RUNNING HEAD: EMPIRICAL ANOMALY ZONE IN PALAEOGNATHS

Phylogenomic analyses resolve historically contentious relationships within the Palaeognathae in the presence of an empirical anomaly zone

Alison Cloutier[1,2*], Timothy B. Sackton[3], Phil Grayson[1,2], Michele Clamp[3], Allan J. Baker[4,5¶], Scott V. Edwards[1,2]

[1]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA USA*

[2]*Museum of Comparative Zoology, Harvard University, Cambridge MA USA*

[3]*Informatics Group, Harvard University, Cambridge MA USA*

[4]*Department of Ecology and Evolutionary Biology, University of Toronto, Toronto ON Canada*

[5]*Department of Natural History, Royal Ontario Museum, Toronto ON Canada*

[¶]*Deceased*

*Corresponding author: alison.cloutier@mail.utoronto.ca

ABSTRACT

Palaeognathae represent one of the two basal lineages in modern birds, and comprise the volant (flighted) tinamous and the flightless ratites. Resolving palaeognath phylogenetic relationships has historically proved difficult, and short internal branches separating major palaeognath lineages in previous molecular phylogenies suggest that extensive incomplete lineage sorting (ILS) might have accompanied a rapid ancient divergence. Here, we investigate palaeognath relationships using genome-wide data sets of three types of noncoding nuclear markers, together totalling 20,850 loci and over 41 million base pairs of aligned sequence data. We recover a fully resolved topology placing rheas as the sister to kiwi and emu + cassowary that is congruent across marker types for two species tree methods (MP-EST and ASTRAL-II). This topology is corroborated by patterns of insertions for 4,274 CR1 retroelements identified from multi-species whole genome screening, and is robustly supported by phylogenomic subsampling analyses, with MP-EST demonstrating particularly consistent performance across subsampling replicates as compared to ASTRAL. In contrast, analyses of concatenated data supermatrices recover rheas as the sister to all other non-ostrich palaeognaths, an alternative that lacks retroelement support and shows inconsistent behavior under subsampling approaches. While statistically supporting the species tree topology, conflicting patterns of retroelement insertions also occur and imply high amounts of ILS across short successive internal branches, consistent with observed patterns of gene tree heterogeneity. Coalescent simulations indicate that the majority of observed topological incongruence among gene trees is consistent with coalescent variation rather than arising from gene tree estimation error alone, and estimated branch lengths for short successive internodes in the inferred species tree fall within the theoretical range encompassing the anomaly zone. Distributions of empirical gene trees confirm

that the most common gene tree topology for each marker type differs from the species tree, signifying the existence of an empirical anomaly zone in palaeognaths.

Keywords: palaeognath, ratite, phylogenomics, retroelement, species tree, multispecies coalescent, incomplete lineage sorting, anomaly zone

INTRODUCTION

The scaling-up of multigene phylogenetic data sets that accompanied rapid advances in DNA sequencing technologies over the past two decades was at first heralded as a possible end to the incongruence resulting from stochastic error associated with single-gene topologies (Rokas et al. 2003, Gee 2003). However, it soon became clear that conflicting, but highly supported, topologies could result from different data sets when sequence from multiple genes was analyzed as a concatenated supermatrix, leading Jeffroy et al. (2006) to comment that phylogenomics – recently coming to signify the application of phylogenetic principles to genome-scale data (Delsuc et al. 2005) – could instead signal 'the beginning of incongruence'. On the one hand, these observations highlighted the need for more sophisticated models to account for nonphylogenetic signal such as convergent base composition or unequal rates that can become amplified in large concatenated data sets (Jeffroy et al. 2006). But at the same time, there was a growing recognition that gene trees and species trees are not equivalent entities and that horizontal gene transfer, gene duplication and loss, and incomplete lineage sorting can result in heterogeneous topologies for gene trees that are all contained within a single overarching species tree history (Maddison 1997). When gene tree discordance arising from these biological factors predominates, concatenation approaches, which implicitly assume that all genes share the same

70 topology, can not only fail to recover the true species tree but can also infer erroneous
71 relationships with strong support (Kubatko and Degnan 2007, Roch and Steel 2015). This
72 recognition of the biological basis for gene tree discordance prompted proposals for a conceptual
73 shift to adopt the multispecies coalescent model as a flexible and biologically realistic
74 framework that incorporates population-level processes into phylogenetic inference (Degnan and
75 Rosenberg 2009, Edwards 2009, Edwards et al. 2016).

76 Among the processes generating variation in gene tree histories, coalescent variation
77 caused by incomplete lineage sorting (ILS) has received much attention, in part because it is the
78 most ubiquitous and common form of gene tree–species tree incongruence (Edwards 2009).
79 Incomplete lineage sorting occurs when ancestral polymorphisms do not reach fixation between
80 successive divergences among taxa or, viewed from the opposite direction as deep coalescence,
81 when alleles of two taxa fail to coalesce in their most recent common ancestor (Maddison 1997,
82 Degnan and Rosenberg 2009). The likelihood of ILS increases when internal branches in the
83 species tree are short relative to the effective population size and this phenomenon is therefore
84 likely to accompany many rapid radiations that are the focus of evolutionary studies (Song et al.
85 2012, Liu et al. 2015b). A surprising outcome of ILS under the multispecies coalescent is that
86 there exist branch lengths in the species tree for which the most probable gene tree topology
87 differs from that of the species tree, producing a zone where these anomalous gene trees (AGTs)
88 predominate in the gene tree distribution (the 'anomaly zone', Degnan and Rosenberg 2006).
89 The anomaly zone carries clear implications for phylogenetic inference, since any method
90 employing a democratic vote among loci will fail in this region of tree space (Degnan and
91 Rosenberg 2006, 2009; Huang and Knowles 2009). Incomplete lineage sorting is documented
92 for many taxa (e.g. Prüfer et al. 2012, Song et al. 2012, Suh et al. 2015), and the theoretical basis

93 of the anomaly zone is well established (Degnan and Rosenberg 2006, 2009; Rosenberg 2013).
94 However, an empirical anomaly zone has thus far only been reported in Scincidae skinks
95 (Linkem et al. 2016) and the general relevance of this phenomenon to empirical studies has been
96 questioned (Huang and Knowles 2009).

97 Short branches that produce high amounts of incomplete lineage sorting are also expected
98 to carry relatively few informative substitutions (Huang and Knowles 2009, Xi et al. 2015).
99 Observed gene tree heterogeneity accompanying an inferred empirical anomaly zone could
100 therefore reflect uninformative gene trees rather than truly anomalous ones (Huang and Knowles
101 2009). More generally, gene tree estimation error has been raised as a potential concern for
102 current methods of species tree inference even outside of the anomaly zone (Liu et al. 2015c,
103 Roch and Warnow 2015). Fully coalescent single-step Bayesian methods that coestimate gene
104 trees with the species tree, despite providing the most seamless statistical environment for testing
105 phylogenetic hypotheses using the coalescent (Xu and Yang 2016), are as yet computationally
106 prohibitive for large data sets (Liu et al. 2015c, Edwards 2016a), motivating the development of
107 many two-step methods that use estimated gene trees as input for species tree inference
108 (reviewed in Liu et al. 2015a, Edwards 2016a). These summary methods can show good
109 performance under simulation (Liu et al. 2010, Mirarab et al. 2016), sometimes even within the
110 anomaly zone (Liu et al. 2010), but assume that gene trees are known without error. Simulation
111 studies have shown that summary species tree methods can still perform well in the presence of
112 gene tree estimation error (e.g. Liu et al. 2010, Mirarab et al. 2016), and recent results from
113 coalescent simulations (Xi et al. 2015) and analysis of empirical data (Blom et al. 2017) indicate
114 that the accuracy of species tree inference can be improved by adding more loci, even if those
115 loci are minimally informative (Xu and Yang 2016). However, whether these observations apply

116 broadly across empirical data sets is as yet unknown, and it is also crucial that gene tree error, if
117 it occurs, is unbiased (Xi et al. 2015). It is therefore imperative for empirical studies to carefully
118 consider the underlying gene tree support for inferred species trees and to assess consistency of
119 results across analysis methods and data types.

120 Here, we use genome-wide data sets of three types of noncoding markers (conserved non-
121 exonic elements [CNEEs], introns, and ultraconserved elements [UCEs]) to investigate
122 relationships within the Palaeognathae, one of the two basal lineages of modern birds (the other
123 being Neognathae, which includes the Galloanserae [gamebirds + waterfowl] and Neoaves [all
124 other birds]). Palaeognaths encompass the flighted tinamous (Tinamiformes) of South and
125 Central America and the flightless ratites, including the African ostrich (Struthioniformes),
126 Australian emu and Australasian cassowaries (Casuariiformes), New Zealand kiwi
127 (Apterygiformes), and South American rheas (Rheiformes) as well as the recently extinct New
128 Zealand moa (Dinornithiformes) and Madagascan elephant birds (Aepyornithiformes).
129 Palaeognath phylogenetic relationships have remained controversial for over a century (reviewed
130 in Houde 1988, Cracraft et al. 2004), and recent molecular studies have added to the debate by
131 placing the tinamous as sister to the moa within a paraphyletic ratite clade (Phillips et al. 2010,
132 Haddrath and Baker 2012, Baker et al. 2014) and recovering an unexpected sister-group
133 relationship between kiwi and elephant birds (Mitchell et al. 2014, Grealy et al. 2017, Yonezawa
134 et al. 2017). These findings have important implications for our understanding of Gondwonan
135 biogeography and the possibility that morphological convergence has accompanied multiple
136 independent losses of flight in the ratites (Harshman et al. 2008, Haddrath and Baker 2012,
137 Baker et al. 2014, Sackton et al. 2018).

138 Although ratite paraphyly is strongly supported by all recent molecular studies (Hackett

139 et al. 2008, Harshman et al. 2008, Phillips et al. 2010, Haddrath and Baker 2012, Smith et al.
140 2013, Baker et al. 2014, Mitchell et al. 2014, Grealy et al. 2017, Yonezawa et al. 2017), rheas
141 have been variously placed as the sister to all other non-ostrich palaeognaths (e.g. Hackett et al.
142 2008, Harshman et al. 2008, Phillips et al. 2010, Mitchell et al. 2014, Prum et al. 2015 and
143 others), to tinamous (Harshman et al. 2008, Smith et al. 2013; note that moa sequences were
144 absent from these analyses), or to a clade containing emu, cassowary, and kiwi (Haddrath and
145 Baker 2012, Prum et al. 2015), with the same data sometimes producing conflicting results under
146 different analysis regimes (e.g. Harshman et al. 2008, Smith et al. 2013, Prum et al. 2015).
147 Alternative placements of rheas are often accompanied by low bootstrap support, and it is
148 suspected that ILS across short internal branches separating major palaeognath lineages could
149 underlie some of the difficulties in resolving phylogenetic relationships within this group
150 (Haddrath and Baker 2012).
151 Newly available whole-genome sequences (Zhang et al. 2014, Le Duc et al. 2015,
152 Sackton et al. 2018) allow investigation of palaeognath relationships at a phylogenomic scale.
153 Recently, we reported fully congruent results using the coalescent species tree method MP-EST
154 (Liu et al. 2010) that recovered rheas as the sister to emu/cassowary + kiwi. These results were
155 consistent across marker type and robust to the effects of missing data, alignment uncertainty,
156 and outgroup choice, but differed from the incongruent results obtained under concatenation
157 (Sackton et al. 2018). Here, we corroborate the species tree topology inferred from sequence-
158 based analysis with genome-wide screening for informative presence/absence insertions of CR1
159 retroelements, and incorporate results from a second species tree method (ASTRAL, Mirarab and
160 Warnow 2015) to further strengthen support for this topology relative to that obtained from
161 concatenated sequence data. Additionally, we employ phylogenomic subsampling to investigate

consistency in the underlying signal for conflicting relationships recovered from species tree versus concatenation approaches, and use likelihood evaluation and coalescent simulation to assess the underlying gene tree support for the recovered species tree topology and the existence of an empirical anomaly zone in palaeognaths. Throughout, we consider the variation in signal among classes of noncoding nuclear markers that are becoming increasingly adopted for genome-scale analyses (Edwards et al. 2017), and contrast the relative performance of these markers under different analysis regimes to resolve historically challenging relationships among Palaeognathae.

MATERIALS & METHODS

Data set compilation

We assembled data sets for three types of noncoding nuclear markers: conserved non-exonic elements (CNEEs), introns, and ultraconserved elements (UCEs) for 14 palaeognath species and a chicken outgroup from publicly available whole-genome sequence assemblies (Suppl. Table S1). We chose to analyze noncoding sequences primarily because coding regions across large taxonomic scales in birds are known to experience more severe among-lineage variation in base composition than noncoding regions, which can complicate phylogenetic analysis (Jarvis et al. 2014, Reddy et al. 2017). On the other hand, noncoding markers can be more challenging to align than coding markers, and the three marker types we analyze here exhibit a range of variation and alignment challenges (Edwards et al. 2017).

We filtered the 1,949,832 conserved elements (CEs) identified statistically by Sackton et al. (2018) to omit elements overlapping annotated exon, gene, or CDS features in the galGal4 chicken genome release (NCBI annotation release 102) to generate a candidate set of 811,696

185 conserved non-exonic elements (CNEEs). This set was further filtered to retain 16,852 elements
186 greater than 250 bp in length. Orthologous sequence from 10 newly published palaeognath
187 genomes (Sackton et al. 2018) as well as ostrich (*Struthio camelus*) and white-throated tinamou
188 (*Tinamus guttatus*) genomes from the Avian Phylogenomics Project (Zhang et al. 2014) was
189 compiled by lifting over reference genome coordinates from chicken to each target palaeognath
190 in an existing 42 species multiway whole-genome alignment (WGA, Sackton et al. 2018) with
191 HAL Tools v.2.1 (Hickey et al. 2013). Liftover output was parsed to retain only uniquely
192 mapping regions between the chicken reference and target palaeognath genomes. CNEEs with
193 no missing taxa from the 12 palaeognaths included in the whole-genome alignment, and with at
194 least as many variable sites as taxa were retained (N= 14,528 loci). Loci overlapping the set of
195 UCEs described below were omitted using BEDTools v. 2.26.0 (Quinlan and Hall 2010), leaving
196 a data set of 12,676 CNEEs.

197 Candidate introns were identified using BEDTools to output coordinates for annotated
198 intron features in the galGal4 genome release that did not overlap with any annotated exon
199 feature. Chicken coordinates for these introns were lifted over to target palaeognaths in the
200 whole-genome alignment as described for CNEEs above, filtered to remove duplicated regions,
201 and expected exon/intron boundaries in each palaeognath sequence were refined by making
202 intron coordinates flush to liftovers for adjacent exon sequences. Candidate introns within each
203 palaeognath species were omitted if shorter than 100 bp, if longer than 100 kb, or if longer than
204 10 kb and more than 50% longer than the chicken reference intron length. After adding
205 sequence from the published North Island brown kiwi (Le Duc et al. 2015) and little bush moa
206 (Cloutier et al. 2018) as described below, one intron per gene was chosen, requiring a minimum
207 pairwise sequence identity of 70% and fewer than 0.5 undetermined sites (e.g. gaps and Ns) per

208 bp aligned sequence, and then choosing among multiple introns per gene based on the smallest
209 number of missing taxa and longest average unaligned input sequence length across taxa. This
210 produced a final data set of 5,016 introns, each originating from a different gene.

