

# Redesigning Plant Breeding Systems

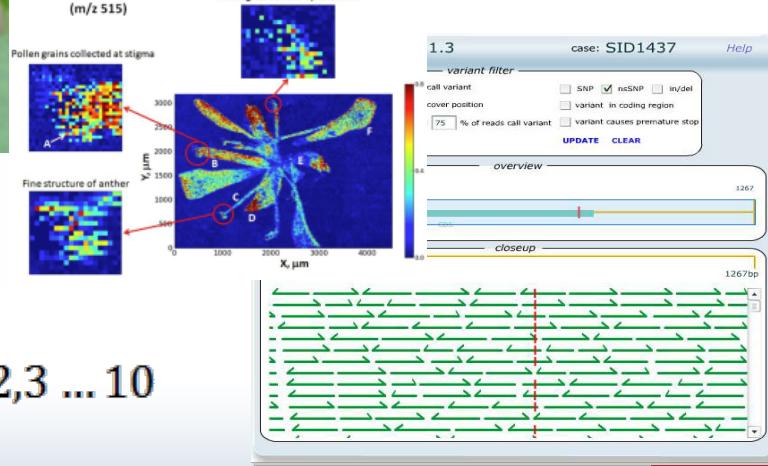


$$a_{i,2} * j - 1 * L_j = x_{i,1}$$

$$a_{i,2} * j * L_j = x_{i,2}$$

$$\sum_{j=1}^{10} L_j = 1 \quad \forall j = 1, 2, 3 \dots 10$$

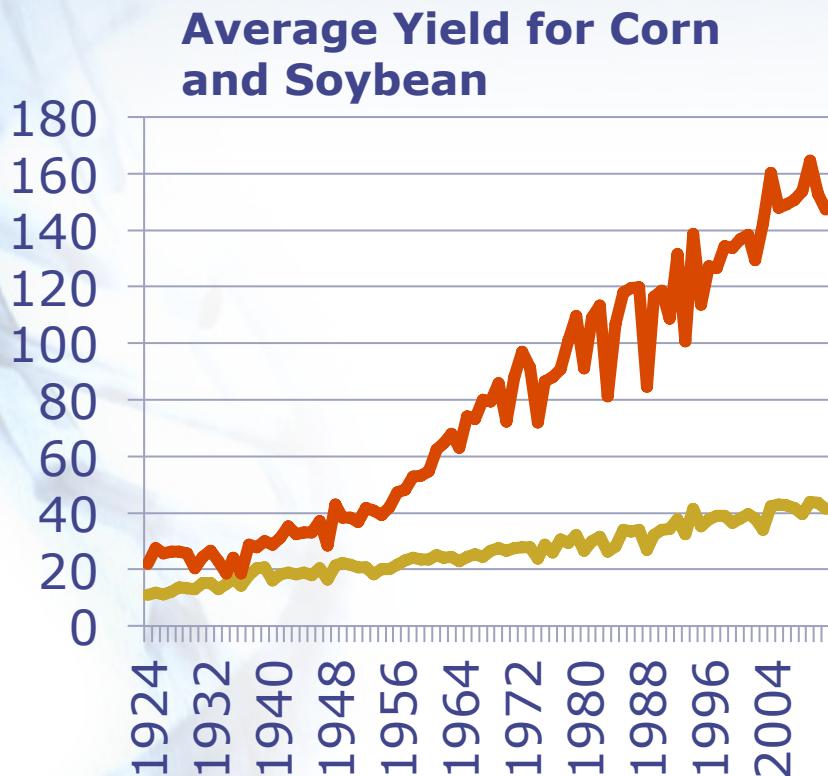
$$\sum_{j=1}^{10} U_j = 1 \quad \forall j = 1, 2, 3 \dots 10$$



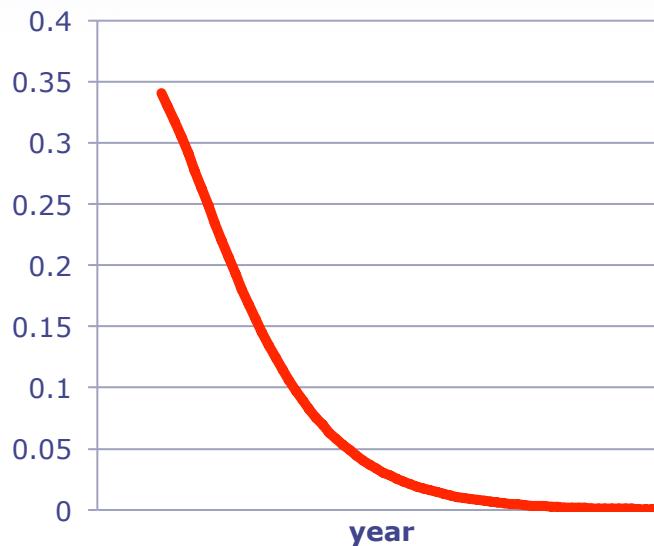
15 February 2019  
Translational Genomics Forum  
Stillwater, Ok

Bill Beavis  
Lizhi Wang  
Guiping Hu  
Sotirios Archontoulis

# Redesigning Plant Breeding Systems: Motivation



**Genetic Gain per \$US  
as a function of time**



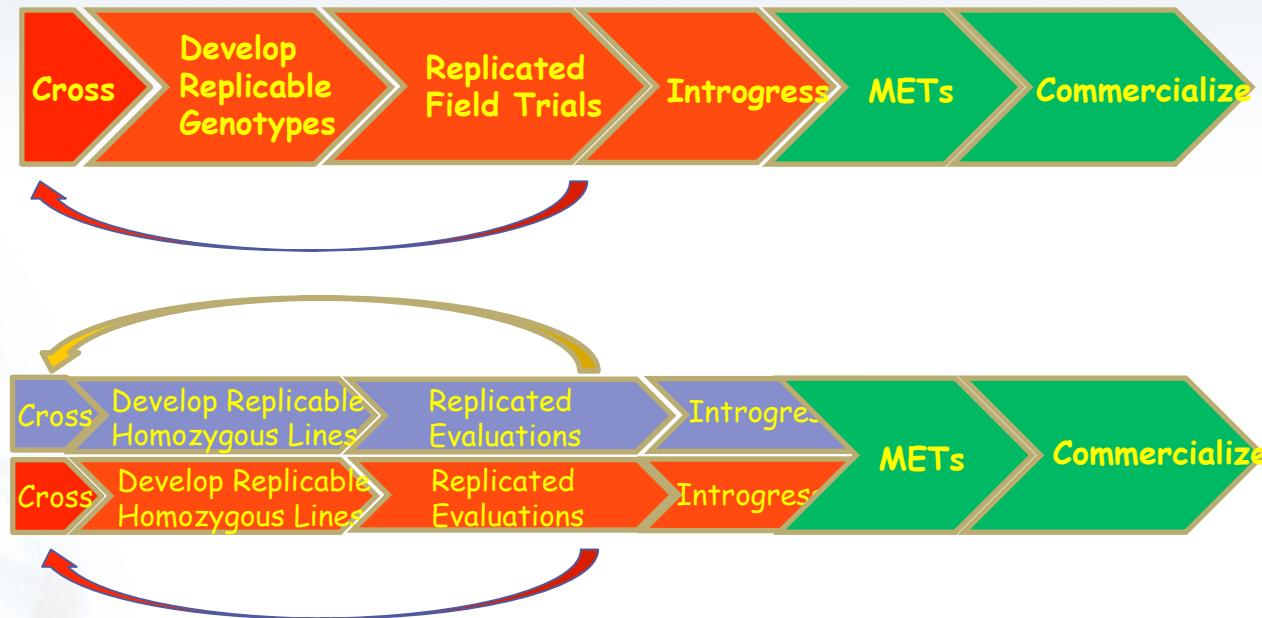
# Pipeline Representation of Cultivar Development



Currently, we integrate new technologies and discoveries into existing breeding systems using *ad hoc* approaches. We do this because we can and we are afraid that our competitors will. We do not know if the resulting design are optimal. – Ted Crosbie, 2009



# Representations of Line and Hybrid Cultivar Development Systems



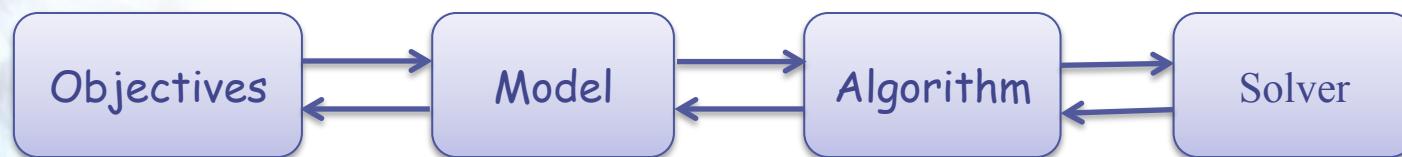
Commercial plant breeding has become an escalating arms race where large seed companies ask farmers to pay for integration of new technologies and ever expanding numbers of field trials. We need to be more efficient not more expensive. Joe Byrum, 2009

# Redesigning Plant Breeding Systems

---

As Biology is transformed from an observational to a predictive science, Plant Breeders have the opportunity to return to their original role as designers of Genetic Improvement Systems.

**Proposal:** Redesign plant breeding systems using principles and methods from Applied Mathematics

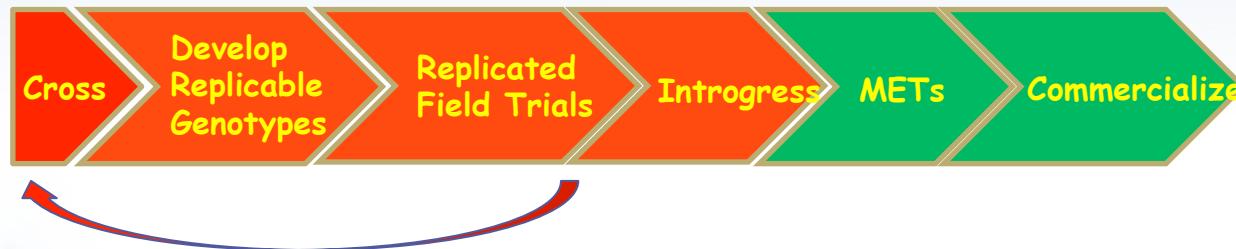


# PREMISE: PLANT BREEDING IS A DECISION MAKING DISCIPLINE

---

- Usually, decision making requires tradeoffs.
- Cultivars should be designed based on knowledge of attributes needed by farmers.
- How to create a cultivar will be constrained by reproductive biology, access to technologies, budgets and time.
- How to create the cultivar should be influenced but **not** constrained by historical designs of breeding systems.

