



Data analysis pipeline in Cassava breeding.

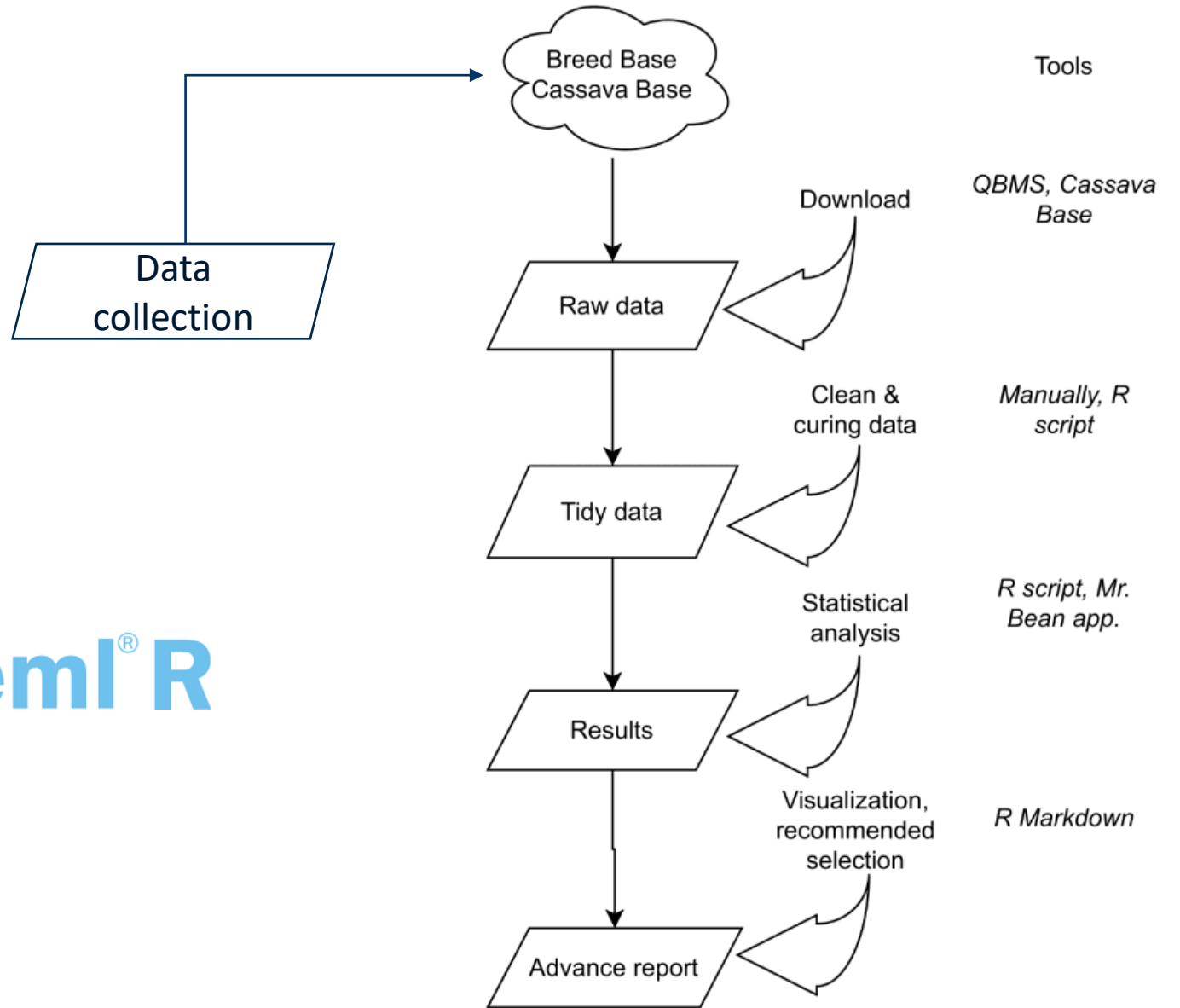
Luis Fernando Delgado Munoz
luis.delgado@cgiar.org
Palmira, September 2022



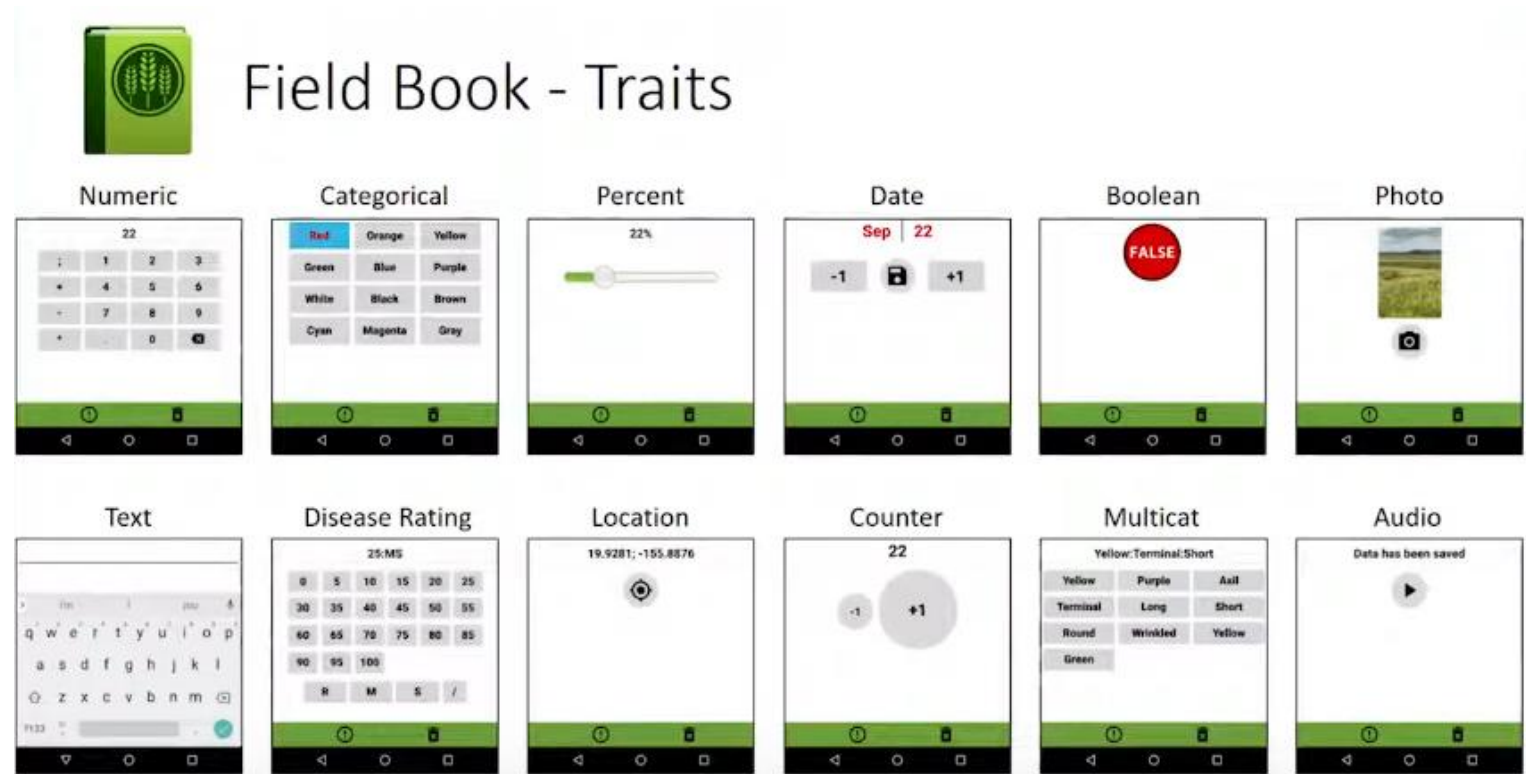
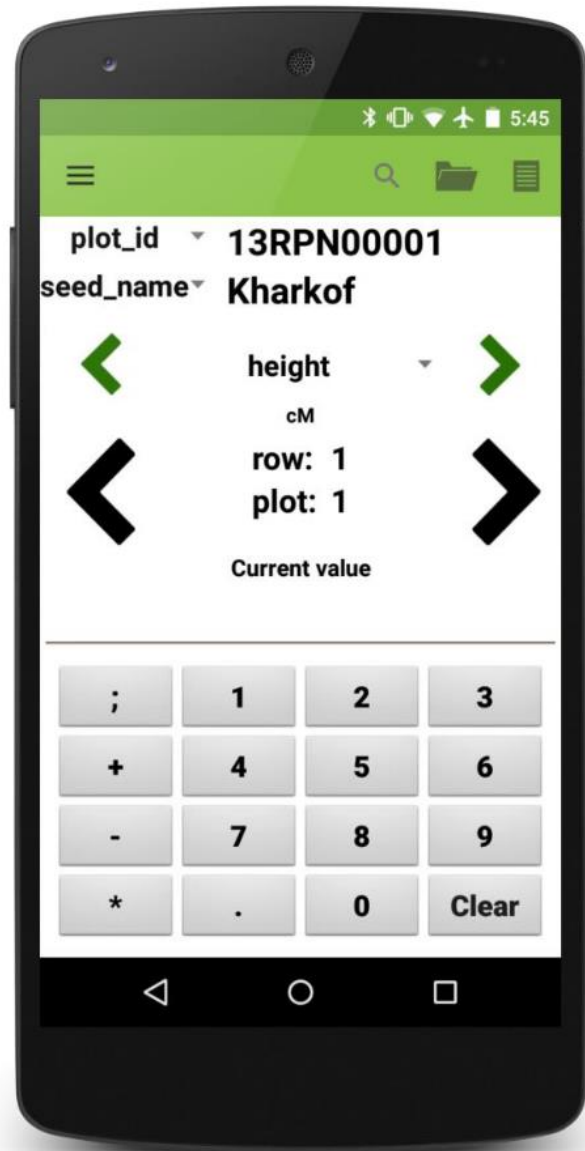
@BiovIntCIAT_eng
@BiovIntCIAT_esp

