BMC Genomics

RESEARCH Open Access

The CNV map construction and ROH analysis of Pinan cattle

Xingya Song[1†], Zijing Zhang[2†], Shengyan Xing[1†], Xian Liu[3†], Yuqiao Zhang[1], Jiamei Wang[1], Shijie Lyu[2], Xiangnan Wang[2], Xingshan Qi[4], Weidong Ma[5], Chuzhao Lei[1], Eryao Wang[2*] and Yongzhen Huang[1*]

Abstract

Pinan cattle, as the progeny of crossbreeding improvement between Nanyang cattle and Piedmontese, have attracted attention for their excellent growth performance. In this study, we constructed a copy number variation map by whole genome resequencing of 132 Pinan cattle. In the genome of Pinan cattle, deletion-type copy number variants occupied a higher proportion and only 3.31% of CNVRs overlapped with exonic regions. It showed that Pinan cattle was clearly distinguishable from other breeds and Pinan cattle was closer to Nanyang cattle by population genetic structure analysis based on CNVRs. The degree of inbreeding in the Pinan cattle population was explored by ROH analysis, which showed that the degree of inbreeding in Pinan cattle was lower than that in European beef cattle, suggesting that the risk of inbreeding was low. Candidate genes related to muscle development (*CADM3*, *CNTFR*, *DOCK3*), reproductive traits (*SCAPER*), embryonic development (*RERE*) and immune traits (*CD84*) were identified by V_{ST} selection analysis, ROH islands and iHS selection analysis, which provided a new scientific basis for the genetic basis of the excellent traits in Pinan cattle.

Keywords Pinan cattle, Copy number variation, Runs of homozygosity, Genetic structure, Inbreeding degree, Seletion analysis

Introduction

The history of the formation of Chinese native cattle breeds is complex. Previous studies using archaeological and genomic methods have determined that the main ancestry sources of Chinese native cattle are East Asian taurine, Eurasian taurine and Chinese indicine [1]. Follow-up studies have made it clearer that Chinese indicine (East Asian indicine) is different from Indian indicine (South Asian indicine), making the origin of the Chinese native cattle clearer [2]. Nanyang cattle is one of the five excellent Chinese native cattle breeds, with advantages of delicate meat, roughage resistance, environmental adaptability [3]. However, due to the long-term breeding as draft cattle, Nanyang cattle has a slow growth rate, low feed conversion rate, low slaughter rate, and poor economic efficiency which are compared with the beef cattle breeds. So, crossbreeding with excellent beef breeds is an

†Xingya Song, Zijing Zhang, Shengyan Xing and Xian Liu are first authors.

*Correspondence:
Eryao Wang
eryaowang@outlook.com
Yongzhen Huang
hyzsci@nwafu.edu.cn
[1]College of Animal Science and Technology, Northwest A&F University, No. 22 Xinong Road, Yangling, Shaanxi 712100, People's Republic of China
[2]Institute of Animal Husbandry, Henan Academy of Agricultural Sciences, Zhengzhou, Henan 450002, People's Republic of China
[3]Henan Provincial Livestock Technology Promotion Station, Zhengzhou, Henan 450008, People's Republic of China
[4]Biyang County Xiananniu Technology Development Co., Ltd, Zhumadian, Henan 463700, People's Republic of China
[5]Shaanxi Agricultural and Animal Husbandry Seed Farm, Fufeng, Shaanxi 722203, People's Republic of China

important method to improve Nanyang cattle [4]. Then, Piedmontese were introduced to crossbreed with Nanyang cattle in Xinye county of Nanyang city. Through more than 30 years of crossbreeding improvement, Pinan cattle with outstanding growth performance have been bred. Copy number variants (CNVs) is a kind of copy number increase/decrease variations of DNA sequences longer than 50 bp [5–7], and the study of copy number variants can reflect important genomic features such as adaptation and selection of animals [8–10]. Runs of homozygosity (ROH) are formed when homozygous haplotypes are passed from parents to offspring, and to a certain extent, they can reflect the population history of the breed, the degree of inbreeding, and the situation of selection [11–14].

Currently, most of the genomic studies on Pinan cattle were based on SNPs, and the study of CNV in Pinan cattle can enrich the mechanism study of the formation of excellent traits in Pinan cattle. Furthermore, the analysis of inbreeding degree is also an important part of breeding beef cattle breeds.

In this study, we described the distribution characteristics of CNVs and ROHs in Pinan cattle, and analyzed the population genetic structure and inbreeding degree of Pinan cattle. Moreover, we explored the relevant genomic selection regions related to the excellent traits of Pinan cattle, such as fast growth rate and strong meat production capacity, and discovered the genes associated with the outstanding production performance of Pinan cattle. This study can provide a scientific and theoretical basis to Pinan cattle for future population breeding strategies and the improvement of key economic traits.

Results

142 new whole genome resequencing data (132 Pinan cattle and 10 Nanyang cattle) were generated in this study, among which a total of 659,832,685 paired-end reads were generated from Pinan cattle, with an average depth of 7.98× and an average alignment rate of 99.68%. The average depth of Nanyang cattle was 11.81×, and the average comparison rate was 99.85%. The average depth of Qinchuan cattle was 13.77×, and the average comparison rate was 99.75%.

We constructed a CNV set of Pinan cattle, and a total of 9,631 copy number variation regions (CNVRs) were detected on 28 autosomes. The total length was 64,302,650 bp, accounting for 2.58% of the reference genome (ARS-UCD1.2), and the average length of CNVR was 6677 bp. Their distribution on chromosomes is shown in Fig. 1A, and it can be seen that the number and distribution of different types of CNVR on chromosomes are not consistent. We found that the deletion CNVRs were the most common in the genome of Pinan cattle, accounting for 77.85% in number and 63.12% in length. Subsequently, 9,631 CNVRs were functionally annotated, and the results showed that 53.88% of CNVRs were located in the intergenic region, 35.99% of the CNVRs were located in the intron region, and only 3.31% of the CNVRs were located in the exon region (Fig. 1B).

We divided CNVRs into five types according to length (< 5 kb, 5-10 kb, 10-20 kb, 20-50kb, > 50 kb). The results showed that CNVRs shorter than 5 kb were the most numerous and longest in total length (Fig. 1C). At the same time, we found that the duplication CNVRs with a length longer than 100 kb was the least numerous, but the total length was the longest among all duplication CNVRs (Fig. 1D).

Population structure analysis based on CNV

We constructed a CNVR set of all individuals of eight breeds and performed principal component analysis and ancestor component analysis based on the CNVs dataset. In principal component analysis, the first and second principal components accounted for 27.4% and 9.7% of the variations (Fig. 2A). The PC2 will clearly divide all the individuals into two parts: one part is Pinan cattle and Nanyang cattle, and the other part is other Chinese native cattle and European beef cattle. PC1 can distinguish between Pinan cattle and Nanyang cattle. The results of ancestry analysis showed that three European beef cattle breeds had similar ancestral components. The four Chinese native cattle breeds showed three types, and Jiaxian red cattle and Luxi cattle showed high consistency. When K = 5, the ancestral components of Pinan cattle and Nanyang cattle that were not found in other breeds appeared (Fig. 2B).

