

Article

Open Access

Transposable elements shape the landscape of heterozygous structural variation in a bird genome

Bo-Ping Li^{1,2,3,#,*}, Na Kang^{2,#}, Zao-Xu Xu^{1,#}, Hao-Ran Luo^{2,4}, Shi-Yu Fan², Xiao-Han Ao², Xing Li², Ya-Peng Han¹, Xiao-Bin Ou^{1,*}, Luo-Hao Xu^{2,*}

¹ Gansu Key Laboratory of Protection and Utilization for Biological Resources and Ecological Restoration in Longdong, Longdong University, Qingyang, Gansu 745000, China

² Integrative Science Center of Germplasm Creation in Western China (Chongqing) Science City, MOE Key Laboratory of Freshwater Fish Reproduction and Development, School of Life Sciences, Southwest University, Chongqing 400715, China

³ College of Medicine, Longdong University, Qingyang, Gansu 745000, China

⁴ Key Laboratory of Ministry of Education for the Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen, Fujian 361102, China

ABSTRACT

Avian genomes exhibit compact organization and remarkable chromosomal stability. However, the extent and mechanisms by which structural variation in avian genomes differ from those in other vertebrate lineages are poorly explored. This study generated a diploid genome assembly for the golden pheasant (*Chrysolophus pictus*), a species distinguished by the vibrant plumage of males. Each haploid genome assembly included complete chromosomal models, incorporating all microchromosomes. Analysis revealed extensive tandem amplification of immune-related genes across the smallest microchromosomes (dot chromosomes), with an average copy number of 54. Structural variation between the haploid genomes was primarily shaped by large insertions and deletions (indels), with minimal contributions from inversions or duplications. Approximately 28% of these large indels were associated with recent insertions of transposable elements, despite their typically low activity in bird genomes. Evidence for significant effects of transposable elements on gene expression was minimal. Evolutionary strata on the sex chromosomes were identified, along with a drastic rearrangement of the W chromosome. These analyses of the high-quality diploid genome of the golden pheasant provide valuable insights into the evolutionary patterns of structural variation in avian genomes.

Keywords: Golden pheasant; Structural variation; Transposable elements; Chromosome evolution

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2025 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

INTRODUCTION

Aves, among the most species-rich vertebrate lineages, occupy nearly every global habitat (Armstrong et al., 2020). Avian genomes are characterized by distinct features, including a relatively small size compared to other vertebrates (Zhang et al., 2014), typically about 1.1 Gb, a feature attributed to a reduced abundance of repetitive sequences (Kapusta & Suh, 2017). Another hallmark of avian genomes is their highly conserved synteny, with very few intra- and interchromosomal rearrangements observed across divergent lineages (O'Connor et al., 2024). Despite these evolutionary constraints, rapid evolutionary changes occur in regions associated with traits such as feather pigmentation, beak morphology, and ecological adaptations (Feng et al., 2020). Notably, the genetic variations underlying such traits are frequently located in non-coding regions, highlighting their potential role in avian evolution (Seki et al., 2017; Yusuf et al., 2020).

Although avian genomes have been the focus of large-scale genomic analyses, most available genome assemblies have relied on short-read sequencing technologies (Bravo et al., 2021; Stiller et al., 2024), which can inherently limit resolution. Key genomic regions, including the smallest microchromosomes (dot chromosomes), have historically been absent from assemblies until recent advancements enabled their successful reconstruction (Huang et al., 2023; Luo et al., 2023). Other complex genomic regions, including the major histocompatibility complex (MHC) (Zhu et al., 2023), segmental duplicates (Wang et al., 2024), subtelomeric

Received: 02 November 2024; Accepted: 26 November 2024; Online: 27 November 2024

Foundation items: This work was supported by the Foundation of Gansu Key Laboratory of Protection and Utilization for Biological Resources and Ecological Restoration in Longdong (LDSWZY202103) and Natural Science Foundation of Gansu Province (22JR5RM210) to B.P.L., and Gansu Ziwuling Ecosystem Observation and Research Station (20JR10RA658)

*Authors contributed equally to this work

*Corresponding authors, E-mail: bopingli0616@126.com; xbou@zju.edu.cn; luohaox@gmail.com

regions (Li et al., 2022), and sex chromosomes (Benham et al., 2024; Xu et al., 2024a), remain underexplored, further emphasizing the need for comprehensive analyses to elucidate avian genomic architecture.

Comparisons between *de novo* genome assemblies are thought to provide more accurate structural variation (SV) callings than read-mapping-based approaches. However, such comparisons require haplotype-resolved genome assemblies, which remain scarce for many taxa. Previous studies using diploid genome assemblies have revealed significant diversity in SVs between haploid genomes in mammals and plants (Sun et al., 2020; Yang et al., 2021). For instance, 11 663 SVs larger than 50 bp have been identified between the haploid genomes of a primate species, including a large 304 kb inversion (Yang et al., 2021). Despite advances in genomic sequencing, very few bird genome assemblies have been resolved to the haplotype level with chromosome-scale phasing. Emerging technologies, such as the integration of High-fidelity (HiFi) and Hi-C sequencing, enable scalable diploid genome assembly (Cheng et al., 2024) and have been applied to species such as turkeys (Barros et al., 2023) and chukars (Zhou et al., 2024). However, in-depth analyses of SVs in bird diploid genomes remain limited. Recent graph-based pangenome studies of 30 chicken genomes have demonstrated the effectiveness of haplotype-resolved genome assemblies for accurate SV calling in avian species.

The golden pheasant (*Chrysolophus pictus*), known for its beautiful plumage in males, has been introduced to various countries since the 18th century and shares its family, Phasianidae, with domesticated chicken (Xu et al., 2024b). Subject to extensive artificial breeding, this species offers a unique model for genomic exploration. Using a combination of Illumina short reads, Nanopore ultra-long reads, PacBio HiFi reads, and Hi-C mapping, coupled with a trio-binning assembly strategy (Koren et al., 2018), we generated near-complete diploid genome assemblies for the golden pheasant, resolving both haploid genomes to the chromosome level. Analyses revealed frequent tandem amplifications of immune-related genes on dot chromosomes. Additionally, intra- and interchromosomal segmental duplications (SDs) were characterized, alongside a landscape of SVs primarily shaped by transposable element (TE) insertions. These findings provide valuable insights into the structural dynamics of avian genomes and highlight the golden pheasant as a genomic model for investigating evolutionary processes.

MATERIALS AND METHODS

Sample collection

An F1 female golden pheasant and her parents were obtained from a local farm in Qingyang, Gansu Province, China, and raised at the Longdong College Agricultural Science and Technology Park under an artificial breeding program. The F1 chick, which died of natural causes, was immediately dissected. Muscle tissue was collected for long-read sequencing, while various tissues were preserved for RNA sequencing (RNA-seq) (Supplementary Table S1). Blood samples (1 mL) were collected from each parent for short-read sequencing. All research protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Southwest University (Approval No. IACUC-20241031-06).

HiFi sequencing

HiFi sequencing was performed on the F1 female golden pheasant following the protocols provided with the SMRTbell®

Prep Kit 3.0 (Pacific Biosciences, USA). DNA quality was assessed using the Femto Pulse system (Agilent Technologies, USA) and DNA shearing and cleanup were conducted using Megaruptor 3 (Diagenode, Belgium). The preparation process included DNA repair, A-tailing, adapter ligation, and cleanup steps, followed by nuclease treatment and library size selection. Sequencing was performed on the PacBio Revio system (Pacific Biosciences, USA), with high-precision HiFi reads generated using CCS software (<https://github.com/PacificBiosciences/ccs>) with default parameters (minimum pass number=3, minimum RQ=0.99), achieving read quality exceeding Q20.

Ultralong Nanopore sequencing

An ultralong-read sequencing library was constructed using the SQK-LSK114 Kit (Oxford Nanopore Technologies, UK) for the same F1 female golden pheasant, with sequencing performed on a PromethION instrument (Oxford Nanopore Technologies, UK). DNA preparation involved dilution, end-repair, and purification. Specifically, after thawing and centrifuging (20°C, 10 s, 500 ×g) the DNA reference (DCS), 1 µg of genomic DNA was diluted to 47 µg with nuclease-free water. The reaction mixture, prepared in a polymerase chain reaction (PCR) tube, consisted of the DNA sample, ligation buffer, fast T4 DNA ligase, and adapter. Each component was gently mixed after addition, followed by gentle centrifugation and incubation at 20°C and 65°C for 5 min each. AMPure XP beads were resuspended and added to the reaction mixture, which was incubated at room temperature for 5 min. The beads were washed twice with 70% ethanol and resuspended in 61 µg of nuclease-free water. The eluate was transferred to a new tube for Qubit quantification (Thermo Fisher Scientific, USA). Ligation adapters and T4 DNA ligase were added to complete the adapter ligation, finalizing the preparation for sequencing.

