RESEARCH

A new duck genome reveals conserved and convergently evolved chromosome architectures of birds and mammals

Jing Li [1], Jilin Zhang [2], Jing Liu[1,3], Yang Zhou[4], Cheng Cai[1], Luohao Xu [1,3], Xuelei Dai[5], Shaohong Feng[4], Chunxue Guo[4], Jinpeng Rao [6], Kai Wei[6], Erich D. Jarvis [7,8], Yu Jiang[5], Zhengkui Zhou[9], Guojie Zhang [10,11,12,13] and Qi Zhou [1,3,6,*]

[1], MOE Laboratory of Biosystems Homeostasis & Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China; [2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 5 Nobels väg, Stockholm 17177, Sweden; [3]Department of Neuroscience and Developmental Biology, University of Vienna, 1 Universitätsring, Vienna 1090, Austria; [4]BGI-Shenzhen, 146 Beishan Industrial Zone, Shenzhen 518083, China; [5]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, 3 Taicheng Road, Yangling 712100, China; [6]Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Zhejiang University, 88 Jiefang Road, Hangzhou 310052, China; [7]Laboratory of Neurogenetics of Language, The Rockefeller University, 1230 York Ave, NY 10065, USA; [8]Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815, USA; [9]Institute of Animal Science, Chinese Academy of Agricultural Sciences, 12 Zhong Guan Cun Da Jie, Beijing, China; [10]China National GeneBank, BGI-Shenzhen, Jinsha Road, Shenzhen 518120, China; [11]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 East Jiaochang Road, Kunming 650223, China; [12]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, 10 Nørregade, DK-2100 Copenhagen, Denmark and [13]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 32 East Jiaochang Road, Kunming 650223, China

*Correspondence address. Qi Zhou, Room 438, College of Life Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China.. E-mail: zhouqi1982@zju.edu.cn http://orcid.org/0000-0002-7419-2047

Abstract

Background: Ducks have a typical avian karyotype that consists of macro- and microchromosomes, but a pair of much less differentiated ZW sex chromosomes compared to chickens. To elucidate the evolution of chromosome architectures between ducks and chickens, and between birds and mammals, we produced a nearly complete chromosomal assembly of a female Pekin duck by combining long-read sequencing and multiplatform scaffolding techniques. **Results:** A major

improvement of genome assembly and annotation quality resulted from the successful resolution of lineage-specific propagated repeats that fragmented the previous Illumina-based assembly. We found that the duck topologically associated domains (TAD) are demarcated by putative binding sites of the insulator protein CTCF, housekeeping genes, or transitions of active/inactive chromatin compartments, indicating conserved mechanisms of spatial chromosome folding with mammals. There are extensive overlaps of TAD boundaries between duck and chicken, and also between the TAD boundaries and chromosome inversion breakpoints. This suggests strong natural selection pressure on maintaining regulatory domain integrity, or vulnerability of TAD boundaries to DNA double-strand breaks. The duck W chromosome retains 2.5-fold more genes relative to chicken. Similar to the independently evolved human Y chromosome, the duck W evolved massive dispersed palindromic structures, and a pattern of sequence divergence with the Z chromosome that reflects stepwise suppression of homologous recombination. **Conclusions:** Our results provide novel insights into the conserved and convergently evolved chromosome features of birds and mammals, and also importantly add to the genomic resources for poultry studies.

Background

Birds have the largest number of species and some of the smallest genome sizes among terrestrial vertebrates. Since the era of cytogenetics, this has attracted extensive efforts to elucidate the diversity of their "streamlined" genomes that give rise to the tremendous phenotypic diversity [1]. The karyotype of birds exhibits 2 major distinctions from that of mammals: first, it comprises ~10 pairs of large- to medium-sized chromosomes (macrochromosomes) and ~30 pairs of much smaller sized chromosomes (microchromosomes) [2]. During the >100 million years (MY) of avian evolution, there were few interchromosomal rearrangements among most species [3–5] except for falcons and parrots (Falconiformes and Psittaciformes) [6–9]. Among the published karyotypes of >800 bird species, the majority have a similar chromosome number ~2n = 80 [10]. These results indicate that the chromosome evolution of birds is dominated by intrachromosomal rearrangements. Genomic comparisons between chicken, turkey, flycatcher, and zebra finch [11, 12] found that birds, similar to mammals [13, 14], have fragile genomic regions that were recurrently used for mediating intrachromosomal rearrangements, and these regions seem to be associated with high recombination rates [15] and low densities of conserved non-coding elements (CNEs) [5]. However, compared with mammals [13, 14, 16], much less is known about the interspecific diversity within avian chromosomes, particularly microchromosomes (but see [5, 12]) at the sequence level, owing to the scarcity of chromosome-level bird genomes.

The other major distinction between the mammalian and avian karyotypes is their sex chromosomes. Birds have a pair of female heterogametic (male ZZ, female ZW) sex chromosomes that originated from a different pair of ancestral autosomes than the eutherian XY [17, 18]. Since their divergence ~300 MY ago, sex chromosomes of birds and mammals have undergone independent stepwise suppression of homologous recombination, and produced a punctuated pattern of pairwise sequence divergence levels between the neighboring regions termed "evolutionary strata" [19–21]. Despite the consequential massive gene loss, both chicken W chromosome (chrW) and eutherian chrYs have been found to preferentially retain dosage-sensitive genes or genes with important regulatory functions [22]. In addition, the human chrY has evolved palindromic sequences that may facilitate gene conversions between the Y-linked gene copies [23], as an evolutionary strategy to limit the functional degeneration under the non-recombining environment [24]. Interestingly, such palindromic structures have also been reported on sex chromosomes of New World sparrows and blackbirds [25],

and more recently in a plant species, the willow [26], suggesting that it is a general feature of evolving sex chromosomes. Both cytogenetic work and Illumina-based genome assemblies of tens of bird species have suggested that bird sex chromosomes comprise an unexpected interspecific diversity regarding both their lengths of recombining regions (pseudoautosomal regions [PARs]) and their rates of gene loss [20, 27]. For example, PARs cover more than two-thirds of the length of ratite (e.g., emu and ostrich) sex chromosomes [28] but are concentrated at the tips of the chicken and eutherian sex chromosomes. However, so far only the chicken chrW has been well assembled using the laborious iterative clone-based sequencing method [22], and the majority of genomic sequencing projects tend to choose a male bird to avoid the repetitive chrW. This has hampered our broad and deep understanding of the composition and evolution of avian sex chromosomes.

The Vertebrate Genomes Project (VGP) has taken advantage of the development of long-read (Pacific Biosciences [PacBio] or Nanopore) sequencing, linked-read (10X), and high-throughput chromatin conformation capture (Hi-C) technologies to empower rapid and accurate assembly of chromosome-level genomes including the sex chromosomes, in the absence of physical maps [29]. Furthermore, Hi-C can uncover the 3D architecture of chromosomes that is segregated in active (A) and inactive (B) chromatin compartments [30], and to a finer genomic scale, topologically associated domains (TADs) as the replication and regulatory units [31]. To elucidate the evolution of avian chromosome architectures in terms of sequence composition, genomic rearrangement, and 3D chromatin structure, here we used a modified VGP pipeline to produce a nearly complete reference genome of a female Pekin duck (*Anas platyrhynchos*, Z2 strain; NCBI:txid8839) with all the cutting-edge technologies mentioned above. We corroborated our reference genome through comparisons to previously published radiation hybrid (RH) [32] and fluorescence *in situ* hybridization (FISH) [33] linkage maps. We chose duck because first, as a representative species of *Anseriformes*, it diverged from *Galliformes* ~72.5 MY ago [34], providing a deep but still trackable evolutionary distance for addressing the functional consequences of genomic rearrangements on chromatin domains. Second, the duck sex chromosomes have diverged to a degree between the highly heteromorphic sex chromosomes of chicken and homomorphic sex chromosomes of emu [20, 27]. The gradient of sex chromosome divergence levels exhibited by the 3 bird species together constitute a chronological order for a comprehensive understanding of the entire avian sex chromosome evolution process. Finally, besides being frequently used for basic evolutionary and developmental

studies [35], the duck is another key poultry species, as well as a natural reservoir of all influenza A viruses [36]. Our new duck genome has anchored >95% of the assembled sequences onto chromosomes, with great improvements in the non-coding regions and chrW sequences. We believe that it will serve as an important genomic resource for future studies into the mechanisms and application of artificial selection.

Data Description

Pekin duck (called duck hereinafter) has a haploid genome size estimated to be 1.41 Gb [37, 38], and a karyotype of 9 pairs of macrochromosomes (chr1–chr8, chrZ/chrW) and 31 pairs of microchromosomes (chr9–chr39) [39]. The Illumina-based genome assembly of the duck (BGI1.0) was produced >7 years ago and has 25.9% of the assembled genome assigned to chromosomes, containing 3.17% of bases as gaps [36]. To *de novo* assemble the new genome, we generated 143× genome coverage of PacBio long reads (read N50, 14.3 kb from 115 single-molecule real-time [SMRT] cells, Supplementary Fig. S1), and 142× genome coverage of 10X linked-read data from a female individual, 56× genome coverage of BioNano map, and 82× genome coverage of Hi-C reads from 2 different male individuals of the same inbred duck strain (Fig. 1, Supplementary Table S1), and assembled the genome with a modified VGP pipeline [29]. To identify the female-specific chrW sequences, we also generated 72× genome coverage Illumina reads from a male individual of the same duck strain to compare to the previously published female reads (SRA accession No.: PRJNA636121). Our primary assembly of PacBio long reads assembles the entire genome into 1,645 gapless contigs (Supplementary Table S2), resulting in a 14-fold reduction of contig number (1,645 vs 227,448) and 212-fold improvement of contig continuity measured by N50 (5.5 Mb vs 26.1 kb) compared with the BGI1.0 genome (Table 1). To scaffold the contigs, we first corrected their sequence errors with 92× genome coverage female Illumina reads, then oriented and scaffolded them into 942 scaffolds with 10X linked reads, BioNano optical maps, and Hi-C reads (see Methods). Because Hi-C data provide linkage but not orientation information, in our final step of chromosome anchoring, we incorporated an RH linkage map [32] and reduced the scaffold number further down to 755. We however detected 69 cases of conflicts of orientation between the RH map and the Hi-C scaffolds, manifested as inversions. By carefully examining the presence/absence of raw PacBio reads, Illumina mate-pairs, and syntenic chicken/goose sequences [40, 41] spanning the breakpoints of such inversions, the majority (54 of 69) supported the Hi-C map. And we have corrected a total of 15 orientation errors within the scaffolds (Supplementary Fig. S2).

