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], Qi Zhou[1,3,6,†]

1. MOE Laboratory of Biosystems Homeostasis & Protection, Life Sciences Institute, Zhejiang University, Hangzhou 310058, China
2. Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm 17177, Sweden
3. Department of Neuroscience and Developmental Biology, University of Vienna, Vienna 1090, Austria
4. BGI-Shenzhen, 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, Yangling 712100, China
6. Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Hangzhou 310052, Zhejiang University
7. Laboratory of Neurogenetics of Language, The Rockefeller University, New York 10065, USA
8. Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.
9. Institute of Animal Science, Chinese Academy of Agricultural Sciences, 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, Kunming 650223, China

12. Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark

13. Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China

†Corresponding author. Email: zhouqi1982@zju.edu.cn

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 chicken. To elucidate the evolution of chromosome architectures between duck and chicken, 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:

The major improvement of genome assembly and annotation quality resulted from 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 the 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 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.

58 **Keywords:** Duck genome, chromosome inversion, topologically associated domain, sex

59 chromosomes

Background

Birds have the largest species number and some of the smallest genome sizes among terrestrial vertebrates. This has attracted extensive efforts since the era of cytogenetics into elucidating the diversity of their 'streamlined' genomes that give rise to the tremendous phenotypic diversity[1]. The karyotype of birds exhibits two major distinctions from that of mammals: first, it comprises about 10 pairs of large to medium sized chromosomes (macrochromosomes) and about 30 pairs of much smaller sized chromosomes (microchromosomes)[2]. During the over 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 over 800 bird species, the majority of them have a similar chromosome number around 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 to 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, due 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 about 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 it is a general feature of evolving sex chromosomes. Both cytogenetic work and Illumina-based genome assemblies of tens of bird species suggested that bird sex chromosomes comprise an unexpected interspecific diversity regarding both their lengths of recombining regions (pseudoautosomal regions, PAR), and their rates of gene loss[20, 27]. For example, PARs cover over 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 (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]. Further, Hi-C can uncover the three-dimensional (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 utilized a modified VGP pipeline to produce a nearly complete reference genome of a female Pekin duck (*Anas*

platyrhynchos, Z2 strain) 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* about 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 three 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 over 95% of the assembled sequences onto chromosomes, with great improvements in the non-coding regions and chrW sequences. We believe it will serve an important genomic resource for future studies into the mechanisms and application of artificial selection.

Data Description

Pekin duck (called duck from here on) has a haploid genome size estimated to be 1.41 Gb[37, 38], and a karyotype of 9 pairs of macrochromosomes (from chr1 to chr8, chrZ/chrW) and 31 pairs of microchromosomes (chr9 to chr39)[39]. The Illumina-based genome assembly of the duck (BGI1.0) was produced over seven 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-X genome coverage of PacBio long reads (read N50 14.3 kb from 115 SMRT cells, **Supplementary Fig. S1**), and 142-X genome coverage of 10x linked-read data from a female individual, 56-X genome coverage of BioNano map and 82-X genome coverage

of Hi-C reads from two different male individuals of the same inbred duck strain (**Figure 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-X genome coverage Illumina reads from a male individual of the same duck strain to compare to the previously published female reads (SRA accession number: 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 (5.5Mb vs. 26.1Kb) compared to the BGI1.0 genome (**Table 1**) . To scaffold the contigs, we first corrected their sequence errors with 92-X 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**). As Hi-C data provides 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 76.3Mb vs. 1.2Mb) compared to the Illumina genome, and is completely consistent with the FISH linkage map previously generated from 155 BAC clones (**Supplementary Fig. S2**)[33, 42]. The entire chrZ exhibits uniformly a 2-fold elevation of Illumina DNA sequencing read coverage in male relative to female, except for the chromosome

tip of pseudoautosomal regions (PAR) (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 about 200Mb 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 (**Figure 1b-c**), and we have assembled majorities of microchromosomes that were all unmapped in the BGI1.0 genome (**Figure 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 (**Figure 2a, Supplementary Table S4-5**), compared to 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 Fig. S3-4**). 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 (TE) 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 with other Galloanserae species (**Figure 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., **Figure 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 50kb, 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 (**Figure 3a-b**; 44.5 % vs. 39.3 % per 50kb, P< 2.2e-16, Wilcoxon test). In addition, all chromosomes but chrZ (**Figure 3a**) show generally equal expression levels between sexes; genes on chrZ are expressed twice the level in males versus females. These chromosome-wide patterns are consistent with those reported 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 (average length 443.3 kb) and telomeres (average length 73.7 kb), which were annotated by a cytogenetically verified *Anseriformes* centromeric repeat (APL-*Hae*III)[50] and conserved telomeric motif sequences (**Supplementary Table S4-5**). We found 22 telomeric sites among the 31 chromosomes, of which 11 were interstitial telomeric repeat (ITR) sites inside the

chromosomes (**Figure 3a-b,** green arrow heads). Consistent with the reported karyotypes of duck and other birds[50, 51], almost all microchromosomes are acrocentric indicated by their positions of 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 repeat GGERVL-A-int (**Figure 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 inter-chromosomal interactions measured by the Hi-C data compared to 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 reported in a companion paper[57]) as the outgroup, and identified 80 inversions on 26 chromosomes (>10kb, median size 1.5Mb, **Supplementary Table S7**) that occurred in the duck or *Anseriformes* lineage after it diverged from chicken in the past 72.5 MY[34] (**Figure 3c-d**). The average inversion rate (1.1 inversion events or 3.1Mb inverted regions per MY) of Pekin duck is lower than that of 1.5-2.0 events or 6.6-7.5Mb 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 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 interests of 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 (**Figure 3c-d**) are enriched for genes that function in immunity-related pathways (**Figure 4a**, e.g., 'defense response to virus', 'G-protein coupled receptor pathway'; P<0.0001, Fisher's Exact test), which may account for the known divergent susceptibility between chicken and duck against 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 if 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 with a significant enrichment of putative 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 (**Figure 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-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] have been recently 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 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 be recently inserted into the TAD boundaries. In addition, all the assembled centromere regions of metacentric chromosomes, and intriguingly 4 out of 11 ITRs (**Figure 2a,b**) coincide with the TAD boundaries (**Supplementary Figs. S7, 17**). This highlighted the uncharacterized role of ITRs in demarcating the functional domains in the chromosomes yet to be functionally tested in 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 (**Figure 4c**). This is probably an underestimate given that different tissues of Hi-C data were used here to identify TADs for the two bird