Hi-C sequencing

Genomic DNA was extracted from the spleen of the F1 female golden pheasant to construct Hi-C libraries, sequenced on the Illumina MGI-2000 platform (MGI, China). Approximately 2 g of tissue was cut into strips and fixed with 2% formaldehyde in NIB buffer at 4°C under vacuum. Samples were subsequently washed, frozen, ground, and filtered. Nuclei were lysed with 0.1% sodium dodecyl-sulfate (SDS) at 65°C, and DNA was digested with DPN II at 37°C. Biotinylated nucleotides were used to label restriction ends, followed by blunt-end ligation at 16°C overnight. Crosslinks were reversed using proteinase K and DNA was purified using a QIAamp DNA Mini Kit (Qiagen, Germany). The purified DNA was then sheared to approximately 400 bp for sequencing.

RNA-seq

Total RNA was extracted from the heart, glandular stomach, kidney, gizzard, liver, bone, and muscle tissues. RNA quality was checked using 1% agarose gels, Qubit® RNA Assay Kit with a Qubit® 2.0 Fluorometer (Life Technologies, USA), and RNA Nano 6000 Assay Kit on the Bioanalyzer 2100 system (Agilent Technologies, USA).

RNA library preparation was performed with 1 µg of input RNA and the TruSeq RNA Library Preparation Kit (Illumina, USA). mRNA was purified from total RNA using poly-T magnetic beads. First-strand cDNA synthesis was carried out with random hexamer primers and M-MuLV reverse transcriptase, followed by second-strand synthesis with DNA

polymerase I and RNase H. Overhangs were converted to blunt ends, and Illumina adapters were ligated after adenylating the 3' ends. cDNA fragments of 150–200 bp were purified using the AMPure XP system, followed by PCR amplification with Phusion High-Fidelity DNA Polymerase and indexed primers. The libraries were purified again and assessed for quality on an Agilent Bioanalyzer 2100 (Agilent Technologies, USA). The libraries were then clustered using the cBot system and sequenced on the Illumina NovaSeq platform (Illumina, USA), generating 150 bp paired-end reads.

Short-read sequencing

Genomic DNA from parental blood samples was extracted using the cetyltrimethylammonium bromide (CTAB) method. Library preparation was performed with the MGIEasy Universal DNA Library Preparation Kit v.1.0 (MGI, China). Genomic DNA (1 µg) was fragmented to an average size of 200–400 bp using a Covaris instrument, and fragments were screened using MGIEasy DNA Clean beads (MGI, China). The selected fragments were subjected to end-repair, adenylation, and adapter ligation, followed by PCR amplification and purification with the same beads. The double-stranded PCR products were heat denatured and circularized to obtain single-stranded circular DNA (ssCir DNA) using MGIEasy circulation software. Final library quality was verified and sequenced using the DNBSEQ-T7RS platform (MGI, China).

Genome assembly and assessment

A high-quality diploid genome assembly for the golden pheasant was constructed using a trio-binning strategy, combining short-read sequencing data from the two parents and long-read sequencing data from the F1 individual. Assembly was performed with Hifiasm (v.0.19.3) in trio mode (Cheng et al., 2021) under default parameters, generating two sets of contig-level haploid genomes. Hi-C data facilitated scaffolding of the contigs using the Juicer (v.1.5) pipeline (Durand et al., 2016b). Visualization of Hi-C heatmaps using Juicebox (v.1.11.08) (Durand et al., 2016a) enabled manual adjustment of contig order and orientation. Genome assembly completeness was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO v.5.5.0) (Di Tommaso et al., 2017) with the *aves_odb10* lineage ($n=8\ 338$) in genome mode.

Genome annotation

Repetitive sequences were annotated by constructing a custom repeat library with RepeatModeler (v.2.0.3) (Flynn et al., 2020) and Satellite Repeat Finder (SRF) (Zhang et al., 2023). The resulting library was combined with the chicken repeat library (Peona et al., 2021) and used to mask repeat sequences with RepeatMasker (v.4.1.2-p1). After soft-masking the repeats, the gene models were annotated based on homologous proteins, *de novo* prediction, and RNA-seq data. RNA-seq data from seven tissues (Supplementary Table S1) were assembled into transcripts using Trinity (v.2.8.5) (Grabherr et al., 2011). Gene predictions incorporated evidence from assembled transcripts and homologous proteins through MAKER (v.2.31.10) (Campbell et al., 2014; Grabherr et al., 2011). Transcripts were further assembled using a genome-guided approach with the HISAT2 (v.2.2.1)-StringTie (v.2.2.1) pipeline (Kim et al., 2015; Sirén et al., 2014). Gene model training was performed with Augustus (v.3.5.0) (Stanke et al., 2008) through the BUSCO (v.5.5.0)

pipeline (Di Tommaso et al., 2017), with gene models predicted using the trained profile. The predicted gene models and aligned transcripts were integrated by EvidenceModeler (EVM, v.1.1.1) (Haas et al., 2008), and gene models were polished using the PASA pipeline (v.2.4.1) (Haas et al., 2008) with the transcripts generated by Trinity. Chromosome 16 was manually annotated using IGV-GSaman (Haas et al., 2008) based on RNA-seq transcript data.

Hi-C chromatin interaction

The BWA-MEM algorithm (v.0.7.17-r1188) was used to align Hi-C reads with the parameters “-A 1 -B 4 -E 50 -L 0 -t 16”. The hicBuildMatrix function from HiCExplorer (v.3.6) was applied with the parameters “-binSize 100000, -inputBufferSize 400000, -restrictionSequence GATC”, producing a Hi-C matrix file as output. This matrix was then normalized using hicCorrectMatrix with default parameters to account for the number of restriction sites per bin and biases related to open chromatin. The resulting h5 file was converted into ginteractions format using hicConvertFormat and subsequently used to calculate chromosomal interaction frequencies with custom Perl scripts (https://github.com/lurebgi/chicken-T2T/blob/main/Hi-C_analysis).

DNA methylation

Nanopore reads were aligned against the reference genome using minimap2 (v.2.26) (Li and Alkan, 2021) with the “map-on” setting. Nanopolish (v.0.13.2) (Loman et al., 2015) was used to detect 5-methylcytosine bases in a CpG context. Methylation frequencies at individual sites were calculated using the script “calculate_methylation_frequency” from the Nanopolish package. Methylation data were further summarized in 500 bp windows to provide an overview of methylation patterns.

Gene expression

RNA-seq datasets used for expression analysis are shown in Supplementary Table S1. The HISAT2-StringTie pipeline (Kim et al., 2015; Sirén et al., 2014) was employed to map raw RNA-seq reads and calculate transcripts per million (TPM) values. For each gene, the mean expression level across tissues was determined to provide a comprehensive view of gene expression patterns.

Phylogenetic analysis

Phylogenetic trees were constructed using the longest protein sequences from the mallard (*Anas platyrhynchos*, GCA_015476345.1), chicken (*Gallus gallus*, GCA_024206055.2), common pheasant (*Phasianus colchicus*) (He et al., 2021), Swinhoe's pheasant (*Lophura swinhoii*, GCA_030408155.1), wild turkey (*Meleagris gallopavo*, GCA_000146605.4), Lady Amherst's pheasant (*Chrysolophus amherstiae*) (Garg et al., 2024), and golden pheasant (this study) (Bian et al., 2024). Single-copy orthologous genes among these species were identified using OrthoFinder (v.2.5.5) (Emms & Kelly, 2019) (Supplementary Table S2). Protein sequences within each orthologous group were aligned using MUSCLE (v.5.1.linux64) (Edgar, 2004), and the alignments were further filtered with the strict model in trimAl (v.1.4.rev15) (Capella-Gutiérrez et al., 2009). The aligned sequences of all orthologous groups for each species were concatenated to construct a species tree using IQ-TREE (v.2.1.2) (Nguyen et al., 2015). Divergence time estimations were determined using the approximate likelihood calculation method in PAML-MCMCTREE (PAML v.4.9j). Initial branch

lengths were estimated with PAMLCodeML, using gradient and Hessian information for the ultrametric tree, which was taken as input for MCMCTREE analysis (Nguyen et al., 2015). Markov chain Monte Carlo (MCMC) chains were run for 20 million steps to estimate divergence times. Fossil records of Galloanserae (66–86.8 million years (Ma)) were used for calibration (Clarke et al., 2005), along with TimeTree5 divergence estimates for chicken and turkey (26.02–31.55 Ma) (Li et al., 2015; Stein et al., 2015).

Single nucleotide polymorphism (SNP) heterozygosity rate analysis

Short reads were mapped to the CPswu2 reference genome using BWA (v.0.7.17) (Li & Durbin, 2009) with default parameters. BAM files were sorted, and duplicate reads were filtered using SAMtools (v.1.18) (Li & Durbin, 2009; Li et al., 2009) and Picard (v.1.56) (<http://broadinstitute.github.io/picard>). SNPs were called using the HaplotypeCaller and GenotypeGVCFs commands in the Genome Analysis ToolKit (v.3.8), then filtered using the VariantFiltration command with the following criteria: "DP<1.0/3.0; DP>3 || QD<2.0 || FS>60.0 || SOR>3.0 || MQ<40 || MQRankSum<-12.5 || ReadPosRankSum<-8.0". Finally, all SNPs were exported in variant call format (VCF) files (Liu et al., 2022). The SNP heterozygosity rate was calculated by dividing the number of heterozygous SNPs by the total number of genome bases (Wu et al., 2023).

