ARTICLES
https://doi.org/10.1038/s41559-019-0850-1
nature ecology & evolution

Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds

Luohao Xu[1,2], Gabriel Auer[2], Valentina Peona[3], Alexander Suh[3], Yuan Deng[4,5], Shaohong Feng[4,5], Guojie Zhang[4,6,7], Mozes P. K. Blom[8,9], Les Christidis[10,11], Stefan Prost[12,13], Martin Irestedt[8*] and Qi Zhou[1,2*]

Songbirds have a species number close to that of mammals and are classic models for studying speciation and sexual selection. Sex chromosomes are hotspots of both processes, yet their evolutionary history in songbirds remains unclear. We characterized genomes of 11 songbird species, with 5 genomes of bird-of-paradise species. We conclude that songbird sex chromosomes have undergone four periods of recombination suppression before species radiation, producing a gradient of pairwise sequence divergence termed 'evolutionary strata'. The latest stratum was probably due to a songbird-specific burst of retrotransposon CR1–E1 elements at its boundary, instead of the chromosome inversion generally assumed for suppressing sex-linked recombination. The formation of evolutionary strata has reshaped the genomic architecture of both sex chromosomes. We find stepwise variations of Z-linked inversions, repeat and guanine–cytosine (GC) contents, as well as W-linked gene loss rate associated with the age of strata. A few W-linked genes have been preserved for their essential functions, indicated by higher and broader expression of lizard orthologues compared with those of other sex-linked genes. We also find a different degree of accelerated evolution of Z-linked genes versus autosomal genes among species, potentially reflecting diversified intensity of sexual selection. Our results uncover the dynamic evolutionary history of songbird sex chromosomes and provide insights into the mechanisms of recombination suppression.

Songbirds (Oscines, suborder Passeri) have over 5,000 species and comprise most passerines and nearly half of all extant bird species[1]. This is because of the largest avian species radiation that occurred about 60 million years (Myr) ago[2]. With the development of genomics, many species besides zebra finch are now becoming important models for studying molecular patterns and mechanisms of speciation[3,4], supergene[5] or cognition[6], out of their long history of ecological or behavioural studies. One major reason for biologists' interest in songbirds is their diversified sexual traits. For example, their ostentatious plumage forms and colours, sophisticated songs and mating rituals, all of which can undergo rapid turnovers even between sister species. Theories predict that sex chromosomes play a disproportionately large role in speciation (the 'large X/Z' effect), sexual selection and evolution of sexually dimorphic traits[7–9]. However, the evolutionary history of songbirds' sex chromosome remains unclear because there were few genomic studies characterizing songbirds' sex chromosomes except for collared flycatcher[10]. Unlike the mammalian XY system, birds have independently evolved a pair of female heterogametic sex chromosomes that are usually heteromorphic in females (ZW) and homomorphic in males (ZZ). A recent cytological investigation of over 400 passerine species found a higher fixation rate of chromosomal inversions on the Z chromosome than autosomes within species, so that gene flow is probably more reduced by hybridization[11,12]. A significantly lower level of introgression in Z-linked genes compared to autosomal genes has been reported from studying pairs of recently diverged songbird species[13–15]. Such a large-Z pattern is probably contributed by several factors that act in an opposite manner in the XY sex system. First, Z chromosomes are more often transmitted in males, thus are expected to have a higher mutation rate than the rest of the genome, due to the 'male-driven evolution' effect[16]. Previous studies[17–19] showed this effect is less pronounced in birds than in mammals, thus the contribution of 'male-driven evolution' to the large-Z pattern may be limited. Second, as sexual selection more frequently targets males, the variation in male reproductive success will further reduce the effective population size of Z chromosomes from three-quarters that of autosomes[20]. The consequential genetic drift effect is expected to fix excessive slightly deleterious mutations on the Z chromosome and lead to its faster evolutionary rate than autosomes (the 'fast-Z' effect)[21]. This has been demonstrated in Galloanserae species (for example, chicken and duck), of which those undergoing stronger sperm competition, which is a more intensive male sexual selection, exhibit a larger difference between Z chromosome and autosomes in their evolution rates[22].

In contrast to the avian Z chromosomes, or more broadly the mammalian XY chromosomes, genomic studies of avian W chromosomes, especially those of songbirds, have only recently been performed[10,23,24]. This is because most genomic projects prefer to

[1]MOE Laboratory of Biosystems Homeostasis & Protection, Life Sciences Institute, Zhejiang University, Hangzhou, China. [2]Department of Molecular Evolution and Development, University of Vienna, Vienna, Austria. [3]Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden. [4]China National Genebank, BGI-Shenzhen, Shenzhen, China. [5]BGI-Shenzhen, Shenzhen, China. [6]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [7]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [8]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden. [9]Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung, Berlin, Germany. [10]National Marin Science Centre, Southern Cross University, Coffs Harbour, New South Wales, Australia. [11]School of BioSciences, University of Melbourne, Parkville, Victoria, Australia. [12]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA. [13]LOEWE-Center for Translational Biodiversity Genomics, Senckenberg, Frankfurt, Germany. *e-mail: Martin.Irestedt@nrm.se; zhouqi1982@zju.edu.cn

choose the homogametic sex (for example, male birds or female mammals) for sequencing, to avoid the presumably gene-poor and highly repetitive Y or W chromosomes. It has been suggested but not yet experimentally shown that the Y/W chromosomes have undergone suppressions of recombination to prevent the sex-determining gene or sexual antagonistic (beneficial to one sex but detrimental to the other) genes being transmitted in the opposite sex[25]. The loss of recombination reduces the efficacy of natural selection and drives the ultimate genetic decay of non-recombining regions of Y/W chromosomes due to the effect of for example, 'Hill–Robertson interference' between linked loci[26]. The degeneration process can be accelerated by selective sweep targeting male-related genes on the Y chromosome, or by background selection, purging the deleterious mutations from highly dosage-sensitive genes[27]. Simulation showed that they play a different role at different stages of Y/W degeneration[28]. Both processes have gained evidence from the analyses of mammalian[29,30] and *Drosophila*[31,32] Y-linked genes. Although purifying selection acting on dosage-sensitive genes has been implicated to maintain the few W-linked genes retained in Galloanserae (for example, chicken and duck)[24,33] or flycatcher[10], little evidence has been found for female-specific positive selection acting on W-linked genes (but see ref. [34]).

In both birds[23] and mammals[35], as well as several plant species such as *Silene latifolia*[36], recombination suppressions have all proceeded in a stepwise manner presumably through chromosomal inversions, leaving a stratified pattern of sequence divergence between sex chromosomes termed 'evolutionary strata'. Eutherian mammalian X and Y chromosomes have been inferred to share at least three strata, with another two more recent ones shared only among catarrhines (old world monkeys and great apes)[30]. We recently discovered from a broad but sparse sampling of diverse bird genomes that the history and tempo of avian sex chromosome evolution are much more complicated than those of mammals[23]. We showed that all birds' sex chromosomes only share the first time of recombination suppression (stratum 0, Aves S0) encompassing the avian male-determining gene *DMRT1*. This was followed by the independent formation of S1 in different basal Palaeognathae species (for example, ratites and tinamou) and in the ancestor of Neognathae (for example, chicken and zebra finch). Ratites have halted any further recombination loss and maintained over two-thirds of the entire sex chromosome pair as the exceptionally long recombining pseudoautosomal regions (PAR). Therefore, their sex chromosomes are homomorphic and gene-rich on the W chromosome. All Neognathae species have suppressed recombination throughout most regions of the sex chromosomes with short and varying sizes of PAR (ref. [37]). However, overall, avian W chromosomes seem to have retained more genes and decayed at a slower rate than the mammalian Y chromosomes. Moreover, sexually monomorphic species (for example, most ratites) seem to differentiate more slowly than sexually dimorphic species (chicken and many Neoaves species) in their sex chromosomes, consistent with the hypothesis that sexual antagonistic alleles have triggered the expansion of recombination suppression between sex chromosomes[38]. However, ratites have a deep divergence from other birds and also a much lower mutation rate as expected from their larger body size. These confounding factors make the actual influence of sexual selection on the rate of sex chromosome evolution unclear. The principal group of Neognathae, Neoaves share one stratum S2, with the more recent history of sex chromosomes of songbirds being unclear. So far, only one songbird (collared flycatcher), has been characterized for its W-linked genes[10], in the range 46–90 reported W-linked genes of other Neoaves species. To explain the evolutionary history of songbirds' sex chromosomes, we produced high-quality female genomes of five bird-of-paradise (BOP) species. Together with six other published female genomes of songbird species, our analyses covered main songbird lineages (Corvida and Passerida) that diverged in the last 50 Myr (refs. [2,39]).

Results

Characterization of songbird sex chromosome sequences. We produced 36- to 150-fold genomic coverage of sequencing data for each BOP species and performed de novo genome assembly followed by chromosome mapping using the genomes of highly continuous or closely related great tit or hooded crow as reference[6]. The high continuity and completeness of the draft genomes are revealed by their scaffold N50 lengths (all longer than 3 Mb, except for Raggiana BOP) and BUSCO scores (92.9–94.0%; Supplementary Table 1). To reconstruct the evolutionary history of sampled songbirds' sex chromosomes, we first identified sequences from putative PARs by their homology to the published PAR sequence of collared flycatcher[40] and confirmed them by their similar read depth level to that of autosomes. Sequences from sexually differentiated regions (SDR) were identified as those that show half the female sequencing depth of autosomes (Fig. 1a and Supplementary Fig. 1). We then separated the Z- and W-linked sequences with the expectation that the latter would diverge much faster than the former from the reference Z chromosome sequence (Methods), and we further confirmed the W-linkage with a clear female-specific pattern in all but one species with sequencing data of both sexes (Fig. 1a and Supplementary Fig. 2). Our method cannot identify recent fusion/translocation of autosomal fragments to the sex chromosome pair (forming so-called 'neo-sex' chromosome), as in some warblers[41]. All the studied songbirds have a short putative PAR ranging from 564 to 781 kilobases (kb). The assembled lengths of the largely euchromatic parts of W chromosomes range from 1.33 to 7.24 megabases (Mb), corresponding to only 1.9–8.5% of the Z chromosome length across species (Fig. 1b and Supplementary Table 2), probably as a result of large deletions and massive invasions of repetitive elements. Indeed, the repeat content of the assembled W chromosomes is 2.5- to 4.9-fold higher than that of Z chromosomes on the chromosome-wide average (Supplementary Fig. 3 and Table 2).

