Whole-genome resequencing reveals genetic characteristics of different duck breeds from the Guangxi region in China

Ran Wang,[1,2,†] Junli Sun,[3,†] Hu Han,[1,2] Yingfei Huang,[3] Tao Chen,[1,2] Manman Yang,[1,2] Qiang Wei,[1,2] Huofu Wan,[3] and Yuying Liao[3,4,*]

[1]BGI Institute of Applied Agriculture, BGI-Shenzhen, Shenzhen, Guangdong 518120, China
[2]ShenZhen Engineering Laboratory for Genomics-Assisted Animal Breeding, BGI-Shenzhen, Shenzhen, Guangdong 518083, China
[3]Guangxi Key Laboratory of Livestock Genetic Improvement, Animal Husbandry Research Institute of Guangxi Zhuang Autonomous Region, Nanning, Guangxi 530001, China
[4]Guangxi Veterinary Research Institute, Nanning, Guangxi 530001, China
†These authors contributed equally to this work.

*Corresponding author: liaoyuying@126.com

Abstract

Distinctive indigenous duck (*Anas platyrhynchos*) populations of Guangxi, China, evolved due to the geographical, cultural, and environmental variability of this region. To investigate the genetic diversity and population structure of the indigenous ducks of Guangxi, 78 individuals from eight populations were collected and sequenced by whole-genome resequencing with an average depth of $\sim$9.40$\times$. The eight indigenous duck populations included four breeds and four resource populations. Moreover, the genome data of 47 individuals from two typical meat-type breeds and two native egg-type breeds were obtained from a public database. Calculation of heterozygosity, nucleotide diversity (π), Tajima's *D*, and F_{ST} indicated that the Guangxi populations were characterized by higher genetic diversity and lower differentiation than meat-type breeds. The highest diversity was observed in the Xilin-Ma ducks. Principal component, structure, and phylogenetic tree analyses revealed the relationship between the indigenous duck populations of Guangxi. A mild degree of differentiation was observed among the Guangxi populations, although three populations were closer to the meat or egg breeds. Indigenous populations are famous for their special flavor, small body size, and slow growth rates. Selective sweep analysis revealed the candidate genes and pathways associated with these growth traits. Our findings provide a valuable source of information regarding genetic diversity, population conservation, and genome-associated breeding of ducks.

Keywords: genetic diversity; indigenous duck; WGS

Introduction

Guangxi (China) has over 180 billion cubic meters of surface water, with over 80% of the land territory that consists of mountains, plateaus, and basins. Pearl River, Yangtze, Duliu, and numerous other smaller rivers run through the mountains of Guangxi. The unique geographical environment results in the expansion of diverse ethnic tribes. Likewise, this region boasts a wide variety of indigenous duck populations such as the Jingxi duck (JXDMD), Longsheng-Cui duck (LSCD), Wenqiao duck (WQD), Donglan duck (DLD), Rongshui-Xiang duck (RSXD), Xilin-Ma duck (XLMD), Yulin-Ma duck (YLMD), and Yulin-Wu duck (YLWD). Among these, XLMD, LSCD, JXDMD, and RSXD have been recorded by Animal Genetic Resources in China Poultry (Chen *et al.* 2010). Geographical isolation rarely allows genomic communication, and the traditional free-range breeding pattern results in low selection intensity. The diversity and origin of some Chinese domestic duck breeds have been studied using molecular markers, including mitochondrial DNA (mtDNA) (Zhang *et al.* 2007; He *et al.* 2008; Li *et al.* 2010b), microsatellite markers (Li *et al.* 2006, 2010a; Liu *et al.* 2008), and single-nucleotide polymorphisms (SNPs) (Gu *et al.* 2020; Wang *et al.* 2020). However, most studies on Guangxi ducks focused on infectious diseases (Xie *et al.* 2012a,b; Peng *et al.* 2013), sequence data (Zhang *et al.* 2014; Xie *et al.* 2016a,b; Zhang *et al.* 2016) or only contained a few Guangxi duck breeds in their breed categories (Liu *et al.* 2008; Li *et al.* 2010b; Zhou *et al.* 2018). A comprehensive study of the genetic characteristics and population admixtures of all indigenous Guangxi populations is therefore important.

Here, we evaluated the genetic diversity of eight indigenous Guangxi duck populations using whole-genome resequencing and performed selective sweep analyses to identify candidate genes for future breeding.

Materials and methods

Production of whole-genome sequencing data

All animal-related handling and sampling procedures were approved by the Animal Experimental Ethical Inspection Form of the Animal Husbandry Research Institute of the Guangxi Zhuang Autonomous Region (Agreement No. 20200810). We handled animals in accordance with the recommendations of the European Commission (1997). All experimental procedures were conducted in accordance with relevant guidelines.

Blood samples from 78 individuals from eight indigenous Guangxi duck populations were collected from duck conservation farms (Figure S1 and Table S1). Blood samples drawn from the wing vein were mixed with ACD anticoagulant (0.85 mol/l trisodium citrate, 0.11 mol/l D-glucose, and 0.071 mol/l citric acid, pH 4.4). Genomic DNA, extracted from blood using a DNA lysate solution (Li *et al.* 2010b) and Proteinase K, was analyzed using Qubit (Invitrogen, Thermo Fisher Scientific, USA, Q32850) and agarose gel electrophoresis. Paired-end libraries were constructed according to the instructions of the MGIEasy FS DNA library prep set (BGI, Shenzhen, China, 1000006988) and sequenced as 100-bp paired-end reads on the MGISEQ-2000 platform (BGI, Shenzhen, China). We obtained over 883.25 Gb data in total. The raw reads are available from NCBI (Bioproject ID: PRJNA658213). In addition, we downloaded the genome data of 47 individuals from four duck breeds from PRJNA419832 (Zhang *et al.* 2018) and PRJNA450892 (Zhou *et al.* 2018). We obtained 1,267.33 Gb of whole-genome sequencing data from 125 individuals across 12 duck populations (Table S2).

