Received: 28 April 2023 | Accepted: 21 September 2023

DOI: 10.1111/age.13364

SHORT COMMUNICATION

ANIMAL GENETICS WILEY

Detection of genomic variations and selection signatures in Wagyu using whole-genome sequencing data

Abstract

Wagyu is recognized for producing marbled beef with high nutritional value and flavor. Reportedly, Wagyu has been widely used to improve the meat quality of local breeds around the world. However, studies on the genetic mechanism of meat quality in Wagyu at the whole-genome level are rarely reported. Here, whole-genome sequencing data of 11 Wagyu and 115 other individuals were used to explore the genomic variations and genes under selection pressure in Wagyu. A total of 31 349 non-synonymous variants and 53 102 synonymous variants were identified in Wagyu. The population structure analysis showed that Wagyu had the closest genetic relationship with Mishima–Ushi cattle and was apparently separated from other cattle breeds. Then, composite likelihood ratio (CLR), integrated haplotype score, fixation index and cross-population composite likelihood ratio (XP-CLR) tests were performed to identify the candidate genes under positive selection in Wagyu. In total, 770 regions containing 312 genes were identified by at least three methods. Among them, 97 regions containing 27 genes were detected by all four methods. We specifically illustrate a list of interesting genes, including *LRP2BP*, *GAA*, *CACNG6*, *CXADR*, *GPCPD1*, *KLF2*, *KLF13*, *SOX5*, *MYBPC1*, *SLC25A10*, *ATP8A1* and *MYH15*, which are associated with lipid metabolism, fat deposition, muscle development, bone development, feed intake and growth traits in Wagyu. This is the first study to explore the genomic variations and selection signatures of Wagyu at the whole-genome level. These results will provide significant help to beef cattle improvement and breeding.

Beef has been of great interest as a high-quality dietary protein. One of the criteria for evaluating the quality of beef quality is marbling (Platter et al., 2005). Wagyu is well known for its ability to produce high-quality marbled beef (Gotoh et al., 2018). It is reported that the number of Wagyu raised in Australia, New Zealand, China and other countries reached about 400 000 cattle in 2020 (Ueda et al., 2022). In China, the frozen semen of Wagyu is used for crossing with native breeds to produce crossbred offspring with good meat quality traits (Li et al., 2013; Wang et al., 2019). Previous studies used SNP arrays (Ishii et al., 2013; Kawaguchi et al., 2022) and transcriptome sequencing (Fang et al., 2021) to identify molecular markers and genes that correlated with meat quality traits. However, no studies have used whole-genome sequencing (WGS) data to identify genomic variations in Wagyu. In this study, a comprehensive analysis of genomic variations and selective signatures was performed using WGS data for Wagyu.

The sperm ($n=9$) and blood ($n=2$) samples were from the commercial bull stations. DNA was extracted using the EasyPure Blood Genomic DNA Kit (TransGen Biotech) for WGS. A total of 428.11 Gb raw data were obtained from DNBSEQ-T7 platform (Table S1). Additionally, the genomic data of 115 individuals were downloaded from the NCBI SRA database (Table S2). All clean reads were mapped to the *Bos taurus* reference assembly ARS-UCD1.2 using BWA-MEM v0.7.13 (Li & Durbin, 2010), the average depth and the average mapping ratio were 14.93× and 98.95% (Tables S1 and S2). The GATK v4.4.1 (McKenna et al., 2010) pipeline was used to perform downstream SNP calling, and the details are provided in Appendix S1. Overall, 22 935 485 high-quality autosomal bi-allelic SNPs were obtained.

There were a total of 16 171 320 (70.51%) transitions (Ts) and 6 764 165 (29.49%) transversions (Tv) in all SNPs. For the Wagyu in this study, 10 155 213 autosomal bi-allelic SNPs were discovered. The average variant rate was 0.41% (this means that there is variant per 245 bases) and the Ts/Tv ratio was 2.35 in Wagyu. In all, 2.03% of the SNPs (containing 31 349 non-synonymous variants and 53 102 synonymous variants) were detected in exons, 49.59% in intergenic, 38.16% in introns and 10.22% upstream and downstream of genes (Table S3 and Figure S1).