Age-dependent genomic impact of evolutionary strata. If recombination was suppressed between sex chromosomes in a stepwise manner, we expect a gradient of Z/W sequence divergence levels along the Z chromosome, such as has been reported along the human X chromosome[42]. However, we have previously showed that the extant synteny of Neognathae Z chromosomes is misleading for inferring evolutionary strata, due to the marked intrachromosomal rearrangements[23]. By contrast, Palaeognathae species (for example, emu and ostrich) have maintained a highly conserved sequence synteny even with reptile species, with over two-thirds of their sex-linked regions still recombining as an approximate of the proto-sex chromosomes of all bird species[23,43]. We are able to identify the reshuffled fragments of the first and the second strata (S0 and S1) shared by all Neognathae species in the studied songbird genomes by their homology to the emu genome. They were mapped as two continuous regions on the emu Z chromosome (Fig. 2a ; Supplementary Figs. 4 and 5). Two recently formed strata (Neoaves S2 and S3) are much more conserved for their synteny across avian species and each shows a significantly different level of Z/W sequence divergence (Fig. 2b and Supplementary Fig. 6), GC3 (GC content at the third codon positions; Supplementary Fig. 7) and Z-linked long terminal repeat (LTR) content (Fig. 2c and Supplementary Fig. 7) from each other. The marked change of Z/W divergence level allows us to precisely map the boundaries between those two strata. In general, series of recombination suppressions have reshaped the genomic architecture of the Z chromosome in chronological order. Regions of younger strata exhibit much less Z-linked intrachromosomal rearrangements between species, suggesting the reduced selective constraints on gene synteny after recombination was suppressed in the older strata[44]. Alternatively, it could also reflect a neutral process that older strata have fixed more genomic rearrangements, as genetic drift has been acting for longer

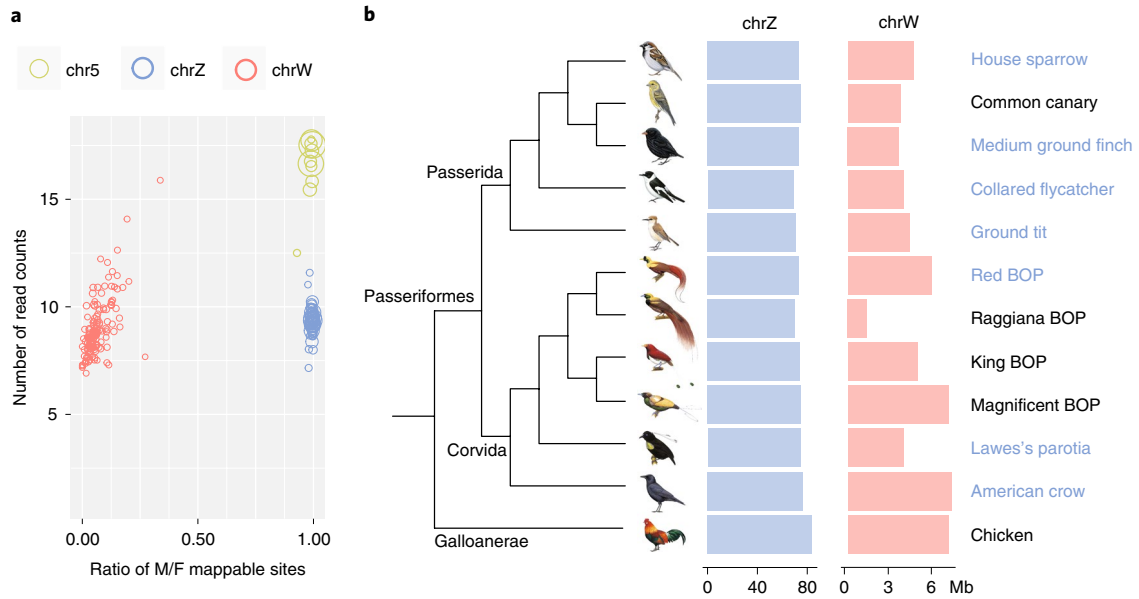

Fig. 1 | The Z and W chromosomes of different songbirds. a, We use medium ground finch as an example to demonstrate our identification and verification of sex-linked sequences. For each scaffold shown as a circle with scaled size to its length, the ratio of nucleotide sites that were mapped by male (M) versus female (F) genomic reads is plotted against the sequencing depth of this scaffold. Scaffold sequences are clustered separately by their derived chromosomes, with W-linked (red circles) and Z-linked (blue circles) sequences showing the expected half the autosome (green) sequencing depth, and W-linked sequences showing almost no mappable sites from male reads. **b**, The lengths of Z and W chromosomes across the studied songbird species. The data of chicken and collared flycatcher are derived from refs. [10,24]. The shorter length of Raggiana BOP W chromosome is probably caused by the low sequencing coverage. Species with Illumina reads of both sexes available are marked in blue. All bird illustrations were ordered from https://www.hbw.com/; ref. [91].

in these regions due to the reduced effective population size. In particular, GC3 content decreases, while the repeat content increases by age of stratum. This is probably because weaker effects of GC-biased gene conversion (gBGC; ref. [45]) and purifying selection against transposable element (TE) insertions[46] have been acting for longer in Z-linked regions of older strata with reduced recombination. Consistently, a similar pattern has also been found contrasting PAR versus the rest Z-linked regions in collared flycatcher[40].

Lineage-specific burst of retrotransposon probably has induced recombination suppression between sex chromosomes. The distribution of long interspersed elements (LINEs), mainly the retrotransposon chicken repeat 1 (CR1) elements, shows an exceptional pattern compared to that of LTR elements (Fig. 2c and Supplementary Fig. 7). The abundance of CR1 is unexpectedly similar in S3 and S0, and much higher than that of the rest of the Z-linked regions. A close examination shows that this is due to the specific accumulation of CR1 spanning the boundary between PAR and S3. Such a burst of CR1, particularly the CR1–E1 subfamily[47], is shared by all the investigated songbirds but absent in the basal passerine rifleman and other Neoaves species. It extends with gradual reduction into about one-third of the entire S3 region (Fig. 2d and Supplementary Fig. 8). The peak region of CR1 accumulation is associated with a large deletion (about 1.5 Mb) in passerines that removes a gene *DCC* (Deleted in Colorectal Carcinoma) highly conserved across other vertebrates[48]. This gene is responsible for axon guidance for brain midline crossing and has been independently lost in some but not all passerines and Galliformes[49].

In addition, the burst of CR1–E1 element seems to coincide with S3 emergence. Almost all the investigated genomes of songbirds have about two-fold more CR1–E1 elements than that of rifleman (Supplementary Table 3). Our phylogenetic reconstruction of Z- and W-linked gametologue sequences shows that only songbird-derived sequences are always grouped by chromosome instead of by species (Fig. 2e and Supplementary Figs. 9–12). This indicates that all songbirds share four evolutionary strata, with the latest S3 formed at the same time with the genome-wide expansion of CR1–E1 elements, after the divergence between all the songbirds and other passerine species. The highly conserved Z-linked synteny of S3 between songbird species and between songbirds and chicken (Fig. 2a and Supplementary Fig. 5) suggests that there was no Z-linked chromosomal inversion at S3. It is likely that the recent burst of CR1–E1 subfamily elements have led to the formation of S3, although we cannot exclude the contribution of W-linked chromosomal inversions. Interestingly, other CR1 subfamilies CR1–E(4–6) have an independent burst both genome-wide and specifically at the PAR/S3 boundary in rifleman (Fig. 2d, Supplementary Fig. 8 and Table 3). Given this boundary region has been shown to have frequent but different degrees of multiple gene loss in different lineages of birds[48,49], it is probably a hotspot for mutations or LINE insertions that have recurrently contributed to the independent formation of S3 in many bird species.

