THE OHIO STATE UNIVERSITY

Utilizing Genomic Selection to Accelerate the Pace of Developing FHB Resistant Varieties

Antonio Cabrera

12/6/2015

Rationale

- Genomic Selection (GS) a new approach for marker-assisted selection.

  - All markers are used in prediction model.
  - Breeding values are estimated from marker information (GEBV).
  - Prediction takes advantage of massive marker data.
  - Cost efficiency is possible (time and money).
  - Shortens breeding cycles: Could improve gain per year.

THE OHIO STATE UNIVERSITY

THE OHIO STATE UNIVERSITY

Prediction accuracy (r)

Phenotype: (Y_{TP})

GS-Models:
RR-BLUP
BL
EN
RF

Estimation of Marker effects (α)

Genotype (high throughput) X_{TP}

Selection Candidates

X_{SC}*α =GEBV

Selected individuals based on GEBV

3

THE OHIO STATE UNIVERSITY

Objectives

Evaluate the prediction accuracy of GS for FHB resistance in a regionally adapted soft winter wheat population.

- Impact of relatedness on accuracy.
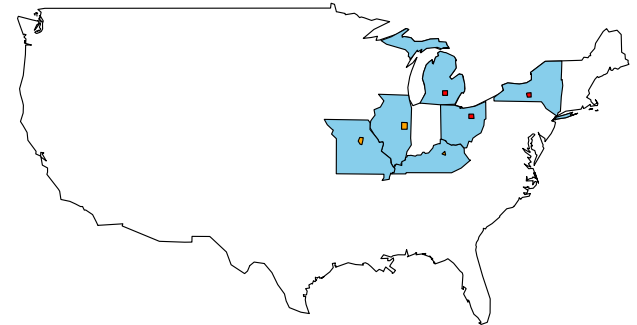- Evaluate the accuracy of several optimization model approaches.

4

Materials and Methods

640 lines:
~100 RILs from each of six programs, tested locally.
   48 checks evaluated in each environment.

Traits: Incidence **(INC)**, Severity **(SEV)**, Index **(IND)**, Fusarium damaged Kernel **(FDK)**, 0.3INC+0.3SEV+0.4FDK **(ISK),** Deoxynivalenol concentration **(DON)**, Principal component phenotype index **(PC1)**.
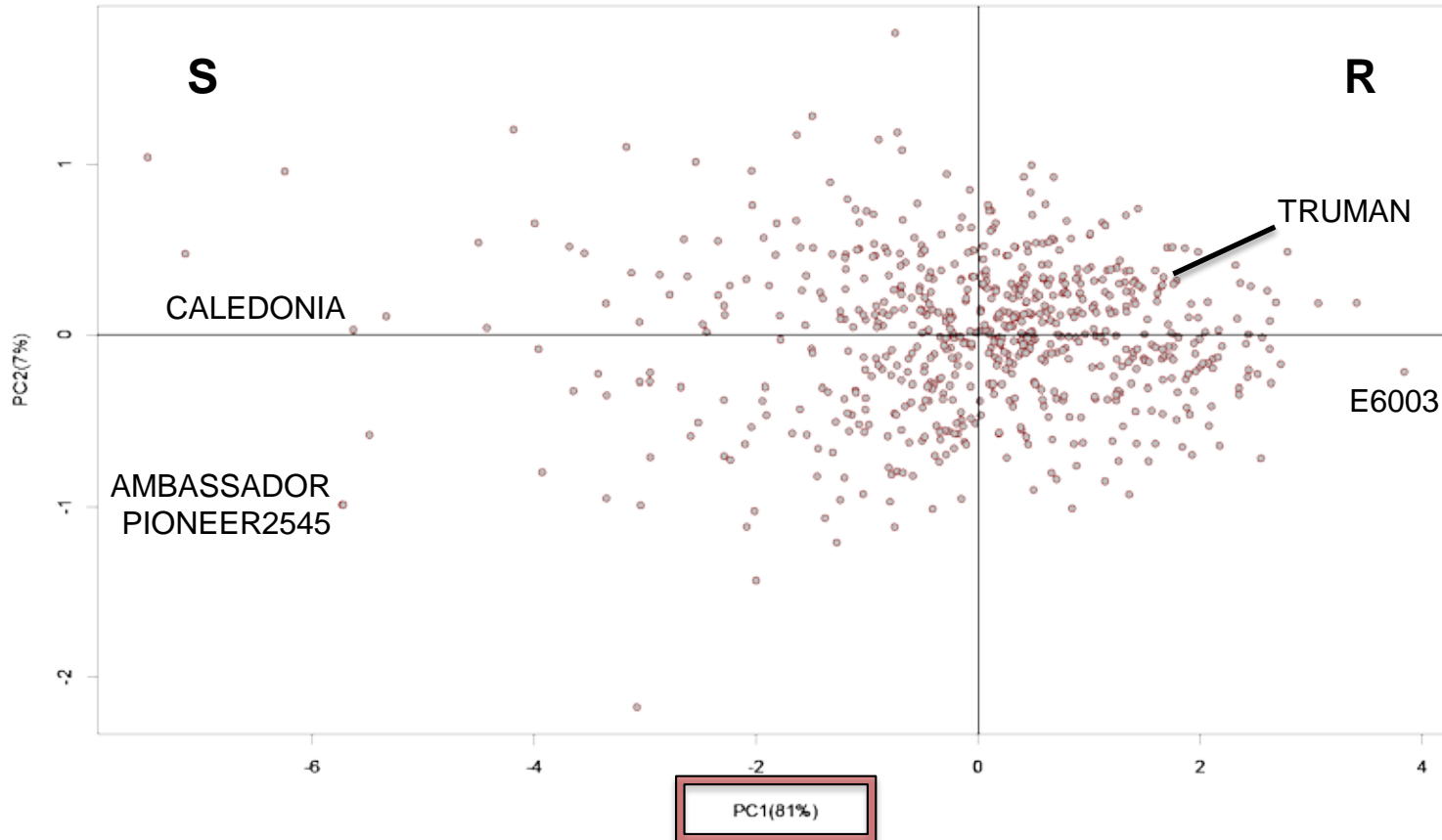
- 1. Checks were used to adjust the RIL data within each location and Heading date (HD) was used as covariate.

