

14.02.2024 - WORKSHOP

Genomic tools for berry pre-breeding material : from genome to new berry cultivars

**An introductory hands-on to the application
of genomic prediction**

INRAE : Juliette Bénéjam, Béatrice Denoyes, Alexandre Prohaska, Pol Rey-Serra

INVENIO : Justine Perrotte, Aurélie Petit

INRAE



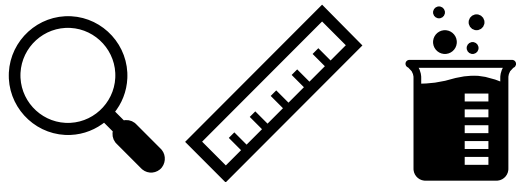
Invenio



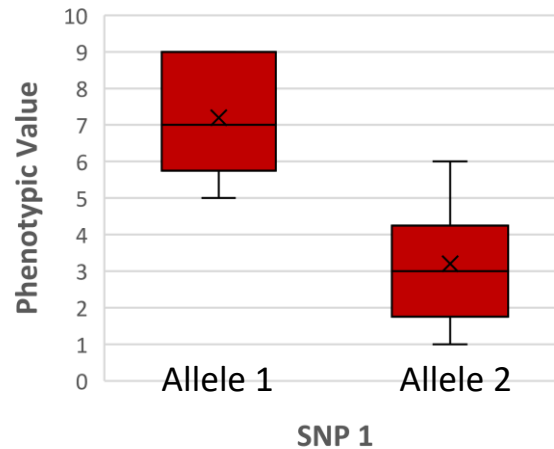
Evolution of breeding strategies : Phenotypic -> Genomic Prediction

Meuwissen 2001

Screening based on phenotypes



Marker-assisted selection



Genotyping of **SNP 1**
Allele 1 = Phenotypic value



New Issues

More than one trait selected
Trait regulated by more than one SNP



Evolution of breeding strategies : Phenotypic -> Genomic Prediction

Selection based on 1 genetic markers

	SNP1
genotype A	1
genotype B	2
genotype C	1
genotype D	1
genotype E	1
genotype F	2
genotype G	1
genotype H	2
genotype I	2
genotype J	1
...	



Genetic regions that regulate 1 trait

Marker-assisted selection

Evolution of breeding strategies : Phenotypic -> Genomic Prediction

Meuwissen *et al.* 2001

Selection based on a large number of genetic markers

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	...	SNP +
genotype A	1	1	1	1	1	1	2	1	1		2
genotype B	2	1	2	1	2	1	1	1	2		2
genotype C	1	1	2	2	2	2	1	1	1		1
genotype D	1	2	2	1	1	1	1	1	2		1
genotype E	1	2	2	2	2	2	1	1	1		2
genotype F	2	2	1	1	2	1	2	1	1		2
genotype G	1	1	1	2	1	2	2	1	2		2
genotype H	2	1	1	2	1	2	2	1	2		1
genotype I	2	2	2	2	1	1	1	1	1		2
genotype J	1	2	1	2	2	2	1	1	2		2
...											



Genomic prediction

What is 'Genomic Prediction' ?

Meuwissen *et al.* 2001

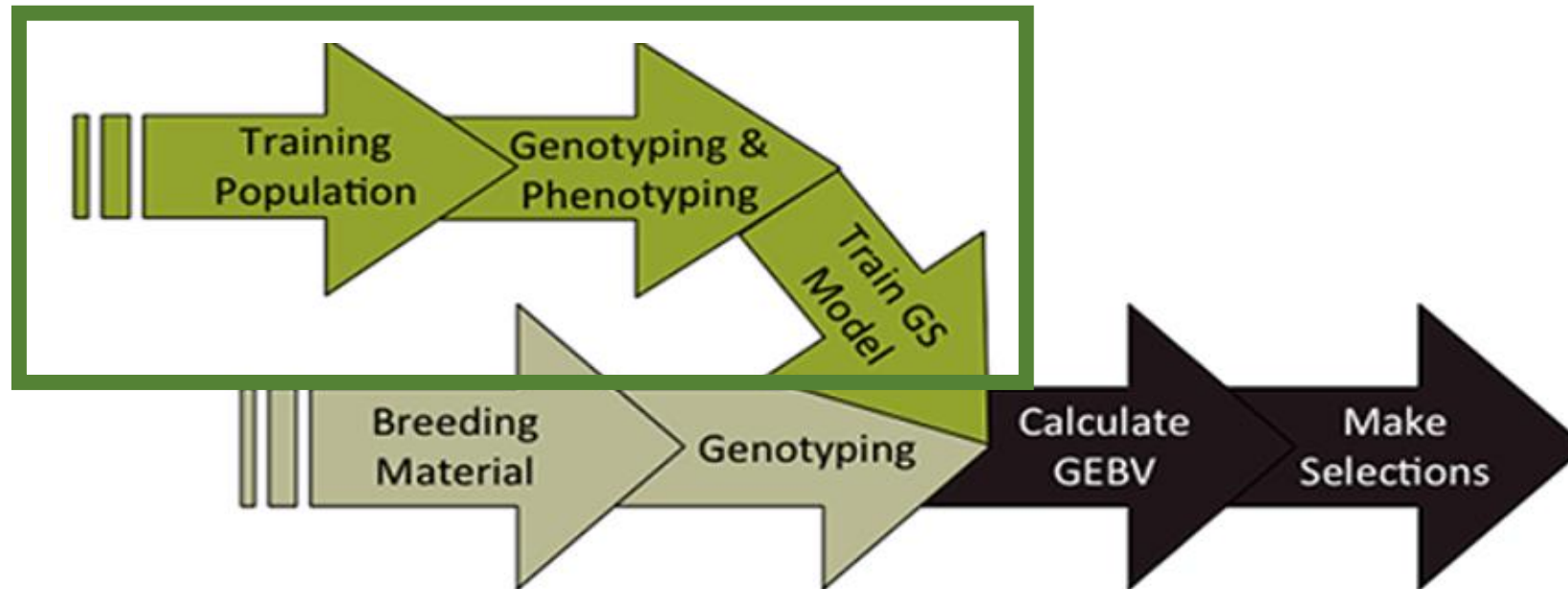
- **Statistical tool** developed to speed up the breeding process
- **Objective: Estimation of a 'Breeding value' based on whole genome information**



