Enviromics: bridging different sources of data, building one framework

Germano Costa-Neto[1,2] and Roberto Fritsche-Neto[2,3]*

Abstract: *Enviromics is the field of applied data science that integrates databases of environmental factors into biostatistics and quantitative genetics. It can leverage plant ecophysiology knowledge to bridge the gaps about environment interactions with systems biology (genes, transcripts, proteins, and metabolites), which also boosts the ability to understand and model the phenotypic plasticity of the main agronomic traits. Recently, the plant breeding community has experienced reduced costs for acquiring environmental sensors to be installed in the field trials while increasing the reliability and resolution of the remote sensing techniques. The combination of those two factors has started the spring of enviromics-aided breeding in recent years. However, the use of environmental information in plant breeding is not a novelty approach developed a few years ago, but a core of efforts originated in the last 60 years, yet some basic ideas traced back to early 20th century attempts to establish a relationship between phenotypic and environmental variation. This review highlights the main concepts surrounding the construction of the "modern enviromics science", tracing back to its origins in the last decades. Finally, we present how this field has helped integrate different data sources in prediction-based models or build one framework.*

Keywords: *Prediction-based breeding, breeding analytics, envirotyping, genotype × environment*

INTRODUCTION

One of the most promising tools of Modern Plant Breeding is the enviromics (Resende et al. 2020, Crossa et al. 2021), yet it less explored omics-based approach (Costa-Neto et al. 2021c) even after several works describing its benefits (Cooper et al. 2014, Xu 2016, Voss-Fels et al. 2019). Most of this neglected usefulness relies on the lack of a clear envirotyping pipeline to provide analytic data to accomplish this field's theoretical concepts. In this context, we envisage the need to review basic concepts and the historical background underlying environmental information in predictive breeding and analytics.

Our objective here is to discuss the differences between envirotyping, enviromics, and the simple use of any environmental information in breeding decisions under G×E prediction scenarios. To achieve this objective, we first highlight some important concepts regarding the genotype-environment interaction and how the environment shapes the phenotypic variation of complex traits. Then we present a historical timeline from the early 20th century until the present days, passing by the benchmark works for elaborating the "enviromics

*Corresponding author:
E-mail: r.fritscheneto@irri.org
ⓘ *ORCID: 0000-0003-4310-0047*

[1] Cornell University, Institute for Genomic Diversity, Ithaca, USA
[2] University of São Paulo, Escola de Agricultura "Luiz de Queiroz", Departamento de Genética, Avenida Pádua Dias, 11, 13418-900, Piracicaba, SP, Brazil
[3] International Rice Research Institute (IRRI), Breeding Analytics and Data Management Unit, Los Baños, Philippines

theory". Finally, we envisage the "treasures" mined to establish enviromics-aided predictive breeding platforms that integrate different data sources in prediction-based models or build one framework.

KEY QUESTIONS FOR UNDERSTANDING "ENVIROMICS"

What is "environment"?

As geneticists, we tend to interpret environment as the core of sources outside of the Central Dogma of Molecular Biology: the effects surrounding the rate of gene expression, driving the process of transcription and its translation in proteins that results in the observed phenotype. It also is related to other processes such as regulating epigenetic factors that affect gene expression. Thus, the environment is an emergent property derived from the balance of inputs and frequency across the plant's lifetime. Therefore, the environment is not a "fixed property", but a gathered effect of multiple factors and its balance driving the quality of the growing conditions for a given species, germplasm, or genotype. From the agriculture or forestry point of view, and translating this idea to a realistic situation, we can define "environment" as a certain time window between planting date and harvesting, that is, experimental treatment or replication of the genetic treatments across the time and the space. At a certain experimental design, the replications conducted in some field trials aim to control micro-environmental effects that are not part of the genetic treatments under study, as pointed by Fisher in 1918. At multi-environment trials (MET), this problem increases, which demands some knowledge about the macro-environmental effects that must be considered to visualize better how the environment shapes the phenotypic variation of plants.

How does the "environment" shape the phenotypic variation?

Understanding how the environment (balance and availability of inputs) regulates growth and development is an important step to understand at which level the enviromics can be used in Systems Biology. Thus, at least three clear levels of environmental acting can be reasonably described.

At the cellular level, the first acting is by regulating gene expression, activating cellular divisions, and cellular growth. Then, it also drives the dynamics of the transcripts within the surrounding cell environment, activating enzymes and determining the rate of protein denaturation during the lifetime of the cell. Secondly, it acts outside of the cell environment, impacting the interaction between cells and plant organs, regulated by the availability of resources and metabolites important for plant physiology. Third, the (eco) physiology of the plants is modulated by external factors surrounding the micro-environments of the plant, that is, its biotic and abiotic interactions. Some of the biotic factors might include the genotype-specific interactions with the soil microbiome and the responsiveness to pests, diseases, and neighbor plants. For the abiotic effects, it consists of the soil-plant-atmosphere continuum, that is: how the ecophysiological responses due to the soil water availability, fertilizer levels, air temperature, solar radiation, and other conditions related to a certain time window of plants development, at a certain cultivation practice, at a certain location in the world. Consequently, the core of those inputs (also referred to as plant stimuli) directly affects the sanity of specific tissues or plant organs, which reflects in the availability of certain metabolites in the cells and finally comes back to the differential gene expression. A reflection of this phenomenon, for instance, is the measured leaf temperature impacting the differential production of hormones and plant architecture (Patel and Franklin 2009, Castroverde and Dina 2021) and the soil water availability in the roots regulating leaf stomata (Buckley 2019), which directly affects the final yield status. Thus, the balance of those factors, in terms of frequency of occurrence in different phenological stages, is also strongly related to the magnitude of those factors in limiting the potential phenotypic expression of the genotypes.

Because of the well-established interactions studied in fields such as plant pathology and crop ecophysiology, it is also possible to trace the interactions between biotic and abiotic factors affecting the phenotypic expression of diverse genotypes. Therefore, quantitative geneticists can take advantage of those studies to establish relationships between the performance of target agronomic traits as a function of the environmental gradient (range of environmental conditions) experienced by the crops as a signal of the possible drivers of genotype by environment interaction. This is why from ecophysiology, we can infer those plants are static in some 'location', and to survive to the balance of inputs (which sometimes affects it negatively), the plasticity of the genotype is a key for understanding the phenotypic variation (Bradshaw 1965). This idea also leads the physiologists back in the 1960's decade to start developing mechanistic crop

growth models capable of measuring those relations, while plant breeders tried to establish linear responsiveness of each genotype for certain key environmental factors – the reaction norm modeling.

What is genotype by environment interaction?

Genotype by environment interaction (G×E) results from the genotype-specific plasticity for the surrounding macro-environment fluctuations in the lifetime of the crops (Hogben 1932, Allard and Bradshaw 1964, Bradshaw 1965, Arnold et al. 2019). If the environment modulates the rate of gene expression (e.g., Jończyk et al. 2017, Liu et al. 2020) and fine-tuning epigenetic variations related to transcriptional responses (Vendramin et al. 2020, Cimen et al. 2021), there are many sources to bridge the genetic diversity and phenotype variation that implies on G×E variation for each genotype at each growing condition. Because of that, the common way to compute the G×E in breeding trials is to assume it as a differential and multiplicative interaction between genetic (G) and macro-environment effects (E), in which micro-environmental factors (plot-level) are corrected using experimental design structures (e.g., replications, blocks) and statistical analysis for the phenotype correction (e.g., linear models, spatial analysis). This last is capture by the residual variance effects into the first stage of statistical analysis of field trials. Thus, all of the unknown processes between the potential phenotypic plasticity and the modeled sources of variation (G, E), and also all model misspecifications, are accounted as G×E and residual variance, respectively.

