



EuGeM

MINI-CURSO

Modelagem da interação $G \times E$ usando dados fenotípicos, genômicos e ambientais



Germano Costa Neto

Universidade de São Paulo, ESALQ/USP

germano.cneto@gmail.com

<https://github.com/gcostaneto>

Agenda (4h de Curso)

TEORIA (1h)

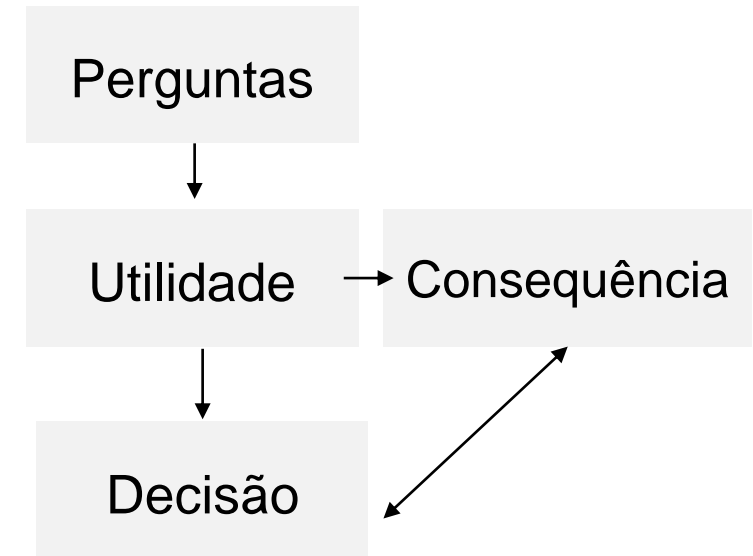
- Questões-Alvo (Teóricas)
- Causas da Variação Fenotípica
- Decomposição da Variação Fenotípica
- Ambiente e sua Tipologia
- Norma de Reação e Plasticidade Fenotípica
- Métodos Empíricos vs Analíticos (biólogicos)
- Adaptabilidade e Estabilidade
- Predição de Fenótipos sob múltiplas condições ambientais

PRÁTICA USANDO O R (3h)

- Questões-Alvo (Práticas)
- Dados Multi-Ambientais
- Norma de Reação via Regressão Linear
- GGE biplot usando Site-Regression Model
- Modelos Marca x Ambiente
- Métodos de Kernel para predição de GxE
- Envirotyping (Tipagem de ambientes) usando o pacote R EnvRtype
- Predições usando dados genômicos e ambientais

Questões-alvo

- O que causa a variação fenotípica de um dado trait?
- O que é Genótipo?
- O que é Ambiente ?
- O que é Interação Genótipo \times Ambiente ($G \times E$)?
- Como descreve-la biologicamente? (conceitos biológicos)
- Como descreve-la numericamente? (conceitos genético-quantitativos)
- É possível interpreta-la?
- Quais são seus impactos no melhoramento de plantas?
- Em quais fases dos programas de melhoramento?
- Como lidar com a $G \times E$? Explorar? Ignorar? Controlar?
- Interrelação entre:
 - norma de reação, plasticidade, adaptabilidade, estabilidade e adaptação
- Como podemos caracterizar $G \times E$?
- Como podemos modelar $G \times E$?
- Quais são os impactos em não-modelar $G \times E$?
- Quais são os benefícios e dificuldades em modelar $G \times E$?



Conceitos-Chave

(para abrir a caixa-preta da interação GxE)

- **Plasticidade Fenotípica**

Variação não-linear da expressão fenotípica em decorrência de flutuações ambientais através do tempo e do espaço

- **Norma de Reação**

Reação linear entre a resposta fenotípica e a variação de um determinado fator ambiental

- **Envirotyping (Tipagem de ambientes, “Ambitipagem”)**

Caracterização da tipologia de um ambiente (ou grupos de ambientes), envolvendo seus padrões climáticos, edáficos e bióticos

- **Enviromics (Tipagem de ambientes em larga-escala, “Ambiômica/Envirômica”)**

Uso de envirotyping em larga-escala (anos, várias épocas de plantio) visando coletar descritores ambientais confiáveis, capazes de relacionar como os impactos ambientais moldam a expressão de um dado fenótipo para uma dada espécie

- **Mega-ambiente**

Conjunto de experimentos cujo padrão ambiental se assemelha e os efeitos da interação GxE entre estes é mínimo ou nulo.

Outros pontos importantes

- Álgebra de Matrizes
- Modelos Lineares
- Ecofisiologia Vegetal (relação planta-ambiente)

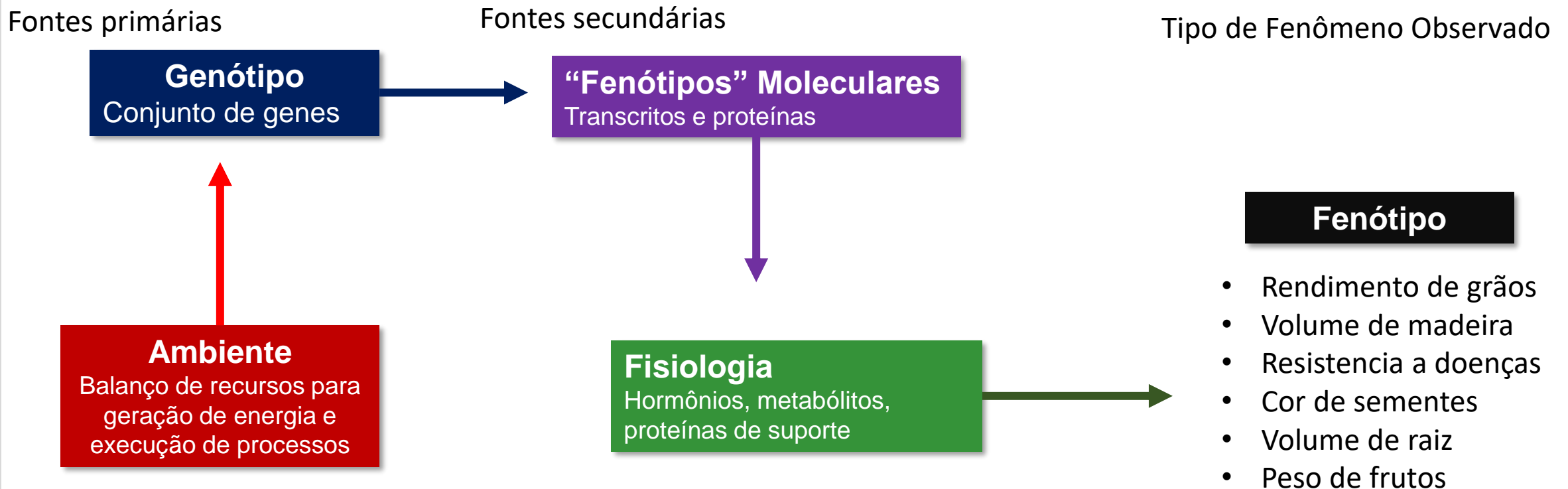


MINI-CURSO

PARTE TEÓRICA

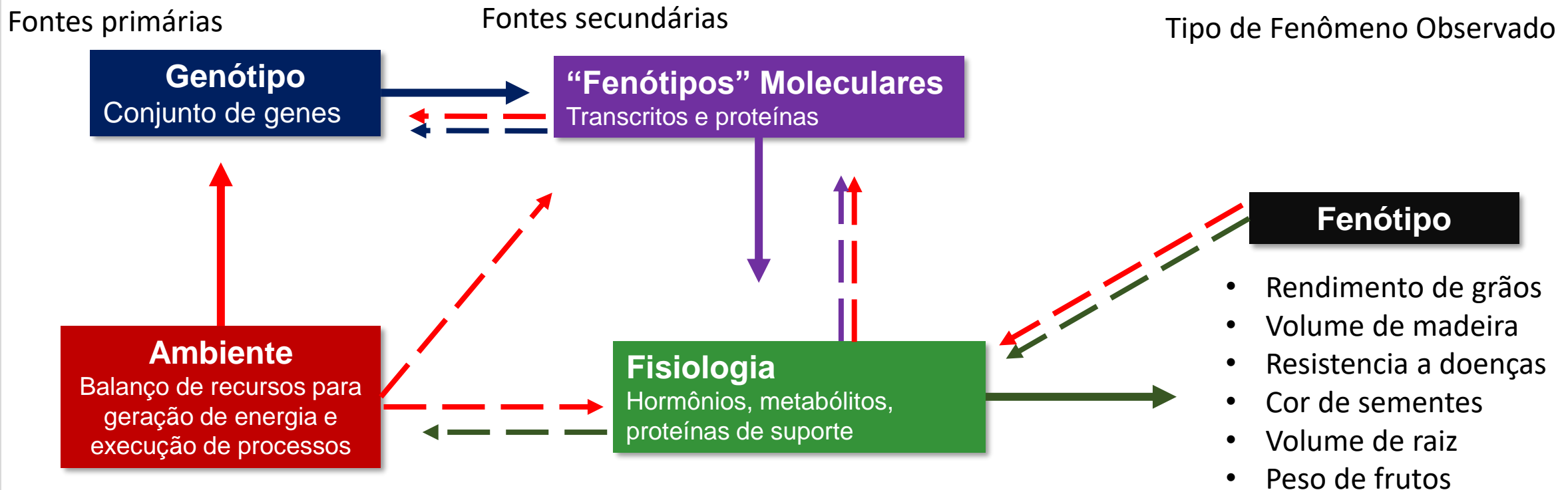
Interrelação entre Genética Molecular e Quantitativa atuando nas causas da variação fenotípica

- **Fenótipo** = Múltipla interação entre genes expressos, epigenoma e ambiente (fatores bióticos, abióticos)
- Retas contínuas : **dogma central da biologia molecular**



Interrelação entre Genética Molecular e Quantitativa atuando nas causas da variação fenotípica

- **Fenótipo** = Múltipla interação entre genes expressos, epigenoma e ambiente (fatores bióticos, abióticos)
- Retas contínuas : **dogma central da biologia molecular**
- Atuação ambiental: nível nuclear, nível celular & nível fisiológico
- Tudo que não seja a estrutura molecular do gene é “ambiente” (linhas tracejadas, seu meio no qual está inserido)
- Não é “via única” ! Existem efeitos compensatorios, homeostase, plasticidade, epigenética....



Decomposição da Variação Fenotípica

Experimentos Agrônomicos/Florestais
Lavouras e florestas!

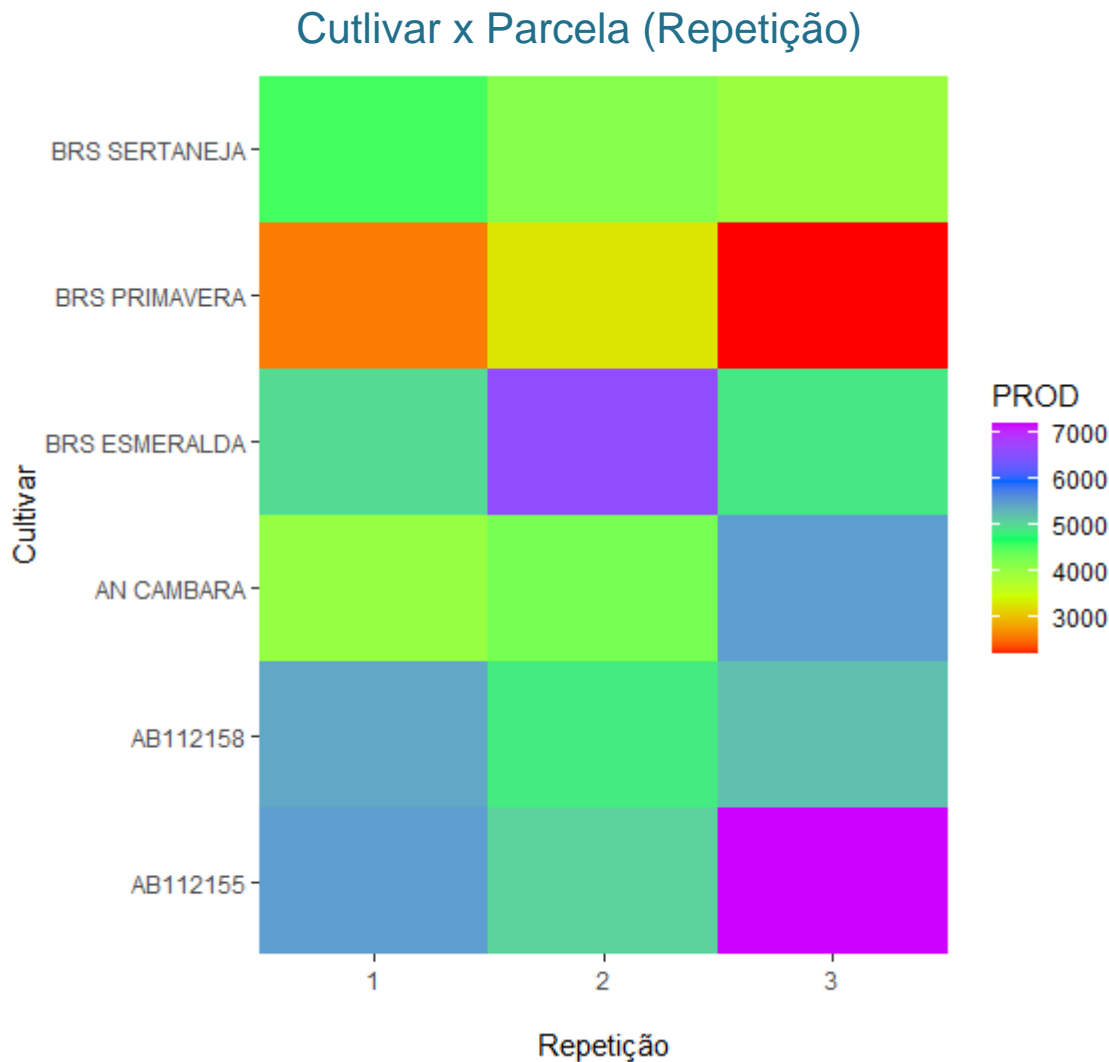
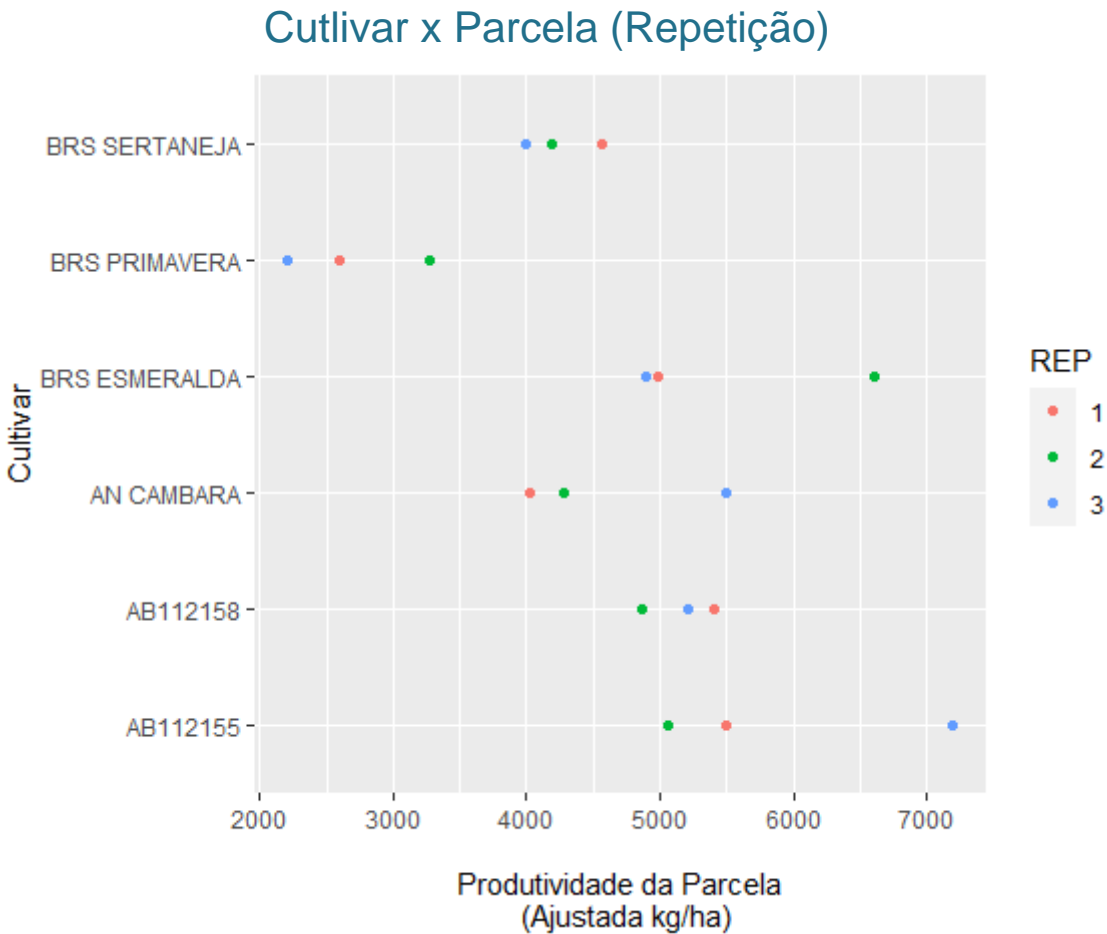
- Talvez a maior contribuição de Sir. Ronald A. Fisher
- Consiste em quantificar a importância das possíveis **fontes de variação (controláveis)** sob a **variação aleatória observada nos fenótipos**
- Específico de cada *trait* (característica/caracter alvo)
- Depende do seu delineamento experimental (e genético)
- **Segue um modelo biométrico** (genético-quantitativo)
- Sua resolução **depende do propósito do estudo**
- **Variação** (fenômeno biológico) **ou Variância** (parâmetro estatístico)?
- **Decomposição da VARIÂNCIA Fenotípica** (componentes de variância)



Fisher (1890 – 1962)

Decomposição da Variação Fenotípica

Dentro de um mesmo “ambiente”