Fast-Z pattern of songbirds suggest their dynamic evolution of sexual selection. The formation of evolutionary strata has subjected the Z chromosome to male-biased transmission and a reduced effective population size, which are expected to produce faster mutation and evolution rates of Z-linked genes, respectively[20]. We found a larger branch-specific synonymous substitution rate (d*S*) of Z-linked genes (statistically not significant) but a significantly smaller d*S* of W-linked genes, compared to that of autosomal genes ($P=0.002165$, Wilcoxon rank sum test; Supplementary Fig. 13), as a result of male-driven evolution[16]. The branch-specific evolution rates (ω) measured by the ratios of non-synonymous substitution rates (d*N*) over d*S* have significantly ($P<0.003$, Wilcoxon rank sum test; Supplementary Fig. 14) increased for both Z- and W-linked gametologues relative to autosomal genes, indicating a 'fast-Z' effect and degeneration of W-linked genes (see below). Previous simulation

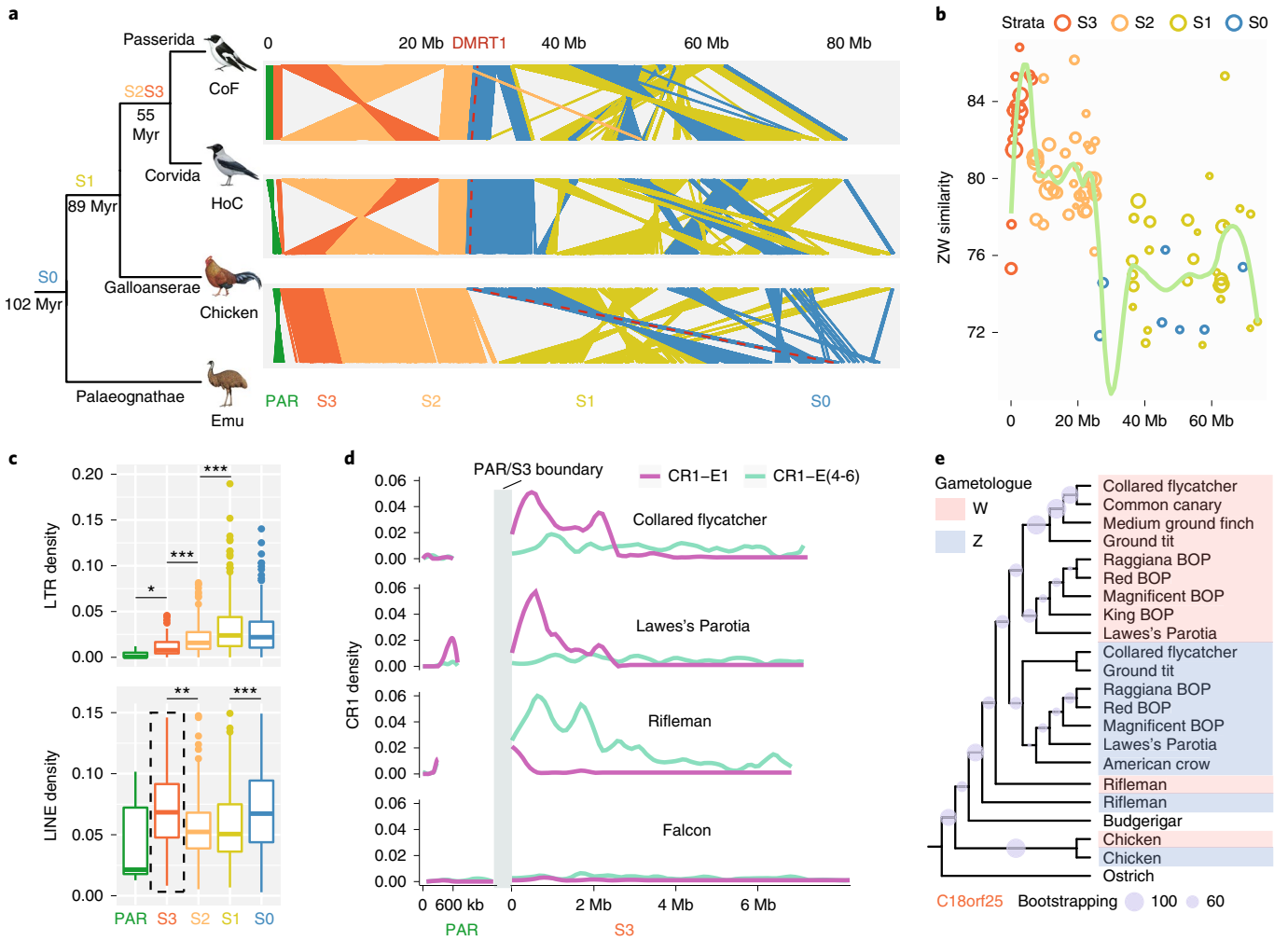

Fig. 2 | Evolutionary strata of songbirds. a, We use chromosomal or nearly chromosomal genome assemblies of four representative bird species (CoF, collared flycatcher; HoC, hooded crow; chicken; and emu) to show their rearrangements on the Z chromosome. Each line represents one pair of aligned fragments between two species, and each colour corresponds to one evolutionary stratum of songbirds. The location of *DMRT1*, the avian male-determining gene, is marked by the red dashed line. On the phylogenetic tree, we also indicate the evolutionary strata at their respective node of origination. Generally, the synteny is more conserved in younger strata between species. **b**, We use Lawes's Parotia as an example to demonstrate the pairwise sequence similarity pattern of evolutionary strata. The size of circles is scaled to the length of sequence alignments between Z/W chromosomes. **c**, Transposable elements (LINEs and LTRs) have accumulated strongly in older strata (S0 is the first stratum), except for LINE at S3. In the boxplots, the horizontal line shows the medium value, the whiskers show the 25 and 75% quartile values of the density of TEs (percentage of TE sequences in every 100 kb window). $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$. **d**, Lineage-specific burst of CR1–E1 (a subtype of CR1 LINE element, purple line) at the boundary of the PAR and S3 in songbirds, since their divergence with other passerine species. Other subtypes of CR1 elements are also plotted with the green line for comparison. **e**, Phylogenetic tree using Z- and W-linked gametologue sequences of the gene *C18orf25* located at S3. Lineages are clustered by chromosomes (red or blue), not by species, suggesting S3 independently formed in rifleman, chicken and the ancestor of songbirds. All bird illustrations were ordered from https://www.hbw.com/; ref. [91].

work and experimental evidence in Galloanserae have suggested that different degrees of sexual selection targeting males will influence the male-mating success, hence the genetic drift effect on the Z chromosome to a different degree[20,22]. Songbirds, especially BOPs have been frequently used as a textbook example of sexual selection[50,51]; however, their evolution history of sexual selection remained unclear. To reconstruct that, we approximated the intensity of sexual selection targeting males by measuring the degree of fast-Z effect (Z/A value, the ratio of branch-specific ω values of Z-linked genes versus autosomal genes) in a phylogenetic context (Fig. 3 and Supplementary Table 4). The varying Z/A values at different lineages suggest a dynamic change of intensity of sexual selection, even among the five BOP species that diverged in 15 Myr (ref. [51]). A social mating system has previously been shown to influence the degree of sexual selection in birds[52] but we did not find a significant (Wilcoxon rank sum test, $P > 0.05$) difference of Z/A values between the monogamous species versus polygynous species, probably because of the few species used for comparison here. While the significant (permutation test, two-sided $P < 0.05$) fast-Z pattern of the sexually monochromatic American crow may reflect the sexual selection acting on the ancestral lineage leading to Corvidae species, a lack of such a pattern in Raggiana and magnificent BOP species is unexpected. These species are known for their lekking behaviours[50,53], with which few males dominate almost all females for copulation through out-competing other males. This produces a strongly biased male-mating success and direct challenge for maintaining genetic variation in the population (the lekking paradox)[54]. Few field quantitative studies have been performed on BOP species; it will be interesting to investigate whether Raggiana and magnificent BOP female

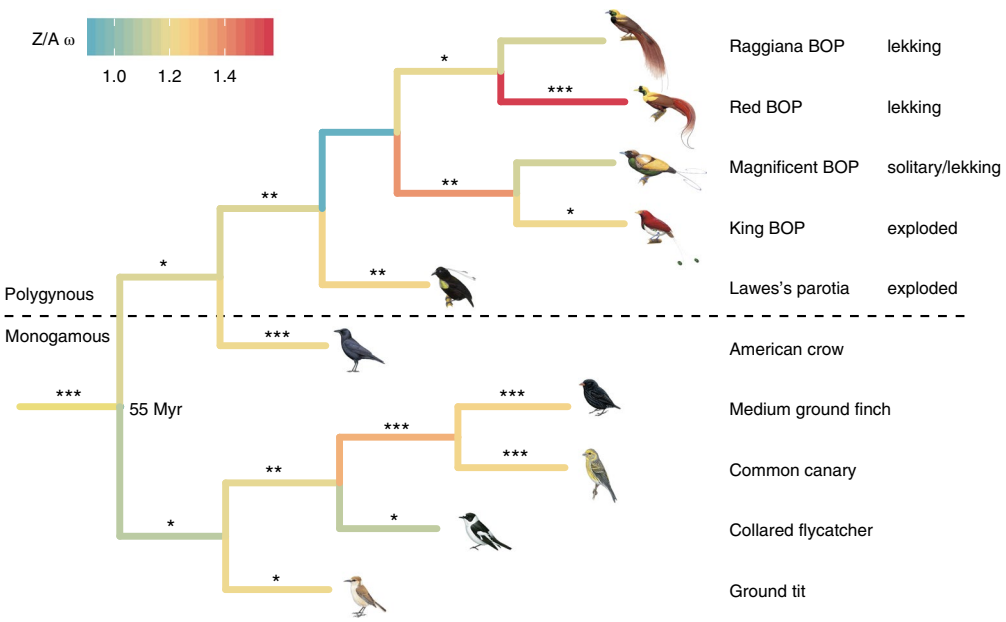

Fig. 3 | Fast-Z evolution of songbirds. Difference of evolution rates are shown for Z-linked genes versus autosome genes (Z/A value) as a measurement of fast-Z effect throughout the lineages of studied songbird species. There are on average 813 Z-linked genes. Genes from chromosome 4 and 5 are used to represent autosomal genes. Chromosome-wide *dN*/*dS* values are compared between chromosome Z and autosomes. The tree length and colour are scaled to the Z/A value, with lineages that show a significant (permutation test, re-sampled 1,000 times, two-sided $P<0.05$) fast-Z pattern labelled with asterisks. ***$P<0.001$, **$P<0.01$, *$P<0.05$. Social mating systems ('monogamous' versus 'polygamous') and male display type[51] ('lekking', 'exploded lekking', 'solitary display') are labelled. All bird illustrations were ordered from https://www.hbw.com/; ref. [91].

individuals solve the lekking paradox by changing mating preference and mate with more males than presumed.

