

Original Research Paper

Genome-Wide Association Study of Single Nucleotide Polymorphisms with Milk Traits in Red Gorbatov Cattle

¹Marina Ivanovna Selionova, ¹Grigorii Sergeyevich Shekhovtsev, ¹Ivan Petrovich Prokhorov, ¹Vladimir Nikolaevich Lukyanov and ²Yulia Vasilevna Shoshina

¹Department of Animal Breeding, Genetics and Biotechnology, Russian State Agrarian University-Moscow Timiryazev Agricultural Academy, Russia

²Department of Dairy and Beef Cattle Breeding, Russian State Agrarian University-Moscow Timiryazev Agricultural Academy, Russia

Article history

Received: 31-05-2024

Revised: 12-07-2024

Accepted: 18-07-2024

Corresponding Author:

Grigorii Sergeyevich Shekhovtsev

Department of Animal Breeding, Genetics and Biotechnology, Russian State Agrarian University-Moscow, Timiryazev Agricultural Academy, Russia
Email: shekhovtsev@rgau-msha.ru

Abstract: In order to preserve the gene pool of the Red Gorbatov breed, it is necessary to use a set of genomic analysis tools that allows more efficient planning in obtaining individuals who meet the economic requirements of dairy farming. The aim of the work was to identify candidate genes associated with milk production traits in Red Gorbatov cattle using a genome-wide association analysis. Genotyping was performed using the BovineSNP50 v3 Bead Chip panel and the Bead Studio program, quality control, data filtering, and identification of associations for each single-nucleotide polymorphism with milk productivity indicators was performed by Multiple linear regression analysis using the PLINK 1.90 software package. The process of data visualization was performed using the qqman package in the R programming language. 44 full-genome and 178 suggestive Single Nucleotide Polymorphisms (SNPs) associated with the milk parameters in Red Gorbatov cows have been identified, while the largest number of SNPs with a full genome were located on chromosomes 18 and 20-6 in both cases. On chromosome 18, one SNP (ARS-BFGL-NGS-18019) was common for two traits (protein and β -casein content). On chromosome 20, 2 SNPs (BTB-01636657 and ARS-BFGL-NGS-91235) were common for two traits (protein and β -casein content). Structural annotation of genomic regions within a window of ± 0.20 Mb revealed 20 genes with the described functions in terms of gene ontology. The genes ODAM, HS3ST2, SLIT2, ADARB2, TBC1D22A, ZFPM2, ARHGEF28, LRRC8D, OTOG, UNC80, GRIA1, C1QTNF3, CTNND2 are involved in processes related to the chemical composition of milk and milk productivity.

Keywords: GWAS, Milk, QTL, Red Gorbatov and SNP

Introduction

The breeding resources of the domestic cattle industry are represented by Various Indigenous breeds, the gene pool of which is characterized by high phenotypic and genotypic diversity, allowing, in particular, adaptation to various environmental conditions, as confirmed by the results of high-throughput genotyping and sequencing of genomes and transcripts of representatives of Turano-Mongolian breeds (Yudin *et al.*, 2021).

At the same time, many domestic cattle breeds have been subjected to introgression from imported genetic resources or so-called improving breeds, which poses a threat to the preservation of the unique domestic gene pool, especially considering that the restoration of the

natural genetic background of the recipient breed is possible only in rare cases (Slagboom *et al.*, 2022).

Nevertheless, recent studies of the allele pool of the Red Gorbatov breed have shown that this population of animals represents an isolated gene pool local breed, the improvement of which by bulls of the angler and red Danish breeds did not lead to the loss of the uniqueness of its allele pool (Volkova *et al.*, 2019).

To preserve the unique "historical" allele variants of the Red Gorbatov breed, in addition to studying the genetic architecture of the population, it is necessary to use a complex of genomic analysis tools that allow for more effective planning of the production of individuals that would meet the economic requirements of dairy cattle farming (Sermyagin *et al.*, 2020). Thus, creating the

conditions for the successful preservation and development of the Red Gorbatov breed by increasing the level of milk productivity of cows and enhancing their competitiveness.

Several researchers provide data on the identification of genetic variations that affect the technological properties of milk. For example, in the work of Bertelsen *et al.* (2016), the most significant SNP affecting milk coagulation, located proximally to the CSN3 gene, which explains 33% of the phenotypic dispersion, was identified. Fang *et al.*, (2019) discovered 10 QTL regions for the relative concentrations of individual phosphorylation isoforms of α S2- and α S1-casein on chromosomes 1, 2, 6, 9, 11, 14, 15, 18, 24 and 28. Sanchez *et al.* (2019) identified genes related to milk properties for cheese production in Montbeliarde cows, among them were well-known genes such as PAEP and DGAT1, as well as new ones - SLC37A1, ALPL, MGST1, SEL1L3, GPT, BRI3BP, SCD, GPAT4, FASN and ANKH, which explained 12 to 30% of the phenotypic dispersion of cheese suitability of milk. In a study on the identification of candidate genes influencing the characteristics of milk protein composition, (Zhou *et al.*, 2019a) identified 194 SNPs on 24 autosomes of cattle.

Other authors have also investigated single nucleotide polymorphisms associated with a wide range of quantitative and qualitative milk indicators. For example, (Zhou *et al.*, 2019b) identified the candidate gene CDH2, involved in adipogenesis, which may affect milk fat production. As a result of GWAS analysis and ROH models for assessing genomic variability for selection traits of Holsteinized black-and-white cattle, Sermyagin *et al.* (2020) identified quantitative trait loci on chromosomes 1, 2, 5, 7, 8, 10, 11, 12, 14, 16, 20, 21 and 26. In particular, a region was found on BTA14 (1.44-1.59 Mb) with the genes ZNF16, ARHGAP39, and ZNF7, associated with increased milk fat yield.

