# Genomic selection

also known as

# Genomewide selection

# Number of loci



**Purpose of QTL mapping**

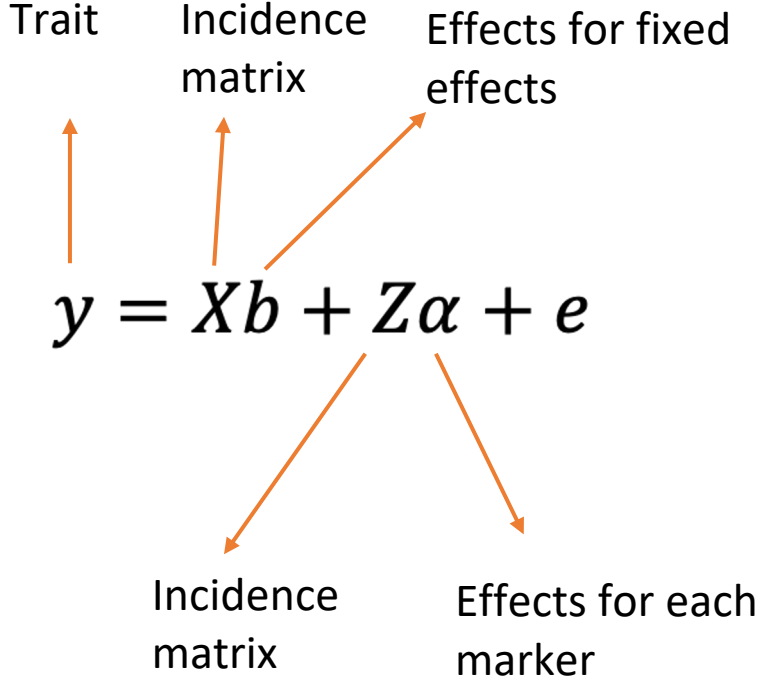Gene discovery

Selection

Few → Many

Clone QTL

Comparative mapping

Association mapping

Candidate genes

Map major QTL

F$_2$ enrichment

MARS

Genomewide selection

**Germplasm**

Unique, diverse

Elite

Introgress major QTL | Combine QTL alleles | Predict performance

**Approach in marker-based selection**

Bernardo 2008

# Genomic selection

Training Population
Calibration Set

DNA marker data

Phenotypic data

$$y = Xb + Z\alpha + e$$

- No QTL mapping
- No testing for significant markers

Selection candidates

Predict and select

Lorenz 2016

Trait    Incidence matrix    Effects for fixed effects

$$y = Xb + Z\alpha + e$$

Incidence matrix    Effects for each marker

# Genomic selection

GS uses all molecular markers to predict genomic estimated breeding values **(GEBV).**

| Genotype | Breeding value |
|----------|----------------|
| A1A1 | $2\alpha_1 = 2q\alpha$ |
| A1A2 | $\alpha_1 + \alpha_2 = (q - p)\,\alpha$ |
| A2A2 | $2\alpha_2 = -2p\alpha$ |

$$GEBV = \sum_{i}^{p} Z_i\,\alpha_i$$
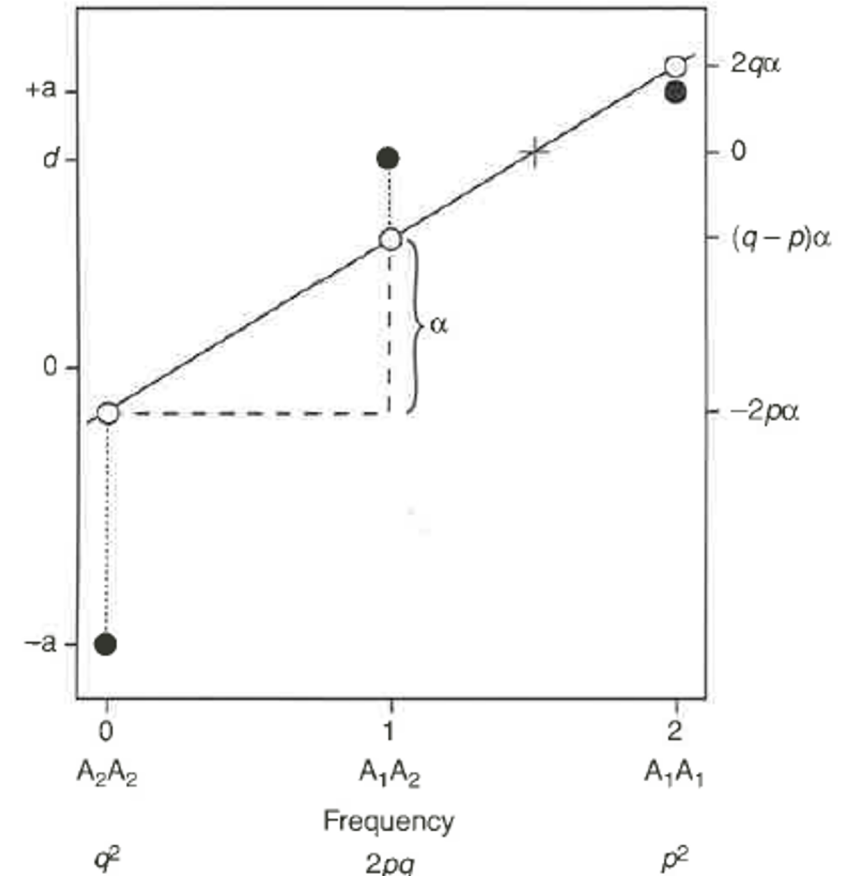
# Genomic selection
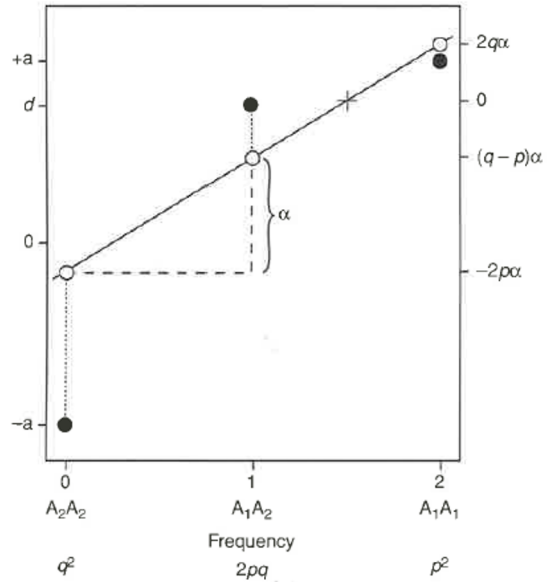


Falconer & Mackay p 115,117



FIGURE 3.2. Average effects of alleles.

