Journal of Genetics and Genomics 49 (2022) 109–119

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Original research

# Phylogeny and sex chromosome evolution of Palaeognathae

Zongji Wang [a, b, c, d, 1], Jilin Zhang [e, 1], Xiaoman Xu [a], Christopher Witt [f], Yuan Deng [d], Guangji Chen [d], Guanliang Meng [d], Shaohong Feng [d], Luohao Xu [b], Tamas Szekely [g, h], Guojie Zhang [d, i, j, k, *], Qi Zhou [a, b, l, *]

[a] MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China
[b] Department of Neuroscience and Developmental Biology, University of Vienna, Vienna 1090, Austria
[c] Institute of Animal Sex and Development, Zhejiang Wanli University, Ningbo, Zhejiang 315100, China
[d] BGI-Shenzhen, Beishan Industrial Zone, Shenzhen 518083, China
[e] Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77 Stockholm, Sweden
[f] Department of Biology and the Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM 87131, USA
[g] State Key Laboratory of Biocontrol, Department of Ecology, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong 510275, China
[h] Milner Center for Evolution, Department of Biology and Biochemistry, University of Bath, Bath BA1 7AY, UK
[i] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[j] Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark
[k] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[l] Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310052, China

ARTICLE INFO

*Article history:*
Received 27 November 2020
Received in revised form
10 June 2021
Accepted 17 June 2021
Available online 13 July 2021

*Keywords:*
Paleognaths
Sex chromosome evolution
Comparative genomics
Sexual selection

ABSTRACT

Many paleognaths (ratites and tinamous) have a pair of homomorphic ZW sex chromosomes in contrast to the highly differentiated sex chromosomes of most other birds. To understand the evolutionary causes for the different tempos of sex chromosome evolution, we produced female genomes of 12 paleognathous species and reconstructed the phylogeny and the evolutionary history of paleognathous sex chromosomes. We uncovered that Palaeognathae sex chromosomes had undergone stepwise recombination suppression and formed a pattern of "evolutionary strata". Nine of the 15 studied species' sex chromosomes have maintained homologous recombination in their long pseudoautosomal regions extending more than half of the entire chromosome length. We found that in the older strata, the W chromosome suffered more serious functional gene loss. Their homologous Z-linked regions, compared with other genomic regions, have produced an excess of species-specific autosomal duplicated genes that evolved female-specific expression, in contrast to their broadly expressed progenitors. We speculate such "defeminization" of Z chromosome with underrepresentation of female-biased genes and slow divergence of sex chromosomes of paleognaths might be related to their distinctive mode of sexual selection targeting females rather than males, which evolved in their common ancestors.



## Introduction

Palaeognathae comprise less than 1% of all avian species but have been intriguing biologists for over a century with an elusive evolutionary history and many extraordinary features that are informative for the origin and evolution of birds. As the basal branch to all other living birds (Neognathae), Palaeognathae includes the flightless ratites (emus, cassowaries, ostriches, kiwis, and rheas), which are predominantly gigantic (except for kiwis), as well as the small pheasant-like, volant tinamous (Cracraft, 1974; Harshman et al., 2008). Recent phylogenomic studies (Baker et al., 2014; Mitchell et al., 2014; Prum et al., 2015; Yonezawa et al., 2017; Cloutier et al., 2019) produced a congruent tree that placed tinamous within ratites, but the rhea node still has low statistical support (Mitchell et al., 2014; Grealy et al., 2017) (Table S1). This phylogenetic

* Corresponding author.
*E-mail addresses:* Guojie.Zhang@bio.ku.dk (G. Zhang), zhouqi1982@zju.edu.cn (Q. Zhou).
[1] These authors contributed equally to this article.

https://doi.org/10.1016/j.jgg.2021.06.013

tree suggested independent evolution of flight loss and gigantism among ratites, which are expected to have a much lower genome-wide evolutionary rate relative to tinamous and other birds of much smaller sizes. Indeed, the ostrich genome is characterized by lower rates of substitutions and insertions/deletions (Jarvis et al., 2014), slower transposon removals (Kapusta et al., 2017), and fewer chromosome rearrangements (O'Connor et al., 2018) than other birds. This parallels the similar comparisons between elephants/whales vs. other mammals (Mank and Ellegren, 2007; Bromham, 2011; Kapusta et al., 2017; Tollis et al., 2019) and can be explained by the fact that large-bodied species tend to have a longer generation time and a lower metabolic rate, hence a reduced genome evolutionary rate per time unit (Bromham, 2011; Berv and Field, 2018). Kiwi is the only small-bodied ratite lineage but has the lowest metabolic rate among all birds (McNab, 1996); therefore, it is also expected to have a low genome evolutionary rate.

The lower genome evolutionary rate of ratites may partially explain the distinctive pattern of homomorphic sex chromosomes found in many ratites (Takagi et al., 1972; Nishida-Umehara et al., 2007) compared with the highly differentiated sex chromosomes of neognathous birds and mammals. We have previously shown that the sex chromosomes of most neognaths have undergone four recombination suppression (RS) events probably through Z- or W-linked inversions (Zhou et al., 2014), with each affected region diverged between the Z/W chromosomes (chrZ/chrW) to a different degree from the nearby regions, together forming a pattern termed "evolutionary strata" (Lahn and Page, 1999). The sex-linked regions that maintained homologous recombination (pseudoautosomal regions [PAR]) of neognaths are mostly very short and concentrated at one chromosome end. Emu and ostrich present a striking exception and have undergone less and smaller sex-linked inversions and have more than two-thirds of the sex-linked region as PAR (Vicoso et al., 2013; Zhou et al., 2014). Interestingly, recent studies (Xu et al., 2019b) reported that some small-bodied tinamous species also have long PARs, whereas our previous work in another tinamou species reported that it has a short PAR (Zhou et al., 2014). This is consistent with cytogenetic results showing a much-diversified pattern of PAR lengths among tinamous (Pigozzi, 2011) and indicates that factors shared by all paleognaths, other than the slow genome evolution rate alone, may account for their homomorphic sex chromosomes.

Once recombination is suppressed, the influenced W-linked regions become transmitted only in females and are expected to undergo functional degeneration because of reduced efficacy of natural selection and the accumulation of deleterious mutations (Charlesworth and Charlesworth, 2000). And the homologous Z-linked regions become transmitted in a male-biased manner and are expected to become "masculinized" or "defeminized" possibly driven by sexually antagonistic selection (Wu and Xu, 2003). Such "sexualization" of sex chromosome, that is, the evolution of sex-linked gene content or expression patterns shaped by their transmission bias toward either sex, has been well documented for the X chromosomes of various *Drosophila* or mammalian species. In addition to sexually antagonistic selection, meiotic sex chromosome inactivation (MSCI), which is the process of transcriptional silencing of the sex (X and Y) chromosomes that occurs during the meiotic phase of spermatogenesis (Lifschytz and Lindsley, 1972; McKee and Handel, 1993; Turner, 2007), plays an important role, and the two factors together drive an excess of RNA- (retrotransposition) or DNA-based duplications from X chromosome onto autosomes in *Drosophila* and mammals (Betrán et al., 2002; Emerson et al., 2004; Bai et al., 2007; Potrzebowski et al., 2008; Meisel et al., 2009; Vibranovski et al, 2009a, 2009b). As a result, testis-biased genes become deficient on the X chromosomes ("demasculinization") (Sturgill et al., 2007) but become enriched among the autosomal duplicated genes derived from the X-linked parental genes (Betrán et al., 2002; Emerson et al., 2004; Meisel et al., 2009; Vibranovski et al., 2009b; Zhang et al., 2010). However, a similar pattern of excessive gene movement off the Z chromosome has not been found in chicken and zebra finch (Toups et al., 2011), probably because of the lack of MSCI in birds (Guioli et al., 2012) and generally, much fewer retrogenes found in the avian genomes compared with those of mammals (Toups et al., 2011). Interestingly, in silkworm, which also has a ZW sex system and lacks MSCI (Traut et al., 2019), there seems to be an excess of retrogenes out of the Z chromosome (Wang et al., 2012) that evolved ovary-biased expression. Overall, whether and if so, why different types of sex chromosomes exhibit different patterns of gene movement remain unclear, as so far, there were much less ZW species than the XY species that have been examined.

To investigate the evolutionary patterns and history of paleognathous sex chromosomes and to explore their potential mechanisms of maintaining large homomorphic sex-linked regions, we produced new female genomes of nine tinamou species and three ratite species and analyzed a total of 15 paleognathous genomes. We reconstructed a new phylogenomic tree using noncoding (NC) sequences spanning nearly 40% of the entire genome of each species, which provides the phylogenetic framework for our subsequent analyses of sex chromosomes. We found that five of the nine investigated tinamou species, despite a much higher evolutionary rate indicated by the longer branch lengths than ratites, share the feature with ratites of not having extensive RS between sex chromosomes. In 8 of the 15 investigated paleognaths, we found an excess of duplicated genes produced by the Z-linked copies onto the autosomes. We finally discussed these distinctive patterns of paleognathous sex chromosomes compared with those of other birds in terms of their different intensities and mode of sexual selection.

