BMC Genomics

RESEARCH

Open Access

Development and validation of a 5K low-density SNP chip for Hainan cattle

Huan Wang[1†], Hui Wu[2†], Wencan Zhang[1], Junming Jiang[1], Hejie Qian[1], Churiga Man[1], Hongyan Gao[1], Qiaoling Chen[1], Li Du[1], Si Chen[1*] and Fengyang Wang[1*]

Abstract

Background This study aimed to design and develop a 5K low-density liquid chip for Hainan cattle utilizing targeted capture sequencing technology. The chip incorporates a substantial number of functional single nucleotide polymorphism (SNP) loci derived from public literature, including SNP loci significantly associated with immunity, heat stress, meat quality, reproduction, and other traits. Additionally, SNPs located in the coding regions of immune-related genes from the Bovine Genome Variation Database (BGVD) and Hainan cattle-specific SNP loci were included.

Results A total of 5,293 SNPs were selected, resulting in 9,837 DNA probes with a coverage rate of 85.69%, thereby creating a Hainan cattle-specific 5K Genotyping by Target Sequencing (GBTS) liquid chip. Evaluation with 152 cattle samples demonstrated excellent clustering performance and a detection rate ranging from 96.60 to 99.07%, with 94.5% of SNP sites exhibiting polymorphism. The chip achieved 100% gender coverage and displayed a heterozygosity rate between 14.20% and 29.65%, with a repeatability rate of 99.65–99.85%. Analyses using Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) revealed the potential regulatory roles of exonic SNPs in immune response pathways.

Conclusion The development and validation of the 5K GBTS liquid chip for Hainan cattle represent a valuable tool for genome analysis and genetic diversity assessment. Furthermore, it facilitates breed identification, gender determination, and kinship analysis, providing a foundation for the efficient utilization and development of local cattle genetic resources.

Keywords Hainan cattle, Targeted sequencing, Genotyping, SNP chip

Introduction

Cattle, as the most widely distributed species of livestock, are primarily raised for their meat, milk, leather, and labor. China, with its vast territory and diverse climate, is home to a rich array of cattle genetic resources, encompassing over 50 indigenous breeds [1]. Based on their physical characteristics and geographical distribution, cattle can be categorized into three distinct lineages: *Bos taurus*, *Bos indicus*, and their hybrids [2]. This genetic diversity is a crucial component of global biodiversity, providing essential genetic material for the development of new cattle breeds and the enhancement of local

†Huan Wang and Hui Wu contributed equally to the work.

*Correspondence:
Si Chen
chensi.ruth@hotmail.com
Fengyang Wang
fywang68@163.com
[1]Hainan Key Lab of Tropical Animal Reproduction, Breeding and Epidemic Disease Research, College of Tropical Agriculture and Forestry, Hainan University, Haikou, Hainan, PR China
[2]Xinjiang Barkol Kazakh Autonomous County Animal Husbandry Veterinary Station, Barkol Kazakh Autonomous County, Xinjiang, PR China

populations [3]. Hainan cattle are predominantly found in various regions of Hainan province, China. Our previous research indicates that Hainan cattle have evolved from Lei Qiong cattle and have successfully adapted to the climatic and environmental conditions of tropical regions [4]. They are recognized for their exceptional heat resistance, strong foraging ability, tolerance for coarse feed, low susceptibility to disease, and strong resistance to parasites. Additionally, Hainan cattle exhibit characteristics such as small body size and a slow growth rate. Consequently, it is imperative to explore effective strategies for the conservation and utilization of the genetic resources of Hainan cattle.

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation, occurring at single nucleotide positions in the DNA sequence, and are valuable markers in genetic research for applications such as genotype identification, pedigree analysis, genetic diversity assessment, and genetic distance estimation, influencing biological processes like gene expression and disease susceptibility [5]. SNP chips are utilized in various genomic studies [6, 7], including population history inference, structure and admixture analysis, effective population size estimation [8, 9], quantitative trait sites (QTL) mapping [10, 11], genome-wide associationstudies (GWAS) [12, 13], genomic selection (GS) [14–16], exploration of genomic variation [17, 18] and linkage disequilibrium (LD) assessment between breeds [19, 20]. Recently, a novel bison SNP panel was developed, comprising 798 autosomal and 13 mitochondrial SNPs, providing a powerful tool for assessing genetic health, informing management decisions, and enabling individual identification within bison populations [21]. Wu et al. utilized high-density SNP arrays, including the Illumina Porcine SNP50 BeadChip, to identify significant SNPs and genomic regions associated with body composition traits in crossbred Yunong-black pigs; moreover, they enhanced genomic prediction accuracy by incorporating dominance effects [22]. However, traditional solid-phase chips, relying on the complementary hybridization between probes and DNA sequences, are analyzed through the fluorescent signals of labeled probes, presenting drawbacks such as high typing costs and challenges in personalized development. Nowadays, Genotyping by Target Sequencing (GBTS) liquid chips are a cost-effective alternative to whole-genome sequencing (WGS) and genotyping by sequencing, and they are widely used in various livestock species [23, 24]. Compared to traditional methods, GBTS liquid chips provide a comprehensive genetic background, enhanced SNP detection accuracy, and cost-effectiveness, along with flexible design and the ability to incorporate new sites, making large-scale breed genotyping more feasible and affordable [25].

GBTS technology integrates second-generation sequencing to capture specific DNA fragments using designed primers or probes, followed by PCR amplification and comprehensive sequencing [26]. It can be categorized into two main types: GenoPlexs, which utilize multiplex PCR for target amplification, and GenoBaits, which employ liquid-phase probe hybridization for selective DNA capture [27]. The SNP chip based on GBTS technology is commonly referred to as a GBTS liquid chip because SNP capture occurs in a liquid-phase environment during its operation. GBTS liquid chips are versatile, compatible with various sequencing platforms, efficient, and simplify the analysis process. Guan et al. developed a low-cost, high-accuracy whole-genome liquid chip for dairy goats using GBTS technology, suitable for large-scale genotyping in GWAS and GS, which aids molecular breeding and genetic improvement [28]. This chip effectively captures SNP-rich regions and provides multiple SNPs (mSNPs) for haplotype analysis and genotype imputation. Additionally, a high-resolution multiple-SNP capture array was created for sheep, identifying 210 K high-quality SNPs and significant genetic loci linked to hair types and teat number, highlighting the potential of low-cost SNP chips in genomic research and breeding [29].

To date, no liquid chip has been specifically designed for Hainan cattle. This study introduces the development of a 5K low-density liquid-phase SNP chip utilizing GBTS technology, incorporating a significant number of functional SNP loci associated with key traits such as immunity, heat stress, meat quality, and reproduction. The successful creation and validation of this 5K GBTS liquid chip provides a valuable tool for genome analysis and genetic diversity assessment, Additionally, it facilitates breed identification, gender determination, and kinship analysis, establishing a robust foundation for the efficient utilization and development of local cattle genetic resources.

Materials and methods

Data collection and SNP calling

The whole-genome resequencing data were collected from 79 cattle belonging to 13 breeds to obtain high-quality SNP sites for subsequent analysis (Supplementary file: Table S1). The raw reads were filtered using fastp software (version 0.20.0) with the following parameters: -n 10 -q 20 -u 40 [30]. All cleaned reads were aligned to the reference genome of *B.taurus* UMD 3.1.1 (GCF_000003055.6) using BWA-MEM with default parameters [31]. For variant detection, the UnifiedGenotyper tool of GATK software (version v3.5-0-g36282e4) was used with the following parameters: -dcov 1000000 -minIndelFrac 0.15 -glm BOTH -l INFO) [32]. The VariantFiltration tool of GATK software was used for

filtering, with the following parameters: --filterExpression "MQ0>=4 && ((MQ0 / (1.0 * DP))>0.1) "--filterName "HARD_TO_VALIDATE" --filterExpression "DP<5 || QD<2" --filterName "LOW_READ_SUPPORT". A Perl-written program (--maf 0.1 --geno 0.2 --het 0.15) was used to filter the SNP loci based on the following criteria: minor allele frequency (MAF)>0.1, deletion rate<20%, and heterozygous rate<15%.