#Alliance4Science

Flow diagram



Data collection using Field book app



Data storage using Cassava Base

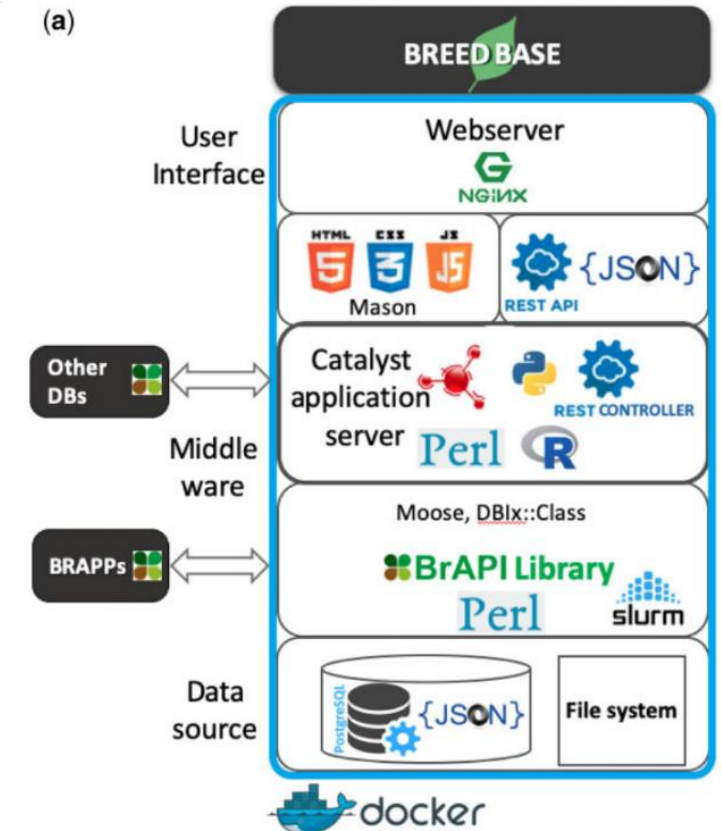


Search Wizard

Don't see your data? Refresh Lists Update Wizard

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

(a)



Source: Morales, et al 2002

Download data from Cassava Base

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](#), [BreedBase](#), and [GIGWA](#) (using [BrAPI](#) calls) and help them to retrieve phenotypic and genotypic data directly into their analyzing pipelines developed in R statistical environment.

Querying the Breeding Management System(s) like [BMS](#), [BreedBase](#), and [GIGWA](#)

Historical Data from Cassava Base

The Cassava program from CIAT has conducted several trials over its history; the following code will show you the way to download the historic data.

Conection with Cassava BreedBase server

```
set_qbms_config("https://cassavabase.org/brapi/v1/calls/",  
               path = "", time_out = 300, no_auth = TRUE,  
               page_size = 10000,  
               engine = "breedbase")
```

List supported crops in the current bms server

```
list_crops()
```

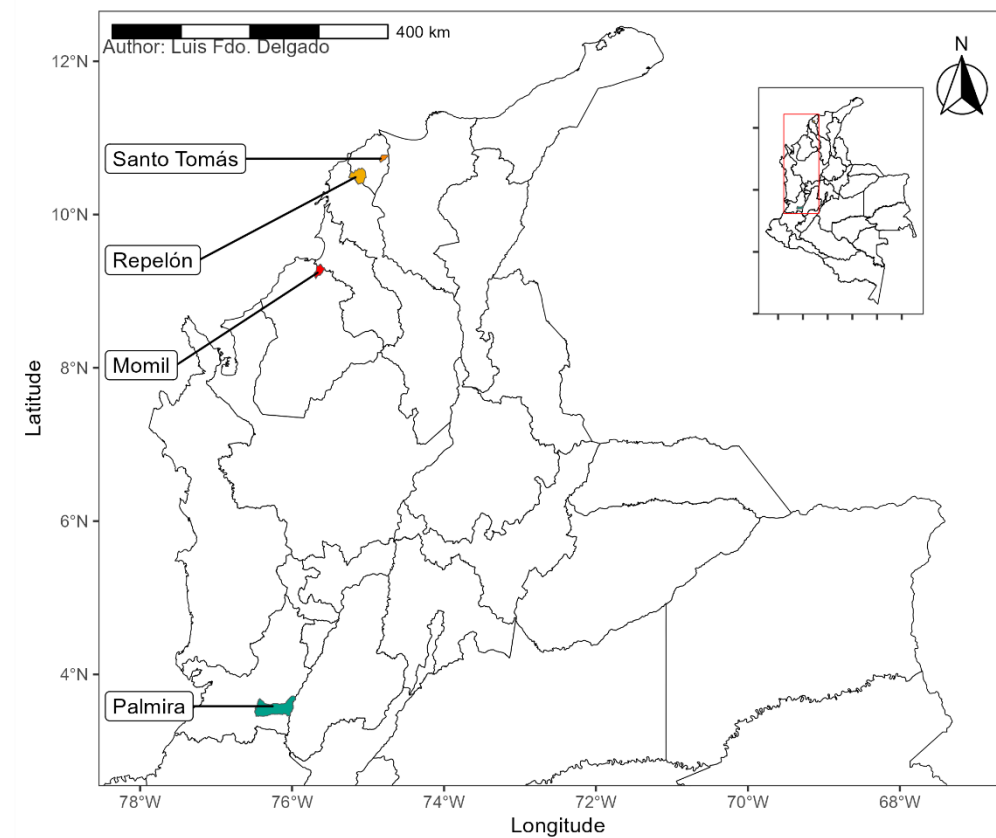
```
[1] "Cassava"
```

Manual: Delgado, 2022 - https://cassava2050.github.io/cassava_base_data/cassava_base_data.html