## Results

### Genome assemblies and phylogenomic tree of paleognaths

We generated new genome assemblies and annotations for 12 paleognathous species, from between 48- and 106-fold of shotgun sequencing per species (Table S2). We specifically chose female samples of all species for sequencing to include the genomic information of W chromosomes, and densely sampled tinamou species (nine species) to form a comprehensive comparison to ratites (five species, including the two published genomes [ostrich and kiwi]) (Duc et al., 2015; Zhang et al., 2014, 2015), as well as for tinamous' diversified sex chromosome composition and unclear genome evolution patterns (Pigozzi, 2011; Xu et al., 2019b). Half of the sequenced species in this work have draft genomes with a scaffold N50 length longer than 1 Mb (Table S3). Their assembled regions comprise on average 82.35% of the estimated genome size (1.0–1.6 Gb), with a gap content lower than 3% (Table S3). For each genome, we annotated from 14,180 to 16,438 protein-coding genes, and approximately 5% of the genome as repetitive elements, with an average BUSCO score (aves_odb10 database) reaching 88.5 (Table S4). We also assembled the complete mitochondrial (MT) genomes and annotated a complete set of 13 MT genes in all the studied species.

To provide a phylogenetic framework for all the evolutionary inferences in this work, we first constructed phylogenomic trees based on the concatenated alignments of whole-genome NC sequences (Fig. 1A) and of entire MT genomes (Fig. S1A) using the maximum likelihood method and also a tree based on fourfold degenerate sites in coding sequences from single-copy orthologous genes using the Bayesian method (4D; Fig. S1B). After removing alignment ambiguities using the published method (Jarvis et al., 2014) (Materials and Methods), these data sets comprise per species 393,165,276 bp NC sequences (~37.56% of the nuclear genome), 15,873 bp or 92.57% of the MT genome, and 588,067 bp from 4387 orthologous genes or

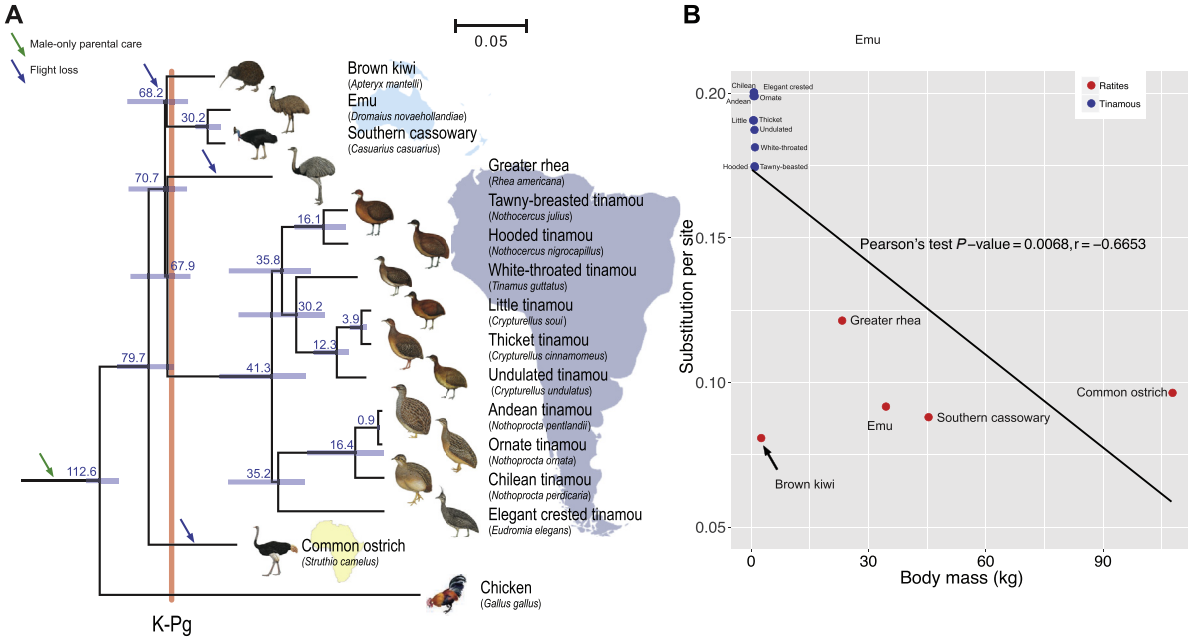


**Fig. 1.** Phylogenomic tree of Palaeognathae. **A**: Maximum likelihood tree reconstructed from genome-wide alignments of NC sequences. Branch lengths indicating the genome substitution rates were estimated by ExaML. All phylogenetic nodes have full bootstrap support and are labeled with estimated species divergence times (in million years) and 95% highest posterior densities (in blue horizontal bars). Based on the tree, all extant paleognaths are grouped by their living landmass shown in the background, and extant ratites probably have had at least three times of independent flight loss marked by blue arrows. All paleognaths have male-only parental care marked by green arrow. The divergence time between rheas and tinamous is very close to the Cretaceous-Paleogene (K-Pg, in red bar) mass extinction about 66 million years ago. All bird illustrations were ordered from https://www.hbw.com/. **B**: Correlation between substitution rate and body mass. Each data point represents a stratum of a certain bird species (dot). Black line depicts a negative linear relationship for substitution per site and body mass.

30.78% of the entire gene repertoire. They represent the largest phylogenomic data set of Palaeognathae known to date (Table S1). All resulting phylogenies have 100% full bootstrapping support for all the nodes of the studied paleognaths, with all the topologies congruent with each other and previous works (Mitchell et al., 2014; Grealy et al., 2017; Yonezawa et al., 2017; Cloutier et al., 2019), except for the phylogenetic position of rhea. NC and MT trees consistently place the greater rhea as the sister species to all tinamou species, with both taxa now living in South America. This result is different from those of 4D tree and previously published trees (Cloutier et al., 2019) using the coalescence (CL tree) method or based on phylogenetic distribution of transposons (Prum et al., 2015), which produced different estimates of minimum times of flight loss and evolution of gigantism among ratites (Fig. S2). The different phylogenetic positions of rheas between these trees can be attributed to the different sizes and types of the input data and also different tree reconstruction methods. Particularly, the extremely short internal branch length of rhea found in our (Fig. 1A) and published works (Parker et al., 2013; Jarvis et al., 2014) is probably the major cause of the different topology. In this work, our NC alignment data comprise nearly 40% of the entire genome and are expected to be much less influenced by the convergent evolution of life history traits than the 4D or exon sites (Braun et al., 2019; Cloutier et al., 2019). We therefore chose the NC tree topology as the guide tree for all the subsequent analyses in this work.

Our estimation of species divergence times with the NC tree is generally consistent with the previous results (Mitchell et al., 2014; Yonezawa et al., 2017) (Fig. S3), with some differences in the tinamou lineages because of the more tinamou species sampled in this study. Particularly, the estimated divergence time (approximately 67.9 million years [MY]) of rhea from its sister group is around the Cretaceous-Paleogene (K-Pg) mass extinction 66 MY ago (Jarvis et al., 2014) when high levels of incomplete lineage sorting (ILS) also occurred during early diversification of the Neoaves. Overall, the branch lengths of ratites are much shorter than those of tinamous ($P = 1e\text{-}07$, relative rate test; Fig. 1A). And there is a significant negative correlation between the genome-wide substitution rate and the body mass after controlling for the phylogeny ($P = 0.0089$, PGLS test; Fig. 1B). This demonstrated the impact of the evolution of gigantism in slowing down the genome evolutionary rates of ratites. One outlier of ratites is the kiwi with a small body size, which shows an even shorter branch length than all the other large-bodied ratites. This is consistent with previous results (Sackton et al., 2019) and is probably because kiwi has the lowest metabolic rate among all birds (McNab, 1996), which has a similar impact with gigantism on the genome evolutionary rate.

## Complex evolutionary history of the slow-evolving paleognathous sex chromosomes

To reconstruct the evolutionary history of paleognathous sex chromosomes, we first identified the sex-linked sequences of each studied species based on their homology to the Z chromosome sequence of ostrich (Zhang et al., 2015; Yazdi and Ellegren, 2018). We then ordered and oriented these scaffolds into pseudochromosome sequences according to their unique mapped positions along the ostrich reference. The lengths of the identified Z chromosome sequences of each paleognaths range from 68.77 to 82.13 megabases (Mb), corresponding to 83.0%–97.6% of the ostrich Z chromosome length (Table S5). We annotated between 550 and 930 Z-linked genes and 12 to 286 W-linked genes across the studied species (Table S6). We then determined the length of PAR of each studied species, which is expected to have a similar level of aligned female read depth with autosomes or a similar read depth between sexes if data of both sexes are available. The lengths of the

identified PARs range from 0.7 to 56.9 Mb, encompassing 15 to 580 Z-linked genes (Table S6). We found no significant association between the proportion of the Z chromosome that is identified as the PAR, that is, the degree of sex chromosome homomorphism, and the species' body mass ($P = 0.2017$, PGLS test) after controlling for the phylogeny. The lack of association can be caused by the small number of the studied species. It can also be because of the fact that 5 of the 15 studied paleognaths are small-bodied tinamou species with a long PAR extending at least half of the entire sex chromosome length (Fig. 2A), similar to the PARs of large-bodied ratites. These results showed that all the 15 investigated paleognaths except for 4 tinamous have much longer PARs than neognathous birds or mammals, regardless of their great variations of genome-wide evolutionary rates (Fig. 1A).

