



Alliance



# Cassava Breeding Year-end Meeting

## Barranquilla 2022-Oct-17



The Alliance of Biodiversity International and the International Center for Tropical Agriculture (CIAT) is part of CGIAR – a global research partnership for a food-secure future.





Alliance



# Breeding tea year-end meeting

**Cassava Breeding Team**

Luis Fernando Delgado Muñoz

[luis.delgado@cgiar.org](mailto:luis.delgado@cgiar.org)

Barranquilla, October 17th, 2022

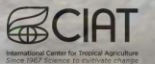
The Alliance of Bioversity International and the International Center for Tropical Agriculture (CIAT) is part of CGIAR – a global research partnership for a food-secure future.



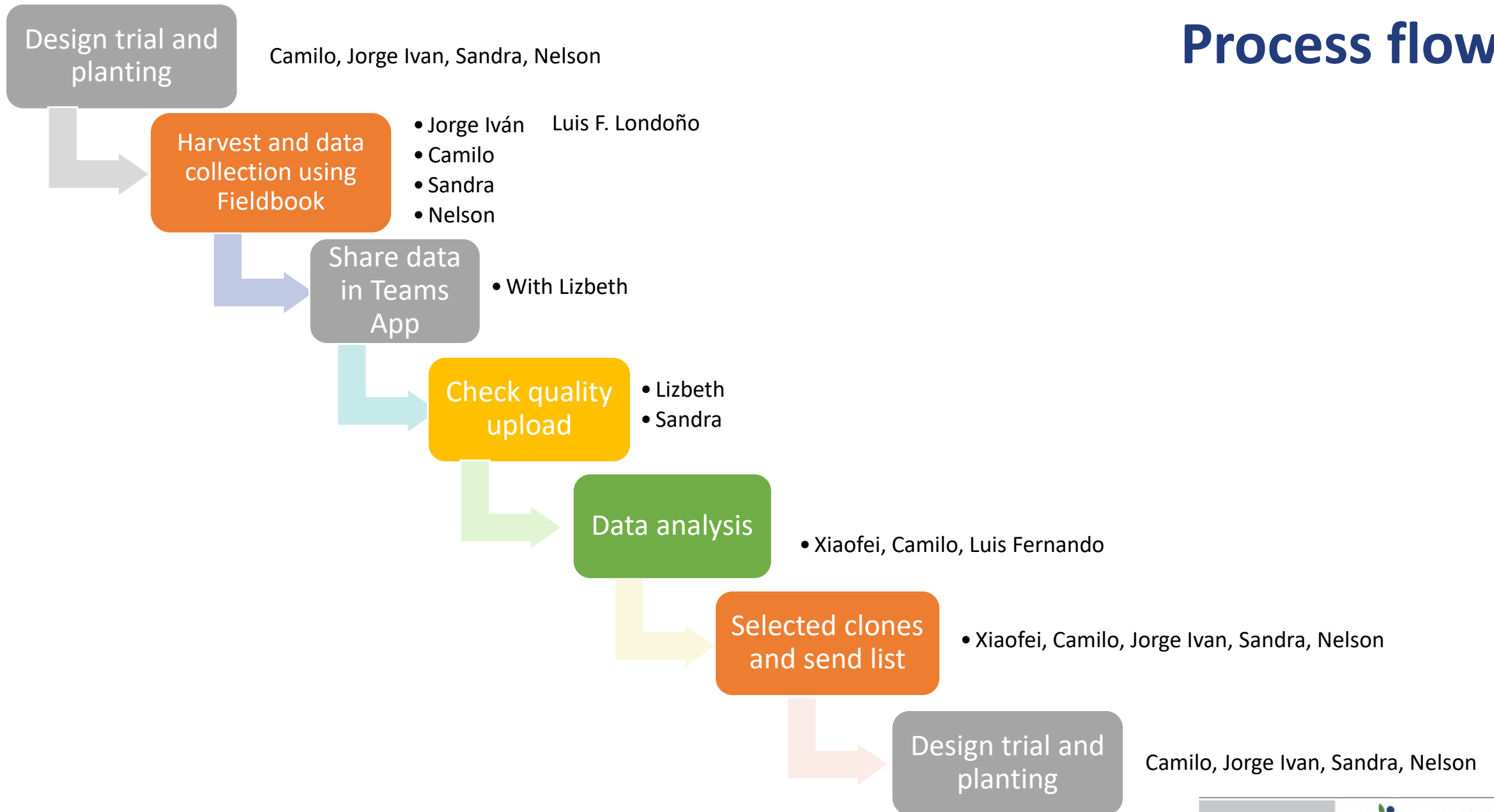
# Content

1. Personal introduction
2. Flow process
3. Data Analysis of Gene bank Clones
4. Genetic gain analysis (pipeline)
5. Trial design pipeline development
6. Challenges and improvement

Alliance



# Process flow



# 1. Data Analysis of Gene bank Clones

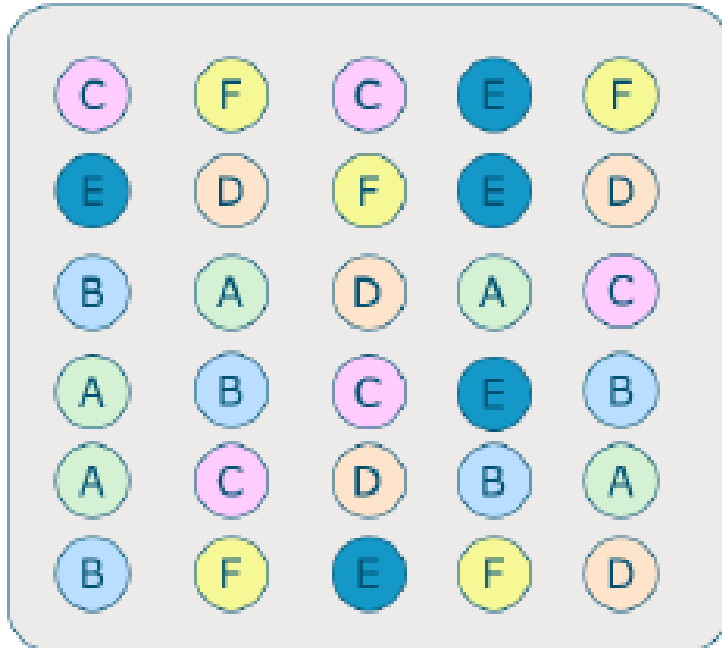
Alliance





# Historical Gene Bank trials (experimental design)

Completely randomized design (CRD)



- Treatments randomly allocated to experimental units
- No restrictions to randomization

# Historical Gene Bank clones

3244 genotypes (landraces)

121 trials

6 Agricultural zones

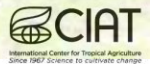
Subhumid lowland tropics  
Semiarid lowland tropics  
Mid-altitude tropics  
High altitude tropics  
Acid soil lowland tropical savannas  
Humid lowland tropics

34 locations (see map)

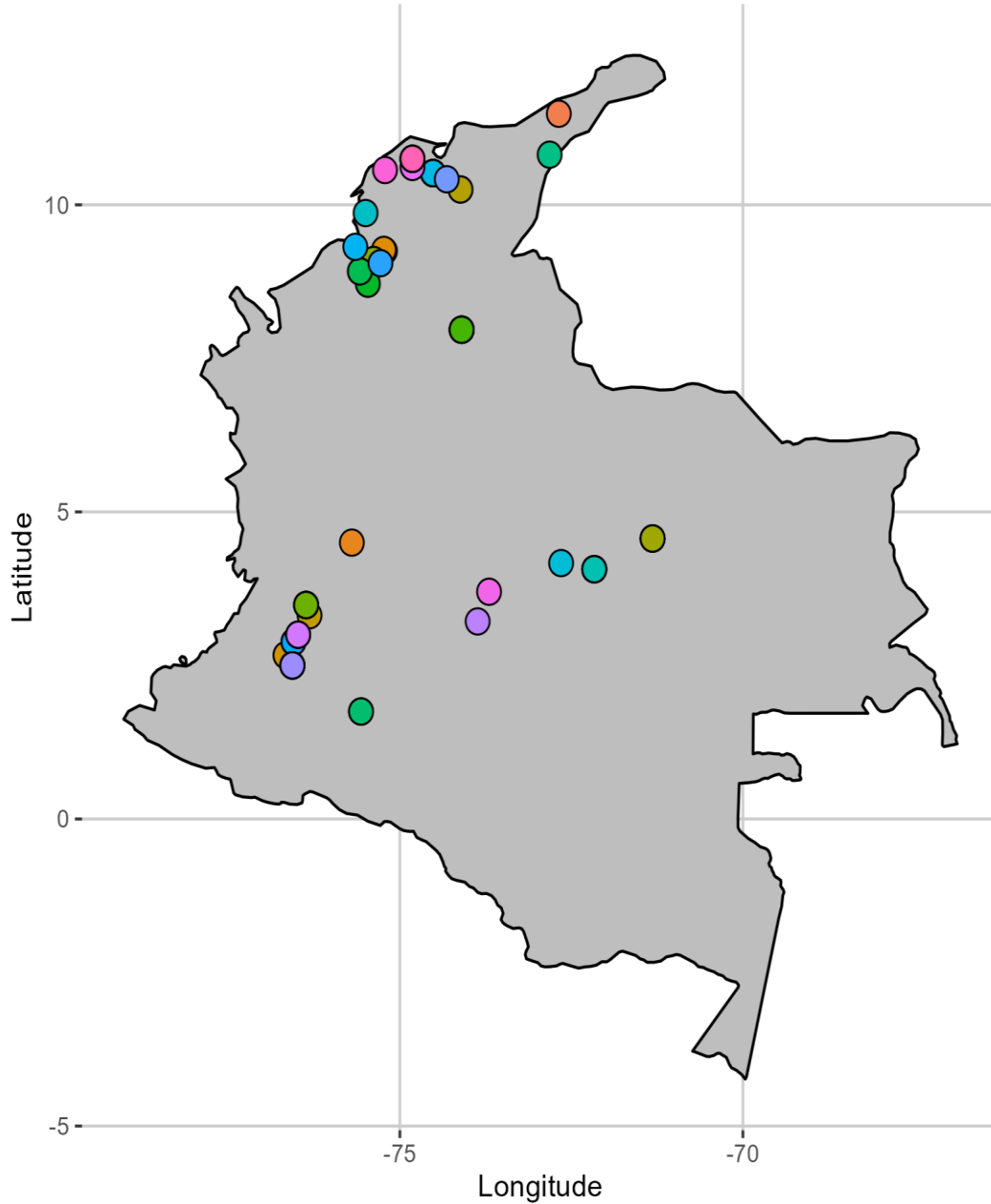
27 years

1980, 1981, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990,  
1991, 1995, 1996, 1997, 1998, 1999, 2003, 2004, 2006, 2007,  
2008, 2009, 2010, 2013, 2014, 2019 & 2020

Alliance



# Historical Gene Bank trials



## Location name

- |                                |                                    |
|--------------------------------|------------------------------------|
| ● Albania. Sucre, Colombia     | ● Laberinto. Sucre, Colombia       |
| ● Aremasain. Guajira, Colombia | ● Matazul. Meta, Colombia          |
| ● Armenia. Quindio, Colombia   | ● Media Luna. Magdalena, Colombia  |
| ● Betulia. Sucre, Colombia     | ● Momil. Cordoba, Colombia         |
| ● Cajibío. Cauca, Colombia     | ● Mondomo. Cauca, Colombia         |
| ● Candelaria. Valle, Colombia  | ● Montanita. Sucre, Colombia       |
| ● Caribia. Magdalena, Colombia | ● Pivijay. Magdalena, Colombia     |
| ● Carimagua. Meta, Colombia    | ● Popayan. Cauca, Colombia         |
| ● Carranzo. Cordoba, Colombia  | ● Puerto Gaitan. Meta, Colombia    |
| ● CIAT. Valle, Colombia        | ● S. de Quilichao. Cauca, Colombia |
| ● El Carmen. Bolivar, Colombia | ● Sabanalarga. Atlantico, Colombia |
| ● El Olivo. Cordoba, Colombia  | ● San Martin. Meta, Colombia       |
| ● El Salado. Cordoba, Colombia | ● Santa Cruz. Atlantico, Colombia  |
| ● Florencia. Caqueta, Colombia | ● Santo Tomas                      |
| ● Fonseca. Guajira, Colombia   | ● Santo Tomas. Atlantico, Colombia |
| ● La Libertad                  | ● Unknown                          |
| ● La Libertad. Meta, Colombia  | ● unknown2                         |

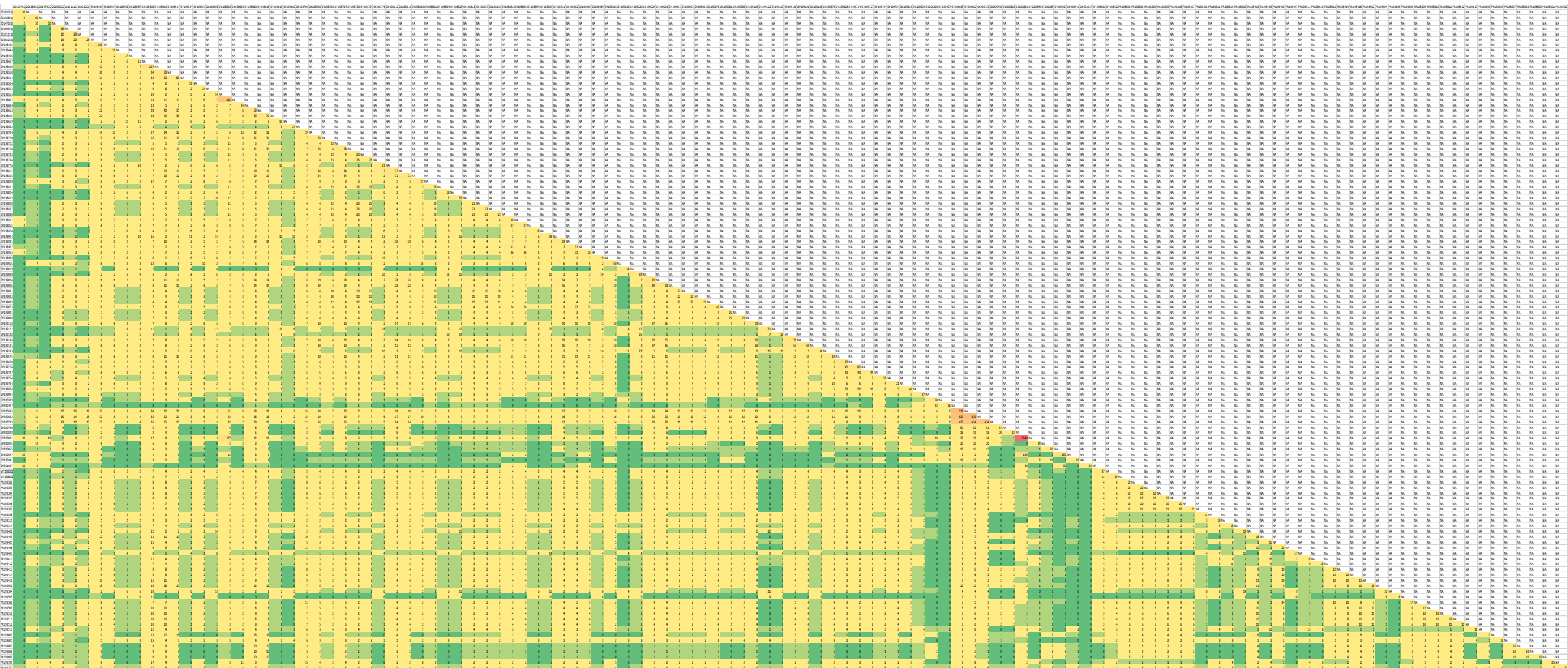
Alliance







# Shared clones between trials



- Yellow color indicates more than 1 shared clone
- Green color indicates less than 2 shared clone

Alliance

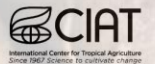




# Next steps

1. Finish curing the meta information from historical Cassava Base
2. Calculate the plot size to make the double check to yield trait.
3. Dig into the Oracle data base (extract the trials with gene bank clones).
4. Dig into the Kawano's data (extract the trials with gene bank clones).
5. Merge all information from tree dataframes.
6. Calculate broad heritability
7. Calculate BLUPs and BLUEs
8. Make the GxE analysis data.

Alliance



# Genetic gain LA population

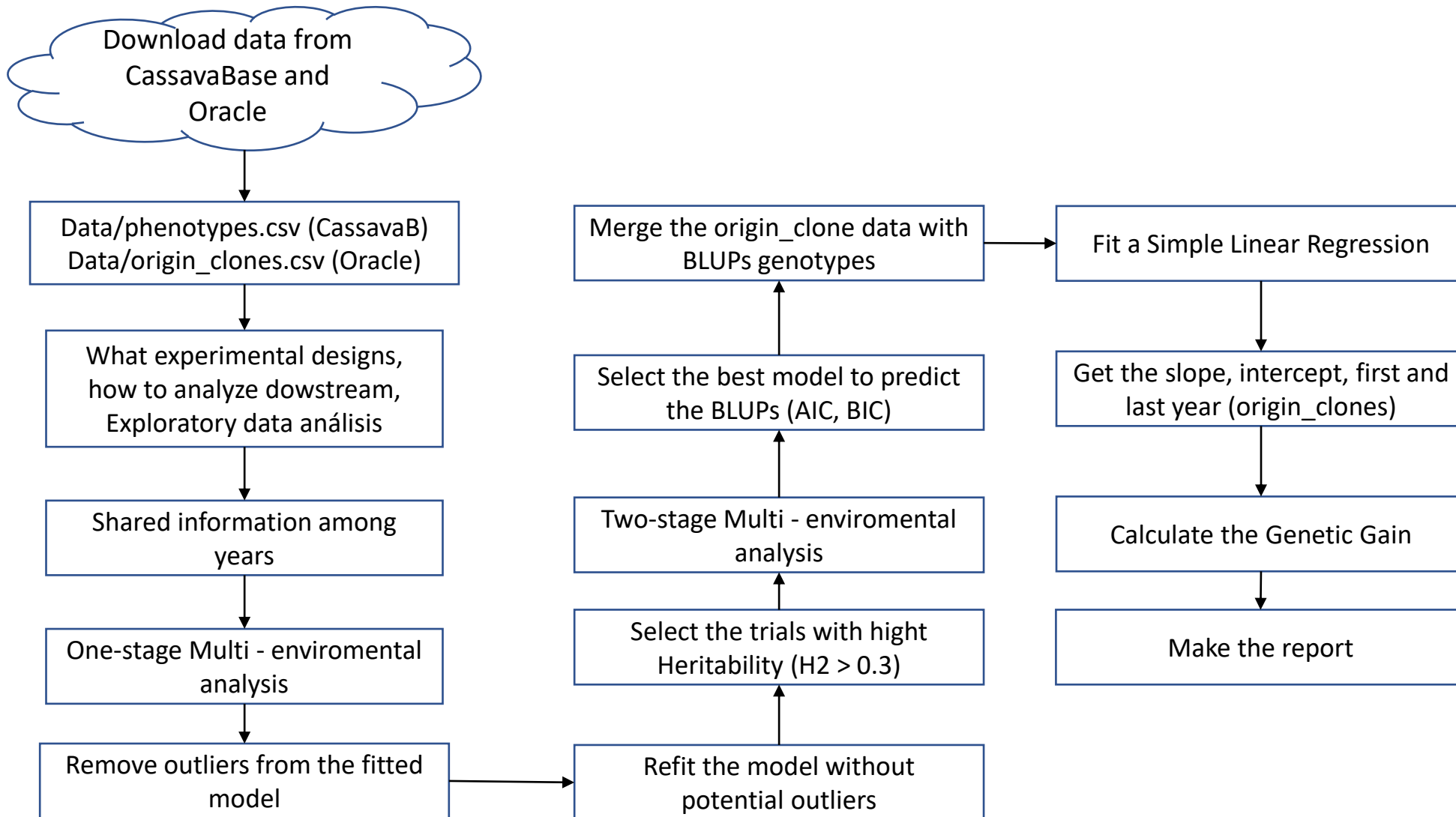
## Yield (t/ha)

