



Incongruence among different mitochondrial regions: A case study using complete mitogenomes



Kelly A. Meiklejohn^{a,1}, Melany J. Danielson^{a,1}, Brant C. Faircloth^b, Travis C. Glenn^c, Edward L. Braun^a, Rebecca T. Kimball^{a,*}

^a Department of Biology, University of Florida, Gainesville, FL, United States

^b Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, United States

^c Department of Environmental Health Science, University of Georgia, Athens, GA, United States

ARTICLE INFO

Article history:

Received 17 February 2014

Revised 31 May 2014

Accepted 2 June 2014

Available online 12 June 2014

Keywords:

Galliformes
Codon models
Mixture models
Noise reduction
Taxon sampling

ABSTRACT

Mitochondrial sequences have long been used to examine vertebrate phylogenetic relationships. The extensive use of mitochondrial data reflects the ease of obtaining mitochondrial sequences and its relatively rapid coalescence time. Mitochondrial genomes typically do not undergo recombination, so the entire mitogenome should have the same underlying gene tree. Thus, given appropriate analyses, conflict among estimates of phylogeny from different mitochondrial regions should not exist. However, estimates of phylogeny based upon different mitochondrial regions can exhibit incongruence. Conflict in phylogenetic signal among mitochondrial regions has been observed in galliform birds for the position of the Odontophoridae (New World quail). To explore this, we expanded sampling to 47 galliform mitogenomes, adding six new mitogenomes, which included representatives of two previously unsampled families. Analyses of complete mitogenomes recovered a well-supported topology that was congruent with expectations from multi-locus studies. However, when analyzing individual regions, we found conflicting positions for the Odontophoridae and several other relationships at multiple taxonomic levels. We tested multiple analytical strategies to reduce incongruence among regions, including partitioning by codon position, using mixture and codon-based models, RY coding, and excluding potentially misleading sites. No approach consistently reduced the conflict among mitochondrial regions at any taxonomic level. The biological attributes of both strongly misleading and non-misleading sites were essentially identical. Increasing taxa actually appeared to increase conflicting signal, even when taxa were selected to break up long branches. Collectively, our results indicate that analyzing mitochondrial data remains difficult, although analyzing complete mitogenomes resulted in a good estimate of the mitochondrial gene tree.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Mitochondrial sequence data has been used extensively in molecular phylogenetics due to its rapid evolutionary rate (Boore, 1999), short coalescence time (Moore, 1997) and the availability of conserved primers to amplify a variety of regions (Kocher et al., 1989; Sorenson et al., 1999). The mitogenome accumulates substitutions more rapidly than the nuclear genome (e.g., Armstrong et al., 2001), leading to a higher probability that synapomorphic substitutions will be present on short internal branches with mitochondrial rather than nuclear data (cf. Braun and Kimball, 2001).

Thus, mitochondrial data may have a strong influence on phylogenetic analyses of concatenated data. The mitochondrial tree is also likely to be better resolved than nuclear gene trees and provide valuable information for coalescent-based methods of species tree estimation. Indeed, including a mitochondrial partition often leads to improved resolution for species tree analyses (Corl and Ellegren, 2013; Sánchez-García and Castresana, 2012). Therefore, mitochondrial data still has the potential to play an important role in multi-locus studies and thus it remains critical to identify analytical approaches that provide the best estimates of the mitochondrial gene tree.

Vertebrate mitogenomes are small (~16–20 kb) circular DNA molecules that do not appear to undergo recombination (Berlin and Ellegren, 2001; Berlin et al., 2004; Boore, 1999; Moritz et al., 1987; but see Sammler et al., 2011 for an example in another avian order). Thus, it seems reasonable to postulate all mitochondrial

* Corresponding author. Address: Department of Biology, P.O. Box 118525, University of Florida, Gainesville, FL 32611, United States.

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

¹ Equal first authorship contributions.

regions are expected have the same evolutionary history and, within the limits of sampling error, should reconstruct the same phylogenetic relationships. However, conflict among mitochondrial regions has been noted for relationships among major tetrapod groups (e.g., Cao et al., 2000; Russo et al., 1996), within bats (Botero-Castro et al., 2013), squamates (Castoe et al., 2009), amphibians (Weisrock et al., 2005; Zhang et al., 2013), and galliform birds (Cox et al., 2007). This suggests that some mitochondrial regions may accumulate substitutions in ways that are difficult to model, resulting in biased estimates of phylogeny.

When incongruence within the mitochondria has been noted, the basis for this conflict has typically remained unclear. However, there are several analytical strategies that can improve phylogenetic estimation with mitochondrial data. First, use of better-fitting models, such as those designed for coding data or that account for heterogeneity in a more biologically realistic manner (e.g., Braun and Kimball, 2002; Powell et al., 2013) can help. However, it may be difficult to estimate large numbers of parameters that may be associated with these biologically realistic models (Steel, 2005). Second, reduction of noise that may obscure evolutionary signal, using strategies such as RY coding (e.g., Braun and Kimball, 2002; Phillips et al., 2010; Pratt et al., 2009) or the exclusion of potentially misleading sites (e.g., Pratt et al., 2009) can be important. Finally, increased taxon sampling has the potential to limit the impact of long branch attraction and it is thought to improve phylogenetic estimation (e.g., Lin et al., 2002). Given that studies in vertebrate systematics are likely to continue using mitochondrial sequence data, it is important to comprehensively explore these different approaches to identify robust methods for alleviating signal incongruence.

The avian order Galliformes (chickens, turkey, quail, and allies) represents a good model for examining the evolutionary signal of the mitochondria, given that conflicts among regions are documented. Cox et al. (2007), using multiple methods, found that 12S rRNA supported a phylogeny consistent with other types of data (placing Odontophoridae [New World quail] sister to the Phasianidae [chickens, turkeys and allies]; Fig. 1A), ND2 supported a rearrangement among two of the five galliform families (Numididae [Guineafowl] sister to all other phasianids; Fig. 1B), and cytochrome-*b* (CYB) was equivocal. Assessing the signal from complete mitochondrial genomes (mitogenomes) for the placement of the Odontophoridae has not been possible to date, as previous studies have largely focused on the most species-rich family (Phasianidae) (Kan et al., 2010; Shen et al., 2010) and mitogenomes from two of the five galliform families (Odontophoridae and Cracidae [guans, currasows, and chachalacas]) have been lacking.

Here we analyze 47 complete mitochondria from galliforms, including representatives from the two galliform families (Odontophoridae and Cracidae) previously lacking mitogenomes. This taxon set also included mitogenome sequences for several other galliform species that are expected to break up long branches to clades involved in the rearrangements seen by Cox et al. (2007). We tried a range of analytical approaches to explore questions of model adequacy, misleading signal and taxon sampling, to gain a better understanding of issues and challenges with reconstructing phylogenies using mitochondrial sequence data. Specifically we ask: (1) is the estimate of the mitochondrial gene tree based upon the complete mitogenome congruent with the results of analyses using other types of data?; (2) is there conflict among different mitochondrial regions regarding the placement of the Odontophoridae, as observed by Cox et al. (2007)?; (3) does taxon sampling have an impact upon phylogenetic estimation using mitochondrial data?; (4) are there conflicts among mitochondrial regions at other taxonomic levels within the Galliformes?; and (5) do commonly implemented analytical strategies reduce incongruence among mitochondrial regions?.