In the sexually differentiated regions (SDR) where recombination has been suppressed, particularly in the older evolutionary strata, we expected to be able to assemble much less W-linked sequence fragments and observe a lower degree of Z/W pairwise sequence similarity than in the younger strata. This is because the former has undergone sequence divergence from the homologous Z-linked regions for a longer time, with more repeats accumulated and more sequence deletions. The density of W-linked fragments and their sequence divergence level exhibit a sharp shift between neighboring regions of their homologous Z chromosome, forming a clear pattern of evolutionary strata (Figs. 2A and S4) and allowing us to define the strata boundaries. To date each stratum, we aligned their boundaries across species and concluded a more ancestral stratum is expected to share its boundaries across multiple species,

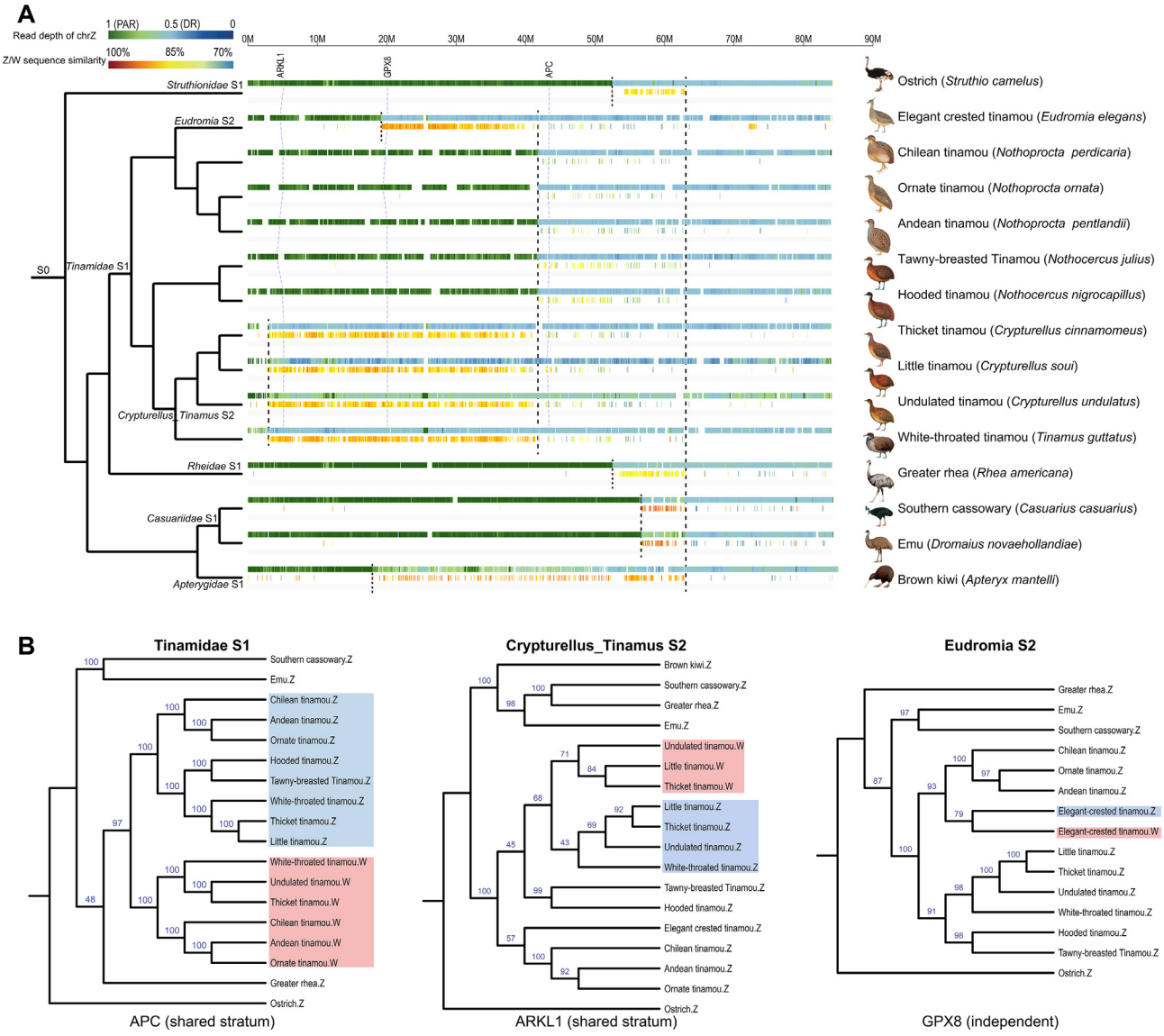


**Fig. 2.** Evolutionary strata in Palaeognathae. **A**: Evolutionary strata of sex chromosomes. Each species has two tracks showing the color-scaled female read depths along the Z chromosome normalized against the median value of the read depths of autosomes and the level of Z/W pairwise sequence divergence. PAR region is expected to show a normalized female read depth around 1 (in green), whereas the SDR whose recombination has been suppressed is expected to show a normalized read depth around 0.5 (in light blue). The faster one SDR region is evolving between sex chromosomes, the less W-linked fragments can be identified, and the more diverged in sequence that homologous Z/W regions are. The boundaries between two neighboring evolutionary strata were inferred by the sharp shift of the Z/W divergence level or the identified W-linked fragments or that of the female read depth. We named each evolutionary stratum according to the phylogenetic branch that it originated and labeled them on the species tree. All bird illustrations were ordered from https://www.hbw.com/. **B**: Examples of gene trees for Z- and W-linked gametologs at different evolutionary strata. If sequences were grouped by their chromosomal origins, their residing evolutionary stratum was inferred to emerge before the species divergence. If sequences were grouped by their species origins, the stratum was inferred to emerge after the species divergence. The chromosomal positions of the example genes were labelled (**A**).

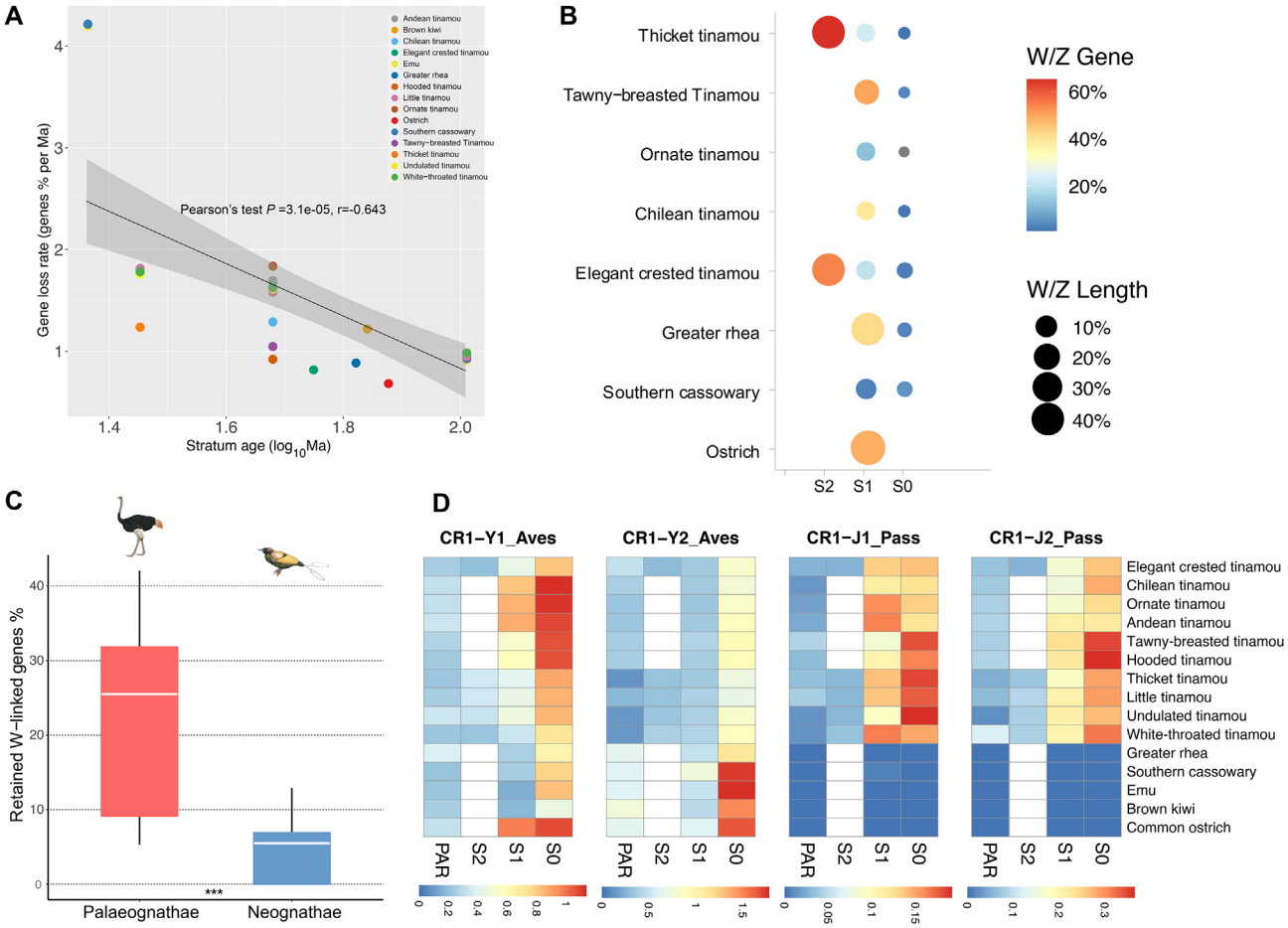


**Fig. 3.** Evolution of sex-linked genes and transposable elements in Paleognathae. **A**: Correlation between gene loss rate and age of stratum. Each data point represents a stratum of a certain bird species. **B**: Comparison of W-linked genes and sequence lengths between different evolutionary strata. The areas of the circles are scaled to the W-to-Z ratios of the assembled lengths, and the colors of the circles are scaled to the W-to-Z ratios of the annotated genes. **C**: Comparison of the percentages of retained W-linked genes at the SDRs of Palaeognathae vs. those of Neognathae. ***, $P = 2.21e{-}05$, Wilcoxon test; **D**: Comparing the CR1 content across Palaeognathae from different evolutionary strata. Each cell corresponds to the average density (measured by copy numbers of certain CR1 subfamily over region length) of certain CR1 subfamilies (shown at the top) at an evolutionary stratum (*X*-axis). Blank cell indicates that there is no corresponding region in that species. CR1, chicken repeat 1; S0, stratum 0; S1, new stratum; S2, recent stratum; PAR, pseudoautosomal regions; SDR, differentiated regions.

