Gene 801 (2021) 145841

Contents lists available at ScienceDirect

Gene

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

Research paper

When good mitochondria go bad: Cyto-nuclear discordance in landfowl (Aves: Galliformes)

Rebecca T. Kimball [a,*], Marisa Guido [a,b], Peter A. Hosner [a,c], Edward L. Braun [a]

[a] Department of Biology, University of Florida, Gainesville, FL, USA
[b] Department of Biological Sciences, Duquesne University, Pittsburgh, PA, USA
[c] Natural History Museum of Denmark & Center for Global Mountain Biodiversity, Copenhagen University, Copenhagen, Denmark

ARTICLE INFO

Keywords:
Phylogenomics
Species trees
Introgression
Incomplete lineage sorting

ABSTRACT

Mitochondrial sequences were among the first molecular data collected for phylogenetic studies and they are plentiful in DNA sequence archives. However, the future value of mitogenomic data in phylogenetics is uncertain, because its phylogenetic signal sometimes conflicts with that of the nuclear genome. A thorough understanding of the causes and prevalence of cyto-nuclear discordance would aid in reconciling different results owing to sequence data type, and provide a framework for interpreting megaphylogenies when taxa which lack substantial nuclear data are placed using mitochondrial data. Here, we examine the prevalence and possible causes of cyto-nuclear discordance in the landfowl (Aves: Galliformes), leveraging 47 new mitogenomes assembled from off-target reads recovered as part of a target-capture study. We evaluated two hypotheses, that cyto-nuclear discordance is "genuine" and a result of biological processes such as incomplete lineage sorting or introgression, and that cyto-nuclear discordance is an artifact of inaccurate mitochondrial tree estimation (the "inaccurate estimation" hypothesis). We identified seven well-supported topological differences between the mitogenomic tree and trees based on nuclear data. These well-supported topological differences were robust to model selection. An examination of sites suggests these differences were driven by small number of sites, particularly from third-codon positions, suggesting that they were not confounded by convergent directional selection. Hence, the hypothesis of genuine discordance was supported.

1. Introduction

Mitochondrial sequence data have a long history as phylogenetic markers in animals (Kocher et al., 1989; Lavrov, 2007) and remains the most extensively sampled type of DNA sequence data in many groups (e.g., birds; Burleigh et al., 2015). Despite the extensive use of mitochondrial sequence data, limitations on its use have been noted (Ballard and Whitlock, 2004). In particular, since the mitochondrion is completely (or largely) non-recombining (e.g., Berlin and Ellegren, 2001; Berlin et al., 2004), it largely represents a single genetic marker. Thus, introgression and incomplete lineage sorting (ILS) may lead to a mitochondrial tree that differs from the species tree (Maddison and Wiens, 1997; Rubinoff and Holland, 2005). Given that next generation sequencing (NGS) methods have made it straightforward to obtain large sets of orthologus nuclear data throughout the genome for non-model systems (e.g., Faircloth et al., 2012; Jarvis et al., 2014; Prum et al., 2015), some researchers may question whether mitochondrial data is still useful in addressing phylogenetic questions given that more reliable estimates of species trees can be readily obtained.

However, mitogenomic data can still provide important contributions to phylogenetic studies for several reasons. First, it is the most commonly sampled marker for many groups (e.g., Burleigh et al., 2015). Hence, there is extensive amounts of mitochondrial sequence data available, including from rare and even extinct taxa (e.g., Mitchell et al., 2014), as well as from extensive population sampling within species that is unlikely to replicated soon using NGS approaches. Continued collection of mitochondrial data, particularly as it can be obtained as a by-product of some NGS approaches (e.g., Meiklejohn et al., 2014; Raposo do Amaral et al., 2015; Tamashiro et al., 2019) allows linking of existing data with newly collected data to provide a broader

Abbreviations: AICc, Akaike information criterion; DNA, deoxyribonucleic acid; ILS, incomplete lineage sorting; ML, maximum likelihood; ND3, NADH dehydrogenase 3; ND4, NADH dehydrogenase 4; RF, Robinson-Foulds distances; rRNAs, ribosomal ribonucleic acids; UCEs, ultraconserved elements.

* Corresponding author.
E-mail address: rkimball@ufl.edu (R.T. Kimball).

https://doi.org/10.1016/j.gene.2021.145841
Received 5 February 2021; Received in revised form 10 June 2021; Accepted 13 July 2021
Available online 16 July 2021

understanding of evolutionary relationships than might be possible if only nuclear sampling is used. Additionally, the mitochondrion is a long, variable region in which the gene tree can also be estimated without consideration of recombination for many taxa (see Springer and Gatesy, 2016, 2018 for a discussion of the recombination issue). Furthermore, the mitochondrion has a shorter coalescent time than nuclear loci, so on average the mitochondrial gene tree is more likely to match the species tree than phylogenies estimated from nuclear loci (Moore, 1995).

Although the mitochondrial gene tree has a higher probability of reflecting the species tree than individual nuclear markers, cyto-nuclear discordance (incongruence between organellar and nuclear topologies) is regularly observed with species trees (Wang et al., 2017; Tamashiro et al., 2019), even when using complete mitogenomes (individual mitochondrial regions often lack the power to robustly resolve relationships; Meiklejohn et al., 2014). However, to effectively leverage what we can learn from the mitochondrion, and ensure it facilitates, not hinders, our overall understanding of evolutionary relationships, it is important to better understand why we may observe cyto-nuclear discordance. Observed discordance may be due to biological processes such as ILS or introgression ("genuine discordance"; Tamashiro et al., 2019) or to errors in estimation of the mitochondrial tree ("inaccurate estimation"; Tamashiro et al., 2019), due to factors such as limited taxon sampling, use of models that have a poor fit to the data (e.g., Braun and Kimball, 2002), or low statistical power (analyzing loci with too few variable sites, such as a single mitochondrial region; Cao et al., 1998; Meiklejohn et al., 2014)).

