



Exploring Genotype-by-Environment Interaction in Plant Breeding: Something Old, Something New, Something Borrowed, Something BLUE

Prof. Kaio Olimpio das Graças Dias (kaio.o.dias@ufv.br)

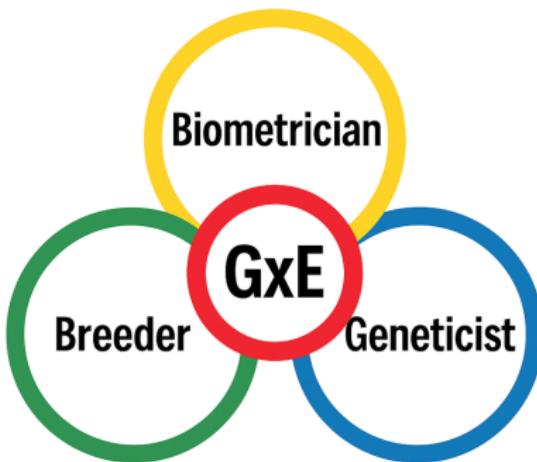
Departmento de Biologia Geral

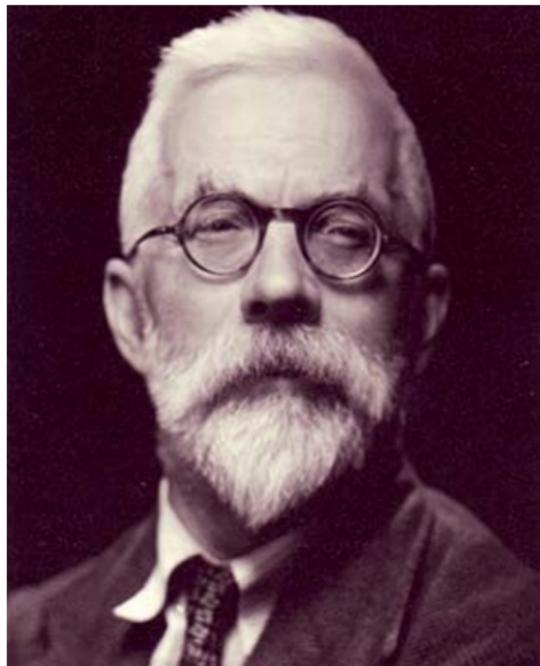
Universidade Federal de Viçosa

Outline

- ▶ Origins of GxE
- ▶ The statistical analysis of MET
 - ▶ Enviromics
 - ▶ GIS-FA
- ▶ Probabilistic Models
 - ▶ ProbBreed
- ▶ References

GxE and Breeding Marriage





Origins of GxE

- R. A. Fisher
- Biometric Concept
- 1923
- “A breakdown of additivity between genotypic and environmental sources of variation”

Origins of GxE

- Lancelot Hogben
- Developmental Concept
- 1933
- “Variation resulting from specific genotype and environment combinations during individual development”



Concept of GxE

Genotype–environment interaction, refers to cases in which different genotypic groups respond differently to the same array of environments

Journal of the History of Biology (2008) 41:717–761
DOI 10.1007/s10729-006-9155-y

© Springer 2008

R. A. Fisher, Lancelot Hogben, and the Origin(s)
of Genotype–Environment Interaction

JAMES TABERY

Department of Philosophy
University of Utah
Salt Lake City, UT 84112
USA
E-mail: tabery@philosophy.utah.edu

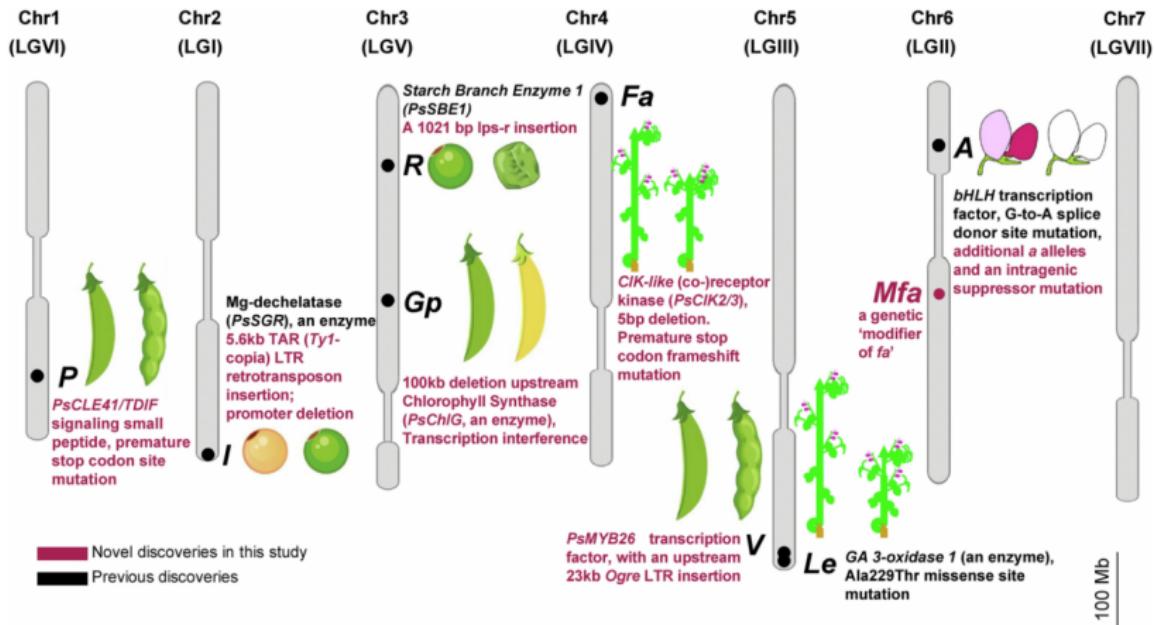
Abstract. This essay examines the origin(s) of genotype–environment interaction, or $G \times E$, “Original” and not “the origin” because the thesis is that there were actually two distinct concepts of $G \times E$ at this beginning: a *biometric* concept, or $G \times E_B$, and a developmental concept, or $G \times E_D$. R. A. Fisher, one of the founders of population genetics, was instrumental in establishing the biometric concept of $G \times E$ as a hereditarian concept as he attempted to resolve one of the main problems in the hereditarian tradition of biology – partitioning the relative contributions of nature and nurture responsible for variation in a population. Lancelot Hogben, an experimental embryologist and also a statistician, introduced the developmental concept as he attempted to resolve one of the main problems in developmental biology – partitioning the relative contributions that developmental relationships between genotype and environment played in the generation of variation. To argue for this thesis, I outline Fisher and Hogben’s separate routes to their respective concepts of $G \times E$; then these separate interpretations of $G \times E$ are drawn to explain a dispute between Fisher and Haldane over the importance of $G \times E$. The ramifications of a persistent controversy. Finally, the concepts of $G \times E_A$ and Hogben’s $G \times E_D$ are traced beyond their own work into mid-20th century population and developmental genetics, and then into the infamous IQ Controversy of the 1970s.

