



Genomic Patterns in Distinct Soybean Breeding Pools: Fst Dynamics and impacts of selection strategies on gains

AGRO 932
Final Exam - May 2025

Arthur Bernardeli
PhD Candidate



UNIVERSITY of NEBRASKA
LINCOLN

Soybean history

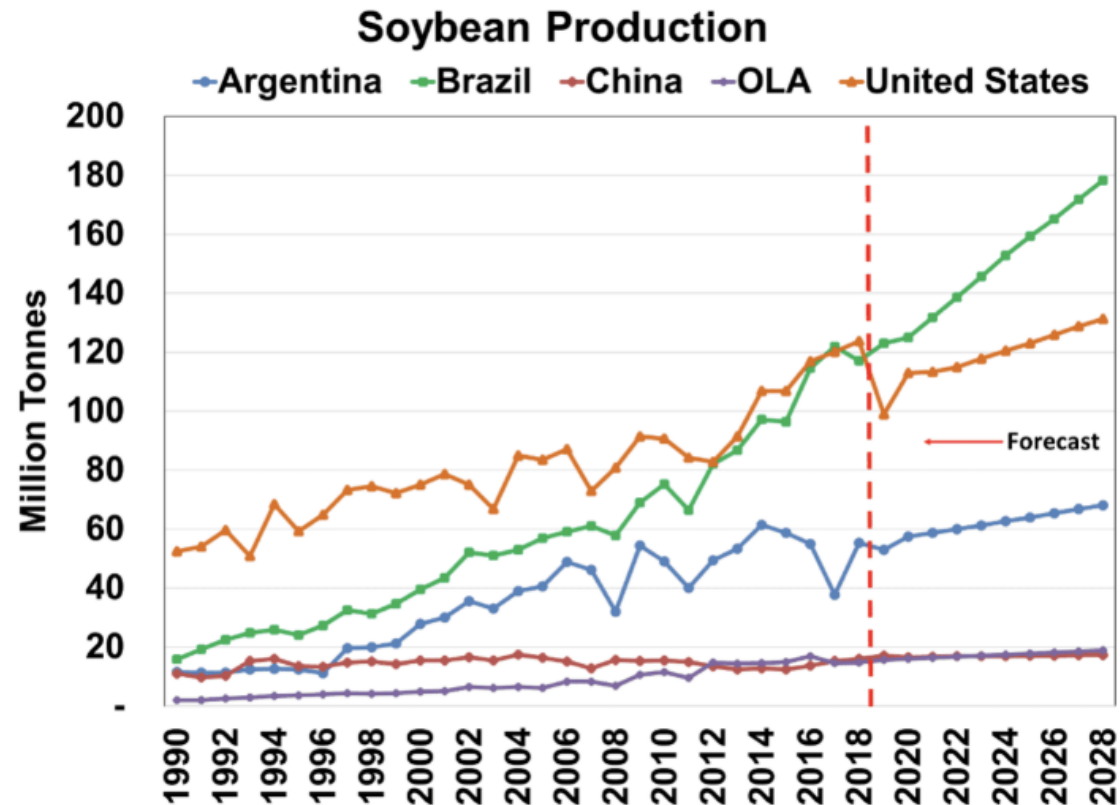
- Soybean was domesticated from *Glycine soja* in East Asia approximately 6,000–9,000 years ago.
- The Huang-Huai Valley in Central China is the primary center of domestication.
- A semi-wild species, *Glycine gracilis*, may have acted as a transitional form.



Sedivy et al. (2017)

