Analysis of a Rapid Evolutionary Radiation Using Ultraconserved Elements: Evidence for a Bias in Some Multispecies Coalescent Methods

Kelly A. Meiklejohn[1], Brant C. Faircloth[2], Travis C. Glenn[3], Rebecca T. Kimball[1,4], and Edward L. Braun[1,4,*]

[1]*Department of Biology, University of Florida, Gainesville, FL;* [2]*Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA;* [3]*Department of Environmental Health Science, University of Georgia, Athens, GA;* [4]*Genetics Institute, University of Florida, Gainesville, FL*
*Correspondence to be sent to: Department of Biology, University of Florida, Gainesville, Florida; Genetics Institute, University of Florida, Gainesville, FL 32611; Email: ebraun68@ufl.edu

Abstract.—Rapid evolutionary radiations are expected to require large amounts of sequence data to resolve. To resolve these types of relationships many systematists believe that it will be necessary to collect data by next-generation sequencing (NGS) and use multispecies coalescent ("species tree") methods. Ultraconserved element (UCE) sequence capture is becoming a popular method to leverage the high throughput of NGS to address problems in vertebrate phylogenetics. Here we examine the performance of UCE data for gallopheasants (true pheasants and allies), a clade that underwent a rapid radiation 10–15 Ma. Relationships among gallopheasant genera have been difficult to establish. We used this rapid radiation to assess the performance of species tree methods, using ~600 kilobases of DNA sequence data from ~1500 UCEs. We also integrated information from traditional markers (nuclear intron data from 15 loci and three mitochondrial gene regions). Species tree methods exhibited troubling behavior. Two methods [Maximum Pseudolikelihood for Estimating Species Trees (MP-EST) and Accurate Species TRee ALgorithm (ASTRAL)] appeared to perform optimally when the set of input gene trees was limited to the most variable UCEs, though ASTRAL appeared to be more robust than MP-EST to input trees generated using less variable UCEs. In contrast, the rooted triplet consensus method implemented in Triplec performed better when the largest set of input gene trees was used. We also found that all three species tree methods exhibited a surprising degree of dependence on the program used to estimate input gene trees, suggesting that the details of likelihood calculations (e.g., numerical optimization) are important for loci with limited phylogenetic information. As an alternative to summary species tree methods we explored the performance of SuperMatrix Rooted Triple - Maximum Likelihood (SMRT-ML), a concatenation method that is consistent even when gene trees exhibit topological differences due to the multispecies coalescent. We found that SMRT-ML performed well for UCE data. Our results suggest that UCE data have excellent prospects for the resolution of difficult evolutionary radiations, though specific attention may need to be given to the details of the methods used to estimate species trees. [Galliformes, gene tree estimation error; Phasianidae; polytomy; supermatrix rooted triplets; Triplec; total evidence.]

Elucidating well-supported phylogenies has proven to be challenging for many groups (Rokas and Carroll 2006), even when multiple loci are used (e.g., Poe and Chubb 2004). In many cases, these problematic groups reflect rapid radiations where the times between cladogenic events were insufficient for the accumulation of changes uniting groups of interest. Discordance among gene trees is also likely under these conditions (Edwards 2009), further exacerbating the difficulty of resolving rapid radiations. Following a radiation, the limited amounts of informative signal that might have arisen can be obscured by subsequent changes (Whitfield and Lockhart 2007; Philippe et al. 2011; Patel et al. 2013), making it even more difficult to resolve a radiation (or determine if it might reflect a hard polytomy).

Phylogenomic studies, which use large data sets and sample the genome broadly, may allow historical signal associated with short internodes to be recovered (McCormack et al. 2012; Jarvis et al. 2014; Sun et al. 2014). However, simply adding sequence data does not always provide more informative signal for resolving difficult phylogenetic problems (Philippe et al. 2011; Kimball et al. 2013; Salichos and Rokas 2013; Kimball and Braun 2014). Indeed, it has long been appreciated that some phylogenetic estimators are inconsistent in parts of parameter space (i.e., they fail to converge on the correct tree topology as data that evolve under the same model

are added; see Felsenstein 1978; Kim 2000; Roch and Steel 2014). In practical terms, this implies that analyses of very large data sets can sometimes result in strong support for incorrect relationships (Jeffroy et al. 2006; Kubatko and Degnan 2007). Thus, the best analytical methods for phylogenomic data remain unclear.

The earliest phylogenomic studies were limited to organisms with genome sequences generated for other reasons (Rokas et al. 2003; Wolf et al. 2004). Sequence capture and next-generation sequencing (NGS) now allow hundreds or thousands of orthologous loci to be collected for a substantially lower cost than whole genome sequencing (e.g., Faircloth et al. 2012; Lemmon et al. 2012; Hedtke et al. 2013; Li et al. 2013; Peñalba et al. 2014). The best analytical approach for phylogenomic data remains unclear, although it is typically assumed that analyses of large data sets comprising many unlinked loci will require methods that incorporate the multispecies coalescent (Edwards 2009).

Here, we examined the potential for data from ultraconserved element (UCE) sequence capture (Faircloth et al. 2012; McCormack et al. 2012) to resolve a difficult phylogenetic question. UCEs are highly conserved regions (>60 base pairs [bp]; Bejerano et al. 2004) with flanks that exhibit increasing variation as the distance from the conserved core increases (Faircloth et al. 2012). The conserved regions facilitate

sequence capture and the flanking regions provide phylogenetic signal (Faircloth et al. 2012). UCE loci are excellent phylogenomic markers because: 1) they are easy to align (McCormack et al. 2012); 2) they exhibit limited saturation, even for relatively deep divergences (McCormack et al. 2012); and 3) they are widely distributed, typically in single copy, throughout vertebrate nuclear genomes (Bejerano et al. 2004). These benefits have led to the use of UCE data to resolve relationships in many different taxonomic groups at various evolutionary depths (Crawford et al. 2012; McCormack et al. 2012; Faircloth et al. 2013; McCormack et al. 2013; Green et al. 2014; Sun et al. 2014; Crawford et al. 2015; Streicher et al. 2016), including within populations (Smith et al. 2014). Despite these promising results, there have been few direct comparisons between UCEs and "traditional" markers (i.e., mitochondrial DNA [Kocher et al. 1989; Sorenson et al. 1999] and nuclear introns [Kimball et al. 2009]). Thus, comparisons between the results of analyses using UCE data and traditional markers using available analytical methods, especially coalescent (or "species tree") methods, are likely to be informative.

Analyses of UCE data obtained by sequence capture have the potential to present challenges for several reasons. First, it is unclear whether phylogenetic data generated by NGS methods, which are assembled using automated pipelines, are of lower quality than the sequences of traditional markers, which are often subjected to manual review. Second, the low degree of variation within individual UCE loci is expected to result in poorly resolved gene trees. This could cause problems with "summary species tree" methods (e.g., MP-EST; Liu et al. 2010). Indeed, the use of gene trees based on relatively low variation markers has been suggested to be problematic for species tree analyses (Gatesy and Springer 2014; Lanier et al. 2014; Xi et al. 2015) and empirical studies using UCE data have indeed produced poorly resolved species trees (McCormack et al. 2013). To further examine these issues we collected a UCE data set focused on a difficult phylogenetic problem, explored the phylogenetic signal in the UCE data, and performed explicit comparisons to nuclear introns and mitochondrial gene regions ("traditional markers").

The pheasant and allies (Phasianidae) underwent numerous rapid radiations (Kimball et al. 1999; Shen et al. 2010; Kimball et al. 2011; Wang et al. 2013; Kimball and Braun 2014; Stein et al. 2015). The "greater gallopheasant clade" (true pheasants and allies, hereafter called "gallophesants") underwent a radiation 10–15 Ma [based on Jetz et al. (2012) and Stein et al. (2015)]. This group comprises six genera (*Catreus, Crossoptilon, Chrysolophus, Lophura, Phasianus,* and *Syrmaticus*), and gallopheasant monophyly is strongly supported (e.g., Crowe et al. 2006; Kimball and Braun 2008; Bao et al. 2010; Kimball et al. 2011; Wang et al. 2013; Kimball and Braun 2014). However, relationships among five of the six genera remain uncertain, even in studies using multiple unlinked loci (e.g., Nadeau et al. 2007; Kimball and Braun 2008; Bao et al. 2010; Kimball

et al. 2011; Wang et al. 2013; Kimball and Braun 2014). Moreover, Kimball and Braun (2014) noted that even the major point of agreement (the position of *Syrmaticus* sister to all other gallopheasants) reflects a bipartition with a remarkably low concordance factor (*sensu* Baum 2007) when multiple nuclear introns are analyzed. Thus, it seems reasonable to postulate that the gallopheasants might highlight any challenges associated with use of UCEs to resolve rapid radiations.

Methods

Taxon Selection and Molecular Methods

To evaluate the performance of UCEs, introns, and mitochondrial data for phylogenetic analysis of the gallopheasants, we selected 18 species, including 17 members of the family Phasianidae and a single outgroup from Odontophoridae (the family sister to Phasianidae; see Cox et al. 2007). This included at least one species from each gallopheasant genus and other phasianids that represent a variety of taxonomic depths. We created a "traditional" phylogenetic data set comprising the 15 nuclear introns and three mitochondrial gene regions used by Kimball and Braun (2014) and then we collected UCEs from the same individuals, with the exception of *Perdix perdix*, *Gallus gallus*, and *Meleagris gallopavo*. We had poor recovery of UCEs from *Perdix perdix* so we substituted *Perdix dauurica* UCE data; *Perdix* species are closely related and monophyly of the genus is uncontroversial (Bao et al. 2010). Finally, we extracted UCE data from the published *Gallus gallus* (chicken) and *Meleagris gallopavo* (turkey) genome sequences (Hillier et al. 2004; Dalloul et al. 2010).

We used the Faircloth et al. (2012) protocol with several modifications to obtain UCE data. Briefly, we prepared Nextera sequencing libraries using the manufacturer's protocol (Illumina Inc., San Diego, CA), except that we used primers with custom index sequences (Faircloth and Glenn 2012), pooled libraries into groups of 8 taxa, and enriched each library pool using 5472 probes (Mycroarray Inc., Ann Arbor, MI) that target 5060 UCE loci. We then amplified the enriched pools using limited-cycle PCR (18 cycles), quantified the resulting pools by qPCR (Kapa Biosystems Inc., Wilmington, MA), and collected a single lane of Illumina HiSeq 2000 (Illumina Inc., San Diego CA) PE75 data at the UC Irvine Genomics High-Throughput Facility. After collecting the data, we generated a *de novo* assembly using Velvet (Zerbino and Birney 2008) before matching the contigs to defined UCE loci using PHYLUCE v1.1 (Faircloth et al. 2012; Faircloth 2014).

