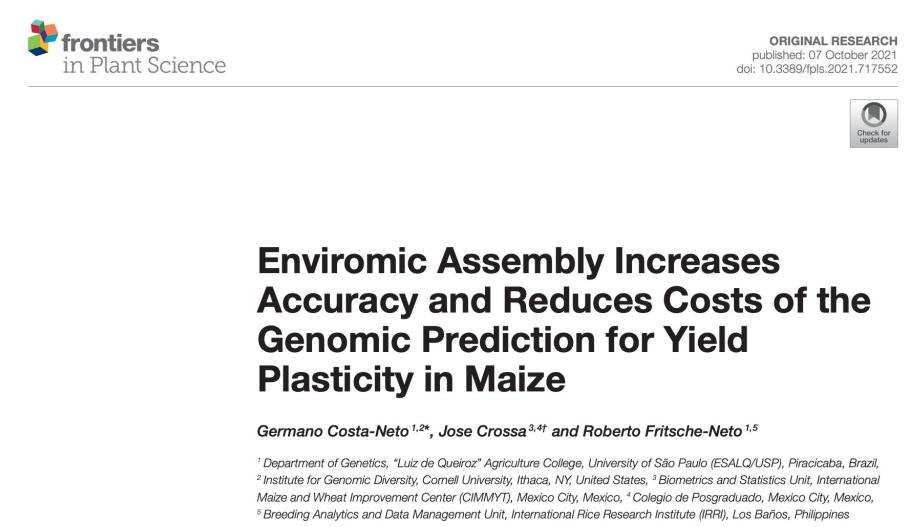


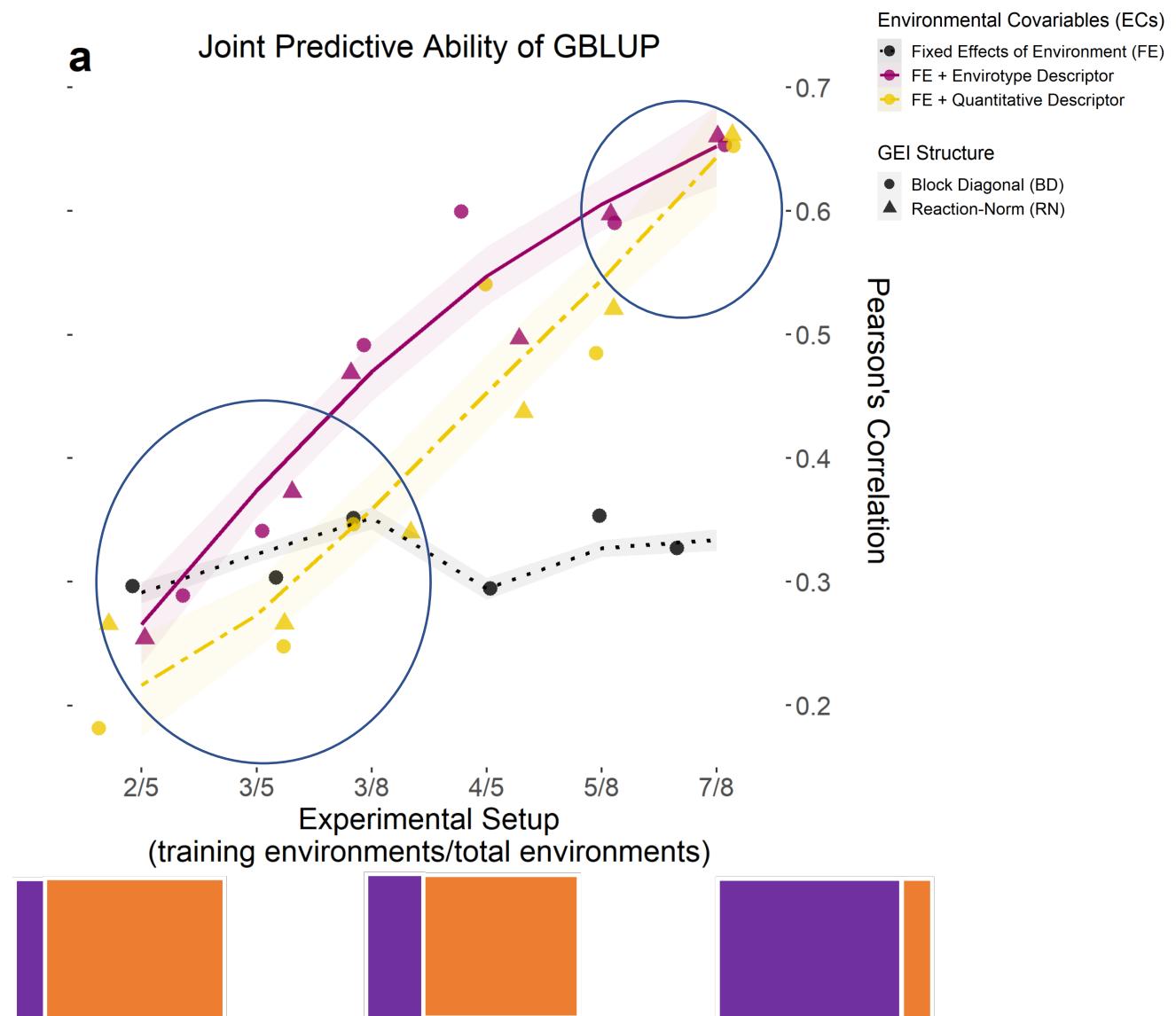
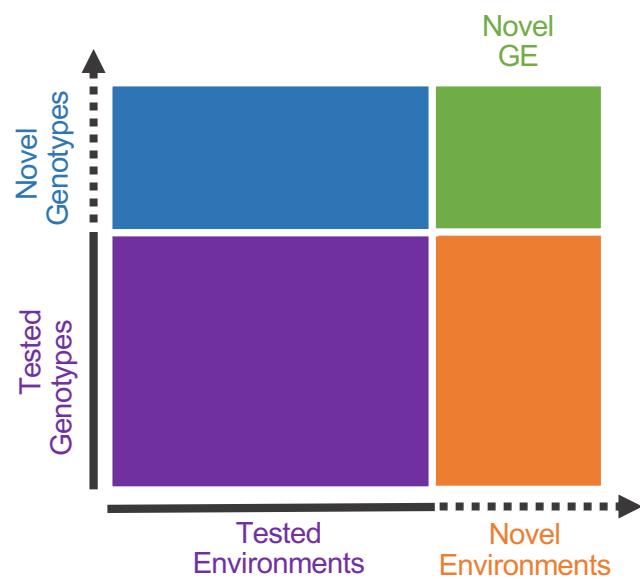
# “Enviromic-aided” Genomic Prediction (E-GP): study case



