104 (2025) 105004

Contents lists available at ScienceDirect

Poultry Science

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

Full-Length Article

Whole-genome resequencing reveals the population structure and domestication processes of endemic endangered goose breeds (*Anser cygnoides*)

Shangzong Qi [a], Teng Wu [a], Hao Wu [a], Yu Liang [a], Wenming Zhao [a], Yang Zhang [a,c,*], Qi Xu [a,b,c], Guohong Chen [a,b,c]

[a] College of Animal Science and Technology, Yangzhou University, Yangzhou, Jiangsu Province 225009, PR China
[b] Joint International Research Laboratory of Agriculture and Agri-Product Safety of the Ministry of Education of China, Yangzhou University, Yangzhou, Jiangsu Province, 225009, PR China
[c] Key Laboratory for Evaluation and Utilization of Livestock and Poultry Resources (Poultry), Ministry of Agriculture and Rural Affairs, PR China

ARTICLE INFO

Keywords:
Endemic endangered geese
Whole-genome resequencing
Population genetic structure
Selective sweep

ABSTRACT

In recent years, the dwindling population of these endangered geese has hindered our understanding of their phenotypic variations and the genes associated with important traits. To investigate the population structure and genetic diversity of this breed, the whole-genome data of 90 individuals from a conservation farm were obtained using the Illumina 6000 paired-end platform. The research results indicate that each locally endangered goose variety has formed a monophyletic population. The Baizi (**BZ**), Lingxian White (**LX**), and Xupu (**XP**) geese exhibiting higher genetic diversity than the other goose breeds. Tree-Mix analysis revealed the presence of five gene flows events between goose populations, with Yangjiang (**YJ**) geese consistently exhibiting significant genetic distance from the other breeds. Under strong pressures from the natural environment and artificial selection, whole-genome selective scanning revealed 394 overlapping genes. Gene Ontology (**GO**) enrichment analysis of the putative candidate genes (**PCGs**) revealed significant enrichment of 20 terms ($P < 0.05$). Similarly, Kyoto Encyclopedia of Genes and Genomes (**KEGG**) enrichment analysis revealed significant enrichment of PCGs in 23 terms ($P < 0.05$). Examination of overlapping genes identified through at least two selection methods revealed a set of genes associated with key traits, including growth and development (*CCND1, DES, CCNO, SMC5, and NUBP1*), immunity (*ABCA2, ABCC8, UHRF2,* and *ABCA1*), and body aging (*KAT6B*). Our findings provide insights into the genetic basis of endangered geese at the whole-genome level, laying the foundation for future molecular research on genetic variation and phenotypic changes. In summary, our results provide invaluable resources for delineating the uniqueness of endangered goose breeds.

Introduction

Livestock and poultry genetic resources are important components of Earth's biodiversity, playing an irreplaceable role in maintaining ecological balance and human well-being. Livestock production in general and domestic chicken production in particular plays a vital socio-economic role for people living in low-income countries of Africa and Asia (Mohammadifar et al. 2014; Moazeni et al. 2016; Mohamadi-nejad et al. 2024). However, owing to global climate change, biological invasions, habitat destruction from human activities, environmental pollution, and land changes, local goose species are facing greater threats than ever before (Watermeyer et al., 2021). Therefore, the top priority in the current research field is how to optimize the population structure through advanced biotechnology, while fully protecting and utilizing the high-quality characteristics of these endangered goose breeds. In recent decades, modern European breeds, such as the **Rhine goose** (*Anser anser*), which has fast growth and high feed conversion, and the **Landes goose** (*Anser anser*), have been introduced into China, becoming the dominant choice for gosling breeding in China (Qu et al., 2016). This has led to a significant reduction in the number of rare local

* Corresponding authors at: No. 48 Wenhui East Road, College of Animal Science and Technology, Yangzhou University, Yangzhou, Jiangsu Province, 225009, PR China.
E-mail address: zyang@yzu.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.psj.2025.105004
Received 2 January 2025; Accepted 6 March 2025
Available online 8 March 2025

geese in China. The number of purebreds is decreasing, causing the extinction of some domestic breeds, such as **Caohai, Wenshan** and **Simao goose** (*Anser cygnoides*) (Ouyang et al., 2022; Qi et al., 2024). These endemic genetic resources are vital to the ecosystem and are the cornerstone of the sustainable development of the waterfowl breeding industry.

To preserve and protect the diversity of rare and endangered goose breeds, the third national resource census found that the number of endemic goose breeds, including the BZ, LX, YJ, YE, WZ, and XP goose (*Anser cygnoides*), had fewer than 1,000 individuals (Chen et al., 2022). Therefore, government departments have launched a national protection plan for native goose breeds (China National Commission of Animal Genetic Resources). Currently, the protection of waterfowl genetic resources in China relies primarily on in vivo preservation. It performs conducts in vivo preservation by establishing in vivo gene banks, conservation farms, and protected areas in the breeds' places of origin (Tscharntke et al., 2024). Due to the limitations of biotechnological methods, in vivo preservation remains the most practical method and will continue for a considerable amount of time. Therefore, optimization of the population genetic structure of endemic endangered goose breeds is paramount.

Evaluation of genetic diversity relies mainly on the use of genetic markers. Currently, DNA molecular markers are the most stable, polymorphic, and widely used genetic markers. Among these, single-nucleotide polymorphisms (**SNPs**) are the most representative. Single nucleotide polymorphism SNP analysis is based on genetic marker genotyping methods widely used at the genomic level. A comparative analysis of SNPs within the whole genome of individuals in a population, genetic diversity of the population, genetic relationships among individuals, and the population structure (Wang et al., 2020). With advancements in sequencing technology and reduced sequencing costs, SNP analysis based on WGS is now widely applied to studies of population genetic diversity and population structure of domestic animals, including horses (Gu et al., 2023), cattle (Zhang et al., 2022), pigs (Yang et al., 2022), and chickens (Cho et al., 2022). Chen et al. (2024) identified genes affecting skin color in **Luning black-bone chickens**, including *ATP5E, EDN3* and *LOC101750371*, using whole-genome resequencing technology. Similarly, Gu used genome-wide association studies analysis technology to explore the genomic sequences of five different duck breeds and they selectively screened positively selected genes involved in physiological processes, such as pigment deposition and muscle contraction (Gu et al., 2020). Research on endemic endangered goose breeds has primarily focused on mitochondrial DNA control region sequences (extrachromosomal genetic material) to study population genetic diversity and origin differentiation. However, no studies have analyzed the genetic diversity and population genetic structure of endangered goose breeds in terms of the genetic material within the cell nucleus.

To achieve this goal, we conducted research at the mitochondrial gene level of endangered goose breeds (Qi et al., 2024). This study represents the first effort to perform resequencing data analysis on the entire genes within the nuclei of local endangered goose breeds, and to examine and assess their current endangered status. In contrast to previous studies, the results of this research have broader application prospects. We can accurately infer the evolutionary history of species, optimize the population genetic structure of genomic markers, eliminate inbreeding individuals, and screen for genes used for physiological functions in endangered geese under selection pressure. Consequently, this study promotes the sustainable utilization of endemic endangered goose breed resources, enhances the understanding of their origin and domestication, and provides valuable references for breeding plans, selection, and mating strategies in waterfowl conservation.

Materials and methods

Ethics statement

All animal experiments were approved by the Institutional Animal Care and Use Committee of Yangzhou University (Approval No: 132-2022). All procedures were performed in accordance with the "Regulations on the Administration of Laboratory Animal Affairs" (Yangzhou University, 2012) and the "Management Standards for Experimental Practices" (Jiangsu, China, 2008).

Animals and environmental conditions

In accordance with the distribution of China's rare and endangered goose breeds, blood samples were collected from six breeds from diverse geographical locations (Fig. 1). Male specimens of LX, YJ, YE, WZ, XP, and BZ geese were selected as research subjects (Table 1). The geese were categorized into six distinct groups based on their specific breeds (n=15 birds/group), and no kinship existed among these groups. After sterilization of the wings of the geese, venous blood was carefully collected from each of the 90 samples. The blood specimens were placed in anticoagulant-containing tubes (EDTA-K2, KJ0501K2E, 13×100mm) and maintained at a temperature of -20 °C pending DNA extraction. All goose breeds were reared by the National Waterfowl Gene Conservation Bank (Taizhou, China), and their body weights were measured from 0 to 10 weeks of age.

DNA sequencing and reference genome

The experimental procedure is illustrated in Fig. 2. A TIANamp Genomic DNA Kit (YDP304-03, TIANGEN Biotech, Beijing, China) was used to extract genomic DNA from the 90 collected goose blood samples according to the manufacturer's protocol (concentration: 227 ± 58.2 ng/μL). The integrity and purity of the DNA were tested using 1 % agarose gel electrophoresis and the A260/280 ratio (1.8-2.0) respectively. All qualified samples were used for library construction (paired-end, 2150 bp), and WGS was performed on an Illumina NovaSeq 6000 platform at Novogene (Beijing, China). All samples were sequenced with an average coverage of 6× (Fig. 3). The reference genome of the goose breed was sourced from the National Center for Biotechnology Information database, and the website is: (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002166845.1/). The obtained original reads from sequencing were aligned to the reference genome using bwa mem (version 0.7.17-r1188) software, and then SNPs were called using GATK (version v4.1.4.1) software with the parameters (Window 4, filter "QD < 2.0 || FS > 60.0 || MQ < 40.0," G-filter "GQ < 20") to filter the previous SNPs.

SNP detection and variant annotation

To maximize the exclusion of artificial biases during base detection and adapter contamination, we used Trimmomatic v0.32 software to discard adapters (>10 nt) and low-quality reads (Q < 30). We used the Burrows-Wheeler Aligner (**BWA**, version 0.7.17-r1188, https://github.com/lh3/bwa) to map the high-quality data of each individual to the reference goose genome with the parameters "mem -t 10 -k 32." Sequence Alignment/Map (**SAM**) tools (version 0.1.19, https://github.com/samtools/samtools) were used to convert the SAM and binary versions of the Binary Alignment/Map (https://gitcode.com/gh_mirrors/ba/bamtools) formats, sorting, and indexing alignment. After mapping, SNP detection was performed for each group using the *Bayesian* method implemented in the SAM tool package. Single SNPs were detected by the command "mpileup." To eliminate SNP detection errors caused by misalignment and Insertion (**InDel**), only high-quality SNPs were used, namely those with a coverage depth ≥ 4 and ≤ 200, a root mean square mapping quality ≥ 20, a distance between adjacent SNPs ≥

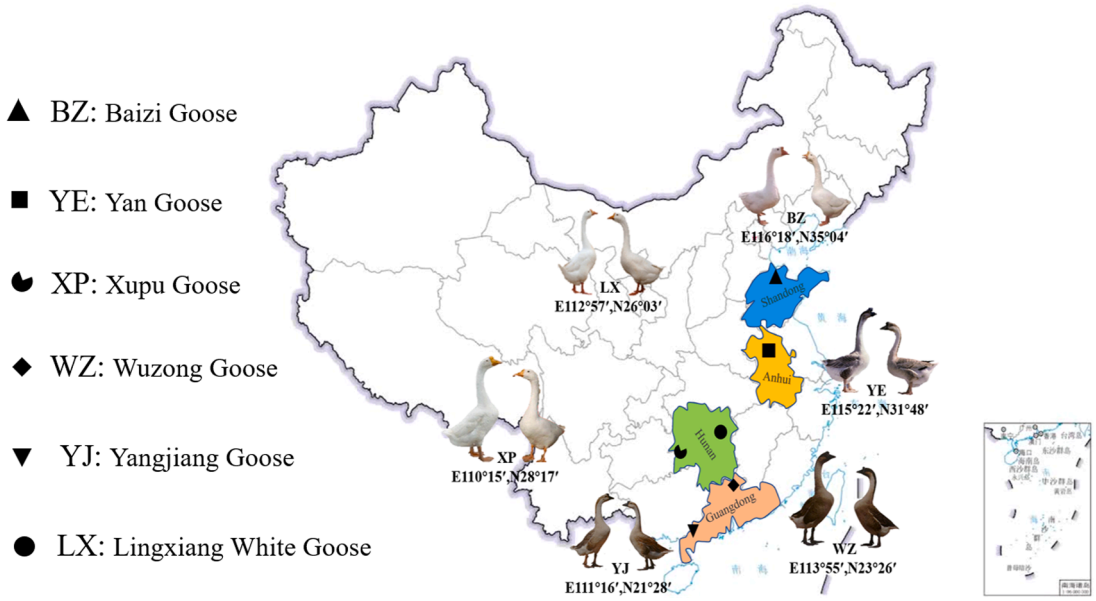

Fig. 1. Geographical collection sites of the six endemic endangered goose breeds. Images of the different breeds of geese were captured using a digital camera (Osaka, Japan).

