

CORESTA Agro-Phyto Joint Study Groups Meeting 2017 (AP2017)

Insights of whole genomic studies in Burley recurrent selection program

Authors:

Pulcinelli C.E.; Ferreira R.A.C.D.; Boaretto L.F.; Barbosa P.K.A.; Weiss V.A.;
Padua J.M.V.

October 2017



Propose

Give some insights of Whole Genomic Prediction (WGP)
in the early steps of a burley recurrent selection program

Burley Varietal Group

❖ Industry importance:

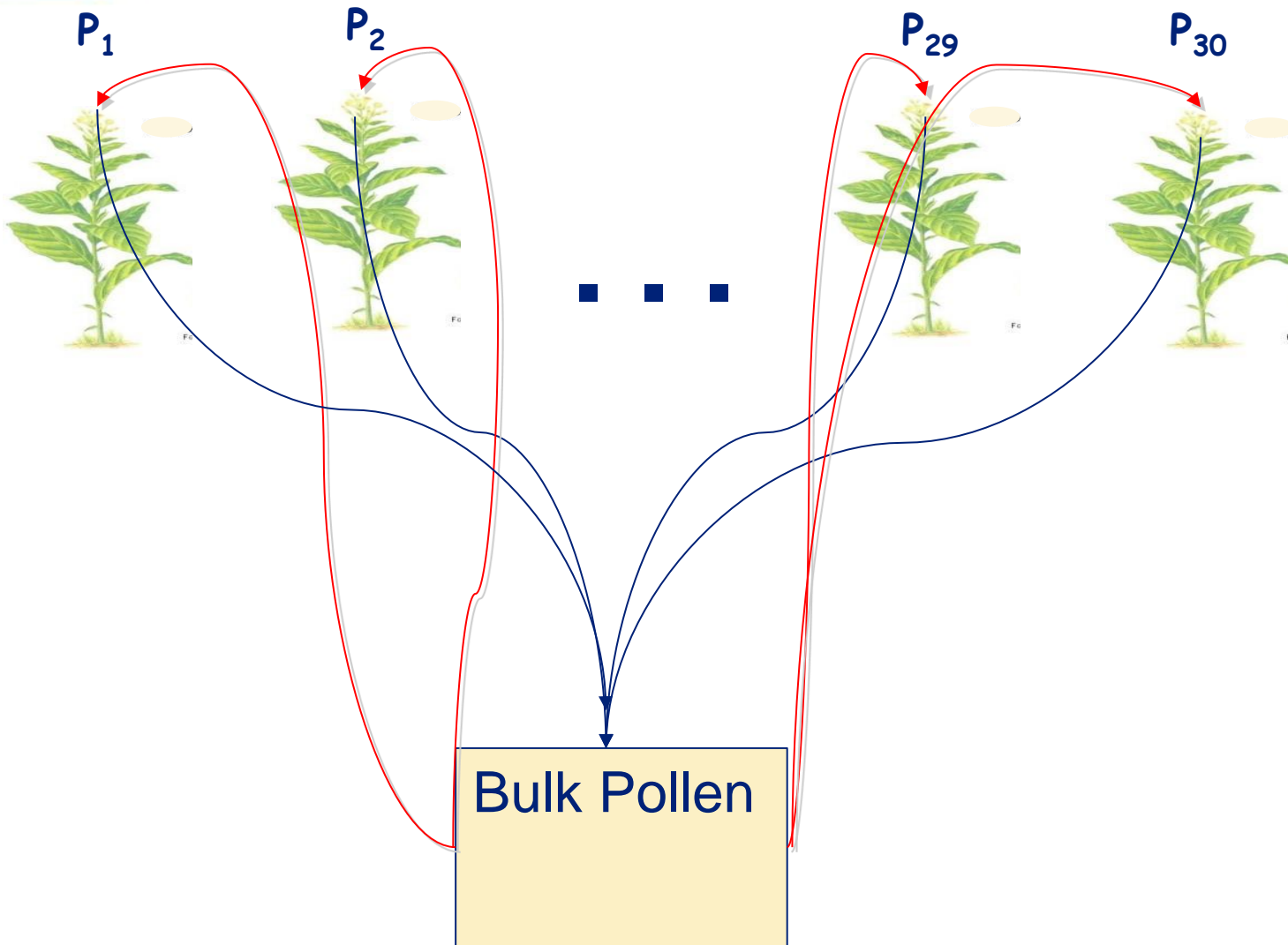
- Second group used in the blend (volume)

❖ Plant Characteristics:

- Air Cured



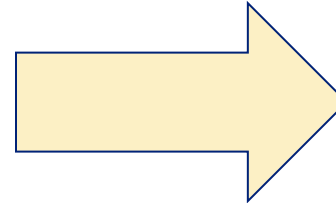
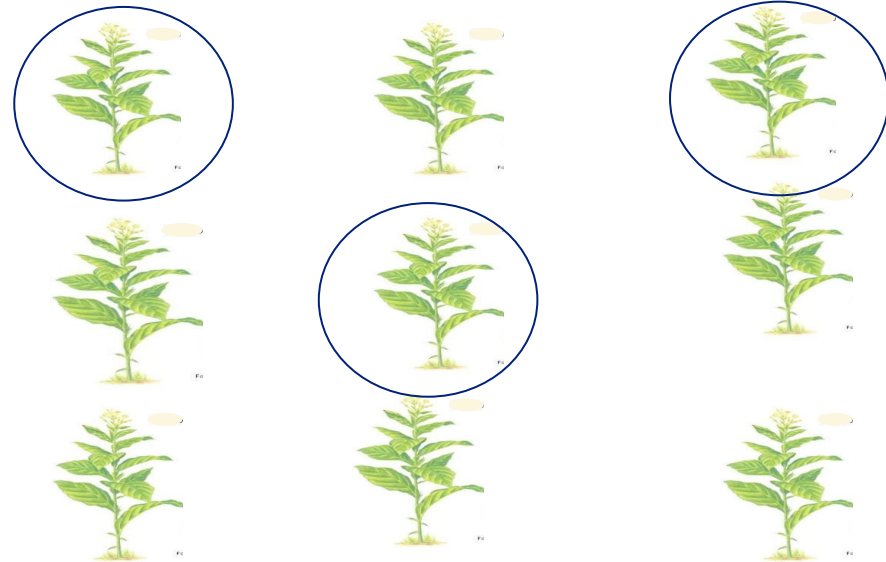
Crossings scheme – Year 2013/14



Seeds
of Pop
 S_0

S_{0:1} Progenies – Year 2015/16

Plants S₀



Progeny 1



Progeny 193

Progenies
S_{0:1}

Evaluation of S_{0:1} – Year 2016/2017

- 193 progenies were evaluated in 2 location using incomplete block design with 3 reps



Phenotypic Data

➤ Summary:

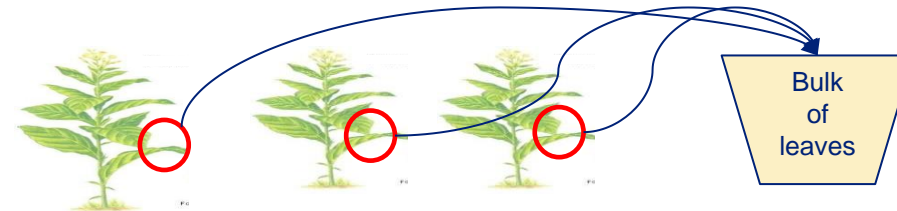
- ❖ 193 progeny $S_{0:1}$
- ❖ 2 Location
- ❖ Traits: 1) Yield (green leaf weight)
2) Total Alkaloids



Genotypic data

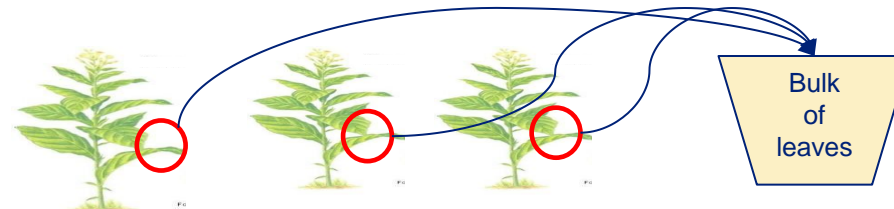
❖ Genotypic Data

Progeny 1



-
-
-

Progeny 193



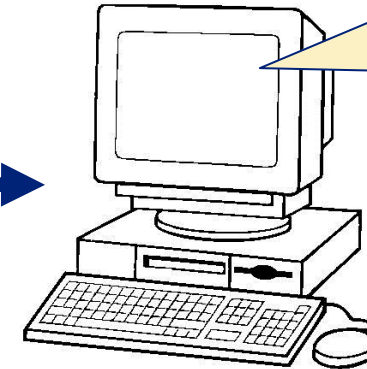
193 bulk of leaves were
shipped to CMT
(Tobacco Breeding
Centre) molecular lab
to DNA extraction and
to GBS sequencing

GBS Library

193
samples
DNA



MiSEQ



Bioinformatics Job:

- Align the sequence;
- Find SNPS

Digest DNA with
enzyme restriction

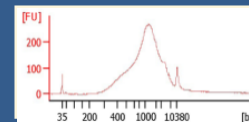


Ligation
Adapters (Ligase)



PCR Amplification

Validate Library



Adds index and adapters
modified

GBS adapted from Elshire et al., 2011

Bioinformatics

Read quality trimming

- Q20

cutadapt

<https://cutadapt.readthedocs.io/en/stable/#>

Sequence Alignment

Parameters:

- Alignment : very sensitive -D 15 -R 2 -N 0 -L 22 -i S,1,1.15

Bowtie2 v2.3.3



<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>

Bioinformatics

GBS analysis

Reference genome : *Nicotiana tabacum*, K326 cultivar
24 pseudo-molecules (diploid representation)
Genome map size: 4.6 GBases

Edwards et al. BMC Genomics (2017) 18:448
DOI 10.1186/s12864-017-3791-6

Tassel parameters:

- mnQS 20

- mnMAF
0.01

SNP filtering:
- callRate 0.7

Raw SNP number:
20.280

After filtering : 3185 SNPs



Whole Genomic Prediction

- Remembering: 193 progeny and 3185 SNPs
- Goal: Compare selection based on phenotype vs WGP

Statistical Models

■ Explicit Models

- ✓ Useful for generating ranking across all sites;
 - ✓ Allows for simplification of GxE term.

Explicit Phenotypic BLUP

Explicit Genomic-BLUP

■ Implicit Models

- ✓ Typically used to understand the dynamics of GxE

Implicit Phenotypic BLUP

Implicit Genomic-BLUP

Statistical Models : Explicit

❖ Phenotypic BLUP

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u}_{BLUP} + \mathbf{Z}\mathbf{ge}_{BLUP} + \mathbf{e}$$

Where:

- \mathbf{b} is the vector of fixed effects of environment, repetition within environment and block/rep/env;
- \mathbf{u}_{BLUP} is the vector of random progeny effect following $\mathbf{u}_{blup} \sim N(0, \sigma_{uBLUP}^2 \mathbf{I})$ which \mathbf{I} is an identity matrix;
- \mathbf{ge}_{BLUP} is the vector of random for genotypes x environment interaction effect following $\mathbf{u}_{blup} \sim N(0, \sigma_{geBLUP}^2 \mathbf{I})$ which \mathbf{I} is an identity matrix.

Statistical Models : Explicit

❖ GBLUP

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u}_{GBLUP} + \mathbf{Z}\mathbf{ge}_{GBLUP} + \mathbf{e}$$

Where:

- \mathbf{b} is the vector of fixed effects of environment, repetition within environment and block/rep/env;
- \mathbf{u}_{GBLUP} is the vector of random progeny effect following $\mathbf{u}_{GBLUP} \sim N(0, \sigma_u^2 \mathbf{G})$ which \mathbf{G} is the realized genomic relationship matrix of VanRaden (2008);
- \mathbf{ge}_{BLUP} is the vector of random for genotypes x environment interaction effect following $\mathbf{u}_{blup} \sim N(0, \sigma_{geGBLUP}^2 \mathbf{I})$ which \mathbf{I} is an identity matrix.

Statistical Models : Implicit

❖ GBLUP

$$Y = Xb + Zu_{env} + e$$

Where:

- b is the vector of fixed effects of environment, repetition within environment and block/rep/env;
 - u_{env} is random vector of genotype within environment where $u_{env} \sim N(0, A \otimes G)$ which A is an unstructured matrix of variance-covariances of genotype-by-environment and G is the realized genomic relationship matrix of VanRaden 2008;
- ❖ Phen. BLUP – Same model above but replacing G for a identity matrix

Statistical Models : Implicit

❖ GBLUP

$$Y = Xb + Zu_{env} + e$$

Where:

- b is the vector of fixed effects of environment, repetition within environment and block/rep/env;
 - u_{env} is random vector of genotype within environment where $u_{env} \sim N(0, A \otimes G)$ which A is an unstructured matrix of variance-covariances of genotype-by-environment and G is the realized genomic relationship matrix of VanRaden 2008;
- ❖ Phen. BLUP – Same model above but replacing G for a identity matrix

$$\begin{pmatrix} \sigma_{g1} & - \\ \sigma_{g12} & \sigma_{g2} \end{pmatrix}$$

Environment Correlation (Implicit models)