Sequence Alignment and Data Quality Assessment

We used the traditional marker alignments from Kimball and Braun (2014), extracting the 18 taxa selected for this study. For UCEs, we filtered the data to produce a "complete" (i.e., no missing loci for any taxon) matrix of 1479 UCE loci, and generated UCE alignments using

the Faircloth et al. (2012) pipeline. We used GENEIOUS (Version 6.1.2; Biomatters, 2013) to examine all UCE alignments manually, allowing us to identify indels. We examined a subset of 10 loci by PCR amplification and Sanger sequencing to determine whether any of the identified indels were assembly artifacts.

Concatenated Phylogenetic Analyses

We conducted analyses of concatenated data to evaluate the phylogenetic signal in each of the three data types (UCEs, nuclear introns, and mitochondrial regions) individually and in two combinations: 1) traditional markers (introns and mitochondrial gene regions) and 2) all three data types combined. We performed most of our analyses using the fast maximum likelihood (ML) program RAxML (version 8.0.20; Stamatakis 2006; Stamatakis 2014), but we also conducted some additional analyses in PhyML (version 3.0; Guindon et al. 2010) and PAUP* (version 4.0b10; Swofford 2003) for comparison (PAUP* was used for maximum parsimony [MP] analyses).

We identified the best partitioning scheme for RAxML using PartitionFinder (Lanfear et al. 2012), using the greedy search for the smaller traditional marker data sets and hcluster (Lanfear et al. 2014) for the larger UCE data set (with equal weighting). We used the Bayesian information criterion (BIC; Schwarz 1978) to identify the best partitioning scheme. We used 10 replicate searches to identify the optimal tree and assessed support using non-parametric bootstraps (with 500 replicates). We also used the Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-like aLRT; Anisimova et al. 2011) implemented in PhyML to assess support; that analysis used the GTR+I+Γ model of evolution.

Coalescent-Based Analyses and Congruence among Gene Trees

Estimates of gene trees.—We generated 500 ML bootstrap gene trees for each locus using RAxML, PhyML, and GARLI (version 2.0; Zwickl 2006). We also generated gene trees using MrBayes (version 3.1.2; Huelsenbeck and Ronquist 2001), running each Markov chain Monte Carlo (MCMC) analysis for 10,000,000 generations and sampling the posterior from two sets of chains every 1000 generations (each set of chains used the default parameters; i.e., four chains, three of which were heated). We discarded the first 30% of the MrBayes trees as burn-in, a conservative value based on examining the results in TRACER (Rambaut et al. 2014). Finally, we generated optimal gene trees for each locus using GARLI because it can produce trees with polytomies, which may be desirable for loci that exhibit limited variation. RAxML analyses used the GTR+Γ model, those conducted in PhyML and MrBAYES used the best-fitting model based on MrAIC (Nylander 2004) for each locus, and GARLI used the best-fitting model based on MODELTEST version 3.7 (Posada and Crandall 1998); PAUP* 4.0b10 (Swofford 2003) was used to estimate likelihoods for

MODELTEST. We used MrAIC to identify the best-fitting model for MrBAYES because the set of models examined by MrAIC corresponds to the models implemented in MrBAYES. We used the MrAIC to identify the best-fitting model for PhyML because MrAIC uses PhyML for likelihood calculations. Finally, we used MODELTEST to identify the best-fitting model for GARLI analyses because it is straightforward to implement the broader set of models examined by MODELTEST in GARLI. For both MrAIC and MODELTEST we used the second order corrected Akaike information criterion (AIC$_c$; Hurvich and Tsai 1989), with the number of sites in the alignment used as the sample size, to identify the best-fitting model for each locus.

Coalescent-based analyses.—To assess the utility of UCEs for species tree estimation from gene trees, we input the bootstrap ML (RAxML, GARLI, PhyML) or Bayesian (MrBayes) gene trees to MP-EST (Liu et al. 2010) and ASTRAL (Mirarab et al. 2014c). MP-EST uses rooted gene trees as input and it generates an estimate of the species tree with branch lengths in coalescent units, whereas ASTRAL uses unrooted input gene trees and it focuses only on the topology of the species tree. We used six different sets of UCE loci: those with 1) $\geqslant 25$ parsimony-informative sites (N = 54); 2) $\geqslant 20$ parsimony-informative sites (N = 93); 3) $\geqslant 15$ parsimony-informative sites (N = 188); 4) $\geqslant 10$ parsimony-informative sites (N = 366); 5) $\geqslant 5$ parsimony-informative sites (N = 747); and 6) all UCEs with at least one parsimony-informative site (N = 1289). As MrBayes produced a large number of trees, we only used one out of every 25 trees (a total of 560 trees) as input for analyses. We used the fully resolved trees produced by the ML bootstrap and Bayesian analyses as input trees for the summary species tree methods MP-EST and ASTRAL. We conducted the MP-EST and ASTRAL analyses for all sets of input gene trees (500 bootstrap trees or 560 Bayesian trees) and then produced a majority rule extended (MRE) consensus (="greedy" consensus) tree, which we used as our estimate of the species tree (with associated support values).

Individual UCEs exhibit limited variation (see below for details), sometimes having a single parsimony-informative site for our taxon sample. Such a UCE only provides information about a single bipartition; any other resolution of the gene tree for that locus would be random. To address this issue, we used Triplec (Ewing et al. 2008), a rooted triplet consensus method that permits the use of gene trees with polytomies and is consistent when discordance among gene trees reflects the multispecies coalescent. Triplec uses quartet puzzling (Strimmer and von Haeseler 1996) to combine rooted triplets and therefore produces quartet puzzling values for each branch (which are not directly comparable to bootstrap values).

We used three different approaches to obtain estimates of gene trees with polytomies for input in Triplec. First, we generated majority rule (MR) bootstrap consensus trees using *consense* from Phylip (Felsenstein 2013) for each ML bootstrap analyses (RAxML, GARLI,

and PhyML; see above). Second, we generated a MR consensus trees from the MrBayes analysis (using all 14,000 trees sampled after burn-in). We feel that collapsing branches with <50% bootstrap support (or a posterior probability <0.5) is reasonable given Berry and Gascuel (1996) and Holder et al. (2008). Finally, we generated optimal trees in GARLI but collapsed very short branches to generate polytomies. All of these strategies should eliminate bipartitions in gene trees without any support, although they are also likely to collapse some correct branches.

To complement the summary species tree methods, we also conducted a SMRT-ML analysis (DeGiorgio and Degnan 2010). Although SMRT-ML is a concatenation approach, it is consistent when gene tree topologies differ due to incomplete lineage sorting (DeGiorgio and Degnan 2010). Briefly, SMRT-ML decomposes a concatenated data matrix into all possible sets of three taxa, obtains the rooted ML tree for each possible rooted triplet, and combines the rooted three taxon trees using a supertree method. We used RAxML without partitioning to conduct the ML analysis for each rooted triplet and placed the root using the outgroup (*Cyrtonyx montezumae*). Then we combined the rooted triplets using matrix representation with parsimony (MRP; Baum 1992; Ragan 1992); we used *clann* (Creevey and McInerney 2005) for MRP coding and PAUP 4.0a136 to identify the MRP tree. This analysis was conducted using a Perl script modified from the program used by Sun et al. (2014); the script is available in the Dryad data submission for this manuscript (http://dx.doi.org/10.5061/dryad.p1m52). We estimated the bootstrap support for clades in the SMRT-ML tree by generating 500 bootstrapped data matrices from the original concatenated data matrix and then running the SMRT-ML script on each bootstrapped data set.

Congruence among estimates of gene trees.—To assess congruence among estimates of gene trees, we generated a MRE consensus (="greedy" consensus) of the bootstrap gene trees using the Phylip *consense* program. Because we wanted to focus on the branches in each gene tree with at least some support, we generated the MRE consensus using the same MR consensus gene trees used as Triplec input (see above). We calculated the number of gene trees supporting or conflicting with each clade directly from the MRE consensus tree or by extracting data from the bipartition table (the "." and "*" notation) generated by *consense*.

RESULTS

Quality of UCE Sequence Data

We generated a complete data matrix of 1479 UCE loci (assembly statistics presented in Supplementary File S1); the length of the aligned UCE loci ranged from 193 bp to 774 bp (median = 400 bp) for a total of 601,191 bp. Manual examination of the alignments revealed two types of potential assembly errors. First, 69 alignments (~5% of UCEs) had at least one sequence with a high degree of mismatch at the 5′ or 3′ end in otherwise relatively conserved regions (e.g., Supplementary Fig. S1). These anomalies probably reflect assembly artifacts so we trimmed nucleotides from the end of the relevant sequence until at least five nucleotides were conserved across the alignment. Second, we identified multiresidue indels in 114 UCE loci. Approximately 75% of these indels were >20 bp in length and many of the longest indels were present in a single taxon. We did not detect any of the five large (>50 bp in length) autapomorphic insertions in Sanger-sequenced amplicons, suggesting that they were assembly artifacts. These errors could result in a minor branch length distortion so they were excluded from analyses. This editing reduced the length of the final edited alignment to 599,627 bp. We initially analyzed both the unedited and edited alignments, and recovered the same topology with similar support (Fig. 1 and see below). For all subsequent analyses, we used the edited alignment (with the terminal mismatches and long autapomorphic insertions removed) because it is likely to best reflect the underlying genomic sequences.

In contrast to the long autapomorphic insertions, the five short insertions that we tested by PCR amplification and Sanger sequencing indicated that these insertions were not assembly artifacts. There were 43 short, multispecies indels in our alignments and 25 matched the likely species tree, whereas 18 showed discordance. Of the discordant indels, 11 involved repetitive regions (microsatellites and homopolymers), so homoplasy was not surprising. The remaining seven discordant indels all conflicted with clades defined by short internodes and therefore likely represent hemiplasy (Avise and Robinson 2008) rather than true homoplasy. Thus, with the exception of those indels involving repetitive regions, indels in UCEs likely define bipartitions in gene trees.

Patterns of Molecular Evolution for Different Data Types

The UCEs contributed >97% of the 599,627 sites in the final edited data matrix (Table 1). Unsurprisingly, the UCE loci had the lowest proportion of variable sites, but still contributed the majority of variable sites to the alignment due to the large number of UCE loci included (Table 1). The proportion of variable sites that were parsimony-informative was highest for the mitochondrial regions, whereas introns and UCEs had similar proportions. The mitochondrial sequences had a lower consistency index (CI) than the introns (Table 1), consistent with many other studies (e.g., Wang et al. 2013), whereas UCEs had a slightly higher CI than the introns (Table 1).

