

## Association of genetic variants in *STAT5B* gene with milk performance and mastitis related traits in dairy cattle

Nawab Ali<sup>1,2,3</sup>, Sadaf Niaz<sup>2</sup>, Irfan Khattak<sup>1</sup>, Naimat U. Khan<sup>1</sup>, Hanpeng Luo<sup>3</sup>, Hazir Muhammad<sup>4</sup>, Tariq Ali<sup>5</sup>, Ying Yu<sup>3</sup>, Yachun Wang<sup>3\*</sup>, and Tahir Usman<sup>1\*</sup>

<sup>1</sup>College of Veterinary Sciences and Animal Husbandry, Abdul Wali Khan University Mardan, Pakistan

<sup>2</sup>Department of Zoology, Abdul Wali Khan University Mardan, Pakistan

<sup>3</sup>College of Animal Sciences and Technology, China Agriculture University Beijing, China

<sup>4</sup>Department of Biotechnology, Abdul Wali Khan University Mardan, Pakistan

<sup>5</sup>Center for Microbiology and Biotechnology, Veterinary Research Institute, Peshawar, Pakistan

---

ALI, N., S. NIAZ, I. KHATTAK, N. U. KHAN, H. LUO, H. MUHAMMAD, T. ALI, Y. YU, Y. WANG, T. USMAN: Association of genetic variants in *STAT5B* gene with milk performance and mastitis related traits in dairy cattle. Vet. arhiv 91, 459-471, 2021.

### ABSTRACT

Udder performance and health are important traits in dairy cattle worldwide. The present study aimed to explore the association of single nucleotide polymorphisms (SNPs) in the *STAT5B* gene in dairy cattle with milk performance and mastitis related traits. The study included 201 cows of three pure breeds (*i.e.* Holstein Friesian, Jersey and Achai) and two crossbred cattle at four established dairy farms in Khyber Pakhtunkhwa, Pakistan. The milk samples were analysed for somatic cell count (SCC) and milk composition (*i.e.* fat, protein and lactose percentages). The generalized linear model was deployed for association analysis using SAS. The pool DNA sequencing showed four (three synonymous and a 3' UTR) SNPs in *STAT5B*. These SNPs were further validated in all DNA samples using SNaPshot assay. The breed-wise analysis showed that most of the SNPs were consistent with Hardy-Weinberg equilibrium ( $P > 0.05$ ). The association analysis revealed a significantly higher protein percentage in TT genotype and lower SCC in CC genotypes of SNP 1 (exon 2, C>T), whereas in SNP 2 (exon 16, T>C) the TT genotypes revealed significantly lower SCC and SCS compared with other genotypes ( $P < 0.05$ ). SNP 4 (3' UTR, C>T) showed significantly lower SCC and frequency of clinical mastitis in the heterozygous (CT) genotype compared to the homozygous genotypes. The *in silico* predictions revealed changes in the RNA secondary structure for SNP 2 and SNP 3. The study suggests that *STAT5B* should be considered as a candidate gene, and the variants identified as useful genetic markers for improved milk composition and udder health.

**Key words:** dairy cattle; mastitis; milk performance; *STAT5B*; SNPs

---

### Introduction

Mastitis is an important inflammatory disease of the udders in lactating cows that causes major economic losses, *i.e.* 80-180 €/cow/year in western

European production systems (SEEGERS et al., 2003). Mastitis results in higher somatic cell count (SCC), thus lowering milk yield due to its profound

---

\*Corresponding author:

Assoc. Prof., Dr. Tahir Usman, College of Veterinary Science and Animal Husbandry, Abdul Wali Khan University Mardan, Pakistan, Phone: +92 313 092 0177; E-mail: tahircau@gmail.com; tahirusman@awkum.edu.pk

Prof. Dr. Yachun Wang, College of Animal Sciences and Technology, China Agriculture University Beijing, China, Phone: +86 158 0159 5851; E-mail: wangyachun@cau.edu.cn

negative impact on the quality and quantity of milk (HAGNESTAM-NIELSEN et al., 2009). Due to the highly positive genetic correlation between clinical mastitis and SCC, lower SCC in cattle is used as a useful indirect selection criteria for resistance against mastitis (RUPP and BOICHARD, 1999; HINRICHS et al., 2005). Different methods, including therapeutic and prophylactic measures, are used to reduce the incidence of mastitis, but each one has certain limitations. Marker-assisted selection is one of the most effective methods of selection of dairy cows with improved natural genetic resistance of the udders to various pathogens (DETILLEUX 2002; PANT et al., 2007). Studies have revealed that genetic variants in immunity-related genes partially mediate the variations in the inflammatory and immune responses to pathogens (LEYVA-BACA et al., 2008). Signal transducer and activator of transcription 5 (*STAT5*) protein is a member of a family of cytoplasmic transcription factors that contribute to signal transduction by cytokines, hormones, and growth factors (BENEKLI et al., 2003) In mammals, seven *STAT* genes have been identified, i.e. *STAT1*, *STAT2*, *STAT3*, *STAT4*, *STAT5A*, *STAT5B* and *STAT6* (COPELAND et al., 1995).

*STAT5B* genes play a key role in the physiological responses associated with cytokines, growth hormone and prolactin in mice (TEGLUND et al., 1998). Various studies have shown that SNPs *STAT5B* have a significant association with milk performance and mastitis resistance traits, such as SCC (HE et al., 2011; USMAN et al., 2015). The present study aimed to investigate the allele and genotype frequencies of SNPs in *STAT5B* genes, and their association with milk production, milk composition and resistance against mastitis in three dairy breeds of cattle in Khyber Pakhtunkhwa, Pakistan.

## Materials and methods

*Study population and phenotypic data.* Milk and blood samples from 214 lactating cows with apparently healthy udder with different parities (ranging from one to six) were collected from four well-established dairy farms in the subtropical conditions of Khyber Pakhtunkhwa, Pakistan.

