Rate of response to selection (genetic gain): concept, methods and recommendations

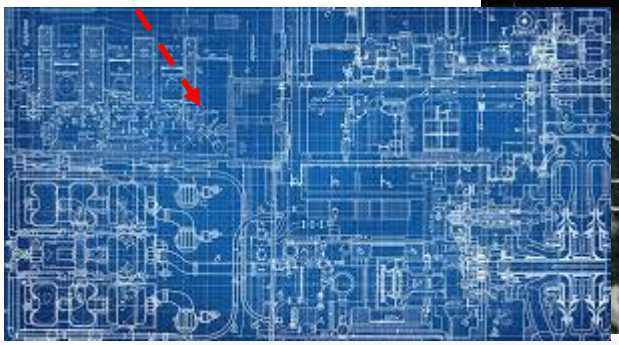
November 18th, 2020

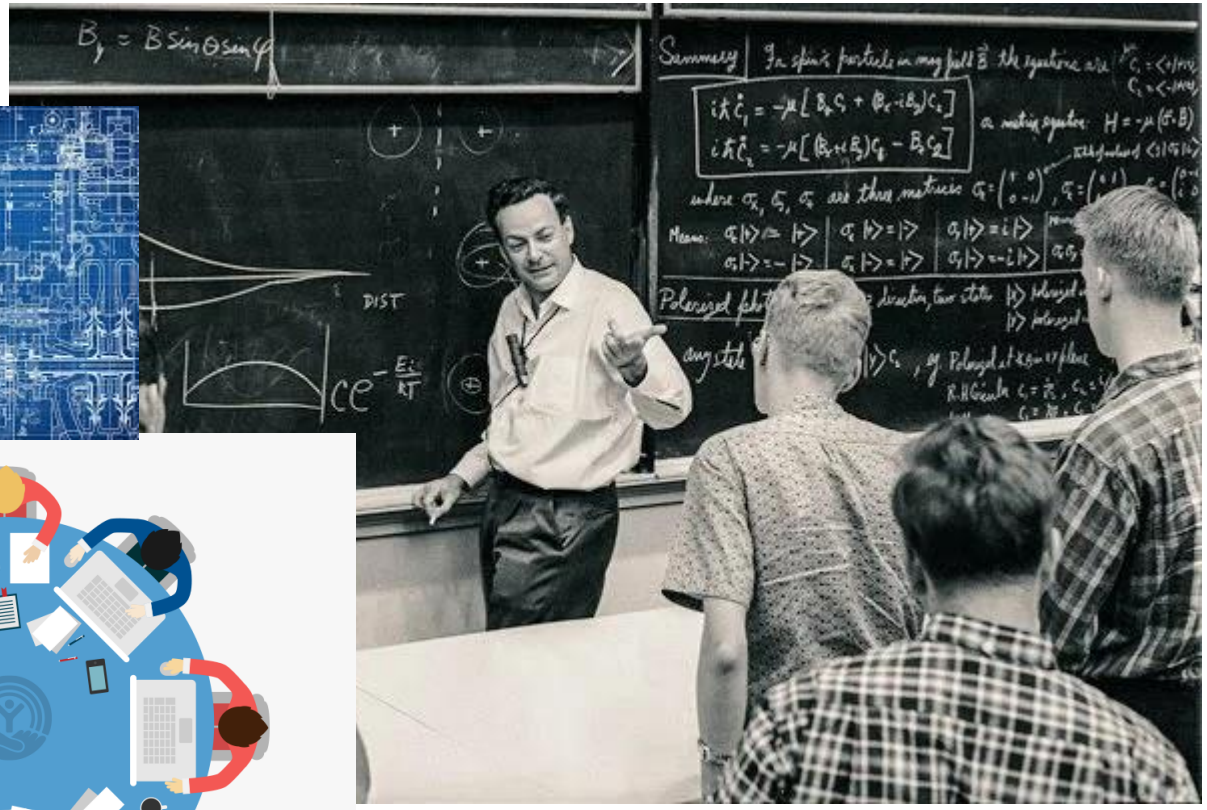
Giovanny E. Covarrubias-Pazaran
Breeding scheme optimization lead