211 Ultraconserved elements (UCEs) were compiled using the set of 3,679 bioinformatically
212 harvested UCEs from the Avian Phylogenomics Project (Jarvis et al. 2014, 2015; accessed from
213 http://dx.doi.org/10.5524/101041). In addition to chicken, we used ostrich and white-throated
214 tinamou included in the Avian Phylogenomics Project data as reference species to lift over to all
215 other palaeognaths in the whole-genome alignment. Liftover output from these multiple
216 reference taxa was parsed to omit duplicated regions and tiled for regions that were consistent
217 across reference species. A maximum of one missing taxon per locus from the 12 palaeognaths
218 in the whole-genome alignment was allowed, resulting in a data set of 3,158 UCEs.

219 Blastn searches with NCBI's default 'somewhat similar' parameters (evalue 1e-10,
220 perc_identity 10, penalty -3, reward 2, gapopen 5, gapextend 2, word size 11) were used with
221 query sequence from each of the three kiwi species included in the whole-genome alignment to
222 identify orthologous regions in the North Island brown kiwi (*Apteryx mantelli*, Le Duc et al.
223 2015), which was not included in the WGA. North Island brown kiwi sequence was added for
224 loci that had consistent top hits across blastn searches, with a single high-scoring segment pair
225 (HSP) covering at least 50% of the input query sequence and minimum 80% sequence identity.
226 Using this approach, we added *A. mantelli* sequence for 11,534/12,676 CNEEs (91% of total),
227 2,560/5,016 introns (51%), and 2,033/3,158 UCEs (64%).

228 Sequence was also added from a reference-based genome assembly of the extinct little
229 bush moa (*Anomalopteryx didiformis*, Cloutier et al. 2018) that was generated by mapping moa
230 reads to the emu genome included in the whole-genome alignment. Emu coordinates from the

231 liftover approach outlined above were used to retrieve the corresponding region in moa, and moa
232 sequence was retained if it spanned at least 30% of the emu reference length or was at least 200
233 bp in length, excluding Ns. Moa sequence was added for 12,643/12,676 CNEEs (99.7%),
234 4,982/5,016 introns (99.3%), and 3,146/3,158 UCEs (99.6%).

235 Sequences for each locus were retrieved from genome assemblies and aligned *de novo*
236 with MAFFT v. 7.245 (Katoh and Standley 2013) using default options for global iterative
237 alignment for CNEEs (option 'ginsi') and local iterative alignment for introns and UCEs (option
238 'linsi'). This produced data sets of 4,794,620 bp of aligned sequence for the 12,676 CNEEs,
239 27,890,802 bp for 5,016 introns, and 8,498,759 bp for 3,158 UCEs.

240 *Gene tree inference*

241 Best maximum likelihood trees were built with RAxML v. 8.1.5 (Stamatakis 2014) from
242 unpartitioned alignments for each locus using a GTR+GAMMA substitution model and 20
243 independent searches beginning with randomly generated starting trees. Topologies were also
244 inferred for 500 bootstrap replicates per locus, with a GTR+GAMMA model on unpartitioned
245 data in RAxML.

246 *Species tree inference*

247 MP-EST v. 1.5 (Liu et al. 2010) analyses were conducted with three full runs per data set
248 beginning with different random number seeds and with ten independent tree searches within
249 each run. The species tree topology was inferred using best maximum likelihood gene trees as
250 input, and node support was estimated from MP-EST runs using gene tree bootstrap replicates as
251 input.

252 ASTRAL-II v. 4.10.9 (Mirarab and Warnow 2015, hereafter 'ASTRAL') was also run
253 using the best maximum likelihood gene trees to infer the species tree topology, and the 500

254 gene tree bootstrap replicates to estimate node support. To permit direct comparison across
255 methods, the default quadripartition support values output by ASTRAL were replaced with
256 traditional bipartition supports by plotting output bootstrap replicates from ASTRAL on the
257 inferred species tree topology using RAxML.

258 To investigate the effects of concatenation on phylogenetic inference, ExaML v. 3.0.16
259 (Kozlov et al. 2015) was used to infer maximum likelihood topologies from fully partitioned (i.e.
260 1 partition per locus) concatenated data sets using a GTR+GAMMA substitution model.
261 Twenty-one full tree searches were run for each data set, with 20 searches beginning from
262 complete random starting trees, and one additional search using a random stepwise addition
263 order parsimony starting tree. A minimum of 50 bootstrap replicates were computed for each
264 data set, with additional replicates performed as necessary until convergence was reached as
265 assessed by bootstopping analysis with a majority-rule consensus tree criterion implemented in
266 RAxML.

267 For all three methods, bootstrap supports were placed on the inferred species tree using
268 RAxML, and trees were outgroup rooted with chicken using ETE v. 3 (Huerta-Cepas et al.
269 2016).

270 *Phylogenomic subsampling*

271 In addition to the traditional bootstrapping approaches outlined above, phylogenomic
272 subsampling (Blom et al. 2016; reviewed in Edwards 2016b) was used to assess the consistency
273 of the underlying support for conflicting clades. For each marker type (CNEEs, introns, and
274 UCEs), loci were randomly sampled with replacement to create subsets of 50, 100, 200, 300,
275 400, 500, 1000, 1500, 2000, 2500, and 3000 loci. This process was repeated ten times to create a
276 total of 110 data sets per marker type (e.g. 10 replicates of 50 loci each, 10 replicates of 100 loci,

277 etc. for each of CNEEs, introns, and UCEs). Topologies were inferred for bootstrap replicates of
278 each data set using MP-EST, ASTRAL, and ExaML as described above. However, for reasons
279 of computational tractability, 200 rather than 500 bootstrap replicates were used for each method
280 (including ExaML), and MP-EST was run once (rather than three times) for each data set,
281 although still with ten tree searches within each run.

282 Support for alternative hypotheses regarding the sister group to the rheas, and the sister to
283 emu + cassowary, was estimated by first counting the number of bootstraps that recovered each
284 alternative topology from the 200 bootstraps run for each replicate, and then calculating the
285 mean value for each hypothesis across the ten replicates within each data set size category.

286 *Analysis of incomplete lineage sorting (ILS) and gene tree heterogeneity*

287 Analyses of gene tree heterogeneity were conducted using both the best maximum
288 likelihood trees inferred by RAxML, as well as majority-rule extended consensus trees generated
289 from RAxML bootstrap replicates. Most analyses of topological incongruence required an
290 identical taxon sampling across gene trees; however, there was a much larger amount of missing
291 data for North Island brown kiwi than for other taxa, in part because this species was added *post*
292 *hoc* to data sets that were compiled for species included in the whole-genome alignment and in
293 part because the assembly quality of this species was lower than for those palaeognaths in the
294 WGA. We therefore pruned North Island brown kiwi from all gene trees, and omitted loci that
295 were missing any other taxa. This approach retained 12,643/12,676 CNEEs (99.7% of loci),
296 4,982/5,016 introns (99.3%), and 2,866/3,158 UCEs (90.8%).

297 Gene support frequency (GSF) and internode certainty 'all' (ICA, Salichos et al. 2014)
298 were calculated with RAxML. Distributions of gene tree topologies were calculated by
299 computing all pairwise Robinson-Foulds (RF) cluster distances on outgroup rooted gene trees

300 with TreeCmp v. 1.1-b308 (Bogdanowicz et al. 2012) and parsing the output into mutually
301 compatible sets of trees with identical topology where RF= 0. Numbers of substitutions
302 occurring on braches conflicting with the species tree topology were inferred under a parsimony
303 criterion in PAUP v. 4.0a151 (Swofford 2002).

304 Relative support for each gene tree topology was assessed by computing ΔAIC from the
305 log-likelihood score (lnL) of the inferred gene tree and lnL when the input alignment for each
306 locus was constrained to the species tree topology. Both likelihood values were calculated in
307 RAxML on unpartitioned alignments with a GTR+GAMMA substitution model, and branch
308 lengths were optimized for each topology, a criterion for applying the ΔAIC approach.

309 Approximately unbiased (AU) tests were run in IQ-TREE v. 1.5.4 (Nguyen et al. 2015).
310 For each locus, we tested the estimated gene tree topology against an *a priori* candidate set of
311 probable trees that enforced monophyly of the five higher-level palaeognath lineages (kiwi, emu
312 + cassowary, rheas, moa + tinamous, ostrich), but allowed all possible rearrangements among
313 those lineages (for a total of 105 trees in the candidate set, plus the gene tree topology itself if it
314 did not occur within this set). We also tested gene trees against a second set of candidate
315 topologies using the same criteria as above, but additionally allowing all possible rearrangements
316 within a monophyletic tinamou clade (for 1,575 candidate topologies). For each gene, in
317 addition to the reported P-value for the fit of the species tree topology, we also calculated the
318 rank of the species tree topology relative to all tested candidates from P-values sorted in
319 ascending order.

320 The proportion of observed gene tree heterogeneity consistent with coalescent variation
321 was estimated through simulation. For each marker type, we estimated ultrametric species tree
322 branch lengths in mutational units (μT, where μ is the mutation rate per generation and T is the

323 number of generations) by constraining concatenated alignments of all loci to the species tree
324 topology with a GTR+GAMMA substitution model and strict molecular clock in PAUP. For this
325 purpose, we ignore the additional mutations incurred due to variation in ancestral lineages in the
326 species tree (Angelis and dos Reis 2015), as well as misestimation of the number of mutations
327 due to deviation of the gene trees from the species tree (Mendes and Hahn 2016). These
328 mutational branch lengths were used in combination with the MP-EST and ASTRAL species tree
329 branch lengths in coalescent units ($\tau = T/4N_e$) to calculate the population size parameter theta (Θ)
330 for each internal branch following Degnan and Rosenberg (2009), where $(\Theta/2) \cdot T/N_e = \mu T$
331 which reduces to $\Theta = \mu T/\tau$. Theta for terminal branches was set to a constant value of 1; this
332 value had no impact on simulated gene trees because we simulated only one allele per species.
333 10,000 gene trees were simulated from species trees using these mutational branch
334 lengths and theta values for each data set (e.g. CNEEs, introns, and UCEs, with theta derived
335 from both MP-EST and ASTRAL branch lengths) using the sim.coaltree.sp function in Phybase
336 v. 1.5 (Liu and Yu 2010). Pairwise Robinson-Foulds cluster and matching cluster distances of
337 each simulated gene tree to the species tree topology were calculated with TreeCmp
338 (Bogdanowicz et al. 2012). The ratio of mean gene tree–species tree distances for simulated
339 gene trees relative to mean distances for empirically estimated gene trees was calculated as a
340 measure of the amount of observed gene tree heterogeneity that can be accounted for by
341 coalescent processes.
342 Rooted triplets were generated by pruning all possible combinations of three ingroup taxa
343 plus the chicken outgroup from each gene tree with ETE. Proportions of the major (species tree)
344 triplet and the two alternative minor triplets were calculated for all species trios, as well as for
345 combinations of higher level lineages (e.g. all trios involving one kiwi, one rhea, and either emu

346 or cassowary). We followed the approach of Song et al. (2012) to account for non-independence
347 of sampled trios when calculating proportions for higher-level lineages by assigning the most
348 commonly occurring triplet combination for each gene and omitting genes with ties before
349 tallying triplet counts across all genes. Observed triplet values were compared to expected
350 proportions calculated using branch lengths in coalescent units from MP-EST and ASTRAL
351 species trees, where the expected frequency of the major triplet equals $1 - 2/3 \exp(-t)$ and the two
352 minor triplets equal $1/3 \exp(-t)$, with t representing the internal branch length in the triplet
353 (Pamilo and Nei 1988).

354 Equation 4 from Degnan and Rosenberg (2006) was used to calculate the value $a(x)$ for
355 each internal branch x measured in coalescent units from MP-EST and ASTRAL species trees.
356 Calculated $a(x)$ values were compared to coalescent branch lengths for each descendent internal
357 branch y, with $y < a(x)$ providing evidence that this region of the species tree falls within the
358 zone where anomalous gene trees (AGTs) are expected – the anomaly zone. In addition to
359 calculations from species tree branch lengths, which are derived from input best maximum
360 likelihood gene trees, we also performed anomaly zone calculations for each of the 500 species
361 tree bootstrap replicates using scripts provided by Linkem et al. 2016 (accessed from
362 http://datadryad.org/resource/doi:10.5061/dryad.sf6s9).

363 *CR1 retroelement insertions*

364 Patterns of CR1 retroelement insertions were used to corroborate findings from sequence-
365 based analyses for all palaeognath species included in the whole-genome alignment (therefore
366 excluding the North Island brown kiwi and little bush moa). Draft genomes for each species
367 were repeat masked using RepeatMasker v. 4.0.5 (Smit et al. 2015) with RepBase Update
368 20140131, the sensitive search option, and query species 'vertebrata metazoa'. RepeatMasker

369 output was parsed to retain simple CR1 insertions at least 50 bp in length, omitting overlapping
370 and nested insertions. Coordinates for up to 500 bp of flanking sequence to either side of CR1
371 insertions in each species were retrieved and lifted over to every other target palaeognath species
372 in the whole-genome alignment with HAL Tools.

373 Liftover coordinates were parsed to exclude duplicated regions, and orthologous loci
374 were compiled across species with BEDTools and custom Perl scripts. Loci with sequence from
375 at least one taxon for each major lineage (kiwi, emu + cassowary, rheas, ostrich, and tinamous)
376 and with phylogenetically informative insertion patterns were retained (i.e. excluding
377 symplesiomorphies shared by all palaeognaths or autapomorphic insertions unique to a single
378 species). Sequence for each locus was extracted from genome assemblies and *de novo* aligned
379 with MAFFT using default parameters for global alignment. Presence/absence scores based on
380 RepeatMasker annotations were bioinformatically reassessed in aligned sequence data, and
381 sequences with gaps spanning the CR1 insertion site were rescored as 'omitted'. Loci with flush
382 insertion boundaries and identifiable target site duplications (TSDs) at least 2 bp in length were
383 retained, and all retained alignments were visually inspected. RepeatMasker output for these loci
384 was parsed to verify that putative orthologous insertions shared the same CR1 subtype, insertion
385 orientation, and degree of fragmentation. Sequence from chicken was added to alignments for
386 CR1 insertions that grouped all non-ostrich palaeognaths to verify insertion boundaries and
387 confirm that CR1 absence was not due to a secondary deletion within this genomic region in
388 ostrich. Statistical support for conflicting patterns of CR1 insertions was calculated with the
389 KKSC insertion significance tests of Kuritzin et al. (2016).
390

391 **RESULTS**

392 *Coalescent species tree methods, but not concatenation, recover congruent palaeognath*
393 *relationships*

394 Using genome-wide data sets of 12,676 CNEEs, 5,016 introns, and 3,158 UCEs, we
395 recover fully congruent topologies across all marker types and for the combined total evidence
396 tree using MP-EST and ASTRAL coalescent species tree approaches (Fig. 1, Suppl. Fig. S1a-h;
397 N.B. following Liu et al. 2015b and Edwards 2016a, we note the distinction between fully
398 coalescent MP-EST and ASTRAL, whose clustering algorithm is not strictly coalescent, but
399 hereafter refer to both methods as 'coalescent'). Maximal support is obtained throughout for
400 MP-EST and for the ASTRAL species tree built from CNEEs, but with reduced support for the
401 placement of rheas as the sister to a clade containing kiwi and emu + cassowary for ASTRAL
402 trees inferred from introns and UCEs (95% and 83% bootstrap support, respectively). In
403 contrast, ExaML inference from concatenated data sets places the rheas as sister to all other non-
404 ostrich palaeognaths with full support for introns and reduced support for CNEEs (60%) and
405 UCEs (90%, Suppl. Fig. S1i-l). ExaML analyses are also inconsistent across marker types in
406 their placement of a casuariiform clade comprising emu and cassowary, where introns recover a
407 sister group relationship to kiwi as is seen for MP-EST and ASTRAL, but with casuariiforms
408 instead placed as the sister to moa + tinamous with 60% support for CNEEs and 96% support for
409 UCEs (Suppl. Fig. S1).