Bernardo 2020

*Average effects of the all*

$$\alpha_1 = q[a + d(q - p)] \qquad \alpha_2 = -p[a + d(q - p)]$$

*Average effects of gene supstit*

$$\alpha = \alpha_1 - \alpha_2 = a + d(q - p)$$
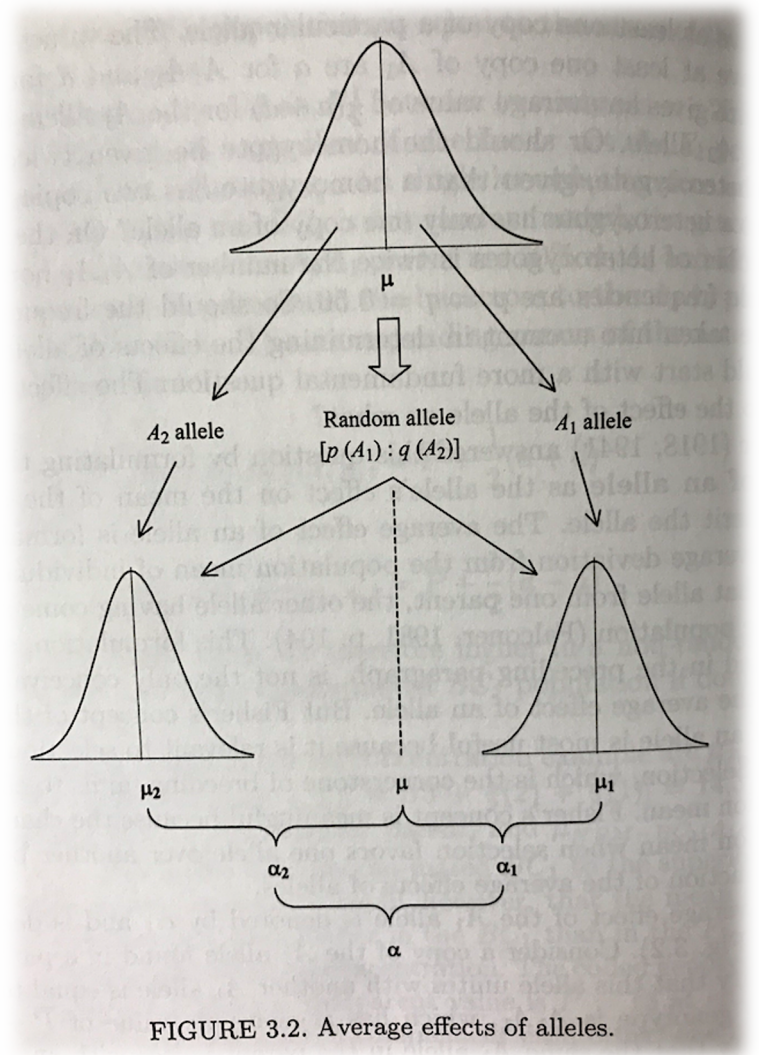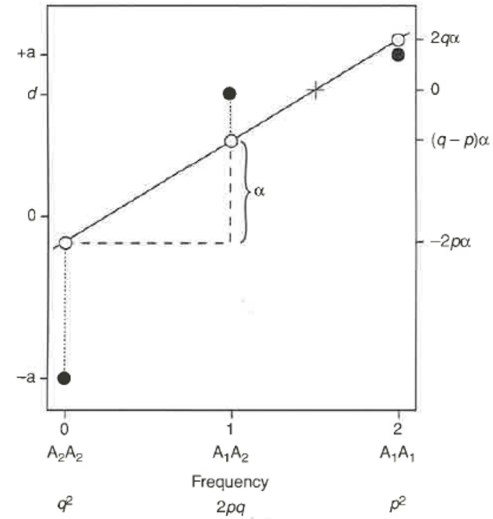
$$GEBV = \sum_i^p Z_i \, \alpha_i$$

