ZOOLOGICAL RESEARCH

Dynamic evolution of transposable elements, demographic history, and gene content of paleognathous birds

Zong-Ji Wang[1,2,3,4], Guang-Ji Chen[4], Guo-Jie Zhang[4,5,6,7,*], Qi Zhou[2,3,8,*]

[1] *Institute of Animal Sex and Development*, *Zhejiang Wanli University*, *Ningbo*, *Zhejiang* 315100, *China*

[2] *MOE Laboratory of Biosystems Homeostasis* & *Protection*, *Life Sciences Institute*, *Zhejiang University*, *Hangzhou*, *Zhejiang* 310058, *China*

[3] *Department of Neuroscience and Developmental Biology*, *University of Vienna*, *Vienna* 1090, *Austria*

[4] *BGI-Shenzhen*, *Beishan Industrial Zone*, *Shenzhen*, *Guangdong* 518083, *China*

[5] *State Key Laboratory of Genetic Resources and Evolution*, *Kunming Institute of Zoology*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650223, *China*

[6] *Section for Ecology and Evolution*, *Department of Biology*, *University of Copenhagen*, *Copenhagen DK*-2100, *Denmark*

[7] *Center for Excellence in Animal Evolution and Genetics*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650223, *China*

[8] *Center for Reproductive Medicine*, *The Second Affiliated Hospital*, *School of Medicine*, *Zhejiang University*, *Hangzhou*, *Zhejiang* 310058, *China*

ABSTRACT

Palaeognathae includes ratite and tinamou species that are important for understanding early avian evolution. Here, we analyzed the whole-genome sequences of 15 paleognathous species to infer their demographic histories, which are presently unknown. We found that most species showed a reduction of population size since the beginning of the last glacial period, except for those species distributed in Australasia and in the far south of South America. Different degrees of contraction and expansion of transposable elements (TE) have shaped the paleognathous genome architecture, with a higher transposon removal rate in tinamous than in ratites. One repeat family, AviRTE, likely underwent horizontal transfer from tropical parasites to the ancestor of little and undulated tinamous about 30 million years ago. Our analysis of gene families identified rapid turnover of immune and reproduction-related genes but found no evidence of gene family changes underlying the convergent evolution of flightlessness among ratites. We also found that mitochondrial genes have experienced a faster evolutionary rate in tinamous than in ratites, with the former also showing more degenerated W chromosomes. This result can be explained by the

Received: 01 July 2020; Accepted: 15 October 2020; Online: 16 October 2020

Foundation items: This study was supported by the National Natural Science Foundation of China (31671319, 31722050, 32061130208), Natural Science Foundation of Zhejiang Province (LD19C190001), and European Research Council Starting Grant (grant agreement 677696) to Q.Z.; the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31020000, XDB13000000), International Partnership Program of Chinese Academy of Sciences (152453KYSB20170002), Carlsberg Foundation (CF16-0663), and Villum Foundation (25900) to G.J.Z.

*Corresponding authors, E-mail: Guojie.Zhang@bio.ku.dk; zhouqi1982@zju.edu.cn

DOI: 10.24272/j.issn.2095-8137.2020.175

Hill-Robertson interference affecting genetically linked W chromosomes and mitochondria. Overall, we reconstructed the evolutionary history of the Palaeognathae populations, genes, and TEs. Our findings of co-evolution between mitochondria and W chromosomes highlight the key difference in genome evolution between species with ZW sex chromosomes and those with XY sex chromosomes.

Keywords: Paleognaths; Demographic history; Transposable elements; Gene families; Mitochondria

INTRODUCTION

Modern birds include Palaeognathae and Neognathae, which diverged from each other over 110 million years ago (Jarvis et al., 2014), with the former currently comprising only 5.7% of extant bird species. Palaeognathae includes the flightless ratites and volant tinamous. Ratites are broadly distributed across the major continents, and include *Apterygidae* (kiwis) and four families of species with much larger body size than any other bird species: i.e., *Struthionidae* (ostriches), *Rheidae* (rheas), *Casuariidae* (cassowaries), and *Dromaiidae* (emus) (Angst & Buffetaut, 2017). As well as in Africa, ostriches also used to be present in a large part of Eurasia (Europe and Asia) (Houde, 1986). Emus, cassowaries, and kiwis are native to Australasia and rheas are distributed in South America. Tinamous consist of 46 extant species in nine genera, which are exclusively located in Central and South America (Figure 1). Despite the evolutionary success of these birds in terms of species diversity and geographic range, their demographic histories remain largely unknown.

Paleognathous species are important for understanding the early evolution of birds due to their unusual features of gigantism and flightlessness, resulting in a lower genome-wide evolutionary rate than that of other birds (Wang et al., 2019). For example, cytogenetic studies and recent genome analyses of ostriches found fewer intragenomic rearrangements compared to that in other birds (O'Connor et al., 2018; Takagi et al., 1972). The loss of flight capability and the increase in body size (gigantism) seem to have occurred independently multiple times among ratites (Sackton et al., 2019). However, the genomic bases for these changes remain poorly understood, except for several recent studies on emus and rheas (Sackton et al., 2019; Young et al., 2019). Tinamous do not exhibit powered flight, but are capable of short, burst flapping flight, which may be due to their smallest heart-to-body size ratio among birds (Altimiras et al., 2017; Bishop, 1997). It has been proposed that the metabolic requirement for powered flight is linked to a reduced genome size (Altimiras et al., 2017; Kapusta et al., 2017; Wright et al., 2014), as evidenced by the smaller genomes of bats and birds relative to those of other mammals and birds. In addition to the different rates of genomic deletions between the two groups of species, the contribution of transposable element (TE) turnover to the variable genome sizes of paleognaths remains to be elucidated, as previous analysis included only ostrich and white-throated tinamou (Kapusta et al., 2017). Whether gene content diversity among species has contributed to the convergent evolution of flightlessness among ratites also remains to be clarified.

Paleognaths provide a unique model for studying the evolutionary history of bird sex chromosomes. Of note, most paleognathous species have homomorphic sex chromosomes with cytologically indistinguishable Z and W chromosomes (Ogawa et al., 1998; Pigozzi, 1999; Pigozzi & Solari, 1999; Shetty et al., 1999; Tsuda et al., 2007), in contrast to the highly heteromorphic sex chromosomes of Neognathae and mammals. The reason for this unusually slow sex chromosome differentiation is unclear. An important distinction between the ZW and XY sex system is that mitochondria, which are non-recombining and are specifically transmitted through females, are genetically linked to the W chromosome (Berlin et al., 2007). The Hill-Robertson interference, i.e., interference between linked loci reducing the efficacy of selection, is therefore expected to act on both mitochondrial DNA (mtDNA) and W chromosomes (Charlesworth & Charlesworth, 2000). However, few genomic studies have examined whether the extent of sex chromosome differentiation is correlated with the mitochondrial gene evolutionary rate, except for an earlier study on collared flycatchers (Smeds et al., 2015).

In this study, we analyzed the whole-genome sequences of 15 paleognathous female birds to reconstruct the evolutionary history of their populations, TEs, and genes. We characterized dynamic expansions and contractions of TE and gene families in these birds. We further demonstrated that the evolutionary rates of mitochondria and W chromosomes were highly correlated.

MATERIALS AND METHODS

Data description

Assembled sequences (both nuclear and mitochondrial genomes) and gene annotations of 15 species of tinamous and ratites (sample information: Supplementary Table S1) were retrieved from our previous study (Wang et al., 2019). All analyses performed were based on our recently produced phylogenomic tree using whole-genome non-coding sequence alignments (Wang et al., 2019).

Here, we presented the genomic annotations of 12 Palaeognathae species (Supplementary Table S1), together with previously reported genomes (Le Duc et al., 2015; Wang et al., 2019; Zhang et al., 2014), representing all extant Palaeognathae orders (Handford & Mares, 1985). Samples were selected to cover the Tinamiformes order, with nine species within four genera represented for the first time (Figure 1). We produced the single nucleotide polymorphism (SNP) sites for each studied species, which were used to

Figure 1 Continental distributions of paleognathous species investigated in this study

Species range information of studied paleognaths (ratites and tinamous) was retrieved from MAP OF LIFE (https://mol.org/). All bird icons were ordered from https://www.hbw.com/. Points indicate their rough distribution range.

reconstruct their demographic history (see below for detailed methods).