To further understand G×E, it is needed first to understand three main concepts: i) phenotypic plasticity; ii) reaction-norm; iii) interrelation between target population of environments (TPE) and multi-environment trials (MET). Figure 1 illustrates the basics to explain that the G×E is not a phenomenon *per se* but a consequence of the environmental range of the multi-environment trials (MET) (and its representativeness) and the lack of knowledge on modeling the sources of phenotypic variation. Let's consider four experimental networks (the core of field trials), conducted under MET conditions for the same TPE, and considering two distinct genotypes (G1 and G2). The curves of each genotype illustrate its particular (and potential) phenotypic plasticity: the ability to shape the phenotypic expression according to the environmental gradient experienced. As discussed in the last section, this plasticity results from the particular system biology dynamics of the genotypes. Excess or deficit of environmental inputs is a stressful factor limiting the potential expression of the genotypes. The vertical solid lines indicate the range of the experimental network conditions:
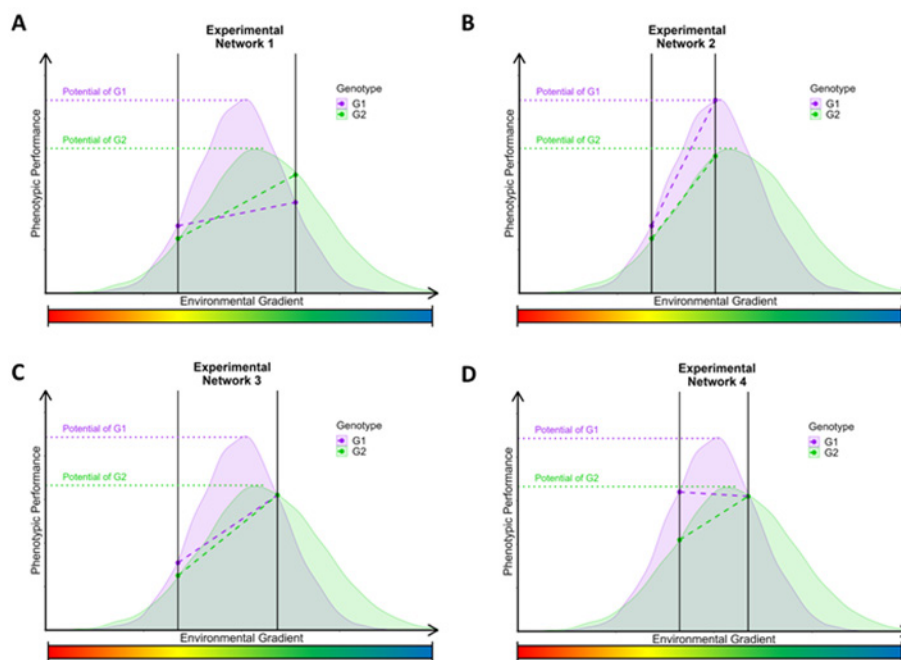

Figure 1. The interplay between phenotypic plasticity and reaction-norm as a function of the observed environmental gradient range.

the maximum and minimum input level that the genotypes experienced in each MET condition.

According to these environmental limits, each genotype expresses a different reaction-norm: nonlinear responsiveness to the gradient of growing conditions. The concept of reaction-norm is an old idea but reflects particular genotype-specific sensibilities for key environmental factors – which defines the adaptability of the genotypes for the MET condition. Finally, at the dashed lines, it is possible to visualize the observed G×E pattern as an expectation of the particular adaptability of the genotypes. As illustrated, for each environmental gradient of each MET, it is possible to visualize a different G×E pattern for the same genotypes. At experimental network 1, the pattern is the crossover, that is, there is a rank change across the environments – also known as qualitative changes. On the other hand, if the range growing conditions differ, the pattern might be non-crossover (or simple), as observed at the experimental network 2,3 and 4 – quantitative changes in the phenotypic expression across the environments without rank changes. Thus, considering this illustration for a real-world breeding program scenario, it is expected to visualize a MET-specific G×E pattern for each germplasm at each MET. Because of that, the plant breeders must define the environmental limits of the TPE to conduct field trials that represent a sample of those conditions – and then explore the G×E to select the most adapted cultivars. The process of identifying TPE and relating it to the MET conditions can only be done using a diverse set of envirotyping analytic tools (Chenu et al. 2011, Heinemann et al. 2019, Crespo-Herrera et al. 2021).

Then, what is the difference between envirotyping, envirome, and enviromics?

Envirotyping gathers the steps of collecting, processing, and associating environmental data with phenotypic data to understand the typology of environments for some MET, which is conducted relating it to the target population of environments (TPE) of the breeding program. This is a two-way process approach, in which the definition of the best MET can also be done by analyzing the typology of the TPE. It is expected to the MET is a representative sample of the TPE, but every MET is composed of some germplasm (well know tested individuals or newly developed or introgressed materials) at yet-to-be-seen growing conditions, in which sometimes some specific field trial might not be so close to the TPE expectations. Furthermore, each germplasm has a unique interaction with the tested environments due to the change in allele frequencies across the years and the arrangement and introgression of variability, which directs impacts G×E relations. Thus, the envirotyping process might be continuously used to update the gene-environment relations for some TPE (and perhaps redefine the TPEs).

The process of deriving typologies relies on identifying the most impactful environmental factors for some specific trait (or core of main traits), in terms of impact and frequency of occurrence at some location and planting date within the TPE. For example, the effect of planting date shaping the climatic impacts is well understood in agrometeorology (Heinemann et al. 2019, Antolin et al. 2021). Thus, we can affirm that the same location with different planting dates might be targeted as different TPEs according to the climatic stability (and occurrence of diseases), for instance. The final goal of envirotyping is then to understand the expected G×E patterns to be achieved for some location, in which the pure environmental data might be associated with historical phenotypic records (or records from some specific MET) to identify the key factors impacting the reaction-norms of the germplasm (and consequently, the end-resulted G×E).

Enviromics is the large-scale envirotyping, based on the collections of environmental data across time and space to establish a global association between the crops envirome (the core of TPEs) and phenotypic variation of key factors driving G×E. From enviromics, it is possible to derive some useful outputs, such as an environmental relatedness matrix, which supposes that the core of typologies is a marker of "environmental polymorphisms" (Resende et al. 2020, Costa-Neto et al. 2021a, Costa-Neto et al. 2021c). Another important goal of enviromics is its further integration in systems biology approaches, which can support the understanding of genomics × enviromics interactions, but also its relation with different omics sources (e.g., transcriptomic, proteomic) and its impact on improving accuracy and resolution of predictive breeding platforms (Costa-Neto et al. 2021a, Costa-Neto et al. 2021b, Rogers et al. 2021).

ENVIRONMENTAL INFORMATION IN PLANT BREEDING HISTORY

The use of environmental information is not a recent idea surrounding the plant breeding community. Perhaps the first attempt of it started with Sir. Ronald Fisher (Fisher 1924) trying to establish a relationship between rainfall and wheat grain yield. However, the idea of establishing a relation between phenotypic variation and environmental

variation started with Woltereck (1909), in which the concept of "reaction norm" (*reaktionsnorm*). Then, Nilsson-Ehle (1914) brings the concept of 'acclimatization or adjustment of plants,' which was the first step surrounding the idea of 'phenotypic plasticity (*plasticitet*) over a certain environmental gradient. Finally, Hogben (1932) recognized a certain 'hereditary source' related to genetic variation at environmental-specific effects – which started to be called genotype × environment interaction.

