Incorporation of Genomic Selection into the University of Illinois' Soft Red Winter Wheat Breeding Program

A.M. Krill[1], M.P. Arruda[1], P.J. Brown[1], A.E. Lipka[1], G. Brown-Guedira[2] and F.L. Kolb[1*]

[1]University of Illinois at Urbana-Champaign, Department of Crop Sciences, Urbana-IL.

[2]ARS-USDA, Eastern Regional Small Grains Genotyping Lab., Raleigh-NC

*Corresponding Author: (217) 333-4256, f-kolb@illinois.edu

Introduction

Genomic Selection (GS) and Marker Assisted Selection (MAS) are useful plant breeding strategies aimed at selecting superior individuals based on genetic data associated with traits of interest. GS uses high density genome wide markers to calculate a genomic estimated breeding value (GEBV) while MAS relies on previously identified QTL. Both strategies are useful for quantitative traits that are time consuming or difficult to measure. Here we compare MAS with GS and show how they are incorporated into the University of Illinois at Urbana-Champaign (UIUC) soft red winter wheat breeding program.

Objectives

1. Compare MAS and GS models for prediction of FHB resistance traits
2. Incorporate GS into the UIUC breeding program for selection of high yield, high test weight and FHB resistant cultivars.

Materials and Methods

Germplasm:

1. 273 diverse lines from, or in use at, the UIUCwheat breeding program.
2. 308 UIUC Preliminary Yield Trial (PYT) lines from 2015

Phenotyping:

1. Field evaluation in a scab nursery in Urbana-IL in 2011, 2013, 2014, and 2015. BLUPs calculated for: severity(SEV), incidence (INC), FHB index (SEV x INC), kernel quality (FDK), ISK index (ISK=0.3xINC+0.3xSEV+0.4xFDK), and deoxynivalenol concentration (DON).
2. Yield and Test Weight over 2 years (4 locations 2015, 2 locations 2014). FHB field evaluation 2015.

Genotyping-by-sequencing (GBS):

- Illumina HiSeq2000 (W.M. Keck Center for Comparative and Functional Genomics)
- Analyzed with TASSEL 4 GBS pipeline, Beagle 4 for imputation after filtering (*maf* > 5%, missing data per marker < 50%, and Fisher's exact test at 0.001 level).
- Pseudo-reference genome developed from *T. aestivum* Chinese spring chromosome survey (IWGSC, 2014)

Statistical models:

1. Prediction accuracy was obtained using a 4-fold cross validation scheme, 75 randomly selected training populations (TP) of size 205 and validation population (VP) of size 68. GS models were based on ridge regression-best linear unbiased predictor (rrBLUP) and MAS models were based on ordinary least square (OLS) regression..
2. Marker effects estimation, GEBVs and progeny simulation were obtained in the PopVar package in R using (rrBLUP) and a 4-fold cross validation. (Mohammadi et al., Crop Science, Vol. 55, 2015)

GBS

Costs: ~ $1,700 per Illumina lane, $150 per 96-plex for reagents *does not include labor

1. 2014 Association Mapping Population 2015
   - 3 two-enzyme combinations (*Pst1-HF-HinP1I, Pst1-HF-BfaI, Pst1-HF-Msp1*)
   - 1 lane x 96 lines/lane (3 plates) = 288 total **~$7.50/line**
   - ~20,000 SNPs
2. 2015 Preliminary Yield Trials (PYT)
   - 1 two-enzyme combination (*Pst1-HF-HinP1I*)
   - 1 lane x 384 lines (4 plates) = 384 total **~$6/line**
   - 6580 SNPs
3. 2016 PYT + Advanced Yield Trials (AYT) + parent lines (*in progress*)
   - 1 two-enzyme combination (*Pst1-HF-HinP1I*)
   - 1 lane x 576 lines (6 plates) = 576 total **~$4.50/line**
4. 2017 $F_{3:5}$ Single Plots (*future work*)
   - 1 two-enzyme combination (*Pst1-HF-HinP1I*)
   - 3 lane x 768 lines (8 plates) = 2304 total **~$3.75/line**

