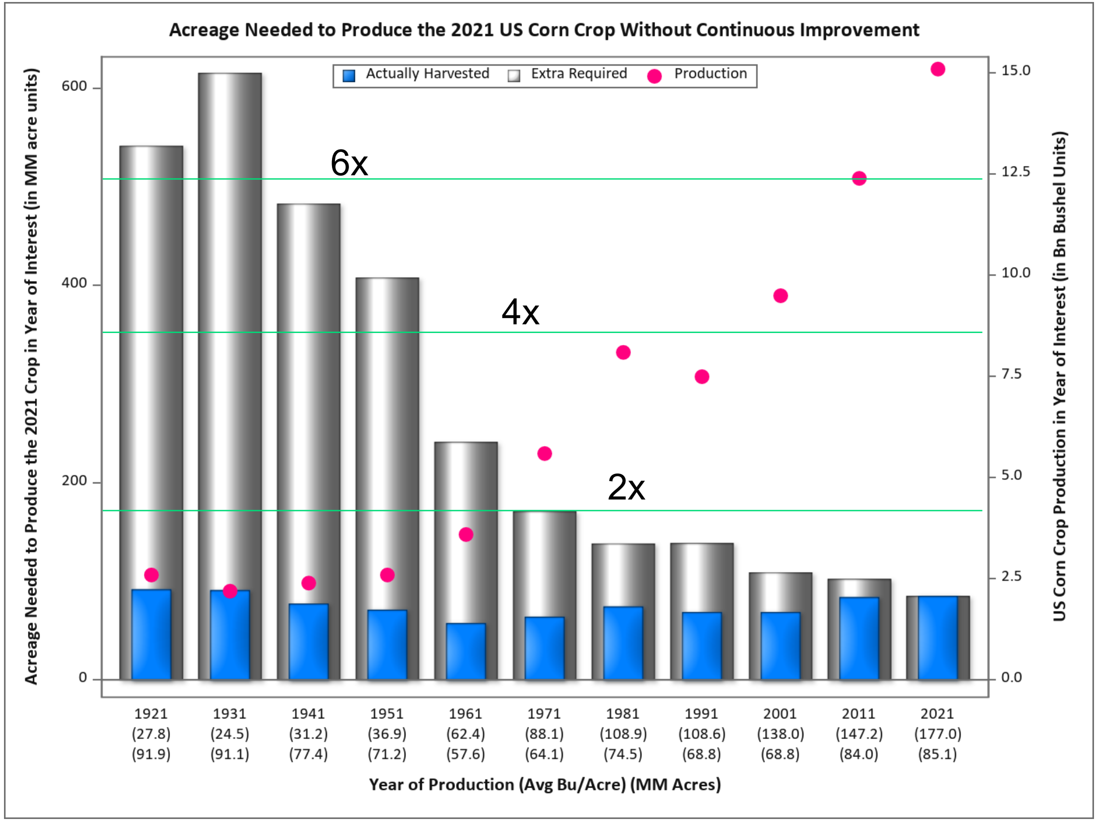
# Intro to Genomic Breeding

**Alencar Xavier**
Breeding Analyst at Corteva
Adjunct professor at Purdue
https://alenxav.github.io/

# What are some implications of continuous Corn Improvement?

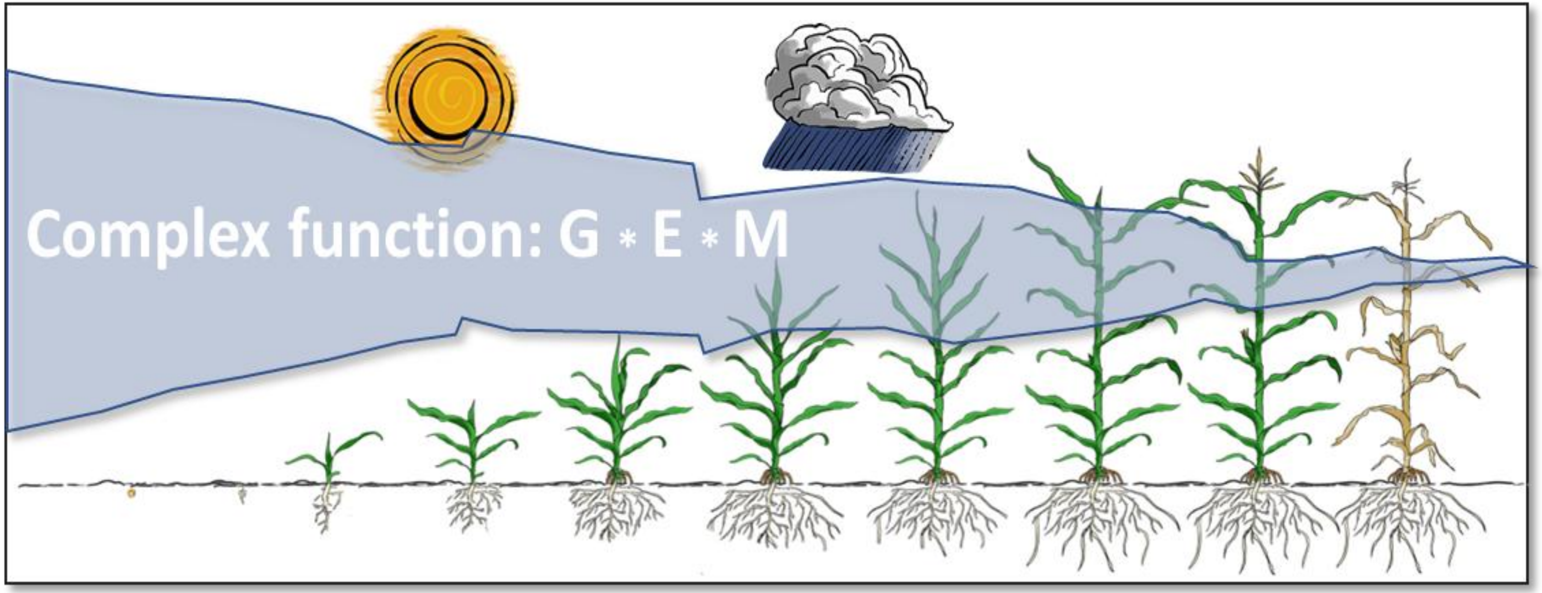



Source: Totir 2021, ASTA

*Based on 2021 USDA NASS data


Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

# Challenges in Corn Improvement

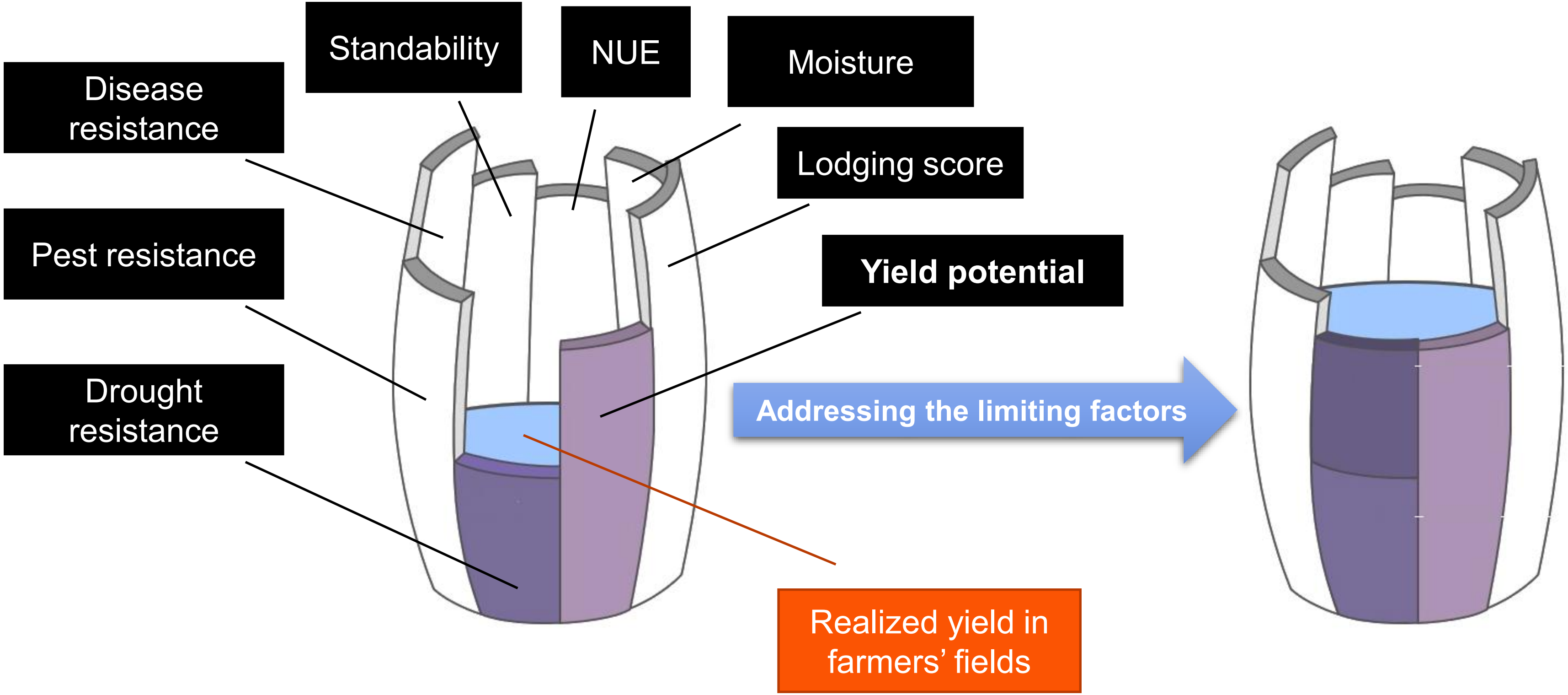



**G = Genetics; E = Environment; M = Management**

Source: Totir 2021, ASTA

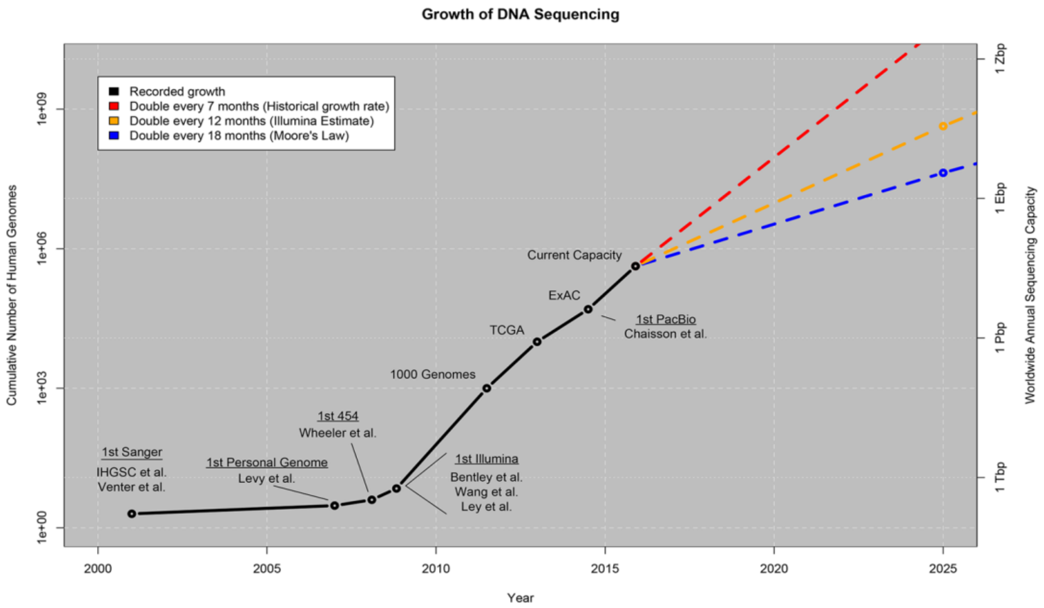

**Alencar.Xavier@Corteva.com**
**Quantitative Geneticist, Breeding Analyst LAAF**

# Law of the minimum

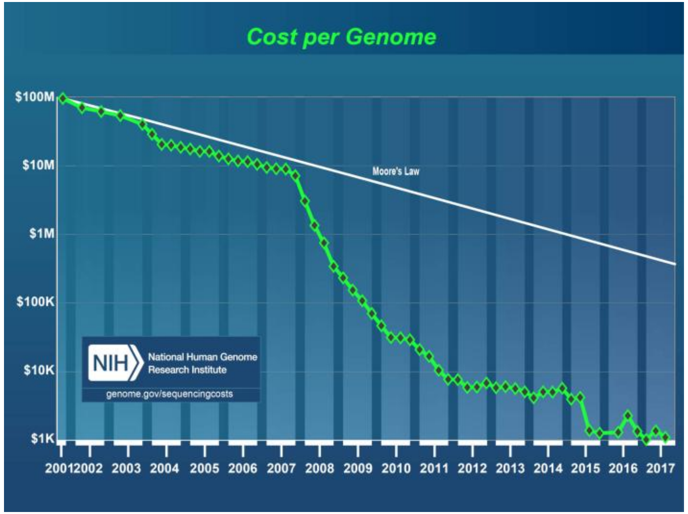





Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF



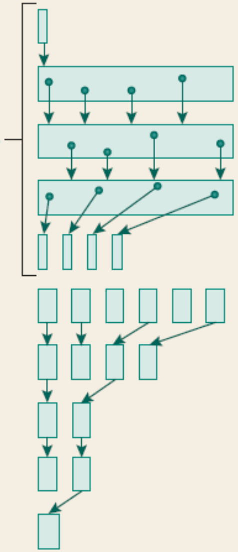

Stephens, Z. D.et al. (2015). Big data: astronomical or genomical? *PLoS biology*, *13*(7), e1002195.




The Cost of Sequencing a Human Genome. NIH. https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/

# PLANT BREEDING PIPELINE

Example from one program, one geography - Varietal Wheat



| Product development | | | |
|---|---|---|---|
| **Year** | **Generation** | **Number of plants** | **Action** |
| 1 | $F_1$ | 124 half-sib families | Increase in greenhouse |
| 2 | $F_2$ | 1,000 plants per family | Bulk 50 plants per family |
| 3 | $F_3$ ×124 | 1,000 plants per family | Bulk 50 plants per family |
| 4 | $F_4$ | 1,000 plants per family | Derive new lines from 50 plants per family |
| 5 | $F_{4:5}$ | 6,200 headrows | Advance 1,000 lines |
| 6 | PYT, $F_{4:6}$ | 1,000 lines | Yield trial, genotype |
| 7 | AYT, $F_{4:7}$ | 100 lines | Yield trial |
| 8 | EYT, $F_{4:8}$ | 10 lines | Yield trial |
| 9 | EYT, $F_{4:9}$ | 10 lines | Yield trial |
| 10 | $F_{4:10}$ | 1 line | Release variety |

Hickey et al. (2017) *Nature genetics* 49(9):1297

Main types of plant breeding

1. Varietal (soy, wheat)
2. Hybrid (corn, sunflower)
3. Clonal (potato, sugar cane)
4. Population (alfalfa, red clover)

CORTEVA agriscience
Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

# “Breeding objective”

$f\,(\text{market segment, farming systems})$

- Set of traits of interest (**TOI**) bred into a

WHAT

Yield, moisture, maturity, disease resistance, stability, producibility

- Target population of genetics (**TPG**) for a given

WHO

Corn 110-112, soybean MG2, winter wheat

- Target population of environments (**TPE**) and management (**TPM**) practices

WHERE

HOW, WHEN

Drought, irrigation, early planting, varying levels of disease pressure, different soil types

$$\rho_{\mathrm{GxExM}} = \rho_{\mathrm{TPG}} \times \rho_{\mathrm{TPE}} \times \rho_{\mathrm{TPM}}$$


Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

# What is genomic prediction?

| DATASET | GENOTYPES | PHENOTYPES |
|---|---|---|
| TRAINING POPULATION | YES | YES |
| PREDICTION TARGET | YES | NO |

**Purpose of GS:**

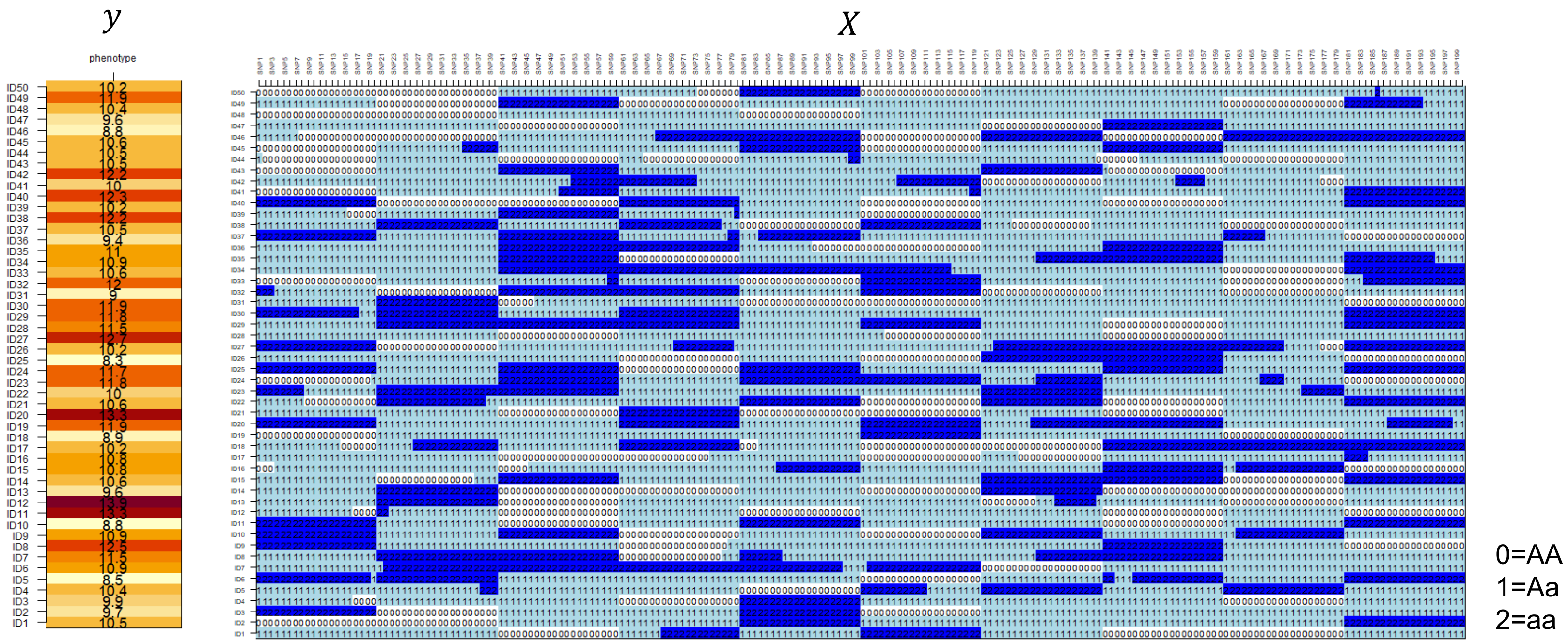
- Improve selection accuracy
- Select material without phenotypes
- Selection of new parents
- Prediction of cross combinations
- Optimize resources
- Stability and genetic architecture

**Genomic prediction accuracy is a function of**

- Heritability of the trait
- Relationship training-prediction population
- GxE between training-prediction environment


Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

# How does GS work?




Regression problem: $y = \mu + X\beta + \epsilon$


Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

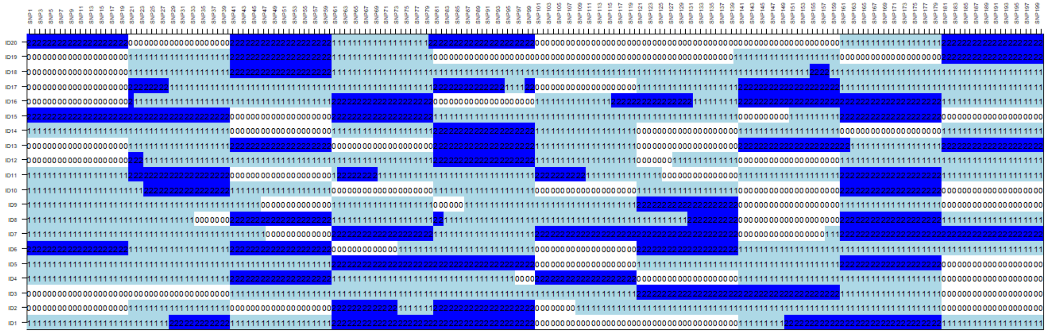
# How does GS work?