Individual UCE loci exhibited limited variation per locus (Supplementary Fig. S2). Thus, it is likely to be difficult to establish the best-fitting model of evolution for many UCE loci; the 95% credible set of models (for the set of 56 models examined by MODELTEST) was quite large for most loci, ranging from 1 to 42, with a median of 16 (Supplementary File S2). Likewise, there
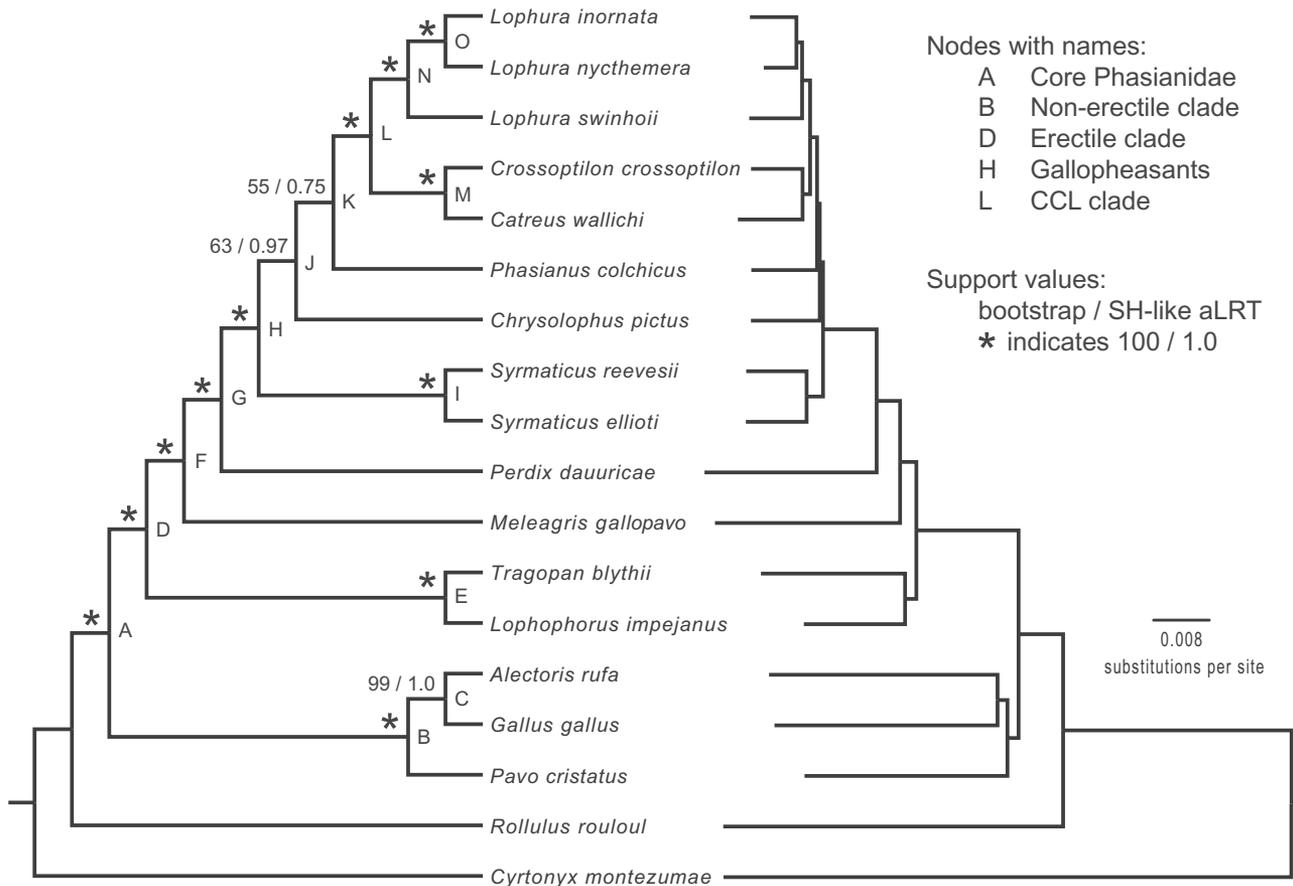
FIGURE 1.   Estimate of phylogeny for gallopheasants and outgroups generated by ML using the UCE data matrix. Identical results were obtained using an unpartitioned analysis with the GTR+Γ model in RAxML (used to generate the phylogram) and the GTR+I+Γ model in PhyML. Support values above branches reflect RAxML bootstrap percentages (from 500 replicates) and support values below branches are from the SH-like aLRT from PhyML. Full support (100% bootstrap support or 1.0 aLRT values) is indicated using an asterisk (*).

TABLE 1.   Patterns of variation for the three data types

Partition	No. of sites	No. of variable sites	No. of inf sites	inf variable sites%	CI
UCEs	599,627	39,848 (6.65%)	10,143 (1.69%)	25.5	0.8331
introns	11,296	3790 (33.6%)	1134 (10%)	29.9	0.8037
Mt	3236	1264 (39.1%)	874 (27%)	69.2	0.4794

Percentages listed after the numbers of variable and parsimony-informative (inf) sites are relative to the number of total sites in each partition; "% inf variable sites" refers to the percentage of variable sites that are parsimony informative. The consistency index (CI) was calculated using the ML tree for the concatenated intron, Mt, and UCE data matrix.

were often differences between the best model obtained by MODELTEST and MrAIC (Supplementary File S2). Nonetheless, there were patterns in the model selection analysis. The best-fitting model most often identified by MODELTEST was HKY+I (25% of UCEs); that model was also included in the 95% credible set for 93% of UCE loci (Supplementary File S2). Thus, allowing different transition and transversion rates appears to be important for analyses of UCEs [Huelsenbeck et al. 2004 reported similar results for other markers], as is including a

proportion of invariant sites to model among-site rate heterogeneity.

The introns also had fairly large credible sets of models (range = 5 to 23; median = 11), and the number of parameters in the best-fitting model for the introns (median = 39) was similar to that for the UCEs (median = 38). No single model was found in the credible sets of all loci, though HKY+Γ, TIM+Γ, and TVM+Γ were in the credible sets for most introns (each was found in 13 of 15 loci; Supplementary File S3). The credible sets for the mitochondrial partitions (which corresponded to the 12S ribosomal RNA and each of the three codon positions in the CYB [cytochrome *b*] and ND2 coding regions) included fewer models (range = 1 to 17; median = 6), and those models were more complex (all credible sets included GTR+I+Γ). However, some model selection uncertainty for the mitochondrial data remained (Supplementary File S3).

The Performance of Phylogenetic Estimation Using Concatenation

Partitioned ML analyses of the UCE data and the combined data matrix resulted in a tree (Fig. 2a)
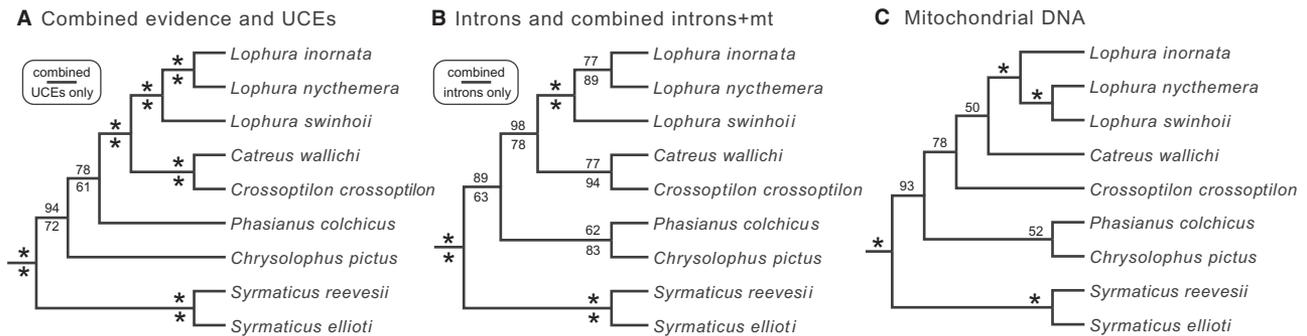
FIGURE 2.   Partitioned ML estimates of phylogeny for the gallopheasant clade obtained using various concatenated data matrices. Topologies for the gallopheasant ingroups are shown for the combined data matrix (introns+mt+edited UCEs) and the edited UCE data matrix (*A*), for the traditional markers (introns+mt) and introns alone (*B*), and for the mitochondrial data (*C*). Bootstrap support is reported as percentages of 500 replicates in RAxML. Full support (100% bootstrap support) is indicated using an asterisk (*).

absolutely congruent with the ML tree for the unpartitioned UCE data alone (Fig. 1). Similar results were found when the MP criterion was used (Supplementary Fig. S3). ML trees estimated using introns (Fig. 2b) and mitochondrial data (Fig. 2c) exhibited some incongruence with the UCE/combined evidence topology, in some cases with moderate support (bootstrap ⩾70%). The topology for the combined intron and mitochondrial data ("traditional markers") was identical to the tree for introns alone (Fig. 2b), although there was a reasonably large change in bootstrap support relative to the introns alone.

All analyses placed *Syrmaticus* sister to the remaining gallopheasants and recovered a clade comprising *Catreus*, *Crossoptilon*, and *Lophura* (node L in Fig. 1; hereafter the CCL clade). All analyses that included the UCE data placed *Chrysolophus* sister to a clade comprising *Phasianus* and the CCL clade (Fig. 2a), whereas analyses based on traditional markers supported a *Chrysolophus–Phasianus* clade (Fig. 2b and c), with that clade sister to the CCL clade. This is a rearrangement of the shortest internode within the gallopheasants (Fig. 1; node K). The mitochondrial data supported two other rearrangements within the CCL clade (Fig. 2c). Outside of gallopheasants, analyses of introns were identical to the UCE topology (Fig. 1) though analyses of the mitochondrial data supported a single rearrangement (node B in Fig. 1). There were several other differences between the mitochondrial and nuclear trees, either reflecting discordance among gene trees or difficulties associated with analyses of mitochondrial data (cf. Braun and Kimball 2002; Meiklejohn et al. 2014).

Estimates of UCE Gene Trees

The majority of UCEs contained fewer parsimony-informative sites than the number of internal branches in the tree given our taxon sample (1101 of the 1289 UCEs with at least one parsimony-informative site have <15 parsimony-informative sites), so estimates of individual gene trees for UCE loci were poorly resolved (Supplementary Fig. S4). The low rate of UCE evolution

means that synapomorphies that define branches in gene trees, particularly short branches, will often be absent (see Braun and Kimball 2001; Slowinski 2001), resulting in poor quality gene trees.

It is very difficult to differentiate between error in estimates of gene trees (i.e., cases where the estimate of the gene tree differs from the true gene tree for any reason) and genuine discordance among true gene trees due to incomplete lineage sorting (ILS) (or other factors; cf. Maddison 1997). However, most species trees are expected to include at least some relatively long branches with limited ILS, and these branches can provide a way to assess error rates for gene trees estimated using UCEs. One such very low ILS clade is the "erectile clade" (Kimball and Braun 2008) (node D in Fig. 1). We recovered this clade in all gene trees based on traditional markers (Fig. 3a), and it has emerged in many previous molecular studies (reviewed by Wang et al. 2013). Moreover, our MP analysis (Supplementary Fig. S3) revealed 1376 unambiguous synapomorphies for the erectile clade. Despite strong evidence that the erectile clade should be present in the majority of true gene trees, we found that it was actually present in fewer than half of the majority rule bootstrap trees for individual UCEs (Fig. 3a and Supplementary File S4). Surprisingly, a large number of UCE gene trees included bipartitions with at least 50% support that conflicted with the erectile clade (Fig. 3b–f) and the proportion of UCEs contradicting the erectile clade increased as the number of parsimony-informative sites per UCE decreased (Fig. 3b–f).