Demographic history

The historical effective population size (N_e) of the golden pheasant and Lady Amherst's pheasant was estimated using the Pairwise Sequentially Markovian Coalescent model (PSMC, v0.6.5) (Liu & Hansen, 2016). Parameters were set to "-N30 -t15 -r5 -p '4+25*2+4+6'". The mutation of the golden pheasant was assumed to be $4.02e-9$, with a generation time of 2 years based on Xu et al. (2024b).

Synteny analysis

Gffread (v.0.12.7) (Pertea & Pertea, 2020) was used to extract the protein sequences of each chromosome from the GGswu1 (Huang et al., 2023) and CPswu2 genomes. Synteny analysis was conducted using MCScan within the JCVI package (Tang et al., 2008). Macrosynteny for all chromosomes was visualized with "Jcvi.graphics.karyotype", while microsynteny, including analyses of chromosome 16 and chrW, was visualized using "jcv.graphics.synteny".

Tandem amplification analysis

Dot chromosome annotation was performed using Miniprot (v.0.12-r237) and Augustus (v.3.5.0), identifying extensive tandem amplification of immune-related genes, including *WC1.1*, *CHIR* (Ig-like receptor) and *VH1* (heavy chain variable region of immunoglobulins). Gffread (v.0.12.7) and Seqkit (v.2.8.0) were used to extract the protein sequence of each gene, which were verified through BLAST on the National Center for Biotechnology Information (NCBI) database. MUSCLE (v.5.1.linux64), trimAl (v.1.4.rev15) and IQ-TREE (v.2.1.2) were used to build the phylogenetic trees (see Phylogenetic analysis). Additional tandem duplications were identified for other genes, such as *ANXA4* (annexin A4) and OR (olfactory receptor), which were analyzed using the same approach.

Segmental duplications

Segmental duplications (SDs) were identified with SDquest

(v.0.1) (Pu et al., 2018) after masking repeats. Criteria for SD detection included sequence identity greater than 90% and lengths exceeding 2 kb. The identified SDs were further classified as interchromosomal or intrachromosomal. In total, 32 unique SDs were found in the golden pheasant, all present as single copies in chicken. For example, on chromosome 15, a segmental duplication involving *P2RX6* and nearby genes was detected in the golden pheasant. Hisat2 (v.2.2.1) was used to map the RNA-seq data of various tissues from the golden pheasant and chicken to the genome and Integrative Genomics Viewer (IGV) (Robinson et al., 2017) was used for visualization (Supplementary Figure S1).

SV identification and analysis

SVs were identified using nucmer (v.4.0.0) to align the two haploid genomes with parameters -b 500 to extend alignment clusters by 500 bp, followed by delta-filtering with stringent criteria (-l -i 90 -l 200) to exclude alignments with <90% similarity or lengths <200 bp. SVs were identified using Assemblytics (v.1.2.1) (Nattestad & Schatz, 2016) and SyRi (v.1.6.4) (Goel et al., 2019), and results were merged for downstream analyses. HiFi read alignments were manually inspected in IGV to validate SV boundaries. RepeatMasker (v.4.1.2-p1) was used to perform genome-wide transposon analysis and parseRM.pl was used to analyze insertion time with the parameters "-p -l 50.1 -m 0.0019 -v" (Zhang et al., 2014). After extracting large indel intervals, the intersect and closest functions in bedtools (v.2.30.0) were applied to search for long terminal repeats (LTRs) and analyze the distance to the nearest exon. Expression analysis of SNPs proximal to large indels across the two haplotypes was conducted using the Genome Analysis ToolKit (v.3.8) and RNA-seq data from various tissues. VCF files were compared and analyzed (please refer to SNP heterozygous rate analysis).

Sex chromosome evolution

Evolutionary strata on the Z and W chromosomes were identified by first masking repetitive elements, followed by alignment of Z and W chromosomes using lastz (v.1.04.03) with parameters ("--step=19 --hsptresh=2 200 --inner=2 000 -ydrop=3 400 --gappedthresh=10 000") (Huang et al., 2022). Sequence similarity between the Z and W chromosomes was calculated using the psiScore script from the UCSC Genome Browser (<https://genome.ucsc.edu/>). Alignments with sequence similarity below 60% were excluded to reduce false positives, while those exceeding 90% were removed to account for potential unmasked repeats. Sequence similarity between the Z and W chromosomes was summarized over 100 kb non-sliding windows along the Z chromosomes.

RESULTS

Diploid genome assembly with complete chromosome models

A female chick from a golden pheasant trio family was sampled for haplotype-resolved diploid genome assembly. Sequencing generated 55.1 Gb of Nanopore, 64.7 Gb of HiFi, and 93.4 Gb of Hi-C sequencing reads for the F1 chick (Supplementary Table S1), along with 28.9 Gb and 36.5 Gb of short reads from the female and male parents, respectively. Genome assembly using Hifiasm in trio mode produced two phased haploid genomes, designated CPswu1 (maternal) and CPswu2 (paternal). Both haploid genomes exhibited high completeness, with total sequence lengths of 1.00 Gb and

1.06 Gb, respectively (Supplementary Table S3), comparable to the previous GCA_003413605.1 assembly (1.02 Gb) (Gao et al., 2018). CPswu1 contained 227 contigs with a N50 size of 33.0 Mb, while CPswu2 contained 198 contigs with a N50 size of 37.8 Mb, significantly improving on 38.2 kb in GCA_003413605.1. Hi-C data further anchored these contigs into chromosome models for both haplotypes. As expected, CPswu1 included a W chromosome, while CPswu2 included a Z chromosome. The assemblies revealed 41 chromosome models for each haploid genome (40 autosomes and one sex chromosome) (Supplementary Figure S2), consistent with the known karyotype of the golden pheasant. Autosomal models were categorized into 11 macrochromosomes, 19 microchromosomes, and 10 dot chromosomes based on size and epigenetic features (Figures 1A, 2A). BUSCO scores of 97.0% and 96.9% confirmed the high completeness of CPswu1 and CPswu2, respectively (Supplementary Figure S3). Repeat content was comparable between the two haploid genomes (15.1% vs. 14.5%, Supplementary Table S4). Long interspersed nuclear elements (LINEs) were dominant among TEs (Jae Lee et al., 2021), accounting for 6.96% and 7.43% of the genomes, respectively (Supplementary Figure S4) (Jin et al., 2022). CPswu1 and CPswu2 contained 18 and 19 chromosome models with no gaps, respectively, while 114 and 99 gaps were scattered across the genomes, predominantly in subtelomeric regions of macrochromosomes and on dot chromosomes (Supplementary Table S5).

Phylogenetic analysis was conducted using 3 192 single-

copy orthologous genes from seven avian species (Supplementary Tables S1, S6) (Zhao et al., 2021), including six Galliformes species (chicken, golden pheasant, Swinhoe's pheasant, common pheasant, Lady Amherst's pheasant, and wild turkey) and one outgroup species (mallard). Results estimated that the golden pheasant diverged from other Galliformes species approximately 7.98 Ma (Figure 1B).

The golden pheasant, which is not considered threatened and can be artificially bred, exhibited an estimated heterozygosity rate of 0.53%, higher than that of most Galliformes species investigated (Figure 1C). Historical effective population size estimates revealed larger population sizes for the golden pheasant compared to its close relative, Lady Amherst's pheasant, over much of their evolutionary history (Figure 1D).