The cows on these farms were fed a standard lactation diet according to their lactation status, as recommended by the Livestock and Dairy Development Department, Pakistan. The cows belonged to three pure breeds, including two high milk-yielding exotic breeds (i.e. Holstein Friesian and Jersey), and an indigenous dairy breed called Achai (A), with a few crossbred cows in addition. The government dairy farms were selected for sampling due to the availability of an authentic data recording system, and a private dairy farm with well-recorded data on production, reproduction and diagnosis/treatment. Secondary data for each cow, i.e. the year of calving, season of calving, the number of parities, and frequency of clinical mastitis (FOM), were obtained from the primary data of stock registers, and then arranged in Microsoft Excel. Out of 214 cows, 201 cows with a complete record in all respects were used for final association analysis. The study was approved by the ethical committee of Abdul Wali Khan University, Mardan, Pakistan.

*Samples collection.* Milk samples (30 mL) were collected in sterilized bottles from all the teats of each cow, and taken in an ice box to the laboratory for analysis of milk composition and somatic cell count (SCC). Milk samples from each cow were analysed for milk composition, such as fat, protein, and lactose percentages, using a milk analyser (Ekomilk). The SCC for each milk sample was measured by the direct microscopic somatic cell count method, using Giemsa stain, following the guidelines of the FDA Laboratory Quality Assurance Team. On the basis of the SCC, the cows were classified into two groups, group 1 included healthy cows (SCC  $\leq$ 200,000 cells/mL of milk) while group 2 included cows with subclinical mastitis (SCC >200,000 cells/mL of milk).

The following formula was then used to convert the SCC of each sample to the somatic cell score (SCS):

$SCS = \log_2 [SCC/100]+3$ , the unit of SCC is 1,000 cells/mL (USMAN et al., 2014)

Blood samples (5 mL) from each cow were collected either from the caudal vein or the jugular vein, and were kept at -20 °C until the next experiment.

**DNA extraction from blood.** Genomic DNA was isolated from whole blood from each sample using ay TianGen blood DNA kit. The extracted DNA was confirmed by gel electrophoresis and kept at -20 °C. For further experiments, the DNA samples were taken to a Chinese agriculture university in Beijing (China). The quality and quantity of each DNA sample was measured by a Nanodrop™ ND-2000c Spectrophotometer (Thermo Scientific, Inc.).

**Primer designing.** A total of 20 primer pairs were designed for the *STAT5B* gene on the basis of the reference sequence of NCBI, using primer 3 and Oligo (version 6.41) software. The primer pairs included primers for 5' upstream (covering 2 kb promoter region), all exons, and their flanking introns, including the flanking 3' UTR. The list of primer pairs that revealed polymorphism in *STAT5B* is shown in Table 1.

Table 1. List of primers of the *STAT5B* gene that showed polymorphisms

Tag	Primer	Amplicon length (bp)	Annealing temperature (°C)
<i>STAT5B</i> E2 F	5'-AAAAACCCAGCGCAAACCAC-3'	227	55
<i>STAT5B</i> E2 R	5'-ACTTCGATATGCTTGTGCCTCA-3'		
<i>STAT5B</i> E16 F	5'-CATGCTTTTGTATGGGGAAT-3'	722	52
<i>STAT5B</i> E16 R	5'-AAAGCCAGGATTCAAACAGA-3'		
<i>STAT5B</i> E17+18 F	5'-CATGCCACACTGTAAGACCA-3'	724	57
<i>STAT5B</i> E17+18 R	5'-AGACTAGCCCTGGTGTCCCTT-3'		
<i>STATB</i> E19+ 3' UTR F	5'-TGGTCTCTGCTGCTCCTTGTT-3'	397	59
<i>STATB</i> E19+3' UTR R	5'-CACTCACACACTCACACATCCTT-3'		

Table 2. Information on the SNPs studied in the *STAT5B* gene

SNP #	Gene	SNP location	Base Position	Variable type	Variant ID	Consequence type
SNP 1	<i>STAT5B</i>	Exon 2	42984480	C>T *1	rs135194045	synonymous variant *5
SNP 2	<i>STAT5B</i>	Exon 16	42965083	T>C *2	rs43706496	synonymous variant *6
SNP 3	<i>STAT5B</i>	Exon 18	42961074	C>A *3	rs209670220	synonymous variant *7
SNP 4	<i>STAT5B</i>	3' UTR	42960341	C>T *4	rs382802804	3' UTR variant

\*1 Analogue to rs135194045 Chr19: g.42984480 G>A + strand according to Bos\_taurus\_UMD\_3.1.1; \*2 Analogue to rs43706496 Chr19: g.42965083 A>G + strand according to Bos\_taurus\_UMD\_3.1.1; \*3 Analogue to rs209670220 Chr19: g.42961074 G>T + strand according to Bos\_taurus\_UMD\_3.1.1; \*4 Analogue to rs382802804 Chr19: g.42960341 G>A + strand according to Bos\_taurus\_UMD\_3.1.1; \*5 codon GGC or GGT codes for the same amino acid (Glycine); \*6 codon GAT or GAC codes for the same amino acid (Aspartic acid); \*7 codon GGC or GGA codes for the same amino acid (Glycine).

**DNA pool and PCR.** Thirty good quality (≥50 ng/μL) DNA samples (10 samples of each breed) of equal amounts (2 μL) were selected randomly to create a DNA pool. PCR was carried out for each primer pair in a final volume of 25 μL (22 μL of ready to use master mix, 1 μL each of forward and reverse primer and 1 μL of pooled DNA). PCR conditions included denaturation at 98 °C for 2 min, followed by 35 cycles including 98 °C for 30 s, respective annealing temperature for 30 s, elongation at 72 °C for 30 s, and final extension at 72 °C for 5 min.

**Sequencing and SNPs validation.** The PCR products were sequenced by Tsingke Biological Technology, Beijing (China). The sequencing results were analysed by a CodonCode Aligner (V 8.0.2) for discovery of SNPs. Four SNPs were identified in the *STAT5B* gene and were further validated in all 201 DNA samples using SNaPshot assay (single base extension and termination method) by Generay Biotechnology, Beijing, China (Table 2 and Fig. 1).