The results of Cruz *et al.* (2019) show that the DGAT1 gene is responsible for most of the variation in the composition of milk fatty acids in Holstein cattle. Additionally, the PLBD1 and MGST1 genes are also important candidate genes. Data from Lijun *et al.* (2021) confirm the genetic association of the AGPAT3 gene with the fatty acid composition of milk.

Silva *et al.* (2020) identified potential candidate genes (TRNAG-CCC, MAPK10, and PTPN3) associated with somatic cells. Ilie *et al.* (2021) reported that the most significant SNP (rs110749552) associated with the number of somatic cells was localized in the HERC3 gene. In the studies of Sanchez *et al.* (2021), 47 positional candidate genes and 45 transcription factors highly expressed in the mammary gland compared to 90 other tissues of cattle were identified, thereby defining the key role of the SLC37A1 and ANKH genes in the mineral content of cow's milk.

Nayeri *et al.* (2019) reported several significant regions associated with predicted mid-infrared spectroscopy concentrations of β -hydroxybutyrate in milk (a subclinical ketosis indicator), with the strongest associations found on chromosomes 6 and 14.

The analysis of whole-genome associations to identify genomic regions and candidate genes determining the milk yield in Brazilian Gir breed cattle revealed the presence of 7 candidate genes (LEP, CLOCK, CASR, LRRRC4, DOCK1, SLC15A2, and SND1) involved in various stages of milk synthesis on chromosomes 1, 4, 6 and 26 (Otto *et al.*, 2020).

Moreover, it is important to note that different breeds are characterized by different sets of candidate genes. For instance, DGAT1 explains a high proportion of variations in milk performance traits in the Holstein breed and fat yield in the Jersey breed. However, in the study by Oliveira *et al.* (2019), no evidence was found for a significant influence of DGAT1 on milk performance indicators and somatic cell count in the Ayrshire breed, even though the gene was identified in all three breeds.

Thus, genome-wide association studies in large ruminants are being conducted in many countries; however, they mainly present data on the Holstein and other commercial breeds, while numerous other breeds have been insufficiently studied to date. This holds true for the Red Gorbatov breed as well, which justifies the relevance of the present study.

The aim of this study was to search for candidate genes that may be associated with milk performance traits in Red Gorbatov cattle based on a genome-wide association analysis.

Materials and Methods

For the research using the random sampling method, 68 lactating Red Gorbatov cows (AO "Ababkovskoe", Nizhny Novgorod region, Pavlovsky district, Ababkovo, Kolhoznaya Street, building 5) were selected and ear tissue samples were collected for DNA isolation. The BovineSNP50 v3 BeadChip (Illumina Inc., San Diego, USA) with a coverage density of 53218 SNPs and the Illumina BeadStudio program was used for the animal genotyping procedure.

Each sample and SNP underwent quality control and genotyping data filtration procedures, performed using the PLINK 1.9 software, with the following filters applied: A call rate of at least 90% for all the examined SNPs for an individual sample (mind); a Call-rate of at least 90% for each SNP examined, for all the genotyped samples (geno); a Minor Allele Frequency (MAF) of more than 0.01 or 0.05 (maf 0.01); deviation of SNP genotypes from the Hardy-Weinberg distribution in the aggregate of tested samples with a p-value less than 10^{-6} (hwe) (Chang *et al.*, 2015). In addition, an assessment of the

disequilibrium coupling of the examined SNPs (LD estimation) with $r^2 < 0.2$ at a 50 kb step (indep-pairwise) was conducted.

The following indicators of milk productivity were used for the associative analysis: Daily milk yield, protein content, fat content, β -casein, lactose, dry matter, dry defatted milk residue, β -hydroxybutyrate, acetone, urea, freezing point, acidity (pH), somatic cell count and differential somatic cell count. During the control milking period, milk samples of the Red Gorbatov cows were individually collected and preserved using Microtabs tablets (USA). The analysis of milk components in Red Gorbatov cows was carried out at the federal research center of livestock farming-VIZH named after academician L.K. Ernst based on ONIS Biotech using the automatic analyzer CombiFoss 7 DC ("FOSS", Denmark).

To identify associations between SNP markers and milk components, a multiple linear regression analysis was conducted using PLINK version 1.90. In order to confirm the significant effects of SNPs and identify significant regions in the genomes of the animals under study, a test for the null hypothesis was performed using Bonferroni correction with a significance threshold of $p < 1.06 \times 10^{-6}$, 0.05/47155. Data visualization was carried out in the qqman package using the R-studio platform. SNPs corresponding to a significance level of $p \leq 0.000001$ were classified as genome-wide and $p \leq 0.00001$ as suggestive (presumed).

The "CattleQTLdb" database was employed for the purpose of identifying Quantitative Trait Loci (QTL) and regions subjected to evolutionary pressure on the chromosomes, which are intricately linked with the functional attributes of animals (Hu *et al.*, 2019).

The elimination of environmental and constant effects using generalized linear models, in order to analyze the normal distribution of the studied features, was carried out using the STATISTICA 10 program.

Results

The use of descriptive statistics for milk parameters studied by the method of infrared spectroscopy helps to better interpret the obtained results. Table (1) presents the data of the statistical processing of the investigated sample of Red Gorbatov cow breeds.