Keywords: analysis of variance (ANOVA), biometry, developmental biology, eugenics, genetics, genotype–environment interaction ($G \times E$), IQ controversy, Lancelot Hogben, nature–nurture debate, population genetics, R. A. Fisher

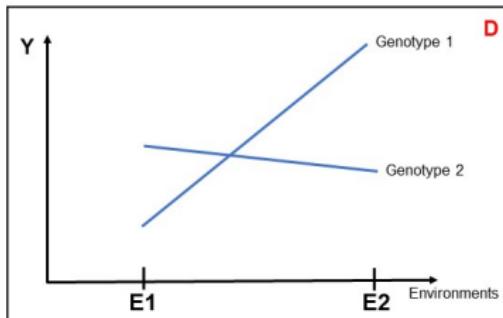
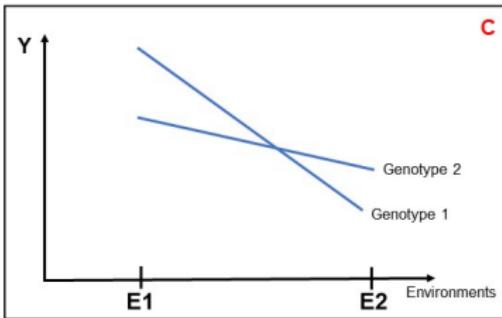
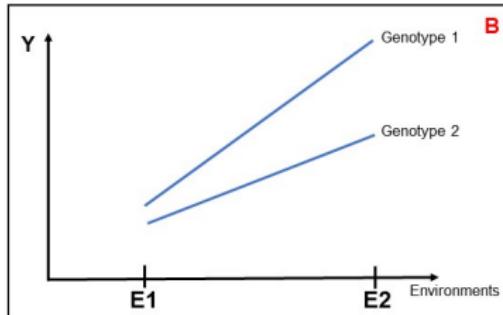
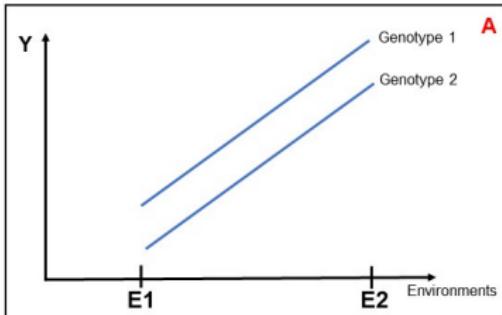
Mendel's laws of heredity

Nature Genetics 2025

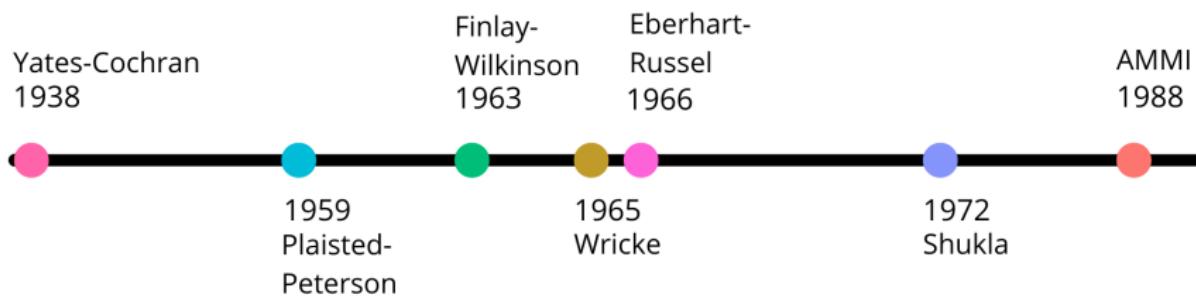
From: [Genomic and genetic insights into Mendel's pea genes](#)