Table 1
Endemic endangered goose breeds resource information.

Breed	Birthplace	Appearance characteristics	Number/w	Protection level	Sample/bird
BZ	Jinxiang, Shandong	Wide and short body with pure white feathers	1.00	Dangerous	15
LX	Zhuzhou, Hunan	Lay a hundred eggs and have pure white feathers	0.04	Endangered	15
XP	Xupu, Hunan	Head sarcoma and black gray feathers	0.60	Endangered	15
YE	Lu'an, Anhui	Flat sarcoma with black and gray feathers	0.09	Endangered	15
YJ	Yangjiang, Guangdong	Large body size and pure white feathers	0.09	Endangered	15
WZ	Qingyuan, Guangdong	Black mane with back feathers and black gray feathers	0.20	Endangered	15

Note: LX: Lingxiang white goose; YE: Yan goose; YJ: Yangjiang goose; WZ: Wuzong goose; XP: Xupu goose; BZ: Baizi goose. w represents 10,000 geese, for instance, the number/w = 1.00 indicates that there are currently 10,000 BZ geese in China. We have included detailed information on six endangered goose species in the supplementary materials (**Figure S1**).

5 bp, no InDel within a 3 bp window, and a missing rate ≥ 50 % in each group (Fig. 3) (Yin et al., 2019). The variants were annotated and predicted using SNP-Eff v5.0 (Cingolani et al., 2012), and the VCF-tools format (https://github.com/vcftools/vcftools) file can be used as both input and output. The SNPs that met these criteria were retained for subsequent analyses.

Population structure and genetic diversity

After filtering, we detected all SNPs were used for phylogenetic and population structure analyses. We used 100 K as a window, counted the number of SNPs within the window, and used the R package CMplot (https://github.com/YinLiLin/CMplot) to draw the density plot. The population sub-program within Stacks software (version 2.59) was used to perform statistical analysis of the population information. The statistical metrics encompass the number of SNPs (**NSNPs**), runs of homozygosity (**ROH**), inbreeding coefficient (**Fis**), linkage disequilibrium (**LD**), and nucleotide diversity ($\boldsymbol{\pi}$). PLINK (version 1.9) was used to estimate the ROH for each individual's autosomal SNPs. Principal component analysis (**PCA**) was performed using GCTA software (version 1.92.1) (http://cnsgenomic.org/software/gcta/#Overview). A phylogenetic tree is a branching diagram or tree that describes the evolutionary sequences among groups and is used to represent the evolutionary relationships. We constructed a phylogenetic tree based on the SNPs using the Maximum Likelihood method in FastTree software (version v2.1.9) (http://www.microbesonline.org/fasttree/). Population structure analysis helps elucidate evolutionary processes. Admixture software (version v1.3.0) (http://software.genetics.ucla.edu/admixture/) was used to conduct the population genetic structure analysis (K = 2–9). PopLDdecay software (Zhang et al., 2019) (https://github.com/BGI-shenzhen/PopLDdecay) was used for the LD analysis. The parameter configurations were as follows: MaxDist 500, Het 0.1, Miss 0.3, and OutPairLD 5. The distribution of LD on chromosomes is typically depicted using an LD decay plot, through which the decay rate of LD in relation to genetic or physical distance can be observed. The Pairwise Sequential Markovian Coalescence model (**PSMC**, version 0.6.5-r67) was used to infer fluctuations in the history of the effective population size based on genomic fragments with different densities of heterozygous sites. The TreeMix software (version 1.13) utilizes allele frequency data across the whole-genome to infer differentiation and gene flow among multiple populations(https://bitbucket.org/nygcresearch/treemix/downloads/).

Selective sweep detection

To identify the positive characteristics of six endemic endangered goose breeds, two statistical methods, namely the population differentiation statistic (**Fst**) and the nucleotide diversity statistic ($\boldsymbol{\theta\pi}$), were adopted to improve the detection effect. The fixation index, also known as the genetic differentiation index or Fst analysis (Hudson et al., 1992), can be used to calculate the degree of differentiation among populations based on genetic information. The formula for calculating the Fst is as follows:

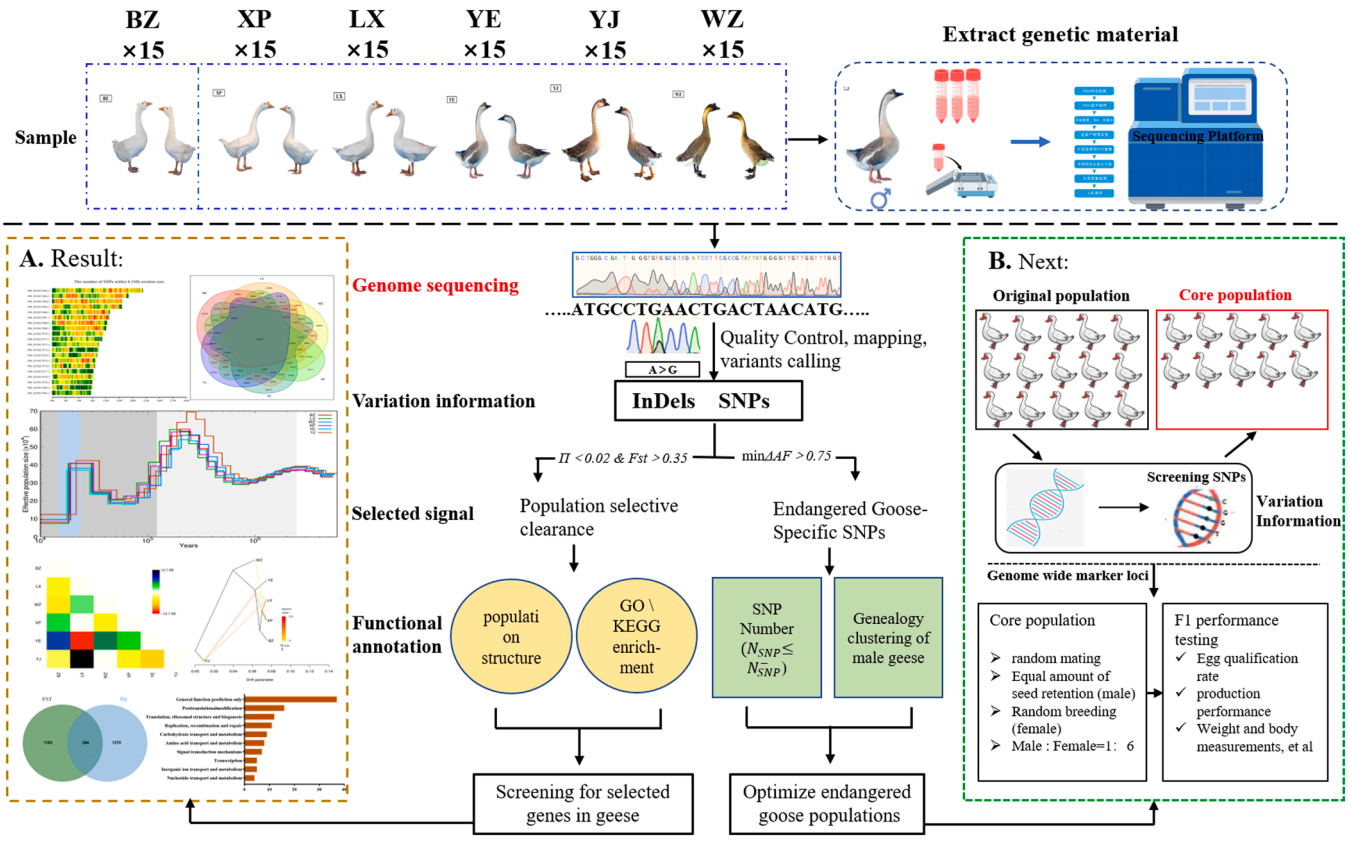

Fig. 2. Experimental workflow.

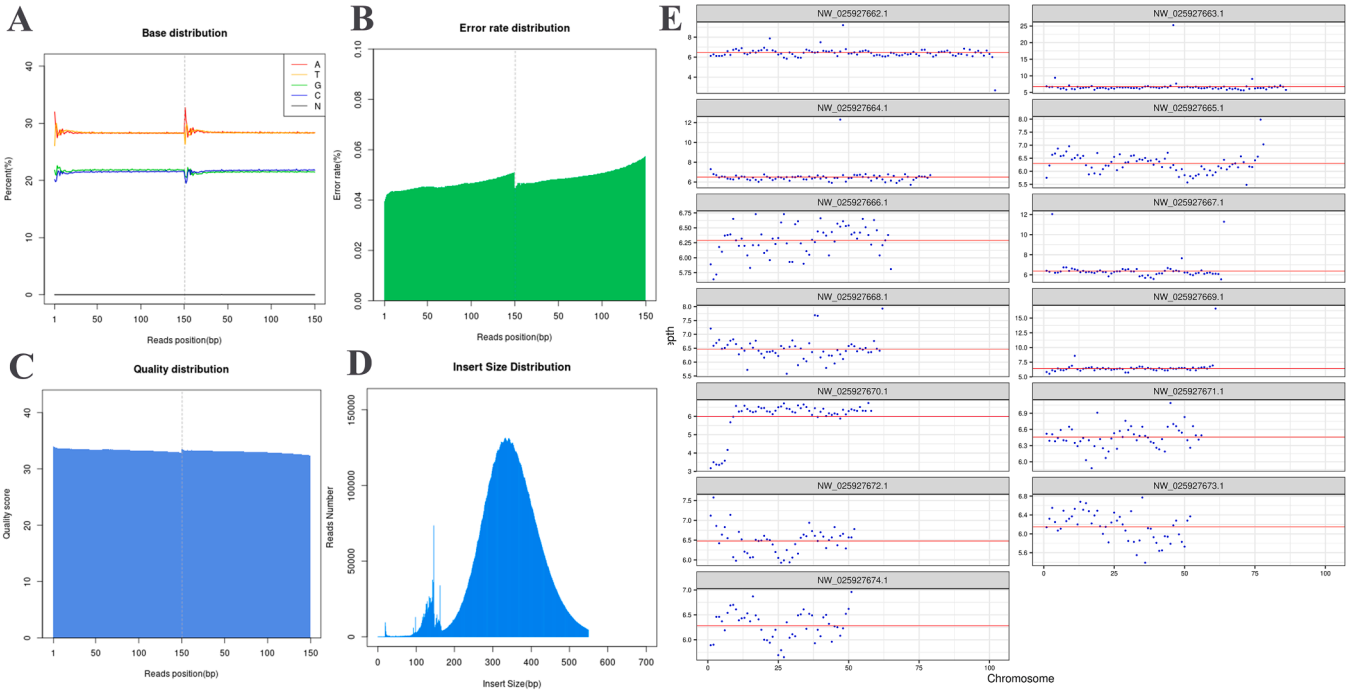

Fig. 3. A. Distribution of the proportion of each base in the original data; B. Distribution of the base error rate in the original data; C. Distribution of the base quality value in the original data; D. Distribution of the insert size; E. Depth distribution of samples on the genome.

$$Fst = \frac{(\pi_{Between} - \pi_{Within})}{\pi_{Between}}$$

π_{Within}: the average number of individual differentially paired bases within the group; $\pi_{Between}$: the average number of individual differentially paired bases between groups.

The top 1 % was selected as the significance threshold for the Fst. Nucleotide diversity reflects the genetic diversity of a population. The θπ

analysis is a method that uses SNPs to calculate the average difference between any two nucleotide sequences in a population (Nei and Li, 1979). The calculation formula of the π ratio is as follows:

$$\theta\pi_{\text{ratio}} = \frac{\pi_{(A)}}{\pi_{(B)}}$$

$\pi_{(A)}$: the π values of the control group; $\pi_{(B)}$: the π values of the selection group.