- 2. Best linear unbiased predictors of phenotypes (BLUPs) were generated from adjusted phenotypes fit in a mixed model.

Genotyping: GBS markers: 4,643 SNPs.

Phenotypic analysis
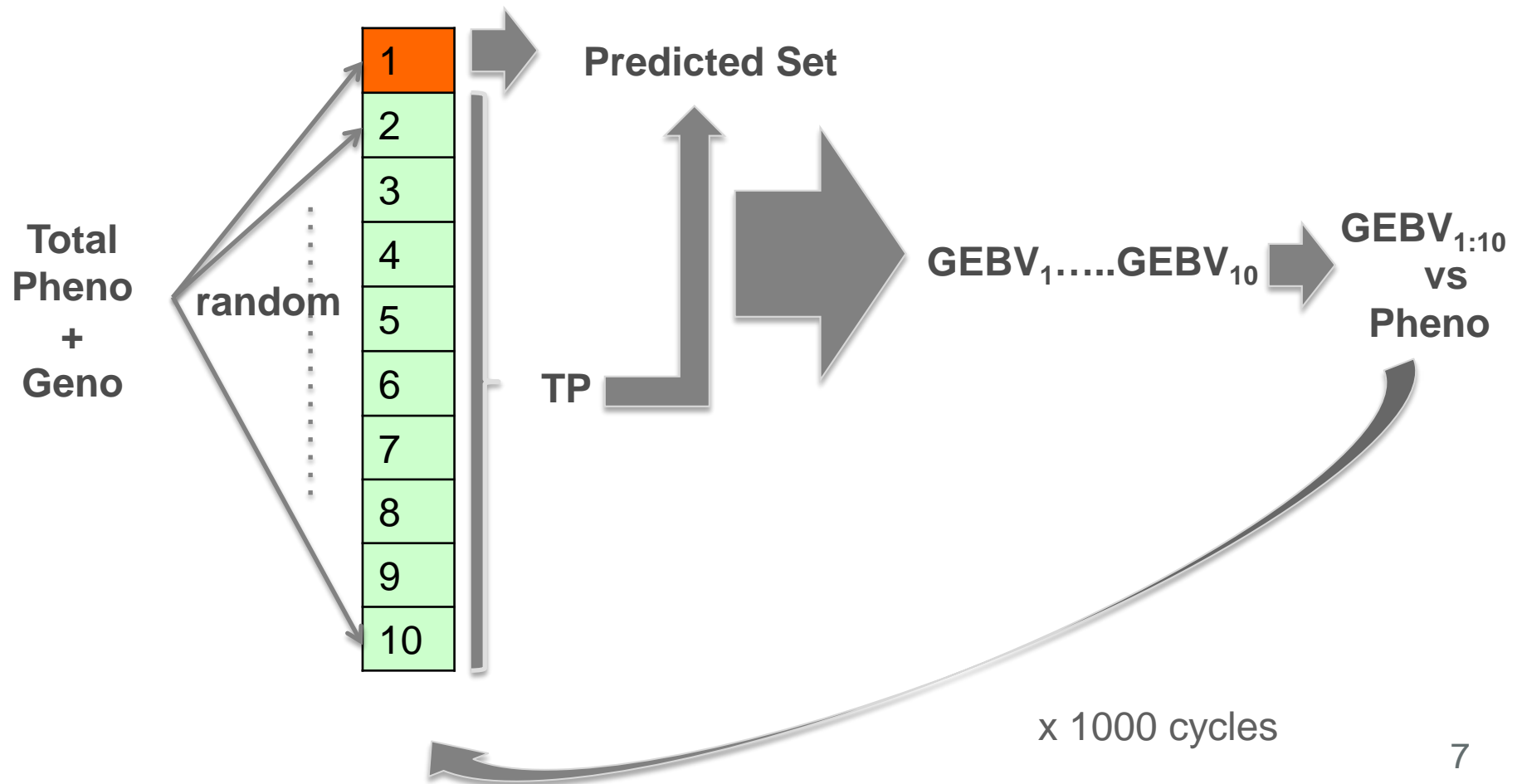
Principal component analysis across all traits

For all traits
$h^2 > 0.8$

Traits were
Correlated:
(from 0.6 to 0.9)