Quality control and reads mapping

Quality control of all raw reads was performed using SOAPnuke1.5.0 (Chen *et al.* 2018, https://github.com/BGI-flexlab/SOAPnuke), with the following criteria: quality value ≤20; low-quality bases >30%; and N bases >5%. Clean reads were then mapped to the reference genome IASCAAS_Peking Duck_PBH1.5 (https://www.ncbi.nlm.nih.gov/genome/?term=IASCAAS_PekingDuck_PBH1.5) using BWA-0.7.12 (Li and Durbin 2009, https://sourceforge.net/projects/bio-bwa/) with default parameters. In the mapping process, a BAM file index was built using SamTools (Li *et al.* 2009, https://github.com/samtools/samtools/releases/), following which the BAM file was sorted using picard-tools-1.105 (https://github.com/broadinstitute/picard/releases). Duplicate reads data were deleted.

Variant calling and annotation

After mapping, the "HaplotypeCaller", "CombineGVCFs", "GenotypeGVCF", and "MergeVcfs" in GATK4 were used to detect SNPs and indels with default parameters. Output was in the form of a variant call format (VCF). We then screened out indels using "SelectVariants" in GATK4 and obtained SNPs. We further reserved the SNPs using "VariantFiltration" of GATK4 according to the following criteria: $QD < 2.0$; $FS > 200.0$; $MQ < 40.0$; and $ReadPosRankSum < -20.0$. Next, we used VCFtools to obtain high-quality SNPs combined with population information. SNPs that met at least one of the following standards were excluded: (i) minor allele frequency (MAF) <0.05; (ii) SNP call rate <90%; and (iii) $P < 0.000001$ for the Hardy–Weinberg equilibrium. Only the loci of two alleles were retained for the subsequent analyses. Following filtering, ~12 million SNPs and ~1.1 million indels remained for further analysis. Annotation files were then downloaded from NCBI. All filtered SNPs and indels were annotated using SnpEff (Cingolani *et al.* 2012) with default parameters. SNPs and indels were divided into eight types according to their location, namely intron, intergenic, upstream, downstream, 3′UTR, 5′UTR, splice, and exon.

Genetic diversity

Heterozygosity (Hp), nucleotide diversity (π), Tajima's D, and F_{ST} were calculated via VCFtools (Danecek *et al.* 2011, http://vcftools.sourceforge.net/) using the VCF filters for SNPs. Hp, π, and Tajima'D were calculated using "-het", "–window-pi" with 10 M windows, and "–TajimaD" with 100 M windows, respectively. Finally, F_{ST} was estimated between each pair of the 12 populations on sliding 100-kb windows with 10-kb step-length using "–fst-windows-size 100,000 -fst-window-step 10,000" in VCFtools. Statistical significance was analyzed using "t.test" in the R package to conduct Welch's two-sample *t*-test for Hp, two-sample *t*-test for π, and Wilcoxon rank sum test with continuity correction for Tajima's D.

Population structure

Three methods were used to estimate population structure: (i) principal component analysis (PCA) by "–pca" order of plink using VCF filters for the SNP (Purcell *et al.* 2007, https://www.cog-genomics.org/plink2), whereby the top 20 PCAs were calculated, which, in turn, indicated that the contributions of PCA1 and PCA2 were 18.2% and 10.16%, respectively; (ii) population structure was inferred using a Bayesian-based approach via the software package Admixture_linux-1.23 (Alexander and Lange 2011, https://dalexander.github.io/admixture/index.html). Cross-validation statistics were performed to choose the optimum K value. (iii) VCF filters were converted into a matrix using VCF2Dis-1.09 (https://github.com/BGI-shenzhen/VCF2Dis), and tree filters were formed using PHYLIPNEW-3.69.650 (https://evolution.genetics.washington.edu/phylip.html). Lastly, a phylogenetic tree based on genetic distance was constructed via the neighbor-joining method by iTOL (Letunic and Bork 2007, http://itol.embl.de). PopLDdecay (Zhang *et al.* 2019, https://github.com/BGI-shenzhen/PopLDdecay) was used to calculate the r^2 of linkage disequilibrium (LD) with the "-MaxDist 300" command.

Detection of selection signatures

F_{ST} and θ_π were performed to evaluate fixation and differentiation. High-quality SNPs were chosen to calculate F_{ST} using VCFtools based on sliding 20-kb windows with a 10-kb step-length. F_{ST} values were then standardized as ZF_{ST} by calculating Z-scores according to the formula $x^* = \frac{x - \bar{x}}{\sigma}$. We also calculated the θπ ratios and transformed $\log_2$ conversion, namely $\log_2(\theta\pi$ ratios). Gene annotation was performed using the top 5% ZF_{ST} and $\log_2(\theta\pi$ ratios), which showed statistical significance and were proven to be effective in previous studies (Lai *et al.* 2016; Wang *et al.*, 2020). Gene ontology and KEGG analyses were performed using Metascape (Zhou *et al.* 2019) and compared to the human genome background.

Data availability

Raw reads of the 78 Guangxi ducks can be obtained from NCBI via BioProject ID: PRJNA658213. The public data were downloaded from NCBI (BioProject ID: PRJNA419832 and PRJNA450892). The IRB number is 20200810, issued from Animal Experimental Ethical Inspection Form of Animal Husbandry Research Institute of Guangxi Zhuang Autonomous Region. Correspondence associated with the samples and database is shown in Table S2. File S1 contains Figures S1–S4, and File S2 contains Tables S1–S11.