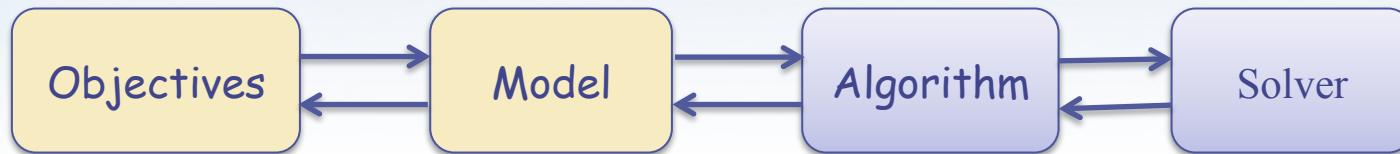
# Proposal: Redesign plant breeding systems using Applied Mathematics



## Example 1: Marker Assisted Introgression

**Hypothesis:** After 20+ years of research on marker assisted backcrossing, applied mathematics will not improve the best known selection strategies for introgressing a single desirable allele from an unrelated donor.

# Define Objectives and Translate to Objective Functions



$$\max G_1,$$

$$\max G_2, \dots$$

$$\max G_k$$

s.t.

$$G_1 = z_1(x)$$

$$G_2 = z_2(x) \dots$$

$$G_k = z_k(x)$$

$$0 \leq g_i(x), i = 1 \dots m$$

$$0 = h_j(x), j = 1 \dots p$$

*find  $\mathbf{x}^*$  to optimize*  $\mathbf{F}(\mathbf{x}^*) = [z_l(\mathbf{x}^*)]_{l=1}^k$

*s.t.*

$$\mathbf{x} = \{x_1, x_2, \dots, x_n\}$$

$$0 \leq g_i(\mathbf{x}^*), i = 1 \dots m$$

$$0 = h_j(\mathbf{x}^*), j = 1 \dots p$$

# Example 1: Define Measurable Objectives

---

## Breeding objectives (Peng et al, 2014):

Minimize costs to create a set of four ILs in less than five generations with an average of no more than 1 cM of the genome from the DL is present within 20 cM of e/ with an average of no more than 8 cM of genome from the DL is present throughout the rest of the genome.

**An alternative metric** (Cameron et al, 2017): Define average outcome as a successful outcome and determine frequencies of successful outcomes.

$$s = \{ll_{,1} \cdot ll_{,2} \geq 19 \cap pl_{,1} \cdot pl_{,2} \geq 1792 \cap g \leq 5\},$$

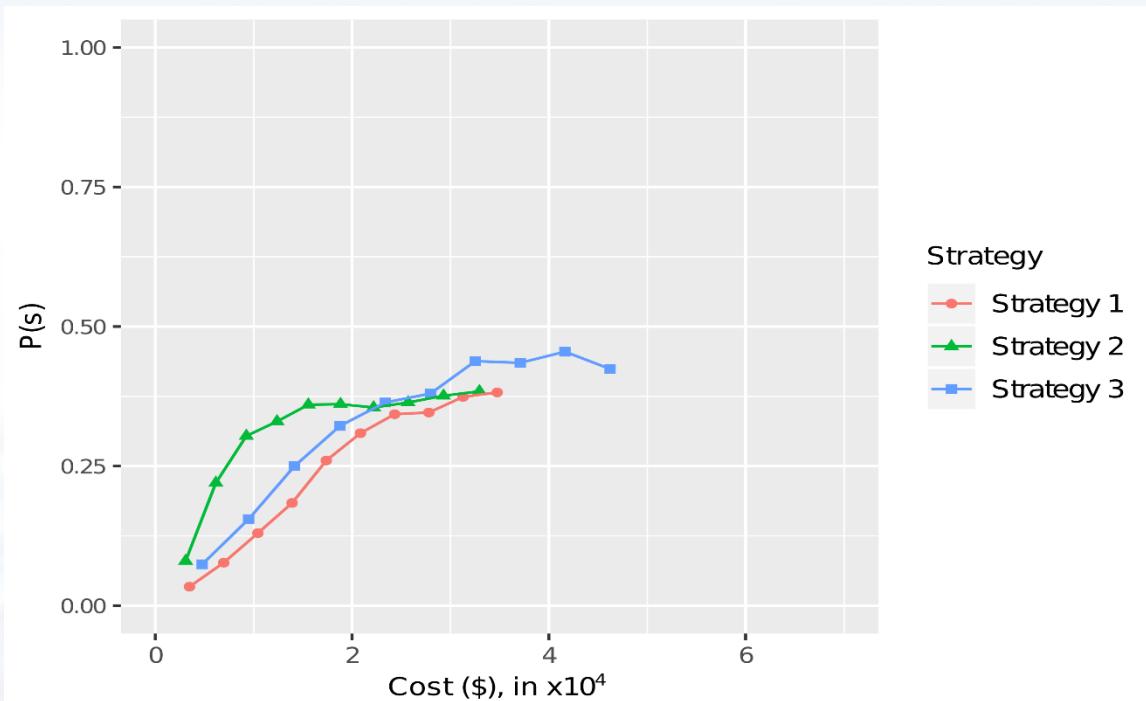
# Example 1: Model the Objectives

---

$$\min(\zeta = c_k^p + c_k^m), \quad c_k^m = \sum_{g=1}^{t^{el}, t^{ul}} \{ [(\$0.05 / ma)(nma_g / p) + \$0.50 / p][np_g] | strategy_k \},$$
$$c_k^p = \sum_{g=1}^{t^{el}, t^{ul}} [(\$5.00 / p)_g(np_g) | strategy_k],$$

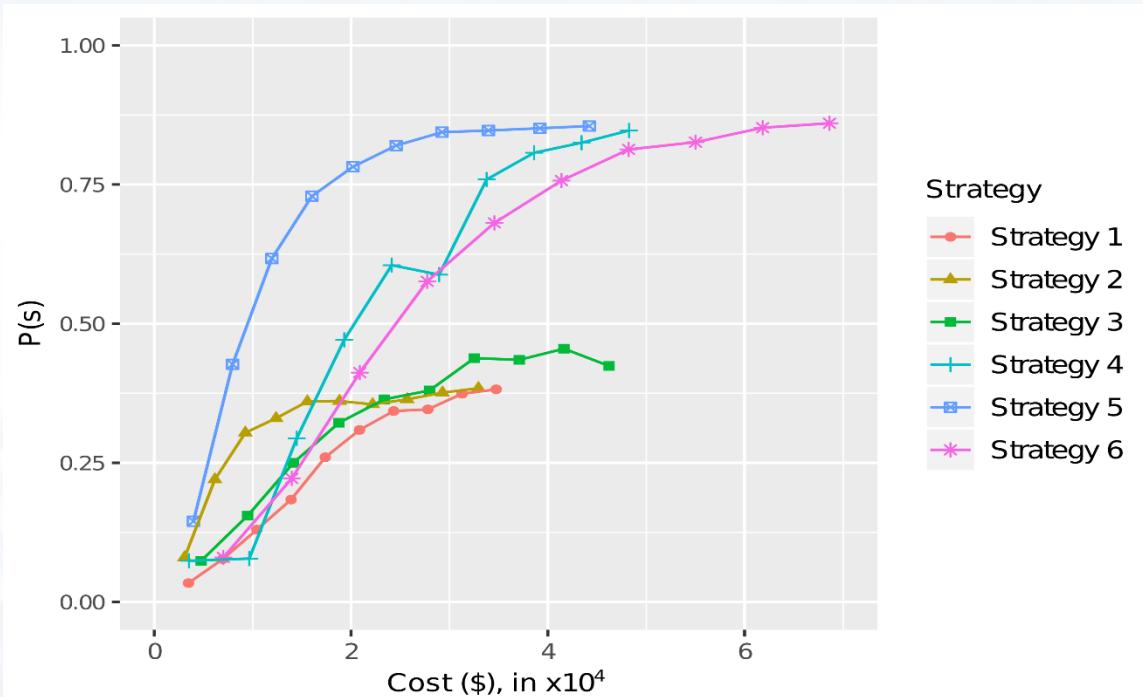
$$\max P(s), \quad P(s) = freq(s | strategy_k, np_g)$$
$$s = \{ \text{ll}_{,1} \bullet \text{ll}_{,2} \geq 19 \cap \text{pl}_{,1} \bullet \text{pl}_{,2} \geq 1792 \cap g \leq 5 \},$$

# Example 1 : Develop at Least Four Successful ILs < 2 years.



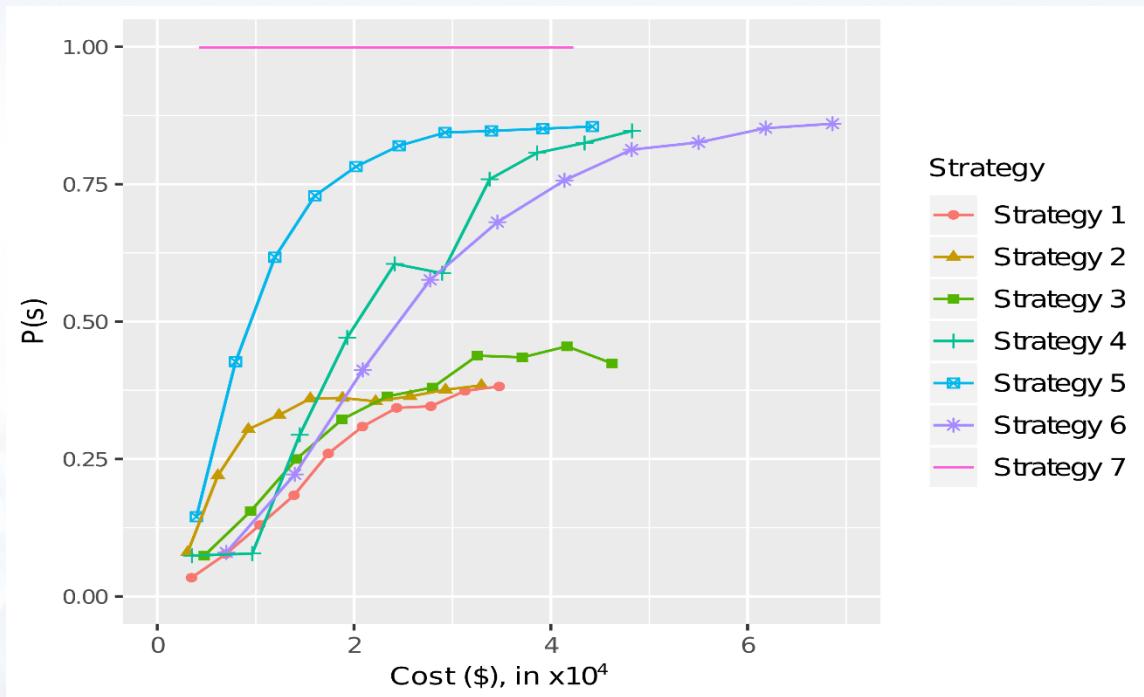
ul markers placed 1/20 cM for backcross strategies 1,2,3

