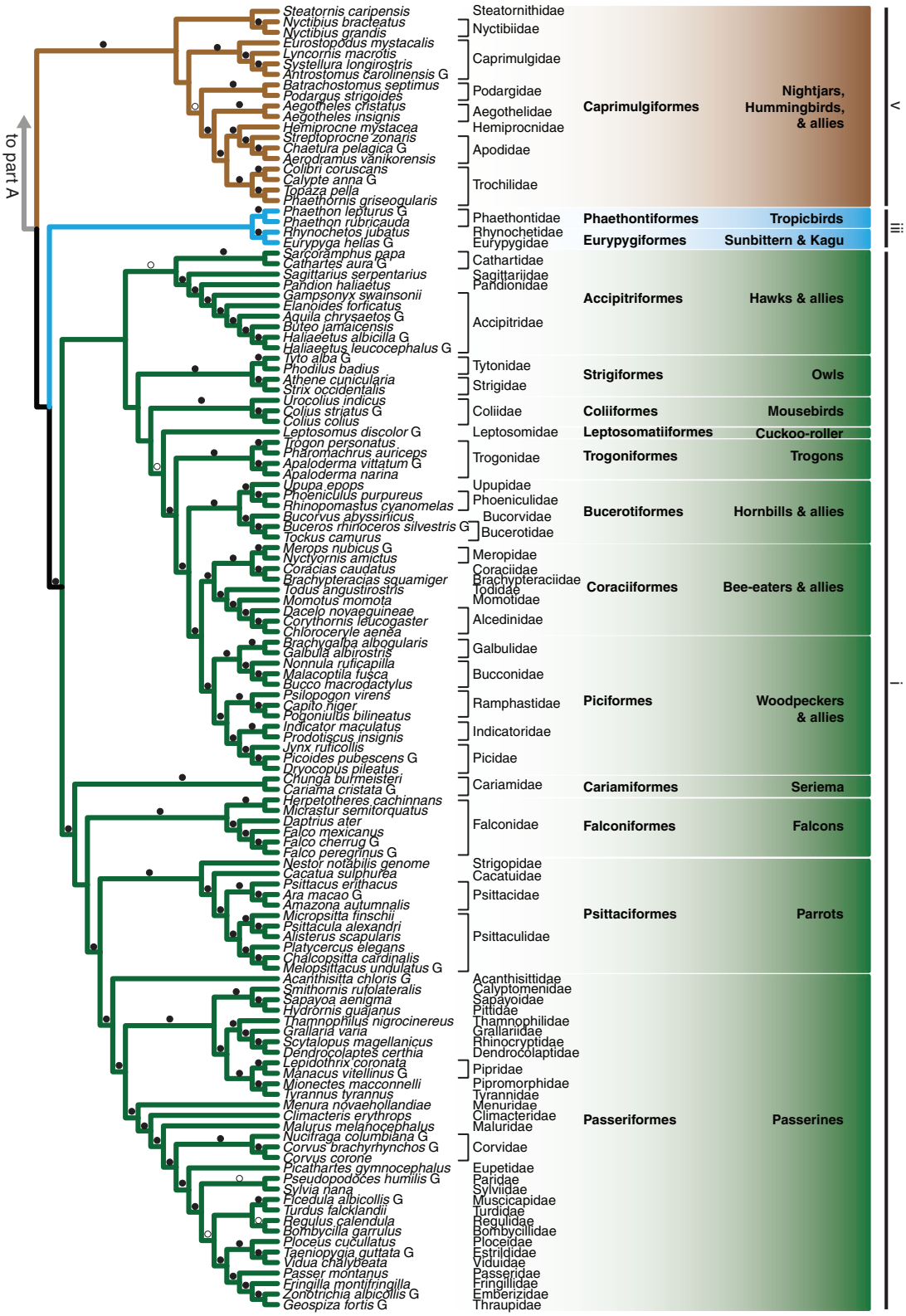


FIGURE 3. Continued

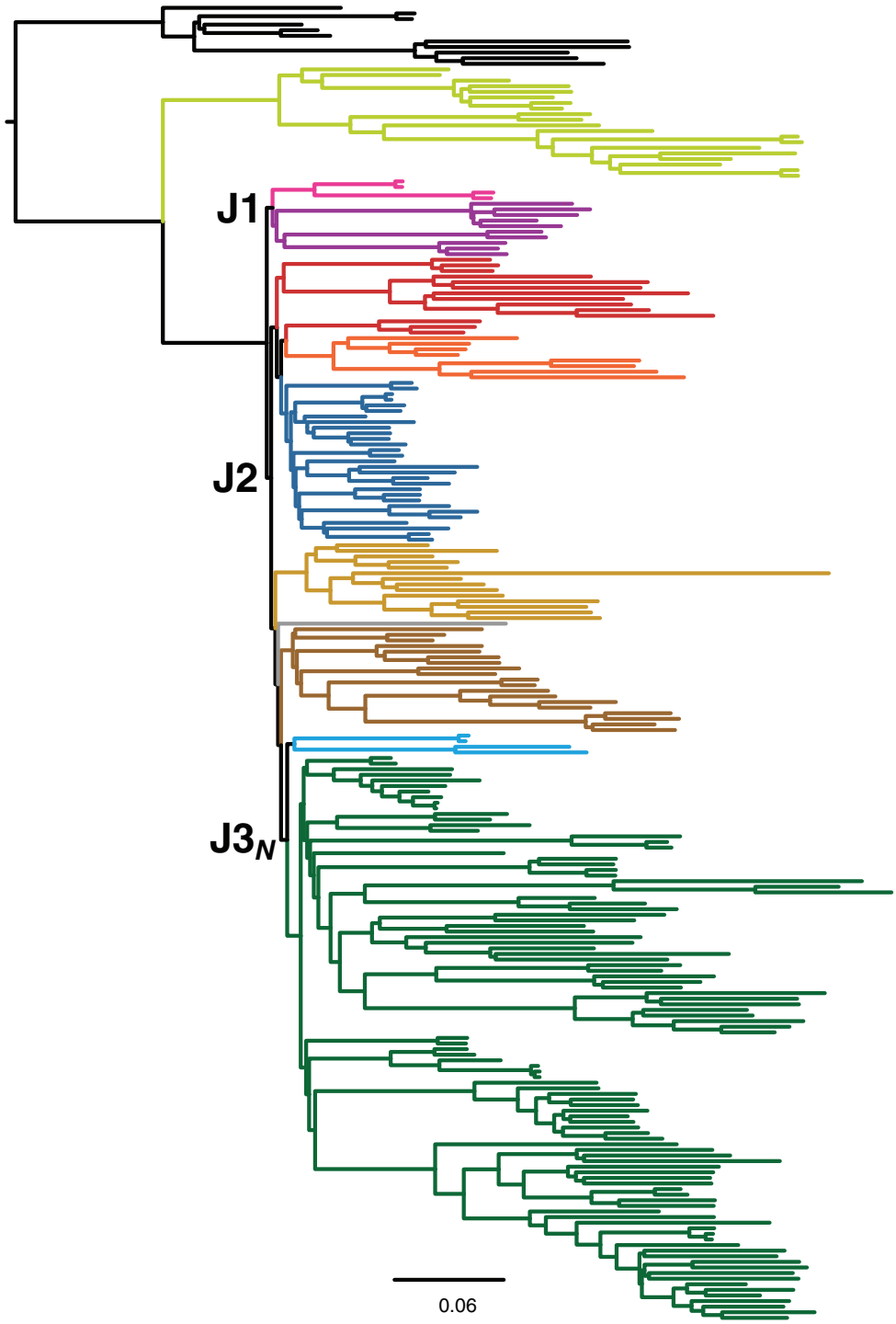


FIGURE 4. The Early Bird II tree presented as a phylogram. Colors are those used in Fig. 3, and they indicate the magnificent seven clades (one of which, iv, was not monophyletic in analyses of these data). The Jarvis indicator clades (i.e., J1, J2, and $J3_N$) are highlighted to the left of the relevant node. The scale bar indicates substitutions per site.

These "Jarvis indicator clades" directly contradict the two "Prum indicator clades" (P1 and P2, see Fig. 1b). These results corroborate the basal topology of the Jarvis TENT (i.e., the division of Neoaves into clades J1 and J2).

