# Accuracy of Genomic Prediction when Accounting for Population Structure and Polygenic Effects

## A.S. Leaflet R2817

Napapan Piyasatian, Research Associate;
Jack C.M. Dekkers, Professor,
Department of Animal Science, ISU

**Summary and Implications**

Accuracy of genomic estimated breeding values obtained using the standard marker effect model was compared with models that account for population structure, either by applying a transmission disequilibrium test (TDT) approach or by fitting polygenic effects. The TDT approach was inferior to the standard model, whereas fitting polygenic effects in addition to marker effects increased the accuracy of estimated breeding values of the progeny of training individuals but also seven generations after training. Thus, fitting polygenic effects enhances utilization of genomic information both in the short and long-term.

**Introduction**

Linkage disequilibrium (LD) between quantitative trait loci (QTL) and genetic markers is an important source of information in genomic prediction, besides co-segregation of QTL and marker alleles and additive-genetic relationships captured by genetic markers. Quantitative trait loci mapping studies have shown that accounting for population structure, either by a transmission disequilibrium test (TDT) or by fitting polygenic effects in addition to markers, reduces spurious associations and thereby decrease the number of false positive QTL. Applying such methods to genomic prediction may reduce prediction errors and therefore increase accuracy of the resulting genomic estimated breeding values (GEBVs). Previous genomic selection studies have shown that modeling polygenic effects in addition to genetic markers results in higher accuracy of GEBVs for progeny of training individuals. However, this advantage may not only be due to a better utilization of LD information but also result from better exploiting relationship information that is not captured by genetic markers. The objective of this study was to compare standard genomic prediction methods with approaches that account for population structure.

**Materials and Methods**

The accuracy of GEBVs from the following four models were compared: 1) the standard marker effects model, 2) the same model with polygenic effects, 3) a genomic TDT (GTDT) model that fits for each marker a parent average effect and a mendelian sampling effect, and 4) the GTDT model with polygenic effects. Stochastic simulations were conducted with varying numbers of QTL and genetic markers, training data size and extent of LD, while simulating an unbalanced population structure with influential sires. Accuracies were obtained for both progeny of the training generation and for individuals seven generations after training.

**Results and Discussion**

As expected, accuracies decreased across generations due to the decay of genetic relationships captured by genetic markers. The decay of accuracy was larger with more QTL because there was less accuracy due to LD and, hence, accuracy due to the decay of relationships has a larger effect.

Fitting polygenic effects increased accuracies for all seven validation generations, for all scenarios, and for both the standard and the GTDT models with five QTL but not for GTDT models with 50 QTL (Table1). Accuracies from models with polygenic effects tended to be higher in early generations after training, because polygenic effects not only enhanced the LD signal but also captured the remaining relationship information that was not exploited by markers, depending on the extent of LD across chromosomes and training size. The increase in accuracy in the first generation after training obtained by the standard model with polygenic effects was higher for the larger training size (Table 1), because markers captured less relationship information as the number of families in the training increased, as shown by Habier et al. (2012). The increase in accuracy was also higher for the low LD case, because accuracy due to LD was lower such that less genetic variation was captured by LD information, leaving more room for relationship information to be picked up by polygenic effects. In the last generation, the increase in accuracy with polygenic effects was due to a better utilization of LD information, which may result from removing spurious LD due to population structure and thereby decreasing false positives markers effects and prediction errors. However, the increase in accuracy of 0.01 on average was rather small.

The GTDT models almost always resulted in lower accuracy than the standard models and did not better account for population structure, as they showed a similar increase in accuracy as the standard models with polygenic effects for a low number of QTLs. In conclusion, the standard markers effects models with polygenic effects improved the utilization of LD information when predicting genomic breeding values.

**Acknowledgments**

This work was supported by USDA NIFA Grant No. 2010-65205-20341.

**Table 1. Accuracy of genomic prediction of the standard marker effects model and for genomic TDT models with and without polygenic effects in the 1st and 7th generation after training using method BayesB with π of 0.95. The simulation scenarios were varied with five and 50 QTL/chromosome, different amounts of LD, training size (2000 individuals) and marker density (1000 markers).**

| Scenario \ Models | Standard marker effects model | | | | Genomic TDT model [1] | | | |
|---|---|---|---|---|---|---|---|---|
| | Without Polygenic effects | | With Polygenic effects | | Without Polygenic effects | | With Polygenic effects | |
| | Generation [2] | | Generation | | Generation | | Generation | |
| | 1 | 7 | 1 | 7 | 1 | 7 | 1 | 7 |
| 5 QTL | 0.73 | 0.60 | 0.75 | 0.61 | 0.67 | 0.51 | 0.69 | 0.52 |
| 5 QTL, low LD | 0.66 | 0.51 | 0.71 | 0.52 | 0.64 | 0.42 | 0.66 | 0.44 |
| 5 QTL, double training size | 0.72 | 0.62 | 0.76 | 0.63 | 0.70 | 0.56 | 0.72 | 0.58 |
| 50 QTL | 0.63 | 0.43 | 0.67 | 0.44 | 0.60 | 0.36 | 0.61 | 0.36 |
| 50 QTL, double marker density | 0.72 | 0.54 | 0.74 | 0.54 | 0.66 | 0.45 | 0.66 | 0.45 |
| 50 QTL, double training size | 0.65 | 0.46 | 0.70 | 0.47 | 0.66 | 0.39 | 0.66 | 0.40 |

[1] Transmission Disequilibrium Test
[2] Generation after training