

Genomic signatures reveal selection in Lingxian white goose

Sumei Zheng,^{*,†} Jing Ouyang,^{*} Siyu Liu,^{*} Hongbo Tang,^{*} Yanpeng Xiong,^{*} Xueming Yan,^{*} and Hao Chen ^{*,1}

^{*}College of Life Science, Jiangxi Science and Technology Normal University, Nanchang, 330013, China; and [†]Fujian Vocational College of Agriculture, Fuzhou, 360119, China

ABSTRACT Lingxian white goose (**LXW**) is a goose breed indigenous to China that is famous for its meat quality and fast growth. However, the genomic evidence underlying such excellent breeding characteristics remains poorly understood. Therefore, we performed whole-genome resequencing of 141 geese from 3 indigenous breeds to scan for selection signatures and detect genomic regions related to breed features of LXW. We identified 5 reproduction-related genes (*SYNE1*, *ESR1*, *NRIP1*, *CCDC170*, and

ARMT1) in highly differentiated regions and 11 notable genes in 26 overlapping windows, some of which are responsible for meat quality (*DHX15*), growth traits (*LDB2*, *SLIT2*, and *RBPJ*), reproduction (*KCNIP4*), and unique immunity traits (*DHX15* and *SLIT2*). These findings provide insights into the genetic characteristics of LXW and identify genes affecting important traits in LXW, which extends the genetic resources and basis for facilitating genetic improvement in domestic geese breeds.

Key words: poultry, Lingxian white goose, selection signature, breeding

2023 Poultry Science 102:102269

<https://doi.org/10.1016/j.psj.2022.102269>

INTRODUCTION

Goose is an economically important agricultural animal. As one of the most important poultry, geese have been domesticated by humans for approximately 7,000 yr (Eda et al., 2022), which has led to a series of behavioral, morphological, and physiological changes. Artificial selection has been performed to improve agriculturally important traits, resulting in a variety of local breeds with different appearances and attributes. Among these geese, Lingxian white goose (**LXW**) is native to Yanling County, Hunan Province. It is a small, domesticated geese breed in China, with white feathers, orange beaks, and webs. Although small in body size, it shows rapid growth, precocious maturity and fattening, and good meat quality, as well as many excellent characteristics, such as resistance to rough feeding, strong adaptability, and disease resistance (Chen et al., 2011).

These breed characteristics are the result of natural and artificial selection pressures, and selective sweeps usually leave a footprint on the genome, leading to long haplotypes, high-frequency derived alleles, and highly

differentiated alleles (Grossman et al., 2010). Identification of selection signatures at the whole-genome level is possible through the development of high-throughput sequencing technology. Recently, various methods have been applied to explore population selection signals. For example, Zhou et al. (2018) used F-statistics (**Fst**) and nucleotide diversity (π) methods to determine that *MITF* regulates the formation of white plumage in Pekin ducks. Jeong et al. (2015) employed cross-population-extended haplotype homozygosity (**XP-EHH**) and cross-population composite likelihood ratio test approaches to detect multiple genes that contribute to intramuscular fat in Berkshire pigs, which might affect the quality of pork. Bortoluzzi et al. (2020) found that the *PITX1* and *TBX5* genes are strongly associated with foot feathering in domestic chickens using a combination of genome-wide association studies and Z-transformed heterozygosity approaches. Although these selective sweeps have been widely used in breeding research and have accelerated the progress of genetic improvement, research on LXW is still lacking.

The physiological characteristics of LXW, an excellent goose breed in China, have received extensive attention. However, the genetic architecture of these important traits remains largely unknown, which may prevent full utilization of their genetic potential. In addition, no national conservation farm has been established for LXW, and the lack of effective protection will cause loss of genetic diversity and unique genomic information.

© 2022 The Authors. Published by Elsevier Inc. on behalf of Poultry Science Association Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Received November 30, 2021.

Accepted October 15, 2022.

¹Corresponding author: haochen-jxau@hotmail.com

Here, we sequenced 45 LXW individuals and compared them with 96 other Chinese geese (Guangfeng white goose [GFW] and Fengcheng gray goose [FCG]) to search for potential genomic evidence related to the breeding features of LXW. As efficient methods for detecting selection signals between 2 different groups, Fst (Weir and Cockerham, 1984), based on allele frequency, and XP-EHH (Sabeti et al., 2007), based on linkage disequilibrium, were used to detect selection effects. We aimed to identify genomic variants specific to LXW and a suite of promising genes that have undergone positive selection during domestication and breeding processes. Genetic research on breed characteristics will contribute to further genetic and breeding research on LXW and other geese breeds.

MATERIALS AND METHODS

Experimental Animals

A total of 141 geese from 3 Chinese indigenous geese breeds were used in this study: Lingxian white goose (LXW; $n = 45$), Fengcheng gray goose (FCG; $n = 50$), and Guangfeng white goose (GFW; $n = 46$). All samples were collected from Jiangxi Province, and LXW was collected from Lianhua County, Jiangxi Province. Whole blood samples were collected from 141 geese from veins under the wings. All experimental procedures were approved by the Animal Ethics Committee of Jiangxi Science and Technology Normal University and strictly followed the Guidelines of Animal Welfare, China.

Whole-Genome Resequencing

We used a standard phenol-chloroform extraction protocol to extract genomic DNA from whole blood of 141 geese and tested the integrity and purity of DNA by agarose gel electrophoresis and the A260/280 ratio (1.8–2.0), respectively. All qualified samples were used for library construction (Paired-end, 2×150 bp) and whole-genome resequencing was performed using the Illumina NovaSeq 6000 platform at Novogene (Beijing, China). The average sequencing depth of geese in this study was approximately 10-fold.

Genome-Wide Variant Calling and Annotation

To minimize artificial bias in the sequencing process, we applied Trimmomatic v0.32 (Bolger et al., 2014) to discard adaptors and low-quality reads ($Q < 30$). The clean reads were then mapped onto the Xingguo gray goose (XGG) reference genome (accession number: GWHBAAW00000000) using Burrows-Wheeler Aligner v0.7.17 (Li and Durbin, 2009) software with default parameters. The mapping results were further converted into BAM format files and sorted by SAMtools v1.10 (Li et al., 2009). Next, SNPs were detected using a custom version of Sentieon v201711.03 (Kendig et al., 2019)

DNaseq pipeline following the best practice algorithms of GATK. LocusCollector and Realigner functions from Sentieon were used to delete PCR duplicates and calibrate them according to the quality score. Finally, GATK v4.0.12 (McKenna et al., 2010) and SAMtools were employed to recalibrate the base quality for further analyses. After data filtering (minor allele frequency [MAF] $> 1\%$ and call rate $> 90\%$), 10,459,469 SNPs were detected in all geese, and LXW yielded 9,756,715 SNPs (MAF $> 1\%$ and call rate $> 90\%$). These variants were annotated using SnpEff v5.0 (Cingolani et al., 2012) software.