Selection analysis of Pinan cattle and Chinese native cattle

We calculated the V_{ST} values of Pinan cattle and Chinese native cattle breeds, taking the top 1% of regions as strongly selected regions (Fig. 3A). Then, a total of 356 candidate genes were annotated in these regions, and we screened for a number of genes associated with important economic traits, including muscle development (*TNNT2*, *NFIC*, *WNT7A*, *MMP9*, *CCND2*, *CASZ1*, *SPEG*), adipogenesis (*NPBWR2*, *WNT10A*), trunk development (*NR6A1*), reproduction (*NR5A1*, *DKKL1*). Then, KEGG pathway analysis was performed on these genes. Four pathways with corrected *P*-value < 0.05 were obtained: "Hypertrophic cardiomyopathy", "Human papillomavirus infection" (Corrected *P*-value = 0.0335), and "Hippo signaling pathway" (Corrected *P*-value = 0.0394), "MAPK signaling pathway" (Corrected *P*-value = 0.0477) (Fig. 3B).

We combined the top differentially expressed genes of the longissimus dorsi muscle of Pinan cattle and Nanyang cattle screened in the existing literature [15] and the annotated 356 genes to obtain two genes, *CADM3*

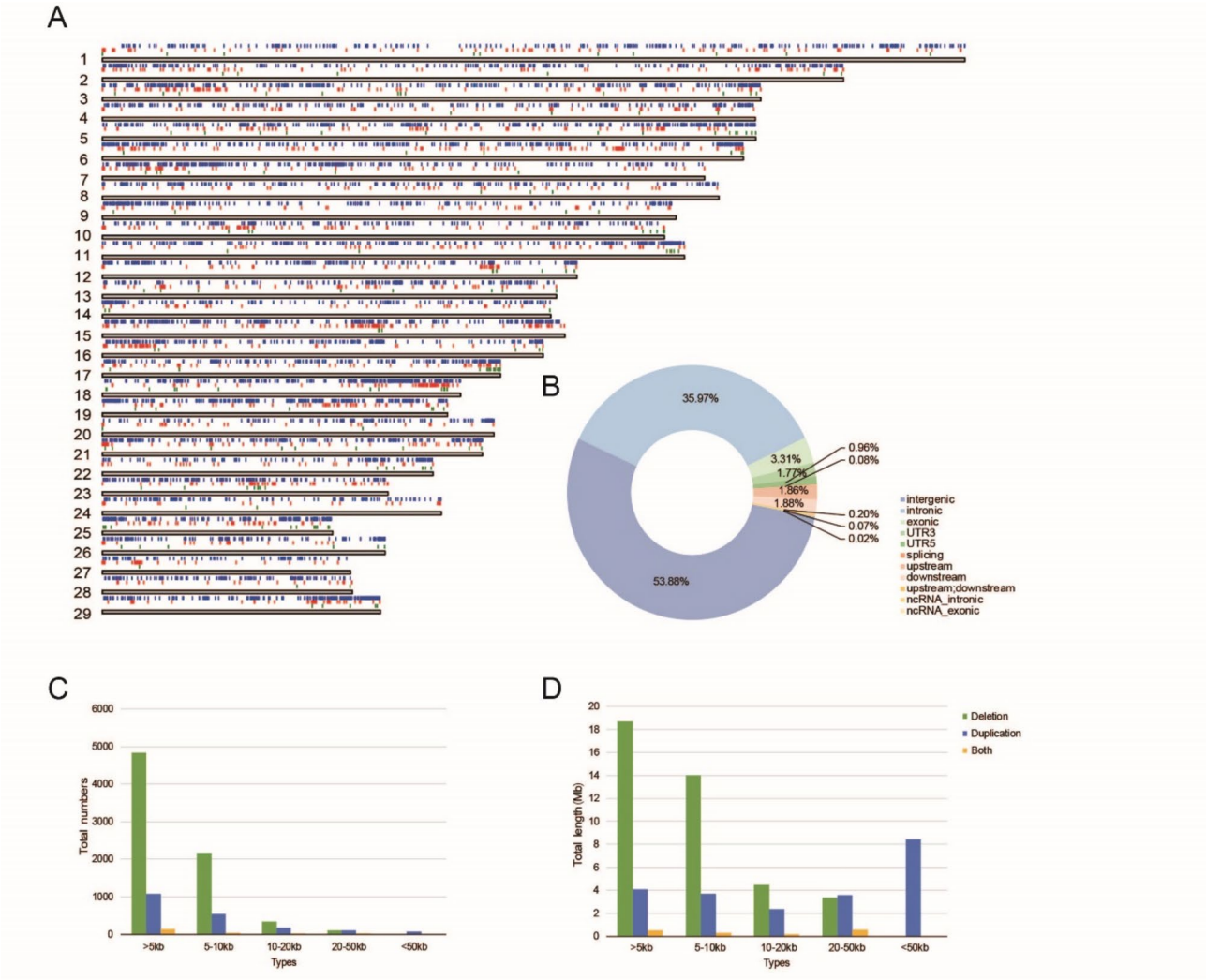

Fig. 1 Distribution characteristics of CNVR in Pinan cattle. **A** the CNVR map of Pinan cattle in autosomes. **B** Functional classification of the detected CNVRs. **C** Total number of different CNVR types. **D** Total length of different CNVR types

and *CNTFR* (Fig. 3C). We calculated the distribution frequency of CNV corresponding to the two genes in different populations. The results showed that the CNV corresponding to the *CADM3* gene was deletion, which had a high frequency in European beef cattle breeds and Pinan cattle, and a low frequency in Chinese native cattle breeds. The CNV corresponding to the *CNTFR* gene was also deletional, with a high frequency in European beef cattle breeds and Pinan cattle, and a low frequency in Chinese native cattle breeds (Fig. 3D and E).

Runs of homozygosity detection and distribution studies

A total of 11,314 ROHs with a total length of 9,964,170.13 kb were detected in this Pinan cattle population, with the average length of 880.69 kb. The shortest ROH is 500.009 kb containing 5407 SNPs and the longest ROH is 12,760.034 kb containing 159,026 SNPs. The average number of ROHs per sample was 85. The average total length of ROHs in each sample was 75,486.14 kb, and the genome coverage of ROHs per sample was 3.03%. Figure 4A shows the distribution of ROH on different chromosomes in the Pinan cattle population. The most distribution of ROHs is on BTA1 (810 ROHs) and the least distribution of ROHs is on BTA25 (50 ROHs), which is similar to the distribution in the previous study of Chinese Simmental beef cattle. The ROH coverage on BTA21 is the highest (4.50%), and the ROH coverage on BTA25 is the lowest (0.65%). Figure 4B depicted the total number and total length of ROHs for each individual. Individuals with a total length of ROH (> 200 Mb) are all European beef cattle or Pinan cattle.

