105 (2026) 106973

Contents lists available at ScienceDirect

Poultry Science

journal homepage: www.elsevier.com/locate/psj

Full-Length Article

De novo chromosome-level genome assembly and annotation of the Muscovy duck (*Cairina moschata*)

Xing Ju [a,b,c,1], Zhijun Wang [e,1], Haiping Xu [f], Liangtian Tan [b,c], Danfeng Cai [b,c], Weijian Zhu [d], Lijin Guo [b,c], Congliang Ji [d], Qinghua Nie [b,c], Wei Han [a,*]

[a] *Poultry Research Institute, Chinese Academy of Agricultural Science/Jiangsu Institute of Poultry Science, Yangzhou 225125, China*
[b] *State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Lingnan Guangdong Laboratory of Agriculture, College of Animal Science, South China Agricultural University, Guangzhou 510642, Guangdong, China*
[c] *Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding and Key Lab of Chicken Genetics, Breeding and Reproduction, Ministry of Agriculture, Guangzhou 510642, Guangdong, China*
[d] *Wens Foodstuff Group Co. Ltd., Yunfu 527400 Guangdong, China*
[e] *College of Animal Science and Technology, Zhejiang Agriculture and Forestry University, Hangzhou 311300, Zhejiang, China*
[f] *Guangzhou College of Technology and Business 510580, Guangzhou, Guangdong, China*

ARTICLE INFO

Keywords:
Cairina moschata
Genome assembly
Hi-C
PacBio
RNA-seq

ABSTRACT

The Muscovy duck (*Cairina moschata*) is an excellent lean-type poultry species characterized by large body weight, rapid growth, high lean meat yield, good liver production performance, and high feed conversion efficiency. It plays an important role in modern poultry industry. However, the quality of the currently available Muscovy duck genome remains limited. Here, we report the de novo sequencing and assembly of the Muscovy duck genome using PacBio and Hi-C technologies. The total length of the genome assembly was 1270.04 Mb, the contig N50 was 6.56 Mb, and the scaffold N50 was 77.04 Mb and it was composed of 39 chromosomes. BUSCO evaluation showed that 95.2% of conserved genes were completely assembled in the Muscovy duck genome, indicating high completeness of the assembly results. The genome annotation results showed that the proportion of repetitive sequences was 22.14%, 16,901 coding genes were predicted, and 93.4% of the genes were functionally annotated. Comparative genomic analysis revealed that expanded gene families were associated with immune-related pathways in *Cairina moschata*. *Cairina moschata* was closely related to *Anas platyrhynchos* and *Aythya fuligula*, both of which belong to the family Anatidae. These results provide valuable resources for poultry genomic breeding programs.

Introduction

Muscovy ducks (*Cairina moschata*) are native to tropical regions of Central and South America. As an important duck br./eed worldwide, Muscovy duck has the advantages of fast growth, ease of management, strong disease resistance, high meat yield, unique meat quality, good performance of fat liver production, and high feed efficiency, which is favored by the majority of farmers and consumers, and has a good market prospect (Bello et al., 2021; Xu et al., 2022b; Ye et al., 2019). Muscovy duck has a higher weight gain rate and feed conversion rate, and it is resistant to rough feeding, easy to raise, more resistant to disease and environmental hazards, and has better economic performance (Bello et al., 2022; Ju et al., 2023; Ye et al., 2017).

PacBio SMRT technology is based on the principle of while sequencing-by-synthesis and uses SMRT cells as sequencing carriers (Deamer et al., 2016; Korlach et al., 2010; Xiao et al., 2019; Zhang et al., 2019b). Fluorescent labeled dNTP are excited by a laser, and bases are identified according to their distinct fluorescence signals. Circular Consensus Sequencing Mode (CCS) generates highly accurate consensus sequences from multiple subreads of the same DNA molecule, with accuracy exceeding 99% (Q20) (Ning et al., 2018; Schmidt et al., 2017; Tan et al., 2018). High-throughput chromosome conformation capture (Hi-C) is a technique used to investigate genome-wide chromatin interactions (Berkum et al., 2010; Choi et al., 2020; Suryamohan et al., 2020; Zhang et al., 2019a). By combining 3C technology and high-throughput sequencing technology, Hi-C enables the analysis of

* Corresponding author.
E-mail address: hanwei830@163.com (W. Han).
[1] These authors equally contributed to this work.

https://doi.org/10.1016/j.psj.2026.106973
Received 2 February 2026; Accepted 16 April 2026
Available online 16 April 2026

spatial chromatin interactions and the three-dimensional organization of the genome, thereby providing genome-wide chromatin interaction information (Austin et al., 2017; Berkum et al., 2010; Jain et al., 2018; Michael et al., 2018; Mondal et al., 2017).

In this study, we performed a de novo assembly of a chromosome-level genome of the Muscovy duck. We aimed to sequence, assemble, and annotate the Muscovy duck genome. This genome assembly will provide a valuable resource for biological research and breeding improvement of this species .

Materials and methods

Ethics Statement

Approval for animal use was obtained from the Animal Protection and Utilization Committee of South China Agricultural University (Guangzhou, People's Republic of China) (Approval number: 2023f271). All procedures were conducted in accordance with the Regulations for the Administration of Affairs Concerning Experimental Animals of The State Council of China. All animals were handled in accordance with institutional guidelines to minimize stress and suffering.

