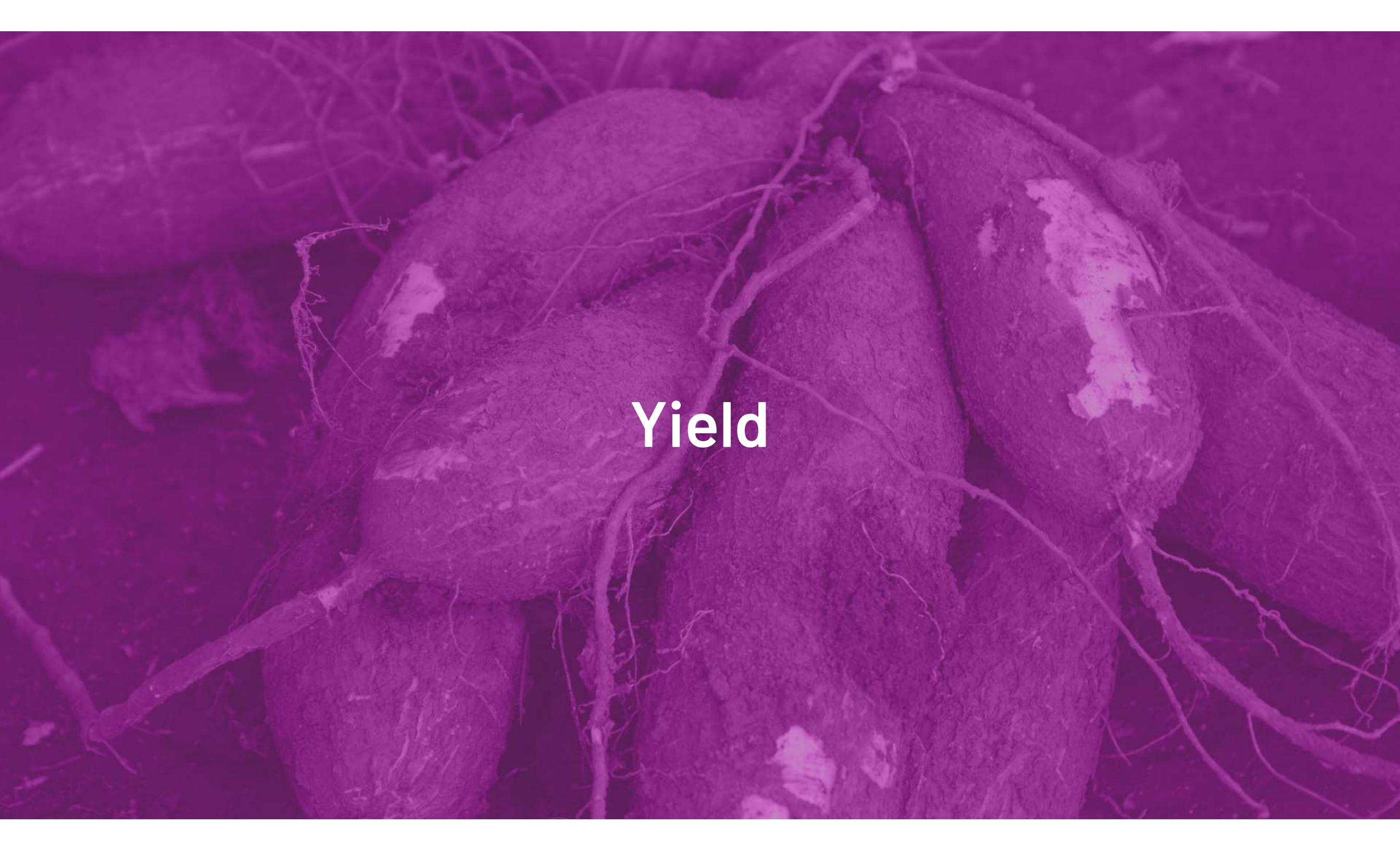




Genetic Gain – Biofortified Cassava

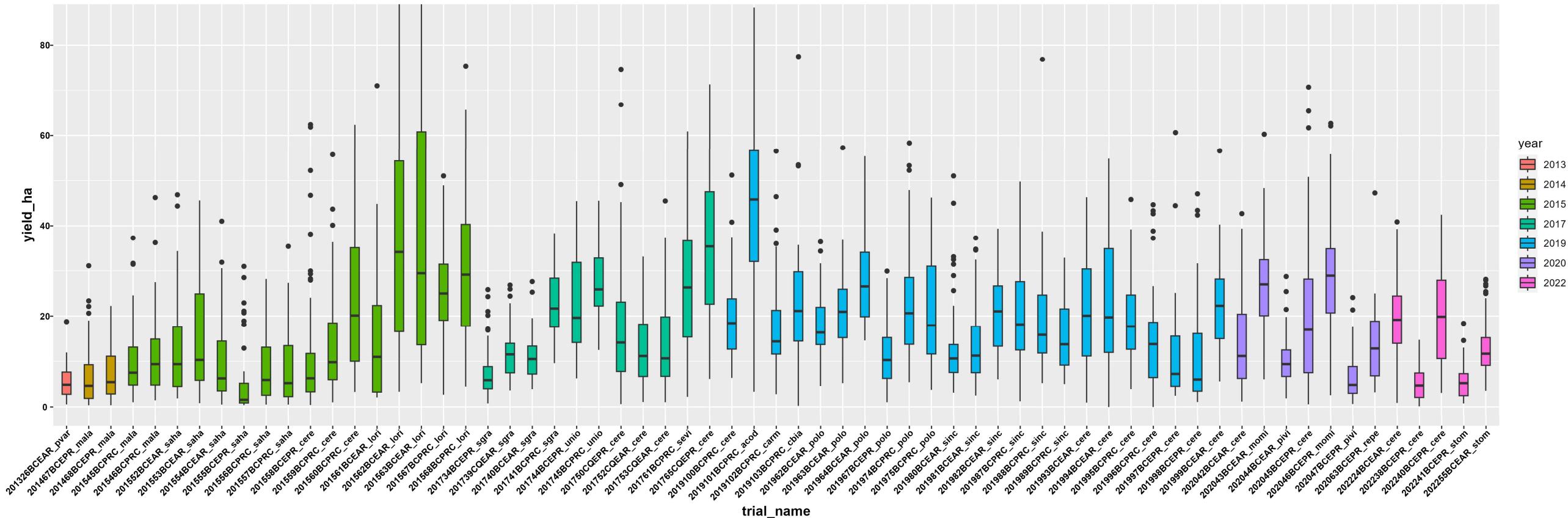
Luis Fernando Delgado
luis.Delgado@cgiar.org

August 2023

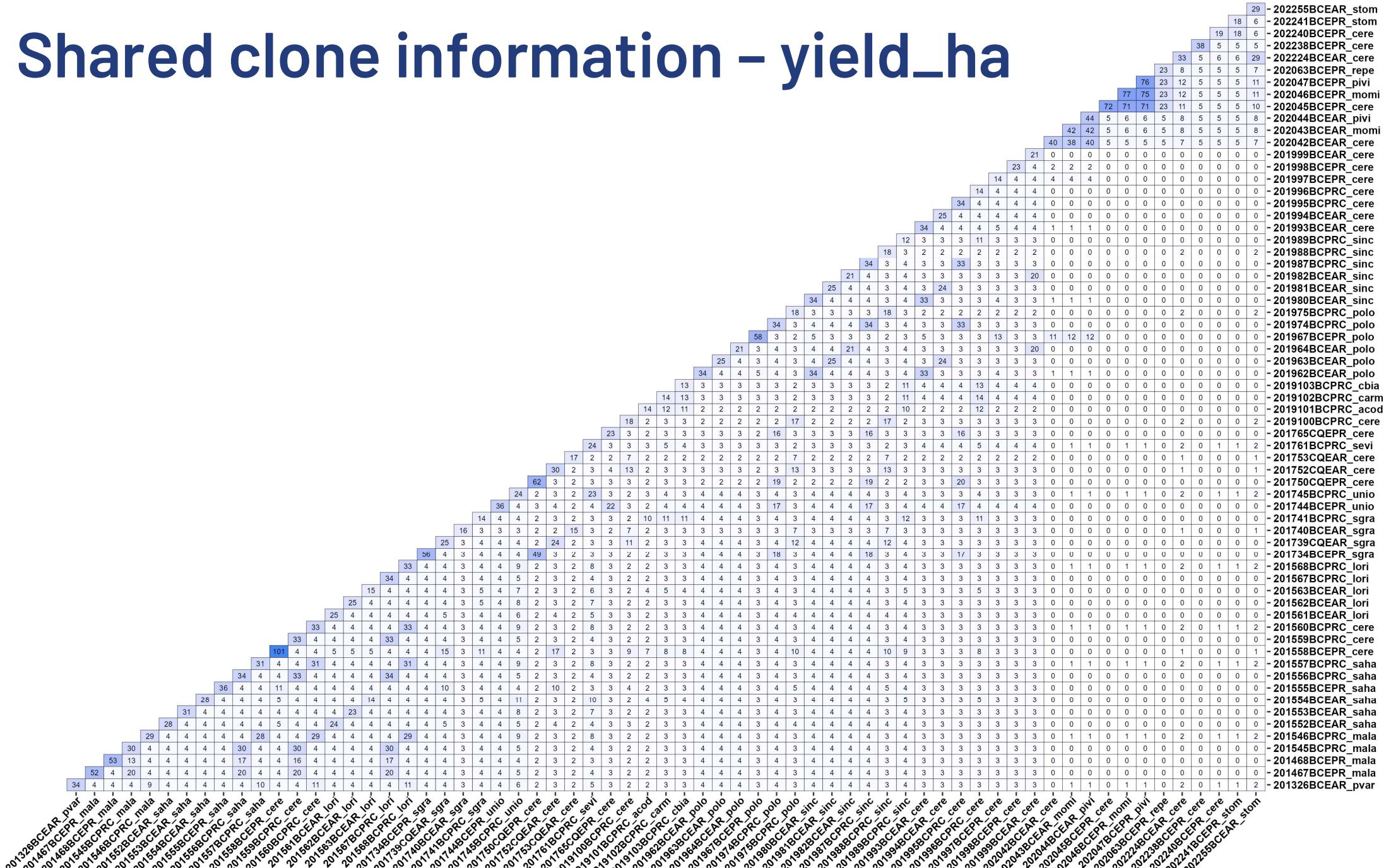


Yield

Raw yield_ha data distribution across trials



Shared clone information - yield_ha



Running a simple linear mixed model

- Removing trials that not converged
- Detecting and removing outliers

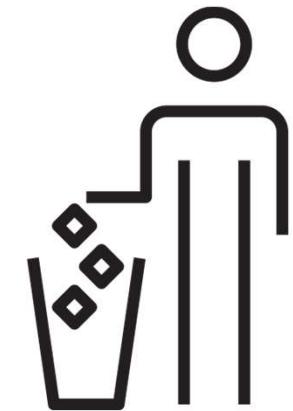
$\text{yield_ha} = \text{rep_number} + (1 | \text{accession_name})$

↓
Fixed term

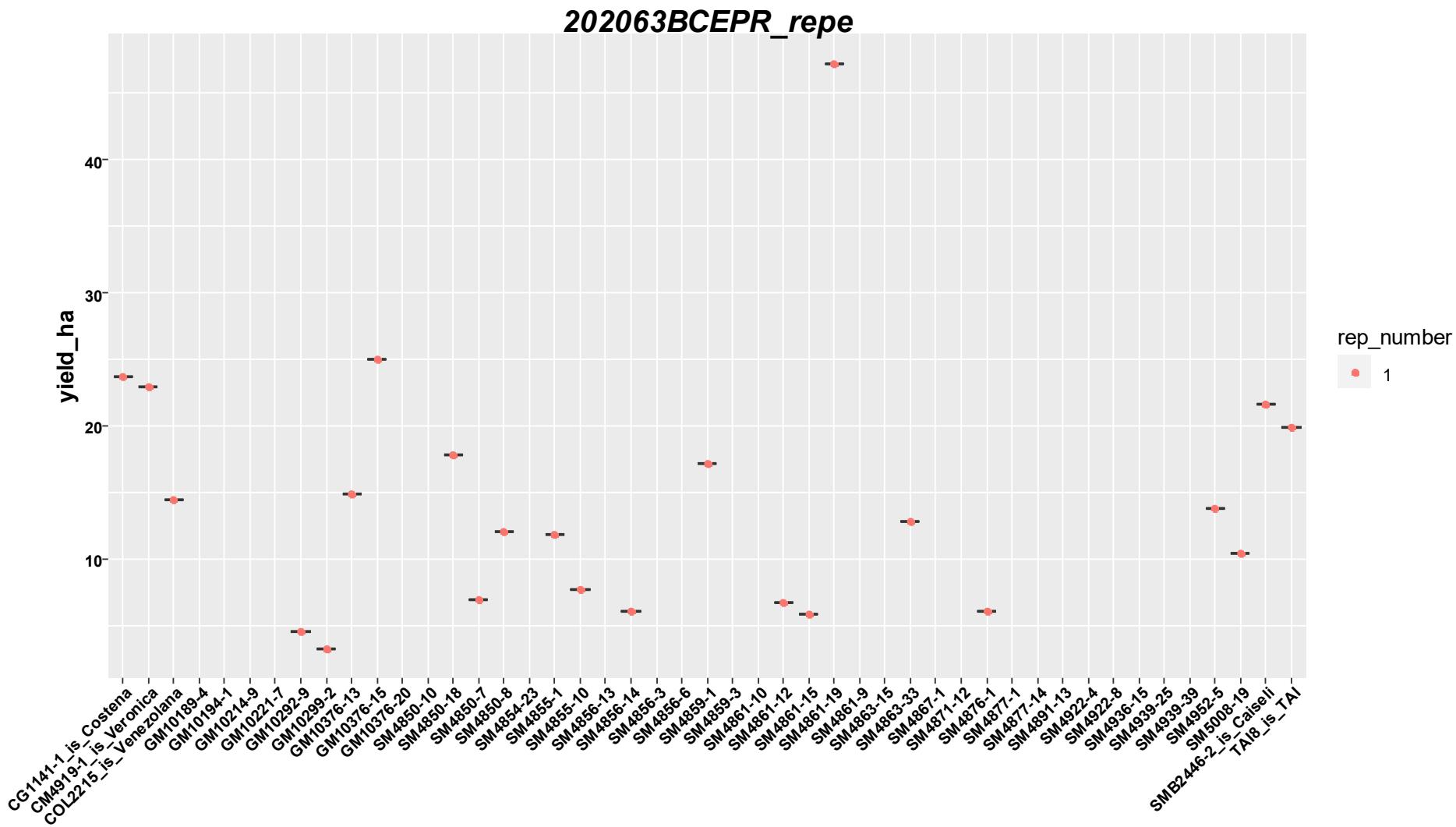
Random term →

After this step, 1 trial model did not converge
and was removed from the analysis.

