

# Increasing the accuracy of genomic prediction for crossbred livestock: examples from dairy cattle\*

I.M. MacLeod<sup>1</sup>, M. Khansefid<sup>1</sup>, M. Haile-Mariam<sup>1</sup>, R. Xiang<sup>1,4</sup>, S. Bolormaa<sup>1</sup>, G. de Jong<sup>2</sup>, E. O’Conner<sup>3</sup>, C. Schrooten<sup>2</sup>, J.E. Pryce<sup>1,5</sup>, M.E. Goddard<sup>1,4</sup>, H.D. Daetwyler<sup>1,5</sup>

<sup>1</sup>Agriculture Victoria, Department of Jobs, Precincts and Regions, Bundoora, VIC 3083, Australia, <sup>2</sup>CRV, 6800 AL Arnhem, the Netherlands, <sup>3</sup>CRV Ambreed, Hamilton, 3216 New Zealand, <sup>4</sup>Faculty of Veterinary & Agricultural Science, University of Melbourne, VIC 3052, Australia, <sup>5</sup>School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia

\* This research is part of a larger study submitted to Frontiers in Genetics.

## Background

Genomic prediction for purebred dairy cattle shows high adoption rates globally. There is now increased demand for genomic predictions in crossbred dairy animals: either systematically crossbred, crossbred to upgrade to a different breed, or offspring of crossbred bulls. To date there are limited studies on optimising genomic prediction in crossbred dairy cattle. National genomic evaluations are often within breed and in most countries one breed is predominant with some minor breed and crossbred use. The aim of this study was to evaluate practical approaches to optimise genomic prediction for crosses and purebreds and in a single evaluation.

## Results & Discussion

The accuracy of genomic prediction was consistently higher for all pure and crossbred validation sets using the Custom50K panel versus the Standard50K panel (Fig 1). The custom panel was developed using multi-omics analyses of sequence data (Xiang et al, ICQG 2020). None of the cows were included in the research to develop the custom 50K panel. The training approach that used separate purebred prediction equations proportionally for crossbreds gave a similar accuracy to combining the two purebred sets in a single training set but showed less bias than the latter (results not shown). Adding crossbreds to the training set considerably increased the accuracy of prediction for crossbreds, and maintained purebred accuracy, compared to using predictions only from purebred sets (Fig 1: set 2 & 3 vs. set 1). Interestingly, there was little difference in accuracy using either all available animals (2) or approximately half the animals (3) where the latter had a balanced proportion of Holstein and Jersey breeds. Most importantly, use of the balanced breed training set consistently reduced bias compared to the full Holstein dominated set (Fig 2: set 3 vs. set 2) and this warrants further investigation.

## Materials & Methods

GBLUP method using genotypes from two different SNP panels:

1. “Standard50K”= commercial industry 50K panel,
2. “Custom50K”= custom 50K panel [1]

Full Training Data Available:

13,985 Holstein (H), 4,484 Jersey (J) and 18,226 crossbred (XBRED) cows (total N=36,695). The crossbred cows were from New Zealand where use of crossbred bulls is not uncommon, so that cows may represent several generations of inter-crossing. They were assigned to breed groups of ~50%H:50%J, or ~75%H:25%J, or ~25%H:75%J based on PCA and pedigree checks.

Training the genomic predictions:

1. Two separate training sets: one pure Holstein and the other pure Jersey. The crossbreds were predicted by combining their two purebred predictions as a weighted average where the weights were based on their breed proportions [2].
2. A single training set with all available purebreds and crossbreds (N=36,695) that was dominated by Holstein purebreds and crosses (“Full Training Data”).
3. A single training set with balanced breed proportions in purebreds and crosses (N=15,734) based on random removal of excess H and ~75%H:25%J animals from 2 above.

Cow Validation Sets:

(a) 1,002 pure H, (b) 863 75%H:25%J, (c) 868 50%H:50%J, (d) 324 25%H:75%J, (e) 532 pure J

Traits:

Milk Yield, Fat Yield and Protein Yield. The average accuracy (correlation between observed and predicted phenotype) and bias (regression coefficient) for these traits is presented here because results were consistent across traits.