Restoration of acrocentric morphology of macrochromosomes

In the absence of CENP-A ChIP-seq data, centromeres in CPswu1 were located by identifying homologous centromeric regions in the chicken genome. Despite the overall high synteny between chicken and golden pheasant chromosomes, frequent telomeric inversions were observed in macrochromosomes (Figure 2B). For instance, telomeric inversions in chr7 and chr8 involved the entire short arm, relocating centromeres to the chromosome ends in the golden pheasant. Additionally, a macrochromosome fission event in chr2 was identified, consistent with findings from previous

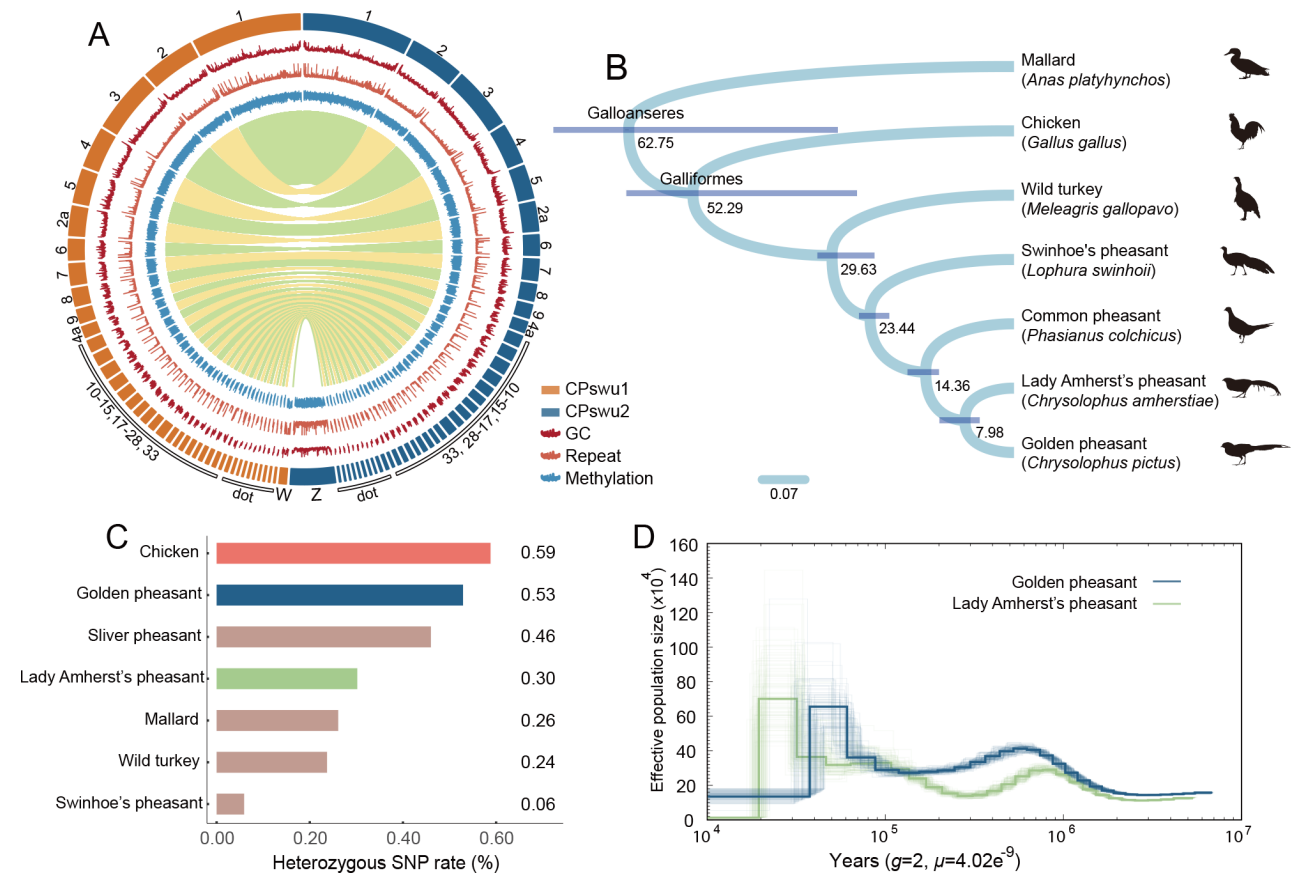


Figure 1 High-quality chromosome assembly and genomic features of golden pheasant

A: Circles show distribution of GC, repeat, and methylation content in CPswu1 and CPswu2 genomes. Synteny between two haplotypes is shown in the center. B: Phylogenetic tree was inferred with single-copy orthologs. Galloanseræ fossil records (66–86.8 Ma) and divergence time for chicken and turkey (26.02–31.44 Ma) were used for calibration. C: Golden pheasant showed relatively high heterozygosity among seven Galloanseræ genomes. D: Golden pheasant showed a higher effective population size than Lady Amherst's pheasant.

comparative cytogenetic analyses. Interestingly, the split sites coincided with chicken centromeres, resulting in two acrocentric chromosomes (Figure 2B).

To assess the potential impact of altered chromosome morphology on 3D genome organization in the nucleus, interchromosomal chromatin interactions were analyzed using Hi-C data. No interchromosomal interactions involving centromeres were detected; however, frequent interactions among macrochromosome telomeres were identified (Figure 2A). Moreover, while expected frequent interactions among microchromosomes were observed (Liu et al., 2021), intensive interactions between macrochromosome euchromatin and microchromosomes were also detected (Figure 2A; Supplementary Figure S5). These findings

suggest that macrochromosome euchromatin may position closer to the nuclear interior (Figure 2C) and potentially contribute to the regulation of microchromosome gene expression.

Tandem amplification of immune genes

The duplication of immune-related genes was analyzed with a focus on chr16, one of the most complex bird chromosomes, known for harboring many immune-related genes including the MHC region (Kulski et al., 2019). The assembled sizes of chr16 in the golden pheasant were 4.61 Mb (CPswu1) and 4.64 Mb (CPswu2), similar to that of the chicken (4.93 Mb). The euchromatic regions of chr16 exhibited high synteny between the two species, except for two inversions near the MHC-Y region (Figure 3A).

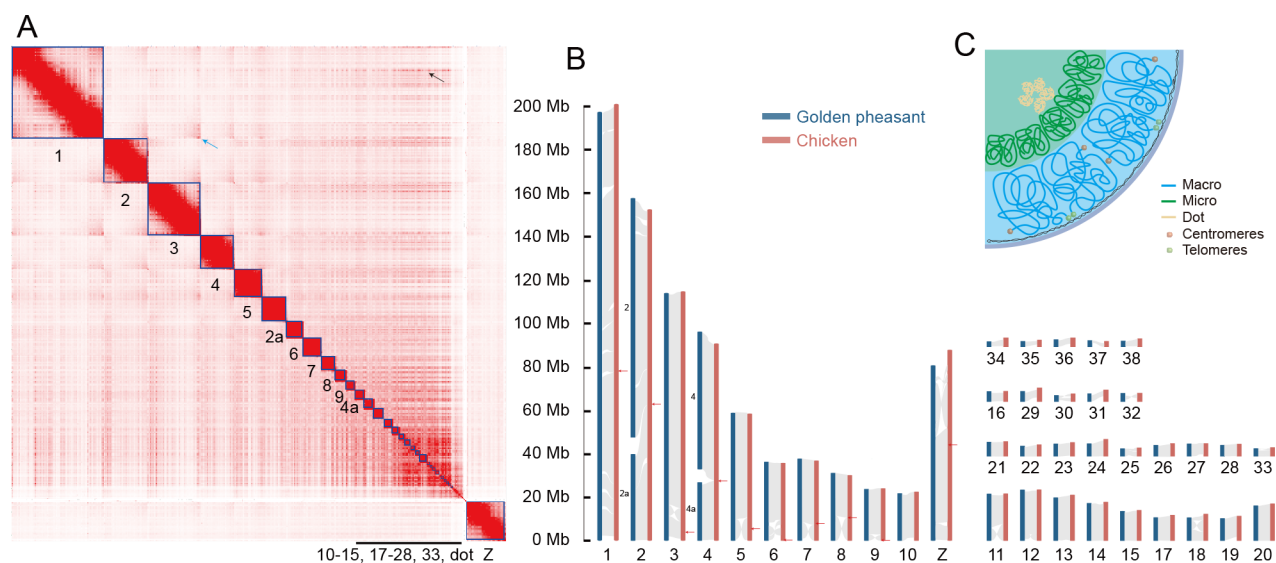


Figure 2 Acrocentric chromosome morphology

A: Hic interaction map showing frequent interactions between macrochromosome telomeric regions (blue arrow) and between macrochromosome euchromatin and microchromosomes (black arrow). B: Conserved chromosomal synteny between chicken and golden pheasant inferred by MCSScan. Red arrow indicates chicken centromeres. C: Proposed chromosome positioning in the nucleus.

