# Whole genome sequencing revealed genetic diversity and selection of Guangxi indigenous chickens

Junli Sun[1¶], Tao Chen[2¶], Min Zhu[1¶], Ran Wang[2], Yingfei Huang[1], Qiang Wei[2], Manman Yang[3]*, Yuying Liao[4*]

[1] Guangxi Key Laboratory of Livestock Genetic Improvement, Animal Husbandry Research Institute of Guangxi Zhuang Autonomous Region, Nanning 530001, Guangxi, China.

[2] BGI Institute of Applied Agriculture, BGI-Shenzhen, Shenzhen 518120, China.

[3] BGI-Shenzhen, Shenzhen 518083, China.

[4] Guangxi Veterinary Research Institute, Nanning 530001, Guangxi, China.

*Corresponding author:

Manman Yang E-mail: yangmanman@genomics.cn

Yuying Liao E-mail: liaoyuying@126.com

# Abstract

Guangxi chickens play an important role in promoting the high-quality development of the broiler industry in China, but their value and potential are yet to be discovered. To determine the genetic diversity and population structure of Guangxi indigenous chicken, we analyzed the whole genomes of 185 chicken from 8 phenotypically and geographically representative Guangxi chicken breeds, together with 12 RJFt, 12 BRA and 12 WL genomes available from previous studies. Calculation of heterozygosity (Hp), nucleotide diversity (π), and LD level indicated that Guangxi populations were characterized by higher genetic diversity and lower differentiation than RJFt and commercial breeds except HGFC. Population structure analysis also confirmed the introgression from commercial broiler breeds. Each population clustered together while the overall differentiation was small, MA has the richest genetic diversity among all varieties. Selective sweep analysis revealed *BCO2*, *EDN3* and other candidate genes have received strong selection in local breeds, these also provided novel breeding visual and data basis for future breeding.

# Introduction

Chickens are the most widely distributed livestock species globally, more than half the total (53%) is found in Asia, one of the largest producers is China [1]. In China, poultry meat consumption accounts for the second largest proportion after pork. People in different regions have different preferences for the appearance, flavor, and cooking methods of chickens. Indigenous chickens with delicious meat quality and unique flavors are more in line with the habits and preferences of domestic consumers.

Guangxi Zhuang Autonomous Region is in mountainous terrain in the far south of China, and its unique climate has created unique and rich chicken germplasm resources. Six breeds are listed in Chinese national catalogue of livestock and poultry genetic resources includes Guangxi Three-yellow chicken (SHC), Guangxima chicken (MA), Nandanyao chicken (NDYC), Xiayan chicken (XYC), Guangxiwu chicken (WC), Longshengfeng chicken (LSFC), and two characteristic populations had a long history of local breeding are Dongzhongai chicken (DZAC) and Cenxi Gudian chicken (GDC). These germplasm resources provide good materials for the cultivation of high-quality chicken breeds and promote the development of poultry industry in Guangxi. However, only sporadic studies were reported on genomic information about them [2, 3].

A comprehensive and deep understanding of the genome diversity of the

indigenous breeds could reveal the genetic diversity and population structure of these breeds. This study therefore investigated genetic diversity, population structure, linkage disequilibrium (LD), and signature selection within Guangxi indigenous chickens using genome-wide single nucleotide polymorphisms (SNPs) generated from the whole genome sequencing.

# Ethics statement

This study was carried out in accordance with the guidelines of the Animal Experimental Ethical Inspection Form of Guangxi Research Institute (20190318).

# Materials and methods

## Sampling and genotyping

A total of 185 blood samples from six breeds and two characteristic populations were investigated from conservation center or breeding farms (S1 Table and S1 Fig) were collected for genomic DNA extracting. Whole-genome sequencing libraries were constructed by MGIEasy Universal DNA Library Prep Set and then sequenced using MGISEQ-2000 with PE100 developed by BGI Genomics Co., Ltd.

## Variant calling and annotation

After quality control, Pair-end reads were mapped onto the *Gallus gallus* GRCg6a using BWA version 0.7.12-r1039[4]. The bam files were sorted using SortSam and duplicated reads were marked using MarkDuplicates from Picard tools version 1.105. SNPs were detected and filtered using HaplotypeCaller and VariantFiltration command in GATK version 4.1.1.0. We applied hard filter command 'VariantFiltration' to exclude potential false-positive variant calls with the parameter "QD < 2.0 || ReadPosRankSum < -8.0 || FS > 60.0 || MQ < 40.0 || SOR > 3.0 || MQRankSum < -12.5 || QUAL < 30". To annotate the SNPs and InDels identified here, Vep (v95.3) was employed.

## Genomic diversity analysis

Genome-wide nucleotide diversity (π) and genetic differentiation (Fst) was performed using VCFtools (v0.1.13) [5] with parameters 40kb sliding window and 20kb step size. Individual heterozygosity (Hp) was calculated by following the formula given by Rubin et al. [6]:

$$Hp = \frac{2\sum n_{MAJ}\sum n_{MIN}}{(\sum n_{MAJ} + \sum n_{MIN})^2}$$

PopLDdecay version 3.40[7] was used to assess patterns in the extent of linkage disequilibrium. ROH analysis was performed using plink version 1.9, the parameters were as follows, the minimum length of ROH > 10kb, the number of SNP per window > 20, and only one heterozygote was allowed, ROH had at least one variant per 1000 kb on average. The results of ROH analysis were visualized with the R package pheatmap, the total ROHs length of each chromosome was centered and scaled in breed's level.

## Population structure analysis

To investigate the genetic background of the chickens, principal component analysis (PCA) and structure analysis were conducted. SNPs in high linkage disequilibrium were removed by PLINK version 1.9[8], based on the pruned SNP data, the individual ancestries were estimated using a maximum likelihood method implemented in ADMIXTURE version 1.23[9]. Using the default parameter (folds = 5) for cross-validation and the lowest cross-validation error was taken as the most probable K value. VCF2Dis (v1.09) software was used to calculate the P distance matrix, then use PHYLIPNEW (v3.69) to constructed a NJ-tree.