Conserved gene content of the songbird W chromosomes. In contrast to the dynamic evolution of Z-linked genes and sequences, W chromosomes of all the studied songbirds have undergone marked gene loss but exhibit an unexpected conservation of the retained gene repertoire across species. The numbers of assembled W-linked genes range from 31 in house sparrow to 63 in the king BOP, compared to about 600–800 Z-linked genes (Fig. 4a and Supplementary Tables 2 and 7). These numbers are probably an underestimate because genes embedded in the highly repetitive regions may be missing from the current W chromosome assemblies. In general, Corvida species have retained more W-linked genes than Passerida species (Supplementary Table 5), probably due to their longer generation time thus lower mutation rate. Most W-linked genes are single-copy without lineage-specific expansion, except for *HINT1W* (Supplementary Fig. 15). Despite rare occasions of gene retroposition in birds[55], we find one W-linked gene that is derived through retroposition from an autosomal gene *NARF* in American crow (Supplementary Fig. 16). It will be interesting to investigate whether this gene shows signatures of female-specific selection, for example, a new pattern of ovary-specific expression, which drives its fixation on the W chromosome. Fifty-seven genes are shared by at least one Corvida and another Passerida species and 23 genes are shared between at least one songbird species and chicken[24]. This suggests they were present on the W chromosome before the divergence of passerine or Neognathae species. Despite the independent origination of S2 in chicken and Neoaves[23], all the chicken W-linked genes but one are also found in passerines, indicating similar underlying evolutionary forces governing their convergent retention since Galloanserae and Neoaves diverged from each other 89 Myr ago.

To examine such forces, we performed gene ontology analyses on the 79 genes that are present on the W chromosome of at least one songbird species. They are enriched ($P<0.01$, Fisher's exact test) for two gene ontology terms of 'DNA binding' and 'transcription factor activity, sequence-specific DNA binding' (Supplementary Table 6). This indicates that, similar to the mammalian Y-linked genes[30], some W-linked genes are retained for their important functions of regulating gene activities elsewhere in the genome. The Z-linked homologues of lost genes evolve significantly faster ($P=0.002165$, Wilcoxon rank sum test) with ω ratios higher than those of the retained genes on the W chromosome (Fig. 4b and Supplementary Fig. 17). This shows a different selective pressure acting on these two sets of genes on the proto-sex chromosomes. As this pattern maybe confounded by the 'faster-Z' effect of hemizygous Z-linked genes, we studied the autosomal orthologues of these genes in green anole lizard. We found that the lizard orthologues of retained genes have significantly higher ($P<1.497\times10^{-5}$, Wilcoxon rank sum test; Fig. 4c) expression levels in all tissues of both males and females, and also a broader expression pattern than those of the lost genes across all the tissues (Fig. 4d). The patterns are consistent among the four songbird evolutionary strata; or if we use emu to infer the ancestral expression pattern (Supplementary Fig. 18), whose sex chromosomes are largely PAR. In addition, the retained genes on the W chromosomes are more likely to be dosage-sensitive than those that have become lost. This is indicated by their significantly higher (Wilcoxon rank sum test, $P<0.001$) predicted haplo-insufficiency scores for the human orthologues[56] of the former than those of the latter (Supplementary Fig. 19). This is consistent with the patterns found for chicken or mammalian W- or Y-linked genes[24,30]. We have not found an excess of ovary-biased lizard orthologues among those of the retained W-linked genes: only 6 out of 72 (8.3%) are ovary-biased while the genome-wide proportion is about 20%. This suggests that female-specific selection may not play an important role in preventing the gene loss, or that certain genes undergo positive selection on the songbirds' W chromosomes, which is consistent with the result of collared flycatcher[10].

Comparing gene loss between avian W chromosomes and mammalian Y chromosomes. Overall, 4.6–9.2% of songbird single-copy W-linked genes, compared to 1.6–3.0% mammalian single-copy

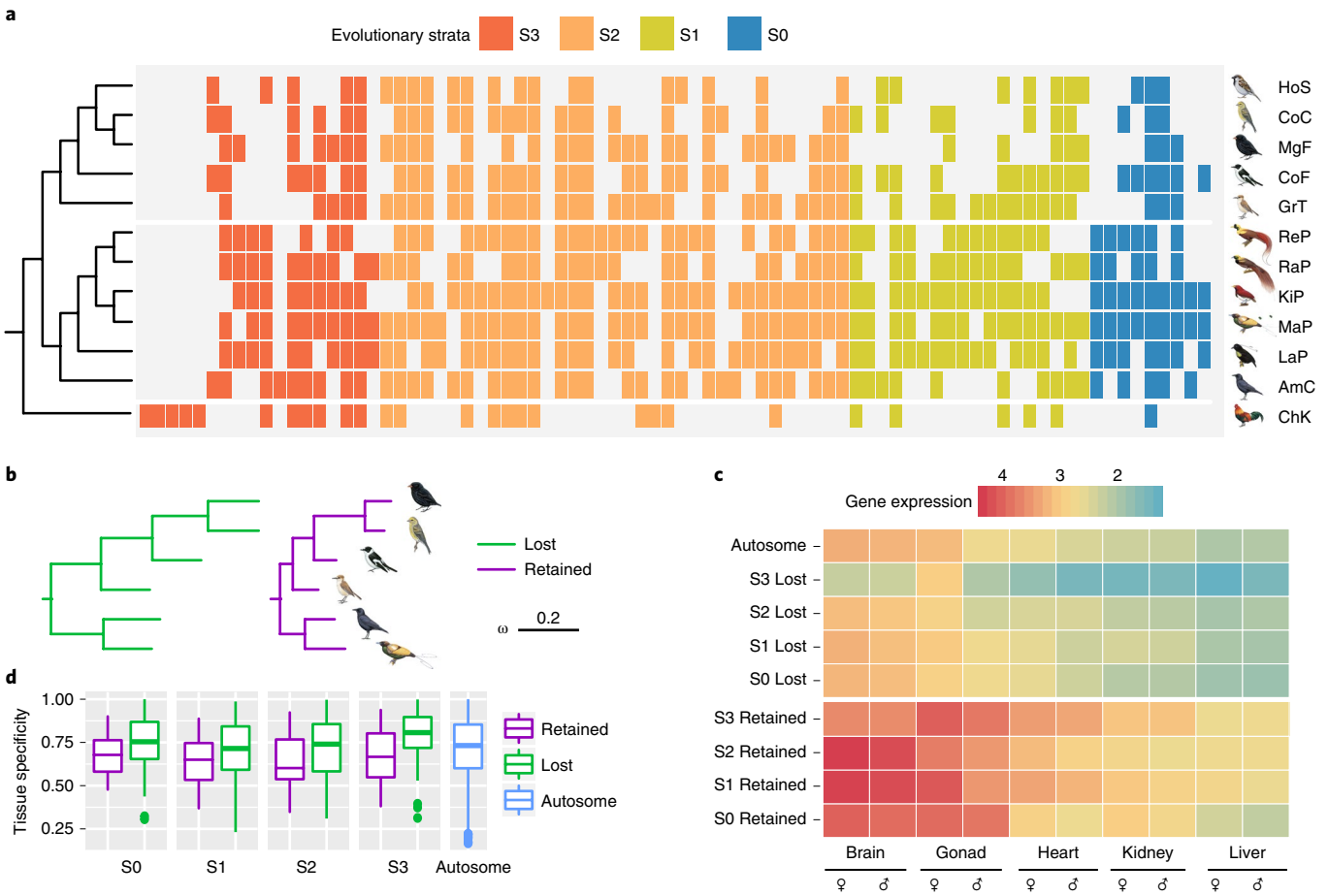

Fig. 4 | The W-linked genes are preserved by purifying selection. a, The retained W-linked genes of each studied songbird species (the species common names are shown with a three-letter code corresponding to those in Fig. 1), as well as those of chicken, with homologous genes aligned vertically. The order of genes follows that of their emu homologues along the Z chromosome. The colours represent the evolutionary strata of songbirds. Species abbreviations: HoS, house sparrow; CoC, Atlantic Canary; MgF, medium ground finch; CoF, collared flycatcher; GrT, ground tit; ReP, red BOP; RaP, Raggiana BOP; KiP, king BOP; MaP, magnificent BOP; LaP, Lawes's parotia; AmC, American crow; ChK, chicken. **b**, The Z-linked genes without W-linked homologues (green, 'Lost') evolve faster than those with W-linked homologues retained (red, 'Retained'), as indicated by their branch lengths scaled to *dN*/*dS* ratios. **c**, The Z-linked genes whose W-linked homologues have become lost (upper panel) tend to have a higher expression level (measured by TPM) in their lizard orthologues than those with W-linked homologues retained (lower panel). The genes are divided further by their resided stratum and the expression level is shown by log-transformed medium expression values of each category as colour-coded heatmap. **d**, Gene expression tissue specificity in green anole lizard for the homologous avian Z-linked genes. In the boxplots, the horizontal line shows the medium value, the whiskers show the 25% and 75% quartile values of expression tissue specificity. All bird illustrations were ordered from https://www.hbw.com/; ref. [91].