# Genomic selection



$$GEBV = \sum_{i}^{p} z_i \, \alpha_i$$
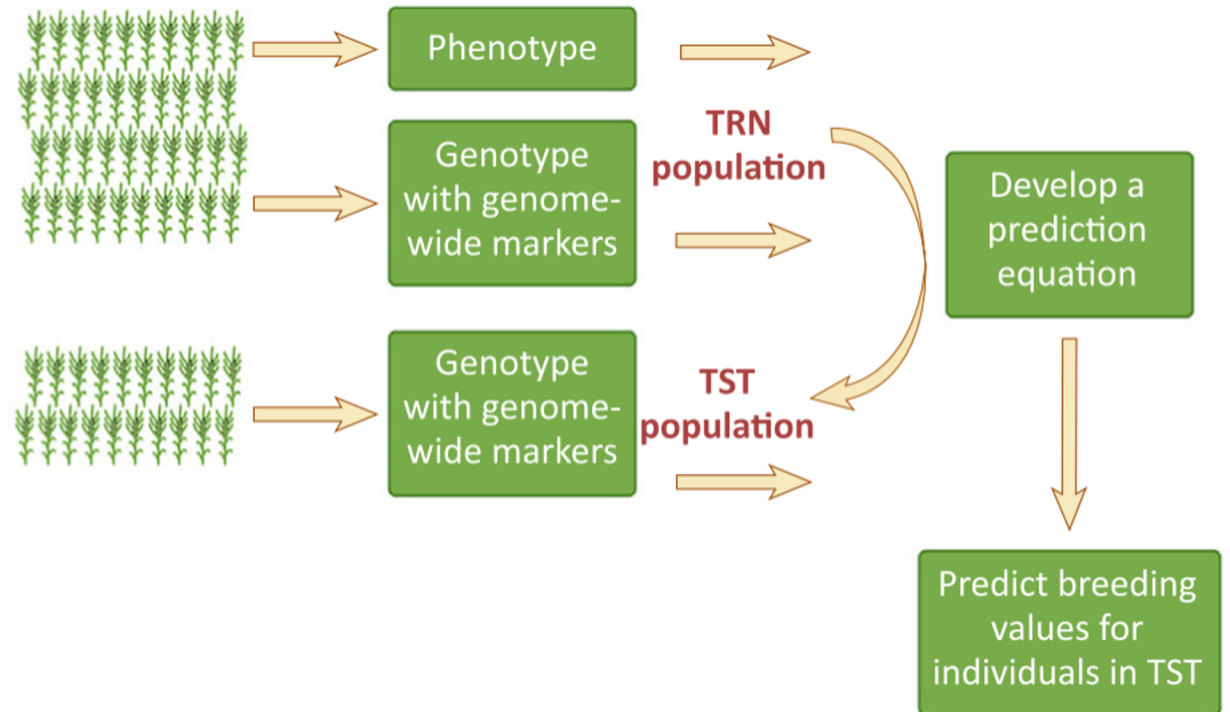
- GS uses all molecular markers to predict genomic estimated breeding values **(GEBV).**

- GS combines molecular and phenotypic data in a training population to obtain the GEBV of individuals in a testing population that have been genotyped but not phenotyped.



Crossa et al 2017

# Genomic selection
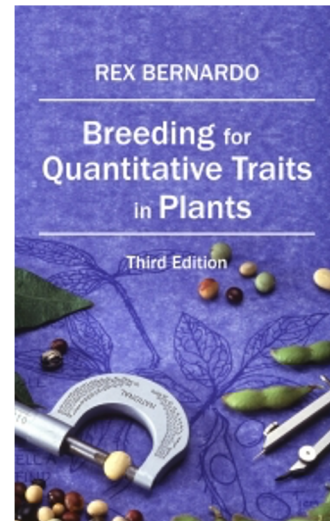
**University of Minnesota**

- Are interested in go deeper in GS basic model?

$$GEBV = \sum_i^p Z_i \, \alpha_i$$

https://www.youtube.com/watch?v=O7KYlSOZhZo&t=1939s

# MAS vs GS
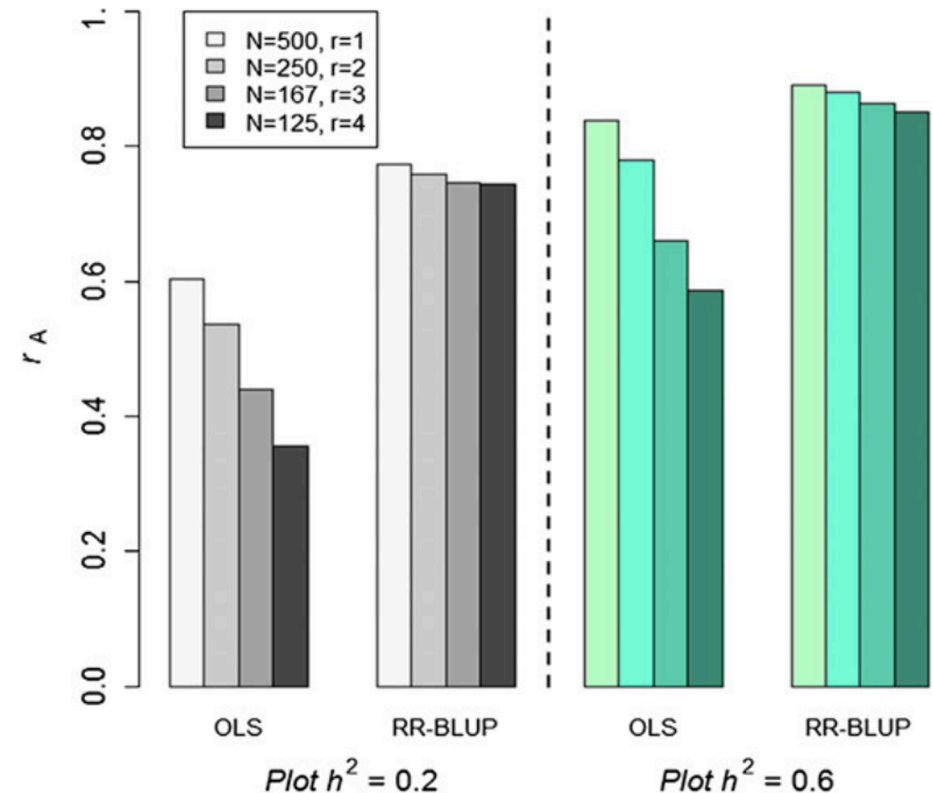
1. MAS VS GS
2. Population size
3. h²

A multiple regression model in which marker effects were estimated with ordinary least squares (OLS)