Supplementary material is available at *figshare* online (https://doi.org/10.6084/m9.figshare.13176029.v1).

Results and discussion

Genome resequencing and variation detection

Seventy-eight individuals were selected for whole-genome resequencing from eight duck populations indigenous to Guangxi. Average data sets of 11.32 Gb (9.40×) per individual, >96.21 Gb (>80.51×) for each population, and >883.25 Gb (>733.40×) were obtained following quality control. We also downloaded data for 47 individuals of two meat-type ducks (PK and CV) and two native egg-type ducks (SM and SX) from the public database (Tables S1 and S2), resulting in 1,267.33 Gb data in total. The average mapping rate was 94.31% (92.75–95.50%), with a coverage rate of 97.84% (94.41–99.19%) (Table S3).

Across samples, we identified 12,740,849 SNPs and 1,168,223 indels across the whole-genome using GATK. High-quality SNPs were defined as those corresponding to the following criteria: QD <2.0; FS >200.0; MQ <40.0; ReadPosRankSum <−20.0; MAF <0.05; SNP call rate <90%; and $P < 0.000001$ of the Hardy–Weinberg equilibrium. The distribution of SNPs is shown (Figure S2). A total of 5,746,447 (45.19%) SNPs, including those in upstream and downstream regions, were located in intergenic regions, while 6,968,869 (54.81%) were located in genes. Only 193,495 (1.52%) SNPs were in the exonic regions, whereas 6,402,187 (50.25%) were in the intronic regions. The distribution of indels was similar to that of SNPs, with 515,430 (44.18%) in the intergenic regions and 651,116 (55.82%) in genes, whereas 612,795 (52.53%) indels were located in the intronic regions, only 1,407 (0.12%) were found in the exonic regions (Figure 1 and Tables S4 and S5). These distribution results were consistent with those of previous studies on ducks (Gu *et al.* 2020) and chickens (Li *et al.* 2017). In summary, most variations were located in non-coding sequences, including intergenic and intronic regions, indicating that non-coding sequences that have the potential to change protein function by regulating gene expression were retained during evolution and domestication.
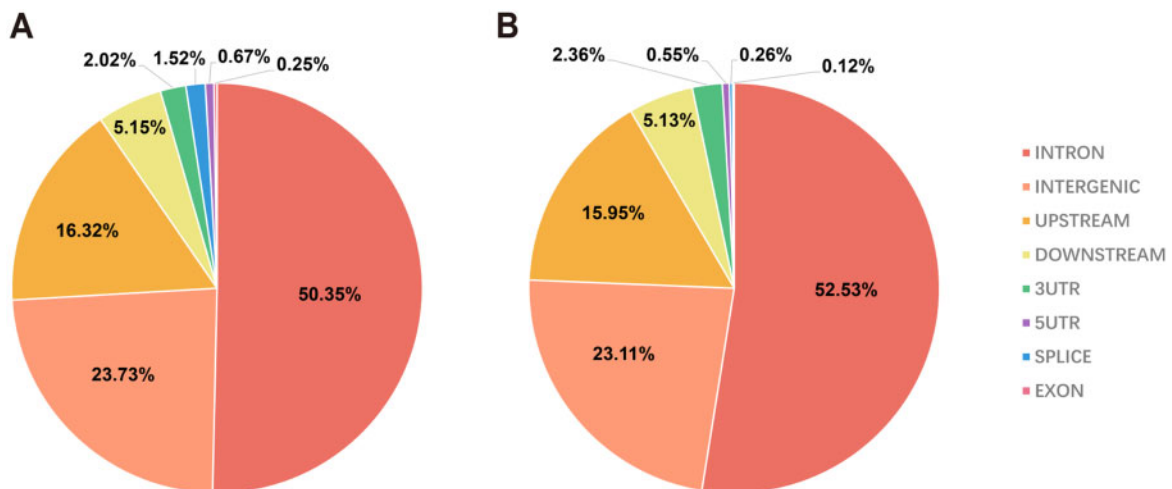
Hp and nucleotide diversity (π) were used to evaluate genetic diversity, where higher Hp and π values indicated richness of diversity. The results of genetic diversity evaluation are shown in Figure 2 and Table S6. Hp among indigenous Guangxi populations ranged from 0.2378 (DLD) to 0.2764 (XLMD), with an average of 0.2608. The Hp values of the eight Guangxi breeds were
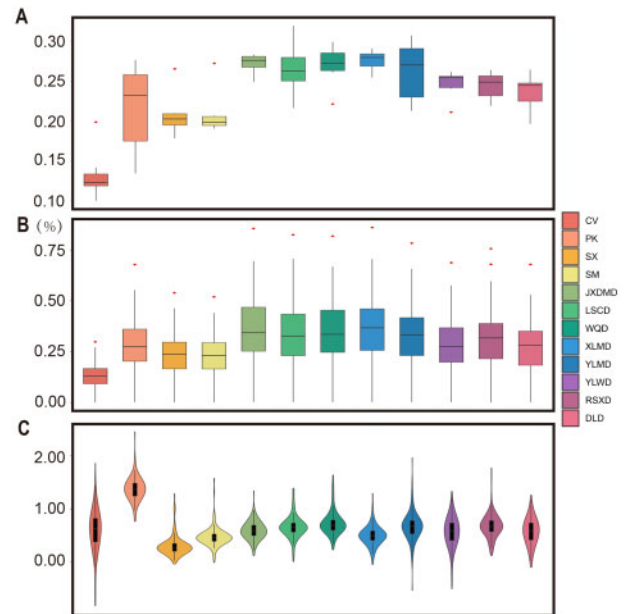