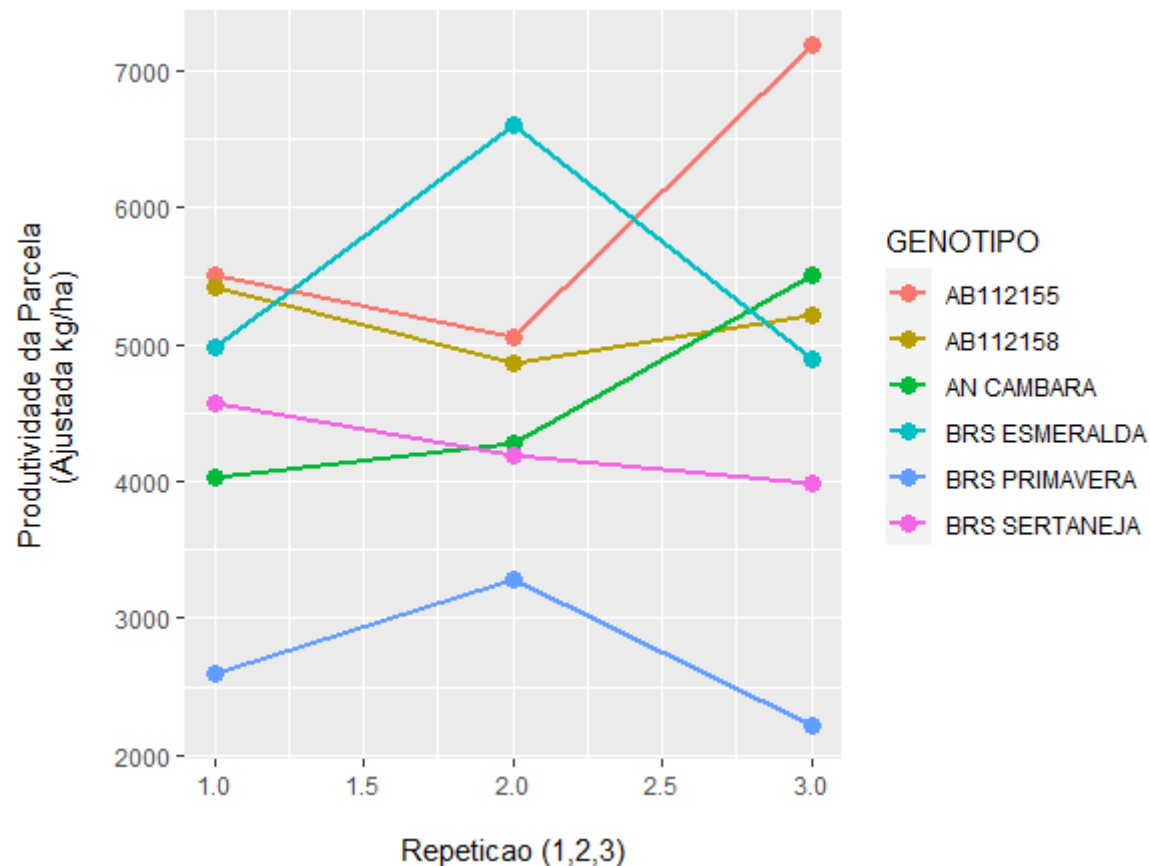


Decomposição da Variação Fenotípica

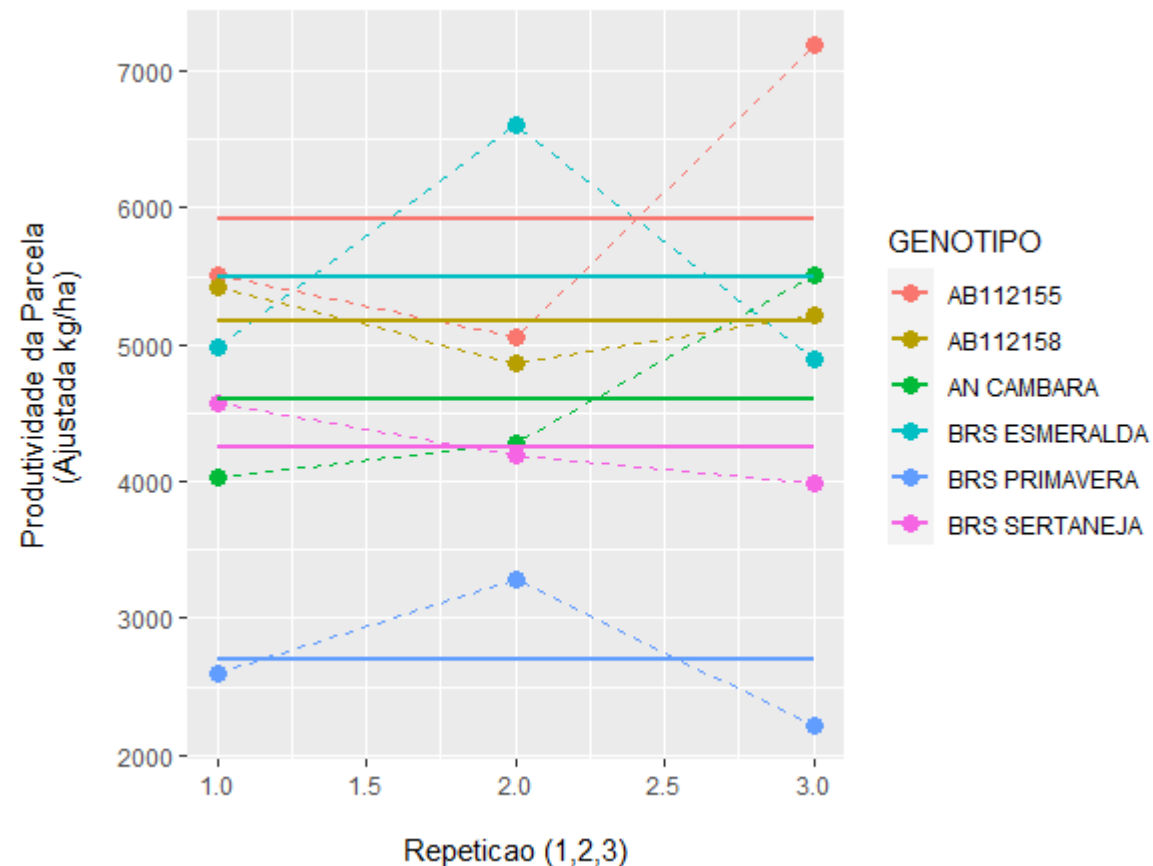
Separar “Padrão” de “Ruído”

Dentro de um mesmo “ambiente”

Dados brutos (raw data):
interação (mas há covariância?) Cultivar x Parcela



Dados ajustados (adjusted means):
Média do Genótipo + Desvios para cada Parcela



Próximo passo?

Testar hipóteses, p.ex: (H_0 : não há diferença entre genótipos)

Decomposição da Variação Fenotípica

Ensaio individuais (único experimento de campo)

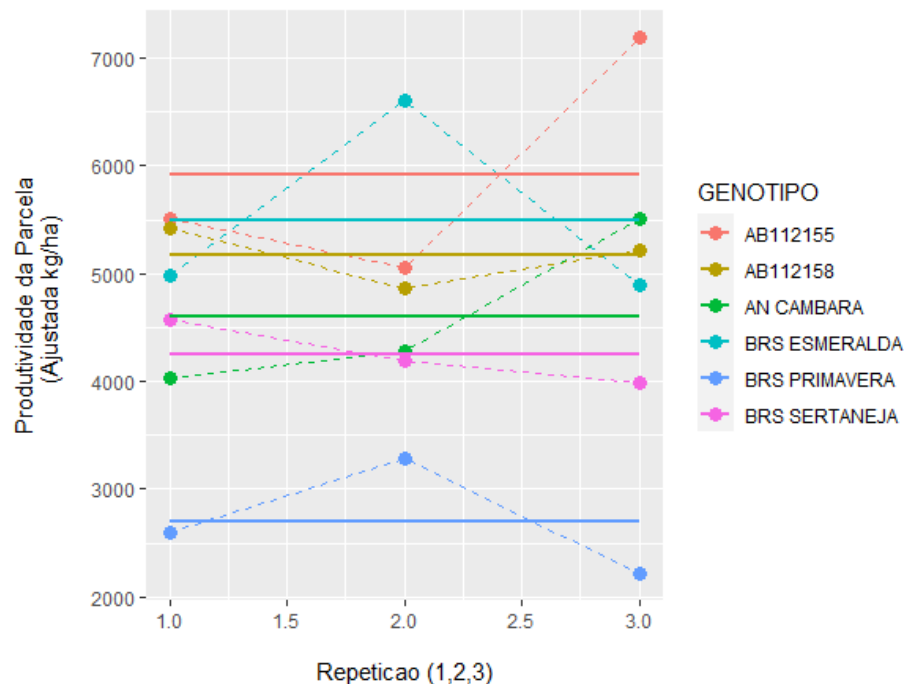
$\text{Var}(Y) = \text{Var}(G) + \text{Cov}(G,E) + \text{Var}(R)$ sendo G = fatores “genéticos” e R = resíduos experimentais

Boa qualidade experimental (condução dos ensaios e ajuste do fenótipo) resulta em $\text{Cov}(G,E) = 0$ e $\text{Var}(R) \rightarrow 0$

Conclusão: $\text{Var}(Y) \approx \text{Var}(G)$, sendo $\text{Cov}(G,E) = 0$ a interação genótipo-parcela experimental

Dados ajustados (adjusted means):

Média do Genótipo + Desvios para cada Parcela



Questão: Essas médias ajustadas são as médias “reais” de cada genótipo?

São amostras aleatórias das possíveis médias que o genótipo pode atingir dada as condições ambientais ocorrentes em seu crescimento

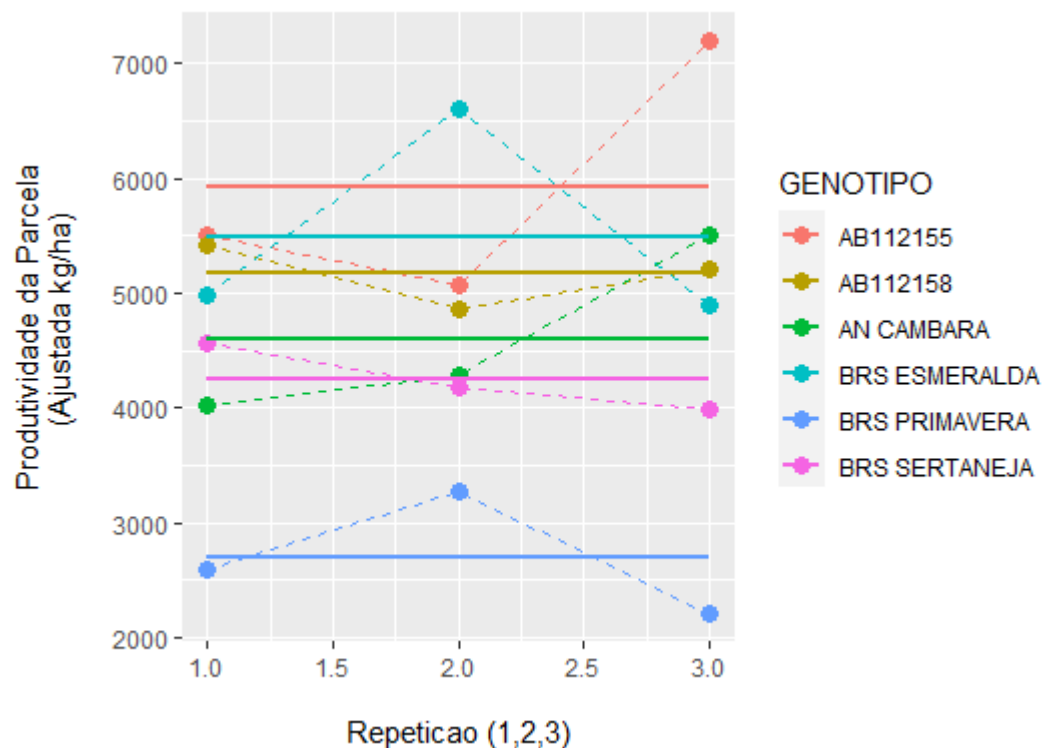
Resultado muito específico do conjunto de múltiplas interações genes x ambiente, além da estabilidade dos transcritos, proteínas e metabólitos presentes durante o ciclo de desenvolvimento da planta

Decomposição da Variação Fenotípica

- Se o genótipo é avaliado em mais de um ambiente, há variação ambiental? (**macro ambiental**)

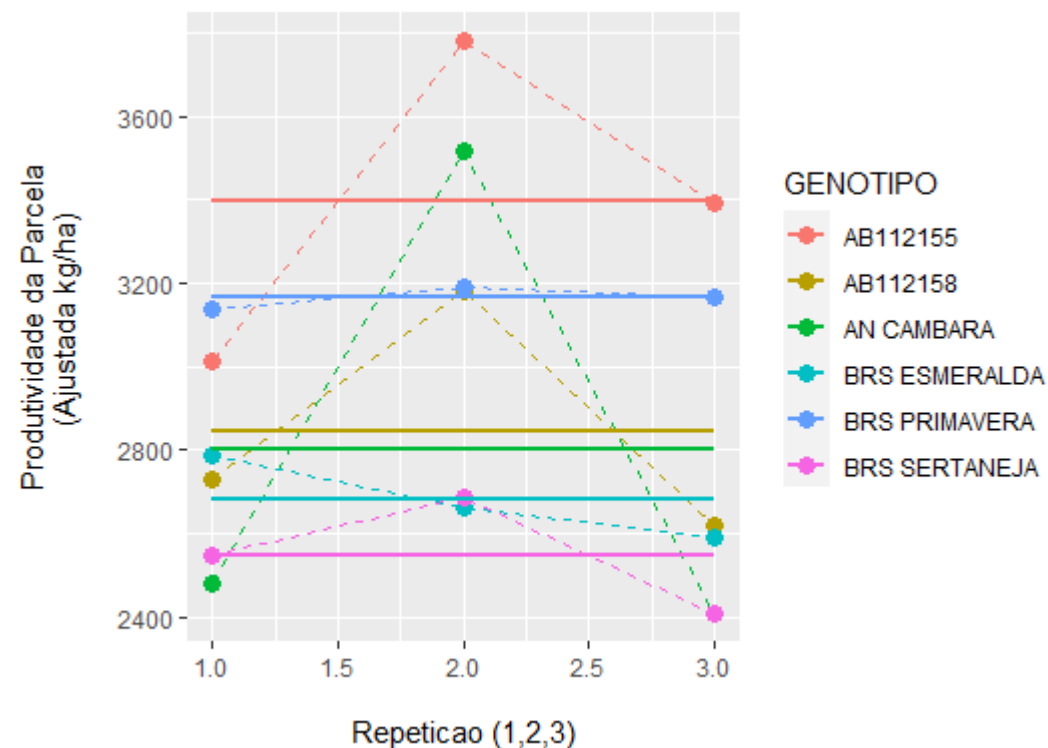
Dados ajustados (adjusted means) para Experimento 1

Média do Genótipo + Desvios para cada Parcela



Dados ajustados (adjusted means) para Experimento 2

Média do Genótipo + Desvios para cada Parcela



- Mudança no ranking dos genótipos (embora o melhor continua sendo o mesmo!)
- Mudança na qualidade experimental (e nível de rendimento, vide eixo y)

Questão: Dá para confiar na repetição 2?
Sugestão: Ver [Revisões sobre análise espacial](#)

Decomposição da Variação Fenotípica

- Se o genótipo é avaliado em mais de um ambiente, há variação ambiental?
- Se há mais genótipos, há variação genética?
- Se há mais genótipos E mais ambientes, é esperado que haja multiplicidade de responsividades genéticas à variação ambiental?

Ensaio Conjuntos (múltiplos experimentos de experimento de campo)

$\text{Var}(Y) = \text{Var}(G) + \text{Var}(E) + \text{Cov}(G, E) + \text{Var}(R)$ sendo G = fatores “genético”s e R = resíduos experimentais

Agora a covariância $\text{Cov}(G, E) = 0$ não mais depende apenas da interação genótipo-parcela experimental

Se $\text{Cov}(G, E) \rightarrow 0$ e $\text{Var}(R) \rightarrow 0$

Conclusão: $\text{Var}(Y) \approx \text{Var}(G) + \text{Var}(E)$,

Porém é possível...

$\text{Var}(G) = \text{Var}(G \text{ geral}) + \text{Var}(G \text{ específica de cada ambiente})$

$\text{Var}(E) = \text{Var}(E \text{ geral}) + \text{Var}(E \text{ específico de cada local})$

- O modelo para decomposição da variação fenotípica vai depender do seu delineamento (de tratamentos, experimental e genético)!

Decomposição da Variação Fenotípica

- Se o genótipo é avaliado em mais de um ambiente, há variação ambiental?
- Se há mais genótipos, há variação genética?
- Se há mais genótipos E mais ambientes, é esperado que haja multiplicidade de responsividades genéticas à variação ambiental?

Ensaio Conjuntos (múltiplos experimentos de experimento de campo)

$\text{Var}(Y) = \text{Var}(G) + \text{Var}(E) + \text{Cov}(G, E) + \text{Var}(R)$ sendo G = fatores “genéticos” e R = resíduos experimentais

Agora a covariância $\text{Cov}(G, E) = 0$ não mais depende apenas da interação genótipo-parcela experimental

Se $\text{Cov}(G, E) \rightarrow 0$ e $\text{Var}(R) \rightarrow 0$

Conclusão: $\text{Var}(Y) \approx \text{Var}(G) + \text{Var}(E)$,

Porém é possível...

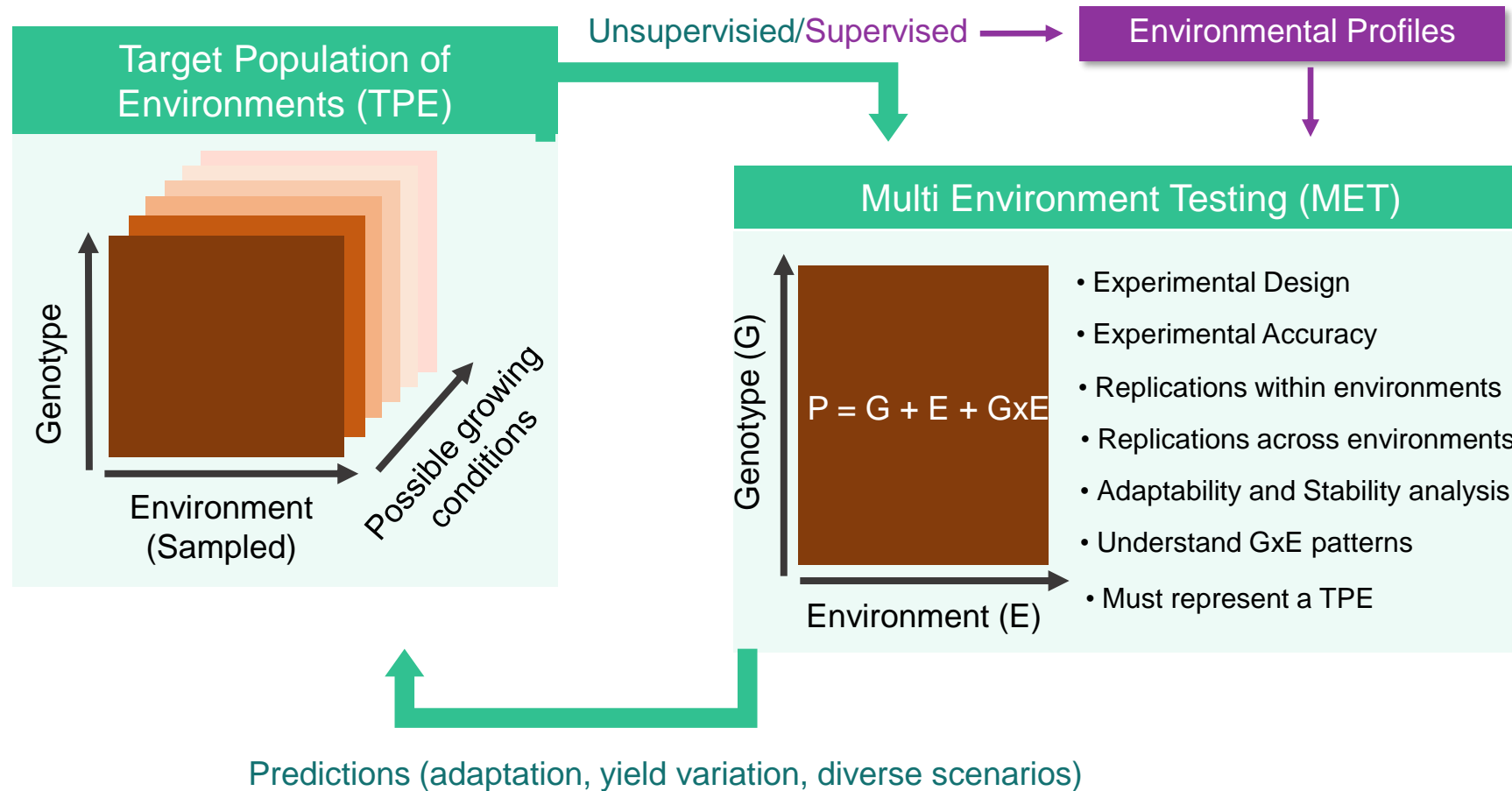
$\text{Var}(G) = \text{Var}(G \text{ geral}) + \text{Var}(G \text{ específica de cada ambiente})$

$\text{Var}(E) = \text{Var}(E \text{ geral}) + \text{Var}(E \text{ específico de cada local})$

Aspectos práticos:

- Tipo de modelo (efeito fixo/aleatório/misto), uso de OLS (Ordinary Least Squares) vs REML **depende do germoplasma e rede experimental**
- Estruturas de variância e covariância “realizadas” (isto é, esperadas a partir de alguma covariável morfológica, HTP, molecular, ambiental, espacial)

Ensaaios Multi-Ambientais e População de Ambientes-Alvo



É importante saber a participação de cada efeito na SQT

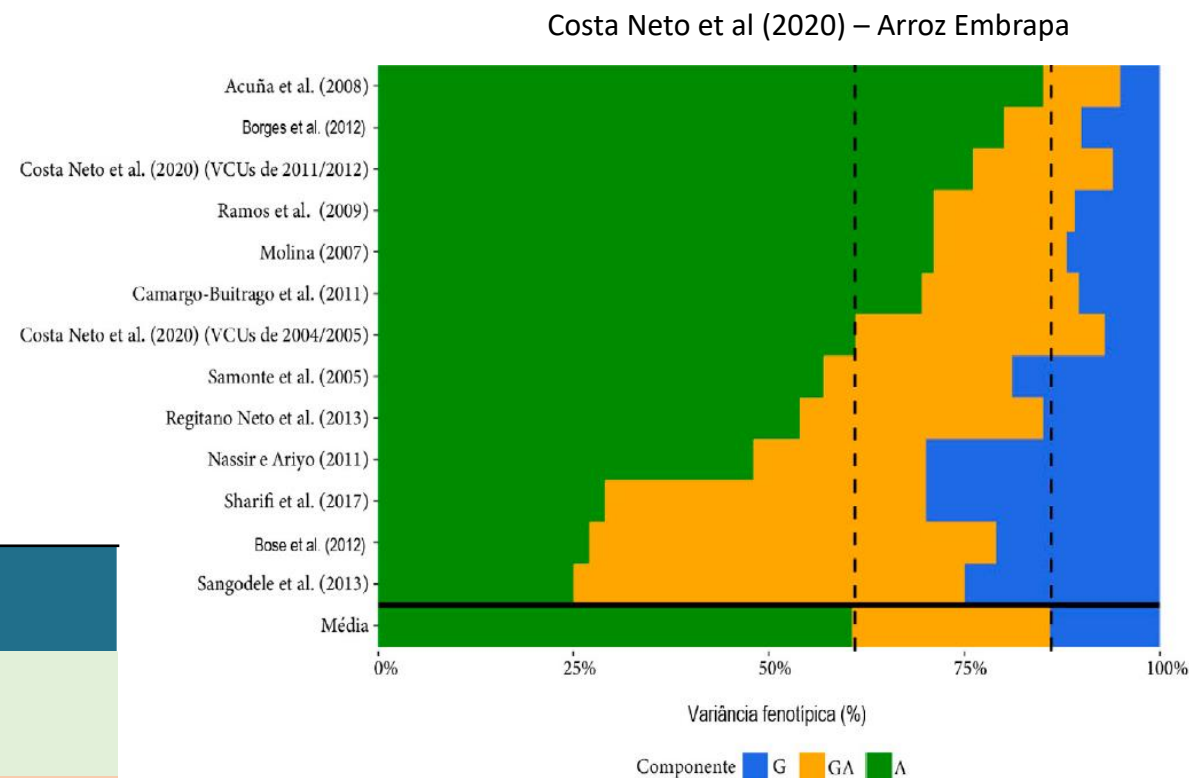
$$Y_{ij} = M + G_i + E_j + GE_{ij} + R_{ij}$$

- Y = fenótipo observado para o genótipo i cultivado no ambiente j
- M = constante (intercepto) da MET (germoplasma e ambientes)
- R = ruído (resíduo) não captado pelas fontes conhecidas
- R = realidade observada – realidade esperada (teorizada)

Adaptação de um dado genótipo i a um dado ambiente j = Ad_{ij}

| Causa (efeito predominante) | Decisão |
|---------------------------------|---|
| $Ad_{ij} = G_i$ | Ampla adaptação dos genótipos ou ambientes muito similares |
| $Ad_{ij} = E_j$ | Ambientes muito heterogêneos e falta de variação genética |
| $Ad_{ij} = GE_{ij}$ | Interação GE complexa, baixa correlação genética entre ambientes |
| $Ad_{ij} = G_i + E_j$ | Interação GE NULA, ou seja, no mesmo mega-ambiente |
| $Ad_{ij} = G_i + GE_{ij}$ | Capitalização da interação GE (complexa ou simples) sob os efeitos genéticos, isto é: $Ad = G + GE > G$ |
| $Ad_{ij} = E_j + GE_{ij}$ | Baixa variabilidade genética (de efeito amplo), porém alta responsividade para ambientes específicos e contrastantes, isto é, $Ad = E + GE > G$ |
| $Ad_{ij} = G_i + E_j + GE_{ij}$ | Modelo completo, adaptação ambiente-específico ou ampla |

Costa Neto (in prep)



E para a sua cultura-alvo?