Sample collection and short-read Illumina sequencing

A Blood sample was collected from a 29-week-old female Muscovy duck (from Renma Farm of Guangdong Wenshi Nanfang Poultry Breeding Co., LTD) for whole genome sequencing. PacBio libraries were constructed and sequenced using the PacBio Sequel II/IIe platform (https://github.com/PacificBiosciences/ccs). A total of 146 GB of clean data were generated and used for genome assembly of *Cairina moschata*, and the raw data has been uploaded to BIG Submission (https://ngdc.cncb.ac.cn/gsub/) (GSA number: CRA013446, Project: PRJCA021188). RNA was isolated from different organs (pituitary, liver, hypothalamus, oviduct, pectoral muscle, ovary, spleen, cerebrum, cerebellum, and leg muscle), and RNA-seq libraries were constructed from the same individual. After quality control, clean reads were used for transcriptome assembly and subsequent genome annotation.

Genome size estimation

Raw sequencing date were processed for quality control, including filtering, decontamination, and error correction. The statistical results include sequencing read number, data yield, sequencing error rate, Q20, Q30, GC content, etc. To obtain a preliminary estimate of genome size, K-mer analysis was used to estimate genome size (Marcais and Kingsford, 2011; Vurture et al., 2017).

PacBio sequencing and Hifiasm assembly

For the de novo assembly of Muscovy duck, PacBio SMRT sequencing was performed. A key to PacBio's SMRT technology is how to distinguish the reaction signal from the strong fluorescent background of the surrounding free bases. Hifiasm (0.16.1) is a fast haplotype-resolved de novo assembly tool designed for PacBio Hifi reads. The assembly mode of Hifiasm refers to the specific assembly strategy and parameter configuration selected based on the sequencing data type, the availability of auxiliary data, and the complexity of the target genome, in order to achieve accurate genome reconstruction and haplotype resolution. Unlike most existing assembly software, Hifiasm starts with an uncollapsed genome and is therefore able to retain haplotype information as much as possible. Hifiasm supports both primary assembly and trio-binning assembly. Genome completeness was assessed using BUSCO (http://busco.ezlab.org/) with Metaeuk and Hmmer pipelines.

Hi-C assembly of the chromosome level genome

Hi-C technology mainly cross-links DNA fragments near the spatial structure, and enriches the cross-linked DNA fragments, and then carries out high-throughput sequencing, and analyzes the sequencing data to reveal the interactions between chromosome fragments in the whole genome. The general structural features of the genome can be revealed by Hi-C technology (Rao et al., 2014; Wingett et al., 2015). By pruning the cross-linking signals between homologous chromosomes, the alleles and homologous sequences were separated into their respective haplotypes and assembled independently. According to the Hi-C data, the interaction signals between contigs were calculated, and the signal strength relationship was used to group. Through genetic algorithm random optimization, contigs in each group were ordered and oriented, and finally chromosome-level assembly was obtained and heat maps were drawn.

Repeat sequence annotation

Repeat sequences can be divided into Interspersed repeats and Tandem repeats. Homologous sequence alignment method of Muscovy duck genome was based on repeat sequence database (commonly used RepBase database, http://www.girinst.org/repbase/), using Repeatproteinmask (http://www.repeatmasker.org/). Known repetitive elements were identified using RepeatMasker and RepeatProteinMask based on the Repbase database. RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html), LTR_FINDER (http://tlife.fudan.edu.cn/ltr_finder/) (Xu and Wang, 2007) and RepeatScout (http://www.repeatmasker.org/) and were used to build the de novo repeat sequence libraries.

Gene structure annotation

Gene models were constructed by integrating homology-based prediction, RNA-seq evidence, and de novo prediction (Haas et al., 2008; Stanke et al., 2006b). Homologous prediction refers to the inference of homologous relationships between different species or different gene sequences in the same species by comparing them. By using Genewise (http://www.ebi.ac.uk/~birney/wise2/) (Birney et al., 2004), Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi), and other than software to predict gene structure. De novo prediction is an important method that can help better understand the structure of genes. Commonly used prediction software include SNAP (http://homepage.mac.com/iankorf/), GlimmerHMM (http://ccb.jhu.edu/software/glimmerhmm/) (Majoros et al., 2004), Augustus (http://bioinf.uni-greifswald.de/augustus/) (Stanke et al., 2006a), and etc. Other evidence supporting the prediction was to use the homologous species of EST or cDNA data, gene structure was predicted by Blat comparison (http://genome.ucsc.edu/cgi-bin/hgBlat).

Gene function annotation

Gene functions was annotated by comparing predicted protein sequences against public databases. Such as KEGG (http://www.genome.jp/kegg/), SwissProt database (http://www.uniprot.org/), Nr (http://www.ncbi.nlm.nih.gov/protein), Pfam (http://pfam.xfam.org/), and InterPro (https://www.ebi.ac.uk/interpro/),.

Noncoding RNA annotation

Non-coding RNAs, including tRNA, rRNA, miRNA, and snRNA, were annotated. Here, we used tRNAscan-SE software to find tRNA in the genome. Search for rRNA in the genome through Blast comparison. Using Rfam family covariance model, adopting Rfam own INFERNAL (http://infernal.janelia.org/) software of microRNAs in the genomes of predictable and snRNA sequence information (Griffiths-Jones et al.,

2005).

Phylogenetic and collinearity analyses

In the study of biological evolution and phylogeny, phylogenetic trees are commonly used to infer evolutionary relationships among various organisms (Abascal et al., 2005; Benton and Donoghue, 2007; Posada, 2008). I In this study, we determined the evolutionary tree by comparing *Cairina moschata* with 11 other species *(Rissa tridactyla, Anas platyrhynchos, Anser cygnoides, Gallus gallus, Cygnus olor, Cygnus atratus, Eurypyga helias, Aquilachrysaetos chrysaetos, Oxyura jamaicensis, Aythya fuligula* and *Meleagris gallopavo)*. Through all the single copy gene families for MUSCLE, (http://www.drive5.com/muscle/) (Edgar, 2004), then combined all the comparison results together, form a super alignment matrix (Li et al., 2003). Finally, we used RAxML 7.2.9 software to build phylogenetic TREE based on the sequence comparison results using maximum likelihood method (ML TREE) (Guindon and Gascuel, 2003; Guindon et al., 2010; Yang, 2007; Zhang et al., 2005).

