Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Evaluating the Effects of QTN for Milk Fat Yield and their Impact on Accuracy and Bias of Genomic Prediction in New Zealand Holstein-Friesian Cows

***M. Hayr*[1], M. Saatchi[1], R. Sherlock[2], D. Johnson[2] and D. Garrick[1]**
[1]Iowa State University, [2]LIC, Hamilton, New Zealand.

ABSTRACT: Five SNPs were analyzed across 4,801 Holstein-Friesian cows, including three QTN for milk fat yield: *DGAT1*, *GHR*, and *AGPAT6*; a QTN for stature: *PLAG1*; and a control SNP with no effect on milk fat yield. Dominance was observed for *DGAT1*, *AGPAT6* and *PLAG1*. A base model of 35,000 SNPs was run in GenSel using BayesB. In addition to the base model 1) SNP dosage was fit as a random covariate, or 2) SNP genotype was fit as a fixed covariate, or 3) SNP dosage was fit as a fixed covariate. Including these QTN as random covariates increased accuracy of direct genomic value prediction. Including QTN as fixed covariates slightly decreased accuracy and increased bias. Including *DGAT1* as a fixed covariate decreased bias. These results suggest inclusion of QTN genotypes can potentially increase accuracy and decrease bias of DGV, although only slightly.
Keywords: Milk fat; QTNl Prediction Accuracy

Introduction

The availability of low cost SNP chips has facilitated use of genomics for estimation of breeding values. One approach generates direct genomic values (DGV) from genotypes at SNPs in concert with their estimated effects. Accurate DGV are essential to prevent buildup of inaccuracies in pedigrees comprising multiple generations of young and unproven parents. Causal mutations have been identified in many species across several traits (Grisart et al. (2002), Thomsen et al. (2004), Johnson et al. (2009)), each of which could be used to increase accuracies of genomic prediction. These causal mutations, which influence quantitative traits, are termed quantitative trait nucleotides (QTN). The integration of data from SNP chips and genotyped QTN has potential to increase accuracy of DGV and improve genetic gain. Milk fat yield is an economically important trait in dairy cattle, because milk fat is in many consumer products. Given its importance, many farmers are paid directly for this trait. The objective of this study was to estimate the effect of five QTN on milk fat yield in the New Zealand Holstein-Friesian population and evaluate the effects on accuracies and bias by including them as a random covariate, fixed covariate or a fixed class effect in addition to 50k SNP routinely used in genomic prediction models.

Materials and Methods

Data. Deregressed Estimated Breeding Values (DEBV) for milk fat yield were collected on 4,801 Holstein-Friesian females. These females had been genotyped at five QTN: three causal mutations for milk fat yield: *DGAT1* (Grisart et al. (2002)), *GHR* (Blott et al (2003)), and *AGPAT6* (Littlejohn et al. (2014)); one causal mutation for stature, a trait that has been shown to be correlated with milk fat yield (Brotherstone (1994)): *PLAG1* (Karim et al. (2011)); and a randomly selected SNP on chromosome 8 that has no known association with any selected trait: *CHR8*. In addition, these cows were genotyped on a parentage panel including 20,000 SNPs that were imputed to 35,000 SNPs (35k) comprising high quality SNPs present on the Illumina BovineSNP50 and BovineHD panels.

Model. The BayesB method was run in GenSel (Fernando and Garrick (2009)) with five-fold cross validation and 2.5% of SNPs assumed to have an effect on the trait, for all the following models. The base model, relied on SNPs in the 35k set that are in LD with the five QTN to pick up their effect. Three models were run in GenSel for each of the QTN independently and for all QTN together: 1) Random Covariate: 35k set plus allele dosage fit as a random covariate; 2) Fixed Covariate: 35k set plus allele dosage fit as a fixed covariate; and 3) Fixed Class: 35k set plus SNP genotype fit as a fixed class.