# Example 1 : Develop at Least Four Successful ILs < 2 years.



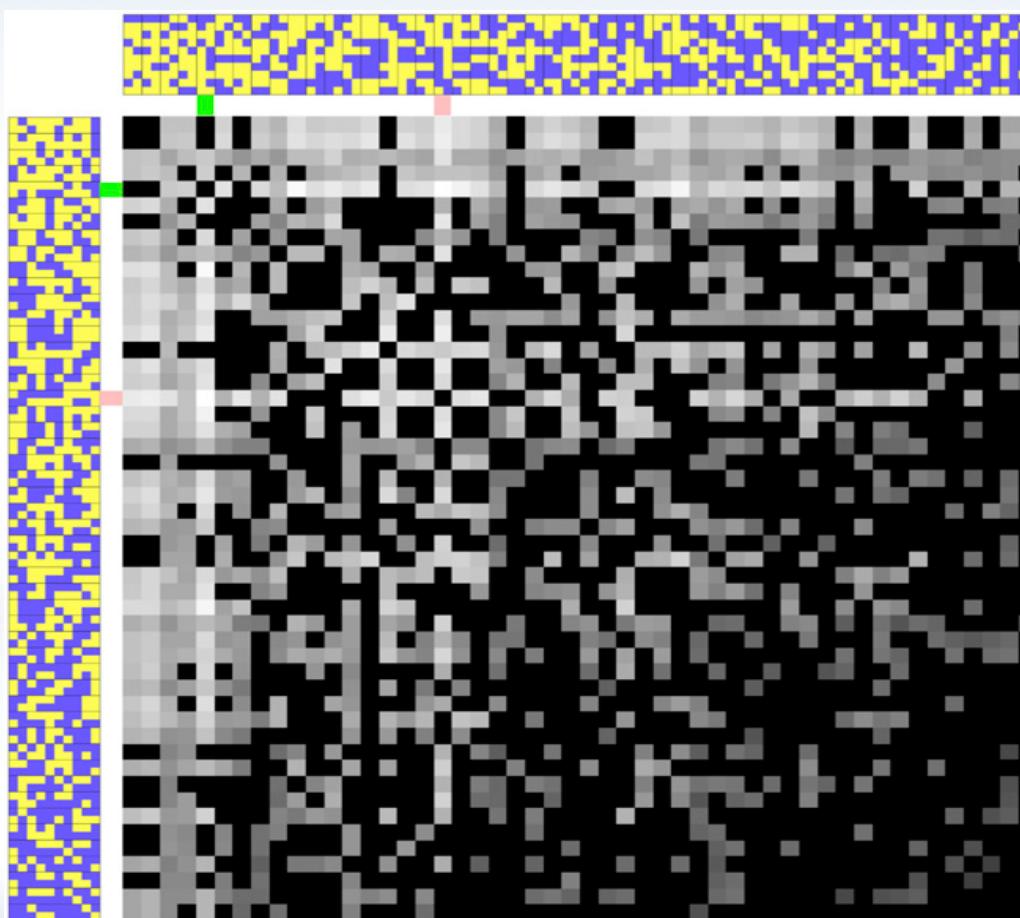
ul markers placed 1/20 cM for backcross strategies 1,2,3  
ul markers placed 1/10 cM for backcross strategies 4,5,6

# Example 1 : Develop at Least Four Successful ILs < 2 years.

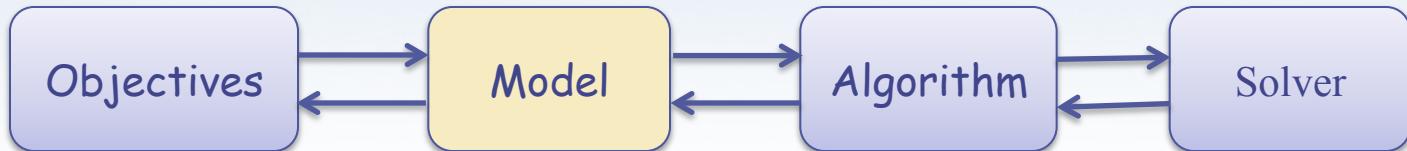


Backcrossing Strategies Are Not Optimal for Introgression!

# Predicted Cross Value for all Possible Crosses Among 50 BC<sub>1</sub>F<sub>1</sub> Progeny



# Example 1: Were there Multiple Objectives?



$$\min(\zeta = c_k^p + c_k^m), \quad c_k^m = \sum_{g=1}^{t^{el}, t^{ul}} \{ [(\$0.05 / ma)(nma_g / p) + \$0.50 / p][np_g] | strategy_k \},$$

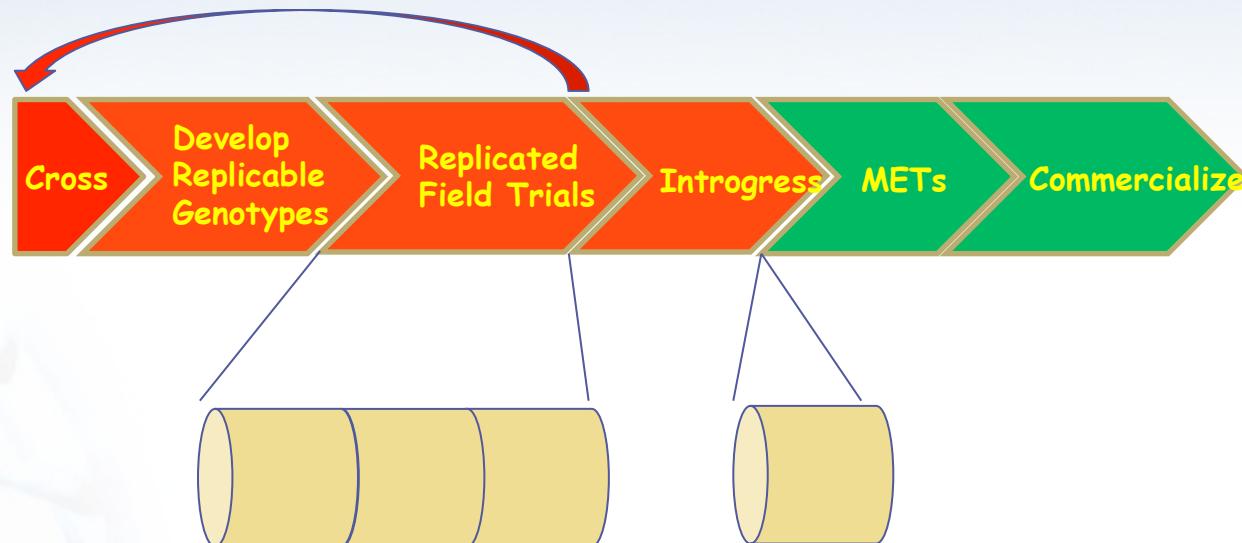
$$c_k^p = \sum_{g=1}^{t^{el}, t^{ul}} [(\$5.00 / p)_g(np_g) | strategy_k],$$

$$\max P(s), \quad P(s) = freq(s | strategy_k, np_g)$$

$$s = \{ ll_{,1} \bullet ll_{,2} \geq 19 \cap pl_{,1} \bullet pl_{,2} \geq 1792 \cap g \leq 5 \},$$

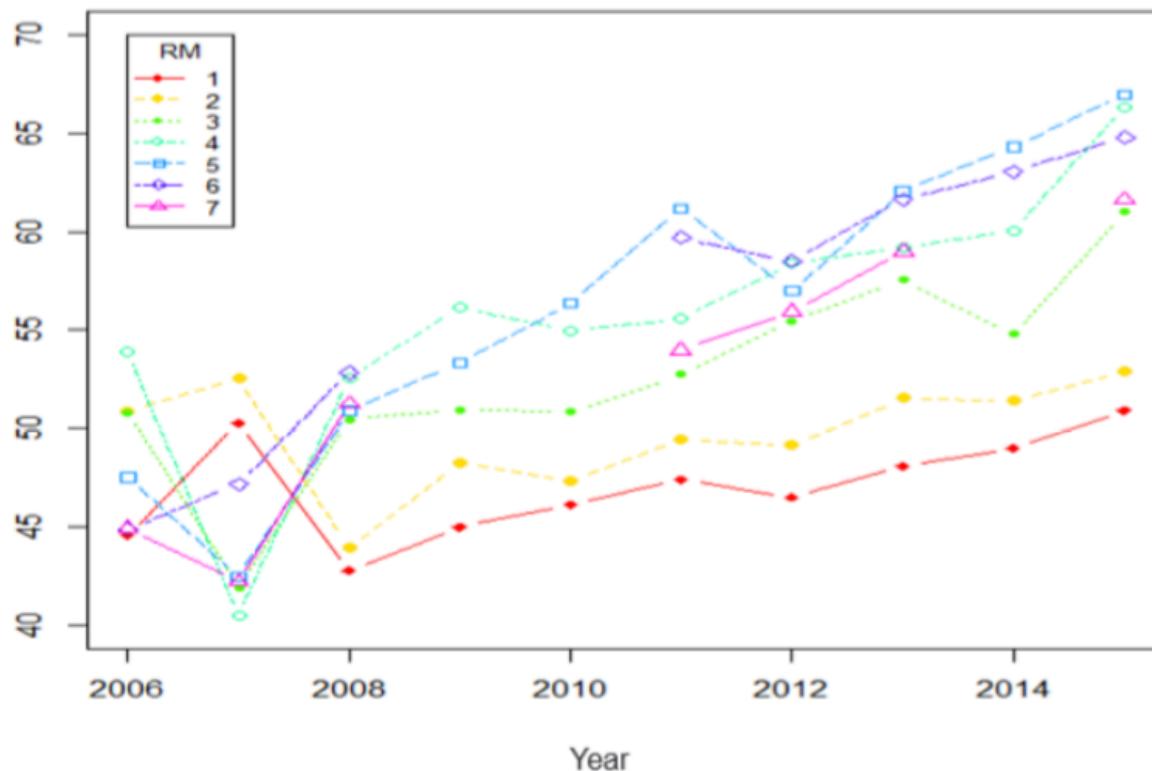
$$s = \{ ll_{,1} \bullet ll_{,2} \geq 19 \cap pl_{,1} \bullet pl_{,2} \geq 1792 \},$$

