



Genomic Selection

Prof. Kaio Olimpio das Graças Dias (kaio.o.dias@ufv.br)

Department of General Biology

Federal University of Viçosa - Brazil

Outline

- ▶ Motivation
 - ▶ From Phenotype to Genotype
- ▶ Genomic Selection
 - ▶ Molecular Markers
 - ▶ Statistical Models
- ▶ Application
 - ▶ Perennial Crops
 - ▶ Annual Crops
- ▶ References

Opportunities

On the road of Breeding 4.0

Breeding 1.0



Breeding 2.0



Breeding 3.0



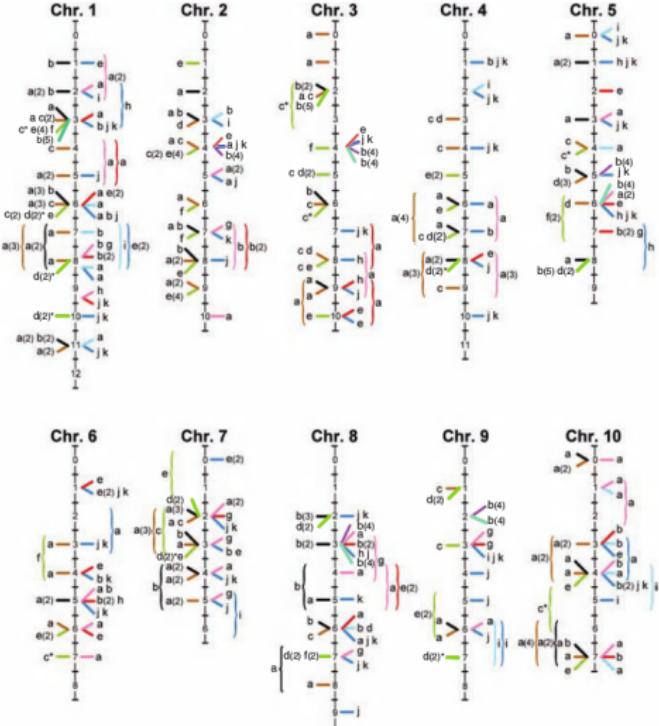
Breeding 4.0



(Wallace, 2018)

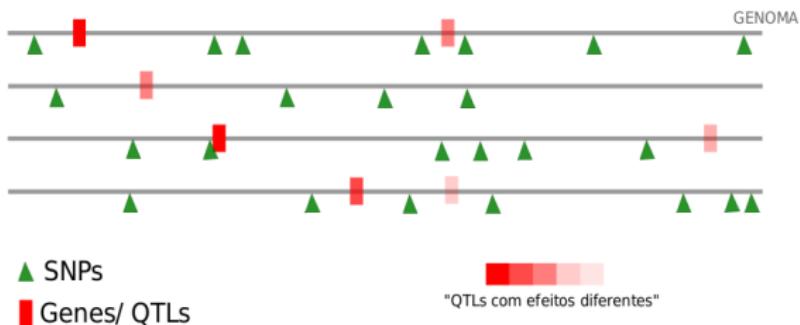
Genetic Architecture of Drought Tolerance

UFV



Meuwissen et al., 2001

"With a dense marker map some markers will be very close to the QTL and probably in linkage disequilibrium with it. Therefore, some marker alleles will be correlated with positive effects on the quantitative trait across all families and can be used for selection without the need to establish linkage phase in each family".



Habier, 2007, 2013

LD SNP and QTL, Cosegregation, Additive-genetic relationships

Linkage Disequilibrium

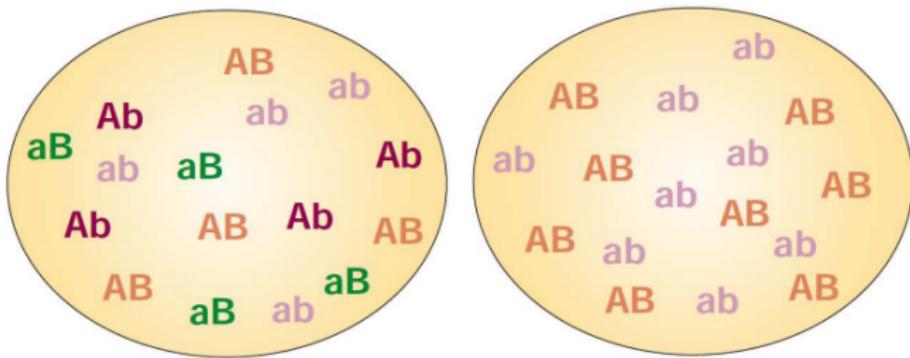
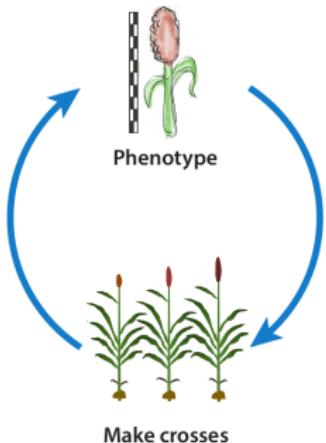


Figura 1 – Considere dois locos A e B , ambos bialélicos (A e a ; B e b). Quatro tipos gaméticos são possíveis: AB , Ab , aB e ab . Se esses locos estiverem em equilíbrio de ligação (à esquerda), os quatro gametas são observados. No entanto, se esses locos estiverem em máximo desequilíbrio de ligação (à direita), apenas os gametas em associação (AB e ab) são observados (MACKAY, 2001b)

