

Genomic Prediction of Milk Fat using Fixed Length Haplotypes

A.S. Leaflet R3071

Melanie Hayr, Ph.D. Student, Animal Science;
Xiaochen Sun, Post-Doc, Animal Science;
Tom Druet, Senior Scientist, University of Liège;
Dorian Garrick, Professor, Animal Science

Summary and Implications

Performance of genomic prediction from haplotype models utilizing fixed length haplotype blocks ranging from 125 kb – 2 Mb and haplotype allele frequency cutoffs ranging from 1-10% were compared to a model using SNP genotypes. Milk fat yield deviations and genotypes at 37,740 SNPs were from 38,385 Holstein Friesian, Jersey and KiwiCross cows (Training=23,907; Validation=14,478) from New Zealand. This study showed slight improvement in accuracy and bias of genomic prediction when using 125 kb haplotype blocks with a 1% filter; however this was associated with a large increase in run-time. Haplotype blocks larger than 1 Mb are not appropriate for genomic prediction in this population.

Introduction

Direct Genomic Values (DGV) must be as accurate as possible to reduce the accumulation of inaccuracies in pedigrees resulting from selection of unproven parents. DGV have traditionally been estimated using Single Nucleotide Polymorphism (SNP) genotypes. With a sufficient number of related animals genotyped, SNP genotypes can be accurately phased to determine which allele was inherited from each parent. Neighboring SNPs that tend to be inherited together can be grouped into haplotype blocks. A haplotype allele is a combination of SNP alleles in a haplotype block. Fitting haplotype alleles as covariates rather than SNP genotypes may improve accuracy, bias or run-time due to haplotypes typically being in higher linkage disequilibrium with QTL or because there may be fewer haplotype alleles than SNPs.

Materials and Methods

Approximately 58,000 Holstein Friesian, Jersey and KiwiCross dairy cattle from New Zealand were genotyped on Illumina Bovine SNP50 or HD panels and 37,740 SNP genotypes were phased using LinkPHASE and DAGPHASE. Haplotype blocks were defined based on

length, with five lengths evaluated: 125 kb, 250 kb, 500 kb, 1 Mb and 2 Mb; corresponding to, on average, 2, 4, 8, 15 and 30 SNPs per block.

The training dataset consisted of a subset of 23,907 genotyped females born before 1 June 2008 with milk fat yield deviation records. The validation set consisted of 14,478 females born 1 June 2008 – 30 June 2010 with milk fat yield deviation records. BayesA was run in GenSel fitting either SNP genotype or haplotype allele dosage. Haplotype alleles with low frequency in the training dataset (1%, 2.5%, 5% or 10%) were excluded in the analysis. Individuals in the validation dataset with an excluded allele have a DGV of zero for that block.

Results and Discussions

The performance of each model was assessed by regressing yield deviations on DGV and inspecting the accuracy (correlation coefficient) and bias (regression coefficient – 1). The SNP model had an accuracy of 0.304 and bias of 0.052. The model that was the most similar to the SNP model in terms of number of covariates fitted, run-time, accuracy and bias used 125 kb haplotype blocks and a 10% filter (Table 1). Haplotype models that included blocks of 1 or 2 Mb had lower accuracies and more extreme biases than the SNP model (Table 1).

The most accurate haplotype model utilized 500 kb haplotype blocks and a frequency filter of 1% (Table 1). The model that performed the best, in terms of both accuracy and bias, utilized 125 kb haplotype blocks and a 1% filter (Table 1). These two models took ~24 hours to run compared to the SNP model, which took 13 hours; because they fitted approximately 2× number covariates as the SNP model (haplotype alleles vs. SNP genotypes).

These results suggest fitting fixed length haplotypes does not provide substantial improvement over fitting SNP genotypes. If fixed block haplotype models are used, it is recommended to keep haplotype blocks < 1 Mb. Using haplotype blocks that more appropriately model inheritance of SNP alleles may increase accuracy and decrease bias of haplotype models.

Acknowledgements

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

Iowa State University Animal Industry Report 2016

Table 1 Accuracy (correlation) and Bias (regression - 1) of Genomic Prediction of Milk Fat Yield using Haplotypes.

Filter Level	Prediction Accuracy					Prediction Bias				
	125 kb	250 kb	500 kb	1 Mb	2 Mb	125 kb	250 kb	500 kb	1 Mb	2 Mb
1%	0.306	0.307	0.308	0.283	0.271	0.048	0.045	0.052	-0.159	-0.154
2.5%	0.306	0.306	0.305	0.300	0.276	0.049	0.049	0.063	0.075	-0.066
5%	0.305	0.305	0.302	0.290	0.263	0.052	0.055	0.076	0.096	0.120
10%	0.305	0.301	0.285	0.253	0.208	0.053	0.065	0.113	0.151	0.175

Cells in light blue represent analyses that took less time to run than the SNP model (13 hours)