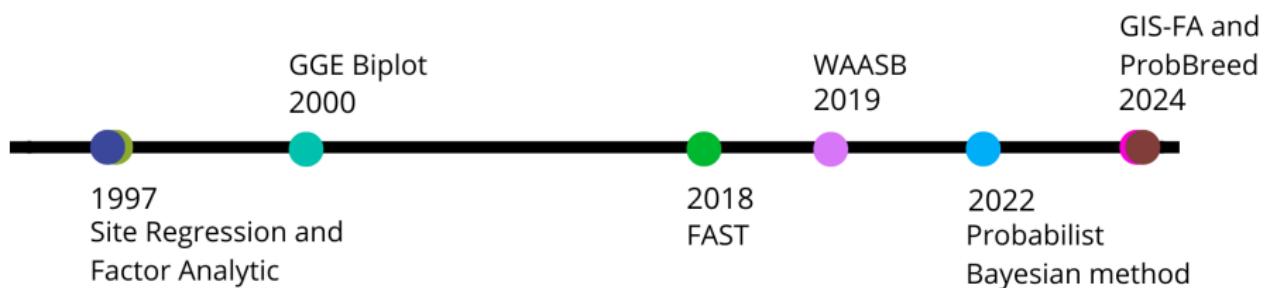
Genotype-by-Environment Interaction



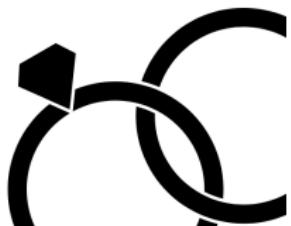
Analysis of MET



Analysis of MET

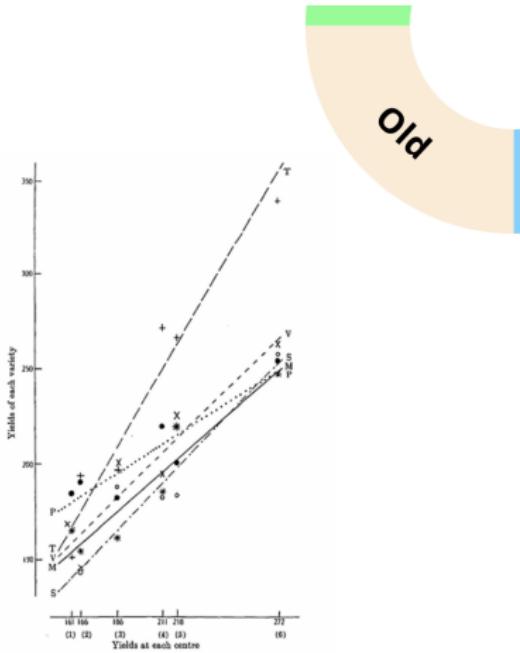


**Something old, something new,
something borrowed, something
blue**



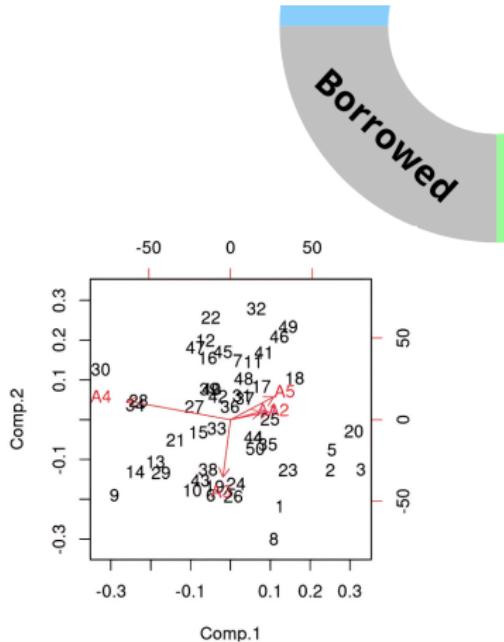
GxE Study

- ANOVA
- Classical tests for detecting GxE through variance components and linear regression methods on the environmental gradient
- Yates-Cochran (1938), Finlay-Wilkinson (1963), Eberhart-Russel (1966)



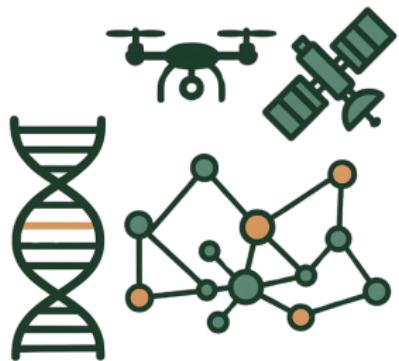
GxE Study

- Mixed Models
- Multivariate methods to dissect the interaction
- Biplot (1971), AMMI (1988), Factor Analytic (1997)



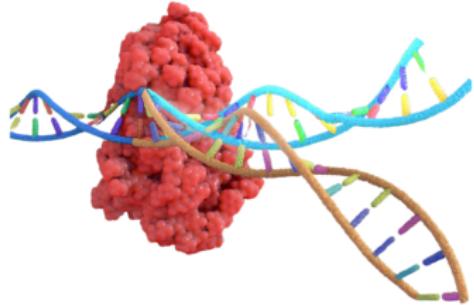
GxE Study

- Integration of Omics, Bayesian, and Machine Learning
- Probabilist Bayesian method (2022), GIS-FA (2024)



GxE Study

- Machine Learning and Genomics at Their Peak
- Use CRISPR-based tools to directly edit or assemble causative variants, enabling the construction of custom allele combinations across the genome
- Bypass lengthy crossing cycles by directly inserting and combining optimal functional SNPs



	The Good	The Bad	The Boring
ANOVA based	Simple and easy to implement	Require homoscedasticity, normality, and independence of residuals. Also, do not handle unbalanced data well.	Assumption checking
Mixed Models based	Allow modeling more complex covariance structures and incorporating pedigree information	Require REML estimation of variance components, which can be computationally heavy on large datasets	Difficulty in convergence when fitting models with many parameters
Bayesian	Allow probabilistic interpretation of risk and uncertainty, such as credibility intervals for GxE	Computationally intensive due to the need for sampling methods like MCMC	Choice of priors and convergence criteria evaluation
Machine Learning	Capture complex and nonlinear interactions	Difficult to extract direct biological interpretations	Hyperparameter tuning (CV is time-consuming)

On the road of breeding 4.0

Breeding 1.0



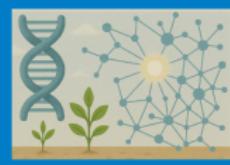
Breeding 2.0



Breeding 3.0



Breeding 4.0



(Wallace, 2018)

On the road of breeding 4.0

How do we adapt crops to better fit agricultural environments?



