A journey from classical to modern cassava breeding: Integrating genomic selection for faster and greater genetic gains

Danilo E. Moreta, Xiaofei Zhang, & Cassava Team

Cassava Team Meeting

Montería, Colombia | 21st November 2023

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

Acknowledgements

Together
Everyone
Achieves
More

Bioversity-CIAT Cassava Team

Jonathan Newby + Research sub team + Management sub team (Ximena, Zulma, Oriana)

Breeding sub team

Xiaofei Zhang
Sean Fenstemaker
Luis Fernando Delgado
Lizbeth Pino
Sandra Salazar
Nelson Morante
Jorge Iván Lenis
Camilo Vargas
Ana María Ensuncho

Genetics sub team

Winnie Gimode
Adriana Bohórquez
Camilo Sánchez
Luz Andrea Gómez
Vianey Barrera (former)
Janeth Gutiérrez
Carmen Bolaños
Carlos Ordóñez

Quality lab sub team

Thierry Tran
Luis Fernando Londoño
María Alejandra Ospina
Jorge Luna
Jhon Larry Moreno
Christian Duarte

Seed systems sub team

Roosevelt Escobar
Adriana Núñez
Natalia Canacuan
Auradela Ríos
Carlos Dorado

Crop protection sub team

Wilmer Cuellar
Juan Manuel Pardo
María Isabel Gómez

Field workers

Outline

PART I: Background & context

- Transition from traditional to modern cassava breeding
  - Overview of the traditional cassava breeding method
    - Recurrent phenotypic selection
  - Genomic selection (GS) as a breeding tool
    - What it is? Why implement? How it works?

PART II: Applications & plans

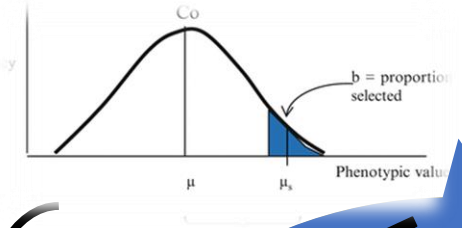
- Incorporating GS in CIAT cassava breeding pipeline
  - Work plan 2023-24, status, challenges & opportunities,
  - Discuss ideas & plans to optimize cassava GS models
  - A glance at the future: Hybrid cassava breeding

Part I

Transition from traditional to modern cassava breeding

The traditional & long cassava breeding cycle

Selections based on phenotypic values



Improve population

SELECT

CROSS

5-YEAR cycle

EVALUATE

Release product (cultivar)

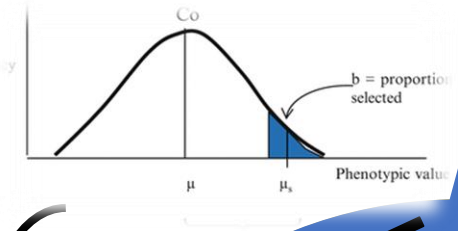
Image credits:
G. Acquaah, 2015; N. Morante; & D. E. Moreta

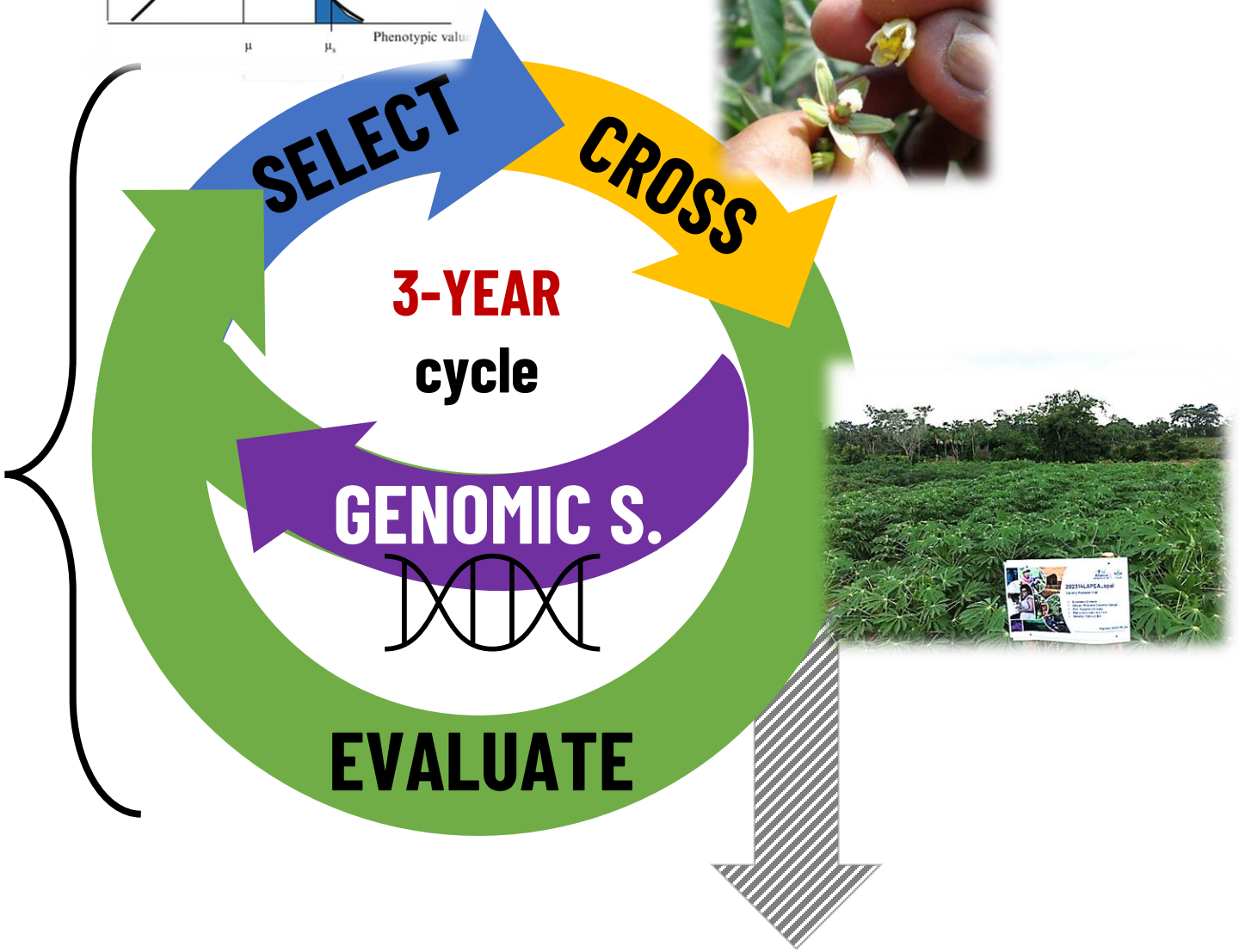
A shorter breeding cycle with genomic selection

