

GENOME-WIDE ASSOCIATION STUDY FOR DAIRY TRAITS IN SLOVENIAN BROWN SWISS BREED

Marija Špehar^{1,2*}, Vesna Mrak², Anamarija Smetko¹, Klemen Potočnik², Gregor Gorjanc^{2,3}

¹Croatian Agricultural Agency, Ilica 101, 10000 Zagreb, Croatia; ²Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Groblje 3, 1230 Domžale, Slovenia; ³The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, Scotland, UK

*Corresponding author, E-mail: mspehar@hpa.hr

Summary: Genome-wide association studies (GWAS) based on thousands of single nucleotide polymorphisms (SNPs) allows exploring genes associated with economically important traits. The objective of this study was to perform GWAS in small population of genotyped bulls of Slovenian Brown Swiss (BSW) breed to identify SNP markers associated with dairy traits: milk yield (MY), fat yield (FY), and protein yield (PY). The daughter yield deviations of 182 progeny tested bulls were used as phenotypes. For each bull, genotypes were scored for 34,450 SNPs across the whole genome. Single SNP analysis was performed to detect significant associations among SNPs and dairy traits across the genome. Two models were considered: 1) linear regression model considering one marker at a time, and 2) linear regression model with admixture components for accounting population stratification. The Bonferroni correction resulted in 52 significant SNPs using model 1). Correction for population stratification based on selected criteria was shown to have great impact on the analysis and then no significant associations among SNPs and dairy traits have been detected. Therefore, the results were sensitive to population stratification and together with the small data set did not give much power to accurately estimate associations. Further improvements should be made on the enlargement of number of genotyped bulls in order to detect association signals and the identification of genes associated with dairy traits in Slovenian BSW breed.

Key words: GWAS; dairy traits; single SNP analysis; population stratification

Introduction

In the dairy cattle production, all economically important traits are influenced by many genes that appear as continuous variation (1). The inclusion of molecular information i.e. thousands of single nucleotide polymorphisms (SNPs) in selection schemes made the use of this information in selection decisions feasible, through a technique commonly called genomic selection (2). Genomic selection improves the level

of genetic gain, compared to traditional selection based on progeny test, increased accuracy of EBV of young or non-phenotyped animals from about 0.60 (parent average) to 0.80 (3). Even though these accuracies are lower than with progeny test the early use of young bulls shortens generation interval considerably and therefore increases genetic gain per unit of time (4). Development of SNP marker panels also enabled genome wide association studies (GWAS) with a purpose to determine associations between DNA polymorphisms and phenotypes of interest and to identify chromosome regions that harbour genes

or regulatory elements related to traits (5). GWAS has been implemented in livestock species, such as cattle, to discover superior variants of all genes contributing to the phenotype of economically important dairy (6, 7, 8), meat (9), and health traits (10). Genomic selection and GWAS can be successfully implemented due to existing linkage disequilibrium (LD) between markers and QTL for the trait of interest (3).

The Brown Swiss (BSW) breed is a cosmopolitan breed selected for dairy characteristics and ability to produce in less favoured areas mostly in Alpine regions of Austria, Germany, Italy, Switzerland, and Slovenia. Original Brown breed (Braunvieh) from Switzerland was exported to the United States in 1869 (11), and after strong selection of this breed, previously classified as a dual-purpose breed, it was established as a dairy type and received a name Brown Swiss breed. Development of Slovenian BSW breed started at the beginning of 20th century when Original Brown breed bulls were imported from Austria and Switzerland with the manner to improve cows of local grey cattle. Lately also breeding heifers were imported. From the 1960s, the American BSW breeding bulls were used for improving cows of Brown breed in order to increase milk production (12). Bulls with 50% of American BSW genotype were imported from neighbouring countries. Moreover, five bulls with 100% American BSW genotype were also used for insemination. From 1976 to 1980 bulls were no longer imported but have been bred in Slovenia. The breed is mainly spread in the western part of Alpine area, as well as southern and northern part of Slovenia with harsh environmental conditions. According to the annual report (13), 6.95% of all active cattle (or 31,087 heads) belong to the BSW breed. It is the second most important breed for milk production (13). The average production is 5,554 kg of milk in the standard lactation with 4.06% of fat and 3.39% of protein content.