Testing between the genuine discordance and inaccurate estimation hypotheses is challenging. A reduction in cyto-nuclear discordance when analyzing (mostly) complete mitogenomes by use of more complex models, such as partitioned analyses (Leavitt et al., 2013; Wang et al., 2017) or with increased taxon sampling (which can improve parameter estimation; Cummings and Meyer, 2005) supports the inaccurate estimation hypothesis (e.g., Tamashiro et al., 2019). However, there are also cases where using more complex models or adding taxa does not alter the topology—yet it remains possible that adding even more taxa (possibly requiring inclusion of long-extinct taxa) or even more accurate molecular evolution models (which may not have been developed or be computationally feasible) would remove the incongruence. So when improved taxon sampling or models do not improve congruence, it remains unclear what drives discordance.

In cases where it can be difficult to support or refute hypotheses, it can be important to explore the issue using multiple approaches. Here we explore an alternative approach to understanding cyto-nuclear discordance by focusing on individual sites within the mitochondrion that might drive topological differences. When relatively similar tree topologies are compared, the number of sites that strongly favor one or the other of those topologies is typically quite small (Evans et al., 2010; Kimball et al., 2013; Shen et al., 2017; Pandey and Braun, 2020). Kimball et al. (2013) referred to those sites as "decisive sites". We hypothesized that examining decisive sites might aid in distinguishing between the genuine discordance and inaccurate estimation hypotheses. Decisive sites in the mitogenome are expected to primarily favor the mitochondrial topology, but they might exhibit different distributions under the two hypotheses. Specifically, under the genuine discordance hypothesis, decisive sites in a mitogenomic alignment are likely to be unique to a single specific incongruent relationship and they are likely to be spread across the possible site positions (i.e., rRNAs and the codon positions) in proportion to the variation expected for those site positions. If we postulate that the sites driving incongruent relationships reflect sites for which evolution is not adequately modeled, then the same decisive sites might reappear in multiple comparisons. In addition, an unexpected distribution of decisive sites relative to site types might be observed, as this could reflect sites that are challenging to model or convergent evolution (e.g., nonsynonymous sites). Both of these observations would corroborate the inaccurate estimation hypothesis.

Using a dataset of 113 galliform mitogenomes (38% of the order, including at least 20% from each family; 42 of these from species not previously represented by a published mitogenome), we assessed which sites contributed to cyto-nuclear discordance by comparing analyses of this mitogenomic data with recent multi-locus (primarily phylogenomic) species trees. We then identified seven strongly supported conflicts (both within and among genera) between our partitioned mitochondrial tree and the nuclear tree and used site likelihoods to identify the decisive sites that strongly supported each of the seven conflicts.

2. Methods

2.1. Sequencing, assembly and alignment

Data for newly assembled mitogenomes came from previous sequence capture studies targeting nuclear ultraconserved elements (UCEs) across Galliformes (Sun et al., 2014; Hosner et al., 2016a; Hosner et al., 2016b; Hosner et al., 2017; Hosner et al., 2020; Meiklejohn et al., 2016; Persons et al., 2016; Wang et al., 2017). Although the majority of sequence data generated in this way corresponds to targeted nuclear regions, some off-target fragments are also sequenced. Since mitochondria are typically at higher copy than the nuclear genome (particularly in tissues such as muscle, brain, or liver), mitochondrial fragments are often obtained in sufficient quantity to allow assembly of complete or mostly complete mitogenomes (e.g., Meiklejohn et al., 2014; Raposo do Amaral et al., 2015; Tamashiro et al., 2019). Since sequence capture approaches also yield many nuclear sequences, they can provide a unique framework to compare the phylogenetic signal of mitogenomes with the signal for large numbers of nuclear loci.

We collected sequence data using two approaches from DNA extracted from fresh tissues (blood, muscle or liver) or toepad clips from dried museum specimens. For our first method, we prepared Nextera sequencing libraries using the manufacturer's protocols (Illumina, Inc., San Diego, CA), except we used primers with custom index tags (Faircloth and Glenn, 2012). We pooled 8 samples together, and enriched for 5,060 UCE loci (Mycroarray, Ann Arbor, MI; http://www.mycroarray.com/mybaits/mybaits-UCEs.html). Enriched libraries were amplified with 18 PCR cycles, quantified using qPCR (quantitative PCR; Kapa Biosystems), and 75 bp paired-end reads were obtained from an Illumina HiSeq 2000 (UC Irvine Genomics High-Throughput Facility). Our second method targeted the same 5,060 UCE loci. For these samples, library construction was performed by RAPiD Genomics (Gainesville, FL). We obtained 150 bp, paired-end reads on an Illumina HiSeq 3000. Regardless of sequencing procedure, reads were de-multiplexed and Trimmomatic (Bolger et al., 2014) was used to remove adaptors and poor quality reads.

To assemble mitogenomes, we mapped cleaned reads onto published mitogenomes using Geneious 6.1.6. We used several different reference genomes, so we mapped onto a closely related species (e.g., we mapped onto a published cracid mitogenome for species in Cracidae). To maximize coverage over more divergent regions (e.g., the control region) we used up to 10 iterations (in this process, the results of one iteration are then used as the reference for the next iteration). Some species (particularly those where DNA was extracted from blood) had few mitochondrial reads (Barker et al., 2015). These were not included in the final dataset, and only species where complete or nearly complete mitogenomes were assembled were retained for final analyses (all gene regions were present in all retained species, though in some cases short segments of unresolved nucleotides remained). After assembly, we examined the consensus sequence to ensure each protein-coding gene began with an appropriate start codon and that there were no unexpected frame shifts. All samples exhibited a single nucleotide frameshift in ND3 (which is known to be present in other galliform mitogenomes; see Mindell et al., 1998). Newly assembled mitogenomes are available as GenBank MW574349-MW574395.

