



The Genomic Selection of *Theobroma cacao*: a new strategy of marker assisted selection to improve breeding efficiency and predict useful traits in new populations

F. Ribeyre (1), O. Sounigo (2), X. Argout (2), C. Cilas (1), B. Efombagn (3), M. Denis (1), J.M. Bouvet (4), O. Fouet (1), C. Lanaud (1)



agropolis fondation

(1) CIRAD, France

(2) CIRAD, Colombia

(3) IRAD, Cameroon

(4) CIRAD, Madagascar

# Why Genomic Selection for cocoa ?

Markers assisted selection: Markers significantly associated with trait



Genomic selection: all genetic markers simultaneously (Meuwissen et al. 2001)

A solution for the prediction of performance in complex traits ?

accuracy of Genomic selection depends on:

- linkage disequilibrium between markers
- the heritability of the trait
- the size of the training population
- the relationships between the training sets and the test sets
- the number of markers
- the statistical method to estimate the GEBV
- the distribution of underlying QTL effects
- the genotype x environment interaction ...

# What is Genomic Selection ?

- Training population: genotyped and phenotyped  
model  GEBV: sum of all markers effects by regressing phenotypic values on all available markers.
- test population: genotyped  
 predictions

# Objectives

Two traits



“Heritable”



“Less heritable”

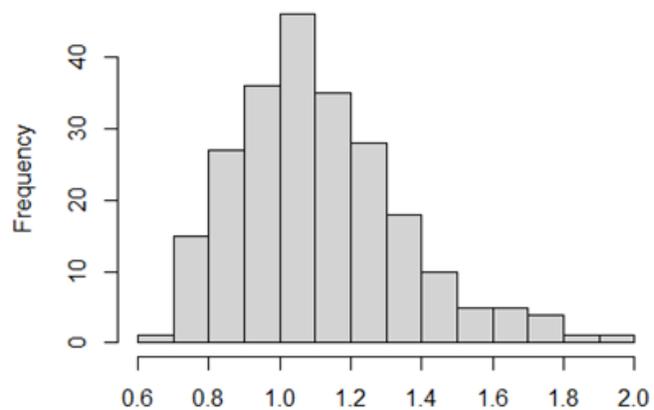
*Phytophthora megakarya*

- Evaluate 2 models
- Evaluate the predictive abilities of models

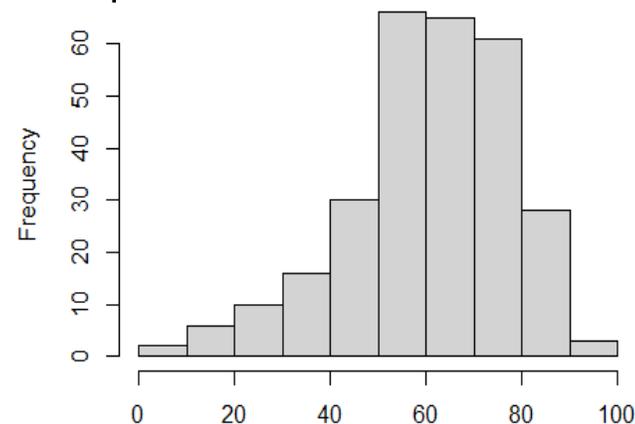
# Data to test genomic selection

A cacao farm plot in Cameroon with a mixture of hybrids

Histogram of the average weight of a bean - 232 trees



Histogram of the percentage of rotten pods - 287 trees during 3 years



Cocoa trees	Marker 1	...	...	...	Marker 50224	Average weight of a bean	% of rotten pods
Tree 1							
...							
Tree 287							

# The statistical challenge

## Best linear unbiased prediction model based on markers (G-BLUP)

Mixed model that suppose a normal distribution of markers effects with same variance

## A large number of QTL with small effects

genetic values are modeled as  $u \sim N(0, U\sigma_u^2)$ ,

where  $U$  is the realized relationship matrix calculated from the markers and  $\sigma_u^2$  is the genetic variance pertaining to model

## Bayesian lasso model (BL)

We suppose a double exponential distribution of markers effects.