From these basic studies, some plant breeders started to pay attention to the implications of G×E in selection decisions. Diverse biometric methods were proposed, but perhaps the most popular came up from Finlay-Wilkinson's (1963) regression-based model for analyzing the adaptation of the cultivars. Four years later, Eberhart and Russell (1966) also implemented a regression of the phenotype variation over the mean-centered yield values as an environmental quality index for running yield adaptability and stability analysis. At that time, the authors suggested that: "…an index independent of the experimental varieties and obtained from environmental factors would be desirable. Unfortunately, our present knowledge of the relationship between these factors and yield does not permit the computation of such an index. Until we can measure such factors to formulate a mathematical relation with yield, the average yield of the varieties in a particular environment must suffice" (Eberhart and Russell 1966, p.37).

Thus, at that time, the lack of reliable environmental data and the knowledge gap in ecophysiology modeling have hampered the implementation of an 'actual' reaction-norm modeling. However, this is not the reality nowadays, with the available remote sensing tools, mechanistic crop growth models, and less expensive costs for environmental sensors. So why most plant breeders still do not use environmental data on MET analysis? Below we show some historical efforts after the Eberhart and Russell (1966) suggestions and how these concepts are implemented under the predictive breeding context. Another important issue is the integrative use of CGM for environmental characterizations, which is also discussed further in this review.

Biological interpretation of the environmental and genetic determinants of G×E

The decade of 1970 started with advances in the crop growth modeling approaches, since the introduction of the called 'School of de Witt' (Bouman et al. 1996). On the other hand, reaction-norm models were introduced to recover some explained G×E patterns using environmental data (Freeman and Perkins 1971, Wood 1976). The first devoted efforts to understanding the mechanistic process of plant-environment interaction: how the resources are captured and allocated into the different plant organs across time (phenological development) to produce biomass and converted it into agronomic yield. These are process-based models in which certain steps must be trained for each cultivar, species, or germplasm. The efforts in CGM continue until these, while the most useful application today is to support crop systems predictions. The second approach is more related to quantitative genetics, while we aimed to recover unknown patterns within the G×E to be understood and perhaps explored as a target trait. In the 1980s, Denis (1980) introduced the concept of factorial regression (FR) that was widespread during the 1990s and 2000s (Epinat-Le Signor et al. 2001, Romay et al. 2010, Nunes et al. 2011, Oliveira et al. 2021). This implementation was mostly a result of the seminal work of Van Eeuwijk (1996) and the innovative efforts from Crossa et al. (1999) and Vargas et al. (1999). Nowadays, it is possible to use these approaches to understand yield adaptability from the main reaction-norms that shape a major impact on MET- or TPE-specific G×E (Ly et al. 2018, Millet et al. 2019, Costa-Neto et al. 2020, Porker et al. 2020). After the step of genetic mapping and QTL (quantitative trait loci) detection, for specific environments (or resulted from a multi-environment QTL mapping model), the resulted QTL×E matrix can also be dissected in terms of reaction-norms at the QTL level (Vargas et al. 2006, Malosetti et al. 2006, Bustos-Korts et al. 2019), which also can be done after association mapping studies (Li et al. 2018, Guo et al. 2020, Li et al. 2021). All of these models were first implemented using linear ordinary least squares (OLS), but nowadays, the same approaches can be made using since partial least squares models (PLS) until machine learning (ML) techniques, such as random forest (Monteverde et al. 2019, Marchal et al. 2019, Porker et al. 2020).

Adding value to the classical predictive breeding platforms

Heslot et al. (2014) introduced a whole-genomic version of the FR using *W* covariates, bringing envirotyping-based outputs to the predictive breeding scenario. This model includes stress covariates derived from running crop growth models (CGM) with raw environmental, that is, resulting in covariables with a reasonable biological meaning. This approach

was expanded by Ly et al. (2017) and Ly et al. (2018), in which in the first study, the CGM was used to characterize G×E while in the second the genotype-specific reaction-norms. It seems that 2014 was an exceptional year for envirotyping because a few months later, Jarquin et al. (2014) introduced the method of building environmental relatedness kernels from W covariates. Unlike the classical FR-based reaction norm models, the goal of this method is to consider a nongenetic source of variation build up with realizations from the envirotyping outputs.

Similar to the classical GBLUP, in which molecular markers shape a genomic relationship matrix (**G**), the environmental relatedness must also be created using the linear covariance among W covariates (**W**). In addition, the authors showed a more intuitive approach to model G×E, in which now the Hadamard product (#) between genomics, $\mathbf{GW} \sim \mathbf{Z}_g \mathbf{GZ}_g^{\mathrm{T}} \,\#\, \mathbf{Z}_e \mathbf{WZ}_e^{\mathrm{T}},$ and what was called later as 'enviromic-based matrix' (Costa-Neto et al. 2021a, Costa-Neto et al. 2021c) is the core of putative reaction-norms driving the expected G×E (Jarquin et al. 2014). This approach was expanded by Morais-Júnior et al. (2018), which involved a wide number of W covariates and another genetic relatedness, both from genomics and pedigree-based records. Then, Costa-Neto et al. (2021b) introduced nonlinear kernels for modeling the **W** matrix and the strategy of computing the predictive ability of each genotype as a sign of the 'model resolution in predict complex G×E'. At this proof-of-concept study, it seems that nonlinear effects better describe the environmental relatedness and improve the model's ability to predict novel G×E. Moreover, with the advent of the EnvRtype software (Costa-Neto et al. 2021c), diverse kernel structures might also be built with W covariates, such as developmental stage-specific W kernels. More detail about enviromics-aided GP models is given in the next section.

Finally, yet in 2014, Cooper et al. (2014) introduced the terminology "envirotyping" (environmental + typing) to the plant breeding community, which was well established only two years later by Xu (2016). The use of "environmental characterizations" with envirotyping and aided by CGM is used at least since the 2000s (Malosseti et al. 2006, Chenu et al. 2011). However, Cooper et al. (2016), followed by Messina et al (2018) and Toda et al (2020), introduced a tan integrative approach that brings the concepts of CGM with the genomic prediction (adapted here as CGM-GP). The potential advantage of this approach is the phenotypic plasticity is better modeled using the mechanistic CGM structure. First, the genotype-specific parameters are trained for a given population, in which the whole-genome regressions are then used to predict those parameters. After an optimization process, the most fitted genotype-specific parameters are used as inputs in the CGM, which takes the raw environmental data to simulate the growing conditions that each genotype must face. Using historical weather and soil data, it is possible to predict the crop's performance across several years and locations (Li et al. 2013, Heinemann et al. 2015, Heinemann et al. 2019, Robert et al. 2020, Antolin et al. 2021), as same for using future climate scenarios (Ramirez-Villegas et al. 2018). This also is related to a useful property of CGM: the design of ideotypes. In this field, the genetic-specific coefficients are optimized for simulated growing scenarios. This approach can be leveraged by marker-assisted procedures (Gu et al. 2014).

ENVIROMIC-AIDED GENOMIC PREDICTION OF G×E

Genomic prediction (GP, Meuwissen et al. 2001) is the most used and powerful predictive breeding tool. It relies on Fisher's Infinitesimal model, in which the sum of whole-genome markers might be a realization of the genetic variation within a given population. Then, the selection made up in silico after training GP models for some traits and populations is also referred to as genomic selection. The GP platforms were first designed to model the genotype-to-phenotype relations (G-to-P) under single environment conditions, e.g., in a breeding program nursery (Lorenzana and Bernardo 2009, Windhausen et al. 2012). Under these conditions, the micro-environmental variations within breeding trials (e.g., spatial gradients in soil properties) are minimized in the phenotypic correction step by separating useful genetic patterns and experimental noises (nongenetic patterns). Thus, it is reasonable to assume that the realized G-to-P relations might capture a large part of the observed phenotypic variation.