## Sweep analysis, GO enrichment and KEGG pathway analysis

The selective sweep analysis was performed using vcftools (v0.1.13) in Guangxi indigenous chickens. Scanning the whole genome selection signal with 40kb as the sliding window and 20kb as the step size, and windows with less than 10 SNPs are excluded [10]. Fst values were Z-transformed: $ZFst = \frac{(Fst - uFst)}{\sigma Fst}$, where μ is the mean of Fst and σ is the standard deviation of Fst. The sliding window with the top 1% of the ZFst value were defined as significant selected region [11]. GO enrichment analysis and KEGG pathway analysis were performed using the packages clusterProfiler, KEGG.db and org.Gg.eg.db in R. We select Benjamini-Hochberg method correction for multiple comparisons, and GO terms with a p-value less than 0.05 were considered significantly enriched.

# Results

## Variant calling and annotation

A total of 9.48 billion clean reads were obtained after quality filtering, corresponding to average depth and coverage is 9.39x and 96.97% (S3 Table). The overall mapping rate is greater than 98.4%. SNPs with MAF < 0.05, call rate < 0.8 and individual call rate < 0.9 were excluded, 13,245,769 high-quality SNPs and 3,790,305 indels were utilized for downstream analysis (Fig 1A). The Guangxi indigenous chickens harbored higher number of SNPs and indels than that of RJFt except for DZAC, while WL exhibited the lowest ones. Vep was used to annotate SNPs, 55.66% of these SNPs were aligned to the intron region, 28.43% were aligned to the intergenic region, and only 2.27% located in the exon region (Fig 1B).

The genome is divided into isochrones with a sliding window of 100kb, and divided into five categories (L1, L2, H1, H2 and H3) according to different GC level to explore the potential impact between GC content and genetic variations (Fig 1C and 1D) [12]. Our results shown that the L2 category has the largest number of isochrones, covering 37% of the genomic region, and the SNPs and Indels counts peak in this category. H1 category with higher GC level also contains a lot of genetic variations (S5 Table). In general, genomic regions with moderate GC content contain more variation.

**Fig1. The number, distribution, and GC content of SNP and INDEL.** (A) The SNP and indel number in different breeds. (B) The distribution of SNPs. (C) Scatter plot of SNP number and GC content in isochrones per 100kb window. (D) Scatter plot of Indel number and GC content in isochrones per 100kb window.

## Genetic diversity，LD and ROH analysis

To provide a more comprehensive understanding and profound insight into the genome diversity of Guangxi indigenous chickens, we incorporated the sequencing data of 12 Red jungle fowl population from Thailand (RJFt) and commercial breeds including 12 white layer (WL) and 12 Broiler A(BRA), which has been previously published [13]. The nucleotide diversity (π) and heterozygosity were calculated to evaluate the genetic diversity of all the chicken breeds. We observed Guangxi indigenous chickens harbored the higher genome-wide π than RJFt (π =0.00334) except for DZAC (π = 0.00332), the lowest genome-wide π in WL (π = 0.00152), followed by BRA (π = 0.0031) (Fig 2A). Unlike the results in nucleotide diversity, BRA harbored the highest heterozygous SNP rate (He), followed by DZAC, while MA harbored the lowest (He = 0.1684) one (Fig 2B). The He in Z chromosome are lower than any autosome among all populations (S2 Fig), probably due to sex chromosome had

undergone higher selective pressure than autosome [14]. Linkage disequilibrium (LD) analysis showed that WL population had the slowest LD decay rate, significantly slower than the followed BRA. MA had a faster LD decay rate than other chicken breeds, DZAC and WC have similar LD level with RJFt in second group (Fig 2C).

The level of ROH reflects the recent inbreeding history of a population [15]. As shown in Fig 2D, the average and total length of ROH of indigenous chicken are short, RJFt had the lowest number of ROH. The difference of mROH among Guangxi breeds is relatively small (ranging from 37.74 kb to 66.78kb), which is very different from the largest value of 167.3 kb in WL (S6 Table). HGFC (347.039Mb) had the same tROH level with BRA (338.084 Mb), and the shortest tROH was observed in LYWC (185.483 Mb).

**Fig2. Genome diversity and LD decay of 11 chicken populations.** (A) Genome nucleotide diversity was calculated with a window size of 40 kb and a step size of 20 kb. (B) Genome heterozygosity within 500kb sliding window across the genome. (C) Linkage disequilibrium (LD) decay, denoted with one line for each population. (D) The ROH of each chromosome in different breeds. The redder color represents longer ROH, the bluer the shorter.

# Population genetics analysis

As expected, the chickens from the same breeds clustered together according to the PCA. The PC1 (26.79% variances explained totally) could separate the commercial layer breed WL from other populations and PC2 (8.86% variances explained totally) displayed the genetic differentiation between commercial broiler breed BRA and other populations except HGFC (Fig 3A). The 12 RJFt gathered with Guangxi fowls and away from commercial chickens. When WL, BRA, RJFt and HGFC were removed from the dataset, DZAC and GDC could be identified as separate clusters, the two populations of LSFC, MC and LSFC, are distributed separately. DZAC and LSFC have a large variation within breed. MA, SHC and XYC are geographically close and tend to get closer to each other.

PCA results could not completely reproduce the phylogenetic relationships, and the neighbor-joining tree corroborates the findings of the PCA (Fig 3B). Individuals from the same breed gathered were consistent with the breeding history and geographical distribution. Cluster 1 consists of breeds of outgroup background (WL, BRA and RJFt), and the hybrid HGFC also was grouped together. In cluster 2, WC , NDYC and LSFC are grouped together. Following this group, the local populations DZAC, GDC and several MA were arranged in the middle of the tree, but were not forming a visually distinct cluster. Cluster 4 consists exclusively of chickens sampled in MA. Cluster 5 branched into two sub-clusters with SHC and XYC, an individual from the WC was grouped with XYC, possibly because of a sampling error.