## A lot of markers with effects near 0 and some with moderate to large effects

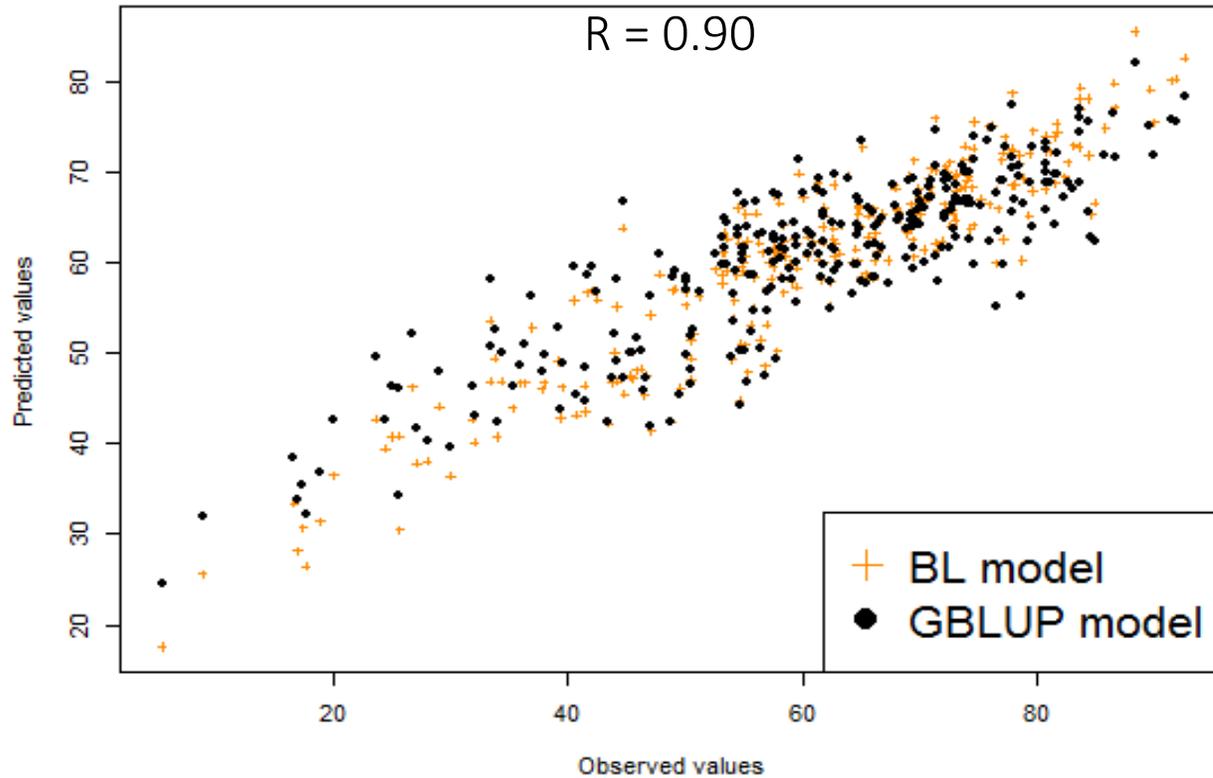
$m \sim N(0, T\sigma^2)$

