Molecular Phylogenetics and Evolution 69 (2013) 1021–1032

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Identifying localized biases in large datasets: A case study using the avian tree of life

Rebecca T. Kimball [a,*], Ning Wang [a,b,1], Victoria Heimer-McGinn [a,2], Carly Ferguson [a,3], Edward L. Braun [a]

[a] Department of Biology, University of Florida, Gainesville, FL 32611, United States
[b] MOE Key Laboratory for Biodiversity Sciences and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 28 September 2012
Revised 12 May 2013
Accepted 29 May 2013
Available online 20 June 2013

Keywords:
Phylogenomics
Incongruence
Localized biases
Gene tree discordance

ABSTRACT

Large-scale multi-locus studies have become common in molecular phylogenetics, with new studies continually adding to previous datasets in an effort to fully resolve the tree of life. Total evidence analyses that combine existing data with newly collected data are expected to increase the power of phylogenetic analyses to resolve difficult relationships. However, they might be subject to localized biases, with one or a few loci having a strong and potentially misleading influence upon the results. To examine this possibility we combined a newly collected 31-locus dataset that includes representatives of all major avian lineages with a published dataset of 19 loci that has a comparable number of sites (Hackett et al., 2008. Science 320, 1763–1768). This allowed us to explore the advantages of conducting total evidence analyses, and to determine whether it was also important to analyze new datasets independent of published ones. The total evidence analysis yielded results very similar to the published results, with only slightly increased support at a few nodes. However, analyzing the 31- and 19-locus datasets separately highlighted several differences. Two clades received strong support in the published dataset and total evidence analysis, but the support appeared to reflect bias at a single locus (β-fibrinogen [FGB]). The signal in FGB that supported these relationships was sufficient to result in their recovery with bootstrap support, even when combined with 49 loci lacking that signal. FGB did not appear to have a substantial impact upon the results of species tree methods, but another locus (brain-derived neurotrophic factor [BDNF]) did have an impact upon those analyses. These results demonstrated that localized biases can influence large-scale phylogenetic analyses but they also indicated that considering independent evidence and exploring multiple analytical approaches could reveal them.

1. Introduction

Improvements in technology have made it possible to collect large, multi-locus datasets for phylogenetic studies. These multi-locus datasets may be extracted from whole genomes (Rokas et al., 2003; Wildman et al., 2007), reflect large-scale *de novo* data collection (Dunn et al., 2008; Hackett et al., 2008), the combination of data from multiple studies into a total evidence analysis (Kimball et al., 2011), or some combination of those approaches. These large datasets have led to increased resolution and support for many nodes in the Tree of Life. However, even with large datasets that appear likely to have the power to robustly identify phylogenetic relationships, conflicts among large-scale datasets have been identified (e.g., compare Dunn et al., 2008; Philippe et al., 2009; Schierwater et al., 2009). Although analyses that can identify some sources of conflict have been proposed (e.g., Philippe et al., 2011), it seems clear that unexpected clades recovered in phylogenetic analyses of large datasets, even those with high support values, should be considered hypotheses that should be subjected to additional tests.

Several phenomena can lead to high support in analyses of large molecular datasets. The simplest, and probably most common, is that the support reflects evolutionary history. However, both systematic and localized biases can result in incorrect estimates of phylogeny, sometimes with high levels of support. There has been substantial attention paid to systematic biases, such as long-branch attraction (Felsenstein, 1978) and convergence in base composition (e.g., Jeffroy et al., 2006; Phillips et al., 2004), which can result in strong non-historical signal, though the use of better-fitting models and noise reduction methods can sometimes address these biases (e.g., Braun and Kimball, 2002; Pratt et al., 2009).

* Corresponding author. Address: Department of Biology, P.O. Box 118525, University of Florida, Gainesville, FL 32611, United States. Fax: +1 352 392 3704.
E-mail address: rkimball@ufl.edu (R.T. Kimball).
[1] Present address: MOE Key Laboratory for Tropical Plant and Animal Ecology, College of Life Sciences, Hainan Normal University, Haikou 571158, China.
[2] Present address: Department of Biochemistry, University College Cork, Cork, Ireland.
[3] Present address: Department of Chemistry & Biochemistry, University of California Los Angeles, Los Angeles, CA, United States.

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.05.029

Likewise, there are specific cases where the majority of gene trees differ from the species tree (Degnan and Rosenberg, 2006). Additional challenges for phylogenetic estimation include alignment errors (Lake, 1991; Liu et al., 2010) and the incorrect identification of orthologs (Philippe et al., 2011). Finally, there are also examples where unexpected phylogenetic signal appears to be limited to individual genes or specific subsets of the genome (Katsu et al., 2009; Rokas et al., 2003). However, it is unclear how often, if at all, these localized biases result in misleading conclusions when large-scale datasets are analyzed, but the possibility that they can be problematic needs to be explored.

Since analyses using large datasets are expected to reduce the variance of the estimated phylogeny there has been limited concern regarding the specific gene regions collected for various studies. Moreover, it has been common to combine data collected as part of previous studies into these larger datasets (e.g., Gatesy et al., 2002; Kimball and Braun, 2008; Pratt et al., 2009; Shen et al., 2012; Thomson and Shaffer, 2010). This practice intrinsically results in datasets with overlapping genes and it could be problematic if one or more genes included in these analyses exhibit strong localized biases. It has been suggested that when enough loci are sampled, any misleading phylogenetic signal localized to specific loci should not affect the conclusions of phylogenetic analyses (Rokas et al., 2003). Indeed, the Rokas et al. (2003) phylogenomic analyses revealed that analyses using most collections of 20 or more genes supported the same phylogeny, despite the existence of substantial, often well-supported, incongruence among estimates of phylogeny based upon individual genes (suggesting some localized biases were likely present). Nonetheless it remains important to examine this more broadly to determine whether localized biases are generally unimportant in large-scale datasets and only systematic biases need to be considered.

The avian tree of life represents an interesting test case for this type of analysis. The topology of the avian tree has been particularly difficult to elucidate due to a rapid radiation at the base of the largest group of birds, Neoaves (which represents over 95% of all avian species; Sibley and Ahlquist, 1990). In fact, Neoaves has been suggested to represent a hard polytomy (Poe and Chubb, 2004), though two large-scale analyses (Ericson et al., 2006; Hackett et al., 2008) have identified supraordinal clades that appear strongly supported. Moreover, there was no incongruence between these two studies for well-supported nodes, though analyses of the larger dataset from Hackett et al. (2008) resulted in more nodes with support than Ericson et al. (2006). However, those two studies used some of the same loci, and thus could be affected by similar localized biases.

Two of the novel and strongly supported relationships in Hackett et al. (2008) have been re-evaluated using datasets that had no overlapping loci (Smith et al., 2013; Wang et al., 2012) and transposable element (TE) insertions (Haddrath and Baker, 2012; Suh et al., 2011). Both Wang et al. (2012) and Smith et al. (2013) searched for misleading phylogenetic signal, and uncovered no evidence that either localized or global biases affected the nodes in question. The conclusions from both of these studies were congruent with Hackett et al. (2008) for the specific relationships being examined (the limited taxon sampling in those studies prevented many additional relationships from being compared). McCormack et al. (2013) used a large number of ultraconserved elements (UCEs) from up to 32 species in Neoaves, providing the ability to test some additional clades identified by Hackett et al. (2008). The strongly supported groups in McCormack et al. (2013) largely corroborated the conclusions of Hackett et al. (2008), with a single conflict in one of two analyses (cf. McCormack et al., 2013, Fig. 2A *versus* B). However, the more limited taxon sampling of these subsequent studies make it difficult to determine whether localized biases can influence the conclusions of large-scale datasets like that used by Hackett et al. (2008).

Here we extended the Hackett et al. (2008) data matrix (hereafter the 19-locus dataset) by adding data from 31 loci, providing a total of 50 loci for analysis. To allow examination of all the higher-level relationships proposed by Hackett et al. (2008), we chose a sample of 77 taxa representing all major avian clades that were selected to break up long branches and to largely overlap with the taxa in Hackett et al. (2008). The additional 31 loci were focused on non-coding regions and resulted in a dataset that was similar in size to Hackett et al. (2008). We concatenated the two datasets into a 50-locus dataset and analyzed this using ML and partitioned ML methods with two different alignment approaches. After conducting the total molecular evidence analysis, we explored whether separate analyses of the 31-locus and 19-locus datasets supported similar clades and exhibited similar levels of bootstrap support relative to each other and to the combined 50-locus dataset. We also searched for localized biases with the potential to drive incongruence by comparing results from the 31- and 19-locus datasets. Finally, we estimated the species tree using individual gene trees. Although our approaches provided corroboration for many of the relationships found by Ericson et al. (2006) and Hackett et al. (2008), it also highlighted a localized bias that affected a small number of relationships that were supported by both studies. These results indicated that exploring independent evidence and multiple analytical strategies may provide useful information that is complementary to total evidence analyses.