**Fig3. Population structure analyses.** (A) Principal component analysis (PCA), with 26.79% and 8.86% variance explained in PC1 and PC2, respectively. (B) Neighbor-joining tree of 221 chickens, constructed with PHYLIPNEW version 3.69.650 (C) Admixture analysis with K values running from 3 to 6.

To assess historical admixture patterns of the chickens, we conducted the ADMIXTURE analysis with K values running from 3 to 18. At K=3, genetic divergency first occurred between commercial breeds and non-commercial ones. HGFC shared the same ancestral lineage with BRA , Guangxi indigenous breeds

shared the same ancestral lineage with RJFt, WL consistent with the above PCA and phylogenetic tree result (Fig 3C). When K=4, the Guangxi indigenous breeds were separated from others (except HGFC). Indigenous chickens gradually separated from each other when K ranged from 5 to 14. MA experienced introgression from SHC, and the ancestral components of SHC, GDC and HGFC are pure. There is differentiation in the breeding programs of LYWC and DLWC, as well as in LSFC and XYC breeds (S3 Fig). According to the calculated cross-validation value, the best fit was k=5, Guangxi indigenous breeds showed two ancestral components that are different from others.

# Selective Sweep Analysis

The skin color influences consumers' preferences, and the yellow skin chickens are more popular than white ones in south of China. We observed the highest ZFst region occurring at chr24: 6.14-6.18 Mb (Fst = 0.63, ZFst = 35.39) between 121 yellow skin and 64 non-yellow skin chickens. The top ten selected window annotated eight genes, among which *BCO2* gene is a classical yellow color gene in chicken. Then we collected the genotypes on this gene and found that the non-yellow skin clusters showed a different genotypic pattern from LSFC and NDYC; as for WC, it has both two patterns might due to the concealment of its black skin that the yellow skin has not been deliberately selected (S4 Fig). We found a missense mutation at chr24: 6155481T>C (rs313409504) was consisted with the previous report [16]. The strongest selective sweep on chromosome 11 we detected was located at 19.12-19.16Mb (Fst = 0.17, ZFst = 9.41) near the gene *MC1R*, which plays a key role in controlling the deposition of melanin.

**Fig4.** The result of ZFst and Log2(pi) of yellow skin.

In oriental countries, nutritional and medicinal benefits have been attributed to the consumption of black-boned chickens. WC is distinguished compared with other breeds because of its black beak, crown, skin, and shank. We compared WC with other Guangxi chickens and scanned the whole genome for selected regions. The strong selected region contained four annotated functional genes which were associated with dermal hyperpigmentation in chickens, beta-1 tubulin (*TUBB1*) and *PRELI* domain containing 3B (*PRELID3B*, also known as *SLMO2*), *GNAS* complex locus (*GNAS*), encoding endothelin 3 (*EDN3*). GO enrichment analysis shows that gene *TYRP1* on chromosome Z and *KITLG* on chromosome 1 were significantly enriched in melanocyte differentiation (GO:0030318) and developmental pigmentation (GO:0048066). *SYK* gene plays a role in regulation of bone resorption (GO:0045124).

XYC and GDC have a similar appearance with SHC, XYC is fat-deposited and GDC is smaller. We compared them with the SHC and scanned the whole genome for selected regions related to fat deposition and body size. We found that the selected genes of fat deposits were located on chromosome 12 and chromosome Z. *HMGCS1* and *OXCT1* are significantly enriched in the pathway of ketone body synthesis and degradation (gga00072), ketone bodies are produced in the liver, mainly from the oxidation of fatty acids, and are exported to peripheral tissues for use as an energy source. *ATG7* (Autophagy Related 7) is a protein coding gene (gga:04140), it has been associated with multiple functions, including axon membrane trafficking, axonal homeostasis, mitophagy, adipose differentiation, and hematopoietic stem cell maintenance. A significant signal peak was detected at 49.32-49.36Mb on chromosome 5 (ZFst=8.5, Fst = 0.51) located in the *DLK1-DIO3* genomic region, which was recognized to be an imprinted domain in placental mammals associated with developmental programming [17].

# Discussion

We assessed the genetic diversity of indigenous chickens from Guangxi provinces using the re-sequencing, meanwhile, we performed selective sweep analysis of phenotypes related to economic traits.

As shown in our study, most Guangxi chicken breed have various genetic diversity on the genome indicated by higher SNPs abundance, slower LD decay rate, lower ROH values and higher heterozygosity. HGFC, as a population of LSFC breed, is an exception, gene introgression from commercial broilers was detected. After further investigation, in order to improve the economic benefit of this group, broilers genes were introduced artificially in the process of breeding. The results of genetic structure are consistent with the origin of breeds, indicating the effectiveness of Guangxi local chicken population in breed protection , even the HGFC also formed its own characteristics because the ancestral composition of it is different from that of BRA when k=6. Compared with commercial breeds, the genetic difference among indigenous chickens is relatively small, and Guangxi indigenous fowls do have a closer relationship with RJFt [18].