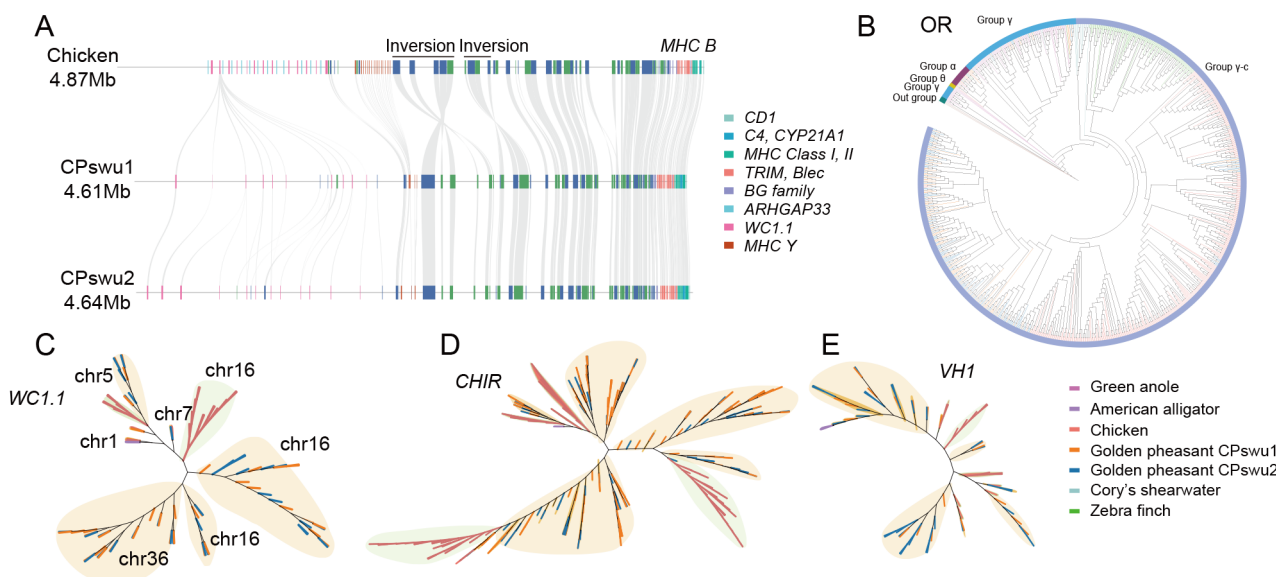


Figure 3 Amplification of immune-related genes on dot chromosomes

A: Microsynteny visualization of chr16 between chicken and diploid golden pheasant genomes. B: Phylogenetic tree for amplified genes of olfactory receptor (OR) re-rooted with adenosine receptor A2b from GenBank. C–E: Phylogenetic trees for amplified genes *WC1.1* (C), *CHIR* (immunoglobulin-like receptor) (D), and *VH1* (immunoglobulin heavy chain variable) (E) re-rooted with American alligator orthologs.

Manual annotation revealed 137 protein-coding genes in chicken chr16, and 111 and 110 genes in CPswu1 and CPswu2, respectively. Although the euchromatic regions of chr16 shared similar gene content between the chicken and golden pheasant (Supplementary Table S7), the heterochromatic regions displayed notable differences (Figure 3A). In the heterochromatic part of chr16, 10 and 12 copies of the *WC1.1* gene were identified in CPswu1 and CPswu2, respectively (Figure 3A), compared to 13 copies in the chicken chr16. Phylogenetic analysis suggested independent amplification of *WC1.1* in the two species (Figure 3C). Furthermore, golden pheasants exhibited further amplification of *WC1.1* on chr36, another dot chromosome, an event absent in chicken (Figure 3C).

Beyond chr16, amplification of other immune-related genes was detected on dot chromosomes. *CHIR* (Ig-like receptor) and *VH1* (heavy chain variable region of immunoglobulins) were amplified on chr31 and chr34, respectively. On chr31, 96 copies of *CHIR* were annotated in the chicken, compared to 136 copies in CPswu1 and 80 copies in CPswu2. Phylogenetic analysis indicated that *CHIR* likely underwent three duplication events before divergence of the chicken and golden pheasant (Figure 3D), followed by species-specific amplifications, resembling the scenario for *WC1.1*. Similarly, the *VH1* gene was duplicated multiple times in the Phasianidae ancestor but experienced more extensive amplification in the golden pheasant compared to the chicken (Figure 3E).

Amplification of other genes

In addition to immune-related genes, several other amplified genes were identified (Supplementary Tables S8, S9), including annexin A4 (*ANXA4*) (Supplementary Figure S6) and the olfactory receptor (OR) gene family (Figure 3B). *ANXA4*, which encodes a protein that potentially interacts with adenosine triphosphate (ATP) and is normally expressed almost exclusively in epithelial cells, was recurrently amplified on four different dot chromosomes in the golden pheasant: chr16, chr30, chr36, and chr37, with 15, five, 15, and five copies, respectively. This amplification was absent in chicken. The OR gene family exhibited significant amplification on chr29 in both chickens and golden pheasants (Supplementary Table S9).

Segmental duplications contribute to gene copy number variation

A total of 155 intrachromosomal SDs (Figure 4A) and 87 interchromosomal SDs (Figure 4B) with alignment identities higher than 90% and lengths larger than 2 kb were identified (Supplementary Tables S10, S11). Among these, 34.7% exhibited sequence identities exceeding 97% between pairs (Figure 4C), suggesting recent duplication events. Gene duplications were associated with 32 (or 13.3%) of the identified SDs (Supplementary Table S12). For instance, *P2RX6* and its nearby genes were duplicated due to a ~15 kb SD (Figure 4D). This duplication was present in both haploid genomes and validated through long-read mapping

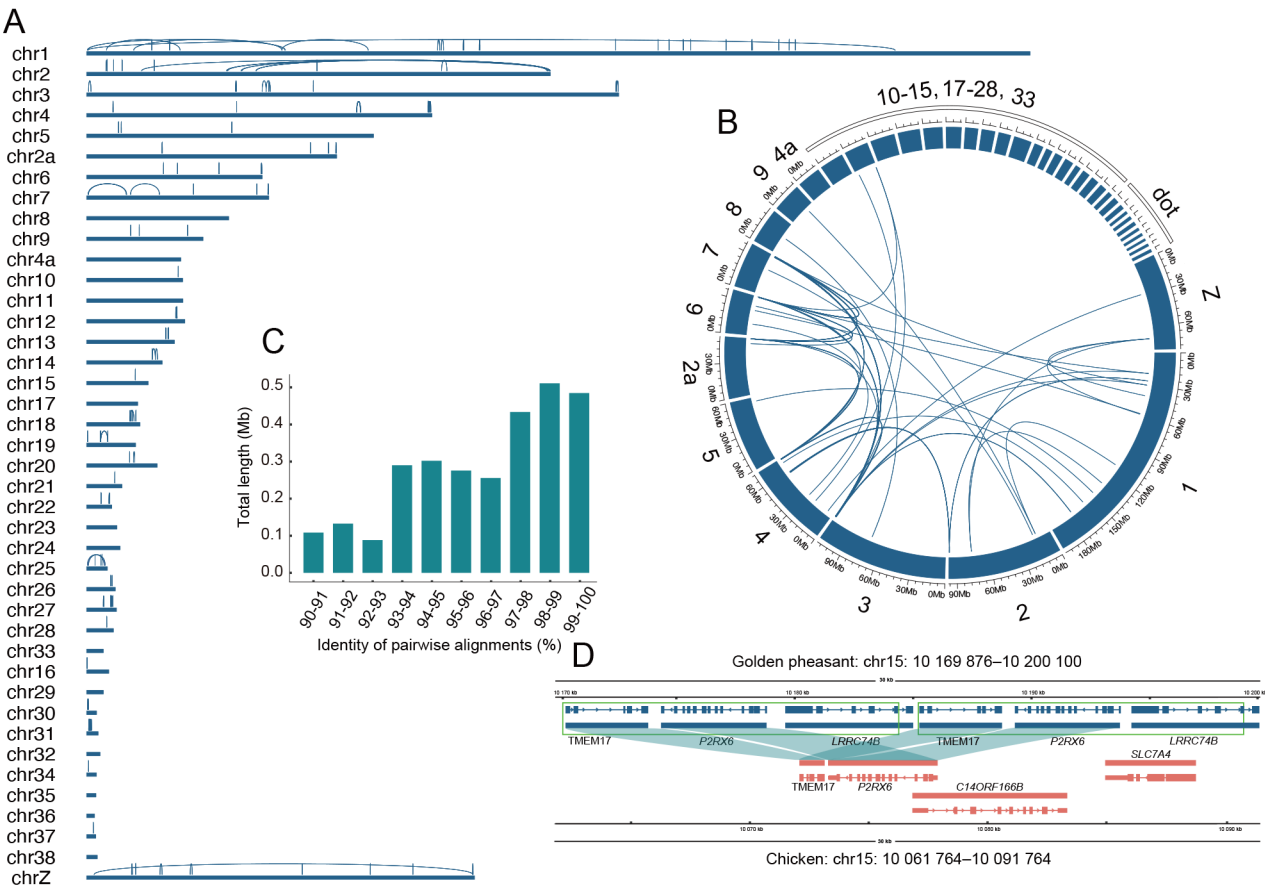


Figure 4 Segmental duplication (SD) landscape

A, B: Intrachromosomal (A) and interchromosomal (B) SDs are shown. C: Histogram for total lengths of SDs of varying identities. D: Example of SD involving *P2RX6*, *TMEM17*, and *LRRCT4B*, with single copies in chicken but two copies in golden pheasant.

(Supplementary Figure S6). The expression patterns of the duplicated *P2RX6* remained consistent with those of the original copy (Supplementary Figure S6), a trend observed for most duplicated genes (Supplementary Table S13).