The analysis of the indicated Coefficient of variation (Cv, %) shows that the least variability was characteristic of the following parameters: Acidity, freezing point, SCC, and lactose, with a relative standard deviation not exceeding 5.5%. The average level of variability was established for such parameters as Milk Fat, β -casein, dry matter, and SCC, with a coefficient of variation not exceeding 17%. Higher values were obtained for urea (26.44%), daily milk yield (38.6%), somatic cell count (40.93%), acetone (53.66%), and BHB (60.94%). The most significant variability was demonstrated by SCC-164.3%. The range of variability for Milk Fat, β -casein, dry matter, SCC, and urea indicated a normal distribution of data characteristics in the investigated sample, while daily milk yield, Somatic Cell Count, acetone, and BHB showed a wider range of variability. Extremely high variability in the Somatic Cell Count (SCC) indicates the presence of subclinical and clinical mastitis in some individuals, with a maximum value for this parameter exceeding 5 million units/mL.

Table 1: The descriptive statistics for the milk productivity parameters of the sample of Red Gorbatov cows under study

Indicators	Min	Max	Average	$\pm m$	Cv, %
Daily milk yield, kg	1.700	15.100	6.400	0.400	38.60
Protein, %	2.570	3.920	3.120	0.060	12.09
Milk fat, %	1.100	6.160	3.160	0.220	40.93
β -casein, %	2.160	3.280	2.590	0.050	12.32
Lactose, %	3.620	5.060	4.640	0.040	5.29
Dry matter, %	9.460	14.800	11.780	0.240	11.63
SNF, %	7.860	9.720	8.730	0.080	5.04
BHB, mmol/l, mg $\times 100$ mL ⁻¹	0.020	0.305	0.080	0.008	60.94
Acetone, mmol/l, mg $\times 100$ mL ⁻¹	0.025	0.290	0.133	0.012	53.66
Urea, mmol/l, mg $\times 100$ mL ⁻¹	10.500	33.000	17.410	0.790	26.44
Freezing point, $-1 \times 10^{-3} \text{ }^\circ\text{C}$	523.500	558.000	534.870	1.380	1.51
pH	6.350	6.750	6.580	0.010	1.16
SCC, thousand units/ml	76.000	5600.500	742.880	209.320	164.30
DSCC, %	32.500	87.500	69.680	2.000	16.71

The obtained data on the normal distribution of the aforementioned parameters allowed for conducting whole-genome association studies to identify SNPs associated with these parameters. After data filtration, 42,275 SNPs were used in the analysis. Figure (1) illustrates the findings of visualizing the positions of statistically significant single-nucleotide polymorphisms (SNPs) on the 29 autosomes of Red Gorbatov cows for some of the studied traits.

GWAS enabled the identification of 44 genome-wide and 178 suggestive single nucleotide polymorphisms (SNPs) associated with the investigated milk parameters

of Red Gorbatov cattle (Table 2). For parameters such as urea, acidity, and daily yield, only suggestive associations were determined, located on chromosomes 1, 2, and 9, respectively. Milk yield and β -casein content were linked to 26 SNPs distributed across 13 chromosomes (2, 5, 7, 8, 12, 13, 15, 16, 18, 19, 20, 23, 27), with genome-wide associations identified for 2 and 3 SNPs, respectively, distributed across 3 chromosomes (15, 18, 20). Eight SNPs were associated with fat content, localized on 6 chromosomes (5, 8, 9, 20, 25, 28), among which only 1 SNP exhibited a genome-wide association, positioned on chromosome 25.

