Categorical edge-based analyses of phylogenomic data reveal conflicting signals for difficult relationships in the avian tree

Ning Wang[1,3,4*], Edward L. Braun[2], Bin Liang[1,4], Joel Cracraft[3], Stephen A. Smith[1]

[1]Department of Ecology & Evolutionary Biology, University of Michigan, 1105 N University Ave, Ann Arbor, MI 48109-1048, USA

[2]Department of Biology, University of Florida, Gainesville, FL 32607, USA

[3]Department of Ornithology, American Museum of Natural History, New York, NY 10024, USA

[4]College of Life Sciences, Inner Mongolia University, Hohhot, 010070, China

***Correspondence**

Ning Wang

Current address: College of Life Sciences, Inner Mongolia University, Hohhot, 010070, China

Email: ningwang83@hotmail.com

ABSTRACT:

Phylogenetic analyses of large-scale datasets sometimes fail to yield a satisfactory resolution of the relationships among taxa for a number of nodes in the tree of life. This has even been true for genome-scale datasets, where the failure to resolve relationships is unlikely to reflect limitations in the amount of data. Gene tree conflicts are particularly notable in studies focused on these contentious nodes in the tree of life, and taxon sampling, different analytical methods, and/or data-type effects are thought to further confound analyses. Observed conflicts among gene trees arise from both biological processes and artefactual sources of noise in analyses. Although many efforts have been made to incorporate biological conflicts, few studies have curated individual genes for their efficiency in phylogenomic studies. Here, we conduct an edge-based analysis of Neoavian evolution, examining the phylogenetic efficacy of two recent phylogenomic bird datasets and three datatypes (ultraconserved elements [UCEs], introns, and coding regions). We assess the potential causes for biases in signal-resolution for three difficult nodes: the earliest divergence of Neoaves, the position of the enigmatic Hoatzin (*Opisthocomus hoazin*), and the position of owls (Strigidae). We observed extensive conflict among genes for all data types and datasets even after we removed potentially problematic loci. Edge-based analyses increased congruence and examined the impact of data type, GC content variation (GC_{CV}), and outlier genes on analyses. These factors had different impact on each of nodes we examined. First, outlier gene signals appeared to drive different patterns of support for the relationships among the earliest diverging Neoaves. Second, the position of Hoatzin was highly variable, but we found that data type was correlated with the signals that support different placements of the Hoatzin. However, the resolution with the most support in our analyses was Hoatzin + shorebirds. Finally, GC_{CV}, rather than data type (i.e., coding vs non-coding) *per se*, was correlated with an owl + Accipitriformes signal. Eliminating high GC_{CV} loci increased the signal for an owl + mousebird relationship. Difficult edges (i.e., characterized by deep coalescence and high gene-tree estimation error) are hard to recover with all methods (including concatenation, multispecies coalescent, and edge-based analyses), whereas "easy" edges (e.g., flamingos + grebes) can be recovered without ambiguity. Thus, the nature of the edges, rather than the methods, is the limiting factor. Categorical edge-based analyses can reveal the nature of each edge and provide a way to highlight especially problematic branches that warrant further examination in future phylogenomic studies. We suggest that edge-based analyses provide a tool that can increase our

understanding about the parts of the avian tree that remain unclear, even with large-scale data. In fact, our results emphasize that the conflicts associated with edges that remain contentious in the bird tree may be even greater than appreciated based on previous studies.

keywords: edge-based analyses; Neoavian resolution; GC content variation; outlier signal

1. Introduction

The increasing availability of genomic datasets is increasingly facilitating the resolution of phylogenetic relationships. However, recent studies have highlighted that as datasets have grown, so has conflict. This may involve gene-tree discordances within a single study (e.g., Wang et al. 2019) or the resolution of difficult and controversial hypotheses found across multiple studies (e.g., Wickett et al. 2014; Puttick et al. 2018). Conflicting gene trees can be the result of biological processes such as hybridization, gene duplication and loss, and incomplete lineage sorting (ILS) or they can be the result of data-set assembly error (Gatesy and Springer 2017) or other artificial noise arising from, for example, orthology inference, sequence alignment, model specification, and insufficient informative variation (Springer and Gatesy 2019). In part, as a result of these conflicts, several areas across the tree of life have consistently been difficult to resolve (Wang et al. 2012; Tarver et al. 2016; Shen et al. 2017; Wang et al. 2019) despite the use of genome-scale datasets (Foley et al. 2016) and increased taxon sampling (Prum et al. 2015; Wang et al. 2017). Additionally, different analytical methods (e.g., concatenation-based or coalescent-based methods), and/or the inclusion or exclusion of multiple data-types (protein coding or noncoding regions, Reddy et al. 2017) can result in conflicting resolutions.

Data limitations such as insufficient sampling of taxa or gene regions can be overcome with increasing effort and worldwide collaboration, as well as advances in DNA sequencing techniques that can obtain DNA from ancient specimens preserved in museums that complement lack of fresh specimen (e.g., Chen et al. 2018). On the other hand, resolution of methodological issues such as inaccurate evolutionary models may be more difficult to address. In particular, difficult-to-resolve edges on the tree of life may be the result of low signal within very short internodes. Concatenation can overcome low signal by combining the signal from multiple genes but also assumes that all genes share a single evolutionary history, yet population genetic and molecular evolutionary processes can result in highly heterogeneous gene trees and violate this assumption (Kolaczkowski and Thornton 2004; Kubatko and Degnan 2007; Edwards 2009). The most commonly used summary coalescent methods (e.g., ASTRAL III, Zhang et al. 2018; MP-EST, Liu et al. 2010), on the other hand, accommodate gene tree conflicts but assume accurately estimated gene trees for neutrally evolving gene regions that are free of intralocus recombination and only exhibit conflict due to ILS (Springer and Gatesy 2016). Typical phylogenomic datasets violate many of these assumptions. If there are numerous gene trees that conflict for reasons

other than ILS (whether this reflect other biological processes or estimation error) the assumptions of coalescent methods are violated and the methods would not be a consistent estimator of the species tree (although many methods will still be relatively robust to these errors). Also, when analyzing large genomic datasets, many genes will have experienced episodic periods of strong selection, or linkage and hitchhiking can also cause violation of neutral evolution for most genomic regions (e.g., Bendall et al. 2016). Some genes may even show within-gene conflict, supporting multiple trees (Kimball et al. 2013; Mendes et al. 2019). These observations highlight the importance of examining individual genes in phylogenetic analyses. Furthermore, while biological processes may drive some conflict, errors in alignment, assembly, and gene-tree estimation may be common, especially as dataset size increases, with larger datasets being more difficult to curate (Gatesy and Springer 2017; Liu et al. 2017).

With ~10,500 living species, birds represent the most species-rich group of tetrapod vertebrates. Despite their importance, several controversial relationships in Neoaves have not been settled, even with genome-scale data (e.g., Jarvis et al. 2014; Prum et al. 2015; Reddy et al. 2017). Challenges to resolving these nodes are not unexpected considering that neoavian birds likely experienced a very rapid radiation (e.g., Ericson et al. 2006; Jarvis et al. 2014; Prum et al. 2015; Houde et al. 2020). Although genome-scale datasets have facilitated the recovery of many relationships (e.g., passerines + parrots, flamingos + grebes, “core” landbirds, “core” waterbirds, sunbittern + tropicbirds, sandgrouse + mesites; Ericson et al. 2006; Hackett et al. 2008; Suh et al. 2011; Wang et al. 2012; McCormack et al. 2013; Jarvis et al. 2014; Prum et al. 2015; Suh et al. 2015), the Neoaves are still well-known for multiple, extremely recalcitrant nodes (Jarvis et al. 2014; Prum et al. 2015). Specifically, the earliest divergences of Neoaves remain one of the most difficult phylogenetic problems in vertebrates, which has even been suggested to represent a hard polytomy (Suh 2016, but see Reddy et al. 2017). Within Neoaves, the positions of Hoatzin (*Ophisthocomus hoazin*) and owls (Strigiformes) remain especially problematic as well as being significant for the analysis of avian development, life history, and morphological traits. The Hoatzin is a bizarre and enigmatic bird species. It converts plant cellulose to simple sugars using microbial foregut fermentation (Domínguez-Bello et al. 1994) and has a modified skeleton (sternum) to accommodate a large crop. Also, Hoatzin chicks have claws on two of their wing digits (Hughes and Baker 1999). These characteristics have confounded its phylogenetic position for decades. Owls are nocturnal predators with raptorial features, but their relationships have

been obscure. Within Neoaves there are at least two distinct raptorial lineages. The first is the falcons (Falconidae), which have now been shown to be related to parrots and passerines, whereas the second, Accipitriformes, includes the hawks, eagles, and vultures. Presently, it remains unclear whether owls represent a third raptorial lineage or whether they are sister to hawks, eagles, and vultures. This uncertainty impacts our understanding of the evolution of raptorial features.

Recent avian studies have suggested that phylogenetic reconstruction is biased by the source of the data (e.g., coding regions [hereafter exons], introns, ultraconserved elements [UCE], Reddy et al. 2017), which we refer to as data-type effects. For example, in Jarvis et al. (2014), concatenated analyses of exons, introns, and UCEs were largely discordant among each other. The resolution of two of the above three difficult relationships (specifically, the position of owls and the topology for the earliest split of Neoaves) has been shown to vary among different data types (Reddy et al. 2017; Braun et al. 2019). The third difficult relationship – the position of the Hoatzin in the bird tree – has been even more variable across studies (e.g., Jarvis et al. 2014; Reddy et al. 2017) and the existence of a correlation with data type is less clear. Due to biased base composition and high GC content variation (hereafter GC_{CV}), exons received the most criticism for model violation and impact on phylogenetic reconstruction (Jarvis et al. 2014; Reddy et al. 2017). Adjusting model heterogeneity to better fit different data (e.g., Galtier and Gouy 1998; Morgan et al. 2013) might be applicable, but the dramatic heterogeneity among gene trees is still notable even within the same data type (e.g., Jarvis et al. 2014). Nevertheless, were exons to be generally misleading in the relationships they reconstruct, this would have far reaching implications for phylogenomic analyses as transcriptomic and targeted sequencing data have become more common (e.g., Prum et al. 2015; Wang et al. 2019).