The top 1 % is selected as the significance threshold of the π ratio. Using BZ as the experimental group and the other geese as the benchmark, a selective sweep analysis was performed based on the population differentiation index Fst. The vcf-tools software (version 0.1.16) was used for the analysis. The parameters were set as follows: taking 100 K as a window and 10 K as the step size to extract a region, and then calculating the population Fst and the π value of the population within this region for comparison. The top 1 % were selected as the selected region. First, the regions were ranked according to the Fst value, and regional information for the top 1 % of the size was selected. Then, the regions were ranked according to the π ratio of the two populations, and the top 1 % and bottom 1 % were taken. Finally, the intersection of the regions screened out based on Fst and the ratio of π was taken as the association result. We annotated the genes in these regions to the reference geese genome and submitted all of the data to the **GO** and **KEGG** databases for enrichment analysis using the Enrichr-Web server. The variations within them were considered reliable signals. Online Metascape was used to conduct GO enrichment analysis of genes that overlapped with the selective windows (Bu et al., 2021).

Results

Regularly measure the body weight of geese at different weeks of age

During the experimental feeding period, significant differences were observed among the six endangered goose breeds of different ages ($P < 0.05$) (Table 2). At 1 d of age, the body weights of the XP geese were significantly higher than those of the other breeds ($P < 0.05$). At four weeks of age, the body weights of the LX and YJ geese were significantly higher than those of the other breeds ($P < 0.05$). At ten weeks of age, the body weights of the YJ, LX, and XP geese were significantly higher than those of the other breeds ($P < 0.05$).

Comparison between the quality of sequencing data and the reference genome

Sequencing data output and quality control

The output and quality control of sequencing data are important foundational steps in research and their results directly affect the reliability of subsequent data analyses and biological conclusions. A total of 90 blood DNA samples (Concentration: 426.7 ± 24.25 ng/μL) were extracted for this study, all of which passed quality inspection. The bands in the agarose gel were clear and bright (**Fig. S2**). Sequencing libraries of the six endangered goose breeds were constructed based on sequencing technology using of the Illumina NovaSeq 6000 platform. Data quality control statistics (**Table S1**) showed that 641.32 Gb of clean data were generated from the reduced genome sequencing of individual geese, among which the total number of paired-end reads was 2,111,506,927. The total number of sequences for each breed ranged from 22,505,812 to 26,229,454. Regarding the total bases in the clean data of the six breeds, the average proportions of the bases A, C, G, and T were 28.24 %, 21.73 %, 21.80 %, and 28.22 % respectively. For the samples of each breed, the average Q30 value across all of the samples exceeded 92.86 %, implying that the sequencing results were relatively accurate, the base error rate was low, the data quality was satisfactory, and conformed to resequencing standards, thus making them suitable for further analysis.

SNP detection and locus development

We adopted the analysis steps of BWA and GATK to identify and genotype the mutation sites in endangered goose breeds. A strict filtering process was applied to the data (**Table S2; Table S3**). In total, 396,527,757 SNP sites were obtained from the 90 samples, with the average number of SNP sites being 4,405,863. In total, 113,910,513 transversion SNP sites were detected and the average number of transversion SNP sites was 1,265,672. The ratio of transitions to transversions ranged from 2.24 to 2.49, and all the transition/transversion values were greater than 1.5. We observed that 30.4 % of the SNPs were located in the intergenic regions of the genome, and 52.9 % of the SNPs were observed to be located in the intron regions of genes after performing functional annotation of SNPs using the SNP-eff v5.1 (Fig. 4A). Functional annotation of SNPs in the protein-coding regions of genes indicated that 34.9 % of the SNPs led to non-synonymous amino acid substitutions, 58.4 % were synonymous substitutions, and 6.7 % were of stop or start codons mutations. PLINK software was used to calculate the ROH of the six endangered goose breeds (Fig. 4B and C). The results demonstrate revealed significant differences among goose breeds (**SROH** *P*-value = 0.0094; **NROH** *P*-value = 0.0154). Among these, WZ had the highest average ROH length (SROH = 6.2 Mb) and number of ROH. In contrast, XP had the lowest average ROH length (< 2 Mb) and number of ROH. These two indicators reflect the different degrees of inbreeding and intensities of artificial selection in these two breeds.

Analysis of population genetic diversity of goose breeds

Statistics of the number of chromosomal SNPs

Genotyping was performed using GATK software. After mutation site detection and filtering, we plotted the chromosome distribution maps (Fig. 5A and B) based on the positions of the SNP sites on the chromosomes. SNPs were observed to be unevenly distributed across chromosomes. The largest number of SNP sites (183, 658 in total) was identified on chromosome 1 (**Chr1** NW_025927662.1). The lowest number of SNP sites, 43,104, was identified on Chr19 (NW_025927680.1). Overall, SNP sites were concentrated on the first four chromosomes. Moreover, the number of SNPs was positively correlated with chromosome length. Relatively more SNPs were observed on the large chromosomes (the number of SNPs on the first eight chromosomes exceeded 100,000). Ultimately, 8,106,645 SNPs were identified on the chromosomes of the

Table 2
Analysis of body weight differences among endangered goose aged 0-10 weeks.

Breed	0*	2	4	6	8	10
BZ	66.79 ± 7.19^{d}	549.38 ± 48.29^{b}	906.23 ± 115.63^{d}	1705.97 ± 184.16^{d}	2297.64 ± 251.46^{c}	2590.29 ± 259.94^{c}
LX	90.57 ± 11.14^{c}	635.97 ± 60.72^{a}	1653.30 ± 103.29^{a}	2112.67 ± 162.13^{c}	3297.07 ± 216.13^{a}	3901.54 ± 295.90^{a}
XP	115.13 ± 11.48^{a}	539.69 ± 39.78^{b}	1291.82 ± 150.00^{c}	2317.45 ± 261.24^{b}	2884.01 ± 301.21^{b}	3905.29 ± 319.79^{a}
YE	105.92 ± 2.07^{b}	512.45 ± 41.50^{b}	1089.77 ± 12.48^{c}	1570.41 ± 129.37^{d}	2169.72 ± 280.82^{c}	3431.31 ± 227.62^{b}
YJ	90.08 ± 7.56^{c}	649.83 ± 34.52^{a}	1489.16 ± 122.51^{b}	2753.66 ± 112.82^{a}	3183.00 ± 308.29^{a}	3893.31 ± 228.33^{a}
WZ	104.24 ± 4.57^{b}	425.06 ± 53.16^{c}	1159.15 ± 163.84^{c}	1963.31 ± 172.45^{c}	2297.19 ± 275.07^{c}	2752.98 ± 265.10^{c}
P-value	0.001	0.001	0.001	0.018	0.046	0.016

Note: The unit of body weight is grams (**g**). "0*" = Week 0 is the first day of birth. There is no significant difference between the same letters ($P > 0.05$), while there is a significant difference between different letters ($P < 0.05$).

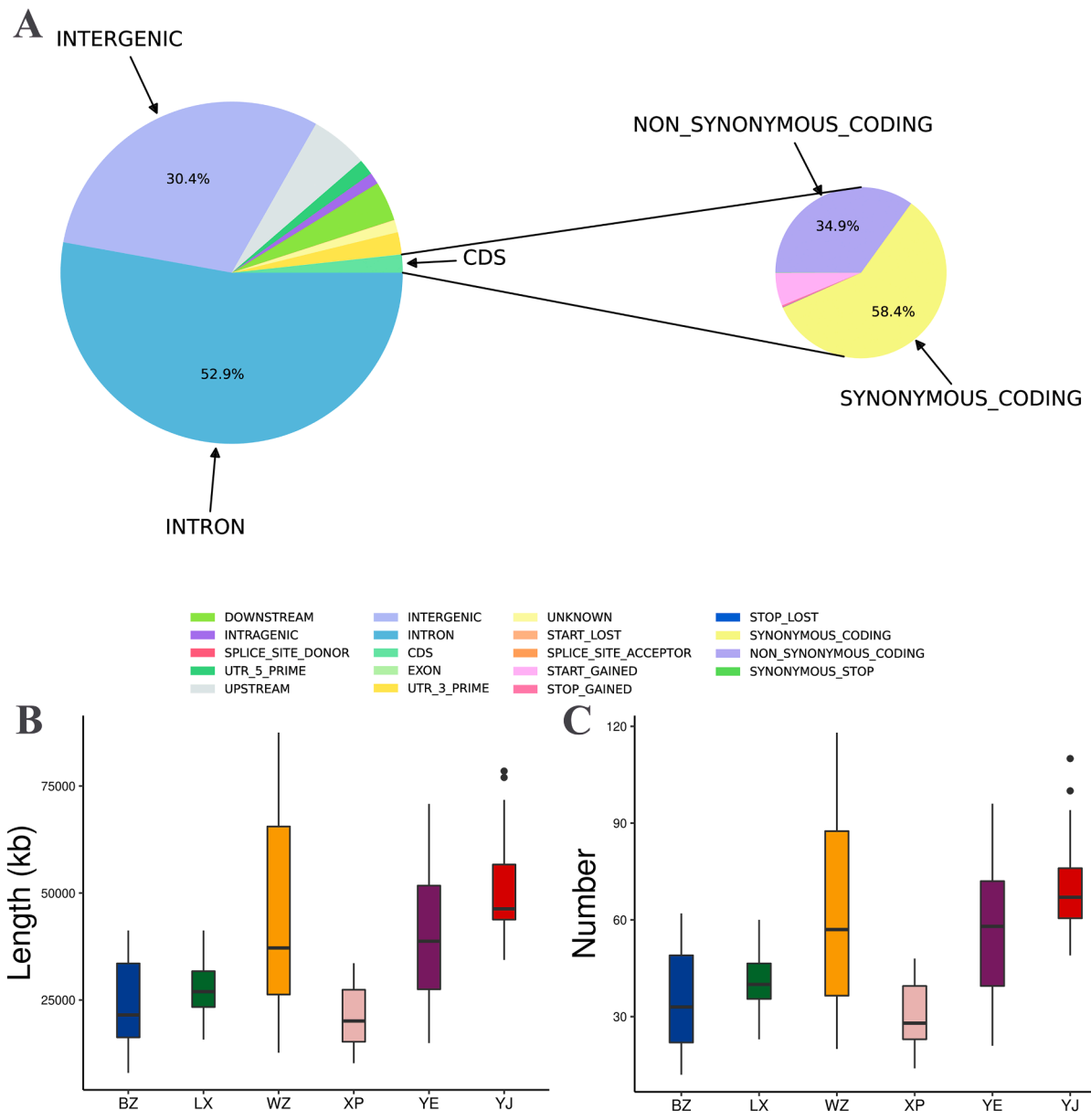

Fig. 4. A. Functional annotation results of the statistical analysis of SNP site; B. Length of homozygous sequences (SROH) in the endangered goose breeds; C. The number of homozygous sequences (NROH) in the endangered goose breeds.

six endangered goose breeds from different regions (**Table S4**). Among them, YJ alone contained 974,604 SNP sites and BZ alone contained 652,248 SNP sites. The six breeds shared 7,370,012 SNP sites (Fig. 5C).

Genetic diversity among populations

The statistical results of the genetic diversity indices and inbreeding coefficients of the six endemic goose breeds (Table 3) demonstrated that the number of polymorphic sites in LX and XP was significantly higher than that in the other goose breeds, with the percentage of polymorphic sites in both reaching over 90 %. Among all breeds, WZ had the highest proportion of minor allele frequencies (**MAF,** MAF = 0.80294). The genetic diversities of the BZ, LX, and XP geese were the highest. The average expected heterozygosities (**He)** were 0.28210, 0.28506, and 0.28972, respectively, and the Pi values were 0.29208, 0.29518, and 0.29999, respectively. YJ had the highest average observed heterozygosity (**Ho**, Ho = 0.24907), and YE had the lowest average (Ho = 0.21956). The heterozygosity of the endangered goose breeds did not reveal any significant characteristics. The inbreeding coefficients of WZ and YJ were the lowest, with Fis values of 0.13641 and 0.10829, respectively, indicating that these two goose breeds have relatively long genetic distances. However, YE had the highest inbreeding coefficient (Fis = 0.18863), possibly due to its relatively small population size, resulting in lower genetic diversity.

