Genome-Wide Assessment of Signatures of Selection in the Pakistan Sahiwal Cattle

Abdul Rahman Sesay1,2, Muhammad Saif-ur-Rehman1*, Faisal Ramzan1 and Faisal Saeed Awan3

1Institute of Animal and Dairy Sciences, Faculty of Animal Husbandry, University of Agriculture, Faisalabad, Pakistan
2Department of Animal Science, School of Agriculture, Njala University, Freetown, Sierra Leone
3Centre of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad, Pakistan

ABSTRACT

The Sahiwal breed of dairy cattle holds significant importance in Pakistan, mostly attributed to its ability to withstand high temperatures, resilience to diseases, and satisfactory performance when fed low-quality roughages. Domestication and breeding of mammals have exerted a consistent selection pressure on a wide range of characteristics in many domesticated species, resulting in discernible genetic modifications at the individual genome level. The study aimed to discover and analyze potential indicators of recent selection in Sahiwal cattle, specifically identifying the genes and quantitative trait loci associated with these selection indicators. The study utilized a sample size of 98 Sahiwal bulls. The genotyping of all animals was conducted using the BovineHD140k BeadChip. After undergoing quality control measures, 87 samples as well as 74,070 SNPs located throughout 29 autosomes were selected as well as included in the study. The selection signatures were examined using the iHS and the Tajima D approach. The result reveals the current positive selections on BTA 1, 2, 6, 11, 12, 15, 17, 21, and 22 with the iHS test, while for the Tajima D test, the current positive selections were detected on BTA 1, 2, 3, 4, 5, 6, 7, 9, 10, 14, 18, 20, 23, and 24. A total of 47 genes were detected within selection regions associated with vital economic traits. The QTL enrichment analysis has shown eight substantial QTLs in BTA19 and BTA20 linked with milk, production, as well as reproduction traits. Therefore, understanding the selection signatures and candidate genes that influence important economic traits can provide foundational knowledge that can be used effectively to gain insight into the underlying mechanisms controlling these traits in Sahiwal cattle.

INTRODUCTION

The Sahiwal cattle breed originated in the central Punjab district of Pakistan. The breed is named after the Sahiwal district, part of its home track (Joshi et al., 2001; Akram and Khan, 2011; Saeed et al., 2020). They are members of a category of giant Zebu breeds known as dual purpose. Due to its higher milk production and growth capacity than other Zebu cattle breeds, it is primarily used for milk and meat production (Ilatia et al., 2011; Iqbal et al., 2015). The Sahiwal breed of dairy cattle holds significant importance in Pakistan, mostly attributed to its ability to withstand high temperatures, resilience to diseases, and satisfactory performance when fed low-quality roughages. Artificial selection has been shown to improve the prevalence of advantageous alleles associated with economic traits, facilitating the enhancement of production parameters (Kim et al., 2015; Mei et al., 2019). Domestication and breeding of mammals have exerted consistent selection pressure on a wide range of characteristics in many domesticated species, resulting in discernible genetic modifications at the individual...
genome level (Jensen, 2015; Jensen and Wright, 2022). The genomic regions under selection pressure shows a functional variation associated with the traits (Driscoll et al., 2009; Stamps and Groothuis, 2010).