Breeding value : A value that **correlates with the phenotypic mean**

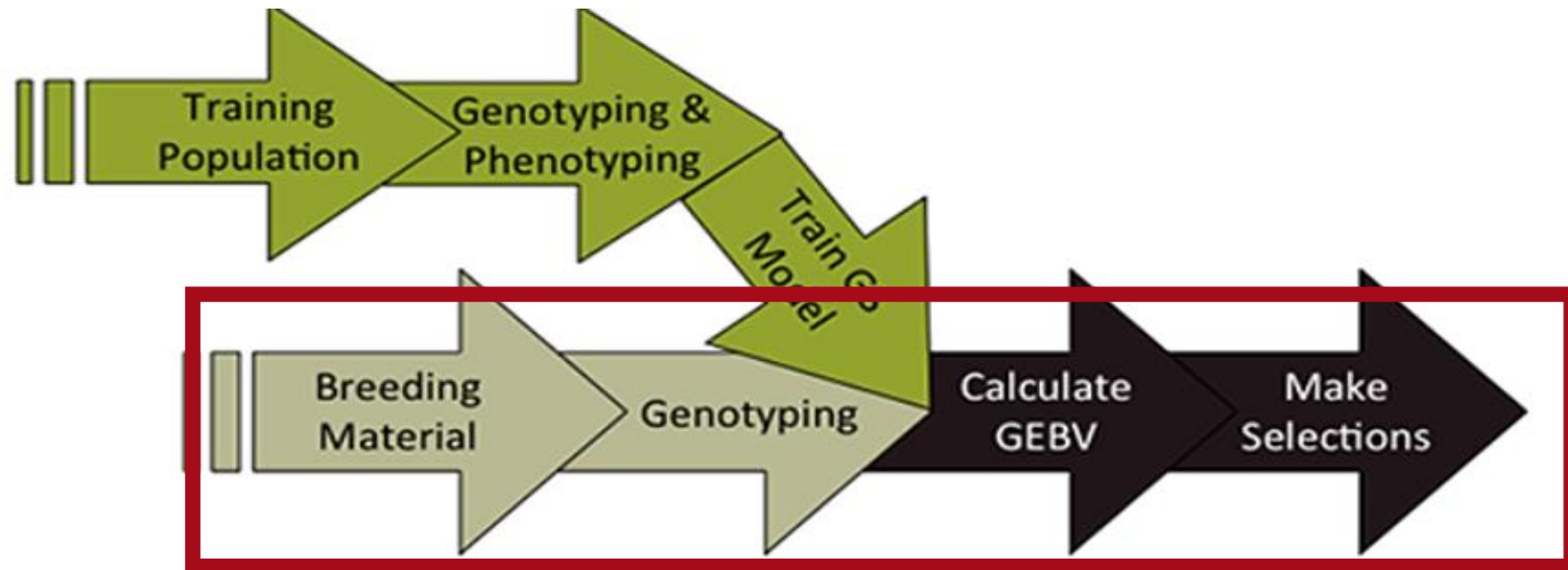
What is 'Genomic Prediction' ?

1- Design of the prediction model using a training population



Heffner *et al.* 2009

What is 'Genomic Prediction' ?



Heffner *et al.* 2009

2- Model application

Selection of the training population



Populations

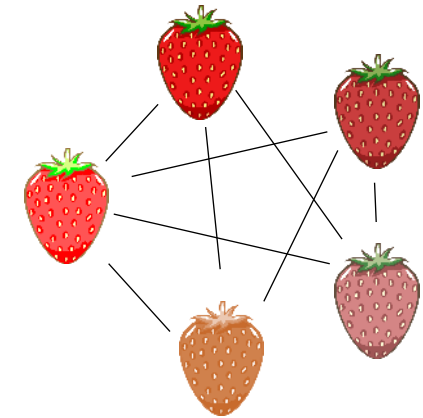


Genetic diversity collection

Progenies



Bi-parental population



Connected population

Data needed for genomic prediction



Populations



Genetic markers = dense genetic map

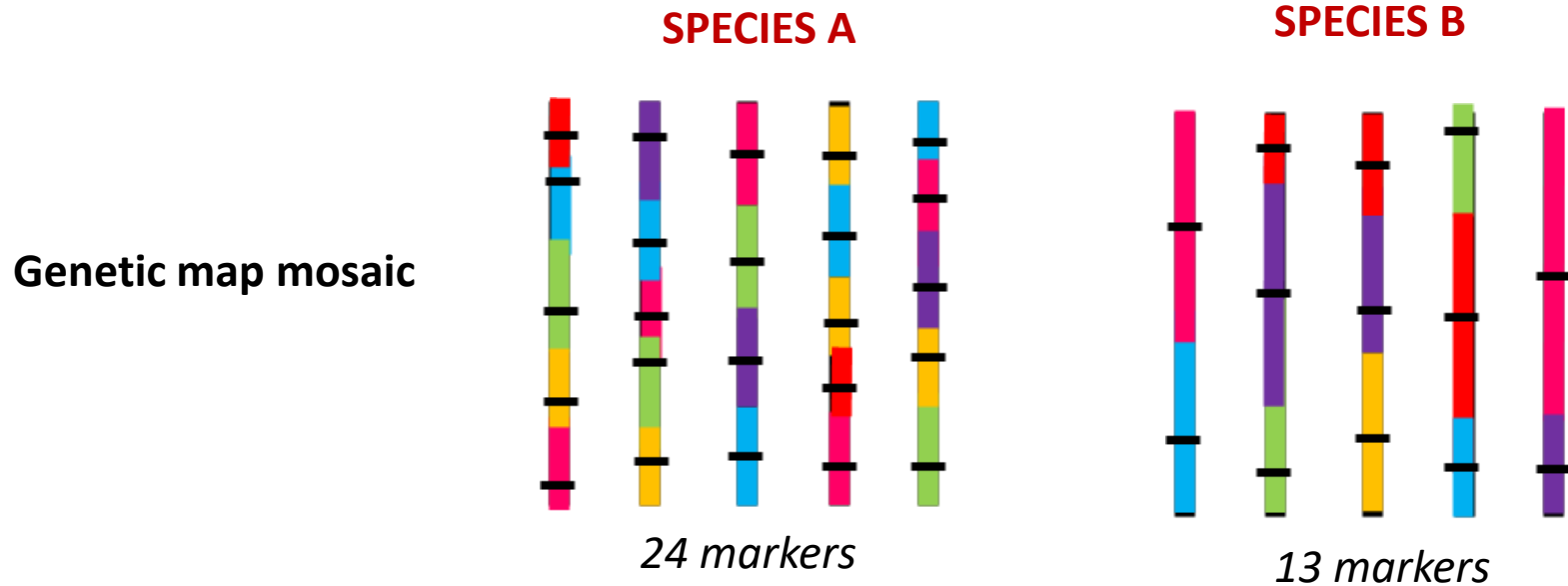
Parameters:

Type of marker,
Density of markers,
Coverage,
Polymorphism,
Minor allele frequency

The number of markers for 'good coverage' depends on the species

Parameters:

Type of marker,
Density of markers,
Coverage,
Polymorphism,
Minor allele frequency



Groups of markers transmitted together : **A < B**

Number of markers needed : **A > B**

Data needed for genomic prediction



Populations



Genetic markers = dense genetic map



Phenotypic scoring



Parameters:
Quantitative scale,
Distribution of the trait,
Variability in the population,
Quality of assessment,
Transformation of raw data
Environmental effect/plasticity

To take **all the SNP information** into account at the same time, we need a **model**

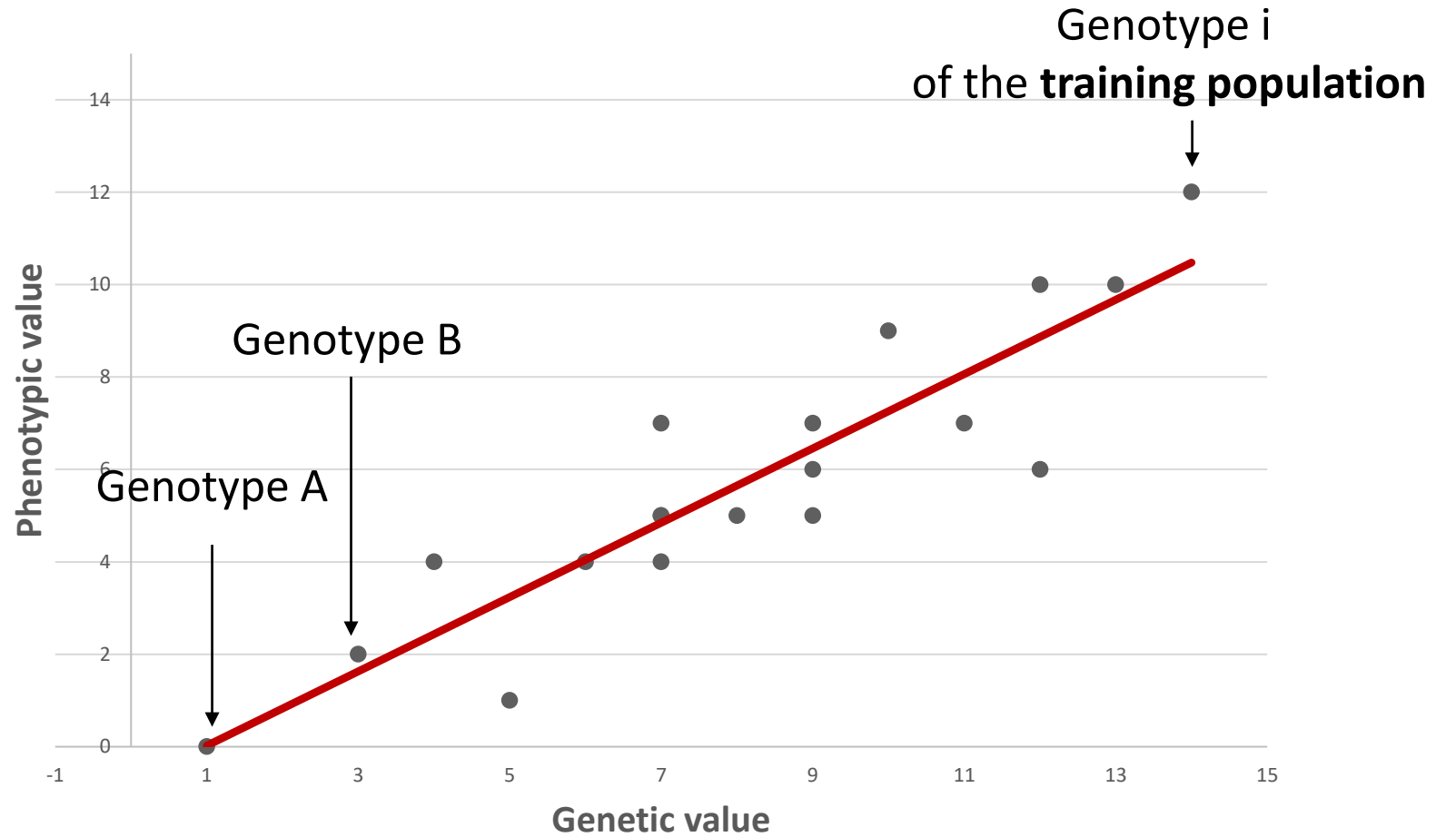


Populations



Genetic markers = dense genetic map

Phenotypic scoring



To take **all the SNP information** into account at the same time, we need a **model**