Estimation of QTN Parameters. The effect of each of the five QTN was estimated based on genetic variance explained by the window when each of these SNPs were added to the base model as random covariates. The allele substitution effect, alpha, was calculated for each QTN based on the regression coefficient for SNP dosage in the Random Covariate model. Classical additive (a) and dominance (d) effects were calculated based on the genotype effect estimates obtained by fitting the Fixed Class model.

Estimation of Prediction Accuracy. Accuracy was defined as the correlation between DGV and DEBV. Bias was represented by the regression coefficient of DEBV on DGV. A paired t-test was performed comparing each correlation and regression coefficient to those for the base model. Each cross-validation set was paired because they comprise the same animals.

Results and Discussion

Estimation of QTN Parameters. Minor allele frequency was > 0.05 for all QTN (Table 1) indicating that QTN effects should be able to be reasonably accurately estimated. *DGAT1*, *GHR*, *AGPAT6* and *PLAG1* all deviate from Hardy Weinberg Equilibrium. This is not surprising because they are QTN for a selected trait. It is common to remove SNPs that fail the test for Hardy Weinberg Equilibrium, however these should remain in a genomic prediction analysis due to having a proven effect on milk fat yield. The SNP on chromosome 8 was a reference SNP, and passes the test for Hardy Weinberg Equilibrium which means that genomic region is likely not under strong selection.

Table 1: Summary statistics for the five quantitative trait nucleotides (QTN)

QTN	Minor Allele Frequency	HWE P-Value
DGAT1	0.49	< 0.001
GHR	0.09	0.042
AGPAT6	0.34	< 0.001
PLAG1	0.26	< 0.001
CHR8	0.22	0.557

Table 2: SNP effects for the five quantitative trait nucleotides (QTN)

QTN	[1]Mean Variance	Window Fit[2]	Alpha[3]	a[3]	d[3]
DGAT1	22.98*	100	9.54*	9.48*	-1.93*
GHR	0.32*	55	4.32*	4.85*	0.60
AGPAT6	0.17*	57	2.16*	1.48*	3.32*
PLAG1	0.04*	38	1.45	2.24*	2.96*
CHR8	0.02*	28	0.20	1.64	1.24

* Value is statistically different from zero with $\alpha = 0.05$

[1] Posterior probability of the mean variance explained by the window containing that QTN in the random covariate model as a percent of genetic variance.

[2] Posterior probability of association for the window containing that QTN in the random covariate model as a percent.

[3] Units are kg/lactation.

The 1Mb window that contains *DGAT1* explains the most genetic variance, followed by windows containing *GHR* and *AGPAT6*, then *PLAG1* with *CHR8* having smallest effect (Table 2). The posterior probability of association followed almost the same order, with *GHR* and *AGPAT6* interchanged. As expected, the three major QTN explain the most variance, the QTN for the correlated trait explains a little genetic variance and the SNP with no known effect explains almost no genetic variance. All values were significantly different from zero.

The allele substitution effect, alpha, is significantly different from zero for all three QTN but not significantly different from zero for *PLAG1* or *CHR8* (Table 2). Alpha is largest for *DGAT1* at 9.54 kg/lactation. The next largest allele substitution effects are *GHR* and *AGPAT6*. The allele substitution effect for *CHR8* is only 0.2 kg/lactation, not significantly different from zero.

The additive effect was significant for all QTN except *CHR8* (Table 2). The additive effect was largest for *DGAT1* but smaller than its allele substitution effect, due to dominance for this QTN. The next largest additive effect was for *GHR*, followed by *PLAG1* then *AGPAT6*. *CHR8* had the smallest additive effect, which was not significant.