Figure 2 Genetic diversity of the 12 duck populations. (A) Boxplot showing the Hp of 12 duck populations. Indigenous Guangxi ducks showed a higher Hp than the other four breeds. The Welch's two-sample *t*-test was used to detect significant differences between the Guangxi *vs* meat breeds (P-value = 1.51×10^{-6}) and Guangxi *vs* egg breeds (P-value = 1.568×10^{-7}). (B) π values of the 12 duck populations indicated that those of the eight indigenous Guangxi duck breeds were significantly higher than those of the meat breeds (P-value = 3.912×10^{-13}) and the egg breeds (P-value = 2.538×10^{-8}). (C) Tajima's *D* of the 12 duck populations. All values were positive indicating that these were under balancing selection or population bottleneck. The data are shown (Table S6).

Figure 1 Distribution of SNPs and indels across the genome of the 12 duck populations via SNPEff. (A) The distribution of SNPs indicates that approximately half (50.35%) were in the intron region, whereas 23.73%, 16.32%, and 5.15% were in the intergenic, upstream, and downstream regions, respectively. In addition, 2.02%, 1.52%, 0.67%, and 0.25% were in the 3′UTR, splice, 5′UTR, and exon regions, respectively. (B) The distribution of indels indicates that 52.53% were in the intron region, 23.11% were in the intergenic region, 15.95% were in the upstream region, 5.13% were in the downstream region, 2.36% were in the 3′UTR region, 0.55% were in the 5′UTR region, 0.26% were in the splice region, and 0.12% were in the exon region (Tables S4 and S5).

significantly higher than that of meat breeds (P-value <0.001) and egg breeds (P-value <0.001). The overall nucleotide diversity (π) of Guangxi populations was 0.32%, ranging from 0.27% in DLD to 0.36% in XLMD among Guangxi populations, which was higher than that of meat and egg breeds (ranging from 0.13% to 0.28%; P-value <0.001). The lowest genetic diversity was observed in CV, consistent with breeding history. The higher Hp and π in the indigenous Guangxi duck populations, compared to those of the meat and egg breeds, indicated that the Guangxi duck populations displayed a higher genetic diversity.

Tajima's D is an index used to evaluate neutral selection. It assumes a negative value under strong purifying selection or a selective sweep and a positive value under balanced selection or a population bottleneck. Tajima's D values corresponding to the 12 populations were all positive, ranging from 0.4971 (XLMD) to 0.7325 (WQD) among the eight indigenous Guangxi duck populations, with an average of 0.6198. The highest was observed in PK (1.378) and the lowest in SX (0.3022). This may be attributed to the average Tajima's D value of the whole-genome being computed, whereas some areas under strong selection were difficult to highlight. In general, the Guangxi populations and egg breeds were under milder selection pressure compared to the meat breeds.

We also calculated the F_{ST} between the 12 duck populations (Table S7). The highest differentiation was between CV and DLD (0.2969), whereas the lowest was between JXDMD and PK (0.0609).

Among the eight indigenous Guangxi duck populations, RSXD and DLD exhibited a medium level of differentiation (0.1239) and also demonstrated medium differentiation when compared with the other six Guangxi populations. The indigenous Guangxi populations showed lower differentiation from egg breeds (SX and SM) compared with that from meat breeds (PK and CV), except JXDMD.

In general, all eight indigenous Guangxi duck populations showed high genetic diversity, with the highest seen in XLMD. An analysis of mtDNA and microsatellite markers conducted previously showed that XLMD had a higher diversity when compared with JXDMD (Li et al. 2006; Li et al. 2010b).