E seu germoplasma?

E na sua região-alvo?

Em quais traits?

Ambiente

Cooper et al., (2014); Xu (2016); Resende et al., (2020); Costa-Neto et al. (2021); Crossa et al (2021 “plant breeding triangle”)

Agronomic Standpoint

- Single-unit (planting date, location, management)
- Location + Planting date = Fixed + Random Factors
- Actual Management = Expected Management + biotic and abiotic interferences

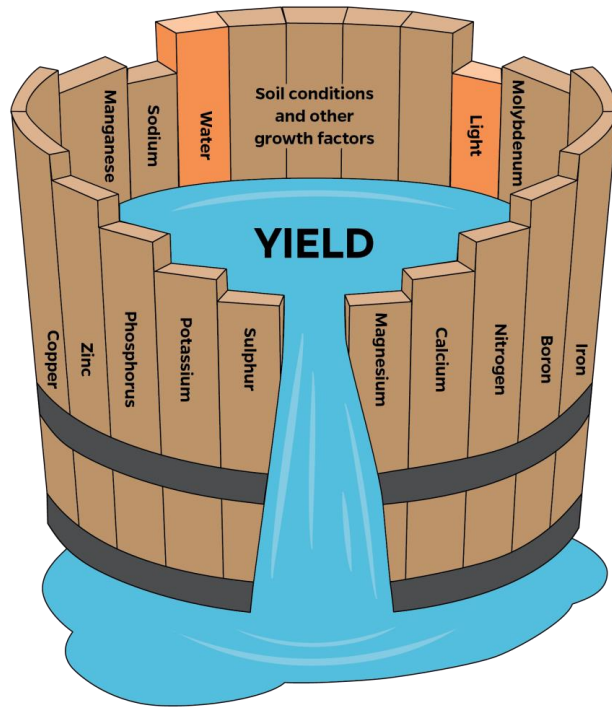
Geneticist Standpoint (Plant Breeders)

- Near-Iso Environments (managed conditions)
- Drought-Stress Trials (screening)
- Nitrogen level conditions (screening)
- A non-genetic source of variation (to be controlled as a fixed factor)

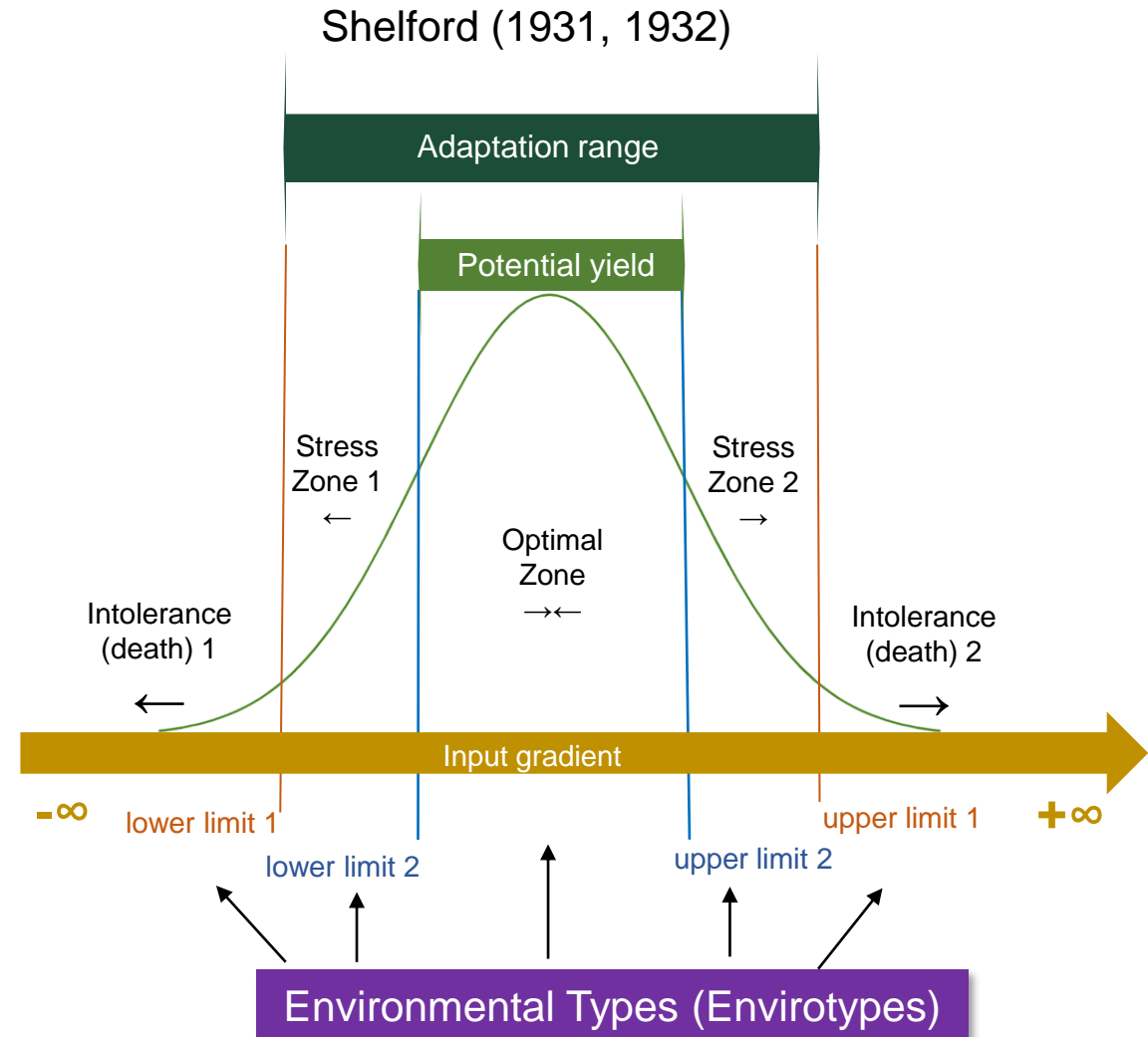
Ecophysiological Standpoint

- Relation between the **availability of inputs** and their resource absorption/allocation
- Availability as function of amount and frequency
- Core of process **linking** soil-plant-atmosphere dynamics (plant-env-pathogen)

Qualidade Ambiental

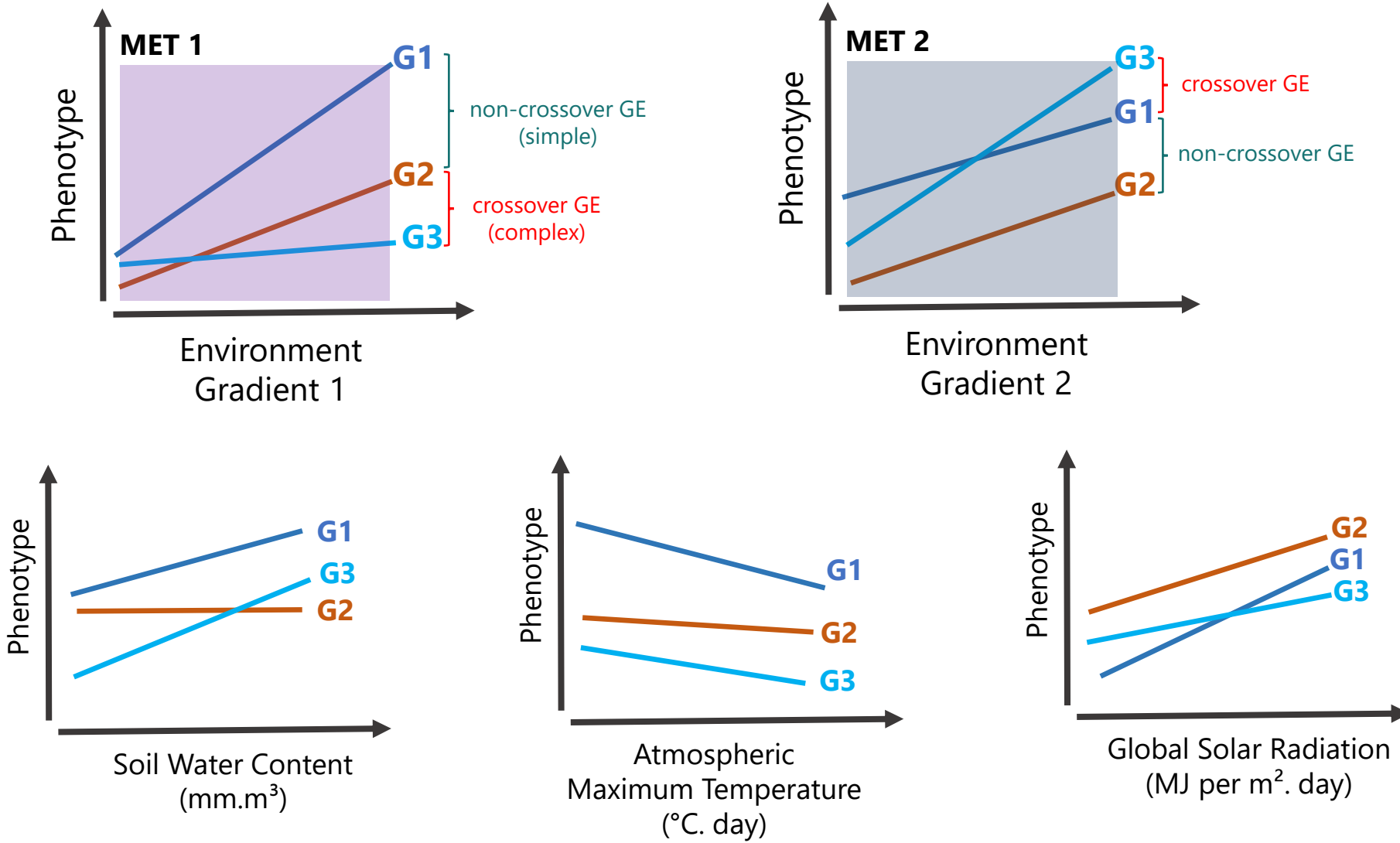


Liebig's Law of Minimum (1840)
Whitson and Walster (1912)



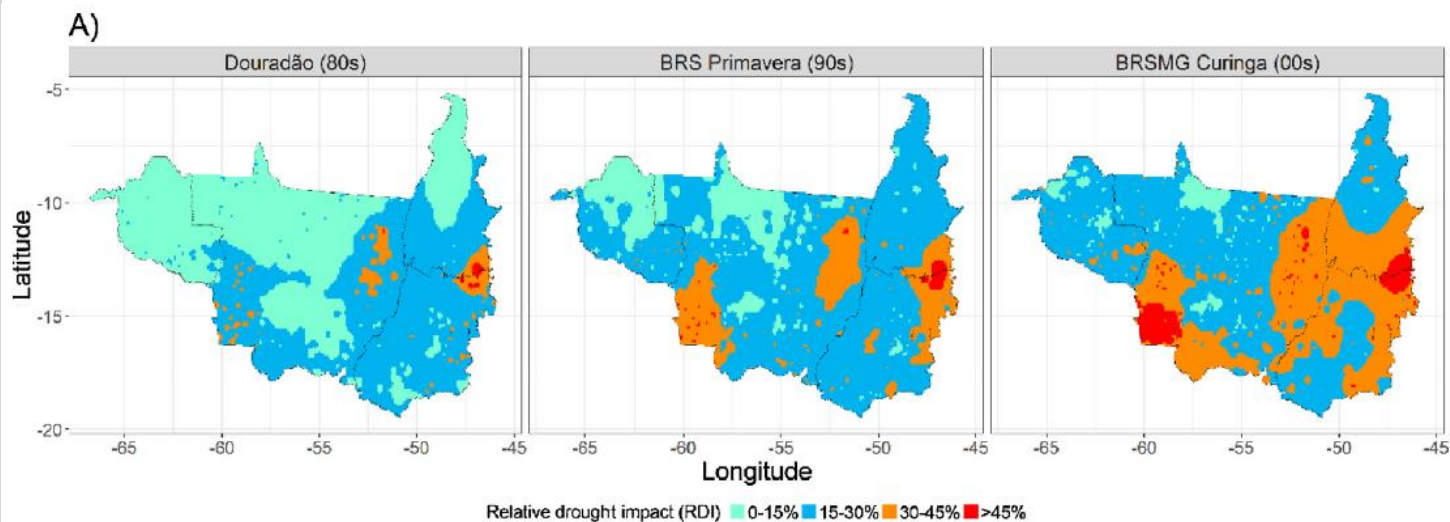
Relação entre Norma de Reação e Tipos de Interação GxE

- Interação GxE Simples (quantitativa), Complexa (Cruzada), Nula
- É ESPECÍFICA do germoplasma e da rede experimental

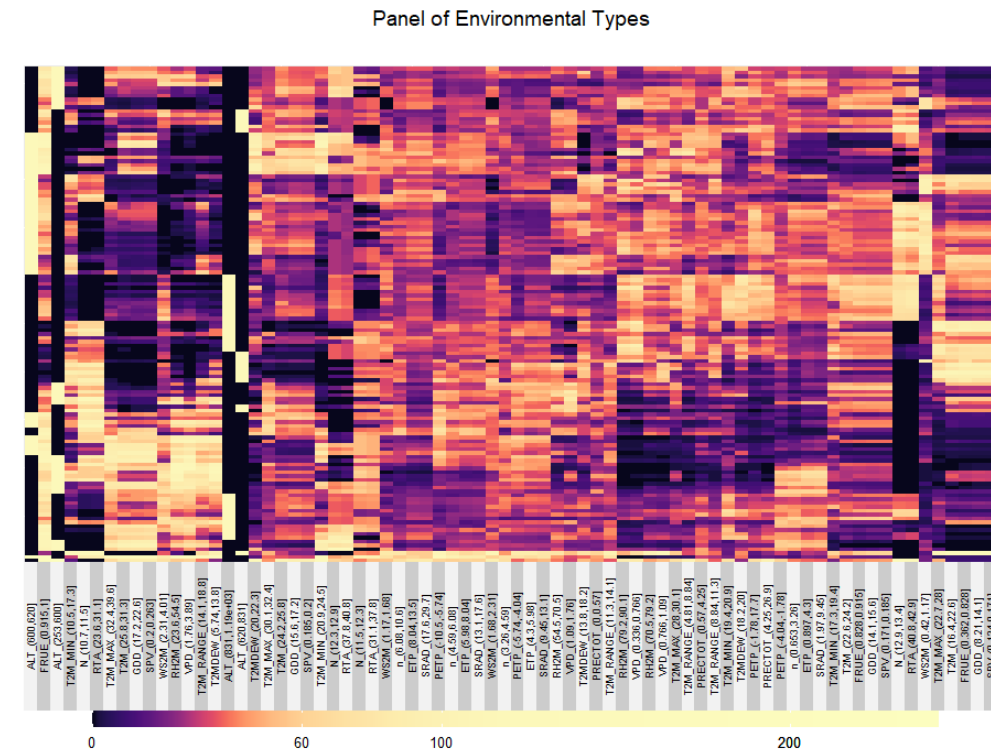


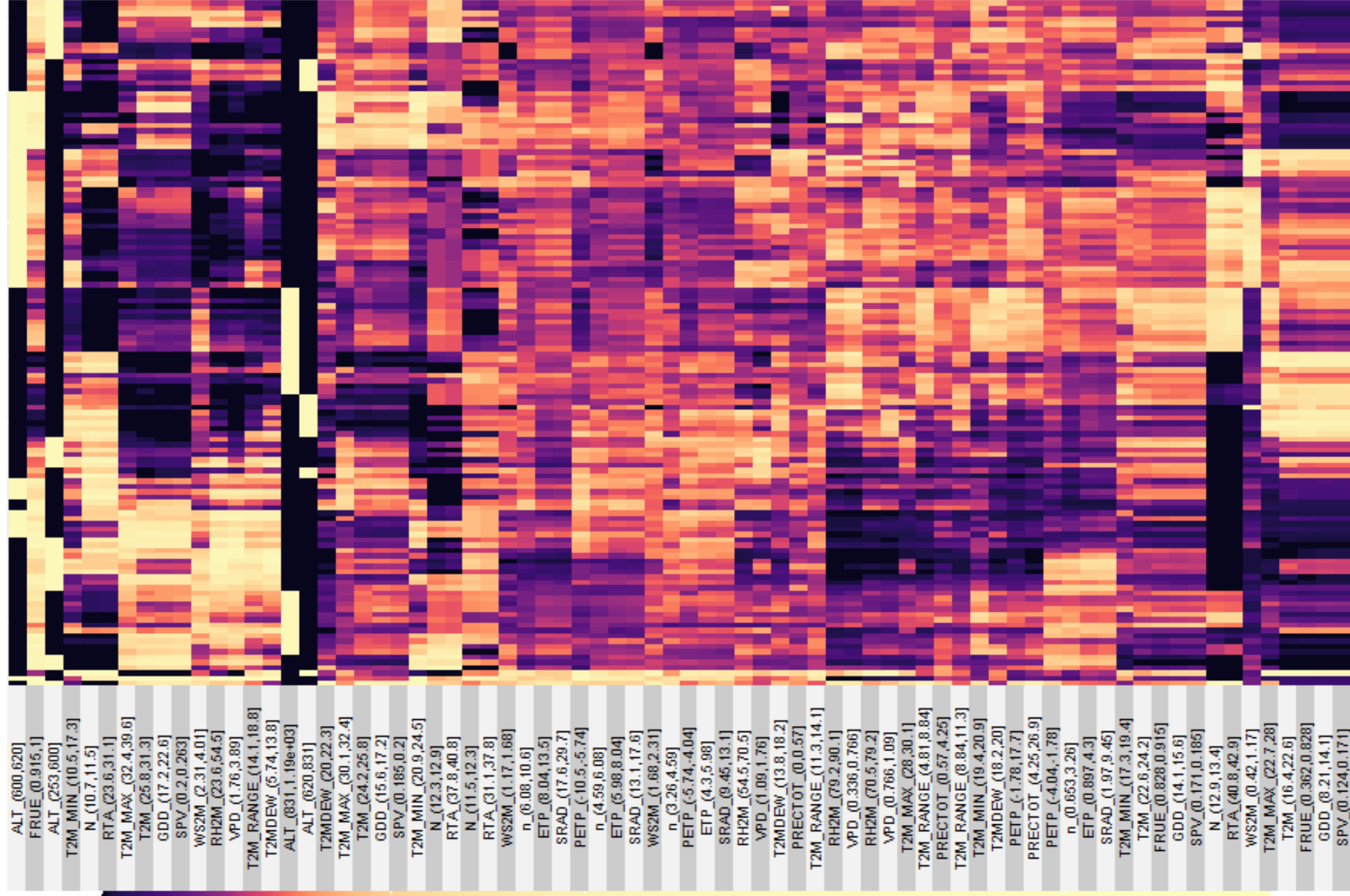
Ambiente e sua Tipologia

- Allard & Brashaw (1963): Padrões previsíveis e não previsíveis – Será?
- Tipologia: frequência e impacto de dado fator ambiental sobre um certo ambiente
- Relação entre tipologia do “berçário” com a da população-alvo do programa
- Se há muita distancia, é necessário mudar o programa!