2. Methods

2.1. Data collection

To generate the 31-locus data matrix, we added sequences to the data that were collected by Braun et al. (2011), Kimball et al. (2009), Smith et al. (2013), and Wang et al. (2012) (loci are listed in Supplementary material Table S1). None of the loci included in the 31-locus dataset were included in Hackett et al. (2008), so this dataset was independent of that study (as well as Ericson et al., 2006). The loci were non-coding regions, primarily introns (with the short segments of coding exon that flanked introns trimmed prior to analyses) but also including two untranslated regions (UTRs). The 50 loci were located on 17 chromosomes in the chicken genome; loci on the same chromosome are separated (e.g., Kimball et al., 2009) and thus unlikely to be linked. Since there appears to be strong conservation of chromosome structure in birds (Griffin et al., 2007), separation in the chicken genome suggests there should also be little or no linkage in other taxa.

The taxa used included all those from Smith et al. (2013), Wang et al. (2012), and the "moderate effort" taxon sample of Kimball et al. (2009), plus additional taxa to subdivide long branches and target the inclusion of at least two species in all major clades when possible (Supplementary material Table S2). Most species were included in Hackett et al. (2008); however, we added several additional species, including the zebra finch (*Taeniopygia guttata*) where data was taken from the draft genome (Warren et al., 2010), the kea (*Nestor notabilis*; added in Wang et al. (2012)), Darwin's rhea (*Pterocnemia pennata*; included in Harshman et al. (2008) and Smith et al. (2013)) and the black-legged seriema (*Chunga burmeisteri*; to provide a second taxon in Cariamidae, a family placed in an unexpected position by Hackett et al. (2008)). For these taxa we downloaded zebra finch data for the Hackett et al. (2008) loci and we amplified and sequenced some loci that were used in Hackett et al. (2008) for the other species added, allowing us to include these taxa in both datasets.

Sequences used for this study were amplified using primers and amplification conditions described by Kimball et al. (2009), Smith et al. (2013), and Wang et al. (2012). PCR products were precipitated using PEG:NaCl (20%:2.5 M) in preparation for direct sequencing. An ABI Prism™ 3100-Avant genetic analyzer (PE Applied Biosystems, Foster City, CA) was used to generate sequences using the ABI BigDye® Terminator v.3.1 chemistry. When necessary, amplicons were cloned into pGEM®-T (Promega) and purified using the Fast Plasmid Mini-kit (5 Prime GmbH). All samples were sequenced in both directions using amplification primers. Sequencher™ 4.1 (Gene Codes Corp.) was used to edit sequences and assemble double-stranded contigs. The novel sequences collected for this study have been deposited in Genbank with accession numbers KF298452-KF299553.

2.2. Sequence alignment

Alignment of non-coding sequences, which are prone to insertions and deletions (indels), is one of the greatest challenges associated with using regions such as nuclear introns and UTRs in phylogenetics (Liu et al., 2010; Morgan-Richards et al., 2008; Pratt et al., 2009; Shapiro and Dunbacher, 2001). Wang et al. (2012) generated a large number of alternative sequence alignments for a subset of these data using many different alignment programs, parameter sets, guide trees, and approaches for excluding sites that are difficult to align. Those approaches demonstrated that the topological differences among analyses of different alignments were largely restricted to regions of low support. Automated methods can be problematic, in part due to the number and size of the indels (Creer, 2007). For example, insertion of transposable elements (TEs) in a single taxon can substantially lengthen (sometimes $\geqslant$500 bp) an intron relative to the orthologous intron in most or all other taxa. Moreover, independent TE insertions can occur at very close locations (Han et al., 2011), introducing homologous sequences at different positions within a locus. Although TE insertions can provide phylogenetic information (e.g., Haddrath and Baker, 2012; Suh et al., 2011), many TE insertions are present in only one or a few taxa so the sites within the insertion are likely to provide little phylogenetic signal. Overall, the large indels due to TE insertions provide limited information (unless a much larger number of TE insertions can be obtained; e.g., Churakov et al., 2009; Nishihara et al., 2009) and they present a major challenge for alignment programs that can alter the quality of the alignment downstream of the insertion.

To avoid biases due to manual alignments (Anisimova et al., 2010) while still avoiding problems due to large insertions we took a two-stage strategy for alignment. Most large ($\geqslant$100 bp) insertions in avian introns correspond to TE insertions (Han et al., 2011) so we focused on excluding these problematic sequences. First, we used the CENSOR software (Kohany et al., 2006) to identify all TE insertions in the sequences. CENSOR conducts a homology search of Repbase, a database of TE sequences (Jurka et al., 2005), and uses the results of that homology search to identify TE insertions in unknown sequences. We restricted our searches of Repbase to those TEs that have been identified in the *Gallus gallus* genome to maximize our ability to identify avian TEs. Then, the TE insertions were replaced with a short uninformative sequence (“NNN”) to allow identification of the original insertion site. Second, we aligned the sequences (after excluding the TE insertions) for each locus using both MUSCLE 3.6 (Edgar, 2004) and MAFFT 6.857beta (Katoh et al., 2009), using the default settings for each program. Wang et al. (2012) analyzed a subset of this data using multiple alignment programs, parameter settings, and strategies, and found that MUSCLE and MAFFT, using default parameters and without the need for post-alignment manipulations, performed well on avian nuclear introns in the absence of *a priori* defined guide trees based upon core scores calculated using T-Coffee (Notredame et al., 2000). This allowed us to obtain two alternative alignments without the constraints of using *a priori* defined guide trees. Finally, after we aligned all loci they were concatenated to yield a total evidence dataset with all 50 loci. We also generated separate datasets using the 19 loci used by Hackett et al. (2008) and the 31 loci that were not included in Hackett et al. (2008).

2.3. Phylogenetic analyses

RAxML (Stamatakis, 2006) was used to obtain the optimal tree using the ML criterion as well as the bootstrap consensus tree for each of the datasets. Bootstrapping used the slow bootstrap option and 500 replicates. We performed both unpartitioned analyses (where all loci were analyzed using a single set of GTR + Γ model parameters) and partitioned analyses in which each locus was considered a unique partition (with distinct GTR + Γ model parameters). This resulted in four analyses using two different alignment methods (MUSCLE/MAFFT) and two different analytical approaches (unpartitioned/partitioned) for each set of loci (the 50-locus dataset, and the separate 19-locus and 31-locus datasets).

We used NJ_{st} (Liu and Yu, 2011) and STAR (Liu et al., 2009) to estimate a species tree from estimates of individual gene trees. Both optimal and bootstrap trees from the RAxML analyses of the MUSCLE alignment were used (see above for RAxML methods). STAR is limited to rooted gene trees so the RAxML analyses used to generate trees were conducted using a single paleognath as an outgroup (*Struthio* was used for all but three loci where *Struthio* was absent and *Dromaius* was used instead). Paleognath sequences could not be obtained for one locus, OPN, so this locus was not included in the STAR analyses. In contrast, NJ_{st} can use unrooted gene trees and it was used to analyze gene trees generated using all taxa as well as those generated using a single outgroup. All of these analyses were conducted using the STRAW server (Shaw et al., 2013).

We explored the differences among trees using several different approaches. First, we extracted bootstrap support values for the two trees under comparison using a perl script written by ELB. We identified the number of nodes that (1) were strongly supported in both ($\geqslant$95% bootstrap support in both), (2) were well-supported in both ($\geqslant$70% bootstrap support [cf. Hillis and Bull, 1993] in both but not $\geqslant$95% in both), (3) were well-supported in one and present in the other ($\geqslant$70% bootstrap support in one and present at <70% in the other), and (4) nodes that conflict ($\geqslant$70% bootstrap support in one while an alternative arrangement receives at least $\geqslant$50% bootstrap support in the other). Second, we quantified differences among trees using the Robinson–Foulds (RF) distances (Robinson and Foulds, 1981) calculated using hashrf Sul and Williams (2008). We multiplied RF distances obtained using hashrf by two to make them comparable to the values typically reported in the literature (e.g., Wang et al., 2012). Finally, we explored the basis of the differences between the 19- and 31-locus datasets by identifying sites with a substantially greater likelihood given each of the two topologies recovered using each alignment and analytical method (separately for the MUSCLE/MAFFT and partitioned/unpartitioned analyses). To do this, we calculated site likelihoods for a combined (50-locus) data matrix using RAxML (via the “-f g” option in the program) based on either the ML tree estimated from the 31-locus data matrix or the ML tree estimated from the 19-locus data matrix. Sites with $\Delta \ln L > 5$ were viewed as “decisive sites” that strongly favor one of the two topologies.

Rates of evolution at each site were estimated using the Meyer and von Haeseler (2003) method, as implemented in IQPNNI version 3.3 (Vinh and von Haeseler, 2004). We used this method for the rate estimates at each site because it does not require any prior assumptions about the rate distribution. Briefly, a search for the optimal tree using the combined (50-locus) dataset was conducted in IQPNNI and

Table 1
Comparison of bootstrap support between analyses of different alignments and analytical strategy. Only nodes receiving at least 50% bootstrap support in at least one tree are considered.