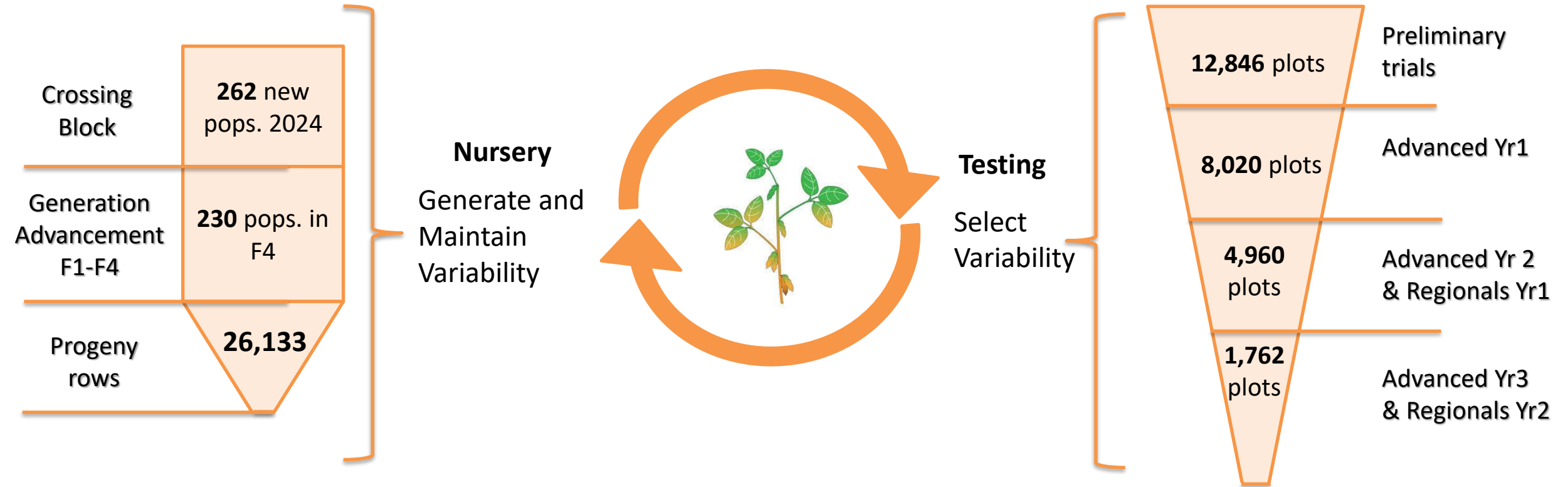
Soybean relevance

- 40% protein and 20% oil
- **NE** vs USA: **57.5** vs 50.7 bu/ac (USDA, 2024)
- 420 mi metric tons: Brazil, US, Argentina (USDA, 2024)



Source: USDA and Doane

Cultivar development outline



- Generate new and elite germplasm (populations)
- Improve selection accuracy (genomic selection or phenotypic selection)
- Reintroduce sources of variability
- Understand limitations

Objectives

- Assess genetic differentiation between high-yield and high-protein populations using F_{ST} to evaluate allele frequency shifts driven by divergent selection.
- Examine how repeated intra-population selection affects genetic diversity and phenotypic means, expecting reduced diversity and increased homogeneity over cycles.
- Compare long-term outcomes of phenotypic vs. genomic selection to guide breeding strategy decisions and identify when to reintroduce genetic variability.

Materials and methods

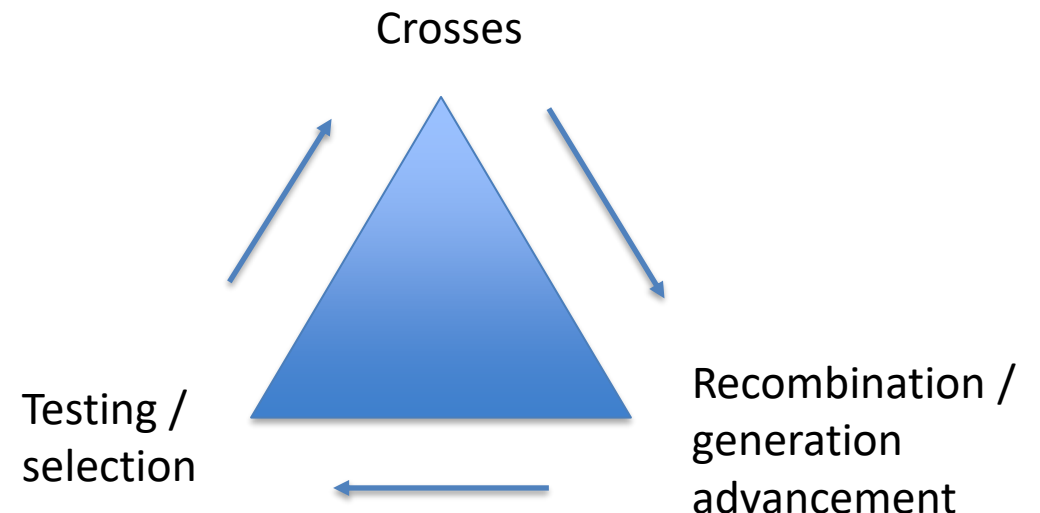
Population Simulation and Selection Pipeline

Founder Population and Trait Architecture

- Simulated 5 founder populations (1/year), each with 40 individuals (20 high-yield, 20 high-protein).
- Used 6,000 SNPs across 20 chromosomes, imputed from the Glycine max reference genome (*Williams 82*).
- Trait simulation used empirical parameters:
 - Yield heritability (H^2) = 0.45
 - Protein heritability (H^2) = 0.70
 - Genetic correlation (r) = -0.45
- AlphaSimR (Gaynor et al., 2021)
- **Unadapted germplasm**

Recurrent Selection

- Within each group, top 10% selected by phenotype
- Selected individuals intermated to form next cycle
- Five cycles simulated to form two divergent **elite** pools



Genomic Analyses and Selection Scenarios

Analyses During Initial Cycles (cycles 1-5)

- Calculated F_{st} each cycle: **unadapted germplasm and elite pool**
- PCA visualized genetic divergence across groups and generations.
- Phenotypic means tracked to assess gains and trade-offs.

