Identification of SNP Markers Associated with Adult Body Weight in Cattle via Whole-Genome Resequencing

Pengpeng Zhang, Wenyan Wu*

School of Life Sciences, Xinyang Normal University, Xinyang 464000, Henan, China
* Corresponding author: Wenyan Wu

Abstract: To address the problems of low adult body weight and insufficient meat production in native beef cattle breeds in China, this study used Nanyang cattle and Pinan cattle as experimental populations, adopted whole-genome resequencing technology combined with genome-wide association study (GWAS) to identify molecular markers associated with adult body weight traits. Whole-genome resequencing and SNP genotyping were performed, and GWAS analysis was conducted using a mixed linear model. A SNP marker (Chr13: 64039609, with C/A polymorphism) significantly associated with adult body weight in cattle was screened out, with an association significance P-value of 4.67E-08. Individuals carrying the C allele had significantly higher adult body weight than those carrying the A allele (effect value: 8.67±1.29). Specific primers were designed based on this SNP marker, and a PCR sequencing detection method was established to achieve accurate genotyping of the target locus. The SNP marker identified in this study provides an efficient molecular tool for the early selection of adult body weight traits in cattle and helps accelerate the genetic improvement process of native beef cattle breeds.

Keywords: cattle; adult body weight; SNP; genome-wide association study

1. Introduction

With the continuous improvement of China's economic level and the upgrading of residents' dietary structure, beef, as an animal food with high protein, low fat and high nutritional value, has shown a steady growth trend in consumer demand. In this context, accelerating the breeding of high-quality beef cattle breeds and improving the production performance of native cattle breeds have become important tasks to ensure the safety of China's beef industry and promote the high-quality development of animal husbandry.

Nanyang cattle is one of the important local yellow cattle breeds in China, with advantages such as strong adaptability, roughage tolerance and good stress resistance [1]. Historically, it has both draft and meat value and has played an important role in agricultural production. However, due to the long-term breeding direction focusing on draft performance and insufficient systematic improvement of meat traits, Nanyang cattle generally have problems such as low adult body weight, slow growth rate and poor meat production performance, which are difficult to fully meet the needs of modern beef cattle industry for high yield, high efficiency and large-scale development [2]. In recent years, leveraging Nanyang cattle—an outstanding local germplasm resource renowned for strong environmental adaptability and roughage tolerance in central China—Pinan cattle, a novel specialized meat-type cattle population, has been successfully bred through systematic crossbreeding and selective breeding programs, with the introduction of frozen semen from imported Italian Piedmontese cattle as the male parent and Nanyang cattle as the female parent [3]. The breeding process initially adopted graded crossing (F1 hybrid females were backcrossed with Piedmontese sires) followed by inter se mating of advanced-generation hybrids, which enabled the new population to integrate the superior production traits of Piedmontese cattle (e.g., rapid growth rate, high carcass yield, and well-developed double-muscle characteristics)[3] while retaining the core adaptive advantages of Nanyang cattle, such as robust stress resistance and low feeding costs. Targeted genetic improvement through multiple generations of selection has allowed Pinan cattle to achieve significant gains in key meat production traits. Pinan cattle have an average dressing percentage of 65.6% and an average net meat percentage of 55.4%, which are 9.5% and 9% higher than those of Nanyang cattle, respectively. Along with lower fat and cholesterol content that aligns with modern healthy dietary demands [4].

As a landmark regional breeding achievement in central China’s beef industry, Pinan cattle has demonstrated remarkable growth potential and broad industrial application prospects in practical production, emerging as a core breed for local beef supply chain upgrading and farmer income enhancement. However, despite its phenotypic superiority, the population’s genetic architecture remains poorly characterized. Specifically, the genetic basis underlying key economic traits (e.g., adult body weight, growth efficiency, and meat quality) has not been systematically dissected, and the functional molecular markers or candidate genes regulating body weight—one of the most critical traits for beef cattle breeding—are largely unknown. This knowledge gap not

only hinders the precision of continuous genetic improvement for Pinan cattle (e.g., accelerating the fixation of favorable weight-related alleles) but also restricts the large-scale popularization and standardized breeding of this elite population, as conventional phenotypic selection alone cannot meet the demands of modern molecular breeding for accuracy and efficiency. Therefore, it is imperative to decipher the genetic mechanisms of body weight traits in Pinan cattle through high-throughput genomic technologies, which will lay a theoretical foundation for establishing genomic selection systems and advancing the sustainable development of the breed.