Heinemann et al. (2015, 2019)

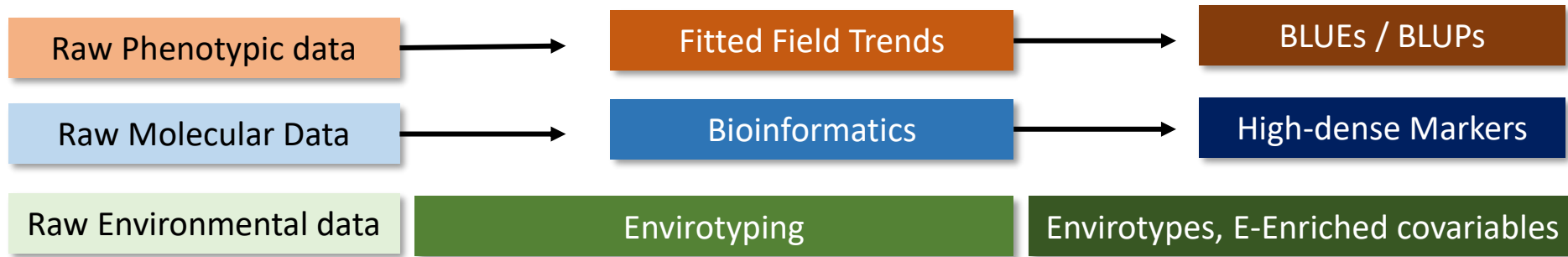




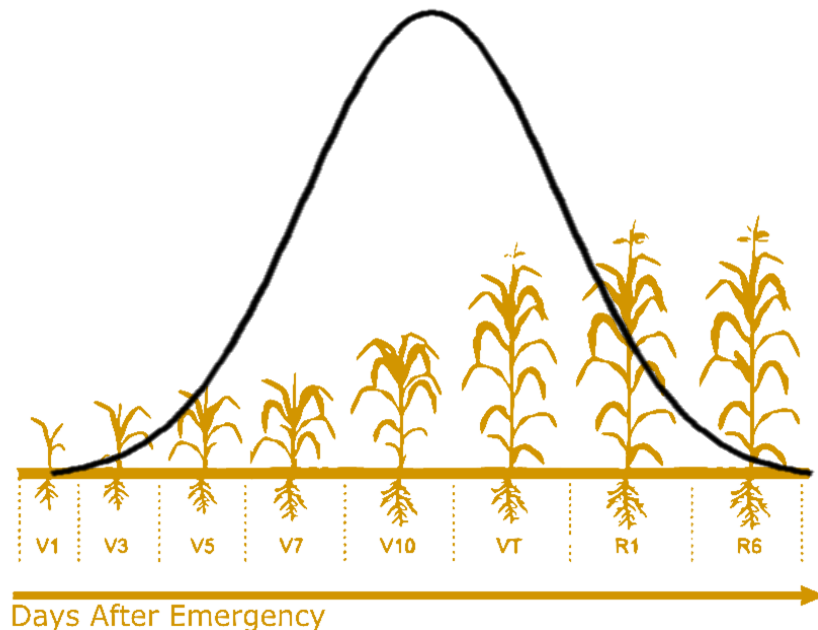
Pipeline para “Ambitipagem” (Envirotyping)



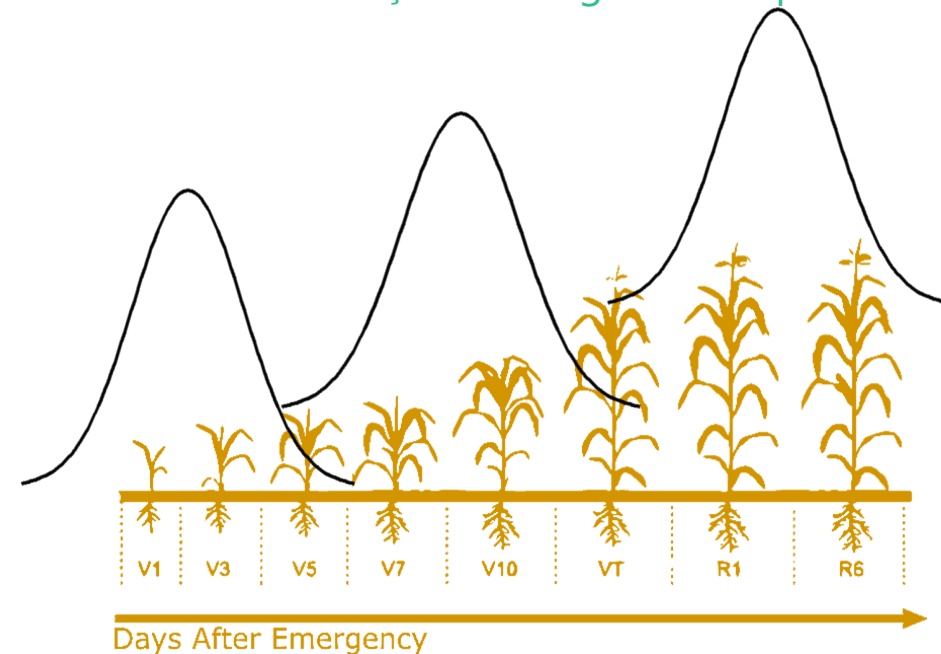
- Disponibilidade de dados e de seu processamento!



Distribuição durante todo o ciclo



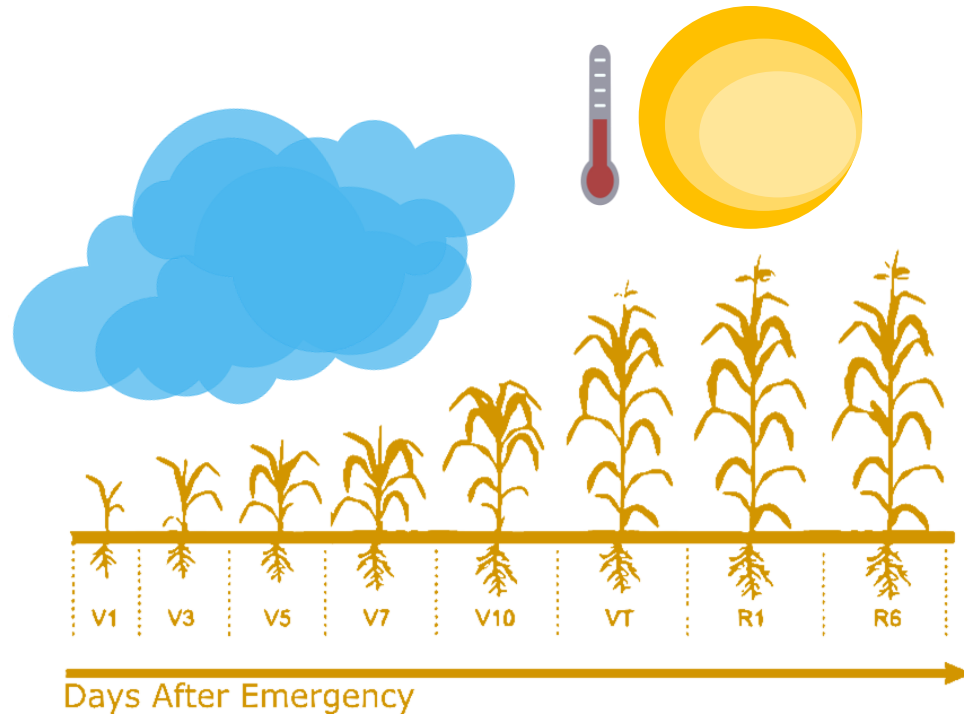
Distribuição ao longo do tempo



Descritores do ambiente

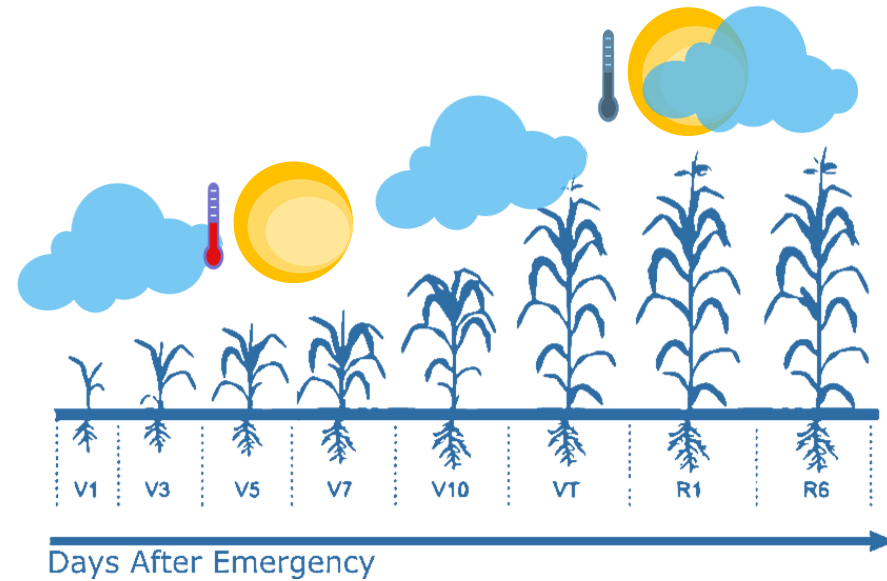
Valores cumulativos (pobre descrição!)

Environment 1



- Precipitation: 560 mm/cycle
- Temperature: 962 °C/cycle
- Radiation: 724 MJ m²/ cycle

Environment 2

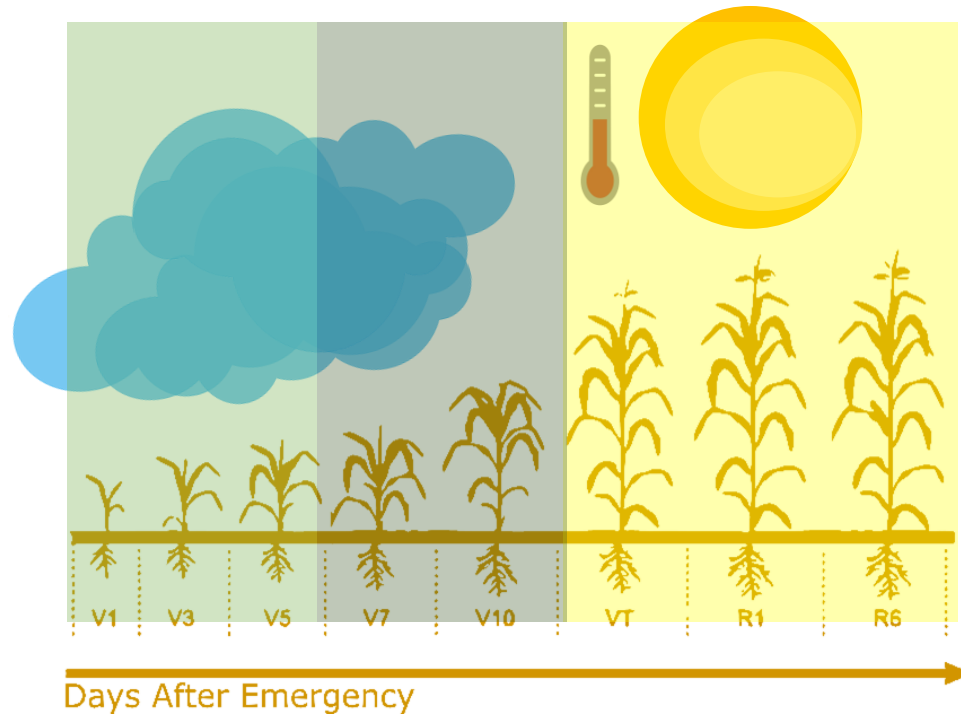


- Precipitation: 560 mm/cycle
- Temperature: 962 °C/cycle
- Radiation: 724 MJ m²/ cycle

Descritores do ambiente

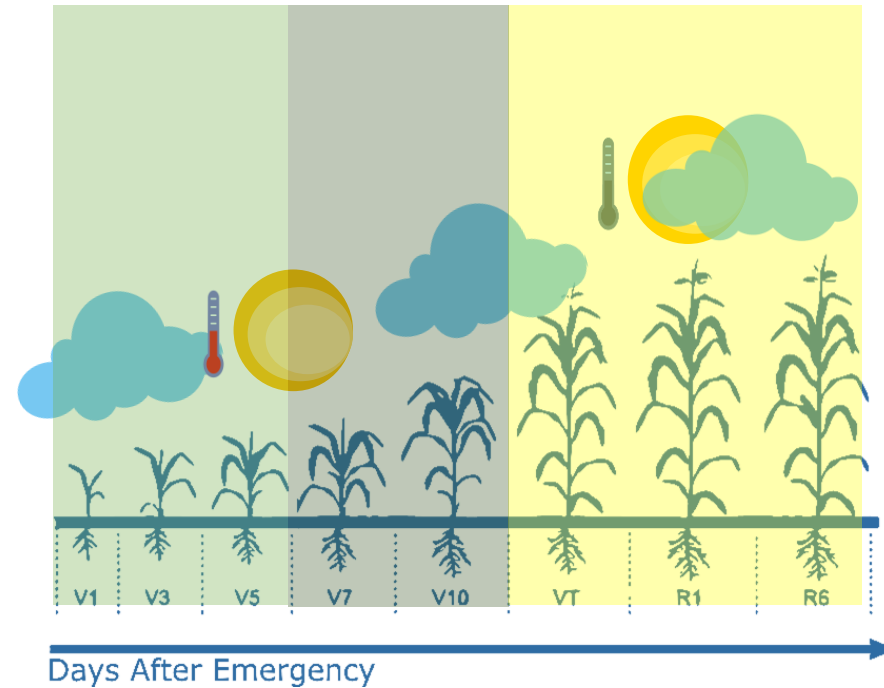
Valores cumulativos por fase fenológica (melhor, mas ainda pobre descrição!)

Environment 1



- Precipitation (mm): 300 (T1); 260 (T2); 0 mm (T3)
- Temperature (°C): 150 (T1); 100 (T2); 712 (T3)
- Radiation (MJ/m2): 144.8 (T1); 115.8 (T2); 463.4 (T3)

Environment 2



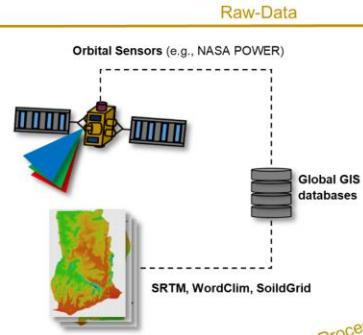
- Precipitation (mm): 224 (T1); 168 (T2); 168 (T3)
- Temperature (°C): 150 (T1); 100 (T2); 712 (T3)
- Radiation (MJ/m2): 144.8 (T1); 115.8 (T2); 463.4 (T3)

Module 1

Remote Data Collection

- get_weather()
- extract_GIS()

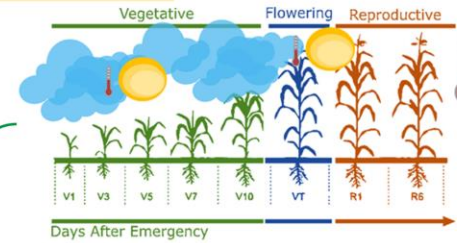
Raw-data from environmental sensors or get_weather() and extract_GIS() can be used in the further steps or processed



Raw-data Processing

- SummaryWTH()
- ProcessWTH()
- param_temperature()
- param_radiation()
- param_atmospheric()

Data processing involves the quality control and computation of additional variables



Divide time intervals in crop development helps to characterize environments and discover a panel of envirotype descriptors

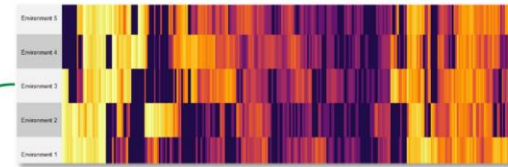
Module 2

Environmental Characterization

- env_typing()
- W_matrix()

Environmental covariables

Envirotype descriptors



Panel of envirotype descriptors or environmental covariables can be incorporated in predictive tools as environmental markers or to study similarity among environments

Module 3

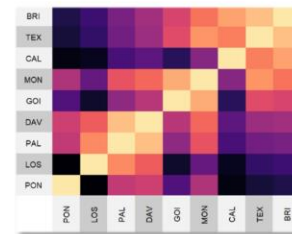
Environmental Similarity

- env_kernel()

Relatedness across environments derived from envirotyping data can be used to group environments with similar pattern

Enviromic Kernels

Kernel Models for phenotype prediction can be made involving genomic and enviromic-based sources derived from environmental similarity kernels



Enviromic-based Kernels Models

- get_kemel()
- kernel_model()

Models:

1. Genotypic Effects (MM and MDs)
2. Enviromic-enriched Main Effects (EMM and EMDs)
3. Enviromic-based Reaction-Norm (RNMM and RNMDs)

Outputs

Reaction-Norm

examples Factorial Regression GxE analysis

Environmnetal Grouping

Clustering (K-mean) TPE definition

Genomic-based Predictions

Genomic-estimated Reaction-Norm Multi-environment Prediction



Remote Sensing (Digital Envirotyping)



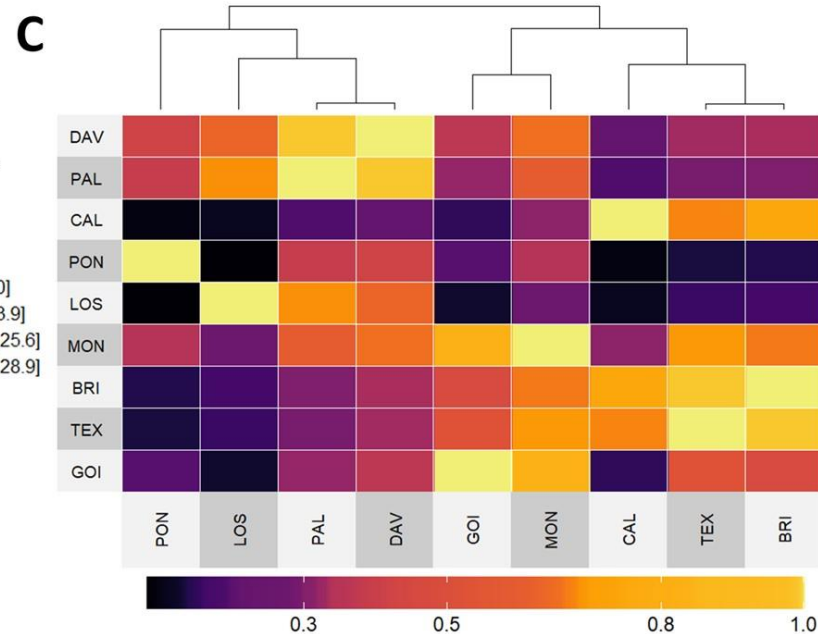
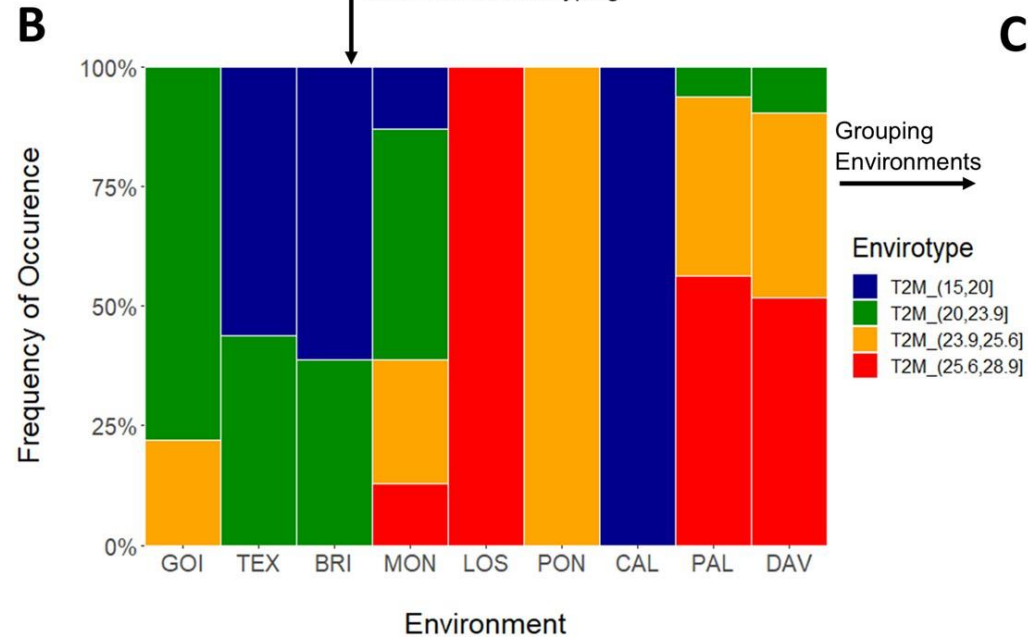
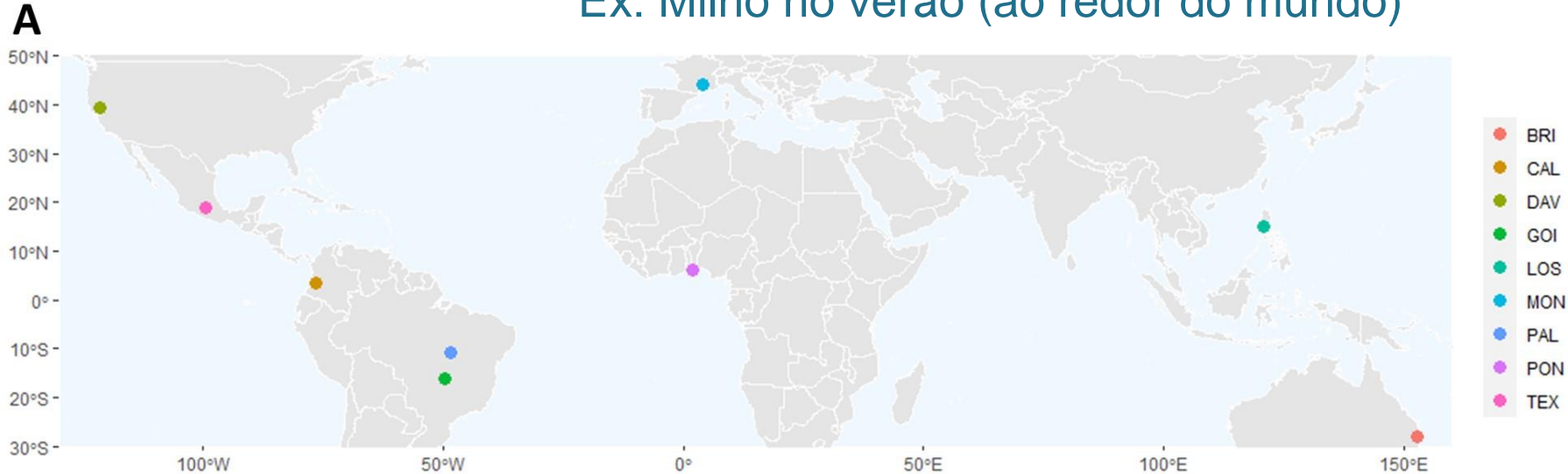
get_weather() and processWTH()