294 species. A comparable level of conservation of human TAD boundaries (53.8%) has also been
295 observed with mouse[66], and expectedly a lower level (26.8%) of conservation has been
296 observed between human and chicken[56]. The other evidence of strong selective constraints
297 acting on the integrity of TADs come from our findings here on the pattern of chromosomal
298 inversion breakpoints of duck, whose TAD insulation scores are significantly (P< 2.2e-16,
299 Wilcoxon test) lower (**Figure 4d**) than the TAD interior regions. That is, inversions more often
300 precisely occurred at the TAD boundaries rather than within the TADs, i.e., disrupting the pre-
301 existing TADs. Only one third of the detected inversions have both their breakpoints located
302 within the TADs, whereas the remaining two thirds have both or one of their breakpoints
303 overlapping with the TAD boundaries (**Figure 4e-g**). Novel TAD boundaries that were created
304 by the duck-specific inversions (e.g., **Figure 4g**) tend to have significantly higher insulation
305 scores, i.e., weaker insulation strengths than those that are conserved between duck and chicken
306 (**Supplementary Fig. S18**). This suggests that natural selection may more frequently target
307 evolutionarily older and stronger TAD boundaries. We have to point out the alternative
308 explanation for the overlap between the TAD boundaries and inversion breakpoints (**Figure 4e**)
309 is that chromatin loop anchors bound by CTCF protein are more likely genomic fragile sites
310 vulnerable for DNA double-strand breaks[73] that induce the inversions. Consistent with this
311 explanation, we found that the TAD boundaries that overlap with inversion breakpoints (**Figure**
312 **4h, bottom**) have a significantly (P<0.001, Chi-square test) higher percentage of loop anchors
313 than others (**Figure 4h, top**).
314 Since the novel TADs generated by chromosome inversions (e.g., **Figure 4g**) may create
315 aberrant or new promoter-enhancer contacts, and consequently divergent gene expression during
316 evolution, we further compared the levels of gene expression divergence in the conserved TADs
317 vs. those novel TADs that encompass inversion breakpoints between chicken and duck.
318 Interestingly, genes that are close to the novel TAD boundaries created by inversions only show
319 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, that 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 differentiation degree of its sex chromosomes is between those of ratites and chicken[27]. Previous comparative cytogenetic work found that the FISH probe of chicken chrZ cannot produce hybridization signals on chicken chrW because of their great sequence divergence, but instead can paint the entire chrW of duck and ostrich, suggesting that substantial sequence homology has been preserved between the Z/W chromosomes of the two species since the recombination was suppressed[27, 66]. The size of duck chrW is nevertheless smaller (estimated size 51Mb)[74, 75] compared to 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 one continuous sequence 84.5Mb long (**Supplementary Fig. S20**). The size of duck chrZ is similar to that of published chicken chrZ (82.5 Mb[76]).

We determined 2.2Mb long PAR at the tip of chrZ (**Figure 5a**), based on its equal read coverage between sexes. This is consistent with previous cytogenetic work showing only one 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 rest Z-linked SDR that do not have recombination in females (**Figure 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 four gene families that are specifically expressed in

346 testis[18]. In contrast, we did not find similar tandem arrays of testis genes on chrZ of duck, and
347 all of the four Z-linked chicken testis gene families are located on the autosomes of duck
348 (**Supplementary Fig. S22**).
349 The assembled duck chrW assembly contains 36 scaffolds with a total length of 16.7Mb (about
350 one third of the estimated size), all of which are almost exclusively mapped by female reads
351 (**Supplementary Fig. S20**). It marks an 8.8-fold increase in size compared to our previous
352 assembly using Illumina reads[20, 78], and is much longer than the most recent assembly of
353 chicken chrW (6.7 Mb)[22]. We have annotated a total of 71 duck W-linked SDR genes, and all
354 of them are single copy genes, compared to 27 single-copy genes and one multicopy gene on the
355 chicken chrW, with 20 genes overlapped between the two (**Figure 5b**). The only multicopy
356 chicken W-linked gene *HINTW* with about 40 copies[22] is present as a single-copy gene on the
357 duck chrW. These results indicate that duck and chicken have independently evolved their sex-
358 linked gene repertoire since their species divergence. The duck chrW retained more genes than
359 chicken, and represents an intermediate stage of avian sex chromosome evolution between those
360 of ratites and chicken.
361 Due to the intrachromosomal rearrangements of chrZ, most birds (including duck) except for
362 ratites have retained few ancestral gene syntenies of their proto-sex chromosomes before the
363 suppression of homologous recombination[20, 78], and exhibit dramatic reshuffling of their old
364 evolutionary strata. In order to accurately reconstruct the history of duck sex chromosome
365 evolution, we used a newly produced chrZ assembly of emu in our group to approximate the
366 avian proto-sex chromosomes. Almost all (15.2Mb, 91%) of the duck chrW sequences can be
367 aligned to the chrZ of emu, and form a clear pattern of four evolutionary strata. This is
368 manifested as a gradient of Z/W pairwise sequence divergence, i.e., a gradient of the age of strata
369 along the chrZ, which is named from the old to the young, as stratum 0, S0 to S3, (**Figure 5a**).
370 Within each stratum, chrW scaffolds of similar levels of sequence divergence are clustered and
371 separated from the neighbouring 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 neighbouring 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 out of 9 boundaries shown in **Figure 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, majorities (30 or 42.9%) of W-linked genes probably have become pseudogenes or long non-coding RNA genes due 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 to the genome average (**Figure 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 (**Figure 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, **Figure 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 (**Figure 5e**). Their lengths range from 15.2 kb to 345.5 kb (**Supplementary Table S9**), together comprising 3.74Mb or 22% of the assembled duck chrW sequence.

Discussion

Birds and mammals diverged over 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% to 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 duck Illumina-based 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 the 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 two 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 slower 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 which 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 many more genes than the chrW of chicken have been preserved.

Methods

Genome assembly

High molecular weight DNA (HMW DNA) was extracted from the liver of a female Pekin duck (*Anas platyrhynchos*, Z2 strain) with Gentra Puregene Tissue Kit (Qiagen #158667). Libraries for SMRT sequencing were constructed as described previously[87]. In total, 115 SMRT cells were sequenced with PacBio RS II and Sequel platform (Pacific Biosciences), and 186 Gb (143-X 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 MGISEQ-2000 platform for sequencing and 185 Gb PE150 (142-X genome coverage) reads were collected. HMW DNA of a male Pekin duck was used to produce the BioNano library with the Enzyme Nt.BspQ1. 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-X 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 Mbol with the protocol described previously[30] and produced a total of 106Gb (82-X genome coverage) pair-end reads of 50bp long on the Illumina HiSeq X Ten platform. We used the published genome resequencing data of 14 female and 11 male duck individuals from[46]. We collected the total RNAs of adult tissues (brain, kidney, gonads) of both sexes using TRIzol® Reagent (Invitrogen #15596-018) following the manufacturers' instructions. Then paired-end libraries were constructed using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) and 3Gb paired-end reads of 150bp were produced for each library.

We generated the genome assembly with the modified Vertebrate Genomes Project (VGP) (v1.0) pipeline[29]. In brief, we produced the contig sequences derived from the PacBio subreads using

476 FALCON[89] (git 12072017) followed by two rounds of assembly polishing by Arrow[90], and
477 then by Purge Haplotigs[91] (bitbucket 7.10.2018) to remove false haplotype and homotypic
478 duplications. The contigs were then scaffolded first with 10x linked reads using Scaff10X
479 (https://github.com/wtsihpag/Scaff10X), then with BioNano optical maps using runBNG[92]
480 (v1.0.3), and finally with Hi-C reads using SALSA[93] (v2.0). We performed gap filling on the
481 scaffolds with the Arrow-corrected PacBio subreads by PBJelly[94], and two rounds of assembly
482 polishing with Illumina reads by Pilon[95] (v1.22). All the scripts used from the VGP assembly
483 pipeline[29] are available at https://github.com/VGP/vgp-assembly. We evaluated the genome
484 completeness using BUSCO[96] (v3.0.2). In brief, 4,915 benchmarking universal single-copy
485 ortholog (BUSCO) proteins of birds from OrthoDB v9 were used in the evaluation.
486
487 **Genome annotation**
488 We combined evidence of protein homology, transcriptome and *de novo* prediction to annotate the
489 protein-coding genes. First, we aligned the protein sequences of human, chicken, duck and zebra
490 finch collected from Ensembl[97] (release 90) to the reference genome using TBLASTN[98]
491 (v2.2.26) with parameters: -F F -p tblastn -e 1e-5. The resulting candidate genes were then refined
492 by GeneWise[99] (v2.4.1). For each candidate gene, only the one with the best score was kept as
493 the representative model. We filtered the candidate genes, if they contain premature stop codons or
494 frameshift mutations reported by GeneWise[99]; or if single-exon genes with a length shorter than
495 100bp, or multi-exon genes with a length shorter than 150bp; or if the repeat content of the CDS
496 sequence is larger than 20%. Second, to obtain the *de novo* gene models, we used the protein
497 queries to train Augustus[100] (v3.3) with default parameters. We also used all available RNA-seq
498 reads to construct transcripts using Trinity[101] (v2.4.0). Finally, all the gene models from the
499 above three resources were merged into a non-redundant gene set with EVidenceModeler[102]
500 (v1.1.1). We used RepeatMasker[103] (v4.0.8) with parameters: -s -pa 4 -xsmall, and the
501 RepBase[104] (v21.01) queries to annotate the repetitive elements.