Population Structure and Genetic Diversity

After filtering, 10,459,469 SNPs were used for the phylogenetic and population structure analyses. A neighbor-joining (NJ) tree was constructed using PHYLIP v3.69 (Retief, 2000) based on the identical-by-state matrix calculated by PLINK v1.9 (Chang et al., 2015) with the parameter “–distance-matrix”, and then visualized via Figtree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). Principal component analysis (PCA) was performed with the parameter of “–make-grm” via GCTA v1.92 (Yang et al., 2011), and the first 2 principal components (PCs) were plotted using in-house R scripts. Admixture v1.3.0 (Alexander et al., 2009) was used to estimate the admixture proportions (from $K = 2$ to 3). Five parameters were calculated to evaluate the genetic diversity of the geese: the number of SNPs (NSNPs), runs of homozygosity (ROH), inbreeding coefficient (F), linkage disequilibrium (LD), and nucleotide diversity (π). ROH was estimated using the autosomal SNPs of each individual using PLINK. F was computed using the command “plink –het”. LD was quantified using the squared correlation coefficient (r^2) between pairs of SNPs as implemented in PLINK with the parameters “–r2 –ld-window-kb 1000 –ld-window-r2 0”. π was calculated using VCFtools v0.1.16 (Danecek et al., 2011) under 10 kb sliding windows.

Detection of Selective Signatures

To identify the positive signatures in LXW against other breeds (GFW and FCG), 2 statistical approaches (Fst and XP-EHH) were applied to improve detection efficacy. Only SNPs in the autosomes were preserved for selective sweeps. First, we estimated the pairwise Fst values in the groups of LXW and other geese using VCFtools v0.1.16 (Danecek et al., 2011), and the average Fst score was used as the summary statistics for each 10 kb non-overlapping window. Next, LXW was used as the experimental group and other geese as the reference to carry out the XP-EHH test with 10 kb non-overlapping sliding windows. The XP-EHH scores were then standardized across the entire genome. The results of the XP-EHH are directional, with a positive score indicating that the test population is highly likely to have experienced positive selection, whereas a negative

score indicates that the reference population has experienced consistent selection during the evolutionary process.

For these two approaches, windows with less than 10 SNPs were excluded, and the top 1% was set as the significance threshold. To reduce the frequency of false-positive signals, the potential candidate regions detected by these 2 methods were considered to be significantly selective regions, and variants located within as more confident signals. Gene Ontology (GO) enrichment analysis was performed for genes overlapping with selective windows using the online Metascape (Zhou et al., 2019).

Haplotype Analysis and Functional Annotation of Candidate Genes

Genecard websites and literature from public databases (for example, NCBI) were used to further investigate the possible biological functions of candidate genes. The π and heterozygosity (Hp) in the target regions were calculated over 10 kb windows via VCFtools (Danecek et al., 2011) and PLINK (Chang et al., 2015), respectively. To further investigate the significantly positively selected genes associated with the breed features of LXW, the resequencing data of 300 individuals from 10 populations, including 89 wild geese and 211 domestic geese (Table S1), in our own bird database (unpublished) were used for haplotype analysis. The details were as follows: 1) First, XGG was used as a reference genome to detect SNPs and InDels (insertions and

deletions of less than 50 bp) in candidate regions of 300 geese; 2) Next, we used the fastPhase function with 1,000 iterations in Beagle v5.1 (Browning et al., 2018) to infer haplotypes of candidate regions, and the haplotype network was constructed using pegas v1.1 (Paradis, 2010) in an R package. Additionally, the statistics for absolute relative allele frequencies (absRAFDif) within the tested populations were computed to verify the target regions, and SnpEff v5.0 (Cingolani et al., 2012) was used to predict the potential effects of SNPs.

RESULTS

Identification of Variants in LXW

In the present study, 141 individuals, including 45 LXW, 50 FCG, and 46 GFW, were selected for whole-genome resequencing with an average depth of $10 \times$. Aligning the sequencing data to the XGG reference genome (GWHBAAW00000000), we detected 10,459,469 SNPs (MAF > 1% and call rate > 90%). Most of these SNPs were located in intergenic regions (6,124,303; 49.196%), followed by 32.936% (4,100,134) of SNPs located in introns, 7.98% in upstream regions, 7.772% in downstream regions, and 1.149% in exons. From these SNPs, we found that 62,139 were specific to LXW (Table S2), with 415 missense mutations covering 278 protein-coding genes. GO enrichment of the SNPs indicated that these genes were involved in protein

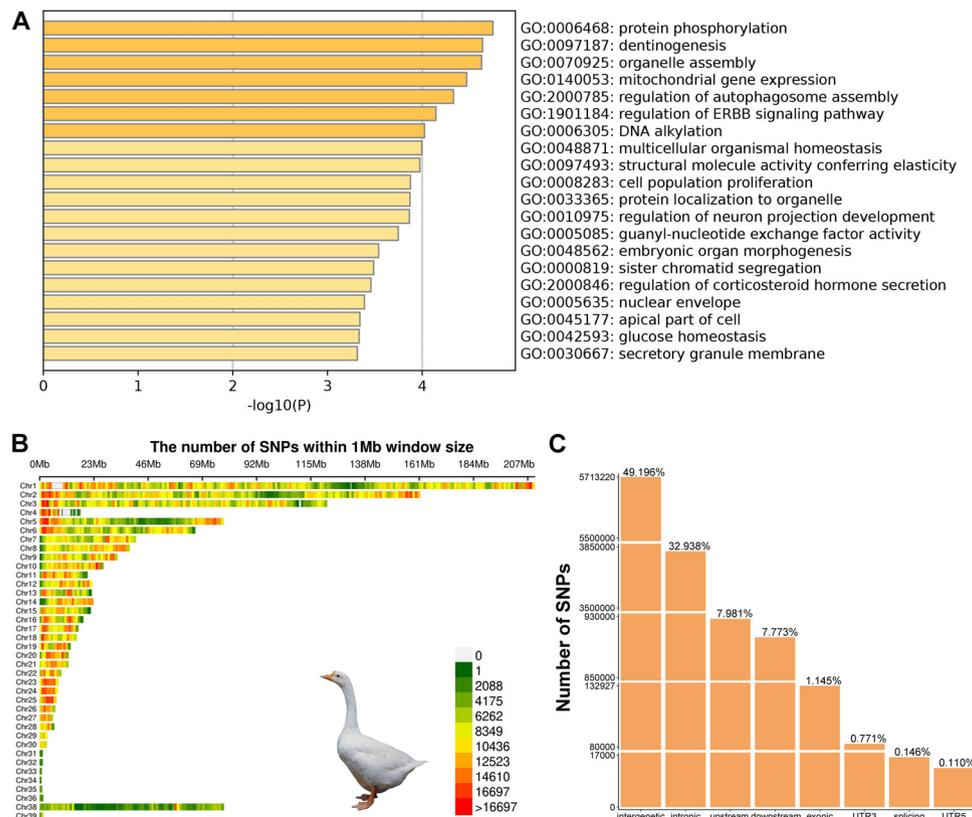


Figure 1. Identification and annotation of SNPs in LXW population. (A) GO enrichment of the 278 protein coding genes where LXW-specific SNPs are located. (B) The SNPs density across the whole genome was estimated in each 1 Mb genome block. (C) The annotation of genome-wide SNPs according to SnpEff.

phosphorylation, dentinogenesis, organelle assembly, and embryonic organ morphogenesis (Figure 1A).

