

Abstract by Joel Ira Weller - Predictive ability of selected subsets of single nucleotide polymorphisms (SNPs) for moderately sized dairy cattle populations

Several studies have shown that computation of genomic estimated breeding values (GEBV) with accuracies significantly greater than parent average EBV requires genotyping of at least several thousand progeny-tested bulls. For all published analyses, GEBV computed from selected samples of markers have lower or equal accuracy than GEBV derived based on all valid SNPs. In the current study we report on four new methods for selection of markers. Milk, fat, protein, somatic cell score, fertility, persistency, herd-life, and the Israeli selection index were analyzed. The 969 Israel Holstein bulls genotyped with EBV for milk production traits computed from daughter records in November, 2011, were assigned into a training set of 829 bulls with progeny test EBV in June, 2008, and a validation set of 140 young bulls. Numbers of bulls in the two sets varied slightly among the nonproduction traits. In Method 1, SNPs were first selected for each trait based on a linear model analysis of the effect of each marker on the bulls' current EBV for each trait. A subset of these SNP then was analyzed by a REML model including relationships. Method 2 was the same as Method 1, except that that the dependent variable was the 2008 EBV. In Method 3, the SNPs with the greatest effects on the 2008 EBV, as determined by the REML analysis were deleted. Of the remaining SNPs, the markers with the greatest effects on 2011 EBV were retained. In Method 4, the SNPs with the greatest effects on the 2008 EBV, as determined by the REML analysis were deleted. Of the remaining SNPs, the markers with the greatest change in allele frequency between the bulls in the training set, and the validation bulls were retained for analysis. For all methods, the numbers of SNPs deleted and retained were varied to obtain a maximum correlation between the GEBV and EBV of the validation bulls. In Methods 1 and 2, the number of SNPs included in the analyses was varied over the range of 400 to 6000. For each trait, except fertility, an optimum number of markers between 600 and 2000 was obtained for Method 1, based on the correlation between the GEBV and current EBV of the validation bulls. For all traits, the difference between the correlation of GEBV and current EBV and the correlation of the parent average and current EBV was >0.1 . Method 2 was inferior to Method 1 and generally no better than parent average EBV, but Method 3 outperformed Method 1. Even Method 4, in which selection of markers is based only on information available at the time the training set is generated, correlations between GEBV and current EBV were on the average 0.042 higher than correlations of parent averages with current EBV. Furthermore, GEBV were less biased than parent averages. It is likely that other methods of SNP selection could improve upon these results.