# Example 2: Redesign Replicated Field Trials

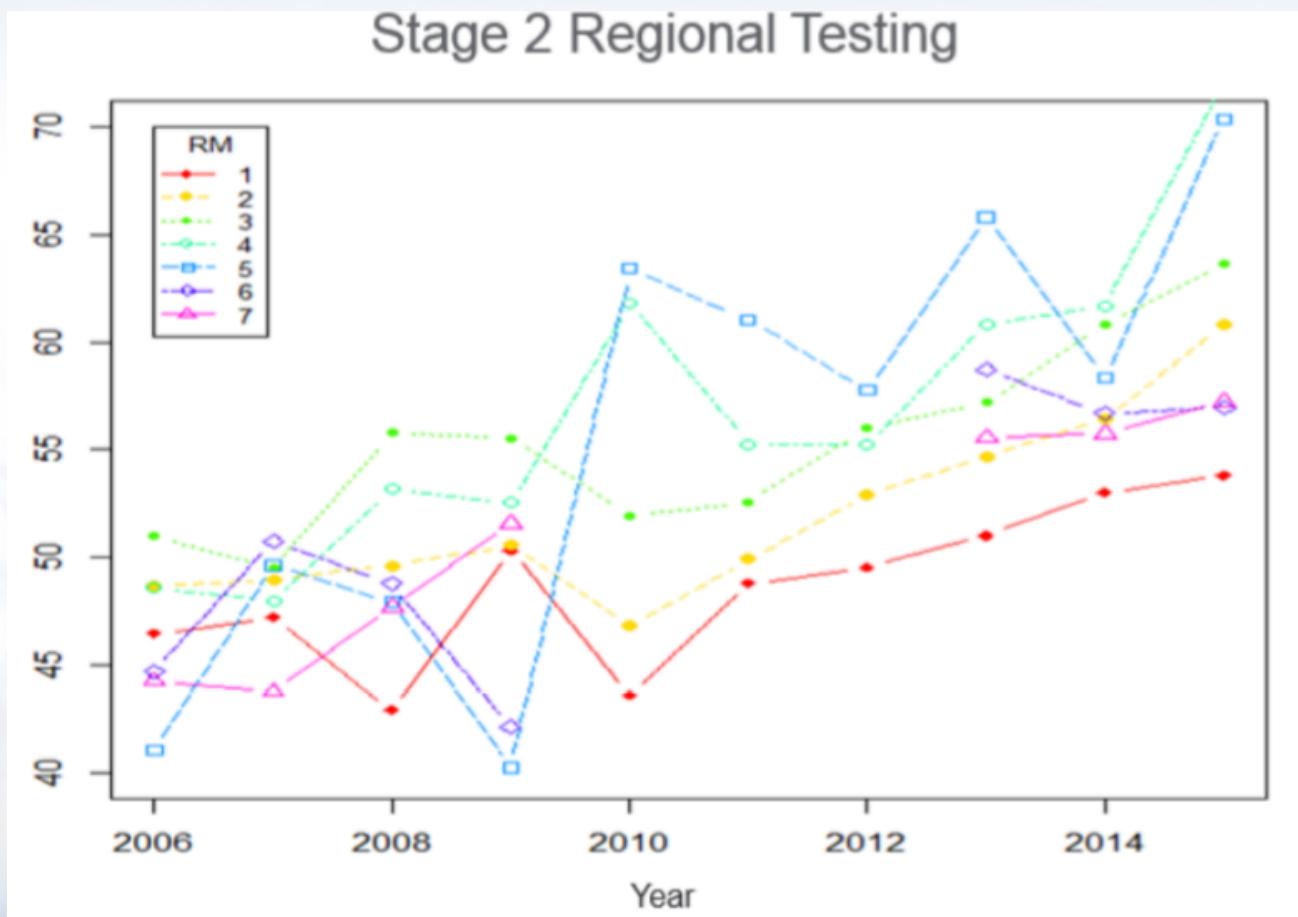


# Outcomes from Redesigned Replicated Field Trials

Stage 1 Regional Testing

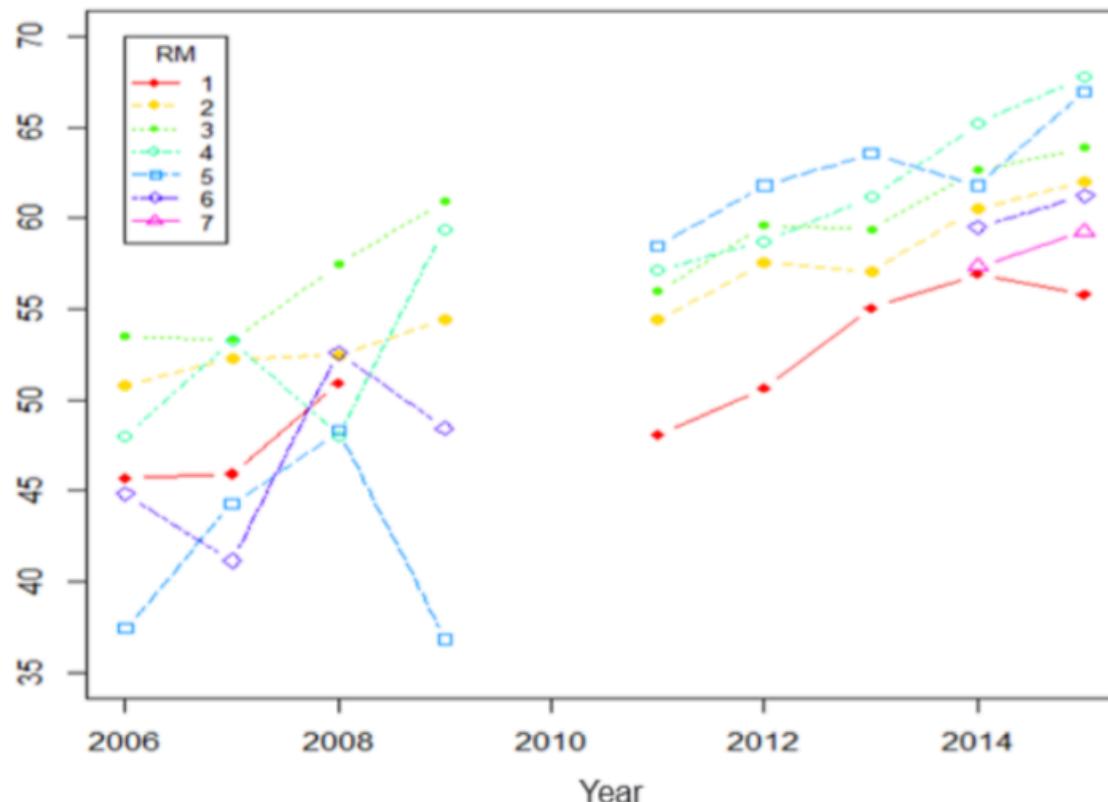


# Outcomes from Redesigned Replicated Field Trials

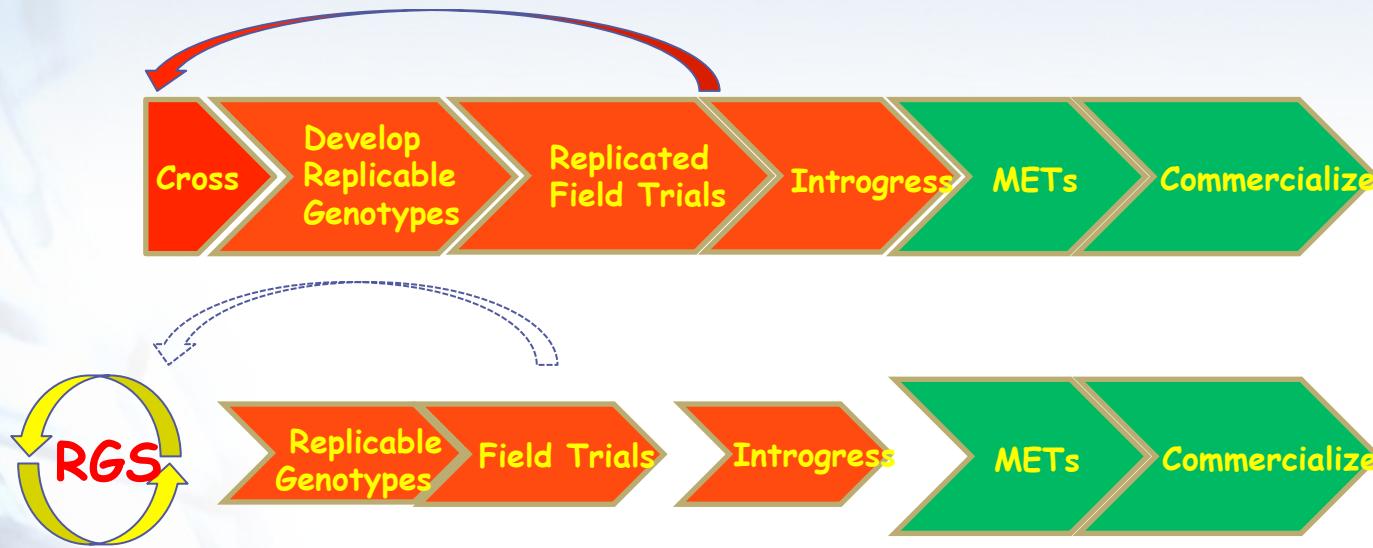


# Outcomes from Redesigned Replicated Field Trials

Stage 3 Regional Testing



# Is the Pipeline Cultivar Development Model Fundamentally Flawed?



Gaynor et al 2017. *Crop Science*  
Goranc et al, 2018. TAG

# RGS Encounters the Trade-offs of Recurrent Phenotypic Selection Faster

---

Recurrent selection will encounter limits to genetic gain that are less than the full genetic potential available in the founders (Robertson, 1960; Hill and Robertson, 1968).

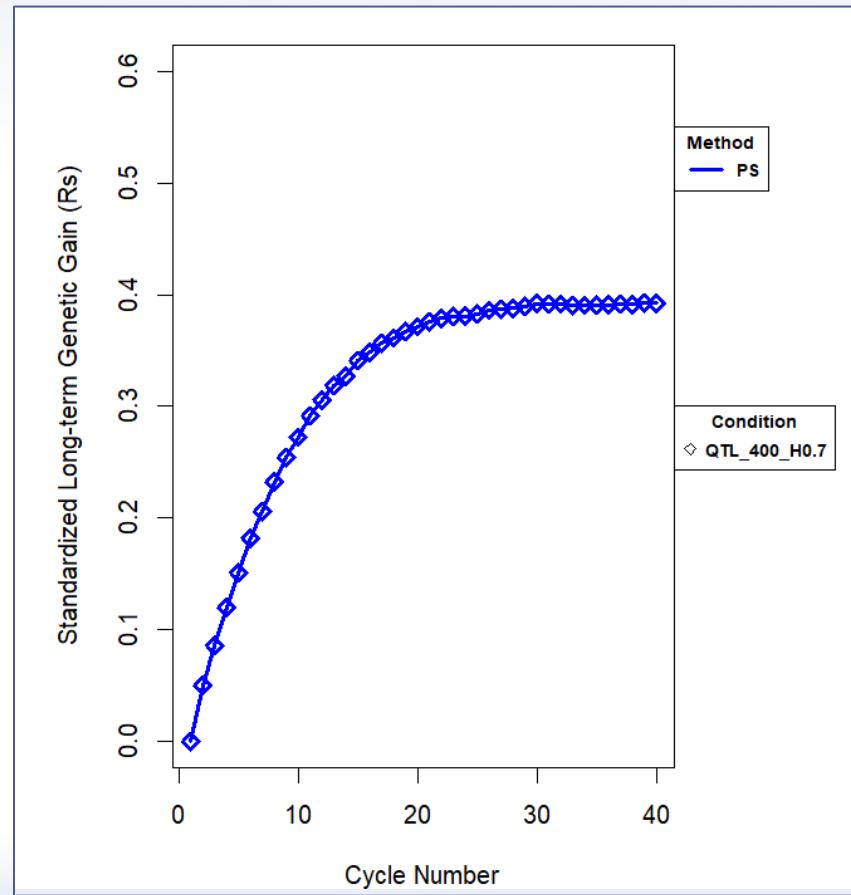
What is the appropriate balance between maintaining genetic diversity (minimizing inbreeding) and maximizing response to selection?