In Pakistan, researchers have been employing the progeny testing program as part of a selection effort to increase the milk production capacity of Sahiwal cattle for decades under the coordination of the Research Centre for Conservation of Indigenous Breeds (RCCIB), Jhang (Moaeen-ud-Din et al., 2014). The program involves registering and documenting institutional and private herds of Sahiwal cattle, recording for genetic evaluation, and identifying superior germplasm for genetic evaluation (McGill, 2015). However, applying for a conventional progeny testing program under Pakistani conditions appears to be difficult due to the small size of the herd, low awareness among the farmer community about pedigree and performance recording, resource limitations, and a lack of fundamental infrastructure (Shah et al., 2008; Moaeen-ud-Din et al., 2014). Thus, the efficiency of the Sahiwal cattle has not improved much over the years (Moaeen-ud-Din et al., 2014; Zawran et al., 2017). The efficiency of the Sahiwal breed can be enhanced by improving nutrition and management practices, estimating genetic parameters, and identifying the genes as well as quantitative trait loci linked with vital economic traits and subsequence genomic selection for such genes (Meuwissen et al., 2001; Rehman and Khan, 2012; Khan et al., 2018). Thus, genomic areas of Sahiwal cattle will continue to experience intense selective pressures for an extended period as the quest to select the best animals with high milk production capability continues in the country (Haskell et al., 2014; Illa et al., 2021). The existing body of literature on the outcome of selection on the Pakistan Sahiwal cattle remains limited at present. Hence, it is imperative to investigate and explore the genomic signatures of selection in Sahiwal cattle to comprehend the molecular mechanisms that influence quantitative as well as other significant traits (Pedrosa et al., 2021; Zhang et al., 2022b; Rajawat et al., 2023). Additionally, annotating the genes and quantitative trait loci (QTL) linked with economically vital traits is crucial (Illa et al., 2021; Zhang et al., 2022b).

Selection signatures refer to distinct genetic variations that occur at the DNA level as a result of deviations in the genomes of the chosen as well as neutral loci within a species that has experienced selection over time (Kreitman, 2000). Selection signatures are found in species subjected to selection during their evolution (Bamshad and Wooding, 2003; Laland et al., 2010). Variants subjected to selection pressure can cause characteristic genomic patterns to emerge, including a change in the distribution of allele frequencies, an increase in the proportion of homozygous genotypes, the prevalence of long haplotypes, and a significant degree of population substructure (Pritchard et al., 2010; Zhang et al., 2015). Modern cattle have undergone extensive selection over the centuries, resulting in dramatic phenotypic changes in the last 40 years (Pitt et al., 2019; Frantz et al., 2020; Brito et al., 2021). The development of affordable genotyping techniques has allowed more individuals to genotype using different densities of single nucleotide polymorphism (SNP) arrays (Boichard et al., 2012; Gorjanc et al., 2015; Cortes et al., 2022). New polymorphism data and the subsequent release of the bovine genome sequence have provided useful new resources for the search for evidence of recent selection in the bovine genome (Utsunomiya et al., 2013; Xu et al., 2015). This has improved precision and accuracy in identifying specific genomic areas in cattle (Hayes et al., 2009; Meuwissen et al., 2022). These advances have also contributed to identifying and analyzing genetic variations subject to natural selection in Homo sapiens as well as other animal species (Oleksyk et al., 2010; Luikart et al., 2019). When an allele experiences positive selection, it experiences a selective sweep when it becomes more common in the population (Moradi et al., 2012; Booker et al., 2017). Genetic hitch-hiking is the process through which closely related alleles increase frequency along with the positively chosen allele (Booker et al., 2017). An area of the genome where the positively selected haplotype is more prevalent due to a strong selection sweep would have less haplotype diversity (Moradi et al., 2012). Measuring LD or checking whether a haplotype is over-represented in a population are good ways to look for evidence of a selective sweep (Hayes, 2007; Zhang et al., 2022a).

Several different statistical models have been created to identify signs of selection. Several studies have been performed to detect the signatures of selection using several statistical techniques, such as the integrated haplotype score (iHS) (Voight et al., 2006), Tajima’s D (Tajima, 1989), fixation index (FST) (Akey et al., 2002), and the extended haplotype homozygosity (EHH) (Sabeti et al., 2002). These studies use methods complementarity to improve statistical power (Illa et al., 2021; Wainema et al., 2022). The iHS and Tajima’s D estimators are especially helpful among the numerous statistics used to recognize signs of positive selection from polymorphism data and would be the techniques of choice in this investigation (Zeng et al., 2007; Chen et al., 2010). The discovery of genomic areas under selection pressure has the potential to improve our comprehension of the underlying biology of certain phenotypes. This knowledge may be used to build techniques to improve selection efficiency (Moradian et al., 2020). Therefore, the study aimed to discover and analyze potential indicators of recent selection in Sahiwal
cattle, specifically identifying the genes and quantitative trait loci (QTL) linked with these selection indicators.