Analysis

A much improved female duck genome

The final polished assembly (ZJU1.0) by Illumina reads exhibits a 62-fold improvement of scaffold continuity (N50, 76.3 Mb vs 1.2 Mb) compared with the Illumina genome and is completely consistent with the FISH linkage map previously generated from 155 bacterial artificial chromosome (BAC) clones (Supplementary Fig. S2) [33, 42]. The entire chrZ uniformly exhibits a 2-fold elevation of Illumina DNA sequencing read coverage in male relative to female, except for the chromosome tip of pseudoautosomal regions (PARs) (see below), confirming that we assembled the Z chromosome and that it does not have chimeric sequences with chrW or the autosomes. This new genome has 95.6% (1.13

Gb) of the assembled sequences assigned to 31 autosomes and the ZW sex chromosomes (Supplementary Table S3). The remaining 4.4% (62.1 Mb) of the genome not anchored, or ~200 Mb unassembled sequences based on the estimated genome size, is likely due to their repetitive sequence composition or lack of linkage markers. In particular, the assembled macrochromosomes have become much more continuous (Fig. 1b and c), and we have assembled majorities of microchromosomes that were all unmapped in the BGI1.0 genome (Fig. 2a).

The ZJU1.0 genome assembly also has a higher level of completeness measured by its almost gapless sequence composition (0.37% vs 3.17%), and substantial numbers of annotated telomeric and centromeric regions (Fig. 2a, Supplementary Tables S4 and S5), compared with the BGI1.0 assembly. We filled in a total of 116.2 Mb sequences of gaps within or between the BGI1.0 scaffolds, which were enriched for repetitive elements and GC-rich sequences (Supplementary Figs S3 and S4). This can be explained by the inability of Illumina reads to span or resolve the repeat regions with high copy numbers or complex structures, and the sequencing bias against the GC-rich regions [43–45]. Indeed, we found specific transposable elements (TEs) that are enriched in the filled gaps (Supplementary Fig. S4). These include the chicken repeat 1 (CR1) retroposon CR1-J2_Pass and the long terminal repeat (LTR) GGLTR8B that have undergone recent lineage-specific bursts in duck after its divergence from other Galloanserae species (Fig. 2b, Supplementary Table S6). These apparent evolutionarily young repeats relative to other repeats of the same family in ducks show a lower level of sequence divergence from their consensus sequences (Supplementary Fig. S5) and tend to insert into other older TEs and form a nested repeat structure (Supplementary Fig. S6).

Assembly of exon sequences embedded in such complex repetitive regions also led to the improvement of gene model annotations in our new assembly (e.g., Fig. 2c). Overall, our new gene annotation combining a total of 17 duck tissue transcriptomes and chicken protein queries has predicted 15,463 protein-coding genes, including 71 newly annotated chrW genes. We have identified 8,238 missing exons in the BGI1.0 assembly in 2,099 genes, including 745 genes that were completely missing. We also corrected 683 partial genes and merged them into 356 genes in the new assembly. The overall quality of our new duck genome is better than that of the previous Sanger-based zebra finch, and comparable to the latest version of chicken [41] and VGP zebra finch genomes [29] (Table 1).

Different genomic landscapes of duck micro- and macrochromosomes

Our high-quality genome assembly and annotation of Pekin duck uncovered a different genomic landscape between the macro- and microchromosomes. Duck microchromosomes have a higher gene density than macrochromosomes per Mb sequence or per TAD domain ($P < 2.2e-16$, Wilcoxon test). The recombination rate estimated from the published population genetic data [46] is also on average 2.3-fold higher on microchromosomes than on macrochromosomes (16.3 vs 7.2 per 50 kb, $P < 2.2e-16$, Wilcoxon test), which drives more frequent GC-biased gene conversion (gBGC) on the microchromosomes [47]. Both factors have resulted in a higher average GC content of the microchromosomes (Fig. 3a and b; 44.5% vs 39.3% per 50 kb, $P < 2.2e-16$, Wilcoxon test). In addition, all chromosomes but chrZ (Fig. 3a) show generally equal expression levels between sexes; genes on chrZ are expressed twice the level in males vs females. These chromosome-wide patterns are consistent with those re-
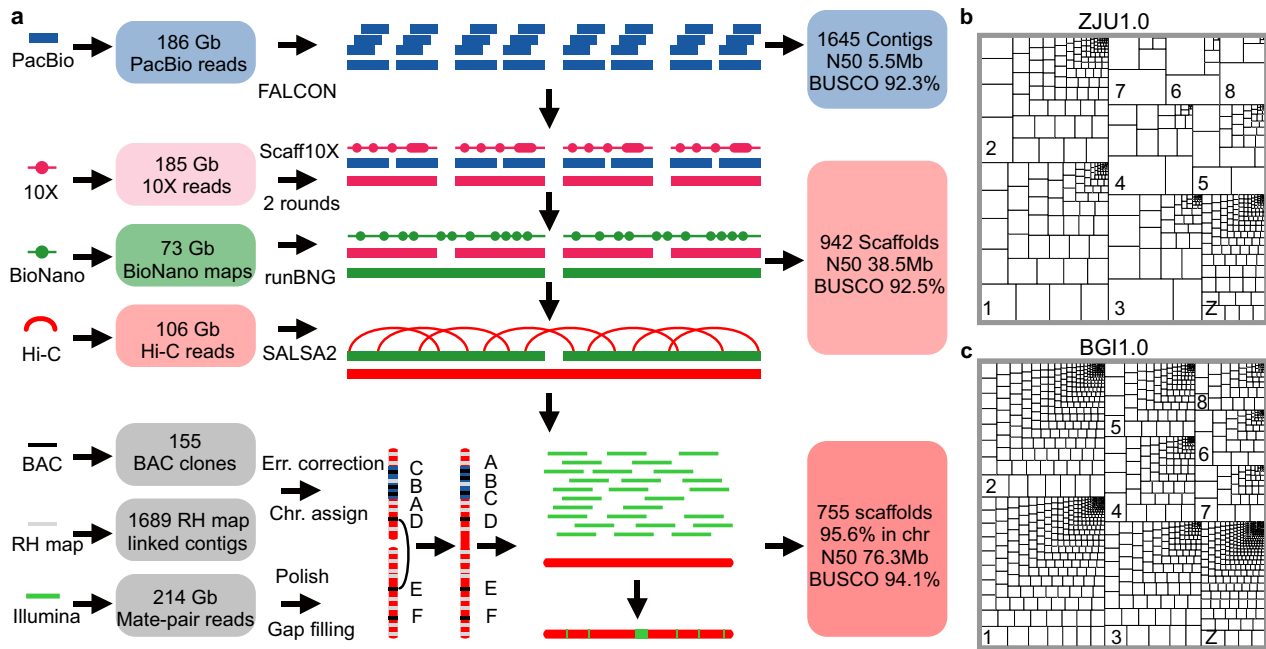
Figure 1: Genome assembly of a female Pekin duck. **a**, Our assembly pipeline uses high-coverage PacBio long reads to generate contigs, which are then sequentially scaffolded with 10X Genomics linked reads, BioNano optical maps, Hi-C paired reads, RH maps, and FISH maps to produce a chromosome-level genome for the Pekin duck. **b, c**, Treemap comparison of contigs between ZJU1.0 and BGI1.0 versions of the duck genome. The size of each rectangle of each chromosome is scaled to that of contig sequence. The bigger and fewer the internal boxes, the more contiguous the contigs.

Table 1: Comparison of genome assemblies of duck vs other birds

Parameter	Pekin duck (BGI1.0)	Pekin duck (ZJU1.0)	Chicken (Ncbi-6a)	Zebra finch (VGP)
Total length (Gb)	1.105	1.189	1.065	1.069
No. contigs	227,448	1,645	1,403	1,053
Total contig length (Gb)	1.07	1.182	1.056	1.047
Maximum contig length (Mb)	0.264	28.519	65.778	29.008
Contig N50 (Mb)	0.026	5.534	17.655	4.378
No. scaffolds	78,487	755	525	205
Longest scaffold length (Mb)	5.998	207.238	197.608	151.897
Scaffold N50 (Mb)	1.234	76.269	82.53	70.879
Total gap length (Mb)	35.08	4.378	9.784	21.569
Anchored into chromosomes (%)	25.9	95.6	98.6	97.2
Gap content (%)	3.17	0.37	0.92	2.02
BUSCO (%)	91.5	94.2	95.1	95.1

ported in other birds regarding the differences between micro- and macrochromosomes, and a lack of global dosage compensation on avian sex chromosomes [1, 48, 49].

The completeness of our new duck genome is also demonstrated by its assembled centromeres (mean length 443.3 kb) and telomeres (mean length 73.7 kb), which were annotated by a cytogenetically verified *Anseriformes* centromeric repeat (APL-HaeIII) [50] and conserved telomeric motif sequences (Supplementary Table S4 and S5). We found 22 telomeric sites among the 31 chromosomes, of which 11 were interstitial telomeric repeat (ITR) sites inside the chromosomes (Fig. 3a and b, green arrowheads). Consistent with the reported karyotypes of duck and other birds [50, 51], almost all microchromosomes are acrocentric, indicated by their positions in the centromeric region. Both macro- and microchromosomes' centromeres are enriched for CR1-J2_Pass repeats (Supplementary Fig. S7), but microchromosome centromeres are specifically enriched for the LTR re-

peat GGERVL-A-int (Fig. 3b, Supplementary Fig. S8). Such an interchromosomal difference of centromeric repeats has been reported in other birds and reptiles [52, 53] and is hypothesized to constitute the genomic basis for the spatial segregation of microchromosomes vs machrochromosomes, respectively, in the interior vs peripheral territories of the nucleus [54, 55]. Given their more aggregated spatial organization in the nuclear interior, microchromosomes exhibit an unusual pattern of more frequent interchromosomal interactions measured by the Hi-C data compared with macrochromosomes (Supplementary Fig. S9), consistent with the reported pattern of microchromosomes of chicken and snakes [56, 57].

