Data-driven breeding decisions for complex traits, governed by complex genomes

Nicholas Santantonio

Cornell University

February 24th, 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 - Wheat)
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 - Wheat)
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)

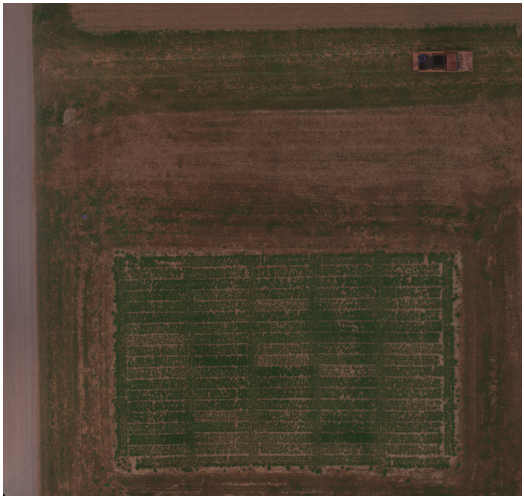
$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{ZQ}\boldsymbol{\gamma} + \sum_l \mathbf{Z}\mathbf{g}_l + \boldsymbol{\varepsilon}$$

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

Cornell University

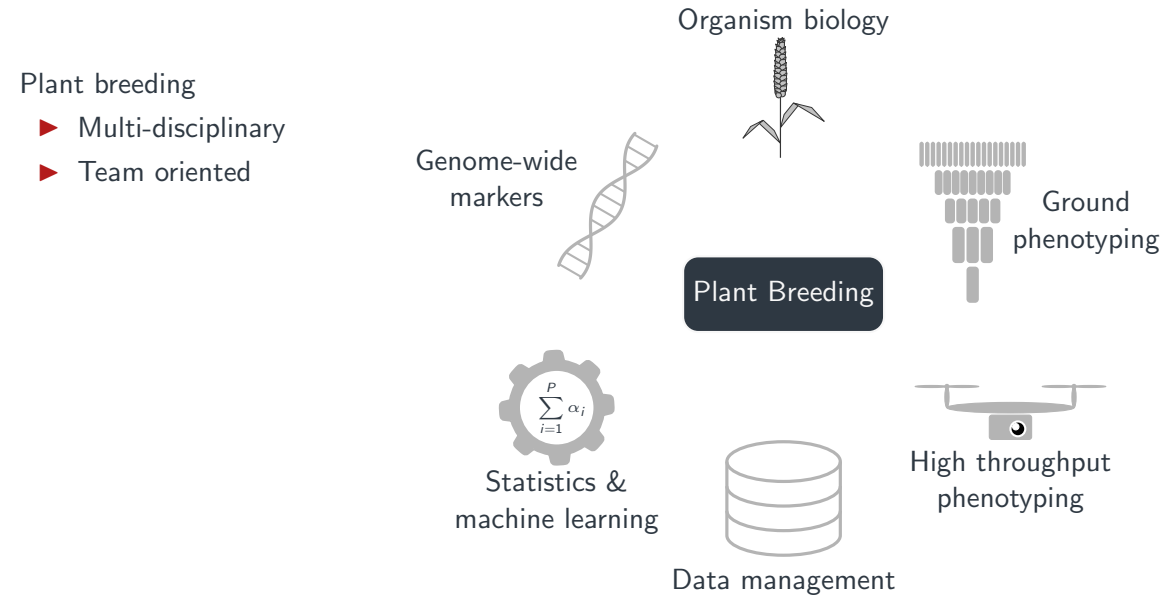
Kelly Robbins (Quantitative Genetics)

- Post Doc, current
  - Crop flexible, equal opportunity
    - Chickpea, maize, alfalfa, simulated
  - Breeding scheme optimization
    - Sparse testing
    - Optimal contributions
  - High-throughput Phenotyping
    - Longitudinal models
    - Genotype specific growth curves



Plant breeding

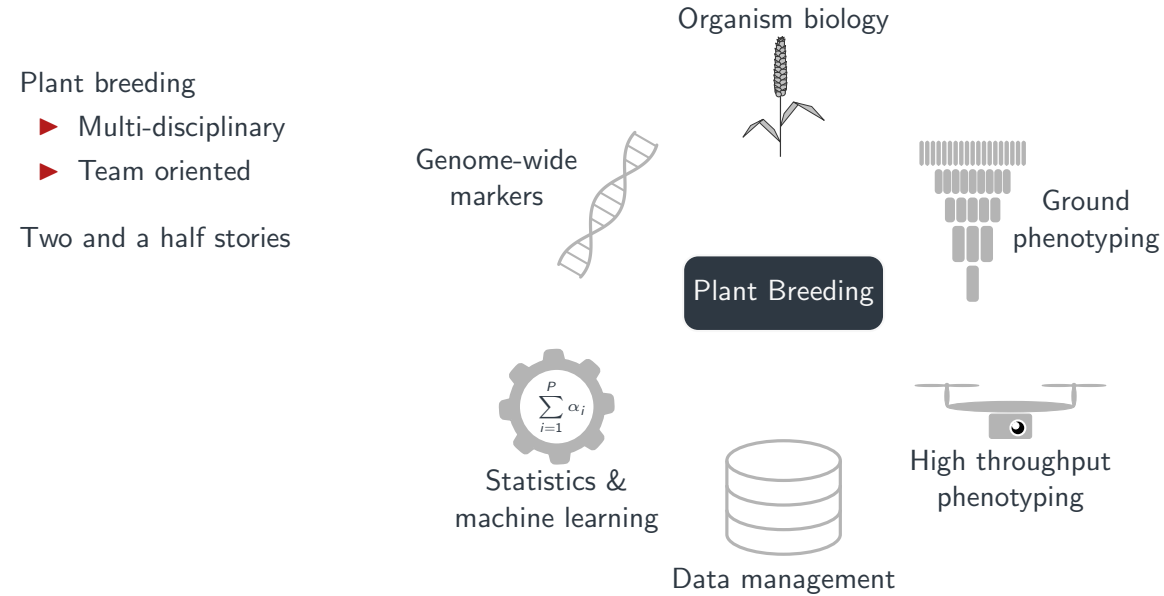
- Multi-disciplinary
- Team oriented



Plant breeding

- Multi-disciplinary
- Team oriented

Two and a half stories



Plant breeding

- Multi-disciplinary
- Team oriented

Two and a half stories

- Subgenome interactions in wheat

Plant Breeding

Plant breeding

- Multi-disciplinary
- Team oriented

Two and a half stories

- Subgenome interactions in wheat
- Transitioning to a 21st Century breeding program

Plant breeding

- Multi-disciplinary
- Team oriented

Two and a half stories

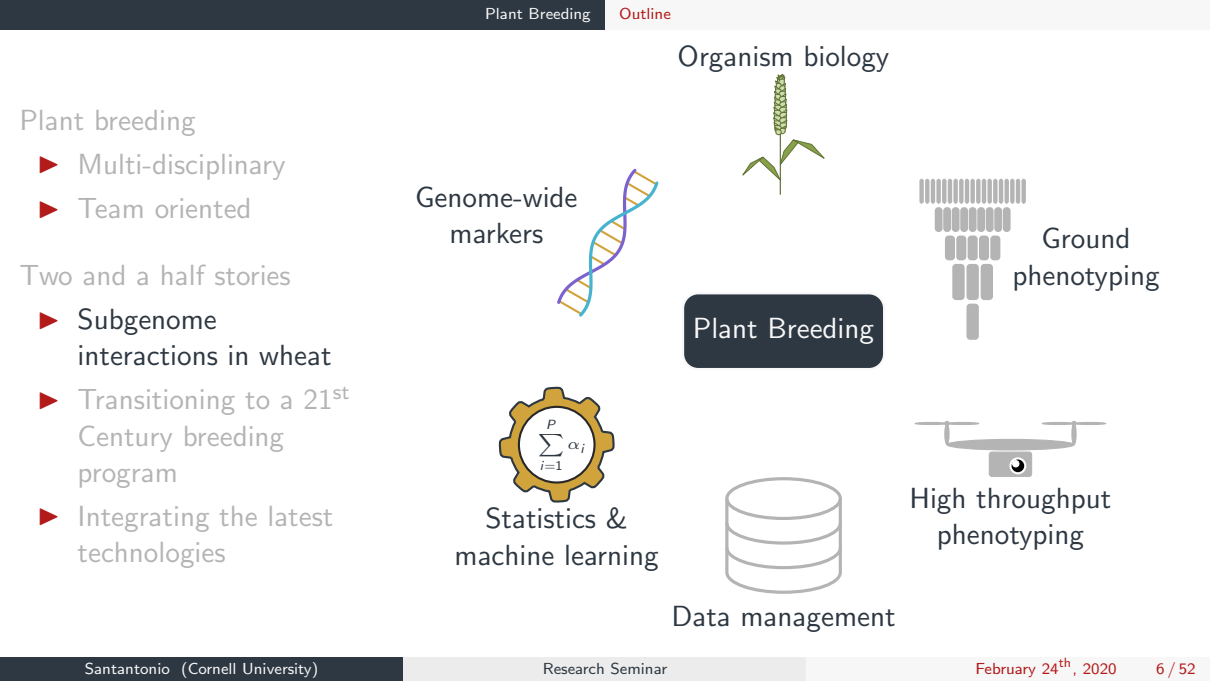
- Subgenome interactions in wheat
- Transitioning to a 21st Century breeding program
- Integrating the latest technologies

Plant breeding

- Multi-disciplinary
- Team oriented

Two and a half stories

- Subgenome interactions in wheat
- Transitioning to a 21st Century breeding program
- Integrating the latest technologies



Mac Key 1970

Hereditas 66: 165—176 (1970)

Significance of mating systems for chromosomes and gametes in polyploids

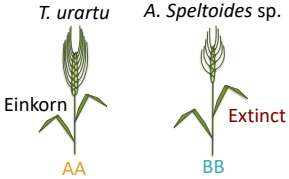
JAMES MAC KEY

Department of Genetics and Plant Breeding, Agricultural College of Sweden, Uppsala[1]

(Received August 10, 1970)

- Evolutionary "balance between new-creating and preserving forces."
- Maintain "homozygosity and heterozygosity ... at different homoeologous loci."
- Allopolyploids preserve through selfing (homo), while maintaining allelic diversity (homeo)

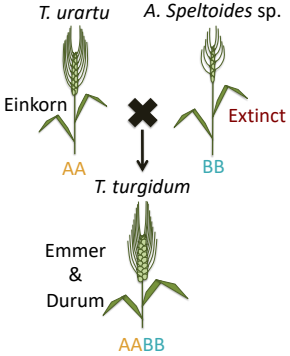
Evolution of allohexaploid wheat



Triticum

- Fertile Crecent
- Neolithic revolution
- AA $\times$ BB $\sim$ 0.5 Mya

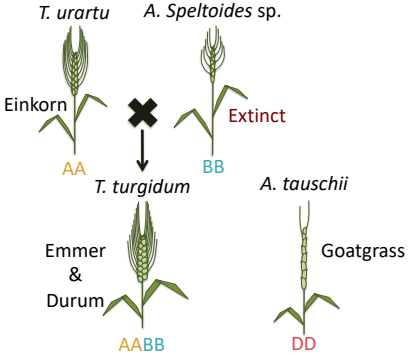
Evolution of allohexaploid wheat



Triticum

- Fertile Crecent
- Neolithic revolution
- AA $\times$ BB $\sim$ 0.5 Mya

Evolution of allohexaploid wheat



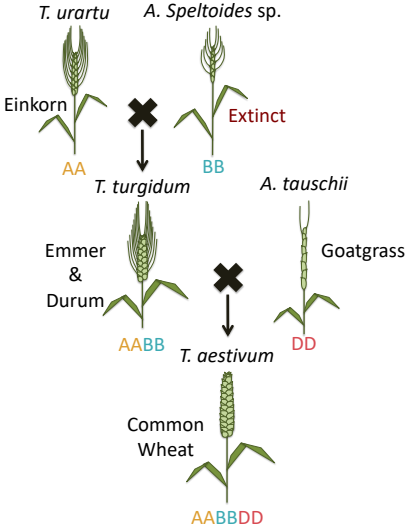
Triticum

- Fertile Crecent
- Neolithic revolution
- AA $\times$ BB $\sim$ 0.5 Mya

Aegelops

- early speciation from Triticum (A $\times$ B)
- AABB $\times$ DD $\sim$ 10,000 ya

Evolution of allohexaploid wheat



Triticum

- Fertile Crecent
- Neolithic revolution
- AA $\times$ BB $\sim$ 0.5 Mya

Aegelops

- early speciation from Triticum (A $\times$ B)
- AABB $\times$ DD $\sim$ 10,000 ya

Allohexaploid wheat

Allopolyploid

Allohexaploid wheat

Allopolyploid

- Disomic inheritance
  - no crossover across homeologous chromosomes

Allohexaploid wheat

Allopolyploid

- Disomic inheritance
  - no crossover across homeologous chromosomes
- Autogamous
  - self-pollinated (outcrossing $< 1\%$)

Allohexaploid wheat

Allopolyploid

- Disomic inheritance
  - no crossover across homeologous chromosomes
- Autogamous
  - self-pollinated (outcrossing $< 1\%$)
- Allelic diversity preserved across subgenomes
  - Most genes have three divergent copies!

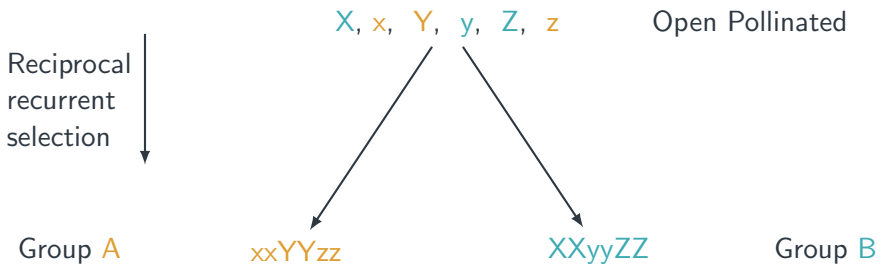
Allohexaploid wheat

Allopolyploid

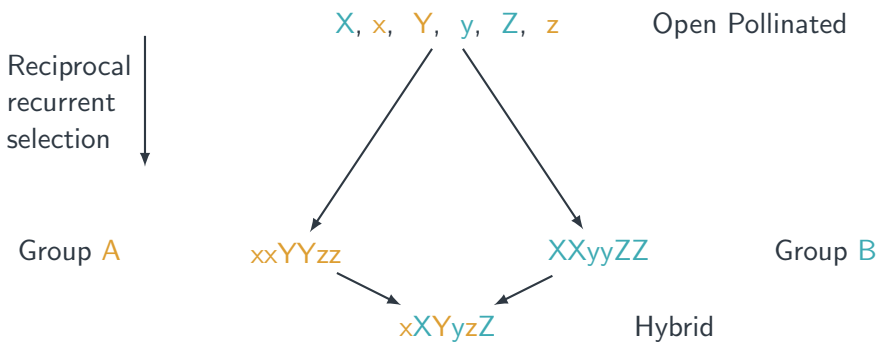
- Disomic inheritance
  - no crossover across homeologous chromosomes
- Autogamous
  - self-pollinated (outcrossing $< 1\%$)
- Allelic diversity preserved across subgenomes
  - Most genes have three divergent copies!

Is wheat an immortalized hybrid?