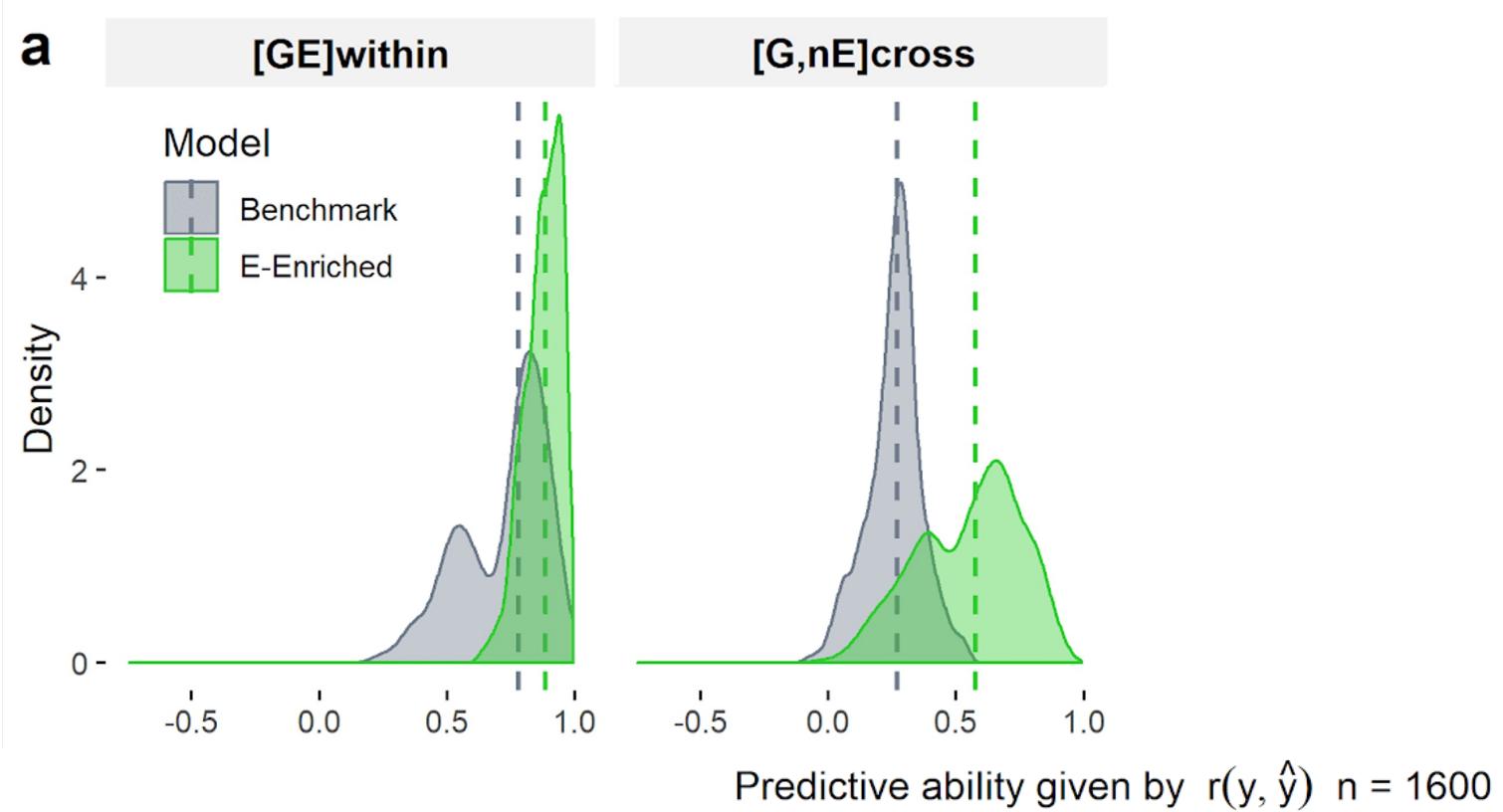
## Questions:

- T-matrix is suitable for prediction?
- T-matrix (frequencies) vs W-matrix (conventional way)
- Kernel-based Reaction-norms really matter?
- Predict new GxE ?

# Good: T-matrix increases GxE prediction accuracy over G+GE model and W-matrix



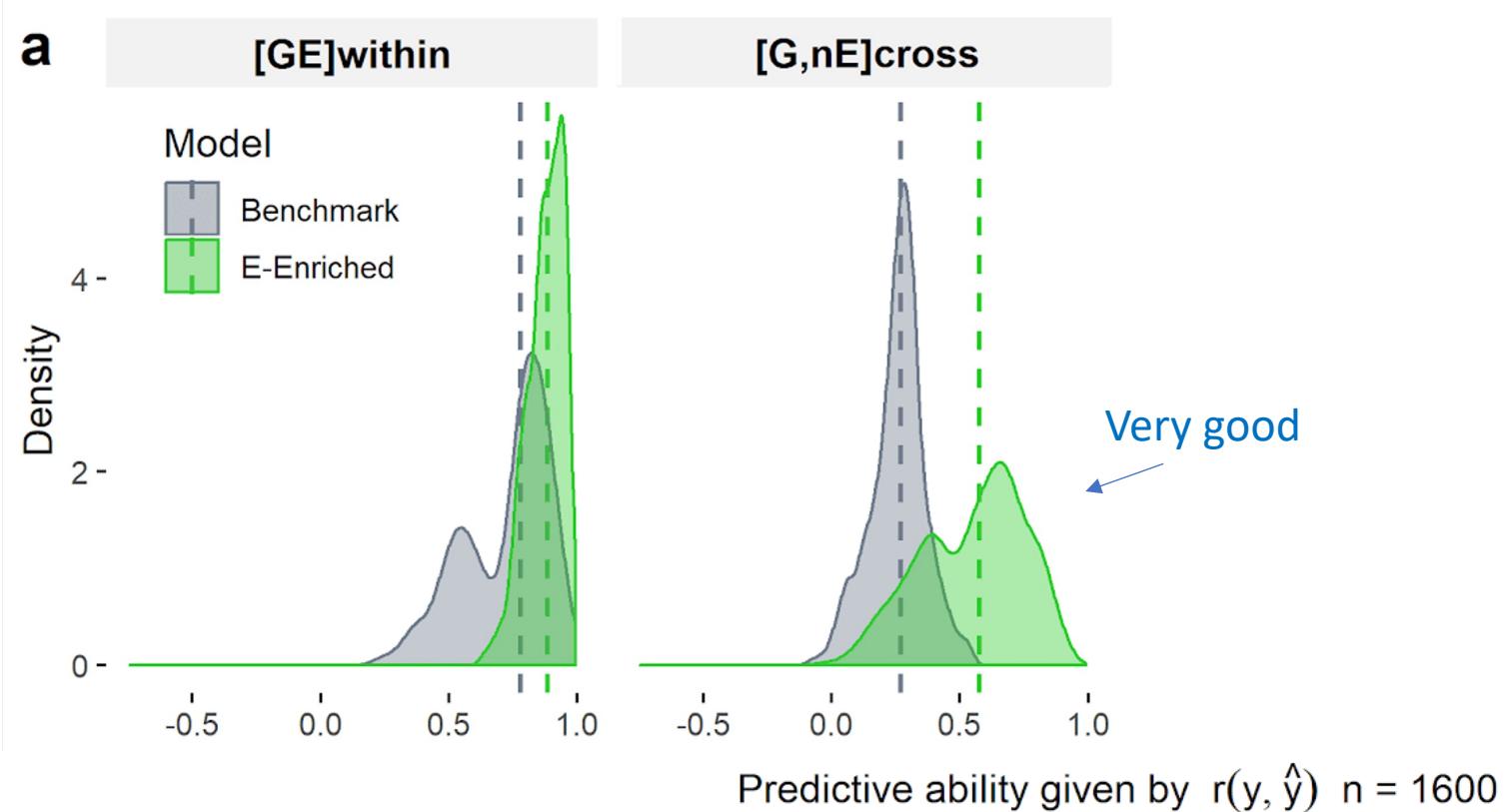
# “No difference between benchmark GBLUP and E-GP...right?”



Split the data set in

- [GE] within (training set)
- [G,nE] know genotypes, new environments (CV0)
- [nG,E] new genotypes, known environments (CV1)
- [nG,nE] entirely new GxE