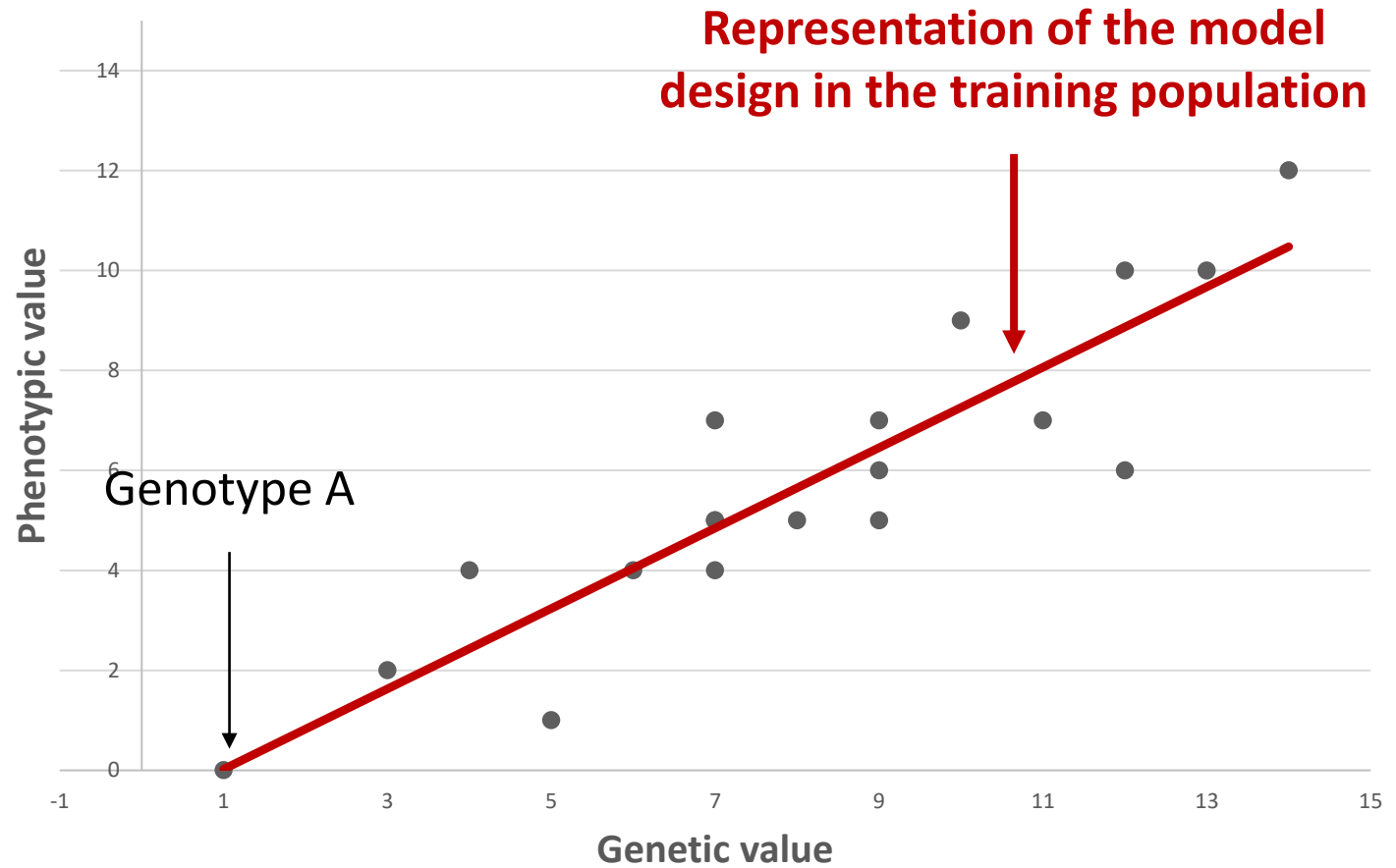


Populations



Genetic markers = dense genetic map

Phenotypic scoring



$$Y = a * x + b$$

$$\text{Phenotype value (genotype A)} = a * \text{Genetic value (genotype A)} + b$$

To take **all the SNP information** into account at the same time, we need a **model**

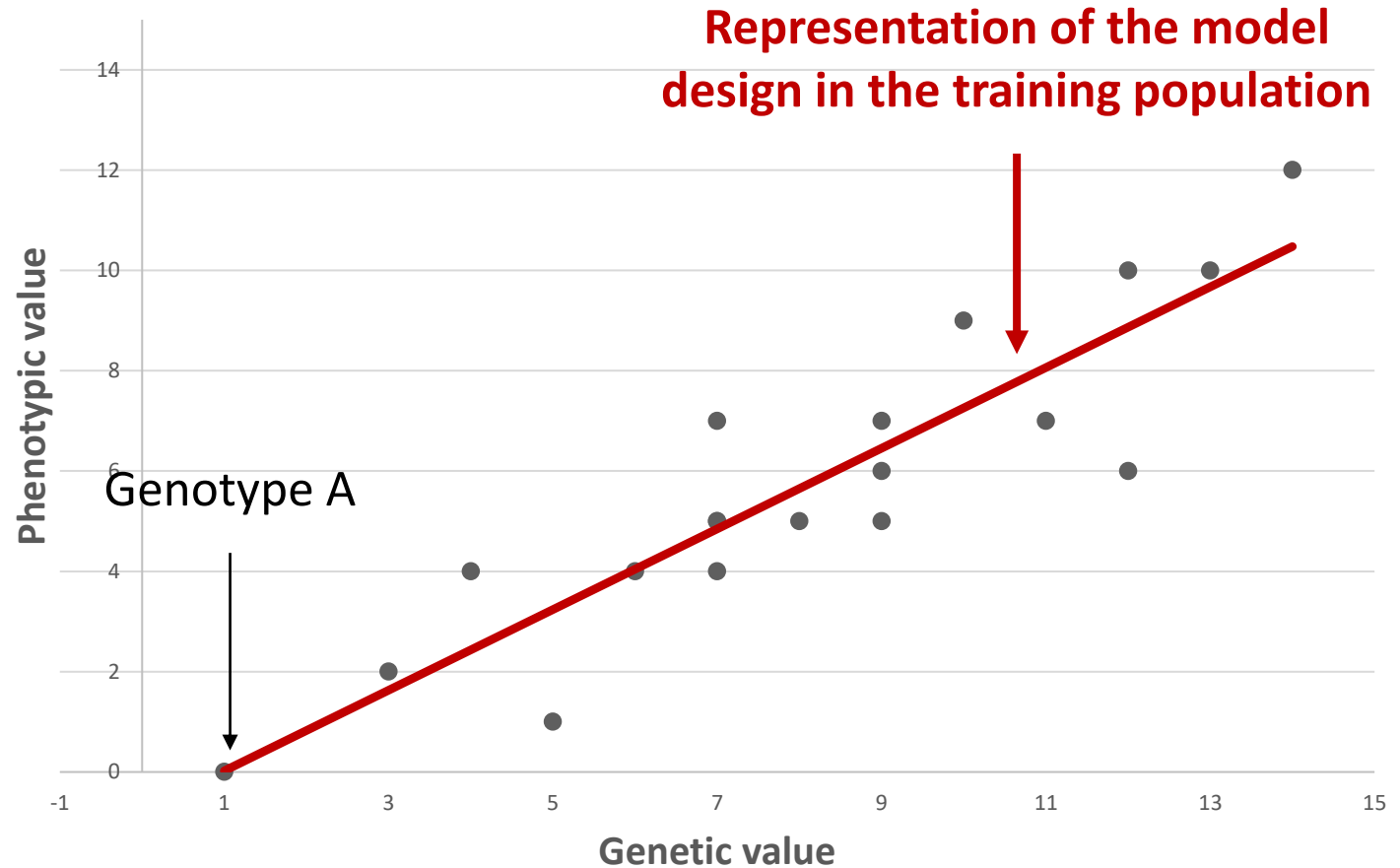


Populations



Genetic markers = dense genetic map

Phenotypic scoring



$$Y = a * x + b$$

$$\text{Phenotype value (genotype A)} = a * \text{Genetic value (genotype A)} + b$$

