



Module 2 – Genomic Selection

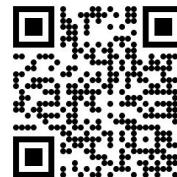
Genomic Selection: lessons from a blueberry breeding program

Felipe Ferrão

Research Scientist

✉ lferrao@ufl.edu

 lfelipeferrao



February, 2023

Motivation



Why genomic selection in blueberry?

Motivation

Why genomic selection in blueberry?

Breeding Design

- UF Breeding program started in the 1940's
- Heavily focused in massal selection, in a recurrent selection design
- Challenges:
 - Time-consuming (15 years -> cultivar)
 - Perennial species
 - Long juvenile time
 - Autotetraploid ($2n=4x$)
 - High level of inbreeding depression

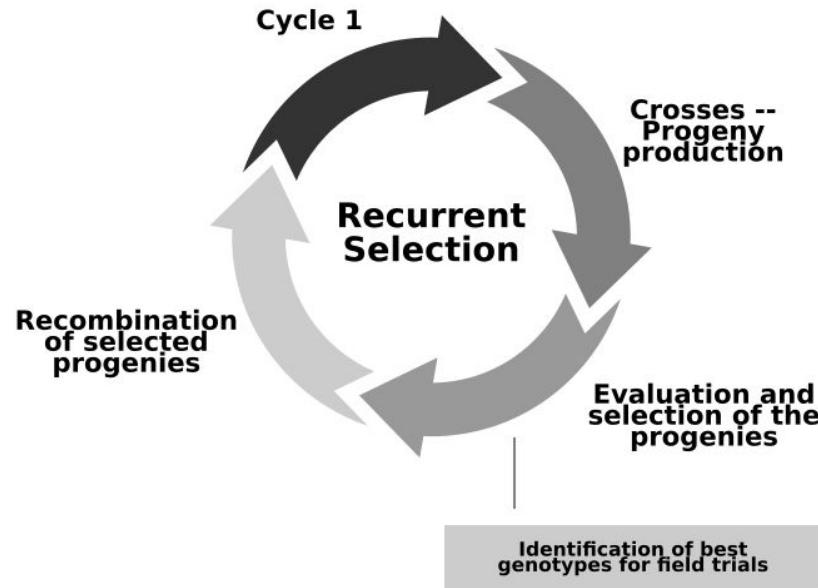


Motivation

Why genomic selection in blueberry?

Recurrent Selection

- Breeding design
- Repeated cycles of crosses & selection
- Goal: systematically increase the number of favorable alleles in the population
- Two main products:
 - Population improvement
 - Cultivar deployment



Motivation

Why genomic selection in blueberry?

Field year	Stage	# Plants	Goal
0	0	2.000.000	Crossing + Seedlings
1	I	20.000	High-density nursery – Single Plant Selection
2-4	II	2.000	Single Plant Selection
5-9	III	200	Farm Condition – Experimental design
10-15	IV	20	Regional Yield Trial – Experimental design
16	V	1-2	Cultivar Release

Motivation

Why genomic selection in blueberry?

Field year	Stage	# Plants	Goal	Genomic Selection
0	0	2.000.000	Crossing + Seedlings	Selecting crosses in early stages
1	I	20.000	High-density nursery – Single Plant Selection	
2-4	II	2.000	Single Plant Selection	
5-9	III	200	Farm Condition – Experimental design	
10-15	IV	20	Regional Yield Trial – Experimental design	Skip stages in a breeding cycle
16	V	1-2	Cultivar Release	

$$\Delta G = \frac{r_{IH} \times i \times \sigma_A}{L}$$

Motivation

Why genomic selection in blueberry?

Questions:

- Is genomic selection better than phenotypic selection?
- How to implement genomic selection in an autotetraploid species?
- What are the best predictive models?
- Can we optimize our phenotypic and genotypic resources? How low can we go?
- Can we unify prediction and discovery in a single framework?



Results



Blueberry Breeding – From theory to practice

Results

Blueberry Breeding – From theory to practice

Material and Methods

- **Population:** large population (~2000 ind) representing our breeding collection
- **Phenotype:** fruit quality traits (firmness, size, brix, acidity and weight)
- **Genotyping:**
 - CaptureSeq (target genotyping by sequencing approach)
 - High-quality reference genome for SNP calling
 - Information at the allele dosage level ($2n=4x$)

Question 1



Is genomic selection better than phenotypic selection?

Results

Is genomic selection better than phenotypic selection?



ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fpls.2021.676326



Genomic Selection in an Outcrossing Autotetraploid Fruit Crop: Lessons From Blueberry Breeding

Luis Felipe V. Ferrão¹, Rodrigo R. Amadeu¹, Juliana Benevenuto¹, Ivone de Bem Oliveira^{1,2} and Patrício R. Munoz^{1*}

¹Blueberry Breeding and Genomics Lab, Horticultural Sciences Department, University of Florida, Gainesville, FL, United States, ²Hortifrut North America, Inc., Estero, FL, United States

Blueberry (*Vaccinium corymbosum* and hybrids) is a specialty crop with expanding production and consumption worldwide. The blueberry breeding program at the University of Florida (UF) has greatly contributed to expanding production areas by developing low-chilling cultivars better adapted to subtropical and Mediterranean climates of the globe. The breeding program has historically focused on recurrent phenotypic selection. As an autopolyploid, outcrossing, perennial, long juvenile phase crop, blueberry breeding cycles are costly and time consuming, which results in low genetic gains per unit of time. Motivated by applying molecular markers for a more accurate selection in the early stages of breeding, we performed pioneering genomic selection studies and optimization for its implementation in the blueberry breeding program. We have also addressed some complexities of sequence-based genotyping and model parametrization for an autopolyploid crop, providing empirical contributions

OPEN ACCESS

Edited by:
Fernando H. Toledo,
International Maize and Wheat
Improvement Center, Mexico

Reviewed by:
Gata Moraes,
Virginia Tech, United States
Danilo Lyra,

 **University of Florida**
HOS6932 | Genomic Selection
Felipe Ferrao



GENOMIC PREDICTION

Genomic Prediction of Autotetraploids; Influence of Relationship Matrices, Allele Dosage, and Continuous Genotyping Calls in Phenotype Prediction