Genetic structure analysis of endangered goose breeds

Phylogenetic tree and ancestral population evolution analysis

Comprehensive analyses were conducted to better understand the evolutionary relationships between the ancestral population evolutions of these six endemic endangered goose breeds. We calculated the whole-genome SNP set to perform in-depth investigations of the population LD decay distance, constructed a phylogenetic tree, and analyzed the population genetic structure. This multifaceted approach allowed for a more detailed exploration of genetic interconnections and evolutionary trajectories within these goose breeds. Population LD decay analysis (Fig. 6A) revealed significant differences in the breeding histories of the

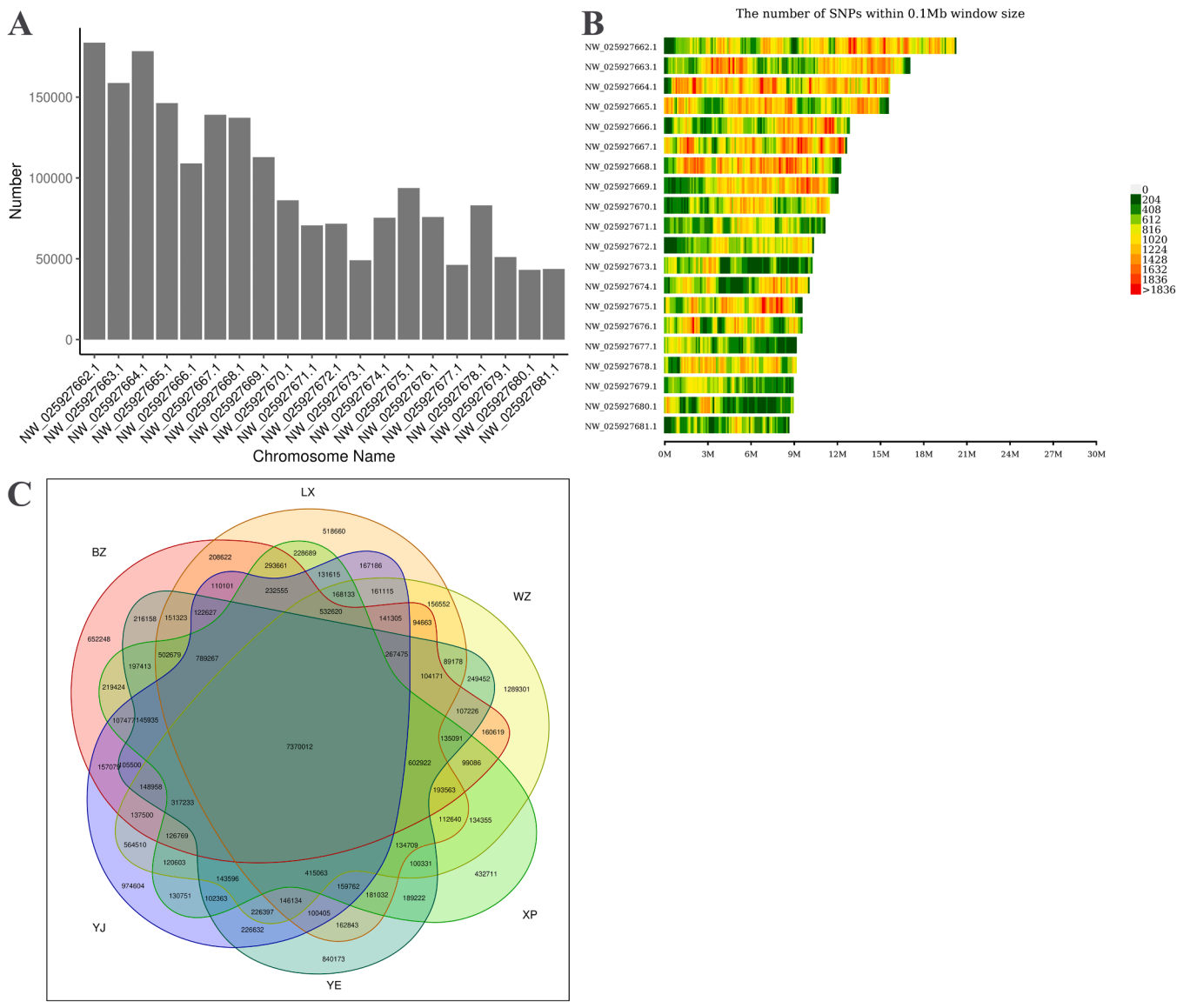

Fig. 5. A. Statistical chart of the number of chromosomal SNPs; B. Distribution map of the density of SNPs on chromosomes; C. Venn diagram of the number of unique and shared SNPs among the six goose groups.

Table 3
Genetic locus information and genetic diversity of endemic endangered goose breeds.

Breed	NP	PL/%	MAF	He	Ho	Pi	Fis
BZ	7143731	88.12	0.79328	0.28210	0.22965	0.29208	0.16576
LX	7318170	90.27	0.79196	0.28506	0.23309	0.29518	0.16419
WZ	6653168	82.07	0.80294	0.26656	0.22421	0.27596	0.13641
XP	7371085	90.92	0.78821	0.28972	0.23356	0.29999	0.17538
YE	7128617	87.94	0.79631	0.27754	0.21956	0.28735	0.18863
YJ	7057080	87.05	0.79438	0.27924	0.24907	0.28901	0.10829

Note: NP: the number of polymorphic loci; PL: the percentage of polymorphic loci; Ho: observed heterozygosity; He: expected heterozygosity; MAF: minor allele frequency; Pi: nucleotide diversity; Fis: inbreeding coefficient.

six endangered goose breeds. Within the genetic distance interval, SNPs in the 0-50 kb range exhibited relatively high r^2 values and rapid decay rate. In contrast, SNPs in the 50-500 kb range displayed lower LD levels that stabilized gradually. The six endangered goose breeds were ranked in ascending sequence with respect to the rate of LD decay: YJ, WZ, YE, BZ, XP, and LX. Among these, the YJ, WZ, and YE geese exhibited relatively slow LD decay speeds, implying a more pronounced degree of linkage. The kinship clustering graph (Fig. 6B) illustrates that all samples from each geographical group clustered independently, forming distinct branches and indicating a favorable clustering outcome. Nevertheless, a minor degree of sample intermixing was observed between the WZ and YE breeds. Consequently, the individuals WZ-8, YE-5, and YE-9 within the conserved population were excluded. These individuals were removed from the protected population samples and disregarded in subsequent statistical analyses to ensure the integrity and accuracy of the research data. To ascertain the proportion of common

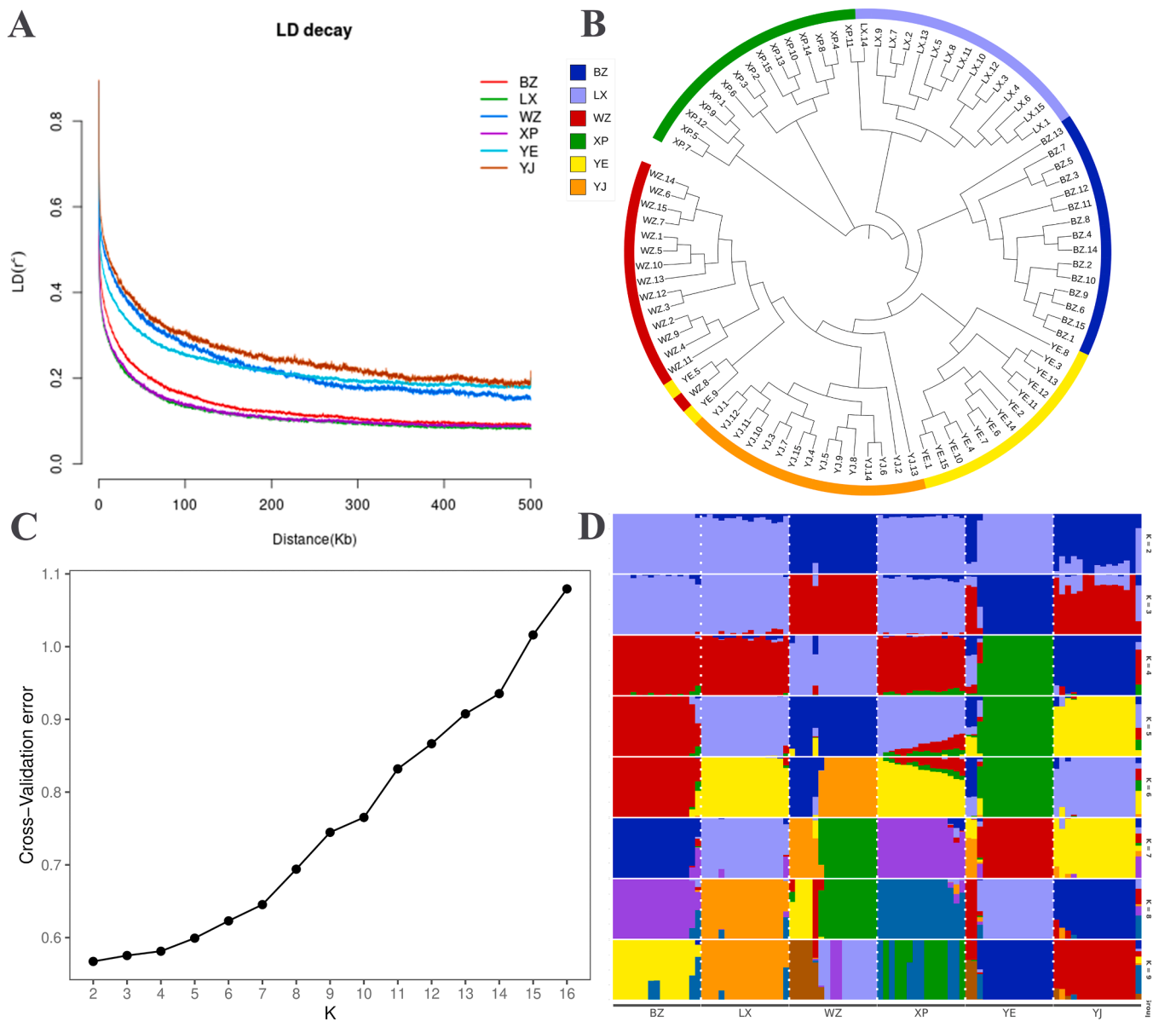

Fig. 6. A. Linkage disequilibrium decay analysis diagram of the six breeds; B. Kinship clustering analysis diagram of the endangered goose breeds, where each vertical line represents an individual, and individuals of the different breeds are divided into different families; C. Genetic structure of the six endangered goose breeds; D. Population cross-validation error values (K = 2-16), with K being the optimal clustering index.

genetic ancestors and admixture level, the genetic structure of the populations was explored (Fig. 6C and D). The cross-validation error rate curve exhibited an overall ascending tendency. When the number of clusters (**K**) was set to 2, the cross-validation error value reached its minimum, suggesting that K = 2 represents the optimal number of clusters. Two genomic clusters were identified at K = 2. The BZ, LX, XP, and YE geese were clustered, whereas the WZ and YJ geese were grouped into separate clusters. This phenomenon implies that two gene pools were shared among all samples of the breeds in question. At K = 7, the endangered goose breeds demonstrated complex pedigree backgrounds and initiated distinct evolutionary pathways. A relatively small number of individuals exhibited a mixed pedigree within a subset of the YE and WZ goose populations. Owing to the influence of living environments and artificial selection, these endangered goose breeds manifested serious differentiation phenomena.

Principal component analysis (PCA)

To evaluate the genetic relationships and structures among the six endangered goose breeds, PCA was performed, and the cumulative contribution rate of each principal component was calculated. The PCA results demonstrated that the first (PC1/2) and third eigenvectors (PC2/3) divided the 90 individuals into four groups (Fig. 7A and C). The BZ, LX, and XP breeds were clustered into one group, while the WZ, YE, and YJ breeds were separated into individual groups. The two principal components accounted for 10.56 % of the total variance and were used to visualize the relationships among the six goose breeds. Nevertheless, in the second eigenvector (PC1/3), a moderate sample overlap of YE individuals within the WZ was observed, although the number of overlapping samples was not substantial. The six goose breeds were divided into three groups (Fig. 7B). Based on the combined results of PC1, PC2, and PC3 (Fig. 7D), the 90 geese were partitioned into four groups. The WZ, YE, and YJ geese constituted individual groups, whereas the BZ, LX, and XP were amalgamated to form a single group.

