Leveraging technology to deploy products and prepare students at Virginia Tech for a new era of data-driven plant breeding

Nicholas Santantonio

Cornell University

March 9th, 2020

About Me: New Mexico State University – Alfalfa (*Medicago sativa*)

BS Genetics 2010
MS Plant Sci. 2013

Ian Ray (Alfalfa)

- Productivity under drought
  - Shoot biomass QTL
    (Ray et al. 2015, Crop Sci)
- Water use efficiency
  - Carbon Isotope Discrimination QTL
    (Santantonio et al. 2018, Crop Sci)
  - MAS at *ERECTA* locus

Cornell University

Mark Sorrells (Small Grains)
PhD Plant Breeding 2018

- G×E
  - `Bilinear` R software package
    - available at github.com/nsantantonio
  - Fructans (Veenstra et al., 2018)
  - Organic wheat (Kucek et al., 2018)
- Genomic Prediction in Allopolyploids
  - Subgenome interactions (Santantonio et al., 2019a)
  - Homeologous epistasis (Santantonio et al., 2019b)
  - Regional epistasis mapping (Santantonio et al., 2019c)

Cornell University

Mark Sorrells (Small Grains)
PhD Plant Breeding 2018

- G×E
  - `Bilinear` R software package
    - available at github.com/nsantantonio
  - Fructans (Veenstra et al., 2018)
  - Organic wheat (Kucek et al., 2018)
- Genomic Prediction in Allopolyploids
  - Subgenome interactions (Santantonio et al., 2019a)
  - Homeologous epistasis (Santantonio et al., 2019b)
  - Regional epistasis mapping (Santantonio et al., 2019c)

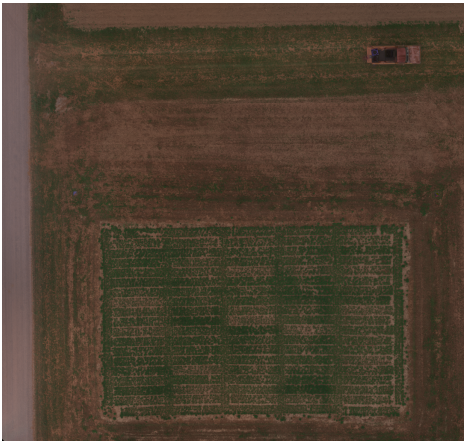
$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{ZQ}\boldsymbol{\gamma} + \sum_l \mathbf{Z}\mathbf{g}_l + \boldsymbol{\varepsilon}$$

$$\text{GEBV} = (\mathbf{I}_n - \bar{\mathbf{J}}_n)(\mathbf{Q}\hat{\boldsymbol{\gamma}}) + \sum_l \hat{\mathbf{g}}_l$$

Cornell University

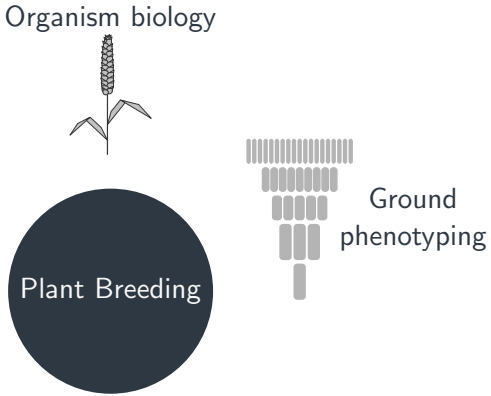
Kelly Robbins (Quantitative Genetics)

- Post Doc, current
  - Crop flexible, equal opportunity
    - Chickpea, maize, alfalfa, simulated
  - Breeding scheme optimization
    - Sparse testing (Santantonio et al. Accepted, FiPS)
    - Optimal contributions (Santantonio et al. In Review, G3)
  - High-throughput Phenotyping
    - Longitudinal models
    - Genotype specific growth curves USAFRI Grant (alfalfa)



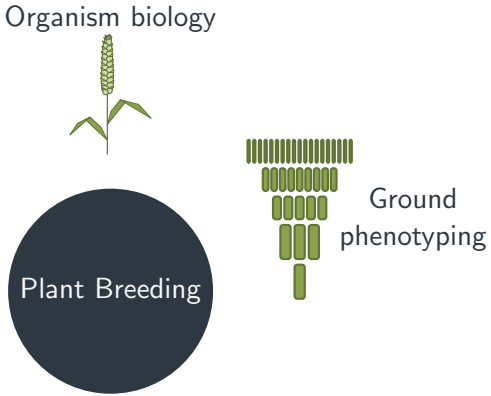
Plant breeding

- Multi-disciplinary
- Team oriented



Plant breeding

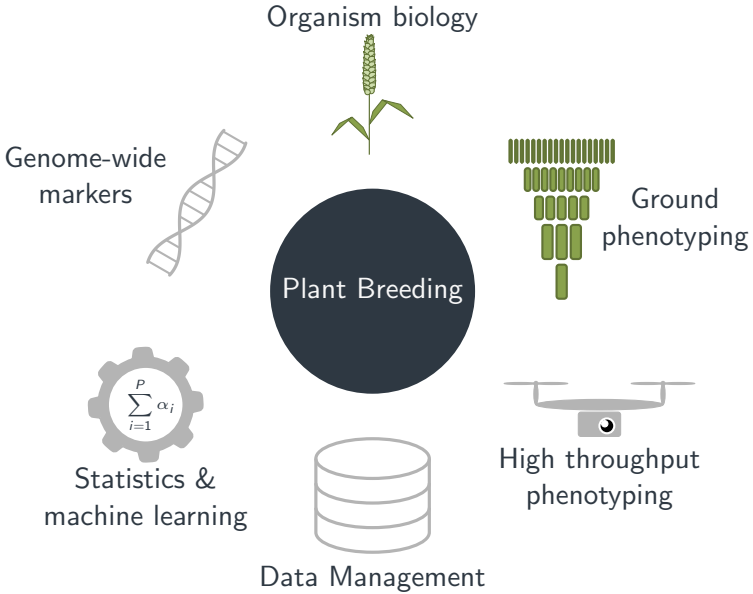
- Multi-disciplinary
- Team oriented



Plant breeding

- Multi-disciplinary
- Team oriented

Era of "Big data"

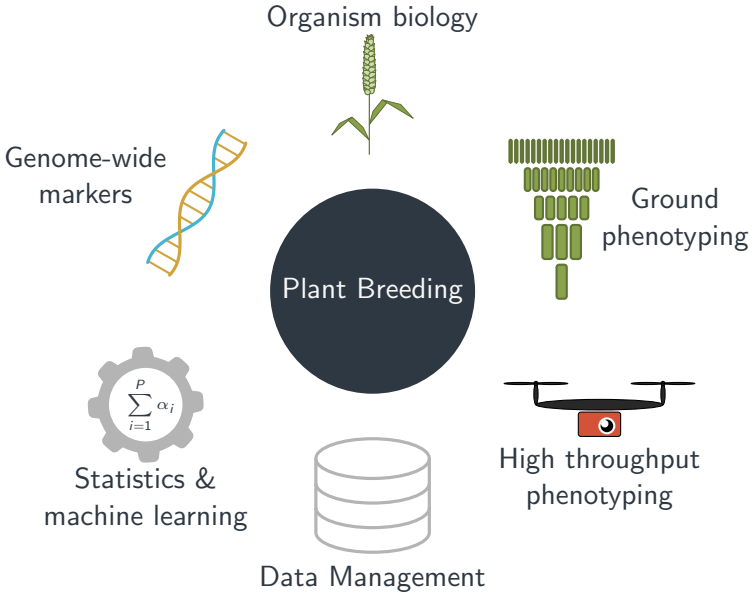


Plant breeding

- Multi-disciplinary
- Team oriented

Era of "Big data"

- Information rich

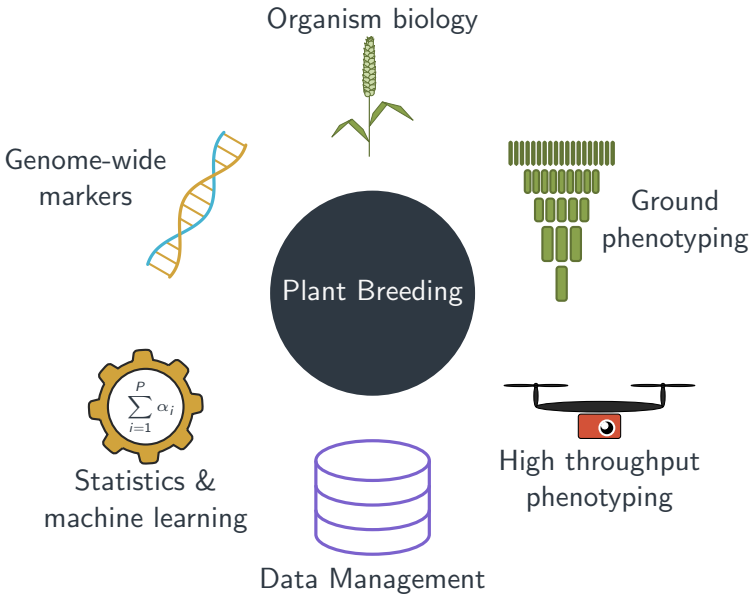


Plant breeding

- Multi-disciplinary
- Team oriented

Era of "Big data"

- Information rich
- Store, access and process



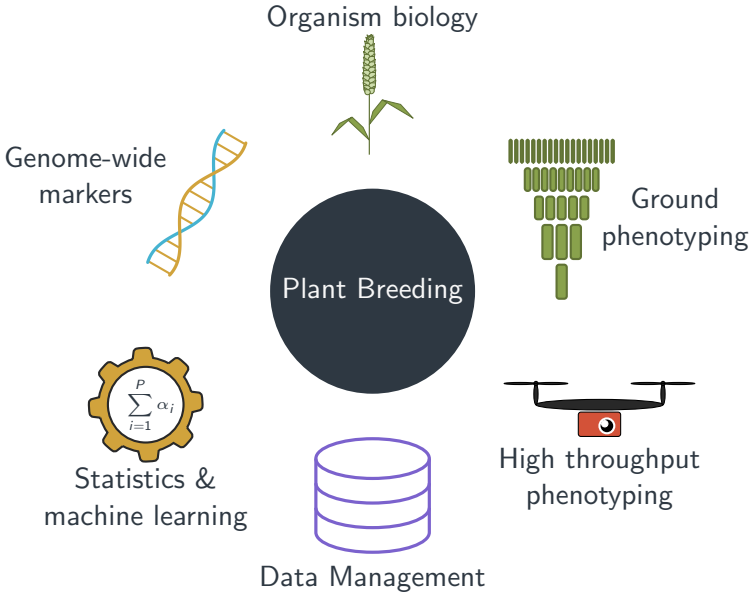
Plant breeding

- Multi-disciplinary
- Team oriented

Era of "Big data"

- Information rich
- Store, access and process

How can we leverage this data to deploy better products to farmers?



Greatest gains in wheat during the 20th Century

Much of the low hanging fruit has been picked

Disease and insect resistances

- Leaf and stem rust, Scab (?), Hessian fly, etc.

Green Revolution

- Improve yield under high N

Greatest gains in wheat during the 20th Century

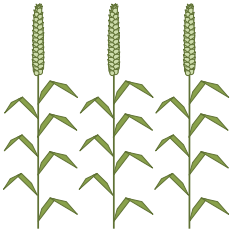
Much of the low hanging fruit has been picked

Disease and insect resistances

- Leaf and stem rust, Scab (?), Hessian fly, etc.

Green Revolution

- Improve yield under high N



Greatest gains in wheat during the 20th Century

Much of the low hanging fruit has been picked

Disease and insect resistances

- Leaf and stem rust, Scab (?), Hessian fly, etc.

Green Revolution

- Improve yield under high N

Greatest gains in wheat during the 20th Century

Much of the low hanging fruit has been picked

Disease and insect resistances

- Leaf and stem rust, Scab (?), Hessian fly, etc.

Green Revolution

- Improve yield under high N
- Semi-dwarfs: Homeologous Rht-1 dwarfing genes on 4B and 4D

Greatest gains in wheat during the 20th Century

Much of the low hanging fruit has been picked

Disease and insect resistances

- Leaf and stem rust, Scab (?), Hessian fly, etc.

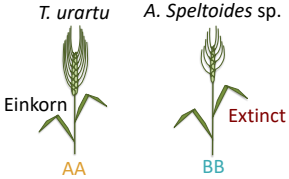
Green Revolution

- Improve yield under high N
- Semi-dwarfs: Homeologous Rht-1 dwarfing genes on 4B and 4D

What other homeologous genes are important?

- *Vrn*, *Ppd*, kernel color
- Can we find and fix beneficial homeologous pairs on a genome-wide scale?