Selections based on genomic estimated breeding values (GEBVs)

$$GEBV = f(\text{DNA})$$

Improve population





Release product (cultivar)

Image credits:
G. Acquaah, 2015; N. Morante; & D. E. Moreta

Genomic selection:
The next (r)evolution in plant breeding

$$\Delta G = \frac{i \cdot r \cdot \sigma_a}{t}$$

ΔG: Genetic gain per unit time
i: Intensity of selection
r: Accuracy of selection
σ_a: Additive genetic variation
t: Duration of the breeding cycle

Impacts of GS on the breeder's equation parameters

		Direct impact	Indirect impact
Introducing genomic information	Add genotyping to increase selection accuracy	r	i $1/t$
	Reduce phenotyping effort with genotyping	i $1/t$	
Management of genetic diversity	Better choice of parents to optimize crossbreeding or preserve genetic diversity	r σ_a	

Source: Fugeray-Scarbel et al., 2021

Alliance Bioversity & CIAT
CGIAR

Accelerated cassava breeding: The major benefit of genomic selection

Genomic Selection (GS, aka genomic prediction, genome-wide selection, genome-wide prediction, etc.) is a genomics-assisted breeding tool.

F_1: seedling nursery
F_1C_1: cloned seedling nursery
SRT: single row trial
PYT: preliminary yield trials
AYT: advanced yield trials
SIT: seed increase trial
TPY: training population yield trials
GWP: genome-wide prediction

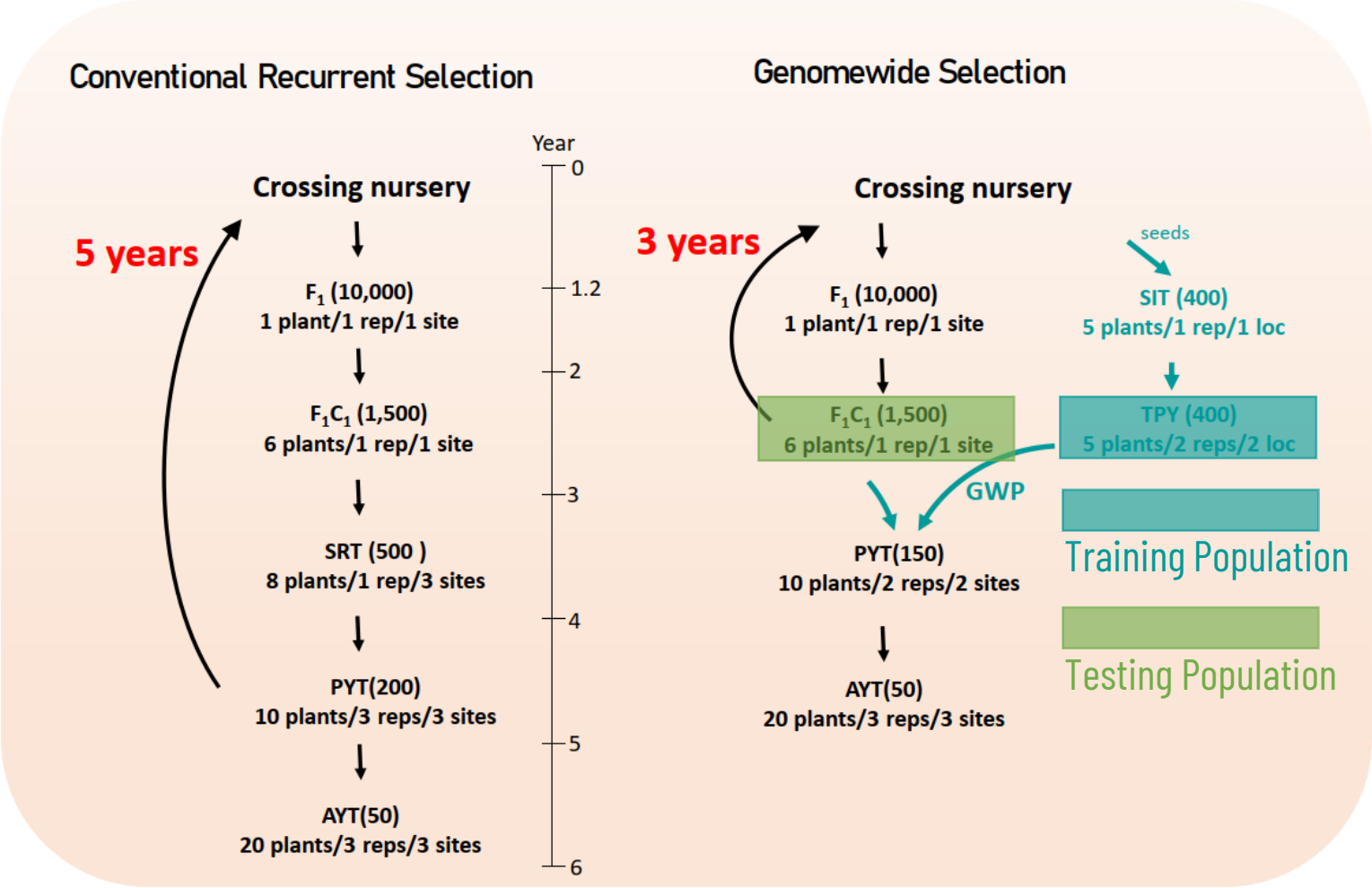


Figure credits: Cassava Breeding Team (Alliance Bioversity-CIAT)