Traditional beef cattle breeding mainly relies on phenotypic selection, which not only has a long breeding cycle and low improvement efficiency, but also is easily interfered by environmental factors, making it difficult to accurately reflect the real genetic potential of individuals. Genomic selection technology can realize early and accurate selection of excellent genotypes by directly using genetic markers closely related to target traits, thus significantly improving breeding efficiency and shortening breeding cycle [1,5,6]. It is a key technical means to promote the genetic improvement of local cattle breeds and newly bred varieties. Genome-wide association study (GWAS), as a core method for analyzing the genetic basis of complex economic traits, can locate genetic loci significantly associated with target traits on a genome-wide scale[7]; whole-genome resequencing technology can break through the limitations of existing foreign commercial SNP chips in insufficient locus coverage and large information bias in local breeds and newly bred populations, and efficiently and systematically explore potential functional SNP markers.

Based on the above background, this study took Pinan cattle and Nanyang cattle as research objects, constructed a basic reference population, and adopted a research strategy of whole-genome resequencing combined with GWAS to systematically screen SNP markers significantly associated with adult body weight, and established an efficient and stable molecular detection method on this basis. The research results will provide theoretical support and technical tools for molecular marker-assisted selection and continuous genetic improvement of body weight traits in Pinan cattle, and also provide a reference technical route and scientific basis for the meat-oriented improvement of Nanyang cattle and other local yellow cattle breeds.

2. Materials and Methods

2.1 Experimental Population and Sample Collection

The experimental populations were from Nanyang Cattle Original Breeding Farm and Nanyang Bull Station in Nanyang City, Henan Province. 10 mL of jugular vein anticoagulated blood was collected from each individual, fully mixed and stored at 4℃ for genomic DNA extraction. At the same time, the adult body weight of all individuals was measured as the basic data for phenotypic association analysis.

2.2 Genomic DNA Extraction

Genomic DNA was extracted from blood samples by phenol-chloroform extraction method, and the specific steps were as follows: 1 mL of blood sample was mixed with an equal volume of PBS buffer, centrifuged at 3500 g for 10 min at room temperature, and the supernatant was discarded; washing was repeated until the supernatant was clear and the precipitate was colorless; 1 mL of DNA extraction solution (10 mmol/L Tris-HCl, pH=8.0; 0.1 mol/L EDTA; 0.5% SDS) was added to the precipitate, incubated in a water bath at 37℃ for 1 h, then proteinase K was added to a final concentration of 60 μg/mL, and incubated overnight at 55℃ until the solution was clear; extraction was performed with an equal volume of Tris-saturated phenol, 0.5 volume of phenol-0.5 volume of chloroform, and an equal volume of chloroform in turn, centrifuged at 12000 g for 10 min at 4℃, and the upper aqueous phase was taken; 2 volumes of pre-cooled absolute ethanol (-20℃) was added to precipitate DNA, placed at -20℃ for 30 min, then centrifuged at 12000 g for 10 min at 4℃, and the ethanol was discarded; the DNA precipitate was rinsed twice with 70% ethanol, dried in vacuum, dissolved with 100~300 μL of ultrapure water, the concentration and purity were determined by a spectrophotometer, and stored at -80℃ for later use.

2.3 Whole-Genome Resequencing and SNP Genotyping

Qualified DNA samples were subjected to paired-end sequencing using DNBSEQ-T7 sequencer, with an average sequencing data volume of 30G per sample. Fastp (V0.23.2) software was used to quality-control the raw sequencing data and remove low-quality reads; BWA (V0.7.17) software was used to align the quality-controlled reads to the cattle reference genome (ARS-UCD1.2) and screen unique aligned sequences; GATK (V4.3.0.0) software was used for SNP genotyping, filtered by conditions such as QD<2.0||FS>60.0||MQ<40.0, and parameters such as maf≥0.05, max-missing≥0.7, minDP ≥4, maxDP≤1000 were set at the same time to obtain high-quality SNP genotyping data sets.