Additionally, we independently extracted the SNPs of LXW and obtained 9,756,715 SNPs (MAF > 1%, call rate > 90%; Table S3). Among these SNPs, substitution mutations G > A and C > T occupied the largest number, followed by T > C and A > G (Table S4). The SNP density across the whole genome was presented in each 1 Mb genome block, among which Chr 23 and 24 showed the highest SNP density, and Chr 32 the lowest SNP density (Figure 1B). We found that the largest number of SNPs was enriched in the intergenic region (49.196%), followed by introns (32.938%), upstream regions (7.981%), downstream regions (7.773%), exons (1.145%), untranslated regions (0.722%), and splice site regions (0.146%) (Figure 1C). With respect to protein-coding genes, we identified 43,245 missense variants, 89,016 synonymous variants, 85 initiator-codon variants, 641 stop-lost variants, 113 start-lost variants, and 1,545 stop-gained variants (Table S5). These potential functional SNPs might be related to the altered traits in geese and would offer valuable genetic resources for in-depth research on the selective signatures of the LXW population.

Population Structure and Genetic Diversity of LXW

To infer the genetic structure of LXW, we used genome-wide variants (10,459,469 SNPs) to construct a phylogenetic tree and performed principal component analysis (PCA). All individuals were clustered together according to their breed, indicating that the individuals

were genetically representative of the breeds (Figure 2A). Principal components 1 (PC1, 4.71% of the total variation) and 2 (PC2, 3.92% of the total variation) also clearly separated the samples into 3 distinct clusters corresponding to FCG, GFW, and LXW (Figure 2B). Additional ADMIXTURE analysis was conducted to estimate the admixture proportions by assuming K ancestral populations (Figure 2C). When K = 2, a clear division was observed between LXW and other geese breeds. When K = 3, there was a split between each breed that matched the phylogenetic tree and the PCA results. Notably, some individuals of LXW had a higher proportion of co-ancestry with others (Figure 2C), possibly due to complex introgression patterns among these geese breeds. Detailed PCA results for LXW also revealed the presence of several subgroups within the LXW population (Figure S1), which may require scientifically effective measures to maintain germplasm uniformity.

We then estimated 5 parameters to evaluate the genetic diversity of LXW. Statistical results showed that the number of SNPs (NSNP) in LXW (9,634,706) was much lower than that in GFW (9,677,929) and FCG (10,087,193; Table S6), implying lower genetic diversity. As expected, low-diversity LXW showed high coverage by ROH (Figure 2D). Among these breeds, we also noticed that LXW ($\pi = 2.51 \times 10^{-3}$) had lower nucleotide diversity than GFW ($\pi = 2.57 \times 10^{-3}$) and FCG ($\pi = 2.68 \times 10^{-3}$), with the highest level of inbreeding ($F = 0.138$; Table S6 and Figure 2E). LD analysis also showed that the LXW genomes have relatively long LD distances ($r_{0.3}^2 = 3.71$ kb) and a slower decay of the pairwise correlation coefficient (r^2) than the others (Figure 2F and Table S6).

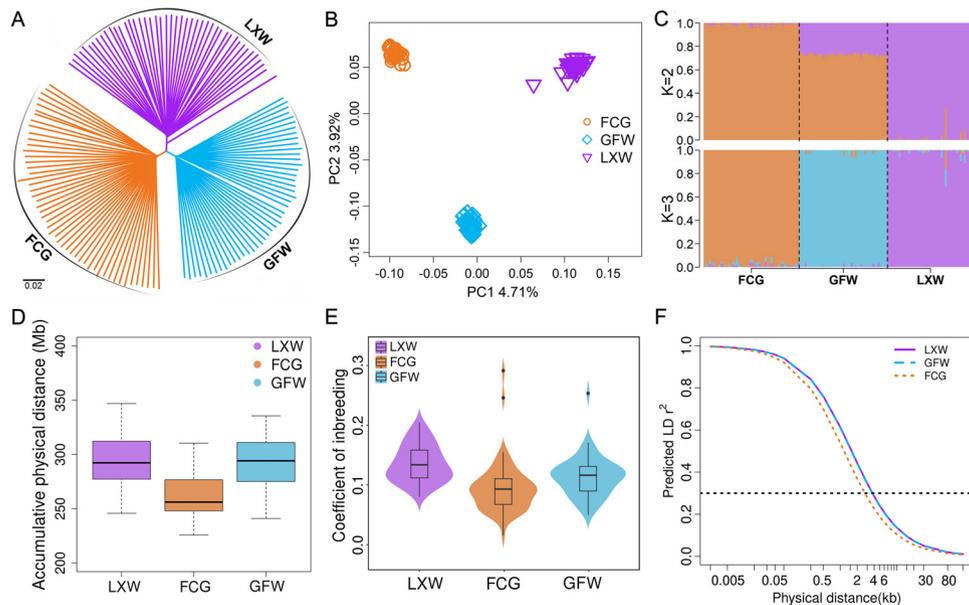


Figure 2. Phylogeny and population genetics of three geese breeds. (A) The neighbor-joining phylogenetic tree of 141 geese was constructed based on 10,459,469 SNPs. (B) Two-dimensional principal component analysis (PCA) plot of three geese breeds. (C) Genetic structure of three geese breeds according to ADMIXTURE, with K = 2–3. (D) The total length of runs of homozygosity (ROH) of each breed. (E) The inbreeding coefficient (F) of each breed. (F) Linkage disequilibrium (LD) analysis for three geese breeds. Abbreviations: FCG, Fengcheng gray goose; GFW, Guangfeng white goose; LXW, Lingxian white goose.

Signatures of Selection in LXW

According to the population-scale genetic differences between LXW and other geese breeds, pairwise F_{st} was estimated to search for genomic regions where allele frequencies changed. We detected 1,059 and 861 regions (top 1% cut-off) with potential selection signatures of LXW against GFW (average $F_{st} = 0.055$) and FCG (average $F_{st} = 0.057$), respectively (Figure 3A). In the comparison of LXW and GFW, we identified 28 highly differentiated regions ($F_{st} > 0.5$) containing five reproduction-related genes, *SYNE1*, *ESR1*, *NRIP1*, *CCDC170*, and *ARMT1* (Figure 3A and Table S7). Comparison of LXW and FCG also found that the region where *NRIP1* gene is located had the strongest selective signal ($F_{st} = 0.55$; Figure 3A and Table S8). A total of 57 overlapping genes were found in the 2 comparisons, showing categories such as “ovulation cycle process” and “triglyceride catabolic process”, which are considered to be potentially involved in the response to reproductive activities and fat deposition processes in geese (Figure 3B). The XP-EHH test in LXW against GFW identified 997 positively selected regions (top 1% cut-off) harboring 154 genes, with the most remarkable signal in the *KCNIP4* gene (Figure 3C and Table S9). A total of 998 potential regions (top 1% cut-off) harboring 133 genes were positively selected in LXW against FCG (Figure 3C and Table S10). The common selective regions of LXW against GFW and FCG contained 66

candidate genes enriched in GO terms associated with “embryonic hemopoiesis”, “reproductive structure development”, and “phospholipid biosynthetic process” (Figure 3D), which might also affect the reproductive and fattening characteristics of LXW.