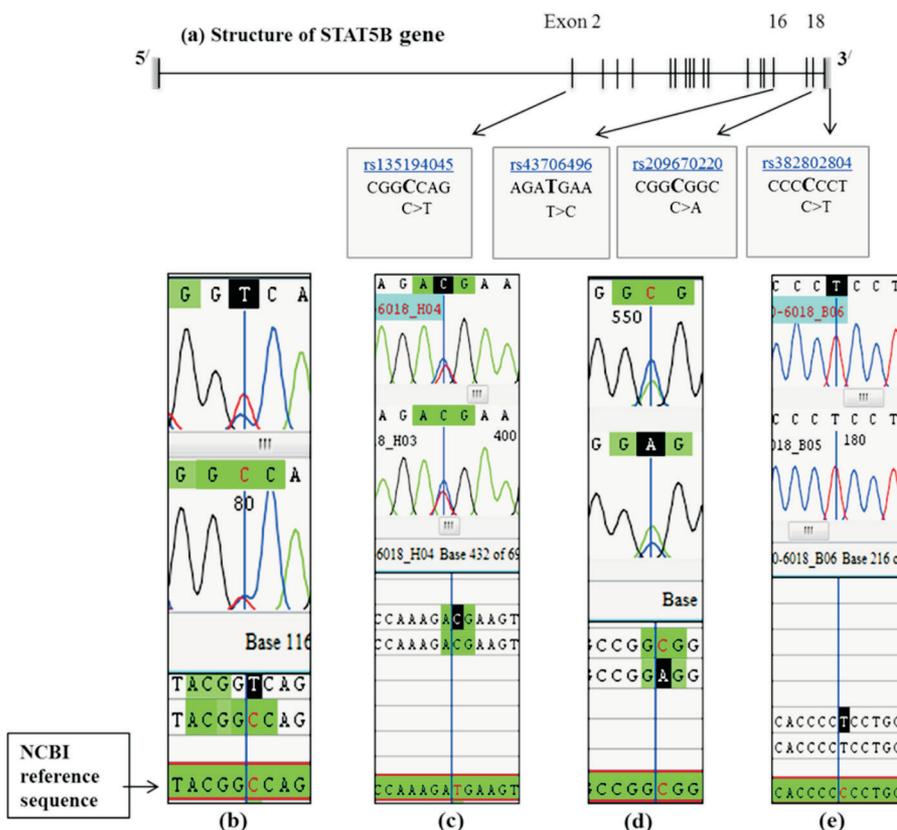


Fig. 1. (a) The structure of the *STAT5B* gene with the positions of four identified SNPs; (b), (c), (d) and (e), shows the sequence alignment (forward & reverse strands) and chromatogram results of CodonCode Aligner (V 8.0.2) for SNP 1, SNP 2, SNP 3, and SNP 4, respectively \*The grey regions represents 5' and 3' UTR, the black line represents exons while intervals represents introns in structure of *STAT5B* gene.

Table 3. Allele frequencies, genotype frequencies and HWE ( ) test of SNPs in *STAT5B* in the overall studied population

SNP #	Variable type	Genotype frequency			Allele frequency			Test (P)*
		Genotype	n	Genotype freq.	Allele	n	Allele freq.	
SNP 1	C>T	CC	107	0.53	C	279	0.69	<0.01
		CT	65	0.32	T	123	0.31	
		TT	29	0.15				
SNP 2	T>C	TT	51	0.25	T	184	0.46	<0.05
		TC	82	0.41	C	218	0.54	
		CC	68	0.34				
SNP 3	C>A	CC	114	0.57	C	281	0.70	<0.01
		CA	53	0.26	A	121	0.30	
		AA	34	0.17				
SNP 4	C>T	CC	56	0.28	C	160	0.40	<0.01
		CT	48	0.24	T	242	0.60	
		TT	97	0.48				

\* P<0.05, not consistent with HWE ( ) test

Table 4. Breed-wise genotype and allele frequencies and HWE () test of SNPs in *STAT5B*

SNP #	SNP 1				SNP 2				SNP 3				SNP 4								
	Genotype		Allele		Genotype		Allele		Genotype		Allele		Genotype		Allele						
HF (n = 104)	CC	0.74	0.24	0.02	0.86	0.14	0.37	0.5	0.13	0.63	0.37	0.77	0.19	0.04	0.14	0.86	0.42	0.31	0.27	0.58	0.42
	CT	0.24	0.02																		
Jersey (n = 42)	CC	0.55	0.43	0.02	0.76	0.24	0.21	0.48	0.31	0.45	0.55	0.65	0.33	0.02	0.19	0.81	0.21	0.31	0.48	0.37	0.63
	CT	0.43	0.02																		
Achai (n = 29)	CC	0	0.17	0.83	0.09	0.91	0	0.03	0.97	0.02	0.98	0	0.14	0.86	0.93	0.07	0	0	1	0	1
	CT	0.17	0.83																		
J × A (n = 19)	CC	0.05	0.84	0.11	0.47	0.53	0.05	0.26	0.69	0.18	0.82	0.05	0.74	0.21	0.58	0.42	0	0.05	0.95	0.03	0.97
	CT	0.84	0.11																		
HF × J (n = 07)	CC	0.86	0.14	0	0.93	0.07	0.29	0.57	0.14	0.57	0.43	0.86	0.14	0	0.07	0.93	0.42	0.29	0.29	0.57	0.43
	CT	0.14	0																		

\* HF means Holstein Friesian breed, \*J × A means Jersey cross Achai breed, \*HF × J means Holstein Friesian cross Jersey breed; \*P>0.05, consistent with HWE () test

Table 5. Overall association of SNPs in the *STAT5B* gene with milk performance traits and mastitis-related traits