Genomes of *Anas platyrhynchos* (GCA_008746955.1) was selected for comparison with the genome of *Cairina Moschata* using BLASTP. Collinear blocks were identified using McscanX to generate circle plots.

Results

Genome size estimation

Genomic DNA from a 29-week-old *Cairina moschata* individual (Fig. 1), which was provided by Renma Farm of Guangdong Wenshi Nanfang Poultry Breeding Co., LTD, was used for genome sequencing (Fig. 2a).

According to the results of survey analysis, the major peak was observed at a sequencing depth of ~57 in Fig. 2d. The estimated genome size was approximately 1,315.47 Mbp, corrected genome size was 1,289.98 Mbp. And heterozygosity rate was 0.78%, proportion of repeated sequences was 40.67%.

Chromosomal-level genome assembly

Using the obtained Hi-C data, contig and scaffold sequences were anchored to the near-chromosome level using Allhic software. After that, sequences were manually corrected using JuiceBox to obtain the genome at the chromosome level (Fig. 2c). The genome was sequenced using the PacBio platform, and de novo assembly was performed on the duck genome using the 146.00 G sequencing data. Then, we used Hi-C technology to assemble *Cairina moschata* genome and made the genome reach chromosome level. The obtained chromosome and genome results were as follows: the total length of contig was 1.27 Gbp, and the length of contig N50 was 6.56 Mbp; The total length of the scaffold was 1.27 Gbp and the length of the scaffold N50 was 77.04 Mbp (Table 1). The genome assembly had 39 chromosomes (Supplementary Table S1), which is more than the number of chromosomes in the *Cairina moschata* (GCA_009194515.1), and the genome mounting rate was 90.68% (Supplementary Table S2). The average GC content was 41.93%, and the gene density of the entire genome was about 3 genes per 100 kb (Fig. 2b).

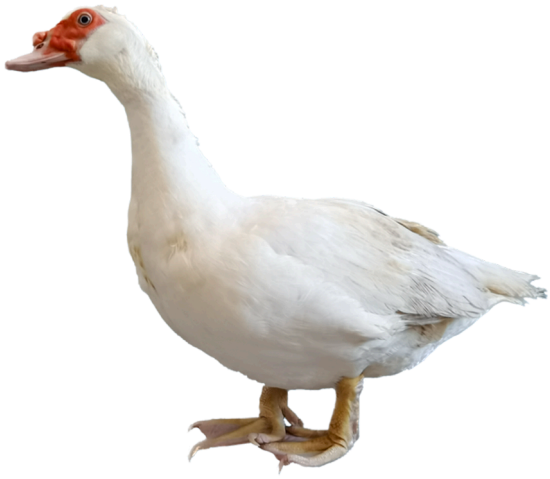

Fig. 1. Photograph of a 29-week-old Cairina Moschata.

BUSCO analysis showed that 95.2% of BUSCO genes were assembled in Muscovy duck genome, indicating that the assembly was of high completeness (Table 2).

Repetitive sequence annotation

Repeatmasker and Repeatproteinmask were used to identify the sequence that were similar to the known repeat sequence, identifying 281.24 Mbp of repetitive sequences, accounting for 22.14% of the whole genome sequence (Table 3). Then, the repetitive sequences comprised DNA transposons in 1.38 Mbp (0.11% of the assembly), long interspersed nuclear elements in 137.07 Mbp (LINEs: 10.79%), long terminal repeats in 168.80 Mbp (LTRs: 13.29%), short interspersed nuclear elements in 0.10 Mbp (SINEs: 0.01%), and unknown repeats in 20.19 Mbp (1.59%) (Table 4).

Gene structure and function annotation

B Based on the results of homology prediction, transcriptome-assisted, and de novo sequencing, 16,901 genes were predicted (Table 5). We compared the statistics of the predicted gene models with *Anas platyrhynchos* and *Cairina moschata* (from NCBI), which showed similar distribution patterns in different genetic elements (Fig. 3). In comparison, the genome we assembled contains a greater number of annotated genes and a longer average transcript length. This indicates that our assembly is more comprehensive and complete, and therefore more reliable as a reference genome.

Protein sequences predicted by gene structure were compared to known protein libraries (Kent, 2002). We also mapped gene sets to KEGG pathways. As a result, 85.90%, 88.80% and 80.10% of the predicted genes received positive hits in the Swissprot, Nr and KEGG databases, respectively. Then, a total of 15783 genes (93.40% of all predicted genes) were successfully annotated (Supplementary Table S3).

Non-coding RNA annotation

For non-coding genes, 330 miRNAs and 386 snRNA were identified by using Rfam own INFERNAL (http://infernal.janelia.org/). In total, 1,110 rRNA were identified by blast. Finally, 458 tRNA were identified (Supplementary Table S4).