# 40 Cycles of Simulated Phenotypic Recurrent Selection

$$R_s = R_t / (R_m - R_0)$$

$R_s$  - Standardized long-term genetic gain  
 $R_t$  - Average genotypic value in  $G_t$  –  $R_0$   
 $R_m$  - Maximal possible genotypic value for the genetic model  
 $R_0$  - Average genotypic value in  $G_0$

21 Soybean Founders adapted to MZ III  
400 QTL ,  $Vg/Vp = 0.7$ ,  
10% Selected

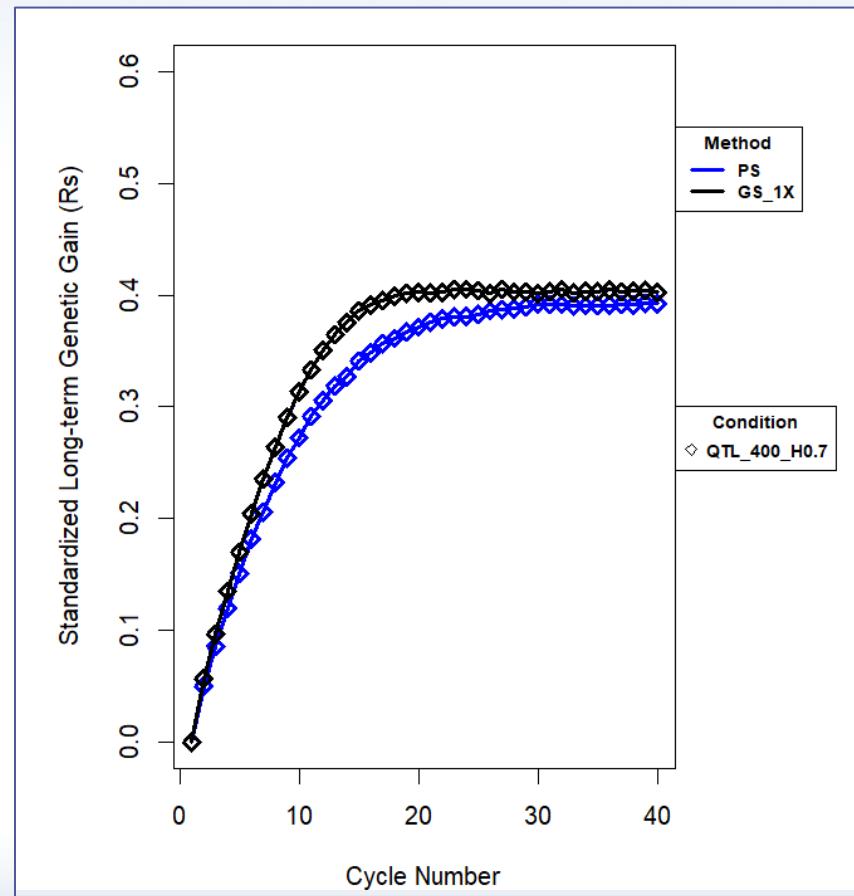


# 40 Cycles of Simulated Recurrent Genomic Selection

$$R_s = R_t / (R_m - R_0)$$

$R_s$  - Standardized long-term genetic gain  
 $R_t$  - Average genotypic value in  $G_t$  –  $R_0$   
 $R_m$  - Maximal possible genotypic value for the genetic model  
 $R_0$  - Average genotypic value in  $G_0$

21 Soybean Founders adapted to MZ III  
400 QTL , Vg/Vp = 0.7,  
10% Selected

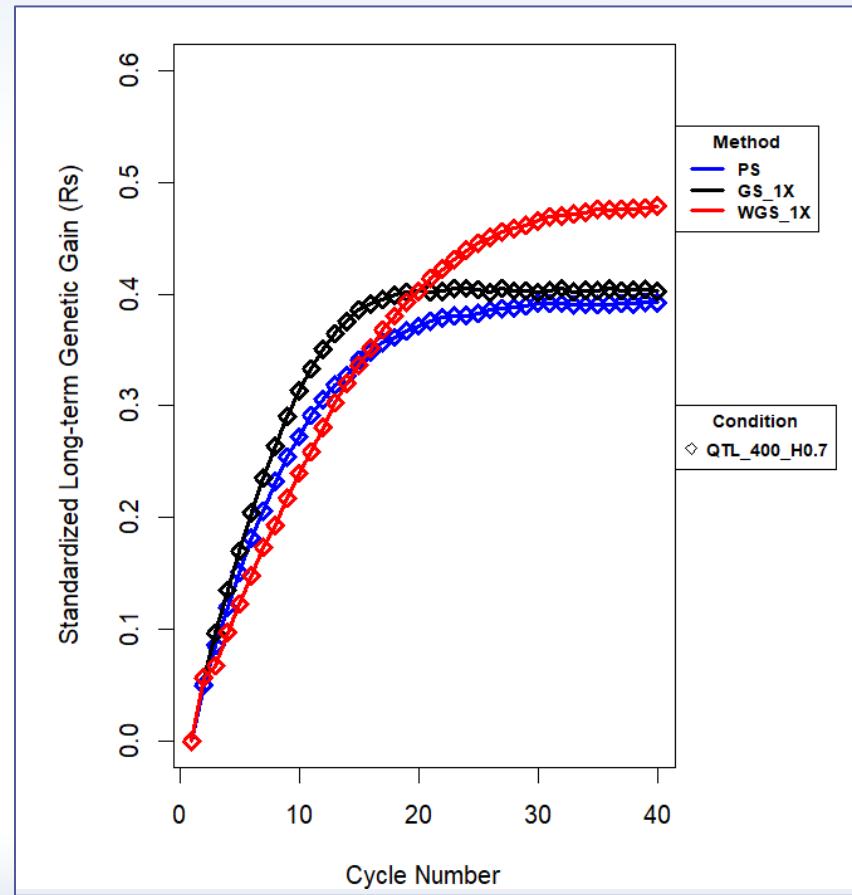


# 40 Cycles of Simulated Recurrent Weighted Genomic Selection

$$R_s = R_t / (R_m - R_0)$$

$R_s$  - Standardized long-term genetic gain  
 $R_t$  - Average genotypic value in  $G_t$  –  $R_0$   
 $R_m$  - Maximal possible genotypic value for the genetic model  
 $R_0$  - Average genotypic value in  $G_0$

21 Soybean Founders adapted to MZ III  
400 QTL , Vg/Vp = 0.7,  
10% Selected

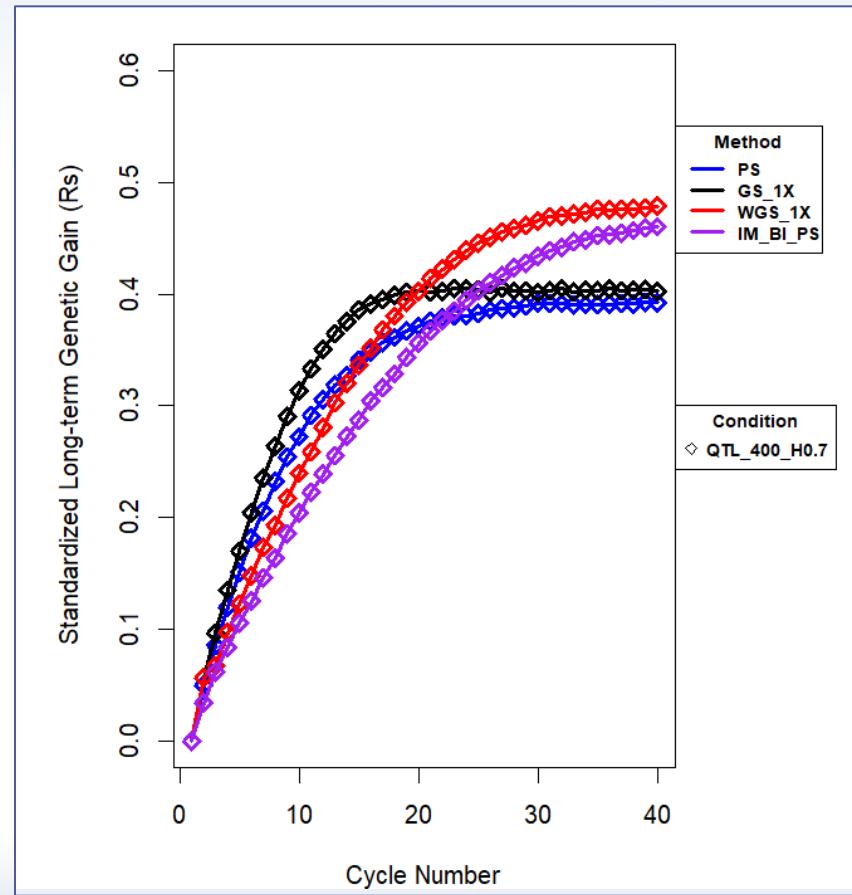


# 40 Cycles of Simulated Recurrent Island Model Phenotypic Selection

$$R_s = R_t / (R_m - R_0)$$

$R_s$  - Standardized long-term genetic gain  
 $R_t$  - Average genotypic value in  $G_t$  –  $R_0$   
 $R_m$  - Maximal possible genotypic value for the genetic model  
 $R_0$  - Average genotypic value in  $G_0$