SNP #	Genotypes	AMY	Fats %	Protein %	Lactose %	SCC	SCS	FOM
SNP 1	CC	3452.65 ± 158.15	3.19 ± 0.32	<b>3.25 ± 0.05<sup>ab</sup></b>	4.84 ± 0.09	<b>154.91 ± 56.13<sup>a</sup></b>	2.94 ± 0.25	0.30 ± 0.09
	CT	3521.05 ± 153.57	3.48 ± 0.31	<b>3.20 ± 0.05<sup>a</sup></b>	4.77 ± 0.09	<b>302.8 ± 54.5<sup>c</sup></b>	3.39 ± 0.24	0.36 ± 0.09
	TT	3496.02 ± 278.93	4.17 ± 0.57	<b>3.49 ± 0.10<sup>b</sup></b>	4.98 ± 0.16	<b>269.83 ± 98.10<sup>ac</sup></b>	3.70 ± 0.44	0.18 ± 0.17
	P Value	0.92	0.35	<b>0.03</b>	0.44	<b>0.04</b>	0.16	0.56
SNP 2	TT	3380.71 ± 182.34	3.16 ± 0.37	3.260 ± 0.06	4.744 ± 0.10	<b>137.22 ± 64.86<sup>b</sup></b>	<b>2.81 ± 0.29<sup>b</sup></b>	0.32 ± 0.11
	TC	3343.70 ± 150.78	3.33 ± 0.31	3.322 ± 0.05	4.851 ± 0.09	<b>156.09 ± 53.64<sup>b</sup></b>	<b>2.98 ± 0.24<sup>ab</sup></b>	0.30 ± 0.09
	CC	3643.15 ± 151.29	3.78 ± 0.30	3.26 ± 0.05	4.88 ± 0.09	<b>331.68 ± 53.82<sup>a</sup></b>	<b>3.66 ± 0.24<sup>a</sup></b>	0.29 ± 0.09
	P Value	0.29	0.37	0.49	0.45	<b>0.02</b>	<b>0.03</b>	0.96
SNP 3	CC	3459.17 ± 156.95	3.41 ± 0.32	3.25 ± 0.05	4.84 ± 0.09	148.64 ± 55.85	2.93 ± 0.25	0.316 ± 0.09
	CA	3474.11 ± 161.09	3.65 ± 0.33	3.24 ± 0.06	4.84 ± 0.09	279.06 ± 57.32	3.42 ± 0.25	0.36 ± 0.09
	AA	3553.04 ± 242.11	3.48 ± 0.50	3.39 ± 0.08	4.87 ± 0.14	324.72 ± 86.15	3.67 ± 0.38	0.19 ± 0.14
	P Value	0.95	0.79	0.30	0.97	0.08	0.14	0.57
SNP 4	CC	3412.59 ± 179.97	3.54 ± 0.37	3.33 ± 0.06	4.90 ± 0.10	<b>176.23 ± 63.83<sup>ab</sup></b>	3.22 ± 0.28	<b>0.41 ± 0.11<sup>a</sup></b>
	CT	3334.22 ± 178.26	3.11 ± 0.36	3.27 ± 0.06	4.84 ± 0.10	<b>104.32 ± 63.22<sup>a</sup></b>	2.81 ± 0.28	<b>0.12 ± 0.10<sup>b</sup></b>
	TT	3554.60 ± 128.19	3.62 ± 0.26	3.27 ± 0.04	4.83 ± 0.07	<b>288.75 ± 45.46<sup>b</sup></b>	3.42 ± 0.20	<b>0.32 ± 0.07<sup>ab</sup></b>
	P Value	0.49	0.37	0.55	0.78	<b>0.02</b>	0.12	<b>0.03</b>

\* Data is displayed as LS means ± SE and P value; \* P values with bold characters are significantly different (P<0.05); \* AMY = Annual milk yield, SCC = Somatic cell count (x10<sup>3</sup> cells/mL), SCS = Somatic cell score, FOM = Frequency of clinical mastitis (per year), \* The Bonferroni *t* test was used for pair-wise comparison in the study; \* The values with no common superscripts are significantly different from each other (P<0.05)

*Gene frequency and Hardy-Weinberg Equilibrium (HWE) test.* Allele frequency, genotype frequency, and HWE () test were calculated in the overall studied population for each SNP in Microsoft Excel (Table 3). Genotype frequency and allele frequency were also calculated for each breed and their crosses (Table 4).

*Linkage disequilibrium (LD) and haplotypes.* LD and haplotypes were analysed by Haploview (V 4.2). Haplogroups for the *STAT5B* gene were constructed on the basis of the identified haplotypes (Fig. 2).

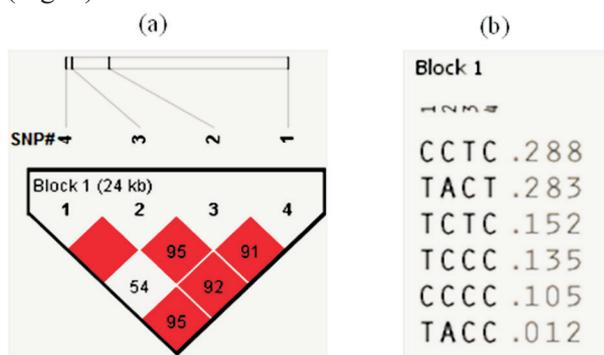


Fig. 2. (a) Linkage Disequilibrium plot of the SNPs in *STAT5B* ( $R^2$  value of 0.31 for SNP 1 & 2, 0.84 for SNP 1 & 3, 0.26 for SNP 1 & 4, 0.33 for SNP 2 & 3, 0.23 for SNP 2 & 4, 0.28 for SNP 3 & 4) (b) Haplotype blocks with expected frequencies

*Association analysis.* The association analysis for each SNP and haplogroups was conducted using the following generalized linear model in SAS studio (V 9.4):

Where,  $Y_{ijklmn}$  is the phenotypic value of annual milk yield (AMY), fat, protein, and lactose percentages, SCC, SCS and FOM,  $\mu$  is population mean,  $G_i$  is the fixed effect of the  $i^{th}$  genotype,  $B_j$  is the fixed effect of  $j^{th}$  breed ( $j = 1, 2, \dots, 5$ ),  $P_k$  is the fixed effect of  $k^{th}$  parity ( $k = 1, 2, \dots, 6$ ),  $H_l$  is the fixed effect of  $l^{th}$  herd ( $l = 1, 2, 3, 4$ ) and is the random error effect of each observation.

*In silico prediction of RNA secondary structure and 3' UTR confirmation.* To find out whether the identified SNPs can cause a change in the secondary mRNA of the gene, *in silico* predictions were performed to calculate the minimum free energy (MFE) by a RNA fold webserver (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>).

For SNP 4, the 3' UTR sequence of the *STAT5B* gene was retrieved from an online UTR database (<http://utrdb.ba.itb.cnr.it/search>).

## Results

*Gene frequency and HWE test.* The observed genotype frequencies of all four SNPs in the overall population studied were found to be inconsistent with the HWE test ( $P < 0.05$ ), and most of them were found consistent in breed-wise analysis. SNP 1 in exon 2 (rs135194045 C>T) showed the lowest (0.15) genotype frequency for the TT genotype in the studied population (Table 3). The breed-wise genotype frequency revealed the highest genotype frequency of the CC genotype and the lowest for the TT genotype in Holstein-Friesian, Jersey and their crosses (Table 4). On the other hand, the Achai breed showed very different genotype frequency from other breeds, with TT as the most prevalent genotype in the studied population, and the complete absence of CC genotype.

