# Genomic prediction for tolerance to Al in a synthetic population of upland rice

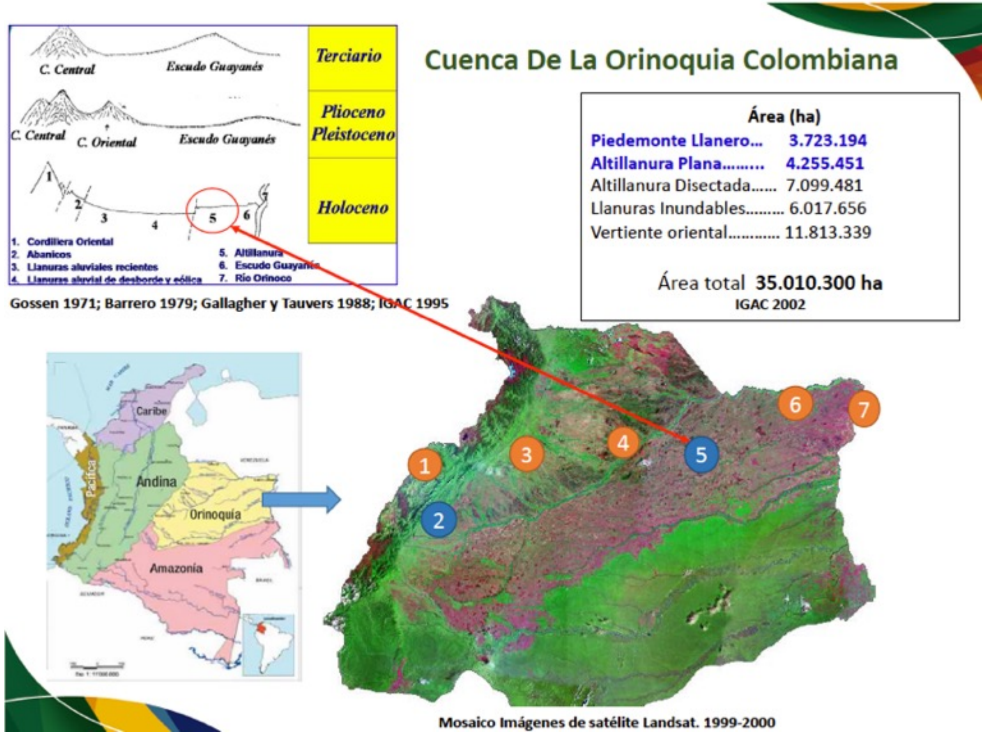
Cécile Grenier

# Why Al tolerance?

In tropical South America, 85% of the soils are acidic and the production of most of staple crops is negatively impacted by acid soils


Investigación en Sistemas Agropastoriles: Antecedentes y Estrategias

Figura 1. Areas de sabanas neotropicales y algunas regiones aledañas de Chaco (Bolivia y Paraguay).
FUENTE: Unidad de Manejo de Tierras, CIAT.


Cuenca De La Orinoquia Colombiana

1. Cordillera Oriental
2. Abanicos
3. Llanuras aluviales recientes
4. Llanuras aluvial de desborde y eólica
5. Altillanura
6. Escudo Guayanés
7. Río Orinoco

Gossen 1971; Barrero 1979; Gallagher y Tauvers 1988; IGAC 1995

| | Área (ha) |
|---|---|
| Piedemonte Llanero... | 3.723.194 |
| Altillanura Plana......... | 4.255.451 |
| Altillanura Disectada...... | 7.099.481 |
| Llanuras Inundables......... | 6.017.656 |
| Vertiente oriental............ | 11.813.339 |
| Área total | 35.010.300 ha |

IGAC 2002

Mosaico Imágenes de satélite Landsat. 1999-2000

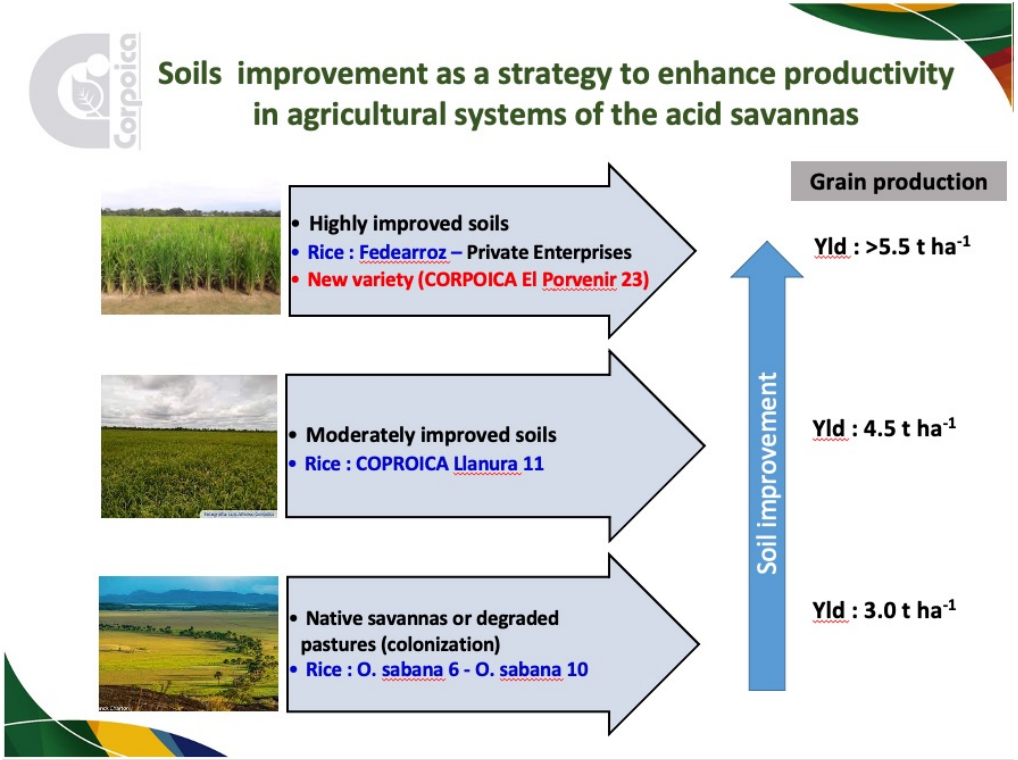
Aluminum (Al) is a metal naturally abundant in the soil, which, when pH values fall below 5.0, becomes highly phytotoxic by adopting a trivalent form ($Al^{3+}$)

# Why upland rice?

Upland rice (tropical japonica) is tolerant to acid soils an important component of rice-pasture cropping systems.