Ivone de Bem Oliveira,^{*†} Marcio F. R. Resende, Jr.,[‡] Luis Felipe V. Ferrão,^{*} Rodrigo R. Amadeu,^{*} Jeffrey B. Endelman,[§] Matias Kirst,^{**} Alexandre S. G. Coelho,[†] and Patrício R. Munoz^{*†}

^{*}Blueberry Breeding and Genomics Lab, and [†]Sweet Corn Genomics and Breeding, Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, [‡]Plant Genetics and Genomics Lab, Agronomy College, Federal University of Goiás, GO, Brazil, 74690-900, [§]Department of Horticulture, University of Wisconsin, Madison, WI 53706, and ^{**}Forest Genomics Lab, School of Forestry Resources and Conservation, University of Florida, Gainesville, FL 32610

ORCID IDs: 0000-0003-3723-9747 (I.d.B.O.); 0000-0002-2367-0766 (M.F.R.R.); 0000-0002-9655-4838 (L.F.V.F.); 0000-0001-5127-4448 (R.R.A.); 0000-0003-0957-4337 (J.B.E.); 0000-0001-8973-9351 (P.R.M.)

Results

Is genomic selection better than phenotypic selection?

BLUP

- ABLUP (pedigree)
- GBLUP (genomic)

Matrix Format: $Y = XB + e$

$$\begin{bmatrix} \text{Y} \\ n \end{bmatrix}_{(\# \text{ rows}) \quad (\# \text{ columns})} = \begin{bmatrix} \text{X} \\ n \end{bmatrix}_{n} \begin{bmatrix} \text{B} \\ g \end{bmatrix}_{g} \begin{bmatrix} \text{u} \\ g \end{bmatrix}_{g} + \begin{bmatrix} \text{e} \\ 1 \end{bmatrix}_{n \quad 1}$$

where,
 $\text{u} \sim \text{MVN}(0, G\sigma_u^2)$

$n = \# \text{ individuals}$
 $g = \# \text{ of breeding values}$

$$\begin{bmatrix} 9.87 \\ 14.48 \\ 9.91 \\ 14.64 \\ 9.55 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \text{GEBV1} \\ \text{GEBV2} \\ \text{GEBV3} \\ \text{GEBV4} \\ \text{GEBV5} \end{bmatrix} + \begin{bmatrix} \text{e} \end{bmatrix}$$

IBS

Genomic Relationship Matrix (\mathbf{G})

Mixed Model Equation

$$\begin{bmatrix} 1'1 \\ 1'X \\ Z'Z + \sigma_e^2 \sigma_u^2 G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ 1'y \end{bmatrix}$$

Genomic Estimated
Breeding Value
(GEBV)

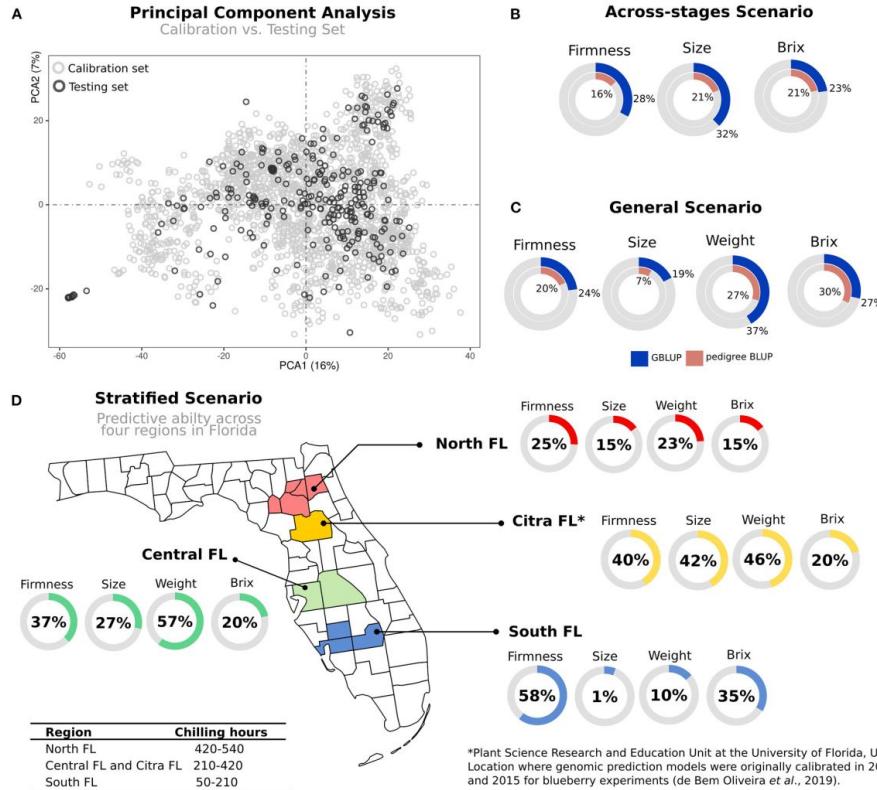
Results

Is genomic selection better than phenotypic selection?

Trait	Matrix	Accuracy	Method
Firmness	A	0.375	Pedigree
Firmness	G2	0.415	GBLUP (2n=2x)
Firmness	G4	0.426	GBLUP (2n=4x)
<hr/>			
Size	A	0.386	Pedigree
Size	G2	0.400	GBLUP (2n=2x)
Size	G4	0.432	GBLUP (2n=4x)

Results

Is genomic selection better than phenotypic selection?



Take home message

Genomic selection has shown consistently higher accuracies than pedigree

Question 2

How to implement genomic selection in an autotetraploid species?

Results

How to implement genomic selection in an autotetraploid species?

RESEARCH

Impact of Dominance Effects on Autotetraploid Genomic Prediction

Rodrigo R. Amadeu, Luis Felipe V. Ferrão, Ivone de Bem Oliveira, Juliana Benevenuto,
Jeffrey B. Endelman, and Patricio R. Munoz*

ABSTRACT

Many commercially important plants are autopolyploid. As a result of the multiple chromosome sets in their genomes, higher orders of allele interactions can occur, implying

R.R. Amadeu, L.F.V. Ferrão, I.D.B. Oliveira, J. Benevenuto, and P.R. Munoz, Blueberry Breeding and Genomics Lab, Horticultural Sciences Dep., Univ. of Florida, Gainesville, FL 32611; J.B. Endelman, Dep. of Horticulture, Univ. of Wisconsin, Madison, WI 53706. Received 28 Feb. 2019. Accepted 22 May 2019. *Corresponding author (p.munoz@ufl.edu). Assigned to Associate Editor Carlos Messina.