In Breeding 3.0

Theoretical and Applied Genetics (2022) 133:1385–1399
<https://doi.org/10.1007/s00122-022-04301-y>

ORIGINAL ARTICLE



Leveraging probability concepts for cultivar recommendation
in multi-environment trials

Kalo O. G. Dias^{1,3} · Jhonathan P. R. dos Santos¹ · Matheus D. Krause² · Hans-Peter Piepho² · Lauro J. M. Guimarães⁴ ·
Maria M. Pastore⁴ · Antonio A. F. Garcia¹

Received: 22 April 2021 / Accepted: 7 January 2022 / Published online: 22 February 2022
© The Author(s), under exclusive license to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Theoretical and Applied Genetics (2024) 137:88
<https://doi.org/10.1007/s00122-024-04579-z>

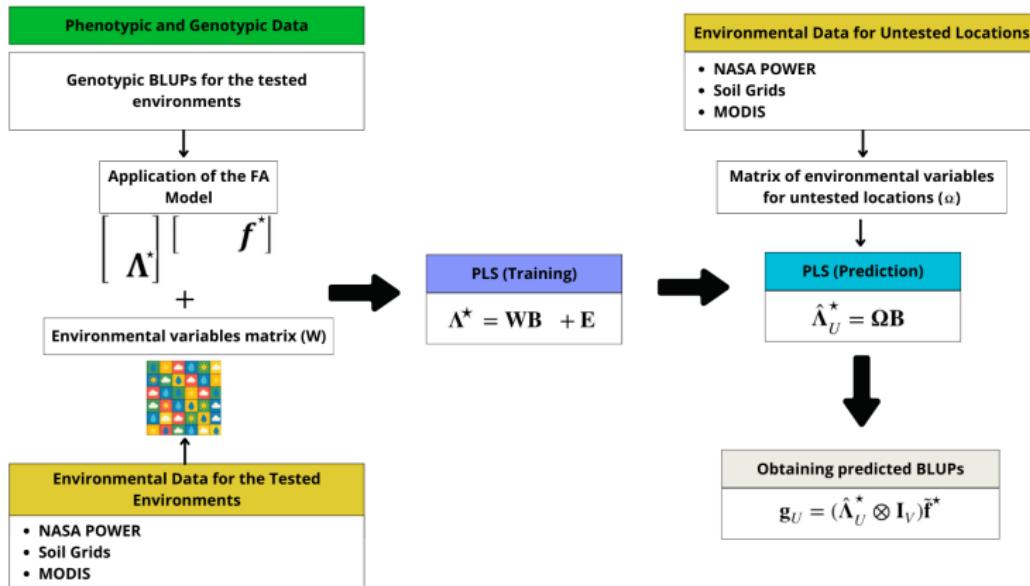
ORIGINAL ARTICLE

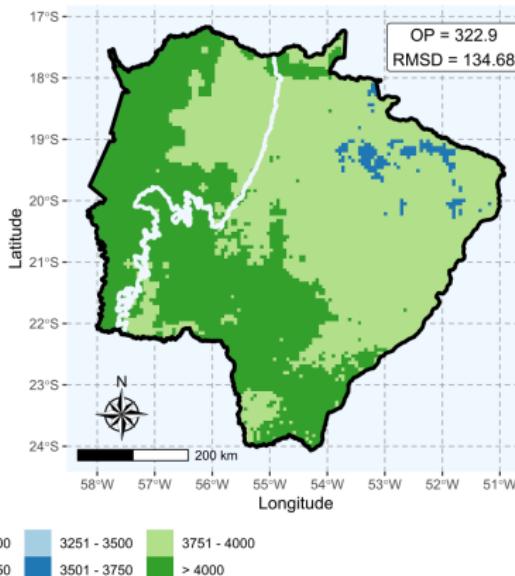
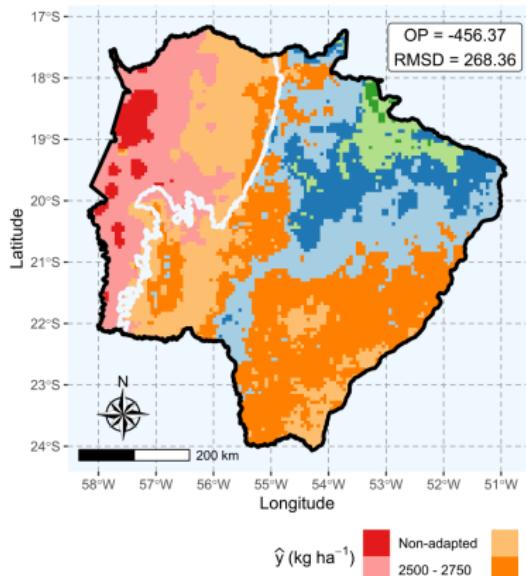


GIS-FA: an approach to integrating thematic maps, factor-analytic,
and envirotyping for cultivar targeting