The difference of deposition location and amount of carotenoid and melanin in chicken skin led to the diversity of chicken skin color. *BCO2* gene encodes beta-carotene dioxygenase 2 could cleave colorful carotenoids to colorless apocarotenoids by an asymmetric cleavage reaction [19], is established as the causal gene for the yellow skin. According to the results of selection scanning, *BCO2* gene is extremely strongly selected in the population, and the SNP shows different patterns in the yellow and non-yellow skin population. Eriksson et al. demonstrate that regulatory mutations that inhibit expression of *BCDO2* in skin caused yellow skin, but not in other tissues [16]. Fallahshahroudi's study showed the down-regulation of *BCO2* in skin, muscle, and adipose tissue was associated with the derived haplotype [20]. Also, *BCO2* has variety variants in different breeds. Wang found a G>A mutation in exon 6 to be associated with the concentration of carotenoids in Guangxi-huang and Qingjiao-ma chicken [21]. A GAG haplotype was fixed in commercial breeds of yellow skin [16]. We also found the missense mutant at chr24:6155481 led to the mutation of threonine to alanine.

The strongest selective sweep region with dermal hyperpigmentation on chromosome 20 was located at 10.64-10.94 Mb consists of a set of 7 genes. *EDN3* is a gene with a known role in promoting melanoblast proliferation by encoding a potent mitogen for melanoblasts/melanocytes. Shinomiya et al. reported that the overexpression of genes in a 130kb duplication region gives rise to the hyperpigmentation in silk chickens [22], and then Dorshorst et al. extended this discovery and strongly suggested that the increase of *EDN3* expression caused by duplication is the cause of *FM* in all breeds of chickens. The expression of two other genes, *SLMO2*, and *TUBB1* were also significantly increased in expression in both skin and muscle tissue from adult fibromelanosis chickens might contribute to the dermal hyperpigmentation phenotype [23]. Analysis of RNA-seq suggested that *SLMO2*, *ATP5e*, and *EDN3* were differentially expressed between the black and yellow skin groups, combined analysis of genomic data found that *EDN3* might interact with the upstream ncRNA *LOC101747896* to generate black skin color during melanogenesis [24]. Wang et al. study indicated that a T2270C mutation in *GNAS* gene promoter in chicken is correlated strongly with the skin color traits [25].

*DLK1* has been reported to be associated with the development and differentiation of adipose and muscle in chicken and *DLK1* promotes muscle development inhibitory

adipogenesis in mammals [26]. Zhang et al.'s study provides strong in vivo evidence that atg7, and by inference autophagy, is critical for normal adipogenesis [27]. *AMACR* coding protein alpha-methylacyl-CoA racemase, this protein is involved in the pathway bile acid biosynthesis, which is part of Lipid metabolism (gga00120). Bile acid is the main component of bile and its main function is to promote the digestion and absorption of fat. *HSD17B4* codes a bifunctional enzyme mediating dehydrogenation and anhydration during β-oxidation of long-chain fatty acids, and a non-synonymous SNP has been reported to be related to meat-quality traits in pig [28]. *PRKAA1* is associated with skeletal muscle lipid accumulation [29].

# Conclusions

In conclusion, a comprehensive analysis of genetic diversity, population genetic structure, LD, and the selection signatures of 8 indigenous chicken breeds distributed in Guangxi. The results suggested that indigenous chickens have abundant genetic diversity and potential, candidate genes related to economic traits can also provide a theoretical basis for breeding. Our analyses provide data for further research and local breeding of Guangxi indigenous chicken.

# Data availability statement

The raw reads data have been submitted to NCBI Sequence Read Archive database with the accession number PRJNA659069. The data reported in this study are also available in the CNGB Nucleotide Sequence Archive (CNSA: https://db.cngb.org/cnsa ; accession number CNP0001716). Furthermore, 36 individuals of commercial breeds and red jungle fowls were downloaded from NCBI at ERP112703 (S2 Table).

# Acknowledgments

We would like to thank the members of Guangxi key laboratory of livestock genetic improvement, BGI Institute of Applied Agriculture, BGI-Shenzhen, and Guangxi veterinary research institute for helpful input on the project.

# References:

1. Scherf BD, Fao R, Pilling D. The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture. 2015.
2. Liao Y, Mo G, Sun J, Wei F, Liao DJ. Genetic diversity of Guangxi chicken breeds assessed with microsatellites and the mitochondrial DNA D-loop region. Mol Biol Rep. 2016;43(5):415-25. http://doi.org/10.1007/s11033-016-3976-0
3. Yang Z, Deng J, Li D, Sun T, Xia L, Xu W, et al. Analysis of Population Structure and Differentially Selected Regions in Guangxi Native Breeds by Restriction Site Associated with DNA Sequencing. G3 (Bethesda). 2020;10(1):379-86. http://doi.org/10.1534/g3.119.400827
4. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95. http://doi.org/10.1093/bioinformatics/btp698
5. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8. http://doi.org/10.1093/bioinformatics/btr330
6. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464(7288):587-91. http://doi.org/10.1038/nature08832
7. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35(10):1786-8. http://doi.org/10.1093/bioinformatics/bty875
8. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. http://doi.org/10.1186/s13742-015-0047-8
9. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655-64. http://doi.org/10.1101/gr.094052.109
10. Qanbari S, Rubin CJ, Maqbool K, Weigend S, Weigend A, Geibel J, et al. Genetics of adaptation

