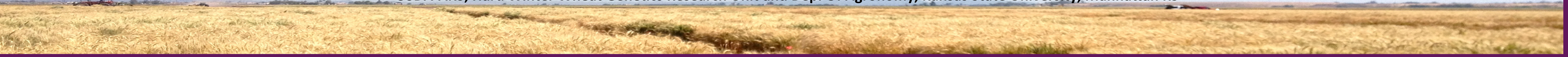


Genomic Selection for Pre-Harvest Sprouting Tolerance in Winter Wheat

Jessica Cooper¹, Eric Storlie¹, Jesse Poland² and Scott Haley¹

¹Soil and Crop Sciences Department, Colorado State University, Fort Collins, CO

²USDA-ARS, Hard Winter Wheat Genetics Research Unit and Dep. Of Agronomy, Kansas State University, Manhattan KS



Introduction



Photo credit: Thomas Lumpkin/CIMMYT

- ◆ **Pre-harvest sprouting (PHS)**
 - premature germination
 - reduces quality
 - lowers grain sale price
- ◆ **Challenge in breeding for PHS tolerance (PHST)**
 - quantitatively inherited
 - very tedious and expensive to phenotype
- ◆ **Genomic Selection (GS)**
 - predict phenotypes using genotypes
 - useful for predicting traits that are costly and time-consuming to assess

OBJECTIVE

- ◆ Develop predictive models for PHST and apply models to CSU wheat-breeding program

Materials & Methods

- ◆ Training population of 399 lines
 - two replications planted at Fort Collins (2012)
 - imbalanced historical data from 2006-2012
- ◆ Validation set of 700 lines
 - harvested in Fort Collins (2013)

GENOTYPING

- ◆ Genotyping-by-sequencing
 - two-enzyme approach
 - 45,757 SNP markers



PHENOTYPING

- ◆ Collected spikes at physiological maturity
- ◆ Threshed samples placed in Percival incubator with sufficient moisture
 - 20°C for 12 hour days, 7 days
- ◆ Germination index (GI) calculated

$$GI = \frac{(7x_{n1} + 6x_{n2} + 5x_{n3} + 4x_{n4} + 3x_{n5} + 2x_{n6} + 1x_{n7})}{\text{total days} * \text{total grains}}$$