To examine whether the different genomic landscape between micro- vs macrochromosomes would underlie different frequencies or molecular mechanisms of intragenomic rearrangements during evolution, we used our newly produced chromosomal genome of emu (with a similar assembly pipeline to be
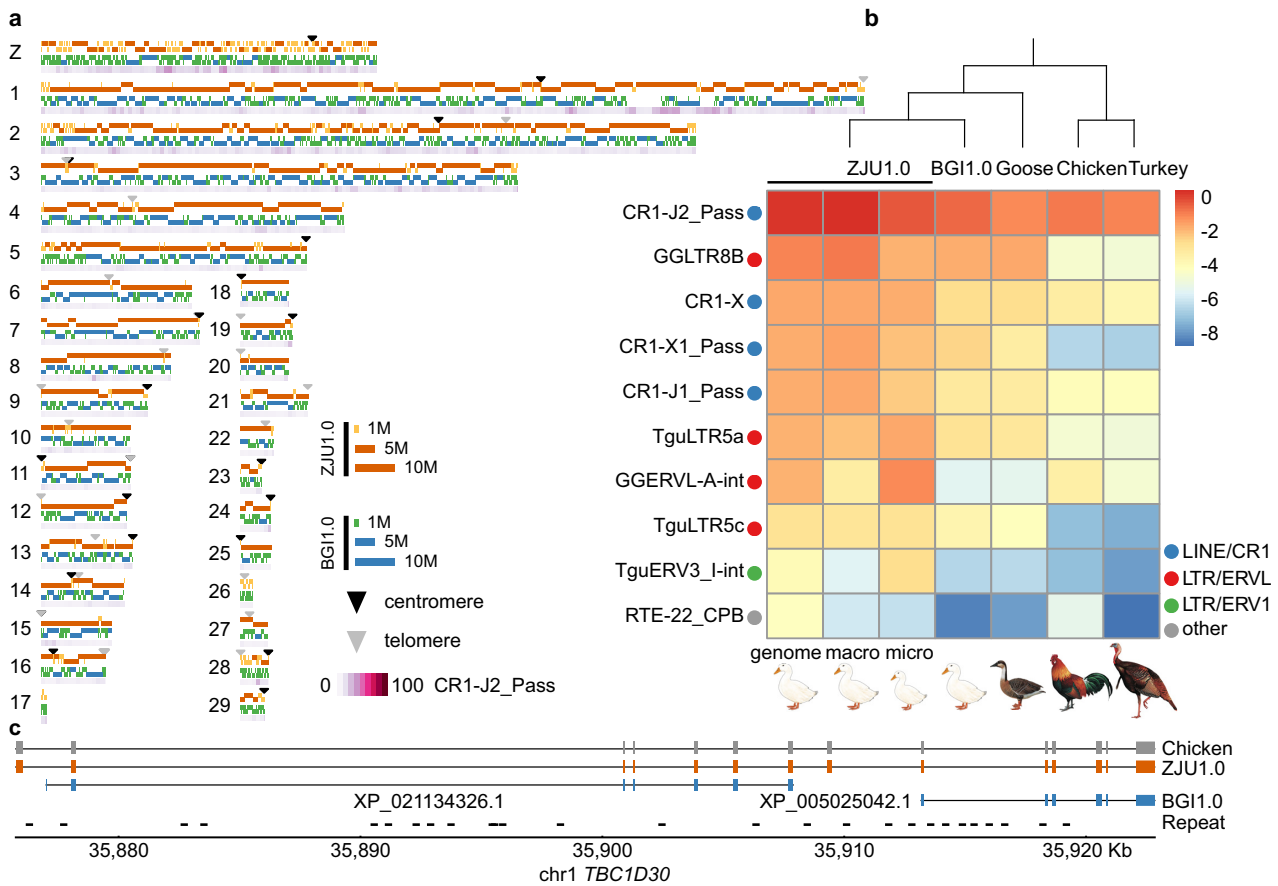
Figure 2: Comparing the new duck genome to other avian genomes. **a,** Schematic plot of each chromosome, showing the mapped contigs of ZJU1.0 (orange/yellow) and BGI1.0 (blue/green), putative centromeres (black triangles), and telomeres or interstitial telomeric sequences (grey triangles), and the most abundant repeat CR1-J2_Pass present in the gap regions of BGI1.0 (purple gradient). **b,** Comparisons of the top 10 most abundant repeats in the duck genome (ZJU1.0 whole genome, macrochromosomes, microchromosomes, and BGI1.0 assembly) to other Galloanseriformes bird genomes (goose, chicken, turkey). The more red, the higher the proportion of assembled repeat content. **c,** An example gene annotation improvement showing that 2 genes in the BGI1.0 genome are really 1 gene in the ZJU1.0 genome and were fragmented into 2 because of low resolution of repeat sequences disrupting the previous genome assembly of exons.

reported in a companion article [57]) as the outgroup, and identified 80 inversions on 26 chromosomes (>10 kb, median size 1.5 Mb, Supplementary Table S7) that occurred in the duck or *Anseriformes* lineage after it diverged from chicken in the past 72.5 MY [34] (Fig. 3c and d). The average inversion rate (1.1 inversion events or 3.1 Mb inverted regions per MY) of Pekin duck is lower than that of 1.5–2.0 events or 6.6–7.5 Mb per MY between flycatcher and zebra finch [12], reflecting more frequent intragenomic rearrangements in the passerines [58, 59]. There are 46 inversions on the duck macrochromosomes and 34 inversions on the microchromosomes, translating to 0.63 and 0.47 inversion events per MY, or 1.96 and 1.09 Mb inverted sequence per MY, respectively. A lower rate and shorter spanned length of inversions on the microchromosomes is probably related to their higher densities of genes and CNEs [60], because of the natural selection against inversions that disrupt these functional elements. Indeed, previous studies examining the breakpoint regions of genomic rearrangements of birds and mammals found that they tend to be devoid of CNEs [5, 61–63]. We also found that different families of TEs are significantly ($P < 2.2e{-}16$) enriched at the inversion breakpoints of macro- vs microchromosomes relative to other genomic regions (Supplementary Fig. S10), suggesting that they play an important role in mediating the inversions. However, we did not find a higher recombination rate at

the breakpoint regions (Supplementary Fig. S11), unlike that reported previously in flycatcher and zebra finch [12, 15].

Comparative analyses of topological chromatin domain architectures

Chromosomal inversions have attracted great interest from evolutionary biologists because they play an important role in local adaptation, speciation, and sex chromosome formation [64]. We found that the duck- or *Anseriformes*-specific inversions (Fig. 3c and d) are enriched for genes that function in immunity-related pathways (Fig. 4a, e.g., "defense response to virus," "G-protein coupled receptor pathway"; $P < 0.0001$, Fisher exact test), which may account for the known divergent susceptibility between chicken and duck to avian influenza virus. Indeed, RNF135 located on chr19, one of the ubiquitin ligases that regulate the RIG-I pathway responsible for the avian influenza virus response in ducks [65], is located in a duck-specific inversion.

To systematically evaluate the functional impacts of the identified duck- or *Anseriformes*-specific inversions, we examined whether there were any relationships with TAD units, as well as their enclosed gene expression patterns compared to chicken. Similar to mammals [66], the boundaries of duck TADs are also characterized by a significant enrichment of putative
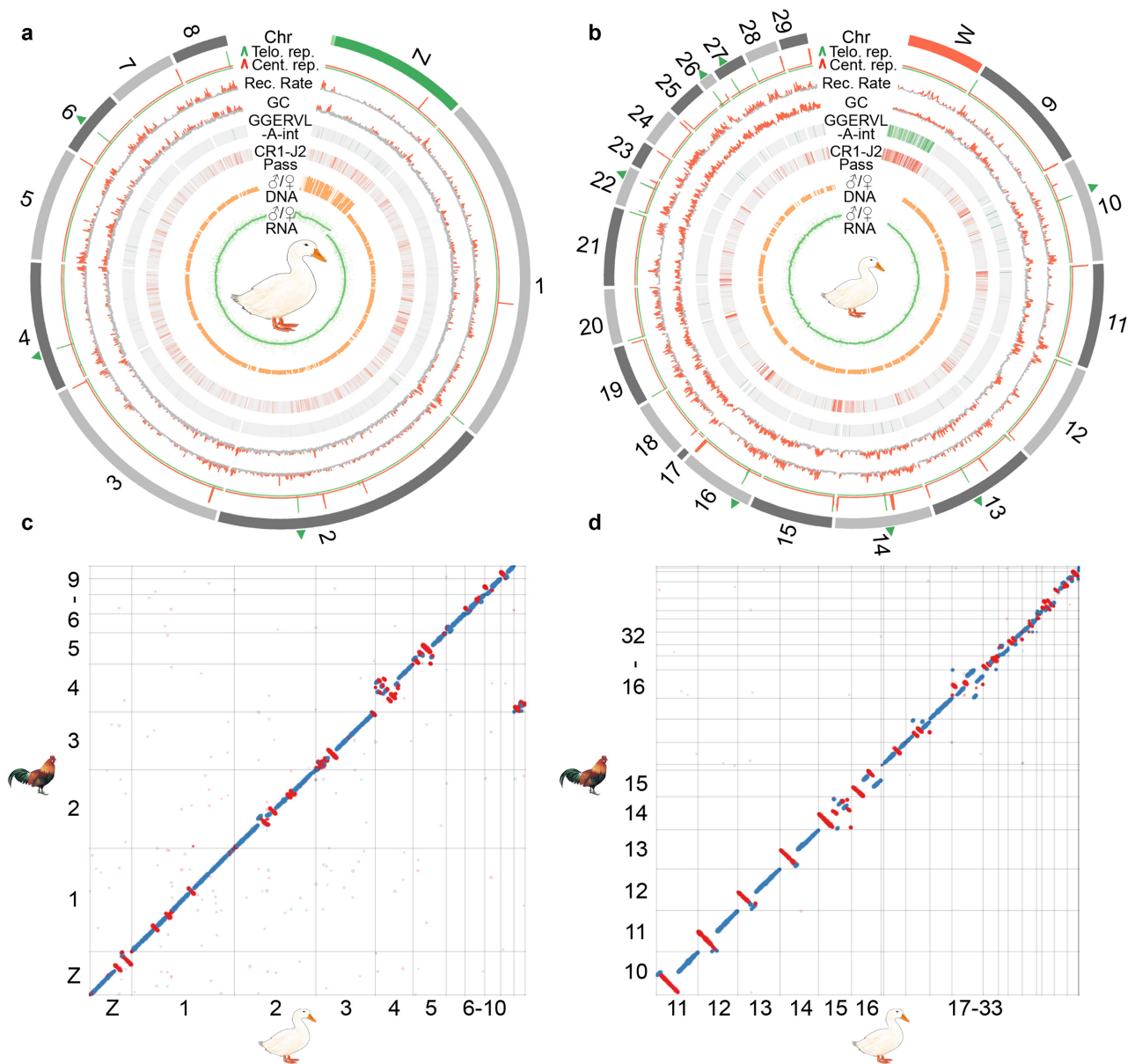
Figure 3: Evolution of the duck macro- and microchromosomes. From the outer to inner rings: the macro- (**a**) and microchromosomes (**b**), together with Z/W chromosomes (green/red), and the pseudoautosomal regions (PARs) labeled with light green at the tip of chrZ. Interstitial telomere sequences are labeled with green triangles on the chromosome. Putative centromeres (red lines) and telomeres (green lines) were inferred by the enrichment of centromeric and telomeric repeat copies, which show a sharp peak. We then show the recombination rate and GC content calculated in non-overlapping 50-kb windows, as well as 2 repeat families (GGERVL-A-int and CR1-J2 Pass) that we identified to be enriched at centromeric regions and chrW. We also show the male vs female (M/F) ratios of Illumina DNA sequencing coverage in non-overlapping 50-kb windows, M/F expression ratios (each green dot as 1 gene) of the adult brain tissue, and the smoothed line. **c, d.** Dot plots show the inversions between chicken and duck genome for both macro- and microchromosomes.

