Genome-wide Mixed Model Association Study in Population of Slovak Pinzgau Cattle

Radovan KASARDA [1(✉)]
Nina MORAVČIKOVÁ [1]
Juraj CANDRÁK [1]
Gábor MÉSZÁROS [2]
Michal VLČEK [1]
Veronika KUKUČKOVÁ [1]
Ondrej KADLEČIK [1]

Summary

The aim of this study was to detect the position of genomic regions associated with milk production and milk components in Pinzgau cattle. The dataset consisted of milk yield records of in total 7729 cows of 35 sires representing active bulls (19 animals) and DNA material stored in gene bank (16 animals) of Pinzgau cattle in Slovakia. In total 130087 test-day records of milk, protein and fat yield were used for the association analysis. The Illumina BovineSNP50 BeadChip V2 was used and after quality control final dataset consisted of 41,487 autosomal loci covering overall length 2,500,315 kb of the genome. Identification of genomic regions associated with milk production and composition were performed using GEMMA software with use of linear mixed model approach based on genetic-relationship matrix estimated from SNP genotypes to model correlation between the phenotypes. Based on this were found SNPs in the regions of important QTLs significantly associated with milk yield, dressing percentage, protein yield, SCS score, marbling score and fat yield.

Key words

test-day, milk yield, dressing percentage, SCS, QTL

[1] Slovak University of Agriculture in Nitra, Faculty of Agrobiology and Food Resources, Department of Animal Genetics and Breeding Biology, Tr. A. Hlinku 2, 94976 Nitra, Slovakia
✉ e-mail: radovan.kasarda@uniag.sk
[2] University of Natural Resources and Life Sciences Vienna, Division Livestock Science, Gregor Mendel Str. 33, 1180 Vienna, Austria

Received: May 2, 2017 | Accepted: August 16, 2017

ACKNOWLEDGEMENTS
This work has been supported by the Slovak Research and Development Agency (IDs No. APVV-14-0054 and SK-AT-2015-0016). Part of the work was done during the scientific stay of authors at BOKU Vienna, supported by the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH).

Introduction

Genome-wide association studies (GWAS) and genome analysis are powerful tools which provide insight into genetic variation of a wide range of traits important for humans, agriculture, conservation, and the evolutionary history of life. Possible use of the genomics and molecular-genetic markers for genetic evaluation, paternity testing, search for lethal recessive, inherited defects and mutations with important influence on production performance is the main aim of the research in present (Mullen et al., 2013). With the increase in availability of markers, there has been a dramatic increase in genome-wide association studies. These studies seek to identify common DNA variants that are associated with variation in the trait being studied (Yang et al., 2010). Such studies are a standard approach used for identifying genomic regions associated with economically important production traits in agricultural species, variation among wild populations, and candidate regions associated with complex genetic diseases (Gondro et al., 2013) with relatively low cost and in short time. This technological progress is allowing understanding genetic basis of phenotypic variance (height, growth intensity, milk production etc.).

According to Manolio (2013), in last 10 years was published more than 1600 scientific papers, identifying in total more than 200 important associations to complex of more than 300 diseases and traits. Today SNP markers are used in whole genome studies (Bolormaa et al., 2011; Schopen et al., 2011); in genetic prediction of breeding values (Meuwissen and Goddard, 2010) as well as for estimation of parameters of genetic diversity and population-genetic parameters (Engelsma et al., 2012). Regarding farm animals this technology is most developed in cattle. Factors like evolutionary history, genetic structure, economy etc, are making cattle the most suitable for genomically assisted selection (Nicolazzi et al., 2014).

The long-term challenge for animal breeders is to improve the productivity of major livestock species to meet the growing demands for livestock products and minimize their impact on the environment and global natural resources (Hume et al., 2015). Genotyping of cattle using SNP arrays has become common practice in dairy cattle breeding programs applying genomic selection (Mulder et al., 2012).

The aim of this study was to detect the position of genomic regions associated with milk production and milk components in Pinzgau cattle.

Material and methods

Phenotype data

The dataset consisted of test day records of in total 7729 cows, daughters of 35 sires representing active breeding bulls (19 animals) and AI doses stored in gene bank (16 animals) of Pinzgau cattle in Slovakia. The phenotypic data were provided by the Breeding Services of SR s.e. The phenotype of the sires was expressed as average daughter yield available for milk production traits including milk yield (kg), fat yield (kg), protein yield (kg), fat (%), and protein (%). In total, 130087 records for milk, protein and fat yield were used.