SNP 2 in exon 16 (rs43706496 T>C) revealed the higher frequency of the T allele (0.63) in the exotic breeds and their crosses, compared to the indigenous Achai cattle (0.02). Similarly, the TT and TC genotype showed higher frequency in the exotic HF and Jersey breeds compared to the indigenous Achai cattle that were noted as having complete absence of the TT genotype.

SNP 3 in exon 18 (rs209670220 C>A) showed a higher frequency of allele C in the exotic breeds (HF and Jersey) compared to allele A, and vice versa for indigenous Achai cattle breed. The genotype frequency also showed a marked difference in the exotic breeds and the Achai breed. The CC genotype was recorded with higher frequency in the exotic breeds and their crosses, whereas it was completely absent in the Achai breed.

SNP 4 in 3' UTR (rs382802804 C>T) showed the complete absence of allele C, genotypes CC and CT in the Achai breed, whereas these allele and genotypes were recorded in higher frequency in the exotic cattle breeds in this study (Tables 3 and 4).

Table 6. Breed-wise association analysis of SNPs in *STAT5B* with studied traits

SNP #	Trait	Genotypes	HF (n = 104)	Jersey (n = 42)	Achai (n = 29)
SNP 1	Protein%	CC	3.09 ± 0.04	3.09 ± 0.07	-
		CT	3.02 ± 0.08	3.09 ± 0.07	<b>2.87 ± 0.15<sup>A</sup></b>
		TT	3.17 ± 0.26	3.31 ± 0.26	<b>3.39 ± 0.07<sup>B</sup></b>
		P Value	0.59	0.69	<b>0.007</b>
SNP 2	SCC	TT	<b>155.43 ± 63.99<sup>A</sup></b>	645.93 ± 126.47	-
		TC	<b>231.92 ± 54.50<sup>A</sup></b>	531.34 ± 109.68	36.54 ± 75.79
		CC	<b>637.95 ± 104.67<sup>B</sup></b>	516.79 ± 119.34	78.38 ± 14.77
		P Value	<b>0.0006</b>	0.67	0.59
SNP 2	SCS	TT	<b>3.00 ± 0.26<sup>a</sup></b>	4.60 ± 0.51	-
		TC	<b>3.50 ± 0.22<sup>ab</sup></b>	4.22 ± 0.45	1.93 ± 1.52
		CC	<b>4.26 ± 0.43<sup>b</sup></b>	4.89 ± 0.49	1.99 ± 0.29
		P Value	<b>0.04</b>	0.45	0.96
SNP 3	SCC	CC	211.13 ± 47.91	548.53 ± 91.06	-
		CA	413.07 ± 88.75	592.83 ± 123.70	87.99 ± 37.64
		AA	515.16 ± 199.28	641.17 ± 372.76	74.25 ± 15.79
		P Value	0.06	0.91	0.74
SNP 4	SCC	CC	263.22 ± 62.30	445.90 ± 131.77	-
		CT	149.52 ± 72.56	523.48 ± 111.11	-
		TT	388.17 ± 75.60	647.13 ± 102.09	76.43 ± 14.11
		P Value	0.06	0.32	-
SNP 4	FOM	CC	0.65 ± 0.11	0.16 ± 0.09	-
		CT	0.24 ± 0.13	0.26 ± 0.08	-
		TT	0.56 ± 0.14	0.27 ± 0.07	0.16 ± 0.05
		P Value	0.05	0.53	-

\* Data is displayed as LS means ± SE and P value; \* The Bonferroni *t* test was used for pair-wise comparison in the study; \* SCC = Somatic cell count (x10<sup>3</sup> cells/mL), SCS = Somatic cell score, FOM = Frequency of clinical mastitis (per year); P values in bold type are significantly different (P<0.05); \* The values with no common superscripts are significantly different from each other (P<0.01 for capital letters and P<0.05 for small letters)

*Association analysis.* The association analysis showed that SNP 1 was significantly associated with protein percentage and SCC in the overall studied population (P<0.05). The TT genotype was recorded in a significantly higher protein percentage than CT. A significantly lower SCC was recorded in the CC genotype than the CT genotype. In SNP 2, the TT genotypes revealed significantly lower SCC and SCS than the CC genotype.

The heterozygous CT genotype in SNP 4 depicted a significant association with SCC and FOM compared to the homozygous TT and CC genotypes, respectively (Table 5 and Fig. 3). The breed-wise association analysis of SNPs revealed that the TT genotype of SNP 1 was found in higher frequency and showed a significant association with protein percentage in the indigenous (Achai) breed.