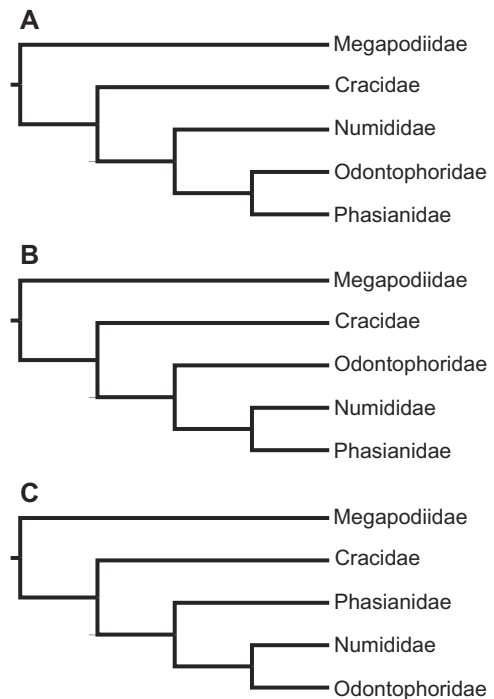


Fig. 1. Different hypotheses for the placement of the Odontophoridae within Galliformes. (A) Odontophoridae derived; (B) Odontophoridae basal in relation to the Numididae; (C) Odontophoridae sister to the Numididae.

2. Methods

2.1. Data collection

In this study we used complete mitogenomes from 47 galliform species, representing all five families (Supplementary Material Table 1). We obtained 41 galliform genomes from GenBank, along with those of five outgroups (Anseriformes [ducks and geese]) (Supplementary Material Table 1). From off-target reads collected from sequence capture experiments run on an Illumina HiSeq 2000 (Illumina Inc., San Diego, CA), we assembled sequences for six additional galliforms. The pipeline that we used to assemble this data included Velvet (Zerbino and Birney, 2008) and local BLASTN searches (Camacho et al., 2009). We used the chicken mitogenome (Desjardins and Morais, 1990) as a reference to identify assemblies that included mitochondrial DNA in BLASTN searches. Among the mitochondria sequences we found from the Illumina data, we identified four genomes that were completely sampled (*Crax rubra*, *C. daubentoni*, *Colinus virginianus* and *Ptilopachus petrosus*) however the remaining two (*Caloperdix ocella* and *Rhynchortyx cinctus*) had small gaps. We filled most of these gaps by amplifying the relevant regions of the mitogenome using previously published primers and protocols (Sorenson et al., 1999). We used agarose gel electrophoresis (1%) to resolve PCR products, with samples containing an amplicon of the expected size cleaned by precipitation using an equal volume of PEG (10%):NaCl (2.5 M). PEG cleaned products were sent to the ICBR (University of Florida, Gainesville, FL) for Sanger sequencing.

2.2. Sequence alignment

To identify gene boundaries and check for possible rearrangements, we aligned all mitogenomes to the annotated chicken mitogenome in Geneious (Version 6.1.2, Biomatters Ltd., 2013). From the mitogenome alignment of all taxa, we extracted individual alignments of the 13 protein coding genes, two rRNAs, and 22 tRNAs. To address concerns surrounding the inclusion of NUMTs in our

analyses, we checked individual protein-coding gene alignments for insertion–deletion (indel) events and premature stop codons using MacClade (Version 4.08; Maddison and Maddison, 2005). Alignments of both the 12S and the 16S rRNAs were also optimized by eye in MacClade. Due to difficulty of aligning the mitochondrial control region across all galliforms (and the anseriform outgroups), we did not include the control region or any of the other short non-coding regions in the analyses described below.

For published data, we used the data and annotations provided in GenBank with a few exceptions. The ND3 sequences of some species (including the new mitogenomes we collected) contain a programmed frameshift (Mindell et al., 1998; Russell and Beckenbach, 2008); this site was excluded from our alignments. Upon checking the ND1 alignment, the published sequence for *Acryllium vulturinum* (Numididae) from Shen et al. (2010) appeared to be highly divergent from the *Numida meleagris* sequence (the other sequence from a member of Numididae). This variation was not restricted to third codon positions and, consistent with the divergent nature of this sequence, ND1 did not support monophyly of Numididae in our preliminary analyses or in the analyses that Shen et al. (2010) presented in their supporting material. We obtained a complete *A. vulturinum* ND1 sequence from our Illumina data and found that our sequence exhibited only 89.7% to the Shen et al. (2010) sequence. However, our ND1 sequence was more similar to the *N. meleagris* sequence than the Shen et al. (2010) sequence (90.3% and 86.6%, respectively) and preliminary phylogenetic analyses found a monophyletic Numididae clade when our sequence was used. Based upon these observations, we replaced the Shen et al. (2010) ND1 sequence for *A. vulturinum* with the one we obtained. No other regions of the published *A. vulturinum* mitogenome appeared unusually divergent.

2.3. Taxon sets

In this study, we used several taxon sets to assess whether evolutionary signal from individual regions and the mitogenome were robust to taxon sampling. We established two main taxon sets (Supplementary Material Table 1): (1) the ‘entire’ taxon set, which included all 47 galliform species with multiple representatives from some genera and (2) a ‘reduced’ taxon set, which was matched as closely as possible to that of Cox et al. (2007) ($N = 13$). We established this ‘reduced’ taxon set to facilitate a more direct comparison to the results of Cox et al. (2007), that revealed conflicting signal between mitochondrial regions in the placement of the Odontophoridae.

Cox et al. (2007) lacked representatives from one major clade within the Phasianidae, the Arborophilinae (the deepest divergence within Phasianidae is between Arborophilinae and all other phasianids, the latter designated ‘core’ Phasianidae by Kimball et al., 2011). To determine whether including a member of the Arborophilinae had an impact on phylogenetic signal by breaking up the branch to the phasianids, we added *Caloperdix* and *Arborophila* species, creating a ‘reduced + arbor’ taxon set. Several recent studies have suggested the African genus *Ptilopachus*, traditionally considered a phasianid, is instead an early diverging Odontophoridae (Cohen et al., 2012; Crowe et al., 2006; Kimball et al., 2011). Thus, including *Ptilopachus* should break up the branch leading to the traditional Odontophoridae. Therefore, we added *P. petrosus* to both the ‘reduced’ and ‘reduced + arbor’ taxon sets, creating ‘reduced + *Ptilopachus*’ and ‘reduced + arbor + *Ptilopachus*’ taxon sets, to explore the impact of this taxon on phylogenetic reconstruction (Supplementary Material Table 1).

We used all five anseriform species for which mitogenomes were available on GenBank at the time of this study as outgroups for each of these five taxon sets (Supplementary Material Table 1). However, for the ‘entire’ and ‘reduced’ taxon sets, we performed

additional sets of analyses without these anseriform outgroups to better compare with Cox et al. (2007) and to evaluate sensitivity to outgroup selection. In these analyses, we rooted the resulting topologies to *Alectura lathamii* (Megapodiidae) since several studies (e.g., Hackett et al., 2008; Wang et al., 2013) indicated that the deepest divergence within Galliformes is between Megapodiidae and all other taxa in the order.

2.4. Phylogenetic analyses

We conducted maximum likelihood (ML) analyses using the GUI interface (Silvestro and Michalak, 2012) of RAxML 7.3.2 (Stamatakis, 2006) under the GTRGAMMA evolutionary model (the general time reversible model with Γ -distributed rate variation among sites). To determine the best likelihood for subsequent AIC/AICc calculations, we performed an initial ML search of 10 replicates. Following this, we performed a 500 replicate bootstrap analysis, with a majority rule consensus tree generated using CONSENSE from the PHYLIP package (Felsenstein, 1989). For all taxon sets, we carried out separate unpartitioned ML analyses for each protein-coding gene, allowing for the assessment of individual evolutionary signal. We partitioned the rRNAs into stems and loops. For the 12S rRNA, we based this on the structure of *Gallus gallus* obtained from the WebRNA server (Cannone et al., 2002) while the 16S rRNA structure was based upon that published by Amaral et al. (2010). tRNA structures were based upon Desjardins and Morais (1990) and we also conducted further partitioning by strand (light or heavy). As the sequences for individual tRNAs were short (~70 bp each), we concatenated the sequences for the 22 tRNAs together and analyzed them as a single dataset.