In dairy cattle, and BSW breed too, the population stratification may exist. It could be result of selection that caused allele frequency changes (14) and artificial insemination which increased the presence of related animals in a randomly selected samples and the presence of large half-sib families (15). A potential challenge in GWAS is the presence of population stratification (family substructure) that can result in false-positive or negative associations and may lead to confounding results (16, 17). In order to take into account population

stratification linear models have been suggested as a method of choice to reduce effect of false-positive or negative associations (18).

In this study, we conducted genome-wide association analysis of dairy traits: milk yield (MY), fat yield (FY), and protein yield (PY) in a small population of genotyped bulls of Slovenian BSW breed. We used the Illumina Bovine SNP50 Bead Chip to identify SNPs associated with dairy traits without and considering population stratification.

Material and methods

Phenotypes used were daughter yield deviations (DYDs) of dairy traits (MY, FY, and PY) taken from the routine national genetic evaluation. Altogether, DYDs for 182 progeny tested and genotyped BSW bulls born between 1990 and 2007 were included in the analysis. Bulls used in analysis predominantly originated from Slovenia. The exceptions were nine bulls imported to Slovenia as live animals from Austria and Germany. These bulls were used only in Slovenia. The DNA was extracted from semen samples and genotyped using the Illumina Bovine SNP50 Bead Chip (Illumina, San Diego, CA). Several quality controls were applied to the genotype data to investigate the integrity and informativeness of SNPs. These controls included genotyping call rate (CR) by SNP and by animal, minor allele frequency (MAF), departure from Hardy–Weinberg equilibrium (HWE) and parent–progeny conflicts (PPC). Data editing was carried out using SAS package (19). Number of SNPs excluded from the analysis differed among filtering criteria. SNPs were included in the analysis if CR was higher than 90% by the SNP. Altogether, 5,975 SNPs failed this criterion and were excluded from the analysis. SNPs with MAF lower than 0.05 (10,619 cases) were also excluded. The departure from HWE with a threshold of $P < 10^{-5}$ excluded 538 SNPs from data set. The mean of PPC was low ($< 0.0001\%$) and below the threshold (< 200 discrepancies) set for the data exclusion. SNPs that could not be mapped in BTA (1,672) or SNPs placed on X chromosome (747) were also excluded from the analysis. The final genotype file included 34,450 SNPs placed on 29 *Bos Taurus* autosomes (BTA) for 182 BSW bulls. Missing genotypes were imputed using the *gpgim* program (20) based on method that approximates the gene content (21). Two types of analyses were conducted to detect SNP association signals.

Initially, a single trait GWAS was used to estimate genome-wide associations between phenotypes and SNPs for each of dairy traits. GWAS between phenotypes and SNP markers for each dairy trait were estimated using the following univariate linear mixed model to estimate association for one marker at a time:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (1)$$

where \mathbf{y} is a vector of phenotype values (DYDs), \mathbf{b} is a vector that holds the intercept and allele substitution effect for the SNP marker of interest, while \mathbf{X} is the incidence matrix linking \mathbf{y} respectively with \mathbf{b} , and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$ is a vector of residuals. Then, admixture components adjusting for population stratification were included in the model as covariates.

Admixture analysis was performed on the data using Admixture software (22) to determine the proportion of genotypes originated from different clusters in the population (K). The proportion of each subpopulation was estimated as unsupervised analysis assuming from two to ten populations. Finally, bulls were divided in four clusters coming from different ancestral populations. Separation was determined using admixture cross-validation procedure, which gives standard error of the cross validation estimate. R software package (23) was used to test associations between SNPs and dairy traits. Significance threshold was calculated using the Bonferroni correction for multiple testing. Estimated effects were transformed to negative logarithm of p -values of significant effects. Correction based on p -value (0.05) resulted in a Bonferroni threshold of 1.45×10^{-6} . SNPs with $-\log(p\text{-values})$ above the Bonferroni threshold were considered significant.