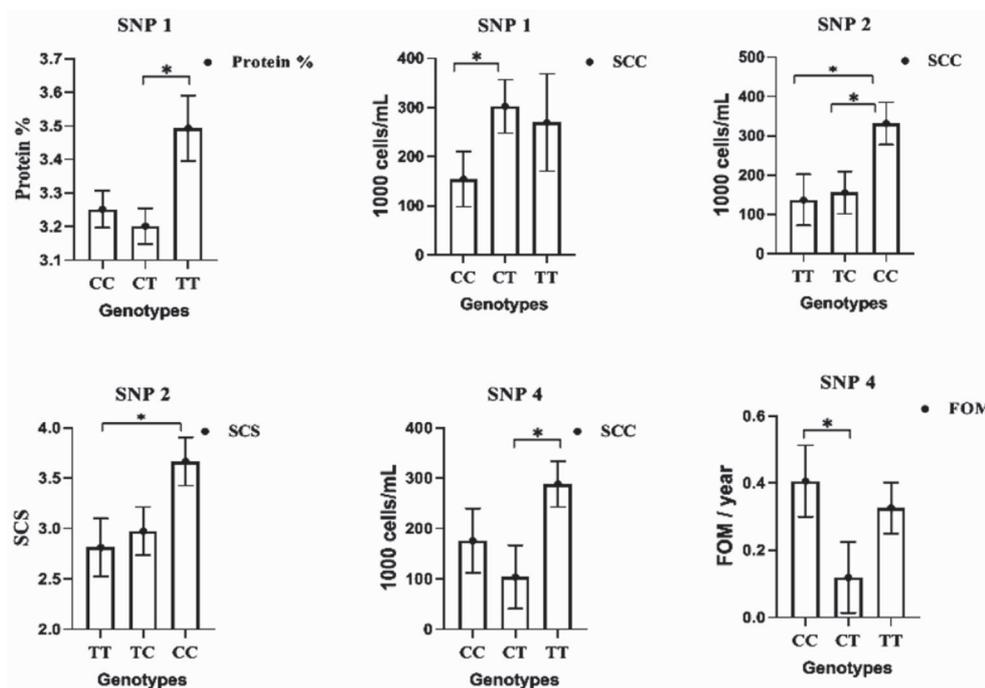


Fig. 3. Significant association of SNPs in *STAT5B* with phenotypes. Asterisk (\*) symbol shows significance association between the two genotypes; SNP 1 ( rs135194045) Chr19 - g.42984480 C>T, synonymous variant; SNP 2 (rs43706496) Chr19 - g.42965083 T>C, synonymous variant; SNP 4 (rs382802804) Chr19 - g.42960341 C>T, 3' UTR variant

In SNP 2, the TT and TC genotypes showed significantly lower SCC, whereas the TT genotype revealed lower SCS and higher frequency in the HF breed compared to other breeds. Similarly a tendency toward significance was observed for the CC genotype in SNP 3, with lower SCC and CT genotype of SNP 4 towards lower SCC and FOM. A higher genotype frequency of these genotypes was recorded in the HF breed (Table 4 and 6).

*In silico prediction of RNA secondary structure and 3' UTR confirmation.* The *in silico* predictions of the secondary structure of mRNA showed an increase in minimum free energy and pseudoknots formation for SNP 2 and SNP 3, showing them to be less stable in a cellular environment (Fig. 4.) The SNP 4 was found 17 nucleotides downstream from the last exon located in the 3' UTR region of the *STAT5B* gene.

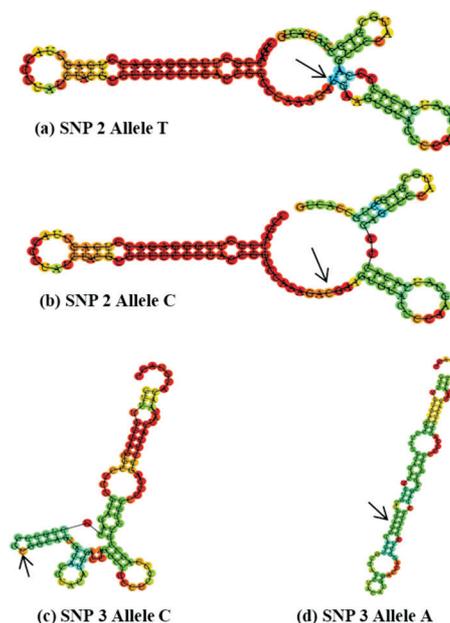


Fig. 4. RNA Secondary structure predictions (a) mRNA for allele T of SNP 2 (MFE = -19.30) (b) mRNA for allele C of SNP 2 (MFE = -19.10) (c) mRNA for allele C of SNP 3 (MFE = -27.00) (d) mRNA for allele A of SNP 3 (MFE = -29.90).

\* The arrows shows the position of variant

## Discussion

Mastitis is the most common inflammatory disease in cows that negatively affects udder health, resulting in lower milk yield and disrupted milk composition. Milk yield, milk composition and mastitis resistance are complex traits that are under the control of many genes.

In the present study, breed-wise genotype frequencies showed that the CC genotype of SNP 1, SNP 3 and SNP 4, the TT genotype of SNP 2, and the CT genotype of SNP 4 were seen to be absent in the studied population of the indigenous Achai breed. The allele frequency of the Achai breed revealed the complete absence of C allele of SNP 4 in the population. A significant association with the protein percentage of the TT genotype in SNP 1 and its lower genotype frequency in the exotic breed indicates the need for selection of this favourable genotype in exotic breeds. The favourable genotypes in SNP 2, SNP 3 and SNP 4 in relation to SCC, SCS and FOM showed higher genotype frequency in the exotic HF breed compared to the indigenous Achai breed. The contradictory results of allele and genotype frequency in exotic breeds and the indigenous breed suggests the outcome of selection pressure of favourable genotypes for production traits and udder health in breeding programs in exotic dairy breeds, *i.e.* HF and Jersey, and the lack of such selection in breeding the indigenous breed (Achai), where normally random mating is practised. The allele and genotype frequencies of the HF breed for SNP 2 in exon 16 (rs43706496 T>C) was found to be in complete agreement with HE et al., (2011) and in partial agreement with USMAN et al., (2015) in their studies focusing on Chinese Holstein.

The SNP 1 and SNP 3 identified in our study had not been studied previously for association analysis with milk performance and mastitis-related traits in dairy cattle. The association analysis of SNP 2 in the overall studied population and the HF breed in breed-wise data revealed a significant association with mastitis-related traits (*i.e.* SCC and SCS). HE et al. (2011) found the significant association of the same SNP with milk and protein yield in Chinese Holstein. KAKOLAKI et al., (2014) also studied SNPs in exon 16 of the *STAT5B* gene in Khorasan

Razavi Holstein cows, and recorded that it was significantly associated with milk yield. USMAN et al., (2015) studied SNP 2 and found a non-significant association with mastitis-related traits in Chinese Holstein cattle.

All three SNPs (SNP 1, 2 and 3) in our study were found to be synonymous variants. These synonymous polymorphisms were formerly known as silent variants in DNA as they were believed to be incapable of causing amino acid changes during translation. However, recent evidence shows that these SNPs may not be neutral. The degeneracy of the genetic code contains some additional information, such as variations in the rate of gene expression, mRNA folding, mRNA alternative splicing, and protein folding and function (PARMLEY et al., 2005; GOODMAN et al., 2013; HUNT et al., 2014). The synonymous variants may cause changes in the thermodynamic stability of mRNA, resulting in changes in the stem loop structure, and hence modification in the mRNA secondary structure. It may affect mRNA decay rates, and may result in disease (CHAMARY and HURST, 2005). Synonymous mutation in humans have been reported to be the cause of haemophilia (SIMHADRI et al., 2017), Schizophrenia and autism (TAKATA et al., 2016), and cystic fibrosis, caused by misfolding and degradation of the altered structure of mRNA (BARTOSZEWSKI et al., 2010).