To determine the population structure of Wagyu, the parameter (--indep-pair-wise 50 25 0.2) in PLINK

Lulu Shi and Mingyue Hu made same contribution to the work and share the first authorship.

v1.9 (Purcell et al., 2007) was used to prune the SNPs with high-level pair-wise linkage disequilibrium. Then 1 975 542 SNPs were used to conduct the admixture analysis using ADMIXTURE (Alexander et al., 2009) with the ancestral compositions (K) set from 2 to 8. The results showed that Wagyu has the greatest ancestry composition from Mishima-Ushi when K=3–4 (Figure 1a). The phylogenetic tree (Figure 1b) reconstructed using the neighbor-joining method in MEGAX (Kumar et al., 2018) and PCA (Figure S2) performed using GCTA v1.92.3 (Yang et al., 2013) showed similar results, with Wagyu having a closer genetic relationship with Mishima-Ushi than with other breeds.

To identify the genomic regions that are under selection in Wagyu, four complementary methods were used to detect the selection signatures. BEAGLE v5.1 (Ayres et al., 2012) was applied to impute and phase our data. After that, the SELSCAN v1.4.0 (Szpiech & Hernandez, 2014) was used to compute the raw integrated haplotype score (iHS) across the genome. The scores were then normalized using the companion program norm of selscan and the fraction of |iHS|>2 was calculated

