BMC Genomics

RESEARCH Open Access

Design and verification of a 25 K multiple-SNP liquid-capture chip by target sequencing for dairy goat

Jianqing Zhao[1], Yaling Wang[1], Jiayidaer Kamalibieke[1], Ping Gong[2], Fuhong Zhang[1], Huaiping Shi[1], Wei Wang[1] and Jun Luo[1*]

Abstract

Background In the genetic breeding research of dairy goats, traditional genotyping methods have limitations, and existing goat chips have shortcomings in functional loci and other aspects, which cannot meet the precise genetic analysis needs of dairy goats. Genotyping by Target Sequencing (GBTS) in the new generation of sequencing technology provides the possibility to solve these problems.

Results A large number of candidate SNP sites related to important economic traits in dairy goats were identified through various analysis and screening methods. The chip ultimately retained 27,396 SNP sites for probe design, which can detect 46,459 SNPs. The site distribution is uniform, and the sequencing data efficiency, base quality, alignment rate, and other indicators are good. The chip SNP detection rate is high and the heterozygosity of gene typing is reasonable. GWAS was performed on 200 dairy goats for litter size and birth weight traits, and multiple genome-wide significantly related SNPs and related genes (litter size trait: *SCAP*, *PTPN23*, *KIF9*, *ANTXRL*, and *GRID1*. birth weight trait: *NALCN*, *LRRN2*, *TMEM132D*, *COL5A2*, and *HS3ST1*) were detected.

Conclusion The 25 K multiplex SNP liquid phase capture chip designed in this study has excellent performance and is of great value for genetic research and breeding of dairy goats, providing strong support for the development of the dairy goat industry.

Keywords Dairy goat, Chip design, Genotyping by target sequencing, Liquid chip

Introduction

In the genetic breeding of dairy goats, advancements in genotyping technologies are pivotal for driving industrial progress [1]. Current large-scale SNP genotyping primarily relies on three approaches: Genotyping by Sequencing (GBS), whole-genome sequencing (WGS), and array-based methods [2–4]. Among these, commercial goat SNP chips (e.g., the International Goat Genome Consortium's medium-density array) [5, 6] face critical limitations. First, their solid-state chip format incurs high production costs and lacks flexibility for updates. Second, their design prioritizes global goat diversity, with

*Correspondence:
Jun Luo
luojun@nwafu.edu.cn
[1]Shaanxi Key Laboratory of Molecular Biology for Agriculture, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi, China
[2]Institute of Animal Husbandry Quality Standards, Xinjiang Academy of Animal Sciences, Urumqi, Xinjiang, China

non-dairy breeds dominating the SNP selection, resulting in poor functional relevance to dairy traits and suboptimal performance in Chinese dairy goat populations [7, 8].

SNPs, as the most abundant genomic variations, are tightly linked to economically vital traits in dairy goats [9–11]. However, traditional genotyping methods struggle to balance throughput, resolution, and cost-effectiveness for large-scale precision breeding [12–14]. Although genomic selection has revolutionized cattle [15], pig [16], and poultry [17] breeding, dairy goats lag significantly in adopting such technologies. This gap underscores an urgent need for innovative solutions.

This study aims to design a high-resolution, multi-SNP liquid capture chip specifically for dairy goats using genomic and other data. The chip will focus on designing regions within the genome of dairy goats that are related to important economic traits. Optimizing the sequence and layout of capture probes will enable the capture of a large number of SNP sites with high sensitivity and specificity. The high-resolution feature of this chip will help us analyze the genetic diversity and population structure of dairy goats more finely, providing more accurate guidance for the genetic breeding of dairy goats and promoting the development of the dairy goat industry toward higher quality and efficiency.

Materials and methods

Sample collection and sequencing

In this study, samples were collected from four dairy goat breeds (aged 2–4 years) in Shaanxi Province, China: Xinong and Guanzhong dairy goat ($n = 142$, 127 females and 15 males; high milk yield and local adaptation), Saanen ($n = 184$, all females; high milk yield), and Alpine ($n = 22$, all females; high quality milk and environmental adaptability). For each goat, 5 mL of blood was drawn from the jugular vein without the need for euthanasia or anesthesia. Subsequently, DNA was extracted using the standard phenol-chloroform method. Samples with determined DNA concentrations were then subjected to whole-genome sequencing by Huada company. The average insert size per individual was 500 bp, and the average read length was 150 bp. Moreover, resequencing data from goat breeds were downloaded from public databases. The study was approved by the Institutional Animal Care and Use Committee of Northwest A&F University (protocol number DK2022008). All the applicable institutional and national guidelines for the care and welfare of animals have been strictly followed for the sampling procedures. We obtained informed consent from the owners of the goats at the farms to use the animals in this study.

Read mapping and variant calling

In this study, Trimmomatic v0.39 [18] was utilized to filter the paired-end sequences. Subsequently, BWA-MEM (v0.7.15-r1140) [19] was utilized to align the clean data with the goat reference genome ARS1.2 (GCF_001704415.2). Then, samtools [20] was used to construct a BAM file index for mapping purposes. The BAM files were sorted, and potential duplicate reads were removed by means of the Picard tool (http://broadinstitute.github.io/picard) tool. After the mapping process, SNP calling was carried out using the "Haplotype Caller," "Genotype GVCFs," and "Select Variants" modules within the GATK genomic analysis tool package (GATK, version 3.8-1-0-gf15c1c3ef) [21]. The initial SNPs were filtered using the "Variant Filtration" module, based on the following parameters: "QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < – 12.5, ReadPosRankSum < – 8.0, and SOR > 3.0", along with an average sequencing depth of variants in the range "<1/3× and > 3×" for all individuals. Finally, the ANNOVAR [22] software was used for the functional annotation of the SNPs.

Identification of specific loci in dairy goats

We classified goats into dairy goat (DG), non-dairy goat (NDG) breeds, and Xinong Dairy Goats, taking into account their characteristics and origins. Next, we compared the genomes of the two dairy goat breeds. To estimate the signal scanning regions, we combined nucleotide diversity (θπ DG/NDG) and fixation index (FST) with VCFtools [23], using 50 kb sliding windows and 25 kb sliding steps. In addition, we employed the cross-population composite likelihood ratio test (XP-CLR) [24] to identify potential regional differences between different breeds. XP-CLR is a likelihood method that detects selective sweeps by jointly modeling multi-locus allele frequency differences between two groups. We annotated the scanned regions from the intersection of the two parameters with the highest 1% threshold to identify candidate regions. Finally, we used Bedtools [25] and in-house written scripts to identify the candidate sites for each interval.