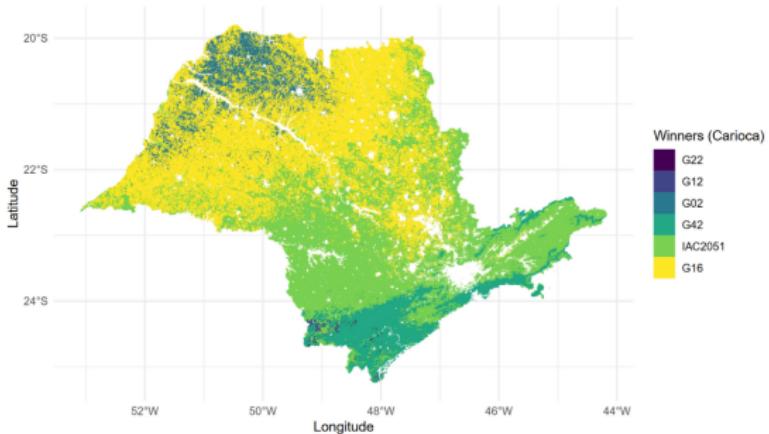
Mauricio S. Araújo¹ · Saúlo F. S. Chaves¹ · Luiz A. S. Dias¹ · Filipe M. Ferreira² · Guilherme R. Pereira³ ·
André I. G. Bezerra² · Rodrigo S. Alves⁴ · Alexandre B. Heinen⁴ · Flávio Brescghello⁵ · Pedro C. S. Carneiro⁵ ·
Matheus D. Krause² · Germano Costa-Neto⁶ · Kalo O. G. Dias¹

Received: 10 July 2023 / Accepted: 6 February 2024 / Published online: 12 March 2024
© The Author(s), under exclusive license to Springer-Verlag GmbH Germany, part of Springer Nature 2024



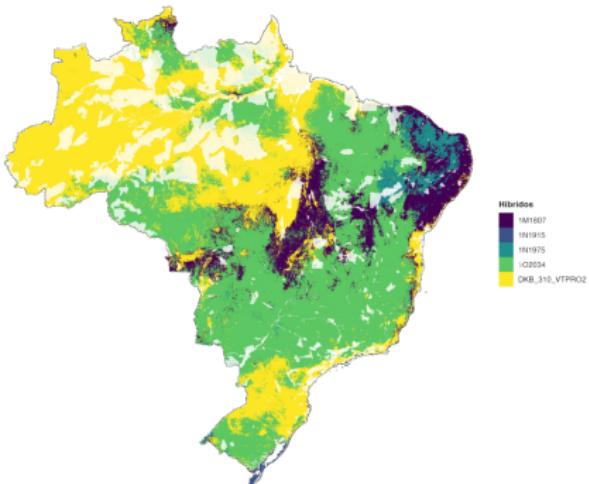


Common Bean, Which-won-where



Gabriel Blasques

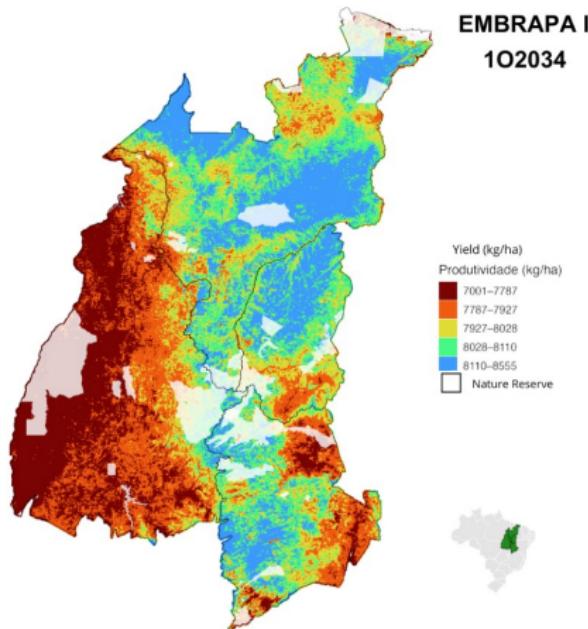
Maize, Which-won-where



Thiago Brommonschenkel

Integrating Environmental and Genomic MATOPIBA

UFV



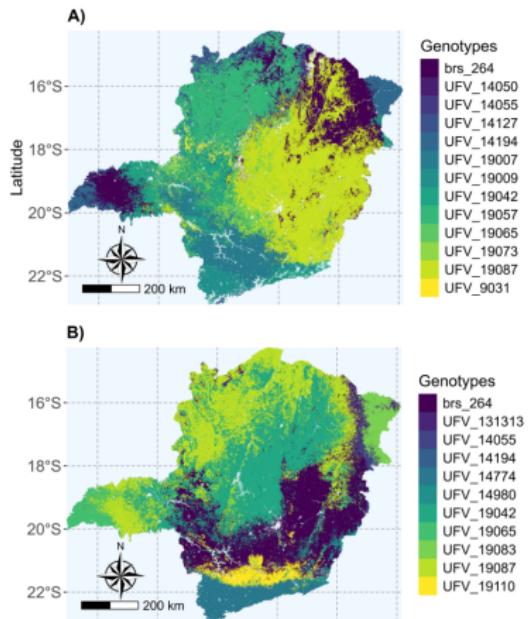
Genotype by Environment by Management (GxExM)