Subsequently, we combined the candidate genes detected using the 2 approaches described above to obtain more confident selective signals. A total of 38 significantly selected genes were detected in LXW (top 1% cut-off for XP-EHH and F_{st} ; Table S11), which were involved in “histone modification”, “cellular response to hormone stimulus”, and “smooth muscle contraction” (Figure S2). We checked the distribution of the selected regions within the genomic windows and found 11 overlapping genes (*CCDC149*, *DHX15*, *KCNIP4*, *LDB2*, *NSG1*, *PACRGL*, *RBPJ*, *RPS6KA5*, *SEPSECS*, *SLIT2*, and *ZBTB49*; Table 1). As expected, the haplotypes of these 11 genes also showed remarkable differences between LXW and other geese breeds as well as wild populations (Figure S3). The published literature was surveyed to further investigate possible biological functions (Table 1). *DHX15* (Pattabhi et al., 2019) and *SLIT2* (Altay et al., 2007) have been found to be involved in immunity; *KCNIP4* (Fan et al., 2017), *DHX15* (Xia et al., 2016), *LDB2* (Wei et al., 2020), and *NSG1* (Lee et al., 2016) were associated with the economic traits of livestock and poultry, such as reproductive capacity, meat quality, growth, and carcass traits; and *RBPJ* and *SLIT2* were considered as potential

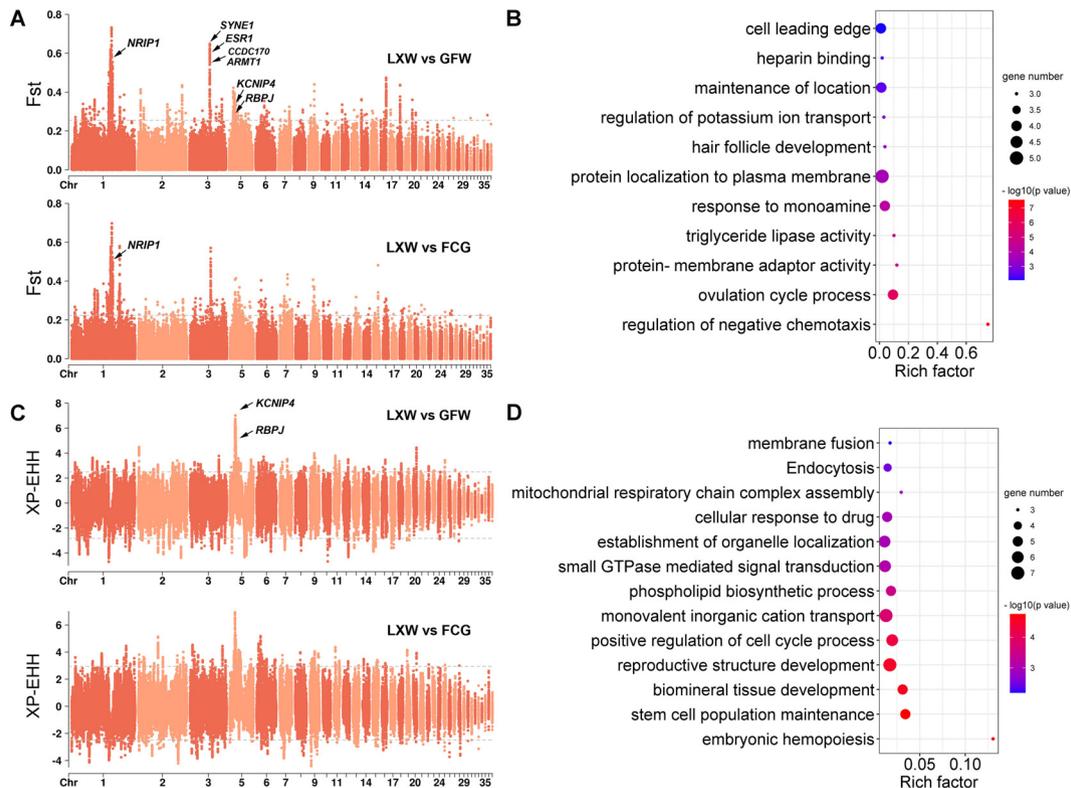


Figure 3. Genomic landscape of positive selection signatures by two statistical tests (XP-EHH and F_{st}). (A) Manhattan plot of F_{st} analyses with LXW against GFW or FCG. (B) Significant enriched GO terms (top 11) of the sharing common genes in F_{st} statistic. The y-axis in plot denotes $-\log_{10}(P$ value). (C) Manhattan plot of XP-EHH analyses with LXW against GFW or FCG. The dashed line refers to the threshold level of top 1%. Each dot denotes the mean values in the 10-kb non-overlapping genomic region. (D) Significant enriched GO terms (top 13) of the sharing common genes in XP-EHH statistic. The y-axis in plot denotes $-\log_{10}(P$ value).

Table 1. The 11 positively selected genes were identified and shared in two comparisons.

Chr	Start	End	Gene	Summary of gene function
5	19,065,083	19,114,966	<i>CCDC149</i>	Beak length trait (Huang et al., 2022)
5	18,966,684	19,013,336	<i>DHX15</i>	Meat tenderness (Xia et al., 2016) and immunity traits (Pattabhi et al., 2019)
5	17,675,084	17,722,057	<i>KCNIP4</i>	Reproductive traits (Fan et al., 2017)
5	16,093,124	16,312,302	<i>LDB2</i>	Body weight (Gu et al., 2011)
5	13,660,999	13,680,953	<i>NSG1</i>	Milk traits (Lee et al., 2016)
5	17,660,787	17,673,411	<i>PACRGL</i>	Growth and carcass traits (Liu et al., 2022)
5	19,477,656	19,537,056	<i>RBPJ</i>	Bone traits (Li et al., 2021)
5	19,177,151	19,203,929	<i>SEPSECS</i>	Neurological disorders (Puppala et al., 2016)
5	17,373,820	17,376,979	<i>SLIT2</i>	Immunity (Altay et al., 2007) and bone traits (Li et al., 2021)
5	13,702,820	13,721,308	<i>ZBTB49</i>	-
6	17,361,914	17,414,185	<i>RPS6KA5</i>	-

Note: Pairwise test (XP-EHH and Fst) analyses were performed with LXW tested against GWF and FCG. Abbreviations: FCG, Fengcheng gray goose; GFW, Guangfeng white goose; LXW, Lingxian white goose.

candidate genes for chicken bone growth and development (Li et al., 2021).