Demographic history analysis

We investigated the demographic histories of the different paleognathous species with whole-genome datasets using the pairwise sequentially Markovian coalescent (PSMC) approach (Li & Durbin, 2011), which can infer changes in effective population size (N_e) within 10 000 to 1 000 000 years. We performed PSMC (Li & Durbin, 2011) with the heterozygous SNP loci produced by the GATK pipeline (DePristo et al., 2011). Parameters were set to "N30 –t5 –r5 –p 4+30*2+4+6+10" and bootstrapping (100 times) was performed for each species to determine variance in N_e estimates. We used the estimated values of generation time and mutation rate to scale the results. We applied branch-specific estimates of the synonymous substitution rate per synonymous site (dS) from our dated phylogeny as proxies for the rate of mutation. The age of sexual maturity (collected from published results) was multiplied by a factor of two as a proxy for generation time (Handford & Mares, 1985) (Supplementary Table S2).

Evolutionary analyses of transposable elements

We used RepeatMasker (v.4.0.7) with RepBase (20160829) and the *de novo* prediction program RepeatModeler (v.open-463 1.0.8) to predict and categorize repetitive elements. We used the Kimura 2-parameter distance to estimate the divergence level between individual TE copies and their consensus sequences. Kimura distances between genome copies and TE consensus were calculated using the RepeatMasker (v.4.0.7) built-in scripts (calcdivergencefromalign.pl). We dated the expansion of certain repeat subfamilies based on their sequence divergence patterns in the phylogeny by parsimony. We inferred more recent or ancestral subfamilies if they showed a sequence divergence pattern concentrated at a lower or higher divergence level and shared among species or specific to certain lineages. To test the influence of assembly quality on the estimation of TE abundances, we performed TE annotation with the same pipeline on a newly produced emu genome using PacBio and Hi-C technologies. We first annotated the genomic copies of AviRTE (a novel family of long interspersed elements) for each Palaeognathae genome assembly using RepeatMasker (v.open-4.0.6) based on its published consensus sequences (Suh et al., 2016). We then used all of the predicted AviRTE sequences from all studied paleognathous genomes to reconstruct the putative ancestral consensus sequence by RepeatModeler (v.open-1.0.8) to calculate the divergence of AviRTE from each species. To date the time of horizontal transfer, we compared the flanking sequences (a block of Ns with the same length) of each AviRTE copy between species by blat (version 36) (Kent, 2002) using those of AviRTEs from *Crypturellus soui* with the highest number of copies as queries. We defined a pair of orthologous AviRTEs if their flanking sequences were aligned to each other.

Evolutionary analyses of gene families

To examine the evolution of gene families in paleognathous birds, genes from ratites, tinamous (15 species in total), and chicken (Ensembl release-87) were clustered into gene families by Treefam (Li et al., 2006) (min_weight=10, min_density=0.34, and max_size=500). Family expansion or contraction analysis was performed by CAFÉ (De Bie et al., 2006). Gene Ontology (GO) enrichment analysis was carried out based on the algorithm implemented in GOstat (Beißbarth & Speed, 2004). The computed *P*-value was then adjusted for multiple tests by specifying a false discovery rate (<0.05) using the Benjamini-Hochberg method (Benjamini et al., 2001).

Evolutionary and association analyses of mitochondrial genes

The mitochondrial protein-coding genes were identified using MITOS (Bernt et al., 2013) and curated by comparison with known sequences of other published ratites and tinamous from GenBank. For the PAML analyses, we first assigned orthologous relationships among all paleognathous birds and the outgroup (chicken) using the reciprocal best blast hit algorithm and syntenic information. We used PRANK (Löytynoja & Goldman, 2005) to align the orthologous gene sequences, which takes phylogenetic information into account when placing a gap into the alignment. We filtered the PRANK alignments by gblocks (Talavera & Castresana, 2007) and excluded genes with a high proportion of low complexity or repetitive sequences to avoid alignment errors. We estimated the evolutionary rates of all individual mitochondrial datasets for each branch using the Codeml program with the free-ratio model in the PAML package (Yang, 2007). We tested the association between the evolutionary rate of mitochondrial genes and sex chromosome differentiation degree using the phylogenetic generalized least squares (PGLS) model implemented in R package APE (Paradis et al., 2004) and nlme (https://CRAN.R-project.org/package=nlme).

RESULTS

Temporal changes in paleognathous species populations

We found large variation in N_e among species, as indicated by their different mean values over time (Supplementary Table S3), which ranged from 20 000–40 000 in emus, Southern cassowaries, and ornate tinamous to approximately 160 000 –1 800 000 in elegant-crested and white-throated tinamous. Similarly, the minimum N_e estimated over the time span analyzed varied between 5 000–9 000 (emus and Southern cassowaries) to >50 000 (elegant-crested and white-throated tinamous), and the maximum N_e varied between 30 000–60 000 (e.g., emus and Andean tinamous) to 700 000–750 000 (undulated tinamous). Similar to the results reported for neognaths (Nadachowska-Brzyska et al., 2015), most species showed a sharp decline in population size since the beginning of last glacial period (LGP) about 100 000 years ago (Figure 2), except for species in the far south of the Southern Hemisphere and Australasia.

Two species (brown kiwis and hooded tinamous) in our study are classified as vulnerable on the IUCN Red List of Threatened Species (http://www.iucnredlist.org/), and another two species (white-throated tinamous and greater rheas) are classified as near threatened (Supplementary Table S3). Their PSMC patterns exhibited a different trend to their historical population changes (Figure 2). Brown kiwis showed a stable N_e during the LGP. Hooded tinamous had a relatively high N_e (~140 000) but experienced a continuous reduction in N_e 30 000 years ago. The white-throated tinamous, which are classified as near threatened, experienced long-term decrease in N_e from 270 000 to 50 000. Greater rheas showed a rather stable N_e for a long time until 40 000 years ago, when they experienced a drastic reduction in N_e to 50 000. Our results suggest that these vulnerable and near threatened species experienced a population size reduction at the end of the LGP, predating the recent declines impacted by human activities.

Temporal evolution of TEs in paleognaths

With similar sequencing coverage, the assembled genome sizes of the ratites were, on average, 27.72% larger (1.276±0.177 vs. 1.005±0.035 Gb, P=0.0012, Wilcoxon test) than those of the flighted tinamous (Supplementary Table S1). Comparison of emu genomes produced by Illumina and PacBio sequences indicated that the annotated repeat sequences, except for the long terminal repeat (LTR) elements, were similar in size, suggesting that most TEs in the studied Palaeognathae genomes were properly annotated (Supplementary Table S4). Repeat library-based annotations of genome samples showed low variance in genomic repeat content and a relatively constant diversity of repeat types. We found that paleognaths generally had a similar level of genome-wide repeats, ranging from 4.89% in thicket tinamous to 5.54% in ostriches, which was largely attributed to their similar long interspersed nuclear element (LINE) content (Supplementary Table S5). Ratites had significantly (P< 0.0007, Wilcoxon test) more DNA transposons (0.730±0.059% vs. 0.472±0.028%, on average 0.55 fold higher) and short interspersed nuclear elements (SINE, 0.178±0.008% vs. 0.077±0.007%, on average 1.25 fold higher) compared with tinamous. They together contributed an additional 13.98 (66.62±7.987 vs. 52.64±3.308) Mb, on average, to the ratite genomes compared to those of tinamous (Supplementary Table S5). When inspecting the sequence divergence patterns of DNA transposons compared to their consensus sequences, we did not find any recent expansions in ratites, whose divergence level was expected to be low (Figure 3A; Supplementary Figure S1). Thus, the higher DNA transposon content of ratites is more likely to be caused by the higher rate of DNA transposon removal in tinamous.

In contrast, the predominant repeat type (about 3% of the genome), i.e., chicken repeat 1 (CR1) LINE elements, showed a similar genomic percentage among paleognaths (Supplementary Table S5). Different CR1 repeat subfamilies exhibited different patterns of genomic composition and