But the efforts of GP at MET only started in 2012 with the marker environment models (Burgueño et al. 2012, Schulz-Streeck et al. 2013). In this scenario, we must consider that all phenotypic records carry the indissoluble effects of macro-environmental fluctuations of certain weather and soil factors during crop growth and development (Costa-Neto et al. 2021a, Costa-Neto et al. 2021c). Because of that, every phenotypic record carries an inner covariance with its growing environment, in which this phenomenon is visualized when the same model is trained for the same germplasm evaluated under different growing conditions. Thus, for multi-environment GP models, the effect of genomic-environment interaction is then maximized. Conversely, if it does not account for it, it results in a lack of predictive ability and an increase of residual variation.

To evolve the current GP platforms for the next level, which includes increasing the accuracy and resolution of predicting particular genotypes for complex future scenarios (e.g., climate change), we envisage that the use of explicit covariates in a wise manner must also evolve. There are three ways to implement it: 1) the classical reaction-norms to consider whole-genome regressions with key-environmental factors, and also hierarchical trait-by-trait interactions (Li et al. 2018, Ly et al. 2018, Millet et al. 2019, Guo et al. 2020, Jarquín et al. 2020); 2) get a better understanding of crops envirome to formulate a mathematical description of the environmental relatedness (Jarquín et al. 2014, Morais-Júnior et al. 2018, Monteverde et al. 2019, de los Campos et al. 2020, Costa-Neto et al. 2021a, Costa-Neto et al. 2021b, Costa-Neto et al. 2021c, Rogers et al. 2021), in which; 3) integrate different prediction approaches, such as crop growth models in the GP platforms (Cooper et al. 2016, Messina et al. 2018, Toda et al. 2020, Robert et al. 2020), in which enviromics is an output of the environmental factors and mechanistic process shaped by genotype-specific parameters. To rethinking the idea of MET GP, perhaps a generalized enviromic-genomic prediction model for modeling the phenotypic variation might be written as:

$$y = 1\mu + X\beta + \sum_{s=1}^{k} g_s + \sum_{s=1}^{k} gE_s + \sum_{r=1}^{l} W_r + \sum_{s=1}^{k} \sum_{r=1}^{k} gW_{sr} + \varepsilon$$

where: $\mu + X\beta$ is the core of fixed effects (intercept + other possible fixed effects); ε is the residual variation of all sources not considered in the model; $\sum_{s=1}^{k} g_s$ is the sum of all k genetic effects, in which might be SNP markers (e.g., coded for additive or dominance effects), a genomic-based relationship kernel (e.g., the **G** matrix for the genomic-enabled realizations of the individuals relatedness) or independent genetic effects (similar to the classical FR model); $\sum_{s=1}^{k} gE_s$ is a the sum of all genetic by environment effects, in which usually is a block-diagonal matrix of the genomic effects or any other structured matrix; $\sum_{r=1}^{l} W_r$ sum of all enviromic-based effects, in which might be W covariates (e.g., air temperature, rainfall, solar radiation) or W-based kernels (the W matrix of enviromic-realized environmental relatedness); $\sum_{s=1}^{k} \sum_{r=1}^{k} gW_{sr}$ is the sum of all interactions between genetic and enviromic effects, in which might be realization of SNP markers by W covariates, Hadamard products between covariance structures for genetic effects (e.g., additive, dominance, epistasis genomic relationships) and enviromic-based effects. This last can also be implemented by projecting the genomic effects over the enviromics by using the Kronecker product ($\otimes$) between enviromics and genomics, that is, $gW \sim N(0, W \otimes G \, \sigma_{gW}^2)$, but only if the enviromics is a matrix of $q \times q$ environments and genomics is a matrix of $p \times p$ genotypes (resulting then into a $n \times n$ dimension kernel for the phenotypic records, where $n = pq$). Otherwise, the direct Haddamard product will be more adequate. In table 1, we summarize some of the possible arrangements of this generalized model. However, if this approach is conducted under the CGM-GP approach (Cooper et al. 2016, Messina et al. 2018), the generalized model is reframed to integrate genomic effects as a predictor of the genotype-specific sensitivities to be inputted in the CGM machinery.

Table 1. Possible model's assumptions derived from the generic genetics (or genomics) with enviromics approach

Model	Effect		
	Genetic (G)	Environment (E)	G×E
Main Genotypic Effects (MM)	$\sum_{s=1}^{p} g_s \neq 0$	$\sum_{r=1}^{l} W_r = 0$	$\sum_{s=1}^{k} gE_s = 0; \sum_{s=1}^{k} \sum_{r=1}^{l} gW_{sr} = 0$
MM + G×E deviation (MDs)	$\sum_{s=1}^{p} g_s \neq 0$	$\sum_{r=1}^{l} W_r = 0$	$\sum_{s=1}^{k} gE_s \neq 0; \sum_{s=1}^{k} \sum_{r=1}^{l} gW_{sr} = 0$
MM + Enviromic effects (EMM)	$\sum_{s=1}^{p} g_s \neq 0$	$\sum_{r=1}^{l} W_r \neq 0$	$\sum_{s=1}^{k} gE_s = 0; \sum_{s=1}^{k} \sum_{r=1}^{l} gW_{sr} = 0$
EMM + G×E deviation (EMDs)	$\sum_{s=1}^{p} g_s \neq 0$	$\sum_{r=1}^{l} W_r \neq 0$	$\sum_{s=1}^{k} gE_s \neq 0; \sum_{s=1}^{k} \sum_{r=1}^{l} gW_{sr} = 0$
EMM + G×W reaction-norm	$\sum_{s=1}^{p} g_s \neq 0$	$\sum_{r=1}^{l} W_r \neq 0$	$\sum_{s=1}^{k} gE_s = 0; \sum_{s=1}^{k} \sum_{r=1}^{l} gW_{sr} \neq 0$
EMDs + G×W reaction-norm	$\sum_{s=1}^{p} g_s \neq 0$	$\sum_{r=1}^{l} W_r \neq 0$	$\sum_{s=1}^{k} gE_s \neq 0; \sum_{s=1}^{k} \sum_{r=1}^{l} gW_{sr} \neq 0$

ENVIROMICS IS LIKE A CHOCOLATE BOX: YOU NEVER KNOW WHAT YOU GONNA GET

This section is devoted to highlighting some suggestions of practical uses of enviromics and envirotyping analytics in plant breeding programs beyond the predictive tools already discussed. But before explain its practical benefits, it is reasonable to discuss the motivation behind incorporating environmental information in plant breeding. In an adaptation

of the famous quote of the character lived by Tom Hanks in Forrest Gump (1994), it seems that "…enviromics is like a chocolate box: you never know what you gonna get". This can lead to several interpretations. The first and most obvious is: you don't know the "taste of enviromics" unless you prove it. Second, suppose the box has "different shapes of chocolates." In that case, that is, different applications and uses of enviromics, perhaps you must prove it and see which type of "enviromic-based practice" most fits your reality and stage of the breeding program. Finally, the exemplified quote also means that there are no black boxes to be opened for implementing envirotyping analytics and enviromics-based predictive tools. If you open a chocolate box, it is expected to have chocolate inside. Thus, the use of enviromic is like an improvement of what has already been done by plant breeders, providing novel options and insights to support the decision making and understand the adaptation of the developed germplasm.