Genome-wide association analysis (GWAS) screening for functional loci

We sourced data from Saanen dairy goats, which had comprehensive genotype information as well as phenotype records. Our aim was to explore the genetic basis of important economic characteristics. To achieve this, we carried out an association analysis of these important economic traits. We utilized the linear mixed model within the Genome-wide Efficient Mixed Model Association (GEMMA) v0.98.4 [26] basing our analysis on the whole-genome-sequenced genotypes of dairy goats. Before proceeding with the analysis, we performed

quality control measures. Quality control was performed using PLINK v1.9 with the following criteria: (1) individual missingness (--mind 0.05), (2) SNP missingness (--geno 0.05), (3) minor allele frequency (--maf 0.05), and (4) Hardy-Weinberg equilibrium (--hwe 1e-6). Additionally, in the R environment, we used the qqman package to visualize Manhattan and Q - Q plots [27]. Finally, for each interval, we determined the threshold through in - house written scripts, and these thresholds were considered as the candidate sites.

Site selection in breed-specific, animal QTL database-related regions and functional sites

To distinguish between different breeds of dairy goats, the genomes of various dairy goat breeds were sequenced and contrasted. The focus was on analyzing the sequence disparities among the genomes of different varieties, with particular emphasis on single nucleotide polymorphisms (SNPs). Through comparing the distribution and frequency of these variations across different breeds, several species-specific markers were pinpointed. Genome-wide screening and functional annotation investigations led to the identification of genomic regions associated with crucial traits in different goat breeds, such as milk components and milk production. Additionally, some of the QTL regions previously recorded in the animal QTL database, which are linked to major economic traits (like growth and development, reproduction, and milk production) in goats, were selected as candidate intervals for chip design. The breed-specific regions and the regions from the goat QTL database were merged into larger regions using the "merge" sub-command of BEDtools. For each of these regions, tagSNPs were identified as the representative sites. Numerous prior studies have linked certain SNPs to important economic traits in dairy goats, including milk fat content, milk production traits, and immune traits. These sites were also incorporated into the chip candidate site dataset.

Probe production and sequencing

The liquid chip operates on the basis of GBTS technology. This technology depends on the target capture achieved through the complementary combination of probes and target sequences. To guarantee the capture efficiency, GenoBaits Probe Designer was utilized for designing two probes. These two probes should have an overlap ranging from 60 to 70% and both must cover the target SNP sites. The length of each designed probe is 110 bp. The GC content of the probes is maintained between 30% and 70%. Probes with non-specific amplification, simple sequence repeats, or gaps are excluded. Genomic DNA was extracted from whole blood following the standard phenol - chloroform method. The libraries were constructed using the GenoBaits DNA - seq Library Prep Kit (MolBreeding Biotechnology Co., Shijiazhuang, Hebei, China) in accordance with the manufacturer's protocol. Subsequently, the probes and hybridization buffer were mixed and allowed to hybridize at 65℃ for 16 h. After that, Dynabeads MyOne Streptavidin C1 and binding buffer were added to the solution. This step was carried out to enrich the target DNA fragments while removing the non - target ones. The target fragments were then amplified using the library amplification primer and DNA polymerase. Next, two rounds of purification were performed using Beckman AMPure Beads. Finally, the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, CA) and qPCR were employed to quantify the library concentration. Sequencing was carried out with PE150 on the MGISEQ-2000 platform (MGI, Shenzhen, China).

Verification of liquid chip

To get a more comprehensive overview of the panel and SNP performance, Samples were collected from the Shaanxi dairy goat breeding farm. A total of 212 samples were collected. DNA was extracted from blood and sequenced by the 25 K liquid chip. We combined the goat reference genome ARS1.2 and used the above methods for mapping, calling, and filtering SNPs. Afterward, evaluate data efficiency, base quality, alignment rate with the reference genome, SNP detection rate, genotype heterozygosity, breed identification, and whole genome association analysis.

Results

Data processing and GBTS liquid chip strategy for dairy goats

A total of 348 dairy goats were chosen for genome resequencing; details can be found in Supplementary Table S1. These datasets were analyzed in combination with previously published whole-genome sequencing (WGS) datasets from 126 individuals belonging to 11 breeds, as presented in Supplementary Table S2. Among the 11 domestic goat breeds, five are Chinese dairy goat breeds: Xinong Dairy Goat (XNG), Chinese Nubian Dairy Goat (CNG), Laoshan Dairy Goat (LSG), Tugenburg Dairy Goat (TGB), and Guanzhong Dairy Goat (GZG). The other six are non-dairy goat breeds, namely Yunnan Black Goat (BBG), Jintang Black Goat (JTY), Matou Black Goat (MTG), Shandong Jining Goat (SJG), Shaanbei White Cashmere Goat (SWG), and Wuzhumuqin Goat (WMG). Additionally, the dataset includes the Iranian wild goat (IWG) and dairy goats from Italy and Russia. From these 474 goats, a total of 11.85 TB of paired-end DNA sequence data was obtained. Each goat had an average genome sequencing depth of 15.2X, with a range from 4.17 to 29.9×, and an average genome coverage of 97.15% (see Supplementary Tables S1–S2). Of the sequenced reads, 99.63% were successfully mapped to the latest goat

reference genome, ARS1.2 (GCA_001704415.2) (Supplementary Tables S1, S2). Among these reads, 0.602% were located in the exon regions. Furthermore, 47,172,477 high-quality single nucleotide polymorphisms (SNPs) were detected, with 1.2% (902,990) of them found in exonic regions (Supplementary Tables S3). The average transition-to-transversion (Ti/Tv) ratio for all goat samples was 2.37. This relatively low ratio indicates a low likelihood of random sequencing errors (Supplementary Table S4). The design strategy and verification experiments of the 25 K GBTS liquid chip for dairy goats are depicted in Fig. 1.

Selection of signal loci and screening of candidate SNP loci for breed identification in dairy goats

To explore the selection signals related to milk production accumulated in the genome layer during the formation of dairy goat breeds, we conducted FST, θπ, and XP-CLR analyses on dairy goat and non-dairy goat breeds. The top 1% threshold outlier windows of the three methods were combined, which was considered to be the specific region of dairy goats under positive selection (Fig. 2) (Supplementary Table S5-7). After gene annotation of SNPs, it was found that they mainly focus on traits such as milk production, reproduction, growth and development, and immunity. This process resulted in the identification of 4117 candidate SNP sites associated with important traits. Next, To enhance the efficiency of distinguishing various breeds of dairy goats, this study employed machine learning algorithms to delve deeper into data features and assess the significance of SNP loci in breed differentiation. A total of 7,030 candidate SNP loci were identified.