| Source | Environmental Factor | Unit |
|-------------------------|---|-------------------------------------|
| Nasa Power ¹ | Top-of-atmosphere insolation | $\text{MJ m}^{-2} \text{d}^{-1}$ |
| | Average insolation incident on a horizontal surface | $\text{MJ m}^{-2} \text{d}^{-1}$ |
| | Average downward longwave radiative flux | $\text{MJ m}^{-2} \text{d}^{-1}$ |
| | Wind speed at 10 m above the surface of the earth | m s^{-1} |
| | Minimum air temperature at 2 m above the surface of the earth | $^{\circ}\text{C d}^{-1}$ |
| | Maximum air temperature at 2 m above the surface of the earth | $^{\circ}\text{C d}^{-1}$ |
| | Dew-point temperature at 2 m above the surface of the earth | $^{\circ}\text{C d}^{-1}$ |
| | Relative air humidity at 2 m above the surface of the earth | % |
| | Rainfall precipitation (P) | mm d^{-1} |
| SRTM ² | Elevation (above sea level) | m |
| Computed ³ | Effect of Temperature on Radiation use Efficiency | - |
| | Evapotranspiration (ETP) | mm d^{-1} |
| | Atmospheric water deficit P-ETP | mm d^{-1} |
| | Deficit of vapor Pressure | kPa d^{-1} |
| | Slope of saturation vapor pressure curve | $\text{kPa } ^{\circ}\text{C}^{-1}$ |
| | Temperature Range | $^{\circ}\text{C d}^{-1}$ |
| | Global Solar Radiation based on Latitude and Julian Day | $\text{MJ m}^{-2} \text{d}^{-1}$ |

Worldwide soil data : <https://soilgrids.org/>

Ambiente: janela entre plantio e colheita em uma dada região para um dado cultivo

Ex: Milho no verão (ao redor do mundo)



B: tipologias para temperatura do ar
C: similaridade (via distancia)

No ano avaliado, Palmas (TO) se assemelha a Davis na California!

Quais conclusões são possíveis através dessa prospecção de tipologias?

Interpretação Analítica vs Empírica

Empírica

- Trait-específica
- Práticas e facilmente implementáveis
- Não demanda informações adicionais
- Distribuições dos fenótipos
- Separa padrão/ruído fenotípico via procedimentos algébricos/computacionais
- Abstração dos efeitos do ambiente
- Dados balanceados e desbalanceados
- Correlações e covariáveis fenotípicas como 'sinal'

Interpretação Analítica vs Empírica

Empírica

- Trait-específica
- Práticas e facilmente implementáveis
- Não demanda informações adicionais
- Distribuições dos fenótipos
- Separa padrão/ruído fenotípico via procedimentos algébricos/computacionais
- Abstração dos efeitos do ambiente
- Dados balanceados e desbalanceados
- Correlações e covariáveis fenotípicas como 'sinal'

Analítica (biológica)

- Trait-específica, cultivar-específica ou ambiente-específico
- Métodos ainda em desenvolvimento/integração multidisciplinar
- Se baseia no uso de informações adicionais
- Distribuições dos fenótipos e outras data bases
- Separa padrão/ruído fenotípico, ambiental, genético via procedimentos algébricos/computacionais e modelagem ecofisiológica
- Tentativa de descrição dos efeitos do ambiente
- Dados balanceados e desbalanceados (alguns modelos)
- Correlações e covariáveis fenotípicas como 'sinal' a ser explorado pelas outras fontes de dados

Interpretação Analítica vs Empírica

Empírica

- Trait-específica
- Práticas e facilmente implementáveis
- Não demanda informações adicionais
- Distribuições dos fenótipos
- Separa padrão/ruído fenotípico via procedimentos algébricos/computacionais
- Abstração dos efeitos do ambiente
- Dados balanceados e desbalanceados
- Correlações e covariáveis fenotípicas como 'sinal'

Analítica (biológica)

- Trait-específica, cultivar-específica ou ambiente-específico
- Métodos ainda em desenvolvimento/integração multidisciplinar
- Se baseia no uso de informações adicionais
- Distribuições dos fenótipos e outras data bases
- Separa padrão/ruído fenotípico, ambiental, genético via procedimentos algébricos/computacionais e modelagem ecofisiológica
- Tentativa de descrição dos efeitos do ambiente
- Dados balanceados e desbalanceados (alguns modelos)
- Correlações e covariáveis fenotípicas como 'sinal' a ser explorado pelas outras fontes de dados

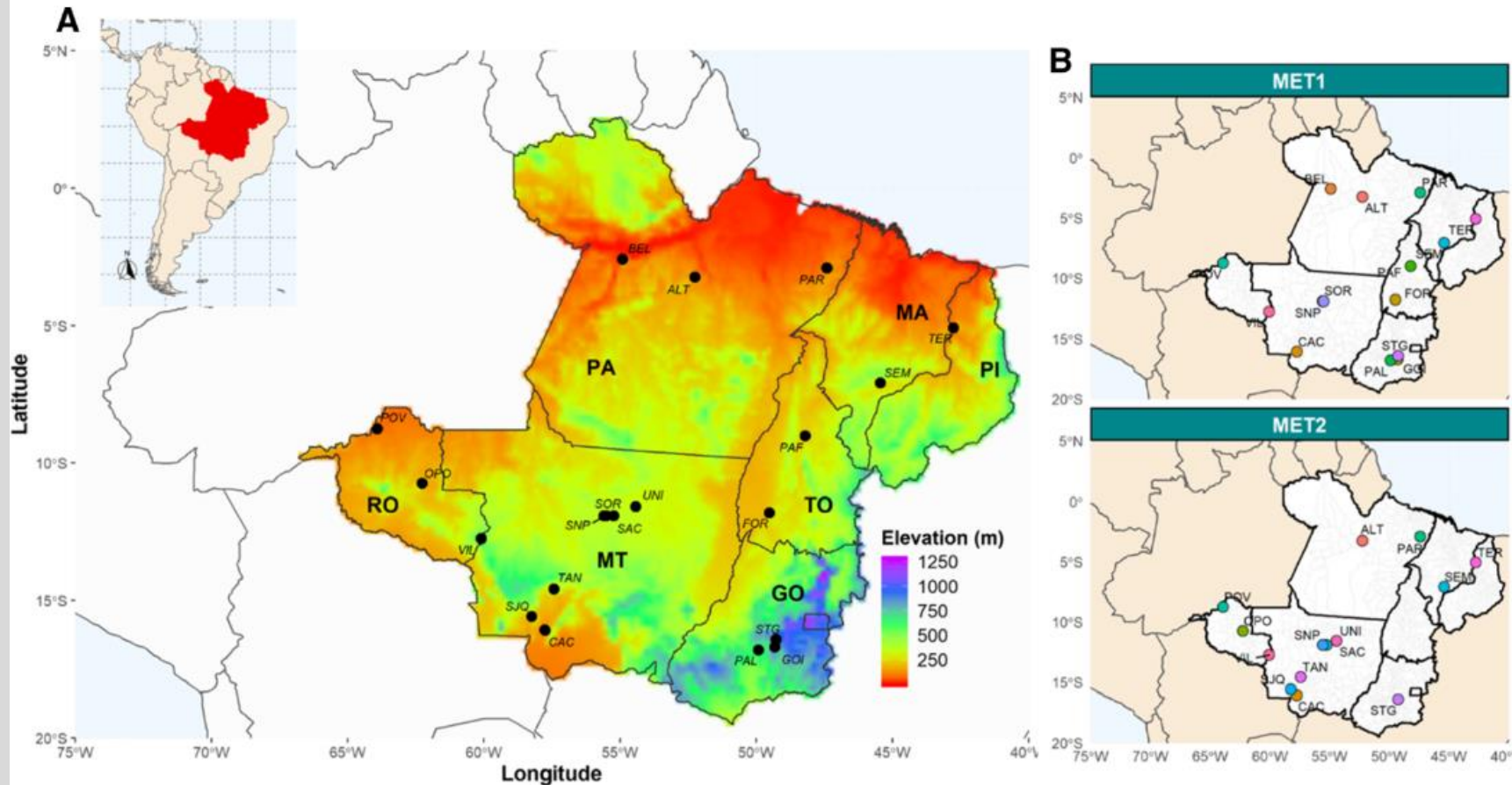
Será que o uso de dados ambientais para entender GE é algo recente e pouco explorado? [\[veja\]](#):

- Fisher, R. A. (1924). Studies in crop variation. III. The influence of rainfall on the yield of wheat at Rothamsted
- Só nos anos 70 tem mais de 10 trabalhos! (Perkins, Freeman, Wood) [\[ver\]](#)
- Nos anos 1980 e começo dos 1990 foi resgatado por Denis, van Eeuwijk e Crossa
- Nos anos 2000 entrou na análise de QTL x E: Malosseti, Reynolds, Crossa, van Eeuwijk, Cooper, Duarte, Martins (mapas geográficos)
- Nos anos 2010 integrou os modelos de crescimento de cultivos (dos anos 80) para caracterização e predição: Cooper, Messina, Chenu, Hammer, Heinemann
- A partir de 2014 finalmente conseguiu chamar a atenção com Jarquin et al (2014) e Heslot et al (2014)

Modelo Multi-Ambiental (análise conjunta)

Estudo de caso em arroz de terras altas

[Costa-Neto et al \(2020\)](#)



Possíveis perguntas:

- Quais são os fatores que podem ser usados?
- Ano, Local, Ambiente?
- Há manejos culturais diferentes?
- Há locais em anos diferentes?
- Há genótipos em locais em anos diferentes?
- Variabilidade climática entre locais?
- Variabilidade climáticas entre anos?
- Doenças que ocorrem em locais específicos?
- Devo modelar conjunto ou ano específico?

Modelo Multi-Ambiental (análise conjunta)

Estudo de caso em arroz de terras altas
[Costa-Neto et al \(2020\)](#)

Como posso montar meu modelo “multi-ambiente”?
Quais são as fontes de variação fenotípica?

MET 1 : $Y_{ij} = M + G_i + L_j + GL_{ij} + R_{ij}$

MET 2 : $Y_{ij} = M + G_i + L_j + GL_{ij} + R_{ij}$

Opção: $Y_{ijk} = M + G_i + L_j + A_k + GL_{ij} + GA_{ik} + LA_{jk} + GLA_{ijk} + R_{ijk}$

Y_{ijk} = fenótipo do i genótipo avaliado no j local dentro do k ano

Opção: $Y_{ij} = M + G_i + E_j + GE_{ij} + R_{ij}$

Y_{ij} = fenótipo do i genótipo avaliado no j ambiente, sendo ambiente combinação de ano-local

| Joint-analysis | DF ^b | SS ^c | MS ^d | Fraction ^c (G + L+G × L) |
|----------------------|-----------------|-----------------|-----------------|-------------------------------------|
| <i>MET1</i> | | | | |
| Location (L) | 14 | 178,975,523 | 12,783,966* | 0.61 |
| Genotype (G) | 19 | 21,458,667 | 1,129,404* | 0.07 |
| G × L | 197 | 93,012,871 | 349,667* | 0.32 |
| Error/r ^a | 880 | 592,031,836 | 108,548 | – |
| <i>MET2</i> | | | | |
| Location (L) | 13 | 1,581,700,166 | 121,669,244* | 0.76 |
| Genotype (G) | 22 | 131,076,146 | 5,958,007* | 0.06 |
| G × L | 286 | 364,974,112 | 1,2761,33* | 0.18 |
| Error/r ^a | 1146 | 729,771,323 | 636,799 | – |

Interpretação “empírica” (clássica)

Fundamenta-se no uso apenas de dados fenotípicos e procedimentos algébricos

Exemplos:

GGEbiplot, AMMI, FGGE, FA mixed models, Wricke (Ecovalencia), Finlay-Wilkinson e mais 50 outros diferentes....

Conceito-Chave para entender:

Visualizar as tabelas de dupla-entrada! (linha por coluna)

$$Y_{pq} = \begin{bmatrix} \hat{Y}_{11} & \cdots & \hat{Y}_{1q} \\ \vdots & \ddots & \vdots \\ \hat{Y}_{p1} & \cdots & \hat{Y}_{pq} \end{bmatrix} = (\mathbf{1}_{pq} \otimes \mu) + (\mathbf{1}_p \otimes E'_q) + (\mathbf{1}'_q \otimes G_p) + GE_{pq} \quad (2)$$

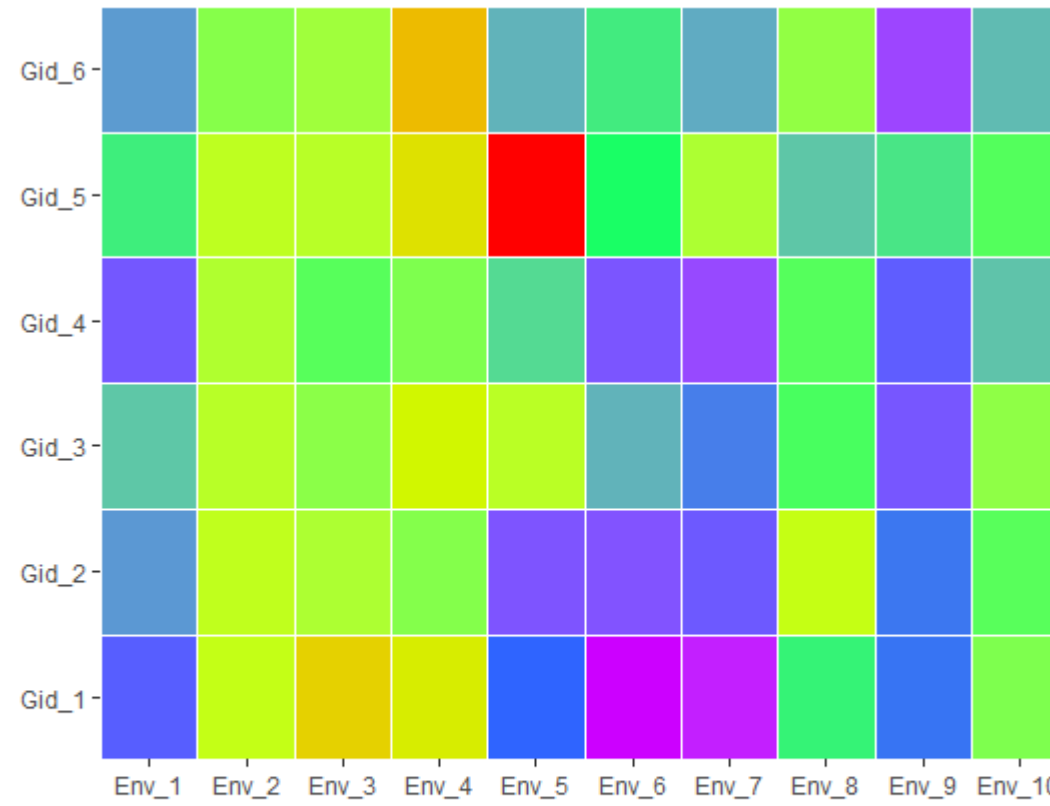
em que $M = (\mathbf{1}_{pq} \otimes \mu)$ é a média geral dos ensaios; $E = (\mathbf{1}_p \otimes E'_q)$ é a matriz de efeitos ambientais; $G = (\mathbf{1}'_q \otimes G_p)$ é a matriz de efeitos genotípicos; e $GE = GE_{pq}$ é a matriz de efeitos de interação GE (todas as matrizes de dimensão $p \times q$).