Comparison	At least one ⩾50%	Both ⩾95%	Both ⩾70% but not both ⩾95%	One ⩾70%, one present <70%	One ⩾70%, absent in the other
Muscle versus Mafft					
50 loci, partitioned	62	44	9	1	1
50 loci, unpartitioned	62	43	8	3	0
31 loci, partitioned	55	41	3	5	1
31 loci, unpartitioned	54	40	6	4	1
19 loci, partitioned	64	42	9	3	0
19 loci, unpartitioned	60	42	9	2	0
Partitioned versus unpartitioned					
50 loci, Muscle	62	45	8	2	0
50 loci, Mafft	60	43	9	2	0
31 loci, Mafft	50	40	6	1	0
31 loci, Muscle	54	41	6	4	0
19 loci, Mafft	62	42	8	2	0
19 loci, Muscle	62	43	10	1	0

the data written to the ".rate" file were used as estimates of the evolutionary rats for each site. The distribution of rates was characterized by calculating the median and various percentiles after excluding both the invariant and the very fast (undefined rate) sites.

3. Results

3.1. Comparison of the datasets, alignments, and partitioning

The 50-locus dataset, with TE insertions removed, had over 57,500 sites (the MUSCLE and MAFFT alignments differed, with the MUSCLE alignment being slightly longer; Supplementary material Table S3). The alignment of the novel 31-locus dataset was slightly shorter than the alignment of the 19-locus dataset from Hackett et al. (2008). However, the 31-locus dataset included almost exclusively more rapidly evolving non-coding data (a small exon in between two introns was included in CALB1) whereas the 19-locus dataset included both coding and non-coding regions. Thus, the differences were smaller when only the number of variable sites in each was considered (Supplementary material Table S3). Although there may be some differences between the datasets, both datasets are similar enough in size that both should have a reasonable chance of recovering short branches in the avian tree of life (Chojnowski et al., 2008).

There were few differences between alignment methods or the use of partitioned ML analyses. In all cases, there were many nodes that were weakly supported in each comparison (for all analyses, more than half of all nodes received less than 50% bootstrap support), as well as many nodes that were strongly supported in each comparison (Table 1). When alignment methods were compared there were only a small number of nodes that would have been considered at least well-supported in one analysis that were not well-supported in the other. In only three cases was there a conflict in topology involving a well-supported node present in one of the topologies, and the conflicting position received very low bootstrap support. Even fewer differences were found between the partitioned and unpartitioned ML analyses of the same dataset and alignment (Table 1). Since the number of differences among these analyses was limited and MUSCLE appeared to outperform MAFFT in previous analyses of a subset of these data (Wang et al., 2012) we will focus on analyses using the MUSCLE alignments hereafter.

3.2. Avian phylogeny based upon analyses of the concatenated 50-locus dataset

The total evidence (50-locus) tree (Fig. 1) is very similar to that of Hackett et al. (2008). Most of the differences between the topology identified for this study and the Hackett et al. (2008) topology (for the taxa in common) reflected nodes that were not strongly supported in either study. There are two well-supported nodes in Hackett et al. (2008) that were not present in this study, and both occur within the waterbirds (these differences are due to exchanging the position of *Gavia* [loon] with that of the *Eudyptula-Oceanodroma* [penguin-tubenose] clade).

3.3. Avian phylogeny based upon analyses of the concatenated 19- and 31-locus datasets

Given that the estimate of phylogeny based upon the 50-locus dataset was very similar to that obtained by Hackett et al. (2008), we expected the separate 31-locus and 19-locus datasets to have similar phylogenetic signal, and that is generally what we observed. Most nodes that were very strongly supported in one dataset (i.e., nodes with ⩾95% bootstrap support) also received similar levels of support in analyses of the other dataset (Fig. 2, Table 2). Likewise, most nodes that received <50% bootstrap support with one dataset also received very low bootstrap support in the other (approximately 58% of nodes had <50% bootstrap support in analyses of both datasets).

However, despite the high degree of congruence between the two datasets (Fig. 2), there were also a few surprising differences. Depending upon the comparison, there were nine or 10 nodes with ⩾70% support in one dataset and <70% support (or absent) in the other (Table 2). Across all four comparisons between the 19 and 31-locus datasets, there were 39 nodes in this category: 38% (15 of 39) of these were nodes present in both trees but with different levels of bootstrap support (Table 2) while the remainder reflect discordance between the bootstrap consensus trees. However, in most cases, the discordance was between one well-supported node and one very weakly (<50%) supported node, with only 13% of nodes exhibiting a greater degree of conflict (Table 2). Thus, the strongly supported nodes were very consistent between the data matrices, with just a handful of cases that resulted in a different set of relationships between datasets (and even in these cases the conflict did not involve conflict between two well-supported alternatives).

3.4. Are estimates of phylogeny for the 50-locus dataset more similar to those for the 19- or 31-locus datasets?

In large part, conclusions based upon analyses of the concatenated 50-locus dataset exhibited more similarities to the conclusions based upon the 19-locus dataset, at least with respect to the placement of clades that are discordant between the datasets.

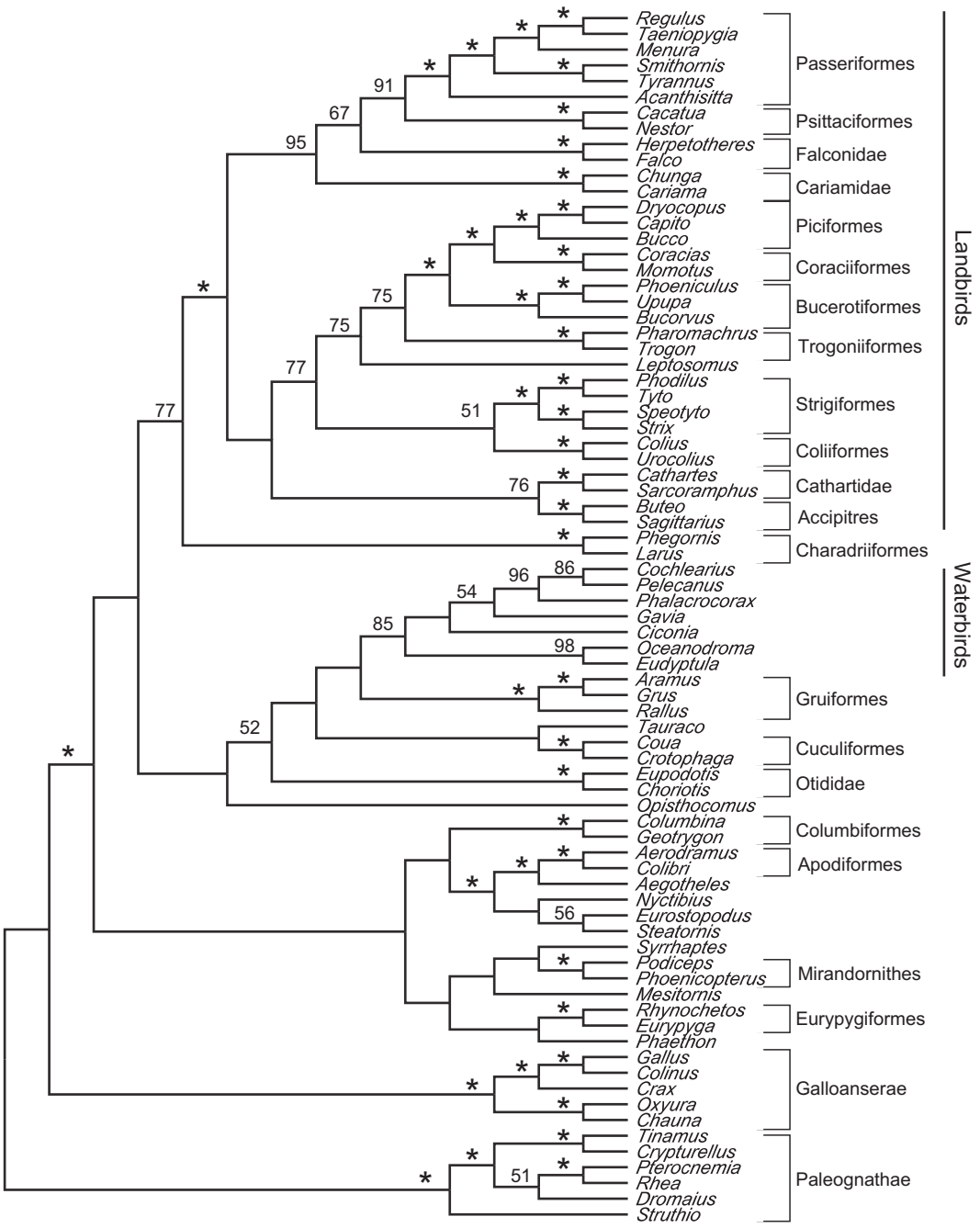


Fig. 1. ML tree estimated using a partitioned analysis of the 50-locus dataset aligned using MUSCLE alignment. Values at nodes represent ML bootstrap percentages. Support values were only reported if they were ⩾50%; * indicates 100% bootstrap support. Other analyses (the unpartitioned analysis using the MUSCLE alignment and both the partitioned and unpartitioned analyses using the MAFFT alignment) were very similar (Supplementary information). Taxonomic groups are indicated to the right of the trees.