Phen. BLUP

$$r_{Blup}^2 = \frac{\sigma_{g12}^2}{\sqrt{\sigma_{g1}^2 * \sigma_{g2}^2}}$$

$u/env \sim N(0, A \otimes I)$

G-BLUP

$$r_{GBlup}^2 = \frac{\sigma_{g12}^2}{\sqrt{\sigma_{g1}^2 * \sigma_{g2}^2}}$$

$u/env \sim N(0, A \otimes G)$

$\begin{pmatrix} \sigma_{g1} & - \\ \sigma_{g12} & \sigma_{g2} \end{pmatrix}$

Heritability (Explicit Models)

Phen. BLUP

$$h_{Blup}^2 = \frac{\sigma_{uBLUP}^2}{\sigma_{uBLUP}^2 + \frac{\sigma_{geBLUP}^2}{2} + \frac{\sigma_e^2}{2 * 3}}$$

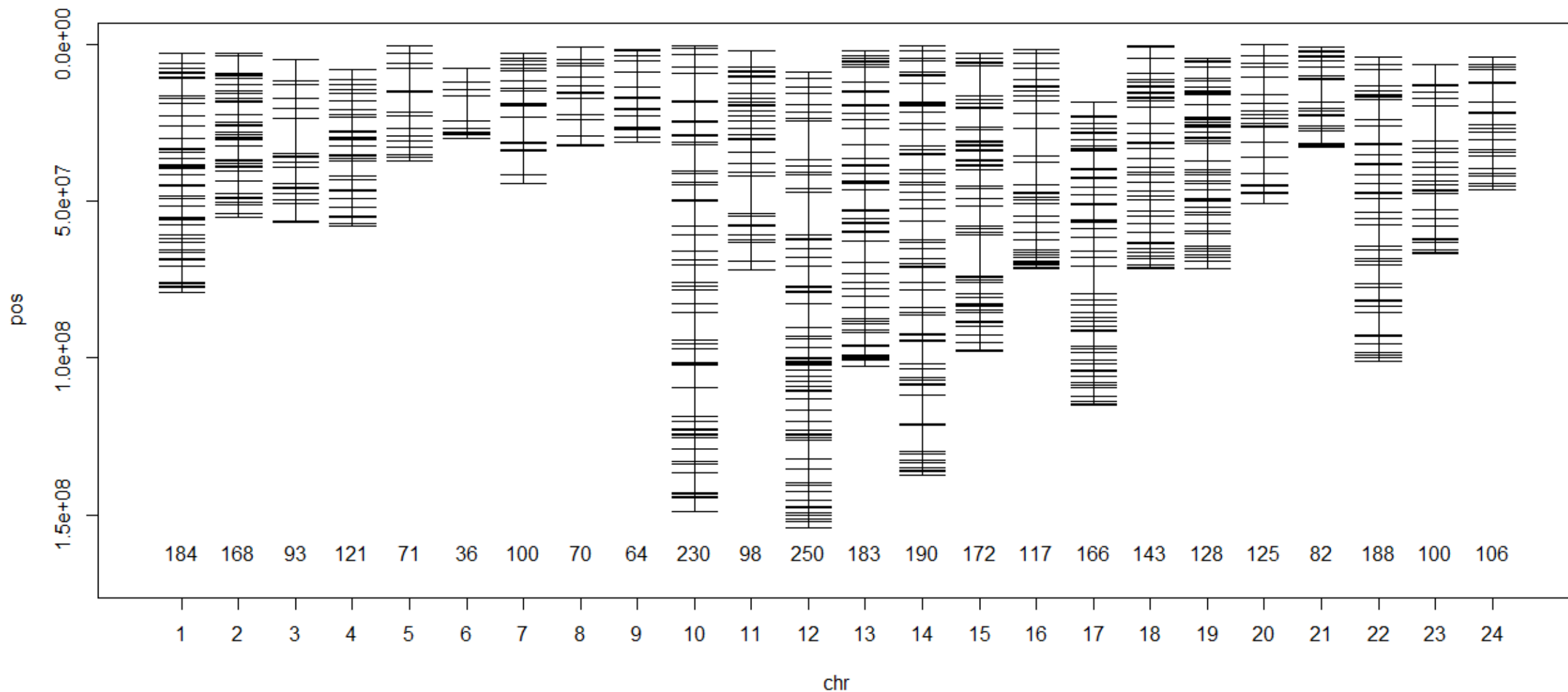
G-BLUP

$$h_{GBlup}^2 = \frac{\sigma_{uGBLUP}^2}{\sigma_{uGBLUP}^2 + \frac{\sigma_{geBLUP}^2}{2} + \frac{\sigma_e^2}{2 * 3}}$$