Gene family clustering and phylogenetic analysis

14,733 gene families with an average of 1.23 genes per family were identified in the *Cairina moschata* genome (Supplementary Table S5). The homologous genes among the above-mentioned genomes were classified by OrthoMCL (http://orthomcl.org/orthomcl/) (Li et al., 2003), and 6,899 single-copy genes were identified in the *Cairina moschata* genome (Supplementary Fig. S1). Based on phylogenetic analysis of single-copy gene families, *Cairina moschata* was closely related to *Anas platyrhynchos* and *Aythya fuligula*, both of which belong to the family Anatidae. Divergence times between *Cairina moschata* and other species were estimated, and we found that *Cairina moschata* diverged approximately 38.8 million years ago from the common ancestor with the lineage of *Aythya fuligula* and *Anas platyrhynchos* (Fig. 4a).

The genome collinearity analysis in *Cairina moschata* and *Anas*

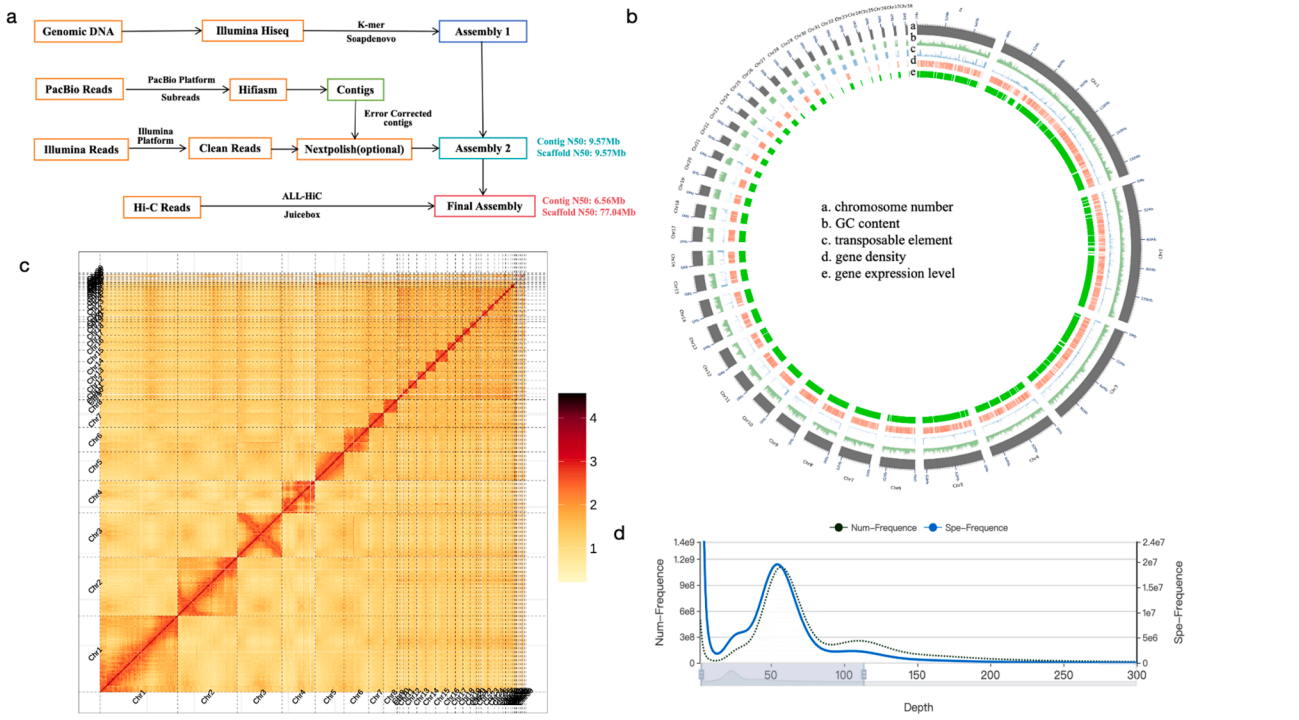

Fig. 2. Genome size estimation and assembly. (a) Pipeline used for assembly of the Muscovy duck genome. (b) The characteristics of morphology and genome of Cairina moschata. (c) Hi-C contact map of Cairina moschata. (d) Graph of k-mer frequency distribution (k = 17) generated from 127Gbp of sequencing data of Cairina moschata.

Table 1
Statistics of the Cairina moschata genome assembly.

Seq type	Contig length	Scaffold length	Contig number	Scaffold number
Total	1,270,037,933	1,270,076,933	723	333
Max	36,662,637	206,624,101	-	-
Number≥2000	-	-	723	333
N50	6,555,063	77,040,014	44	5
N60	4,462,428	36,711,912	66	8
N70	2,906,860	22,387,498	101	12
N80	1,669,589	15,170,058	158	19
N90	871,516	4,548,245	261	35

Table 2
BUSCO analysis results of the Cairina moschata genome.

Percentage	
Complete BUSCOs (C)	95.2%
Complete and single-copy BUSCOs (S)	94.3%
Complete and duplicated BUSCOs (D)	0.9%
Fragmented BUSCOs (F)	1.8%
Missing BUSCOs (M)	3.0%
Total BUSCO group searched	100

platyrhynchos was shown in Fig. 4b. A total of 296 synteny blocks and 25696 genes in all blocks were detected in *Cairina moschata* and *Anas platyrhynchos* versus (Supplementary Table S6). Thus, these data suggested that the genetic order of these two species was highly conserved.

Discussion

With economic development and the improvement of people's living standards, waterfowl meat, characterized by high protein and low fat content, is increasingly favored by consumers, thereby promoting the rapid development of waterfowl breeding industry in our country

Table 3
Statistics of repetitive sequences in the Cairina moschata genome.

