

**Corresponding author: Tajda Mirič**

**E-mail: tajda.miric@gmail.com**

---

## **IDENTIFICATION OF GENES ASSOCIATED WITH MILK FAT YIELD IN BROWN SWISS CATTLE POPULATION**

Tajda Mirič<sup>1</sup>, Klemen Potočnik<sup>1</sup>, Jurij Krsnik<sup>1</sup>, Miran Štepec<sup>1</sup>, Barbara Luštrek<sup>1</sup>, Tanja Kunej<sup>1</sup>

<sup>1</sup> University of Ljubljana, Biotechnical Faculty, Department of Animal Science, Slovenia

### **IZVLEČEK**

S proizvodnimi lastnostmi mleka so povezani številni geni in kromosomske regije. Lokusi kvantitativnih lastnosti (QTL) za količino maščobe v mleku se pri govedu nahajajo na vseh kromosomih. Vendar so si objavljeni podatki nasprotujoči in kandidatni geni za številne QTL-e še niso znani. Namen te raziskave je bil določiti gene za količino maščobe v mleku v populaciji rjavega goveda (BSW). Analiza je bila izvedena na podatkih zadnjega mednarodnega obračuna plemenskih vrednosti mlečnih pasem goveda Interbull-centra. Študijska skupina je vključevala 26.203 genotipiziranih živali, od katerih je 6.542 živali imelo tudi mednarodno oceno plemenske vrednosti, preračunano na slovensko merilo. Genotipizacija je bila izvedena z uporabo različnih tipov DNA-mikromrež, med njimi BovineSNP50 BeadChip. Analizirali smo genomske razporeditve 20 polimorfizmov, najbolj značilno povezanih s količino maščobe v mleku. Rezultati so razkrili štiri genomske regije z večjo gostoto polimorfizmov na kromosomih 5, 6, 9 in 25. Šest od 20 analiziranih polimorfizmov se nahaja znotraj petih genov: *SLC25A43*, *NAV3*, *BFAR*, *HS3ST6* in *TMPRSS11F*. Gen *NAV3* je posebej zanimiv, saj vključuje dva polimorfizma, povezana s količino maščobe v mleku. Nekateri signali so se prekrivali s predhodno poročanimi QTL-i za podobne lastnosti pri drugih populacijah goveda. Za določanje zanesljivejših bioznačevalcev bi bilo potrebno analizo opraviti tudi pri drugih pasmah goveda.

**Ključne besede:** govedo, DNA-mikromreža, količina maščob v mleku

### **ABSTRACT**

Several genes and chromosomal regions have been associated with milk traits. Reported QTLs associated with milk fat yield are distributed on all cattle chromosomes. However, published data are opposing and several QTL candidate genes are unknown. The aim of the present study was to identify genes, associated with milk fat yield in Brown Swiss (BSW) cattle population. The analysis was performed using the data available from the international genetic evaluation of dairy cattle at Interbull Centre. Study group included 26,203 genotyped animals including 6,542 with international breeding values on Slovenian scale. Genotyping was performed with various types of DNA microarrays, including BovineSNP50 BeadChip. We analyzed genomic distribution of 20 polymorphisms, most significantly associated with milk fat yield. The results revealed four genomic regions with higher density of polymorphisms on chromosomes 5, 6, 9, and 25. Six out of 20 polymorphisms are located within five genes: *SLC25A43*, *NAV3*, *BFAR*, *HS3ST6*, and *TMPRSS11F*. Gene *NAV3* is of special importance, since it includes two polymorphisms associated with milk fat yield. Some signals overlapped previously reported QTLs for similar traits in other cattle populations. For identification of more robust markers, the analysis should also be performed in other cattle breeds.