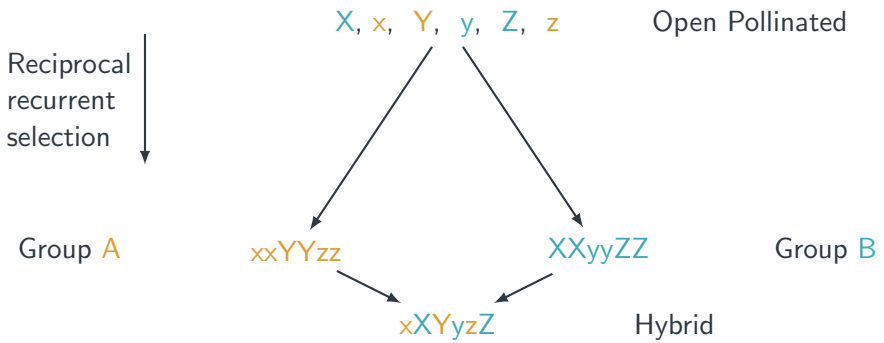
Hybrid generation



Hybrid generation

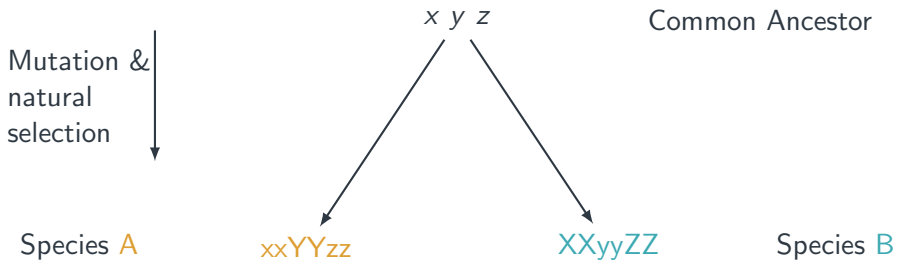


Hybrid generation

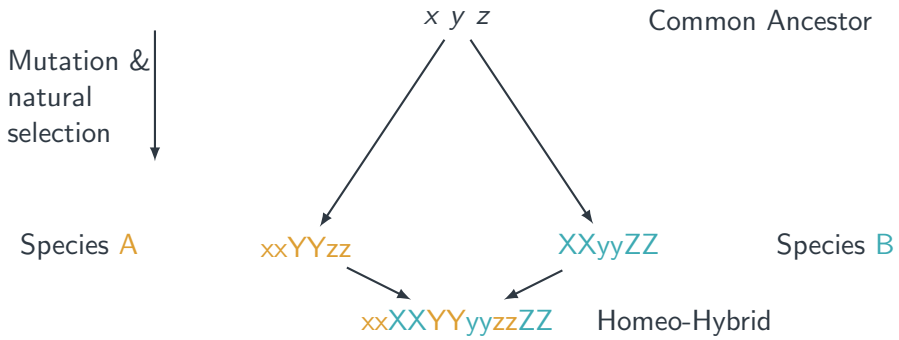


allelic interactions! (dominance)

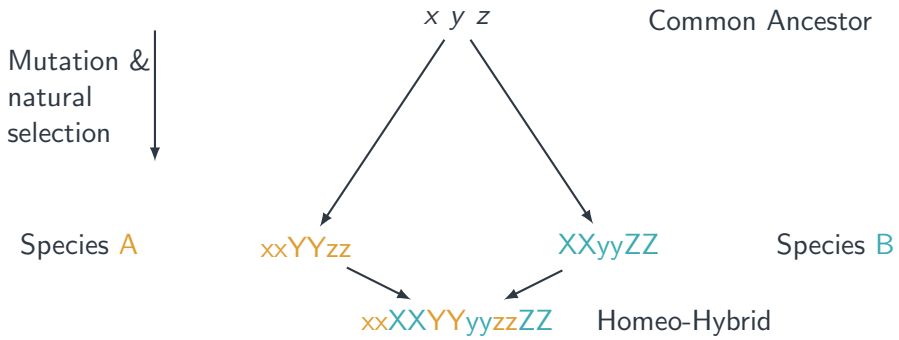
Allopolyploid formation



Allopolyploid formation

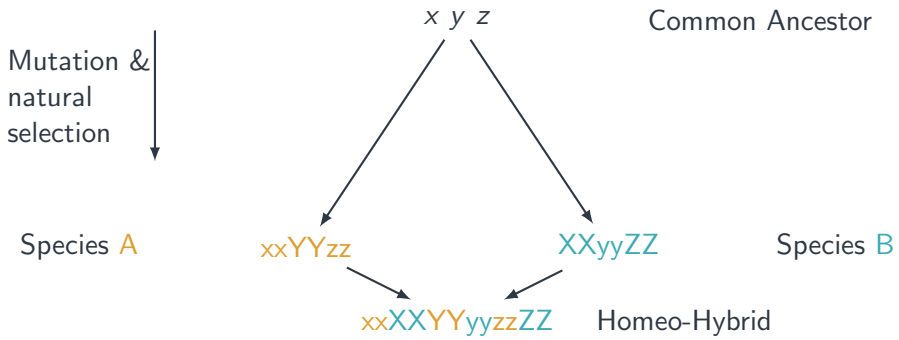


Allopolyploid formation



homeoallelic interactions? (homeologous epistasis)

Allopolyploid formation



homeoallelic interactions? (homeologous epistasis)

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

Cornell Winter Wheat Master Population

Cornell winter wheat breeding population

Cornell Winter Wheat Master Population

Cornell winter wheat breeding population

- 8,692 phenotypic records
  - 1,447 lines
  - 26 NY trials
  - 10 years (2007 - 2016)
  - 2-3 locations / year
  - 11,604 GBS markers

Cornell Winter Wheat Master Population

Cornell winter wheat breeding population

- 8,692 phenotypic records
  - 1,447 lines
  - 26 NY trials
  - 10 years (2007 - 2016)
  - 2-3 locations / year
  - 11,604 GBS markers
- 4 traits
  - Grain Yield (GY)
  - Test Weight (TW)
  - Heading Date (HD)
  - Plant Height (PH)

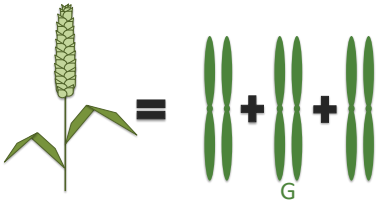
Cornell Winter Wheat Master Population

Cornell winter wheat breeding population

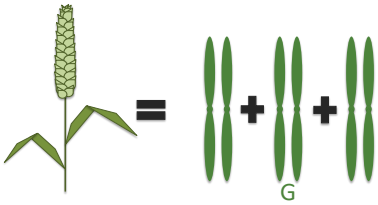
- 8,692 phenotypic records
  - 1,447 lines
  - 26 NY trials
  - 10 years (2007 - 2016)
  - 2-3 locations / year
  - 11,604 GBS markers
- 4 traits
  - Grain Yield (GY)
  - Test Weight (TW)
  - Heading Date (HD)
  - Plant Height (PH)

- Align markers to RefSeq v1.0
  - separate markers by subgenome
  - calculate genetic covariance for each subgenome: $\mathbf{K}_A$, $\mathbf{K}_B$ and $\mathbf{K}_D$
  - estimate subgenome breeding values

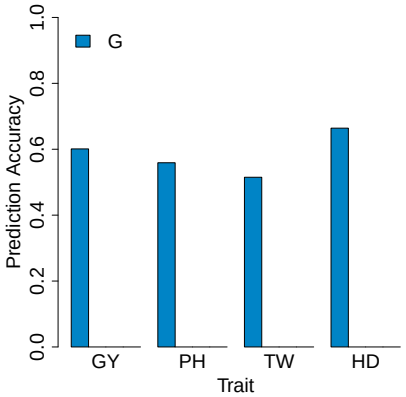
Can predict subgenome breeding values



Can predict subgenome breeding values

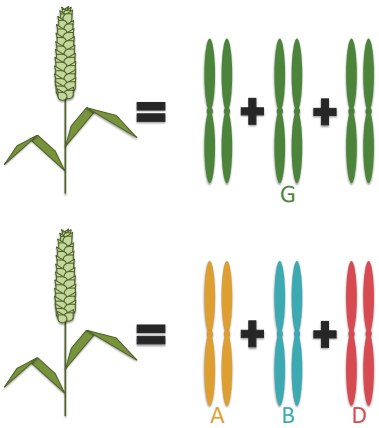


Santantonio, Jannink and Sorrells (2019a; G3)

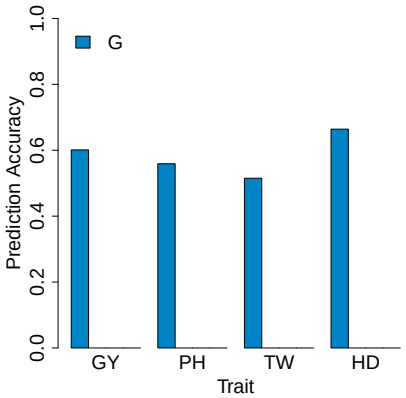


Use genomic prediction to evaluate genetic signal

Can predict subgenome breeding values

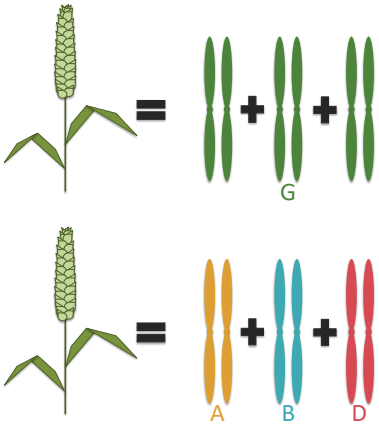


Santantonio, Jannink and Sorrells (2019a; G3)

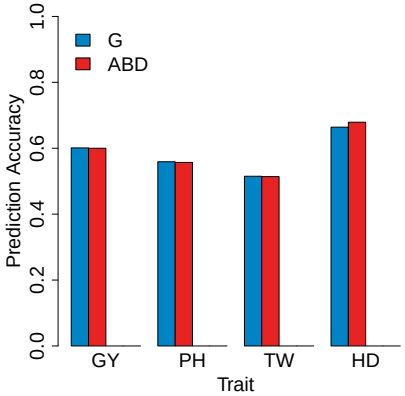


Use genomic prediction to evaluate genetic signal

Can predict subgenome breeding values

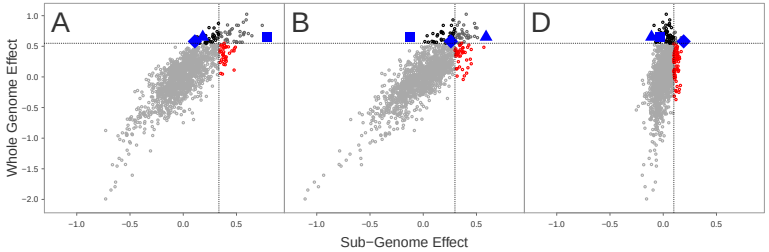


Santantonio, Jannink and Sorrells (2019a; G3)

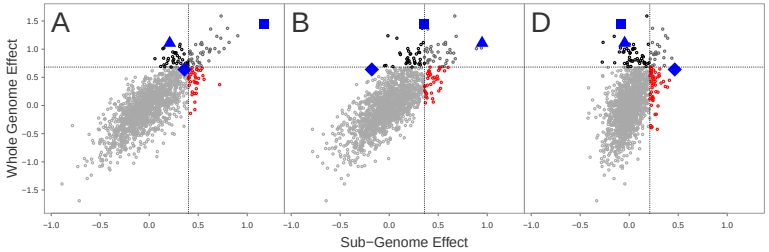


Use genomic prediction to evaluate genetic signal

GY Grain Yield



TW Test Weight

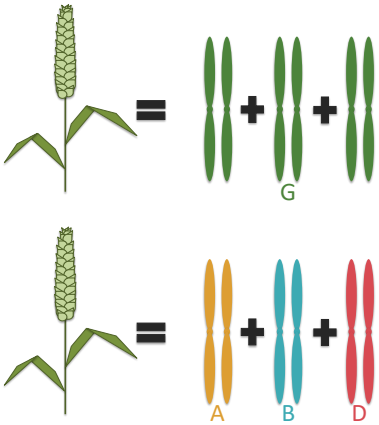


best A
best B
best D

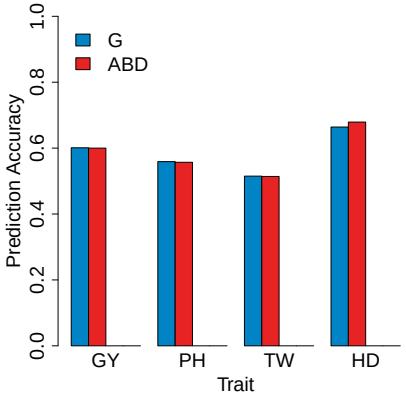
- Best individuals dont have the best subgenomes
- Can select on specific subgenomes
- Can select parents with complementary subgenomes

Santantonio, Jannink and Sorrells (2019a; G3)

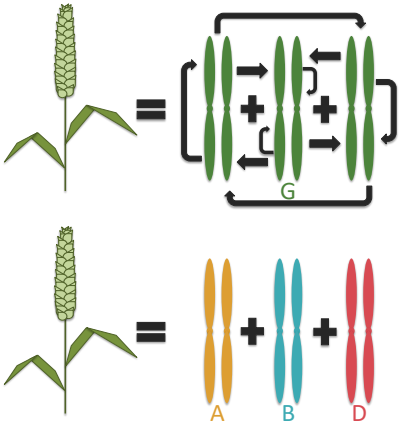
Subgenome interactions aid genomic prediction



Santantonio, Jannink and Sorrells (2019a; G3)



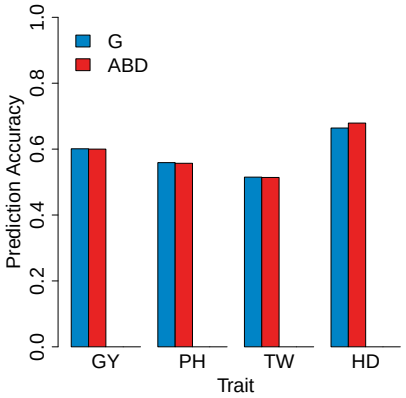
Subgenome interactions aid genomic prediction



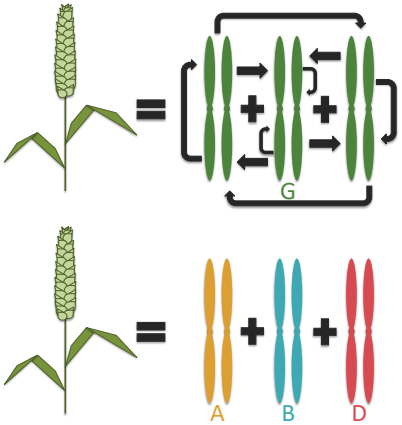
Hadamard product for epistatic covariance

$\mathbf{K}_G \odot \mathbf{K}_G$

Santantonio, Jannink and Sorrells (2019a; G3)



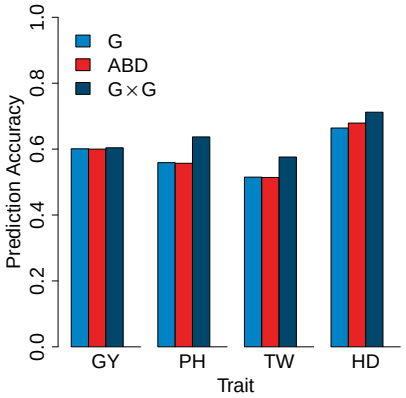
Subgenome interactions aid genomic prediction



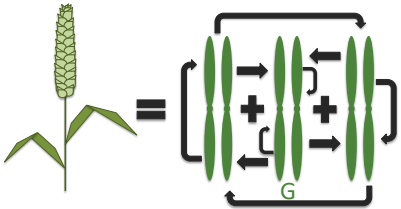
Hadamard product for epistatic covariance

