ISSN 0032-7786

DOI Link: https://doi.org/10.61586/OpROH
Vol.97, Issue.9, Part.1, Sep 2025, PP. 2-21

Validation of ERBB2 and HSPA8 Gene Polymorphisms for Milk Productivity in endemic Zebu Cattle (Sahiwal breed)

Asma Younis[a], Dmitrii E. Aleshin[b], Imtiaz Hussain[c], Syeda Nadia Ahmad*[d], Iram Inayat[a], Muhammad Ali Kanwal[a], Khawaja Raees Ahmad[e], Naila Kanwal[a], Saima Matloob[a], Muhammad Atif Kamran[a], Rameen Khalid[a], Sadia Suleman[e] and Mehwish Nasir[a]

[a] Department of Zoology, University of Sargodha, Sargodha, Pakistan.
[b] Candidate of Sciences (Biology), Russian State Agrarian University – Moscow Timiryazev Agricultural Academy.
[c] Department of Animal Sciences, University of Sargodha, Sargodha, Pakistan
[d] Faculty of Sciences, Department of Zoology, University of Chakwal, Punjab, Pakistan
[e] Ex-Professor, Department of Zoology, University of Sargodha, Sargodha, Punjab, Pakistan.
*Correspondence: Syeda Nadia Ahmad, Faculty of Sciences, Department of Zoology, University of Chakwal, Punjab, Pakistan, Email: nadia.ahmad@uoc.edu.pk

Received Feb 2025 Accepted Aug 2025 Published Sep 2025

Abstract:

It has been shown that SNPs in both the ERBB2 and HSPA8 significantly affect milk yield in dairy cattle. In present study the reported taurine SNPs of these two genes were explored for their presence and significance in terms of milk productivity in famous Zebu cattle milk breed (Sahiwal). DNA from alveolar epithelial cells separated from the milk of 560 dams, and tested for the presence of selected nine *Bos taurus* SNPs (7 form ERBB2 and 2 from HSPA8 genes) using specially designed cited specific primers employing NCBI database.

The Five SNPs rs133031530 (A/T), rs109763505 (A/G), rs109122971 (T/C), rs1107735562 (A/G) and rs109017161 in ERBB2 gene and SNPs rs133632043 (T/G), rs132976221 (A/C) in HSPA8 gene were significantly associated with milk yield. Results of the present study confirm the SNPs in ERBB2 and HSPA8 that effect the milk protein concentration also enhance the per capita milk production. Thus these SNPs may be considered as potential genetic marker of milk productivity in Sahiwal Zebu dairy cattle breed.

Keywords: Sahiwal Cattle, Single Nucleotide Polymorphism, Milk Production

ISSN 0032-7786

1. Introduction

Milk production capacity in dairy cattle breeds has great economic importance (Gross, 2022, Kumaresan, A., & Srivastava, 2022). Unlike a simple Mendelian trait, milk production is a complex quantitative trait influenced by multiple genes and their interactions. In recent years, molecular genetics approaches, particularly single nucleotide polymorphisms (SNPs) studies followed by genome-wide association studies (GWAS), have become powerful tools for identifying genetic markers that can assist in the selection of superior sires and heifers by animal breeders, thereby enhancing genetic gain in milk production traits." (Križanac et al., 2025). In the current scenario of molecular genetics marker (SNPs) studies followed by gemone wide association can be a very effective tool for selection of sires and heifers by the animal breeders.

Along with many other the ERBB2 and HSPA8 genes are promising candidate for milk production traits (Do et al., 2017). Recent large-scale association studies in Chinese Holstein populations have confirmed that multiple SNPs in ERBB2 are significantly associated with milk protein percentage and other milk production traits, with some SNPs explaining phenotypic variances above 1%, making them valuable genetic markers for breeding programs (Li et al., 2019).The ERBB2 gene encodes tyrosine kinases receptor. It is located on the bovine autosomal chromosome 19. It consists of 24,682 DNA base pairs having 27 exons and 26 introns. This gene encodes 1255 amino acids (Anderson et al., 2007). The ERBB2 protein directly binds to the PI3K regulatory subunit p85, phosphorylating tyrosine residues to activate the PI3K signaling pathway. This signaling pathway is involved in the regulation of milk protein synthesis. Therefore, SNPs in the ERBB2 gene are expected to influence both the quality and quantity of milk production (Li et al., 2019).

The HSPA8 gene is located on bovine autosome 15. It consist of 4426 DNA base pairs having 8 introns and 9 exons. It encodes 650 amino acid (Al-Thuwaini et al., 2022). From the evolutionary point of view it is highly conserved protein in Swine and Bovine (Elliott et al., 2009). HSPA8 is involved in a variety of physiological functions, including protein synthesis, protein folding, and protein degradation. The HSPA8 gene is highly expressed in lactating mammary glands (Li et al., 2016; Paten et al., 2015).

One important role of HSPA8 is in the regulation of protein folding and processing in the endoplasmic reticulum. This function is particularly crucial for proteins involved in milk production, such as beta-casein (Lee, 2014).