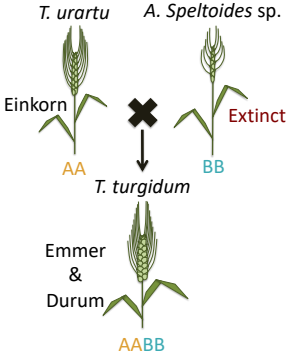
Evolution of allohexaploid wheat



Triticum aestivum

- AA $\times$ BB $\sim$ 0.5 Mya
- AABB $\times$ DD $\sim$ 10,000 ya

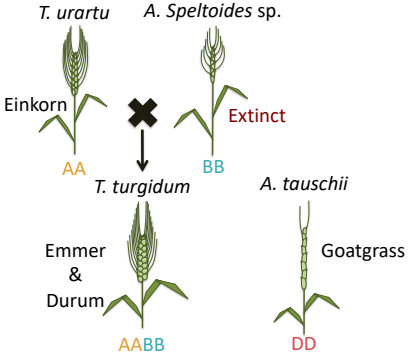
Evolution of allohexaploid wheat



Triticum aestivum

- AA $\times$ BB $\sim$ 0.5 Mya
- AABB $\times$ DD $\sim$ 10,000 ya

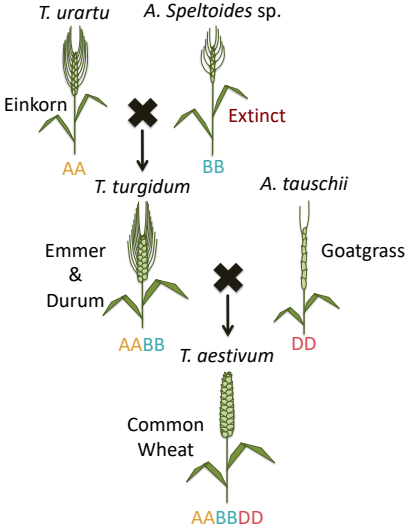
Evolution of allohexaploid wheat



Triticum aestivum

- AA $\times$ BB $\sim$ 0.5 Mya
- AABB $\times$ DD $\sim$ 10,000 ya

Evolution of allohexaploid wheat



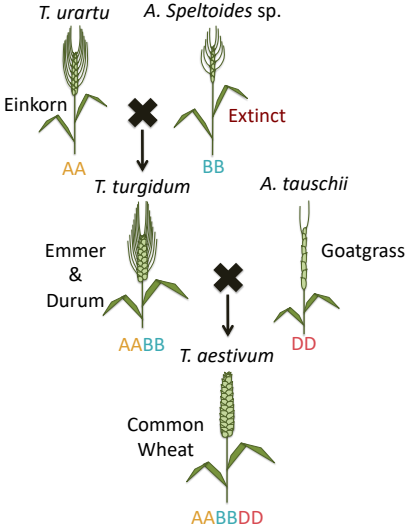
Triticum aestivum

- AA $\times$ BB $\sim$ 0.5 Mya
- AABB $\times$ DD $\sim$ 10,000 ya

Allohexaploid

- Disomic inheritance
- Autogamous
- Allelic diversity preserved across subgenomes

Evolution of allohexaploid wheat



Triticum aestivum

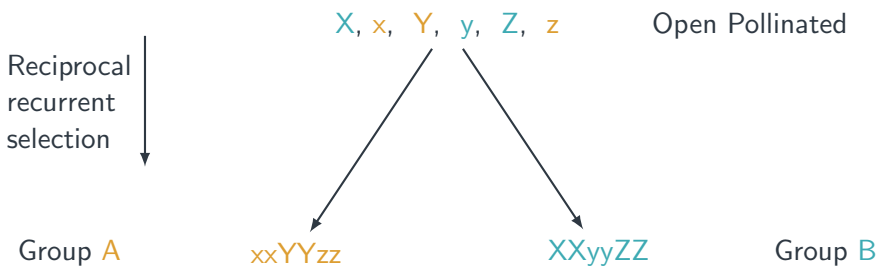
- AA $\times$ BB $\sim$ 0.5 Mya
- AABB $\times$ DD $\sim$ 10,000 ya

Allohexaploid

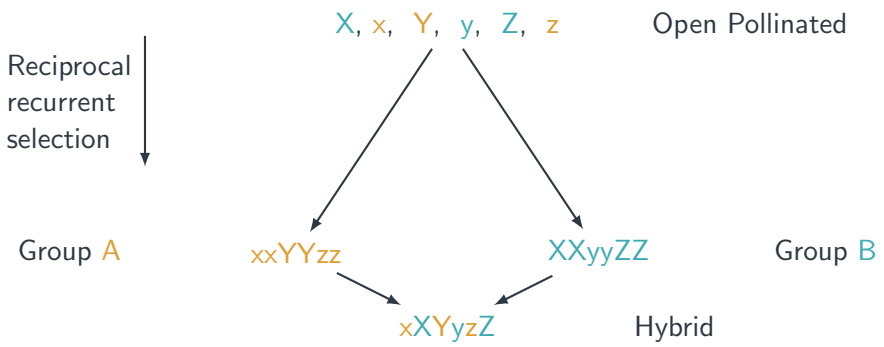
- Disomic inheritance
- Autogamous
- Allelic diversity preserved across subgenomes

Is wheat an immortalized hybrid?

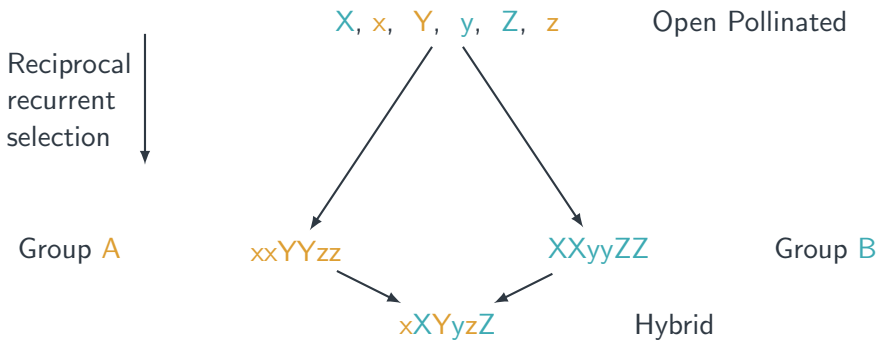
Hybrid generation



Hybrid generation

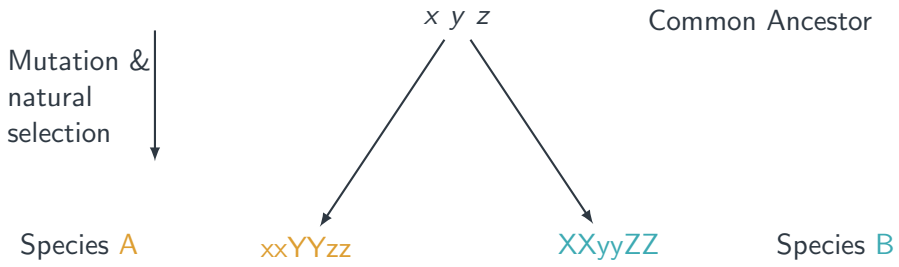


Hybrid generation

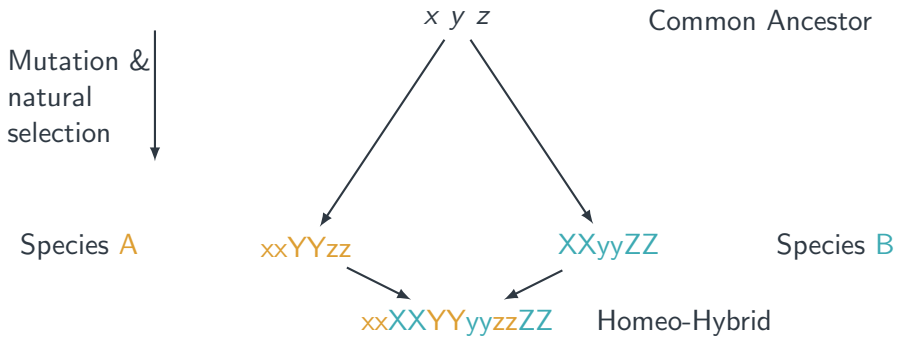


allelic interactions! (dominance)

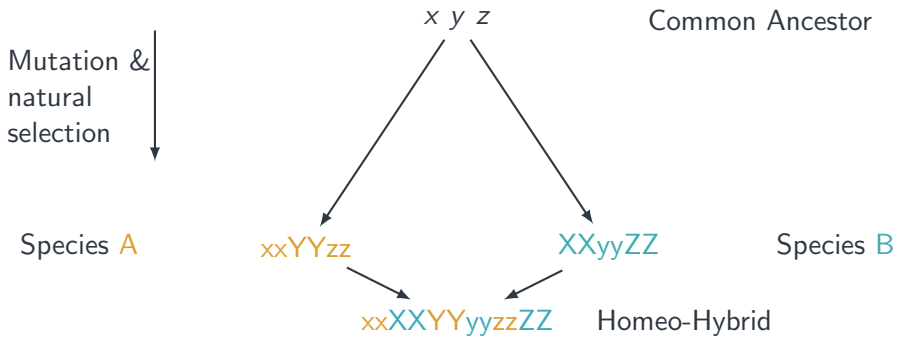
Allopolyploid formation



Allopolyploid formation

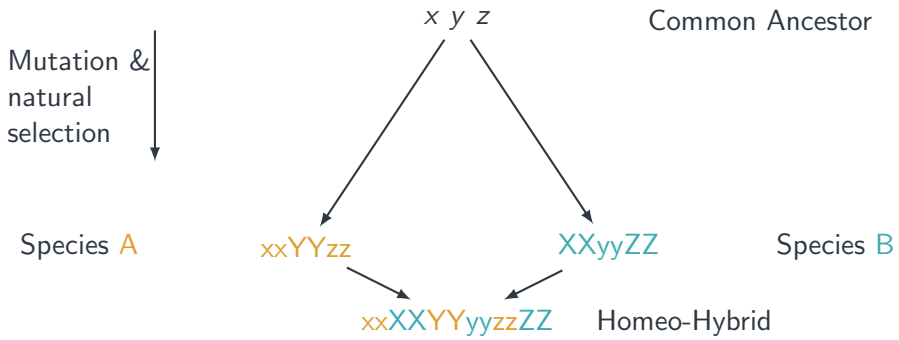


Allopolyploid formation



homeoallelic interactions? (homeologous epistasis)

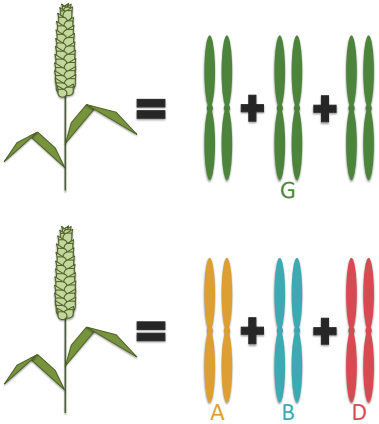
Allopolyploid formation

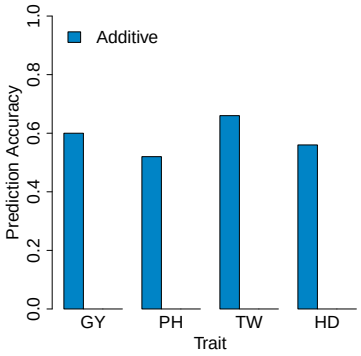


homeoallelic interactions? (homeologous epistasis)

With markers and a genome, we can now ask this question!

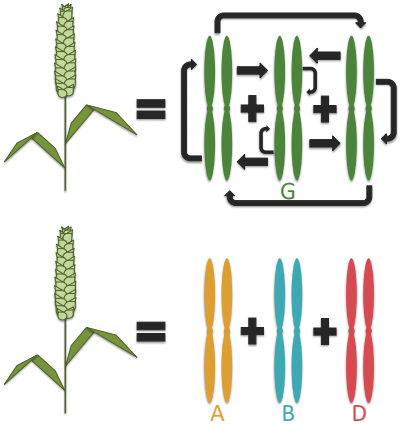
Homeologous interactions explain much of non-additive genetic signal

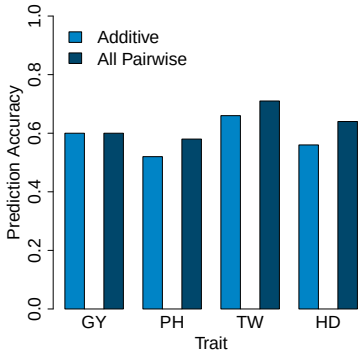




Use genomic prediction to evaluate genetic signal

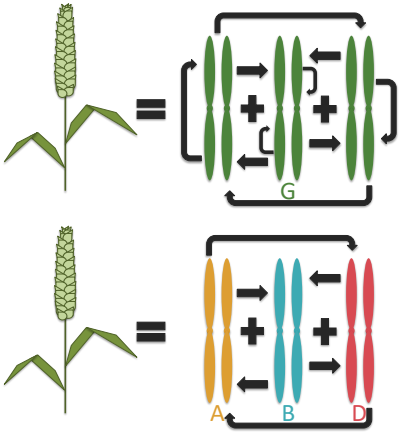
Homeologous interactions explain much of non-additive genetic signal



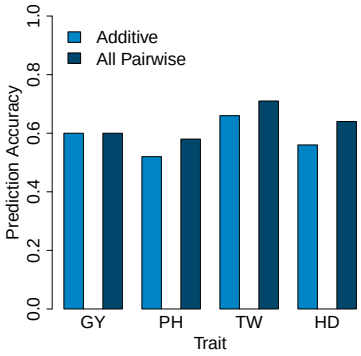


Use genomic prediction to evaluate genetic signal

Homeologous interactions explain much of non-additive genetic signal

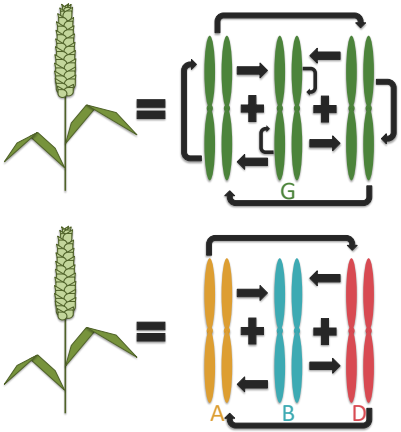


- Tag homeologous gene sets with markers

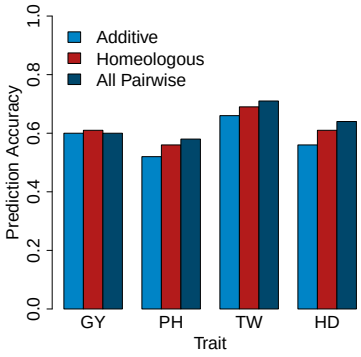


Use genomic prediction to evaluate genetic signal

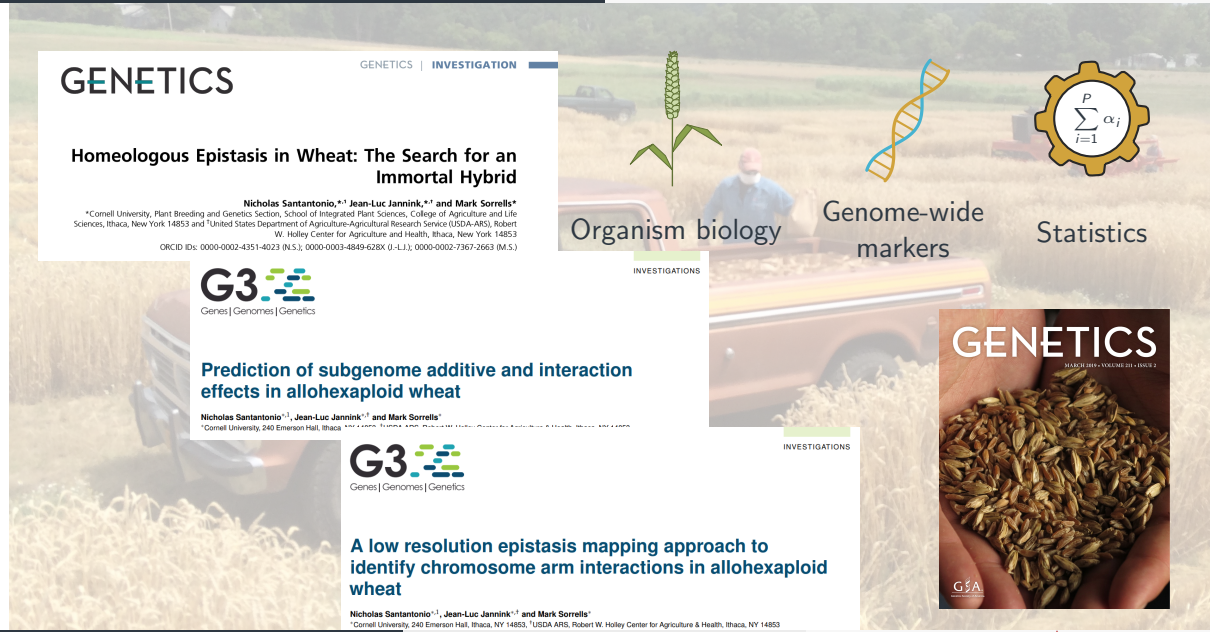
Homeologous interactions explain much of non-additive genetic signal



- Tag homeologous gene sets with markers
- $\sim$ 60-75% non-additive genetic signal
- Can find and fix!