**Key words:** cattle, DNA microarray, milk fat yield

## INTRODUCTION

Genome-wide association study (GWAS) is a powerful approach for identification of genes and chromosomal regions associated with dairy phenotypes and is useful for gene-based and genome selection (Harder et al., 2006). Several genes and chromosomal regions have been associated with a large number of milk traits. Reported QTLs associated with milk fat are distributed on all cattle chromosomes and are systematically catalogued in the QTLdb (Hu et al., 2016). For example, in Holstein cattle a QTL for milk fat yield has been identified in BTA13 (Cole et al., 2011). Additionally, *PPARGC1A* gene has been suggested to be involved in genetic variation underlying the QTL for milk fat synthesis on BTA6 (Weikard et al., 2005). The strongest signal for milk production traits, including fat yield in Brown Swiss cattle was identified on chromosome 25 (Guo et al., 2012). This study also included samples from Slovenian Brown Swiss cattle population. Published data are opposing and several QTL candidate genes are still unknown. The aim of the present study was to identify genes, associated with milk fat yield in Brown Swiss cattle population.

## MATERIAL AND METHODS

**Databases and bioinformatics tools.** Ensembl release 90 was used to retrieve genomic locations of polymorphisms and overlapping protein-coding genes (<http://www.ensembl.org/index.html>). Conversion of commercial BovineSNP50 BeadChip polymorphism names to reference SNP (rs) ID was performed using SNPchiMp tool (<http://bioinformatics.tecnoparco.org/SNPchimp/>) (Nicolazzi et al., 2014). Quantitative trait loci data were downloaded from QTLdb, release 33 (<https://www.animalgenome.org/cgi-bin/QTLdb/index>) (Hu et al., 2016). Ensembl genes which have not yet been manually annotated were identified using information of orthologous genes from Ensembl genomic browser and NCBI database (<https://www.ncbi.nlm.nih.gov/>).

**Samples, genotyping and statistical analysis.** The analysis was performed using the data available from the international genetic evaluation of dairy cattle at Interbull Centre (Jorjani et al., 2011). Study group included 26,203 genotyped animals including 6,542 with international breeding values on Slovenian scale (analysis performed in March, 2017). Genotyping was performed using various types of DNA microarrays, including BovineSNP50 BeadChip, Geneseek GPP, IDV v.2 and v.3) and all results were unified according to the Interbull SNP index. Statistical analysis was performed as described previously (Guo et al., 2012). Estimated regression coefficients from SNP equation for milk fat yield on Slovenian scale were used. Each regression coefficient represent additive genetic effect of each SNP on milk fat yield.

## RESULTS AND DISCUSSION

**Results of the genome-wide association analysis.** In the present study we performed DNA-microarray analysis for identification of genes associated with milk fat yield in Brown Swiss cattle breed. The results revealed that the first 280 polymorphisms, most significantly associated with milk fat yield contributed to phenotypic variability in 3% and the first 100 most significantly associated polymorphisms contributed 1.22% to the researched trait. In the present study we analyzed genomic



distribution of 20 polymorphisms most significantly associated with milk fat yield, which explained 0.29% of the phenotypic variability.

**Genomic distribution of polymorphisms associated with milk fat yield.** The highest number of SNPs associated with milk fat yield is located on the chromosomes: 5 (four SNPs), 6 (four SNPs), 25 (three SNPs), 9 (two SNPs), and 20 (two SNPs). One polymorphism per chromosome is located on chromosomes 13, 14, 17, 28, and X. Genomic location of twenty SNPs significantly associated with milk fat yield is presented in Table 1. The table also includes information of the nearest protein-coding genes, their orientation (sense, antisense) and the location of polymorphisms relative to the gene (upstream, downstream or within the gene).

Six out of these 20 polymorphisms are located in five genes: rs42197808 (in the *SLC25A43* gene), rs42774123 and rs43710106 (both in the *NAV3* gene), rs110426340 (in the *BFAR* gene), rs109557202 (in the *HS3ST6* gene), and rs41655005 (in the *TMPRSS11F* gene). These six polymorphisms are intronic.