To evaluate the signal of the mitogenome, we combined the rRNAs, tRNAs and protein coding genes and performed a series of ML analyses for all datasets using eight *a priori* defined partitioning schemes: (1) unpartitioned; (2) partitioned into rRNAs, tRNAs and protein coding genes; (3) same as 2, but with further partitioning by codon position for protein coding regions; (4) partitioned into rRNAs, tRNAs, ND6 (since this is the only protein coding gene encoded on the heavy strand) and remaining protein coding genes; (5) same as 4, but with further partitioning by codon position for protein coding regions; (6) partitioned by region (i.e., separate partitions for the 12S rRNA, the 16S rRNA, the tRNAs, and for each of the protein coding genes); (7) same as 6, but with further partitioning by codon position for protein coding regions; and (8) same as 7, but also partitioning the rRNAs into stems and loops, and the tRNAs into those on the light versus heavy strand. We also used the approach recommended by Powell et al. (2013) and ran a principal components analysis on the log-transformed model parameters for each of the partitions defined in the last scheme, and clustered the principal components to define new partitions based upon regions with similar parameter estimates. From this, we compared five additional partitioning schemes that used different levels of clustering (Supplementary Material Fig. 1 and Supplementary Material Table 2).

2.5. Comparing analytical approaches to resolve conflict among mitochondrial regions

For the two focal taxon sets, the ‘entire’ and ‘reduced’, we employed a range of commonly implemented analytical strategies to assess whether data partitioning and model choice affected phylogenetic estimation. First, we partitioned protein-coding genes by codon position and re-estimated the phylogenetic relationships, since there can be substantial differences among codon positions for mitochondrial genes (Boore, 1999; Moritz et al., 1987). Expanding on this, we also analyzed protein-coding regions using the codon models implemented in GARLI 2.0 (Zwickl, 2006). We examined a set of six models, all of which used codon frequencies

calculated by the $F3 \times 4$ method (codon frequencies reflect the base frequencies at each codon position). The models were two substitution type (HKY-like) and six substitution type (GTR-like) with one, three, or four dN/dS ratios. To determine the best-fitting codon model for each region, we initially performed an ML search of 10 replicates using all six of these models and identified the best model using the AIC. Then we performed a 500 replicate bootstrap analysis and generated a majority rule consensus tree using CONSENSE. All of analyses were run on the CIPRES Science Gateway (Miller et al., 2010). Second, we RY coded protein-coding genes at third codon positions (using a Perl script written by KAM; available on request), to assess the impact of possible saturation biases. We then performed analyses of the individual genes and whole mitogenomes using these RY coded alignments. Third, to account for among-site heterogeneity in evolutionary processes, we used the CAT–GTR mixture model implemented in PHYLOBAYES (Version 3.3; Lartillot and Philippe, 2004) to analyze all individual regions. For all analyses, we used two independent chains run for 5000 points each. We discarded the first 10% of trees from each run as burn-in and a majority-rule consensus tree was generated, using CONSENSE, with the remaining trees.

To determine if rapidly evolving sites provided misleading evolutionary signal, we used two approaches to identify and exclude potentially misleading sites: observed variability (OV) (tree-independent) and evolutionary rate (tree-dependent). We calculated the OV of each site in the alignment by the method outlined in Goremykin et al. (2010), using a Perl script generously provided by Gordon Burleigh. We utilized BASEML from the PAML package (Version 4.7; Yang, 2007) to calculate the evolutionary rate for each site in the alignment, using a partitioned input file (strategy 7) and the best ML tree for this partitioning strategy. We implemented the default settings for BASEML except Malpha was set to 1 and Mgene was set to 4. For both methods, we sorted sites in ascending order based on OV and evolutionary rate, with the highest 5%, 8% and 10% of sites across the entire alignment identified for each method. We then excluded these sites from the separate alignments for each individual region using PAUP* (Version 4.0a128; Swofford, 2002), with the remaining sites retained for analysis. For both exclusion methods, we performed additional 500 replicate bootstrap ML analyses for each region as described above. It is important to note that we only excluded such 'noisy' sites for the 'entire' taxon set, due to concerns with underestimating the OV and evolutionary rate with the smaller 'reduced' taxon set (Tourasse and Gouy, 1997).

Biological attributes of a site can also be used to identifying misleading signal, so we initially determined the number of 'decisive' sites (e.g., Kimball et al., 2013) supporting the expected placement of the Odontophoridae and the most plausible alternative (Fig. 1A and B, respectively). These 'decisive' sites were identified based on the magnitude of difference in the per site log-likelihoods ($\ln L$) for the alternative topologies. These per-site $\ln L$ values were calculated in GARLI for the mitogenome alignment from the 'entire' taxon set. We determined the numbers of both non-synonymous and synonymous 'decisive' sites and established the biochemical properties of such sites. To do this, a Perl script written by KAM was used to recode each amino acid residue in each translated protein coding alignment as either hydrophobic (L, F, I, C, W, V, Y, M) or polar (A, H, T, G, P, D, E, Q, R, S, N, K), based on the Wang and Wang (1999) groupings of amino acids. Using the recoded alignment, a consensus amino acid sequence was then obtained for each gene, and subsequently used to assign each 'decisive' site to one of three categories (hydrophobic, polar, or mixed).

2.6. Comparing the power of different mitochondrial gene regions

To assess the overall performance of the different mitochondrial gene regions, we estimated the ML tree using RAXML for each

mitochondrial region (for protein coding genes, this was partitioned by codon position). The Robinson–Foulds distance (Robinson and Foulds, 1981) between the optimal tree estimated from the mitogenome (all protein-coding and RNA regions) and the trees obtained from each individual region was obtained. This was compared with the numbers of parsimony informative sites in each region.

2.7. Comparing the galliform species tree to the mitochondrial gene tree

The goal of this study is the analysis of incongruence among mitochondrial regions, which are expected to have the same gene tree (see Section 1), rather than a direct comparison between the mitochondrial gene tree and the species tree. However, we also wanted to establish the topology of the galliform species tree to examine the congruence of the mitochondrial tree for a few specific nodes (see Section 3). Given the more rapid coalescent of the mitochondrial genome the true mitogenomic tree is likely to be congruent with the species tree for clades defined by long branches. Thus, a comparison of the signal on long branches between mitochondrial and nuclear data should be informative. The estimates of individual gene trees are the most important data for this comparison, since long branches in the species tree (where the true mitochondrial tree should be congruent) are expected to be present in the majority of gene trees (Degnan et al., 2009). However, many prior analyses of galliform phylogeny have used concatenation and did not include estimates of individual loci. Therefore, we also estimated individual gene trees for loci used by Hackett et al. (2008). Briefly, data for all galliforms in that study and four anseriform outgroups were exported from the alignment used by Hackett et al. (2008), posted at <http://www.biology.ufl.edu/earlybird>. Then we used RAXML to conduct a bootstrap analysis (with 500 replicates) of each locus using the GTRGAMMA model.

3. Results and discussion

Analysis of the complete mitogenome (protein coding regions, rRNAs, and tRNAs) resulted in a phylogeny that has high bootstrap support for many nodes (Fig. 2). The different partitioning schemes yielded similar topologies and levels of bootstrap support (Supplementary Material Table 2). The best-fitting partitioning scheme corresponded to our 'fully partitioned' (mitogenome 8) model (Supplementary Material Table 2), which was partitioned by gene region, by codon position within the coding genes, and by stems and loops within the RNAs (with additional partitioning of tRNAs into those encoded by the heavy and light strands). However, we note that two models of similar complexity, mitogenome 7 and Powell Method A (based on the Powell et al., 2013) method, yielded an identical topology with similar levels of support. We used this topology as our reference estimate of the mitochondrial gene tree (hereafter called the 'mitogenomic tree') for comparison to the trees generated by analyses of individual mitochondrial regions. As expected, given the rapid coalescent of the mitochondrial genome, this mitogenomic tree is highly congruent to recent multi-locus (multiple nuclear and mitochondrial regions) studies of galliforms (Cohen et al., 2012; Kimball and Braun, 2008, 2014; Kimball et al., 2011; Wang et al., 2013) wherever taxon sampling permitted comparisons.