Ridge regression best linear unbiased prediction (RR-BLUP)
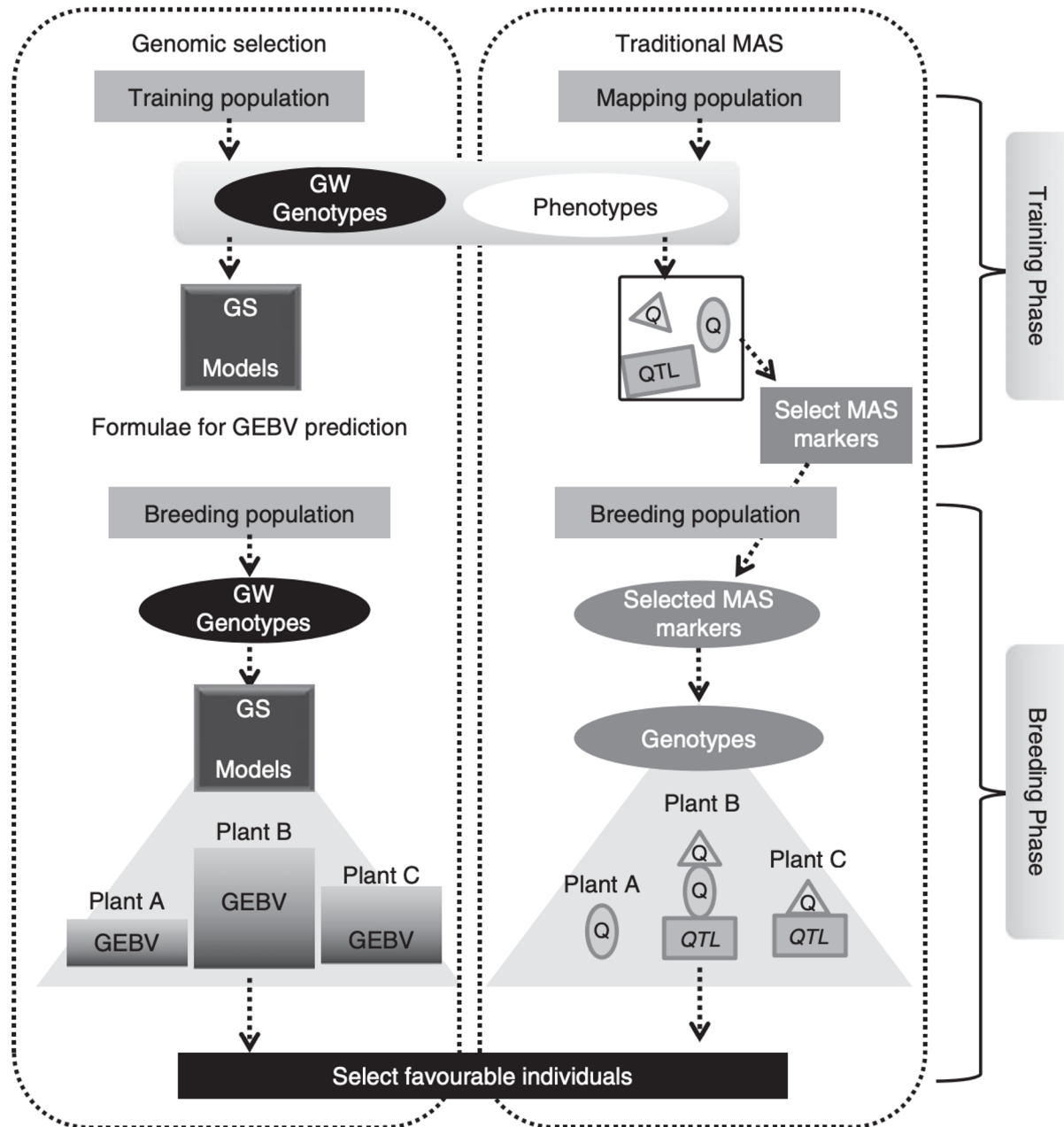
markers > population size

h² and number of molecular markers to set the phenotypic variance that can be explained in the model



Lorenz 2013

# MAS vs GS

Schemes of GS and traditional MAS for the selection of quantitative traits (right).



Nakaya and Isobe 2021

# GS Models



Gaussian — rr-BLUP

Thick Tail — Bayes A
- Double-Exponential (Bayesian Lasso)
- Scaled-t with 4DF (Bayes A)