Herein we examine conflicting signals within two major avian datasets: Jarvis et al. (2014) and Prum et al. (2015). Specifically, we conducted constraint phylogenetic analyses (Smith et al. 2020, hereafter edge-based analyses) to examine the three confounding questions regarding relationships within the large clade Neoaves (i.e., the earliest divergence within Neoaves, the position of Hoatzin, and the position of owls). We explore the efficacy of large datasets to resolving these relationships by carefully curating individual genes and assessing their contribution to the support for alternative hypotheses. In addition to individual gene evaluations, we focused on dissecting the phylogenetic signal associated with different data types (exon,

147 intron, and UCE) and we examined other factors, such as variation in GC content (Foster and
148 Hickey 1999) and gene evolutionary rate (Chojnowski et al. 2008) that also might influence
149 phylogenetic estimation. The inclusion of multiple data types facilitates examination of whether
150 phylogenetic signal from protein-coding regions were different from that of non-coding regions.
151 The edge-based analyses (see detailed description below) examine phylogenetic signal by
152 investigating alternative resolutions for individual edges, and similar approaches have been
153 successfully applied to determine support for specific controversial relationships (e.g., Walker et
154 al. 2018; Wang et al. 2019; Smith et al. 2020). Our use of edge-based analysis is not designed to
155 resolve these contentious relationships but explore a method for identifying the nature of difficult
156 relationships, especially around short internodes. Through these analyses, we seek to better
157 characterize the ability of large datasets to address difficult phylogenetic questions in birds and
158 other groups.
159
160 **2. Materials and methods**
161 ***2.1 Data acquisition and filtering***
162 We obtained the alignment data including 8251 exons, 2516 introns, and 3769 UCEs from Jarvis
163 et al. (2015). In order to increase the phylogenetic signal from individual genes and eliminate
164 potential systematic error or noise (e.g., incorrect orthologous inference from the original study),
165 we conducted several filtering steps. First, genes having a short alignment (< 500bp) and/or
166 missing taxa (< 42 species after deleting sequences with > 70% gaps in the alignment) were
167 excluded. Second, we reconstructed ML gene trees using IQTREE (v. 1.6.3, Nguyen et al. 2014)
168 with the GTR + Γ model and 1000 ultrafast bootstrap (UFBS) replicates. Trees having outgroups
169 that clustered within ingroups (i.e., non-bird species nested within birds, and/or species from
170 Palaeognathae and Galloanserae nested within Neoaves) were treated as systematic error and
171 excluded from additional analyses. Finally, trees with extremely long branch lengths (> 1
172 expected substitutions per site) were also excluded due to potential error in ortholog
173 identification. Scripts are available at https://bitbucket.org/ningwang83/avian_conflict_assess. In
174 addition, the original 259 sequence alignments and gene trees from Prum et al. (2015) were
175 downloaded and filtered following a same procedure.
176
177 ***2.2 Examination of gene tree resolution diversity***

178 In general, we focused on three controversial edges on the avian tree of life: the earliest
179 divergence in Neoaves and those encompassing the positions of the Hoatzin and owls (Table 1).
180 To assess the diversity of resolutions in the gene trees, we conducted several analyses. First, we
181 examined the topological distributions of ML gene trees (Fig. 1) in relation to the three focal
182 edges under different support cutoffs (i.e., ≥ 0%, or 50%, or 70%, or 95% support from either
183 regular bootstrap [BS] analyses or Ultra Fast Bootstrap [UFBS], Minh et al. 2013, Hoang et al.
184 2018). Second, we examined Robinson–Foulds (RF, Robinson and Foulds 1981) distances
185 among gene-trees for each data type, as well as between each gene tree and the concatenated
186 TENT ExaML species tree from Jarvis et al. (2014) and the concatenated RAxML tree from
187 Prum et al. (2015). RF distances were calculated using the bp script
188 (https://github.com/FePhyFoFum/gophy), which allows non-overlapping taxa in the gene trees.
189 Finally, to quantify the relative strength of the phylogenetic signal for a certain edge, we
190 calculated the ICA (internode certainty all; the degree of certainty for a bipartition in a given set
191 of trees in conjunction with all conflicting bipartitions in the same tree set) values (Salichos et al.
192 2014) on the species trees (same as above) from the filtered gene trees of each data type
193 respectively. These analyses were conducted using RAxML 8.2.11 (Stamatakis 2014) with the
194 GTRCAT model (Stamatakis 2006) and lossless support to accommodate partial gene trees
195 (Kobert et al. 2016).
196
197 ***2.3 Assessment for constraint topologies for three focal edges***
198 We selected several alternative topologies from previous large-scale studies (Table 1) for each of
199 the three focal edges. For the Jarvis et al. (2014) data, gene trees were reconstructed in IQTREE
200 by giving one constraint edge at a time while letting the remainder of the tree to be optimized
201 with the best ML topology. This analysis was conducted for each locus and every focal edge (see
202 example flow chart in Fig. 2), but genes were filtered out if their gene tree was incompatible with
203 any testing constraint under a 95% UFBS cutoff. For example, any ML gene tree showing
204 Hoatzin + Loons, which is conflict with all the seven test hypotheses (Table 1), would be
205 counted in the edge analysis (Fig. 2B) if Hoatzin + Loons shows UFBS support < 95% (i.e., if it
206 is unsupported) in the ML tree. However, if the UFBS support for Hoatzin + Loons was ≥ 95%,
207 we would remove the gene from analyses due to its strong conflict with all of the test hypotheses.
208 To identify the best hypothesis, we conducted the following analysis:

209 (1) We compared ln$L$ scores among conflicting alternative topologies for a single edge
210 for each gene and identified its supporting (i.e., the best) topology.
211 (2) We calculated the difference in ln$L$ (Δln$L$) between the best and the second-best
212 resolutions for each gene. At this stage we also identified potential outlier genes
213 (cases where Δln$L$ fell out of the smooth distribution among other genes) by eye.
214 (3) We summed Δln$L$ for each constraint from supporting or significantly supporting
215 (Δln$L > 2$) genes ($\sum\Delta$ln$L$ or $\sum$sigΔln$L$).
216 (4) We identified the overall best topology either supported or significantly supported
217 (Δln$L > 2$) using two criteria: the largest number of genes ($\sum$Ng, note that this value
218 is simply the number of genes supporting a specific topology) and the highest $\sum\Delta$ln$L$
219 (or $\sum$sigΔln$L$) scores (Fig. 2).
220 Genes for which the ln$L$ could not be calculated given specific constraint topologies (this occurs
221 when relevant taxa are missing) were excluded from the calculation. For genes that strongly
222 support one hypothesis versus others (i.e., outlier genes, Fig. 2), we manually examined them for
223 misalignment. We also collected information about the functions of the genes. We did this
224 because positive selection can give rise to convergent evolution, at least in principle, and genes
225 with a specific function might group corresponding species together and mislead phylogenetic
226 inferences (e.g., Stern 2013). For example, hearing gene *Prestin* grouped echolocating taxa
227 (microbats and toothed whales), presumably reflecting convergence (Li et al. 2010). Thus, we
228 identified exons with outlier topologies using BLAST to see if those outlier genes exhibit
229 specific functions that might be relevant to the topology they strongly supported. However, we
230 did not conduct lineage-specific selection analyses for outlier exons as their alignments
231 contained several premature stop codons.
232 To examine taxon-sampling effect, we also conducted edge-based analyses using genes
233 from the large sampling of Prum et al. (2015), following similar processes and criteria indicated
234 above. In order to limit CPU time, the RAxML 8.2.11 (Stamatakis 2014) with GTRCAT model
235 was used. The ML gene-trees including BS support built with RAxML were directly retrieved
236 from Prum et al. (2015). We also reduced the number of constraint topologies for each focal edge
237 by including only the two published hypotheses that are best supported by Jarvis et al. (2014)
238 data, as well as the one shown in Prum et al. (2015). Results were organized after filtering out
239 loci with gene-tree incompatibilities with any testing constraints under 70% BS cutoff. In

240 addition, to examine the sampling effect on the hypotheses examined here, we focused on the
241 earliest diverging Neoaves and assessed five additional hypotheses by randomly referring one
242 single lineage (i.e., caprimulgiforms [nightjars], Hoatzin, charadriiforms [represented by plovers],
243 gruiforms [cranes], or cuculiforms [cuckoos], Fig. 3) as sister to other Neoaves and conducted
244 edge-based analyses with Jarvis et al. (2014) data using the same criteria mentioned above.
245 In order to assess influence of GC_{CV} and tree length (i.e., gene evolutionary rate) on the
246 phylogenetic reconstruction using different data types, we sorted the filtered genes from Jarvis et
247 al. (2014) based on their GC_{CV} and tree length. Specifically, we expanded the method from
248 Reddy et al. (2017) by limiting GC content calculations to variable sites and comparing only
249 within bird species (i.e., excluding nonavian outgroups) to generate the ΔGC_{50} (the GC
250 difference between the inter-quartile GC-rich taxa), as well as the ΔGC_{90} (the GC difference
251 between the 95 and 5 percentile GC-rich taxa) and the ΔGC_{100} (the GC difference between the
252 most and least GC-rich taxa). We ordered genes based on ΔGC_{90} to balance the GC difference
253 and similarity among data types (Fig. 4a) and examined the gene trees in the upper and lower
254 quartiles of the variation for their supporting pattern to each testing edge (Fig. 4b). We
255 conducted similar analyses by sorting genes based on their tree length and compared the edge
256 assessment results between gene sets from the upper and lower quartile of the tree lengths (Fig.
257 4c and d).
258
259 ***2.4 Coalescent simulation – assessing the performance of edge-based analyses for deep***
260 ***coalescence***
261 We conducted simulation analyses to assess the performance of edge-based analyses in
262 recovering the potential true topology at edges under differing level of ILS. The Jarvis et al.
263 (2014) MP-EST* TENT tree, which was built with a binned MP-EST* method (Mirarab et al.
264 2014), was selected as the true tree. 500 gene trees were then simulated under the neutral
265 coalescent model using the *treesim.contained coalescent_tree* module in Dendropy 4.2.0
266 (Sukumaran and Holder 2010). We rescaled the internal branches (hereafter inbrlen) of these
267 gene trees with 0.01 that is estimated by comparing the coalescent branch length with the
268 substitution branch length in the empirical data of Jarvis et al. (2014). The tip branch-lengths
269 (hereafter tipbrlens) were rescaled based on species' tipbrlen distribution of the empirical data of
270 Jarvis et al. (2014). A recent study (Duchêne et al. 2018) found that a proportionally linked

271 model is the best fit for gene tree branch length differences, meaning the tipbrlens are
272 proportional among species across gene trees. Under this condition, we first obtained the branch-
273 lengths distribution for each species from the empirical ML gene trees of Jarvis et al. (2014).
274 Then, we generated tipbrlens for each simulated gene by selecting the same percentile tipbrlens
275 value from the empirical distribution of each species. This rescaling method ensures a
276 proportional tipbrlens across all gene trees. It is noteworthy that the branch length distributions,
277 as well as the modeling parameters (see below) were estimated from exons, introns and UCEs,
278 respectively, to simulate different types of data. Here, we conducted ASTRAL analyses
279 (ASTRAL III, Zhang et al. 2018) using these rescaled true gene trees (for each datatype) to build
280 a coalescent species tree (hereafter simAstral). Moreover, the average RF distance (ARFD)
281 between the true species tree and the true gene trees was used to reflect the level of overall ILS.
282 We generated one sequence alignment for each tree with length of 800-3000bp. Seq-Gen
283 (Rambaut and Grass 1997), implemented in Dendropy, was used for alignment-generation under
284 the GTR + Γ model. The base compositional frequency, the alpha parameter for gamma
285 distribution, and the GTR substitution rate were also selected randomly from the density
286 distribution of the Jarvis et al. (2014) empirical data of exons, introns, and UCEs, respectively.
287 Simulations were repeated once for each data type to validate results, so there were six sets of
288 500 simulated genes in total. Using these datasets, we re-estimated gene trees in IQTREE using
289 GTR + Γ model with and without bipartition constraints (i.e., ML gene trees). In total, 184
290 bipartition constraints were generated from 28 Jarvis species-trees representing the variety of
291 potential hypotheses as to bird relationships from the different data types, so that all the
292 bipartitions in the true tree were included. The edge-based analyses were conducted following
293 the above-method on the three difficult focal edges that exhibit higher level of ILS as well as
294 some commonly accepted edges that exhibit lower levels of ILS, including the core landbirds (**I**),
295 the core waterbirds (**II**), the tropicbirds + sunbittern (i.e., Phaethontimorphae, **III**), and the
296 nightjar + swifts + hummingbirds (i.e., Caprimulgimorphae, **V**) from the "magnificent seven"
297 superordinal groups (Reddy et al. 2017). At this point, we also conducted ASTRAL analyses
298 (hereafter MLAstral) using the re-estimated ML gene trees. We also calculated the mean gene-
299 tree estimation error (GTEE) by conducting a normalized RF distance between the true and
300 estimated gene trees, averaged across all genes, using Dendropy. Moreover, in order to also
301 compare the performance of concatenation analyses, we concatenated the simulated alignments