DGAT1, *AGPAT6* and *PLAG1* all showed significant dominance (Table 2). While Grisart et al. (2002) did not find significant dominance in *DGAT1* in this population, these results are consistent with dominance observed by Kuehn et al. (2007) in German Holstein cows, showing that heterozygotes have lower milk fat yield than the midpoint of the two homozygotes. There is evidence that the *DGAT1* heterozygote has a different mean milk fat yield than the qq homozygote ($p << 0.001$) so *DGAT1* shows partial dominance. Due to *AGPAT6* only recently being identified as a causal mutation this is the first instance of its dominance being reported. There is evidence *AGPAT6* heterozygotes have different mean milk fat yield than QQ homozygotes ($p = 0.03$), suggesting over-dominance. The results for *PLAG1* are consistent with Littlejohn et al. (2012) who showed there may be dominance in *PLAG1* in this population in respect to stature. There is no evidence the *PLAG1* heterozygote has different milk fat yield than the QQ homozygote ($p = 0.51$), indicating complete dominance. The presence of dominance in *GHR* for milk fat yield has not been well-reported, however in this study it is not significant ($p = 0.60$).

Estimation of Prediction Accuracy. The base model has an accuracy of 0.406 and a regression coefficient of 1.154 (Table 3).

Compared to the base model, accuracy of DGV is increased slightly, but significantly, by addition of all QTN simultaneously as random covariates in the model, but decreased significantly by including them as fixed covariates (Table 3). There is no significant change in accuracy when these genotypes were fit as fixed classes. The regression coefficient is significantly different from the base model when all QTN are fit as a fixed covariate but not when fit as a random covariate or a fixed class. However, bias is higher when all QTN are fit as covariates. When all QTN are fit as random covariates the model can take into account any variation not captured by LD with the 35k markers, so accuracy increases. When fitting QTN as fixed covariates dominance at those loci is not taken into account which may cause the decrease in accuracy and increase in bias.

Including *DGAT1* as a random covariate, fixed covariate or fixed class does not impact accuracy of DGV prediction, however bias is decreased when *DGAT1* is fit as a fixed covariate (Table 3). The bias when *DGAT1* genotype is fit as a fixed class is actually lower than when

DGAT1 is fit as a covariate, however the bias is not significantly different from that in the base model. These results suggest that while including *DGAT1* may not have an effect on accuracy, it may slightly reduce bias of DGV estimates. One possible reason why such a large QTN (Table 2) does not have any impact on accuracy when smaller QTN do (Table 3) is because there is a SNP in the 35k subset that is almost in complete LD with *DGAT1* in this population (Figure 1) and this SNP may be picking up the effect of *DGAT1*.

Table 3: Accuracy and bias of genomic prediction

Model[1,2]	Accuracy (s.e.)[3]	Bias (s.e.)[4]
Base	0.406 (0.009)	1.154 (0.035)
1) All	0.407 (0.009)*	1.155 (0.035)
2) All	0.372 (0.008)*	0.726 (0.021)*
3) All	0.408 (0.006)	1.102 (0.039)
1) DGAT1	0.406 (0.009)	1.154 (0.035)
2) DGAT1	0.406 (0.009)	1.144 (0.034)*
3) DGAT1	0.406 (0.008)	1.135 (0.037)
1) GHR	0.407 (0.009)	1.156 (0.035)
2) GHR	0.409 (0.007)	1.141 (0.031)
3) GHR	0.409 (0.007)	1.140 (0.032)
1) AGPAT6	0.407 (0.009)	1.156 (0.035)
2) AGPAT6	0.407 (0.009)	1.146 (0.036)
3) AGPAT6	0.407 (0.009)	1.145 (0.037)
1) PLAG1	0.406 (0.009)	1.155 (0.035)
2) PLAG1	0.405 (0.008)	1.146 (0.037)
3) PLAG1	0.406 (0.008)	1.145 (0.035)
1) CHR8	0.406 (0.009)	1.155 (0.035)
2) CHR8	0.406 (0.008)	1.153 (0.035)
3) CHR8	0.406 (0.008)	1.148 (0.032)

* Value is significantly different from base model at $\alpha = 0.01$ for a paired t-test, pairing cross-validation sets.