Selection of specific and functional SNP sites

To identify Hainan cattle-specific SNP loci, the F_{ST} value for each SNP site between Hainan cattle and non-Hainan cattle was calculated using VCFtools-0.1.13 (--vcf snp.vcf --weir-fst-pop hnc.list --weir-fst-pop nohnc.list --out snpfst) [33]. Hainan cattle-specific SNP loci were defined as those exhibiting consistent genotypes within the Hainan cattle population, while showing inconsistent genotypes in the non-Hainan cattle population. Specifically, in non-Hainan cattle, at least 80% of individuals did not match the genotype of the Hainan cattle. We selected SNPs that are evenly distributed across the chromosomes and have high polymorphism (MAF>0.1) as supplementary background loci. Functional SNPs include candidate SNPs of important traits from literature sources and SNPs of immune-related genes from the Bovine Genome Variation Database and Selective Signatures (BGVD) database (http://animal.nwsuaf.edu.cn/code/index.php/BosVar) [34]. PubMed (https://pubmed.ncbi.nlm.nih.gov/) and CNKI (https://www.cnki.net/) were used to search the literatures. NCBI (https://www.ncbi.nlm.nih.gov/) and Ensembl (http://www.ensembl.org/) were used to blast the candidate DNA fragments. The positions of these SNPs were determined by aligning their flanking sequences to the reference genome of *B. taurus* UMD 3.1.1 (GCF_000003055.6). In the BGVD databases, SNPs located within the coding regions of 71 immune-related genes were selected and mapped to their corresponding physical locations on the reference genome. Moreover, the *RBMY* gene and *SRY* gene were selected as candidate genes for sex determination.

Design and synthesis of probe

The probes were designed for the above selected SNP loci to facilitate further detection. The probe design principle involved excluding duplicated SNP loci and designing one or multiple probes that cover the selected SNP loci and their flanking sequence. The probe length is approximately 110 base pairs (bp), with a GC content ranging from 30 to 70%. The number of homologous regions (areas with high sequence similarity) is less than 5. After probe design evaluation, the probes were synthesized and tested.

Loci distribution and function analysis

The distribution of SNP sites was analyzed by counting the number of target sites on different chromosomes. The analysis for uniform distribution of SNP sites involved counting the number of SNPs within a sliding statistical window of 1 Mb on the chromosome. The gap distribution was analyzed across all sections. The analysis principle is as follows: subtract the end position of the previous section from the start position of the subsequent section to obtain the distance between the two sections. If the distance between two segments is greater than the genome size divided by the number of target segments multiplied by 10, it is considered a gap. Followed by visualization of the results using R package ggplot2. For the screening of SNP loci in the exon region, the database for annotation, visualization and integration discovery (DAVID, https://david.ncifcrf.gov/) was used for gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis [35].

Preparation of DNA samples and genotyping by targeted sequencing

The experimental cattle were cared for normally after sampling and were not euthanized or slaughtered. According to the reference [36, 37], cattle were restrained in a squeeze chute in the presence of a veterinarian. After the ear was shaved and disinfected, 2% lidocaine hydrochloride was administered for surface anesthesia. Following this, a 2–3 mm tissue sample was collected from the outer edge of the ear using an ear punch. The ear tissues of cattle were collected from local farms in Ding'an County, Hainan Province, China. Ear tissue collection procedures were specifically referenced from Carrier et al. [38] and Guan et al. [28]. The genomic DNA extraction using the TIANamp Genomic DNA Kit (Tiangen, Beijing, China). The genomic DNA was assessed using 1% (w/v) agarose gel electrophoresis and quantified with the Qubit 2.0 Fluorometer (Invitrogen, Shanghai, China). Genomic DNA libraries were prepared using the GenoBaits DNA-seq Library Prep Kit (MolBreeding Biotechnology Co., Shijiazhuang, Hebei, China) following the manufacturer`s protocol and then hybridized with the biotin-labeled target probe at 65 ℃ for 16 h. Subsequently, Dynabeads MyOne Streptavidin C1 and binding buffer were introduced to capture the hybridized targets while removing non-target fragments. The captured fragments were amplified using library amplification primers and DNA polymerase. Following two rounds of purification using Beckman AMPure Beads, the libraries were quantified with the Qubit 2.0 Fluorometer (Invitrogen, Shanghai, China) and sequenced using PE150 on the MGISEQ-2000 platform (MGI, Shenzhen, China). The SNP site of the chip is composed of three parts. The

design strategy and verification experiments of 5K GBTS liquid chip for Hainan cattle were shown in Fig. 1.

Verification of GBTS liquid chip

We tested 102 qualified cattle DNA samples, including 88 Hainan cattle and 14 Angus cattle (Supplementary file: Table S2), using the liquid chip detection process. To assess the detection rate of the chip, we calculated the ratio of the number of SNP loci detected in each sample to the total number of SNP loci available. To verify SNP consistency, we selected 14 Hainan cattle previously analyzed by whole genome resequencing. We aligned the probe sequences to their genomes to obtain SNP base information and assessed the consistency by cross-referencing with the the liquid chip data. In addition, four bovine DNA samples (CS02_D1F, CS02_X1F, CS02_X7F, and CS02_X8F) were selected for intra- and inter-group replicates to verify the repeatability of the liquid chip. In the case of intra-group replication, the same liquid chip was used for duplicate testing, while inter-group replication was assessed using liquid chips from different batches. Subsequently, PLINK software was used to analyze and process the liquid chip detection data.

To evaluate the capability of this chip for breed identification, we first pooled the liquid chip detection data from 96 DNA samples and the whole genome resequencing data from 56 non-Hainan cattle. Then, the specific base information of 5,292 targeted SNP sites in the liquid chip was collected into a VCF file using BCFtools software from the above two data sets. Perform Principal component analysis (PCA) analysis using GCTA (version 1.92.4), and construct the phylogenetic tree using MEGA-CC (Version: 10.0.5). Subsequently, visualize the results using the R packages ggtree, ggplot2, and scatterplot3d.

Result

Identification of 3,238 specific SNP sites in Hainan cattle

A total of 47,243,114 SNPs were identified from the whole-genome resequencing data of seventy-nine cattle.