University of Florida

HOS6932 | Genomic Selection
Felipe Ferrao

Results

How to implement genomic selection in an autotetraploid species?

Background

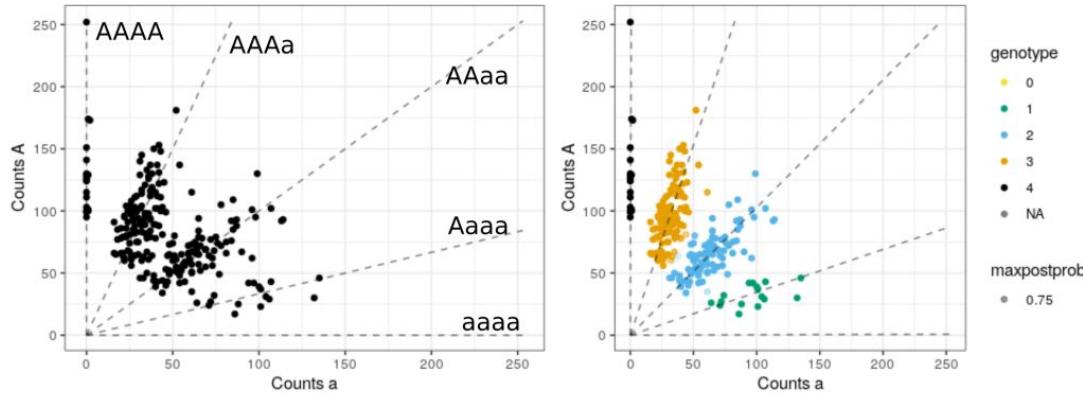
- Diploid: two sets of homologous chromosomes
- Genotyping: measure allelic variations within a locus
- Illumina sequencing to access such variation
- Not very complex in diploids



Results

How to implement genomic selection in an autotetraploid species?

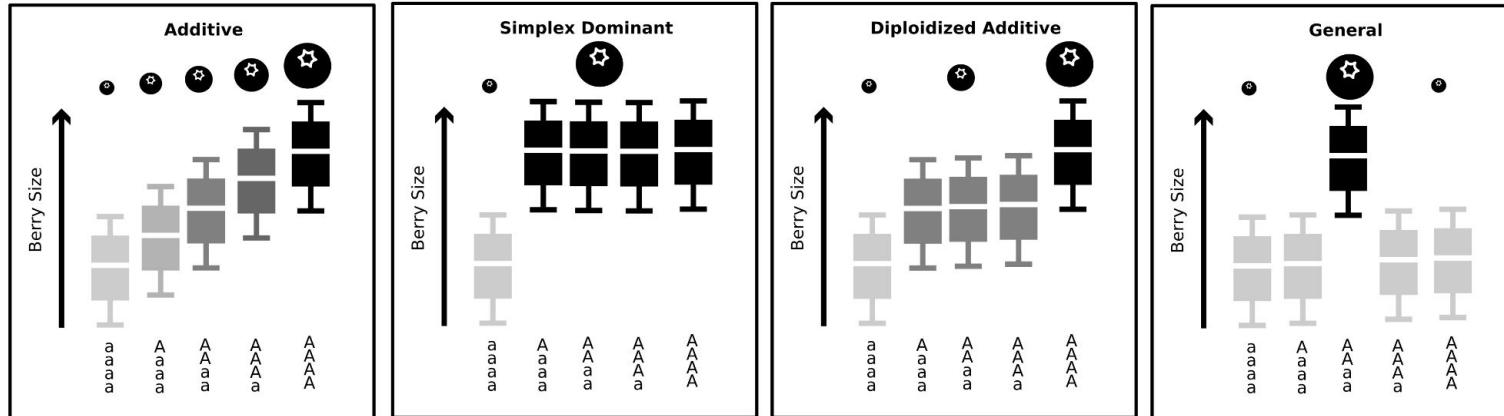
- Polyploid: allelic dosage range from zero up to the ploidy level
- Polyploid: allelic dosage range from zero up to the ploidy level
- $2n=4x$: aaaa(nulliplex) 7 → AAAA (tetraplex)



Results

How to implement genomic selection in an autotetraploid species?

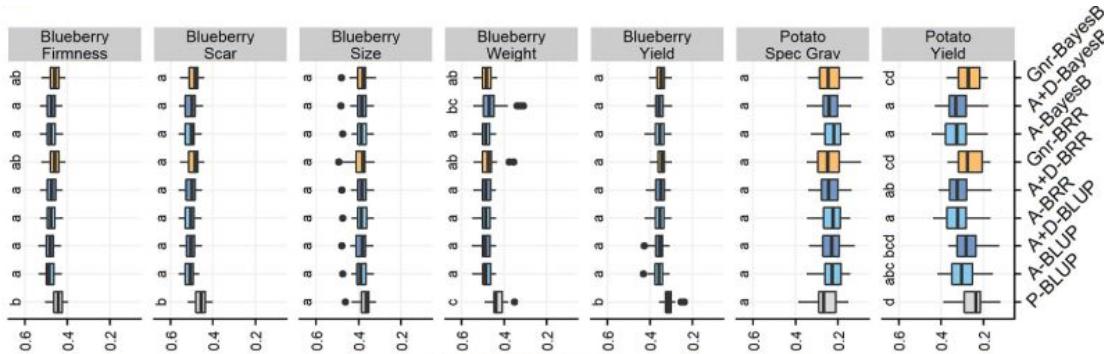
- Multiple gene actions
- Different dominance levels



Results

How to implement genomic selection in an autotetraploid species?

- Bayesian vs. Mixed Models
- Additive vs. Dominance vs. General Models
- Blueberry (5 traits) vs. Potato (2 traits)



Take home message

On the relevance of additive models

Question 3



What are the best predictive models?

Results

What are the best predictive models?