$T = \text{diag}(\tau_1^2, \dots, \tau_j^2, \dots, \tau_p^2)$

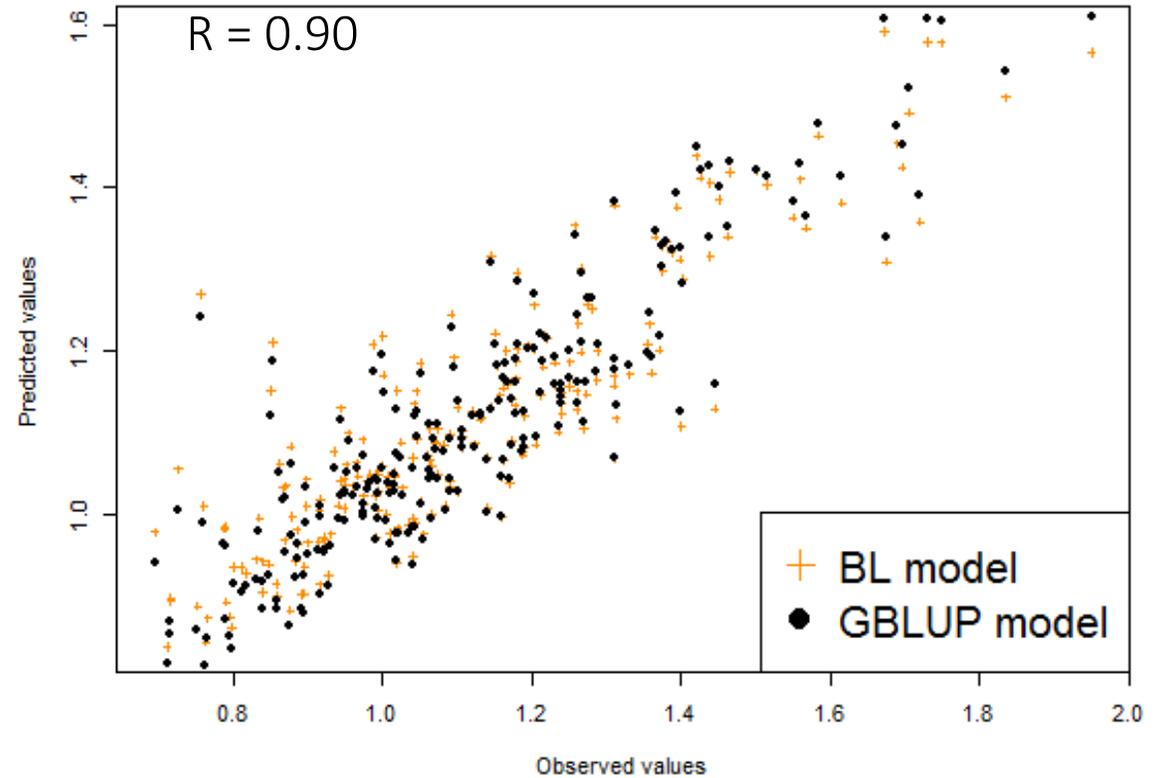
# Results using R-package synbreed

(Wimmer et al., 2012)