Trials removed
202063BCEPR_repe

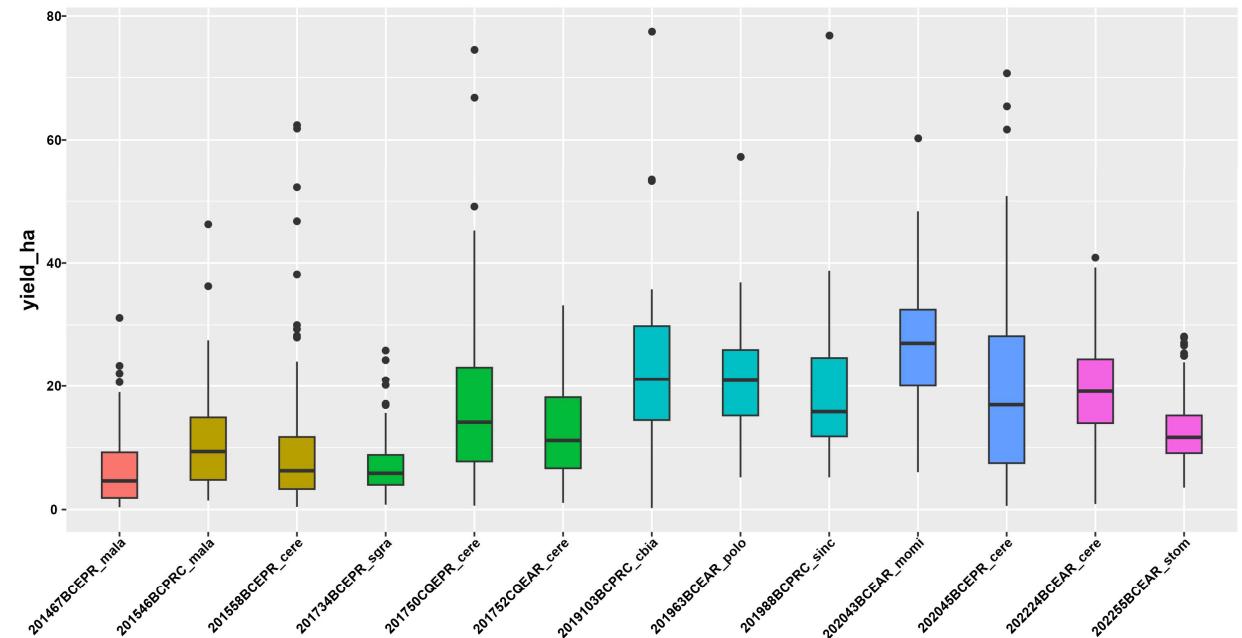


202063BCEPR_repe -> not converged.



Yield_ha outliers

Definition: an *outlier* is an observation that lies an abnormal distance from other values in a random sample from a population.

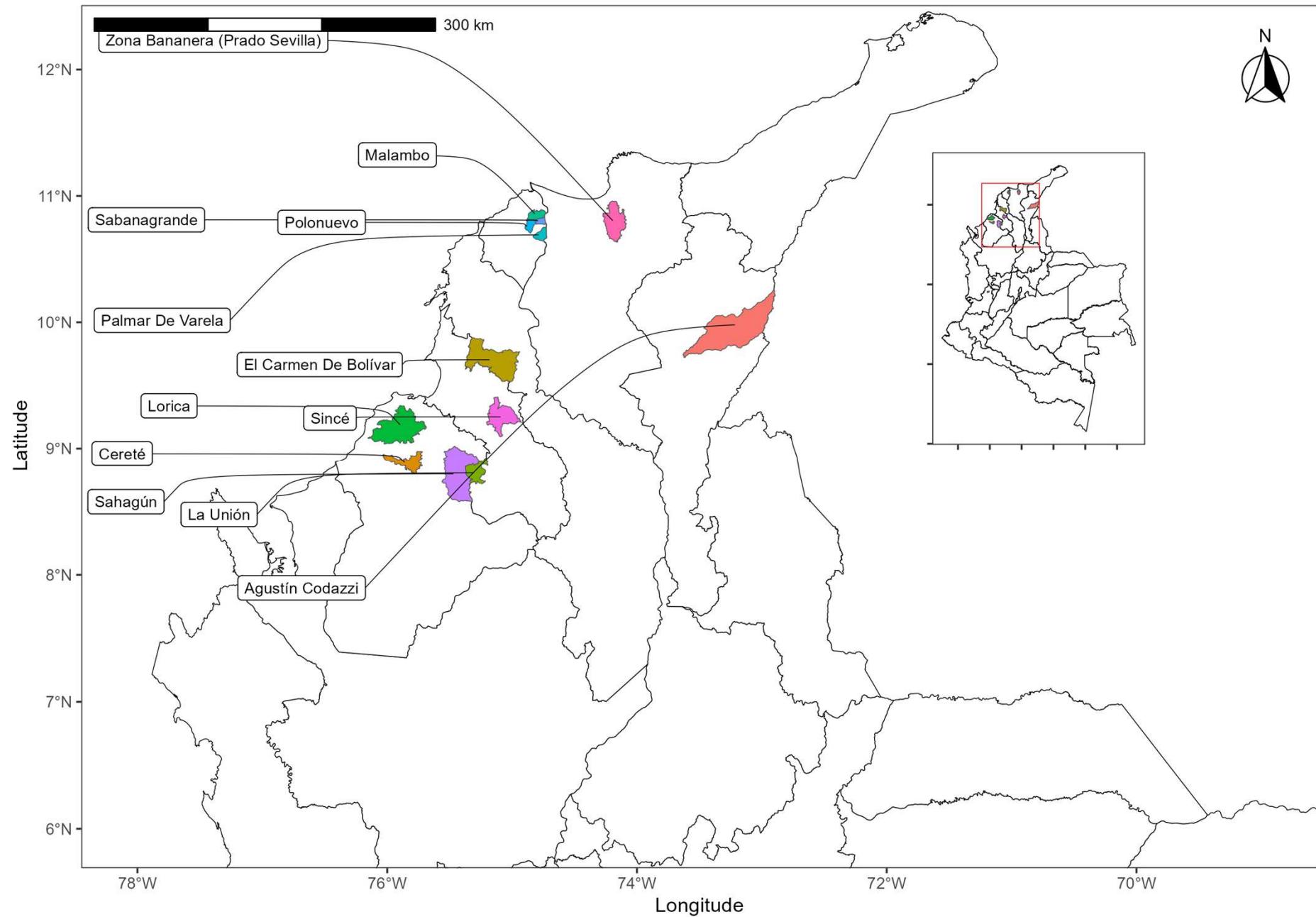


Standardized residuals:
Data points with residuals \leq or \geq equal to 3 times SD were classified as outliers.

Experiment	yield_ha	rep_number	accession_name	res	Classify
20146BCEPR_mala	31.12	2	TAI8	3.11	Outlier
201546BCPRC_mala	46.25	2	TAI8	3.36	Outlier
201558BCEPR_cere	52.25	2	TAI8	3.54	Outlier
201734BCEPR_sgra	24.3	2	GM8978-15	4.03	Outlier
201750CQEPR_cere	74.58	2	GM9019-7	3.74	Outlier
201752CQEPR_cere	33.17	1	SM3882-76	3.13	Outlier
2019103BCPRC_cbia	77.5	2	CG1141-1	3.32	Outlier
201963BCEAR_polo	57.29	1	GM9458-6	3.12	Outlier
201988BCPRC_sinc	76.88	1	TAI8	4.05	Outlier
202043BCEAR_momi	60.26	2	TAI8_is_TAI	3.21	Outlier
202045BCEPR_cere	70.67	2	SM4939-25	3.4	Outlier
202224BCEAR_cere	39.29	3	SM4891-13	3.02	Outlier
202255BCEAR_stom	23.96	1	GM3426-5	3.13	Outlier
202255BCEAR_stom	28.12	3	SMB2446-2_is_Caiseli	3.16	Outlier

Municipalities yield_ha trials

Agustín Codazzi
Cereté
El Carmen De Bolívar
La Unión
Lorica
Malambo
Palmar De Varela
Polonuevo
Sabanagrande
Sahagún
Sincé
Zona Bananera (Prado Sevilla)



Yield across trials kept

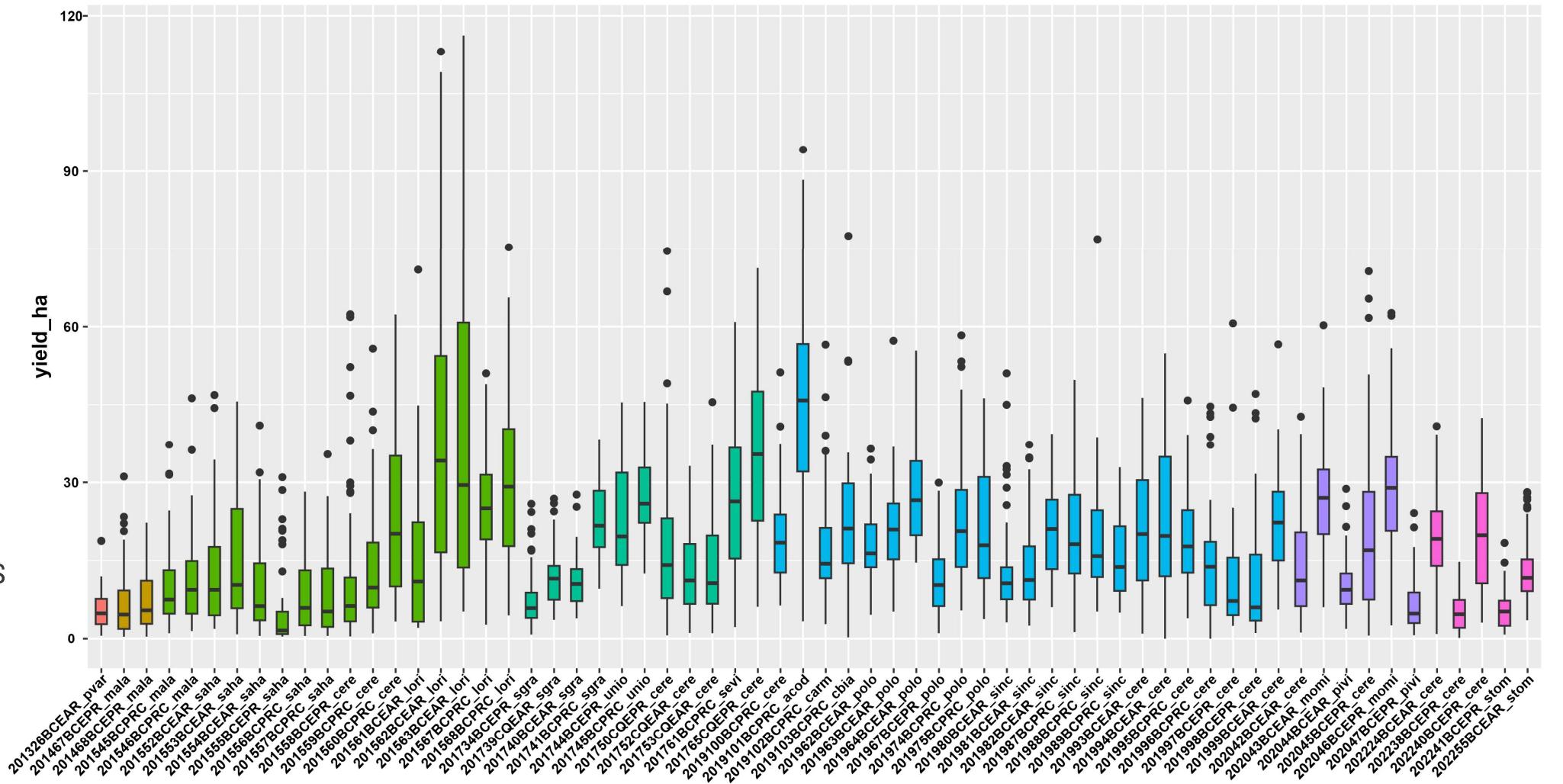
52 trials

5 Years

16 Locations

RCBD - 2/3 blocks

CRD



Models

1.