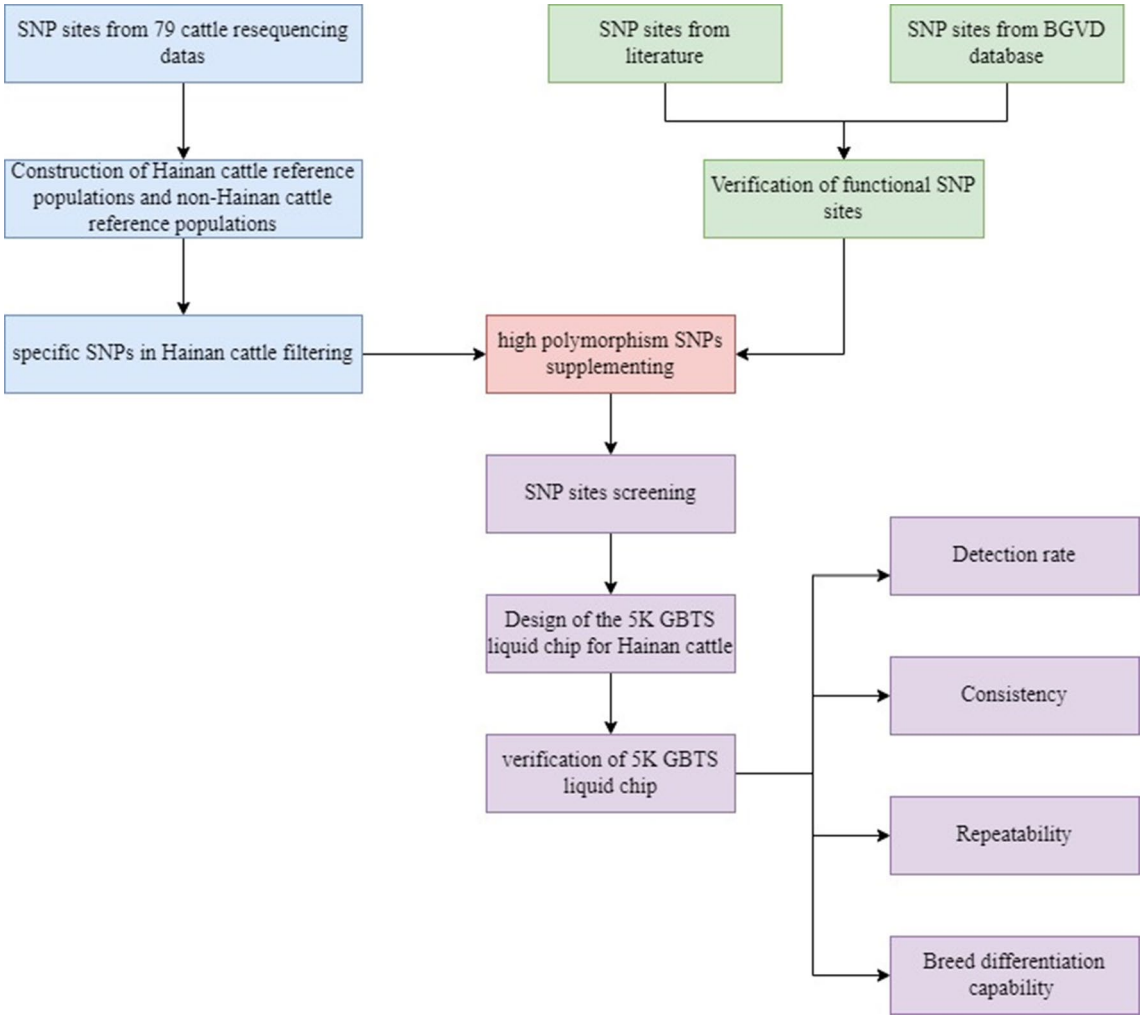

Fig. 1 Roadmap for the development and validation of the 5K liquid phase chip for Hainan cattle

The PCA analysis results indicated that the reference populations for Hainan cattle and non-Hainan cattle, comprising 21 Hainan cattle and 58 non-Hainan cattle as shown in Fig. 2, were successfully established. In Hainan Cattle, 3,238 specific sites was selected, comprising 619 Hainan cattle-specific SNP sites and 2,619 additional background loci. An essential criterion for selecting background loci is the uniform distribution of SNPs across the entire genome.

Selection of 2,939 functional SNP sites from literatures

We conducted searches on PubMed and CNKI and reviewed SNP literature related to important traits including cattle reproduction, meat quality, immune response, and heat stress up until September 1, 2022. After removing duplicate and incomplete sites, the information of SNP sites from 226 literature sources was recorded (Supplementary file: Table S3), which yielded a total of 2,653 SNPs. In addition, a comprehensive analysis of the BGVD database revealed the identification of 429 SNPs within 71 immune-related genes. This process resulted in the identification of 2,939 candidate SNP sites associated with important traits.

Analysis of SNP loci types captured by 5K GBTS liquid chip

A total of 6,177 SNPs sites, encompassing 3,238 specific sites and 2,939 functional sites, were assessed and used for probe design. The uniform distribution of these SNP sites was observed, as depicted in Fig. 3.

According to the principles of probe design, 5,293 sites passed the evaluation, including 518 Hainan cattle-specific SNPs, 1,983 additional highly polymorphic sites, 2,791 functional sites and one sex determination gene fragment (Supplementary file: Table S4). Figure 4A illustrates that functional sites comprised 52.74%, Hainan cattle-specific SNPs accounted for 9.79%, and highly polymorphic supplementary sites represented 37.47%. The coverage of the probe design segment was 85.69%. A total of 9,837 probes, encompassing 5,293 candidate SNP loci and one gene fragment, was utilized to construct the 5K liquid chip for Hainan cattle. The percentages of probes for functional sites, Hainan cattle-specific sites, and highly polymorphic supplementary sites were 54.6%, 9.3%, and 36.1%, respectively. The chip site annotation revealed that the majority of SNPs (74.7% of total core SNP sites) were situated between genes or within introns, with 14.1% located within exons (Fig. 4B).

Functional analysis of 5K GBTS liquid chip for Hainan cattle

Based on the annotation results of SNPs sites on the 5K GBTS liquid chip for Hainan cattle, SNPs annotated in the exon regions were used for the functional analysis. GO enrichment analysis identified 309 terms of three categories, including 229 biological processes (BP), 53 molecular functions (MF) and 27 cellular components (CC). Subsequently, the terms of three categories were merged, and the top 20 significantly enriched terms were shown as the scatter plot (Fig. 5A). The genes with SNPs annotation in the exon region mainly involved in extracellular space (GO: 0005615), inflammatory response (GO: 0006954), extracellular region (GO: 0005576), immune response, extracellular region (GO: 0005576), cytokine activity (GO: 0005125), positive regulation of NF-kappa B transcription factor activity (GO: 0051092), and positive regulation of interleukin-8 production (GO: 0032757). According to KEGG analysis, a total of 63 pathways were significantly enriched and the top 20 significantly enriched pathways were selected (Fig. 5B). It was shown that some genes were involved in immune-related signaling pathways, such as cytokine-cytokine receptor interaction (bta04060), JAK-STAT signaling pathway (bta04630), Toll-like receptor signaling pathway (bta04620), and IL-17 signaling pathway (bta04657). This indicates that the 5K GBTS liquid chip facilitates the discovery and study of SNPs associated with immune traits in Hainan cattle.

Verification of 5K GBTS liquid chip

After read filtering and quality control, the effective rate of the data was 90.62-96.21%, with an average effective rate of 94.02%. The percentage of bases with Phred values greater than 20 and 30 in the total bases were 96.19 −97.73% and 88.58 −92.86%, respectively. The clean reads after quality control were compared with the reference genome, reaching an average comparison rate of 89.33%. These results showed that the sequencing quality