MATERIALS AND METHODS

Animal resources and SNP genotyping

The study’s sample size consisted of 98 Sahiwal cattle bulls, which were sourced from public as well as private livestock farms in the Punjab area of Pakistan. Blood samples were collected from all the 98 Sahiwal cattle bulls. The researchers employed a salting-out method to extract genomic DNA from blood samples, as described by Miller et al. (1988). The NanoDrop ND-1000 spectrophotometer, manufactured by NanoDrop Technologies in Wilmington, DE, was utilized to quantify the concentration of the isolated DNA. Good quality DNA samples were sent for genotyping utilizing the BovineHD140k Bead Chip (Illumina Inc. in San Diego, California, USA) following the standard operating procedure described by the manufacturer. Raw data was processed to generate ped files that provide genotypes and map files having a genomic location of markers.

Quality control

The PLINK program (Purcell et al., 2007) conducted quality control assessments on genotyping data. The research study opted for a call rate exceeding 95% for the collection of study data as well as subsequent analysis. The study employed SNPs with minor allele frequencies (MAFs) below 0.05. The analysis of the study included the examination of markers and animals, which did not exhibit a substantial divergence from Hardy-Weinberg proportions (P > 0.001, Bonferroni corrected). The analysis exclusively incorporated SNPs that were identified within autosomal chromosomes. Furthermore, samples that exhibited a missing genotyping rate exceeding 10% were excluded from the analysis. SNPs were subjected to a filtering process in order to exclude loci that were allocated to unmapped contigs and chromosomes associated with sexual determination. After undergoing quality control procedures, a total of 87 samples and 74,070 SNPs were deemed suitable for further analysis. These SNPs were distributed over the 29 autosomes.

Calculation of the integrated haplotype score

The rehh package (Gautier and Vitalis, 2012) in R Software was utilized to analyze the integrated haplotype score (iHS) test. The iHS score is derived from the assessment of extended haplotype homozygosity (EHH) linked to each allele. The computation of single-site iHS values was performed for each animal throughout the whole genome. These values were then averaged in non-overlapping windows of 500 kb throughout the genome. The window size was adapted on the bases of the extent of LD as defined by Qanbari et al. (2011). The unstandardized iHS can be obtained by using the following calculation:

\[
\text{Unstandardized } iHS = \ln \left( \frac{|HR_A|}{|HR_D|} \right) \quad (1)
\]

Where iHH\textsubscript{A} as well as iHH\textsubscript{D} signify the integrated EHH score for ancestral as well as derived core alleles, separately. This value has been normalized such that the mean is 0 and the standard deviation is 1, and this is done regardless of the allele frequency at the core SNP (Voight et al., 2006).

The standardized iHS was computed as follows:

\[
\text{iHS} = \frac{1}{n} \left( \ln \left( \frac{|HR_A|}{|HR_D|} \right) - E \left[ \ln \left( \frac{|HR_A|}{|HR_D|} \right) \right] \right) \quad (2)
\]

In this context, iHH\textsubscript{A} and iHH\textsubscript{D} denote the integrated EHH score for ancestral (A) and derived (D) core alleles, while E and SD reflect the expectation and standard deviation of unstandardized iHS, respectively. To determine the P value at the genome level, the iHS scores for each single nucleotide polymorphism (SNP) were subjected to a subsequent transformation. This transformation involved calculating the p\textsubscript{iHS} = − log [1 − 2Φ(iHS) − 0.5]]. The function Φ(x) denotes the Gaussian cumulative distribution function in the context of neutrality, whereas p iHS refers to the two-sided P-value linked to the neutral hypothesis (Gautier and Naves, 2011).

Tajima’s D statistics

Tajima’s D statistics were also computed for each chromosome utilizing the vcftools software (Danecek et al., 2011). The Tajima D index was calculated using nonoverlapping sliding windows of 100 MB as a parameter (Tajima D 100). Within each bin, the p-values were calculated and given to each SNP. Zero was substituted for any missing values.