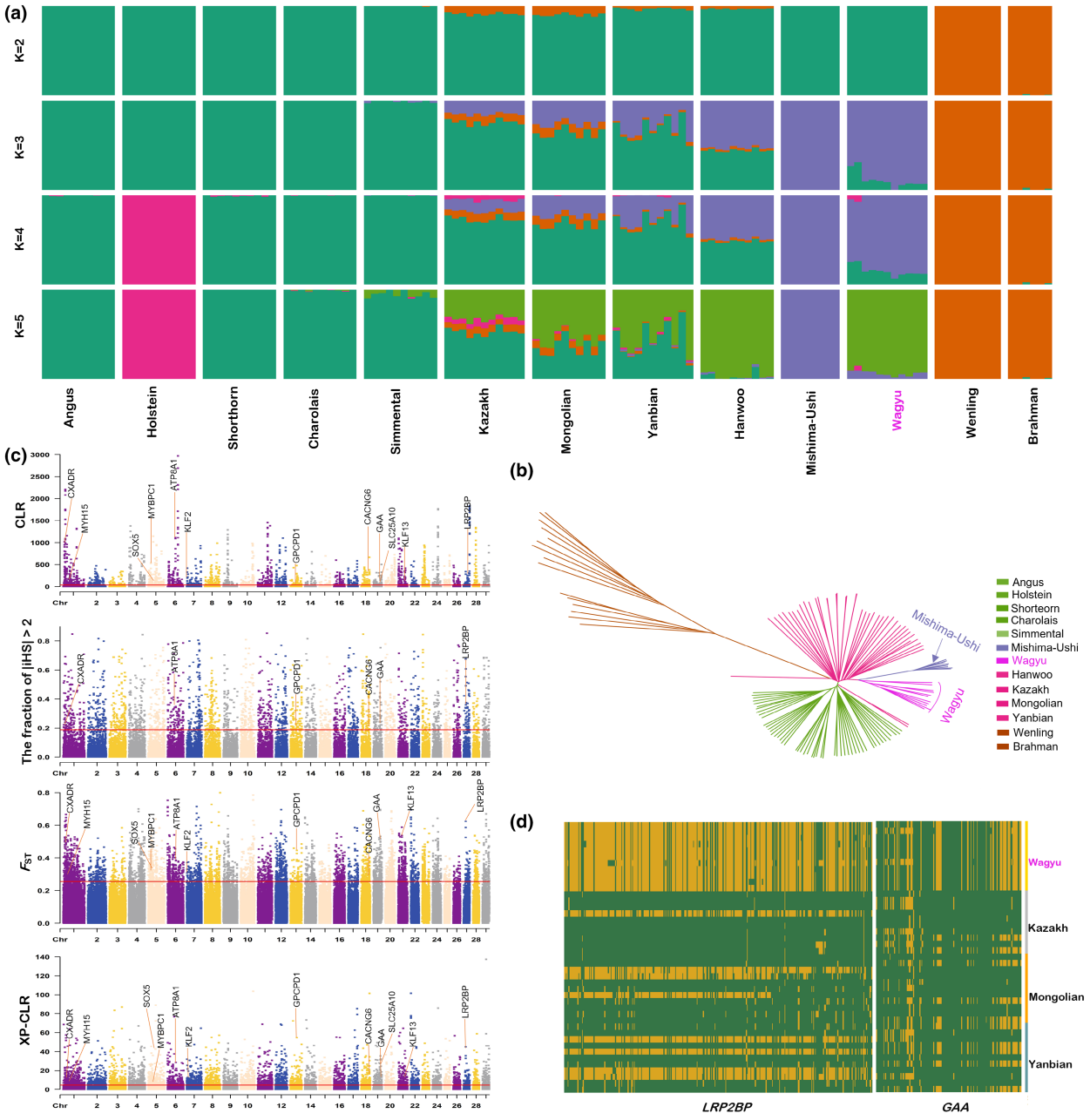

FIGURE 1 Population structure and selection signatures of Wagyu. (a) Model-based clustering of cattle breeds using ADMIXTURE with K=2–5. (b) Neighbor-joining tree of the relationship between 13 cattle breeds. (c) Manhattan plot of selective sweeps used composite likelihood ratioR, integrated haplotype score, fixation index and cross-population composite likelihood ratio tests in Wagyu. The red line means the thresholds of the top 5% of windows with extreme values. (d) Haplotype patterns of *GAA* and *LRP2BP* genes.

in each 50 kb window. SWEED (Pavlidis et al., 2013) was used to perform the composite likelihood ratio (CLR) with 50 kb windows. The fixation index (F_{ST}) and cross-population composite likelihood ratio (XP-CLR) test were used to perform comparisons between Wagyu and Chinese indigenous breeds (Mongolian, Kazakh, and Yanbian). The F_{ST} was calculated using VCFTOOLS (Danecek et al., 2011) with 50 kb windows and 25 kb steps. The xpclr (https://github.com/hardingnj/xpclr) software was used to conduct XP-CLR analysis with the same sliding windows strategy. Then, the top 5% of windows with extreme values were considered as potential candidate regions in four methods (Figure 1c). To reduce the false-positive results of the potential candidate regions, we only focused on the regions detected by at least three methods. The candidate regions were annotated using BIOMART tools (http://asia.ensembl.org/index.html).

A total of 312 genes were obtained in the 770 candidate regions which were identified by at least three methods (Table S4). The genes identified by at least three methods were used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. From the KOBAS enrichment results, there are 15 and 2 significant (corrected $p<0.05$) GO terms and KEGG pathways (Table S5), respectively. The most significant GO term and KEGG pathway were 'Synapse assembly, GO:0007416' and 'Calcium signaling pathway, bta04020', respectively.

It is noteworthy that there are 97 candidate regions containing 27 genes that were detected by all four methods (Table 1; Table S6). Among them, *LRP2BP* took part in the regulation of lipid metabolism (May et al., 2007); *GAA* was associated with the regulation of adipose tissues (Zhu et al., 2018); *CACNG6* and *CXADR* were related to feed efficiency (de Oliveira et al., 2014; Olivieri et al., 2016); and *GPCPD1* was a fatty acid metabolism-related gene in chicken (Yuan et al., 2022). A haplotype comparison analysis of *LRP2BP* and *GAA* indicated an obvious differentiation between Wagyu and Chinese indigenous breeds (Figure 1d). It is worth noting that the frequency of two missense variants (BTA19:g.52494418C>T and BTA19:g.52495927A>G) in *GAA* showed significant differences in Wagyu and Chinese indigenous breeds (Figure S3a,b).

