LOUISIANA STATE UNIVERSITY College of Agriculture School of Plant, Environmental, and Soil Sciences AGRO 7075 Prediction-based Breeding

G x E

Prof. Roberto Fritsche-Neto

rfneto@agcenter.lsu.edu

Baton Rouge, April 24th, 2023

Basic concepts

- **Environment - all not genetic features affecting the plant development**
- ***edaphoclimatic conditions, harvest, year, management, diseases...***

- **Agriculture has evolved into greater environmental control**
- ***fertilizers, pest and disease control, irrigation, GMOs...***

- **One cultivar can be extremely productive in one environment but not in another**
- **Genotype x Environment: differential performance of genotypes as a function of environment variations**
- **P = G + E + GE**

Main causes of GxE

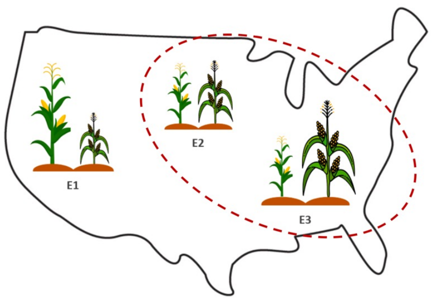
- Predictable factors
- Agricultural practices
- Soil physical composition
- Sowing method
- Photoperiod
- Soil fertility
- Aluminum toxicity

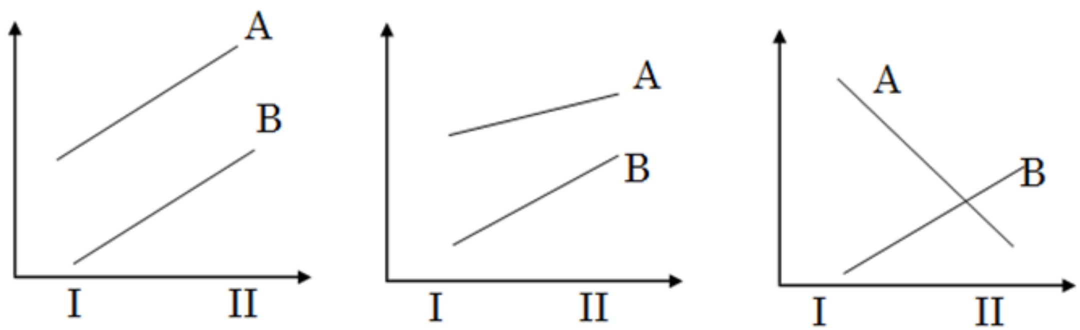
- Unpredictable factors
- Relative humidity of the air
- Rainfall distribution
- Atmospheric and soil temperature
- Pathogens
- Insects
- Diseases

Genetic composition
E.g., Hybrids vs Inbreeds

Improvement for biotic and abiotic stress conditions

Genotype x Environment types





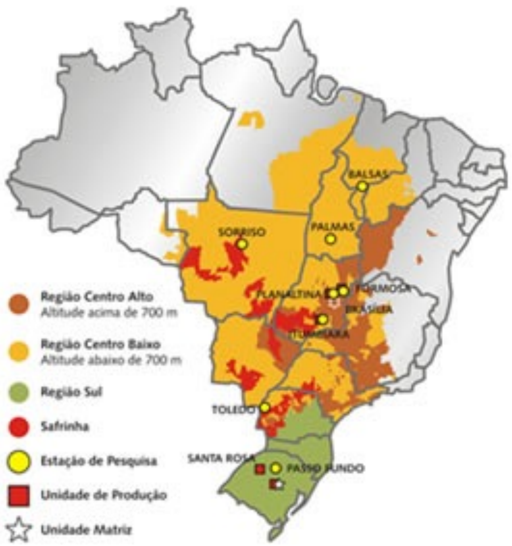
Absence Non-crossover Crossover

- **Which is the most common one?**
- **It depends on the crops and trait**
- **What is the desirable type?**
- **It depends on the target trait (*productivity vs. diseases vs. flowering*)**
- **the objective (specific or mega environments),**
- **And the seed market**