To know the inbreeding level in the Pinan cattle population, we calculate the inbreeding coefficient for all populations, and the inbreeding coefficient for the Pinan population is 0.0303. It was found that the inbreeding coefficient of Pinan cattle was lower than that of

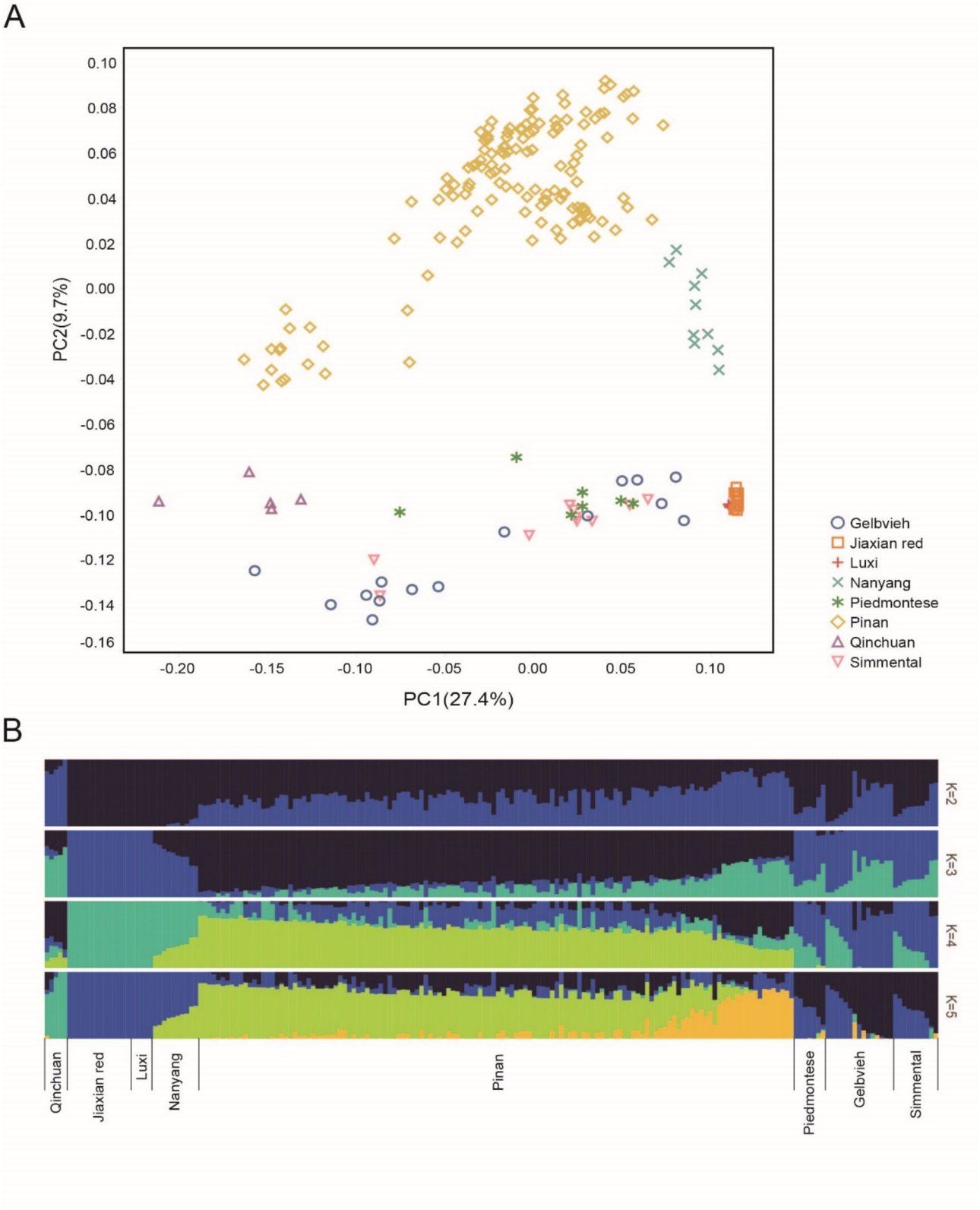

Fig. 2 Population structure analysis. **A** Principal component analysis. **B** Ancestral component analysis (K = 2, 3, 4, 5)

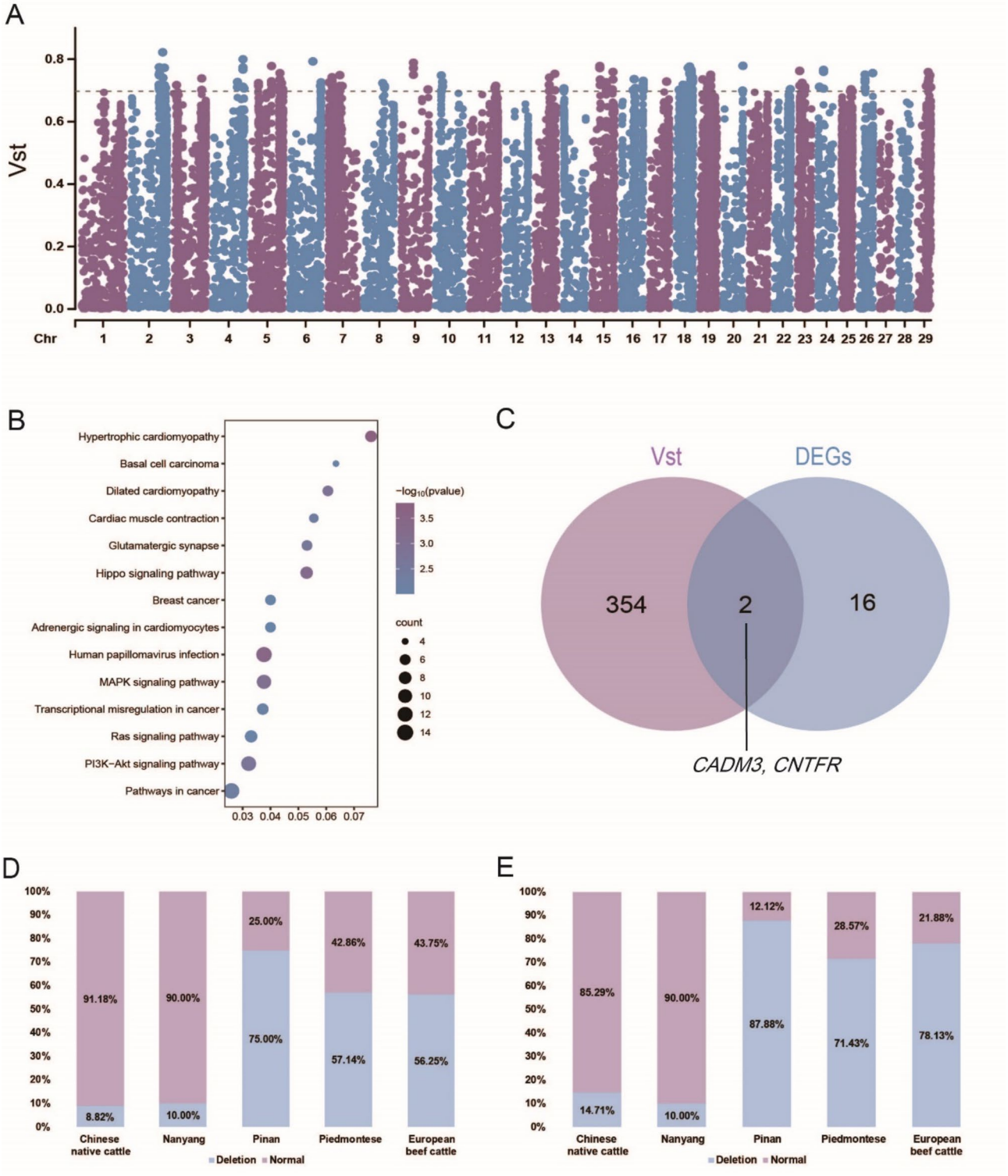

Fig. 3 Selection analysis based on CNVs and candidate genes analysis. **A** Manhattan plot of V_{ST} in Pinan cattle and Chinese native cattle. **B** KEGG pathways from the enrichment analysis. **C** Venn diagram of the candidate genes in this study and the DEGs in the previous study. **D** Frequency of *CADM3*-CNVRs in different populations. **E** Frequency of *CNTFR*-CNVRs in different populations