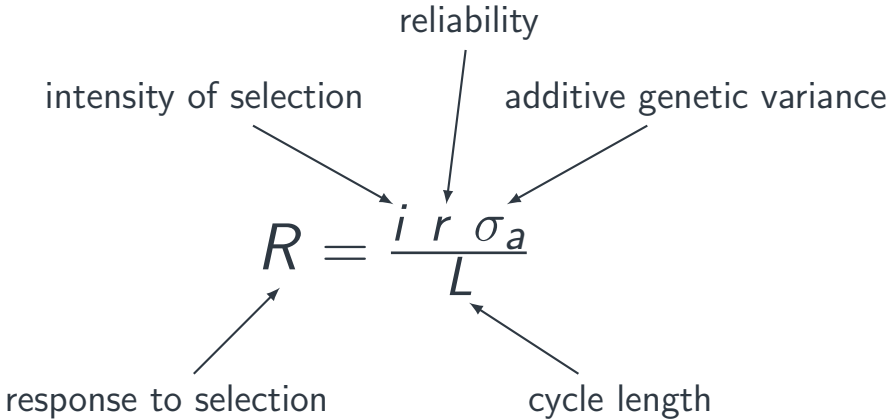
Use genomic prediction to evaluate genetic signal


GENETICS
GENETICS | INVESTIGATION
Homeologous Epistasis in Wheat: The Search for an Immortal Hybrid
Nicholas Santantonio,*,† Jean-Luc Jannink,*,† and Mark Sorrells*
*Cornell University, Plant Breeding and Genetics Section, School of Integrated Plant Sciences, College of Agriculture and Life Sciences, Ithaca, New York 14853 and †United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Robert W. Holley Center for Agriculture and Health, Ithaca, New York 14853
ORCID IDs: 0000-0002-4351-4023 (N.S.); 0000-0003-4849-628X (J.-L.J.); 0000-0002-7367-2663 (M.S.)
Organism biology
Genome-wide markers
Statistics
$\sum_{i=1}^{P} \alpha_i$
G3
Genes | Genomes | Genetics
INVESTIGATIONS
Prediction of subgenome additive and interaction effects in allohexaploid wheat
Nicholas Santantonio*,1, Jean-Luc Jannink*,† and Mark Sorrells*
*Cornell University, 240 Emerson Hall, Ithaca
G3
Genes | Genomes | Genetics
INVESTIGATIONS
A low resolution epistasis mapping approach to identify chromosome arm interactions in allohexaploid wheat
Nicholas Santantonio*,1, Jean-Luc Jannink*,† and Mark Sorrells*
*Cornell University, 240 Emerson Hall, Ithaca, NY 14853, †USDA ARS, Robert W. Holley Center for Agriculture & Health, Ithaca, NY 14853
GENETICS
MARCH 2019 • VOLUME 211 • ISSUE 2
GSA

Let's start with the breeder's equation

$$R = \frac{i\ r\ \sigma_a}{L}$$

Let's start with the breeder's equation



A traditional breeding program

Variety Development Pipeline

10^4, 1 Env, unrep

Year 1

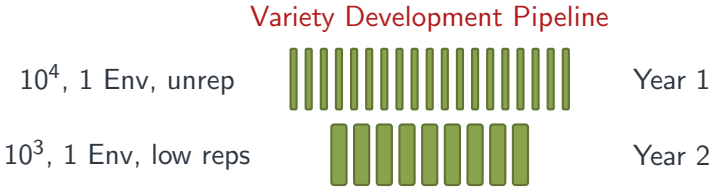
Traditional Breeding

- Accurate (high r)
  - Extensive testing

A traditional breeding program

Traditional Breeding

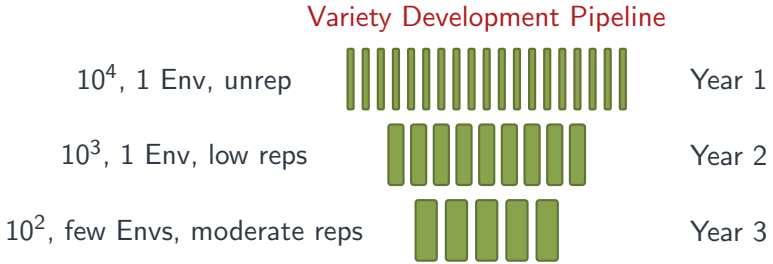
- Accurate (high r)
  - Extensive testing



A traditional breeding program

Traditional Breeding

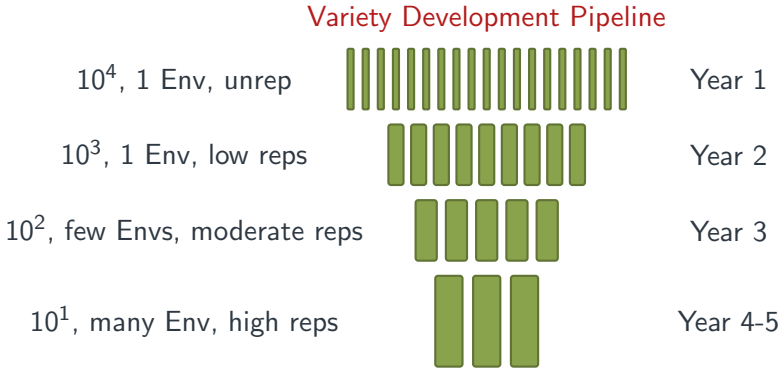
- Accurate (high r)
  - Extensive testing



A traditional breeding program

Traditional Breeding

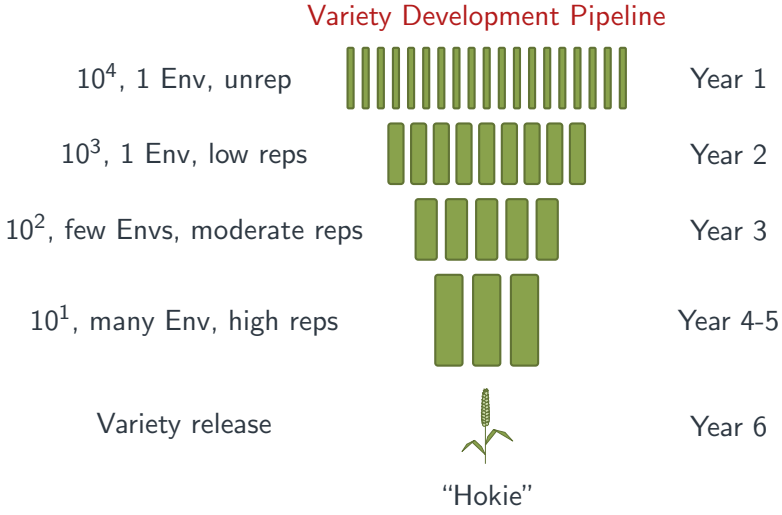
- Accurate (high r)
  - Extensive testing



A traditional breeding program

Traditional Breeding

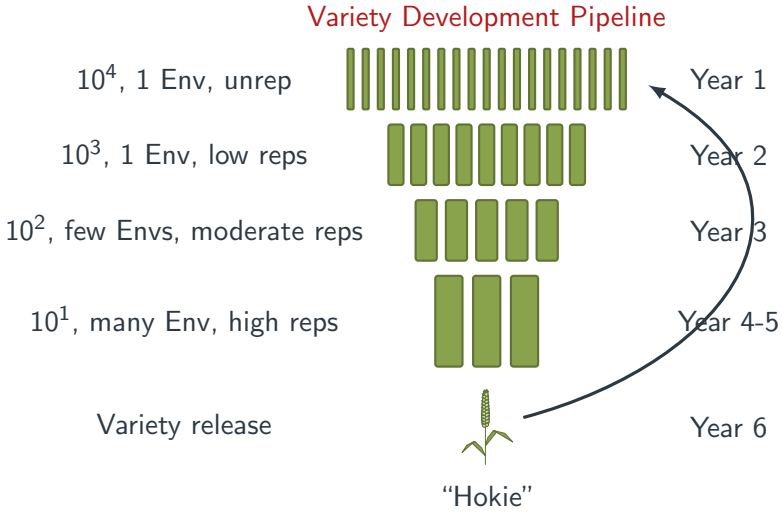
- Accurate (high r)
  - Extensive testing



A traditional breeding program

Traditional Breeding

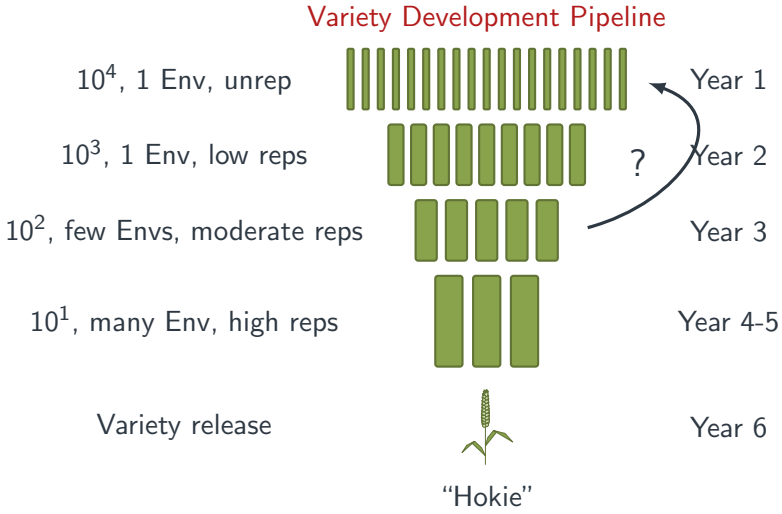
- Accurate (high r)
  - Extensive testing
- Slow ()
  - Multiple trial years
  - Long generation intervals



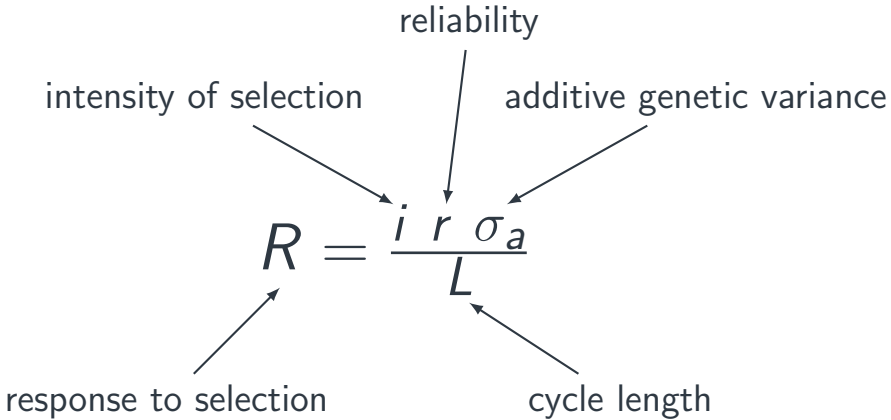
A traditional breeding program

Traditional Breeding

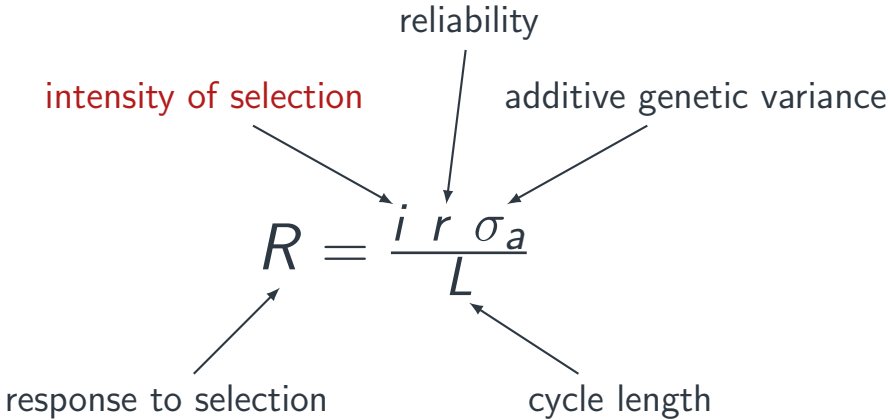
- Accurate (high r)
  - Extensive testing
- Slow ()
  - Multiple trial years
  - Long generation intervals



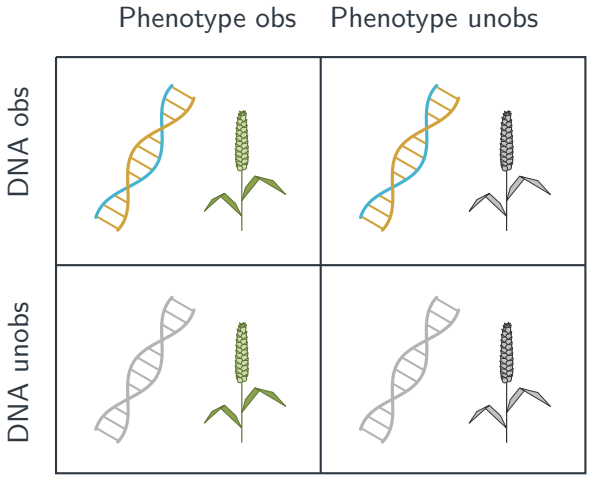
Which terms can we easily exploit?



Which terms can we easily exploit?