GxE implications

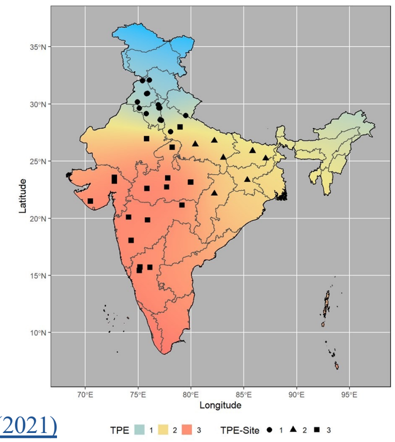
- A material recommended for one location, is not for another
- An improvement program at each location - high cost
- It makes it difficult to obtain cultivars of wide adaptability
- Preserves genetic variability - prevents only one genotype in large areas - genetic vulnerability

The more advanced the breeding program and the market, the greater the tendency to capitalize on GxE



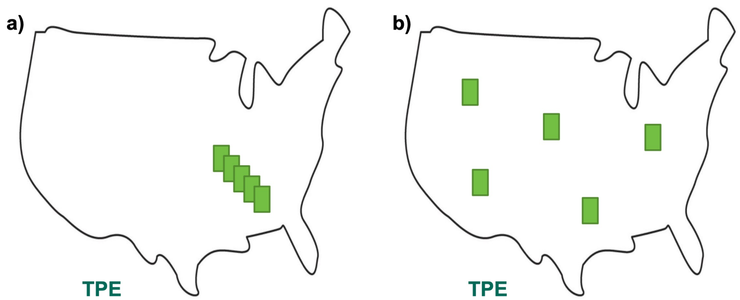
Target Population of Environments (TPE)

- **The set of farms and future seasons in which the varieties released by a breeding program will be grown**
- **Environment (predictable and non-predictable ones)**
- **Socio-economic factors**
- **Management / Production system (DRS or transplanted; available technology; hybrids, lines, clones; use of fertilizer and plant protection)**
- **Market preferences (type of grain, color, composition, …)**
- **Frontiers (transit and trade are not allowed)**
- **What is the difference between mega-environment and TPE?**
- ***Clusters that minimize GxE intra ME, then combine the other layers***



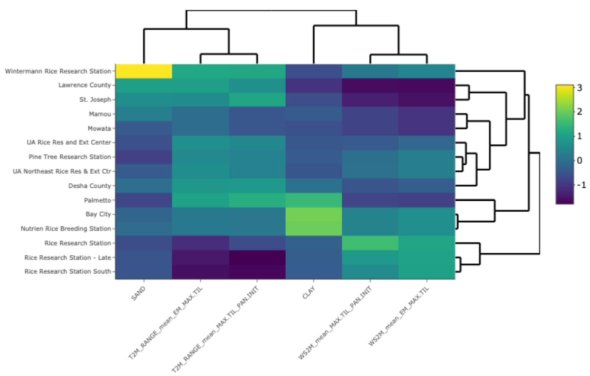
Crespo-Herrera et al (2021)

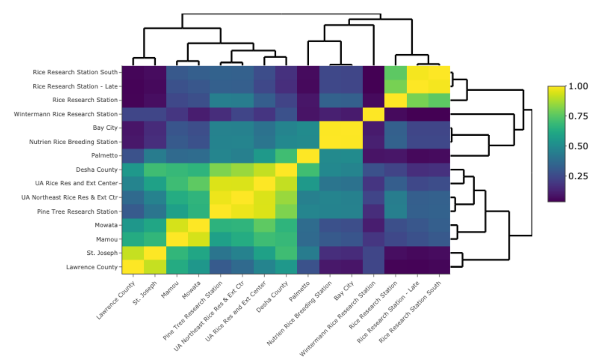
Number of locations *vs.* number of environments