Support for most clades in the EB2 tree was relatively insensitive to details of the model and partitioning scheme used. All analyses of the EB2 taxon sample resulted in similar levels of support for most clades (Fig. 5 and Supplementary File S7, available on Dryad). As expected (Minh et al. 2013), the ultrafast bootstrap support values were higher than those estimated using the standard bootstrap or the RAxML fast bootstrap,

| Analysis and data matrix | Taxon Sample: Clade J1 EB2 | Clade J1 KIM | Clade J1 JAR | Clade J2 EB2 | Clade J2 KIM | Clade J2 JAR | Clade J3$_N$ EB2 | Clade J3$_N$ KIM | Clade J3$_N$ JAR |
|---|---|---|---|---|---|---|---|---|---|
| **RAxML, fast bootstrap, GTR+Γ model** | | | | | | | | | |
| Early Bird II data, fully partitioned | 99 | 99 | 99 | 73 | 55 | | 60 | 76 | 57 |
| Early Bird II data, optimal rcluster (PF) | 97 | 97 | 100 | 71 | 49 | | 70 | 55 | 53 |
| Early Bird II data, unpartitioned | 98 | 97 | 98 | 64 | 52 | | 58 | 56 | 44 |
| Non-coding data, fully partitioned | 95 | 95 | 99 | 56 | 45 | | 66 | 67 | 56 |
| **RAxML standard bootstrap, GTR+Γ model** | | | | | | | | | |
| Early Bird II data, optimal rcluster (PF) | 96 | 97 | 98 | 53 | 51 | 35 | 51 | 56 | 37 |
| **IQ-TREE ultrafast bootstrap, I+Γ rates** | | | | | | | | | |
| Early Bird II data, fully partitioned | 100 | 99 | 99 | 79 | 64 | | 80 | 88 | 57 |
| Early Bird II data, optimal rcluster (PF) | 100 | 100 | 100 | 84 | 63 | | 78 | 90 | 53 |
| Early Bird II data, unpartitioned | 100 | 100 | 100 | 85 | 64 | | 77 | 91 | 66 |
| Non-coding data, fully partitioned | 100 | 100 | 100 | 81 | 59 | | 78 | 82 | 71 |
| **IQ-TREE ultrafast bootstrap, free rates** | | | | | | | | | |
| Early Bird II data, fully partitioned | 100 | 100 | 100 | 90 | 72 | | 94 | 86 | 66 |
| Non-coding data, fully partitioned | 100 | 100 | 100 | 86 | 57 | | 72 | 75 | 68 |

FIGURE 5. Heat map showing support for the Jarvis indicator clades for different analyses and taxon samples. Colors indicate support values (dark green: ⩾95%; light green: ⩾70%; yellow: ⩾50%; uncolored: <50%; light red: not present in optimal tree). Clade J2 was not present in the optimal tree for the JAR taxon set but the branch contradicting that clade had <50% support (see trees in Supplementary File S7).

but the differences were fairly modest (Fig. 5). Bayesian support values were even higher (Supplementary File S7, available on Dryad); moreover, the topology that resulted from Bayesian analyses was virtually identical to the ML topologies (Fig. 3 and Supplementary File S7, available on Dryad) and clades J1 or J2 both had posterior probabilities of 1.0. Thus, we recovered J1 and J2 with support comparable to that reported for P1 and P2 by Prum et al. (2015) using comparable (i.e., Bayesian MCMC) methods. Taken as a whole, these analyses indicate that the basal division of the EB2 tree into clades J1 and J2 is largely insensitive to the details of the model or analytical method used to estimate phylogeny.

### *Taxon Sampling Does Not Explain Incongruence between Avian Phylogenomic Studies*

A fundamental prediction of the Prum et al. (2015) taxon-sampling hypothesis ($H_3$) is that analysis of phylogenomic data should converge on the Prum tree as the taxon sample increases. However, the EB2 tree (Fig. 3) showed greater similarity to the Jarvis TENT than to the Prum tree (compare the trees in Figs. 1–3). These observations falsify the hypothesis ($H_3$) that taxon sampling is the primary explanation for the differences between the Jarvis TENT and the Prum tree.

We emphasize that these analyses do not represent a general test of the benefits of increasing taxon sampling; the goal of these analyses was testing whether increased taxon sampling would result in an estimate of Neoaves phylogeny closer to the Prum tree (i.e., a tree containing clade P1 and/or clade P2/P2$_J$). Instead, increased taxon sampling improved support for Jarvis indicator clades (Fig. 5 and Supplementary File S7, available on Dryad), particularly clade J2 (Fig. 5). Analyses that used the JAR taxon set typically did not recover clade J2 (Fig. 5 and Supplementary File S7, available on Dryad), suggesting that taxon sampling is indeed important. In addition, reducing the number of taxa resulted in some modest topological rearrangements that are likely to be errors (Supplementary File S7, available on Dryad). Thus, increased taxon sampling had the expected behavior of improving phylogenetic estimation. However, the observation that increased taxon sampling increased support for a "Jarvis-like" tree and did not shift the tree toward a "Prum-like" topology provided additional evidence that the observed differences between the trees recovered by Jarvis et al. (2014) and Prum et al. (2015) did not reflect the more extensive taxon sampling in the latter study (contra $H_3$).

### *"Data-Type Effects" Explain Conflict between Avian Phylogenomic Studies*

Our analyses suggest that there is a "data-type effect" (i.e., $H_4$) driving topological differences among trees produced in phylogenomic studies of birds. Trees based primarily on exon data (e.g., Fig. 1b and 1d) include clade P1 and clade P2 or P2$_J$. In contrast, trees that are based primarily or exclusively on non-coding sequence data contain clades J1 and J2. The EB2 tree also has clade J3$_N$, which appears to be a potential indicator of non-coding data because it is present in the Jarvis intron tree (Fig. 1c) and UCE tree (Fig. 4b in Jarvis et al. 2014). Support for clade J3$_N$ was insensitive to taxon sampling and removal of the small amount of coding data in EB2 (Fig. 5 and Supplementary File S7, available on Dryad). Taken as a whole, these observations corroborate the data-type hypothesis ($H_4$) and suggest that data type has a greater

influence on topology than does taxon sampling for this phylogenetic problem.

### *Insufficient Signal, Hard Polytomies, and the Avian Species Tree*

The observed support for many deep branches in the EB2 tree (Figs. 3 and 4) falsifies $H_1$ (the insufficient signal hypothesis). The congruence of the EB2 tree with the Jarvis TENT and Prum tree with respect to relatively difficult clades (e.g., the magnificent seven) suggests that there is indeed a consistent signal in avian genomes that emerges in analyses of as few as 54 loci. This congruent signal contradicts the predictions of the hard polytomy hypothesis ($H_2$), which predicts that we would either observe no support for difficult branches (consistent with $H_1$ or $H_2$) or a random resolution reflecting the star tree paradox (consistent with $H_2$).

Another important question is whether estimates of the basal topology of Neoaves would change if "species tree" analyses (i.e., methods that are consistent given the multispecies coalescent) were used. The goal of this study was to explore the differences between the Jarvis TENT and the Prum tree, both of which reflect analyses of concatenated data. However, both Jarvis et al. (2014) and Prum et al. (2015) report multispecies coalescent analyses. Species tree analysis of the TENT data and intronic data support clades J1 and J2 (cf. Fig. 3b, and Supplementary Figs. S9b and S9d in Jarvis et al. 2014), but not of clade $J3_N$. Recovery of clades J1 and J2 using species tree methods indicates that these clades do not simply reflect the analysis of concatenated data. In contrast to Jarvis et al. (2014), the multispecies coalescent trees from Prum et al. (2015) are poorly resolved (e.g., Supplementary Fig. S3 in Prum et al. 2015) and none of them include clade P1 or P2 (or $P2_J$).

Our estimates of the avian species tree using ASTRAL also exhibited limited support for most basal relationships among Neoaves (Supplementary File S8, available on Dryad), similar to other studies of deep avian phylogeny that have employed species tree methods (e.g., Kimball et al. 2013; McCormack et al. 2013; Prum et al. 2015). Our multispecies coalescent trees were also sensitive to the source of the input gene trees (i.e., whether RAxML or IQ-TREE was used to estimate gene trees), a phenomenon noted in other studies (Meiklejohn et al. 2016). Nonetheless, most of our species tree analyses with the KIM and EB2 taxon samples recovered clade J1, albeit with less than 50% bootstrap support. Clade $J3_N$ was only recovered in two trees, both of which used IQ-TREE estimates of gene trees for input. However, failure to recover clade $J3_N$ in other species trees largely reflected failure to recover clade iii (Fig. 1); most analyses placed the sunbittern (along with the kagu when it was included) as sister to landbirds (though with low support). None of our species tree estimates included clade J2. Regardless, the presence of clades J1 and $J3_N$ in some of our ASTRAL trees, albeit with limited bootstrap support, suggests that a phylogenetic signal congruent with two of the Jarvis indicator clades can emerge in multispecies coalescent analyses with as few as 54 loci, as long as the data are largely non-coding.

To date, the only large-scale study that has found high levels of bootstrap support for relationships at the base of Neoaves using a species tree method was Jarvis et al. (2014), which used thousands of genes. Recently, Sayyari and Mirarab (2016) showed that the Prum et al. (2015) study is unlikely to have enough loci to resolve the bird tree using species tree methods that use gene tree reconciliation, like ASTRAL or MP-EST. Thus, given that the EB2 data set is smaller, our recovery of clade J1 in many of our species tree analyses (albeit with bootstrap support <50%) is certainly provocative and points away from the "Prum-like" tree topology. Collectively, these results strongly support the inference that recovery of clades J1 and J2 does not reflect a straightforward bias due to gene tree–species tree discordance.