Genotyping and data quality control

Genomic data from 35 bulls were obtained by using the Illumina BovineSNP50 BeadChip V2. Markers assigned to unmapped regions or with unknown chromosomal position according to the latest bovine genome assembly (Btau 4.0) and SNPs positioned to sex chromosomes were removed. Filtering of data was performed as described in Purcell et al. (2007) to exclude any autosomal loci with call rate lower than 90 %, minor allele frequency lower than 0.05 and HWE limit of 0.00001. The final dataset was composed of 41,487 autosomal loci covering overall length 2,500,315 kb of the genome.

Statistical analysis

Identification of genomic regions associated with milk production and composition were performed using GEMMA software (Zhou and Stephens, 2012) that uses a linear mixed model approach based on genetic-relationship matrix estimated from SNP genotypes to model correlation between the phenotypes. Testing was done using log likelihood ratio statistic to test the alternative hypothesis H_1: $\beta \neq 0$ against the null hypothesis H_0: $\beta = 0$ for each SNP. An association test using the univariate linear mixed model was performed for each trait. The model in the following form was used:

$$y = W\alpha + x\beta + u + \varepsilon$$

where y is the vector of quantitative traits for individuals, W is a matrix of covariates (fixed effects), α is a vector of the corresponding coefficients including the intercept, x is a vector of marker genotypes, β is the effect size of the markers, u is a vector of random effects, and ε is a vector of errors (Zhou et al., 2012).

To show a significant deviation from the null hypothesis the quantile–quantile (Q–Q) plot was constructed using R package qqman (Turner, 2014). The genome-wide significance thresholds for SNP p-values were determined by Bonferroni correction (Holm, 1979; Yang et al., 2005). A SNP was considered to have genome-wide significance at $P < 0.05/N$, where N is the total number of tested SNPs. A SNP with appropriate level of genome-wide significance have been assigned to the genomic QTL location according to the Bovine Genome Database (http://bovinegenome.org).

Results and discussion

The phenotype of the sires expressed as average daughter yield available for milk production traits including milk yield (kg), fat yield (kg), protein yield (kg), fat (%), and protein (%) is in Table 1. Due to variation between sires according their year of birth (1983 – 2005) higher variance of average values was observed.

Table 1. Phenotypes in the sires expressed as daughter average yield

	Phenotypic trait	No. of records	Mean	SD
Sires	Milk yield (kg)	130087	10.566	6.344
	Fat yield (kg)	130087	0.425	0.265
	Protein yield (kg)	130087	0.358	0.214
	Fat %	130087	4.065	0.797
	Protein %	130087	3.407	0.358