Y-linked genes[30] have been retained for their essential or sex-specific functions. A higher retention ratio of W-linked genes in birds can be partially attributed to the generally much lower mutation rate of W chromosome relative to Y chromosome by male-driven evolution effect (Supplementary Fig. 13), assuming a similar generation time between mammals and birds. In addition, a more frequent and stronger sex-specific selection acting on the Y chromosome than on the W chromosome, sometimes driving the massive expansion of Y-linked gene copies with male-related function[29], probably also contributed to a faster rate of Y chromosome gene loss by hitchhiking effect. To examine the tempo of gene loss throughout the evolution of songbirds sex chromosomes, we conservatively reconstructed the numbers of retained W-linked gametologues at each phylogenetic node of avian tree (Fig. 5a and Supplementary Table 8) by identifying the genes present on any of the studied avian W chromosomes. We found that in each stratum, the percentage of gene loss is always much larger at an earlier evolutionary time point than the recent ones and this is consistent between birds and mammals (Fig. 5b). Thus, most gene loss probably occurred during the early stages of recombination suppression, and the rate of gene loss markedly decreases by the less retained genes. Although convergent gene loss may cause an overestimate of lost genes at more ancestral time points (for example, in S0 region), this probably has little influence on our estimate in the most recent songbird-specific stratum S3 which has already lost 69.8% of the W-gametologues in 50 Myr. We also found that the retained genes of songbird W chromosomes are often close to each other (Supplementary Fig. 20), indicating large sequence deletions have contributed to marked gene loss.

The decrease of gene loss rate on Y/W chromosomes over evolutionary time can be explained by a weaker Hill–Robertson effect that the less retained genes can induce, which has been previously shown by simulation study[28]. In addition, the size, that is, the ancestral gene number of older evolutionary strata which would have undergone more serious gene loss must have a larger influence on the extant number of retained genes. Thus, a lower rate of retained mammalian Y-linked genes relative to avian W-linked genes can be attributed to the fact that the first two or three mammalian evolutionary strata occurred before the divergence of eutherians; together

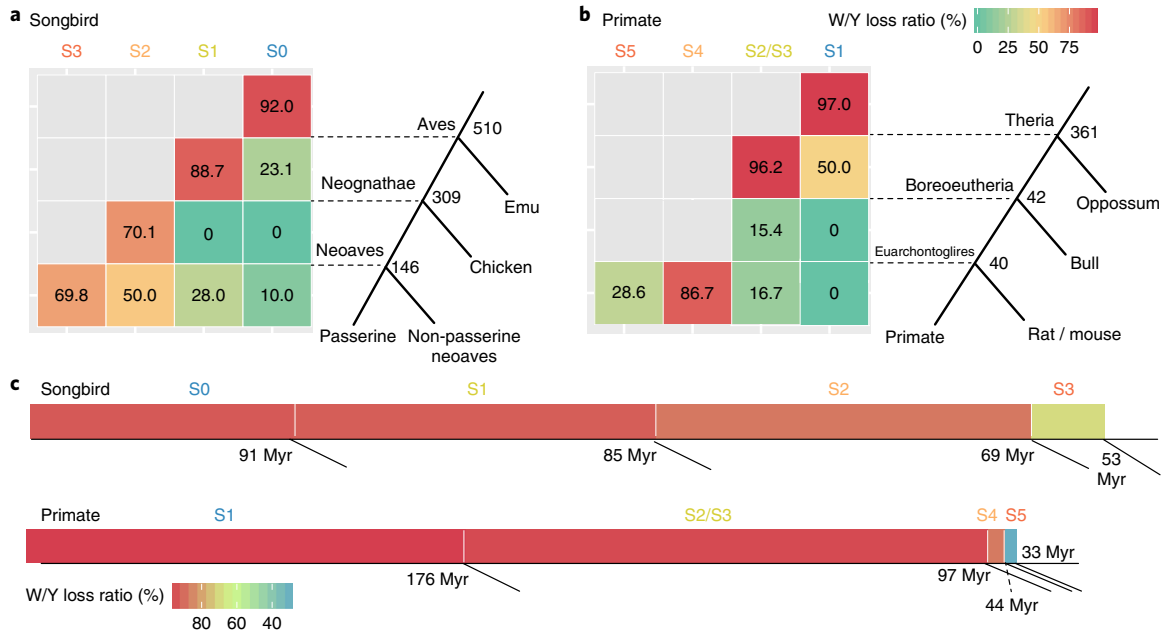

Fig. 5 | Comparison of gene loss between W chromosomes of songbirds and Y chromosomes of primates. a, Percentage of gene loss and ancestral gene number for each evolution stratum at each phylogenetic node. **b**, Similar analyses for the Y-linked gametologues of primates on the basis of the data of ref. [24], with S1 as the first stratum of eutherian mammals. **c**, Lengths of songbird W or primate Y chromosomes scaled to the ancestral gene number of each evolutionary stratum, with the colour scaled to the overall percentage of gene loss. The ages of evolutionary strata are indicated by the numbers (in million years) at the nodes below the bars. As eutherian mammals have much larger ancestral evolutionary strata than those of birds, they probably have more gene loss on the Y chromosome.

these account for over 93.2% of the entire gene content of ancestral Y chromosome, while those of birds only account for 53.3% of the entire ancestral W-linked gene content (Fig. 5c).

Discussion

The evolution of sex chromosomes is often but not always (for example, in frogs[57], ratites[23] and python[58]) marked with episodes of recombination suppressions that eventually restrict the recombining region in one or two small PARs at the end of chromosome. The resulting patterns of evolutionary strata have been widely reported in many animal and plant species, with the formation mechanism presumed to be chromosomal inversions[59]. Indeed, footprints of inversions in the latest two strata between human X and Y chromosomes have been found by examining the synteny order between X/Y, and particularly the homologous X-linked PAR boundary (PAB) sequence on the Y chromosome that has been disrupted into two dispersed sequences[60]. The Y-linked PAB is defined by an insertion of *Alu* element[61], with similar insertions of various types of TE elements at PAB reported among other mammals such as cattle and pig (reviewed in ref. [62]). Such TE insertions were probably due to the reduction of recombination rate at PAB, after chromosome inversion suppressed the recombination in the youngest stratum bordering the PAR. In the case of birds, a W-linked chromosome inversion giving birth to a young evolutionary stratum would be fixed within the population, given its advantage of linking sexual antagonistic genes (beneficial to female but detrimental to male in the case of ZW system) to the sex-determining genes. However, neither W-linked sexual antagonistic alleles nor sex-determining genes have been identified so far in birds.

Alternatively, we propose that the TE insertion may have occurred before the chromosome inversion and initiated the recombination suppression[62]. In this case, the recombination loss between sex chromosomes proceeded gradually rather than immediately by chromosome inversion. We found a genome-wide burst of CR1 subfamilies specifically concentrated at PAB (Fig. 2d and Supplementary Table 3). This songbird PAB has undergone genomic deletions or rearrangements independently in many other bird species[48,49] and exhibits a different subfamily of CR1 insertion in rifleman, therefore it is probably a mutation hotspot. It has been reported in many species that local recombination rate and abundance of TE elements generally have a negative association, with their causal relationship being difficult to disentangle[46]. However, several patterns suggest that the CR1–E accumulation is the cause rather than the result of recombination suppression at songbird PAB: first, in several species (for example, Lawes's parotia and King BOP; Fig. 2d and Supplementary Fig. 8), the CR1–E repeats are also enriched in the part of PAR close to PAB, where there is supposed to be frequent recombination. Second, the abundance of CR1–E is gradually decreased further away from the PAB. Third, only the CR1–E repeats but not any other type of CR1 or repeat families have accumulated at the PAB. These patterns are distinct from that of *Alu* insertion at the human PAB[61], which does not extend into PAR or show a specific enrichment at certain regions at a chromosome-wide level. They are unlikely if chromosome inversion occurred before the CR1–E accumulation, which predicts a more uniformly distributed accumulation of various kinds of repeat elements (for example, LTR elements in this study; Supplementary Fig. 21) that would not extend into the PAR. In addition, our comparative analyses between species indicate there was no Z-linked inversion in S3 (Fig. 2a and Supplementary Fig. 5), although we cannot exclude the possibility of a W-linked inversion that may have contributed to the formation of S3. Verification of the latter requires improvements of genome assembly using, for example, PacBio/Nanopore sequencing technology to assemble the highly repetitive W-linked sequence.

We propose that TEs probably reduced the recombination rate in PAR through, for example, changing the chromatin structure or disrupting the recombination hotspot[63,64]. TE accumulation was probably selected against at the beginning because it disrupted

gene functions. This has been demonstrated by results showing that at the PAB where CR1–E has accumulated, several genes have been partially or completely deleted in songbirds[48,49]. However, the resulting reduction of recombination rate can provide the selective advantage of accelerating the fixation of pre-existing sexual antagonistic polymorphic alleles in PAR through sex-biased transmission, or subjecting the PAR for the 'fast-Z' evolution by male-driven evolution effect (Fig. 3) and increasing its exposure for male-biased selection, so that new sexual antagonistic alleles may more frequently emerge and become fixed. The latter has been implicated by the recent findings in songbirds that male-specific trait genes, for example those related to sperm morphology[65] or plumage colours[66], which have recently diverged within or between populations, are enriched on the Z chromosome. In addition, TE accumulation is likely to increase the chance of chromosome inversions through ectopic recombination or by reducing the selective constraints on gene synteny. The latter is supported by our result that older evolutionary strata have undergone more Z-linked genomic rearrangements between songbird species than the younger ones (Fig. 2a), which creates a positive feedback once the recombination suppression was initiated. This provides a mechanistic explanation for a more frequent fixation of Z-linked inversions among passerines.