binding sites of insulator protein CTCF (Supplementary Fig. S12), an enrichment of broadly expressed housekeeping genes (Supplementary Fig. S13), and coincide with the transitions between active (A) and inactive (B) chromatin compartments (Supplementary Fig. S14). The diverse types of TAD boundaries of duck are not mutually exclusive (Fig. 4b) and suggest conserved mechanisms of TAD formation between birds and mammals [31]. The presence of putative CTCF binding sites, particularly with excessive pairs of binding sites in convergent orientation ("loop anchors") at the duck TAD boundaries (Supplementary Fig. S15a and b), suggested an active "loop extrusion" mechanism involving both the extruding factors cohesin protein complex along

chromatin and the counteracting CTCF protein [67]. In support of this, TAD boundaries that overlap with DNA loops have a significantly higher density of putative CTCF binding sites than any other TAD boundaries (Supplementary Fig. S15c). The overlap pattern between the TAD boundaries with the active/inactive compartment transition implies that self-organization of different chromatin types, probably driven by heterochromatin [68], underlies TAD formation. Finally, active transcription of genes [69] or TEs [70] has recently been discovered to account for TAD formation in mammals. We indeed found that various TEs located at the TAD boundaries have a significantly higher expression level (P < 0.01, Wilcoxon test) than their copies elsewhere
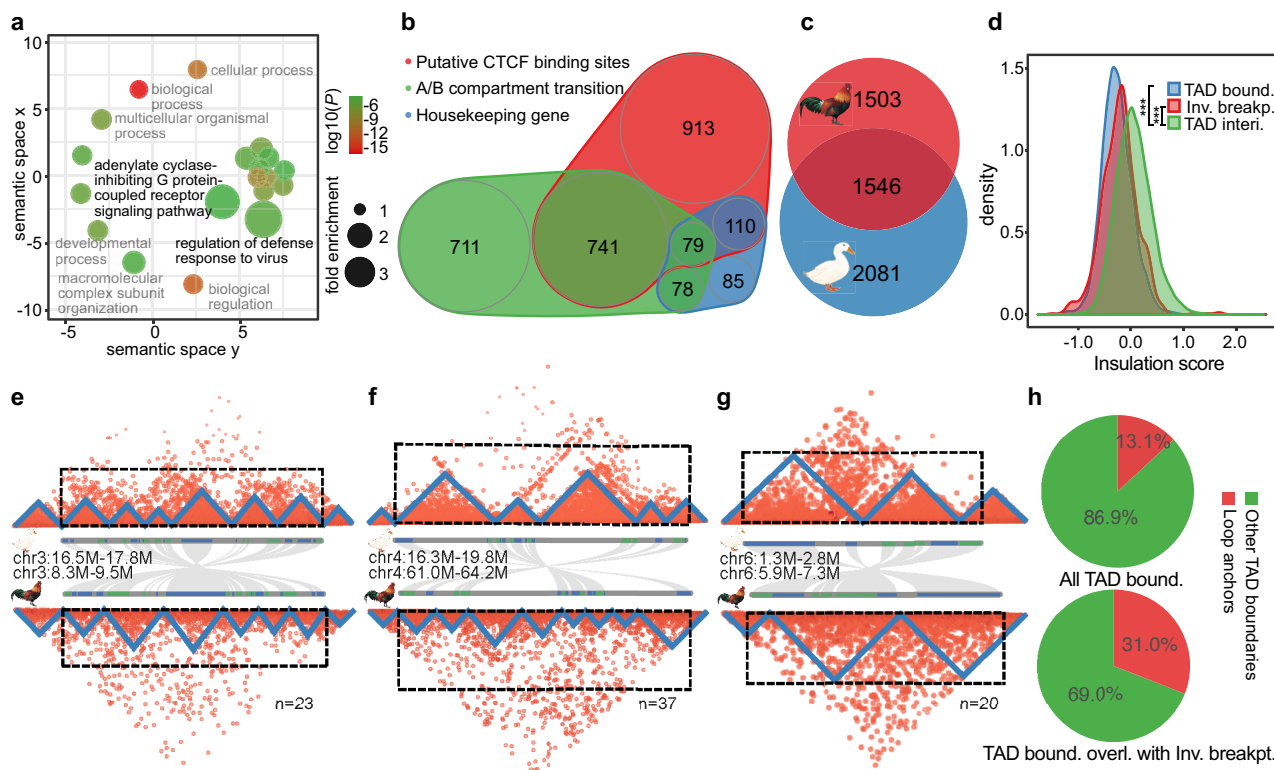
Figure 4: Genome inversions and topologically associated domains. **a,** Enriched GO terms of the genes included in the duck-specific inversions. The x- and y-axes measure the GO term semantic similarities, which are used to remove the GO redundancies. **b,** Scaled Venn diagram shows the different compositions of TAD boundaries in duck. **c,** Scaled Venn diagram shows the TAD boundaries shared between chicken and duck. **d,** Inversion breakpoint regions tend to show a significantly lower insulation score than the TAD interior regions. **e–g,** Hi-C heat maps, with each triangle structure indicating 1 TAD, along with the gene (blue or green bars) synteny plot between chicken and duck. Three examples are presented to show the impact of inversions between duck and chicken on TAD structure, with both inversion breakpoints (e), 1 inversion breakpoint (f), and no breakpoint (g), overlapped with the TAD boundaries. We also show the numbers of inversions that fit into each category. **h,** Pie charts showing that TAD boundaries that overlap with inversion breakpoints (bottom) have a higher percentage of loop anchors than others (top).

in the genome. However, these boundary TEs generally show a lower population frequency and a higher level of segregating sequence polymorphism ($P < 0.05$, Wilcoxon test) in their flanking sequences compared to the same families of TEs elsewhere (Supplementary Fig. S16), indicating that they are not under selection to fixation and may have been recently inserted into the TAD boundaries. In addition, all the assembled centromere regions of metacentric chromosomes, and intriguingly 4 of 11 ITRs (Fig. 2a and b), coincide with the TAD boundaries (Supplementary Figs. S7 and S17). This highlighted the uncharacterized role of ITRs in demarcating the functional domains in the chromosomes yet to be functionally tested in the future.

We hypothesize that the TAD units or TAD boundaries are probably under strong selective constraint during evolution. This is suggested by some congenital diseases and cancer cases caused by disruptions of TADs through structural variations [71], and also sharing of TAD boundaries between distantly related species [66, 72]. A substantial proportion (42.6%) of duck TAD boundaries are shared with those of chicken (Fig. 4c). This is probably an underestimate given that different tissues of Hi-C data were used here to identify TADs for the 2 bird species. A comparable level of conservation of human TAD boundaries (53.8%) has also been observed with mouse [66], and expectedly a lower level (26.8%) of conservation has been observed between human and chicken [56]. The other evidence of strong selective constraints acting on the integrity of TADs comes from our findings here on the pattern of chromosomal inversion break-

points of duck, whose TAD insulation scores are significantly ($P < 2.2e-16$, Wilcoxon test) lower (Fig. 4d) than for the TAD interior regions. That is, inversions more often precisely occurred at the TAD boundaries rather than within the TADs, i.e., disrupting the pre-existing TADs. Only one-third of the detected inversions have both their breakpoints located within the TADs, whereas the remaining two-thirds have both or 1 of their breakpoints overlapping with the TAD boundaries (Fig. 4e–g). Novel TAD boundaries that were created by the duck-specific inversions (e.g., Fig. 4g) tend to have significantly higher insulation scores, i.e., weaker insulation strengths than those that are conserved between duck and chicken (Supplementary Fig. S18). This suggests that natural selection may more frequently target evolutionarily older and stronger TAD boundaries. We have to point out that the alternative explanation for the overlap between the TAD boundaries and inversion breakpoints (Fig. 4e) is that chromatin loop anchors bound by CTCF protein are more likely genomic fragile sites vulnerable to DNA double-strand breaks [73] that induce the inversions. Consistent with this explanation, we found that the TAD boundaries that overlap with inversion breakpoints (Fig. 4h, bottom) have a significantly ($P < 0.001$, χ^2 test) higher percentage of loop anchors than others (Fig. 4h, top).

Because the novel TADs generated by chromosome inversions (e.g., Fig. 4g) may create aberrant or new promoter-enhancer contacts, and consequently divergent gene expression during evolution, we further compared the levels of gene expression divergence in the conserved TADs vs those novel TADs that

encompass inversion breakpoints between chicken and duck. Interestingly, genes that are close to the novel TAD boundaries created by inversions only show slightly but not significantly higher levels of expression divergence than the genes located in the conserved TADs, except for certain tissues (Supplementary Fig. S19). This reflects that the TAD boundary changes have only affected a few genes' expression patterns. It can be also explained by other regulatory divergences (e.g., in *cis*-elements) within the conserved TADs during the long-term divergence between chicken and duck, which have increased the target genes' expression divergence to the same degree as that in the novel TADs.

Sex chromosome evolution of Pekin duck

The Pekin duck provides a great model for understanding the process of avian sex chromosome evolution because the degree of differentiation of its sex chromosomes is between those of ratites and chicken [27]. Previous comparative cytogenetic work found that using FISH to probe chicken chrZ cannot produce hybridization signals on chicken chrW because of their great sequence divergence but in contrast can paint the entire chrW of duck and ostrich, suggesting that substantial sequence homology has been preserved between the Z/W chromosomes of the 2 species since the recombination was suppressed [27, 66]. The size of duck chrW is nevertheless smaller (estimated size 51 Mb) [74, 75] compared with chrZ, probably because of extensive large deletions.

Our new duck genome has assembled most of its chrZ derived from 53 scaffolds, except for 1.3 Mb unanchored sequences, into 1 continuous sequence 84.5 Mb long (Supplementary Fig. S20). The size of duck chrZ is similar to that of published chicken chrZ (82.5 Mb [76]).

We determined a 2.2-Mb-long PAR at the tip of chrZ (Fig. 5a), based on its equal read coverage between sexes. This is consistent with previous cytogenetic work showing only 1 recombination nodule concentrated at the tip of the female duck sex chromosomes [77]. Consistently, the PAR shows a significantly ($P < 2.2e-16$, Wilcoxon test) higher rate of recombination than the remaining Z-linked sex-differentiated regions (SDRs) that do not have recombination in females (Fig. 5a). The distribution of GC content also exhibits a sharp shift at the PAR boundary because of the effect of gBGC (Supplementary Fig. S21). The evolution of chicken chrZ is marked by the acquisition of large tandem arrays of 4 gene families that are specifically expressed in the testes [18]. In contrast, we did not find similar tandem arrays of testis genes on chrZ of duck, and all of the 4 Z-linked chicken testis gene families are located on the autosomes of duck (Supplementary Fig. S22).