Rice-pasture cropping system, a productive option for the savannas

Since 1992 the Cirad-CIAT rice breeding program has worked on improving upland rice for the acid savannas of Colombia and LAC



**Soils improvement as a strategy to enhance productivity in agricultural systems of the acid savannas**

**Grain production**

- **Highly improved soils**
- **Rice : Fedearroz – Private Enterprises**
- **New variety (CORPOICA El Porvenir 23)**

**Yld : >5.5 t ha$^{-1}$**

- **Moderately improved soils**
- **Rice : COPROICA Llanura 11**

**Yld : 4.5 t ha$^{-1}$**

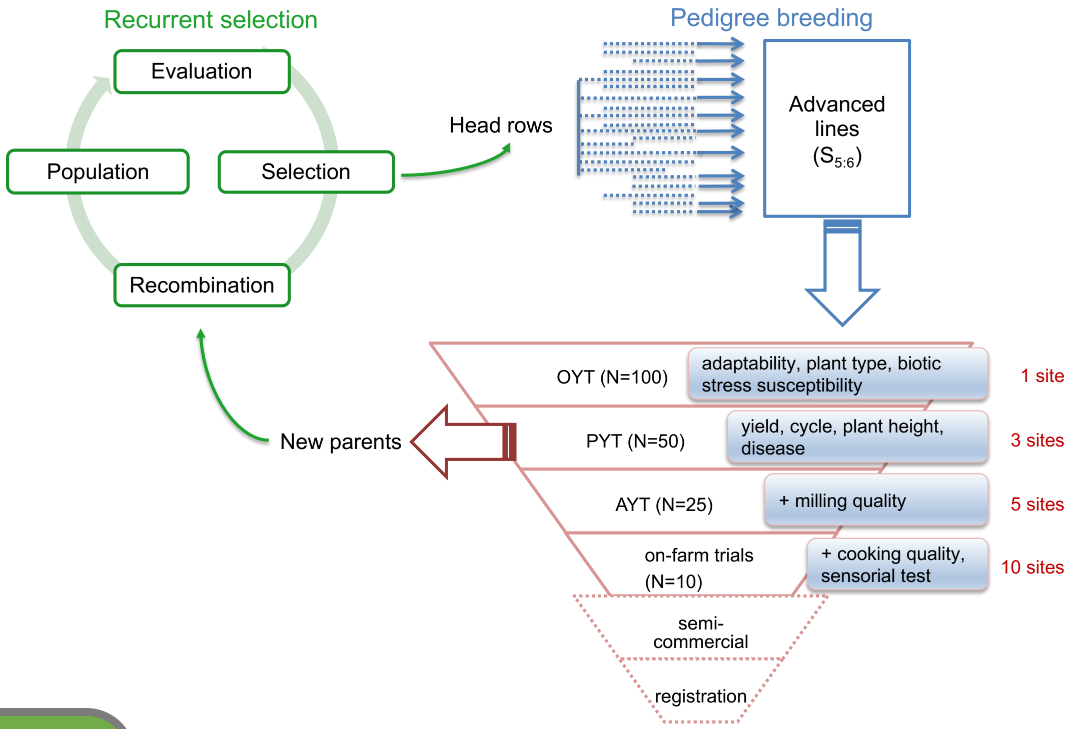
- **Native savannas or degraded pastures (colonization)**
- **Rice : O. sabana 6 - O. sabana 10**

**Yld : 3.0 t ha$^{-1}$**

**Soil improvement**

# The Cirad-CIAT rice breeding program

- Adaptation to upland conditions with acid soils (pH 4 - 5.2)
- Tolerance to water shortage
- Good yield
- Disease resistance (blast)



Variety benchmark

- Ozyzica sabana 6 (1991)
- Oryzica sabana 10 (1995)
- Corpoica llanura 11 (L-30) (2000-2011)

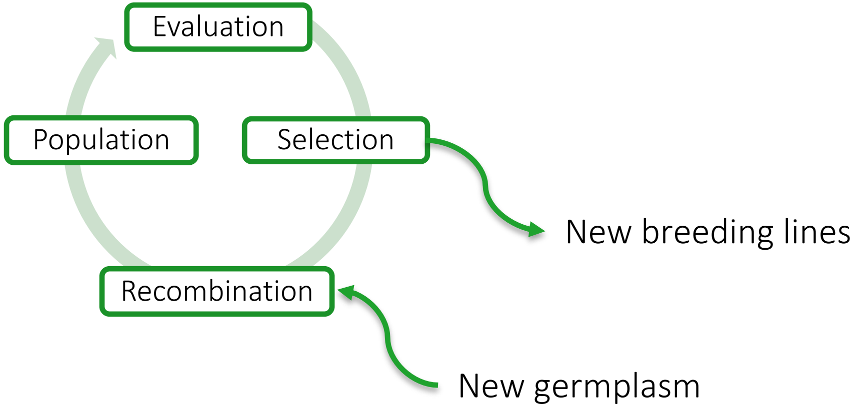
# Why Recurrent Selection on synthetic populations?

To combine many required trait of interest.

Synthetic populations can thus be improved with a particular focus for adaptation to the savanna soils.

The recurrent selection method of plant breeding carries the advantage to target the improvement of quantitative traits such as those usually controlling adaptative characters.

Recombination among best progenies to improve the population value



Evaluation

Population

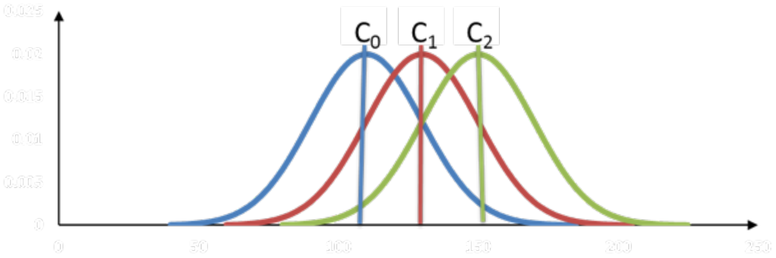
Selection

Recombination

New breeding lines

New germplasm

Mean value increased without compromising the genetic variability



# Why Genomic Prediction aided RS?

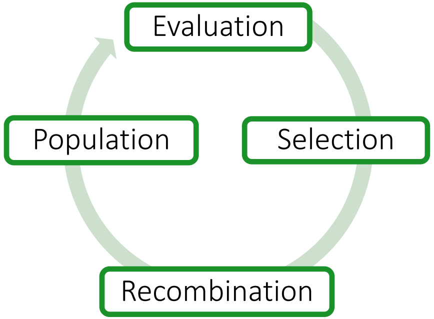
Accelerate the development of elite material