Selection of functional SNP sites from GWAS, goat QTLdb, and literatures

In order to obtain loci related to important economic traits in dairy goats, this study collected traits such as milk production, milk composition (milk fat percentage, milk protein percentage, lactose percentage), somatic cell count, growth and development, breasts, and reproduction. Combined with the resequencing data of 298 dairy

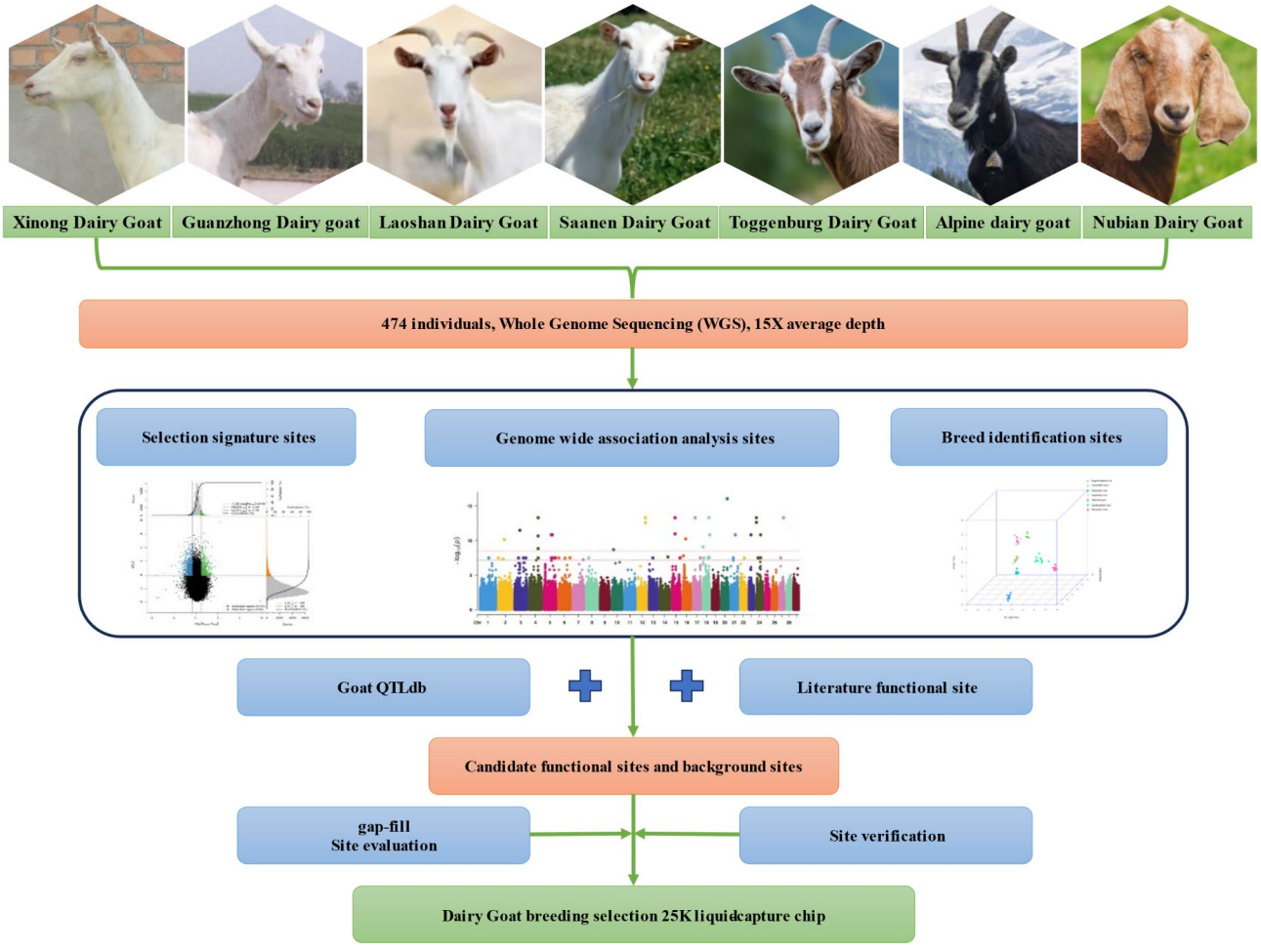

Fig. 1 Roadmap for the development and validation of the 25 K liquid phase chip for dairy goat