How genomic selection works?

$$GEBVs = f(\text{DNA})$$

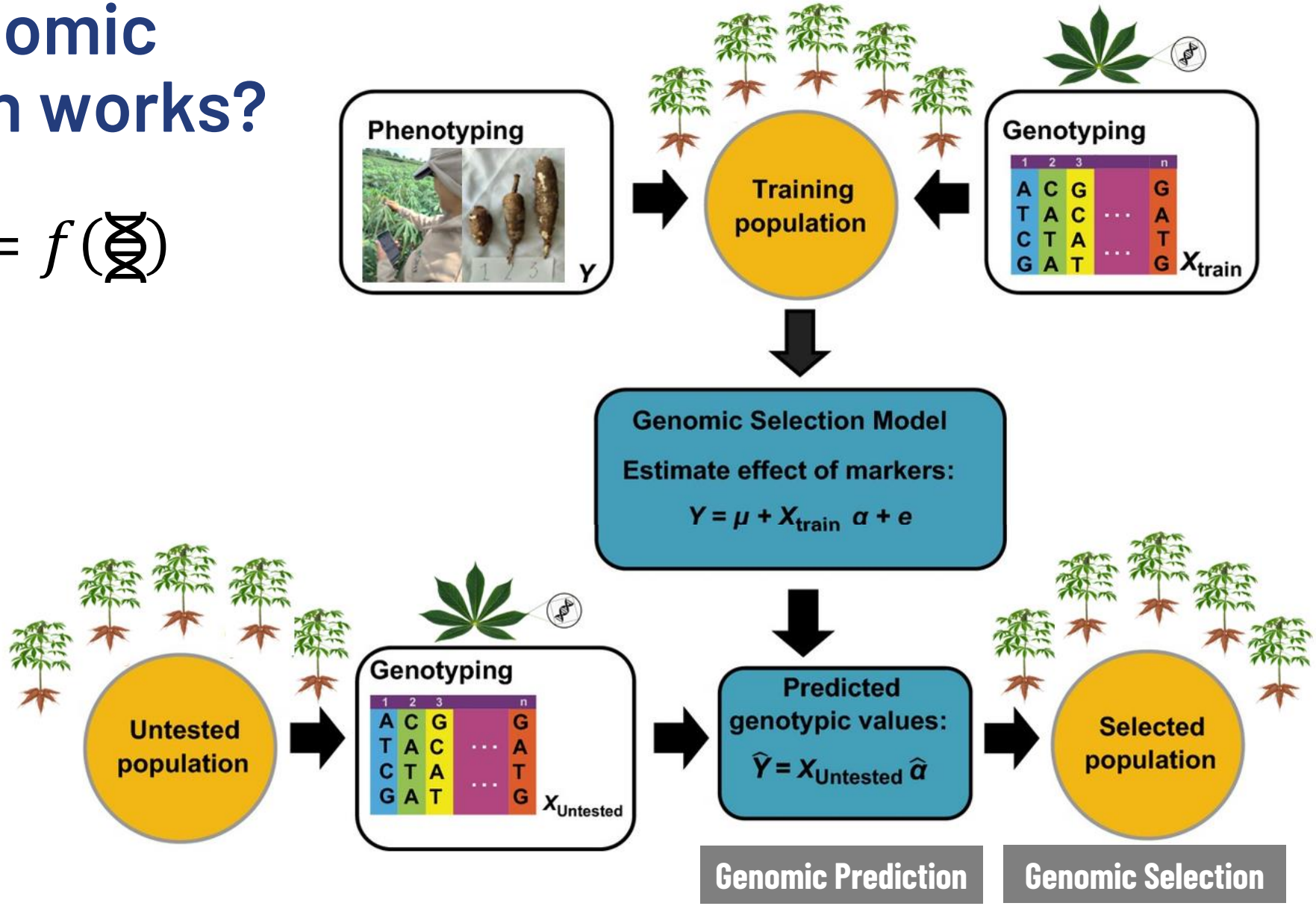


Figure adapted from Zhao et al., 2015 | Photo credits: S. Salazar

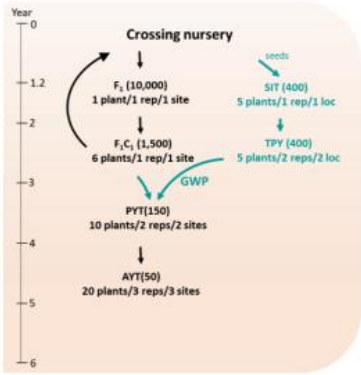
GS as a breeding tool to accelerate cassava genetic gains

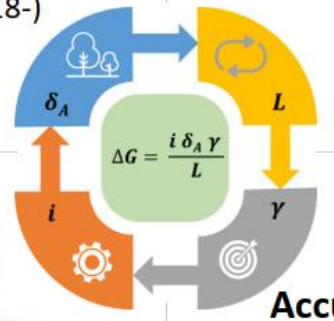
Genetic Diversity

- New traits, e.g., CBSD res., CMD res., good cooking quality et al.
- Sequencing of progenitors (2020-)
- Hybrid Breeding (2018-)

Duration of Selection Cycle

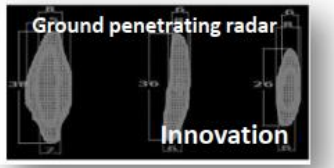
- Flower Inducing (2016-)
- Genomewide Prediction (2019-)





Intensity

- High throughput phenotyping



Accuracy

- CassavaBase, Fieldbook & Barcode (2018-)
- Quality control and MAS (2020-)
- TPE, ≥ 2 Environments (2020-)
- ≥ 5 Checks, BLUP and GBLUP (2020-)
- Selection Index (2012-)
- NIRS & Image Analysis (2012-)
- Stage&Gate System (2020-)
- Operational Excellence (2019-)

Figure credits: X. Zhang

Alliance Bioversity & CIAT

CGIAR

Why implement genomic selection?