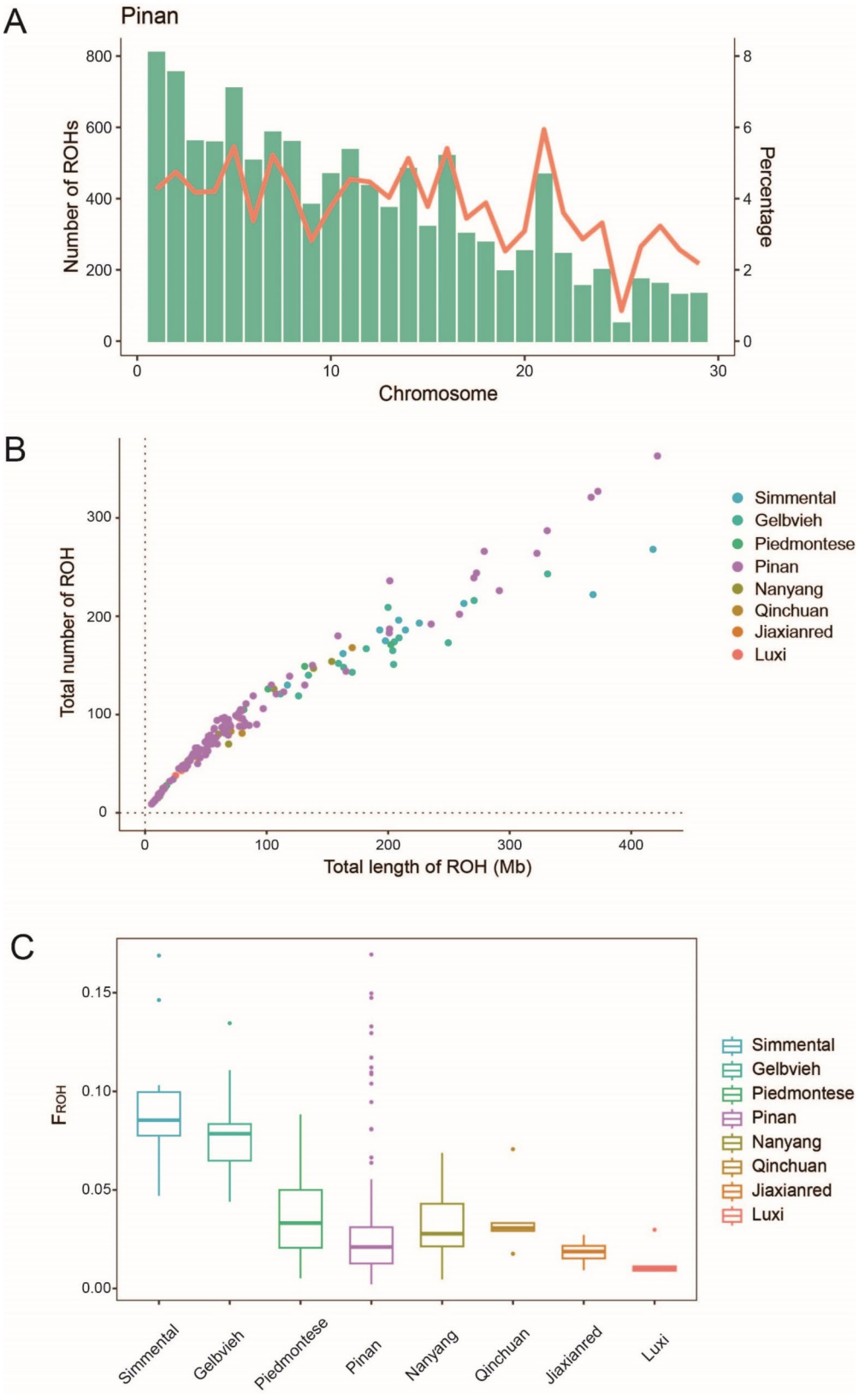

Fig. 4 The total number and coverage of ROH on each autosomes in the genome of Pinan cattle. **B** Scatter plot of the total number of ROHs and the total length of ROHs for each individual within each breed. **C** Box plot of F_{ROH} in each breed

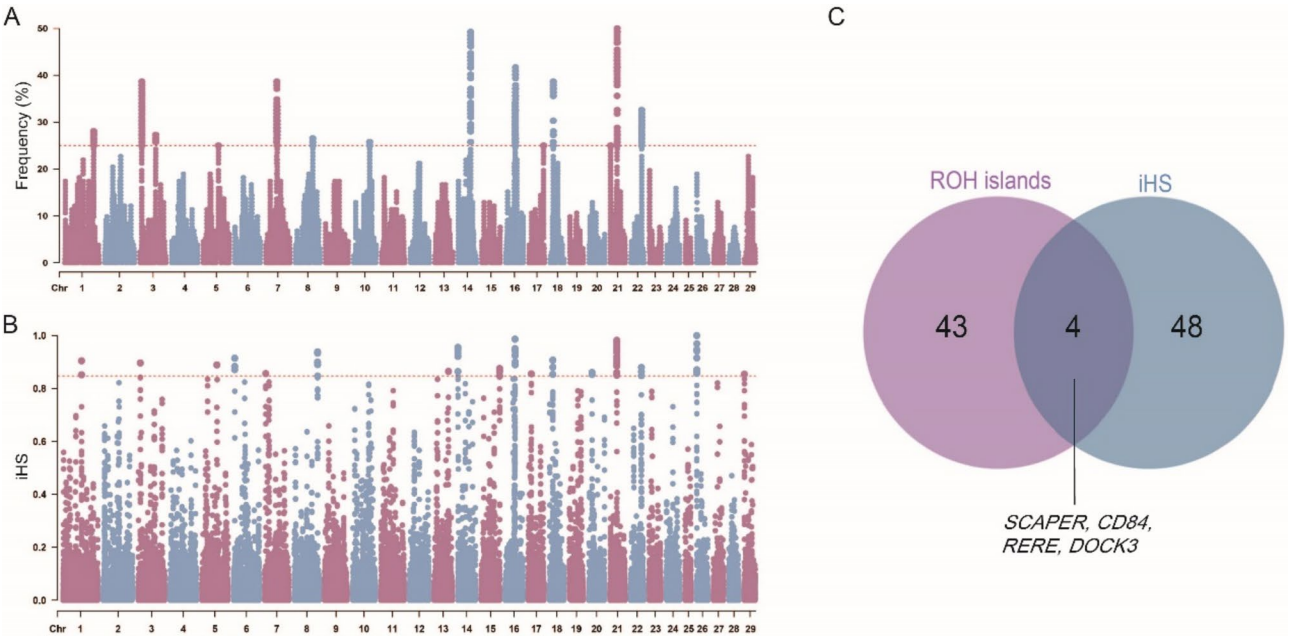

Fig. 5 **A** the ROH islands of Pinan cattle. **B** Selection analysis by integrated haplotype score (iHS). **C** Venn diagram of the candidate genes by ROH islands and iHS analysis

European beef cattle breeds, and it was similar to that of Nanyang cattle (Fig. 4C). It's shown that the inbreeding risk of Pinan cattle population was low, but there are also some individuals with high inbreeding level.