While Z chromosome is predicted to accumulate dominant male-beneficial mutations, W chromosome is expected to accumulate female-beneficial mutations responding to the female-specific transmission. However, both previous works in chicken and flycatcher[10,24], as well as our study have not found evidence for such 'feminization' of W chromosome. This is in contrast to reported 'masculinization' cases of ancestral Y chromosomes of mammals[29] or of recently evolved Y chromosome of *Drosophila miranda*[31]. Y-linked genes specifically expressed in male germline have either greatly amplified their copy numbers or upregulated their expression levels in these systems. Such a difference can be explained by the fact that, regardless of the sex chromosome type, sexual selection is more often targeting males in most species. Thus Z/Y chromosome is more frequently influenced than the W/X chromosome due to male-biased transmission, although the X chromosome is nevertheless expected to accumulate recessive male-beneficial alleles[9]. The convergently evolved pattern shared between the mammalian Y and avian W chromosomes is largely attributed to the essential genes that have important regulatory functions and are preferentially retained over the long period of recombination suppression (Fig. 4). Besides a weaker Hill–Robertson effect by the course of Y or W chromosome evolution, these essential genes probably have also contributed to the decreased rate of gene loss, as they are under stronger selective constraints than other genes that became lost at earlier stages. However, previous transcriptome comparison of chicken breeds selected for egg-laying versus fighting, that is, female-specific versus male-specific traits, has found most W-linked genes are upregulated in the former[34]. Few high-quality avian W chromosome sequences are available except for that of chicken. Songbirds provide a rich resource with many species (for example, blue crow) having a reversed direction of sexual selection and ornamented females. Application of long-read sequencing technology will help to elucidate the role of W chromosome in sexual selection and speciation of birds[67].

Methods

Genome assembly and annotation. Genomic DNA was extracted from fresh tissue samples of female BOP species *Cicinnurus regius* (museum catalogue number ANWC B24969), *Cicinnurus magnificus* (ANWC B27061), *Paradisaea raggiana* (USNM638608) and both sexes of *Paradisaea rubra* (YPM84686; ref. [68]) and *Parotia lawesii* (ANWC B26535 and ANWC B15265), using Thermo Scientific KingFisher Duo Prime purification system or EZNA SQ Tissue DNA Kit. Paired-end and mate pair libraries for these samples were prepared by SciLifeLab in Stockholm, Sweden. All libraries were sequenced on Illumina HiSeq 2500 or Hiseq X v4 at SciLifeLab or BGI. We also used the published female genomes of *Corvus brachyrhynchos*, *Serinus canaria, Passer domesticus, Geospiza fortis, Ficedula albicollis* and *Pseudopodoces humilis* for analysis in this work (Supplementary Table 9). The BOP genomes were assembled using ALLPATHS-LG (52488; ref. [69]) with 'HAPLOIDIFY = True'. For *P. raggiana*, due to the lack of overlapping paired-end reads, SOAPdenovo2 (v.2.04; ref. [70]) was used instead (K-mer 23). Gaps of the SOAPdenovo2 scaffolds were filled using Gapcloser (v.1.12) with default parameters. Gene models were annotated using the MAKER pipeline (v.2.31.9) in two rounds[71]. The reference protein sequences of zebra finch, great tit, hooded crow, American crow, collared flycatcher and chicken were downloaded from NCBI RefSeq (Supplementary Table 9). Using the reference protein sequences and chicken HMM (Hidden Markov Models), an initial set of gene models was obtained by using MAKER and those models were taken for SNAP (v.2013.11.29) model training[72]. In addition, 3,000 gene models with top AED (Annotation Edit Distance) scores were selected for Augustus (v.3.2.3) training[73]. The trained gene models and the protein sequences were taken as input for MAKER in the second run. To annotate repeats, first we used RepeatModeler (http://www.repeatmasker.org/, v.1.0.10) with default parameters to identify and classify repeat elements for each species. Then we combined each individual library with an avian repeat library[68] to annotate repeats using RepeatMasker (v.4.0.7) with the parameter '-a -xsmall -gccalc'. This repeat annotation pipeline was also applied to other published avian species for re-annotation to allow for a direct comparison.

Identification of sex-linked sequences. We used the published Z chromosome sequences of two species, great tit[6] and hooded crow[4], as references to maximize the identification of the sex-linked sequences in the studied species. The great tit Z chromosome assembly is the most complete among the published genomes of songbirds and hooded crow is the closest species to BOP with male genome data available. We first used nucmer from MUMMer package (v.4.0; ref. [74]) for genome-wide pairwise sequence alignment with the parameter '-b 400' between all the studied species versus the reference species. Only the best one-to-one alignments were retained (delta-filter-1). Any scaffold longer than 10 kb that has over 60% of the sequence aligned to reference Z chromosome was identified as a candidate Z-linked scaffold in that species. All the scaffolds, including the candidate Z-linked scaffolds were examined for whether their female sequencing coverage values are about half of those of autosomes. The raw female reads were mapped to the scaffold sequences by using bwa (v.0.7.16a, ref. [75]; default parameters of the maximal exact matches algorithm), subsequently the average sequencing coverage of every 50-kb window was calculated. We then plotted the distribution of coverage values of all the windows (Supplementary Fig.1) to decide whether a scaffold is a candidate Z- or W-linked scaffold by showing the half-coverage value of autosomes. Sequencing coverage of each nucleotide position was also counted using 'samtools depth' before calculating window-based coverage. Any alignment with low mapping quality (lower than 60) was not counted to exclude the effect of probable misalignments. Additionally, any site with high coverage (three times larger than average) was excluded, as it was probably derived from repetitive sequences.

To identify the W-linked scaffolds, first we focused on all the half-coverage scaffolds that aligned with the reference Z chromosome and inspected their distribution of the proportion of aligned sequences to decide a cutoff. This cutoff was used to separate the candidate Z- and W- linked scaffolds, and varies from species to species, possibly due to the varying assembly quality and/or the divergence from the reference genomes. We excluded the candidate W-linked scaffolds that over 10% of the sequences were aligned to reference autosomes, or a larger portion of sequences aligned to the autosomes than the Z. Finally, only the scaffolds longer than 10 kb were kept because shorter scaffold often show ambiguous coverage patterns. For species that have both male and female sequencing reads available, we directly verified the candidate W-linked sequences by mapping the male reads. Specifically, for each scaffold, the number of nucleotide sites that were mapped by male and female sequencing data were counted as N_m and N_f respectively, with their ratios as N_m/N_f. W-linked scaffolds are expected to have Nm/Nf ratios close to zero, while autosome or Z-linked scaffolds tend to have an expected ratio of 1 (this is the ratio of mappable sites; for the ratio of coverage between sexes, the expected value would be 2). Given the short divergence time of BOP species[51], we are able to map substantial numbers of male reads of a red BOP to the three BOP species (magnificent BOP, king BOP and Raggiana BOP; over 95% of the genomes can be mapped) lacking the male data to verify their W-linked sequences. We used the known PAR sequences of zebra finch[76] and flycatcher[40] to infer those of other species using nucmer (-b 400) and then confirmed their similar levels of female coverage value to autosomes.

Demarcation of evolutionary strata. We ordered and oriented the identified Z-linked scaffolds of all BOP species into one pseudo-chromosomal sequence (pseudo-chrZ) on the basis of their alignments against the Z chromosome of great tit. For Fig. 2a, we used chromosomal or nearly chromosomal assemblies of four species but not BOPs. For example, hooded crow has 15 Z-linked scaffolds and ten of them are larger than 1 Mb. We determined the relative order and orientation of the crow scaffolds according to their alignment with the great tit Z chromosome. We used nucmer for pairwise alignment of the Z or pseudo-Z chromosomes between species. Alignments short than 2 kb were excluded to avoid probable

misalignments. Similarly, for BOP species, we created pseudo-chromosome Z using the great tit Z chromosome as reference. The pseudo-Z chromosome of emu was built using ostrich Z chromosome[77] as reference and was available from Xu et al.[78]. The W-linked scaffolds were then aligned to the pseudo-chrZ of the same species using lastz (v.1.04; ref. [79]) with '--step = 19 --hspthresh = 2200 --inner = 2000 --ydrop = 3400 --gappedthresh = 10000', after masking the repetitive sequences. Sequence similarity of the alignments between the Z and W chromosomes was calculated by the script pslScore from UCSC Genome Browser (https://genome.ucsc.edu/). Individual alignments that have sequence identities lower than 60 or higher than 96, or alignment lengths shorter than 65% were removed as those are probably derived from misalignments or unmasked repeats. After that, we ordered the W-linked scaffolds by their aligned positions along the pseudo-chrZ. We then extracted Z/W pairwise alignments for every non-overlapping sliding window of 100 kb along the pseudo-chrZ and calculated the sequence divergence level for the windows whose lengths of ZW alignment are longer than 2 kb. The window-based sequence divergence levels were then plotted along the pseudo-chrZ, with the shift of divergence level to demarcate the boundaries between evolutionary strata. Since few W-linked sequences have been assembled for the most ancient stratum S0, we mapped its reshuffled fragments in songbirds on the basis of their homology with the emu S0. Our previous study showed that emu has a recent species-specific stratum (S1), while the oldest stratum (S0) is shared by all birds[23]. This allows for the demarcation of S1 and S0 by detecting their differential degree of Z/W differentiation in emu. Specifically, by using a relatively relaxed mapping criteria (bwa mem) to map the female sequencing reads, only S0 showed reduced coverage relative to autosomes or PAR (Supplementary Fig. 4), while S1 showed reduced coverage when stringent mapping was applied (bwa sampe –a 900 –n 1 –N 0 –o 10000). To examine the accumulated LINE (mostly CR1) elements at the PAB, we first divided them into each subtype according to the RepeatMasker annotation. Among all the subtypes, CR1–E1 was usually ranked with the highest or second highest copy number at the S3 region across all songbird species. Other high-ranking subtypes included CR1–E3, CR1–E5, CR1–E4, CR1–E6, CR1–J2 and CR1–Y2. Then we plotted each subtype's abundance with a 100 kb non-overlapping window along the Z chromosome, in all the studied songbirds, as well as outgroup species rifleman and falcon, to identify the burst of CR1–E1.