Exploratory data analysis

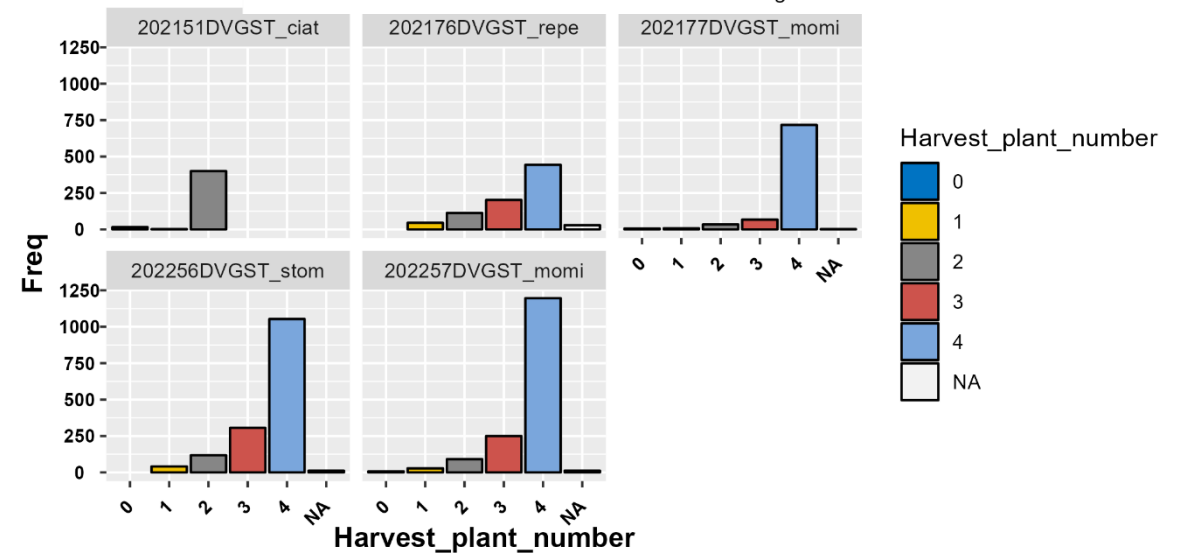
General information about trials conducted

use_trial_name	use_plant_date	use_harvest_date	harvesting_time	use_location	n_gen
202151DVGST_ciat	2021-August-12	2022-June-30	10month 18day	CIAT. Valle, Colombia	407
202176DVGST_repe	2021-July-07	2022-April-04	8month 28day	Repelon. Atlantico, Colombia	411
202177DVGST_momi	2021-July-03	2022-March-30	8month 27day	Momil. Cordoba, Colombia	410
202256DVGST_stom	2022-July-15	2023-April-03	8month 19day	Santo Tomas. Atlantico, Colombia	758
202257DVGST_momi	2022-July-22	2023-April-17	8month 26day	Momil. Cordoba, Colombia	787