Signatures of Selection in Two Well-Characterized Growth Genes

LXW is small in body size and is characterized by early maturity. *KCNIP4* (Fan et al., 2017) and *RBPJ* (Li et al., 2021) have been identified as targets for the selection of reproductive and bone traits in poultry, respectively. We calculated the π and H_p values for each group in the region containing *KCNIP4*. The results showed that LXW was strongly selected in this gene region (Figure 4A), which was verified in haplotype heatmap analysis as having nearly fixed haplotypes (Figure 4B). A haplotype network of *KCNIP4* showed that haplotypes in domestic geese were completely different from those in wild geese. Among these domestic geese, LXW had the least haplotype

type ($n = 2$) and the smallest difference (nucleotide difference = 1) within the different haplotypes (Figure 4C and Figure S4). We also noted that 5 loci had extreme differences in absolute relative allele frequencies (abs-RAFdif > 0.9) between LXW and wild geese in *KCNIP4*. Two of these loci were found only in domestic geese, and the frequency in LXW was remarkably higher than that in other domestic geese breeds (abs-RAFdif > 0.5; Table S12). Evidence of strong selection for LXW was also observed in the *RBPJ* region (Figure 5A). The haplotypes of *RBPJ* were constructed and four main haplotypes were found in LXW. The most common haplotypes appeared 42 times in the 3 geese breeds, with 29 times in LXW, 9 times in FCG, and 4 times in GFW (Figures 5B and 5C). Furthermore, we noted that 2 main haplotypes were only found in LXW but were absent in GFW and FCG. Given the function of these genes, we speculated that this may explain the differences in body size and reproductive performance among distinct domestic geese breeds.

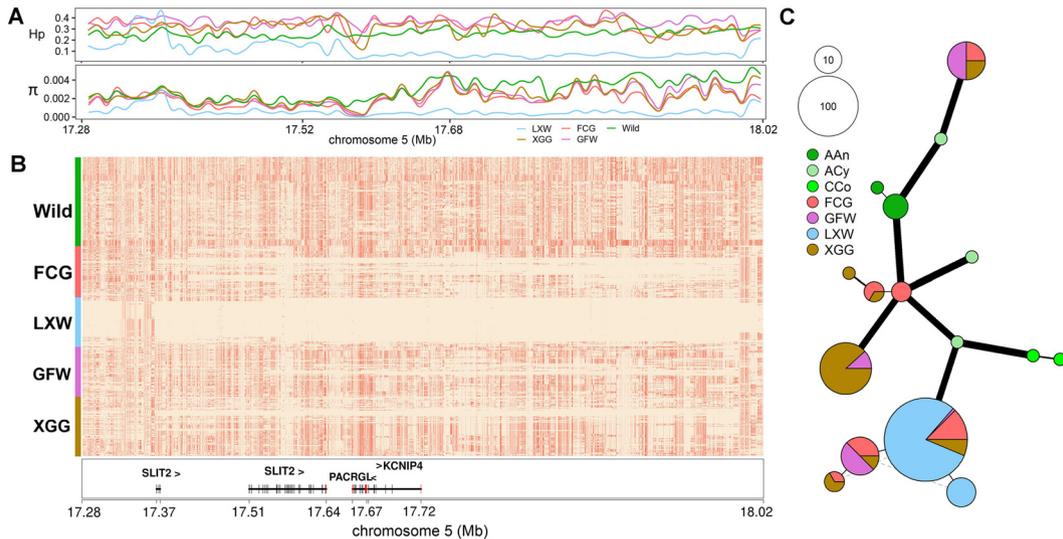


Figure 4. Population differentiation of *KCNIP4* gene (chr5: 17,675,084–17,722,057 bp) between LXW and other geese. (A) The Nucleotide diversity (π) and heterozygosity (H_p) values are plotted surrounding *KCNIP4* gene. (B) Haplotypes comparison between LXW and other geese. Major and minor alleles are labeled in beige and coral, respectively. (C) The haplotype network diagram of *KCNIP4* gene. Each circle denotes a haplotype, the circle size and haplotype frequency are proportional, and the width and length of the line represent the difference between haplotypes. Abbreviations: FCG, Fengcheng gray goose; GFW, Guangfeng white goose; LXW, Lingxian white goose; XGG, Xingguo gray goose; Wild, wild species (ACy: *Anser cygnoides*, AAn: *Anser anser*, AAl: *Anser albifrons*, AEr: *Anser erythropus*, AFa: *Anser fabalis*, CCo: *Cygnus columbianus*).

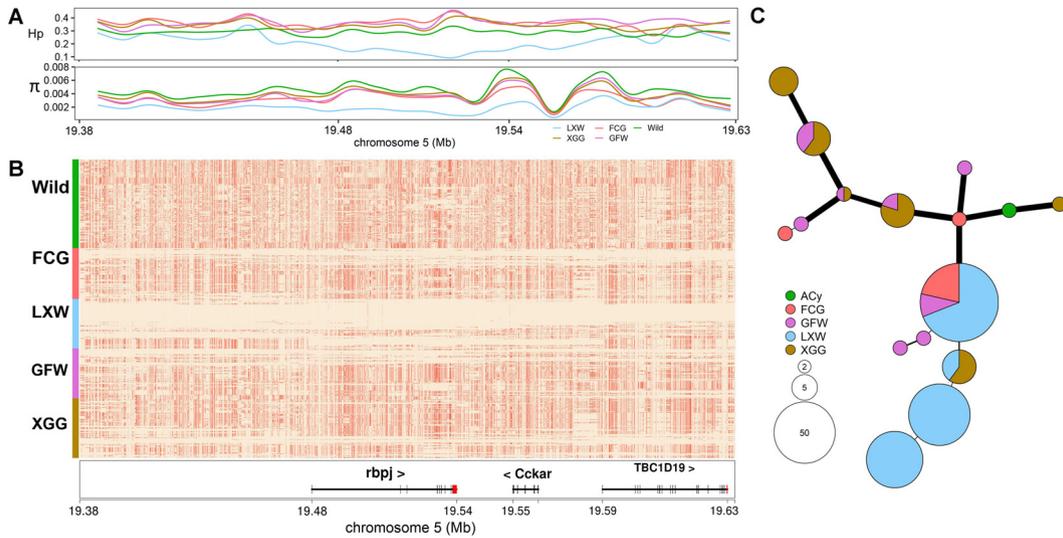


Figure 5. Population differentiation of *RBPJ* gene (chr5: 19,477,656–19,537,056 bp) between LXW and other geese. (A) The Nucleotide diversity (π) and heterozygosity (Hp) values are plotted surrounding *RBPJ* gene. (B) Haplotypes comparison between LXW and other geese. Major and minor alleles are labeled in beige and coral, respectively. (C) The haplotype network diagram of *RBPJ* gene. Each circle denotes a haplotype, the circle size and haplotype frequency are proportional, and the width and length of the line represent the difference between haplotypes. Abbreviations: FCG, Fengcheng gray goose; GFW, Guangfeng white goose; LXW, Lingxian white goose; XGG, Xingguo gray goose; Wild, wild species (ACy: *Anser cygnoides*, AAn: *Anser anser*, AAl: *Anser albifrons*, AEr: *Anser erythropus*, AFa: *Anser fabalis*, CCc: *Cygnus columbianus*).