Sex-linked gene analyses. We used BLAT (v.35.1; ref. [80]) to align the annotated coding sequence of W-linked genes to the Z chromosome to search for their homologous pairs after removing the LTR genes. Then we produced pairwise gametologue alignments using MUSCLE (v.3.8.31) with the default parameters[81] and manually inspected the alignments to remove genes with short or ambiguous alignments. For species other than BOPs, gene models of the W chromosomes were directly retrieved from the RefSeq genome annotation (https://www.ncbi.nlm.nih.gov/refseq/), with some of them also subjected to manual inspections. To determine the orthologous relationship among the studied species, we first extracted the sequence of the longest protein of each gene. Those protein sequences were subjected to all-versus-all BLAST search that was implemented through the program proteinortho (v.5.16; ref. [82]). BLAST hits with sequence identity lower than 50% or aligned percentage lower than 50% were removed. We also took gene synteny information into account when grouping orthologous genes. Besides the 12 female genomes for which we studied the sex chromosomes, we also included high-quality genomes of great tit, hooded crow and ostrich (Supplementary Table 9). We retained those orthologous groups if they contain sequences of at least ten species. To estimate the substitution rates of coding sequences, first we performed multiple sequence alignment for orthologous genes. We used the guidence2 pipeline (http://guidance.tau.ac.il/ver2/source.php) which used PRANK (v.170427) to align sequences of codons, with the default parameters. To filter low-quality sites in the alignments, we ran trimal (http://trimal.cgenomics.org/) with '-gt 0.8'. The phylogeny of the birds was extracted from Jetz. et al.[83]. We used codeml from the PAML package (v.4.9e; ref. [84]) to estimate the synonymous substitution rates (d*S*) and non-synonymous substitution rates (d*N*). To estimate the chromosome-wide d*N* and d*S*, sums of synonymous or non-synonymous substitutions were divided by those of total synonymous or non-synonymous sites, as applied in Wright et al.[22]. Individual genes with abnormal d*N* (higher than 0.1, in total 179 genes) or d*S* (higher than 0.8, in total 135 genes) out of 111,748 orthologous gene groups were removed, as those were probably caused by misalignments or misassignment of orthologues. Confidence intervals were calculated by 100 bootstrappings. Chromosome-wide d*N*/d*S* (ω) was calculated by the ratios of chromosome-wide d*N* to chromosome-wide d*S*. The fast-Z effect was measured by Z/A value as the ratio of ω values of Z-linked genes to autosomal genes and we calculated the Z/A value for each terminal branch and internal branch. To determine if the difference of ω between Z-linked and autosomal genes is significant, we performed permutation test by resampling 1,000 times to calculate the two-sided *P*-values. The genes of chromosome 4 and chromosome 5 were used to represent autosomal genes as the sizes of those two chromosomes are similar to the Z chromosome. For each gametologue pair, we grouped together Z-linked genes and assembled W-linked genes and performed multiple sequence alignment. The same guidence2 pipeline was used as in sequence divergence analysis. For S3 genes, we also included rifleman[85] to infer the origination time of S3. We used IQ-TREE (v.1.6.1; ref. [86]) to construct phylogenetic trees. The best substitution model was automatically selected in by IQ-TREE. We ran bootstrapping 100 times to evaluate the confidence levels of phylogenies with ostrich as outgroup to root the tree. The gene ontology annotations for both studied gametologue-pair genes (list) and all Z-linked genes (background) of chicken was analysed by DAVID v.6.8 (ref. [87]). Gene ontology term enrichment was estimated by comparing the numbers of appearance of GO terms of 'list' gene versus 'background' gene. The GC content of the third position of codons (GC3) was calculated using codonW (http://codonw.sourceforge.net//culong.html) for the longest isoform of each gene.

Gene loss analysis. We identified 673 Z-linked orthologous genes shared between chicken and emu as the putative ancestral genes on the proto-sex chromosomes of birds. For the gene cluster that was lost in chicken at the *DCC* loci on the S3 and ancestral gene content was inferred on the basis of Fig. 3 of Patthey et al.[49]. They were then grouped into four evolutionary strata according to the strata annotation of songbird Z chromosomes. At each node of the avian phylogenetic tree, we calculated the ratio of lost genes to ancestral genes of that node. For the nodes leading to Passerida and Corvida, if there was at least one species retaining a W-linked gene, we inferred that that gene was present in their ancestor. Similarly, we defined the presence of ancestral genes in Passeriformes, Neoaves and Neognathae according to the presence/absence of W-linked genes in other published avian species[23,24].

Gene expression analysis. We downloaded the raw RNA-seq reads of green anole (brain, gonad, liver, heart and kidney) and emu (brain, gonad and spleen) from SRA (Supplementary Table 9). In addition, we collected the transcriptomes of adult emu kidneys of both sexes. We used the RSEM pipeline (v.1.3.0; ref. [88]) to quantify the gene expression levels. The RSEM pipeline used STAR (v.2.5.30; ref. [89]) with default parameters to map raw reads to the transcriptomes which was constructed on the basis of gene annotations. The expectation–maximization (EM) algorithm was used to estimate the abundance of transcripts by RSEM. The expression levels were estimated at the gene level, in the form of TPM (transcripts per million). The mean TPM value of biological replicates was calculate for each gene. Tissue specificity of gene expression were estimated by calculating tau[90].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Genome sequencing and RNA-seq data generated in this study have been deposited in the NCBI SRA under PRJNA491255. The raw genomic reads of *Paradisaea raggiana* are available in the CNGB Nucleotide Sequence Archive (https://db.cngb.org/cnsa; accession number CNP0000186). The genome assemblies are available under NCBI BioProject portal (PRJNA491255). The IDs of W-linked scaffolds are included in Supplementary Table 10.

Code availability

Custom scripts and pipelines used in this study have been deposited at Github (https://github.com/lurebgi/BOPsexChr).

Received: 29 October 2018; Accepted: 22 February 2019;
Published online: 1 April 2019

References

1. World Bird List v.8.2 (IOC) (2018); https://www.worldbirdnames.org/ioc-lists/master-list-2/
2. Barker, F. K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. Phylogeny and diversification of the largest avian radiation. *Proc. Natl Acad. Sci. USA* **101**, 11040–11045 (2004).
3. Ellegren, H. et al. The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**, 756–760 (2012).
4. Poelstra, J. W. et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**, 1410–1414 (2014).
5. Tuttle, E. M. et al. Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344–350 (2016).
6. Laine, V. N. et al. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* **7**, 10474 (2016).
7. Mank, J. E. Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. *Am. Nat.* **173**, 141–150 (2009).
8. Coyne, J. A. Genetics and speciation. *Nature* **355**, 511 (1992).
9. Charlesworth, B., Coyne, J. & Barton, N. The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**, 113–146 (1987).
10. Smeds, L. et al. Evolutionary analysis of the female-specific avian W chromosome. *Nat. Commun.* **6**, 7330 (2015).
11. Hooper, D. M. & Price, T. D. Chromosomal inversion differences correlate with range overlap in passerine birds. *Nat. Ecol. Evol.* **1**, 1526–1534 (2017).
12. Hooper, D. M. & Price, T. D. Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. *Evolution* **69**, 890–903 (2015).