Our final dataset contained 119 species, of which 114 were

galliforms and five were anseriform outgroups. This was done by combining our 47 newly assembled mitogenomes (46 galliforms and one anseriform outgroup) with 72 published mitogenomes (68 galliforms and four anseriforms). We used only one mitogenome for each species (using the IOC 10.2; Gill et al., 2020), so when multiple complete or nearly complete mitogenomes were available, we selected just one of these (see Supplementary Table S1 for GenBank and voucher details for each sample included). Published mitogenome sequences were already available for four of the species we sequenced; we analyzed our mitogenome assemblies in those cases.

Since the control region did not always assemble completely or unambiguously, we extracted the 13 protein coding genes and the two rRNAs for analysis (Alignment in Supplementary File S1). Alignment of protein coding genes was mostly straightforward, with two exceptions. First, we excluded from analyses the extra nucleotide in ND3. Second, there was an additional nucleotide in ND4 near the 3′ end of the published *Arborophila rufipectus* sequence. We observed a likely homologous nucleotide in *Rollulus rouloul*, but in our sequence it was preceded by a stop codon that led to a sequence four amino acids shorter. We excluded the additional site from analyses, but retained the homologous sites in *Rollulus* after the stop codon for phylogenetic estimation. For the rRNA regions, we aligned using Muscle 3.8.31 (Edgar, 2004) implemented in Mesquite 3.40 (Maddison and Maddison, 2019).

2.2. Analyses

We estimated the maximum likelihood (ML) topology in RAxML 8.2.10 (Stamatakis, 2014) using GTRGAMMA and 25 start trees as implemented on CIPRES (Miller et al., 2010). Support was estimated using 1000 rapid bootstrap replicates. We conducted both unpartitioned and partitioned analyses. To identify partitions, we ran PartitionFinder2 using the greedy algorithm (Lanfear et al., 2012; Lanfear et al., 2017; Guindon et al., 2010) on CIPRES. The input file separated the two rRNAs; each protein coding gene was separated into codon positions (for a maximum of 41 partitions). Based on the AICc, the best partitioning strategy included 28 partitions.

To ensure our mitochondrial topology was robust, we also estimated the ML topology and ultrafast bootstrap support using multiple different approaches in IQ-TREE 2.0.6 (Chernomor et al., 2016; Hoang et al., 2018; Minh et al., 2020). For the unpartitioned dataset, we implemented two analyses: 1) "standard model set" analysis where IQ-TREE identified the best model from the set of models examined by JModelTest (Posada, 2008); and 2) "complete model set" analysis where IQ-TREE identified the best model from a set that includes all implemented models (including free rate models); we refer to those analyses as. IQ-TREE Standard and IQ-TREE Free, respectively. We also had IQ-TREE estimate best partitioning strategies for partitioned analyses, using 1) to the set of models used by PartitionFinder and 2) using all models implemented in IQ-TREE; we refer to these as IQ-TREE Standard partitioned and IQ-TREE Free partitioned, respectively.

No single published nuclear topology matched the taxon sampling included in our analyses. The majority of taxa were included in the phylogenomic study of Hosner et al. (2017), but we supplemented this with other phylogenomic and multi-locus studies (Hosner et al., 2015; Hosner et al., 2020; Wang et al., 2017) to generate a nuclear topology for the taxa we sampled. Most relationships in these studies were strongly supported (typically 100% bootstrap support), there were no clear conflicts between species trees estimated under the multi-species coalescent versus concatenation, and there were no topological disagreements among these studies for taxa included in this analysis. Thus we deduced that our consensus nuclear topology (see Supplementary Fig. S1 and File S2) was a credible estimate of the species tree.

We compared the mitochondrial RAxML partitioned ML topology with the nuclear topology to identify strongly supported differences (where the mitochondrial topology showed at least 80% support for an alternative relationship in partitioned analyses). For each difference identified, we then generated a modified topology in which the partitioned mitochondrial ML tree was rearranged to match one of the conflicting nuclear relationships (i.e., the minimum number of rearrangements were made so that, with the exception of one set of relationships, the mitochondrial tree was unchanged). For one node (non-erectile clade), the modification was to cluster into a clade, rather than a grade. However, relationships were not modified within that clade, so the modified tree contained relationships within the non-erectile clade that differed from the nuclear species tree (though conflicting relationships were not well supported within this clade). All topologies, including modified topologies, are available in Supplementary File S2. After generation of the modified topologies, we estimated site likelihoods for the partitioned ML tree, and each of the modified topologies.

We extracted the site likelihoods from the partitioned topology, and subtracted the site likelihood from the modified topologies (so positive values reflected sites supporting the mitochondrial topology, while negative values indicated stronger support for the nuclear topology). To identify sites that appeared to make strong contributions to the differences ("decisive sites"; Kimball et al., 2013) we identified sites where the difference in log likelihood scores was greater than 5 standard deviations from the mean (based on differences at all sites for a specific modification). We then determined the gene region and codon position for each of the decisive sites identified in each of our seven comparisons.

3. Results

The estimate of mitochondrial genealogy from the partitioned RAxML analysis was largely well supported and it identified the five recognized families as monophyletic clades (Fig. 1). Symmetric distances (i.e., twice the Robinson-Foulds [RF] distance [Robinson and Foulds, 1981]) between the RAxML partitioned tree with IQ-TREE Standard and Free partitioned trees was 10 for both comparisons (since these were fully resolved trees this distance corresponds to five differences). Among the unpartitioned analysis, the two IQ-TREE topologies were identical (though the best-fitting model differed) and had an RFx2 distance to the RAxML unpartitioned tree of 2. Differences between partitioned and unpartitioned analyses were slightly greater (RFx2 distances 12 or 14), although all estimates of the mitogenomic tree were quite similar overall.

