A Nested Mixture Model for Genomic Prediction Using Whole-Genome SNP Genotypes

A.S. Leaflet R3060

Jian Zeng, Postdoc Fellow;
Dorian Garrick, Professor;
Jack Dekkers, Professor;
Rohan Fernando, Professor,
Department of Animal Science

Summary and Implications

We propose a novel model (BayesN) for genomic prediction, where multiple markers in a small segment are simultaneously fitted to jointly capture the effect of major genes (QTL) in the segment. Compared with BayesB, in which the effects of neighboring markers are *a priori* assumed to be independent, BayesN gave higher accuracies of prediction and required less computing effort. BayesN is an accurate and practical method for analyzing high-density markers, especially for traits influenced by rare QTL alleles.

Introduction

Genomic prediction exploits single nucleotide polymorphism (SNP) markers across the whole genome for predicting genetic merit of selection candidates. It has been successfully integrated into breeding programs for many species. However, accuracy of prediction is often low for traits with low minor allele frequency (MAF) or so-called rare alleles at quantitative trait loci (QTL). This is because a SNP that has a high MAF cannot be very informative for a QTL with a low MAF. In most models for genomic prediction, e.g. BayesA, B, C, R, the effects of neighboring SNPs are assumed to be independent, which results in their predictive ability being low when individual SNPs are not very informative.

We proposed to address this problem by using a prior that allows multiple SNPs surrounding a QTL to jointly capture the effect of that QTL. Thus, a nested mixture model (BayesN) was developed, where the effects of SNPs in every 200 kb non-overlapping window of the genome are collectively considered in the model. The objective was to compare the predictive and computational performance of BayesN with that of BayesB.

Materials and Methods

Illumina 777K BovineHD genotypes from 948 Angus cattle were used to simulate 5,000 offspring, with 4,000 used for training and 1,000 for validation. Scenarios with phenotypes generated using 300 common (MAF > 0.05) or rare (MAF < 0.05) QTL randomly selected from segregating SNPs were replicated 8 times. SNPs corresponding to QTL were masked from a 600k panel comprising SNPs with MAF > 0.05 or a 50k evenly spaced subset of these, which was used for training. The QTL effects were sampled from a standard normal distribution. Trait phenotypes with heritability 0.5 were simulated by adding random standard normal deviates to the true breeding values (TBV).

Genomic estimated breeding values (GEBV) were obtained from BayesB or BayesN with a prior assumption that 600 SNPs had nonzero effects for either the 50k or 600k panels. Accuracies of prediction were calculated from the correlation between GEBV and TBV for validation individuals.

Results and Discussion

Compared with BayesB, BayesN improved the accuracy of prediction up to 2.0% with 50k SNPs and up to 7.0% with 600k SNPs, most improvements occurring in the rare QTL scenario (Table 1). Computing time was reduced up to 60% with 50k SNPs and up to 75% with 600k SNPs. Due to the use of a prior that allows dependence of SNP effects in a window, effects of SNPs with small effects were shrunk more heavily towards zero, while real signals were shrunk less. In conclusion, BayesN is an accurate and computationally efficient method for genomic prediction with high-density SNP genotypes, especially for traits with rare QTL.

Acknowledgments

This work was supported by National Institutes of Health grant R01GM099992.

Table 1. Predictive and computational performance of BayesB and BayesN

	Common (high MAF) QTL				Rare (low MAF) QTL			
	50k Panel		600k Panel		50k Panel		600k Panel	
	BayesB	BayesN	BayesB	BayesN	BayesB	BayesN	BayesB	BayesN
Accuracy of prediction	0.723	0.728	0.785	0.786	0.579*	0.590*	0.652*	0.695*
Run time[1] (hr)	1.1	0.8	12.6	3.0	1.1	0.8	10.6	2.9

*Significant difference between BayesB and BayesN at the 0.01 level
[1]ISU CyEnce cluster with 2.0 GHz 8-Core Intel E5 289 2650 processors