Frequent large indels driven by recent TE insertions

The diploid genome assembly enabled the identification of SVs between the two haplotypes of the golden pheasant genome. A total of 11 432 SVs were detected, comprising 9 576 large indels (>50 bp), 1 747 copy-number variants, 64 translocations, eight inversions, and 37 inverted translocations (Figure 5A; Supplementary Table S14). Large indels (83.8%) represented the majority of SVs, while inversions (0.1%) were rare. A disproportionate number of SVs were observed in the heterochromatin of dot chromosomes (Supplementary Figure S7), possibly due to the low recombination rates in these regions.

Over 91.4% of the large indels were shorter than 1 kb, with a notable enrichment of indels around 7 kb in length (Figure 5B). Indels measuring 5–8 kb accounted for 15.8% of the total indel length (Figure 5C). These larger indels were predominantly associated with TEs, particularly LTRs. The large LTRs linked to these indels ranged from 5–10 kb in size (Figure 5C), consistent with full-length LTRs. Supporting this, the insertion times of indel-associated LTRs were more recent than the genome-wide average, with multiple waves occurring in the last 30 Ma (Figure 5D; Supplementary Table S15).

To evaluate the functional impact of large indels, their distance to gene bodies was analyzed. Large indels were found to occur further away from gene bodies than expected, with lower proportions located within gene bodies or their flanking regions (Figure 5E). Furthermore, few genes near large indels exhibited biased allele expression, suggesting minimal disruption to gene regulation (Supplementary Figure S8).

Structurally rearranged W chromosome with conserved gene expression

The assembled W chromosome of the golden pheasant was 12.9 Mb in size, comparable to the 14.2 Mb chicken W chromosome. Similar to the chicken W assembly, the golden pheasant W assembly contained 14 gaps, suggesting that the sequence remains incomplete. Repetitive sequences accounted for 87.45% of the golden pheasant W chromosome, with ERVK being the dominant TE family in both species (Figure 6E). Annotation identified 29 non-redundant W-linked genes in the golden pheasant, 28 of which have homologs in the chicken (Figure 6A; Supplementary Table S16). The chicken W chromosome contained only two additional genes (Figure 6D), highlighting the high conservation of W chromosome gene content between the two species. However, the chromosomal locations of these conserved genes were not collinear, likely due to frequent inversions.

Gene expression analysis revealed no significant differences in W-linked gene expression between the two birds. Consistent with earlier findings (Xu et al., 2019), W-linked genes in both species were widely expressed across tissues, with most exhibiting high expression levels (Figure 6B). Remarkably, the expression patterns of W-linked genes remained highly conserved, despite ~50 Ma of divergence and differences in sexual dimorphism between the two species. Finally, an analysis of sex chromosome evolutionary strata due to recombination suppression (Figure 6C) indicated that the two birds share a common history of sex chromosome degeneration, reflected in the similar number and length of these strata.

DISCUSSION

The SV landscape in avian genomes remains poorly understood, largely due to the limited availability of haplotype-

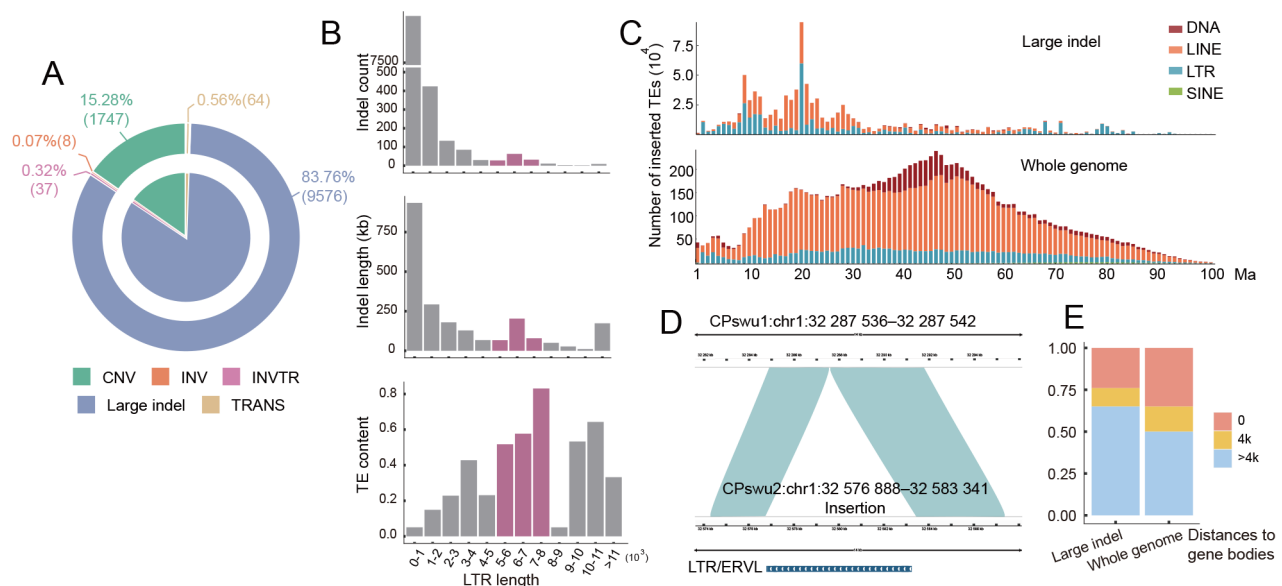


Figure 5 Frequent large indels caused by recent TE insertions

A: Proportion and number of each SV type, including copy number variants (CNV), single nucleotide variant (SNVs), inversions (INVs), translocations (TRANS), and inverted translocations (INVTRs). B: Number and total length of large indels, and TE content over different lengths of LTRs. Results showed enrichment of large LTRs in the 5–8 kb range, with those LTRs exhibiting higher TE content. C: More recent TE insertions in large indel regions relative to whole-genome background. D: Example of a large indel involving a full-length LTR in chr1. E: Large indels tended to be distant from gene bodies or gene-flanking regions. X-axis refers to distance from gene bodies.

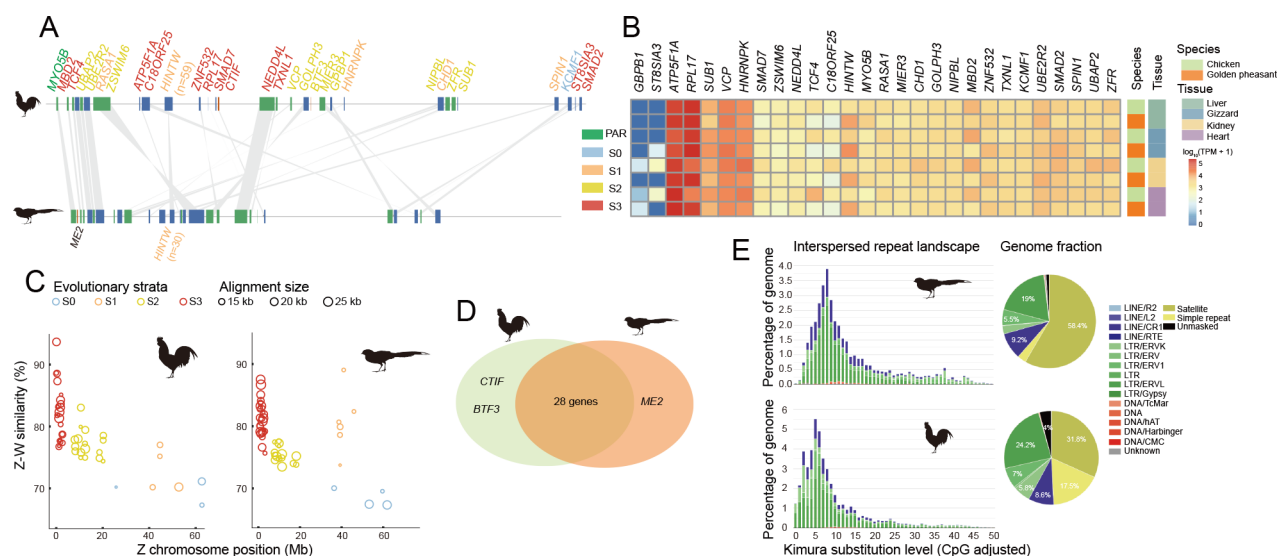


Figure 6 Evolutionary history of golden pheasant sex chromosomes

A: Gene order of W chromosome was extensively rearranged. B: Gene expression profile of the W-linked genes was conserved between both birds. C: Sequence divergence of Z and W chromosomes revealed a pattern of evolutionary strata with shared evolutionary history between the two species. X-axis shows the position of the Z chromosome, Y-axis shows the sequence similarity of Z and W. D: A large proportion of W-linked genes were shared by both birds. E: Divergence of TEs on W chromosomes of chicken and golden pheasant.

resolved genome assemblies or long-read-based pangenomes (Fang & Edwards, 2024). By employing a trio-binning approach, this study assembled a near-complete diploid genome of the golden pheasant, achieving telomere-to-telomere completeness for nearly half of its chromosomes. This level of resolution is unprecedented for avian genome assemblies, where small microchromosomes (dot chromosomes) are frequently missing (Ko et al., 2022). While all dot chromosomes were identified in both haploid assemblies, they retained more assembly gaps and higher SV frequencies than other chromosomes, suggesting that their genetic diversity may be underappreciated. The challenges associated with sequencing dot chromosomes are likely attributed to their repetitive nature and high GC-rich regions (Huang et al., 2023). To bridge the remaining gaps, additional Oxford Nanopore Technology (ONT) sequencing is needed, as GA-rich simple repeats are difficult to sequence using HiFi sequencing technology.