302 of each data type and conducted ML analyses in RAxML 8.2.11 using GTR + Γ model that
303 partitioned by locus, and with node support assessed by 200 bootstrap replicates.
304
305 ***2.5 Comparison of tree topologies within genetic regions***
306 Within-genetic region comparisons were conducted using exons and UCEs from Jarvis et al.
307 (2014), which have high alignment quality without long indels, in addition to the Prum et al.
308 (2015) dataset. We split the genes into 1000bp chunks, omitting the short leftovers. We
309 conducted ML analyses in IQTREE based on GTR + Γ model with uncertainty measured using
310 both SH-aLRT (Guindon et al. 2010) and UFBS (Minh et al. 2013), each with 1000 replicates.
311 Edges with < 80% SH-aLRT or < 95% UFBS support were collapsed. We note that this
312 requirement is extremely conservative but limits false positives. We also re-estimated gene trees
313 using these same criteria for consistency among these tests. We then compared trees from gene
314 region fragments to the ML gene trees based on the entire gene region and noted conflicts.
315
316 **3. Results and Discussion**
317 ***3.1 Gene informativeness***
318 To increase phylogenetic signal and reduce noise, we carefully curated data before analyses. For
319 genes from Jarvis et al. (2014), 246 introns and 924 exons were excluded due to having a short
320 alignment (< 500bp) and/or a large number of missing taxa (< 42 species, Table S1). After
321 removing genes with potential systematic error (non-monophyletic outgroups and extremely long
322 terminal branches), we obtained 5062 exons, 1409 introns and 3370 UCEs (Table S1). All data
323 from Prum et al. (2015) passed the filtering processes and were therefore retained.
324 We examined the extent to which each genomic partition had information to resolve
325 nodes within the tree. We found that each gene in the Jarvis et al. (2014) dataset supported, on
326 average, 24 (≥ 70% UFBS) or 15 (≥ 95% UFBS, Minh et al. 2013) of the 47 internal nodes,
327 respectively, whereas the Prum et al. (2015) data supported (≥ 70% BS, Hillis and Bull 1993)
328 ~110 out of 197 internal nodes on each gene tree, with a maximum of 144 and minimum of 22
329 (Fig. 1A). Although the low support on many nodes can have many causes, it is likely that there
330 was relatively low phylogenetic signal for individual genes that are subject to high gene-tree
331 estimation error (GTEE, Molloy and Warnow 2018). In fact, this is not unexpected as single
332 genes are unlikely to be informative across such broad relationships. Although introns generally

333 supported more relationships, the number of supporting nodes appeared to correspond to
334 alignment length (Fig. S1). For the three difficult edges of interest, only 23%– 33% exons, 38%–
335 51% introns, and 34%–43% UCEs in the Jarvis et al. (2014) data obtained ≥ 70% UFBS on ML
336 gene trees after excluding genes that were missing owls or Hoatzin. Similarly, when we used the
337 70% BS cutoff for Prum et al. (2015) data, we only found four loci obtained supports on their
338 ML gene trees for the resolution of Hoatzin, two genes showed supports for the earliest diverging
339 clade of Neoaves, whereas no gene supported the position of owls. We also identified 113 genes
340 in the Prum et al. (2015) dataset that showed non-monophyly of owls (i.e., the two families of
341 owl had distinct positions in the tree). Taken as a whole, these results suggest that most loci have
342 limited information for resolving the contentious edges and therefore a high potential for GTEE
343 from these genome data, regardless of the sources of conflict.
344
345 ***3.2 Diversity of resolutions***
346 Many gene regions across datasets and datatypes did not demonstrate a strong preference for any
347 specific resolution regarding the three focal relationships (Fig. S2). However, to ascertain the
348 nature of the support for alternative resolutions when support was indicated, we conducted
349 several analyses. First, we examined how many different resolutions were supported by different
350 genes. We observed, for example in the Jarvis et al. (2014) exons, there were more than 1500
351 alternative resolutions for each focal relationship. Most of these were unsupported (~ five
352 resolutions obtained ≥ 95% UFBS and ~40 resolutions obtained ≥ 50% UFBS) or were present in
353 fewer than 10 gene trees (i.e., rare resolutions, Fig. 1B). Even the best-supported topology for
354 each data type was still present in less than 5% of gene trees. The situation was similar for
355 introns and UCEs, and not limited to the Jarvis et al. (2014) dataset (Fig. S3A). The Prum et al.
356 (2015) dataset had 259 loci, and we identified 110 or more topologies for each focal edge, but
357 very few of those topologies had high (> 70%) bootstrap support and most were found in a single
358 gene (Fig. S3B).
359 The uncertainty of focal relationships has also been confirmed by the extremely low and
360 negative ICA values on the species tree. For example, the TENT_ExaML tree in Jarvis et al.
361 (2014) showed Phoenicopterimorphae (**VII**: flamingos and grebes) + Columbimorphae (**VI**:
362 doves, sandgrouse and mesites) as the earliest diverging clade within Neoaves (Fig. 2A).
363 Analyses of the non-coding (intronic and UCE) data in Jarvis et al. (2014) and of indels in Houde

364 et al. (2019) strongly suggest that one or both of the two groups are sister to all other Neoaves.
365 However, this clustering obtained -0.051 (intron), -0.272 (UCE), and -0.317 (exon) ICA values,
366 indicating substantial conflicting signals among gene trees, and a better supported alternative
367 conflicting relationship as shown in the ML topology distribution (Fig. S3). The very different
368 ICA values are also indicative of a data type effect, with introns showing a very different signal
369 than UCEs or exons. Similar situations were also found for the other two focal edges (i.e., for the
370 Hoatzin and owls) and in both datasets (not shown).
371 The diversity of resolutions can also be observed by the RF distances among gene trees.
372 Trees based on the exons and UCEs from Jarvis et al. (2014) exhibited more among-gene tree
373 differences (~65, 69% of the edges, Fig. S4) than do trees based on introns (~58, 62% of the
374 edges), implying the potential of high GTEE in these gene trees (Simmons et al. 2016). Similar
375 patterns were also found when comparing gene trees to the species-tree topologies. The RF
376 distances among gene trees for the Prum et al. (2015) data were usually larger than 150 (38% of
377 the edges) and as high as 300 (76% of the edges, Fig. S4), suggesting high topological diversity
378 in these data; the larger number of topologies reflects the denser taxon sampling of the Prum et al.
379 (2015) study.
380
381 ***3.3 Edge analyses***
382 Examining conflicting signals in large-scale datasets is perhaps the foremost challenge in the
383 field of systematics now that we have entered the phylogenomic era. One analytical framework
384 that is commonly used is concatenation (e.g., Jarvis et al. 2014; Prum et al. 2015). Although
385 concatenated analyses have been criticized for their failure to model the multispecies coalescent
386 (Edwards et al. 2016) it is still a useful methodology in many contexts (Bryant and Hahn 2020).
387 However, the utility of concatenated analyses for the examination of conflict is limited because
388 they generally yield a single tree. Low support values in analyses of concatenated data can
389 provide evidence that conflict might exist, but low support can also reflect the absence of any
390 signal in the data (e.g., due to the existence of a polytomy). Fundamentally, there are only two
391 ways to examine conflict using the concatenated approach. First, one can subdivide data based
392 on some criterion, such as coding vs. non-coding data (Jarvis et al. 2014; Reddy et al. 2017) or
393 protein structure (e.g., Pandey and Braun 2020), and conduct concatenated tree searches on the
394 data subsets. Second, one can examine differences in the tree selection criterion (typically the

395 likelihood) for various topologies at the level of the genes (Shen et al. 2017) or the individual site
396 level (Kimball et al. 2013). Both approaches have limitations; the first approach only allows one
397 to assess data-type effects whereas the second approach can only yield information about a small
398 number of topologies due to the limited availability of computational time.

399 An alternative way to examine conflicts involves comparisons of gene trees; in fact,
400 comparisons among gene trees represent an obvious way to assess conflict because individual
401 gene trees can have any topology. However, analyses of individual genes are subject to high
402 GTEE, so many bipartitions in estimated gene trees may be spurious. Moreover, there can be
403 conflicts within gene regions due to intralocus recombination and other factors (see below,
404 section 3.6). Ultimately, it is difficult to determine whether the extensive conflict typically
405 observed when estimated gene trees are compared reflects a mixture of meaningful biological
406 signals (regardless of whether those signals are genuine conflicts among gene trees or biases due
407 to some other factor) or simply random resolutions due to GTEE. We have chosen to use a
408 method that lies somewhere between concatenation and simple gene-tree analysis: edge-based
409 analyses of genes. This approach limits the consideration of gene-tree topologies to those
410 containing a specific bipartition. In this study we have limited the bipartitions under
411 consideration to those present either in published analyses that were identified using
412 concatenated analyses (see Table 1) or those present in large numbers of gene trees. We used two
413 complementary criteria to examine the results of these edge-based analyses: $\sum\Delta\ln L$ and $\sum$Ng.
414 Comparison of the results given these criteria can provide evidence for outlier genes; when
415 $\sum\Delta\ln L$ for a specific edge is much larger than $\sum$Ng it implies that a relatively small number of
416 genes are contributing a large amount of the $\ln L$ differences among alternative resolutions.
417 Overall, this edge-based approach should limit GTEE while still retaining the advantages of gene
418 tree-based analyses and provide information about conflicting signals that might be present in the
419 phylogenetic datasets.

420 We examined gene-wide support for alternative resolutions of the three focal edges using
421 the filtered Jarvis et al. (2014) data (Table S1). We primarily focused on the dominant alternative
422 resolutions that have been previously published (Table 1). However, we found that the ML gene
423 trees typically exhibited topologies different from published dominant relationships (Figs. S2 and
424 S3). Thus, for each edge, we determined whether to exclude a gene by examining whether its
425 corresponding node on the ML tree is strong conflict with all the testing relationships using 95%

426 UFBS (Minh et al. 2013) threshold (see above, in section 2.3 of the Methods, for additional
427 details). After examining the compatibility of each ML gene tree with our test hypotheses, we
428 obtained 4194 – 5019 exons, 1134 – 1274 introns, and 3097 – 3295 UCEs for each focal
429 relationship (Table S1).
430 When we tested the set of five published hypotheses regarding the earliest diverging
431 Neoaves (Fig. 3A and Table 1) we found that all data types had the highest support for placing
432 doves sister to all other Neoaves (Figs 2B and S2). When we expanded the set of hypotheses
433 regarding the earliest diverging Neoaves to include the most prevalent topologies in the gene
434 trees (Fig. 3A, hypotheses 6–10) we found that support was more evenly distributed across the
435 hypotheses (Fig. 3A). For the other two edges, we found that Hoatzin were sister to shorebirds
436 (*S*), and owls were sister to mousebirds (Figs. 3B and 3C). Previous studies (e.g., Hackett et al.
437 2008; Jarvis et al. 2014) have seldom provided support for the best supported hypotheses in this
438 study. Exceptions include owls + mousebirds (*M*), found in a reanalysis of shallow-filtered Jarvis
439 UCE data (Gilbert et al. 2018), doves sister to all other Neoaves in the UCE indel tree (Houde et
440 al. 2019), and Hoatzin + *S* in the complete UCE data matrix (McCormack et al. 2013). Moreover,
441 the relationships published in the Jarvis et al. (2014) TENT tree (i.e., **VI+VII** as earliest Neoaves,
442 *S*+*CR* [shorebirds + cranes and rails] sister to Hoatzin, and *M*+*W* [mousebirds + woodpeckers
443 and allies] sister to owls) were only weakly supported (Figs. 2B and 3B and 3C).
444 To control for the impact of taxon sampling on topological preference, we conducted
445 similar analyses using the Prum et al. (2015) data, which include 198 bird species. To reduce the
446 CPU time necessary for analyses of the Prum data we only used the hypotheses from Prum et al.
447 (2015) (i.e., **V** as earliest Neoavian branch, **I** sister to Hoatzin, and *M*+*W* sister to owls) and the
448 top two published hypotheses supported by the Jarvis et al. (2014) data (i.e., **VII** and doves sister
449 to all other Neoaves, *S* and **V** sister to Hoatzin, and *M* and *EHV* sister to owls). After examining
450 the compatibility of the ML gene tree with the testing constraints based on 70% BS (Hillis and
451 Bull 1993), we obtained 231 – 258 loci in the edge-based analyses for each focal edge. The best
452 relationships supported by the Prum data were also supported in our analyses of the Jarvis data
453 (e.g., Hoatzin + *S*; owls + *M*, Table S2), indicating a limited taxon-sampling effect on edge-
454 based analyses. Taken as a whole, edge-based analyses increase congruence between analyses of
455 the Prum and Jarvis data. Moreover, analyses focused on the earliest diverging Neoaves using
456 the Prum data revealed a major impact of genes with a large $\Delta \ln L$. For example, we noted that