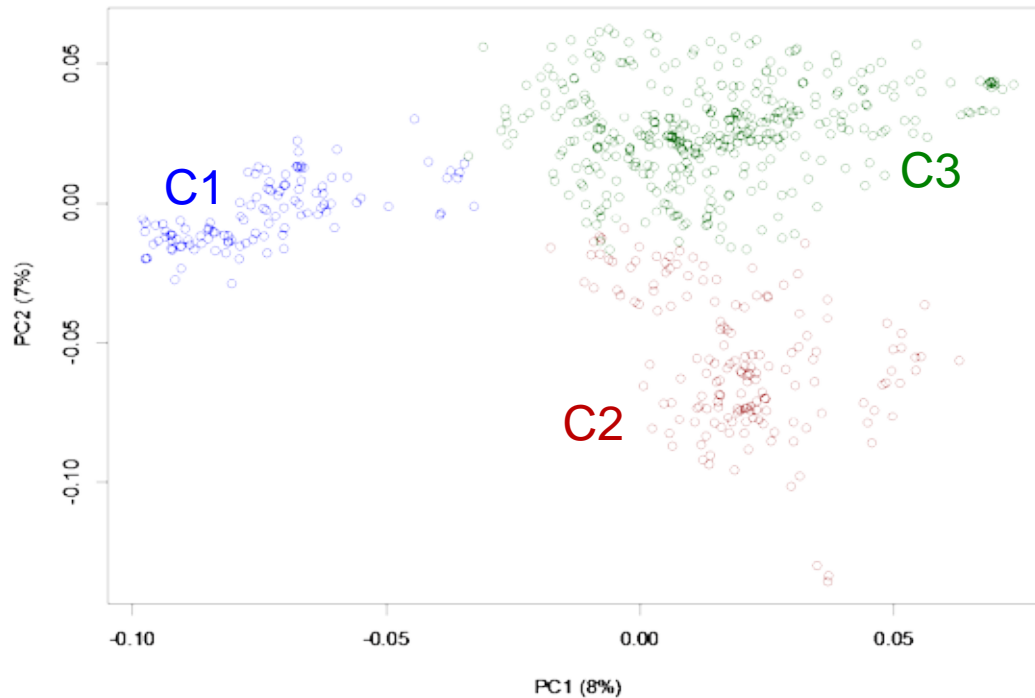# Assessing Model Performance

- **10 fold Cross validation**

Total Pheno + Geno **random**

| 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| 6 |
| 7 |
| 8 |
| 9 |
| 10 |

Predicted Set

TP

$GEBV_1.....GEBV_{10}$

$GEBV_{1:10}$ **vs Pheno**

x 1000 cycles

7

Prediction Accuracy (r)

10 fold Cross validation

Pearson's correlation (r of GEBV with BLUP)

#of genotypes	INC	SEV	IND	FDK	ISK	DON	PC1
640	0.45	0.57	0.49	0.45	0.52	0.49	0.53

Clustering of lines using marker data

Within cluster prediction accuracy

Cluster (#genotypes)	INC	SEV	IND	FDK	ISK	DON	PC1
All (n=640)	0.45	0.57	0.49	0.45	0.52	0.49	0.53
C1 (n=113)	0.46	0.61	0.49	0.38	0.49	0.47	0.52
C2 (n=161)	0.26	0.53	0.45	0.48	0.52	0.51	0.53
C3 (n=366)	0.46	0.44	0.42	0.43	0.49	0.41	0.48

Accuracy was similar despite variation in n

C1 accuracy : Small n but high relatedness
C2 accuracy: Moderate. n, mod. relatedness
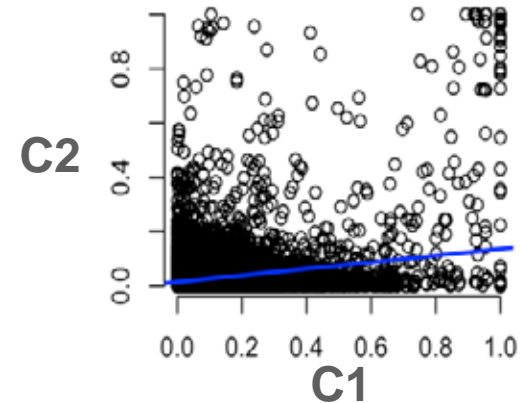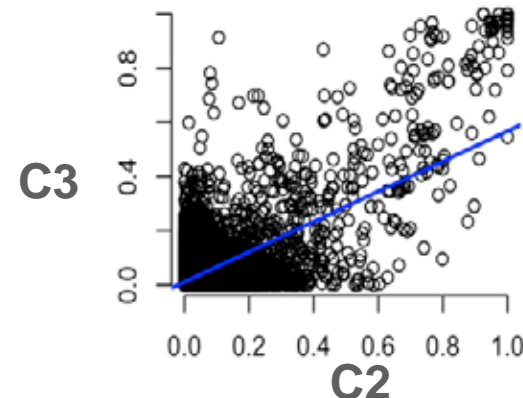C3 accuracy: Large n, but less related

Kinship
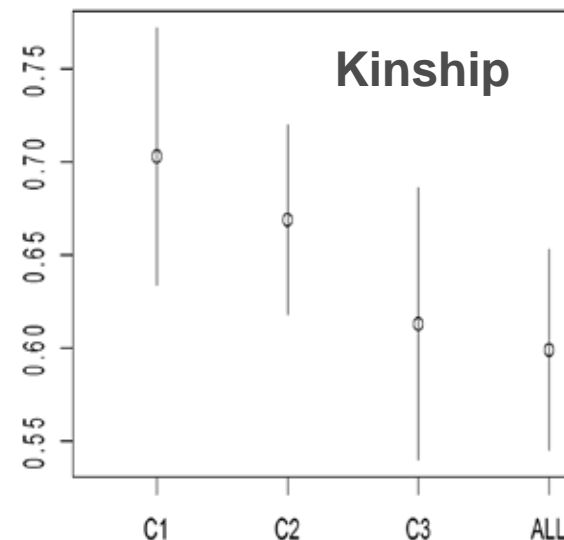
Can model trained with data from one cluster predict phenotype of another cluster?