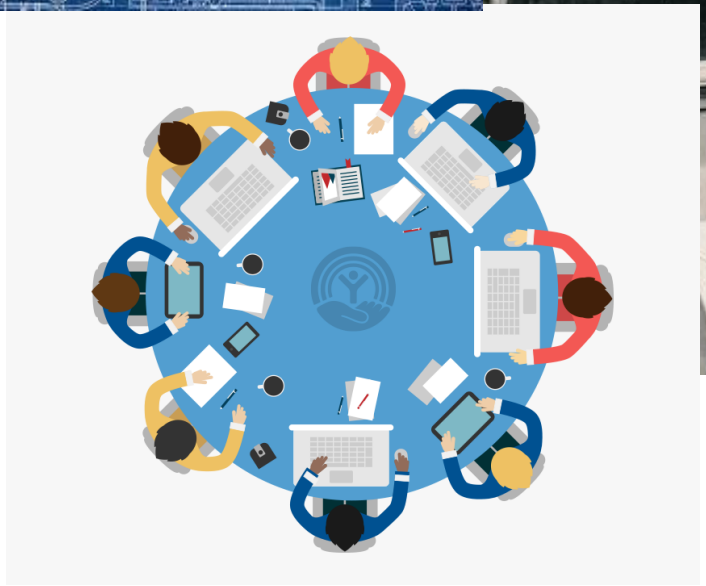
CGIAR

Excellence in Breeding PLATFORM

You better ask if there's doubts







CGIAR

Excellence in Breeding PLATFORM

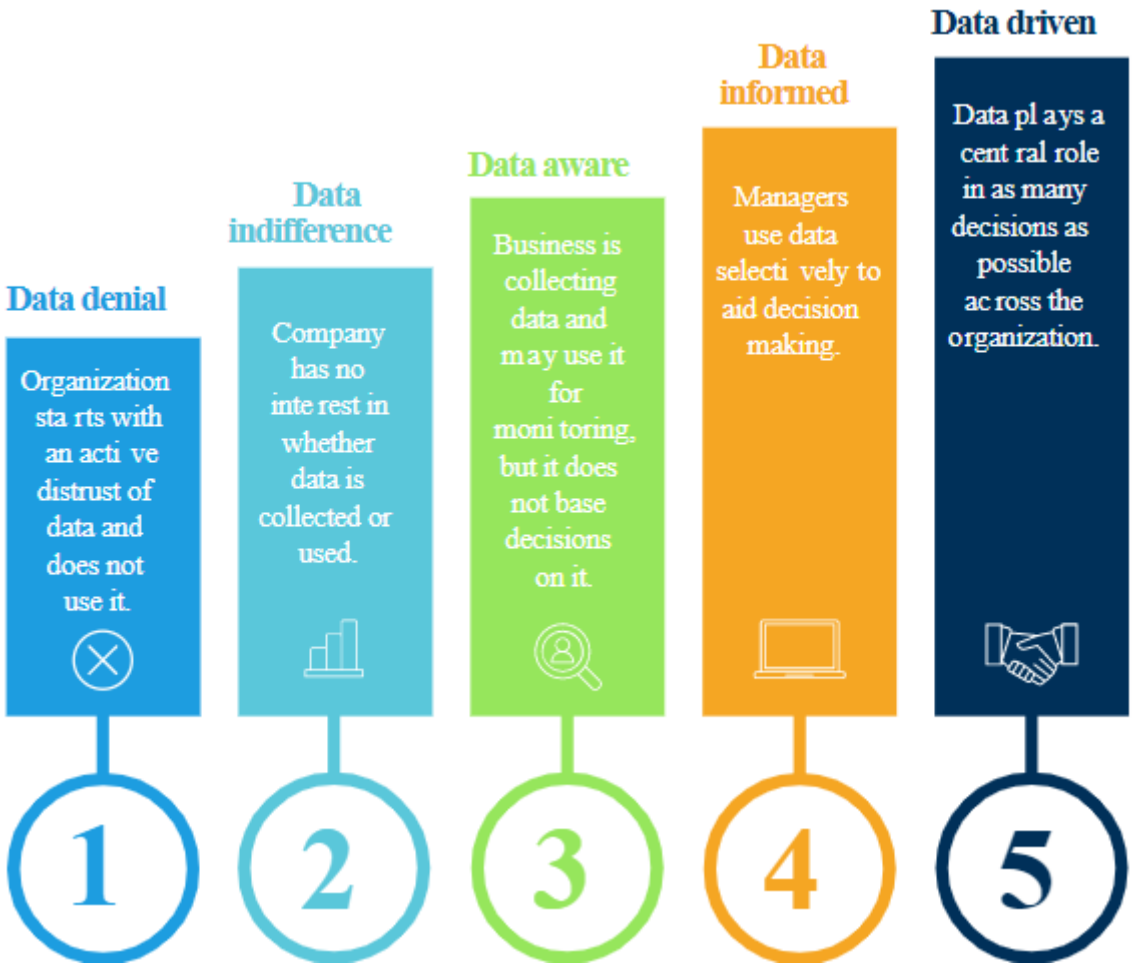
Outlook

- Breeding as a process