410 Robustness of the underlying signal for these inconsistently recovered relationships was
411 further assessed with phylogenomic subsampling, which uses a variation of double bootstrapping
412 (Seo et al. 2008) to generate replicate data sets with increasing numbers of randomly sampled
413 loci, each of which is then subject to traditional bootstrap resampling across sites. MP-EST
414 analyses rapidly accumulate support for a sister-group relationship between rheas and

415 emu/cassowary + kiwi, where 100% support is consistently recovered across replicates with an
416 intermediate number of loci (e.g. with > 500 loci, Fig. 2a-c, Fig. 3a,j), and with support for
417 alternative hypotheses sharply dropping off for replicates with greater than 200 loci. Support
418 accumulates more slowly for ASTRAL, but the hypothesis of rheas as sister to emu/cassowary +
419 kiwi clearly dominates and support for alternatives declines in replicates with more than 1000
420 loci for all markers (Fig. 2d-f, Fig. 3). In contrast, subsampling replicates are less consistent for
421 relationships inferred under concatenation with ExaML (Fig. 2g-i, Fig. 3). In particular, CNEEs
422 oscillate between recovering rheas as the sister to moa + tinamous, or the sister to all other non-
423 ostrich palaeognaths, although always with low bootstrap support (Fig. 2g, Fig. 3g). The other
424 two marker types more clearly support the topology recovered from full data sets with ExaML
425 that place rheas as sister to the remaining non-ostrich palaeognaths, although bootstrap support
426 for UCE replicates is generally weak (Fig. 2h,i; Fig. 3).

427 Subsampling provides even more robust support for emu + cassowary as the sister to
428 kiwi, with both MP-EST and ASTRAL quickly accumulating support for this clade and with
429 rapidly declining support for all other hypotheses (Suppl. Fig. S2a-f, Suppl. Fig. S3). ExaML
430 intron replicates also steadily accumulate support for this relationship with an increasing number
431 of loci (Suppl. Fig. S2h, Suppl. Fig. S3h,q). The alternative hypothesis of emu + cassowary as
432 sister to moa + tinamous, which is favored by CNEEs and UCEs analyzed within a concatenation
433 framework, is not well supported by subsampling, where conflicting topologies characterized by
434 low support are recovered across ExaML replicates (Suppl. Fig. S2g,i; Suppl. Fig. S3).

435 *CR1 retroelement insertions corroborate findings from sequence-based analyses*

436 Patterns of CR1 retroelement insertions corroborate both the inferred species tree
437 topology from MP-EST and ASTRAL and the existence of substantial conflicting signal

438 consistent with incomplete lineage sorting across short internal branches. In total, 4,301
439 informative CR1 insertions were identified from multispecies genome-wide screens, the vast
440 majority of which (4,274 of 4,301, or 99.4%) are entirely consistent with relationships inferred
441 from sequence-based analyses (Fig. 4a; analysis here was restricted to species in the whole-
442 genome alignment, and little bush moa and North Island brown kiwi are therefore not included).
443 Not surprisingly, we identify many more insertion events occurring along shallower branches
444 with longer estimated lengths and fewer insertions along shorter branches that form the backbone
445 of the inferred species tree (refer to Fig. 1 for estimated branch lengths).

446 Of the 27 (0.6%) CR1s that are inconsistent with the species tree topology, two conflict
447 with the inferred relationships within kiwi, and nine contradict relationships among tinamous
448 (Fig. 4b,c). However, in each case, conflicting CR1s are far outweighed by markers that support
449 the species tree topology, providing robust support from statistical tests of insertion significance
450 (i.e. the KKSC tests of Kuritzin et al. 2016, P= $1.1e^{-66}$ in support of inferred kiwi relationships,
451 P= $8.8e^{-24}$ for Chilean tinamou + elegant crested tinamou, and P= $8.4e^{-152}$ for thicket + white-
452 throated tinamous). A further five insertions are inconsistent with tinamous as sister to a clade
453 containing rheas, kiwi, and emu + cassowary and would instead place tinamous as sister to just
454 one of these other lineages (Fig. 4d). Again, species tree relationships are statistically supported
455 for comparisons of all trios despite these contradictory markers. For example, while
456 casuariiforms (emu and cassowary) and kiwi each share one CR1 insertion exclusively with
457 tinamous, they have 12 insertions shared with each other to the exclusion of tinamous, therefore
458 supporting emu/cassowary + kiwi with P= $8.2e^{-05}$. Similar calculations support rheas with
459 emu/cassowary to the exclusion of tinamous at P= 0.0018, and rheas with kiwi to the exclusion
460 of tinamous at P= 0.004.

461 Perhaps most strikingly, we identified six CR1s shared by rheas and emu/cassowary to
462 the exclusion of all other palaeognaths, including kiwi, and five CR1s shared by rheas and kiwi
463 alone (Fig. 4e). Together with the 12 insertions shared by emu, cassowary and kiwi, these
464 patterns are still sufficient to support the inferred relationship of rheas as the sister to an
465 emu/cassowary + kiwi clade (P= 0.048), but indicate substantial conflict in this region of the
466 tree, mirroring the results from sequence-based analysis. Symmetric numbers of CR1s
467 supporting the two conflicting topologies further suggest that ILS, rather than ancestral
468 hybridization, underlies the observed conflicts. Insertion significance tests accordingly support
469 the bifurcating species tree topology rather than a reticulate hybridization network or unresolved
470 polytomy among these three lineages (Kuritzin et al. 2016). We also note that no CR1 insertions
471 were recovered to support the alternative placement of rheas as the sister to all other non-ostrich
472 palaeognaths as produced by concatenation-based analyses (i.e. no CR1s were shared by emu,
473 cassowary, kiwi, and tinamous, but absent from rheas and ostrich).

474 *Patterns of gene tree heterogeneity suggest substantial incomplete lineage sorting (ILS) and an*
475 *empirical anomaly zone*

476 We investigated gene tree heterogeneity using both the maximum likelihood estimates of
477 gene trees and the majority rule extended consensus of bootstrap replicates. These analyses
478 produced similar results and are therefore reported for consensus gene trees only.

479 Distributions of estimated gene tree topologies illustrate that the most common topology
480 for each marker type is not the species tree topology inferred by MP-EST and ASTRAL, thereby
481 suggesting the existence of an empirical anomaly zone (Fig. 5a-c). While the ranking of specific
482 gene tree topologies differs across marker types, common to these anomalous gene trees that
483 occur at higher frequency than the species tree topology is the fact that both the shallowest clades

484 as well as the deepest split between the ostrich and all other palaeognaths are maintained
485 throughout (Fig. 5d). Rearrangements of AGTs relative to the MP-EST and ASTRAL species
486 tree topology instead involve the two short internal branches forming the common ancestor of
487 emu and cassowary with kiwi, and with this clade to rheas (Fig. 5d).

488 To more fully investigate the observed gene tree heterogeneity, we considered all
489 estimated gene trees on a clade-by-clade basis by calculating the gene support frequency (GSF),
490 which tallies the number of genes that recover each bipartition in the inferred species tree, and
491 the internode certainty 'all' statistic (ICA, Salichos et al. 2014), which incorporates the
492 frequencies of all commonly occurring bipartitions (Fig. 6). Clades that are recovered
493 consistently in both coalescent and concatenation approaches (e.g. ostrich sister to other
494 palaeognaths, emu + cassowary, moa + tinamous, interrelationships within kiwi, and within
495 tinamous) are also typically supported by the majority of individual gene trees, although single-
496 gene support is weaker for CNEEs (Fig. 6a). However, low (near zero) internode certainty 'all'
497 values for emu/cassowary + kiwi indicate that alternative bifurcations occur at roughly equal
498 frequency in the input set of gene trees, and the negative ICA for the relationship of this clade
499 with rheas means that the emu/cassowary + kiwi + rhea clade is actually recovered less often
500 than alternatives.

501 We next considered whether topological differences between estimated gene trees and the
502 species tree are well supported, or are instead likely to primarily reflect gene tree estimation
503 error. Mean bootstrap support for estimated gene trees is relatively high, especially for introns
504 and UCEs (83.9% and 82.8% respectively, Fig. 7a-c). However, average support falls by about
505 10% for each marker type when gene tree bootstrap replicates are constrained to the species tree
506 topology, with $P < 0.0001$ for paired t-tests of each data set. These results suggest that

507 differences from the species tree are broadly supported by variation in sequence alignments for
508 individual loci. To test this further, we compared the difference in Akaike information criterion
509 (AIC) for estimated gene trees to the AIC obtained when the sequence alignment for each gene
510 was constrained to the species tree topology. Approximately 80% of CNEEs have ΔAIC (gene
511 tree–species tree) less than -2, indicating substantial support in favor of the gene tree topology
512 relative to that of the species tree (Burnham and Anderson 2002), while the proportion was even
513 greater for introns and UCEs (approximately 90% with ΔAIC < -2, Fig. 7d). Despite this result,
514 approximately unbiased (AU) tests typically failed to reject a hypothesis that the data fit the
515 species tree topology, with only about 20% of introns and UCEs, and 30% of CNEEs rejecting
516 the species tree topology at $P < 0.05$ (Fig. 7d). However, the species tree topology is also not
517 commonly amongst the top 5% of candidate alternative topologies when these alternatives are
518 ranked according to increasing AU test P-value within each locus (Fig. 7d).

519 In keeping with the results for all loci, gene tree topologies are also generally supported
520 for loci falling within AGT groups. Support for individual gene trees is somewhat weak for
521 CNEEs, with low median bootstrap support and few substitutions occurring along branches that
522 conflict with the species tree topology (Suppl. Fig. S4a,d), which is consistent with the shorter
523 average alignment length and lower variability of these loci (Suppl. Fig. S5). However, support
524 is much stronger for introns and UCEs, with most loci having bootstrap support above 50% for
525 conflicting clades and ΔAIC < -2 indicating much stronger likelihood support for the recovered
526 gene tree topology relative to that obtained if sequence alignments are constrained to match the
527 inferred species tree (Suppl. Fig. S4).

528 Simulations were used to further assess what proportion of total gene tree heterogeneity
529 is likely attributable to coalescent processes rather than to gene tree estimation error (Fig. 8).

530 Using either Robinson-Foulds (RF) distances or the matching cluster distance, which is
531 influenced less by the displacement of a single taxon than is the RF metric (Bogdanowicz et al.
532 2012), and simulating from coalescent branch lengths estimated with either MP-EST or
533 ASTRAL for each marker type, we find that coalescent processes alone can account for more
534 than 70% of the observed gene tree heterogeneity in most comparisons, and >90% for introns
535 when gene trees are simulated from MP-EST coalescent branch lengths (Fig. 8).

536 *Observed gene tree heterogeneity and anomalous gene trees (AGTs) are consistent with*
537 *expectations from coalescent theory*

538 Analysis of rooted triplets, where gene trees are decomposed into all possible
539 combinations of three species plus the outgroup, corroborate relationships inferred by MP-EST,
540 which also relies on rooted triplets (Liu et al. 2010) and ASTRAL, which instead uses unrooted
541 quartets (Mirarab and Warnow 2015), with the major, or most frequent, topology for all species
542 trios matching that of the species tree. Observed counts of the major and two minor topologies
543 for the rhea, kiwi, and emu+cassowary lineages (i.e. from sampling all trios that contain one
544 representative of each lineage) also occur in proportions that are consistent with coalescent
545 theory, based on the estimated length of the internal branch within the trio (Pamilo and Nei 1988,
546 Fig. 9 Fisher's exact $P > 0.9$ for all comparisons of observed proportions to those expected from
547 either MP-EST or ASTRAL branch lengths). Similar to the results from conflicting CR1
548 insertions, the two minor topologies for rooted triplets occur at almost equal frequencies, which
549 is consistent with coalescent variation arising from incomplete lineage sorting rather than
550 alternative processes such as introgression. Also notable are the similar triplet proportions
551 observed for all three marker types, which suggest that gene tree estimation error and/or
552 systematic biases within datasets alone are unlikely to explain the observed results.

553 The identity and frequency of observed anomalous gene trees are also consistent with
554 expectations from coalescent theory. Following Degnan and Rosenberg (2006) and Rosenberg
555 (2013), branch lengths in coalescent units estimated with either MP-EST (Fig. 10) or ASTRAL
556 (Suppl. Fig. S6) have values expected to produce AGTs across the short successive internal
557 branches separating rheas, emu+cassowary, and kiwi for all marker types. Branch lengths for
558 CNEEs, but not introns or UCEs, are also consistent with AGTs arising across branches
559 separating moa + tinamous from the remaining non-ostrich palaeognaths. Inference of these
560 anomaly zones applied equally when using species tree branch lengths, or for branch lengths
561 from each bootstrap replicate (i.e. every bootstrap replicate that recovers the species tree
562 topology also places these regions of tree space within the anomaly zone).
563

564 **DISCUSSION**

565 High-throughput sequencing has revolutionized the scale at which phylogenetic inference
566 is being made, facilitating a phylogenomics era with unparalleled opportunity to not only resolve
567 difficult taxonomic relationships but also to better understand the underlying evolutionary
568 processes that have given rise to current biodiversity. In birds, recent large-scale efforts using
569 sequence capture methodologies (McCormack et al. 2013, Prum et al. 2015) and whole-genome
570 sequencing (Jarvis et al. 2014) have helped resolve higher-level interrelationships within
571 Neoaves. Still, some relationships remain poorly resolved or have discordant results across
572 analyses (Suh et al. 2015, Suh 2016, Reddy et al. 2017). Retroposon insertion patterns suggest
573 that a large amount of incomplete lineage sorting (ILS) accompanied a rapid Neoavian radiation
574 (Suh et al. 2015), prompting some authors to propose a hard polytomy involving nine lineages at
575 the base of this group (Suh 2016). Others have stressed the importance of data-type effects in

576 explaining the observed topological incongruence for some clades, and urge researchers to
577 incorporate evidence from as many different marker types as possible when addressing difficult
578 phylogenetic questions (Edwards et al. 2017, Reddy et al. 2017). Together, these results
579 illustrate the advances made with genome-scale data, but also highlight that large amounts of
580 data alone are not necessarily sufficient to resolve deep divergences in the Avian Tree of Life
581 (Philippe et al. 2011).

582 In parallel, there has been a renewed focus on palaeognath phylogeny throughout the past
583 decade largely motivated by reports that the volant (flighted) tinamous are nested within
584 palaeognaths rather than forming the sister to a clade of flightless ratites (Hackett et al. 2008,
585 Harshman et al. 2008, Phillips et al. 2010, Haddrath and Baker 2012, Baker et al. 2014) and
586 more recent findings that the extinct elephant birds of Madagascar are sister to the New Zealand
587 kiwi (Mitchell et al. 2014, Grealy et al. 2017, Yonezawa et al. 2017). These relationships are
588 consistently recovered across recent molecular phylogenies, but conflicting results for other
589 relationships suggest that ILS across short internal branches in palaeognaths could mirror the
590 discord arising from rapid ancient divergences seen in Neoaves.

591 We recover a topology placing rheas as sister to emu/cassowary + kiwi that is congruent
592 across all analyses for MP-EST and ASTRAL species tree methods, but that differs from the
593 placement of rheas as sister to the remaining non-ostrich palaeognaths when concatenated data
594 are analyzed. These conflicting topologies are recovered with maximal support in at least some
595 data sets for both coalescent and concatenation analyses. However, subsampling approaches that
596 incorporate resampling across genes as well as sites (double bootstrapping, Seo et al. 2008,
597 Edwards 2016b) provide more robust support for the coalescent species tree, and this topology is
598 further corroborated by patterns of CR1 retroelement insertions from multiway genome-wide

599 screening. However, conflicting CR1 insertions suggest extensive ILS across short internal
600 branches separating major palaeognath lineages, and coalescent lengths for these pairs of
601 branches fall within the theoretical range expected to produce anomalous gene trees. We indeed
602 find that the most common gene tree for each marker type does not match the inferred species
603 tree topology, consistent with an empirical anomaly zone in palaeognaths.