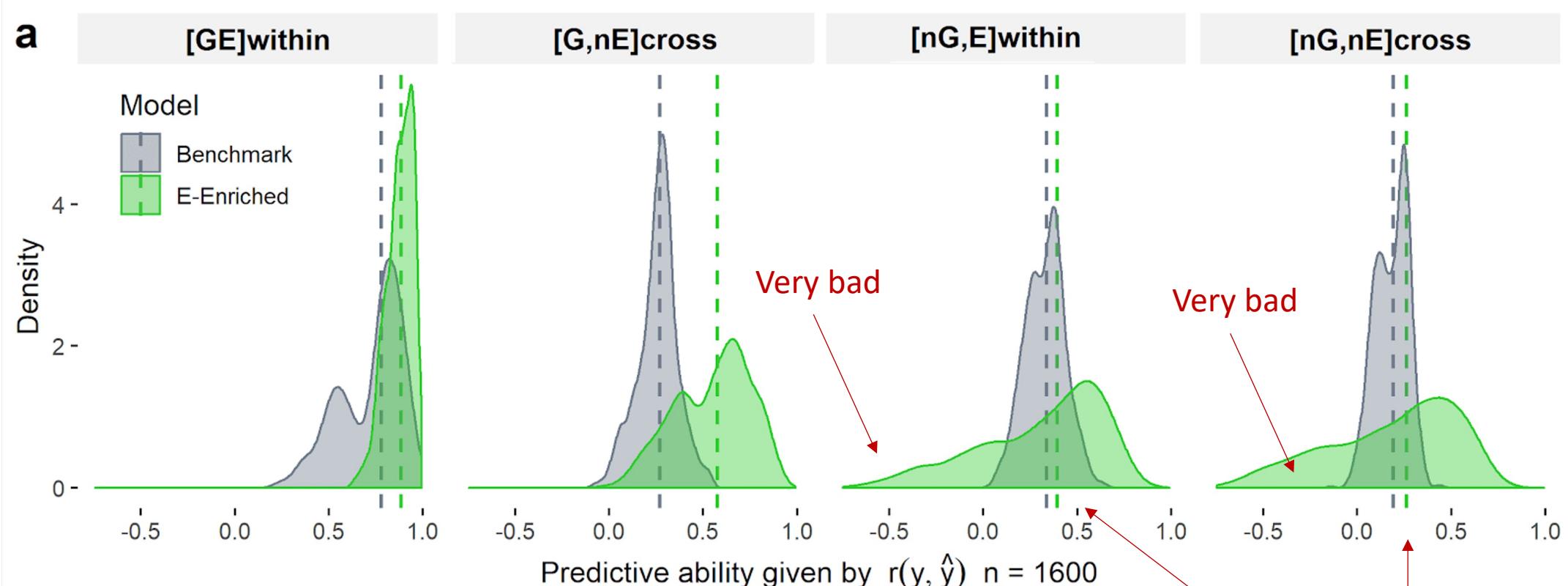
# “No difference between benchmark GBLUP and E-GP...right?”



Split the data set in

- [GE] within (training set)
- [G,nE] know genotypes, new environments (CV0)
- [nG,E] new genotypes, known environments (CV1)
- [nG,nE] entirely new GxE

# “No difference between benchmark GBLUP and E-GP...right?”

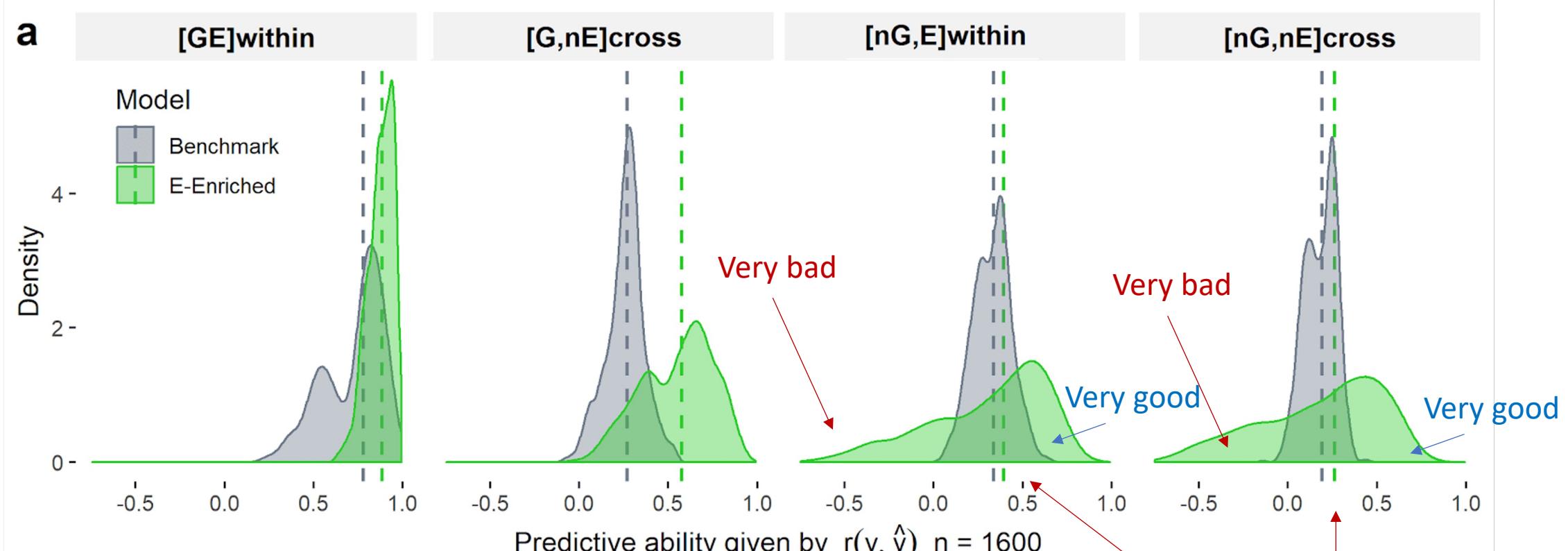


Split the data set in

- [GE] within (training set)
- [G,nE] know genotypes, new environments (CV0)
- [nG,E] new genotypes, known environments (CV1)
- [nG,nE] entirely new GxE

Almost no difference  
in the median

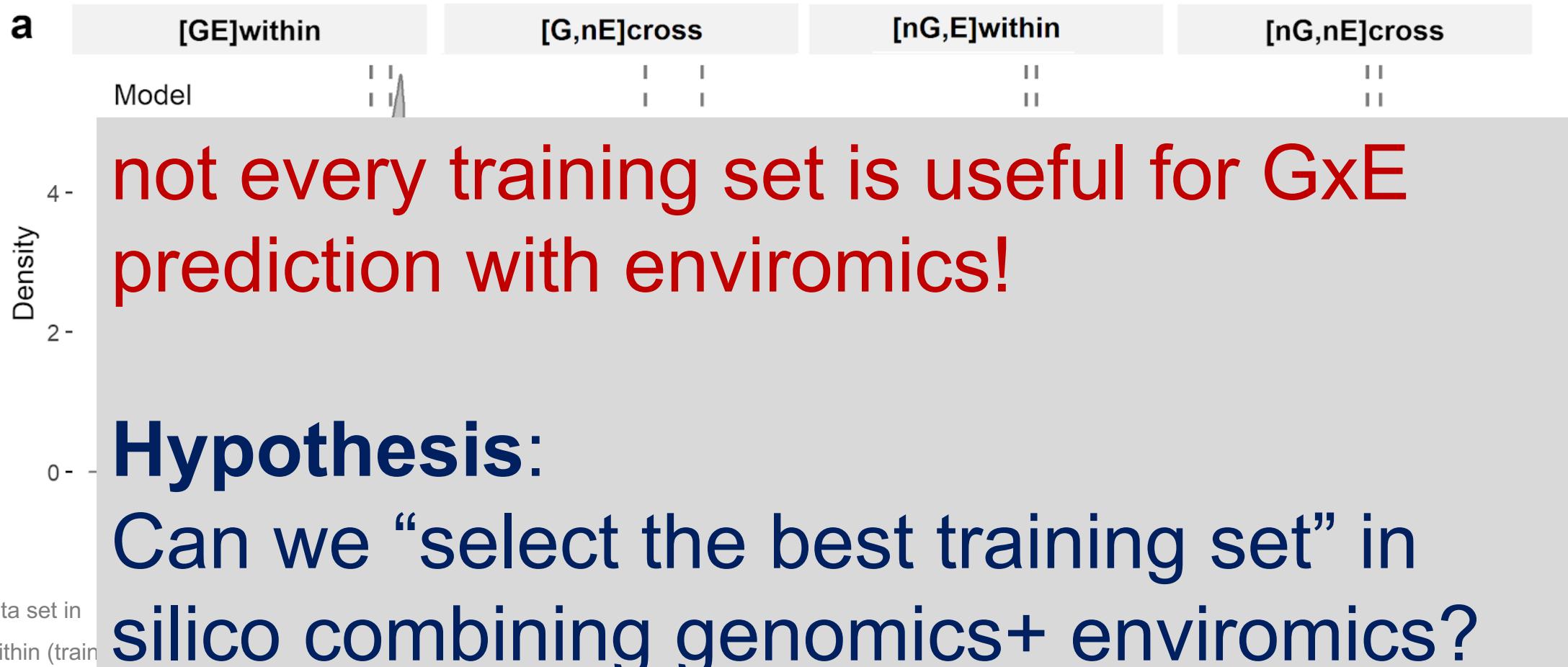
# “No difference between benchmark GBLUP and E-GP...right?”