Results

Genetic Map Using The SNPs



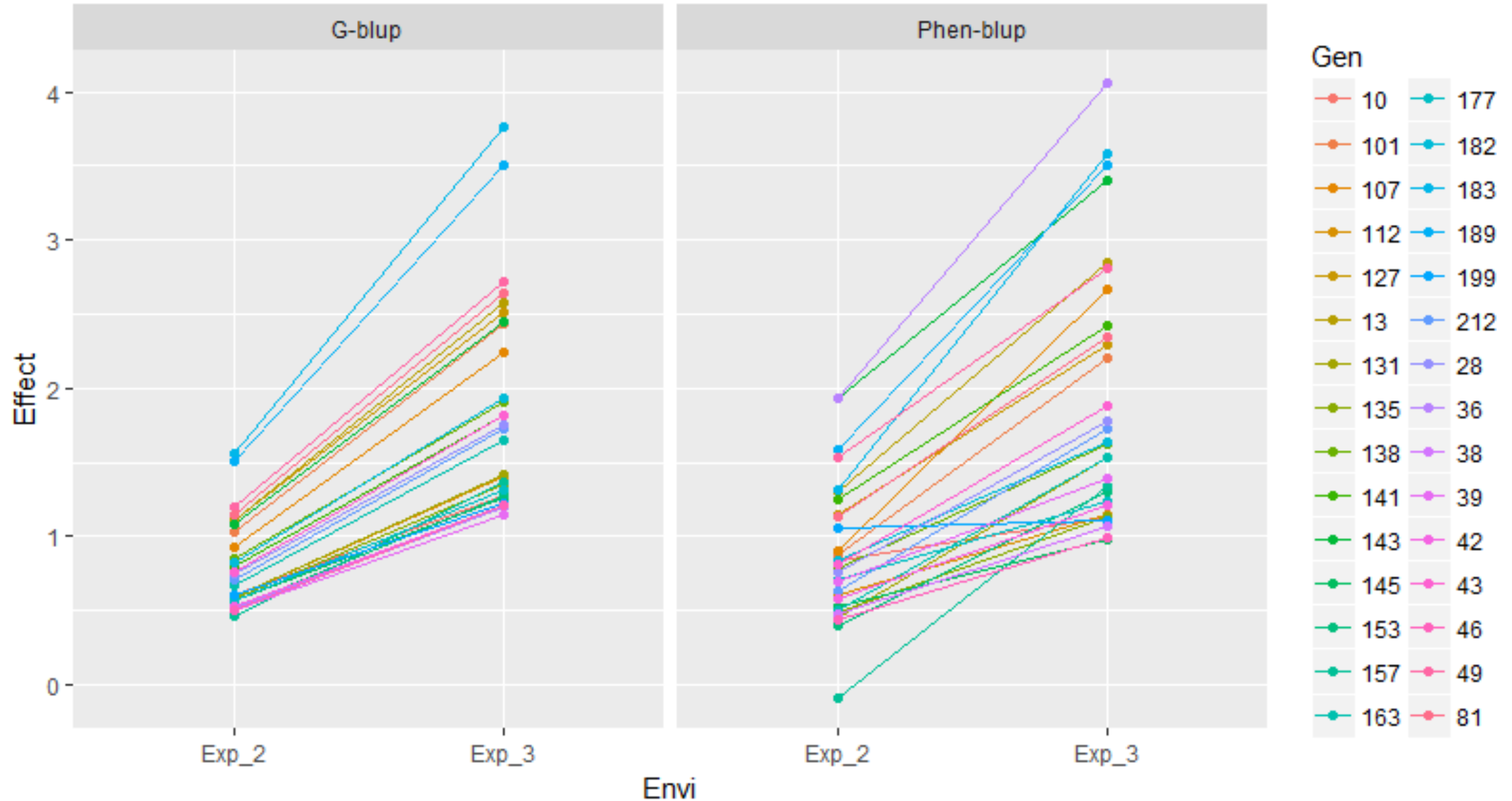
Correlation – Implicit models

1. Genetic Correlation between environments

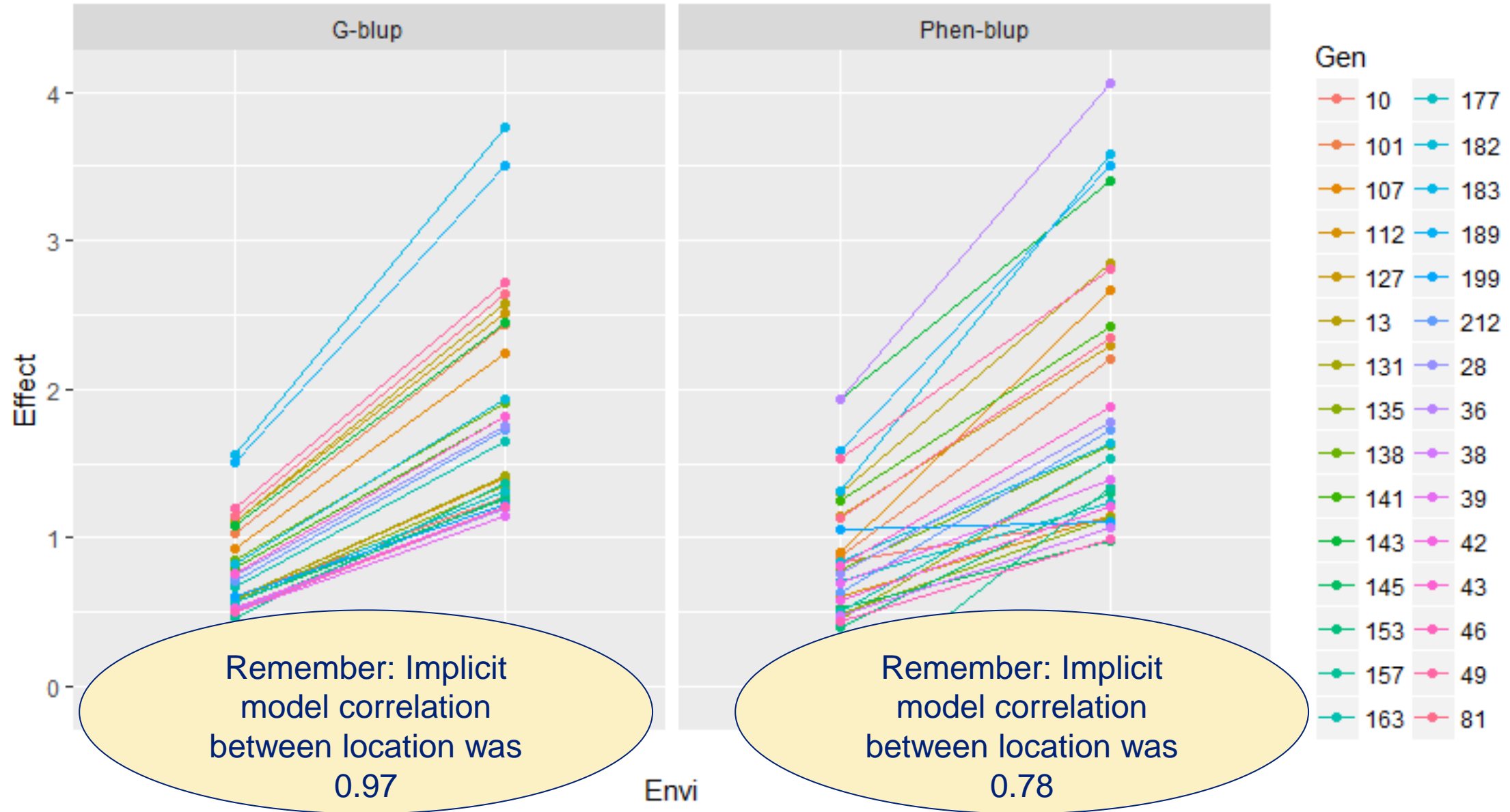
Model	Yield (kg/plot)	Alkaloid
G-BLUP	0.97	0.84
Phen. BLUP	0.78	0.75

- ✓ High correlation between environments – Non cross over interaction
- ✓ We can select rank across the environments
- ✓ G-BLUP had high correlation compared to Phenotypic BLUP