Differences between the nuclear topology and the mitochondrial topologies were larger than among the trees that resulted from analyses of mitochondrial data (RFx2 distances of 32 or 34), though the majority of nodes were still identical. We identified seven nodes that were well-supported (>80% bootstrap support in the RAxML partitioned analysis; Table 1) that conflicted with the nuclear topology (Fig. 2). For these discordant nodes, all of the mitochondrial analyses matched the relationships of the RAxML partitioned topology, typically with strong support (Table 1). Thus, these discordances were robust to method of analyzing the mitochondrial data.

In comparing the site likelihoods of the modified topologies with the unmodified RAxML partitioned ML tree, a total of 407 sites were identified as "decisive", in that the difference in site likelihoods (partitioned ML - modified tree) was greater than 5 standard deviations from the average for a particular tree comparison. There were a total of seven comparisons and the majority of decisive sites (383 of. 407; 94%) were identified in a single comparison (Table 2); the remaining 24 decisive sites (6%) were identified in more than one comparison. Of the shared sites, all but one was shared by only two comparisons (the remaining decisive site was shared by three comparisons). Decisive sites were found in all types of partitions (rRNA or codon position). Although more decisive sites occurred in 3rd positions, this likely reflected greater variability for 3rd positions (nearly 50% of variable sites in the alignment were 3rd position sites and just over 50% of decisive sites were at 3rd positions).

As expected, we found more positive differences (positive site

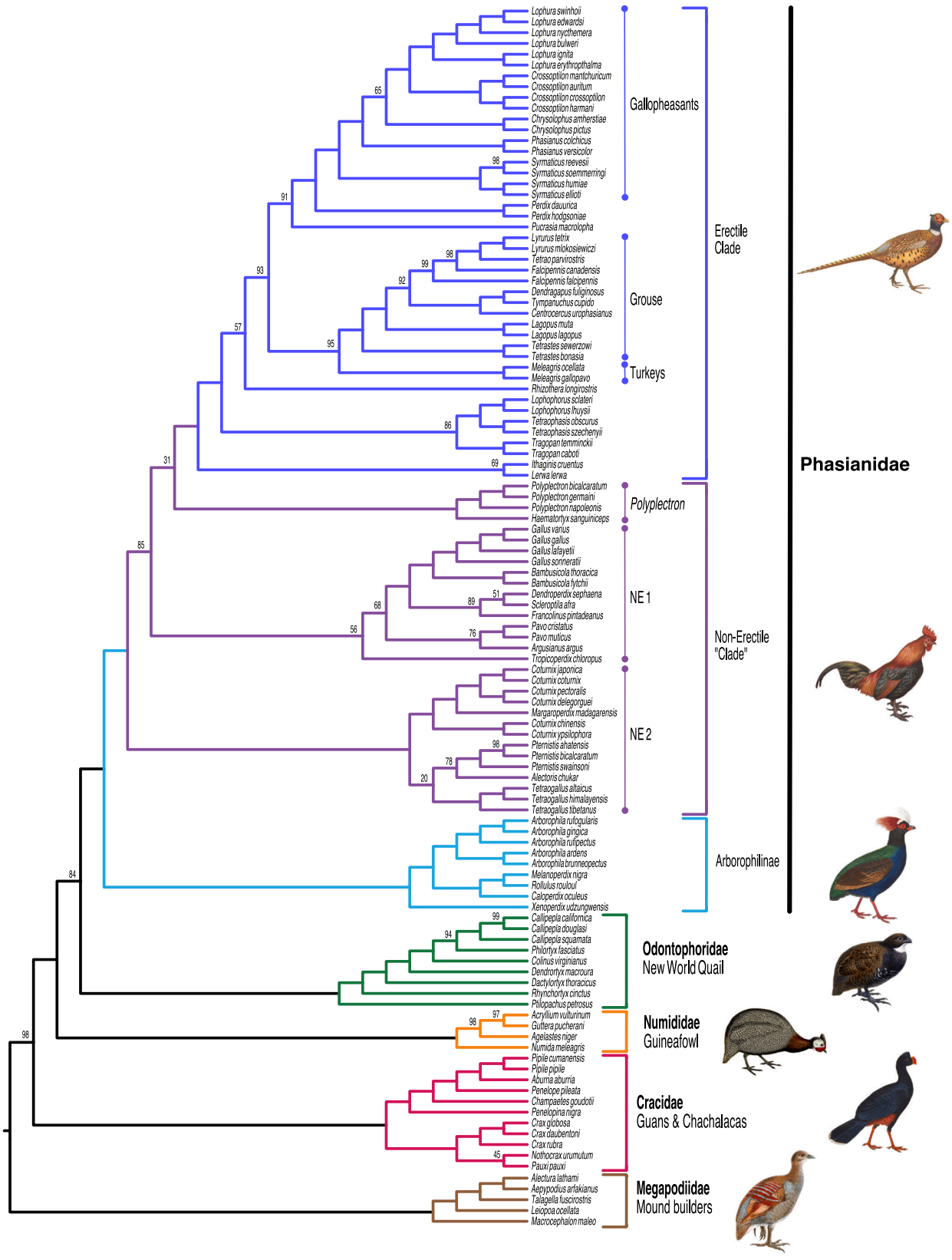

Fig. 1. Partitioned RAxML tree with bootstrap support. The members of the Non-erectile "clade" are monophyletic in the best available estimates of the galliform species tree (e.g., Hosner et al., 2016a; Hosner et al., 2016b; Hosner et al., 2017), but they form three lineages that are successive sister groups of the erectile clade in the mitogenomic tree. Nodes without values had 100% support. Images based on 19th century drawings are used to illustrate major groups within galliforms (see Supplementary File S3 for details).