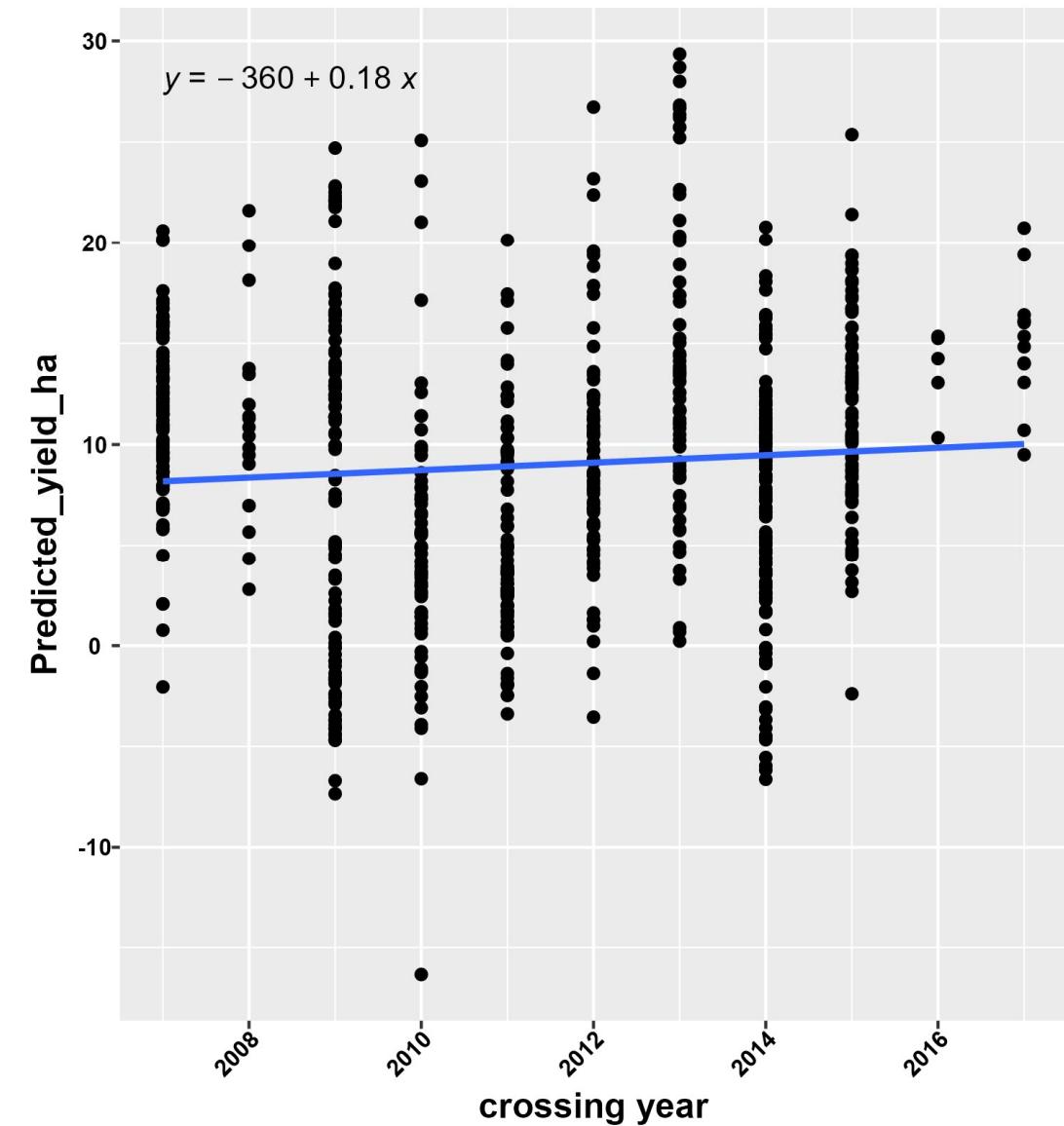
fixed = yield_ha ~ 1 + accession_name + year,
random = ~ trial_name + location + year:location + accession_name:year +
accession_name:location + trial_name:rep_number,
residual = ~ units

2.

fixed = yield_ha ~ 1 + accession_name + year,
random = ~ trial_name + **diag**(trial_name):accession_name + **at**(trial_name):rep_number,
residual = ~ **dsum**(~ units | trial_name)

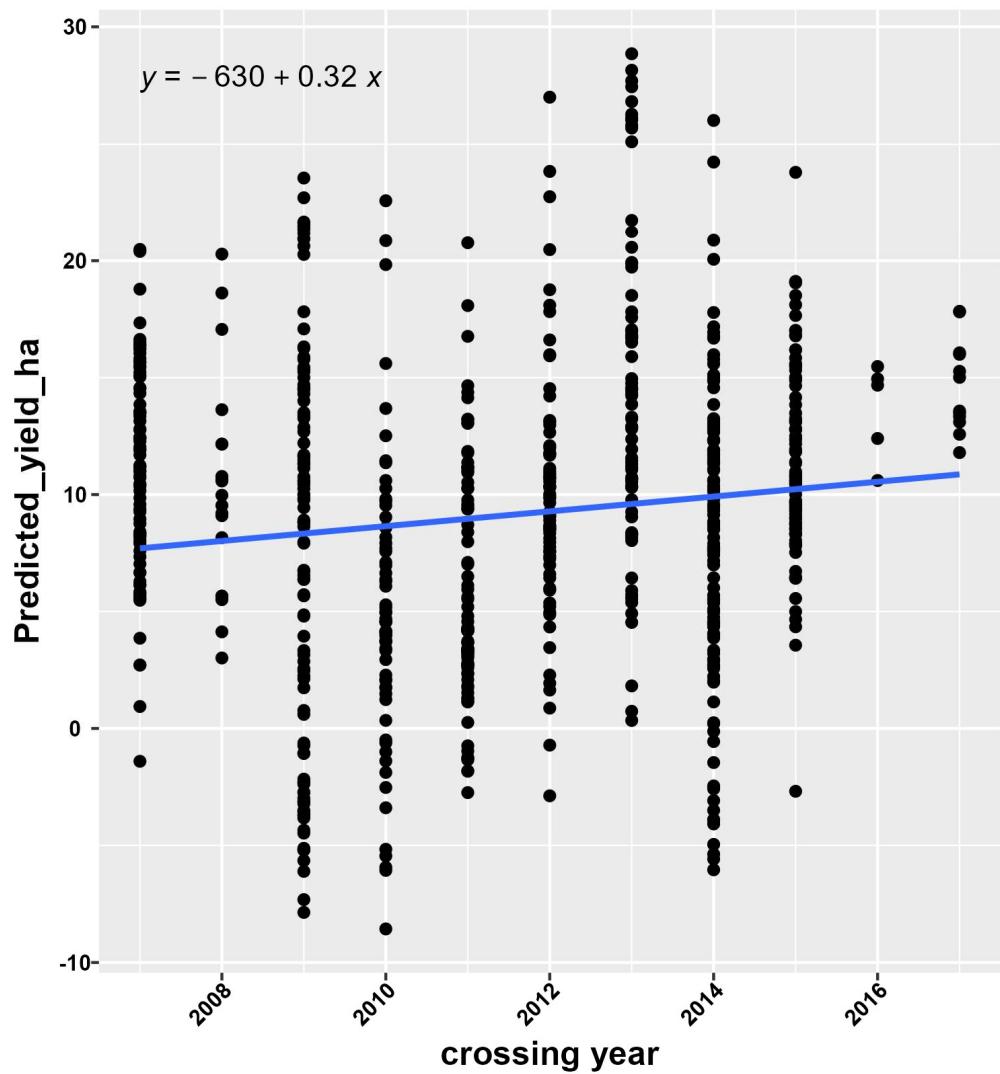
Model 1 Estimation

trait	yield_ha
first_year	2007
last_year	2017
Slope	0.1839
se_Slope	0.0946
Intercept	-361.0011
r2	0.0052
Pr(>F)	5.21E-02
Genetic_Gain%	2.2477



Model 2 Estimation

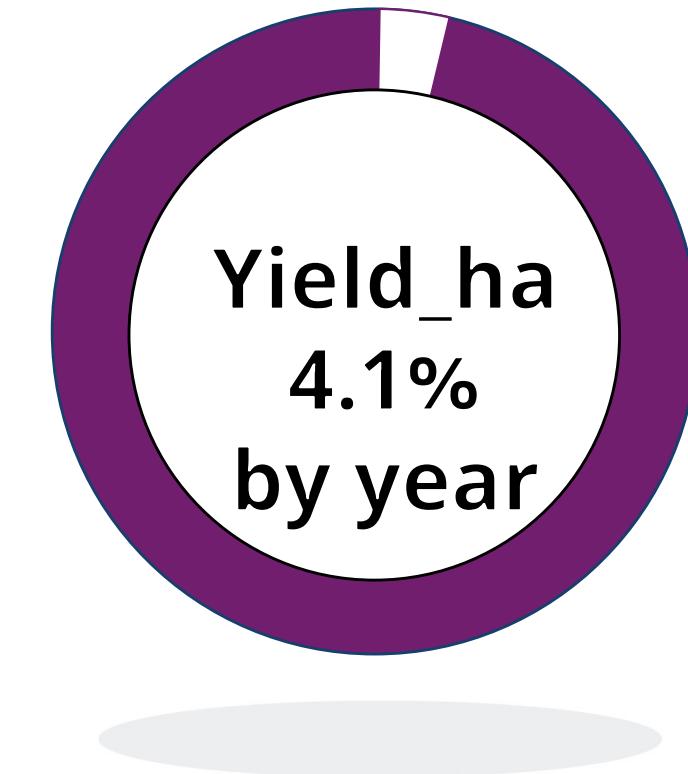
trait	yield_ha
first_year	2007
last_year	2017
Slope	0.3164
se_Slope	0.0915
Intercept	-627.4015
r2	0.0162
Pr(>F)	6.00E-04
Genetic_Gain%	4.1098



Genetic Gain yield_ha



0.31 t/ha

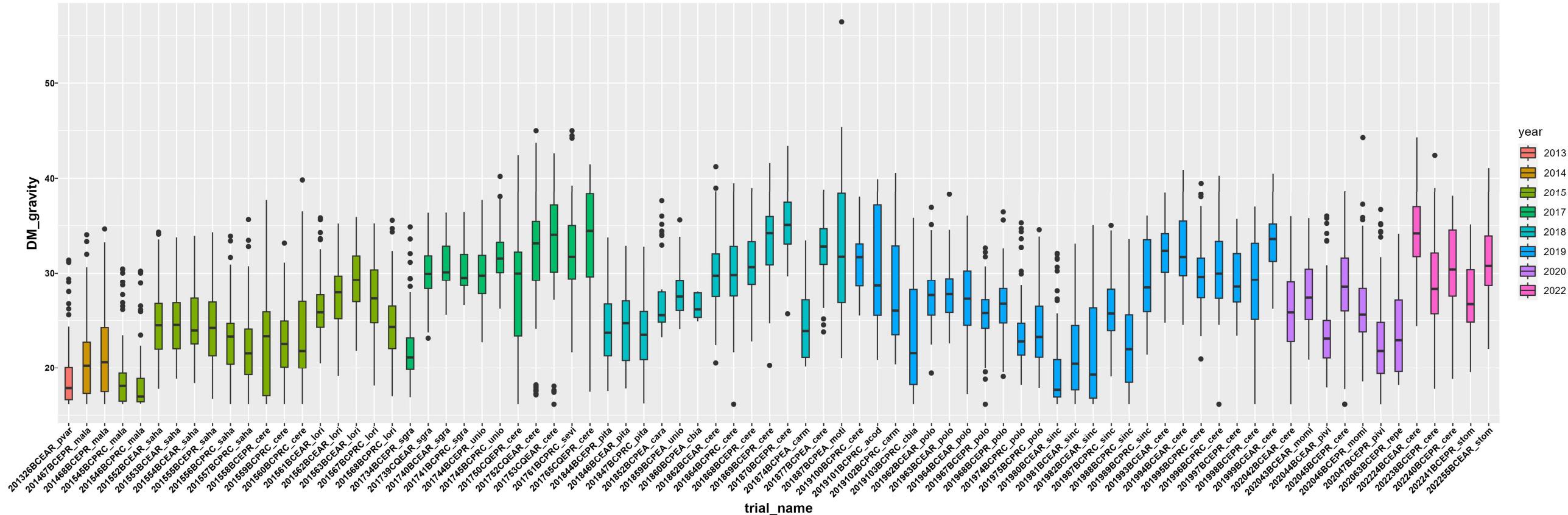


$$\frac{slope}{firstYear} = \frac{0.31}{intercept + (2007 * slope)} = \frac{0.31}{7.70} = 0.041$$