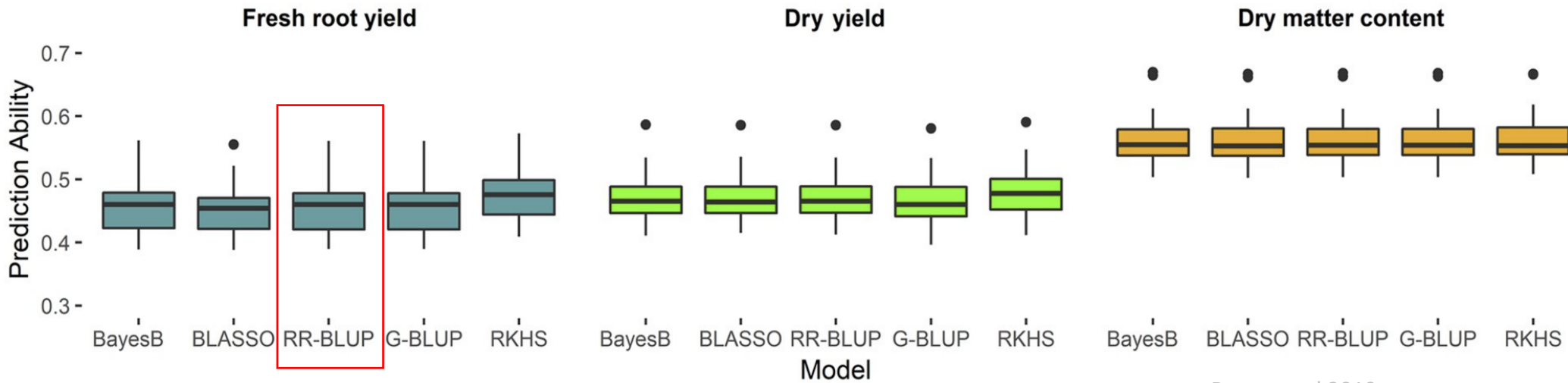Point of Mass & Slab — Bayes C
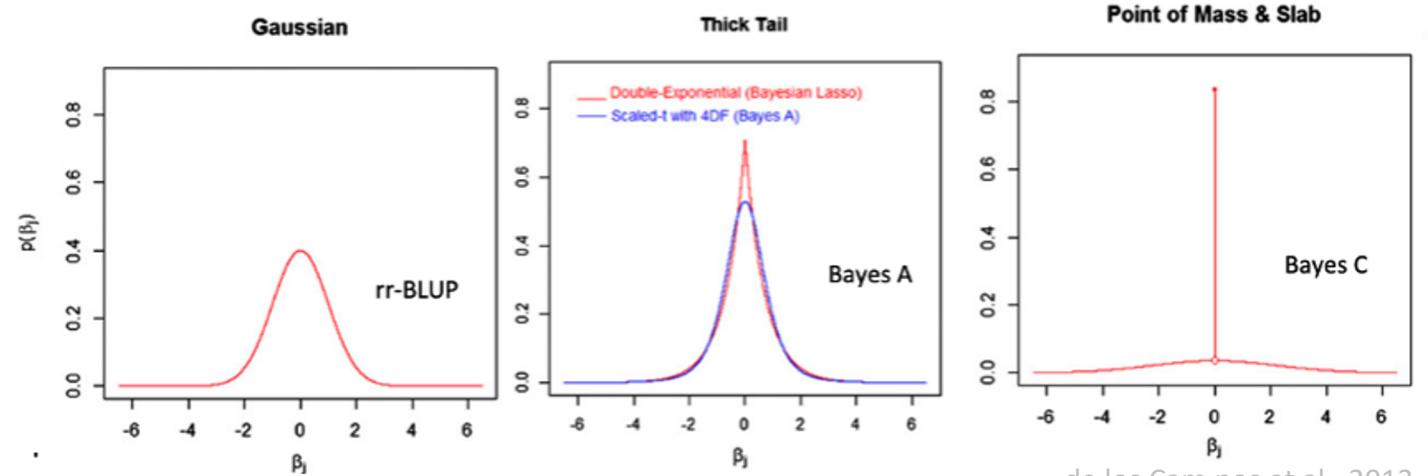
de los Cam,pos et al., 2013

In general, different genomic selection models tend to have similar predictive ability.

BayesC :     Small number QTL
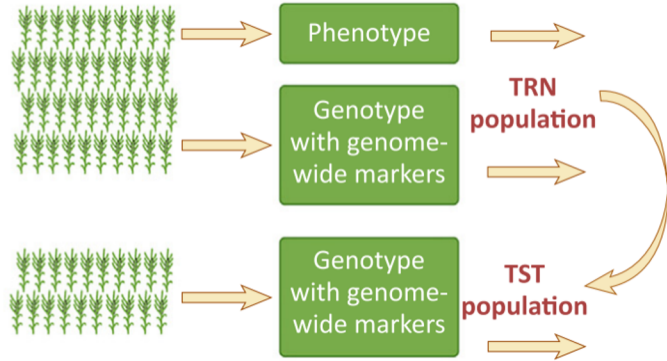BayesA:      Moderate number of QTL
RR-BLUP:     Extreme number of QTL

Wang et al., 2015



Fresh root yield — Dry yield — Dry matter content
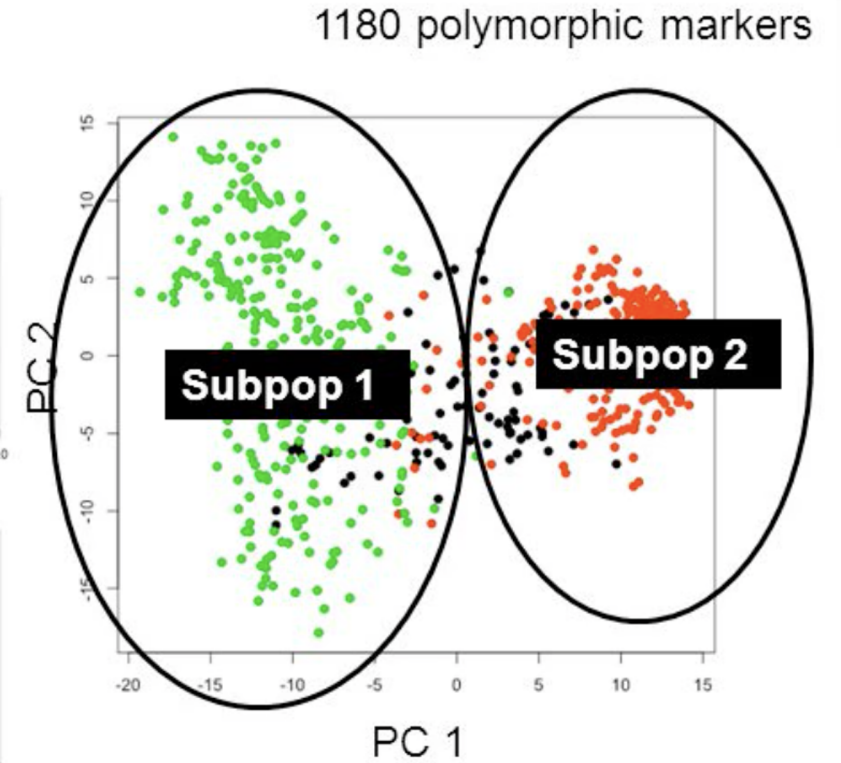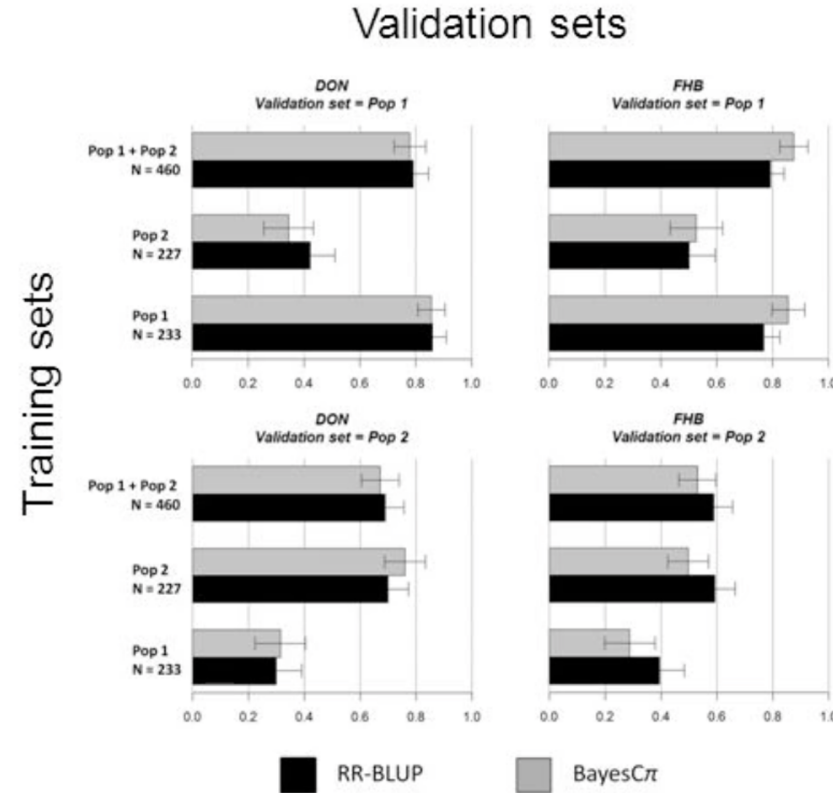Prediction Ability
BayesB   BLASSO   RR-BLUP   G-BLUP   RKHS
Model