$\mathbf{K}_G \odot \mathbf{K}_G$

Santantonio, Jannink and Sorrells (2019a; G3)



Subgenome interactions aid genomic prediction



Hadamard product for epistatic covariance

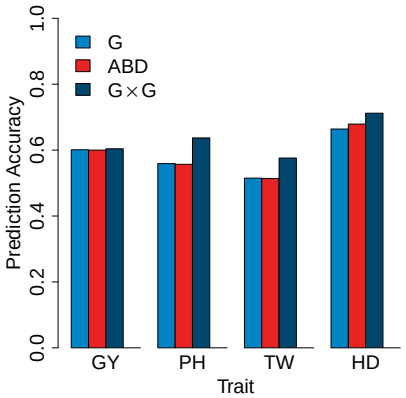
$\mathbf{K}_G \odot \mathbf{K}_G$

$\mathbf{K}_A \odot \mathbf{K}_B$

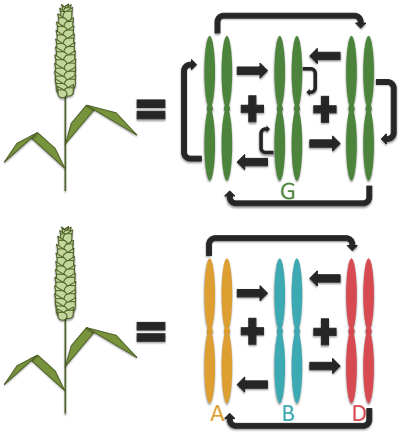
$\mathbf{K}_A \odot \mathbf{K}_D$

$\mathbf{K}_B \odot \mathbf{K}_D$

Santantonio, Jannink and Sorrells (2019a; G3)



Subgenome interactions aid genomic prediction



Hadamard product for epistatic covariance

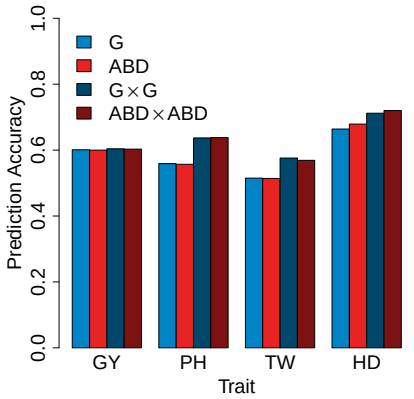
$\mathbf{K}_G \odot \mathbf{K}_G$

$\mathbf{K}_A \odot \mathbf{K}_B$

$\mathbf{K}_A \odot \mathbf{K}_D$

$\mathbf{K}_B \odot \mathbf{K}_D$

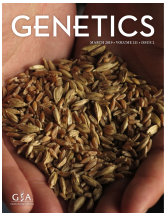
Santantonio, Jannink and Sorrells (2019a; G3)



Suggests all epistasis is homeologous?

Two-Locus Epistasis

Santantonio, Jannink and Sorrells (2019b; Genetics)



Consider the two locus model (from Hill et al. 2008):

$$\mathrm{E}[y] = \mu + B\alpha_B + C\alpha_C + BC(\alpha\alpha)_{BC}$$

Additive $\times$ Additive
(statistical epistasis)

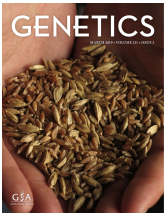
	CC	cc
BB	$2a$	0
bb	0	$2a$

Duplicate Factor
(functional epistasis)

	CC	cc
BB	a	a
bb	a	0

Two-Locus Epistasis

Santantonio, Jannink and Sorrells (2019b; Genetics)



Consider the two locus model (from Hill et al. 2008):

$$\mathrm{E}[y] = \mu + B\alpha_B + C\alpha_C + BC(\alpha\alpha)_{BC}$$

Additive × Additive
(statistical epistasis)

	CC	cc
BB	$2a$	0
bb	0	$2a$

Duplicate Factor
(functional epistasis)

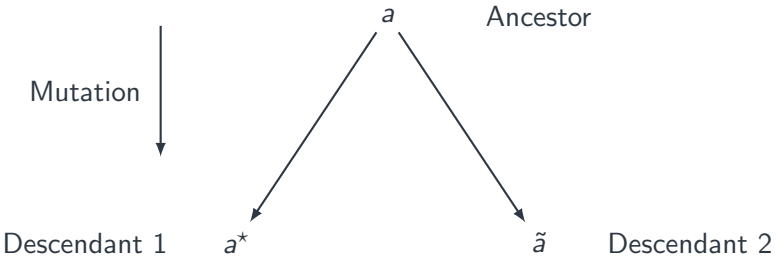
	CC	cc
BB	a	a
bb	a	0

Subfunctionalization
(subfunctional epistasis)

	CC	cc
BB	$s(a^\star + \tilde{a})$	$a^\star$
bb	$\tilde{a}$	0

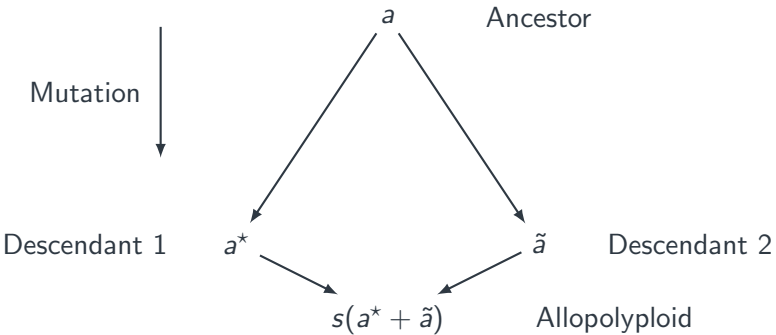
Subfunctionalization Epistasis

Let a be the effect of a functional allele (or haplotype),



Subfunctionalization Epistasis

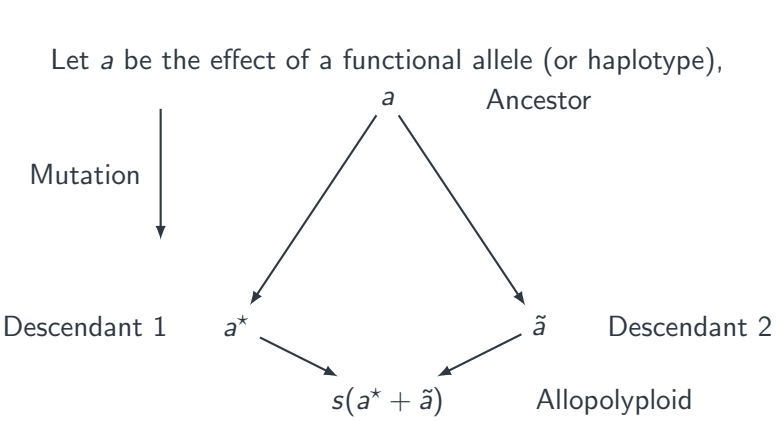
Let a be the effect of a functional allele (or haplotype),



s = subfunctionalization coefficient

Subfunctionalization Epistasis

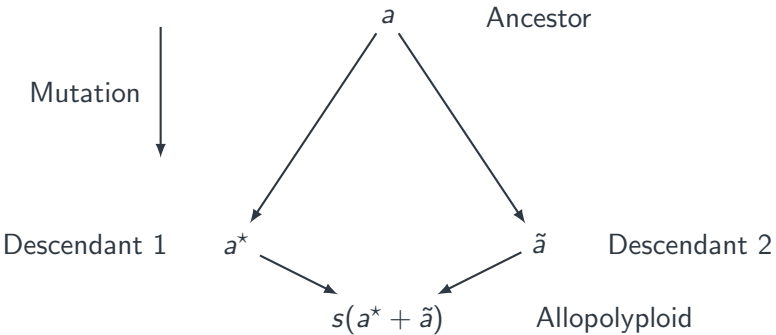
Let a be the effect of a functional allele (or haplotype),



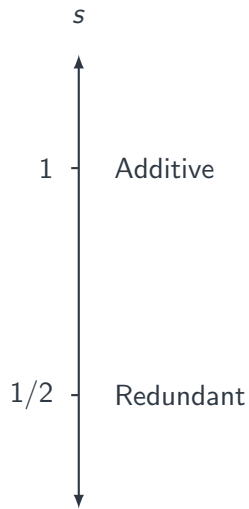
s = subfunctionalization coefficient

Subfunctionalization Epistasis

Let a be the effect of a functional allele (or haplotype),

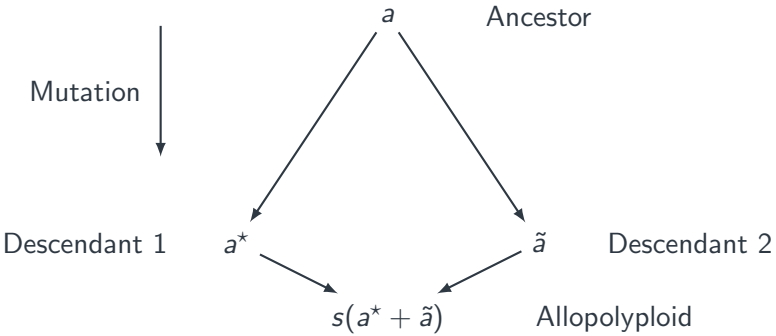


s = subfunctionalization coefficient

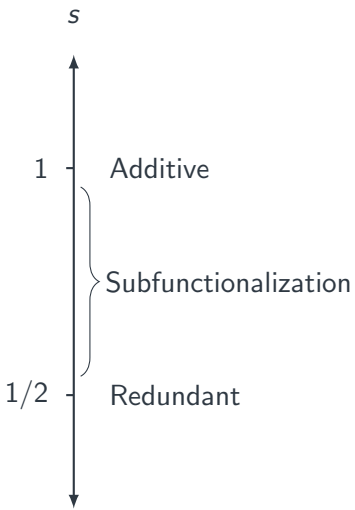


Subfunctionalization Epistasis

Let a be the effect of a functional allele (or haplotype),

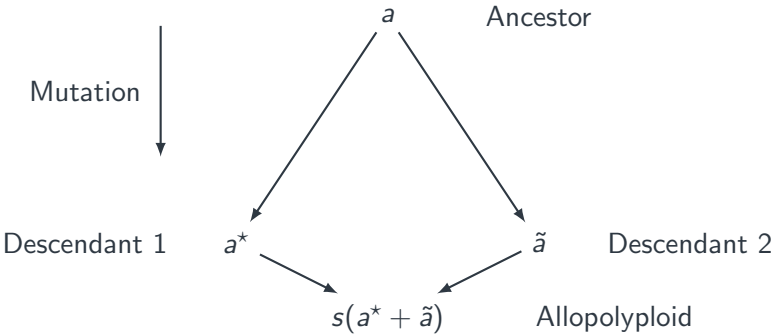


s = subfunctionalization coefficient

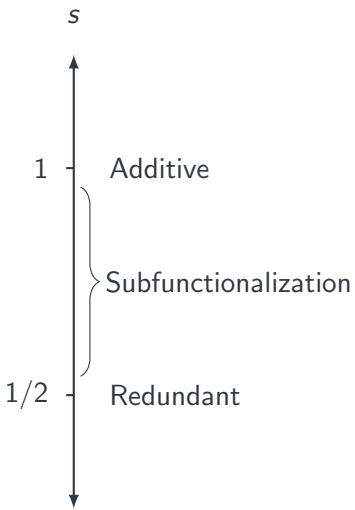


Subfunctionalization Epistasis

Let a be the effect of a functional allele (or haplotype),



s = subfunctionalization coefficient

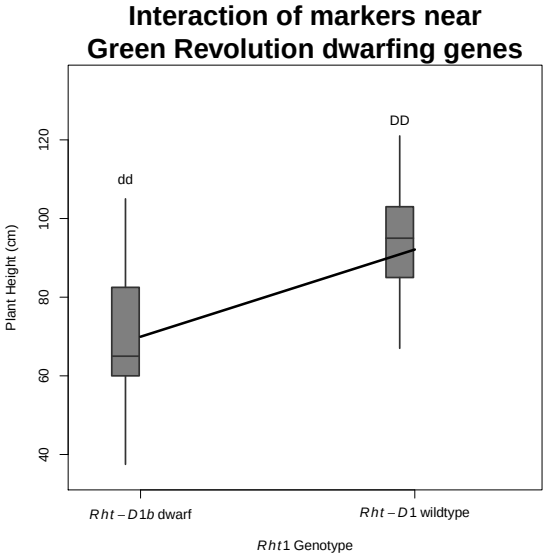


Caledonia $\times$ NY8080
Rht-B1 $\times$ *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes

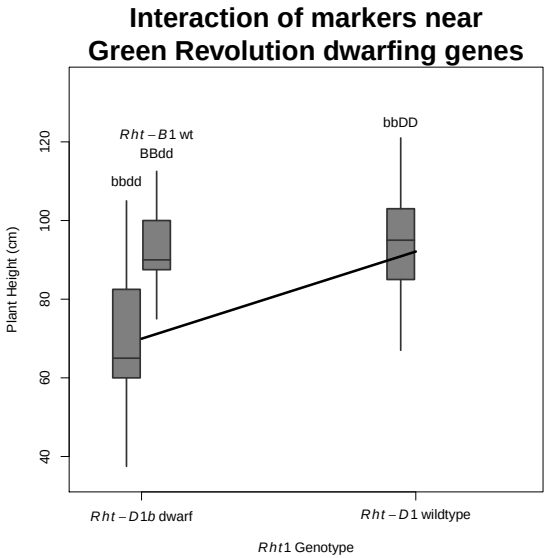
Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes



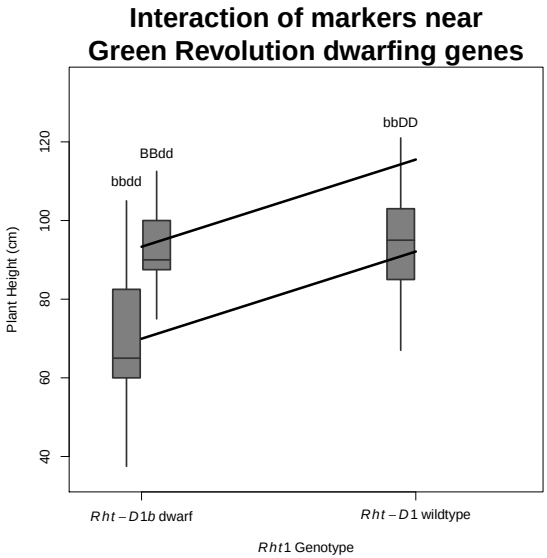
Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes



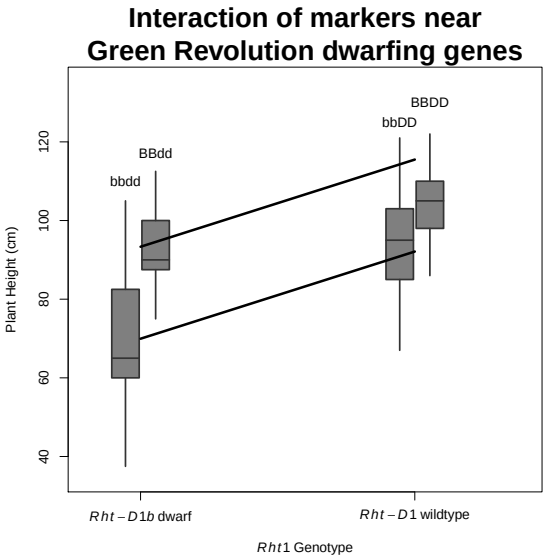
Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes



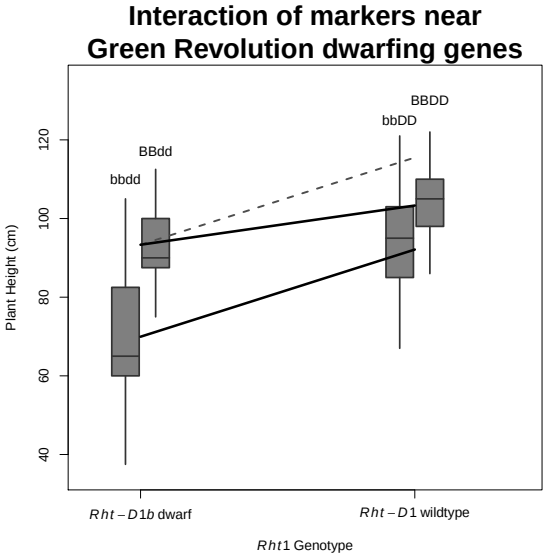
Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes



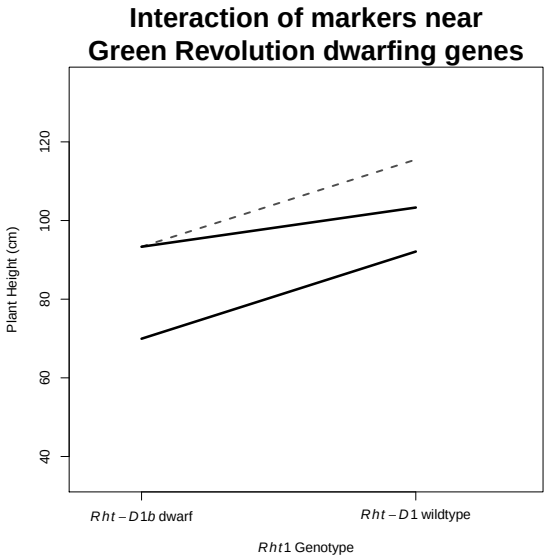
Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes



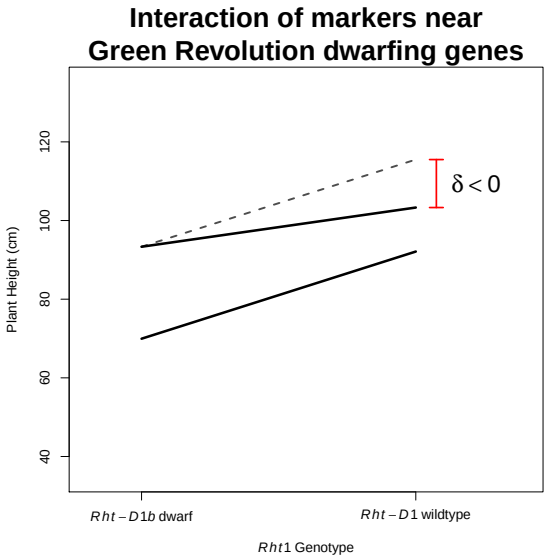
Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes



Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes

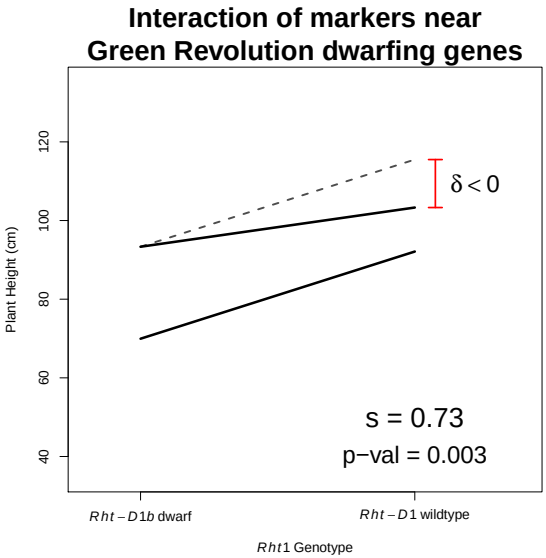


Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes

Subfunctionalization of Dwarfing Genes

- $1 + 1 \neq 2$
- functional redundancy

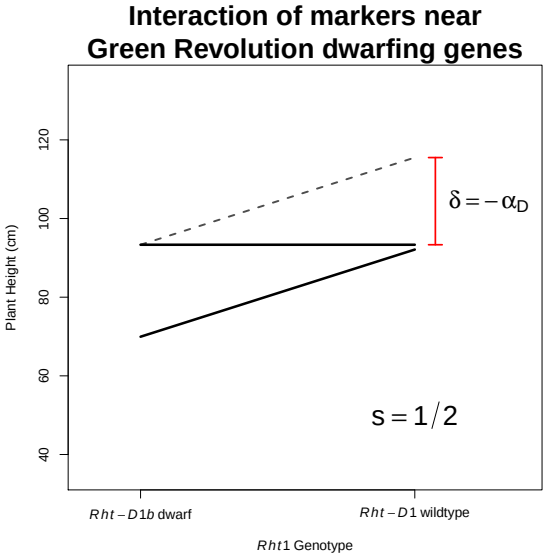


Caledonia × NY8080
Rht-B1 × *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes

Subfunctionalization of Dwarfing Genes

- $1 + 1 \neq 2$
- functional redundancy



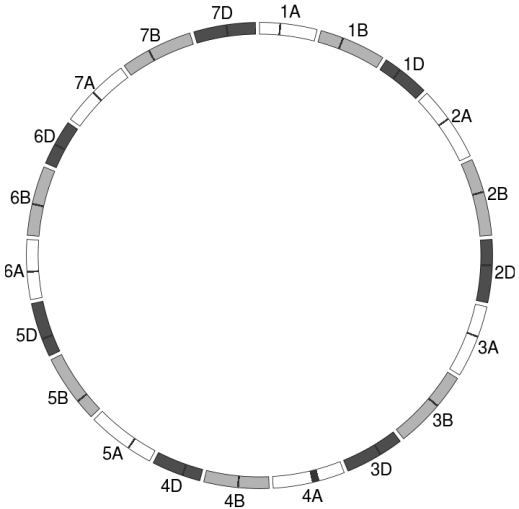
Caledonia $\times$ NY8080
Rht-B1 $\times$ *Rht-D1*

- 158 RILs
- Segregating for two homeologous dwarfing genes

Subfunctionalization of Dwarfing Genes

- $1 + 1 \neq 2$
- functional redundancy

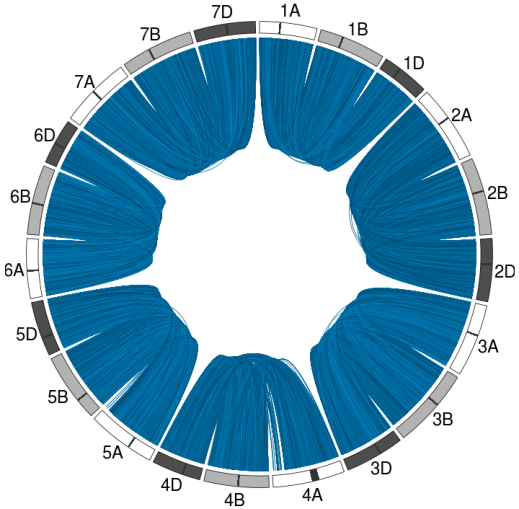
Homeoallelic Marker Sets



IWGSC RefSeq v1.0 genome

- 110,790 coding sequences
- Align CDS to self

Homeoallelic Marker Sets



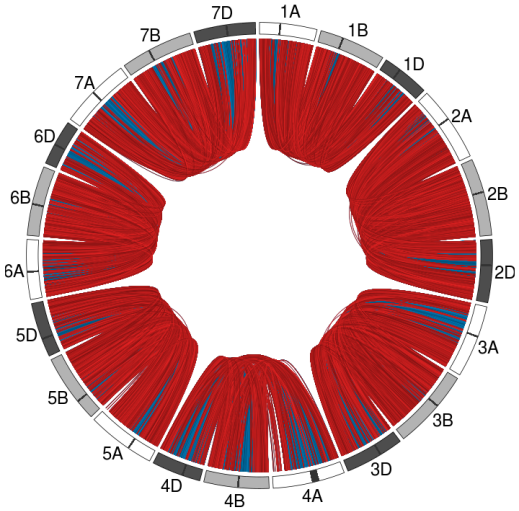
IWGSC RefSeq v1.0 genome

- 110,790 coding sequences
- Align CDS to self

Alignments

- 24,695 singletons, 20,319 multi-align
- 23,796 homeologous gene sets
  - 18,184 triplicates
  - 5,612 duplicates
  - $\sim$ 60% gene space

Homeoallelic Marker Sets



IWGSC RefSeq v1.0 genome

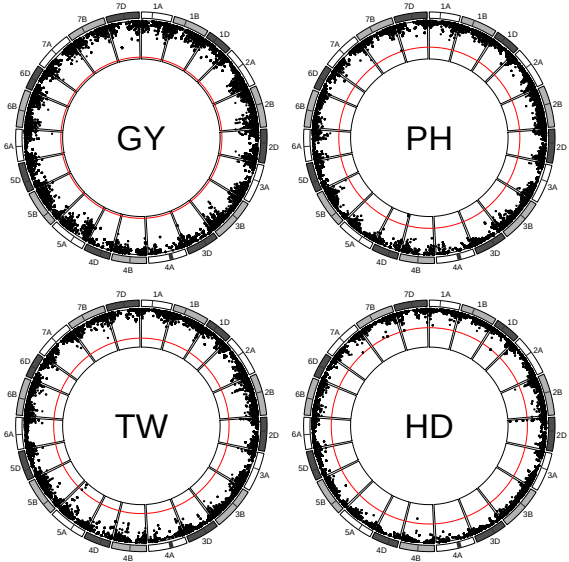
- 110,790 coding sequences
- Align CDS to self

Alignments

- 24,695 singletons, 20,319 multi-align
- 23,796 homeologous gene sets
  - 18,184 triplicates
  - 5,612 duplicates
  - $\sim$ 60% gene space

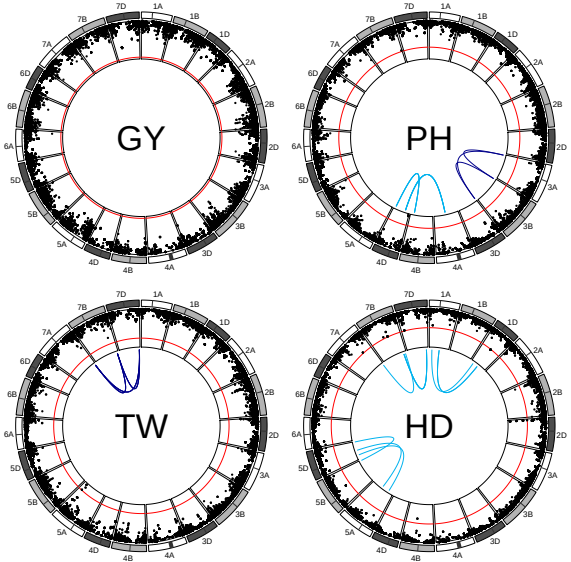
Anchor closest GBS marker

- 10,172 marker sets
  - 6,142 triplicates
  - 3,985 duplicates



Additive GWAS

- Few large effect QTL
- But high prediction accuracy
  - Many small effect loci

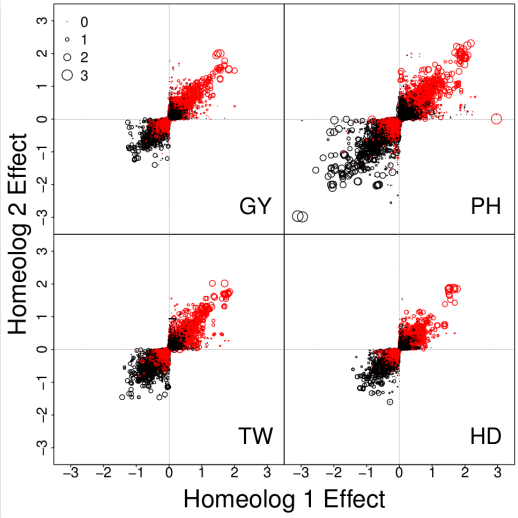


Additive GWAS

- Few large effect QTL
- But high prediction accuracy
  - Many small effect loci

Homeologous Epistasis GWAS

- Few large effect interactions
- Pattern genome-wide?
- Increase in prediction accuracy?



Additive vs 2-way Interactions

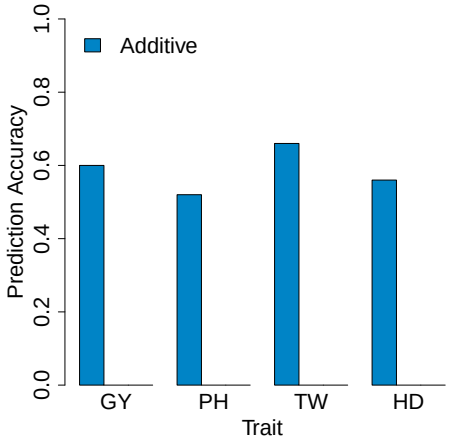
- Negative trend
- $\frac{1}{2} < d < 1$
- "less than additive"

Subfunctionalization

- redundant function

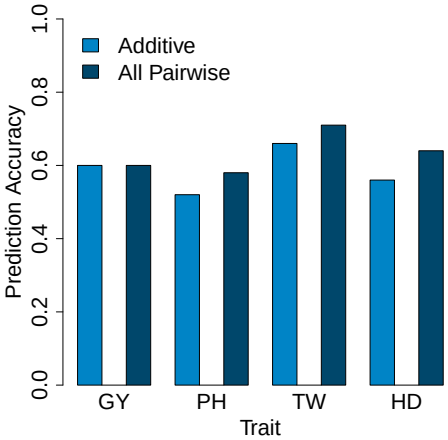
Homeologous interactions explain much of non-additive genetic signal

- How much non-additive genetic signal is explained by homeologous interactions?



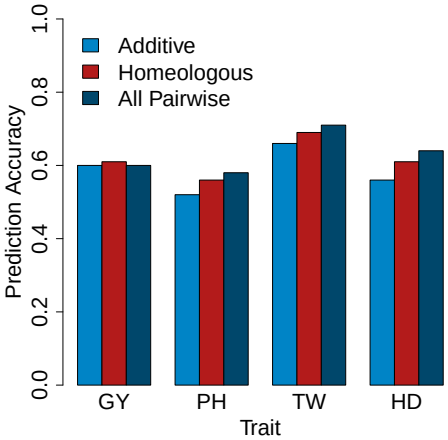
Homeologous interactions explain much of non-additive genetic signal

- How much non-additive genetic signal is explained by homeologous interactions?