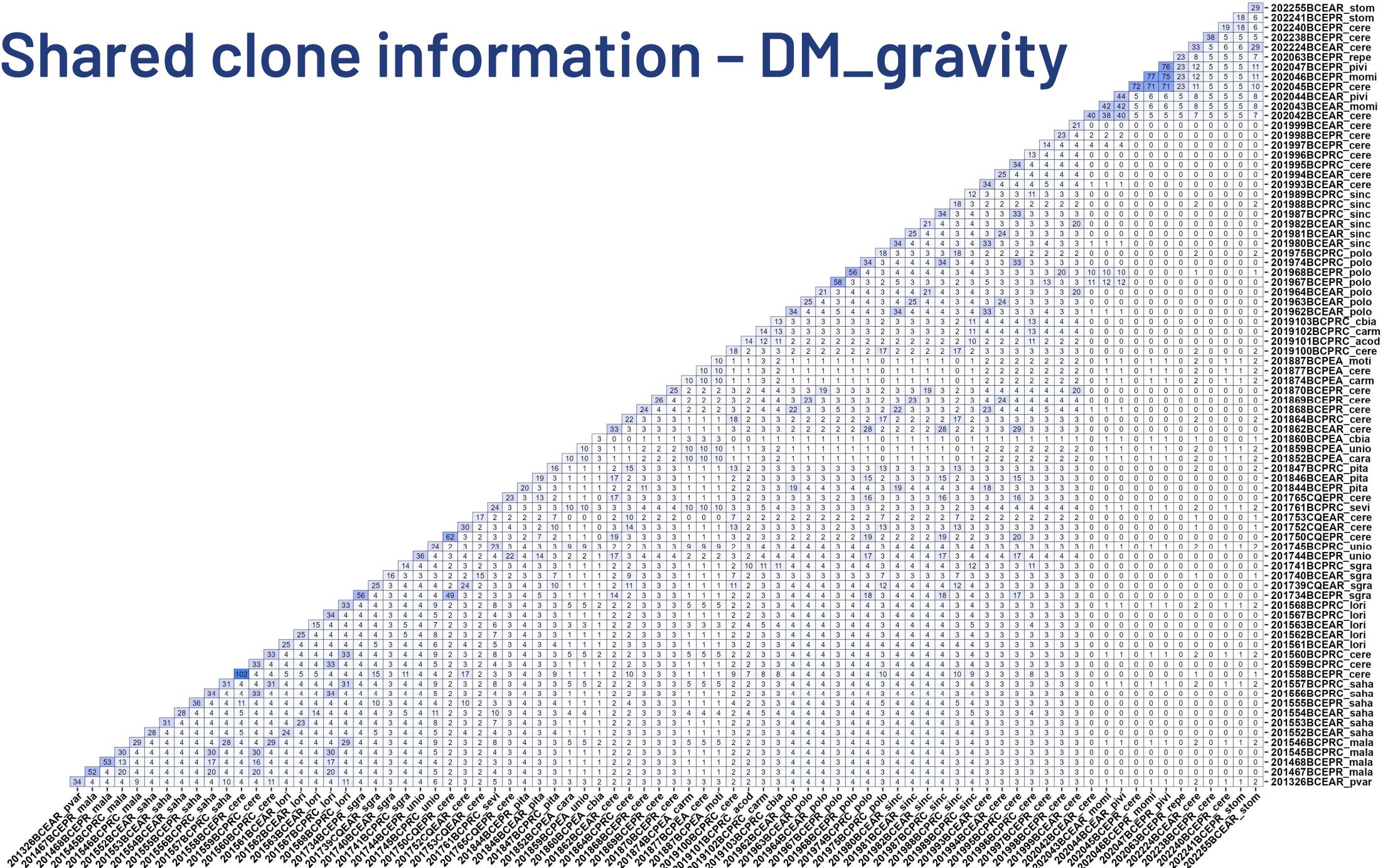


Dry matter

Raw DM_gravity data distribution across trials

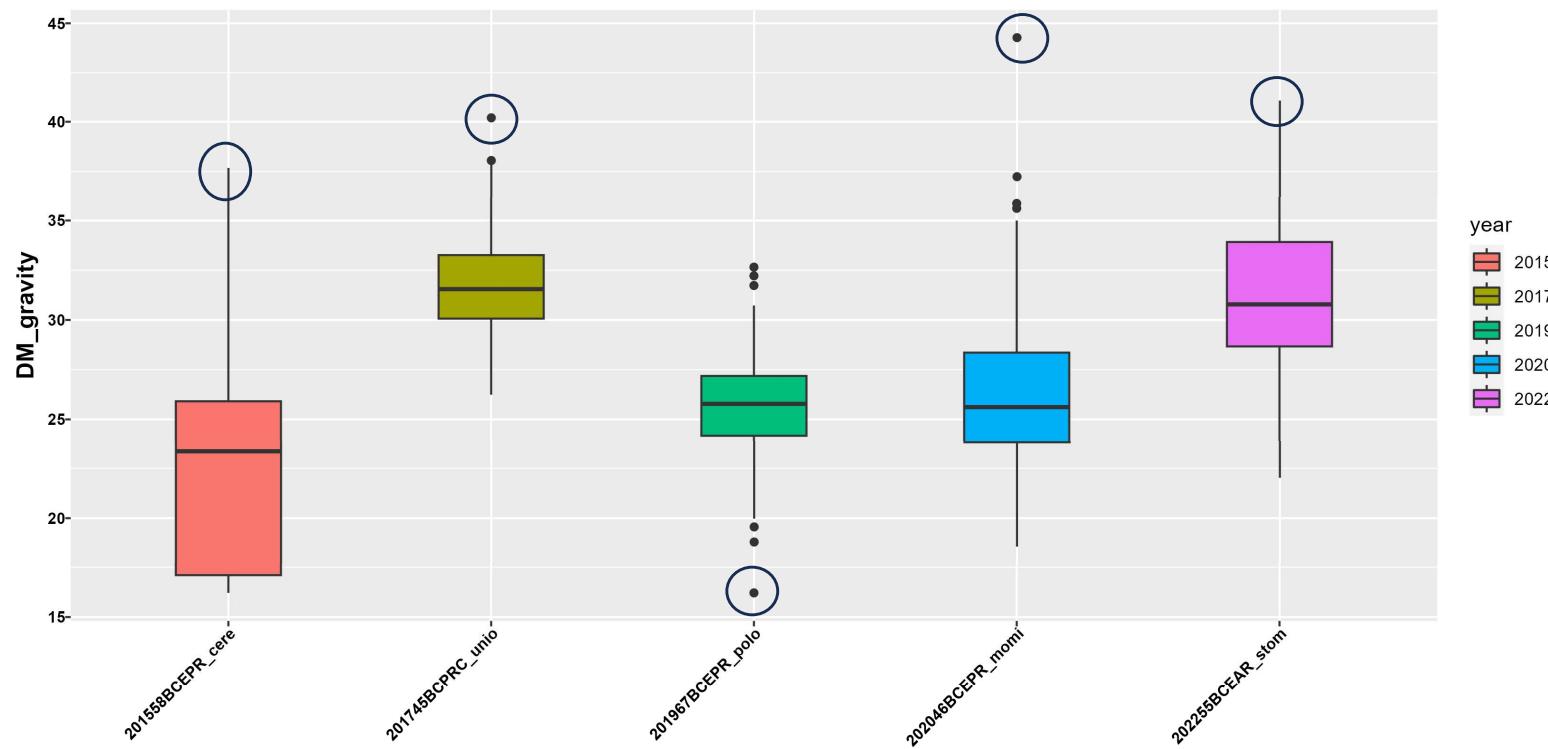


Shared clone information - DM_gravity

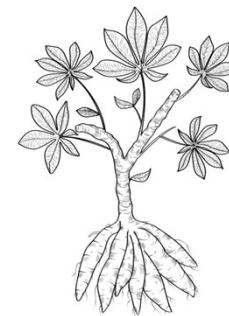
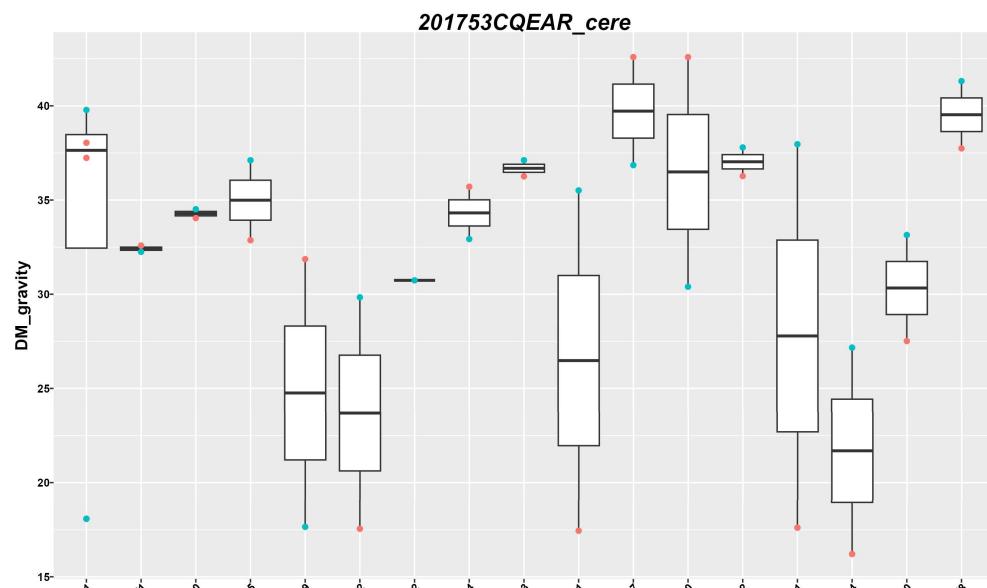


DM_gravity outliers

Experiment	DM_gravity	rep_number	accession_name	res	Classify
201558BCEPR_cere	37.69	1	GM5204-11	3.04	Outlier
201745BCPRC_unio	40.2	2	CM4919-1	3.26	Outlier
201967BCEPR_polo	16.21	2	GM10004-2	-3.23	Outlier
202046BCEPR_momi	44.28	2	COL2215_is_Venezolana	3.16	Outlier
202255BCEAR_stom	41.06	2	SM3536-44	3.86	Outlier

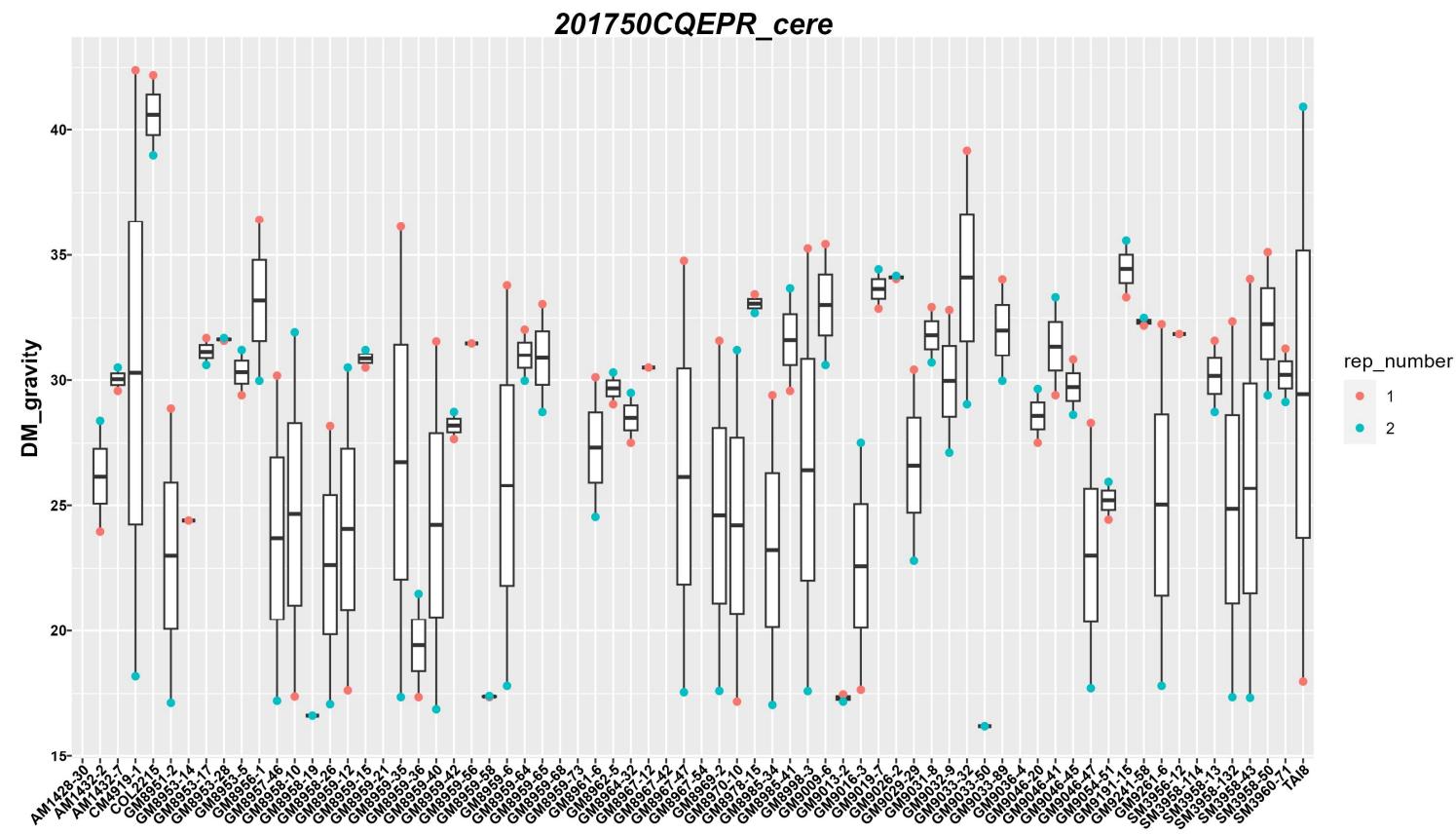


Single heritability = 0



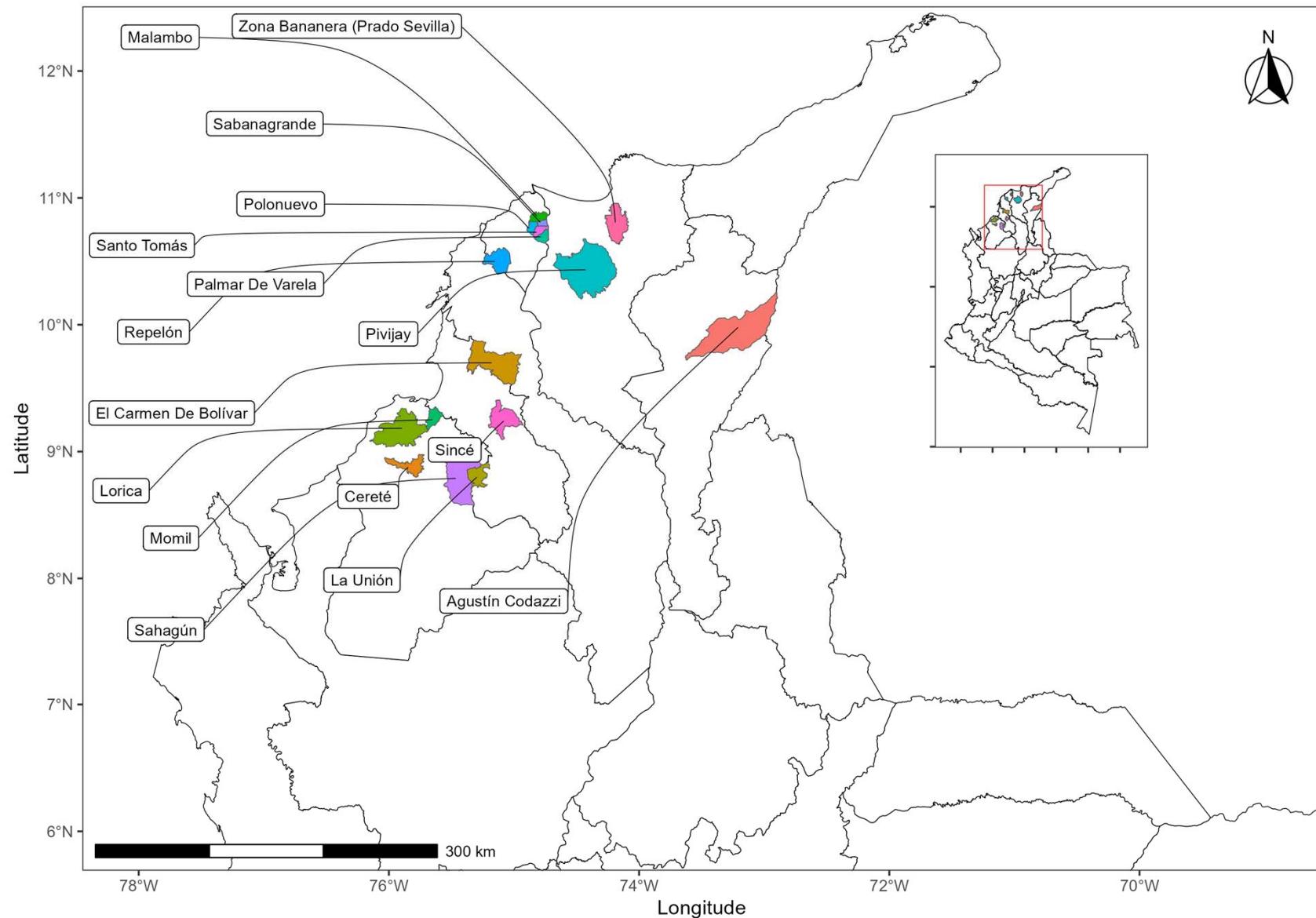
- 201750CQEPR_cere
- 201753CQEPR_cere"