2.4 Genome-Wide Association Analysis

GWAS analysis was conducted using a mixed linear model (LMM) implemented in the GEMMA software package under a Linux environment. The specific model is as follows:

$$y = W\alpha + x\beta + u + \varepsilon \quad (1)$$

where y is an n-dimensional vector representing the phenotypic values of n individuals; W is an n × c matrix of fixed effects, including a column of ones; α is the corresponding c-dimensional vector of coefficients, including the intercept; x is an n-dimensional vector of SNP genotypes; β denotes the SNP effect size; u is an n-dimensional vector of random effects; and ε is an n-dimensional vector of residual errors.

The random effects follow a multivariate normal distribution:

$$u \sim MVN_n(0, \lambda\tau^{-1}K) \quad (2)$$

and the residual errors follow:

$$\varepsilon \sim MVN_n(0, \lambda\tau^{-1}I_n) \quad (3)$$

where MVN denotes the multivariate normal distribution, $\lambda\tau^{-1}$ represents the residual error variance, λ is the ratio between the two variance components, K is the kinship matrix calculated from SNP data, and I_n is the n × n identity matrix.

2.5 Detection of SNP Markers

According to the sequences upstream and downstream of the significant SNP locus screened by GWAS, specific PCR primers were designed: upstream primer 5'-GCTGCTCCCTCACTCCTACCC-3', downstream primer 5'-GCACTGCCCGTCTGTCTCATT-3'. PCR amplification system (20 μL): 10 μL PCR Mix, 0.5 μmol/L upstream primer, 0.5 μmol/L downstream primer, 50 ng template DNA, supplemented with deionized water to the volume. Amplification program: pre-denaturation at 94 ℃ for 4 min; denaturation at 94 ℃ for 20 s, annealing at 58 ℃ for 20 s, extension at 72 ℃ for 20 s, 35 cycles. After agarose gel electrophoresis detection, the amplification products were purified by Takara DNA purification kit and subjected to Sanger sequencing to determine the base polymorphism of SNP loci.

3. Results

3.1 DNA Extraction and Sequencing Quality Control Results

The quality of genomic DNA extracted from cattle blood samples was first evaluated by agarose gel electrophoresis (Figure 1). Electrophoresis was performed using a 1.0% agarose gel (stained with ethidium bromide), and the rightmost lane corresponds to a DNA molecular weight marker (DL10000). Lanes 1–10 (representing individual cattle samples) display bright, single high-molecular-weight bands, with no obvious smearing or degraded fragments observed (Figure 1). This band pattern indicates that the extracted genomic DNA has excellent integrity, with no significant degradation during the extraction process. The OD260/OD280 ratio of genomic DNA of cattle was between 1.8 and 2.0, and the purity met the sequencing requirements. After quality control of the raw data of whole-genome resequencing, the proportion of effective reads was higher than 95%, and the alignment rate of reads with the reference genome was ⩾98%. After strict filtration of SNP genotyping data, high-quality SNP loci were finally obtained for subsequent GWAS analysis.

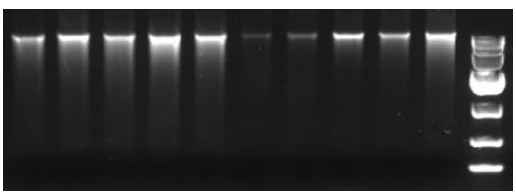

Figure 1. Agarose gel electrophoresis profile of Nanyang cattle genomic DNA

3.2 Screening of SNP Markers Associated with Adult Body Weight by GWAS

The results of GWAS analysis based on LMM model showed that a SNP marker significantly associated with adult body weight of Nanyang cattle was detected at locus Chr13: 64039609 on cattle chromosome 13, with an association significance P-value of 4.67E-08, reaching the genome-wide significant level. This locus had C/A base polymorphism, and the effect value of C allele was 8.67±1.29 (Table 1), indicating that Nanyang cattle carrying C allele had significantly higher adult body weight than those carrying A allele.

Table 1. SNP marker significantly associated with adult body weight in cattle.