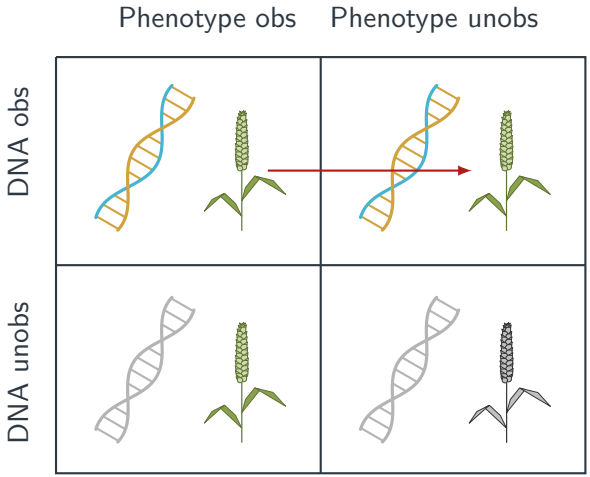
Incorporating Genomic Selection



Genomic prediction

- Estimate all marker effects
- Predict unobserved lines based on marker scores
  - with some accuracy < 1

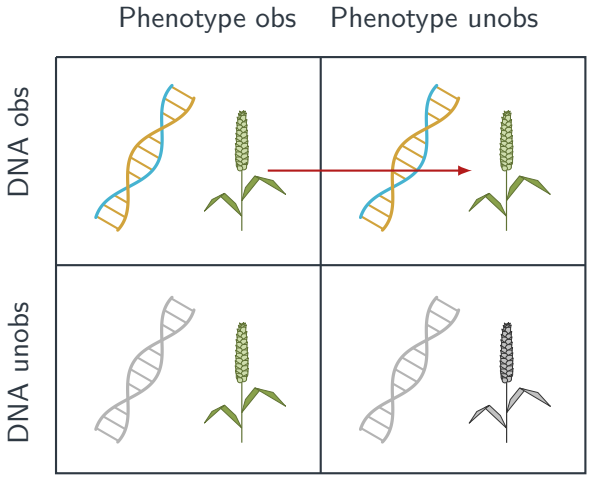
Incorporating Genomic Selection



Genomic prediction

- Estimate all marker effects
- Predict unobserved lines based on marker scores
  - with some accuracy < 1

Incorporating Genomic Selection



Genomic prediction

- Estimate all marker effects
- Predict unobserved lines based on marker scores
  - with some accuracy < 1

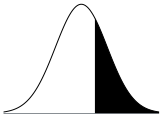
Genomic selection

- Make breeding decisions based on genomic predictions
  - increase selection intensity
  - reduce cycle time

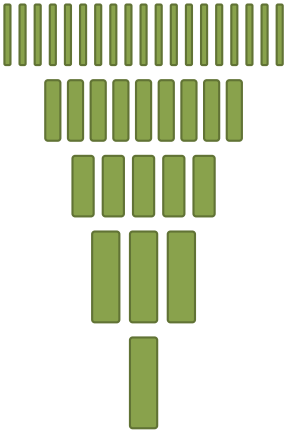
Increasing population size increases intensity

Genomic Selection

- Less accurate
  - Less phenotypic information



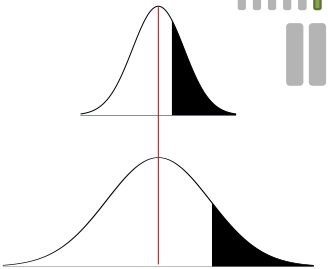
Variety Development Pipeline



Increasing population size increases intensity

Genomic Selection

- Less accurate
  - Less phenotypic information
- More intense
  - Increased number of selection candidates in early trials

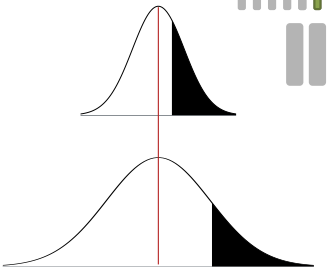


Variety Development Pipeline

Increasing population size increases intensity

Genomic Selection

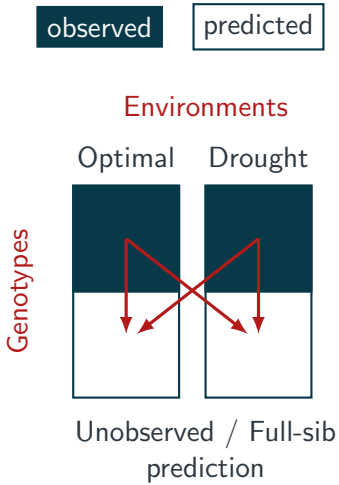
- Less accurate
  - Less phenotypic information
- More intense
  - Increased number of selection candidates in early trials
- 1-2k markers $<$ \$10/line
  - Need to make up for extra costs
  - Reduce # of plots, less reps/line
  - Trade for replication at a genetic level



Variety Development Pipeline

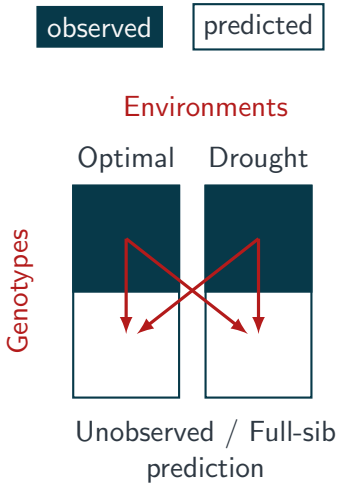
Current model for genomic prediction works...

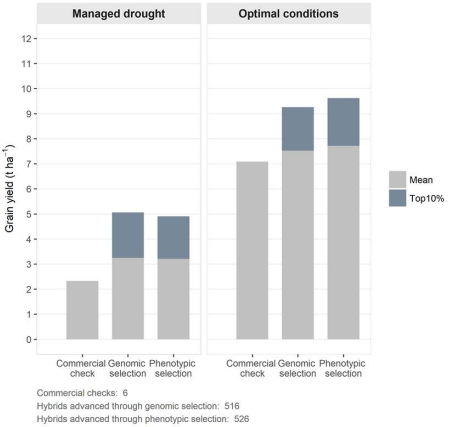
1. Observe half of population in both environments
2. Estimate genetic correlation of environments
3. Predict other half



Current model for genomic prediction works...

1. Observe half of population in both environments
2. Estimate genetic correlation of environments
3. Predict other half



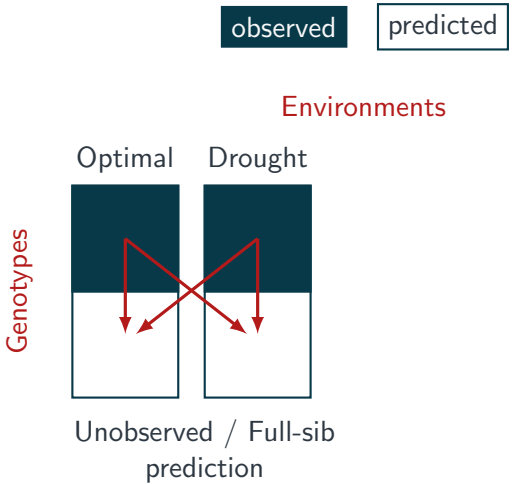


Beyene et al. 2020

... but can we do better?

Genetic relationships known

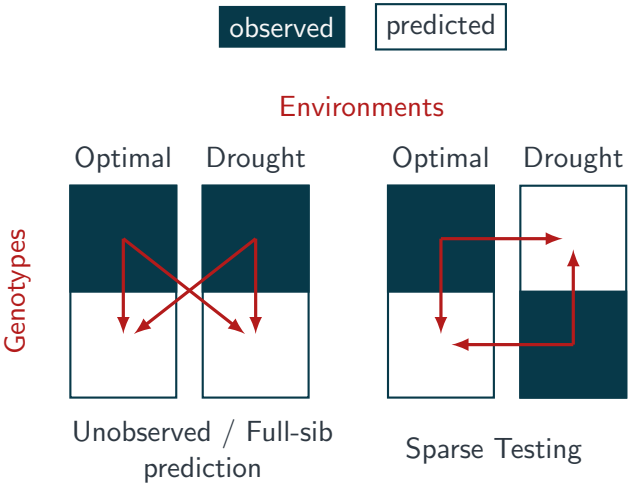
- estimate genetic correlation of environments without replicating across
- Get to observe every line



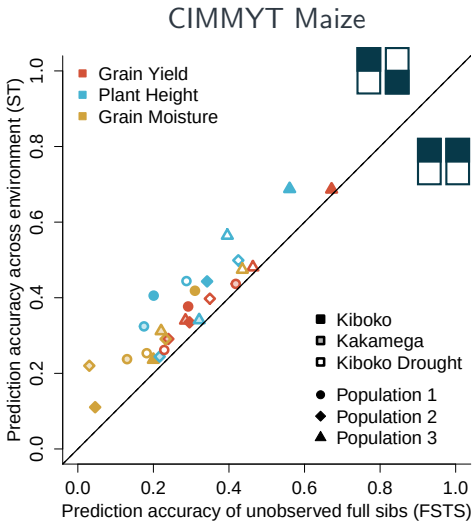
... but can we do better?

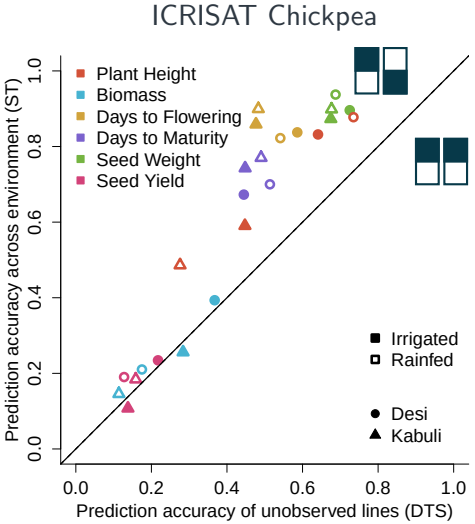
Genetic relationships known

- estimate genetic correlation of environments without replicating across
- Get to observe every line



Sparse testing is (almost) uniformly superior to unobserved prediction

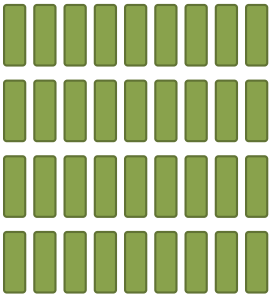




Marginal savings from plot reduction

Still have to...

- Plant all locations
- Take notes all locations
- Harvest all locations



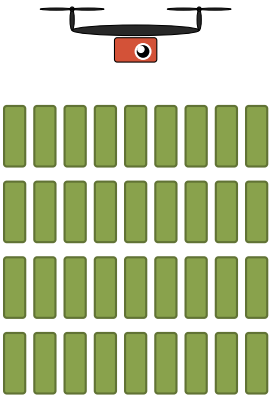
Marginal savings from plot reduction

Still have to...

- Plant all locations
- Take notes all locations
- Harvest all locations

Use drone in place of a note taker or harvester?

- All plots imaged
- Only some plots harvested
- Use multi-trait models



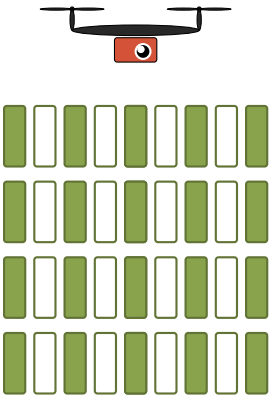
Marginal savings from plot reduction

Still have to...

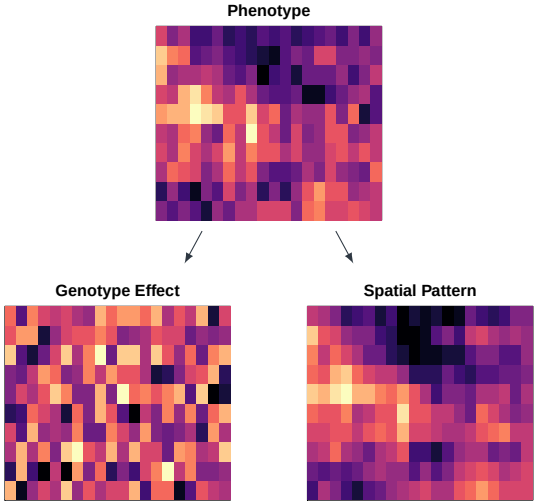
- Plant all locations
- Take notes all locations
- Harvest all locations

Use drone in place of a note taker or harvester?

- All plots imaged
- Only some plots harvested
- Use multi-trait models



Use HTPs to correct for spatial variability, increase intensity



High throughput phenotypes

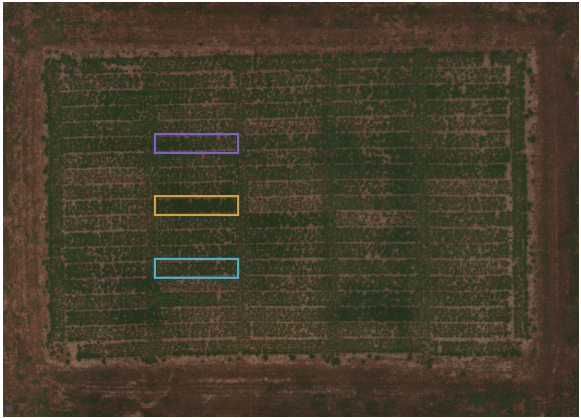
- Track genetic & spatial variation
- Monitor growth, disease progression (scab?)

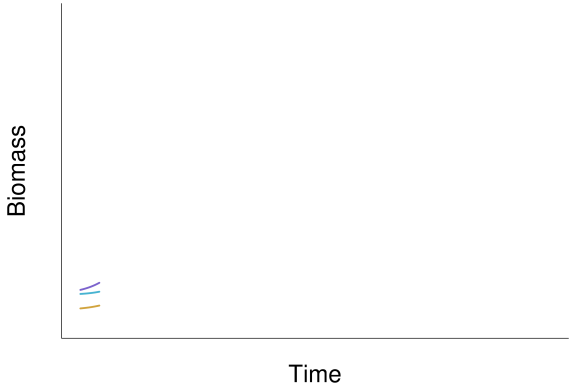
Can use to.

- Reduce replication
- Increase entries (i.e. increase i)
- Reduce time to Multi-Env. Trials

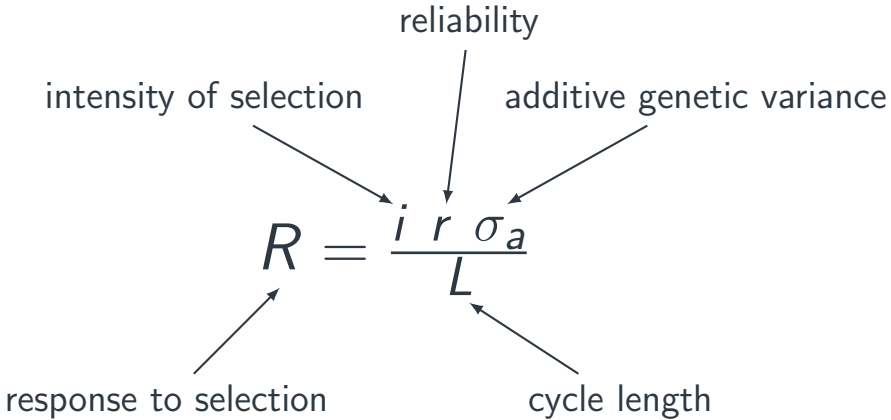
Potential SmartFarm Collaboration?