Percentage of rotten pods



Average weight of a bean



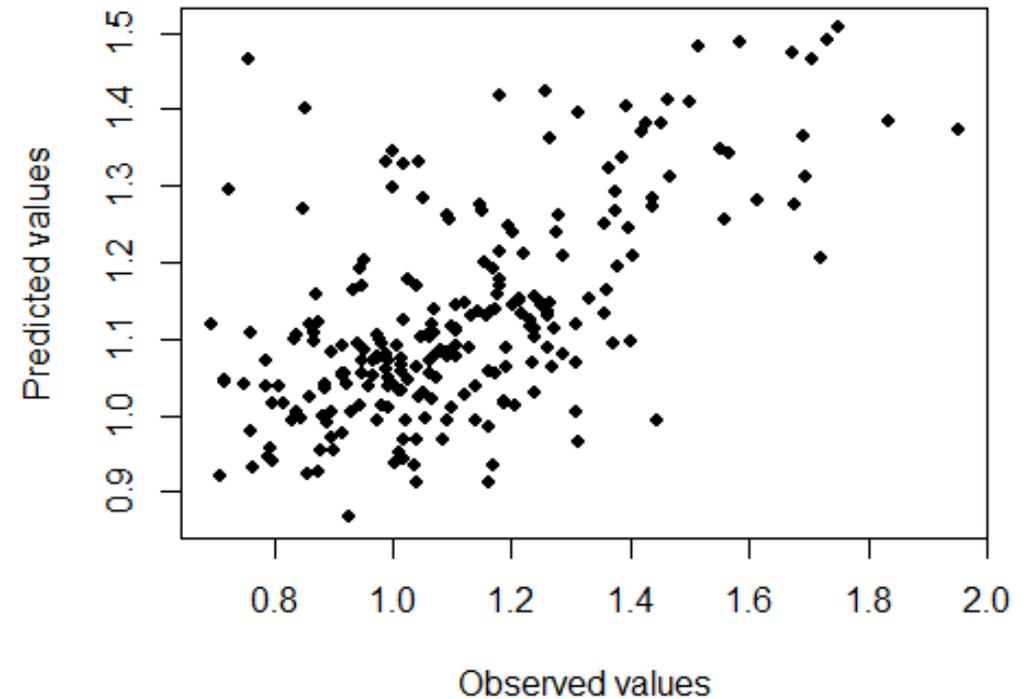
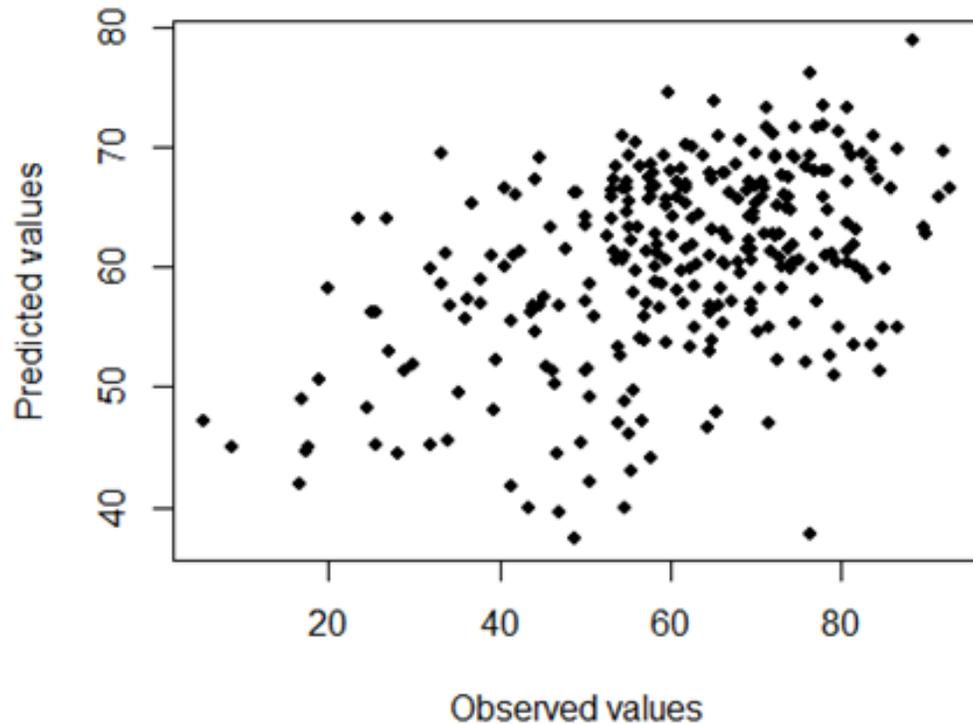
# Evaluate models : cross validation

## % rotten

Predictive ability = **0,42** (GBLUP 2 folds)  
= **0,37** (BL 2 folds)