Results

The distribution of observed distributions of $-\log(p\text{-value})$ for each SNP was compared to the expected distribution under no-association (red line) in a quantile-quantile (Q-Q) plot for each dairy trait (Figure 1). Q-Q plot of data without accounting population stratification (black line) for all analysed traits showed an early separation of the observed and expected distribution (red line) indicating that the model which neglected family relationship was problematic. The fact that the black dots are above the diagonal axis on a Q-Q plot shows an elevated amount of statistical

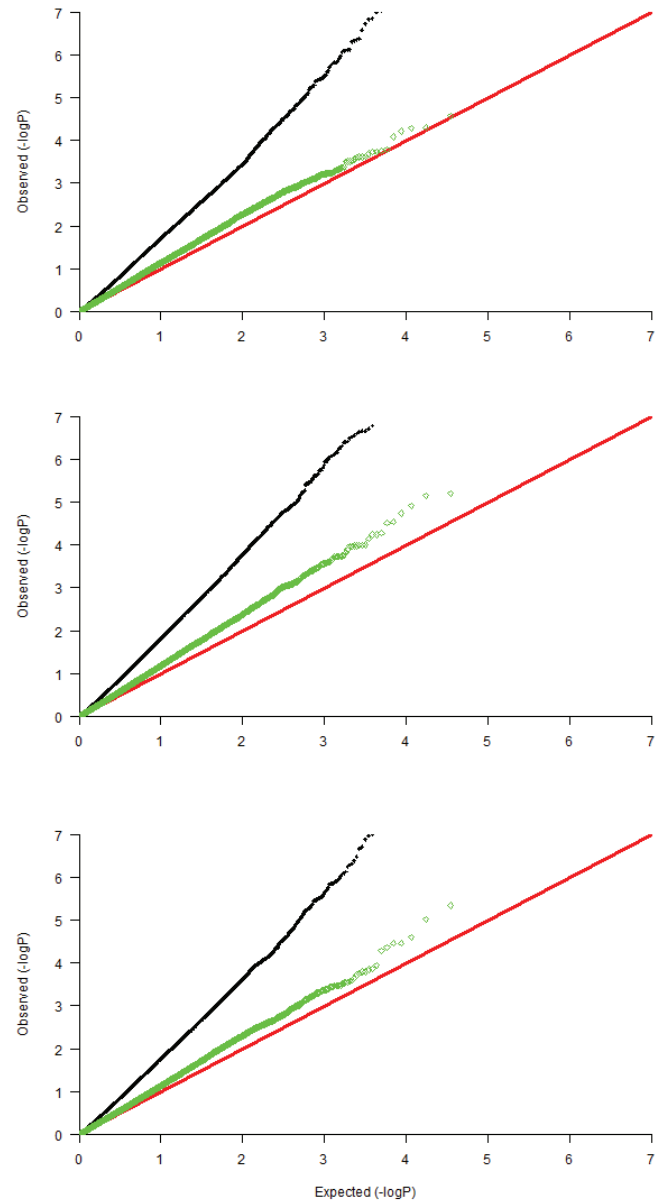


Figure 1: Q-Q plot for MY, FY, and PY. Red line is expected distribution, black dots are GWAS estimates without stratification, and green dots are GWAS estimates with stratification added to analysis