ORIGINAL RESEARCH
published: 06 February 2020
doi: 10.3389/fpls.2020.00025



Exploring Deep Learning for Complex Trait Genomic Prediction in Polyploid Outcrossing Species

Laura M. Zingaretti^{1*}, Salvador Alejandro Gezan², Luis Felipe V. Ferrão³, Luis F. Osorio⁴, Amparo Monfort^{1,5}, Patricio R. Muñoz³, Vance M. Whitaker⁴ and Miguel Pérez-Enciso^{1,6*}

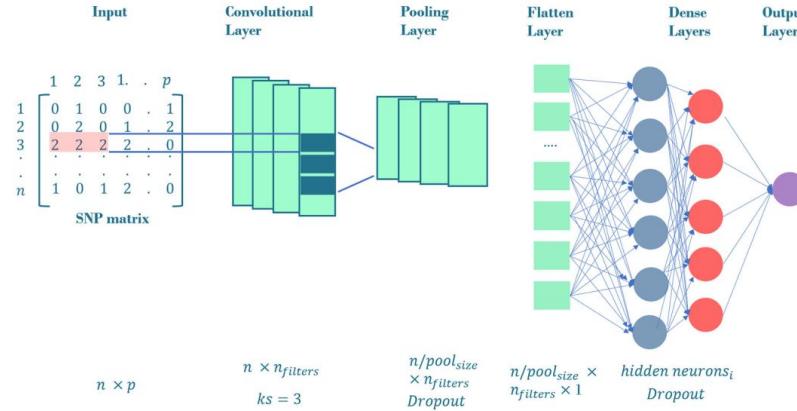
¹ Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Barcelona, Spain, ² School of Forest Resources and Conservation, University of Florida, Gainesville, FL, United States, ³ Blueberry Breeding and Genomics Lab, Horticultural Sciences Department, University of Florida, Gainesville, FL, United States, ⁴ IFAS Gulf Coast Research and Education Center, University of Florida, Wimauma, FL, United States, ⁵ Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Barcelona, Spain, ⁶ ICREA, Passeig de Lluís Companys 23, Barcelona, Spain

Results

What are the best models?

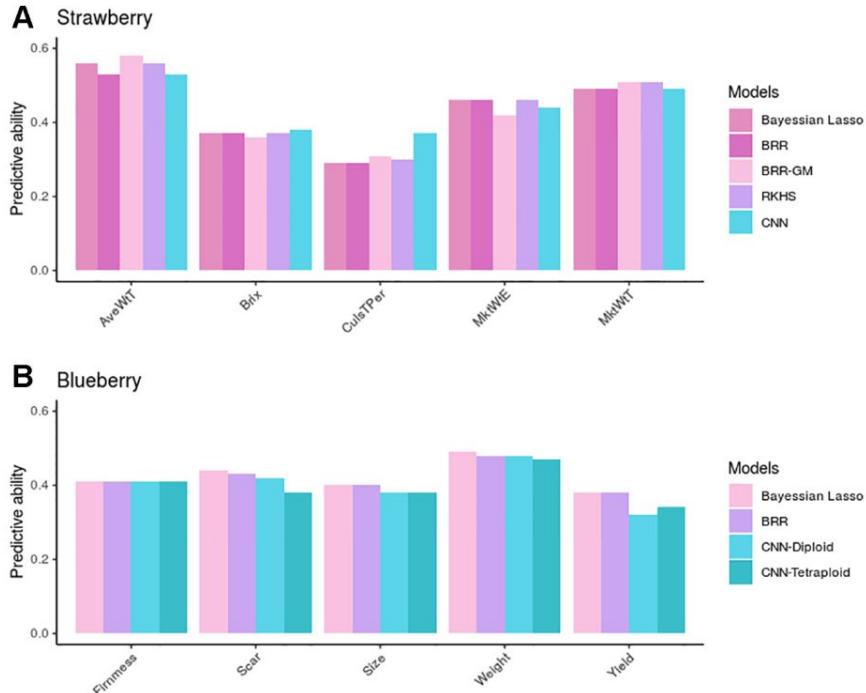
Background

- Polyplloid is complex !
- Most of methods are based on linear methods
- More elaborate models (Deep learning) can incorporate non-linearity



Results

What are the best models?



Take home message

On the relevance of additive models² !!!

Question 4



Can we optimize our phenotypic and genotypic resources? How low can we go?

Results

How low can we go?

Heredity
<https://doi.org/10.1038/s41437-020-00357-x>

the
geneticssociety

ARTICLE



Optimizing whole-genomic prediction for autotetraploid blueberry breeding

Ivone de Bem Oliveira ¹ · Rodrigo Rampazo Amadeu ¹ · Luis Felipe Ventorim Ferrão¹ · Patricio R. Muñoz¹

Received: 23 January 2020 / Revised: 11 August 2020 / Accepted: 11 August 2020

© The Author(s) 2020. This article is published with open access

Results

How low can we go?

Background

- Genomic Prediction relies on LD between markers and QTL
- In the literature, there are evidences showing that predictive ability does not increase linearly with large number of markers
- Can we reduce the number of markers ?
- Can we optimize our training set?

Results

How low can we go?