Population structure

To evaluate population stratification and admixture, we constructed a phylogenetic tree and performed PCA (Figure 3A) as well as population structure analysis of whole-genome SNPs of the 12 duck populations. The meat-type breeds (PK and CV) and egg breeds (SM and SX) formed separate clusters, substantiating the results of Zhang et al. (2018) and Zhou et al. (2018). The eight Guangxi populations were divided into different groups. JXDMD was closer to the meat breeds (PK and CV), whereas LSCD and WQD clustered with egg-type breeds (SX and SM). RSXD and DLD were distant from the other groups, consistent with F_{ST}. The first eigenvector distinguished growth traits in terms of body size, with CV and PK showing large body sizes, JXDMD showing
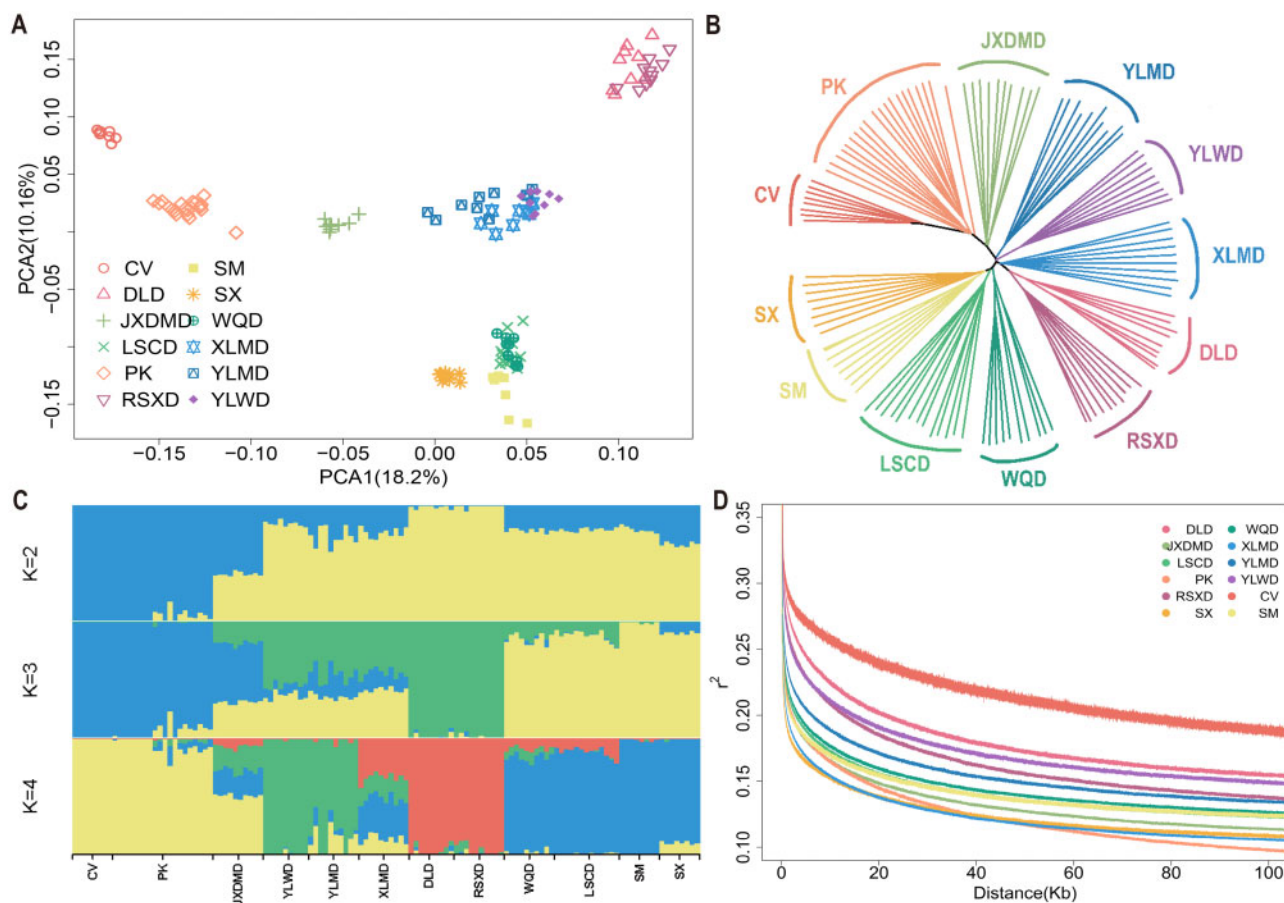

Figure 3 The structures of eight indigenous Guangxi duck populations and four control breeds. (A) PCA. Principal components 1 (18.2%) and 2 (10.16%) for the 125 ducks. (B) Phylogenetic tree. Phylogenetic relationships were estimated using the Neighbor-Joining method. (C) Population genetic structure of the 125 ducks inferred from the program Admixture_linux-1.23. The length of each segment represents the proportion of the individual genome inferred from ancestral populations (K = 2–4). (D) Decay of LD in the 12 duck populations, with one line per population.

medium sizes, and other populations showing small sizes. The phylogenetic relationship is shown in Figure 3B. The 12 populations were clustered separately into three distinct genetic groups. In general, JXDMD was clustered with the meat breeds (PK and CV), whereas LSCD and WQD were closer to the egg breeds (SM and SX), with the other five Guangxi populations forming a third group. We performed a population structure analysis using Admixture software to estimate individual ancestry. When K = 2, the meat breeds (PK and CV) were separated from other native populations, with JXDMD occupying the middle (Figure 3C). When K = 3, five indigenous Guangxi populations showed ancestral compositions that differed from those of egg breeds, consistent with that shown in the phylogenetic tree constructed by Zhou *et al.* (2018). DLD and RSXD had the same ancestral composition when *K* ranged from 2 to 4, suggesting that these had evolved from the same ancestral population. Three of the indigenous Guangxi populations clustered with meat- or egg-type breeds, implying that these populations were targeted for selection. The LD decay rates among indigenous Guangxi duck populations and egg breeds were similar, with the fastest LD decay observed in the meat-type breed CV (Figure 3D).

In summary, the eight Guangxi populations were separate from each other. JXDMD was similar to the meat breeds (PK and CV). This may be due to the selection of larger body size in JXDMD or the possible introduction of meat breeds such as PK. LSCD and WQD were closer to egg populations (SX and SM), which may be because these were all Chinese native duck breeds that were previously clustered together (Zhou *et al.* 2018). DLD, RSXD, YLMD, and YLWD were clustered together. These underwent a free-range raising pattern, and their close geographical location may have caused genetic communication.

Selective sweep and functional enrichment analysis

Natural and artificial selection may leave characteristic footprints on the genome. Selection of beneficial mutations can affect the patterns of genetic variation at surrounding loci, causing a reduction in Hp, skewed allele frequency distribution and excess high-frequency derived alleles near the selected allele (Smith and Haigh 1974). Genes and pathways related to traits may be located via genome scanning.