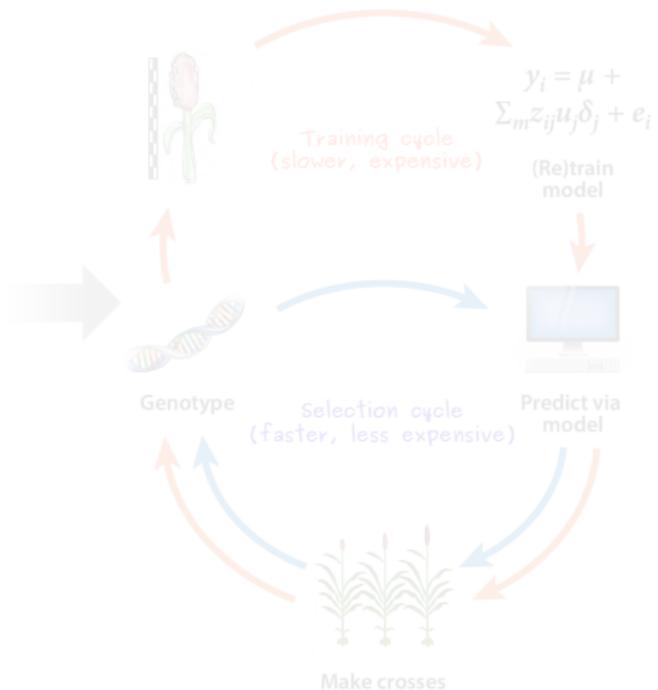
Introduction

Genomic prediction

a Traditional selection



b Genomic selection

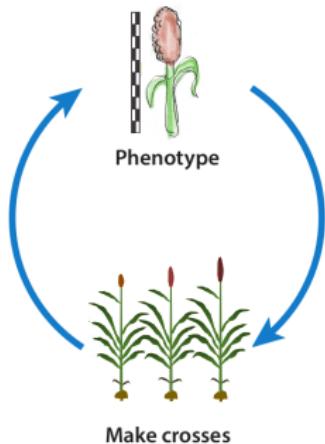


Source: adapted from Wallace, J. G. et al. 2019 and Xi, L. 2018

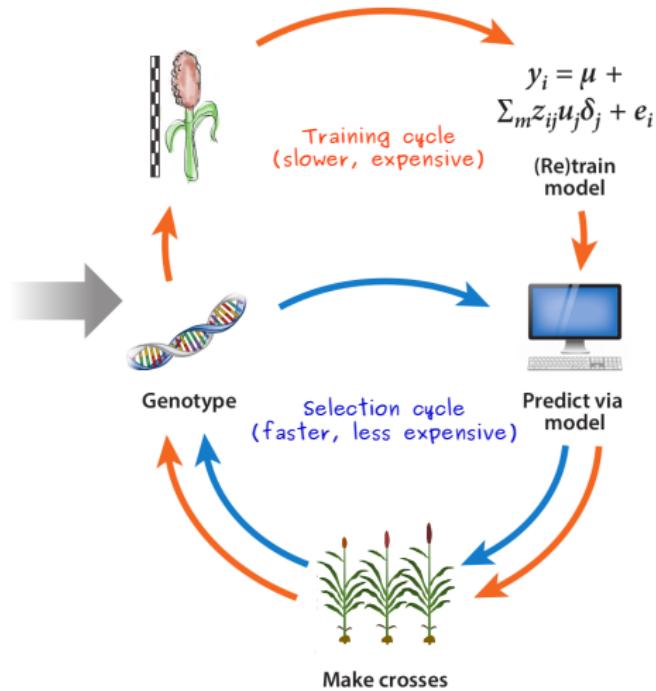
Introduction

Genomic prediction

a Traditional selection



b Genomic selection



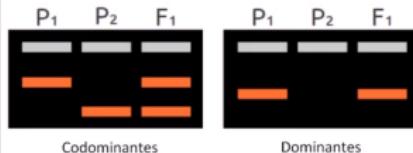
Source: adapted from Wallace, J. G. et al. 2019 and Xi, L. 2018



Foto: Dr. Fernando Correr

Marcadores moleculares

- PCR (replicação)
⇒ RAPD, AFLP, SSR ...
- Hibridização (complementaridade)
⇒ RFLP
- Sequenciamento (Estrutura DNA)
⇒ SNPs

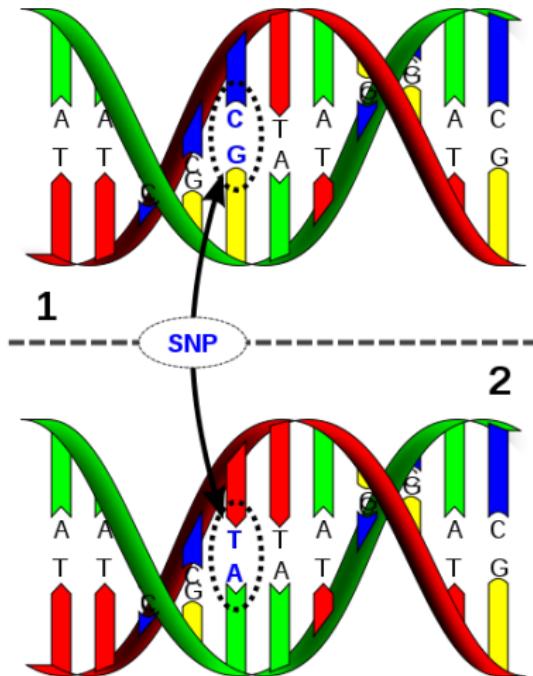


Molecular Markers

SNP - Single Nucleotide Polymorphisms

- Vantagens
 - Abundância
 - Baixo custo
 - Automatização

Fonte: Tanigutti, 2019



Model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

y_i : Dependent Variable

β_0 : Intercept

β_1 : Slope Coefficient

x_i : Explanatory Variable (e.g. SNP coded as 0, 1 e 2)