Analysis of selection characteristics of Pinan cattle

A total of 3,484 ROH islands were detected in the Pinan cattle. Among them, 38 high-frequency ROH enrichment regions (frequency greater than 25%) were found. In these 38 islands, 47 candidate genes and 64 QTLs associated with important traits were identified (Fig. 5A). At the same time, we calculated the iHS (Integrated Haplotype Score) of the Pinan cattle population, and selected the top 1% regions as the strong selection regions (Fig. 5B). Then we obtained 52 selected genes after annotation, and jointly screened four key candidate genes (*SCAPER*, *CD84*, *RERE*, *DOCK3*) (Fig. 5C).

Discussion

Genetic variation is a specific manifestation of artificial selection in the genome of domestic animals, and CNV is one of the main constituents. In recent years, CNV atlases of many livestock have been constructed [9, 10, 16–21], and a large number of CNVs associated with important traits in livestock have been identified [22–24]. In this study, 9,631 CNVRs were detected in the Pinan cattle population, and the CNV map of Pinan cattle is similar to that of the previous study in Chinese cattle, and it also showed that the deletion type was the majority [22]. It suggested that deletions are more likely to be present in the genome than duplications. This may be due to the fact that deletions are more likely to occur during DNA replication. It may also be affected by read depth and CNV detection software, which exhibits lower sensitivity for identifying duplication [25]. Previous studies have analyzed the population structure of Pinan cattle based on SNP data, and have found that Pinan cattle are closer to Piedmontese [26], but the results of this study based on CNV show that Pinan cattle are closer to Nanyang cattle. It may relate to the different selection pressures of SNP and CNV in the process of artificial selection.

High meat yield is an important goal for breeding of Pinan cattle. In the comparison of Pinan cattle and Chinese native cattle breeds, we noticed that the "Hippo signaling pathway" and "MAPK signaling pathway" in the significant pathways of candidate genes are related to skeletal muscle development [27–30], suggesting that the two pathways may play an important role in the high meat yield traits of Pinan cattle. Among these candidate genes, we found some genes involved in muscle development. WNT7A (Wingless-related integration site 7 A) is a member of the WNT family, and it has been found that intramuscular injection of WNT7A protein can increase muscle mass and muscle strength in mdx mice (a mouse model of Duchenne muscular dystrophy), and produce muscle fiber hypertrophy and decreased muscle fiber necrosis [31]. Subsequent deletion and salvage experiments demonstrated that WNT7A is required for effective muscle regeneration in mdx mice [32]. As cellular transcription factors and DNA replication factors, the Nuclear factor I (NFI) family plays an important role in

mammalian development. There was a study found that *NFIC* gene was highly expressed in bovine muscle tissue, and knockdown of NFIC gene would promote the proliferation of bovine myoblasts, and found that CENPF, as a downstream target gene of NFIC, could affect the expression of CDK1 and CCNB1, actively regulate cell cycle pathways and cell proliferation, and finally found that NFIC acts on the CENPF/CDK1 axis to regulate the mechanism of bovine myoblast proliferation [33]. Moreover, we noted two key candidate genes, *CADM3* and *CNTFR*, which were also found as the top differentially expressed genes for the longissimus dorsi muscle of Pinan cattle and Nanyang cattle [15]. CADM3 is a member of the cell adhesion factor family and plays a role primarily in the development of neurons, regulating synapse formation [34–36]. The *CNTFR* gene encodes a member of the type 1 cytokine receptor family. The encoded protein is a ligand-specific component of the ciliary neurotrophin triple receptor and plays a key role in neuronal cell survival, differentiation, and gene expression. There was a previous study found that SNPs in *CNTFR* gene were associated with changes in muscle strength [37, 38]. A beef cattle SNP panel study found that *CNTFR* had an effect on increasing average daily gain (ADG) in beef cattle [39]. These genes are likely to be associated with the high meat yield of Pinan cattle.

In the process of breeding livestock breeds, the genome is affected by factors such as parenting, selection intensity, and mating mode, so the number, length and distribution frequency of ROH in the population also show certain differences [12, 40–42]. In this study, the ROH length of European beef cattle breeds was longer than that of Chinese native cattle breeds, suggesting that European beef cattle breeds had been more strongly selected. There are large differences in the coverage of ROHs in different chromosomes, which indicates that different chromosomes are subjected to different selection pressures. The calculation of inbreeding coefficient showed that the degree of inbreeding of Pinan cattle was lower than that of European beef cattle breeds, and it was similar to that of Nanyang cattle, and the risk of inbreeding was smaller, but there were still individuals with inbreeding. It also showed that the utilization rate of excellent individuals can be appropriately improved and the selection efforts can be strengthened in the breeding of Pinan cattle. Several studies have confirmed that the homozygosity within the genome of livestock after selection has been greatly improved, resulting in more ROH-rich regions within the population, ROH islands [14, 43–45]. Based on these regions, QTL annotation was carried out, and four candidate genes were screened based on ROH islands and iHS. DOCK (dedicator of cytokinesis) is an 11-member family of typical guanine nucleotide exchange factors (GEFs) expressed in the brain, spinal cord, and skeletal muscle. DOCK3 is a member of DOCK family which play an important role in skeletal muscle development. The knockout of DOCK3 in mice showed that the muscle structure of the knocked mice was damaged, muscle fiber regeneration was impaired and metabolic dysfunction was impaired, which proved the important role of DOCK3 in skeletal muscle [46]. S-phase cyclin A-associated protein in the endoplasmic reticulum (SCAPER) interacts with cyclin A and functions as a feedback loop regulator in the G1/S and G2/M phases of the cell cycle [47]. SCAPER has been found to be associated with male sterility in multiple species (human, cattle, sheep, mice, fruit flies) [48–51]. It may be related to the stronger reproductive performance of Pinan cattle. CD84-mediated signaling regulates diverse immunological processes, including T cell cytokine secretion, natural killer cell cytotoxicity [52]. Previous studies have found that the region in which this gene is located is strongly selected in Chinese local cattle, suggesting that this gene may be related to better disease resistance in Chinese native cattle [53, 54]. Arginine-glutamic acid dipeptide repeats (RERE) is associated with embryonic development, and mutations in RERE can lead to asymmetric defects in mouse embryos [55]. It is possible that these genes on ROH islands play important roles in the formation of excellent traits in Pinan cattle.

The detection of CNV, especially short fragments of CNVs and complex variants, was limited by depth of next-generation sequence. Long-read sequencing can be used to improve the accuracy of CNV identification and molecular experiments can further verify the function of genes in the next step. In addition, the identification of ROH can be affected by software parameters, and comparisons between different groups within a single study were relatively accurate. There is a need for uniform standard in ROH study of livestock.