Homeologous interactions explain much of non-additive genetic signal

- How much non-additive genetic signal is explained by homeologous interactions?
- $\sim$ 60-75%



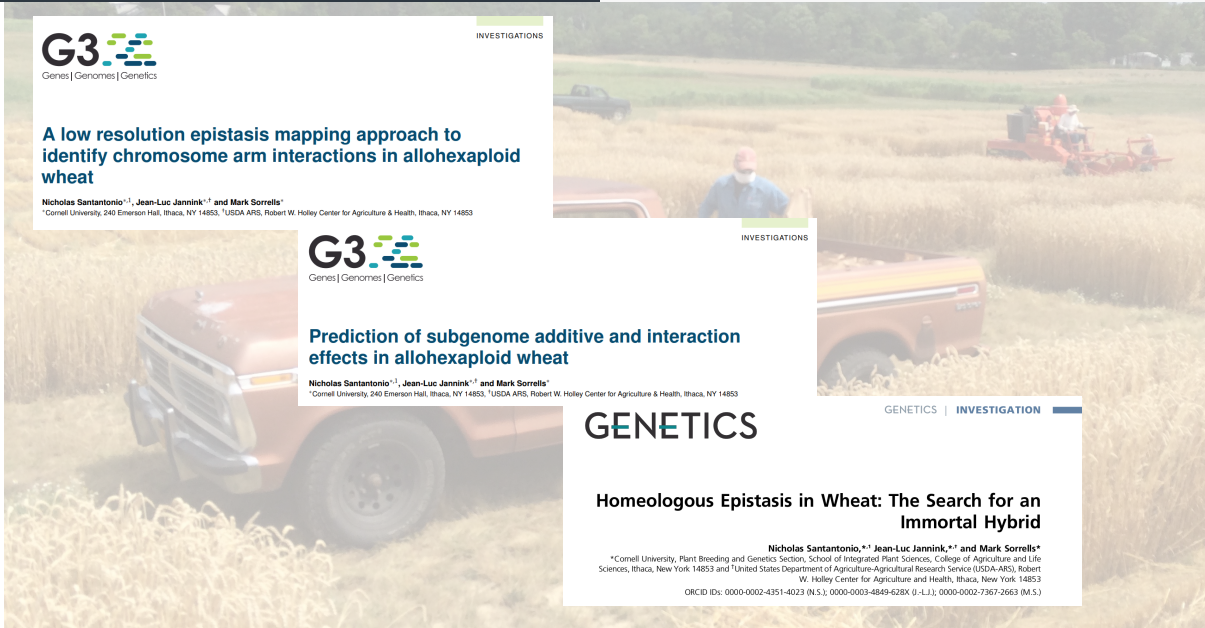
Summary of Homeoallelic Epistasis

Opportunity to fix advantageous homeoallelic pairs

- Establish heterozygosity across subgenomes
- Immortalize through inbreeding

Capsicum, a large genome in flux?

- Find and fix interacting loci?
  - Is *Phytopthera* resistance an epistatic network?
  - What about capsaicinoid variability?
- What role do transposons play?
  - Somatic mutations driving variability?



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

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, NY 14853, [†]USDA ARS, Robert W. Holley Center for Agriculture & Health, Ithaca, NY 14853

GENETICS

GENETICS | INVESTIGATION

Homeologous Epistasis in Wheat: The Search for an Immortal Hybrid

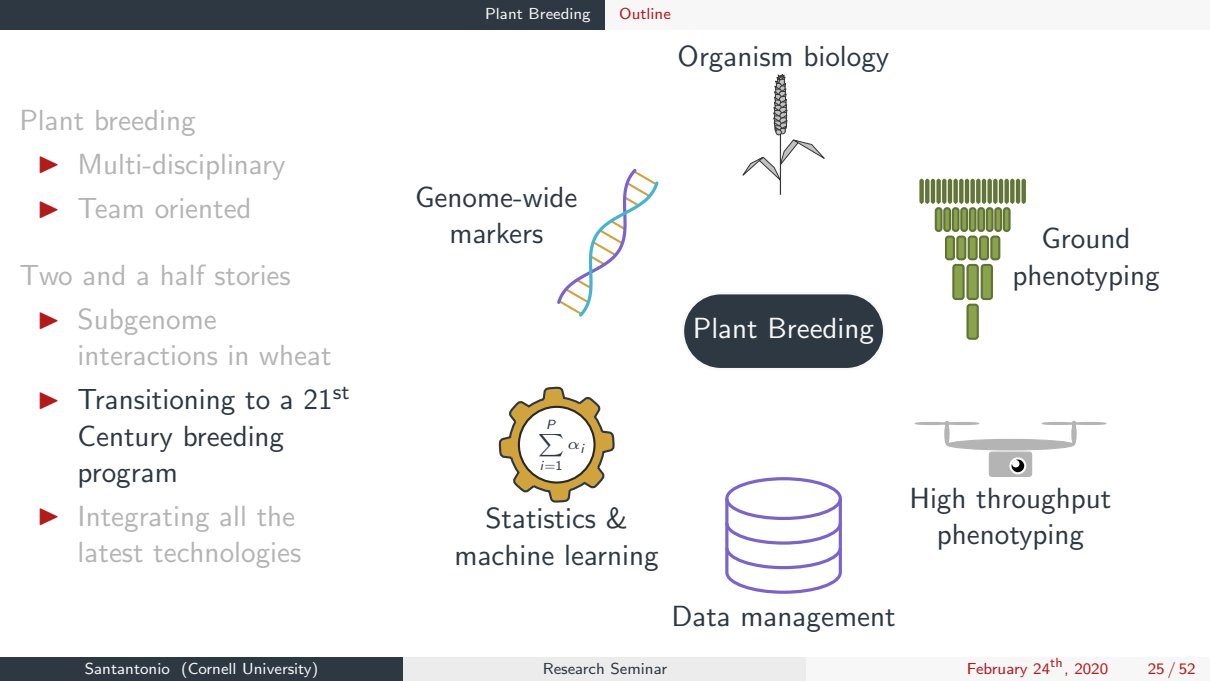
Nicholas Santantonio,*,1 Jean-Luc Jannink,*,† and Mark Sorrells*
*Cornell University, Plant Breeding and Genetics Section, School of Integrated Plant Sciences, College of Agriculture and Life Sciences, Ithaca, New York 14853 and [†]United States Department of Agriculture-Agricultural Research Service (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.)

Plant breeding

- Multi-disciplinary
- Team oriented

Two and a half stories

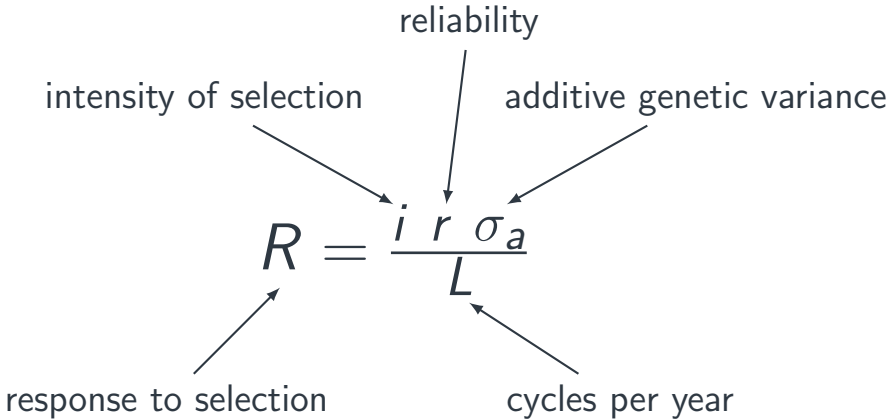
- Subgenome interactions in wheat
- Transitioning to a 21st Century breeding program
- Integrating all the latest technologies



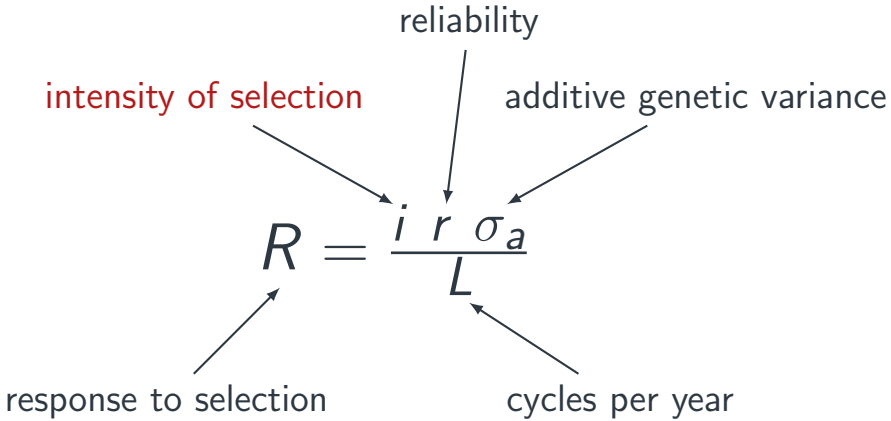
How can we best exploit the breeder's equation?

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

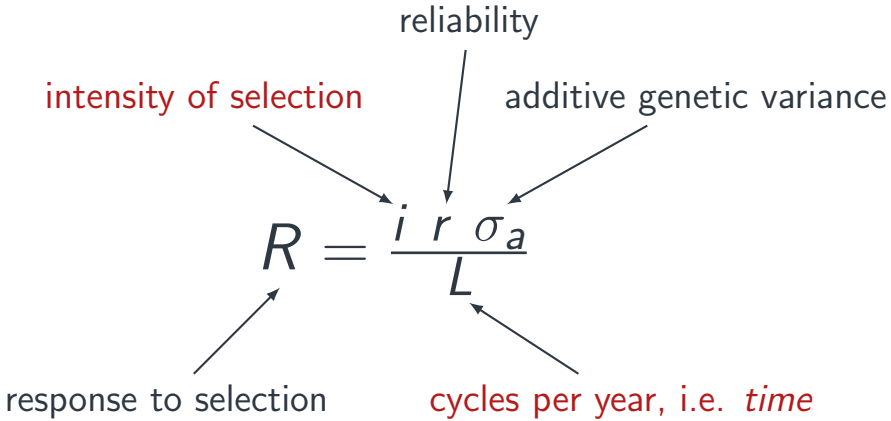
How can we best exploit the breeder's equation?



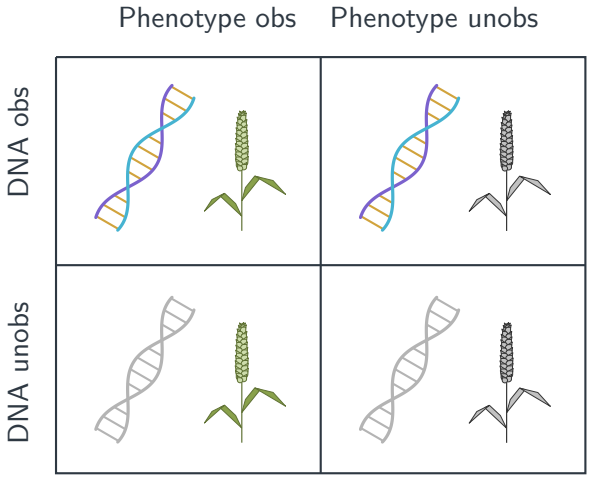
How can we best exploit the breeder's equation?



How can we best exploit the breeder's equation?