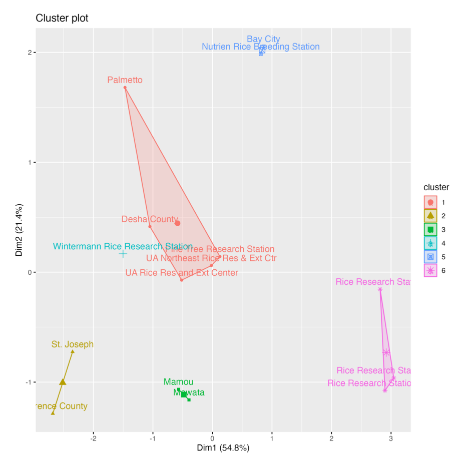


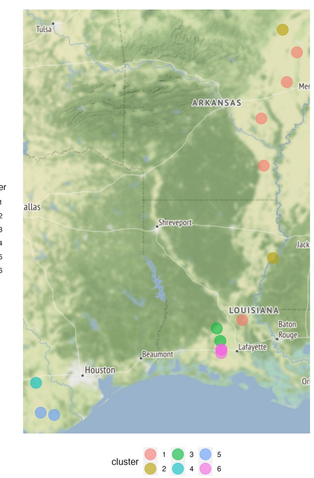
Overall trend:

- **Minimize the number of replicates and maximize the number of "unique" locations**
- **Sparse designs (later…)**



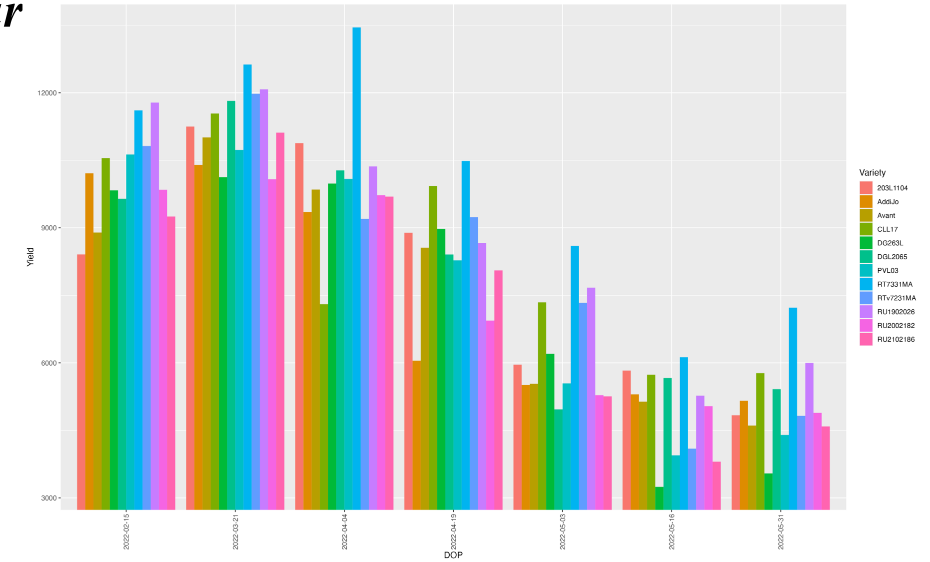






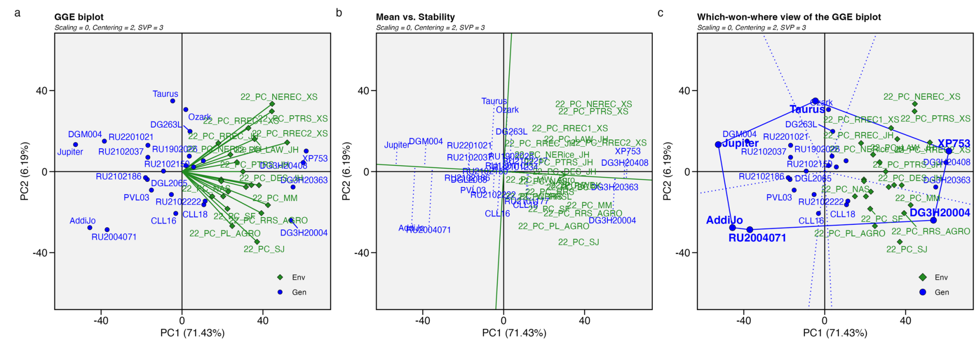
Choosing locations for MET

- **The ideal environment should:**
- ***a) promote the expression of the genotype's performances***
- ***b) maximize the genetic variance***
- ***c) minimize the variance of environment and GxE interaction***
- ***d) have consistent characteristics from year to year***
- ***e) represent the actual farming environment***
- ***f) Be accessible***
- ***g) represent the actual market***