Methods

Sample collection and whole genome sequencing

The 132 Pinan cattle in this study were females between 2 and 6 years old selected from the core breeding area of Pinan Cattle in Xinye County, Nanyang City, Henan Province, China. DNAs were extracted from blood to construct a 300 bp library, which were sequenced by BGI for whole genome resequencing.

The study also used data from 10 Nanyang cattle and 5 Qinchuan cattle collected by our lab. In addition, public data of 2 Chinese native cattle breeds (14 Jiaxian red cattle and 5 Luxi cattle) and 3 South-central European beef cattle breeds (7 Piedmontese, 10 Simmental and 15 Gelbvieh) were downloaded.

Genomic data processing and CNVR identification

The raw data was filtered using Trimmomitic v0.38 with the parameters: "LEADING:20, TRAILING:20, SLIDINGWINDWOE: 3:15, AVGQUAL:20, MINLEN:35, TOPHRED33" [56], followed by alignment of reads to the reference genome (ARS-UCD1.2) using BWA-MEM (version 0.7.13-r1126) with default parameters [57] and deduplication using the "BaseRecalibrator" and "ApplyBQSR" modules in GATK (version 4.3.0.0). CNVcaller [58] was used to detect the CNVs. Subsequently, we used a 1500 bp window and a 750 bp step size to count the GC, repeat, and gap content of each window in the reference genome, and calculated the absolute copy number of each window for each individual to determine the boundaries of the CNV region (parameter: -f 0.1 -h 3 -r 0.3). CNVR is the region with a uniform boundary merged from CNVs originating from different individuals. We classified CNVRs into three types: deletion, duplication and both. The CNVRs were filtered by silhouette coefficient and length: (1) Length: the length of deletion and both CNVRs was ≤ 50 kb, and the length of duplication CNVRs was < 500 kb; (2) Silhouette coefficient: The silhouette coefficient of duplication and deletion is required to be higher than 0.25, and the group silhouette coefficient of both is lower than 0.75. ANNOVAR [59] was used to annotate the function regions of CNVRs.

Population structure analysis

Principal component analysis was performed on all individuals using PLINK v1.9 (--pca 10) [60]. Ancestor component analysis was performed using ADMIXTURE [61], with K values ranging from 2 to 5. Pophelper [62] was used for visualization of stacked graphs.

Selection analysis of Pinan cattle and Chinese native cattle breeds

We used a 50 kb window and a 20 kb step size to calculate the V_{ST} for each window for the selected analysis of Pinan cattle and Chinese native cattle breeds. V_{ST} is a common method for interpopulation selection based on CNV, similar to F_{ST}. The formula is $V_{ST} = (Vt\text{-}Vs) / Vt$. Vt represents the standard deviation of the copy number size of the region for all samples, and Vs represents the value of standard deviation after each population weighted according to the size of their respective populations [9].

The top 1% of areas are defined as areas that have received strong selection. We used ANNOVAR [59] to annotate candidate genes involved in these regions. In order to screen the genes associated with the high meat yield of Pinan cattle, we intersected these candidate genes with the DEGs of the longissimus dorsi muscle of Pinan cattle and Nanyang cattle in the previous study [15], and obtained two key candidate genes, and examined the distribution of CNVs of these two genes in different populations.

ROH detection and inbreeding coefficient calculation

The detection and filtration of SNPs using GATK was based on the previous research of our team. PLINK was used to detect ROH on each individual autosome, and the following criteria were used: (1) a minimum length of ROH of 500 kb, (2) at least 1 SNP in the range of 50 kb in ROH, (3) a minimum of 50 SNPs in ROH, (4) a sliding window size of 50 SNPs, (5) a maximum of 3 SNPs in the sliding window that were heterozygous, and (6) a maximum of 5 SNP deletions in the sliding window.

All ROHs are divided into four types according to length: 500 kb – 1000 kb, 1000–2000 kb, 2000–4000 kb, > 4000 kb. Subsequently, the genomic inbreeding coefficient F_{ROH} within each population was calculated as the method in a previous study [63], as the following formula is $F_{ROH} = L_{ROH} / L_{Genome}$. L_{ROH} is the length of all ROH, and L_{Genome} is the length of all autosomes.

Identification and selection characteristics of ROH Islands

The ROH-enriched region in the genome of Pinan cattle was detected by "--homozyg" in PLINK, and the top 1% of the ROH-enriched region was selected as ROH regions, ROH islands with the high-frequency, and the threshold line was 25%. In order to better understand the selection characteristics of the genome of Pinan cattle, we used the selscan (version1.3) [64] to calculate the iHS on the genome of Pinan cattle using a 50 kb window and a 20 kb step size, and then normalized the scores using the norm module, and also selected the top 1% of regions as regions subject to strong selection for gene annotation.

Enrichment analysis and QTL annotation

In this study, KEGG and GO pathway analysis were performed on the candidate genes and KOBAS3.0 [65] for these genes, and significant enrichment pathways were screened based on corrected p-values less than 0.05. Quantitative trait loci (QTL) data of cattle was obtained from AnimalQTLdb [66].

Conclusions

In this study, the CNVs and ROHs of Pinan cattle were analyzed by whole genome sequencing, and the CNV map of Pinan cattle was constructed, and the characteristics of genome CNV and individual inbreeding degree of Pinan cattle population were understood. Candidate genes that may be related to excellent traits such as high meat yield, good disease resistance and strong fecundity of Pinan cattle were screened. Further molecular experimentation is warranted to confirm the functional roles of these genes, which could serve as molecular genetic markers for improved Chinese native cattle in the future.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11626-6.

Supplementary Material 1: **Additional file**: **Table S1**. A summary of new sequencing data in this study. **Table S2**. List of additional cattle samples for analysis in this study. **Table S3**. The number of different CNVR types in Pinan cattle. **Table S4**. The total length of different CNVR types in Pinan cattle. **Table S5**. Coefficient of variation (CV) errors for ADMIXTURE ancestry models with K value from 2 to 9. **Table S6**. A summary of top differentiated genes screened by V_{ST} method between Pinan cattle and Chinese native cattle. **Table S7**. KEGG pathway analysis of candidate genes in V_{ST} analysis. **Table S8**. The ROH statistics for each individual. **Table S9**. The ROH number of four types in different groups based on length. **Table S10**. The gene and QTL annotation of top ROH islands (more than 25%). **Table S11**. The gene annotation of top 1% regions in iHS analysis

Acknowledgements
We would like to thank Yu Jiang for his technical support. We thank the High-Performance Computing platform of Northwest A&F University for providing computing resources.

Author contributions
YH conceived and designed the experiments. XS and YZ performed the statistical analysis and data upload. SX and JW performed the sample DNA extraction. ZZ, XL, XW, SL, and EW provided suggestions for the revision of the manuscript. YJ, CL, and SQ provided technical assistance. XQ, WM and EW contributed to the sample collections. YH provided the laboratories for DNA extraction and statistical analysis. XS drafted the manuscript. All authors read and approved the final manuscript.