Chromosome	Position	Ref SNP	Alt SNP	P-value	Effect value
13	64039609	C	A	4.67E-08	8.67±1.29

3.3 PCR Detection and Sequencing Verification of SNP Markers

PCR amplification was performed on Nanyang cattle DNA samples using SEQ ID NO.2~3 primers. The results of agarose gel electrophoresis (Figure 2A) showed that the amplification products had a single band, the length was consistent with the expected (197bp), and there were no miscellaneous bands and non-specific amplification, indicating that the primers had good specificity. The purified PCR products were subjected to Sanger sequencing, and the sequencing peak map could clearly identify the base type of the target locus. Among them, the heterozygous samples showed C/A double peaks at the target locus (Figure 2B), and the homozygous samples showed single C peaks or A peaks, which verified the accuracy and feasibility of the detection method.

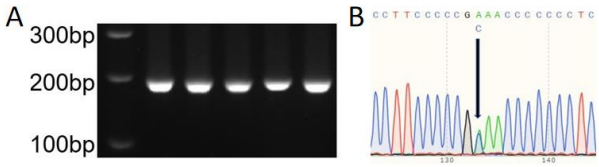

Figure 2. PCR amplification and sequencing verification of SNP markers

4. Discussion

Improving adult body weight is a central objective in beef cattle breeding because it directly affects growth efficiency, carcass yield, and overall production profitability. However, genetic improvement of indigenous Chinese cattle breeds has historically been constrained by reliance on phenotypic selection and the lack of effective molecular markers. In this study, we employed whole-genome resequencing combined with GWAS to identify a SNP marker significantly associated with adult body weight in Nanyang and Pinan cattle, providing both theoretical insight and practical tools for molecular breeding.

4.1 Advantages of Whole-Genome Resequencing–Based GWAS in Local Cattle Breeds

Most commercially available SNP chips were developed using European taurine cattle populations [8,9], leading to suboptimal marker coverage and allele frequency bias when applied to Chinese indigenous breeds or newly formed composite populations. This limitation reduces the power to detect breed-specific functional variants. By using whole-genome resequencing, this study effectively overcame these constraints and enabled comprehensive detection of genetic variation across the genome. This strategy allowed for the identification of a genome-wide significant SNP locus that would likely be missed by chip-based approaches, highlighting the methodological advantage of resequencing-based GWAS for local and novel cattle populations.

4.2 Biological and Breeding Significance of the Chr13:64039609 SNP

The significant SNP identified on chromosome 13 (Chr13:64039609) showed a strong association with adult body weight, with cattle carrying the C allele exhibiting significantly higher body weight than those carrying the A allele. The estimated allelic effect (8.67 ± 1.29) indicates that this locus has a moderate but meaningful contribution to phenotypic variation, which is particularly valuable for complex traits such as body weight that are controlled by multiple genes.

Notably, this SNP is located upstream of the TP53 gene. Although TP53 is best known for its role in cell cycle regulation, apoptosis, and genomic stability [10,11], increasing evidence suggests that it can influence growth and metabolism indirectly through regulation of cell proliferation and energy balance. While direct functional evidence linking this locus to body weight regulation in cattle remains to be established, its genomic location suggests potential regulatory effects that merit further investigation. This finding provides an important starting point for fine mapping and functional validation studies, such as expression quantitative trait locus (eQTL) analysis or gene-editing experiments.

4.3 Implications for Marker-Assisted and Genomic Selection

From a breeding perspective, the SNP identified in this study represents a promising molecular marker for early selection of body weight traits. Unlike traditional phenotypic selection, which requires animals to reach maturity before evaluation, marker-assisted selection enables genotype-based decision-making at an early developmental stage. Incorporation of this SNP into breeding programs could substantially shorten the selection cycle and improve genetic gain per generation.

Furthermore, the PCR-based genotyping method developed in this study offers a cost-effective, accurate, and scalable approach for routine application in breeding stations. Compared with methods such as PCR-RFLP, direct sequencing avoids enzyme digestion steps, reduces operational complexity, and minimizes genotyping errors. This makes the marker particularly suitable for large-scale breeding programs targeting Pinan cattle, Nanyang cattle, and potentially other related local breeds.

4.4 Limitations and Future Perspectives

Despite the promising results, several limitations should be acknowledged. First, this study primarily establishes a statistical association between the SNP marker and adult body weight; causal relationships and underlying biological mechanisms have not yet been experimentally verified. Second, the effect of this SNP in different genetic backgrounds or environmental conditions remains unclear. Therefore, validation in independent populations and across multiple breeding environments will be essential before large-scale application.