UFV

Wheat, Which-won-where

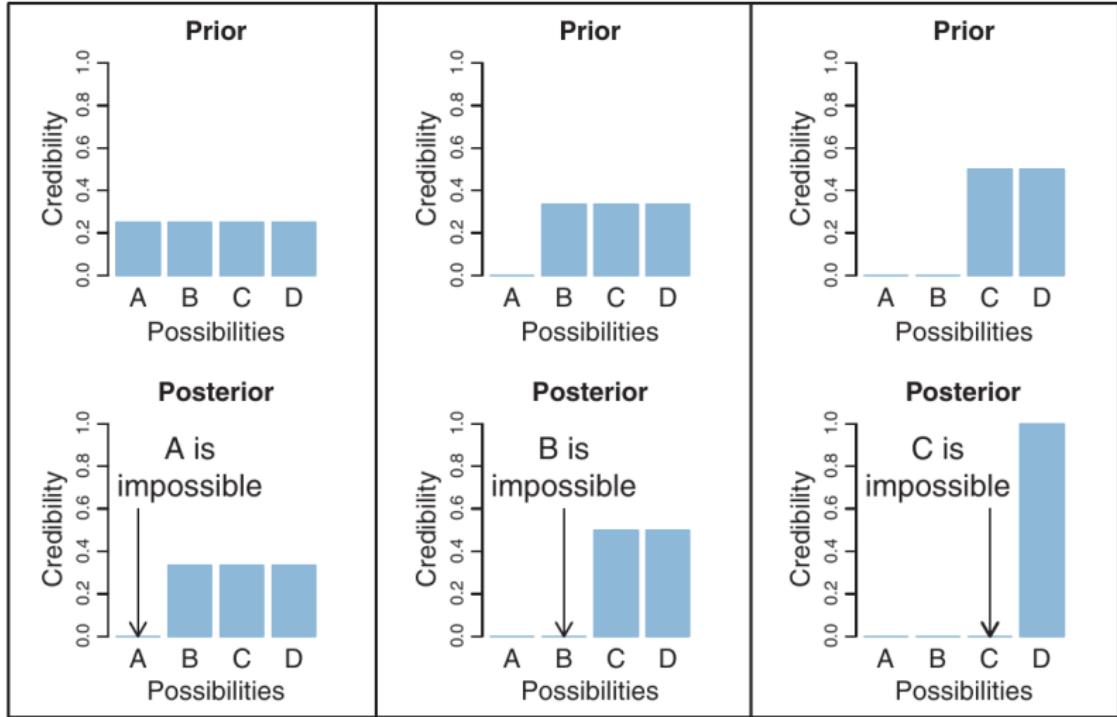


João Amario



- ▶ Probability of Superior Performance
- ▶ Probability of Superior Stability
- ▶ Which candidates are the top performers?
- ▶ What is the risk of recommending a given candidate (performance)?
- ▶ Is candidate x better than candidate y ?
- ▶ What is the probability that x performs better than y in the TPE?
- ▶ ...

Bayesian Inference



- ▶ Fitting a Bayesian model

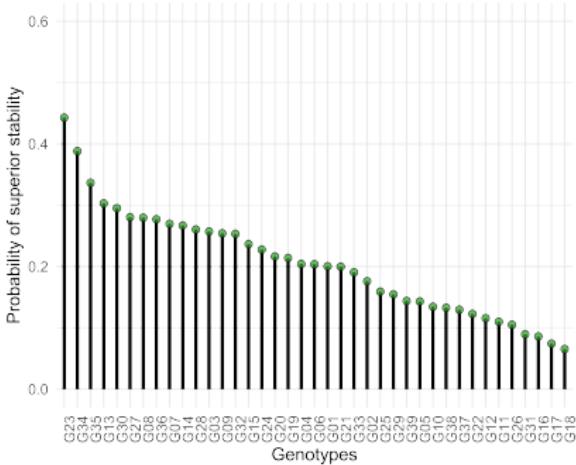
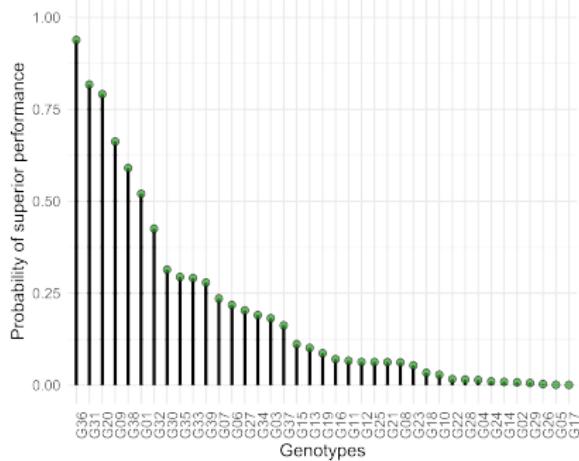
$$y_{jklm} \sim N(E[y_{jklm}], \sigma)$$

$$E[y_{jklm}] = \mu + g_j + e_k + r_{m(k)} + b_{l(mr)} + (ge)_{jk}$$

$$Pr(\hat{g}_j \in \Omega | y) = \frac{1}{S} \sum_{s=1}^S I(\hat{g}_j^{(s)} \in \Omega | y)$$

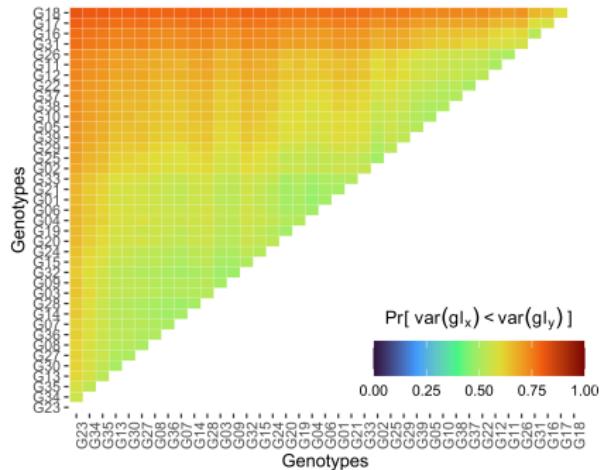
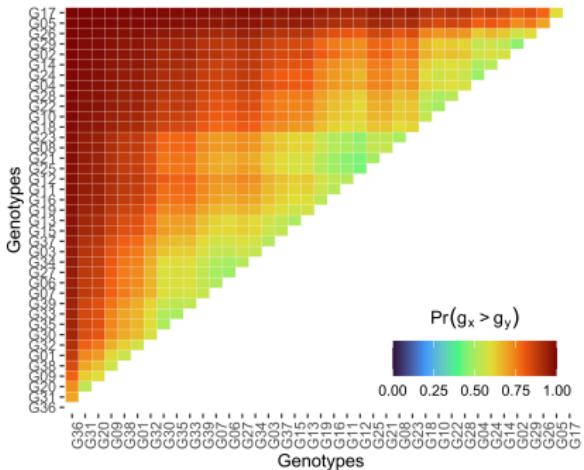
$$\begin{cases} \hat{g}_j \in \Omega \rightarrow I(\hat{g}_j^{(s)} \in \Omega | y) = 1 \\ \hat{g}_j \notin \Omega \rightarrow I(\hat{g}_j^{(s)} \in \Omega | y) = 0 \end{cases}$$