Analysis of the historical dynamics of population genetics

To investigate the domestication process and domestication time of

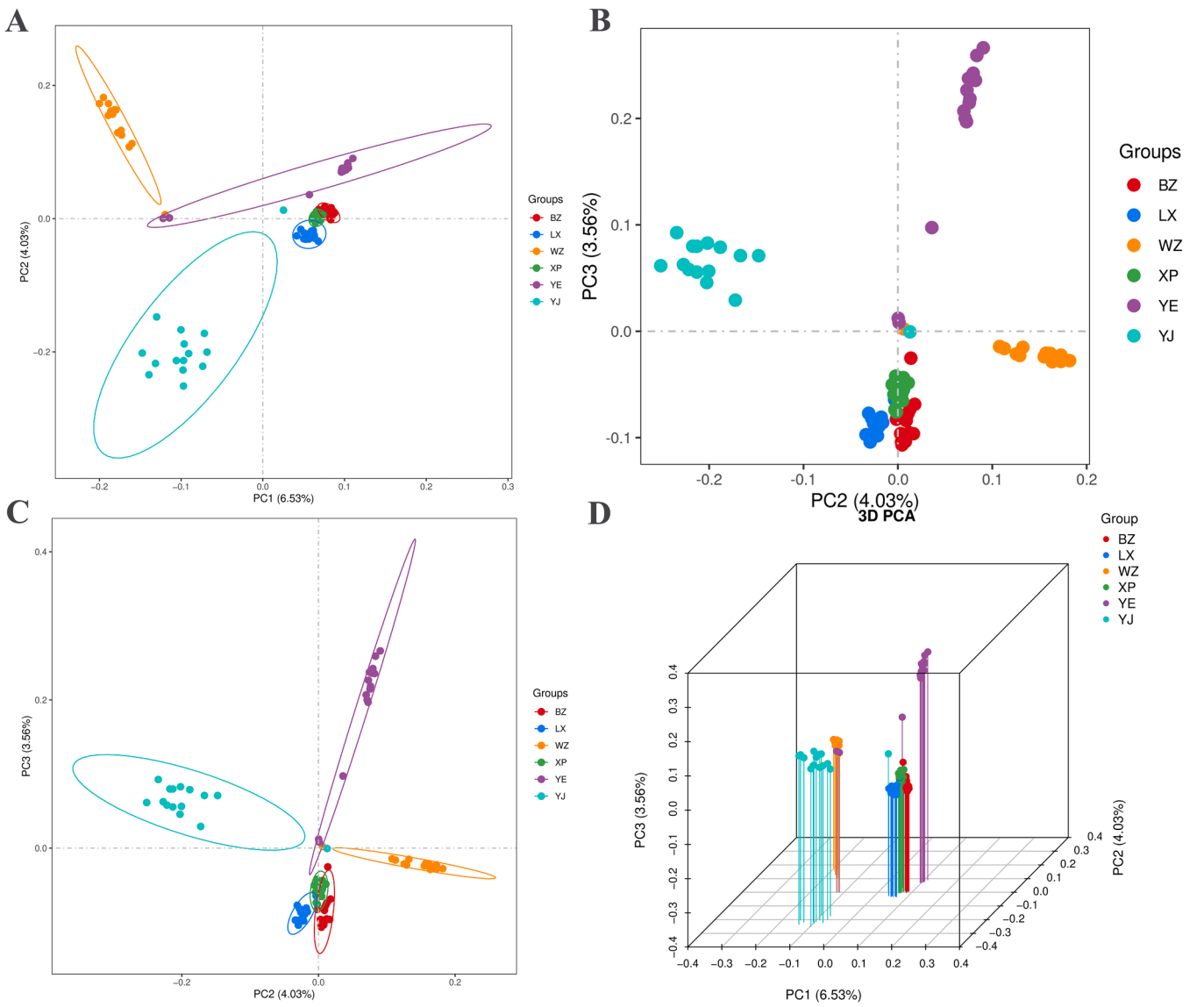

Fig. 7. Principal component analysis diagram of the six endangered goose breeds; A. Result analysis diagram of the first principal component (PCA1) and the second principal component (PCA2); B. Result analysis diagram of the first principal component (PCA1) and the third principal component (PCA3); C. Result analysis diagram of the second principal component (PCA2) and the third principal component (PCA3); D. Result analysis diagram of 3D principal components. " . ": The principal component clustering situation of the samples was acquired.

the endangered goose breeds, we estimated the sizes of the goose populations using PSMC (Fig. 8A). For nearly one million years, the effective population size (**Ne**) of the six endangered goose breeds has undergone periodic fluctuations and has been subject to two episodes of bottleneck effects. Following the Last Glacial Period (**LGP**), which occurred between 2.3×10^4 and 1.1×10^5 years ago, a relatively warm geological epoch, the Ne of the six breeds began a phase of continuous decline, followed by a progressive increase. During the Last Glacial Maximum (**LGM**) (1.5×10^4 and 2.1×10^4 years ago, a frigid geological period), the Ne value exhibited a precipitous decline, reaching its minimum. This nadir persisted until 10,000 years ago, when the Ne of the endangered goose breeds entered a period of stasis and reached a plateau. To examine the gene flow among diverse populations, YJ was selected as the root population for gene flow analysis. Upon performing the analysis with a model that precluded gene flow, the resultant maximum likelihood tree was analogous to the phylogenetic tree and outcomes of the PCA analysis (Fig. 8C). The YE, BZ, LX, and XP geese were situated on the same branch, whereas the genetic distance between YJ and the other goose varieties was comparatively significant. As shown in the heat map (Fig. 8B), the values corresponding to the colors among the different goose populations were above zero, suggesting multiple gene flows. Under the gene flow model, both the maximum and minimum values of the color scale on the heat map were reduced compared to the values when there was no gene flow, and there may have been a total of five gene flows among the goose populations (Fig. 8D). From the maximum likelihood tree (Fig. 8A), except for YJ, the five goose breeds were all under the same major branch under the gene flow model, and the YJ goose maintained a relatively long genetic distance from the other goose breeds.

In addition to other analyses, we calculated the Fst and gene flow (**Nm**) coefficients of the six populations and drew correlation maps (Fig. 9), which were intended to visually represent the relationships among the relevant genetic factors. Among these breeds, there was a certain degree of gene flow (Nm > 1). Population genetic exchanges transpired at a remarkable frequency (Nm = 11.1358 and 12.4183, respectively) between the BZ and XP and the LX and XP pairs of goose breeds. A plausible explanation is that interspecific hybridization ensued between these two pairs of populations, which may have led to

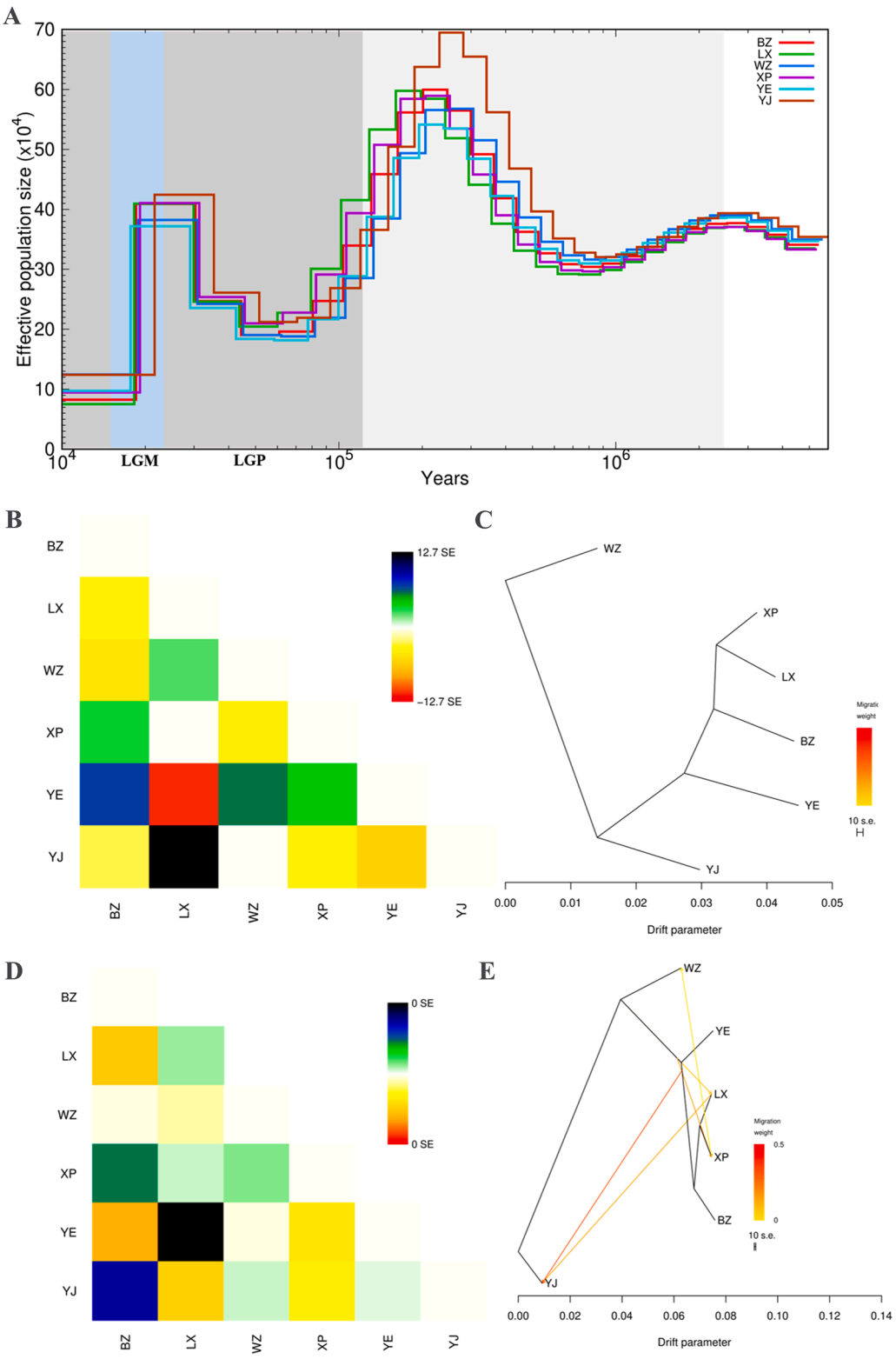

Fig. 8. A. PSMC model reveals the changes in the effective population size of endangered goose breeds in different historical periods; Examples of PSMC estimate changes in the effective population size over time, representing variation in inferred Ne dynamics. The lines represent inferred population sizes, and the gray shaded areas indicate the Pleistocene period, with the Last Glacial Period (LGP) shown in darker gray, and the Last Glacial Maximum (LGM) shown as light-blue areas. B. Residual fitting heat map of goose breeds under the no-gene-flow model; C. Phylogenetic tree of goose breeds under the no-gene-flow model. Events related to migration are represented as colored weighted arrows. The length of parallel branches is proportional to the degree of genetic drift occurring on each branch. The standard deviation shown on the scale bar is 10 times the average of the values in the sample covariance matrix. D. Maximum likelihood tree along migration events. E. Residual fitting heat map of goose breeds under the gene flow model.

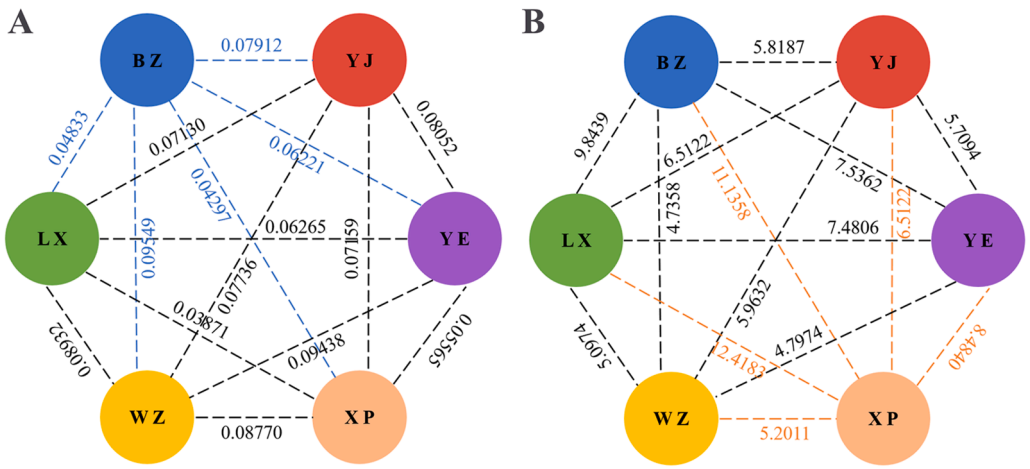

Fig. 9. A. Map of the genetic differentiation index (Fst) for the endemic endangered goose breeds; B. Map of the gene flow (Nm) for the endangered goose breeds.

the corresponding transformations in their genotypes. The Fst values among the diverse groups within the population fall within the range of 0.038-0.095, which aligns with the scope of the differentiation index. The extent of genetic differentiation among the BZ, LX, and XP goose breeds was minimal (Fst < 0.05). In contrast, there was a moderate level of genetic differentiation among the other three breeds (0.05 < Fst < 0.15).