Environmental stratification - mega-environments

- **Tests:** the costliest phase
- **Mega-environment:** minimum GxE variance within group and maximum between
- **Optimize:** trials, replicates, and allocation
- **KPI**: broad-sense heritability (repeatability)



Observed phenotype

Confounding factors

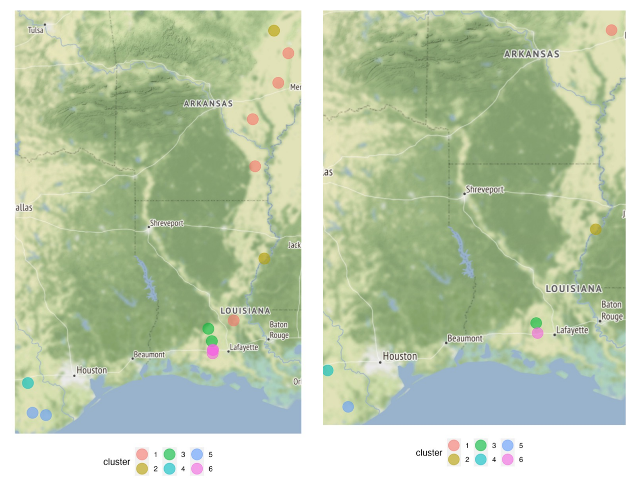
Random error
(what we want to reduce)

$$y_{ijk} = \mu + g_i + t_j + gt_{jk} + e_{ijk}$$

Surrogate of genetic value (estimate we select on)

Systematic confounding factors
- Correct for the environmental effect
- Correct for (and use) GxE interaction effect

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2} = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gxe}^2}{\boldsymbol{nEnv}} + \frac{\sigma_e^2}{\boldsymbol{nEnv * nRep}}}$$



Estimating GxE

- **At least two genotypes and two environments**

Yield	Env 1	Env 2	Total (Ti.)
Genotype A	10	09	19
Genotype B	08	12	20
Total (Y.j)	18	21	39 (Y..)

SV	DF	E(MS)	F
Blocks:Env	(b-1)a	Ve + gVb	
Enviroments (A)	a-1	Ve + gVb + gbVa	MSA / MSB
Genotypes (G)	g-1	Ve + ba*Vga + baVg	MSG / MSGA
G x A	(a-1)(g-1)	Ve + ba*Vga	MSGA / MSR
Residual	(a-1)(b-1)a	Ve	

$$C = \frac{Y_{..}^2}{N} \qquad SQA = \frac{\sum Y_{.j}^2}{I} - C \qquad SQG = \frac{\sum Y_{i.}^2}{J} - C$$

$$SQTotal = \sum Y_{ij}^2 - C \qquad SQ_{GxA} = SQtotal - SQA - SQG$$

- **The principle is the same for DOP, year, location, N, etc.**
- **What happens when we have a 100% crossover GxE?**
- **Genotypes, environments and GxE interaction - fixed or random?**

Decompose GxE into simple and complex

- **MS_{GxE} = simple + complex**

$$MS_{GxE} = 1.72$$

- Simple

$$S = \frac{1}{2}\left(\sqrt{Q_1} - \sqrt{Q_2}\right)^2$$

$$S = 0.01 = 0.06\%$$

- Complex

$$C_1 = (1 - r)\sqrt{Q_1 . Q_2}$$

$$C = 1.71 = 99.4\%$$

- ***Example:***

SV	DF	MS	
		ENV 1	ENV 2
Genotypes	255	5.73	4.96
Residual	224	1.47	1.53

$(r = 0.68)$

SV	DF	MS
Environments	1	113.30
Genotypes	255	5.09
G x E	255	6.52
Residual	450	1.50