604 Although we contrast results from concatenation and coalescent species tree methods, we
605 reiterate previous statements that concatenation should be viewed as a specialized case of the
606 multispecies coalescent rather than in strict opposition to it (Liu et al. 2015a,c; Edwards 2016a;
607 Edwards et al. 2016). Concatenation is expected to perform well when coalescent variation
608 among gene trees is low (Liu et al. 2015a,b; Tonini et al. 2015; Mirarab et al. 2016), and we find
609 that most palaeognath relationships are robustly supported by both concatenation and coalescent
610 methods as well as by CR1 insertions. Of particular note, we recover congruent topologies
611 across all sequence-based analyses that nest the tinamous as the sister group to moa within a
612 paraphyletic ratite clade. This moa-tinamou clade is supported even at the level of individual
613 loci, where 94% of both introns and UCEs recover this relationship. Ratite paraphyly is further
614 corroborated by 18 retroelement insertions that are shared by tinamous and non-ostrich
615 palaeognaths, whereas no CR1s supporting ratite monophyly were found.

616 Unlike the case described above, gene tree heterogeneity resulting from coalescent
617 processes can yield topologies with high support for erroneous clades under concatenation, since
618 this method assumes that all genes evolve with a common topology (Kubatko and Degnan 2007;
619 Liu et al. 2015a,c; Roch and Steel 2015; Edwards et al. 2016). In contrast, the multispecies
620 coalescent models gene trees as conditionally independent variables and, by accommodating
621 gene tree heterogeneity, can accurately infer species tree topologies under high levels of ILS and

622 even in the extreme case of the anomaly zone (Liu et al. 2010, 2015c; Edwards et al. 2016;
623 Mirarab et al. 2016). We believe this fundamental difference in approaches to species tree
624 inference underlies the discordant placements of rheas in our concatenation and coalescent
625 analyses. While recognizing caveats pertaining to possible data-type effects of coding sequence
626 (Reddy et al. 2017), we also note that ASTRAL analysis of anchored hybrid enrichment loci by
627 Prum et al. (2015) produced an identical topology to our coalescent species tree for palaeognaths,
628 whereas Bayesian estimation from concatenated loci was identical to our concatenated analyses
629 of CNEEs and UCEs under maximum likelihood.

630 Symmetrical counts for the two minor triplet topologies of rheas, emu + cassowary, and
631 kiwi, as well as the symmetry of conflicting retroelement insertions for these alternative
632 hypotheses, suggest a role for ILS in producing gene tree heterogeneity. Importantly, though,
633 both the major triplet topology and statistical tests of CR1 insertions support rheas as sister to
634 emu/cassowary + kiwi within an overarching bifurcating species tree topology as recovered by
635 MP-EST and ASTRAL (Pamilo and Nei 1988, Kuritzin et al. 2016). Retroelement support for
636 the alternative placement of rheas as sister to all other non-ostrich palaeognaths as found by
637 concatenation is entirely absent, whereas CR1 support for a sister-group relationship of emu and
638 cassowary with tinamous, as found under concatenation for CNEEs and UCEs, is strongly
639 rejected in favor of emu/cassowary + kiwi as occurs in all other analyses, including concatenated
640 introns. Instances of retroposon homoplasy through exact parallel insertions or precise deletion
641 are known, but are believed to be rare (Han et al. 2011, Ray et al. 2006, Suh et al. 2015). These
642 results thus strongly corroborate both the inferred species tree topology and the existence of
643 substantial ILS across short internodes separating several palaeognath lineages. We emphasize
644 that the large amount of gene tree heterogeneity in this study is mostly generated by only two

645 short internodes. Our study thus underscores how small numbers of short branches in the species
646 tree can nonetheless generate substantial heterogeneity of largely similar gene trees differing by
647 a small number of nodes.

648 A question raised for empirical studies is the degree to which gene tree estimation error
649 impacts the performance of summary species tree methods such as MP-EST and ASTRAL that
650 use estimated gene trees as input, since these methods assume gene tree heterogeneity arises
651 from coalescent variation rather than from analytical errors associated with low information
652 content of individual loci, alignment uncertainty, mutational bias, or inappropriate models of
653 sequence evolution (Huang and Knowles 2009, Roch and Warnow 2015, Xi et al. 2015, Blom et
654 al. 2017). We find high average bootstrap support for estimated gene trees and high gene
655 support frequencies for most palaeognath relationships, indicating that marker choice is
656 appropriate to address these questions and gene tree uncertainty is largely confined to the short
657 internal branches inferred to lie within the anomaly zone. Despite this uncertainty, we find that
658 most loci (80-90%) contain variation that is sufficient to support their inferred gene tree topology
659 relative to that of the species tree in likelihood comparisons, and simulations indicate that, for
660 most comparisons, more than 70% of the observed gene tree heterogeneity can be accounted for
661 by coalescent processes alone. While gene tree estimation error certainly occurs in our data set,
662 the consistency of rooted triplet proportions across the three marker types further suggests that,
663 on the whole, estimated gene trees for these data sets provide an adequate and unbiased
664 representation of the coalescent variation in gene tree histories.

665 Simulations have shown that gene tree summary methods can accurately infer the species
666 tree topology despite gene tree estimation error if the sampling of gene trees is sufficiently large
667 and gene tree inference is unbiased (Liu et al. 2015a,c; Xi et al. 2015; Mirarab et al. 2016; Xu

668 and Yang 2016). Adding weak genes to data sets decreased the performance of summary
669 methods such as MP-EST relative to data sets of strong genes alone, but these methods are
670 nevertheless expected to converge upon the correct estimate of the species tree with an
671 increasing number of genes, even if those genes are minimally informative (Xi et al. 2015, Xu
672 and Yang 2016). These inferences could explain observed patterns for CNEEs, which are on
673 average much shorter and less variable than introns or UCEs and have lower bootstrap support
674 for clades in anomalous gene trees that conflict with the species tree, but which still yield
675 congruent estimates of the species tree topology for coalescent methods. Despite lower per-locus
676 information content, the higher alignment certainty and lower estimated homoplasy of CNEEs
677 relative to other marker types (Edwards et al. 2017) could allow CNEEs to converge upon the
678 correct species tree topology for palaeognaths with high support even for intermediate numbers
679 of loci, but only when methods that accommodate gene tree heterogeneity are used. Simulations
680 investigating the effects of gene tree estimation error also assume there is no bias in gene tree
681 estimation, which can cause summary methods to produce inconsistent results (Xi et al. 2015).
682 We doubt that biased gene trees explain the results for palaeognaths since that explanation would
683 necessitate a similar bias across data sets of three different marker types, would require bias
684 prevalent enough to yield consistent results from even small subsampling replicates within
685 marker types, and would have to generate results that are also consistent with patterns of CR1
686 insertions.

687 An important and related question is whether we can confidently detect an empirical
688 anomaly zone given that the same short internal branches expected to produce anomalous gene
689 trees also increase the likelihood of gene tree estimation error due to the short time interval for
690 informative substitutions to accumulate in individual loci of finite length (Huang and Knowles

691 2009). Using highly variable markers might not ameliorate the situation for deep divergences
692 because of the increased probability for subsequent mutations along long terminal branches to
693 obscure informative substitutions on short antecedent branches (Philippe et al. 2011, Haddrath
694 and Baker 2012, Liu et al. 2015b). Huang and Knowles (2009) demonstrated through simulation
695 that the range of species tree branch lengths producing AGTs expands when the mutational
696 variance of estimated gene trees is considered and that unresolved gene trees, rather than AGTs,
697 would predominate within this region of tree space. This expanded anomaly zone might be what
698 is observed for CNEEs in palaeognaths; although ILS between tinamous and other non-ostrich
699 lineages is clearly evident from CR1 retroelements, only CNEEs infer an anomaly zone in this
700 region of the species tree. In contrast, all three marker types support an anomaly zone across
701 branches separating rheas from emu + cassowary and kiwi. For introns and UCEs, the relative
702 likelihood of anomalous gene trees in most cases is substantially greater than if alignments are
703 constrained to the species tree topology and clades that conflict with the species tree tend to have
704 at least 50% median bootstrap support, suggesting that gene tree heterogeneity is real rather than
705 reflecting gene tree estimation error alone. However, we concur that short internodes pose
706 substantial challenges to accurate gene tree inference, and investigations of the empirical
707 anomaly zone should greatly benefit from algorithms that make 'single-step' coestimation of
708 gene trees and species trees scalable to phylogenomic data sets.

709 In conclusion, we find strong evidence that past difficulty in resolving some palaeognath
710 relationships is likely attributable to extensive incomplete lineage sorting within this group, and
711 that species tree methods accommodating gene tree heterogeneity produce robustly supported
712 topologies despite what appears to be an empirical anomaly zone. We echo the sentiments of
713 other authors that high bootstrap support alone is an inadequate measure of confidence in

714 inferred species trees given increasingly large phylogenomic data sets (Edwards 2016b, Sayyari
715 and Mirarab 2016). Congruence across marker types and corroboration from rare genomic
716 changes such as retroposon insertions, as well as phylogenomic subsampling strategies to assess
717 the underlying phylogenetic signal, will increase our confidence in recovered topologies and
718 could also highlight which conflicts are robustly supported and merit further investigation
719 (Tonini et al. 2015, Edwards 2016b, Mirarab et al. 2016, Reddy et al. 2017). Congruence is
720 certainly not a new idea in phylogenetics, but an increasing emphasis on its importance in the era
721 of species trees will continue to advance the field beyond reports of highly supported, but often
722 discordant, 'total evidence' topologies toward a more nuanced 'sum of evidence' approach that
723 considers not just which topologies are produced but also what they can tell us about the
724 underlying evolutionary processes and our attempts to model them.
725

726 **ACKNOWLEDGEMENTS**

727 We thank Liang Liu for advice regarding coalescent simulation, and members of the
728 Edwards lab for helpful discussions. All analyses were run on the Odyssey cluster supported by
729 the FAS Division of Science, Research Computing Group at Harvard University. Support was
730 provided by NSF grant DEB 1355343 (EAR 1355292) to AJB, MC, and SVE.
731

732 **LITERATURE CITED**

733 Angelis, K., Dos Reis, M. 2015. The impact of ancestral population size and incomplete lineage
734 sorting on Bayesian estimates of species divergence times. Curr. Zool. 61:874–885.
735 Baker, A.J., Haddrath, O., McPherson, J.D., Cloutier, A. 2014. Genomic support for a moa-
736 tinamou clade and adaptive morphological convergence in flightless ratites. Mol. Biol.

737 Evol. 31:1686–1696.
738 Blom, M.P.K., Bragg, J.G., Potter, S., Moritz, C. 2017. Accounting for uncertainty in gene tree
739 estimation: summary-coalescent species tree inference in a challenging radiation of
740 Australian lizards. Syst. Biol. 66:352–366.
741 Bogdanowicz, D., Giaro, K., Wrobel, B. 2012. TreeCmp: comparison of trees in polynomial
742 time. Evol. Bioinform. 8:475–487.
743 Burnham, K.P., Anderson, D.R. 2002. Model selection and multimodel inference: a practical
744 information-theoretic approach. Second Edition. New York: Springer.
745 Cloutier, A., Sackton, T.B., Grayson, P., Edwards, S.V., Baker, A.J. 2018. First nuclear genome
746 assembly of an extinct moa species, the little bush moa (*Anomalopteryx didiformis*).
747 bioRxiv [DOI pending].
748 Cracraft, J., Barker, F.K., Braun, M., Harshman, J., Dyke, G.J., Feinstein, J., Stanley, S., Cibois,
749 A., Schikler, P., Beresford, P., García-Moreno, J., Sorenson, M.D., Yuri, T., Mindell,
750 D.P. 2004. Phylogenetic relationships among modern birds (Neornithes): toward an avian
751 tree of life. In: Cracraft, J., Donoghue, M.J. (eds.), Assembling the Tree of Life. Pp. 468–
752 489. New York: Oxford University Press.
753 Degnan, J.H., Rosenberg, N.A. 2006. Discordance of species trees with their most likely gene
754 trees. PLoS Genet. 2:762–768.
755 Degnan, J.H., Rosenberg, N.A. 2009. Gene tree discordance, phylogenetic inference and the
756 multispecies coalescent. Trends Ecol. Evol. 24:332–340.
757 Delsuc, F., Brinkmann, H., Philippe, H. 2005. Phylogenomics and the reconstruction of the tree
758 of life. Nat. Rev. Genet. 6:361–375.
759 Edwards, S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution.

760 63:1–19.
761 Edwards, S. 2016a. Species trees, inference of. In: Kliman, R.M. (ed.), Encyclopedia of
762 Evolutionary Biology. Vol. 4, Pp. 236–244. Oxford: Academic Press.
763 Edwards, S.V. 2016b. Phylogenomic subsampling: a brief review. Zool. Scripta 45:63–74.
764 Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu,
765 S., Lemmon, E.M., Lemmon, A.R., Leache, A.D., Liu, L., Davis, C.C. 2016.
766 Implementing and testing the multispecies coalescent model: a valuable paradigm for
767 phylogenomics. Mol. Phylogenet. Evol. 94:447–462.
768 Edwards, S.V., Cloutier, A., Baker, A.J. 2017. Conserved nonexonic elements: a novel class of
769 marker for phylogenomics. Syst. Biol. 66:1028–1044.
770 Eisen, J., Fraser, C. 2003. Phylogenomics: intersection of evolution and genomics. Science
771 300:1706–1707.
772 Gee, H. 2003. Evolution – Ending incongruence. Nature 425:782.
773 Grealy, A., Phillips, M., Miller, G., Gilbert, M.T.P., Rouillard, J.-M., Lambert, D., Bunce, M.,
774 Haile, J. 2017. Eggshell palaeogenomics: palaeognath evolutionary history revealed
775 through ancient nuclear and mitochondrial DNA from Madagascan elephant bird
776 (*Aepyornis* sp.) eggshell. Mol. Phylogenet. Evol. 109:151–163.
777 Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J., Chojnowski,
778 J.L., Cox, W.A., Han, K.-L., Harshman, J., Huddleston, C.J., Marks, B.D., Miglia, K.J.,
779 Moore, W.S., Sheldon, F.H., Steadman, D.W., Witt, C.C., Yuri, T. 2008. A
780 phylogenomic study of birds reveals their evolutionary history. Science 320:1763–1768.
781 Haddrath, O., Baker, A.J. 2012. Multiple nuclear genes and retroposons support vicariance and
782 dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. Proc. R.