- What is genetic gain?
- Decisions for estimating genetic gain
- Guidelines for calculation

CGIAR

Excellence in Breeding PLATFORM

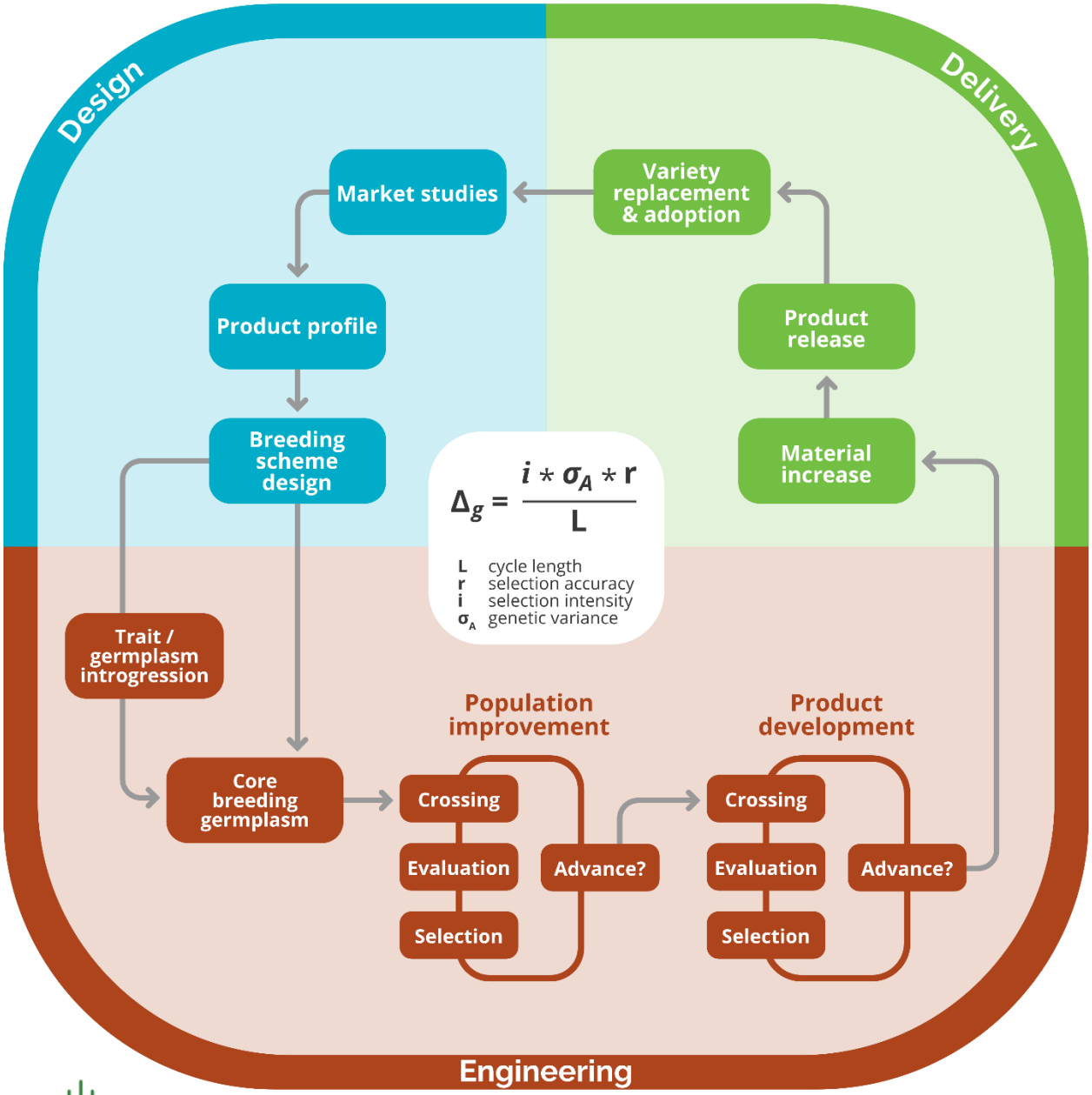
Why we want to estimate the rate of genetic gain?