Table 1
Bootstrap support for different analyses for nodes that differed from nuclear topology; in some cases one node changed while in others the rearrangement involved two nodes. See Fig. 2 for node identities.

	Partitioned Analyses			Unpartitioned Analyses		
	RAxML	IQ-TREE Standard	IQ-TREE Free	RAxML	IQ-TREE Standard	IQ-TREE Free
Lophura-1	100	100	100	100	100	100
Lophura-2	100	100	100	97	99	100
Pucrasia-1	91	98	100	97	98	99
Grouse-1	92	99	99	99	100	99
Gallus-1	100	100	100	100	100	100
Gallus-2	100	99	99	66	94	93
Non-Erectile-1	85	97	94	95	98	96
NWQ-1	94	85	80	83	98	98
Guinea-1	97	100	100	92	100	100
Guinea-2	98	100	99	98	100	100

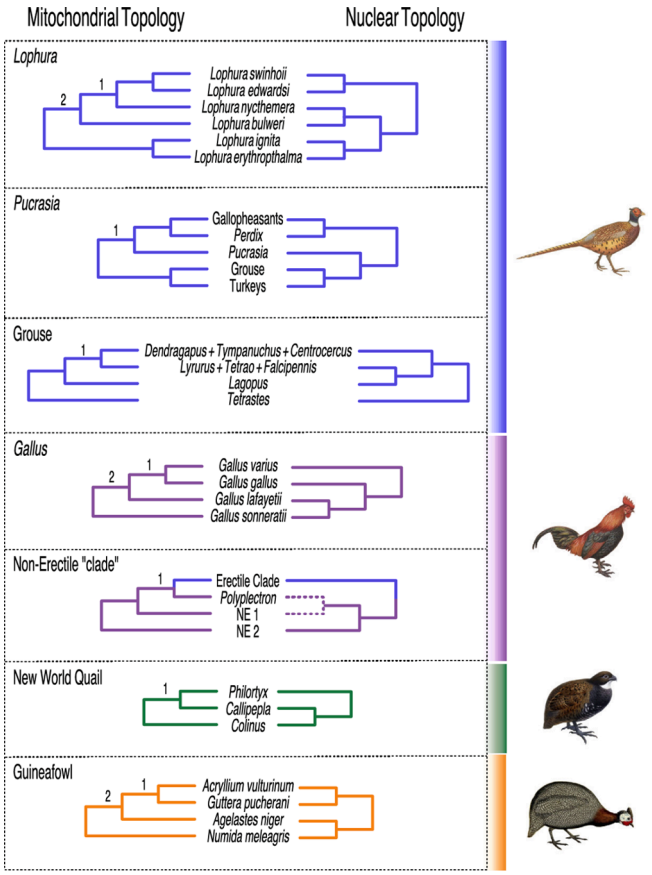

Fig. 2. Strongly supported differences between the mitogenomic topologies and the rearrangement based on the nuclear topology that was used in each tree comparison. The non-erectile clade comparison involves the rearrangement of major lineages (see Fig. 1). In the comparison we used a tree with the minimal rearrangement to yield monophyly of the non-erectile clade; for the nuclear topology, the *Polyplectron* clade nests within N1 (represented by dotted lines). Numbers at nodes refer to focal nodes referred to in Table 1. Information about the approximate positions of these conflicts in the overall phylogeny is provided using the illustrations used in Fig. 1.

differences support the optimal topology for the mitogenome) in the unique sites (71% overall). This was true for all comparisons except the rearrangement in the New Word Quail. For that comparison, just over 50% of decisive sites that were unique to that comparison (37 of 72) were negative, and all types of sites (those in each codon position and those in rRNAs) had either more or the same number of negative as compared to positive values.

4. Discussion

We obtained a phylogeny that was largely congruent with nuclear phylogenies (e.g., Hosner et al., 2016a; Hosner et al., 2016b; Hosner et al., 2017). We also found very similar results to other studies that have focused on analysis of complete galliform mitogenomes (e.g., Kan et al., 2010a; Kan et al., 2010b; Shen et al., 2010; Shen et al., 2014; Meiklejohn et al., 2014). However, by combining all available galliform mitogenomes, our analysis included more than twice the number of galliform species as any previous study. Most of the differences between our results and previous studies involved relationships that were not strongly supported in those studies. Our results were more congruent with the nuclear topology than previous mitogenomic studies, such as for the position of *Phasianus* and *Chrysolophus*, which form a clade in earlier mitogenomic studies (e.g., Meiklejohn et al., 2014; Shen et al., 2014). This suggests that taxon sampling can be important when analyzing mitogenomic data (e.g., Braun and Kimball, 2002; Tamashiro et al., 2019).

In spite of the improved taxon sampling and use of more complex models than in earlier mitogenomic papers of galliforms (Kan et al., 2010a; Kan et al., 2010b; Shen et al., 2010; Shen et al., 2014; Meiklejohn et al., 2014), there were still some relationships that were discordant with our best understanding of the underlying species tree as determined by multilocus (primarily phylogenomic) studies (Hosner et al., 2016a; Hosner et al., 2016b; Hosner et al., 2017; Hosner et al., 2020; Persons et al., 2016; Wang et al., 2017). Some of these differences were poorly supported in the mitogenomic tree (e.g., some relationships within the non-erectile clade), and thus may reflect cases where the mitogenome lacked the signal to robustly resolve relationships. These were often nodes that differed among our analyses, indicating that they were sensitive to the model and highlighting that there was little signal in the data to reliably resolve relationships. Excluding those weakly supported relationships, there still remained the seven discordant nodes (~6% of all nodes in the tree) that we focused on in our analyses.

