Genomic Prediction for Fusarium Head Blight Resistance: the Experience of the University of Illinois Wheat Breeding Program

M.P. Arruda[1], P.J. Brown[1], A.M. Krill[1], A.E. Lipka[1], G. Brown-Guedira[2], C. Thurber[3], and F.L. Kolb[1*]

[1]University of Illinois at Urbana Champaign, Depart Crop Sciences, Urbana-IL. [2]ARS-USDA, Eastern Regional Small Grains Genotyping Lab., Raleigh-NC. [3]Abraham Baldwin Agricultural College, School of Science and Mathematics, Tifton-GA. Corresponding Author: f-kolb@uiuc.edu .

Introduction and objective

Breeding for Fusarium Head Blight (FHB) resistance is challenging due to its quantitative inheritance and difficulties in obtaining high quality phenotypes. At the University of Illinois soft red winter wheat breeding program, genomic selection (GS) models have been tested for predicting multiple parameters associated with FHB resistance. Then, the objectives of this study were: i) to assess the effect of genotypic imputation methods, statistical models, marker density, and training population size on GS accuracy; ii) to compare GS and marker-assisted selection (MAS) in their ability to predict phenotypes associated with FHB resistance

Materials and Methods

➢ ***Study 1:***

Germplasm: 273 lines from, or in use at the University of Illinois.

Phenotyping: Field evaluation in a scab nursery in Urbana-IL in 2011, 2013, and 2014. BLUPs calculated for: severity (SEV), incidence (INC), FHB index [(FHBdx=(SEV x INC)/100), Fusarium-damaged kernel (FDK), incidence-severity-kernel index (ISK=0.3xINC+0.3xSEV+0.4xFDK), and deoxinivalenol (DON).

Genotyping: Genotyping-by-sequencing (GBS) used with three two-enzyme combinations: *PstI-MspI*, *PstI-HinPI*, and *PstI-BfaI*. Sequence data obtained from Illumina HiSeq2000, and then analyzed with UNEAK (*maf* = 5%, missing data per marker ≤ 20%)

Imputation methods: mean imputation (MNI), singular value decomposition (SDVI), random forest regression (RFI), and expectation maximization (EMI).

Genetic structure: Assessed through principal component analysis (PCAs) using all SNPs in JMP Genomics.

Statistical models: ridge regression best linear unbiased predictor – RR-BLUP, least absolute shrinkage and operator selector – LASSO, and elastic-net. The R packages "rr-BLUP" and "glmnet" were used.

Marker density: random samples (*p* = 500, 1500, 3000, and 4500 SNPs) obtained from a total of 5054 SNPs.

Training population size (n_{TP}): random samples of n_{TP}=96, 144, 192, and 218 were obtained from the 273 lines.

Accuracy: Calculated as $r(GEBV:PEBV)/\sqrt{h^2}$, where r = Pearson's correlation between genomic and phenotypically breeding values (GEBVs and PEBVs); h^2 = broad-sense heritability.

Mean comparison: Each analysis was repeated 300 times and the mean accuracies were compared using the Ryan-Einot-Gabriel-Welch multiple comparison test at 0.05 level with SAS PROC GLM.

➢ ***Study 2:***

Germplasm and phenotypes: same as in study 1.

Genotyping: Sequence data from study 1 was analyzed with TASSEL GBS 4 (*maf* = 5%, missing data per marker ≤ 50%).

QTL detection: Marker-trait associations tested in GAPIT using a mixed model controlling for relatedness and population structure.

Comparing GS and MAS: 5 models were compared: GS1 = genomic selection with all SNPs treated as random effects; GS2 = all SNPs as random effects + *Fhb1* treated as fixed effect; GS3 = all SNPs as random effects + QTL as fixed effect (Table 1); MAS1 = marker-assisted selection with *Fhb-1* as fixed effect; MAS2 = marker-assisted selection with QTL as fixed effect. Marker effects were estimated with RR-BLUP for GS models, and with multiple linear regression for MAS models.

Accuracy and mean comparison: same as in study 1.

Selection differential: For each trait, the 273 breeding lines were ranked based on their genomic estimated breeding values (GEBVs), obtained from the 5 models tested. Then, the selection differential was calculated as the difference between the mean PEBV of the best 5, 10, 15, 20, and 25% lines and the mean of the reference group (all lines).