Five stages towards data drive culture

CGIAR

Excellence in Breeding PLATFORM

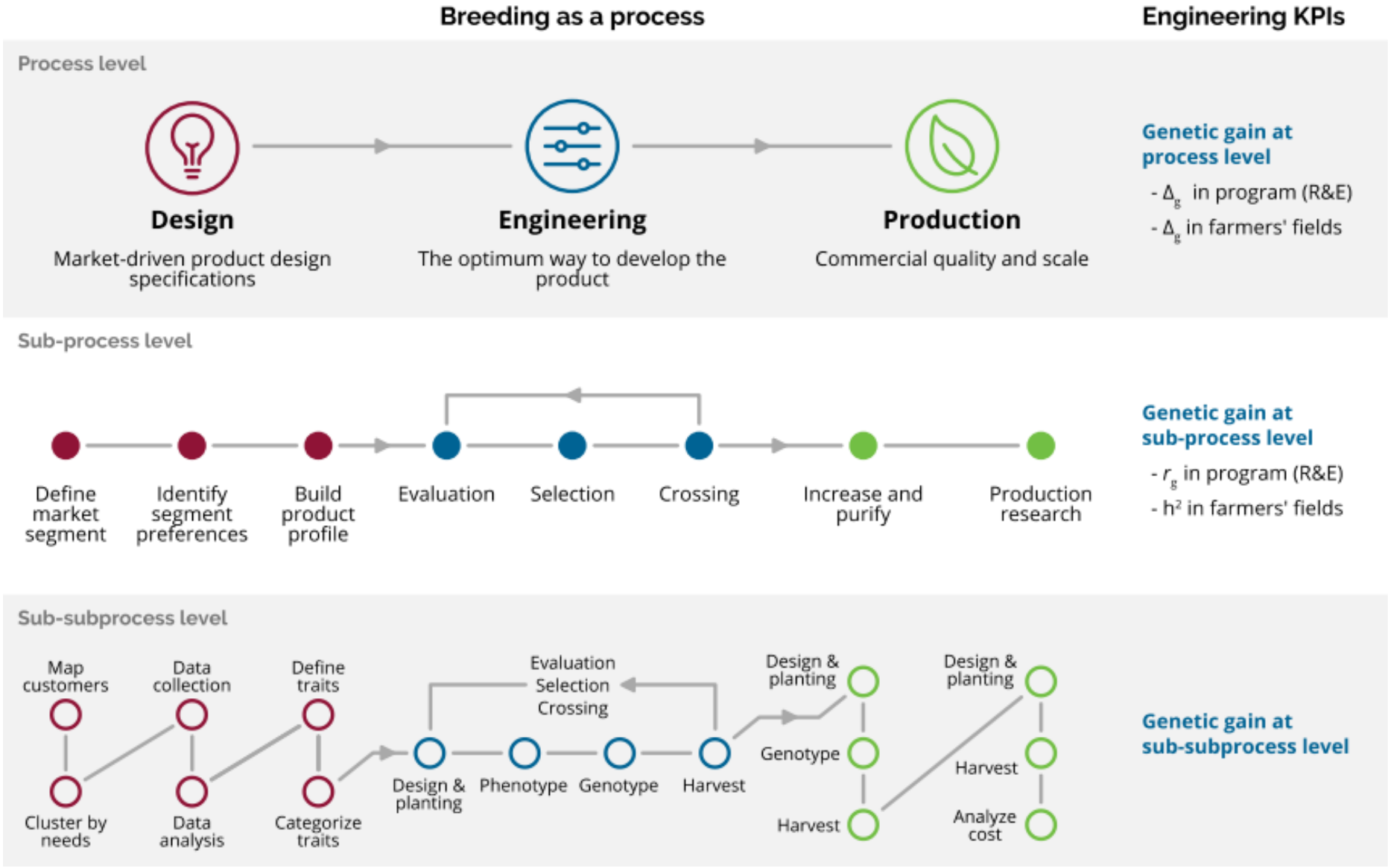


Breeding as a process

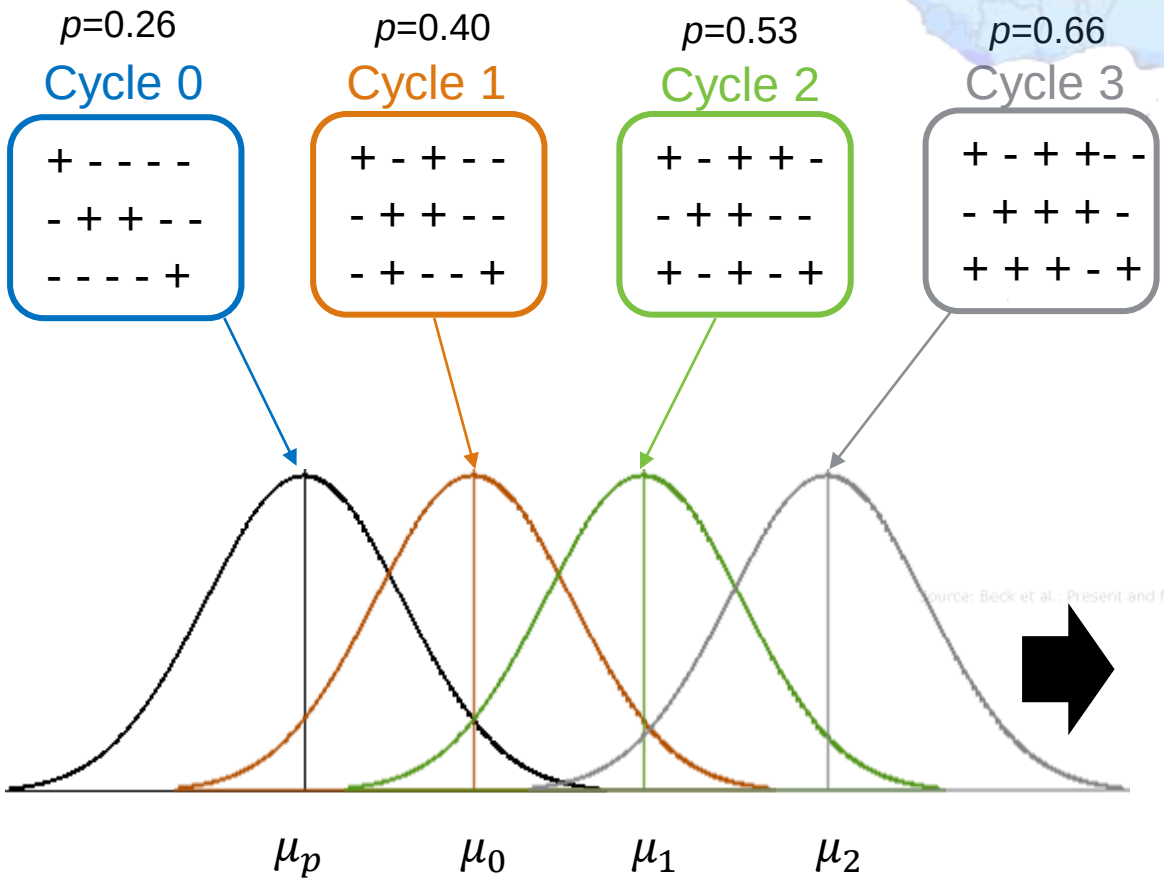
CGIAR

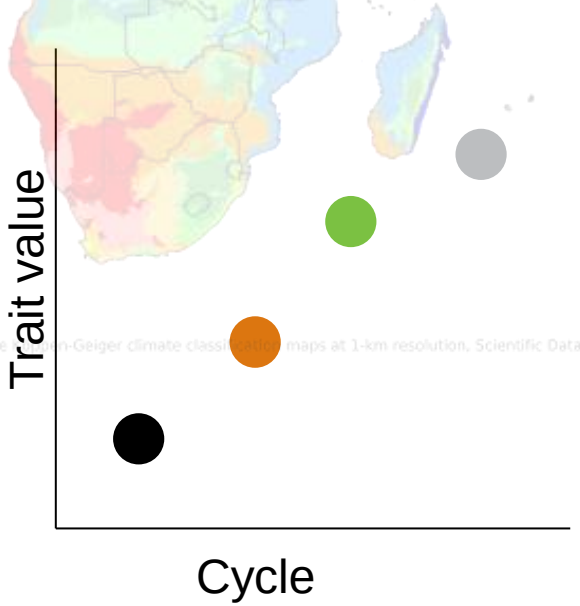
Excellence in Breeding PLATFORM

Any process requires KPIs



Where genetic gain comes from?