In the genes detected by three methods, some genes have important functions (Table 1). *KLF2* and *KLF13* as members of Krüppel-like factor family were considered as positive and negative organizers in the context of mammalian adipogenesis (Raza et al., 2022). *SLC25A10* was associated with fat deposition in pigs (Chen et al., 2019). The *SOX5* gene was necessary for the formation and development of bones (Lefebvre et al., 2001). Zhang et al. demonstrated that the copy number variation of *SOX5* was associated with growth traits in yak (Zhang et al., 2022). *MYBPC1*, one of the subtypes of myosin-binding protein C, has been demonstrated to be located within genomic regions of quantitative trait loci (QTL) that are associated with growth-related traits (Offer et al., 1973; Takasuga et al., 2007). *ATP8A1* is considered to be associated with feed intake, feed conversion ratio and weight gain in African Ankole cattle and Dehong humped cattle (Li et al., 2020; Taye et al., 2017). Zhang et al. found that the myosin heavy chain family gene (*MYH15*) was highly expressed in high-weight chickens, indicating that it is associated with growth and development (Zhang et al., 2019).

Furthermore, the defined candidate regions were mapped to the latest cattle QTL database (release 49,

TABLE 1 Important genes within the candidate regions of selection in Wagyu.

Chromosome	Position (bp)[a]	Gene	Methods	Function
1	19 500 001-19 525 000	*CXADR*	CLR, iHS, XP-CLR, F_{ST}	Feed efficiency
1	46 875 001–46 900 001	*ZPLD1*	CLR, iHS, XP-CLR, F_{ST}	Obesity
1	53 225 001–53 250 001	*MYH15*	CLR, F_{ST}, XP-CLR	Growth and development
5	65 450 001–65 475 000	*MYBPC1*	CLR, F_{ST}, XP-CLR	Growth
5	86 225 001–86 228 571	*SOX5*	CLR, F_{ST}, XP-CLR	Growth
6	61 476 401–61 500 000	*ATP8A1*	CLR, F_{ST}, XP-CLR	Feed intake
7	6 675 001–6 700 000	*KLF2*	CLR, F_{ST}, XP-CLR	Adipogenesis
13	47 700 001–47 725 000	*GPCPD1*	CLR, iHS, XP-CLR, F_{ST}	Lipid metabolism
13	60 198 382–60 200 001	*ANGPT4*	CLR, iHS, XP-CLR, F_{ST}	Body weight
18	61 664 273–61 675 000	*CACNG6*	CLR, iHS, XP-CLR, F_{ST}	Feed efficiency
19	51 100 001–51 144 030	*SLC25A10*	CLR, F_{ST}, XP-CLR	Fat deposition
19	52 494 353–52 500 000	*GAA*	CLR, iHS, XP-CLR, F_{ST}	Regulation of adipose tissues
21	27 700 001–27 725 000	*KLF13*	CLR, F_{ST}, XP-CLR	Adipogenesis
27	15 600 001–15 623 771	*LRP2BP*	CLR, iHS, XP-CLR, F_{ST}	Fat deposition

Abbreviations: CLR, Composite likelihood ratio; F_{ST}, fixation index; iHS, integrated haplotype score; XP-CLR, cross-population composite likelihood ratio.
[a]This column presents the position of candidate genes that are within or overlap with the potential regions of selection.

28 December 2022). There were 3284 cattle QTLs overlapped with the candidate regions (Table S7). Notably, 50.37% of these QTLs were associated with production, meat quality, carcass and reproduction traits, suggesting that these traits were selected during the breeding of Wagyu and left a trace on the genome.

In conclusion, a comprehensive overview of sequence variations in Wagyu was provided using WGS data. Four complementary methods (F_{ST}, CLR, iHS and XP-CLR) were applied to identify the potential candidate regions. A total of 312 genes supported by at least three methods were detected. Our results provided some candidate genes including *LRP2BP*, *GAA*, *CACNG6*, *CXADR*, *GPCPD1*, *KLF2*, *KLF13*, *SOX5*, *MYBPC1*, *SLC25A10*, *ATP8A1* and *MYH15*, which were associated with lipid metabolism, fat deposition, muscle development, bone development, feed intake and growth traits in Wagyu. We provided the first study to explore the genomic variations and selection signatures of Wagyu at the whole-genome level. Our results will provide a valuable genomic resource for the improvement and breeding of beef cattle in the future.