whereas a more recent species-specific stratum does not. In addition, if the Z- and W-linked genes (also called "gametologs") of an evolutionary stratum started to diverge before a lineage split, their gametolog trees should include separate groups of Z and W sequences. In contrast, a lineage-specific evolutionary stratum is indicated when the Z/W gametologs of the same species group together. Based on these principles and the species divergence times, we first determined the evolutionary stratum shared by all birds (stratum 0, S0), thus also shared by all paleognaths, which encompasses the avian male-determining gene *DMRT1* (Zhou et al., 2014) and emerged over 100 MY ago. After their divergence, ostrich, the ancestor of all tinamous, rhea, the ancestor of emu and cassowary, and kiwi have each independently experienced a second RS event that each formed a new stratum (S1); these 4 events are dated to different times, ranging from 23 to 75 MY ago (Table S6). As we have only one species of rhea and kiwi, this S1 could have formed either in the ancestors of all rheas and all kiwis, or specifically in the species that we studied here. Some tinamou lineages (the ancestor of *Crypturellus* genus, including thicket, little, undulated, and white-throated tinamous) and also the elegant crested tinamou have independently formed a more recent stratum (S2), at times estimated to be between 26 and 56 MY ago (Fig. 2A; Table S6). As expected, the analyses of such trees for genes with positions known in the ostrich reference assembly yielded ages of evolutionary strata consistent with those inferred from inferring their boundaries from W-Z divergence of individual genes (Figs. S5–S12). That is, an evolutionary stratum that was dated to the ancestor of multiple species is expected to share their boundaries within these species. For example, all tinamou species share one evolutionary stratum (Tinamidae S1, named after its phylogenetic node when it originated), whose boundaries are aligned between tinamous by the nearby PAR or S2 sequences. Consistently, gametologs of *APC* gene from this stratum are grouped separately by Z and W chromosomes (Fig. 2B). Gametologs of *ARKL1* gene are also grouped by sex chromosomes, suggesting the evolutionary stratum (Crypturellus_Tinamus S2) started to diverge before speciation (Fig. 2B). While for the gene *GPX8*, whose Z/W recombination was suppressed after elegant crested tinamou diverged from its sister species in Eurdomia S2 region, its gametologs are grouped by species (Fig. 2B).

## Genetic degeneration of sex chromosome and evolution of sex-linked transposable elements

The difference in the numbers of genes between Z and W chromosomes reflects the degeneration of W chromosomes after

recombination was suppressed. Our analysis found that older evolutionary strata had expectedly lost more genes because of the longer time of evolving under a nonrecombining environment (Table S6). The rate of gene loss (percentage of lost genes per million years) however becomes much reduced in the older strata (Fig. 3A). This is consistent with our previous finding in other birds (Zhou et al., 2014) or the work in *Silene latifolia* (Krasovec et al., 2018) and reflects a weaker effect of Hill-Robertson interference caused by less genes retained in the older strata. Both the assembled W-linked lengths and the numbers of annotated W-linked genes relative to those of their Z-linked homologs decrease in the older evolutionary strata (Fig. 3B). Also, the synonymous substitution rates (*Ks*) between the Z/W gametologs are generally higher in the older evolutionary strata, although low numbers of genes (e.g., there are less than five genes assembled in S0 in many species) and gene conversion that reduce the *Ks* values are expected to confound this pattern (Fig. S13; Table S7). In general, paleognaths not only retained a longer PAR (Fig. 2A) but also retained a significantly ($22.39 \pm 13.628$ vs. $4.429 \pm 4.389$, $P = 2.21e\text{-}05$, Wilcoxon test) higher percentage of W-linked genes at SDRs, relative to neognaths (Fig. 3C). This indicated that the Palaeognathae sex chromosomes not only underwent less frequent RS but also diverged from each other at a lower rate after the recombination was suppressed. Retained W-linked genes at SDRs of S0 and S1 regions show a decreased expression level relative to their Z-linked homologs among all the examined tissues in emu and Chilean tinamou with transcriptomes of both sexes available (Fig. S14). This provides direct evidence for functional degeneration of W-linked genes after the RS.

The formation of evolutionary strata has also reshaped the TE landscape of the Z chromosomes because of the reduction of effective population size and average recombination rate from different time points when the influenced region of Z chromosomes has stopped recombination in females. This has led to the differential accumulation of TEs on the Z chromosome so that different evolutionary strata are predicted to have different TE densities. Indeed, the Z-linked SDRs in all studied species have higher TE content ($7.337 \pm 3.691\%$ vs. $5.396 \pm 2.121\%$, $P < 0.01$, Wilcoxon test) than PARs or autosomes and in the older strata compared with the younger ones (Fig. S15). Chicken repeat 1 (CR1) elements, a family of non-long terminal repeat retrotransposons, that were active throughout the Palaeognathae evolution history (Wang et al., 2021) are expectedly enriched on the Z chromosomes across all species, whereas some families (e.g., CR1-J1/2_Pass) are tinamou specific and are enriched only on tinamou Z chromosomes (Fig. 3D).

## An excess of gene movements off the paleognathous Z chromosomes, but not from PAR

We hypothesized that if sexually antagonistic selection is shaping the gene movement on the male-biasedly transmitted Z chromosomes of paleognaths, their SDRs but not PARs are expected to either harbor an excess of duplicated genes derived from autosomes (A-to-Z duplicates) that evolve male- or testis-biased gene expression ("masculinization") or to produce an excess of duplicated genes on autosomes (Z-to-A duplicates) that evolve female- or ovary-biased gene expression ("defeminization"). To test this, we first identified all the interchromosomal duplicated genes of each species, annotated each of them as either DNA- or RNA-based duplication, then divided them into three categories of "A-to-A," "Z-to-A," and "A-to-Z" according to their source and target chromosomes. We found that 8 of the 15 studied species show a significant ($P < 0.05$, chi-squared test) excess of Z-to-A gene movements (Fig. 4A), which is almost exclusively contributed by their old evolutionary strata S0, and S1 in tinamous, but not from the younger strata S1 in ratites and S2 in tinamous, or PAR (Fig. 4B). In contrast, we have not found a significant excess of A-to-Z gene movements in any species. The Z-to-A duplicated genes are predominantly (22.0%–93.8% across species) derived from DNA-based duplications. If we only considered RNA-based duplications, there are no significant Z-to-A movements (Table S8) because of the very low numbers of such duplications identified at the genome-wide level.

We further compared the gene expression patterns of duplicated genes on the autosomes vs. those of their parental genes on the Z chromosome in ostrich and Chilean tinamou with multiple tissue transcriptomes of both sexes available. Interestingly, between 33.3% and 66.7% of the autosomal gene copies have evolved a female-specific gene expression pattern in both gonad and somatic tissues, in contrast to their broadly expressed Z-linked parental genes