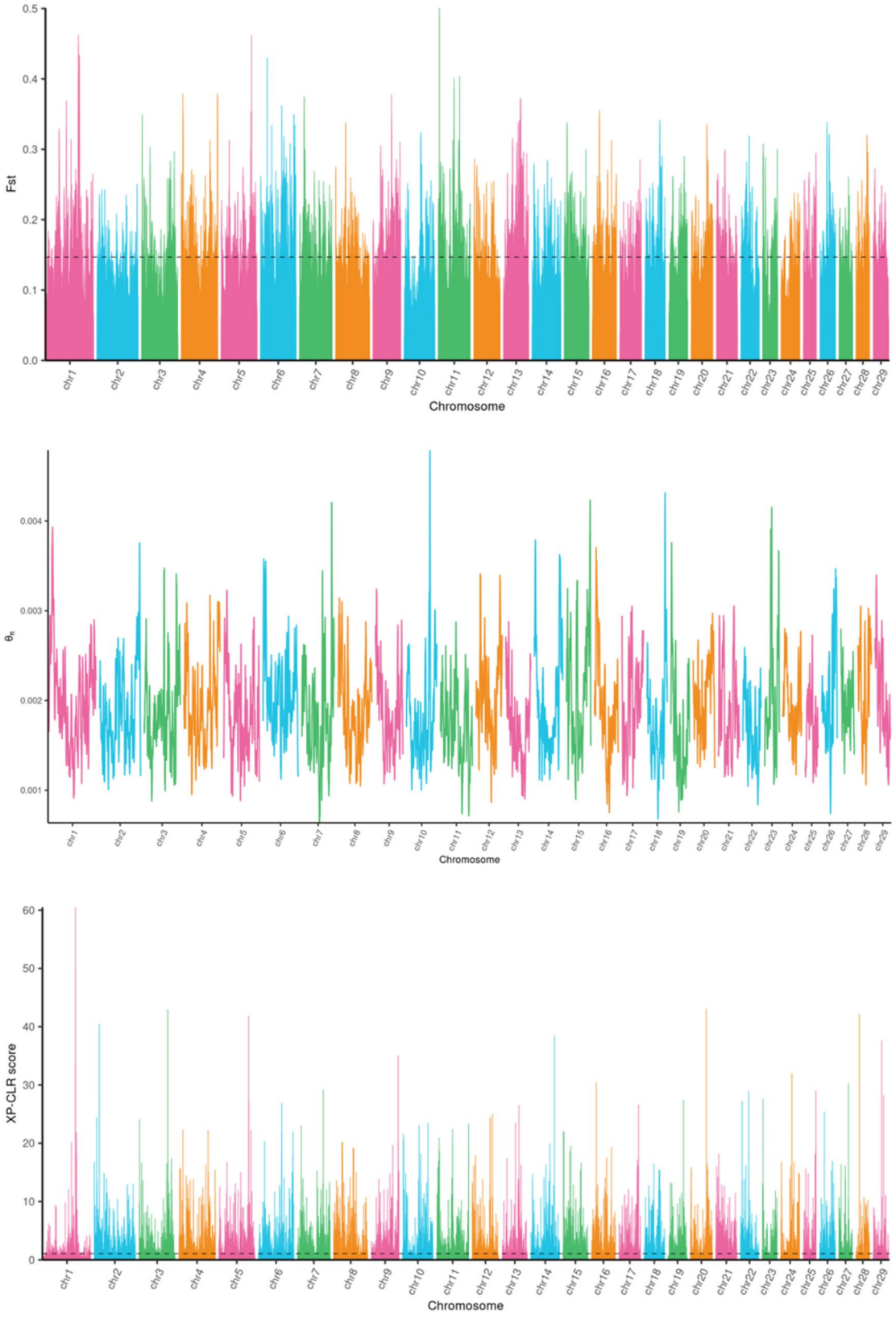

Fig. 2 Selective sweep analysis of dairy goat and non-dairy goat

goats, whole genome association analysis was conducted to obtain SNP loci related to important economic traits in dairy goats. A search was conducted on PubMed to review SNP literature related to important traits of dairy goats, including milk production traits, casein, milk fat content, immune traits (False tuberculosis, pyoderma, brucellosis), environmental adaptability, coat color, mastitis, keratinization, fatty acids, and semen quality. In addition, a comprehensive analysis of the Goat QTLdb database revealed the discovery of SNPs in genes related to economic traits such as milk production in goats. After removing duplicate and incomplete sites, which yielded a total of 6172 SNPs (Supplementary Table S8).

Analysis and verification of SNP loci types captured by 25 K GBTS liquid chip

In order to achieve uniform distribution of chip sites, background sites 14,887 were selected. A total of 32,293 loci were formed. These SNP loci were evaluated and 27,396 SNP loci were ultimately retained (Fig. 3A) for probe design. Due to the probe's ability to cover multiple sites, the number of SNPs in this chip can reach 46,459. From the figure, it can be seen that SNP loci are evenly distributed on chromosomes and there are no large GAP (Fig. 3B, C). The chip site annotation revealed that the majority of SNPs were situated between genes or within introns, next is the exon region (Fig. 3D). After

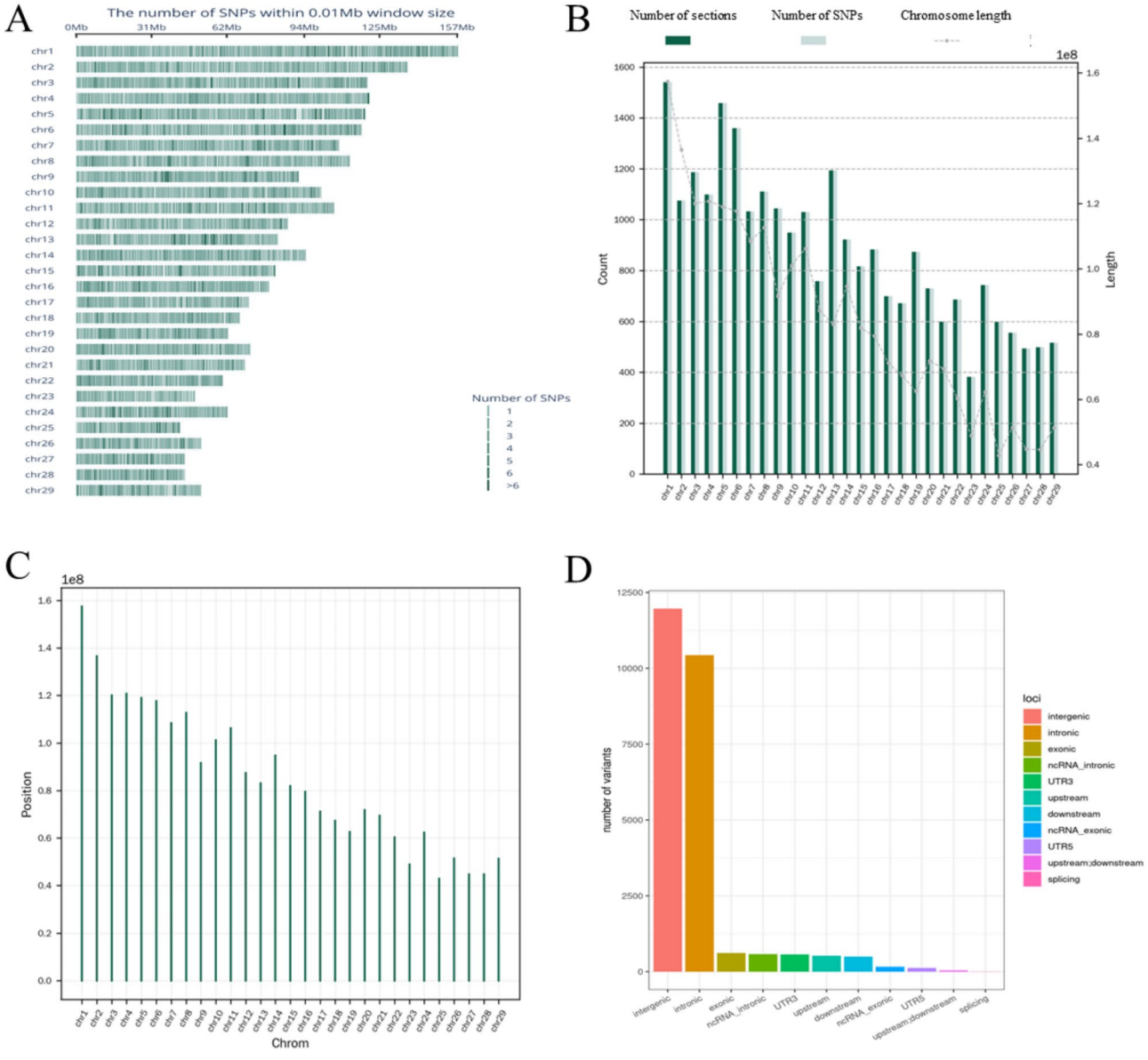

Fig. 3 The distribution of sites of 25 K GBTS liquid chip for dairy goats. (**A**) The uniform distribution of SNP sites on different chromosomes. (**B**) The distribution of SNP sites on different chromosomes. (**C**) Genomic distribution of target sites across 29 chromosomes in dairy goats. (**D**) The annotation results of core sites on the dairy goats 25 K GBTS liquid chip