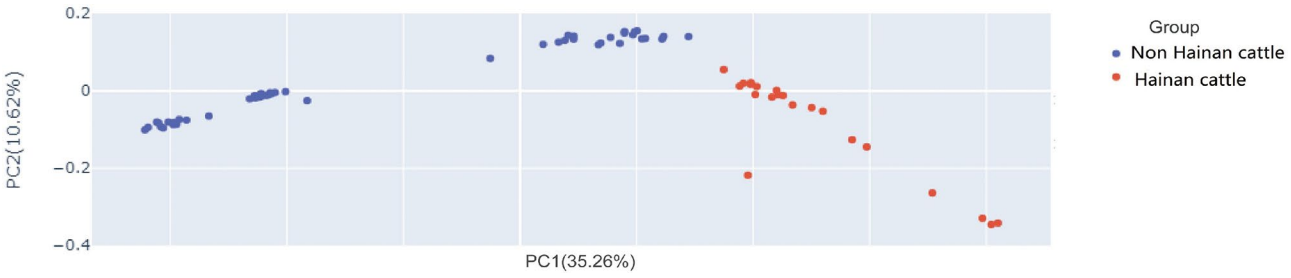

Fig. 2 Principal component analysis of seventy-nine cattle

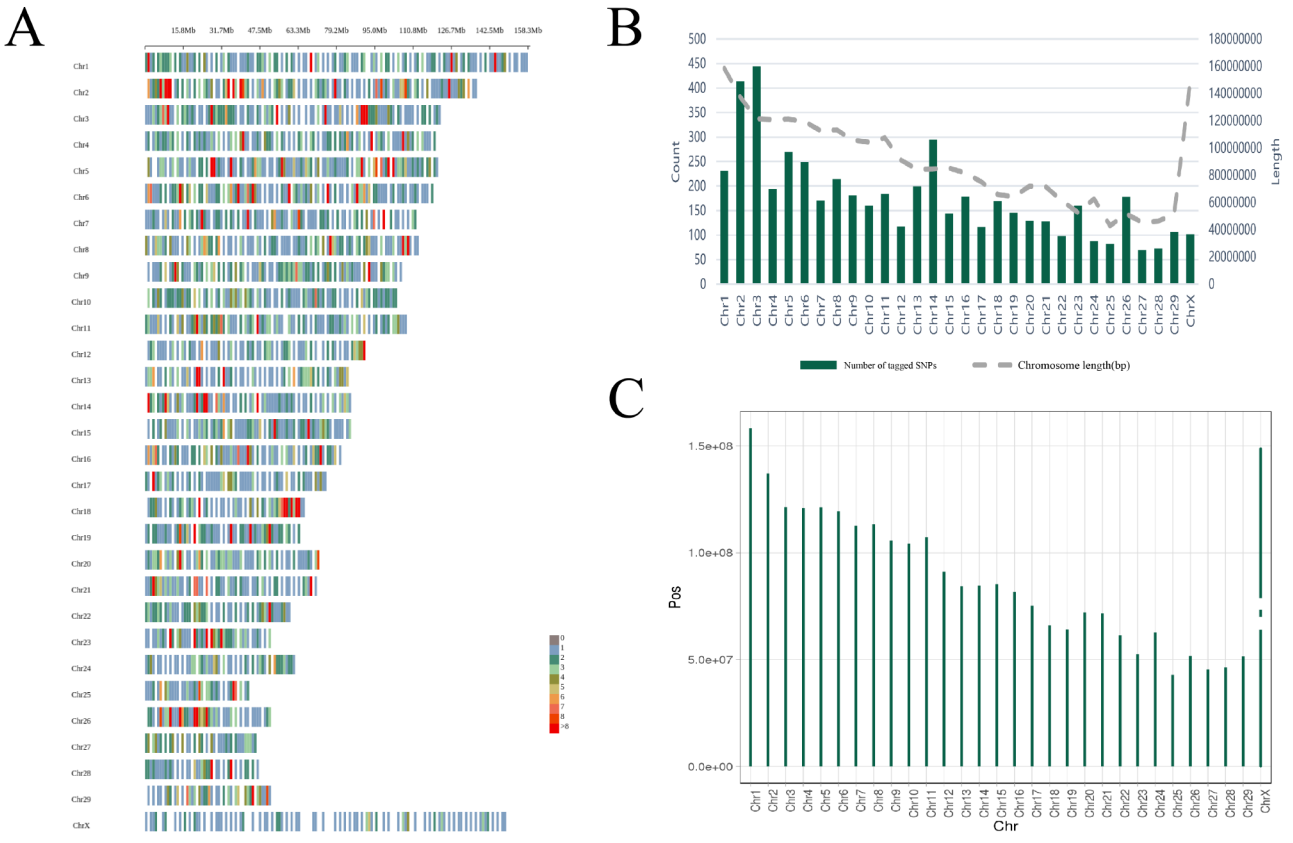

Fig. 3 The distribution of sites of 5K GBTS liquid chip for Hainan cattle. **(A)** The uniform distribution of SNP sites on different chromosomes. **(B)** The distribution of SNP sites on different chromosomes. **(C)** Gap analysis of the target sites on the genome

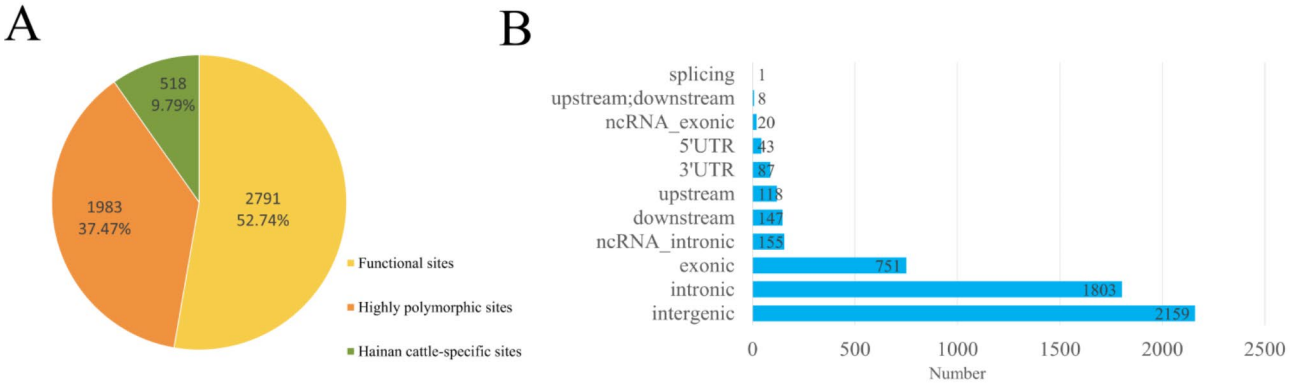

Fig. 4 The site of 5K GBTS liquid chip for Hainan cattle. **(A)** The site information of 5K GBTS liquid chip for Hainan cattle. Yellow dots represent functional sites, green dots represent Hainan cattle-specific sites, and orange dots represent highly polymorphic sites. **(B)** The annotation results of core sites on the 5K GBTS liquid chip

was reliable and met the requirements for subsequent analysis.

The 5K liquid chip exhibited a high SNP detection rate for Hainan cattle, ranging from 96.60 to 99.07% (Fig. 6A), with an average of 98.73%, thereby meeting the required standards (Supplementary file: Table S5). In two rounds of genotyping, 94.5% of the SNPs were found to be polymorphic, and the heterozygosity rate was 14.20-29.65%. The distribution map of MAF sites revealed that the highest concentration of SNP sites fell within the 0.1 to 0.2 range, while the lowest was observed between 0.4 and 0.5. Most SNP sites had an MAF greater than 0.01, and 90.80% of the SNP sites met the established requirements (Fig. 6B). However, 487 SNP sites exhibited MAF values below 0.01, indicating the need to expand the sample size

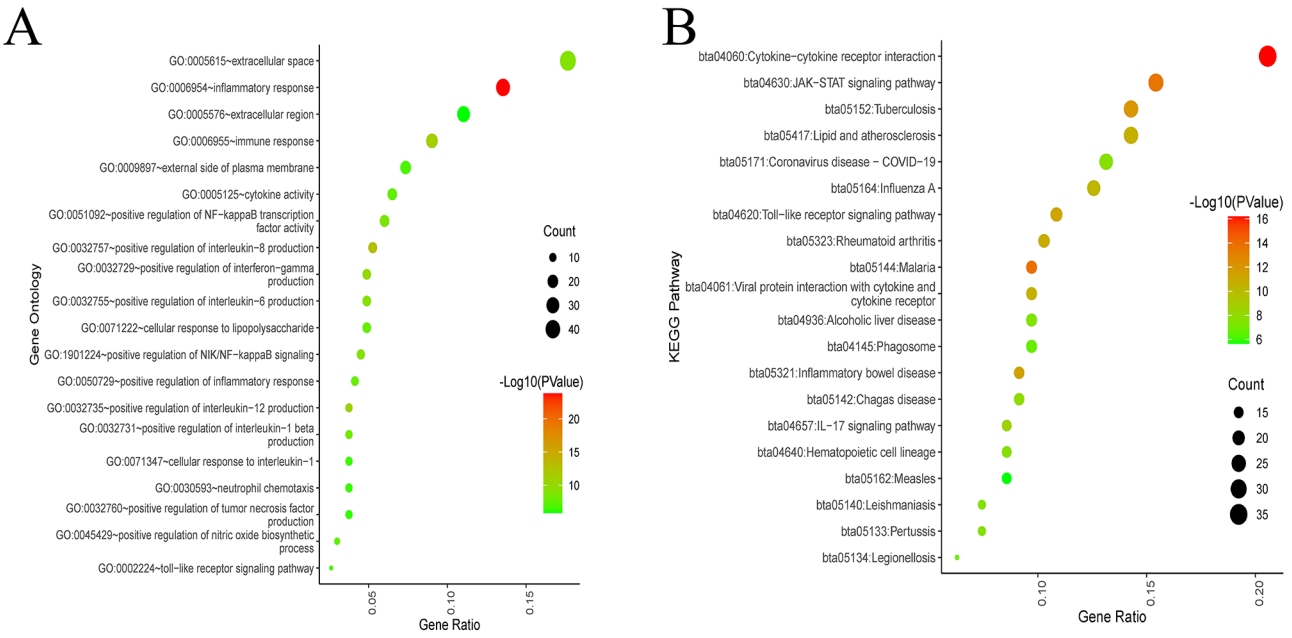

Fig. 5 Analysis and KEGG analysis of genes with SNPs annotation in the exon region on the 5K GBTS liquid chip for Hainan cattle. **(A)** Scatter plot based on the top 20 significantly enriched term of GO analysis. **(B)** Scatter plot based on the top 20 significantly enriched pathways of KEGG analysis

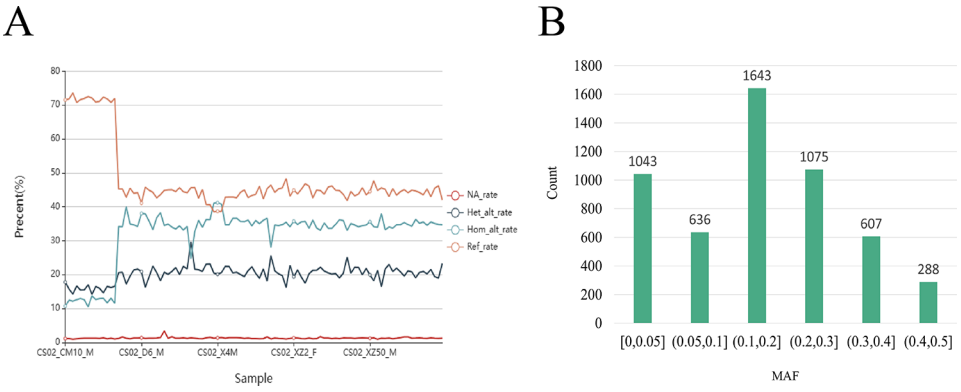

Fig. 6 Statistics of SNP sites on the 5K liquid chip for Hainan cattle and MAF statistics of SNP sites on the 5K liquid chip for Hainan cattle detection results. **(A)** Statistical results of SNP sites on 5K liquid chip Hainan cattle. Red represents the deletion rate, black represents the ratio of the number of heterozygous sites to the total number of non deletion sites, blue represents the ratio of the number of homozygous sites to the total number of non deletion sites, and orange represents the ratio of the number of reference consensus sites to the total number of non deletion sites. **(B)** MAF statistics of SNP sites in all Hainan cattle detected by 5K liquid chip

and adjust the SNP sites on the 5K liquid chip for Hainan cattle.