Y_{pq} é uma matriz com genótipos nas linhas e ambientes nas colunas

Conceito-Chave para entender (análise multi-ambiental ou multi-trait):

Visualizar as tabelas de dupla-entrada! (linha por coluna)

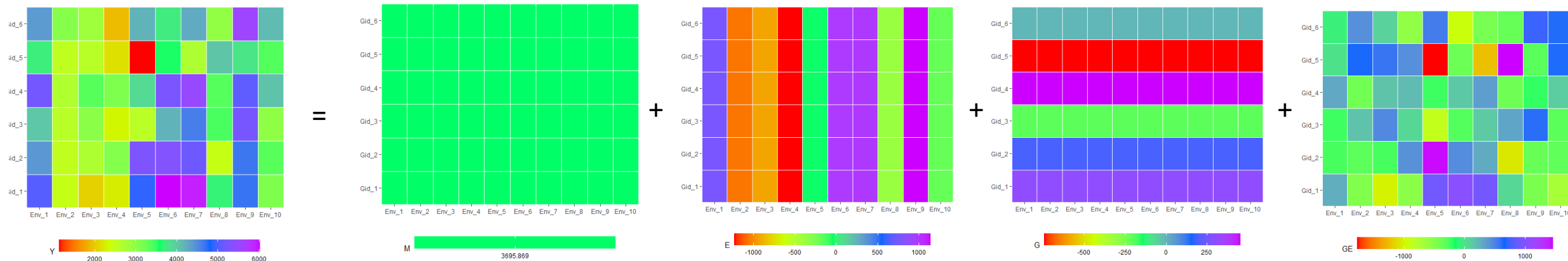
$$Y_{pq} = \begin{bmatrix} \hat{Y}_{11} & \cdots & \hat{Y}_{1q} \\ \vdots & \ddots & \vdots \\ \hat{Y}_{p1} & \cdots & \hat{Y}_{pq} \end{bmatrix} = (\mathbf{1}_{pq} \otimes \boldsymbol{\mu}) + (\mathbf{1}_p \otimes \mathbf{E}'_q) + (\mathbf{1}'_q \otimes \mathbf{G}_p) + \mathbf{GE}_{pq}$$



Conceito-Chave para entender (análise multi-ambiental ou multi-trait):

Visualizar as tabelas de dupla-entrada! (linha por coluna)

$$Y_{pq} = \begin{bmatrix} \hat{Y}_{11} & \cdots & \hat{Y}_{1q} \\ \vdots & \ddots & \vdots \\ \hat{Y}_{p1} & \cdots & \hat{Y}_{pq} \end{bmatrix} = (\mathbf{1}_{pq} \otimes \boldsymbol{\mu}) + (\mathbf{1}_p \otimes \mathbf{E}'_q) + (\mathbf{1}'_q \otimes \mathbf{G}_p) + \mathbf{GE}_{pq}$$



As observações fenotípicas de um mesmo genótipo em diferentes ambientes podem ser vistas como “traits” diferentes

Por esse motivo podemos modelar “correlações” (covariâncias normalizadas)

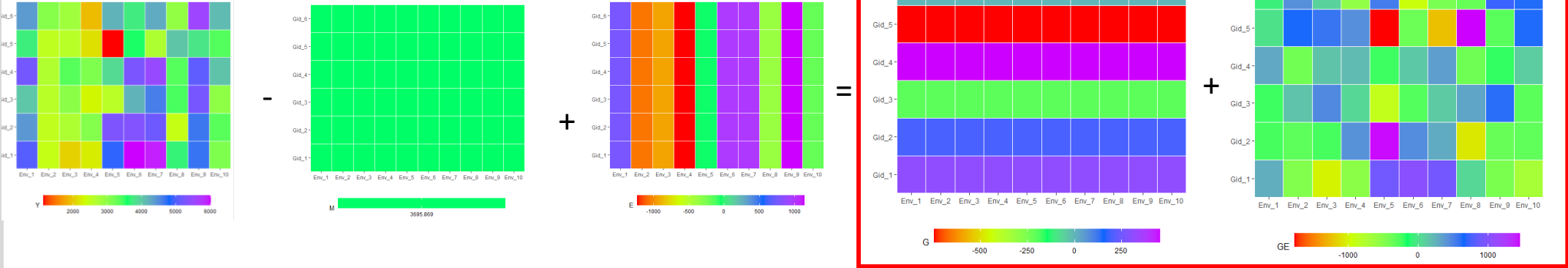
Para ensaio multi-ambiente = interação GxE

Para ensaio multi-trait = pleiotropia e epistasia! (GxG)

Podemos montar a tabela também como genótipo x trait-ambiente

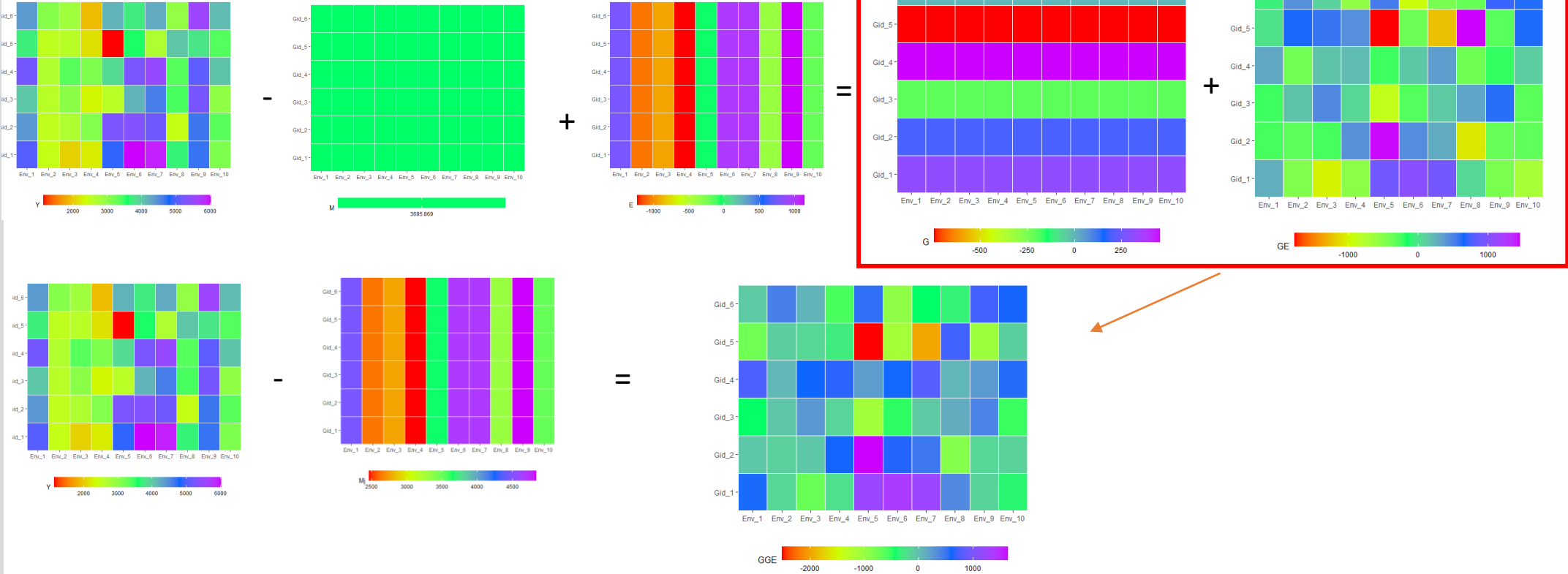
Modelo Site-Regression (SREG, G+GE)

Visualizar as tabelas de dupla-entrada! (linha por coluna)



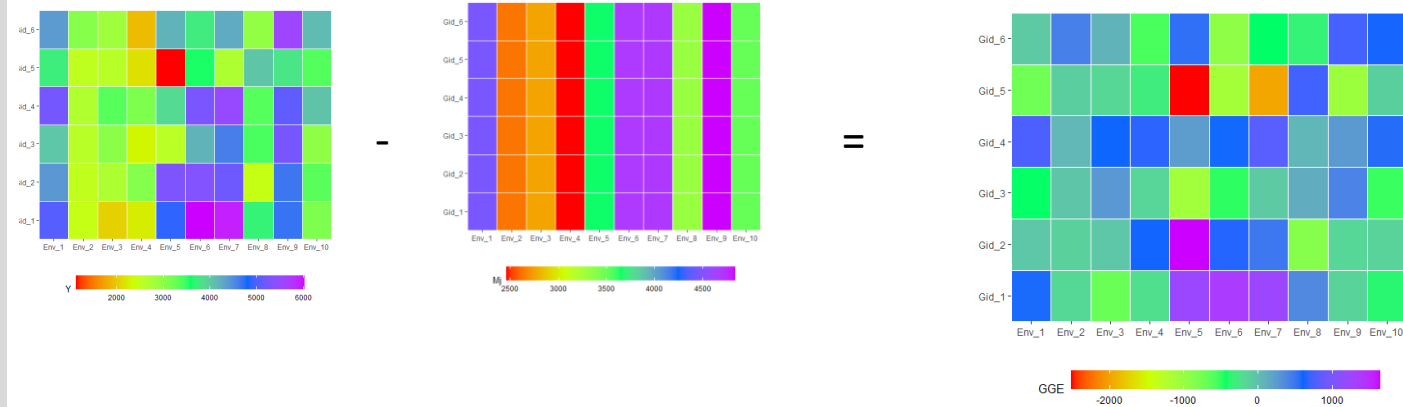
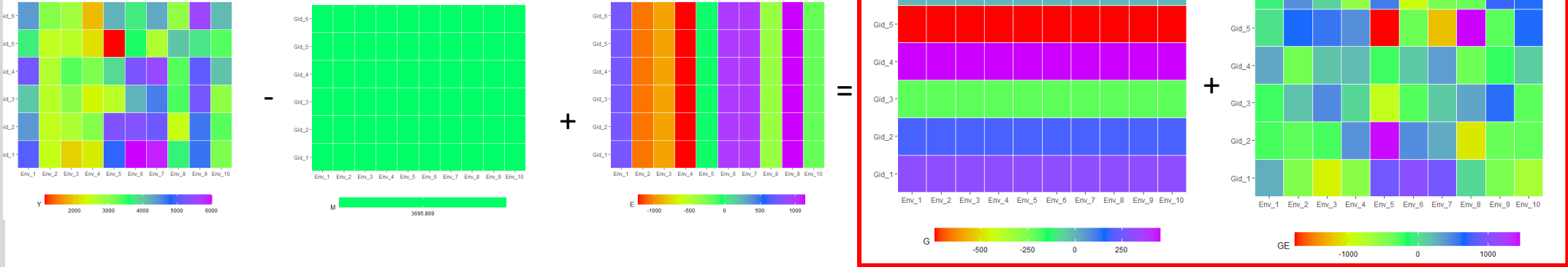
Modelo Site-Regression (SREG, G+GE)

Visualizar as tabelas de dupla-entrada! (linha por coluna)



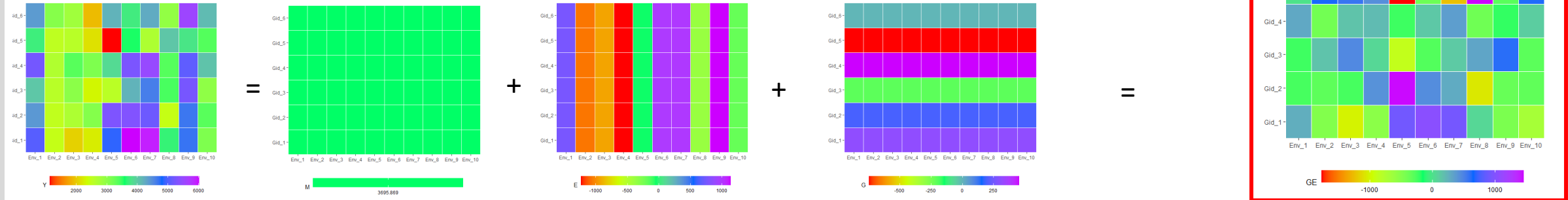
Modelo Site-Regression (SREG, G+GE)

Visualizar as tabelas de dupla-entrada! (linha por coluna)



Modelo Additive main effects and multiplicative interaction (AMMI, double-centered GE)

Visualizar as tabelas de dupla-entrada! (linha por coluna)



Decomposição dos efeitos multiplicativos

AMMI clássico: usa SVD (singular value decomposition), também é possível usar PCA (principal component analysis)

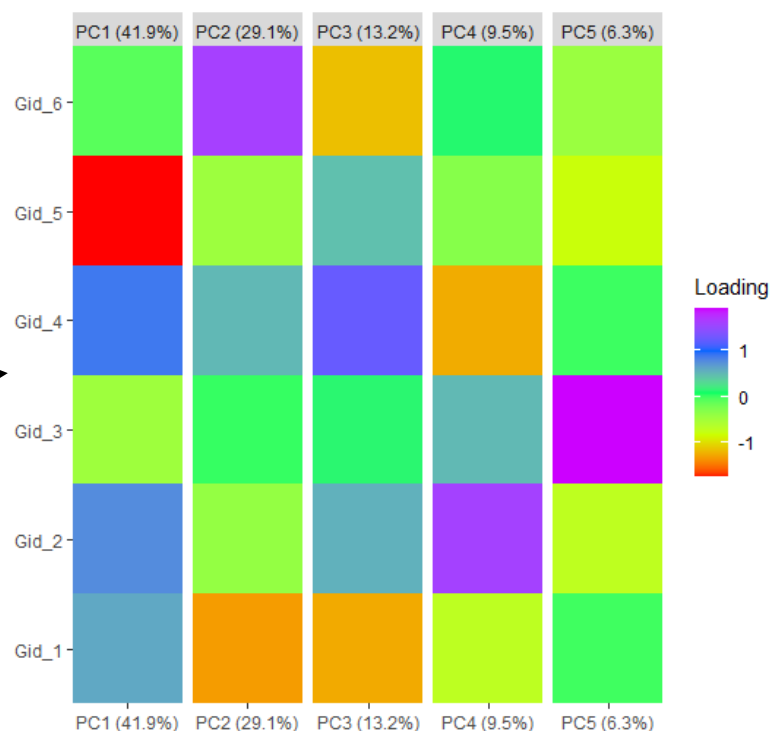
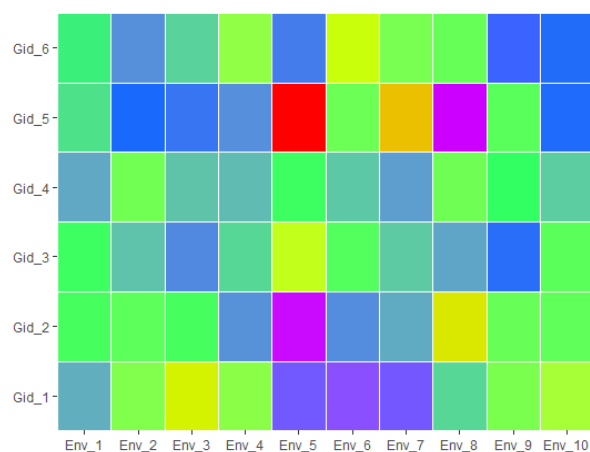
GGEbiplot (software): usa SVD

FGGE: usa análise de fatores (factor analysis)

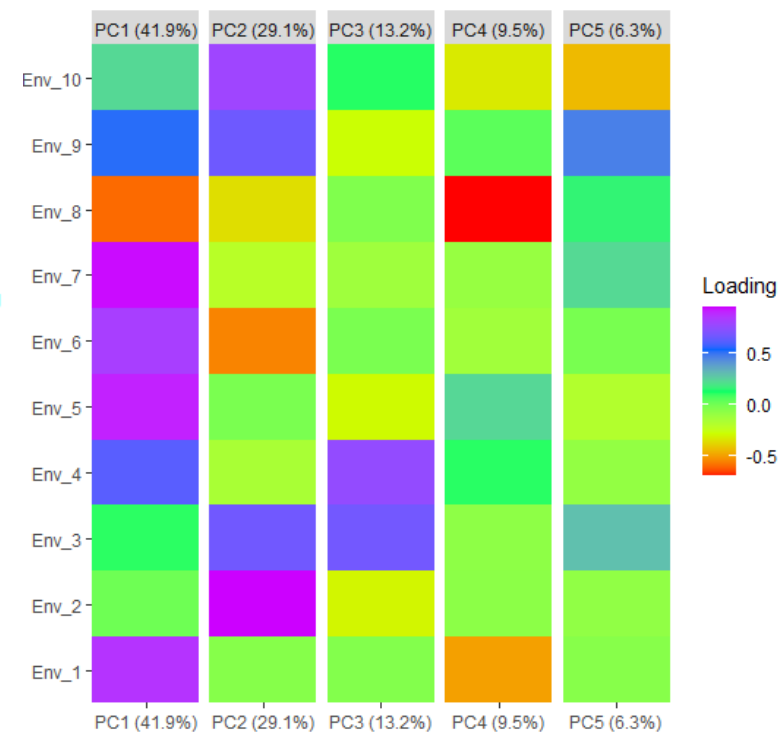
Passo a Passo:

- (1) isolar padrões genéticos e os padrões ambientais em *componentes principais*
- (2) Construir uma figura (visualização gráfica) desses padrões (os biplots!)
- (3) Selecionar genótipos com base na adaptação (ampla ou ambiente-específico)
- (4) Identificar grupos de ambientes que resultam num padrão de adaptação similar (isto é, mega-ambientes)

$$Y_{ij} = \mu + e_j + \sum \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij},$$



Pesos (vetores) de genótipo (linha)



Pesos (vetores) de ambiente (coluna)

Decomposição dos efeitos multiplicativos

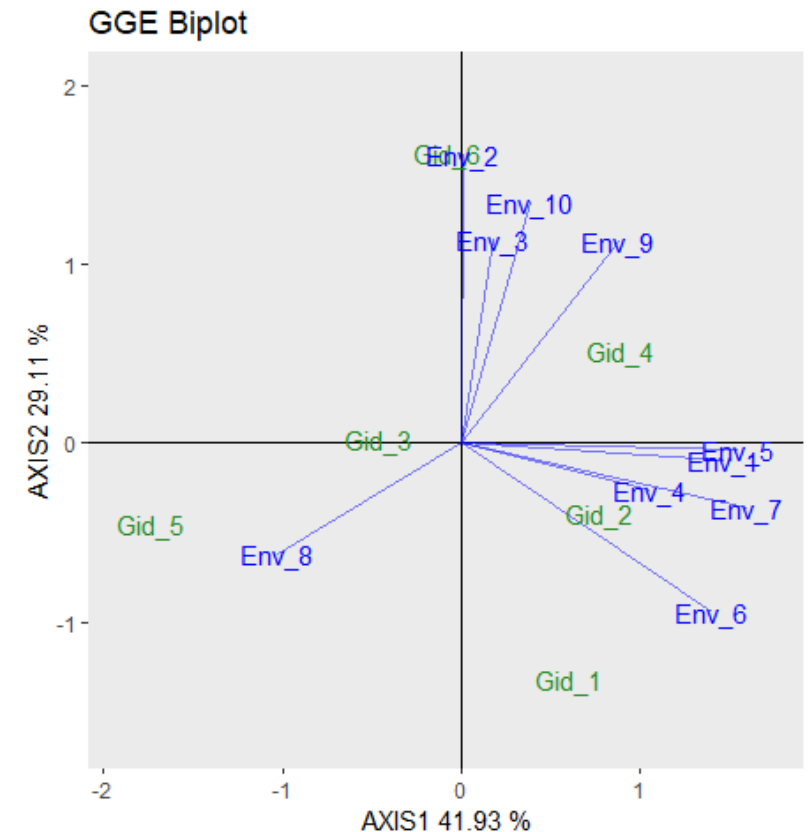
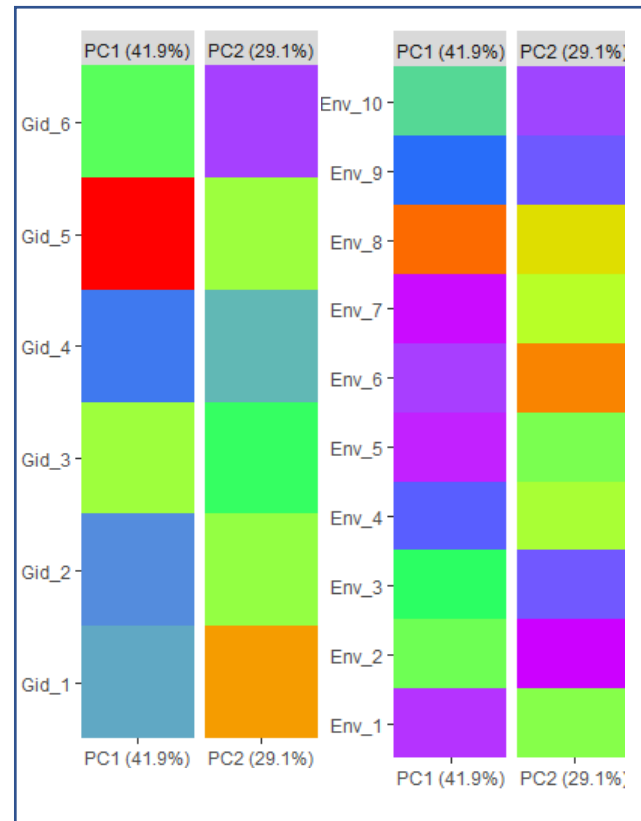
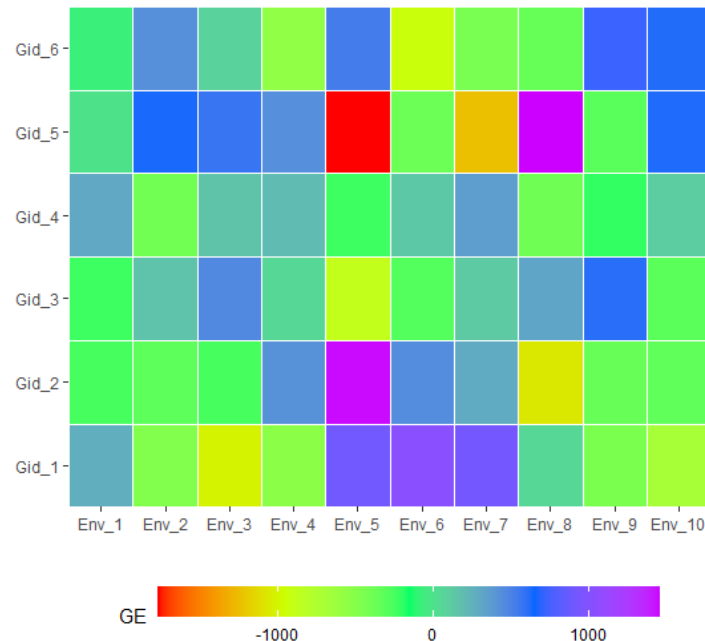
AMMI clássico: usa SVD (singular value decomposition), também é possível usar PCA (principal component analysis)

GGEbiplot (software): usa SVD

FGGE: usa análise de fatores (factor analysis)