Funding
This research has been supported by STI2030-Major Projects (2023ZD040480206); National Key R&D Plan (2022YFD1602310); Breeding and production of cattle and sheep by scientific and technological innovation team of Henan Academy of Agricultural Sciences (2023TD25); Major Science and Technology Projects in Henan Province (221100110200); Henan Beef Cattle Industrial Technology System (HARS-22-13-S); China Agriculture Research System of MOF and MARA (CARS-37).

Data availability
Sequences are available from the National Center of Biotechnology Information (NCBI) database. Bioproject accession number is PRJNA1173901.

Declarations

Ethics approval and consent to participate
All cattle were handled following the guidelines established by the Council for Animal Welfare of China. The protocols for sample collection and animal handling have been approved by the Faculty of Animal Policy and Welfare Committee of Northwest A&F University (FAPWCNWAFU, Protocol number, NWAFAC 1008). The study was carried out in compliance with the ARRIVE guidelines.

Consent for publication
This publication was obtained consent from all authors.

Competing interests
The authors declare no competing interests.

Received: 25 November 2024 / Accepted: 22 April 2025
Published online: 14 May 2025

References

1. Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, Hu S, Huang S, Zhang H, Zheng Z, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat Commun. 2018;9(1):2337.
2. Chen N, Xia X, Hanif Q, Zhang F, Dang R, Huang B, Lyu Y, Luo X, Zhang H, Yan H, et al. Global genetic diversity, introgression, and evolutionary adaptation of indicine cattle revealed by whole genome sequencing. Nat Commun. 2023;14(1):7803.
3. Zhang Y, Wei Z, Zhang M, Wang S, Gao T, Huang H, Zhang T, Cai H, Liu X, Fu T et al. Population structure and selection signal analysis of Nanyang cattle based on Whole-Genome sequencing data. Genes 2024, 15(3).
4. Song X, Yao Z, Zhang Z, Lyu S, Chen N, Qi X, Liu X, Ma W, Wang W, Lei C, et al. Whole-genome sequencing reveals genomic diversity and selection signatures in Xia'nan cattle. BMC Genomics. 2024;25(1):559.
5. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annu Rev Genom Hum Genet. 2009;10:451–81.
6. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The database of genomic variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 2014;42(Database issue):D986–992.
7. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. Nat Genet. 2007;39(7 Suppl):S37–42.
8. Hollox EJ, Zuccherato LW, Tucci S. Genome structural variation in human evolution. Trends Genet. 2022;38(1):45–58.
9. Huang Y, Li Y, Wang X, Yu J, Cai Y, Zheng Z, Li R, Zhang S, Chen N, Asadollahpour Nanaei H, et al. An atlas of CNV maps in cattle, goat and sheep. Sci China Life Sci. 2021;64(10):1747–64.
10. Shi H, Li T, Su M, Wang H, Li Q, Lang X, Ma Y. Identification of copy number variation in Tibetan sheep using whole genome resequencing reveals evidence of genomic selection. BMC Genomics. 2023;24(1):555.
11. Zhang Q, Guldbrandtsen B, Bosse M, Lund MS, Sahana G. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genomics. 2015;16(1):542.
12. Xu L, Zhao G, Yang L, Zhu B, Chen Y, Zhang L, Gao X, Gao H, Liu GE, Li J. Genomic patterns of homozygosity in Chinese local cattle. Sci Rep. 2019;9(1):16977.
13. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. BMC Genet. 2012;13:70.
14. Mulim HA, Brito LF, Pinto LFB, Ferraz JBS, Grigoletto L, Silva MR, Pedrosa VB. Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. BMC Genomics. 2022;23(1):209.
15. Wei X, Zhu Y, Zhao X, Zhao Y, Jing Y, Liu G, Wang S, Li H, Ma Y. Transcriptome profiling of mRNAs in muscle tissue of Pinan cattle and Nanyang cattle. Gene. 2022;825:146435.
16. Dang D, Zhang L, Gao L, Peng L, Chen J, Yang L. Analysis of genomic copy number variations through whole-genome scan in Yunling cattle. Front Veterinary Sci. 2024;11:1413504.
17. Zheng X, Zhao P, Yang K, Ning C, Wang H, Zhou L, Liu J. CNV analysis of Meishan pig by next-generation sequencing and effects of AHR gene CNV on pig reproductive traits. J Anim Sci Biotechnol. 2020;11:42.
18. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M, Taylor JF, et al. Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res. 2012;22(4):778–90.
19. Couldrey C, Keehan M, Johnson T, Tiplady K, Winkelman A, Littlejohn MD, Scott A, Kemper KE, Hayes B, Davis SR, et al. Detection and assessment of copy number variation using PacBio long-read and illumina sequencing in new Zealand dairy cattle. J Dairy Sci. 2017;100(7):5472–8.
20. Yang L, Han J, Deng T, Li F, Han X, Xia H, Quan F, Hua G, Yang L, Zhou Y. Comparative analyses of copy number variations between swamp buffaloes and river buffaloes. Anim Genet. 2023;54(2):199–206.
21. Rao YS, Li J, Zhang R, Lin XR, Xu JG, Xie L, Xu ZQ, Wang L, Gan JK, Xie XJ, et al. Copy number variation identification and analysis of the chicken genome using a 60K SNP BeadChip. Poult Sci. 2016;95(8):1750–6.
22. Mei C, Junjvlieke Z, Raza SHA, Wang H, Cheng G, Zhao C, Zhu W, Zan L. Copy number variation detection in Chinese Indigenous cattle by whole genome sequencing. Genomics. 2020;112(1):831–6.
23. Wu J, Wu T, Xie X, Niu Q, Zhao Z, Zhu B, Chen Y, Zhang L, Gao X, Niu X et al. Genetic Association Analysis of Copy Number Variations for Meat Quality in Beef Cattle. Foods (Basel, Switzerland). 2023, 12(21).
24. da Silva JM, Giachetto PF, da Silva LO, Cintra LC, Paiva SR, Yamagishi ME, Caetano AR. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. BMC Genomics. 2016;17:454.