Post-Cycle 5 Selection:

Phenotypic and genomic selection **of improved populations from the elite pool**

- Top 5 individuals from Y and P were diallel-crossed to generate 1,000 individuals.
- Compared phenotypic vs. genomic selection over 10 more cycles.
- Genomic selection via RR-BLUP:
 - Model: $y = \mu + Zu + e$
 - $GEBV_i = \sum(Z_{ij} * \hat{u}_j)$
- Training/testing split: 250/250 for accuracy evaluation.
- First half: germplasm adaptation
- Second half: germplasm improvement (cultivar development context)

Results

Genomic Divergence and Phenotypic Gains (Cycles 0–5)

•Initial State (Cycle 0):

- Minimal genetic differentiation between protein (P) and yield (Y) groups (low F_{ST}).
- PCA showed no clustering — similar genetic background.

•Divergence Over Cycles:

- F_{ST} increased steadily (Y: 0.238→0.626, P: 0.243→0.589).
- PCA revealed clear genetic separation by Cycle 5.

•Phenotypic Responses:

- Group P: Higher protein over time.
- Group Y: Increased yield.
- F_{ST} increases aligned with trait improvement.

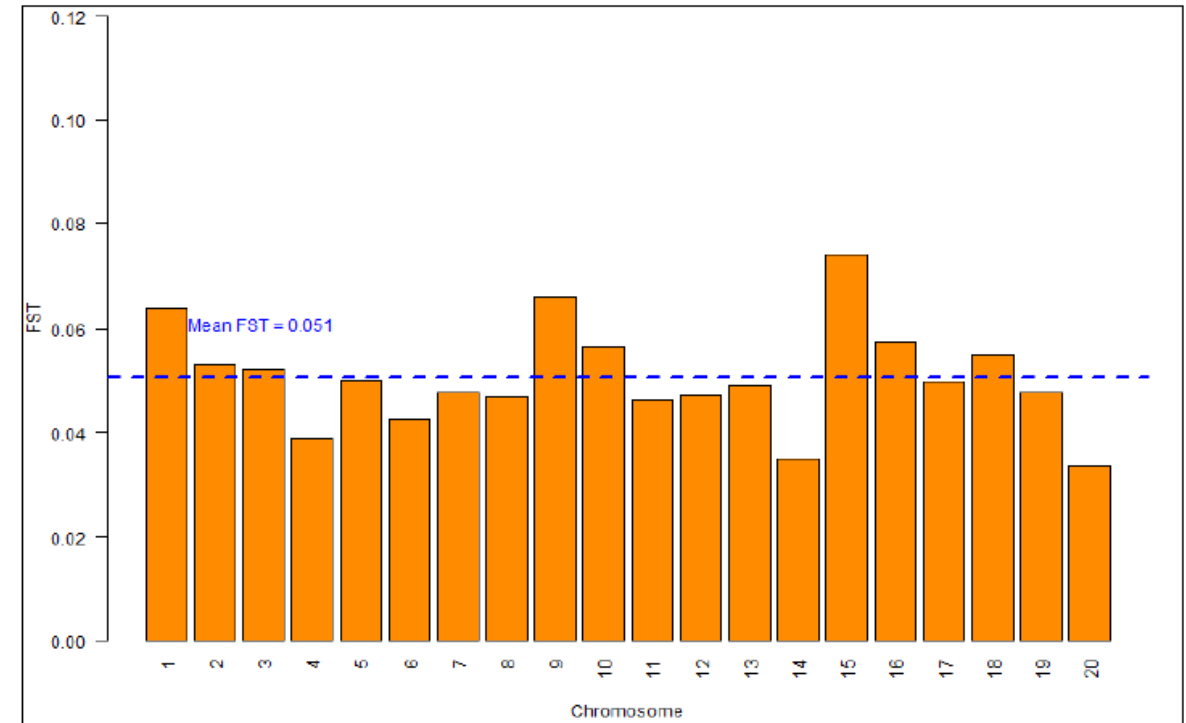
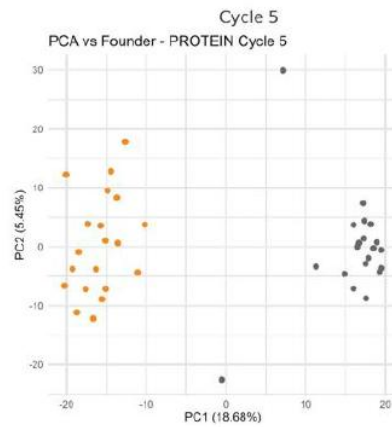
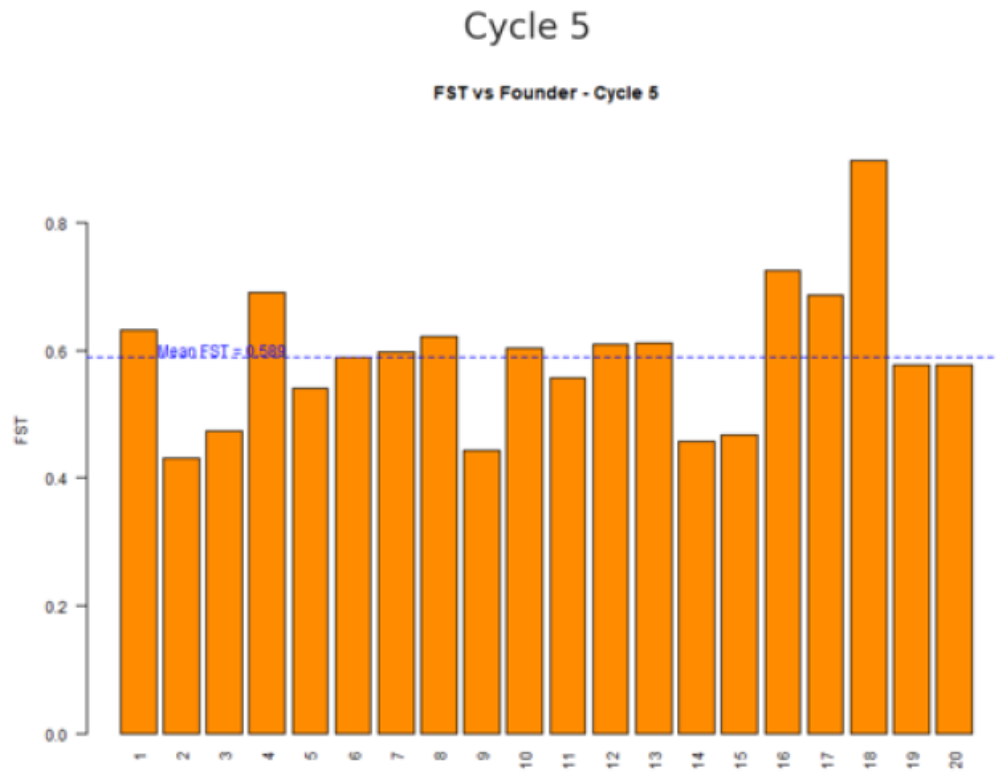
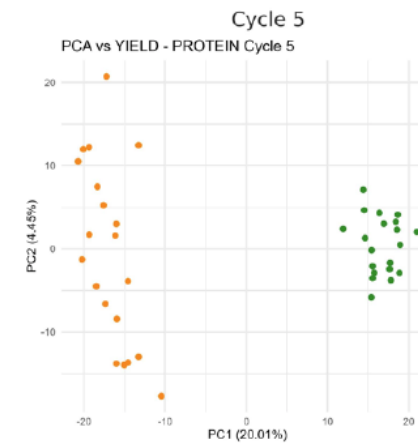
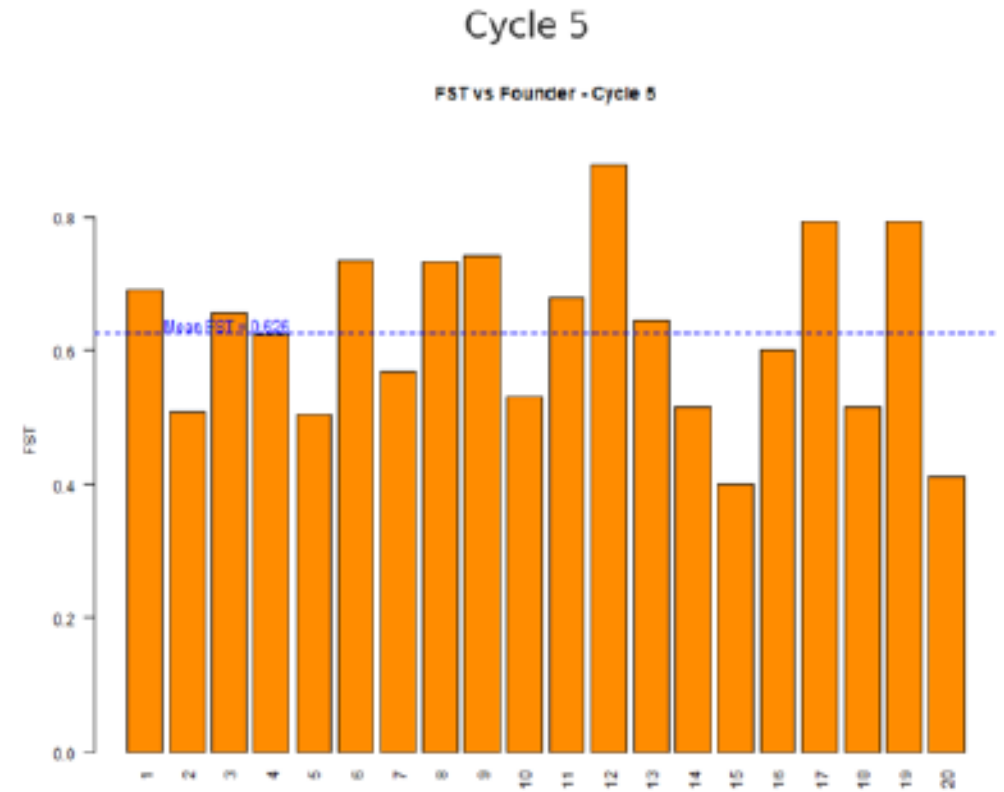