$$GEBV = f(\text{DNA})$$

- Shorten lengthy breeding cycles
- Increase breeding efficiency: greater ΔG per unit time
- Seed shortages
- Discard poor clones earlier
- Reduce manual labor (phenotyping)
- Hard-to-measure traits

What factors affect GS accuracy?

$$\boldsymbol{Accuracy} = corr(phenotype, prediction)$$

- Marker density
- Size & composition of training population
- Number of QTLs
- Heritability
- Linkage disequilibrium (LD)
- Model used

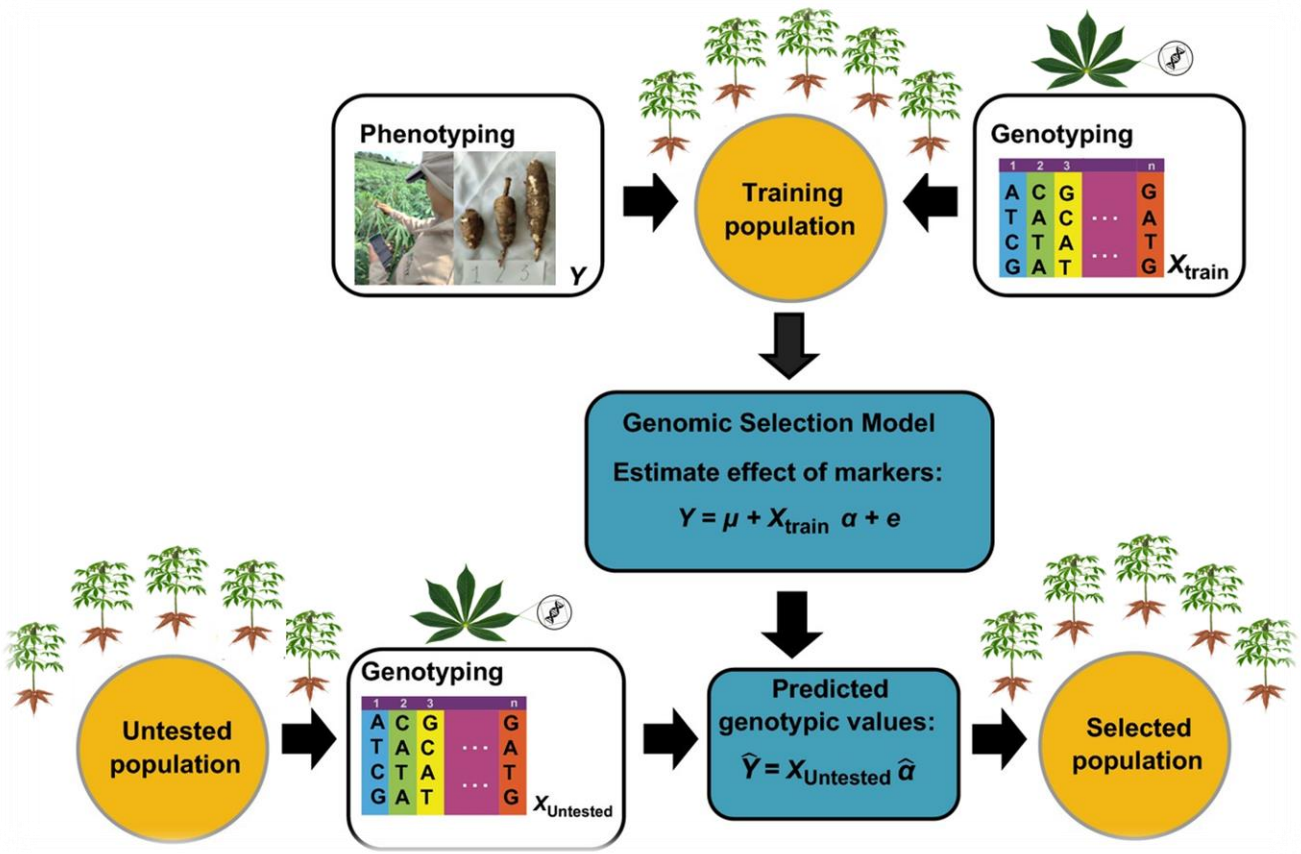


Figure adapted from Zhao et al., 2015 | Photo credits: S. Salazar

Genomic selection models

All models are approximations and hence wrong, but some are useful. ~ George Box

The basic genetic model: $\boldsymbol{P} = G + E + (GxE)$

$$GEBV_{GS} = f(\text{DNA})$$

Ridge-regression best linear unbiased prediction (RR-BLUP)

Assumption: Infinitesimal model of genetic architecture (all markers have an equal effect)

Step 1: Estimate marker effects in training population (TP)

Step 2: Use marker effects & genotypes of selection candidates to predict GEBVs

Genomic best linear unbiased prediction (G-BLUP)

Step 1: Use markers to quantify genomic relationships

Step 2: Use genetic relatedness to TP of unevaluated to predict GEBVs

Bayesian models

- Better model marker effects of differing sizes (Hayes, 2007)
- Separate variance estimated for each marker (Meuwissen et al., 2001)