where n1,n2...n7 = number germinated grains on day1, day2, ...day7



Applied Genomic Selection in Wheat Program

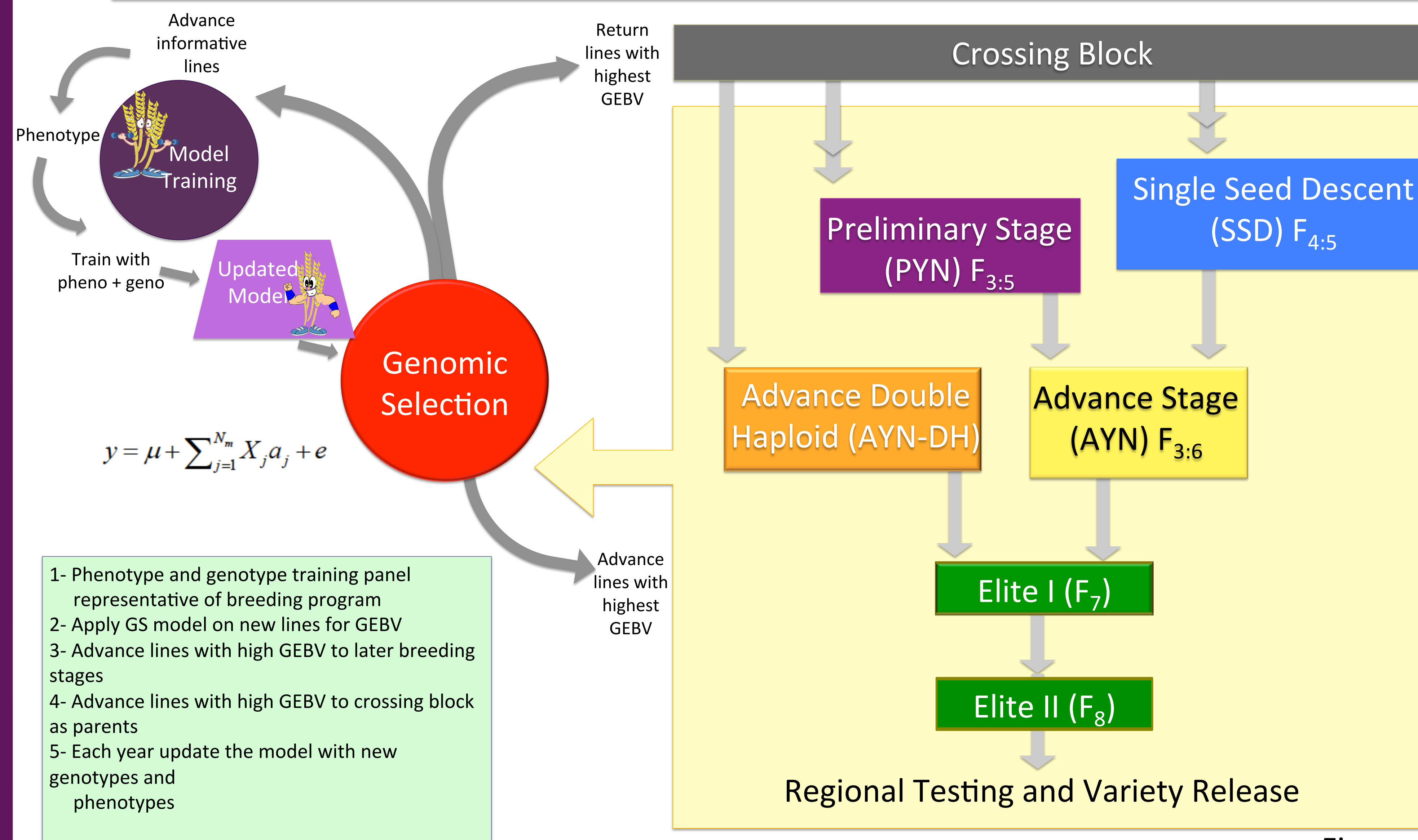


Figure 1

Procedures

- ◆ 5-fold cross validation was used to determine accuracy of different GS models and Bayesian Ridge Regression (5,000 iter, 1,000 burn) was selected as the best model due to high prediction accuracy and low computing time.
- ◆ 2013 breeding lines were divided into 5 validation sets based on breeding schemes (PYN, AYN, AYN-DH, Elite; Fig. 1).
- ◆ The 399 lines (2006-2012) were used as the training panel to predict 2013 breeding lines (Fig. 2A).
- ◆ The “updated model” included all available data (2006-2013) to train the model, excluding the validation set used for prediction accuracy (Fig. 2B).
- ◆ Principal component analysis (PCA) of marker data was used to estimate population structure of all lines (Fig. 3).

Discussion

- ◆ GS models were *most accurate* at predicting *Elite* lines and *least accurate* at predicting *Doubled Haploids*.
- ◆ PCA showed Elite lines to be most similar to the 399 training panel which may explain the high accuracy as the trained model was representative of the breeding material.
- ◆ Low accuracy of Doubled Haploid lines may be due to small phenotypic and genotypic *variation*.
 - Few parent lines were used to create Doubled Haploids leading to low genetic diversity.
 - Most phenotypes scored very low for GI.
- ◆ The updated model increased accuracy for early stage lines (SSD) but decreased accuracy for Elite lines. More analysis will be done to determine the best method for selecting lines to update models.