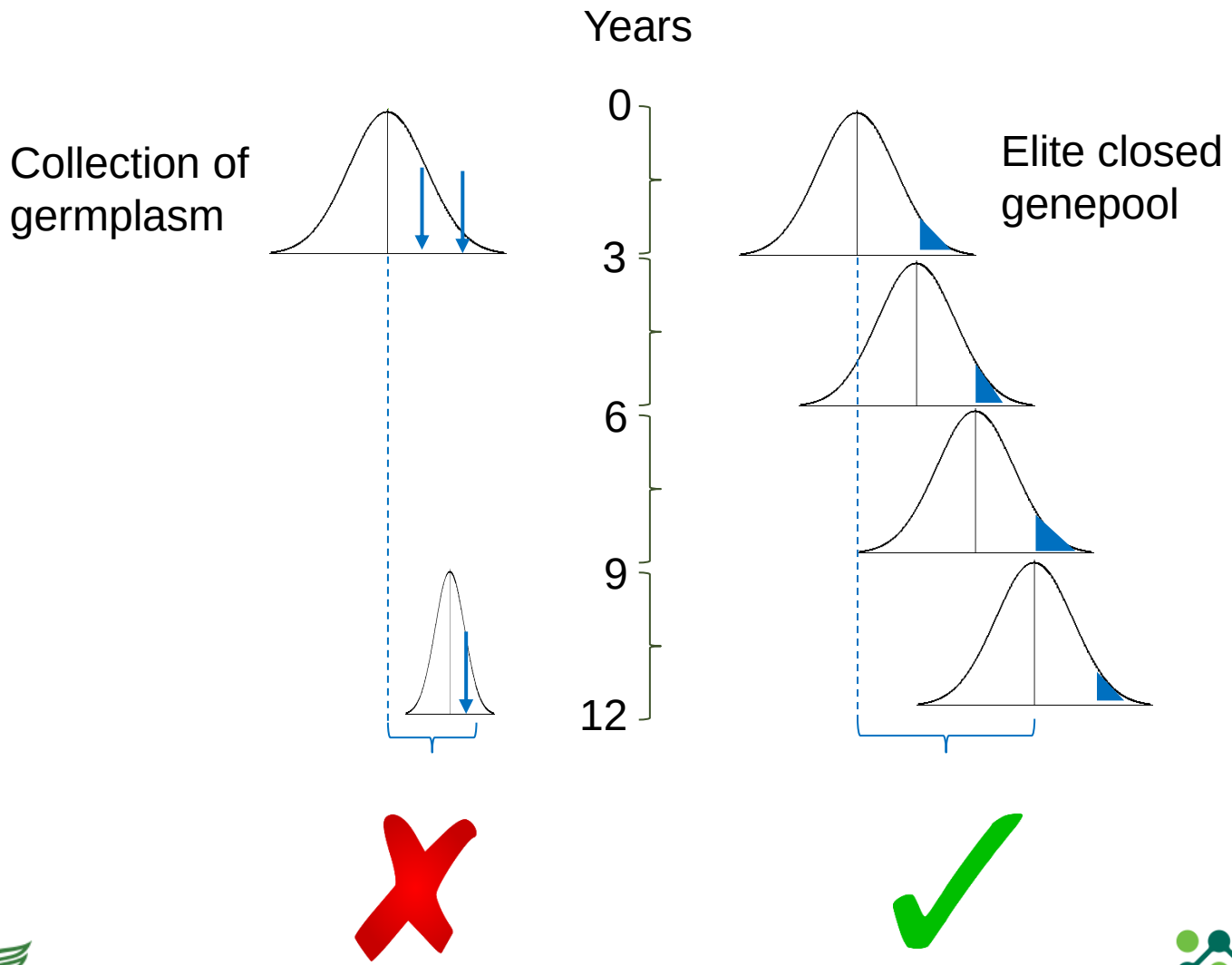




1) Understood in a population basis
2) Comes from a recurrent selection system (relatively closed system)
3) Comes from the increase in the frequency of alleles responding positively to natural and/or simulated environmental conditions.

CGIAR

What is not “rate of response to selection” or “genetic gain”?



CGIAR

Excellence in Breeding PLATFORM


~


OR


~

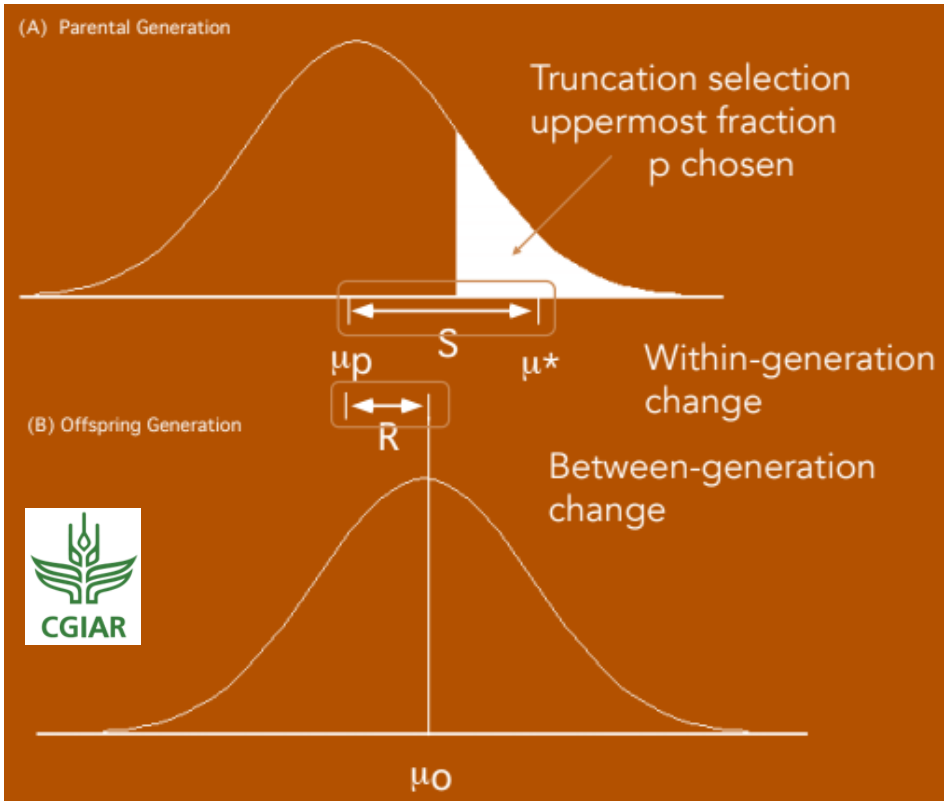

How to derive the expected genetic gain?