In fact the RF distance between the 50- and 19-locus trees generated by partitioned analyses of the MUSCLE alignment was 16 whereas the RF distance between the 50- and 31-locus trees was 44; this observation was general, with RF distances between 50- and 19-locus trees ranging from 12 to 36 and those between the 50- and 31-locus trees ranging from 40 to 52 (fairly similar to the RF distances between the 19- to 31-locus datasets, which ranged from 48 to 54). Whether this pattern reflects the slightly larger number of variable sites in the 19-locus dataset, or whether, by chance, those loci used by Hackett et al. (2008) have stronger signal than the 31-loci is not clear from these analyses.

Since the 50-locus dataset is nearly double the size of the 19- and 31-locus datasets, it would be expected to exhibit higher bootstrap support on average, and therefore a larger number of well-resolved nodes. When considering nodes with at least 50% bootstrap support in tree, the bootstrap values per node are greater in the 50-locus tree by an average of 1.0–7.0%, depending upon which data matrix, alignment, or analysis is being compared. The improvement is lower for the 19- to 50-locus dataset comparisons (1.0–2.7%) than it is for the 31- to 50-locus dataset comparisons (4.5–7.0%), consistent with the suggestion that there may be greater phylogenetic signal in the 19-locus dataset. However, combining the datasets did not result in a greater number of well-supported nodes relative to the individual 19 or 31-locus datasets, suggesting that it will require substantially more data to resolve many of the weakly supported nodes (assuming the relevant node does not reflect a hard polytomy and thus has the potential to be resolved).

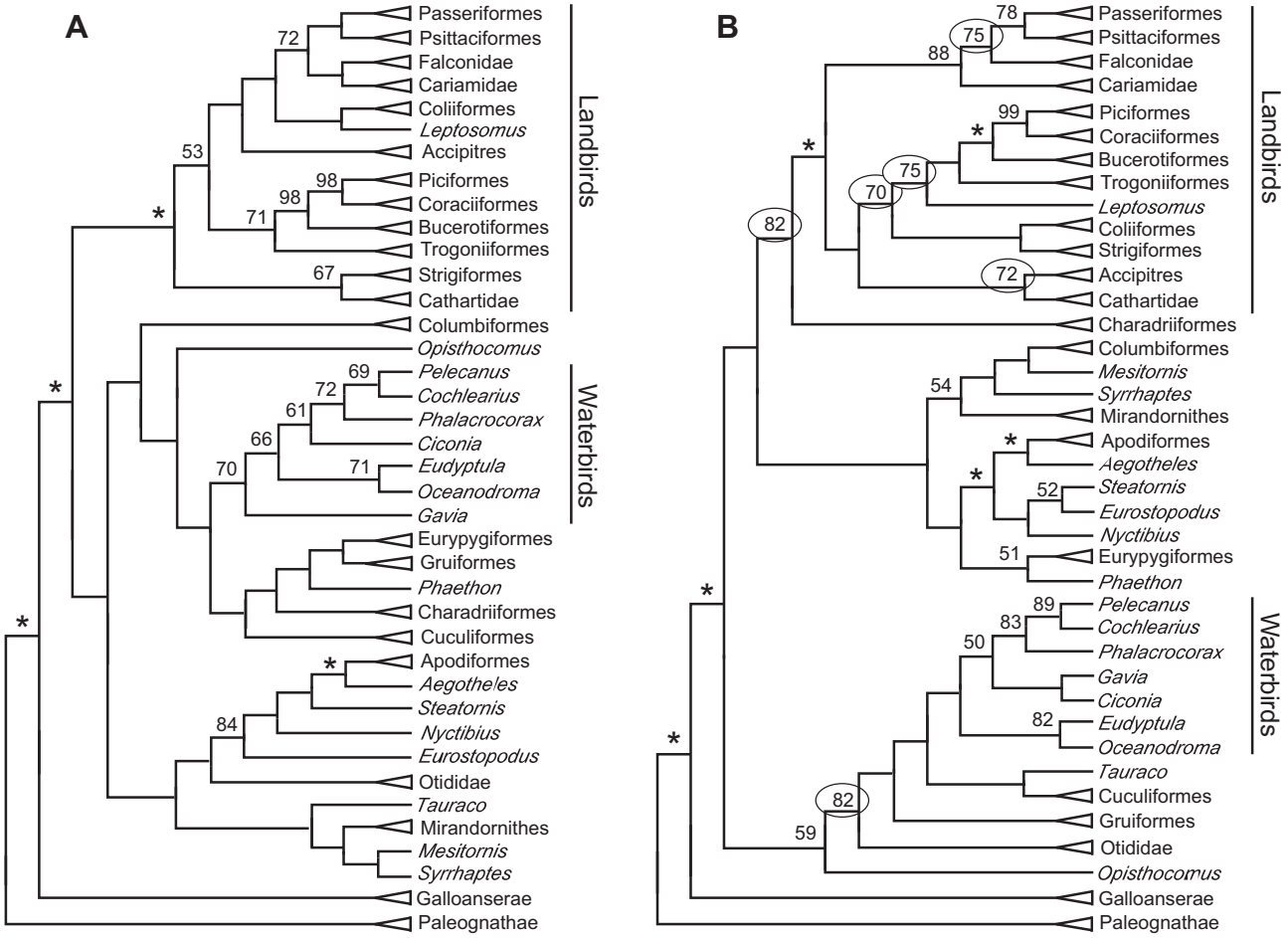


Fig. 2. ML trees estimated using the partitioned analysis of the 31-locus dataset (A) and the 19-locus dataset (B) aligned using MUSCLE. Values at nodes represent ML bootstrap percentages. Support values were only reported if they were ⩾50%; * indicates 100% bootstrap support. Strongly supported clades from Fig. 1 that correspond to orders and families are collapsed. Nodes present at ⩾70% bootstrap support in one topology but absent in the other are circled. Other analyses (the unpartitioned analysis of the MUSCLE alignment and both the partitioned and unpartitioned analyses of the MAFFT alignment) were very similar (Supplementary information).

Table 2
Comparison of bootstrap support between different data partitions. Only nodes receiving at least 50% bootstrap support in at least one tree are considered.

Comparison	At least one ⩾50%	Both ⩾95%	Both ⩾70% but not both ⩾95%	One ⩾70%, one present <70%	One ⩾70%, absent in other (# of those supported ⩾50%)
31 versus 19 loci					
Muscle, partitioned	66	42	4	4	6 (2)
Muscle, unpartitioned	63	41	6	3	7 (1)
Mafft, partitioned	65	40	4	4	5 (1)
Mafft, unpartitioned	63	40	4	4	6 (1)

3.5. A single locus had a major impact upon analyses of concatenated datasets

To understand why there might be different signal in the two datasets, we identified "decisive sites" – those that strongly supported either the 19- or 31-locus topology (Fig. 3). Most loci have only a small number of sites that strongly support one of the two topologies, a finding consistent with the relatively modest differences between the trees (most differences involve rearrangements of short branches). The decisive sites have a relatively high rate of substitution with a median rate (calculated using the Meyer and von Haeseler (2003) method) approximately four times that for variable sites overall. However, there were a large number of these decisive sites in a single locus (FGB), suggesting that much of the difference in phylogenetic signal between the 19- and 31-locus datasets reflects the influence of that locus specifically.

Excluding FGB alters two key conclusions (compare Fig. 1 to Fig. 4), indicating that this locus may explain some of the conflicts between datasets that we observed. Both the 19- and 50-locus topologies place the Charadriiformes (shorebirds) as sister to the landbirds (e.g., Figs. 1 and 2B; see also Hackett et al., 2008), though this node is not present in analyses of the 31-locus dataset or when either FGB or the decisive sites are excluded from the combined 50-locus dataset (e.g., Figs. 2A and 4) or the 19-locus dataset (not shown). Another well-supported clade in the 19-locus dataset (Fig. 2B, see also Hackett et al., 2008) that was still present (albeit with lower support) in analysis of the 50-locus dataset places the waterbirds within a larger clade (defined as Insolitaves by Yuri et al., 2013) that also includes Cuculiformes (cuckoos),

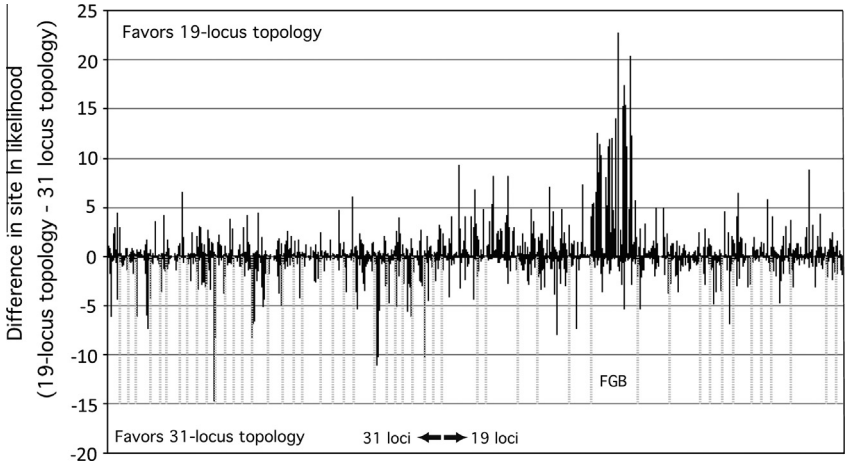


Fig. 3. Differences in site likelihoods for a partitioned analysis of the 50-locus dataset given the optimal topologies for the 19- and 31-locus datasets aligned using MUSCLE. Positions in the alignment are arranged from the left (site 1) to right (site 57818), with loci concatenated in alphabetical order for each (31- and 19-locus) dataset. Dashed gray lines indicate boundaries between loci in the alignment.