Suggestion 1 (Basic step): check out the sources for collecting reliable environmental data

The use of environmental data is a long-term approach neglected by most breeders, despite diverse efforts to call attention to its benefits (Cooper et al. 2014, Xu 2016, Costa-Neto et al. 2021c, Crossa et al. 2021 and so many others). Perhaps it was because of: 1) lack of equipment; 2) data availability; 3) not consider the importance of the environment; 4) a mainstream creed that by modeling only genetics would be enough; 5) communication gaps between biometrics and crop ecophysiologists. There are at least two main strategies to collect raw data that can be used for envirotyping analytics. The first and most obvious is the data sources derived from environmental sensors installed in the field, that is, sensors of soil water contents, sampled fertilizer status of the soil profile, structural properties of the soils (e.g., texture), weather stations, and scored ranks of phenotyped pests and diseases severity, in which the average value of those scores in the certain environment might be an indicator of the environmental sanity. The limitation of this source is the costs for acquiring the remote sensors, which might be expensive for some public breeding programs in certain regions of the world. In addition, if every field trial does not have the same core of sensors, there is a bias of using different equipment that must be considered. Another important factor is that some equipment has the data resolution of minutes, hours, or days and it must be standardized for all experimental networks. However, this might be worth the investment because it might be one of the most reliable data sources describing environmental conditions. The limitations of implementation costs, data availability, or logistics issues are overcome by the second source of data: remote sensing-based tools and public databases, such as SoilGrids (https://soilgrids.org/), WorldClim (https://www.worldclim.org/) and NASA POWER (https://power.larc.nasa.gov/). These two latter can be easily implemented by the EnvRtype software (Costa-Neto et al 2021c). The main limitation of this technique is the physical and temporal resolution, which varies depending on the data source used and are mostly given at a daily or monthly scale.

Suggestion 2. Integration of envirotyping analytics for breeding diagnosis using past or future data

Historical field trial data provides a reliable source to understand the G×E for a certain trait and TPE. In this context, the investigation of the G×E is used as a diagnosis of breeding's program efficiency in delivering adapted cultivars to attempt the consumer's needs (Chenu et al. 2011, Chenu et al. 2018, Heinemann et al. 2019). It can be done, for instance, by running reaction-norm models across the historical yield trials and then checking the genetic progress of reaction-norm. It also provides the data sources for understanding the static and nonstatic environmental patterns for a given region. A second possibility is to use CGM to first process outputs from the raw environmental data. For instance, Heinemann et al. (2019) used historical predictions enabled by CGM to simulate the impact of direct selection for grain yield in the upland rice tolerance for drought-stress in Central Brazil. Another use is given by Cooper et al. (2020), where genotype-specific parameters (GSP) of CGM are evaluated to diagnose the germplasm's selection responses. Finally, CGM using simulated raw data (e.g., future scenarios) provides the foundations for defining plant ideotypes that can be screened (Chenu et al. 2018) also using marker-assisted tools (Gu et al. 2014). Thus, after founding some significant QTL or genes related to the certain GSP, the breeding program germplasm might be screened to select the most promising parentals capable of generating a desirable ideotype for expected growing conditions (Martre et al. 2014, Rötter et al. 2015).

Suggestion 3. Use of envirotyping networks to define crops envirome and mega-environments

Before being associated with some phenotype, at a specific MET or TPE, the raw environmental data (or its processed data for a given crop species) is a global source of information shared among breeding programs and TPEs. Think about the genome of a certain species: despite being global, every breeding program has its unique allelic pool anchored to

the species' genome. The same basic idea can be implemented: after running an envirotyping analytics of a certain TPE, considering its variations across the space (diverse regions) or time (different planting dates), the plant breeders might define mega-environments and their global TPEs. If the envirotyping analytics pass by to a second step, where the envirotyping data is associated with some trait-specific phenotype, then it's possible to see a breeding program-specific TPE (Crespo-Herrera et al. 2021). The use of remote sensing-based envirotyping analytic tools (e.g., EnvRtype) easily implements those approaches as diverse global sites might share similar environmental patterns according to their planting dates (Costa-Neto et al. 2021a, Costa-Neto et al. 2021b, Costa-Neto et al. 2021c). A next step might be implementing well-consolidated enviromes for each crop species, considering the all-possible typologies that some crop species might face. Then, the matrix of envirotype markers by the environment will be global, and each entry will be preached by the frequencies of occurrence of each marker typology (Costa-Neto et al. 2021a). Our recent study demonstrates the benefits of studying the environment as a core of typologies. Then, it is possible to compute the frequencies of occurrence of each typology as an actual environmental marker (also known as an envirotype descriptor). The use of ensemble machine learning techniques might help define the thresholds for each environmental factor, but ecophysiology knowledge about the crop and agronomic expertise must also suffice in some cases. After this, optimization algorithms might be used to select the most representative environments of the TPE (Rincent et al. 2017).

Suggestion 4. Selection of genotypes based on its reaction-norms and further genomic architecture of the G×E effects

As commonly done in classical adaptability and stability analysis, the genotype-specific sensibilities for the key environmental factors must be used to rank the most adapted genotypes. This will take the suggestion of Eberhart and Russell (1966) from almost 60 years ago. However, a genomic-wide association study (GWAS) can find another interesting approach over these genotype-specific coefficients. If these coefficients indicate the genotype-specific reaction norms, it can be treated as a trait evaluated at single environment conditions – so dealing with the environmental dimensionality in GWAS (Li et al. 2021). Thus, it facilitates a biologically informed dissection of complex traits, in which genomic regions related to reaction-norms must be identifying and explored by plant breeders. Before this work, Gage et al. (2017) conducted a similar approach over the classical Finlay-Wilkinson parameters of broad adaptation, adaptability, and stability. These authors investigated how artificial selection indirectly affects the yield plasticity at different genomic regions for eleven traits in maize. This is also a potential analytical tool useful for plant breeders because it might help explain 'what is going on with the germplasm across several years of selection over a nursery-specific growing condition. After this type of study, it is possible to rethink breeding strategies and optimize the selection of alleles that confer adaptation of the genotypes according to the breeding goals.

FREQUENT QUESTIONS IN THE FIELD

Does enviromics is used in environmental data in any manner? The answer is no. Environmental information needs to be processed, organized, and make agronomic sense in explaining the plant's growth and development biology. Otherwise, there is a huge possibility of spurious and unrealistic relationships between the environmental factor and phenotype. Take as an example the factor rainfall precipitation. The accumulated rainfall values during the growing cycle of the crop do not represent the biology of the crop at all, but the frequency of rainfall is already a more accurate indication of the environment quality for this factor.

Does enviromics is the only solution for increased accuracy on the G×E prediction? That is a common mistake that some research waves most cause in the scientific society. And as with other waves, we must take care of these affirmations. No, there are also different techniques whose uses will vary according to the MET-specific conditions and the breeder's target. For example, suppose the G×E interactions have small importance on the phenotypic variation (less than 10%, for instance). In that case, the modeling of G×E will be a concern, and the use of enviromics will just help the plant breeder to extract novel insights from its germplasm – how the germplasm responds to air temperature, radiation, etc. This might be very helpful for tracing strategies for selecting higher adapted genotypes and know the potential genetic factors underlying the adaptation for its environmental aspects. In addition, it helps the agronomic transfer after the advanced yield testing stage, optimizing then the recommendation of superior cultivars for given locations. On the other hand, if G×E is very important to the phenotypic variation (higher than 60%, for instance), it seems that envirotyping

analytic might help group environments, optimize the experimental network and provide enviromic data for increasing the resolution and accuracy of the prediction models. However, for prediction novel G×E the use of some enviromic-based strategy is still the main option.

Why sometimes the use of environmental data did not increase accuracy in predictive breeding? Perhaps the answer to this question may be related to MET-specific factors (e.g., the magnitude of G×E), genomic architecture of the trait (e.g., qualitative traits). And issues with processing environmental data (environmental data not representative of ambient conditions) or finding key development stages with a higher sensibility for the target trait under analysis. Another possibility is the misspecification of some environmental or management factors not considered in the envirotyping analytics (e.g., some differential agronomic practices across field trials) and issues in the phenotype correction step of the field trial single analysis.