In summary, HSPA8 plays an important role in the proper development and function of the mammary gland during lactation. Its involvement in the regulation of protein synthesis, folding, and degradation, especially for proteins important for

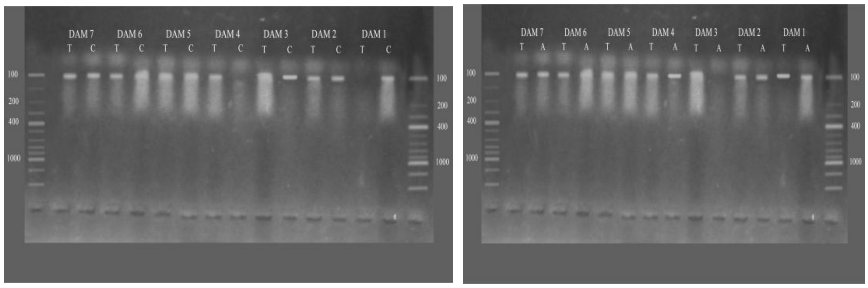
milk production, highlights its importance for milk productivity. The high expression of the HSPA8 gene in lactating mammary glands indicates its critical role in these processes. Thus the HSPA8 is also considered as potential candidate for milk protein synthesis (Li et al., 2019) that may also affect milk productivity in dairy cattle. In present study 9 reported taurine SNPs (rs133632043 (T/G), rs132976221 of HSPA8 and rs133031530, rs109763505, rs109122971, rs110133654, rs109941438, rs1107735562 and rs109017161 of ERBB2 genes) were explored for their presence and association with the milk productivity traits in Sahiwal breed of Zebu cattle.

2. Materials and Methods

2.1 Sampling of Animals and Extraction of Genomic DNA

A total of 560 clinically healthy lactating cows of pure Sahiwal breed were selected from LES Khiz-rabad Farm, PLDDB, Punjab, Pakistan. The animals represented a range of parities, with particular emphasis on cows in their second to fifth lactation stages, and their ages ranged from 3 to 8 years. All cows were managed under uniform feeding and husbandry conditions to minimize environmental variation. Milk samples of approximately 50 mL were aseptically collected from each cow during mid-lactation to ensure consistency in physiological status. Samples were collected using sterile techniques to avoid contamination and were immediately stored at 4°C until DNA extraction.

Genomic DNA was extracted from somatic cells, specifically udder epithelial cells present in the milk, following the protocol described by Pokorska et al. (2016). This method involves centrifugation to pellet somatic cells, cell lysis, and purification steps that yield high-quality genomic DNA suitable for downstream molecular analyses. The presence and integrity of the extracted DNA were confirmed by electrophoresis on 1% agarose gels prepared with TBE buffer. DNA concentration and purity were measured and standardized using a PicoDrop 200 spectrophotometer to ensure consistency across all samples.



Pcr Amplificaaation in Dams

2.2. Selection of Genes and Genetic Markers

Candidate genes and associated single nucleotide polymorphisms (SNPs) were selected based on criteria outlined by Ogorevc et al. (2009). Markers were chosen if they showed significant association with milk yield traits in previous studies, had gene expression profiles linked to the milk production phenotype under study, and were located within known quantitative trait loci (QTL) regions relevant to milk yield in cattle.

2.3. SNP Genotyping Using Allele-Specific PCR (AS-PCR)

SNP genotyping was performed using the allele-specific polymerase chain reaction (AS-PCR) technique. Primers were designed based on *Bos taurus* genome data from NCBI, targeting biallelic SNPs. Each SNP was genotyped using three primers: two allele-specific forward primers and one common reverse primer. Primer lengths ranged from 18 to 24 nucleotides, producing PCR products approximately 100 base pairs to facilitate efficient amplification.

Table 1. List of Primers

1	rs110735562	ERBB2	Fwd	AGCGAGGTCCACACTGGTG
			Rev	TTGGGACTCTTGACCAGT
			Rev	TTGGGACTCTTGACCAGC
2	rs133031530	ERBB2	Fwd	GATGTCTACATGATCATGG
			Rev	ATGCCCCTTGTAGGCCCAA
			Rev	ATGCCCCTTGTAGGCCCAT
3	rs109763505	ERBB2	Fwd	GCATGGTGATATGGTTAATGGA
			Fwd	GCATGGTGATATGGTTAATGGG
			Rev	CAGTCTTTTTTCCTAGCTGG
4	rs109017161	ERBB2	Fwd	TTATGATGGGATCCCAG
			Rev	GACCATGATCATGTAGACA
			Rev	GACCATGATCATGTAGACG
5	Rs109122971	ERBB2	FWD	AGCGAGGACCCCACAGTAC
			REV	AGCTCGAGCCTGACTTCA
			REV	AGCTCGAGCCTGACTTCG
6	rs136632043	HSPA8	Fwd	TTGTGTGGAACAATGCTA
			Fwd	TTGTGTGGAACAATGCTC
			Rev	AGCTCCTCCATCAGGT
7	rs132976221	HSPA8	Fwd	GACAGCGTTGGTAACCGT
			Rev	CAAGGGGTGAGTGGTAAAAA
			Rev	CAAGGGGTGAGTGGTAAAAC