read filtering and quality control, the effective rate of the data was 93.35–95.39%, with an average effective rate of 94.27%. The percentage of bases with Phred values greater than 20 and 30 in the total bases was 98.41% and 96.07%, respectively (Supplementary Table S9). The clean reads after quality control were compared with the reference genome, reaching an average comparison rate of 98.52% (Supplementary Table S10). These results showed that the sequencing quality was reliable and met the requirements for subsequent analysis. The 25 K liquid chip exhibited a high SNP detection rate for dairy goats, ranging from 99.66 to 99.75%, with an average of 99.71%, thereby meeting the required standards. In genotyping, the heterozygosity rate was 30.92–35.93% (Supplementary S11).

Breed identification and GWAS verification

To evaluate the performance of this chip, seven breeds of dairy goats were validated using chips, and the principal component analysis (PCA) obtained is shown in Fig. 4. The results clearly demonstrate that these SNP loci act as effective discriminators among individual dairy goat breeds. This highlights the potential of these SNP loci to serve as a characteristic database for dairy goats,

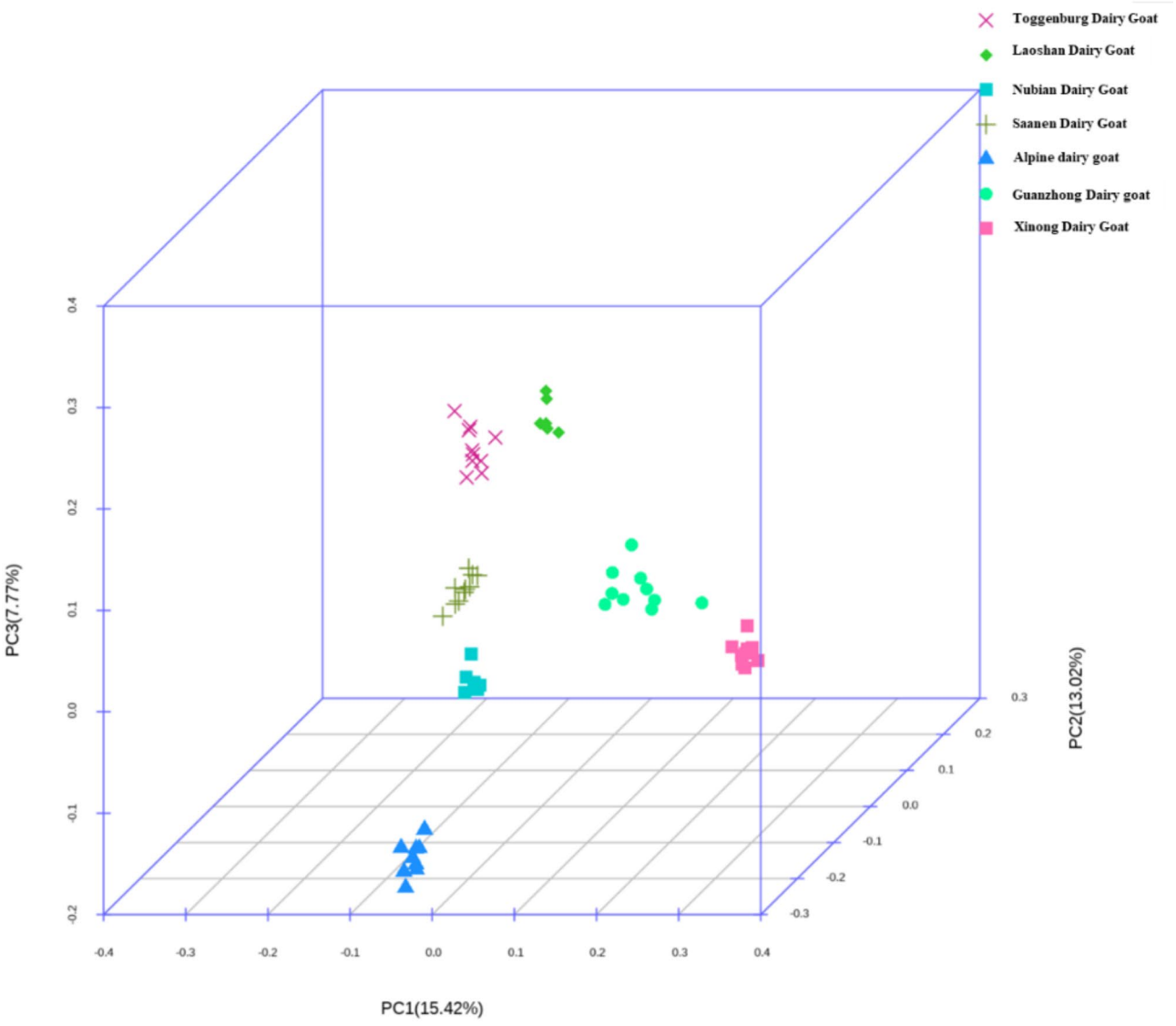

Fig. 4 Identification of dairy goat breeds using 3DPCA

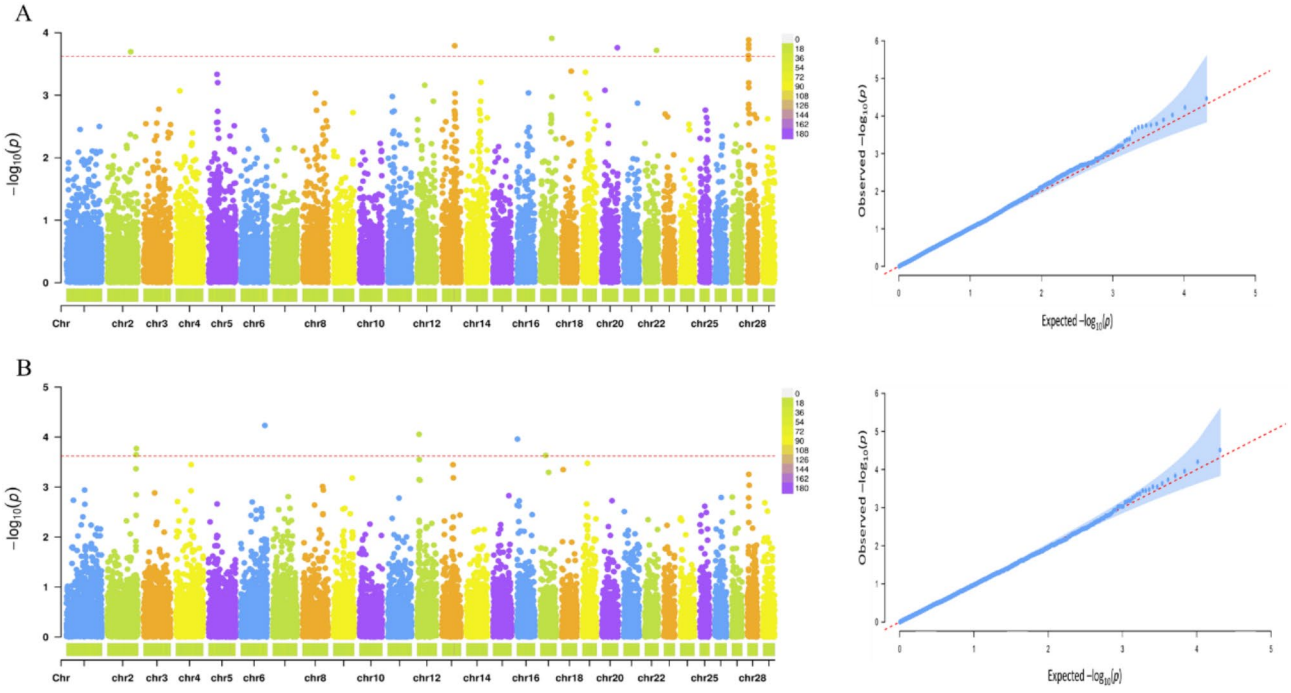

Fig. 5 Manhattan and QQ plot of GWAS for litter size and birth weight. (**A**) Manhattan and QQ plot for litter size. (**B**) Manhattan and QQ plot for birth weight

laying the groundwork for key genetic markers that could aid in the development of efficient breed identification methods in the future. Next, we conducted GWAS on litter size and birth weight traits using 200 dairy goats (Fig. 5). GWAS for litter size trait detected 11 associated SNPs with a genome-wide significance (Fig. 5A). These SNPs are located on chromosomes 17, 22, and 28, spanning a total of 1.03 Mb. There are a total of 10 genes in this region, including *PCDH18*, *SCAP*, *PTPN23*, *LOC102173049*, *KLHL18*, *KIF9*, *LOC102173511*, *LOC108633335*, *ANTXRL*, *GRID1*. GWAS for birth weight trait detected 7 associated SNPs with a genome-wide significance (Fig. 5B). These SNPs are located on chromosomes 2, 6, 12, 16, and 17, spanning a total of 1.44 Mb. There are a total of 7 genes in this region, including *NALCN*, *TRNAC-GCA*, *LRRN2*, *TMEM132D*, *COL5A2*, *HS3ST1*, and *TRNAC-GCA*.

Discussion

In the field of genetic breeding of dairy goats, accurate and efficient genotyping techniques are crucial for a deeper understanding of their genetic characteristics and promoting industrial development. In recent years, with the rapid application of genome resequencing technology and GBTS technology in the field of animal and plant genetics and breeding, new directions have been expanded for low-cost and efficient genotyping technology [12, 28, 29]. This study selected 348 dairy goats for genome resequencing and integrated published whole genome sequencing data of 126 individuals from 11 breeds, obtaining large-scale sequencing data of 474 goats. This multi breed and large-scale data source provides a solid foundation for comprehensively revealing the genetic information of dairy goats. The average genome depth of 15.20 times and coverage rate of 97.15% ensure the accuracy and completeness of the data, creating conditions for precise mutation detection and functional site screening in the future. Based on the sequencing data above, this study focuses on designing a 25 K multiplex SNP liquid phase capture chip suitable for dairy goats, dedicated to the high-quality and sustainable development of China's dairy goat industry.