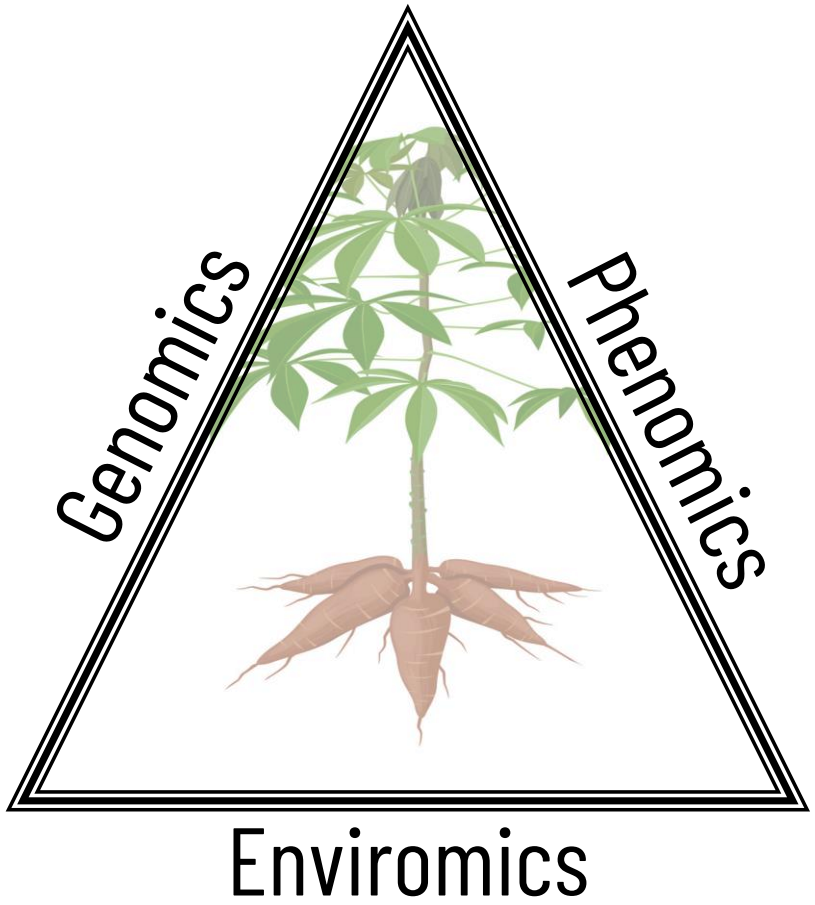
Modern plant breeding beyond molecular markers. Integrating and connecting disciplines

$$P = G + E + (GxE)$$

Phenomics

Genomics

Enviromics

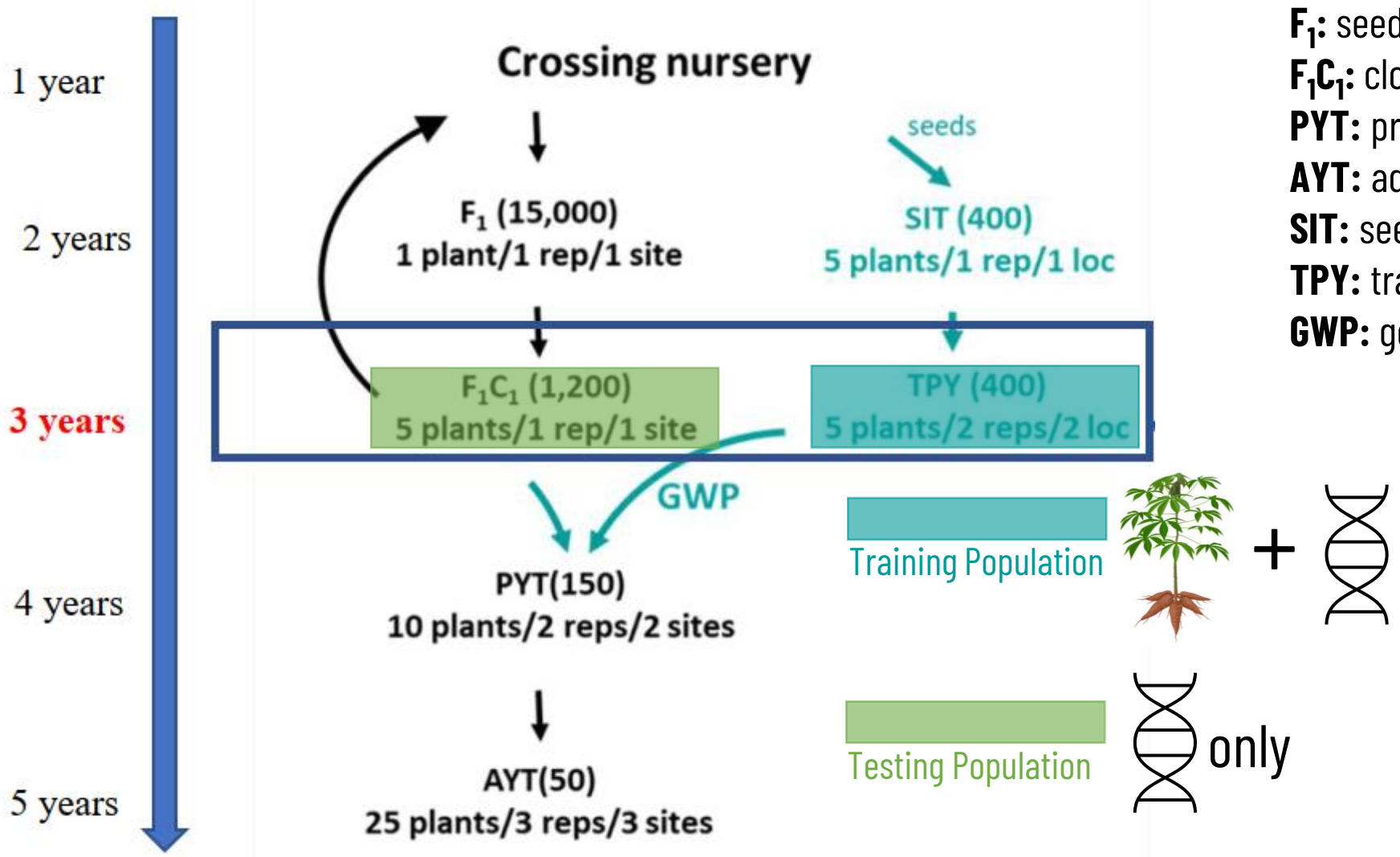


GS will not replace traditional breeding & field work, instead it will help optimize the system.