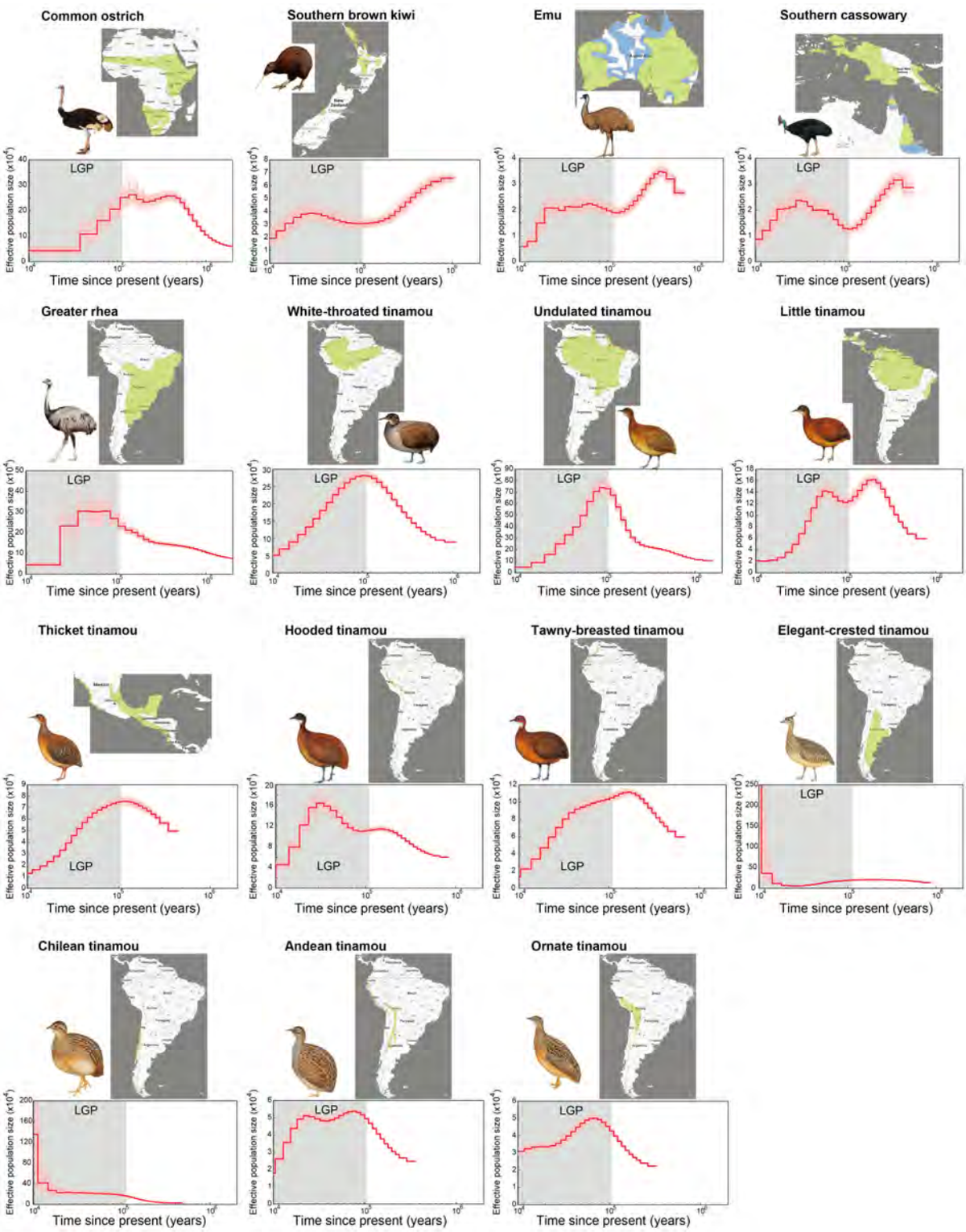


Figure 2 Dynamic changes in effective population size

Red curve of each species is the population size dynamics inferred from PSMC analyses, with pink curves indicating variation in population size derived from 100 bootstraps. Gray shaded areas indicate last glacial period (LGP). We also aligned species range information (https://mol.org/) to each panel.

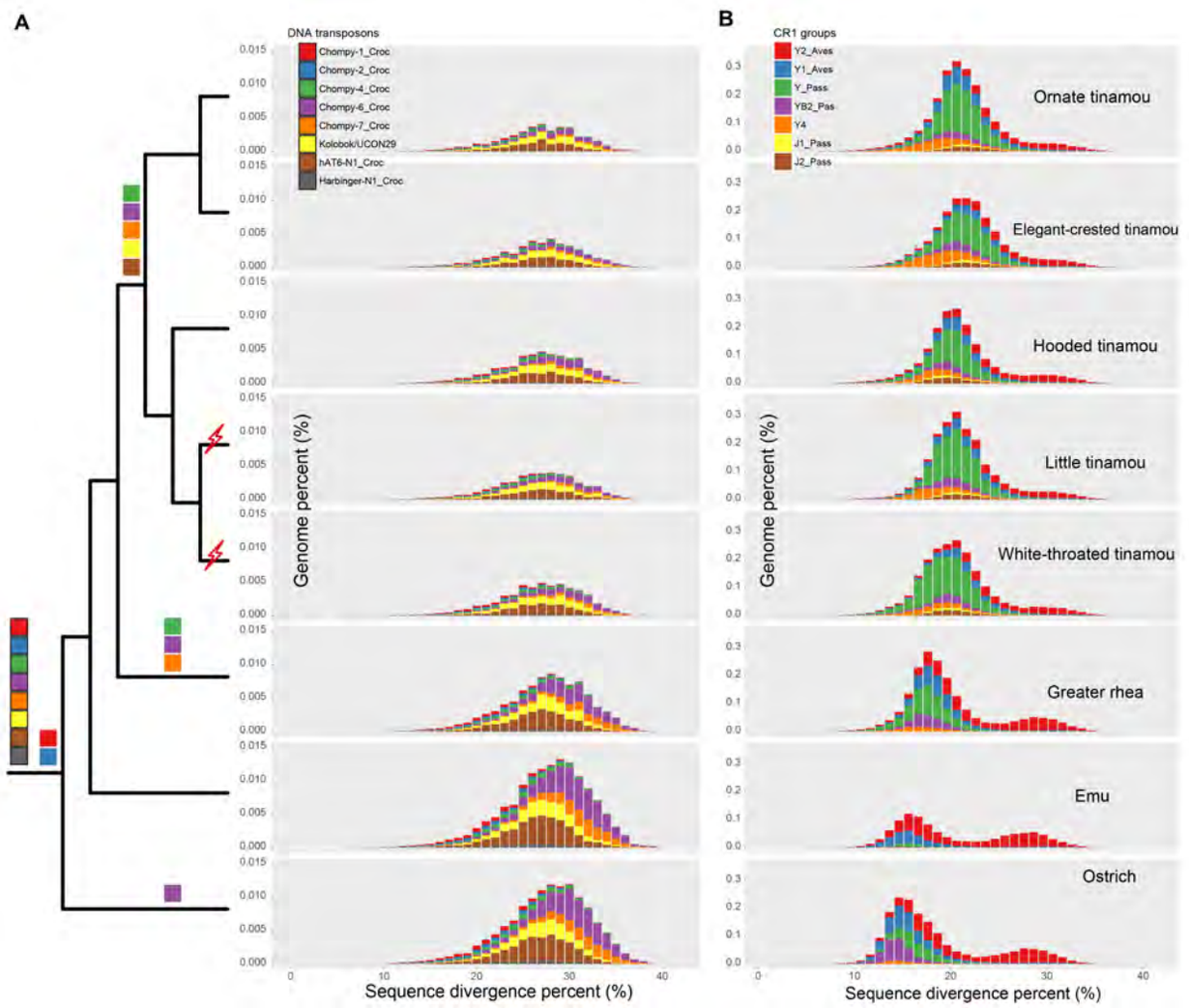


Figure 3 Temporal evolution of transposable elements of paleognaths

Patterns of eight out of 15 studied Palaeognathae species are shown here. The remaining species, whose patterns were very similar to the eight species, are included in Supplementary Figures S1, S3. A, B**:** Inferred bursts of certain subfamilies of DNA transposons (in black-framed color squares) and CR1 LINEs (in squares without black frames) are labelled at corresponding phylogenetic nodes. Histograms show distributions of sequence divergence between each subfamily vs. their consensus sequences. We dated the expansion of certain repeat subfamilies based on their sequence divergence patterns in phylogeny by parsimony. We also dated two horizontal transfers of AviRTE retroposon, one of which has been reported previously (Suh et al., 2016), indicated here in red.

temporal activity between the ratites and tinamous. Certain subfamilies, for example, CR1-Y1_Aves and -Y2_Aves, are thought to have been active throughout the evolutionary history of paleognathous species, and probably throughout that of all birds, with previous studies showing activity in both chickens and zebra finches (Suh et al., 2011). This accords with our chronological analyses of nested TEs, with the expectation that younger and active TEs are more likely to be nested within older and inactive TEs (‘transposition in transposition’, TinT (Churakov et al., 2010)) than the opposite scenario. Based on TinT analyses, subfamilies CR1-Y_Pass, -Y4, and -YB2_Pass showed lineage-specific activity throughout the tinamou lineage as well as in some ratites (e.g., CR1-YB2_Pass in ostriches) (Supplementary Figure S2). This result was corroborated by their sequence divergence patterns compared to their consensus sequences (Figure 3B, Supplementary Figure S3): i.e., recently active CR1 subfamilies (e.g., CR1-Y4) were specifically enriched in tinamou species and showed a lower sequence divergence pattern than those (e.g., CR1-Y2_Aves) active at an earlier time point. Overall, the lower contraction rate of ancient CR1 elements in ratites and more recent expansion of certain CR1 subfamilies in tinamous resulted in the similar level of LINE elements between the two groups of species.