783 Soc. Lond. B. 279:4617–4625.
784 Han, K.-L., Braun, E.L., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, M.J., Chojnowski,
785 J.L., Hackett, S.J., Harshman, J., Huddleston, C.J., Marks, B.D., Miglia, K.J., Moore,
786 W.S., Sheldon, F.H., Steadman, D.W., Witt, C.C., Yuri, T. 2011. Are transposable
787 element insertions homoplasy free?: an examination using the avian tree of life. Syst.
788 Biol. 60:375–386.
789 Harshman, J., Braun, E.L., Braun, M.J., Huddleston, C.J., Bowie, R.C.K., Chojnowski, J.L.,
790 Hackett, S.J., Han, K.-L., Kimball, R.T., Marks, B.D., Miglia, K.J., Moore, W.S., Reddy,
791 S., Sheldon, F.H., Steadman, D.W., Steppan, S.J., Witt, C.C., Yuri, T. 2008.
792 Phylogenomic evidence for multiple losses of flight in ratite birds. Proc. Natl. Acad. Sci.
793 U. S. A. 105:13462–13467.
794 Hickey, G., Paten, B., Earl, D., Zerbino, D., Haussler, D. 2013. HAL: a hierarchical format for
795 storing and analyzing multiple genome alignments. Bioinformatics 29:1341–1342.
796 Houde, P.W. 1988. Palaeognathous birds from the early tertiary of the northern hemisphere.
797 Paynter, R.A. (ed.). Cambridge, MA: Publications of the Nuttall Ornithological Club.
798 Huang, H., Knowles, L.L. 2009. What is the danger of the anomaly zone for empirical
799 phylogenetics? Syst. Biol. 58:527–536.
800 Huerta-Cepas, J., Serra, F., Bork, P. 2016. ETE 3: reconstruction, analysis, and visualization of
801 phylogenomic data. Mol. Biol. Evol. 33:1635–1638.
802 Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C.,
803 Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Li, J., Zhang, F., Li,
804 H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M.S., Zavidovych,
805 V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B.,

806 Munch, K., Schierup, M., Lindow, B., Warren, W.C., Ray, D., Green, R.E., Bruford,
807 M.W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E.P., Bertelsen, M.F.,
808 Sheldon, F.H., Brumfield, R.T., Mello, C.V., Lovell, P.V., Writhlin, M., Schneider,
809 M.P.C., Prosdocimi, F., Samaniego, J.A., Velazquez, A.M.V., Alfaro-Núñez, A.,
810 Campos, P.F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce,
811 M., Lambert, D.M., Zhou, Q., Perelman, P., Driskell, A.C., Shapiro, B., Xiong, Z., Zeng,
812 Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yingi, X., Zheng, Q., Zhang, Y., Yang, H.,
813 Wang, J., Smeds, L., Rheindt, F.E., Braun, M., Fjeldsa, J., Orlando, L., Barker, F.K.,
814 Jønsson, K.A., Johnson, W., Koepfli, K.-P., O'Brien, S., Haussler, D., Ryder, O.A.,
815 Rahbek, C., Willerslev, E., Graves, G.R., Glenn, T.C., McCormack, J., Burt, D., Ellegren,
816 H., Alström, P., Edwards, S.V., Stamatakis, A., Mindell, D.P., Cracraft, J., Braun, E.L.,
817 Warnow, T., Jun, W., Gilbert, M.T.P., Zhang, G. 2014. Whole-genome analyses resolve
818 early branches in the tree of life of modern birds. Science 346:1320–1331.
819 Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C.,
820 Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Alfaro-Nunez, A.,
821 Narula, N., Liu, L., Burt, D., Ellegren, H., Edwards, S.V., Stamatakis, A., Mindell, D.P.,
822 Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T.P., Zhang, G. 2015.
823 Phylogenomic analyses data of the avian phylogenomics project. GigaScience 4:4.
824 Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H. 2006. Phylogenomics: the beginning of
825 incongruence? Trends Genet. 22:225–231.
826 Katoh, K., Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7:
827 improvements in performance and usability. Mol. Biol. Evol. 30:772–780.
828 Kozlov, A.M., Aberer, A.J., Stamatakis, A. 2015. ExaML version 3: a tool for phylogenomic

829 analyses on supercomputers. Bioinformatics 31:2577–2579.
830 Kubatko, L.S., Degnan, J.H. 2007. Inconsistency of phylogenetic estimates from concatenated
831 data under coalescence. Syst. Biol. 56:17–24.
832 Kuritzin, A., Kischka, T., Schmitz, J., Churakov, G. 2016. Incomplete lineage sorting and
833 hybridization statistics for large-scale retroposon insertion data. PLoS Comput. Biol.
834 12:e1004812.
835 Le Duc, D., Renaud, G., Krishnan, A., Almén, M.S., Huynen, L., Prohaska, S.J., Ongyerth, M.,
836 Bitarello, B.D., Schiöth, H.B., Hofreiter, M., Stadler, P.F., Prüfer, K., Lambert, D., Kelso,
837 J., Schöneberg, T. 2015. Kiwi genome provides insights into evolution of a nocturnal
838 lifestyle. Genome Biol. 16:147.
839 Linkem, C.W., Minin, V.N., Leache, A.D. 2016. Detecting the anomaly zone in species trees and
840 evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae).
841 Syst. Biol. 65:465–477.
842 Liu, L., Yu, L. 2010. Phybase: an R package for species tree analysis. Bioinformatics 26:962–
843 963.
844 Liu, L., Yu, L., Edwards, S.V. 2010. A maximum pseudo-likelihood approach for estimating
845 species trees under the coalescent model. BMC Evol. Biol. 10:302.
846 Liu, L., Wu, S., Yu, L. 2015a. Coalescent methods for estimating species trees from
847 phylogenomic data. J. Syst. Evol. 53:380–390.
848 Liu, L., Xi, Z., Davis, C.C. 2015b. Coalescent methods are robust to the simultaneous effects of
849 long branches and incomplete lineage sorting. Mol. Biol. Evol. 32:791–805.
850 Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S.V. 2015c. Estimating phylogenetic trees from
851 genome-scale data. Ann. N.Y. Acad. Sci. 1360:36–53.

852 Maddison, W. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.
853 McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield, R.T.
854 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and
855 high-throughput sequencing. PLOS ONE 8:e54848.
856 Mendes, F.K., Hahn, M.W. 2016. Gene tree discordance causes apparent substitution rate
857 variation. Syst. Biol. 65:711–721.
858 Mirarab, S., Warnow, T. 2015. ASTRAL-II: coalescent-based species tree estimation with many
859 hundreds of taxa and thousands of genes. Bioinformatics 31:44–52.
860 Mirarab, S., Bayzid, M.S., Warnow, T. 2016. Evaluating summary methods for multilocus
861 species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65:366–
862 380.
863 Mitchell, K.J., Llamas, B., Soubrier, J., Rawlence, N.J., Worthy, T.H., Wood, J., Lee, M.S.Y.,
864 Cooper, A. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and
865 clarifies ratite bird evolution. Science 344:898–900.
866 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q. 2015. IQ-TREE: a fast and
867 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol.
868 Biol. Evol. 32:268–274.
869 Pamilo, P., Nei, M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol.
870 5:568–583.
871 Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, T.J., Manuel, M., Wörheide, G.,
872 Baurain, D. 2011. Resolving difficult phylogenetic questions: why more sequences are
873 not enough. PLoS Biol. 9:e1000602.
874 Phillips, M.J., Gibb, G.C., Crimp, E.A., Penny, D. 2010. Tinamous and moa flock together:

875 mitochondrial genome sequence analysis reveals independent losses of flight among
876 ratites. Syst. Biol. 59:90–107.
877 Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J.R., Walenz, B., Koren, S., Sutton, G.,
878 Kodira, C., Winer, R., Knight, J.R., Mullikin, J.C., Meader, S.J., Ponting, C.P., Lunter,
879 G., Higashino, S., Hobolth, A., Dutheil, J., Karakoç, E., Alkan, C., Sajjadian, S.,
880 Catacchio, C.R., Ventura, M., Marques-Bonet, T., Eichler, E.E., André, C., Atencia, R.,
881 Mugisha, L., Junhold, J., Patterson, N., Siebauer, M., Good, J.M., Fischer, A., Ptak, S.E.,
882 Lachmann, M., Symer, D.E., Mailund, T., Schierup, M. H., Andrés, A.M., Kelso, J.,
883 Pääbo, S. 2012. The bonobo genome compared with the chimpanzee and human
884 genomes. Nature 486:527–531.
885 Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon,
886 A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation
887 DNA sequencing. Nature 526:569–573.
888 Quinlan, A.R., Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic
889 features. Bioinformatics 26:841–842.
890 Ray, D.A., Xing, J., Salem, A.-H., Batzer, M.A. 2006. SINEs of a nearly perfect character. Syst.
891 Biol. 55:928–935.
892 Reddy, S., Kimball, R.T., Pandey, A., Hosner, P.A., Braun, M.J., Hackett, S.J., Han, K.-L.,
893 Harshman, J., Huddleston, C.J., Kingston, S., Marks, B.D., Miglia, K.J., Moore, W.S.,
894 Sheldon, F.H., Witt, C.C., Yuri, T., Braun, E.L. 2017. Why do phylogenomic data sets
895 yield conflicting trees? Data type influences the avian tree of life more than taxon
896 sampling. Syst. Biol. 66:857–879.
897 Roch, S., Steel, M. 2015. Likelihood-based tree reconstruction on a concatenation of aligned

898 sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100:56–62.
899 Roch, S., Warnow, T. 2015. On the robustness to gene tree estimation error (or lack thereof) of
900 coalescent-based species tree methods. Syst. Biol. 64:663–676.
901 Rokas, A., Williams, B., King, N., Carroll, S. 2003. Genome-scale approaches to resolving
902 incongruence in molecular phylogenies. Nature 425:798–804.
903 Rosenberg, N.A. 2013. Discordance of species trees with their most likely gene trees: a unifying
904 principle. Mol. Biol. Evol. 30:2709–2713.
905 Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J., Wheeler, N., Gardner, P., Clarke, J.,
906 Baker, A.J., Clamp, M., Edwards, S.V. 2018. Convergent regulatory evolution and the
907 origin of flightlessness in palaeognathous birds. bioRxiv doi:
908 https://doi.org/10.1101/262584.
909 Salichos, L., Stamatakis, A., Rokas, A. 2014. Novel information theory-based measures for
910 quantifying incongruence among phylogenetic trees. Mol. Biol. Evol. 31:1261–1271.
911 Sayyari, E., Mirarab, S. 2016. Fast coalescent-based computation of local branch support from
912 quartet frequencies. Mol. Biol. Evol. 33:1654–1668.
913 Seo, T.-K. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence
914 data. Mol. Biol. Evol. 25:960–971.
915 Smit, A.F.A., Hubley, R., Green, P. 2015. RepeatMasker Open-4.0.
916 http://www.repeatmasker.org.
917 Smith, J.V., Braun, E.L., Kimball, R.T. 2013. Ratite nonmonophyly: independent evidence from
918 40 novel loci. Syst. Biol. 62:35–49.
919 Song, S., Liu, L., Edwards, S.V., Wu, S. 2012. Resolving conflict in eutherian mammal
920 phylogeny using phylogenomics and the multispecies coalescent model. Proc. Natl. Acad.

921 Sci. U. S. A. 109:14942–14947.
922 Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
923 large phylogenies. Bioinformatics 30:1312–1313.
924 Suh, A., Smeds, L., Ellegren, H. 2015. The dynamics of incomplete lineage sorting across the
925 ancient adaptive radiation of Neoavian birds. PLoS Biol. 13:e1002224.
926 Suh, A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of
927 Neoaves. Zool. Scripta 45:50–62.
928 Swofford, D.L. 2002. Phylogenetic analysis using parsimony (*and other methods). Version 4.
929 Sinauer Associates, Sunderland, MA.
930 Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Orti, G. 2015. Concatenation and species
931 tree methods exhibit statistically indistinguishable accuracy under a range of simulated
932 conditions. PLoS Currents 7:10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be.
933 Xi, Z., Liu, L., Davis, C.C. 2015. Genes with minimal phylogenetic information are problematic
934 for coalescent analyses when gene tree estimation is biased. Mol. Phylogenet. Evol.
935 92:63–71.
936 Xu, B., Yang, Z. 2016. Challenges in species tree estimation under the multispecies coalescent
937 model. Genetics 204:1353–1368.
938 Yonezawa, T., Segawa, T., Mori, H., Campos, P.F., Hongoh, Y., Endo, H., Akiyoshi, A., Kohno,
939 N., Nishida, S., Wu, J., Jin, H., Adachi, J., Kishino, H., Kurokawa, K., Nogi, Y., Tanabe,
940 H., Mukoyama, H., Yoshida, K., Rasoamiaramanana, A., Yamagishi, S., Hayashi, Y.,
941 Yoshida, A., Koike, H., Akishinonomiya, F., Willerslev, E., Hasegawa, M. 2017.
942 Phylogenomics and morphology of extinct paleognaths reveal the origin and evolution of
943 the ratites. Curr. Biol. 27:68–77.

944 Zhang, G., Li, B., Li, C., Gilbert, M.T.P., Jarvis, E.D., Wang, J. 2014. Comparative genomic data
945 of the avian phylogenomics project. GigaScience 3:26.
946

947 **FIGURE CAPTIONS**

948 **Figure 1** Palaeognath relationships inferred from the total evidence data set (TENT) of 20,850
949 loci using MP-EST (a) and ASTRAL (b) species tree methods, and maximum likelihood analysis
950 of concatenated data with ExaML (c). Bootstrap support values are indicated for clades with <
951 100% support. Scale bars indicate branch length in coalescent units for MP-EST and ASTRAL
952 and in substitutions/site for ExaML. Terminal branches in MP-EST and ASTRAL species trees
953 are uninformative and are drawn as a constant value across taxa.

954 **Figure 2** Support for alternative hypotheses of the sister group to rheas from phylogenomic
955 subsampling using MP-EST (a-c), ASTRAL (d-f), and ExaML (g-i). Plots display the mean
956 bootstrap support for each hypothesis from 10 replicates of randomly sampled loci within each
957 data set size category (e.g. 50–3000 loci, shown on x-axis).

958 **Figure 3** Heatmap of the number of replicates from phylogenomic subsampling that support
959 each alternative hypothesis for the sister group to rheas at a minimum bootstrap support of 70%
960 (a-i) and 90% (j-r). Rows labelled H1-H5 within each panel correspond to the five alternative
961 hypotheses outlined in the legend. Columns within each panel labelled 50-3000 indicate the
962 number of loci subsampled at random with replacement from all loci for each marker type.
963 Coloring of cells indicates the number of replicates (of 10 in total for each data set size category)
964 that support each hypothesis at the given bootstrap cut off, corresponding to the colouring
965 scheme outlined in the legend.

966 **Figure 4** CR1 retroelements corroborate the inferred species tree topology from MP-EST and

967 ASTRAL (a), and also display conflicting insertion patterns consistent with incomplete lineage
968 sorting (b-e). Numbers in colored balls indicate the number of shared retroelement insertions for
969 each clade. Bird illustrations were obtained under the Creative Commons License from
970 Wikimedia Commons (https://commons.wikimedia.org; full details are given in Supplemental
971 Table S2).

972 **Figure 5** Distributions of the 50 most common gene tree topologies for CNEEs (a), introns (b),
973 and UCEs (c), showing that the most common gene tree topology for each marker type does not
974 match the inferred species tree topology. Symbols and bar coloring in parts a-c correspond to the
975 topologies shown in (d).

976 **Figure 6** Gene tree support for clades recovered in the MP-EST and ASTRAL species trees for
977 CNEEs (a), introns (b), and UCEs (c). Gene support frequency (GSF, the percent of gene trees
978 containing each clade) is given beneath branches, with heatmap coloring of branches according
979 to GSF as indicated in the legend. Values above branches give the internode certainty 'all'
980 statistic (ICA, Salichos et al. 2014), indicating support for each clade relative to all other
981 conflicting bipartitions in the set of gene trees.