ISSN 0032-7786

2.4 PCR Amplification Protocol

The allele-specific PCR (AS-PCR) protocol was employed to accurately detect SNPs, addressing the limitation of simple PCR amplification. In this study, primers were designed in triplets for each SNP: two allele-specific primers and one common primer. For SNPs located on the forward strand, two forward allele-specific primers and one reverse common primer were designed and vice versa (list of primers). This design allows selective amplification of each allele based on perfect complementarity at the 3' end of the allele-specific primers, enabling discrimination between different SNP variants.All primers were designed to amplify a target product size of approximately 100 base pairs

PCR reactions were performed in a total volume of 50 µL, which is larger than the conventional volume (typically 20–25 µL). This increased volume was chosen to improve amplification efficiency and yield sufficient product for downstream analyses such as gel electrophoresis and sequencing. Each reaction contained 25 µL of 2X Taq PCR Master Mix (Bioshop Canada, Cat. No TAQ006D), 0.5 µL each of forward and reverse primers (corresponding to approximately 10 pmol of each primer), 1 µL of genomic DNA (containing approximately 50 ng of DNA), and 23 µL of nuclease-free water.

The allele-specific PCR (AS-PCR) protocol was standardized with all primers annealing at 55°C to simplify the procedure and ensure uniform amplification conditions. The thermal cycling conditions were as follows: initial denaturation at 94°C for 3 minutes; 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 1 minute; followed by a final extension at 72°C for 10 minutes. This protocol was optimized to achieve specific and efficient amplification of the target SNP regions.PCR products (100bp long) were separated on a 1% agarose gel and visualized under ultraviolet light using a UV transilluminator.

2.6.Statistical Analysis

The association between individual SNP genotypes and milk yield (per day yield) was analyzed using a single-marker linear regression model:

$Y=\beta 0 +\beta\ 1\ X+\epsilon$

where Y is the dependent variable representing the predicted milk yield, X is the independent variable representing the SNP genotype coded numerically (e.g., 0, 1, 2 for different genotypes), β_0 is the intercept, β_1 is the regression coefficient indicating the effect size of the SNP on milk yield, and ε is the random error term. The model's goodness-of-fit was evaluated using the adjusted coefficient of determination (R^2_adj).

To control for confounding factors, a factorial analysis of covariance (ANCOVA) was performed including animal age (in years) as a continuous covariate. This allowed assessment of SNP effects on milk yield while adjusting for age-related variability. All statistical analyses were conducted using IBM SPSS Statistics version 23, with significance set at $p<0.05$.

2.7.Haplotype Construction and Analysis

Haplotypes were constructed from genotyping data of five SNPs within the ERBB2 gene using SNPSTAT software, which applies an expectation-maximization algorithm to estimate haplotype frequencies and assign haplotype pairs to individuals. Allelic frequencies and linkage disequilibrium (LD) between SNP loci were calculated to support accurate haplotype inference.

2.8. Hardy-Weinberg Equilibrium Testing

Each SNP was tested for Hardy-Weinberg equilibrium (HWE) using a chi-square test to compare observed and expected genotype frequencies, with significance threshold set at $P<0.05$.

2.9. Linkage Disequilibrium Analysis

Pairwise LD between SNPs was quantified using the squared correlation coefficient (r2) calculated by SHESIS software. LD strength was visualized through a triangular heatmap matrix, where darker red shades indicate stronger LD. Standard quality control filters were applied, including genotyping call rates above 95%, minor allele frequency greater than 0.01, and HWE $p>0.0001$.

3. Results

3.1 Erb-b2 receptor tyrosine kinase 2 (ERBB2):

Seven SNPs within the ERBB2 gene (rs133031530, rs109763505, rs109122971, rs110133654, rs109941438 (excluded from analysis because of low minor allele frequency MAF< 5% of these SNPs), rs1107735562, and rs109017161) were analyzed in the Sahiwal cattle population. Among these, three SNPs (rs133031530, rs109763505, rs109122971) were located in intronic regions, while two (rs1107735562, rs109017161) were synonymous exonic substitutions. Genotype and allele frequencies for each SNP are presented in Table 2.

Hardy-Weinberg equilibrium (HWE) testing indicated that all five SNPs rs109763505, rs109122971 rs110735562, rs109017161, and rs133031530 conformed to equilibrium expectations ($P > 0.05$)

The association between each SNP and milk yield was evaluated using single-marker linear regression. The strength of association was quantified by the adjusted

ISSN 0032-7786

R^2 value, representing the proportion of variance in milk yield explained by each SNP (Table 3). All five SNPs tested showed statistically significant associations with milk yield ($P < 0.05$).

The SNP rs133031530 demonstrated the strongest association with milk productivity, exhibiting the highest adjusted R^2 value (0.814), indicating that approximately 81.4% of the variance in milk yield could be attributed to this variant. Other SNPs, including rs110735562, rs109763505, rs109017161, and rs109122971, also showed significant associations, with adjusted R^2 values ranging from 0.359 to 0.741 (Table 2).

Table 2. Genotype and Allelic frequency and Hardy-Weinberg equilibrium test of ERBB2 SNPs in Bos indicus (Sahiwal breed).