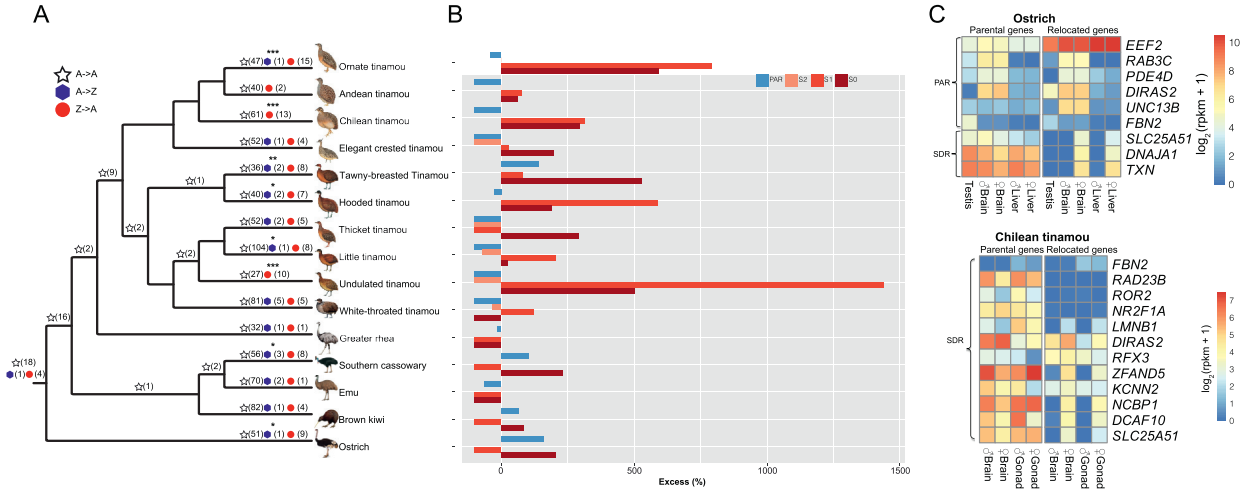


**Fig. 4.** Gene movement off the Z chromosome in Paleognathae. **A**: Distribution of DNA-based gene movements in the Palaeognathae phylogenetic tree. Inferred ancestral gene movements are labeled at corresponding phylogenetic nodes. Three different shapes and colors correspond to the three types of movement directions (presented as legend). *, $P < 0.05$; **, $P < 0.005$; ***, $P < 0.0005$, chi-squared test. **B**: Comparison of excess across different evolutionary strata. Excess = ([O − E]/E) × 100, where E is expected and O is observed. **C**: The heatmap shows the expression levels of duplicated genes on the autosomes (right: relocated genes) vs. those of their parental genes on the Z chromosome (left: parental genes) across different tissues in ostrich and Chilean tinamou. We grouped these genes by their chromosomal regions shown on the left of the *Y*-axis. S0, stratum 0; S1, new stratum; S2, recent stratum; PAR, pseudoautosomal regions; SDR, differentiated regions.

(Fig. 4C). This diverged expression pattern between parental and daughter gene copies was not found in the PARs of the two species. These results together suggested that in some paleognathous species, sexually antagonistic selection probably accounts for the stepwise defeminization of paleognathous Z chromosomes, that is, relocation of female-biased gene duplications onto the autosomes from the Z chromosome, during its stepwise suppression of recombination with the W chromosome.

## Discussion

Birds harbor a tremendous diversity in their extent of sex chromosome differentiation (Zhou et al., 2014; Xu et al., 2019a; Xu and Zhou, 2020) and therefore provide a great model for understanding the evolutionary factors influencing the tempo of sex chromosome evolution. Such diversity is not observed among the XY sex chromosomes of eutherian mammals (Graves, 2006; Bachtrog, 2013; Cortez et al., 2014), whose Y chromosomes have all become highly degenerated or even completely lost. The "male-driven evolution" effect (Li et al., 2002) predicts that the male-specific Y chromosomes have the highest mutation rate, whereas the female-specific W chromosomes have the lowest mutation rate throughout the genome. However, the "male-driven evolution" effect alone cannot explain the different evolutionary patterns of sex chromosomes between birds and mammals. This is because the majority of birds, such as chicken and zebra finch, also have a highly degenerated W chromosome similar to the mammalian Ys. Here, we first reconstructed the phylogenomic tree of Paleognathae species. Because of the different data sets and reconstruction methods used in this study, particularly the impact of ILS, we acquired a tree topology that differs from the previous results for the position of rheas (Fig. 1). This however does not change our major conclusions on the patterns of evolutionary strata (Fig. 2) and gene loss rate (Fig. 3); because majorities of ratite lineages formed their evolutionary strata independently because of the species radiation, and rheas only contributed one data point on the pattern of gene loss rate. In addition, we found that many tinamous, which have a much higher genome-wide substitution rate, but a similar homomorphic sex chromosome pair compared with the ratites, also suggest that other factors may be responsible for the generally lower sex chromosome evolution rate of paleognaths relative to other birds.

We speculate that one of the possible factors could be the distinctive mode of paternal care originated in the paleognathous ancestor. Paleognaths have predominant (e.g., in ostrich) or exclusive (in tinamous) paternal care behaviors that have been reported in only ~1% of extant bird species (Handford and Mares, 1985; Cockburn, 2006). Despite such rarity, paternal care likely has evolved before the divergence of paleognaths and their sex chromosomes (Handford and Mares, 1985; Wesolowski, 1994; Tullberg et al., 2002; Cockburn, 2006; Varricchio et al., 2008). Meta-analyses of life history traits of birds showed that postmating paternal care is usually associated with reduced premating sexual selection targeting males, manifested as reduced levels of male-biased sexual size dimorphism and extrapair paternity (Remeš et al., 2015). Indeed, the reversal of sex roles in parental care seems to result in a significantly different level ($P = 3.816e{-}05$, Wilcoxon test) of sexual size dimorphism between Palaeognathae and Neognathae (Fig. S16), with most paleognaths having larger females than males (Table S9). Unlike most other vertebrates, it has been reported in emu that females would compete vigorously for access to males during mating season (Coddington and Cockburn, 1995). These results led us to propose that the evolution of paternal care in the Palaeognathae ancestor probably "resolved" the postmating sexual conflict over care (Szekely, 2014) and reduced the intensity of premating sexual selection, eventually producing the unique pattern of homomorphic sex chromosome pattern of Palaeognathae. However, why some tinamous nevertheless suppressed recombination throughout almost the entire sex chromosome pair and have extensive differentiation between sex chromosomes still remains unclear. We have to point out that this hypothesis cannot be directly tested in paleognaths because paternal care only originated one time in their ancestor. More insights can be gained regarding the relationship between the parental care mode and sex chromosome evolution by comparing species, for example, shorebirds that have recently evolved diverse modes of parental care (Reynolds and Székely, 1997).

Another interesting and relevant finding is that in more than half of the studied paleognaths, there are an excess of duplicated genes derived from the Z chromosome onto the autosomes, which evolved female-biased gene expression. A similar pattern of extra "X-to-A" duplicated genes that evolved testis expression has been well documented in *Drosophila* and mammalian species (Betrán et al., 2002; Emerson et al., 2004; Vibranovski et al., 2009b). Such redistribution of sex-biased genes between X chromosome and autosomes was interpreted to be caused by MSCI and/or sexually antagonist selection. Birds do not seem to have MSCI and have very few retrogenes, which may explain a lack of extra "Z-to-A" duplicated genes in chicken and zebra finch. However, as there is probably stronger sexual selection targeting females in paleognaths, female-biased duplicated genes maybe more frequently fixed on the autosomes than other avian species, resulting in the "defeminization" pattern of Z chromosomes of many paleognaths observed in this work.

## Materials and Methods

### Genome assembly and annotation

All animal procedures were carried out with the approval of the China National GeneBank animal ethics committee. We extracted 12 female genomic DNAs from vouchered museum specimens for 10 of the studied species and from zoo samples for 2 of the studied species (Table S10) using Gentra Puregene Tissue Kit (QIAGEN). Illumina paired-end DNA libraries were constructed and sequenced (Hiseq 4000) following the manufacturer's protocol with varying insert sizes (250 bp, 800 bp, 2 or 5 kb) for each species. We used customized perl scripts to remove the reads containing more than 10% ambiguous nucleotides or with more than 40% low-quality (Phred score $\leq$ 7) bases or derived from adapter sequences and PCR duplications. We further trimmed the low-quality ends of reads if abnormal base compositions or low read qualities were detected. We then used the cleaned reads for *de novo* genome assembly using SOAPdenovo2 (Luo et al., 2012) (version 2.04). Different k-mers were tested for SOAPdenovo2 for each species to achieve the largest contig/scaffold N50 lengths. We then polished the assembly by Gapcloser (v1.12). We used Tandem Repeats Finder (Benson, 1999) (TRF Version 4.09), RepeatProteinMask and RepeatMasker (version 4.0.6) with RepBase (Bao et al., 2015) (20160829), and the *de novo* prediction program RepeatModeler (version open-1.0.8) together to predict and categorize the repetitive elements. We used the Kimura 2-parameter distance to estimate the divergence level between the individual TE copies vs. their consensus sequences. Kimura distances between genome copies and TE consensus were calculated using the RepeatMasker built-in scripts (calcdivergencefromalign.pl). We dated the expansion of certain repeat subfamilies based on their sequence divergence pattern in the phylogeny by parsimony. We inferred more recent or ancestral subfamilies if they show a sequence divergence pattern concentrated at a lower or higher divergence level and shared among species or specific to certain lineages. For gene annotation, we first defined the candidate gene regions after aligning the query protein sequences of zebra finch, chicken, and human (Ensembl

release-87) to the targeted genomes using TBLASTN (E-value ≤ 1e−5). Then we refined the candidate gene regions by GeneWise (Birney et al., 2004). One thousand such homology-based genes with a GeneWise score of 100 were randomly selected to train Augustus (Stanke et al., 2004) (version 3.2.1) for *de novo* prediction. And we combined gene models from these two sources into a nonredundant gene set. The resulting gene models (measured by gene length, mRNA length, exon number, and exon length) are comparable to those of other vertebrates. Gene names were further translated to those of their orthologs in the query species based on the reciprocal best BLAST hit.