982 **Figure 7** Support for observed topological heterogeneity in estimated gene trees. (a-c)
983 Distributions for the average bootstrap support for all clades recovered in majority rule extended
984 consensus gene trees ('MRE consensus topology', drawn as solid bars), and the average support
985 when bootstrap replicates for each gene are constrained to the inferred species tree topology
986 (open bars). Values above each panel indicate the decrease in the grand mean of average support
987 across genes when bootstrap replicates are constrained to the species tree topology. (d) Support
988 that observed gene tree topologies differ from the inferred species tree. Dark gray bars indicate
989 the difference in Akaike information criterion (AIC) from values calculated using the MRE

990 consensus topology for each gene, relative to that obtained when the sequence alignment is
991 constrained to the species tree topology. The baseline at 0 corresponds to ΔAIC = -2, thus values
992 beneath this line indicate loci where likelihood values support the gene tree topology
993 substantially better than that of the species tree. Results of approximately unbiased (AU) tests
994 are indicated with medium gray bars showing the proportion of loci that reject (below baseline)
995 and fail to reject (above baseline) the species tree topology at a P-value cut off of 0.05, and light
996 gray bars showing the proportion of loci where the species tree topology occurs among the top
997 5% of candidate topologies (above baseline), or is within the bottom 95% of tested topologies
998 (below baseline) for AU P-values ranked in ascending order. AU tests for the 105 and 1575
999 candidate tree sets produced similar results and are shown for the 105 candidate set only.
1000 **Figure 8** Pairwise distances between gene trees and the MP-EST/ASTRAL species tree
1001 topology. Distances between each gene tree and the species tree topology were calculated as the
1002 matching cluster distance (a) or Robinson-Foulds cluster distance (b) on outgroup rooted trees,
1003 and mean values for all pairwise gene tree–species tree distances within each category are
1004 shown. Error bars indicate the 95% confidence interval of the mean. Distances were calculated
1005 for empirically estimated gene trees and for data sets of 10,000 gene trees that were simulated
1006 using coalescent branch lengths from the MP-EST and ASTRAL species trees. Values above
1007 bars for simulated data sets indicate the ratios of means for simulated data sets compared to the
1008 mean for empirically estimated gene trees.
1009 **Figure 9** Proportions of rooted triplets for the three alternative topologies involving rheas, kiwi,
1010 and emu + cassowary. (a-c) Observed proportions of the major and two minor topologies from
1011 empirically estimated gene trees. (d-f) Expected proportions based on the length of the internal
1012 branch within the triplet (e.g. the common ancestor of kiwi + emu/cassowary) in coalescent

1013 units, following Pamilo and Nei (1988). For d-f, pie charts on the left and right give expected
1014 proportions based on branch lengths from MP-EST and ASTRAL species trees, respectively.
1015 **Figure 10** Pairs of short successive internal branches are consistent with expectations for the
1016 anomaly zone. Species tree topologies for CNEEs (a), introns (b), and UCEs (c) are shown with
1017 internal branch lengths in coalescent units estimated from best maximum likelihood gene trees
1018 with MP-EST. Terminal branch lengths are uninformative and are drawn as a constant value
1019 across taxa. Coalescent branch lengths for all pairs of branches (x and y) are given in (d), with
1020 a(x) calculated following Equation 4 from Degnan and Rosenberg (2006). Anomalous gene trees
1021 are expected when $y < a(x)$. Clades fulfilling this anomaly zone criterion are shaded in (d), with
1022 the corresponding branches indicated in a-c.
1023 **Suppl. Figure S1** Palaeognath phylogenetic relationships inferred with MP-EST (a-d) and
1024 ASTRAL (e-h) coalescent-based species tree methods and maximum-likelihood inference of
1025 fully partitioned concatenated alignments with ExaML (i-l). Topologies were inferred from
1026 12,676 conserved non-exonic elements (CNEEs), 5,016 introns, 3,158 ultraconserved elements
1027 (UCEs), and the total evidence data set of 20,850 loci combined across marker types (TENT).
1028 Bootstrap supports are drawn for clades with < 100% support.
1029 **Suppl. Figure S2** Support for alternative hypotheses of the sister group to emu + cassowary
1030 from phylogenomic subsampling using MP-EST (a-c), ASTRAL (d-f), and ExaML (g-i). Plots
1031 display the mean bootstrap support for each hypothesis from 10 replicates of randomly sampled
1032 loci within each data set size category (e.g. 50–3000 loci, shown on x-axis).
1033 **Suppl. Figure S3** Heatmap of the number of replicates from phylogenomic subsampling that
1034 support each alternative hypothesis for the sister group to emu + cassowary at a minimum
1035 bootstrap support of 70% (a-i) and 90% (j-r). Rows labelled H1-H4 within each panel

1036 correspond to the four alternative hypotheses outlined in the legend. Columns within each panel
1037 labelled 50-3000 indicate the number of loci subsampled at random with replacement from all
1038 loci for each marker type. Coloring of cells indicates the number of replicates (of 10 in total for
1039 each data set size category) that support each hypothesis at the given bootstrap cut off,
1040 corresponding to the colouring scheme outlined in the legend.

1041 **Suppl. Figure S4** Support for the four most common gene tree topologies for each marker type.
1042 Median bootstrap support (a-c) and the median number of substitutions under a parsimony
1043 criterion (d-f) for branches in gene tree topologies that conflict with the species tree. Letters
1044 below each group refer to topologies shown in (j), with numbers referring to branches labelled on
1045 those topologies. Reference lines in (a-c) are drawn at 50% bootstrap support. (g-i) Violin plots
1046 of ΔAIC for loci belonging to each AGT category. Values show the difference in AIC for
1047 sequence alignments given the majority rule extended consensus gene tree topology, relative to
1048 the AIC when sequence alignments are constrained to the species tree topology. A reference line
1049 is drawn at $\Delta AIC_{(\text{gene tree–species tree})} = -2$, with values beneath this cut off indicating substantially
1050 stronger support in favor of the gene tree topology (Burnham and Anderson 2002).

1051 **Suppl. Figure S5** Summary measures of aligned sequence data for each marker type (CNEEs,
1052 introns, UCEs), showing the total aligned sequence length per locus (a-c), variable alignment
1053 columns as a percent of the total alignment length (d-f), and the percent of parsimony
1054 informative alignment columns per locus (g-i).

1055 **Suppl. Figure S6** Pairs of short successive internal branches are consistent with expectations for
1056 the anomaly zone. Species tree topologies for CNEEs (a), introns (b), and UCEs (c) are shown
1057 with internal branch lengths in coalescent units estimated from best maximum likelihood gene
1058 trees with ASTRAL. Terminal branch lengths are uninformative and are drawn as a constant

1059 value across taxa. Coalescent branch lengths for all pairs of branches (x and y) are given in (d),
1060 with a(x) calculated following Equation 4 from Degnan and Rosenberg (2006). Anomalous gene
1061 trees are expected when $y < a(x)$. Clades fulfilling this anomaly zone criterion are shaded in (d),
1062 with the corresponding branches indicated in a-c.

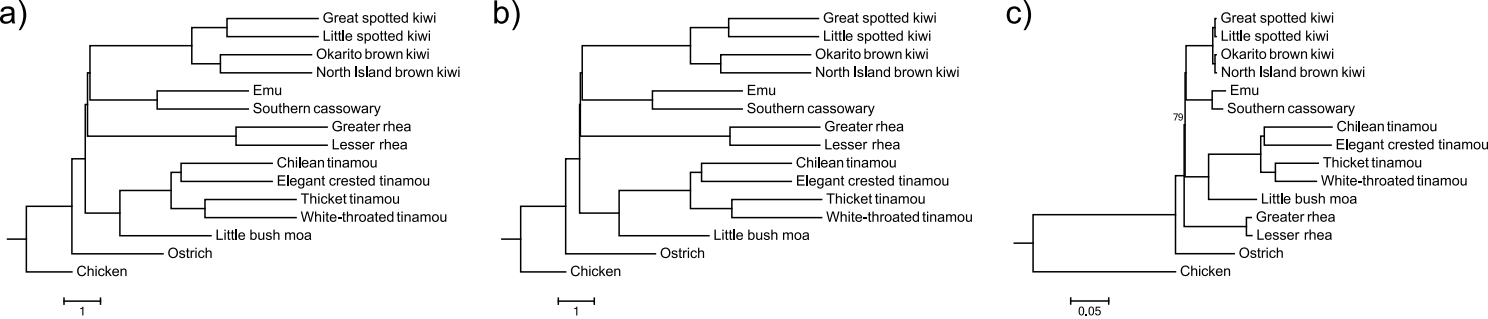

a)
Great spotted kiwi
Little spotted kiwi
Okarito brown kiwi
North Island brown kiwi
Emu
Southern cassowary
Greater rhea
Lesser rhea
Chilean tinamou
Elegant crested tinamou
Thicket tinamou
White-throated tinamou
Little bush moa
Ostrich
Chicken
1
b)
Great spotted kiwi
Little spotted kiwi
Okarito brown kiwi
North Island brown kiwi
Emu
Southern cassowary
Greater rhea
Lesser rhea
Chilean tinamou
Elegant crested tinamou
Thicket tinamou
White-throated tinamou
Little bush moa
Ostrich
Chicken
1
c)
Great spotted kiwi
Little spotted kiwi
Okarito brown kiwi
North Island brown kiwi
Emu
Southern cassowary
79
Chilean tinamou
Elegant crested tinamou
Thicket tinamou
White-throated tinamou
Little bush moa
Greater rhea
Lesser rhea
Ostrich
Chicken
0.05

Fig. 1

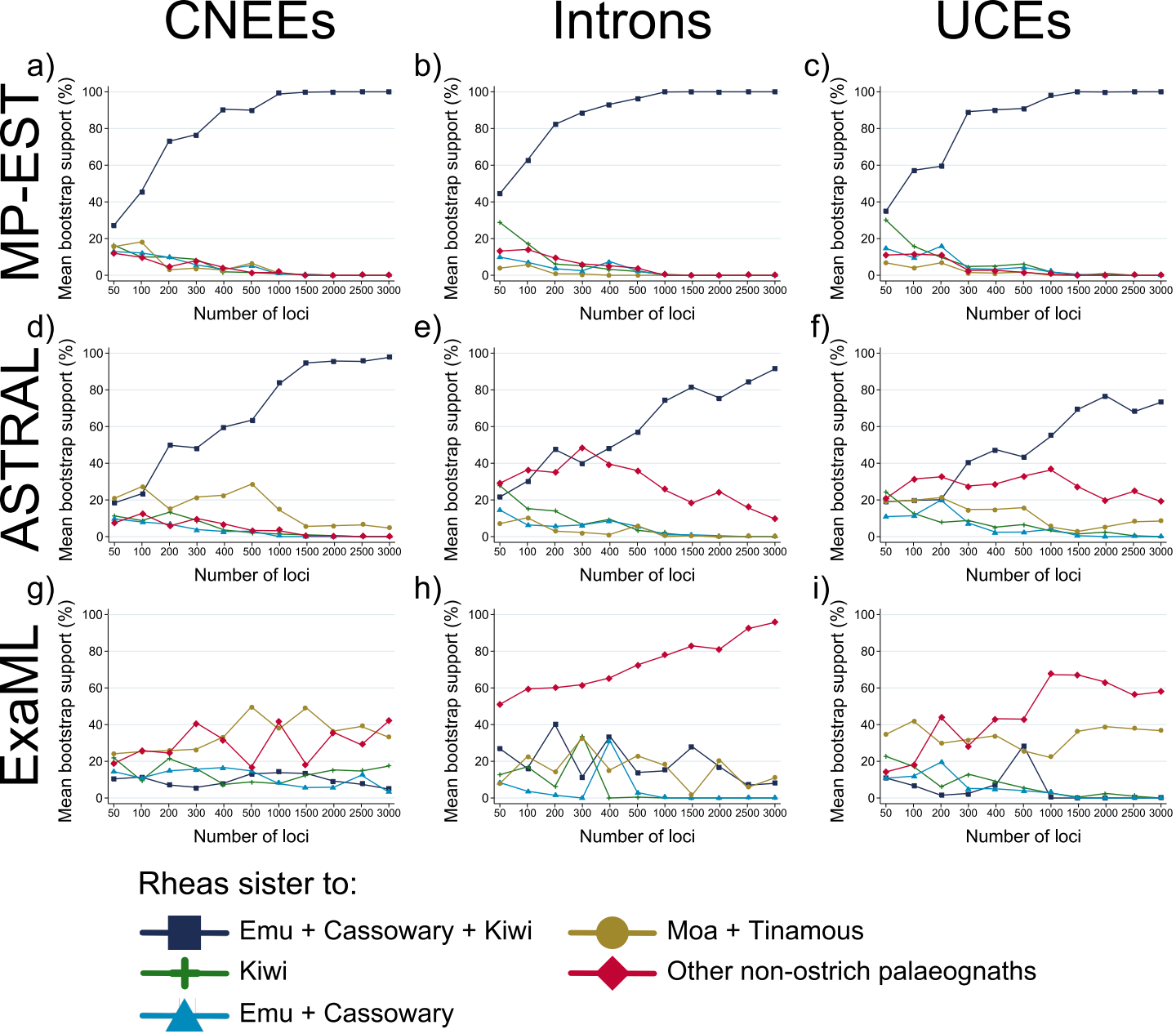

CNEEs
Introns
UCEs
MP-EST
ASTRAL
ExaML
a)
b)
c)
d)
e)
f)
g)
h)
i)
Mean bootstrap support (%)
Number of loci
0
20
40
60
80
100
50
100
200
300
400
500
1000
1500
2000
2500
3000
Rheas sister to:
Emu + Cassowary + Kiwi
Kiwi
Emu + Cassowary
Moa + Tinamous
Other non-ostrich palaeognaths

Fig. 2

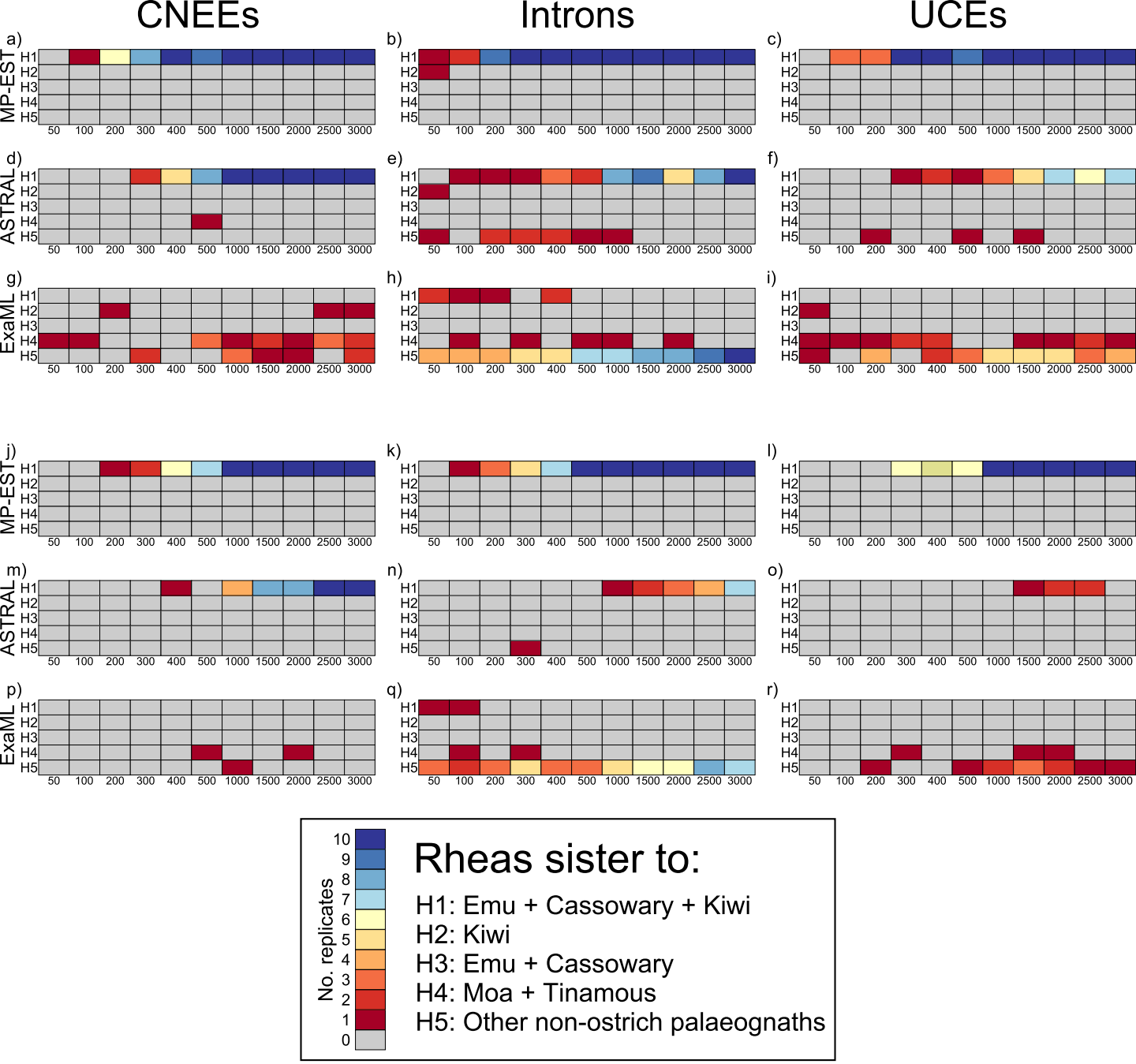

CNEEs
Introns
UCEs
MP-EST
ASTRAL
ExaML
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
o)
p)
q)
r)
H1
H2
H3
H4
H5
50
100
200
300
400
500
1000
1500
2000
2500
3000
No. replicates
10
9
8
7
6
5
4
3
2
1
0
Rheas sister to:
H1: Emu + Cassowary + Kiwi
H2: Kiwi
H3: Emu + Cassowary
H4: Moa + Tinamous
H5: Other non-ostrich palaeognaths