Bayesian Probabilistic Methods



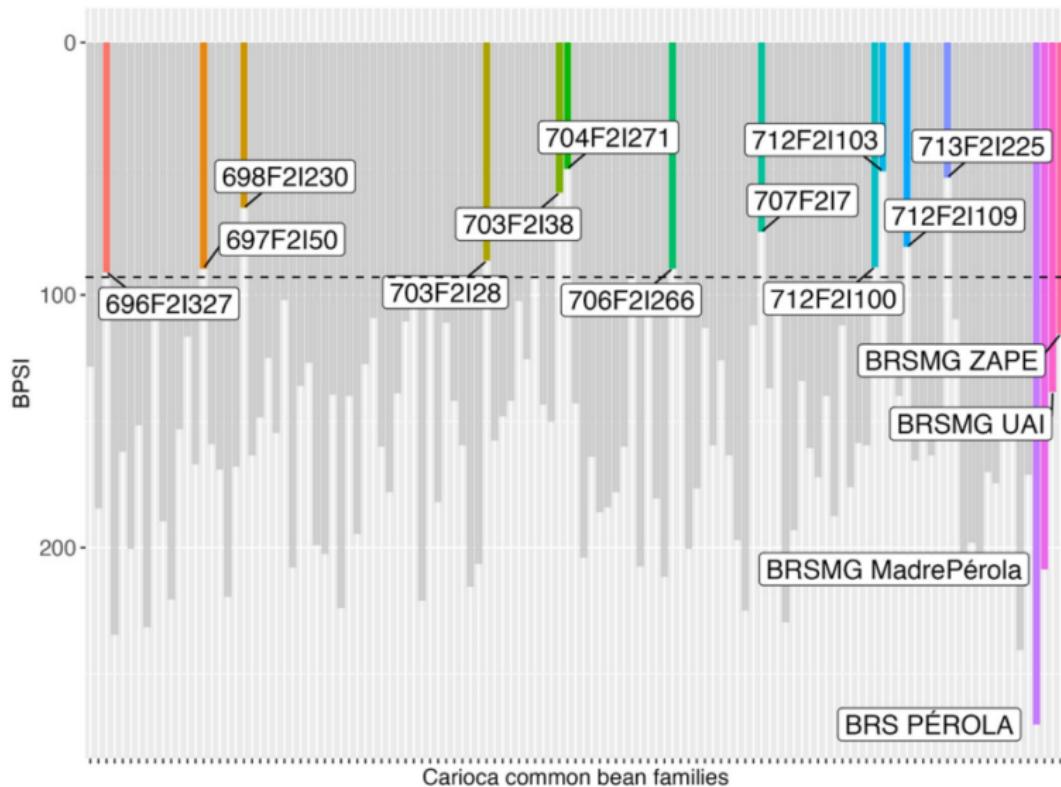
Bayesian Probabilistic Methods

UFV



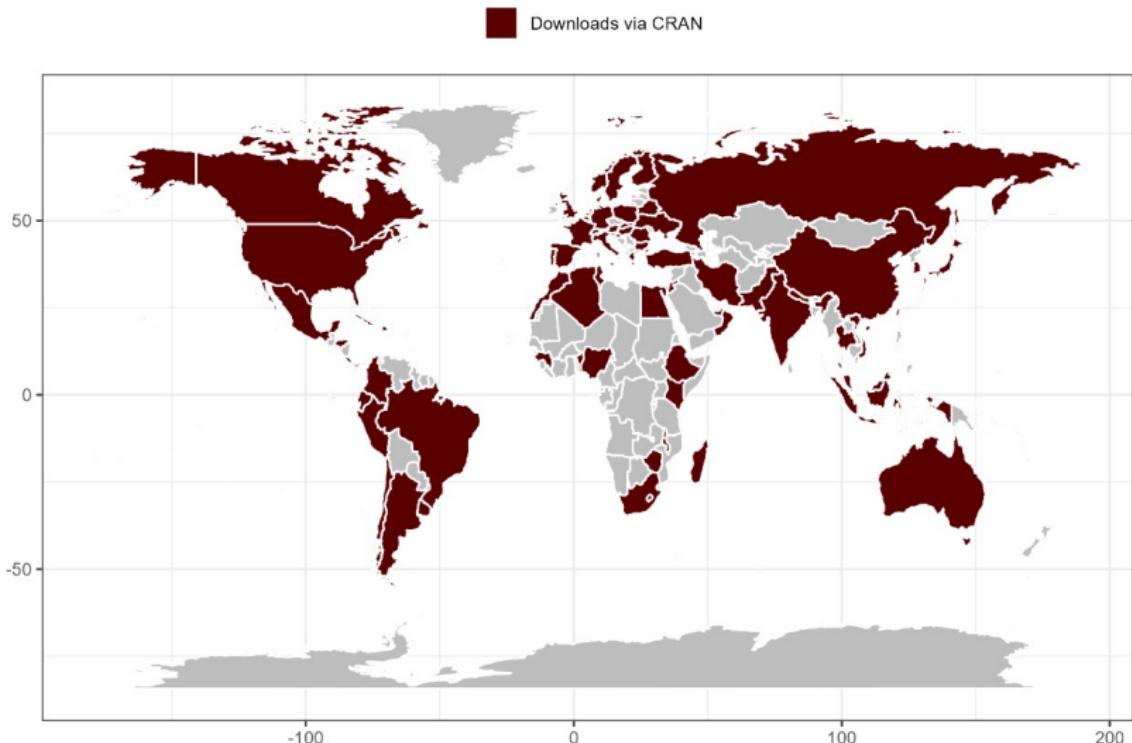
Bayesian probabilistic selection index

UFV

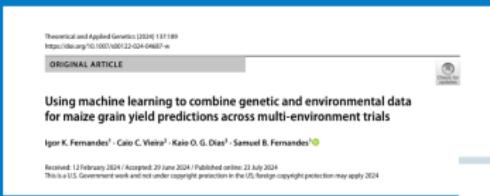


ProbBreed

ProbBreed is a free package, available from CRAN



In Breeding 4.0



Theoretical and Applied Genetics (TARG) 133(1) 137–149
https://doi.org/10.1007/s00122-024-04887-w
ORIGINAL ARTICLE

Using machine learning to combine genetic and environmental data for maize grain yield predictions across multi-environment trials

Igor K. Fernandes¹ · Caió C. Vieira² · Káte O. G. Dias³ · Samuel B. Fernandes⁴ ·

Received: 12 February 2024 / Accepted: 29 June 2024 / Published online: 23 July 2024
This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024



www.nature.com/scientificreports/

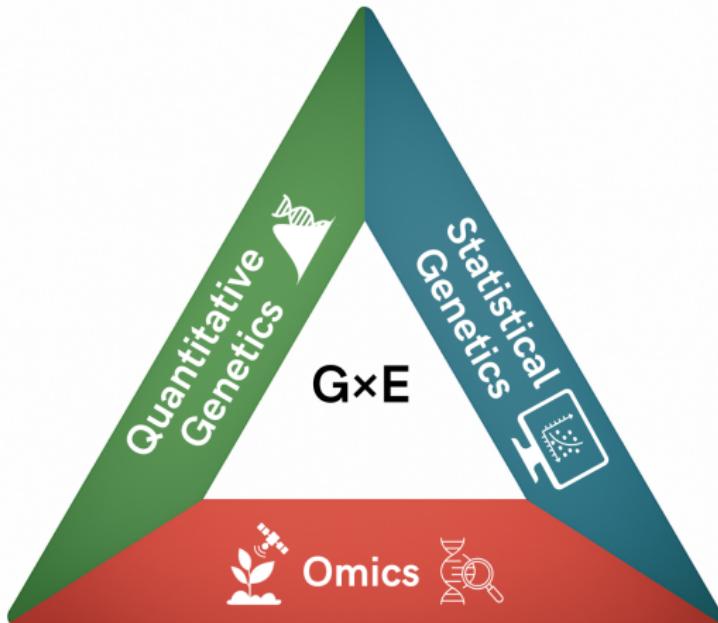
scientific reports

Open Genomic prediction in multi-environment trials in maize using statistical and machine learning methods

Cettito Aparecida Valenti Barreto¹, Káte Olímpia dos Gracis Dias²,
Itálio Coelho de Souza³, Cândia Ferreira Alvesnido⁴, Ana Carolina Campagni Nascimento⁵,
Lauro José Moreira Guimarães⁶, Cláudia Teixeira Guimarães⁷, Maria Marta Penteado⁸ &
Miguel Nascimento⁹

Take-Home Message

The Modern G×E Triangle



Acknowledgments

Embrapa Milho e Sorgo

Dr. Maria Marta Pastina
Dr. Lauro José Moreira Guimarães
Dr. Cláudia Teixeira Guimarães