Alliance





## 2. Genetic gain analysis (pipeline, Process Map)



# Experimental designs, how to analyze downstream

| trial_name       | n_gen | n_reps | n_cols | n_rows | n_total | n_missing | n_percent | col_row | rcbd | design  |
|------------------|-------|--------|--------|--------|---------|-----------|-----------|---------|------|---------|
| 201983LAEPR_sinc | 32    | 3      | 1      | 1      | 96      | 0         | 0         | FALSE   | TRUE | rcbd    |
| 202113LAEAR_repe | 52    | 3      | 12     | 13     | 156     | 0         | 0         | TRUE    | TRUE | row_col |
| 202114LAEAR_momi | 52    | 3      | 12     | 13     | 156     | 0         | 0         | TRUE    | TRUE | row_col |
| 202123LAEPR_polo | 94    | 3      | 18     | 16     | 288     | 0         | 0         | TRUE    | TRUE | row_col |
| 202124LAEPR_momi | 93    | 3      | 18     | 16     | 288     | 0         | 0         | TRUE    | TRUE | row_col |
| 202125LAEPR_repe | 83    | 3      | 15     | 17     | 255     | 0         | 0         | TRUE    | TRUE | row_col |
| 202119LAEAR_polo | 51    | 3      | 12     | 13     | 156     | 4         | 0.026     | TRUE    | TRUE | row_col |
| 201551LAEAR_saha | 33    | 3      | 1      | 1      | 102     | 5         | 0.049     | FALSE   | TRUE | rcbd    |
| 202066LAEPR_momi | 192   | 3      | 26     | 24     | 612     | 30        | 0.049     | TRUE    | TRUE | row_col |
| 202061LAEPR_repe | 179   | 3      | 32     | 18     | 576     | 36        | 0.063     | TRUE    | TRUE | row_col |
| 201843LAEPR_pita | 82    | 3      | 1      | 1      | 246     | 16        | 0.065     | FALSE   | TRUE | rcbd    |
| 201738LAEPR_cara | 111   | 3      | 1      | 1      | 333     | 22        | 0.066     | FALSE   | TRUE | rcbd    |
| 202058LAEPR_pivi | 190   | 3      | 34     | 18     | 612     | 55        | 0.090     | TRUE    | TRUE | row_col |
| 201955LAEPR_mlun | 16    | 3      | 1      | 1      | 48      | 5         | 0.104     | FALSE   | TRUE | rcbd    |
| 201857LAEAR_sinc | 67    | 3      | 1      | 1      | 201     | 27        | 0.134     | FALSE   | TRUE | rcbd    |
| 201572LAEAR_sinc | 42    | 3      | 1      | 1      | 126     | 22        | 0.175     | FALSE   | TRUE | rcbd    |
| 201853LAEAR_cara | 24    | 3      | 1      | 1      | 72      | 13        | 0.181     | FALSE   | TRUE | rcbd    |
| 201417LAEPR_mala | 72    | 3      | 1      | 1      | 216     | 63        | 0.292     | FALSE   | TRUE | rcbd    |
| 201419LAEPR_mala | 33    | 3      | 1      | 1      | 99      | 30        | 0.303     | FALSE   | TRUE | rcbd    |
| 201418LAEPR_mala | 75    | 3      | 1      | 1      | 225     | 82        | 0.364     | FALSE   | TRUE | rcbd    |
| 201420LAEPR_mala | 21    | 3      | 1      | 1      | 63      | 27        | 0.429     | FALSE   | TRUE | rcbd    |
| 201321LAEPR_pvar | 124   | 4      | 1      | 1      | 500     | 255       | 0.510     | FALSE   | TRUE | rcbd    |
| 201856LAEPR_sinc | 78    | 3      | 1      | 1      | 234     | 120       | 0.513     | FALSE   | TRUE | rcbd    |
| 201427LAEPR_mala | 21    | 3      | 1      | 1      | 63      | 63        | 1.000     | FALSE   | TRUE | rcbd    |
| 201944LAEPR_palm | 20    | 3      | 1      | 1      | 60      | 60        | 1.000     | FALSE   | TRUE | rcbd    |
| 201951LAEPR_mlun | 89    | 3      | 1      | 1      | 267     | 267       | 1.000     | FALSE   | TRUE | rcbd    |
| 201952LAEPR_mlun | 20    | 3      | 1      | 1      | 60      | 60        | 1.000     | FALSE   | TRUE | rcbd    |
| 201961LAEPR_polo | 89    | 3      | 1      | 1      | 267     | 267       | 1.000     | FALSE   | TRUE | rcbd    |
| 201969LAEPR_polo | 20    | 3      | 1      | 1      | 60      | 60        | 1.000     | FALSE   | TRUE | rcbd    |
| 201978LAEPR_sinc | 89    | 3      | 1      | 1      | 267     | 267       | 1.000     | FALSE   | TRUE | rcbd    |
| 201984LAEPR_sinc | 20    | 3      | 1      | 1      | 60      | 60        | 1.000     | FALSE   | TRUE | rcbd    |
| 201991LAEPR_sinc | 89    | 3      | 1      | 1      | 267     | 267       | 1.000     | FALSE   | TRUE | rcbd    |



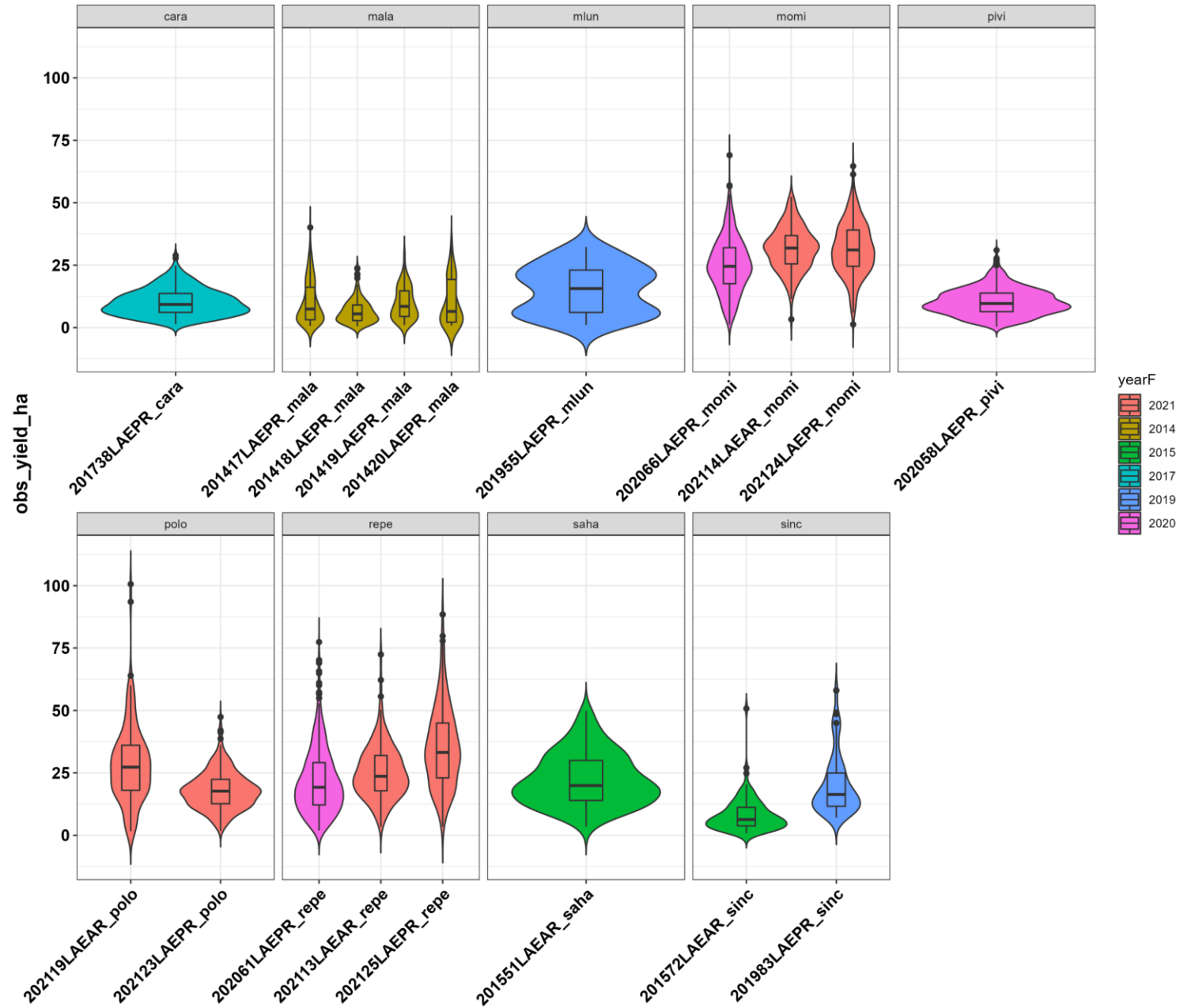
# Waxy population genetic gain

6 years

9 locations

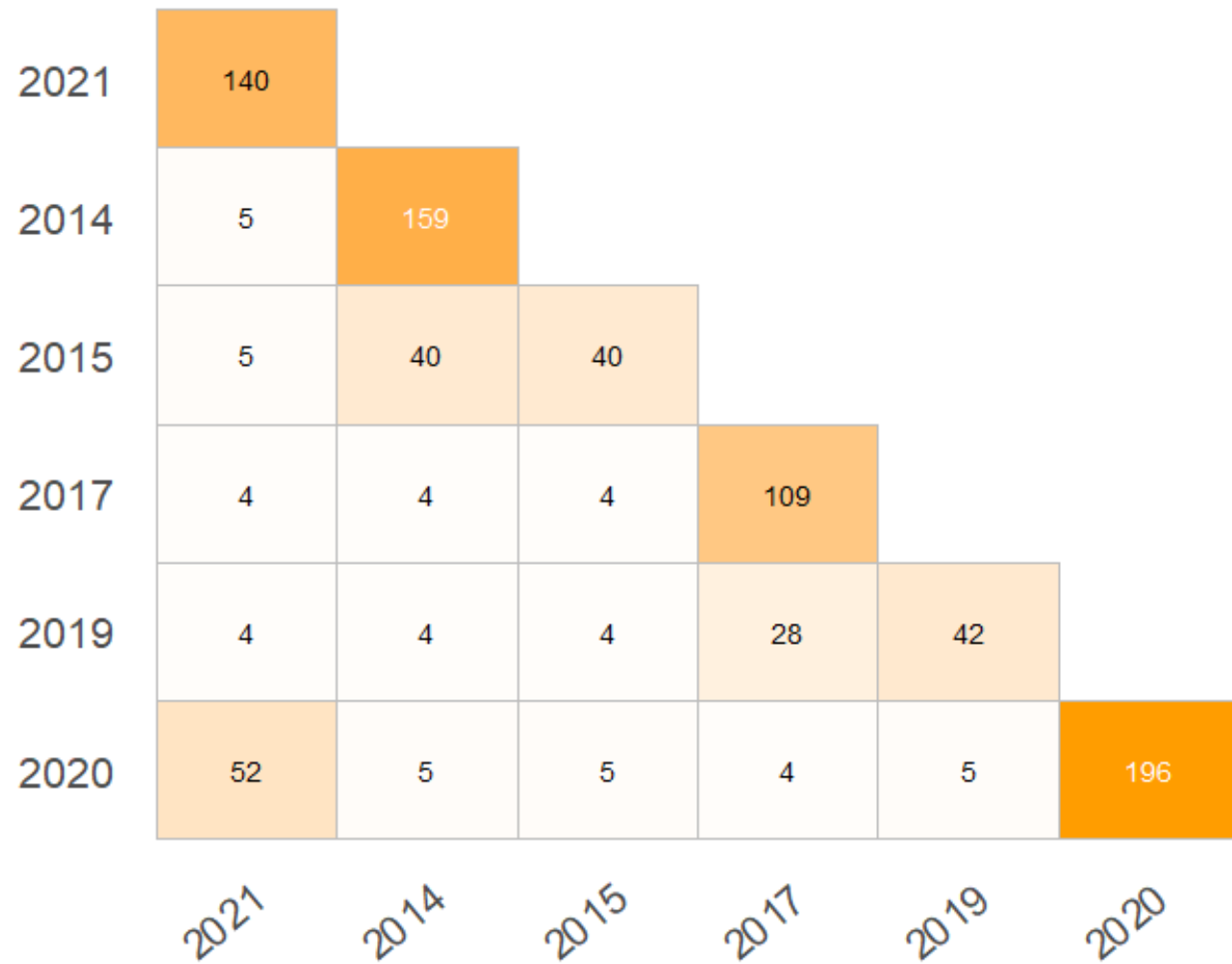
588 waxy clones

18 trials



## Shared information between years

## Shared information



Alliance



# Models outliers

| Experiment       | obs_yield_ha | rep_number | accession_name | Classify |
|------------------|--------------|------------|----------------|----------|
| 202119LAEAR_polo | 100.625      | 1          | SM4698-3       | Outlier  |
| 202125LAEPR_repe | 88.46153846  | 2          | CM4919-1       | Outlier  |
| 202061LAEPR_repe | 77.40384615  | 3          | BELLOTI        | Outlier  |
| 202113LAEAR_repe | 72.43589744  | 1          | SM4754-18      | Outlier  |
| 202066LAEPR_momi | 69.04761905  | 3          | SM4756-19      | Outlier  |
| 202113LAEAR_repe | 62.17948718  | 3          | SM4715-20      | Outlier  |
| 201572LAEAR_sinc | 50.78125     | 3          | GM5542-15      | Outlier  |
| 202123LAEPR_polo | 47.39583333  | 1          | GM11138-3      | Outlier  |
| 202123LAEPR_polo | 41.92708333  | 2          | GM11103-49     | Outlier  |
| 201417LAEPR_mala | 40.10416667  | 2          | GM5393-6       | Outlier  |
| 201738LAEPR_cara | 27.83203125  | 1          | GM8745-1       | Outlier  |
| 201738LAEPR_cara | 18.75        | 1          | SM4080-6       | Outlier  |
| 202061LAEPR_repe | 11.85897436  | 1          | BELLOTI        | Outlier  |
| 202066LAEPR_momi | 5.677655678  | 2          | TAI8           | Outlier  |

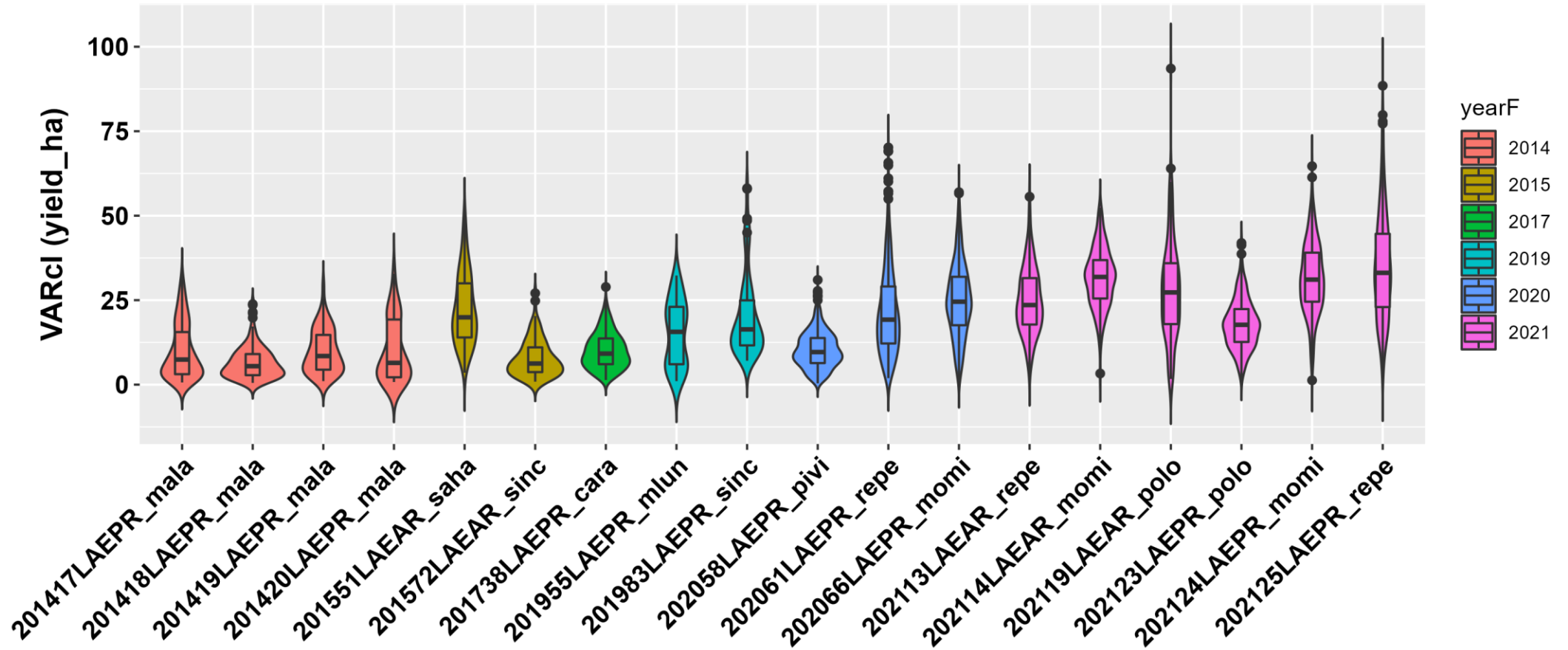
Alliance



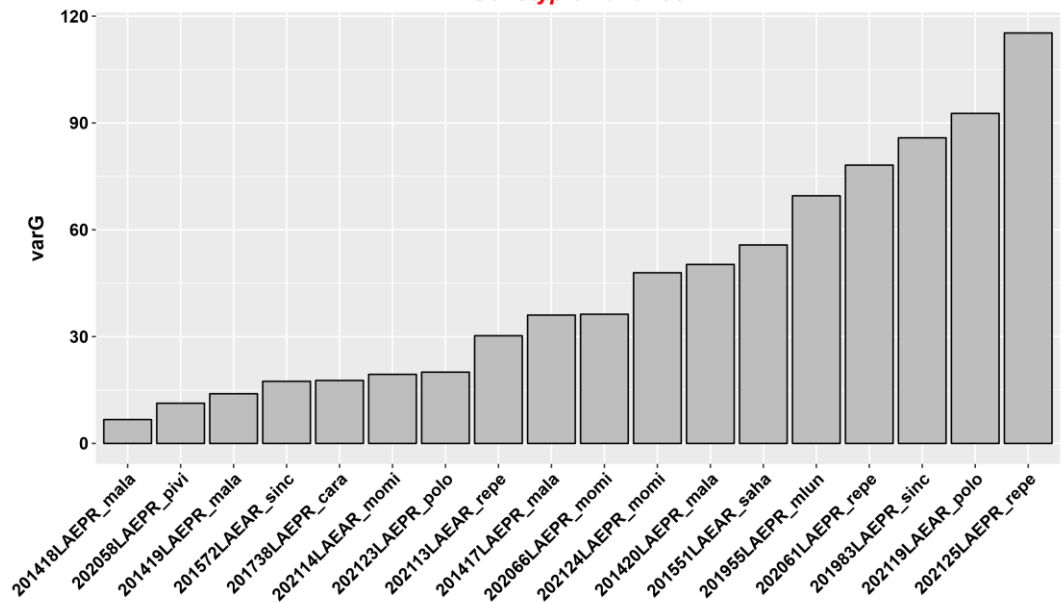
## Mix model results by trial

| Experiment       | Trait        | varG    | varE    | h2    | outliers | design |
|------------------|--------------|---------|---------|-------|----------|--------|
| 201955LAEPR_mlun | obs_yield_ha | 69.555  | 25.810  | 0.874 | 0        | rcbd   |
| 201738LAEPR_cara | obs_yield_ha | 17.688  | 10.155  | 0.826 | 0        | rcbd   |
| 201983LAEPR_sinc | obs_yield_ha | 85.827  | 59.121  | 0.813 | 0        | rcbd   |
| 201572LAEAR_sinc | obs_yield_ha | 17.447  | 10.584  | 0.799 | 0        | rcbd   |
| 202061LAEPR_repe | obs_yield_ha | 78.160  | 62.577  | 0.77  | 0        | spats  |
| 201420LAEPR_mala | obs_yield_ha | 50.260  | 27.797  | 0.767 | 0        | rcbd   |
| 201551LAEAR_saha | obs_yield_ha | 55.737  | 53.803  | 0.752 | 0        | rcbd   |
| 202124LAEPR_momi | obs_yield_ha | 47.934  | 46.232  | 0.75  | 0        | spats  |
| 201417LAEPR_mala | obs_yield_ha | 36.045  | 28.431  | 0.736 | 0        | rcbd   |
| 202058LAEPR_pivi | obs_yield_ha | 11.291  | 11.198  | 0.73  | 0        | spats  |
| 202125LAEPR_repe | obs_yield_ha | 115.273 | 117.181 | 0.71  | 0        | spats  |
| 202119LAEAR_polo | obs_yield_ha | 92.674  | 115.709 | 0.69  | 0        | spats  |
| 202113LAEAR_repe | obs_yield_ha | 30.243  | 37.406  | 0.68  | 0        | spats  |
| 202123LAEPR_polo | obs_yield_ha | 20.020  | 26.971  | 0.67  | 0        | spats  |
| 202066LAEPR_momi | obs_yield_ha | 36.279  | 65.589  | 0.61  | 0        | spats  |
| 201419LAEPR_mala | obs_yield_ha | 13.952  | 27.412  | 0.531 | 0        | rcbd   |
| 202114LAEAR_momi | obs_yield_ha | 19.395  | 48.042  | 0.53  | 0        | spats  |
| 201418LAEPR_mala | obs_yield_ha | 6.705   | 15.589  | 0.478 | 0        | rcbd   |