1a) $y = \mu + g + e$

1b) $g = b(y - \mu)$

1c) Assuming no dominance, g = a (the additive component), remember that:

$$b = \frac{cov(a,y)}{var(y)} = \frac{cov(a,a+e)}{var(y)} = \frac{cov(a,a)}{var(y)} = \frac{var(a)}{var(y)} = h^2$$



If we apply selection (*) in the parental generation based in 1b):

1d) $g_* = \mu_p + b(y_* - \mu_p)$

Take the expectation for individuals selected in 1d)

1e) $\mu_0 = \mu_p + b(\mu_* - \mu_p)$

Re-arrange

1f) $\mu_0 - \mu_p = b(\mu_* - \mu_p)$

Remember 1c) and:

$$S = \mu_* - \mu_p$$

1g)

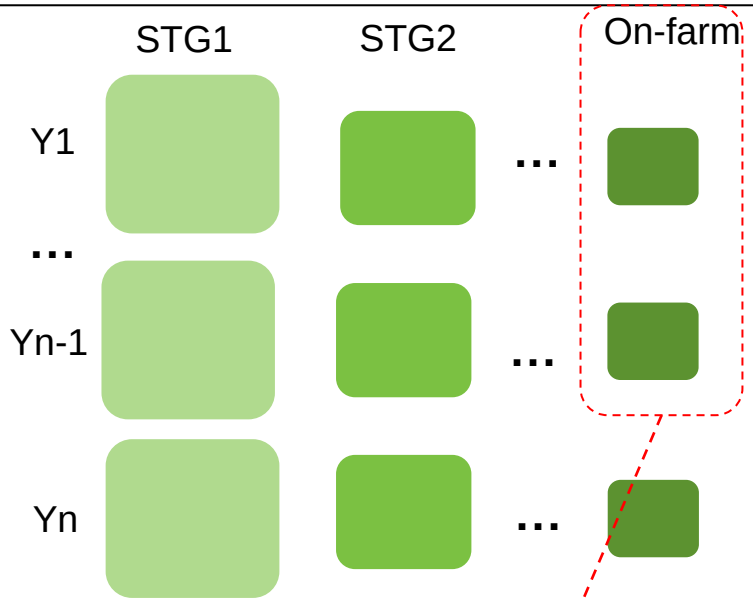
1h) $R = h^2 S$

1i) Now remember:

$$h^2 = \frac{\sigma_a^2}{\sigma_y^2} = \frac{\sigma_a\,\sigma_a}{\sigma_y \sigma_y} \quad \& \quad i = \frac{S}{\sigma_y}$$

1j) $R = \frac{\sigma_a\,\sigma_a}{\sigma_y\,\sigma_y}\,S = \frac{\sigma_a\,\sigma_a}{\sigma_y}\,i = h\,\sigma_a\,i = r\,\sigma_a\,i$

How to derive the realized genetic gain?

	Input	Method	Output
Data generation	Target segment information + Resources + Breeding scheme	Breeding trials method OR Era trial method	
Data reduction	Phenotypes	Trait(s) + Stage + Time	
Data analysis	Phenotypes subset	Linear models $y = X\beta + Zu + e$	
Realized gain calc	Adjusted means	Linear models $y = X\beta$	Δ_G

CGIAR

Excellence in Breeding PLATFORM

Methods for calculating the rate of genetic gain

Table 1. Summary of methods to estimate the response to selection and rate of genetic gain.

Method	Formula used	Data required	Sample	Factors to be considered	Connectivity** / TPE* coverage	Recommendations
Expected	$R = h^2 S = i\, r\, \sigma_g$ Lush (1942), Burrows (1972), Walsh (2004)	Any trial information	Any generation material	The heritability used will have an important effect in under- or over-estimating the metric.	Low after first selection cycle.	Use across-trials heritability. Do not use to take complex decisions
Realized	$y = X\beta + Z_d u_d + Z_g u_g + \varepsilon$ β: vector of fixed effects. u_d, u_g, ε: vector of random non-genetic, genetic and error effects. X,Z_d,Z_g: incidence matrices connecting observations with vectors of fixed and random effects Laidig *et al.* (2014), Piepho *et al.* (2014), Mackay (2011), Garrick (2010)	Era trial information	Early generation material	1) TPE* coverage is low (usually some locations & a couple of years). 2) Connectivity** among entries is maximum (cohorts evaluated at the same time). 3) Sample can overestimate the metric.	High / Low	1) Evaluate the material in representative environments for more than one year. 2) Use a replicated design. 3) Take a representative sample from each cohort if an estimate of evolution of genetic variance is required.
			Advanced material			
			Released varieties			
			On-farm			
		Historical trial information	Early generation material	1) TPE* coverage can be low (early), intermediate (advanced) or high (varieties). 2) Connectivity among entries depends on checks and the use of methods like EBV. 3) Sample can overestimate the metric.	Variable / Low	1) Use 4-10 checks depending on the stage*** to increase the connectivity of the data. 2) Use early generation trials for better estimate of evolution of genetic variance and advanced material for better estimates of the rate of genetic gain.
			Advanced material		Variable / Intermediate	
			Released varieties		Variable / High	
			On farm		Variable / High	