### Phylogenomic analyses

The protein-coding genes from three amniotic species (*Gallus gallus*, *Anolis carolinensis*, and *Homo sapiens*) were downloaded from Ensembl (release-87). The gene sets from brown kiwi (PRJEB6383), white-throated tinamou, and common ostrich (PRJNA212876) were obtained from National Center for Biotechnology Information (NCBI). For gene loci with alternative splicing isoforms, only the longest transcript was retained. Then we used Treefam (Li et al., 2006) to cluster genes from different species into gene families. (1) We performed all-vs.-all alignments using BLASTP with E-value <1e−7 and joined the fragmental alignments using Solar (a program in Treefam). The alignments were used to calculate the distance between two genes. Next, a hierarchical clustering algorithm was used to cluster all the genes, with the following parameters: min_weight = 10, min_density = 0.34, and max_size = 500. (2) Multiple sequence alignment (MSA) for each gene family was performed by MUSCLE (Edgar, 2004) (v3.8.1551), and fourfold degenerate sites were extracted and concatenated to generate superalignments. Our final data set contained a total of 54,013 fourfold degenerate sites. We used MrBayes (Ronquist and Huelsenbeck, 2003) (v3.1.2) to reconstruct the phylogeny with the (GTR + gamma) model using the following parameters: lset nst = 6 rates = gamma; mcmc ngen = 100000 printfreq = 100 samplefreq = 100 nchains = 4 savebrlens = yes. For the NC tree, we first generated the whole genome alignments with LASTZ (1.04.00) using the genome of the ostrich as the reference. In detail, each species' genome was first built into pseudochromosome sequences using ostrich as reference, using a LASTZ parameter set designed for distant species comparison (–step = 19 –hspthresh = 2200 –inner = 2000 –ydrop = 3400 –gappedthresh = 10000 –seed = 12of19 for alignment, –minScore = 1000 –linearGap = loose, with the scoring matrix HOXD55 for the chain step). The pairwise alignments were further processed by MULTIZ (Blanchette et al., 2004) with the ROAST approach to construct multiple genome alignment, using a guided tree fixing the known phylogenetic positions of ostrich, and chicken as an outgroup. We then used the alignment filtering code (filter_alignment_maf_v1.1B.pl –window 36 –minidentity 0.55) described in the study (Jarvis et al., 2014) to remove the poorly aligned regions. The MSAs were then passed to MAFFT (Katoh and Standley, 2013) (–maxiterate 1000 –localpair) to obtain further locally refined alignments. To construct the ML tree, we first preprocessed the noninformative sites with RAxML. This step resulted in 393,165,276 whole-genome aligned sites with 10.34% gaps and undetermined sites removed, of which 15,565,579 distinct patterns can be used to estimate the phylogenetic topology. The final MSA was first randomized into 100 subsets, and each subset was used to generate a parsimony tree as the starting topology with RAxML (Stamatakis, 2014) (8.2.11). Each starting tree was used together with its corresponding sequence subset to estimate a maximum likelihood tree with ExaML (Kozlov et al., 2015) (3.0.19) under the GTRGAMMA model. Of the 100 ML trees constructed above, the one with maximum ML was further used to evaluate the consistency and obtain the bootstrap value for the final tree. To further avoid the bias introduced by the protein-coding sequences, a genome-wide NC tree was further reconstructed based on the above whole genome alignments by excluding the coding sequences using msa_view in PHAST (Hubisz et al., 2011) (v1.3). In total, 348,449,087 usable sites were obtained to construct the NC tree with the same approaches applied to the whole genome tree. The MCMCtree program (version 4.4) implemented in the Phylogenetic Analysis by Maximum Likelihood (PAML) (Yang, 2007) package was used to estimate the species divergence time. Calibration time was obtained from the TimeTree database (http://www.timetree.org/). Two calibration points were applied in this study as normal priors to constrain the age of the nodes described below: 105−118 MA for the most recent common ancestor (TMRCA) of ostrich-chicken and 271−286 MA for TMRCA of Aves and Dactyloidae. MitoZ (Meng et al., 2019) was used with the default settings to assemble the whole MT genome for all studied Palaeognathae using a random subset (5 Gb) of paired-end reads from small insert-size library sequencing. The genes were identified using MITOS (Bernt et al., 2013) and curated by comparison with known sequences of other published ratites and tinamous from GenBank. Complete MT genomes were also used to construct a maximum likelihood phylogenetic tree using the same method as the NC tree. We tested the association between the body mass vs. the degree of sex chromosome homomorphism using Phylogenetic Generalized Least Squares (PGLS) models implemented in the R packages APE (Paradis et al., 2004) and nlme (https://CRAN.R-project.org/package=nlme).

### Identification of sex-linked sequences

We identified the Z chromosome sequences out of the draft genomes (14 females and 4 males from 14 species) based on their alignments to the Z chromosome sequences of ostrich (Zhang et al., 2015; Yazdi and Ellegren, 2018) and also a female-specific reduction of the read depth. Scaffold sequences of each species were aligned with LASTZ (version 1.02.00) to the ostrich Z chromosome sequence with parameter set "–step = 19 –hspthresh = 2200 –inner = 2000 –ydrop = 3400 –gappedthresh = 10000 –format = axt" and a score matrix set for distant species comparison. Alignments were converted into a series of syntenic "chains," "net," and "maf" results with different levels of alignment scores using UCSC Genome Browser's utilities (http://genomewiki.ucsc.edu/index.php/). Based on the whole genome alignments, we first identified the best aligned scaffolds within the overlapping regions on the reference genome, according to their alignment scores with a cutoff of at least 50% of the whole scaffold length aligned in the LASTZ net results. We further estimated the overall identity and coverage distributions with a 10 kb nonoverlapped sliding window for each scaffold along the reference sequence and obtained the distributions of identity and coverage. Scaffolds within the lower 5% region of each distribution were removed to avoid spurious alignments. Finally, scaffolds were ordered and oriented into pseudochromosome sequences according to their unique positions on the reference chrZ in ostrich. Scaffolds were linked with 600 "N"s as a mark of separation. W-linked sequences are expected to also form an alignment with the reference Z chromosome of ostrich but with lower numbers of aligned sequences and lower levels of sequence identity than their homologous Z-linked sequences because of the accumulation of deleterious mutations after recombination was suppressed on the W. We also expect that there are still certain degrees (at least 70% as a cutoff) of sequence similarities between the Z- and W-linked sequences for discriminating the true W-linked sequences from spurious alignments. After excluding the Z-linked sequences from the draft genome, we performed a second round of LASTZ alignment against the Z chromosome sequences of each species built from the previously mentioned step. Then we excluded the spurious alignments with the cutoff of the pairwise sequence identity to be higher than 70% but lower than 95% and with the aligned sequences spanning at least 50% of the

scaffold length. We further verified the W-linked sequences in species with sequencing data of both sexes available (Fig. S17). We identified the missing Z-linked sequences (Table S11), if any because of the lack of substantial sequence similarity with the ostrich Z chromosome, in species with sequencing data of both sexes available (greater rhea, southern cassowary, emu, Chilean tinamou, and elegant crested tinamou). They are expected to show a reduced coverage in female than in male. We first aligned the female and male reads to all the scaffold sequences separately with BWA-MEM. We then calculated the mean depth for each scaffold using the single-base depths produced by SAMtools. We defined a putative Z-linked scaffold with the male vs. female ratio to be within the range of 1.8–2.2. Then we compared the lengths of all putative Z-linked scaffolds with those identified from alignments with the ostrich Z chromosome. We aligned the protein sequences of Z-linked genes with BLAT to the W-linked sequences and then annotated the W-linked genes with GeneWise (Birney et al., 2004). To construct the gametolog trees, CDS sequences of single-copy genes' Z/W gametologs were aligned by MUSCLE (Edgar, 2004), and the resulting alignments were cleaned for ambiguous alignments by gblocks (-b4 = 5, -t = c, -e = -gb). Only the alignments longer than 300bp were used for constructing maximum likelihood trees by RAxML (Stamatakis, 2014) to infer whether their residing evolutionary stratum is shared among species or specific to certain lineages. Then all gametologs trees were manually checked and categorized as supporting the evolutionary stratum to be originated either before or after the speciation or simply as noninformative because of the low bootstrapping values (<50; Table S12).

### Characterization of PARs and evolutionary strata

We aligned the raw reads of each species to their pseudochromosome sequences by BWA (Li and Durbin, 2009) (0.7.12-r1039) with BWA-MEM algorithm. After removing PCR duplicates, we calculated read depth using SAMtools (version: 1.3.1) within each 100 kb nonoverlapping window and normalized it against the median value of depths per single base pair throughout the entire genome to allow comparison among species. Boundaries of PAR were determined by a sharp shift of depth values between neighboring windows. We first plotted all the depth values along the chrZ for each species. Then the sharp shift between PAR and differential region is directly visible, which helps us to define the boundaries of the strata. Then we performed Wilcoxon's test to test its significance. Similarly, we then scanned the Z/W pairwise alignment along the Z chromosome with a nonoverlapping 100 kb window to determine the boundaries between the neighboring strata, where there expected to be a sharp shift of the pairwise sequence identities or the occurrences of W-linked scaffolds. The information of PAR, SDR, and W-linked gene numbers from Neognathae were retrieved from our previous study (Zhou et al., 2014; Xu et al., 2019a). To determine the heterozygosity level of the evolutionary strata, we used the GATK (DePristo et al., 2011) (version 3.6-0-g89b7209) for Single Nucleotide Polymorphism (SNP) calling in both sexes for elegant crested tinamou and Chilean tinamou. SNPs whose read depths were too low (<100) or qualities lower than 100 were excluded. Boundaries of the very young stratum were determined by a sharp shift of heterozygosity ratio between female and male.