Fourteen out of 20 polymorphisms are intergenic and located in the vicinity of 24 protein-coding genes: *SLC4A4*, *GC*, *ENSBTAG00000046399*, *CPPED1*, *RBM19*, *TBX5*, *DCK*, *SOCS2*, *ENSBTAG00000005107*, *SRGN*, *ENSBTAG00000014558*, *ENSBTAG0000013213*, *STK3*, *SLC1A3*, *RANBP3L*, *KCNJ8*, *LDBH*, *KIAA0408*, *ECHDC1*, *MAFB*, *TOP1*, *NPFFR2*, *SLIT3* and *ENSBTAG00000047333*. For five of these protein-coding genes the official gene name is not yet available in the Ensembl database. They are named with the Ensembl ID (ENSBTA#). Names for four out of these five genes were obtained using the information of their orthologous genes: *ENSBTAG00000046399* (*SHISA9*), *ENSBTAG00000014558* (*DDX21*), *ENSBTAG0000013213* (*OSR2*), and *ENSBTAG00000005107* (*CRADD*). For the Ensembl gene *ENSBTAG00000047333* 34 orthologous genes are known, but all of them have only Ensembl names, so it was not possible to determine annotated gene name.

The results revealed that polymorphisms are not uniformly distributed throughout the genome. There are four genomic regions with higher density of polymorphisms associated with milk fat yield; on chromosomes 5, 6, 9 and 25. Four SNPs are located on the chromosome 5, two of them are located within the gene *NAV3* (rs42774123 and rs43710106). On the chromosome 6 four polymorphisms are located. One polymorphism is located within the *TMPRSS11F* gene. Three polymorphisms (rs110527224, rs110354582, and rs42766480) are located in vicinity of four genes *DCK*, *SLC4A4*, *GC*, and *NPFFR2* gene. These three polymorphisms are located at the distance of 0.78 Mbp (Figure 1).



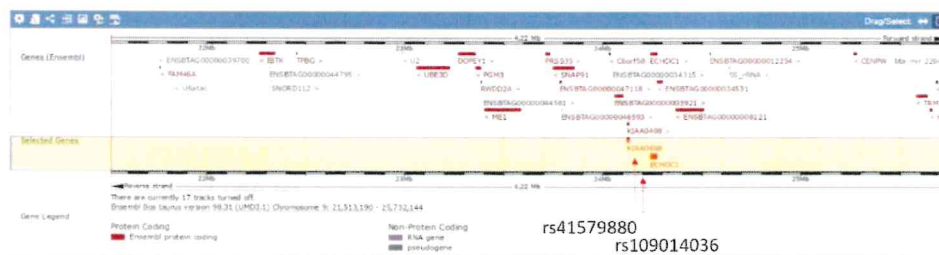
**Figure 4:** Presentation of the chromosome BTA6 with four polymorphisms associated with milk fat yield in Brown Swiss cattle breed. One polymorphism is located within the *TMPRSS11F* gene and three polymorphisms are located in vicinity of the *DCK*, *SLC4A4*, *GC*, and *NPFFR2* genes.