## Weight

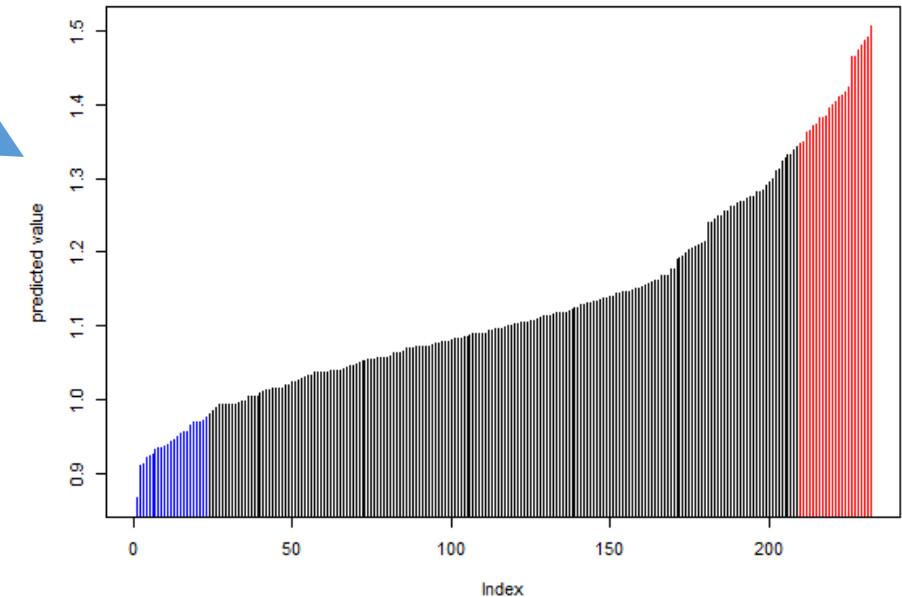
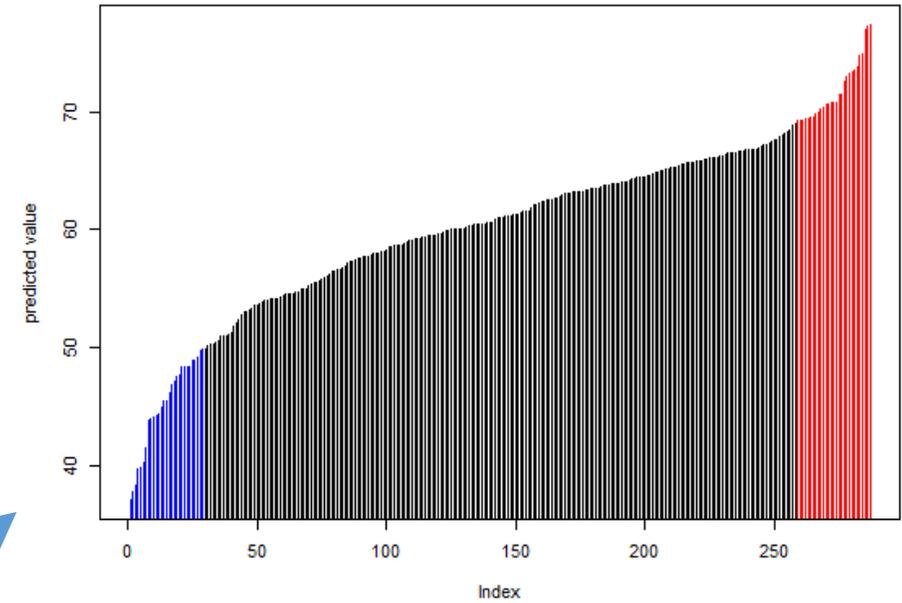
Predictive ability = **0.59** (GBLUP 2 folds)  
= **0,58** (BL 2 folds)



# Evaluate models

	Observed mean of 10% higher predicted	Observed mean of 10% lower predicted
Percentage of rotten pods	67	48
Average weight of a bean	1.6	0.96

Good differentiation



# Conclusion

- Good predictive ability of models
- A promising method to improve these cocoa traits
- predict tolerant cocoa trees to disease only present in another environment ?