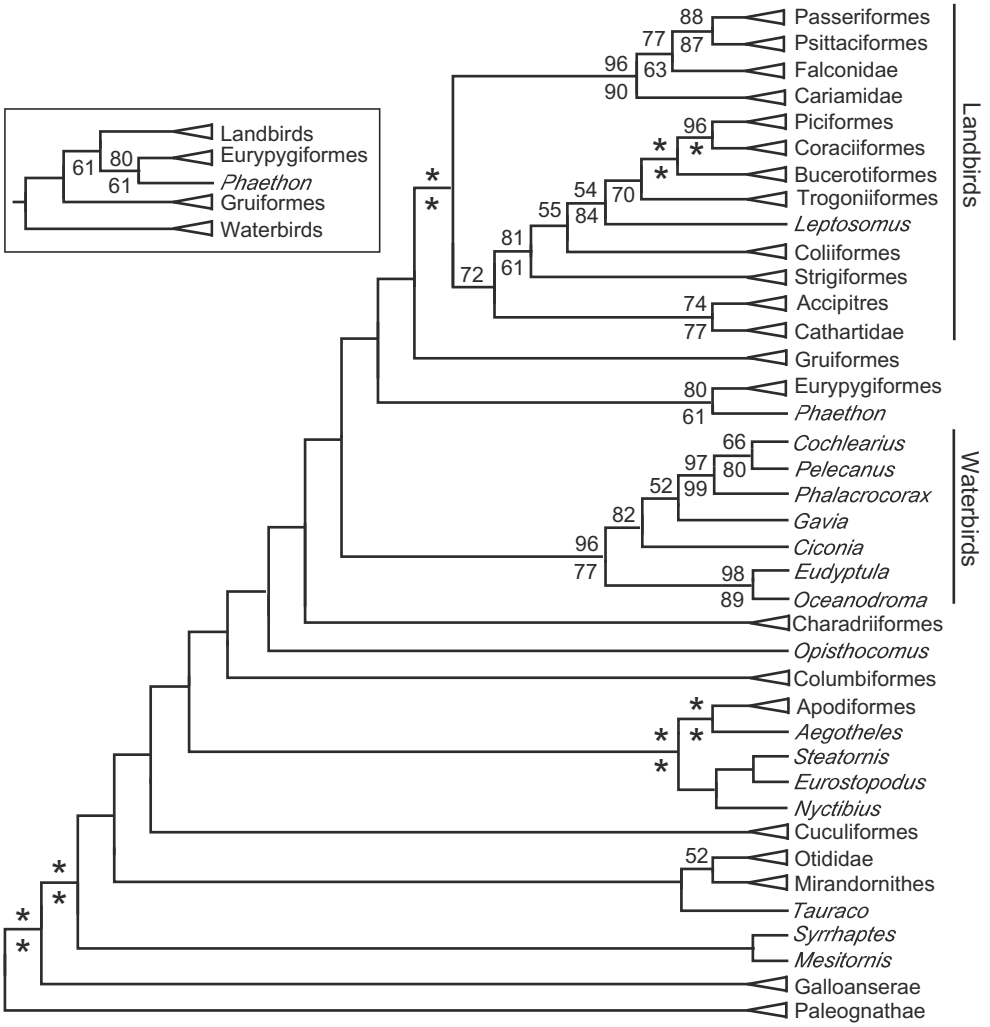


Fig. 4. ML tree estimated using a partitioned analysis of the 49-locus (FGB excluded) dataset aligned using MUSCLE. Bootstrap values are reported for nodes that received ⩾50% support; as above, * indicates 100% bootstrap support. Values above the line reflect the analysis of the 49-locus dataset, while those below the line are bootstrap values for the 50-locus dataset with the decisive sites excluded. The decisive sites excluded topology shows a single difference from the 49-locus (FGB excluded) topology with ⩾50% bootstrap support (see inset for the decisive sites excluded topology; only bootstrap values from the decisive sites excluded analysis are shown). Strongly supported clades corresponding to orders and families are collapsed, as above. Other analyses (the unpartitioned analysis and both partitioned and unpartitioned analyses using the MAFFT alignment) were very similar (Supplementary information).

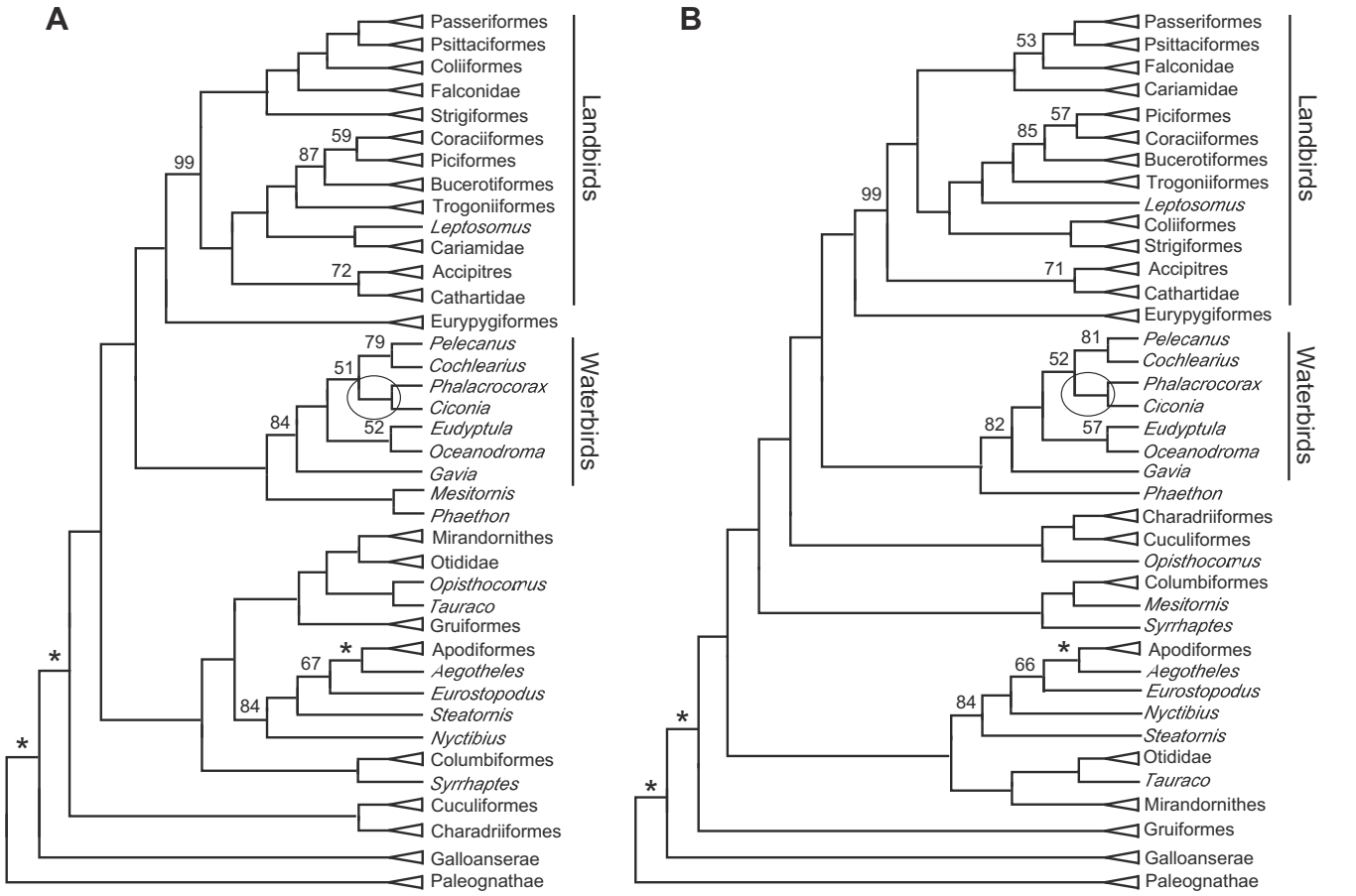


Fig. 5. Species trees estimated in NJ_{st} using the optimal ML gene trees from the MUSCLE alignments. Values at nodes represent the percentage of species trees estimated from the bootstrapped gene trees (only values ⩾50% are shown); * indicates present in 100% of species trees. Strongly supported clades from Fig. 1 that correspond to orders and families are collapsed. Circled nodes represent rearrangements between the species trees estimated from the optimal and bootstrap gene trees; in all cases, support was <70%. (A) All 50 loci analyzed. (B) 49 loci (BDNF excluded).

