World Academy of Science, Engineering and Technology
International Journal of Animal and Veterinary Sciences
Vol:13, No:02, 2019

Genomic Prediction Reliability Using Haplotypes Defined by Different Methods

Authors : Sohyoung Won, Heebal Kim, Dajeong Lim

Abstract : Genomic prediction is an effective way to measure the abilities of livestock for breeding based on genomic estimated breeding values, statistically predicted values from genotype data using best linear unbiased prediction (BLUP). Using haplotypes, clusters of linked single nucleotide polymorphisms (SNPs), as markers instead of individual SNPs can improve the reliability of genomic prediction since the probability of a quantitative trait loci to be in strong linkage disequilibrium (LD) with markers is higher. To efficiently use haplotypes in genomic prediction, finding optimal ways to define haplotypes is needed. In this study, 770K SNP chip data was collected from Hanwoo (Korean cattle) population consisted of 2506 cattle. Haplotypes were first defined in three different ways using 770K SNP chip data: haplotypes were defined based on 1) length of haplotypes (bp), 2) the number of SNPs, and 3) k-medoids clustering by LD. To compare the methods in parallel, haplotypes defined by all methods were set to have comparable sizes; in each method, haplotypes defined to have an average number of 5, 10, 20 or 50 SNPs were tested respectively. A modified GBLUP method using haplotype alleles as predictor variables was implemented for testing the prediction reliability of each haplotype set. Also, conventional genomic BLUP (GBLUP) method, which uses individual SNPs were tested to evaluate the performance of the haplotype sets on genomic prediction. Carcass weight was used as the phenotype for testing. As a result, using haplotypes defined by all three methods showed increased reliability compared to conventional GBLUP. There were not many differences in the reliability between different haplotype defining methods. The reliability of genomic prediction was highest when the average number of SNPs per haplotype was 20 in all three methods, implying that haplotypes including around 20 SNPs can be optimal to use as markers for genomic prediction. When the number of alleles generated by each haplotype defining methods was compared, clustering by LD generated the least number of alleles. Using haplotype alleles for genomic prediction showed better performance, suggesting improved accuracy in genomic selection. The number of predictor variables was decreased when the LD-based method was used while all three haplotype defining methods showed similar performances. This suggests that defining haplotypes based on LD can reduce computational costs and allows efficient prediction. Finding optimal ways to define haplotypes and using the haplotype alleles as markers can provide improved performance and efficiency in genomic prediction.

Keywords : best linear unbiased predictor, genomic prediction, haplotype, linkage disequilibrium
Conference Title : ICABG 2019 : International Conference on Animal Breeding and Genetics
Conference Location : Sydney, Australia
Conference Dates : February 27-28, 2019