The association analysis showed that the CT genotype in SNP 4 was associated with significantly lower SCC and FOM compared to the other genotypes, which suggests increased genetic resistance to mastitis. The *in silico* prediction of mRNA secondary structure and pseudoknots formation in our study also showed the decreased stability of mRNA due to the identified SNPs, that alter the expression level. USMAN et al., (2014) reported a non-significant association of this SNP with all the mastitis-related traits under study in Chinese Holstein cattle. Studies showed that 3' UTR is an important regulatory region that plays a role in the activity and expression of genes. Like the enhancer, the 5' and 3' UTR regions may contain the regulatory motifs for a gene and thus can affect gene regulation in different ways, such as nuclear export, stability, and the translational efficiency of mRNA

(HUGHES, 2006). The 3' UTR regions may contain binding sites for small interfering RNAs (siRNA), which include microRNAs (miRNA) (OLSEN and AMBROS, 1999). ALAM et al., (2019) found a significant association of SNPs in 3' UTR of the goat *STAT5B* gene with resistance to *Haemonchus contortus* infection in goats. Synonymous and 3' UTR variants were formerly neglected but the current study has revealed that these SNPs could play an important role in milk composition and mastitis resistance traits.

### Conclusion

The study showed noticeable differences between the allele and genotype frequencies of exotic and indigenous breeds. These changes suggest extensive selection for favourable alleles against mastitis resistance in exotic cattle breeds compared to the indigenous Achai breed, in which random mating is common. The identified genetic variants and their significant association with mastitis indicator and milk composition traits suggest that these could be useful molecular markers, and should be considered in dairy cattle breeding schemes for improved udder health and milk yield.

### Conflict of Interest

The authors declare that they have no conflict of interest.

### Funding

The research was funded by the start up research project of HEC Pakistan (SRGP-536), Modern Agro-industry Technology Research System (CARS-36); the Program for Changjiang Scholar and Innovation Research Team in University (IRT\_15R62); and the joint project of Pakistan Science Foundation and Nature Science Foundation of China PSF-NSFC III /Agr/KP/AWKUM/ (20) and (31961143009). The funders had no role in study design, data collection, analysis, decision to publish, and preparation of the manuscript.

### Acknowledgments

We are thankful to the manager and staff of the dairy farms (*i.e.* University of Agriculture, Peshawar Dairy Farm, Government Cattle Breeding and Dairy Farm Harichand, Livestock Research and Development Station Surezai and Dairy Farm Khazana, Charsada) for providing the milk and blood samples and relevant data of the animals.

### References

- ALAM, M. B. B., A. I. OMAR, M. O. FARUQUE, D. R. NOTTER, K. PERIASAMY, M. M. H. MONDAL, M. J. U. SARDER, M. SHAMSUDDIN, J. CAO, X. DU (2019): Single nucleotide polymorphisms in candidate genes are significantly associated with resistance to *Haemonchus contortus* infection in goats. *J. Anim. Sci. Biotechnol.* 10, 30. DOI: 10.1186/s40104-019-0327-8
- BARTOSZEWSKI, R. A., M. JABLONSKY, S. BARTOSZEWSKA, L. STEVENSON, Q. DAI, J. KAPPES, J. F. COLLAWN, Z. BEBOK (2010): A synonymous single nucleotide polymorphism in  $\Delta F508$  CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J. Biol. Chem.* 285, 28741-28748. DOI: 10.1074/jbc.M110.154575
- BENEKLI, M., M. R. BAER, H. BAUMANN, M. WETZLER (2003). Signal transducer and activator of transcription proteins in leukemias. *Blood.* 101, 1940-1954. DOI: 10.1182/blood-2002-04-1204.
- CHAMARY, J. V., L. D. HURST (2005): Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol.* 6, R75. DOI: 10.1186/gb-2005-6-9-r75
- COPELAND, N. G., D. J. GILBERT, C. SCHINDLER, Z. ZHONG, Z. WEN, J. E. DARNELL JR, A. L.-F. MUI, A. MIYAJIMA, F. W. QUELLE, J. N. IHLE (1995): Distribution of the mammalian Stat gene family in mouse chromosomes. *Genomics.* 29, 225-228. DOI: 10.1006/geno.1995.1235
- DETILLEUX, J. C. (2002): Genetic factors affecting susceptibility of dairy cows to udder pathogens. *Vet. Immunol. Immunop.* 88, 103-110. DOI: 10.1016/S0165-2427(02)00138-1
- GOODMAN, D. B., G. M. CHURCH, S. KOSURI (2013): Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475-479. DOI: 10.1126/science.1241934
- HAGNESTAM-NIELSEN, C., U. EMANUELSON, B. BERGLUND, E. STRANDBERG (2009): Relationship between somatic cell count and milk yield in different stages of lactation. *J. Dairy. Sci.* 92, 3124-3133. DOI: 10.3168/jds.2008-1719
- HE, Y., Q. CHU, P. MA, Y. WANG, Q. ZHANG, D. SUN, Y. ZHANG, Y. YU, Y. ZHANG (2011): Association of bovine *CD4* and *STAT5b* single nucleotide polymorphisms with somatic cell scores and milk production traits in Chinese Holsteins. *J. Dairy Res.* 78, 242-249. DOI: 10.1017/S0022029911000148
- HINRICHS, D., E. STAMER, W. JUNGE, E. KALM (2005): Genetic analyses of mastitis data using animal threshold models and genetic correlation with production traits. *J. Dairy Sci.* 88, 2260-2268. DOI: 10.3168/jds.S0022-0302(05)72902-7