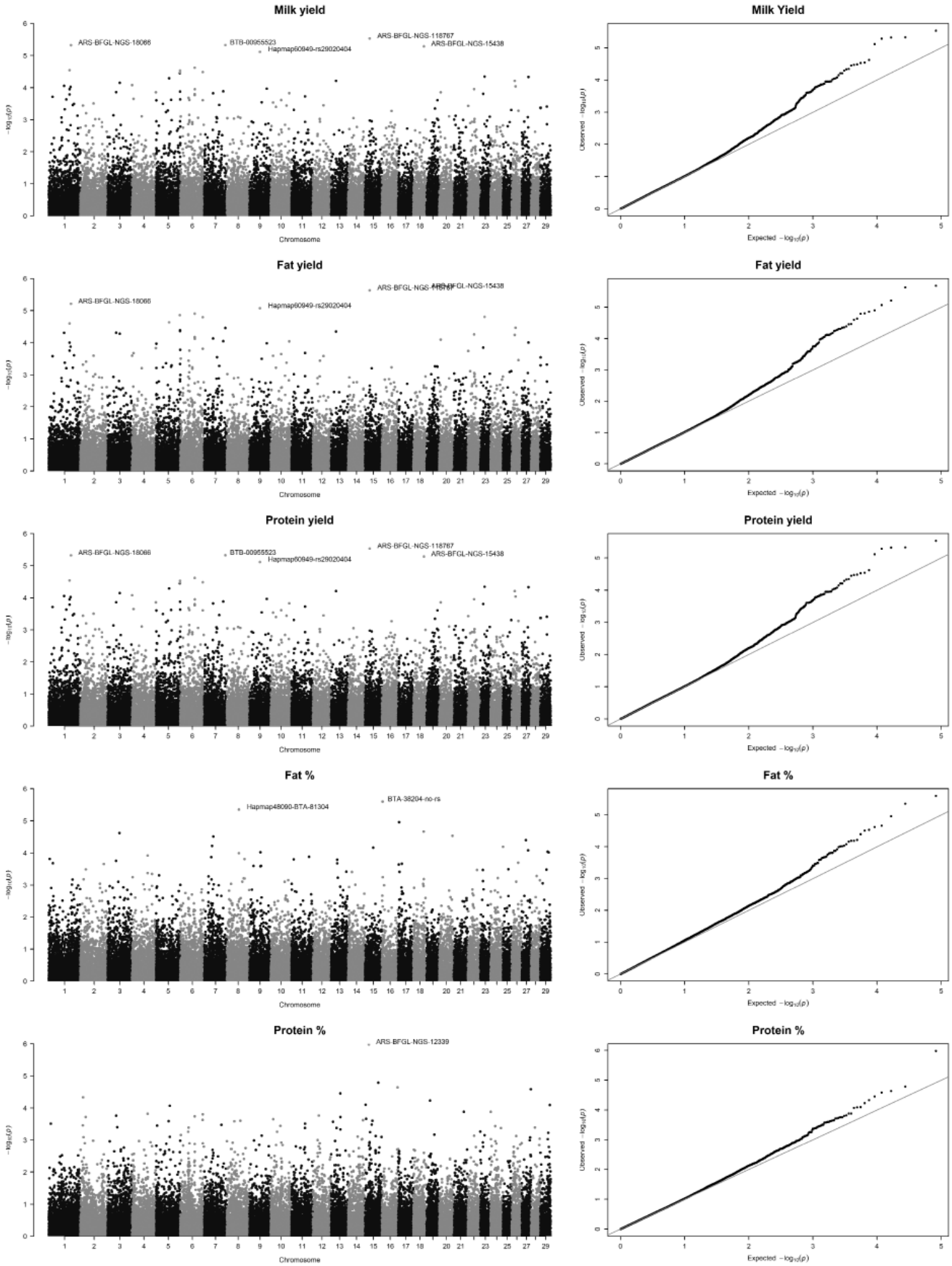

Figure 1. Manhattan and Quantile–quantile (Q–Q) plots for univariate genome-wide association studies

Table 2. Pearson correlations between phenotypes for the sire

	Milk yield	Fat yield	Protein yield	Fat %	Protein %
Milk yield	1				
Fat yield	0.9295***	1			
Protein yield	0.98225***	0.92812***	1		
Fat %	-0.09334***	0.22218***	-0.04995***	1	
Protein %	-0.10109***	-0.01905***	0.05663***	0.30527***	1

Conclusion

Genome wide association tests are part of the breeding management of livestock populations. Results of such studies could directly influence the estimation of breeding values and provide additional value to reliability of estimation in case of small population like Slovak Pinzgau cattle. Pinzgau cattle in Slovakia are under active conservation and farmers expect positive genetic progress. In present nucleus, population is spread in two directions with dairy production or suckling cow system. Our results are confirming dual purpose character of selection as established in breeding goal. The identified genomic regions could be used in future to support the selection of Slovak Pinzgau in both directions. Further analysis with more fixed effects as well as multitrait mixed model could benefit in higher number of genomic signals.

Table 3. SNPs significantly associated with milk production

CHR	Illumina ID	Position	P value	QTL traits
1	ARS-BFGL-NGS-18066	111357945	4.77e-06	Milk Yield, Dressing percentage
7	BTB-00955523	105621232	4.74e-06	Milk Yield, Protein Yield, SCS score
8	Hapmap48090-BTA-81304	60269047	4.47e-06	
9	Hapmap60949-rs29020404	52283151	7.68e-06	Marbling Score, Milk, Protein and Fat yield
15	ARS-BFGL-NGS-12339	20018872	1.05e-06	
	ARS-BFGL-NGS-118767	24021537	2.93e-06	
16	BTA-38204-no-rs	3075859	2.52e-06	
18	ARS-BFGL-NGS-15438	53224638	5.18e-06	

In Table 2 correlations between phenotypes are presented. High Pearson correlations were observed between milk, fat and protein yield mainly due to fact that within each test day, milk yield is observed directly together with fat and protein percentage. Fat and protein yield are just mathematical expression of fat resp. protein content.

Results of the association test are presented in Table 3. After Bonferroni correction of p-values statistically significant signals were observed on 7 SNPs. Some of the SNPs are located in the interesting genomic QTL regions responsible for milk yield and composition, marbling score and fat yield. As visible from the results in Table 3, significant signals confirm existing fact that Pinzgau cattle are dual – purpose (milk-meat) which is steadily selected in both directions.