Removed because of
 $H^2 = 0$



Municipalities DM_gravity trials

Agustín Codazzi
Cereté
El Carmen De Bolívar
La Unión
Lorica
Malambo
Momil
Palmar De Varela
Pivijay
Polonuevo
Repelón
Sabanagrande
Sahagún
Santo Tomás
Sincé
Zona Bananera (Prado Sevilla)



DM_gravity across trials kept

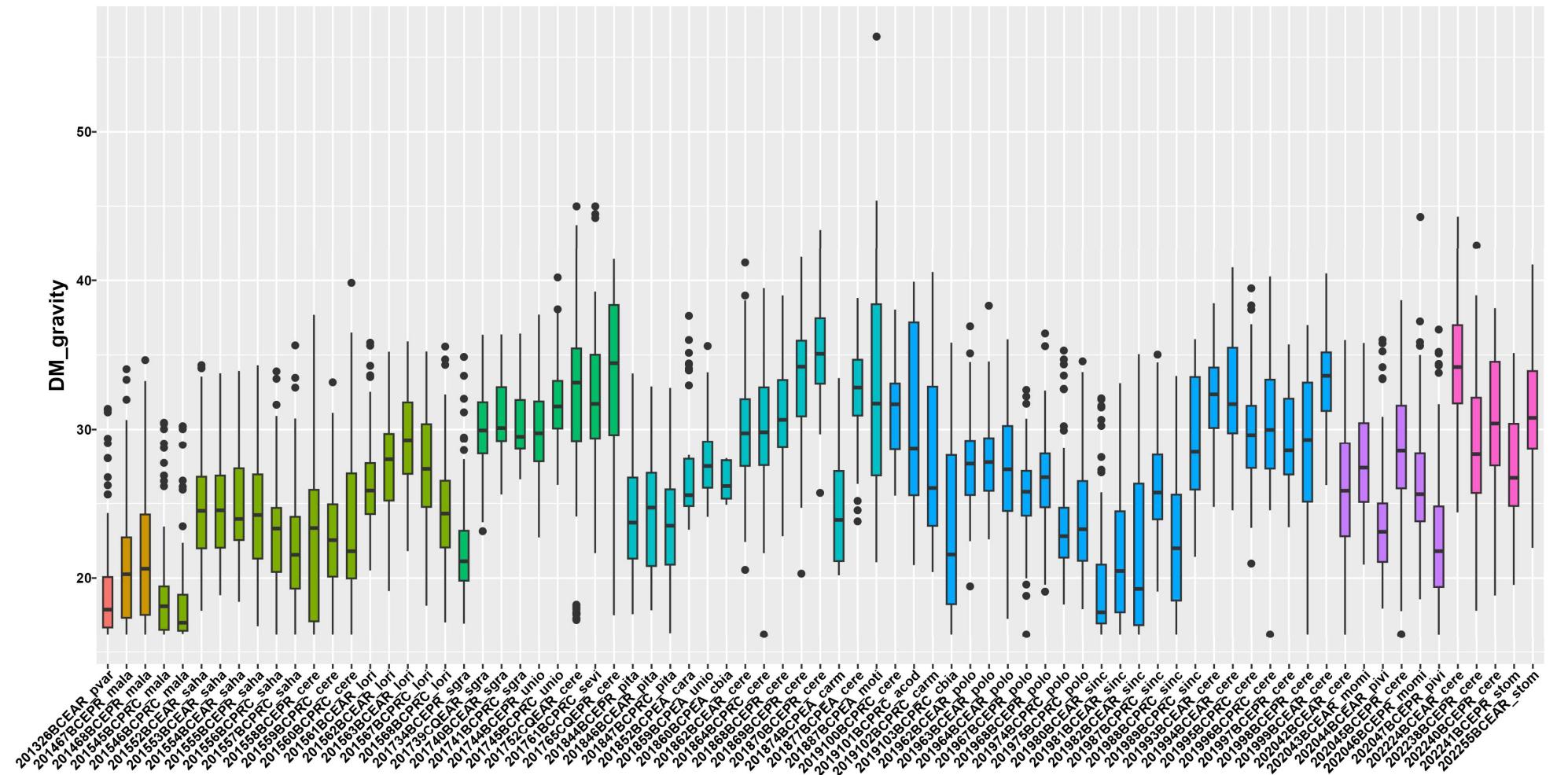
77 trials

8 Years

19 Locations

RCBD - 2/3 bloc

RCD



Models

1.

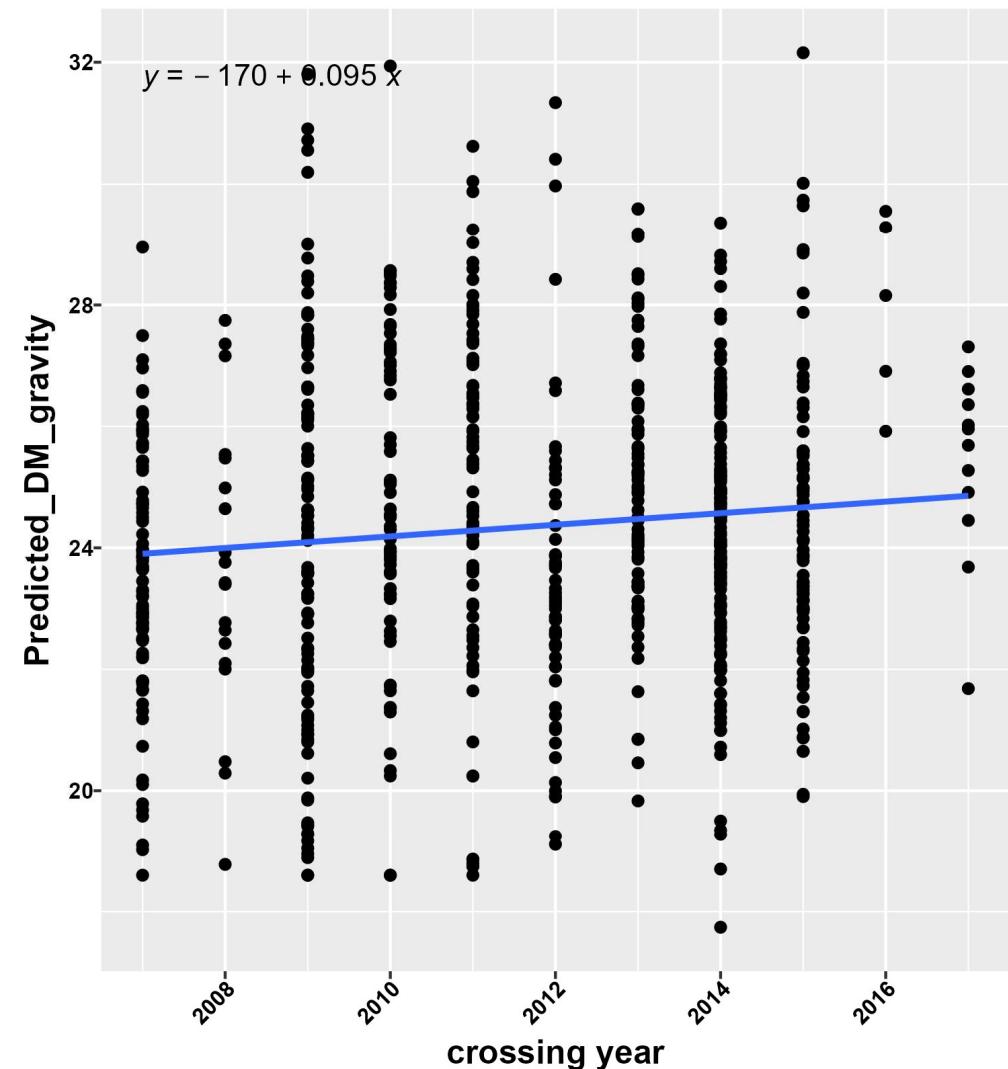
fixed = yield_ha ~ 1 + accession_name + year,
random = ~ trial_name + location + year:location + accession_name:year +
accession_name:location + trial_name:rep_number,
residual = ~ units

2.

fixed = yield_ha ~ 1 + accession_name + year,
random = ~ trial_name + **diag**(trial_name):accession_name + **at**(trial_name):rep_number,
residual = ~ **dsum**(~ units | trial_name)

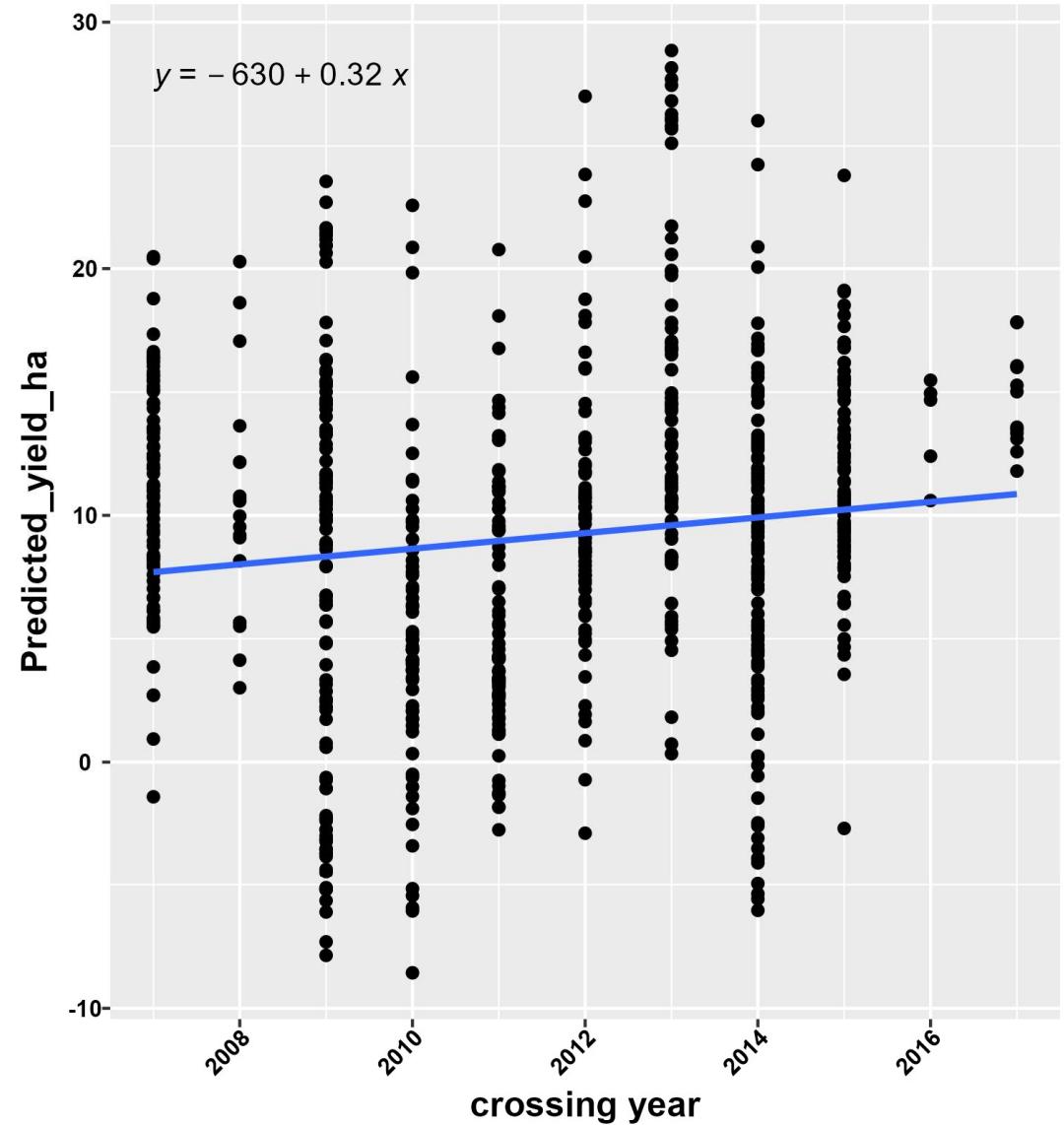
Model 1 Estimation

trait	DM_gravity
first_year	2007
last_year	2017
Slope	0.0954
se_Slope	0.0343
Intercept	-167.6628
r2	0.0102
Pr(>F)	5.50E-03
Genetic_Gain%	0.3993



Model 2 Estimation

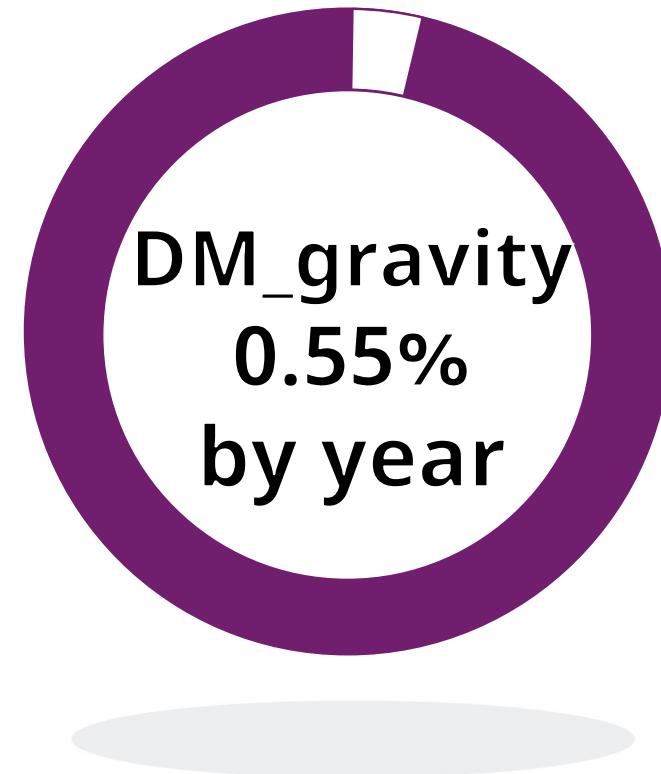
trait	DM_gravity
first_year	2007
last_year	2017
Slope	0.1316
se_Slope	0.0345
Intercept	-240.3764
r2	0.0191
Pr(>F)	1.00E-04
Genetic_Gain%	0.5543