To meet the diversity and specificity of chip sites, this study screened chip sites through multiple pathways. Firstly, site mining is based on selection signals. Through FST, θπ, and XP-CLR analysis of dairy and non-dairy goat breeds, 4117 candidate SNP loci related to important economic traits in dairy goats were successfully identified. These loci are concentrated in key trait-related regions such as milk production, reproduction, growth and development, and immunity, indicating that these genomic regions are subject to selection pressure during the formation of dairy goat breeds and may play an important role in the breed-specific characteristics of dairy goats [30, 31]. Secondly, machine learning screening of a breed of identification sites. Using machine learning algorithms to screen 7030 candidate SNP loci for breed identification. The PCA analysis results confirmed

that these loci can effectively distinguish different dairy goat breeds, providing strong support for establishing a molecular marker database for dairy goat breed identification, and helping to provide a more accurate basis for breed protection, breeding, and management [32]. Finally, integrate multiple sources of data to obtain relevant loci. By combining whole genome association analysis, literature research, and information from the Goat QTLdb database, 6172 SNP loci related to important economic traits in dairy goats were identified. This multi-source data integration method can comprehensively cover known and potential functional loci, improving the chip's ability to detect genetic variations related to economic traits in dairy goats [14, 33].

In the chip design process, 14,887 background sites were selected to achieve a uniform distribution of sites, resulting in 32,293 sites. After evaluation, 27,396 SNP sites were retained for probe design, and 46,459 SNPs were detected. The uniform distribution of loci on chromosomes ensures effective coverage of the entire genome, avoiding excessive or insufficient representation of local regions, thereby improving the accuracy and comprehensiveness of chip detection. Meanwhile, the annotation results of the loci showed that most SNPs were located in intergenic or intronic regions, which is consistent with the distribution characteristics of functional elements in the genome and helps to discover potential regulatory elements and intergenic interaction sites [34].

In order to verify the sequencing quality and detection efficiency of the chip, we strictly filtered and quality-controlled the sequencing data, and the effective rate reached 93.35–95.39%, with an average effective rate of 94.27%. The base quality index was good, and the average alignment rate with the reference genome was as high as 98.52%, indicating that the sequencing quality is reliable and meets the requirements of subsequent analysis. The detection rate of SNPs in dairy goats by the chip is extremely high, with an average of 99.71%. This result fully demonstrates the rationality of chip design and the high sensitivity of detection, which can accurately capture target SNP sites and provide guarantees for precise genotyping. In genotyping, the heterozygosity rate is 30.92–35.93%, which reflects the genetic diversity level of the dairy goat population. Appropriate heterozygosity helps maintain the adaptability and evolutionary potential of the population, while also providing abundant genetic variation resources for genetic breeding [35]. Through genome-wide association analysis of traits such as litter size and birth weight, this study has preliminarily identified SNP loci and genes associated with these traits. Previous literature has shown that *SCAP* [36], *PTPN23* [37], *KIF9* [38, 39], *ANTXRL* [40], and *GRID1* [41–43] are associated with the reproductive capacity of poultry and sheep [44]. Moreover, *NALCN* [45], *LRRN2* [46], *TMEM132D* [47], *COL5A2* [48], and *HS3ST1* [49] are associated with growth and development [50].

