Improving Accuracy of Genomic Prediction in Holstein Friesians

A.S. Leaflet R2784

Melanie Hayr, Master's Student, Animal Science;
Mahdi Saatchi, Post Doc, Animal Science;
Dave Johnson, Senior Scientist, LIC;
Dorian Garrick, Professor, Animal Science

Summary and Implications

Three statistical models were considered to assess the advantage of including information of known causative mutations when estimating genomic breeding values. Data included phenotypic records and 50k genotypes from 5,661 Holstein Friesian cows. This study showed that when a known causative mutation for milk traits, DGAT1, was fit as a fixed effect in genomic prediction, an increase in accuracy was seen compared to fitting it as either a random effect or not explicitly fitting it and relying on linked markers fitted as random effects. The regression coefficients of genomic prediction on phenotype were near one for all estimates, indicating that no major bias was present in the estimates. These results suggest that, when calculating genomic predictions, it is beneficial to include information from known major genes in the analysis to increase the accuracy of prediction.

Introduction

It is necessary for Genomic Estimated Breeding Values (GEBV) to be as accurate as possible to reduce the accumulation of inaccuracies in pedigrees resulting from selection of unproven parents. Major genes or Quantitative Trait Loci (QTL) have been identified for many traits and have the potential for aiding selection decisions compared to using anonymous markers. This study determined whether including causal genotypes when calculating GEBVs increases their accuracy.

A mutation in Diacylglycerol Acyltransferase 1 (DGAT1) on chromosome BTA14 has been shown to have a large effect on milk traits such as milk, fat, and protein yields in both *Bos taurus* and *Bos indicus*. There are two alleles: $DGAT1^{K}$ causes an increase in fat yield and a decrease in protein yield and milk yield compared to $DGAT1^{A}$.

Materials and Methods

The data set used in this study consisted of 5,661 New Zealand Holstein Friesian cows with Illumina BovineSNP50 (50k) genotypes and deregressed estimated breeding values (DEBV) for fat yield. DGAT1 genotypes were available for 1,133 of these cows and were imputed for the remaining 4,528 cows using BEAGLE.

Three models were run in GenSel using Bayes B with 2.5% of SNPs assumed to have an effect on the trait. Five-fold cross-validation was used to test the following three models: 1) a model fitting only 50k genotypes as random effects; 2) a model fitting 50k genotypes and DGAT1 as random effects; and 3) a model fitting 50k genotypes as random effects and DGAT1 genotypes as a fixed effect. These three models were separately fitted to individuals directly genotyped for DGAT1 and for all individuals, both directly genotyped and imputed for DGAT1.

Results and Discussions

The accuracy of GEBV was assessed by regressing DEBV on GEBV and inspecting the regression and correlation coefficients. The regression coefficients were approximately one in all cases, suggesting little to no bias in GEBV. Furthermore, the regression coefficients also got closer to one when DGAT1 was included in the model compared to when it was not, and was closest to one when DGAT1 was fit as a fixed effect, suggesting that fitting DGAT1 as a fixed effect results in less bias.

There was a consistent increase in accuracy of GEBV when DGAT1 was included in the model compared to when it was not included, and fitting DGAT1 as a fixed effect resulted in the most accurate estimate of the DEBV. The increase in accuracy of GEBV between fitting DGAT1 as a fixed effect rather than a random effect is greater for the data that had DGAT1 directly genotyped than when DGAT1 was imputed, perhaps due to inaccuracies in imputation.

These results indicate that fitting a causative mutation as a fixed effect in the model when calculating GEBVs increases the accuracy of prediction.

Acknowledgements

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

Table 1. Accuracy of genomic prediction of phenotype.

Model	Direct Genotypes		Direct and Imputed Genotypes	
	b	cor	b	cor
50k	1.104	0.402	0.908	0.377
50k + DGAT1 as a Random Effect	1.102	0.406	0.911	0.381
50k + DGAT1 as a Fixed Effect	1.014	0.425	0.917	0.389