Operate more precise selection

Save cost in the breeding program

Increase genetic gain

$$\Delta G_{Yr} = \frac{i_{Ns} . r_{g\hat{g}} . {\sigma^2}_A}{L}$$

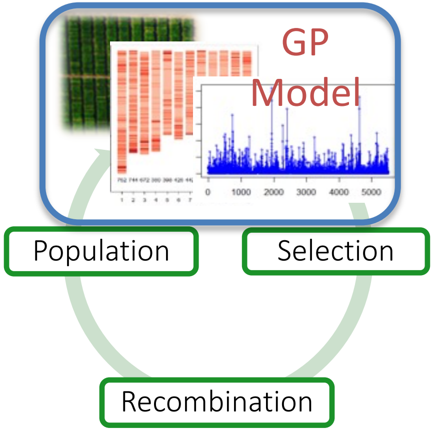


Evaluation

Population

Selection

Recombination



GP Model

Population

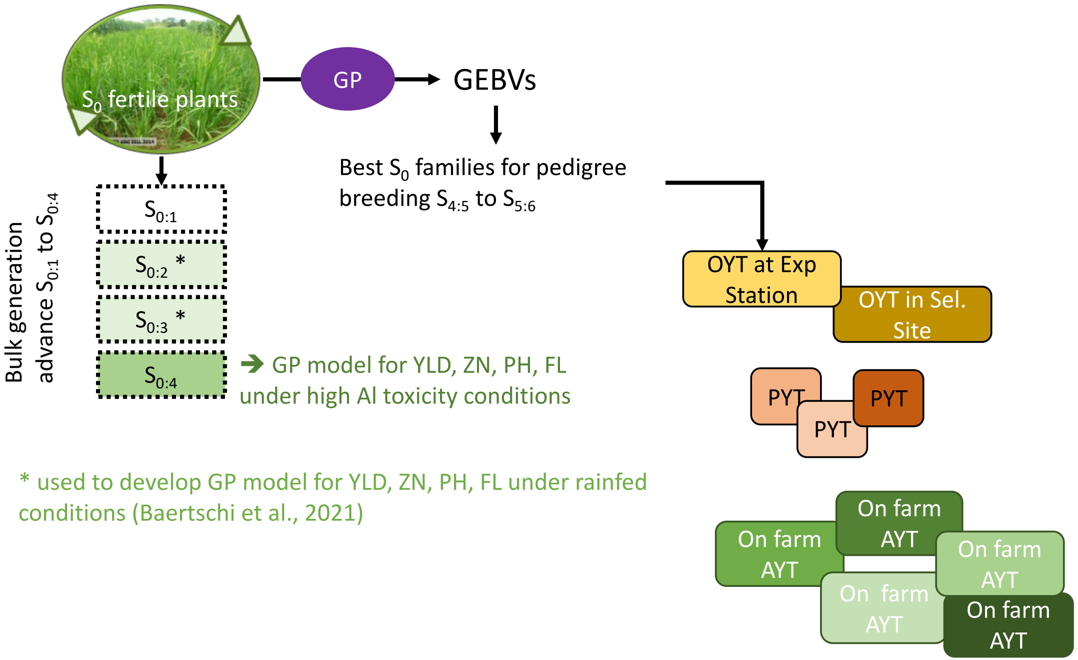
Selection

Recombination

# Our approach

Families derived from a synthetic population which is the base of the upland rice breeding program

Tropical japonica germplasm derived from the recombination of elites upland rice from Asia, Africa and Latin America and three decades of recombination



$S_0$ fertile plants

GP → GEBVs

Best $S_0$ families for pedigree breeding $S_{4:5}$ to $S_{5:6}$

Bulk generation advance $S_{0:1}$ to $S_{0:4}$

$S_{0:1}$

$S_{0:2}$ *

$S_{0:3}$ *

$S_{0:4}$

➔ GP model for YLD, ZN, PH, FL under high Al toxicity conditions

OYT at Exp Station

OYT in Sel. Site

PYT

PYT

PYT

On farm AYT

On farm AYT

On farm AYT

On farm AYT

On farm AYT

* used to develop GP model for YLD, ZN, PH, FL under rainfed conditions (Baertschi et al., 2021)

PCT27 generated from the mixture of for population (Baertschi et al., 2021)

Genotyping performed at $S_0$ generation

334 advanced generation families ($S_{0:4}$) randomly drawn from the PCT27

# Experimental set up

334 $S_{0:4}$ lines (grey colour)

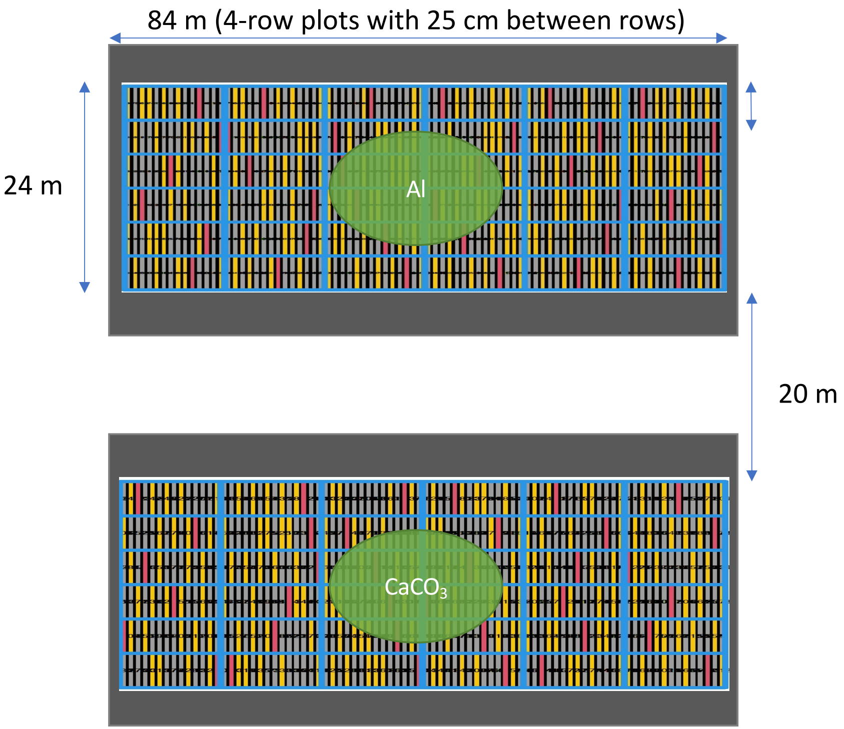
+ 25% repetitions = 84 $S_{0:4}$ lines (yellow colour)

+ 50 $S_{0:2}$ lines (for cross year comparisons & GxE-Genomic Prediction)

+ 6 checks repeated 6 times (red colour)

➔ 504 plots (6 ranges with 6 blocks of 14 plots)

4 3m long row
25 cm between rows
Sowing density: 1 g/m



# Data Analysis

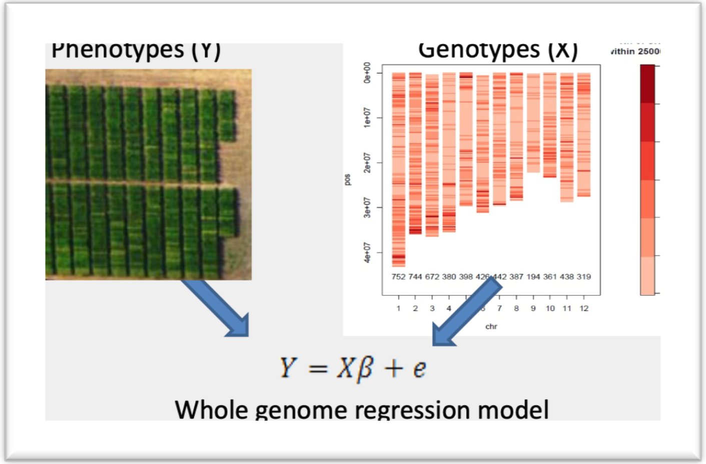
BLUEs estimation using mixed linear model