The assembled duck chrW assembly contains 36 scaffolds with a total length of 16.7 Mb (approximately one-third of the estimated size), all of which are almost exclusively mapped by female reads (Supplementary Fig. S20). It marks an 8.8-fold increase in size compared to our previous assembly using Illumina reads [20, 78] and is much longer than the most recent assembly of chicken chrW (6.7 Mb) [22]. We have annotated a total of 71 duck W-linked SDR genes, and all of them are single-copy genes, compared to 27 single-copy genes and 1 multicopy gene on the chicken chrW, with 20 genes overlapped between the two (Fig. 5b). The only multicopy chicken W-linked gene *HINTW* with ~40 copies [22] is present as a single-copy gene on the duck chrW. These results indicate that duck and chicken have independently evolved their sex-linked gene repertoire since their species divergence. The duck chrW retained more genes than chicken and represents an intermediate stage of avian sex chromosome evolution between those of ratites and chicken.

Owing to the intrachromosomal rearrangements of chrZ, most birds (including duck) except for ratites have retained few ancestral gene syntenies of their proto-sex chromosomes before the suppression of homologous recombination [20, 78], and exhibit dramatic reshuffling of their old evolutionary strata. To accurately reconstruct the history of duck sex chromosome evolution, we used a newly produced chrZ assembly of emu in our group to approximate the avian proto-sex chromosomes. Almost all (15.2 Mb [91%]) of the duck chrW sequences can be aligned to the chrZ of emu, and form a clear pattern of 4 evolutionary strata. This is manifested as a gradient of Z/W pairwise sequence divergence, i.e., a gradient of the age of strata along the chrZ, which is named from the old to the young, as stratum 0, S0 to S3 (Fig. 5a). Within each stratum, chrW scaffolds of similar levels of sequence divergence are clustered and separated from the neighboring strata with different divergence levels (Supplementary Fig. S23). The genes enclosed in each stratum are consistent with our previous annotation of the duck evolutionary strata based on the BGI1.0 genome and show a consistent gradient of synonymous substitution rates (Supplementary Fig. S24) between the Z- and W-linked alleles according to the age of the strata where they reside. We did not find any chrW scaffolds that span the boundaries of neighboring strata, probably because of some complex repeat sequences (e.g., CR1-J2_Pass) that accumulate at the boundary. Interestingly, the inferred boundaries between evolutionary strata on chrZ, i.e., the breakpoints between the inverted regions within or between the strata (8 of 9 boundaries shown in Fig. 5a) tend to have a low TAD insulation score, i.e., to overlap with TAD boundaries or loop anchors (Supplementary Fig. S25). This again strongly supports the idea that loop anchors or TAD boundaries are likely the genomic fragile regions that induced inversions.

Because of the lack of recombination, (30 [42.3%]) of W-linked genes probably have become pseudogenes or long non-coding RNA genes owing to frameshift mutations or premature stop codons (Supplementary Fig. S26). The other pronounced signature of functional degeneration of chrW is accumulation of TEs. The duck chrW shows a much higher genomic proportion (46.5% vs 10.1%) and a different composition of TEs compared with the genome average (Fig. 5c). The W-linked repeats are concentrated in those families that have specifically expanded their copy numbers in the duck after it diverged from other *Anseriformes* (Supplementary Fig. S27, Supplementary Table S8). Among them, different TE families exhibit opposing trends of colonizing the different evolutionary strata of different ages (Fig. 5d, Supplementary Fig. S28). TE families that have been propagating since the ancestor of Neoaves (e.g., CR1-J2_Pass, Supplementary Fig. S6) [79] are more enriched in the older strata, while TE families that were specifically propagated in the duck (e.g., TguERV3_I-int, Fig. 2b) are more enriched in the younger strata. This suggests that older evolutionary strata might be saturated for old TEs relative to TEs with recent activities. Particularly, duck or *Anseriformes* enriched repeats are nested with each other and form 38 palindromes dispersed across the entire chrW (Fig. 5e). Their lengths range from 15.2 to 345.5 kb (Supplementary Table S9), together comprising 3.74 Mb (22%) of the assembled duck chrW sequence.
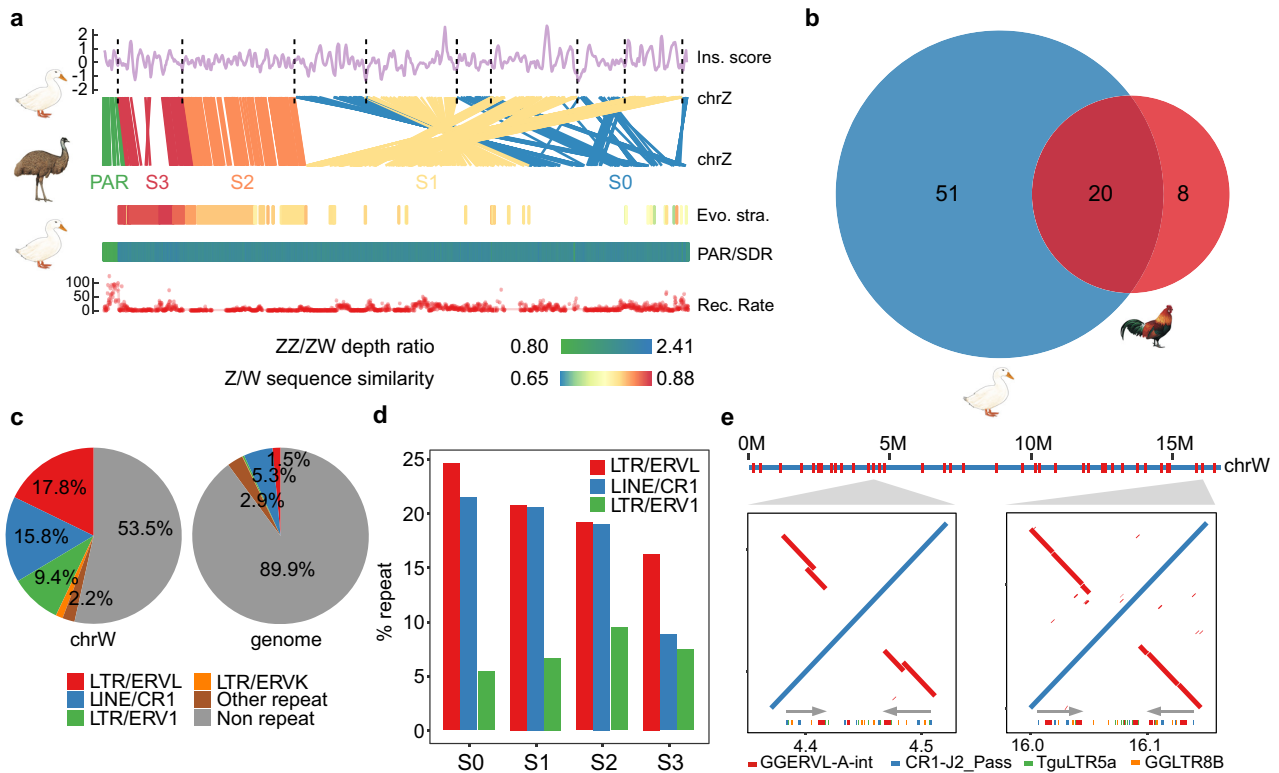
Figure 5: Sex chromosome evolution in Pekin duck. **a**, Evolutionary strata analyses of the duck sex chromosomes. From top to bottom: the breakpoints of genomic rearrangements (dashed lines) between emu and duck chrZ tend to have a lower insulation score; gene synteny between the emu and duck Z chromosomes; alignment of the duck chrW scaffolds against the emu chrZ reveals a pattern of evolutionary strata, with each scaffold showing the color-scaled sequence divergence levels between the duck chrW vs the emu chrZ; PAR (light green)/SDR (dark green) composition inferred by the ratio of male vs female Illumina DNA sequencing depth with the color scaled to the ratio value; a higher recombination rate in the duck PAR than in SDR. **b**, Scaled Venn diagram shows the chrW genes shared between duck and chicken. **c**, Comparing the repeat content of the duck chrW to the whole genome. **d**, Different enrichment trends of chrW repeats at different evolutionary strata. **e**, Palindrome structure of duck chrW. Palindromes are labeled across the entire chrW (red), ordered according to the duck chrZ. Shown are alignment plots of 2 zoomed-in examples of palindromes (red inversions and grey arrows) for their repeat content (colors below grey arrows). LINE: long interspersed nuclear element; LTR: long terminal repeat.

Discussion

Birds and mammals diverged >300 MY ago and are known to have a very different chromosomal composition [1]. Our comparative analyses of the nearly complete genome of the Pekin duck revealed that TADs are conserved functional and evolutionary chromosome units in both birds and mammals. The 40–50% of the TADs shared between chicken and duck is comparable to the proportions shared between human and mouse [66]. This is also consistent with the highly conserved pattern of replication domains between human and mouse [80], which have a nearly one-to-one correspondence with TADs [81]. The interspecific overlap of TADs implies strong selection on TAD integrity during evolution. In this work, we identified many chromosomal inversions between chicken and duck that were previously uncharacterized because of the fragmented Illumina-based duck genome. Consistent with selection against the genome rearrangements disrupting the TADs, there are disproportionately more chromosome inversions that occurred at the TAD boundaries than within the TADs. This extensive overlap between TAD boundaries and inversion breakpoints likely reflects the susceptibility of TAD boundaries to DNA double-strand breaks. TADs can form either by self-organization of genomic regions of the same epigenetic state or by active loop extrusion involving the cohesin and insulator protein CTCF [67]. This is indicated by the transition between active and inactive chromatin compartments or

the enrichment of CTCF binding sites at the TAD boundaries of duck (this study), chicken [56], and mammals [66]. It has been recently shown that type II topoisomerase B (TOP2B), which releases DNA torsional stress by transiently breaking and rejoining DNA double-strands, physically interacts with cohesin and CTCF and colocalizes with the TAD boundaries with convergent CTCF binding site pairs (loop anchors) [73]. This probably frequently exposes the TAD boundaries to double-strand breaks and induces chromosomal inversions involving the entire TAD. This mechanism may also account for the common genomic fragile sites found in both birds and mammals that have been reused during evolution to mediate genomic rearrangements [7, 11, 13, 82]. Overall, despite divergent chromosomal composition, our results suggested conserved mechanisms of chromosome folding and rearrangements between birds and mammals.

The 2 clades of vertebrates also evolved convergent sex chromosome architectures. Our finding that the duck chrW has suppressed recombination with chrZ in a stepwise manner is similar to the pattern of evolutionary strata between the human X and Y chromosomes [19]. As the result of recombination suppression, the duck chrW has accumulated massive TEs, some of which formed dispersed palindromes along the chromosome. Unlike other sex-specific palindromes reported in primates, birds, and willow [25, 26, 83–85], the duck palindromes do not seem to contain functional genes that have robust gene

expression. This suggests that the gene copies contained in the palindromes may have nevertheless become pseudogenes, despite the repair mechanism mediated by gene conversions between gene copies within the palindromes. Or the involved genes have already become a pseudogene before being amplified by the palindromes. An interesting contrast is that we did not find palindromes on our recently assembled emu chrW with a similar dataset and pipeline, which evolves much more slowly than chrWs of chicken and duck. Palindromes were also not reported in the recently evolved *Drosophila miranda* chrY [86]. These results suggest that sex-linked palindromes are a feature of strongly differentiated sex chromosomes that have accumulated abundant TEs. The palindromes may retard the functional degeneration of Y- or W-linked genes but can also promote large sequence deletions by intrachromosomal recombination. The latter probably contributed to the much smaller size of chrW relative to the chrZ of duck, despite the fact that many more genes than in the chrW of chicken have been preserved.