Figure 1. Chromosome-wise F_{ST} between individuals of the founder population (cycle 0).

Cycle 5 – Protein pop. and Founder pop.



Cycle 5 – Yield pop. and Founder pop.



- Distinct elite yield and protein populations

Initial
germ.

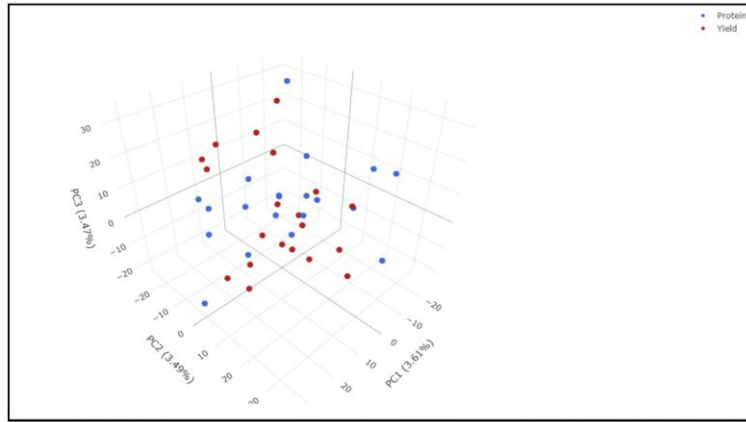


Figure 2. Principal component analyses plot (PCA) between individuals of the founder population (cycle 0).

Protein: initial
germ. vs Cycle 5

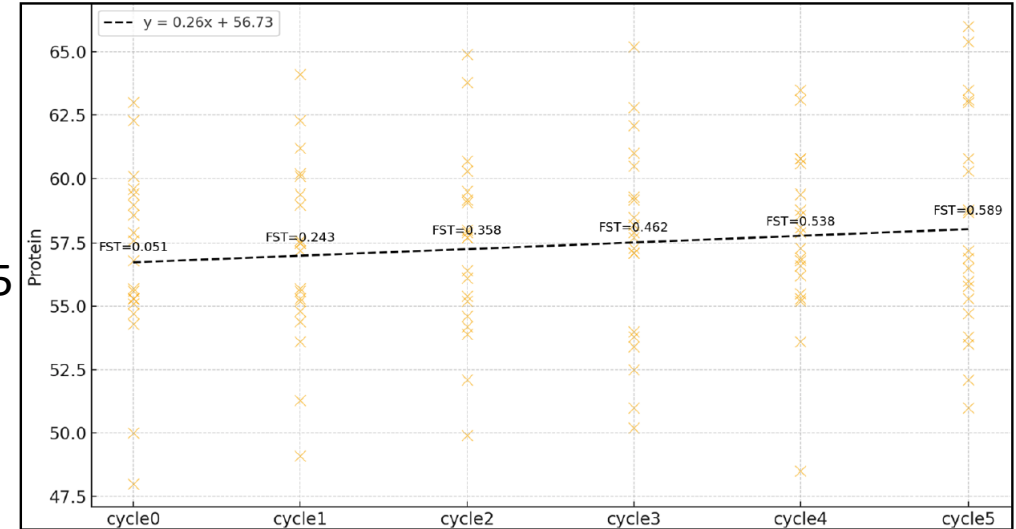
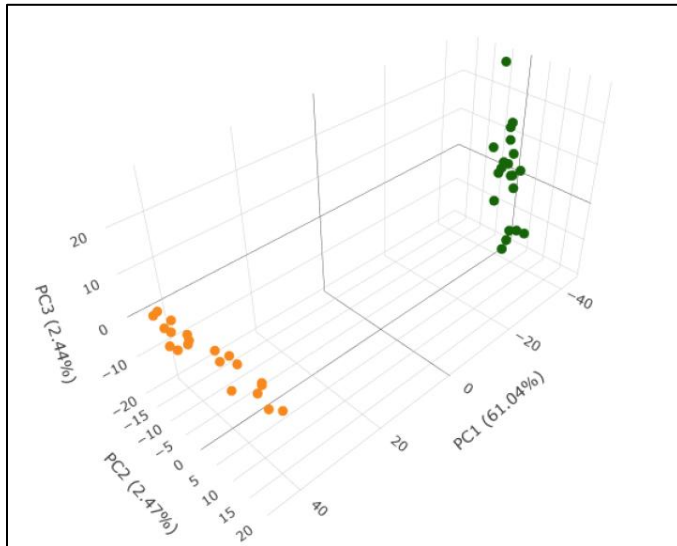