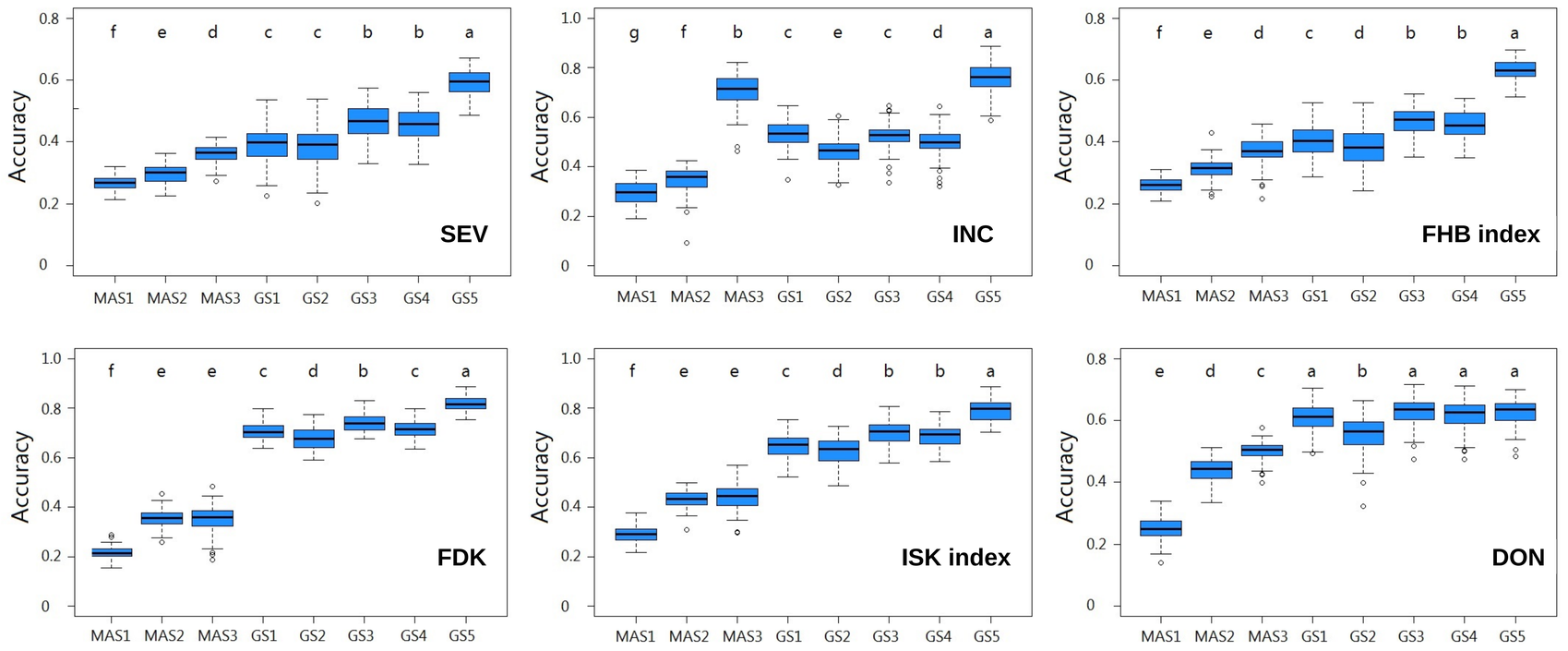

Figure 1. Prediction accuracies for six traits corresponding to Fusarium head blight (FHB) resistance: Severity (**SEV**), Incidence (**INC**), **FHB index**, Fusarium damaged kernels (**FDK**), **ISK index**, and deoxynivalenol concentration (**DON**). **MAS1**: Marker assisted selection with *Fhb-1* using linear regression. **MAS2**: Marker assisted selection with "*independent*" QTL using multiple linear regression. **MAS3**: Marker assisted selection with "*in house*" QTL using multiple linear regression. **GS1**: genomic selection with 19,992 SNPs. **GS2**: genomic selection with 19,992 SNPs + 5 randomly selected SNPs treated as fixed effects. **GS3**: genomic selection with 19,992 SNPs + *Fhb-1* treated as fixed effect. **GS4**: genomic selection with 19,992 SNPs + "*independent*" QTL treated as fixed effects. **GS5**: genomic selection with 19,992 SNPs + "*in house*" QTL treated as fixed effects. Treatments with the same letter are not significantly different at α = 0.05 level, according to the Ryan-Einot-Gabriel-Welch Q (REGWQ) range test

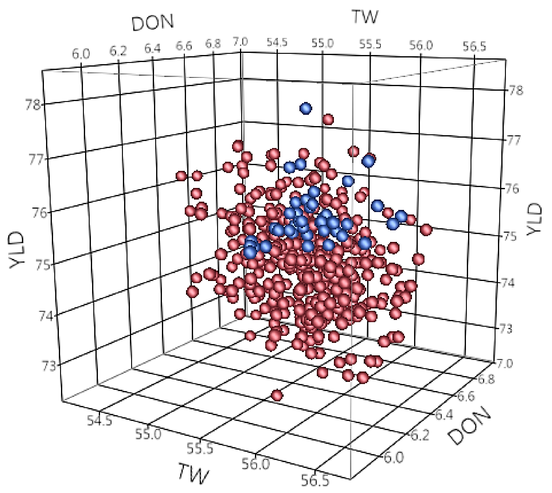

Figure 2. Genomic estimated breeding values (GEBVs) for 453 elite lines (PYT 2015 + UIUC breeding lines from the association mapping population. The 45 lines in blue indicate those in the top 30% for Yield (YLD), top 65 % for test weight (TW) and lowest 65 % for Deoxynavalanol concentration (DON). These lines were used as parents for PopVar Progeny simulations.