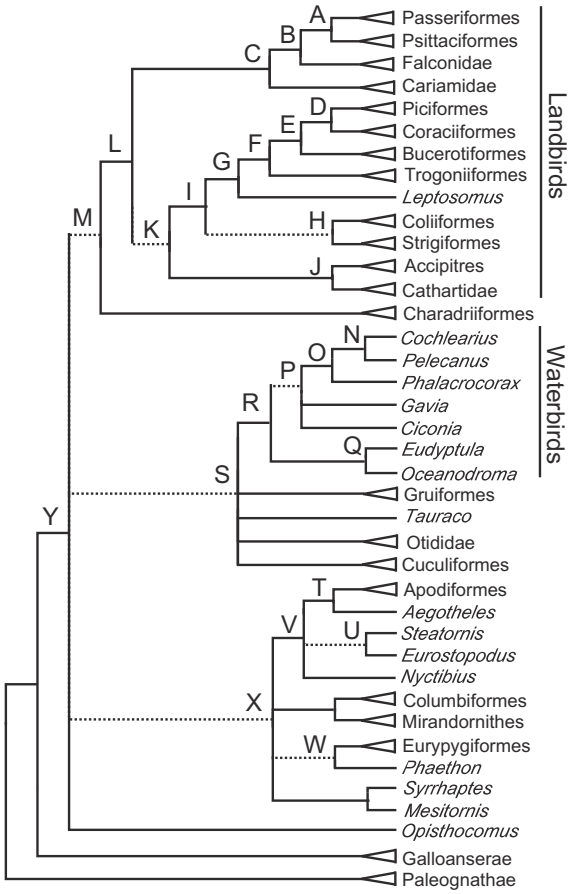
Gruiformes (rails, storks, and allies), Otididae (bustards) and *Tauraco* (turacos) (Fig. 1). Insolitaves was also absent in the 31-locus, FGB excluded (49-locus), and decisive sites excluded (from all 50 loci) trees (Figs. 2A and 4). Thus, the strong signal in one locus (FGB) can drive relationships even when combined with 49 other loci lacking that signal. The concentration of decisive sites in a single locus is unexpected so these results are most consistent with the hypothesis that FGB is providing at least some misleading signal.

3.6. A different locus had a major impact upon NJ_{st} analyses

The large impact of FGB on concatenated analyses raises the question of whether methods that estimate species trees from individual gene trees might also be biased by unusual and strong phylogenetic signal limited to one or a few loci. The species tree estimated using NJ_{st} resulted in another distinctive topology (Fig. 5A), albeit with lower overall bootstrap support than the analyses of the concatenated dataset. There were multiple conflicts between well-supported nodes in both the 50-locus and the 49-locus analysis of the concatenated data (Figs. 1 and 4) and the 50-locus NJ_{st} analysis, particularly within the landbirds. Surprisingly, the results of analyses using STAR (Supplementary material Treefile) were more congruent with analyses of the concatenated datasets (particularly Fig. 4, where FGB was excluded). A major difference between STAR and NJ_{st} is that the former required the use of a single paleognath sequence as an outgroup, making it difficult to determine whether the topological differences between the trees estimated using STAR and NJ_{st} could reflect differences between algorithms or taxon sampling. To differentiate between these possibilities we analyzed the gene trees with a single outgroup using NJ_{st} (Supplementary material Treefile), revealing the major topological differences reflected taxon sampling rather than algorithmic differences.

We hypothesized that the topological differences between the 50-locus, NJ_{st} tree with all taxa and the NJ_{st} tree with a single paleognath outgroup reflected the inclusion of the gene tree for the BDNF locus. In previous studies, analyses of BDNF alone failed to recover many well-corroborated monophyletic clades, including Passeriformes, Neoaves, and Palaeognathae (Hackett et al., 2008). All other individual loci included in Hackett et al. (2008) were able to recover these groups, as do individual loci in other studies (e.g., Ericson et al., 2006). Exclusion of the BDNF gene tree did increase congruence between the NJ_{st} tree (Fig. 5B) and the concatenated 49-locus tree (Fig. 4), corroborating this hypothesis, though exclusion of the BDNF locus did not have a substantive impact upon concatenated analyses (Fig. 6).

There did appear to be a modest difference when the FGB locus was excluded from NJ_{st} analyses. The Eurypygiformes-*Phaethon* clade (Node W, Fig. 6) was present in the NJ_{st} tree when FGB was excluded but absent when it was included. Exclusion of FGB from analyses of the concatenated data resulted in a substantial increase in the bootstrap support for that clade (compare Fig. 1 and Fig. 4). The Eurypygiformes-*Phaethon* clade is also present in analyses of UCE data (McCormack et al., 2013) and in an analysis of indels in



	Concatenation			NJ_{st}		
Node	50-locus	no FGB	no BDNF	50-locus	no FGB	no BDNF
A	91	88	82			
B	67	77	63		50	53
C	95	96	95			
D	100	96	98	59		57
E	100	100	100	87	83	85
F	75		72			
G	75	54	75			
H	51					
I	77	81	77			
J	76	74	70	72	69	71
K		72				
L	100	100	100	99	98	99
M	77		74			
N	86	66	87	69	74	81
O	96	97	96	53	58	55
P	54	82	60			
Q	98	98	98	52	53	57
R	85	96	89	84	76	82
S	52		60			
T	100	100	100	100	100	100
U	56					
V	100	100	100	84	80	84
W		80	54			
X						
Y	100	100	100	100	100	100

Fig. 6. Summary phylogeny showing nodes of interest, and their support in different analyses. Shaded cells indicate the presence of the clade in the optimal tree for each analysis, while white cells indicate the absence of the clade. Values are bootstrap support for those clades (only values ⩾50% are shown). For Node O, two of the optimal NJ_{st} trees lacked the clade but it was present (albeit with low support) in the corresponding bootstrap consensus trees. The tree is based upon the analysis of the concatenated 50-locus dataset. With two exceptions (Node W, Eurypygiformes-*Phaethon*, and Node X, the "Metaves" clade in Hackett et al., 2008), the nodes with <50% bootstrap support are collapsed. Branches absent in the trees based upon analysis of concatenated data excluding FGB and NJ_{st} analyses excluding BDNF are dashed.

a subset of the data analyzed here (Yuri et al., 2013). Thus, NJ_{st} largely accommodated the incongruent signal in FGB but was strongly affected by the phylogenetic signal in BDNF, while analyses of concatenated data were strongly biased by the inclusion of FGB but not BDNF.

4. Discussion

4.1. Independent evidence corroborates a difficult phylogenetic problem

Although the total evidence analysis (Fig. 1) included many well-supported nodes, it was the independent analyses of the 31- and 19-locus datasets that emphasized that distinct phylogenetic signals exist in different parts of the genome and highlighted which relationships in Hackett et al. (2008) were robust and were likely to represent the species tree. For example, the clade comprising Cathartidae (New World vultures) and Accipitres (hawks and allies) was present in Hackett et al. (2008) but had limited bootstrap support (61%). This clade conflicted with suggestions that the Cathartidae are allied to Ciconiidae (storks), a hypothesis based on DNA-DNA hybridization data (Sibley and Ahlquist, 1990) and corroborated by behavioral and ecological data (Ligon, 1967). Analyses of the concatenated 50-locus dataset and NJ_{st} analyses had higher (>70%) bootstrap support for the Accipitres–Cathartidae clade. However, despite the increased support for this clade when the new data were added, the Accipitres–Cathartidae clade was not present in the analysis of the 31-locus dataset when all sites were included (Fig. 2A). It was present (albeit weakly supported) in analyses of the 31-locus analyses after excluding decisive sites (Supplementary material Treefile) and it clearly emerges in the context of the other data (e.g., UCE data; McCormack et al., 2013). Thus, phylogenetic signal supporting the Accipitres–Cathartidae clade is likely spread across the genome but it is clearly not universal due to incomplete lineage sorting, variation in patterns of molecular evolution, and/or other processes. Regardless, these results emphasized that careful exploration of the data motivated by examining analyses of independent datasets can reveal this complexity and increase our confidence in phylogenetic conclusions.

While most well-supported relationships from Hackett et al. (2008) were corroborated by comparison of the 31- and 19-locus analyses, our comparisons also revealed relationships that were relatively well supported in Hackett et al. (2008) that may not be valid. This included the two relationships that appeared driven by FGB – the landbird-Charadriiformes clade (Node M, Fig. 6) and Insolitaves (Node S, Fig. 6). In addition, some relationships within well-supported clades (i.e., relationships within the landbirds and waterbirds) also varied substantially in levels of support and may also be worthy of further study.

4.2. Identification of loci with strong impact upon analyses

Our results showed that even in large, multi-locus datasets, strong signal from a single partition (FGB or BDNF) has the

potential to drive specific relationships in different analyses. FGB had a major impact upon analyses of concatenated data (Fig. 4) whereas BDNF appeared to have an impact upon the species tree estimated from the gene trees (Fig. 5). Excluding BDNF from concatenated analyses had essentially no impact upon the topologies resulting from those analyses (Fig. 6) and excluding FGB from NJ_{st} had a limited impact upon that analysis (see above, Section 3.6). This suggests that each analytical approach was largely affected by only one of these two loci. Only by careful examination of analyses using different subsets of the data (e.g., 31- *versus* 19-loci) and the different analytical approaches (analyses of concatenated data and estimates using individual gene trees) was it possible to determine that even when a relatively large number of loci are used one locus can drive specific conclusions.