Calibration model (BGGE R-CRAN) by Gustavo de los Campos & Paulino Perez-Rodriguez)

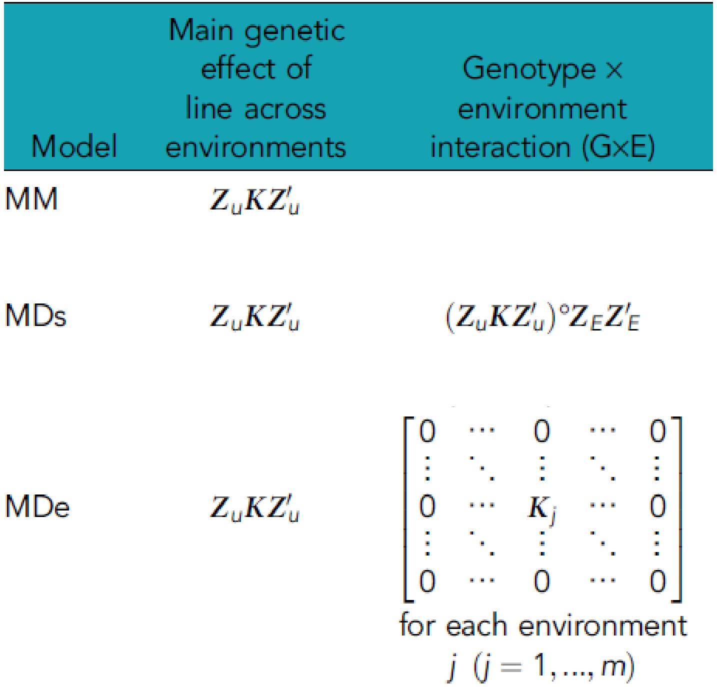
```
pTst <- 0.3
folds <- 100
nIter <- 70000;  burnIn <- 2000
```

➔ GEBVs

Predictive ability (PA) = r (GEBVs, BLUEs) from fold 1 to 100



# **BGGE** – the Bayesian linear mixed models (Granato et al., 2018)



| Model | Main genetic effect of line across environments | Genotype × environment interaction (G×E) |
|---|---|---|
| MM | $Z_uKZ'_u$ | |
| MDs | $Z_uKZ'_u$ | $(Z_uKZ'_u)^\circ Z_EZ'_E$ |
| MDe | $Z_uKZ'_u$ | $\begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & K_j & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$<br>for each environment $j$ $(j = 1, \ldots, m)$ |

SM : single-environment, main genotypic effect model -- fits the data for a single environment, and only one kernel is produced

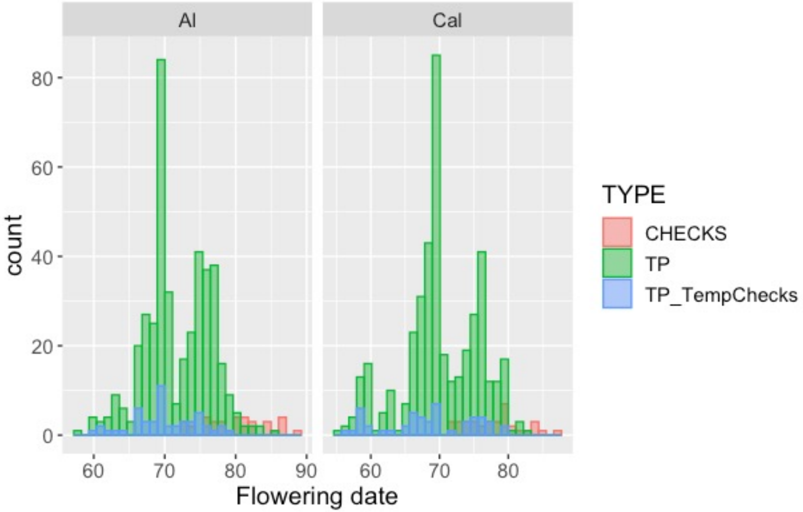
MM : multi-environment, main genotypic effect model -- considers the main random genetic effects across environments.

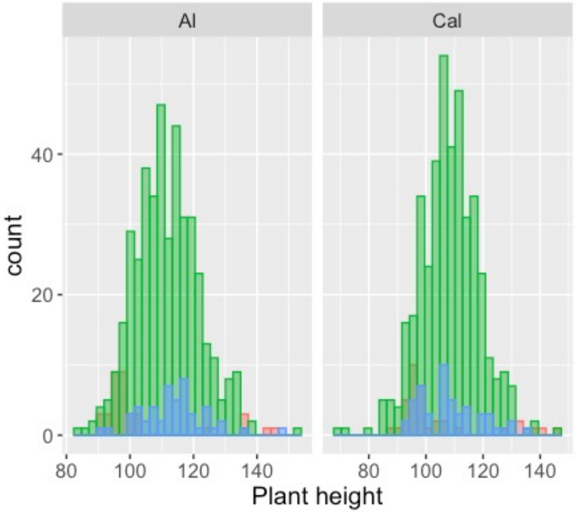
MDs : multi-environment, single variance GxE deviation mode -- extension of MM by adding the random interaction effect of environments with genotype information.

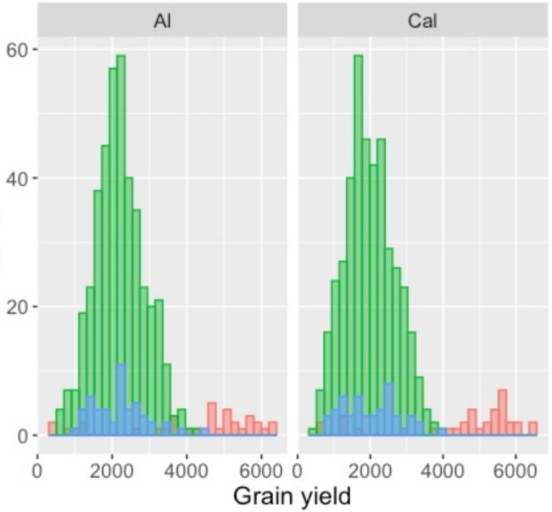
MDe : multi-environment, environment-specific variance GxE deviation model -- separates the genetic effects into the main genetic effects and the specific genetic effects (for each environment).