### *Analyses of Independent Data Corroborate $H_4$ (Data-Type Effects)*

Since neither the EB2 nor Prum data matrix are completely independent of the Jarvis TENT data we wanted to test whether trees based on truly independent data included the indicator clades. Analyses of "EB2 noJAR", a 45-locus data matrix excluding sites that overlapped with the Jarvis TENT data, did recover clades J1 and $J3_N$ (Supplementary File S9, available on Dryad). However, the EB2 noJAR trees contradicted clade J2 with fairly high (>70%) bootstrap support; instead, the EB2 noJAR tree included a "J2-contradicting clade" that resembled Metaves (Fain and Houde 2004; see Supplementary Fig. S2, available on Dryad). Metaves is a clade found in many analyses that include FGB, a locus with strong phylogenetic signal (Kimball et al. 2013). To test the hypothesis that FGB was responsible for the J2-contradicting clade we analyzed a data matrix that excluded FGB (EB2 noJAR/noFGB). Analyses of EB2 noJAR/noFGB recovered clades J1 and J2 but not clade $J3_N$ (Supplementary File S9, available on Dryad). The sensitivity of tree building to inclusion or exclusion of FGB (after excluding Jarvis-overlapping data) suggests the EB2 data set could be near the lower size limit for data matrices that can reliably recover the indicator clades. However, the more important conclusion is that the sites that overlap between the EB2 and Jarvis TENT data do not explain the recovery of the indicator clades in the EB2 data.

We also analyzed a 104-locus "Prum noJAR" data matrix that excluded loci present in the Jarvis TENT data. The Prum noJAR trees included clade P1 but it did not include P2 or $P2_J$ (Supplementary File S10 and Supplementary Fig. S3, available on Dryad). The absence of P2/$P2_J$ in the Prum noJAR trees reflected a shift in the position of clade iii that resulted in the recovery of clade $J3_N$. Recovery of clade $J3_N$ in the Prum noJAR trees, which were largely based on coding sequences, was surprising and provocative. However, we stress that all

Prum noJAR trees exhibited limited bootstrap support, which is not surprising given the low bootstrap support in the original Prum tree (Fig. 1b). However, the most important result is that clade P1 emerges in analyses of two truly independent exonic data matrices (i.e., the Jarvis exon trees and our Prum noJAR trees).

We also assessed whether trees based on the same data type converged on similar topologies, regardless of the presence or absence of the indicator clades. NJ of symmetric RF distances among avian phylogenies (Fig. 6) revealed a deep division between coding exon trees and trees that are largely (or completely) based on non-coding data. This result strongly corroborated $H_4$. Interestingly, the two trees based on rare genomic changes (RGCs) were placed in the same cluster as the non-coding data; the tree based on indels actually fell within the non-coding trees (Fig. 6). Sequence data and RGCs have different strengths and weaknesses as sources of phylogenic information. The fact that trees based on non-coding sequence data and RGCs cluster in treespace suggests that those trees could be closer to the true evolutionary history of birds.

### *What Drives Data-Type Effects?*

We suggest using the term "data-type effects" as a way to discuss different signals associated with analyses of subsets of the genome that can be defined *a priori* using non-phylogenetic criteria. Here, we noted a strong contrast in the phylogenetic signals associated with coding and non-coding (largely intronic) data, but similar phenomena might be found for more finely subdivided subsets of the genome (e.g., globular proteins vs. transmembrane proteins, long introns vs. short introns, and so forth). Indeed, two studies have reported that the phylogenetic signal associated with ribosomal proteins conflict with the signal associated with other proteins (Nosenko et al. 2013; Whelan et al. 2015), and another noted a difference between introns and UTRs (Bonilla et al. 2010). There are two potential explanations for data-type effects: 1) different data types may have different underlying gene trees; or 2) model violations prevent accurate reconstruction from one (or more) data types.

The hypothesis that different data types are associated with distinct sets of gene trees is unlikely because many data types will reflect collections of unlinked gene regions that may even be intermixed in the genome (e.g., exons are interspersed with introns). Thus, estimates of trees based on concatenated data sets of each data type should reflect the same sets of underlying gene trees, and analyses of data that were generated on those gene trees are expected to converge on the same topology (as long as the model used does not exhibit a bias for any of the data types being analyzed). Even if there is substantial recombination within loci, as some have suggested to be common (Gatesy and Springer 2014; Springer and Gatesy 2016; Scornavacca and Galtier 2017), there is no expectation that the spectra of gene trees for intermixed data types (such as coding and non-coding sequences) would differ.

There are two possible exceptions to the expectation that the spectrum of gene trees would be similar across a genome: 1) cases in which a data type is limited to a specific genetic unit that has a distinct pattern of inheritance (e.g., a sex chromosome or a uniparentally inherited organelle genome); and 2) cases where data types are globally subject to different patterns of selection. The basis for the first is obvious, and it does not apply to this study (the two different data types we considered here were sampled both from autosomes and the Z chromosome). The second issue can occur when consistent purifying selection over long evolutionary timescales leads to local reductions in effective population size ($N_e$) (Comeron et al. 2008). The reduction of $N_e$ due to selection is most famously articulated in the case of the Hill and Robertson (1966) effect, but it can occur under a number of conditions. Reduction in $N_e$ decreases the time to coalescence for relevant subsets of the genome (i.e., sites that are linked to sites subject to selection) and this results in increased coalescent branch lengths and different mixtures of gene trees. It is unlikely that this effect could represent a major explanation for the data-type effects we observed here. Hill–Robertson effects have been invoked to explain the correlation between efficiency of selection on coding regions and localized differences in $N_e$ (Axelsson et al. 2005; Künstner et al. 2010), although the spectrum of gene trees for the regions that were suggested to have low $N_e$ was not assessed in those studies. Regardless, analyses of UCEs, which are non-coding regions that appear to be ultra-selected (Katzman et al. 2007), yield a tree more similar to intronic trees than to coding exon trees (Fig. 6), despite the fact that both UCEs and exons are expected to be subject to stronger purifying selection than introns. Taken as a whole, these observations suggest that localized variation in $N_e$ does not explain the data-type effects that we observe here.

We believe that most data-type effects reflect poor model fit. The slower accumulation and known patterns of substitutions in coding regions (Supplementary Table S1, available on Dryad) might give the impression that their evolution is easier to model than that of non-coding regions. However, the complexity of the factors that determine rates of amino acid substitution (for recent reviews, see Chi and Liberles 2016; Echave et al. 2016) suggests this is unlikely to be the case. For example, constraints that result in slower accumulation of substitutions in coding regions could drive site-specific biases in nucleotide frequencies due to correlation between physicochemical properties of amino acids and the structure of the genetic code (e.g., Supplementary Fig. S4, available on Dryad). For example, Naylor et al. (1995) describe a site-specific bias that reflects the correlation between hydrophobicity and second codon position pyrimidine content. Interactions among amino acid residues after the polypeptide folds into