Effects on Response to Selection

- **You can increase the decrease in gains with the selection**

$$RS = i.h.\sigma_a \qquad \sigma^2_{G(i1)} = \sigma^2_{Gi} + \sigma^2_{GE(i1)}$$

- **Selection and response in the same environment - capitalizes on GxE**

$$\sigma^2_{G(i1)} = \sigma^2_{Gi} + \sigma^2_{GE(i1)} \qquad \sigma^2_{G(i2)} = ? \qquad COV(Yij; Yij') = \sigma^2_G - \sigma^2_{GE} \qquad r_g = \frac{\hat{\sigma}^2_g - \hat{\sigma}^2_{ge}}{\sqrt{\hat{\sigma}^2_{g1}.\hat{\sigma}^2_{g2}}}$$

- **Selection in one and response in another contrasting environment**

Average environments – minimize GxE

$$H^2 = \frac{\sigma^2_g}{\sigma^2_p} = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_{gxe}}{\boldsymbol{nEnv}} + \frac{\sigma^2_e}{\boldsymbol{nEnv * nRep}}}$$

Specific environment – capitalize GxE

$$V_{(y)} = Vg + Ve/r$$

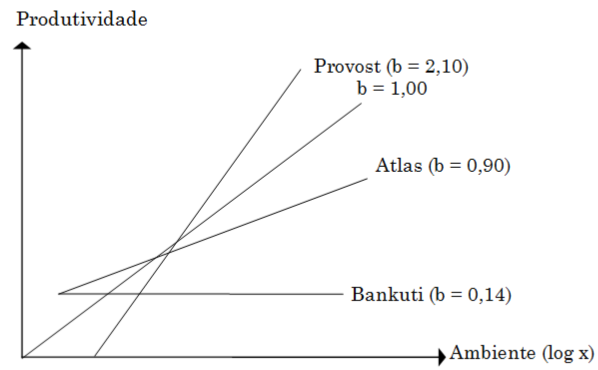
$$Vg = Vg + Vg * year + Vg * loc + Vg * season \; + Vg * M\ldots$$

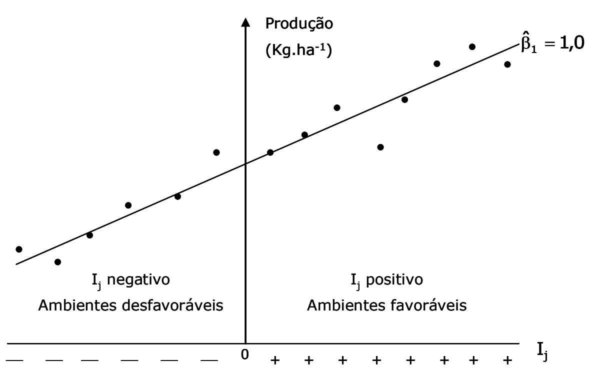
Adaptability and stability – Regression methods

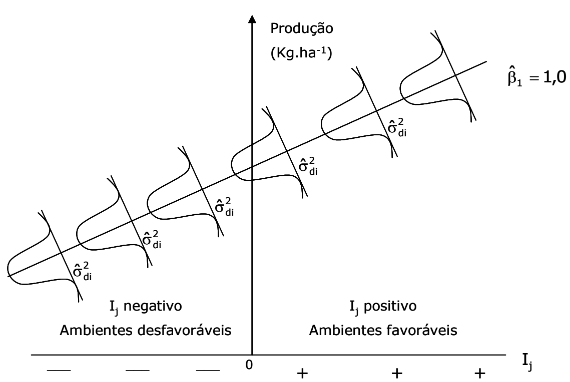
- **Adaptability:** ability to take advantage of environmental variations
- **Stability:** predictable behavior in the face of variations in the environment
- **Finlay K, Wilkinson G (1963):** Linear regression coefficient and the variance of the regression deviations