Combining the results of F_{ST} and population structure indicated that the genetic distance between indigenous Guangxi populations and meat-type breeds was larger than that between indigenous Guangxi populations and egg-type breeds, indicating that these had a greater potential for breeding. To identify candidate genes for breeding, we compared genome-wide variations in five indigenous Guangxi populations (DLD, YLMD, RSXD, XLMD, and YLWD) with those of meat breeds (PK and CV). ZF_{ST} and calculated $\log_2(\theta\pi$ ratios) were used to evaluate fixation and differentiation (Figure 4). A total of 106 genes were identified in the top 5% areas of ZF_{ST} and $\log_2(\theta\pi$ ratios) (Table S8), including *PHC1*, *MC2R*, *MC5R*, *MYLK*, *NCAM1*, *HOXB1* to *HOXB7*, and *MITF*. All 106 genes were used for GO and KEGG analyses and 79 terms were gathered (Figure S3 and Table S9). Most terms were associated with growth and development, including embryonic skeletal system development, glucagon signaling pathway, and response to nutrient levels. *PHC1* and its homologous genes are evolutionarily conserved in humans, chickens and zebrafish. They encode nuclear proteins that are components of the II PcG complex. The PcG complex regulates the transcription of *Hox* cluster genes in mammals (Isono *et al.* 2005), indicating that *PHC1* may be a candidate gene involved in regulating the growth of ducks. *MC2R* and *MC5R* were in the same cluster. These are both melanocortin receptors and mediate melanocortin signaling in different tissues
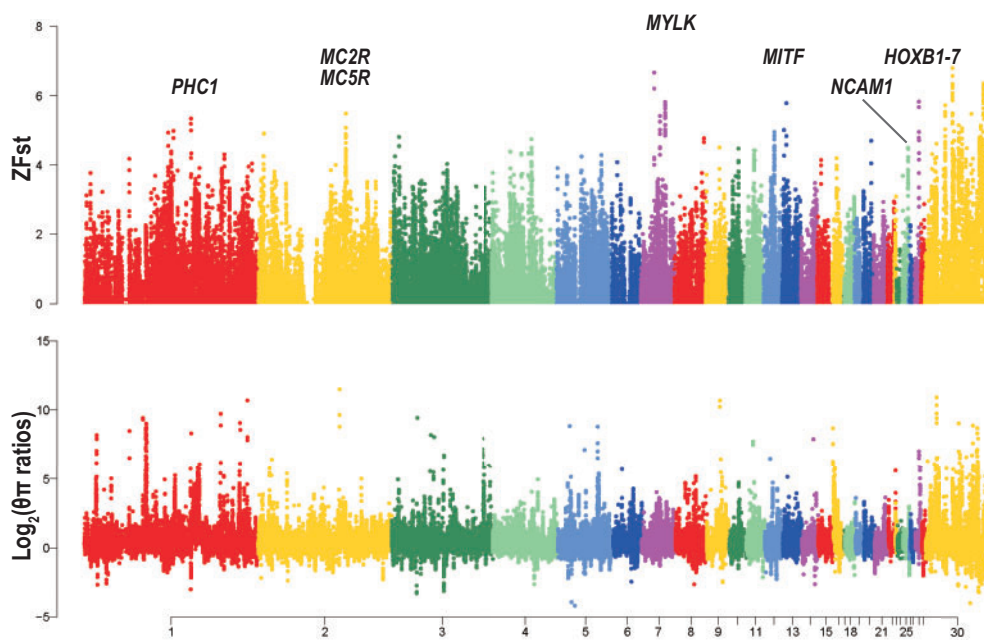

Figure 4 ZF_{ST} values and $\log_2(\theta\pi$ ratios) in Guangxi populations and meat breeds. The *x*-axis represents the chromosomes, where 1–29 are autosomes; and 30 is the Z chromosome. y-Axis represents ZF_{ST} values and $\log_2(\theta\pi$ ratios) for non-overlapping 20-kb windows with a 10-kb step-length.

(Boswell and Takeuchi 2005). MCRs are associated with energy homeostasis in mammals (Seeley *et al.* 2004). *MC2R* reportedly binds with ACTH rather than with α-MSH and is expressed in the adrenal gland and spleen, suggesting a potential function for ACTH endocrine and immunoregulatory functions in chicken (Takeuchi *et al.* 1998). *MC5R* is expressed in the brain and numerous peripheral tissues; however, its mRNA has not been detected in the spleen and skeletal muscle of chickens, indicating that its function may involve the regulation of secretions from exocrine glands (Takeuchi and Takahashi 1998). Selective signaling by *MC2R* and *MC5R* may be related to resistance, anti-stress or even energy balance during intensive livestock farming. It is also important to coordinate the selection of meat traits for immunity (Van Der Most *et al.* 2011). *MYLK* regulates smooth muscle contraction, actin–myosin interaction, inflammatory response, cell motility and morphology, and the nervous system in humans. It was reported as a candidate gene for abdominal fat deposition (Zhang *et al.* 2014) and high-altitude adaptation (Zhang *et al.* 2019) in chickens. Neural cell adhesion molecule 1 (*NCAM1*) encodes a cell surface marker in precursor cells carrying a high myogenic potential and tends to be downregulated during normal skeletal muscle development in pigs (Perruchot *et al.* 2015). Furthermore, it was suggested as a candidate gene for higher egg production in chickens (Shiue *et al.* 2006) and geese (Luan *et al.* 2014). The ∼90-kb region of *HOXB1* to *HOXB7* located 150-kb upstream of *IGF2BP1* on chromosome 28 showed both high ZF_{ST} and $\log_2(\theta\pi$ ratios) and was reported as the causative *cis*-regulatory locus related to body size in Peking ducks (Zhou *et al.* 2018). *MITF* was also notably correlated with plumage color differentiation (Zhang *et al.* 2018; Zhou *et al.* 2018). In conclusion, *PHC1*, *MC2R*, *MC5R*, *MYLK*, and *NCAM1* may act as potentially important breeding candidate genes for meat production.