Training Pop	Predicted Pop	Accuracy PC1
C1	C2	0.01
C1	C3	0.02
C2	C1	-0.09
C3	C1	-0.05
C2	C3	0.07
C3	C2	0.36

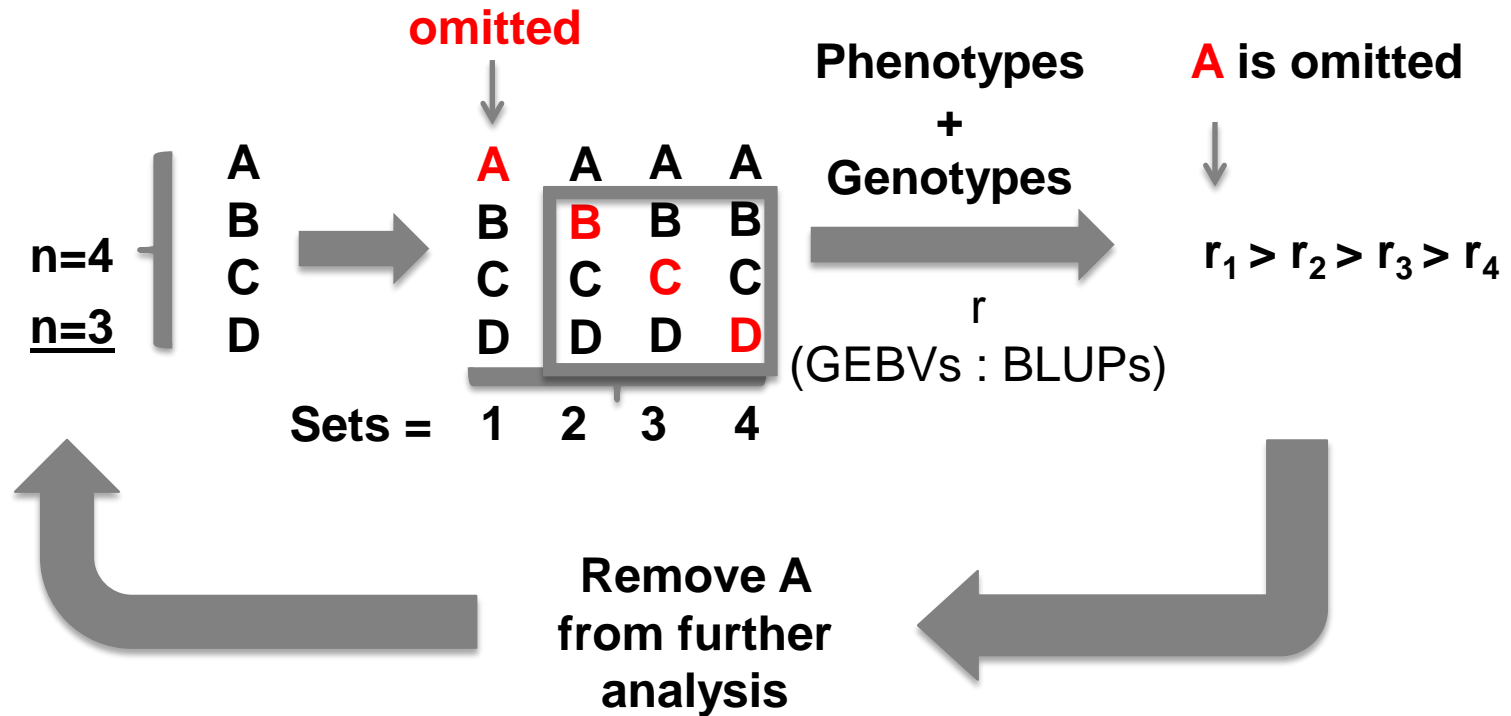**Correlation of pairwise LD r2 values**

LD in C1 ≠ LD in C2

LD in C3 ≈ LD in C2

Prediction accuracy is low:

- C3 predicts C2 while C2 does not predict C3
- Same significant LD relationship however C3 presents a larger TP than does C2.
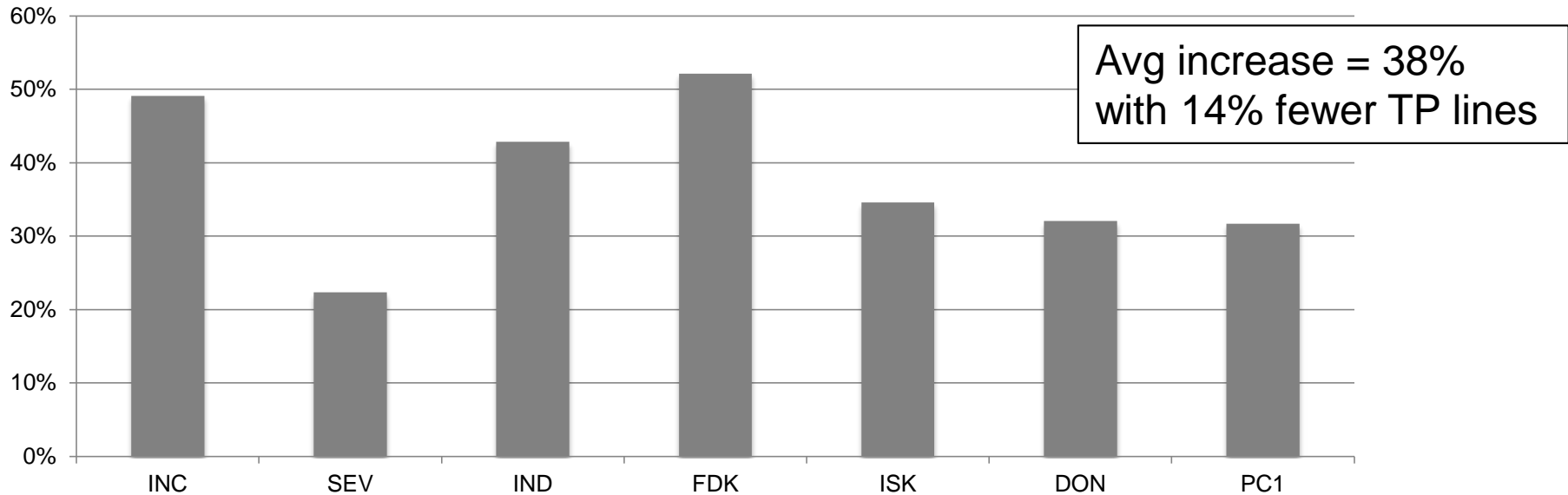
11

Elimination of less predictive individuals

omitted

$n=4$

$\underline{n=3}$

A
B
C
D

A A A A
B B B B
C C C C
D D D D

Sets = 1 2 3 4

r
(GEBVs : BLUPs)

**Phenotypes
+
Genotypes**

A is omitted

$r_1 > r_2 > r_3 > r_4$

**Remove A
from further
analysis**

n=640

Elimination until no further improvement is obtained

12

% increase in accuracy after optimization

Avg increase = 38% with 14% fewer TP lines

Improvement in predictive ability after elimination of less predicted individuals

	INC	SEV	IND	FDK	ISK	DON	PC1
Before	0.45	0.57	0.49	0.45	0.52	0.49	0.53
After*	0.67	0.70	0.70	0.68	0.70	0.65	0.70
% increase	49%	22%	43%	52%	35%	32%	32%
n after elimination	483	489	583	574	576	589	581

Relative Efficiency of GS for FHB Resistance

$$RE_{cycle} = \frac{r}{\sqrt{H}}$$

$$RE_{year} = RE_{cycle} \frac{Y_{ps}}{Y_{gs}}$$

Where Y_{ps} and Y_{gs} are years to complete a cycle of phenotypic selection and GS, respectively

$Y_{ps} \approx 7$ $Y_{gs} \approx 1$

Par -> F1	F1 -> F2
F1 -> F2	**F2** -> F1
F2 -> F3	
F3 -> F4	
F4 -> F4:5	
F4:5 -> F4:6	
YR1 Test	
YR2 Test	

	RE$_{cycle}$	RE$_{year}$
INC	0.71	4.3
SEV	0.74	4.4
IND	0.73	4.4
FDK	0.71	4.3
ISK	0.73	4.4
DON	0.69	4.1

14

CD, PEV and Random and Stratified

Optimization of the training population when phenotyping capacity is limited.
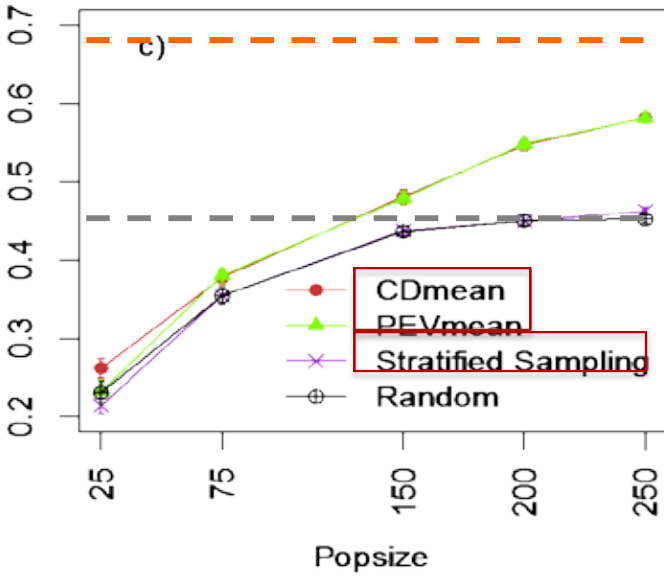
2 ways:

- Capture most of the variability (stratified sampling of clusters)