Regression problem:

$$y = \mu + X\beta + \epsilon$$

```
> fit = bWGR::mrr(Y,X)
> fit$mu # intercept
[1] 10.80188
> head( fit$b ) # coefficients
              [,1]
[1,] 0.01093033
[2,] 0.01169121
[3,] 0.01169235
[4,] 0.01126376
[5,] 0.01126586
[6,] 0.01126662
> fit$h2 # heritability
[1] 0.1917349
```

New set of individuals, genotyped but not phenotype?



Regression model (SNP-BLUP)

Intercept: $\hat{\mu}$

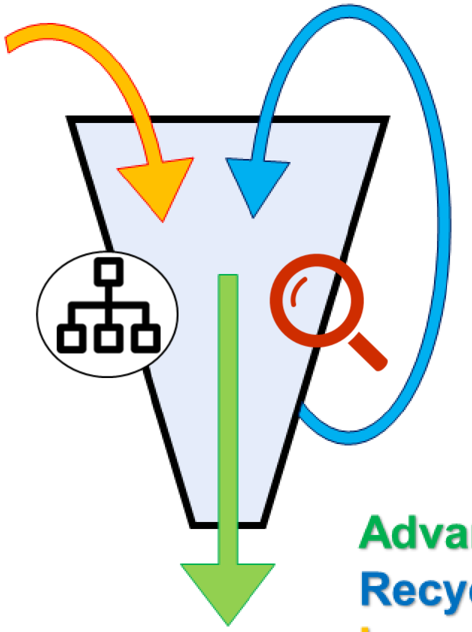
Coefficients: $\hat{\beta}$

Heritability: $\hat{h}^2$

Breeding values: $\hat{g} = X\hat{\beta}$

$$\boxed{\hat{g}_{new} = \hat{\mu} + X_{new}\hat{\beta}}$$

# Where is genomic information used for breeding?



Advancement
Recycling
Incorporation

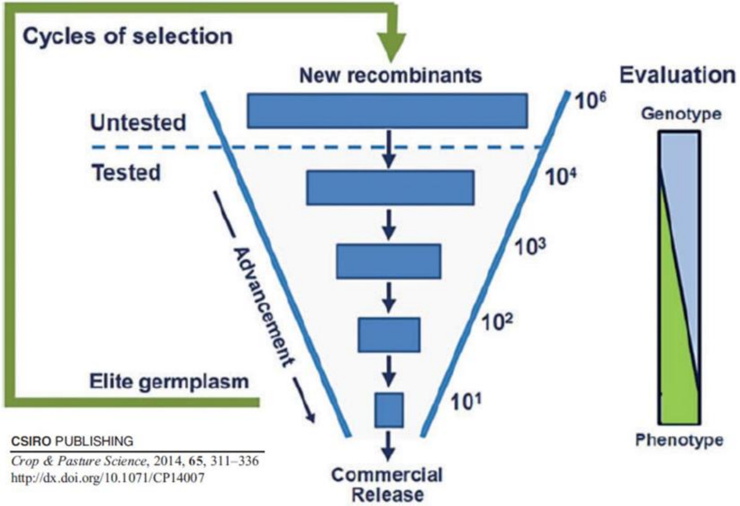
Monitoring

Classification

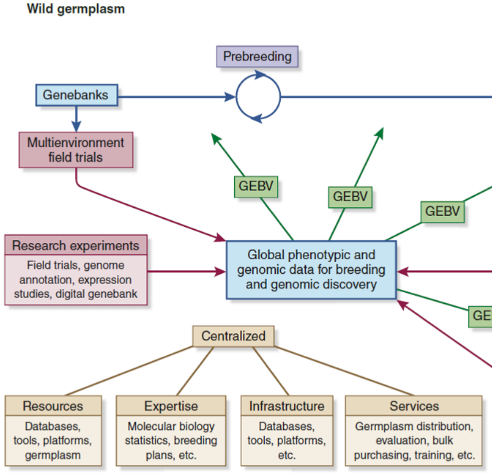


Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF

# Training population theory




CSIRO PUBLISHING
*Crop & Pasture Science*, 2014, **65**, 311–336
http://dx.doi.org/10.1071/CP14007