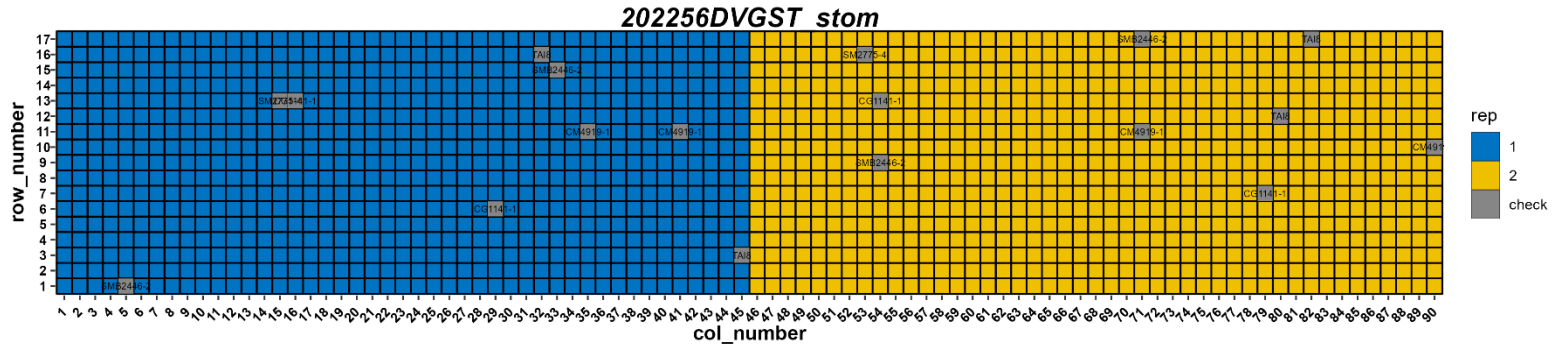
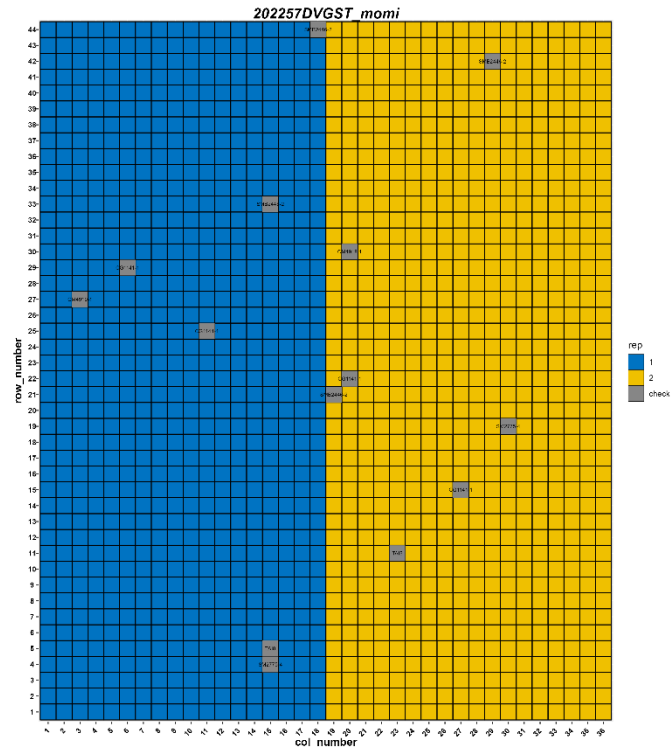
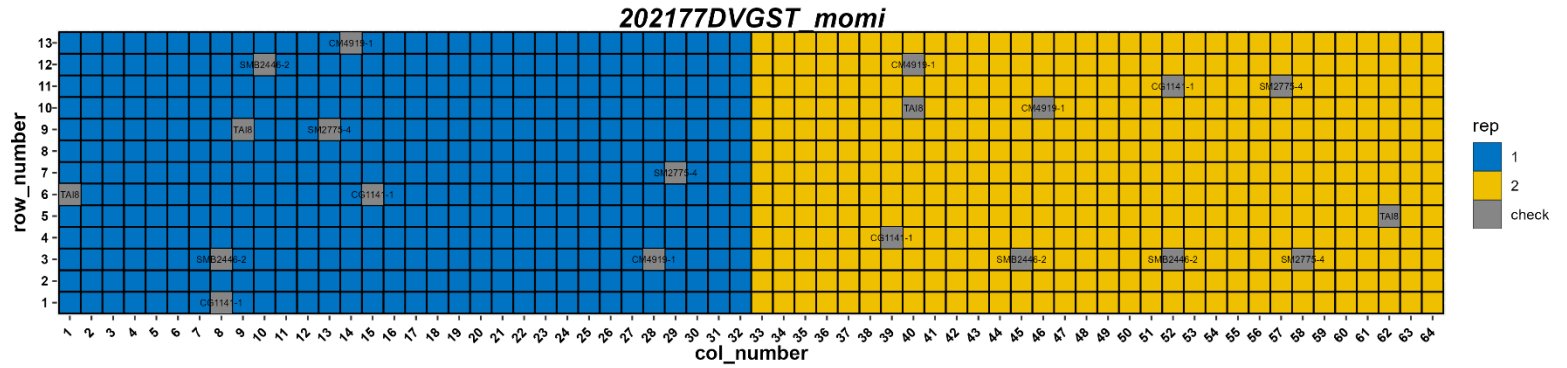
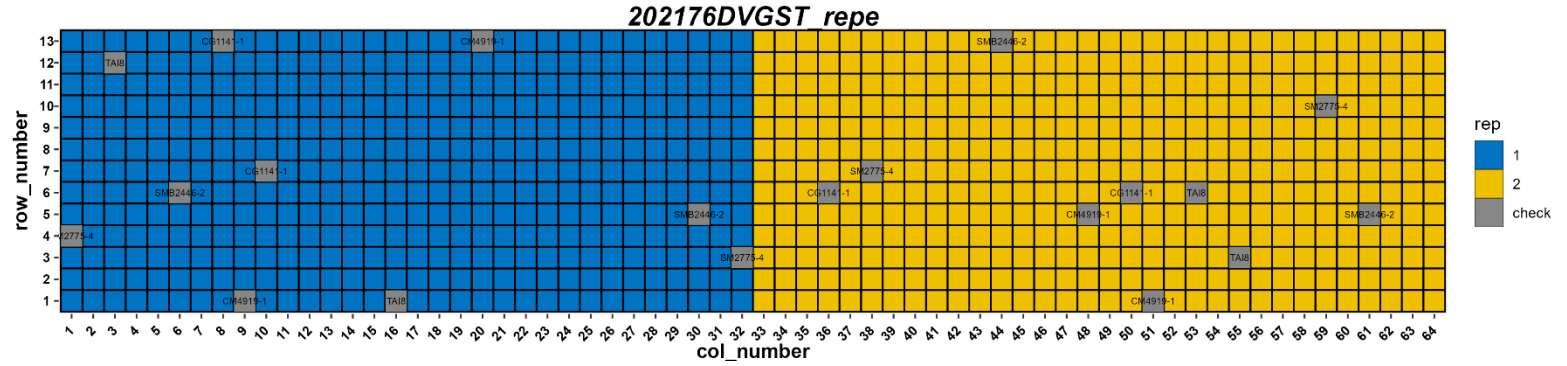
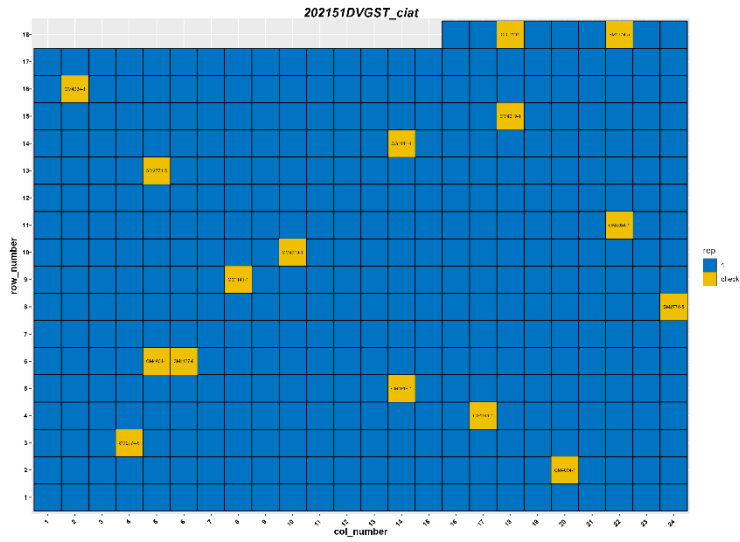
Plot size:

- 202151DVGST_ciat, **4** plants per plot, harvesting **2** plants for **root_type**
- 202176DVGST_repe, **4** plants per plot, harvesting **4** plants for **yield**
- 202177DVGST_momi, **4** plants per plot, harvesting **4** plants for **yield**
- 202256DVGST_stom, **5** plants per plot, harvesting **4** plants for **yield**
- 202257DVGST_momi, **5** plants per plot, harvesting **4** plants for **yield**



Trial Design – Row col

Check varieties are completely randomized



Check genotypes

Released varieties

1. CG1141-1_is_Costena
2. CM4919-1_is_Veronica
3. COL2215_is_Venezolana
4. HMC1P12_is_HMC1
5. SM1127-8_is_Cubana
6. SM2775-4_is_Bellotti
7. SMB2446-2_is_Caiseli
8. TAI8_is_TAI
9. GM4034-1_is_Reina_wx
10. CM6438-14_is_Vergara