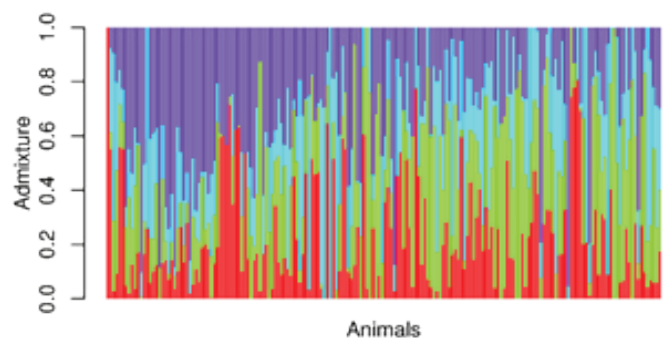


Figure 2: The proportion of subpopulations in each BSW bull estimated by Admixture analysis

significance for a huge number of variants, which is probably the result of population stratification. After correction, the data were well-behaved (green line) however for the majority of SNPs there was no more statistical significance.

For inclusion of population stratification components to the model, four ancestral populations were accepted as the best match describing population stratification due to lowest cross-validation error in $K=4$. Different genotype proportions per animal divided population in four clusters (Figure 2) which were added as a fixed effect, affecting SNP influence in GWAS analysis.

In total, 52 SNPs with a significant effect on MY, FY, and PY were identified by single trait analysis without considering population stratification.

Among these, 11 were significant for all dairy traits. The associations were spread over 14 chromosomes. The majority of SNPs associated with MY were placed on the BTA21 followed by BTA9. The majority of genome-wide significant SNPs for FY and PY were also located on BTA 21 and BTA9. Manhattan plots based on the negative logarithm of p -value (y axis) of each SNP versus its chromosomal locations (x axis) for MY, FY; and PY are shown in Figure 3. Horizontal dashed line indicates the Bonferroni-corrected significance level.

Correction for population stratification based on the selected criteria was shown to have large impact on the results of analysis. No significant associations between SNPs and dairy traits have been detected (Figure 3 - right side).