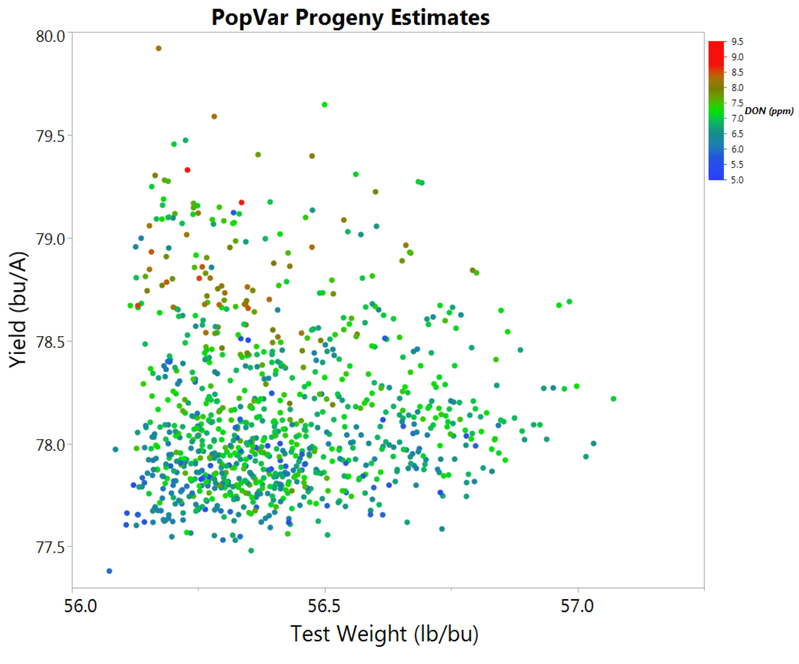

Figure 3. Predicted mean of the best 5% progeny from bi-parental crosses among the 45 lines chosen for high yield, test weight and DON (Figure 2) using PopVar package in R.

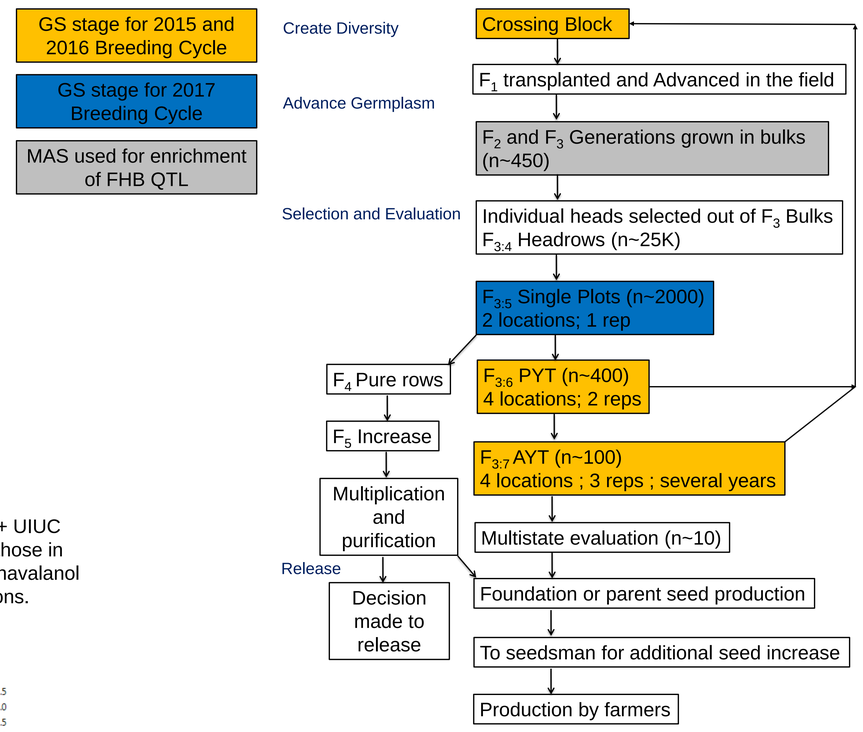

Figure 4. Breeding scheme for UIUC soft red winter wheat program indicating areas where MAS or GS are used or being incorporated. MAS has historically been used in F_1 or F_2 populations for enrichment of *Fhb1*. This is the first year GS has been incorporated in the breeding program. A combination of GEBVs and phenotype data was used in selection of lines to advance from the PYTs. GEBV from PYT and the AYT lines were used to simulate progeny and select parent lines for the crossing block.

Conclusion

GS models outperformed MAS models for prediction of traits related to FHB resistance. Treating significant QTL as fixed effects resulted in higher accuracies. All GS models predicted similar ranking of lines. They did not in the MAS models. The use of GS in the breeding program should provide higher response to selection in early stages where limited phenotype data are available. This is especially useful for prediction of data such as DON, where high quality accurate data are not immediately available.

Acknowledgments

This material is based upon work supported by the U.S. Department of Agriculture, under Agreement No. 59-0206-4-029. This is a cooperative project with the U.S. Wheat & Barley Scab Initiative.