Type	Repeat Size(bp)	% of genome
Trf	56,652,251	4.46
Repeatmasker	266,720,668	21.00
Proteinmask	67,217,426	5.29
Total	281,239,572	22.14

Note: TRF: Results of tandem repeats predicted using TRF; Repeatmasker: Through LTR_FINDER, RepeatScout and RepeatModeler, the De novo repeat sequence library was built, and then predicted by Repeatmasker software; Proteinmask: Compare the genome sequence with the transposon library that comes with RepeatProteinMasker to predict the repeat sequence; Total: The results obtained by the above various methods are non-redundant results after removing the overlapping parts between them.

(Fournier et al., 1997; Xu et al., 2022a; Yang et al., 2017). Muscovy duck is one of the main duck breeding breeds in China after Beijing duck and Cherry Valley duck. Waterfowl breeding in China has obvious regional characteristics. At the same time, waterfowl breeding and breeding have broad development prospects and progress potential. In this study, we generated a high-quality chromosome-level genome assembly for Muscovy duck using a combination of Illumina, PacBio, and Hi-C sequencing technologies. By using Hi-C, total length of the genome assembly was 1270.04 Mbp, the contig N50 was 6.56 Mbp, the scaffold N50 was 77.04 Mbp and it was composed of 39 chromosomes. The BUSCO analysis results were superior to those of the previous assembly (95.2%) (Xu et al., 2022b), indicating that the Muscovy duck genome assembled in this study exhibits higher continuity and completeness. Similarly, compared with previous studies, we annotated a greater number of protein-coding genes (16,901). In addition, the numbers of non-coding RNA genes identified—including miRNAs (330), snRNAs (386), rRNAs (1,110), and tRNAs (458)—were all higher than those reported in previous studies (Xu et al., 2022b), suggesting that our annotation is more comprehensive and informative.

Table 4
Repeat sequence classification result statistics.

	Denovo+Repbase		TE Proteins		Combined TEs	
	Length(bp)	%in Genome	Length(bp)	%in Genome	Length(bp)	%in Genome
DNA	1,344,249	0.11	90,579	0.01	1,378,833	0.11
LINE	112,286,119	8.84	58,006,120	4.57	137,067,967	10.79
SINE	102,563	0.01	0	0	102,563	0.01
LTR	167,043,880	13.15	9,145,650	0.72	168,799,401	13.29
Unknown	20,191,243	1.59	0	0	20,191,243	1.59
Total	266,720,668	21.00	5.29	5.29	270,342,080	21.29

Note: LINE (long interspersed nuclear elements): long interspersed repeat sequence, the length of the repeat sequence unit is more than 1000 bp; SINE (short interspersed nuclear elements): short interspersed repeat sequence, the length of the repeat sequence unit is less than 50 bp; LTR(long terminal repeats): Long terminal repeats exist on both sides; Unknown: indicates that the repeat sequence cannot be classified by RepeatMasker; Total: The results obtained by various categories, excluding the overlapping parts between them.

Table 5
Statistics of gene predictions in the Cairina moschata.

	Gene set	Number	Average gene length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
De novo	Augustus	16,762	19,320.12	1,514.22	8.76	172.77	2,293.22
	GlimmerHMM	143,236	6,484.01	536.87	3.17	169.44	2,742.48
	SNAP	54,263	32,840.92	716.33	6.1	117.41	6,297.81
	Geneid	28,588	26,538.54	1,282.4	6.91	185.48	4,270.53
	Genscan	37,626	22,390.98	1,408.83	8.31	169.45	2,868.74
Homolog	Apla	9,304	17,911.71	1,433.82	8.5	168.69	2,197.15
	Cmos	9,286	20,644.67	1,597.44	9.34	171.06	2,284.32
RNAseq	transcripts	35,960	50,294.63	5,130.57	13.28	386.28	3,677.27
	PASA	27,394	20,076.32	1,293.2	7.49	172.59	2,892.96
EVM		19,066	21,427.01	1,489.77	8.91	167.27	2,521.65
Pasa-update		18,899	24,082.42	1,532.94	9.1	168.4	2,782.87
Final set		16,901	26,153.89	1,637.78	9.84	166.43	2,773.16

Note: Denovo: Using Augustus, GlimmerHMM, SNAP, Geneid and Genscan software to predict gene structure; RNAseq: Annotating gene structure using transcriptome data; PASA was predicted by combining Trinty assembly results of transcriptome. The Transcripts were predicted by the transcriptome combined with the reference genome; EVM: The result of integrating the results of Denovo annotation, homologous annotation and RNA annotation; Pasa-update: result of correction according to PASA; Final set: The final result of functional gene annotation.