Methods

Genome assembly

High molecular weight (HMW) DNA was extracted from the liver of a female Pekin duck (*Anas platyrhynchos*, Z2 strain, from Pekin duck breeding farm, Beijing, China) with Gentra Puregene Tissue Kit (Qiagen No. 158667). Libraries for SMRT sequencing were constructed as described previously [87]. In total, 115 SMRT cells were sequenced with PacBio RS II (PacBio RS II Sequencing System, RRID:SCR_017988) and Sequel platform (PacBio Sequel System, RRID:SCR_017989) (PacBio), and 186 Gb (143× genome coverage) subreads with an N50 read length of 14,262 bp were produced. The same DNA was used to generate a linked-reads library following the protocol on the 10X Genomics Chromium platform (Genome Library Kit & Gel Bead Kit v2 PN-120258, Genome HT Library Kit & Gel Bead Kit v2 PN-120261, Genome Chip Kit v2 PN-120257, i7 Multiplex Kit PN-120262). This 10X library was subjected to the DNBSEQ-G400 platform (DNBSEQ-G400, RRID:SCR_017980) for sequencing and 185 Gb PE150 (142× genome coverage) reads were collected. HMW DNA of a male Pekin duck was used to produce the BioNano library with the enzyme Nt.BspQI. After the enzyme digestion, segments of the DNA molecules were labeled and counterstained following the IrysPrep Reagent Kit protocol (Bionano Genomics) as described previously [88]. Libraries were then loaded into IrysChips and run on the Irys imaging instrument, and a total of 73 Gb (56× genome coverage) optical map data were generated. We used the HMW DNA from the breast muscle of a male Pekin duck to prepare the Hi-C library using the restriction enzyme MboI with the protocol described previously [30] and produced a total of 106 Gb (82× genome coverage) paired-end reads of 50 bp long on the Illumina HiSeq X Ten platform (Illumina HiSeq X Ten, RRID:SCR_016385). We used the published genome resequencing data of 14 female and 11 male duck individuals from Zhou et al. [46]. We collected the total RNAs of adult tissues (brain, kidney, gonads) of both sexes using TRIzol® Reagent (Invitrogen No. 15596–018) following the manufacturer's instructions. Then paired-end libraries were constructed using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) and 3-Gb paired-end reads of 150 bp were produced for each library.

We generated the genome assembly with the modified VGP (v1.0) pipeline [29]. In brief, we produced the contig sequences derived from the PacBio subreads using FALCON [89] (FALCON, RRID:SCR_016089) (git 12072017) followed by 2 rounds of assem-

bly polishing by Arrow [90], and then by Purge Haplotigs [91] (bitbucket 7.10.2018) to remove false haplotype and homotypic duplications. The contigs were then scaffolded first with 10X linked reads using Scaff10X [92], then with BioNano optical maps using runBNG [93] (v1.0.3), and finally with Hi-C reads using SALSA [94] (v2.0). We performed gap filling on the scaffolds with the Arrow-corrected PacBio subreads by PBJelly (PBJelly, RRID:SCR_012091) [95], and 2 rounds of assembly polishing with Illumina reads by Pilon (Pilon, RRID:SCR_014731) [96] (v1.22). All the scripts used were from the VGP assembly pipeline [29]. We evaluated the genome completeness using BUSCO (BUSCO, RRID:SCR_015008) [97] (v3.0.2). In brief, 4,915 BUSCO proteins of birds from OrthoDB v9 were used in the evaluation.

Genome annotation

We combined evidence of protein homology, transcriptome, and *de novo* prediction to annotate the protein-coding genes. First, we aligned the protein sequences of human, chicken, duck, and zebra finch collected from Ensembl (Ensembl, RRID:SCR_002344) [98] (release 90) to the reference genome using TBLASTN v2.2.26 (TBLASTN, RRID:SCR_011822) [99] with the following parameters: -F F -p tblastn -e 1e-5. The resulting candidate genes were then refined by GeneWise v2.4.1 (GeneWise, RRID:SCR_015054) [100]. For each candidate gene, only the one with the best score was kept as the representative model. We filtered the candidate genes if they contained premature stop codons or frameshift mutations reported by GeneWise [100], if they were single-exon genes with a length shorter than 100 bp or multi-exon genes with a length shorter than 150 bp, or if the repeat content of the CDS sequence was >20%. Second, to obtain the *de novo* gene models, we used the protein queries to train Augustus v3.3 (Augustus, RRID:SCR_008417) [101] with default parameters. We also used all available RNA-seq reads to construct transcripts using Trinity v2.4.0 (Trinity, RRID:SCR_013048) [102]. Finally, all the gene models from the above 3 resources were merged into a nonredundant gene set with EVidenceModeler v1.1.1 (EVidenceModeler, RRID:SCR_014659) [103]. We used RepeatMasker v4.0.8 (RepeatMasker, RRID:SCR_012954) [104] with the following parameters: -s -pa 4 -xsmall, and RepBase [105] (v21.01) queries to annotate the repetitive elements.

To annotate the putative centromeres, we searched the genome with the reported 190-bp duck centromeric repeats [50] using TRFinder [106] (v4.09) with the following parameters: 2 5 7 80 10 50 2000. A genome-wide distribution of the 190-bp sequences was generated by binning the genome with a 50-kb non-overlapping window to find the local enrichment of copy numbers, which was defined as the putative centromeres. For telomeres, we used the known vertebrate consensus sequence [107] "TTAGGG/CCCTAA" to search for the clusters of consensus sequence on both strands from the above tandem repeat annotation. Consensus sequence–enriched genomic blocks in a 50-kb window were then defined as the putative telomere regions.

Building the chromosomal sequences and identifying the sex-linked sequences

To anchor Pekin duck scaffolds onto chromosomes, we first collected the ordered 1,689 RHmap linked contigs [32] and 155 BAC clone sequences [33] from the previous studies. We aligned these sequences, as well as the Illumina duck genome [36] (BGI1.0), to the new duck scaffolds that we generated by nucmer [108] (v3.23) and only kept the best hits for each sequence. Scaffolds were oriented and ordered first on the basis of the RHmap con-

tigs that span >1 scaffold, then by BAC sequences whose order was determined previously by FISH, and finally by the syntenic relationship with the BGI1.0 genome. We also corrected scaffolding errors using the raw PacBio reads if the order of our scaffolds conflicted with that of RHmap or BAC sequence order (Supplementary Fig. S2).

To identify the sex-linked sequences, Illumina reads from both sexes were aligned to the scaffold sequences using BWA ALN [109] with default parameters. Read depth of each sex was then calculated using SAMtools (Samtools, RRID:SCR_002105) [110] in 5-kb non-overlapping windows, and normalized against the median value of depths per single base pair throughout the entire genome, respectively, to enable the comparison between sexes. To identify the Z-linked sequences, the depth ratio of male-vs-female (M/F) was calculated for the genomic regions mapped by reads for each sequence, with a minimum 80% coverage in both sexes, and sequences with a depth ratio ranging from 1.5 to 2.5 were assigned as Z-linked. To identify the W-linked sequences, we calculated M/F depth ratio, as well as M/F coverage ratio, and assigned scaffolds to W-linked when either ratio was within the range 0.0–0.25 as W-linked sequences (Supplementary Fig. S21). Because we do not have linkage markers on the W chromosome, we ordered the W scaffolds based on their unique aligned position with the Z chromosome using RaGOO [111] (v1.1) with default parameters. This does not reflect the actual order of W-linked sequences, which probably have rearrangements with the homologous Z chromosome, but allows us to examine the pattern of evolutionary strata.

To identify the inversions in the duck genome, genomic syntenic blocks between chicken and duck and between emu and duck were constructed using nucmer (v3.1) with the parameters -b 500 -l 20. Then inversions between chicken and duck were manually checked by plotting the dot plot between the 2 species. The duck-specific inversions were identified by excluding chicken-specific inversion, using emu as the outgroup.

Hi-C analyses

Hi-C read mapping, filtering, correction, binning, and normalization were performed by HiC-Pro v2.10.0 (HiC-Pro, RRID:SCR_017643) [112] with the default parameters. In brief, Hi-C reads of chicken [113] (sourced from FR-AgENCODE project) and duck were mapped to the respective reference genome and only uniquely mapped reads were kept. Then each uniquely mapped read was assigned to a restriction fragment and invalid ligation products were discarded. Data were then merged and binned to generate the genome-wide interaction maps at 10- and 50-kb resolution. TADs were identified by HiCExplorer [114] (v3.0) with the application hicFindTADs. First, HiC-Pro interaction maps were transformed to h5 format matrix by hicConvertFormat with the following parameters: –inputFormat hicpro –outputFormat h5. Then the h5 matrix was imported to hicFindTADs with the parameters –outPrefix TAD –numberOfProcessors 32 –correctForMultipleTesting fdr. The TAD boundaries were identified by hicFindTADs through an approach that computes a TAD insulation score. Genomic bins with low insulation scores relative to neighboring regions were defined as local minima and called the TAD boundaries. Human CTCF [115] motif was used as a query for FIMO in MEME [116] (v4.12.0) to identify the putative CTCF binding sites. CTCF density in every 10-kb non-overlapping sliding window along the genome was calculated to check its enrichment at the TAD boundaries. We identified the A/B compartments using the pca.hic function from the HiTC [117] (High Throughput Chromosome Conformation Cap-

ture analysis) R package with default parameters, and the 10-kb matrix generated by HiC-Pro as the input. We identified the chromatin loops by means of Mustache [118] with the parameters -p 32 -r 10 kb -pt 0.05, after converting the h5 format matrix to mcool matrix format by hicConvertFormat with parameters –inputFormat h5 –outputFormat mcool.

Evolutionary strata

To demarcate the evolutionary strata, all the repeat-masked duck W-linked scaffolds were aligned to emu Z chromosome using LASTZ v0.9 (LASTZ, RRID:SCR_018556) [119] with the following parameters: –step = 19 –hspthresh = 2200 –inner = 2000 –ydrop = 3400 –gappedthresh = 10 000 –format = axt, and a score matrix set for the distant species comparison. Alignments were converted into "net" and "maf" results using the UCSC Genome Browser's utilities [120]. Based on "net" and "maf" results, the identity of the aligned sequence was calculated for each alignment block with a 10-kb non-overlapped window, and then we oriented the aligned W-linked sequences along the Z chromosomes. Then we color-coded the pairwise sequence divergence level between the Z/W sequences to demarcate the evolutionary strata.

