Proceedings of the World Congress on Genetics Applied to Livestock Production, 11.(289)

Combining SNP-chip and whole genome sequence data towards the identification of causal mutations underlying feet and leg disorders in cattle

T. Suchocki[1,2], Ch. Egger-Danner[3], H. Schwarzenbacher[3], M. Mielczarek[1,2] & J. Szyda[1,2]

[1] Wroclaw University of Environmental and Life Sciences, Department of Genetics, Biostatistics group, Kozuchowska 7, 51-631 Wroclaw, Poland
tomasz.suchocki@upwr.edu.pl (Corresponding Author)
[2] National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland
[3] ZuchtData EDV-Dienstleistungen GmbH, Dresdner Straße 89/19, 1200 Vienna, Austria

Summary

Feet and leg disorders are traits of increasing importance in dairy cattle because of their impact on welfare of animals. The main aim of this study is identification of particular genes responsible for the disorder of feet and leg health. We used data set consisted of 2,223 cows from Fleckviev and Braunvieh breeds genotyped using GeneSeek®Genomic ProfilerTM HD panel. After editing, 74,762 SNPs were used for statistical modelling of the total number of hoof diseases scored for a cow till day in milk 100 with the average MAF of 0.31 and the average call rate equal to 99.48%. Additionally 78 whole genome sequences were used to identify the polymorphisms in coding sequences (i.e. exons of genes) located in the proximity of SNPs from the panel. We discovered two significant SNPs at the level 5% located on chromosome 7 and 14. Marker rs109798552 located on chromosome 7 had p-value after FDR correction equal to 0.034, while rs110532594 located on BTA14 had corrected p-value equal to 0.022.Additionally we found the “aggregate genotypes” built based on sequence and panel data which significantly influence the trait.

Keywords: cattle, feet and leg disorders, genome-wide association study, SNP microarray, whole genome sequence

Introduction

Feet and leg disorders are traits of increasing importance in dairy cattle because of their impact on welfare of animals. On one hand, genetic or genomic selection can be an important tool for selecting against susceptibility of such traits, but on the other hand, often positive genetic correlations with production traits may impair the selection response. Another way towards the genetic improvement of feet and leg health would be the identification of particular genes responsible for the disorder – an approach that was attempted in the current study.

Material and methods

Dataset

The data set originated from the Austrian Braunvieh and Fleckvieh cow population and comprised records from several parities (1-13). Records of the total number of hoof diseases

scored for a cow till day in milk (DIM) 100 for 1,469 Fleckvieh and 754 Braunvieh cows with deep pedigree (41,431 individuals from 17 generations) were used in the analysis (0-5). The number of records per cow varying between 1 and 3. Each cow was genotyped using the GeneSeek®Genomic ProfilerTM HD panel, which consists of 76,934 SNPs. We applied SNP selection criteria comprised a minor allele frequency (MAF) of at least 0.01 and the technical quality of a SNP quantified by a minimal call rate of 99%. After editing, 74,762 SNPs were used for statistical modelling with the average MAF of 0.31 and the average call rate equal to 99.48%.

DNA sequences of 30 Fleckvieh and 48 Braunvieh bulls were obtained within the framework of Gene2Farm project by Illumina HiSeq 2000 Next Generation Sequencing platform. Filtered reads were aligned to the UMD3.1 reference genome using BWA. SNPs were detected using several software, including GATK.

Estimation of variance components

Variance components were estimated using the following linear mixed model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_a a} + \boldsymbol{Z_d d} + \boldsymbol{Z_p p} + \boldsymbol{e}, \quad (1)$$