ϵ_i : Residual

Multiple Linear Regression

Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_j x_{ji} + \epsilon_i$$

$$y_i = \beta_0 + \sum_{j=1}^p x_{ji} \beta_j + \epsilon_i$$

y_i : Dependent Variable

β_0 : Intercept

β_1 : Slope Coefficient

x_i : Explanatory Variables (e.g. SNPs coded as 0, 1 e 2)

ϵ_i : Residual

p >> n

Fitting a model with many markers covariates (not possible with least squares)

- Mean squared error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2$$

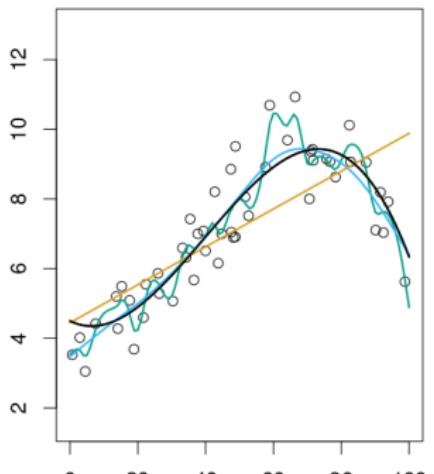
$$E(MSE) = Var(\hat{f}_{(x)}) + [Bias(\hat{f}_{(x)})]^2 + Var(e))$$

- ▶ Cause of dimensionality
- ▶ Bias-Variance Trade-off

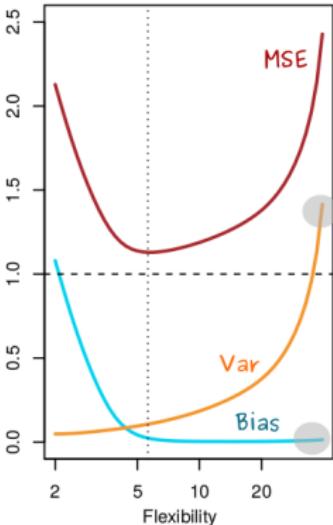
Bias-Variance Trade-off

* Modelos:

- i) Dados reais
- ii) Regressão Linear
- iii) Regressão Splines
- iv) Regressão Splines



Bias-Variance Trade-off Test set



* Fácil ter um modelo com baixo viés e alta flexibilidade, porém não são bons modelos preditivos.

* Desafio é encontrar um bom balanço entre Bias x Var, que reduza MSE.

Application

Breeder Equation

Breeder Equation:

$$R = \frac{i \times r_a \times \sigma_a}{L}$$

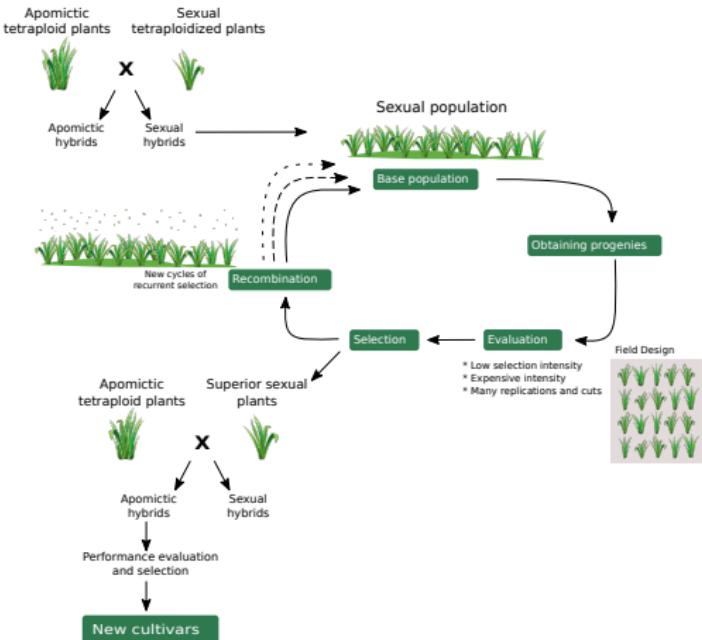
R = Response to selection

i = Selection intensity

r_a = Selection accuracy

σ_a = Square root of genetic variance

L = Generation interval



Breeder Equation

Breeder Equation:

$$R = \frac{i \times r_a \times \sigma_a}{L}$$

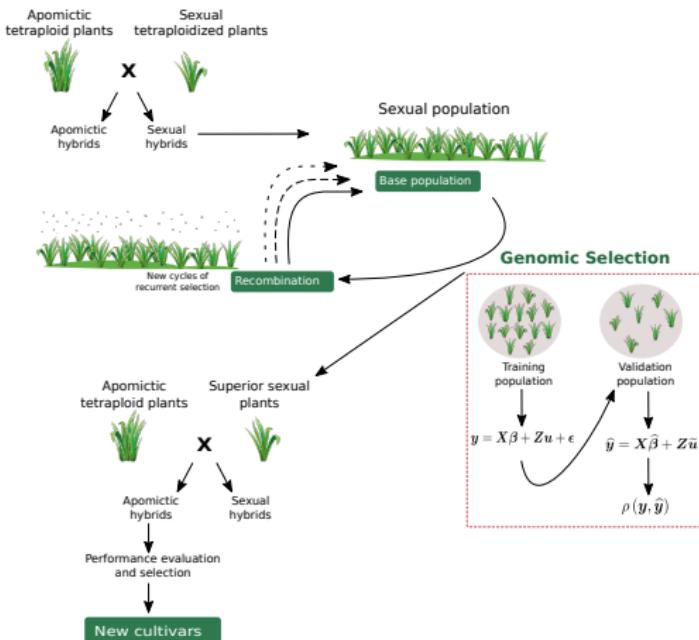
R = Response to selection

i = Selection intensity

r_a = Selection accuracy

σ_a = Square root of genetic variance

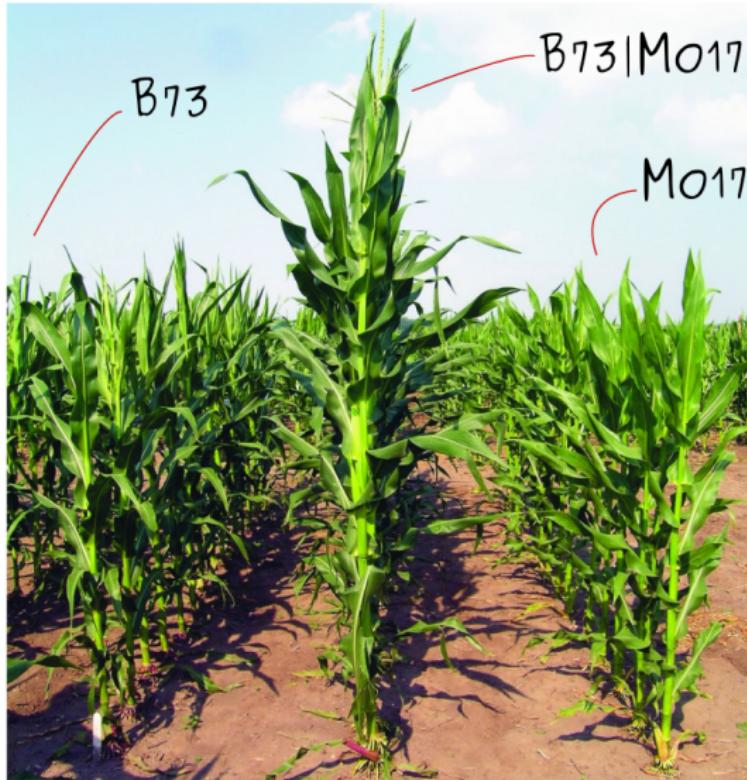
L = Generation interval



Source: Dr. Leticia Lara

Breeding Challenge

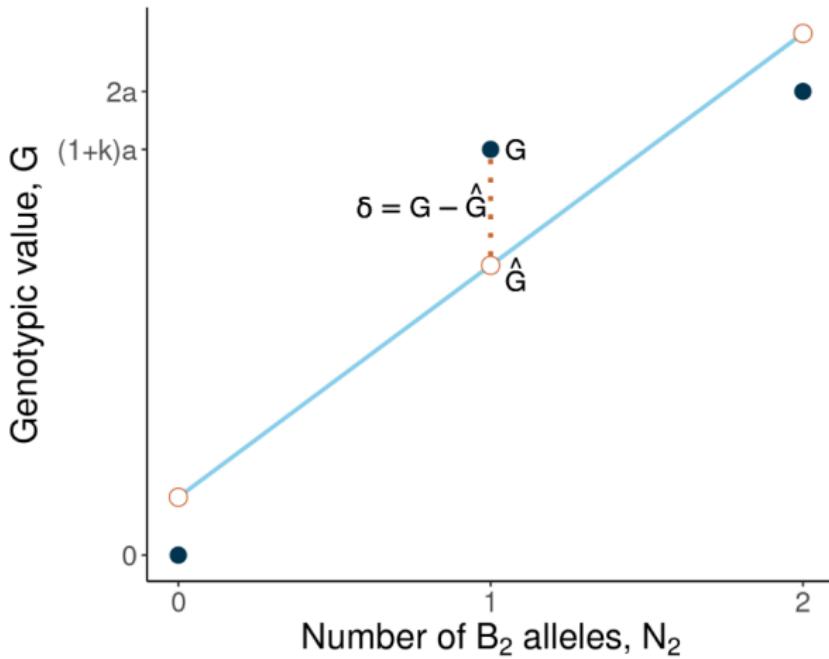
Heterosis



Source: Adapted from nativetraits.blogspot.com.br

Fisher's Infinitesimal Model (1918)

Cockerham's Model (1954)