Estimates of the Species Tree and Individual Gene Trees

Several summary species tree methods (e.g., Liu et al. 2010) are known to be consistent estimators of the species tree when used with true gene trees if discordance among gene trees reflects deep coalescence (i.e., ILS resolved in a manner incongruent with the species tree). However, we found evidence for a relatively high error rate in gene tree estimates (Fig. 3b–f; see above) so we are not using true gene trees. Indeed, the error rate we observed makes it unclear whether we could test
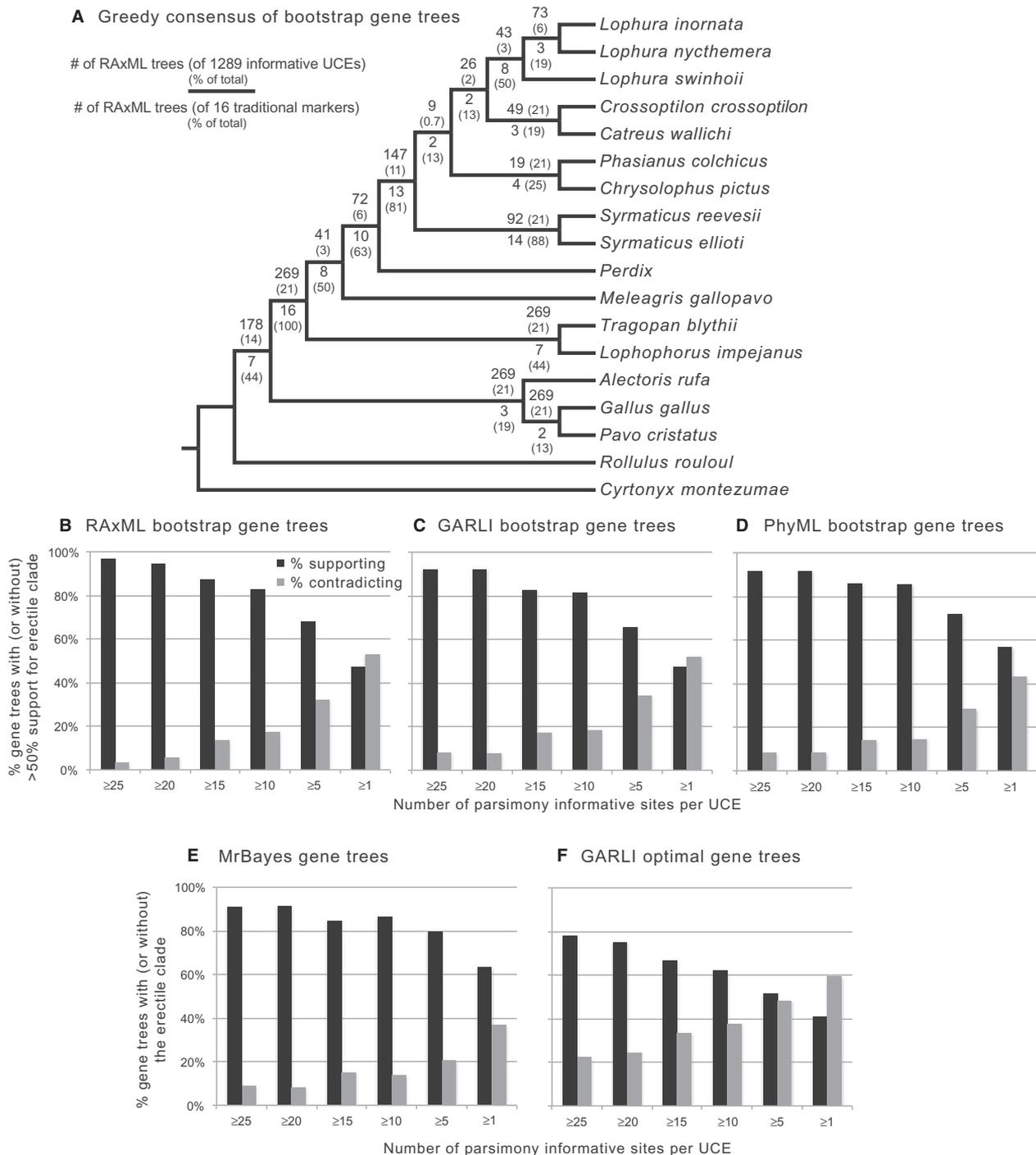
FIGURE 3.   Similarities among estimates of gene trees obtained using different programs and UCEs. The greedy (MRE) consensus of gene trees generated after collapsing nodes with <50% support to form polytomies (A). MRE consensus trees for all sets of gene trees are provided in Supplementary File S4. The number of trees with resolved branches that agree or disagree with the erectile clade (the best supported clade in the tree) is shown as histograms; results are shown for RAxML (B), GARLI (C), PhyML (D), MrBayes (E), and the optimal trees obtained using GARLI (F).

the fit of our data to the multispecies coalescent (e.g., using the frequency test described by Zwickl et al. 2014). This prompted us to explore the impact of restricting species tree analyses to UCEs with different numbers of parsimony-informative sites in a systematic manner, because loci with more parsimony-informative sites

appear to provide more accurate estimates of the gene trees (e.g., Fig. 3b–f).

Analyses of UCEs using MP-EST and ASTRAL.—Estimates of species trees obtained using both MP-EST and
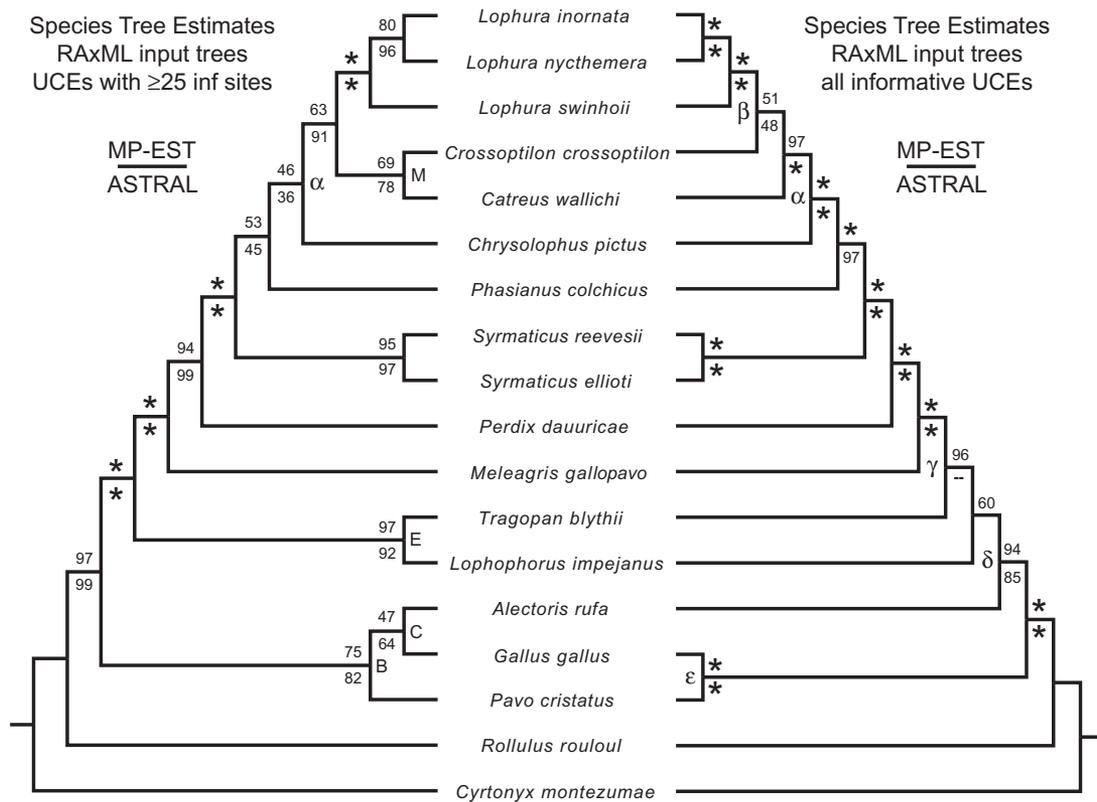
FIGURE 4.   MP-EST-RAxML and ASTRAL-RAxML estimates of the species tree. The species tree estimates for the smallest (left) and largest (right) sets of gene trees (the 54 UCEs with ⩾25 parsimony-informative sites and all 1289 informative UCEs, respectively) are shown. Nodes of interest that differ from those in the trees estimated using concatenated data are indicated with Greek letters. Bootstrap support is reported as percentages of 500 replicates in RAxML (MP-EST above branches and ASTRAL below). Full support (100% bootstrap support) is indicated using an asterisk (*). Failure to recover a branch is indicated using "–" (ASTRAL analysis of all informative UCEs recovered node E instead of node γ).

ASTRAL were largely congruent with concatenated analyses when restricted to gene trees estimated using "highly informative UCEs" that have a large number of parsimony-informative sites (Fig. 4 and Supplementary Files S5 and S6). The topologies of species tree estimates obtained using highly informative UCE were also stable with respect to the program used to estimate gene trees (Supplementary Files S5 and S6). Other than the lower support for species tree analyses, the only difference between estimates of the species tree generated using the highly informative UCEs and the concatenated analyses was the positions of *Phasianus* and *Chrysolophus*, which were swapped relative to their positions in the concatenated tree (Fig. 4, left, and Supplementary Files S5 and S6). This rearrangement involves a branch with limited bootstrap support in analyses of the UCE data (see Fig. 3), although the MP analysis revealed 43 unambiguous synapomorphies uniting *Phasianus* with the CCL clade (Supplementary Fig. S3).

MP-EST analyses using all informative loci recovered several unexpected relationships when we input gene trees generated by several different programs (Fig. 4, right, and Supplementary File S5), including *Lophophorus+Tragopan* (node E vs. γ, Fig. 4) and *Catreus+Crossoptilon* (node M vs. β, Fig. 4). Nodes E and

M were present in the concatenated UCE tree (Fig. 1) and the clades defined by those nodes were supported by 180 and 111 unambiguous synapomorphies, respectively, in our MP analyses (Supplementary Fig. S3). Nodes E and M were also present in MP-EST trees based on highly informative UCEs (Fig. 4, left and Supplementary File S5) and in other studies that have used species tree methods as well as concatenation (Wang et al. 2013; Kimball and Braun 2014). In contrast to the MP-EST-RAxML analysis with all informative loci, the ASTRAL-RAxML analysis with all loci did support node E, albeit with only 61% bootstrap support (Supplementary File S6). We also observed a shift in the position of *Alectoris* (node C, within the non-erectile clade) to one sister to a *Gallus+Pavo* clade (node ε) and finally to one sister to the erectile clade (Fig. 4, node δ) as input trees based on less informative UCE loci were added (Table 2).