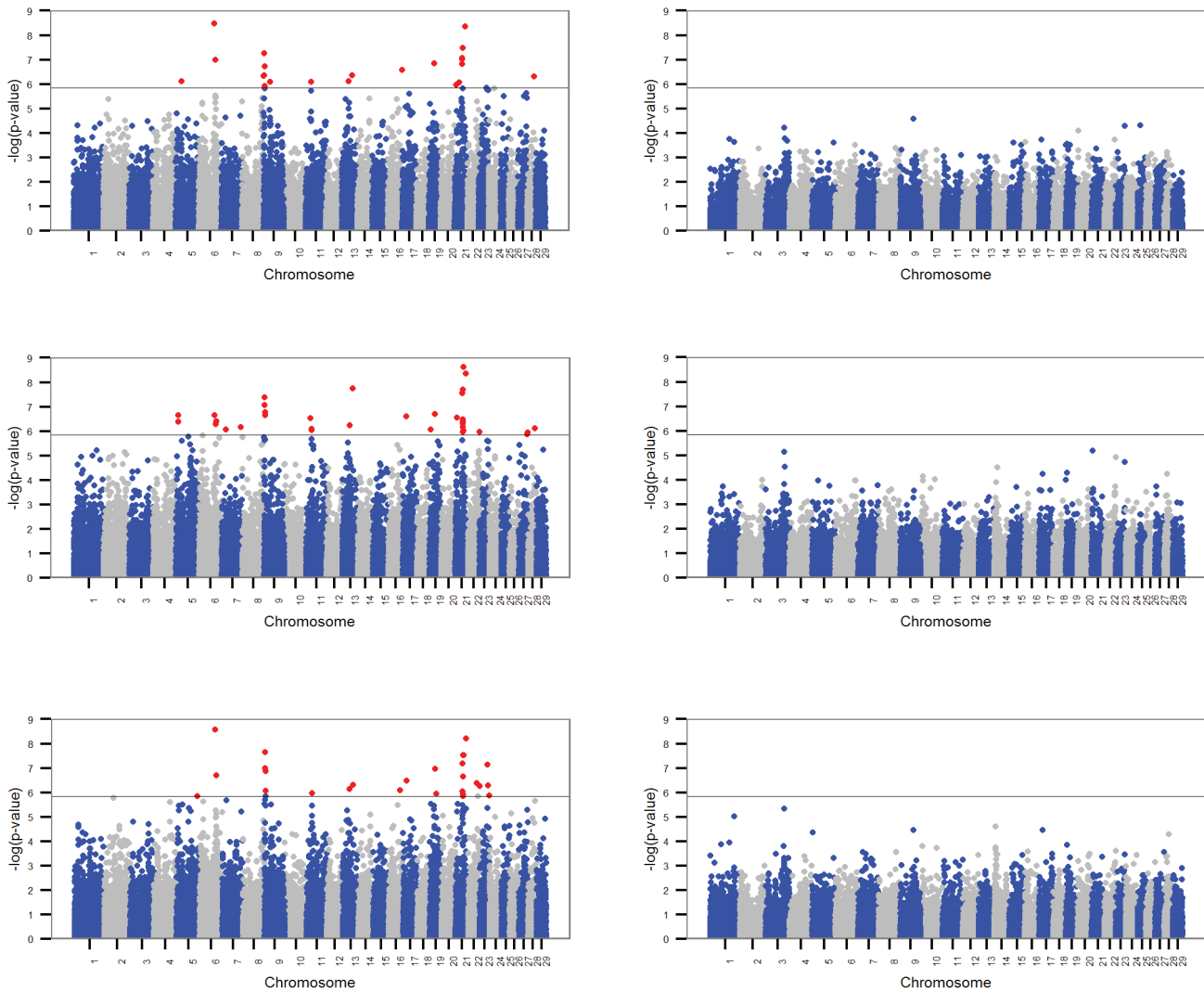


Figure 3: Manhattan plots for MY, FY, and PY using single SNP analysis without (left side) and with (right side) population stratification

Discussion

Single trait genome-wide analysis for dairy traits was presented to address the aspects of the GWAS for implementation of genomic selection in Slovenian BSW breed. The approach was used to estimate genome-wide associations between phenotypes and SNP markers based on data from national genetic evaluation. The effect of population stratification was taken into account since it was known from earlier publications (24, 26) that BSW breed is dealing with population stratification which has an impact on the results of GWAS. In the mentioned studies, principal component analysis was used to adjust the effect of population stratification. For studied population, animals were divided from two to ten ancestral clusters and according to the lowest cross validation error it was decided to take four clusters as best representation of stratification. After correction for population stratification, the data fitted more closely to the expected distribution on Q-Q plot (see Figure 1, green line) but then there were no statistically significant SNPs for investigated traits. Therefore, the results were sensitive to population stratification and together with the small sample size did not give much power to accurately estimate associations. Sample size must be increased by a factor of $1/r^2$ (where r^2 is a measure of linkage disequilibrium between pairs of biallelic markers) to detect an ungenotyped QTL, compared with the sample size for testing the QTL itself (25). Similar study but with higher number of bulls (554) of German Braunvieh (26) found five SNPs with the significant effect on dairy traits. Two SNPs which affected MY were placed on the BTA4, two SNPs associated FY were on the BTA14 and BTA23, while significant SNP with an effect on the fat content was on the BTA1. The largest data set (7,038 bulls) considering BSW breed (24) was based on large scale data from the international genetic evaluation. The total number of significant SNPs was 16 that affected dairy traits, with the strongest signal on the BTA25 (24). SNP associations for MY and PY were located on the same chromosome like in studies of Finnish Ayrshire (27) and Holstein (28) breed. GWAS in the dairy cattle mainly focused on the Holstein breed due to worldwide spreading (7). For MY, significant SNPs were located on BTA8, BTA9, BTA10, BTA11, BTA13, BTA25, and BTA29

(6, 8, 29). BTA14 has also been in focus since many studies reported DGAT1 as a major gene affecting milk traits (8, 29, 30). In this study, the association between SNPs and dairy traits was not detected in the DGAT1 region.