To annotate the putative centromeres, we searched the genome with the reported 190bp duck centromeric repeats[50] using TRFinder[105] (v4.09) with the parameters: 2 5 7 80 10 50 2000. A genome-wide distribution of the 190bp sequences was generated by binning the genome with a 50kb 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[106] '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 50kb 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 1689 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 we generated by nucmer[107] (v3.23) packages (http://mummer.sourceforge.net) and only kept the best hits for each sequence. Scaffolds were orientated and ordered first based on the RHmap contigs that span more than one 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 had conflicts 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[108] with default parameters. Read depth of each sex was then calculated using SAMtools[109] in 5kb 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 sequences, 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 from 0.0 to 0.25 as W-linked sequences (**Supplementary Fig. S21**). Since 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[110] (v1.1) with default parameters (https://github.com/malonge/RaGOO). 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 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 two 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[111] (v2.10.0) with the default parameters. In brief, Hi-C reads of chicken[112] (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 reads were assigned to a restriction fragment and invalid ligation products were discarded. Data was then merged and binned to generate the genome-wide interaction maps at 10kb and 50kb resolution. TADs were identified by HiCExplorer[113] (v3.0) with the application hicFindTADs. First, HiC-Pro interaction maps were transformed to h5 format matrix by hicConvertFormat with parameters: --inputFormat hicpro --outputFormat h5. Then the h5 matrix was imported to hicFindTADs with parameters:--outPrefix TAD --numberOfProcessors 32 --correctForMultipleTesting fdr. hicFindTADs identifies the TAD boundaries through an approach that computes a TAD insulation score. Genomic bins with low

554 insulation scores relative to neighboring regions were defined as local minima and called as the
555 TAD boundaries. Human CTCF[114] motif was used as a query for FIMO in MEME[115]
556 (v4.12.0) to identify the putative CTCF binding sites. CTCF density in every 10kb non-overlapping
557 sliding window along the genome was calculated to check its enrichment at the TAD boundaries.
558 We identified the A/B compartments using the pca.hic function from HiTC[116] (High Throughput
559 Chromosome Conformation Capture analysis) R package with default parameters, and the 10kb
560 matrix generated by HiC-Pro as the input. We identified the chromatin loops by Mustache[117]
561 with the parameters: -p 32 -r 10kb -pt 0.05, after converting the h5 format matrix to mcool matrix
562 format by hicConvertFormat with parameters: --inputFormat h5 --outputFormat mcool.

563 **Evolutionary strata**

564 To demarcate the evolutionary strata, all the repeat masked duck W-linked scaffolds were aligned to
565 emu Z chromosome using LASTZ[118] (v0.9) with parameters: --step=19 --hspthresh=2200 --
566 inner=2000 --ydrop=3400 --gappedthresh=10000 --format=axt, and a score matrix set for the distant
567 species comparison. Alignments were converted into 'net' and 'maf' results using UCSC Genome
568 Browser's utilities (http://genomewiki.ucsc.edu/index.php/). Based on 'net' and 'maf' results, the
569 identity of the aligned sequence was calculated for each alignment block with a 10kb non-
570 overlapped window and then we oriented the aligned W-linked sequences along the Z
571 chromosomes. Then we color-coded the pairwise sequence divergence level between the Z/W
572 sequences to demarcate the evolutionary strata.

573 **Gene expression analyses**

574 RNA-seq reads were mapped to the duck genome by HISTA2[119] with default parameters. Only
575 uniquely mapped RNA-seq reads were kept and used to calculate the RPKM expression level.
576 DESeq2[120] was applied to normalize the RPKM values across different samples and finally
577 generated an expression matrix. For each gene, we used the median expression value in each tissue

to calculate the tissue specificity index TAU[121, 122]. Expression levels of TE elements were calculated using SQUIRE[123] (v0.9.9.92) (https://github.com/wyang17/SQuIRE) with default parameters.

Data availability

The assembly and annotation of Pekin duck has been deposited in GenBank under the Bioproject accession code PRJNA636121 (accession number JACGAL000000000) and the emu under PRJNA638233 (accession number JABVCD000000000).

Code availability

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

Acknowledgment

Q.Z. is supported by the National Natural Science Foundation of China (31722050, 31671319), the Natural Science Foundation of Zhejiang Province (LD19C190001) and the European Research Council Starting Grant (grant agreement 677696). We thank BGI-Shenzhen for providing the 10x linked reads data of duck.

Conflict of interest statement

None declared.

Authors' contributions

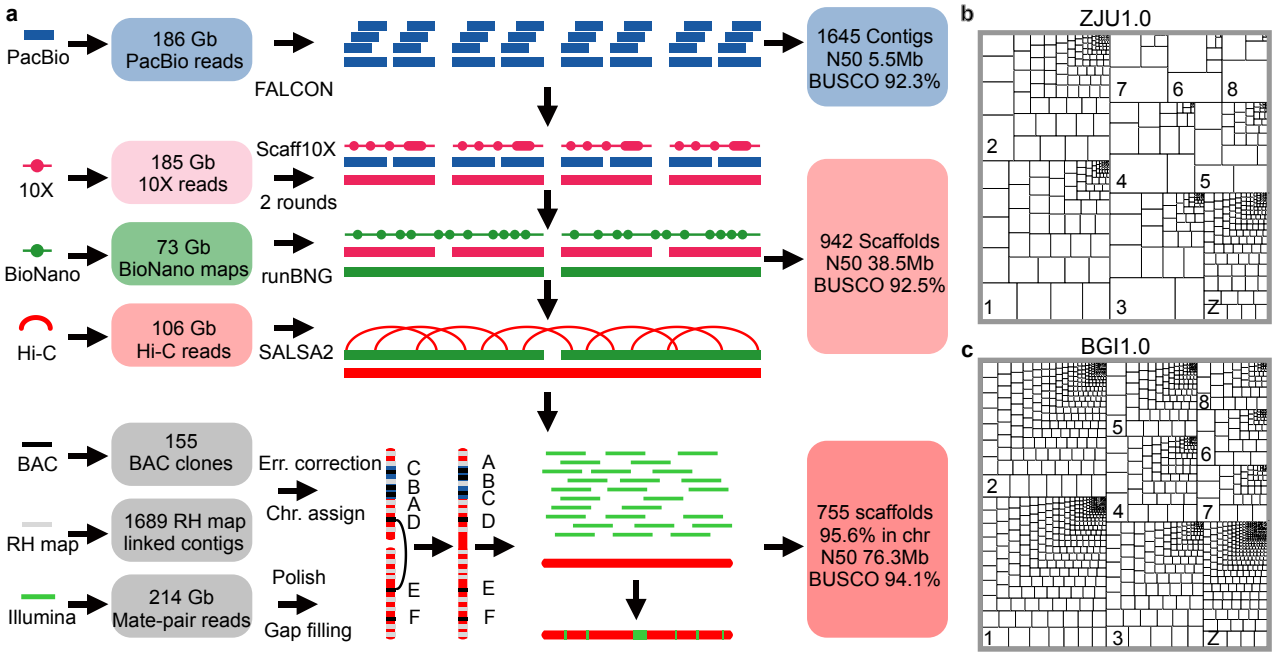
Q. Z. conceived the project and acquired the funding; J. L., X. D., S. F., C. G., J. R., K. W., acquired the samples and produced the data; J. L., J. Z., J. L., Y. Z., C. C., L. X., Q. Z. performed the analyses.; J. L., Y. J. , Z. Z., G. Z., E. J. and Q. Z. wrote the paper.