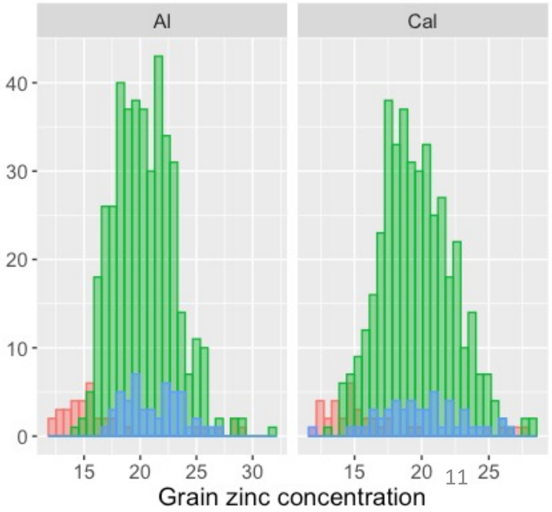
$Z_u$ is a known incidence matrix that relates the genotypes to the observations in the environments

# Trait distribution in the whole set







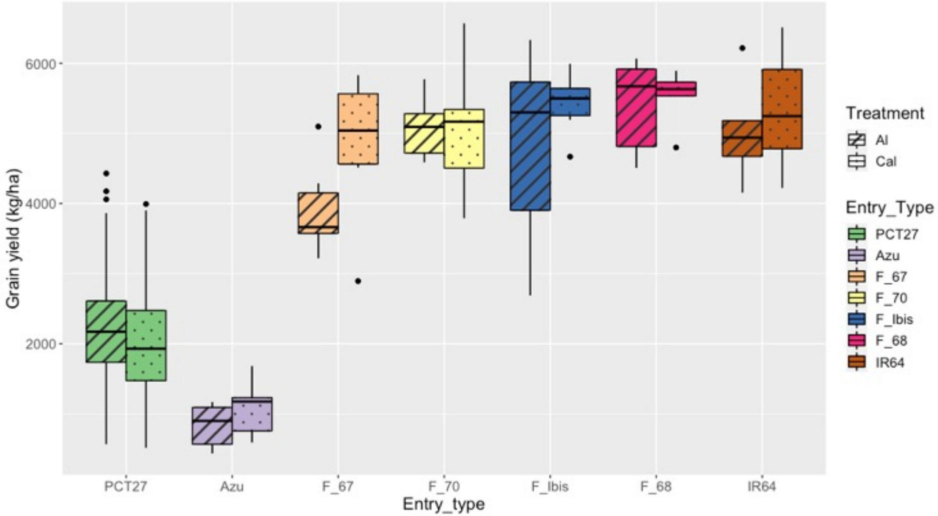


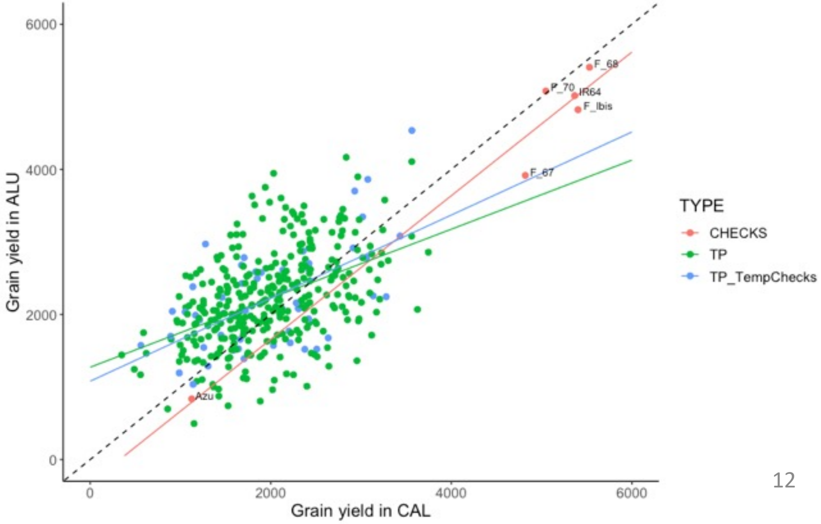
# Effect of treatment on grain yield

Greater general performance of indica rice under any treatment

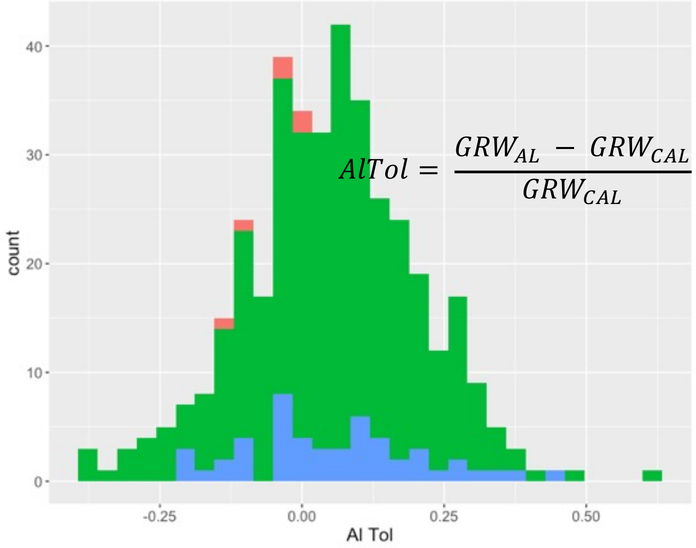
Only within the PCT27 grain yield significantly greater under Al than Cal

→ Upland rice families have lower yield than the indica rice but tolerance to aluminum seems to exist in this set of germplasm