DISCUSSION

LXW is an invaluable component of the genetic resources of geese in China. It is recognized for its small body size, rapid growth, strong disease resistance, tender meat, and excellent reproductive performance (Chen et al., 2011). Dissection of the genetic architecture of these traits is conducive to the genetic improvement of domestic geese, conservation of germplasm resources, and development of the local economy. Here, whole-genome resequencing data with more comprehensive and reliable information was used for further selective sweeps. We adopted a feasible strategy that combines the pairwise tests of F_{st} , which depend on allele frequency, and XP-EHH based on linkage disequilibrium patterns across genomes to explore candidate genes associated with breed characteristics at the genome-wide level. These 2 approaches are necessary and helpful for revealing the genetic mechanisms of phenotypic diversity.

Selection can increase beneficial allele frequency over time and fix it within a population. Our 2 comparative analyses of F_{st} statistics (LXW vs. GFW and LXW vs. FCG) identified extremely significant genetic differentiation ($F_{st} > 0.5$) of LXW in some regions that contained 5 important functional genes. Of these, *NRIP1* has been reported to be essential for ovulation and female fertility (White et al., 2000). *ESR1* is an estrogen receptor that participates in ovarian follicular development and has a potential effect on egg production traits in laying hens (Niu et al., 2017) and litter size in pigs (Munoz et al., 2007). *CCDC170* and *SYNE1* are involved in sex steroid hormone pathways (Sapkota et al., 2017). *ARMT1* is highly homologous to the dominantly expressed genes in *Drosophila* testes (Tang et al., 2018). All these genes and the pathways involved directly or indirectly play an

important role in reproductive function, which provides strong evidence for the excellent reproductive performance of LXW.

Genomic regions identified by F_{st} and XP-EHH (top 1% cut-off) identified only 11 common genes that potentially harbored selective signals. One possible explanation for this is that these 2 methods are based on different principles. F_{st} is a detection method based on population differentiation, which is mainly affected by migration and genetic drift (Weir and Cockerham, 1984). However, when the mutation is subjected to artificial or natural selection, its frequency increases to improve the level of population differentiation, and the F_{st} values also increase. Therefore, the F_{st} method is sensitive to signature selection between different populations. The XP-EHH test aims to assess haplotype differences between 2 populations and is more effective for some artificially selected signatures, considering the change in haplotypes (Simonson et al., 2010). Combining the results of 2 different statistical analyses provides more powerful information than the results of a single test. Furthermore, stringent screening criteria were used to obtain high-quality data; therefore, less stringent criteria may yield better results. Despite these limitations, the results partially explain the selective pressures that LXW underwent during evolution.

Among the 11 commonly selected genes, we found that *DHX15* (Pattabhi et al., 2019) and *SLIT2* (Altay et al., 2007) play roles in immune regulation. The cell-intrinsic innate immune response serves as the first line of defense against viral infection, and RNA helicases play an important role in the response to microbial infection. *DHX15* is a pre-mRNA processing factor that participates in the disassembly of spliceosomes after the release of mature mRNA. In human cells, *DHX15* is a co-receptor required for innate immune responses to

regulate and control RNA viral infection (Pattabhi et al., 2019). *SLIT2* has been previously identified as a ligand for the Robo family of immunoglobulin receptors. A study in mice found that *SLIT2* reduces leukocyte migration to cortical venules after global cerebral ischemia (Altay et al., 2007). Accordingly, both genes may be of great importance for the innate immunity of LXW.

We also identified a set of genes related to economic traits. For example, *NSG1* is associated with milk-related traits (milk yield, fat, and protein) of Holstein (Lee et al., 2016). *DHX15* plays an essential role in the shear force measured by the tenderness of Simmental beef cattle (Xia et al., 2016); thus, it may also affect the meat tenderness of LXW. Compared to the GFW and FCG populations, LXW displayed a faster growth rate and excellent reproductive performance. Our findings also illustrate these differences. *KCNIP4* has an important effect on egg weight in 300-day-old Jinghai Yellow chickens (Fan et al., 2017). We noted that the haplotype in *KCNIP4* had absolute differences between wild and domestic geese, and the genetic diversity of LXW around this gene was significantly reduced in comparison to other domestic geese. Given that the age of LXW at the first egg was earlier than that of the other 2 varieties (Chen et al., 2011), we speculated that the gene may play a critical role in the reproductive traits of LXW. These findings further confirmed that breeding LXW in the direction of egg and meat production might have achieved some results, which would lay a foundation for future research on the genomic characteristics of LXW and other poultry to develop improved varieties with excellent production performance.

Several genes have been suggested to be associated with growth traits. For example, the *LDB2* gene was found to be related to the body weight of chickens at wk 7 to 12 and average daily gain at wk 6 to 12 (Gu et al., 2011) and has been repeatedly reported to be associated with growth and carcass traits in chickens (Liu et al., 2013; Zhang et al., 2021). In addition, *SLIT2* participates in bone metabolism and inhibits osteoclastogenesis and bone resorption in vitro (Park et al., 2019). *RBPJ* is a key regulator of osteoclastogenesis and bone resorption and plays an important upstream negative regulatory role in osteoclast formation (Li et al., 2014). A recent genome-wide study suggested that *SLIT2* and *RBPJ* may be predominantly selected for bone traits in chickens (Li et al., 2021). Bones and the skeletal system play an important role in supporting and protecting the animal body and are key factors that affect the body size of animals. LXW is known for its small body size, whereas GFW, FCG, and XGG are medium-sized goose breeds. The strong selection of bone-related genes may explain, to a certain extent, the genetic basis of the difference in body size between LXW and other geese.

CONCLUSIONS

In summary, natural and artificial selection have remarkably shaped variation in the goose genome during

breed formation. We integrated Fst and XP-EHH tests to detect alleles that rapidly increased to a high frequency in the LXW population. A series of potential genes associated with meat quality, reproductive performance, and other functions, which could have strong impacts on the breed characteristics of LXW, were identified. To our knowledge, this study is one of the few to identify the selection signatures in LXW and could help better understand the selection mechanisms and breeding practices of LXW and other goose breeds.

ACKNOWLEDGMENTS

This study is supported by the National Natural Science Foundation of China (31860622 and 32060735) and the technology research project of Jiangxi Provincial Education Department (GJJ180595)

Data availability statement: The Xingguo gray goose genome used in this study has been deposited in the Genome Warehouse in BIG Data Center (<https://bigd.big.ac.cn/gwh/>) under accession number GWHBAA-W00000000. The genome resequencing data has been deposited in the NCBI database as BioProject No. PRJNA678815.

DISCLOSURES

The authors declare that they have no competing financial interests.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.psj.2022.102269](https://doi.org/10.1016/j.psj.2022.102269).