Decisions for measuring Δ_G

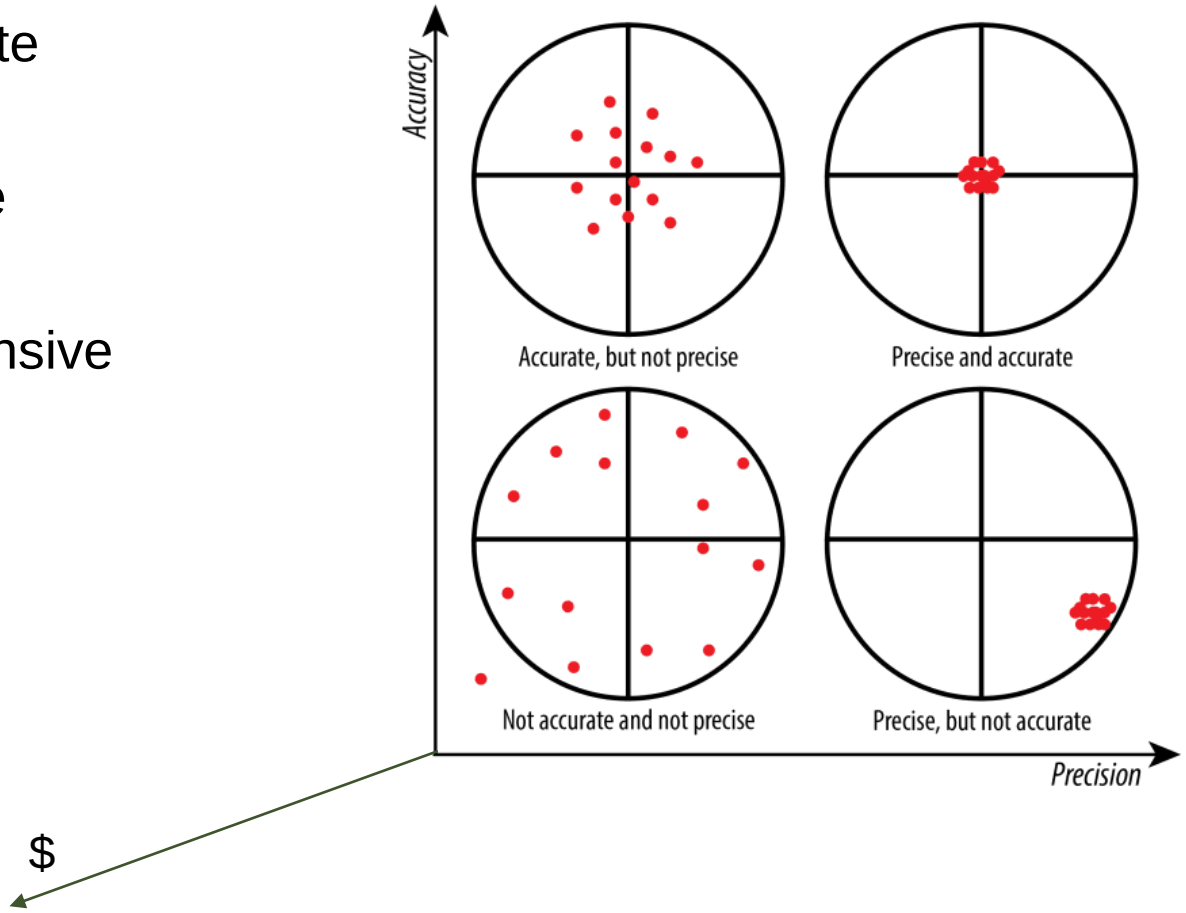
1. Associated target (Market segment: TPE + product description)
2. Trait of interest [productivity (yield), other trait or an index; BV or genetic value?]
3. Germplasm stage sample (PYT, IYT, AYT, On-Farm, … trials)
4. Time period of interest (i.e. last 10 years)
5. Method to produce the data (explicit era trial or historical information)
6. Methodology to calculate Δ_G (mixed models; 1-step, 2-step, etc.)

CGIAR

Excellence in Breeding PLATFORM

What makes a good estimate of rate of genetic gain?