Fig. 3

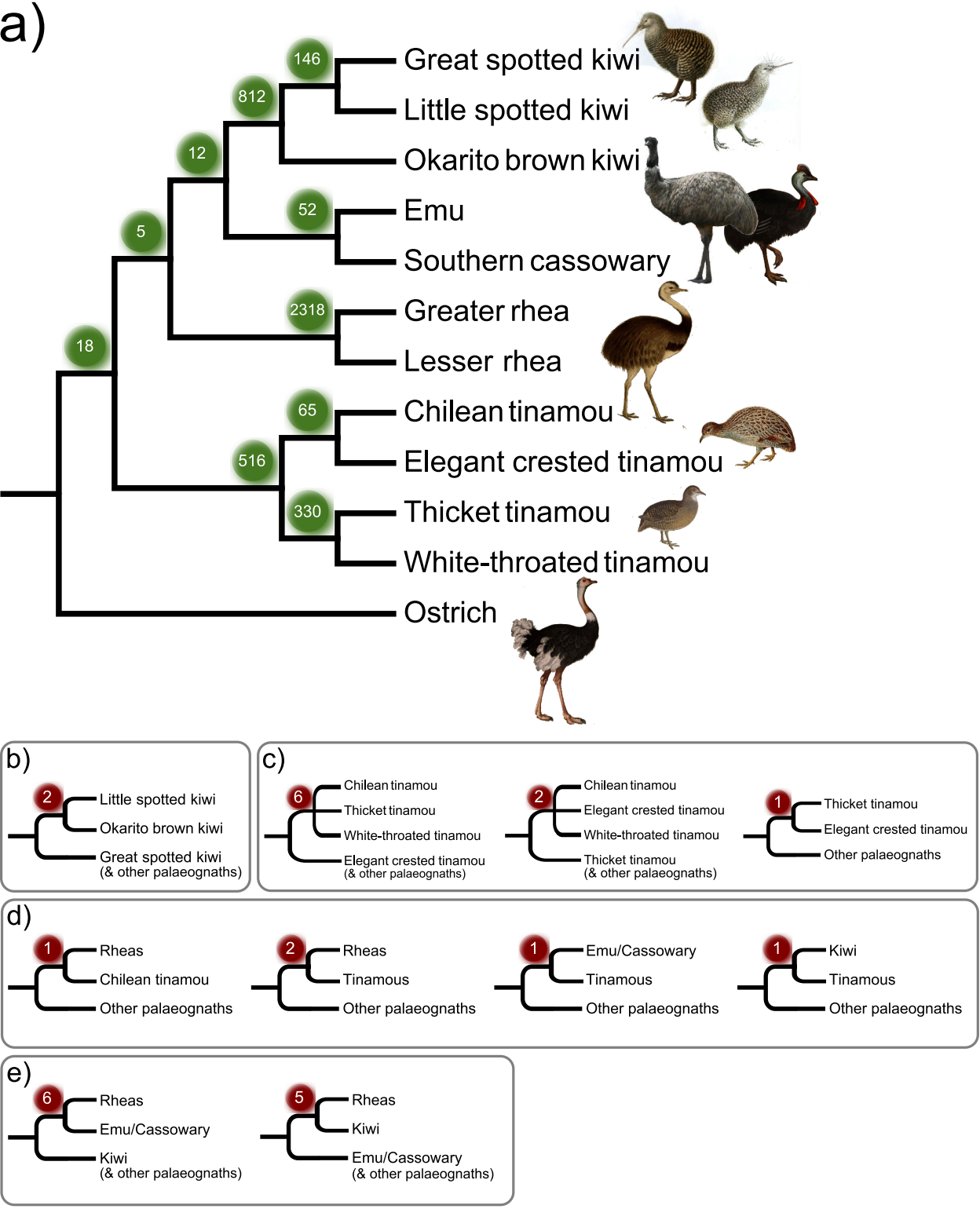

a)
146
812
12
52
5
2318
18
65
516
330
Great spotted kiwi
Little spotted kiwi
Okarito brown kiwi
Emu
Southern cassowary
Greater rhea
Lesser rhea
Chilean tinamou
Elegant crested tinamou
Thicket tinamou
White-throated tinamou
Ostrich
b)
2
Little spotted kiwi
Okarito brown kiwi
Great spotted kiwi
(& other palaeognaths)
c)
6
Chilean tinamou
Thicket tinamou
White-throated tinamou
Elegant crested tinamou
(& other palaeognaths)
2
Chilean tinamou
Elegant crested tinamou
White-throated tinamou
Thicket tinamou
(& other palaeognaths)
1
Thicket tinamou
Elegant crested tinamou
Other palaeognaths
d)
1
Rheas
Chilean tinamou
Other palaeognaths
2
Rheas
Tinamous
Other palaeognaths
1
Emu/Cassowary
Tinamous
Other palaeognaths
1
Kiwi
Tinamous
Other palaeognaths
e)
6
Rheas
Emu/Cassowary
Kiwi
(& other palaeognaths)
5
Rheas
Kiwi
Emu/Cassowary
(& other palaeognaths)

Fig. 4

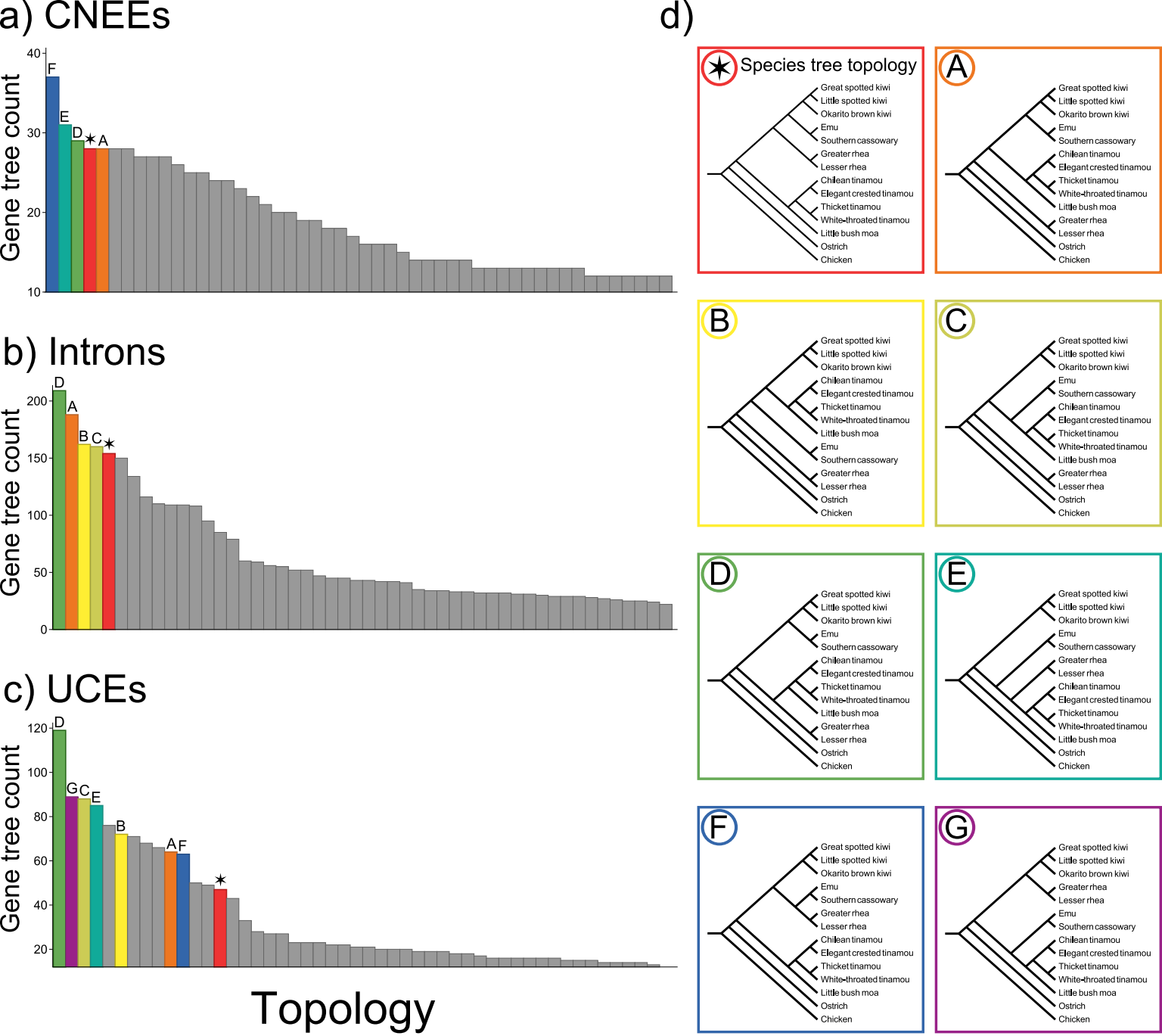

a) CNEEs
Gene tree count
40
30
20
10
F
E
D
A
b) Introns
Gene tree count
200
150
100
50
0
D
A
B
C
c) UCEs
Gene tree count
120
100
80
60
40
20
D
G
C
E
B
A
F
Topology
d)
Species tree topology
A
B
C
D
E
F
G
Great spotted kiwi
Little spotted kiwi
Okarito brown kiwi
Emu
Southern cassowary
Greater rhea
Lesser rhea
Chilean tinamou
Elegant crested tinamou
Thicket tinamou
White-throated tinamou
Little bush moa
Ostrich
Chicken

Fig. 5

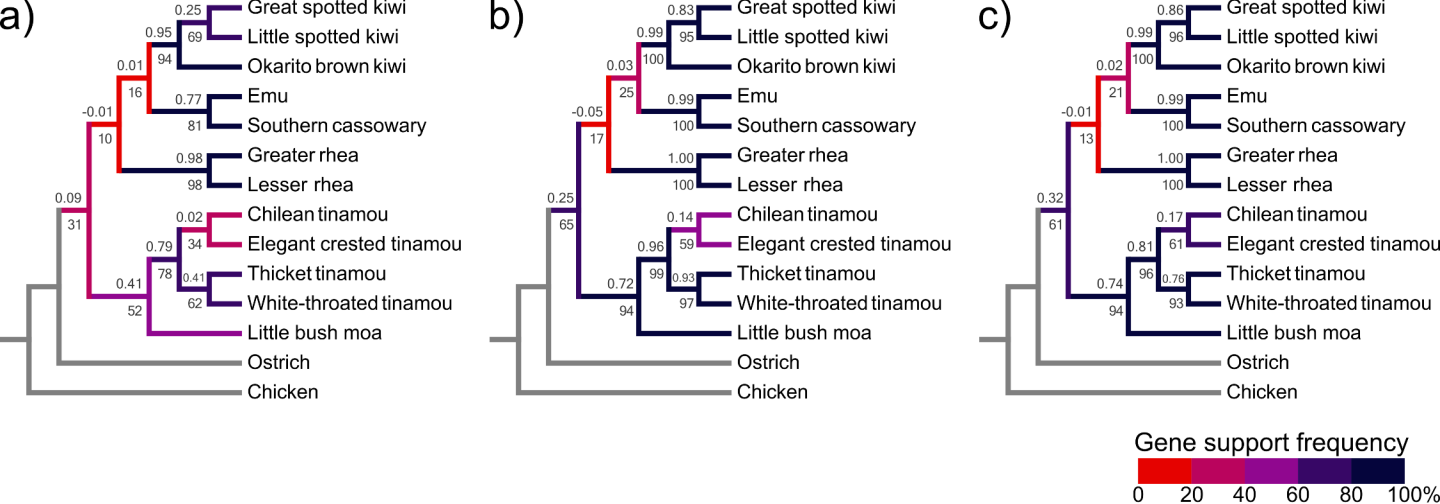
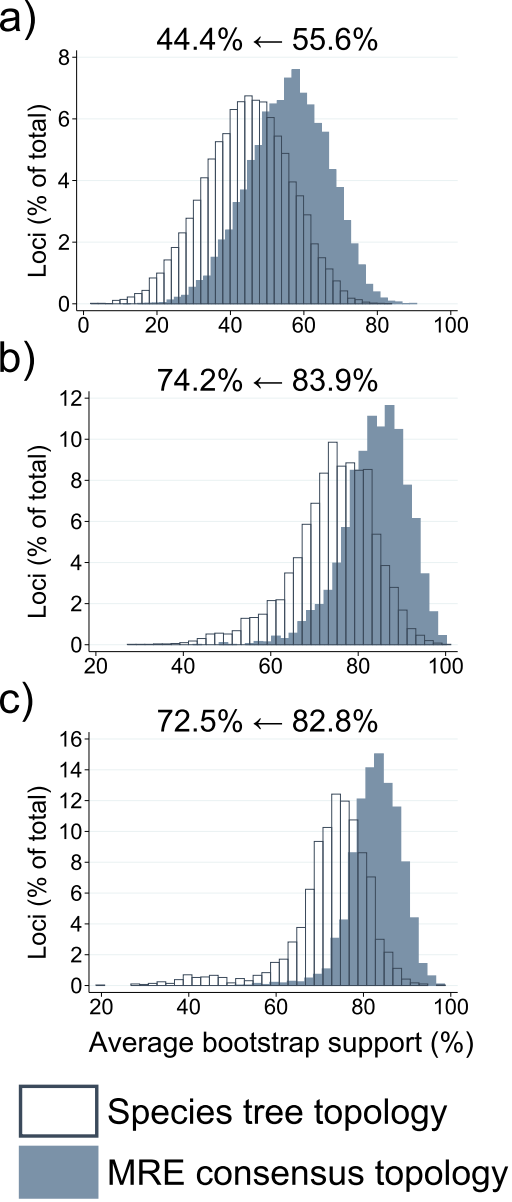

a)
Great spotted kiwi
Little spotted kiwi
Okarito brown kiwi
Emu
Southern cassowary
Greater rhea
Lesser rhea
Chilean tinamou
Elegant crested tinamou
Thicket tinamou
White-throated tinamou
Little bush moa
Ostrich
Chicken
b)
Great spotted kiwi
Little spotted kiwi
Okarito brown kiwi
Emu
Southern cassowary
Greater rhea
Lesser rhea
Chilean tinamou
Elegant crested tinamou
Thicket tinamou
White-throated tinamou
Little bush moa
Ostrich
Chicken
c)
Great spotted kiwi
Little spotted kiwi
Okarito brown kiwi
Emu
Southern cassowary
Greater rhea
Lesser rhea
Chilean tinamou
Elegant crested tinamou
Thicket tinamou
White-throated tinamou
Little bush moa
Ostrich
Chicken
Gene support frequency
0
20
40
60
80
100%