$$I_j = \overline{I}._j - \overline{Y}..$$

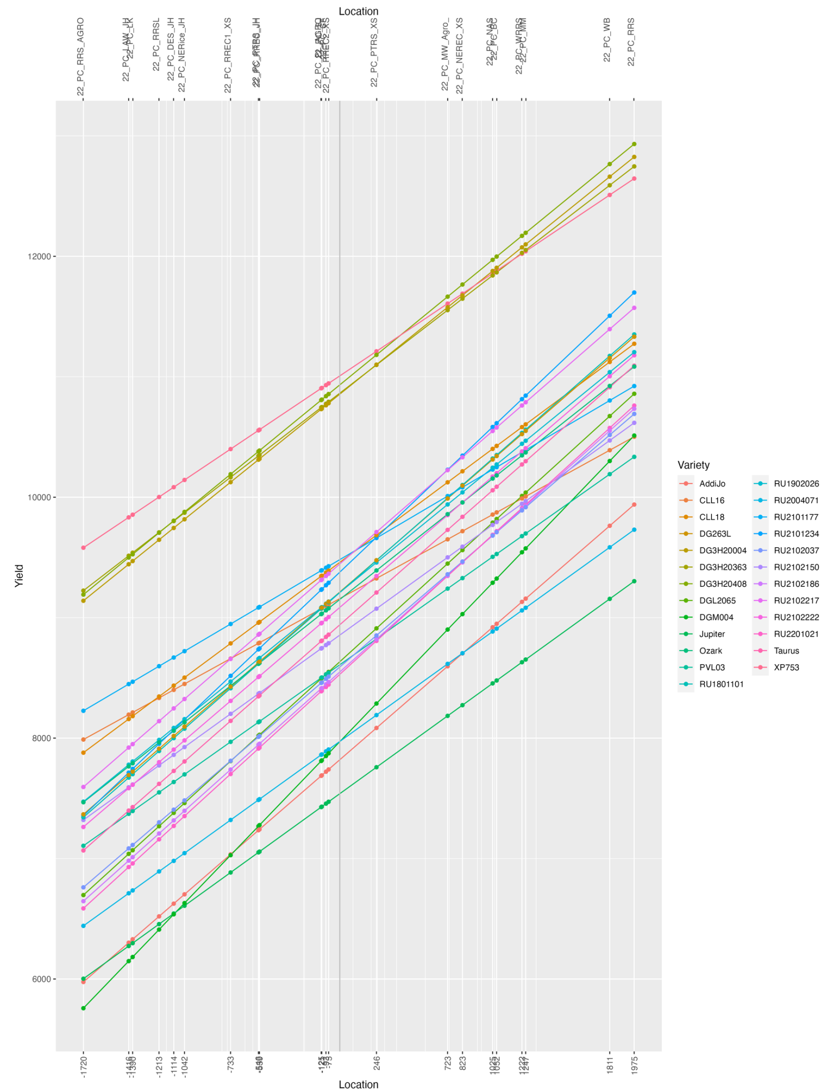
$$Y_{ij} = m_i + b_i I_j + d_{ij}$$







- ***d_{ij}: regression deviations predictability (stability)***
- ***What would be the ideal cultivar?***
- ***Y_{ij}: high overall performance***
- ***b_i =1.0***
- ***d_{ij}= 0***



Location
Yield
Location
Variety
AddiJo
CLL16
CLL18
DG263L
DG3H20004
DG3H20363
DG3H20408
DGL2065
DGM004
Jupiter
Ozark
PVL03
RU1801101
RU1902026
RU2004071
RU2101177
RU2101234
RU2102037
RU2102150
RU2102186
RU2102217
RU2102222
RU2201021
Taurus
XP753

HMRPGV

- **Harmonic Mean of Relative Performance of Genotypic Values - *Linn & Binns (1988)* via Mixed Models (Resende, 2007)**
- **First, estimate the genotypic values for each individual in each environment**
- **Then, estimate the harmonic average (*HMRPGV*):**

$$H = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots+\frac{1}{x_n}}$$

$$HMRPGV_i = \frac{E}{\sum_{j=1}^{E} \frac{1}{Gv_{ij}/\mu_j}}$$

- **Select individuals with a good baseline performance and minimum variation**
- **Also good for many levels of stress conditions**