Despite the expectation that mitochondrial data should have greater power to resolve relationships than nuclear data, we found two nodes that have been strongly supported in recent multi-locus studies (e.g., Cox et al., 2007; Hackett et al., 2008; Kimball and Braun, 2008, 2014; Kimball et al., 2011; Wang et al., 2013) that were present, but with low support, in the mitogenomic tree. The first of

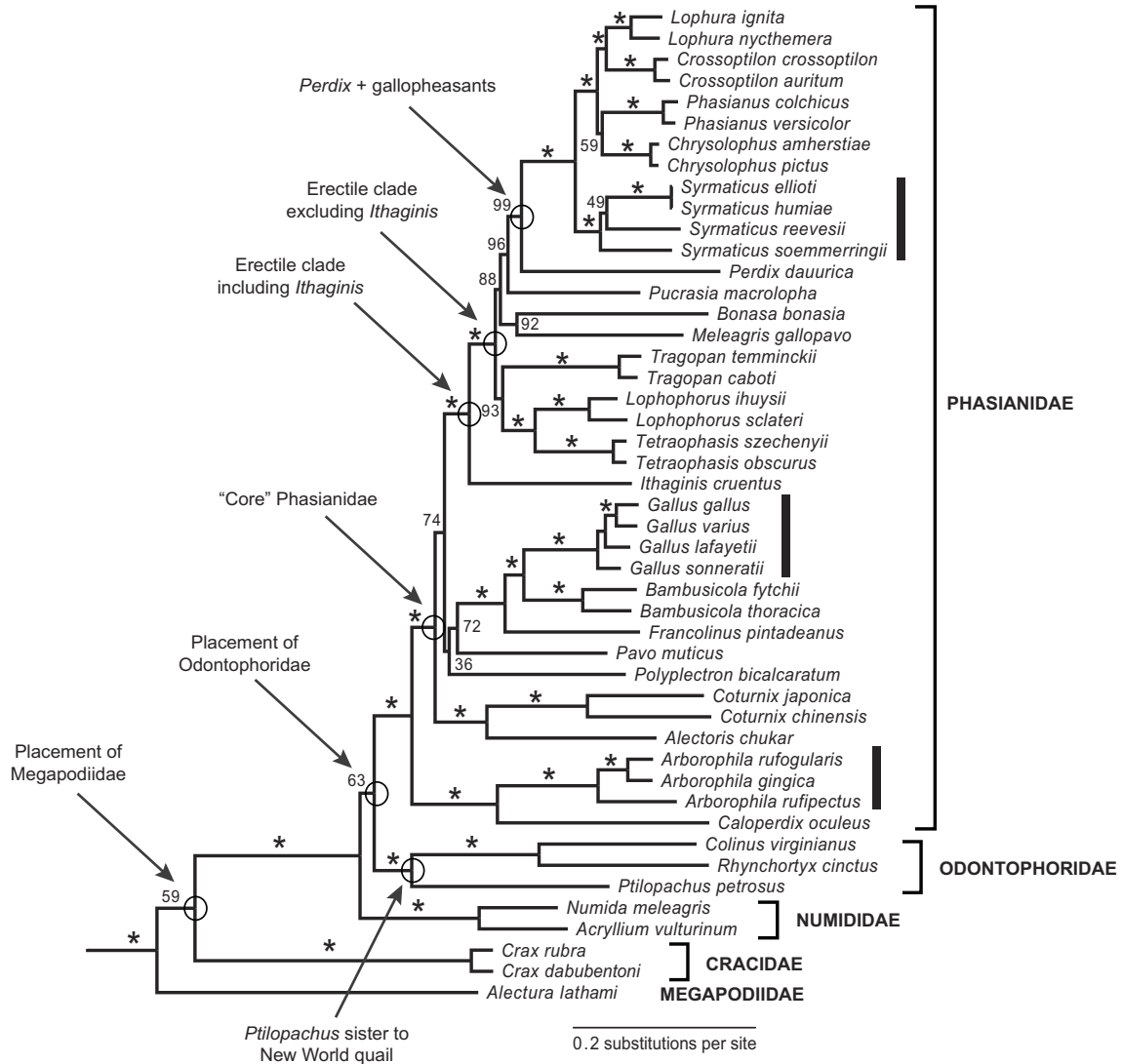


Fig. 2. Maximum likelihood (ML) phylogeny of the mitogenome for the 'entire' taxon set. This analysis was conducted with the partitioned by region and further partitioned based upon codon position for protein coding genes (the optimal partitioning strategy based upon the AIC; this corresponds to strategy 7 in Table 1). Numbers at nodes represent ML bootstrap support as a percentage; * denotes 100%. Nodes used to assess the performance of various analyses are circled (cases where relationships among genera were evaluated) or highlighted using thick black bars on the right hand side (cases where relationships within genera were evaluated).

these relatively poorly supported nodes defined the clade comprising Odontophoridae plus Phasianidae whereas the other reflected the position of the root for the order, which is between Megapodiidae (mound builders) and all other galliforms in the mitogenomic tree. We found that these relationships were generally recovered in analyses of individual nuclear loci (Supplementary Material Table 3; see also Cox et al., 2007) despite the limited power (reflecting the lower rates of evolution and the more limited number of sites) of those individual nuclear loci. Indeed, more than 75% of the genes analyzed supported each clade (Table S2). This result strongly suggests that both nodes are present in the galliform species tree because, when a number of gene trees are sampled, those clades present in more than half of gene trees are expected to be present in the species tree (Degnan et al., 2009). Given the rapid coalescent of the mitochondrial genome, and the observation that the mitogenomic tree also includes these clades (albeit with less than 100% support) suggests that those two specific clades are very likely to be present in the true mitochondrial gene tree. Thus, the limited support for these two nodes even when large amounts of mitochondrial data are analyzed suggests that there are conflicting signals within the mitogenome.

3.1. Position of the Odontophoridae

Based on conflict among mitochondrial regions in Cox et al. (2007) and the limited bootstrap support for this clade in the mitogenomic tree, we examined the placement of the Odontophoridae by all individual mitochondrial regions, and the mitogenome, across a range of taxon sets and analytical approaches. When analyzing sequences from all protein coding genes, rRNAs and tRNAs together (Fig. 2), the position of the Odontophoridae matched Fig. 1A, though the bootstrap support varied among partitioning strategies and taxon sampling (Table 1).

3.1.1. Varying positions among mitochondrial regions

The position of the Odontophoridae found in analyses of the mitogenome (Fig. 1A) was only recovered in analyses of the five taxon sets by the protein coding genes ND3, ND4 and ND5, and even then it was sometimes with poor support (<50% bootstrap support; Table 1). Interestingly, CYB and ND2, regions commonly used in avian phylogenetics, differed from each other, though within a region, there was a consistent topology obtained across all taxon sets (Fig. 1B and C respectively; Table 1), in agreement with other studies

Table 1

Placement of the Odontophoridae for the five taxon sets for each mitochondrial region and the different partitioned analyses of the mitogenome.^a Values in brackets are ML bootstrap percentages. Bold terms indicate the topology (A) expected based on the mitogenome.