Genetic Gain DM_gravity



0.13 %



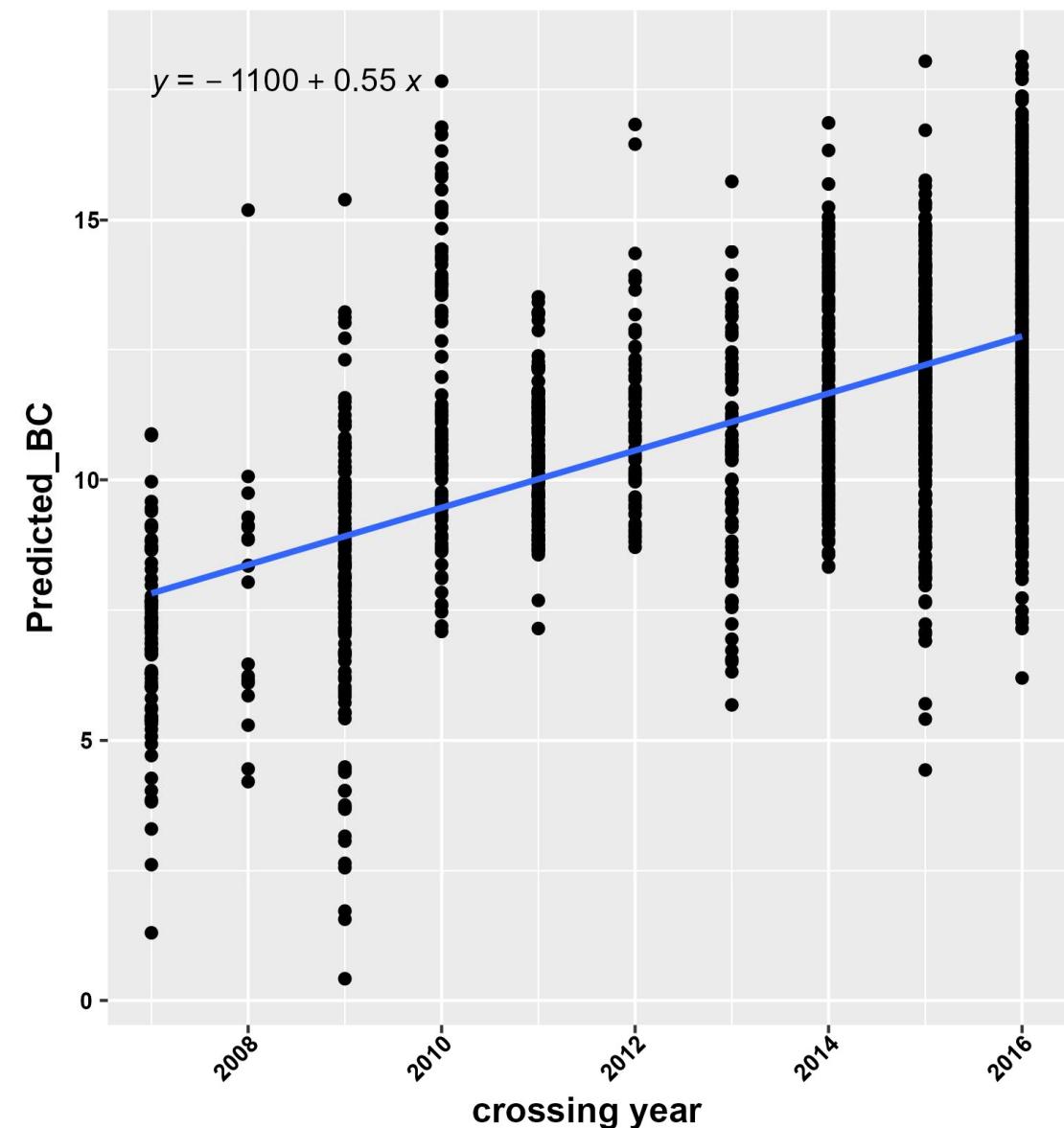
$$\frac{slope}{firstYear} = \frac{0.13}{intercept + (2007 * slope)} = \frac{0.13}{23.74} = 0.005$$



Betacoretenes

Model 2 Estimation

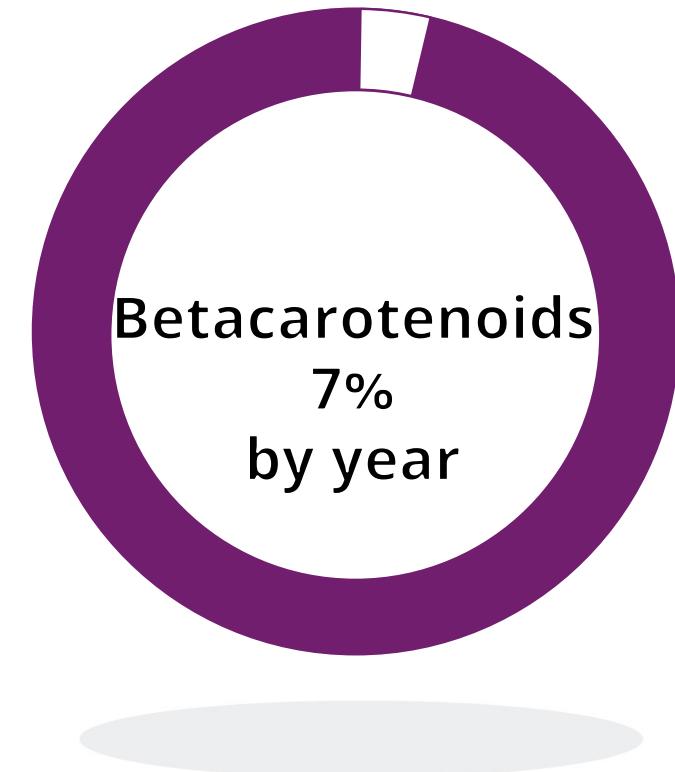
trait	BC
first_year	2007
last_year	2016
Slope	0.5481
se_Slope	0.0205
Intercept	-1092.251
r2	0.3344
Pr(>F)	0
Genetic_Gain%	7.0075



Genetic Gain Betacarotenes



0.55 ug/g

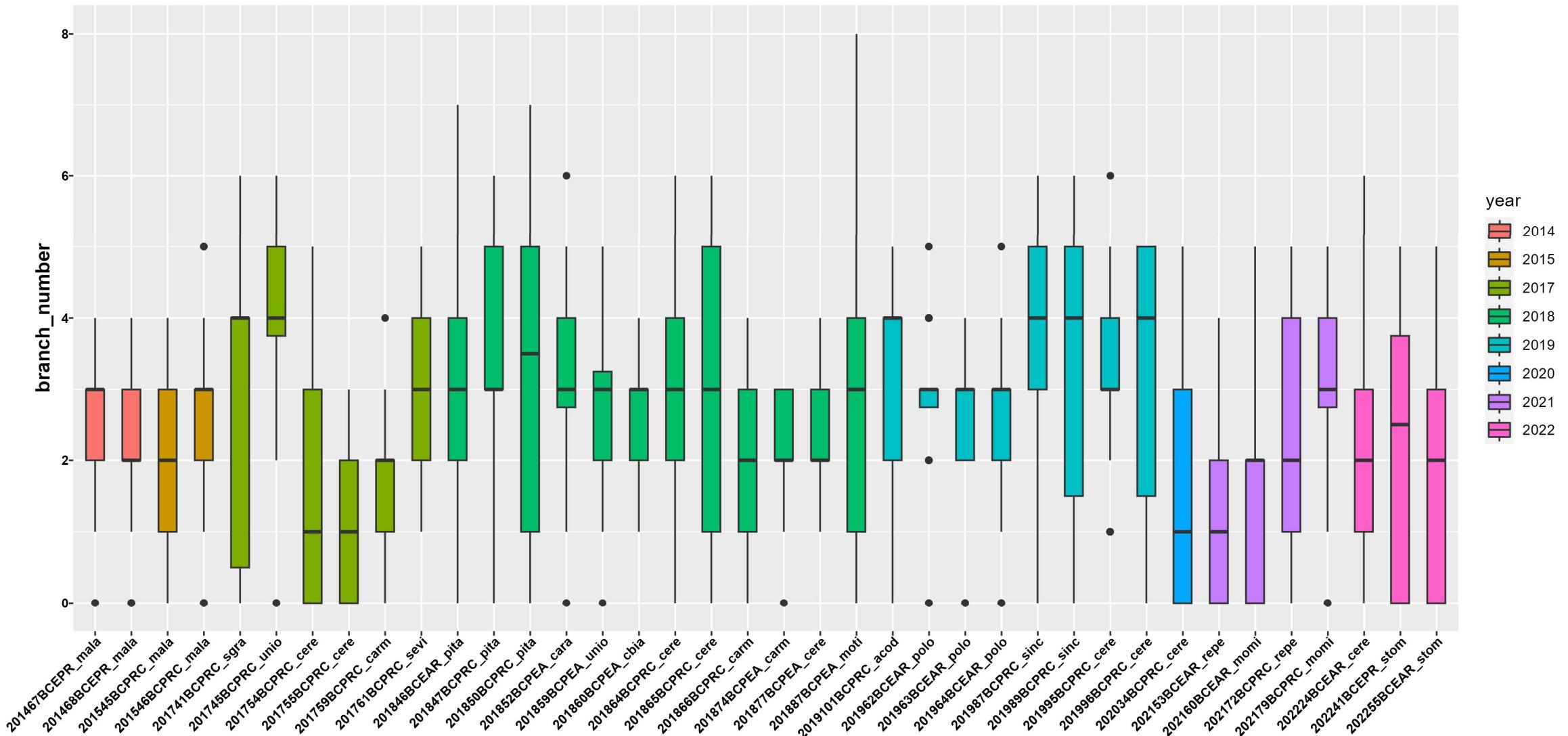


$$\frac{slope}{firstYear} = \frac{0.55}{intercept + (2007 * slope)} = \frac{0.55}{7.82} = \mathbf{0.07}$$

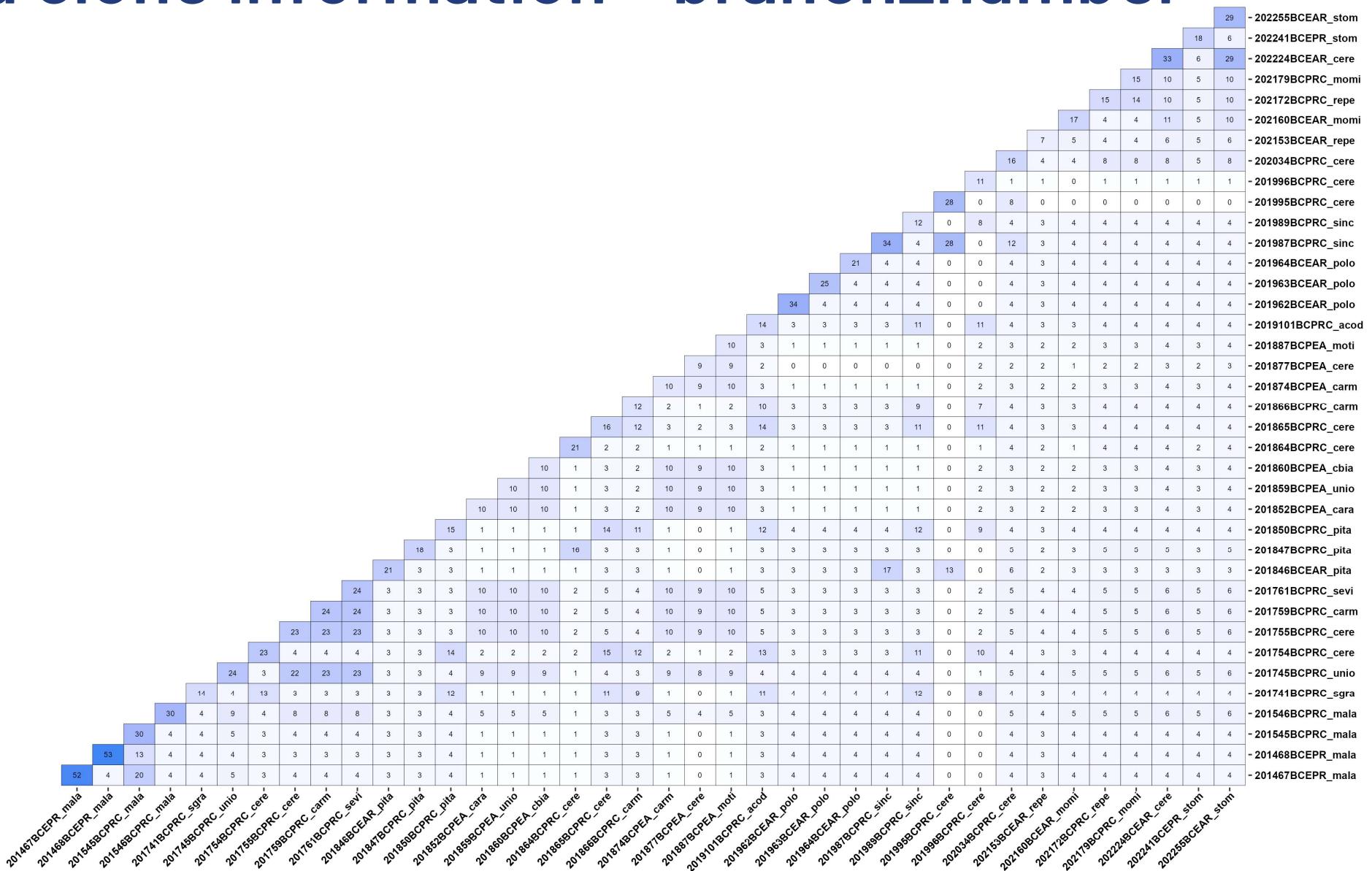


Branches number

Raw branch_number data distribution across trials



Shared clone information - branch_number



Models

1.