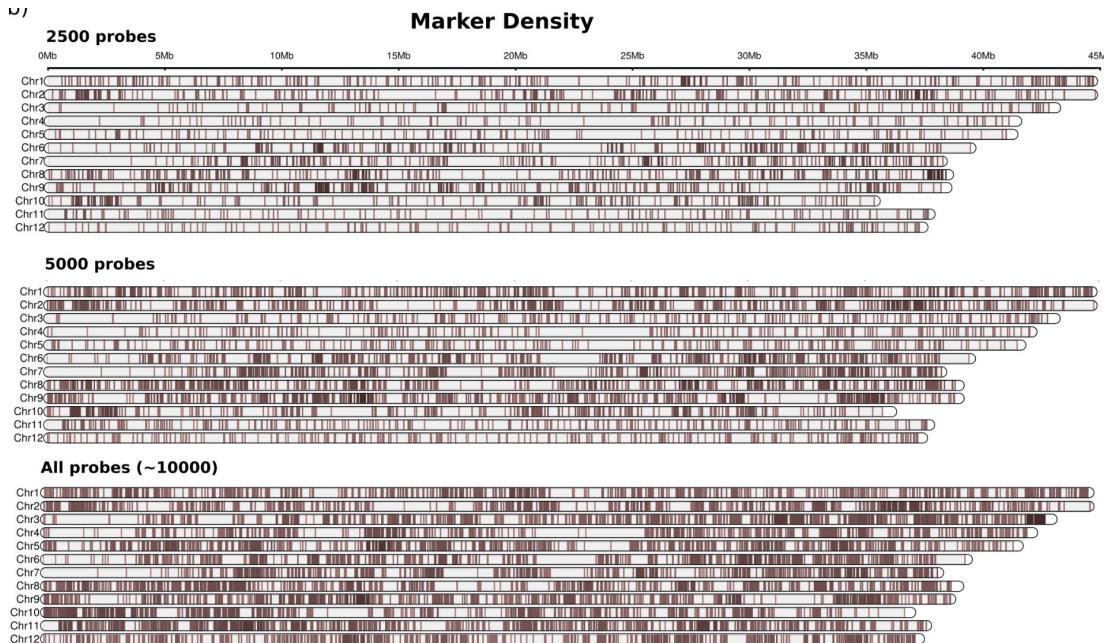
Prediction Ability

# probes	Brix	Firmness	TTA
All	0.31	0.50	0.42
2500	0.32	0.48	0.40
5000	0.32	0.49	0.41

- Probes are regions in the genome where SNPs will be mapped
- On average, each probe has ~6 SNPs
- A total of 10.000 probes was designed by blueberry
- 2500 probes (~25% of the total costs) results in good predictive ability for multiple traits

Results

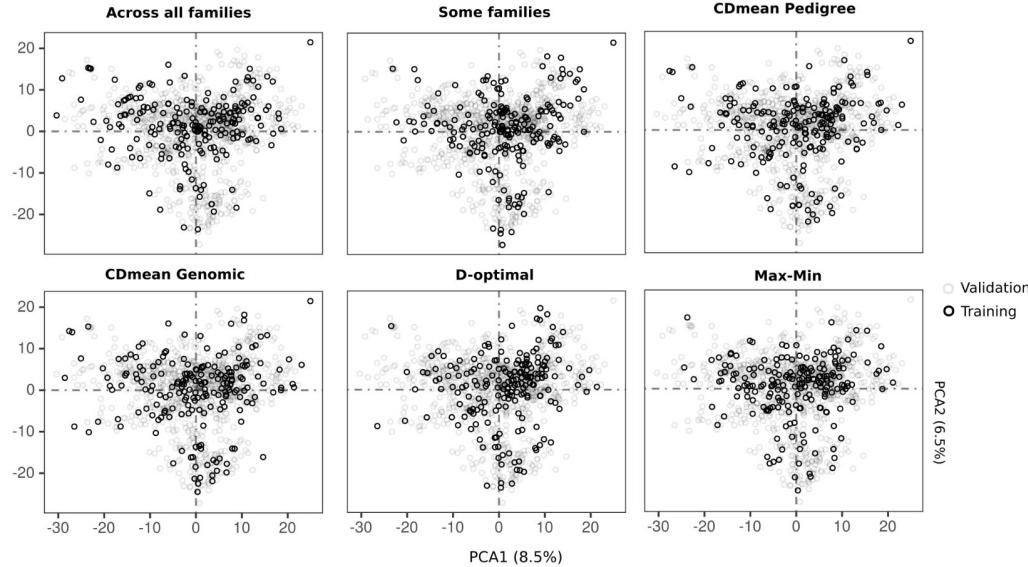
How low can we go?



Results

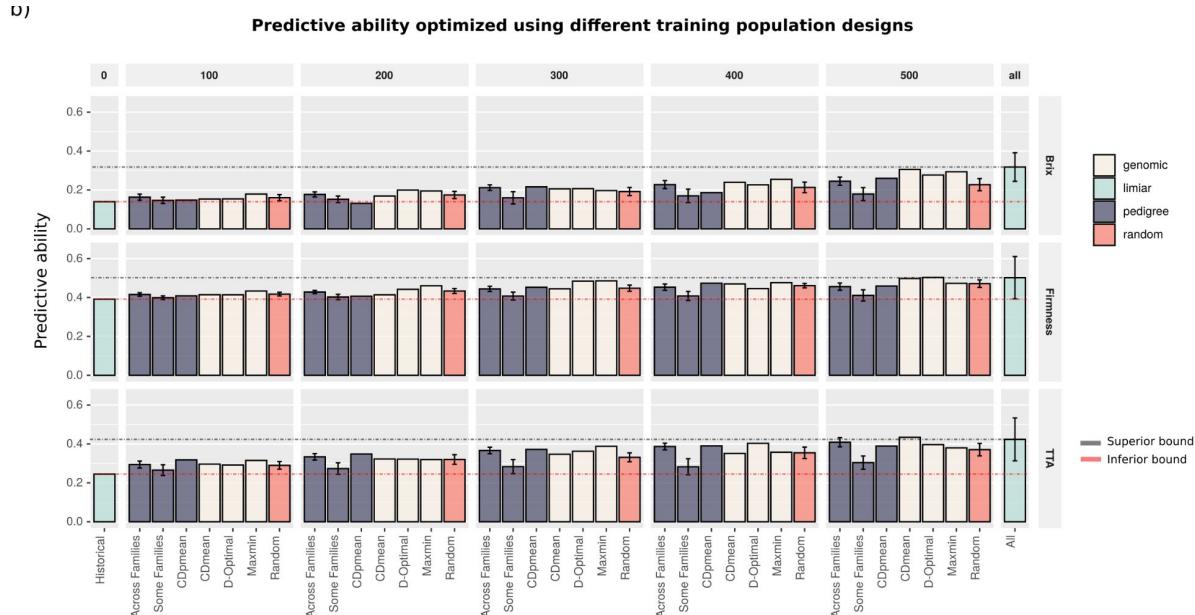
How low can we go?

What is the impact on selecting a TRN population?



Results

How low can we go?



