Incorporating evolutionary constraints in genomic selection improved prediction accuracies for heterosis in maize

Jinliang Yang[1], Sofiane Mezmouk[1,2], and Jeffrey Ross-Ibarra[1]

[1]*Department of Plant Sciences, University of California, Davis, CA 95616*
[2]*Institute of strange sciences, Somewhere, Street of Something*

ABSTRACT

Genomic selection (GS) gains its popularity recently as it become more easier to access genome-wide markers over a population. Current methods weight all of the available SNPs equally to train GS model. This normally leads to an over fitting problem especially when the number of markers are greatly larger than the number of individuals (p » n problem). Over fitted models are prone to have lower prediction accuracies and, therefore, would be less useful for breeding purpose. To overcome the over fitting problem, we proposed to weight more on evolutionary constraint sites for GS, where tend to be deleterious if a SNP was detected. As a proof of the concept, a half diallel population founded from 12 diverse inbred lines was constructed, from which seven phenotypic traits were collected. After sequencing of the 12 founder lines, 14M SNPs were discovered and the SNPs were used to identify 50K identity by descent (IBD) blocks. With the sum of the conservation scores in the IBD blocks as the explanatory variables in the diallel population, we show that the prediction accuracies are statistically significant better than random shuffled variables using a 5-fold cross-validation approach. The result demonstrates the importance of incorporating evolutionary information in GS and plant breeding.