Rooted triplet consensus.—Although many available summary species tree methods require fully resolved gene trees, Triplec accepts partially resolved gene trees, which may be advantageous with low-quality gene trees. The pattern we observed with Triplec appeared to be the opposite of that for MP-EST and

TABLE 2. Support for the non-erectile clade (node B) in species tree analyses

UCE set	MP-EST				ASTRAL			
	RAxML	GARLI	PhyML	MrBayes	RAxML	GARLI	PhyML	MrBayes
⩾25	75/—*	79/—*	79/—*	86/—*	89/—*	90/—*	90/—*	94/—*
⩾20	71/—*	73/—*	78/—*	71/—*	82/—*	85/—*	85/—*	88/—*
⩾15	62/—	64/—	75/—	80/—*	88/—*	89/—*	94/—*	96/—*
⩾10	—/64	—/62	72/—	85/—	81/—	83/—	94/—	98/—
⩾5	—/91	—/96	67/—	72/—	61/—	61/—	96/—	99/—
⩾1	—/94	—/100	—/69	—/78	—/85	—/87	64/—	85/—

This table highlights the support for the non-erectile clade (and the position of *Alectoris*) in MP-EST and ASTRAL analyses of the indicated sets of UCE loci (numbers indicate the minimum number of parsimony-informative sites for the UCE loci). A support value on the left indicates node B (the non-erectile clade) was present; a support value on the right indicates that node δ (*Alectoris* + the erectile clade) was present. Cells with node B are shaded, and they are broken into those with an *Alectoris* + *Gallus* clade (node C, indicated with an asterisk) and those with a *Pavo* + *Gallus* clade (node ε).

ASTRAL: estimates of the species tree based on a small number of highly informative UCEs exhibited incongruence with the concatenated tree (for the gallopheasants) and estimates of the species tree based on the largest number of gene trees were more congruent (Supplementary File S7). Like the MP-EST and ASTRAL trees based on the highly informative UCEs (e.g., Fig. 4, left), the Triplec trees with all UCEs placed *Phasianus* sister to a clade comprising *Chrysolophus* and the CCL clade (i.e., node α in Fig. 4). Triplec support values for node α ranged from 75% (Triplec-MrBayes) to 100% (Triplec-GARLI), though using the optimal GARLI trees with short branches collapsed to polytomies reduced the support for this node to 50% (Supplementary File S7).

In contrast to the relatively unstable topology within gallopheasants, the topology outside of gallophesants was more robust and generally well supported regardless of the input trees we used in Triplec. The only point of incongruence outside the gallopheasants was the recovery of the *Gallus-Pavo* clade (i.e., node ε in Fig. 4) instead of the *Gallus-Alectoris* clade (node C in Fig. 1) when large numbers of UCEs were analyzed.

Supermatrix rooted triplet analyses.—Although standard ML analyses of concatenated data are positively misleading in specific parts of species tree space (Roch and Steel 2014), separate concatenated analyses of rooted triplets are consistent (DeGiorgio and Degnan 2010). Because the SMRT-ML method estimates topologies for all possible rooted triplets from a concatenated data matrix, SMRT-ML might be especially useful for UCE data (e.g., Sun et al. 2014).

The estimate of tree topology for gallopheasants in SMRT-ML analyses depended on the model used to estimate the rooted triplet topologies (Fig. 6). The SMRT-ML topology based on the GTR+Γ model was identical to that from standard concatenated analyses of UCEs (i.e., Figs. 1 and 2a) whereas the topology based on the GTR+I+Γ model was identical to the MP-EST and ASTRAL tree estimated from the highly informative UCE loci (e.g., Fig. 4) and the Triplec analyses using the complete set of loci (see Supplementary File S7).
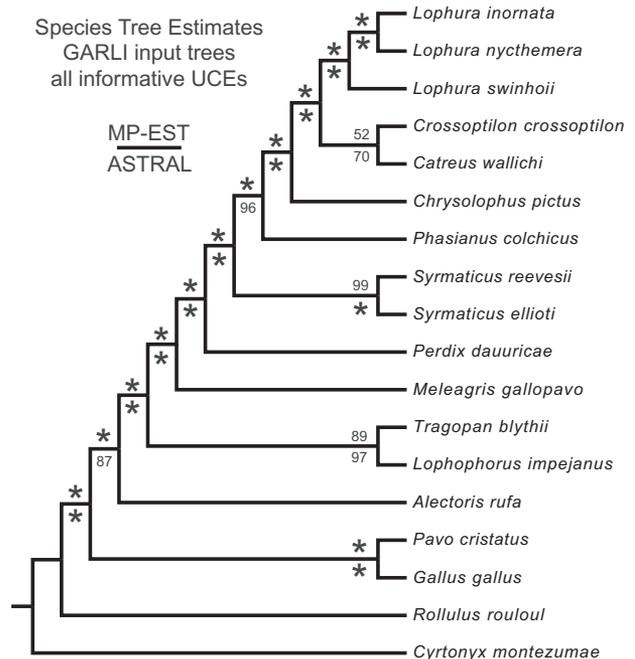

FIGURE 5. MP-EST-GARLI and ASTRAL-GARLI estimate of the gallopheasant species tree with outgroups. This estimate of the species tree used the largest set of gene trees (all 1289 informative UCEs). Full support (100% bootstrap support) is indicated using an asterisk (*).

Bootstrap support for each of these alternatives was limited (Fig. 6), possibly reflecting the conservative nature of SMRT-ML (DeGiorgio and Degnan 2010; also see Sun et al. 2014).

Outside of the gallopheasants the SMRT-ML topologies exhibited a single rearrangement relative to standard analysis of concatenated data (Fig. 1): they recovered the *Gallus-Pavo* clade (i.e., node ε in Fig. 4) with ⩾95% bootstrap support (Supplementary File S8). The SMRT-ML trees also showed strong (99%) bootstrap support for the non-erectile clade (node B in Fig. 1 and Fig. 4), unlike the highly rearranged MP-EST trees that were based on all informative UCE loci (e.g., Fig. 4, right).
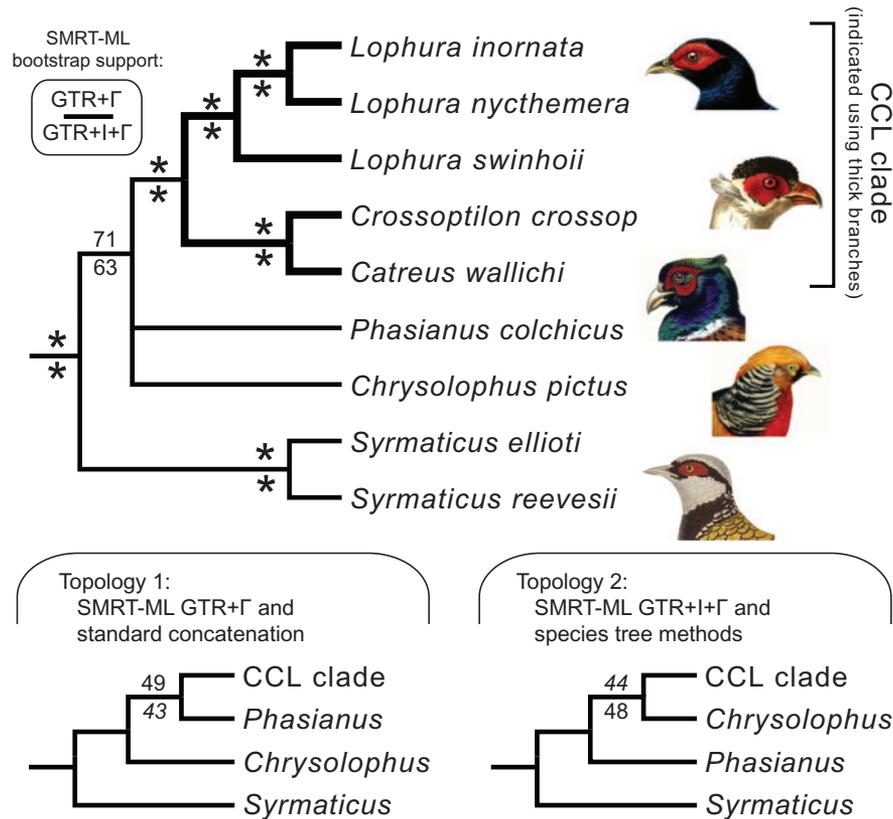
FIGURE 6. SMRT-ML estimate of the gallopheasant species tree based on all UCE loci. A strict consensus of the topologies obtained using the GTR+Γ and GTR+I+Γ models. Bootstrap support is reported as percentages of 500 replicates and 100% bootstrap support is indicated using an asterisk (*). Two simplified topologies at the bottom of the figure are used to indicate the conflicting positions of *Chrysolophus* and *Phasianus* in the GTR+Γ (topology 1, left) and GTR+I+Γ (topology 2, right) trees. The number of bootstrap replicates supporting each topology is shown as presented above; italics indicate that the majority of bootstrap replicates using the relevant model did not recover the topology in question. Complete SMRT-ML bootstrap consensus trees are presented in Supplementary File S8. Images of taxa used for this figure were modified by E.L. Braun from the sources listed in Supplementary File S9.

DISCUSSION

It is now straightforward and relatively inexpensive to collect large amounts of sequence data for phylogenetic analyses using sequence capture with UCE probes (Faircloth et al. 2012), and this approach is becoming increasingly popular for phylogenetic studies (e.g., Crawford et al. 2012; McCormack et al. 2012; Faircloth et al. 2013; McCormack et al. 2013; Green et al. 2014; Sun et al. 2014; Crawford et al. 2015; Streicher et al. 2016). Our results indicate that UCEs exhibit slightly less homoplasy than nuclear introns and showed that both UCE loci and introns exhibit less homoplasy than mitochondrial sequences. This is promising, because introns are known to provide excellent phylogenetic signal in birds (e.g., Chojnowski et al. 2008; Hackett et al. 2008) and other vertebrates (e.g., Fujita et al. 2004; Matthee et al. 2007). However, UCE data is much easier to collect than intronic data, underlining the value of UCE data.

The size of the UCE data matrix and the limited homoplasy of those markers suggests that they should provide strong support for relationships, but we found that our analyses were unable to resolve all relationships

for the gallopheasants with full (100%) bootstrap support. Perhaps surprisingly, analyses of the UCE data contradicted a relationship with moderate (83%) bootstrap support in the combined analysis of a 14.5 kb combined intron and mitochondrial data matrix (with 5055 variable and 2008 parsimony-informative sites). This suggests that nodes with moderate support in analyses of small to moderately sized data sets (e.g., Fig. 2b) should be approached with caution.