Selective sweep analysis among populations

To accurately identify the genomic footprints left by natural and artificial selection forces during the domestication of endangered goose breeds, we conducted a selective sweep analysis based on the population Fst and θπ. BZ geese were designated as the control group. Using a region with a window size of 100 K and a step length of 10 K, we performed a genome-wide selection signal analysis for the breeds within this specific region. The top 1 % of the genomic regions emerging within the ROH fragments were considered the threshold (black line) for screening high-frequency ROH regions. Subsequently, unique high-frequency ROH fragments within the selected population were retained for further investigation and analysis. A total of 1,581 candidate genes were identified and isolated through the Fst screening criteria, while 1,683 candidate genes were likewise singled out from θπ (Fig. 10A and B; Tables S5 and S6). After identifying the overlapping genes using the two statistical methods, we obtained a cumulative total of 394 overlapping selective genes, which were deemed candidate genes for further analysis (Fig. 10C and D). In parallel, we discovered 425 overlapping selective genes between the BZ and WZ geese (**Fig. S3; Tables S9-S11**); 332 overlapping selective genes between the BZ and XP geese (**Fig. S4; Tables S12-S14**); 378 overlapping selective genes between the BZ and YE geese (**Fig. S5; Tables S15-S17**); and 401 overlapping selective genes between the BZ and YJ geese (**Fig. S6; Tables S18-S20**).

Through comprehensive analysis, following the standardization and correlation of the Fst and θπ, 744 selected regions were meticulously screened and identified on the autosomes. The gene function annotation analysis implemented on BZ-LX revealed that 154 PCGs identified through GO function enrichment analysis were conspicuously enriched within 249 GO terms ($P < 0.05$). These GO terms comprised included 86 cellular components, 94 molecular functions, and 69 biological processes. The principal foci of these genes were centered on processes such as receptor binding, transportation, modulation of responses to stimuli, negative regulation of cellular processes, and regulation of biological mass. Twenty terms were significantly enriched in 20 terms ($P < 0.05$), of which eight had extremely strong significance. We observed that 37 PCGs were primarily used for general functional prediction of physiological processes (Fig. 10E and Table S7). An in-depth examination of the overlapping genes identified using at least two selection methods revealed a cluster of genes associated with crucial traits. These traits encompass health through specific genes, including *CEP95, APOB, CCDC158, DHX29, MTREX, MTUS1, LOC106043735, NUDT12, XPA* and *LOC125181894*. Moreover, genes relevant to growth and development, such as *CCND1, DES, CCNO, SMC5,* and *NUBP1* were also detected. Additionally, genes related to immunity, namely *ABCA2, ABCC8, UHRF2,* and *ABCA1*, as well as those related to fat deposition, *LOC106039054* and *ACHE*, and the *KAT6B* gene associated with organism aging, were all part of this significant discovery. KEGG functional enrichment analysis demonstrated that the PCGs were predominantly and significantly enriched in 23 terms ($P < 0.05$) (**Table S8**). Overall, the PCGs are mainly involved in the regulation of translation, ribosomal structure and biogenesis (such as methionine-tRNA ligase activity [GO:0004825]; ATP binding [GO:0005524] and methionyl-tRNA aminoacylation [GO:0006431], replication, recombination and repair (lipid transport [GO:0006869]); posttranslational modification, protein turnover, chaperones (Biological Process: ubiquitin-dependent ERAD pathway [GO:0030433]; proteasome-mediated ubiquitin-dependent protein catabolic process [GO:0043161] and biological process: pronephric nephron tubule development [GO:0039020]), amino acid transport and metabolism and inorganic ion transport and metabolism. This may have a causal relationship with the specific phenotypic characteristics that developed during the domestication process.

Discussion

Genetic diversity is a key index for evaluating genetic resources, and population structure constitutes the basis for dissecting the genetic relationships and origins among diverse breeds. Exploring the population structure and genetic diversity of endemic endangered geese in China is beneficial for appraising their poultry genetic resources and plays a significant role in safeguarding and capitalizing on trait enhancement in endangered goose breeds. This study focused on six endangered goose breeds, all local breeds unique to central China. Currently, there are relatively few reports on the genetic structure and diversity of locally endangered geese, based on WGS data. Therefore, this study used WGS to conduct an in-depth exploration of the genetic structure and genomic diversity of endemic endangered goose breeds.

Population genetic diversity

Genetic diversity is paramount for the adaptability and survival of all species. Higher genetic diversity implies a better ability to adapt to environmental changes and cope with various pressures (Hameed et al.,

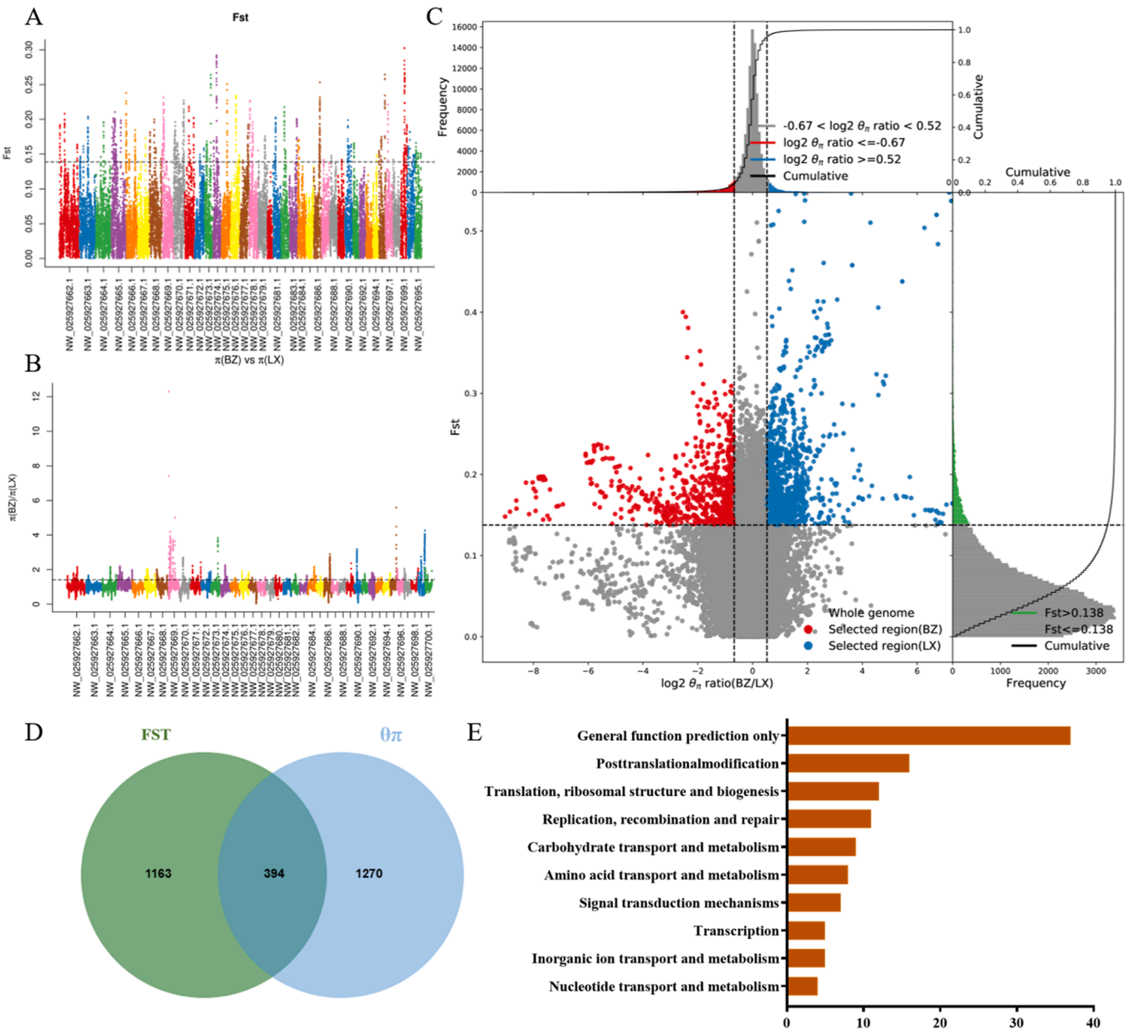

Fig. 10. Selection signal and functional enrichment analysis of the endangered BZ-LX goose breed. A. Manhattan plot based on the top 1 % threshold of the Fst ratio. This plot visually represents the distribution of genetic variation in terms of the Fst ratio, with the top 1 % values being of particular significance in identifying regions potentially under selection. B. Manhattan plot based on the top 1 % threshold of the θπ ratio. Similarly, this plot depicts the θπ ratio distribution, and the top 1 % threshold serves as a crucial indicator for detecting genomic regions with differential variation patterns. C. Target regions identified through Fst and θπ selective scanning. The Fst and θπ ratios are precisely plotted along the X-axis and Y-axis respectively. The horizontal and vertical gray dotted lines demarcate the top 1 % values of the Fst (0.138) and θπ ratio (0.52), which are key reference points for screening significant regions. The red dots denote the top 5 % regions of group 1 that are under selection in the PI analysis and also possess a top 5 % Fst value. The blue dots correspond to the top 5 % regions of group 2 under selection and with a top 5 % Fst. The green regions represent the top 5 % of regions selected by the Fst. D. Venn diagram of overlapping genes within the selective scanning regions. This diagram effectively illustrates the intersection and overlap of genes identified by the different selective scanning methods. E. Results of functional enrichment analysis of candidate genes.

2022). With the continuous advancement of sequencing technologies, a multitude of methods have emerged to directly computer inbreeding coefficients based on whole-genome SNP data incessantly (Machová et al., 2022). ROH estimation is an efficacious instrument that plays a pivotal role in probing the genetic diversity of populations, providing statistical appraisals of historical population information, and identifying potential genomic structures (Shi et al., 2023). ROH can reveal regions with adverse phenotypic effects in the case of homozygosity and can detect associations between economic traits and genes in these regions (Szmatoła et al., 2019). Furthermore, selection pressure and unregulated breeding management of endangered goose genomes may influence SROH. Forutan et al. (2018) reported a high correlation between the inbreeding coefficient calculated based on the ROH and that calculated using pedigree records. In our study, YJ and WZ geese had the lowest inbreeding coefficients (Fis = 0.108 and 0.132, respectively). For BZ and XP, the SROH was mainly less than 2Mb, potentially indicating a higher accumulation of harmful mutations. Relatively short ROHs are highly likely to be correlated with ancestral genetics or potential bottlenecks (Purfield et al., 2012). In contrast, the longer ROHs in the WZ were more likely associated with relatively recent inbreeding activities.

Thus, inbreeding small populations on a specific scale boosts purifying selection, facilitating the removal of harmful mutations. The ratio of transitions to transversions within the gene regions of the filtered SNPs set was observed to vary from 2.24 to 2.49, with all the transition/-transversion values exceeding 1.5. This suggests that, similar to most vertebrates, endangered goose breeds exhibit a transition-type skew. This is consistent with the findings of Yan et al. (2024) who mined the characteristic loci of **Luokeng chickens** using WGS data.

Heterozygosity is a crucial parameter for assessing the genetic diversity in populations. As the heterozygosity of a population increases, the variety encompasses more genetic information. Expected heterozygosity represents the probability that any individual within a population shows a heterozygous state at a particular locus. The observed heterozygosity denotes the proportion of individuals in a population that displays a heterozygous state at a specific locus relative to the total number (Mastrangelo et al., 2018; Dzomba et al., 2021). In the current study, we observed that the expected He of endangered goose breeds was significantly higher than the observed Ho (He > Ho), suggesting that inbreeding events may have occurred within the population or that the population was affected by natural selection. Abdel-Kafy and his colleagues, during their research on **Egyptian geese** (*Anser anser*), detected an expected He value of 0.352 (Abdel-Kafy et al., 2021). This value was considerably higher than the average expected He (He = 0.280) of the endangered Chinese goose breeds (*Anser cygnoides*). This finding implies that the overall genetic diversity of endangered Chinese goose breeds is relatively low, and that there is a risk of loss of species genetic diversity.