Dr. Jurandir Vieira Magalhães

ESALQ/USP

Dr. Antonio Augusto F. Garcia

University of Florida

Dr. Marcio Resende
Dr. Diego Jarquin

VSN International

Dr. Salvador Gezan

University of Arkansas

Dr. Samuel Fernandes

Universität Hohenheim

Dr. Hans-Peter Piepho

University of Nebraska

Dr. George Graef
Dr. Arthur Bernadelli

UFV

Dr. Guilherme da Silva Pereira
Dr. Maicon Nardino
Dr. Pedro Crescêncio S. Carneiro
Dr. Rodrigo Oliveira de Lima
Dr. Leonardo Lopes Bhering
Dr. José Marcelo Soriano Viana
Dr. Luiz Antônio dos Santos Dias
Dr. Moyses Nascimento
Dr. Luiz Peternelli
Dr. Maximiller Dal-Bianco L. Costa

Acknowledgments

Instituto de Inteligência Artificial e Computacional da UFV (IDATA)



Leonardo Lopes Bhering
Diretor do IDATA



References



[Bernardo, 2020]

Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE
Heredity volume 125, pages375–385



[Dias et al. 2022]

Leveraging probability concepts for cultivar recommendation in multi-environment trials Theoretical and Applied Genetics 135 (4), 1385-1399



[Araujo et al. (2024)]

GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting

Theoretical and Applied Genetics 137 (4), 1-23



[Dias et al. 2018]

Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials

Heredity volume 121, pages24–37



[Chaves et al. 2024]

ProbBreed: a novel tool for calculating the risk of cultivar recommendation in multienvironment trials

G3: Genes, Genomes, Genetics 14 (3), jkae013