Identification of functional genes and QTL

SNPs exhibiting statistically significant iHS and Tajima D values were analyzed utilizing the ARS-UCD1.2 assembly (Rosen et al., 2018) as well as Ensembl Genome Browser (Zerbino et al., 2018) databases. This analysis aimed to discover QTL genes or neighboring genes associated with these SNPs. The genes and QTLs were acquired from the ARS-UCD1.2 assembly (Rosen et al., 2018) as well as the Animal QTL database (Hu et al., 2016), respectively, in the gtf and gff formats. The study utilized Panther databases (Mi and Thomas, 2009) to document the molecular functions and biological processes associated with the discovered genes. The QTL animal database (available at https://www.animalgenome.
org/cgi-bin/QTLdb/BT/index) (Hu et al., 2016) was utilized to ascertain the QTL reported in the literature for each specific candidate region. The gene as well as QTL annotations were conducted with the R package GALLO (Fonseca et al., 2020). The enrichment analysis of QTLs was performed on all QTLs annotated using the chromosome-based technique, utilizing the GALLO program. Implementing a bootstrap approach involved conducting a correlation analysis between the observed as well as expected number of QTLs per characteristic, using data obtained from the cattle QTL database. The p-values derived from the enrichment analysis were adjusted utilizing the false discovery rate (FDR) approach. A significance threshold of less than 5% was used to accommodate for the numerous tests conducted. Gene ontology and pathway analyses were performed using the TOPGO and KEGGREST packages in R software.

RESULTS

Genome-wide distribution of iHS

The iHS distribution exhibited a close approximation to normality, as indicated by $iHS \sim N(0, 1)$. Consequently, it was possible to make comparisons between the markers and the chromosomes (Fig. 1). The unstandardized iHS within frequency bins (IHS-b) was used to identify the selection signatures in dairy cattle (Fig. 2). The IHS-b was calculated for each SNP within each frequency bin. The frequency bins were created by dividing the range of possible frequencies into equal intervals. The IHS-b was then calculated for each frequency bin. The SNPs with the highest IHS-b values were identified as signature SNPs.

Figure 3 shows the distribution of iHS by chromosome, while Figure 4 shows the alteration of analogous markers into piHS. The iHS statistic was plotted against the genomic location of the breed to visually represent the distribution of outlier signals across chromosomes. Since the iHS test found many important signals, we used the maximum iHS and piHS values to assess genes in the target genomic regions. Considering the iHS values, the genomic regions under the recent signatures of selection were found on BTA 1, 2, 6, 11, 12, 15, 17, 21, and 27 (Table I). The most convincing evidence of selection in the Sahiwal cattle is on BTA6 with an iHS score of -5.23 as well as on BTA27 with an iHS score of -5.58.

Table I. Genomic autosomal regions and potential
Signatures of Selection in the Pakistan Sahiwal Cattle