30 highest value ranked by G-BLUP mean of 2 location – Yield



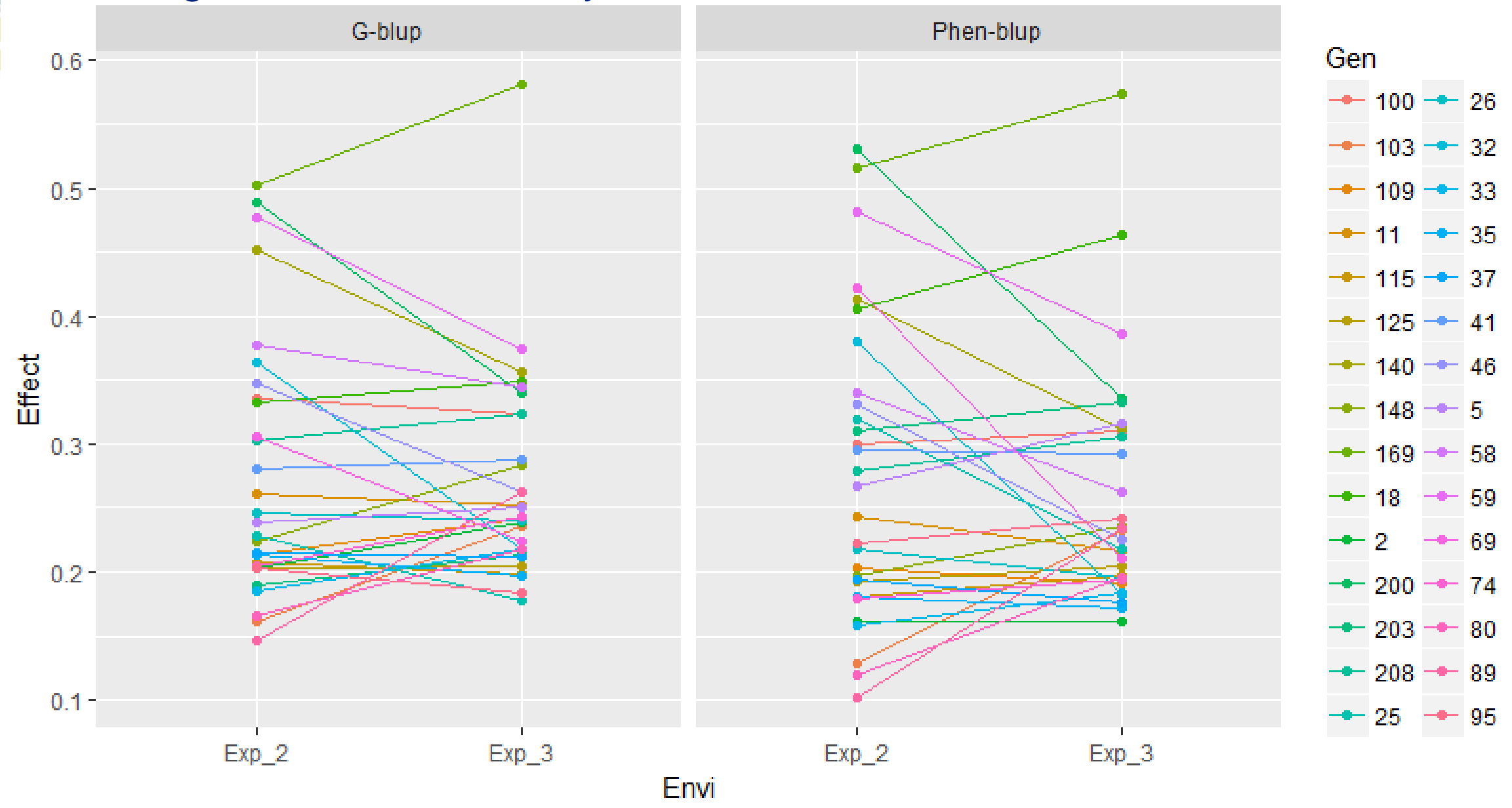
30 highest value ranked by G-BLUP mean of 2 location – Yield