Is there a unique and correct model or computational to use enviromics in G×E analysis? It seems that in terms of computational implementation, the best model is the model that solves your issues. However, it always must consider agronomic aspects and the biological reality (Hammer et al. 2016). An exception perhaps is the kernel method used for estimating environmental relatedness from *W* covariates. Some proof-of-concept studies seem that nonlinear kernel methods (Gaussian kernel and Deep kernel) are better than using the classical linear covariance matrix, that is WW^T (Jarquín et al. 2014).

CONNECTING THE DOTS: BUILDING ONE FRAMEWORK

Plant breeding is an interdisciplinary field, involving and evolving more areas every day. The main reason for that is that most revolutions happen within the boundaries of the sciences. Consequently, the amount and nature of data have exponentially increased over the years. However, in quantitative genetics, we are still targeting the same: understand the genetic basis of quantitative traits, predict performances, and optimize breeding programs.

In this context, breeders have added new layers of data, one by one in prediction models, such as with markers, image-based phenotypes, molecular phenotypes, and enviromics. The next step is to integrate two different worlds, quantitate genetics and crop-growth models. There are many attempts to combine them. Based on our experience, we propose one of the possible ways (Figure 2).

All individuals will be genotyped and phenotyped for important traits, such as grain yield in the training set, moreover, characterized for molecular phenotypes (e.g., transcriptome, proteome, etc.; Morgante et al. 2020) and image-based
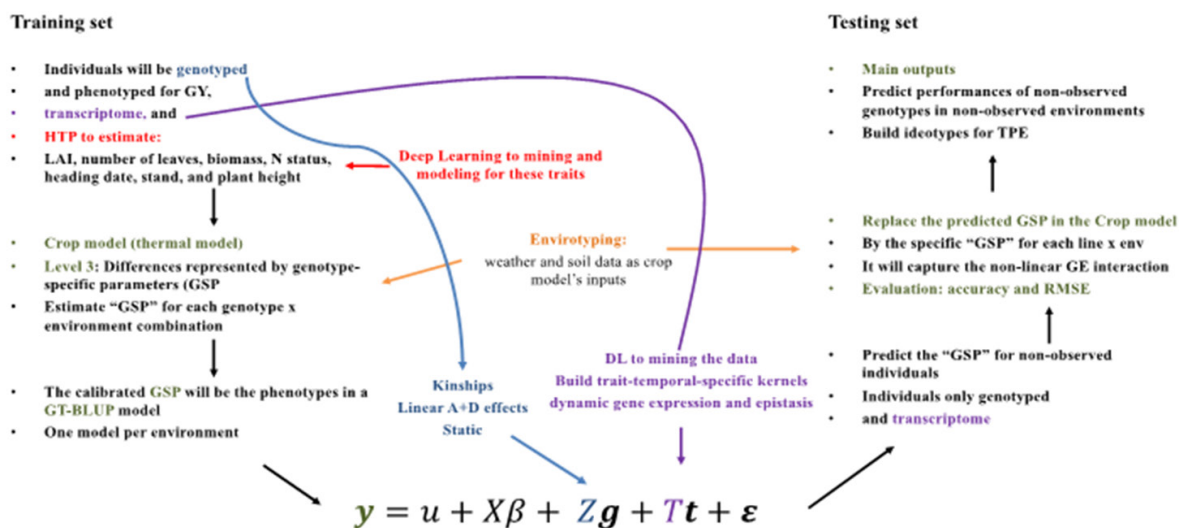

$$y = u + X\beta + Zg + Tt + \varepsilon$$

Figure 2. A proposed framework to combine different data sources and statistical approaches, with one aim, predicts performances.

phenotypes (number of leaves, Leaf index area (LAI), biomass, etc.; Li et al. 2018, Gano et al. 2021). However, for the latter, with the aid of new tools, for instance, deep learning (DL), it is possible to improve even more the accuracy for these image-based traits.

The second step is to estimate the genotype-specific parameters (GSP) in crop models (Acharya et al. 2017), leveraging the genotypic variation in these mechanistic approaches (Oliveira et al. 2021). For that, in the begging, we recommend using a radiation-based thermal model due to its simplicity. These GSPs are latent phenotypes, representing the nonlinear slopes of genotypes' reactions norms of genotypes to environmental inputs, based on putative physiological responses. For that, a good description of the weather and soil growing conditions is essential – the basics of envirotyping analytics.

Then, we can use the calibrated GSP as phenotypes in a univariate single-environment GT-BLUP model (Morgante et al. 2020). Besides the markers, it is possible to include transcriptome-based kernels to capture the dynamic gene expression over the crop cycle and epistasis. In this context, DL might play an important role in selecting the best features and optimize the molecular phenotype information (Morgante et al. 2020).

Finally, we can predict the GSP for each non-tested genotype and replacing them with the base crop model, tuning the crop model for specific genotype-environment combinations. Consequently, we can predict performances on non-observed genotypes in non-observed environments, optimize training sets in multi-environment-trials (MET), and build ideotypes per target population of environments (TPE).

ACKNOWLEDGEMENTS

REFERENCES

Acharya S, Correll M, Jones JW, Boote KJ, Alderman PD, Hu Z and Vallejos CE (2017) Reliability of genotype-specific parameter estimation for crop models: Insights from a Markov chain Monte-Carlo estimation approach. **Transactions of the ASABE 60**: 1699-1712.

Allard RW and Bradshaw AD (1964) Implications of genotype-environmental interactions in applied plant breeding. **Crop Science 4**: 503-508.

Antolin LAS, Heinemann AB and Marin FR (2021) Impact assessment of common bean availability in Brazil under climate change scenarios. **Agricultural Systems 191**: 103174.

Arnold PA, Kruuk LEB and Nicotra AB (2019) How to analyze plant phenotypic plasticity in response to a changing climate. **New Phytologist 222**: 1235-1241.

Bouman BAM, Van Keulen H, Van Laar HH and Rabbinge R (1996) The "School of de Wit" crop growth simulation models: A pedigree and historical overview. **Agricultural Systems 52**: 171-198.

Bradshaw AD (1965) Evolutionary significance of phenotypic plasticity in plants. **Advances in Genetics 13**: 115-155.

Buckley TN (2019) How do stomata respond to water status? **New Phytologist 224**: 21-36.

Burgueño J, de los Campos G, Weigel K and Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. **Crop Science 52**: 707-719.

Bustos-Korts D, Malosetti M, Chenu K, Chapman S, Boer MP, Zheng B and van Eeuwijk FA (2019) From QTLs to adaptation landscapes: using genotype-to-phenotype models to characterize G×E over time. **Frontiers in Plant Science 10**: 1540.

Castroverde CDM and Dina D (2021) Temperature regulation of plant hormone signaling during stress and development. **Journal of Experimental Botany** (doi:10.1093/jxb/erab257).

Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer MF and Chapman SC (2011) Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. **Journal of Experimental Botany 62**: 1743-1755.

Chenu K, Van Oosterom EJ, Mclean G, Deifel KS, Fletcher A, Geetika G, Tirfessa A, Mace ES, Jordan DR, Sulman R and Hammer GL (2018) Integrating modelling and phenotyping approaches to identify and screen complex traits – Illustration for transpiration efficiency in cereals. **Journal of Experimental Botany 69**: 3181-3194.

Cimen E, Jensen SE and Buckler ES (2021) Building a tRNA thermometer

to estimate microbial adaptation to temperature. **Nucleic Acids Research 48**: 12004-12015.

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A and Hausmann NJ (2014) Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. **Crop and Pasture Science 65:** 311-336.

Cooper M, Powell O, Voss-Fels KP, Messina CD, Gho C, Podlich DW, Technow F, Chapman SC, Beveridge CA, Ortiz-Barrientos D and Hammer GL (2020) Modelling selection response in plant breeding programs using crop models as mechanistic gene-to-phenotype (CGM-G2P) multi-trait link functions. *in silico* **Plants 3**: diaa016.

