Improving the Accuracy of Genomic Prediction of Milk Fat

A.S. Leaflet R2870

Melanie Hayr, Master's Student; Mahdi Saatchi, Postdoctoral Fellow; Dave Johnson, Senior Scientist, Livestock Improvement Corporation, NZ; Dorian Garrick, Professor, Department of Animal Science

Summary and Implications

Four statistical models were considered to quantify any advantage of including the genotype of known causative mutations when calculating direct genomic values. Data included 50k genotypes from 5,661 Holstein Friesian cows and 2,287 bulls. This study showed that when a known QTL for milk traits, DGAT1, was fit as a fixed class or fixed covariate in genomic prediction, an increase in accuracy was seen compared to fitting it as either a random covariate or relying on linked 50k markers fit as random covariates. The regression coefficients of genomic prediction on phenotype were near one for all estimates, indicating no major bias was in the estimates. These results suggest it is beneficial to the accuracy of prediction to include information from known major QTL in genomic analyses.

Introduction

It is necessary for Direct Genomic Values (DGV) to be as accurate as possible to increase genetic gain. Major genes or Quantitative Trait Loci (QTL) have been identified for many traits and have the potential for aiding prediction compared to using anonymous markers. This study determined whether including QTL genotype when calculating DGV increases their accuracy.

A mutation in Diacylglycerol Acyltransferase 1 (DGAT1) on chromosome BTA14 has been shown to have a large effect on milk, fat and protein yields in *Bos taurus* and *indicus* cattle where the $DGAT1^{K}$ allele causes an increase in fat yield and decrease in protein yield and milk yield compared to $DGAT1^{A}$.

Materials and Methods

The data set consisted of 5,661 Holstein Friesian cows and 2,287 bulls with Illumina BovineSNP50 (50k) genotypes and deregressed estimated breeding values (DEBV) for fat yield. Actual DGAT1 genotypes were available for 1,133 cows and 655 bulls and were imputed for the remaining animals using BEAGLE software.

Four models were run using five-fold cross-validation in GenSel using Bayes-B with 2.5% of SNPs assumed to have an effect on the trait: 1) a control model relying on 50k markers to pick up the effect of DGAT1; 2) a model that included DGAT1 dosage fit as a random covariate; 3) a model that included the three DGAT1 genotypes fit as a fixed class; and 4) a model that included DGAT1 dosage fit as a fixed covariate. These four models were separately fitted for males and females and for individuals directly genotyped for DGAT1 and for all individuals.

Accuracy was defined as the correlation between DEBV and DGV while bias was represented by regression coefficient of DEBV on DGV deviating from unity.

Results and Discussions

Similar results were obtained using models 1 or 2. Models 3 or 4 were similar but consistently better than models 1 or 2.

Accuracy was higher in males when all animals were included while in females accuracy was higher when only animals with true DGAT1 genotype were included. Bias was lowest when true and imputed DGAT1 genotypes were included in the model.

These results suggest that including DGAT1 genotype as a fixed class or a fixed covariate when calculating DGVs both increases accuracy and reduces bias.

Acknowledgements

We would like to thank Livestock Improvement Corporation (LIC) in New Zealand for the use of their data for this research.

Table 1. Regression and Accuracy of DEBV for Prediction of Milk Fat Yield in Male and Female Holstein Friesians.

Sex	Model	Direct DGAT1		Direct and Imputed DGAT1	
		b	r	b	r
Males	50k	1.246	0.553	1.012	0.697
	50k+DGAT1(Random Covariate)	1.246	0.552	1.010	0.696
	50k+DGAT1(Fixed Class)	1.191	0.660	1.007	0.737
	50k+DGAT1(Fixed Covariate)	1.055	0.536	1.008	0.737
Females	50k	1.141	0.399	1.048	0.385
	50k+DGAT1(Random Covariate)	1.140	0.400	1.045	0.384
	50k+DGAT1(Fixed Class)	1.083	0.503	1.039	0.453
	50k+DGAT1(Fixed Covariate)	1.038	0.463	1.040	0.455