

# Bayesian Methods for Genomic Prediction and Genome-Wide Association Studies combining Information on Genotyped and Non-Genotyped Individuals

## A.S. Leaflet R2865

Rohan Fernando, Professor; Jack Dekkers, Professor;  
Dorian Garrick, Professor,  
Department of Animal Science, Iowa State University

### Summary and Implications

Genomic prediction involves using high-density marker genotypes to characterize the impact on performance of every region of the genome, and using that information to predict performance of genotyped selection candidates. This is a relatively new technology and is now gaining traction in personalized medicine and in various livestock industries. Our new approach promises to overcome serious limitations with existing techniques for genomic prediction.

### Introduction

Genomic prediction uses high-density genotypes to predict the likely breeding performance of selection candidates based on results relating the performance of historical animals to their marker genotypes. Genome-wide association studies (GWAS) use high-density genotypes and performance information to identify regions of the genome that demonstrate association with traits of interest.

Genomic prediction and GWAS are relatively straightforward when all individuals have both phenotypic and genotypic data. In many breeding applications, however, genotyped individuals have descendants or other relatives that have been phenotyped but not genotyped. One common approach to use information on these relatives is to use all pedigree and performance data to estimate breeding values (EBV) and then deregress this information on genotyped animals to use as data in genomic prediction or GWAS. Many different statistical models can be applied in that analysis, including BayesA, BayesB, BayesCpi, and Bayesian Lasso. In a second step, the direct genomic values (DGV) produced by one of these methods are then commonly blended with pedigree based EBV predictions to produce genome-enhanced EBV (GE-EBV).

Another approach is to use a single-step analysis that involves creation of a genetic variance-covariance matrix among all animals using information on both genotyped and non-genotyped animals. That approach requires brute force inversion of two matrices of order equal to the number of genotyped animals, and requires the genomic and pedigree-based relationship matrices to have the same definition of base or founder animals. The method is also limited to approaches that use all markers (e.g. GBLUP), and not to variable selection models like BayesB.

Neither the two-step blending nor the current single-step analysis is suitable for routine use in the long-term as the number of genotyped individuals increases. A new approach is required.

### Materials and Methods

We have proposed an alternative formulation for this problem, that does not require matrix inversion, and is therefore practical for problems with hundreds of thousands of genotyped individuals. This formulation includes the conventional single-step model as a special case, and provides identical solutions when so parameterized. However, our method is easily extended to other classes of models, including those that use variable selection.

### Results and Discussion

The new method partitions the breeding value (BV) of a non-genotyped individual into a part that can be predicted from its genotyped relatives and an independent deviation that is correlated with such effects on other non-genotyped relatives. The part that can be predicted from its genotyped relatives is akin to using regression to impute the marker genotypes on the non-genotyped individuals, and using these as if observed, provided the independent deviations are also fitted.

We are now testing this approach with simulated and field data and are developing parallel computing to reduce processing times.