Gene expression analyses

RNA-seq reads were mapped to the duck genome by HISTA2 [121] with default parameters. Only uniquely mapped RNA-seq reads were kept and used to calculate the RPKM expression level. DESeq2 (DESeq2, RRID:SCR_015687) [122] was applied to normalize the RPKM values across different samples and finally generated an expression matrix. For each gene, we used the median expression value in each tissue to calculate the tissue specificity index TAU [123, 124]. Expression levels of TE elements were calculated using SQUIRE (v0.9.9.92) [125] with default parameters.

Data Availability

The assembly and annotation of Pekin duck has been deposited in GenBank under the Bioproject accession code PRJNA636121 (accession No. JACGAL000000000) and the emu under PRJNA638233 (accession No. JABVCD000000000). All supporting data and materials are available in the *GigaScience* GigaDB database [126].

Code Availability

Scripts used in this study are shared on GitHub at https://github.com/ZhouQiLab/DuckGenome under a MIT license.

Additional Files

Supplementary Figure S1. Length distribution of one representative Pacbio RSII SMRT cell from all 115 SMRT cells.
Supplementary Figure S2. A representative case of assembly error correction.
Supplementary Figure S3. High GC content at the gap regions of BGI1.0.
Supplementary Figure S4.Transposable elements (TE) are enriched in the filled gap regions.
Supplementary Figure S5. Comparison of repeat composition between duck and other Galloanseriformes birds.
Supplementary Figure S6. Transposition in transposition (TinT) analyses of repeats enriched in the filled gap regions.

Supplementary Figure S7. An example centromere overlapped with the TAD boundary and enriched for CR1-J2_Pass repeats.

Supplementary Figure S8. Different repeats are enriched at centromeres of macro- and micro-chromosomes.

Supplementary Figure S9. Inter-chromosomal interactions.

Supplementary Figure S10. Repeat enrichment at the inversion breakpoints.

Supplementary Figure S11. Recombination rate at the inversion breakpoints.

Supplementary Figure S12. CTCF is enriched at TAD boundaries in the duck and chicken genome.

Supplementary Figure S13. TAD boundaries tend to be enriched for broadly expressed housekeeping genes in the Pekin duck.

Supplementary Figure S14. An example of diverse types of TAD boundaries that overlap.

Supplementary Figure S15. DNA loop regions have a higher CTCF density and a higher percentage of paired CTCF sites in convergent orientation.

Supplementary Figure S16. TEs at the TAD boundaries.

Supplementary Figure S17. An example of interstitial telomere sequences that are overlapped with a TAD boundary.

Supplementary Figure S18. Disrupted TADs tend to have a higher insulation score.

Supplementary Figure S19. No significant difference of expression divergence between disrupted and conserved TAD domains.

Supplementary Figure S20. Sex chromosomes show a different coverage pattern between sexes.

Supplementary Figure S21. GC shift along PAR boundary.

Supplementary Figure S22. Chicken Z-amplicon orthologs' expression in Pekin duck.

Supplementary Figure S23. Sequence similarity between duck chrW and emu chrZ

Supplementary Figure S24. The distribution of pairwise dS values of duck sex chromosomes.

Supplementary Figure S25. Rearrangement breakpoints between duck and emu chrZ tend to have a low insulation score.

Supplementary Figure S26. Expression pattern of duck chrW genes.

Supplementary Figure S27. Repeats enriched in duck chrW.

Supplementary Figure S28. Repeats enriched at different evolutionary strata.

Supplementary Table S1. Sequencing data.

Supplementary Table S2. Statistics of contig assemblies.

Supplementary Table S3. Chromosome anchoring in ZJU1.0 assembly

Supplementary Table S4. Centromere location.

Supplementary Table S5. Telomere location.

Supplementary Table S6. Repeat content comparison between different bird genomes.

Supplementary Table S7. Inversions between duck and chicken.

Supplementary Table S8. Repeat enrichment at chrW.

Supplementary Table S9. Palindromes in duck chrW.

Abbreviations

BAC: bacterial artificial chromosome; bp: base pairs; BUSCO: Benchmarking Universal Single-Copy Orthologs; BWA: Burrows-Wheeler Aligner; CNE: conserved non-coding element; FISH: fluorescence in situ hybridization; Gb: gigabase pairs; gBGC: GC-biased gene conversion; GC: guanine-cytosine; ITR: interstitial telomeric repeat; HMW: high molecular weight; kb: kilobase pairs; LTR: long terminal repeat; Mb: megabase pairs; MY:

million years; NCBI: National Center for Biotechnology Information; PacBio: Pacific Biosciences; PAR: pseudoautosomal region; RH: radiation hybrid; RPKM: reads per kilobase of transcript per million mapped reads; SDR: sex-differentiated region; SMRT: single-molecule real-time; TAD: topologically associated domain; TE: transposable element; UCSC: University of California Santa Cruz; VGP: Vertebrate Genomes Project.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Q.Z. conceived the project and acquired the funding; J.Li, X.D., S.F., C.G., J.R., and K.W. acquired the samples and produced the data; J. Li, J.Z., J. Liu, Y.Z., C.C., L.X., and Q.Z. performed the analyses.; J.Li, Y.J., Z.Z., G.Z., E.D.J., and Q.Z. wrote the manuscript.

References

1. Zhang G, Li C, Li Q, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 2014;**346**(6215):1311–20.

2. Burt DW. Origin and evolution of avian microchromosomes. Cytogenet Genome Res 2002;**96**(1–4):97–112.

3. Burt DW, Bruley C, Dunn IC, et al. The dynamics of chromosome evolution in birds and mammals. Nature 1999;**402**(6760):411–3.

4. Griffin DK, Robertson LBW, Tempest HG, et al. The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenet Genome Res 2007;**117**(1–4):64–77.

5. Damas J, Kim J, Farré M, et al. Reconstruction of avian ancestral karyotypes reveals differences in the evolutionary history of macro- and microchromosomes. Genome Biol 2018;**19**(1):155.

6. Nanda I, Karl E, Griffin DK, et al. Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. Cytogenet Genome Res 2007;**117**(1–4):43–53.

7. O'Connor RE, Farré M, Joseph S, et al. Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. Genome Biol 2018;**19**(1):171.

8. Nishida C, Ishijima J, Kosaka A, et al. Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and delineation of chromosome rearrangements during their differentiation. Chromosome Res 2008;**16**(1):171–81.

9. Jarvis ED, Mirarab S, Aberer AJ, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 2014;**346**(6215):1320–31.

10. Les C , ed. Animal Cytogenetics, Vol. 4: Chordata 3: B. Aves. Berlin, Germany: Gebrüder Borntraeger; 1990:55–7.

11. Skinner BM, Griffin DK. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity 2012;**108**(1):37–41.

12. Kawakami T, Smeds L, Backström N, et al. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol Ecol 2014;**23**(16):4035–58.

13. Pevzner P, Tesler G. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci U S A 2003;**100**(13):7672–7.

14. Larkin DM, Pape G, Donthu R, et al. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res 2009;**19**(5):770–7.

15. Völker M, Backström N, Skinner BM, et al. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. Genome Res 2010;**20**(4):503–11.

16. Lemaitre C, Zaghloul L, Sagot M-F, et al. Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. BMC Genomics 2009;**10**:335.

17. Irwin DE. Sex chromosomes and speciation in birds and other ZW systems. Mol Ecol 2018;**27**(19):3831–51.

18. Bellott DW, Skaletsky H, Pyntikova T, et al. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature 2010;**466**(7306):612–6.

19. Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. Science 1999;**286**(5441):964–7.

20. Zhou Q, Zhang J, Bachtrog D, et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 2014;**346**(6215):1246338.

21. Cortez D, Marin R, Toledo-Flores D, et al. Origins and functional evolution of Y chromosomes across mammals. Nature 2014;**508**(7497):488–93.

22. Bellott DW, Skaletsky H, Cho T-J, et al. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. Nat Genet 2017;**49**(3):387–94.

23. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 2003;**423**(6942):825–37.

24. Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. Phil Trans R Soc Lond B 2000;**355**(1403):1563–72.

25. Davis JK, Program NCS, Thomas PJ, et al. A W-linked palindrome and gene conversion in New World sparrows and blackbirds. Chromosome Res 2010;**18**(5):543–53.

26. Zhou R, Macaya-Sanz D, Carlson CH, et al. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. Genome Biol 2020;**21**(1):38.

27. Nanda I, Schlegelmilch K, Haaf T, et al. Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination. Cytogenet Genome Res 2008;**122**(2):150–6.

28. Xu L, Wa Sin SY, Grayson P, et al. Evolutionary dynamics of sex chromosomes of paleognathous birds. Genome Biol Evol 2019;**11**(8):2376–90.

29. Rhie A, McCarthy SA, Fedrigo O, et al. Towards complete and error-free genome assemblies of all vertebrate species. bioRxiv 2020, doi:10.1101/2020.05.22.110833.

30. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009;**326**(5950):289–93.

31. Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Sci Adv 2019;**5**(4):eaaw1668.

32. Rao M, Morisson M, Faraut T, et al. A duck RH panel and its potential for assisting NGS genome assembly. BMC Genomics 2012;**13**(1):513.

33. Skinner BM, Robertson LBW, Tempest HG, et al. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. BMC Genomics 2009;**10**:357.

34. Claramunt S, Cracraft J. A new time tree reveals Earth history's imprint on the evolution of modern birds. Sci Adv 2015;**1**(11):e1501005.

35. Herrera AM, Brennan PLR, Cohn MJ. Development of avian external genitalia: interspecific differences and sexual differentiation of the male and female phallus. Sex Dev 2015;**9**(1):43–52.

36. Huang Y, Li Y, Burt DW, et al. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat Genet 2013;**45**(7):776–83.

37. Nakamura D, Tiersch TR, Douglass M, et al. Rapid identification of sex in birds by flow cytometry. Cytogenet Cell Genet 1990;**53**(4):201–5.

38. Tiersch TR, Wachtel SS. On the evolution of genome size of birds. J Hered 1991;**82**(5):363–8.

39. Takagi N, Makino S. A revised study on the chromosomes of three species of birds. Caryologia 1966;**19**(4):443–55.

40. Lu L, Chen Y, Wang Z, et al. The goose genome sequence leads to insights into the evolution of waterfowl and susceptibility to fatty liver. Genome Biol 2015;**16**:89.

41. Warren WC, Hillier LW, Tomlinson C, et al. A new chicken genome assembly provides insight into avian genome structure. G3 (Bethesda) 2017;**7**(1):109–17.

42. Islam FB, Uno Y, Nunome M, et al. Comparison of the chromosome structures between the chicken and three anserid species, the domestic duck (*Anas platyrhynchos*), Muscovy duck (*Cairina moschata*), and Chinese goose (*Anser cygnoides*), and the delineation of their karyotype evolution by comparative chromosome mapping. J Poult Sci 2014;**51**(1), doi:10.2141/jpsa.0130090.