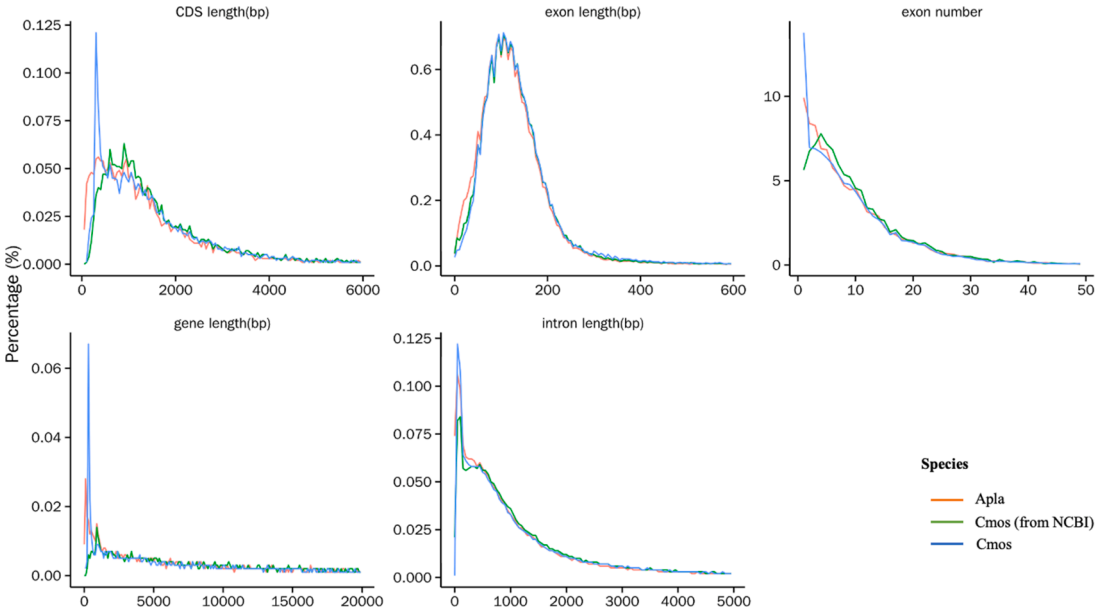

Fig. 3. Comparisons of the predicted gene models between the Anas platyrhynchos, Cairina moschata (from NCBI) and Cairina moschata.

Similarly, several studies have assembled the genomes of other duck breeds. Compared with previous studies, our contig N50 and scaffold N50 values were slightly lower than those reported for Liancheng ducks, and both the BUSCO scores and the number of annotated protein-coding genes were also lower (Wang et al., 2025). In contrast, for the Pekin duck, our contig N50 and scaffold N50 values were slightly higher than those reported previously, and the BUSCO scores were also improved (Yu et al., 2023). These differences may be attributed to breed-specific genomic characteristics as well as variations in sequencing technologies and assembly strategies.

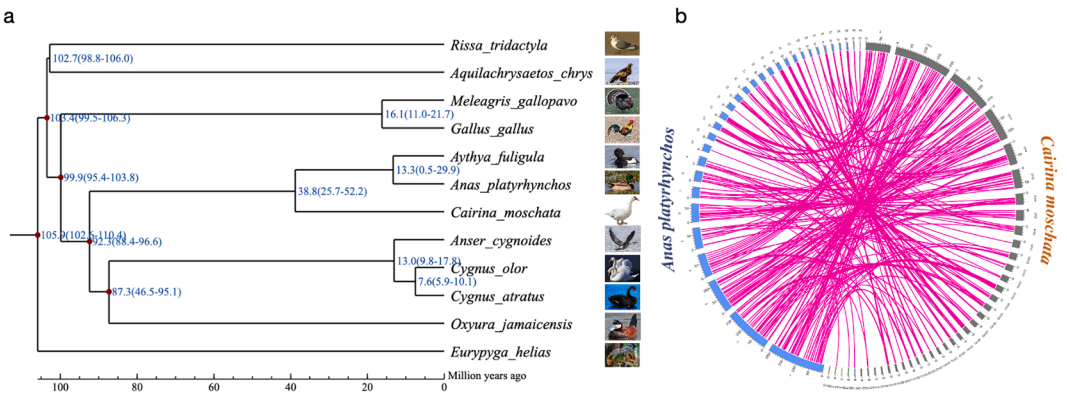

Fig. 4. Comparative genomic analyses of Cairina moschata with other grouper species. (a) The phylogenetic relationships of Cairina moschata with other species. The upper value represents the number of Bootstrap supports. Branch length represents the rate of evolution, and the longer the branch length, the faster the evolution. (b) Chromosomal collinearity between Cairina moschata and Anas platyrhynchos.

In addition, in *Cairina moschata*, lipid metabolism, liver development and immune regulation are coordinately controlled by several key genes. For lipid metabolism, *ACACA,SCD*, and *CPT1A* are involved in fatty acid synthesis, desaturation, and β-oxidation, respectively (Zhang et al., 2023). For liver development, *PCK1* is involved in the process of liver gluconeogenesis (Zhang et al., 2023). For immune response, *DDX58, TLR3*, and *Mx1* are involved in innate immune signaling and antiviral defense (Huang et al., 2013). Future studies will leverage the newly assembled Muscovy duck genome to further elucidate the functions and regulatory mechanisms of these key genes. In addition, comprehensive comparative genomic analyses will be performed, including synteny-based functional enrichment and the characterization of species-specific gene families, to provide deeper insights into the evolutionary features and molecular basis of trait differentiation.

Conclusion

Our genome assembly and annotation provides the genome of the Muscovy duck, these genomic variations and their functional annotation will maybe facilitate accurate genetic analysis and accelerate Muscovy duck genome breeding programs.

Financial support statement

This work was supported by Frontier Technologies R&D Program of Jiangsu (BF2025306) and Natural Science Foundation of Jiangsu Province (BK20241980).

CRediT authorship contribution statement

Xing Ju: Writing – original draft, Formal analysis, Conceptualization. **Zhijun Wang:** Writing – review & editing, Methodology. **Haiping Xu:** Data curation. **Liangtian Tan:** Resources, Conceptualization. **Danfeng Cai:** Methodology. **Weijian Zhu:** Investigation. **Lijin Guo:** Investigation. **Congliang Ji:** Resources. **Qinghua Nie:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Wei Han:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Disclosures

There was no conflict of interest during submission of this manuscript for publication. We hope for positive responses for our submission.

Acknowledgements

We are grateful to all members who gave us constructive advice of this work, and thank Novogene Co., Ltd. (Beijing, China) for providing us the sequencing platform.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psj.2026.106973.

Supplemental Table 1. Size of the chromosomes in the *Cairina moschata* genome.

Supplemental Table 2. Genomic mounting rate of *Cairina moschata.*

Supplemental Table 3. Gene function annotation statistical results of *Cairina moschata* genome.