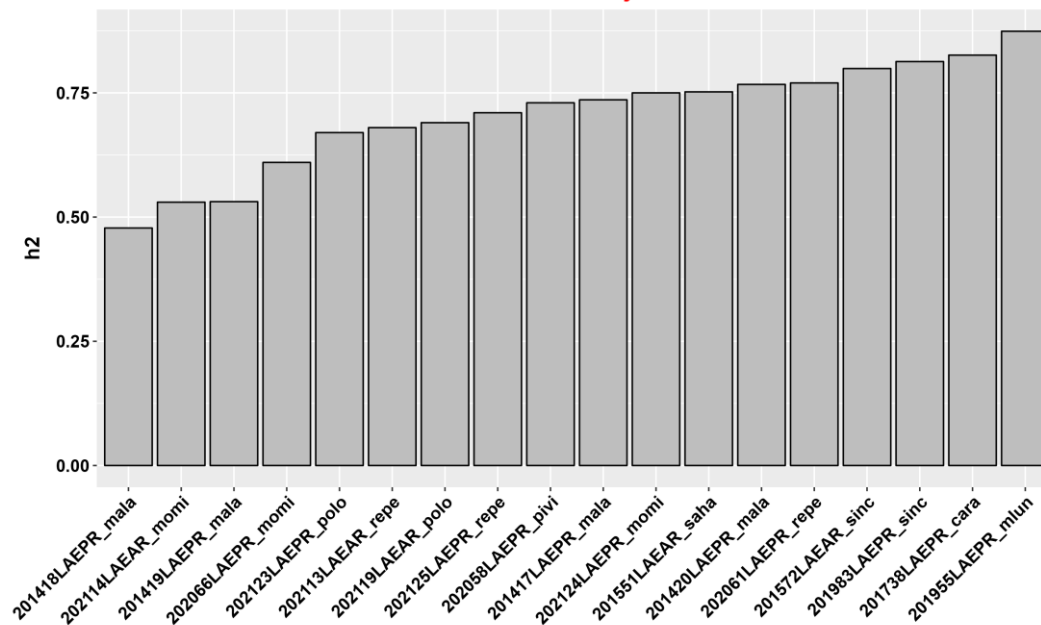
# Tidy data



### Genotypic Variance



### Heritability





# Models

1.

**fixed** = obs\_yield\_ha ~ 1 + accession\_name + yearF,  
**random** = ~ trial\_name + trial\_name:accession\_name,  
**residual** = ~ units

2.

**fixed** = VARcl ~ 1 + accession\_name + yearF,  
**random** = ~ trial\_name + trial\_name:accession\_name + trial\_name:rep\_number +  
at(trial\_name, exp\_rowcol):rep\_number:ROWf +  
at(trial\_name,exp\_rowcol):rep\_number:COLf,  
**residual** = ~ units

3.

**fixed** = VARcl ~ 1 + accession\_name + yearF,  
**random** = ~ trial\_name + trial\_name:accession\_name + trial\_name:rep\_number +  
at(trial\_name, exp\_rowcol):rep\_number:ROWf + at(trial\_name,  
exp\_rowcol):rep\_number:COLf,  
**residual** = ~ dsum(~ units | trial\_name)

4.

**fixed** = VARcl ~ 1 + accession\_name + yearF,  
**random** = ~ trial\_name + diag(trial\_name):accession\_name + trial\_name:rep\_number +  
at(trial\_name, exp\_rowcol):rep\_number:ROWf + at(trial\_name,  
exp\_rowcol):rep\_number:COLf,  
**residual** = ~ dsum(~ units | trial\_name)

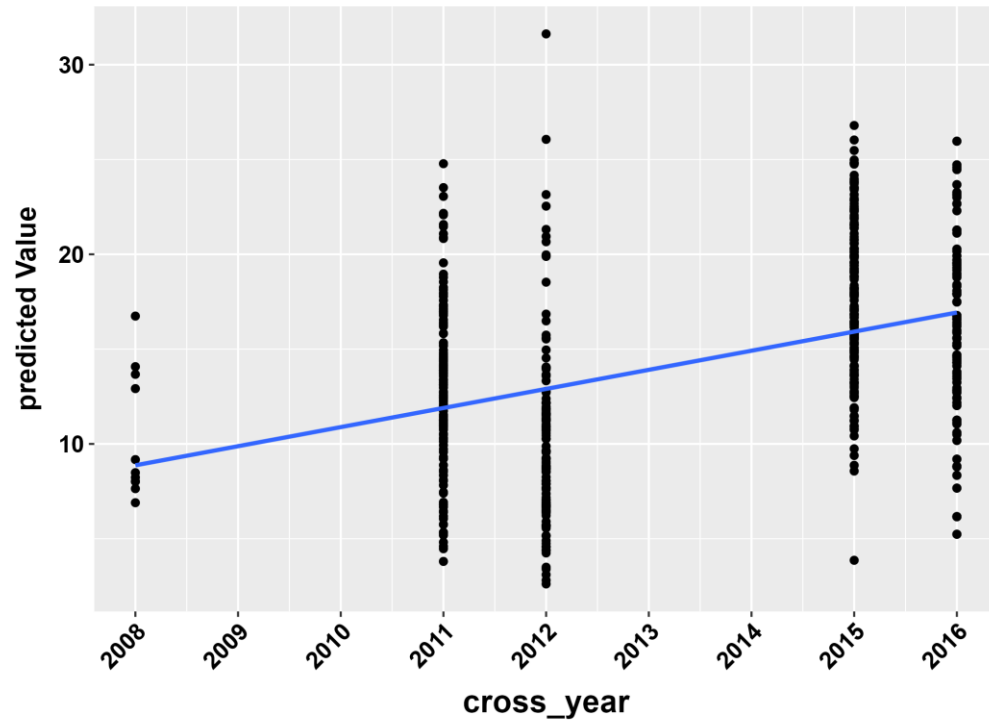
| Model   | AIC      | BIC      |
|---------|----------|----------|
| model 4 | 18231.62 | 18575.77 |
| model 3 | 18248.45 | 18488.13 |
| model 2 | 18894.75 | 19029.96 |
| model 1 | 19132.99 | 19151.43 |

Alliance



## Model 4

Linear Regression

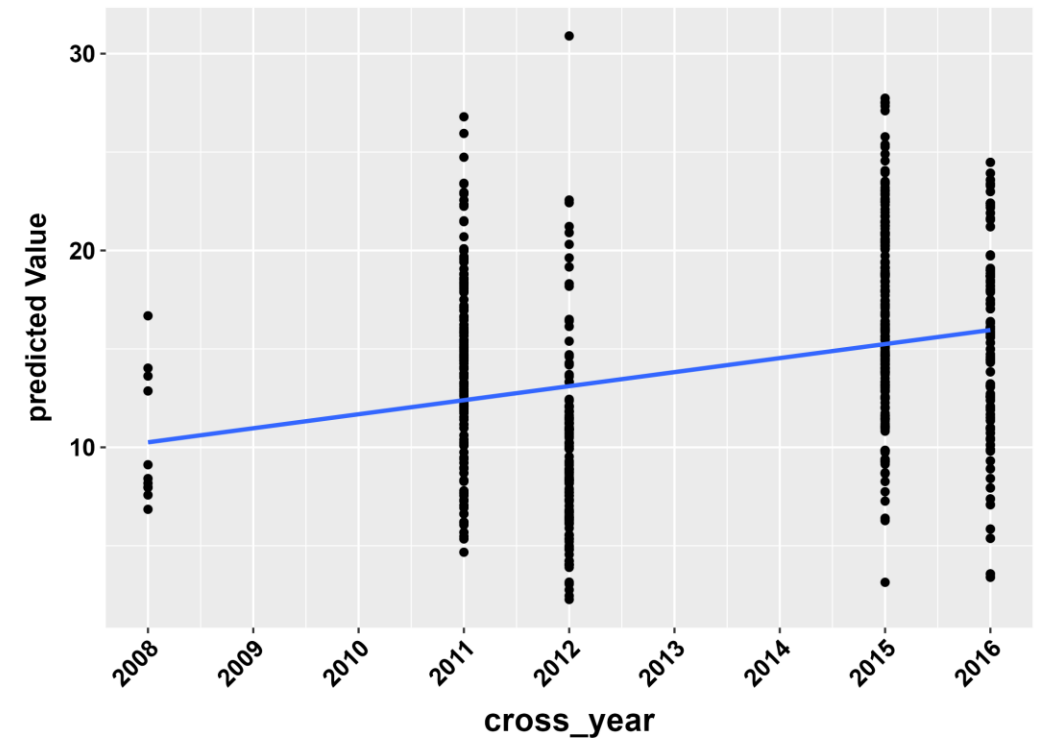


|             |       |
|-------------|-------|
| (Intercept) | slope |
| -2013.3     | 1.007 |

| Model           | R <sup>2</sup> | p-value   | SE       |
|-----------------|----------------|-----------|----------|
| m4 ~ cross_year | 0.1743         | < 2.2e-16 | 0.008727 |

## Model 3

Linear Regression



|             |        |
|-------------|--------|
| (Intercept) | slope  |
| - 1421.17   | 0.7129 |

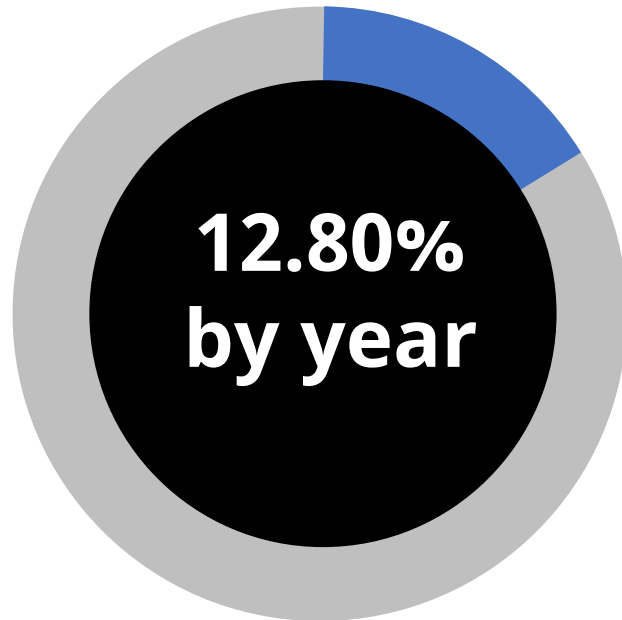
| Model           | R <sup>2</sup> | p-value   | SE     |
|-----------------|----------------|-----------|--------|
| m3 ~ cross_year | 0.114          | 9.377e-14 | 0.1254 |

Alliance

# Genetic Gain Yield

Model 4

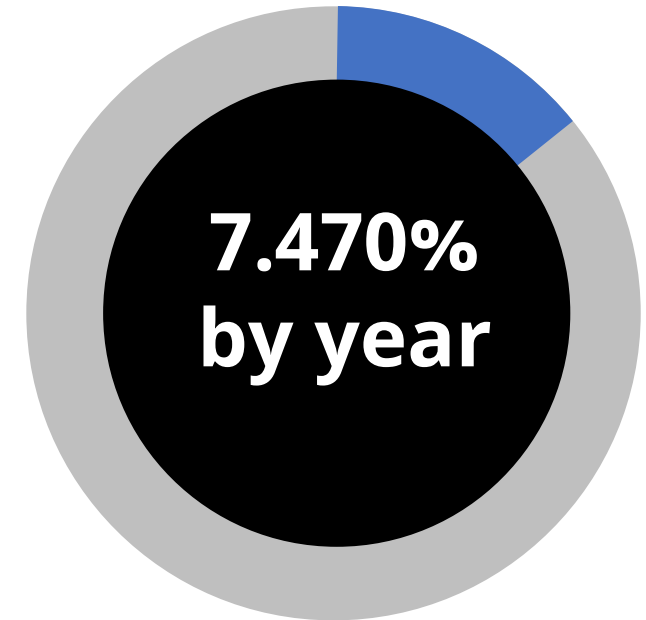
**1.007** t/ha



$$\frac{\text{slope}}{\widehat{\text{firstYear}}} = \frac{1.007}{\text{intercept} + (2007 * \text{slope})} = \frac{1.007}{7.865704} = 0.1280$$

Model 3

**0.713** t/ha



$$\frac{\text{slope}}{\widehat{\text{firstYear}}} = \frac{0.7129}{\text{intercept} + (2007 * \text{slope})} = \frac{0.9643}{9.5426} = 0.0747$$