Gene	Position	Ref SNP	Location	Genotype	N	Frequency	Allele	Frequency	Hardy-Weinberg Equilibrium X2 Test
ERBB2	Rev	rs110735562	Synonymos variant	TT CT CC	217 175 168	0.3 0.31 0.39	C T	0.54 0.46	P>0.05
ERBB2	Rev	rs133031530	Intron-23	AA AT TT	252 196 112	0.45 0.35 0.2	A T	0.62 0.38	P>0.05
ERBB2	Fwd	rs109763505	Intron-23	AA GA GG	175 238 147	0.31 0.42 0.26	G A	0.52 0.48	P>0.05
ERBB2	Rev	rs109017161	Synonymo us variant	TT TC CC	182 196 182	0.32 0.35 0.32	T C	0.5 0.5	P>0.05
ERBB2	Rev	rs109122971	Intron-26	TT TC CC	189 238 133	0.32 0.45 0.22	C T	0.55 0.45	P>0.05

Table 3. Variation in milk yield predicted by Adjusted R2 value.

rs#	R	R square	Adjusted R2	SES	P value
rs110735562	.861	.742	**.**738	.787	.000
Rs133031530	.904	.816	.814	.663	.000
rs109763505	.762	581	.575	1.00	.000
rs109017161	.863	.744	.741	.783	.000
rs109122971	.606	.367	.359	1.23	.000

SES: Standered error of the estimates.

3.2 Single Locus Based analysis

Findings of this study show that the genotype AG (heterozygous) for SNP rs110735562 was associated with significantly higher milk productivity (10.43 ± 0.01 L) compared to the homozygous genotypes AA (9.33 ± 0.03 L) and GG (7.05 ± 0.09 L) ($P < 0.05$). Similarly, for SNP rs133031530, animals with the AT genotype showed higher cumulative milk production (10.40 ± 0.01 L) than those with the AA (9.03 ± 0.04 L) and TT (6.44 ± 0.09 L) genotypes, respectively ($P < 0.05$). The results also revealed that for SNP rs109763505, animals in the heterozygous condition AG were more productive (10.06 ± 0.04 L) than those with the AA (9.10 ± 0.04 L) and GG (7.12 ± 0.03 L) genotypes ($P < 0.05$).

Regarding SNP rs109017161, the genotype TC was associated with higher milk yield (10.40 ± 0.01 L) compared to either of the homozygous genotypes TT (7.17 ± 0.09 L) and CC (9.29 ± 0.03 L) ($P < 0.05$). Furthermore, animals homozygous for the CC allele also showed relatively high milk yield, while the TT genotype showed the lowest milk productivity. These results suggest that the C allele has a dominant effect on milk yield.

For SNP rs109122971, animals homozygous for the CC genotype produced significantly more milk (10.41 ± 0.02 L) than those homozygous for TT (9.16 ± 0.04 L) and heterozygous TC (8.06 ± 0.12 L) genotypes ($P < 0.05$) (Table 4).

ISSN 0032-7786

Table 4. Single locus based analysis.

Genes	Locus	Levels	Mean milk yield±SEM	P value
ERBB2	rs110735562	AA	9.33±0.03	<0.05
		AG	10.43±0.01	
		GG	7.05±0.09	
ERBB2	rs133031530	AA	9.03±0.04	<0.05
		AT	10.4±0.01	
		TT	6.44±0.09	
ERBB2	rs109763505	AA	9.1±0.04	<0.05
		AG	10.06±0.04	
		GG	7.12±0.03	
ERBB2	rs109017161	TT	7.17±0.09	<0.05
		TC	10.4±0.01	
		CC	9.29±0.03	
ERBB2	rs109122971	TT	9.16±0.04	<0.05
		TC	8.06±0.12	
		CC	10.41±0.02	

(Mean of milk Yield ± standered error). P value refers to the results of association each SNP with milk yield. (Covariates appearing in the model are evaluated at the values: Age in years = 3.6919).

ISSN 0032-7786

3.3. Linkage Disequilibrium

Measurements were taken of the linkage disequilibrium between different pairs of loci, ranging from complete linkage disequilibrium to no linkage disequilibrium. As the distance between pairs of loci increased, the level of linkage disequilibrium decreased, and vice versa. It was observed that certain regions with high levels of linkage disequilibrium were not evenly distributed throughout the chromosome. The observed clustering of high LD value suggests the presence of functional haplotype blocks that may collectively influence milk production traits (Fig. A).