457 gene L111 encoding the FH2 domain containing 1 (FHDC1, whose homolog may play a role in
458 both cell shape and microtubule polymerization in humans), disproportionately contributed 171
459 $\Delta \ln L$ to support **VII** as the earliest divergence of Neoaves. We did not detect alignment error or
460 lineage specific positive selection for this gene. Moreover, eliminating FHDC1 would not shift
461 the best supported hypothesis given the $\sum \Delta \ln L$ criterion (Table S2). Regardless of the details, the
462 best-supported hypothesis in these edge-based analyses of the Prum data were never the
463 relationships present in the Prum et al. (2015) tree; they were hypotheses found in edge-based
464 analyses of the Jarvis data.
465
466 ***3.4 Complex data type effects revealed by categorical edge-based analyses***
467 The relatively even distribution of support for various hypotheses regarding the earliest diverging
468 Neoaves when we examined 10 alternatives (Fig. 3A) suggests a very high degree of conflict
469 among gene trees. However, this result also suggests a hypothesis-sampling-effect (i.e., the more
470 hypotheses tested, the more widely distributed the support). It also provides evidence for a data
471 type effect, where identified that the best hypotheses varied among datatypes; specifically, exons
472 supported cuckoos sister, introns continued to support doves sister, and UCEs supported either
473 clade **VII** (flamingos + grebes; best supported using the $\sum \Delta \ln L$ criterion) or cuckoos (best
474 supported using the $\sum$Ng criterion). It is noteworthy that Jarvis et al. (2014) placed **VII** sister to
475 all other Neoaves in their concatenated analysis of UCEs and the $\sum \Delta \ln L$ criterion, as the sum of
476 the differences in log-likelihood, is similar to the criterion used for tree selection in ML analyses.
477 However, the observation that UCEs supported cuckoos sister when we used our alternative
478 criterion, $\sum$Ng (the number of supporting genes), suggests support for **VII** sister in analyses of
479 UCEs reflects a limited number of UCE loci with large $\Delta \ln L$ values. Indeed, support given the
480 UCE data and the $\sum \Delta \ln L$ criterion was highest for the **VII** sister hypothesis with 16.5% of the
481 total $\Delta \ln L$ whereas the $\sum$Ng criterion placed **VII** sister in fifth place with 12.6% of loci. There
482 were other differences among data types in the rank order of suboptimal hypotheses. For
483 example, when the $\sum \Delta \ln L$ criterion is used **VI + VII** as the earliest Neoavian branch received
484 much higher support in analyses of non-coding data, especially introns, than it did in analyses of
485 coding data (10.9% for introns, 7.4% for UCEs, and 2% for exons). These results agree with the
486 data type effects reported by Reddy et al. (2017), although the differences among data types also
487 appear to be more complex than the results of that study suggested.

488 Reddy et al. (2017) pointed out that data type had a major impact on the topology for the
489 deepest divergence in Neoaves, with analyses of coding data supporting a **V** sister topology and
490 analyses of non-coding data supporting either **VI+VII** sister (introns) or **VII** sister (UCEs). Both
491 Jarvis et al. (2014) and Reddy et al. (2017) pointed out that avian coding data typically exhibit
492 greater GC_{CV} than do non-coding data. Previous studies (e.g., Foster 2004; Jermiin et al. 2004)
493 have demonstrated that high GC_{CV}, when fit to a homogeneous model, can results in biased
494 phylogenetic estimation. Applying models with more heterogeneous assumptions (e.g., Galtier
495 and Gouy 1998; Holland et al. 2013) has the potential to improve estimates of phylogeny,
496 although the available programs that implement these models are challenging to apply to
497 phylogenomic datasets (see additions discussion in Reddy et al. 2017). Since our edge-based
498 analyses provide a sensitive tool to examine conflicting signals in phylogenomic datasets we
499 assessed the impact of GC_{CV} in the Jarvis et al. (2014) data on the conflicts among data types that
500 we observed.
501 Here, we examined parsimony-informative sites (as in Reddy et al. 2017 and Braun et al.
502 2019) using three schemes, which we called ΔGC_{50}, ΔGC_{90}, and ΔGC_{100}. These values reflect the
503 difference between taxa with different GC contents (i.e., ΔGC_{90} is the GC difference between the
504 95 and 5 percentile GC-rich taxa). ΔGC_{50} was similar among data types but ΔGC_{90} and ΔGC_{100}
505 variation was larger for exons than for introns and UCEs, consistent with findings in Reddy et al.
506 (2017) (Fig. 4a). We ordered genes based on ΔGC_{90} to maximize differences while eliminating
507 outlier signals (Fig. 4a) and compared the edge analyses results between genes from either the
508 upper (i.e., high GC_{CV}) or the lower (e.g., low GC_{CV}) quartiles of the variation. This results in
509 about 1266 exons, 352 introns and 842 UCEs in each category. If GC_{CV} has an impact on
510 inference we would expect different patterns of support between GC_{CV} categories within the
511 same data type. Indeed, owls + *EHV* (eagles, hawks and vultures) showed much higher support
512 (both $\sum$Ng and $\sum\Delta\ln L$) in all high GC_{CV} makers (Figs. 4b and S5), and such signal is strongest in
513 EH and IH. It seems the characteristics of high GC-variation genes tend to correlate with owls +
514 *EHV*. Braun et al. (2019) suggested that the position of owls in concatenated analyses was
515 subject to a data type effect, with analyses of coding data having a tendency to yield the owls +
516 *EHV* topology. We found that support for owls + *EHV* was similar for the EL, IL, and UL
517 datasets (low GC_{CV} exons, introns, and UCEs, respectively) but much higher in the EH, IH, and
518 UH (the high GC_{CV} datasets). However, the signal for owls + *EHV* was still stronger in the high

519 GC_{CV} exons than in the high GC_{CV} introns and UCEs. Thus, our results imply that an interaction
520 between higher GC_{CV} and data type– not coding vs non-coding data type *per se* – is correlated
521 with the owls + *EHV* signal.
522 As previously noted, the position of Hoatzin has been highly variable across studies and
523 data type effects have not been noted in prior studies (reviewed by Braun et al. 2019). The
524 dominant signal in our categorical edge-based analyses was Hoatzin + *S* (shorebirds), which is
525 also present in the Jarvis exon c123 tree and the McCormack et al. (2013) UCE tree. In the
526 categorical edge-based analyses, there was higher support for Hoatzin + *S* in high GC_{CV} markers
527 (compare EH vs EL, IH vs IL, and UH vs UL; Fig. 4B), but the support differences are quite
528 trivial. In fact, support for alternative relationships of the Hoatzin show more variations among
529 data types (Figs. 4B and S4) regardless of GC_{CV} and tree length. The most notable difference
530 was a slightly lower signal supporting Hoatzin + *S* and a higher signal for Hoatzin + **V**
531 (Caprimulgimorphae) in the introns (Fig 4B and 4D). Provocatively, Hoatzin + **V** is present in
532 the Jarvis intron tree (albeit with only 70% bootstrap support) but it is not in the Jarvis UCE tree
533 (which included Hoatzin + *S* + *CR*, like the Jarvis total evidence nucleotide tree). These results
534 are not consistent with a simple coding vs noncoding effect, as (1) the signal for Hoatzin + *S* is
535 very similar in exons and UCEs (Fig. 4B) whereas the Hoatzin + **V** relationship was elevated in
536 both noncoding data types, and (2) Hoatzin + **I** (core landbirds) – which was found in the Prum
537 et al. (2015) study that largely used coding sequences – received higher support in introns and
538 UCEs (Figs. 4B and S5) than it did in Jarvis' exons (Fig. 4). Thus, in agreement with many
539 previous studies, this study finds that the position of the Hoatzin is very difficult to resolve.
540 However, our edge-based analyses revealed a previously unappreciated whereas potentially more
541 complex data type effect and further showed that Hoatzin + *S* was the strongest signal in all three
542 data types.
543 The most prominent data type effect hypothesized by Reddy et al. (2017) was the **VI +**
544 **VII** sister to all other Neoaves topology (their non-coding topology) and **V** sister to all other
545 Neoaves (their coding hypothesis). Using our categorical edge-based analyses, **V** as sister to all
546 other Neoaves did not appear to have better support in exons than in introns or UCEs (Figs 4B
547 and S5). The data subset with the highest support for **V** as the earliest neoavian branch was the
548 high GC_{CV} UCEs (UH in Fig 4B). This support for **V** sister appears to result from one outlier
549 gene (chr1_28895_s) that contributes almost 40% of the $\sum \Delta \ln L$ (Table S3) for this relationship.

550 Unsurprisingly, given the evidence for an outlier gene in the UH dataset, the total number of
551 genes (∑Ng) that support **V** did not vary much among GC_{CV} categories or data types (Fig. S5).
552 As described above **VI+VII** did appear to be better supported in introns and UCEs, especially in
553 the data subsets with high GC_{CV} and high tree lengths (i.e., IH and UH, Figs 4B and 4D).
554 However, the number of genes (∑Ng) supporting **VI+VII** did not vary much among categories
555 and data types (Fig. S5). Thus, the higher support for **VI+VII** sister appears to reflect a relatively
556 small number of genes that make high $\Delta \ln L$ contributions; given the much higher proportion of
557 signal for **VI+VII** sister in the IH and UH subsets (Fig. 4B) these genes with a high $\Delta \ln L$
558 contribution must be especially common in the high GC_{CV} data subsets.
559 Evolutionary rate is another major factor in phylogenetic resolution (Chojnowski et al.
560 2008). Tree length (sum of branch lengths) variation suggests that different overall rates of
561 evolution and different data types exhibited different patterns, with introns having larger tree
562 length/internode lengths and less variation than UCEs and exons (Fig. 4C). Tree length in these
563 datasets is primarily the result of tip lengths as internodes were very short across data types (Fig.
564 4C) as a consequence of sampling at high taxonomic levels. We sorted genes based on their total
565 tree length and compared support for focal edges with the upper and lower quartile sets. In
566 general, tree lengths did not have an obvious impact on support for different hypotheses for each
567 focal edge. However, we noticed two interesting patterns with respect to the earliest diverging
568 Neoaves (Fig. 4D). First, **VI+VII** as sister to all other Neoaves usually received higher support
569 in categories with high GC_{CV} and high tree lengths, but we noted that **VI+VII** also received high
570 support (i.e., the best) in the intron subset of low tree lengths (IL, Fig. 4D). Within this category,
571 we identified one gene (e.g., 11117.intron, Table S3) contributing ~21% of the $\sum \Delta \ln L$ supporting
572 this relationship. We did not detect any alignment issue in this gene so it is possible that the
573 biological characteristics of this gene drove an outlier signal to support this relationship, which
574 may not be directly related to evolutionary rate as the tree length of this gene is around the upper
575 edge of this quartile set. Second, **VII** as sister to all other Neoaves received the highest support
576 from the UCE subset of high tree length (UH, Fig. 4D), consistent with the Jarvis et al. (2014)
577 UCE tree. Again, an outlier gene (chr8_1881_s, that contributed 1/3 (666 $\Delta \ln L$) to the $\sum \Delta \ln L$,
578 Table S3) was identified, and its support for **VII** as sister may be driven by certain characteristics
579 that may or may not be directly related to its high total tree length.