in modern chicken. Plos Genet. 2019;15(4):e1007989. http://doi.org/10.1371/journal.pgen.1007989
11. Liu LL, Fang C, Liu WJ. Identification on novel locus of dairy traits of Kazakh horse in Xinjiang. Gene. 2018;677:105-10. http://doi.org/10.1016/j.gene.2018.07.009
12. Li D, Li Y, Li M, Che T, Tian S, Chen B, et al. Population genomics identifies patterns of genetic diversity and selection in chicken. Bmc Genomics. 2019;20(1):263. http://doi.org/10.1186/s12864-019-5622-4
13. Qanbari S, Rubin C, Maqbool K, Weigend S, Weigend A, Geibel J, et al. Genetics of adaptation in modern chicken. Plos Genet. 2019;15(4):e1007989. http://doi.org/10.1371/journal.pgen.1007989
14. Corl A, Ellegren H. The genomic signature of sexual selection in the genetic diversity of the sex chromosomes and autosomes. Evolution. 2012;66(7):2138-49. http://doi.org/10.1111/j.1558-5646.2012.01586.x
15. Traspov A, Deng W, Kostyunina O, Ji J, Shatokhin K, Lugovoy S, et al. Population structure and genome characterization of local pig breeds in Russia, Belorussia, Kazakhstan and Ukraine. Genet Sel Evol. 2016;48:16. http://doi.org/10.1186/s12711-016-0196-y
16. Eriksson J, Larson G, Gunnarsson U, Bed'Hom B, Tixier-Boichard M, Stromstedt L, et al. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. Plos Genet. 2008;4(2):e1000010. http://doi.org/10.1371/journal.pgen.1000010
17. Da RS, Edwards CA, Ito M, Ogata T, Ferguson-Smith AC. Genomic imprinting at the mammalian Dlk1-Dio3 domain. Trends Genet. 2008;24(6):306-16. http://doi.org/10.1016/j.tig.2008.03.011
18. Wang MS, Thakur M, Peng MS, Jiang Y, Frantz L, Li M, et al. 863 genomes reveal the origin and domestication of chicken. Cell Res. 2020;30(8):693-701. http://doi.org/10.1038/s41422-020-0349-y
19. Kiefer C, Hessel S, Lampert JM, Vogt K, Lederer MO, Breithaupt DE, et al. Identification and characterization of a mammalian enzyme catalyzing the asymmetric oxidative cleavage of provitamin A. J Biol Chem. 2001;276(17):14110-6. http://doi.org/10.1074/jbc.M011510200
20. Fallahshahroudi A, Sorato E, Altimiras J, Jensen P. The Domestic BCO2 Allele Buffers Low-Carotenoid Diets in Chickens: Possible Fitness Increase Through Species Hybridization. Genetics. 2019;212(4):1445-52. http://doi.org/10.1534/genetics.119.302258
21. Wang Y, Gan SY, Luo CL, Liu TF, Zhao XL, Ma J, et al. Variations in BCO2 Coding Sequence Causing a Difference in Carotenoid Concentration in the Skin of Chickens. Research Square. 2020.
22. Shinomiya A, Kayashima Y, Kinoshita K, Mizutani M, Namikawa T, Matsuda Y, et al. Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (Fibromelanosis) in silky chickens. Genetics. 2012;190(2):627-38. http://doi.org/10.1534/genetics.111.136705
23. Dorshorst B, Molin AM, Rubin CJ, Johansson AM, Stromstedt L, Pham MH, et al. A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. Plos Genet. 2011;7(12):e1002412. http://doi.org/10.1371/journal.pgen.1002412
24. Li D, Sun G, Zhang M, Cao Y, Zhang C, Fu Y, et al. Breeding history and candidate genes responsible for black skin of Xichuan black-bone chicken. Bmc Genomics. 2020;21(1):511. http://doi.org/10.1186/s12864-020-06900-8
25. Wang HH, Chen ML, Lou LF, Zhang L, Zhang CX, Chen XH, et al. GNAS gene promoter mutation in chicken and the correlation with skin color traits. Acta Veterinaria et Zootechnica Sinica. 2016;12(47):2354-61.
26. Shin J, Lim S, Latshaw JD, Lee K. Cloning and expression of delta-like protein 1 messenger ribonucleic acid during development of adipose and muscle tissues in chickens. Poult Sci. 2008;87(12):2636-46. http://doi.org/10.3382/ps.2008-00189
27. Zhang Y, Goldman S, Baerga R, Zhao Y, Komatsu M, Jin S. Adipose-specific deletion of autophagy-related gene 7 (atg7) in mice reveals a role in adipogenesis. Proc Natl Acad Sci U S A. 2009;106(47):19860-5. http://doi.org/10.1073/pnas.0906048106
28. Jo JL, Hwang JH, Kwon SG, Park DH, Kim TW, Kang DG, et al. Association between a non-synonymous HSD17B4 single nucleotide polymorphism and meat-quality traits in Berkshire pigs. Genet Mol Res. 2016;15(4). http://doi.org/10.4238/gmr15048970
29. Wu W, Xu Z, Zhang L, Liu J, Feng J, Wang X, et al. Muscle-specific deletion of Prkaa1 enhances skeletal muscle lipid accumulation in mice fed a high-fat diet. J Physiol Biochem. 2018;74(2):195-205. http://doi.org/10.1007/s13105-017-0604-y

# Supporting information

**S1 Fig. Geographic distribution and appearances of typical female chickens.**

**S2 Fig. Boxplot showing heterozygous SNP rate of autosomes (left) and Z**

**chromosome (right) between each chicken population.**

**S3 Fig. Admixture analysis with K values running from 7 to 18.** Each population separated by white dotted line.

**S4 Fig. The genotype of fixed SNPs in chr24: 6.14Mb~6.18Mb of individuals.** The row represents the SNP position and the column represents the individual. Light blue denotes reference alleles while red indicates alternative homozygous alleles, yellow means heterozygous and dark blue means missing.