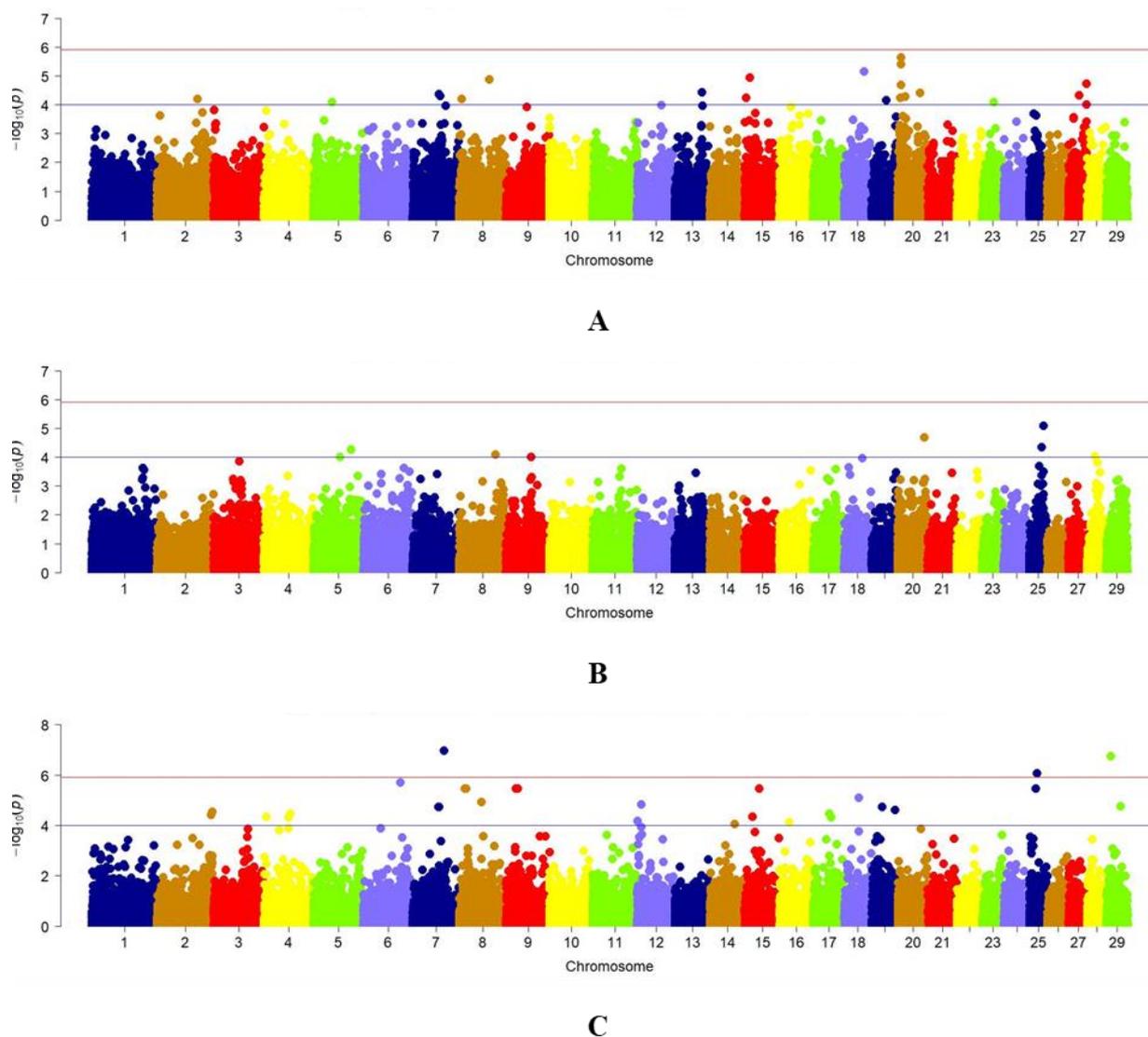


Fig. 1: The distribution of statistically significant single nucleotide polymorphisms (SNPs) in 29 autosomes of Red Gorbatov cattle for traits: A-mass fraction of protein; B-mass fraction of fat; C-somatic cell count. The logarithm of the significance level (q-value) is plotted on the Y-axis for each chromosome on the X-axis. The lower line on the Y-axis corresponds to a significance level of $p \leq 0.00001$ and the upper line to $p \leq 0.000001$

Table 2: Number of SNPs significantly associated with milk productivity parameters in Red Gorbatov cattle

Index	Genome-wide		Suggestive	
	Number	Chromosome distribution	Number	Distribution by chromosomes
Daily milk yield	-	-	9	3, 5, 13, 18, 20, 24
Protein	3	18, 20	20	2, 5, 7, 8, 13, 15, 19, 20, 23, 27
Milk fat	1	25	7	5, 8, 9, 20, 25, 28
β-casein	5	15, 18, 20	18	5, 7, 8, 12, 13, 15, 16, 19, 20, 27
Lactose	4	4, 12, 17, 18	8	5, 9, 12, 18, 20, 25, 29
Dry matter	1	20	11	5, 8, 18, 20, 22, 25, 28, 29
SNF	1	20	10	2, 3, 7, 11, 15, 20, 27
BHB	8	2, 4, 8, 12, 17, 25, 26	35	2, 3, 4, 5, 6, 7, 10, 18, 19, 20, 24, 25, 27
Acetone	2	6, 14	28	2, 5, 6, 7, 8, 10, 11, 13, 15, 16, 18, 19, 24, 27, 28
Urea	-	-	1	28
Freezing point	3	3, 26	7	1, 4, 14, 17, 20, 21, 27
Acidity	-	-	2	16, 22
SCC	15	6, 7, 8, 9, 13, 15, 18, 25, 29	19	2, 4, 7, 8, 12, 14, 15, 16, 17, 19, 29
DSCC	1	23	3	7, 26, 28

Associations with the content of metabolic products, such as BHBA and acetone in the milk of Red Gorbatov cattle, were detected for 43 and 30 SNPs, respectively, distributed across 23 chromosomes (for BHBA: 2, 3, 4, 5, 6, 7, 8, 10, 12, 17, 18, 19, 20, 24, 25, 26, 27; for acetone: 2, 5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 18, 19, 24, 27, 28). Specifically, 8 genome-wide SNPs were associated with BHBA content and 2 SNPs were associated with acetone content, localized across 2, 4, 8, 12, 17, 25, 26, and 6, 14 chromosomes, respectively. Associations with the ketosis indicator were demonstrated by 34 SNPs, including 15 genome-wide SNPs, distributed across 9 chromosomes (6, 7, 8, 9, 13, 15, 18, 25, 29).

The largest number of genome-wide SNPs for the studied traits was identified on chromosomes 18 and 20 - 6 in both cases. Of the 6 significant SNPs on chromosome 18, 1 SNP (ARS-BFGL-NGS-18019) was common to two traits (Milk Fat and β-casein content), while the remaining 4 SNPs (ARS-BFGL-NGS-34817, BTA-42999-no-rs, UA-IFASA-5348, ARS-BFGL-NGS-29060) were not shared among the examined parameters.

Among the 6 significant SNPs on chromosome 20, 2 SNPs (BTB-01636657 and ARS-BFGL-NGS-91235), similar to the previous case, were common for two traits (Milk Fat and β-casein content). Additionally, 2 other SNPs (ARS-BFGL-NGS-6953 and ARS-BFGL-NGS-14107) were associated with parameters such as dry matter and SCC (Table 3).

The structural annotation of genomic regions covering the window of ±0.20 Mb around the identified SNP was performed using the Ensembl Genome Browser (<http://oct2018.archive.ensembl.org/index.html>) within the biological library, revealing the presence of 20 functional genes (Table 4).