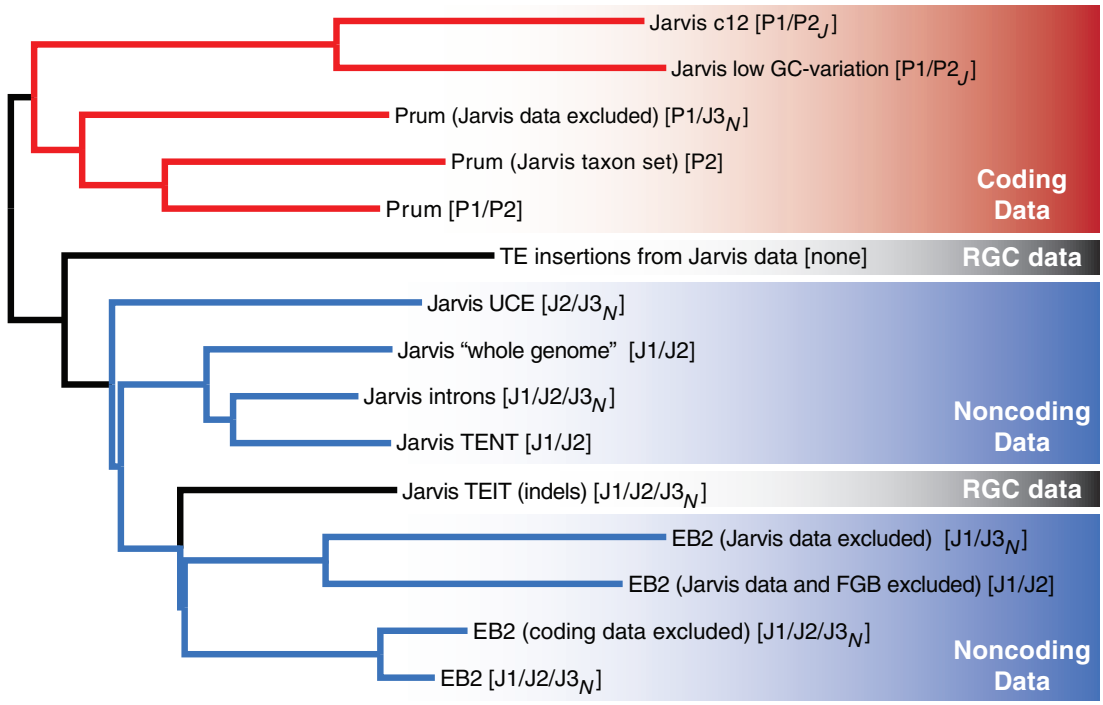


FIGURE 6. Clustering of the Early Bird II trees and other estimates of avian phylogeny using tree distances. We generated this dendrogram by NJ of symmetric RF distances followed by midpoint rooting. The indicator clades present in each tree are provided in square brackets. Coding exon trees used for these comparisons were: 1) Jarvis c12 (Fig. 1d); 2) Jarvis low GC variation (Fig. 6a in Jarvis et al. 2014); 3) Prum (Fig. 1b); 4) Prum, Jarvis taxon set (Supplementary Fig. S8 in Prum et al. 2015); and 5) Prum, Jarvis data excluded (Supplementary File S10). Non-coding trees were: 1) Jarvis TENT (Fig. 1a); 2) Jarvis whole-genome tree (Supplementary Fig. S4d in Jarvis et al. 2014); 3) Jarvis intron (Fig. 1c); 4) Jarvis UCE (Fig. 4b in Jarvis et al. 2014); 5) EB2 (Fig. 3); 6) EB2, coding data excluded (Supplementary File S7); 7) EB2, Jarvis data excluded (Supplementary File S9); and 8) EB2, Jarvis data and FGB excluded (Supplementary File S9). RGC trees were: 1) Jarvis TEIT (total evidence indel tree; Supplementary Fig. S12 in Jarvis et al. 2014); and 2) TE insertions from Jarvis data (Fig. 1B in Suh et al. 2015). All of these trees are available in Supplementary File S11.

a tertiary structure have the potential to create even more complex patterns of constraint (Penny 2017). Finally, dependencies among codon positions can causes standard phylogenetic models to underestimate branch lengths, potentially resulting in biased estimates of topology (Whelan 2008). These issues represent some of the reasons why the commonly used GTR+I+Γ model (and its submodels) is likely to exhibit a poor fit to the underlying patterns of evolution of aligned protein-coding regions.

The GTR model is also expected to exhibit poor fit to regions where base composition has not remained stationary. Although models that relax the assumption that base composition remains stationary over the tree have been developed (e.g., Lake 1994; Lockhart et al. 1994; Galtier and Gouy 1998; Foster 2004; Holland et al. 2013), the behavior of methods based on these models remains relatively poorly explored, and in some cases their use results in the recovery of unexpected (and probably incorrect) clades (e.g., Katsu et al. 2009; Holland et al. 2013). Therefore, some authors advocate using loci with limited base compositional variation for tree building (e.g., Collins et al. 2005; Romiguier et al. 2013). Jarvis et al. (2014) reported striking variation in GC-content in some avian genomic regions (specifically coding regions; cf. Fig. 6 in Jarvis et al. 2014). This finding suggests that deviations from compositional stationarity could lead to problems for analyses that depend on avian coding regions.

Motivated by these observations, we examined the GC-content variation in the EB2 and Prum data (Fig. 7 and Supplementary Fig. S5, available on Dryad). The Prum data and the coding data of Jarvis et al. (2014) exhibited more GC-content variation than the EB2 data (Fig. 7). Moreover, the three EB2 loci with the greatest compositional variation (BDNF, NGF, and NTF3) were those for which we collected only exonic data. We conducted analyses of the EB2 matrix after excluding those three loci (and other coding exon sequences) and found that our conclusions were unaltered (Fig. 5 and Supplementary File S7, available on Dryad). GC-content variation in the Prum data was less than that exhibited by all exons in the Jarvis et al. (2014) data, but it was similar to that in the Jarvis et al. (2014) c12 data, which comprises first and second codon positions. However, GC-content variation alone is unlikely to explain the recovery of P1 and P2/P2$_J$ in trees based on coding data, since the "Jarvis low GC-variation" tree, which was estimated using the least GC-variable coding regions in Jarvis et al. (2014), also contains those clades (Fig. 6). Regardless, it is clear that the non-coding subset of the EB2 data (like the Jarvis et al. 2014 introns) exhibited less

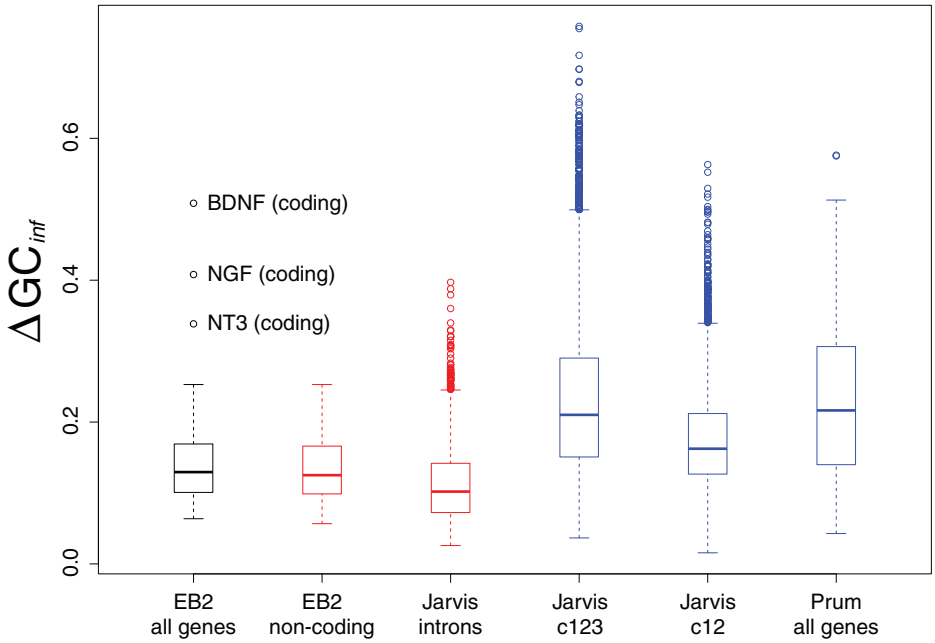


FIGURE 7. GC-content variation among taxa in phylogenomic data sets of birds showing the distribution of differences between the minimum and maximum GC-content at each locus ($\Delta GC_{inf}$). Values that fell outside the interquartile range by >1.5-fold were designated outliers. We limited GC-content calculations to parsimony-informative sites from regions with $\geqslant$50 parsimony-informative sites. The identity of three outliers in the EB2 data is noted next to the relevant data points. We used the JAR taxon sets for EB2 and Prum et al. (2015) so analyses of those data set more comparable to the data described by Jarvis et al. (2015). Analysis using the complete taxon sets (Supplementary Fig. S3) were very similar.

GC-content variation across species than any of the exon data sets examined here. Thus, it is reasonable to postulate that reduced base compositional variation is one reason GTR+I+Γ and its submodels often fit non-coding data better than they do coding data.

Since all three data sets include some loci with high GC variation, we tested whether a method (squangles; Holland et al. 2013) that is consistent under the general (i.e., non-stationary) Markov model would improve the tree estimated from the EB2 locus with the greatest GC variation (BDNF). This effort resulted in limited improvement to the BDNF gene tree (Supplementary File S12, available on Dryad). Use of squangles with the JAR taxon set of the EB2 data (the current implementation of squangles was too slow for the larger taxon sets) resulted in distinct topologies when different quartet assembly methods (NJ and MRP, see Materials and Methods) were used. However, clade J1 was present in trees based on both quartet assembly methods (Supplementary File S12, available on Dryad). Squangles + NJ did succeed in recovering all of the magnificent seven (Supplementary File S12, available on Dryad), unlike our ML analyses (e.g., Fig. 3) and squangles + MRP analysis (Supplementary File S12, available on Dryad). However, both squangles analyses of the EB2 data exhibited some unexpected (and probably incorrect) rearrangements (e.g., within landbirds; Supplementary File S12, available on Dryad). Continued development of methods based on the GMM may be fruitful, but the method we used clearly has some limitations.