How do you calculate genetic value from markers???

$$Y = a * x + b$$



Whittaker *et al.* 2000

$$y_i = \mu + \sum_k w_{ik} \beta_k + e_i$$

Y = vector of phenotypes
 y_i for individual i

Y_i = phenotypic value (genotypeA)

μ overall mean

w_{ik} is the genotype at locus k {0, 1, 2}

β_k the effect of this allele

e_i residual vector

$$e_i \sim \tilde{N}(0, \sigma_e^2)$$

Allele SNP1 (genotypeA) * SNP1 effect + Allele SNP2 (genotypeA) * SNP2 effect +

Genomic Best Linear Unbiased Prediction (GBLUP)

Parameters: Statistical framework (frequentist or bayesian), Resolution method, Input data (genetic matrix RRBLUP or kinship GBLUP), Shrinkage (RR, LASSO, intermediates Bayes alphabet/Elastic Net),

New variety

= **unknown phenotype**



Genotyping of the genetic markers (same as training population) in the **new variety**



$$Y = a * x + b$$



Solving the model with the genetic markers of the new variety
Calculation of y_i



Predicted phenotypic value of the new variety

Key point: Evaluation of new genotypes using only genetic information (earlier + faster assessment, more accurate, takes all genetic information into account)

Model design

Application

Optimization

Perspectives

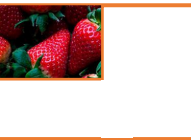


Whole population



75% of whole pop
Training population TS

25% of whole pop
Validation population VS



Examples of results:
Powdery mildew resistance
prediction on strawberry

Cross-validation

For testing the efficiency of genomic prediction



Examples of results:
Powdery mildew resistance prediction on strawberry



Whole population

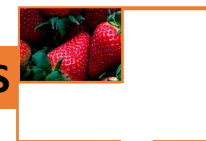


75% of whole pop

Training population TS

25% of whole pop

Validation population VS



Design of model



Prediction of pheno.pred in VS

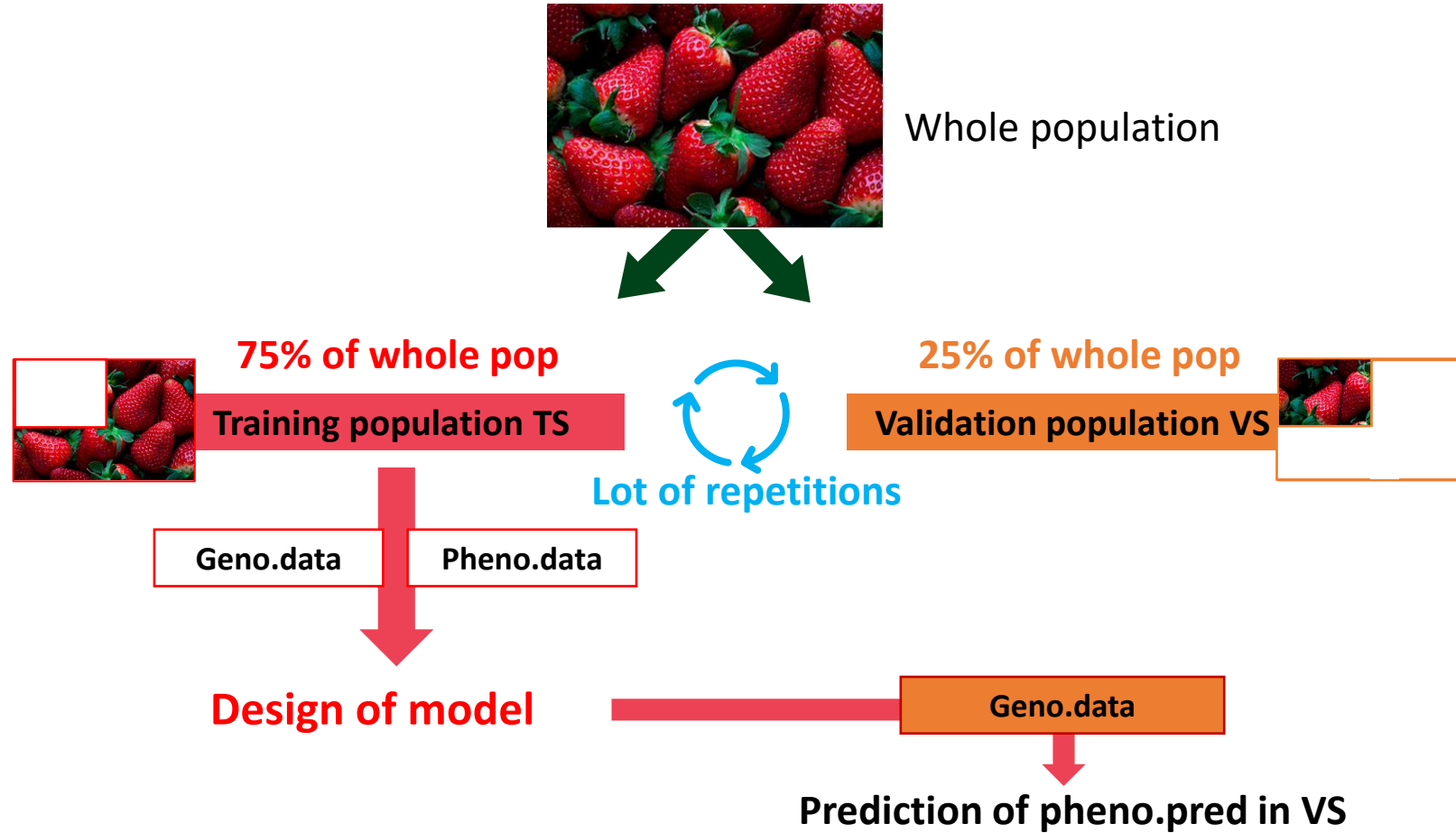
$$\text{Prediction ability} = \text{cor}(\text{predicted pheno}, \text{observed pheno})$$

Cross-validation

For testing the efficiency of genomic prediction



Examples of results:
Powdery mildew resistance prediction on strawberry



$$\text{Prediction ability} = \text{cor}(\text{predicted pheno}, \text{observed pheno})$$