KEYWORDS
candidate gene, genomic variation, selection signature, Wagyu

ACKNOWLEDGEMENTS
This study was supported by the Natural Science Foundation of Jilin Province (no. 20200201035JC).

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Sequences are available from GenBank with the Bioproject accession number PRJNA932924.

Lulu Shi[1]
Mingyue Hu[1]
Weining Lai[1]
Wenfeng Yi[1]
Zhengxi Liu[1]
Hao Sun[1]
Feng Li[2]
Shouqing Yan[1]

[1]*College of Animal Science, Jilin University, Changchun, China*
[2]*Shandong Binzhou Animal Science and Veterinary Medicine Academy, Binzhou, China*

Correspondence
Shouqing Yan, College of Animal Science, Jilin University, Changchun 130062, China.
Email: yansq@jlu.edu.cn

Feng Li, Shandong Binzhou Animal Science and Veterinary Medicine Academy, Binzhou 256600, China.
Email: lf951207@163.com

ORCID
Lulu Shi https://orcid.org/0000-0001-5762-770X
Hao Sun https://orcid.org/0000-0002-0542-9477
Shouqing Yan https://orcid.org/0000-0001-9321-3790

REFERENCES
Alexander, D.H., Novembre, J. & Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.

Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O. et al. (2012) BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Systematic Biology*, 61, 170–173.

Chen, G.T., Cheng, X.F., Shi, G.L., Zou, C., Chen, L., Li, J.X. et al. (2019) Transcriptome analysis reveals the effect of long intergenic noncoding RNAs on pig muscle growth and fat deposition. *BioMed Research International*, 2019, 2951427.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., De Pristo, M.A. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

de Oliveira, P.S., Cesar, A.S., do Nascimento, M.L., Chaves, A.S., Tizioto, P.C., Tullio, R.R. et al. (2014) Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genetics*, 15, 100.

Fang, X.B., Xia, L.X., Yu, H.B., He, W., Bai, Z.T., Qin, L.H. et al. (2021) Comparative genome-wide alternative splicing analysis of longissimus dorsi muscles between Japanese Black (Wagyu) and Chinese Red Steppes cattle. *Frontiers in Veterinary Science*, 8, 634577.

Gotoh, T., Nishimura, T., Kuchida, K. & Mannen, H. (2018) The Japanese wagyu beef industry: current situation and future prospects – a review. *Asian-Australasian Journal of Animal Sciences*, 31, 933–950.

Ishii, A., Yamaji, K., Uemoto, Y., Sasago, N., Kobayashi, E., Kobayashi, N. et al. (2013) Genome-wide association study for fatty acid composition in Japanese black cattle. *Animal Science Journal*, 84, 675–682.

Kawaguchi, F., Nakamura, M., Kobayashi, E., Yonezawa, T., Sasazaki, S. & Mannen, H. (2022) Comprehensive assessment of genetic diversity, structure, and relationship in four Japanese cattle breeds by Illumina 50 K SNP array analysis. *Animal Science Journal*, 93, e13770.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547–1549.

Lefebvre, V., Behringer, R.R. & de Crombrugghe, B. (2001) L-Sox5, Sox6 and Sox9 control essential steps of the chondrocyte differentiation pathway. *Osteoarthritis and Cartilage*, 9, S69–S75.

Li, H. & Durbin, R. (2010) Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*, 26, 589–595.

Li, R., Li, C., Chen, H., Li, R., Chong, Q., Xiao, H. et al. (2020) Genome-wide scan of selection signatures in Dehong humped cattle for heat tolerance and disease resistance. *Animal Genetics*, 51, 292–299.