The analysis of biological functions of candidate genes revealed their categorization into several main functions related to:

- Mineral content in milk: Calcium content in milk (ODAM, HS3ST2), magnesium (ODAM), potassium (SLIT2)
- Reproductive functions: Ease of calving, calving ease (AAGAB, DTNA), stillbirth (HMCN1, DTNA), insemination rate (GALK2), first insemination post-calving (USH2A), conception insemination (USH2A, GNAQ, UNC80), the interval from first to last insemination (SLIT2)
- Milk production traits and technological qualities: Mass fraction of milk protein (ADARB2, TBC1D22A, SLIT2, ZFPM2, ARHGEF28), milk protein yield, milk yield (LRRC8D), comprehensive milk production index (OTOG), conjugated linoleic acid content in milk (UNC80), milk fat yield (ARHGEF28), C14 milk index (GRIA1), myristoleic acid content (GRIA1), somatic cell counts evaluation (C1QTNF3), pentacyclic acid content, milking speed (CTNND2)
- Meat production traits: Average daily gain, subcutaneous fat thickness (HMCN1), carcass weight (CTNNBL1)
- Disease resistance: Hoof and leg diseases (PDE6A), susceptibility of cattle to respiratory diseases, susceptibility to *M. Paratuberculosis* (ANKRA2), susceptibility of cattle to tuberculosis (ABCA13, CTNND2)
- Conformation and phenotype: Rump angle (AAGAB), udder cleft (AAGAB, DTNA), udder index (SWAP70), rear leg set - side view (HMCN1, DTNA), body depth, mammary shapes (type), teat length, udder height (DTNA), eye area pigmentation (ATRN), conformation assessment (CTNND2)
- Thyroxine level in blood (CTNND2)
- Productive life duration (AAGAB, HMCN1, DTNA)
- General breeding value (AAGAB, HMCN1)
- Residual feed intake (HMCN1)

Table 3: Genome-wide SNPs associated with indicators composition of milk of the Red Gorbatov breed

CHR	SNP	Position	P
Protein			
18	ARS-BFGL-NGS-18019	48538054	6.88E-06
20	BTB-01636657	7626481	3.98E-06
20	ARS-BFGL-NGS-91235	7543450	2.23E-06
Milk fat			
25	ARS-BFGL-NGS-32883	33773540	7.90E-06
β-casein content			
15	ARS-BFGL-NGS-114291	14242668	9.85E-06
15	ARS-BFGL-NGS-117603	14272339	9.85E-06
18	ARS-BFGL-NGS-18019	48538054	8.03E-06
20	BTB-01636657	7626481	3.10E-06
20	ARS-BFGL-NGS-91235	7543450	3.02E-06
Lactose content			
4	ARS-BFGL-NGS-38567	60288494	7.81E-07
12	BTB-01678450	1681182	1.79E-06
17	BTB-00677011	44164743	7.81E-07
18	ARS-BFGL-NGS-34817	8384452	2.76E-06
Dry matter			
20	ARS-BFGL-NGS-6953	64423007	1.56E-06
SNF			
20	ARS-BFGL-NGS-14107	7007017	4.36E-06
BHB			
2	ARS-BFGL-NGS-67929	104555063	4.91E-06
2	ARS-BFGL-NGS-23688	33023842	9.10E-06
4	ARS-BFGL-NGS-38567	60288494	2.59E-09
8	BTA-92138-no-rs	14394671	7.29E-06
12	BTB-01678450	16811820	1.14E-09
17	BTB-00677011	44164743	2.59E-09
25	Hapmap48251-BTA-21977	17998799	5.88E-06
26	Hapmap39763-BTA-61122	30678267	4.28E-06
Acetone			
6	Hapmap33631-BTC-043555	87327708	2.54E-06
14	Hapmap54994-rs29026820	19849178	3.28E-06
Freezing point			
3	BTA-120086-no-rs	101879077	9.74E-07
26	Hapmap52045-rs29025349	41544425	8.27E-06
26	Hapmap36060-SCAFFOLD71490_809	41154335	8.27E-06
SCC			
6	BTB-01946648	89572513	2.05E-06
7	BTB-00327936	76349643	1.11E-07
8	Hapmap39518-BTA-82980	13929541	3.51E-06
8	BTA-83016-no-rs	16099898	3.51E-06
9	ARS-BFGL-NGS-10167	23630089	3.51E-06
9	ARS-BFGL-NGS-110238	28954785	3.51E-06
13	ARS-BFGL-NGS-7342	60714542	7.28E-20
15	BTB-01444568	36211338	3.51E-06
15	ARS-BFGL-NGS-67096	35480232	3.51E-06
18	BTA-42999-no-rs	35085008	8.21E-06
18	UA-IFASA-5348	35027876	8.21E-06
18	ARS-BFGL-NGS-29060	35163513	8.21E-06
25	Hapmap48251-BTA-21977	17998799	8.79E-07
25	Hapmap25008-BTA-109486	14464084	3.51E-06
29	Hapmap49260-BTA-66294	99631080	1.77E-07
DSCC			
23	Hapmap40179-BTA-56624	46063916	3.74E-06

Table 4: Candidate genes associated with milk composition parameters of Red Gorbatov cattle