- Accurate
- Precise
- Inexpensive

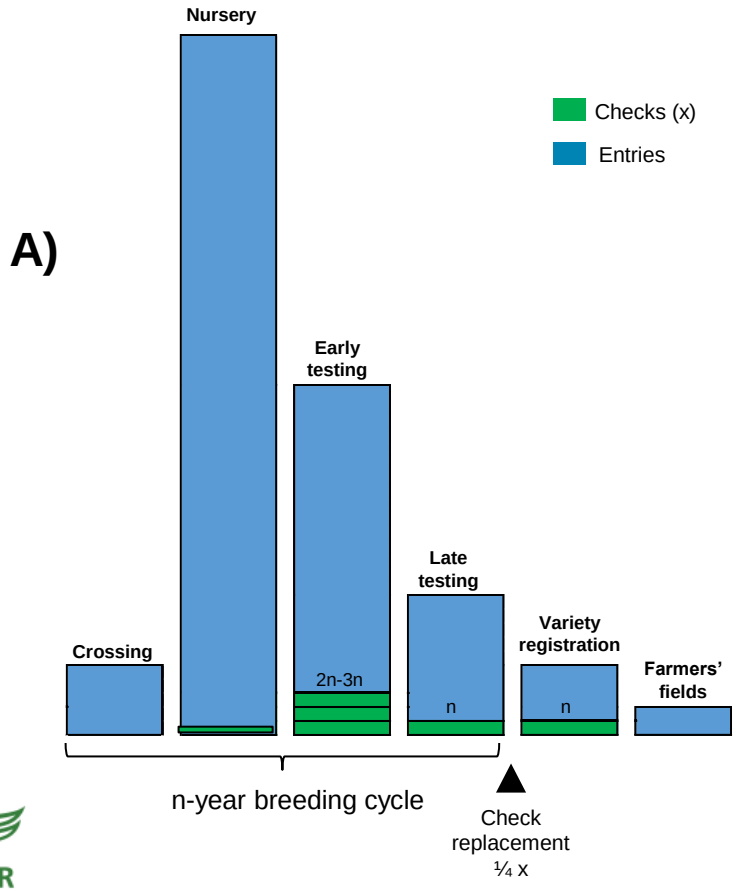


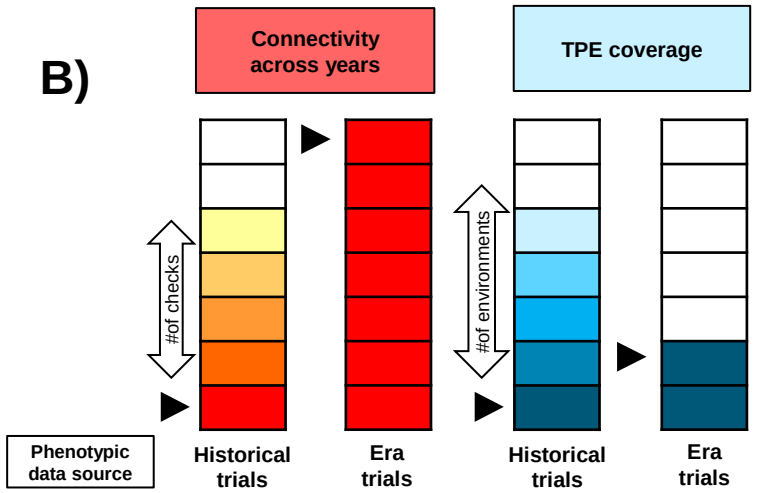
CGIAR

Excellence in Breeding PLATFORM

Guidelines to maximize accuracy and precision of Δ_G

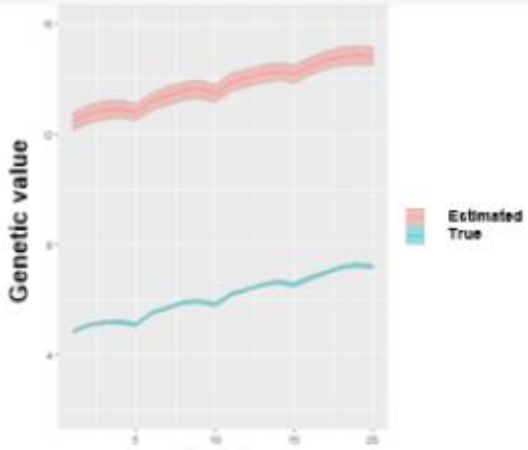
- Make sure you can **connect your data (A)** to remove year effects. Make sure you can **cover the target (i.e. TPE) properly (B)**.

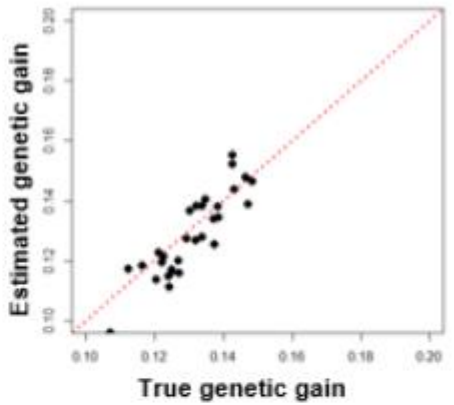




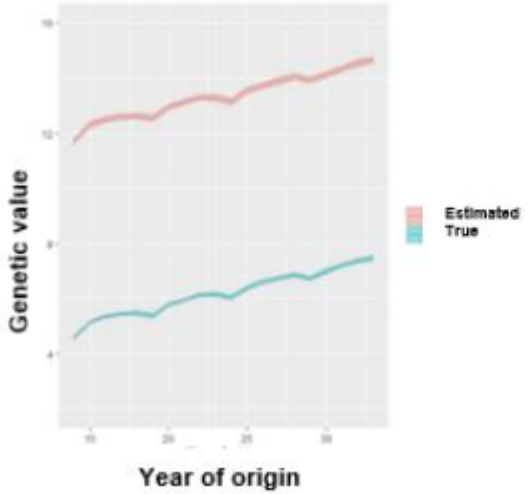
Guidelines to minimize cost of estimating the Δ_G

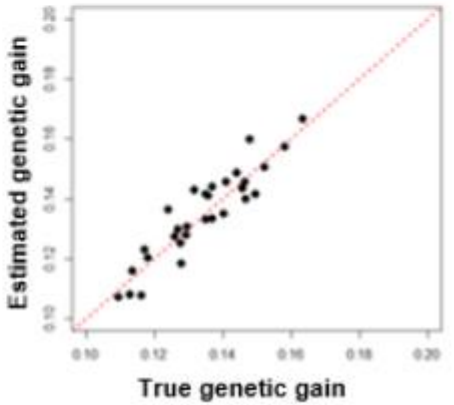
Era trial information (Advanced)





Historical trial information (Varieties)





Adoption of Δ_G KPI estimation

Progress

0 cycles	1 cycle	2 cycles	3 cycles	Adoption time
1) Era trial OR Historical-EBV 2) Expected Δ_g method	1) Era trial OR Historical-EBV 2) Expected Δ_g method	1) Historical 2) Expected Δ_g method		Δ_G used
Start	Continues	Adopted		Recommendations
Low	Low	High		Accuracy per dollar invested

CGIAR

Excellence in Breeding PLATFORM

Thank you for your interest!

CGIAR

Excellence in Breeding PLATFORM

Recipe to calculate Δ_G

1. Define target associated (i.e. Drought conditions environments).
2. Define time period of interest (i.e. last 20 years).
3. Define trait of interest (i.e. yield).
4. Define germplasm stage sample (PYT & AYT).
5. Produce field data:
    1. Run an era trial with PYT&AYT materials from last 20 years under drought environments and record yield.
    2. Collect historical information from PYT&AYT materials from last 20 years under drought environments and extract yield)
6. Fit a MET analysis with a mixed model and extract adjusted means for genotypes.
7. Merge year of origin to the adjusted means.
8. Fit a linear model of the form *adjusted.mean~year.origin*
9. The slope is the rate of genetic gain.

CGIAR

Excellence in Breeding PLATFORM

Typical structure of a breeding program