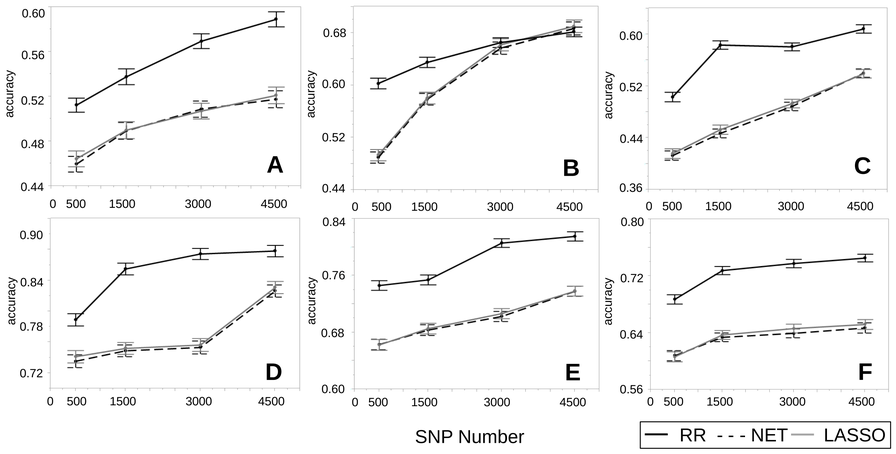

Figure 1. Five-fold cross validated genomic selection accuracies for six FHB-related parameters as a function of genomic selection models and SNP numbers. A = SEV (severity), B = INC (incidence), C = FHBdx (FHB index), D = FDK (Fusarium-damaged kernels), E = ISK (ISK index), F = DON (deoxynivalenol concentration). RR = ridge regression-best unbiased linear predictor, NET = elastic-net, LASSO = least absolute shrinkage and selection operator. Error bars represent ± one standard error of the mean.

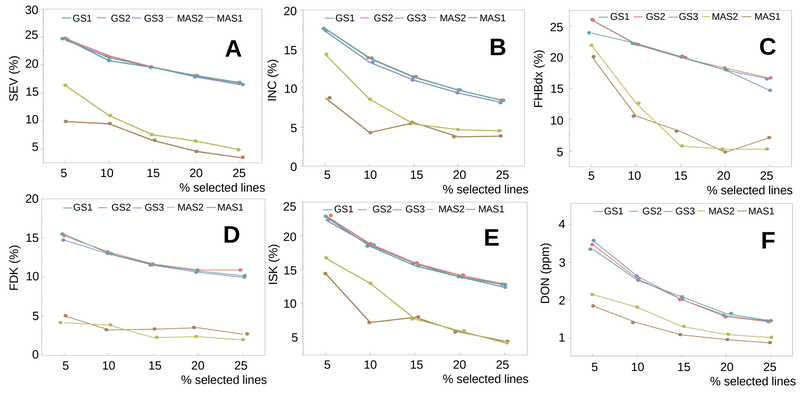

Figure 5. Selection differential for six parameters associated with Fusarium head blight (FHB) resistance: severity (A), incidence (B), FHB index (C), Fusarium-damaged kernel (D), incidence-severity-kernel index (E), and deoxynivalenol concentration (F). The numbers on the x-axis represent the percentage of selected individuals after ranking 273 breeding lines according to breeding values estimated from five different models: MAS1: Marker assisted selection with *Fhb-1* using linear regression, MAS2: Marker assisted selection with QTL using multiple linear regression, GS1: genomic selection with 19,992 SNPs using ridge regression-best linear unbiased predictor (RR-BLUP) model, GS2: genomic selection with 19,992 SNPs + *Fhb-1* using RR-BLUP, GS3: genomic selection with 19,992 SNPs + QTL using RR-BLUP.