Part II

Incorporating GS in CIAT cassava breeding pipeline

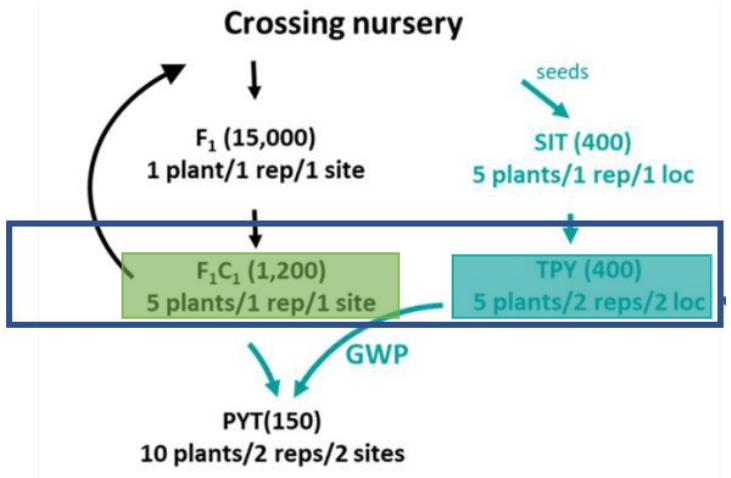
First steps to implementing GS in the cassava breeding pipeline at CIAT



F_1: seedling nursery
F_1C_1: cloned seedling nursery
PYT: preliminary yield trials
AYT: advanced yield trials
SIT: seed increase trial
TPY: training population yield trials
GWP: genome-wide prediction

Figure credits: Cassava Breeding Team (Alliance Bioversity-CIAT)

Training & testing (breeding) populations
C1.1 (Cycle 1, Cohort 1) | 2020-2021



Training Population Testing Population

Germplasm: Full-sibs

Trial names: 2021DVGST (n = 399) 2020DVF1C (n = 651)

Locations: 3 locs

2 in north coast*: Repelón, Momil | Ag traits + Pest & disease resistance

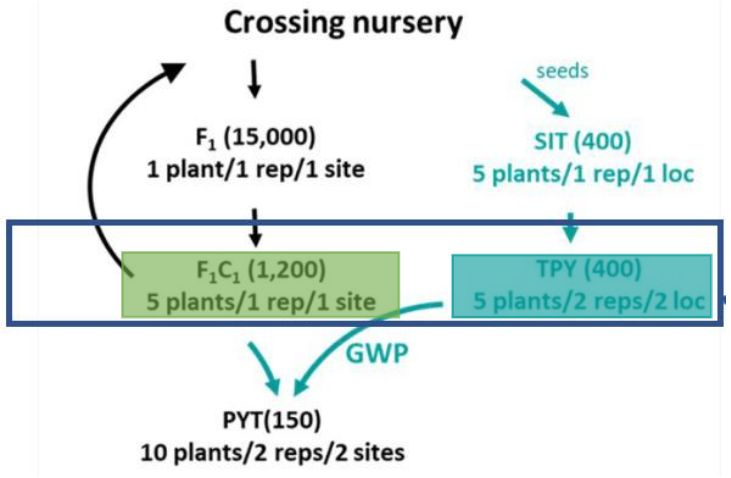
1 in Palmira | Only quality traits

---

*Semi-arid & sub-humid environments

Training & testing (breeding) populations
C1.2 (Cycle 1, Cohort 2) | 2021-2022

Training Population Testing Population



Germplasm: Full & half-sibs

Trial names: 2022DVGST (n = ~873) 2021DMF1C (n = ~823)
2021CQF1C (n = ~672)

GS for waxy cassava (COMING SOON!)
n = ~200 clones
Ingredion agreement

Locations: 3 locs

2 in north coast*: Santo Tomás, Momil | Ag traits + Pest & disease resistance

1 in Palmira | Only quality traits

---

*Semi-arid & sub-humid environments

Genotyping the training & testing (breeding) populations

C1.1 (Cycle 1, Cohort 1) | 2020-2021

$$Phenotype = Genotype + Environment$$

10X WGS

Whole genome sequencing for 1052 materials

399 training population materials

TPY (400)
5 plants/2 reps/2 loc

651 testing population.

F_1C_1 (1,200)
5 plants/1 rep/1 site

About 230K high-quality SNPs were selected to compute the genomic estimated breeding values (GEBVs) for 21 traits.

Figure credits: C. Vargas

Alliance Bioversity & CIAT

CGIAR

Prediction ability

C1.1 (Cycle 1, Cohort 1) | 2020-2021

$$Accuracy = corr(phenotype, prediction)$$

Cross-Validation

How well do we predict individuals without phenotypes?