Genetic structure and ancestral domestication

Analyzing the genetic structure and diversity of populations by leveraging whole-genome SNP data is an efficient approach for probing and assessing the circumstances of locally bred populations. The NJ-tree, admixture analysis, and PCA revealed a common clustering pattern (Paskov et al., 2023). In this study, the NJ-tree and admixture analysis showed that WZ and YJ shared the same bloodline origin and might have a common ancestor. Many individuals, particularly male geese, possessed only one ancestral component. Consequently, emphasis should be placed on these factors in subsequent conservation and utilization procedures. In contrast, the other three breeds have complex genetic backgrounds. This implies that during their long-term rearing processes, they might have experienced relatively frequent hybridization and improvement due to the introduction of foreign goose breeds (*Anser anser*). Wen et al. (2023) observed gene flow between domestic geese and their wild ancestors in a study revealing the origin and evolution of Chinese domestic geese (*Anser cygnoides*) and mentioned bloodline intrusion events in other goose breeds. The integrated outcomes of PC1, PC2, and PC3 accounted for 14.12 % of the total variance and were applied to a methodological analysis of the relationships among endangered goose breeds. The BZ, LX, and XP geese samples overlapped and were categorized into a single group. These three goose breeds may have experienced relatively frequent hybridization and improvement. The PCA results agreed with the admixture analysis. However, in a study on endemic goose breeds in Guangdong**,** WZ and YJ could be clustered into a small group of goose breeds (Lin, 2023). This may be related to specific genetic characteristics, historical evolution, and environmental adaptability of the group. These results reflect the complexity and diversity of the different endangered geese groups during their evolution. Based on the data acquired in the current study, we determined that the endangered goose breeds could be partitioned into three subpopulations. Such an association is presumably ascribable to their shared distribution within plains regions characterized by analogous topographical and climatic circumstances (Li et al., 2011).

Effective population size is an important parameter in population genetics, as it reflects historical changes and genetic diversity variations in a population (Honka et al., 2022). To explore the evolutionary process and domestication time of endangered goose breeds, we used PSMC software to estimate the size of goose populations dating back to one million years. The principal Chinese endangered goose breeds have experienced two substantial alterations in population size. These changes corresponded to specific geological periods. The first large-scale population shift occurred 100,000 years ago. In the Pleistocene epoch (denoted by the grey shaded region, which was a relatively warm geological period), the effective population number manifested periodic oscillations. Population size initially increased in a stepwise manner and subsequently declined. A likely explanation for this is that geese, as warm-blooded creatures, can adapt more effectively to climate alterations within a relatively narrow range of fluctuations. The second large-scale population change happened 10,000 years ago. With the advent of the LGM (a frigid geological period), the effective population number of goose breeds declined sharply and then entered a plateau phase. Once the population size drops below a plateau threshold, which is lower than the level required for the self-sustainability of the population, they quickly become extinct. Changes in the Earth's environment have profound effects on goose populations. Studies have reported that the LGM has led to the extinction of many mammalian species (Nie et al., 2008).

In conclusion, the effective population size of each species constantly undergoes periodic fluctuations due to climate, food supply, and other factors. Based on the results of the TreeMix analysis, gene flow exists among all the endangered goose breeds. In an experiment in which low-coverage whole-genome sequencing was used to study the population genetic structure of *Setophaga ruticilla*, genome-wide analysis of Fst showed that weak genetic differentiation among populations was interrupted by regions with elevated genetic differentiation (DeSaix et al., 2023). In this study, a significant degree of extensive gene flow (Nm = 11.1358 and 12.4183) was detected between BZ and XP geese and between LX and XP geese, respectively. The extent of genetic contribution differed among individual breeds. More precisely, a certain portion of YJ geese exhibited gene flow originating from LX geese. Similarly, a segment of WZ geese demonstrated gene flow stemming from XP geese. Admixture analysis identified an intermixture of lineages from other goose breeds. This could potentially be attributed to the historical scenario in which, in antiquity, endemic goose breeds were predominantly reared in a free-range fashion by individual farmers. Because of commercial trading activities, geese from diverse regions are transported and sold to other localities. They are subsequently acquired by farmers and incorporated into endemic goose breeds, facilitating crossbreeding and mixed husbandry practices (Heikkinen et al., 2020). On the one hand, this has given rise to complex gene flow among endemic goose breeds. However, foreign genetic introgression has emerged to boost economic benefits. Endemic goose breeds in certain regions have been gradually supplanted by hybrid varieties. Moreover, their breeding numbers have decreased significantly, to the extent that they have almost vanished.

Selection signatures of candidate genes

Genetic differentiation is caused by long-term natural and artificial selection and is influenced by multiple factors such as mating, geographical isolation, and gene flow. Selective sweep analysis can explore selected regions within a population and subsequently identify genes under selection, thereby accounting for the adaptive mechanisms underlying population evolution and domestication at the molecular level. A selective sweep analysis was conducted on the genomes of endangered goose breeds. Compared to their wild ancestors, the behavior of waterfowl has changed significantly, and this transformation has evolved during the domestication process (Jensen, 2014). Ochoa and Storey (2021) have claimed that the adaptive evolution of a population to a specific environment increases population differentiation. When the Fst value was greater than 0.25, the degree of genetic differentiation among the populations was even greater. In the current study, the Fst values among the six breeds were less than 0.25, suggesting a

comparatively low degree of genetic differentiation among the endangered goose breeds. This phenomenon can be attributed to the varying intensities and directions of the local selection. Consequently, these breeds display remarkable genetic disparities in appearance and production capabilities. Zhu et al. (2024) carried out an assessed the population Fst and π ratios through by means of selection scan analysis based on SNP data. They then identified five functional genes, *AQP3, PIK3C3, NOL6, RPP25* and *DCTN3*, that are potentially associated with crucial economic traits in laying ducks. These traits include encompass vasopressin-regulated water reabsorption, ribosome biogenesis, and the PI3K signaling pathway. Feng (2018) identified 17 common domestication genes in populations of **Shaoxing, Shanma** and **Cherry Valley ducks**, which are closely related to physiological activities such as cardiovascular development, smooth muscle proliferation, and insulin function regulation. Based on the results of our study (BZ-LX), 154 PCGs were identified and screened. These PCGs were prominently enriched in 20 GO functional pathways. The main areas of enrichment encompassed "Posttranslational modification, protein turnover, chaperones," "Translation, ribosomal structure and biogenesis," "Replication, recombination and repair," "Carbohydrate transport and metabolism," as well as "Amino acid transport and metabolism." Furthermore, these genes were significantly enriched in 23 terms within KEGG functional enrichment pathways. The main areas of enrichment encompassed "Posttranslational modification, protein turnover, chaperones," "RNA processing and modification," "Posttranslational modification, protein turnover, chaperones," "vesicular transport," as well as "Intracellular trafficking, secretion," among others. These signaling pathways provide a theoretical underpinning for the purpose of effecting the directional selection of genes that have been selected and are associated with significant economic traits of endangered goose breeds in the context of prospective breeding endeavors.

Cell proliferation is the fundamental basis for organismal growth and development. Notably, among the selection signals of the different goose breeds, we identified five genes, namely *CCND1, DES, CCNO, SMC5*, and *NUBP1*, that are actively involved in regulating the biological process of cell proliferation and that play crucial roles in cell cycle control. Several studies have indicated that overexpression or dysfunction of the *CCND1* gene is typically associated with the progression of multiple types of cancers, including breast cancer and lymphoma (Wang et al., 2024). *SMC5* may be involved in the cellular response to certain viral infections and assists cells in resisting viral invasion (O'Leary et al., 2024). During the screening of PCGs in endemic endangered goose breeds, we identified that the selected gene (*KAT6B*) was mostly related to animal body aging and regeneration, indicating that endangered goose breeds are under strong selection in terms of body senescence and lifespan traits. Histone lysine acetyltransferase *KAT6B (MYST4, MORF, QKF)* is a target of recurrent chromosomal translocations that cause hematological malignancies with a poor prognosis. *KAT6B* gene deletion results in a cell-intrinsic gene-dose-dependent multilineage hematopoietic defect, including impaired multilineage hematopoietic reconstitution in lethally irradiated transplant recipients (Bergamasco et al., 2024). Additionally, we identified and screened candidate genes, namely *UHRF2, ABCA2, ABCC8 and ABCA1*, that endangered goose breeds and are implicated in animal immune regulation and cancer processes. *UHRF2*, a ubiquitin-like protein 2 with PHD and RING finger domains, encodes a protein that participates in multiple cellular processes, with particular emphasis on the modulation of gene expression and DNA methylation (Li et al., 2024). This gene may influence cancer progression by regulating key tumor suppressor genes and oncogenes, and is being investigated as a potential therapeutic target. Increasing evidence has suggested that *UHRF2* is associated with the development and occurrence of various types of cancer. Numerous studies have shown that *UHRF2* modulates histones via epigenetic alteration of DNA methylation, thereby potentiating breast cancer oncogenesis. The molecular interaction between *UHRF2* and the transcription factor E2F1 triggers apoptosis in lung cancer cells (Luo et al., 2013). Additionally, *UHRF2* augments tumorigenesis in liver cancer cells by stabilizing the acetyltransferase TIP60 and regulating the acetylation of lysine residues at positions 9 and 14 of histone H3, which are crucial epigenetic modification implicated in transcriptional regulation and chromatin remodeling (Zeng et al., 2017). These findings provide genetic information for further research on traits such as genetic reproduction and disease resistance in endemic endangered goose breeds.

Conclusion

This study examined the genetic diversity, genetic associations, population architecture, and selection traits of six endangered geese breeds. These findings suggest that endemic endangered goose breeds harbor genetic diversity and exhibit remarkable utilization potential. There is an urgent need to implement effective conservation measures. Furthermore, genome-wide selective sweep analysis has provided profound insights into assorted evolutionary scenarios in the natural environment and artificial selection. In total 154 PCGs were identified as candidate genes that modulate essential growth and development processes in endangered goose breeds. Therefore, this study provides a deeper understanding of the genetic structure differences among endangered goose populations in local areas and the candidate genes selected under selective pressure. This will be conducive to the development of more effective strategies and measures for the conservation and utilization of goose germplasm resources.

Funding

This work was supported by the National Key Research and Development Program of China [2021YFD1200302], Key and General Projects of Modern Agriculture in Jiangsu Province [BE2022350] and a Jiangsu Province Seed Industry Revitalization Announcement and Leading Project [JBGS [2021] 023].

Declaration of competing interest

None of the authors involved in this research have any conflicts of interest to report.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psj.2025.105004.

References

Abdel-Kafy, E.M., Ramadan, S.I., Ali, W.H., Youssef, S.F., Shabaan, H.A., El-Deighadi, A., Inoue-Murayama, M., 2021. Genetic and phenotypic characterization of domestic geese (*Anser anser*) in Egypt. Animals (Basel) 11, 3106–3110.

Bergamasco, M.I., Ranathunga, N., Abeysekera, W., Li-Wai-Suen, C.S.N., Garnham, A.L., Willis, S.N., McRae, H.M., Yang, Y., D'Amico, A., Di Rago, L., Wilcox, S., Nutt, S.L., Alexander, W.S., Smyth, G.K., Voss, A.K., Thomas, T., 2024. The histone acetyltransferase *KAT6B* is required for hematopoietic stem cell development and function. Stem. Cell. Rep. 19, 469–485.

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., Wu, Y., Zhao, L., Liu, J., Guo, J., Fang, S., Cao, W., Yi, L., Zhao, Y., Kong, L., 2021. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic. Acids. Res. 49, 317–325.

Chen, C., Li, J., Li, Z., Nong, Y., Wang, J., Wang, Z., Li, Z., 2024. Whole-genome resequencing reveals melanin deposition candidate genes of *Luning* chicken. Bmc Genomics [Electronic Resource] 25, 858–860.

Chen, G.H., Liu, J.L, Xu, Q., 2022. The current situation and prospects of the protection and utilization of goose genetic resources in China. China Poultry Ind. Guide 39, 6–12 (in Chinese).