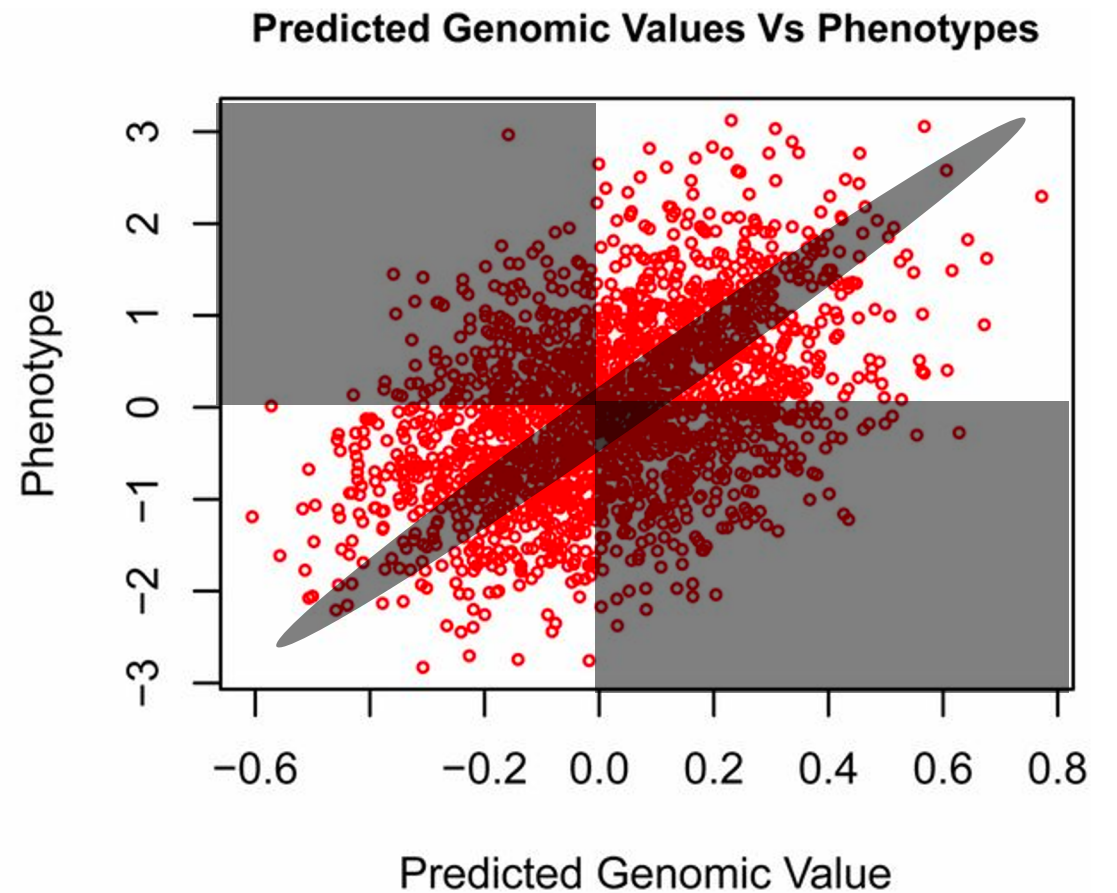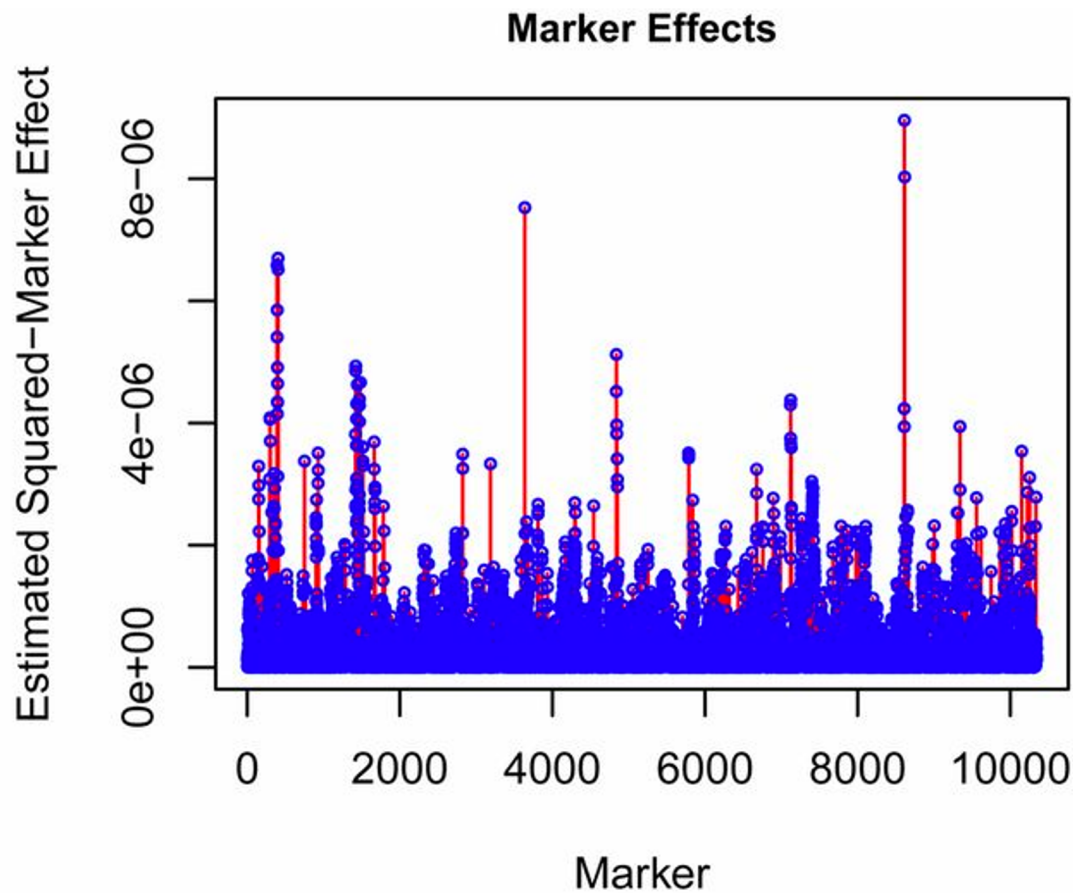Braatz et al 2019