Heredity (2018) 121:24–37
<https://doi.org/10.1038/s41437-018-0053-6>

ARTICLE

the
genetics society



Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials

GxE- FA

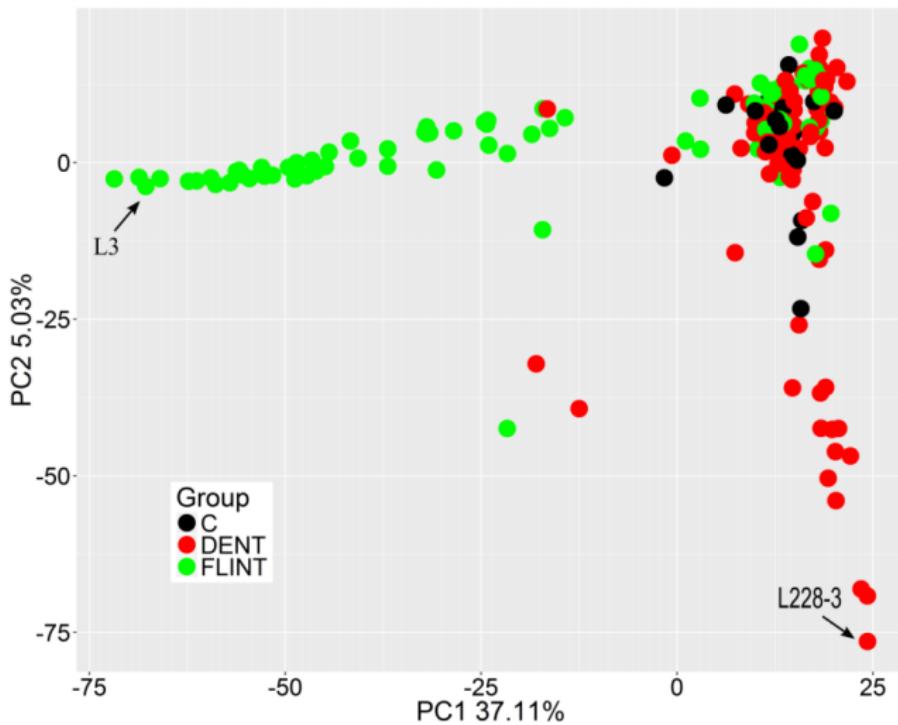
- To evaluate the accuracy of genomic selection of additive (A) against additive+dominance (AD) models to predict the performance of untested maize single-cross hybrids for drought tolerance in MET
- To explore the stability of hybrids via latent regression plots using AD models

Our focus were:

- ① Predict hybrids that have not been evaluated in any trial/environment
- ② Predict hybrids that are evaluated in some trials but not measured in others

Genomic Selection

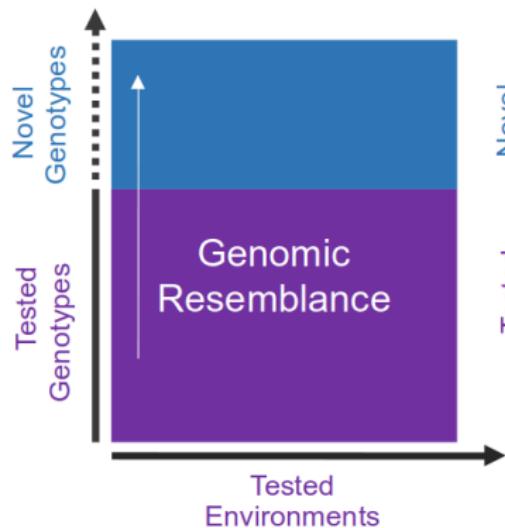
Population Structure



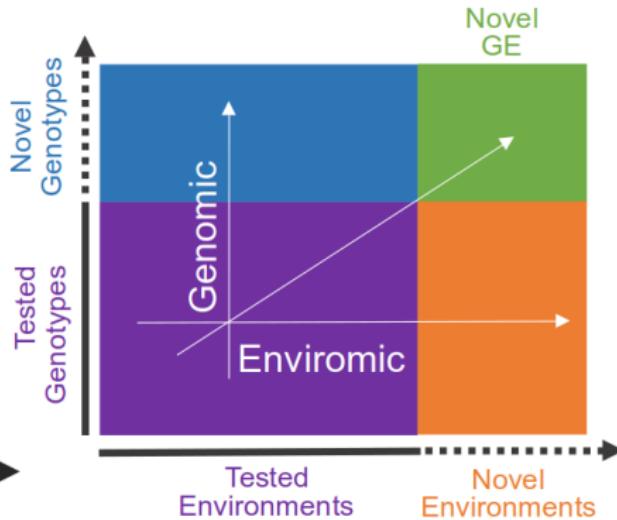
Genomic Selection

Cross-Validation

MET Prediction (Pure Genomic)



MET Prediction (Enviromic + Genomic)

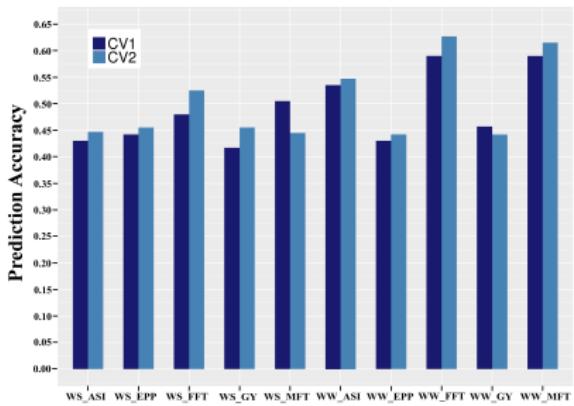
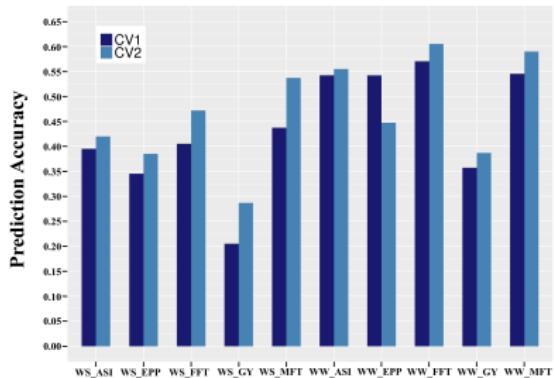


Adapted from Malossetti et al., 2016 (Slide: Germano Costa)