We used the method of assuming the molecular clock to infer the age of evolutionary strata as well as relying on the species divergence time. In brief, we calculated branch-specific synonymous substitution rates (dS) of genes on the macrochromosomes by PAML (Yang, 2007) (free-ratio model) for each species. The species-specific mutation rate to that of chicken ($2.5 \times 10^{-9}$/site/year) was scaled by their ratio of median values of the branch-specific synonymous substitution rates. We calculated the degree of male-biased mutation ($\alpha$) by using the equation $\alpha = 3\ (Z/A - 2)/(4 - 3Z/A)$, where "Z" and "A" represent the median substitution rates of intronic regions linked to autosomes and Z chromosome. The mutation rate of the W chromosome is then estimated from the equation $\alpha = 2A/W - 1$ (Miyata et al., 1987), where "W" refers to the mutation rate of the W chromosome. The sum of Z and W chromosome mutation rates in different strata results in the local divergence rate between the sex chromosomes. For the Z/W divergence level, the coding regions were removed from the pairwise Z/W genome alignments generated previously from LASTZ by customized perl scripts, and then we estimated the divergence level within NC regions by baseml in the PAML package after 1000 bootstraps. The age of a stratum was finally estimated by dividing the divergence level by divergence rate. Alternatively, the age of the strata was also estimated by the species divergence time when we can map the strata to certain phylogenetic nodes.

We downloaded the transcriptomes of green anole lizard (brain, gonad: PRJNA381064) and Chilean tinamou (brain and gonad: PRJNA433114) from SRA. In addition, we collected the transcriptomes of adult emu brain, kidney, and gonad from both sexes. We collected the total RNAs of adult tissues (brain, kidney, and gonad) of both sexes using TRIzol Reagent (Invitrogen #15596-018) following the manufacturers' instructions. Then paired-end libraries were constructed using NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, USA), and approximately 3.5 Gb paired-end reads of 150 bp were produced for each library. Overall, we used the data from three emu individuals, two lizard individuals for each sex and one tinamou individual here. We used HISAT2 (Kim et al., 2015) (version 2.1.0) for aligning the RNA-seq reads against reference genomes. Gene expression was measured by reads per kilobase of gene per million mapped reads (RPKM). To minimize the influence of different samples, RPKMs across tissues were adjusted by a scaling method based on TMM (trimmed mean of M values; M values mean the log expression ratios) (Robinson and Oshlack, 2010), which assumes that the majority of genes have similar expression levels across all samples. The mean RPKM value of biological replicates was calculated for each gene. Then we calculated the fold-change difference (>2) in gene expression between both sexes to detect differential expression.

### Identification of interchromosomal gene movements

For DNA-based gene relocations, we used MCScanX (Wang et al., 2012) to detect intraspecies collinearity blocks between each studied species and green anole lizard and chicken with default settings (5 genes required to call a collinear block) based on the all-to-all BLASTP result (E-value $\leq$ 1e−5 and identity > 75). Then we used MCScanX-transposed (Wang et al., 2013) to detect transposed gene duplications. To further filter out false transposed gene duplications, we only kept gene pairs in which parental and transposed duplicates have the same gene name based on the SwissProt annotation. For RNA-based gene relocations, we first aligned all peptide sequences (retrieved from our homology-based gene prediction) against all peptide sequences using BLASTP with E-value < 1e−7. We screened out gene pairs with a blast score ratio of best hit over second best hit < 0.8. We further screened out potential paralog pairs generated by retrotransposition, which contain one single-exon gene (potential retrogene) and one multiple-exon gene (potential parental gene). Then we confirmed the intronlessness of candidate retrogene by aligning the protein sequence of potential parental gene against the DNA sequence of potential retrogene with GeneWise (Birney et al., 2004). We only kept TE-free paralog pairs for further analyses, in which neither of the two genes had >50% CDS occupied by TEs. We defined more conservative paralog pairs with >50% align rate and >50% identity. We used the same way

(Vibranovski et al., 2009b) to calculate the expected number of interchromosomal gene movements.

## Data availability

All the genomic reads, the genome assemblies, and annotations generated in this study have been deposited in NCBI SRA and GenBank under the project accession number PRJNA545868. The previously mentioned data have also been deposited in the CNGBdb (https://db.cngb.org/cnsa/) with accession number CNP0000505. All customized codes used in this study are available at https://github.com/zj-wien/Paleognaths.

## CRediT authorship contribution statement

**Zongji Wang**: Conceptualization, Methodology, Software, Writing - Original draft, Visualization, Investigation, Formal analysis. **Jilin Zhang**: Methodology, Visualization, Investigation, Formal analysis. **Xiaoman Xu**: Investigation. **Christopher Witt**: Resources, Writing - Review & Editing. **Yuan Deng**: Investigation, Data curation. **Guangji Chen**: Investigation, Methodology. **Guanliang Meng**: Methodology. **Shaohong Feng**: Investigation, Methodology. **Luohao Xu**: Validation. **Tamas Szekely**: Resources. **Guojie Zhang**: Writing - Review & Editing, Supervision, Funding acquisition. **Qi Zhou**: Conceptualization, Writing - Original draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

## Conflict of interest

The authors declare that they have no competing interests.

## Acknowledgements

We also would like to thank 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, and Mariel L. Campbell and Ariel M. Gaffney from the Museum of Southwestern Biology, University of New Mexico, for providing bird DNA samples for this work. Q.Z. was supported by the National Natural Science Foundation of China (32061130208, 32170415), the Natural Science Foundation of Zhejiang Province (LD19C190001), and the European Research Council Starting Grant (grant agreement 677696). G.Z. was supported by 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).

## Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2021.06.013.