Estimates of the Species Tree and Individual Gene Trees

Standard analyses of concatenated data can result in misleading estimates of the species tree (Kubatko and Degnan 2007; Roch and Steel 2014; Warnow 2015). However, simulations (reviewed by Gatesy and Springer 2014) and detailed consideration of the underlying theory (reviewed by Warnow 2015) both reveal a more complex situation. Simulations show that multispecies coalescent methods ("species tree methods") do outperform analyses of concatenated data in some parts of parameter space (e.g., Mirarab et al. 2014a; Mirarab et al. 2014b; Liu et al. 2015; Edwards et al. 2016), but they also show that concatenated analyses

provide a better estimate of the species tree in other parts of parameter space (e.g., Bayzid and Warnow 2013; Patel et al. 2013; Mirarab et al. 2014b). Still other analyses suggest that the performance of both approaches is statistically indistinguishable in many parts of parameter space (Tonini et al. 2015). Finally, we note that the statistical guarantees for summary species tree methods do not hold when gene trees have error (Roch and Warnow 2015; Warnow 2015). Thus, empirical evaluation of species tree methods in various parts of parameter space remains important.

We explored the quality of our gene trees by examining if the individual gene trees include a single, specific node (the erectile clade) with at least 50% bootstrap support or whether the gene trees have at least 50% bootstrap support for a bipartition that conflicts with that node. Although this is only a single measure of gene tree quality, our results suggest many gene trees conflicted with the expected erectile clade. Estimated gene trees may differ from the true gene trees of orthologous loci for two reasons. First, there could be misleading phylogenetic signal (e.g., base compositional convergence [Jeffroy et al. 2006; Katsu et al. 2009] or long-branch attraction [Felsenstein 1978]). Second, the number of sites sampled might be insufficient to provide an accurate estimate of the gene tree (Braun and Kimball 2001; Chojnowski et al. 2008). The second possibility is clearly important for our UCE data because only 52% of the RAxML UCE gene trees either supported or contradicted the erectile clade with ⩾50% bootstrap support. Additionally, our observations that loci with few parsimony-informative sites were much more likely to exhibit conflict than those with many parsimony-informative sites and that there were differences among programs in the number of trees that contradicted the erectile clade (Fig. 3b–f) provide evidence that error in phylogenetic estimation due to insufficient phylogenetic information is a major explanation for the observed incongruence among gene trees.

Three general patterns emerged in the more commonly used summary species tree analyses (MP-EST and ASTRAL). First, using gene trees estimated using less informative UCEs as input for the species tree method appeared to cause a shift toward more asymmetric (pectinate) species tree topologies (Fig. 4 and Supplementary Files S5 and S6); this could reflect known biases in the expected shape of gene trees when branches in the species tree are short (cf. Harshman et al. 2008; Smith et al. 2013; Rosenberg 2013). Second, the estimate of the species tree exhibited a surprising degree of dependence on the specific program used to estimate the input gene trees. Finally, we observed different results when MP-EST and ASTRAL were compared. Like Simmons and Gatesy (2015), we found that ASTRAL appeared to be more robust than MP-EST (Table 2; also compare Supplementary Files S5 and S6). This could reflect the fact that MP-EST obtains joint estimates of branch lengths (in coalescent units) as part of the tree search, whereas ASTRAL focuses only on topology. Errors in the input gene tree will lead to the

underestimation of branch lengths in the species tree (Liu et al. 2010) and that could make MP-EST more problematic when the input gene trees are based on loci with very low information content.

In contrast to MP-EST and ASTRAL, Triplec appeared to be relatively robust to the use of poorly resolved gene trees. Instead, our results suggest that Triplec requires a large number of gene trees to converge on an accurate estimate of the species tree. However, similar to the results for MP-EST and ASTRAL, the program used to estimate input gene trees did have an impact on support values.

The basis for the differences in support observed for species tree methods using input gene trees estimated with different programs is unclear. Different models were used for analyses, but the observed patterns probably reflected more than the differences among the models used (e.g., overparamaterization of some analyses). The results of MP-EST-RAxML analyses (which used the GTR+Γ model) were more similar to those from MP-EST-PhyML and MP-EST-MrBayes (Supplementary File S5) than to MP-EST-GARLI (Fig. 5 and Supplementary File S5), despite the use of less parameter-rich models for many gene trees in GARLI, PhyML, and MrBayes. It seems likely that differences among programs in the details of the tree search and/or numerical optimization routines (e.g., parameter and branch length estimation) may play a role. Although ASTRAL appeared to be somewhat more robust to differences among the input gene trees, it did not appear to be immune to these effects (Table 2 and Supplementary File S6). Regardless of the details, the program used to estimate input gene trees clearly had a major impact on the results of our species tree analyses using UCE data.

Considerations for Species Tree Estimation using UCEs

Unlike analyses of UCE data using concatenated data, where a single topology emerged using different analytical approaches (i.e., MP, unpartitioned ML, and partitioned ML), analyses using species tree methods resulted in very different topologies depending on the details of the analysis. Perhaps more problematic, support values for contradictory nodes in the estimated species trees were often quite high and there was a surprisingly high degree of dependence on the program used to estimate gene trees. These findings raise questions about the use of summary species tree methods with low-information-content loci. Method development for species tree approaches is likely to profit from improving the extraction of historical signal despite error in gene tree estimates.

SMRT-ML has the potential to address concerns that species trees may lie in the anomaly zone while simultaneously taking advantage of the increased power from using concatenated data. However, the reliance of SMRT-ML on rooted triplets (when an outgroup is used to root the triplets, as we did here, they actually correspond to quartets) raises concerns regarding the

potential impact of long-branch attraction. Similar concerns probably apply to other species tree methods based on analyses of concatenated data for rooted triplets or quartets of taxa (e.g., SVDquartets [Chifman and Kubatko 2015]). However, loci with low rates of evolution are less likely to exhibit long-branch attraction than more rapidly evolving loci, suggesting that UCEs may be especially useful for SMRT-ML analyses (especially when the other advantages of using conserved loci in phylogenetic analyses [see Betancur-R. et al. 2014] are considered). An additional advantage of SMRT-ML (and related methods) is the expectation of consistency regardless of the recombination landscape. This is not true for standard species tree methods (including co-estimation methods) because individual regions that are implicitly viewed as having a single gene tree may actually represent a mosaic of histories (see Hobolth et al. 2011). The potential for intragenic recombination to have a negative impact on summary species tree methods is a subject of lively debate, with some authors suggesting that recombination represents a major challenge for those methods (Gatesy and Springer 2013; Gatesy and Springer 2014; Springer and Gatesy 2016) and others arguing that recombination is unlikely to be problematic (Lanier and Knowles 2012; Wu et al. 2013; Edwards et al. 2016). Unfortunately, SMRT-ML has seldom been used in empirical studies [recent exceptions are DeGiorgio et al. 2014, Sun et al. 2014, and Hosner et al. 2015]. Thus, the parts of parameter space where SMRT-ML behaves well remain poorly characterized.

There are several species tree methods that we did not explore. Bayesian co-estimation methods are computationally impractical for data sets of this size. Even if they were practical, it is reasonable to be concerned that the priors will dominate the gene tree topologies when low-information loci are analyzed, though the impact this might have on species tree estimation is unclear. Binning loci with similar histories has also been suggested to improve the estimated distribution of gene trees when phylogenetic signal is limited (Bayzid and Warnow 2013; Mirarab et al. 2014a). However, we believe that our analyses suggest binning methods are poorly suited to UCE data. The limited variation of individual UCEs could result in random binning if the tree compatibility threshold is high. However, a lower threshold of support could amplify phylogenetic errors given the surprisingly large number of trees that conflicted with our "gold standard" branch (the erectile clade; Fig. 3b–f). This could result in the need for complex "fine tuning" of the compatibility threshold. Finally, we did not examine methods that use branch lengths in the gene trees (e.g., STEM; Kubatko et al. 2009) because we felt that the low variation of UCEs makes the recovery of accurate branch lengths impractical.

There are a number of statistical issues that deserve consideration when applying species tree methods to UCE data (or any type of molecular data). The most commonly cited rationale for using species tree methods is the inconsistency of concatenation in parts of parameter space (e.g., Kubatko and Degnan 2007;

Edwards 2009; Liu and Edwards 2009; Edwards et al. 2016). However, consistency is not the only criterion that should be used to choose among methods of phylogenetic estimation (Sanderson and Kim 2000) and proofs of consistency (as well as simulations to assess the performance of methods with increasing amounts of data) often rely on assumptions that are unlikely to be "biologically realistic" (Roch and Warnow 2015; Warnow 2015; Springer and Gatesy 2016). Typical assumptions for species tree methods include the absence of hybridization and population subdivision within species, because both distort the frequency of gene trees relative to expectation under the multispecies coalescent (Holland et al. 2008; Slatkin and Pollack 2008). It seems reasonable to view all of the methods we employed, including concatenation, as approaches that are based on approximations to the true underlying model (which is expected to include discordance among gene trees due to the multispecies coalescent). It is known that analyses based on good approximating models often provide useful estimates of phylogeny even when the data violate some assumptions of those models (cf. Sullivan and Swofford 2001). Unfortunately, the best approximating model for analyses of rapid radiations when many individual loci have limited information (Supplementary Fig. S4) and estimates of gene trees from those loci have a high error rate (Fig. 3b–f) remains unclear; further research (and method development) in this area is necessary.

Relationships among Gallopheasants and Other Phasianids

The only difference between the species tree estimate from highly informative UCEs and concatenated analyses was the swapping of the positions of *Phasianus* and *Chrysolophus* relative to the concatenated tree (Fig. 4, left, and Supplementary Files S5 and S6). This node also varied between UCEs and introns and combined introns plus mitochondrial data. Both topologies have been recovered in previous multilocus studies (e.g., Kimball and Braun 2008; Wang et al. 2013).

Could the base of the gallopheasant radiation be in the anomaly zone? The recovery of asymmetric species tree topologies has been used to argue that the results of concatenated analyses do not reflect an anomaly zone problem (Degnan and Rosenberg 2006; Harshman et al. 2008; Rosenberg 2013; Smith et al. 2013). The same arguments apply here, because our species tree analyses resulted in an asymmetric "swapped tree" topology [using the Kubatko and Degnan 2007 nomenclature]. Could relationships among *Phaisanus, Chrysolophus*, and the CCL clade reflect a hard polytomy? Most SMRT-ML bootstrap replicates supported one of two specific topologies (Fig. 6). The third topology consistent with the strict consensus topology of both SMRT-ML analyses (i.e., presence of a *Chrysolophus-Phasianus* clade) was not recovered in any SMRT-ML bootstrap replicates (see Supplementary File S8). This suggests that there is not a hard polytomy in the gallopheasants, because that would predict approximately equal phylogenetic

signal for each resolution. However, it remains possible that the cladogenic events at the base of gallopheasants reflect a complex speciation involving gene flow [i.e., similar to that suggested for humans and chimpanzees by Patterson et al. 2006] or population subdivision for one or more ancestral species (cf. Slatkin and Pollack 2008).