604 **Table 1. Comparing genome assemblies of duck vs. other birds**

	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
#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
#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

605

606 **Figures**



607

608 **Figure 1. Genome assembly of a female Pekin duck. a**. Our assembly pipeline uses high
609 coverage PacBio long reads to generate contigs, which are then sequentially scaffolded with 10X
610 Genomics linked reads, BioNano optical maps, Hi-C paired reads, RH maps and FISH maps, to
611 produce a chromosome-level genome for the Pekin duck. **b, c**. Treemap comparison of contigs
612 between ZJU1.0 and BGI1.0 versions of the duck genome. The size of each rectangle of each
613 chromosome is scaled to that of contig sequence. The bigger and fewer the internal boxes, the
614 more contiguous the contigs.

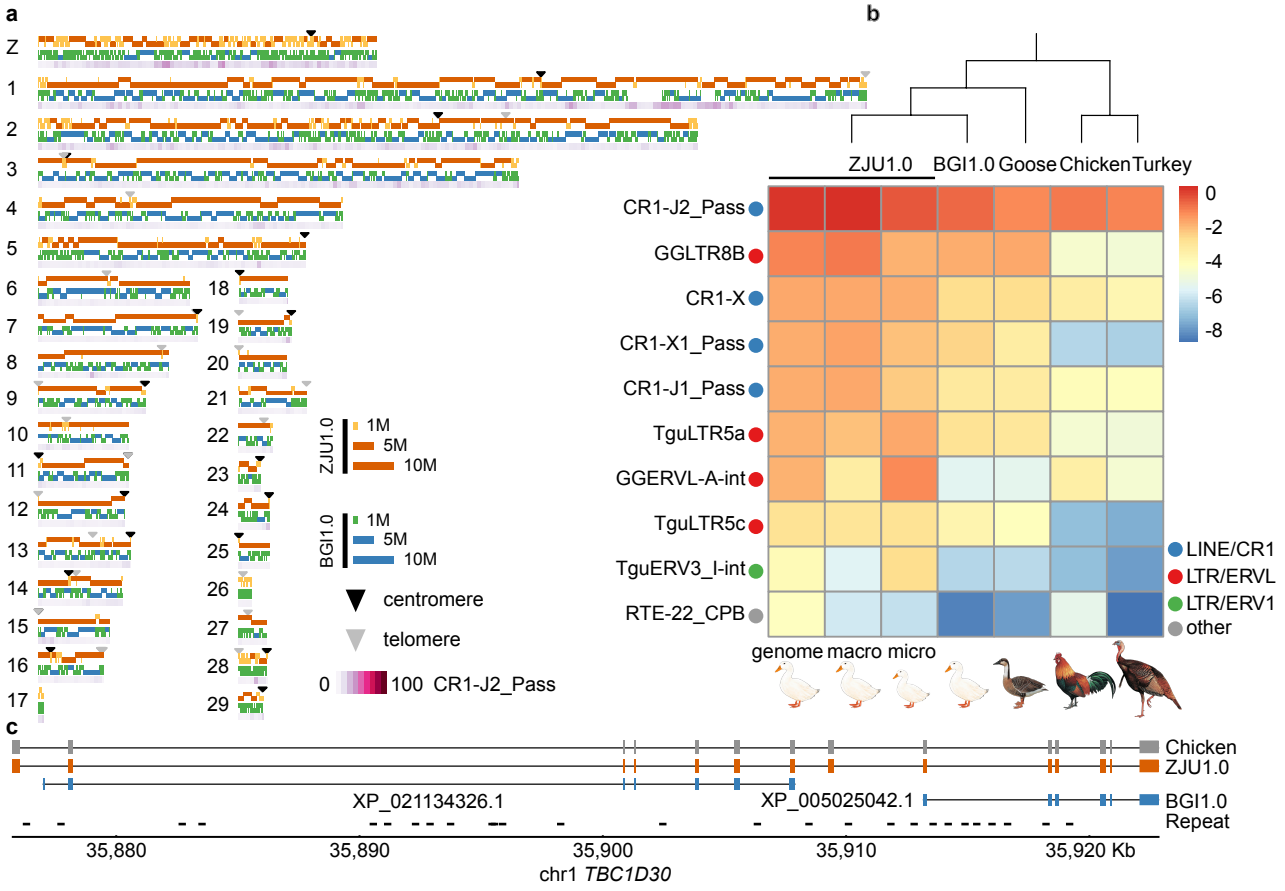
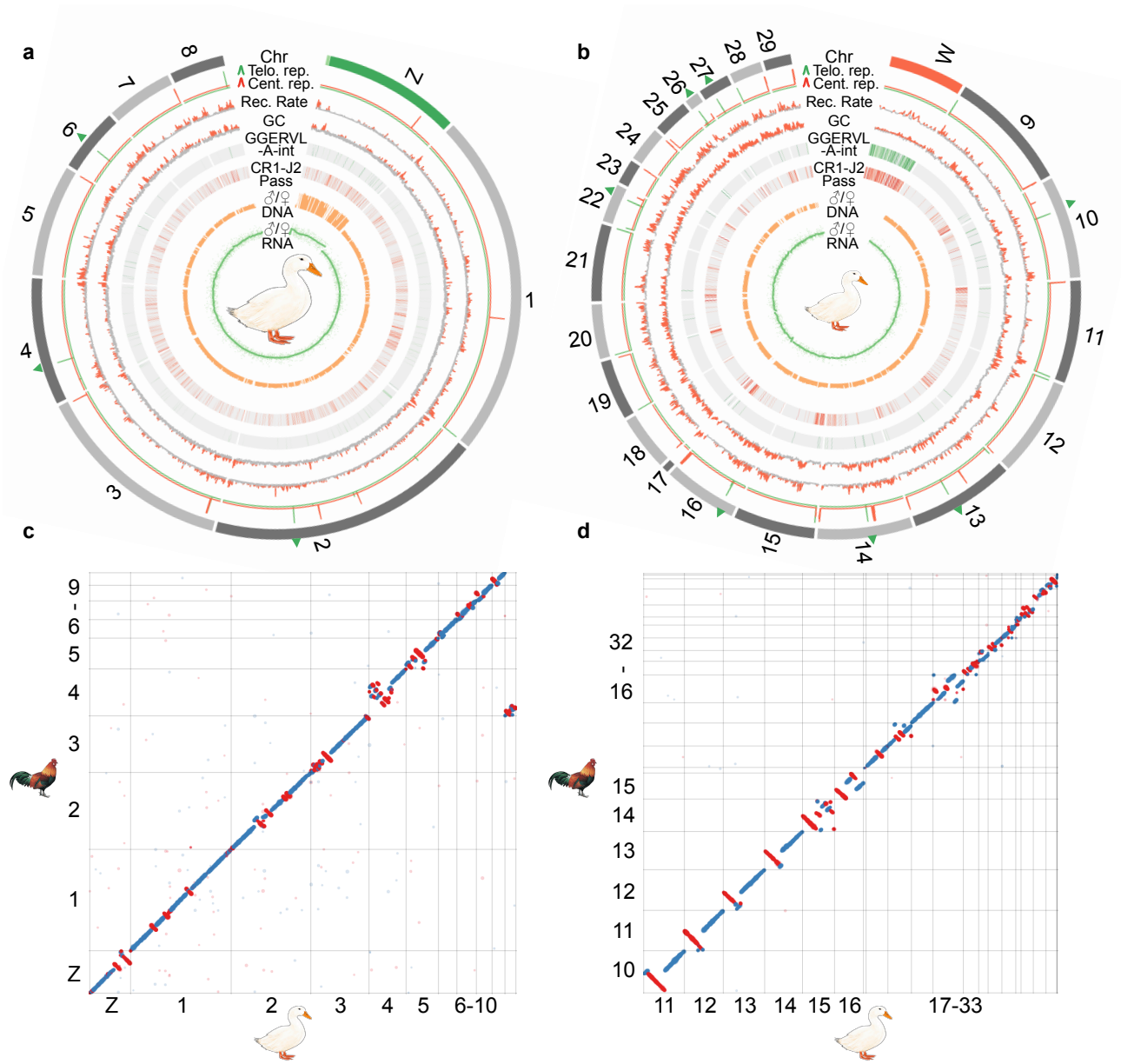