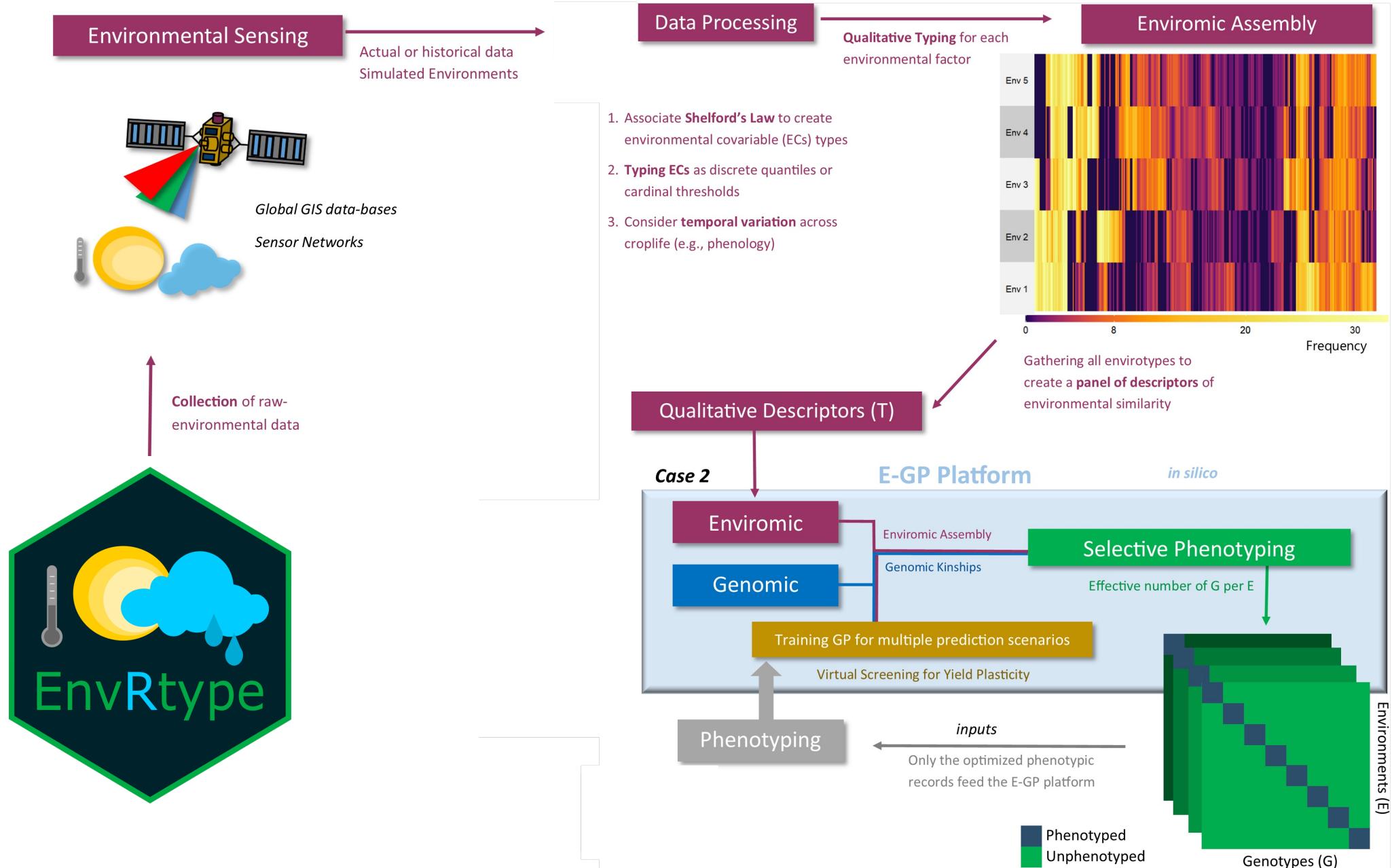


Split the data set in

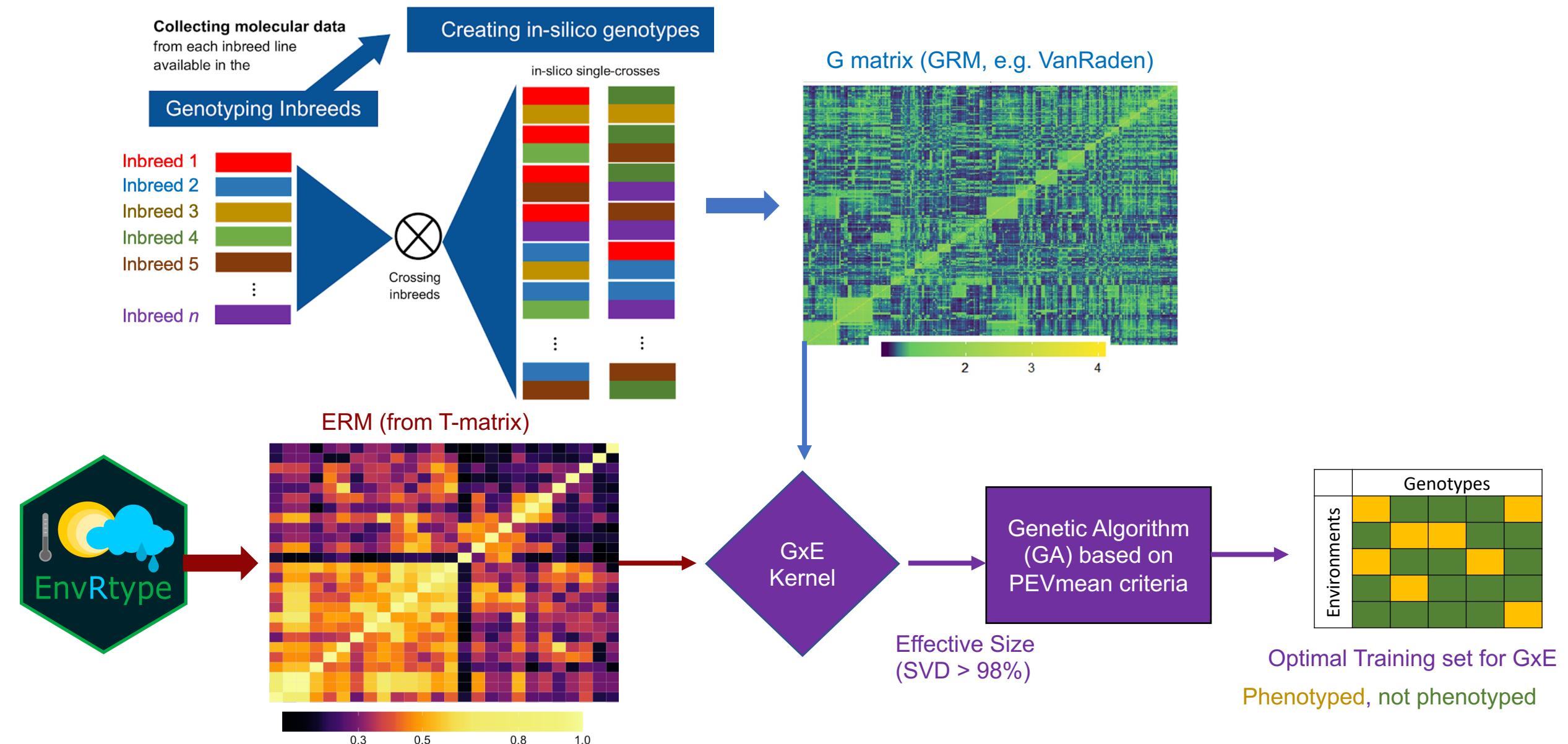
- [GE] within (training set)
- [G,nE] know genotypes, new environments (CV0)
- [nG,E] new genotypes, known environments (CV1)
- [nG,nE] entirely new GxE