REFERENCES

- Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.
- Altay, T., B. McLaughlin, J. Y. Wu, T. S. Park, and J. M. Gidday. 2007. Slit modulates cerebrovascular inflammation and mediates neuroprotection against global cerebral ischemia. *Exp. Neurol.* 207:186–194.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bortoluzzi, C., H. J. Megens, M. Bosse, M. F. L. Derks, B. Dibbitts, K. Laport, S. Weigend, M. A. M. Groenen, and R. Crooijmans. 2020. Parallel genetic origin of foot feathering in birds. *Mol. Biol. Evol.* 37:2465–2476.
- Browning, B. L., Y. Zhou, and S. R. Browning. 2018. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103:338–348.
- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
- Chen, K. W., N. Yang, G. Y. Wang, Z. Y. Wang, K. H. Wang, J. Y. Wang, J. W. Wang, J. Wen, L. Z. Lu, H. F. Li, L. Zhang, X. Y. Zhang, X. Q. Zhang, Y. T. Chen, J. L. Chen, G. F. Gong, G. Y. Xu, and X. T. Kang. 2011. *Animal Genetic Resources in China: Poultry*. China Agriculture Press, China.

- Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 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.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and G. Genomes Project Analysis. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
- Eda, M., Y. Itahashi, H. Kikuchi, G. Sun, K. H. Hsu, T. Gakuhari, M. Yoneda, L. Jiang, G. Yang, and S. Nakamura. 2022. Multiple lines of evidence of early goose domestication in a 7,000-y-old rice cultivation village in the lower Yangtze River, China. *Proc. Natl. Acad. Sci. U S A* 119 e2117064119.
- Fan, Q. C., P. F. Wu, G. J. Dai, G. X. Zhang, T. Zhang, Q. Xue, H. Q. Shi, and J. Y. Wang. 2017. Identification of 19 loci for reproductive traits in a local Chinese chicken by genome-wide study. *Genet. Mol. Res.* 16:1–8.
- Grossman, S. R., I. Shlyakhter, E. K. Karlsson, E. H. Byrne, S. Morales, G. Frieden, E. Hostetter, E. Angelino, M. Garber, O. Zuk, E. S. Lander, S. F. Schaffner, and P. C. Sabeti. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 327:883–886.
- Gu, X., C. Feng, L. Ma, C. Song, Y. Wang, Y. Da, H. Li, K. Chen, S. Ye, C. Ge, X. Hu, and N. Li. 2011. Genome-wide association study of body weight in chicken F2 resource population. *PLoS One* 6:e21872.
- Huang, J., C. Wang, J. Ouyang, H. Tang, S. Zheng, Y. Xiong, Y. Gao, Y. Wu, L. Wang, X. Yan, and H. Chen. 2022. Identification of key candidate genes for beak length phenotype by whole-genome resequencing in geese. *Front. Vet. Sci.* 9:847481.
- Jeong, H., K. D. Song, M. Seo, K. Caetano-Anolles, J. Kim, W. Kwak, J. D. Oh, E. Kim, D. K. Jeong, S. Cho, H. Kim, and H. K. Lee. 2015. Exploring evidence of positive selection reveals genetic basis of meat quality traits in Berkshire pigs through whole genome sequencing. *BMC Genet.* 16:104.
- Kendig, K. I., S. Baheti, M. A. Bockol, T. M. Drucker, S. N. Hart, J. R. Heldenbrand, M. Hernaez, M. E. Hudson, M. T. Kalmbach, E. W. Klee, N. R. Mattson, C. A. Ross, M. Taschuk, E. D. Wieben, M. Wepert, D. E. Wildman, and L. S. Mainzer. 2019. Sentieon DNaseq variant calling workflow demonstrates strong computational performance and accuracy. *Front. Genet.* 10:736.
- Lee, Y. S., D. Shin, W. Lee, M. Taye, K. Cho, K. D. Park, and H. Kim. 2016. The prediction of the expected current selection coefficient of single nucleotide polymorphism associated with holstein milk yield, fat and protein contents. *Asian-Australas. J. Anim. Sci.* 29:36–42.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing. 2009. The sequence alignment/map format and samtools. *Bioinformatics* 25:2078–2079.
- Li, S., C. H. Miller, E. Giannopoulou, X. Hu, L. B. Ivashkiv, and B. Zhao. 2014. RBP-J imposes a requirement for ITAM-mediated costimulation of osteoclastogenesis. *J. Clin. Invest.* 124:5057–5073.
- Li, Y. D., X. Liu, Z. W. Li, W. J. Wang, Y. M. Li, Z. P. Cao, P. Luan, F. Xiao, H. H. Gao, H. S. Guo, N. Wang, H. Li, and S. Z. Wang. 2021. A combination of genome-wide association study and selection signature analysis dissects the genetic architecture underlying bone traits in chickens. *Animal* 15:100322.
- Liu, R., Y. Sun, G. Zhao, F. Wang, D. Wu, M. Zheng, J. Chen, L. Zhang, Y. Hu, and J. Wen. 2013. Genome-wide association study identifies Loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. *PLoS One* 8: e61172.
- Liu, Z., C. Bai, L. Shi, Y. He, M. Hu, H. Sun, H. Peng, W. Lai, S. Jiao, and Z. Zhao. 2022. Detection of selection signatures in South African Mutton Merino sheep using whole-genome sequencing data. *Anim. Genet.* 53:224–229.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303.
- Munoz, G., C. Ovilo, J. Estelle, L. Silio, A. Fernandez, and C. Rodriguez. 2007. Association with litter size of new polymorphisms on ESR1 and ESR2 genes in a Chinese-European pig line. *Genet. Sel. Evol.* 39:195–206.
- Niu, X., T. L. Tyasi, N. Qin, D. Liu, H. Zhu, X. Chen, F. Zhang, S. Yuan, and R. Xu. 2017. Sequence variations in estrogen receptor 1 and 2 genes and their association with egg production traits in Chinese Daggu chickens. *J. Vet. Med. Sci.* 79:927–934.
- Paradis, E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26:419–420.
- Park, S. J., J. Y. Lee, S. H. Lee, J. M. Koh, and B. J. Kim. 2019. SLIT2 inhibits osteoclastogenesis and bone resorption by suppression of Cdc42 activity. *Biochem. Biophys. Res. Commun.* 514:868–874.
- Pattabhi, S., M. L. Knoll, M. Gale Jr, and Y. M. Loo. 2019. DHX15 is a coreceptor for RLR signaling that promotes antiviral defense against RNA virus infection. *J. Interferon Cytokine Res.* 39:331–346.
- Puppala, A. K., R. L. French, D. Matthies, U. Baxa, S. Subramaniam, and M. Simonovic. 2016. Structural basis for early-onset neurological disorders caused by mutations in human selenocysteine synthase. *Sci. Rep.* 6:32563.
- Retief, J. D. 2000. Phylogenetic analysis using PHYLIP. *Methods Mol. Biol.* 132:243–258.
- Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E. H. Byrne, S. A. McCarroll, R. Gaudet, S. F. Schaffner, E. S. Lander, C. International HapMap, K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal, S. Pasternak, D. A. Wheeler, T. D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, H. Zhao, J. Zhou, S. B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R. C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, Y. Shen, W. Sun, H. Wang, Y. Wang, Y. Wang, X. Xiong, L. Xu, M. M. Wayne, S. K. Tsui, H. Xue, J. T. Wong, L. M. Galver, J. B. Fan, K. Gunderson, S. S. Murray, A. R. Oliphant, M. S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J. F. Olivier, M. S. Phillips, S. Roumy, C. Sallee, A. Verner, T. J. Hudson, P. Y. Kwok, D. Cai, D. C. Koboldt, R. D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L. C. Tsui, W. Mak, Y. Q. Song, P. K. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, A. Sekine, T. Tanaka, T. Tsunoda, P. Deloukas, C. P. Bird, M. Delgado, E. T. Dermitzakis, R. Gwilliam, S. Hunt, J. Morrison, D. Powell, B. E. Stranger, P. Whittaker, D. R. Bentley, M. J. Daly, P. I. de Bakker, J. Barrett, Y. R. Chretien, J. Maller, S. McCarroll, N. Patterson, I. Pe'er, A. Price, S. Purcell, D. J. Richter, P. Sabeti, R. Saxena, S. F. Schaffner, P. C. Sham, P. Varilly, D. Altshuler, L. D. Stein, L. Krishnan, A. V. Smith, M. K. Tello-Ruiz, G. A. Thorisson, A. Chakravarti, P. E. Chen, D. J. Cutler, C. S. Kashuk, S. Lin, G. R. Abecasis, W. Guan, Y. Li, H. M. Munro, Z. S. Qin, D. J. Thomas, G. McVean, A. Auton, L. Bottolo, N. Cardin, S. Eyheramendy, C. Freeman, J. Marchini, S. Myers, C. Spencer, M. Stephens, P. Donnelly, L. R. Cardon, G. Clarke, D. M. Evans, A. P. Morris, B. S. Weir, T. Tsunoda, T. A. Johnson, J. C. Mullikin, S. T. Sherry, M. Feolo, A. Skol, H. Zhang, C. Zeng, H. Zhao, I. Matsuda, Y. Fukushima, D. R. Macer, E. Suda, C. N. Rotimi, C. A. Adebamowo, I. Ajayi, T. Aniagwu, P. A. Marshall, C. Nkwodimmah, C. D. Royal, M. F. Leppert, M. Dixon, A. Peiffer, R. Qiu, A. Kent, K. Kato, N. Niikawa, I. F. Adewole, B. M. Knoppers, M. W. Foster, E. W. Clayton, J. Watkin, R. A. Gibbs, J. W. Belmont, D. Muzny, L. Nazareth, E. Sodergren, G. M. Weinstock, D. A. Wheeler, I. Yakub, S. B. Gabriel, R. C. Onofrio, D. J. Richter, L. Ziaugra, B. W. Birren, M. J. Daly, D. Altshuler, R. K. Wilson, L. L. Fulton, J. Rogers, J. Burton, N. P. Carter, C. M. Clee, M. Griffiths, M. C. Jones, K. McLay, R. W. Plumb, M. T. Ross, S. K. Sims, D. L. Willey, Z. Chen, H. Han, L. Kang, M. Godbout, J. C. Wallenburg, P. L'Archeveque, G. Bellemare, K. Saeki,