Figure 2. Comparing the new duck genome to other avian genomes a. Schematic plot of each chromosome, showing the mapped contigs of ZJU1.0 (orange/yellow) and BGI1.0 (blue/green), putative centromeres (black triangles), and telomeres or interstitial telomeric sequences (grey triangles), and the most abundant 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 proportion of assembled repeat content. **c.** An example gene annotation improvement showing two genes in the BGI1.0 genome are really one gene in the ZJU1.0 genome, and were fragmented into two because of low resolution of repeat sequences disrupting the previous genome assembly of exons.



627

628 **Figure 3. Evolution of the duck macro- and microchromosomes**. From the outer to inner
629 rings: the macro- (**a**) and microchromosomes (**b**), together with Z/W chromosomes (green/red
630 color), and the pseudoautosomal regions (PARs) labelled with light green color at the tip of chrZ.
631 Interstitial telomere sequences were labelled with green triangles on the chromosome. Putative
632 centromeres (red lines) and telomeres (green lines) were inferred by the enrichment of
633 centromeric and telomeric repeat copies, which show a sharp peak. We then show the
634 recombination rate and GC content calculated in non-overlapping 50kb windows, as well as two
635 repeat families (GGERVL-A-int and CR1-J2 Pass) that we identified to be enriched at
636 centromeric regions and chrW. We also show the male vs. female (M/F) ratios of Illumina DNA

637 sequencing coverage in non-overlapping 50kb windows, M/F expression ratios (each green dot
638 as one gene) of the adult brain tissue and the smoothed line. **c-d.** Dot plots show the inversions
639 between chicken and duck genome for both macro and micro chromosomes.

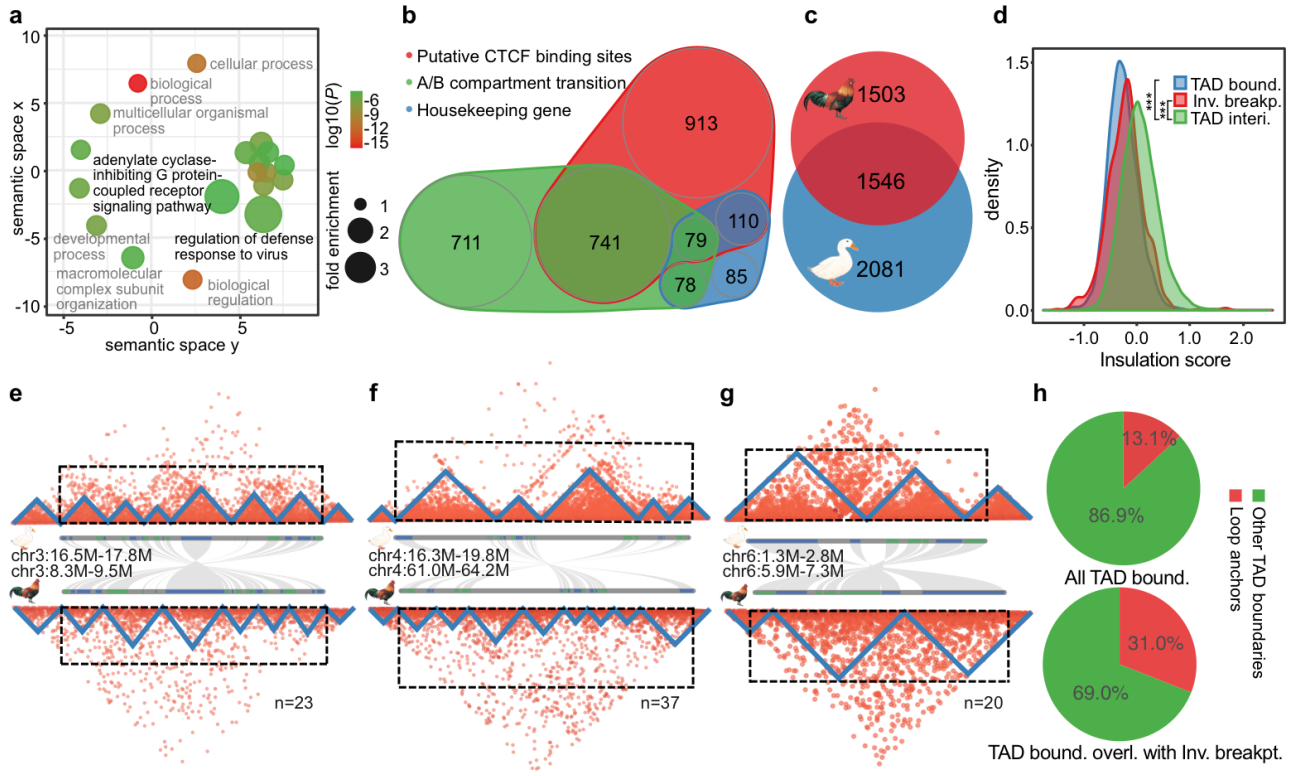


Figure 4. Genome inversions and topologically associated domains. a. Enriched GO terms of the genes included in the duck specific inversions. The x- and y-axes measure the GO term semantic similarities, which are used to remove the GO redundancies. **b**. Scaled Venn diagram shows the different compositions 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**. We show the Hi-C heatmaps with each triangle structure indicating one 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), one 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).