# “No difference between benchmark GBLUP and E-GP...right?”



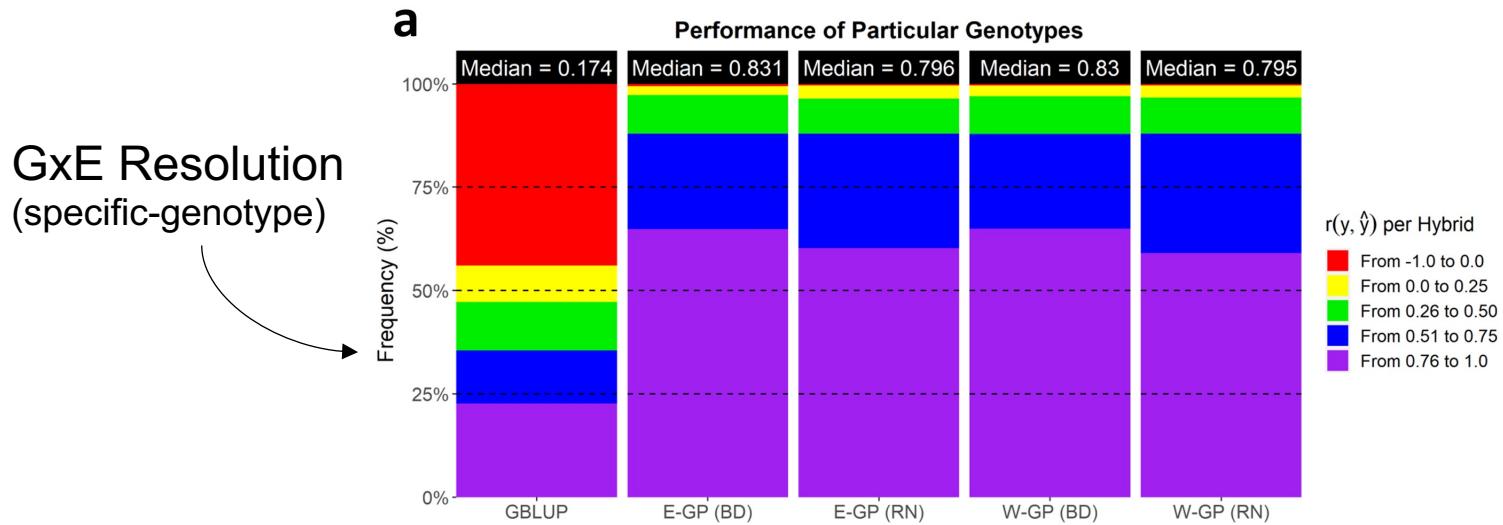


# Optimizing training sets for GP



## Good: E-GP using T-matrix allows an early screening for plasticity

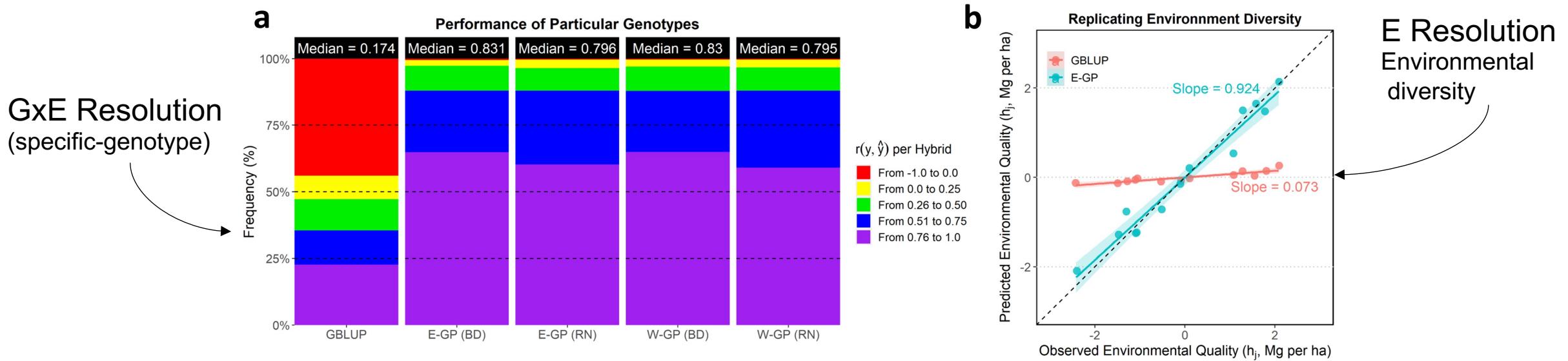
After optimization (some few genotypes, some environments)  $N < 10\%$  MET size



Correlation between  
observed x predict values  
for each environment

# Good: E-GP using T-matrix allows an early screening for plasticity

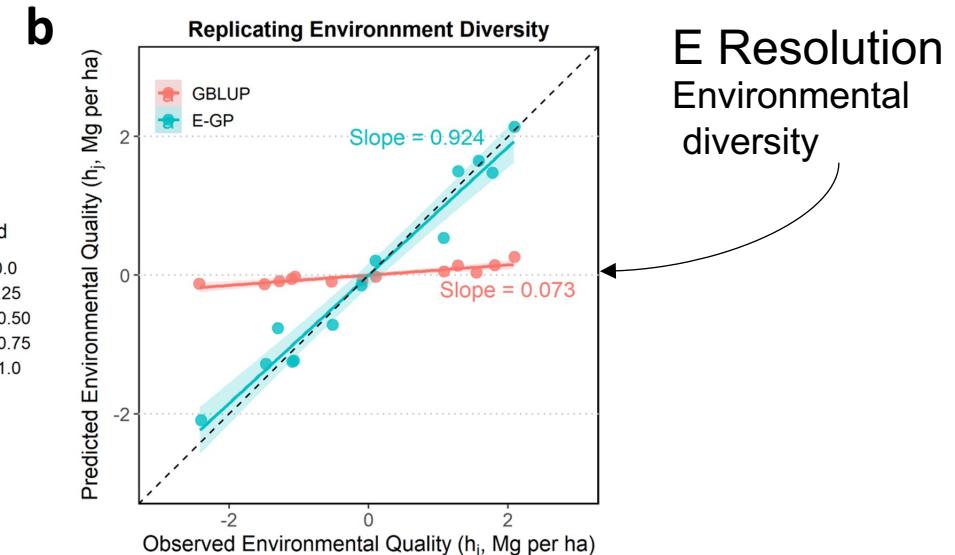
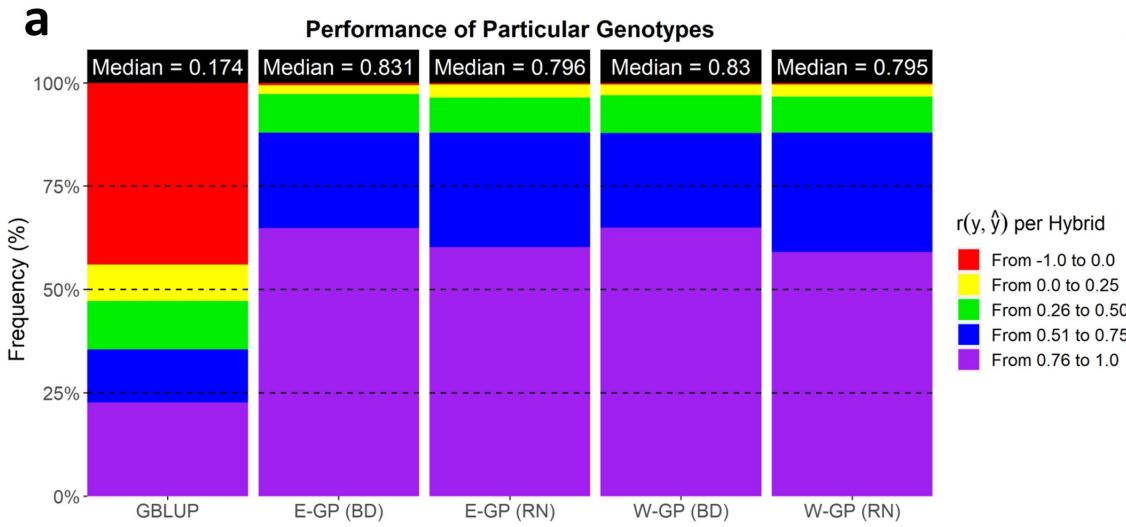
After optimization (some few genotypes, some environments)  $N < 10\%$  MET size