Consistency and repeatability analysis of the chip

In order to verify the consistency of the genotyping results of the liquid chip, we used 14 DNA samples for genotype detection by the liquid chip. Then, the genotyping results from liquid chip were compared with those from resequencing data. The consistency rate was between 83.36% and 87.43%. And the average consistency rate was 85.59% (Supplementary file: Table S6). To eliminate the contingency of liquid chip detection, the repetition between groups was detected with different batches of liquid chips, showing that the repeatability rate was 99.60% and 99.82% (Supplementary file: Table S6). The repeatability rate of the same samples in the same batch group was between 99.65% and 99.85% (Table 1). The average repeatability rate was 99.75%, indicating that the repeatability was good, and the liquid chip for Hainan cattle met the detection requirements.

Validation of breed and gender identification

To assess the breed differentiation capability of the 5K liquid chip for Hainan cattle, we conducted cluster analysis on 96 samples (82 Hainan cattle and 14 Angus cattle) along with resequencing data from 56 non-Hainan

Table 1 Statistical table of genotype concordance rate in duplicate samples

Sample ID	Number of discordant SNPs	Number of concordant SNPs	Concordance rate
CS02_D1F_01&CS02_D1F_02	20	5678	99.65%
CS02_X8F_01&CS02_X8F_02	19	6165	99.68%
CS02_X7_F_01&CS02_X7_F_02	12	5936	99.80%
CS02_X1F_01&CS02_X1F_02	9	5945	99.85%

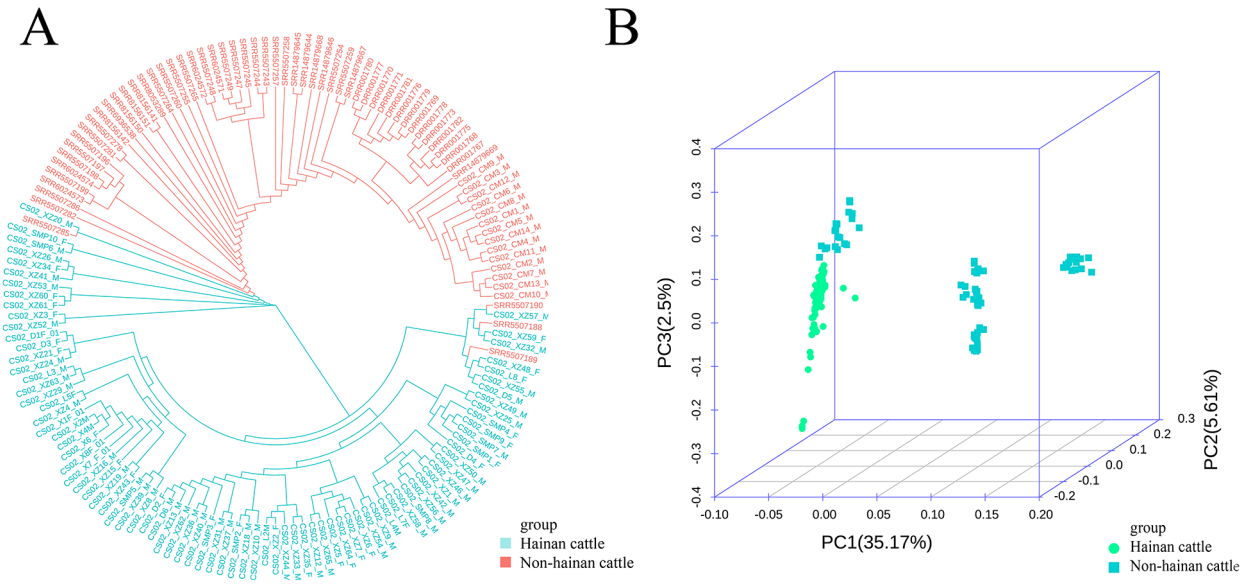

Fig. 7 Phylogenetic tree and PCA analysis results of the 5K liquid chip for Hainan cattle detection results. **(A)** Phylogenetic tree of the chip assay results for 152 cattle DNA samples. Cyan represents Hainan cattle, and red represents Non-Hainan cattle. **(B)** PCA analysis of chip assay results for 152 cattle DNA samples. Green dots represent Hainan cattle, and blue dots represent Non-Hainan cattle

cattle. Phylogenetic tree and PCA results indicated distinct clustering of Hainan cattle with other breeds, effectively demonstrating the clustering function (Fig. 7). Notably, Hainan cattle from different regions did not form distinct clusters but were intermixed, reflecting the high genetic diversity of the breed. To evaluate the sex identification capability of the 5K liquid chip for Hainan cattle, we tested the *RBMY* and *SRY* gene regions. The results indicated that the Hainan cattle genome matched only the *RBMY* gene region, not the *SRY* region. Gender identification for 102 samples achieved 100% accuracy (Supplementary file: Table S7), confirming that the liquid chip effectively identifies gender in Hainan cattle.

Discussion

Our 5K liquid chip is a custom SNP chip designed for Hainan cattle. In its design, we prioritized SNPs specific to Hainan cattle and included previously reported loci associated with key economic traits. This strategy enhances the accuracy of subsequent GWAS and GS studies. Moreover, a second important criterion for our liquid chip design is the uniform distribution of the SNPs across the genome, as this greatly facilitates finding associations between markers and phenotypes [39].

The SNP sites selected for the 5K liquid chip designed for the Hainan cattle were divided into three parts. The first part included 2,654 SNPs sourced from the literature, selected based on two criteria: (1) studies providing direct evidence of associations between SNPs in specific genes and important traits in cattle; and (2) GWAS studies identifying significant correlations between SNP markers and economically relevant traits. We focused on SNPs associated with key traits such as meat quality, reproduction, heat stress tolerance, disease resistance, and immunity, as these may influence genomic selective breeding. GWAS are crucial for understanding the genetic basis of complex traits and diseases through genotype-phenotype associations [40]. They can identify strong signals of pathogenic variation and correlate various economic traits in cattle [41]. For instance, Liu et al. [42] demonstrated that using selected WGS variants as loci for a 54 K SNP chip improved the prediction reliability of milk production traits in Danish Jersey cattle. Similarly, Liu et al. [3] developed a 55K SNP genotyping array for chicken, selecting candidate SNPs based on GWAS of 15 economically relevant traits. Additionally, during the design of a 34 K SNP genotyping array for *Populus trichocarpa* [43], candidate genes were primarily identified through comprehensive literature and database searches.

The second part comprised 429 SNPs sourced from the BGVD database. Among the comprehensive databases that contain SNP information, AnimalQTLdb [44] includes SNPs associated with traits, while dbSNP [45] and EVA [46] encompass all genetic variations across the entire genome. However, it is important to note that these databases cannot directly search for genes, which greatly hinders the utility of these data [47]. Additionally, they rely on published literature, which may not be comprehensive and requires regular updates over time [48]. For target SNP screening, we selected the BGVD database, which includes SNPs with allele frequency data. The availability of allele frequency data will greatly facilitate population genetic research and aid in the design of molecular markers for cattle breeding projects [49]. According to our previous research results, we found that most of the selected SNP sites in the Hainan cattle population were clustered in immune-related signaling pathways, suggesting that the Hainan cattle population has strong genetic characteristics of disease resistance [4]. Within the BGVD, our study focused on a targeted search for SNP sites within several immunity genes, namely *CCL19*, *CD163*, *CXCR4*, *IL10*, *IL18*, *IL1RN*, *IL2RA*, *IL6*, *IRF7*, *NLRP3*, *SOS1*, *TLR2*, *TLR4*, *VPS13D*, and other relevant genes. The identified SNP sites proved to be valuable for the subsequent exploration and identification of disease-resistant genes in Hainan cattle.