Conclusion

In the present study, the preliminary results from GWAS analysis were reported for dairy traits based on the small population of genotyped bulls of Slovenian BSW breed. Altogether, 52 SNPs with a significant effect on MY, FY, and PY were identified using single SNP analysis. After adjusting for population stratification, no significant associations between SNPs and dairy traits have been detected. Based on the analysis that considered afore mentioned criteria we could conclude that population stratification exists in the analysed data set and the model without adjustment was not dealing properly with the stratification in the data. Further improvements should be made on the enlargement of the number of genotyped bulls in order to detect association signals and the identification of genes associated with dairy traits. Results from GWAS will facilitate the understanding of the genetic architecture of complex traits.

References

1. Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Harlow; Essex: Longman Green, 1996: 98–20.
2. Meuwissen THE, Hayes BH, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 2001; 157: 1819–29.
3. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 2009; 92: 433–43.
4. Schaeffer LR. Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet* 2006; 123: 218–23.
5. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet* 2009; 10: 381–391.
6. Bolormaa S, Pryce JE, Hayes BJ, Goddard ME. Multivariate analysis of a genome-wide asso-

- ciation study in dairy cattle. *J Dairy Sci* 2010; 93: 3818–33.
7. Pryce JE, Bolormaa S, Chamberlain AJ, et al. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J Dairy Sci* 2010; 93: 3331–45.
8. Mai MD, Sahana G, Christiansen FB, Gulbrandsen B. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *J Anim Sci* 2010; 88: 3522–8.
9. Bolormaa S, Porto Neto LR, Zhang YD, et al. A genome-wide association study of meat and carcass traits in Australian cattle. *J Anim Sci* 2011; 89: 2297–309.
10. Murdoch BM, Murdoch GK, Settles M, McKay S, Williams JL, Moore SS. Genome-wide scan identifies loci associated with classical BSE occurrence. *PLoS ONE* 2011; 6(11): e26819 (7 p.) <http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0026819&representation=PDF> (10. Oct. 2014)
11. Yoder DM, Lush JL. A genetic history of the Brown Swiss cattle in the United States. *J Hered* 1937; 28: 154–60.
12. Ferčej J, Pogačar J, Lovšin N, Nendl B. Oplemenjevanje rjave pasme z ameriško rjavo. *Sodob Kmet* 1980; 4: 136–9.
13. Sadar M, Jenko J, Jeretina J, et al. Rezultati kontrole prireje mleka in mesa 2013. Ljubljana: Kmetijski inštitut Slovenije, 2014: 89 str. http://www.govedo.si/files/cpzgss/knjiznica/porocila/kontrola_porocila/REZULTATI_KONTROLE_2013.pdf (28. Oct. 2014)
14. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; 55: 997–1004.
15. Ma L, Wiggans GR, Wang S, et al. Effect of sample stratification on dairy GWAS results. *BMC Genomics* 2012; 13: e536 (18 p.) <http://www.biomedcentral.com/content/pdf/1471-2164-13-536.pdf> (16. Oct. 2014)
16. Kang HM, Zaitlen NA, Wade CM, et al. Efficient control for population structure in model organism association mapping. *Genetics* 2008; 178:1709–23.
17. Pritchard JK, Stephens M, Rosenberg N, Donnelly P. Association mapping in structured populations. *Am J Hum Genet* 2000; 67: 170–81.
18. Sonstegard TS, Ma L, Van Tassell CP, et al. Forty years of artificial selection in U.S. Holstein cattle had genome-wide signatures. In: 9th World Congress on Genetics Applied to Livestock Production: poster presentation. Leipzig, 2010. https://www.aipl.arsusda.gov/publish/presentations/WC9_10/WC9_10_yang_da.pdf (28. Oct. 2014)
19. SAS Institute. SAS/STAT Software, Version 9.3. Cary, NC, 2011.
20. Strandén I. Manual for gpig program: pedigree based imputation of genotypes. Jokioinen, Finland : Biotechnology and Food Research, Biometrical Genetics, MTT Agrifood Research Finland, 2010: 13 p.
21. Gengler N, Mayeres P, Szydlowski M. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal* 2007; 1: 21–8.
22. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009; 19: 1655–64.
23. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2011 ISBN 3-900051-07-0. <http://www.R-project.org> (28 Oct. 2014).
24. Guo J, Jorjani H, Carlborg O. 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: 82–92.
25. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 2001; 69: 1–14.
26. Maxa J, Neuditschko M, Russ I, Förster M, Medugorac I. Genome-wide association mapping of milk production traits in Braunvieh cattle. *J Dairy Sci* 2012; 95: 5357–64.
27. Viitala SM, Schulman NF, Koning DJ, et al. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *J Dairy Sci* 2003; 86: 1828–36.
28. Harder B, Bennewitz J, Reinsch N, et al. Mapping of quantitative trait loci for lactation persistency traits in German Holstein dairy cattle. *J Anim Breed Genet* 2006; 123: 89–96.
29. Jiang L, Liu J, Sun D, et al. Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS ONE* 2010; 5(10): e13661 (12 p.) <http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0013661&representation=PDF> (28.