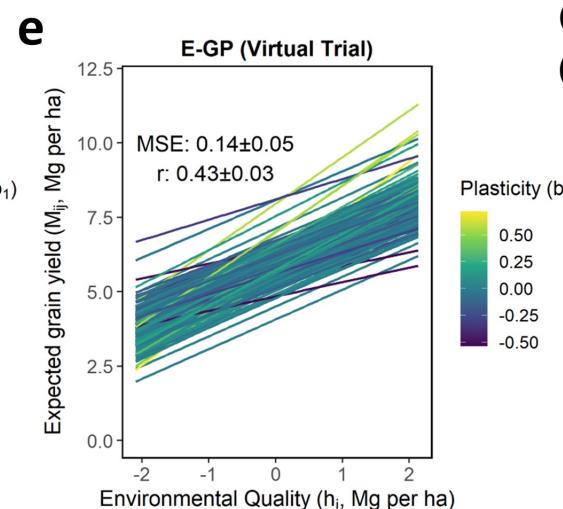
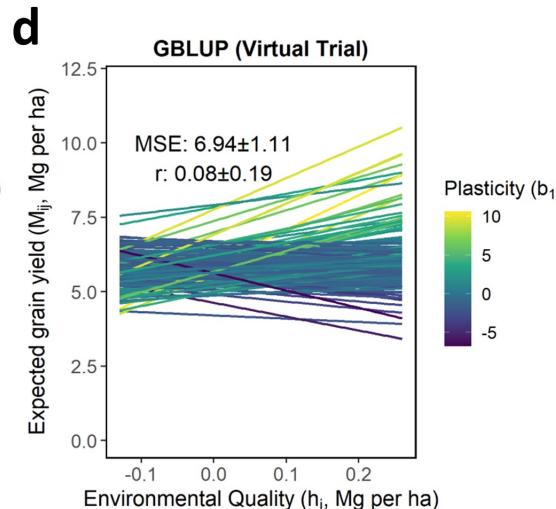
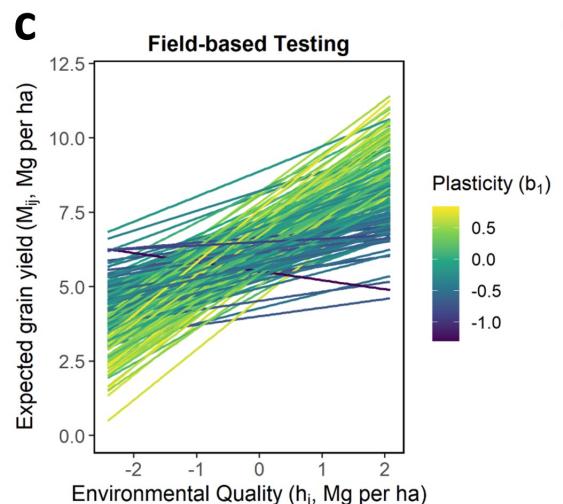
# Good: E-GP using T-matrix allows an early screening for plasticity

After optimization (some few genotypes, some environments)  $N < 10\%$  MET size

GxE Resolution  
(specific-genotype)



E Resolution  
Environmental diversity

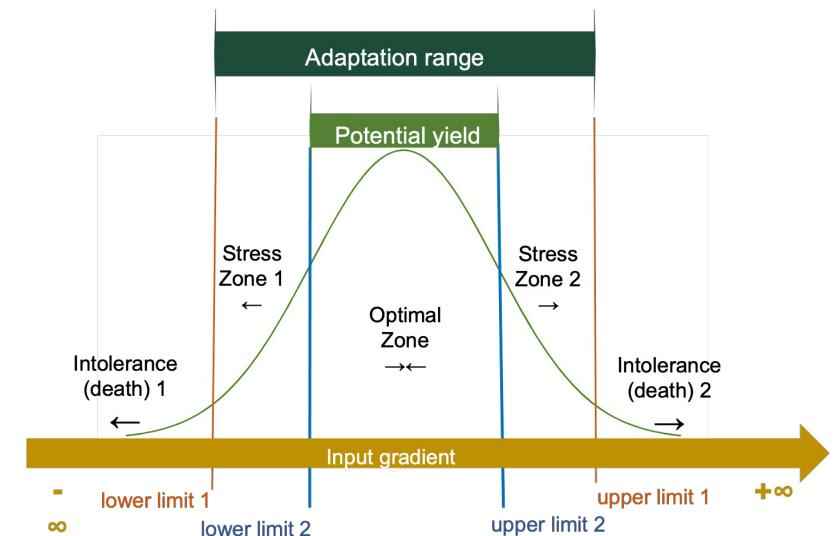
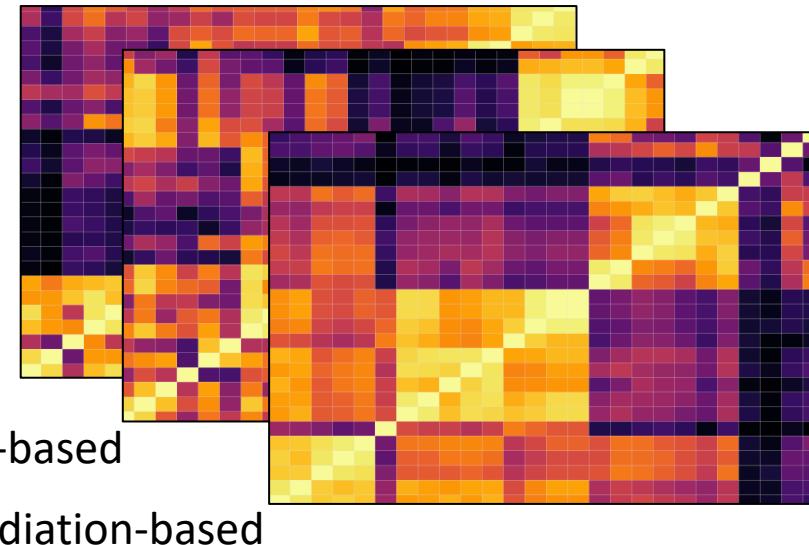
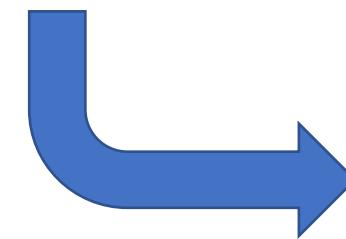
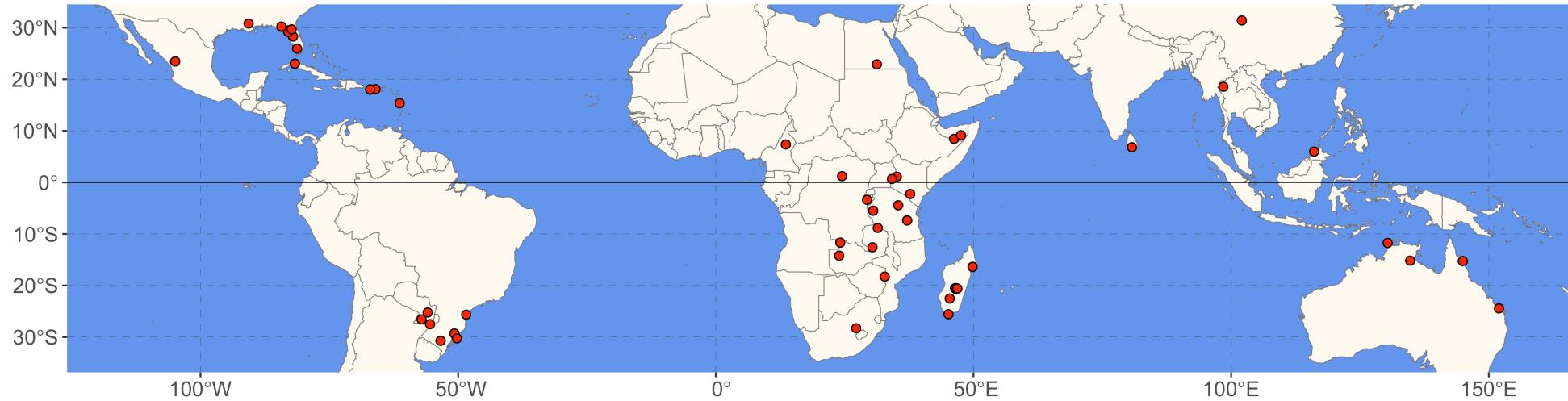


