Genomic selection in admixed and crossbred populations

A. Toosi, R. L. Fernando, and J. C. M. Dekkers

Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames 50011

Abstarc :

In livestock, genomic selection (GS) has primarily been investigated by simulation of purebred populations. Traits of interest are, however, often measured in crossbred or mixed populations with uncertain breed composition. If such data are used as the training data for GS without accounting for breed composition, estimates of marker effects may be biased due to population stratification and admixture. To investigate this, a genome of 100 cM was simulated with varying marker densities (5 to 40 segregating markers per cM). After 1,000 generations of random mating in a population of effective size 500, 4 lines with effective size 100 were isolated and mated for another 50 generations to create 4 pure breeds. These breeds were used to generate combined, F1, F2, 3- and 4-way crosses, and admixed training data sets of 1,000 individuals with phenotypes for an additive trait controlled by 100 segregating QTL and heritability of 0.30. The validation data set was a sample of 1,000 genotyped individuals from one pure breed. Method Bayes-B was used to simultaneously estimate the effects of all markers for breeding value estimation. With 5 (40) markers per cM, the correlation of true with estimated breeding value of selection candidates (accuracy) was greatest, 0.79 (0.85), when data from the same pure breed were used for training. When the training data set consisted of crossbreds, the accuracy ranged from 0.66 (0.79) to 0.74 (0.83) for the 2 marker densities, respectively. The admixed training data set resulted in nearly the same accuracies as when training was in the breed to which selection candidates belonged. However, accuracy was greatly reduced when genes from the target pure breed were not included in the admixed or crossbred population. This implies that, with high-density markers, admixed and crossbred populations can be used to develop GS prediction equations for all pure breeds that contributed to the population, without a substantial loss of accuracy compared with training on purebred data, even if breed origin has not been explicitly taken into account. In addition, using GS based on high-density marker data, purebreds can be accurately selected for crossbred performance without the need for pedigree or breed information. Results also showed that haplotype segments with strong linkage disequilibrium are shorter in crossbred and admixed populations than in purebreds, providing opportunities for QTL fine mapping.

Key Word :
admixture, crossbreeding, genomic selection, marker-assisted selection