Cross-validation

For testing the efficiency of genomic prediction



Examples of results:
Powdery mildew resistance
prediction on strawberry

Training population =
collection of 165 genotypes



Y = vector of susceptibility values
X = matrix with 33K SNPs



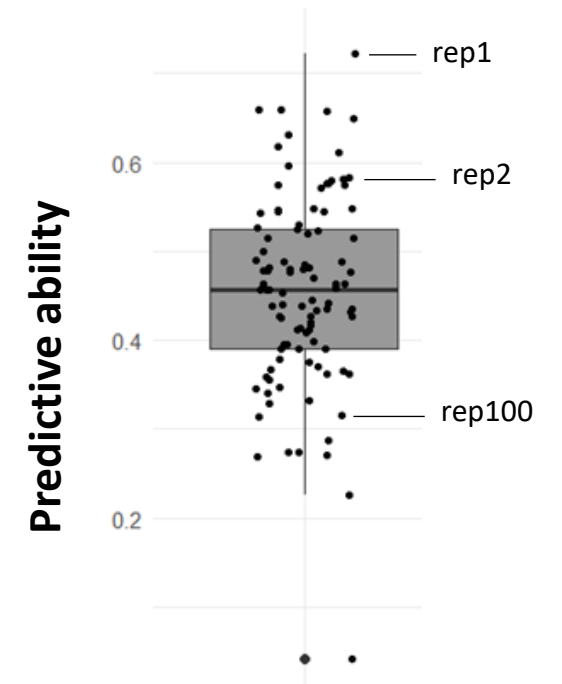
Model : RRBLUP (x100 repetitions)

Validation population =
55 genotypes



Calculation of **predictive ability**
~ **Model efficiency / accuracy**

Boxplot of result





Examples of results:
Powdery mildew resistance prediction on strawberry

Training population =
collection of 165 genotypes



Y = vector of susceptibility values
 X = matrix with 33K SNPs



Model : RRBLUP (x100 repetitions)

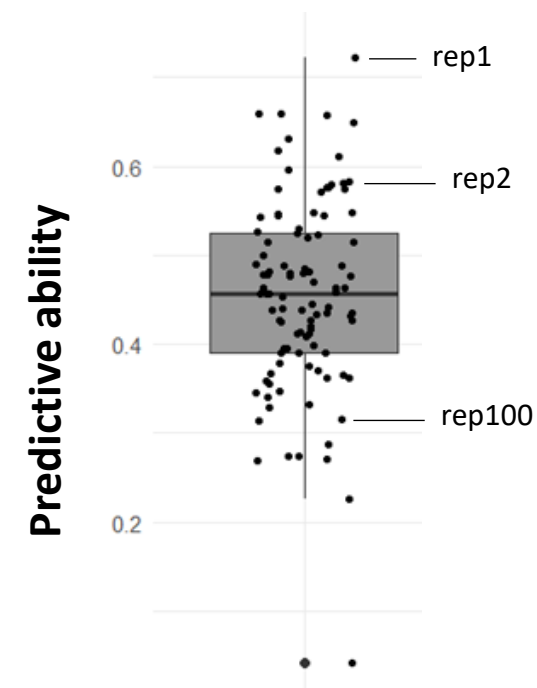
Validation population =
55 genotypes



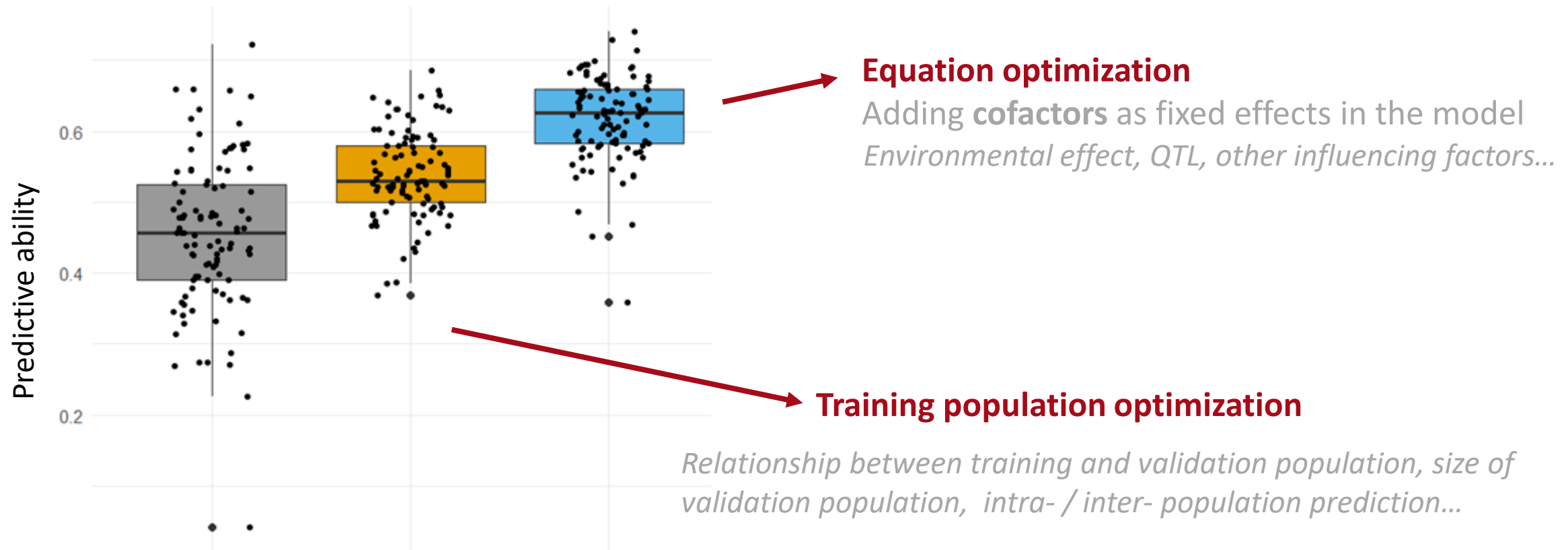
Calculation of predictive ability
~ Model efficiency / accuracy

Improvement of model efficiency ?