Use imagery to model growth, predict G×E

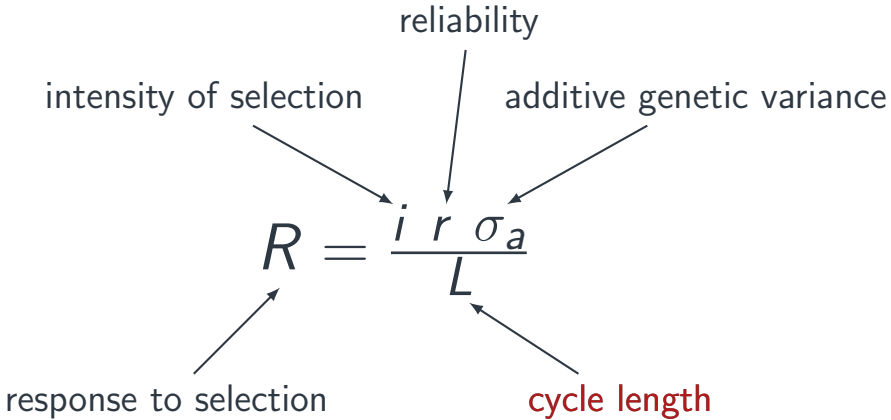




Time is on our side



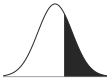
Time is on our side



Rapid cycling greatly decreases L

Rapid Cycling

- Assume 3 cycles/year in greenhouse



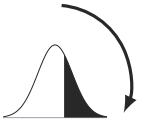
Variety Development Pipeline

Year 1

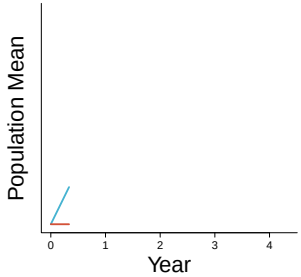
Rapid cycling greatly decreases L

Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

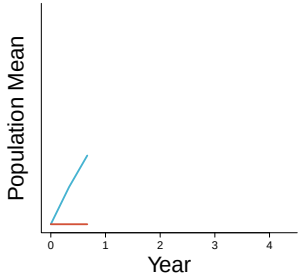


Rapid cycling greatly decreases L

Rapid Cycling

- Assume 3 cycles/year in greenhouse

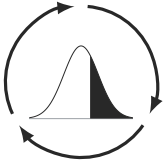
Variety Development Pipeline



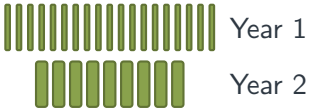
Rapid cycling greatly decreases L

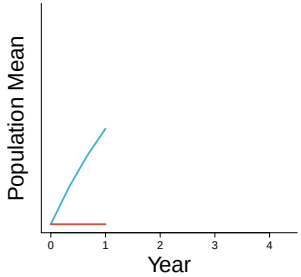
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

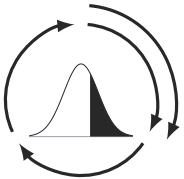




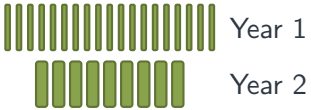
Rapid cycling greatly decreases L

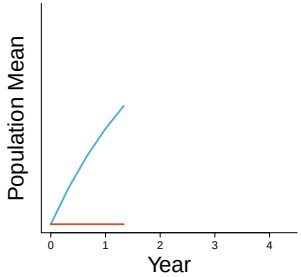
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

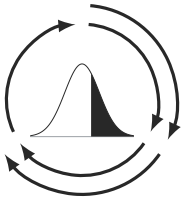




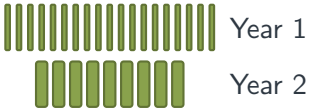
Rapid cycling greatly decreases L

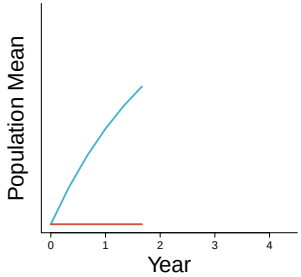
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

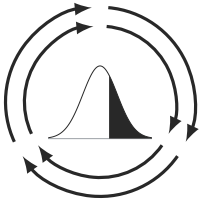




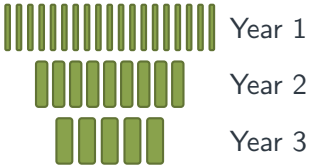
Rapid cycling greatly decreases L

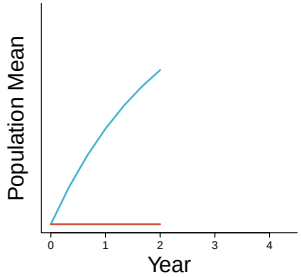
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

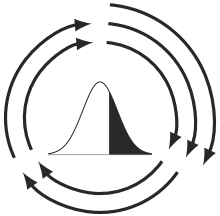




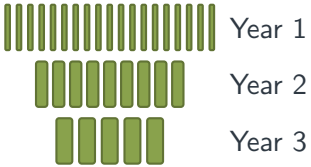
Rapid cycling greatly decreases L

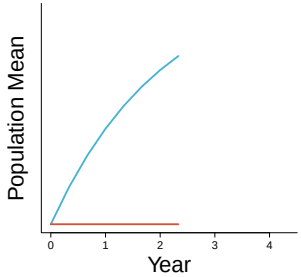
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

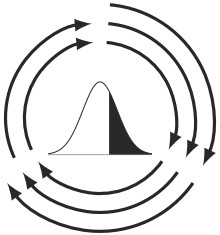




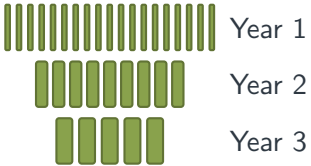
Rapid cycling greatly decreases L

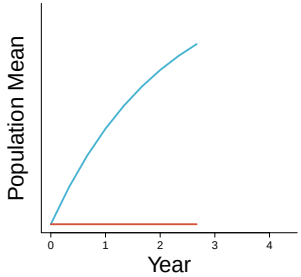
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

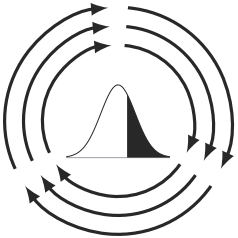




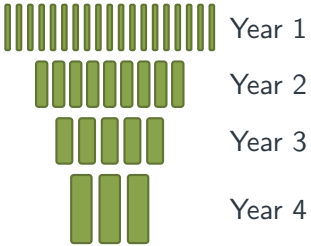
Rapid cycling greatly decreases L

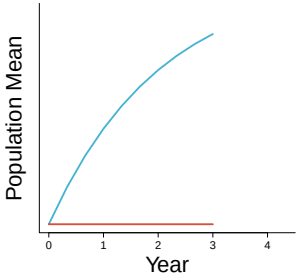
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

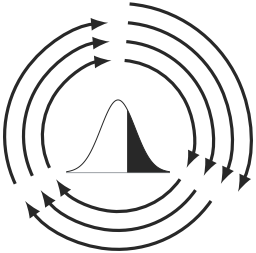




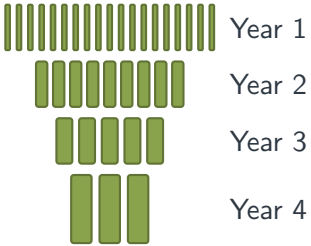
Rapid cycling greatly decreases L

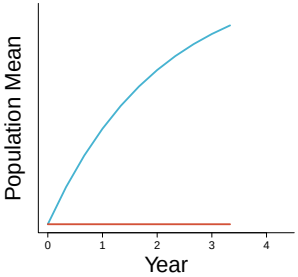
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

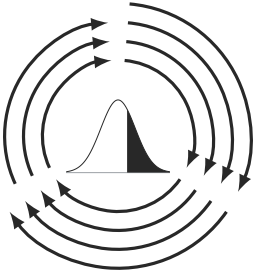




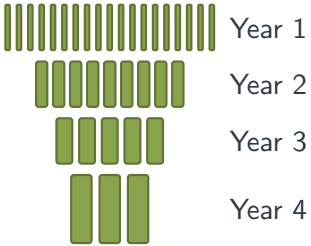
Rapid cycling greatly decreases L

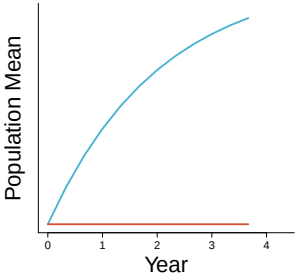
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

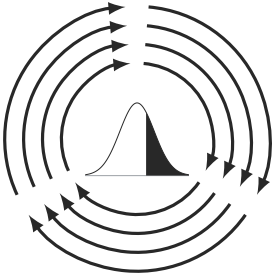




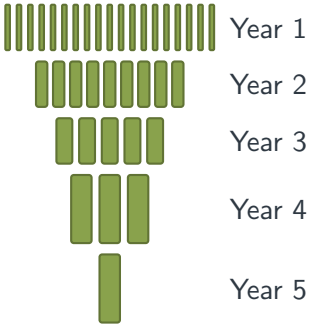
Rapid cycling greatly decreases L

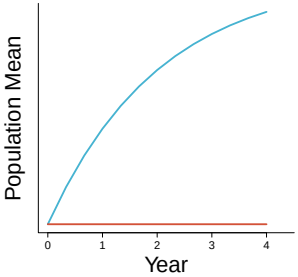
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

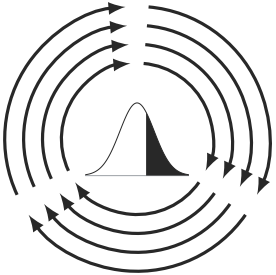




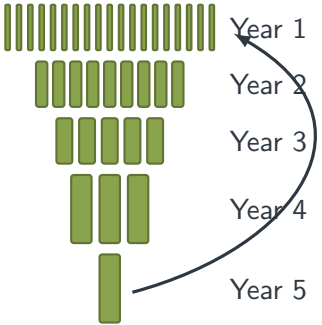
Rapid cycling greatly decreases L

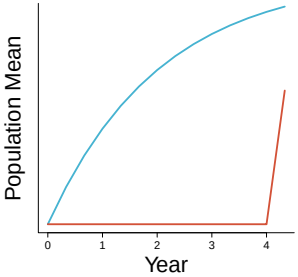
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

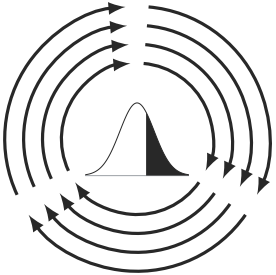




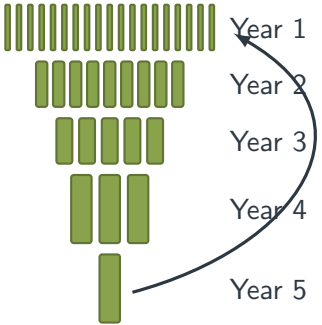
Rapid cycling greatly decreases L

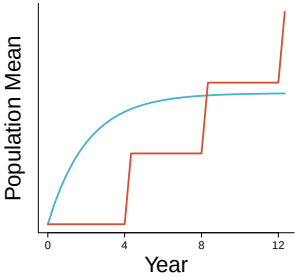
Rapid Cycling

- Assume 3 cycles/year in greenhouse

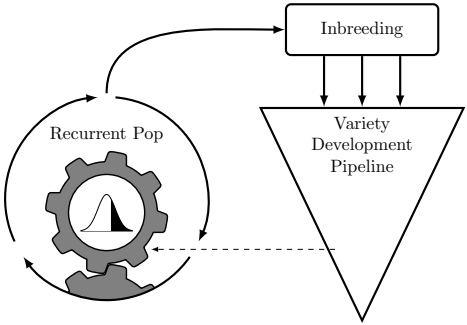


Variety Development Pipeline

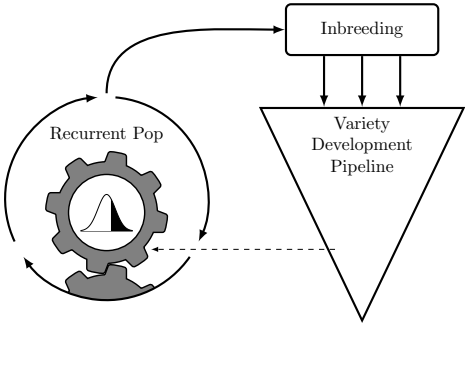


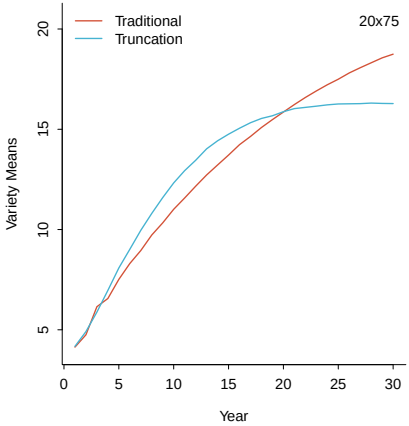


Rapid cycle recurrent truncation beats traditional, at first

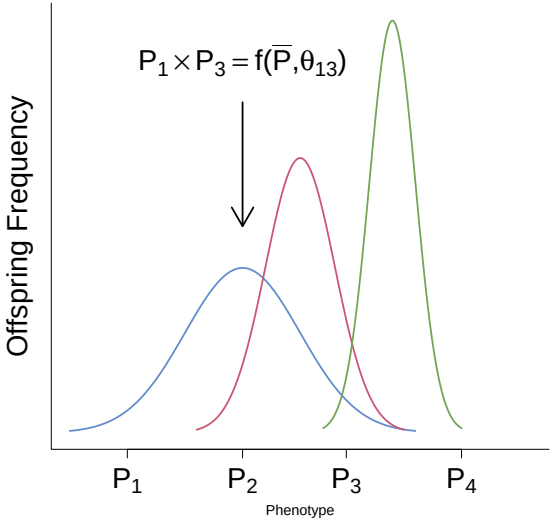


Rapid cycle recurrent truncation beats traditional, at first

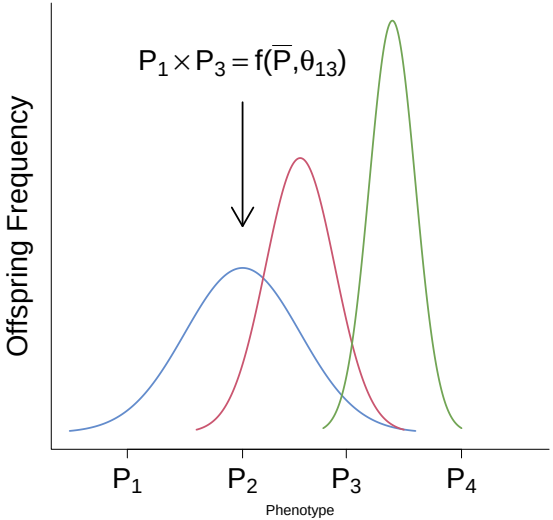


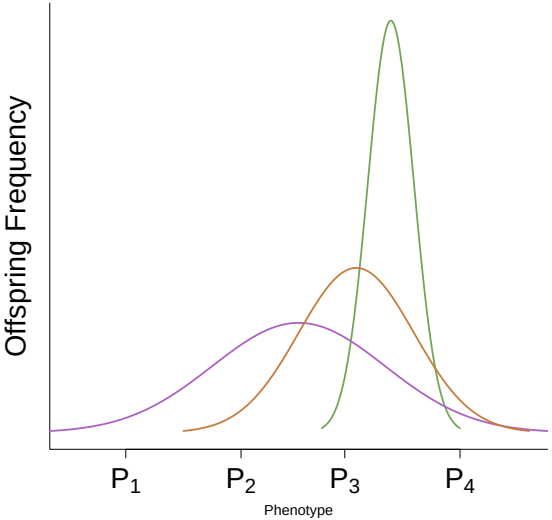


Parent selection must balance mean and variance



Parent selection must balance mean and variance





Optimal contributions of parents

Balance genetic gain and inbreeding

- Minimize inbreeding given desired gain
- Maximize gain given acceptable inbreeding
- See Meuwissen, 1997