Multiplicative models: GGE-Biplot

- **GGE stands for genotype main effect (G) + the interaction (GE), which is the only source of variation relevant to cultivar evaluation**
- **Mathematically, GGE is the genotype by environment data matrix after the environment means are subtracted**
- **The biplot is constructed by the first two symmetrically scaled principal components (PC1 and PC2) derived from singular value decomposition (SVD) of environment-centered MET data**
- **SVD creates three matrices: the singular value matrix, the entry eigenvector matrix, and the tester eigenvector matrix**
- **A biplot graphically displays the interrelationship among the entries (e.g., genotypes), testers (e.g., environments), and the interaction between them**

Multiplicative models - GGE-Biplot

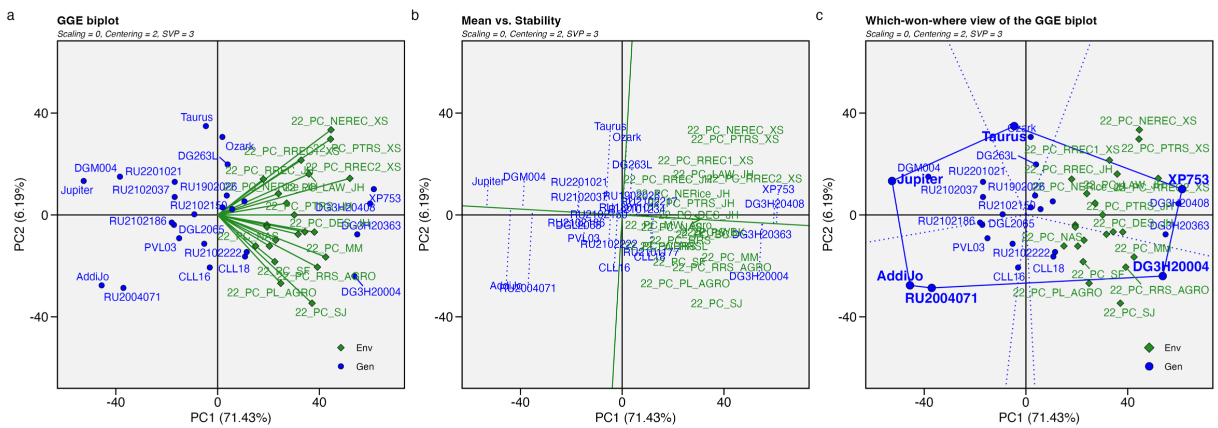
- The mean yield of genotype *i* in environment *j* is commonly described by a general linear model:

$$\hat{y}_{ij} + \mu + \alpha_i + \beta_j + \phi_{ij}$$

- **Environment means are subtracted**

$$\phi_{ij} = \hat{y}_{ij} - \mu - \beta_j = \sum_{k=1}^{p} \xi_{ik}^{*} \eta_{jk}^{*}$$

- **Apply SVD**



where $\xi_{ik}^{*} = \lambda_k^{\alpha} \xi_{ik}$; $\eta_{jk}^{*} = \lambda_k^{1-\alpha} \eta_{jk}$ being λ_k the kth eigenvalue from the SVD ($k = 1, \ldots p$), with $p \leq min(e, g)$; α is the the singular value partition factor for the Principal Component (PC) k; ξ_{ik}^{*} and η_{jk}^{*} are the PC scores for genotype *i* and environment *j*, respectively.

Always use phenotypes in analyses!!!

Sparse-Testing

- **Experimental designs:** **understand and control the effect of the environment (year x location combination) on the phenotype**
- **Replications within environments, row and column: control micro-environmental spatial variation within locations. Estimate the blocks, spatial trends, and residual effects.**
- **Trials are replications across locations and years: control macro-environmental spatial and seasonal variation. Estimate the effects of years, locations, and GxE interaction effects.**
- **<u>When is it possible? From which breeding stage?</u>**
- **Sparse designs: alleles are replicated across locations, not individuals**
- **Sparse + GS + W: a way to select for GxE in the early stages**