	Entire	Reduced	Reduced + <i>Ptilopachus</i>	Reduced + arbor	Reduced + arbor + <i>Ptilopachus</i>
ATP6	*C (45)	A (56)	*C (32)	A (52)	*C (29)
ATP8	O	*B (28)	*B (23)	*B (26)	*B (27)
CYB	*B (57)	B (82)	*B (61)	B (63)	*B (50)
COI	*B (24)	* A (38)	*C (52)	O	C (40)
COII	B (32)	B (47)	B (26)	B (45)	B (27)
COIII	O	A (56)	O	C (39)	O
ND1	C (60)	C (38)	C (59)	A (33)	C (46)
ND2	*C (43)	C (52)	*C (45)	C (41)	*C (33)
ND3	A (35)	A (25)	A (31)	A (23)	A (30)
ND4	* A (48)	A (92)	A (54)	A (80)	A (45)
ND4L	*C (3)	B (53)	B (28)	B (36)	B (16))
ND5	A (58)	A (64)	A (43)	A (60)	A (44)
ND6	O	O	*B (23)	O	O
tRNAs	A (34)	C (45)	B (31)	O	O
12S	O	A (48)	A (36)	O	O
16S	*C (47)	C (72)	*C (67)	C (46)	*C (50)
Mitogenome 8 ^b	A (63)	A (57)	A (67)	A (65)	A (70)

^a Letters refer to topology in Fig. 1 ('O' denotes placement other than depicted in Fig. 1).

^b Best partitioning scheme for mitogenome analyses.

* The topology was similar to that noted (e.g., *Ptilopachus* clustered within rather than sister to the Odontophoridae).

using those regions (e.g., Cox et al., 2007; Crowe et al., 2006; Kimball et al., 1999). The remaining mitochondrial regions failed to recover the same topology across all taxon sets, with some even resolving a topology other than those shown in Fig. 1 (denoted by 'O' in Table 1). In contrast to the expectation that increased taxon sampling (in particular the addition of taxa that break up long branches, as done in this study), individual mitochondrial regions using the reduced taxon set found the expected placement of Odontophoridae more often than did any of the more taxon-rich data sets.

When we removed the anseriform outgroups from the 'entire' and 'reduced' taxon sets and rooted analyses to *A. lathamii* (Megapodiidae), there was little effect on the placement of the Odontophoridae or bootstrap support (Table 2). For the few cases where removing the anseriform outgroups resulted in changing the position of the Odontophoridae to match the mitogenome topology, these were not consistent across genes for both taxon sets (Table 2). In some cases the mitogenomic placement of the Odontophoridae was obtained when using the anseriform outgroups but an alternate topology was obtained when rooting to *A. lathamii*.

3.2. Examining conflicting signal at various taxonomic levels

To explore incongruence at other nodes, we examined relationships at three broad levels: among families, within phasianids, and within genera. The relationships examined are shown in Fig. 2 and Table 3 provides an overall performance assessment of the individual gene regions across taxonomic levels.

Across all five taxon sets, CYB, COIII, ND2 and ND5 recovered Megapodiidae as the most basal galliform lineage (Supplementary Material Table 4), in agreement with the mitogenome (Fig. 2). Interestingly, Megapodiidae and Cracidae formed a clade (see also Sibley and Ahlquist, 1990) using most of the other mitochondrial regions (Supplementary Material Table 4) and in all but the best-fitting partitioning strategies for the mitogenomic analyses. This could reflect the narrow taxon sampling of these families. Although we added two sequences from the previously unsampled family Cracidae, sequencing more members of Cracidae and Megapodiidae will be necessary in order to determine whether this incongruence reflects

Table 2

Placement of the Odontophoridae for analyses of the Entire and Reduced taxon sets for various analyses.^a Values in brackets are ML bootstrap percentages or Bayesian posterior probabilities (for PHYLOBAYES). To allow comparison, the "Basic" column is the same as "Entire" from Table 1.

	Entire taxon set						Reduced taxon set				
	Basic	No outgroup	Coding	RY	PHYLO-BAYES	GARLI codon	Basic	No outgroup	Coding	RY	PHYLO-BAYES
ATP6	*C (45)	*C (56)	*C (46)	*C (49)	*C (0.62)	*C (35)	A (56)	A (50)	A (59)	A (55)	A (0.61)
ATP8	O	O	O	O	O	O	*B (28)	O	O	O	O
CYB	*B (57)	A (50)	*B (75)	*B (64)	*B (0.61)	*B (72)	B (82)	B (50)	B (81)	B (64)	B (0.59)
COI	*B (24)	*B (38)	C (32)	*B (20)	*C (0.40)	B (30)	* A (38)	* A (41)	* A (49)	A (40)	A (0.22)
COII	B (32)	C (49)	B (36)	B (59)	B (0.42)	B (39)	B (47)	C (29)	B (60)	B (55)	B (0.69)
COIII	O	O	O	O	O	O	A (56)	A (37)	A (51)	A (49)	A (0.67)
ND1	C (60)	C (25)	C (52)	C (67)	C (0.70)	A (42)	C (38)	A (64)	A (53)	C (38)	A (0.62)
ND2	*C (43)	* A (46)	* A (50)	*C (46)	* A (0.38)	A (36)	C (52)	C (46)	C (47)	C (44)	C (0.62)
ND3	A (35)	O	A (36)	A (38)	A (0.80)	A (20)	A (25)	A (30)	A (22)	A (28)	A (0.64)
ND4	* A (48)	* A (44)	A (40)	*A (39)	A (0.44)	A (72)	A (92)	A (89)	A (80)	A (84)	A (0.86)
ND4L	*C (3)	B (22)	B (13)	B (31)	O	O	B (53)	B (53)	B (55)	B (40)	B (0.61)
ND5	A (58)	A (38)	A (58)	A (57)	A (0.49)	A (45)	A (64)	B (49)	A (63)	A (61)	A (0.44)
ND6	O	O	O	O	*B (0.67)	O	O	B (30)	B (33)	O	B (0.47)
tRNAs	A (34)	A (37)	–	–	A (0.72)	–	C (45)	A (53)	–	–	A (0.48)
12S	O	O	–	–	O	–	A (48)	A (39)	–	–	A (0.48)
16S	*C (47)	*C (52)	–	–	*C (0.88)	–	C (72)	C (58)	–	–	B (0.42)
Mitogenome 8	A (63)	A (70)	–	A (63)	–	–	A (57)	A (69)	–	A (60)	–

^a Letters match the topologies in Fig. 1 and Table 1. Analyses conducted are: (1) basic (GTRGAMMA in RAxML; only RNAs were partitioned); (2) no outgroup (anseriform outgroups excluded); (3) coding (partitioning by codon position in RAxML); (4) RY (third codon positions RY coded in RAxML); (5) PHYLOBAYES (GTR-CAT analyses in Phylobayes); and (6) GARLI codon (codon models implemented in GARLI 2.0).

* The topology was similar to that noted (e.g., *Ptilopachus* clustered within rather than sister to the Odontophoridae).

Table 3
Overall performance of mitochondrial regions and the mitogenome at varying taxonomic levels.