Unlike the metacentric or submetacentric macrochromosomes in chickens, formed due to fusions of ancestral acrocentric chordate chromosomes, most golden pheasant chromosomes exhibited acrocentric morphology, including many macrochromosomes that appear to have reverted to an acrocentric state, except for two large macrochromosomes retaining their original structure. Chromosome synteny analysis identified the centromeric breakpoints of inversions and fissions responsible for the formation of acrocentric chromosomes in the golden pheasant. These chromosomal rearrangements likely contributed to alterations in 3D chromatin architecture, as evidenced by the observed telomeric clustering (Bista et al., 2024) near the nuclear periphery (Figure 2C).

The golden pheasant genome revealed a striking depletion of inversions but a comparable number of copy number variations to mammalian genomes and an unexpectedly high prevalence of large indels (Supplementary Table S14). Avian genomes are typically characterized by smaller sizes, lower TE accumulation, and stronger evolutionary constraints (Zhang et al., 2014). The prevalence of extensive

heterozygous large indels driven by recent TE insertions suggests that avian genomes may lack strong selective pressure for purging TEs, indicating that TE insertion polymorphism may be more common than previously thought. While most TE insertions were located in the intergenic region (68.4%), a considerable proportion also occurred in the introns (26.6%), with a smaller proportion found in the exons (5.0%), suggesting that some TE insertions may have important functional implications. Comprehensive pangenome analyses are essential for characterizing the dynamics of TE insertion polymorphisms (He et al., 2024) and determining whether TEs have been co-opted or exapted for regulatory functions, a question requiring future expression and epigenetic studies (Capy, 2021).

Our study identified a surprisingly high number of duplicated genes, primarily immune-related, on dot chromosomes—a genomic feature often overlooked in previous research. Chicken chr16, one such dot chromosome, is well known for containing MHC genes (He et al., 2021; O'connor et al., 2019). Duplication of MHC and other immune-related genes is an important source of genetic diversity, enabling a rapid and robust response to a wide range of pathogens (Conrad & Antonarakis, 2007). The relatively frequent TE activity observed on dot chromosomes raises the intriguing possibility that TEs may have contributed to immune gene duplications, warranting further investigation (Gozashti et al., 2024). Our findings provide new opportunities to deepen our understanding of genetic variation in avian immune genes and their responses to avian pathogens.

The W chromosome remains one of the most complex and evolutionarily dynamic regions of bird genomes (Peona et al., 2020; Peona et al., 2021; Xu et al., 2019; Xu & Zhou, 2020). Due to its highly repetitive and heterochromatic nature, a complete sequence assembly has yet to be achieved, consistent with other degenerated W chromosomes in birds (Huang et al., 2022; Wang et al., 2021). Previous studies have revealed extensive rearrangements in the W chromosome between chickens and songbirds (Peona et al., 2021), which diverged over 66 Ma. Remarkably, our findings demonstrated

that this pattern of rapid chromosomal reorganization persisted even between closely related species, suggesting that the W chromosome in birds evolved rapidly in terms of chromosomal organization, although its sequence evolution appears to have progressed more slowly compared to autosomes (Zhou et al., 2014). In the golden pheasant, we found that gene content and expression patterns of the W chromosome were almost identical to those of the chicken, despite extensive rearrangements in gene order and significant species divergence. Thus, our analysis suggests that the gene repertoire of the Galliformes W chromosome was likely shaped by ancient W chromosome decay, with minimal gene losses or gains following species divergence, and that gene synteny is not required to maintain the conserved gene expression patterns.

CONCLUSIONS

Using a trio-binning approach, we successfully assembled a near-complete diploid genome for the golden pheasant, achieving unprecedented resolution in genome completeness and quality. The high-quality assemblies enabled comprehensive analyses of SVs, segmental duplications, gene amplifications, and sex chromosomes. Our findings highlighted the critical role of TEs in driving heterozygous SVs in bird genomes, corroborating their functional significance in shaping avian genome architecture (Kapusta & Suh, 2017). Chromosomal rearrangements, including inversions and fissions, were shown to influence chromosome morphology and 3D chromosome configuration. However, these structural changes appear to have limited regulatory impact on the expression of W-linked genes. This study highlights the utility of trio-binning and similar approaches for exploring genome diversity and evolution in non-model organisms.

DATA AVAILABILITY

All sequencing data were deposited in the NCBI database under BioProjectID PRJNA857339 and the Genome Sequence Archive (Chen et al., 2021) at the National Genomics Data Center (CNCB-NGDC Members and Partners, 2022), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA021136). A full list of accession IDs is available in Supplementary Table S1. The genome assembly data were deposited at the Science Data Bank (<https://doi.org/10.57760/sciencedb.18296>) and NCBI under BioProject ID PRJNA1048143 and PRJNA1048144.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

L.H.X. and X.B.O. conceived and designed the project. B.P.L. and Y.P.H. collected experimental samples. N.K., Z.X.X., H.R.L., S.Y.F., X.H.A., and X.L. performed data analysis. L.H.X. and N.K. wrote the manuscript. B.P.L., X.B.O., and Z.X.X. participated in revising the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

Armstrong J, Hickey G, Diekhans M, et al. 2020. Progressive cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, **587**(7833): 246–251.

Barros CP, Derks MFL, Mohr J, et al. 2023. A new haplotype-resolved

turkey genome to enable turkey genetics and genomics research. *GigaScience*, **12**: giad051.

Benham PM, Cicero C, Escalona M, et al. 2024. Remarkably high repeat content in the genomes of sparrows: the importance of genome assembly completeness for transposable element discovery. *Genome Biology and Evolution*, **16**(4): evae067.

Bian C, Li R H, Ruan Z Q, et al. 2024. Chromosome-level genome assembly of the glass catfish (*Kryptopterus vitreolus*) reveals molecular clues to its transparent phenotype. *Zoological Research*, **45**(5): 1027–1036.

Bista B, González-Rodelas L, Álvarez-González L, et al. 2024. De novo genome assemblies of two cryptodiran turtles with ZZ/ZW and XX/XY sex chromosomes provide insights into patterns of genome reshuffling and uncover novel 3D genome folding in amniotes. *Genome Research*, **34**(10): 1553–1569.

Bravo GA, Schmitt CJ, Edwards SV. 2021. What have we learned from the first 500 avian genomes? *Annual Review of Ecology, Evolution, and Systematics*, **52**: 611–639.

Campbell MS, Holt C, Moore B, et al. 2014. Genome annotation and curation using MAKER and MAKER - P. *Current Protocols in Bioinformatics*, **48**: 4–11.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15): 1972–1973.

Capy P. 2021. *Taming, domestication and exaptation*: trajectories of transposable elements in genomes. *Cells*, **10**(12): 3590.

Chen TT, Chen X, Zhang SS, et al. 2021. The genome sequence archive family: toward explosive data growth and diverse data types. *Genomics, Proteomics & Bioinformatics*, **19**(4): 578–583.

Cheng HY, Asri M, Lucas J, et al. 2024. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nature Methods*, **21**(6): 967–970.

Cheng HY, Concepcion GT, Feng XW, et al. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**(2): 170–175.

Clarke JA, Tambussi CP, Noriega JI, et al. 2005. Definitive fossil evidence for the extant avian radiation in the cretaceous. *Nature*, **433**(7023): 305–308.

CNCB-NGDC Members and Partners. 2022. Database resources of the national genomics data center, china national center for bioinformation in 2022. *Nucleic Acids Research*, **50**(D1): D27–D38.

Conrad B, Antonarakis SE. 2007. Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annual Review of Genomics and Human Genetics*, **8**: 17–35.

Di Tommaso P, Chatzou M, Floden EW, et al. 2017. Nextflow enables reproducible computational workflows. *Nature Biotechnology*, **35**(4): 316–319.

Durand NC, Robinson JT, Shamim MS, et al. 2016a. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, **3**(1): 99–101.

Durand NC, Shamim MS, Machol I, et al. 2016b. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, **3**(1): 95–98.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5): 1792–1797.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**(1): 238.

Fang B, Edwards SV. 2024. Fitness consequences of structural variation inferred from a House Finch pangenome. *Proceedings of the National Academy of Sciences of the United States of America*, **121**(47): e2409943121.