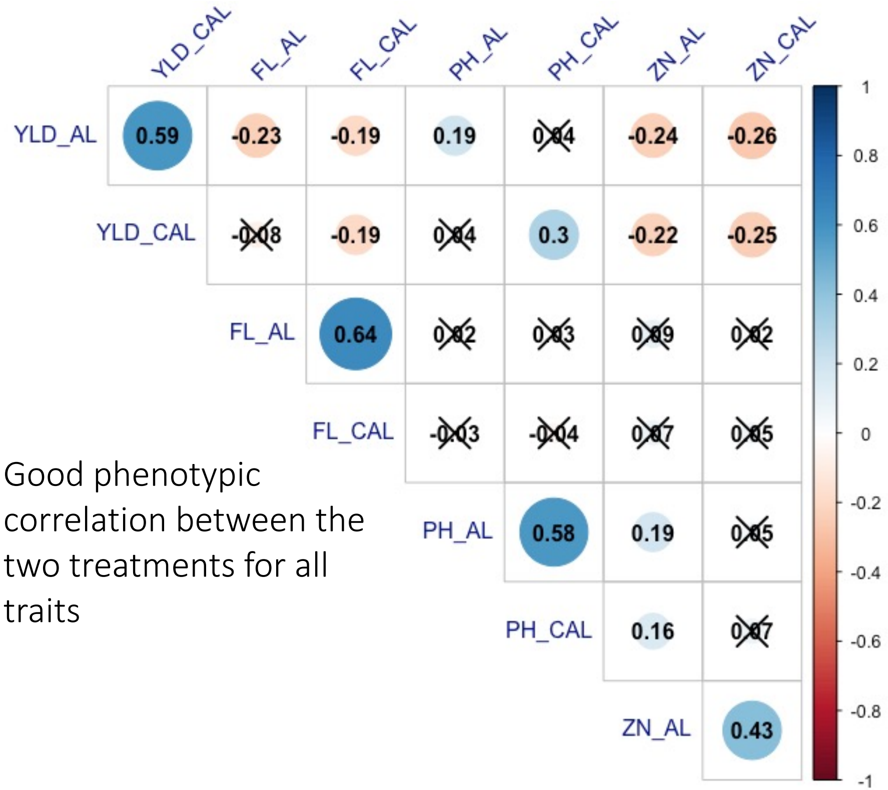




# High trait correlation and variability in AlTol index



$$AlTol = \frac{GRW_{AL} - GRW_{CAL}}{GRW_{CAL}}$$

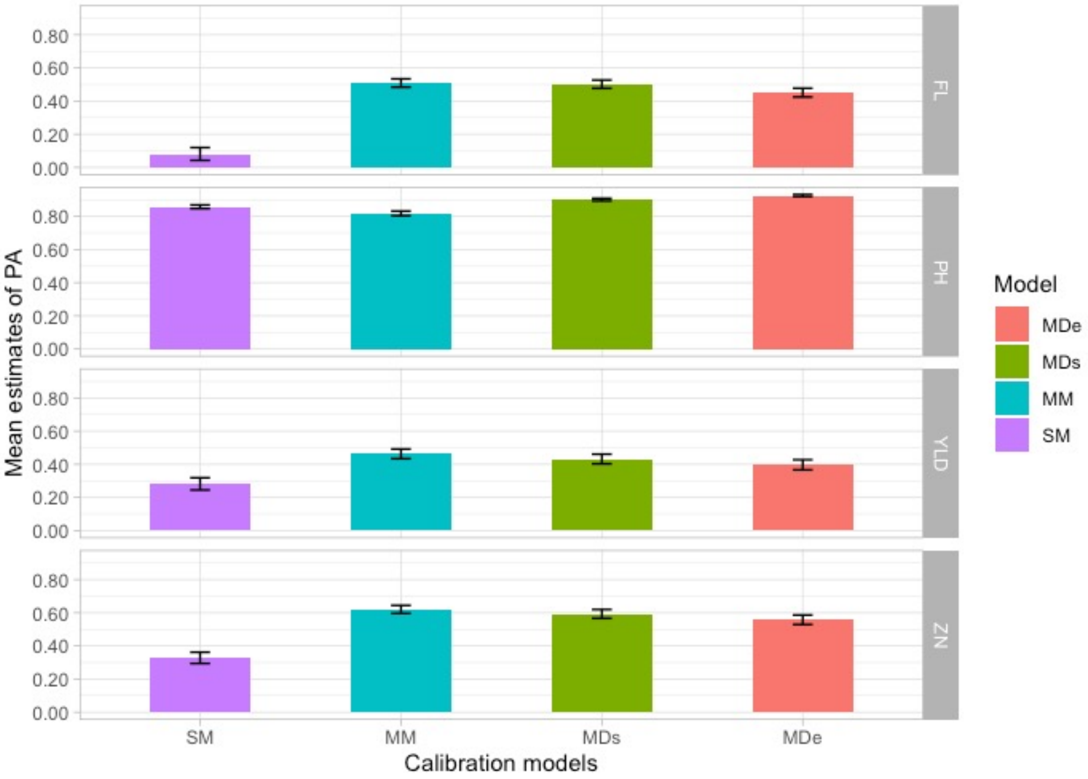


| | YLD_CAL | FL_AL | FL_CAL | PH_AL | PH_CAL | ZN_AL | ZN_CAL |
|---|---|---|---|---|---|---|---|
| YLD_AL | 0.59 | -0.23 | -0.19 | 0.19 | ~~0.04~~ | -0.24 | -0.26 |
| YLD_CAL | | ~~-0.08~~ | -0.19 | ~~0.04~~ | 0.3 | -0.22 | -0.25 |
| FL_AL | | | 0.64 | ~~0.02~~ | ~~0.03~~ | ~~0.09~~ | ~~0.02~~ |
| FL_CAL | | | | ~~-0.03~~ | ~~-0.04~~ | ~~0.07~~ | ~~0.05~~ |
| PH_AL | | | | | 0.58 | 0.19 | ~~0.05~~ |
| PH_CAL | | | | | | 0.16 | ~~0.07~~ |
| ZN_AL | | | | | | | 0.43 |

Good phenotypic correlation between the two treatments for all traits

Potential to exploit AlTol among the families

# GP on the 334 $S_{0:4}$ families for trait in Al condition



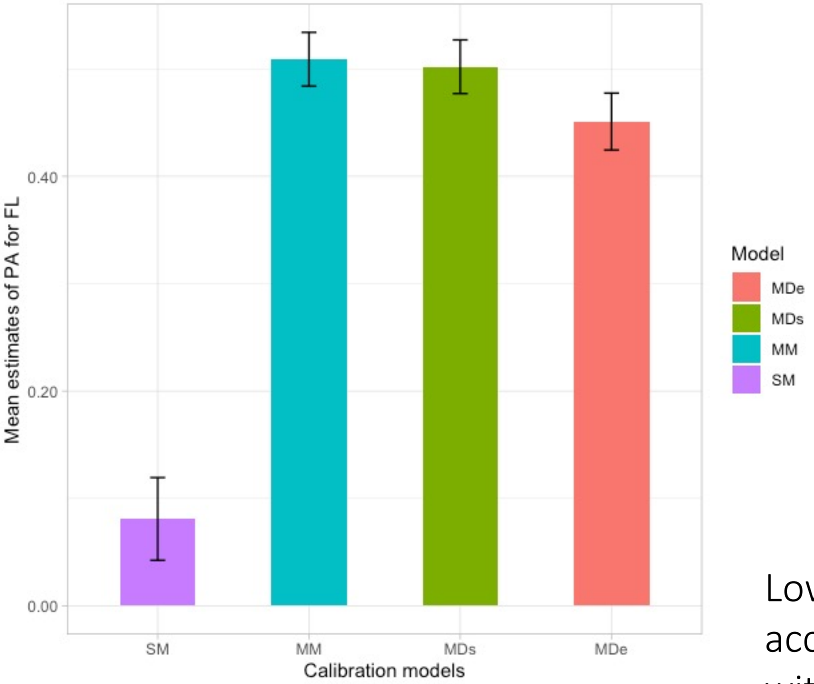
Predictive ability as the Pearson correlation between predicted values (GEBVs) and the estimated values (BLUEs)