It is worth noting that we also dated two independent horizontal transfers of retroposon AviRTE from filarial nematodes to tinamou species, which is absent in any other paleognaths. One was previously reported in white-throated tinamous (Suh et al., 2016). Based on comparison of the orthologous flanking sequences of AviRTE among species, we inferred another horizontal transfer in the ancestor of little and undulated tinamous within 30 million years (Supplementary Figure S4 and Table S6).

Evolution of gene families

Based on the comparison of orthologous gene families among the 15 paleognathous species, with chickens as the outgroup, we characterized the expanded and contracted gene families across all phylogenetic branches (Figure 4A). We found that genes enriched in the GO terms 'olfactory receptor activity', 'G-protein coupled receptor signaling pathway', or 'spermatogenesis' frequently exhibited expansion and contraction for both internal and external branches in the Palaeognathae species tree (Dataset 1, 2). These immune- or reproduction-related pathways also show rapid turnover in gene families in other species (Demuth et al., 2006; Hahn et al., 2007; Sackton et al., 2007) (Figure 4B, Dataset 1). Interestingly, the ancestor of tinamous showed significant expansions of gene families enriched in GO terms 'oxidoreductase activity' and 'developmental maturation' (Figure 4C), which may be related to their small body size, and thus different metabolic rate and developmental speed compared with ratites. Some gene families also showed expansions and were related to sexual development within Tinamiformes: i.e., 'male genitalia development', 'testosterone biosynthetic process', and 'steroid biosynthetic process' (Dataset 2). Taken together, these results suggest that different species' adaptation processes to different environments and different degrees of sexual selection may have driven changes in gene family size. Interestingly, gene families showing significant contractions in ratites were not enriched in GO terms involved in limb development. This suggests that gene loss is probably not a major force underlying the loss of flight in ratites.

Rapid evolution of mtDNA sequences in birds with highly degenerated W chromosomes

The sex chromosomes of Palaeognathae species were generally less differentiated than those of Neognathae species, with ratite sex chromosomes being even less differentiated than those of tinamous. This may be related to the unique male-only parental care exhibited in paleognathous species, and the low degree of sexual selection targeting males (Wang et al., 2019). To test whether mtDNA genes showed an associated sequence evolutionary pattern with the differentiation degree of sex chromosomes, we calculated the branch-specific cumulative evolutionary rates based on the nonsynonymous to synonymous substitution rate ratio (dN/dS) of 13 mitochondrial genes across the 15 paleognathous birds. We also measured the differentiation degree of sex chromosomes for each species by the length ratio of the Z-linked region that suppressed recombination with the homologous W-linked region (sexually differentiated region, SDR) to the entire Z chromosome. Results demonstrated a significant correlation between the evolutionary rate of mitochondrial genes and sex chromosome differentiation degree (PGLS test, P=0.0079, Figure 5A). That is, tinamous with a more differentiated sex chromosome pair (a higher SDR/chrZ ratio) tended to fix an excess of slightly deleterious mutations (Supplementary Figure S5) in their mitochondrial genes due to Hill-Robertson interference between the W chromosome and mitochondrial genome. We also found that tinamou mitochondrial genes had a significantly higher (P<2.2e–16, Wilcoxon rank sum test) (Figure 5B) synonymous substitution rate, which evolved largely under neutrality, than that of the ratites. This is consistent with the generally smaller body size and higher basal metabolic rate of tinamous (Figure 5C) compared to ratites. The higher metabolic rate associated with a smaller body size probably has a higher mutagenic effect on mtDNA and nuclear genomes, thus producing a higher mutation rate, as reflected by the synonymous substitution rate.

DISCUSSION

Tinamous are widely distributed throughout the South American continent and occupy different ecological niches, which probably diversified their demographic histories (Figure 2). Here, the trajectories of temporal N_e changes were generally different between the ratites and tinamous (Figure 2; Supplementary Table S3). Most species did not show a strong phylogenetic signal regarding N_e changes, which may reflect recent population size changes impacted by their separate ecological niches. The closely related Australasian species, i.e., emus, cassowaries, and kiwis, were the exceptions, demonstrating very similar PSMC patterns despite having different extant geographic distributions (Australia, New Guinea, and New Zealand, respectively). This suggests that their living area used to be largely overlapping. In contrast, the Chilean and elegant-crested tinamous are distantly related but showed very similar population expansions compared to other paleognaths. This may be due to their shared distribution in high-altitude shrubland in southern South America, which may have avoided serious impact from by the LGP. Similarly, white-throated and undulated tinamous shared very similar PSMC patterns and geographic ranges.

Our comparative analyses of the 15 paleognathous nuclear and mitochondrial genomes highlighted that the different sequence evolutionary patterns between ratites and tinamous are likely associated with their different body mass and flight capabilities.

Differences in TE content in paleognathous species can be impacted by their different effective population sizes (Lynch, 2007). Previous studies have demonstrated that recent bottlenecks in endangered mammals can fix excessive TEs in the genome by genetic drift (Abascal et al., 2016; Li & Durbin, 2011). Here, population size had a significant association

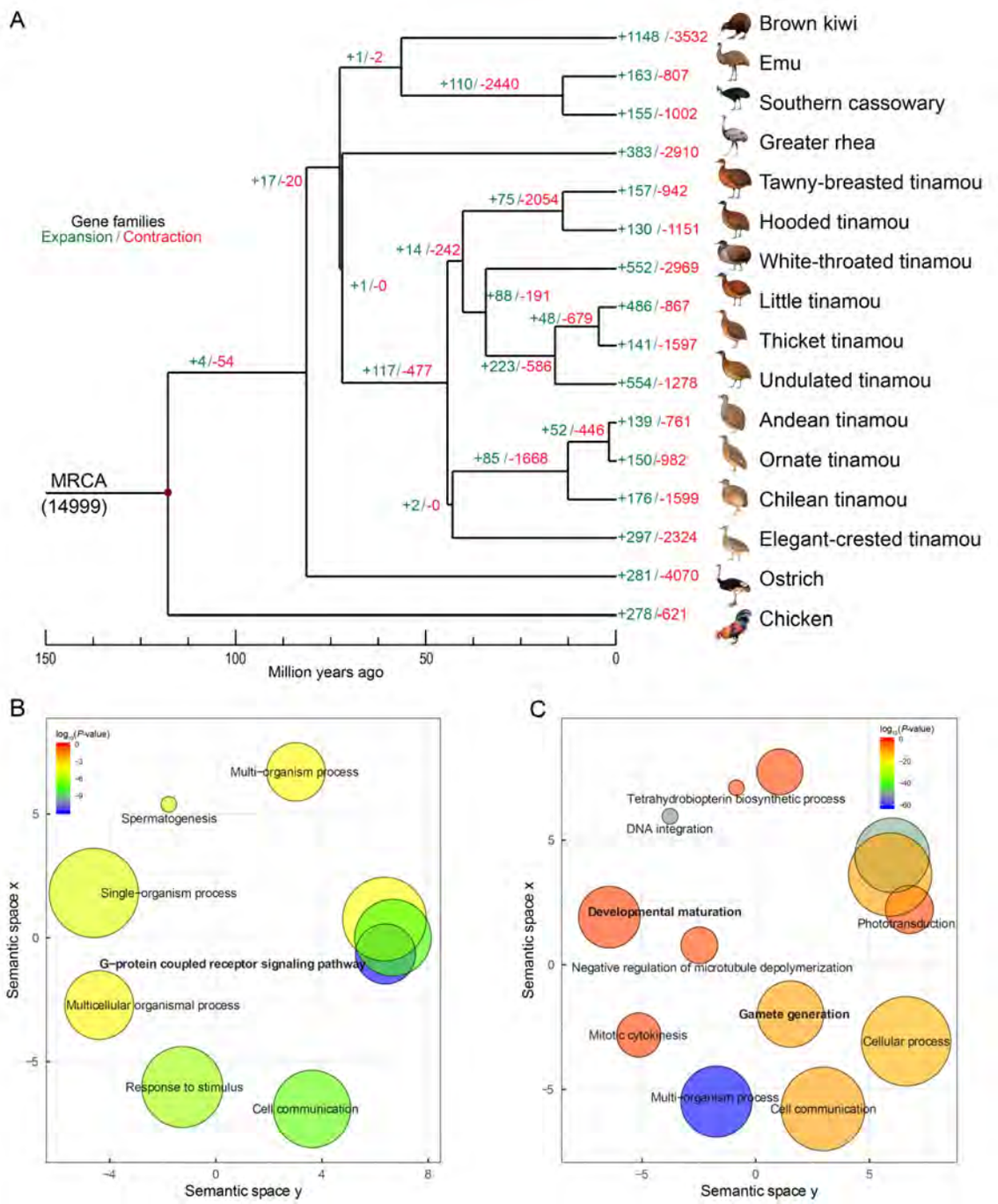


Figure 4 Gene family evolution across Palaeognathae tree

A: Numbers designate number of gene families that have expanded (green) or contracted (red) since split from common ancestor. Most recent common ancestor (MCRA) has 14 999 gene families. Phylogenetic tree based on genome-wide alignments of non-coding sequences, adapted from Wang et al. (2019). B, C: GO-term enrichment analysis of contracted gene families in ostriches (B) and expanded gene families in ancestor of tinamous (C). Bubble color indicates log10 (*P*-value) (legend in upper left-hand corner). Size of bubble indicates frequency of the GO term in the underlying GOA database (larger ones denote more general terms). Scatterplots were drawn by REVIGO (http://revigo.irb.hr/).