Fig. 6


a)
44.4% ← 55.6%
Loci (% of total)
b)
74.2% ← 83.9%
Loci (% of total)
c)
72.5% ← 82.8%
Loci (% of total)
Average bootstrap support (%)
Species tree topology
MRE consensus topology

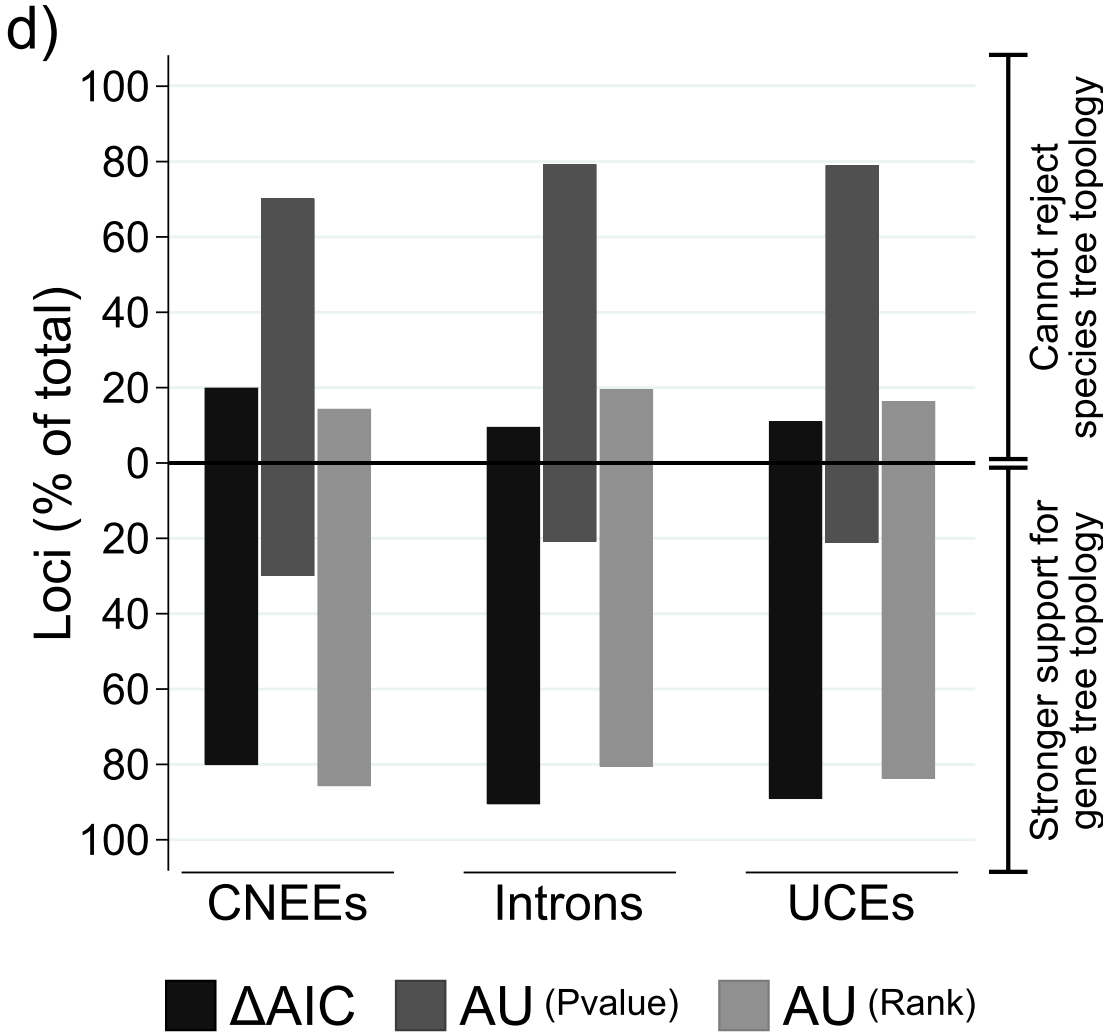

d)
Loci (% of total)
Cannot reject species tree topology
Stronger support for gene tree topology
CNEEs
Introns
UCEs
ΔAIC
AU (Pvalue)
AU (Rank)

Fig. 7

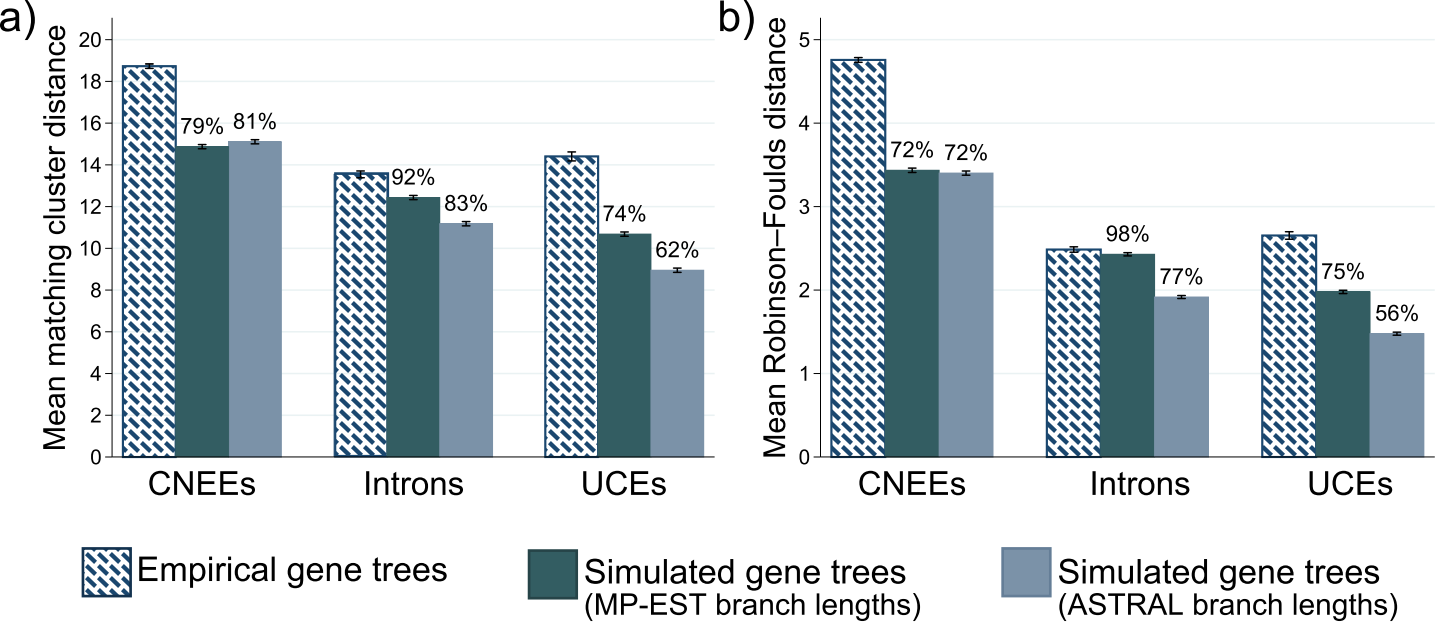

a)
Mean matching cluster distance
0
2
4
6
8
10
12
14
16
18
20
79%
81%
92%
83%
74%
62%
CNEEs
Introns
UCEs
b)
Mean Robinson–Foulds distance
0
1
2
3
4
5
72%
72%
98%
77%
75%
56%
CNEEs
Introns
UCEs
Empirical gene trees
Simulated gene trees
(MP-EST branch lengths)
Simulated gene trees
(ASTRAL branch lengths)

Fig. 8

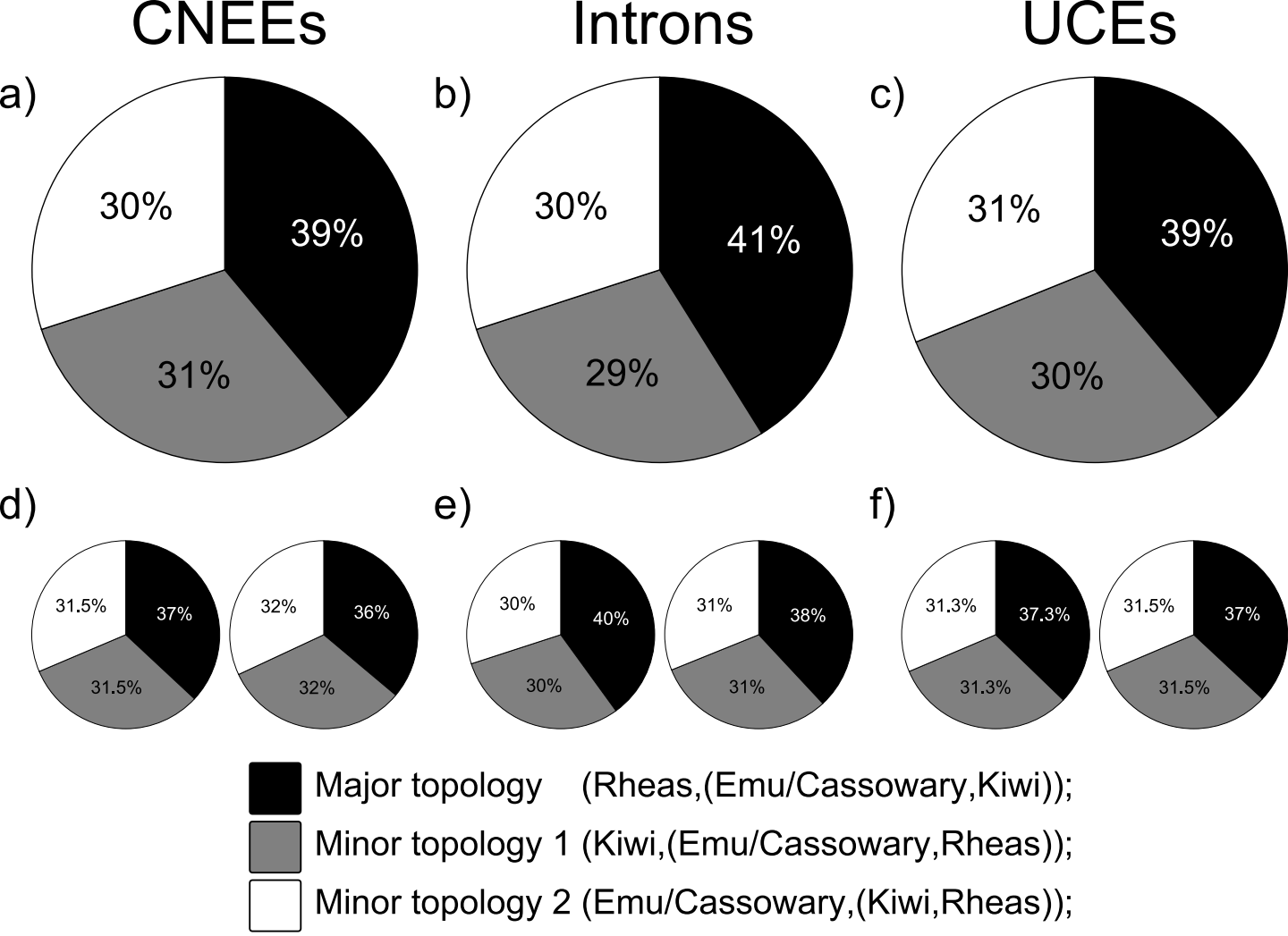
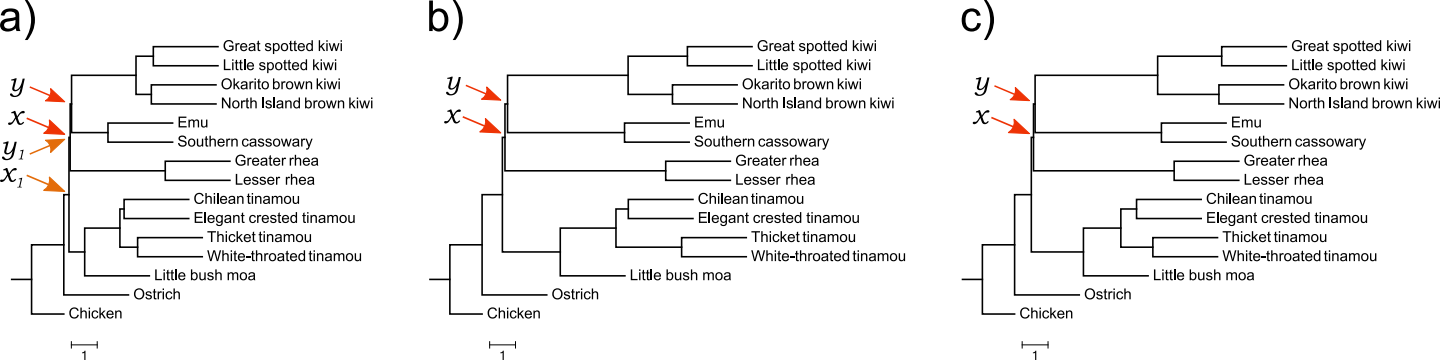

CNEEs
Introns
UCEs
a)
30%
39%
31%
b)
30%
41%
29%
c)
31%
39%
30%
d)
31.5%
37%
31.5%
32%
36%
32%
e)
30%
40%
30%
31%
38%
31%
f)
31.3%
37.3%
31.3%
31.5%
37%
31.5%
Major topology (Rheas,(Emu/Cassowary,Kiwi));
Minor topology 1 (Kiwi,(Emu/Cassowary,Rheas));
Minor topology 2 (Emu/Cassowary,(Kiwi,Rheas));

Fig. 9



d)

Clade 1	Clade 2	CNEEs			Introns			UCEs		
		x	*a(x)*	*y*	*x*	*a(x)*	*y*	*x*	*a(x)*	*y*
All Kiwi	Spotted kiwi	2.490	-0.383	0.667	4.661	-0.403	2.290	4.737	-0.403	2.475
All Kiwi	Brown kiwi	2.490	-0.383	0.592	4.661	-0.403	1.716	4.737	-0.403	2.369
Emu/Cassowary + Kiwi	Emu/Cassowary	0.057	0.595	1.407	0.098	0.335	4.525	0.062	0.550	4.887
Emu/Cassowary + Kiwi	Kiwi	0.057	0.595	2.490	0.098	0.335	4.661	0.062	0.550	4.737
Emu/Cassowary + Kiwi + Rheas	Rheas	0.045	0.729	3.676	0.099	0.330	6.213	0.087	0.387	5.438
Emu/Cassowary + Kiwi + Rheas	Emu/Cassowary + Kiwi	0.045	0.729	0.057	0.099	0.330	0.098	0.087	0.387	0.062
All Tinamous	Thicket & White-throated tinamou	1.359	-0.326	0.682	2.150	-0.373	2.537	1.443	-0.334	1.236
All Tinamous	Chilean & Elegant crested tinamou	1.359	-0.326	0.165	2.150	-0.373	0.479	1.443	-0.334	0.610
Moa + Tinamous	Tinamous	0.608	-0.191	1.359	2.239	-0.376	2.150	2.017	-0.368	1.443
Non-ostrich palaeognaths	Emu/Cassowary + Kiwi + Rheas	0.204	0.074	0.045	0.781	-0.239	0.099	0.643	-0.202	0.087
Non-ostrich palaeognaths	Moa + Tinamous	0.204	0.074	0.608	0.781	-0.239	2.239	0.643	-0.202	2.017

Fig. 10