Signs	Chromosome	SNP	Genes ^{Positions}
Protein, casein	2	ARS-BFGL-NGS-3990 ⁹⁸²¹⁷⁵⁹⁸	<i>UNC80</i> ^{98024633...98250962}
	7	BTB-00314778 ^{63255889...66921835}	<i>PDE6A</i> ^{63190542...63267311}
	20	Hapmap25537-BTA-134848 ^{7993223...8042175}	<i>ARHGGEF28</i> ^{7604086...7716625}
Casein, protein, SNF	7	BTA-79668-no-rs ⁶⁶⁹²¹⁸³⁵	<i>GRIA1</i> ^{66766525...67109288}
	20	ARS-BFGL-NGS-101215 ^{8042175...7993223}	<i>ANKRA</i> ^{7992719...8003011}
Casein, SCC	4	BTB-01580567 ⁷¹⁸⁶⁹⁷¹	<i>ABCA13</i> ^{7096630...7487703}
Acetone	6	BTB-00264506 ⁷⁹⁷³²¹²⁹	<i>ODAM</i> ^{87327028...87335880}

Table 4: Cont.

	10	BTB-00410309 ¹⁴⁰³⁷⁹⁰⁷	<i>AAGAB</i> ^{13992197...14074945}
	13	ARS-BFGL-NGS-35887 ⁴⁶⁴³³⁶⁹⁷	<i>ADARB2</i> ^{46383731...46630127}
	15	ARS-BFGL-NGS-115625 ⁴³⁵³⁸⁸⁶⁶	<i>SWAP70</i> ^{43509670...43585976}
	16	BTB-00654373 ⁶⁸⁴¹⁴⁶³⁷	<i>HMCN1</i> ^{68192901...68733392}
	24	Hapmap51000-BTA-57638 ²²⁵¹⁹⁰⁶⁰	<i>DTNA</i> ^{22445691...22767026}
Casein	13	ARS-BFGL-NGS-65111 ^{66623395...67264571}	<i>CTNBL1</i> ^{67257419...67431111}
Lactose	5	ARS-BFGL-NGS-29924 ¹¹⁸⁰⁸⁹²⁷⁵	<i>TBC1D22A</i> ^{118086468...118343833}
pH	10	ARS-BFGL-NGS-101983 ⁶¹²⁰¹⁰²³	<i>GALK2</i> ^{61139588...61295379}
	16	BTA-40330-no-rs ²⁰⁰⁷⁷⁹³⁷	<i>USH2A</i> ^{19573856...20502175}
SCC	6	Hapmap33430-BTC-037618 ⁴¹⁵⁸⁸⁸⁴⁷	<i>SLIT2</i> ^{41236589...41640789}
	8	BTA-38269-no-rs ⁵⁴¹²⁹³⁵⁸	<i>GNAQ</i> ^{53970972...54280697}
	14	ARS-BFGL-BAC-21623 ⁶¹⁰³⁴¹¹³	<i>ZFPM2</i> ^{60994139...61520208}
	15	BTB-01444568 ^{36211338...35480232}	<i>OTOG</i> ^{35409567...35496594}

Discussion

The search for candidate genes associated with performance traits is an important step in further breeding work to preserve the gene pool of the Red Gorbatov breed. The discovery of new genome-wide associations with milk production traits will allow for more efficient selection of these animals at the genetic level.

A genome-wide analysis of different cattle breeds conducted by Saravanan *et al.* (2021) identified several candidate genes linked to various traits. In particular, the ADARB2 gene was linked to milk production in breeds such as Ayrshire, Holstein, and Jersey. Our study found that this gene, located on chromosome 13, was associated with the concentration of acetone in milk. We also found that the ARHGEF28 gene is associated with indicators of milk production, such as protein mass fraction, casein content, and the number of somatic cells. Previous work by Forutan *et al.* (2022) suggested a connection between this gene and fertility in Brahman cattle. In the study of associations between single nucleotide polymorphisms (SNPs) and carcass traits in Nellore cattle, Espigolan *et al.* (2015) found a number of SNPs located close to the CTNBL1 gene. This gene is involved in the formation of the spliceosome, which plays a role in basal metabolism. In our study, one specific SNP, ARS-BFGL-NGS-65111, located close to the CTNBL1 gene, was associated with casein content in milk. A genome-wide analysis of the Swedish Red breed (Tarekegn *et al.*, 2021), which, like the Red Gorbatov, belongs to the group of red cattle breeds, found that the GNAQ gene was linked with a fertility trait, such as the first luteal phase length. At the same time, we have identified an association between this candidate gene and the number of somatic cells. SNP BTB-00318021 (BTA7), located within the GRIA1 gene and associated with milk fat yield, was identified in Holstein cattle by Zielke *et al.* (2013). Genome-wide association studies in the Red Gorbatov breed identified SNP BTA-79668, located close to the GRIA1 gene on chromosome 7, which is also associated with milk productivity traits such as protein mass fraction, casein

content, and SNF. The candidate gene ZFPM2 is of particular interest, as a study by Cai *et al.* (2020) of Nordic Holstein cattle found a link between this gene and resistance to mastitis in cows, while our data indicates that the SNP ARS-BFGL-BAC-21623 located within this gene on chromosome 14 is also associated with somatic cell counts in milk.

Conclusion

Genome-wide association studies have identified 44 genome-wide and 178 suggestive SNPs associated with milk traits in Red Gorbatov cattle. The highest number of genome-wide SNPs were located on chromosomes 18 and 20-6 in both cases. On chromosome 18, one SNP (ARS-BFGL-NGS-18019) was common for two traits (Milk Fat and β -casein content). On chromosome 20, 2 SNPs (BTB-01636657 and ARS-BFGL-NGS-91235) were common for two traits (Milk Fat and β -casein content).