4.2.1. Localized biases and analyses of concatenated datasets

Previous studies have demonstrated that excluding FGB from concatenated analyses resulted in some topological changes (Hackett et al., 2008; Mayr, 2011). However, FGB was one of the longer and more variable loci in both of those studies. Thus, excluding FGB could have reduced the power to resolve relationships leading to the observed topological differences. The observation that exclusion of FGB leads to a loss of relationships in this larger dataset (Fig. 4) indicates that this is not the case. Indeed, we demonstrated that FGB has an unusually large number of sites that strongly supported one of our two candidate topologies (Fig. 2) providing a clear illustration of the unusual signal present in this locus.

There are several alternative explanations for the observed pattern. First, FGB may have an unusually discordant gene tree that simply reflects the stochastic nature of the coalescent (see above, Section 4.2). Second, FGB might have a discordant gene tree if ancestral polymorphisms at that locus were maintained for unusually long period of time by balancing selection. Finally, the FGB gene tree may differ from the species tree due to an introgression event shortly after the radiation of Neoaves, a scenario that has been suggested to explain incongruence observed in deep mammalian phylogeny (e.g., Churakov et al., 2009; Hallström and Janke, 2010). Regardless, the observation that this signal is present in only one of the 50 loci suggests it is unlikely to reflect the species tree even if it does reflect the FGB gene tree. Indeed, we emphasize that the first large-scale analysis of the FGB locus (Fain and Houde, 2004) acknowledged the possibility that the FGB gene tree might differ from the species tree for specific clades of interest (like the gene tree for any individual locus), although it was not clear whether that was the case since that was a single-locus study.

Alternatively, it is possible that the decisive sites represent misleading phylogenetic signal, in which case the estimate of the FGB gene tree is likely to be inaccurate. Although it is difficult to differentiate between the hypothesis that the FGB gene tree actually differs substantially from the avian species tree and the hypothesis that there is misleading information present in that locus, the higher rate of substitution for the decisive sites may be more consistent with the latter hypothesis. Moreover, we observed that deleting the fastest sites in the 50-locus dataset greatly reduced support for the landbird-Charadriiformes clade (Node M, Fig. 6) while retaining support for many other nodes present in both the 19- and 31-locus datasets (Supplementary material Treefile). Regardless of the specific explanation for the unexpected concentration of the decisive sites in FGB, it seems unlikely that their signal is representative of the avian species tree.

4.2.2. Incongruent gene trees and NJ_{st} analyses

The unique and almost certainly misleading phylogenetic signal present in BDNF has also been noted previously (Hackett et al., 2008; Harshman et al., 2008; Smith et al., 2013), though in none of these studies did it appear that inclusion of BDNF in analyses of concatenated datasets had an impact upon the estimate of phylogeny. The base composition of BDNF exhibits strong deviations from stationarity (Harshman et al., 2008; Smith et al., 2013), strongly suggesting that at least some of the observed relationships in the BDNF gene tree reflect convergence in base composition rather than historical signal. Thus, the idea that BDNF might have a negative impact upon phylogenetic analyses might be expected. Nonetheless, the absence of a detectable impact upon analyses using concatenated data (see also Hackett et al., 2008; Harshman et al., 2008; Smith et al., 2013) combined with the observable impact upon the 50-locus NJ_{st} analysis is surprising.

Although analyses of concatenated data may be misleading (Kubatko and Degnan, 2007), in our case it is the species tree estimated from the estimates of gene trees for all 50 loci (Fig. 5A) rather than the analysis of concatenated data that lacked specific clades that are very likely to be present in the avian species tree based upon other types of data. Suh et al. (2011) used TE insertions to identify the sister group of Passeriformes. Using fairly broad sampling of Neoaves, they found that the largest number of insertions (seven) supported Node B (Fig. 6), providing highly significant support for the hypothesis that this clade is present in the species tree (based upon the Waddell et al. (2001) test). Suh et al. (2011) also reported two TE insertions that supported Node C (Fig. 6). Although this did not provide significant support for the hypothesis that Node C is present in the species tree, the existence of a microinversion that maps at the base of this group (Braun et al., 2011) combined with the absence of any conflicting TE insertions suggests that clade is likely to be present in the species tree. This suggests that the estimates of avian phylogeny obtained by analysis of concatenated data (i.e., Figs. 1 and 4) and by NJ_{st} excluding BDNF (Fig. 5B), both of which include these clades, are likely to represent better estimates of the avian species tree than the NJ_{st} tree based upon all 50 loci. Taken as a whole, our results emphasize the importance of careful data exploration and the use of multiple analytical methods to identify localized biases that may result in misleading estimates of phylogeny.

4.3. Further improvement in the avian tree of life?

Nearly doubling the size of the data matrix analyzed by adding 31 new loci to the Hackett et al. (2008) data did not substantially increase our confidence in specific relationships relative to Hackett et al. (2008) or identify additional well-supported clades. However, the strength of the phylogenetic signal supporting specific nodes (measured using the bootstrap) varied between the two datasets. For example, the waterbird clade, which has also been independently corroborated by mitogenomic analyses (e.g., Morgan-Richards et al., 2008; Pacheco et al., 2011), was present with <50% bootstrap support in analyses of the 19-gene data matrix but with 70% bootstrap support in analyses of the 31-gene data matrix. Likewise, Eucavitaves (Node F, Fig. 6) was well supported in analyses of the 31-locus dataset but not the 19-locus dataset whereas the opposite was true for the more inclusive Cavitaves (Node G, Fig. 6). Both groups received support in the analyses of the concatenated data (Figs. 1 and 4), probably reflecting the combination of the signal supporting Node F in the 31-locus data matrix and the signal supporting Node G in the 19-locus data matrix. These results further emphasize the heterogeneity of the phylogenetic signal in different parts of the genome. Moreover, both clades were present in the NJ_{st} tree excluding BDNF (Fig. 5B), albeit without bootstrap support. Oliver (2013) suggests that gene tree-species tree discordance might be especially problematic for the position of *Leptosomus*; our NJ_{st} analyses suggest this discordance is unlikely to have an impact upon the analyses of concatenated data that strongly

support placing *Leptosomus* sister to Node F (in agreement with Hackett et al. (2008)).

It is unclear whether the short internodes in the avian tree of life that were not recovered in these analyses can be resolved by using substantially larger datasets (e.g. 10- to 100-fold greater than this study). Although there were substantial improvements to the resolution of the avian tree of life when amount of data used for analyses increased from individual loci (e.g., Fain and Houde, 2004) to 19 loci (Hackett et al., 2008), the increase to 50 loci in this study did not result in much additional improvement. Although McCormack et al. (2013) analyzed more than 1500 loci, their alignment actually had slightly fewer parsimony informative sites than this study (24,703 sites *versus* 30,512 sites), as the UCEs contain many fewer variable sites than the introns used in this study. It may prove to be the case that the expected number of synapomorphic substitutions along the branches in the avian tree of life that remain unresolved is so small that those few mutations that did occur along the branches have been obscured by subsequent homoplasy. Moreover, it seems likely that there was substantial incomplete lineage sorting along these branches given their length (Oliver, 2013). Lineage sorting has been suggested as an explanation for the distribution of TE insertions for these taxa (Suh et al., 2011; Matzke et al., 2012). Indeed, rare genomic changes (RGCs), such as TE insertions or microinversions, may be promising for resolving deep divergences that are subject to lineage sorting, although the probability of synapomorphic RGCs accumulating along short internodes is low and RGCs can be subject to rare instances of homoplasy in addition to lineage sorting (Braun et al., 2011; Han et al., 2011). If the branches at the base of Neoaves are actually as short as they appear, it is possible that deep avian relationships reflect either a soft polytomy that is effectively unresolvable (possibly even with genome-scale data) or even a true hard polytomy (simultaneous speciation events).

5. Conclusions

This study demonstrates that phylogenetic analyses of large datasets can yield robust (and reproducible) results, though a small proportion of well-supported nodes may be driven by the unique characteristics of the specific loci sampled rather than representing the underlying species tree. Since all nodes that were strongly supported in Hackett et al. (2008) were also strongly supported by the independent 31-locus dataset (most also with >95% bootstrap support) it seems likely that those nodes receiving the highest levels of bootstrap support accurately reflect the species tree as long as the method used to analyze the data is unbiased. In contrast, nodes with moderate support in analyses of concatenated data (i.e., >70% but <95%) were occasionally discordant when analyses of different sets of genes are compared, suggesting that they can reflect misleading signal. Two of the strongly supported but discordant nodes evident in comparisons of the independent datasets appeared to reflect the influence of a single locus. The use of a method that estimates the species tree by combining gene trees (NJ_{st}) was largely able to accommodate the incongruent signal associated that locus; however, NJ_{st} performed poorly when a different locus was included. Based upon the ability of the approaches used here to reveal the incongruent signal associated with different loci we propose that conducting analyses of independent datasets and exploring a variety of analytical approaches can be a powerful tool to identify similar cases and eventually develop a robust estimate of the tree of life.