Variation in milk production traits and functional traits in dairy cattle have a major genetic component. Out of 36,693 quantitative trait loci (QTL) for 492 traits archived in the cattle QTL database (Cattle QTLdb database), about 5,815 are QTLs for milk fat composition, 3,157 for milk protein composition, 1,324 for milk yield, 550 for fatty acid content and 1,246 for mastitis. These QTLs are spread on most bovine chromosomes but only a few of the causative genes have been identified (Ibeagha-Awemu et al., 2016). GWAS studies have been successful for identifying genomic regions which associate with these traits but few have led to identification of the underlying mutation (Raven et al., 2014). Three significant SNPs on BTA1, BTA7, and BTA9 detected in this study were found within a region that contain possible candidate genes associated with milk production and composition. A connection between BTA1 and BTA7 and milk yield was reported in many studies (Meredith et al. 2012; Buitenhuis et al., 2014; Olsen et al., 2017; El-Halawany et al., 2017).

References

Bolormaa S., Hayes B. J., Savin K., Hawken R., Barendse W., Arthur P.F., Herd R.M., Goddard M.E. (2011). Genome-wide association studies for feedlot and growth traits in cattle. J Anim Sci 89: 1684-1697

Buitenhuis B., Janss L.L.G., Poulsen N.A., Larsen L.B., Larsen M.K., Sørensen P. (2014). Genome-wide association and biological pathweys analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15: 1112

El-Halawany N., Abdel-Shafy H., Shawky A.A., Abdel-Latif A., Al-Tohamy A.F.M., El-Moneim O.M.A. (2017). Genome-wide association study for milk production in Egyptin buffalo. Livest sci 198: 10-16

Engelsma K.A., Veerkamp R.F., Calus M.P.L., Bijma P., Windig J.J. (2012). Pedigree and marker-based methods in the estimation of genetic diversity in small groups of Holstein cattle. J Anim Breed Genet 129: 195-205

Gondro C., Lee S.H., Lee H.K., Porto-Neto L.R. (2013). Quality control for genome-wide association studies. Methods Mol Biol 1019: 129-147

Holm S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 6: 65-70.

Hume D.A., Whitelaw C.B.A., Archibald A.L. (2011). The future of animal production: improving productivity and sustainability. J Agr Sci 149: 9-16

Ibeagha-Awemu E.M., Peters, S.O., Akwanji K.A., Imumorin I.G., Zhao X. (2016). High density genome wide genotyping-by-sequencing and association identified common and low frequency SNPs, and novel candidate genes influencing cow milk traits. Scientific Reports 6: 31109

Manolio T.A., Chisholm R.L., Ozenberger B., Roden D.M., Williams M.S., Wilson R., Bick D., Bottinger E.P., Brilliant M.H., Eng C., Frazer K.A., Korf B., Ledbetter D.H., Lupski J.R., Marsh C., Mrazek D., Murray M.F., O'Donnell P.H., Rader D.J., Relling M.V., Shuldiner A.R., Valle D., Weinshilboum R., Green E.D., Ginsburg G.S. (2013). Implementing genomic medicine in the clinic: the future is here. Genet Med 15(4): 258-267

Meredith B., Kearney F., Finlay E., Bradley D., Fahey A., Berry D., Lyn D.J. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet 13:21

Meuwissen T., Goddard M. (2010). Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing. Genetics 185: 623-631

Mulder H.A., Calus M.P., Druet T., Schrooten C. (2012). Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. J Dairy Sci 95: 876-889

Mullen M.P., Hanrahan J.P., Howard D.J., Powell R. (2013). Investigation of Prolific Sheep from UK and Ireland for Evidence on Origin of the Mutations in BMP15 (FecXG, FecXB) and GDF9 (FecGH) in Belclare and Cambridge Sheep. PLoS ONE 8(1): e53172

Nicolazzi E.L., Picciolini M., Strozzi F., Schnabel R.D., Lawley C., Pirani A., Brew F, Stella A. (2014). SNPchiMp: a database to disentangle the SNPchip jungle in bovine livestock. BMC Genomics 15: 123

Olsen H.G., Knutsen T.M., Kohler A., Svendsen M., Gidskehaug L., Grove H., Nome T., Sodeland M., Sundsaasen K.K., Kent M.P., Martens H., Lien S. (2017). Genome-wide association mapping for milk fat composition and fine mapping of QTL for de novo synthesis of milk fatty acids on bovine chromosome 13. Gen Sel Evol 49:20

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 81: 559-575

Raven L.A., Cock B.G., Hayes B.J. (2014). Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15: 62

Schopen G.C., Visker M.H., Koks P.D., Mullaart E., van Arendonk J.A., Bovenhuis H. (2011). Whole-genome association study for milk protein composition in dairy cattle. J Dairy Sci 94: 3148-3158

Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. biorXiv

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E., Visscher P.M. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42(7): 565-569

Yang Q., Cui J., Chazaro I., Cupples L.A., Demissie, S. (2005). Power and type I error rate of false discovery rate approaches in genome-wide association studies. BMC Genet 6(Suppl 1): S134

Zhou X., Stephens M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44: 821-824

acs82_53