Supplemental Table 4. Statistical results of non-coding RNA of *Cairina moschata* genome.

Supplemental Table 5. Summary of gene family clustering between *Cairina moschata* and other species.

Supplemental Table 6. Collinearity comparison analysis of *Cairina moschata* and *Anas platyrhynchos.*

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Austin, C.M., Tan, M.H., Harrisson, K.A., Lee, Y.P., Croft, L.J., Sunnucks, P., Pavlova, A., Gan, H.M., 2017. De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (Maccullochella peelii), from Illumina and Nanopore sequencing read. Gigascience 6, 1–6.

Bello, S.F., Xu, H., Guo, L., Li, K., Zheng, M., Xu, Y., Zhang, S., Bekele, E.J., Bahareldin, A. A., Zhu, W., Zhang, D., Zhang, X., Ji, C., Nie, Q., 2021. Hypothalamic and ovarian transcriptome profiling reveals potential candidate genes in low and high egg production of white Muscovy ducks (Cairina moschata). Poult. Sci. 100, 101310.

Bello, S.F., Xu, H., Li, K., Guo, L., Zhang, S., Ahmed, R.O., Bekele, E.J., Zheng, M., Xian, M., Abdalla, B.A., Adeola, A.C., Adetula, A.A., Lawal, R.A., Zhu, W., Zhang, D., Zhang, X., Ji, C., Nie, Q., 2022. Research note: association of single nucleotide polymorphism of AKT3 with egg production traits in White Muscovy ducks (Cairina moschata). Poult. Sci. 101, 102211.

Benton, M.J., Donoghue, P.C., 2007. Paleontological evidence to date the tree of life. Mol. Biol. Evol. 24, 26–53.

Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. Genome Res. 14, 988–995.

Choi, J.Y., Lye, Z.N., Groen, S.C., Dai, X., Rughani, P., Zaaijer, S., Harrington, E.D., Juul, S., Purugganan, M.D., 2020. Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. Genome Biol. 21, 21.

Deamer, D., Akeson, M., Branton, D., 2016. Three decades of nanopore sequencing. Nat. Biotechnol. 34, 518–524.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids. Res. 32, 1792–1797.