(*P*<0.05, PGLS test) with overall and certain types of TE content in the studied paleognathous species (Supplementary Table S7), with some exceptions. For example, the dramatic recent expansion of population size in elegant-crested

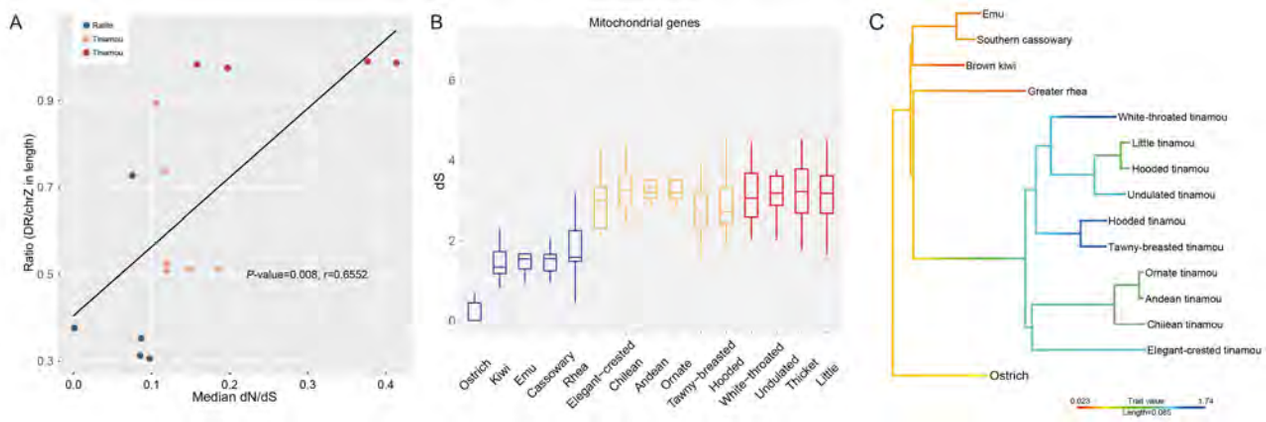


Figure 5 Correlated evolution of mitochondrial genes and W chromosomes among paleognathous species

A**:** Correlation between evolutionary rates of mitochondrial genes and length ratios of sexually differentiated regions (SDR) over entire chrZ. Each data point represents one paleognathous species. Pearson correlation coefficients (r) are shown for comparison. B**:** Synonymous substitution rates (dS) of mitochondrial genes (13 genes). Colors refer to ratites (blue), tinamous with moderately degenerated chrW (yellow), and tinamous with highly degenerated chrW (red). C**:** Branch-specific basal metabolic rates (BMR). Branches are colored according to BMR values obtained through ancestral character estimation using *phytools* (http://github.com/liamrevell/phytools) package as well as dependent package *ape*. Red indicates low BMR value and blue indicates high BMR values.

tinamous and contraction of population size in emus have not resulted in dramatically different genome-wide TE content between the two species (5.02% vs. 4.98%). However, the recent population size change may impact the frequencies of TE families within the population, as reported in *Drosophila* (Barrón et al., 2014). All the studied ratites had slightly larger TE content and genome size (Supplementary Tables S1, S5) than that of the studied tinamous. Although species with a smaller estimated N_e (e.g., emus and cassowaries) may be expected to fix an excess of TEs relative to those with a larger N_e (e.g., ostriches) due to less effective natural selection purging TE insertions (Lynch, 2007), all ratites had very similar TE content. Their higher TE content compared with tinamous is consistent with recent comparison between ostriches and white-throated tinamous (Kapusta et al., 2017), which suggested that ostriches, similar to other flightless bird species like penguins, have undergone fewer sequence deletions than flighted birds since their divergence from the common ancestor of birds. Consistent with this pattern, we found that the differences in TE content and genome size between ratites and tinamous were partially attributed to the more severe loss of DNA transposons in tinamous (Figure 3A, Supplementary Figure S1).

Associated with independent loss of flight, gigantism has also evolved multiple times among ratites. We previously showed that ratites have a lower nuclear genome-wide substitution rate than tinamous, which is associated with their body size (Wang et al., 2019). Here, we found a similar pattern for synonymous substitution rates of mitochondrial genes (Figure 5B). This may explain the less differentiated pattern of sex chromosomes in ratites compared with that in tinamous, i.e., lower rate of genomic rearrangements would suppress homologous recombination between the sex chromosomes. Indeed, a recent comparative analysis of ostrich, falcon, and budgerigar genomes found that ostriches show the lowest genomic rearrangements when all three species are compared to chickens (O'Connor et al., 2018). Among tinamou species with a similar body size, their different estimated N_e values do not seem to predict the degree of sex chromosome differentiation. For example, in our previous work (Wang et al., 2019), we showed that thicket and little tinamous have very similar patterns and degrees of sex chromosome differentiation, but a nearly 2-fold difference in N_e estimated from this work. Interestingly, we found a significant correlation between the extent of sex chromosome differentiation and the mitochondrial gene evolutionary rate. This reflects the effect of Hill-Robertson interference, which would fix slightly deleterious mutations on mtDNA that are genetically linked to the degenerating W chromosome. Tinamous with more degenerated W chromosomes therefore had a higher rate of mitochondria gene evolution than ratites. This could also explain the lower sequence diversity of mitochondria in birds (Berlin et al., 2007) compared to that in mammals, whose mitochondria are not influenced by the paternally inherited Y chromosome. One might expect that strong natural selection on mitochondrial genes due to high metabolic rates could, in turn, retard the degeneration of W chromosomes. However, this does not seem to be the case in hummingbirds: we previously found hummingbirds to have a highly degenerated W chromosome (Zhou et al., 2014), probably because their high metabolic rates evolved after the loss of functional genes on the W chromosome.

CONCLUSIONS

In summary, we delineated the temporal evolution of historical population size, transposable elements, and gene families of

paleognathous species in the context of their phylogeny. Their varying degrees of sex chromosome differentiation also allowed us to demonstrate the correlated evolution of maternally-inherited mitochondria and W chromosomes across species, which is expected to only exist in species with female heterogametic sex chromosomes, such as birds and butterflies.

DATA AVAILABILITY

All genomic reads, genome assemblies, and annotations used in this study have been deposited in CNGBdb (https://db.cngb.org/cnsa/) with accession No. CNP0000505.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Z.J.W. and G.J.C. conducted bioinformatics analyses. Z.J.W., Q.Z., and G.J.Z. wrote and revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank China National Genebank at BGI for contributing to the sequencing. We would like to thank Christopher C. Witt, Mariel L. Campbell and Ariel M. Gaffney from the Museum of Southwestern Biology, Gary Graves from Smithsonian Institute, Robb T. Brumfield and Donna L. Dittman from Louisiana State University Museum of Natural Science, Jack Withrow and Andy Kratter from Florida Museum of Natural History, University of New Mexico for providing bird DNA samples for this work.

REFERENCES

Abascal F, Corvelo A, Cruz F, Villanueva-Cañas JL, Vlasova A, Marcet-Houben M, et al. 2016. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered iberian lynx. *Genome Biology*, **17**: 251.

Altimiras J, Lindgren I, Giraldo-Deck LM, Matthei A, Garitano-Zavala Á. 2017. Aerobic performance in tinamous is limited by their small heart. *A novel hypothesis in the evolution of avian flight. Scientific Reports*, **7**(1): 15964.

Angst D, Buffetaut E. 2017. Paleobiology of Giant Flightless Birds. Oxford: Elsevier.

Barrón MG, Fiston-Lavier AS, Petrov DA, González J. 2014. Population genomics of transposable elements in Drosophila. *Annual Review of Genetics*, **48**: 561−581.