580 As found previously, genes that have disproportionally high support for specific
581 relationships (i.e., outlier genes) can bias phylogenetic reconstruction (Kimball et al. 2013; Shen
582 et al. 2017, Walker et al. 2018). We noted several potential outliers within these datasets (Table
583 S3). As indicated above, some outliers can significantly support certain hypotheses. However,
584 removal of most single outlier genes did not change the results significantly nor could we detect
585 any alignment errors with these genes. The diversity of alternative hypotheses noted earlier made
586 it difficult to detect the underlying reasons for outliers biased support. One biological process
587 that might cause biased phylogenetic resolution is selection. However, we could not test
588 selection as most exons contained premature stop codons (e.g., single mutation likely caused by
589 sequencing or assembly error). Both introns and UCEs also contain outliers. Although it was
590 impossible to link gene specific traits to the topology they support without annotation, our
591 categorical edge-based analyses indeed implied that the earliest diverging taxa in Neoaves might
592 be influenced by outlier signal. In contrast to the case for the earliest diverging Neoaves, we did
593 not detect as strong of an impact from outlier genes on the other two focal edges (the positions of
594 Hoatzin and owls).
595
596 ***3.5 Simulations***
597 Most of the gene-tree topologies resulted in a single species/lineage as sister to the remaining
598 lineages (Figs. 1B and S3). To determine if this may be expected given that short internodes
599 should have a higher prevalence of ILS, we conducted coalescent simulations using the
600 *TENT.MP-EST.binned* tree from Jarvis et al. (2014). This tree has a clade **VI+VII** as the earliest
601 diverged Neoaves, the Hoatzin groups with a large clade including **II**+**III**+ *S+CR*, and owls are
602 sister to *EHV*. Simulations were conducted twice, with parameters drawn from the same
603 parameter distribution for each hypothetical data type (see details in method, Table 2). We found
604 the overall level of ILS to be relatively high (~58% ARFD between the simulated true gene tree
605 and the true species tree [*TENT.MP-EST.binned* tree]). The GTEE appears to be high as well,
606 showing ~68% normalized RF distances between the true gene tree and the estimated ML gene
607 tree, which implies low phylogenetic signal per gene for reconstructing this rapid Neoavian
608 radiation. Even though coalescent species-trees reconstructed with ASTRAL III using simulated
609 data did recover > 90% of the true relationships across the tree, the concatenation analyses only
610 recovered ~85% of the true relationships (Table S4). However, the positions of the Hoatzin and

611 the earliest diverging neoavian lineage differed from the true topologies 61% and 44% of the
612 time, respectively, possibly due to the high GTEE and ILS.
613 After examining the estimated ML gene-trees, we found that the most common
614 topologies for each focal edge were likely to have a single species/lineages sister to the
615 remaining clades (Fig. S6), implying that ILS may be one source of this preference. There were
616 ~20% estimated ML gene-trees showing non-monophyletic Palaeognathae and Galloanserae.
617 The coalescent branch length on the true species tree (*TENT.MP-EST.binned* tree) is long (2.03)
618 and even longer using indel analyses (2.57, Houde et al. 2020), so the expected number of deep
619 coalescences that separate Galloanserae and Palaeognathae should be ~5% – 9%. Under this
620 condition, the high occurrence of non-monophyletic Palaeognathae and Galloanserae in the
621 simulated ML gene-trees is presumably caused by GTEE. We conducted edge-based analyses
622 with the 184 bipartitions from the 28 Jarvis et al. (2014) species-trees representing the extent of
623 the variation of potential species-tree hypotheses. Among these bipartitions, there were 63
624 (earliest Neoaves), 45 (Hoatzin sister), and 14 (owl sister) conflicting hypotheses for each focal
625 edge respectively (Table 2). The best supported hypotheses usually differed from the true
626 relationships, which is possibly caused by high GTEE and ILS (Table 2). For example, for the
627 earliest diverging Neoaves, 25 and 28 hypotheses obtained some support from the two sets of
628 simulated exons, respectively. Among them, the true topology (i.e., **VI+VII** as earliest Neoaves)
629 is only supported by one or five genes and ranked as the 14th or 18th among the resolutions. The
630 situation was similar for simulated introns and UCEs (Table 2) and for the other two focal edges:
631 the true topology ranks 10th – 17th among the resolutions of the position of the Hoatzin, and
632 ranks second for the position of owls (Table 2) across all data types.
633 In contrast to the focal edges that exhibited extremely short internodes and dramatic gene
634 tree conflicts, we also examined less contentious edges with longer internodes (i.e., those with a
635 low level of ILS). For example, clades including the core landbirds (**I**), the core waterbirds (**II**),
636 the Caprimulgimorphae (**V**), and the tropicbirds + sunbittern (**III**), have been consistently
637 recovered by previous studies (e.g., Reddy et al. 2017), and are connected by long internal
638 branches (stems) in the true tree. These received high support across the simulations regardless
639 of data type (Table 2), which suggests that the edge-based analyses were able to recover the
640 correct relationships when there was enough phylogenetic signal and weak ILS. These analyses

641 suggest that strong ILS and high GTEE remain a difficult problem, and may result in biased
642 resolutions for gene trees.
643
644 ***3.6 Within gene region conflict***
645 To determine whether gene tree conflict was the result of conflicting signal within genomic
646 regions, we further examined phylogenetic signal across regions within each gene. The average
647 gene-region length of the Jarvis et al. (2014) was 1998 bp (501 – 15819 bp) for exons, 2509 bp
648 (1787 – 3527 bp) for UCEs, and s 9309 bp (506 – 66,762 bp) for introns. For Prum et al. (2015),
649 the average gene region length was 1524 bp (361 – 2316 bp). We focused on exon and UCE
650 alignments as they were smaller and therefore less likely than introns to suffer from mixed signal.
651 We conducted analyses on each gene region to determine whether 1000 bp segments conflicted
652 with the ML tree calculated for the entire region (see methods for detailed splitting strategy). We
653 found extensive conflicting signal after collapsing edges with < 95% UFBS (Minh et al. 2013)
654 and < 80% SH-aLRT (approximate likelihood-ratio test; Guindon et al. 2010) support. Although
655 this is an extremely conservative measure, and likely to have a high false-negative rate (Guindon
656 et al. 2010), it was chosen to prevent false positives. For loci with at least 2 × 1000 bp segments,
657 we found that 10% of Jarvis et al. (2014) exons (195 of 1985), 6% (207 of 3367) of UCEs, and
658 50% (8 of 15) of gene regions in Prum et al. (2015) had conflicts. It is noteworthy that exons in
659 Jarvis et al. (2014) are usually a combination of multiple exonic regions of one gene, with some
660 regions separated by long introns. These regions might exhibit different evolutionary histories
661 due to recombination, a different selection level, and/or for other reasons. Although intragenic
662 conflicts within the Prum et al. (2015) data were much higher they could reflect sampling error
663 due to the small number (i.e., only 15) of long gene regions in the data (the larger number of taxa
664 in the Prum et al. 2015 dataset is also likely to have had an impact).
665 To determine whether within gene conflict corresponded to obvious data "dirtiness", we
666 examined alignment patterns within the conflicting gene regions. In general, some loci that
667 exhibited substantial intragenic conflicts can still have alignments that appear largely
668 unambiguous (e.g., neat alignments in locus 9838.exons, Fig. S7), which suggests there may be
669 biological processes underlying these intragenic conflicts. However, other intragenic conflicts
670 might be due to misalignment patterns. For example, the 1424.exon gene region from Jarvis et al.
671 (2014) had an alignment that was obviously "messy" alignments (Fig. S7); analyses of

problematic segments placed Rifleman sister to pelicans, mesites sister to cranes, grebes sister to sandgrouse, and fulmars sister to seriemas.

While it is possible for these intragenic conflicts to be the result of biological processes (Scornavacca and Galtier 2017; Mendes et al. 2019), the most likely issue is data curation given the volume of both datasets (Springer and Gatesy 2018; Braun et al. 2019). All large datasets suffer from some amount of noisy data. However, because of the biological nature of the bird radiation with short intervals among many early speciation events, when the true phylogenetic signal is likely to be weak, the importance of data cleanliness is even more important.

4. Conclusions

The categorical edge-based analyses conducted herein provide a way to examine differences in signal associated with subsets of the data and assess the best-supported resolution of specific nodes. The best-corroborated resolutions of our focal nodes were: 1) owl + *M* (mousebirds); 2) Hoatzin + S (shorebirds); and 3) doves are the earliest diverging Neoaves. However, the edge-based analyses revealed substantial conflict; indeed, they provided evidence for a larger amount of conflict than the data-type analyses conducted by Jarvis et al. (2014) or Reddy et al. (2017). First, the owls + *EHV* signal reflects a combination of higher GC_{CV}, and data type rather than data type *per se*. Second, the position of Hoatzin is highly variable but there is a data type effect that was not evident in previous studies. Specifically, support for Hoatzin + **V** was strongest in introns, intermediate in UCEs, and weakest for exons. Finally, data type effects were evident for the earliest diverging Neoaves, as expected based on Reddy et al. (2017), but outlier genes also appeared to play an important role in this part of the tree. Perhaps surprisingly, there was not a single factor (e.g., GC_{CV} or data type) that played an important role in the resolution of all three focal nodes. Thus, simple solutions like focusing on a specific data type or eliminating genes with high GC_{CV} are unlikely to be a panacea for analyses of avian phylogeny (or, by extension, other nodes in the Tree of Life that have proven to be recalcitrant). We believe that our findings will facilitate further exploration of phylogenetic conflicts using an edge-by-edge approach.

Perhaps one of the most surprising results of this study is the fact that the results of some of our edge-based analysis contradict the magnificent seven – the superordinal groups that have been consistently identified on the Neoavian tree by recent large-scale studies (Reddy et al. 2017). The relatively consistent recovery of these groups in large-scale analyses has led to the

703 view that most or all of these clades are present in the true avian species tree (Suh 2016; Braun et
704 al. 2019). However, edge-based analyses indicate that at least some of those groups might have
705 less support than we currently believe. For example, signal placing either cuckoos or doves sister
706 to all other Neoaves is stronger than many other alternatives (Figs. 3 and 4). Both signals
707 contradict magnificent seven clades; specifically, doves as the earliest divergence contradicts
708 Columbimorphae (group **VI**; doves + mesites + sandgrouse; Fig. 2A) and placing cuckoos sister
709 to all other Neoaves contradicts the Otidimorphae (group **V**; cuckoos + turacos + bustards; Fig.
710 2A). There are several interpretations of this result. First, both magnificent seven clades could be
711 present in the true tree but our edge-based analyses reflect the fact that the branches uniting those
712 clades are especially short and, therefore, "noisy." Arguably, this is the simplest hypothesis
713 given that previous studies (e.g., Jarvis et al. 2014; Prum et al. 2015) agree that the relevant
714 edges are quite short. Second, the fact that signals contradicting those magnificent seven clades
715 are stronger than signals consistent with those clades could provide evidence that the
716 Columbimorphae and Otidimorphae are present but there is something unusual about the
717 branches uniting those clades. In fact, Houde et al. (2020) examined indel data and they reported
718 that conflicts among indels for Columbimorphae were higher than expected given the relevant
719 edge length in the Jarvis et al. (2014) timetree; they interpreted this as evidence for a transient
720 increase in the effective population size for the common ancestor of that clade. Finally, our edge-
721 based analyses can be viewed as evidence that Columbimorphae (if the doves sister hypothesis is
722 correct) or Otidimorphae (if the doves sister hypothesis is correct) are not present as clades in the
723 true bird tree. Regardless of the best interpretation of the edge-based analyses, it is clear that they
724 provide a way to highlight especially problematic branches that warrant further examination in
725 future phylogenomic studies.
726 Given the evidence for conflict evident in our edge-based analyses we are left with a
727 fundamental question: what is the best-way forward for resolving the nodes in the bird tree that
728 remain uncertain? Obviously, collecting even more data will be important and efforts like the
729 B10K project (Zhang 2015; Stiller and Zhang 2019) and the OpenWings (Kimball et al. 2019)
730 will provide these larger datasets for birds. However, there is ample evidence that simply
731 expanding the size of phylogenomic datasets will not solve the problem of recalcitrant nodes in
732 avian phylogeny (or the difficult nodes in other parts of the tree of life). Rare genomic changes
733 (Rokas and Holland 2000) may provide an alternative data type that sidesteps the complexities

evident in the analyses we present here. In fact, rare genomic changes can avoid (or at least reduce) problems like intralocus recombination, natural selection, and GTEE (Springer and Gatesy 2019). The availability of genome-scale datasets will only increase availability of rare genomic change data and allow the community to move beyond retroposon insertions, which were used even before the phylogenomic era (e.g., Suh et al. 2011), to types of rare genomic changes like numts (Liang et al. 2018) and microinversions (Braun et al. 2011) that have not been used in many phylogenetic studies. These types of data, which are very different from nucleotide sequence data, have the potential to corroborate (or falsify) the hypothesis that the most common resolution in edge-based analyses represents a correct resolution of the species tree. Regardless of whether the hypothesis with the greatest support (whether judged using $\sum\Delta\ln L$ and $\sum$Ng) is correct, it is clear that the edge-based analyses reveal substantial conflicts within the data. Understanding those conflicts is an important problem in evolutionary biology even if we do not know the true resolution of the tree. We can learn even more if the information about conflicts among genes, like those revealed by our edge-based analyses, can be combined with an accurate estimate of the species tree.