- H. Wang, D. An, H. Fu, Q. Li, Z. Wang, R. Wang, A. L. Holden, L. D. Brooks, J. E. McEwen, M. S. Guyer, V. O. Wang, J. L. Peterson, M. Shi, J. Spiegel, L. M. Sung, L. F. Zacharia, F. S. Collins, K. Kennedy, R. Jamieson, and J. Stewart. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.
- Sapkota, Y., V. Steinthorsdottir, A. P. Morris, A. Fassbender, N. Rahmioglu, I. De Vivo, J. E. Buring, F. Zhang, T. L. Edwards, S. Jones, D. O. D. Peterse, K. M. Rexrode, P. M. Ridker, A. J. Schork, S. MacGregor, N. G. Martin, C. M. Becker, S. Adachi, K. Yoshihara, T. Enomoto, A. Takahashi, Y. Kamatani, K. Matsuda, M. Kubo, G. Thorleifsson, R. T. Geirsson, U. Thorsteinsdottir, L. M. Wallace, P.-S. S. I. B. G. i, J. Yang, D. R. Velez Edwards, M. Nyegaard, S. K. Low, K. T. Zondervan, S. A. Missmer, T. D’Hooghe, G. W. Montgomery, D. I. Chasman, K. Stefansson, J. Y. Tung, and D. R. Nyholt. 2017. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat. Commun.* 8:15539.
- Simonson, T. S., Y. Yang, C. D. Huff, H. Yun, G. Qin, D. J. Witherspoon, Z. Bai, F. R. Lorenzo, J. Xing, L. B. Jorde, J. T. Prchal, and R. Ge. 2010. Genetic evidence for high-altitude adaptation in Tibet. *Sci.* 329:72–75.
- Tang, J. B., H. W. Cao, R. Xu, D. D. Zhang, and J. Huang. 2018. [Mutant generation of the testis genes and phenotype analyses in *Drosophila*]. *Yi Chuan* 40:478–487.
- Wei, C., D. Hou, Y. Feng, T. Li, Z. Jing, W. Li, R. Han, G. Li, G. Sun, Y. Tian, X. Liu, X. Kang, and Z. Li. 2020. Molecular characterization and a duplicated 31-bp indel within the LDB2 gene and its associations with production performance in chickens. *Gene* 761:145046.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- White, R., G. Leonardsson, I. Rosewell, M. Ann Jacobs, S. Milligan, and M. Parker. 2000. The nuclear receptor co-repressor nr1p1 (RIP140) is essential for female fertility. *Nat. Med.* 6:1368–1374.
- Xia, J., X. Qi, Y. Wu, B. Zhu, L. Xu, L. Zhang, X. Gao, Y. Chen, J. Li, and H. Gao. 2016. Genome-wide association study identifies loci and candidate genes for meat quality traits in Simmental beef cattle. *Mamm. Genome* 27:246–255.
- Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88:76–82.
- Zhang, Y., Y. Wang, Y. Li, J. Wu, X. Wang, C. Bian, Y. Tian, G. Sun, R. Han, X. Liu, R. Jiang, Y. Wang, G. Li, W. Li, X. Hu, and X. Kang. 2021. Genome-wide association study reveals the genetic determinism of growth traits in a Gushi-Anka F2 chicken population. *Heredity (Edinb)* 126:293–307.
- Zhou, Y., B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, and S. K. Chanda. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523.
- Zhou, Z., M. Li, H. Cheng, W. Fan, Z. Yuan, Q. Gao, Y. Xu, Z. Guo, Y. Zhang, J. Hu, H. Liu, D. Liu, W. Chen, Z. Zheng, Y. Jiang, Z. Wen, Y. Liu, H. Chen, M. Xie, Q. Zhang, W. Huang, W. Wang, S. Hou, and Y. Jiang. 2018. An intercross population study reveals genes associated with body size and plumage color in ducks. *Nat. Commun.* 9:2648.