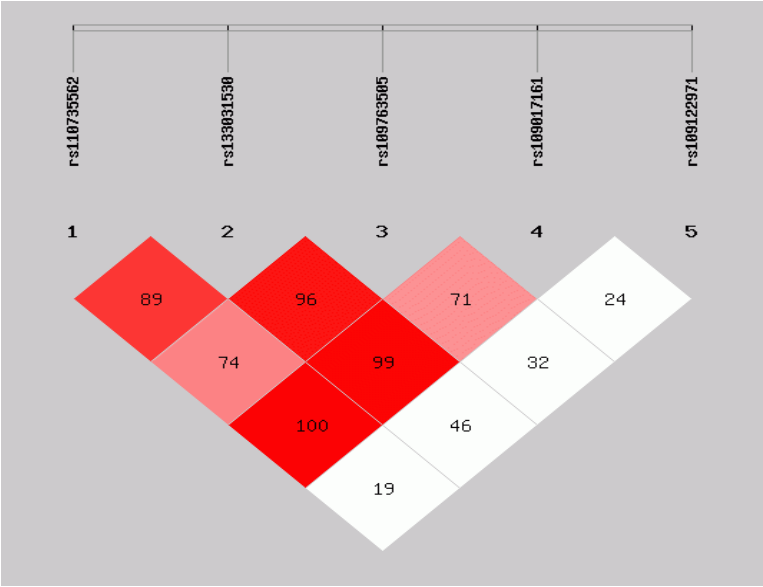


Fig A. Linkage Disequilibrium Plot of ERBB2 SNPs in Sahiwal Cattle

3.4. Haplotype Construction

Fourteen unique haplotypes were identified in the ERBB2 gene among the Bos indicus (Sahiwal breed) population. The most prevalent haplotypes were AAACT (frequency: 0.2935), GTGTC (0.2435), AAACC (0.144), and GTGTT (0.1127). In contrast, haplotypes such as ATGTC (0.0125), AAATC (0.0062), and ATATC (0.0062) were observed at much lower frequencies, while ATGCC was not detected in the studied population (Table 5).

ISSN 0032-7786

Table 5. ERBB2 Haplotype and their frequency observed in ERBB2 gene.

Haplotype	rs110735562	rs133031530	rs109763505	rs109017161	rs109122971	Total
AAACT	A	A	A	C	T	0.2935
GTGTC	G	T	G	T	C	0.2435
AAACC	A	A	A	C	C	0.144
GTGTT	G	T	G	T	T	0.1127
GAATT	G	A	A	T	T	0.0562
AAGCT	A	A	G	C	T	0.0438
GAGTT	G	A	G	T	T	0.025
AAATT	A	A	A	T	T	0.0188
AAGCC	A	A	G	C	C	0.0188
GAGTC	G	A	G	T	C	0.0188
ATGTC	A	T	G	T	C	0.0125
AAATC	A	A	A	T	C	0.0062
ATATC	A	T	A	T	C	0.0062
ATGCC	A	T	G	C	C	0

3.5. Haplotype based association analysis

Haplotype-based association analysis was conducted to evaluate the relationship between specific ERBB2 gene haplotypes and milk production in the studied population. Among the identified haplotypes, "GTGTC" and "GTGTT" were observed at frequencies of 0.2562 and 0.1000, respectively. The "GTGTC" haplotype was significantly associated with increased milk production, showing an average increase of 1.22 units (95% CI: 0.89–1.56; $P < 0.0001$). In contrast, the "GTGTT" haplotype was associated with a significant decrease in milk production, with an average reduction of 4 units (95% CI: -4.38 to -3.62; $P < 0.0001$). Additionally, the "GAATT" haplotype (frequency: 0.0562) was associated with a moderate decrease in milk production, corresponding to a mean reduction of 0.37 units (95% CI: -0.68 to -0.05; $P < 0.025$) Table 6.

Table 6. ERBB2 Haplotype significantly associated with milk productivity in Sahiwal cattle.

	rs110735562	rs133031530	rs109763505	rs109017161	Rs109122971	Frequency	Difference (95% CI)	P-value
GTGTC	G	T	G	T	C	0.2562	1.22 (0.89 - 1.56)	<0.0001
GTGTT	G	T	G	T	T	0.1	-4 (-4.38 to -3.62)	<0.0001
GAATT	G	A	A	T	T	0.0562	-0.37 (-0.68 to -0.05)	<0.025

3.6. Haplotype not associated with milk productivity

Haplotype AAACC, AAGCT, GAGTT, AAATT, AAGCC, GAGTC and ATGTC were not associated with milk productivity P>0.05 (Table 7).

Table 7. ERBB2 Haplotype not associated with milk productivity.

Haplotype	rs110735562	rs133031530	rs109763505	rs109017161	rs109122971	Frequency	P-value
AAACC	A	A	A	C	C	0.1312	P>0.05
AAGCT	G	A	T	G	C	0.0438	P>0.05
GAGTT	A	G	A	A	T	0.025	P>0.05
AAATT	G	A	A	G	C	0.0188	P>0.05
AAGCC	G	G	A	G	C	0.0188	P>0.05
GAGTC	A	A	A	G	T	0.0188	P>0.05
ATGTC	G	A	A	G	T	0.0125	P>0.05

3.7. Heat shock protein 8 (HSPA8):

SNP Identification

A total of two SNPs were studied in HSPA8 gene. One SNP rs133632043 was in 3UTR and other SNP rs132976221 was intronic. One SNP of HSPA8 rs132976221 was in Hardy- Weinberg equilibrium (P> 0.05), while SNP rs136632043 was not in Hardy- Weinberg equilibrium (P< 0.05) (Table 8).

Table 8. Genotype and Allelic frequency and Hardy-Weinberg equilibrium test of HSPA8 SNPs in *Bos indicus* (Sahiwal breed).