genes identified by the integrated haplotype score (iHS) test under the indication of the signature of selection in Sahiwal cattle.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position</th>
<th>Candidate gene</th>
<th>iHS</th>
<th>PiHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>132552990</td>
<td>PPP2R3A</td>
<td>-4.019491072</td>
<td>4.234152759</td>
</tr>
<tr>
<td>2</td>
<td>26259557</td>
<td>UBR3</td>
<td>-4.035439883</td>
<td>4.263602754</td>
</tr>
<tr>
<td>2</td>
<td>59837978</td>
<td>THSD7B</td>
<td>-4.533780388</td>
<td>5.237041985</td>
</tr>
<tr>
<td>2</td>
<td>92293678</td>
<td>PARD3B</td>
<td>-4.268678484</td>
<td>4.706340383</td>
</tr>
<tr>
<td>6</td>
<td>13502824</td>
<td>AP1AR</td>
<td>-5.238324881</td>
<td>6.790376077</td>
</tr>
<tr>
<td>6</td>
<td>21884446</td>
<td>CENPE</td>
<td>-4.852971281</td>
<td>5.914976041</td>
</tr>
<tr>
<td>6</td>
<td>22334879</td>
<td>MANBA</td>
<td>-4.526024536</td>
<td>5.221099479</td>
</tr>
<tr>
<td>6</td>
<td>22363715</td>
<td>MANBA</td>
<td>-4.266206770</td>
<td>4.701635120</td>
</tr>
<tr>
<td>6</td>
<td>22695313</td>
<td>SLC39A8</td>
<td>-4.210045351</td>
<td>4.592916130</td>
</tr>
<tr>
<td>11</td>
<td>45961342</td>
<td>CHCHD5</td>
<td>-4.824651776</td>
<td>5.853103064</td>
</tr>
<tr>
<td>11</td>
<td>45985907</td>
<td>CHCHD5</td>
<td>-4.829806865</td>
<td>5.864309063</td>
</tr>
<tr>
<td>11</td>
<td>47097519</td>
<td>EIF2AK3</td>
<td>-4.145736881</td>
<td>4.701558085</td>
</tr>
<tr>
<td>12</td>
<td>82530142</td>
<td>NALF1</td>
<td>-4.489584574</td>
<td>5.146331708</td>
</tr>
<tr>
<td>15</td>
<td>1.35E+08</td>
<td>ANAPC13</td>
<td>1.35E+08</td>
<td>1.35E+08</td>
</tr>
<tr>
<td>21</td>
<td>47961935</td>
<td>LOC107132255</td>
<td>4.81811888</td>
<td>4.701635120</td>
</tr>
<tr>
<td>27</td>
<td>13371357</td>
<td>TENM3</td>
<td>-4.738578572</td>
<td>5.667112173</td>
</tr>
</tbody>
</table>

Fig. 5. Manhattan plot of Tajima’s D values.

Tajima’s D statistics

Most autosomes exhibited at least one significant signal of selection (Fig. 5), and high Tajima D values were found to be under positive selection in various contexts. The study only considered negative values representing current positive selection signals. Tajima’s D test found significant selection signals in 51 genomic regions. Current signals of positive selection were detected on BTA 1, 2, 3, 4, 5, 6, 7, 9, 18, 20, 23, and 24. Tajima’s D signals comprised 31 candidate genes (Table II). Furthermore, the 31 genes identified within the selection regions are linked with milk composition traits like milk protein yield, milk alpha casein percentage, milk fat percentage, milk fat yield, milk alpha-lactalbumin percentage, milk protein percentage, as well as milk beta-casein percentage according to previous reports.

Table II. Genomic autosomal region and potential genes identified by Tajima’s D test under the indication of a positive signature of selection in Sahiwal cattle.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Start position</th>
<th>End position</th>
<th>Candidate gene</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29105747</td>
<td>29418948</td>
<td>GBE1</td>
<td>Milk protein yield</td>
</tr>
<tr>
<td>1</td>
<td>1.35E+08</td>
<td>1.35E+08</td>
<td>ANAPC13</td>
<td>Milk alpha casein percentage</td>
</tr>
<tr>
<td>2</td>
<td>1.35E+08</td>
<td>1.35E+08</td>
<td>LOC107132255</td>
<td>Milk fat content</td>
</tr>
<tr>
<td>2</td>
<td>1.27E+08</td>
<td>1.27E+08</td>
<td>STMN1</td>
<td>Milk fat yield</td>
</tr>
<tr>
<td>3</td>
<td>51855689</td>
<td>51880658</td>
<td>CDC7</td>
<td>Milk protein yield</td>
</tr>
<tr>
<td>3</td>
<td>30266039</td>
<td>30268283</td>
<td>LRIG2</td>
<td>Milk protein percentage</td>
</tr>
<tr>
<td>5</td>
<td>98993095</td>
<td>99048504</td>
<td>LOC101902742</td>
<td>Milk alpha lactalbumin percentage</td>
</tr>
<tr>
<td>6</td>
<td>55443414</td>
<td>55635282</td>
<td>ARAP2</td>
<td>Milk protein percentage</td>
</tr>
<tr>
<td>7</td>
<td>328408</td>
<td>329337</td>
<td>LOC107131408</td>
<td>Milk fat yield</td>
</tr>
<tr>
<td>9</td>
<td>32590633</td>
<td>33013293</td>
<td>SLC35F1</td>
<td>Milk fat yield</td>
</tr>
<tr>
<td>18</td>
<td>43093119</td>
<td>43130616</td>
<td>ANKRD27</td>
<td>Milk fat yield</td>
</tr>
<tr>
<td>20</td>
<td>42774735</td>
<td>42774841</td>
<td>LOC112443073</td>
<td>Milk fat yield</td>
</tr>
<tr>
<td>24</td>
<td>27152147</td>
<td>27152251</td>
<td>LOC112444247</td>
<td>Milk fat content</td>
</tr>
</tbody>
</table>