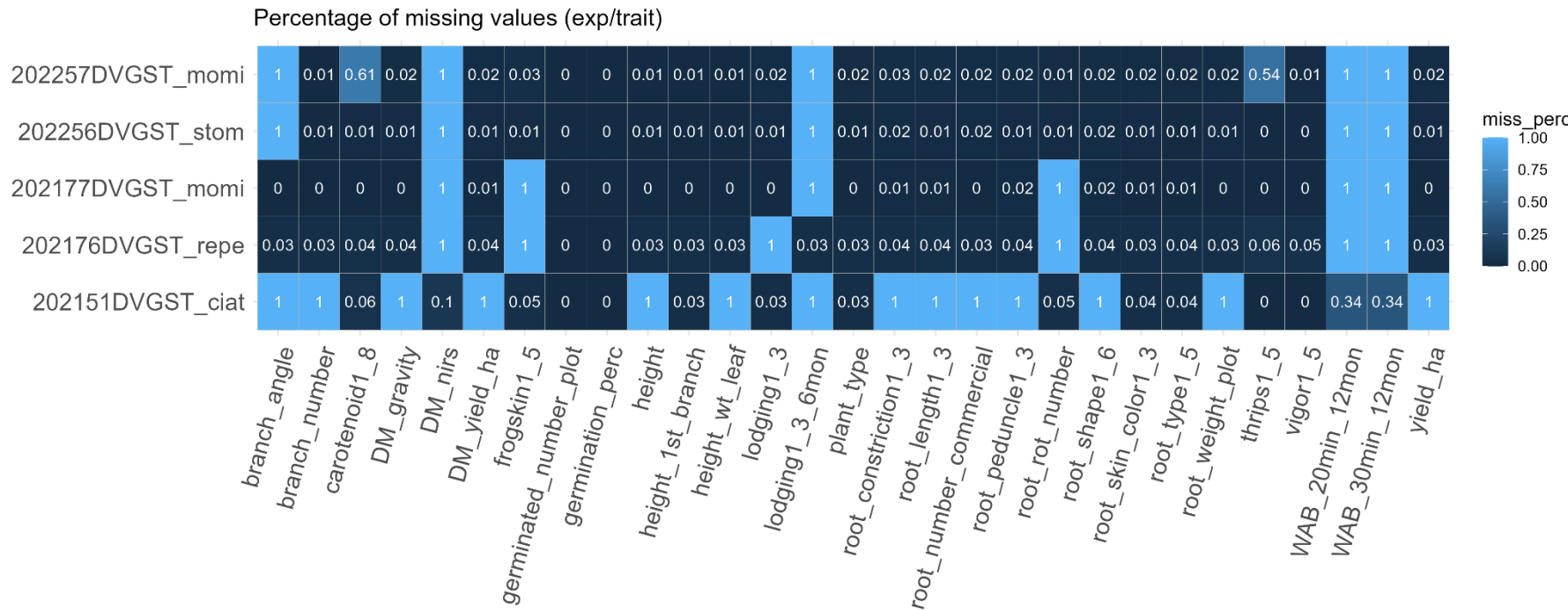


3 - Unloading cassava. Women processors, Korogho Côte d'Ivoire. Patricio Mendez del Villar / 3 - Déchargement de manioc Femmes transformatrices, Korogho Cote d'Ivoire. Patricio Mendez del Villar

Traits evaluated

Assessed agronomic traits

WAB_20min_12mon
WAB_30min_12mon
branch_angle
branch_number
DM_gravity
DM_nirs
height_1st_branch
yield_ha
germination_perc
DM_yield_ha
root_weight_plot
frogskin1_5
vigor1_5
lodging1_3
lodging1_3_6mon
root_number_commercial
plant_type
height
height_wt_leaf
root_skin_color1_3
root_constriction1_3
root_type1_5
root_length1_3
root_rot_number
germinated_number_plot
root_peduncle1_3
root_shape1_6
thrips1_5
carotenoid1_8



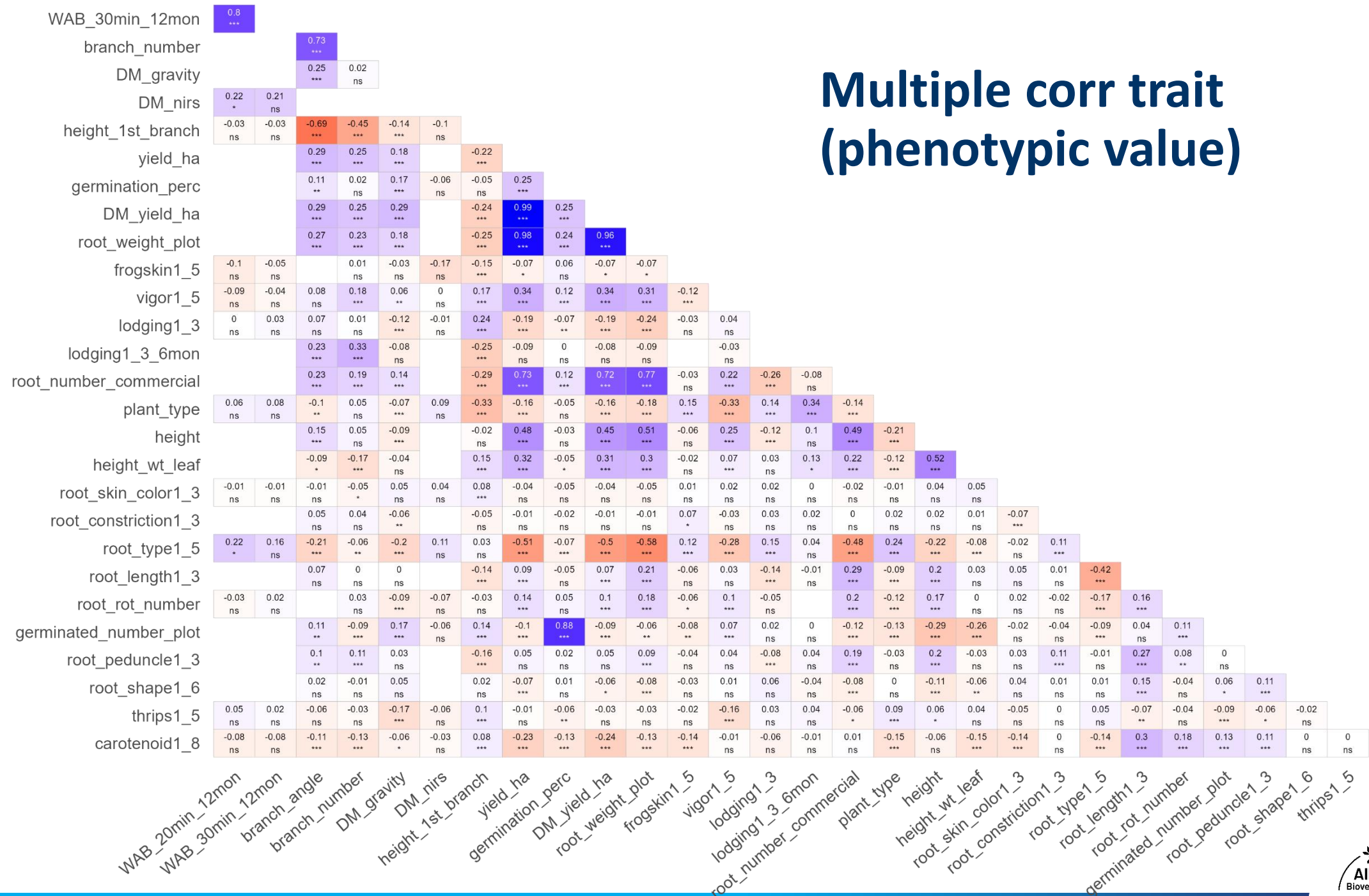
Shared genotypes

2 common genotype groups found.