Structural annotation of the genomic regions revealed 20 genes with described functions in terms of gene ontology. Genes ODAM, HS3ST2, SLIT2, ADARB2, TBC1D22A, ZFPM2, ARHGEF28, LRRC8D, OTOG, UNC80, GRIA1, C1QTNF3, CTNND2 were associated with the chemical composition of milk and animal milk productivity.

Acknowledgment

The authors would like to express their sincere gratitude to the Russian State Agrarian University Moscow Timiryazev Agriculture Academy, which generously covered the publication costs for this article.

Funding Information

The study was conducted as part of the comprehensive project "Scientific and Technological Frontiers" of the Strategic Academic Leadership Program "Priority-2030" on the topic "Biotechnological methods of reproduction and genomic technologies in the breeding of farm animals and the preservation of the gene pool of indigenous breeds."

Author's Contributions

Marina Ivanovna Selionova: Conceptualization and funding acquisition, methodology condition the study.

Grigorii Sergeevich Shekhovtsev: Methodology, conducting the study, data curation, writing original draft preparation.

Ivan Petrovich Prokhorov and Vladimir Nikolaevich Lukyanov: Writing review and editing supervision.

Yulia Vasilevna Shoshina: Project administration.

Ethics

This article is original and has never been published before. The author also confirmed with all authors involved in the study that they had read and agreed to the contents of the article and that there were no ethical issues involved.

Conflict of Interest

All authors declare that they have no conflicts of interest.

References

- Bertelsen, H. P., Gregersen, V. R., Poulsen, N., Nielsen, R. O., Das, A., Madsen, L. B., Buitenhuis, A. J., Holm, L.-E., Panitz, F., Larsen, L. B., & Bendixen, C. (2016). Detection of Genetic Variation Affecting Milk Coagulation Properties in Danish Holstein Dairy Cattle by Analyses of Pooled Whole-Genome Sequences from Phenotypically Extreme Samples (Pool-Seq). *Journal of Animal Science*, *94*(4), 1365–1376. <https://doi.org/10.2527/jas.2015-9884>
- Cai, Z., Dusza, M., Guldbrandtsen, B., Lund, M. S., & Sahana, G. (2020). Distinguishing pleiotropy from linked QTL between milk production traits and mastitis resistance in Nordic Holstein cattle. *Genetics Selection Evolution*, *52*, 1-15. <https://doi.org/10.1186/s12711-020-00538-6>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *Giga Science*, *4*(1), s13742-015. <https://doi.org/10.1186/s13742-015-0047-8>
- Cruz, V. A. R., Oliveira, H. R., Brito, L. F., Fleming, A., Larmer, S., Miglior, F., & Schenkel, F. S. (2019). Genome-Wide Association Study for Milk Fatty Acids in Holstein Cattle Accounting for the DGAT1 Gene Effect. *Animals*, *9*(11), 997. <https://doi.org/10.3390/ani9110997>
- Espigolan, R., Baldi, F., Boligon, A. A., Souza, F. R., Fernandes Júnior, G. A., Gordo, D. G., & Albuquerque, L. G. (2015). Associations between single nucleotide polymorphisms and carcass traits in Nelore cattle using high-density panels. *Genet Mol Res*, *14*, 11133-11144. <https://doi.org/10.4238/2015.September.22.7>
- Fang, Z. H., Bovenhuis, H., van Valenberg, H. J. F., Martin, P., Duchemin, S. I., Huppertz, T., & Visker, M. H. P. W. (2019). Genome-Wide Association Study for α S1- and α S2-Casein Phosphorylation in Dutch Holstein Friesian. *Journal of Dairy Science*, *102*(2), 1374–1385. <https://doi.org/10.3168/jds.2018-15593>
- Forutan, M., Engle, B., Goddard, M. E., & Hayes, B. J. (2022, December). A conditional multi-trait sequence GWAS of heifer fertility in tropically adapted beef cattle. In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges* (pp. 1106-1109). Wageningen Academic Publishers. <https://doi.org/10.3920/978-90-8686-940-4>
- Hu, Z.-L., Park, C. A., & Reecy, J. M. (2019). Building a Livestock Genetic and Genomic Information Knowledgebase through Integrative Developments of Animal QTLdb and CorrDB. *Nucleic Acids Research*, *47*(D1), D701–D710. <https://doi.org/10.1093/nar/gky1084>
- Ilie, D. E., Mizeranschi, A. E., Mihali, C. V., Neamț, R. I., Goilean, G. V., Georgescu, O. I., Zaharie, D., Carabaș, M., & Huțu, I. (2021). Genome-Wide Association Studies for Milk Somatic Cell Score in Romanian Dairy Cattle. *Genes*, *12*(10), 1495. <https://doi.org/10.3390/genes12101495>
- Lijun, S., Xin, W., Yuze, Y., Zhu, M., Xiaoqing, L., Lin, L., Yanhua, L., Feng, Z., Bo, H., & Dongxiao, S. (2021). A post-GWAS Confirming the Genetic Effects and Functional Polymorphisms of AGPAT3 Gene on Milk Fatty Acids in Dairy Cattle. *Journal of Animal Science and Biotechnology*, *12*(1), 24. <https://doi.org/10.1186/s40104-020-00540-4>
- Nayeri, S., Schenkel, F., Fleming, A., Kroezen, V., Sargolzaei, M., Baes, C., Cánovas, A., Squires, J., & Miglior, F. (2019). Genome-Wide Association Analysis for β -Hydroxybutyrate Concentration in Milk in Holstein Dairy Cattle. *BMC Genetics*, *20*(1), 58. <https://doi.org/10.1186/s12863-019-0761-9>
- Oliveira, H. R., Cant, J. P., Brito, L. F., Feitosa, F. L. B., Chud, T. C. S., Fonseca, P. A. S., Jamrozik, J., Silva, F. F., Lourenco, D. A. L., & Schenkel, F. S. (2019). Genome-wide association for milk production traits and somatic cell score in different lactation stages of Ayrshire, Holstein, and Jersey dairy cattle. *Journal of Dairy Science*, *102*(9), 8159–8174. <https://doi.org/10.3168/jds.2019-16451>
- Otto, P. I., Guimarães, S. E. F., Calus, M. P. L., Vandenplas, J., Machado, M. A., Panetto, J. C. C., & da Silva, M. V. G. B. (2020). Single-Step Genome-Wide Association Studies (GWAS) and Post-GWAS Analyses to Identify Genomic Regions and Candidate Genes for Milk Yield in Brazilian Girolando Cattle. *Journal of Dairy Science*, *103*(11), 10347–10360. <https://doi.org/10.3168/jds.2019-17890>