# Genetic gain LA population

## Dry matter

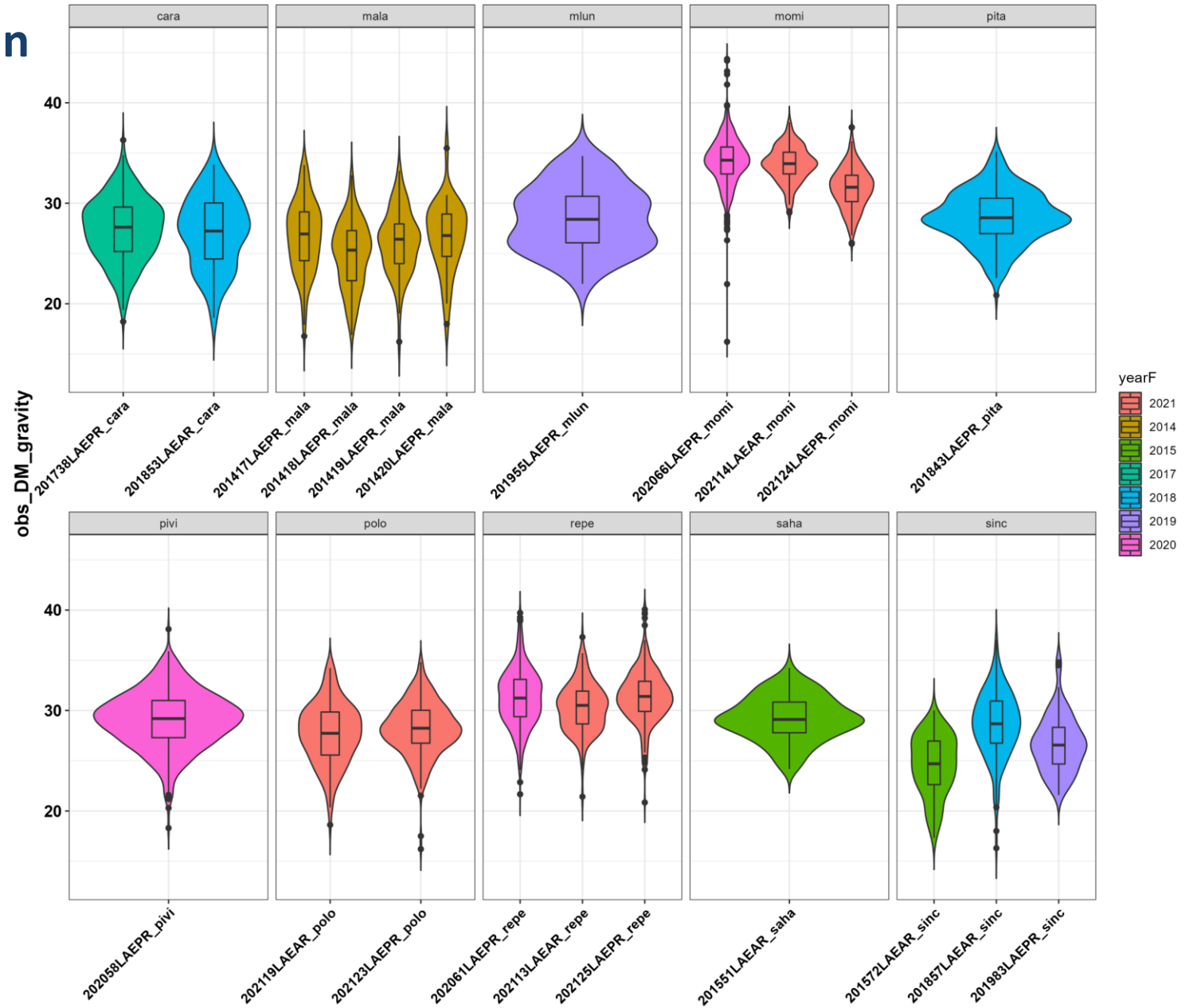
# Waxy population genetic gain

7 years

10 locations

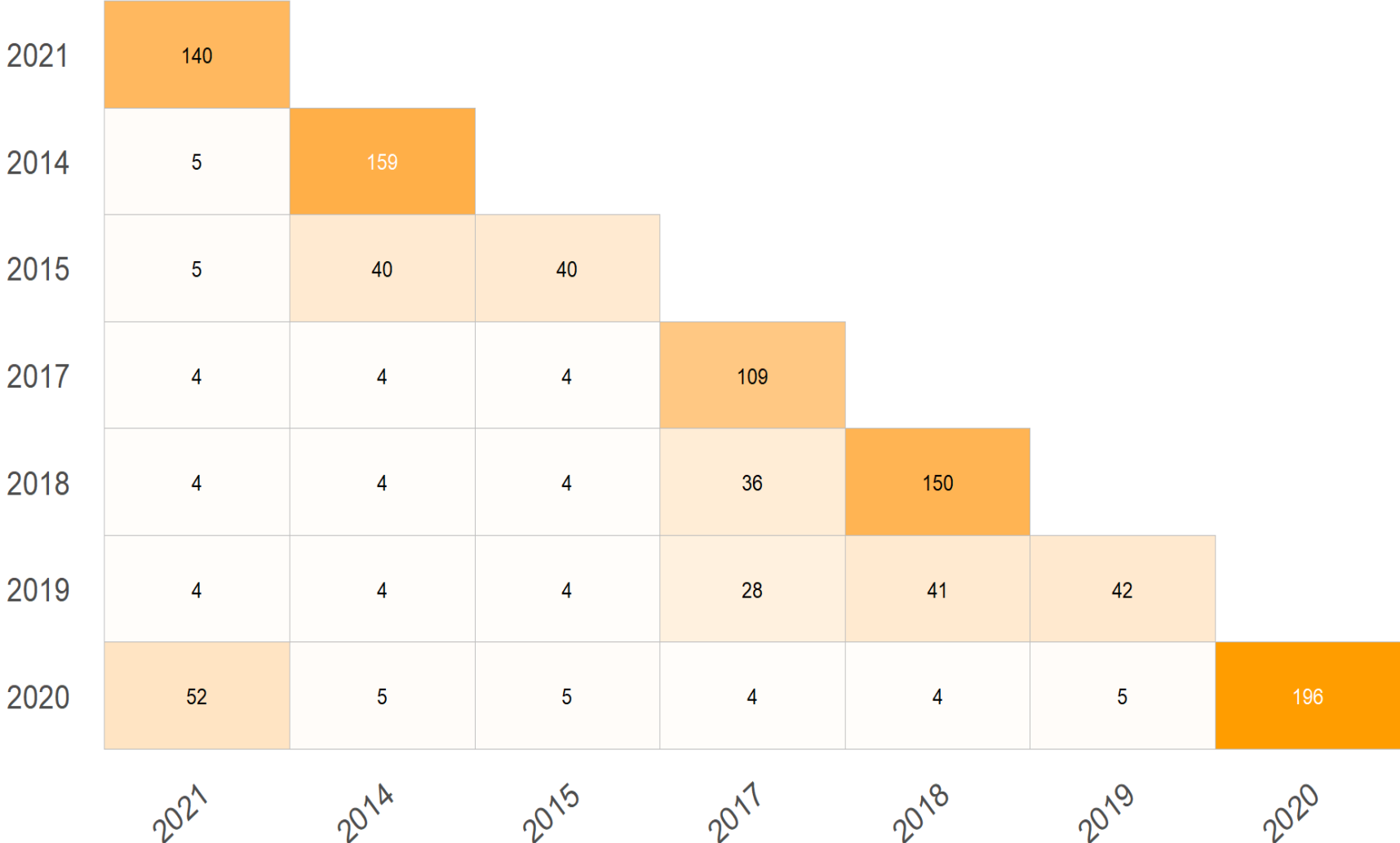
693 waxy clones

21 trials



# Shared information between years

## Shared information



## Models outliers

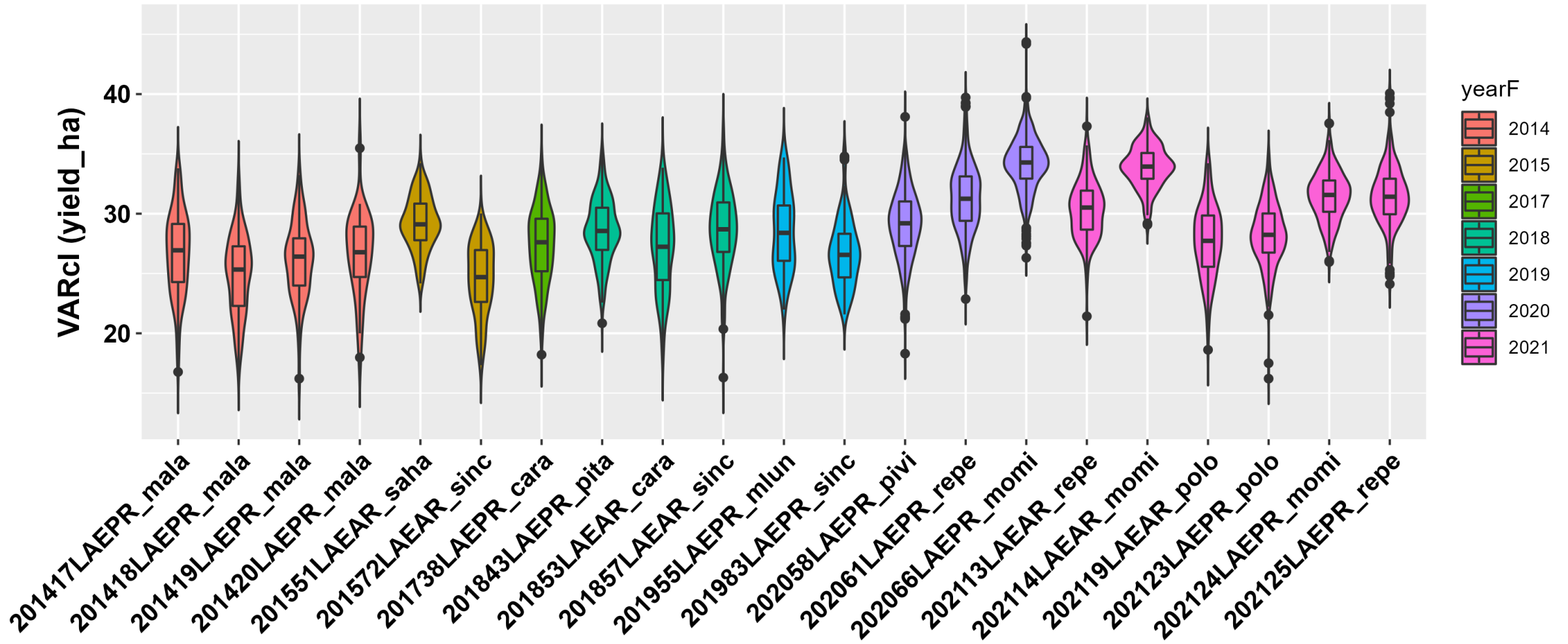
| Experiment       | obs_DM_gravity | rep_number | accession_name | Classify |
|------------------|----------------|------------|----------------|----------|
| 202066LAEPR_momi | 43.1397        | 2          | SM4844-16      | Outlier  |
| 202066LAEPR_momi | 42.8332        | 1          | SM4848-18      | Outlier  |
| 202066LAEPR_momi | 41.8194        | 3          | SM4726-22      | Outlier  |
| 202066LAEPR_momi | 37.1436        | 3          | SM4832-6       | Outlier  |
| 201738LAEPR_cara | 36.2897        | 3          | GM8745-1       | Outlier  |
| 201738LAEPR_cara | 33.2380        | 3          | SM4114-6       | Outlier  |
| 202124LAEPR_momi | 32.7853        | 3          | GM11159-2      | Outlier  |
| 202066LAEPR_momi | 21.9500        | 3          | SM4844-29      | Outlier  |
| 202061LAEPR_repe | 21.6672        | 2          | SM4793-2       | Outlier  |
| 202125LAEPR_repe | 20.8485        | 1          | GM11111-32     | Outlier  |
| 202058LAEPR_pivi | 20.3000        | 2          | SM4772-1       | Outlier  |
| 201857LAEAR_sinc | 18.0067        | 1          | SM4178-5       | Outlier  |
| 202066LAEPR_momi | 16.2100        | 2          | SM4832-6       | Outlier  |



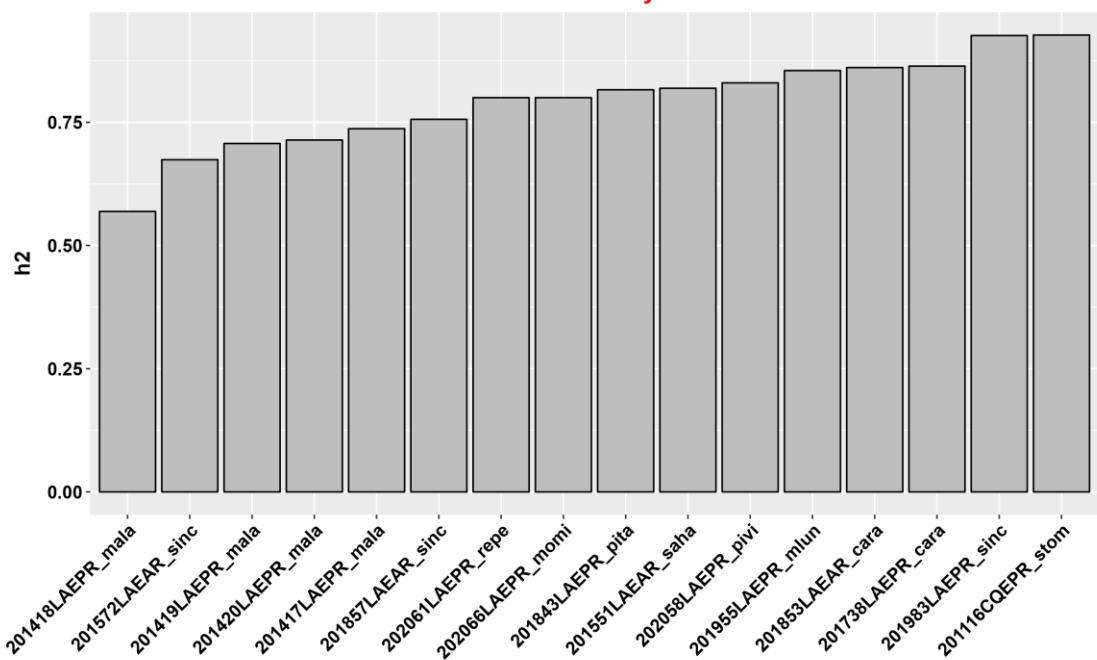
# Mix model results by trial

| Experiment       | y              | varG   | varE   | h2     | outliers | design |
|------------------|----------------|--------|--------|--------|----------|--------|
| 201983LAEPR_sinc | obs_DM_gravity | 6.2798 | 1.5037 | 0.9260 | 0        | rcbd   |
| 202124LAEPR_momi | obs_DM_gravity | 3.1550 | 0.7820 | 0.9200 | 0        | spats  |
| 202114LAEAR_momi | obs_DM_gravity | 2.1770 | 0.7420 | 0.8900 | 0        | spats  |
| 202125LAEPR_repe | obs_DM_gravity | 4.0740 | 1.1940 | 0.8900 | 0        | spats  |
| 201738LAEPR_cara | obs_DM_gravity | 6.9015 | 2.9558 | 0.8640 | 0        | rcbd   |
| 201853LAEAR_cara | obs_DM_gravity | 8.7577 | 3.5749 | 0.8610 | 0        | rcbd   |
| 202123LAEPR_polo | obs_DM_gravity | 3.9130 | 1.6980 | 0.8600 | 0        | spats  |
| 201955LAEPR_mlun | obs_DM_gravity | 7.6481 | 3.2579 | 0.8550 | 0        | rcbd   |
| 202058LAEPR_pivi | obs_DM_gravity | 4.1380 | 2.1580 | 0.8300 | 0        | spats  |
| 201551LAEAR_saha | obs_DM_gravity | 2.9725 | 1.9125 | 0.8190 | 0        | rcbd   |
| 201843LAEPR_pita | obs_DM_gravity | 4.2002 | 2.6414 | 0.8160 | 0        | rcbd   |
| 202066LAEPR_momi | obs_DM_gravity | 2.7040 | 1.9710 | 0.8000 | 2        | spats  |
| 202061LAEPR_repe | obs_DM_gravity | 3.8970 | 2.5870 | 0.8000 | 0        | spats  |
| 202113LAEAR_repe | obs_DM_gravity | 2.8940 | 2.1440 | 0.7800 | 0        | spats  |
| 202119LAEAR_polo | obs_DM_gravity | 4.4330 | 3.7590 | 0.7600 | 0        | spats  |
| 201857LAEAR_sinc | obs_DM_gravity | 5.8074 | 4.7557 | 0.7560 | 0        | rcbd   |
| 201417LAEPR_mala | obs_DM_gravity | 6.7850 | 5.3696 | 0.7370 | 0        | rcbd   |
| 201420LAEPR_mala | obs_DM_gravity | 6.5211 | 4.8183 | 0.7140 | 0        | rcbd   |
| 201419LAEPR_mala | obs_DM_gravity | 5.1599 | 4.5989 | 0.7070 | 0        | rcbd   |
| 201572LAEAR_sinc | obs_DM_gravity | 4.0303 | 4.9278 | 0.6740 | 0        | rcbd   |
| 201418LAEPR_mala | obs_DM_gravity | 4.2119 | 6.7085 | 0.5690 | 0        | rcbd   |

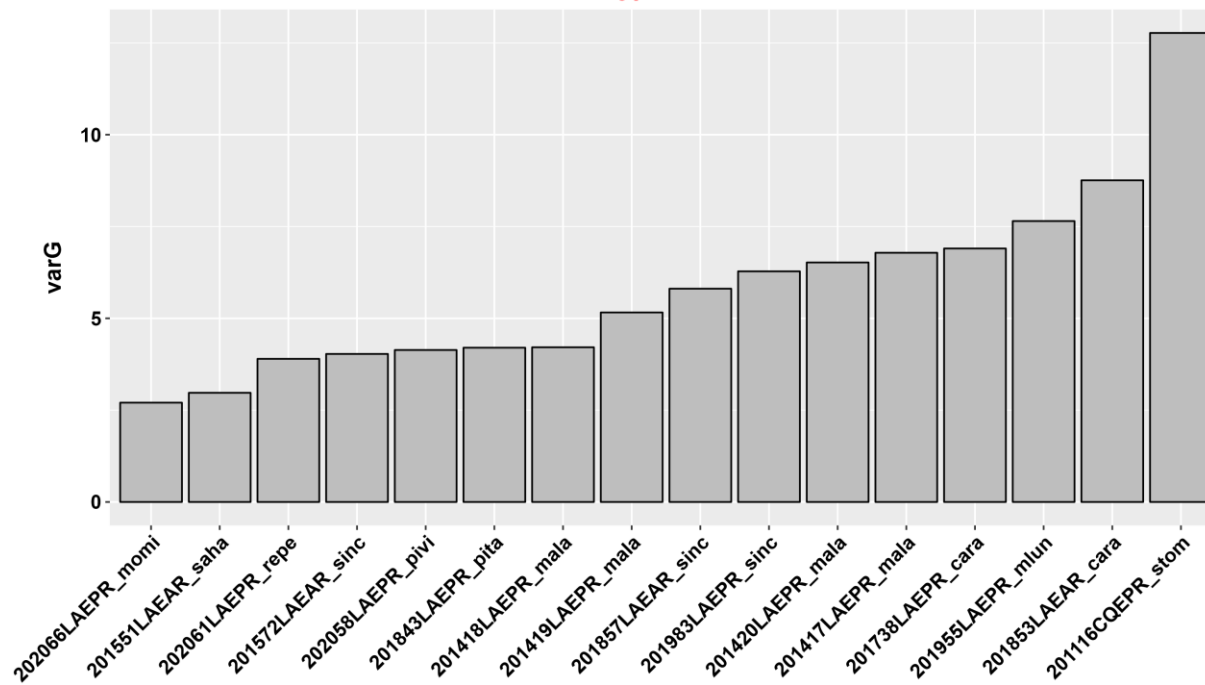
# Tidy data



### Heritability



### Genotypic Variance



Alliance



# Models

1.

**fixed** = obs\_yield\_ha ~ 1 + accession\_name + yearF,  
**random** = ~ trial\_name + trial\_name:accession\_name,  
**residual** = ~ units

2.

**fixed** = VARcl ~ 1 + accession\_name + yearF,  
**random** = ~ trial\_name + trial\_name:accession\_name + trial\_name:rep\_number +  
at(trial\_name, exp\_rowcol):rep\_number:ROWf +  
at(trial\_name,exp\_rowcol):rep\_number:COLf,  
**residual** = ~ units

3.

**fixed** = VARcl ~ 1 + accession\_name + yearF,  
**random** = ~ trial\_name + trial\_name:accession\_name + trial\_name:rep\_number +  
at(trial\_name, exp\_rowcol):rep\_number:ROWf + at(trial\_name,  
exp\_rowcol):rep\_number:COLf,  
**residual** = ~ dsum(~ units | trial\_name)

4.

**fixed** = VARcl ~ 1 + accession\_name + yearF,  
**random** = ~ trial\_name + diag(trial\_name):accession\_name + trial\_name:rep\_number +  
at(trial\_name, exp\_rowcol):rep\_number:ROWf + at(trial\_name,  
exp\_rowcol):rep\_number:COLf,  
**residual** = ~ dsum(~ units | trial\_name)

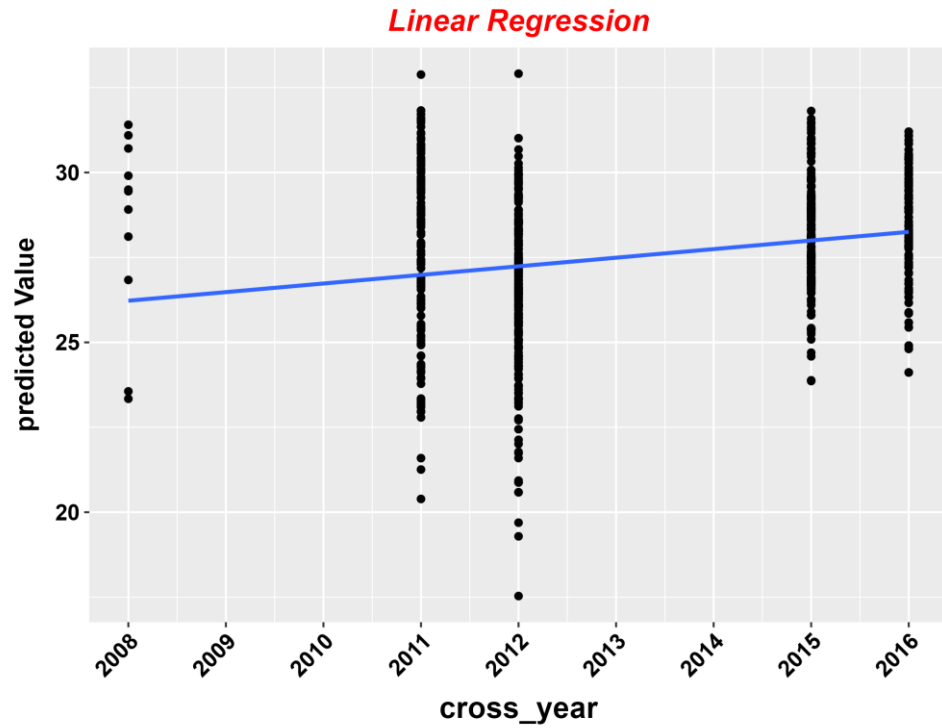
| Model   | AIC      | BIC      |
|---------|----------|----------|
| model 4 | 9453.894 | 9841.157 |
| model 3 | 9473.065 | 9735.404 |
| model 2 | 9779.287 | 9916.703 |
| model 1 | 9873.999 | 9892.737 |