We also explored the impact of assumptions regarding among-site rate variation (+I+Γ vs. free rates). Free rate models fit the data better, suggesting that among-site variation may be more complex than expected given the +I+Γ model (Supplementary File S7, available on Dryad). Their use also resulted in substantially increased branch length estimates for the KIM and EB2 taxon sets. However, they had essentially no impact on topology or support values (Fig. 5 and Supplementary File S7, available on Dryad). Several well-characterized sources of site constraint in non-coding regions have the potential to drive among-site rate variation. This is especially true for UTRs, which are known to regulate translation and mRNA stability (Mazumder et al. 2003). Visual examination of alignments also revealed an ultraconserved region in one of the EB2 introns (a BLAST search revealed that the region corresponded to a small nucleolar RNA). However, most of the EB2 data were intronic, and the majority of intronic sites do not exhibit significant deviations from neutrality (e.g., see Keightley and Gaffney 2003). Of course, Hill–Robertson effects can reduce $N_e$ for introns due to selection on adjacent exons; this could explain the observation that introns have lower nucleotide diversity ($\pi$) than anonymous regions (Lee and Edwards 2008). Despite such factors, it is likely that patterns of sequence evolution for introns are less complex than for coding regions. This difference makes it reasonable to postulate that standard analytical methods (i.e., those using GTR+I+Γ and its submodels) are likely to fit non-coding regions, especially introns, better than coding regions.

*What Has Phylogenomics Taught us about the Avian Tree of Life?*

Falsification of $H_3$ (taxon sampling) and corroboration of $H_4$ (data-type effects) makes establishing the most accurate estimate of avian phylogeny challenging. If analytical models have a better fit to non-coding data, as may be likely, then the trees based largely or completely on analyses of non-coding data [i.e., the Jarvis TENT, the Jarvis et al. (2014) non-coding trees, and the EB2 tree] are likely to be the best available estimates of avian phylogeny. On the other hand, the Prum tree, the Jarvis et al. (2014) exon trees, and the Prum noJAR tree are likely to be closer to the true bird tree if the models used for analyses are more appropriate for coding sequences. The clustering of non-coding and RGC trees (Fig. 6) may indicate that the non-coding trees are closer to the true avian tree, corroborating our arguments that it is likely to be more difficult to model evolution for coding data than non-coding data. Nevertheless, the community should remain open minded to the possibility that the coding sequences are actually the better source of phylogenetic information. Ultimately, establishing the best estimate of avian phylogeny requires establishing the data type most likely to exhibit a good fit to the models of sequence evolution used for phylogenetic estimation.

Many superordinal relationships are recovered in the Jarvis TENT, Prum tree, and the EB2 tree. The majority rule consensus of these three trees is actually quite similar to Hackett et al. (2008), although it is better resolved (Fig. 8). This consensus tree includes the magnificent seven along with clades J1 and J2. There is also substantial agreement among the aforementioned studies and other studies (e.g., McCormack et al. 2013) regarding relationships within the magnificent seven; most or all of those relationships are very likely to reflect the true avian phylogeny.

Resolving the remaining questions that surround basal avian phylogeny will likely require a better understanding of the underlying patterns of sequence evolution for birds. Indeed, the situation actually becomes more confusing if we look to TE insertions. In contrast to indels as a whole, analyses of TE insertions (Suh et al. 2015) revealed a tree that does not include two of the magnificent seven (clades iv and vi), has several rearrangements within landbirds, and (most importantly) lacks all of the indicator clades (i.e., J1, J2, $J3_N$, P1, and $P2/P2_J$). Although the TE insertion tree clustered with the non-coding trees (Fig. 6), it is clearly fairly different from the non-coding trees and the tree for all indels. This seems surprising since TE insertions are often assumed to be essentially homoplasy-free. The differences between the TE tree and other estimates of avian phylogeny suggest that avian TE insertions can exhibit true homoplasy, corroborating a previous study (Han et al. 2011). Overall, the conclusions for TE insertions are similar to our conclusions for analyses of introns and coding exons: better models of genomic change will be necessary to better understand avian phylogeny.

There is certainly room for the addition of taxa to further improve avian phylogenomic studies. Indeed, our results indicate that analyses of the EB2 235-taxon matrix yielded better estimates than the 120-taxon KIM matrix despite the fact that the KIM taxa were selected, as much as possible, to bisect long branches (see Kimball et al. 2013). However, some of the instability at the base of Neoaves involves lineages with no close living relatives (e.g., hoatzin) and it is impossible to break up those long branches in a meaningful way. Expanding the number of species with underlying sequence data available for large-scale "synthetic" phylogenies (e.g., Jetz et al. 2012; Burleigh et al. 2015) will no doubt be invaluable for analyses of biogeography and traits (e.g., Wang et al. 2017). However, simply adding taxa is unlikely to be the key to resolving the most difficult nodes in the bird tree. Instead, the key is likely to be improving models of evolution for data types used in analyses. Finally, we also need to remain open minded to the possibility that we have reached the limits of resolution of the avian tree of life; perhaps birds (or, more specifically, some Neoaves) are indeed perched in a phylogenetic bush.

*Broader Implications for Phylogenomic Studies*

> More taxa and characters may not guarantee a "satisfying" answer, by which we mean having resolution of nodes with sufficiently strong branch support that additional data will merely confirm what has already been found.—Cracraft et al. (2004).

The question of whether large-scale data sets, potentially even complete genomes, will yield a "satisfying" estimate of phylogeny is fundamental to all phylogenomic studies. Our analyses suggest that simply collecting more data and conducting analyses using the same models is unlikely to resolve avian phylogeny with confidence. Our results further suggest conclusions from other phylogenomic studies that have examined recalcitrant nodes in the tree of life should be reexamined for data-type effects and interpreted with caution.

When large-scale data matrices are used to inspect the difficult nodes, relatively subtle model violations may be sufficient to mislead analyses (Hahn and Nakhleh 2016). Those violations may not be obvious unless multiple data types are analyzed, as in Jarvis et al. (2014). The hypothesis testing approach we used here may be a desirable way to explore the sources of conflicting signals in phylogenomic analyses. We believe that data-type effects represent a fundamental challenge for phylogenomics and urge investigators to attack difficult problems in the tree of life using as many data types as possible. If it is established that there are data-type effects, it becomes critical to use the best-fitting model for analyses. Of course, it may be necessary to develop substantially more complex and "biologically realistic" models to achieve this goal.

Achieving "biological realism" may require very complex models, which in turn present fundamental challenges (cf. Steel 2005). First, complex models are