13. Backström, N. et al. A high-density scan of the Z chromosome in Ficedula flycatchers reveals candidate loci for diversifying selection. *Evolution* **64**, 3461–3475 (2010).
14. Elgvin, T. O. et al. Hybrid speciation in sparrows II: a role for sex chromosomes?. *Mol. Ecol.* **20**, 3823–3837 (2011).
15. Storchová, R., Reif, J. & Nachman, M. W. Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. *Evolution* **64**, 456–471 (2010).
16. Li, W. H., Yi, S. & Makova, K. Male-driven evolution. *Curr. Opin. Genet. Dev.* **12**, 650–656 (2002).
17. Wang, Z. et al. Temporal genomic evolution of bird sex chromosomes. *BMC Evol. Biol.* **14**, 250 (2014).
18. Ellegren, H. Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc. Biol. Sci.* **274**, 1–10 (2007).
19. Smeds, L., Qvarnström, A. & Ellegren, H. Direct estimate of the rate of germline mutation in a bird. *Genome Res.* **26**, 1211–1218 (2016).
20. Vicoso, B. & Charlesworth, B. Effective population size and the faster-X effect: an extended model. *Evolution* **63**, 2413–2426 (2009).
21. Mank, J. E., Nam, K. & Ellegren, H. Faster-Z evolution is predominantly due to genetic drift. *Mol. Biol. Evol.* **27**, 661–670 (2010).
22. Wright, A. E. et al. Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. *Mol. Ecol.* **24**, 1218–1235 (2015).
23. Zhou, Q. et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* **346**, 1246338 (2014).
24. Bellott, D. W. et al. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nat. Genet.* **49**, 387–394 (2017).
25. Charlesworth, D., Charlesworth, B. & Marais, G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**, 118–128 (2005).
26. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
27. Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. B* **355**, 1563–1572 (2000).
28. Bachtrog, D. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* **179**, 1513–1525 (2008).
29. Soh, Y. Q. et al. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **159**, 800–813 (2014).
30. Bellott, D. W. et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
31. Zhou, Q. & Bachtrog, D. Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* **337**, 341–345 (2012).
32. Koerich, L. B., Wang, X., Clark, A. G. & Carvalho, A. B. Low conservation of gene content in the *Drosophila* Y chromosome. *Nature* **456**, 949–951 (2008).
33. Wright, A. E., Harrison, P. W., Montgomery, S. H., Pointer, M. A. & Mank, J. E. Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution* **68**, 3281–3295 (2014).
34. Moghadam, H. K., Pointer, M. A., Wright, A. E., Berlin, S. & Mank, J. E. W chromosome expression responds to female-specific selection. *Proc. Natl Acad. Sci. USA* **109**, 8207–8211 (2012).
35. Cortez, D. et al. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014).
36. Bergero, R., Forrest, A., Kamau, E. & Charlesworth, D. Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* **175**, 1945–1954 (2007).
37. Gorelick, R. et al. Abrupt shortening of bird W chromosomes in ancestral Neognathae. *Biol. J. Linn. Soc. Lond.* **119**, 488–496 (2016).
38. Rice, W. R. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41**, 911–914 (1987).
39. Claramunt, S. & Cracraft, J. A new time tree reveals Earth historys imprint on the evolution of modern birds. *Sci. Adv.* **1**, e1501005–e1501005 (2015).
40. Smeds, L. et al. Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nat. Commun.* **5**, 5448 (2014).
41. Pala, I. et al. Evidence of a neo-sex chromosome in birds. *Heredity* **108**, 264–272 (2012).
42. Ross, M. T. et al. The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
43. O'Connor, R. E. et al. Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. *Genome Biol.* **19**, 171 (2018).
44. Wright, A. E., Dean, R., Zimmer, F. & Mank, J. E. How to make a sex chromosome. *Nat. Commun.* **7**, 12087 (2016).
45. Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**, 285–311 (2009).
46. Kent, T. V., Uzunović, J. & Wright, S. I. Coevolution between transposable elements and recombination. *Philos. Trans. R. Soc. B* **372**, 20160458 (2017).
47. Suh, A. et al. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Commun.* **2**, 443 (2011).
48. Friocourt, F. et al. Recurrent DCC gene losses during bird evolution. *Sci. Rep.* **7**, 37569 (2017).
49. Patthey, C., Tong, Y. G., Tait, C. M. & Wilson, S. I. Evolution of the functionally conserved DCC gene in birds. *Sci. Rep.* **7**, 42029 (2017).
50. Ligon, R. A. et al. Evolution of correlated complexity in the radically different courtship signals of birds-of-paradise. *PLoS Biol.* **16**, e2006962 (2018).
51. Irestedt, M., Jonsson, K. A., Fjeldsa, J., Christidis, L. & Ericson, P. G. An unexpectedly long history of sexual selection in birds-of-paradise. *BMC Evol. Biol.* **9**, 235 (2009).
52. Dunn, P. O., Whittingham, L. A. & Pitcher, T. E. Mating systems, sperm competition, and the evolution of sexual dimorphism in birds. *Evolution* **55**, 161–175 (2001).
53. Diamond, J. Biology of birds of paradise and bowerbirds. *Annu. Rev. Ecol. Syst.* **17**, 17–37 (1986).
54. Kirkpatrick, M. & Ryan, M. J. The evolution of mating preferences and the paradox of the lek. *Nature* **350**, 33–38 (1991).
55. Suh, A. The specific requirements for CR1 retrotransposition explain the scarcity of retrogenes in birds. *J. Mol. Evol.* **81**, 18–20 (2015).
56. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6**, e1001154 (2010).
57. Stock, M. et al. Ever-young sex chromosomes in European tree frogs. *PLoS Biol.* **9**, e1001062 (2011).
58. Vicoso, B., Emerson, J. J., Zektser, Y., Mahajan, S. & Bachtrog, D. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol.* **11**, e1001643 (2013).
59. Charlesworth, D. Evolution of recombination rates between sex chromosomes. *Philos. Trans. R. Soc. Lond. B* **372**, 20160456 (2017).
60. Lemaitre, C. et al. Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. *Genome Biol. Evol.* **1**, 56–66 (2009).
61. Ellis, N. A. et al. The pseudoautosomal boundary in man is defined by an Alu repeat sequence inserted on the Y chromosome. *Nature* **337**, 81–84 (1989).
62. Raudsepp, T. & Chowdhary, B. P. The eutherian pseudoautosomal region. *Cytogenet. Genome Res.* **147**, 81–94 (2015).
63. Zamudio, N. et al. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev.* **29**, 1256–1270 (2015).
64. Ben-Aroya, S., Mieczkowski, P. A., Petes, T. D. & Kupiec, M. The compact chromatin structure of a Ty repeated sequence suppresses recombination hotspot activity in *Saccharomyces cerevisiae*. *Mol. Cell* **15**, 221–231 (2004).
65. Knief, U. et al. Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biol.* **17**, 199 (2016).
66. Campagna, L. et al. Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Sci. Adv.* **3**, e1602404 (2017).
67. Irwin, D. E. Sex chromosomes and speciation in birds and other ZW systems. *Mol. Ecol.* **27**, 3831–3851 (2018).
68. Prost, S. et al. Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. *GigaScience* https://doi.org/10.1093/gigascience/giz003 (2018).
69. Gnerre, S. et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl Acad. Sci. USA* **108**, 1513–1518 (2011).
70. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
71. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
72. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
73. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–ii225 (2003).
74. Marçais, G. et al. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
75. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics (q-bio.GN)* https://arxiv.org/abs/1303.3997 (2013).
76. Singhal, S. et al. Stable recombination hotspots in birds. *Science* **350**, 928–932 (2015).
77. Zhang, J., Li, C., Zhou, Q. & Zhang, G. Improving the ostrich genome assembly using optical mapping data. *GigaScience* **4**, 24 (2015).
78. Xu, L. et al. Evolutionary dynamics of sex chromosomes of palaeognathous birds. Preprint at *bioRxiv* https://doi.org/10.1101/295089 (2018).
79. Harris, R. S. Improved pairwise alignment of genomic DNA. PhD thesis, The Pennsylvania State University (2017).
80. Kent, W. J. BLAT — the BLAST-like alignment tool. *Genome Res.* **12**, 656–64 (2002).
81. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

82. Lechner, M. et al. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).
83. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
84. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
85. Zhang, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
86. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
87. Jiao, X. et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* **28**, 1805–1806 (2012).
88. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
89. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
90. Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
91. del Hoyo, J, . & Elliot, A. & Sargatal, J. & Christie, D. & de Juana, E. *Handbook of the Birds of the World Alive* (Lynx Editions, 2019).

Acknowledgements

We thank the Smithsonian Institute (G. Graves), Australian National Wildlife Collection, CSIRO Sustainable Ecosystems (L. Joseph), Museum Victoria, Australia (J. Sumner), Division of Vertebrate Zoology Yale University, Peabody Museum of Natural History (K. Zyskowski) for tissue samples; and E. Scholes for discussions on BOP. We also acknowledge the support from Science for Life Laboratory, the National Genomics Infrastructure (NGI), Uppmax. L.X. is supported by the uni:docs fellowship programme from University of Vienna. M.I. is supported by the Swedish Research Council (grant no. 621-2014-5113). Q.Z. is supported by National Natural Science Foundation of China (grant nos. 31722050 and 31671319), the Fundamental Research Funds for the Central Universities (grant no. 2018XZZX002-04) and start-up funds from Zhejiang University. The computational analyses were performed on CUBE cluster from Department of Computational System Biology of University of Vienna and Vienna Scientific Cluster.

Author contributions

Q.Z. and M.I. conceived the project. L.X., Q.Z., G.A., V.P., Y.D., S.F., G.Z., M.B. and S. P. performed the analyses. Q.Z., L.X., A.S., L.C. and M.I. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-019-0850-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.I. or Q.Z.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Qi Zhou, Martin Irestedt

Last updated by author(s): Feb 19, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used sratoolkit (2.8.2-1) to download raw reads from NCBI SRA.
Data analysis	For genome assembly, ALLPATHS-LG (52488) and SOAPdenovo2 (2.04) were used. For genome annotation, we used MAKER (2.31.9), Augustus (3.2.3), SNAP (2013.11.29), exonerate (2.2.0), RMBlast (2.2.28), RepeatMasker (4.0.7) and RepeatModeler (1.0.10). For sequence alignment, we used MUMMER (4.0), BLAT (35.1), PRANK (170427) , MUSCLE (3.8.31) and lastz (1.04). For gene expression analysis, we used STAR (2.5.30) and RSEM (1.3.0). IQ-TREE (1.6.1) was used for phylogeny construction. Custom codes are available at https://github.com/lurebgi/BOPsexChr

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the raw reads and genome assemblies generated in this study have been deposited in NCBI under the accession number PRJNA491255. The raw genomic reads of Paradisaea raggiana are available in the CNGB Nucleotide Sequence Archive (https://db.cngb.org/cnsa; accession number CNP0000186).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed.
Data exclusions	No data were excluded from the analyses.
Replication	This is not applicable as there is no experiments involved.
Randomization	This is not applicable as there is no experiments involved.
Blinding	This is not applicable as there is no experiments involved.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	The study did not involve wild animals.
Field-collected samples	The samples of birds-of-paradise we used were all collected from museum specimen. Information of specimen are available under the accession number of PRJNA491255 at NCBI. Tissue samples of one female and one male emu were collected from an emu farm in Fujian, China, both at the age of one year.
Ethics oversight	No ethical approval was required, as the study only involves museum specimen and blood samples of poultry species.

Note that full information on the approval of the study protocol must also be provided in the manuscript.