With the development of more trait-related research, this chip will help construct a genetic map of important economic traits in dairy goats, explore key genes and regulatory elements, and achieve a transition from traditional breeding to precision breeding. Recent advances in genomic technologies, such as machine learning-driven SNP prioritization and multi-omics integration, highlight the potential for further refining trait-associated loci in dairy goats. Precision breeding can improve breeding efficiency, accelerate the cultivation process of excellent breeds of dairy goats, meet the demand of the dairy goat industry for high-yield, high-quality, and stress-resistant breeds, and promote the development of the dairy goat industry towards higher quality and efficiency. Although significant achievements have been made in chip design and verification in this study, there are still certain limitations. For example, although the selection of chip loci is based on multiple data sources, it may not cover all functional loci related to important economic traits in dairy goats. In addition, this study mainly focuses on specific breeds and populations of dairy goats, and its applicability to other regions or breeds needs further verification. However, based on the flexibility of liquid-phase chips, further optimization of loci can be carried out according to the expansion of the sequencing population size and the mining of important traits. Future research should focus on integrating real-time phenotyping systems and artificial intelligence-driven genomic prediction models, which are emerging as transformative tools in livestock genomics.

Conclusions

The 25 K multiplex SNP liquid-phase capture chip designed in this study demonstrates high resolution, sensitivity, and specificity, enabling efficient genome-wide SNP detection in dairy goats. Compared to traditional solid-state chips, this platform offers lower costs, dynamic updatability of functional loci, and enhanced trait-specific targeting. Its modular design allows scalable integration of new SNPs, ensuring adaptability to evolving breeding objectives. Practical applications include genome selection of important economic traits, directly promoting precise breeding and genetic improvement of dairy goat populations. This tool provides critical technical support for driving high-quality development in the dairy goat industry.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11576-z.

Supplementary Material 1

Acknowledgements
This work received support from the high-performance computing (HPC) resources at Northwest A&F University (NWAFU).

Author contributions
JL and JQZ designed the experiment and wrote and typeset the manuscript. YLW and FHZ carried out the experiments and analyzed the data. JK and WW contributed to data collection. WW revised the manuscript. PG, XFL and HPS managed the experiments, performed and supervised the research. All authors read and approved the final manuscript.

Funding
This research was jointly supported by the following projects: the Shaanxi Livestock and Poultry Breeding Double-chain Fusion Key Project of China (2022GD-TSLD-46-0201), the Hohhot City Science and Technology Plan Project (Major Science and Technology Special Project) in China (2023150103000025), the National Key Research and Development Program of China (2021YFD1600704), and the Tianchi Talent Special Appointed Expert Project.

Data availability
The data that support the findings of this study are openly available in NCBI Sequence Read Archive at https://www.ncbi.nlm.nih.gov/bioproject/1077730, Accession number: PRJNA1077730. Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA017705) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa. Other data that supported the discovers could been obtained from the corresponding authors.

Declarations

Ethics approval and consent to participate
The study was approved by the Institutional Animal Care and Use Committee of Northwest A&F University (protocol number DK2022008). All the applicable institutional and national guidelines for the care and welfare of animals have been strictly followed for the sampling procedures. We obtained informed consent from the owners of the goats at the farms to use the animals in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 23 December 2024 / Accepted: 7 April 2025
Published online: 15 April 2025

References

1. Hm H, Aa S, Nnam H, Ama R, Ma S, Az M et al. The evolution and role of molecular tools in measuring diversity and genomic selection in livestock populations (Traditional and up-to-date insights): a comprehensive exploration. Veterinary Sci. 2024;11.
2. Manimekalai R, Suresh G, Govinda Kurup H, Athiappan S, Kandalam M. Role of NGS and SNP genotyping methods in sugarcane improvement programs. Crit Rev Biotechnol. 2020;40:865–80.
3. Li Y-H, Wang H-P. Advances of genotyping-by-sequencing in fisheries and aquaculture. Rev Fish Biol Fisheries. 2017;27:535–59.
4. Ott A, Liu S, Schnable JC, Yeh C-T, Eddy, Wang K-S, Schnable PS. tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Res. 2017;45:e178.
5. Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, et al. Design and characterization of a 52K SNP chip for goats. PLoS ONE. 2014;9:e86227.
6. Li R, Yang P, Dai X, Asadollahpour Nanaei H, Fang W, Yang Z, et al. A near complete genome for goat genetic and genomic research. Genet Sel Evol. 2021;53:74.
7. Vijh RK, Sharma U, Kapoor P, Raheja M, Arora R, Ahlawat S, et al. Design and validation of high-density SNP array of goats and population stratification of Indian goat breeds. Gene. 2023;885:147691.
8. Waineina RW, Ngeno K, Okeno TO, Ilatsia ED. Genetic diversity and population structure among indigenous and imported goat breeds in Kenya. Genetic Resour. 2021;2:25–35.
9. Scholtens M, Jiang A, Smith A, Littlejohn M, Lehnert K, Snell R, et al. Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. J Anim Sci Biotechnol. 2020;11:55.
10. Brito LF, Kijas JW, Ventura RV, Sargolzaei M, Porto-Neto LR, Cánovas A, et al. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC Genomics. 2017;18:1–20.
11. Lai F-N, Zhai H-L, Cheng M, Ma J-Y, Cheng S-F, Ge W, et al. Whole-genome scanning for the litter size trait associated genes and SNPs under selection in dairy goat (Capra hircus). Sci Rep. 2016;6:38096.
12. Guo Y, Bai F, Wang J, Fu S, Zhang Y, Liu X, et al. Design and characterization of a high-resolution multiple-SNP capture array by target sequencing for sheep. J Anim Sci. 2023;101:skac383.
13. Meng Y, Zhang W, Cheng Y, Wu Y, Wu H, He M, et al. Development and verification of a 10K liquid chip for Hainan black goat based on genotyping by pinpoint sequencing of liquid captured targets. BMC Genom Data. 2024;25:44.
14. Guan S, Li W, Jin H, Zhang L, Liu G. Development and validation of a 54K genome-wide liquid SNP chip panel by target sequencing for dairy goat. Genes (Basel). 2023;14:1122.
15. Weller JI, Ezra E, Ron M. *Invited review*: a perspective on the future of genomic selection in dairy cattle. J Dairy Sci. 2017;100:8633–44.
16. Knol EF, Nielsen B, Knap PW. Genomic selection in commercial pig breeding. Anim Front. 2016;6:15–22.
17. Tan X, Liu R, Li W, Zheng M, Zhu D, Liu D, et al. Assessment the effect of genomic selection and detection of selective signature in broilers. Poult Sci. 2022;101:101856.
18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30:2114–20.
19. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25:1754–60.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009;25:2078–9.
21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
23. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and vcftools. Bioinformatics. 2011;27:2156–8.
24. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. Genome Res. 2010;20:393–402.
25. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
26. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4.
27. Turner SD. Qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. J Open Source Softw. 2018;3:731.
28. Liu S, Xiang M, Wang X, Li J, Cheng X, Li H et al. Development and application of the genobaits WheatSNP16K array to accelerate wheat genetic research and breeding. Plant Comm. 2024;0.
29. Huang Y, Li Z, Li M, Zhang X, Shi Q, Xu Z. Fish genomics and its application in disease-resistance breeding. Reviews Aquaculture. 2025;17:e12973.
30. Zhang B, Chang L, Lan X, Asif N, Guan F, Fu D, et al. Genome-wide definition of selective sweeps reveals molecular evidence of trait-driven domestication among elite goat (Capra species) breeds for the production of dairy, cashmere, and meat. GigaScience. 2018;7:giy105.
31. Qanbari S, Rubin C-J, Maqbool K, Weigend S, Weigend A, Geibel J, et al. Genetics of adaptation in modern chicken. PLoS Genet. 2019;15:e1007989.
32. Zhao C, Wang D, Teng J, Yang C, Zhang X, Wei X, et al. Breed identification using breed-informative SNPs and machine learning based on whole genome sequence data and SNP chip data. J Anim Sci Biotechnol. 2023;14:85.
33. Zhao J, Mu Y, Gong P, Liu B, Zhang F, Zhu L et al. Whole-genome resequencing of native and imported dairy goat identifies genes associated with productivity and immunity. Front Vet Sci. 2024;11.