Acknowledgments

This paper was improved by feedback from members of the Kimball-Braun group and an anonymous reviewer. We thank the museums and collectors listed in Supplemental Table S2 for the loan of samples. This research was funded in part by the US National Science Foundation Assembling the Tree of Life Program (Grant DEB-0228682 to RTK, ELB, and David W. Steadman). NW was supported on a fellowship from the China Scholarship Council and a Singer Seed Grant (University of Florida) to RTK and ELB. Padmaksi Tester, Tara Conway, Elon Fernandez, and Kathleen Zelle assisted with data collection. We are grateful to the University of Florida Genetics Institute for providing computational resources and to Tim Shaw and Liang Liu for helpful discussions regarding methods to estimate the species tree using information from gene trees. Photographs in the graphical abstract were downloaded from Wikimedia Commons and were adapted for this work in accordance with the Creative Commons Attribution License. Photographers should be attributed to: (1) Zebra Finch: Keith Gerstung from McHenry, IL, United States, (2) Pacific Gull: JJ Harrison, (3) Virginia Rail: Mike Baird from Morro Bay, USA, (4) Brown Pelican: Terry Foote, and (5) Sparkling Violetear: The Lilac Breasted Roller from Sullivan's Island, United States.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.05.029.

References

Anisimova, M., Cannarozzi, G., Liberles, D.A., 2010. Finding the balance between the mathematical and biological optima in multiple sequence alignment. Trends Evol. Biol., e7.

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., 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.

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.

Churakov, G., Kriegs, J.O., Baertsch, R., Zemann, A., Brosius, J., Schmitz, J., 2009. Mosaic retroposon insertion patterns in placental mammals. Genome Res. 19, 868–875.

Creer, S., 2007. Choosing and using introns in molecular phylogenetics. Evol. Bioinform. 3, 99–108.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, e68.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sørensen, M.V., Haddock, S.H., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–749.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Ericson, P.G.P., Anderson, C.L., Britton, T., Elzanowski, A., Johansson, U.S., Källersjö, 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.

Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Gatesy, J., Matthee, C., DeSalle, R., Hayashi, C., 2002. Resolution of a supertree/supermatrix paradox. Syst. Biol. 41, 652–664.

Griffin, D., Robertson, L., Tempest, H., Skinner, B., 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. Chrom. Res. 15, 29.

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.

Haddrath, O., Baker, A.J., 2012. Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. Proc. R. Soc. B 279, 4617–4625.

Hallström, B.M., Janke, A., 2010. Mammalian evolution may not be strictly bifurcating. Mol. Biol. Evol. 27, 2804–2816.
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.
Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.
Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22, 225–231.
Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462–467.
Katoh, K., Asimenos, G., Toh, H., 2009. Multiple alignment of DNA sequences with MAFFT. Method. Molec. Biol. 537, 39–64.
Katsu, Y., Braun, E.L., Guillette Jr, 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.
Kimball, R.T., Braun, E.L., 2008. A multigene phylogeny of Galliformes supports a single origin of erectile ability in non-feathered facial traits. J. Avian Biol. 39, 438–445.
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., Mary, C.M.S., Braun, E.L., 2011. A macroevolutionary perspective on multiple sexual traits in the Phasianidae (Galliformes). Int. J. Evol. Biol., 423938.
Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and CENSOR. BMC Bioinformatics 7, 474.
Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24.
Lake, J.A., 1991. The order of sequence alignment can bias the selection of tree topology. Mol. Biol. Evol. 8, 378–385.
Ligon, J.D., 1967. Relationships of the cathartid vultures. Occas. Pap. Univ. Mich. Mus. Zool. 651.
Liu, L., Yu, L., 2011. Estimating species trees from unrooted gene trees. Syst. Biol. 60, 661–667.
Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58, 468–477.
Liu, K., Linder, C.R., Warnow, T., 2010. Multiple sequence alignment: a major challenge to large-scale phylogenetics. PLoS Curr.: Tree of Life. http://dx.doi.org/10.1371/currents.RRN1198.
Matzke, A., Churakov, G., Berkes, P., Arms, E.M., Kelsey, D., Brosius, J., Kriegs, J.O., Schmitz, J., 2012. Retroposon insertion patterns of Neoavian birds: strong evidence for an extensive incomplete lineage sorting era. Mol. Biol. Evol. 29, 1497–1501.
Mayr, G., 2011. Metaves, Mirandornithes, Strisores and other novelties – a critical review of the higher-level phylogeny of neornithine birds. J. Zool. Syst. Evol. Res. 49, 58–76.
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 (1), e54848.
Meyer, S., von Haeseler, A., 2003. Identifying site-specific substitution rates. Mol. Biol. Evol. 20, 182–189.
Morgan-Richards, M., Trewick, S.A., Bartosch-Härlid, A., Kardailsky, O., Phillips, M.J., McLenachan, P.A., Penny, D., 2008. Bird evolution: testing the Metaves clade with six new mitochondrial genomes. BMC Evol. Biol. 8, 20.
Nishihara, H., Maruyama, S., Okada, N., 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. Proc. Natl. Acad. Sci. USA 106, 5235–5240.
Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217.
Oliver, J.C., 2013. Microevolutionary processes generate phylogenomic discordance at ancient divergences. Evolution. http://dx.doi.org/10.1111/evo.12047.
Pacheco, M.A., Battistuzzi, F.U., Lentino, M., Aguilar, R.F., Kumar, S., Escalante, A.A., 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. Mol. Biol. Evol. 28, 1927–1942.
Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Queinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D.J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Worheide, G., Manuel, M., 2009. Phylogenomics revives traditional views on deep animal relationships. Curr. Biol. 19, 706–712.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602.
Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21, 1455–1458.
Poe, S., Chubb, A.L., 2004. Birds in a bush: five genes indicate explosive evolution of avian orders. Evolution 58, 404–415.
Pratt, R.C., Gibb, G.C., Morgan-Richards, M., Phillips, M.J., Hendy, M.D., Penny, D., 2009. Toward resolving deep Neoaves phylogeny: data, signal enhancement, and priors. Mol. Biol. Evol. 26, 313–326.
Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53, 131–147.
Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.
Schierwater, B., Eitel, M., Jakob, W., Osigus, H.J., Hadrys, H., Dellaporta, S.L., Kolokotronis, S.O., DeSalle, R., 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "Urmetazoon" hypothesis. PLoS Biol. 7, 36–44.
Shapiro, L.H., Dunbacher, J.P., 2001. Adenylate kinase Intron 5: a new nuclear locus for avian systematics. Auk 118, 248–255.
Shaw, T.I., Ruan, Z., Glenn, T., Liu, L., 2013. Webserver for species tree reconstruction. Nucleic Acids Res. 41, W238–W241.
Shen, X.-X., Liang, D., Zhang, P., 2012. The development of three long universal nuclear protein-coding locus markers and their application to Osteichthyan phylogenetics with nested PCR. PLoS One 7, e39256.
Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds: A Study in Molecular Evolution. Yale University Press, New Haven.
Smith, J.V., Braun, E.L., Kimball, R.T., 2013. Ratite non-monophyly: independent evidence from 40 novel loci. Syst. Biol. 62, 35–49.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
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, n443.
Sul, S.-J., Williams, T.L., 2008. An experimental analysis of Robinson–Foulds distance matrix algorithms. In: European Symposium on Algorithms (ESA'08). Springer-Verlag, pp. 793–804.
Thomson, R.C., Shaffer, H.B., 2010. Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. Syst. Biol. 59, 42–58.
Vinh, L.S., von Haeseler, A., 2004. IQPNNI: moving fast through tree space and stopping in time. Mol. Biol. Evol. 21, 1565–1571.
Waddell, P.J., Kishino, H., Ota, R., 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inform. 12, 141–154.
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.
Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Kunstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., Heger, A., Kong, L., Ponting, C.P., Jarvis, E.D., Mello, C.V., Minx, P., Lovell, P., Velho, T.A.F., Ferris, M., Balakrishnan, C.N., Sinha, S., Blatti, C., London, S.E., Li, Y., Lin, Y.-C., George, J., Sweedler, J., Southey, B., Gunaratne, P., Watson, M., Nam, K., Backstrom, N., Smeds, L., Nabholz, B., Itoh, Y., Whitney, O., Pfenning, A.R., Howard, J., Volker, M., Skinner, B.M., Griffin, D.K., Ye, L., McLaren, W.M., Flicek, P., Quesada, V., Velasco, G., Lopez-Otin, C., Puente, X.S., Olender, T., Lancet, D., Smit, A.F.A., Hubley, R., Konkel, M.K., Walker, J.A., Batzer, M.A., Gu, W., Pollock, D.D., Chen, L., Cheng, Z., Eichler, E.E., Stapley, J., Slate, J., Ekblom, R., Birkhead, T., Burke, T., Burt, D., Scharff, C., Adam, I., Richard, H., Sultan, M., Soldatov, A., Lehrach, H., Edwards, S.V., Yang, S.-P., Li, X., Graves, T., Fulton, L., Nelson, J., Chinwalla, A., Hou, S., Mardis, E.R., Wilson, R.K., 2010. The genome of a songbird. Nature 464, 757–762.
Wildman, D.E., Uddin, M., Opazo, J.C., Liu, G., Lefort, V., Guindon, S., Gascuel, O., Grossman, L.I., Romero, R., Goodman, M., 2007. Genomics, biogeography, and the diversification of placental mammals. Proc. Natl. Acad. Sci. USA 104, 14395–14400.
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.L., 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 2, 419–444.