Analysis of QTL identification.
The study identified significant genomic regions consisting of 25.08% of milk-type QTLs in Sahiwal cattle as well as other QTLs associated with traits like meat as well as carcass, production, health, reproduction, as well as exterior, which were annotated and represented 21.54, 18.72, 13.85, 11.05, and 9.76%, respectively (Fig. 6). Milk-type QTLs included loci associated with milk protein yield, milk yield, milking speed, milk beta-casein content, milk energy yield, stearic acid content, milk oleic acid content, milk kappa-casein percentage, curd firming rate, milk caproic acid content, milk alpha-casein content, milk alpha-S2-casein percentage, milk casein content, milk color, milk kappa-casein content, milk protein content, milk odd-chain fatty acid percentage, milk palmitoleic acid content, milk saturated to unsaturated fatty acid ration, and milk whey protein content (Fig. 7).

The QTL enrichment analysis was done to obtain unprejudiced data on the significant QTLs in the population instead of doing the QTL annotation. The QTL enrichment analysis has shown 8 substantial QTLs on BTA19 and BTA20, which are linked with milk, production, as well as reproduction traits (Table III). The utmost substantial QTLs were mapped on BTA19 and BTA20, linked with fat percentage, fertilization rate, milk fat yield, early embryonic survival, body depth, body weight, and average daily gain (Fig. 8). Intriguingly, the highest significant QTL identified on BTA19 was linked with milk fat yield as well as fat percentage.

Table III. The enriched QTLs were annotated in the assumed genomic areas.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Chromosome</th>
<th>Number of QTLs</th>
<th>Number of annotated QTLs</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>19</td>
<td>16</td>
<td>321</td>
<td>0.040234</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>12</td>
<td>321</td>
<td>0.046226</td>
</tr>
<tr>
<td>Production</td>
<td>19</td>
<td>22</td>
<td>321</td>
<td>0.000171</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>19</td>
<td>210</td>
<td>0.008069</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>11</td>
<td>321</td>
<td>0.013937</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>8</td>
<td>321</td>
<td>0.045253</td>
</tr>
<tr>
<td>Reproduction</td>
<td>19</td>
<td>12</td>
<td>321</td>
<td>0.009393</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>10</td>
<td>321</td>
<td>0.020660</td>
</tr>
</tbody>
</table>

Fig. 8. Quantitative trait loci (QTL) enrichment analysis resolve the vital traits enriched in the substantial genomic areas.

Gene enrichment analysis for iHS

The gene ontology (GO) enrichment analysis of the significant genes of the iHS scores was categorized.
into biological processes (Fig. 9A), molecular function (Fig. 9B), and cellular components (Fig. 9C). The GO enrichment analysis revealed 30 biological processes, 6 molecular functions, and 22 cellular components for the iHS test.

Fig. 9. Biological process (A), molecular function (B) and cellular component (C) of gene network computed for the iHS significant potential genes. The cut off lines drawn at equivalents of p=0.05, p=0.01, and p=0.001. KS means the p-value of Kolmogorov-Smirnov test implemented in the R package ‘topGO.’
Fig. 10. Biological process (A), molecular function (B) and cellular component (C) of gene network computed for the Tajima’s D significant potential genes. The cut off lines drawn at equivalents of p=0.05, p=0.01, and p=0.001. KS means the p-value of Kolmogorov-Smirnov test implemented in the R package ‘topGO.’
**DISCUSSION**