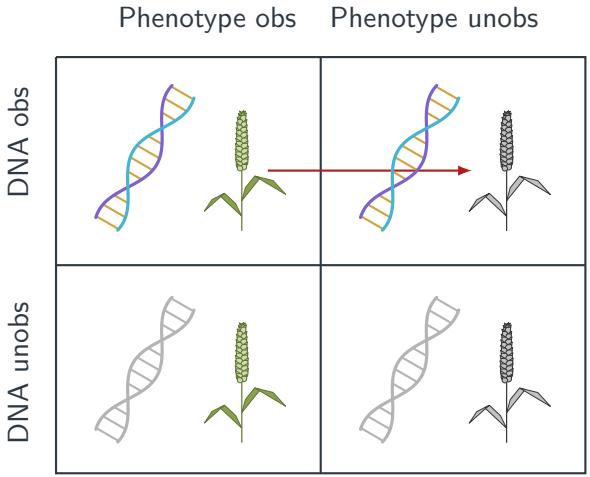
Incorporating Genomic Selection



Genomic prediction

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

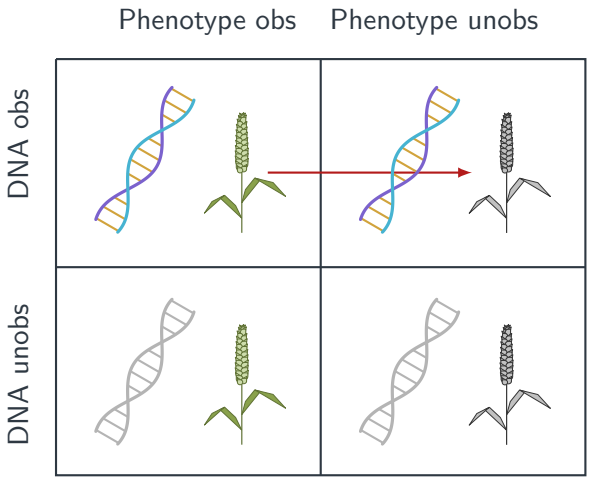
Incorporating Genomic Selection



Genomic prediction

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

Incorporating Genomic Selection



Genomic prediction

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

Genomic selection

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

A traditional breeding program

Variety Development Pipeline

10,000's, 1 Env, unrep  Year 1

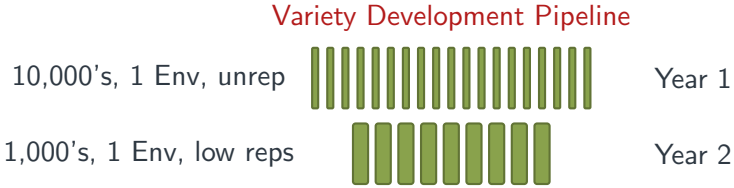
Traditional Breeding

- Accurate
  - Extensive testing

A traditional breeding program

Traditional Breeding

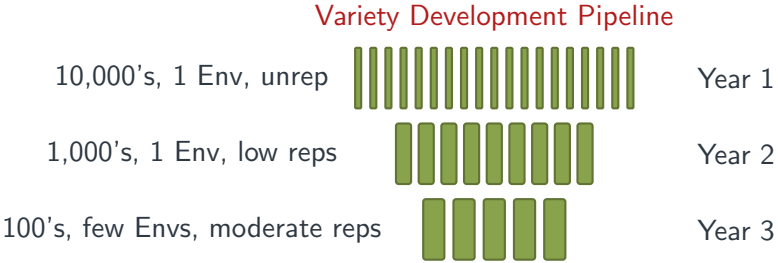
- Accurate
  - Extensive testing



A traditional breeding program

Traditional Breeding

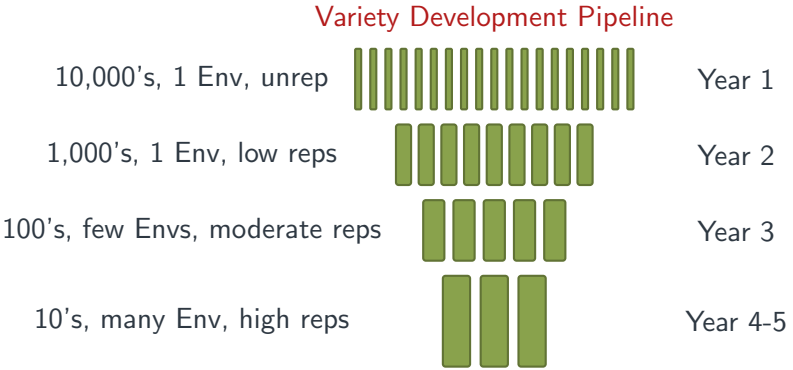
- Accurate
  - Extensive testing



A traditional breeding program

Traditional Breeding

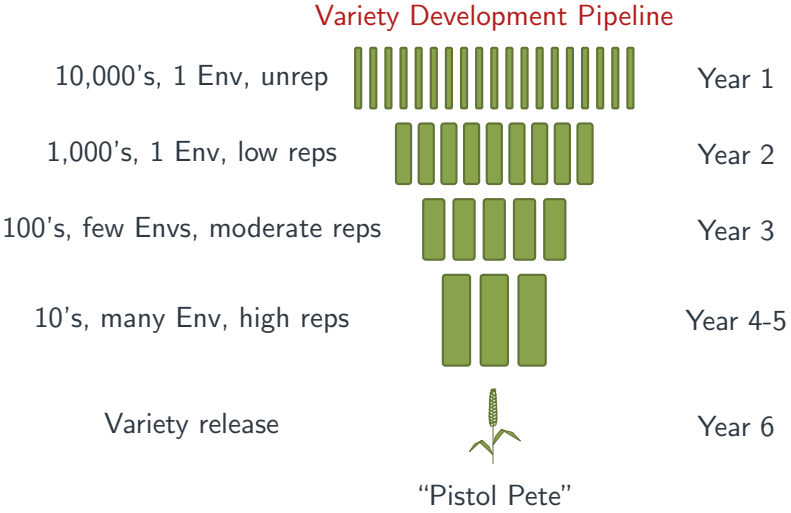
- Accurate
  - Extensive testing



A traditional breeding program

Traditional Breeding

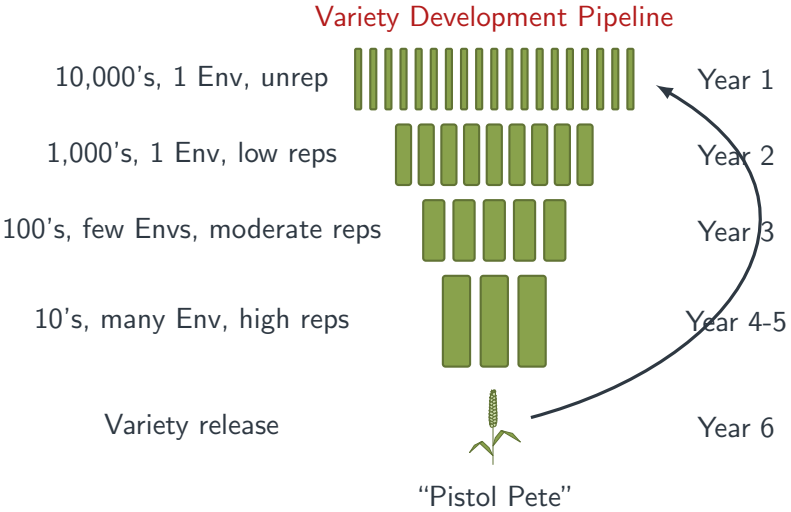
- Accurate
  - Extensive testing



A traditional breeding program

Traditional Breeding

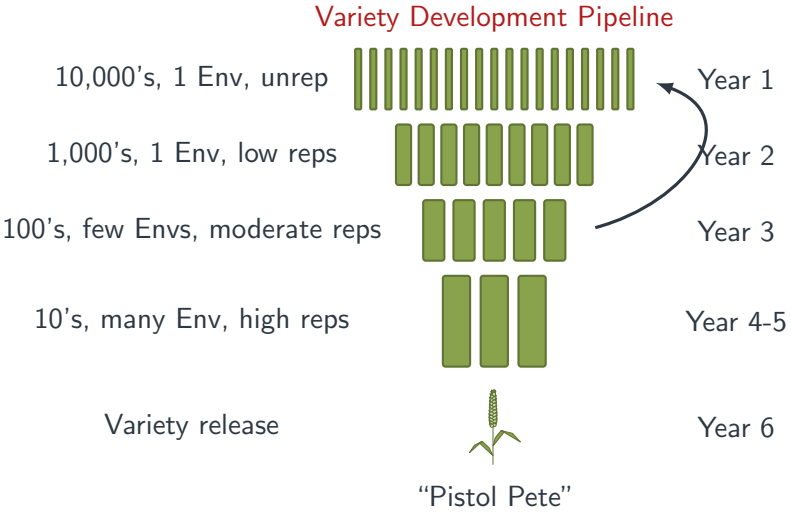
- Accurate
  - Extensive testing
- Slow
  - Multiple trial years
  - Long generation intervals



A traditional breeding program

Traditional Breeding

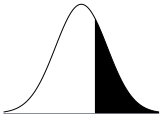
- Accurate
  - Extensive testing
- Slow
  - Multiple trial years
  - Long generation intervals



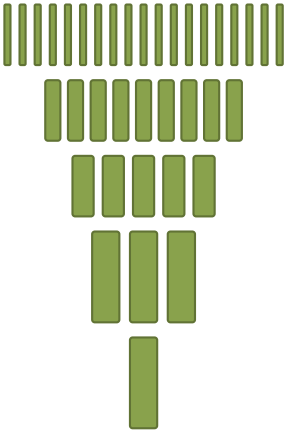
Increasing population size increases intensity

Genomic Selection

- Less accurate
  - Less phenotypic information



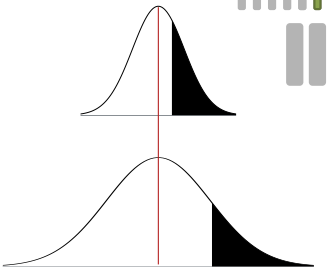
Variety Development Pipeline



Increasing population size increases intensity

Genomic Selection

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

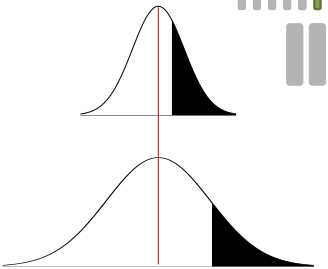


Variety Development Pipeline

Increasing population size increases intensity

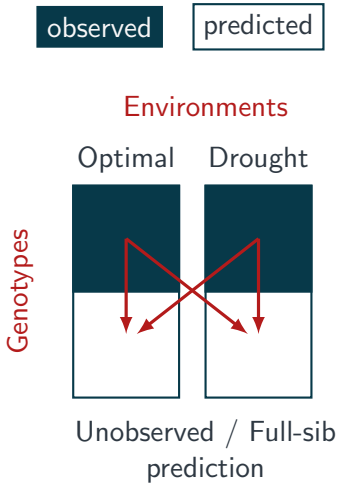
Genomic Selection

- 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

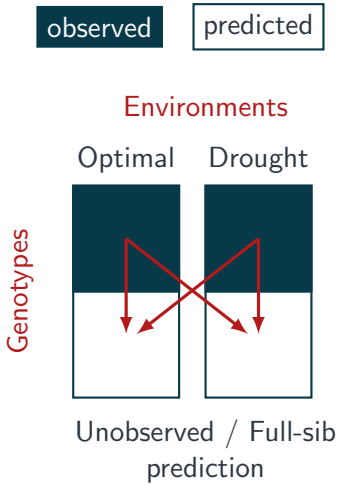
Current model for genomic prediction works...

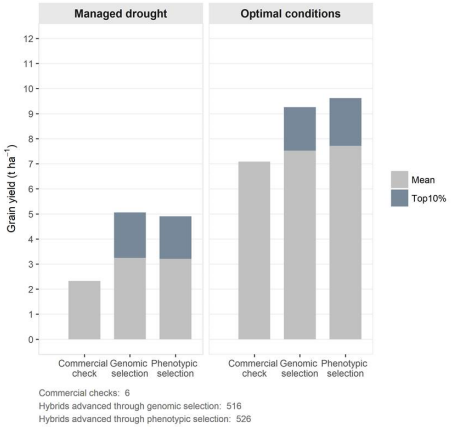


1. Observe half of population in both environments
2. Estimate genetic correlation of environments
3. Predict other half

Current model for genomic prediction works...

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



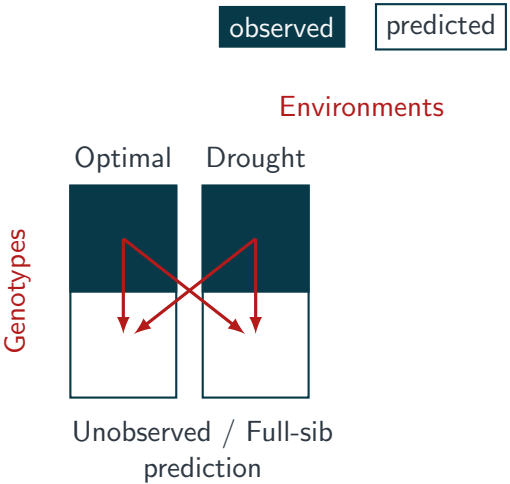


Beyene et al. 2020

... but can we do better?

Genetic relationships known

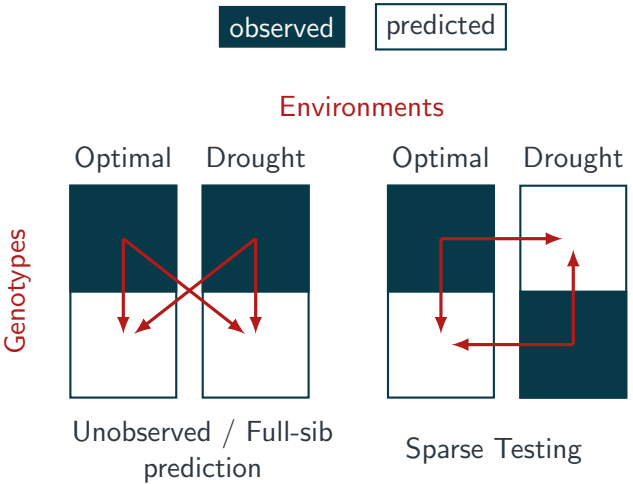
- estimate genetic correlation of environments without replicating across
- Get to observe every line



... but can we do better?

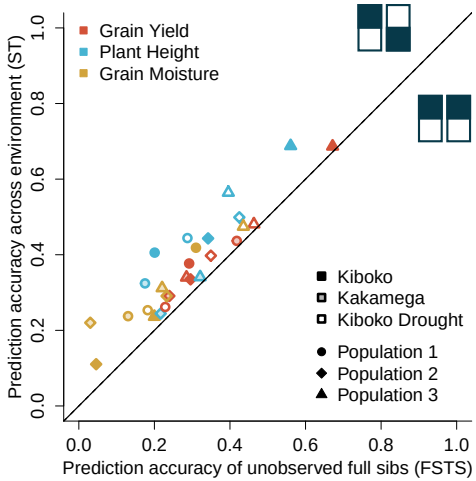
Genetic relationships known

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



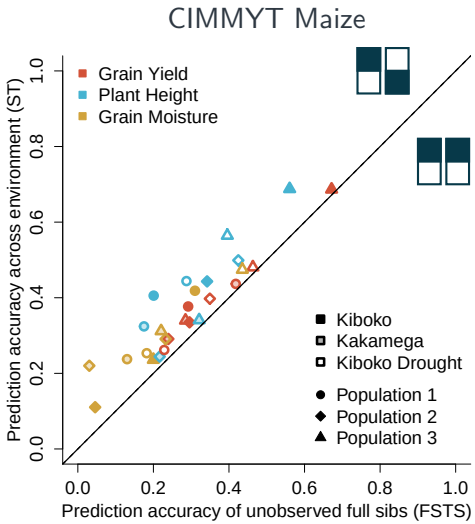
Sparse testing is (almost) uniformly superior to unobserved prediction

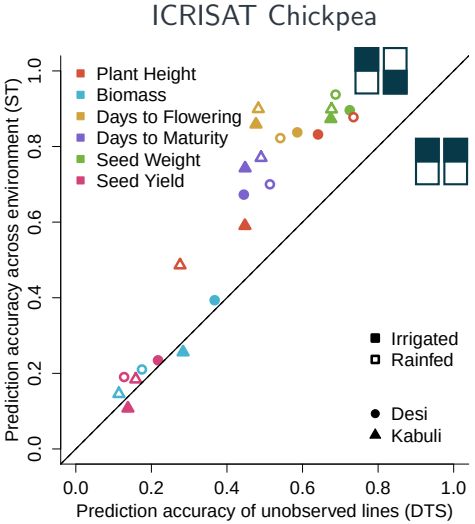
CIMMYT Maize



ICRISAT Chickpea

Sparse testing is (almost) uniformly superior to unobserved prediction

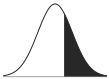




Rapid cycling greatly decreases L

Rapid Cycling

- Assume 3 cycles/year in greenhouse



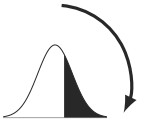
Variety Development Pipeline

Year 1

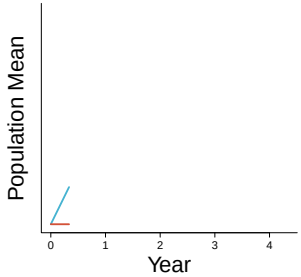
Rapid cycling greatly decreases L

Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

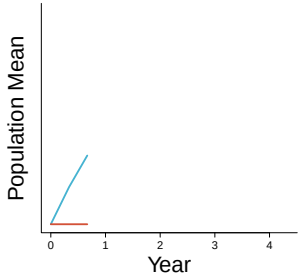


Rapid cycling greatly decreases L

Rapid Cycling

- Assume 3 cycles/year in greenhouse

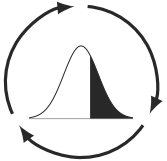
Variety Development Pipeline



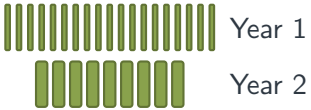
Rapid cycling greatly decreases L

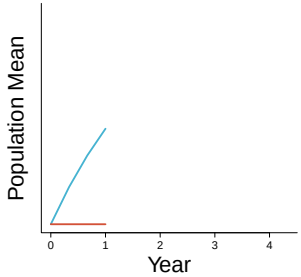
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

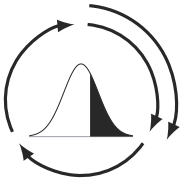




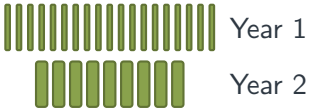
Rapid cycling greatly decreases L

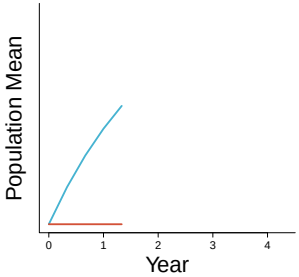
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

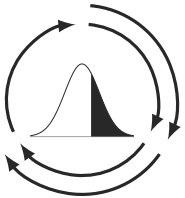




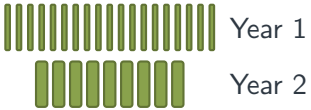
Rapid cycling greatly decreases L

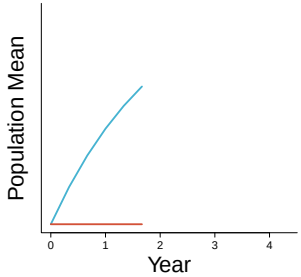
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

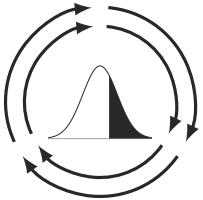




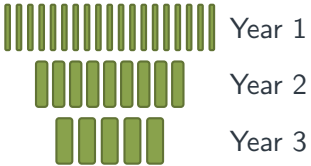
Rapid cycling greatly decreases L

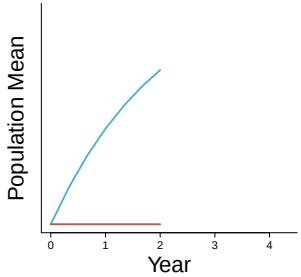
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

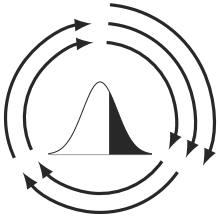




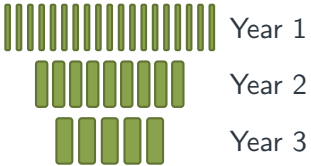
Rapid cycling greatly decreases L

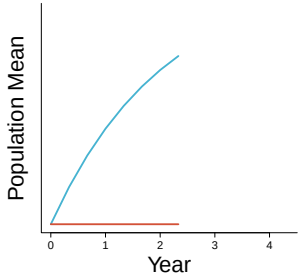
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