We also scanned the selective sweep of indigenous Guangxi duck populations with egg breeds (SX and SM). A total of 84 genes and 35 terms were in the top 5% regions with a high ZF_{ST} and $\log_2(\theta\pi$ ratios) (Figure S4 and Tables S10 and S11). Most terms were associated with the nervous system, including the neuronal system, neuron recognition, and the synaptic vesicle cycle. The nervous system affects reproduction via the endocrine system (Pohl and Knobil 1982) and regulates behavior during domestication. Additionally, changes in neuronal development, common characteristics of animal domestication, have been reported in ducks (Zhou *et al.* 2018; Zhang *et al.* 2018; Wang *et al.* 2020), rabbits (Carneiro *et al.* 2014), and dogs (Axelsson *et al.* 2013).

Conclusion

We analyzed the genetic characteristics of indigenous Guangxi duck populations. The results suggested that indigenous ducks have abundant genetic diversity and potential for breeding. We adapted selective signals to detect candidate genes and pathways for breeding. This study is a data-rich resource and establishes a theoretical basis to further explore the genetic composition and drift of indigenous Guangxi ducks as well as accelerate modern duck breeding.

Acknowledgments

Literature cited

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 12:246.doi:10.1186/1471-2105-12-246.

Axelsson E, Ratnakumar A, Arendt M-L, Maqbool K, Webster MT, *et al.* 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. 495:360–364. doi: 10.1038/nature11837.

Boswell T, Takeuchi S. 2005. Recent developments in our understanding of the avian melanocortin system: Its involvement in the regulation of pigmentation and energy homeostasis. Peptides. 26:1733–1743. doi:10.1016/j.peptides.2004.11.039.

Carneiro M, Rubin CJ, Palma FD, Albert FW, Alfoldi J, *et al.* 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science. 345:1074–1079. doi: 10.1126/science.1253714.

Chen K-W, Yang N, Wang G-Y. 2010. Animal Genetic Resources in China Poultry. Beijing: China Agriculture Press,.

Chen Y-X, Chen Y-S, Shi C-M, Huang Z-B, Zhang Y, *et al.* 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 7:1–6. doi:10.1093/gigascience/gix120.

Cingolani P, Platts A, Wang L-L, Coon M, Nguyen T, *et al.* 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms. Fly (Austin). 6:80–92. doi:10.4161/fly.19695.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, *et al.*; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158. doi:10.1093/bioinformatics/btr330.

Gu L-H, Wang F, Lin Z-M, Xu T-S, Lin DJ, *et al.* 2020. Genetic characteristics of Jiaji duck by whole genome re-sequencing. PLoS One. 15:e0228964. doi:10.1371/journal.pone.0228964.

He D-Q, Zhu Q, Chen S-Y, Wang H-Y, Liu Y-P, *et al.* 2008. A homogenous nature of native Chinese duck matrilineal pool. BMC Evol Biol. 8:298–298. doi:10.1186/1471-2148-8-298.

Isono K, Fujimura Y, Shinga J, Yamaki M, O-Wang J, *et al.* 2005. Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes. Mol Cell Biol. 25:6694–6706. doi:10.1128/MCB.25.15.6694.

Lai F-N, Zhai H-L, Cheng M, Ma J-Y, Cheng S-F, *et al.* 2016. Whole-genome scanning for the litter size trait associated genes and SNPs under selection in dairy goat (Capra hircus). Sci Rep. 6: 38096. doi:10.1038/srep38096.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 23: 127–128. doi:10.1093/bioinformatics/btl529.

Li D, Che T, Chen B, Tian S, Zhou X, *et al.* 2017. Genomic data for 78 chicken from 14 populations. Gigascience. 6:1–5. doi:10.1093/gigascience/gix026.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760. doi:10.1093/bioinformatics/btp324.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, *et al.*; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078–2079. doi:10.1093/bioinformatics/btp352.

Li H, Song W, Shu J-T, Chen K-W, Zhu W-Q, *et al.* 2010a. Genetic diversity and population structure of 10 Chinese indigenous egg-type duck breeds assessed by microsatellite polymorphism. J Genet. 89:65–72. doi:10.1007/s12041-010-0012-3.

Li H, Yang N, Chen K, Chen G, Tang Q, *et al.* 2006. Study on molecular genetic diversity of native duck breeds in China. World's Poult Sci J. 62:603–611. doi:10.1017/S0043933906001152.

Li H, Zhu W, Song W-T, Shu J-T, Han W, *et al.* 2010b. Origin and genetic diversity of Chinese domestic ducks. Mol Phylogenet Evol. 57:634–640. doi:10.1016/j.ympev.2010.07.011.

Liu W, Hou Z, Qu L-J, Huang Y-H, Yao J-F, *et al.* 2008. Population structure and biodiversity of Chinese Indigenous duck breeds revealed by 15 microsatellite markers. Asian Australas J Anim Sci. 21:314–319. doi:10.5713/ajas.2008.70100.

Luan X, Cao Z, Li R, Liu M, Hu J. 2014. Differential expression profiling of hypothalamus genes in laying period and ceased period Huoyan geese. Mol Biol Rep. 41:3401–3411. doi:10.1007/s11033-014-3202-x.

Peng Y, Xie Z-X, Liu J-B, Pang Y-S, Deng X-W, *et al.* 2013. Epidemiological surveillance of low pathogenic avian influenza virus (LPAIV) from poultry in Guangxi Province, Southern China. PLoS One. 8:e77132. doi:10.1371/journal.pone.0077132.