Figure 10. Changes in F_{st} and protein content mean for each selection cycle for group P. Protein content (y-axis) is represented on a percentage basis.

Cycle 5



Yield: initial
germ. vs Cycle 5

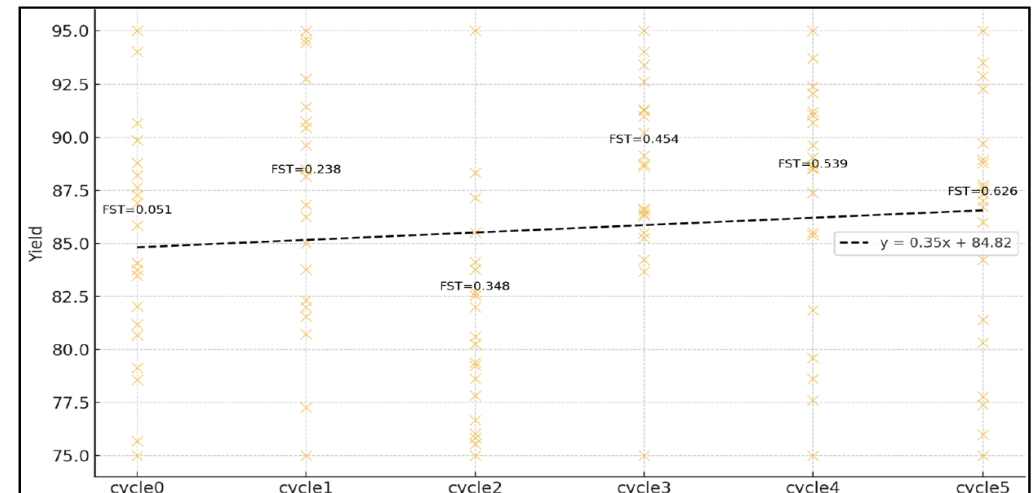


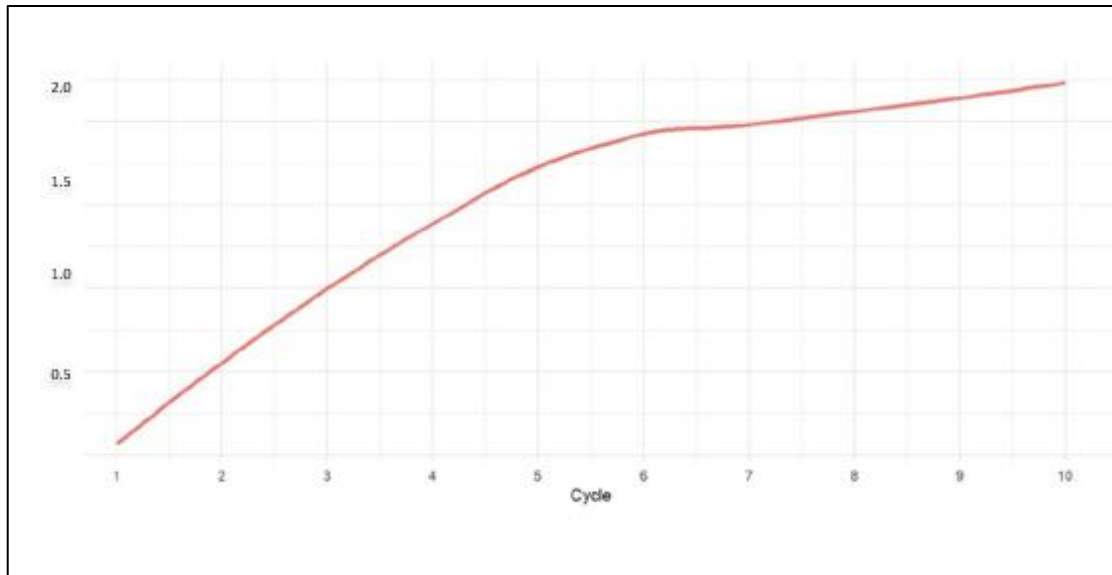
Figure 11. Changes in F_{st} and yield content mean for each selection cycle for group Y. Yield (y-axis) is represented on a bushels per acre basis.