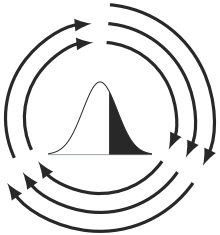




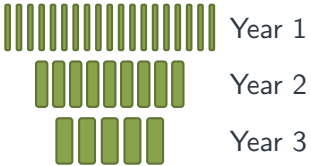
Rapid cycling greatly decreases L

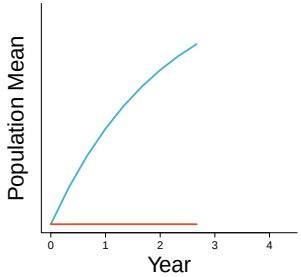
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

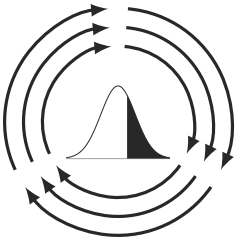




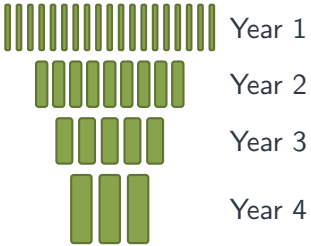
Rapid cycling greatly decreases L

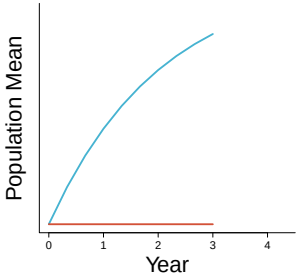
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

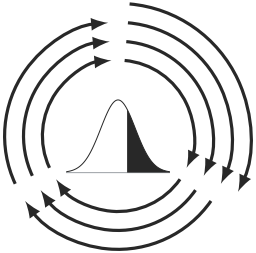




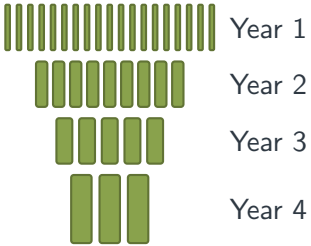
Rapid cycling greatly decreases L

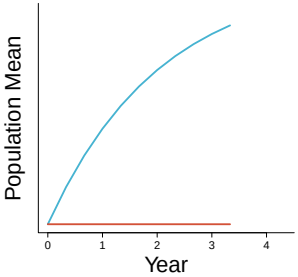
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

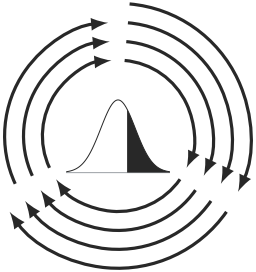




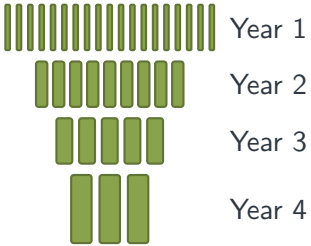
Rapid cycling greatly decreases L

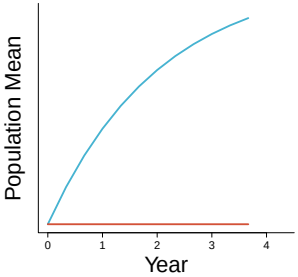
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

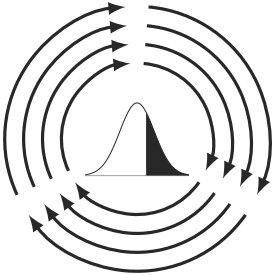




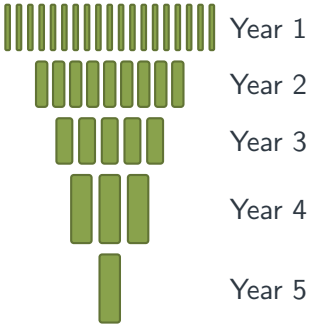
Rapid cycling greatly decreases L

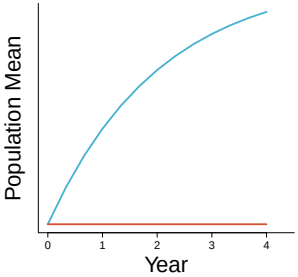
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

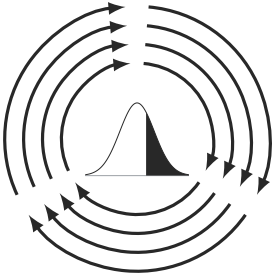




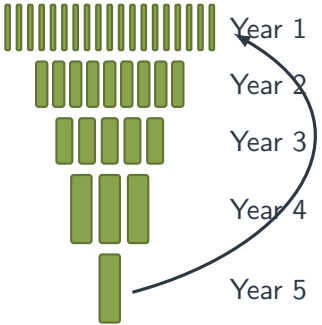
Rapid cycling greatly decreases L

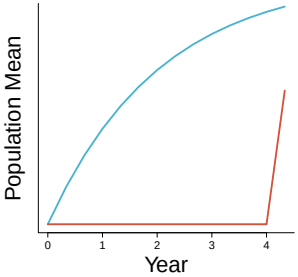
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline

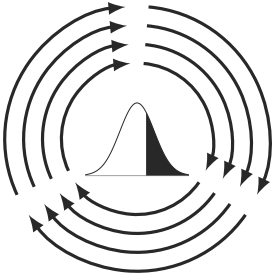




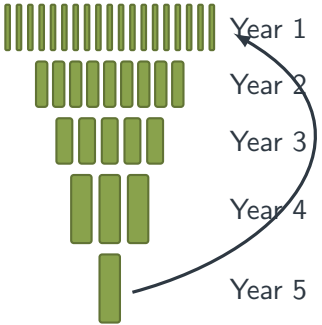
Rapid cycling greatly decreases L

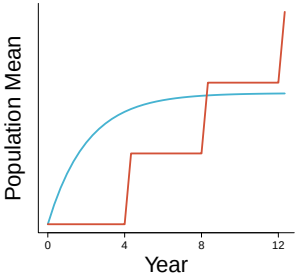
Rapid Cycling

- Assume 3 cycles/year in greenhouse



Variety Development Pipeline





Can this be implemented at scale?

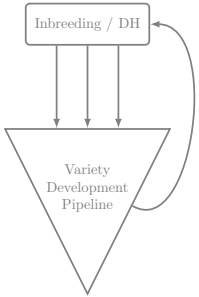
Can this be implemented at scale?

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

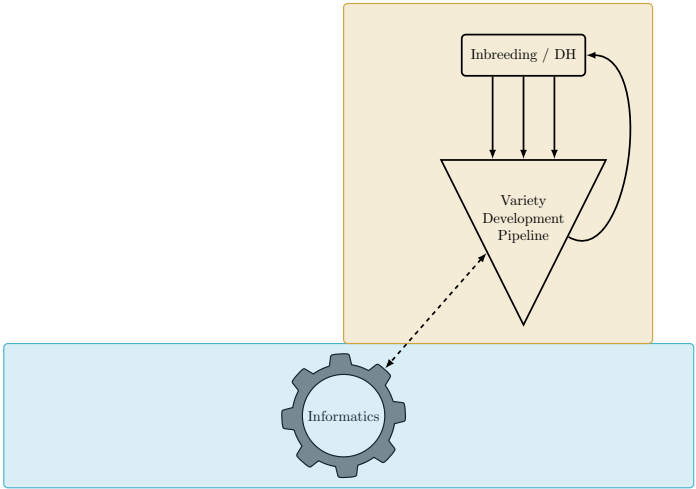


Phase 1: Informatics

- 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

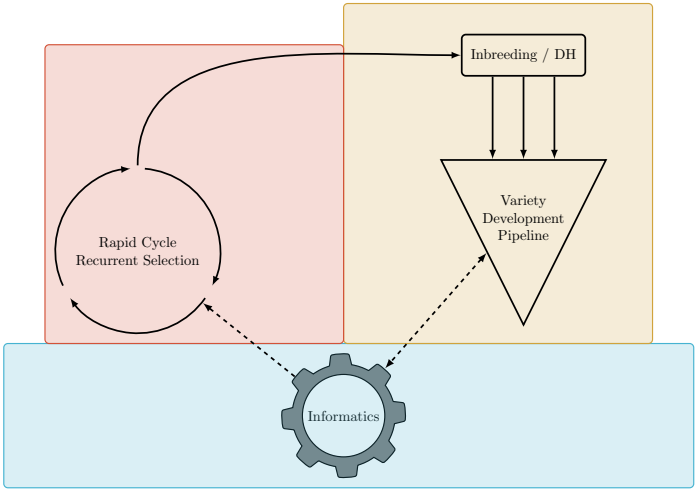
- Informatics, genotyping platform, SOPs and QC
- Genotyping all phenotyped entries to build training set
- Offset with trial designs that exploit genotypic information

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



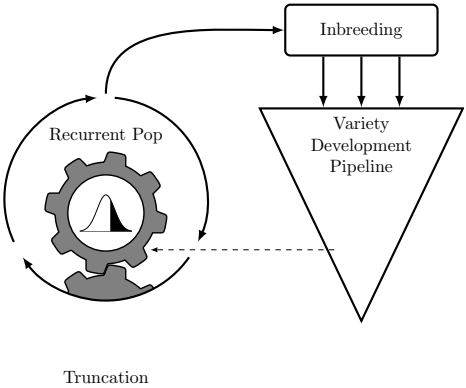
frontiers

Technology Driven Crop Improvement for Africa and South Asia.

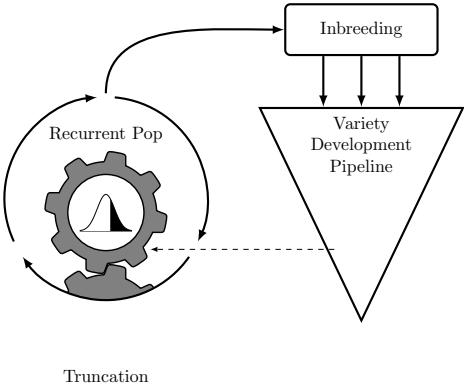
Nicholas Santantonio[1], Sikiru A. Atanda[2, 3, 1], Yoseph Beyene[3], Rajeev K. Varshney[4], Michael S. Olsen[3], Elizabeth Jones[5], Manish Roorkiwal[6], Xuecai Zhang[7], BHARADWAJ CHELLAPILLA[8], Pooran M. Gaur[9], Manje Gowda[3], Kate Dreher[7], Cludio A. Hernandez[7], Jose Crossa[7], Paulino Pérez-Rodríguez[10],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

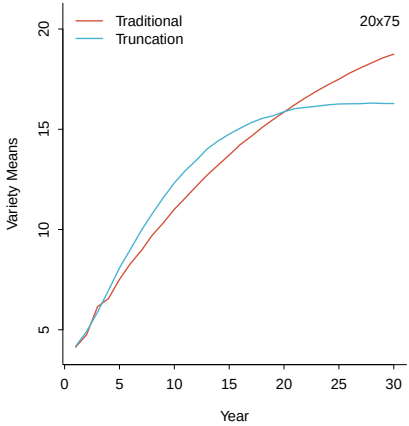
Accepted pending minor revision

Rapid cycle recurrent truncation beats traditional, at first

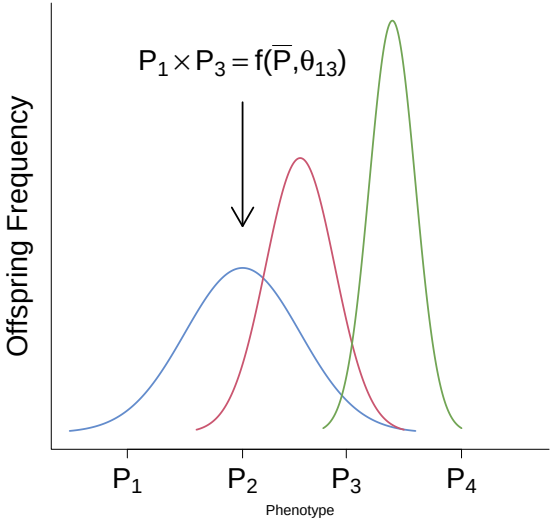


Rapid cycle recurrent truncation beats traditional, at first

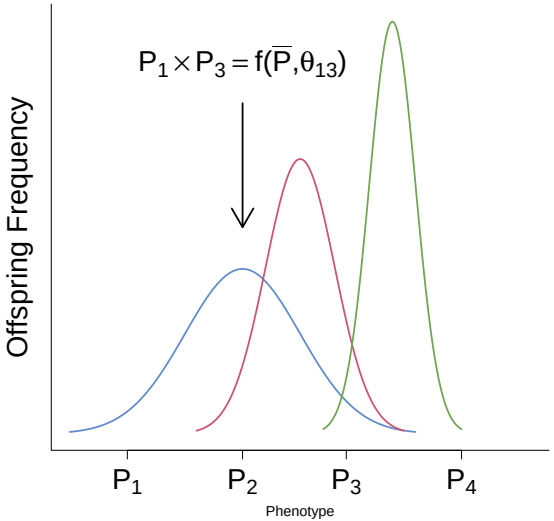


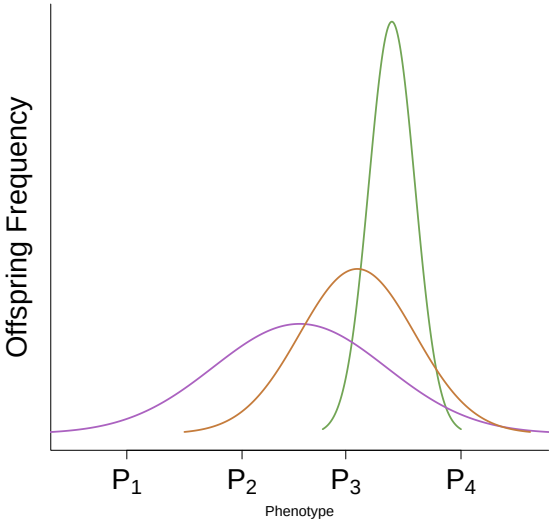


Parent selection must balance mean and variance



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

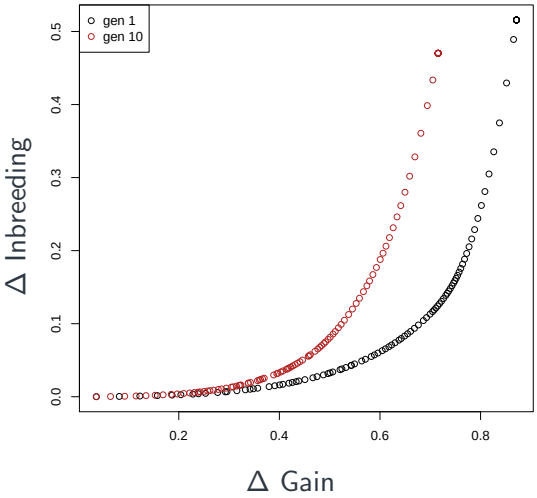
Optimal contributions of parents

Balance genetic gain and inbreeding

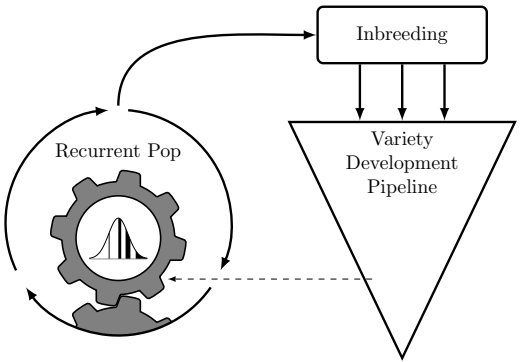
- Minimize inbreeding given desired gain
- Maximize gain given acceptable inbreeding
- See Meuwissen, 1997