Oct. 2014)

30. Grisart B, Coppieters W, Farnir F, et al. Positional candidate cloning of a QTL in dairy cattle:

identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res* 2002; 12: 222–31.

ASOCIACIJSKE ŠTUDIJE NA CELOTNEM GENOMU ZA LASTNOSTI MLEČNOSTI PRI SLOVENSKI RJAVI PASM

M. Špehar, V. Mrak, A. Smetko, K. Potočnik, G. Gorjanc

Povzetek: Asociacijske študije na celotnem genomu (GWAS), ki temeljijo na več tisoč polimorfizmih na posameznih nukleotidih (SNP), nam omogočajo, da lahko najdemo gene povezane z gospodarsko pomembnimi lastnostmi. Cilj raziskave je bil narediti GWAS z namenom identifikacije označevalcev SNP, ki so povezani z lastnostmi mlečnosti: količina mleka (MY), količina maščob (FY) in količina beljakovin (PY) v majhni populaciji genotipiziranih bikov slovenske rjave pasme. Kot fenotipske meritve smo uporabili odstopanja hčera 182 progeno testiranih bikov. Genotip za vsakega bika je bil poznan na 34,450 označevalcih SNP porazdeljenih po celem genomu. V študiji je bila narejena analiza polimorfizmov na posameznem označevalcu za odkrivanje označevalcev SNP, povezanih z lastnostmi mlečnosti na celotnem genomu. Uporabljena sta bila dva modela: 1) linearni regresijski model ob upoštevanju enega označevalca naenkrat in 2) komponente admixture analize, ki so bile vključene v prvi model, da se je upoštevala razdeljenost populacije. Bonferroni korekcija je pokazala 52 statistično značilnih označevalcev SNP, ki so temeljili na modelu 1). Korekcija na razdeljenost populacije na podpopulacije na podlagi izbranih kriterijev se je pokazala kot ključna. Z uporabo modela 2) nismo odkrili značilnih povezav med označevalci SNP in lastnostmi mlečnosti. Zaradi tega so bili rezultati občutljivi na razdeljenost populacije na podpopulacije in skupaj z majhnim številom podatkov niso veliko pripomogli k bolj natančni oceni povezav. Nadaljnje izboljšave bi lahko naredili s povečanjem števila genotipiziranih bikov, z namenom, da bi odkrili povezane signale in identificirali gene, povezane z lastnostmi mlečnosti pri slovenski rjavi pasmi.

Ključne besede: GWAS; lastnosti mlečnosti; analiza polimorfizmov na posameznem nukleotidu; razdeljenost populacije