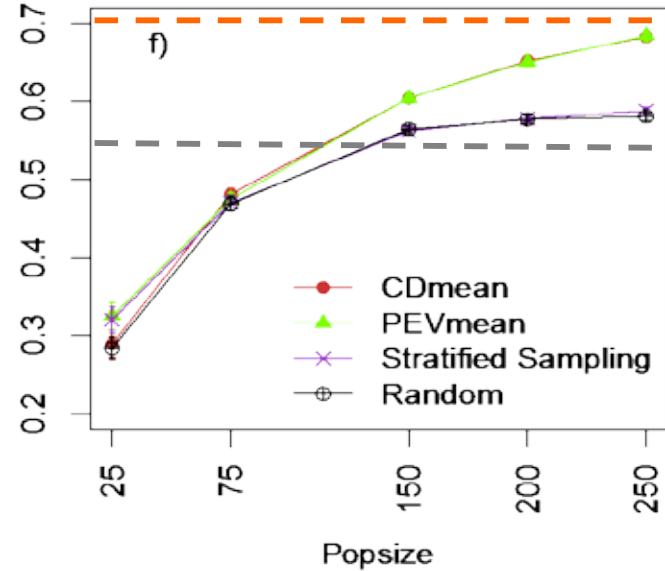- Subset that minimizes the prediction error variance of genetic effects (PEV)
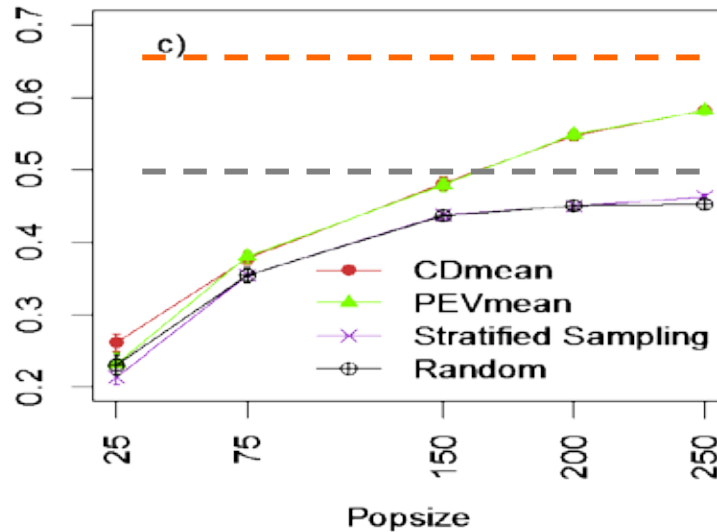