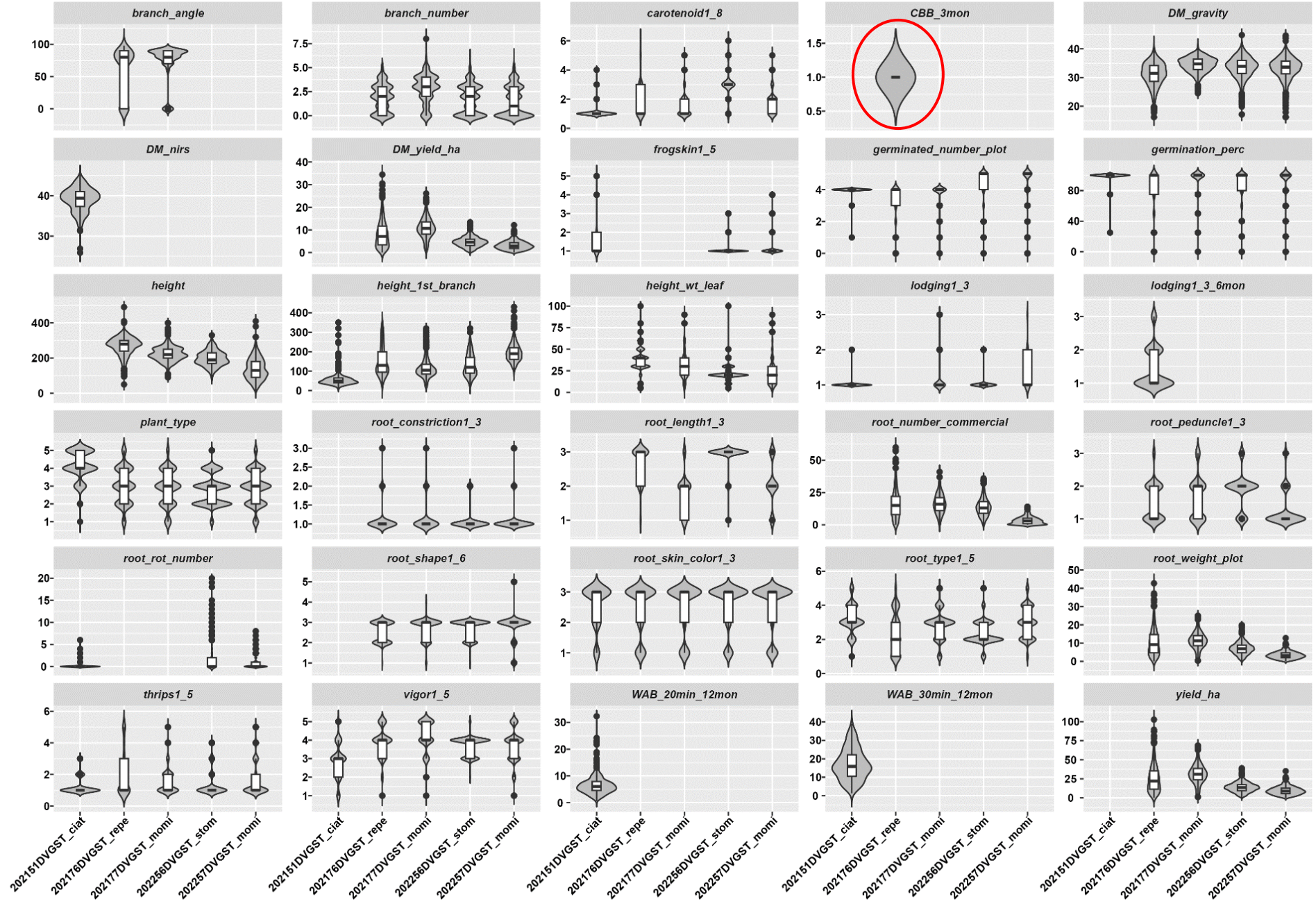
Connectivity Matrix

202257DVGST_momi	17	20	20	753	787
202256DVGST_stom	17	20	20	758	753
202177DVGST_momi	400	410	410	20	20
202176DVGST_repe	400	411	410	20	20
202151DVGST_ciat	407	400	400	17	17
	202151DVGST_ciat	202176DVGST_repe	202177DVGST_momi	202256DVGST_stom	202257DVGST_momi

Multiple corr trait (phenotypic value)



Trait variation across trials (phenotypic value)



Statistical data analysis

Dropped traits due to no variation across trials

CBB_3mon } These traits were dropped from the analysis because they had no variation within themselves.

5. Severity of Cassava Bacterial Blight CBB - CBBSev_VisScg_1to5_T3m (CO_334:0000175)

Cassava Base Ontology

i) Definition

The symptoms of CBB are caused for *Xanthomonas*



1 = Healthy plant



2 = Points on the leaves, angular spots



3 = Leaf wilting and defoliation



4 = Exudation on stems



5 = Descending death of the plant



Angular spots

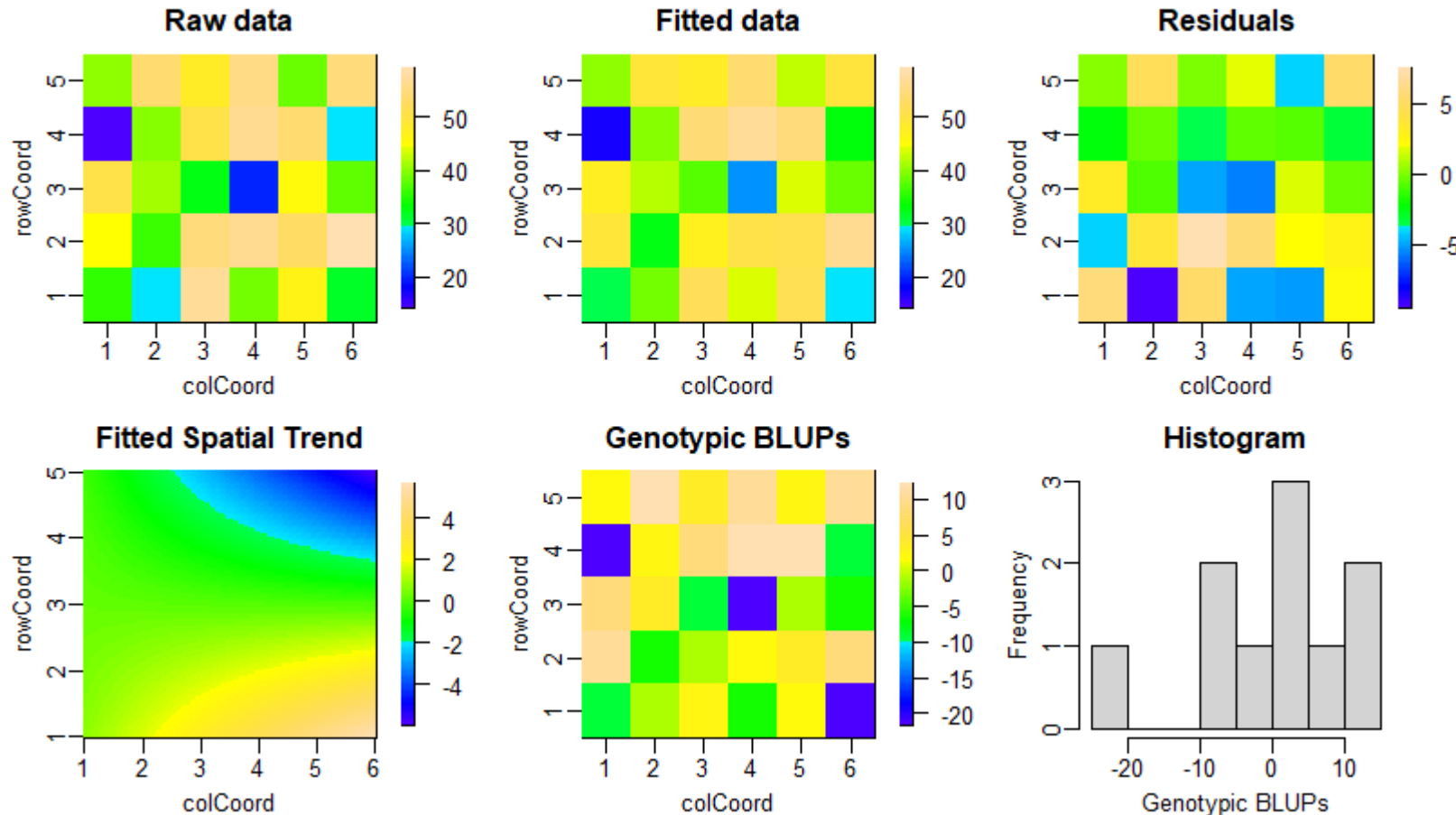


Source: Salazar 2022 - <https://cgspace.cgiar.org/bitstream/handle/10568/129729/presentation.pdf?sequence=1&isAllowed=y>

Single trial analysis – Spatial analysis



Trait: yield_ha - Trial: 202071CQPEA_cpai



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

Heritability calculus

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

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

σ_g^2 = genotypic variance

σ_ϵ^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

Single heritability

Moderate to high heritability among two locations.