**S5 Fig. ZFst values and Log 2 (pi).** (A) WC and other indigenous population. (B) XYC and SHC. (C) GDC and SHC

**S1 Table. The character of chickens in this study.**

**S2 Table. The public data information.**

**S3 Table. The sequencing information of samples.**

**S4 Table. The distribution of variant, nucleotide diversity and HE.**

**S5 Table. The distribution of isochrones.**

**S6 Table. ROH of chicken breeds.**

**S7 Table. Selective sweep of dermal hyperpigmentation.**

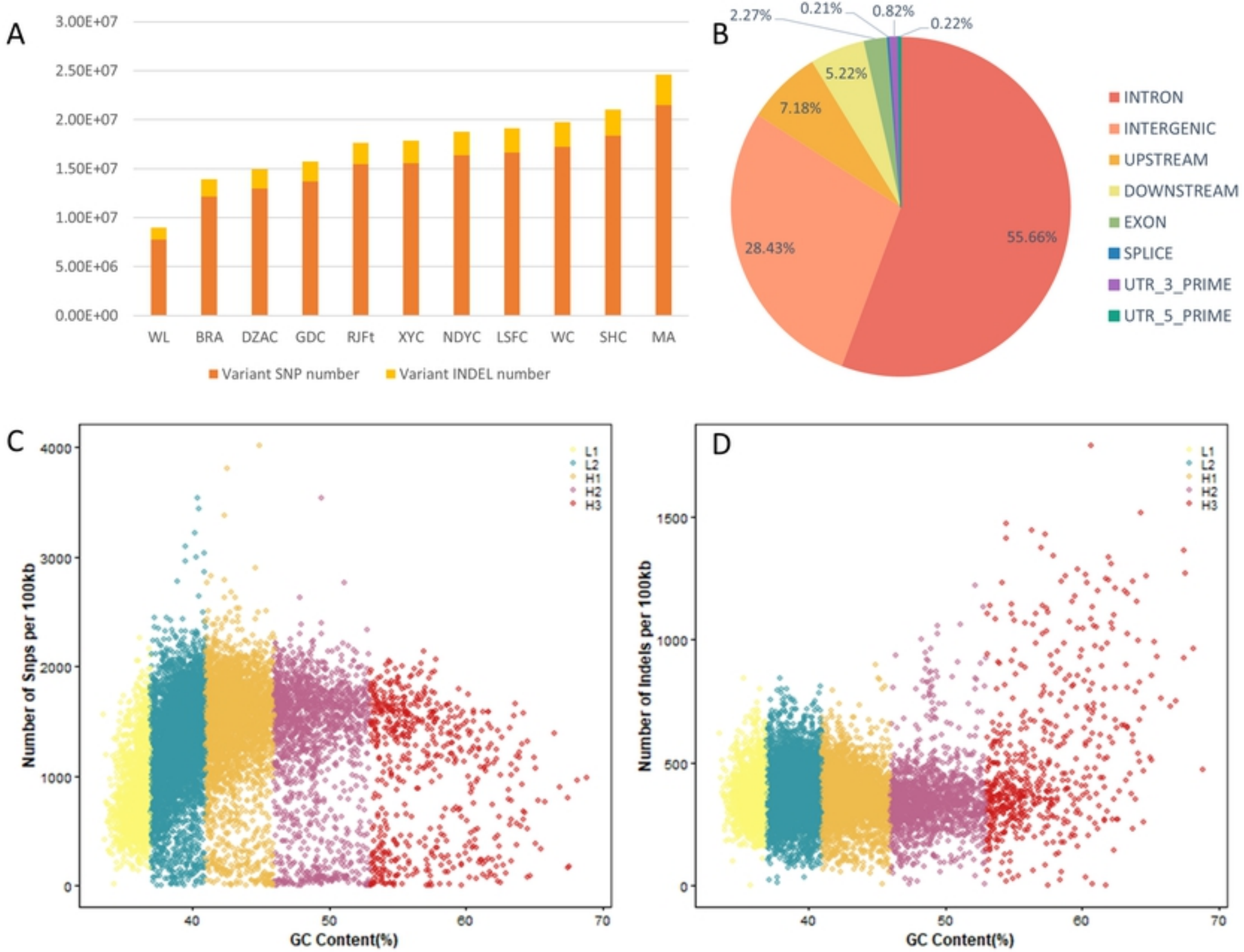
**S8 Table. Selective sweep of yellow skin.**

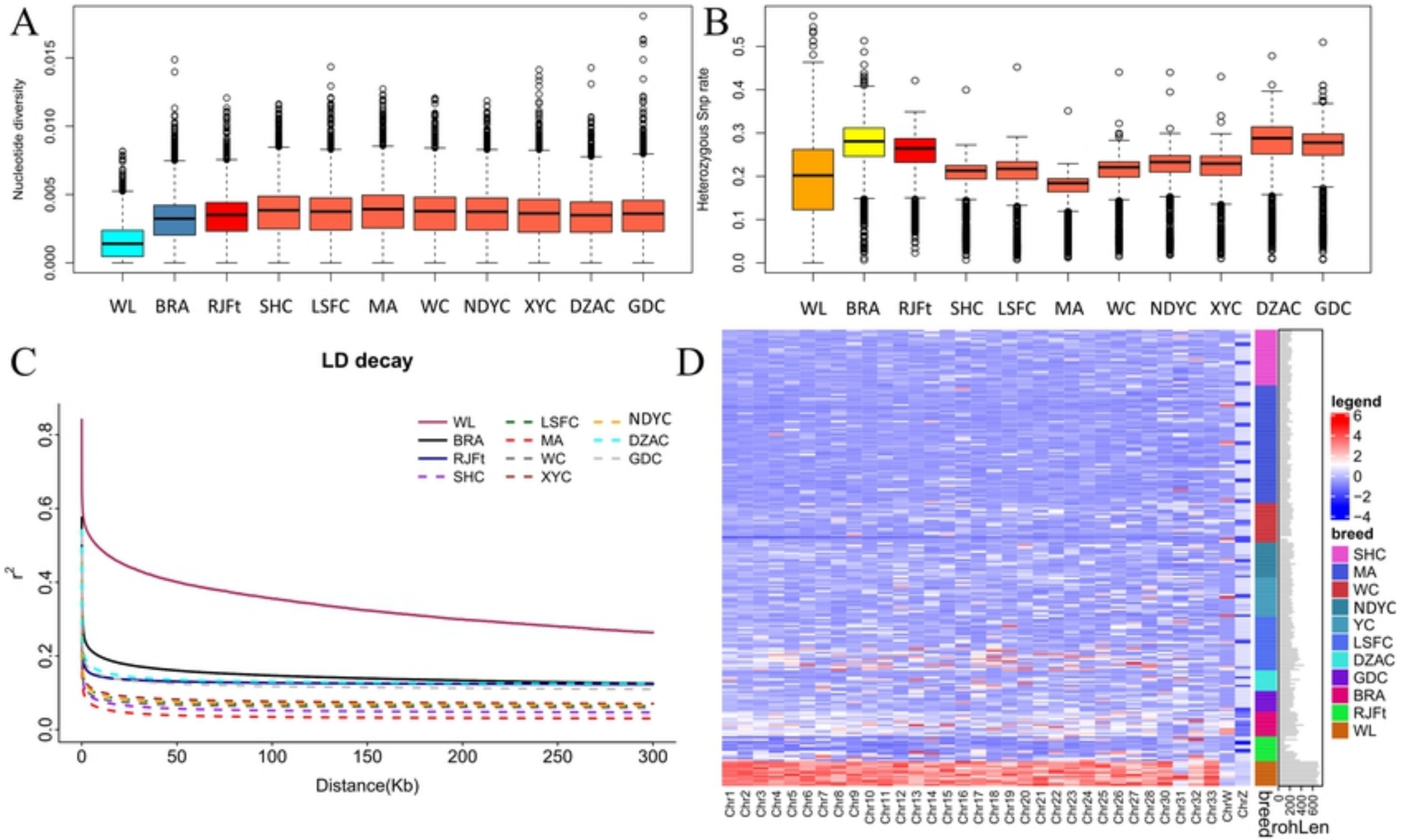
**S9 Table. Selective sweep of body size.**