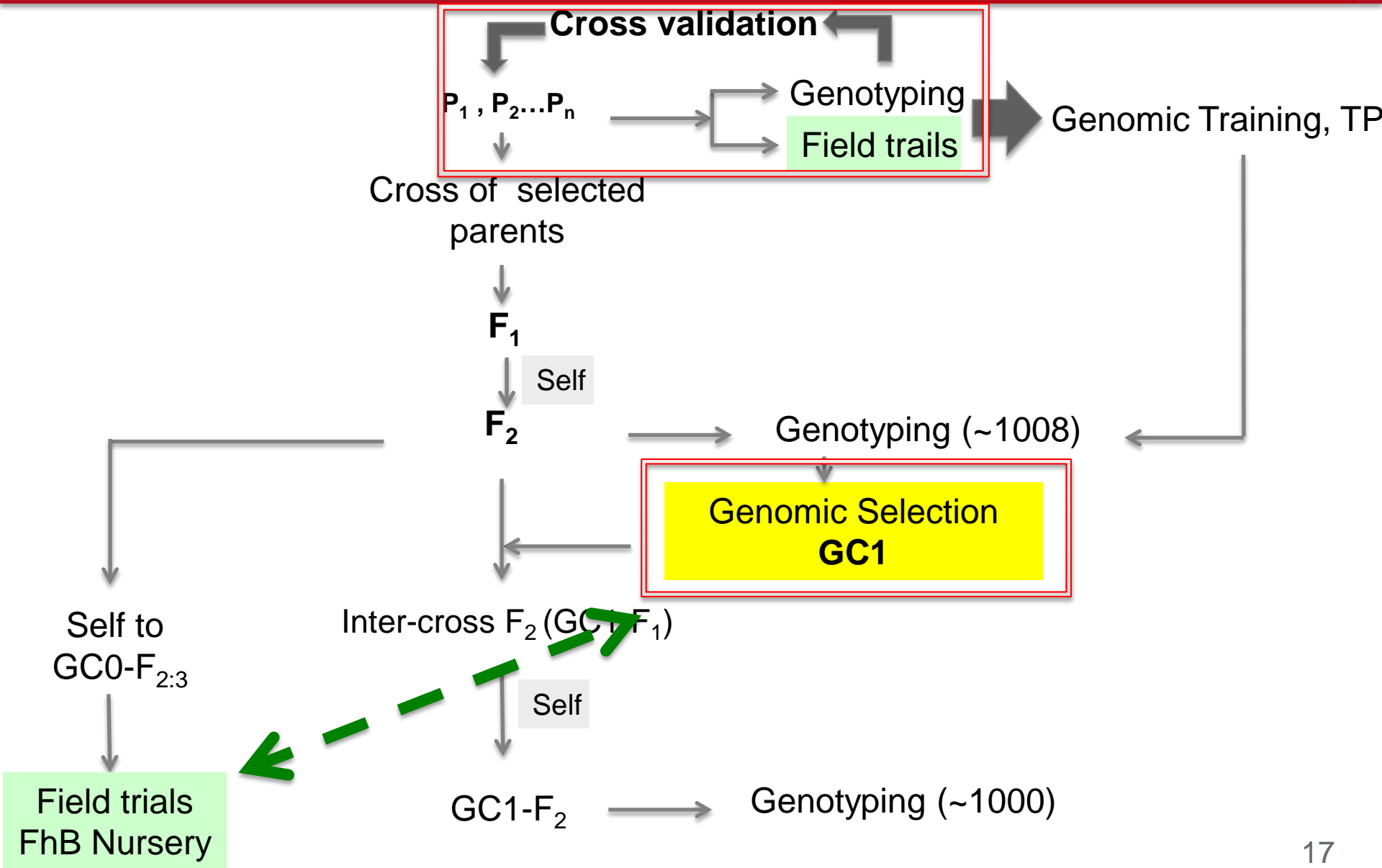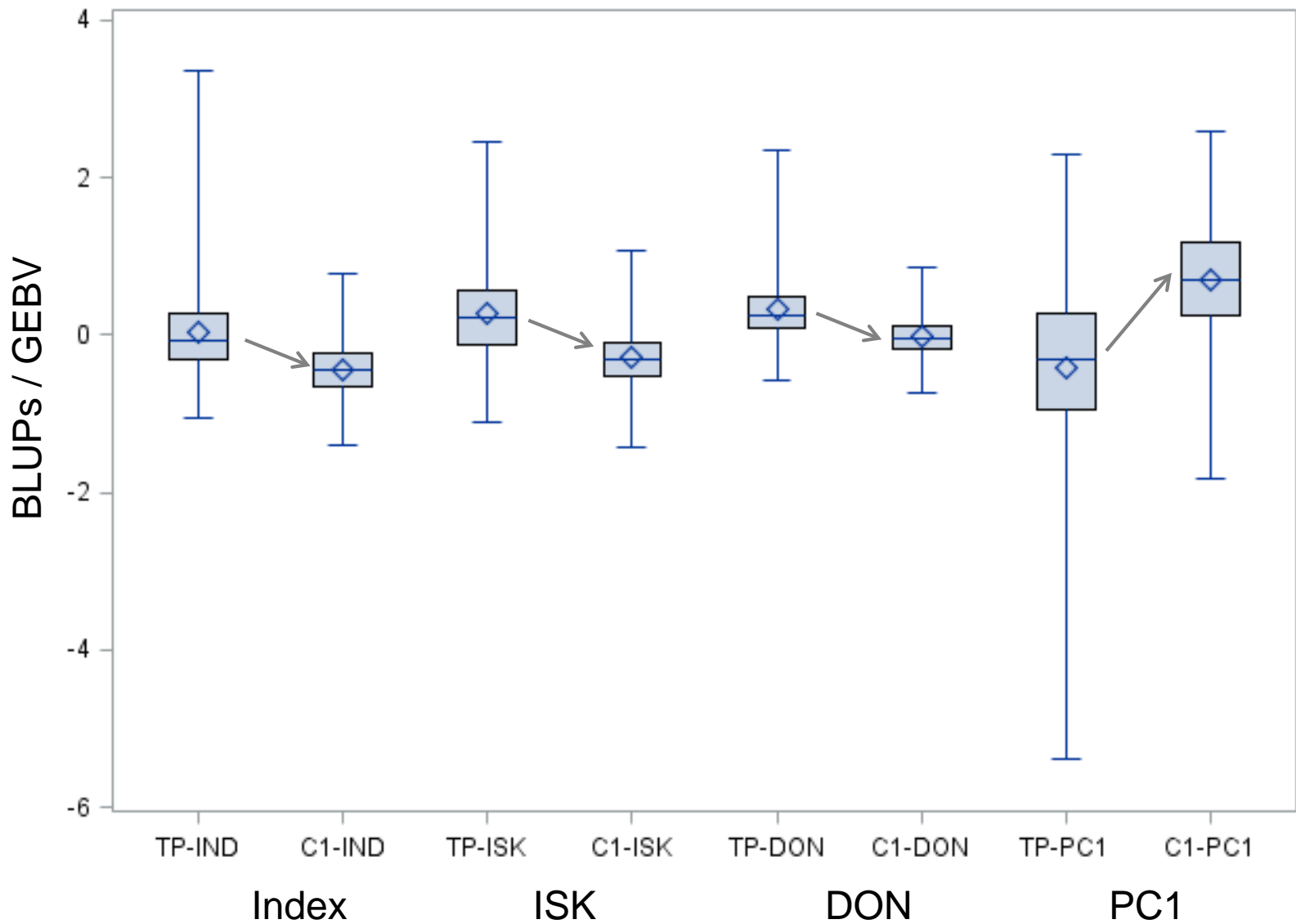
(Rincent et al., 2012, Isidro et al., 2015)