Portfolio optimization problem

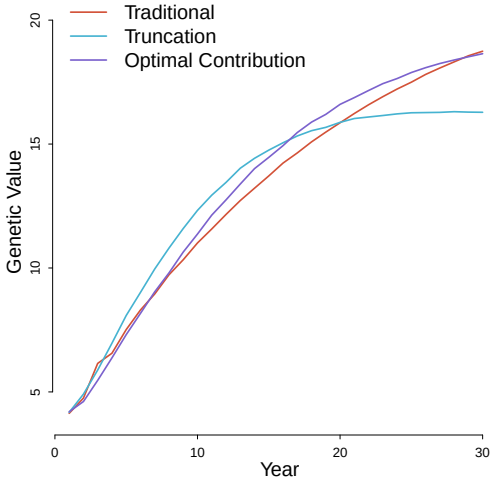
- Solve with quadratic programming
- Doesn't always pick the "best" individuals
- Parental contributions vary



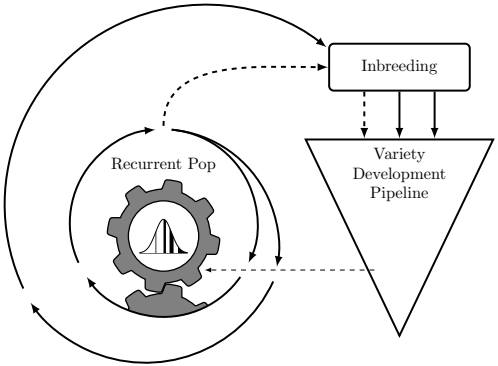
Optimal contributions can maintain V_g at the cost of short-term gains



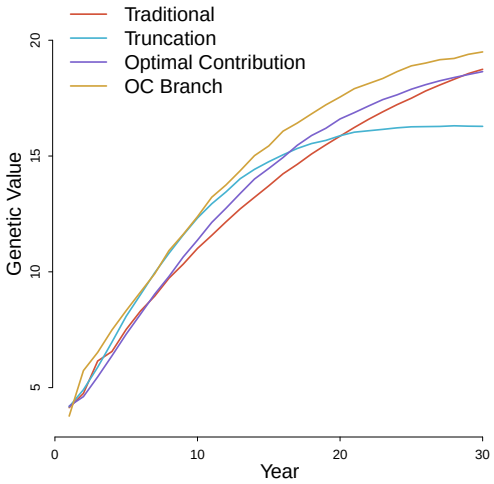
Optimal Contribution



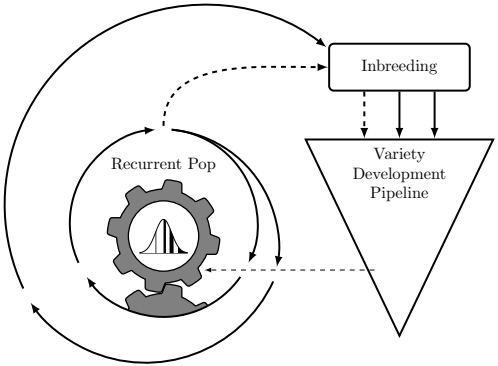
Can we have our cake and eat it too?



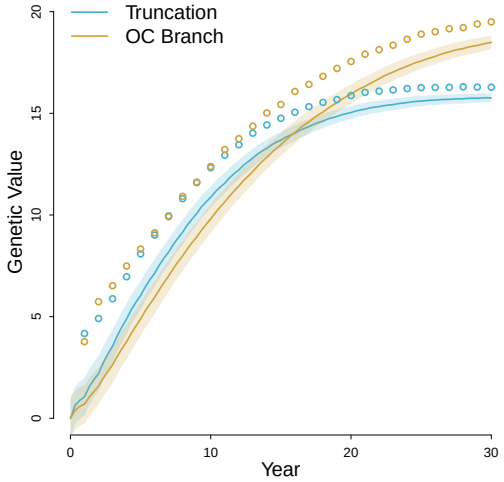
Optimal Contribution with Branching



Can we have our cake and eat it too?



Optimal Contribution with Branching



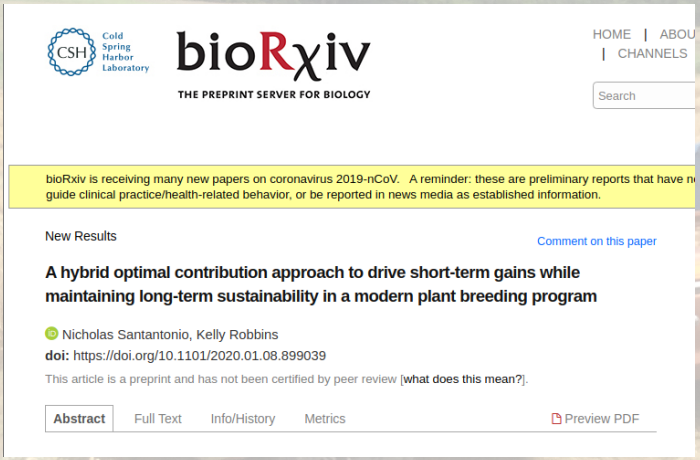
Rapid cycling

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 *Capsicum*

- Start by genotyping!
- Build foundational capacities
- Economic index for rapid improvement?


Cold Spring Harbor Laboratory
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY
HOME | ABOU
| CHANNELS
Search
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

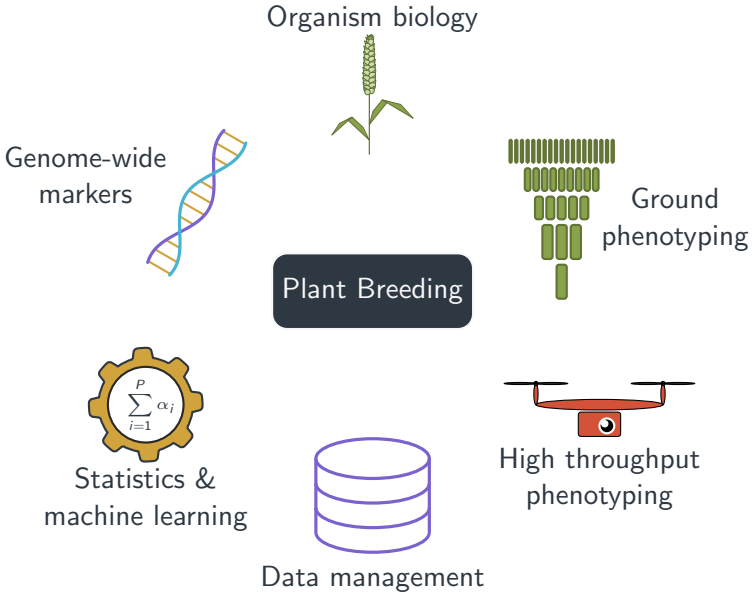
G3: In Review

Plant breeding

- Multi-disciplinary
- Team oriented

Two and a half stories

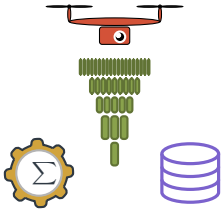
- Subgenome interactions in wheat
- Transitioning to a 21st Century breeding program
- Integrating the latest technologies



Alfalfa

- Autotetraploid ($4x = 2n = 32$)
- Perennial forage (3-4 years)
- 3-10 cuts per year
- Self-incompatible
- High inbreeding depression
- Varieties are (synthetic) populations!

Alfalfa



High phenotypic burden

- Low h^2 = High reps
- Multi-harvest, multi-year
- High throughput phenotyping
- Genomic selection

How to genotype?

- Varieties/lines are populations
- Which/how many individuals?

US Alfalfa Farmer Research Initiative (USAFRI) Grant

Evaluating Approaches to High-Throughput Phenotyping and Genotyping for Genomic Selection in Alfalfa

Population-level genotyping

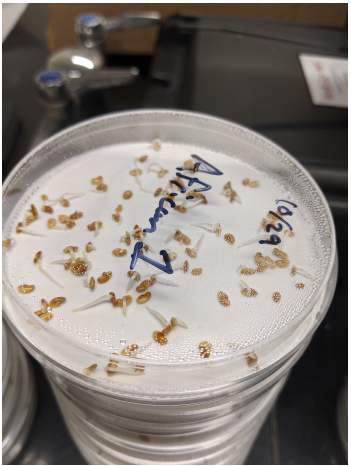
- Pool/bulk seeds for sequencing

Whole genome re-sequencing 50×

- 9 historic germplasm sources
- 8 Cornell varieties

Align to alfalfa genome

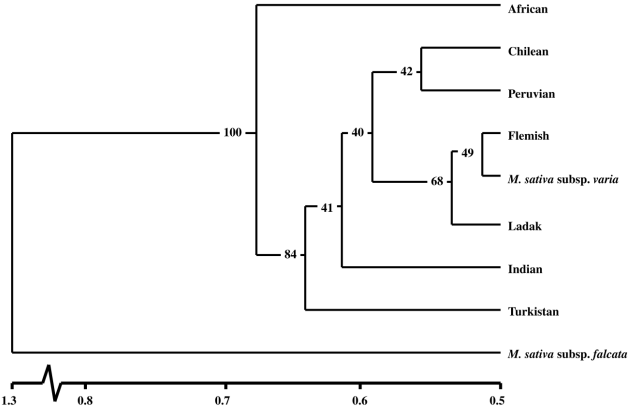
- Call variants, count alleles in population
- Calculate relationship from allele frequencies



Validation in diallel of 9 historic germplasm sources

AFLPs (Segovia-Lerma 2003, 2004)

Fig. 1. UPGMA dendrogram based on genetic distances as estimated by 1541 AFLPs among nine alfalfa populations. Numbers at branch points indicate support for groupings to the right of the number; values are percent of bootstrap datasets that exhibited the cluster.



Kinship in the diallel using AFLPs

Half-diallel

- Forage Yield
  - 1997–1998
  - 5 cuts/yr
- 1,544 AFLPs

Diallel

- 9 parent populations
- 36 hybrids families

Kinship in the diallel using AFLPs

Half-diallel

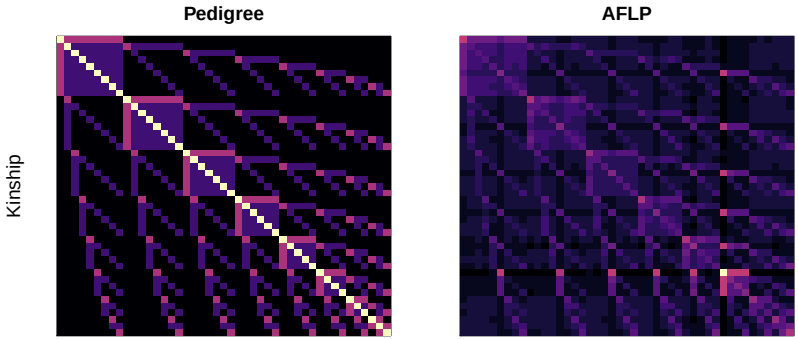
- Forage Yield
  - 1997–1998
  - 5 cuts/yr
- 1,544 AFLPs

Diallel

- 9 parent populations
- 36 hybrids families

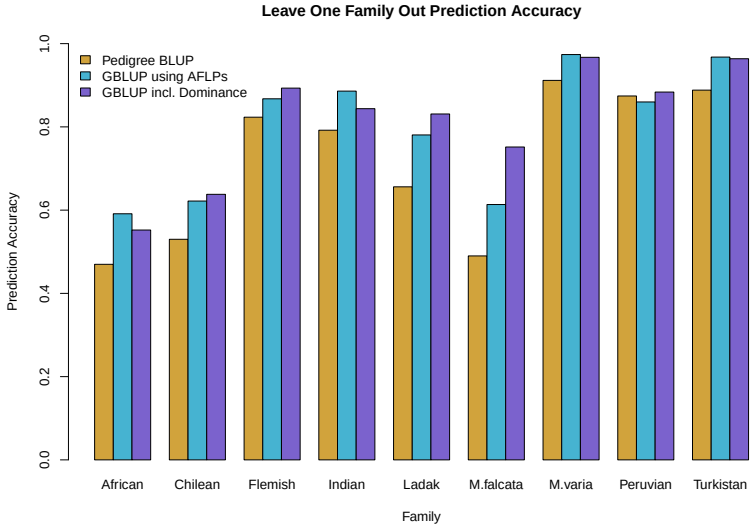
AFLPs track pedigree well

- Dominant
- No dosage



Genomic prediction with AFLPs

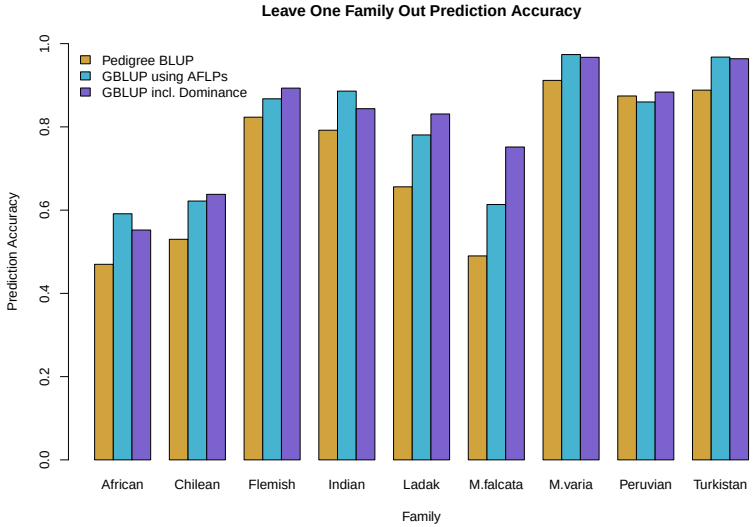
AFLPs increase prediction accuracy



Genomic prediction with AFLPs

AFLPs increase prediction accuracy

- Does dosage help prediction?
- Validate PopGS strategy

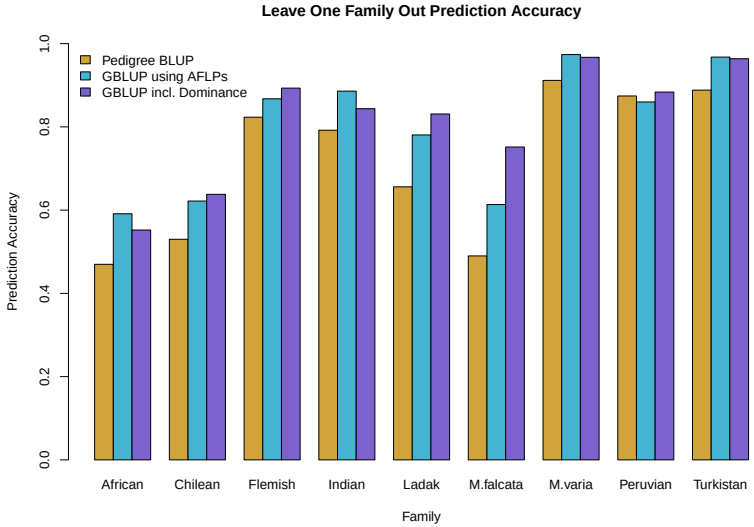


Genomic prediction with AFLPs

AFLPs increase prediction accuracy

- Does dosage help prediction?
- Validate PopGS strategy

What about population allele frequencies?
TBD



Building genotype specific growth curves to understand G×E

Moving Forward

https://www.pinterest.com/pin/473722454525419585/

Changing *Capsicum* plant architecture

- Open canopy structure
- Coordinated fruit set

Machine learning plant growth and development

- Ground and aerial images
- 3D architecture reconstruction
- Train on mechanical harvesting output

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