Results

Figure 2

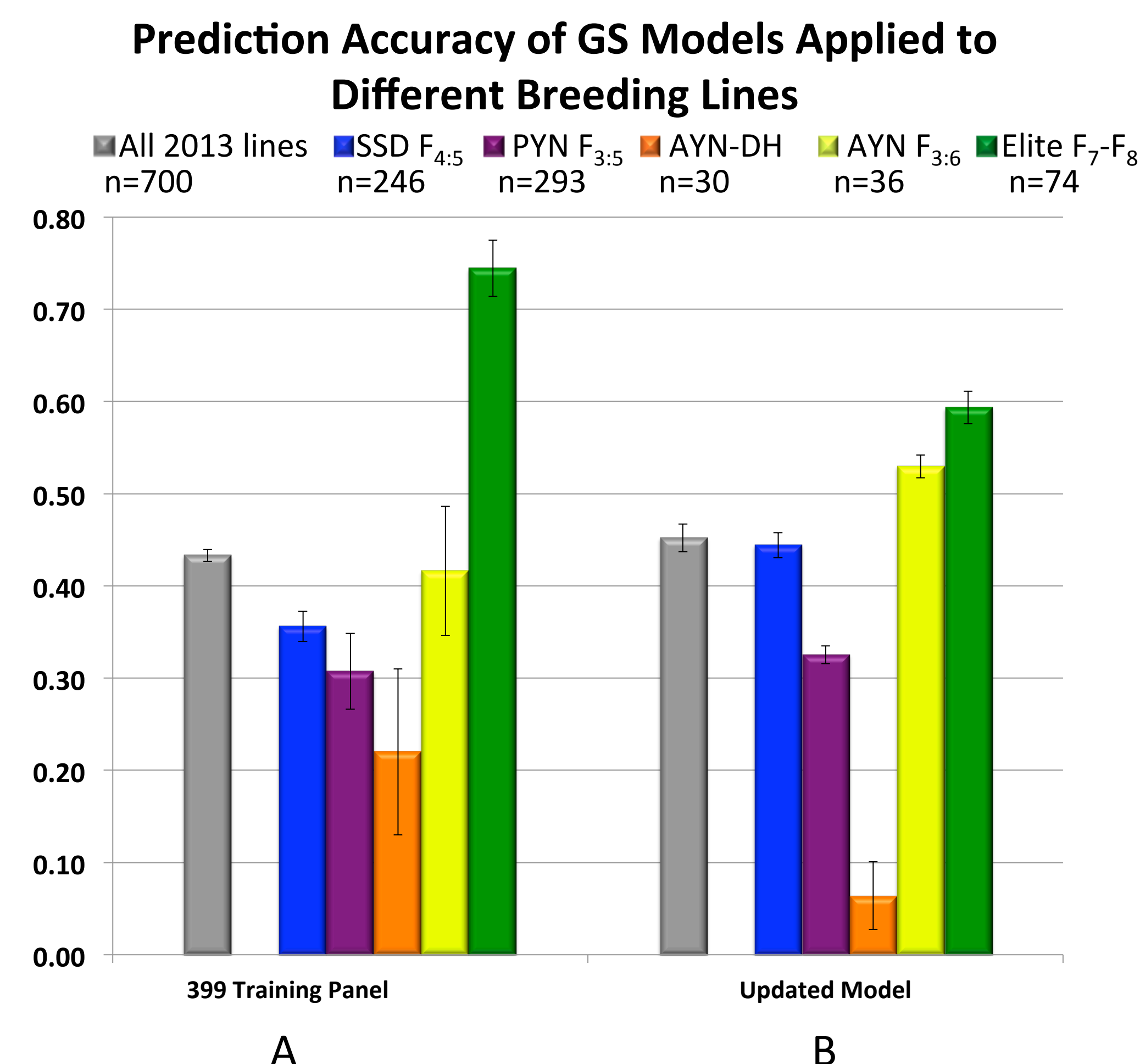
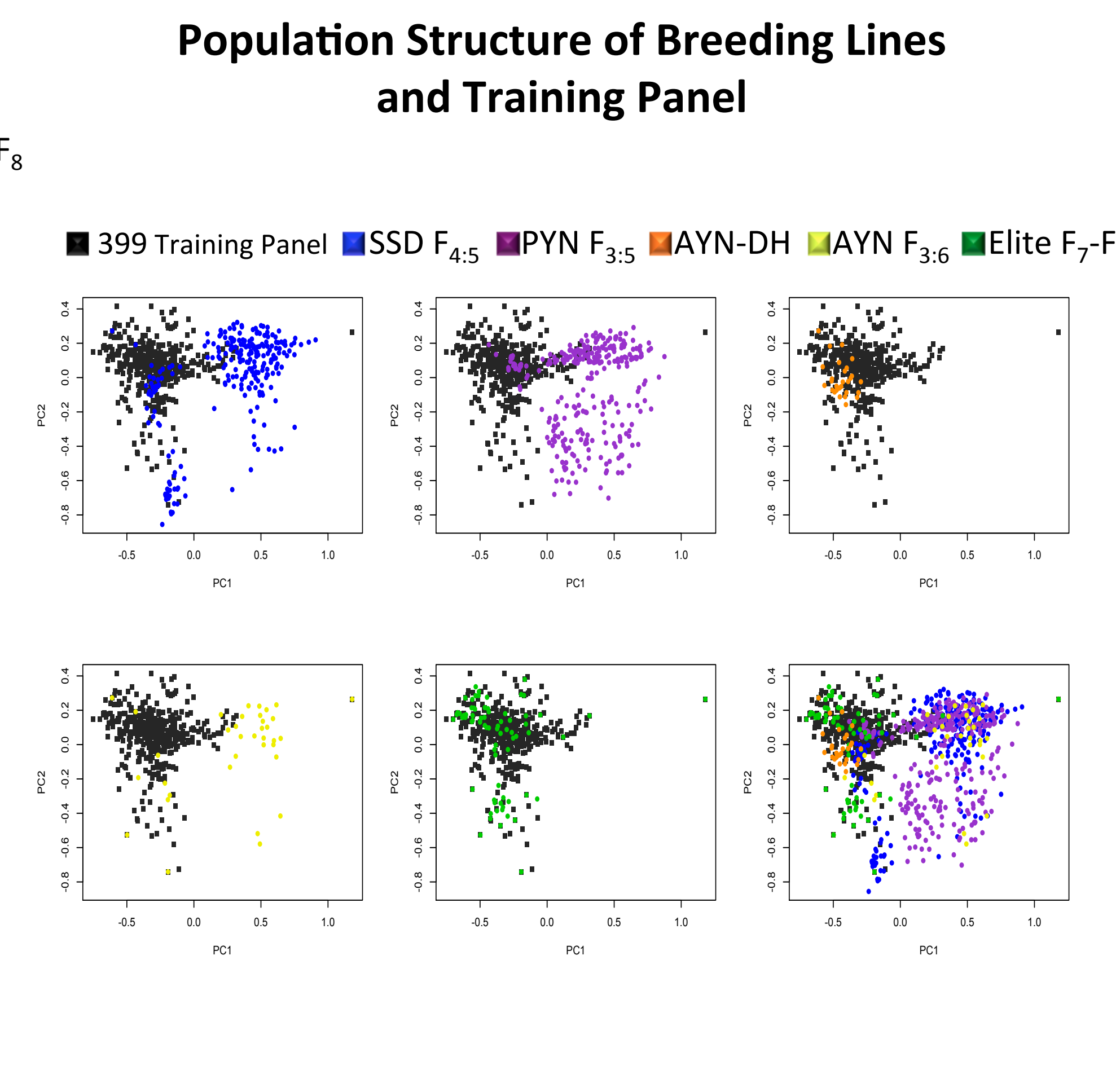


Figure 3



Summary

A training population of wheat lines from 2006 – 2012 was used to develop predictive models for PHST. New lines from 2013 were used as validation sets to determine prediction accuracy of models applied to different breeding stages in a wheat breeding program. Accuracies were similar to findings in the literature ($r^2 = 0.3 - 0.7$), supporting the use of GS to improve pre-harvest sprouting tolerance in wheat.