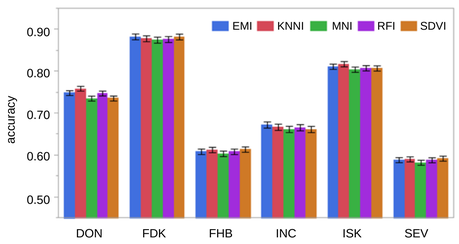

Figure 2. Five-fold cross validated genomic selection accuracies for six parameters associated with FHB resistance and four imputation methods. Methods with the same letter do not differ according to the Ryan-Einot-Gabriel-Welch comparison test at 0.05 level. Error bars represent ± one standard error of the mean.

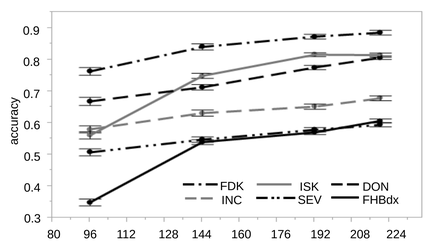

Figure 3. The effect of training population (TP) size on genomic selection accuracy for six FHB related traits. The analysis were performed using RR-BLUP and 5054 SNPs, imputed by the EMI method. Error bars represent ± one standard error of the mean.

Table 1. SNPs associated with FHB resistance in a panel of 273 breeding lines.

Trait	SNP	C	cM	*p*[a]	r²	Adj p[b]	effects
SEV	WCSS1_contig10676713_3B_7175	3B	18.32	5.14	0.08	0.050	-9.54
	WCSS1_contig10352272_3B_5482[c]	3B	10.19	3.83	0.05	0.980	-7.21
	WCSS1_contig10698462_3B_2332[c]	3B	6.86	3.6	0.04	0.980	6.02
	WCSS1_contig10699215_3B_3620[c]	3B	18.32	3.37	0.04	0.980	-4.99
INC	WCSS1_contig3876750_7DS_2023	7D	70.84	11.57	0.16	<0.001	6.74
	WCSS1_contig5780077_6AL_12152	6A	134.15	4.94	0.06	0.097	3.60
	WCSS1_contig2300354_4DS_4482	4D	0	4.72	0.06	0.097	-4.56
	WCSS1_contig7146617_4AL_11335	4A	78.35	3.54	0.06	0.070	-3.02
	WCSS1_contig4132011_7AS_1400	7A	22.82	4.54	0.05	0.097	-2.62
FHBdx	WCSS1_contig10676713_3B_7175	3B	18.32	5.14	0.07	0.052	-8.96
FDK	WCSS1_contig3314747_1AS_2298[c]	1A	27.24	4.55	0.04	0.563	4.84
	WCSS1_contig4427089_3AL_5739[c]	3A	131.87	4.05	0.04	0.569	-4.96
	WCSS1_contig5938251_4AS_8590[c]	4A	0	3.92	0.04	0.569	-5.44
	WCSS1_contig3109240_7BS_581[c]	7B	77.11	3.72	0.03	0.569	-5.31
	WCSS1_contig7110688_4AL_882[c]	4A	78.35	3.70	0.03	0.569	4.45
ISK	WCSS1_contig10676713_3B_7175	3B	18.32	5.14	0.07	0.052	-5.55
	WCSS1_contig3876750_7DS_2023	7D	70.84	11.57	0.16	0.000	7.47
DON	WCSS1_contig10413672_3B_4839	3B	73.67	5.14	0.07	0.052	1.75
	WCSS1_contig10676713_3B_7175	3B	18.32	5.14	0.07	0.052	-1.36
	WCSS1_contig1879930_1DS_3352	1D	19.04	5.10	0.06	0.052	-1.99
	WCSS1_contig10764618_3B_2168	3B	73.67	4.99	0.06	0.052	2.08

[a] *p* value reported in a $-\log_{10}$ scale; [b] Adj p = FDR-adjusted *p* value; [c] SNP not significant at FDR-adjusted *p* value = 0.10. Marker-trait associations were tested using a compressed mixed linear model with control for population structure and relatedness. A total of 19,992 SNPs were used in the analysis.