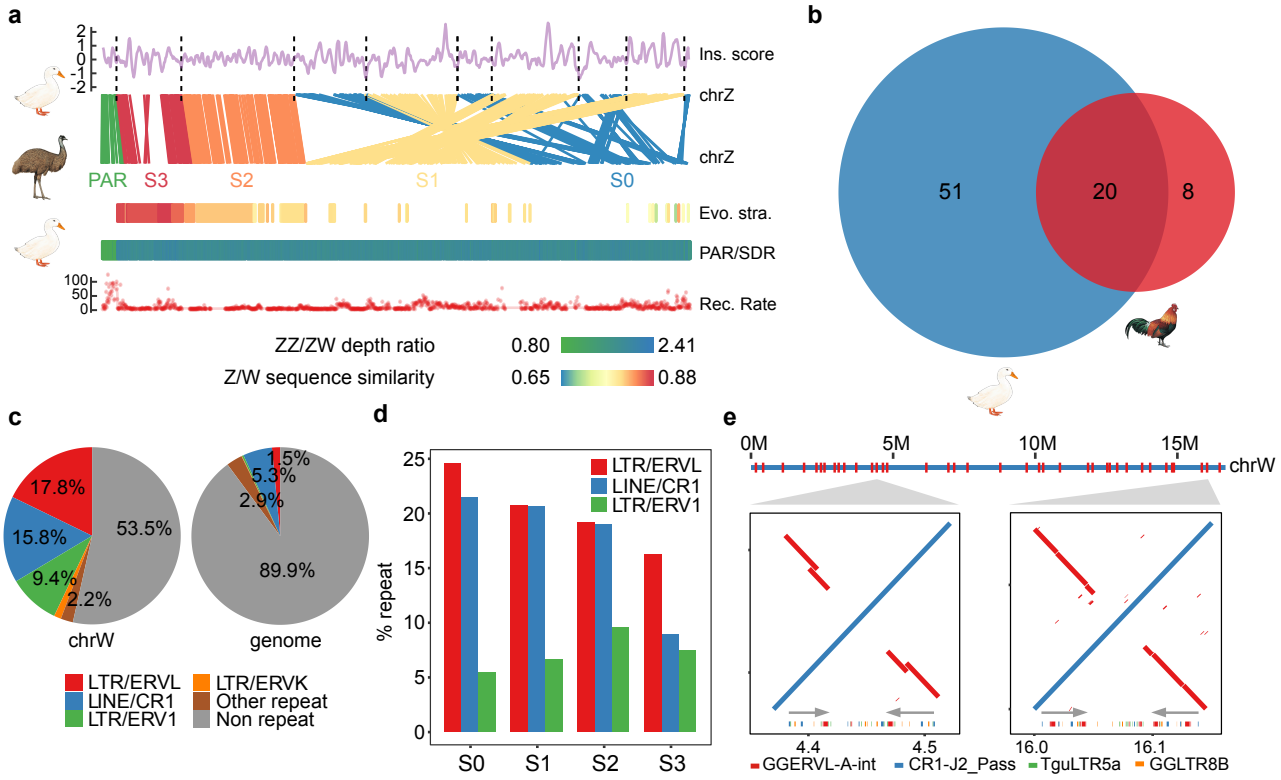


Figure 5. Sex chromosome evolution in Pekin duck. a, Evolutionary strata analyses of the duck sex chromosomes. From top to bottom: the breakpoints of genomic rearrangements between emu and duck chrZ tend to have a lower insulation score; gene synteny between the emu and duck Z chromosomes; alignment 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 showed 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 labelled across the entire chrW (red), ordered according to the duck chrZ. Shown are alignment plots of two zoomed-in examples of palindromes (red inversions and grey arrows) for their repeat content (colors below grey arrows).

669 **Reference**

670 1. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights
671 into avian genome evolution and adaptation. Science. 2014;346 6215:1311-20.
672 doi:10.1126/science.1251385.
673 2. Burt DW. Origin and evolution of avian microchromosomes. Cytogenet Genome Res.
674 2002;96 1-4:97-112. doi:10.1159/000063018.
675 3. Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, et al. The dynamics of
676 chromosome evolution in birds and mammals. Nature. 1999;402 6760:411-3.
677 doi:10.1038/46555.
678 4. Griffin DK, Robertson LBW, Tempest HG and Skinner BM. The evolution of the avian
679 genome as revealed by comparative molecular cytogenetics. Cytogenet Genome Res.
680 2007;117 1-4:64-77. doi:10.1159/000103166.
681 5. Damas J, Kim J, Farré M, Griffin DK and Larkin DM. Reconstruction of avian ancestral
682 karyotypes reveals differences in the evolutionary history of macro- and
683 microchromosomes. Genome Biol. 2018;19 1:155. doi:10.1186/s13059-018-1544-8.
684 6. Nanda I, Karl E, Griffin DK, Schartl M and Schmid M. Chromosome repatterning in three
685 representative parrots (Psittaciformes) inferred from comparative chromosome painting.
686 Cytogenet Genome Res. 2007;117 1-4:43-53. doi:10.1159/000103164.
687 7. O'Connor RE, Farré M, Joseph S, Damas J, Kiazim L, Jennings R, et al. Chromosome-level
688 assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich,
689 genomes. Genome Biol. 2018;19 1:171. doi:10.1186/s13059-018-1550-x.
690 8. Nishida C, Ishijima J, Kosaka A, Tanabe H, Habermann FA, Griffin DK, et al.
691 Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves)
692 by chromosome painting and delineation of chromosome rearrangements during their
693 differentiation. Chromosome Research. 2008;16 1:171-81. doi:10.1007/s10577-007-1210-6.
694 9. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses
695 resolve early branches in the tree of life of modern birds. Science. 2014;346 6215:1320-31.
696 10. Volume 4: Chordata 3: B. Aves. In: Les C, editor. Animal Cytogenetics. Berlin, Germany:
697 Gebrüder Borntraeger; 1990. p. 55-7.
698 11. Skinner BM and Griffin DK. Intrachromosomal rearrangements in avian genome evolution:
699 evidence for regions prone to breakpoints. Heredity. 2012;108 1:37-41.
700 doi:10.1038/hdy.2011.99.
701 12. Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, et al. A high-
702 density linkage map enables a second-generation collared flycatcher genome assembly and
703 reveals the patterns of avian recombination rate variation and chromosomal evolution.
704 Molecular Ecology. 2014;23 16:4035-58. doi:10.1111/mec.12810.
705 13. Pevzner P and Tesler G. Human and mouse genomic sequences reveal extensive breakpoint
706 reuse in mammalian evolution. Proc Natl Acad Sci U S A. 2003;100 13:7672-7.
707 doi:10.1073/pnas.1330369100.
708 14. Larkin DM, Pape G, Donthu R, Auvil L, Welge M and Lewin HA. Breakpoint regions and
709 homologous synteny blocks in chromosomes have different evolutionary histories. Genome
710 Res. 2009;19 5:770-7. doi:10.1101/gr.086546.108.
711 15. Völker M, Backström N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H, et al. Copy
712 number variation, chromosome rearrangement, and their association with recombination
713 during avian evolution. Genome Res. 2010;20 4:503-11. doi:10.1101/gr.103663.109.
714 16. Lemaitre C, Zaghloul L, Sagot M-F, Gautier C, Arneodo A, Tannier E, et al. Analysis of
715 fine-scale mammalian evolutionary breakpoints provides new insight into their relation to
716 genome organisation. BMC Genomics. 2009;10:335. doi:10.1186/1471-2164-10-335.
717 17. Irwin DE. Sex chromosomes and speciation in birds and other ZW systems. Mol Ecol.
718 2018;27 19:3831-51. doi:10.1111/mec.14537.

18. Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, et al. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature. 2010;466 7306:612-6. doi:10.1038/nature09172.
19. Lahn BT and Page DC. Four evolutionary strata on the human X chromosome. Science. 1999;286 5441:964-7. doi:10.1126/science.286.5441.964.
20. Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science. 2014;346 6215:1246338. doi:10.1126/science.1246338.
21. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional evolution of Y chromosomes across mammals. Nature. 2014;508 7497:488-93. doi:10.1038/nature13151.
22. Bellott DW, Skaletsky H, Cho T-J, Brown L, Locke D, Chen N, et al. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. Nat Genet. 2017;49 3:387-94. doi:10.1038/ng.3778.
23. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 2003;423 6942:825-37. doi:10.1038/nature01722.
24. Charlesworth B and Charlesworth D. The degeneration of Y chromosomes. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2000;355 1403:1563-72. doi:10.1098/rstb.2000.0717.
25. Davis JK, Program NCS, Thomas PJ and Thomas JW. A W-linked palindrome and gene conversion in New World sparrows and blackbirds. Chromosome Research. 2010;18 5:543-53. doi:10.1007/s10577-010-9134-y.
26. Zhou R, Macaya-Sanz D, Carlson CH, Schmutz J, Jenkins JW, Kudrna D, et al. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. Genome Biol. 2020;21 1:38. doi:10.1186/s13059-020-1952-4.
27. Nanda I, Schlegelmilch K, Haaf T, Schartl M and Schmid M. Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination. Cytogenetic and Genome Research. 2008;122 2:150-6. doi:10.1159/000163092.
28. Xu L, Wa Sin SY, Grayson P, Edwards SV and Sackton TB. Evolutionary Dynamics of Sex Chromosomes of Paleognathous Birds. Genome Biol Evol. 2019;11 8:2376-90. doi:10.1093/gbe/evz154.
29. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, 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, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326 5950:289-93. doi:10.1126/science.1181369.
31. Szabo Q, Bantignies F and Cavalli G. Principles of genome folding into topologically associating domains. Science Advances. 2019;5 4:eaaw1668. doi:10.1126/sciadv.aaw1668.
32. Rao M, Morisson M, Faraut T, Bardes S, Fève K, Labarthe E, et al. A duck RH panel and its potential for assisting NGS genome assembly. BMC Genomics. 2012;13 1:513. doi:10.1186/1471-2164-13-513.
33. Skinner BM, Robertson LBW, Tempest HG, Langley EJ, Ioannou D, Fowler KE, et al. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. BMC Genomics. 2009;10:357. doi:10.1186/1471-2164-10-357.
34. Claramunt S and Cracraft J. A new time tree reveals Earth history's imprint on the evolution of modern birds. Science Advances. 2015;1 11:e1501005. doi:10.1126/sciadv.1501005.

35. Herrera AM, Brennan PLR and 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. doi:10.1159/000364927.
36. Huang Y, Li Y, Burt DW, Chen H, Zhang Y, Qian W, et al. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat Genet. 2013;45 7:776-83. doi:10.1038/ng.2657.
37. Nakamura D, Tiersch TR, Douglass M and Chandler RW. Rapid identification of sex in birds by flow cytometry. Cytogenet Cell Genet. 1990;53 4:201-5. doi:10.1159/000132930.
38. Tiersch TR and Wachtel SS. On the evolution of genome size of birds. J Hered. 1991;82 5:363-8. doi:10.1093/oxfordjournals.jhered.a111105.
39. Takagi N and Makino S. A Revised Study on the Chromosomes of three Species of Birds. Caryologia. 1966;19 4:443-55. doi:10.1080/00087114.1966.10796235.
40. Lu L, Chen Y, Wang Z, Li X, Chen W, Tao 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. doi:10.1186/s13059-015-0652-y.
41. Warren WC, Hillier LW, Tomlinson C, Minx P, Kremitzki M, Graves T, et al. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. G3. 2017;7 1:109-17. doi:10.1534/g3.116.035923.
42. Islam FB, Uno Y, Nunome M, Nishimura O, Tarui H, Agata K, 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. The Journal of Poultry Science. 2013:0130090.
43. Peona V, Blom MPK, Xu L, Burri R, Sullivan S, Bunikis I, et al. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. doi:10.1101/2019.12.19.882399.
44. Botero-Castro F, Figuet E, Tilak M-K, Nabholz B and Galtier N. Avian Genomes Revisited: Hidden Genes Uncovered and the Rates versus Traits Paradox in Birds. Mol Biol Evol. 2017;34 12:3123-31. doi:10.1093/molbev/msx236.
45. Korlach J, Gedman G, Kingan SB, Chin C-S, Howard JT, Audet J-N, 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, Fan W, Yuan Z, Gao Q, et al. An intercross population study reveals genes associated with body size and plumage color in ducks. Nat Commun. 2018;9 1:2648. doi:10.1038/s41467-018-04868-4.
47. Duret L and Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet. 2009;10:285-311. doi:10.1146/annurev-genom-082908-150001.
48. International Chicken Genome Sequencing C. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004;432 7018:695-716. doi:10.1038/nature03154.
49. McQueen HA, McBride D, Miele G, Bird AP and Clinton M. Dosage compensation in birds. Current Biology. 2001;11 4:253-7. doi:10.1016/s0960-9822(01)00070-7.
50. Uno Y, Nishida C, Hata A, Ishishita S and Matsuda Y. Molecular cytogenetic characterization of repetitive sequences comprising centromeric heterochromatin in three Anseriformes species. PLoS One. 2019;14 3:e0214028. doi:10.1371/journal.pone.0214028.
51. Wójcik E and Smalec E. Description of the mallard duck (Anas platyrhynchos) karyotype. Folia Biol. 2007;55 3-4:115-20. doi:10.3409/173491607781492588.
52. Matzke MA, Varga F, Berger H, Schernthaner J, Schweizer D, Mayr B, 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. doi:10.1007/bf01735329.

53. Tanaka K, Suzuki T, Nojiri T, Yamagata T, Namikawa T and Matsuda Y. Characterization and chromosomal distribution of a novel satellite DNA sequence of Japanese quail (Coturnix coturnix japonica). J Hered. 2000;91 5:412-5. doi:10.1093/jhered/91.5.412.

54. Maslova A, Zlotina A, Kosyakova N, Sidorova M and Krasikova A. Three-dimensional architecture of tandem repeats in chicken interphase nucleus. Chromosome Res. 2015;23 3:625-39. doi:10.1007/s10577-015-9485-5.

55. Zlotina A, Maslova A, Kosyakova N, Al-Rikabi ABH, Liehr T and Krasikova A. Heterochromatic regions in Japanese quail chromosomes: comprehensive molecular-cytogenetic characterization and 3D mapping in interphase nucleus. Chromosome Res. 2019;27 3:253-70. doi:10.1007/s10577-018-9597-9.

56. Fishman V, Battulin N, Nuriddinov M, Maslova A, Zlotina A, Strunov A, 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. doi:10.1093/nar/gky1103.

57. Schield DR, Card DC, Hales NR, Perry BW, Pasquesi GM, Blackmon H, et al. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome Res. 2019;29 4:590-601. doi:10.1101/gr.240952.118.

58. Hooper DM and Price TD. Chromosomal inversion differences correlate with range overlap in passerine birds. Nat Ecol Evol. 2017;1 10:1526-34. doi:10.1038/s41559-017-0284-6.

59. Knief U, Hemmrich-Stanisak G, Wittig M, Franke A, Griffith SC, Kempenaers B, et al. Fitness consequences of polymorphic inversions in the zebra finch genome. Genome Biol. 2016;17 1:199. doi:10.1186/s13059-016-1056-3.

60. Craig RJ, Suh A, Wang M and Ellegren H. 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. doi:10.1111/mec.14462.

61. Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, Kent WJ, et al. Reconstructing contiguous regions of an ancestral genome. Genome Res. 2006;16 12:1557-65. doi:10.1101/gr.5383506.

62. Damas J, O'Connor R, Farré M, Lenis VPE, Martell HJ, Mandawala A, et al. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. Genome Research. 2017;27 5:875-84. doi:10.1101/gr.213660.116.

63. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491 7424:393-8. doi:10.1038/nature11622.