After elimination of less predictive

All lines

Cross validation

P_1 , P_2…P_n → Genotyping

Field trails

→ Genomic Training, TP

Cross of selected parents

F_1

Self

F_2 → Genotyping (~1008)

Genomic Selection **GC1**

Self to GC0-$F_{2:3}$

Inter-cross F_2 (GC1-F_1)

Self

Field trials FhB Nursery

GC1-F_2 → Genotyping (~1000)

17

Summary

1. Prediction accuracy of GS for FHB traits was high.

2. RE_{year} is quite favorable.

3. Prediction over entire populations was slightly more effective than accuracy within clusters.

4. Prediction accuracy between clusters was low.

5. Elimination of less predictive individuals increased the prediction accuracy by 22-49%.

6. Results from CD and PEV demonstrate high accuracies can be obtained with a reduced set of individuals.

7. After one cycle, average GEBVs are changing in desired direction, very susceptible are decreased, individuals with superior GEBVS than the TP are identified.

Acknowledgments

Wheat Genetics Lab (OSU):

Dr. Clay Sneller

Cassi Sewell

Duc hua

Brian Sugarman

Amber Hoffstetter

Mao Huang

Nelly Arguello

Cereal pathology Lab (OSU):
Dr. Pierce Paul

Cornell University
Dr. Mark Sorrells
Dr. Julio Isidro

Michigan State University
Dr. Eric Olson

University of Missouri
Dr. Anne McKendry

University of Illinois
Dr. Fred Kolb

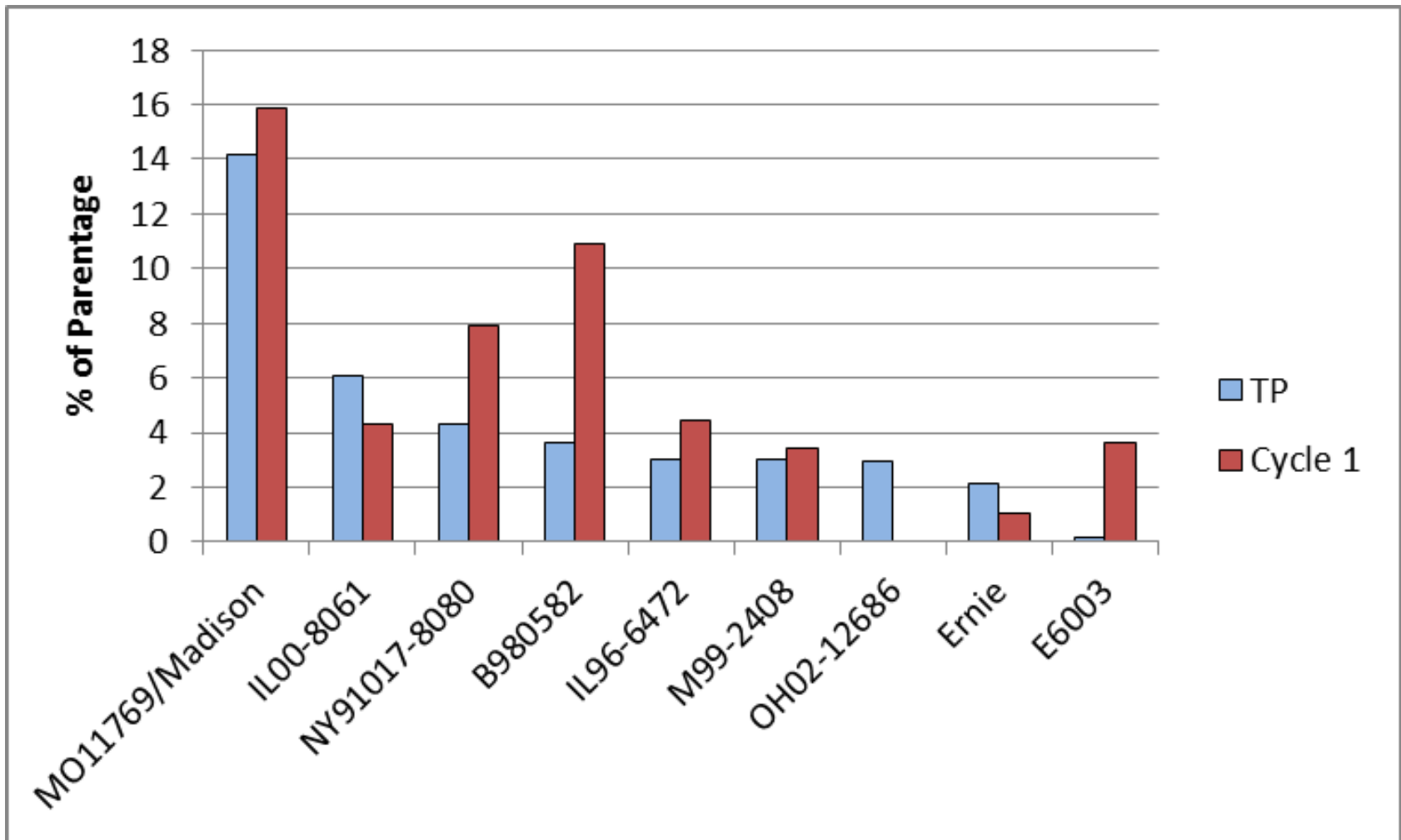University of Kentucky
Dr. David Van Sanford

Thanks!

Rationale

- ## Development of resistant cultivars using traditional methods is inefficient.

  - Time consuming, many years
  - Major genes are not common (absent?) for controlling FHB resistance in SWW.
  - Selection based on few significant QTL alone will not provide optimum resistance levels (i.e. *Fhb1*).
  - Impact of Introgression of *Fhb1* and other QTL is variable in soft winter wheat (SWW): Inconsistent resistance

Shift in percentage of parentage

Phenotype: (Y_{TP})

GS-Models:
RR-BLUP
BL
EN
RF

Genotype (high throughput) X_{TP}

Selection Candidates

Estimation of Marker effects (α)

$X_{SC}*α = GEBV$

Selected individuals based on GEBV