Alliance



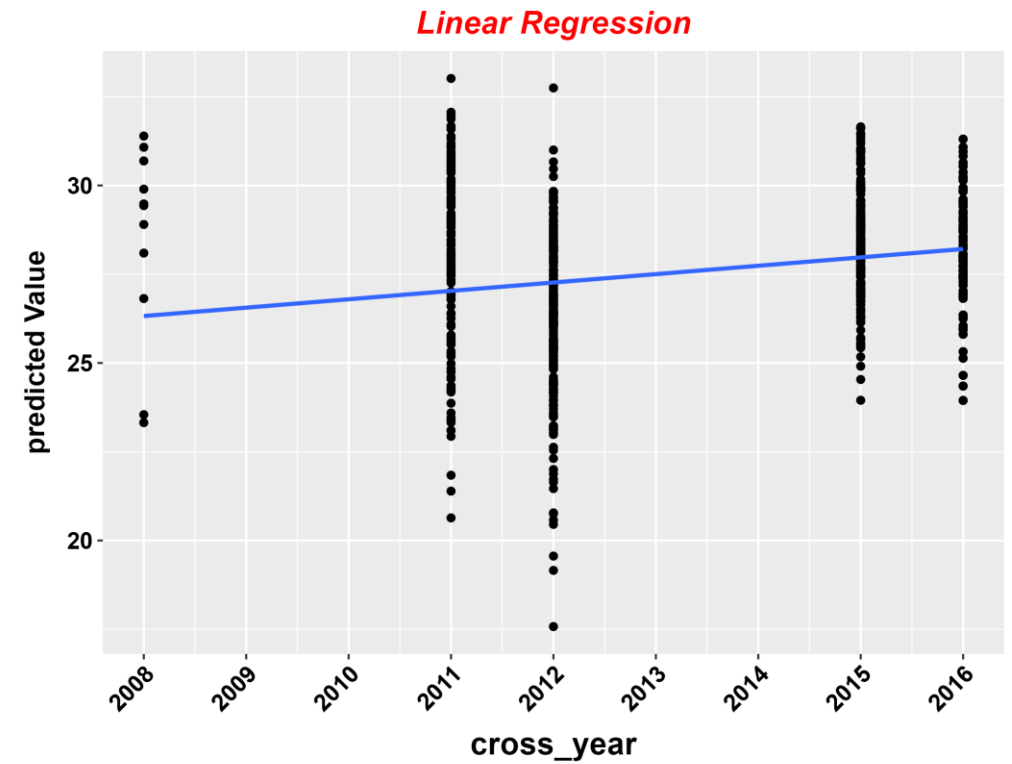


## Model 4



|             |        |
|-------------|--------|
| (Intercept) | slope  |
| -482.263    | 0.2532 |

## Model 3



|             |       |
|-------------|-------|
| (Intercept) | slope |
| -447.497    | 0.236 |

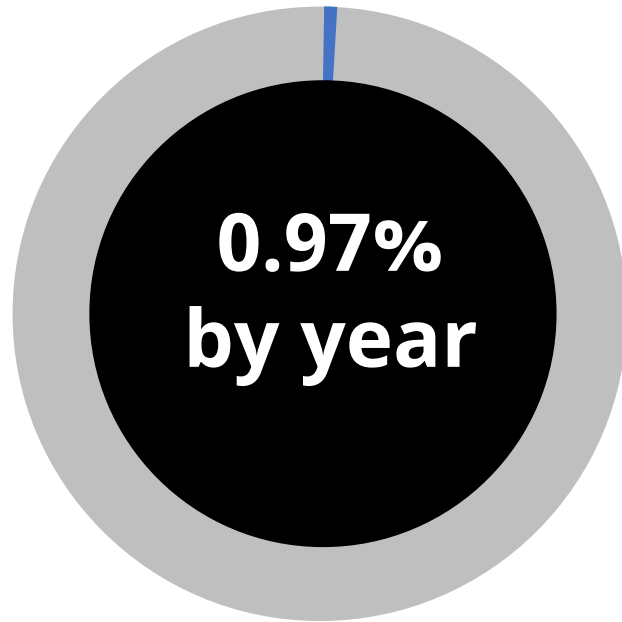
| Model           | R <sup>2</sup> | p-value  | SE      |
|-----------------|----------------|----------|---------|
| m4 ~ cross_year | 0.0568         | 5.93e-11 | 0.03811 |

| Model           | R <sup>2</sup> | p-value   | SE      |
|-----------------|----------------|-----------|---------|
| m3 ~ cross_year | 0.0485         | 1.644e-09 | 0.03864 |

# Genetic Gain Dry Matter

Model 4

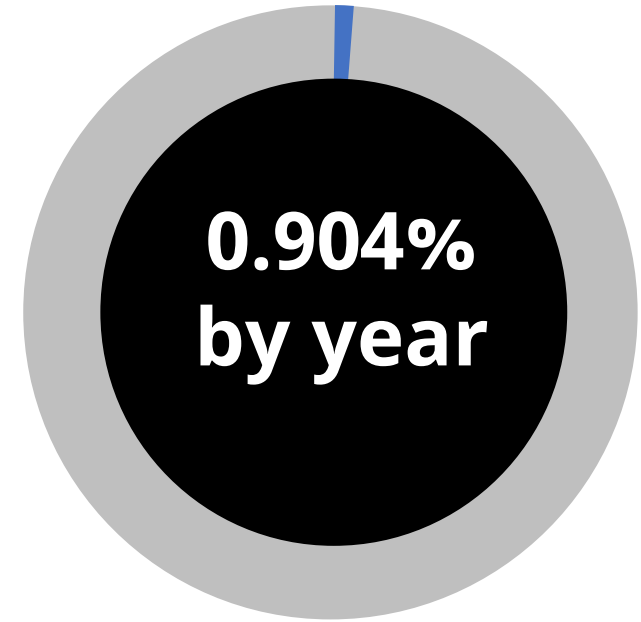
**0.25 %**



$$\frac{\text{slope}}{\widehat{\text{firstYear}}} = \frac{0.253}{\text{intercept} + (2007 * \text{slope})} = \frac{0.253}{25.972} = 0.0097$$

Model 3

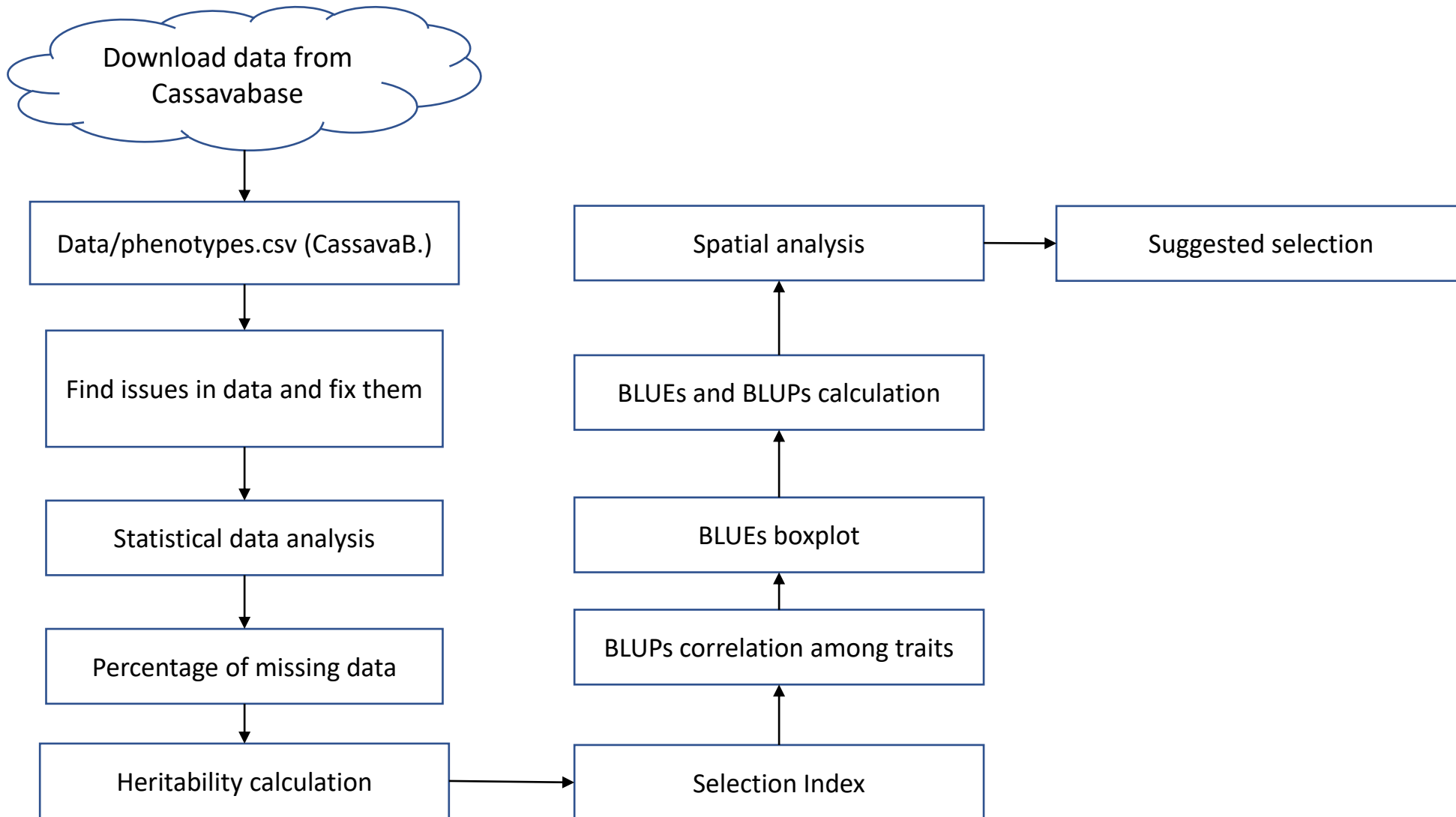
**0.23 %**



$$\frac{\text{slope}}{\widehat{\text{firstYear}}} = \frac{0.236}{\text{intercept} + (2007 * \text{slope})} = \frac{0.236}{26.082} = 0.0090$$

# 3. Trial design pipeline development

# 3. Trial design (pipeline, Process Map)





# Download the data from Cassava Base

The screenshot shows the Cassava Base Search Wizard interface. At the top, there is a navigation bar with 'CASSAVABASE' and menu items: Search, Manage, Analyze, Maps, About. A search bar contains 'Ifdelgadam' and there are buttons for 'Lists' and 'Calendar'. Below the navigation bar, the 'Search Wizard' section has a heading 'Search Wizard' and a sub-heading 'Don't see your data?' with buttons for 'Refresh Lists' and 'Update Wizard'. The main area is divided into four columns. The first column is 'Breeding Programs' with a search bar and a list of programs: SCP, BTI, CARI, CH, CNRA, and CIAT. The second column is 'Trials' with a search bar and a list of trials: 201011CQCOB\_stom, 201012CQCOB\_stom, 201013CQEPR\_pita, 201014CQEPR\_pita, and 201015CQEPR\_pita. The third and fourth columns are 'Select Column Type' with search bars and 'Select All' buttons. At the bottom, there are 'Match ANY ALL' and 'Add to List...' buttons.

Source: <https://cassavabase.org/>



## QBMS















































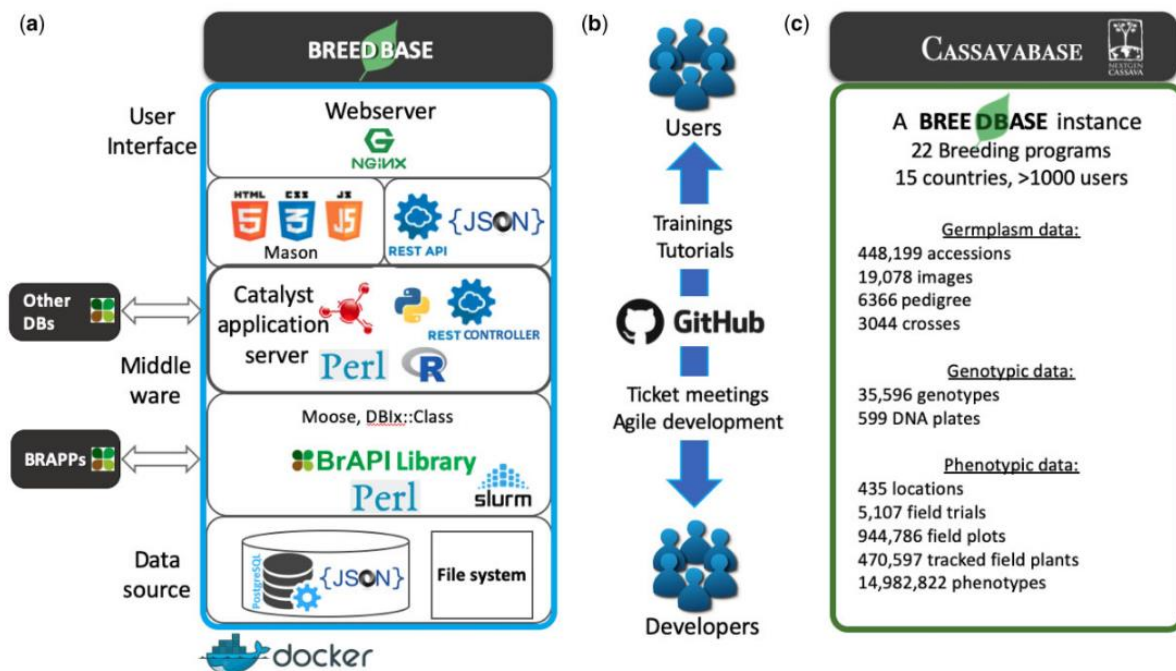
## Overview

Linking data management systems to analytics is an important step in breeding digitalization. Breeders can use this R package to Query the Breeding Management System(s) like [BMS](#), [BreeBase](#), and [GIGWA](#) (using [BrAPI](#) calls) and help them to retrieve phenotypic and genotypic data directly into their analyzing pipelines developed in R statistical environment.

Source: <https://cloud.r-project.org/web/packages/QBMS/QBMS.pdf>

## Breedbase: a digital ecosystem for modern plant breeding

Nicolas Morales <sup>1,2,†</sup> Alex C. Ogonna <sup>1,2,†</sup> Bryan J. Ellerbrock<sup>1,†</sup> Guillaume J. Bauchet <sup>1,16,†</sup> Titima Tantikanjana,<sup>1</sup> Isaak Y. Teclé <sup>1</sup> Adrian F. Powell <sup>1</sup> David Lyon <sup>1,17</sup> Naama Menda <sup>1</sup> Christiano C. Simoes,<sup>1</sup> Surya Saha <sup>1</sup> Prashant Hosmani <sup>1,18</sup> Mirella Flores <sup>1</sup> Naftali Panitz <sup>1</sup> Ryan S. Preble,<sup>1</sup> Afolabi Agbona <sup>3</sup> Ismail Rabbi <sup>3</sup> Peter Kulakow <sup>3</sup> Prasad Peteti <sup>3</sup> Robert Kawuki <sup>4</sup> Williams Esuma <sup>4</sup> Micheal Kanaabi,<sup>4</sup> Doreen M. Chelangat <sup>4</sup> Ezenwanyi Uba,<sup>5</sup> Adeyemi Olojede <sup>5</sup> Joseph Onyeka,<sup>5</sup> Trushar Shah <sup>6</sup> Margaret Karanja <sup>6</sup> Chiedozi Egesi <sup>1,3,5</sup> Hale Tufan <sup>2</sup> Agre Paterne <sup>3</sup> Asrat Asfaw <sup>7</sup> Jean-Luc Jannink <sup>2,8</sup> Mamin Wolfe <sup>2</sup> Clay L. Birkett <sup>2,8</sup> David J. Waring <sup>2,8</sup> Jenna M. Hershberger <sup>2</sup> Michael A. Gore <sup>2</sup> Kelly R. Robbins <sup>2</sup> Trevor Rife <sup>9</sup> Chaney Courtney,<sup>9</sup> Jesse Poland <sup>9</sup> Elizabeth Arnaud <sup>10</sup> Marie-Angélique Laporte <sup>10</sup> Heneriko Kulembeka,<sup>11</sup> Kasele Salum,<sup>11</sup> Emmanuel Mrema,<sup>11</sup> Allan Brown <sup>3</sup> Stanley Bayo,<sup>3</sup> Brigitte Uwimana <sup>3</sup> Violet Akech,<sup>3</sup> Craig Yencho <sup>12</sup> Bert de Boeck <sup>13</sup> Hugo Campos <sup>13</sup> Rony Swennen <sup>14</sup> Jeremy D. Edwards <sup>15</sup> Lukas A. Mueller <sup>1,\*</sup>



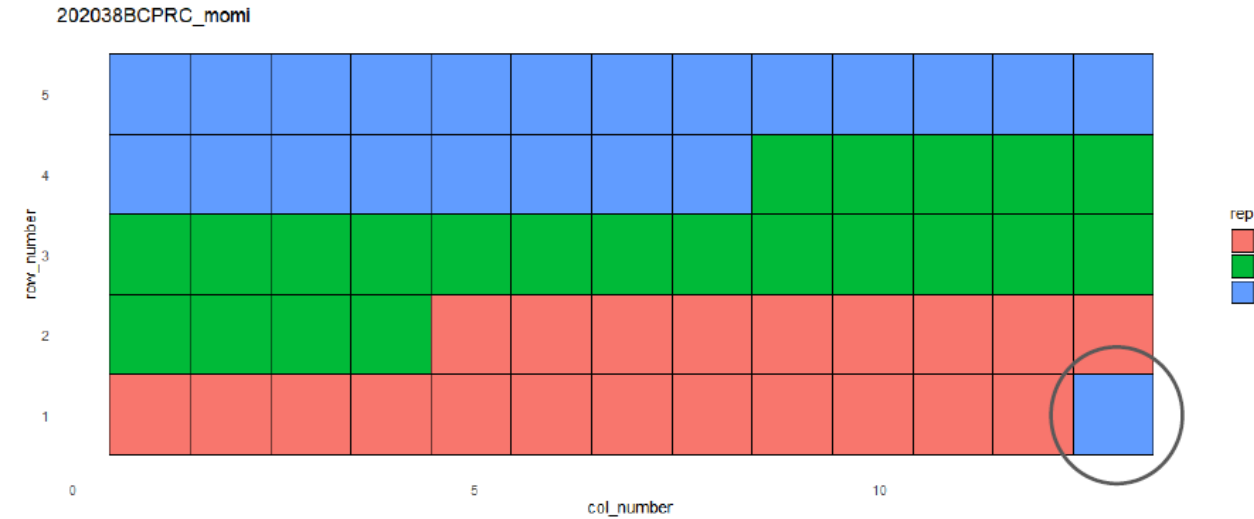
# Find issues in data and fix them

| use plot name                        | use col number | use row number | use trial name   | use plot number |
|--------------------------------------|----------------|----------------|------------------|-----------------|
| 201942LACOB_palm_rep1_SM4821-31_1389 | 47             | 29             | 201942LACOB_palm | 1389            |
| 201942LACOB_palm_rep1_SM4821-35_1390 | 48             | 29             | 201942LACOB_palm | 1390            |
| 201942LACOB_palm_rep1_SM4821-36_1391 | 49             | 29             | 201942LACOB_palm | 1391            |
| 201942LACOB_palm_rep1_SMB2446-2_1392 | 50             | 29             | 201942LACOB_palm | 1392            |
| 201942LACOB_palm_rep1_SM4824-35_1393 | 50             | 29             | 201942LACOB_palm | 1393            |
| 201942LACOB_palm_rep1_SM4826-37_1394 | 49             | 29             | 201942LACOB_palm | 1394            |
| 201942LACOB_palm_rep1_SM4838-15_1395 | 48             | 29             | 201942LACOB_palm | 1395            |
| 201942LACOB_palm_rep1_SM4840-40_1396 | 47             | 29             | 201942LACOB_palm | 1396            |
| 201942LACOB_palm_rep1_SM4838-15_1395 | 48             | 29             | 201942LACOB_palm | 1395            |
| 201942LACOB_palm_rep1_SM4840-40_1396 | 47             | 29             | 201942LACOB_palm | 1396            |