Large variation in PA according to the traits

Considering GxE has different impact on the PA

Generally better PA when using the data from both sites to calibrate the model

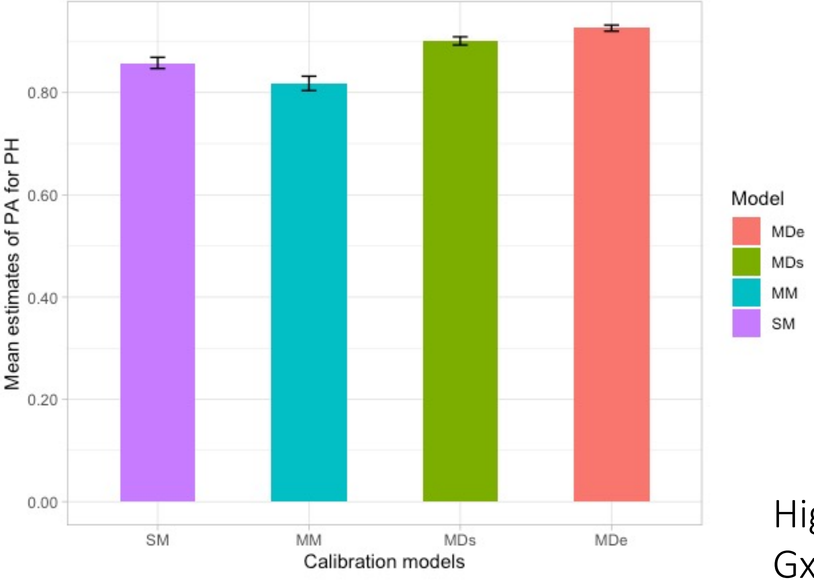
# High PA for FL considering multisite models

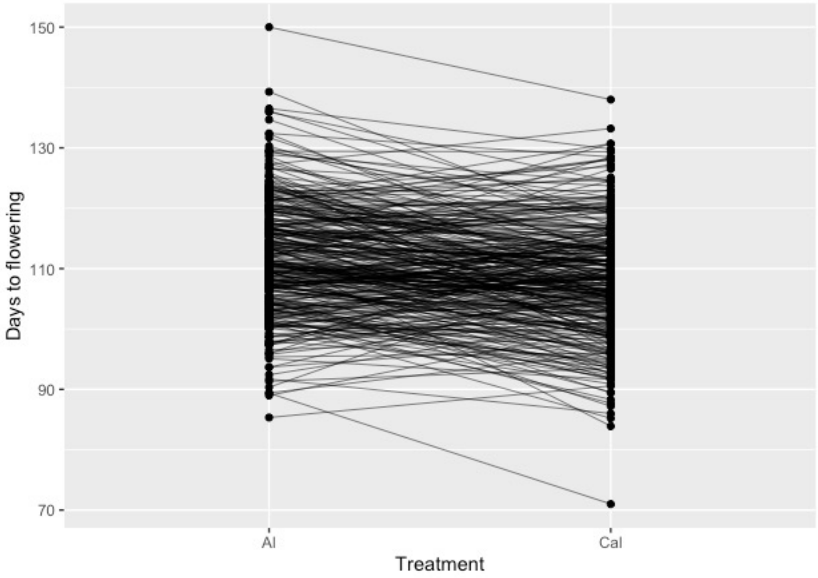


| FL | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| DESIGN | 383 | 18837.1 | 49.18 | 18.7591 | 0.051905 | . |
| SITE | 1 | 652.9 | 652.91 | 249.0299 | 0.003992 | ** |
| DESIGN:SITE | 383 | 4688.9 | 12.24 | 4.6695 | 0.192681 | |
| SITE:R | 10 | 476.2 | 47.62 | 18.1612 | 0.053289 | . |
| SITE:C | 156 | 1497.1 | 9.6 | 3.6603 | 0.238695 | |
| Residuals | 2 | 5.2 | 2.62 | | | |

Low PA with the single-environment, best when accounting for the GxE and genotypic as main effect or with single GxE variance effect

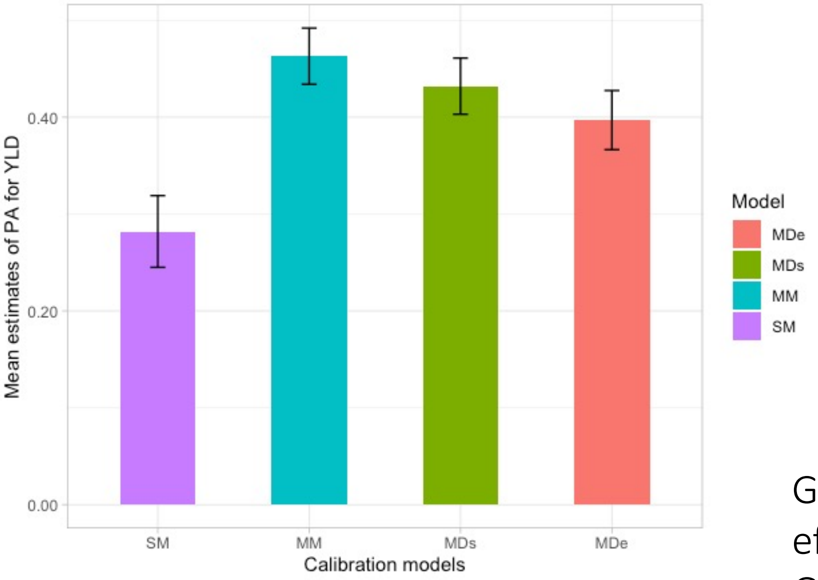
# High PA for PH with any models





High PA regardless of the model probably due to the low GxE contribution and high phenotypic correlation between sites

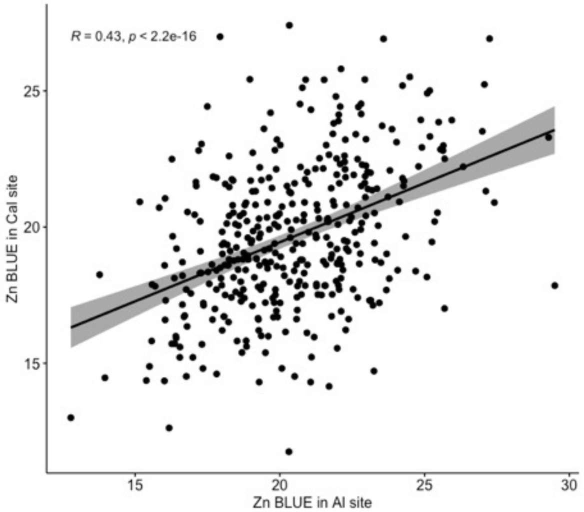
# Good PA for YLD with main genetic effect