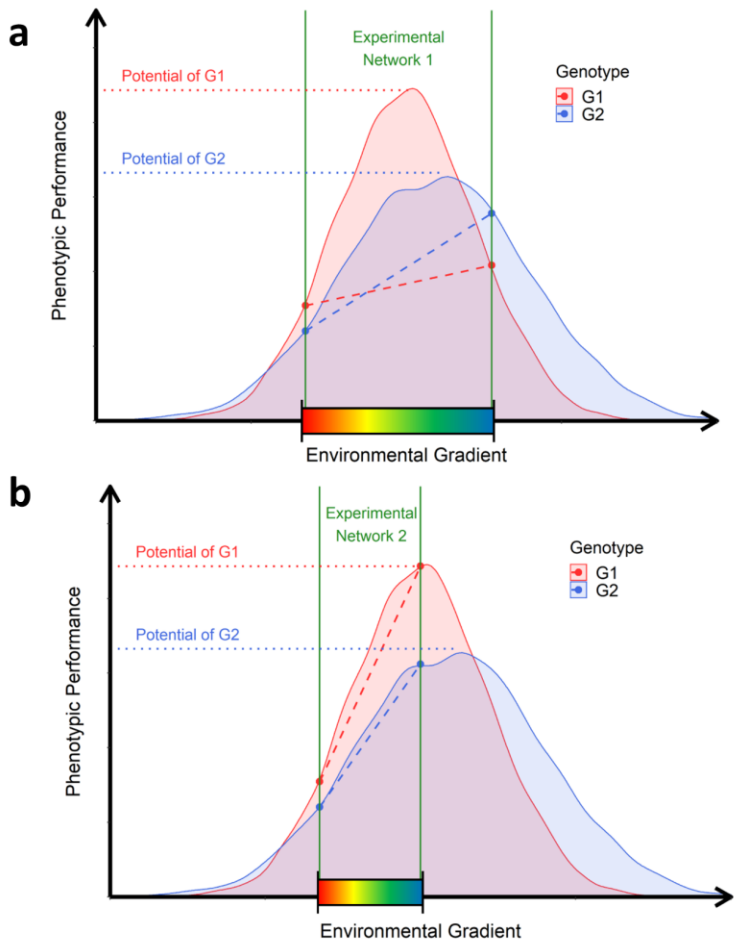
Passo a Passo:

- (1) isolar padrões genéticos e os padrões ambientais em *componentes principais*
- (2) Construir uma figura (visualização gráfica) desses padrões (os biplots!)
- (3) Selecionar genótipos com base na adaptação (ampla ou ambiente-específico)
- (4) Identificar grupos de ambientes que resultam num padrão de adaptação similar (isto é, mega-ambientes)



Adaptabilidade, Estabilidade, Plasticidade e Norma de Reação

- Conceitos que derivam da biologia (plasticidade), porém carregam conotação prática para o melhoramento de plantas
- São parâmetros emergentes de um dado germoplasma sob uma dada variação ambiental
- O melhoramento tem efeito no genoma indiretamente selecionando regiões que minimizam a variabilidade de GE (Gage et al. 2017)
- Biometricistas desenvolveram mais de 50 métodos desde a década de 50 (de 1929 e 1938, se considerar trabalhos do Sir Fisher)



Estabilidade: Definições Históricas

| Estabilidade | Autores | Conceito |
|-------------------------|--|---|
| Biológica | Becker (1981) | Performance constante do material em todos os ambientes. Homeostase de desenvolvimento |
| Estática | Allard & Bradshaw (1964) Waddington (1953) Lerner (1954), Gregg (1973) Becker & Leon (1988) | |
| Tipo 1 | Lin et al. (1986) | |
| Agrônômica | Becker (1981) | Capacidade genotípica em expressar um comportamento fenotípico previsível através dos ambientes |
| Dinâmica | Becker & Leon (1988) | |
| Comportamento | Mariot et al. (1976) Morais (1980) | |
| Tipo 2 | Lin et al. (1986) | |
| Fenotípica | Lewis (1954) | Mínima variância entre a expressão fenotípica do indivíduo e a média ambiental. |
| Comportamento Relativo | Coffman et al. (1976) | |
| Tipo 3 | Lin et al. (1986) Yue et al. (1997) | |
| Produção | Heinrich et al. (1963) | Capacidade genotípica em evitar flutuações na produção em vários ambientes |
| Plasticidade fenotípica | Bradshaw (1965) | |

Interpretação “empírica” (clássica)

Fundamenta-se no uso apenas de dados fenotípicos e procedimentos algébricos

Estabilidade (e instabilidade): Definições Históricas

| Estabilidade | Autores | Conceito |
|-------------------------|-----------------------------|---|
| Biológica | Becker (1981) | Performance constante do material em todos os ambientes. Homeostase de desenvolvimento |
| Estática | Allard & Bradshaw (1964) | |
| | Waddington (1953) | |
| | Lerner (1954), Gregg (1973) | |
| Tipo 1 | Becker & Leon (1988) | Capacidade genotípica em expressar um comportamento fenotípico previsível através dos ambientes |
| | Lin et al. (1986) | |
| | Becker (1981) | |
| Agronômica | Becker & Leon (1988) | |
| Dinâmica | Mariot et al. (1976) | Capacidade genotípica em evitar flutuações na produção em vários ambientes |
| | Morais (1980) | |
| | Lin et al. (1986) | |
| Comportamento | Lin et al. (1986) | |
| Tipo 2 | Lewis (1954) | Mínima variância entre a expressão fenotípica do indivíduo e a média ambiental. |
| | Coffman et al. (1976) | |
| | Lin et al. (1986) | |
| Fenotípica | Yue et al. (1997) | |
| Comportamento Relativo | Heinrich et al. (1963) | Capacidade genotípica em evitar flutuações na produção em vários ambientes |
| | Bradshaw (1965) | |
| | | |
| Tipo 3 | | |
| Produção | | |
| Plasticidade fenotípica | | |

Interrelação com a origem biométrica do fenômeno

| Causa (efeito predominante) | Decisão |
|---------------------------------|---|
| $Ad_{ij} = G_i$ | Ampla adaptação dos genótipos ou ambientes muito similares |
| $Ad_{ij} = E_j$ | Ambientes muito heterogêneos e falta de variação genética |
| $Ad_{ij} = GE_{ij}$ | Interação GE complexa, baixa correlação genética entre ambientes |
| $Ad_{ij} = G_i + E_j$ | Interação GE NULA, ou seja, no mesmo mega-ambiente |
| $Ad_{ij} = G_i + GE_{ij}$ | Capitalização da interação GE (complexa ou simples) sob os efeitos genéticos, isto é: $Ad = G + GE > G$ |
| $Ad_{ij} = E_j + GE_{ij}$ | Baixa variabilidade genética (de efeito amplo), porém alta responsividade para ambientes específicos e contrastantes, isto é, $Ad = E + GE > G$ |
| $Ad_{ij} = G_i + E_j + GE_{ij}$ | Modelo completo, adaptação ambiente-específico ou ampla |

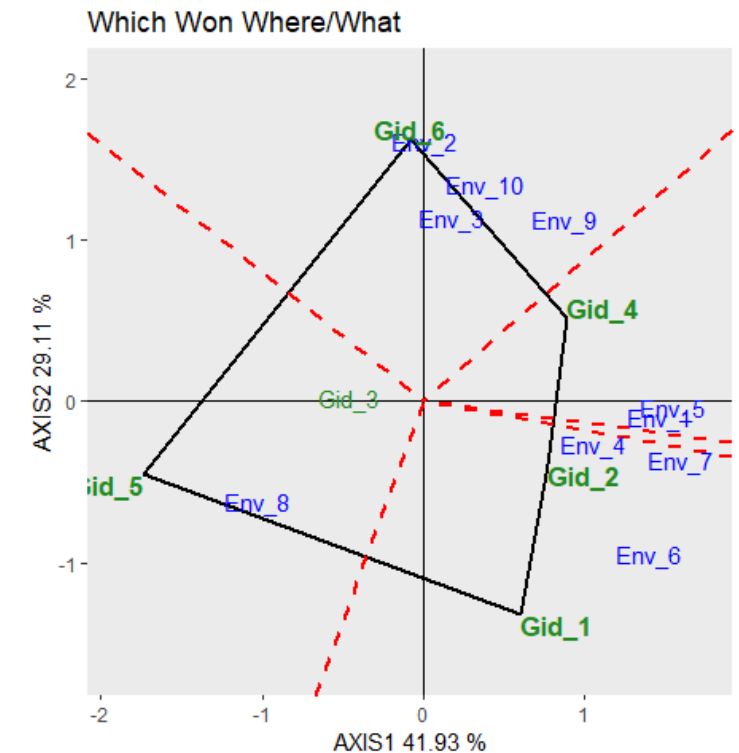
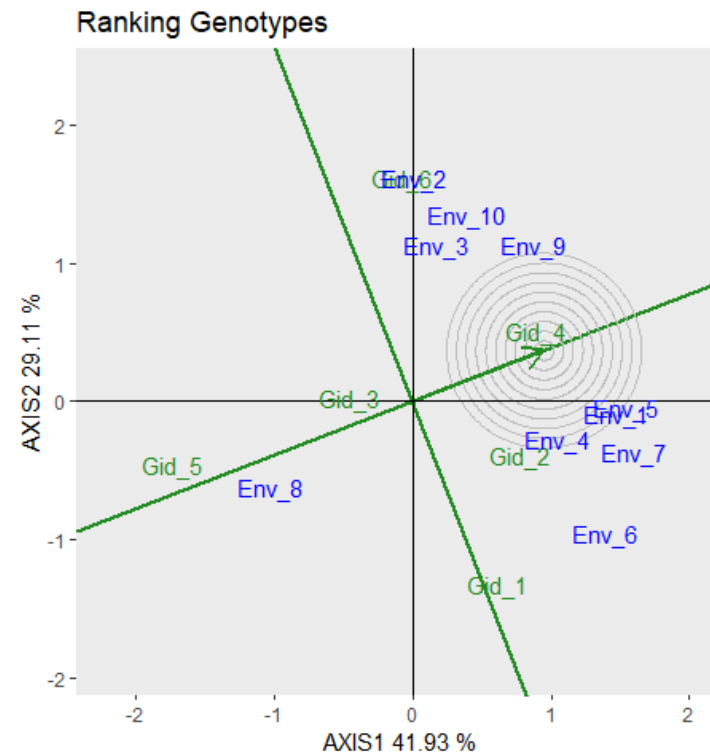
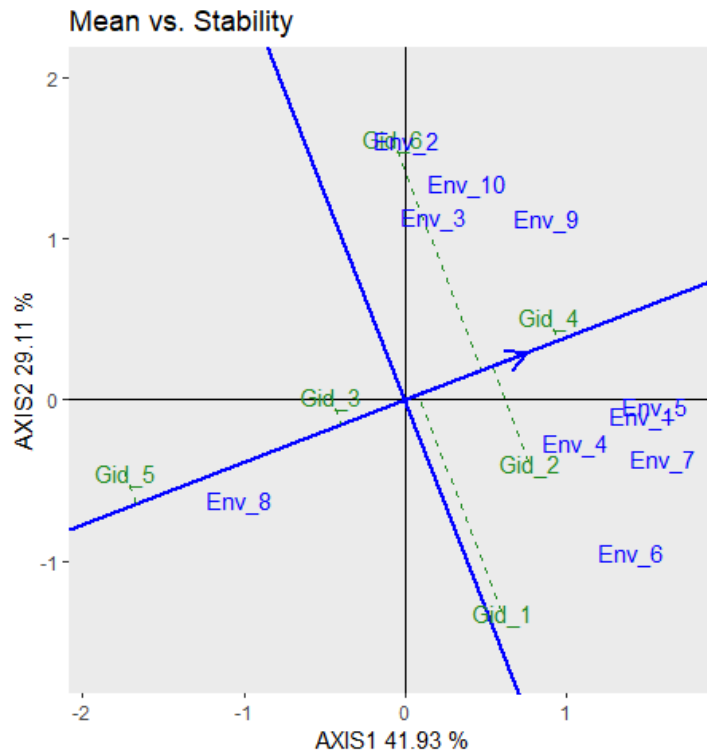
Visualização gráfica do GGEbiplot

Vejam: [Material GGEbiplot](#)

“Perfumes” e “cerejas do bolo” para facilitar a interpretação (e tomada de decisão)!

Gosta de gráficos? E Planos cartesianos? Então:

- (1) Repare nas posições dos genótipos
- (2) Repare nas posições dos ambientes
- (3) Repare nos quadrantes
- (4) Retas são geralmente “padrões médios”, isto é, tendências da população na rede experimental
- (5) Repare na variância explicada (variância dos componentes, isto é, variância dos eixos x + y)



Diagnóstico de Adaptabilidade, Estabilidade, Estratificação Ambiental, Genótipos similares Ou qualquer outra dupla-entrada que usar!

Regressão sob um “índice ambiental”

Índice que representa a qualidade de um ambiente

Em termos práticos, a média de um dado trait (centrada na média geral, isto é: $h_j = E$)

Finlay & Wilkinson (1963), Eberhart & Russel (1966), Perkins and Jinks (1968)

An index independent of the experimental varieties and obtained from environmental factors such as rainfall, temperature, and soil fertility would be desirable. Our present knowledge of the relationship of these factors and yield does not permit the computation of such an index. Until we can measure such factors in order to formulate a mathematical relation with yield, the average yield of the varieties in a particular environment must suffice.

$$\begin{aligned}y_{ij} &= \mu + G_i + E_j + GE_{ij} \\y_{ij} &= \mu + G_i + E_j + GE_{ij} \\y_{ij} &= \mu + G_i + E_j(1 + G_i) \\y_{ij} &= \mu + G_i + (1 + G_i)E_j\end{aligned}$$

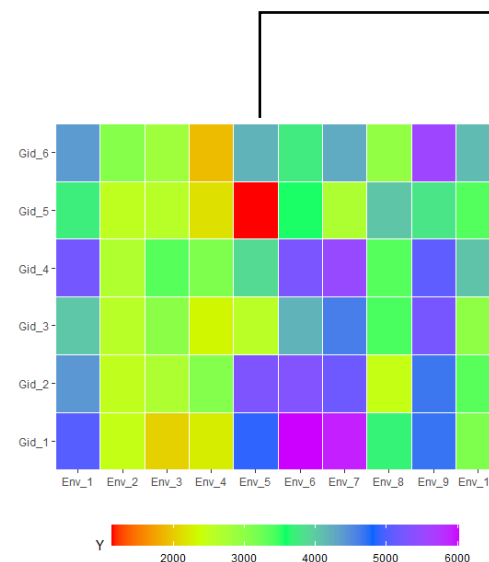
Lembrando que:

$$\bar{y}_{i.} = \mu + G_i$$

$$\bar{y}_{.j} = \mu + E_j$$

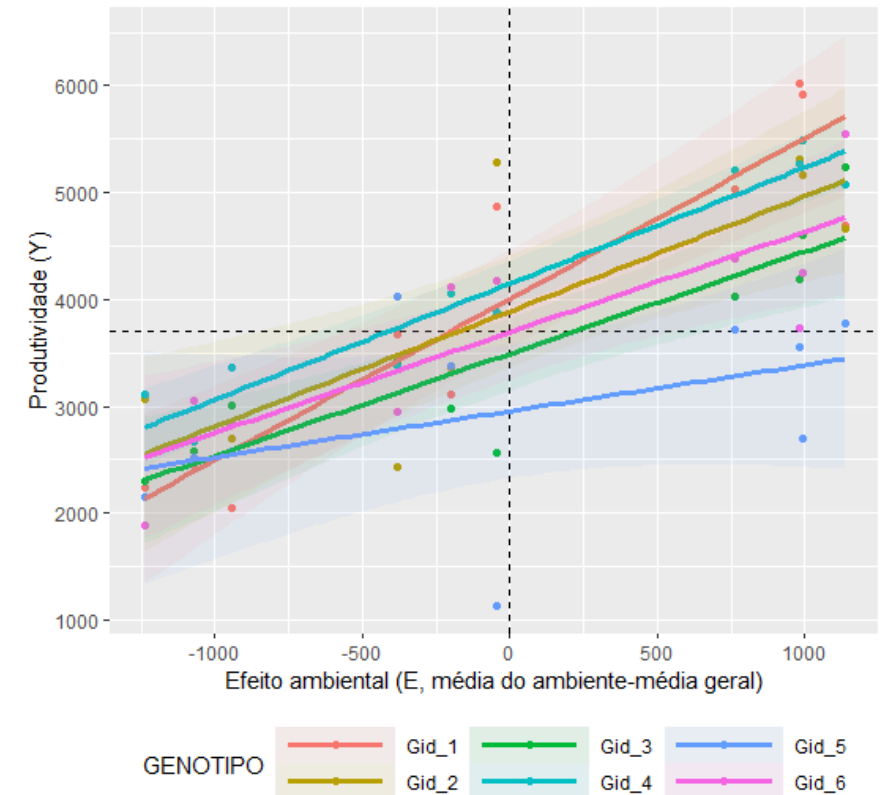
$$y_{ij} = \mu + G_i + b_i E_j$$

$$\hat{y}_{ij} = \bar{y}_{i.} + b_i E_j + d_{ij}$$



Ângulo (b_i)

(Slope, coeficiente de regressão, b)
é a responsividade linear



Desvio (d_{ij})

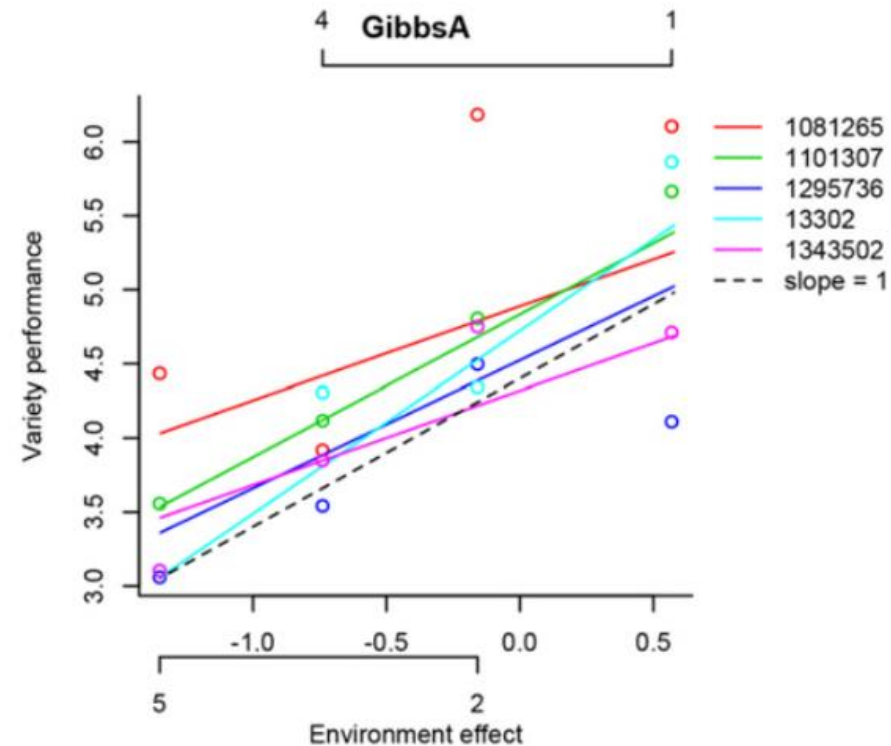
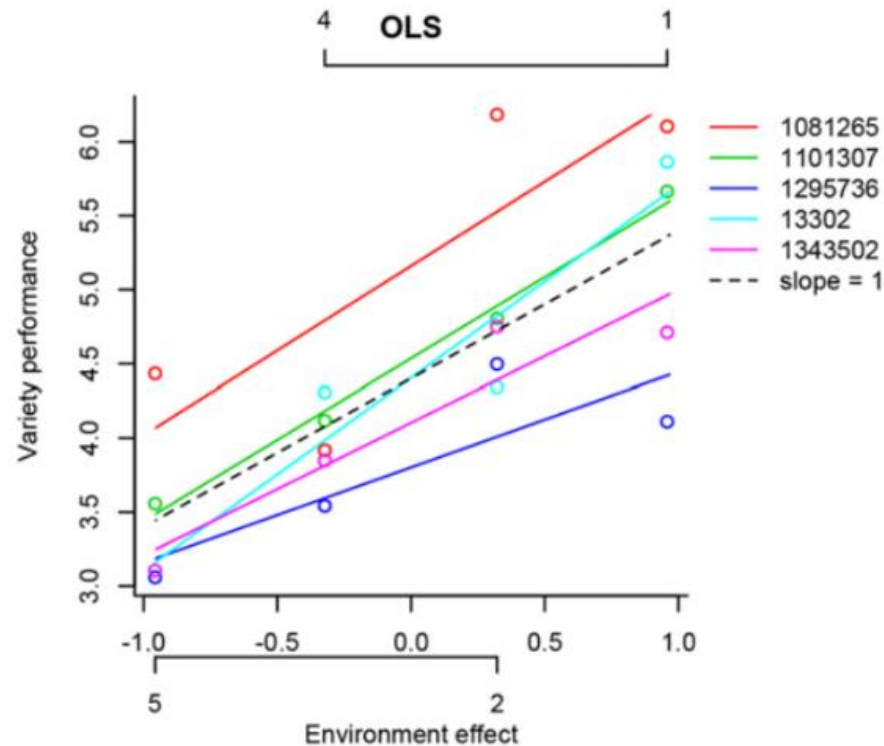
(em relação ao padrão linear)
é estabilidade (nuvem)

Regressão sob um “índice ambiental”

Amostrador de Gibbs (Bayesiano) versus Quadrados mínimos ordiários (OLS)

[Lian and de los Campos \(2016\)](#)

- OLS: efeitos fixos, regressão linear de cada genótipo sobre a média (centrada) ambiental
- Bayesiano: amostragem de Gibbs para inferencia dos parametros do modelo (g , b e h)
- Será que usar uma matrix de covariância **A** capta melhor a interação GE complexa (cruzada)?