Future studies should integrate functional genomics approaches, including transcriptomic profiling and epigenetic analyses, to elucidate how this locus influences growth-related pathways. Additionally, combining this marker with other significant loci identified through genomic selection models may further enhance prediction accuracy for body weight and related traits.

5. Conclusion

In this study, a SNP marker significantly associated with the adult body weight of Nanyang cattle was successfully identified by whole-genome resequencing combined with GWAS, and an efficient PCR sequencing detection method was established. Nanyang cattle carrying the C allele have significantly higher adult body weight, and this marker can be used as a core tool for molecular breeding of cattle body weight traits, providing important theoretical support and technical guarantee for the genetic improvement of native beef cattle breeds, and has important practical significance for improving the self-sufficiency capacity of China's beef industry.

Acknowledgment

This work was supported in part by the Key Scientific Research Project of Higher Education Institutions in Henan Province (24A230016).

References

[1] Zhang, Y.; Liu, X.; Liu, J.; Fu, T.; Huang, H.; Han, M.; Liang, D.; Gao, T.; Zhang, Y.; Liu, X.; et al. Genomic Insights into Genetic Diversity and Adaptation of Nanyang Cattle: Implications for Conservation and Breeding. Animals 2025, 15, doi:10.3390/ani15203033.

[2] Guo, X.; Zhang, S.; Yang, H.; Pei, J.; Wu, X.; Bao, P.; Liang, C.; Xiong, L.; Chu, M.; Lan, X.; et al. Bovine TMEM95 Gene: Polymorphisms Detecting in Five Chinese Indigenous Cattle Breeds and Their Association with Growth Traits. Electron. J. Biotechnol. 2021, 51, 58–66, doi:10.1016/j.ejbt.2021.03.004.

[3] Zhang, S.; Yao, Z.; Li, X.; Zhang, Z.; Liu, X.; Yang, P.; Chen, N.; Xia, X.; Lyu, S.; Shi, Q.; et al. Assessing Genomic Diversity and Signatures of Selection in Pinan Cattle Using Whole-Genome Sequencing Data. BMC Genomics 2022, 23, 460, doi:10.1186/s12864-022-08645-y.

[4] Qiaozhen Li Study on Precocity and Meat Production Performance of Pinan Cattle. Jiangxi J. Anim. Husb. Vet. Med. 2012, 8–10.

[5] Goddard, M.E.; Hayes, B.J. Genomic Selection. J. Anim. Breed. Genet. 2007, 124, 323–330, doi:10.1111/j.1439-0388.2007.00702.x.
[6] Miller, S. Genomic Selection in Beef Cattle Creates Additional Opportunities for Embryo Technologies to Meet Industry Needs. Reprod. Fertil. Dev. 2022, 35, 98–105, doi:10.1071/RD22233.
[7] Sanchez, M.-P.; Tribout, T.; Kadri, N.K.; Chitneedi, P.K.; Maak, S.; Hozé, C.; Boussaha, M.; Croiseau, P.; Philippe, R.; Spengeler, M.; et al. Sequence-Based GWAS Meta-Analyses for Beef Production Traits. Genet. Sel. Evol. 2023, 55, 70, doi:10.1186/s12711-023-00848-5.
[8] Neumann, G.B.; Korkuć, P.; Arends, D.; Wolf, M.J.; May, K.; König, S.; Brockmann, G.A. Genomic Diversity and Relationship Analyses of Endangered German Black Pied Cattle (DSN) to 68 Other Taurine Breeds Based on Whole-Genome Sequencing. Front. Genet. 2023, 13, 993959, doi:10.3389/fgene.2022.993959.
[9] Neumann, G.B.; Korkuć, P.; Arends, D.; Wolf, M.J.; May, K.; Reißmann, M.; Elzaki, S.; König, S.; Brockmann, G.A. Design and Performance of a Bovine 200 k SNP Chip Developed for Endangered German Black Pied Cattle (DSN). BMC Genomics 2021, 22, 905, doi:10.1186/s12864-021-08237-2.
[10] Wang, H.; Guo, M.; Wei, H.; Chen, Y. Targeting P53 Pathways: Mechanisms, Structures and Advances in Therapy. Signal Transduct. Target. Ther. 2023, 8, 92.
[11] Engeland, K. Cell Cycle Regulation: P53-P21-RB Signaling. Cell Death Differ. 2022, 29, 946–960.