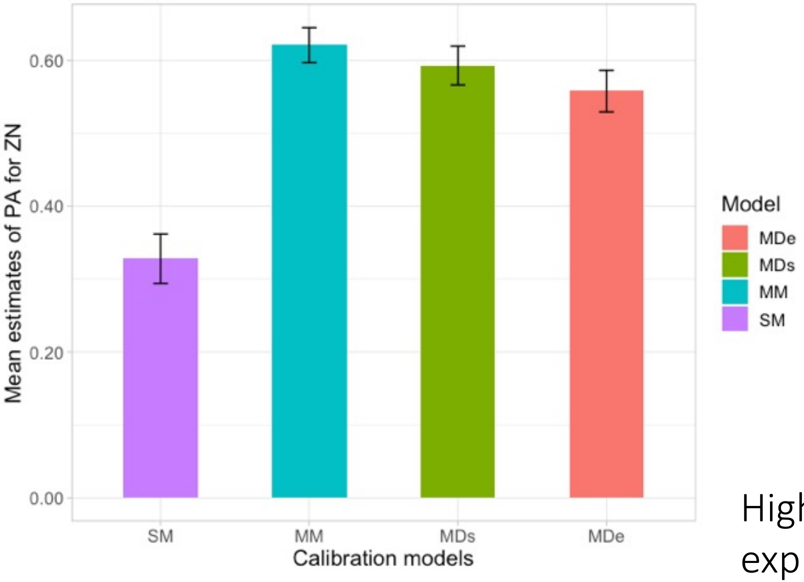
| | Groups | Variance | Std.Dev. | $\sigma^2_{ge}/\sigma^2_g$ |
|---|---|---|---|---|
| **FL** | DESIGN:SITE | 1.9938 | 1.412 | 0.13 |
| | DESIGN | 15.0419 | 3.8784 | |
| | Residual | 4.0519 | 2.0129 | |
| **PH** | DESIGN:SITE | 2.51E-12 | 1.58E-06 | 4E-14 |
| | DESIGN | 55.97 | 7.48 | |
| | Residual | 32.16 | 5.67 | |
| **YLD** | DESIGN:SITE | 0 | 0 | **0.00** |
| | DESIGN | 212260 | 460.72 | |
| | Residual | 149032 | 386.05 | |
| **ZN** | DESIGN:SITE | 1.3394 | 1.1573 | 1.22 |
| | DESIGN | 1.0947 | 1.0463 | |
| | Residual | 4.2442 | 2.0602 | |

Greater PA with multi-environment, main genotypic effect model probably because variance explained by GxE remains low compared to the genetic variance

# Good PA of ZN with multisite model



| Trait | $H^2$_Al | $H^2$_Cal |
|---|---|---|
| **FL** | 0.733 | 0.743 |
| **PH** | 0.568 | 0.396 |
| **GRW** | 0.566 | 0.513 |
| **YLD** | 0.598 | 0.486 |
| **ZN** | 0.647 | 0.634 |



High PA for ZN with GxE and main genotypic effect explained as high $H^2$

# Discussion

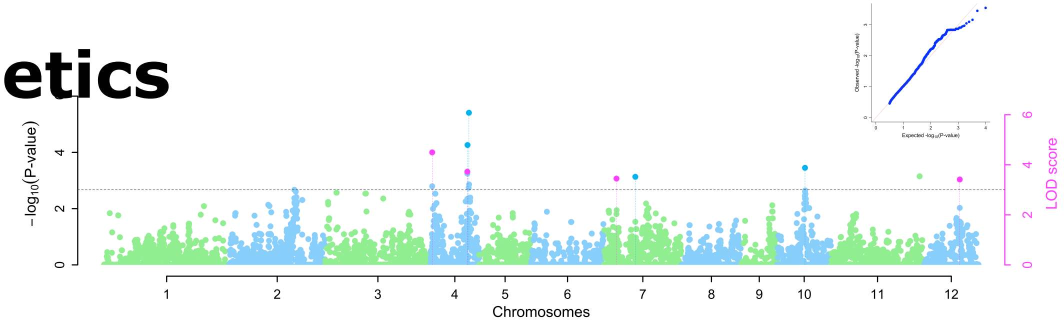
Although lower in productivity than the indica rice the potential of japonica is important to consider for acid soils.

Japonica rice can have a function for degraded soil restauration, crop rotation with pasture/leguminous.

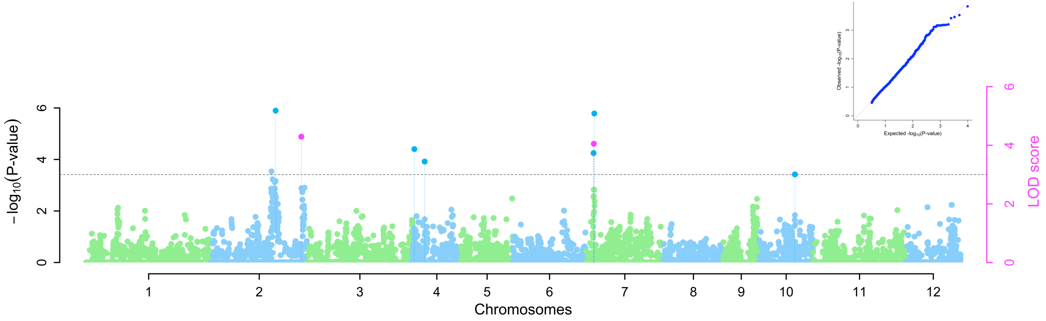
Japonica rice can bring interesting gene to high yielders indica rice.

Good perspective to increase genetic gain in the breeding program through genomic prediction aided recurrent selection.

# Next step -- Genetics



The GWAS combined with the genomic prediction could potential improve the fixation of certain key genes



Need of more markers (GBS pending) to detect more closely the responsible genetic control

Finer evaluation of GP and estimate on intra family variance to further the optimization of the breeding scheme to adaptation to acid soils

# Next steps -- Connections

The 50 elite lines evaluated for drought and Al tolerance are derived from populations which conformed PCT27. Validation of genetic control in the 50 lines can be done on this larger population

→ direct application of MAS with targeted regions / combination of GP and single marker MAS to combine several traits of interest

→ discovery of new mechanisms (under more quantitative control)

General poor performance of the tropical japonicas for lack of resistance to RHBV

→ should be taking advantage of the work on other P6 for general improvement of upland rice

Population available for further phenotyping of quantitative traits and that is of direct use for breeding

→ consider a subset for phenotyping climate resilient traits

# Acknowledgments

All the co-authors and even further than the group who have collaborated to this work

Jérôme Bartholome, José Omar Ospina, Mario Sandoval, Natalia Espinosa, Jairo Arcos, Thaura Ghneim

The funding CRP GRiSP, Rice and Harvest Plus for the development of the population and the genotyping of $S_0$, and Omicas for the field phenotyping and advanced generation sequencing

CGIAR
Research Program on Rice
Global Rice Science Partnership

HarvestPlus
Better Crops • Better Nutrition

CGIAR
RESEARCH PROGRAM ON Rice

ómicas

agAP INSTITUT
RÉPUBLIQUE FRANÇAISE
cirad
INRAE
L'INSTITUT agro
UNIVERSITÉ DE MONTPELLIER
Alliance Bioversity & CIAT
CGIAR