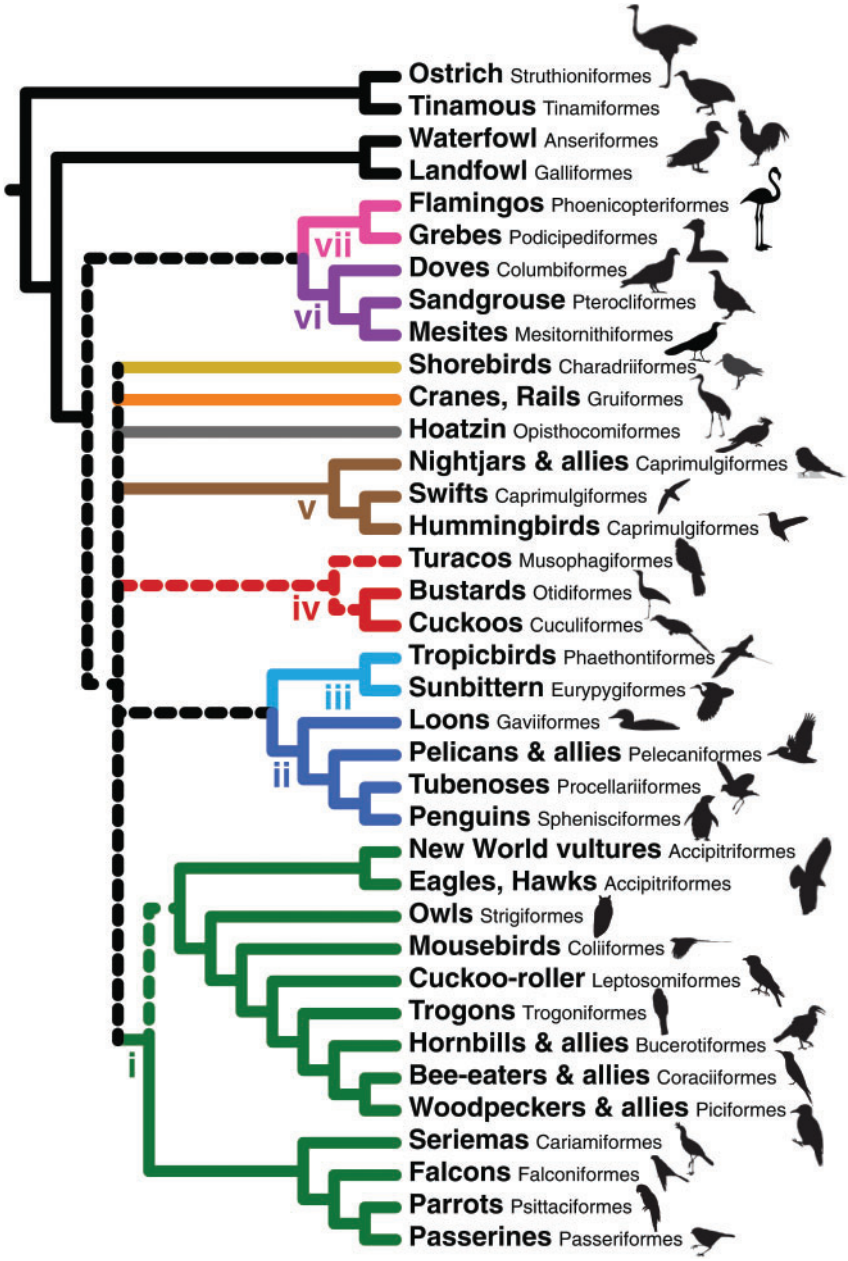


FIGURE 8. Consensus phylogeny of birds. This topology reflects a majority rule consensus tree of the Jarvis TENT, the Prum tree, and the Early Bird II tree; branches present in only two out of the three trees are presented as dashed lines. The ordinal classification follows Cracraft (2013). Only two paleognath orders are shown because Jarvis et al. (2014) only sampled those orders; all other orders (as circumscribed by Cracraft 2013) were sampled by all three studies. The magnificent seven are identified by lower case Roman numerals next to the relevant nodes and the colors are identical to those in Fig. 3. Silhouettes were drawn by Sushma Reddy, Edward L. Braun, or obtained from http://phylopic.org.

likely to impose a substantial computational burden. Even with standard models, Jarvis et al. (2014) required more than 400 years of CPU time to conduct their analyses. Second, adding a large number of free parameters may create problems with parameter identifiability (cf. Ponciano et al. 2012). A potential solution to this problem might be finding semi-parametric approaches that add biological complexity in other ways (e.g., penalized likelihood; Kim and Sanderson 2008).

One final problem, which applies to both complex and simpler models, is the accuracy of numerical optimization routines. When analyzing very large data matrices, the limitations of floating point arithmetic can have a profound impact on analyses (cf. Sainudiin and Yoshida 2005; Darriba et al. 2015; Meiklejohn et al. 2016). This potential certainly exits for programs that use different libraries for likelihood calculations [e.g., the phylogenetic likelihood library (Flouri et al. 2015) for IQ-TREE and the BEAGLE library (Ayres et al. 2012)

for MrBayes]. However, differences in numerical optimization can even exist for the same program compiled in different environments (e.g., Darriba et al. 2015). If estimates of phylogeny were generated using different programs then any observed conflicts could reflect the failure to identify the true optimal trees due to the impact of numerical optimization routines. The most straightforward way to examine the potential impact of this pitfall is running the same (or very similar) analyses in multiple programs that employ distinct numerical optimization routines. Herein, we tested the impact of this issue on the EB2 phylogeny by using multiple programs and found that they had no major effect on our conclusions.

A final approach to solving the problem of data-type effects might be the identification of data types that yield unbiased estimates of phylogeny when analyzed using available models and software. Some methods to accomplish this have been proposed (e.g., Chen et al. 2015; Doyle et al. 2015), although the proportion of the genome that we can realistically expect to exhibit a sufficiently good fit to commonly used models remains unclear. No doubt, phylogenomic data to address many different questions will continue to be generated (e.g., Jarvis et al. 2014; Misof et al. 2014; Wickett et al. 2014). The real test of whether phylogenomics can fulfill the promise to resolve the tree of life will depend on careful scrutiny of the data for patterns of sequence evolution that might lead to bias and understanding the impact of those patterns on the results of phylogenetic analyses.

## Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.6536v.

## Funding

This work was supported by the U.S. National Science Foundation [grants DEB-0228682 and DEB-1118823 to R.T.K. and E.L.B.; DEB-0228675 to S.J.H.; DEB-0228688 to F.H.S.; DEB-0733029 to M.J.B.; and DEB-0228617 to W.S.M.].

## Acknowledgements

This manuscript was improved by comments from Erich Jarvis, Jay McEntee, and the members of the Braun–Kimball laboratory group. They are grateful to John McCormack, Brant Faircloth, and two anonymous reviewers for many helpful suggestions that improved the final version of this manuscript. They thank the museums and collectors who donated the tissues used for additional data collection (we list full information about the museums and collectors for each specimen in Supplementary File S3).

## References

Aberer A.J., Kobert K., Stamatakis A. 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31:2553–2556.

Alfaro M.E., Zoller S., Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20:255–266.

Axelsson E., Webster M.T., Smith N.G., Burt D.W., Ellegren H. 2005. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Res. 15:120–125.

Ayres D.L., Darling A., Zwickl D.J., Beerli P., Holder M.T., Lewis P.O., Huelsenbeck J.P., Ronquist F., Swofford D.L., Cummings M.P., Rambaut A., Suchard M.A. 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst. Biol. 61:170–173.

Baum B.R. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41:3–10.

Bonilla A.J., Braun E.L., Kimball R.T. 2010. Comparative molecular evolution and phylogenetic utility of 3′-UTRs and introns in Galliformes. Mol. Phylogenet. Evol. 56:536–542.

Braun E.L., Kimball R.T. 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.

Braun E.L., Kimball R.T. 2002. Examining basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. Syst. Biol. 51:614–625.

Braun E.L., Kimball R.T., Han K.L., Iuhasz-Velez N.R., Bonilla A.J., Chojnowski J.L., Smith J.V., Bowie R.C.K., Braun M.J., Hackett S.J., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Reddy S., Sheldon F.H., Witt C.C., Yuri T. 2011. Homoplastic microinversions and the avian tree of life. BMC Evol. Biol. 11:141.

Buckley T.R. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51:509–523.

Burleigh J.G., Kimball R.T., Braun E.L. 2015. Building the avian tree of life using a large-scale, sparse supermatrix. Mol. Phylogenet. Evol. 84:53–63.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. 2008. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Card D.C., Schield D.R., Reyes-Velasco J., Fujita M.K., Andrew A.L., Oyler-McCance S.J., Fike J.A., Tomback D.F., Ruggiero R.P., Castoe T.A. 2014. Two low coverage bird genomes and a comparison of reference-guided versus de novo genome assemblies. PLoS One 9:e106649.

Chen M.Y., Liang D., Zhang P. 2015. Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. Syst. Biol. 64:1104–1120.

Chi P.B., Liberles D.A. 2016. Selection on protein structure, interaction, and sequence. Protein Sci. 25:1168–1178.

Chojnowski J.L., Kimball R.T., Braun E.L. 2008. Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. Gene 410:89–96.

Clements J.F., Schulenberg T.S., Iliff M.J., Roberson D., Fredericks T.A., Sullivan B.L., Wood C.L. 2015. The eBird/Clements checklist of birds of the world: v2015. Available from: URL http://www.birds.cornell.edu/clementschecklist/download.

Collins T.M., Fedrigo O., Naylor G.J. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. Syst. Biol. 54:493–500.

Comeron J.M., Williford A., Kliman R.M. 2008. The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity (Edinb) 100:19–31.

Cracraft J. 2013. Avian higher-level relationships and classification: nonpasseriforms. In: Dickinson E.C., Remsen J.V., editors. The Howard and Moore complete checklist of the birds of the world. Eastbourne, UK: Aves Press. p. xxi–xliii.

Cracraft J., Barker F.K., Braun M., Harshman J., Dyke G.J., Feinstein J., Stanley S., Cibois A., Schikle P., Beresford P., García-Moreno J.,