Portfolio optimization problem

- Solve with quadratic programming
- Doesn't always pick the "best" individuals
- Parental contributions vary

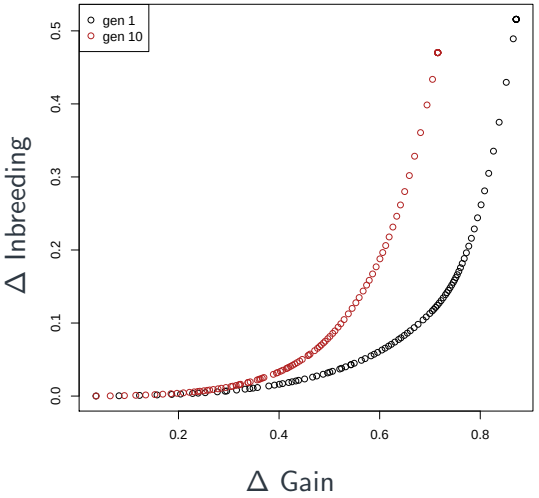
Optimal contributions of parents

Balance genetic gain and inbreeding

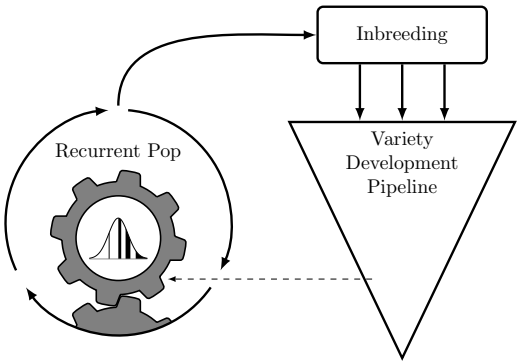
- Minimize inbreeding given desired gain
- Maximize gain given acceptable inbreeding
- See Meuwissen, 1997

Portfolio optimization problem

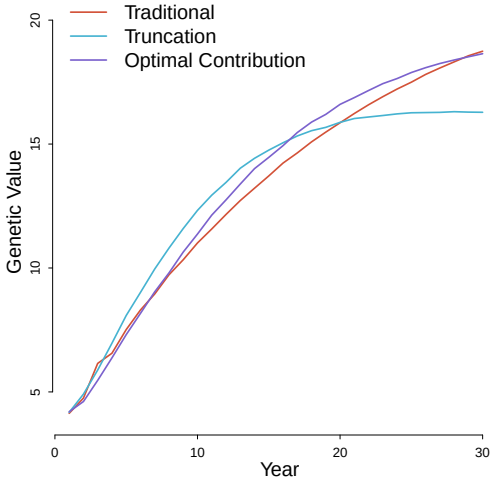
- Solve with quadratic programming
- Doesn't always pick the "best" individuals
- Parental contributions vary



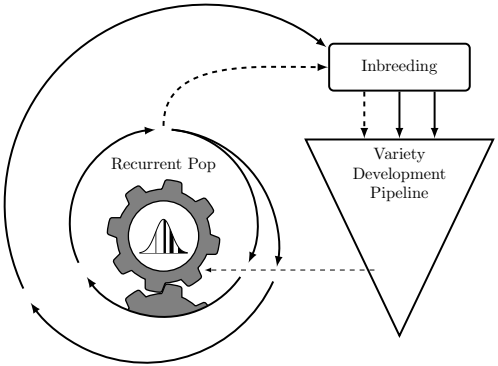
Optimal contributions can maintain V_g at the cost of short-term gains



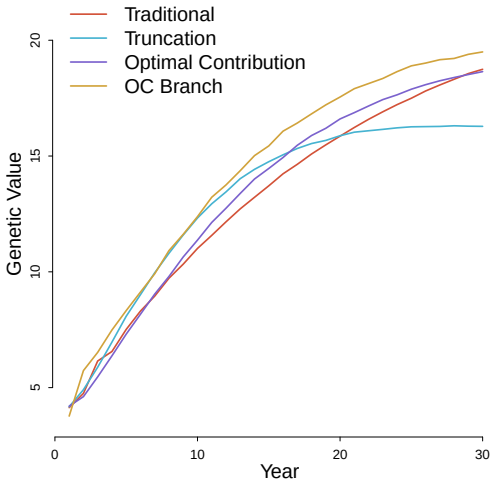
Optimal Contribution



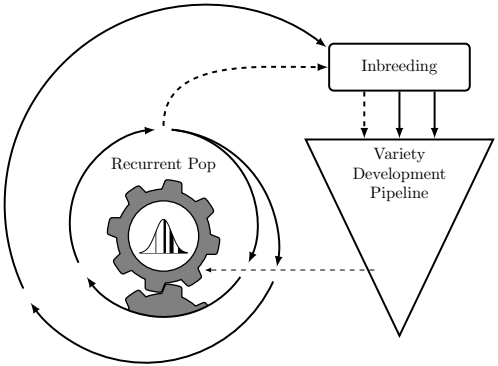
Can we have our cake and eat it too?



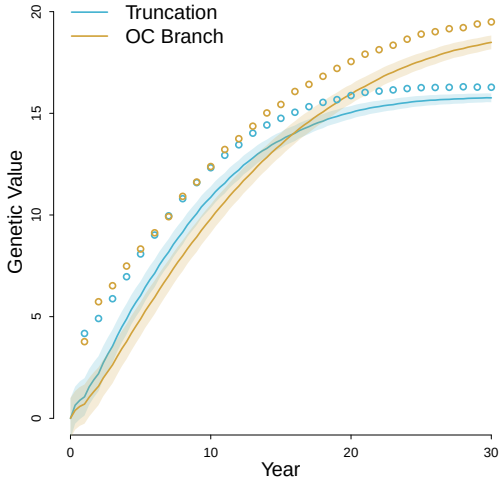
Optimal Contribution with Branching



Can we have our cake and eat it too?



Optimal Contribution with Branching



Rapid cycling

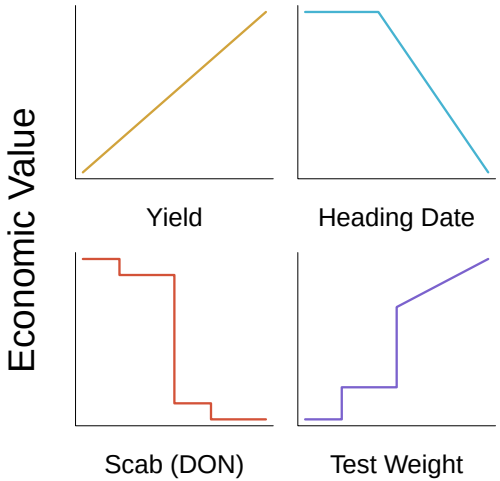
Rethinking the selection pipeline

- Traditionally used for selection and validation
- Can also serve as an information generator!
- Optimizing VDP to maximize products may not be intuitive

Rapid cycling in small grains at VT: 2025

- Start by genotyping material that has been phenotyped!
- Build foundational capacities
- Economic index for rapid improvement

Selecting for varietal value



*Jessica Rutkoski (2019)

Smith Hazel index

- Based on grain elevators
- Select on value (i.e. $)
- Improve multiple traits at once

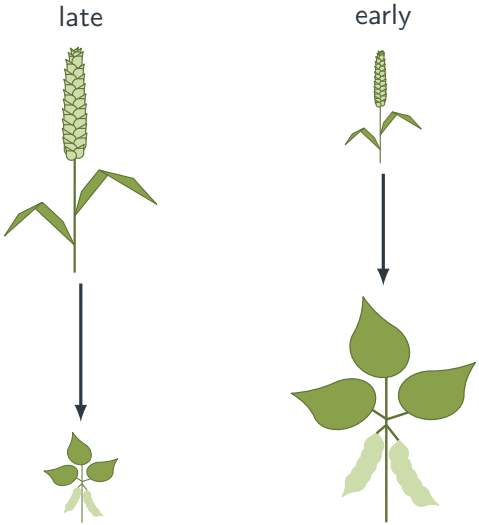
Could include other traits

- Disease resistance
- Pre-harvest sprouting
- Soybean double cropping?

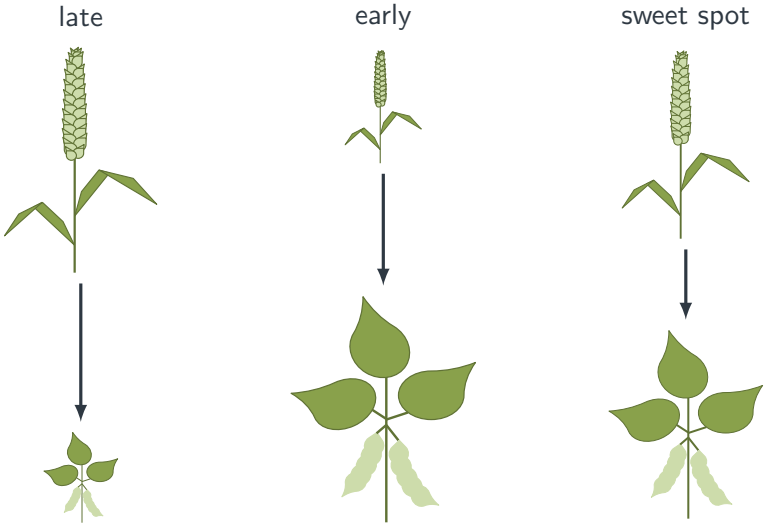
Double cropping system

late

Double cropping system



Double cropping system



What else might be optimized across species?

- No till establishment
- Soil microbiome
- Nitrogen use

G×G is a special case of G×E

Need an estimate of genetic correlation of environments

- Cov(Genotypes)
  - Can estimated with markers
- Cov(Environments)
  - Usually estimated with phenotypic data.

Wheat

G×G is a special case of G×E

Need an estimate of genetic correlation of environments

- Cov(Genotypes)
  - Can estimated with markers
- Cov(Environments)
  - Usually estimated with phenotypic data.
  - If E is another G, then estimate with markers

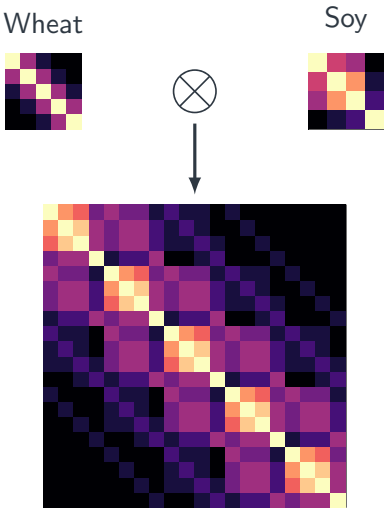
Wheat

Soy

G×G is a special case of G×E

Need an estimate of genetic correlation of environments

- Cov(Genotypes)
  - Can estimated with markers
- Cov(Environments)
  - Usually estimated with phenotypic data.
  - If E is another G, then estimate with markers

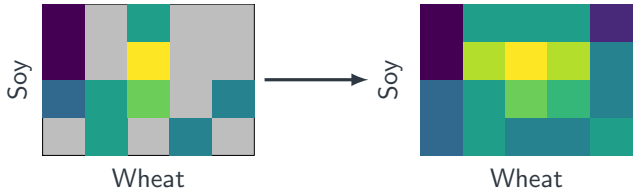


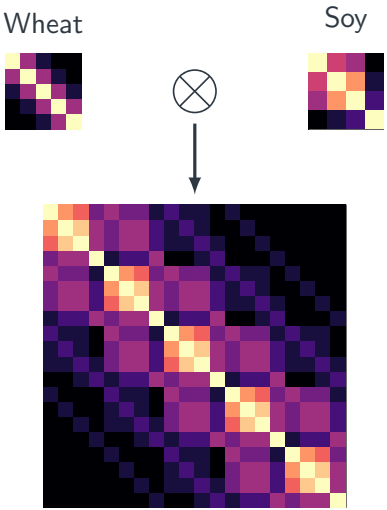
G×G is a special case of G×E

Need an estimate of genetic correlation of environments

- Cov(Genotypes)
  - Can estimated with markers
- Cov(Environments)
  - Usually estimated with phenotypic data.
  - If E is another G, then estimate with markers

Too many pairs to test, but can be predicted





Systems integration

Synergistic breeding: breed varieties for ecosystem interactions

- Soil microbiome testing for variety recommendations
- Specialized yeasts for malting barley cultivars
- Specific forages for specialized animals

Potential to build “packaged” products

Varietal release is complicated...

Few important traits

- Under direct selection

Many "threshold" traits

- Must meet market expectation
- Difficult to directly select
- Cull or advance

Varietal release is complicated...