The study investigated the signature of selection in the genomes of Sahiwal cattle utilizing two methods (iHS and Tajima D). Considering iHS values, the regions on BTA 1, 2, 6, 11, 12, 15, 17, 21, and 27 are under the recent signatures of selection (Table I). The study observed recent signatures of selection on BTA 1, 2, 3, 4, 5, 6, 7, 9, 14, 18, 20, 23, and 24 with Tajima’s D test (Table II). This study also noticed results that were similar to those observed by other researchers. In Thai dairy cattle, Buaban et al. (2022) found that regions associated with milk production traits were located on BTA 1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 16, 20, 21, 26, 27, and 29. Kolbehdari et al. (2009) performed a whole genome scan on Canadian Holstein bulls to identify QTL for milk production characteristics and somatic cell score. Their research identified map QTL significantly associated with protein yield in six SNP at the genome level, and nine at the chromosome level on BTA 1, 4, 7, 8, 9, 11, 14, 18, 21, 23, 26 and 28. 13 SNP significantly related to fat yield at the genome level, and 7 at the chromosome level on BTA 1, 4, 5, 7, 10, 11, 14, 21, 23, 24 and 28; 6 and 12 SNP that were significantly associated with the fat percentage in the genome were found on BTA 2, 3, 6, 9, 10, 14, 17, 21, 23, and 26 and one SNP at genomic and 9 SNPs at the chromosome level were found on BTA 3, 4, 5, 10, 13, 17, 22, and 23 for the percentage of protein. In a whole genome evaluation of recent selection signatures in Sarabi cattle from Iran, Moradian et al. (2020) identified statistically significant SNPs on BTA5, BTA7, BTA10, BTA14, and BTA17. Kadri et al. (2015) found potential genes on BTA20 that are related to milk production, percentage of proteins, and resistance to mastitis.

The common genomic region identified by both tests (iHS and Tajima’s D) was on BTA 1, 2, and 6. Pitt et al. (2019) identified positive selection in Creole cattle breeds on BTA1 that harbored genes associated with Polled, milk production, and reproduction. Alshawi et al. (2019) found the most compelling evidence of selection in Jenoubi cattle in BTA1 with an iHS score of -5.40 as well as in BTA26 with an iHS score of -5.0. Iraqi Rustaqi cattle notice a clear selection indication at BTA1 with an iHS score of -5.00 as well as BTA18 with an iHS score of -5.03. Hayes et al. (2008) used an LD-based iHS technique and found many QTLs associated with milk production traits. They found selection signatures present on BTA6 in Norwegian red cattle. In a study on selection signatures employing ROH patterns in four different cattle breeds, Szmatoła et al. (2016) discovered that a homozygous area under selection on BTA2 was related to QTL for the muscling trait in the Limousin breed. At least three QTLs that influence milk traits are located on BTA6 (Khatkar et al., 2004; Ogorevc et al., 2009), and it has been hypothesized that dairy breeders are selecting these areas to improve milk production (Schwarzenbacher et al., 2012). Using iHS, FST, as well as XP-EHH methods, Maiorano et al. (2018) demonstrated the existence of QTLs that impact milk as well as meat quality traits in dual-purpose Gir cattle populations on BTA6. Lee et al. (2016) also found a gene on BTA6 that was under selection and related to milk production parameters like milk yield, fat composition, as well as protein yield in Holstein dairy cattle.