Take home message

We can reduce number of markers and optimize our TRN population to maximize the accuracy

Question 5



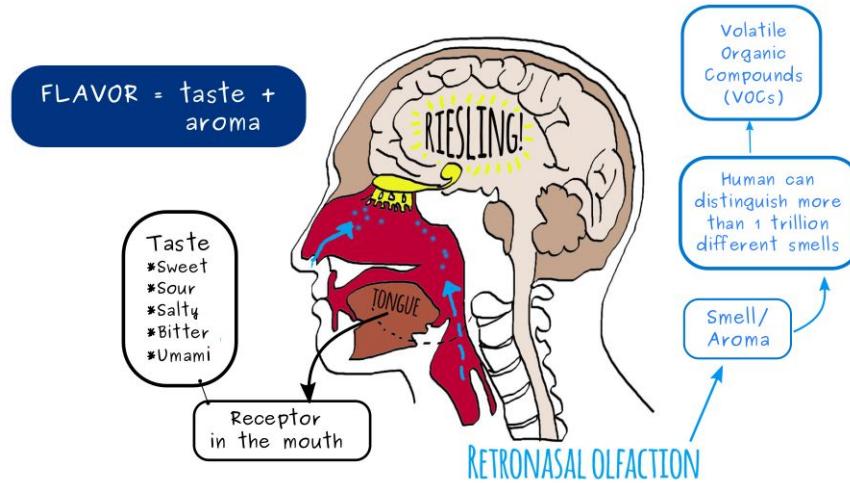
Can we unify prediction and discovery in a single framework?

Results

Can we unify prediction and discovery in a single framework?

What is Flavor?

- Flavor is the sum of inputs from multiple senses that inform our brain what we are eating



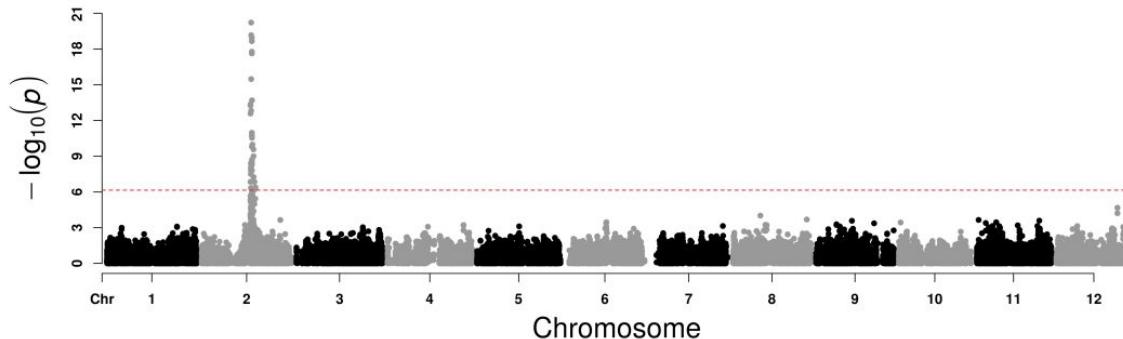
Results

Can we unify prediction and discovery in a single framework?

Evidences based on the GWAS analysis:

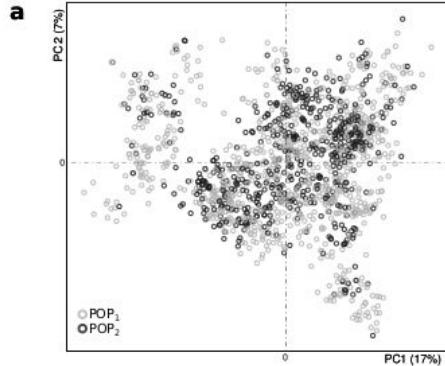
- Significant SNPs converging to a tower-like structure
- Single markers explaining large portions of the phenotypic variation
- Hypothesis: VOCs are traits with simple genetic architecture

2-undecanone



Results

Can we unify prediction and discovery in a single framework?

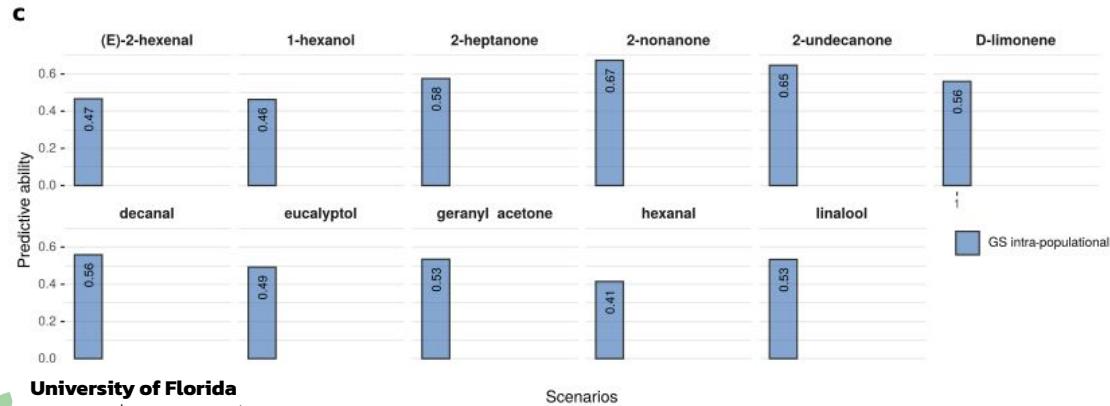
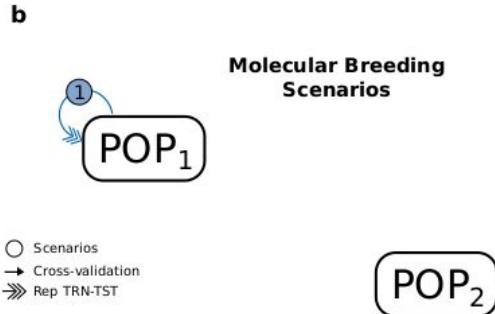
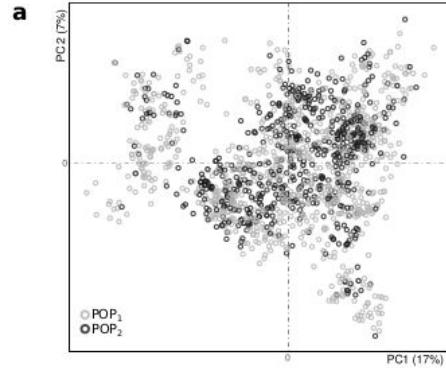