Beißbarth T, Speed TP. 2004. GOstat: find statistically overrepresented gene ontologies within a group of genes. *Bioinformatics*, **20**(9): 1464−1465.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. 2001. Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, **125**(1-2): 279−284.

Berlin S, Tomaras D, Charlesworth B. 2007. Low mitochondrial variability in birds may indicate hill–robertson effects on the w chromosome. *Heredity*, **99**(4): 389−396.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, **69**(2): 313−319.

Bishop CM. 1997. Heart mass and the maximum cardiac output of birds and mammals: Implications for estimating the maximum aerobic power input of flying animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **352**(1352): 447−456.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **355**(1403): 1563−1572.

Churakov G, Grundmann N, Kuritzin A, Brosius J, Makałowski W, Schmitz J. 2010. A novel web-based tint application and the chronology of the primate *alu* retroposon activity. *BMC Evolutionary Biology*, **10**: 376.

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, **22**(10): 1269−1271.

Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. *PLoS One*, **1**(1): e85.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**(5): 491−498.

Hahn MW, Han MV, Han SG. 2007. Gene family evolution across 12 *Drosophila* genomes. *PLoS Genetics*, **3**(11): e197.

Handford P, Mares MA. 1985. The mating systems of ratites and tinamous: an evolutionary perspective. *Biological Journal of the Linnean Society*, **25**(1): 77−104.

Houde P. 1986. Ostrich ancestors found in the northern hemisphere suggest new hypothesis of ratite origins. *Nature*, **324**(6097): 563−565.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**(6215): 1320−1331.

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(8): E1460−E1469.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Research*, **12**(4): 656−664.

Le Duc D, Renaud G, Krishnan A, Almén MS, Huynen L, Prohaska SJ, et al. 2015. Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biology*, **16**(1): 147.

Li H, Coghlan A, Ruan J, Coin LJ, Hériché JK, Osmotherly L, et al. 2006. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Research*, **34**(S1): D572−D580.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*, **475**(7357): 493−496.

Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(30): 10557−10562.

Lynch M. 2007. The Origins of Genome Architecture. Sunderland, MA:

Sinauer Associates.

Nadachowska-Brzyska K, Li C, Smeds L, Zhang GJ, Ellegren H. 2015. Temporal dynamics of avian populations during pleistocene revealed by whole-genome sequences. *Current Biology*, **25**(10): 1375−1380.

O'Connor RE, Farré M, Joseph S, Damas J, Kiazim L, Jennings R, et al. 2018. Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. *Genome Biology*, **19**(1): 171.

Ogawa A, Murata K, Mizuno S. 1998. The location of Z-and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(8): 4415−4418.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in r language. *Bioinformatics*, **20**(2): 289−290.

Pigozzi MI. 1999. Origin and evolution of the sex chromosomes in birds. *Biocell*, **23**(2): 79−95.

Pigozzi MI, Solari AJ. 1999. The ZW pairs of two paleognath birds from two orders show transitional stages of sex chromosome differentiation. *Chromosome Research*, **7**(7): 541−551.

Sackton TB, Grayson P, Cloutier A, Hu ZR, Liu JS, Wheeler NE, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science*, **364**(6435): 74−78.

Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007. Dynamic evolution of the innate immune system in Drosophila. *Nature Genetics*, **39**(12): 1461−1468.

Shetty S, Griffin DK, Graves JAM. 1999. Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Research*, **7**(4): 289−295.

Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, et al. 2015. Evolutionary analysis of the female-specific avian w chromosome. *Nature Communications*, **6**: 7330.

Suh A, Paus M, Kiefmann M, Churakov G, Franke FA, Brosius J, et al. 2011. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nature Communications*, **2**: 443.

Suh A, Witt CC, Menger J, Sadanandan KR, Podsiadlowski L, Gerth M, et al. 2016. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. *Nature Communications*, **7**: 11396.

Takagi N, Itoh M, Sasaki M. 1972. Chromosome studies in four species of *Ratitae* (*Aves*). *Chromosoma*, **36**(3): 281−291.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, **56**(4): 564−577.

Tsuda Y, Nishida-Umehara C, Ishijima J, Yamada K, Matsuda Y. 2007. Comparison of the z and w sex chromosomal architectures in elegant crested tinamou (*Eudromia elegans*) and ostrich (*Struthio camelus*) and the process of sex chromosome differentiation in palaeognathous birds. *Chromosoma*, **116**(2): 159−173.

Wang ZJ, Zhang JL, Xu XM, Witt C, Deng Y, Chen GJ, et al. 2019. Phylogeny, transposable element and sex chromosome evolution of the basal lineage of birds. *bioRxiv*, doi: 10.1101/750109.

Wright NA, Gregory TR, Witt CC. 2014. Metabolic 'engines' of flight drive genome size reduction in birds. *Proceedings of the Royal Society B: Biological Sciences*, **281**(1779): 20132780.

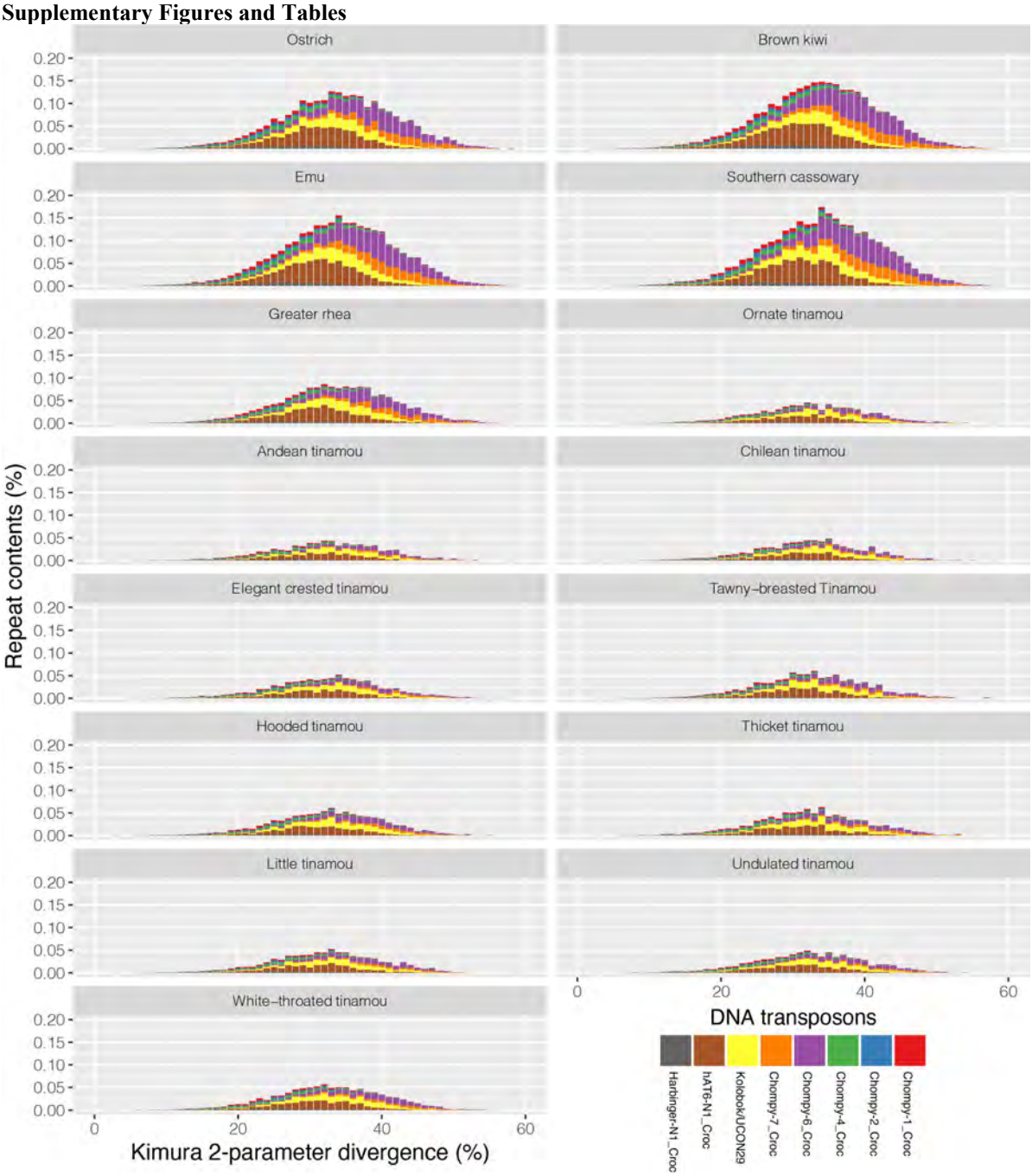
Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586−1591.