21 Soybean Founders adapted to MZ III  
400 QTL , Vg/Vp = 0.7,  
10% Selected

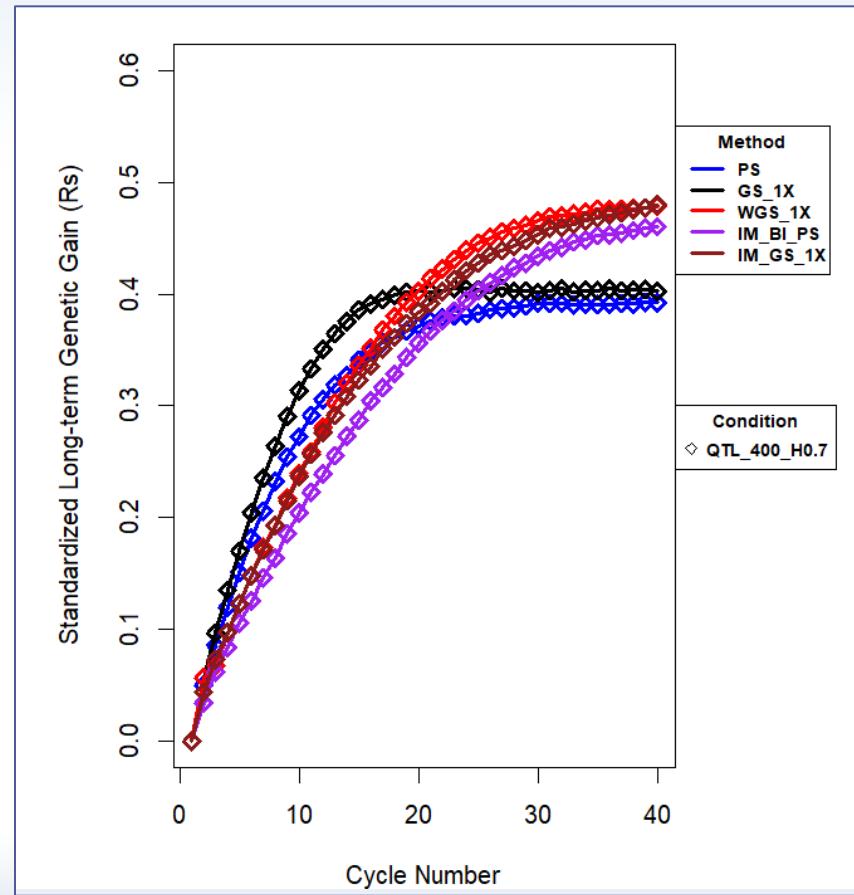


# 40 Cycles of Simulated Recurrent Island Model Genomic Selection

$$R_s = R_t / (R_m - R_0)$$

$R_s$  - Standardized long-term genetic gain  
 $R_t$  - Average genotypic value in  $G_t$  –  $R_0$   
 $R_m$  - Maximal possible genotypic value for the genetic model  
 $R_0$  - Average genotypic value in  $G_0$

21 Soybean Founders adapted to MZ III  
400 QTL , Vg/Vp = 0.7,  
10% Selected

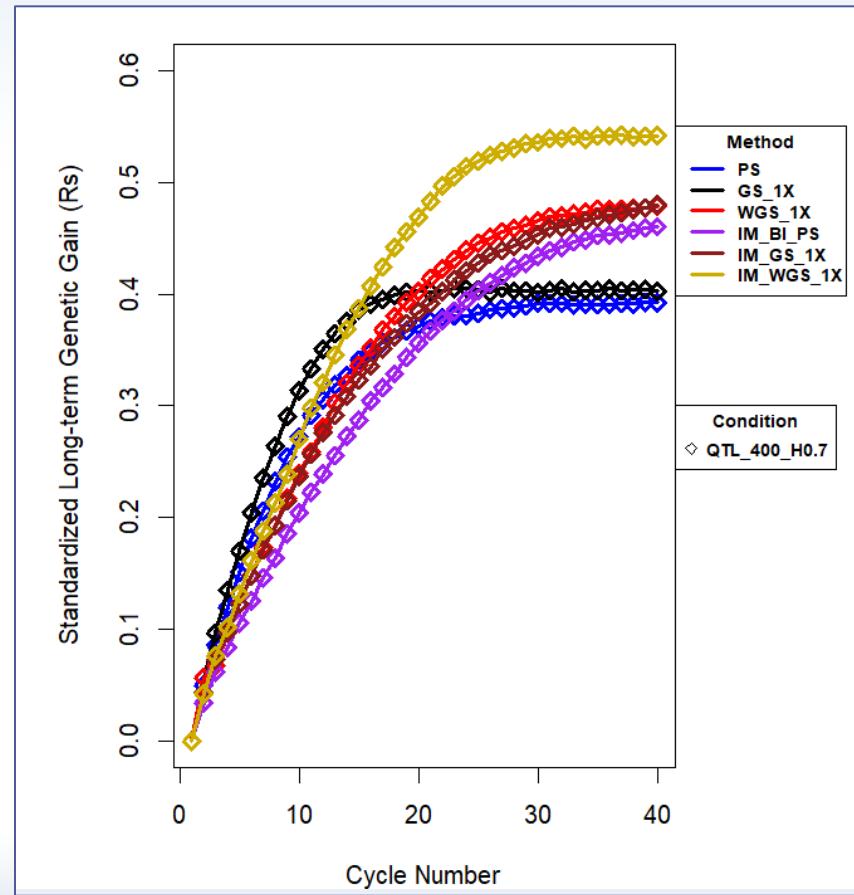


# 40 Cycles of Simulated Recurrent Island Model Weighted Genomic Selection

$$R_s = R_t / (R_m - R_0)$$

$R_s$  - Standardized long-term genetic gain  
 $R_t$  - Average genotypic value in  $G_t$  –  $R_0$   
 $R_m$  - Maximal possible genotypic value for the genetic model  
 $R_0$  - Average genotypic value in  $G_0$

21 Soybean Founders adapted to MZ III  
400 QTL , Vg/Vp = 0.7,  
10% Selected



# A More Difficult Challenge for RGS

---

Maximize genetic gain for multiple traits while minimizing inbreeding

Historical plant breeding approaches:

- Tandem Selection
- Independent Culling (Hazel, 1943)
- Multi-stage Selection
- Selection Index (Hazel and Lush, 1942)

Applied mathematics approach:

- Translate  $k$  traits into  $k$  objective functions
- Simultaneously optimize  $k+1$  objective functions

# Multi-trait Selection: Selection Index

---

$$H_i = \sum_{i=1}^n w_i G_i,$$

$$G_i = \sum_{l=1}^Q \alpha_l;$$

$$I_i = \sum_{i=1}^n b_i P_i,$$

$$P_i = G_i + E_i;$$

$$\mathbf{b} = \boldsymbol{\sigma}_P^{-1} \boldsymbol{\sigma}_G \mathbf{w}$$

Optimum Selection Index:

$$\text{Max}(\rho(I, H))$$

$$\text{Min}(E(I - H)^2)$$

# Translate $k$ Traits into $k$ Objective Functions

---

$$\max G_1,$$

$$\max G_2, \dots$$

$$\max G_k$$

s.t.

$$G_1 = z_1(x)$$

$$G_2 = z_2(x) \dots$$

$$G_k = z_k(x)$$

$$0 \leq g_i(x), i = 1 \dots m$$

$$0 = h_j(x), j = 1 \dots p$$

*find  $\mathbf{x}^*$  to optimize  $\mathbf{F}(\mathbf{x}^*) = [z_l(\mathbf{x}^*)]_{l=1}^k$*

*s.t.*

$$\mathbf{x} = \{x_1, x_2, \dots, x_n\}$$

$$0 \leq g_i(\mathbf{x}^*), i = 1 \dots m$$

$$0 = h_j(\mathbf{x}^*), j = 1 \dots p$$

# Approaches to Simultaneously Optimize $k$ Objective Functions

---

- Scalar methods, e.g., weighted sum
- $\infty$  constraint method
- Identify Pareto optimal solution sets and their associated Pareto front

# Weighted Sum Scalar Method

$$\max G_1,$$

$$\max G_2, \dots$$

$$\max G_k$$

s.t.

$$G_1 = z_1(x)$$

$$G_2 = z_2(x) \dots$$

$$G_k = z_k(x)$$

$$0 \leq g_i(x), i = 1 \dots m$$

$$0 = h_j(x), j = 1 \dots p$$

$$z = w_1 z'_1(x) + w_2 z'_2(x) + \dots + w_k z'_k(x)$$

$$\max z$$

If KKT conditions exist, then a best solution exists.

# Identify Pareto Optimal Sets

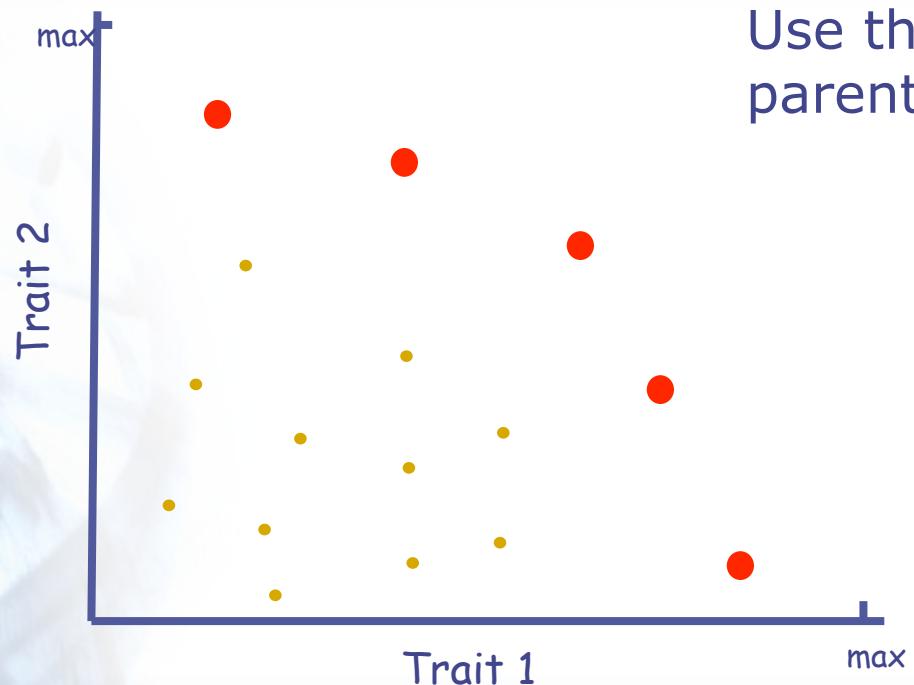
A Pareto optimal set consists of solutions that are non-dominated with respect to each other.



More formally for a set of maximization objectives:

$$(\mathbf{x} \neq \mathbf{y}) \text{ iff } \forall j (\mathbf{x}_j > \mathbf{y}_j)$$

# Genetic Contributions to the Next Generation



Use the non-dominated set as parents for the next generation.

# Genetic Contributions to the Next Generation



Parental contributions  
based on distance from a  
feasible ideal

# Contributions for Negative Correlated Traits and Minimize Inbreeding

$$\max \mathbf{F}(\mathbf{c}) = [z_l(\mathbf{c})]_{l=1}^{k+1}$$

s.t.