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

65. Evseev D and 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, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485 7398:376-80. doi:10.1038/nature11082.

67. Mirny LA, Imakaev M and Abdennur N. Two major mechanisms of chromosome organization. Curr Opin Cell Biol. 2019;58:142-52. doi:10.1016/j.ceb.2019.05.001.

68. Falk M, Feodorova Y, Naumova N, Imakaev M, Lajoie BR, Leonhardt H, et al. Heterochromatin drives compartmentalization of inverted and conventional nuclei. Nature. 2019;570 7761:395-9. doi:10.1038/s41586-019-1275-3.

69. Busslinger GA, Stocsits RR, van der Lelij P, Axelsson E, Tedeschi A, Galjart N, et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature. 2017;544 7651:503-7. doi:10.1038/nature22063.

70. Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nature Genetics. 2019;51 9:1380-8. doi:10.1038/s41588-019-0479-7.
71. Ibrahim DM and Mundlos S. Three-dimensional chromatin in disease: What holds us together and what drives us apart? Curr Opin Cell Biol. 2020;64:1-9. doi:10.1016/j.ceb.2020.01.003.
72. Harmston N, Ing-Simmons E, Tan G, Perry M, Merkenschlager M and Lenhard B. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. Nat Commun. 2017;8 1:441. doi:10.1038/s41467-017-00524-5.
73. Canela A, Maman Y, Jung S, Wong N, Callen E, Day A, et al. Genome Organization Drives Chromosome Fragility. Cell. 2017;170 3:507-21.e18. doi:10.1016/j.cell.2017.06.034.
74. Rutkowska J, Lagisz M and Nakagawa S. The long and the short of avian W chromosomes: no evidence for gradual W shortening. Biology Letters. 2012;8 4:636-8. doi:10.1098/rsbl.2012.0083.
75. Hammar BO. THE KARYOTYPES OF NINE BIRDS. Hereditas. 2009;55 2-3:367-85. doi:10.1111/j.1601-5223.1966.tb02056.x.
76. Schneider V and Church D. Genome reference consortium. The NCBI Handbook [Internet] 2nd edition. National Center for Biotechnology Information (US); 2013.
77. Solari AJ and 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. doi:10.1159/000133591.
78. Xu L, Auer G, Peona V, Suh A, Deng Y, Feng S, et al. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. Nat Ecol Evol. 2019;3 5:834-44. doi:10.1038/s41559-019-0850-1.
79. Suh A, Paus M, Kiefmann M, Churakov G, Franke FA, Brosius J, et al. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. Nat Commun. 2011;2:443. doi:10.1038/ncomms1448.
80. Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang 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. doi:10.1101/gr.099655.109.
81. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, et al. Topologically associating domains are stable units of replication-timing regulation. Nature. 2014;515 7527:402-5. doi:10.1038/nature13986.
82. Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, et al. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science. 2005;309 5734:613-7. doi:10.1126/science.1111387.
83. Malcolm S and Abu-Amero S. Faculty Opinions recommendation of Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature. 2012; doi:10.3410/f.14079956.15778060.
84. Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature. 2012;483 7387:82-6. doi:10.1038/nature10843.
85. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, et al. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature. 2003;423 6942:873-6. doi:10.1038/nature01723.
86. Mahajan S, C. Wei KH, Nalley MJ, Gibilisco L and Bachtrog D. De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. PLOS Biology. 2018;16 7:e2006348. doi:10.1371/journal.pbio.2006348.

87. Pendleton M, Sebra R, Pang AWC, Ummat A, Franzen O, Rausch T, et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat Methods. 2015;12 8:780-6. doi:10.1038/nmeth.3454.
88. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan 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. doi:10.1038/ng.3802.
89. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016;13 12:1050-4. doi:10.1038/nmeth.4035.
90. Melissa LS, Delany N, Hepler. N L, Alexander D, Katzenstein D, Brown M, 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 and Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19 1:460. doi:10.1186/s12859-018-2485-7.
92. Yuan Y, Bayer PE, Lee H-T and Edwards D. runBNG: a software package for BioNano genomic analysis on the command line. Bioinformatics. 2017;33 19:3107-9. doi:10.1093/bioinformatics/btx366.
93. Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 2019;15 8:e1007273. doi:10.1371/journal.pcbi.1007273.
94. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One. 2012;7 11:e47768. doi:10.1371/journal.pone.0047768.
95. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9 11:e112963. doi:10.1371/journal.pone.0112963.
96. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol Biol Evol. 2018;35 3:543-8. doi:10.1093/molbev/msx319.
97. Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, et al. Ensembl 2017. Nucleic Acids Res. 2017;45 D1:D635-D42. doi:10.1093/nar/gkw1104.
98. Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215 3:403-10.
99. Birney E, Clamp M and Durbin R. GeneWise and Genomewise. Genome Res. 2004;14 5:988-95. doi:10.1101/gr.1865504.
100. Stanke M, Schöffmann O, Morgenstern B and Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7:62. doi:10.1186/1471-2105-7-62.
101. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 2011;29 7:644-52. doi:10.1038/nbt.1883.
102. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9 1:R7. doi:10.1186/gb-2008-9-1-r7.
103. Tarailo-Graovac M and Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009;Chapter 4:Unit 4.10. doi:10.1002/0471250953.bi0410s25.
104. Bao W, Kojima KK and Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015;6 1 doi:10.1186/s13100-015-0041-9.

105. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research. 1999;27 2:573-80. doi:10.1093/nar/27.2.573.
106. Meyne J, Ratliff RL and Moyzis RK. Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci U S A. 1989;86 18:7049-53. doi:10.1073/pnas.86.18.7049.
107. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5 2:R12. doi:10.1186/gb-2004-5-2-r12.
108. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25 14:1754-60. doi:10.1093/bioinformatics/btp324.
109. 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. doi:10.1093/bioinformatics/btr509.
110. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20 1:224. doi:10.1186/s13059-019-1829-6.
111. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015;16:259. doi:10.1186/s13059-015-0831-x.
112. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. Transcriptome and chromatin structure annotation of liver, CD4 and CD8 T cells from four livestock species. doi:10.1101/316091.
113. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature Communications. 2018;9 1 doi:10.1038/s41467-017-02525-w.
114. Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, et al. DNA-binding specificities of human transcription factors. Cell. 2013;152 1-2:327-39.
115. Bailey TL, Elkan C, University of California SDDoCS and Engineering. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. 1994.
116. Servant N, Lajoie BR, Nora EP, Giorgetti L, Chen C-J, Heard E, et al. HiTC: exploration of high-throughput 'C' experiments. Bioinformatics. 2012;28 21:2843-4. doi:10.1093/bioinformatics/bts521.
117. Ardakany AR, Gezer HT, Lonardi S and Ay F. Mustache: Multi-scale Detection of Chromatin Loops from Hi-C and Micro-C Maps using Scale-Space Representation. bioRxiv. 2020.
118. Harris RS. Improved pairwise Alignmnet of genomic DNA. 2007.
119. Kim D, Langmead B and Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12 4:357-60. doi:10.1038/nmeth.3317.
120. Love MI, Huber W and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15 12:550. doi:10.1186/s13059-014-0550-8.
121. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005;21 5:650-9. doi:10.1093/bioinformatics/bti042.
122. Kryuchkova-Mostacci N and Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform. 2017;18 2:205-14. doi:10.1093/bib/bbw008.
123. Yang WR, Ardeljan D, Pacyna CN, Payer LM and Burns KH. SQuIRE reveals locus-specific regulation of interspersed repeat expression. Nucleic Acids Res. 2019;47 5:e27. doi:10.1093/nar/gky1301.