Fournier, E., Peresson, R., Guy, G., Hermier, D., 1997. Relationships between storage and secretion of hepatic lipids in two breeds of geese with different susceptibility to liver steatosis. Poult. Sci. 76, 599–607.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., Bateman, A., 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic. Acids. Res. 33, D121–D124.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to assemble spliced alignments. Genome Biol. 9, R7.
Huang, Y., Li, Y., Burt, D.W., Chen, H., Zhang, Y., Qian, W., Kim, H., Gan, S., Zhao, Y., Li, J., Yi, K., Feng, H., Zhu, P., Li, B., Liu, Q., Fairley, S., Magor, K.E., Du, Z., Hu, X., Goodman, L., Tafer, H., Vignal, A., Lee, T., Kim, K.W., Sheng, Z., An, Y., Searle, S., Herrero, J., Groenen, M.A.M., Crooijmans, R., Faraut, T., Cai, Q., Webster, R.G., Aldridge, J.R., Warren, W.C., Bartschat, S., Kehr, S., Marz, M., Stadler, P.F., Smith, J., Kraus, R.H.S., Zhao, Y., Ren, L., Fei, J., Morisson, M., Kaiser, P., Griffin, D. K., Rao, M., Pitel, F., Wang, J., Li, N., 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat. Genet. 45, 776–783.
Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A. D., Dilthey, A.T., Fiddes, I.T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B.S., Rhie, A., Richardson, H., Quinlan, A.R., Snutch, T.P., Tee, L., Paten, B., Phillippy, A.M., Simpson, J.T., Loman, N.J., Loose, M., 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. 36, 338–345.
Ju, X., Wang, Z., Cai, D., Xu, H., Bello, S.F., Zhang, S., Zhu, W., Ji, C., Nie, Q., 2023. TAT gene polymorphism and its relationship with production traits in Muscovy ducks (Cairina Moschata). Poult. Sci. 102, 102551.
Kent, W.J., 2002. BLAT–the BLAST-like alignment tool. Genome Res. 12, 656–664.
Korlach, J., Bjornson, K.P., Chaudhuri, B.P., Cicero, R.L., Flusberg, B.A., Gray, J.J., Holden, D., Saxena, R., Wegener, J., Turner, S.W., 2010. Real-time DNA sequencing from single polymerase molecules. Method. Enzym. 472, 431–455.
Li, L., Stoeckert Jr., C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189.
Majoros, W.H., Pertea, M., Salzberg, S.L., 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–2879.
Marcais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770.
Michael, T.P., Jupe, F., Bemm, F., Motley, S.T., Sandoval, J.P., Lanz, C., Loudet, O., Weigel, D., Ecker, J.R., 2018. High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. Nat. Commun. 9, 541.
Mondal, T.K., Rawal, H.C., Gaikwad, K., Sharma, T.R., Singh, N.K., 2017. First de novo draft genome sequence of Oryza coarctata, the only halophytic species in the genus Oryza. F1000Res 6, 1750.
Ning, H., Ao, S., Fan, Y., Fu, J., Xu, C., 2018. Correlation analysis between the karyotypes and phenotypic traits of Chinese cymbidium cultivars. Hortic. Environ. Biotechnol. 59, 93–103.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680.
Schmidt, M.H., Vogel, A., Denton, A.K., Istace, B., Wormit, A., van de Geest, H., Bolger, M.E., Alseekh, S., Mass, J., Pfaff, C., Schurr, U., Chetelat, R., Maumus, F., Aury, J.M., Koren, S., Fernie, A.R., Zamir, D., Bolger, A.M., Usadel, B., 2017. De novo assembly of a new solanum pennellii accession using nanopore sequencing. Plant Cell. 29, 2336–2348.
Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006a. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic. Acids. Res. 34, W435–W439.
Stanke, M., Schoffmann, O., Morgenstern, B., Waack, S., 2006b. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinform. 7, 62.
Suryamohan, K., Krishnankutty, S.P., Guillory, J., Jevit, M., Schroder, M.S., Wu, M., Kuriakose, B., Mathew, O.K., Perumal, R.C., Koludarov, I., Goldstein, L.D., Senger, K., Dixon, M.D., Velayutham, D., Vargas, D., Chaudhuri, S., Muraleedharan, M., Goel, R., Chen, Y.J., Ratan, A., Liu, P., Faherty, B., de la Rosa, G., Shibata, H., Baca, M., Sagolla, M., Ziai, J., Wright, G.A., Vucic, D., Mohan, S., Antony, A., Stinson, J., Kirkpatrick, D.S., Hannoush, R.N., Durinck, S., Modrusan, Z., Stawiski, E.W., Wiley, K., Raudsepp, T., Kini, R.M., Zachariah, A., Seshagiri, S., 2020. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nat. Genet. 52, 106–117.
Tan, M.H., Austin, C.M., Hammer, M.P., Lee, Y.P., Croft, L.J., Gan, H.M., 2018. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (Amphiprion ocellaris) genome assembly. Gigascience 7, 1–6.
van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., Dekker, J., Lander, E.S., 2010. Hi-C: a method to study the three-dimensional architecture of genomes. J. Vis. Exp. 39, e1869.
Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204.
Wang, Z., Guo, Z., Liu, H., Liu, T., Liu, D., Yu, S., Tang, H., Zhang, H., Mou, Q., Zhang, B., Cao, J., Schroyen, M., Hou, S., Zhou, Z., 2025. A high-quality assembly revealing the PMEL gene for the unique plumage phenotype in Liancheng ducks. Gigascience 14, giae114.
Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., Andrews, S., 2015. HiCUP: pipeline for mapping and processing hi-C data. F1000Res 4, 1310.
Xiao, Y., Xiao, Z., Ma, D., Liu, J., Li, J., 2019. Genome sequence of the barred knifejaw Oplegnathus fasciatus (Temminck & Schlegel, 1844): the first chromosome-level draft genome in the family Oplegnathidae. Gigascience 8, giz013.
Xu, L., Shi, Z., Li, H., He, J., Chen, B., Tao, Z., Tian, Y., Chen, L., Li, G., Tao, Z., Gu, T., Xu, W., Lu, L., 2022a. Genome-wide DNA methylation differences between conservation and breeding populations of Shaoxing ducks. Heliyon 8, e11644.
Xu, M.M., Gu, L.H., Lv, W.Y., Duan, S.C., Li, L.W., Du, Y., Lu, L.Z., Zeng, T., Hou, Z.C., Ma, Z.S., Chen, W., Adeola, A.C., Han, J.L., Xu, T.S., Dong, Y., Zhang, Y.P., Peng, M. S., 2022b. Chromosome-level genome assembly of the Muscovy duck provides insight into fatty liver susceptibility. Genomics 114, 110518.
Xu, Z., Wang, H., 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic. Acids. Res. 35, W265–W268.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yang, Z., Wang, Z.Y., Yang, H.M., Xu, L., Gong, D.Q., 2017. Effects of dietary methionine and betaine on slaughter performance, biochemical and enzymatic parameters in goose liver and hepatic composition. Anim. Feed. Sci. Technol. 228, 48–58.
Ye, P., Li, M., Liao, W., Ge, K., Jin, S., Zhang, C., Chen, X., Geng, Z., 2019. Hypothalamic transcriptome analysis reveals the neuroendocrine mechanisms in controlling broodiness of Muscovy duck (Cairina moschata). PLoS. One 14, e0207050.
Ye, Q., Xu, J., Gao, X., Ouyang, H., Luo, W., Nie, Q., 2017. Associations of IGF2 and DRD2 polymorphisms with laying traits in Muscovy duck. PeerJ 5, e4083.
Yu, S., Liu, Z., Li, M., Zhou, D., Hua, P., Cheng, H., Fan, W., Xu, Y., Liu, D., Liang, S., Zhang, Y., Xie, M., Tang, J., Jiang, Y., Hou, S., Zhou, Z., 2023. Resequencing of a Pekin duck breeding population provides insights into the genomic response to short-term artificial selection. Gigascience 12, giad016.
Zhang, B.W., Xu, L.L., Li, N., Yan, P.C., Jiang, X.H., Woeste, K.E., Lin, K., Renner, S.S., Zhang, D.Y., Bai, W.N., 2019a. Phylogenomics reveals an ancient hybrid origin of the Persian walnut. Mol. Biol. Evol. 36, 2451–2461.
Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22, 2472–2479.
Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C.M., Zhang, C., Tian, Y., Liu, G., Gul, H., Wang, D., Tian, Y., Yang, C., Meng, M., Yuan, G., Kang, G., Wu, Y., Wang, K., Zhang, H., Wang, D., Cong, P., 2019b. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat. Commun. 10, 1494.
Zhang, X., Tang, B., Li, J., Ouyang, Q., Hu, S., Hu, J., Liu, H., Li, L., He, H., Wang, J., 2023. Comparative transcriptome analysis reveals mechanisms of restriction feeding on lipid metabolism in ducks. Poult. Sci. 102, 102963.