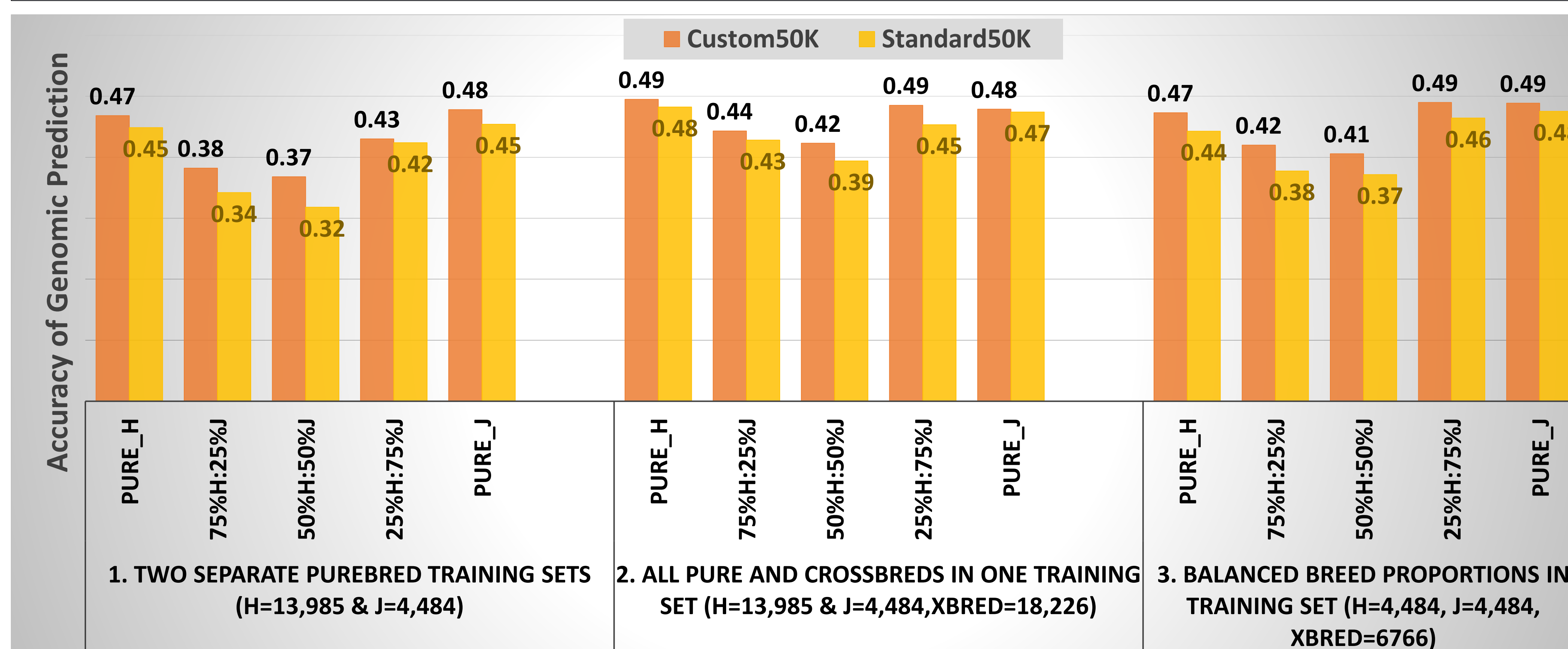


Figure 1. Accuracy of genomic prediction in pure and crossbred validation sets (pure H, 75%H:25%J, 50%H:50%J, 25%H:75%J and pure J) using 2 different 50K SNP panels and three approaches for training predictions.



Figure 2. Bias of genomic prediction in pure and crossbred validation sets (pure H, 75%H:25%J, 50%H:50%J, 25%H:75%J and pure J) using 2 different 50K SNP panels and three approaches for training predictions.

## References

1. Xiang, R., Berg, I.v.d., MacLeod, I.M., et al. (2019). Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. Proc Nat Acad Sci 116, 19398
2. VanRaden, P.M., Tooker, M.E., Chud, et al. (2020). Genomic predictions for crossbred dairy cattle. J Dairy Sci 103, 1620