The third part involved selecting 3,238 SNPs from the whole-genome resequencing data of Hainan cattle. This included 2,619 SNPs exhibiting high polymorphism across all cattle and 619 SNPs specific to Hainan cattle. During the screening process, sites with high frequency distribution in the population were prioritized. This approach ensures that low-frequency sites, which may be difficult to detect in subsequent analyses, are avoided, as their limited impact can undermine the reliability of GWAS results [29]. Using the F_{ST} value allows for more accurate screening of unique SNP sites in Hainan cattle, facilitating a better identification of the breed. Fan et al. [5] measured the F_{ST} values and heterozygosity of all SNP sites in the reference populations of sika deer and red deer. They identified 1,000 SNP sites with high F_{ST} values(F_{ST}>0.95) to create a 1 K sika deer SNP chip.

To explore the functional characteristics of SNPs, we conducted GO and KEGG enrichment analyses on genes associated with SNPs in the exonic regions, based on the 5K liquid chip annotation of Hainan cattle. This analysis identified enriched terms that could serve as potential targets for enhancing disease resistance. Among the top 20 significant GO terms, over half were related to immune response and inflammation, including key terms like inflammatory response and neutrophil chemotaxis, particularly relevant to mastitis [50]. In the top 20 enriched KEGG pathways, genes linked to SNPs were also involved in immune response and disease-related signaling. Notably, the most prominent pathway identified was bta04060: Cytokine-cytokine receptor interaction. A study [50] found that differentially expressed genes (DEGs) during Bovine viral diarrhea virus (BVDV-1) infection were significantly enriched in this pathway, with seven DEGs (*TGFB2*, *CCL20*, *NGFR*, *CXCR1*, *CXCR2*, *IL1R2*, and *CSF3*) showing downregulated expression associated with immune processes and cell death [51]. These findings indicate that the chip contains a substantial number of SNPs related to immune response and inflammation, highlighting its significance for research on immune traits and for selecting disease-resistant breeding strategies in Hainan cattle.

In the application of chips, we found that the chip has a high detection rate due to the ingenious design and strict filtering of the probe. The detection rate was between 96.60% and 99.07%. The repeatability rate was between 99.65% and 99.85%. And the consistency rate between liquid chip genotyping results and resequencing genotyping results ranged from 83.36 to 87.43%. One possible explanation we considered for the discrepancy in the consistency of the genotyping results between the two methods is the difference in sequencing depth. For instance, the verification results of the 200 K SNP array developed by Kang Wei et al. [52] demonstrated the consistency rate of SNP genotyping between SNP array and resequencing data ranged from 64.14 to 91.93%, with an average of 84.07%. To analyze the reasons for this phenomenon, they randomly selected 18 inconsistent SNPs and performed sanger sequencing. The results showed that neither resequencing nor SNP array could guarantee 100% correct results [52]. Torkamaneh et al. [53] found that SNP catalogues from different pipelines or sequencing technologies may be inconsistent. Therefore, the subsequent mutual verification by different methods is very important. Compared to whole genome resequencing technology, liquid chip genotyping offers higher accuracy and facilitates the efficient acquisition of genotyping information for targeted SNP sites across a large number of individuals.

Furthermore, Hainan cattle, and other non-Hainan cattle were selected to verify the clustering ability of the chip. The results showed that 94.5% of the SNP sites were polymorphic and the heterozygosity rate was between 14.20% and 29.65%. It indicated that the 5K liquid chip can be used to determine the genetic variation of Hainan cattle breeds. A multi-species SNP array designed by Mastrochirico-Filho et al. [54] was used to genotype 94 and 58 samples of *Piaractus mesopotamicus* (pacu) and *Colossoma macropomum* (tambaqui), respectively. The analysis revealed that 74.17% and 71.25% of SNPs were identified as polymorphic in pacu and tambaqui, respectively. Iamartino et al. [55] found that 70% of the SNPs

were high quality and polymorphic in the design and verification of 90 K SNP genotyping in *Bubalus bubalis* (buffalo). The phylogenetic tree was constructed using the genetic distance between individuals belonging to the analyzed population. This method is often used for genetic diversity analysis and parental line selection [56]. The results of the phylogenetic tree (Fig. 7A) and PCA analysis (Fig. 7B) showed that Hainan cattle and non-Hainan cattle were clustered in different positions, which basically realized the distinguishing function. This is consistent with the result that the phylogenetic tree of 152 cattle showed that Hainan cattle and Leiqiong cattle were closely related.

Currently, research on bovine sex identification chips has been conducted, including the work by Neumann et al. [23]. In the design of the bovine 200 K SNP chip developed for the endangered German Black Spotted Cattle (DSN), specific SNPs located on the X and Y chromosomes are incorporated within the chip to enable sex verification. However, studies have shown that the morphological structure of the Y chromosome of *Bos taurus* and *Bos indicus* is different. Mukherjee et al. [57] detected the gene copy number variation of four spermatogenesis-related genes on the Y chromosome of hybrid cattle and zebu, and found that *SRY* copy number and *TSPY* (Testis-specific protein Y-encoded) copy number were significantly higher in hybrid cattle than in zebu. Therefore, we designed probes for the *SRY* gene and *RBMY* gene unique to the Y chromosome, and identified the sex of the cattle through the test results. The sex determination region Y (*SRY*) is an intronless gene located on the Y chromosome with only one exon [58]. *RBMY* is a repetitive gene located in the azoospermia factor (*AZF*) region on the long arm of the Y chromosome [59]. Under normal conditions, RBMY is mainly expressed in male germ cells [60]. The results of this study showed that the Hainan cattle genome could only match the *RBMY* gene region, but could not match the *SRY* gene region. It is suggested that the sex identification genes of Hainan cattle belonging to zebu cattle cannot be completely applied to the sex identification genes of ordinary cattle. The design of the *RBMY* gene for sex identification of Hainan cattle not only increases the accuracy of sex identification but also has more advantages than other SNP chips.

Six different cattle SNP chips, ranging from low to high density, are commercialized on two different genotyping platforms [61]. Among them, five SNP chips were sequenced using the Illumina sequencing platform (San Diego, CA), including the Golden Gate Bovine3K BeadChip (Bovine3k), Infinium BovineLD BeadChip (BovineLD), Infinium BovineSNP50 v.1 BeadChip (BovineSNP50v.1), Infinium BovineSNP50 v.2 BeadChip (BovineSNP50v.2), and Infinium BovineHD BeadChip (BovineHD). They contain 2,900, 6,909, 54,001, 54,609, and 777,962 SNPs, respectively. Affymetrix's new product, the Axiom Genome-Wide BOS 1 Array Plate, enables genotyping of over 640,000 SNPs from a diverse range of *Bos taurus* and *Bos indicus* breeds, providing comprehensive genetic coverage for Holstein, Angus, and Nelore. However, these SNP chips are tailored for commercial breeds and may not encompass the specific SNP information relevant to local cattle. Given that several studies indicate a certain degree of *B. indicus* lineage in Hainan cattle [62–64], it becomes crucial to develop an appropriate SNP chip for this valuable genetic resource in tropical regions, ensuring accurate genetic diversity research and evaluation of the Hainan cattle breed. When comparing genotyping costs, the GBTS liquid chip technology significantly reduces expenses compared to traditional chip-based genotyping for the same markers and samples [65, 66].

Conclusion

The development of the 5K low-density liquid chip for Hainan cattle represents a promising advancement in genomic analysis and genetic diversity assessment. By integrating various functional SNP loci associated with important traits like immunity and meat quality, the chip aims to enhance our understanding and support targeted breeding strategies. Its high detection rates and strong performance suggest potential reliability for applications in breed identification, gender determination, and kinship analysis. Continued validation will be important to fully explore its contributions to the improvement and sustainability of Hainan cattle and related breeds.