**Error 1: different plot numbers with the same row and column information**

**How can we change the error in CassavaBase?**

**Trial layout with issue: col number 13 – row number 1 have 2 assigned plots**



**Data to fix: col number 13 – row number 1 have 2 assigned plots**

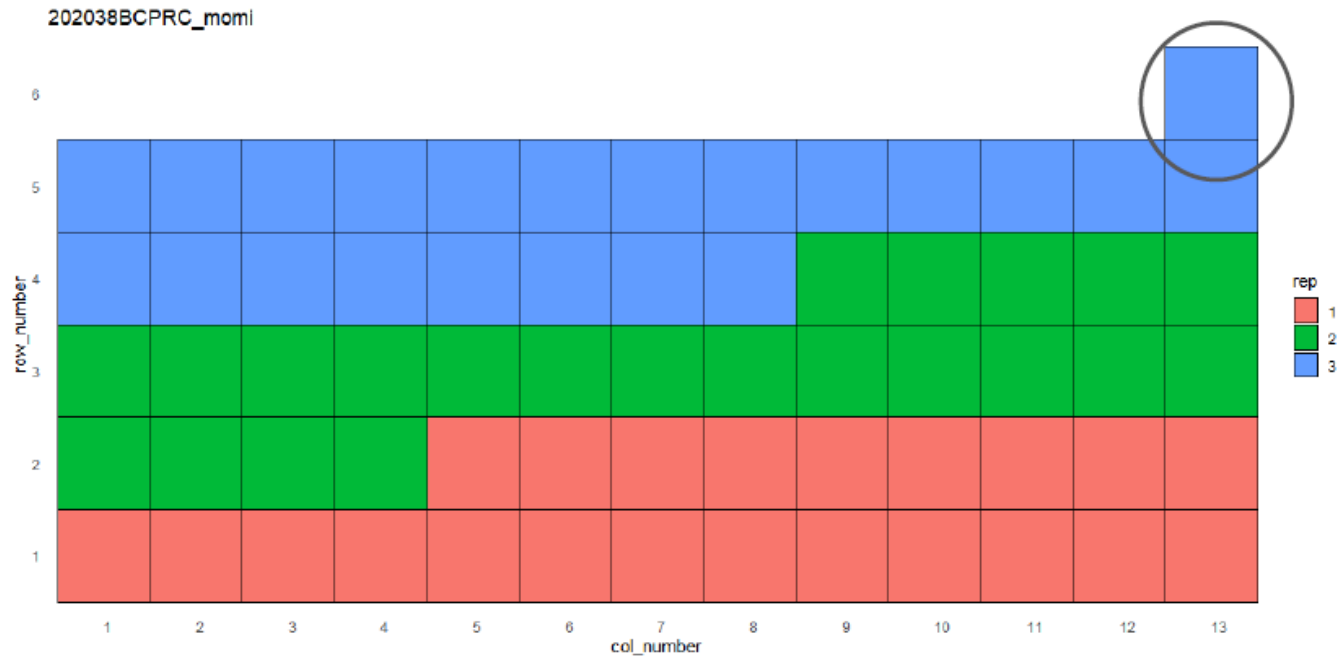
| observationUnitName               | replicate | blockNumber | plotNumber | rowNumber | colNumber |
|-----------------------------------|-----------|-------------|------------|-----------|-----------|
| 202038BCPRC_momi_rep1_Tai8_13     | 1         | 1           | 13         | 1         | 13        |
| 202038BCPRC_momi_rep3_GM8956-1_66 | 3         | 3           | 66         | 1         | 13        |

**Error 2: Wrong replication number**

**How can we change the error in CassavaBase?**

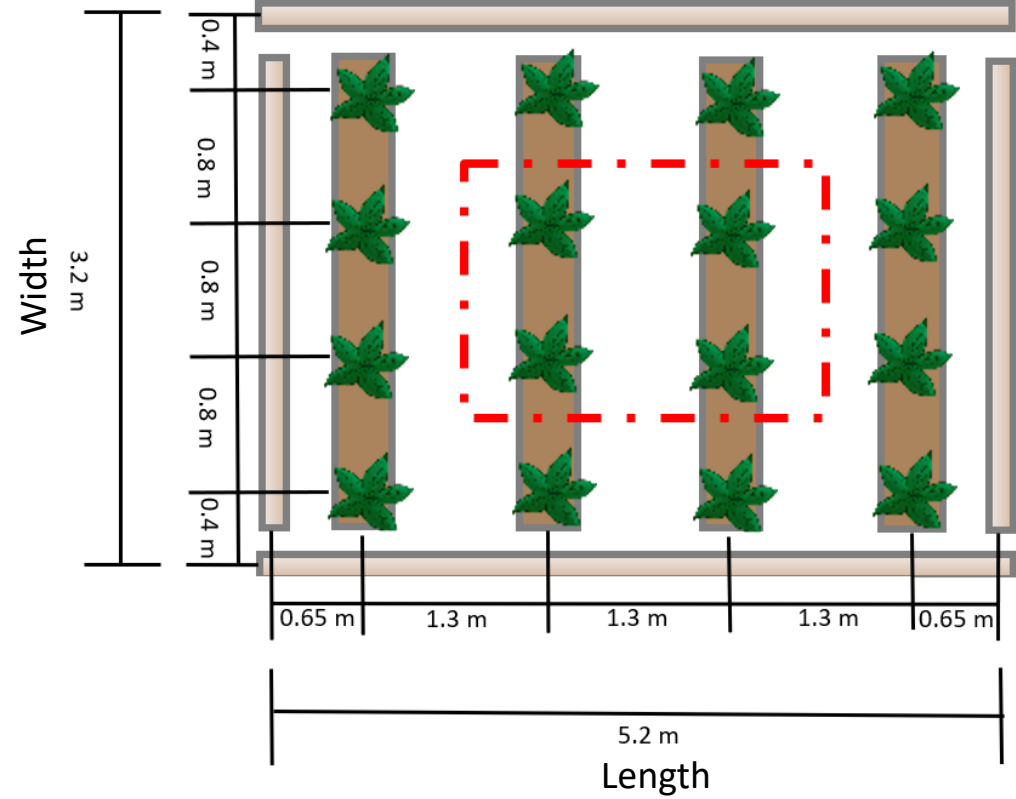
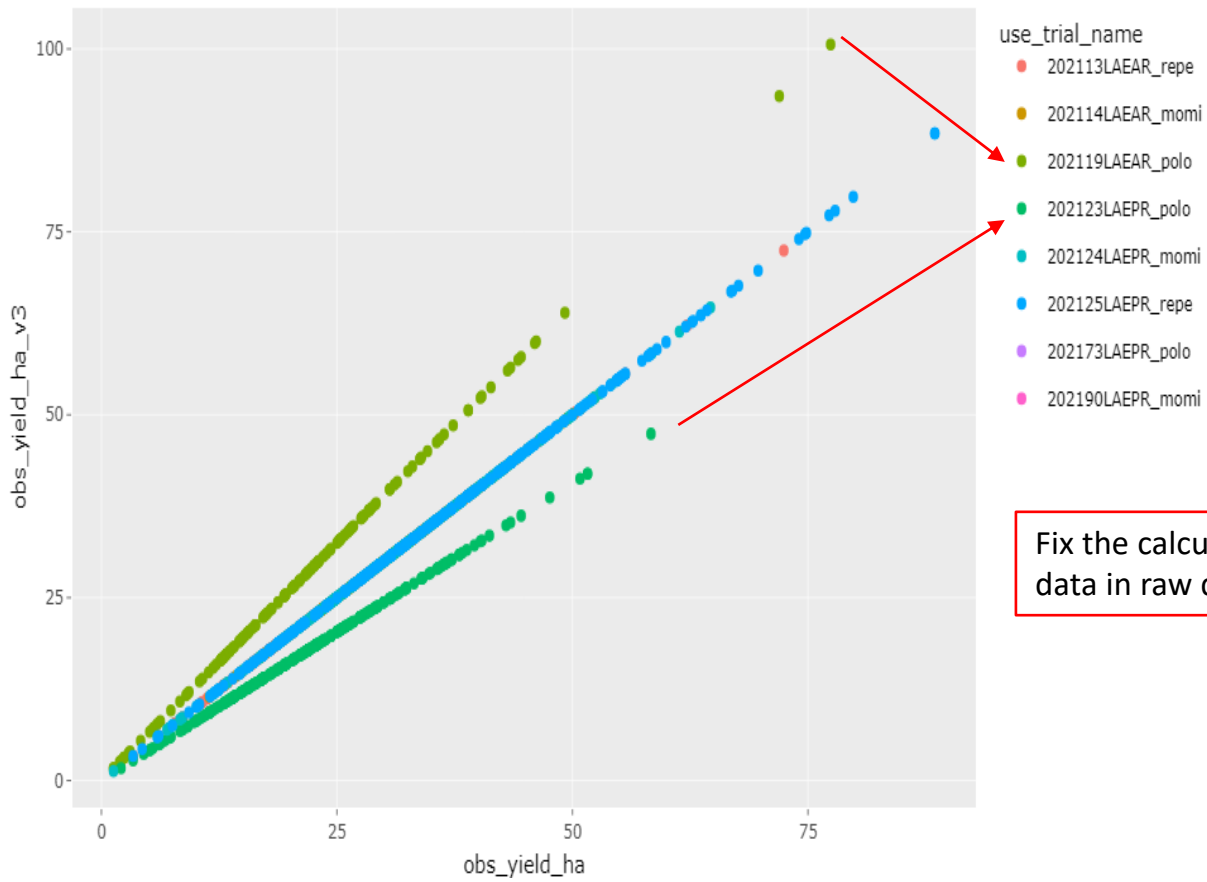


## Trial layout fixed



## Data fixed

| observationUnitName               | replicate | blockNumber | plotNumber | rowNumber | colNumber |
|-----------------------------------|-----------|-------------|------------|-----------|-----------|
| 202038BCPRC_momi_rep1_Tai8_13     | 1         | 1           | 13         | 1         | 13        |
| 202038BCPRC_momi_rep3_GM8956-1_66 | 3         | 3           | 66         | 6         | 13        |

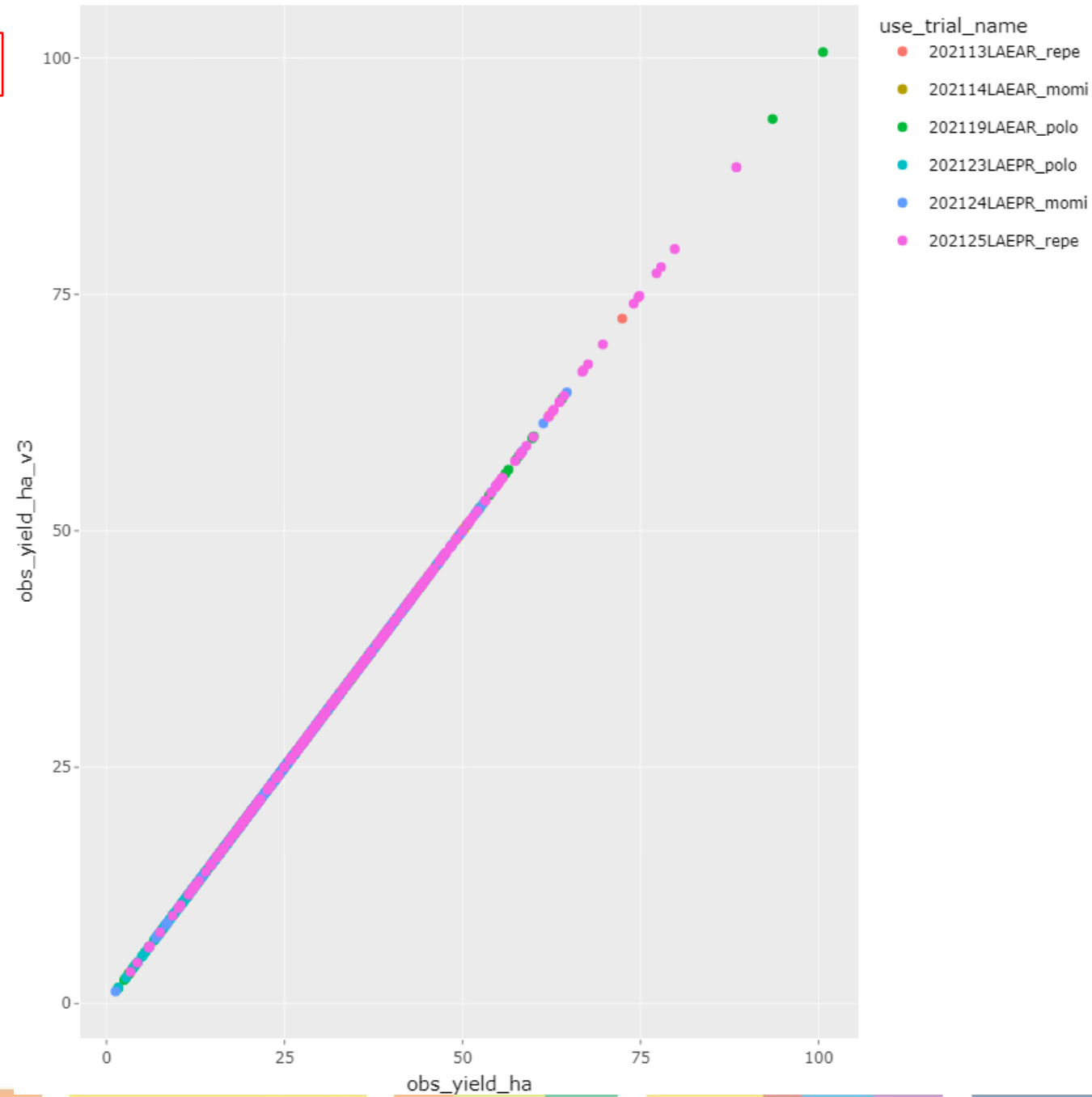


**Error 3: used wrong formula**  
**How can we change the error in CassavaBase?**

$$Area\ per\ plant = \frac{plot_{length} * plot_{width}}{obs_{planted}_{number}}; \frac{m^2}{plants}$$

$$obs_{yield}_{ha_v3} = \frac{obs_{root_{weight}_{plot}(kg)}}{Area\ per\ plant\ (m^2) * obs_{harvest}_{number}_{plan}} * \frac{10,000m^2}{1ha} * \frac{1\ t}{1000kg}; \frac{t}{ha}$$

Issue fixed!!!

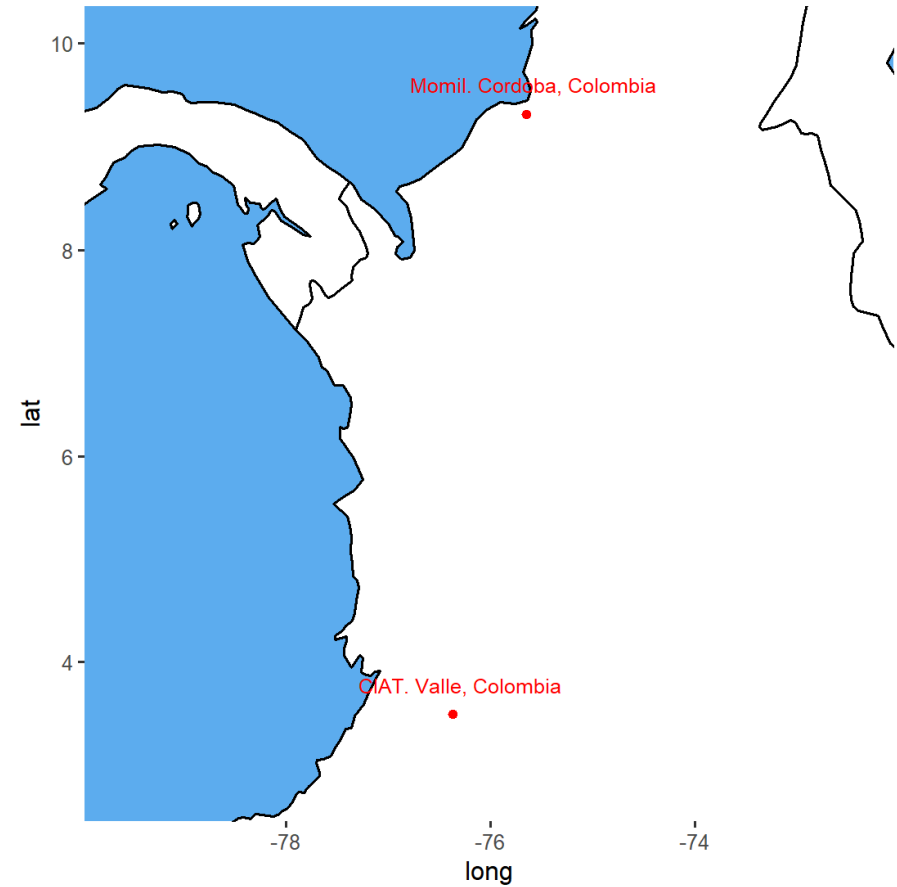
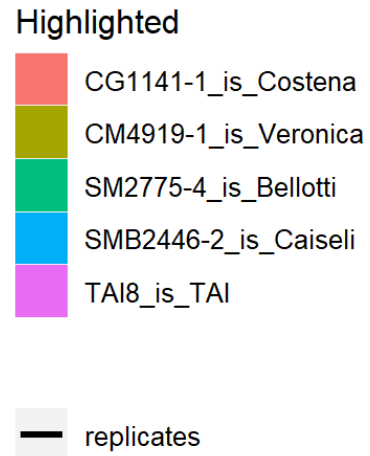
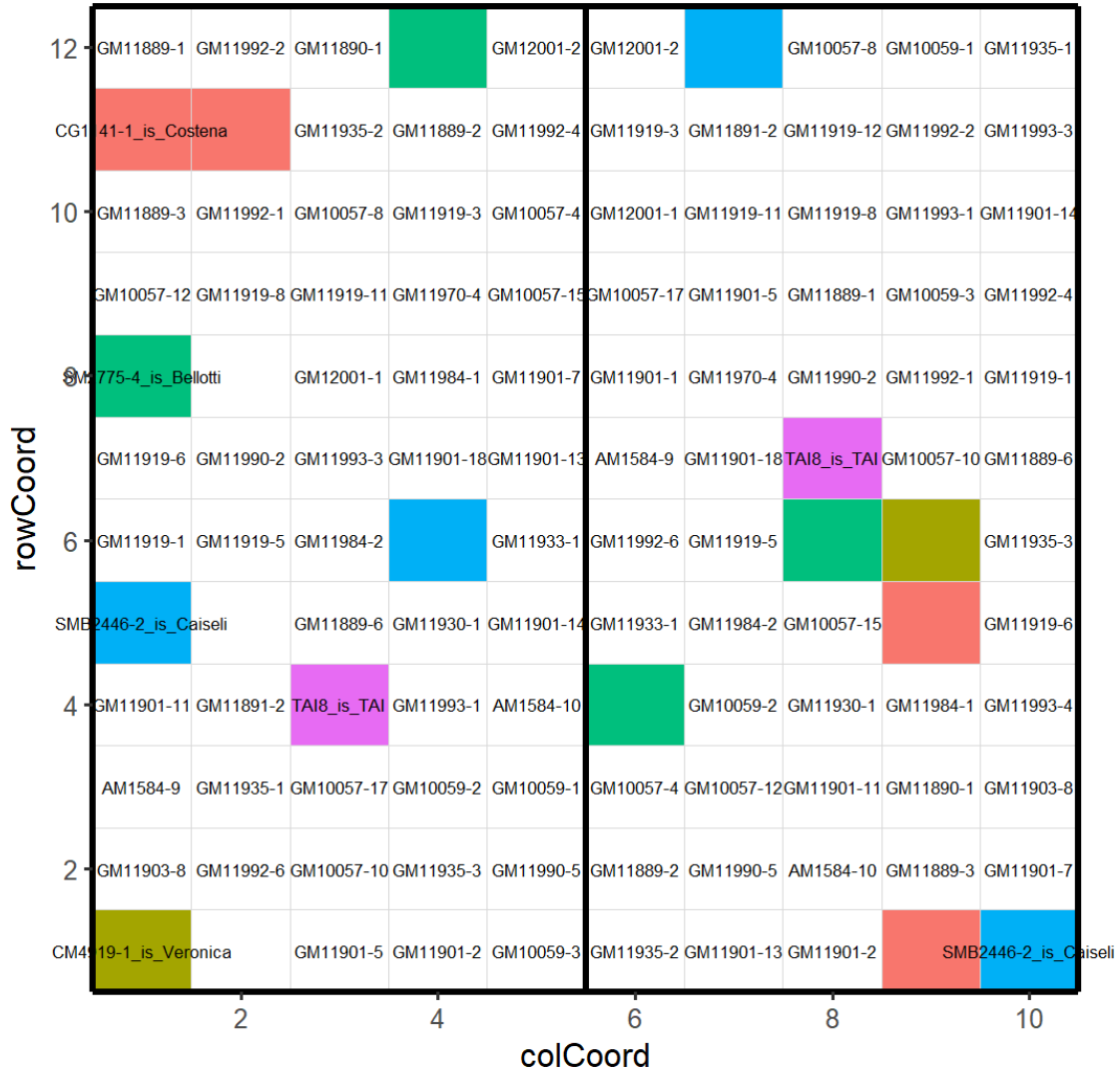