Interpretação Biológica (GE ou G+GE)

Algoritmo

1. Ajustamos as médias de cada genótipo em cada ambiente (análise conjunta)
2. Para cada genótipo, ajustamos suas médias ao longo dos ambientes usando o modelo FR e dados ambientais
3. Temos a soma de quadrados original (sem decomposição). Comparamos com a capturada pelo modelo.
4. Para cada genótipo teremos um coeficiente de responsividade a cada fator ambiental. E o quanto que este fator explicou para o dado genótipo.
5. A soma das variancias para cada genótipo resulta na soma total explicada pelo modelo
6. Opcional: bootstrap/leave-one-out

OBS: (1) é feito com [Adij](#), lembrando que este depende do seu germoplasma e da sua rede experimental

| Specie | Region | Author | Env. Information | %SS |
|--------------|------------|-------------------------------|--------------------------------|-------|
| Potato | Germany | Baril et al. (1995) | DFLo, Tmax, Tmed | 44,40 |
| Ryegrass | France | Balfourier et al. (1997) | Tmin | 72,80 |
| Maize | USA | Magari et al. (1997) | GDv, GDf, Pracm | 45,80 |
| Maize | France | Epinat Le Signor et al (2001) | Tmax, Tavg, GDv, GDf | 60,00 |
| Winter Wheat | Spain | Voltas et al. (2005) | EHv, GDv, Tmaxf | 71,10 |
| Winter Wheat | Spain | Voltas et al. (2005) | GDv, Tminv | 58,70 |
| Tomato | Caribbean | Ortiz et al. (2007) | Prac, DFLo, Tavg, pH | 91,16 |
| Wheat | India | Joshi et al. (2010) | Urf, Tmaxv, Tminv,Ur, SoilChar | 60,26 |
| Wheat | Mexico | Verhulst et al. (2011) | Tminf, Racf | 90,2 |
| Melon | Brazil | Nunes et al. (2011) | Tmin, Tmax, Prac | 79,89 |
| Sugarcane | Brazil | Sá (2013) | Tavg, Prac | 55,56 |
| Soybean | Brazil | Cardoso Junior (2014) | Pracf, Etacm, LAT | 41,00 |
| Upland Rice | Brazil | Costa-Neto et al. (2020) | LAT, LON, ELE | 53-59 |
| Barley | Austrailia | Porker et al. (2020) | Tavg, Photoperiod | 69,00 |

Regressão Fatorial (GE ou G+GE)

Estudo de caso em arroz de terras altas

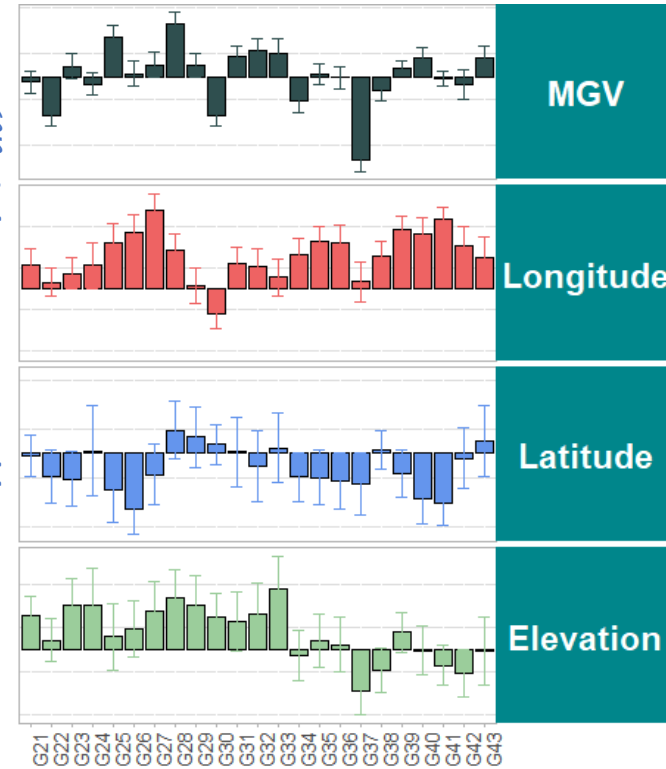
[Costa-Neto et al \(2020\)](#)

Passo 1: FR em Ad = G+GE

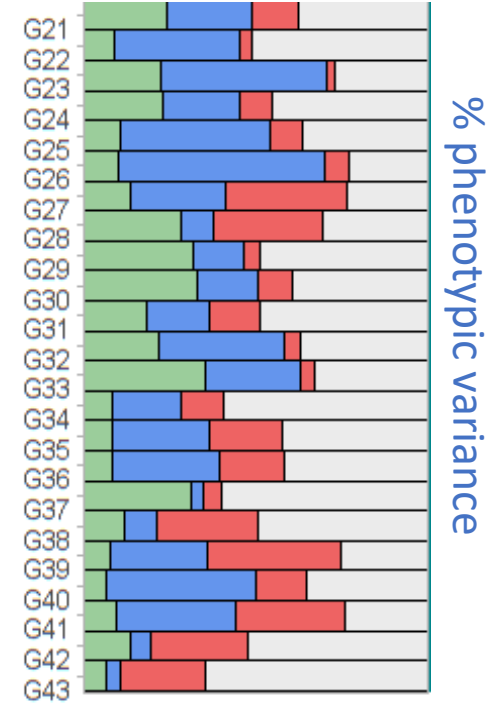
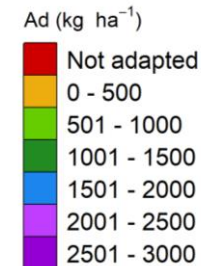
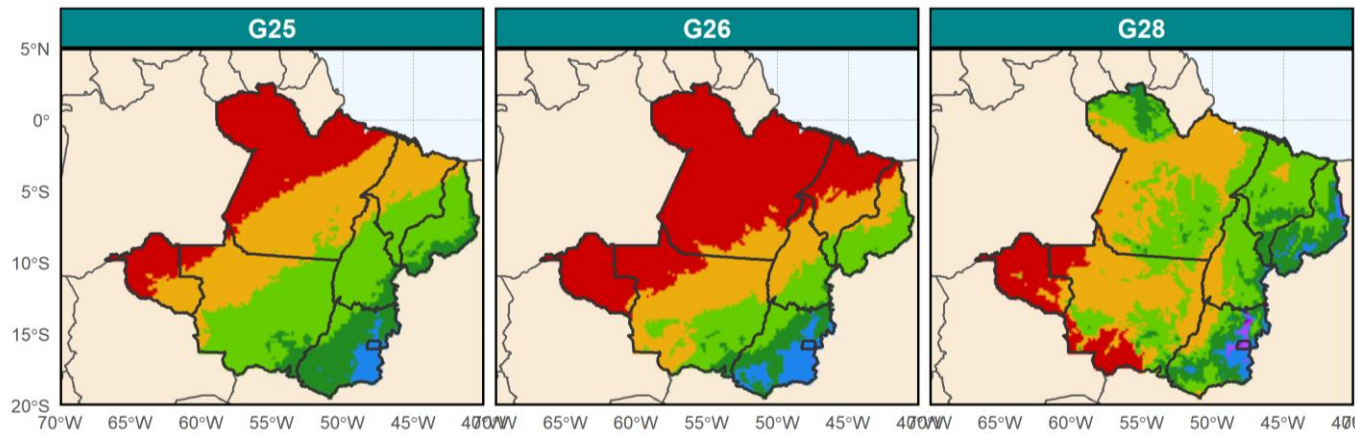
$$y_{ij} - \bar{y}_{.j} = G_i + \sum_{k=1}^v b_k z_{jk} + u_{ij} (\text{noise})$$

$$Ad_{ij} = G_i + \sum_{k=1}^v b_{ik} z_{jk}$$

Genotypic Sensitivity (b_{ik})



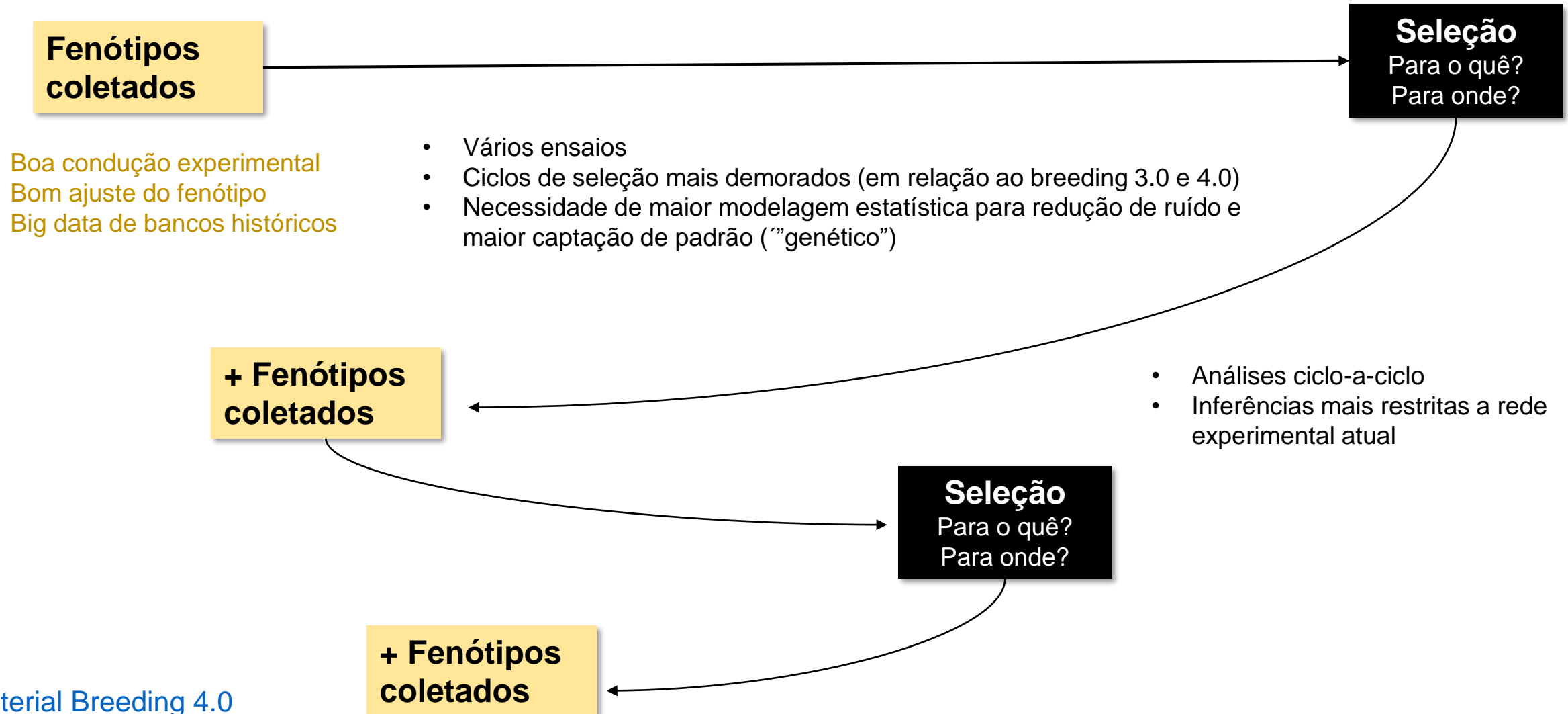
Passo 2: Interpolação Espacial



Predição de Fenótipos

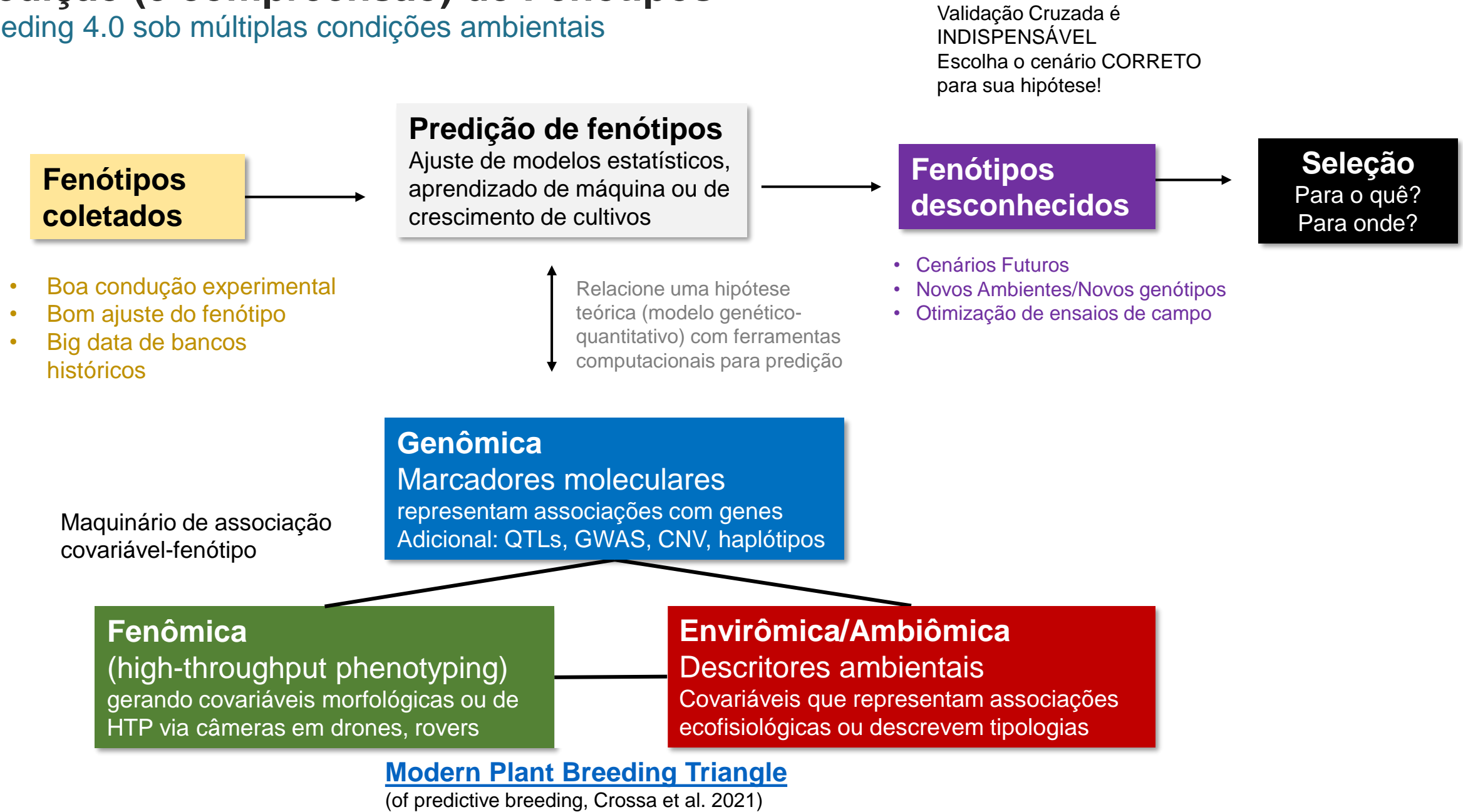
sob múltiplas condições ambientais

Breeding 2.0



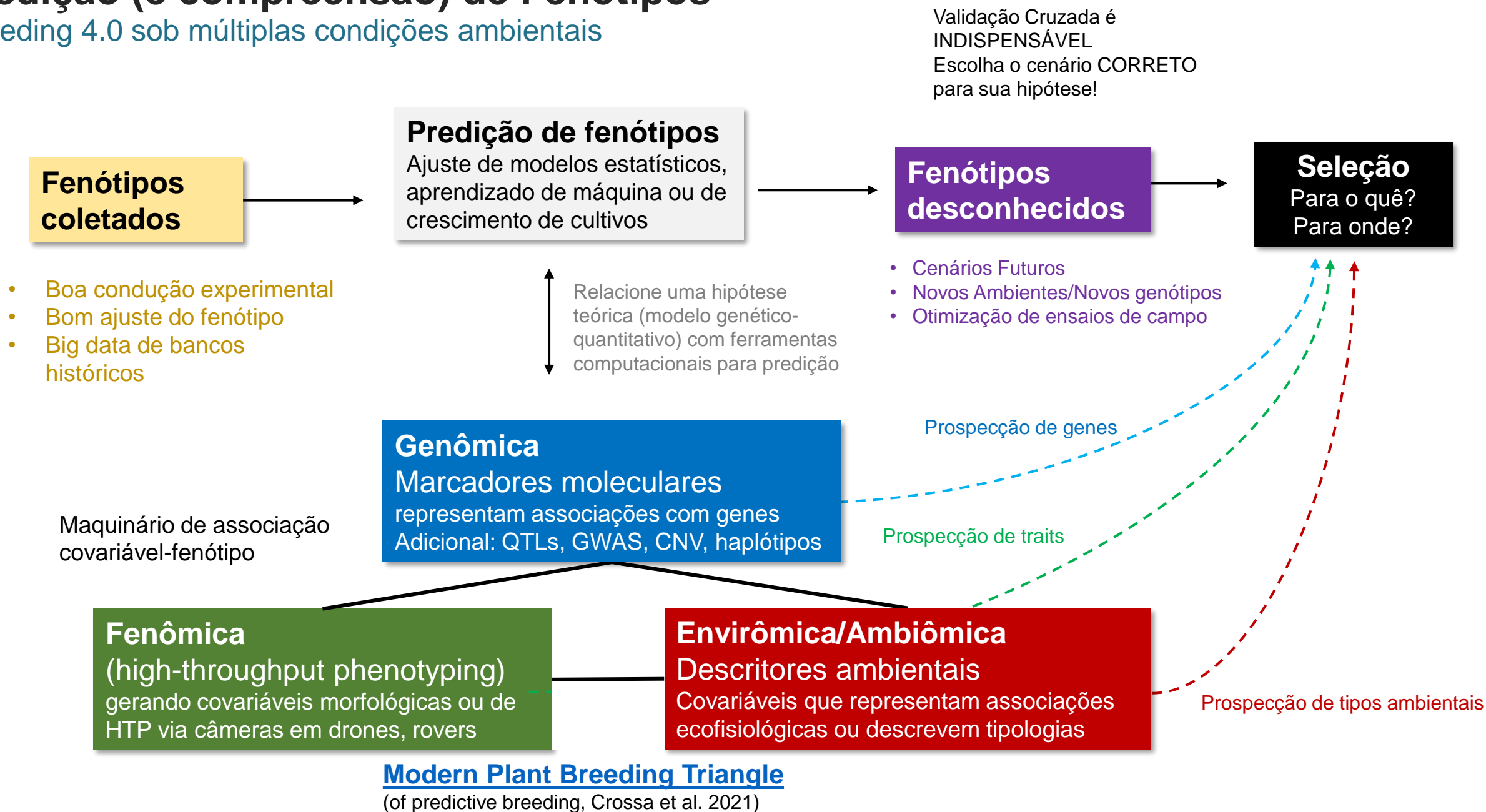
Predição (e compreensão) de Fenótipos

Breeding 4.0 sob múltiplas condições ambientais



Predição (e compreensão) de Fenótipos

Breeding 4.0 sob múltiplas condições ambientais



Tendências & Conclusões pessoais

- Migrar totalmente da “empírica” para “analítica/biológica” em algumas fases do programa de melhoramento;
- Trabalhar com dados históricos (VCU, striptests) de pelo menos 5 anos;
- Explorar GE via outras ômicas (e.g., transcriptômica, proteômica);
- Predição genômica para multi-ambientes sob cenários futuros e otimização de ensaios de campo;
- Em muitos cenários, o uso de envirotyping será cada vez mais frequente;
- Machine Learning & Deep Learning conectando “camadas” envirotyping, genoma e fenótipo;
- CNN, Random Forest (Resende et al. 2020), Support Vector Machine, Gradient Boost Machine (GBM);
- Modelos de crescimento de cultivo (crop models) sendo enriquecidos por genômica (e outras ômicas);
- Genômica de Associação (GWAS) para fenótipos vinculados a norma de reação e plasticidade;
- GWAS com fenótipos “putativos” de crop models;
- FR para compreensão da GE; kernels de covariância ambiental para predição da GE
- GGEbiplot não é modelo, é software. SREG, AMMI, GREG, COMM são modelos “bilineares”
- Matrizes de dupla entrada (e.g., GxE) podem ser decompostas de diversas formas para diversos propósitos
- Em muitos casos, modelos mistos com Factor Analytic (FA) podem ser a saída mais simples e prática;
- Tenha coerência e entenda seu germoplasma para escolher o melhor método;
- Não existe “competição de modelos”
- Se entender os conceitos poderá analisar os seus dados DO SEU JEITO, sem estar preso a métodos “em caixinhas”
- **Use a ferramenta GE que te deixe mais confortável e que te permita tirar as melhores conclusões possíveis para resolver o seu problema**