**Table 3:** Genomic location of twenty SNPs significantly associated with milk fat yield and associated genes.

rs SNP ID	Chromosome	Nearest gene 1			Nearest gene 2		
		Gene name	Orientation (sense/antisense)	SNP location	Gene name	Orientation (sense/antisense)	SNP location
rs110527224	6	<i>SLC4A4</i>	Sense	Downst.	<i>GC</i>	Antisense	Downst.
rs42055812	25	<i>ENSBTAG00000046399 (SHISA9)</i>	Sense	Downst.	<i>CPPED1</i>	Antisense	Upst.
rs41625122	17	<i>TBX5</i>	Sense	Downst.	<i>RBM19</i>	Sense	Upst.
rs42197808	X	<i>SLC25A43</i>	Sense	In	/	/	/
rs42774123	5	<i>NAV3</i>	Sense	In	/	/	/
rs110354582	6	<i>DCK</i>	Sense	Downst.	<i>SLC4A4</i>	Sense	Upst.
rs41654779	5	<i>SOCS2</i>	Sense	Downst.	<i>ENSBTAG00000005107 (CRADD)</i>	Sense	Upst.
rs110103383	28	<i>ENSBTAG00000014558 (DDX21)</i>	Sense	Downst.	<i>SRGN</i>	Sense	Upst.
rs110405331	14	<i>ENSBTAG00000013213 (OSR2)</i>	Antisense	Upst.	<i>STK3</i>	Sense	Upst.
rs41943459	20	<i>SLC1A3</i>	Antisense	Upst.	<i>RANBP3L</i>	Sense	Upst.
rs109014036	9	<i>KIAA0408</i>	Sense	Downst.	<i>ECHDC1</i>	Sense	Upst.
rs109628799	5	<i>KCNJ8</i>	Sense	Downst.	<i>LDHB</i>	Sense	Upst.
rs41579880	9	<i>KIAA0408</i>	Sense	Downst.	<i>ECHDC1</i>	Sense	Upst.
rs110426340	25	<i>BFAR</i>	Sense	In	/	/	/
rs109557202	25	<i>HS3ST6</i>	Antisense	In	/	/	/
rs110573257	13	<i>MAFB</i>	Antisense	Downst.	<i>TOP1</i>	Sense	Upstr.
rs41655005	6	<i>TMPRSS11F</i>	Antisense	In	/	/	/
rs43710106	5	<i>NAV3</i>	Sense	In	/	/	/
rs42766480	6	<i>GC</i>	Antisense	Upst.	<i>NPFRR2</i>	Sense	Upst.
rs109346620	20	<i>ENSBTAG00000047333</i>	Sense	Downst.	<i>SLIT3</i>	Antisense	Downst.

Downst. – downstream; Upst. – upstream; In – in the gene

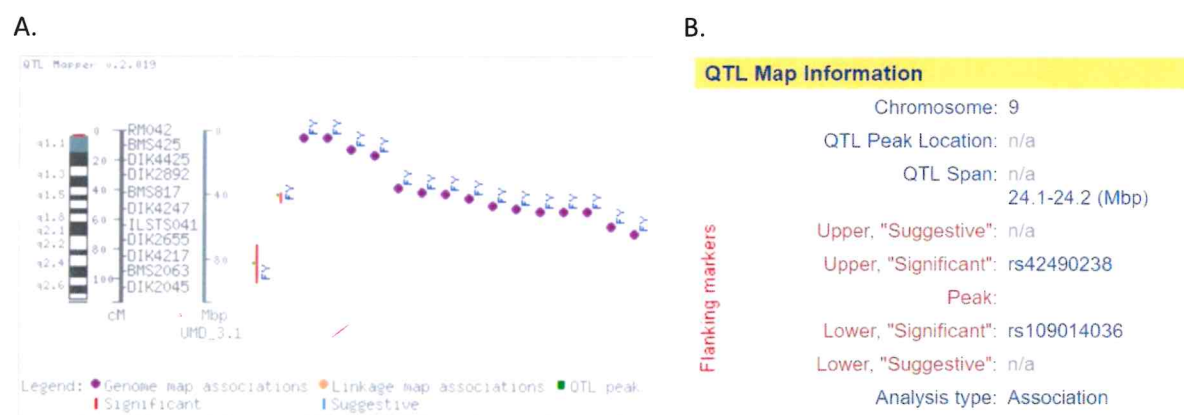
Two polymorphisms located in the chromosome 9 (rs109014036 and rs41579880) are located in vicinity of two genes; *KIAA0408* and *ECHDC1* (Figure 2). The distance between these two polymorphisms is 27,879 bps. Three polymorphisms are located on the chromosome 25, including two at the distance of 1.88 Mbp (rs42055812 and rs110426340).