## References

Bachtrog, D., 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14, 113–124.
Bai, Y., Casola, C., Feschotte, C., Betrán, E., 2007. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. Genome Biol. 8, R11.
Baker, A.J., Haddrath, O., McPherson, J.D., Cloutier, A., 2014. Genomic support for a Moa–tinamou clade and adaptive morphological convergence in flightless ratites. Mol. Biol. Evol. 31, 1686–1696.
Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6, 11.
Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580.
Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319.
Berv, J.S., Field, D.J., 2018. Genomic signature of an avian lilliput effect across the K-Pg extinction. Syst. Biol. 67, 1–13.
Betrán, E., Thornton, K., Long, M., 2002. Retroposed new genes out of the X in *Drosophila*. Genome Res. 12, 1854–1859.
Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. Genome Res. 14, 988–995.
Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al., 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 14, 708–715.
Braun, E.L., Cracraft, J., Houde, P., 2019. Resolving the Avian Tree of Life from Top to Bottom: The Promise and Potential Boundaries of the Phylogenomic Era, in: Kraus RHS (ed). Avian Genomics in Ecology and Evolution. Springer, Cham, pp. 151–210.
Bromham, L., 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 366, 2503–2513.
Charlesworth, B., Charlesworth, D., 2000. The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 355, 1563–1572.
Cloutier, A., Sackton, T.B., Grayson, P., Clamp, M., Baker, A.J., Edwards, S.V., 2019. Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. Systematic. Biol. 68, 937–955.
Cockburn, A., 2006. Prevalence of different modes of parental care in birds. Proc. Biol. Sci. 273, 1375–1383.
Coddington, C., Cockburn, A., 1995. The mating system of free-living emus. Aust. J. Zool. 43, 365.
Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P.D., Grützner, F., Kaessmann, H., 2014. Origins and functional evolution of Y chromosomes across mammals. Nature 508, 488–493.
Cracraft, J., 1974. Phylogeny and evolution of the ratite birds. Ibis 116, 494–521.
DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Angel, G. del, Rivas, M.A., Hanna, M., et al., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.
Duc, D.L., Renaud, G., Krishnan, A., Almén, M.S., Huynen, L., Prohaska, S.J., Ongyerth, M., Bitarello, B.D., Schiöth, H.B., Hofreiter, M., et al., 2015. Kiwi genome provides insights into evolution of a nocturnal lifestyle. Genome Biol. 16, 147.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Emerson, J.J., Kaessmann, H., Betrán, E., Long, M., 2004. Extensive gene traffic on the mammalian X chromosome. Science 303, 537–540.
Graves, J.A.M., 2006. Sex chromosome specialization and degeneration in mammals. Cell 124, 901–914.
Grealy, A., Phillips, M., Miller, G., Gilbert, M.T.P., Rouillard, J.-M., Lambert, D., Bunce, M., Haile, J., 2017. Eggshell palaeogenomics: palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. Mol. Phylogenet. Evol. 109, 151–163.
Guioli, S., Lovell-Badge, R., Turner, J.M.A., 2012. Error-prone ZW pairing and No evidence for meiotic sex chromosome inactivation in the chicken germ line. PLoS Genet. 8, e1002560.
Handford, P., Mares, M.A., 1985. The mating systems of ratites and tinamous: an evolutionary perspective. Biol. J. Linn. Soc. 25, 77–104.
Harshman, J., Braun, E.L., Braun, M.J., Huddleston, C.J., Bowie, R.C.K., Chojnowski, J.L., Hackett, S.J., Han, K.-L., Kimball, R.T., Marks, B.D., et al., 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. Proc. Natl. Acad. Sci. U. S. A. 105, 13462–13467.
Hubisz, M.J., Pollard, K.S., Siepel, A., 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. Briefings Bioinf. 12, 41–51.
Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., et al., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346, 1320–1331.
Kapusta, A., Suh, A., Feschotte, C., 2017. Dynamics of genome size evolution in birds and mammals. Proc. Natl. Acad. Sci. U. S. A. 114, E1460–E1469.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360.
Kozlov, A.M., Aberer, A.J., Stamatakis, A., 2015. ExaML version 3: a tool for phylogenomic analyses on supercomputers. Bioinform. Oxf. Engl. 31, 2577–2579.
Krasovec, M., Chester, M., Ridout, K., Filatov, D.A., 2018. The mutation rate and the age of the sex chromosomes in Silene latifolia. Curr. Biol. 28, 1832–1838.
Lahn, B.T., Page, D.C., 1999. Four evolutionary strata on the human X chromosome. Science 286, 964–967.
Li, H., Coghlan, A., Ruan, J., Coin, L.J., Hériché, J.-K., Osmotherly, L., Li, R., Liu, T., Zhang, Z., Bolund, L., et al., 2006. TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res. 34, D572–D580.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760.
Li, W.-H., Yi, S., Makova, K., 2002. Male-driven evolution. Curr. Opin. Genet. Dev. 12, 650–656.
Lifschytz, E., Lindsley, D.L., 1972. The role of X-chromosome inactivation during spermatogenesis. Proc. Natl. Acad. Sci. U. S. A. 69, 182–186.
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Yunjie, et al., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience 1, 18.
Mank, J.E., Ellegren, H., 2007. Parallel divergence and degradation of the avian W sex chromosome. Trends Ecol. Evol. 22, 389–391.
McKee, B.D., Handel, M.A., 1993. Sex chromosomes, recombination, and chromatin conformation. Chromosoma 102, 71–80.
McNab, B.K., 1996. Metabolism and temperature regulation of kiwis (Apterygidae). The Auk 113, 687–692.
Meisel, R.P., Han, M.V., Hahn, M.W., 2009. A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. Genome. Biol. Evol. 1, 176–188.
Meng, G., Li, Y., Yang, C., Liu, S., 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 47, gkz173.
Mitchell, K.J., Llamas, B., Soubrier, J., Rawlence, N.J., Worthy, T.H., Wood, J., Lee, M.S.Y., Cooper, A., 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. Science 344, 898–900.
Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K., Yasunaga, T., 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold. Spring. Harb. Sym. 52, 863–867.
Nishida-Umehara, C., Tsuda, Y., Ishijima, J., Ando, J., Fujiwara, A., Matsuda, Y., Griffin, D.K., 2007. The molecular basis of chromosome orthologies and sex chromosomal differentiation in palaeognathous birds. Chromosome Res. 15, 721–734.
O'Connor, R.E., Farré, M., Joseph, S., Damas, J., Kiazim, L., Jennings, R., Bennett, S., Slack, E.A., Allanson, E., Larkin, D.M., et al., 2018. Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. Genome Biol. 19, 171.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Parker, J., Tsagkogeorga, G., Cotton, J.A., Liu, Y., Provero, P., Stupka, E., Rossiter, S.J., 2013. Genome-wide signatures of convergent evolution in echolocating mammals. Nature 502, 228–231.
Pigozzi, M.I., 2011. Diverse stages of sex-chromosome differentiation in tinamid birds: evidence from crossover analysis in *Eudromia elegans* and *Crypturellus tataupa*. Genetica 139, 771–777.
Potrzebowski, L., Vinckenbosch, N., Marques, A.C., Chalmel, F., Jégou, B., Kaessmann, H., 2008. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biol. 6, e80.
Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, 569–573.
Remeš, V., Freckleton, R.P., Tökölyi, J., Liker, A., Székely, T., 2015. The evolution of parental cooperation in birds. Proc. Natl. Acad. Sci. U. S. A. 112, 13603–13608.
Reynolds, J.D., Székely, T., 1997. The evolution of parental care in shorebirds: life histories, ecology, and sexual selection. Behav. Ecol. 8, 126–134.
Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J.S., Wheeler, N.E., Gardner, P.P., Clarke, J.A., Baker, A.J., Clamp, M., et al., 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. Science 364, 74–78.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinform. Oxf. Engl. 30, 1312–1313.
Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B., 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32, W309–W312.
Sturgill, D., Zhang, Y., Parisi, M., Oliver, B., 2007. Demasculinization of X chromosomes in the *Drosophila* genus. Nature 450, 238–241.
Szekely, T., 2014. Sexual conflict between parents: offspring desertion and asymmetrical parental care. Csh. Perspect. Biol. 6, a017665.
Takagi, N., Itoh, M., Sasaki, M., 1972. Chromosome studies in four species of Ratitae (Aves). Chromosoma 36, 281–291.
Tollis, M., Robbins, J., Webb, A.E., Kuderna, L.F.K., Caulin, A.F., Garcia, J.D., Bèrubè, M., Pourmand, N., Marques-Bonet, T., O'Connell, et al., 2019. Return to the sea, get huge, beat cancer: an analysis of cetacean genomes including an assembly for the humpback whale (*Megaptera novaeangliae*). Mol. Biol. Evol. 36, 1746–1763.
Toups, M.A., Pease, J.B., Hahn, M.W., 2011. No excess gene movement is detected off the avian or Lepidopteran Z chromosome. Genome. Biol. Evol. 3, 1381–1390.
Traut, W., Schubert, V., Daliková, M., Marec, F., Sahara, K., 2019. Activity and inactivity of moth sex chromosomes in somatic and meiotic cells. Chromosoma 128, 533–545.
Tullberg, B.S., Ah–King, M., Temrin, H., 2002. Phylogenetic reconstruction of parental–care systems in the ancestors of birds. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 357, 251–257.
Turner, J.M.A., 2007. Meiotic sex chromosome inactivation. Development 134, 1823–1831.
Varricchio, D.J., Moore, J.R., Erickson, G.M., Norell, M.A., Jackson, F.D., Borkowski, J.J., 2008. Avian paternal care had dinosaur origin. Science 322, 1826–1828.
Vibranovski, M.D., Lopes, H.F., Karr, T.L., Long, M., 2009a. Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. PLoS Genet. 5, e1000731.
Vibranovski, M.D., Zhang, Y., Long, M., 2009b. General gene movement off the X chromosome in the *Drosophila* genus. Genome Res. 19, 897–903.
Vicoso, B., Kaiser, V.B., Bachtrog, D., 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc. Natl. Acad. Sci. U. S. A. 110, 6453–6458.
Wang, J., Long, M., Vibranovski, M.D., 2012. Retrogenes moved out of the Z chromosome in the silkworm. J. Mol. Evol. 74, 113–126.
Wang, Z.-J., Chen, G.-J., Zhang, G.-J., Zhou, Q., 2021. Dynamic evolution of transposable elements, demographic history, and gene content of paleognathous birds. Zool. Res. 42, 51–61.
Wang, Y., Li, J., Paterson, A.H., 2013. MCScanX-transposed: detecting transposed gene duplications based on multiple colinearity scans. Bioinformatics 29, 1458–1460.
Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., et al., 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40, e49.
Wesolowski, T., 1994. On the origin of parental care and the early evolution of male and female parental roles in birds. Am. Nat. 143, 39–58.
Wu, C.-I., Xu, E.Y., 2003. Sexual antagonism and X inactivation – the SAXI hypothesis. Trends Genet. 19, 243–247.
Xu, L., Auer, G., Peona, V., Suh, A., Deng, Y., Feng, S., Zhang, G., Blom, M.P.K., Christidis, L., Prost, S., Irestedt, et al., 2019a. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. Nat. Ecol. Evol. 3, 834–844.
Xu, L., Sin, S.Y.W., Grayson, P., Edwards, S.V., Sackton, T.B., 2019b. Evolutionary dynamics of sex chromosomes of paleognathous birds. Genome. Biol. Evol. 11, 2376–2390.
Xu, L., Zhou, Q., 2020. The female-specific W chromosomes of birds have conserved gene contents but are not feminized. Genes (Basel) 11, 1126.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yazdi, H.P., Ellegren, H., 2018. A genetic map of ostrich Z chromosome and the role of inversions in avian sex chromosome evolution. Genome. Biol. Evol. 10, 2049–2060.
Yonezawa, T., Segawa, T., Mori, H., Campos, P.F., Hongoh, Y., Endo, H., Akiyoshi, A., Kohno, N., Nishida, S., Wu, J., Jin, H., Adachi, et al., 2017. Phylogenomics and morphology of extinct paleognaths reveal the origin and evolution of the ratites. Curr. Biol. 27, 68–77.
Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W., et al., 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346, 1311–1320.
Zhang, J., Li, C., Zhou, Q., Zhang, G., 2015. Improving the ostrich genome assembly using optical mapping data. GigaScience 4, 24.
Zhang, Y.E., Vibranovski, M.D., Landback, P., Marais, G.A.B., Long, M., 2010. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. PLoS Biol. 8, e1000494.
Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert, M.T.P., Zhang, G., 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 346, 1246338.