Li, Y.B., Ni, S.H., Chen, B.X., Feng, F. & Chen, M.L. (2013) Study on the fattening experiment of wagyu and leiqiong crossbreed cattle. *Acta Ecologiae Animalis Domastici*, 34, 24–28.

May, P., Woldt, E., Matz, R.L. & Boucher, P. (2007) The LDL receptor-related protein (LRP) family: an old family of proteins with new physiological functions. *Annals of Medicine*, 39, 219–228.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A. et al. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303.

Offer, G., Moos, C. & Starr, R. (1973) A new protein of the thick filaments of vertebrate skeletal myofibrils. Extractions, purification and characterization. *Journal of Molecular Biology*, 74, 653–676.

Olivieri, B.F., Mercadante, M.E., Cyrillo, J.N., Branco, R.H., Bonilha, S.F., de Albuquerque, L.G. et al. (2016) Genomic regions associated with feed efficiency indicator traits in an experimental Nellore cattle population. *PLoS One*, 11, e0164390.

Pavlidis, P., Zivkovic, D., Stamatakis, A. & Alachiotis, N. (2013) SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, 30, 2224–2334.

Platter, W.J., Tatum, J.D., Belk, K.E., Koontz, S.R., Chapman, P.L. & Smith, G.C. (2005) Effects of marbling and shear force on consumers' willingness to pay for beef strip loin steaks. *Journal of Animal Science*, 83, 890–899.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559–575.

Raza, S.H.A., Pant, S.D., Wani, A.K., Mohamed, H.H., Khalifa, N.E., Almohaimeed, H.M. et al. (2022) Kruppel-like factors family regulation of adipogenic markers genes in bovine cattle adipogenesis. *Molecular and Cellular Probes*, 65, 101850.

Szpiech, Z.A. & Hernandez, R.D. (2014) Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, 31, 2824–2827.

Takasuga, A., Watanabe, T., Mizoguchi, Y., Hirano, T., Ihara, N., Takano, A. et al. (2007) Identification of bovine QTL for growth and carcass traits in Japanese black cattle by replication and identical-by-descent mapping. *Mammalian Genome*, 18, 125–136.

Taye, M., Kim, J., Yoon, S.H., Lee, W., Hanotte, O., Dessie, T. et al. (2017) Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. *BMC Genetics*, 18, 11.

Ueda, S., Takashima, Y., Gotou, Y., Sasaki, R., Nakabayashi, R., Suzuki, T. et al. (2022) Application of mass spectrometry for determining the geographic production area of wagyu beef. *Metabolites*, 12, 777.

Wang, L.M., Wang, D.B., Wang, X.D., Kang, L.H., Guo, T.L., Te, M.Q. et al. (2019) Comparative analysis of beef quality between Japanese wagyu and Simmental crossbreed cattle and pure Simmental cattle. China cattle. *Science*, 45, 17–20+57.

Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. (2013) Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods in Molecular Biology*, 1019, 215–236.

Yuan, X., Cui, H., Jin, Y., Zhao, W., Liu, X., Wang, Y. et al. (2022) Fatty acid metabolism-related genes are associated with flavor-presenting aldehydes in Chinese local chicken. *Frontiers in Genetics*, 13, 902180.

Zhang, Z.L., Chu, M., Bao, Q., Bao, P.J., Guo, X., Liang, C.N. et al. (2022) Two different copy number variations of the SOX5 and SOX8 genes in yak and their association with growth traits. *Animals*, 12, 1587.

Zhang, Z.R., Du, H.R., Yang, C.W., Li, Q.Y., Qiu, M.H., Song, X.Y. et al. (2019) Comparative transcriptome analysis reveals regulators mediating breast muscle growth and development in three chicken breeds. *Animal Biotechnology*, 30, 233–241.

Zhu, J., Jin, J., Ding, J.X., Li, S.Y., Cen, P.P., Wang, K.Y. et al. (2018) Ganoderic acid a improves high fat diet-induced obesity, lipid accumulation and insulin sensitivity through regulating SREBP pathway. *Chemico-Biological Interactions*, 290, 77–87.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.