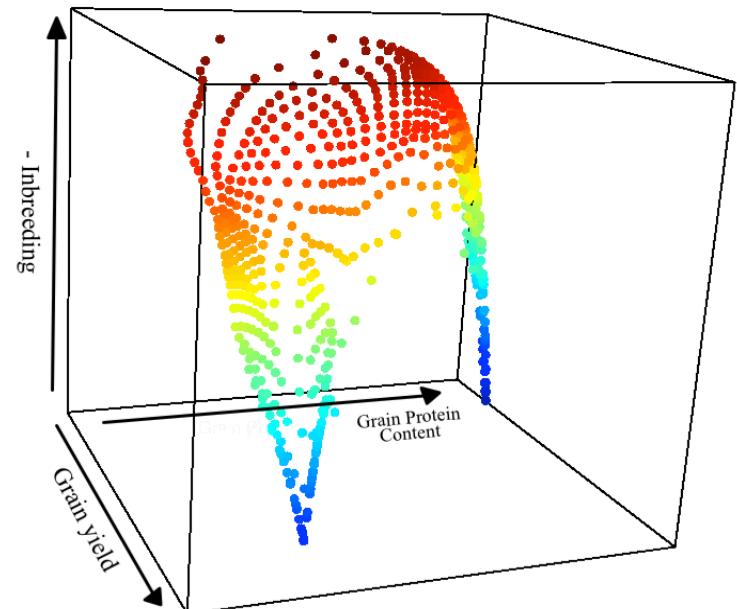
$$z_l(\mathbf{c}) = \mathbf{c}' \mathbf{g}_l \quad \forall l = 1, 2, \dots, k$$

$$\mathbf{g}_l = \mathbf{c}' \mathbf{v}_l$$

$$z_{k+1}(\mathbf{c}) = -\frac{1}{2} \mathbf{c}' \boldsymbol{\psi}' \mathbf{A} \mathbf{c}$$

$$c_i \geq 0, i = 1, \dots, m$$

$$\sum_{i=1}^m c_i = 1$$

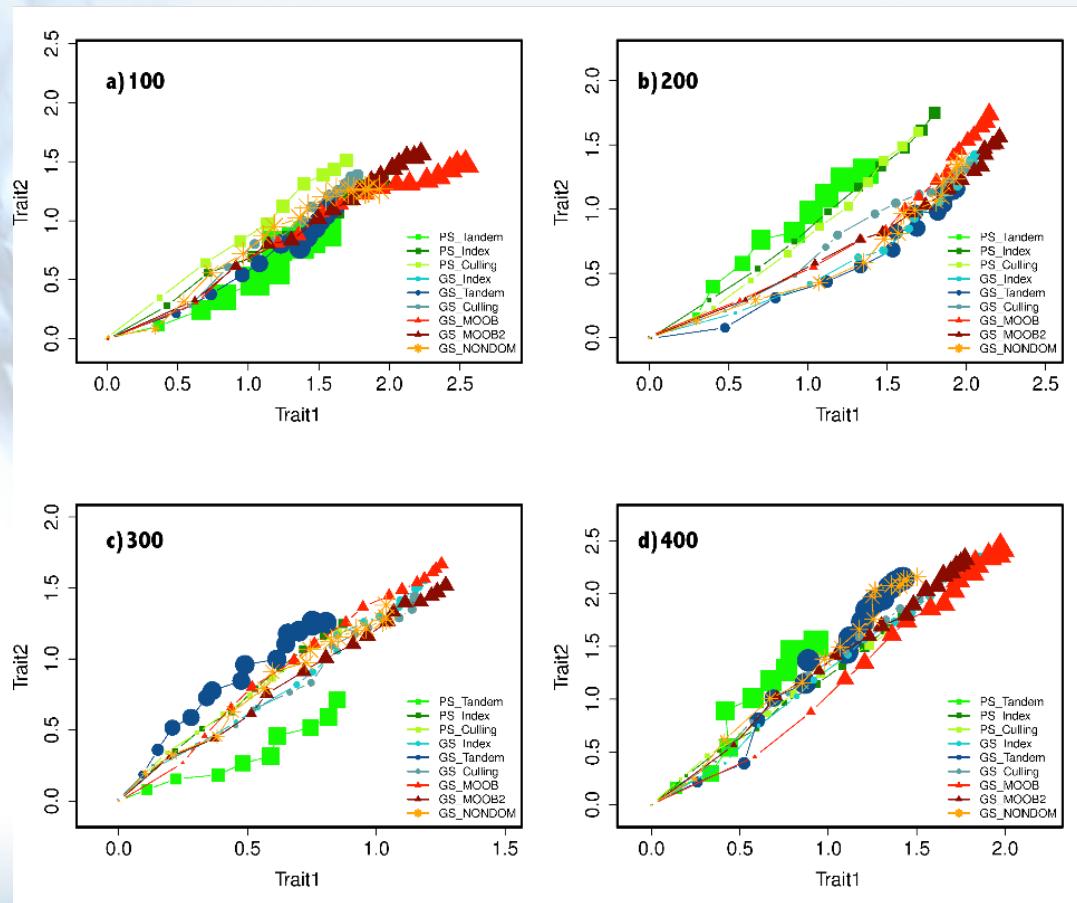


# Simulated Comparison of Multi-trait Selection Approaches

---

- 3000 marker alleles with initial  $f(a)=0.5$  in diploid individuals
- Random mate (100, 200, 300, 400) individuals for 100 generations. Tandem PS for 10 generations
- $k=2$  traits,  $h^2=0.5$ ,  $\rho(t_1, t_2) = -0.5$ 
  - 600 uniformly distributed QTL
  - $a_Q=+/- (1/600)$
- 30 replicates
- Breeding objectives:
  - Max  $T_1$ , Max  $T_2$ , Max –Inbreeding
- Selection strategies:
  - PS(SI), PS(Tandem), PS(Cull), GS(SI), GS(SI), GS(Tandem), GS(Cull), GS(NonD), OMOD<sub>1</sub>, OMOD<sub>1</sub>  $\Psi$
  - $i=0.05$  ( $N=100, 200$ ),  $i=0.10$  ( $N=300, 400$ )

# Simulated Responses to 16 Cycles of Recurrent Selection



$$1.2 \times GS(SI) < OMOD_1$$

# The Most Difficult Challenge for RGS (For all Plant Breeding Systems)

---

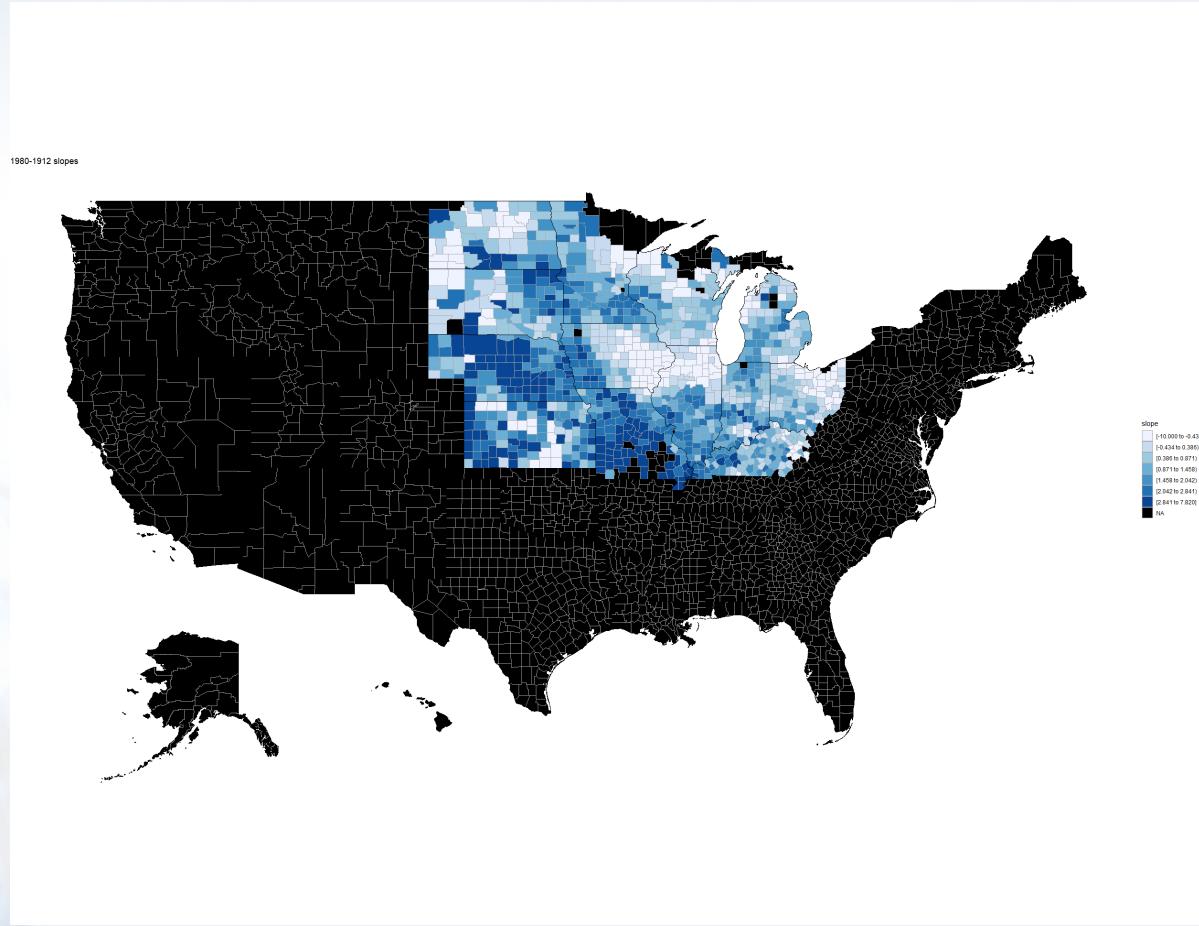
Select multiple traits needed for adaptation to emerging future environments.