Phenotypic vs. Genomic Selection (Cycles 6–15)

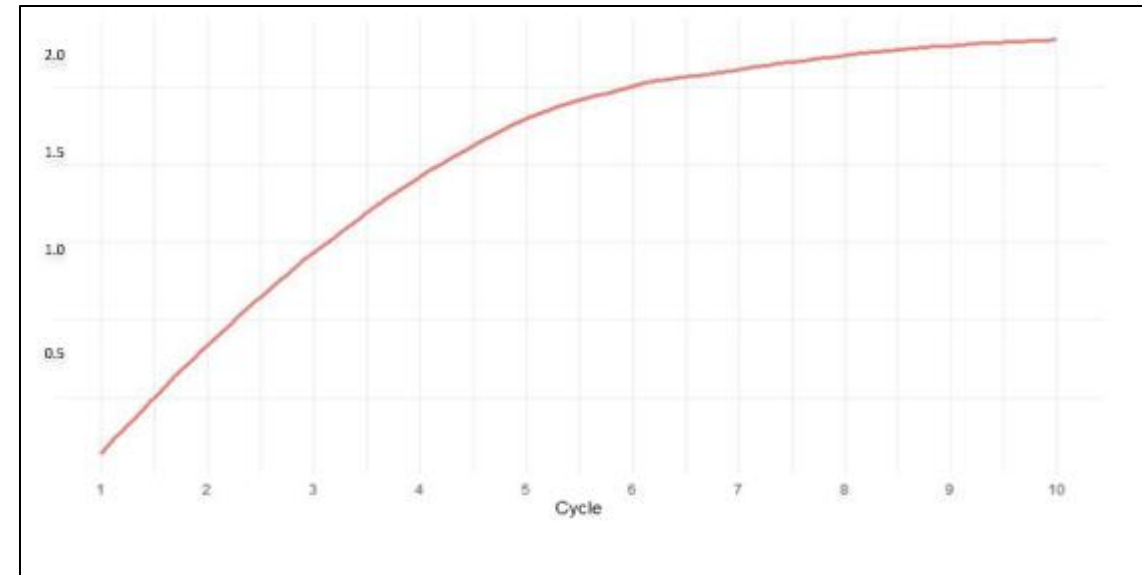
Phenotypic Selection

- Yield and protein gains plateaued after Cycle 6.
 - Early cycles saw strong gains due to favorable allele fixation.