Prediction of single-cross hybrids

FA model

- $G_{SA} = \Delta_A \Delta_A^T + \Psi_A$; $G_{SD} = \Delta_D \Delta_D^T + \Psi_D$
- $\text{cov}(\mu_g) = A_g \otimes G_{SA} + D_g \otimes G_{SD}$
- $\text{cov}(\mu_g) = A_g \otimes (\Delta_A \Delta_A^T + \Psi_A) + D_g \otimes (\Delta_D \Delta_D^T + \Psi_D)$





[Fernandes, SB; Dias, KOG; Ferreira, DF; Brown, PF. 2018]

Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theoretical and Applied Genetics volume 131, pages 747–755

Plant Material

- 435 inbred lines of biomass sorghum
- 3 locations
- Lines were genotyped using GBS
- Six traits were phenotyped
 - ▶ 30 (H1), 60 (H2), 90 (H3), 120 (H4), M, and Y



Photo: Embrapa Maize and Sorghum

Multi-trait Prediction

Fernandes et al. 2018

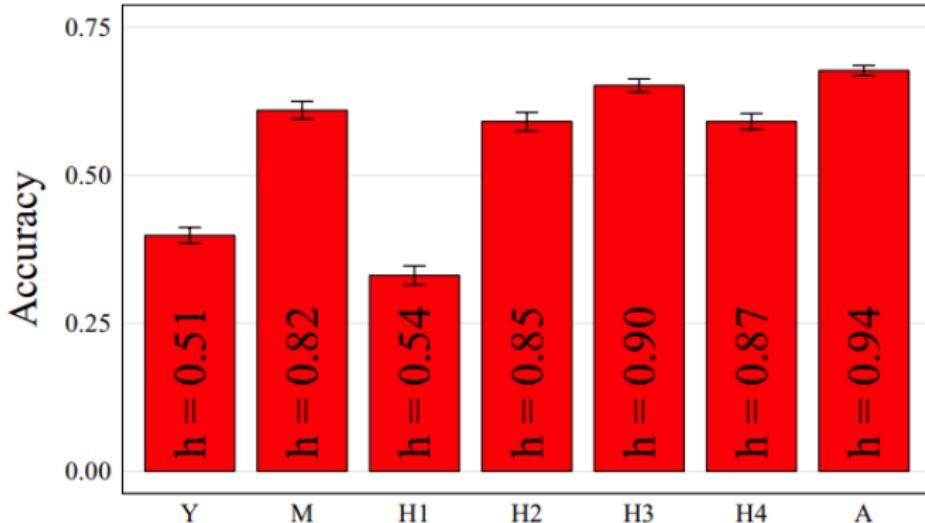


Fig. 1 Prediction accuracy of standard GS for biomass (Y), moisture (M), height at 30 (H1), 60 (H2), 90 (H3), 120 (H4) DAP and the area under growth progress curve (A). Standard deviations across 30 cross-validation runs are shown. The square root of the heritability (h) is shown inside each bar

Multi-trait Prediction

Fernandes et al. 2018

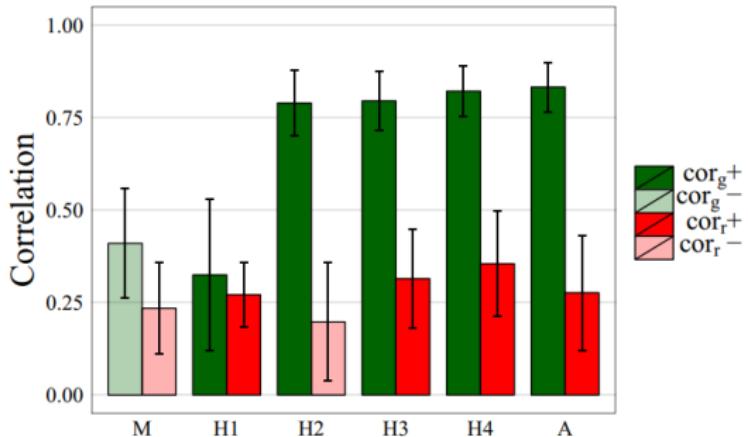
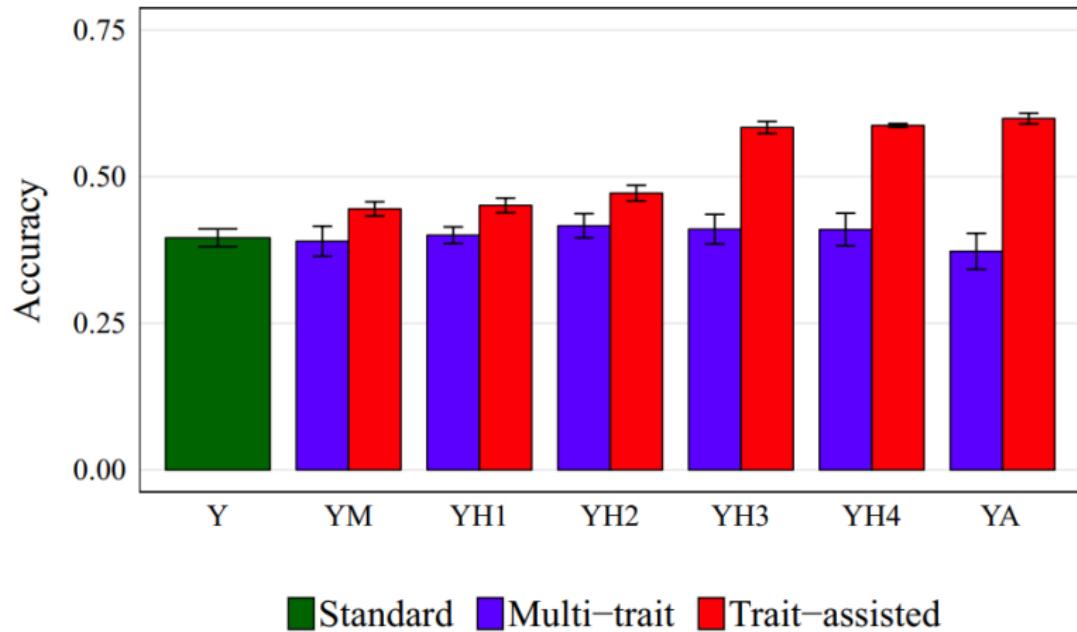