# Percentage of missing data

| trial_name       | DM_gravity | height_1st_branch | yield_ha | root_weight_plot | frogskin1_5 | vigor1_5 |
|------------------|------------|-------------------|----------|------------------|-------------|----------|
| 202138MDEPR_ciat | 0          | 0                 | 0        | 0                | 1           | 1        |
| 202152MDEPR_momi | 1          | 1                 | 1        | 1                | 0           | 1        |

# Trial layout and map

202152MDEPR\_momi





# Heritability calculation (broad sense)

| trial            | DM_gravity | height_1st_branch | lodging1_3 | plant_type | root_skin_color1_3 | root_type1_5 | germination_perc | yield_ha_v3 | DM_yield |
|------------------|------------|-------------------|------------|------------|--------------------|--------------|------------------|-------------|----------|
| 202152MDEPR_momi | 0.83       | 0.66              | 0.59       | 0.68       | 0.97               | 0.69         | 0.2              | 0.78        | 0.81     |

Measure of the extent to which **phenotypic variation** is due to **genetic variation** between individuals in that population

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\epsilon^2}{n}}$$

$\sigma_g^2$  = genotypic variance

$\sigma_\epsilon^2$  = error variance (residual)

$n$  = number of replicates



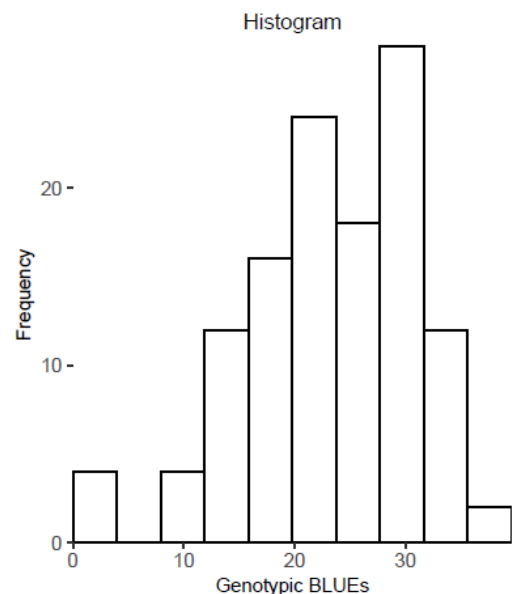
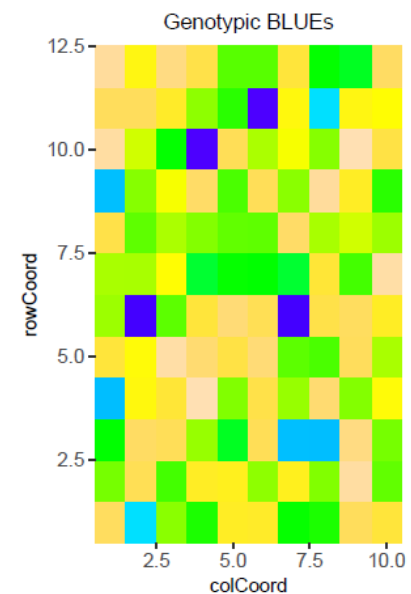
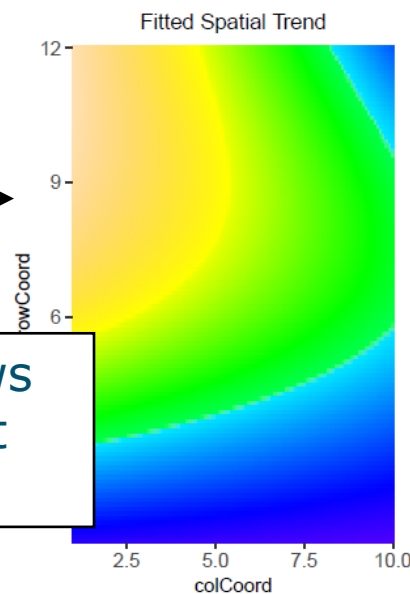
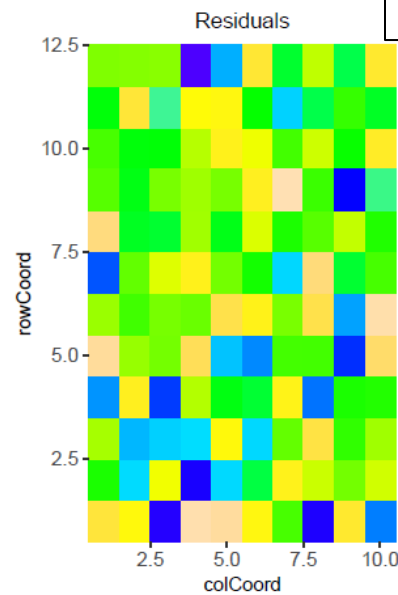
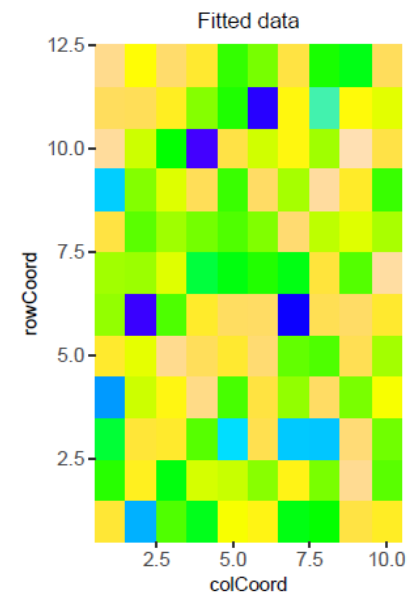
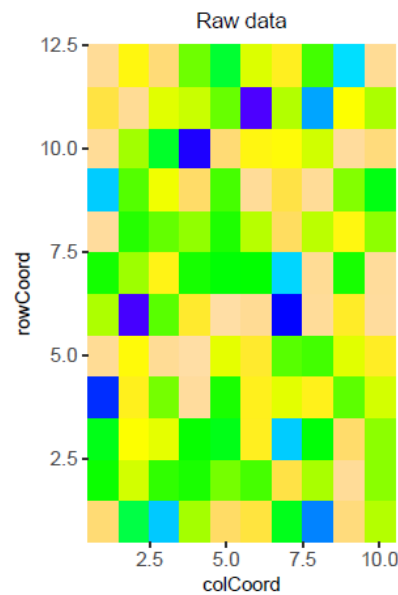
**$H^2 = 0$**   
variation is not explained by the genotypes

**$H^2 = 1$**   
variation is fully explained by the genotypes

# Spatial analysis

Trial: 202152MDEPR\_momi Trait: yield\_ha\_v3

Yield variation that is not due to genetics and does not depend on the row and column positions



Yield variation along rows and columns that is not due to genetics



ELSEVIER

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: [www.elsevier.com/locate/spasta](http://www.elsevier.com/locate/spasta)



## Correcting for spatial heterogeneity in plant breeding experiments with P-splines

María Xosé Rodríguez-Álvarez<sup>a,b,\*</sup>, Martin P. Boer<sup>c</sup>,  
Fred A. van Eeuwijk<sup>c</sup>, Paul H.C. Eilers<sup>d</sup>

<sup>a</sup> BCAM – Basque Center for Applied Mathematics, Alameda de Mazarredo, 14. E-48009 Bilbao, Basque Country, Spain

<sup>b</sup> IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

<sup>c</sup> Biometris, Wageningen University & Research, Wageningen, The Netherlands

<sup>d</sup> Erasmus University Medical Centre, Rotterdam, The Netherlands



## Package ‘SpATS’

February 26, 2021

**Type** Package

**Title** Spatial Analysis of Field Trials with Splines

**Version** 1.0-12

**Date** 2021-02-26

**Imports** stats, grDevices, graphics, fields, plot3Drgl, spam,  
data.table, methods

**Description** Analysis of field trial experiments by modelling spatial trends using two-dimensional Penalised spline (P-spline) models.

Theor Appl Genet (2017) 130:1375–1392  
DOI 10.1007/s00122-017-2894-4



CrossMark

ORIGINAL ARTICLE

## Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model

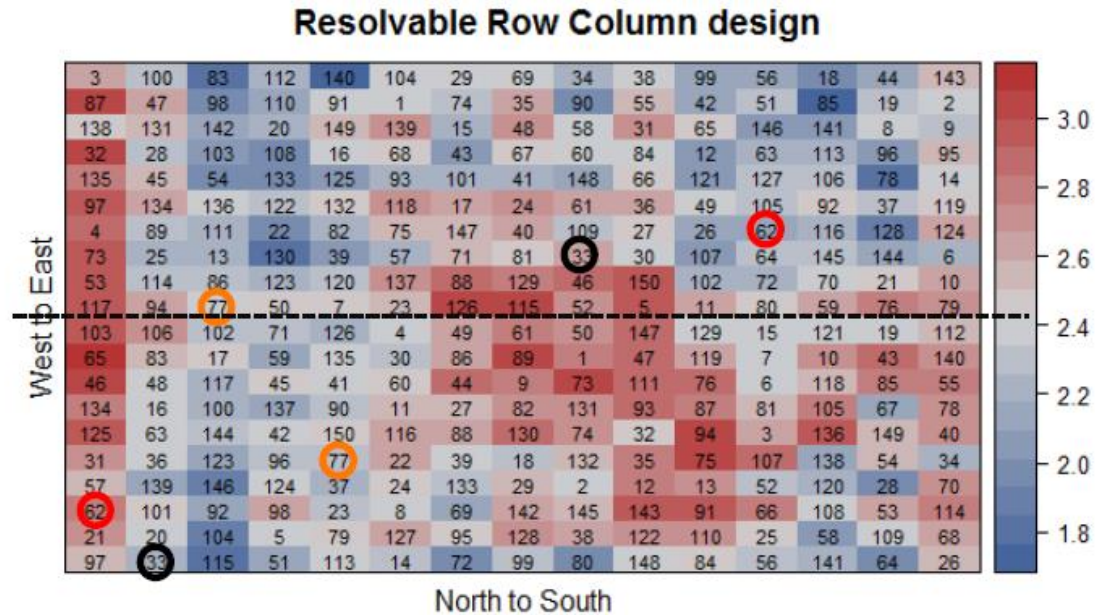
Julio G. Velazco<sup>1,2</sup> · María Xosé Rodríguez-Álvarez<sup>3,4</sup> · Martin P. Boer<sup>1</sup> ·  
David R. Jordan<sup>5</sup> · Paul H. C. Eilers<sup>6</sup> · Marcos Malosetti<sup>1</sup> · Fred A. van Eeuwijk<sup>1</sup>

Received: 2 November 2016 / Accepted: 18 March 2017 / Published online: 3 April 2017  
© The Author(s) 2017. This article is an open access publication

Alliance



# Randomization: resolvable row-column



- All 150 in each half of the field.
- Genotypes within same row (column) in rep 1 will not be in the same row (column) in rep 2.

# Analysis Resolvable Row-Column design (genotypes random > BLUPs)

$$Y_{ijklm} = \mu + Rep_m + R_{j(m)} + C_{k(m)} + G_i + \epsilon_{ijklm}$$

$$G_i \sim N(0, \sigma_g^2)$$
$$R_j \sim N(0, \sigma_r^2)$$
$$C_k \sim N(0, \sigma_c^2)$$
$$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

```
m4 <- lmer(yield ~ Rep + (1|Rep:Row) + (1|Rep:Col) + (1|Geno),
           data = rowcol2$Book)
```

Random effects:

| Groups   | Name        | Variance | Std.Dev. |
|----------|-------------|----------|----------|
| Geno     | (Intercept) | 0.02652  | 0.1629   |
| Rep:Row  | (Intercept) | 0.03792  | 0.1947   |
| Rep:Col  | (Intercept) | 0.01986  | 0.1409   |
| Residual |             | 0.02464  | 0.1570   |

Number of obs: 300, groups: Geno, 150; Rep:Row, 30; Rep:Col, 20

Fixed effects:

|             | Estimate | Std. Error | df       | t value | Pr(> t )   |
|-------------|----------|------------|----------|---------|------------|
| (Intercept) | 2.37774  | 0.06968    | 42.04000 | 34.123  | <2e-16 *** |
| Rep1        | 0.12587  | 0.09673    | 39.15000 | 1.301   | 0.201      |

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\epsilon^2}{n}}$$
$$H^2 \cong \frac{0.02652}{0.02652 + \frac{0.02464}{2}} = 0.68$$

- Note that rows and columns are **nested** within replicates (the : operator used in R).
- Heritability is the highest among all the designs.



## Analysis resolvable row-column (genotypes fixed > BLUEs)

$$y_{ijklm} = \mu + Rep_m + R_{j(m)} + C_{k(m)} + G_i + \epsilon_{ijklm}$$

$$R_j \sim N(0, \sigma_r^2)$$
$$C_k \sim N(0, \sigma_c^2)$$
$$\epsilon_{ijklm} \sim N(0, \sigma_\epsilon^2)$$

```
m4f <- lmer(yield ~ Rep + (1|Rep:Row) + (1|Rep:Col) + Geno,  
           data = rowcol2$Book)
```

Least Squares Means table:

|         | Rep | Geno | Estimate | Standard Error | DF  | t-value | Lower CI | Upper CI | p-value    |
|---------|-----|------|----------|----------------|-----|---------|----------|----------|------------|
| Geno 1  | NA  | 1    | 2.379    | 0.136          | 140 | 17.6    | 2.11     | 2.65     | <2e-16 *** |
| Geno 2  | NA  | 63   | 2.379    | 0.135          | 140 | 17.6    | 2.11     | 2.65     | <2e-16 *** |
| Geno 3  | NA  | 74   | 2.355    | 0.134          | 139 | 17.5    | 2.09     | 2.62     | <2e-16 *** |
| Geno 4  | NA  | 85   | 2.241    | 0.135          | 139 | 16.6    | 1.97     | 2.51     | <2e-16 *** |
| Geno 5  | NA  | 96   | 2.632    | 0.135          | 140 | 19.5    | 2.36     | 2.90     | <2e-16 *** |
| Geno 6  | NA  | 107  | 2.312    | 0.136          | 140 | 17.0    | 2.04     | 2.58     | <2e-16 *** |
| Geno 7  | NA  | 118  | 2.347    | 0.135          | 140 | 17.4    | 2.08     | 2.61     | <2e-16 *** |
| Geno 8  | NA  | 129  | 2.382    | 0.133          | 139 | 17.9    | 2.12     | 2.65     | <2e-16 *** |
| Geno 9  | NA  | 140  | 2.218    | 0.135          | 139 | 16.5    | 1.95     | 2.48     | <2e-16 *** |
| Geno 10 | NA  | 2    | 2.484    | 0.135          | 139 | 18.4    | 2.22     | 2.75     | <2e-16 *** |

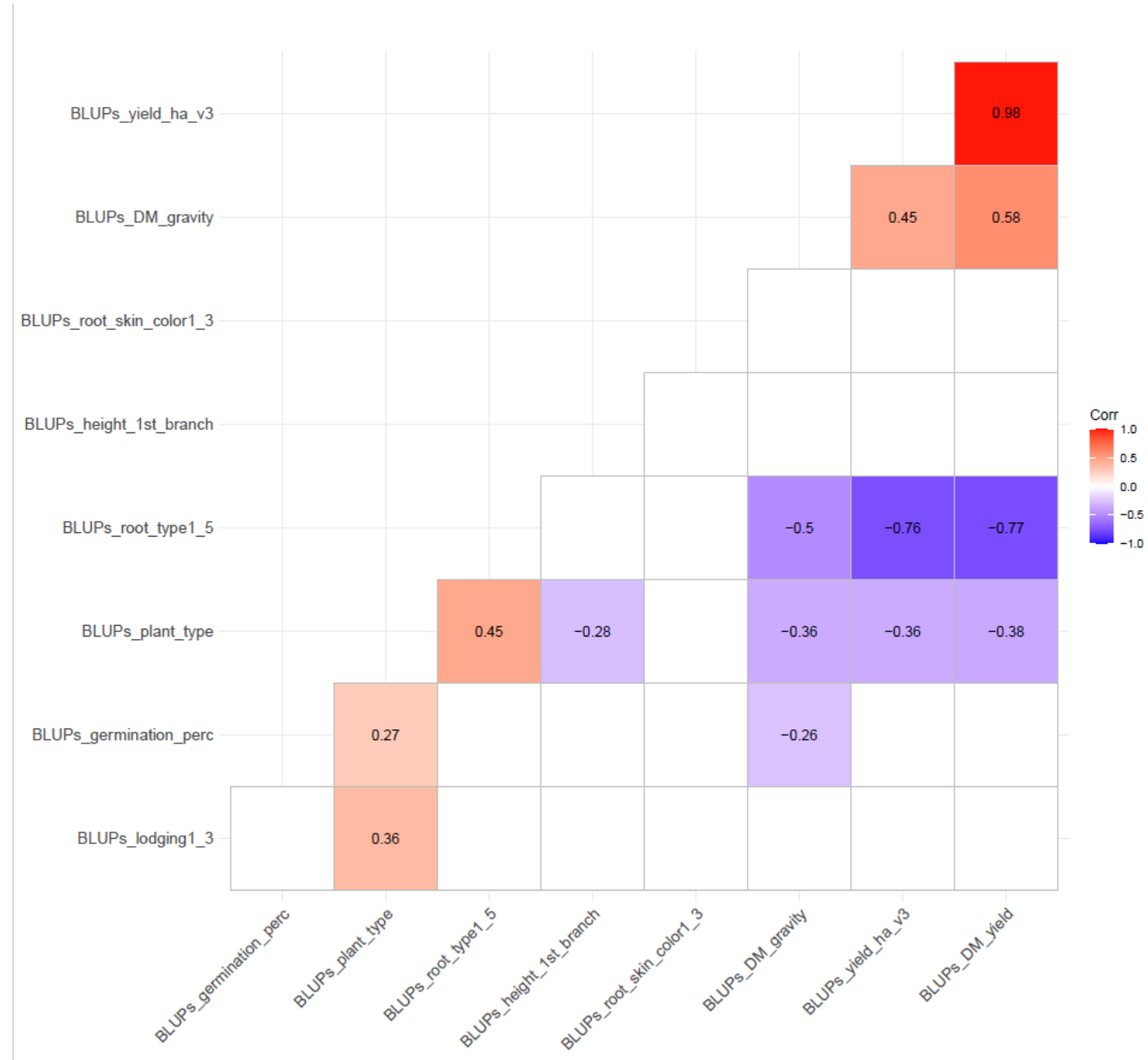
# BLUPs and BLUEs calculation

| genotype             | BLUEs_DM_gravity_202152MDEPR_mon | BLUPs_DM_gravi | BLUEs_height_1st_branch_202152MDEPR_mon | BLUPs_height_1st_bran | BLUEs_lodging1_3_202152MDEPR_mon | BLUPs_lodging1 |
|----------------------|----------------------------------|----------------|---|-----------------------|----------------------------------|----------------|
| SM2775-4_is_Bellotti | 38.63870605                      | 38.19343632    | 109.9790267                             | 118.2416134           | 1.065871057                      | 1.125391404    |
| SMB2446-2_is_Caiseli | 37.7490234                       | 37.56105742    | 204.7053786                             | 190.7546429           | 1.388283868                      | 1.314616352    |
| GM10057-4            | 38.11353771                      | 37.31365851    | 90.64496676                             | 114.9620107           | 1.040414082                      | 1.184555676    |
| GM11919-1            | 37.83976495                      | 36.93832357    | 106.8114559                             | 124.8877244           | 1.481196615                      | 1.468986055    |
| AM1584-10            | 37.22007321                      | 36.63290128    | 187.8587352                             | 167.9084304           | 1.858858674                      | 1.742600812    |
| GM11891-2            | 36.77861673                      | 36.37046251    | 106.8376822                             | 126.6911167           | 2.203528831                      | 1.745780123    |
| CM4919-1_is_Veronica | 37.40253366                      | 36.20242004    | 296.6663949                             | 238.2675521           | 1.290119342                      | 1.203392798    |
| GM11935-2            | 36.22448794                      | 35.86873554    | 45.76290013                             | 87.34264962           | 1.556713445                      | 1.487989535    |
| GM11933-1            | 35.81802775                      | 35.36994682    | 98.18489511                             | 118.3231004           | 2.007428038                      | 1.760577835    |
| TAI8_is_TAI          | 35.72997461                      | 35.28480704    | 143.4750958                             | 143.2455477           | 1.094675962                      | 1.196720581    |
| GM10059-3            | 35.67467228                      | 35.12405304    | 136.9114752                             | 136.1513527           | 1.17275983                       | 1.46853712     |
| GM10057-8            | 35.30425834                      | 35.10323216    | 54.01987574                             | 94.89147286           | 1.605283079                      | 1.484826466    |
| GM11970-4            | 35.38143447                      | 35.099533      | 97.44000642                             | 120.0150882           | 1.049055779                      | 1.170504292    |
| CG1141-1_is_Costena  | 34.72456813                      | 34.87129947    | 198.4793708                             | 193.4317351           | 0.936194084                      | 1.114394676    |
| GM10057-12           | 34.98010918                      | 34.81387413    | 222.9565375                             | 209.3533544           | 1.458384231                      | 1.467058635    |