Sorensen M.D., Yuri T., Mindell D.P., Cracraft J., Barker F.K., Braun M., Harshman J., Dyke G.J., Feinstein J., Stanley S., Cibois A., Schikle P., Beresford P., García-Moreno J. 2004. Phylogenetic relationships among modern birds (Neornithes): towards an avian tree of life. In: Cracraft J., Donoghue M.J., editors. Assembling the tree of life. New York: Oxford University Press. p. 468–489.
Creevey C.J., McInerney J.O. 2005. Clann: investigating phylogenetic information through supertree analyses. Bioinformatics 21:390–392.
Darriba D., Flouri T., Stamatakis A. 2015. The state of software in evolutionary biology. BioRxiv: http://dx.doi.org/10.1101/031930.
Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:762–768.
Delsuc F., Brinkmann H., Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6:361–375.
Dickinson E.C., Christidis L. 2014. The Howard and Moore complete checklist of the birds of the world. 4th ed., Vol. 2. Passerines. Eastbourne, UK: Aves Press.
Dickinson E.C., Remsen J.V. 2013. The Howard and Moore complete checklist of the birds of the world. 4th ed. Vol. 1. Non-passerines. Eastbourne, UK: Aves Press.
Doyle V.P., Young R.E., Naylor G.J., Brown J.M. 2015. Can we identify genes with increased phylogenetic reliability? Syst. Biol. 64:824–837.
Echave J., Spielman S.J., Wilke C.O. 2016. Causes of evolutionary rate variation among protein sites. Nat. Rev. Genet. 17:109–121.
Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.
Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.
Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leache A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.
Ericson P.G.P., Anderson C.L., Britton T., Elzanowski A., Johansson U.S., Kallersjo M., Ohlson J.I., Parsons T.J., Zuccon D., Mayr G. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biol. Lett. 2:543–547.
Fain M.G., Houde P. 2004. Parallel radiations in the primary clades of birds. Evolution 58:2558–2573.
Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717–726.
Felsenstein J. 1978a. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.
Felsenstein J. 1978b. The number of evolutionary trees. Syst. Zool. 27: 27–33.
Flouri T., Izquierdo-Carrasco F., Darriba D., Aberer A.J., Nguyen L.-T., Minh B.Q., Von Haeseler A., Stamatakis A. 2015. The phylogenetic likelihood library. Syst. Biol. 64:356–362.
Foster P.G. 2004. Modeling compositional heterogeneity. Syst. Biol. 53:485–495.
Galtier N., Gouy M. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15:871–879.
Gatesy J., Springer M.S. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80:231–266.
Gee H. 2003. Evolution: ending incongruence. Nature 425:782.
Gill F., Donsker D. 2016. IOC World Bird List (v 6.2). Available from: URL http://www.worldbirdnames.org.
Groth J.G., Barrowclough G.F. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12:115–123.
Hackett S.J., Kimball R.T., Reddy S., Bowie R.C.K., Braun E.L., Braun M.J., Chojnowski J.L., Cox W.A., Han K.L., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Steadman D.W., Witt C.C., Yuri T. 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320:1763–1768.
Hahn M.W., Nakhleh L. 2016. Irrational exuberance for resolved species trees. Evolution 70:7–17.
Han K.L., Braun E.L., Kimball R.T., Reddy S., Bowie R.C.K., Braun M.J., Chojnowski J.L., Hackett S.J., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Steadman D.W., Witt C.C., Yuri T. 2011. Are transposable element insertions homoplasy free? An examination using the avian tree of life. Syst. Biol. 60: 375–386.
Harshman J., Braun E.L., Braun M.J., Huddleston C.J., Bowie R.C.K., Chojnowski J.L., Hackett S.J., Han K.L., Kimball R.T., Marks B.D., Miglia K.J., Moore W.S., Reddy S., Sheldon F.H., Steadman D.W., Steppan S.J., Witt C.C., Yuri T. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. Proc. Natl Acad. Sci. USA. 105:13462–13467.
Heath T.A., Hedtke S.M., Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 48:239–257.
Hendy M.D., Penny D. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38:297–309.
Hill W.G., Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.
Hillis D.M. 1996. Inferring complex phylogenies. Nature 383:130–131.
Hillis D.M., Pollock D.D., McGuire J.A., Zwickl D.J. 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52:124–126.
Holland B.R., Jarvis P.D., Sumner J.G. 2013. Low-parameter phylogenetic inference under the general Markov model. Syst. Biol. 62:78–92.
Hosner P.A., Faircloth B.C., Glenn T.C., Braun E.L., Kimball R.T. 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. 33:1110–1125.
Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Alfaro-Nunez A., Narula N., Liu L., Burt D., Ellegren H., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T., Zhang G.; The Avian Phylogenomics Consortium. 2015. Phylogenomic analyses data of the avian phylogenomics project. Gigascience 4:4.
Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J.W., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldon T., Capella-Gutierrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X.J., Dixon A., Li S.B., Li N., Huang Y.H., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M.V., Alfaro-Nunez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z.J., Zeng Y.L., Liu S.P., Li Z.Y., Liu B.H., Wu K., Xiao J., Yinqi X., Zheng Q.M., Zhang Y., Yang H.M., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldsa J., Orlando L., Barker F.K., Jonsson K.A., Johnson W., Koepfli K.P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alstrom P., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G.J. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320–1331.
Jeffroy O., Brinkmann H., Delsuc F., Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.
Jetz W., Thomas G.H., Joy J.B., Hartmann K., Mooers A.O. 2012. The global diversity of birds in space and time. Nature 491:444–448.
Katsu Y., Braun E.L., Guillette L.J., Iguchi T. 2009. From reptilian phylogenomics to reptilian genomes: analyses of c-*Jun* and *DJ-1* proto-oncogenes. Cytogenet. Genome Res. 127:79–93.
Katzman S., Kern A.D., Bejerano G., Fewell G., Fulton L., Wilson R.K., Salama S.R., Haussler D. 2007. Human genome ultraconserved elements are ultraselected. Science 317:915.
Keightley P.D., Gaffney D.J. 2003. Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. Proc. Natl Acad. Sci. USA. 100:13402–13406.

Kim J., Sanderson M.J. 2008. Penalized likelihood phylogenetic inference: bridging the parsimony-likelihood gap. Syst. Biol. 257:665–674.

Kimball R.T., Braun E.L., Barker F.K., Bowie R.C.K., Braun M.J., Chojnowski J.L., Hackett S.J., Han K.L., Harshman J., Heimer-Torres V., Holznagel W., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Reddy S., Sheldon F.H., Smith J.V., Witt C.C., Yuri T. 2009. A well-tested set of primers to amplify regions spread across the avian genome. Mol. Phylogenet. Evol. 50:654–660.

Kimball R.T., Wang N., Heimer-McGinn V., Ferguson C., Braun E.L. 2013. Identifying localized biases in large datasets: a case study using the avian tree of life. Mol. Phylogenet. Evol. 69:1021–1032.

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Künstner A., Wolf J.B., Backström N., Whitney O., Balakrishnan C.N., Day L., Edwards S.V., Janes D.E., Schlinger B.A., Wilson R.K., Jarvis E.D., Warren W.C., Ellegren H. 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. Mol. Ecol. 19(Suppl. 1):266–276.

Lake J.A. 1994. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. Proc. Natl Acad. Sci. USA. 91:1455–1459.

Lanfear R., Calcott B., Ho S.Y., Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Lee J.Y., Edwards S.V. 2008. Divergence across Australia's Carpentarian barrier: statistical phylogeography of the red-backed fairy wren (*Malurus melanocephalus*). Evolution 62:3117–3134.

Lewis P.O., Holder M.T., Holsinger K.E. 2005. Polytomies and Bayesian phylogenetic inference. Syst. Biol. 54:241–253.

Lockhart P.J., Steel M.A., Hendy M.D., Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11:605–612.

Matsen F.A., Steel M. 2007. Phylogenetic mixtures on a single tree can mimic a tree of another topology. Syst. Biol. 56:767–775.

Mazumder B., Seshadri V., Fox P.L. 2003. Translational control by the 3′-UTR: the ends specify the means. Trends Biochem. Sci. 28:91–98.

McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8:e54848.

Meiklejohn K.A., Danielson M.J., Faircloth B.C., Glenn T.C., Braun E.L., Kimball R.T. 2014. Incongruence among different mitochondrial regions: a case study using complete mitogenomes. Mol. Phylogenet. Evol. 78:314–323.

Meiklejohn K.A., Faircloth B.C., Glenn T.C., Kimball R.T., Braun E.L. 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements (UCEs): evidence for a bias in some multispecies coalescent methods. Syst. Biol. 65:612–627.

Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA: IEEE, p. 1–8.

Minh B.Q., Nguyen M.A., von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30: 1188–1195.

Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548.

Misof B., Liu S., Meusemann K., Peters R.S., Donath A., Mayer C., Frandsen P.B., Ware J., Flouri T., Beutel R.G., Niehuis O., Petersen M., Izquierdo-Carrasco F., Wappler T., Rust J., Aberer A.J., Aspock U., Aspock H., Bartel D., Blanke A., Berger S., Bohm A., Buckley T.R., Calcott B., Chen J., Friedrich F., Fukui M., Fujita M., Greve C., Grobe P., Gu S., Huang Y., Jermiin L.S., Kawahara A.Y., Krogmann L., Kubiak M., Lanfear R., Letsch H., Li Y., Li Z., Li J., Lu H., Machida R., Mashimo Y., Kapli P., McKenna D.D., Meng G., Nakagaki Y., Navarrete-Heredia J.L., Ott M., Ou Y., Pass G., Podsiadlowski L., Pohl H., von Reumont B.M., Schutte K., Sekiya K., Shimizu S., Slipinski A., Stamatakis A., Song W., Su X., Szucsich N.U., Tan M., Tan X., Tang M., Tang J., Timelthaler G., Tomizuka S., Trautwein M., Tong X., Uchifune T., Walzl M.G., Wiegmann B.M., Wilbrandt J., Wipfler B., Wong T.K., Wu Q., Wu G., Xie Y., Yang S., Yang Q., Yeates D.K., Yoshizawa K., Zhang Q., Zhang R., Zhang W., Zhang Y., Zhao J., Zhou C., Zhou L., Ziesmann T., Zou S., Li Y., Xu X., Zhang Y., Yang H., Wang J., Wang J., Kjer K.M., Zhou X. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767.