where $\boldsymbol{y}$ was a vector of the number of hoof diseases per cow and parity; $\boldsymbol{\beta}$ was a vector of fixed effects comprising: a general mean, breed, parity (from 1 to 4 and ≥ 5), calving year-season (years between 2012 and 2015, season 1 – between October and March, season 2 – between April and September), percent of non Holstein-Friesian genes, type of disease evaluation (four levels); $\boldsymbol{a}$ was a random additive polygenic effects of a cow; $\boldsymbol{p}$ was a random permanent environmental effect; $\boldsymbol{d}$ was a random smith effect and $\boldsymbol{e}$ was a vector of residuals. $\boldsymbol{X}$, $\boldsymbol{Z_a}$, $\boldsymbol{Z_d}$ and $\boldsymbol{Z_p}$ were design matrices respectively for β, $\boldsymbol{a}$, $\boldsymbol{d}$ and $\boldsymbol{p}$. The assumed covariance structure was defined as $\boldsymbol{a} \sim N(0, \boldsymbol{A}\sigma_a^2)$, $\boldsymbol{d} \sim N(0, \boldsymbol{I}\sigma_d^2)$, $\boldsymbol{p} \sim N(0, \boldsymbol{I}\sigma_p^2)$ and $\boldsymbol{e} \sim N(0, \boldsymbol{I}\sigma_e^2)$, with $\boldsymbol{A}$ being an additive relationships matrix among cows. The ASReml software (Gilmour *et al.*, 2009) was used for the estimation of model parameters and variance components.

Estimation of SNP effects

In order to estimate the additive effects of SNPs a single SNP effect was added to the right-hand side of equation (1) with a design matrix parameterized as -1, 0, or 1 for a homozygous, heterozygous and an alternative homozygous genotypes, respectively. The procedure was repeated for each SNP separately. For testing the significance of SNP effects the Wald test was used. Under H_0, this test statistic asymptotically follows the standard normal distribution. The multiple testing problem was addressed by the false discovery rate (FDR) correction (Benjamini and Hohberg,1995) of each nominal P value.

Updating SNP chip information by SNPs from whole genome sequence

In order to identify polymorphisms in coding sequences (i.e. exons of genes) located in the proximity of SNPs with significant effects on phenotype, the SNPs were annotated to the whole genome sequence data of the Gene2Farm project. Separately for each individual with whole genome sequence, "aggregated genotypes" were constructed as combined genotype constellations of the significant SNP from model (1), polymorphism located in coding sequence

and flanking SNPs occurring both in sequence and SNP panel. From each "aggregated genotypes" unique values with their frequency were used in the further analysis.

Estimation of aggregate genotype effects

The effects of "aggregated genotypes" were estimated using model (1) with additional effect of each possible "aggregate genotype". For each individual with phenotype we calculated probability of carrying each possible "aggregate genotype". Then this probability matrix was used as a design matrix for "aggregate genotype" effects.

Results

Heritability

The heritability of the total number of hoof diseases scored for a cow till DIM 100 was 0.28. All estimated variance parameters are presented in Table (1).

Table 1. Variance components estimated using ASReml software for model (1)

Additive polygenic	Permanent environmental	Smith effect	Residual
0.243642	0.0000001	0.168146	0.460228

Effects of SNPs and aggregated genotypes

Effects of each SNP effect estimated using Model (1) were presented on Figure (1). Out of them two SNPs located on chromosome 7 and 14 were significant. Marker rs109798552 located on chromosome 7 had p-value after FDR correction equal to 0.034, while rs110532594 located on BTA14 had corrected p-value equal to 0.022. Around those two SNPs "aggregate genotypes" were built using SNPs from chip panel and WGS data. The "aggregates genotypes" are presented on Table (2).