Few important traits

- Under direct selection

Many "threshold" traits

- Must meet market expectation
- Difficult to directly select
- Cull or advance

We will always need a human at the helm

Can this be implemented at scale?

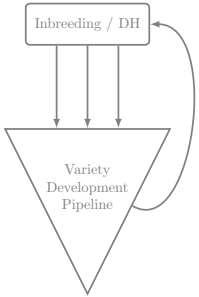
Can this be implemented at scale?

Not without foundational capacities!

Need a phased-in approach

Phase 1: Informatics

- Informatics, genotyping platform, SOPs and QC
- Genotyping all phenotyped entries to build training set
- Offset with trial designs that exploit genotypic information

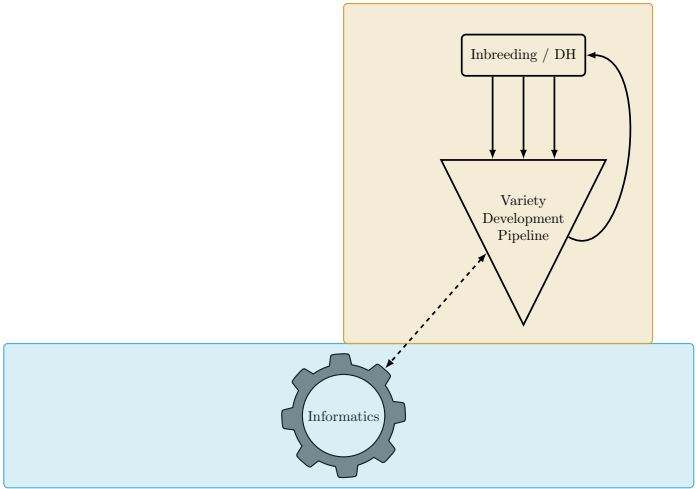


Phase 1: Informatics

- Informatics, genotyping platform, SOPs and QC
- Genotyping all phenotyped entries to build training set
- Offset with trial designs that exploit genotypic information

Phase 2: Optimize VDP

- Reduce number of years for testing
- Recycle lines earlier in the VDP
- Increase selection intensity using genomic prediction



Phase 1: Informatics

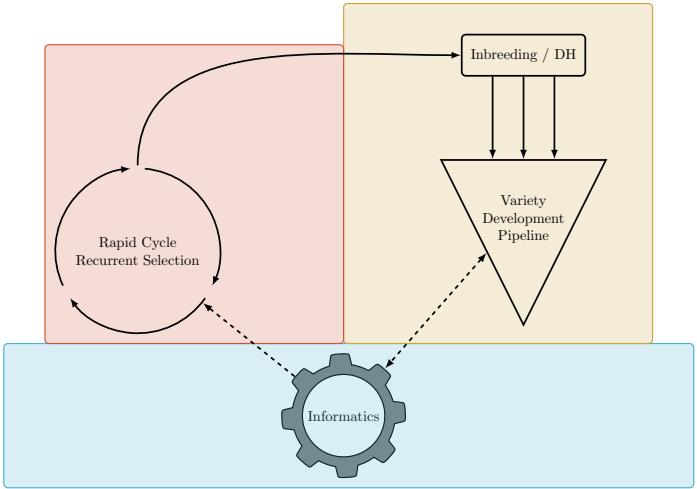
- Informatics, genotyping platform, SOPs and QC
- Genotyping all phenotyped entries to build training set
- Offset with trial designs that exploit genotypic information

Phase 2: Optimize VDP

- Reduce number of years for testing
- Recycle lines earlier in the VDP
- Increase selection intensity using genomic prediction

Phase 3: Rapid Cycling

- Rapid cycle genomic selection
- Drive generation intervals toward biological limits
- Re-optimize VDP for rapid cycling



Resources: part of a breeding informatics ecosystem

Breeding

BREEDBASE

Genomics

GOBii

API

BrAPI

ImageBreed

Multiple systems will need to to communicate and exchange information

Excellence in Breeding PLATFORM

Excellence in Breeding: Define common standards for open-source breeding informatics systems

frontiers

Technology Driven Crop Improvement for Africa and South Asia.

Nicholas Santantonio[1], Sikiru A. Atanda[2, 3, 1], Yoseph Beyene[3], Rajeev K. Varshney[4], Michael S. Olsen[3], Elizabeth Jones[5], Manish Roorkiwal[6], Xuecai Zhang[7], BHARADWAJ CHELLAPILLA[8], Pooran M. Gaur[9], Manje Gowda[3], Kate Dreher[7], Cludio A. Hernandez[7], Jose Crossa[7], Paulino Pérez-Rodríguez[10], Abhishek Rathore[6], Star Y. Gao[11], Susan McCouch[1], Kelly R. Robbins[1*]

Accepted pending minor revision

CSH Cold Spring Harbor Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

bioRxiv is receiving many new papers on coronavirus 2019-nCoV. A reminder: these are preliminary reports that have n
guide clinical practice/health-related behavior, or be reported in news media as established information.

New Results

Comment on this paper

A hybrid optimal contribution approach to drive short-term gains while maintaining long-term sustainability in a modern plant breeding program

Nicholas Santantonio, Kelly Robbins

doi: https://doi.org/10.1101/2020.01.08.899039

This article is a preprint and has not been certified by peer review [what does this mean?].

Abstract | Full Text | Info/History | Metrics | Preview PDF

G3: In Review

How can we prepare students for this change?

Multiple paths for instruction

- Generalists
- Specialists
  - Quantitative

Need to introduce "complex" ideas earlier

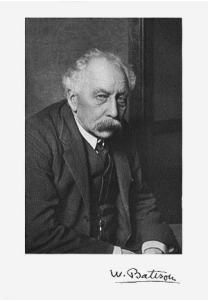
- statistics
- programming
- machine learning

Learning through doing

- Simulation as a learning tool
- Term projects

Introducing complex ideas through historical perspectives

Mendelians



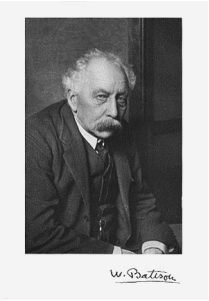
William Bateson

Biometricians

Karl Pearson

Introducing complex ideas through historical perspectives

Mendelians



William Bateson

Quantitative Genetics is born

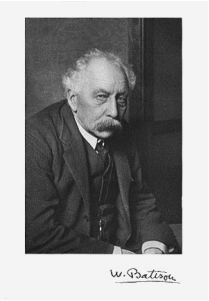
Ronald Fisher

Biometricians

Karl Pearson

Introducing complex ideas through historical perspectives

Mendelians



William Bateson

Quantitative Genetics is born

Ronald Fisher

Biometricians

Karl Pearson

- 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance
- Used Mendelian genetics to explain continuous variation

Genetic Variance

Let n = number of individuals
Let P = frequency of AA's
Let 2Q = frequency of Aa's
Let R = frequency of aa's

such that $P + 2Q + R = 1$

$$\mathrm{E}[\mathbf{y}] = \frac{1}{n}\sum_{i=1}^{n} y_i$$
$$\mu = Pa + 2Qd - Ra$$

$$\mathrm{Var}(\mathbf{y}) = \frac{1}{n}\sum_{i=1}(y_i - \mu)^2$$
$$\alpha = P(a-\mu)^2 + 2Q(d-\mu)^2 + R(-a-\mu)^2$$

Genetic Variance

Let n = number of individuals
Let P = frequency of AA's
Let 2Q = frequency of Aa's
Let R = frequency of aa's

such that $P + 2Q + R = 1$

$$\mathrm{E}[\mathbf{y}] = \frac{1}{n}\sum_{i=1}^{n} y_i$$
$$\mu = Pa + 2Qd - Ra$$

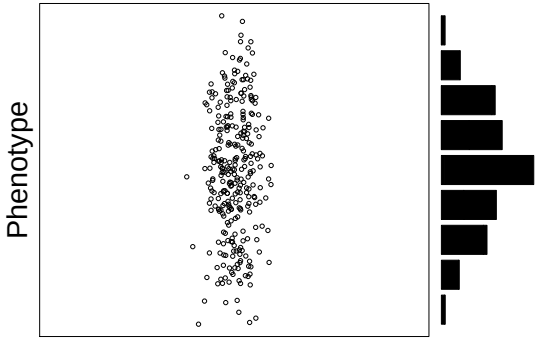
$$\mathrm{Var}(\mathbf{y}) = \frac{1}{n}\sum_{i=1}(y_i - \mu)^2$$
$$\alpha = P(a-\mu)^2 + 2Q(d-\mu)^2 + R(-a-\mu)^2$$

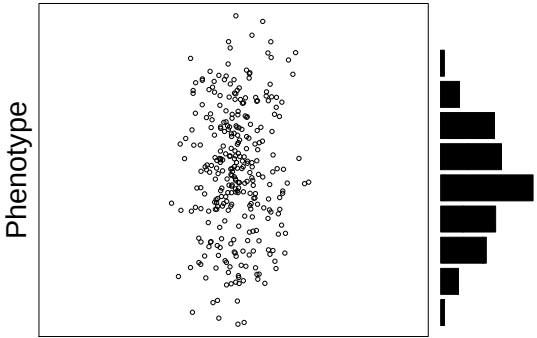
If p independent factors:

$$\mathrm{BV} = \sum_{i=1}^{p} x_i a_i$$

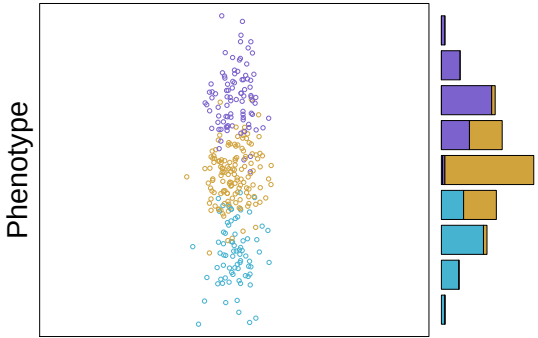
$$\mathrm{V}_g = \sum_{i=1}^{p} \alpha_i$$