Moyle R.G., Andersen M.J., Oliveros C.H., Steinheimer F.D., Reddy S. 2012. Phylogeny and biogeography of the core babblers (Aves: Timaliidae). Syst. Biol. 61:631–651.

Naylor G.J., Collins T.M., Brown W.M. 1995. Hydrophobicity and phylogeny. Nature 373:565–566.

Nguyen L.T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Nosenko T., Schreiber F., Adamska M., Adamski M., Eitel M., Hammel J., Maldonado M., Muller W.E.G., Nickel M., Schierwater B., Vacelet J., Wiens M., Worheide G. 2013. Deep metazoan phylogeny: when different genes tell different stories. Mol. Phylogenet. Evol. 67: 223–233.

Penny D. 2017. Phylogenetics: tertiary protein structures needed. Nat. Ecol. Evol. 1:0080.

Poe S., Chubb A.L. 2004. Birds in a bush: five genes indicate explosive evolution of avian orders. Evolution 58:404–415.

Poe S., Swofford D.L. 1999. Taxon sampling revisited. Nature 398: 299–300.

Pollock D.D., Zwickl D.J., McGuire J.A., Hillis D.M. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51:664–671.

Ponciano J.M., Burleigh J.G., Braun E.L., Taper M.L. 2012. Assessing parameter identifiability in phylogenetic models using data cloning. Syst. Biol. 61:955–972.

Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526:569–573.

Ragan M.A. 1992. Phylogenetic inference based on matrix representation of trees. Mol. Phylogenet. Evol. 1:53–58.

Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

Roch S., Steel M. 2014. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100:56–62.

Rokas A., Carroll S.B. 2006. Bushes in the tree of life. PLoS Biol. 4: 1899–1904.

Romiguier J., Ranwez V., Delsuc F., Galtier N., Douzery E.J. 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. Mol. Biol. Evol. 30:2134–2144.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Hohna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Rosenberg N.A. 2013. Discordance of species trees with their most likely gene trees: a unifying principle. Mol. Biol. Evol. 30:2709–2713.

Sainudiin R., Yoshida R. 2005. Applications of interval methods to phylogenetics. In: Pachter L., Sturmfels B., editors. Algebraic statistics for computational biology. New York: Cambridge University Press. p. 359–374.

Saitou N., Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Sanderson M.J., Wojciechowski M.F. 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-Astragalus (Leguminosae). Syst. Biol. 49:671–685.

Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33: 1654–1668.

Scornavacca C., Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. Syst. Biol. 66:112–120.

Simmons M.P., Pickett K.M., Miya M. 2004. How meaningful are Bayesian support values? Mol. Biol. Evol. 21:188–199.

Slowinski J.B. 2001. Molecular polytomies. Mol. Phylogenet. Evol. 19:114–120.
Smith J.V., Braun E.L., Kimball R.T. 2013. Ratite nonmonophyly: independent evidence from 40 novel loci. Syst. Biol. 62:35–49.
Soltis D.E., Albert V.A., Savolainen V., Hilu K., Qiu Y.L., Chase M.W., Farris J.S., Stefanovic S., Rice D.W., Palmer J.D., Soltis P.S. 2004. Genome-scale data, angiosperm relationships, and "ending incongruence": a cautionary tale in phylogenetics. Trends Plant Sci. 9:477–483.
Springer M.S., Gatesy J. 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94:1–33.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.
Stamatakis A., Hoover P., Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57:758–771.
Steel M. 2005. Should phylogenetic models be trying to "fit an elephant"? Trends Genet. 21:307–309.
Steel M., Matsen F.A. 2007. The Bayesian "star paradox" persists for long finite sequences. Mol. Biol. Evol. 24:1075–1079.
Stresemann E. 1959. The status of avian systematics and its unsolved problems. Auk 76:269–280.
Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. Zool. Scripta 45:50–62.
Suh A., Paus M., Kiefmann M., Churakov G., Franke F.A., Brosius J., Kriegs J.O., Schmitz J. 2011. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. Nat. Commun. 2:443.
Suh A., Smeds L., Ellegren H. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of Neoavian birds. PLoS Biol. 13:e1002224.
Sukumaran J., Holder M.T. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571.
Susko E. 2008. On the distributions of bootstrap support and posterior distributions for a star tree. Syst. Biol. 57:602–612.
Suzuki Y., Glazko G.V., Nei M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl Acad. Sci. USA. 99:16138–16143.
Swofford D.L. 2016. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.0a149. Available from: URL http://people.sc.fsu.edu/~dswofford/paup_test.
Swofford D.L., Waddell P.J., Huelsenbeck J.P., Foster P.G., Lewis P.O., Rogers J.S. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50:525–539.
Wang N., Braun E.L., Kimball R.T. 2012. Testing hypotheses about the sister group of the passeriformes using an independent 30-locus data set. Mol. Biol. Evol. 29:737–750.
Wang N., Kimball R.T., Braun E.L., Liang B., Zhang Z.W. 2017. Ancestral range reconstruction of Galliformes: the effects of topology and taxon sampling. J. Biogeogr. 44:122–135.
Wheeler T.J., Eddy S.R. 2013. nhmmer: DNA homology search with profile HMMs. Bioinformatics 29:2487–2489.
Whelan N.V., Kocot K.M., Moroz L.L., Halanych K.M. 2015. Error, signal, and the placement of Ctenophora sister to all other animals. Proc. Natl Acad. Sci. USA. 112:5773–5778.
Whelan S. 2008. The genetic code can cause systematic bias in simple phylogenetic models. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:4003–4011.
Wickett N.J., Mirarab S., Nguyen N., Warnow T., Carpenter E., Matasci N., Ayyampalayam S., Barker M.S., Burleigh J.G., Gitzendanner M.A., Ruhfel B.R., Wafula E., Der J.P., Graham S.W., Mathews S., Melkonian M., Soltis D.E., Soltis P.S., Miles N.W., Rothfels C.J., Pokorny L., Shaw A.J., DeGironimo L., Stevenson D.W., Surek B., Villarreal J.C., Roure B., Philippe H., dePamphilis C.W., Chen T., Deyholos M.K., Baucom R.S., Kutchan T.M., Augustin M.M., Wang J., Zhang Y., Tian Z., Yan Z., Wu X., Sun X., Wong G.K., Leebens-Mack J. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci USA. 111:E4859–E4868.
Yang Z. 1995. A space-time process model for the evolution of DNA sequences. Genetics 139:993–1005.
Yang Z., Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst. Biol. 54:455–470.
Yuri T., Kimball R.T., Harshman J., Bowie R.C.K., Braun M.J., Chojnowski J.L., Han K.L., Hackett S.J., Huddleston C.J., Moore W.S., Reddy S., Sheldon F.H., Steadman D.W., Witt C.C., Braun E.L. 2013. Parsimony and model-based analyses of indels in avian nuclear genes reveal congruent and incongruent phylogenetic signals. Biology (Basel) 2:419–444.
Zhang G., Li C., Li Q., Li B., Larkin D.M., Lee C., Storz J.F., Antunes A., Meredith R.W., Ödeen A., Cui J., Zhou Q., Xu L., Pan H., Wang Z., Jin L., Zhang P., Hu H., Yang W., Hu J., Xiao J., Yang Z., Liu Y., Xie Q., Lian J., Wen P., Zhang F., Li H., Zeng Y., Xiong Z., Liu S., Zhou L., Huang Z., An N., Wang J., Zheng Q., Xiong Y., Wang G., Wang B., Wang J., Fang Y., da Fonseca R., Alfaro-Núñez A., Schubert M., Orlando L., Mourier T., Howard J., Ganapathy G., Smith J., Farré M., Narayan J., Slavov G., Romanov M.N., Borges R., Machado J.P., Khan I., Springer M.S., Gatesy J., Hoffmann F.G., Opazo J.C., Håstad O., Greenwold M.J., Sawyer R.H., Kim H., Kim K.-W., Li N., Huang Y., Bruford M.W., Zhan X., Dixon A., Bertelsen M., Derryberry E., Warren W., Li S., Ray D.A., Green R.E., O'Brien S.J., Griffin D., Johnson W.E., Haussler D., Ryder O.A., Willerslev E., Graves G., Alström P., Fjeldså J., Mindell D., Edwards S.V., Braun E.L., Rahbek C., Burt D.W., Houde P., Zhang Y., Yang H., Wang J., Avian Genome Consortium, Jarvis E.D., Gilbert M.T.P., Wang J. 2014. Comparative genomics across modern bird species reveal insights into pan-avian genome evolution and trait biodiversity. Science 346:1311–1320.
Zwickl D.J., Hillis D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51:588–598.