Our observations that these nodes were consistently and strongly discordant across all of our analyses make it unlikely that the observed incongruences were due to inaccurate estimation. We hypothesized that examining decisive sites might be reveal several signatures of inaccurate estimation. If a subset of sites were poorly modeled, this might lead to the case where those sites are associated with multiple cases of incongruence. If strong selection led to convergent changes (e.g., Castoe et al., 2009), first and second codon positions would be likely to dominate the decisive sites supporting the mitogenomic tree with more decisive third codon positions supporting the nuclear tree. In contrast to these expectations, our results were more consistent with the genuine discordance hypothesis for all of the conflicts that we observed, with most decisive sites being unique, and occurring commonly at third codon positions. The one possible exception was the New World quail, where there was a slight majority of decisive sites supporting the nuclear topology.

Table 2
Numbers of decisive sites for each modified topology. Unique sites were also separated by type of site (rRNA or codon position). Numbers in parentheses are the sites that were positive (greater likelihood value in partitioned ML tree than for nuclear topology).

	Shared	Total	Total Unique	rRNA	Position 1	Position 2	Position 3
Lophura	12	67 (59)	55 (49)	10 (9)	5 (3)	5 (4)	35 (33)
Pucrasia	2	23 (21)	21 (19)	3 (2)	4 (4)	3 (3)	11 (10)
Grouse	5	55 (48)	50 (43)	9 (8)	17 (15)	1 (1)	23 (19)
Gallus	10	99 (65)	89 (59)	21 (15)	26 (14)	6 (6)	36 (24)
Non-Erec	4	53 (32)	49 (29)	6 (3)	14 (9)	5 (3)	24 (14)
NWQ	7	79 (39)	72 (35)	16 (7)	17 (7)	7 (5)	32 (16)
Guinea	8	55 (45)	47 (39)	6 (5)	4 (3)	1 (1)	36 (30)
TOTAL	24 sites		383 (273)	71 (49)	87 (55)	28 (23)	197 (146)

Genuine discordance is most likely due to either mitochondrial introgression (i.e., hybridization) or to incomplete lineage sorting. Hybridization has been observed among many extant galliforms (McCarthy, 2006). In all cases, our mitochondria exhibited many differences from each other, which suggests recent hybridization (which might be particularly common in captivity) is not likely, even though a number of the sequenced samples were from captive individuals. However, hybridization in the past is also likely to have occurred, and there are examples of galliform phylogenies where the nuclear genome exhibits signatures of historical introgression (e.g., whole-genome phylogenies; Tiley et al., 2020). Although mitochondrial introgression might be selected against over time, as it may lead to mitonuclear incompatibility, there are situations in which mitochondrial introgression may persist even if there is some fitness loss due to incompatibilities (Hill, 2019). Thus, it is possible that historical introgression underlies some or all of the discordances we observed.

However, we cannot rule out incomplete lineage sorting as an alternative explanation. While galliform species trees estimated using approaches that incorporate the multi-species coalescent have not suggested that any relationships are within the anomaly zone (Degnan and Rosenberg, 2006), different loci will still have distinct evolutionary histories, which could drive some or all of the discordance we observed. Although the shorter coalescent time for mitochondria may mean that the mitogenomic tree is more likely, on average, to match the species tree than a nuclear gene tree (e.g., Zink and Barrowclough, 2008), stochastic variation may lead to cases where the mitogenome differs (McKay and Zink, 2010). Unfortunately, our results provide no specific suggestion as to whether hybridization or lineage sorting may be more likely as an explanation for the observed discordances.

Overall, it is challenging to identify the source of cyto-nuclear discordance. However, a first step is to identify whether it is more likely due to inaccurate estimation or genuine discordance. Some cases of inaccurate estimation are easy to identify because they resolve upon addition of taxa or the use of improved models (e.g., Tamashiro et al., 2019). However, inaccurate estimation always remains a possible for apparent cyto-nuclear discordance even if the observed incongruence remains after improvements to taxon sampling and model fit. For example, it remains possible that complete taxon sampling (possibly also including extinct taxa) could remove at least some of these incongruences, as might use of even more complex models (once those are developed). Thus, in cases such as this, using an alternative approach to explore the data may be important before simply concluding genuine discordance exists. Indeed, an approach similar to the one we employed did provide evidence for inaccurate estimation driven by convergent evolution in squamates (Castoe et al., 2009). Thus, we view the examination of decisive sites as a tool that will help further refine our understanding of whether cyto-nuclear discordance exists in specific cases and allow us to make inferences regarding the basis for the discordance.

This approach could be applied in other groups to further our understanding of evolution and phylogenetic history of many groups of organisms as well as to better understand the limitations in the use of mitochondrial data. For example, our results highlight that, in groups like galliforms, over 5% of relationships identified in mitochondrial phylogenies may disagree with the species tree. Given that inaccurate estimation may lead to additional discordance in studies with sparse taxon sampling, insufficient models, or the use just one or a few mitochondrial regions, phylogenies based exclusively on mitochondrial data may exhibit an even greater disagreement with the underlying species tree than we observed here. Thus, while the majority of relationships are the same, it will still be important to revisit mitochondrial-only studies when feasible.

CRediT authorship contribution statement

Rebecca T. Kimball: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Marissa Guido:** Investigation, Data curation, Writing - review & editing. **Peter A. Hosner:** Investigation, Resources, Data curation, Writing - review & editing. **Edward L. Braun:** Conceptualization, Methodology, Resources, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank museums that provided tissues for our research (American Museum of Natural History, Australian National Wildlife Collection, Burke Museum of Natural History and Culture, Field Museum of Natural History, Florida Museum of Natural History, KU Natural History Museum, Louisiana State University Museum of Natural Sciences, Museum of Southwestern Biology, Sam Noble Oklahoma Museum of Natural History, Yale Peabody Museum, Zoological Museum Copenhagen); voucher information is included in Table S1. We also thank an anonymous reviewer for helpful comments that improved this manuscript.