# Population structure



Crossa et al 2017

Lorenz 2016

# Framework for GS

# Framework for GS



Marker Effects
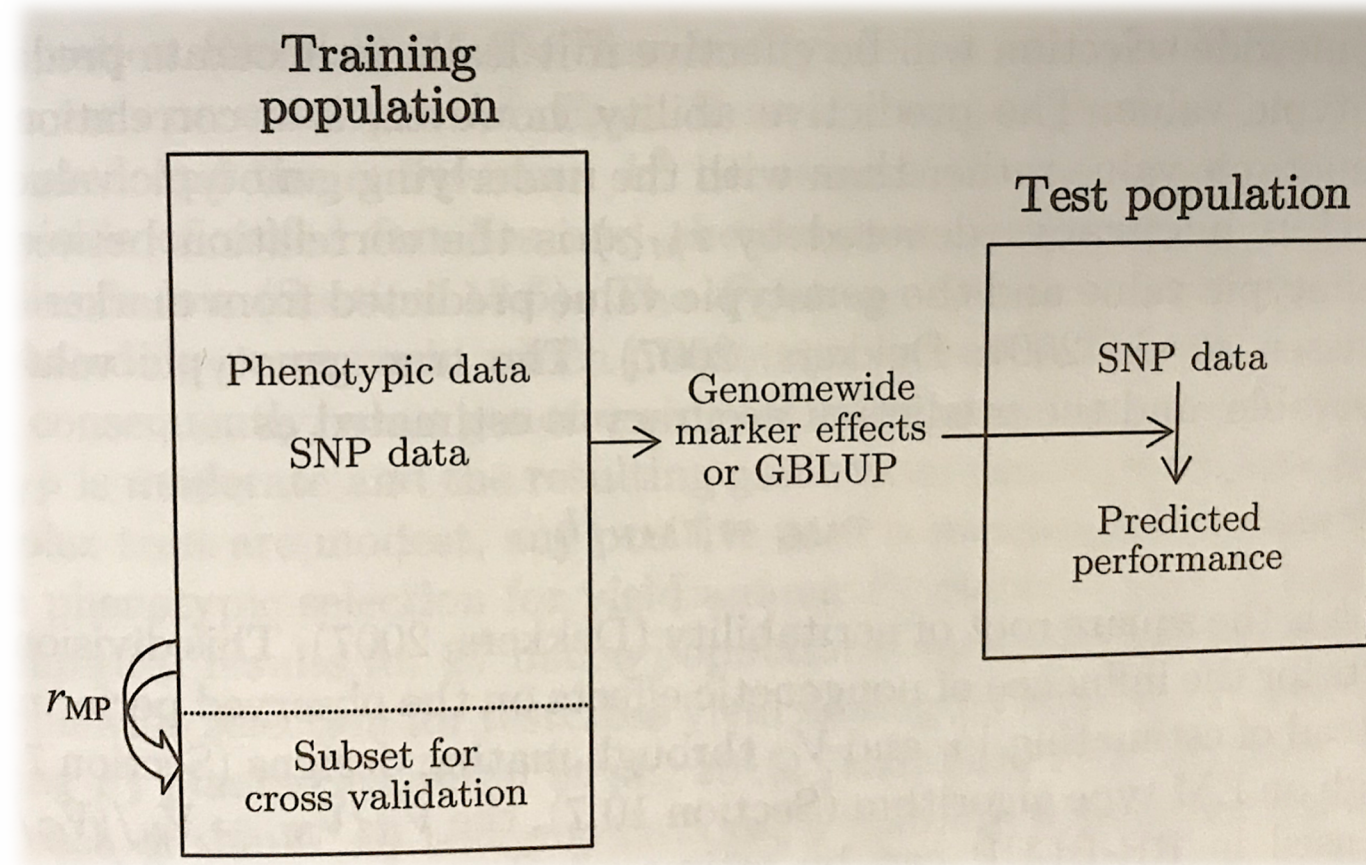
Predicted Genomic Values Vs Phenotypes