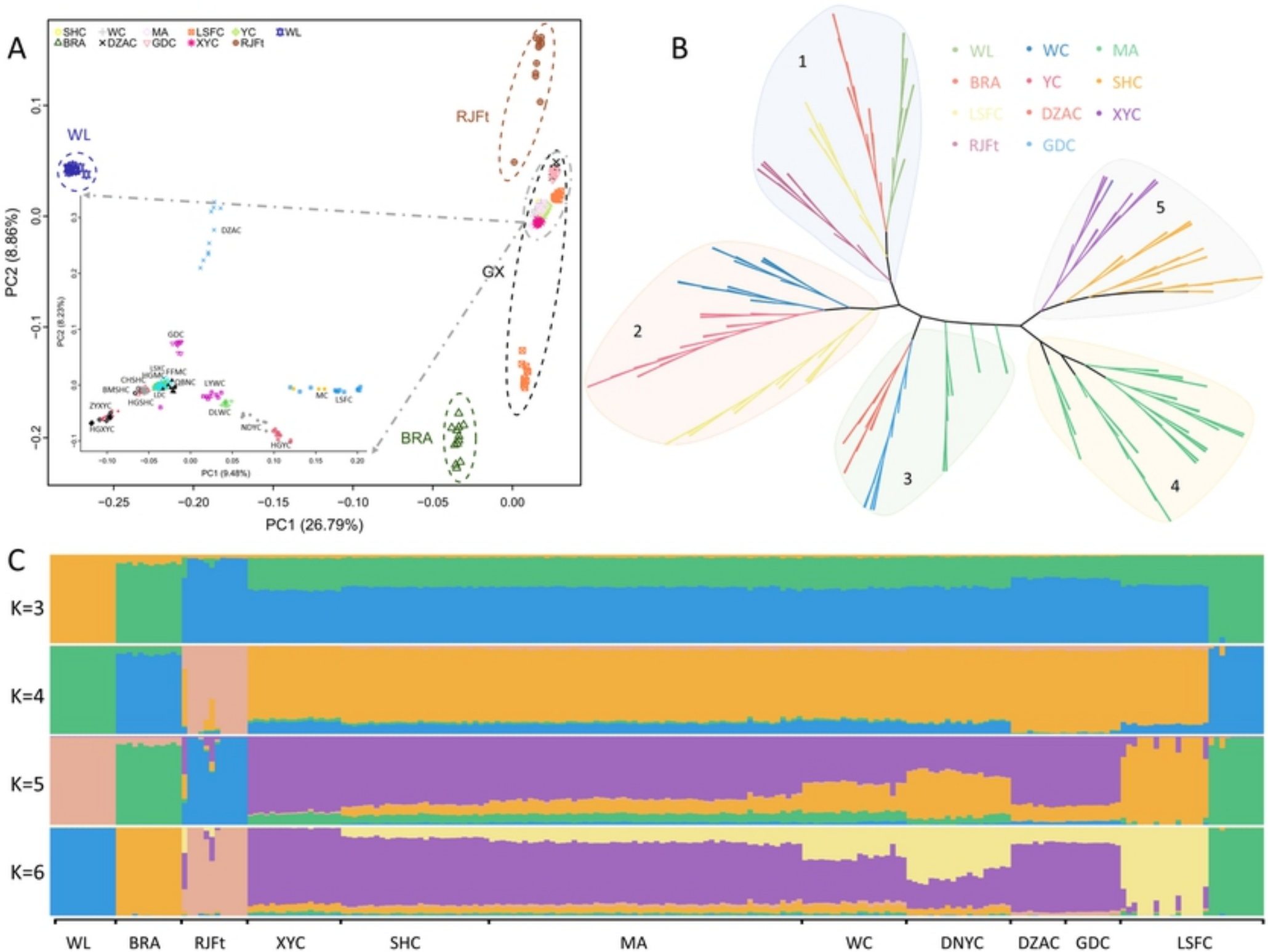
**S10 Table. Selective sweep of fat deposition.**