Funding

This research was funded by grants from the US National Science Foundation to RTK and ELB (DEB-1118823 and DEB-1655683), with an REU supplement that supported MG. PAH acknowledges the support of VILLUM FONDEN for the Center for Global Mountain Biodiversity (grant no 25925)

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2021.145841.

References

Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13 (4), 729–744. https://doi.org/10.1046/j.1365-294X.2003.02063.x.

Barker, F.K., Oyler-McCance, S., Tomback, D.F., 2015. Blood from a turnip: tissue origin of low-coverage shotgun sequencing libraries affects recovery of mitogenome sequences. Mitochon. DNA 26 (3), 384–388. https://doi.org/10.3109/19401736.2013.840588.

Berlin, S., Ellegren, H., 2001. Evolutionary genetics. Clonal inheritance of avian mitochondrial DNA. Nature 413 (6851), 37–38. https://doi.org/10.1038/35092623.

Berlin, S., Smith, N.G.C., Ellegren, H., 2004. Do avian mitochondria recombine? J. Mol. Evol. 58 (2), 163–167. https://doi.org/10.1007/s00239-003-2537-z.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

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. https://doi.org/10.1080/10635150290102294.

Burleigh, J.G., Kimball, R.T., Braun, E.L., 2015. Building the avian tree of life using a large-scale, sparse supermatrix. Mol. Phylogenet. Evol. 84, 53–63. https://doi.org/10.1016/j.ympev.2014.12.003.

Cao, Y., Janke, A., Waddell, P.J., Westerman, M., Takenaka, O., Murata, S., Okada, N., Pääbo, S., Hasegawa, M., 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. J. Mol. Evol. 47 (3), 307–322. https://doi.org/10.1007/PL00006389.

Castoe, T.A., de Koning, A.P.J., Kim, H.-M., Gu, W., Noonan, B.P., Naylor, G., Jiang, Z.J., Parkinson, C.L., Pollock, D.D., 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. Proc. Natl. Acad. Sci. 106 (22), 8986–8991. https://doi.org/10.1073/pnas.0900233106.

Chernomor, O., von Haeseler, A., Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. 65 (6), 997–1008. https://doi.org/10.1093/sysbio/syw037.

Cummings, M.P., Meyer, A., 2005. Magic bullets and golden rules: data sampling in molecular phylogenetics. Zoology (Jena) 108 (4), 329–336. https://doi.org/10.1016/j.zool.2005.09.006.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2 (5), e68. https://doi.org/10.1371/journal.pgen.0020068.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32 (5), 1792–1797.

Evans, N.M., Holder, M.T., Barbeitos, M.S., Okamura, B., Cartwright, P., 2010. The phylogenetic position of Myxozoa: exploring conflicting signals in phylogenomic and ribosomal data sets. Mol. Biol. Evol. 27 (12), 2733–2746. https://doi.org/10.1093/molbev/msq159.

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 (8), e42543. https://doi.org/10.1371/journal.pone.0042543.

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. https://doi.org/10.1093/sysbio/sys004.

Gill, F., Donsker, D., Rasmussen, P., 2020. IOC World Bird List (v. 10.2). https://doi.org/10.14344/IOC.ML.10.2.

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. https://doi.org/10.1093/sysbio/syq010.

Hill, G.E., 2019. Reconciling the mitonuclear compatibility species concept with rampant mitochondrial introgression. Integr. Comp. Biol. 59, 912–924. https://doi.org/10.1093/icb/icz019.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. Ufboot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

Hosner, P.A., Braun, E.L., Kimball, R.T., 2015. Land connectivity changes and global cooling shaped the colonization history and diversification of New World quail (Aves: Galliformes: Odontophoridae). J. Biogeogr. 42 (10), 1883–1895. https://doi.org/10.1111/jbi.12555.

Hosner, P.A., Braun, E.L., Kimball, R.T., 2016a. Rapid and recent diversification of curassows, guans, and chachalacas (Galliformes: Cracidae) out of Mesoamerica: phylogeny inferred from mitochondrial, intron, and ultraconserved element sequences. Mol. Phylogenet. Evol. 102, 320–330. https://doi.org/10.1016/j.ympev.2016.06.006.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2016b. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. 33 (4), 1110–1125. https://doi.org/10.1093/molbev/msv347.

Hosner, P.A., Owens, H.L., Braun, E.L., Kimball, R.T., 2020. Phylogeny and diversification of the gallopheasants (Aves: Galliformes): Testing roles of sexual selection and environmental niche divergence. Zool. Scr. 10.1111/zsc.12441.

Hosner, P.A., Tobias, J.A., Braun, E.L., Kimball, R.T., 2017. How do seemingly non-vagile clades accomplish trans-marine dispersal? Trait and dispersal evolution in the landfowl (Aves: Galliformes). Proc. R. Soc. B 284, 20170210. https://doi.org/10.1098/rspb.2017.0210.

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., 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., Dixon, A., Li, S., Li, N., Huang, Y., 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., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., 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., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346 (6215), 1320–1331. https://doi.org/10.1126/science:1253451.

Kan, X.Z., Li, X.F., Lei, Z.P., Chen, L., Gao, H., Yang, Z.Y., Yang, J.K., Guo, Z.C., Yu, L., Zhang, L.Q., Qian, C.J., 2010a. Estimation of divergence times for major lineages of galliform birds: evidence from complete mitochondrial genome sequences. Afr. J. Biotechnol. 9, 3073–3078.

Kan, X.Z., Yang, J.K., Li, X.F., Chen, L., Lei, Z.P., Wang, M., Qian, C.J., Gao, H., Yang, Z.Y., 2010b. Phylogeny of major lineages of galliform birds (Aves: Galliformes) based on complete mitochondrial genomes. Genet. Mol. Res. 9, 1625–1633. https://doi.org/10.4238/vol9-3gmr898.

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 (3), 1021–1032. https://doi.org/10.1016/j.ympev.2013.05.029.

