Improved Accuracy of Across-breed Genomic Prediction Using Haplotypes in Beef Cattle Populations

A.S. Leaflet R3065

Xiaochen Sun, Postdoctoral Research Associate;
Hailin Su, Postdoctoral Research Associate;
Dorian Garrick, Professor, Department of Animal Science,
Iowa State University

Summary and Implications

Genomic prediction uses a reference population of animals with SNP genotypes and phenotypes to predict the merit of selection candidates that typically do not have observed phenotypes. Accuracy of genomic prediction from models that fitted 50K SNP genotypes was low when selection candidates were from a breed only distantly related with the breeds in the reference population. That accuracy was not improved by increasing SNP density from 50K to a ten-fold higher density using imputation. This indicates that the accuracy of genomic prediction mainly came from family-wise co-segregation information. In this study, a genomic prediction model that fitted genome-wide 100 kilo-bp (Kbp) haplotypes improved accuracy for breeds that were not in the reference population. The haplotype model is a more accurate alternative to the SNP model for genomic prediction when animals of the same breed as the prediction candidates are not available for the reference population.

Introduction

Genomic prediction in both simulated and field datasets has shown that prediction accuracy could be improved by assembling a large reference population including multiple breeds. However, improved accuracy using the SNP model in multi-breed prediction has mainly been observed in candidate breeds that were included in the reference population, whereas the accuracy for distantly related breeds not included in the reference population has remained low or even zero. Evidence suggests that the level and consistency of linkage disequilibrium (LD) between quantitative trait loci (QTL) and SNPs was low among distantly related breeds, and hence prediction accuracy of the SNP model mainly came from co-segregation information among related individuals that was implicitly captured by SNP genotypes, rather than LD that was persistent across breeds.

Studies of LD between multi-allelic SNP haplotypes and bi-allelic SNP loci implied that complete LD existed between haplotypes and QTL with SNP density similar to the Bovine HD 770K panel. Therefore, fitting haplotypes could improve prediction accuracy relative to the SNP model for candidate breeds that were not in the reference population. Simulation studies showed that, with SNP density similar to 600K panel across the bovine genome, the haplotype model had significantly higher accuracy than SNP models when LD between SNP and QTL was incomplete. In this study, genomic prediction was performed for growth and carcass traits in a beef cattle population composed of five breeds, using the SNP model or a model fitting genome-wide 100 Kbp haplotypes, to verify that fitting haplotypes improves accuracy over SNP genotypes when the candidate breed was not included in the reference population.

Materials and Methods

The beef cattle datasets included 1,905 Angus (AAN), 1,449 Charolais (CHA), 1,214 Gelbvieh (GVH), 1,500 Hereford (HER) and 1,500 Simmental (SIM) animals that had been genotyped with the Bovine 50K SNP chip. Within each breed, the 50K genotypes were imputed to ten-fold higher density HD panel density using samples of HD-genotyped animals from each of the same breeds as imputation reference populations. The imputed HD genotypes were phased for haplotype analysis. A total of 607,039 HD SNPs on chromosomes 1 ~ 29 were used in analysis after removing SNPs with MAF ≤ 0.01.

Genomic prediction was performed for birth weight (BW), weaning weight (WW), yearling weight (YW), back fat thickness (BF) and ribeye area (RA). Deregressed breeding values from national cattle evaluations with parent average added back were used as phenotypes. To predict the genomic estimated breed values (GEBV) of animals in each breed, the animals in the other four breeds were used as the reference population.

Each chromosome was divided into consecutive non-overlapping windows with length 100 Kbp. The haplotype model fitted the effect of each common haplotype (allele frequency ≥ 0.01) at each chromosome window as an independent random effect. Method BayesC was used to estimate haplotype effects (or SNP allele substitution effects) in the haplotype (or SNP model).

Results and Discussion

For traits BW, BF and RA, the haplotype model had significantly higher accuracy than the SNP model in each breed (except for RA in breed SIM), while for traits WW and YW, the advantage of the haplotype model over the SNP model varied across breeds (Table 1). The average improvement in accuracy of the haplotype model over the SNP model was 17.1%, 34.8%, 42.4%, 8.5% and 28.7% in breeds AAN, CHA, GVH, HER and SIM, respectively. A possible reason for the relatively low improvement for breed HER is that HER animals are more distantly related and share less haplotypes with the other four breeds.

In conclusion, with SNP density similar to the Bovine HD panel, the haplotype model improves prediction accuracy for a candidate breed that is not included in the reference population. The haplotype model can be used to improve prediction accuracy when animals of the same breed as prediction candidates are not available for use in the reference population.

Acknowledgments

This work was supported by the Zoetis Cattle Call 2014 Research Grant.

Table 1. Accuracy of genomic predictions for Angus (AAN), Charolais (CHA), Gelbvieh (GVH), Hereford (HER) and Simmental (SIM) breeds using the other four breeds in the reference population for each of five traits

Trait	Breeds									
	AAN		CHA		GVH		HER		SIM	
	SNP[1]	Haplo[2]	SNP	Haplo	SNP	Haplo	SNP	Haplo	SNP	Haplo
BW[3]	0.309	0.367	0.290	0.366	0.284	0.342	0.302	0.317	0.388	0.420
WW	0.474	0.467	0.263	0.219	0.212	0.214	0.226	0.146	0.168	0.164
YW	0.560	0.590	0.101	0.156	0.273	0.277	0.247	0.131	0.191	0.210
BF	0.165	0.372	0.001	0.118	0.181	0.438	0.018	0.185	0.074	0.333
RA	0.415	0.455	0.071	0.120	0.065	0.174	0.130	0.222	0.167	0.145
Mean	0.385	0.450	0.145	0.196	0.203	0.289	0.185	0.200	0.198	0.254

[1]Prediction accuracy of the SNP model
[2]Prediction accuracy of the haplotype model
[3]BW=birth weight; WW=weaning weight; YW=yearling weight; BF=back fat; RA=ribeye area