Cooper M, Technow F, Messina C, Gho C and Radu Totir L (2016) Use of crop growth models with whole-genome prediction: Application to a maize multienvironment trial. **Crop Science 56**: 2141-2156.

Costa-Neto G, Crossa J and Fritsche-Neto R (2021a) Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity. https://www.biorxiv.org/content/10.1101/2021.06.04.447091v2

Costa-Neto G, Fritsche-Neto R and Crossa J (2021b) Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. **Heredity 126**: 92-106.

Costa-Neto G, Galli G, Carvalho HF, Crossa J and Fritsche-Neto R (2021c) EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. **G3: Genes, Genomes, Genetics 11**: jkab040.

Costa-Neto GMF, Morais Júnior OP, Heinemann AB, de Castro AP and Duarte JB (2020) A novel GIS-based tool to reveal spatial trends in reaction norm: upland rice case study. **Euphytica 216**: 37.

Crespo-Herrera LA, Crossa J, Huerta-Espino J, Mondal S, Velu G, Juliana P, Vargas M, Pérez-Rodríguez P, Joshi AK, Braun HJ and Singh RP (2021) Target population of environments for wheat breeding in India: Definition, prediction and genetic gains. **Frontiers in Plant Science 12**: 638520.

Crossa J, Fritsche-Neto R, Montesinos-lopez OA, Costa-Neto G, Dreisigacker S, Montesinos-lopez A and Bentley AR (2021) The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. **Frontiers in Plant Science 12**: 651480.

Crossa J, Vargas M, Van Eeuwijk FA, Jiang C, Edmeades GO and Hoisington D (1999) Interpreting genotype x environment interaction in tropical maize using linked molecular markers and environmental covariables. **Theoretical and Applied Genetics 99**: 611-625.

de los Campos G, Pérez-Rodríguez P, Bogard M, Gouache D and Crossa J (2020) A data-driven simulation platform to predict cultivars' performances under uncertain weather conditions. **Nature Communications 11**: 4876.

Denis JB (1980) Analyse de régression factorielle. **Biometrics 20**: 1-34.

Eberhart SA and Russell WA (1966) Stability parameters for comparing varieties. **Crop Science 6**: 36-40.

Epinat-Le Signor C, Dousse S, Lorgeou J, Denis J-B, Bonhomme R, Carolo P and Charcosset A (2001) Interpretation of genotype x environment interactions for early maize hybrids over 12 years. **Crop Science 41**: 663-669.

Finlay KW and Wilkinson GN (1963) The analysis of adaptation in a plant breeding programme. **Journal of Agricultural Research 14**: 742-754.

Fisher RA (1924) Studies in crop variation III - The influence of rainfall on the yield of wheat at Rothamsted. **Philosophical Transactions of the Royal Society of London 213**: 89-142.

Freeman GH and Perkins JM (1971) Environmental and genotype–environmental components of variability: VIII - Relations between genotypes grown in different environments and measures of these environments. **Heredity 27**: 15-23.

Gage JL, Jarquin D, Romay C, Lorenz A, Buckler ES, Kaeppler S, Alkhalifah N, Bohn M, Campbell DA, Edwards J, Ertl D, Flint-Garcia S, Gardiner J, Good B, Hirsch CN, Holland J, Hooker DC, Knoll J, Kolkman J, Kruger G, Lauter N, Lawrence-Dill CJ, Lee E, Lynch J, Murray SC, Nelson R, Petzoldt J, Rocheford T, Schnable J, Schnable PS, Scully B, Smith M, Springer NM, Srinivasan S, Walton R, Weldekidan T, Wisser RJ, Xu W, Yu J and De Leon N (2017) The effect of artificial selection on phenotypic plasticity in maize. **Nature Communications 8**: 1348.

Gano B, Dembele JSB, Ndour A, Luquet D, Beurier G, Diouf D and Audebert A (2021) Using uav borne, multi-spectral imaging for the field phenotyping of shoot biomass, leaf area index and height of West African sorghum varieties under two contrasted water conditions. **Agronomy 11**: 850.

Gu J, Yin X, Zhang C, Wang H and Struik PC (2014) Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (Oryza sativa) under drought stress. **Annals of Botany 114**: 499-511.

Guo T, Mu Q, Wang J, Vanous AE, Onogi A, Iwata H, Li X and Yu J (2020) Dynamic effects of interacting genes underlying rice flowering-time phenotypic plasticity and global adaptation. **Genome Research 30**: 673-683.

Hammer G, Messina C, van Oosterom E, Chapman S, Singh V, Borrell A, Jordan D and Cooper M (2016) Molecular breeding for complex adaptive traits: how integrating crop ecophysiology and modelling can enhance efficiency. In Yin X and Struik P (eds) **Crop systems biology**. Springer, Cham, p. 147-162.

Heinemann AB, Barrios-Perez C, Ramirez-Villegas J, Arango-Londoño D, Bonilla-Findji O, Medeiros JC and Jarvis A (2015) Variation and impact of drought-stress patterns across upland rice target population of environments in Brazil. **Journal of Experimental Botany 126**: 1-14.

Heinemann AB, Ramirez-Villegas J, Rebolledo MC, Costa Neto GMF and Castro AP (2019) Upland rice breeding led to increased drought sensitivity in Brazil. **Field Crops Research 231**: 57-67.

Heslot N, Akdemir D, Sorrells ME and Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment

interactions. **Theoretical and Applied Genetics 127**: 463-480.

Hogben L (1932) **Genetic principles in medicine and social science**. Alfred A. Knopf, New York, 230p.

Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Pérez P, Calus M, Burgueno J and de los Campos G (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. **Theoretical and Applied Genetics 127**: 595-607.

Jarquín D, Kajiya-Kanegae H, Taishen C, Yabe S, Persa R, Yu J, Nakagawa H, Yamasaki M and Iwata H (2020) Coupling day length data and genomic prediction tools for predicting time-related traits under complex scenarios. **Scientific Reports 10**: 13382.

Jończyk M, Sobkowiak A, Trzcinska-Danielewicz J, Skoneczny M, Solecka D, Fronk J and Sowiński P (2017) Global analysis of gene expression in maize leaves treated with low temperature. II. Combined effect of severe cold (8 °C) and circadian rhythm. **Plant Molecular Biology 95**: 279-302.

Li J, Shi Y, Veeranampalayam-Sivakumar AN and Schachtman DP (2018) Elucidating sorghum biomass, nitrogen and chlorophyll contents with spectral and morphological traits derived from unmanned aircraft system. **Frontiers in Plant Science 9**: 1406.

Li T, Raman AK, Marcaida M, Kumar A, Angeles O and Radanielson AM (2013) Simulation of genotype performances across a larger number of environments for rice breeding using ORYZA2000. **Field Crops Research 149**: 312-321.

Li X, Guo T, Mu Q, Li Xianran and Yu J (2018) Genomic and environmental determinants and their interplay underlying phenotypic plasticity. **PNAS 115**: 6679-6684.

Li X, Guo T, Wang J, Bekele WA, Sukumaran S, Vanous AE, McNellie JP, Cortes LT, Lopes MS, Lamkey KR, Westgate ME, McKay JK, Archontoulis SV, Reynolds MP, Tinker NA, Schnable PS and Yu J (2021) An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. **Molecular Plant 14**: 874-887.

Liu S, Li C, Wang H, Wang S, Yang S, Liu X, Yan J, Li B, Beatty M, Zastrow-Hayes G, Song S and Qin F (2020) Mapping regulatory variants controlling gene expression in drought response and tolerance in maize. **Genome Biology 21**: 163.

Lorenzana RE and Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and Applied Genetics 120**: 151-161.

Ly D, Chenu K, Gauffreteau A, Rincent R, Huet S, Gouache D, Martre P, Bordes J and Charmet G (2017) Nitrogen nutrition index predicted by a crop model improves the genomic prediction of grain number for a bread wheat core collection. **Field Crops Research 214**: 331-340.