	Among families			Within Phasianidae				Within genera		
	Position of Odontophoridae	Megapodes basal	<i>Ptilopachus</i> + Odontophoridae	“core” phasianids	Erectile including <i>Ithaginis</i>	Erectile excluding <i>Ithaginis</i>	<i>Perdix</i> + gallopheasants	<i>Arborophila</i>	<i>Gallus</i>	<i>Syrmaticus</i>
ATP6	M	M	P	G	M	P	G	G	P	P
ATP8	P	M	P	P	P	P	P	G	M	P
CYB	P	G	P	G	G	G	G	G	M	P
COI	M	P	M	P	G	G	P	G	P	P
COII	P	P	G	G	M	M	P	G	P	P
COIII	M	G	G	P	P	P	P	G	P	P
ND1	P	P	G	P	M	M	P	P	P	M
ND2	P	G	P	G	G	G	M	M	M	P
ND3	G	P	G	P	P	P	M	P	M	M
ND4	G	P	M	G	G	G	M	G	P	P
ND4L	P	P	G	P	P	P	M	M	P	P
ND5	G	G	G	G	G	G	G	G	G	M
ND6	P	P	P	P	P	P	P	G	P	M
tRNAs	P	P	G	G	G	G	G	G	P	G
12S	P	P	G	G	P	P	P	M	P	P
16S	P	P	M	G	G	G	P	M	G	P
Mitogenome	G	P	G	G	G	G	G	G	–	–

‘G’ denotes good (bold; expected relationship was recovered in all analyses); ‘M’ denotes moderate (expected relationship was recovered in $\geq 50\%$ but $< 100\%$ of analyses); and ‘P’ denotes poor (expected relationship was recovered in $\leq 50\%$ of analyses). There is no consensus in the literature regarding the phylogeny within *Arborophila*, *Gallus* and *Syrmaticus*, so we considered the best estimate for those genera to be the topology in analyses of the mitogenome (Fig. 2).

taxon sampling or difficulties associated with modeling the molecular evolution of the mitogenome.

The mitogenomic analysis placed the African genus *Ptilopachus*, traditionally classified as a member of Phasianidae, as the earliest diverging Odontophoridae with high support (Fig. 2). This result is consistent with recent multi-locus studies (Cohen et al., 2012; Crowe et al., 2006; Kimball et al., 2011). While the two members of Odontophoridae that we sampled were always reciprocally monophyletic, *Ptilopachus* was only recovered as part of the Odontophoridae by some mitochondrial regions (Supplementary Material Table 5).

We also examined the ability of different mitochondrial regions to resolve specific clades present in the mitogenomic tree (Fig. 2) using the ‘entire’ taxon set. These clades included (1) the “core” phasianids (Supplementary Material Table 6); (2) the “erectile clade”, both including and excluding the deepest-branching genus in that group, *Ithaginis* (Supplementary Material Tables 7 and 8, respectively); and (3) the genus *Perdix* as sister to the gallopheasants (Supplementary Material Table 9). The mitogenome, along with CYB, ND5 and the combined tRNAs, consistently recovered all these phasianid groups. There were however, a number of regions (ATP8, COIII, ND3, ND4L and ND6) that never recovered these expected phasianid clades, while other regions did occasionally, depending upon analysis.

Intragenetic relationships showed a similar pattern of variation among mitochondrial regions. Most individual regions consistently recovered the mitogenomic (Fig. 2) topology for *Arborophila* (Supplementary Material Fig. 2), with moderate to high support (Supplementary Material Table 10). Estimates of phylogeny for both *Gallus* and *Syrmaticus* obtained using different mitochondrial regions exhibited a greater degree of variation, with many alternate topologies recovered, sometimes with only moderate support (Supplementary Material Tables 11 and 12 and Supplementary Material Figs. 3 and 4, respectively). Although support for relationships within *Gallus* was high in the mitogenomic tree (Fig. 2), no mitochondrial region recovered that (or any other) topology in all analyses, likely due to the short internodes within this genus (Fig. 2). Only two regions consistently recovered one topology across analyses within *Syrmaticus* (ND2 and ND4), but these two

regions disagreed with each other (Supplementary Material Table 12); other regions exhibited substantial variation in the recovered topology. Collectively, these results indicate that incongruence among regions occurs at a range of taxonomic levels.

3.3. Resolving conflicts in evolutionary signal among mitochondrial regions

Many mitochondrial regions did not consistently recover the topology found in analyses of the mitogenome (Fig. 2). Therefore, we examined a range of commonly employed analytical strategies to determine whether any provided a robust method to resolve this incongruence. We evaluated the impact of these approaches for the same nodes discussed above to determine whether some strategies appeared useful at one, but possibly not all, taxonomic depths.

Analyses can be affected by use of alternative models, particularly those that might better accommodate the patterns of molecular evolution evident in the mitogenome, especially the differences among sites due to functional constraints. Since codon positions vary in their constraints, a common strategy is to either partition based on codon position or to use a codon model for analyses. Neither of these two strategies reduced incongruence among regions in the position of the Odontophoridae (Table 2); nor did they have an impact upon the estimates of phylogeny for relationships at other taxonomic levels (Supplementary Material Tables 4–12).

The CAT–GTR mixture model implemented in PHYLOBAYES (Lartillot et al., 2006) allows even more flexibility in the patterns of molecular evolution among sites than the models implemented above. Thus, we examined the relationships supported by each mitochondrial region for the ‘entire’ and ‘reduced’ taxon sets using the CAT–GTR model. This did not result in greater congruence among regions in the placement of the Odontophoridae (Table 2), or at other taxonomic levels (Supplementary Material Tables 4–12). Thus, use of models that would be expected to better fit the data do not appear to consistently result in improved signal.

Removal of sites that provide misleading signal (phylogenetically ‘noisy’ sites) is another approach to improving phylogenetic reconstruction (e.g., Pratt et al., 2009). These misleading sites can

exacerbate long-branch attraction and obscure signal in studies trying to infer deep divergences. Rapidly evolving sites are more likely to exhibit homoplasy or saturation, and thus may be more problematic. We identified potentially misleading sites using both tree-independent (OV) and tree-dependent methods (evolutionary rate). As expected, a high proportion of the sites excluded with either approach were from the more rapidly evolving 3rd codon position (from 88% to 100% of sites, depending on method and taxon set). The mitogenomic position of the Odontophoridae was recovered for an additional two regions when sites were excluded (albeit with poor to moderate support; [Supplementary Material Table 13](#)), though in other cases exclusion of sites changed the topology from the mitogenomic to an alternative topology. Thus, for most regions, removal of sites by either approach eliminated valuable evolutionary signal, resulting in reduced resolution of relationships at a range of taxonomic levels ([Supplementary Material Tables 4–13](#)).

RY coding, which excludes rapidly accumulating transition mutations, can decrease saturation and enhance the recovery of evolutionary signal, and/or reduce any bias due to changes in base composition over time ([Braun and Kimball, 2002](#); [Delsuc et al., 2003](#); [Phillips and Penny, 2003](#); [Phillips et al., 2004](#)). In this study however, RY coding of third positions had little effect upon the estimated phylogeny for different gene regions at all taxonomic levels ([Table 2](#); [Supplementary Material Tables 4–12](#)). However, for the mitogenome, use of RY coding increased the number of analyses (particularly for the reduced taxon set) that correctly rooted the galliform tree (e.g., [Fig. 2](#)) rather than placing Megapodiidae as the sister group of Cracidae ([Supplementary Material Table 4](#)).

3.4. Correlation between 'decisive' sites and biological properties

In attempts to understand the biological basis of conflicting signal among mitochondrial regions, we identified the decisive sites (cf. [Kimball et al., 2013](#)) that provide high support for either of the two most plausible placements of the Odontophoridae ([Fig. 1A](#) and [B](#)). We then determined whether the decisive sites were disproportionately located in specific mitochondrial regions or, for those decisive sites in protein coding regions, associated with specific amino acid properties. Given that all analyses of complete mitogenomes recovered the expected placement of the Odontophoridae ([Fig. 1A](#)), it was not surprising that we identified more decisive sites supporting this topology ($N = 67$) than the most plausible alternate relationship ([Fig. 1B](#); $N = 44$).

If there were a biochemical basis underlying the distribution of misleading sites, then we would expect those decisive sites supporting topology B to be concentrated in a specific partition. We compared the proportion of decisive sites in (1) RNAs versus coding sequences and (2) between first and second versus third positions in the coding sequences. In both cases, the proportions of decisive sites supporting each topology were about equal and a two-tailed Fisher's Exact test indicated there were no significant differences in the decisive sites favoring topology A versus C ($p = 1.0$ in each case).