- HUGHES, T. A. (2006): Regulation of gene expression by alternative untranslated regions. *Trends Genet.* 22, 119-122.  
DOI: 10.1016/j.tig.2006.01.001
- HUNT, R. C., V. L. SIMHADRI, M. IANDOLI, Z. E. SAUNA, C. KIMCHI-SARFATY (2014): Exposing synonymous mutations. *Trends Genet.* 30, 308-321.  
DOI: 10.1016/j.tig.2014.04.006
- KAKOLAKI, M. R., M. B. M. TORBATI, H. FARHANGFAR, A. OMIDI (2014): Study of polymorphism in *Stat5b* gene and its association with milk production traits in holstein cows of Khorasan Razavi Province. *Genetics in the 3<sup>rd</sup> Millennium.* 12, 3512-3519.
- LEYVA-BACA, I., F. SCHENKEL, J. MARTIN, N. A. KARROW (2008): Polymorphisms in the 5' upstream region of the *CXCR1 chemokine receptor* gene, and their association with somatic cell score in Holstein cattle in Canada. *J. Dairy. Sci.* 91, 407-417.  
DOI: 10.3168/jds.2007-0142
- OLSEN, P. H., V. AMBROS (1999): The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* 216, 671-680.
- PANT, S. D., F. S. SCHENKEL, I. LEYVA-BACA, B. S. SHARMA, N. A. KARROW (2007): Identification of single nucleotide polymorphisms in bovine *CARD15* and their associations with health and production traits in Canadian Holsteins. *BMC Genomics.* 8, 421.  
DOI: 10.1186/1471-2164-8-421
- PARMLEY, J. L., J. V. CHAMARY, L. D. HURST (2005): Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Molecular biology and evolution.* 23, 301-309.  
DOI: 10.1093/molbev/msj035
- RUPP, R., D. BOICHARD (1999): Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. *J. Dairy. Sci.* 82, 2198-2204.  
DOI: 10.3168/jds.S0022-0302(99)75465-2
- SEEGERS, H., C. FOURICHON, F. BEAUDEAU (2003): Production effects related to mastitis and mastitis economics in dairy cattle herds. *Vet. Res.* 34, 475-491.  
DOI: 10.1051/vetres:2003027
- SIMHADRI, V. L., N. HAMASAKI-KATAGIRI, B. C. LIN, R. HUNT, S. JHA, S. C. TSENG, A. WU, A. A. BENTLEY, R. ZICHEL, Q. LU (2017): Single synonymous mutation in factor IX alters protein properties and underlies haemophilia B. *J. Med. Genet.* 54, 338-345.  
DOI: 10.1136/jmedgenet-2016-104072
- TAKATA, A., I. IONITA-LAZA, J.A. GOGOS, B. XU, M. KARAYIORGOU (2016): De novo synonymous mutations in regulatory elements contribute to the genetic etiology of autism and schizophrenia. *Neuron.* 89, 940-947.  
DOI: 10.1016/j.neuron.2016.02.024
- TEGLUND, S., C. MCKAY, E. SCHUETZ, J.M. VAN DEURSEN, D. STRAVOPODIS, D. WANG, M. BROWN, S. BODNER, G. GROSVELD, J.N. IHLE (1998): *Stat5a* and *Stat5b* proteins have essential and nonessential, or redundant, roles in cytokine responses. *Cell.* 93, 841-850.  
DOI: 10.1016/S0092-8674(00)81444-0
- USMAN, T., Y. WANG, C. LIU, X. WANG, Y. ZHANG, Y. YU (2015): Association study of single nucleotide polymorphisms in *JAK 2* and *STAT5B* genes and their differential mRNA expression with mastitis susceptibility in Chinese Holstein cattle. *Anim. Genet.* 46, 371-380.  
DOI: 10.1111/age.12306
- USMAN, T., Y. WANG, Y. YU, C. LIU, X. WANG, Y. ZHANG (2014): Association study of single nucleotide polymorphisms in *STAT5A/B* with mastitis susceptibility in Chinese Holstein cattle. In *Proceedings of the 10<sup>th</sup> World Congress of Genetics Applied to Livestock Production.*

Received: 6 April 2020

Accepted: 25 June 2020

---

**ALI, N., S. NIAZ, I. KHATTAK, N. U. KHAN, H. LUO, H. MUHAMMAD, T. ALI, Y. YU, Y. WANG, T. USMAN: Povezanost varijacija gena *STAT5B* sa proizvodnjom mlijeka i svojstvima mastitisa u mliječnim goveda. *Vet. arhiv* 91, 459-471, 2021.**

#### SAŽETAK

Proizvodnja i zdravlje vimena su zahtjevna svojstva u uzgoju mliječnih goveda. Cilj je ovog rada bio je istražiti povezanost polimorfizma pojedinačnih nukleotidnih (SNPs) gena *STAT5B* u mliječnim goveda sa svojstvima proizvodnje mlijeka i mastitisa. Istraživanje je uključilo 201 kravu s četiri mliječne farme iz pokrajine Khyber Pakhtunkhwi u Pakistanu, među kojima su bile jedinke čistokrvnih pasmina (Holstein Friesian, Jersey i Achai) i križanci. U uzorcima mlijeka analizirani su broj somatskih stanica (SCC) i sastav mlijeka (postotak masnoće, proteina

i laktoze). Generalizirani linearni model, uz uporabu SAS statističkog programa, poslužio je za analizu povezanosti. Sekvenciranje DNA pokazalo je fond od četiri (tri sinonimna i jedan 3' UTR) SNP-a u genu *STAT5B*. Ovi su pojedinačni nukleotidni polimorfizmi dodatno potvrđeni u svim uzorcima DNA primjenom testa SNaPshot. Analiza pasmina pokazala je da većina SNP-a odgovara Hardy-Weinbergovoj ravnoteži ( $P > 0,05$ ). Istraživanje povezanosti pokazalo je znakovito veći postotak proteina u genotipu TT i manji broj SCC-a u CC genotipu SNP-a 1 (ekson 2, C > T), dok je u SNP-u 2 (ekson 16, T > C), genotip TT pokazao znakovito manji broj SCC-a i SCS-a u usporedbi s drugim genotipovima ( $P < 0,05$ ). SNP 4 (3' UTR, C>T) pokazao je znakovito manji broj SCC-a i učestalost kliničkog mastitisa u heterozigotnog (CT) genotipa u usporedbi s homozigotnim genotipovima. In silico predikcija pokazala su promjene u RNA sekundarnoj strukturi za SNP 2 i SNP 3. Rezultati pokazuju da bi se *STAT5B* trebao uzeti u obzir kao kandidatni gen, a njegove varijante kao koristan genski biljeg za poboljšanje sastava mlijeka i zdravlja vimena.

**Ključne riječi:** mliječne krave; mastitis; svojstva mlijeka; *STAT5B*; SNP

---