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 (16), 6196–6200. https://doi.org/10.1073/pnas.86.16.6196.

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 (6), 1695–1701. https://doi.org/10.1093/molbev/mss020.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773. https://doi.org/10.1093/molbev/msw260.

Lavrov, D.V., 2007. Key transitions in animal evolution: a mitochondrial DNA perspective. Int. Comp. Bio. 47, 734–743. https://doi.org/10.1093/icb/icm045.

Leavitt, J.R., Hiatt, K.D., Whiting, M.F., Song, H., 2013. Searching for the optimal data partitioning strategy in mitochondrial phylogenomics: a phylogeny of Acridoidea (Insecta: Orthoptera: Caelifera) as a case study. Mol. Phylogenet. Evol. 67 (2), 494–508. https://doi.org/10.1016/j.ympev.2013.02.019.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536. 10.1093/sysbio/46.3.523.

Maddison, W.P., Maddison, D.R., 2019. Mesquite: A modular system for evolutionary analysis. Version 3, 40. http://www.mesquiteproject.org.

McCarthy, E., 2006. Handbook of Avian Hybrids of the World. Oxford University Press, New York, NY.

McKay, B.D., Zink, R.M., 2010. The causes of mitochondrial DNA gene tree paraphyly in birds. Mol. Phylogenet. Evol. 54 (2), 647–650. https://doi.org/10.1016/j.ympev.2009.08.024.

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. https://doi.org/10.1016/j.ympev.2014.06.003.

Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T., Braun, E.L., 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. Syst. Biol. 65 (4), 612–627. https://doi.org/10.1093/sysbio/syw014.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). In: Presented at the 2010 Gateway Computing Environments Workshop (GCE), IEEE, pp. 1–8. https://doi.org/10.1109/GCE.2010.5676129.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. Mol. Biol. Evol. 15 (11), 1568–1571. https://doi.org/10.1093/oxfordjournals.molbev.a025884.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. 10.1093/molbev/msaa015.

Mitchell, K.J., Llamas, B., Soubrier, J., Rawlence, N.J., Worthy, T.H., Wood, J., Lee, M.S.Y., Cooper, A., 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. Science 344 (6186), 898–900. https://doi.org/10.1126/science:1251981.

Moore, W.S., 1995. Inferring phylogenies from MtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. Evolution 49, 718–726. https://doi.org/10.1111/j.1558-5646.1995.tb02308.x.

Pandey, A., Braun, E.L., 2020. Phylogenetic analyses of sites in different protein structural environments result in distinct placements of the Metazoan root. Biology 9, 64. https://doi.org/10.3390/biology9040064.

Persons, N.W., Hosner, P.A., Meiklejohn, K.A., Braun, E.L., Kimball, R.T., 2016. Sorting out relationships among the grouse and ptarmigan using intron, mitochondrial, and ultra-conserved element sequences. Mol. Phylogenet. Evol. 98, 123–132. https://doi.org/10.1016/j.ympev.2016.02.003.

Posada, D., 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256. https://doi.org/10.1093/molbev/msn083.
Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526 (7574), 569–573. https://doi.org/10.1038/nature15697.
Raposo do Amaral, F., Neves, L.G., Resende, M.F.R., Mobili, F., Miyaki, C.Y., Pellegrino, K.C.M., Biondo, C., Yue, G.H., 2015. Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. PLoS ONE 10 (9), e0138446. https://doi.org/10.1371/journal.pone.0138446.
Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53 (1-2), 131–147. https://doi.org/10.1016/0025-5564(81)90043-2.
Rubinoff, D., Holland, B.S., 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54, 952–961. 10.1080/10635150500234674.
Shen, X.X., Hittinger, C.T., Rokas, A., 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evol. 1, 1–10. https://doi.org/10.1038/s41559-017-0126.
Shen, Y.-Y., Dai, K., Cao, X., Murphy, R.W., Shen, X.-J., Zhang, Y.-P., Bai, Y., 2014. The updated phylogenies of the phasianidae based on combined data of nuclear and mitochondrial DNA. PLoS ONE 9 (4), e95786. https://doi.org/10.1371/journal.pone.0095786.
Shen, Y.-Y., Liang, L.u., 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 (1), 132. https://doi.org/10.1186/1471-2148-10-132.
Springer, M.S., Gatesy, J., 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94, 1–33. https://doi.org/10.1016/j.ympev.2015.07.018.
Springer, M.S., Gatesy, J., 2018. Delimiting coalescence genes (C-Genes) in phylogenomic data sets. Genes 9, 123. https://doi.org/10.3390/genes9030123.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
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. Proc. R. Soc. B 281, 20140823. https://doi.org/10.1098/rspb.2014.0823.
Tamashiro, R.A., White, N.D., Braun, M.J., Faircloth, B.C., Braun, E.L., Kimball, R.T., 2019. What are the roles of taxon sampling and model fit in tests of cyto-nuclear discordance using avian mitogenomic data? Mol. Phylogenet. Evol. 130, 132–142. https://doi.org/10.1016/j.ympev.2018.10.008.
Tiley, G.P., Pandey, A., Kimball, R.T., Braun, E.L., Burleigh, J.G., 2020. Whole genome phylogeny of *Gallus*: introgression and data-type effects. Avian Res. 11, 7. https://doi.org/10.1186/s40657-020-00194-w.
Wang, N., Hosner, P.A., Liang, B., Braun, E.L., Kimball, R.T., 2017. Historical relationships of three enigmatic phasianid genera (Aves: Galliformes) inferred using phylogenomic and mitogenomic data. Mol. Phylogenet. Evol. 109, 217–225. https://doi.org/10.1016/j.ympev.2017.01.006.
Zink, R.M., Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. Mol. Ecol. 17, 2107–2121. https://doi.org/10.1111/j.1365-294X.2008.03737.x.