Yield - phenotypic selection



Protein - phenotypic selection

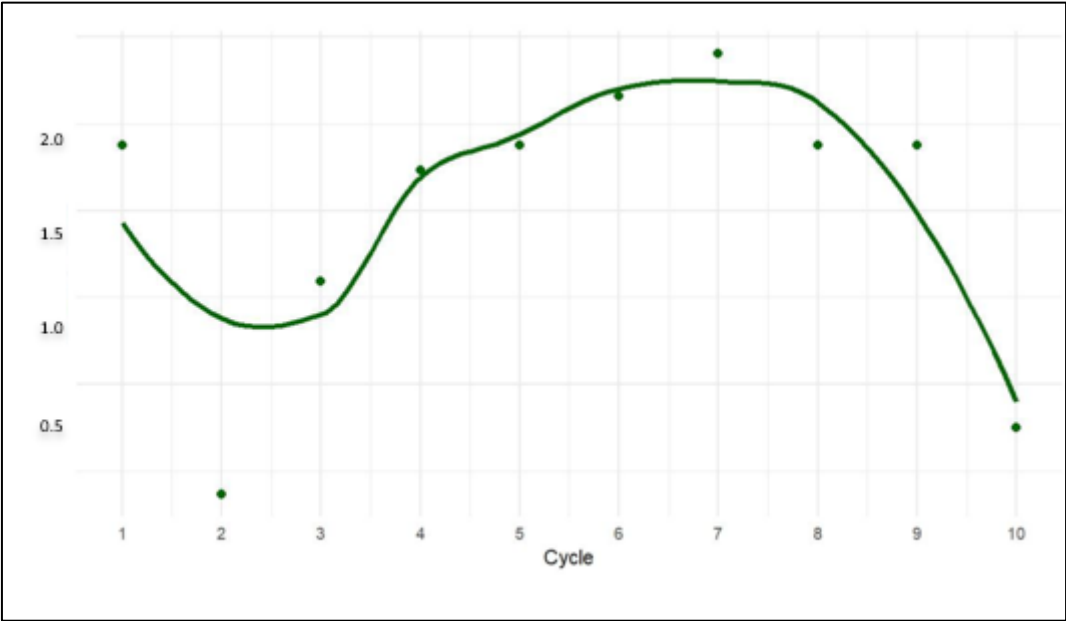


Genomic Selection (RR-BLUP):

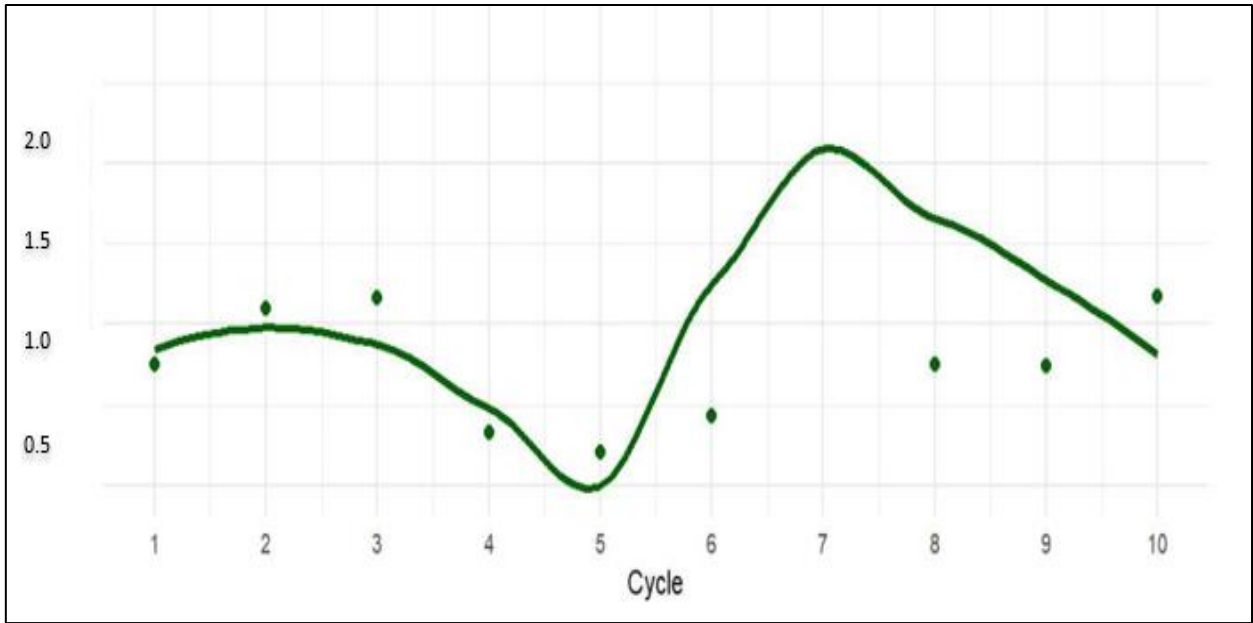
- Initial gains were low or negative, poor training optimization (Akdemir et al, 2015)
- Gains rose mid-cycle, then plateaued/decreased again.
- Prediction accuracy ranged 33–44% (yield), 39–56% (protein).
- RR-BLUP favored related individuals, narrowing diversity (Jannink, 2010).

Prediction accuracies		
Cycle	Yield	Protein
1	43.05	49.01
2	40.01	53.97
3	33.99	48.97
4	38.99	39
5	40.92	38.01
6	43.98	55.99
7	41.01	56.02
8	40.57	50.99
9	39.98	49.99
10	33.05	47.97

Yield - RR-BLUP



Protein - RR-BLUP



Implications for Breeding Programs

Key Outcomes:

- Divergent selection on shared founders leads to rapid genomic separation.
- Both selection methods improved target traits, but dynamics differed.

Lessons for Breeders:

- F_{st} is effective for short to mid-term changes.
- Genomic selection requires optimized training sets and continuous diversity input.
- Long-term gains require balancing with genetic variability conservation.