43. Peona V, Blom MPK, Xu L, et al. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. Mol Ecol Resour 2020, doi:10.1111/1755-0998.13252.

44. Botero-Castro F, Figuet E, Tilak M-K, et al. Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. Mol Biol Evol 2017;**34**(12):3123–31.

45. Korlach J, Gedman G, Kingan SB, et al. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. Gigascience 2017;**6**(10):gix085.

46. Zhou Z, Li M, Cheng H, et al. An intercross population study reveals genes associated with body size and plumage color in ducks. Nat Commun 2018;**9**(1):2648.

47. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genom Hum Genet 2009;**10**:285–311.

48. Hillier LW, Miller W, Birney E, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 2014;**423**(10):695–777.

49. McQueen HA, McBride D, Miele G, et al. Dosage compensation in birds. Curr Biol 2001;**11**(4):253–7.

50. Uno Y, Nishida C, Hata A, et al. Molecular cytogenetic characterization of repetitive sequences comprising centromeric heterochromatin in three *Anseriformes* species. PLoS One 2019;**14**(3):e0214028.

51. Wójcik E, Smalec E. Description of the mallard duck (*Anas platyrhynchos*) karyotype. Folia Biol (Krakow) 2007;**55**(3–4):115–20.

52. Matzke MA, Varga F, Berger H, et al. A 41–42 bp tandemly repeated sequence isolated from nuclear envelopes of chicken erythrocytes is located predominantly on microchromosomes. Chromosoma 1990;**99**(2):131–7.

53. Tanaka K, Suzuki T, Nojiri T, et al. Characterization and chromosomal distribution of a novel satellite DNA sequence of Japanese quail (*Coturnix coturnix japonica*). J Hered 2000;**91**(5):412–5.

54. Maslova A, Zlotina A, Kosyakova N, et al. Three-dimensional architecture of tandem repeats in chicken interphase nucleus. Chromosome Res 2015;**23**(3):625–39.

55. Zlotina A, Maslova A, Kosyakova N, et al. Heterochromatic regions in Japanese quail chromosomes: comprehensive molecular-cytogenetic characterization and 3D mapping in interphase nucleus. Chromosome Res 2019;**27**(3):253–70.

56. Fishman V, Battulin N, Nuriddinov M, et al. 3D Organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. Nucleic Acids Res 2019;**47**(2):648–65.

57. Schield DR, Card DC, Hales NR, et al. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome Res 2019;**29**(4):590–601.

58. Hooper DM, Price TD. Chromosomal inversion differences correlate with range overlap in passerine birds. Nat Ecol Evol 2017;**1**(10):1526–34.

59. Knief U, Hemmrich-Stanisak G, Wittig M, et al. Fitness consequences of polymorphic inversions in the zebra finch genome. Genome Biol 2016;**17**(1):199.

60. Craig RJ, Suh A, Wang M, et al. Natural selection beyond genes: identification and analyses of evolutionarily conserved elements in the genome of the collared flycatcher (*Ficedula albicollis*). Mol Ecol 2018;**27**(2):476–92.

61. Ma J, Zhang L, Suh BB, et al. Reconstructing contiguous regions of an ancestral genome. Genome Res 2006;**16**(12):1557–65.

62. Damas J, O'Connor R, Farré M, et al. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. Genome Res 2017;**27**(5):875–84.

63. Groenen MAM, Archibald AL, Uenishi H, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 2012;**491**(7424):393–8.

64. Kirkpatrick M. How and why chromosome inversions evolve. PLoS Biol 2010;**8**(9), doi:10.1371/journal.pbio.1000501.

65. Evseev D, Magor KE. Innate immune responses to avian influenza viruses in ducks and chickens. Vet Sci China 2019;**6**(1), doi:10.3390/vetsci6010005.

66. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012;**485**(7398):376–80.

67. Mirny LA, Imakaev M, Abdennur N. Two major mechanisms of chromosome organization. Curr Opin Cell Biol 2019;**58**:142–52.

68. Falk M, Feodorova Y, Naumova N, et al. Heterochromatin drives compartmentalization of inverted and conventional nuclei. Nature 2019;**570**(7761):395–9.

69. Busslinger GA, Stocsits RR, van der Lelij P, et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature 2017;**544**(7651):503–7.

70. Zhang Y, Li T, Preissl S, et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat Genet 2019;**51**(9):1380–8.

71. Ibrahim DM, Mundlos S. Three-dimensional chromatin in disease: What holds us together and what drives us apart? Curr Opin Cell Biol 2020;**64**, doi:10.1016/j.ceb.2020.01.003.

72. Harmston N, Ing-Simmons E, Tan G, et al. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. Nat Commun 2017;**8**(1):441.

73. Canela A, Maman Y, Jung S, et al. Genome organization drives chromosome fragility. Cell 2017;**170**(3):507–21.e18.

74. Rutkowska J, Lagisz M, Nakagawa S. The long and the short of avian W chromosomes: no evidence for gradual W shortening. Biol Lett 2012;**8**(4):636–8.

75. Hammar BO. The karyotypes of nine birds. Hereditas 2009;**55**(2–3):367–85.

76. http://pagelabsupplement.wi.mit.edu/papers/Bellott_et_al_2010/.

77. Solari AJ, Pigozzi MI. Recombination nodules and axial equalization in the ZW pairs of the Peking duck and the guinea fowl. Cytogenet Cell Genet 1993;**64**(3–4):268–72.

78. Xu L, Auer G, Peona V, et al. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. Nat Ecol Evol 2019;**3**(5):834–44.

79. Suh A, Paus M, Kiefmann M, et al. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. Nat Commun 2011;**2**:443.

80. Ryba T, Hiratani I, Lu J, et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res 2010;**20**(6):761–70.

81. Pope BD, Ryba T, Dileep V, et al. Topologically associating domains are stable units of replication-timing regulation. Nature 2014;**515**(7527):402–5.

82. Murphy WJ, Larkin DM, Everts-van der Wind A, et al. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 2005;**309**(5734):613–7.

83. Malcolm S, Abu-Amero S. Faculty Opinions Recommendation of [Hughes JF, Skaletsky H, Brown LG, et al. Nature 2012;483(7387):82–6]. Fac Opin 2012, doi:10.3410/f.14079956.15551056.

84. Hughes JF, Skaletsky H, Brown LG, et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature 2012;**483**(7387):82–6.

85. Rozen S, Skaletsky H, Marszalek JD, et al. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 2003;**423**(6942):873–6.

86. Mahajan S, Wei KHC, Nalley MJ, et al. De novo assembly of a young *Drosophila* Y chromosome using single-molecule se-

quencing and chromatin conformation capture. PLoS Biol 2018;**16**(7):e2006348.

87. Pendleton M, Sebra R, Pang AWC, et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat Methods 2015;**12**(8):780–6.

88. Bickhart DM, Rosen BD, Koren S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet 2017;**49**(4):643–50.

89. Chin C-S, Peluso P, Sedlazeck FJ, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 2016;**13**(12):1050–4.

90. Melissa LS, Delany N, Hepler N L, et al. An improved circular consensus algorithm with an application to detect HIV-1 drug resistance associated mutations (DRAMs). 2016.

91. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 2018;**19**(1):460.

92. Zemin N, Francesca G, Ed H. Scaff10X. https://github.com/wtsi-hpag/Scaff10X. 2019.

93. Yuan Y, Bayer PE, Lee H-T, et al. runBNG: a software package for BioNano genomic analysis on the command line. Bioinformatics 2017;**33**(19):3107–9.

94. Ghurye J, Rhie A, Walenz BP, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol 2019;**15**(8):e1007273.

95. English AC, Richards S, Han Y, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One 2012;**7**(11):e47768.

96. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;**9**(11):e112963.

97. Waterhouse RM, Seppey M, Simão FA, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 2018;**35**(3):543–8.

98. Aken BL, Achuthan P, Akanni W, et al. Ensembl 2017. Nucleic Acids Res 2017;**45**(D1):D635–D42.

99. Altschul SF, Gish W, Miller W, et al. Basic Local Alignment Search Tool. J Mol Biol 1990;**215**(3):403–10.

100. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res 2004;**14**(5):988–95.

101. Stanke M, Schöffmann O, Morgenstern B, et al. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 2006;**7**:62.

102. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;**29**(7):644–52.

103. Haas BJ, Salzberg SL, Zhu W, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol 2008;**9**(1):R7.

104. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics 2009, doi:10.1002/0471250953.bi0410s25.

105. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 2015;**6**(1), doi:10.1186/s13100-015-0041-9.

106. Benson G. Tandem Repeats Finder: a program to analyze DNA sequences. Nucleic Acids Res 1999;**27**(2):573–80.

107. Meyne J, Ratliff RL, Moyzis RK. Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci U S A 1989;**86**(18):7049–53.

108. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes. Genome Biol 2004;**5**(2):R12.

109. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;**25**(14):1754–60.

110. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;**27**(21):2987–93.

111. Alonge M, Soyk S, Ramakrishnan S, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 2019;**20**(1):224.

112. Servant N, Varoquaux N, Lajoie BR, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 2015;**16**:259.

113. Foissac S, Djebali S, Munyard K, et al. Transcriptome and chromatin structure annotation of liver, CD4 and CD8 T cells from four livestock species. bioRxiv 2019, doi:10.1101/316091.

114. Ramírez F, Bhardwaj V, Arrigoni L, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun 2018;**9**(1), doi:10.1038/s41467-017-02525-w.

115. Jolma A, Yan J, Whitington T, et al. DNA-binding specificities of human transcription factors. Cell 2013;**152**(1–2):327–39.

116. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Proc Int Conf Intell Syst Mol Biol 1994;**2**:28–36.

117. Servant N, Lajoie BR, Nora EP, et al. HiTC: exploration of high-throughput 'C' experiments. Bioinformatics 2012;**28**(21):2843–4.

118. Ardakany AR, Gezer HT, Lonardi S, et al. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. Genome Biol 2020;**21**(1):256.

119. Harris RS. Improved pairwise alignment of genomic DNA. Ph.D. Thesis. Pennsylvania State University; 2007.

120. UCSC Genome Browser Utilities. http://systemsbiology.cau.edu.cn/util.html., 2019.

121. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;**12**(4):357–60.

122. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;**15**(12):550.

123. Yanai I, Benjamin H, Shmoish M, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 2005;**21**(5):650–9.

124. Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform 2017;**18**(2):205–14.

125. Yang WR, Ardeljan D, Pacyna CN, et al. SQuIRE reveals locus-specific regulation of interspersed repeat expression. Nucleic Acids Res 2019;**47**(5):e27.

126. Li J, Zhang J, Liu J, et al. Supporting data for "A new duck genome reveals conserved and convergently evolved chromosome architectures of birds and mammals." GigaScience Database 2020; http://dx.doi.org/10.5524/100831.