These genomic regions identified by both tests in the study harbor candidate genes linked with milk, production and reproduction. PARD3B identified on BTA2 in the study has been found to be linked with bovine development and neural development in red Angus beef cattle (Smith et al., 2022). The gene has also been found to be linked with fat percentage in Danish Holstein cattle (Buitenhuis et al., 2014). PARD3B in red Angus beef cattle has been found to be associated with bovine tuberculosis traits (Raphaka et al., 2017). The PARD3B gene has also been identified in pigs as a candidate gene for body weight (Xu et al., 2020). The CENPE identified on BTA6 in the study has been found to be a candidate gene linked with milk composition traits in Holstein cattle (Jiang et al., 2016). The CENPE gene has also been identified in cattle to be linked with residual feed intake (Rathert et al., 2020). CSMD1 and TENM3 have also been detected on BTA27 in the study. Gonzalez et al. (2020) found CSMD1 to be associated with the rear udder height trait in Holstein cows. In the context of Hanwoo cattle, it was observed that CSMD1 exhibited higher expression levels in muscle samples derived from animals with elevated carcass weight, particularly in relation to intramuscular fat content and eye muscle area (Lee et al., 2011). Hoff et al. (2019) identified a region in the CSMD1 gene that regulates the complement system that controls inflammatory responses in Holstein cattle. CSMD1 has also been identified in goats as associated with goat fertility (Li et al., 2022). TENM3 has been found to be a candidate gene within the most significant QTL, which is associated with wither height or stature in beef cattle (Doyle et al., 2020). TENM3 has also been identified in sheep as associated with milk production traits (Sutera et al., 2019). TENM3, among others, affects the metabolic pathways of cell differentiation.
and proliferation and is linked with the regulation of the immune system in goats (Krivoruchko et al., 2022).

The identification of QTLs in the study discovered that significant genomic areas consist of 25.08% of milk-type QTLs in Sahiwal cattle as well as other types like meat as well as carcass, production, health, reproduction, as well as exterior, which were annotated and represented 21.54, 18.72, 13.85, 11.05, and 9.76%, respectively (Fig. 6). The study revealed a low percentage of milk-type QTLs in Pakistani Sahiwal cattle compared to Indian Sahiwal cattle, where Illa et al. (2021) observed 54.6% milk-type QTLs. This may suggest that Pakistani Sahiwal cattle have not been subjected to intense selection for milk production traits. The QTL enrichment analysis has shown 8 substantial QTLs on chromosomes BTA 19 and BTA20, associated with milk, production, as well as reproduction traits. The utmost significant QTLs were assigned to BTA19 and BTA20, associated with milk fat yield, fat percentage, fertilization rate, early embryonic survival, body weight, body depth, and average daily gain (Fig. 8). Illa et al. (2021) conducted signature selection in Indian Sahiwal cattle, and the QTL enrichment analysis painted 14 substantial locations in BTA 1, 3, 6, 11, 20, as well as 21. They found that the top most three enriched QTLs were in BTA 6, 20, as well as 23, linked to the exterior, milk production, health, as well as reproduction traits. They obtained a more significant region, which may be due to Indian Sahiwal cattle being under intense selection for milk yield; they have a milk-type QTL of 54.6% as compared to the Pakistani Sahiwal cattle, which have just 25.08% milk-type QTLs.

CONCLUSION

The study findings revealed many genomic regions and several novel genes that exhibited positive selection in the Sahiwal cattle for both the test (iHS and Tajima’s D) used in the study. Considering iHS values, the regions on BTA 1, 2, 6, 11, 12, 15, 17, 21, and 27 are under the recent signatures of selection. While genomic regions on BTA 1, 2, 3, 4, 5, 6, 7, 9, 14, 18, 20, 23, and 24 are under recent signatures of selection with Tajima’s D test. The common genomic region identified by both tests under positive selection signatures was BTA 1, 2, and 6. These genomic regions harbor candidate genes linked with milk composition traits. The study also observed that the Sahiwal cattle population has low milk-type QTLs of 25.08%, indicating that the population is not under intense selection for milk production traits. Therefore, understanding the selection signatures and candidate genes that influence important economic traits can provide foundational knowledge that can be used effectively to gain insight into the underlying mechanisms controlling these traits in Sahiwal cattle. Additionally, the results might serve as a basis for further investigation of economically vital traits linked to milk production in Sahiwal cattle.

Funding

This study was funded by the Higher Education, Pakistan through a NRPU project No. Prediction of genetic values of economic traits of Sahiwal cattle.

Ethical approval

Ethical approval was given by the Ethics Committee of University Agriculture Faisalabad, Pakistan.

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES


Signatures of Selection in the Pakistan Sahiwal Cattle


https://doi.org/10.1093/nar/gkv1233