fixed = branch_number ~ 1 + accession_name + year,

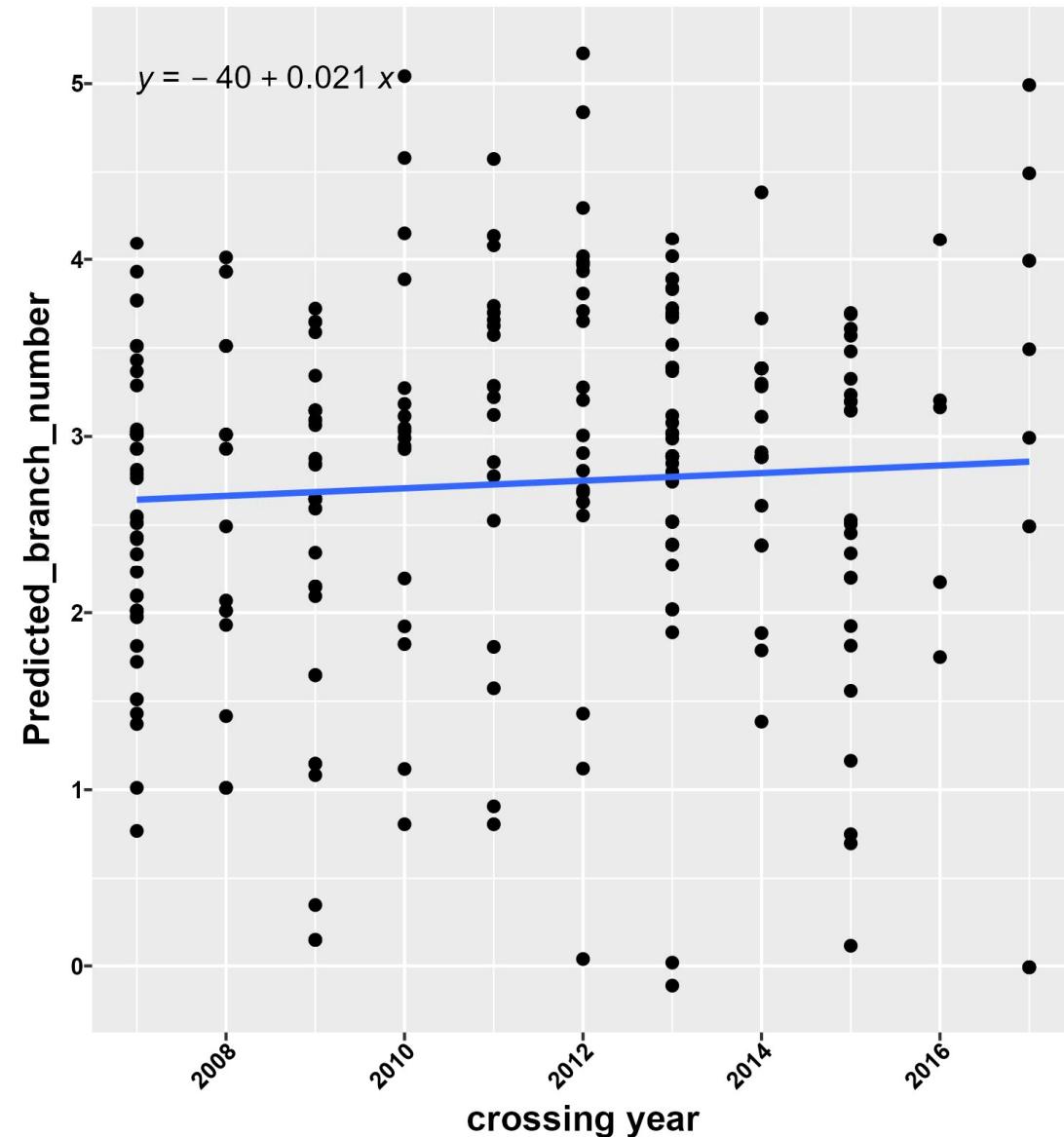
random = ~ trial_name + **diag**(trial_name):accession_name + at(trial_name):rep_number,

residual = ~ **dsum**(~ units | trial_name)

Model 1 Estimation

trait	branch_number
first_year	2007
last_year	2017
Slope	0.0214
se_Slope	0.0181
Intercept	-40.3796
r2	0.0044
Pr(>F)	0.2369
Genetic_Gain%	0.811

Unfortunately, genetic gain turned out positive, however the number of branches remain among **2 - 3** since the first year of regression.



Genetic Gain branch_number



0.0214 branches

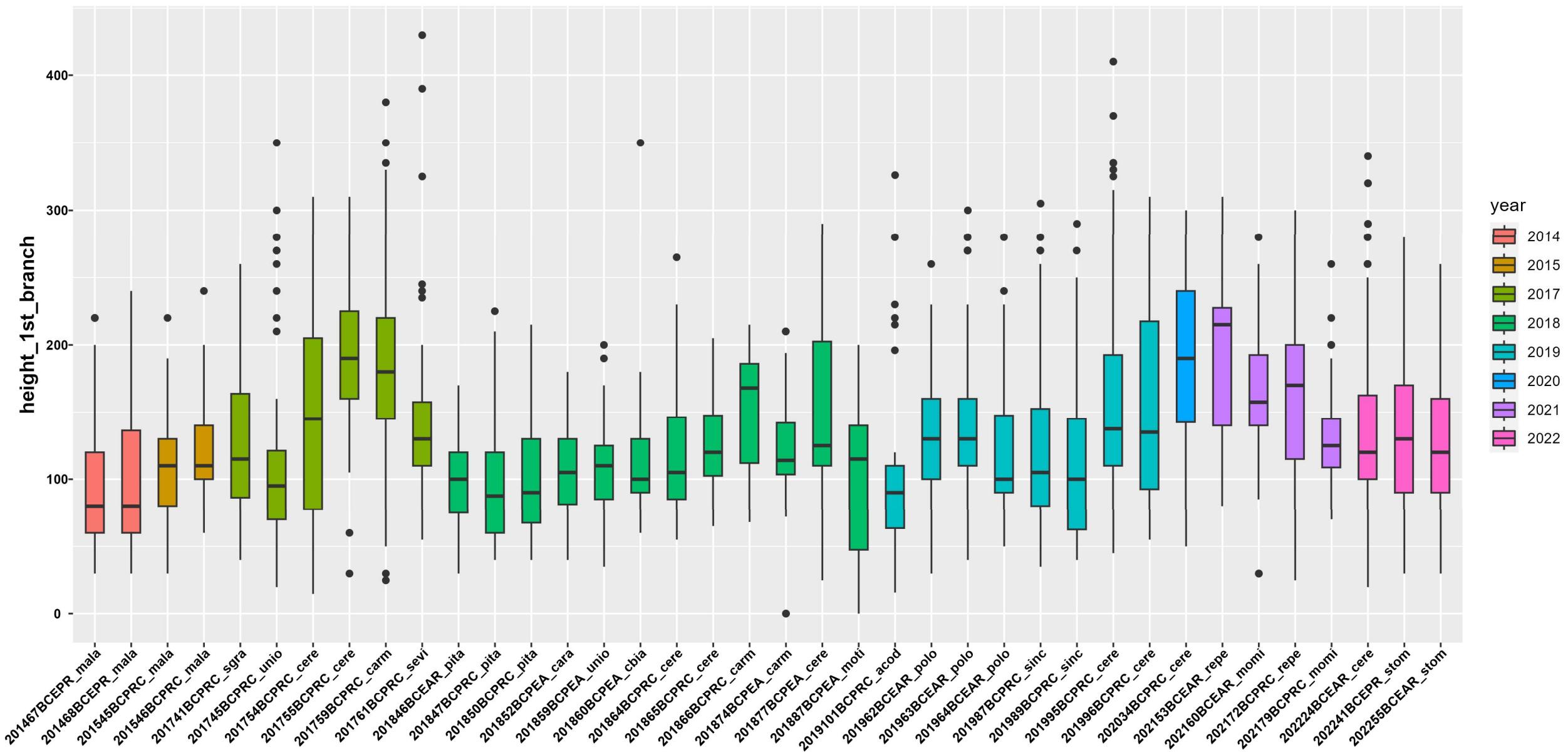
Branch_number
0.81 %
by year

$$\frac{slope}{firstYear} = \frac{0.021}{intercept + (2007 * slope)} = \frac{0.021}{2.643} = \mathbf{0.008}$$

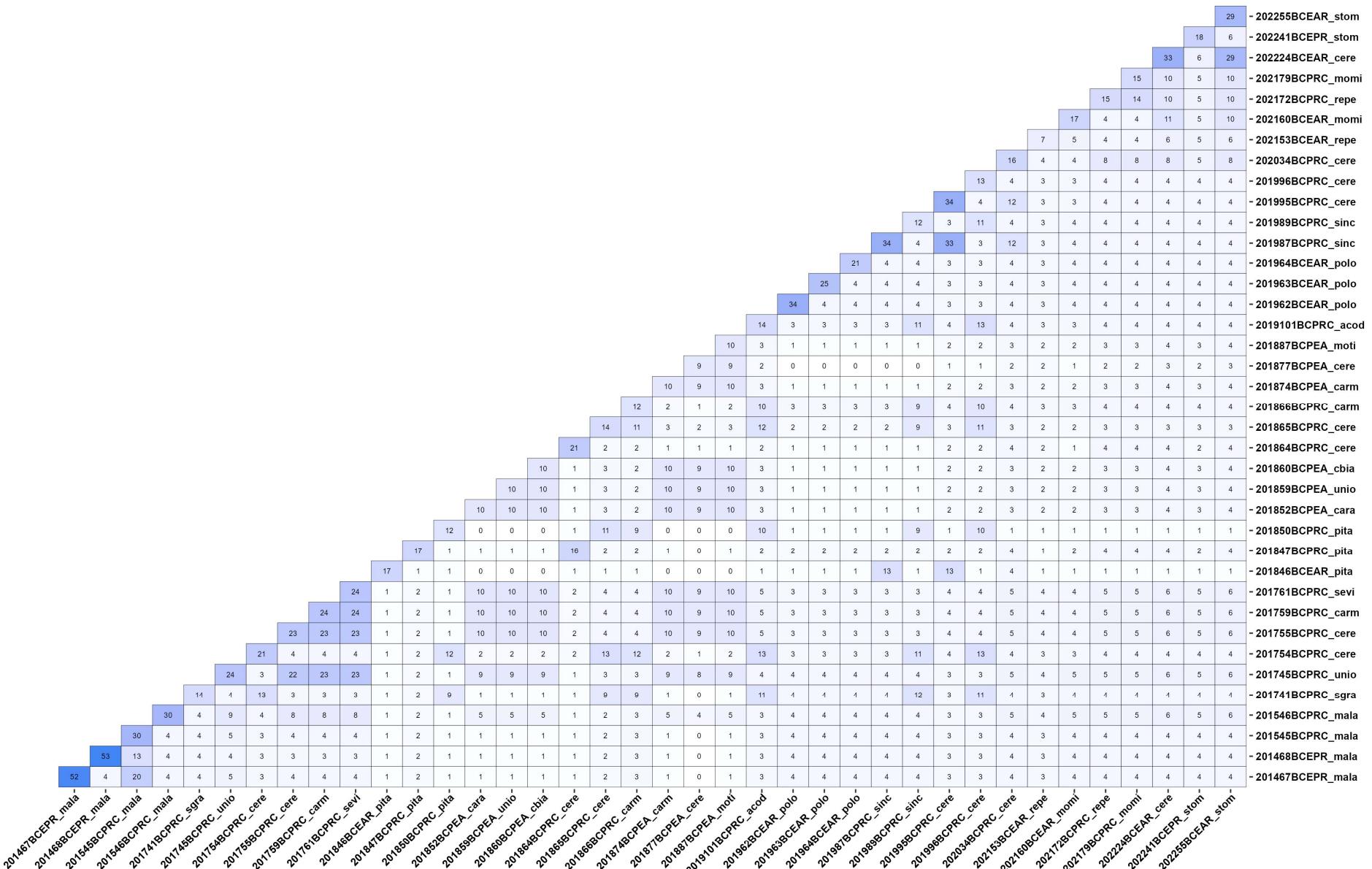


Height 1st branch

Raw height_1st_branch data distribution across trials



Shared clone information - heiaht_1st_branch



Models

1.