# GS vs traditional breeding

| Entries | Across eight populations | | | Five WEMA populations[†] | | |
|---|---|---|---|---|---|---|
| | GY | AD | PH | GY | AD | PH |
| | Mg ha⁻¹ | days | cm | Mg ha⁻¹ | days | cm |
| C₀ | 2.286 | 63.910 | 179.900 | 2.212 | 64.060 | 174.300 |
| C₁ | 2.420 | 64.080 | 181.900 | 2.482 | 64.160 | 178.100 |
| C₂ | 2.438 | **64.410** | **184.600** | 2.474 | 64.160 | **179.400** |
| C₃ | **2.593** | 64.100 | 182.200 | **2.581** | 63.730 | 175.600 |
| Pedigree | 2.417 | 64.400 | 181.700 | 2.417 | **64.400** | 181.700 |
| F₁ | 2.394 | 63.930 | 175.600 | 2.394 | 63.930 | 175.600 |
| Parents | 2.361 | 64.000 | 178.400 | 2.431 | 63.950 | 176.600 |



Beyene et al., 2015

14
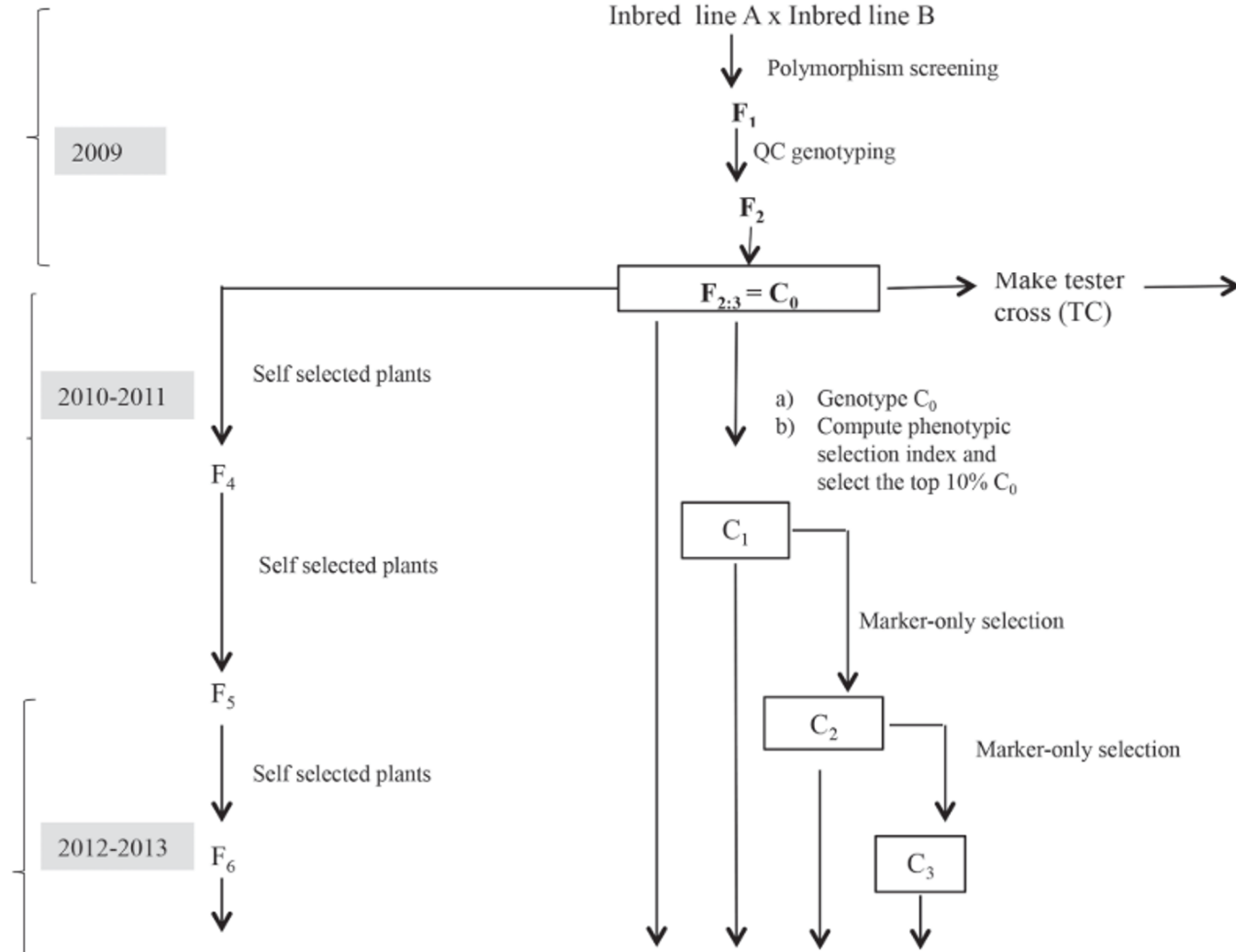
# GS is not the solution of all our problems
## ...but it is helpful

- Reduce breeding cycle
- Increase selection intensity
- Increase genetic variance
- Reduce costs

$$\Delta G_{year} = \frac{i \; r_{AI} \; \sigma_A}{L}$$

$i$ = Selection intensity

$r_{AI}$ = Accuracy

$\sigma_A$ = Genetic standard deviation

$L$ = Generation interval

In the next presentation we are going to talk about it!