- Sanchez, M.-P., Ramayo-Caldas, Y., Wolf, V., Laithier, C., El Jabri, M., Michenet, A., Boussaha, M., Taussat, S., Fritz, S., Delacroix-Buchet, A., Brochard, M., & Boichard, D. (2019). Sequence-Based GWAS, Network and Pathway Analyses Reveal Genes Co-Associated with Milk Cheese-Making Properties and Milk Composition in Montbéliarde Cows. *Genetics Selection Evolution*, 51(1), 34.
<https://doi.org/10.1186/s12711-019-0473-7>
- Sanchez, M.-P., Rocha, D., Charles, M., Boussaha, M., Hozé, C., Brochard, M., Delacroix-Buchet, A., Groperrin, P., & Boichard, D. (2021). Sequence-Based GWAS and Post-GWAS Analyses Reveal a Key Role of SLC37A1, ANKH, and Regulatory Regions on Bovine Milk Mineral Content. *Scientific Reports*, 11(1), 7537. <https://doi.org/10.1038/s41598-021-87078-1>
- Saravanan, K. A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G. K., ... & Singh, R. K. (2021). Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. *Genomics*, 113(3), 955-963.
<https://doi.org/10.1016/j.ygeno.2021.02.009>
- Sermyagin, A. A., Bykova, O. A., Loretts, O. G., Kostyunina, O. V., & Zinovieva, N. A. (2020). Genomic Variability Assess for Breeding Traits in Holsteinized Russian Black-And-White Cattle Using Gwas Analysis and Roh Patterns. *Agricultural Biology*, 55(2), 257–274.
<https://doi.org/10.15389/agrobiology.2020.2.257eng>
- Silva, A. A., Silva, D. A., Silva, F. F., Costa, C. N., Silva, H. T., Lopes, P. S., Veroneze, R., Thompson, G., & Carnevali, J. (2020). GWAS and Gene Networks for Milk-Related Traits from Test-Day Multiple Lactations in Portuguese Holstein Cattle. *Journal of Applied Genetics*, 61(3), 465–476.
<https://doi.org/10.1007/s13353-020-00567-3>
- Slagboom, M., Milkevych, V., Liu, H., Thomasen, J. R., Kargo, M., & Schmidtman, C. (2022). Conservation of Local Red Cattle breeds by Collaboration with a Mainstream Red Dairy Cattle Breed. *Livestock Science*, 260, 104936.
<https://doi.org/10.1016/j.livsci.2022.104936>
- Tarekegn, G. M., Strandberg, E., Andonov, S., Båge, R., Ask-Gullstrand, P., Rius-Vilarrasa, E., ... & Berglund, B. (2021). Single-step genome-wide association study uncovers known and novel candidate genomic regions for endocrine and classical fertility traits in Swedish Red and Holstein dairy cows. *Livestock Science*, 253, 104731.
<https://doi.org/10.1016/j.livsci.2021.104731>
- Volkova, V. V., Romanenkova, O. S., Deniskova, T. E., Mishina, A. I., Kostyunina, O. V., & Zinovieva, N. A. (2019). *Characteristics of the Allele Pool of the Kholmogorsk Breed of Cattle Using Str Markers*.
<https://doi.org/10.33943/mms.2019.7.42224>
- Yudin, N. S., Yurchenko, A. A., & Larkin, D. M. (2021). Traces of selection and candidate genes for adapting to extreme environmental factors in the genomes of Turano-Mongolian breeds of cattle. *Vavilov Journal of Genetics and Breeding*, 25(2), 190-201.
<https://doi.org/10.18699/VJ21.023>
- Zhou, C., Li, C., Cai, W., Liu, S., Yin, H., Shi, S., ... & Zhang, S. (2019a). Genome-wide association study for milk protein composition traits in a Chinese Holstein population using a single-step approach. *Frontiers in genetics*, 10, 72.
<https://doi.org/10.3389/fgene.2019.00072>
- Zhou, J., Liu, L., Chen, C. J., Zhang, M., Lu, X., Zhang, Z., ... & Shi, Y. (2019b). Genome-wide association study of milk and reproductive traits in dual-purpose Xinjiang Brown cattle. *BMC genomics*, 20, 1-11.
<https://doi.org/10.1186/s12864-019-6224-x>
- Zielke, L., Bortfeldt, R., Tetens, J., Thaller, G., & Brockmann, G. (2013). The role of obesity genes for milk fat yield in Holstein dairy cattle. In *10th World Congress on Genetics Applied to Livestock Production*.