[Naylor and Brown \(1998\)](#) showed that misleading phylogenetic signal in chordate mitochondrial protein-coding genes are associated with rapidly evolving hydrophobic residues. Based on this, we expected that the majority of the 'decisive' sites supporting the expected topology ([Fig. 1A](#)) would encode for polar amino acids while those supporting the alternate topology ([Fig. 1B](#)) would encode for hydrophobic amino acids. We did not see this trend and both topologies were supported by nearly equal numbers of hydrophobic and polar 'decisive' sites (one-tailed Fisher's Exact test, $p = 0.31$). Our results were more in accordance with the observations of [Takezaki and Gojbori \(1999\)](#), who also did not find a

clear correlation between hydrophobicity and misleading phylogenetic signal, instead, they found that incorporating rate heterogeneity was important (the models we used incorporated among-sites rate heterogeneity, however, so that cannot explain the conflicts we identified). These results highlight that multiple phenomena may drive the accumulation of phylogenetic noise, the importance of which may differ depending upon the time of divergence. Regardless, our results indicate that a straightforward correlation between hydrophobicity and misleading signal is not applicable for all groups, and so isolating misleading signal using this approach should be used with caution.

3.5. The effects of taxon sampling

Increased taxon sampling generally improves phylogenetic estimation (e.g., [Zwickl and Hillis, 2002](#)). However, as noted above ([Table 1](#)), a greater number of mitochondrial regions supported the mitogenomic position of the Odontophoridae ([Fig. 2](#)) with the reduced taxon set. This was true for all of the alternative models examined and when rooting to the Megapodiidae rather than the anseriform outgroups ([Table 2](#)). Even the addition of taxa that are expected to break up long branches (*Ptilopachus*, *Arborophila*, and *Caloperdix*) actually appeared to increase noise. However, [Sanderson and Wojciechowski \(2000\)](#) noted a general decrease in bootstrap support when adding taxa and attributed this to a statistical bias in the bootstrap that reflects, at least in part, random homoplasy distributed among taxa. Thus, our results could reflect the high degree of homoplasy typical for mitochondrial data (and the fact that none of the models tested adequately accommodated that homoplasy).

3.6. Phylogenetic signal, power, and numbers of sites

The performance of gene regions depends in part upon the number of variable or informative sites. Not unexpectedly, the shortest region we examined (ATP8) did not appear to be very phylogenetically informative at any taxonomic level whereas the longest region (ND5) exhibited the greatest degree of congruence among analyses. However, the patterns we observed were not simply a function of the numbers of sites, since other long regions (COI and 16S) did not perform as well as ND5, while other short regions (e.g., ND4L and ND3) performed much better than ATP8. To examine this more rigorously, we compared the number of informative sites in a region with the Robinson–Foulds distance ([Robinson and Foulds, 1981](#)) between the tree estimated for each region and the mitogenomic tree ([Fig. 2](#)). Outside of the poor performance of ATP8, both 12S and CYB performed worse than expected whereas ND2, ND4 and ND5 all performed better ([Fig. 3](#)). We reached the same conclusions if we used only parsimony informative sites, or

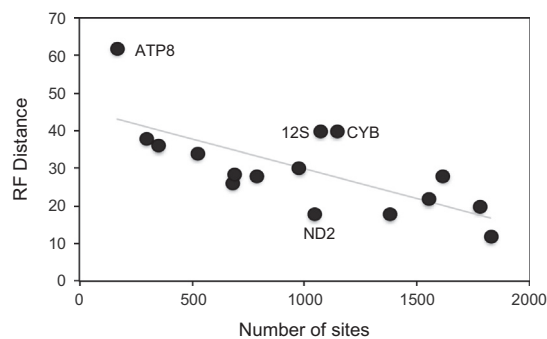


Fig. 3. Robinson–Foulds distance from the mitogenomic tree compared to the tree from individual partitions relative to the number of sites in the individual partitions.

the K tree score (Soria-Carrasco et al., 2007) as a distance (Supplementary Material Fig. 5). Within those broad patterns, however, specific patterns at each node vary (e.g., ND2 did not always identify the same relationships as in the mitogenomic tree, although it performs well in most cases). This suggests that the issue was more complex than the power differences associated with sequence length (and variation), and that the selection of a mitochondrial region can influence the results obtained.

4. Conclusions

This study assessed the phylogenetic signal in the mitogenome, with a particular focus on comparing different mitochondrial regions. Analyses of the mitogenome resulted in a well-supported phylogeny that was generally congruent with prior multi-locus studies. This was expected given both the rapid coalescent of the mitochondria, which results in a high probability of matching the species tree for most nodes, and the relatively large amount of data in the complete mitogenome. However, disagreement among mitochondrial regions, including some with relatively large numbers of informative sites, was documented at a range of taxonomic levels. We also found that the conflicting topologies sometimes had poor support despite the expectation that analyses of mitochondrial data should have high power. Although we explored many recommended analytical strategies to improve phylogenetic resolution, none of these approaches consistently alleviated the issues with incongruence among mitochondrial regions, emphasizing the difficulties associated with modeling the evolution of mitochondrial sequences.

The increased accessibility of next-generation sequencing technologies for data collection is expected to make the sequencing of complete mitogenomes even more common, sometimes as by-products of other studies (e.g., Barker et al., 2014; Miller et al., 2012; Nabholz et al., 2010). Based on our results, we advocate the use of mitogenomes (or at the very least multiple mitochondrial regions) for phylogenetic analyses. Individual regions should be used with extreme caution; although ND5 performed best in this study, it may not perform as well in other groups of organisms. However, even mitogenomes were unable to recover some relationships consistently in all analyses (e.g., the basal divergence within galliforms and the position of Odontophoridae). Instead, it appears that the conflicting signal among regions has led to low support for some relationships. Moreover, there was a modest decrease in performance of analyses using the expanded taxon set, probably reflecting the impact of homoplasy (cf. Sanderson and Wojciechowski, 2000). It seems reasonable to postulate that improved models of sequence evolution (possibly combined with corrections to the bootstrap) will alleviate this effect. Taken as a whole, our results suggest that further model development will be necessary to maximize the phylogenetic information from mitochondrial data in the future, so that when combined with nuclear data, better estimates of species trees can be obtained.

Acknowledgments

We thank anonymous reviewers for helpful suggestions that improved this manuscript. We also thank Ping Huang for help with some analyses. This work was supported by the US National Science Foundation (Grant Number DEB 1118823 to RTK and ELB). MJD was supported by a National Science Foundation Research Experiences for Undergraduates (NSF REU) supplement, HHMI Science for Life (University of Florida), and University Scholars (University of Florida). Tissues for some species were kindly donated by Louisiana State University Museum of Natural Science (LSUMNS) and the Florida Museum of Natural History (FLMNH).

Northern Bobwhite tissues were salvaged from hunter-killed birds and donated to BCF by William E. Palmer and Brad Dabbert. Complete mitochondria sequenced for this study have been submitted to GenBank (Accessions # KJ914543–KJ914548; see also Supplementary Material Table 1).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version, at <http://dx.doi.org/10.1016/j.ympev.2014.06.003>.