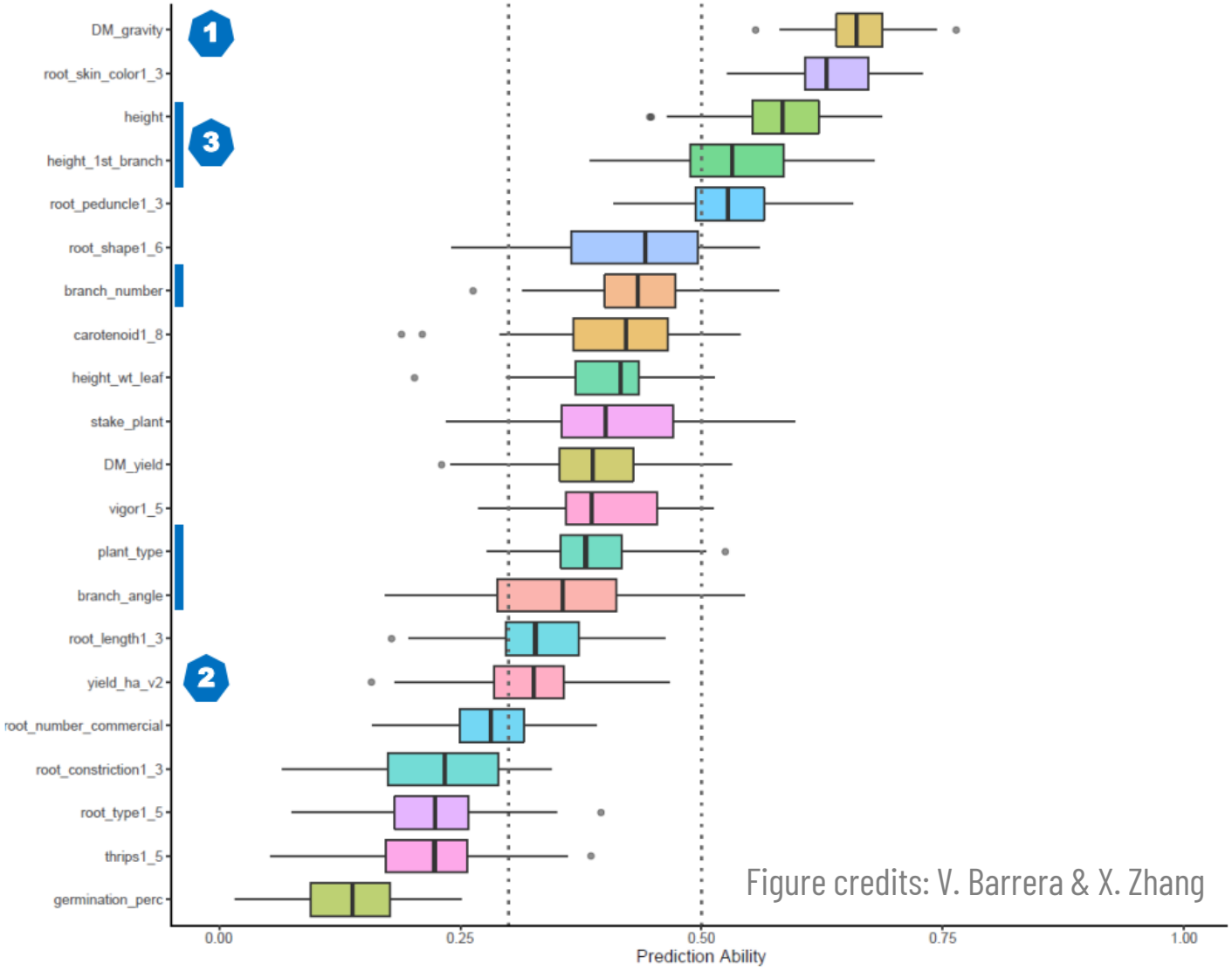
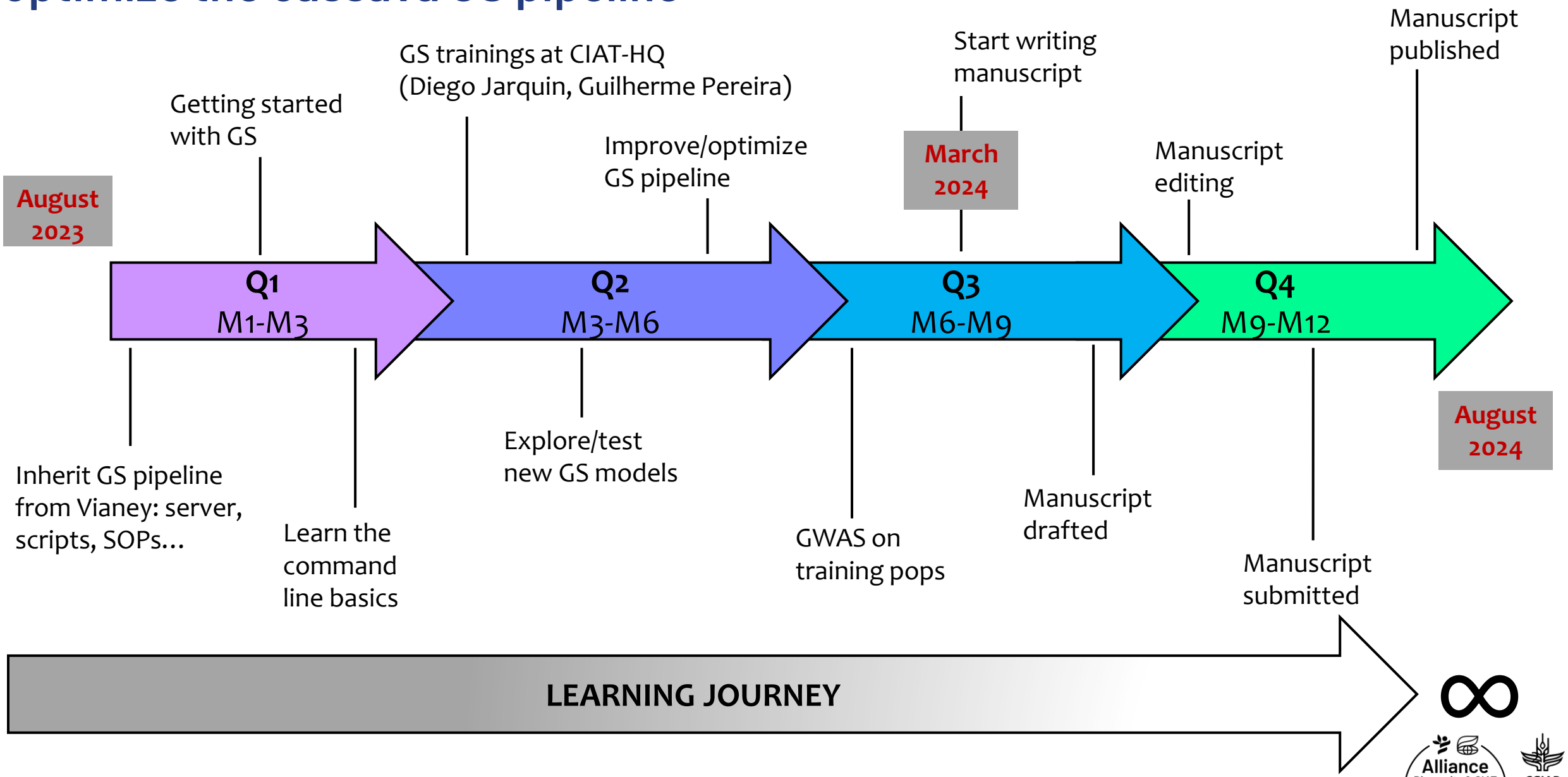


Figure credits: V. Barrera & X. Zhang

My timeline/roadmap to implement & optimize the cassava GS pipeline

Q: quarter of a year (3-month period)
M: month



Status: Trying to reproduce what others did

VCFtutorial_statsVisualization

$dEmc^2$

Cali, CO | 17th October 2023

GUIDE/REFERENCE: https://speciationgenomics.github.io/filtering_vcfs/

$Notes_{dEmc^2}$:

DNA sequencing info

10x whole genome sequencing

VCF file source: **Group1_11.chromosome01.vcf**

$n_{samples}$ = 395

n_{SNPs} = 588,247 (original/raw)

2021: Group 1-8

Group

Genom

Learning by doing!
Running demos & generating reports

GP_GS_demo: rrBLUP

$dEmc^2$

Cali, CO | 25th October 2023

GUIDE/REFERENCE: **GS_GEBVs_Prediction_Tutorial_dEmc2.jpynb**

$Notes_{dEmc^2}$:

Working files

1. $Genotypes_{SNPs}$ = **gs_2023.reduced.txt** | Should contain training (gsT) + breeding pops (gsB)
2. $Phenotypes_{BLUPs}$ = **GS_2023_second_cohort_blups.csv** | Should be only gsT
3. $Accessions_{names}$ = **group12_group12_accessions.csv** | Use to match genos & phenos

NEED TO TRACE BACK ALL THESE FILES IN VIANEY'S FOLDERS. They a

Alliance Bioversity & CIAT
CGIAR

Operational challenges & opportunities

Challenges	Opportunities
Unorganized files on server (hard to track and find)	Systematic organization of files on server
Lack of repository for GS scripts and related	Feed/update **Cassava2050 GitHub repository** (reproducible research)
IT micromanagement & lack of communication with researchers	Creation of a **server committee**: “Cassava byters” (Sean, Xiaofei, Winnie, Camilo, Danilo)

Ideas to optimize (make it more accurate) the cassava GS models

$$\boldsymbol{P} = G + E + (GxE)$$

Let's not miss these parts!

$$GEBV_{GS} = f(\text{DNA})$$

Phenos & environment

- Plug environmental covariates: weather & soil
- Account for GxE effects x
- Combine populations to increase power
- Include secondary traits (easier to predict?)
  - Relative yield
  - Total disease resistance
  - Harvest index, etc.
- High-throughput phenotyping for better model training (ongoing collab. w/ Mike Selvaraj)

DNA markers

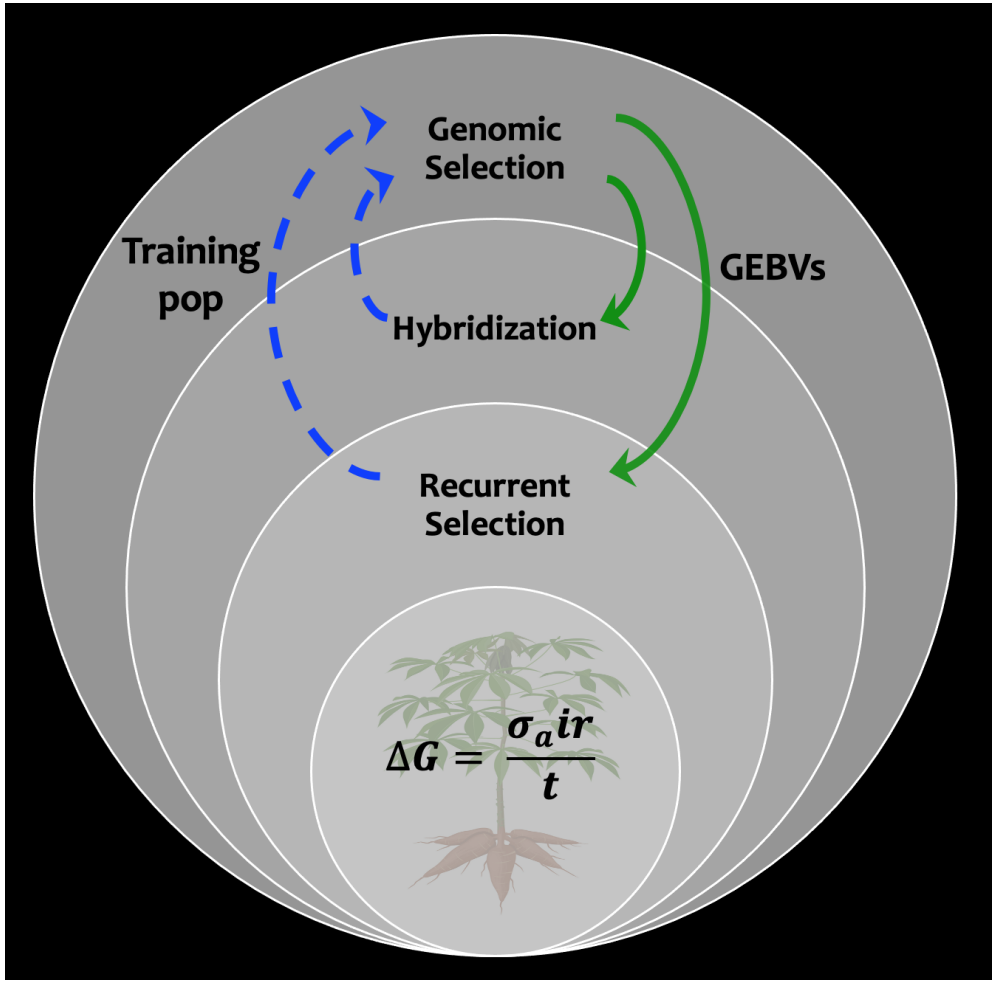
- Try different TP sizes & number of markers
- Dual purpose of TP: predictions + association (GWAS)
- Special treatment to some markers
  - Most significant & GWAS hits as fixed effects

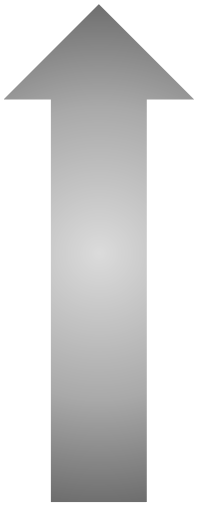
Moving up cassava improvement & production to the next level through hybrid breeding

How can GS help achieve this?

Inbreds: Combining ability

Hybrids: Performance





Efficacy of ΔG per unit time and cost

Alliance
Bioversity & CIAT

CGIAR

Thanks!