Boxplot of result



2 examples of ways to improve predictive ability



Many other options

Statistical model -> depends on trait architecture

Genetic markers -> number and filter for quality

Polyploids -> Consider markers with allelic frequency

Cross validation scheme

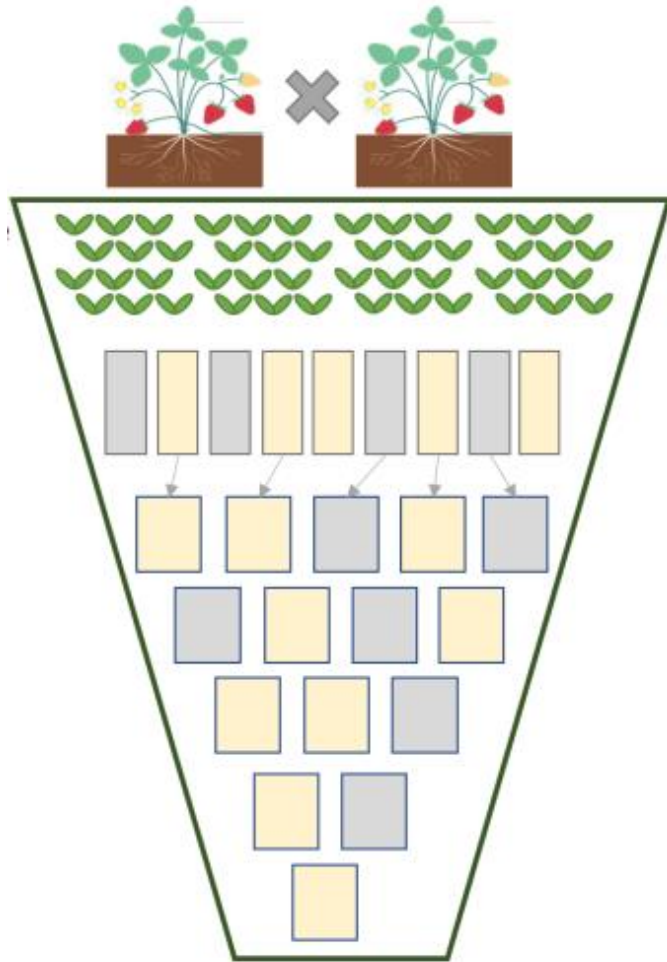
Population structure

Kinship for non-additive effects

Multivariate models

... Infinite and linked to the breeding objectives ...

Perspectives in breeding process



Prediction of **parental values**

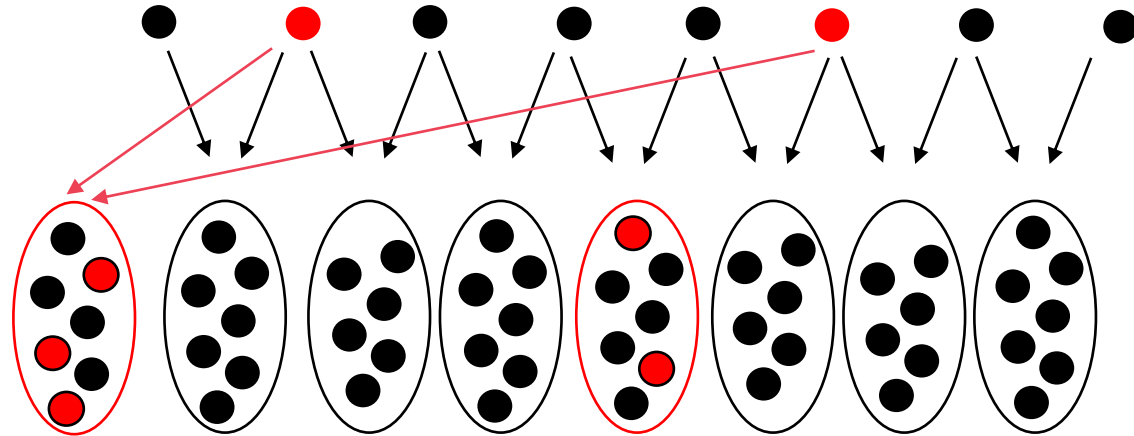
Prediction of the 'best' crosses by **simulations of progenies**