G+GE Resolution  
(linear plasticity)

Slope =  $1+b$

# Multiple enviromic “layers” for adaptive allele mining

Andropogoneae



# “Enviromics Network” – Collaborators, Support & History Back in Brazil



Alexandre B. Heinemann  
(Embrapa)



Roberto Fritsche-Neto  
(ESALQ/USP, IRRI)



João Batista Duarte (EA/UFG)



PhD thesis funded by



Conselho Nacional de Desenvolvimento  
Científico e Tecnológico

BILL & MELINDA  
GATES foundation

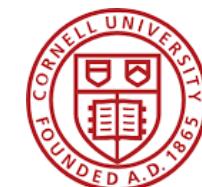


Jose Crossa  
(BSU/ CIMMYT)

Now:

Postdoc at Buckler's Lab (2021-date)

ArC CoE for Plant and Nature Success



Cornell University®

# Q & A

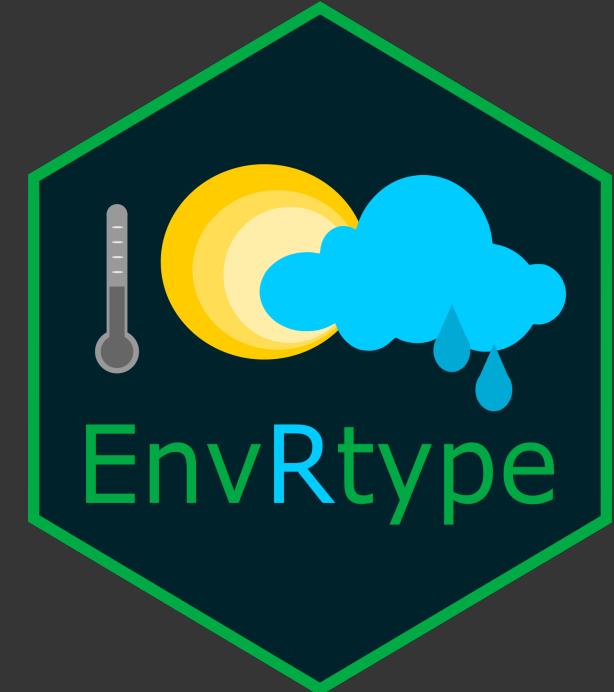
Germano Costa Neto

Postdoctoral Research Associate

Cornell University, Ithaca, NY

[gmc222@cornell.edu](mailto:gmc222@cornell.edu)

[github.com/gcostaneto](https://github.com/gcostaneto)



<https://github.com/allogamous/EnvRtype>

# EnvRtype Pipeline

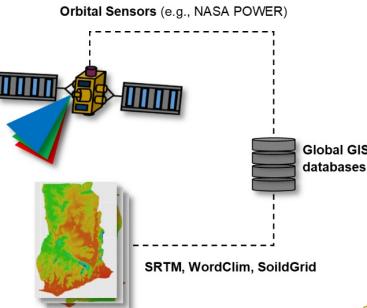
## Module 1

### 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



### Raw-data Processing

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

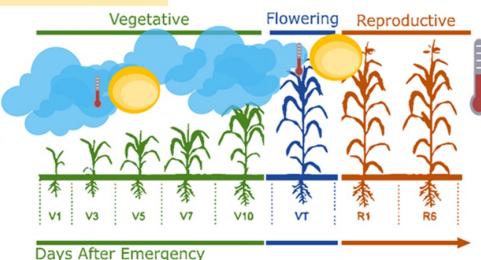
**Data processing** involves the quality control and computation of additional variables

## Module 2

### Environmental Characterization

- `env_typering()`
- `W_matrix()`

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



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

## Outputs

### Reaction-Norm

examples  
Factorial Regression GxE analysis

### Environmental Grouping

Clustering (K-mean)  
TPE definition

Environmental covariables

Envirotype descriptors

### Module 3

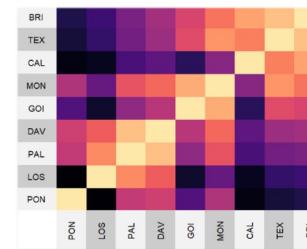
### Environmental Similarity

- `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_kernel()`
- `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)

### Genomic-based Predictions

Genomic-estimated Reaction-Norm Multi-environment Prediction

# Enviromics enhance GS accuracy in historical yield trials

Application in wheat trials (CIMMYT, 2011-2018, India) predicting unbalanced GxY interactions

**Ugly:** Aim to predict future GxE using the known W-matrix

**Bad:** in reality, we don't know the future **W-matrix ( $\Omega$ )**

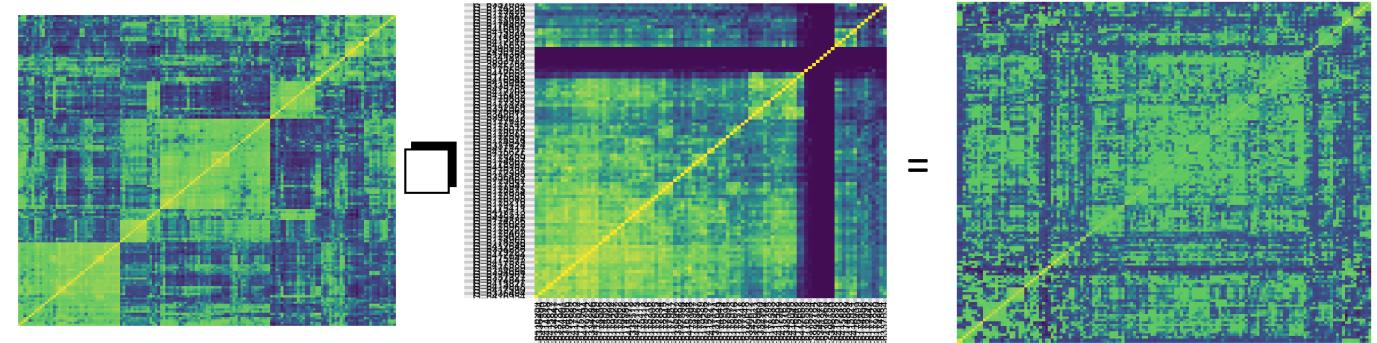
**Good:** historical data may provide good priors