Supplementary material

Supplementary figures and tables can be found online.

Acknowledgements

We would like to thank Benjamin Winger, Joseph Walker, Nathanael Walker-Hale and Diego Alvarado-Serrano for their constructive comments on the earliest version of the paper. We appreciate other comments from members of the Smith lab. This work was supported by grants from the US National Science Foundation (DBI-1458466 and DEB-1207915 to S.A.S., DEB-1241066 and DEB-1146423 to J.C., and DEB-1655683 to E.L.B.) and by the Frank M. Chapman Memorial Fund of the American Museum of Natural History.

Author contributions

Study conception and design: N.W., S.A.S.; Acquisition of data: N.W.; Analysis and interpretation of data: N.W., S.A.S., E.L.B., J.C., B.L.; Drafting of manuscript: N.W., E.L.B., J.C., S.A.S.; Image illustration: B.L., N.W., E.L.B., J.C.

Competing interests

The authors have no competing interests declared.

References

Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, Froula J. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME* 10(7):1589.

Braun EL, Kimball RT, Han KL, Iuhasz-Velez NR, Bonilla AJ, Chojnowski JL, Smith JV, Bowie RC, Braun MJ, Hackett SJ, Harshman J. 2011. Homoplastic microinversions and the avian tree of life. *BMC Evol Biol.* 11(1):141.

Braun EL, Cracraft J, Houde P. 2019. Resolving the avian tree of life from top to bottom: the promise and potential boundaries of the phylogenomic era. In Avian Genomics in Ecology and Evolution (pp. 151-210). Springer, Cham.

Bryant D, Hahn MW. 2020. The Concatenation Question. Scornavacca, Celine; Delsuc, Frédéric; Galtier, Nicolas. Phylogenetics in the Genomic Era, No commercial publisher | Authors open access book, pp.3.4:1–3.4:23. hal-02535651f.

Chen D, Braun EL, Forthman M, Kimball RT, Zhang Z. 2018. A simple strategy for recovering ultraconserved elements, exons, and introns from low coverage shotgun sequencing of museum specimens: Placement of the partridge genus *Tropicoperdix* within the galliformes. *Mol Phylogenet Evol.* 129:304–314.

Chojnowski JL, Kimball RT, Braun EL, 2008. Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. *Gene* 410(1):89–96.

Dominguez-Bello MG, Michelangeli F, Ruiz MC, Garcia A, Rodriguez E. 1994. Ecology of the folivorous hoatzin (*Opisthocomus Hoatzin*) on the Venezuelan plains. *Auk* 111(3):643–651.

Duchêne DA, Tong KJ, Foster CS, Duchene S, Lanfear R, Ho SY. 2018. Linking Branch Lengths Across Loci Provides the Best Fit for Phylogenetic Inference. *bioRxiv* 467449.

Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63(1):1-19.

795 Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM,
796 Lemmon AR, Leaché AD. 2016. Implementing and testing the multispecies coalescent model: a
797 valuable paradigm for phylogenomics. *Mol Phylogenet Evol.* 94: 447-462.
798 Ericson PG, Anderson CL, Britton T, Elzanowski A, Johansson US, Källersjö M, Ohlson JI, Parsons TJ,
799 Zuccon D, Mayr G. 2006. Diversification of Neoaves: integration of molecular sequence data and
800 fossils. *Biol Lett.* 2(4):543–547.
801 Foster PG. 2004. Modeling compositional heterogeneity. *Syst Biol.* 53:485–495.
802 Foster PG, Hickey DA. 1999. Compositional bias may affect both DNA-based and protein-based
803 phylogenetic reconstructions. *J Mol Evol.* 48:284–290.
804 Galtier N, Gouy M. 1998. Inferring pattern and process: maximum-likelihood implementation of a
805 nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol Biol Evol.*
806 15(7): 871-879.
807 Gatesy J, Springer MS. 2017. Phylogenomic red flags: Homology errors and zombie lineages in the
808 evolutionary diversification of placental mammals. *Proc Natl Acad Sci USA.* 114(45):E9431–
809 E9432.
810 Gilbert PS, Wu J, Simon MW, Sinsheimer JS, Alfaro ME. 2018. Filtering nucleotide sites by
811 phylogenetic signal to noise ratio increases confidence in the Neoaves phylogeny generated from
812 ultraconserved elements. *Mol Phylogenet Evol.* 126:116–128.
813 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and
814 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML
815 3.0. *Syst Biol.* 59(3):307–321.
816 Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han KL,
817 Harshman J, Huddleston CJ. et al. 2008. A phylogenomic study of birds reveals their evolutionary
818 history. *Science* 320(5884):1763–1768.
819 Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in
820 phylogenetic analysis. *Syst Biol.* 42(2):182–192.
821 Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018 UFBoot2: Improving the ultrafast
822 bootstrap approximation. *Mol Biol Evol.* 35:518–522.
823 Holland BR, Jarvis PD, Sumner JG. 2013. Low-parameter phylogenetic inference under the general
824 Markov model. *Syst Biol.* 62:78-92.
825 Hughes JM, Baker AJ. 1999. Phylogenetic relationships of the enigmatic hoatzin (*Opisthocomus Hoatzin*)
826 resolved using mitochondrial and nuclear gene sequences. *Mol Biol Evol.* 16(9):1300–1307.
827 Houde P, Braun EL, Narula N, Minjares U, Mirarab S. 2019. Phylogenetic signal of indels and the
828 neoavian radiation. *Diversity* 11(7):108.

Houde P, Braun EL, Zhou L. 2020. Deep-time demographic inference suggests ecological release as driver of Neoavian adaptive radiation. *Diversity* 12:164.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, Suh A. et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346(6215):1320–1331.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, Suh A. 2015. Phylogenomic analyses data of the avian phylogenomics project. *GigaScience* 4(1):4.

Jermiin LS, Ho SY, Ababneh F, Robinson J, Larkum AW. 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst Boil*. 53(4):638-643.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. *Nature* 491(7424):444.

Kimball RT, Wang N, Heimer-McGinn V, Ferguson C, Braun EL. 2013. Identifying localized biases in large datasets: a case study using the avian tree of life. *Mol Phylogenet Evol.* 69(3): 1021–1032.

Kimball RT, Oliveros CH, Wang N, White ND, Barker FK, Field DJ, Ksepka DT, Chesser RT, Moyle RG, Braun MJ, Brumfield RT, Faircloth BC, Smith BT, Braun EL. 2019. A phylogenomic supertree of birds. *Diversity* 11(7):109

Kobert K, Salichos L, Rokas A, Stamatakis A. 2016. Computing the internode certainty and related measures from partial gene trees. *Mol Biol Evol.* 33(6):1606-1617.

Kolaczkowski B, Thornton JW. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* 431(7011):980.

Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol.* 56(1):17–24.

Li Y, Liu Z, Shi P, Zhang J. 2010. The hearing gene Prestin unites echolocating bats and whales. *Curr Biol.* 20(2):R55-R56

Liang B, Wang N, Li N, Kimball RT, Braun EL. 2018. Comparative genomics reveals a burst of homoplasy-free numt insertions. *Mol Biol Evol.* 35(8):2060–2064.

Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol.* 10(1):302.

Liu L, Zhang J, Rheindt FE, Lei F, Qu Y, Wang Y, Zhang Y, Sullivan C, Nie W, Wang J, Yang F. 2017. Reply to Gatesy and Springer: Claims of homology errors and zombie lineages do not compromise the dating of placental diversification. *Proc Natl Acad Sci USA.* 114(45):E9433–E9434.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.* 22(4):746–754.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One* 8(1):e54848.

Mendes FK, Livera AP, Hahn MW. 2019. The perils of intralocus recombination for inferences of molecular convergence. *Philos Trans Royal Soc B.* 374(1777):20180244.

Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 30(5):1188–1195.

Mirarab S, Bayzid MS, Boussau B, Warnow T. 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science* 346(6215).

Molloy EK, Warnow T. 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. *Syst Biol.* 67(2):285-303.

Morgan CC, Foster PG, Webb AE, Pisani D, McInerney JO, O'Connell MJ. 2013. Heterogeneous models place the root of the placental mammal phylogeny. *Mol Biol Evol.* 30(9):2145–2156.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.

Pandey A, Braun EL. 2020. Phylogenetic Analyses of Sites in Different Protein Structural Environments Result in Distinct Placements of the Metazoan Root. *Biology* 2020, *9*:64.

Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526(7574):569.

Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH, Schneider H, Pisani D, Donoghue PC. 2018. The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr Biol.* 28(5):733–745.

Rambaut A, Grass NC. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics* 13(3):235–238.

Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, Han KL, Harshman J, Huddleston CJ, Kingston S, Marks BD. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Syst Biol.* 66(5):857–879.

Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. *Math Biosci.* 53(1-2):131–147.

Rokas A, Holland PW. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol.* 15(11):454–459.

895 Salichos L, Stamatakis A, Rokas A. 2014. Novel information theory-based measures for quantifying
896 incongruence among phylogenetic trees. *Mol Biol Evol.* 31:1261–1271.
897 Scornavacca C, Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Syst Biol.*
898 66(1):112–120.
899 Shen XX, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven
900 by a handful of genes. *Nat Ecol Evol.* 1(5): 0126.
901 Simmons M.P., Sloan D.B., Gatesy J. 2016. The effects of subsampling gene trees on coalescent methods
902 applied to ancient divergences. *Mol Phylogenet Evol.* 97:76–89.
903 Smith SA, Walker-Hale N, Walker JF, Brown J. 2020. Phylogenetic Conflicts, Combinability, and Deep
904 Phylogenomics in Plants. *Syst Biol.* 69(3):579–592.
905 Springer MS, Gatesy J. 2016. The gene tree delusion. *Mol Phylogenet Evol.* 94:1–33.
906 Springer MS, Gatesy J. 2018. On the importance of homology in the age of phylogenomics. *System*
907 *Biodivers.* 16(3): 210-228.
908 Springer MS, Gatesy J 2019. Retroposon insertions within a multispecies coalescent framework suggest
909 that ratite phylogeny is not in the 'Anomaly Zone'. *bioRxiv* p.643296.
910 Stamatakis A. 2006, April. Phylogenetic models of rate heterogeneity: a high performance computing
911 perspective. In Proceedings 20th IEEE international parallel & distributed processing symposium
912 Rhodos, Greece (pp. 8-pp). IEEE.
913 Stamatakis A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large
914 Phylogenies. *Bioinformatics* 30 (9):1312-1313.
915 Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet.* 14(11):751-764.
916 Stiller J, Zhang G. 2019. Comparative Phylogenomics, a stepping stone for bird biodiversity studies.
917 *Diversity* 11: 115.
918 Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zool*
919 *Scr.* 45:50–62.
920 Suh A, Paus M, Kiefmann M, Churakov G, Franke FA, Brosius J, Kriegs JO, Schmitz J. 2011. Mesozoic
921 retroposons reveal parrots as the closest living relatives of passerine birds. *Nat Commun.* 2:443.
922 Suh A, Linnéa S, Hans E. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive
923 radiation of neoavian birds. *PLoS Biol.* 13(8):e1002224.
924 Sukumaran J, Holder MT. 2010. DendroPy: A Python library for phylogenetic computing. *Bioinformatics*
925 26:1569–1571.
926 Tarver JE, Dos Reis M, Mirarab S, Moran RJ, Parker S, O'Reilly JE, King BL, O'Connell MJ, Asher RJ,
927 Warnow T, Peterson KJ. 2016. The interrelationships of placental mammals and the limits of
928 phylogenetic inference. *Genome Biol Evol.* 8(2):330–344.