- *Original Population
- *GWAS
- *90 full-sib families

- *New Population
- *Validation
- *Genetic related to POP1

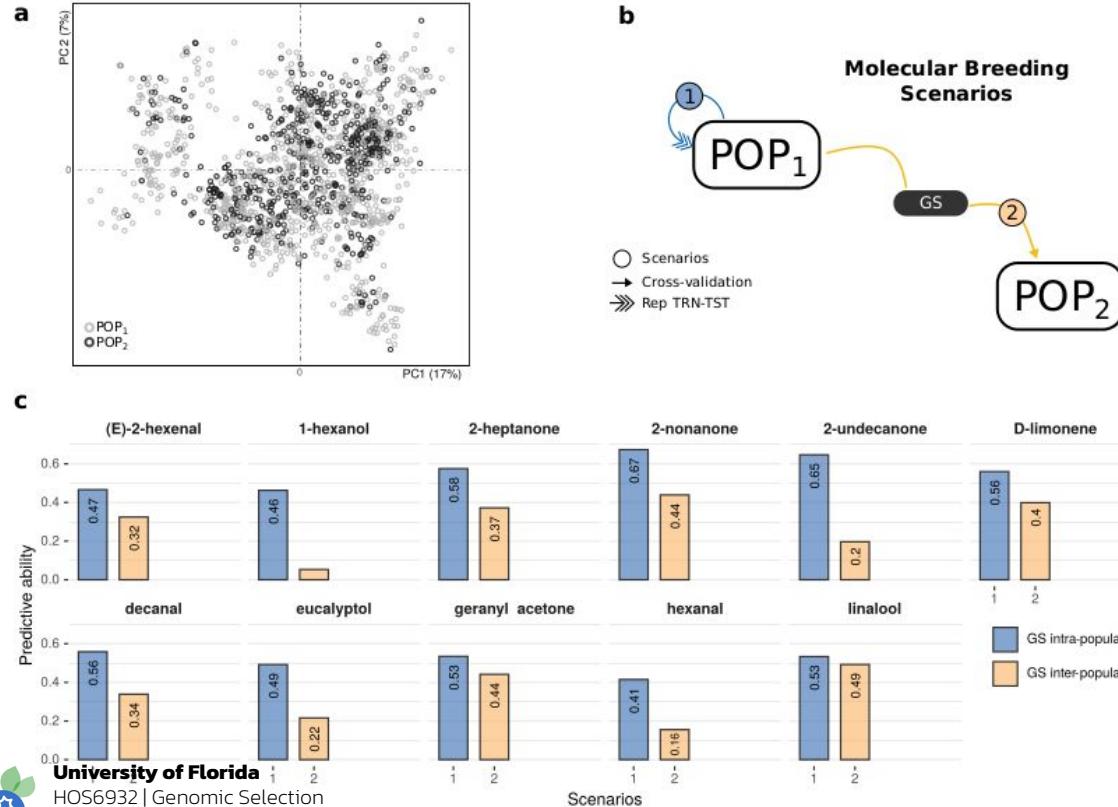
Results

Can we unify prediction and discovery in a single framework?



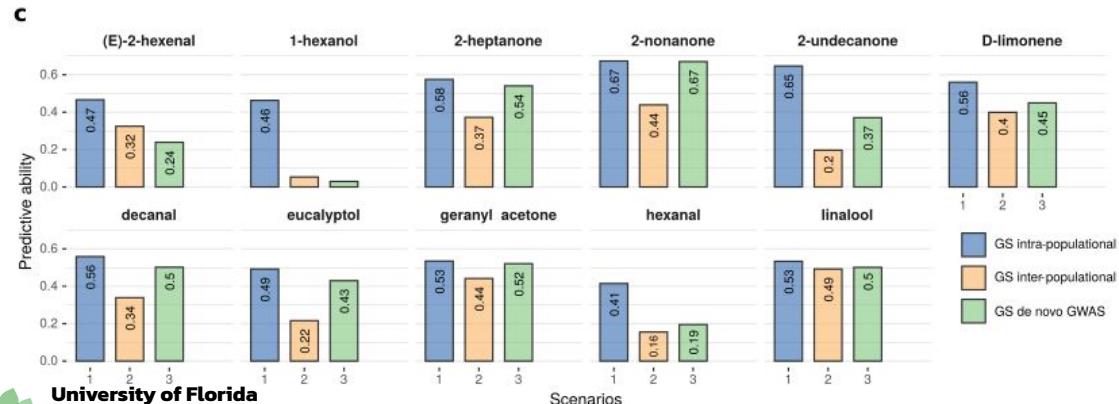
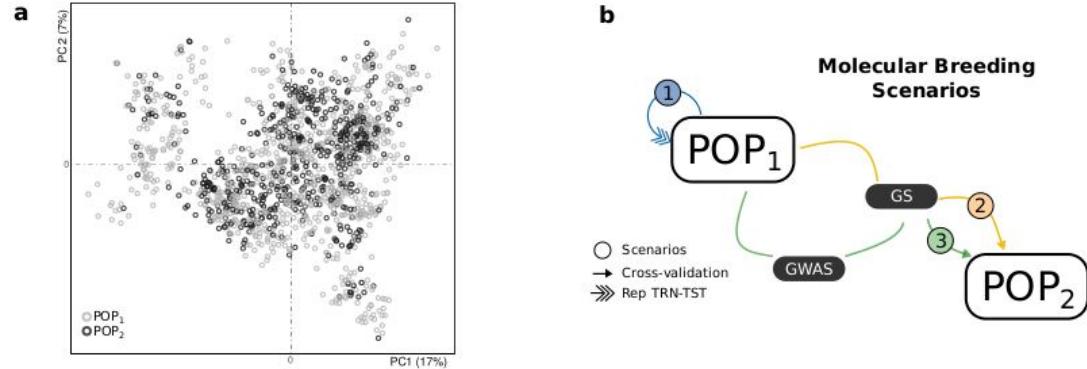
Results

Can we unify prediction and discovery in a single framework?



Results

Can we unify prediction and discovery in a single framework?