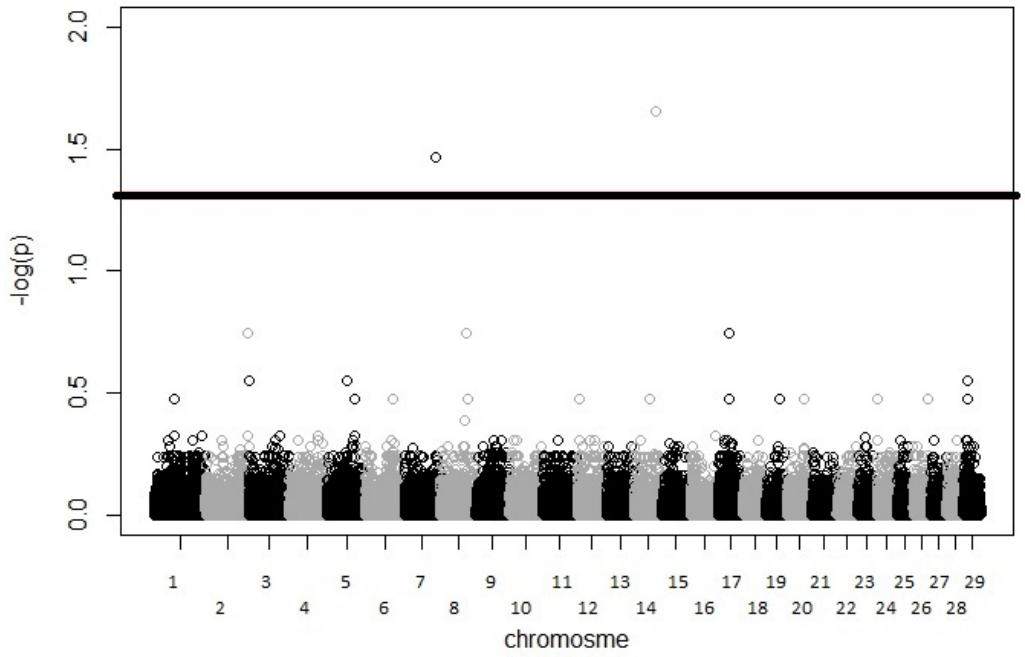

Figure 1. Transformed p-values after FDR correction occurred using model (1).

Table 2. "Aggregate genotypes" created on chromosome 7 and 14 by aggregating SNP located on chip panel and whole genome sequence (WGS).

SNP	Position	Function	SNP panel or WGS
rs109798552	100,343,329	most significant SNP on BTA7	SNP panel
SNP1	100,446,177	SNP located in exon of RGMB gene	WGS
SNP2	100,446,197	SNP located in exon of RGMB gene	WGS
SNP3	100,446,401	SNP located in exon of RGMB gene	WGS
rs109507183	100,465,392	Flanking SNP located on BTA7	SNP panel
rs136813945	67,768,919	Flanking SNP located on BTA14	SNP panel
SNP4	67,781,949	SNP located in exon of STK3 gene	WGS
SNP5	67,901,541	SNP located in exon of STK3 gene	WGS
SNP6	67,901,601	SNP located in exon of STK3 gene	WGS
SNP7	67,901,616	SNP located in exon of STK3 gene	WGS
rs110532594	67,911,958	most significant SNP on BTA14	SNP panel
SNP8	67,986,610	SNP located in exon of STK3 gene	WGS
SNP9	67,986,929	SNP located in exon of STK3 gene	WGS
SNP10	67,987,260	SNP located in exon of STK3 gene	WGS
rs136884351	67,997,855	Flanking SNP located on BTA14	SNP panel

31 possible "aggregate genotype" were located on BTA7 and 44 on BTA14 based on WGS data. For each animal with phenotype we calculated probability of being either each of 31 possible "aggregate genotype" on BTA7 or 44 on BTA14. Then we calculated effect of each possible "aggregate genotype". TTCCAATTTT and TTCCAGTTTT were significant on BTA7 and AATTGGAATTTTGGTTGG, AATTAAAATTGGAATTGG, AATTAGAATTGTGGTGGG, AATTAGAATTGTGGTGAG, AATTAGAATTTTGGTTAG and AGTTAGAATTGTAGTTAG on BTA14. Additionally we found some positive influence

on SNP effect when marker rs109798552 has genotype TT and negative influence on SNP effect otherwise. For BTA14 we found possible epistasis between markers rs136813945 and rs110532594.

List of References

Benjamini, Y., & Y. Hohberg, 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. of the Royal Statistical Society. Series B. 57: 289-300.

Gilmour, A. R., R. Thompson, & B. R. Cullis, 1995. Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. Biometrics 51:1440-1450.