34. Groenen MAM, Megens H-J, Zare Y, Warren WC, Hillier LW, Crooijmans RPMA, et al. The development and characterization of a 60K SNP chip for chicken. BMC Genomics. 2011;12:274.
35. Talebi R, Szmatoła T, Mészáros G, Qanbari S. Runs of homozygosity in modern chicken revealed by sequence data. G3 Genes|Genomes|Genetics. 2020;10:4615–23.
36. Wang Y, Guo Y, Duan C, Li J, Ji S, Yan H, et al. LncGSAR controls ovarian granulosa cell steroidogenesis via sponging MiR-125b to activate SCAP/SREBP pathway. Int J Mol Sci. 2022;23:12132.
37. Tanaka K, Kondo K, Kitajima K, Muraoka M, Nozawa A, Hara T. Tumor-suppressive function of protein-tyrosine phosphatase non-receptor type 23 in testicular germ cell tumors is lost upon overexpression of miR142-3p MicroRNA. J Biol Chem. 2013;288:23990–9.
38. Miyata H, Shimada K, Morohoshi A, Oura S, Matsumura T, Xu Z, et al. Testis-enriched Kinesin KIF9 is important for progressive motility in mouse spermatozoa. FASEB J. 2020;34:5389–400.
39. Meng Z, Meng Q, Gao T, Zhou H, Xue J, Li H, et al. Identification of bi-allelic KIF9 loss-of-function variants contributing to asthenospermia and male infertility in two Chinese families. Front Endocrinol (Lausanne). 2022;13:1091107.
40. Zhang J, Duan Z, Wang X, Li F, Chen J, Lai X, et al. Screening and validation of candidate genes involved in the regulation of egg yolk deposition in chicken. Poult Sci. 2021;100:101077.
41. Sánchez-Ramos R, Trujano-Chavez MZ, Gallegos-Sánchez J, Becerril-Pérez CM, Cadena-Villegas S, Cortez-Romero C. Detection of candidate genes associated with fecundity through genome-wide selection signatures of Katahdin Ewes. Anim (Basel). 2023;13:272.
42. Asadollahpour Nanaei H, Ayatollahi Mehrgardi A, Esmailizadeh A. Whole-genome sequence analysis reveals candidate genomic footprints and genes associated with reproductive traits in thoroughbred horse. Reprod Domest Anim. 2020;55:200–8.
43. Mekonnen KT, Lee D-H, Cho Y-G, Son A-Y, Seo K-S. Genome-wide association studies and runs of homozygosity reveals genetic markers associated with reproductive performance in Korean Duroc, landrace, and Yorkshire breeds. Genes (Basel). 2024;15:1422.
44. Abdoli R, Zamani P, Mirhoseini S, Ghavi Hossein-Zadeh N, Nadri S. A review on prolificacy genes in sheep. Reprod Domest Anim. 2016;51:631–7.
45. Belal M, Mucha M, Monteil A, Winyard PG, Pawlak R, Walker JJ, et al. The background sodium leak channel NALCN is a major controlling factor in pituitary cell excitability. J Physiol. 2024. https://doi.org/10.1113/JP284036
46. Andreae LC, Lumsden A, Gilthorpe JD. Chick Lrrn2, a novel downstream effector of Hoxb1 and Shh, functions in the selective targeting of rhombomere 4 motor neurons. Neural Dev. 2009;4:27.
47. Tarsani E, Kranis A, Maniatis G, Avendano S, Hager-Theodorides AL, Kominakis A. Discovery and characterization of functional modules associated with body weight in broilers. Sci Rep. 2019;9:9125.
48. Mohammadi H, Khaltabadi Farahani AH, Moradi MH, Moradi-Shahrbabak H, Gholizadeh M, Najafi A, et al. Genome-wide scan for selective sweeps reveals novel loci associated with prolificacy in Iranian sheep breeds in comparison with highly prolific exotic breed. Anim (Basel). 2024;14:3245.
49. Chen J, Sun T, You Y, Lin B, Wu B, Wu J. Genome-wide identification of potential odontogenic genes involved in the dental epithelium-mesenchymal interaction during early odontogenesis. BMC Genomics. 2023;24:163.
50. Talebi R, Ghaffari MR, Zeinalabedini M, Abdoli R, Mardi M. Genetic basis of muscle-related traits in sheep: a review. Anim Genet. 2022;53:723–39.

Publisher's note