Feng SH, Stiller J, Deng Y, et al. 2020. Dense sampling of bird diversity

- increases power of comparative genomics. *Nature*, **587**(7833): 252–257.
- Flynn JM, Hubley R, Goubert C, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(17): 9451–9457.
- Gao GQ, Xu M, Bai CL, et al. 2018. Comparative genomics and transcriptomics of *Chrysolophus* provide insights into the evolution of complex plumage coloration. *GigaScience*, **7**(10): giy113.
- Garg KM, Dovih P, Chattopadhyay B. 2024. Hybrid *de novo* genome assembly of the sexually dimorphic Lady Amherst's pheasant. *DNA Research*, **31**(1): dsae001.
- Goel M, Sun HQ, Jiao WB, et al. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, **20**(1): 277.
- Gozashti L, Hartl DL, Corbett-Detig R. 2024. Universal signatures of transposable element compartmentalization across eukaryotic genes. *Nature*, **620**(7973): 123–130.
- Grabherr MG, Haas BJ, Yassour M, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**(7): 644–652.
- Haas BJ, Salzberg SL, Zhu W, et al. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biology*, **9**(1): R7.
- He K, Minias P, Dunn PO, et al. 2021. Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biology and Evolution*, **13**(2): evaa270.
- He X, Qi ZY, Liu ZP, et al. 2024. Pangenome analysis reveals transposon-driven genome evolution in cotton. *BMC Biology*, **22**(1): 92.
- Huang Z, De O, Furo I, Liu J, et al. 2022. Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nature Communications*, **13**(1): 944.
- Huang Z, Xu ZX, Bai H, et al. 2023. Evolutionary analysis of a complete chicken genome. *Proceedings of the National Academy of Sciences of the United States of America*, **120**(8): e2216641120.
- Jae Lee S, Kim JH, Jo E, et al. 2021. Chromosomal assembly of the Antarctic toothfish (*Dissostichus mawsoni*) genome using third-generation DNA sequencing and Hi-C technology. *Zoological Research*, **42**(1): 124–129.
- Jin W, Cao XJ, Ma XY, et al. 2022. Chromosome-level genome assembly of the freshwater snail *Bellamya purificata* (Caenogastropoda). *Zoological Research*, **43**(4): 683–686.
- Kapusta A, Suh A. 2017. Evolution of bird genomes—a transposon's - eye view. *Annals of the New York Academy of Sciences*, **1389**(1): 164–185.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**(4): 357–360.
- Ko BJ, Lee C, Kim J, et al. 2022. Widespread false gene gains caused by duplication errors in genome assemblies. *Genome Biology*, **23**(1): 205.
- Koren S, Rhie A, Walenz BP, et al. 2018. *De novo* assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology*, **36**(12): 1174–1182.
- Kulski JK, Shiina T, Dijkstra JM. 2019. Genomic diversity of the major histocompatibility complex in health and disease. *Cells*, **8**(10): 1270.
- Li H. 2021. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, **37**(23): 4572–4574.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**(14): 1754–1760.
- Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078–2079.
- Li M, Sun CJ, Xu NY, et al. 2022. *De novo* assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on microchromosomes and subtelomeric regions. *Molecular Biology and Evolution*, **39**(4): msac066.
- Li XJ, Huang Y, Lei FM. 2015. Comparative mitochondrial genomics and phylogenetic relationships of the *Crossoptilon* species (Phasianidae, Galliformes). *BMC Genomics*, **16**(1): 42.
- Liu BJ, Zhang K, Zhang SF, et al. 2022. Chromosome-level genome assembly of the dotted gizzard shad (*Konosirus punctatus*) provides insights into its adaptive evolution. *Zoological Research*, **43**(2): 217–220.
- Liu J, Wang ZJ, Li J, et al. 2021. A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. *Genome Research*, **31**(3): 497–511.
- Liu SL, Hansen MM. 2017. PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. *Molecular Ecology Resources*, **17**(4): 631–641.
- Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature Methods*, **12**(8): 733–735.
- Luo HR, Jiang XR, Li BP, et al. 2023. A high-quality genome assembly highlights the evolutionary history of the great bustard (*Otis tarda*, Otidiformes). *Communications Biology*, **6**(1): 746.
- Nattestad M, Schatz MC. 2016. Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics*, **32**(19): 3021–3023.
- Nguyen LT, Schmidt HA, Von Haeseler A, et al. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**(1): 268–274.
- O'Connor EA, Westerdahl H, Burri R, et al. 2019. Avian MHC evolution in the era of genomics: phase 1.0. *Cells*, **8**(10): 1152.
- O'Connor RE, Kretschmer R, Romanov MN, et al. 2024. A bird's-eye view of chromosomal evolution in the class aves. *Cells*, **13**(4): 310.
- Peona V, Blom MPK, Xu LH, et al. 2020. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird - of - paradise. *Molecular Ecology Resources*, **21**(1): 263–286.
- Peona V, Palacios-Gimenez OM, Blommaert J, et al. 2021. The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **376**(1833): 20200186.
- Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. *F1000Research*, **9**: 304.
- Pu LR, Lin Y, Pevzner PA. 2018. Detection and analysis of ancient segmental duplications in mammalian genomes. *Genome Research*, **28**(6): 901–909.
- Robinson JT, Thorvaldsdóttir H, Wenger AM, et al. 2017. Variant review with the integrative genomics viewer. *Cancer Research*, **77**(21): e31–e34.
- Seki R, Li C, Fang Q, et al. 2017. Functional roles of Aves class-specific *cis*-regulatory elements on macroevolution of bird-specific features. *Nature Communications*, **8**(1): 14229.
- Sirén J, Välimäki N, Mäkinen V. 2014. Indexing graphs for path queries with applications in genome research. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **11**(2): 375–388.
- Stanke M, Diekhans M, Baertsch R, et al. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*, **24**(5): 637–644.
- Stein RW, Brown JW, Mooers AØ. 2015. A molecular genetic time scale demonstrates cretaceous origins and multiple diversification rate shifts within the order Galliformes (Aves). *Molecular Phylogenetics and Evolution*, **92**: 155–164.
- Stiller J, Feng SH, Chowdhury AA, et al. 2024. Complexity of avian evolution revealed by family-level genomes. *Nature*, **629**(8013): 851–860.
- Sun XP, Jiao C, Schwaninger H, et al. 2020. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nature Genetics*, **52**(12): 1423–1432.

- Tang HB, Bowers JE, Wang XY, et al. 2008. Synteny and collinearity in plant genomes. *Science*, **320**(5875): 486–488.
- Wang SN, Shen Y, Lin ZC, et al. 2024. New genes driven by segmental duplications share a testis - specific expression pattern in the chromosome - level genome assembly of tree sparrow. *Integrative Zoology*, **19**(5): 1004–1008.
- Wang ZJ, Chen GJ, Zhang GJ, et al. 2021. Dynamic evolution of transposable elements, demographic history, and gene content of paleognathous birds. *Zoological Research*, **42**(1): 51–61.
- Wu XL, Mu DP, Yang QS, et al. 2023. Comparative genomics of widespread and narrow-range white-bellied rats in the *Niviventer niviventer* species complex sheds light on invasive rodent success. *Zoological Research*, **44**(6): 1052–1063.
- Xu LH, Auer G, Peona V, et al. 2019. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nature Ecology & Evolution*, **3**(5): 834–844.
- Xu LH, Zhou Q. 2020. The female-specific W chromosomes of birds have conserved gene contents but are not feminized. *Genes*, **11**(10): 1126.
- Xu LL, Ren YD, Wu JH, et al. 2024a. Evolution and expression patterns of the neo-sex chromosomes of the crested ibis. *Nature Communications*, **15**(1): 1670.
- Xu X, Wang C, Xu CZ, et al. 2024b. Genomic evolution of island birds from the view of the Swinhoe's pheasant (*Lophura swinhoii*). *Molecular Ecology Resources*, **24**(2): e13896.
- Yang CT, Zhou Y, Marcus S, et al. 2021. Evolutionary and biomedical insights from a marmoset diploid genome assembly. *Nature*, **594**(7862): 227–233.
- Yusuf L, Heatley MC, Palmer JPG, et al. 2020. Noncoding regions underpin avian bill shape diversification at macroevolutionary scales. *Genome Research*, **30**(4): 553–565.
- Zhang GJ, Li C, Li QY, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, **346**(6215): 1311–1320.
- Zhang YJ, Chu J, Cheng HY, et al. 2023. De novo reconstruction of satellite repeat units from sequence data. *Genome Research*, **33**(11): 1994–2001.
- Zhao N, Guo HB, Jia L, et al. 2021. High-quality chromosome-level genome assembly of redlip mullet (*Planiliza haematocheila*). *Zoological Research*, **42**(6): 796–799.
- Zhou H, Huang XH, Liu JJ, et al. 2024. De novo phased genome assembly, annotation and population genotyping of alectoris chukar. *Scientific Data*, **11**(1): 162.
- Zhou Q, Zhang JL, Bachtrog D, et al. 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*, **346**(6215): 1246338.
- Zhu F, Yin ZT, Zhao QS, et al. 2023. A chromosome-level genome assembly for the Silkie chicken resolves complete sequences for key chicken metabolic, reproductive, and immunity genes. *Communications Biology*, **6**(1): 1233.