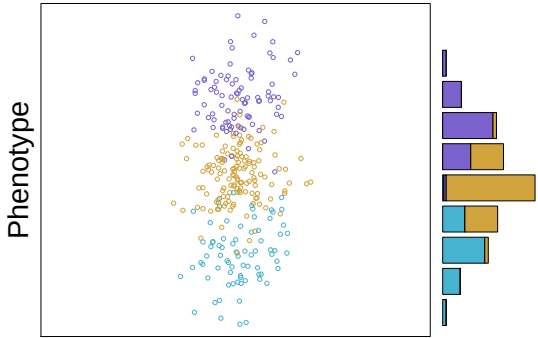
Additive only



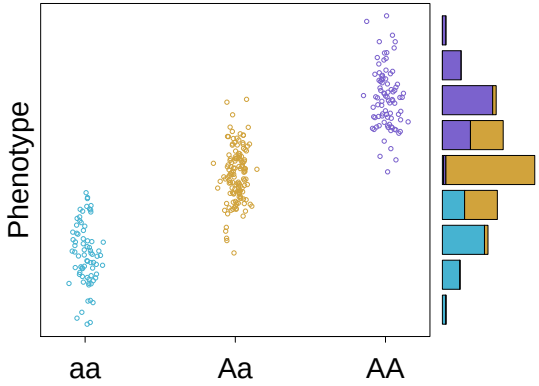


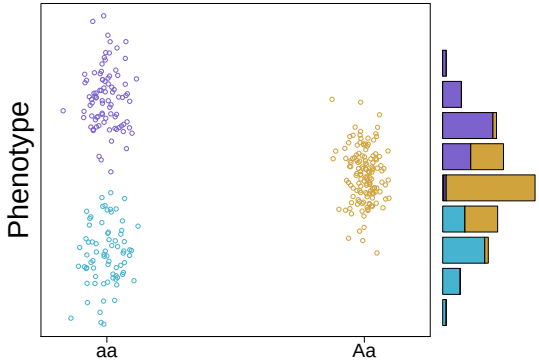
Additive only



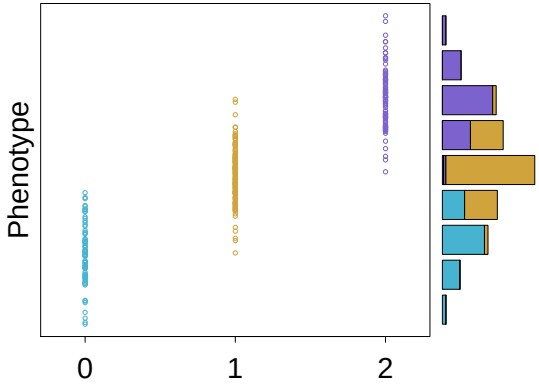


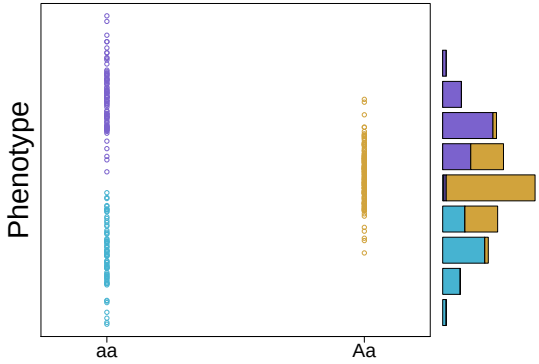
Additive only



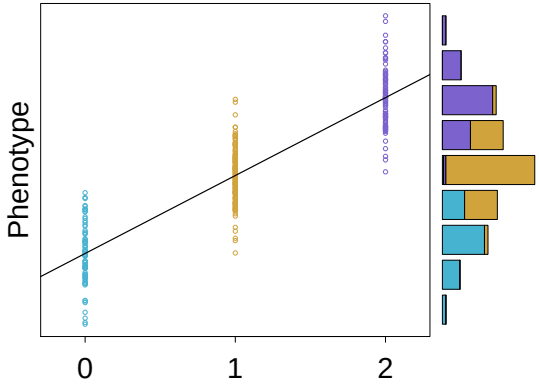


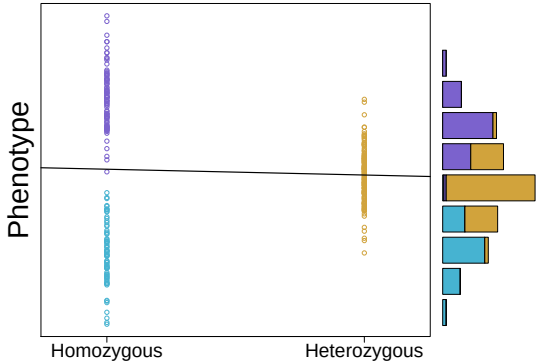
Additive only



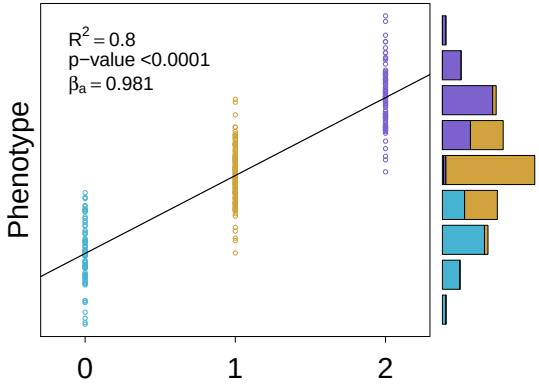


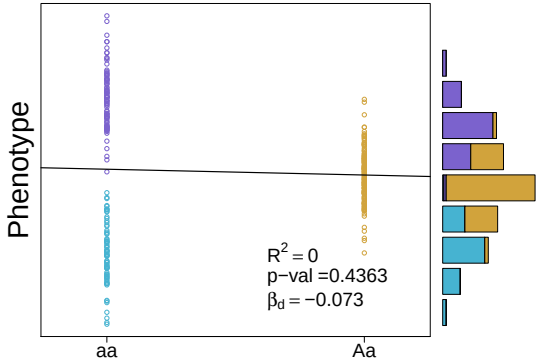
Additive only



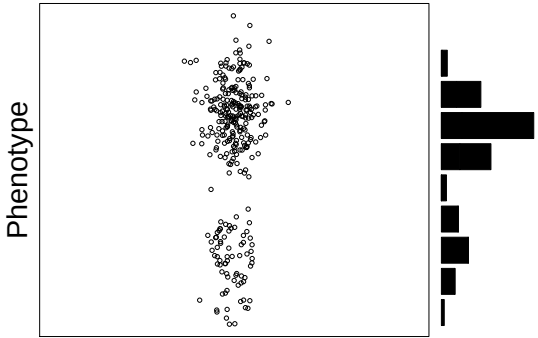


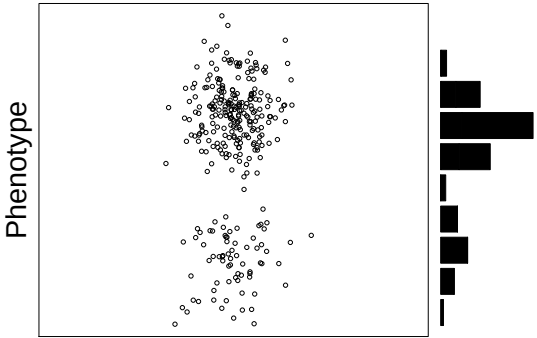
Additive only



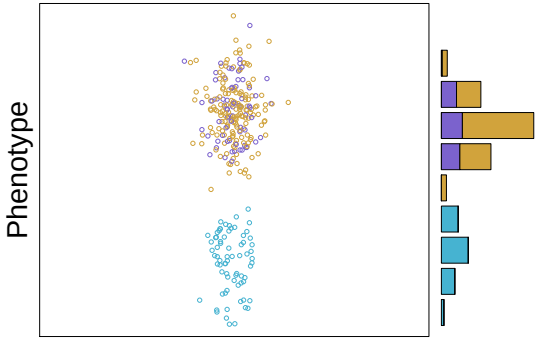


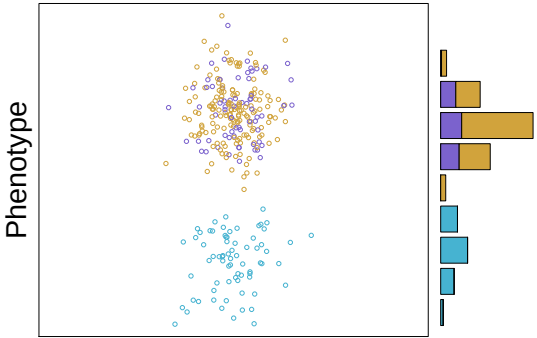
Additive with Dominance



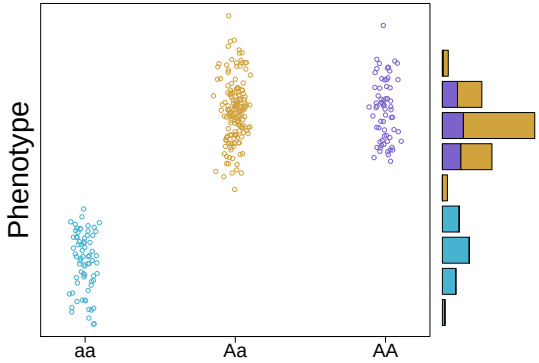


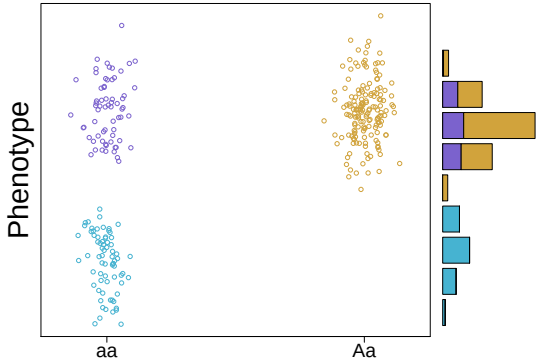
Additive with Dominance



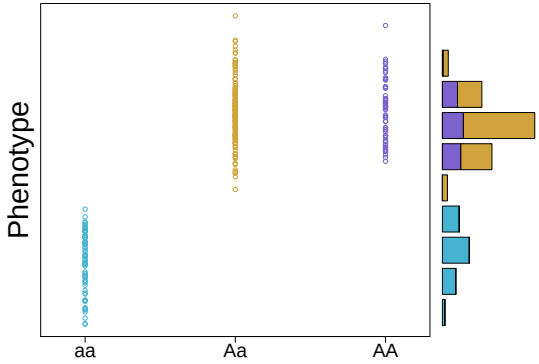


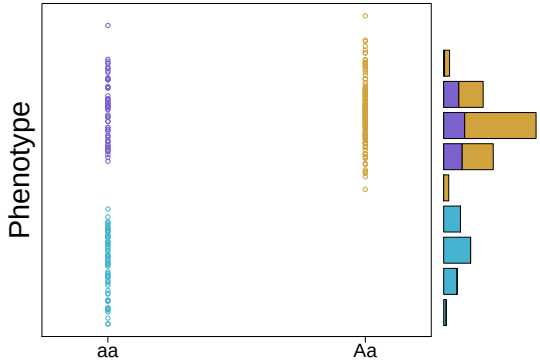
Additive with Dominance



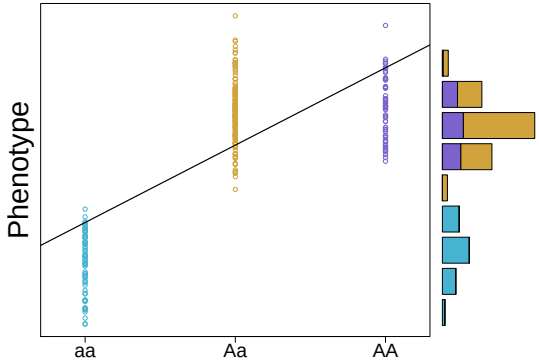


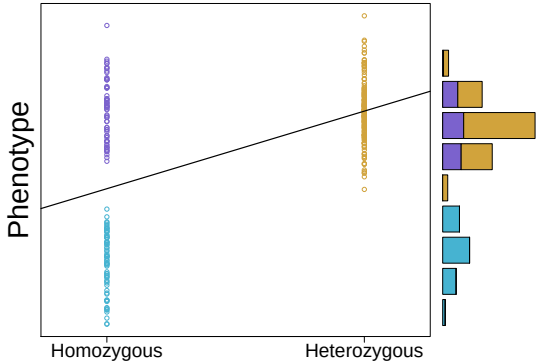
Additive with Dominance



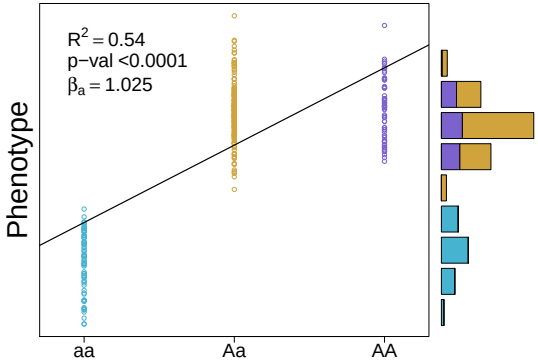


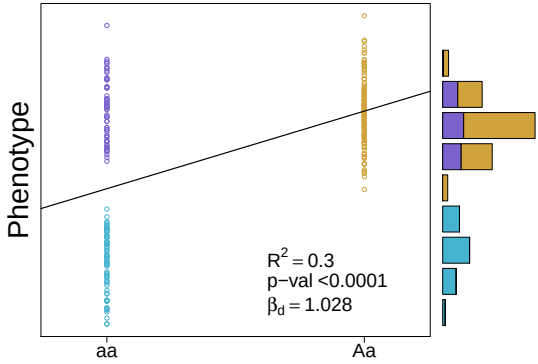
Additive with Dominance





Additive with Dominance





Using simulation as a teaching tool

Let's start with the single locus: nsantantonio.shinyapps.io/singlelocus/

Using simulation as a teaching tool

Let's start with the single locus: nsantantonio.shinyapps.io/singlelocus/

What about when there are many loci: nsantantonio.shinyapps.io/quantitative/

Genetic Variance

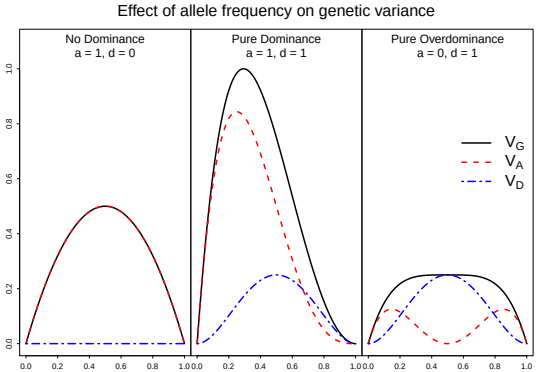
Breeder's Equation:

$$\Delta_R = \frac{ir\sigma_a}{c}$$

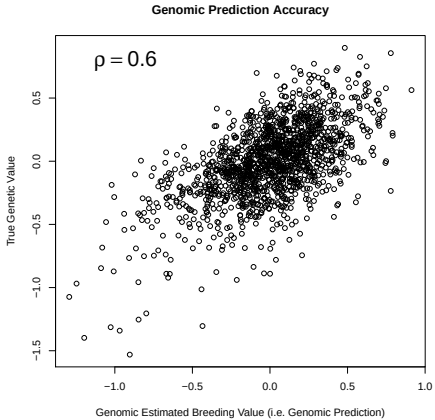
Genetic Variance

Breeder's Equation:

$$\Delta_R = \frac{ir\sigma_a}{c}$$



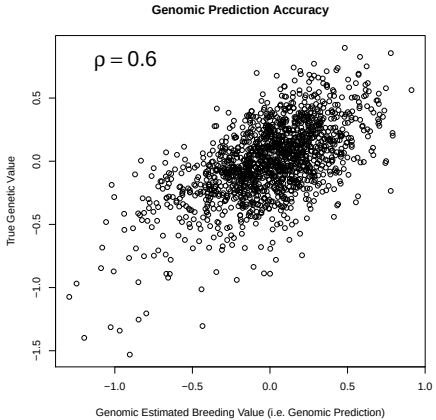
Genomic Prediction of Grain Yield



Once marker effects estimated

- sum them to predict a breeding value!
- breeding decisions based on BV

Genomic Prediction of Grain Yield



Once marker effects estimated

- sum them to predict a breeding value!
- breeding decisions based on BV

Lots of caveats...

- Start simple
- work toward more complex

Outreach

New ways to reach out

Outreach

New ways to reach out

Social media presence

- Twitter
- Instagram
- Science blogs

Outreach

New ways to reach out

Social media presence

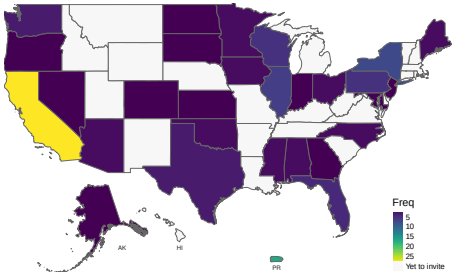
- Twitter
- Instagram
- Science blogs

Involving young people

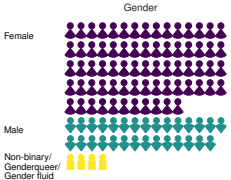
- High school summer internships
- Earlier breeding related undergrad courses

Diversity Preview Weekend

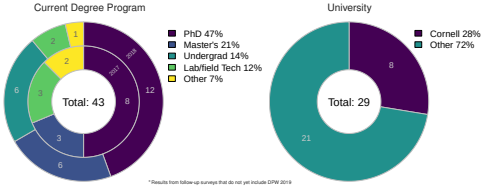
Where we come from (2017–2019)



Who we are (2017–2019)

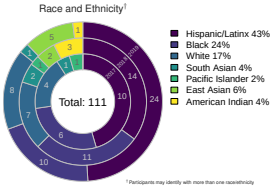


Our continuing education since DPW (2017–2018)*



*Results from follow-up surveys that do not yet include DPW 2019

Race and Ethnicity†



† Participants may identify with more than one race/ethnicity

Recognition

Robbins Lab
Kelly Robbins
Peter Selby
Sikiru Atanda
Mahlet Anche
Nicolas Morales
Evan Long

Sorrells/Jannink Labs
Lisa Kissing Kucek
Lynn Veenstra
Itaraju Brum
Uche Godfrey Okeke
Marnin Wolfe
Roberto Lonzano Gonzalez
David Benscher
Amy Fox
Jesse Chavez
James Tanaka

Other Cornell Collaborators
Susan McCouch
Mike Gore
Nick Kaczmar

Ed Buckler
Jean-Luc Jannink

Lukas Mueller

CGIAR
Mike Olsen
Yoseph Beyene
Jose Crossa
Manish Roorkiwal
Rajeev Varshney
Abhishek Rathore
many more...

New Mexico State University
Ian Ray
Chris Pierce
Chris Cramer
Champa Sengupta Gopalan
Jinfa Zhang
Robert Steiner

Funding
USDA
BILL & MELINDA GATES foundation
the Genomes to Fields initiative
IOWA CORN PROMOTION BOARD