Hickey et al. (2017) *Nature genetics* 49(9):1297

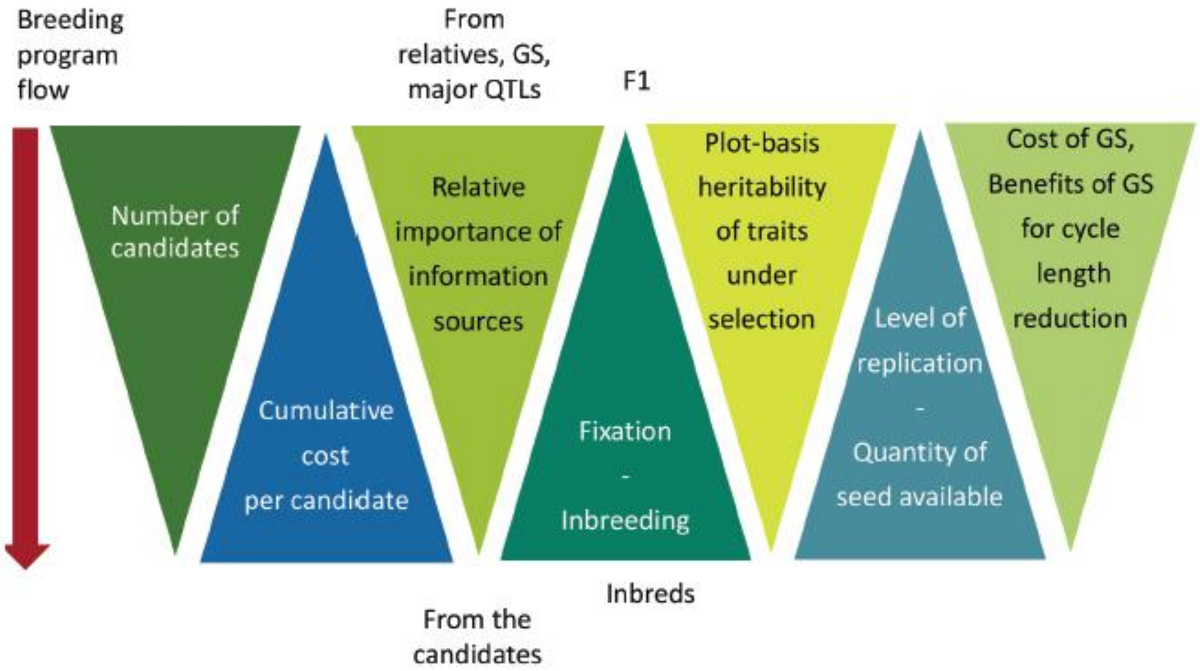



Figure 1. Key parameters and changes during a breeding cycle, to consider in implementing genomic selection (GS). The triangles indicate increase or decrease of the quantity considered. QTL, quantitative trait loci.

Heslot, N., Jannink, J. L., & Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Science*, *55*(1), 1-12.

# Concluding Remarks

1. GS is utilized differently for advancement, recycling and incorporations
2. Experimental settings and breeding design play key role in GS
3. Breeding pipeline is dynamic and constantly improved

**Thank you for your attention!**

**Questions??**

***Alencar Xavier***
alencar.xavier@corteva.com
https://alenxav.github.io/

# References and additional information

- Variance components - Purdue lecture 2022
- Walking through the statistical black boxes of plant breeding
- Bases for Genomic Prediction
- Genomic selection in plant breeding: from theory to practice
- Perspectives for Genomic Selection Applications and Research in Plants
- Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery
- Prediction-based breeding: Modern tools to optimize and reshape programs
- Genomic selection and reproducibility - are complex models distracting us from true scientific validity?
- Optimizing Plant Breeding Programs for Genomic Selection
- Expanding genomic prediction in plant breeding
- Statistical Approach for Improving Genomic Prediction Accuracy
- Megavariate methods capture complex genotype-by-environment interactions


Alencar.Xavier@Corteva.com
Quantitative Geneticist, Breeding Analyst LAAF