References

- Amaral, F.R.D., Sheldon, F.H., Wajntal, A., 2010. Towards an assessment of character independence in avian RNA phylogenetics: a general secondary structure model for the avian mitochondrial 16S rRNA. *Mol. Phylogenet. Evol.* 56, 498–506.
- 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 the Galliformes. *Auk* 118, 799–804.
- Barker, F., Oyler-McCance, S., Tomback, D., 2014. Blood from a turnip: tissue origin of low-coverage shotgun sequencing libraries affects recovery of mitogenome sequences. *Mitochondr. DNA*. <http://dx.doi.org/10.3109/19401736.19402013.19840588>.
- Berlin, S., Ellegren, H., 2001. Evolutionary genetics: clonal inheritance of avian mitochondrial DNA. *Nature* 413, 37–38.
- Berlin, S., Smith, N.G.C., Ellegren, H., 2004. Do avian mitochondria recombine? *J. Mol. Evol.* 58, 163–167.
- Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780.
- Botero-Castro, F., Tilak, M., Justy, F., Catzeflis, F., Delsuc, F., Douzery, E.J.P., 2013. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol. Phylogenet. Evol.* 69, 728–739.
- 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.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
- Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., Pande, N., Shang, Z., Yu, N., Gutell, R.R., 2002. The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3, 2.
- Cao, Y., Fujiwara, M., Nikaido, M., Okada, N., Hasegawa, M., 2000. Interordinal relationships and timescale of Eutherian evolution as inferred from mitochondrial genome data. *Gene* 259, 149–158.
- Castoe, T.A., de Koning, A.P., 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. USA* 106, 8986–8991.
- Cohen, C., Wakeling, J.L., Mandiwana-Neudani, T.G., Sande, E., Dranzoa, C., Crowe, T.M., Bowie, R.C.K., 2012. Phylogenetic affinities of evolutionarily enigmatic African galliforms: the Stone Partridge *Ptilopachus petrosus* and Nahan's Francolin *Francolinus nahani*, and support for their sister relationship with New World quails. *Ibis* 154, 768–780.
- Cori, A., Ellegren, H., 2013. Sampling strategies for species trees: the effects on phylogenetic inference of the number of genes, number of individuals, and whether loci are mitochondrial, sex-linked, or autosomal. *Mol. Phylogenet. Evol.* 67, 358–366.
- 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.
- 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.
- Degnan, J.H., DeGiorgio, M., Bryant, D., Rosenberg, N.A., 2009. Properties of consensus methods for inferring species trees from gene trees. *Syst. Biol.* 58, 35–54.
- Delsuc, F., Phillips, M.J., Penny, D., 2003. Comment on “Hexapod origins: monophyletic or paraphyletic?”. *Science* 301, 1482.
- Desjardins, P., Morais, R., 1990. Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *J. Mol. Biol.* 212, 599–634.
- Felsenstein, J., 1989. PHYLIP – phylogeny inference package (version 3.2). *Cladistics* 5, 164–166.
- Goremykin, V.V., Nikiforova, S.V., Bininda-Emonds, O.R., 2010. Automated removal of noisy data in phylogenomic analyses. *J. Mol. Evol.* 71, 319–331.
- Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C., 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, Sheldon, F.H., Steadman, D.W., Witt, C.C., Yuri, T., 2006. A phylogenomic study of birds reveals their evolutionary history. *Science* 320, 1763–1768.
- Kan, X.Z., Yang, J.K., Li, X.F., Chen, L., Lei, Z.P., Wang, M., Qian, C.J., Gao, H., Yang, Z.Y., 2010. Phylogeny of major lineages of galliform birds (Aves: Galliformes) based on complete mitochondrial genomes. *Genet. Mol. Res.* 9, 1625–1633.
- 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., Zwartjes, P., Crowe, T.M., 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., 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., Pääbo, 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. USA* 86, 6196–6200.
- Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095–1109.
- Lartillot, N., Brinkmann, H., Philippe, H., 2006. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* 7, S4.
- Lin, Y.H., McLenachan, P.A., Gore, A.R., Phillips, M.J., Ota, R., Hendy, M.D., Penny, D., 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. *Mol. Biol. Evol.* 19, 2060–2070.
- Maddison, D.R., Maddison, W.P., 2005. *MacClade 4: Analysis of Phylogeny and Character Evolution*. Available at: <http://macclade.org>.
- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, pp. 1–8.
- Miller, J.M., Malenfant, R.M., Moore, S.S., Coltman, D.W., 2012. Short reads, circular genome: skimming solid sequence to construct the bighorn sheep mitochondrial genome. *J. Hered.* 103, 140–146.
- 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, 1568–1571.
- Moore, W.S., 1997. Mitochondrial-gene trees versus nuclear-gene trees, a reply to Hoelzer. *Evolution* 51, 627–629.
- Moritz, C., Dowling, T.E., Brown, W.M., 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Ann. Rev. Ecol. Syst.* 18, 269–292.
- Nabholz, B., Jarvis, E.D., Ellegren, H., 2010. Obtaining mtDNA genomes from next-generation transcriptome sequencing: a case study on the basal Passerida (Aves: Passeriformes) phylogeny. *Mol. Phylogenet. Evol.* 57, 466–470.
- Naylor, G.J.P., Brown, W.M., 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* 47, 61–76.
- Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* 28, 171–185.
- Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* 21, 1455–1458.
- Phillips, M.J., Gibb, G.C., Crimp, E.A., Penny, D., 2010. Tinamou and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. *Syst. Biol.* 59, 90–107.
- Powell, A.F.L.A., Barker, F.K., Lanyon, S.M., 2013. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study. *Mol. Phylogenet. Evol.* 66, 69–79.
- Pratt, R.C., Gibb, G.C., Morgan-Richards, M., Phillips, M.J., Hendy, M.D., Penny, D., 2009. Toward resolving deep Neoaves phylogeny: data, signal enhancement, and priors. *Mol. Biol. Evol.* 26, 313–326.
- Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53, 131–147.
- Russell, R.D., Beckenbach, A.T., 2008. Recoding of translation in turtle mitochondrial genomes: programmed frameshift mutations and evidence of a modified genetic code. *J. Mol. Evol.* 67, 682–695.
- Russo, C.A., Takezaki, N., Nei, M., 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* 13, 525–536.
- Sammler, S., Bleidorn, C., Tiedemann, R., 2011. Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genomics* 12, 35.
- Sánchez-García, A., Castresana, J., 2012. Impact of deep coalescence on the reliability of species tree inference from different types of DNA markers in mammals. *PLoS ONE* 7, e30239.
- Sanderson, M.J., Wojciechowski, M.F., 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-Astragalus (leguminosae). *Syst. Biol.* 49, 671–685.
- 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.
- Sibley, C.G., Ahlquist, J.E., 1990. *Phylogeny and Classification of Birds: A Study in Molecular Evolution*. Yale University Press, New Haven.
- Silvestro, D., Michalak, I., 2012. RaxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol.* 12, 335–337.
- 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.
- Soria-Carrasco, V., Talavera, G., Igea, J., Castresana, J., 2007. The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. *Bioinformatics* 23, 2954–2956.
- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Steel, M., 2005. Should phylogenetic models be trying to “fit an elephant”? *Trends Genet.* 21, 307–309.
- Swofford, D.L., 2002. *PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Takezaki, N., Gojobori, T., 1999. Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences. *Mol. Biol. Evol.* 16, 590–601.
- Tourasse, N.J., Gouy, M., 1997. Evolutionary distances between nucleotide sequences based on the distribution of substitution rates among sites as estimated by parsimony. *Mol. Biol. Evol.* 14, 287–298.
- Wang, J., Wang, W., 1999. A computational approach to simplifying the protein folding alphabet. *Nat. Struct. Biol.* 6, 1033–1038.
- Wang, N., Kimball, R.T., Braun, E.L., Liang, B., Zhang, Z., 2013. Assessing phylogenetic relationships among Galliformes: a multigene phylogeny with expanded taxon sampling in Phasianidae. *PLoS ONE* 8, e64312.
- Weisrock, D.W., Harmon, L.J., Larson, A., 2005. Resolving deep phylogenetic relationships in salamanders: analyses of mitochondrial and nuclear genomic data. *Syst. Biol.* 54, 758–777.
- Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829.
- Zhang, P., Liang, D., Mao, R.L., Hillis, D.M., Wake, D.B., Cannatella, D.C., 2013. Efficient sequencing of anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. *Mol. Biol. Evol.* 30, 1899–1915.
- 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., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51, 588–598.