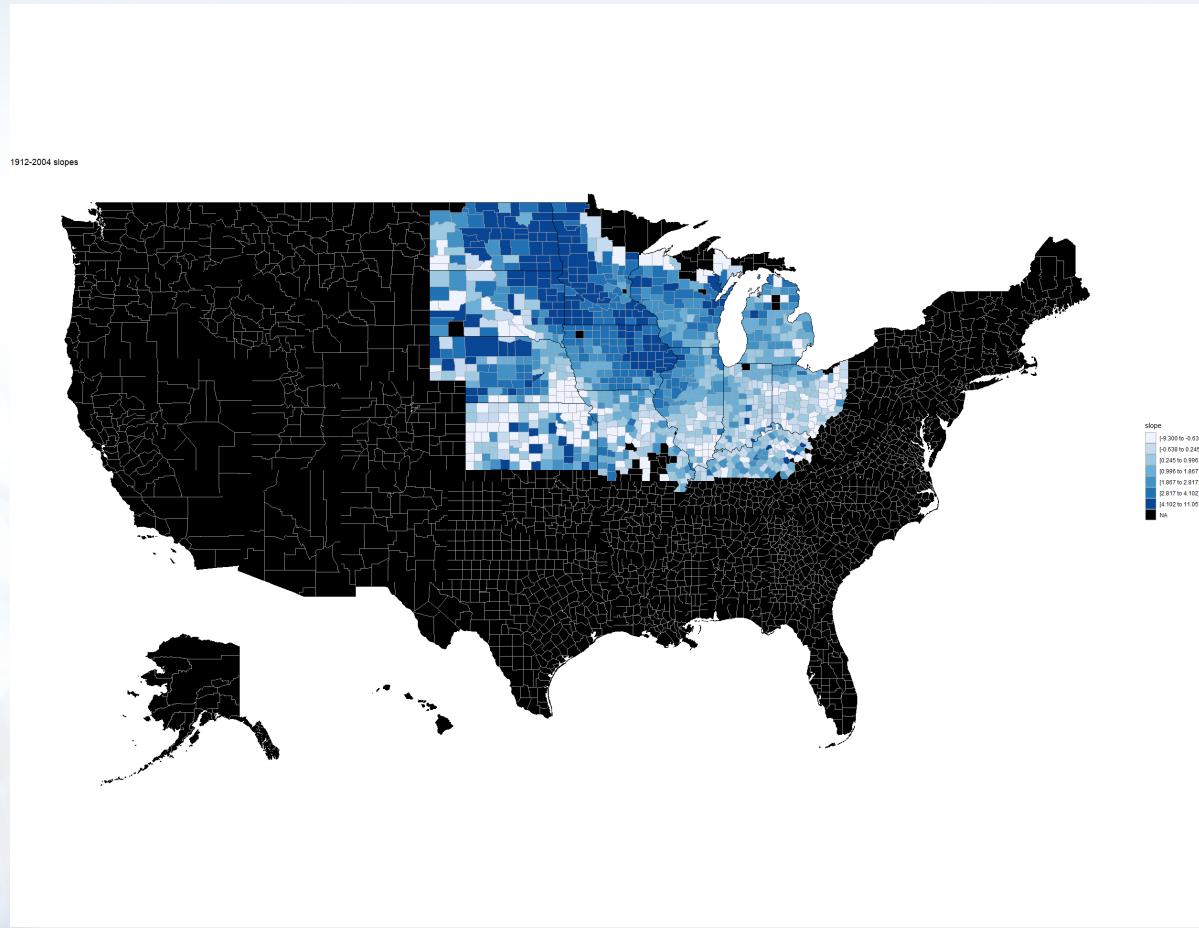
- What are the most likely distributions of future environments?
- Are we measuring the right traits for adaptation to future environments?
  - Crop Modeling
- What product profile will provide a successful future cultivar?
- Do we know the genetics underlying adaptation traits?
- Do we have alleles that will enable adaptation to predicted environments?

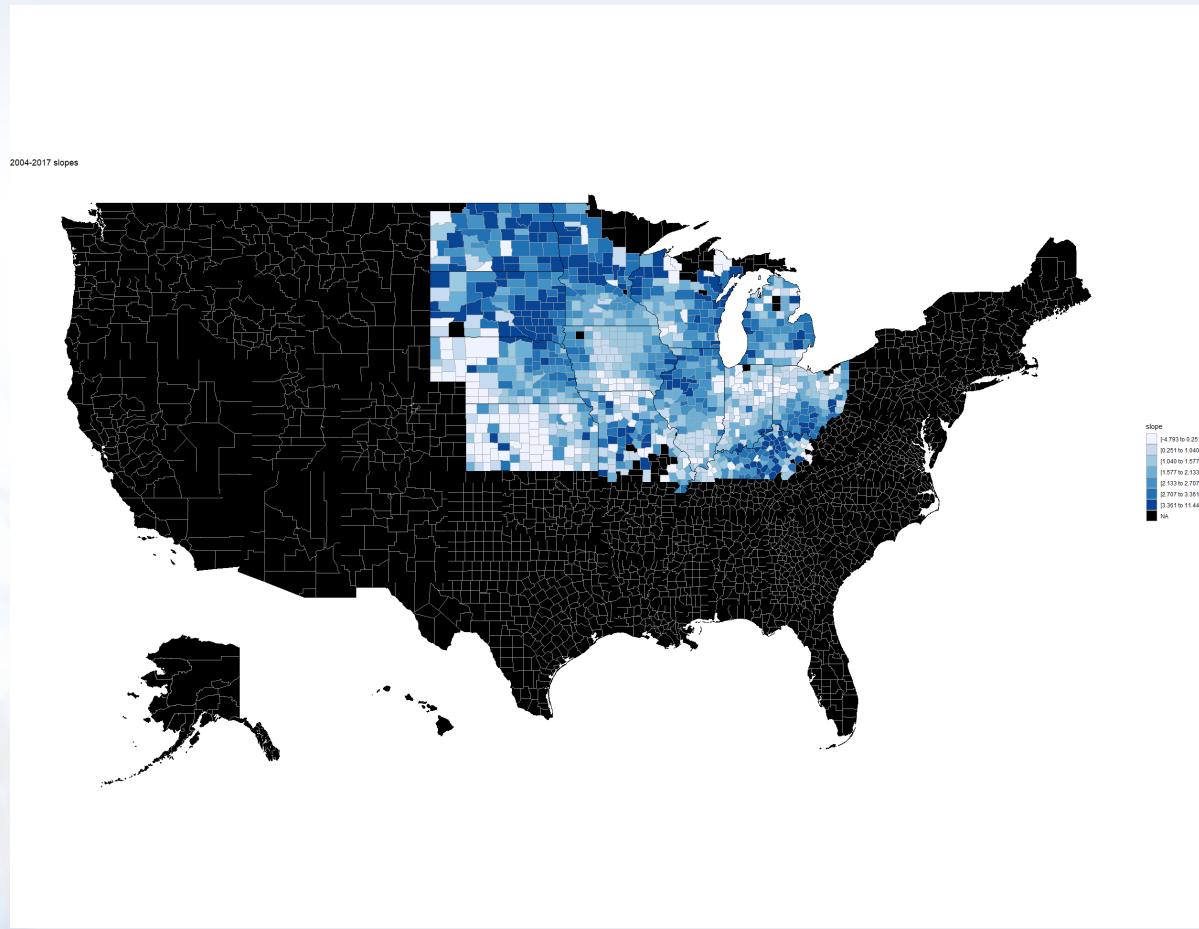
# Summary:

---

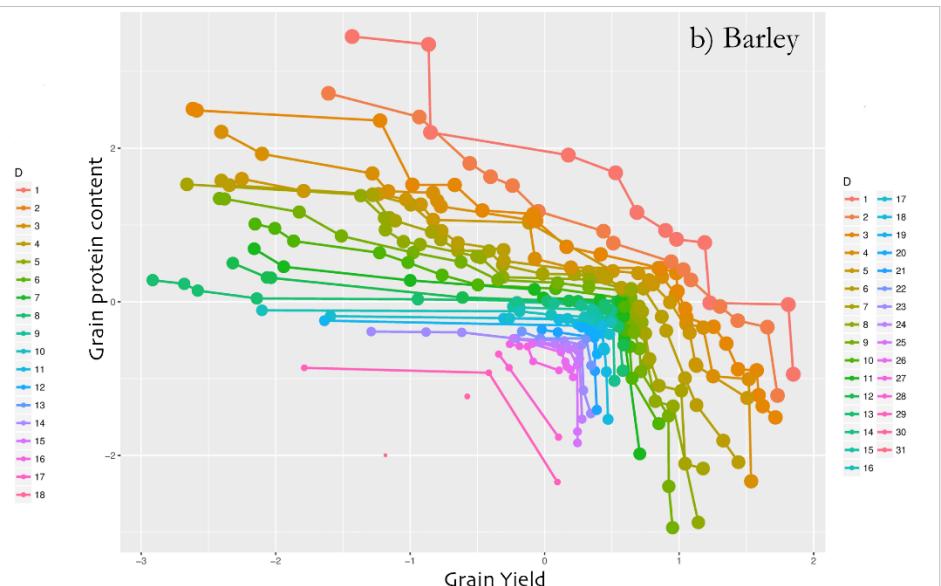
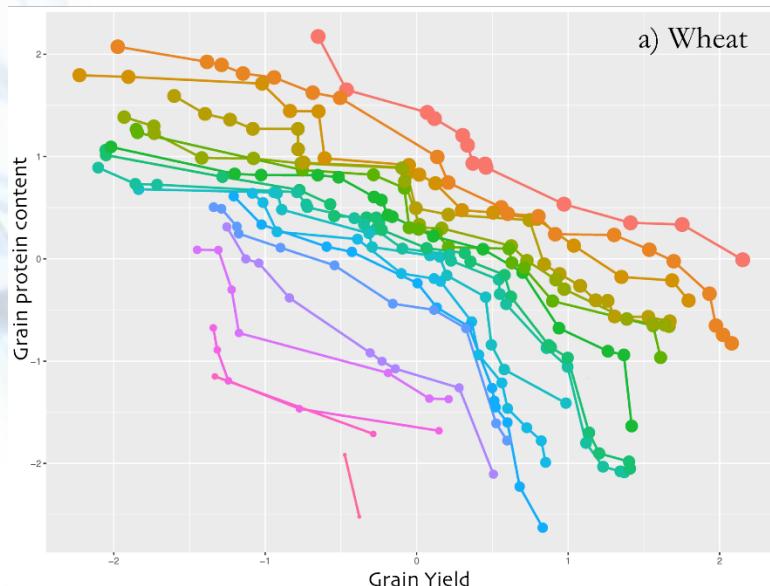
- If we are going to reclaim our birthright as designers of plant breeding systems, then we will need to learn to think like an engineer.
  - Clearly define measurable objectives: What product profile will be recognized as a successful outcome?
  - Learn how to quantify and communicate trade-offs among multiple objectives.
  - Pressure to produce adapted high quality cultivars as fast as possible, for minimal cost. We must communicate that two of the three are possible.







# Pareto Front for yield and protein in wheat and barley



# What are an Optimal Plant Breeding Systems?

---

**Optimization:** A process or methodology of making a design, system or decision as effective and efficient as possible, specifically through mathematical procedures. – Merriam Webster.

# Parental Contributions and Minimize Inbreeding

$$\max \mathbf{F}(\mathbf{c}) = [z_l(\mathbf{c})]_{l=1}^{k+1}$$

s.t.

$$z_l(\mathbf{c}) = \mathbf{c}' \mathbf{g}_l \quad \forall l = 1, 2, \dots, k$$

$$\mathbf{g}_l = \mathbf{c}' \mathbf{v}_l$$

$$z_{k+1}(\mathbf{c}) = -\frac{1}{2} \mathbf{c}' \mathbf{A} \mathbf{c}$$

$$c_i \geq 0, i = 1, \dots, m$$

$$\sum_{i=1}^m c_i = 1$$