# Estimates: surrogates of true value

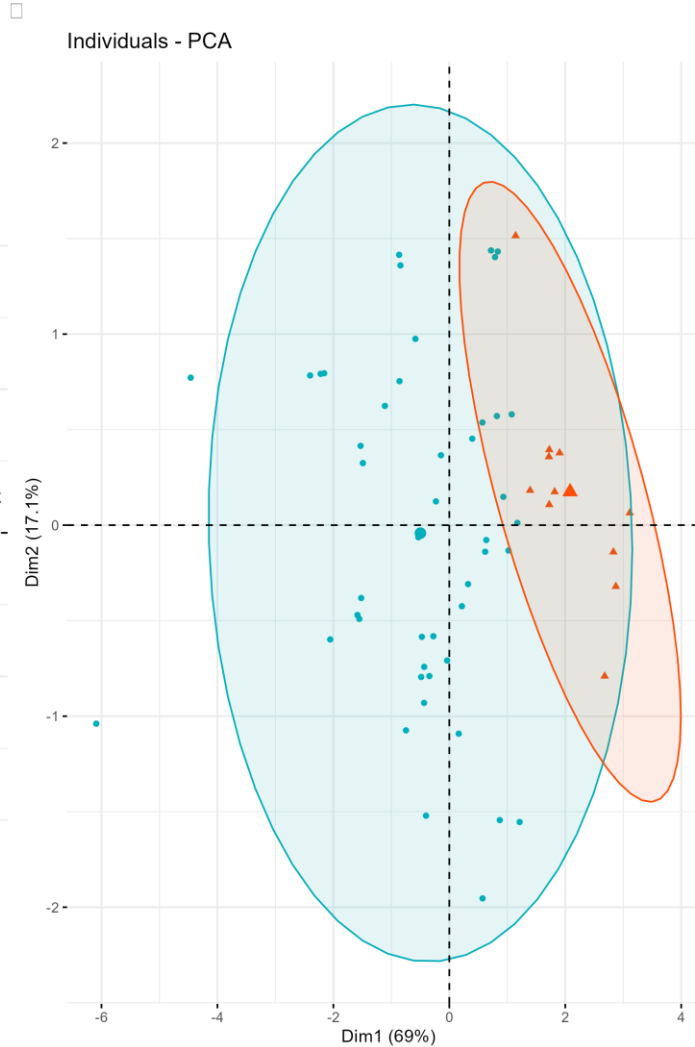
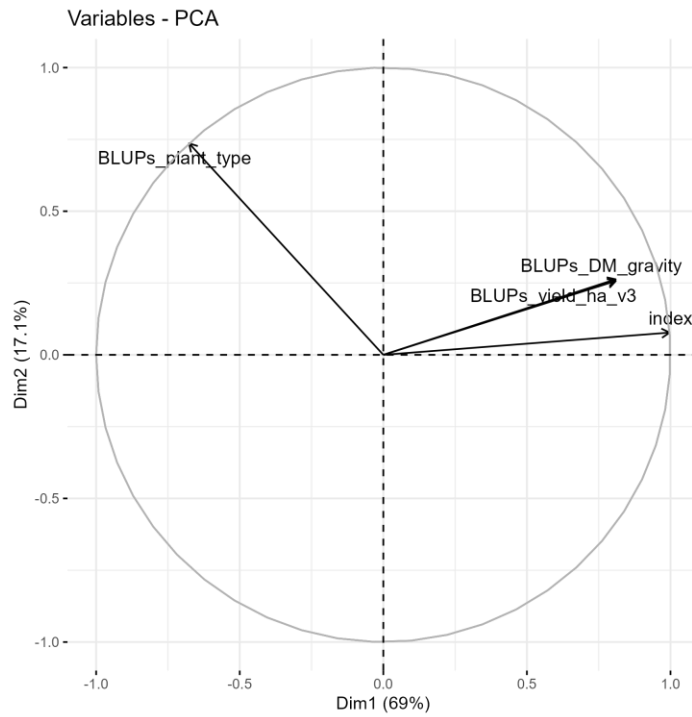
| Phenotype   | Phenotypic mean   | BLUE  | BLUP   | PBLUP  | GBLUP |   |   |   |   |      |   |     |   |  |  |   |   |   |   |      |   |      |   |
|---|---|---|--|--|-------|---|---|---|---|------|---|-----|---|--|--|---|---|---|---|------|---|------|---|
| <p>The phenotypic value of a single observation (e.g. plant or plot) is a surrogate of the observation's true genetic value.</p> <p>However, the phenotype deviates from the true genetic value due to non-genetic effects like environment and random chance. How much phenotypes deviate from the true genetic value can be estimated by measuring heritability.</p> <p>If phenotypes have high heritabilities, they are relatively accurate estimates of true value. If not, they are relatively inaccurate, and other surrogates can increase accuracy.</p> | <p>Phenotypic means of multiple observations are generally more accurate than single phenotypic values.</p> <p>This is because the error component of phenotypes is random, so the mean of the phenotypes is closer to the true genetic value than a single phenotype.</p> <p>Phenotypic means derived from phenotypes with high heritabilities still have higher accuracies than phenotypic means derived from phenotypes with low heritabilities.</p> <p>The increase in accuracy from using the phenotypic mean vs. phenotype is higher if the phenotypes have low heritabilities.</p> | <p>Best linear unbiased estimates can be obtained from linear regression of phenotypes on fixed effects, e.g.:</p> $\text{Pheno} \sim \text{mean} + \text{block} + \text{genotype} + \text{error}$ <p>BLUEs allow more complex evaluation strategies which can increase accuracy (e.g. blocking) and accounting for non-genetic effects.</p> <p>In absence of non-genetic effects besides error and with equal replication, BLUEs of genetic value are equal to phenotypic means.</p> <p>BLUEs can increase accuracy compared to using phenotypic means indirectly by enabling use of experimental designs which increase accuracy.</p> | <p>Best linear unbiased predictions can also be obtained from linear regression of phenotypes, but the regression is on random effects or a mix of fixed and random effects, e.g.:</p> $\text{Pheno} \sim \text{mean} + \text{block} + \text{genotype} + \text{error}$ <p>The random genotype effects, which estimate genetic value, are shrunk to the mean by an estimate of heritability for the genotype.</p> <p>Genotypes with fewer replications (as in unbalanced data) have lower heritabilities, so estimates of their genetic value are shrunk more to anticipate their regression to the mean. This can increase accuracy.</p> | <p>Pedigree best linear unbiased predictions are similar to BLUPs, but the random genetic effects are not assumed to be independent (since genotypes are related):</p> $\text{Pheno} \sim \text{mean} + \text{block} + \text{genotype} + \text{error}$ <table border="1" data-bbox="1589 686 1839 872"> <thead> <tr> <th></th> <th>A</th> <th>B</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>1</td> <td>0.50</td> </tr> <tr> <th>B</th> <td>0.5</td> <td>1</td> </tr> </tbody> </table> <p>The genetic effects are not only shrunk by heritability, but also weighted by the pedigree expectation of their relatedness to other observations. This borrowing of information from relatives can increase accuracy in some scenarios.</p> |       | A | B | A | 1 | 0.50 | B | 0.5 | 1 | <p>Genomic best linear unbiased predictions are quite similar to PBLUPs, except that genotype relatedness is estimated from markers rather than pedigree expectations.</p> $\text{Pheno} \sim \text{mean} + \text{block} + \text{genotype} + \text{error}$ <table border="1" data-bbox="1913 719 2165 905"> <thead> <tr> <th></th> <th>A</th> <th>B</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>1</td> <td>0.54</td> </tr> <tr> <th>B</th> <td>0.54</td> <td>1</td> </tr> </tbody> </table> <p>Genotype relatedness deviates from pedigree expectations because of recombination (Mendelian sampling), so accounting for this deviation can increase accuracy compared to PBLUP in some scenarios.</p> |  | A | B | A | 1 | 0.54 | B | 0.54 | 1 |
|   | A   | B   |  |  |       |   |   |   |   |      |   |     |   |  |  |   |   |   |   |      |   |      |   |
| A   | 1   | 0.50  |  |  |       |   |   |   |   |      |   |     |   |  |  |   |   |   |   |      |   |      |   |
| B   | 0.5   | 1   |  |  |       |   |   |   |   |      |   |     |   |  |  |   |   |   |   |      |   |      |   |
|   | A   | B   |  |  |       |   |   |   |   |      |   |     |   |  |  |   |   |   |   |      |   |      |   |
| A   | 1   | 0.54  |  |  |       |   |   |   |   |      |   |     |   |  |  |   |   |   |   |      |   |      |   |
| B   | 0.54  | 1   |  |  |       |   |   |   |   |      |   |     |   |  |  |   |   |   |   |      |   |      |   |

# Correlation among traits



# Selection index

Selection: 20%  
Weights: (-5, 10, 10)



Groups ● FALSE ▲ TRUE



# Suggested selection

| rowname              | index        | select | BLUPs_DM_gravi | BLUPs_height_1st_bran | BLUPs_lodging1 | BLUPs_plant_ty | BLUPs_root_skin_color1 | BLUPs_root_type1 | seBLUPs_root_type1 |
|----------------------|--------------|--------|----------------|-----------------------|----------------|----------------|------------------------|------------------|--------------------|
| SM2775-4_is_Bellotti | 36.43849604  | TRUE   | 38.19343632    | 118.2416134           | 1.125391404    | 2.252729678    | 2.969910764            | 1.580064617      | 0.279808625        |
| SMB2446-2_is_Caiseli | 32.9321129   | TRUE   | 37.56105742    | 190.7546429           | 1.314616352    | 2.137294633    | 2.989842575            | 1.878556656      | 0.279917429        |
| GM10057-4            | 32.76929839  | TRUE   | 37.31365851    | 114.9620107           | 1.184555676    | 2.221779431    | 2.968111408            | 2.899475684      | 0.363052442        |
| GM11919-1            | 13.61182414  | FALSE  | 36.93832357    | 124.8877244           | 1.468986055    | 2.860122935    | 2.012360149            | 2.277752173      | 0.362910532        |
| AM1584-10            | 11.14320002  | FALSE  | 36.63290128    | 167.9084304           | 1.742600812    | 2.706052313    | 2.048296667            | 2.97368316       | 0.362796527        |
| GM11891-2            | 11.7597177   | FALSE  | 36.37046251    | 126.6911167           | 1.745780123    | 3.262600168    | 2.991141475            | 2.522501916      | 0.362703387        |
| CM4919-1_is_Veronica | 29.83616766  | TRUE   | 36.20242004    | 238.2675521           | 1.203392798    | 1.978019799    | 2.921453909            | 1.776274693      | 0.362869767        |
| GM11935-2            | 11.01720769  | FALSE  | 35.86873554    | 87.34264962           | 1.487989535    | 3.291153729    | 3.010304738            | 2.483923309      | 0.363052875        |
| GM11933-1            | 16.11084252  | TRUE   | 35.36994682    | 118.3231004           | 1.760577835    | 3.243303148    | 2.976938085            | 2.537085138      | 0.363052442        |
| TAI8_is_TAI          | 16.58691342  | TRUE   | 35.28480704    | 143.2455477           | 1.196720581    | 2.634418354    | 1.051647379            | 2.195740101      | 0.362796602        |
| GM10059-3            | 13.72201809  | FALSE  | 35.12405304    | 136.1513527           | 1.46853712     | 2.604697171    | 1.504207284            | 1.835677777      | 0.362845146        |
| GM10057-8            | -2.473923583 | FALSE  | 35.10323216    | 94.89147286           | 1.484826466    | 2.919896933    | 2.95259727             | 2.55813067       | 0.3627966          |
| GM11970-4            | 21.52640654  | TRUE   | 35.099533      | 120.0150882           | 1.170504292    | 2.550603976    | 2.957545772            | 2.035839533      | 0.363003563        |
| CG1141-1_is_Costena  | 20.31682317  | TRUE   | 34.87129947    | 193.4317351           | 1.114394676    | 2.540353993    | 3.003638779            | 2.266518934      | 0.284138308        |
| GM10057-12           | -6.420457149 | FALSE  | 34.81387413    | 209.3533544           | 1.467058635    | 2.593976977    | 2.946526274            | 3.266652786      | 0.362445534        |

**DataBase**



**Raw data**



**Tidy data**



**Clean data**



**Results**



**Report**

**Download**

**Curation**

**Clean**

**Analysis**

**Visualization**

## Tools

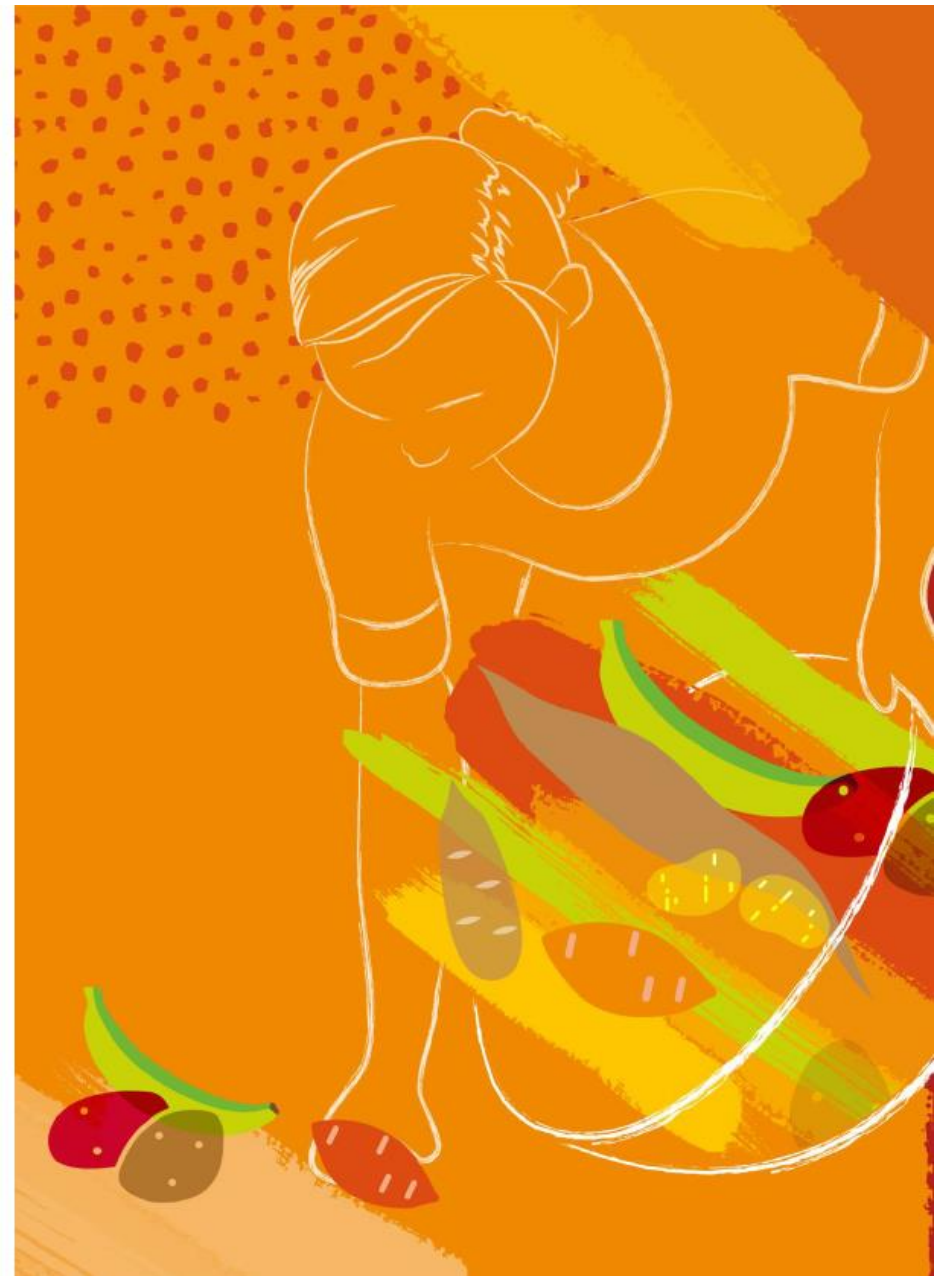
Manually, QBMS, MrBean

Manually, R script

Manually, R script

MrBean, BreedBase, R script

MrBean, ViTSel, BreedBase,  
Spotfire, R script



# 6. Challenges and improvement



- Non replicate experimental design data analysis (p-rep, augmented)
- Curation of data from 1980 2009 in CassavaBase (pipeline design)
- Start to do the analysis including the GBLUP estimate from markers (accuracy increases)
- Do the entire analysis with ASReml - R library
- Development of analysis pipelines for genome-wide association and genome-wide prediction.
- Keep growing with data science for plant breeding and

# Gratefulness

- Johan Aparicio from beans programm (codes, libraries and Mr.bean shiny app)
- Xiaofei Zhang (advices, suggestions and economical support for trainings)
- Sandra, Nelson, Jorge Ivan and Camilo (how to take the data on field)
- Lizbeth (data management)
- Thierry and Luis Fernando Londoño (allow me to work in lab during almost 2 years)



Alliance



Thank you!!!