**Figure 2:** Presentation of the chromosome BTA9 with two polymorphisms associated with milk fat yield in Brown Swiss cattle breed. Polymorphisms are located between *KIAA0408* and *ECHDC1* genes.



**Comparison of genomic distribution between polymorphisms and QTL.** We also performed comparison of the genomic location of 20 researched polymorphisms with location of previously identified QTLs associated with milk fat yield. The analysis revealed genomic overlaps. For example, on the chromosome BTA9 two polymorphisms (rs41579880 and rs109014036) overlap the region of two previously identified QTLs (Figure 3). QTL#1687 spans the region 18.2–26.7 Mbp and it has been identified using genome-wide scan of 221 microsatellite loci in Holstein breed (Schnabel et al., 2005). QTL #121799 spans the region 24.1–24.2 Mbp and it has been identified in a genome wide CNV analysis in Holstein breed (Xu et al., 2014). Candidate genes associated with this QTL are not yet known. These two QTL regions span large genomic regions. Additional studies are needed to identify the associated loci.



**Figure 3:** Presentation of QTLs, associated with milk fat yield (FY) in chromosome BTA9. Two polymorphisms (rs41579880 and rs109014036) are located within this region and they overlap with two QTL. QTL#1687 spans the region 18.2–26.7 Mbp (A). QTL #121799 spans the region 24.1–24.2 Mbp (B).

**Conclusion.** Six out of 20 analyzed polymorphisms are located within five genes (*SLC25A43*, *NAV3*, *BFAR*, *HS3ST6*, and *TMPRSS11F*). Gene *HS3ST6* has previously been reported to be associated with milk fat yield (Guo et al., 2012). To our knowledge, the other four genes have not yet been associated with milk fat yield in cattle. In this study we therefore confirmed one and identified four novel candidate genes for this trait. Gene *NAV3* is of special importance since it includes two polymorphisms associated with milk fat yield. In the present study we analyzed genomic location of 20 polymorphisms most significantly associated with milk fat yield. The study should be extended to other polymorphisms, for example to 280 polymorphisms explaining 3% of the phenotypic variability. For identification of more robust markers, the analysis should also be performed on other cattle breeds.

## REFERENCES

1. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ Jr, Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK, Da Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics*. 2011, 12, 408.
2. Guo J, Jorjani H, Carlborg Ö. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. *BMC Genet*. 2012, 13, 82.
3. Harder B, Bennewitz J, Reinsch N, Thaller G, Thomsen H, Kühn C, Schwerin M, Erhardt G, Förster M, Reinhardt F, Kalm E. Mapping of quantitative trait loci for lactation persistency traits in German Holstein dairy cattle. *J Anim Breed Genet*. 2006, 123, 89–96.
4. Hu ZL, Park C, Reecy JM. Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res*. 2016, 44 (D1), D827–33.

5. Jorjani H, Jakobsen J, Nilforooshan MA, Hjerpe E, Zumbach B, Palucci V, Dürr J. Genomic evaluation of BSW populations, InterGenomics: Results and Deliverables. *Interbull Bulletin*. 2011, 43, 5–8.
6. Nicolazzi EL, Picciolini M, Strozzi F, Schnabel RD, Lawley C, Pirani A, Brew F, Stella A. SNPchipMp: a database to disentangle the SNPchip jungle in bovine livestock. *BMC Genomics*. 2014, 15, 123.
7. Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Anim Genet*. 2005, 36, 408–16.
8. Weikard, R., C. Kühn, T. Goldammer, G. Freyer, and M. Schwerin. The bovine PPARGC1A gene: molecular characterization and association of an SNP with variation of milk fat synthesis. *Physiol Genomics*. 2005, 21, 1–13.
9. Xu L, Cole JB, Bickhart DM, Hou Y, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP, Liu GE. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. *BMC Genomics*. 2014, 15, 683.