Fig. 2 Genetic (cor_g) and residual (cor_r) correlations between biomass and moisture (M), height at 30 (H1), 60 (H2), 90 (H3) and 120 (H4) DAP. Positive (+) and negative (-) correlations are indicated by shading, and standard errors of correlations are shown

Multi-trait Prediction

Fernandes et al. 2018

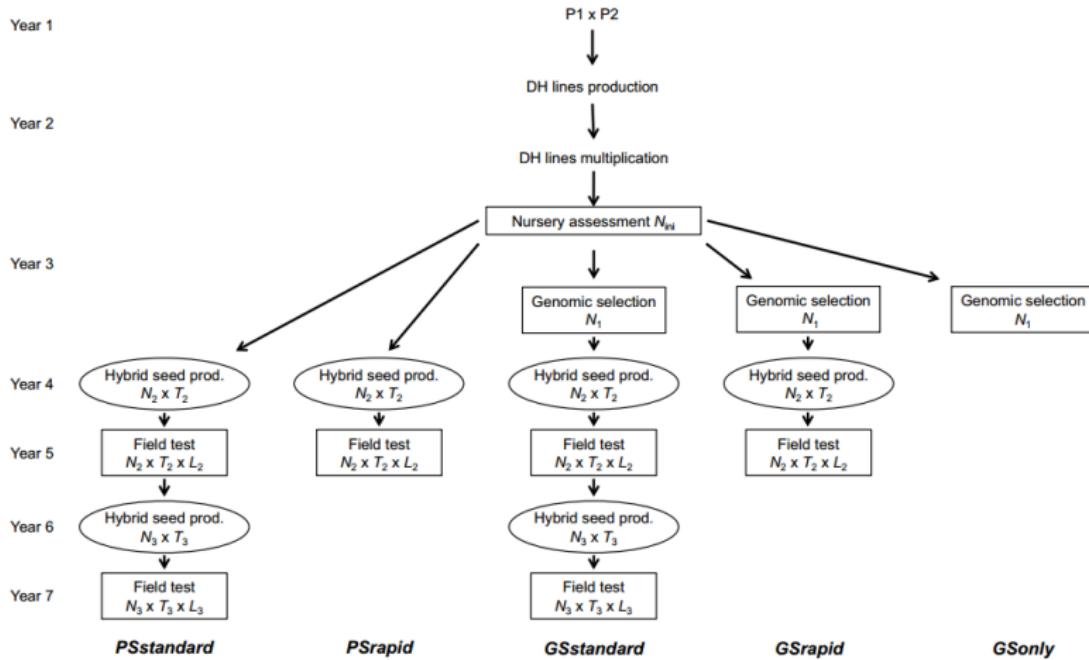


Genomic Selection for Annual Crops

Murulanda *et al.*, 2016

Theor Appl Genet (2016) 129:1901–1913

1903

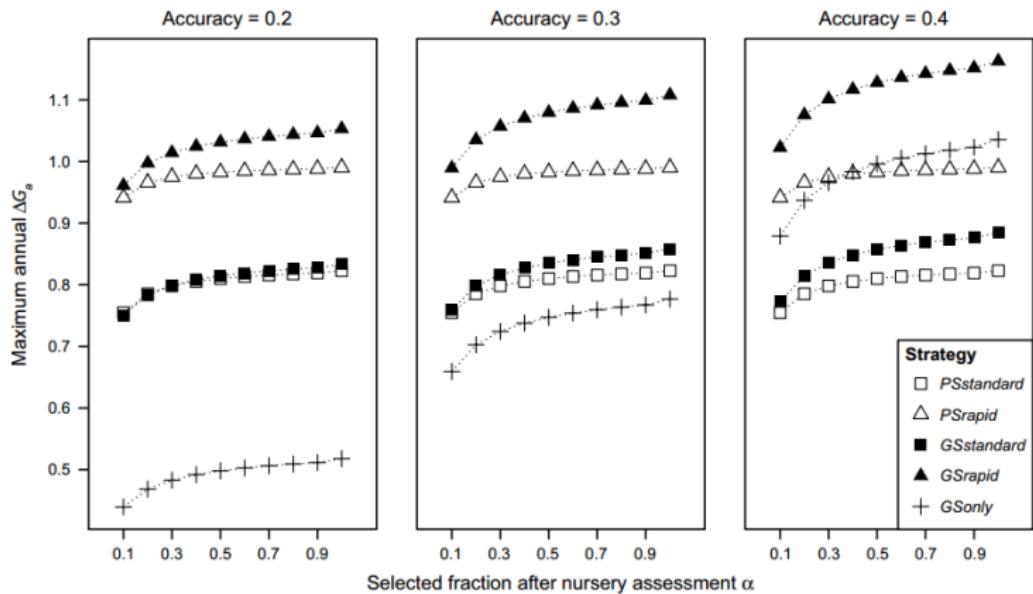


Genomic Selection for Annual Crops

UFV

Theor Appl Genet (2016) 129:1901–1913

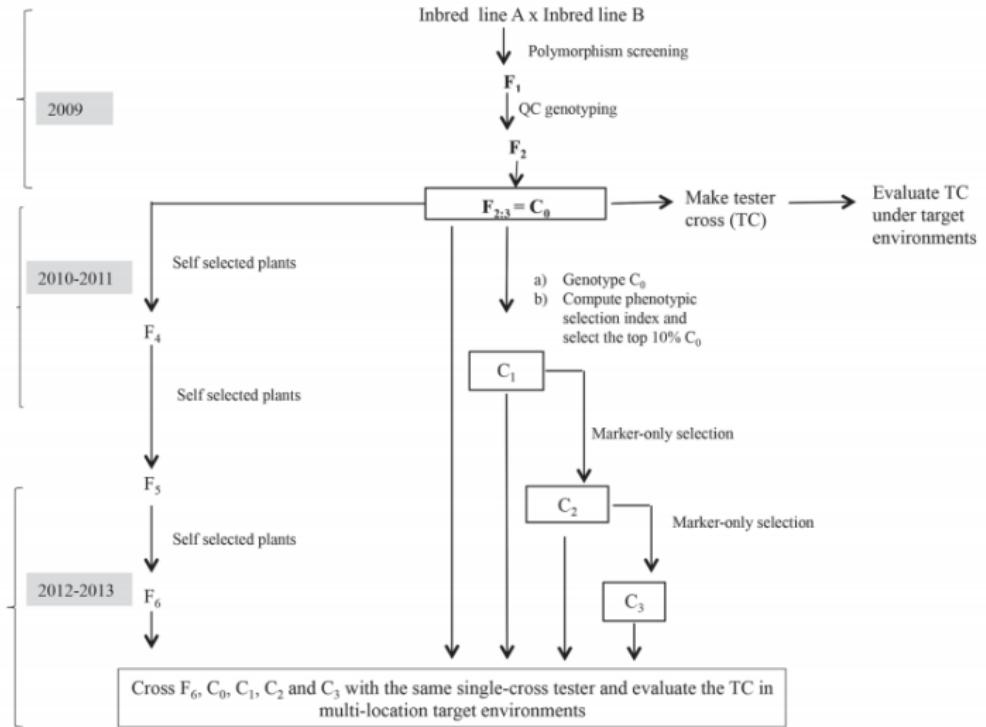
1905



Genomic Selection for Annual Crops

UFV

Beyene et al., 2015



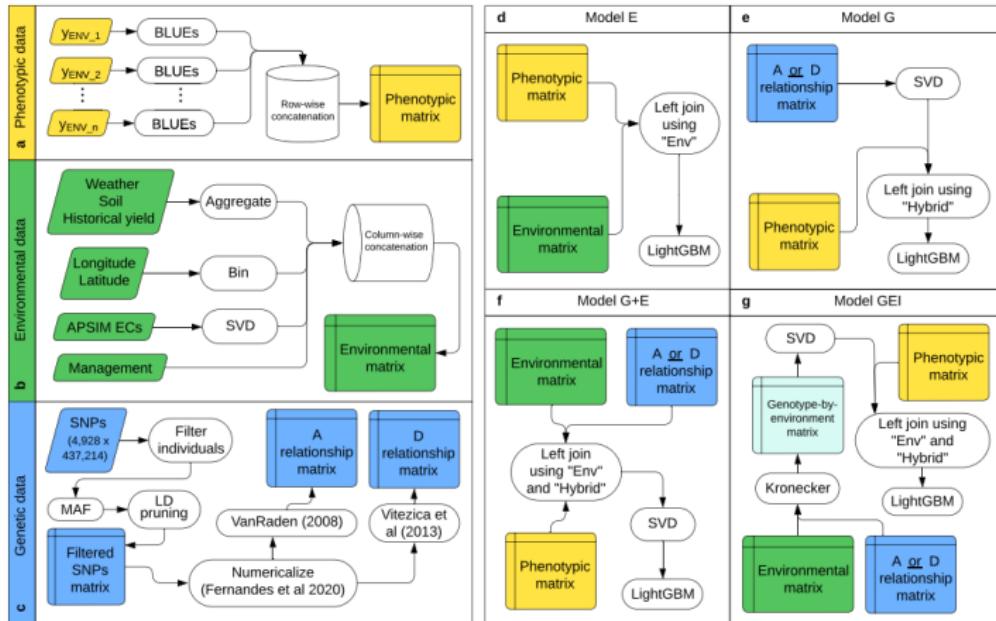
Genomic Selection

UFV

Using machine learning to combine genetic and environmental

189 Page 4 of 13

Theoretical and Applied Genetics (2024) 137:189



References



[dos Santos et al. 2020]

Novel Bayesian networks for genomic prediction of developmental traits in biomass sorghum. *G3 Genes|Genomes|Genetics*, Volume 10, Issue 2, 1 February 2020, Pages 769–781



[Gezan et al. 2017]

An experimental validation of genomic selection in octoploid strawberry. *Horticulture Research*, Volume 4, 2017, 1607



[Jarquin et al. 2021]

Genomic prediction enhanced sparse testing for multi-environment trials. *G3 Genes|Genomes|Genetics*, Volume 10, Issue 8, 1 August 2020, Pages 2725–2739



[Dias et al. 2022]

Leveraging probability concepts for cultivar recommendation in multi-environment trials. *Theoretical and Applied Genetics* (2022) 135:1385–1399



[Burgueño et al. 2012]

Genomic prediction of breeding values when modeling genotype \times environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52:707–712