Perruchot MH, Lefaucheur L, Louveau I, Mobuchon L, Palin MF, *et al.* 2015. Delayed muscle development in small pig fetuses around birth cannot be rectified by maternal early feed restriction and subsequent overfeeding during gestation. Animal. 9:1996–2005. doi:10.1017/S1751731115001202.

Pohl CR, Knobil E. 1982. The role of the central nervous system in the control of ovarian function in higher primates. Annu Rev Physiol. 44:583–593. doi:10.1146/annurev.ph.44.030182.003055.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, *et al.* 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81:559–575. doi:10.1086/519795.

Seeley RJ, Drazen DL, Clegg DJ. 2004. The critical role of the melanocortin system in the control of energy balance. Annu Rev Nutr. 24:133–149. doi:10.1146/annurev.nutr.24.012003.132428.

Shiue Y-L, Chen L-R, Chen C-F, Chen Y-L, Ju J-P, *et al.* 2006. Identification of transcripts related to high egg production in the chicken hypothalamus and pituitary gland. Theriogenology. 66:1274–1283. doi:10.1016/j.theriogenology.2006.03.037.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23:23–35. doi:10.1017/S0016672308009579.

Takeuchi S, Kudo T, Takahashi S. 1998. Molecular cloning of the chicken melanocortin 2 (ACTH)-receptor gene. Biochim Biophys Acta. 1403:102–108. doi:10.1016/s0167-4889(98)00022-6.

Takeuchi S, Takahashi S. 1998. Melanocortin receptor genes in the chicken tissue distributions. Gen Comp Endocrinol. 112:220–231. doi:10.1006/gcen.1998.7167.

Van Der Most PJ, de Jong B, Parmentier HK, Verhulst S. 2011. Trade-off between growth and immune function: a meta-analysis of selection experiments. Funct Ecol. 25:74–80. doi:10.1111/j.1365-2435.2010.01800.x.

Wang L, Guo J-Z, Xi Y, Ma S-C, Li Y-Y, *et al.* 2020. Understanding the genetic domestication history of the Jianchang duck by genotyping and sequencing of genomic genes under selection. G3 (Bethesda). 10:1469–1476. doi:10.1534/g3.119.400893.

Xie L-J, Xie Z-X, Zhao G-Y, Liu J-B, Pang Y-S, *et al.* 2012a. Complete genome sequence analysis of a duck circovirus from Guangxi pockmark ducks. J Virol. 86:13136–13136. doi:10.1128/JVI.02494-12.

Xie Z-X, Xie L-J, Zhou C-Y, Liu J-B, Pang Y-S, *et al.* 2012b. Complete genome sequence analysis of an H6N1 avian influenza virus isolated from Guangxi pockmark ducks. J Virol. 86:13868–13869. doi:10.1128/JVI.02700-12.

Xie Z-X, Zhang Y-F, Xie L-J, Deng X-W, Xie Z-Q, *et al.* 2016a. Sequence and gene organization of the Xilin small partridge duck (Anseriformes, Anatidae, Anas) mitochondrial DNA. Mitochondrial DNA A: DNA Mapp Seq Anal. 27:1579–1580. doi:10.3109/19401736.2014.958677.

Xie Z-X, Zhang Y-F, Xie L-J, Liu J-B, Deng X-W, *et al.* 2016b. The complete mitochondrial genome of the Jingxi duck (Anas platyrhynchos). Mitochondrial DNA A: DNA Mapp Seq Anal. 27:809–810. doi:10.3109/19401736.2014.919453.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, *et al.* 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 10:1523. doi:10.1038/s41467-019-09234-6.

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. doi:10.1093/bioinformatics/bty875.

Zhang H, Du Z-Q, Dong J-Q, Wang H-X, Shi H-Y, *et al.* 2014. Detection of genome-wide copy number variations in two chicken lines divergently selected for abdominal fat content. BMC Genomics. 15:517. doi:10.1186/1471-2164-15-517.

Zhang T-J, Li H-F, Chen K-W, Chang H, Tang Q-P, *et al.* 2007. Genetic diversity and systematic evolution of Chinese domestic ducks along the Yangtze-Huai River. Biochem Genet. 45:823–837. doi:10.1007/s10528-007-9121-y.

Zhang Y, Gou W, Zhang Y, Zhang H, Wu C. 2019. Insights into hypoxic adaptation in Tibetan chicken embryos from comparative proteomics. Comp Biochem Physiol D: Genomics Proteomics. 31:100602. doi:10.1016/j.cbd.2019.100602.

Zhang Y-F, Xie Z-X, Xie L-J, Deng X-W, Xie Z-Q, *et al.*, 2016. Analysis of the Rongshui Xiang duck (Anseriformes, Anatidae, Anas) mitochondrial DNA. Mitochondrial DNA A: DNA Mapp Seq Anal. 27:1867–1868. doi:10.3109/19401736.2014.971263.

Zhang Y-F, Xie Z-X, Xie L-J, Tan W, Liu J-B, *et al.* 2016. Genetic characterization of the Longsheng duck (Anas platyrhynchos) based on the mitochondrial DNA. Mitochondrial DNA A: DNA Mapp Seq Anal. 27:1146–1147. doi:10.3109/19401736.2014.936411.

Zhang Z, Jia Y, Almeida P, Mank J-E, van Tuinen M, *et al.* 2018. Whole-genome resequencing reveals signatures of selection and timing of duck domestication. GigaScience. 7. doi:10.1093/gigascience/giy027.

Zhou Z, Li M, Cheng H, Fan W, Yuan Z, *et al.* 2018. An intercross population study reveals genes associated with body size and plumage color in ducks. Nat Commun. 9:2648. doi:10.1038/s41467-018-04868-4.