Young JJ, Grayson P, Edwards SV, Tabin CJ. 2019. Attenuated fgf signaling underlies the forelimb heterochrony in the emu *Dromaius novaehollandiae*. *Current Biology*, **29**(21): 3681−3691.e5.

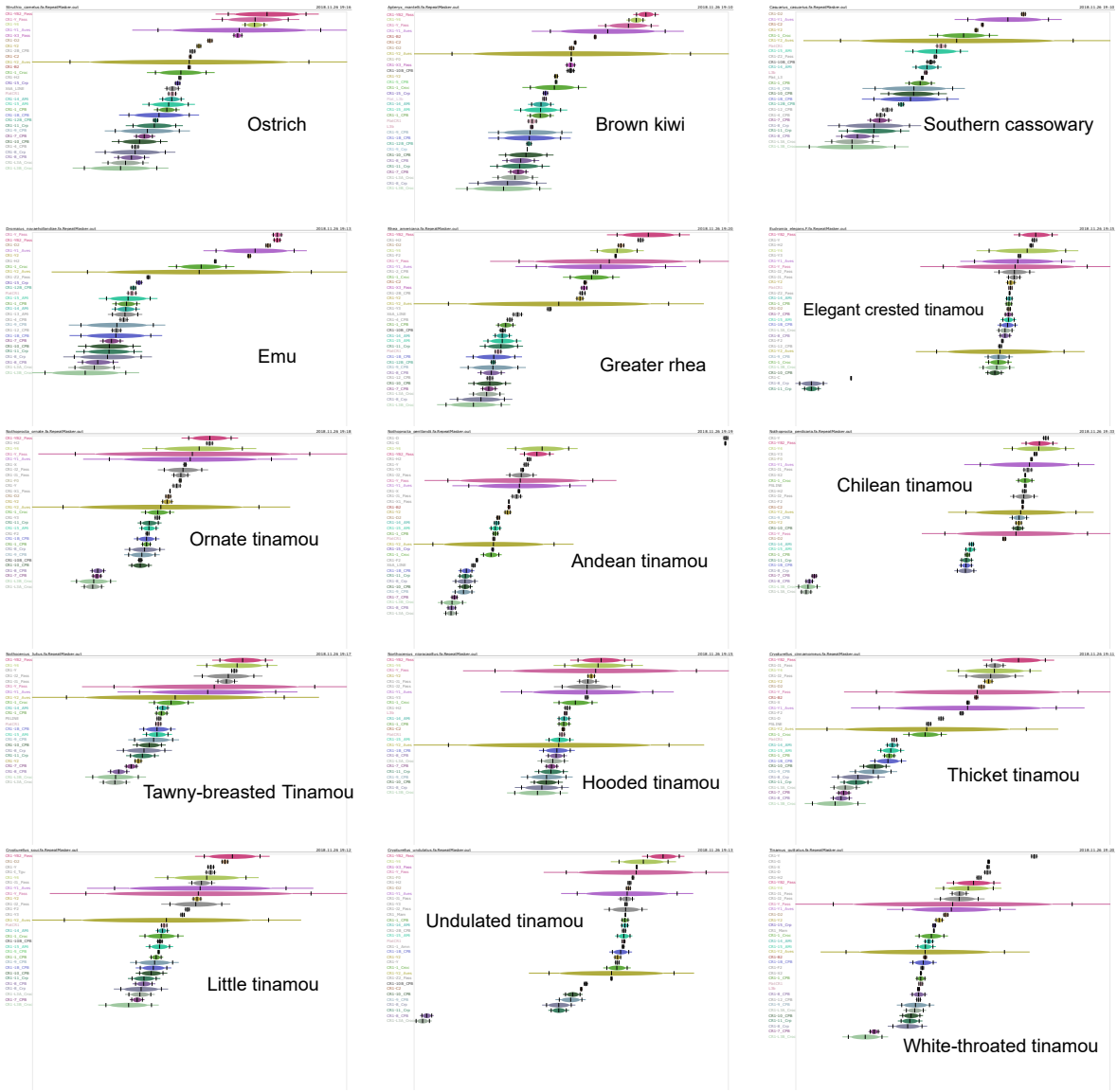
Zhang GJ, Li C, Li QY, Li B, Larkin DM, Lee C, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, **346**(6215): 1311−1320.

Zhou Q, Zhang JL, Bachtrog D, An N, Huang QF, Jarvis ED, et al. 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*, **346**(6215): 1246338.