Trial	branch_angle	branch_number	carotenoid1_8	DM_gravity	DM_yield_ha	frogskin1_5	germinated_number_plot	germination_perc	height	height_1st_branch	height_wl_t_leaf	lodging1_3	lodging1_3_6mon	plant_type	root_constriction1_3	root_length1_3	root_number_commercial	root_peduncle1_3	root_rot_number	root_shape1_6	root_skin_color1_3	root_tuber1_5	root_weight_plot	thrips1_5	vigor1_5	yield_ha
202176DVGST_repe	0.82	0.87	0.61	0.74	0.72		0.69	0.69	0.86	0.82	0.58		0.57	0.72	0.25	0.42	0.62	0.41		0.35	0.96	0.64	0.72	0.1	0.07	0.72
202177DVGST_momi	0.83	0.77	0.58	0.86	0.83		0.54	0.54	0.8	0.85	0.58	0.5		0.63	0.51	0.54	0.67	0.76		0.67	0.96	0.69	0.82	0.49	0.53	0.82
202256DVGST_stom		0.87	0.67	0.82	0.71	0.81	0.68	0.68	0.88	0.86	0.53	0.37		0.66	0.04	0.42	0.07	0.61	0.36	0.03	0.96	0.62	0.71	0.88	0.52	0.72
202257DVGST_momi		0.79		0.71	0.69	0.3	0.69	0.69	0.63	0.73	0.52	0.01		0.01	0.04	0.28	0.49	0.16	0.15	0.19	0.95	0.58	0.7		0.58	0.7

Traits with heritability lower than 0.1 were removed from the GxE analysis.

GxE analysis was performed without following traits - trials:

- **lodging1_3**: 202257DVGST_momi
- **plant_type**: 202257DVGST_momi
- **lodging1_3_6mon**: (only one trait to fit).
- **root_constriction1_3**: 202257DVGST_momi & 202256DVGST_stom
- **root_number_commercial**: 202256DVGST_stom
- **root_shape1_6**: 202256DVGST_stom
- **vigor1_5**: 202176DVGST_repe

Heritability GxE across trials

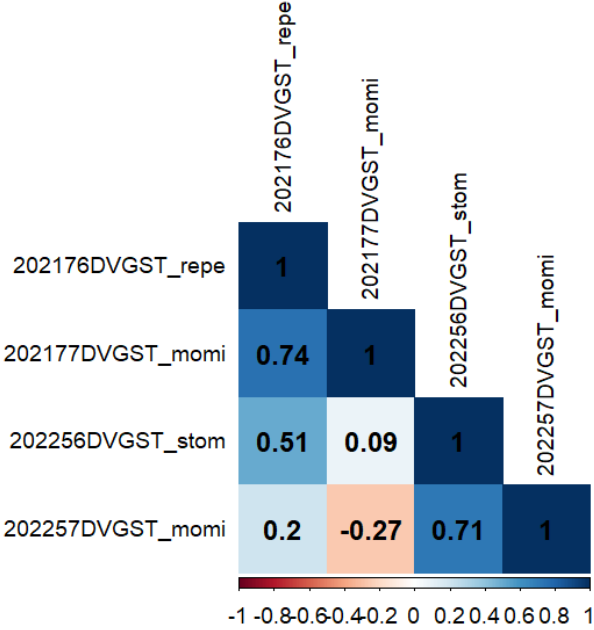
trait	h2
root_skin_color1_3	0.97
branch_angle	0.84
height	0.78
DM_gravity	0.75
germinated_number_plot	0.73
thrips1_5	0.73
germination_perc	0.72
frogskin1_5	0.71
branch_number	0.65
carotenoid1_8	0.61
height_1st_branch	0.59
root_peduncle1_3	0.51
vigor1_5	0.47
height_wt_leaf	0.44
plant_type	0.43
root_length1_3	0.4
lodging1_3	0.26
DM_yield_ha	0.17
root_type1_5	0.17
root_weight_plot	0.15
yield_ha	0.12
root_shape1_6	0
root_constriction1_3	0
root_number_commercial	0
root_rot_number	0

For traits with low heritability, please be **careful**. (blue square)

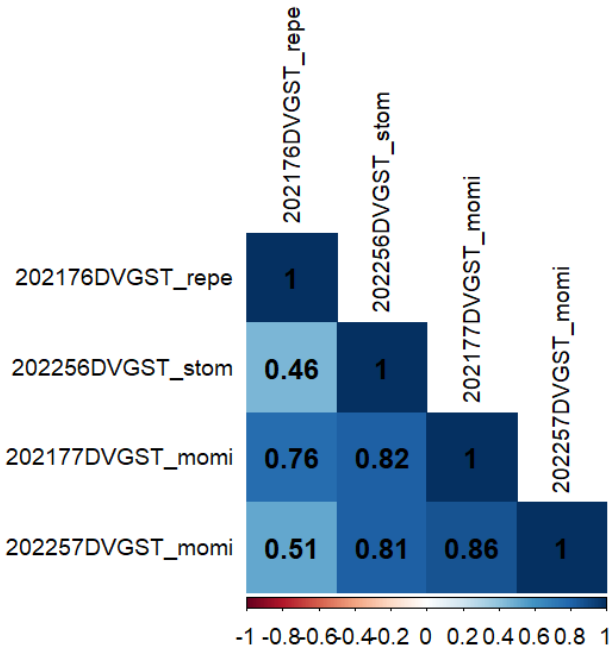
When I fix thrips1_5 data, single heritability increase, but GxE heritability decrease.

