

Incorporating Evolutionary Constraints in Genomic Selection Improved Prediction Accuracies for Heterotic Traits in Maize

JINLIANG YANG¹, SOFIANE MEZMOUK^{1,2}, RITA MUMM³, AND JEFFREY ROSS-IBARRA¹

¹*Department of Plant Sciences, University of California, Davis, CA 95616, USA*

²*Current address: Institute of sciences, Somewhere, Street of Something*

³*Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

ABSTRACT

Genomic selection (GS) gains its popularity recently as it becomes more easier to access genome-wide markers in a training population. Current methods for GS weigh all the available SNPs equally for model training, without considering their functional differences. Genetic variations detected at evolutionary conserved sites tend to be deleterious and, therefore, would be more informative for GS. To build this kind of information as a prior into the GS model, we proposed a method to put more weight on evolutionarily constrained sites. As a proof of concept, a half diallel population founded from 12 diverse inbred lines was used, from which seven phenotypic traits were collected. After sequencing the 12 founder lines, ~14M SNPs were discovered and the SNPs were used to identify ~50K shared haplotype blocks using identity by descent (IBD). A five-fold cross-validation experiment was conducted using the summary statistics of the SNPs' conservation scores in the IBD blocks as explanatory variables. The results show that the prediction accuracies are significantly better than shuffled data with randomly assigned conservation scores. This study demonstrates the importance of incorporating evolutionary information in GS and it is potential use for plant breeding.