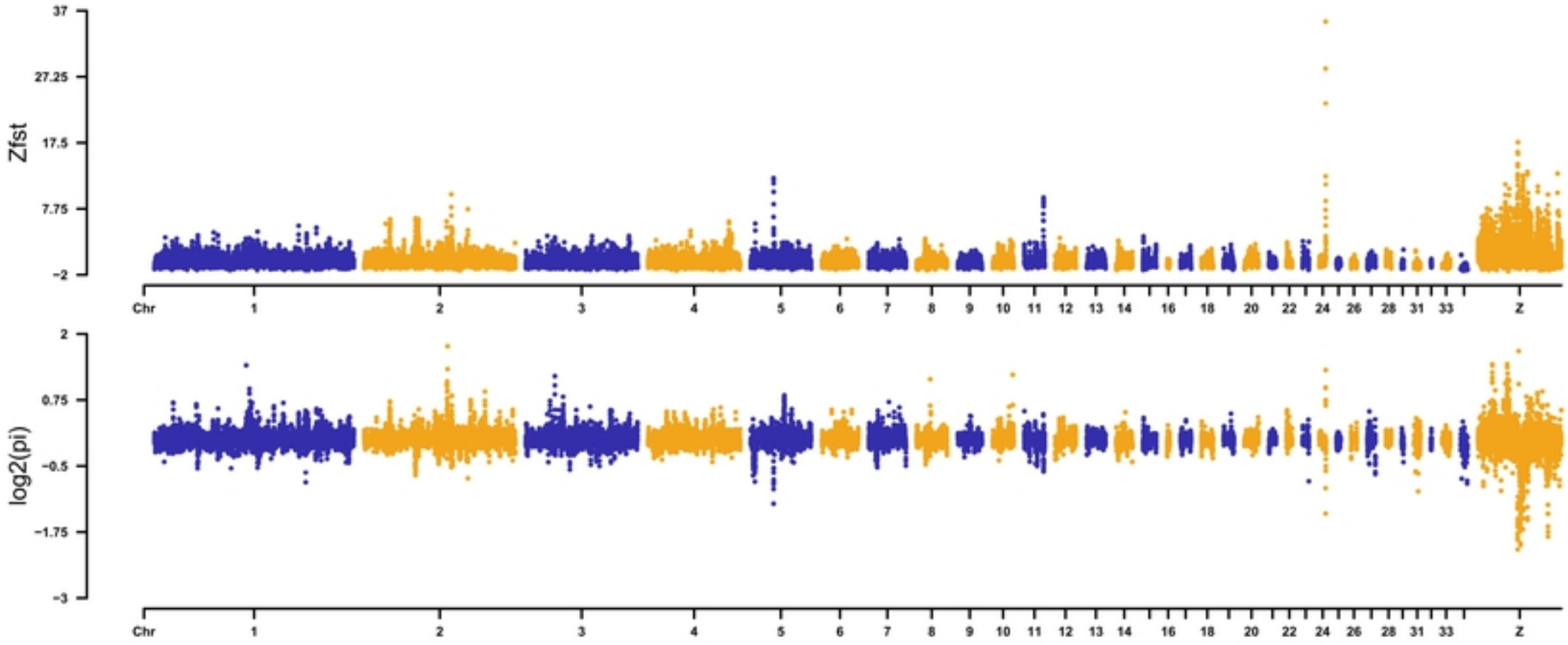
**S11 Table. Functional gene categories enriched for genes under selection.**

**S12 Table. KEGG pathway analysis of genes.**


A
3.00E+07
2.50E+07
2.00E+07
1.50E+07
1.00E+07
5.00E+06
0.00E+00
WL
BRA
DZAC
GDC
RJFt
XYC
NDYC
LSFC
WC
SHC
MA
Variant SNP number
Variant INDEL number
B
2.27%
0.21%
0.82%
0.22%
5.22%
7.18%
28.43%
55.66%
INTRON
INTERGENIC
UPSTREAM
DOWNSTREAM
EXON
SPLICE
UTR_3_PRIME
UTR_5_PRIME
C
Number of Snps per 100kb
GC Content(%)
L1
L2
H1
H2
H3
D
Number of Indels per 100kb
GC Content(%)
L1
L2
H1
H2
H3


A
Nucleotide diversity
0.000
0.005
0.010
0.015
WL BRA RJFt SHC LSFC MA WC NDYC XYC DZAC GDC
B
Heterozygous Snp rate
0.0
0.1
0.2
0.3
0.4
0.5
WL BRA RJFt SHC LSFC MA WC NDYC XYC DZAC GDC
C
LD decay
WL
BRA
RJFt
SHC
LSFC
MA
WC
XYC
NDYC
DZAC
GDC
r2
0.0
0.2
0.4
0.6
0.8
0 50 100 150 200 250 300
Distance(Kb)
D
legend
6
4
2
0
−2
−4
breed
SHC
MA
WC
NDYC
YC
LSFC
DZAC
GDC
BRA
RJFt
WL
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16 Chr17 Chr18 Chr19 Chr20 Chr21 Chr22 Chr23 Chr24 Chr25 Chr26 Chr27 Chr28 Chr30 Chr31 Chr32 Chr33 ChrW ChrZ
breed
rohLen
0 200 400 600


A
SHC
WC
MA
LSFC
YC
WL
BRA
DZAC
GDC
XYC
RJFt
WL
RJFt
GX
BRA
PC2 (8.86%)
PC1 (26.79%)
PC2 (8.23%)
PC1 (9.48%)
B
WL
WC
MA
BRA
YC
SHC
LSFC
DZAC
XYC
RJFt
GDC
1
2
3
4
5
C
K=3
K=4
K=5
K=6
WL
BRA
RJFt
XYC
SHC
MA
WC
DNYC
DZAC
GDC
LSFC


Zfst
37
27.25
17.5
7.75
−2
Chr
1
2
3
4
5
6
7
8
9
10
11
12
13
14
16
18
20
22
24
26
28
31
33
Z
log2(pi)
2
0.75
−0.5
−1.75
−3

Figure