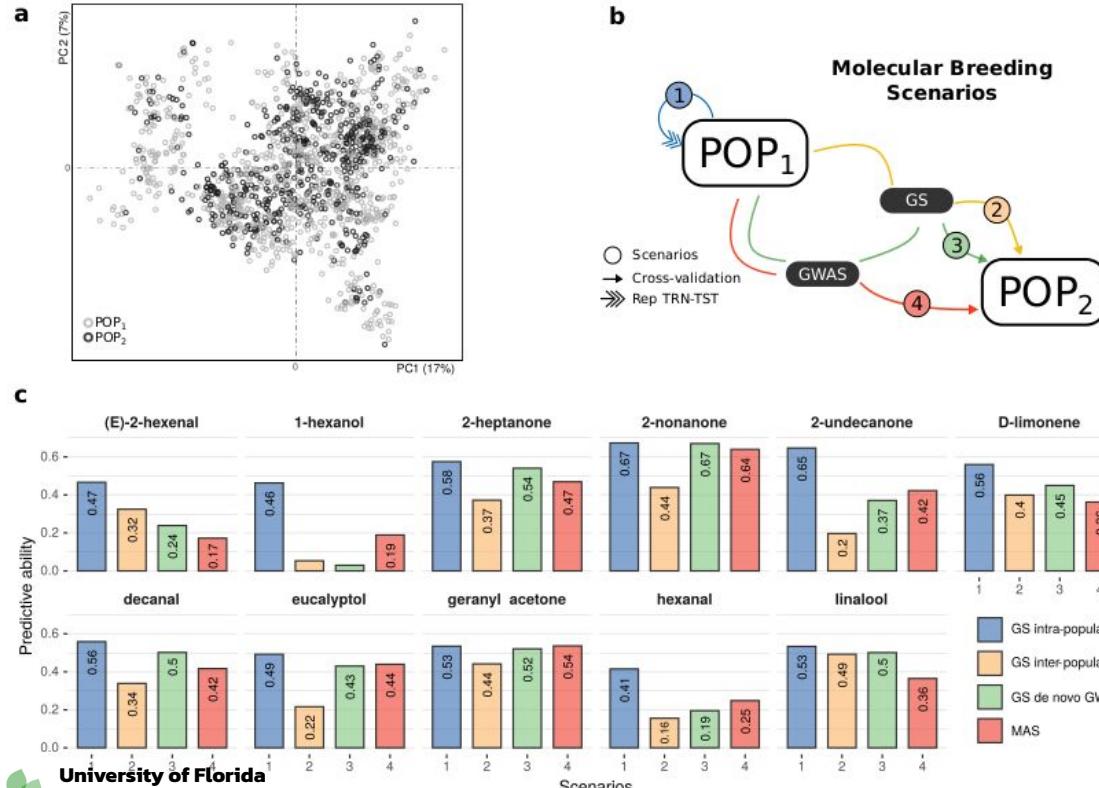
University of Florida

HOS6932 | Genomic Selection

Felipe Ferrao

Results

Can we unify prediction and discovery in a single framework?



Take home message

Prior biological information can improve prediction ability

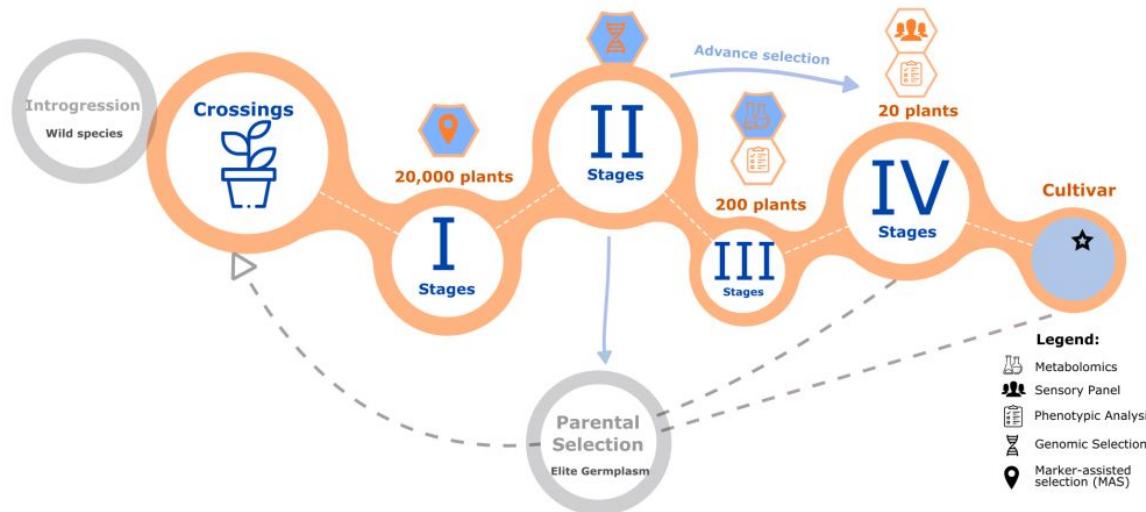
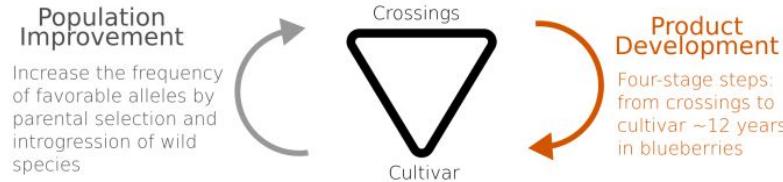
Conclusions



Final considerations !

Conclusions

Final considerations



Conclusions

Final considerations

Five main lessons from the last 5 years

- Solid gains using data driven methods
- Good reference genome (and molecular tools) is very important
- Prefer statistical genetics methods designed for polyploid systems
- "Simplicity is the Ultimate Sophistication"¹: on the Relevance of Additive GBLUP Models
- GS can be optimized, after having a very good understanding about your breeding pipeline

1 Quote by Leonardo da Vinci

Conclusions

Final considerations

My personal opinion for practical implementation:

- Genomic selection is a tool to assist breeding and not the other way around
- Implementation require a solid breeding program
- Genotyping might be not so easy and cheap
- Collect good phenotype is imperative
- Test new methods is valid. But don't forget to include an additive GBLUP as a benchmark
- Genomic Selection is multidisciplinary topic
- Biological discovery and prediction can run side-by-side

Thank you !!!

Question ??

About Me:

SCAN ME