Abbreviations

GBTS	Genotyping by Target Sequencingsingle
SNP	Single nucleotide polymorphism
WGS	Whole-genome sequencing
MAF	Minor allele frequency
F_{ST}	Fixation index
BGVD	Bovine Genome Variation Database
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
PCA	Principal component analysis
BP	Biological processe
MF	Molecular function
CC	Cycellular component
GWAS	Genome-wide association analysis
GS	Genome selection analysis
QTL	Quantitative trait sites
LD	Linkage disequilibrium

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10753-w.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions
HW wrote and typeset the manuscript. HW carried out the experiments and analyzed the data. WZ contributed to data collection. JJ and QH revised the manuscript. CM and HG managed the experiments. QC, and LD performed and supervised the research. SC reviewed the data and authored the manuscript. FW designed the study and revised the manuscript. All authors read and approved the final manuscript.

Funding
This work was financially supported by Hainan Innovation Center for Academician of Zhang Yong.

Data availability
The datasets generated and/or analysed during the current study are available in the NCBI SRA repository, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA774120. All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
All samples were collected in accordance with the Regulations for the Administration of Affairs Concerning Experimental Animals (Ministry of Science and Technology, China, 2004) and were approved by the owner's informed consent and the Academic Committee of the College of Animal Science and Technology of Hainan University, in accordance with the regulations on the use of experimental animals and institutional safety procedures.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 6 December 2023 / Accepted: 2 September 2024
Published online: 18 September 2024

References

1. 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.
2. 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.
3. Liu R, Xing S, Wang J, Zheng M, Cui H, Crooijmans R, Li Q, Zhao G, Wen J. A new chicken 55K SNP genotyping array. BMC Genomics. 2019;20(1):410.
4. Chen S, Chang C, Cui K, Yang W, Li B, Ni S, Zhang W, Li S, Li X, Wu G, et al. Whole-genome analyses reveal the genomic diversity and selection signatures of Hainan cattle. Livest Sci. 2023;275:105311.
5. Fan H, Wang T, Li Y, Liu H, Dong Y, Zhang R, Wang H, Shang L, Xing X. Development and validation of a 1 K sika deer (Cervus nippon) SNP chip. BMC Genomic data. 2021;22(1):35.
6. Hu M, Jiang H, Lai W, Shi L, Yi W, Sun H, Chen C, Yuan B, Yan S, Zhang J. Assessing Genomic Diversity and Signatures of Selection in Chinese Red Steppe Cattle Using High-Density SNP Array. Animals 2023;13(10).
7. Perini F, Cendron F, Wu Z, Sevane N, Li Z, Huang C, Smith J, Lasagna E, Cassandro M, Penasa M. Genomics of Dwarfism in Italian Local Chicken breeds. Genes 2023;14(3).
8. Sudrajad P, Kusminanto RY, Volkandari SD, Cahyadi M. Genomic structure of Bali cattle based on linkage disequilibrium and effective population size analyses using 50K single nucleotide polymorphisms data. Veterinary World. 2022;15(2):449–54.
9. Tolone M, Sardina MT, Criscione A, Lasagna E, Senczuk G, Rizzuto I, Riggio S, Moscarelli A, Macaluso V, Di Gerlando R, et al. High-density single nucleotide polymorphism markers reveal the population structure of 2 local chicken genetic resources. Poult Sci. 2023;102(7):102692.
10. Palti Y, Vallejo RL, Purcell MK, Gao G, Shewbridge KL, Long RL, Setzke C, Fragomeni BO, Cheng H, Martin KE, et al. Genome-wide association analysis of the resistance to infectious hematopoietic necrosis virus in two rainbow trout aquaculture lines confirms oligogenic architecture with several moderate effect quantitative trait loci. Front Genet. 2024;15:1394656.
11. Ma C, Liu L, Liu T, Jia Y, Jiang Q, Bai H, Ma S, Li S, Wang Z. QTL mapping for important agronomic traits using a Wheat55K SNP array-based genetic map in tetraploid wheat. Plants 2023;12(4).
12. More M, Veli E, Cruz A, Gutiérrez JP, Gutiérrez G. Ponce de León FA: Genome-Wide Association Study of Fiber Diameter in Alpacas. *Animals* 2023, 13(21).
13. Heidaritabar M, Bink MCAM, Dervishi E, Charagu P, Huisman A, Plastow GS. Genome-wide association studies for additive and dominance effects for body composition traits in commercial crossbred Piétrain pigs. J Anim Breed Genet. 2023;140(4):413–30.
14. Garcia A, Tsuruta S, Gao G, Palti Y, Lourenco D, Leeds T. Genomic selection models substantially improve the accuracy of genetic merit predictions for fillet yield and body weight in rainbow trout using a multi-trait model and multi-generation progeny testing. Genet Selection Evol. 2023;55(1):11.
15. Mastrangelo S, Ben-Jemaa S, Perini F, Cendron F, Biscarini F, Lasagna E, Penasa M, Cassandro M. Genome-wide mapping of signatures of selection using a high-density array identified candidate genes for growth traits and local adaptation in chickens. Genet Selection Evol. 2023;55(1):20.
16. Guo X, Puritz JB, Wang Z, Proestou D, Allen S, Small J, Verbyla K, Zhao H, Haggard J, Chriss N, et al. Development and evaluation of high-density SNP arrays for the Eastern Oyster Crassostrea virginica. Mar Biotechnol. 2023;25(1):174–91.
17. Linhuan H, Danlei C, Zhiming H, Shu K, Jiayi C, Jiayi P, Chuqi S, Yinghong Y, Ding W, Yingjun X, et al. The use of high-resolution SNP arrays to detect congenital cardiac defects. J maternal-fetal Neonatal Medicine: Official J Eur Association Perinat Med Federation Asia Ocean Perinat Soc Int Socie. 2024;37(1):2301831.
18. Kumar H, Panigrahi M, Saravanan KA, Rajawat D, Parida S, Bhushan B, Gaur GK, Dutt T, Mishra BP, Singh RK. Genome-wide detection of copy number variations in Tharparkar cattle. Animal Biotechnol. 2021;34(2):448–55.
19. Saravanan KA, Panigrahi M, Kumar H, Parida S, Bhushan B, Gaur GK, Kumar P, Dutt T, Mishra BP, Singh RK. Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in Tharparkar cattle breed of India. Animal Biotechnol. 2020;33(2):297–311.
20. Alipanah M, Mazloom SM, Gharari F. Detection of selective sweep in European wild sheep breeds. 3 Biotech. 2024;14(4):122.
21. Stroupe S, Derr JN. Development and evaluation of a novel single nucleotide polymorphism panel for north American bison. Evol Appl. 2024;17(2):e13658.
22. Wu Z, Dou T, Bai L, Han J, Yang F, Wang K, Han X, Qiao R, Li XL, Li XJ. Genomic prediction and genome-wide association studies for additive and dominance effects for body composition traits using 50 K and imputed high-density SNP genotypes in Yunong-black pigs. J Anim Breed Genet. 2023;141(2):124–37.
23. Neumann GB, Korkuc P, Arends D, Wolf MJ, May K, Reissmann M, Elzaki S, Konig S, Brockmann GA. Design and performance of a bovine 200 k SNP chip developed for endangered German black Pied cattle (DSN). BMC Genomics. 2021;22(1):905.
24. Suratannon N, van Wijck RTA, Broer L, Xue L, van Meurs JBJ, Barendregt BH, van der Burg M, Dik WA, Chatchatee P, Langerak AW, et al. Rapid low-cost microarray-based genotyping for genetic screening in primary immunodeficiency. Front Immunol. 2020;11:614.
25. Balog K, Mizeranschi AE, Wanjala G, Sipos B, Kusza S, Bagi Z. Application potential of chicken DNA chip in domestic pigeon species – preliminary results. Saudi J Biol Sci. 2023;30(3):103594.
26. Samorodnitsky E, Datta J, Jewell BM, Hagopian R, Miya J, Wing MR, Damodaran S, Lippus JM, Reeser JW, Bhatt D, et al. Comparison of custom capture for targeted next-generation DNA sequencing. J Mol Diagnostics: JMD. 2015;17(1):64–75.
27. Guo Z, Wang H, Tao J, Ren Y, Xu C, Wu K, Zou C, Zhang J, Xu Y. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. Mol Breeding. 2019;39(3):37.
28. 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 2023;14(5).
29. Guo Y, Bai F, Wang J, Fu S, Zhang Y, Liu X, Zhang Z, Shao J, Li R, Wang F et al. Design and characterization of a high-resolution multiple-SNP capture array by target sequencing for sheep. J Anim Sci. 2023;101.
30. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinf (Oxford England). 2018;34(17):i884–90.

31. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In.; 2013: arXiv:1303.3997.
32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
33. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinf (Oxford England). 2011;27(15):2156–8.
34. Chen N, Fu W, Zhao J, Shen J, Chen Q, Zheng Z, Chen H, Sonstegard TS, Lei C, Jiang Y. BGVD: an Integrated Database for Bovine Sequencing Variations and selective signatures. Genom Proteom Bioinform. 2020;18(2):186–93.
35. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.
36. Skarda RT. Local and regional anesthesia in ruminants and swine. Veterinary Clin North Am Food Anim Pract. 1996;12(3):579–626.
37. Greene SA. Protocols for anesthesia of cattle. Veterinary Clin North Am Food Anim Pract. 2003;19(3):679–93. vii.
38. Carrier A, Prunier J, Poisson W, Trottier-Lavoie M, Gilbert I, Cavedon M, Pokharel K, Kantanen J, Musiani M, Côté SD, et al. Design and validation of a 63K genome-wide SNP-genotyping platform for caribou/reindeer (Rangifer tarandus). BMC Genomics. 2022;23(1):687.
39. Groenen MA, Megens HJ, Zare Y, Warren WC, Hillier LW, Crooijmans RP, Vereijken A, Okimoto R, Muir WM, Cheng HH. The development and characterization of a 60K SNP chip for chicken. BMC Genomics. 2011;12(1):274.
40. Bush WS, Moore JH. Chap. 11: genome-wide association studies. PLoS Comput Biol. 2012;8(12):e1002822.
41. Wu X, Guldbrandtsen B, Nielsen US, Lund MS, Sahana G. Association analysis for young stock survival index with imputed whole-genome sequence variants in Nordic Holstein cattle. J Dairy Sci. 2017;100(8):6356–70.
42. Liu A, Lund MS, Boichard D, Karaman E, Fritz S, Aamand GP, Nielsen US, Wang Y, Su G. Improvement of genomic prediction by integrating additional single nucleotide polymorphisms selected from imputed whole genome sequencing data. Heredity. 2020;124(1):37–49.
43. Geraldes A, Difazio SP, Slavov GT, Ranjan P, Muchero W, Hannemann J, Gunter LE, Wymore AM, Grassa CJ, Farzaneh N, et al. A 34K SNP genotyping array for Populus trichocarpa: design, application to the study of natural populations and transferability to other Populus species. Mol Ecol Resour. 2013;13(2):306–23.
44. 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.
45. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res. 1999;9(8):677–9.
46. Cezard T, Cunningham F, Hunt SE, Koylass B, Kumar N, Saunders G, Shen A, Silva AF, Tsukanov K, Venkataraman S, et al. The European variation archive: a FAIR resource of genomic variation for all species. Nucleic Acids Res. 2022;50(D1):D1216–20.
47. Elsik CG, Unni DR, Diesh CM, Tayal A, Emery ML, Nguyen HN, Hagen DE. Bovine genome database: new tools for gleaning function from the Bos taurus genome. Nucleic Acids Res. 2016;44(D1):D834–839.
48. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, Bennett R, Bhai J, Billis K, Boddu S, et al. Ensembl 2019. Nucleic Acids Res. 2019;47(D1):D745–51.
49. Chen SH, Lo CZ, Su SY, Kuo BH, Hsiung CA, Lin CY. UPS 2.0: unique probe selector for probe design and oligonucleotide microarrays at the pangenomic/genomic level. BMC Genomics. 2010;11(Suppl 4):S6.
50. Naserkheil M, Ghafouri F, Zakizadeh S, Pirany N, Manzari Z, Ghorbani S, Banabazi MH, Bakhtiarizadeh MR, Huq MA, Park MN, et al. Multi-omics Integration and Network Analysis Reveal Potential Hub Genes and genetic mechanisms regulating bovine Mastitis. Curr Issues Mol Biol. 2022;44(1):309–28.
51. Ma Y, Wang L, Jiang X, Yao X, Huang X, Zhou K, Yang Y, Wang Y, Sun X, Guan X, et al. Integrative Transcriptomics and proteomics analysis provide a deep insight into bovine viral diarrhea virus-host interactions during BVDV infection. Front Immunol. 2022;13:862828.
52. Wei K, Wang X, Hao X, Qian Y, Li X, Xu L, Ruan L, Wang Y, Zhang Y, Bai P, et al. Development of a genome-wide 200K SNP array and its application for high-density genetic mapping and origin analysis of Camellia sinensis. Plant Biotechnol J. 2022;20(3):414–6.
53. Torkamaneh D, Laroche J, Belzile F. Genome-wide SNP calling from genotyping by sequencing (GBS) data: a comparison of seven pipelines and two sequencing technologies. PLoS ONE. 2016;11(8):e0161333.
54. Mastrochirico-Filho VA, Ariede RB, Freitas MV, Borges CHS, Lira LVG, Mendes NJ, Agudelo JFG, Caceres P, Berrocal MHM, Sucerquia GAL, et al. Development of a multi-species SNP array for serrasalmid fish Colossoma macropomum and Piaractus mesopotamicus. Sci Rep. 2021;11(1):19289.
55. Iamartino D, Nicolazzi EL, Van Tassell CP, Reecy JM, Fritz-Waters ER, Koltes JE, Biffani S, Sonstegard TS, Schroeder SG, Ajmone-Marsan P, et al. Design and validation of a 90K SNP genotyping assay for the water buffalo (Bubalus bubalis). PLoS ONE. 2017;12(10):e0185220.
56. Chen H, Xie W, He H, Yu H, Chen W, Li J, Yu R, Yao Y, Zhang W, He Y, et al. A high-density SNP genotyping array for rice biology and molecular breeding. Mol Plant. 2014;7(3):541–53.
57. Mukherjee A, Dass G, G JM, Gohain M, Brahma B, Datta TK, De S. Absolute copy number differences of Y chromosomal genes between crossbred (Bos taurus x Bos indicus) and Indicine bulls. J Anim Sci Biotechnol. 2013;4(1):15.
58. Lu W, Rawlings N, Zhao J, Wang H. Amplification and application of the HMG box of bovine SRY gene for sex determination. Anim Reprod Sci. 2007;100(1–2):186–91.
59. Tsuei DJ, Lee PH, Peng HY, Lu HL, Su DS, Jeng YM, Hsu HC, Hsu SH, Wu JF, Ni YH, et al. Male germ cell-specific RNA binding protein RBMY: a new oncogene explaining male predominance in liver cancer. PLoS ONE. 2011;6(11):e26948.
60. Kido T, Tabatabai ZL, Chen X, Lau YC. Potential dual functional roles of the Y-linked RBMY in hepatocarcinogenesis. Cancer Sci. 2020;111(8):2987–99.
61. Nicolazzi EL, Picciolini M, Strozzi F, Schnabel RD, Lawley C, Pirani A, Brew F, Stella A. SNPchiMp: a database to disentangle the SNPchip jungle in bovine livestock. BMC Genomics. 2014;15:123.
62. Dong Z, Ning Q, Liu Y, Wang S, Wang F, Luo X, Chen N, Lei C. Comparative transcriptomics analysis of testicular miRNA from indicine and taurine cattle. Animal Biotechnol 2022:1–11.
63. Liu Y, Xu L, Yang L, Zhao G, Li J, Liu D, Li Y. Discovery of Genomic Characteristics and Selection Signatures in Southern Chinese local cattle. Front Genet. 2020;11:533052.
64. Nie L, Yu Y, Zhang XQ, Yang GF, Wen JK, Zhang YP. Genetic diversity of cattle in south China as revealed by blood protein electrophoresis. Biochem Genet. 1999;37(7–8):257–65.
65. Puiggalí-Jou A, Pérez-Madrigal MM, Del Valle LJ, Armelin E, Casas MT, Michaux C, Perpète EA, Estrany F, Alemán C. Confinement of a β-barrel protein in nanoperforated free-standing nanomembranes for ion transport. Microbiol Resource Announcements. 2016;8(38):16922–35.
66. Senapati S, Slouka Z, Shah SS, Behura SK, Shi Z, Stack MS, Severson DW, Chang HC. An ion-exchange nanomembrane sensor for detection of nucleic acids using a surface charge inversion phenomenon. Biosens Bioelectron. 2014;60:92–100.

Publisher's note