Gene	Position	Ref SNP	Location	Genotype	N	Frequency	Allele	Frequency	Hardy-Weienberg Equilibrium X2 Test
HSPA8	Fwd	rs13663 2043	UTR3	TT	203	0.36	G	0.51	P>0.05
				TG	140	0.25	T	0.49	
				GG	217	0.39			
HSPA8	Fwd	rs13297 6221	Intron 3	AA	182	0.32	A	0.53	P>0.05
				AC	231	0.41	C	0.47	
				CC	147	0.26			

Variation in milk yield

For the SNPs of HSPA8 adjusted R2 value for the SNP rs136632043 was .520. It strongly indicates the SNP rs136632043 was effecting the milk yield. And the adjusted R2 value for the SNP rs132976221 was .371 (Table 9).

Table 9. Variation in milk yield predicted by Adjusted R^2 value.

rs#	R	R square	Adjusted R Square	Std. Error of the Estimates	P value
Rs132976221	.616	.379	.371	1.22	.000
Rs136632043	.725	.526	.520	1.06	.000

Single Locus Based analysis

Regarding the population of Bos indicus (Sahiwal breed) both of these SNPs of HSPA8 were significantly associated with milk productivity. For the SNP rs136632043, animals with the genotype GG were most productive than the animals with the genotype TG and TT respectively. For the SNP rs132976221animals with genotype AC show high milk yield CC and AA show low milk yield respectively (Table 10).

Table 10. Single locus based analysis.

Gene	Locus	Genotype	Milk Yield
HSPA8	rs136632043	TT TG GG	7.67 ± 0.32 8.99 ± 0.07 10.23 ± 0.07 P<0.05
HSPA8	rs132976221	AA AC CC	7.5 ± 0.33 9.77 ± 0.16 9.61 ± 0.07 P<0.05

(Mean of milk Yield ± standered error**).** P vale refers to the results of association each SNP with milk yield. Age was covariate.

4. Discussion

ERBB2 is a promising candidate gene for milk protein composition and regulation. Two synonymous SNPs, rs109017161 and rs110735562, although not altering the amino acid sequence directly (Raschia et al., 2018), may influence milk protein traits by affecting mRNA stability, splicing, translational accuracy, and protein expression (Walsh et al., 2020). Additionally, these SNPs could be in linkage disequilibrium (LD) with causal variants affecting milk production (Raschia et al., 2018).

The genotype frequencies of five ERBB2 SNPs in Sahiwal cattle show balanced polymorphism (allele frequencies 0.38–0.62), supporting their utility in association studies. Four SNPs deviate from Hardy-Weinberg equilibrium (HWE; $P < 0.05$), possibly due to selection or population structure, while two conform to HWE, indicating stable allele distributions (Li et al., 2019). Such patterns align with findings in Bos indicus breeds under dairy selection.

All five ERBB2 SNPs significantly explain milk yield variation ($P = 0.000$), with rs133031530 showing the strongest association (adjusted $R^2 = 0.814$), surpassing other SNPs. This highlights its potential as a primary marker for selection. Strong

ISSN 0032-7786

LD (D' ≥ 0.89) among rs110735562, rs133031530, and rs109763505 suggests these SNPs tag a shared causal variant, reducing redundant genotyping (Mustafa et al., 2018). Weaker LD for rs109017161 and rs109122971 implies independent effects or recombination hotspots, consistent with previous Holstein studies linking LD patterns to milk yield haplotypes (De Roos et al., 2008).

SNPs within the HSPA8 gene were also significantly associated with milk yield, reinforcing the gene's role in lactation. Linear regression and adjusted R^2 values remain effective for genetic association analyses (Wang et al., 2016). These findings have practical implications for dairy breeding, though further validation and mechanistic studies are needed.

Locus-based analysis revealed that the CC genotype of rs109017161 and AG/GG genotypes of rs110735562 were associated with higher milk yield, consistent with reports in Chinese Holsteins linking these SNPs with milk protein and fat yield (Li et al., 2019). These results support the predictive value of these markers for selective breeding (Yang et al., 2021).

Three intronic SNPs (rs110735562, rs109122971, rs109763505) may affect mRNA splicing, potentially generating alternative isoforms or regulatory RNAs that influence gene expression (García-Ruiz et al., 2025; Jacobs & Elmer, 2021). In this study, certain genotypes at these loci correlated with higher milk production, consistent with previous findings in Canadian Holsteins and Sahiwal cattle (Do et al., 2017; Ahlawat et al., 2020). Such conserved genetic effects across breeds suggest a fundamental role in milk protein regulation.

Heat shock protein 8 (HSPA8):

SNPs in the 3' untranslated region (3'UTR) have been reported to affect post-transcriptional regulation and mRNA stability (Ma et al., 2021). In the present study, two SNPs (rs136632043 and rs132976221) in the 3'UTR were investigated for their association with milk productivity in Bos indicus (Sahiwal breed) population. The results showed that the G polymorphism of rs136632043 was significantly associated with higher milk yield, while cows carrying the TG and TT genotypes had lower milk productivity. A previous study also reported the significant association of this SNP with five milk yield traits (milk yield, fat yield, protein yield, fat percentage, and milk percentage) (Li et al., 2019). These findings suggest that rs136632043 may serve as a potential genetic marker for selecting high milk-producing dams.