- Static Location effects (environmental-wide association between yield and long-term envirotyping) – **S-matrix ( $\Upsilon$ )**
- Combined **S-W matrix ( $\Phi$ )**, envirotyping data + environmental weights as prior for a known location)
- “Adaptability pedigree” of the tested genotypes (shared past G+GE reaction-norms)

$$\Phi_{\tau,\omega} = \Upsilon + \begin{bmatrix} \Upsilon_{12}\Upsilon_{22}^{-1}(\Omega_{22} - \Upsilon_{22})\Upsilon_{22}^{-1}\Upsilon_{21} & \Upsilon_{12}\Upsilon_{22}^{-1}(\Omega_{22} - \Upsilon_{22}) \\ (\Omega_{22} - \Upsilon_{22})\Upsilon_{22}^{-1}\Upsilon_{21} & (\Omega_{22} - \Upsilon_{22}) \end{bmatrix}$$

with  $\Phi_{22} = [\tau\Omega^{-1} + (1 - \omega)\Upsilon_{22}^{-1}]^{-1}$

3 TPEs (Crespo-Herrera et al., 2021)



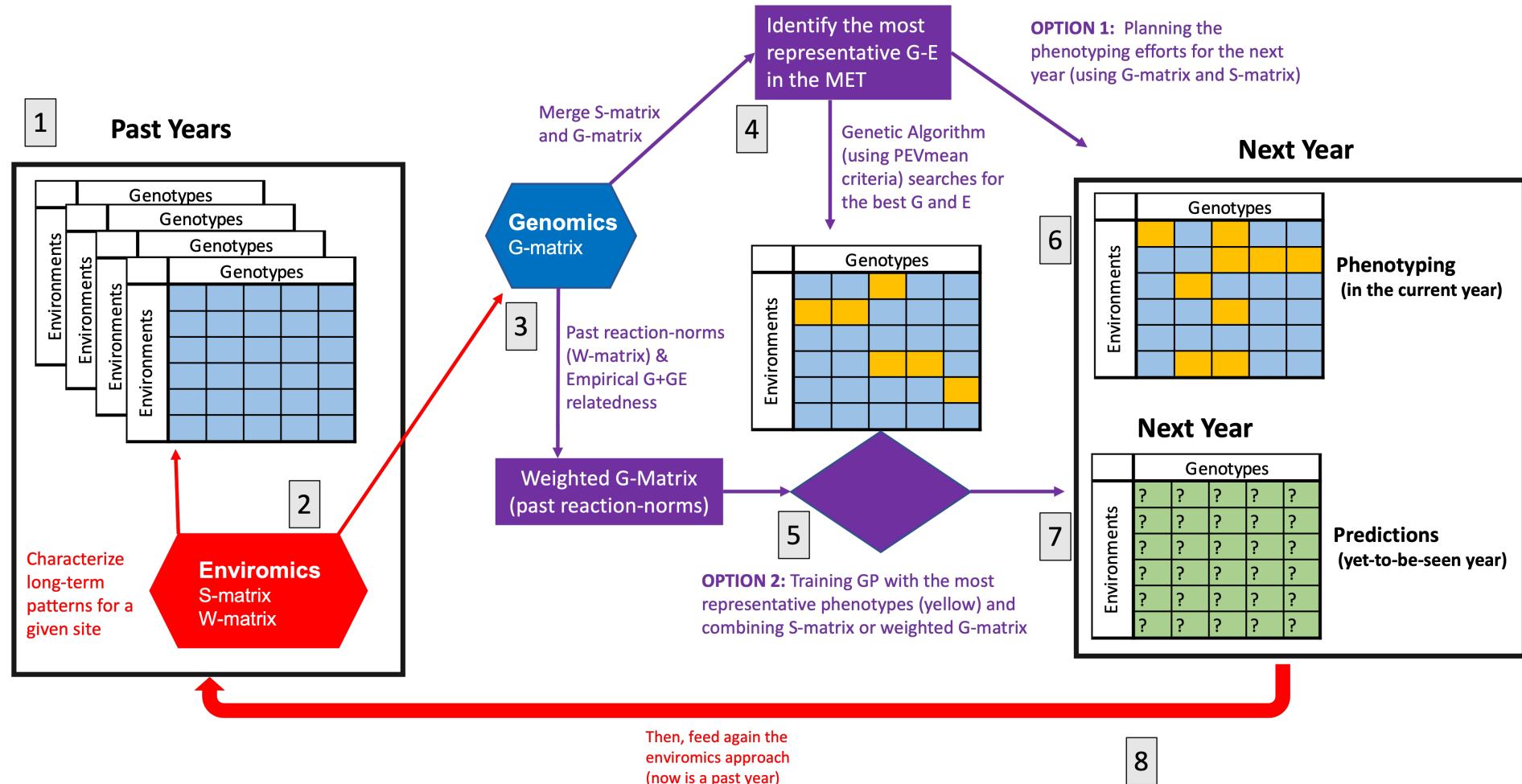
Static Location Diversity  
( $\Upsilon$  from S-matrix)

$\Omega$  Environment Diversity  
(from known W-matrix)

$\Phi$  Connecting past and future

# Reinforcement Learning for GxY prediction and MET optimization

## Wheat (yet-to-be-seen years)



# CIMMYT's Wheat Prediction (India, 2011-2018)

Costa-Neto et al (in prep)

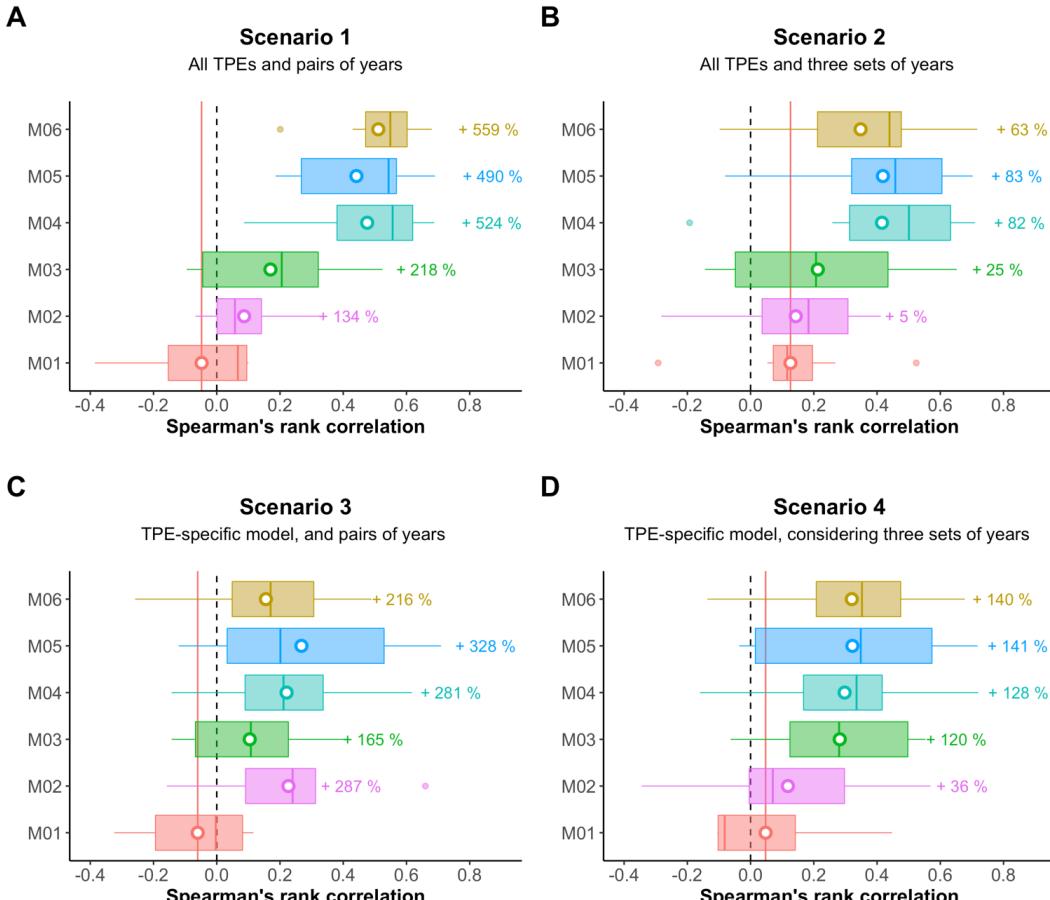
M01: no enviromics

M02: linear reaction-norms

M03: nonlinear reaction-norms

M04-M05: new propositions including historical env-wide associations with yield

## Prediction of new GxY



## Resolution in predicting new GxY (correlation by genotype across locations)