929 Walker JF, Brown JW, Smith SA. 2018. Analyzing contentious relationships and outlier genes in
930 phylogenomics. *Syst Biol.* 67(5):916–924.
931 Wang N, Braun EL, Kimball RT. 2012. Testing hypotheses about the sister group of the Passeriformes
932 using an independent 30-locus data set. *Mol Biol Evol.* 29(2):737–750.
933 Wang N, Kimball RT, Braun EL, Liang B, Zhang Z. 2017. Ancestral range reconstruction of Galliformes:
934 the effects of topology and taxon sampling. *J Biogeogr.* 44(1):122–135.
935 Wang N, Yang Y, Moore MJ, Brockington SF, Walker JF, Brown JW, Liang B, Feng T, Edwards C,
936 Mikenas J, Olivieri J. et al. 2019. Evolution of Portulacineae marked by gene tree conflict and
937 gene family expansion associated with adaptation to harsh environments. *Mol Biol Evol.*
938 36(1):112–126.
939 Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS,
940 Burleigh JG, Gitzendanner MA, Ruhfel BR. et al. 2014. Phylotranscriptomic analysis of the
941 origin and early diversification of land plants. *Proc Natl Acad Sci USA.* 111(45):E4859–E4868.
942 Zhang GJ. 2015. Bird sequencing project takes off. *Nature* 522:34-44.
943 Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: Polynomial Time Species Tree
944 Reconstruction from Partially Resolved Gene Trees. *BMC Bioinformatics* 19(6):153.
945

Table 1. Three focal edges and major testing hypotheses that published before.

Focal Edges	Hypotheses	Major References
Earliest Neoaves	1. **VI+VII**	Jarvis et al. 2014; Reddy et al. 2017
	2. **V**	Jarvis et al. 2014, Prum et al. 2015
	3. **VII**	Jarvis et al. 2014
	4. **III+V+VI+VII**	Hackett et al. 2008
	5. Doves	McCormack et al. 2013
Hoatzin's sister	1. **I+III+V**	Reddy et al. 2017
	2. **I**	Prum et al. 2015
	3. *S*+*CR*	Jarvis et al. 2014
	4. **V**	Jarvis et al. 2014
	5. *S*	McCormack et al, 2013, Jarvis et al. 2014
	6. *TB*	Jarvis et al. 2014
	7. **II+III**+*S*+*CR*	Jarvis et al. 2014
Owls' sister	1. *M*+*W*	Jarvis et al. 2014; Prum et al. 2015; Reddy et al. 2017
	2. *W*	Jetz et al. 2012; McCormack et al. 2013; Jarvis et al. 2014
	3. *EHV*	Jarvis et al. 2014
	4. *M*	Hackett et al. 2008

NOTE.— the Greek numbers (bolded) correspond to the "magnificent seven" of Reddy et al. (2017), including **I**: core Landbirds; **II**: core Waterbirds; **III**: Sunbittern and Tropicbirds; **IV**: Cuckoos, Bustards, and Turacos; **V**: Nightjars, Swifts, Hummingbirds, and allies; **VI**: Doves, Mesites, and Sandgrouse; **VII**: Flamingos and Grebes. Other italic acronyms including *S*: Shorebirds; *CR*: Cranes and Rails; *TB*: Turacos and Bustards; *M*: Mousebirds; *W*: Woodpeckers and allies; *EHV*: Eagles, Hawks, and New World Vultures.

Table 2. Simulation results showing the best supported edge

Simulations	earliest Neoaves[1] (63[#])	TR	Hoatzin Sister[2] (45)	TR	Owls Sister[3] (14)	TR	**I** (33)	**II** (4)	**III** (16)	**V** (27)
exon1	Doves + Turacos	14	Sunbittern + Turacos	12	Owls + Mousebirds	2	**5432***	**19746**	**8497**	**5581**
exon2	Cuckoos + Sandgrouse	18	Sunbittern + Turacos	17	Owls + Mousebirds	2	**6438**	**15763**	**7757**	**5416**
intron1	Cuckoos + Doves	8	Sunbittern + Turacos	10	Owls + Mousebirds	2	**4639**	**21835**	**7672**	**6130**
intron2	Cuckoos + Sandgrouse	14	Sunbittern + Turacos	13	Owls + Mousebirds	2	**3805**	**14934**	**7122**	**6537**
UCE1	Cuckoos + Sandgrouse	14	Cranes + Flamingos + Grebes	13	Owls + Mousebirds	2	**4329**	**18031**	**7621**	**5733**
UCE2	Doves + Turacos	9	Hoatzin + Nightjars	17	Owls + Mousebirds	2	**5906**	**20108**	**7616**	**6508**

NOTE.— TR: the ranking of the true topology of each focal edge in the edge-based analyses. [#]: The number in the parenthesis represents the total number of conflicting constraints for each testing edge. [1]: showing the best hypothesis that is conflict with (**VI + VII**) + other Neoaves. [2]: showing the best hypothesis that is conflict with Hoatzin + **II** + **III** + *S* + *CR*. [3]: showing the best hypothesis that is conflict with Owls + *EHV*. Control edges include **I**, **II**, **III**, and **V**, which were all best supported by the edge-based analyses. *Bolded numbers: The difference of $\sum \Delta \ln L$ between the best hypotheses (i.e., **I, II, III, and V** respectively) and conflicting second best hypotheses for each edge. These large numbers indicate that the control edges can all be recovered with strong support. Clade keys can be found in Table 1 and Figure 2.

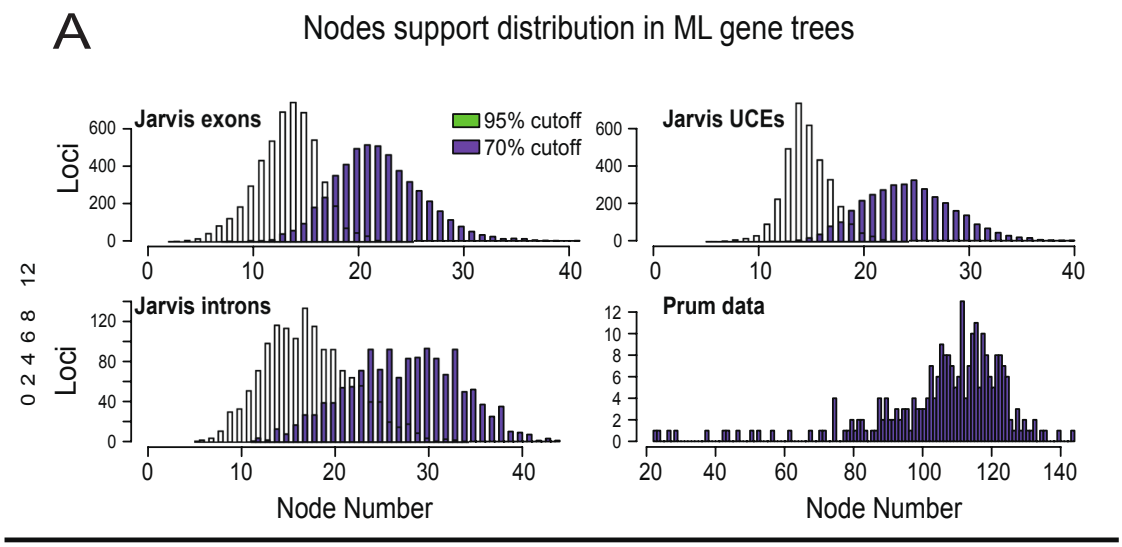
Figure Legends

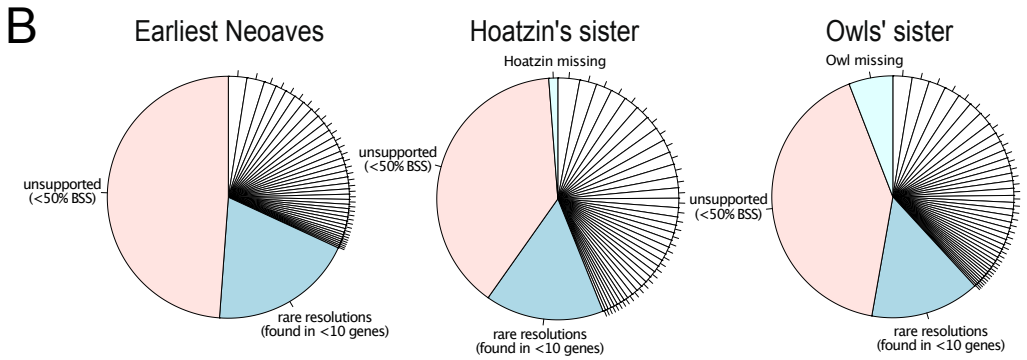
Figure 1. A: Node support distribution among ML gene trees. B: Topology distribution of the ML gene trees for the three focal edges using Jarvis et al. (2014) exon data. Analyses for Introns and UCEs can be found in Figure S3. unsupported: relationships shown have less than 50% UFBS (ultrafast bootstrap support, we use lower support cutoff to show potential topology distributions). The Top 20 supported topologies are shown. rare resolutions: relationships shown in less than 10 gene trees. Hoatzin missing or Owl missing: Hoatzin or Owl is missing in the gene trees.

Figure 2. A: Keys (**I-VII**) for each major clade in the hypotheses being tested. B: flow chart on how to conduct edge-based analyses by using the earliest diverging Neoaves as an example.

Figure 3. The three focal edges and the testing hypotheses in the edges-based analyses using exons (upper), introns (middle) and UCEs (lower). A: The earliest diverging clade in Neoaves, including the five dominant hypotheses published before (1–5), and the five additional hypotheses prevalent observed in the gene trees for bias assessment (6–10). B: the sister of Hoatzin, including seven previously identified hypotheses. C: the sister group of Owls, including four commonly found hypotheses. $\sum\Delta\ln L$: the sum of log likelihood score differences; $\sum$Ng: the sum of number of genes. Hypotheses keys can be found in Figure 2.

Figure 4. A: The GC content variation among exons, introns, or UCEs, including ΔGC_{50} (the GC difference between the inter-quartile GC-rich taxa), ΔGC_{90} (the GC difference between the 95 and 5 percentile GC-rich taxa), and ΔGC_{100} (the GC difference between most GC-rich taxon and the least GC-rich taxon). B: Comparison of hypotheses supporting pattern between gene sets from the lower and upper quartile of ΔGC_{90} among data types. C: Tree length (left) and internode length (right) variation among different types of data. D: Comparison of hypotheses supporting pattern between gene sets from the lower and upper quartile of tree length variation. For x axis, EL/EH: exons with lower or higher ΔGC_{90}/tree length; IL/IH: introns with lower or higher ΔGC_{90}/tree length; UL/UH: UCEs with lower or higher ΔGC_{90}/tree length.