CAN

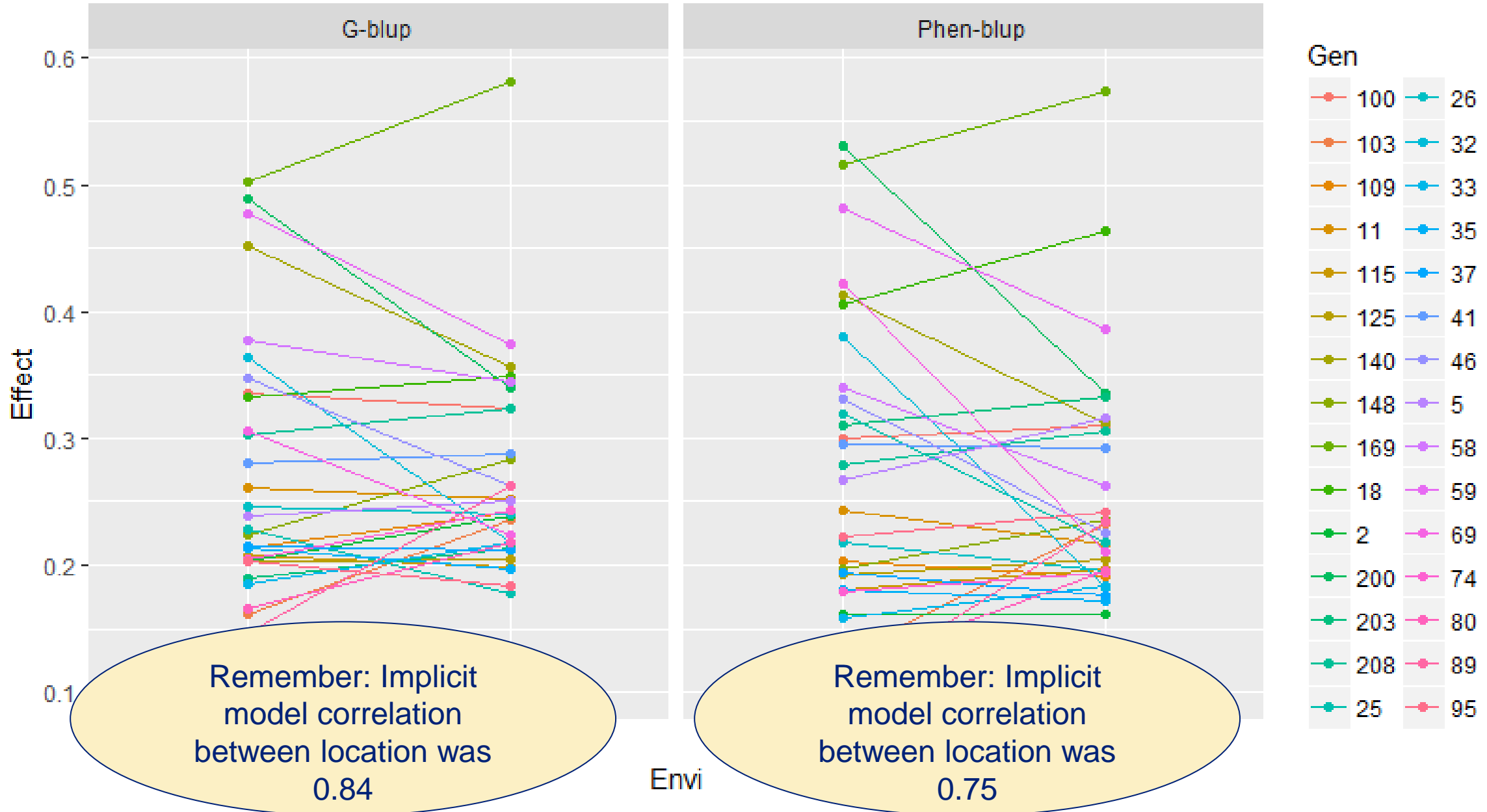
30 highest value ranked by G-BLUP mean of 2 location – Alkaloid





CAN

30 highest value ranked by G-BLUP mean of 2 location – Alkaloid



Explicit Model – Select Genotype Across Locations

2. Heritability mean of 2 locations

Model	Yield (kg/plot)	Alkaloid
G-BLUP	0.66	0.73
Phen. BLUP	0.55	0.61

- ✓ G-BLUP increase the heritability (INCREASE THE GENETIC GAIN) for both traits.

GBLUP vs Phen BLUP Coincidence

3. GBLUP vs Phen. BLUP Coincidence

Model	Yield (kg/plot)	Alkaloid
30 Highest	0.77	0.73
30 lowest	0.9	0.83

- ✓ The ranking are not the same
- ✓ The heritability of G-BLUP
- ✓ The genetic correlation between environments

GBLUP vs Phen BLUP Coincidence

3. GBLUP vs Phen. BLUP Coincidence

Model	Yield (kg/plot)	Alkaloid
30 Highest	0.77	0.73
30 lowest	0.9	0.83

77% of 30 best genotypes ranked by G-BLUP an Phen. BLUP are coincident

- ✓ The ranking are not the same
- ✓ The heritability of G-BLUP
- ✓ The genetic correlation between environments

Prediction

4. G-BLUP prediction

Number of prediction (individuals)	Yield (kg/plot)	Alkaloid
Predicting 30	0.28	0.30
Predicting 60	0.21	0.18
Predicting 100	0.12	0.11

Summary of Results

- The Implicit models is useful to plant breeders giving some perspectives about de GXE interaction;
- The G-BLUP mitigate the GXE (high genetic correlation) facilitating the selection across environments;
- The G-BLUP heritability for both traits were higher than Phen. BLUP increasing the genetic gain and accuracy for selection.
- The prediction of progeny base on molecular markers had poor performance. So we need to keep making good test varieties trials to have phenotype to make selection.

Thank you

- BAT – GLR&D – Global Leaf Research and Development

Ricardo Ferreira: Scientific Researcher (Plant Breeding)
at Souza Cruz/BAT

Contact: ricardo_ferreira@souzacruz.com.br