Genotypic Correlation: Locations

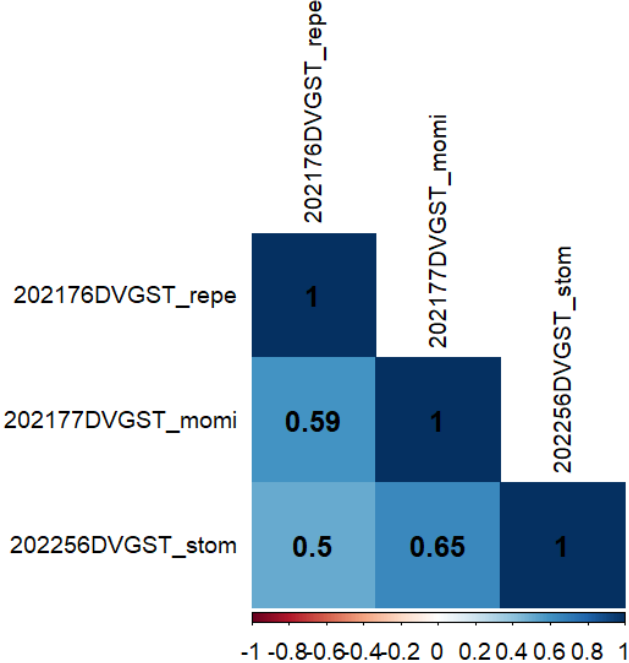
Yield



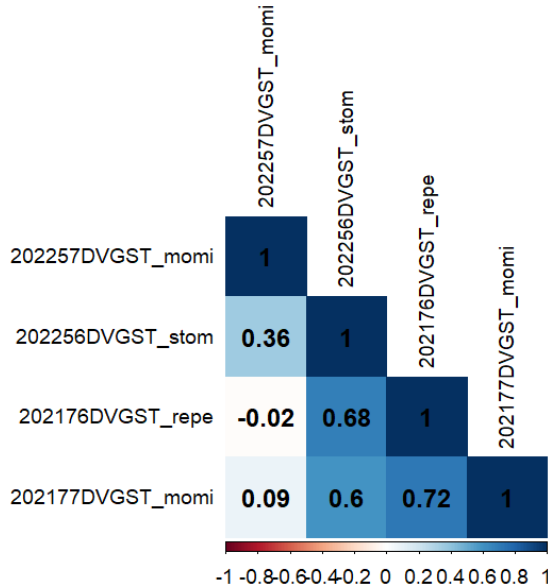
Dry Matter



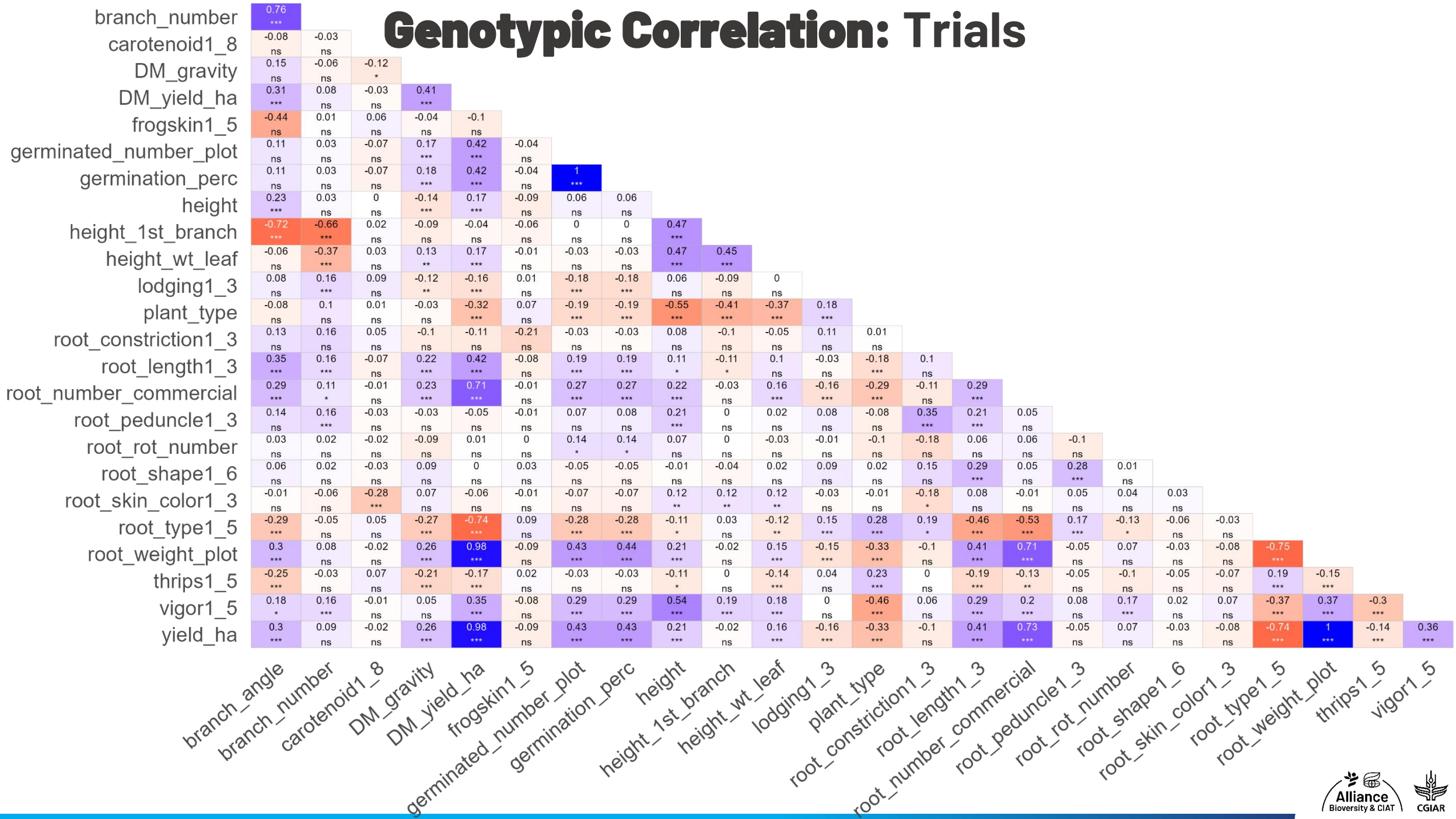
Plant type



height_1st_branch

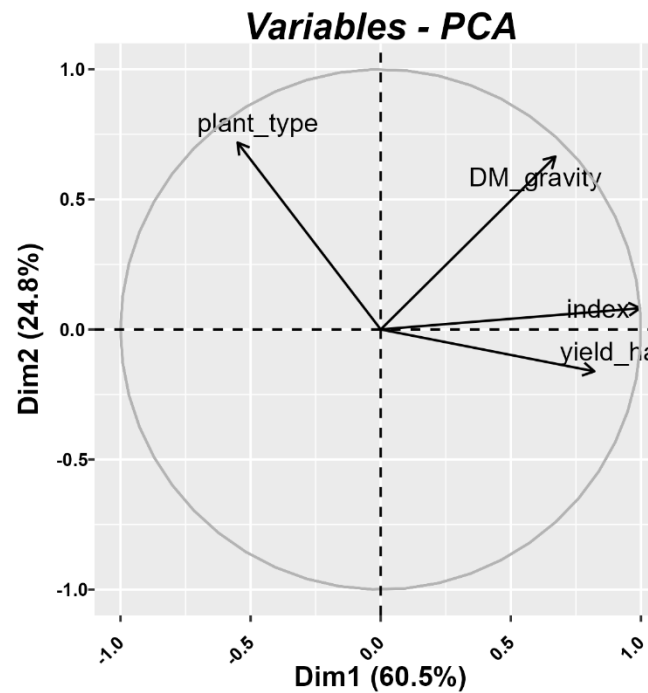


Genotypic Correlation: Trials



Index selection

- 1) **DM_gravity**, **plant_type** & **yield_ha** were used to calculate index selection.
- 2) Scores given were:
 - DM_gravity = 10
 - plant_type = -5
 - yield_ha = 10
- 3) Percentage to be selected: **20%**



Results storage – Cassava2050 GitHub



Cassava2050

Cassava breeding | Mixed models
Genomic selection | Genomic prediction

Edit profile

11 followers · 14 following

Alliance CIAT Bioversity
km 17 recta Cali - Palmira

Cassava2050 / README.md

Cassava Breeding Program at the Alliance of Bioversity and CIAT (CIAT)

Introduction

This repository contains the code and documentation developed by the Cassava breeding program at the Alliance of Bioversity International and the International Center for Tropical Agriculture (CIAT). The mission of the Cassava Program is to improve the livelihoods of cassava-growing smallholders and their communities, the Cassava Program will co-develop innovative solutions for specific regions, farming systems, and markets that boost farm productivity, enhance consumer benefits and drive sustainability throughout the value chain.

Contributing

We welcome contributions from the community! If you would like to contribute to the Cassava program, please follow these steps:

1. Fork this repository.
2. Create a new branch for your changes.
3. Make your changes and commit them.
4. Submit a pull request to this repository.



Source: <https://github.com/Cassava2050?tab=repositories&q=&type=&language=&sort=name>

Summary

DataBase



Download

Raw data



Curation

Tidy data



Clean

Clean data



Analysis

Results



Visualization

Report

Tools

Manually, QBMS, MrBean

Manually, R script

Manually, R script

MrBean, BreedBase, R script

MrBean, ViTsel, BreedBase, Spotfire, R script



Thank you!!!