25. Teo SM, Pawitan Y, Ku CS, Chia KS, Salim A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. Bioinf (Oxford England). 2012;28(21):2711–8.
26. 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(1):460.
27. Keren A, Tamir Y, Bengal E. The p38 MAPK signaling pathway: a major regulator of skeletal muscle development. Mol Cell Endocrinol. 2006;252(1–2):224–30.
28. Liu SY, Chen LK, Jhong YT, Chen CW, Hsiao LE, Ku HC, Lee PH, Hwang GS, Juan CC. Endothelin-1 impairs skeletal muscle myogenesis and development via ETB receptors and p38 MAPK signaling pathway. Clinical science (London, England: 1979) 2024;138(12):711–723.
29. Hulmi JJ, Oliveira BM, Silvennoinen M, Hoogaars WM, Ma H, Pierre P, Pasternack A, Kainulainen H, Ritvos O. Muscle protein synthesis, mTORC1/MAPK/Hippo signaling, and capillary density are altered by blocking of myostatin and activins. Am J Physiol Endocrinol Metabolism. 2013;304(1):E41–50.
30. Watt KI, Goodman CA, Hornberger TA, Gregorevic P. The Hippo signaling pathway in the regulation of skeletal muscle mass and function. Exerc Sport Sci Rev. 2018;46(2):92–6.
31. von Maltzahn J, Renaud JM, Parise G, Rudnicki MA. Wnt7a treatment ameliorates muscular dystrophy. Proc Natl Acad Sci USA. 2012;109(50):20614–9.
32. Gurriaran-Rodriguez U, Kodippili K, Datzkiw D, Javandoost E, Xiao F, Rejas MT, Rudnicki MA. Wnt7a is required for regeneration of dystrophic skeletal muscle. Skelet Muscle. 2024;14(1):34.
33. Wang J, Guo J, Yu S, Yu H, Kuraz AB, Jilo DD, Cheng G, Li A, Jia C, Zan L. Knockdown of NFIC promotes bovine myoblast proliferation through the CENPF/CDK1 Axis. J Agric Food Chem. 2024;72(22):12641–54.
34. Rebelo AP, Cortese A, Abraham A, Eshed-Eisenbach Y, Shner G, Vainshtein A, Buglo E, Camarena V, Gaidosh G, Shiekhattar R, et al. A CADM3 variant causes Charcot-Marie-Tooth disease with marked upper limb involvement. Brain. 2021;144(4):1197–213.
35. Tanabe Y, Fujita E, Hayashi YK, Zhu X, Lubbert H, Mezaki Y, Senoo H, Momoi T. Synaptic adhesion molecules in Cadm family at the neuromuscular junction. Cell Biol Int. 2013;37(7):731–6.
36. Sukhanov N, Vainshtein A, Eshed-Eisenbach Y, Peles E. Differential contribution of Cadm1-Cadm3 cell adhesion molecules to peripheral myelinated axons. J Neuroscience: Official J Soc Neurosci. 2021;41(7):1393–400.
37. Khanal P, Morse CI, He L, Herbert AJ, Onambélé-Pearson GL, Degens H, Thomis M, Williams AG, Stebbings GK. Polygenic models partially predict muscle size and strength but not low muscle mass in older women. Genes 2022, 13(6).
38. Homma H, Kobatake N, Sekimoto Y, Saito M, Mochizuki Y, Okamoto T, Nakazato K, Nishiyama T, Kikuchi N. Ciliary neurotrophic factor receptor rs41274853 polymorphism is associated with weightlifting performance in Japanese weightlifters. J Strength Conditioning Res. 2020;34(11):3037–41.
39. Abo-Ismail MK, Lansink N, Akanno E, Karisa BK, Crowley JJ, Moore SS, Bork E, Stothard P, Basarab JA, Plastow GS. Development and validation of a small SNP panel for feed efficiency in beef cattle. J Anim Sci. 2018;96(2):375–97.
40. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 2018;19(4):220–34.
41. Liu SH, Ma XY, Hassan FU, Gao TY, Deng TX. Genome-wide analysis of runs of homozygosity in Italian mediterranean Buffalo. J Dairy Sci. 2022;105(5):4324–34.
42. Falchi L, Cesarani A, Mastrangelo S, Senczuk G, Portolano B, Pilla F, Macciotta NPP. Analysis of runs of homozygosity of cattle living in different climate zones. J Anim Sci 2023, 101.
43. Nosrati M, Asadollahpour Nanaei H, Javanmard A, Esmailizadeh A. The pattern of runs of homozygosity and genomic inbreeding in world-wide sheep populations. Genomics. 2021;113(3):1407–15.
44. Wang X, Li G, Ruan D, Zhuang Z, Ding R, Quan J, Wang S, Jiang Y, Huang J, Gu T, et al. Runs of homozygosity uncover potential Functional-Altering mutation associated with body weight and length in two duroc pig lines. Front Veterinary Sci. 2022;9:832633.
45. Signer-Hasler H, Henkel J, Bangerter E, Bulut Z, Drögemüller C, Leeb T, Flury C. Runs of homozygosity in Swiss goats reveal genetic changes associated with domestication and modern selection. Genet Selection Evolution: GSE. 2022;54(1):6.
46. Samani A, Karuppasamy M, English KG, Siler CA, Wang Y, Widrick JJ, Alexander MS. DOCK3 regulates normal skeletal muscle regeneration and glucose metabolism. FASEB Journal: Official Publication Federation Am Soc Experimental Biology. 2023;37(10):e23198.
47. Tsang WY, Wang L, Chen Z, Sánchez I, Dynlacht BD. SCAPER, a novel Cyclin A-interacting protein that regulates cell cycle progression. J Cell Biol. 2007;178(4):621–33.
48. Wormser O, Levy Y, Bakhrat A, Bonaccorsi S, Graziadio L, Gatti M, AbuMadighem A, McKenney RJ, Okada K, El Riati S, et al. Absence of SCAPER causes male infertility in humans and Drosophila by modulating microtubule dynamics during meiosis. J Med Genet. 2021;58(4):254–63.
49. Tatour Y, Bar-Joseph H, Shalgi R, Ben-Yosef T. Male sterility and reduced female fertility in SCAPER-deficient mice. Hum Mol Genet. 2020;29(13):2240–9.
50. Serrano M, Ramón M, Calvo JH, Jiménez M, Freire F, Vázquez JM, Arranz JJ. Genome-wide association studies for sperm traits in Assaf sheep breed. Animal: Int J Anim Bioscience. 2021;15(2):100065.
51. Ghoreishifar M, Vahedi SM, Salek Ardestani S, Khansefid M, Pryce JE. Genome-wide assessment and mapping of inbreeding depression identifies candidate genes associated with semen traits in Holstein bulls. BMC Genomics. 2023;24(1):230.
52. Cuenca M, Sintes J, Lányi Á, Engel P. CD84 cell surface signaling molecule: an emerging biomarker and target for cancer and autoimmune disorders. Clin Immunol (Orlando Fla). 2019;204:43–9.
53. Xia X, Zhang S, Zhang H, Zhang Z, Chen N, Li Z, Sun H, Liu X, Lyu S, Wang X, et al. Assessing genomic diversity and signatures of selection in Jiaxian red cattle using whole-genome sequencing data. BMC Genomics. 2021;22(1):43.
54. Ma X, Cheng H, Liu Y, Sun L, Chen N, Jiang F, You W, Yang Z, Zhang B, Song E et al. Assessing genomic diversity and selective pressures in Bohai black cattle using Whole-Genome sequencing data. Animals: Open Access J MDPI 2022, 12(5).
55. Vilhais-Neto GC, Maruhashi M, Smith KT, Vasseur-Cognet M, Peterson AS, Workman JL, Pourquié O. Rere controls retinoic acid signalling and Somite bilateral symmetry. Nature. 2010;463(7283):953–7.
56. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinf (Oxford England). 2014;30(15):2114–20.
57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinf (Oxford England). 2009;25(14):1754–60.
58. Wang X, Zheng Z, Cai Y, Chen T, Li C, Fu W, Jiang Y. CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. GigaScience. 2017;6(12):1–12.
59. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
60. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
61. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry Estimation. BMC Bioinformatics. 2011;12:246.
62. Francis RM. Pophelper: an R package and web app to analyse and visualize population structure. Mol Ecol Resour. 2017;17(1):27–32.
63. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008;83(3):359–72.
64. Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014;31(10):2824–7.
65. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Res. 2021;49(W1):W317–25.
66. Hu ZL, Fritz ER, Reecy JM. AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. Nucleic Acids Res. 2007;35(Database issue):D604–609.

Publisher's note