[1] 1 = Random Covariate, 2 = Fixed Covariate, 3 = Fixed Class

[2] All refers to all five QTN fit simultaneously

[3] Correlation between DGV and DEBV.

[4] Regression coefficient for DGV on DEBV.

While including *GHR* in the model does not significantly change results from the base model its inclusion in any form gives the highest accuracy of any of the models. It is possible that, given the relatively low minor allele frequency of *GHR* (Table 1), that there were insufficient qq homozygotes to get significant results for this QTN.

Including *AGPAT6* does not have significant effect on either accuracy or bias of DGV prediction (Table 3). While the difference is not significant, including *AGPAT6* gives slightly higher accuracy and lower bias for all models.

Fitting either *PLAG1* or *CHR8* did not have significant effect on accuracy or bias. These results were expected from fitting two SNPs that do not have a large effect on milk fat yield.

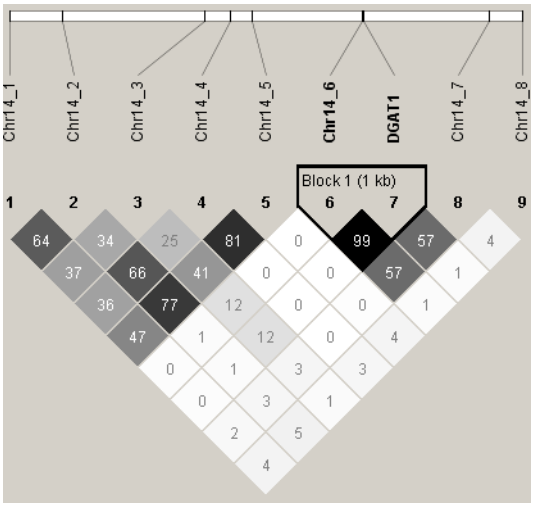

Figure 1. LD plot from Haploview for the 1Mb window that includes the *DGAT1* mutation with R^2 values

Conclusion

Allele substitution and additive effects were both consistent with estimates of genetic variance explained by each of the five QTN. Significant dominance was observed at three QTN: *DGAT1* (partial), *AGPAT6* (over) and *PLAG1* (complete).

Including all five QTN as random covariates slightly increased the accuracy of prediction compared to the base model but fitting them as fixed covariates significantly decreased accuracy of prediction and increased bias. While not significant, it appears that including QTN as either a fixed covariate or fixed class will increase the accuracy and decrease the bias, although only to a small extent, however more animals will be needed to confirm these findings. While some of these changes are significant, including these QTN does not have a sufficient increase in accuracy or bias to make a difference in overall genetic gain.

Literature Cited

Blott, S., Kim, J.-J., Moisio S. et al (2003) *Genetics* 163:253-266.

Brotherstone, S. (1994) *Animal Prod*., 59:183-187.

Cohen-Zinder, M., Seroussi, E., Larkin D.M. et al (2005) *Genome Res.*, 15:936-944.

Fernando, R.L., and Garrick, D.J. (2009) GenSel – User Manual.

Johnson, P.L., Dodds, K.G., Bain, W.E., et al. (2009) *J. Anim. Sci.*, 87(6):1856-1864.

Grisart, B., Coppieters, W., Farnir, F. et al. (2002) *Genome Res.*, 12:222-231.

Karim, L., Takeda, H., Lin, L. et al. (2011) *Nat. Genet.*, 43(5):405-413.

Keuhn, C., Edel, C., Weikard, R., et al. (2007) *BMC Genet.*, 8(1):62.

Littlejohn M.D., Grala, T., Tiplady K., et al. (2012) *Anim. Genet*., 43(5):591-594.

Littlejohn M.D., Tiplady K., Lopdell T., et al. (2014) *PLOS ONE*.

Thomsen, H., Lee, H.K., Rothschild, M.F. (2004) *J. Anim. Sci.*, 82(8):2213-2228.