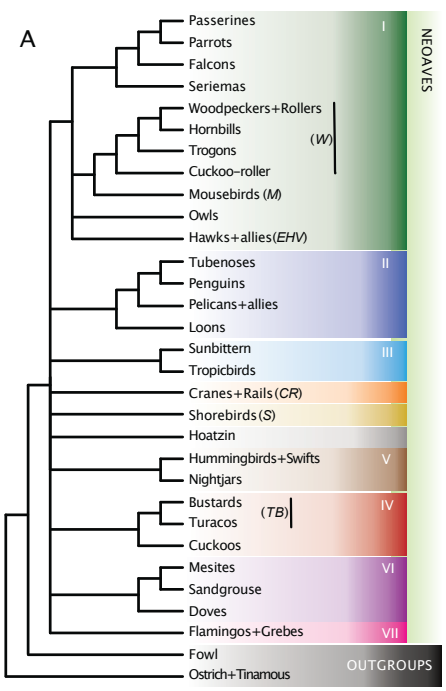




Top 20 topologies for each focal edge

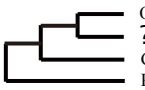
Earliest diverging Neoaves	No. exons	Hoatzin's sister	No. exons	Owls' sister	No. exons
Hummingbirds & Swifts	125	Sunbittern	148	Eagles	130
Cuckoos	119	Plovers	115	Falcons	116
Doves	83	Cranes	107	Cuckoo-roller	114
Bustards	72	Turacos	106	Seriemas	106
Sandgrouse	70	Sandgrouse	100	Trogons	94
Mesites	70	Doves	95	New World Vultures	92
Parrots	70	Nightjars	93	Mousebirds	91
Woodpeckers	65	Bustards	93	Parrots	75
Tropicbirds	61	Cuckoos	89	Hornbills	64
Nightjars	58	Tropicbirds	85	Tropicbirds	63
Hoatzin	55	Mesites	84	Nightjars	56
Mousebirds	52	Falcons	76	Cranes	53
Cranes	52	Trogons	70	Bee-eaters	52
Passerines	46	Loons	64	Sunbittern	51
Plovers	46	Herons	62	Plovers	51
Turacos	43	Cuckoo-roller	56	Bustards	48
Falcons	42	Cormorants	52	Sandgrouse	47
Sunbittern	42	Seriemas	52	Mesites	45
Loons	38	Flamingos & Grebes	50	Hoatzin	45
Bee-eaters	36	Flamingos	50	Loons	44
Rare resolutions	976	Rare resolutions	803	Rare resolutions	740
Unsupported	2472	Unsupported	1975	Unsupported	2092
		Hoatzin missing	61	Owl missing	298

Figure 1



B

Example Focal node



Hypotheses

1. VI+VII
2. V
3. VII
4. III+V+VI+VII
5. Doves

Edge-based analyses: Constraint focal bipartition; ML resolve other nodes

Loc/Hypo	1	2	3	4	5	**Best hypo**	**Δln*L****
gene1	-34363.52	**-34337.91**	*-34339.23*	-34408.11	-34343.69	2	1.33
gene2	-9928.07	-9916.19	*-9915.47*	-9937.53	**-9914.19**	5	1.28
gene3**	-30698.49	-30651.37	**-30039.67**	-30761.06	*-30432.68*	3	**393.01**
gene4	-8882.78	-8934.74	*-8881.29*	-8955.03	**-8876.95**	5	4.3397
…	…	…	…	…	…	…	…
Summarize							
∑Ng	73	1012	1561	3	**2413**		
∑Δln*L*	1497.43	6917.37	13922.42	0	**23684.10**		

*Δln*L*: ln*L* difference between the best (bolded) and the second best (italic) topology
**gene3: potential outlying gene supporting hypothesis 3
∑Ng: sum of number of gene supporting each hypothesis
∑Δln*L*: sum of Δln*L* supporting each hypothesis

Figure 2

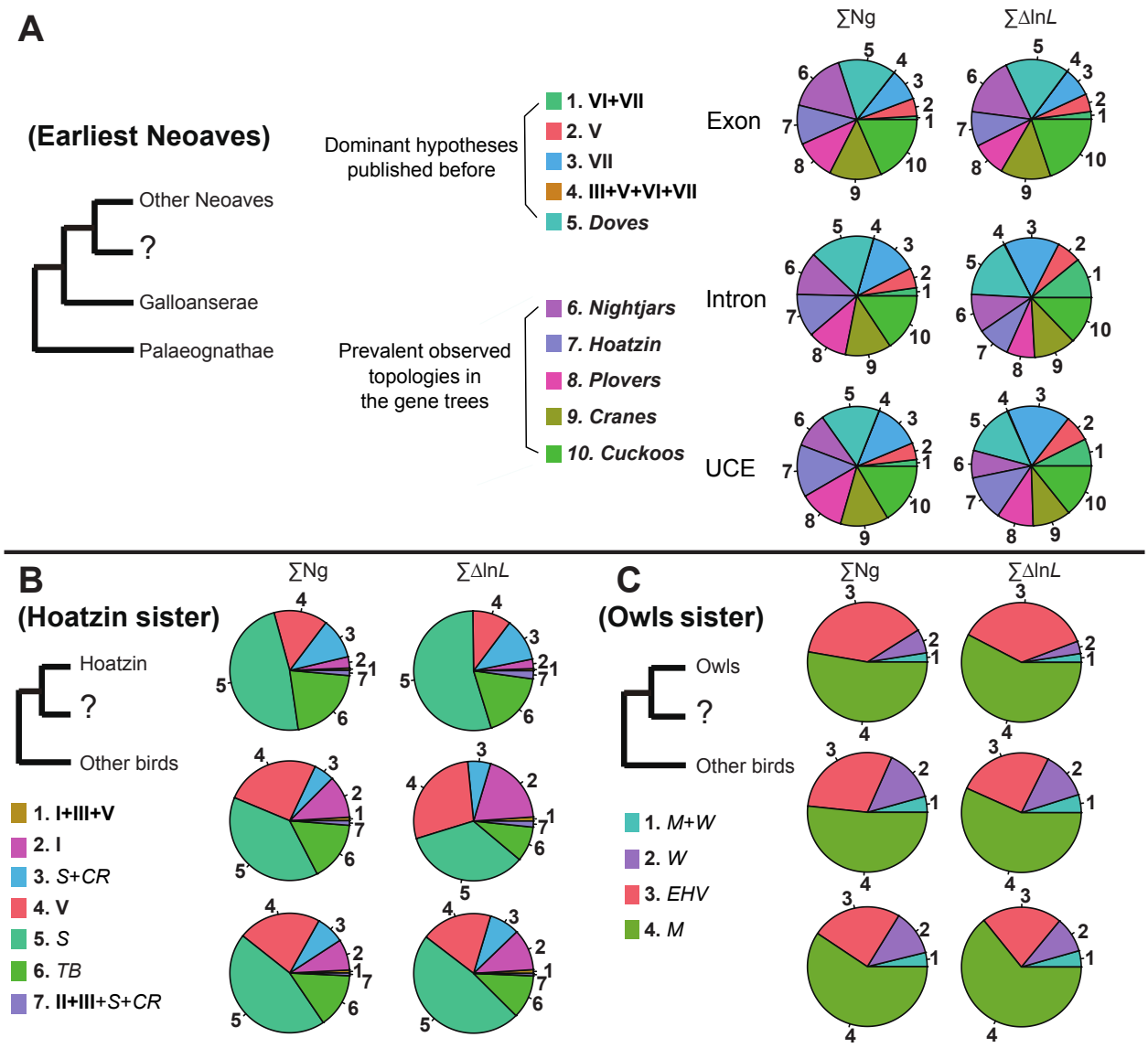

A
(Earliest Neoaves)
Other Neoaves
?
Galloanserae
Palaeognathae
Dominant hypotheses published before
1. VI+VII
2. V
3. VII
4. III+V+VI+VII
5. Doves
Prevalent observed topologies in the gene trees
6. Nightjars
7. Hoatzin
8. Plovers
9. Cranes
10. Cuckoos
∑Ng
∑ΔlnL
Exon
Intron
UCE
B
(Hoatzin sister)
Hoatzin
?
Other birds
1. I+III+V
2. I
3. S+CR
4. V
5. S
6. TB
7. II+III+S+CR
C
(Owls sister)
Owls
?
Other birds
1. M+W
2. W
3. EHV
4. M

Figure 3

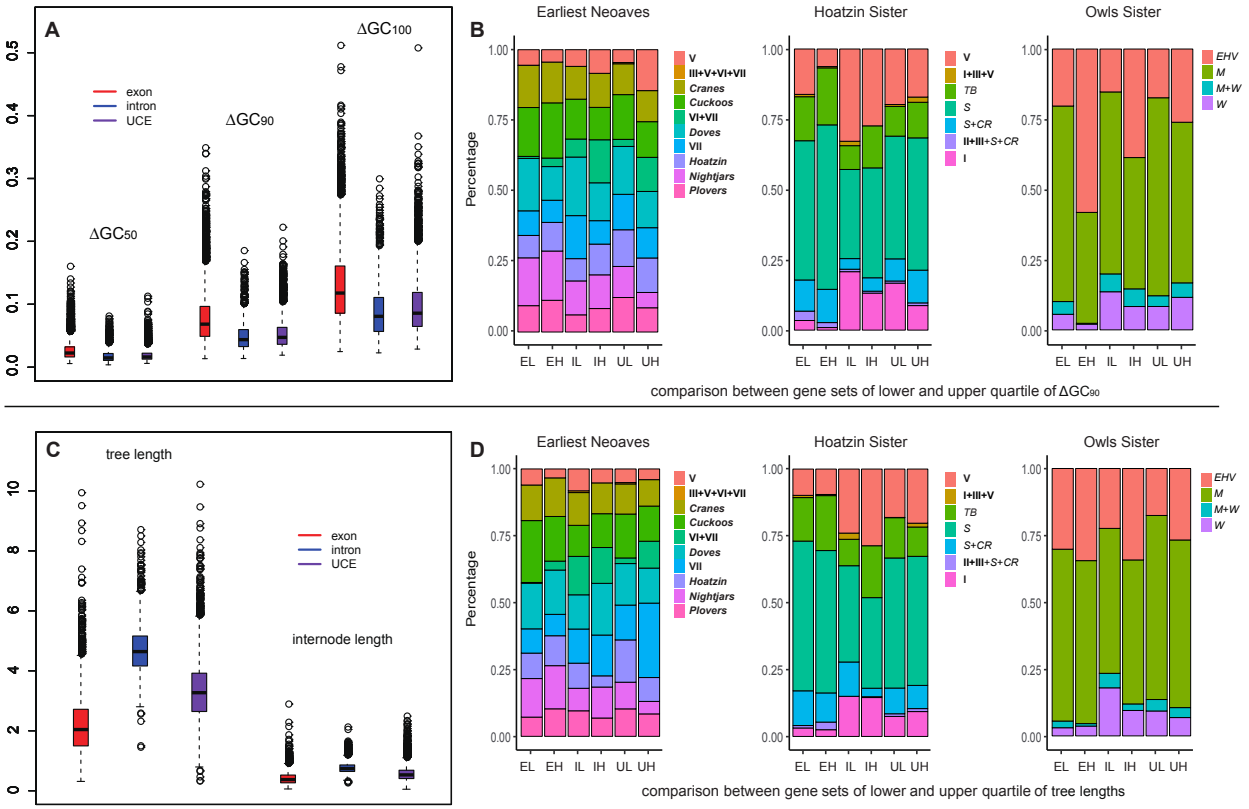

A
exon
intron
UCE
ΔGC50
ΔGC90
ΔGC100
B
Earliest Neoaves
Hoatzin Sister
Owls Sister
Percentage
V
III+V+VI+VII
Cranes
Cuckoos
VI+VII
Doves
VII
Hoatzin
Nightjars
Plovers
I+III+V
TB
S
S+CR
II+III+S+CR
I
EHV
M
M+W
W
EL EH IL IH UL UH
comparison between gene sets of lower and upper quartile of ΔGC90
C
tree length
internode length
D
comparison between gene sets of lower and upper quartile of tree lengths

Figure 4