We also observed differences outside the erectile clade, and these were very sensitive to analytical differences. The topology outside the erectile clade has been difficult to establish (e.g., Crowe et al. 2006; Kimball and Braun 2008; Bonilla et al. 2010; Wang et al. 2013; Kimball and Braun 2014), and the taxon sample we used here also includes a few short internodes (Fig. 1; also see Kimball and Braun 2014). The limited taxon sampling outside the focal clade for this study may have further limited our ability to resolve these relationships.

The use of UCEs to address a variety of phylogenetic questions is increasing, and many UCE studies have employed summary species tree methods (e.g., McCormack et al. 2012; Faircloth et al. 2013; McCormack et al. 2013; Crawford et al. 2015; Streicher et al. 2016), so it is important to understand the best practices for analyzing this type of data. Here, we largely focused on the use of species tree methods, and we found that using these analyses appeared to be more challenging than concatenated analyses. Although our results suggest UCEs provide valuable phylogenetic information we also found that gene trees estimated from low-signal UCEs appear to have a relatively high error rate. Two of the summary species tree methods (MP-EST and ASTRAL) performed better when gene trees based on the more variable UCEs were used, so it may be possible to ameliorate the problem of errors in gene trees by generating longer UCE contigs with improved laboratory and/or sequence assembly methods. However, extending the UCE contigs could also introduce sites with more homoplasy. Moreover, some error is likely to remain for any reasonable sequence length [see Patel et al. 2013, who conducted simulations using parameters based on intronic data]. This suggests that the benefits of extending UCE contigs relative to improving analytical methods should also be evaluated carefully. Developing species tree methods that are better able to tolerate errors in individual gene trees may ultimately prove to be a better approach. Indeed, methods like SMRT-ML may perform very well with markers that exhibit limited variation but low homoplasy; those methods should be investigated more broadly in this context. Despite the challenges that UCE data present for species tree methods, they clearly remain very promising markers with the potential to resolve many difficult phylogenetic questions.

SUPPLEMENTARY INFORMATION

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.p1m52.

REFERENCES

Anisimova M., Gil M., Dufayard J.F., Dessimoz C., Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60:685–699.

Armstrong M.H., Braun E.L., Kimball R.T. 2001. Phylogenetic utility of avian ovomucoid intron G: A comparison of nuclear and mitochondrial phylogenies in Galliformes. Auk 118:799–804.

Avise J.C., Robinson T.J. 2008. Hemiplasy: A new term in the lexicon of phylogenetics. Syst. Biol. 57:503–507.

Bao X.K., Liu N.F., Qu J.Y., Wang X.L., An B., Wen L.Y., Song S. 2010. The phylogenetic position and speciation dynamics of the genus *Perdix* (Phasianidae, Galliformes). Mol. Phylogenet. Evol. 56:840–847.

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.

Baum D.A. 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56:417–426.

Bayzid M.S., Warnow T. 2013. Naive binning improves phylogenomic analyses. Bioinformatics 29:2277–2284.

Bejerano G., Pheasant M., Makunin I., Stephen S., Kent W.J., Mattick J.S., Haussler D. 2004. Ultraconserved elements in the human genome. Science 304:1321–1325.

Berry V., Gascuel O. 1996. On the interpretation of bootstrap trees: Appropriate threshold of clade selection and induced gain. Mol. Biol. Evol. 13:999–1011.

Betancur-R. R., Naylor G.J., Orti G. 2014. Conserved genes, sampling error, and phylogenomic inference. Syst. Biol. 63:257–262.

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.

Bush K.L., Strobeck C. 2003. Phylogenetic relationships of the Phasianidae reveals possible non-pheasant taxa. J. Hered. 94:472–489.

Chifman J., Kubatko, L. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J. Theor. Biol. 374:35–47.

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.

Cox W.A., Kimball R.T., Braun E.L. 2007. Phylogenetic position of the New World quail (Odontophoridae): Eight nuclear loci and three mitochondrial regions contradict morphology and the Sibley-Ahlquist tapestry. Auk 124:71–84.

Crawford N.G., Faircloth B.C., McCormack J.E., Brumfield R.T., Winker K., Glenn T.C. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. 8:783–786.

Crawford N.G., Parham J.F., Sellas A.B., Faircloth B.C., Glenn T.C., Papenfuss T.J., Henderson J.B., Hansen M.H., Simison W.B. 2015. A phylogenomic analysis of turtles. Mol. Phylogenet. Evol. 83: 250–257.

Creevey C.J., McInerney J.O. 2005. Clann: Investigating phylogenetic information through supertree analyses. Bioinformatics 21: 390–392.

Crowe T.M., Bowie R.C.K., Bloomer P., Mandiwana T.G., Hedderson T.A.J., Randi E., Pereira S.L., Wakeling J. 2006. Phylogenetics, biogeography and classification of, and character evolution in, gamebirds (Aves: Galliformes): Effects of character exclusion, data partitioning and missing data. Cladistics 22:495–532.

Dalloul R.A., Long J.A., Zimin A.V., Aslam L., Beal K., Blomberg L., Bouffard P., Burt D.W., Crasta O., Crooijmans R.P.M.A., Cooper K., Coulombe R.A., De S., Delany M.E., Dodgson J.B., Dong J.J., Evans C., Frederickson K.M., Flicek P., Florea L., Folkerts O., Groenen M.A.M., Harkins T.T., Herrero J., Hoffmann S., Megens H.J., Jiang A., de Jong P., Kaiser P., Kim H., Kim K.W., Kim S., Langenberger D., Lee M.K., Lee T., Mane S., Marcais G., Marz M., McElroy A.P., Modise T., Nefedov M., Notredame C., Paton I.R., Payne W.S., Pertea G., Prickett D., Puiu D., Qioa D., Raineri E., Ruffier M., Salzberg S.L., Schatz M.C., Scheuring C., Schmidt C.J., Schroeder S., Searle S.M.J., Smith E.J., Smith J., Sonstegard T.S., Stadler P.F., Tafer H., Tu Z.J., Van Tassell C.P., Vilella A.J., Williams K.P., Yorke J.A., Zhang L.Q., Zhang H.B., Zhang X.J., Zhang Y., Reed K.M. 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome assembly and analysis. PLoS Biol. 8:e1000475.

DeGiorgio M., Degnan J.H. 2010. Fast and consistent estimation of species trees using supermatrix rooted triples. Mol. Biol. Evol. 27:552–569.

DeGiorgio M., Syring J., Eckert A.J., Liston A., Cronn R., Neale D.B., Rosenberg N.A. 2014. An empirical evaluation of two-stage species tree inference strategies using a multilocus dataset from North American pines. BMC Evol. Biol. 14:67.

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:762–768.

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., Leaché 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.

Ewing G.B., Ebersberger I., Schmidt H.A., von Haeseler A. 2008. Rooted triple consensus and anomalous gene trees. BMC Evol. Biol. 8:118.

Faircloth B.C. 2014. phyluce: Phylogenetic estimation from ultraconserved elements. Available from: URL https://github.com/faircloth-lab/phyluce.

Faircloth B.C., Glenn T.C. 2012. Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. PLoS One 7:e42543.

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.

Faircloth B.C., Sorenson L., Santini F., Alfaro M.E. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS One 8:e65923.

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

Felsenstein J. 2013. PHYLIP (Phylogeny Inference Package) version 3.695. Available from: URL http://evolution.genetics.washington.edu/phylip.html, Department of Genome Sciences, University of Washington, Seattle, WA.

Fujita M.K., Engstrom T.N., Starkey D.E., Shaffer H.B. 2004. Turtle phylogeny: Insights from a novel nuclear intron. Mol. Phylogenet. Evol. 31:1031–1040.

Gatesy J., Springer M.S. 2013. Concatenation versus coalescence versus "concatalescence". Proc. Natl. Acad. Sci. 110:E1179.

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.

Green R.E., Braun E.L., Armstrong J., Earl D., Nguyen N., Hickey G., Vandewege M.W., St John J.A., Capella-Gutierrez S., Castoe T.A., Kern C., Fujita M.K., Opazo J.C., Jurka J., Kojima K.K., Caballero J., Hubley R.M., Smit A.F., Platt R.N., Lavoie C.A., Ramakodi M.P., Finger J.W., Suh A., Isberg S.R., Miles L., Chong A.Y., Jaratlerdsiri W., Gongora J., Moran C., Iriarte A., McCormack J., Burgess S.C., Edwards S.V., Lyons E., Williams C., Breen M., Howard J.T., Gresham C.R., Peterson D.G., Schmitz J., Pollock D.D., Haussler D., Triplett E.W., Zhang G., Irie N., Jarvis E.D., Brochu C.A., Schmidt C.J., McCarthy F.M., Faircloth B.C., Hoffmann F.G., Glenn T.C., Gabaldoon T., Paten B., Ray D.A. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346:1254449.

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

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.

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. 105:13462–13467.

Hedtke S.M., Morgan M.J., Cannatella D.C., Hillis D.M. 2013. Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. PLoS One 8:e67908.

Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Hillier L.W., Miller W., Birney E., Warren W., Hardison R.C., Ponting C.P., Bork P., Burt D.W., Groenen M.A.M., Delany M.E., Dodgson J.B., Chinwalla A.T., Cliften P.F., Clifton S.W., Delehaunty K.D., Fronick C., Fulton R.S., Graves T.A., Kremitzki C., Layman D., Magrini V., McPherson J.D., Miner T.L., Minx P., Nash W.E., Nhan M.N., Nelson J.O., Oddy L.G., Pohl C.S., Randall-Maher J., Smith S.M., Wallis J.W., Yang S.P., Romanov M.N., Rondelli C.M., Paton B., Smith J., Morrice D., Daniels L., Tempest H.G., Robertson L., Masabanda J.S., Griffin D.K., Vignal A., Fillon V., Jacobbson L., Kerje S., Andersson L., Crooijmans R.P.M., Aerts J., van der Poel J.J., Ellegren H., Caldwell R.B., Hubbard S.J., Grafham D.V., Kierzek A.M., McLaren S.R., Overton I.M., Arakawa H., Beattie K.J., Bezzubov Y., Boardman P.E., Bonfield J.K., Croning M.D.R., Davies R.M., Francis M.D., Humphray S.J., Scott C.E., Taylor R.G., Tickle C., Brown W.R.A., Rogers J., Buerstedde J.M., Wilson S.A., Stubbs L., Ovcharenko I., Gordon L., Lucas S., Miller M.M., Inoko H., Shiina T., Kaufman J., Salomonsen J., Skjoedt K., Wong G.K.S., Wang J., Liu B., Wang J., Yu J., Yang H.M., Nefedov M., Koriabine M., deJong P.J., Goodstadt L., Webber C., Dickens N.J., Letunic I., Suyama M., Torrents D., von Mering C., Zdobnov E.M., Makova K., Nekrutenko A., Elnitski L., Eswara P., King D.C., Yang S., Tyekucheva S., Radakrishnan A., Harris R.S., Chiaromonte F., Taylor J., He J.B., Rijnkels M., Griffiths-Jones S., Ureta-Vidal A., Hoffman M.M., Severin J., Searle S.M.J., Law A.S., Speed D., Waddington D., Cheng Z., Tuzun E., Eichler E., Bao Z.R., Flicek P., Shteynberg D.D., Brent M.R., Bye J.M., Huckle E.J., Chatterji S., Dewey C., Pachter L., Kouranov A., Mourelatos Z., Hatzigeorgiou A.G., Paterson A.H., Ivarie R., Brandstrom M., Axelsson E., Backstrom N., Berlin S., Webster M.T., Pourquie O., Reymond A., Ucla C., Antonarakis S.E., Long M.Y., Emerson J.J., Betran E., Dupanloup I., Kaessmann H., Hinrichs A.S., Bejerano G., Furey T.S., Harte R.A., Raney B., Siepel

A., Kent W.J., Haussler D., Eyras E., Castelo R., Abril J.F., Castellano S., Camara F., Parra G., Guigo R., Bourque G., Tesler G., Pevzner P.A., Smit A., Fulton L.A., Mardis E.R., Wilson R.K. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716.

Hobolth A., Dutheil J.Y., Hawks J., Schierup M.H., Mailund T. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res. 21:349–356.

Holder M.T., Sukumaran J., Lewis P.O. 2008. A justification for reporting the majority-rule consensus tree in Bayesian phylogenetics. Syst. Biol. 57:814–821.

Holland B.R., Benthin S., Lockhart P.J., Moulton V., Huber K.T. 2008. Using supernetworks to distinguish hybridization from lineage-sorting. BMC Evol. Biol. 8:202.

Hosner P.A., Faircloth B.C., Glenn T.C., Braun E.L., Kimball R.T. Forthcoming 2015. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. doi: 10.1093/molbev/msv347).

Huelsenbeck J.P., Larget B., Alfaro M.E. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Mol. Biol. Evol. 21:1123–1133.

Huelsenbeck J.P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Hurvich C.M., Tsai C.L. 1989. Regression and time-series model selection in small samples. Biometrika 76:297–307.

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.

Kim J. 2000. Slicing hyperdimensional oranges: The geometry of phylogenetic estimation. Mol. Phylogenet. Evol. 17:58–75.

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. 2014. Does more sequence data improve estimates of galliform phylogeny? Analyses of a rapid radiation using a complete data matrix. PeerJ 2:e361.

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., Braun E.L., Ligon J.D., Randi E., Lucchini V. 2006. Using molecular phylogenetics to interpret evolutionary changes in morphology and behavior in the Phasianidae. Acta Zoologica Sinica 52 (supplement):362–365.

Kimball R.T., Braun E.L., Zwartjes P.W., Crowe T.M., Ligon J.D. 1999. A molecular phylogeny of the pheasants and partridges suggests that these lineages are not monophyletic. Mol. Phylogenet. Evol. 11:38–54.

Kimball R.T., St.Mary C.M., Braun E.L. 2011. A macroevolutionary perspective on multiple sexual traits in the Phasianidae (Galliformes). Int. J. Evol. Biol. 2011:423938.

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.

Kocher T.D., Thomas W.K., Meyer A., Edwards S.V., Paabo S., Villablanca F.X., Wilson A.C. 1989. Dynamics of mitochondrial DNA evolution in animals - Amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. 86:6196–6200.

Kubatko L.S., Carstens B.C., Knowles, L.L. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25:971–973.

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

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

Lanfear R., Calcott B., Kainer D., Mayer C., Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14.

Lanier H.C., Huang H.T., Knowles L.L. 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Mol. Phylogenet. Evol. 70:112–119.

Lanier, H.C., Knowles, L.L. 2012. Is recombination a problem for species-tree analyses? Syst. Biol. 61:691–701.

Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61:727–744.

Li C.H., Hofreiter M., Straube N., Corrigan S., Naylor G.J.P. 2013. Capturing protein-coding genes across highly divergent species. Biotechniques 54:321–326.

Liu L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24:2542–2543.

Liu L., Edwards S.V. 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58:452–460.

Liu L., Xi Z., Davis C.C. 2015. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Mol. Biol. Evol. 32:791–805.

Liu L.A., Yu L.L., Edwards S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10:302.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Matthee C.A., Eick G., Willows-Munro S., Montgelard C., Pardini A.T., Robinson T.J. 2007. Indel evolution of mammalian introns and the utility of non-coding nuclear markers in eutherian phylogenetics. Mol. Phylogenet. Evol. 42:827–837.

McCormack J.E., Faircloth B.C., Crawford N.G., Gowaty P.A., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 22:746–754.

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.

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, 1–8.

Mirarab S., Bayzid M.S., Boussau B., Warnow T. 2014a. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science 346:1250463.

Mirarab S., Bayzid M.S., Warnow T. Forthcoming 2014b. Evaluating summary methods for multilocus species tree estimation in

the presence of incomplete lineage sorting. Syst. Biol. doi: 10.1093/sysbio/syu063).

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

Nadeau N.J., Burke T., Mundy N.I. 2007. Evolution of an avian pigmentation gene correlates with a measure of sexual selection. Proc. R. Soc. B 274:1807–1813.

Nylander J.A.A. MrAIC.pl. 2004. Program distributed by the author. Available from: URL http://www.abc.se/~nylander/mraic/mraic.html, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Patel S., Kimball R.T., Braun E.L. 2013. Error in phylogenetic estimation for bushes in the Tree of Life. J. Phylogenetics Evol. Biol. 1:110.

Patterson N., Richter D.J., Gnerre S., Lander E.S., Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature 441:1103–1108.

Peñalba J.V., Smith L.L., Tonione M.A., Sass C., Hykin S.M., Skipwith P.L., McGuire J.A., Bowie R.C.K., Moritz C. 2014. Sequence capture using PCR-generated probes: A cost-effective method of targeted high-throughput sequencing for nonmodel organisms. Mol. Ecol. Resour. 14:1000–1010.

Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Worheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. PLoS Biol. 9:e1000602.

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

Posada D., Crandall K.A. 1998. MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817–818.

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

Rambaut A., Suchard M.A., Xie D., Drummond A.J. Tracer v1.6. Available from: URL http://beast.bio.ed.ac.uk/Tracer. 2014.

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.

Roch S., Warnow T. 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Syst. Biol. 64:663–676.

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

Rokas A., Williams B.L., King N., Carroll S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804.

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

Salichos L., Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497:327–331.

Sanderson M.J., Kim J. 2000. Parametric phylogenetics? Syst. Biol. 49:817–829.

Schwarz G. 1978. Estimating the dimension of a model. Ann. Statist. 6:461–464.

Shen Y.Y., Liang L., Sun Y.B., Yue B.S., Yang X.J., Murphy R.W., Zhang Y.P. 2010. A mitogenomic perspective on the ancient, rapid radiation in the Galliformes with an emphasis on the Phasianidae. BMC Evol. Biol. 10:132.

Simmons M.P., Gatesy J. 2015. Coalescence vs. concatenation: Sophisticated analyses vs. first principles applied to rooting the angiosperms. Mol. Phylogenet. Evol. 91: 98–122.

Slatkin M., Pollack J.L. 2008. Subdivision in an ancestral species creates asymmetry in gene trees. Mol. Biol. Evol. 25:2241–2246.

Slowinski J.B. 2001. Molecular polytomies. Mol. Phylogenet. Evol. 19:114–120.

Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C., Brumfield R.T. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63:83–95.

Smith J.V., Braun E.L., Kimball R.T. 2013. Ratite nonmonophyly: Independent evidence from 40 novel loci. Syst. Biol. 62:35–49.

Sorenson M.D., Ast J.C., Dimcheff D.E., Yuri T., Mindell D.P. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol. Phylogenet. Evol. 12:105–114.

Springer M.S., Gatesy J. 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94:1–33.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stein R.W., Brown J.W., Mooers A.Ø. 2015. A molecular genetic time scale demonstrates Cretaceous origins and multiple diversification rate shifts within the order Galliformes (Aves). Mol. Phylogenet. Evol. 92:155–164.

Streicher J.W., Schulte II J.A., Wiens J.J. 2016. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. Syst. Biol. 65:128–145.

Strimmer K., von Haeseler A. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964–969.

Sullivan J., Swofford D.L. 2001. Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated?. Syst. Biol. 50:723–729.

Sun K.P., Meiklejohn K.A., Faircloth B.C., Glenn T.C., Braun E.L., Kimball R.T. 2014. The evolution of peafowl and other taxa with ocelli (eyespots): A phylogenomic approach. P. Roy. Soc. B-Biol. Sci. 281:20140823.

Swofford D.L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates. Sunderland, MA.

Tonini J., Moore A., Stern D., Shcheglovitova M., Ortí G. 2015 Mar 9. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Curr.: Tree of Life; doi: 10.1371/currents.tol.34260cc27551a34527b34124ec34265f36334b34266be.

Wang N., Kimball R.T., Braun E.L., Liang B., Zhang Z.W. 2013. Assessing phylogenetic relationships among Galliformes: A multigene phylogeny with expanded taxon sampling in Phasianidae. PLoS One 8:e64312.

Warnow T. 2015 May 22. Concatenation analyses in the presence of incomplete lineage sorting. PLoS Curr.: Tree of Life. doi: 10.1371/currents.tol.1378d1341ac1370f1313d1371abedf1374c1374a1359f1375d1317b1371f1377.

Whitfield J.B., Lockhart P.J. 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22:258–265.

Wolf Y.I., Rogozin I.B., Koonin E.V. 2004. Coelomata and not ecdysozoa: Evidence from genome-wide phylogenetic analysis. Genome Res. 14:29–36.

Wu S., Song S., Liu L., Edwards S.V. 2013. Reply to Gatesy and Springer: The multispecies coalescent model can effectively handle recombination and gene tree heterogeneity. Proc. Natl. Acad. Sci. 110:E1180.

Xi Z., Liu L., Davis C.C. 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol. Phylogenet. Evol. 92:63–71.

Zerbino D.R., Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Zwickl D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin.

Zwickl D.J., Stein J.C., Wing R.A., Ware D., Sanderson M.J. 2014. Disentangling methodological and biological sources of gene tree discordance on Oryza (Poaceae) chromosome 3. Syst. Biol. 63: 645–659.