In the present study, the effects of SNP rs132976221 located in the HSPA8 gene on milk yield were investigated. The results showed that animals carrying the AC and CC genotypes had higher milk yield compared to those carrying the AA genotype. Previous studies have also reported the association of HSPA8 gene with milk yield in different cattle breeds, including Chinese Holstein and Sahiwal (Silpa et al.,

2021). These results suggest that SNP rs132976221 may be a promising genetic marker for animal selection with high milk yield. However, further studies are needed to validate these findings and to determine the underlying mechanisms of this association. The focus of this study was to investigate the impact of SNPs on milk production in the endemic Zebu cattle breed (Sahiwal). While the present study has revealed that these SNPs are significantly associated with milk production, previous research has already established their effects on desirable milk traits. Therefore, based on these collective findings, it can be inferred that these SNPs may serve as promising candidates for marker-assisted selection programs.

5. Conclusion:

This study stands as a pioneering investigation into the genetic determinants of milk productivity in the endemic Zebu cattle (Sahiwal breed), revealing significant associations between specific SNPs and haplotypes within the ERBB2 and HSPA8 genes and milk yield. Through detailed genotype, allele frequency, and haplotype analyses, the research highlights that certain genetic variants—such as rs133031530 and the "GTGTC" haplotype in ERBB2, as well as the GG genotype of rs136632043 and AC genotype of rs132976221 in HSPA8—are strongly correlated with enhanced milk production in Sahiwal cattle. Conversely, other haplotypes like "GTGTT" and "GAATT" were linked to reduced yields. These findings provide foundational knowledge of the mechanisms by which these genetic markers influence milk productivity and underscore their potential utility in future selection programs. As a crucial initial step, this research lays the groundwork for the development of informed breeding strategies aimed at improving milk yield and, ultimately, increasing the profitability of the dairy industry at the national level.

Funding: This research work was funded by Office of Research Innovation and Commercialization (ORIC) University of Sargodha, Sargodha, project number (UOS/ORIC/2016/57).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

Acknowledgments: The author is thankful to the management of Khizrabad Farm, Sargodha, Punjab, Pakistan for their cooperation in providing the milk samples and their milking information. I would like to extend my gratitude to the University of Sargodha for their support and funding of project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Ahlawat, S., Vijh, R. K., Sharma, A., Sharma, U., Girdhar, Y., Kaur, M., and Arora, R. (2020). Milk somatic cell derived transcriptome analysis identifies regulatory genes and pathways during lactation in Indian Sahiwal cattle (Bos indicus). Molecular Biology Reports, 47: 7029-7038.

Al-Thuwaini, T. M., Al-Shuhaib, M. B. S., and Hussein, Z. M. (2020). A novel T177P missense variant in the HSPA8 gene associated with the low tolerance of Awassi sheep to heat stress. Tropical Animal Health and Production, 52(5): 2405-2416.

Anderson, S. M., Rudolph, M. C., McManaman, J. L., and Neville, M. C. (2007). Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis! Breast Cancer Research, 9(1): 1-14.

Daniel, J. B., Friggens, N. C., Chapoutot, P., Van Laar, H., and Sauvant, D. (2016). Milk yield and milk composition responses to change in predicted net energy and metabolizable protein: a meta-analysis. Animal, 10(12): 1975-1985.

De Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics*, ***179***(3): 1503-1512.

Do, D. N., Bissonnette, N., Lacasse, P., Miglior, F., Sargolzaei, M., Zhao, X., and Ibeagha-Awemu, E. M. (2017). Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. Journal of Dairy Science, 100(3): 1955-1970.

Elliott, R. M., Lloyd, R. E., Fazeli, A., Sostaric, E., Georgiou, A. S., Satake, N., and Holt, W. V. (2009). Effects of HSPA8, an evolutionarily conserved oviductal protein, on boar and bull spermatozoa. Reproduction, 137(2): 191-203.

Križanac, Ana-Marija, Christian Reimer, Johannes Heise, Zengting Liu, Jennie E. Pryce, Jörn Bennewitz, Georg Thaller, Clemens Falker-Gieske, and Jens Tetens. "Sequence-based GWAS in 180,000 German Holstein cattle reveals new candidate variants for milk production traits." Genetics Selection Evolution 57, no. 1 (2025): 3.

Ghoreishifar, S. M., Eriksson, S., Johansson, A. M., Khansefid, M., Moghaddaszadeh-Ahrabi, S., Parna, N., and Javanmard, A. (2020). Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. Genetics Selection Evolution, 52(1): 1-15.

https://www.snpstats.net/start.html

Gross, J. J. (2022). Limiting factors for milk production in dairy cows: perspectives from physiology and nutrition. Journal of animal science, 100(3), skac044.

Jacobs, A., and Elmer, K. R. (2021). Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. Molecular Ecology, 30(20): 4955-4969.

Kumaresan, A., and Srivastava, A. K. (2022). Reproduction Efficiency in Dairy Bovine: Trends and Targets. In Current Concepts in Bovine Reproduction (pp. 7-24). Springer, Singapore.

Lee, X., Shao, Z., Sheng, G., Pepinsky, B., and Mi, S. (2014). LINGO-1 regulates oligodendrocyte differentiation by inhibiting ErbB2 translocation and activation in lipid rafts. Molecular and Cellular Neuroscience, 60: 36-42.

Li, C., Cai, W., Zhou, C., Yin, H., Zhang, Z., Loor, J. J., and Zhang, S. (2016). RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. Scientific reports, 6(1): 1-11.