Cho, Y., Kim, J.Y., Kim, N., 2022. Comparative genomics and selection analysis of *Yeonsan Ogye black* chicken with whole-genome sequencing. Genomics 114, 110298.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80–92.

DeSaix, M.G., Anderson, E.C., Bossu, C.M., Rayne, C.E., Schweizer, T.M., Bayly, N.J., Narang, D.S., Hagelin, J.C., Gibbs, H.L., Saracco, J.F., Sherry, T.W., Webster, M.S.,

Smith, T.B., Marra, P.P, Ruegg, K.C., 2023. Low-coverage whole genome sequencing for highly accurate population assignment: mapping migratory connectivity in the American Redstart (*Setophaga ruticilla*). Mol. Ecol. 32, 5528–5540.
Dzomba, E.F., Chimonyo, M., Pierneef, R., Muchadeyi, F.C., 2021. Runs of homozygosity analysis of South African sheep breeds from various production systems investigated using OvineSNP50k data. Bmc Genomics [Electronic Resource] 22, 7–12.
Feng, P.S., 2018. Whole Genome Resequencing and Transcriptome Sequencing Reveal the Molecular Basis of Duck Population Structure and Domestication Adaptation, 14. South China Agricultural University, pp. 1142–1151 (in Chinese).
Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F.S., Sargolzaei, M., 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. Bmc Genomics [Electronic Resource] 19 (1), 98–108.
Gu, J., Li, S., Zhu, B., Liang, Q., Chen, B., Tang, X., Chen, C., Wu, D.D., Li, Y., 2023. Genetic variation and domestication of horses revealed by 10 chromosome-level genomes and whole-genome resequencing. Mol. Ecol. Resour. 23, 1656–1672.
Gu, L., Wang, F., Lin, Z., Xu, T., Lin, D., Xing, M., Yang, S., Chao, Z., Ye, B., Lin, P., Hui, C., Lu, L., Hou, S., 2020. Genetic characteristics of Jiaji Duck by whole genome re-sequencing. PLoS One 15, e0228964.
Hameed, A., Schlecht, E., Tariq, M., Buerkert, A., Scheper, C., König, S., Roessler, R., 2022. Phenotypic and genetic diversity of domestic yak (*Bos grunniens*) in high-altitude rangelands of Gilgit-Baltistan, Pakistan. J. Anim. Breed. Genet. 139, 723–737.
Heikkinen, M.E., Ruokonen, M., White, T.A., Alexander, M.M., Gündüz, İ., Dobney, K.M., Aspi, J., Searle, J.B, Pyhäjärvi, T., 2020. Long-term reciprocal gene flow in wild and domestic geese reveals complex domestication history. G3 10, 3061–3070.
Honka, J., Baini, S., Searle, J.B., Kvist, L., Aspi, J., 2022. Genetic assessment reveals inbreeding, possible hybridization, and low levels of genetic structure in a declining goose population. Ecol. Evol. 12, e8547.
Hudson, R.R., Slatkin, M., Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132, 583–589.
Jensen, P., 2014. Behavior genetics and the domestication of animals. Annu Rev Anim Biosci 2, 85–104.
Li, H.F., Zhu, W.Q., Chen, K.W.HY., Xu, W.J., Song, W., 2011. Two maternal origins of Chinese domestic goose. Poult. Sci. 90, 2705–2710.
Li, Z., Xue, Y., Huang, X., Xiao, G., 2024. Stratifying osteosarcoma patients using an epigenetic modification-related prognostic signature: implications for immunotherapy and chemotherapy selection. Transl. Cancer Res 13, 3556–3574.
Lin, Y., 2023. Genetic study of local goose population in Guangdong Province based on whole genome resequencing data. Zhong-kai College Agric. Eng. 16, 27–31 (in Chinese).
Luo, T., Cui, S., Bian, C., Yu, X., 2013. Uhrf2 is important for DNA damage response in vascular smooth muscle cells. Biochem. Biophys. Res. Commun. 441, 65–70.
Machová, K., Marina, H., Arranz, J.J., Pelayo, R., Rychtářová, J., Milerski, M., Vostrý, L., Suárez-Vega, A., 2022. Genetic diversity of two native sheep breeds by genome-wide analysis of single nucleotide polymorphisms. Animal 17, 100690.
Mastrangelo, S., Ciani, E., Sardina, M.T., Sottile, G., Pilla, F., Portolano, B., Bi, O.v., 2018. Ita Consortium. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. Anim. Genet. 49, 71–81.
Mohammadifar, A., Faghih Imani, S.A., Mohammadabadi, M.R., Soflaei, M., 2014. The effect of *TGFb3* gene on phenotypic and breeding values of body weight traits in Fars native fowls. Agric. Biotech. J 5, 125–136.
Moazeni, S., Mohammadabadi, M.R, Sadeghi, M., 2016. Association between UCP gene polymorphisms and growth, brreeding value of growth and reproductive traits in Mazandaran indigenous chicken. Open J. Anim. Sci. 6, 1–8.
Mohamadinejad, F., Mohammadabadi, M., Roudbari, Z., Eskandarynasab, S.S., Babenko, O., Klopenko, N., Borshch, O., Starostenko, I., Kalashnyk, O., Assadi, S.E., 2024. Analysis of liver transcriptome data to identify the genes affecting lipid metabolism during the embryonic and hatching periods in ROSS breeder broilers. Journal of Livestock Science and Technologies 12, 61–67.
Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. P. Natl. Acad. Sci. USA 76, 5269–5273.
Nie, Z.S., Li, H., Ma, B.Q., 2008. Late pleistocene fossil fauna in the Hetao Basin, Inner Mongolia. Quaternary Studies 22, 14–25 (in Chinese).
Ochoa, A., Storey, J.D., 2021. Estimating FST and kinship for arbitrary population structures. PLos Genet. 17, 1009241.
O'Leary, D.R., Hansen, A.R., Fingerman, D.F., Tran, T., Harris, B.R., Hayer, K.E., Fan, J., Chen, E., Tennakoon, M., DeWeerd, R.A., Meroni, A., Szeto, J.H., Weitzman, M.D., Shalem, O., Bednarski, J., Vindigni, A., Zhao, X., Green, A.M., 2024. The SMC5/6 complex prevents genotoxicity upon APOBEC3A-mediated replication stress. Biorxiv 24, 568952.
Ouyang, J., Zheng, S., Huang, M., Tang, H., Qiu, X., Chen, S., Wang, Z., Zhou, Z., Gao, Y., Xiong, Y., 2022. Chromosome-level genome and population genomics reveal evolutionary characteristics and conservation status of Chinese indigenous geese. Commun. Biol. 7, 1187–1191.
Paskov, K., Chrisman, B., Stockham, N., Washington, P.Y., Dunlap, K., Jung, J.Y, Wall, D. P., 2023. Identifying crossovers and shared genetic material in whole genome sequencing data from families. Genome Res. 33, 1747–1756.
Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G., 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13, 70–85.
Qi, S., Fan, S., Li, H., He, Y., Zhang, Y., Zhao, W., Xu, Q, and Chen, G. 2024. Analysis of genetic diversity and population structure of endemic endangered goose (*Anser cygnoides*) breeds based on mitochondrial *CYTB*. Animals (Basel). 14:1480-1488.
Qu, X.Y., Peng, Y.D, He, J., 2016. Comparative study on growth curve and production performance of *Daozhou Grey* Goose × *Lande* Goose Hybrid F_1 generation and *Lande* Goose. Chinese Poultry. 38, 9–14 (in Chinese).
Shi, H., Li, T., Su, M., Wang, H., Li, Q., Lang, X., Ma, Y., 2023. Whole genome sequencing revealed genetic diversity, population structure, and selective signature of *Panou* Tibetan sheep. Bmc Genomics [Electronic Resource] 24, 50–58.
Szmatoła, T., Gurgul, A., Jasielczuk, I., Ząbek, T., Ropka-Molik, K., Litwińczuk, Z., Bugno-Poniewierska, M., 2019. A comprehensive analysis of runs of homozygosity of eleven cattle breeds representing different production types. Animals (Basel) 9, 1024–1029.
Tscharntke, T., Batáry, P., Grass, I., 2024. Mixing on- and off-field measures for biodiversity conservation. Trends Ecol. Evol. 39, 726–733.
Wang, M.S., Thakur, M., Peng, M.S., Jiang, Y., Frantz, LAF., Li, M., Zhang, J.J., Wang, S., Peters, J., Otecko, N.O., Suwannapoom, C., Guo, X., Zheng, Z.Q., Esmailizadeh, A., Hirimuthugoda, N.Y., Ashari, H., Suladari, S., Zein, MSA., Kusza, S., Sohrabi, S., Kharrati-Koopaee, H., Shen, Q.K., Zeng, L., Yang, M.M., Wu, Y.J., Yang, X.Y., Lu, X. M., Jia, X.Z., Nie, Q.H., Lamont, S.J., Lasagna, E., Ceccobelli, S., Gunwardana, H.G.T. N., Senasige, T.M., Feng, S.H., Si, J.F., Zhang, H., Jin, J.Q., Li, M.L., Liu, Y.H., Chen, H.M., Ma, C., Dai, S.S., Bhuiyan, A., Khan, M.S., Silva, G., Le, T.T., Mwai, O.A., Ibrahim, M., Supple, M., Shapiro, B., Hanotte, O., Zhang, G., Larson, G., Han, J.L., Wu, D.D, Zhang, Y.P., 2020. 863 genomes reveal the origin and domestication of chicken. Cell. Res. 30, 693–701.
Wang, W., Li, D., Xu, Q., Cheng, J., Yu, Z., Li, G., Qiao, S., Pan, J., Wang, H., Shi, J., Zheng, T., Sui, G., 2024. G-quadruplexes promote the motility in MAZ phase-separated condensates to activate *CCND1* expression and contribute to hepatocarcinogenesis. Nat. Commun. 15, 1045–1051.
Watermeyer, K.E., Guillera-Arroita, G., Bal, P., Burgass, M.J., Bland, L.M., Collen, B., Hallam, C., Kelly, L.T., McCarthy, M.A., Regan, T.J., Stevenson, S., Wintle, B.A, Nicholson, E., 2021. Using decision science to evaluate global biodiversity indices. Conserv. Biol. 35, 492–501.
Wen, J., Li, H., Wang, H., Yu, J., Zhu, T., Zhang, J., Li, X., Jiang, Z., Ning, Z., Qu, L., 2023. Origins, timing and introgression of domestic geese revealed by whole genome data. J. Anim. Sci. Biotechnol. 14, 20–26.
Yang, W., Liu, Z., Zhao, Q., Du, H., Yu, J., Wang, H., Liu, X., Liu, H., Jing, X., Yang, H., Shi, G., Zhou, L., Liu, J., 2022. Population genetic structure and selection signature analysis of Beijing black pig. Front Genet. 13, 860669.
Yan, X., Mo, D.I., Kong, S., 2024. Genetic structure analysis and feature locus mining of *Luokeng* chicken based on whole genome resequencing. Chinese J. Animal Husbandry 12, 17–22 (in Chinese).
Yin, H., Li, D., Wang, Y., Zhu, Q., 2019. Whole-genome resequencing analysis of *Pengxian Yellow* Chicken to identify genome-wide SNPs and signatures of selection. Bio. Tech 9, 383–390.
Zeng, S., Wang, Y., Zhang, T., Bai, L., Wang, Y., Duan, C., 2017. E3 ligase *UHRF2* stabilizes the acetyltransferase TIP60 and regulates H3K9ac and H3K14ac via RING finger domain. Protein. Cell. 8, 202–218.
Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., Yang, T.L., 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35, 1786–1788.
Zhang, S., Yao, Z., Li, X., Zhang, Z., Liu, X., Yang, P., Chen, N., Xia, X., Lyu, S., Shi, Q., Wang, E., Ru, B., Jiang, Y., Lei, C., Chen, H., Huang, Y., 2022. Assessing genomic diversity and signatures of selection in Pinan cattle using whole-genome sequencing data. Bmc Genomics [Electronic Resource] 23, 460–466.
Zhu, Z., Lin, R., Zhao, B., Shi, W., Cai, Q., Zhang, L., Xin, Q., Li, L., Miao, Z., Zhou, S., Huang, Z., Huang, Q., Zheng, N., 2024. Whole-genome resequencing revealed the population structure and selection signal of 4 indigenous Chinese laying ducks. Poult. Sci. 103, 103832.