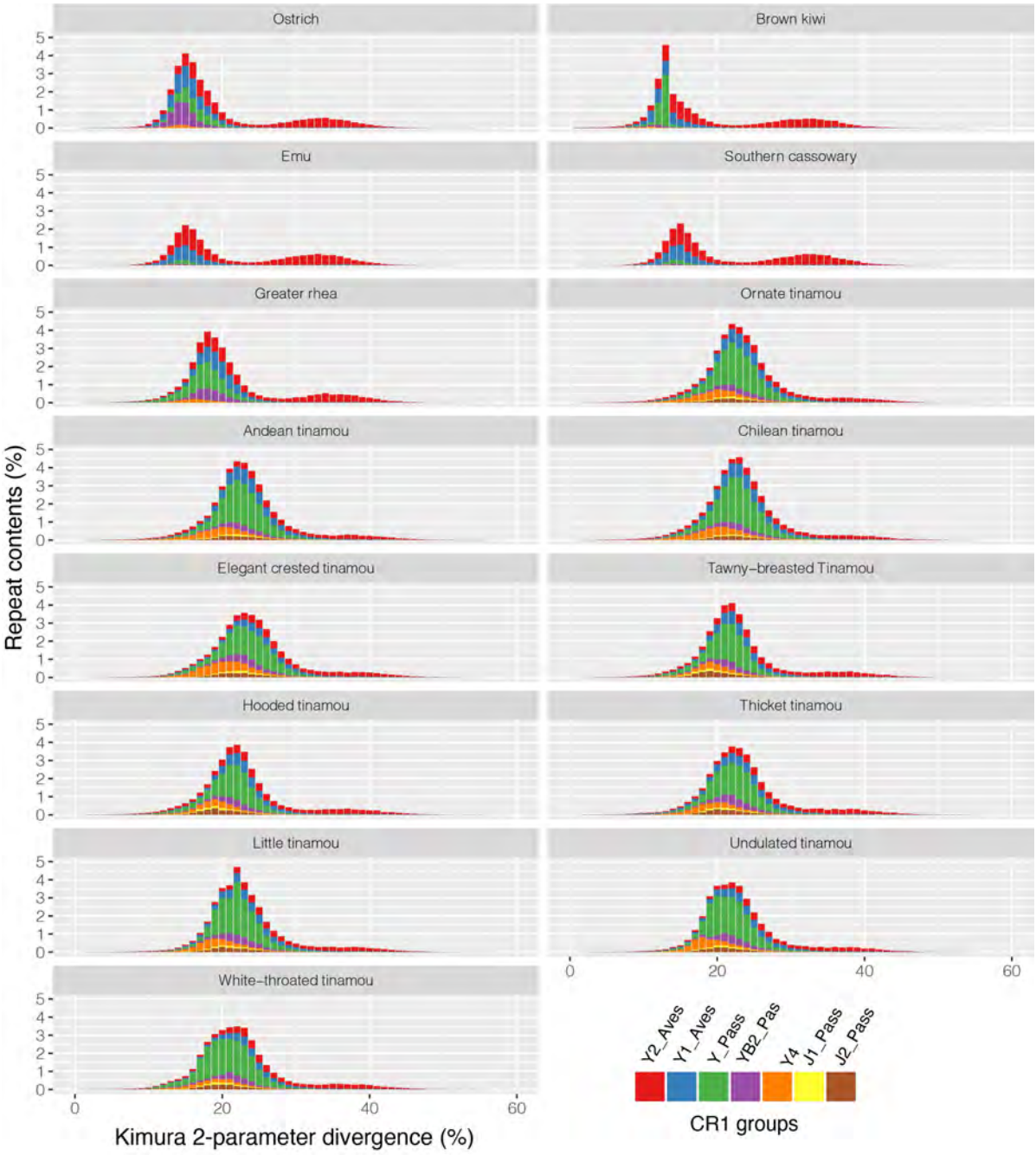
Supplementary Figures and Tables



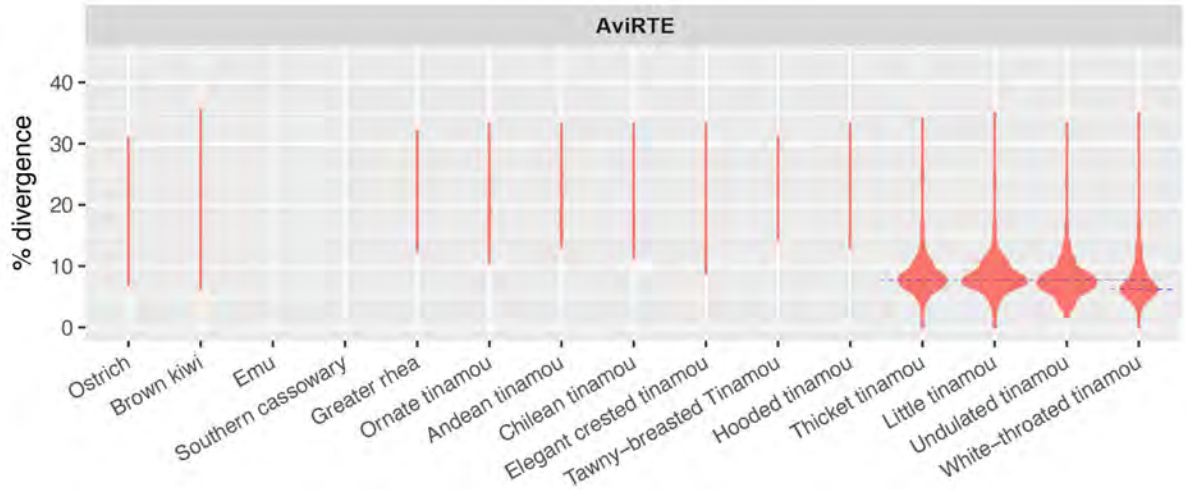
Supplementary Figure S1. Distribution of sequence divergence patterns of paleognathous DNA transposons compared to their consensus sequences.
We showed divergence pattern for each subfamily of DNA transposons here, with the fraction of the genome of each family shown on the y-axis, and Kimura 2-Parameter sequence divergence between individual TE copies and consensus references on the x-axis.



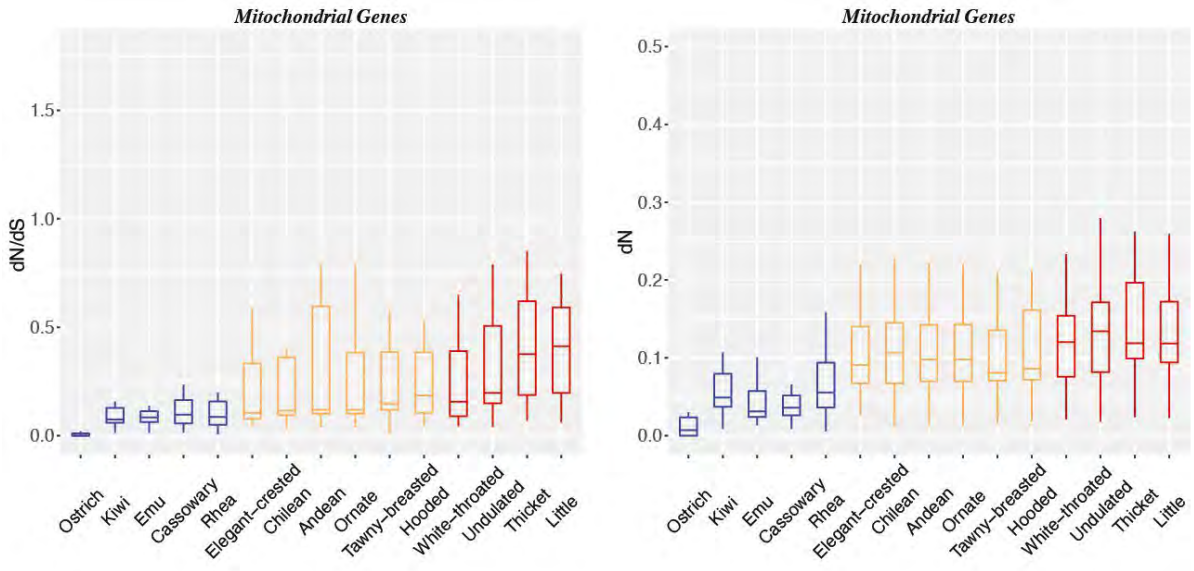
Supplementary Figure S2. TinT analyses of different CR1 subfamilies in Palaeognathae. Each oval represents one CR1 subfamily, sorted by peak time of activity (line in the middle of the oval), from top to bottom, from most recent to most ancient time period. Each oval contains 75% (at oval end), 95% (vertical lines), and 99% (ends of each line) of the probable activity period range



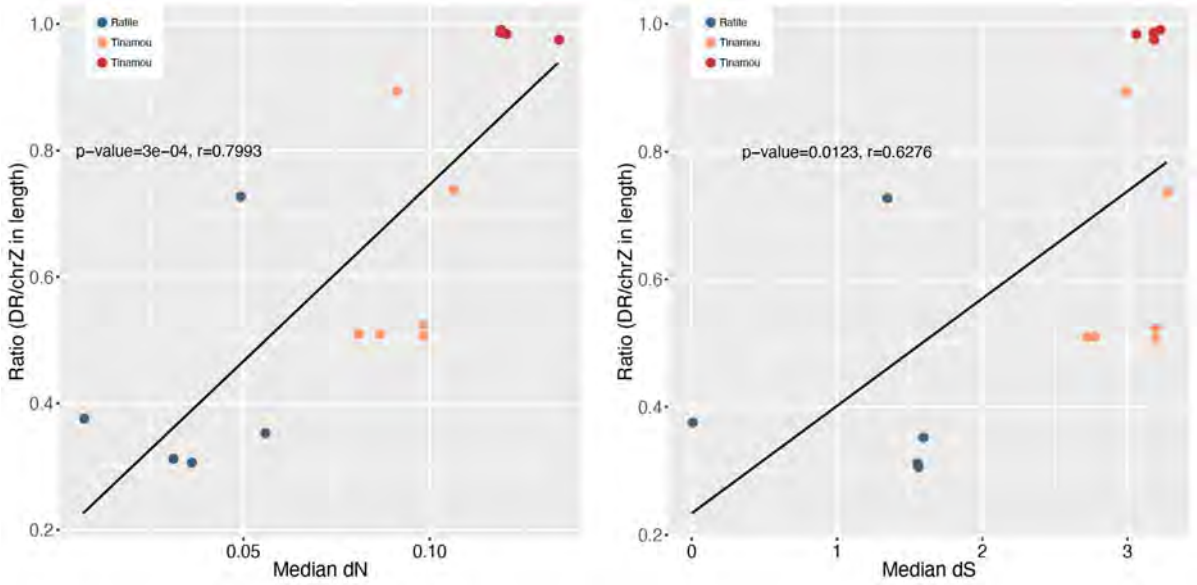
Supplementary Figure S3. Distribution of sequence divergence patterns of paleognathous CR1 LINEs compared to their consensus sequences.
We showed divergence pattern for each subfamily of CR1 LINE elements here, with the fraction of the genome of each family shown on the y-axis, and Kimura 2-Parameter sequence divergence between individual TE copies and consensus references on the x-axis.



Supplementary Figure S4. Violin plot showing AviRTEs' frequency distribution of sequence divergence level from inferred ancestral consensus sequences. Clustering of TEs with similar divergence levels, manifested as 'bout' of the violin, corresponds to the burst of AviRTE amplification. Blue dashed lines indicate position of burst.



Supplementary Figure S5. Evolutionary patterns of mitochondrial (MT) genes among different groups.
Evolutionary patterns in MT genes (13 genes) at their ratio of non-synonymous over synonymous sequence divergence (dN/dS) [left panel], and their non-synonymous substitutions (dN) [right panel]. Colors refer to ratites (blue), tinamous with moderately degenerated chrW (yellow), and tinamous with highly degenerated chrW (red).



Supplementary Figure S6. Correlation between non-synonymous sites (dN) [left panel] and synonymous sites (dS) [right panel] and ratio of differentiated region (DR) over chrZ in length. Each data point represents a certain bird species. Pearson correlation coefficients (r) are shown for comparison. Colors refer to ratites (blue), tinamous with moderately degenerated chrW (yellow), and tinamous with highly degenerated chrW (red).

Supplementary Tables

Supplementary Table S1. Information on genome assemblies used in this study
Supplementary Table S2. Summary of ecological features in Paleognaths
Supplementary Table S3. Temporal dynamics of avian populations
Supplementary Table S4a. Comparison of assembly contiguity statistics in emus
Supplementary Table S4b. Comparison of repeat content in emus (PacBio emu genome was recently produced in the lab)
Supplementary Table S5. Comparison of repeat content across Palaeognathae
Supplementary Table S6. Statistics of AviRTE in tinamous
Supplementary Table S7. Correlation analysis between mean effective population size and percentage of transposable elements
Supplementary Tables S1-S7 are listed as a separate xlsx file due to their large size.