Li, C., Wang, M., Cai, W., Liu, S., Zhou, C., Yin, H., and Zhang, S. (2019). Genetic analyses confirm SNPs in HSPA8 and ERBB2 are associated with milk protein concentration in Chinese Holstein cattle. Genes, 10(2): 104.

Liang, Y., Gao, Q., Zhang, Q., Arbab, A. A. I., Li, M., Yang, Z., and Mao, Y. (2020). Polymorphisms of the ACSL1 gene influence milk production traits and somatic cell score in Chinese Holstein cows. Animals, 10(12): 2282.

Liu, J., Sun, Y., Yang, C., Zhang, Y., Jiang, Q., Huang, J., and Wang, C. (2016). Functional SNPs of INCENP affect semen quality by alternative splicing mode and binding affinity with the target bta-mir-378 in Chinese Holstein bulls. PLoS One, 11(9): e0162730.

Ma, Y., Khan, M. Z., Xiao, J., Alugongo, G. M., Chen, X., Chen, T., and Cao, Z. (2021). Genetic markers associated with milk production traits in dairy cattle. Agriculture, 11(10): 1018.

Metin Kiyici, J., Arslan, K., Akyuz, B., Kaliber, M., Aksel, E. G., and Çinar, M. U. (2019). Relationships between polymorphisms of growth hormone, leptin and myogenic factor 5 genes with some milk yield traits in Holstein dairy cows. International journal of dairy technology, 72(1): 1-7.

Mustafa, H., Ahmad, N., Heather, H. J., Eui-Soo, K., Khan, W. A., Ajmal, A., and Sonstegard, T. S. (2018). Whole genome study of linkage disequilibrium in Sahiwal cattle. *South African Journal of Animal Science*, ***48***(2): 353-360.

Ogorevc, J., Kunej, T., Razpet, A., and Dovc, P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. Animal genetics, 40(6): 832-851.

Paten, A. M., Duncan, E. J., Pain, S. J., Peterson, S. W., Kenyon, P. R., Blair, H. T., and Dearden, P. K. (2015). Functional development of the adult ovine mammary gland—insights from gene expression profiling. BMC genomics, **16**(1): 1-13.

Pizarro Inostroza, M. G., Navas González, F. J., Landi, V., León Jurado, J. M., Delgado Bermejo, J. V., Fernández Álvarez, J., & Martínez Martínez, M. D. A. (2020). Software-automatized individual lactation model fitting, peak and persistence and Bayesian criteria comparison for milk yield genetic studies in Murciano-Granadina goats. Mathematics, **8**(9): 1505.

Pokorska, J., Kułaj, D., Dusza, M., Żychlińska-Buczek, J., and Makulska, J. (2016). New rapid method of DNA isolation from milk somatic cells. Animal biotechnology, 27(2): 113-117.

Rajawat, D., Ghildiyal, K., Nayak, S. S., Sharma, A., Parida, S., Kumar, S., and Panigrahi, M. (2024). Genome-wide mining of diversity and evolutionary signatures revealed selective hotspots in Indian Sahiwal cattle. *Gene*, ***901:*** 148178.

García-Ruiz, S., Zhang, D., Gustavsson, E. K., Rocamora-Perez, G., Grant-Peters, M., Fairbrother-Browne, A., & Ryten, M. (2025). Splicing accuracy varies across human introns, tissues, age and disease. Nature Communications, 16(1), 1068.

Shah, S., Bonder, M. J., Marioni, R. E., Zhu, Z., McRae, A. F., Zhernakova, A., & Visscher, P. M. (2015). Improving phenotypic prediction by combining genetic and epigenetic associations. The American Journal of Human Genetics, 97(1): 75-85.

Silpa, M. V., König, S., Sejian, V., Malik, P. K., Nair, M. R. R., Fonseca, V. F., and Bhatta, R. (2021). Climate-resilient dairy cattle production: Applications of genomic tools and statistical models. Frontiers in Veterinary Science, 8: 625189.

Solé, Xavier, Elisabet Guinó, Joan Valls, Raquel Iniesta, and Víctor Moreno. "SNPStats: a web tool for the analysis of association studies." Bioinformatics 22, no. 15 (2006): 1928-1929.

Walsh, I. M., Bowman, M. A., Soto Santarriaga, I. F., Rodriguez, A., and Clark, P. L. (2020). Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. Proceedings of the National Academy of Sciences, 117(7): 3528-3534.

Wang, Shi-Bo, Jian-Ying Feng, Wen-Long Ren, Bo Huang, Ling Zhou, Yang-Jun Wen, Jin Zhang, Jim M. Dunwell, Shizhong Xu, and Yuan-Ming Zhang. "Improving power and accuracy of genome-wide association studies via a

ISSN 0032-7786

multi-locus mixed linear model methodology." Scientific reports 6, no. 1 (2016): 1-10.

Yang, Z., Lian, Z., Liu, G., Deng, M., Sun, B., Guo, Y., and Li, Y. (2021). Identification of genetic markers associated with milk production traits in Chinese Holstein cattle based on post genome-wide association studies. Animal Biotechnology, 32(1): 67-76.

Yong, Y. O. N. G., and He, L. (2005). SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. Cell research, 15(2): and 97.