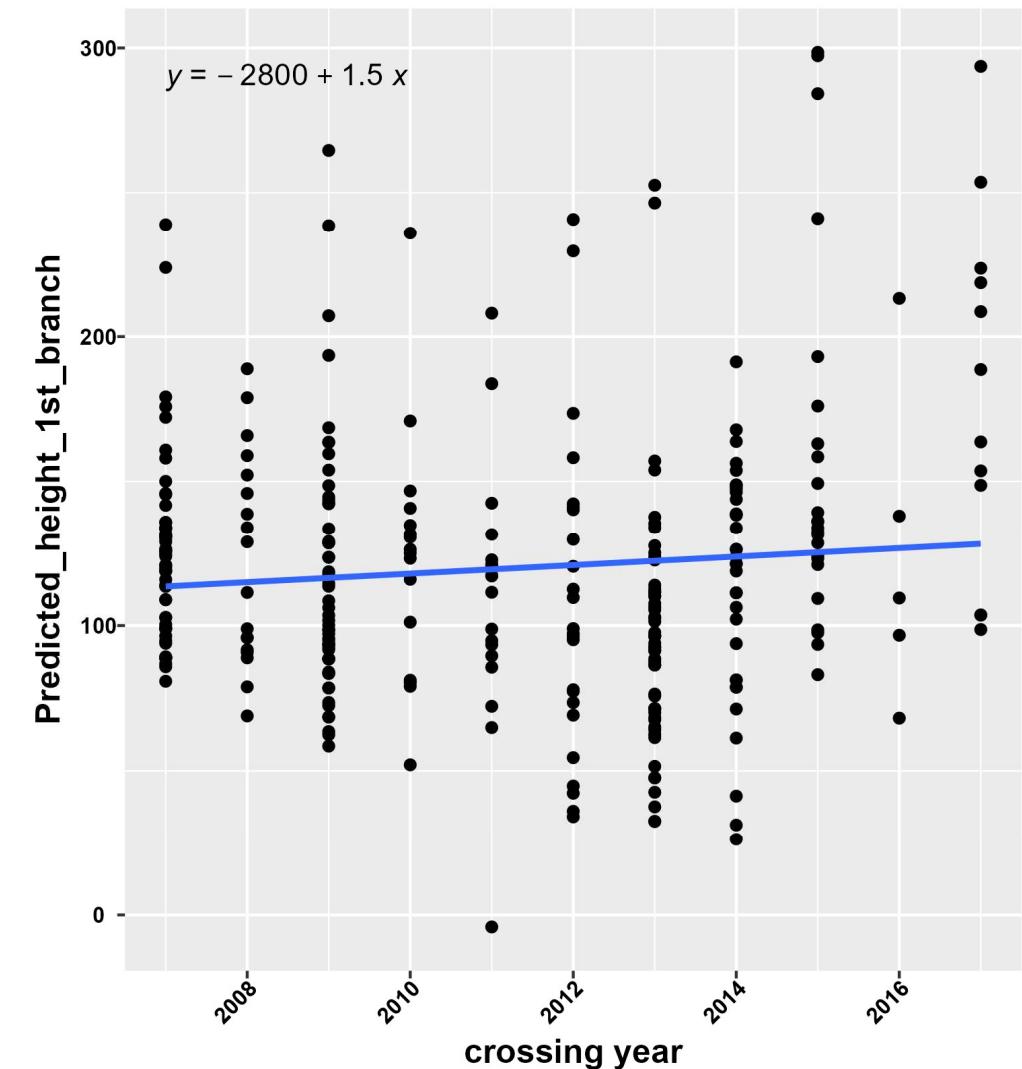
fixed = height_1st_branch~1 + accession_name + year,

random = ~ trial_name + **diag**(trial_name):accession_name + at(trial_name):rep_number,

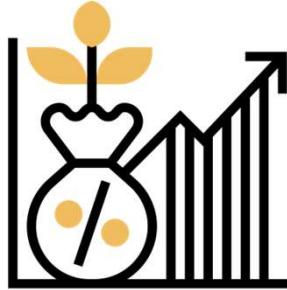
residual = ~ **dsum**(~ units | trial_name)

Model 1 Estimation

trait	height_1st_branch
first_year	2007
last_year	2017
Slope	1.4725
se_Slope	0.8921
Intercept	-2841.919
r2	0.0085
Pr(>F)	0.0998
Genetic_Gain%	1.2976



Genetic Gain height_1st_branch

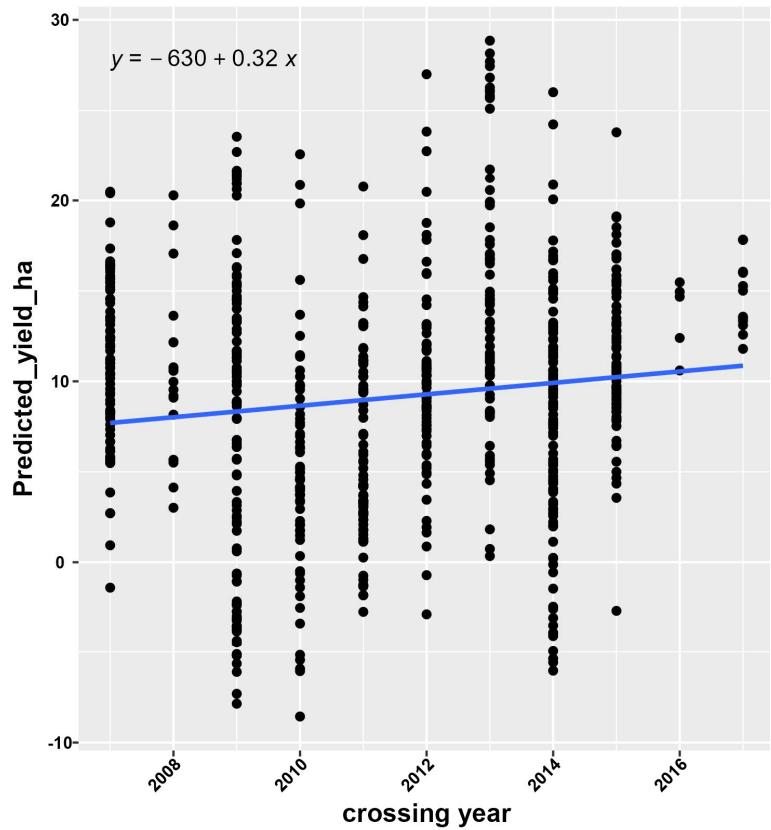


1.472 cm

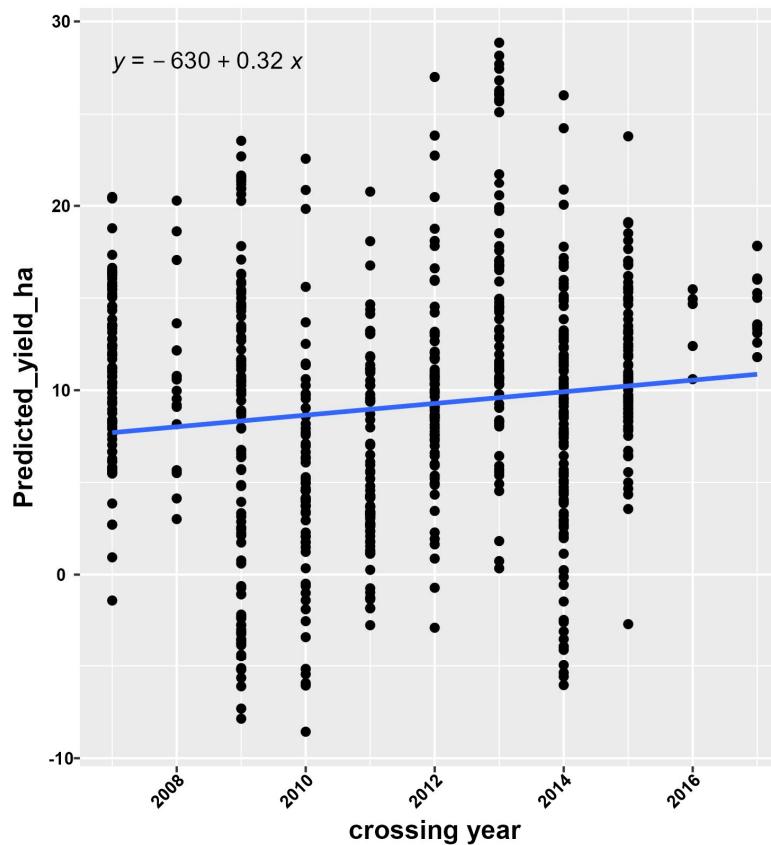
Height_1st_branch
1.30 %
by year

$$\frac{slope}{firstYear} = \frac{1.472}{intercept + (2007 * slope)} = \frac{1.472}{113.48} = 0.130$$

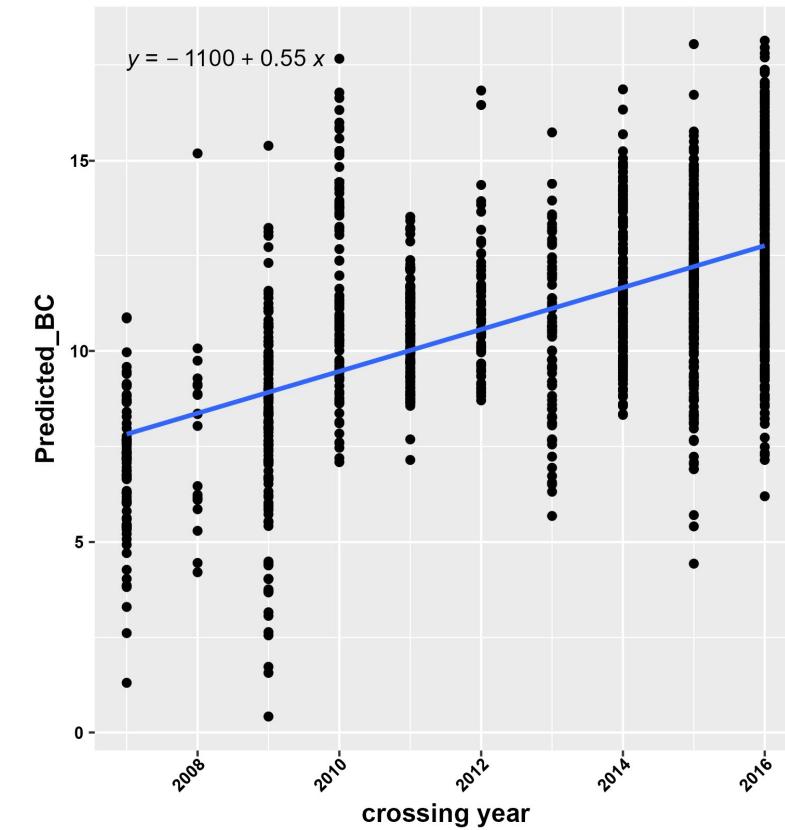
Genetic gain



yield
4.1% by year

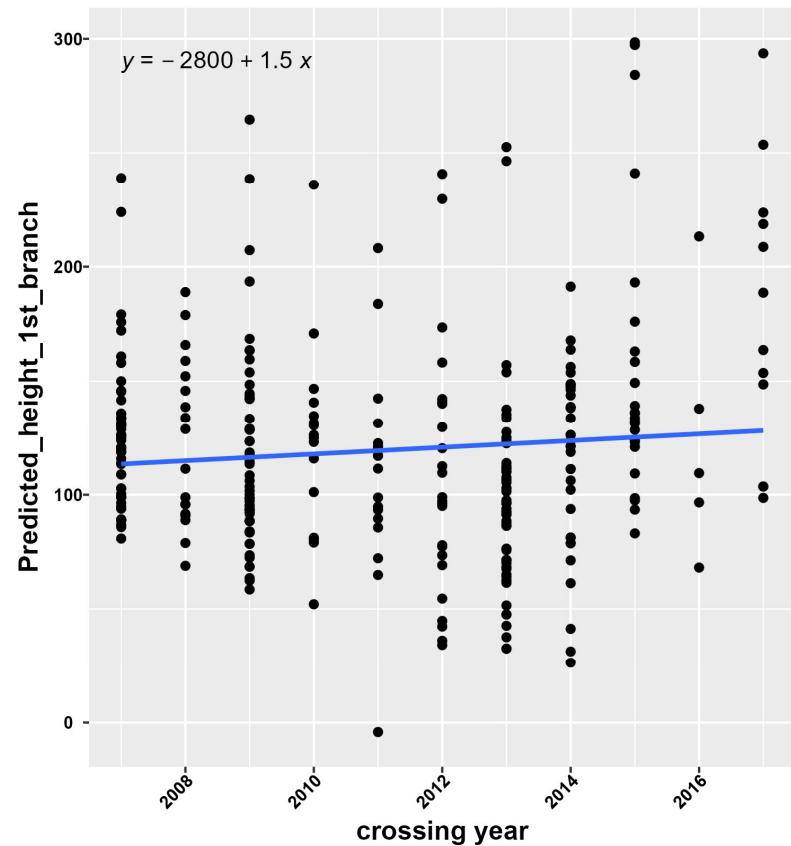


DM_gravity
0.55% by year



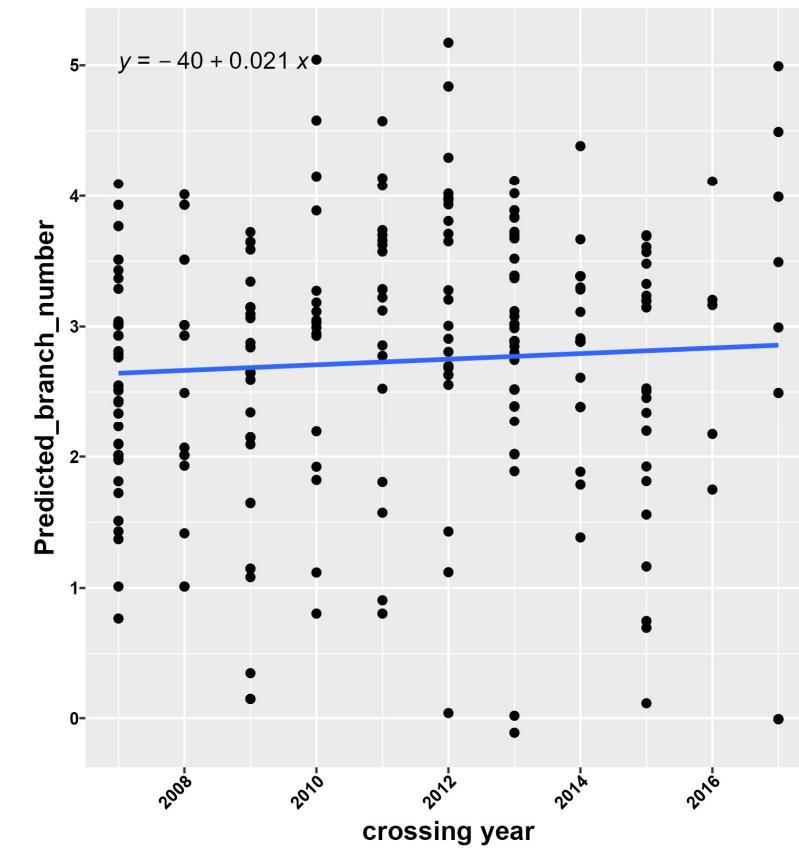
Betacarotenoids
7 % by year

Genetic gain



Height_1st_branch

1.30% by year



Branch_number

0.81% by year



Thanks!