Parents

1- Best breeding values

Progenies

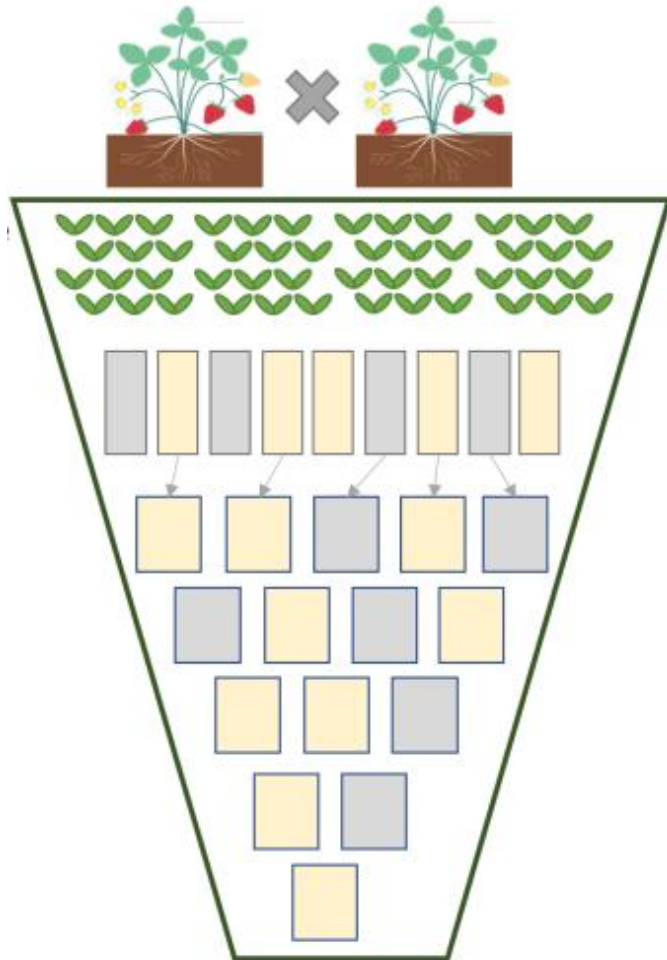
Simulations
of crosses



2-Best parental values

= favorable combination of alleles in progeny

Perspectives in breeding process



Prediction of **parental values**

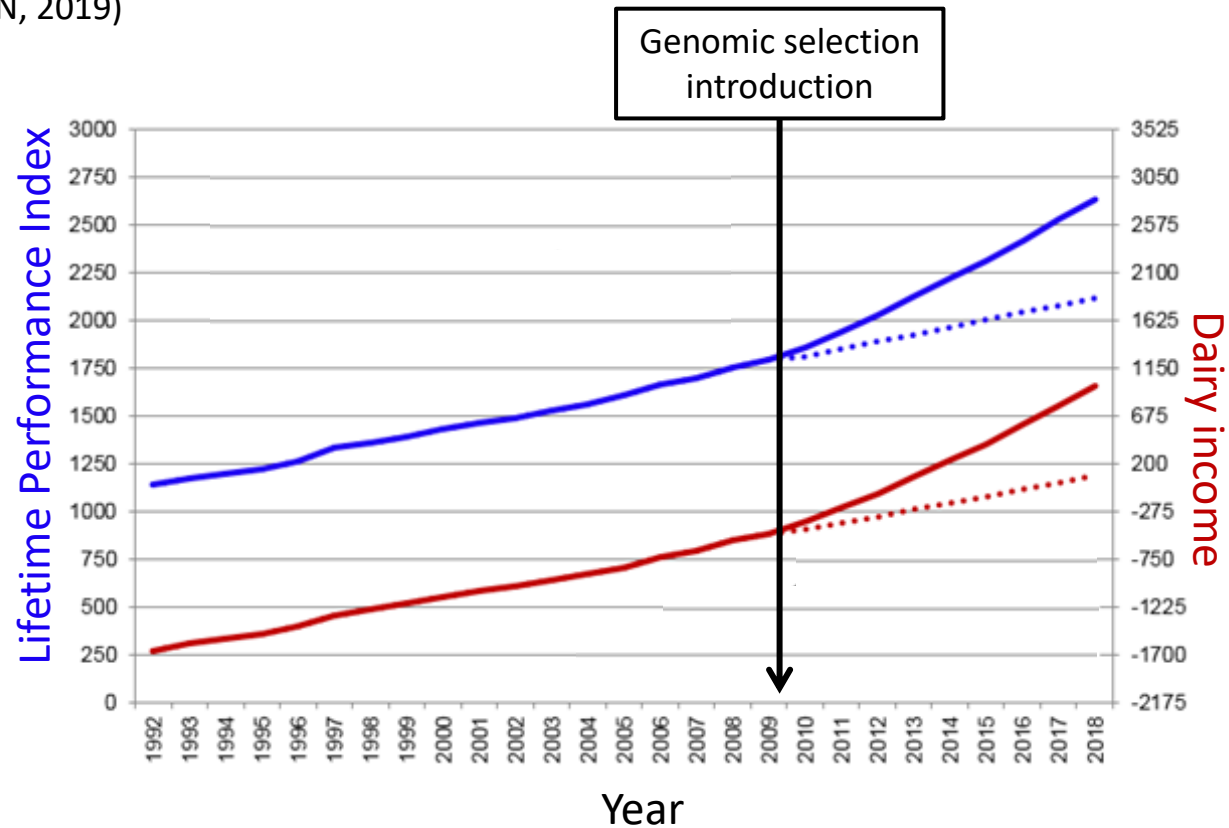
Prediction of the 'best' crosses by **simulations of progenies**

Prediction of phenotypic values of new hybrids in screening steps (replacing a part of phenotypic scoring)

Perspectives in breeding process

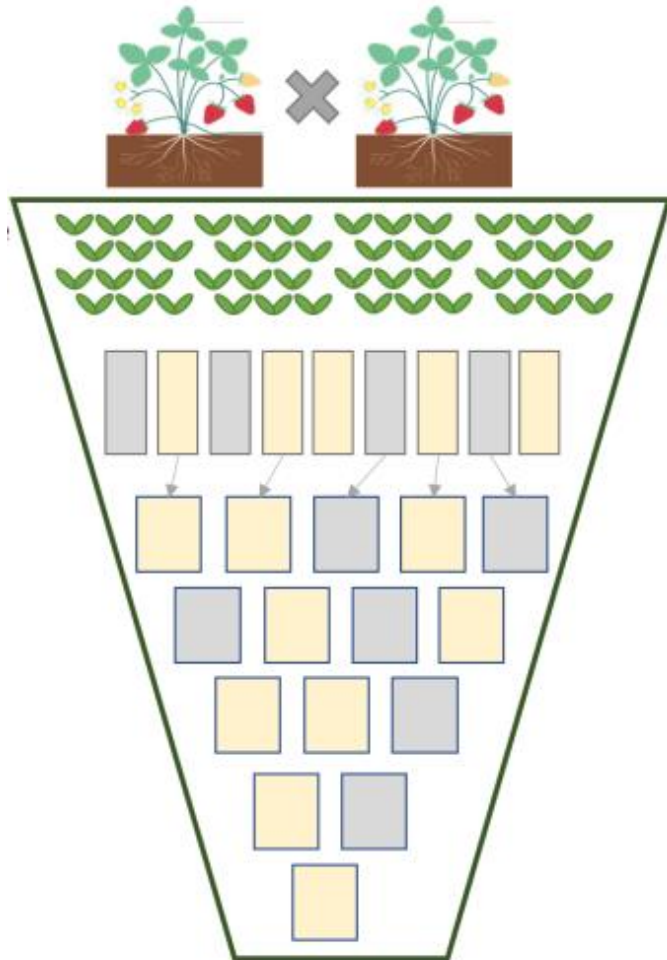
Gains index accross year in Holsteins breed

(Adatped from CDN, 2019)



**Gains = x2 / year
with genomic selection**

Perspectives in breeding process



Prediction of **parental values**

Prediction of the 'best' crosses by **simulations of progenies**



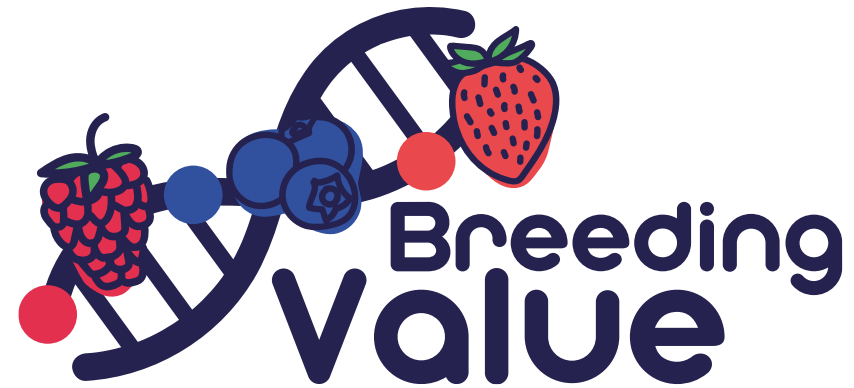
Prediction of phenotypic values of new hybrids in screening steps (replacing a part of phenotypic scoring)

Conclusion: Genomic prediction is a method which needs time to be improved but which is very promising for the selection of complex and integrated traits.

Acknowledgements

INRAE *BFP*

Team FDFE



Breeding value, European project

Invenio