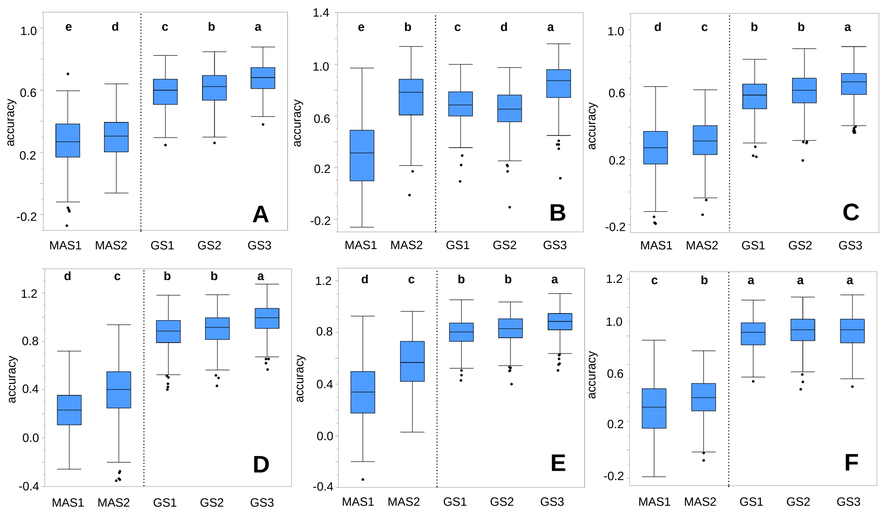

Figure 4. Five-fold cross validated prediction accuracies for six parameters associated with Fusarium head blight (FHB) resistance: severity (A), incidence (B), FHB index (C), Fusarium-damaged kernel (D), incidence-severity-kernel index (E), and deoxynivalenol concentration (F). The dotted line divides marker-assisted selection (left) and genomic selection (right). MAS1: Marker-assisted selection with *Fhb-1* using linear regression. MAS2: Marker-assisted selection with QTL using multiple linear regression. GS1: genomic selection with 19,992 SNPs using ridge regression-best linear unbiased predictor (RR-BLUP) model. GS2: genomic selection with 19,992 SNPs + *Fhb-1* using RR-BLUP. GS3: genomic selection with 19,992 SNPs + QTL using RR-BLUP. All marker estimates were obtained using a 5-fold cross validation scheme, 60 randomly selected training populations (TP) of size 218 and validation population (VP) of size 55. Treatments with the same letter at not significantly different according to the Ryan-Einot-Gabriel-Welch-q test at α = 0.05 level.

Results

➢ ***Study 1:***

- ✓ Moderate to high accuracies were obtained for the traits evaluated in this study. The lowest accuracies were observed for SEV, and the highest values were observed for FDK.
- ✓ RR-BLUP outperformed the LASSO and ELASTIC-NET for all traits.
- ✓ Marker density had a significant effect on prediction accuracies, with diminishing increments after 1500 SNPs.
- ✓ The imputation methods performed equally well in terms of GS accuracy, with a numerical advantage for EMI in two out of six traits.
- ✓ A significant increase on accuracies was observed as the training population size increased. A plateau was observed for all traits when n_{TP}>192, except for ISK.

➢ ***Study 2:***

- ✓ GS provided higher prediction accuracies than MAS in nearly all situations. In addition, the accuracies were more variable for MAS than for GS models.
- ✓ For INC, MAS2 outperformed GS1 and GS2, but not GS3.
- ✓ Treating QTL as fixed effects resulted in higher accuracy for all traits except DON.
- ✓ Treating only *Fhb-1* as fixed effect did not improve prediction accuracies, except for SEV.
- ✓ GS provided higher selection differentials than MAS in all situations; however, the differences in accuracy among the GS models were not sufficiently higher to substantially change the order of the breeding lines.

Conclusions

These results of these studies showed that a complex trait such as FHB resistance can be predicted with moderate to high precision with GS. Prediction accuracy was impacted by different parameters, which should be analyzed before performing GS. In addition, GS outperformed MAS in nearly all situations, and treating QTL as fixed effect resulted in even higher accuracies. At the moment, GS is being used as a complementary tool at the University of Illinois wheat breeding program, mainly for parental selection and academic training purposes. With the reducing genotyping costs, GS could be fully implemented in our breeding pipeline.

Acknowledgments

This material is based upon work supported by the U.S. Department of Agriculture, under Agreement No. 59-0206-9-057. This is a cooperative project with the U.S. Wheat & Barley Scab Initiative. The first author gratefully acknowledges Monsanto for financial support (Monsanto Fellowship).