Ly D, Huet S, Gauffreteau A, Rincent R, Touzy G, Mini A, Jannink J-L, Cormier F, Paux E, Lafarge S, Le Gouis J and Charmet G (2018) Whole-genome prediction of reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.) by genomic random regression. **Field Crops Research 216**: 32-41.

Malosetti M, Visser RGF, Celis-Gamboa C and Van Eeuwijk FA (2006) QTL methodology for response curves on the basis of nonlinear mixed models, with an illustration to senescence in potato. **Theoretical and Applied Genetics 113**: 288-300.

Marchal A, Schlichting CD, Gobin R, Balandier P, Millier F, Muñoz F, Pâques LE and Sánchez L (2019) Deciphering hybrid larch reaction norms using random regression. **G3: Genes, Genomes, Genetics 9**: 21-32.

Martre P, Quilot-Turion B, Luquet D, Memmah MMOS, Chenu K and Debaeke P (2014) Model-assisted phenotyping and ideotype design. **Crop Physiology: Applications for Genetic Improvement and Agronomy 2:** 349-373.

Messina CD, Technow F, Tang T, Totir R, Gho C and Cooper M (2018) Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP). **European Journal of Agronomy 100**: 151-162.

Meuwissen THE, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics 157**: 1819-1829.

Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, Charcosset A, Welcker C, van Eeuwijk F and Tardieu F (2019) Genomic prediction of maize yield across European environmental conditions. **Nature Genetics 51**: 952-956.

Monteverde E, Gutierrez L, Blanco P, Pérez de Vida F, Rosas JE, Bonnecarrère V, Quero G and McCouch S (2019) Integrating molecular markers and environmental covariates to interpret genotype by environment interaction in rice (*Oryza sativa* L.) grown in subtropical areas. **G3: Genes, Genomes, Genetics 9**: 1519-1531.

Morais Júnior OP, Duarte JB, Breseghello F, Coelho ASG and Magalhães AM (2018) Single-step reaction norm models for genomic prediction in multienvironment recurrent selection trials. **Crop Science 58**: 592-607.

Morgante F, Huang W, Sørensen P, Maltecca C and Mackay TFC (2020) Leveraging multiple layers of data to predict drosophila complex traits. **G3: Genes, Genomes, Genetics 10**: 4599-4613.

Nilsson-Ehle H (1914) Vilka erfarenheter hava hittills vunnits rorande mojligheten av vaxters acklimatisering. **Kgl Landtbruks-Akad Handl Tidskr 63**: 537-572.

Nunes GHS, Andrade Neto RC, Costa Filho JH and Melo SB (2011) Influência de variáveis ambientais sobre a interação genótipos x ambientes em meloeiro. **Revista Brasileira de Fruticultura 33**: 1194-1199.

Oliveira FAA, Jones JW, Pavan W, Bhakta M, Vallejos CE, Correll MJ, Boote KJ, Fernandes JMC, Hölbig CA and Hoogenboom G (2021) Incorporating a dynamic gene-based process module into a crop simulation model. *in silico* **Plants 3**: diab011.

Patel D and Franklin KA (2009) Temperature-regulation of plant architecture. **Plant Signaling and Behavior 4**: 577-579.

Porker K, Coventry S, Fettell NA, Cozzolino D and Eglinton J (2020) Using

a novel PLS approach for envirotyping of barley phenology and adaptation. **Field Crops Research 246**: 1-11.

Ramirez-Villegas J, Heinemann AB, Castro AP, Breseghello F, Navarro-Racines C, Li T, Rebolledo MC and Challinor AJ (2018) Breeding implications of drought stress under future climate for upland rice in Brazil. **Global Change Biology 24**: 2035-2050.

Resende RT, Piepho HP, Rosa GJM, Silva-Junior OB, Silva FF, Resende MDV and Grattapaglia D (2020) Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. **Theoretical and Applied Genetics 134:** 95-112.

Rincent R, Kuhn E, Monod H, Oury FX, Rousset M, Allard V and Le Gouis J (2017) Optimization of multi-environment trials for genomic selection based on crop models. **Theoretical and Applied Genetics 130**: 1735-1752.

Robert P, Le Gouis J and Rincent R (2020) Combining crop growth modeling with trait-assisted prediction improved the prediction of genotype by environment interactions. **Frontiers in Plant Science 11**: 827.

Rogers AR, Dunne JC, Romay C, Bohn M, Buckler ES, Ciampitti IA, Edwards J, Ertl D, Flint-Garcia S, Gore MA, Graham C, Hirsch CN, Hood E, Hooker DC, Knoll J, Lee EC, Lorenz A, Lynch JP, McKay J, Moose SP, Murray SC, Nelson R, Rocheford T, Schnable JC, Schnable PS, Sekhon R, Singh M, Smith M, Springer N, Thelen K, Thomison P, Thompson A, Tuinstra M, Wallace J, Wisser RJ, Xu W, Gilmour AR, Kaeppler SM, De Leon N and Holland JB (2021) The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. **G3: Genes, Genomes, Genetics 11**: jkaa050.

Romay MC, Malvar RA, Campo L, Alvarez A, Moreno-González J, Ordás A and Revilla P (2010) Climatic and genotypic effects for grain yield in maize under stress conditions. **Crop Science 50**: 51-58.

Rötter RP, Tao F, Höhn JG and Palosuo T (2015) Use of crop simulation modelling to aid ideotype design of future cereal cultivars. **Journal of Experimental Botany 66**: 3463-3476.

Schulz-Streeck T, Ogutu JO, Gordillo A, Karaman Z, Knaak C and Piepho HP (2013) Genomic selection allowing for marker-by-environment interaction. **Plant Breeding 132**: 532-538.

Toda Y, Wakatsuki H, Aoike T, Kajiya-Kanegae H, Yamasaki M, Yoshioka T, Ebana K, Hayashi T, Nakagawa H, Hasegawa T and Iwata H (2020) Predicting biomass of rice with intermediate traits: Modeling method combining crop growth models and genomic prediction models. **PLoS One 15**: e0233951.

van Eeuwijk F, Denis J and Kang MS (1996) Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In Kang MS and Gauch Jr. HG (eds) **Genotype-by-environment interaction**. CRC Press, Boca Raton, p. 15-50.

Vargas M, Crossa J, Van Eeuwijk FA, Ramírez ME and Sayre K (1999) Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype x environment interaction. **Crop Science 39**: 955-967.

Vargas M, Van Eeuwijk FA, Crossa J and Ribaut JM (2006) Mapping QTLs and QTL x environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. **Theoretical and Applied Genetics 112**: 1009-1023.

Vendramin S, Huang J, Crisp PA, Madzima TF and McGinnis KM (2020) Epigenetic regulation of ABA-induced transcriptional responses in maize. **G3: Genes, Genomes, Genetics 10**: 1727-1743.

Voss-Fels KP, Cooper M and Hayes BJ (2019) Accelerating crop genetic gains with genomic selection. **Theoretical and Applied Genetics 132**: 669-686.

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink JL and Sorrells ME (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. **G3: Genes, Genomes, Genetics 2**: 1427-1436.

Woltereck R (1909) Weitere experimentelle Vntersuchungen iiber Artverandenmg, spcziell fiber das Wesen quantitativer Artunterschiede bei Daphniden. **Verhandlungen derDeutschen Zoologischen Gesellschaf 19**: 110-170.

Wood JT (1976) The use of environmental variables in the interpretation of genotype-environment interaction. **Heredity 37**: 1-7.

Xu Y (2016) Envirotyping for deciphering environmental impacts on crop plants. **Theoretical and Applied Genetics 129**: 653-673.